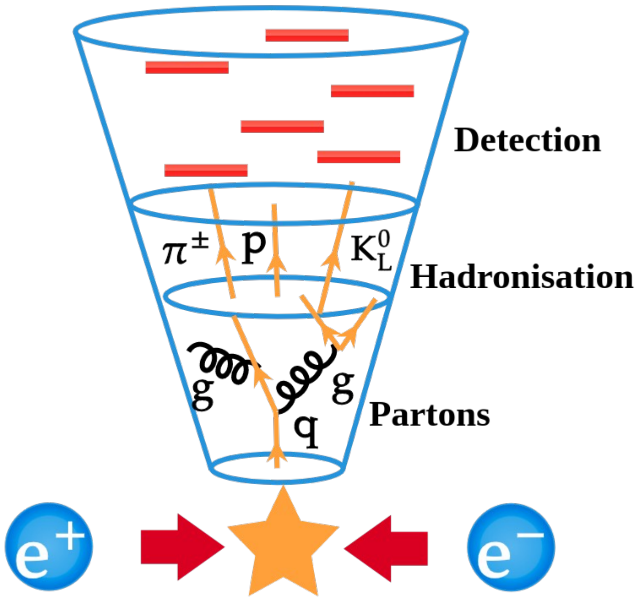
Software Compensation Using Machine Learning for AHCAL

Jack Rolph, Erika Garutti
Department of Physics, University of Hamburg
High-D Consortium Meeting

5/09/2022

Particle Flow Calorimetry

- **Future e^+-e^- linear collider experiments** (e.g International Linear Collider, ILC) require **exceptional jet-energy resolution**.
- **'Interesting' Higgs/weak sector physics** measurements **often involve decays to quarks;**
  - example: ~40% reduction of uncertainty on Higgs invariant mass measurement from Higgstrahlung by including statistics from decays to quarks for 500 fb^{-1} of integrated luminosity at ILC [1]
- However, **quarks do not exist in isolation in nature** due to colour charge conservation.
- Quarks produced in high energy lepton collisions:
  - first radiate soft gluons, which may decay into new $q\bar{q}$ pairs;
  - quarks/gluons form hadrons (hadronisation) almost instantaneously;
  - **result in relativistic, highly collimated beams of final state particles called 'jets'.**
- **Average jet composed of** [2]:
  - **charged hadrons and leptons** (~60%);
  - **photons** (~30%),
  - **neutral hadrons** (~10%)
  - **small proportion of neutrinos**



[1] Abdelhak Djouadi et al. "International Linear Collider Reference Design Report Volume 2: PHYSICS AT THE ILC". Arxiv: 0709.1893

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

Particle Flow Calorimetry

- **Jet energy resolution** at future linear colliders must produce a **invariant mass resolution comparable to the decay width of the weak vector bosons**:

$$\frac{\sigma_m}{m} = 2.7\% \approx \frac{\Gamma_W}{m_W} \approx \frac{\Gamma_Z}{m_Z}$$

→ **3.6-sigma separation** of the invariant masses of the W/Z boson [2];

- **Requires a jet energy resolution of** $\frac{\sigma_{E_{\text{jet}}}}{E_{\text{jet}}} \lesssim \frac{0.3}{\sqrt{E_{\text{jet}}}}$

- **Problem**: typical jet energy resolution of ‵traditional′ calorimetry is **much worse than required at ILC**!

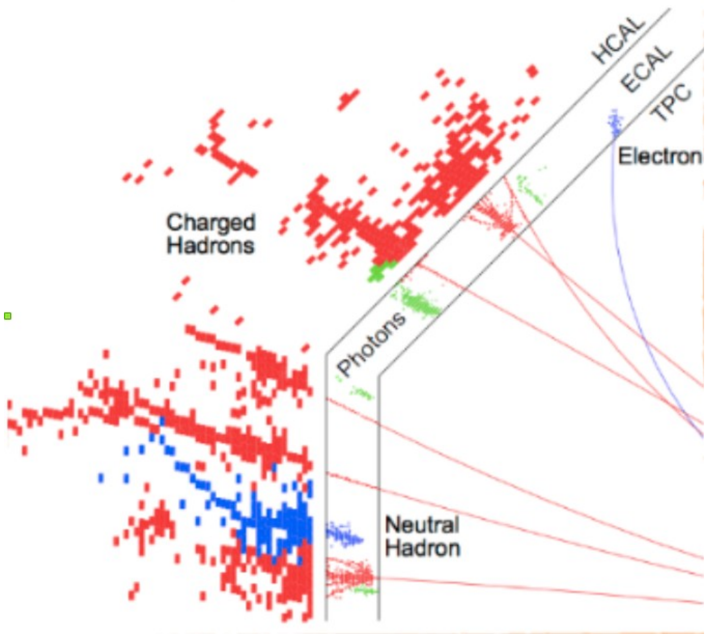
- **Solution: Particle Flow Calorimetry (PFC) [2]**:
  - measure **momentum of charged particles** (~60% of jet energy) very precisely **using tracker**;

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

  - use **highly granular calorimeters** (i.e. calorimeters with many individual readout cells) 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**;
  - **PFC jet energy resolution** depends primarily on the **accuracy of associating energy deposits to tracks**.

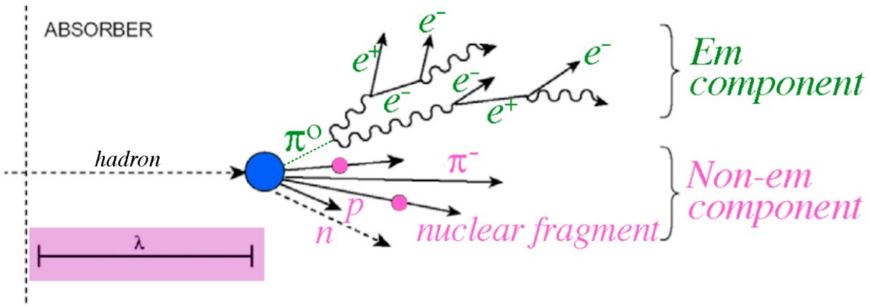


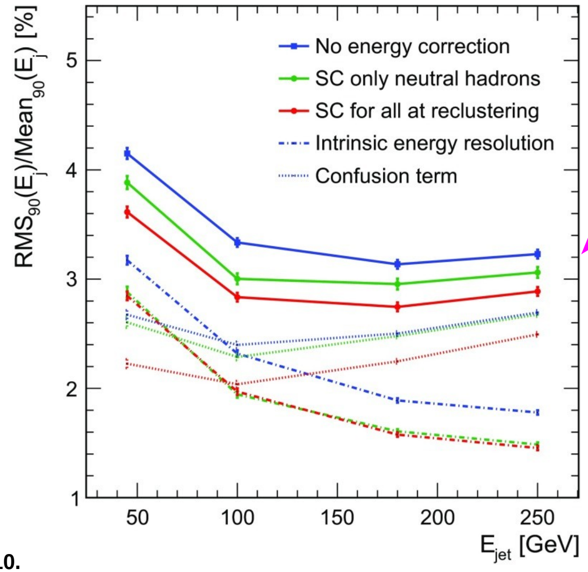
Missing Energy and Software Compensation

- When a **hadron interacts with matter**, it typically initiates **a hadron shower** with two components: an **electromagnetic (EM)** and a **hadronic (HAD) fraction**.
- **Up to 40% of the deposited energy in the HAD fraction is 'invisible'**
  - **binding energy losses from spallation of nuclei not observable;**
- **EM/HAD fraction fluctuates significantly** from event to event.
- Calorimeter response to hadrons is split into two components**:**
  - an **EM response (e)** and;
  - a **HAD response (h**). **Typically, e>h because of 'invisible' energy.**
- **Software compensation** (SC) is the **equalisation of e and h**, using:
  - information from a recorded event and;
  - a specially designed algorithm;
- **SC found to improve jet energy resolution** achieved by Pandora PFA by **enabling more accurate association of tracks to energy deposits** [3];

TAKE HOME MESSAGE:

better SC →
more accurate energy clustering for PFC→
better jet energy resolution →
more precise measurements at future linear colliders





Red line < blue line → compensation improves jet energy resolution.

[3] CALICE Collaboration, "A new approach to software compensation for the CALICE AHCAL," tech. rep., 2010.

CALICE Analogue Hadronic CALorimeter (AHCAL)

- **AHCAL** is a **Fe-Sc highly granular calorimeter** prototype **designed for Particle Flow.**
- **Highly granular ➝ 22,000 individual channels** (SiPM-on-Chip readout)
- **AHCAL is a non-compensating hadron calorimeter→ software compensation required.**
- **AHCAL is a five dimensional calorimeter**:

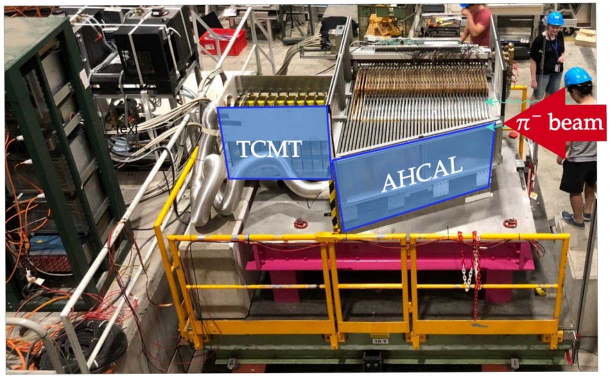
  i.e. **AHCAL measures energy density of hadron showers in space and time**, with up to 100 ps (0.1 ns) time resolution**.**

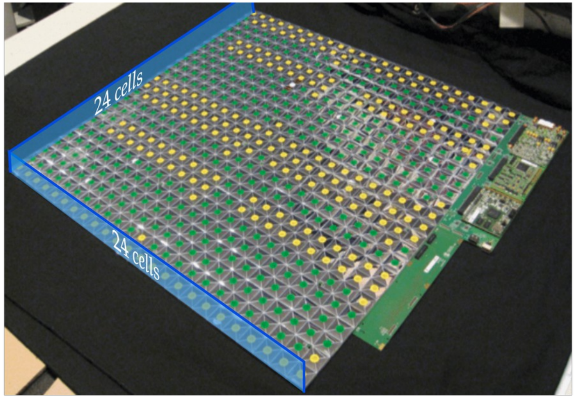
- **Highly-granular spatial energy density** is related to EM/HAD fraction**:**
  - average hadron shower has **'energy-dense' EM 'core'** and **'energy-sparse' HAD 'halo';**
  - **highly granular calorimeter can resolve shower development.**
- **Temporal energy density is** also related to EM/HAD fraction**:**
  - **Neutrons produced in abundance** during hadron shower**;**
  - **Number of neutrons proportional to hadronic fraction of event;**
  - **Indirect energy depositions from neutrons** (neutron elastic scattering/neutron capture) **delayed in steel by ~10-100 ns** from instantaneous EM component**.**
  - **timing information offers improved sensitivity to HAD fraction;**

TAKE-HOME MESSAGE:

spatial & temporal energy density information available

in a highly granular calorimeter ought to improve software compensation.

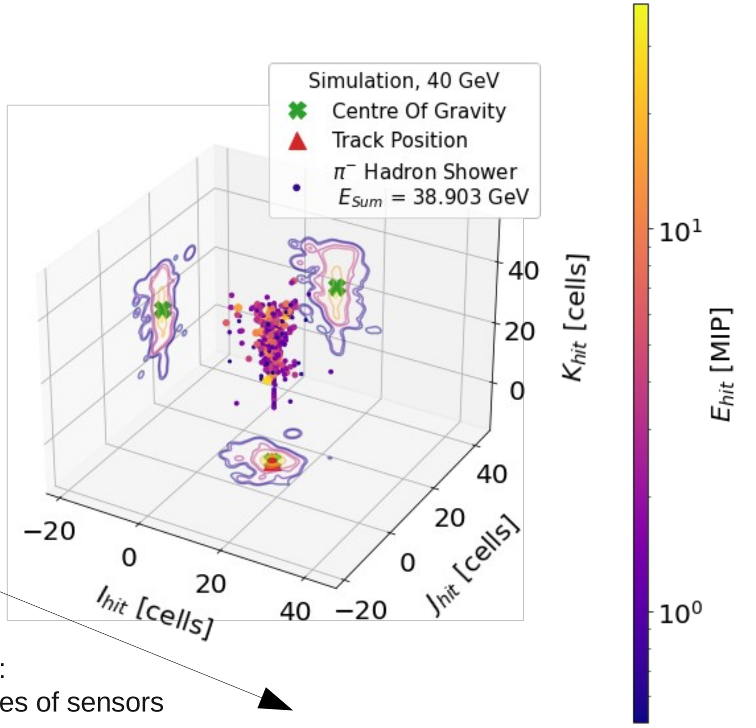




What constitutes an ‘event’ for AHCAL?”

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.

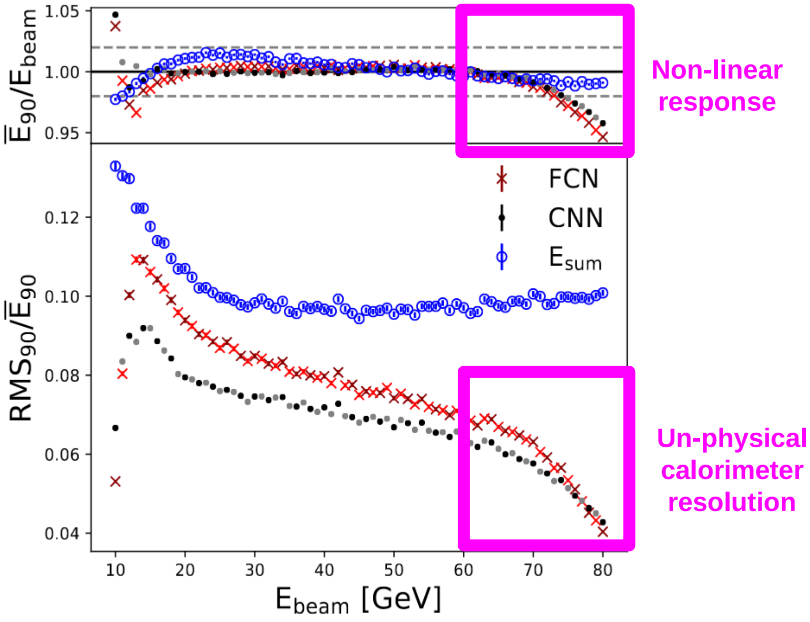
Machine Learning + Software Compensation

- **Machine learning (ML)** can be used build a **bespoke software compensation algorithm for a specific detector**.

- **Previous study** using **ML** to build an **SC algorithm for AHCAL** [4] showed **two undesirable effects**:
  - **Interpolation Failure**: biasing of model to specific particle energies;
  - **Extrapolation Failure**: biasing to the lowest and highest of the training range of energies (bottom plot);

- Why does this happen, exactly?
  - **available simulation/data** are **always biased** to **specific particle momenta**;
  - We do not know the EM/HAD fraction a priori;
  - an **assumption must be made** to train a SC algorithm:

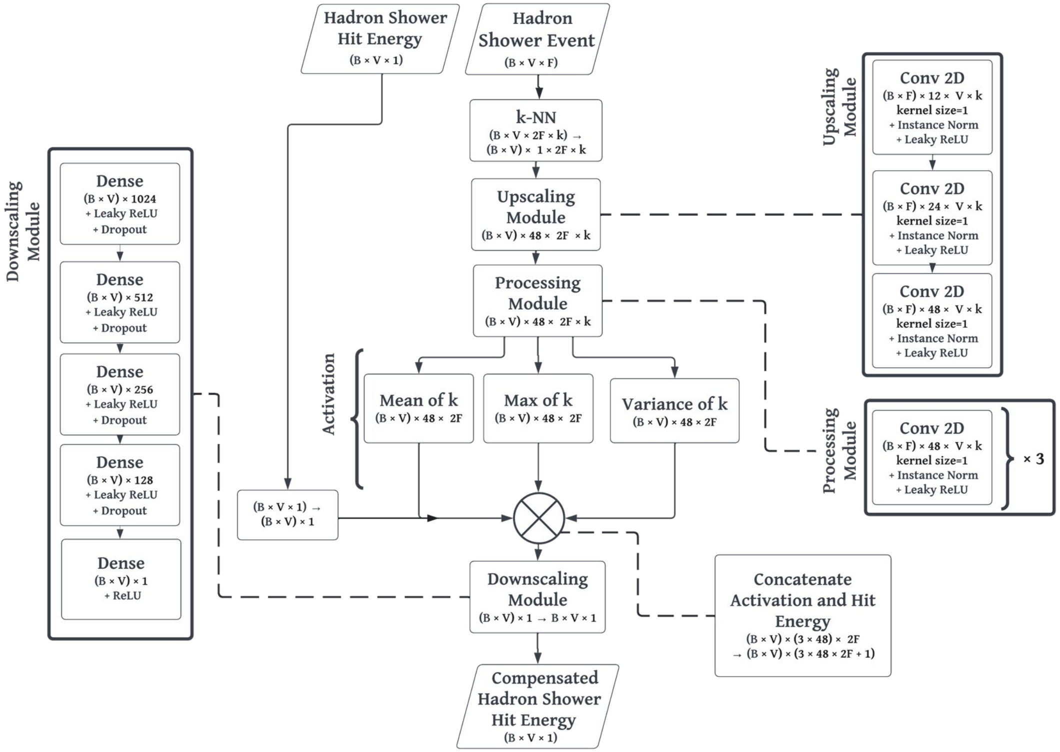
    if average difference between calorimeter response and particle momentum reduces, then e/h must have improved;
  - **Not always so ➝ bias to the training energies can be learned!**
  - Visual examples of biasing available in backup slides.



[4] E. Buhmann and E. Garutti, "Deep learning based energy reconstruction for the CALICE AHCAL"

Proposed Method

- A **machine learning SC algorithm** was implemented **using a graph neural network (GNN) technique**;
- Uses **spatial (and temporal) energy density to perform compensation**;
- **How it works:**
  - **What goes in →**
    - hadron shower event information (position, time, energy of each active sensor)
  - **What comes out →**
    - attenuated/enhanced cell energies.
  - However, **only the information from the 20 neighbouring cells** is used to compensate the **energy of each individual cell.**
  - **The sum of the 'locally' compensated cells → reconstructed energy of the hadron shower.**



TAKE-HOME MESSAGES:

- **The model cannot learn the response of the calorimeter by design;**
- **The model is necessarily focused on weighting the local spatial/temporal energy density, without imposing artificial constraints;**
- **The model should be able to interpolate/extrapolate compensation to energies not used for training.**

Control Method

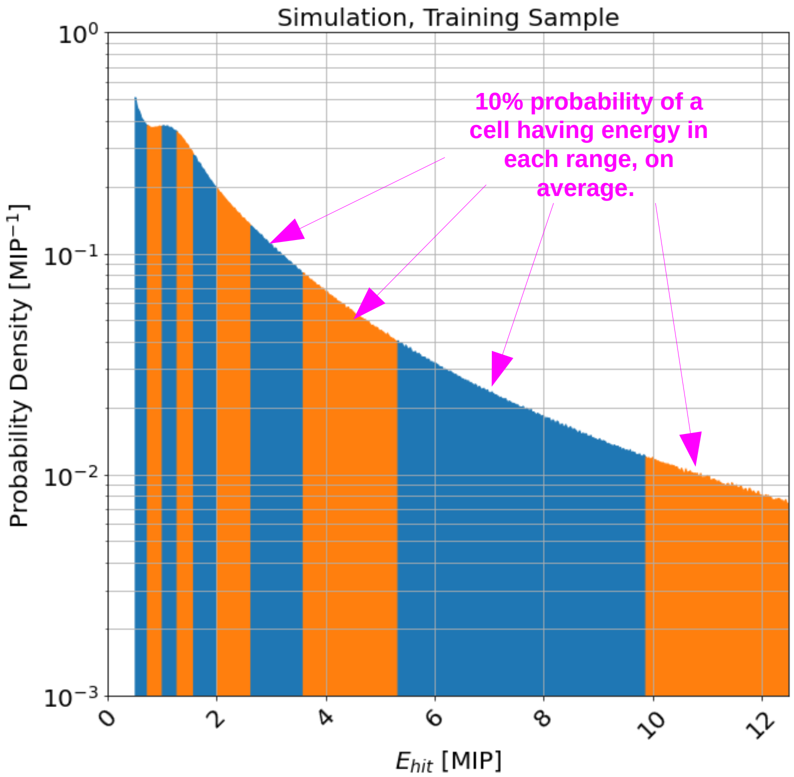
- **Standard CALICE SC method used as a control** [3];
- **Control method** is also designed to **estimate the energy density** of the hadron shower, event-by-event, **but relies on calorimeter response to do so**:
  - **Individual cell energy distribution binned in deciles** (10% probability an active cell energy will fall in any given bin, on average)
  - **Compensation weight** for each bin is calculated using a **Chebyshev polynomial function approximator as a function of calorimeter response**:

$$\omega_b(E_{\mathrm{sum}}; S, \alpha_b, \beta_b, \gamma_b) = \alpha_b + \beta_b \cdot \frac{E_{\mathrm{sum}}}{S} + \gamma_b \cdot \left(2\left(\frac{E_{\mathrm{sum}}}{S}\right)^2 - 1\right)$$

$$S = 150\mathrm{GeV} \qquad b = \mathrm{bin} \qquad \alpha_b, \beta_b, \gamma_b = \mathrm{weights}$$

  - Energy falling in each bin scaled by weight;

$$\widehat{E}_{\mathrm{sum}} = \sum_b^{\mathrm{bins}} \omega_b \cdot E_{\mathrm{sum},b}$$



- **SUMMARY**: **algorithm weights fraction of energy falling into each bin as a function of the calorimeter response to hadrons.**

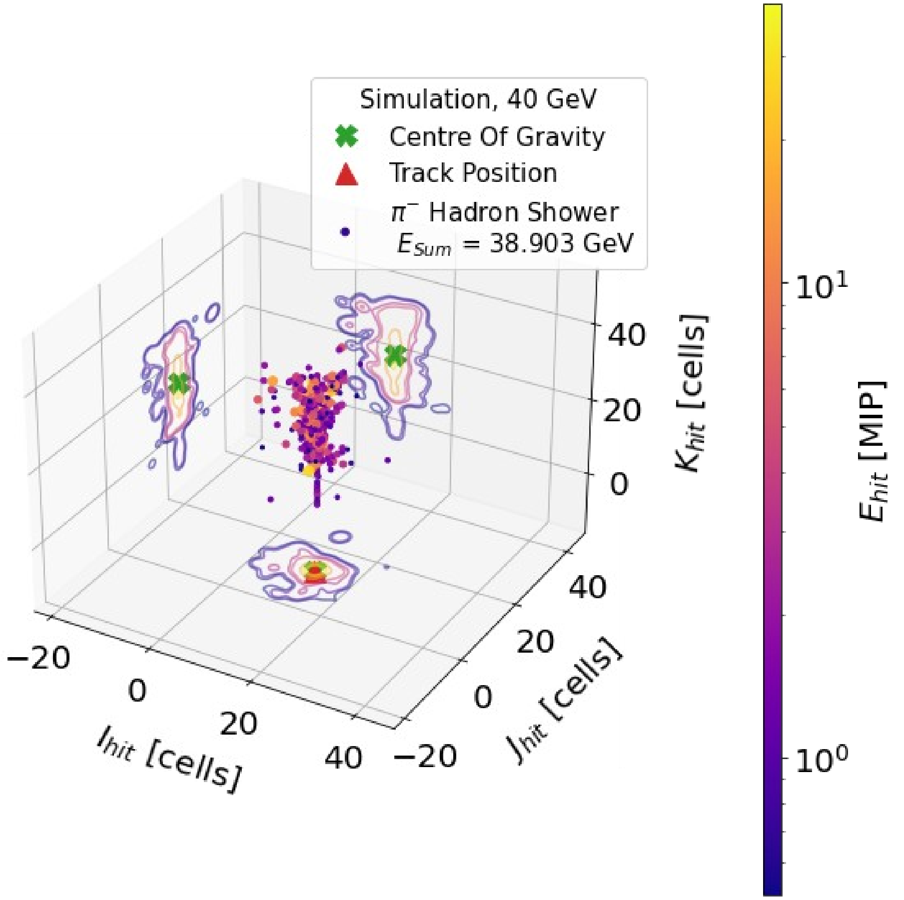
[3] CALICE Collaboration, "A new approach to software compensation for the CALICE AHCAL," tech. rep., 2010.

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Training Details

HIGH-D

- Both **network and control methods were implemented in Python/Pytorch;**
- Networks were trained **with and without timing information** to assess its influence on SC;
- **Training/testing samples simulated with Geant4** used for training control/ML methods:
  - **Training**:
    - Energy Range: 10-80 GeV
    - Binning: 10 GeV steps;
    - Events per bin: ~20,000 events
  - **Testing**:
    - Energy Range: 10-120 GeV (higher upper limit than training set)
    - Binning: 5 GeV steps (2 x # of bins than training set)
    - Cut applied using tail-catcher detector to reduce the influence of leakage;
    - **Capacity for each model to interpolate and extrapolate the compensation can be tested:**
- χ^2 **of measured energy to known particle momentum used as cost function**;
- **Many more important details are available in the backup slides.**



Results: Distributions of Reconstructed Energy

What we expect:

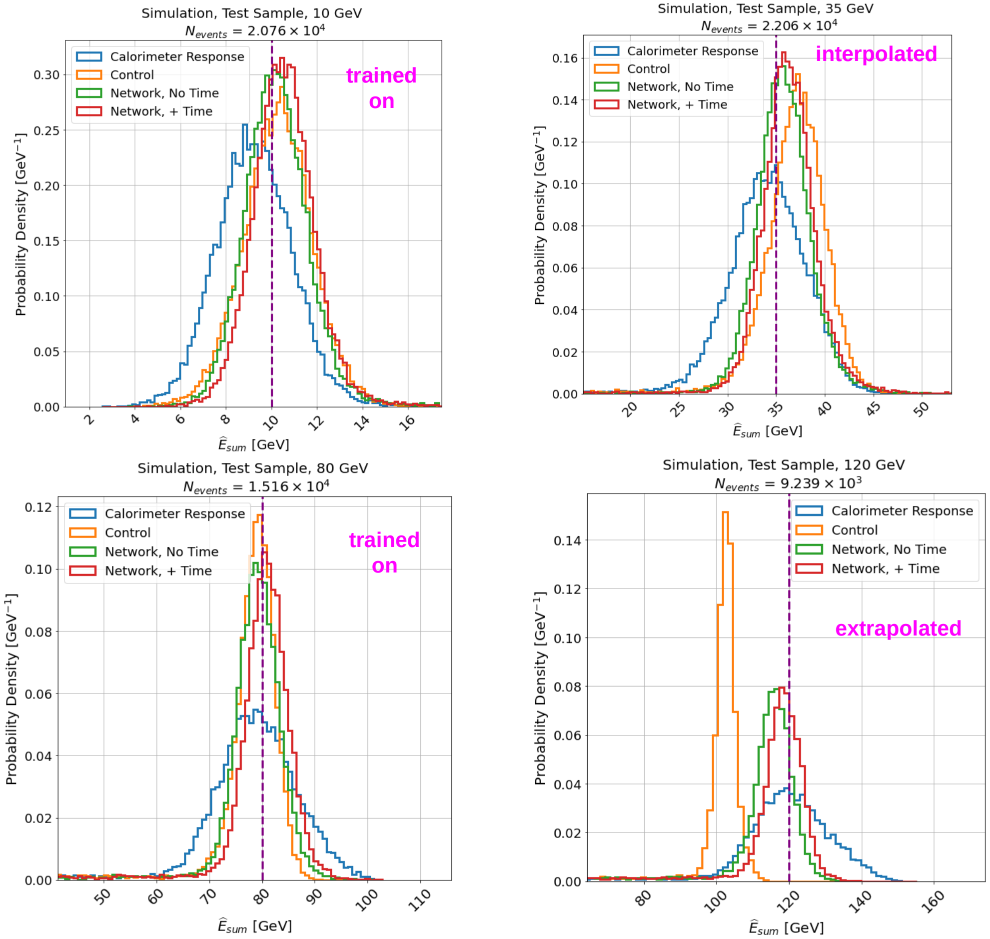
- The **resolution of the calorimeter should improve** by application of SC algorithms **➞ the width of the response distribution should reduce.**
- **The ML methods should be able to interpolate/extrapolate SC;**

What is shown:

- The **distributions of AHCAL response** to pions at different energies:
    - Intrinsic calorimeter response **(blue);**
    - control SC method applied (**orange**);
    - neural network SC method applied, no timing information **(green);**
    - neural network SC method applied, with timing information (**red).**
- **Test samples of simulation:**
    - **these are events the models have never been exposed to.**

What we learn:

- **Machine learning method** produces **superior resolution for most of training range below < 60 GeV;**
- **Superior ´compensation´ by control method in range 60-80 GeV;**
- **Above 80 GeV, control method seen to bias to training range;**
- **Neural networks able to compensate effectively above the edge of the training range;**



Results: Linearity of Response

What we expect:

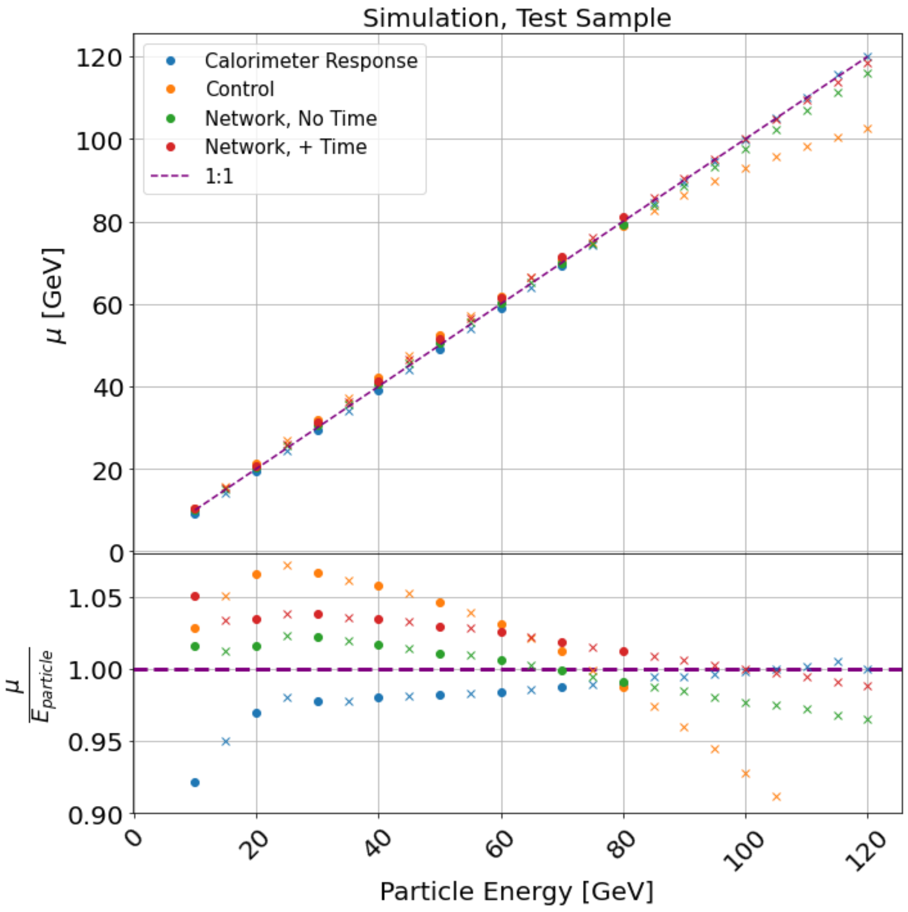
- On average, the **mean compensated shower energy should be equal to the known particle momentum** (linearity of response)

What is shown:

- **Extracted µ of two standard deviation Gauss fit** to the **energy response**, vs. **known particle momentum**, in GeV;
- Colour coding as in previous slide.

What we learn:

- **Machine learning methods** show **superior linearity of response than control method** by around **2% on average**;
- **All SC methods over-correct** the reconstructed energy;
- The **control method fails to reconstruct** the **correct response** energy **above the upper edge of the training range**, at 80 GeV;
- The **neural network methods can correctly reconstruct the energy of the hadron shower at energies above the training range** within 2%.



Results: Calorimeter Resolution

What we expect:

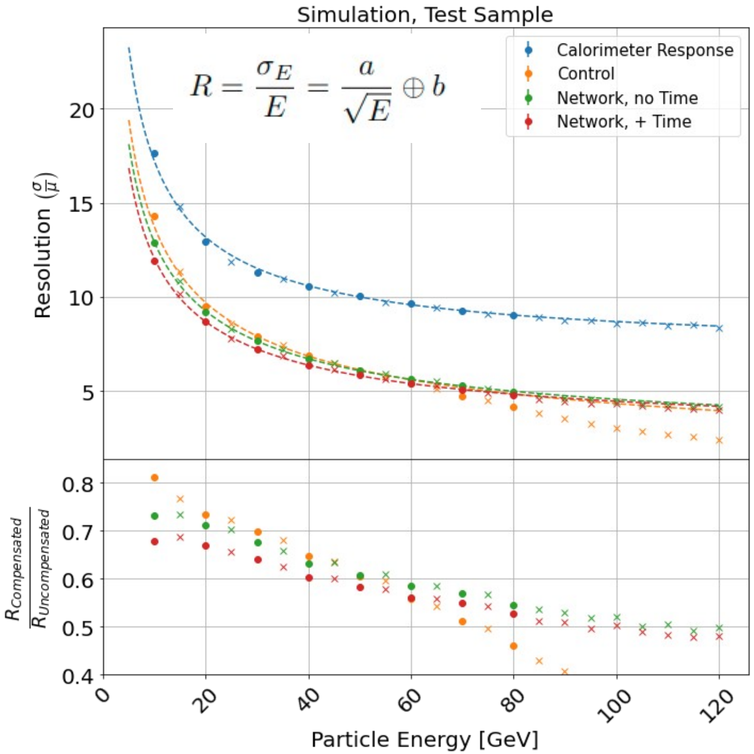
- **The energy resolution of the calorimeter should improve by applying SC;**
- Resolution function has two terms:
  - a **'stochastic term', a** → contribution of invisible energy fluctuations
  - a **'constant term', b** → calibration quality of AHCAL detector.
  - **both should reduce because of compensation.**

What is shown:

- **Extracted σ/μ** of **two standard deviation Gauss fit to response**, **vs known particle momentum**, in GeV;
- **Fits of resolution equation** shown in **upper right of figure**;
- Control fitted only up to 60 GeV – **bias results in unphysical resolution.**

What we learn:

- **Neural networks show significant improvement in stochastic term, a:**
  - - 3 % vs control using spatial information;
  - - 6 % vs control using spatial + 100 ps timing information;
- **Improvement in the constant term, b by ~5%**
  - **suggests method also calibrates detector;**
- **Neural networks show excellent agreement with theory by comparison to control.**



	a [%]	b [%]	$\frac{\chi^2}{NDF}$
Calorimeter Response	49.516 ± 0.401	7.147 ± 0.067	4.575
Control	43.387 ± 0.119	0.010 ± 2.873	14.333
Network, No Time	40.236 ± 0.217	2.158 ± 0.087	0.857
Network, + Time	37.275 ± 0.208	2.448 ± 0.070	1.440

Conclusion

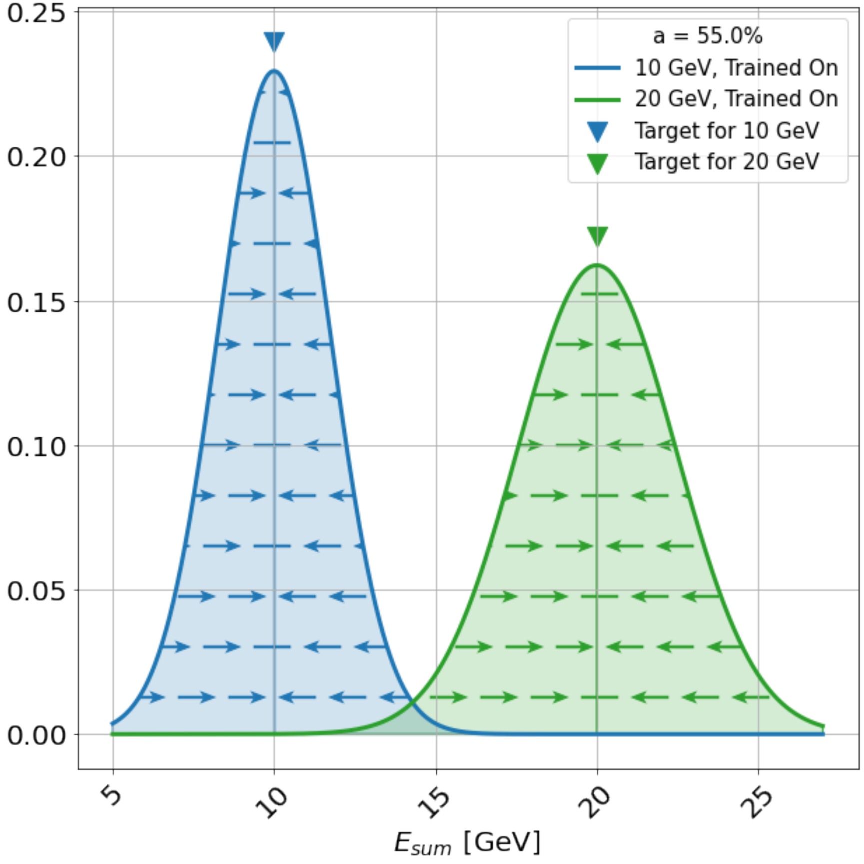
- **Machine Learning SC algorithm** developed using a **graph neural network (GNN) technique**;
- **ML method** was designed to be **unbiased to the particle momenta, overcoming previous limitations on ML-based SC.**
- **ML outperformed the control** in both compensation metrics:
  - **superior linearity of response than control method by around 2%** with and without timing information;
  - **superior stochastic resolution improved vs control method** by:
    - **3% using spatial energy-density event information;**
    - **6% using spatial + 100 ps resolution timing information**
  - **superior constant term by around 5% than intrinsic response distribution**

    → suggests model performs detector calibration too.

- **ML method can interpolate and extrapolate beyond training range → we succeeded in our goal!**

- Not presented here, but in the backup slides:
  - **ML method also learns leakage correction and MIP track subtraction;**
  - **suggests the method can be used as a general-purpose detector calibration tool, not just for SC!**

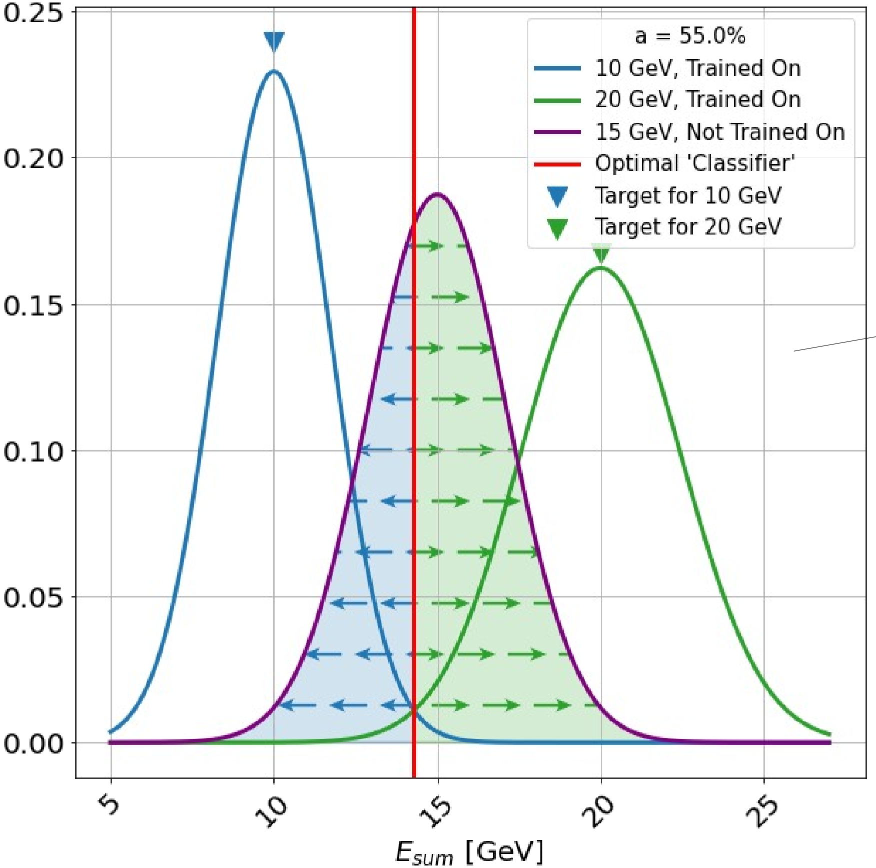
Backup: how SC algorithms are optimised

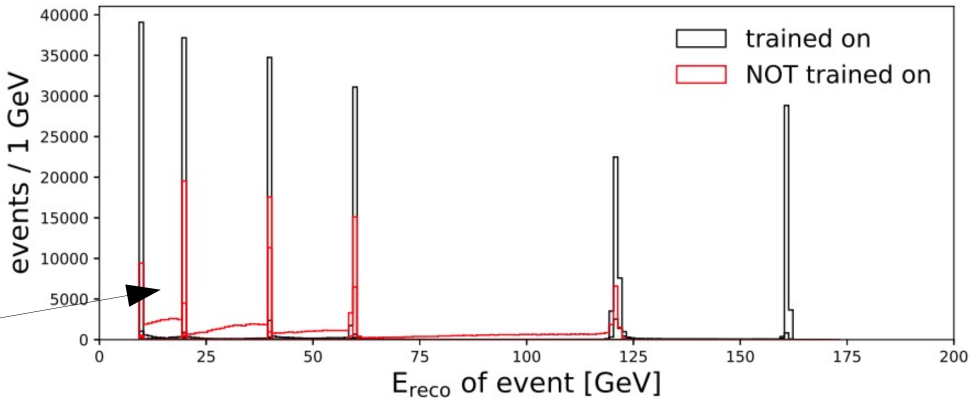


- We do not know the EM/HAD fraction a-priori.
- We do know:
  - HAD fraction ↑ → energy measured by AHCAL less than mean;
  - HAD fraction ↓ → energy measured by AHCAL more than mean;
  - Response of calorimeter is normally distributed about particle energy **(given no leakage and linearity of response!)**
- **Minimise χ^2 between target energy and distribution, using event information.**
- The width of the distribution is proportional to sum of 'stochastic' and 'sampling' fraction fluctuations, times the root of the energy of the particle (Poisson fluctuations in the number of particles
- **Underlying assumption: if influence of stochastic fluctuations reduced, then width of response distribution will reduce.**

$$\frac{\sigma_E}{E} \simeq \frac{a}{\sqrt{E_h}}$$

Backup: Interpolation Failure



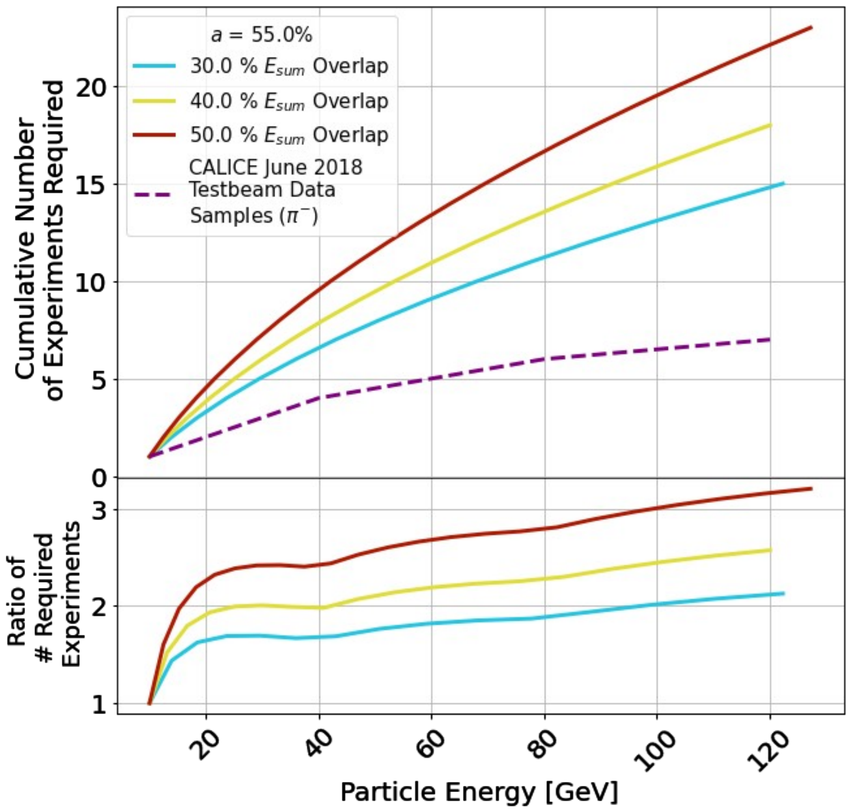


:

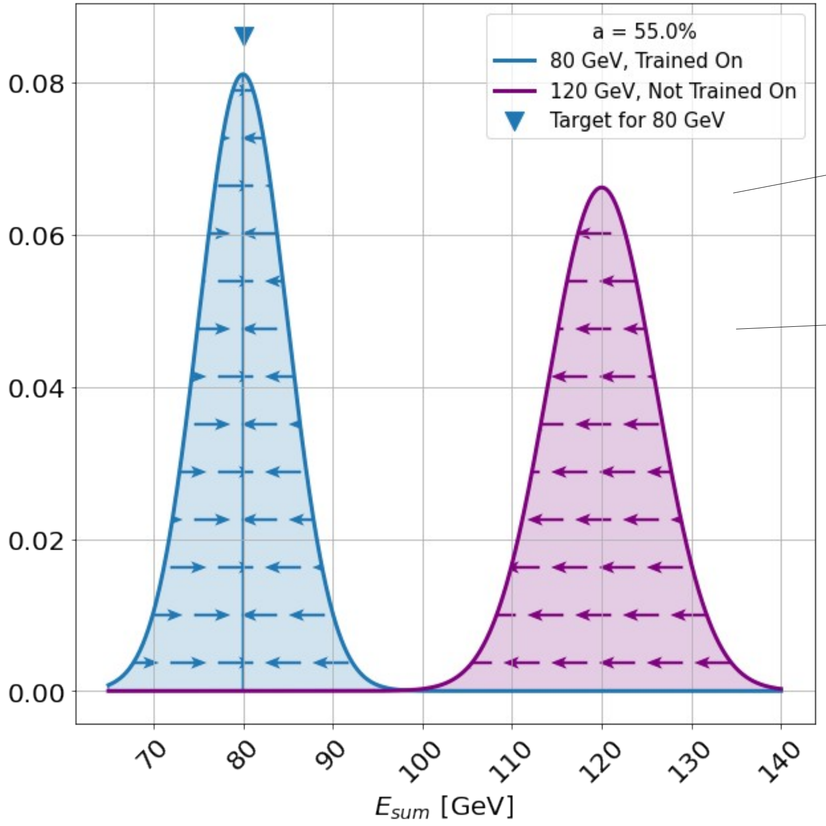
- If AHCAL records an shower energy of less than ~14 GeV, it must always be a 10 GeV hadron shower.
- If the AHCAL records a shower energy of more than ~14 GeV and less than ~30 GeV, it must be always be a 20 GeV hadron shower;
- **Result: 15 GeV showers are artificially compensated to be a 10 or a 20 GeV hadron shower → algorithm cannot be used in experiment.**

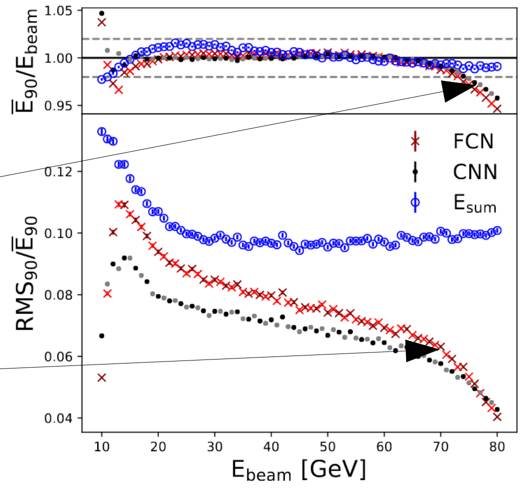
Backup: so, why not just take more data?

- We can calculate how many steps in energy (data samples) we would acquire for the calorimeter responses to overlap by a certain percentage.
  - i.e. if the responses overlapped, then it would not be possible for the model to take the ˎeasyˊ way out by applying biased thresholds.
- **Assumptions:**
  - Normally distributed AHCAL response to hadrons;
  - AHCAL detector resolution of around 55%
  - Linear response.
- June 2018 Testbeam at Super Proton Synchrotron have taken data samples at 10, 20, 30, 40, 60, 80 and 120 GeV particle energies. (7 samples)
- To calibrate the detector from 10 to 120 GeV (under these assumptions), we would need
  - ~2 x available samples for 30% overlap;
  - ~2.5 x available samples for for a 40% overlap;
  - ~3 x available samples for a 50% overlap;
  - **Test-beams are expensive.**
  - **Beam time is limited.**
  - **Energy consumption for data-taking increased.**



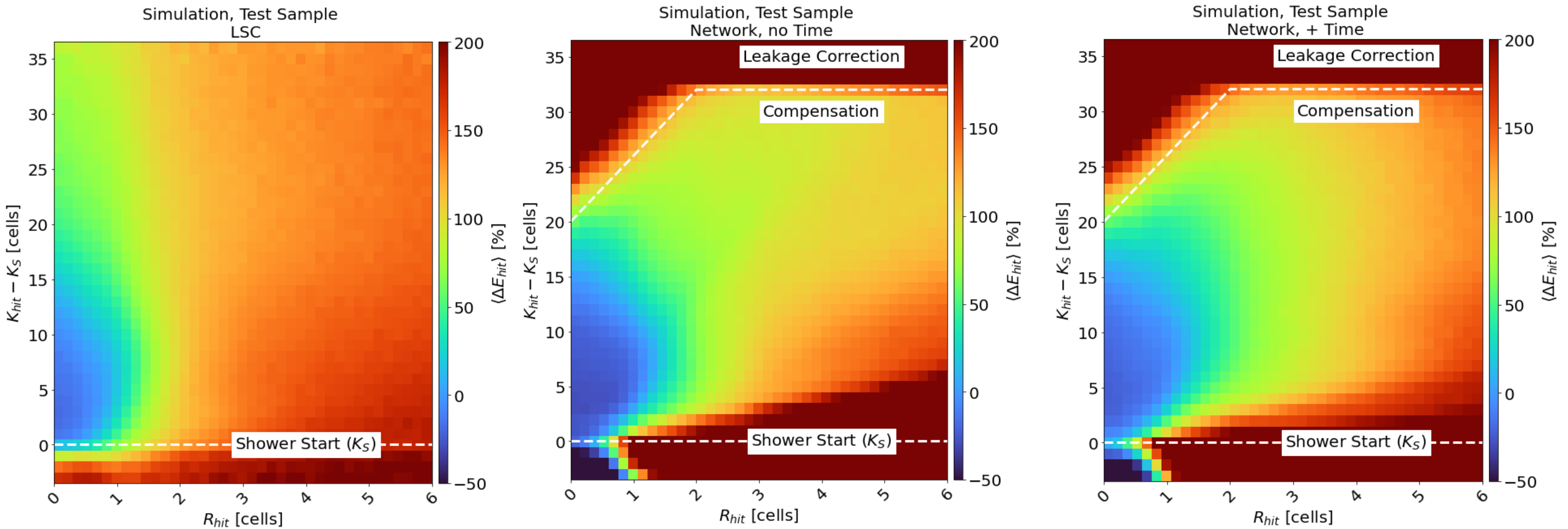
Backup: Extrapolation Failure





- Highest/lowest sample in distribution not penalise for weighting down/up the 'edges of the world' – affects the whole algorithm!
- **Result: showers with energy higher than edge of training range are always weighted down in energy, because there is no penalty for choosing this solution.**

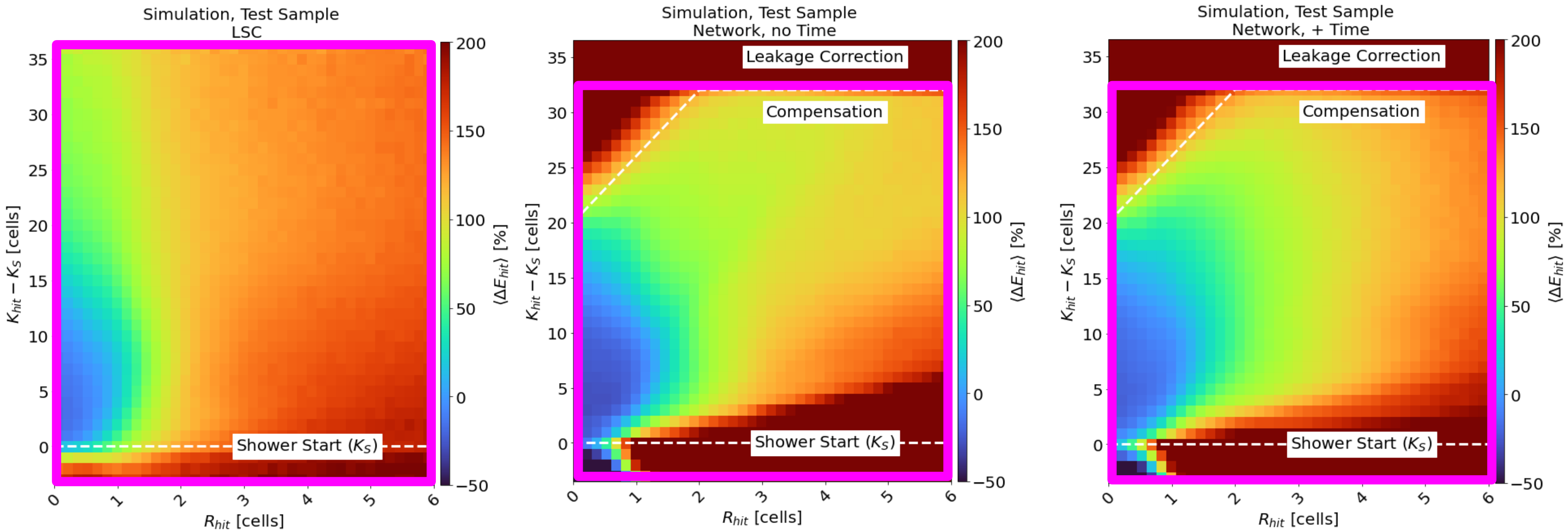
Backup: Mean Change in Cell Energy, Spatial Dependence



What is shown:
the average change in active cell energy due to compensation for each studied method as a function of:

- radial distance from shower axis (x-axis) and
- depth relative to shower start (y-axis).

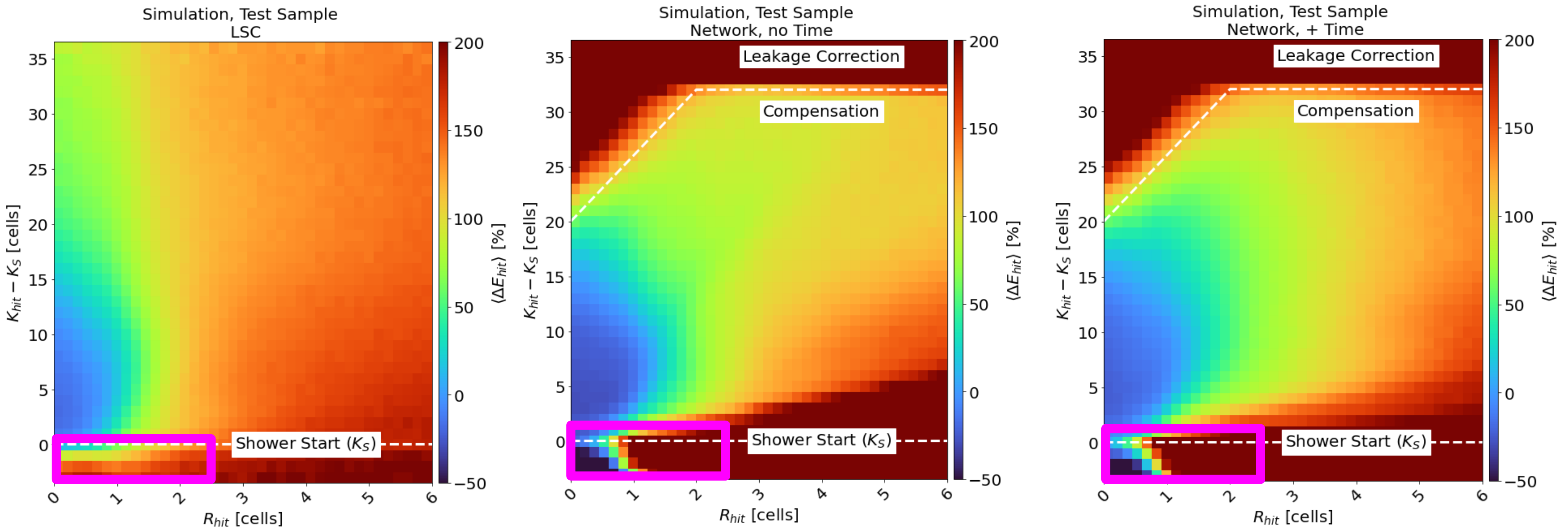
Backup: Mean Change in Cell Energy, Spatial Dependence



What we learn:

- neural networks learn a shower-development oriented compensation by comparison to control method;
- **Conclusion: spatial energy density information can be used to compensate the hadron shower.**

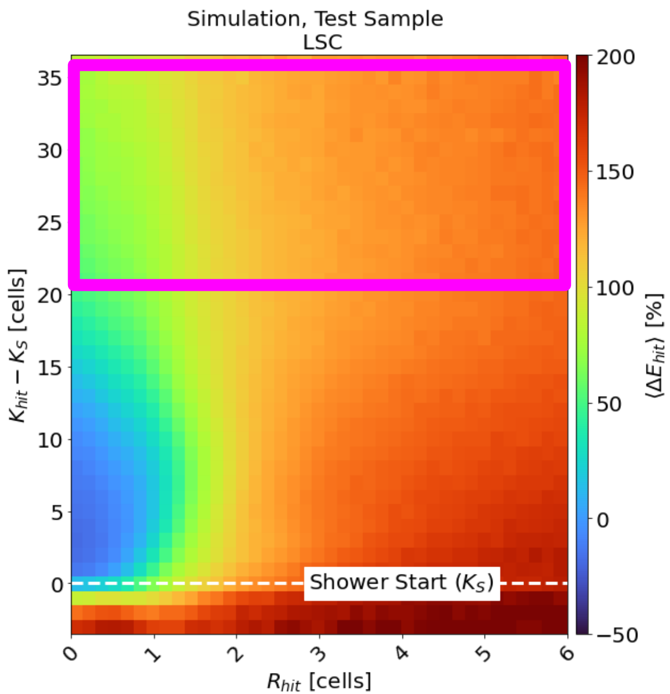
Backup: Mean Change in Cell Energy, Spatial Dependence

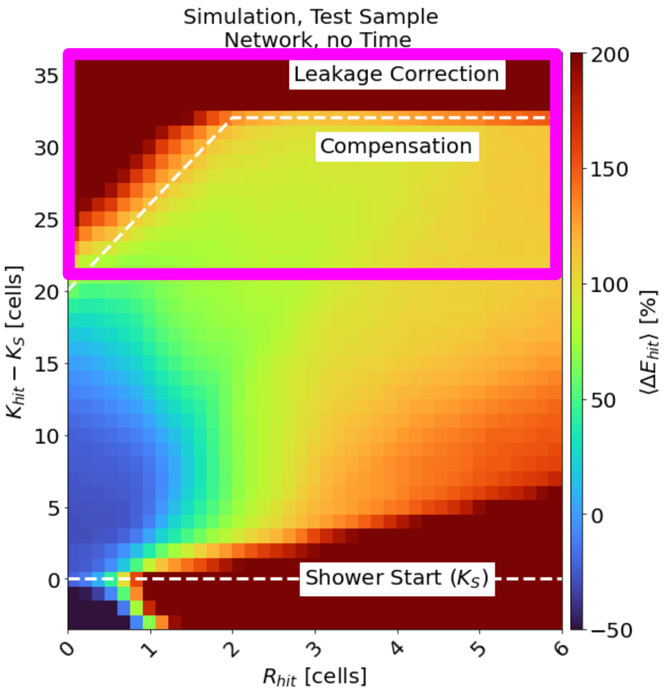


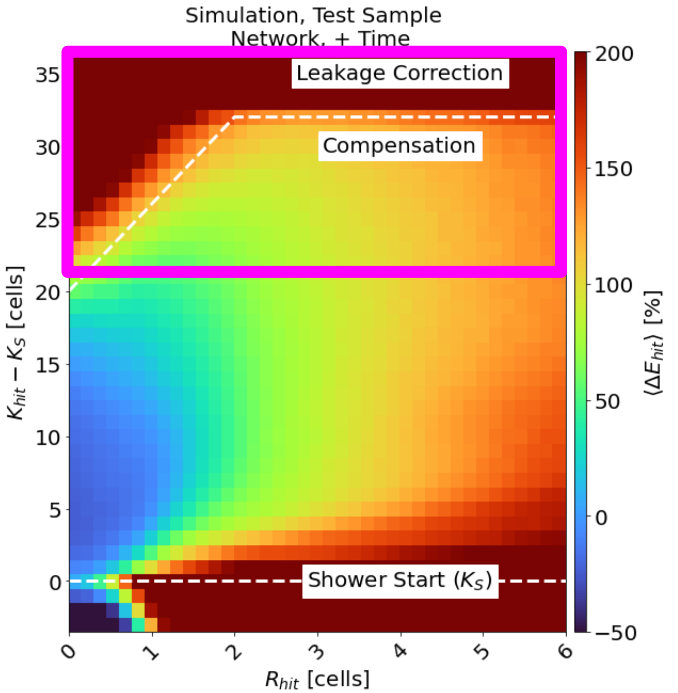
What we learn:

- neural networks to subtract the MIP track energy during energy reconstruction;
- control method shows no such effect → **must be a consequence of spatial energy density information**

Backup: Mean Change in Cell Energy, Spatial Dependence



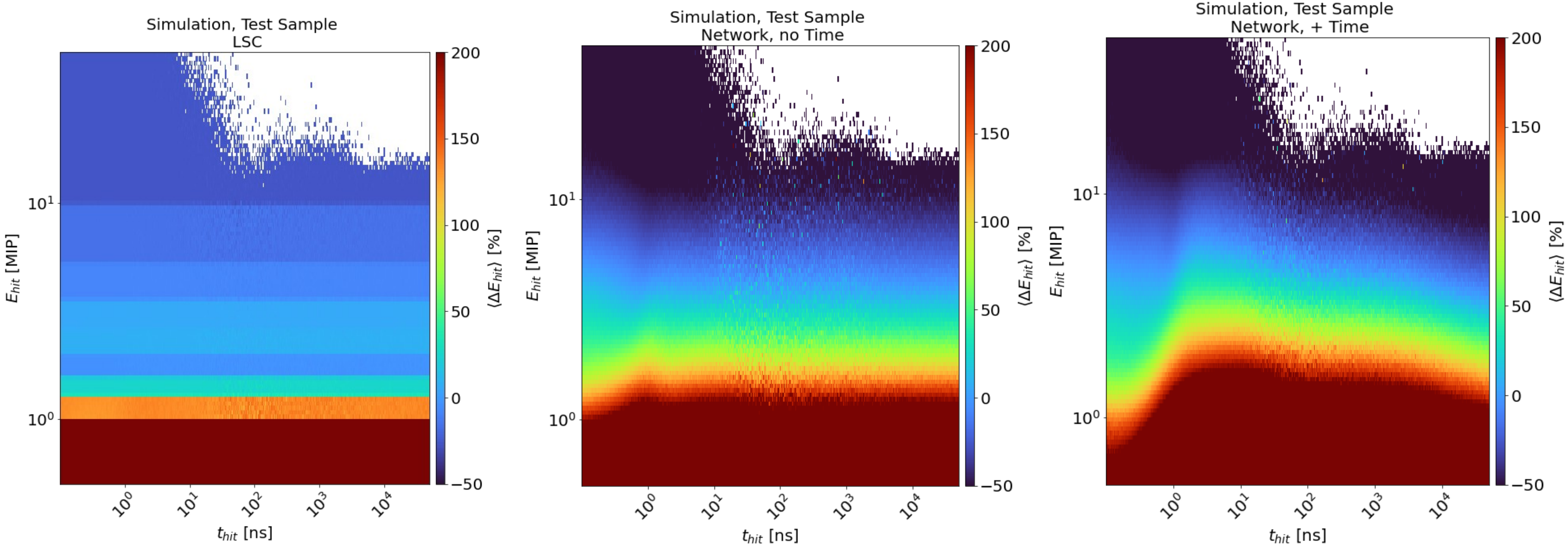




What we learn:

- Networks learn longitudinal leakage correction → sharp increase in weighting for hits deeper in the calorimeter.
- LSC shows no such effect.

Backup: Mean Change in Cell Energy, Energy/Temporal Dependence

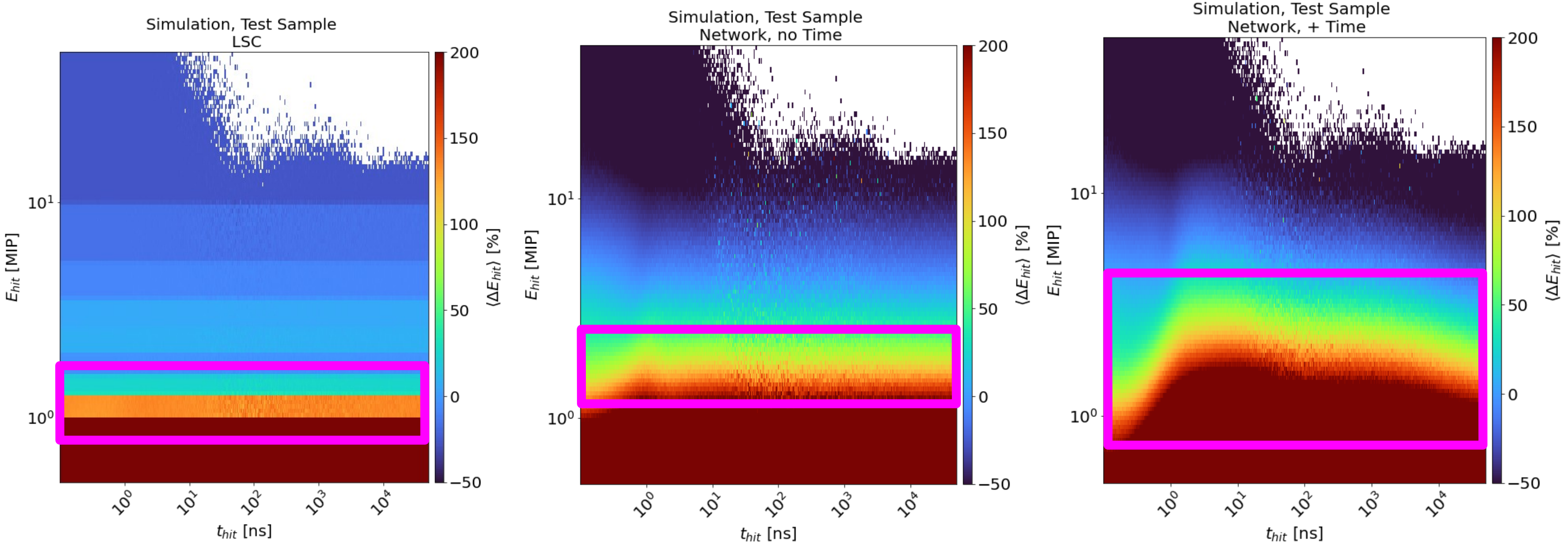


What is shown:

the average change in active cell energy due to compensation for each studied method as a function of:

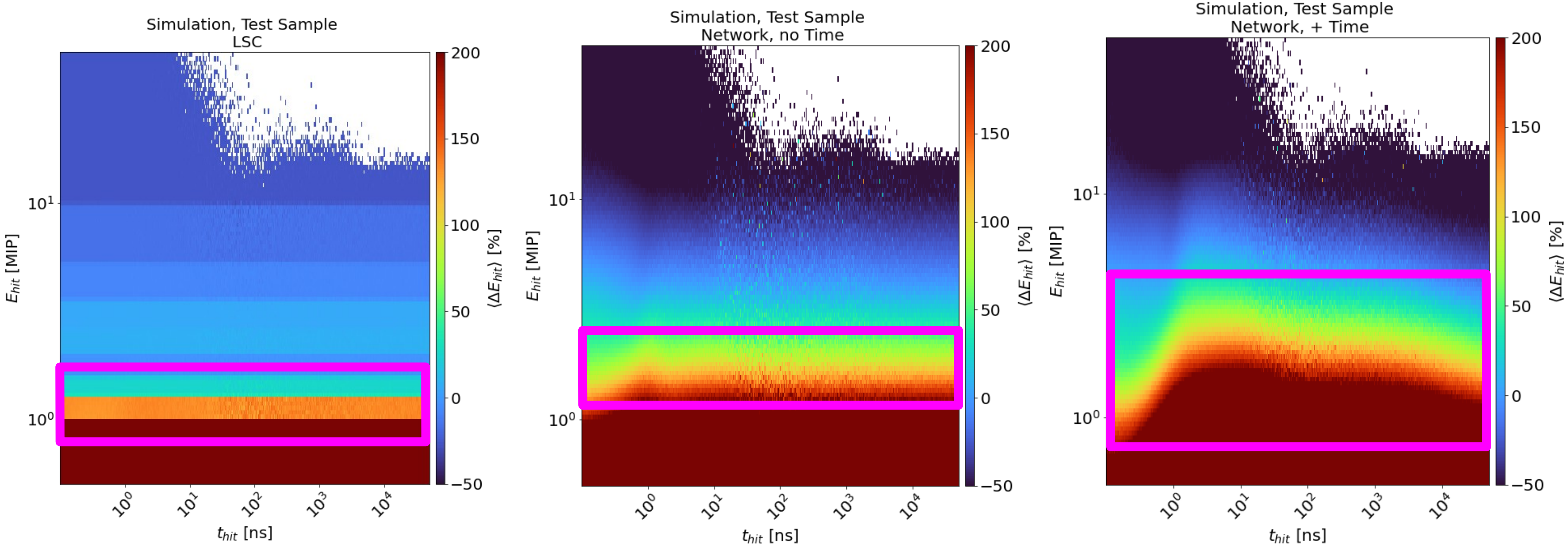
- original hit energy and;
- hit time

Backup: Mean Change in Cell Energy, Energy/Temporal Dependence



What we learn:
- Binning structure clearly visible for LSC method;
- Neural networks learn continuous energy weighting by comparison;

Backup: Mean Change in Cell Energy, Energy/Temporal Dependence

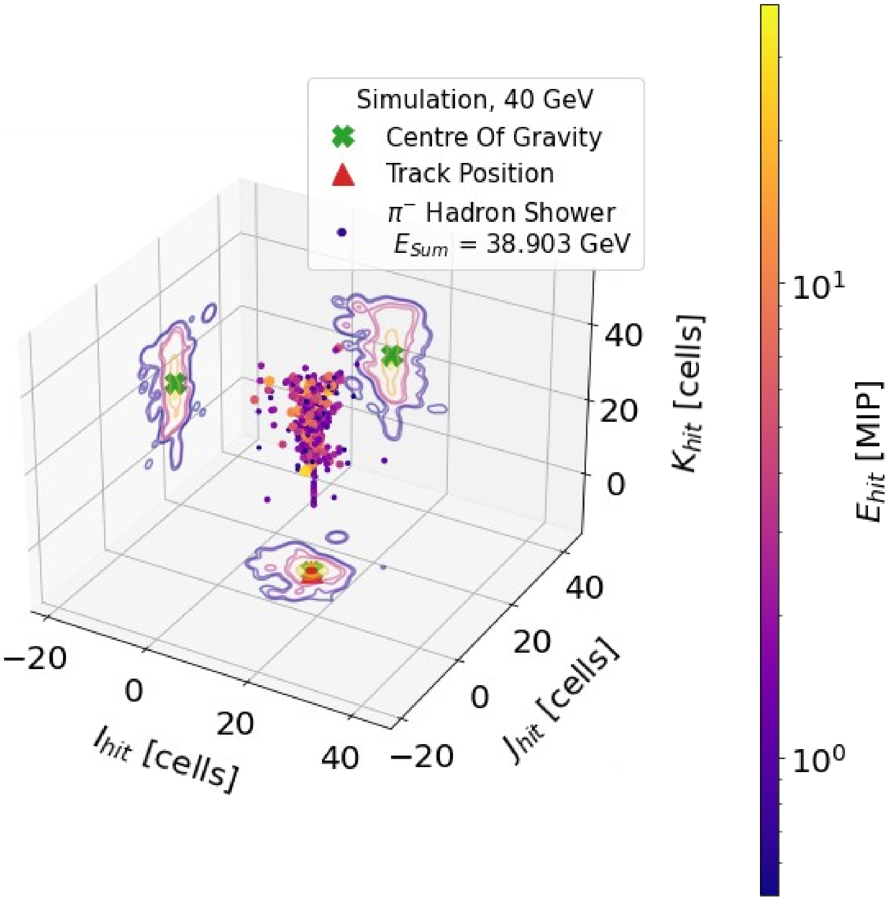


What we learn:

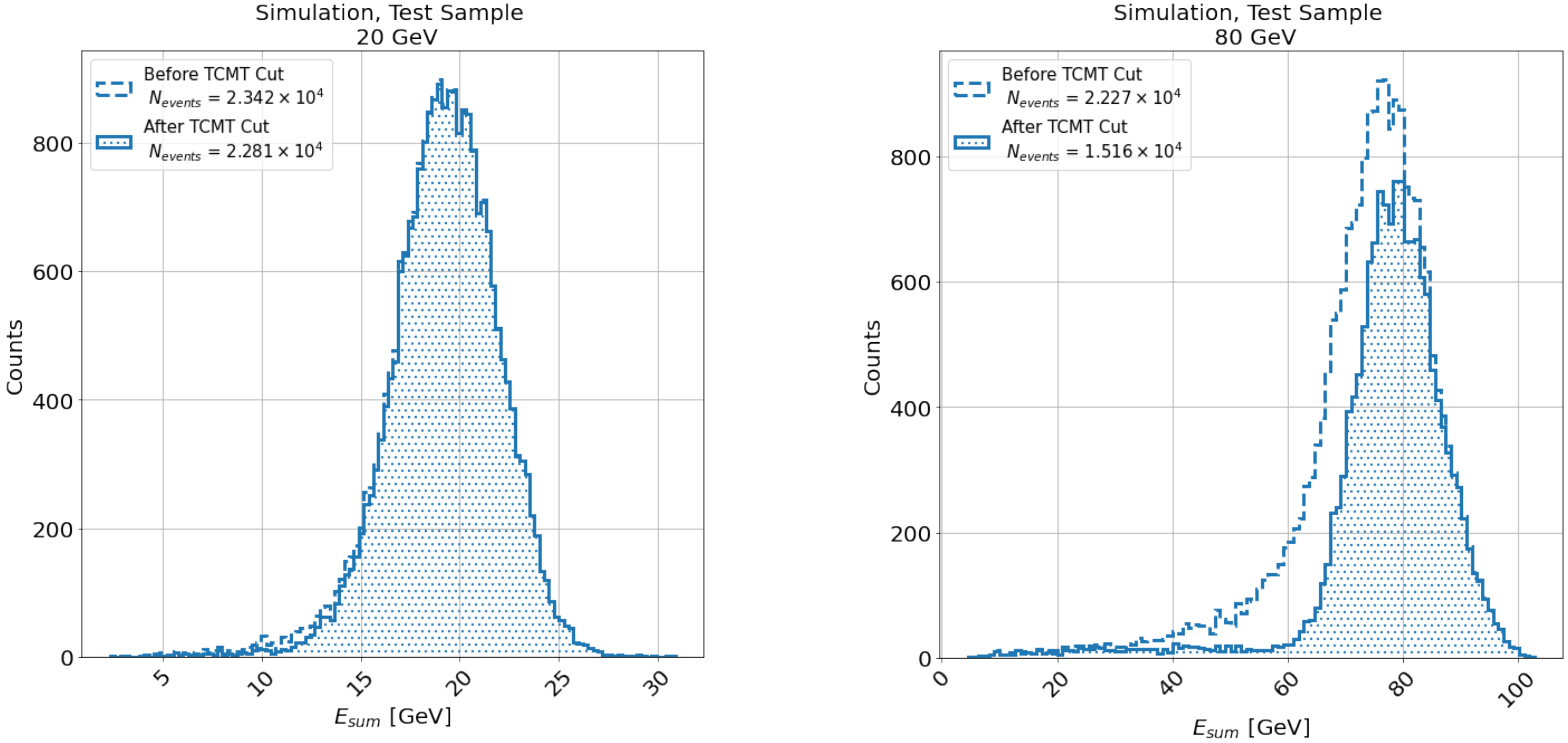
- No strong timing correlations observed for methods without timing information
- The addition of timing information results in a greater threshold for energy enhancement at the order of around 10 ns (Order of evaporation neutrons)

Backup: Simulation Details

- Physics list: QGSB_BERT_HP
- Particle: π^-
- Cuts:
  - + Shower Start: $1 < K_S \leq 5$
  - + Single track, with position 'inside' calorimeter: $1 < I_{\text{Track}}/J_{\text{Track}} \leq 24$
  - + PID MIP Cut: $P_{\mu-\text{like}} < 0.5\%$
  - (+ Tail Catcher Leakage Cut: $E_{\text{sum}}^{TC} < 25\text{MIP}$, for resolution/linearity measurement only)
- Events:
  - Training/Validation Sample:
    - 10-80 GeV, steps of 10 GeV;
    - Training: ~185,000 events (~20,000 events per step) (90% available)
    - Validation: ~ 20,000 events (~2,500 events per step) (10% available)
  - Testing Sample:
    - 10-120 GeV, steps of 5 GeV
    - Testing: ~ 497,000 events (~20,000 events per step) (all available)
    - Testing, + TC cut, ~384,000 events



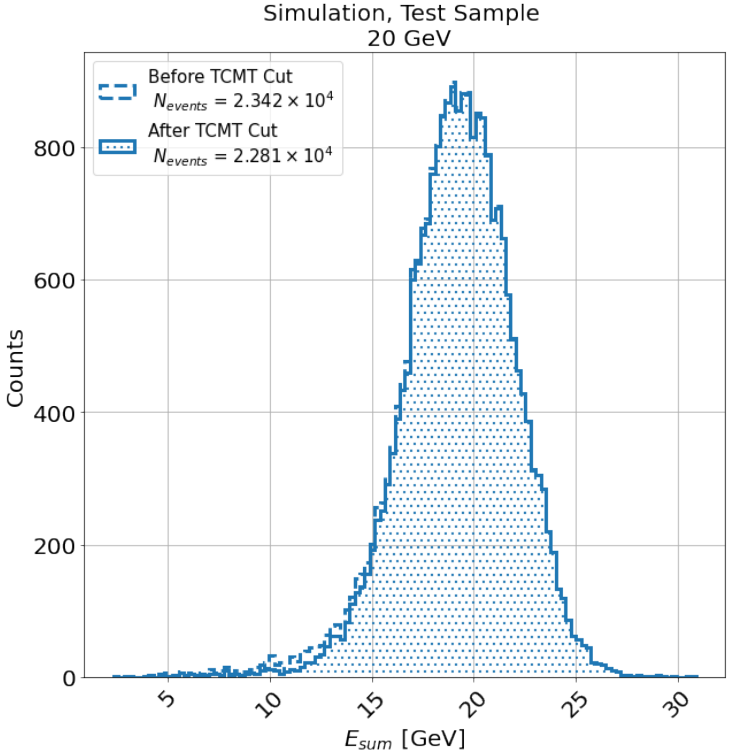
Backup: TCMT Cut

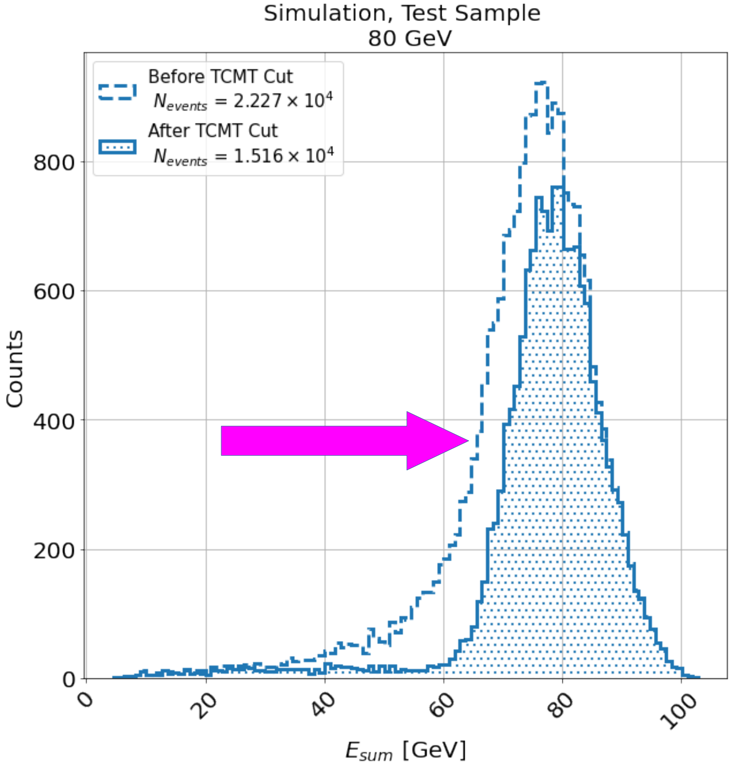


What is shown:

Distributions of uncompensated calorimeter response, before and after TCMT cut.

Backup: TCMT Cut





What we learn:

- At 20 GeV, barely any effect: hadron showers well contained.
- At 80 GeV, leakage tail greatly reduced → normally distributed response + minor leakage tail recovered.

Backup: choosing a co-ordinate system

Why is the co-ordinate system important?

- **Network is incapable of learning shower shapes by design→ not a problem if we use an appropriate co-ordinate system for shower development.**
  - hadron showers develop isotropically around a `**shower axis**´
    → **use radial distance of cell to shower axis instead of cell position**
  - Event by event, hadron showers can, however, deposit energy asymmetrically:
    → **shower asymmetry encoded in azimuthal angle of active cell**
  - A shower begins only after the first inelastic collision with the calorimeter absorber
    → **use layer, offset by the shower starting position.**
  - **All this information can be predicted from a single hadron shower event, a-priori.**
  - **This co-ordinate system remains the same, regardless of where the shower starts and where the particle enters**

TAKE-HOME MESSAGES:

- **Co-ordinate system is important for this method;**
- **algorithm `sees´ all hadron showers in terms of its development.**