SOFTWARE COMPENSATION FOR HIGHLY GRANULAR CALORIMETERS USING MACHINE LEARNING

The CALICE Collaboration

S. Lai,[a], J. Utehs,[a], A. Wilhahn[a], O. Bach,[b], E. Brianne,[b], A. Ebrahimi,[b], K. Gadow,[b], P. Göttlicher,[b], O. Hartbrich,[*b,1], D. Heuchel,[b], A. Irles, [†b,2], K. Krüger,[b], J. Kvasnicka [‡b,3], S. Lu,[b], C. Neubüser,[b], A. Provenza,[b], M. Reinecke,[b], F. Sefkow,[b], S. Schuwalow,[§b,4], M. De Silva,[b], Y. Sudo,[b], H.L. Tran,[b], E. Buhmann,[c], E. Garutti,[c], S.Ḣuck,[c], G. Kasieczka,[c], S. Martens,[c], J. Rolph[¶5], J. Wellhausen[c], G. C. Blazey[d], A. Dyshkant,[d], K. Francis,[d], V. Zutshi,[d], B. Bilki,[e], D. Northacker,[e], Y. Onel,[e], F. Hummer,[f], F. Simon,[f], K. Kawagoe,[g], T. Onoe,[g], T. Suehara,[‖g, 6], S. Tsumura,[g], T. Yoshioka,[g], M.C. Fouz,[h], L. Emberger,[i], C. Graf,[i], M. Wagner,[i], R. Pöschl,[j], F. Richard,[j], D. Zerwas,[j], V. Boudry,[k], J-C. Brient,[k], J. Nanni,[k], H. Videau,[k], L. Liu,[l], R. Masuda,[l], T. Murata,[l], W. Ootani,[l], T. Takatsu,[l], N. Tsuji,[l], M. Chadeeva,[m], M. Danilov,[m], S. Korpachev,[m], and V. Rusinov[m]

[a]II. Physikalisches Institut, Georg-August-Universität Göttingen, Friedrich-Hund-Platz 1, D-37077 Göttingen, Germany, [b]DESY, Notkestrasse 85, D-22603 Hamburg, Germany, [c]Univ. Hamburg, Physics Department, Institut für Experimentalphysik, Luruper Chaussee 149, 22761 Hamburg, Germany, [d]NICADD, Northern Illinois University, Department of Physics, DeKalb, IL 60115, USA, [e]University of Iowa, Dept. of Physics and Astronomy, 203 Van Allen Hall, Iowa City, IA 52242-1479, USA, [f]Karlsruhe Institute of Technology, Institute for Data Processing and Electronics, Kaiserstr. 12, D-76131 Karlsruhe, Germany, [g]Department of Physics and Research Center for Advanced Particle Physics, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan, [h]CIEMAT, Centro de Investigaciones Energeticas, Medioambientales y Tecnologicas, Madrid, Spain, [i]Max-Planck-Institut für Physik, Föhringer Ring 6, D-80805 Munich, Germany, [j]Université Paris-Saclay, CNRS/IN2P3, IJCLab, 91405 Orsay, France, [k]Laboratoire Leprince-Ringuet (LLR), CNRS, École polytechnique, Institut Polytechnique de Paris, F-91120 Palaiseau, France, [l]ICEPP, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan, [m]Affiliated with an institute that has signed the CALICE MOU

ABSTRACT

A neural network for software compensation was developed for the highly granular CALICE Analogue Hadronic Calorimeter (AHCAL). The neural network uses spatial and temporal event information from the AHCAL and energy information, which is expected to improve sensitivity to shower development and the neutron fraction of the hadron shower. The neural network method produced a depth-dependent energy weighting and a time-dependent threshold for enhancing energy deposits consistent with the timescale of evaporation neutrons. Additionally, it was observed to learn an energy-weighting indicative of longitudinal leakage correction. In addition, the method produced a linear detector response and outperformed a published control method regarding resolution for every particle energy studied.

[*] now at Oak Ridge National Laboratory, 1 Bethel Valley Road, Oak Ridge, TN 37830, USA

[†] now at Instituto de Física Corpuscular, Parque Científico, Catedrático José Beltrán, 2 | E-46980 Paterna, España

[‡] also at Institute of Physics, The Czech Academy of Sciences

[§] deceased

[¶] corresponding author

[‖] now at ICEPP, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo

1 Introduction

To fulfil the requirements for BSM physics searches and Higgs precision measurements at future linear colliders, a challenging final state jet-energy resolution must be achieved. For example, for ILC operating at $\sqrt{s}$= $0.5-1\,\mathrm{TeV}$ where typical di-jet energies for interesting physics processes will be in the range 150–350 GeV, a jet energy resolution of $2.7\,\%$ is crucial [1]. Particle Flow (PF) is a method expected to provide this resolution, which relies upon accurate tracking of charged particles in a jet, sophisticated event reconstruction techniques, and highly granular sampling calorimeters. A prototype of such a detector is the CALICE Analogue Hadronic Calorimeter (AHCAL) [2], a highly-granular steel-scintillator sampling calorimeter designed for PF, with $24 \times 24 \times 38$ individual silicon photomultiplier (SiPM) readout cells. The AHCAL is notable for its capacity to measure a timestamp for each readout channel.

The response of calorimeters to hadrons may be described in terms of two components: an electromagnetic component (produced mainly by $\pi^0/\eta \to \gamma\gamma$, contributed to by nuclear γ), and a hadronic component, which contains the remainder of energy depositing processes. The calorimeter response is therefore split into an EM response (e) and a HAD response (h). A hadron shower in a calorimeter exhibits an EM-dominated, energy-dense 'core' that propagates over a short longitudinal and lateral range and a HAD-dominated, diffuse energy-sparse 'halo', which propagates over a wider range [3]. Part of the energy deposited by a hadron shower cannot be detected and is called 'invisible energy' (e.g. neutrinos, nuclear binding energy losses). This fraction also experiences significant stochastic fluctuations from event to event, contributing to the calorimeter's resolution.

Compensation describes a method to equalise e and h, typically by attenuating e and enhancing h to improve the resolution. Hardware compensation requires careful tuning of the composition and proportions of active and passive material in the calorimeter. This method is difficult to implement in highly-granular calorimeters, which require a high degree of longitudinal segmentation. Therefore, software compensation (SC) algorithms are employed for this purpose and operate by estimating the EM fraction of a shower using information measured in each event.

Notably, spatial and temporal readout information available from highly granular calorimeter may be used for SC:

- A highly granular calorimeter may be able to resolve the hadron shower core and halo, and therefore exploit spatial energy density for SC;

- The number of neutrons produced in nuclear interactions is proportional, on average, to the invisible energy of the hadron shower. Energy deposits from neutrons can be measured indirectly using ionisation by recoil protons from neutron elastic scattering in hydrogenous active material such as plastic scintillator and photons from neutron capture. Energy deposits induced by neutrons are delayed by 10-100 ns in steel [4]. A time-sensitive hadron calorimeter may therefore exploit temporal information for SC.

Artificial neural network models have already been demonstrated to effectively exploit the spatial development of hadron showers to improve SC. For example, a study performed in Ref. [5] demonstrated that a deep neural network was found to improve the response of a highly-granular hadron calorimeter system from $48\,\%/\sqrt{E_{\mathrm{particle}}} \oplus 2.2\,\%$ to $37\,\%/\sqrt{E_{\mathrm{particle}}} \oplus 1\,\%$ using simulation.

However, a similar studies performed for AHCAL in Ref. [6] and Ref. [7], which trained and compared the performance of neural networks trained on both simulation and testbeam data, demonstrated the inability of similar machine learning-based SC algorithms to interpolate or extrapolate compensation from the limited hadron shower data typically available for such studies. In other words, the SC algorithm was biased to the training range of energies and its binning. This result is problematic as it indicates that experimental data of hadron showers from testbeams cannot be used for training SC algorithms because the available samples are typically binned too coarsely in particle energy to prevent bias. Additionally, while simulation samples can be used to train the algorithm with no constraints on particle energy binning or ranges, producing and storing these samples is presently an unsustainable practice. These limitations therefore motivate the development of an algorithm that can exploit the spatial and temporal information from AHCAL and simultaneously remain unbiased to the training particle energies.

In the presented study, a neural network is designed to perform SC on simulated π^- hadron showers observed with the AHCAL calorimeter, using the local spatial and temporal energy density from the event rather than just the sum of energy deposits. This information was expected to reduce the effect of stochastic fluctuations by improving sensitivity to the shower development and the neutron fraction of the event. Importantly, the neural network was carefully structured to reduce the effect of energy biasing. Finally, the neural network is compared to the standard CALICE SC method, which is used as a control algorithm. The results are then compared.

2 Methods and Tools

The CALICE AHCAL is a non-compensating, highly granular steel-scintillator calorimeter prototype designed for future precision e^+-e^- collider experiments. It has a highly granular structure, consisting of $24 \times 24 \times 38$ plastic scintillator cells of $30 \times 30 \times 3\,\mathrm{mm}^3$ volume each, read out by SiPMs. These cells indicate the spatial position, magnitude and timestamp of energy deposition with a minimum operating time resolution of up to $100\,\mathrm{ps}$ allowed by hardware. The detector has a depth of approximately 4.2 nuclear interaction lengths (λ_I). The hadronic calorimeter is complemented by a steel-scintillator Tail Catcher/Muon Tracker (TCMT) detector, composed of 320 extruded scintillator strips of $50 \times 5\,\mathrm{mm}^2$ area packaged in $16 \times 1\,\mathrm{m}^2$ planes interleaved between steel plates corresponding to an additional depth of $1.1\,\lambda_\mathrm{I}$ [8]. The TCMT is used in this analysis to tag leakage. Pictures of the AHCAL calorimeter are shown for reference in Fig. 1.

Event information from AHCAL consists of the position of an active cell with an energy deposit in the AHCAL cell matrix (I_hit, J_hit, K_hit), its energy in calibrated MIP units (E_hit), and its timestamp in nanoseconds, relative to the time at which deposited energy in a given cell cross a pre-defined threshold (t_hit). I_hit and J_hit indicate the lateral spatial position of an active cell relative to the longitudinal axis of the calorimeter (I_hit, $J_\mathrm{hit} \in [1, 24]$ in units of cell index). The longitudinal spatial position (depth in layers) is denoted K_hit ($K_\mathrm{hit} \in [1, 38]$ in units of layer index). The energy of an active cell is denoted E_hit, measured in Analogue-to-Digital counts, calibrated to the energy deposited by a minimum ionising particle (MIP) in one cell [9]. E_hit takes a value between a noise threshold at $0.5\,\mathrm{MIP}$ and the energy corresponding to the SiPM saturation value. The t_hit is bounded between the time at which the energy deposited in a given cell crosses a pre-defined threshold (normalised to $0\,\mathrm{ns}$ in this study), smeared by the resolution, and the chosen gate length for the measurement of an event. This study considers the ultimate $100\,\mathrm{ps}$ timing resolution for AHCAL. No charge integration gate length is considered in this study. The calorimeter response is measured as the sum of the individual active cells (hits) in an event, $E_\mathrm{sum} = \sum^\mathrm{event} E_\mathrm{hit}$. Additionally, the incident position of a charged particle in lateral coordinates is reconstructed using four delay wire chambers (DWC) of $10 \times 10\,\mathrm{cm}^2$ size, which is denoted as a vector $[I_\mathrm{track}, J_\mathrm{track}]$ [10]. The track information is only relevant to event selection cuts described in Section 2.3.

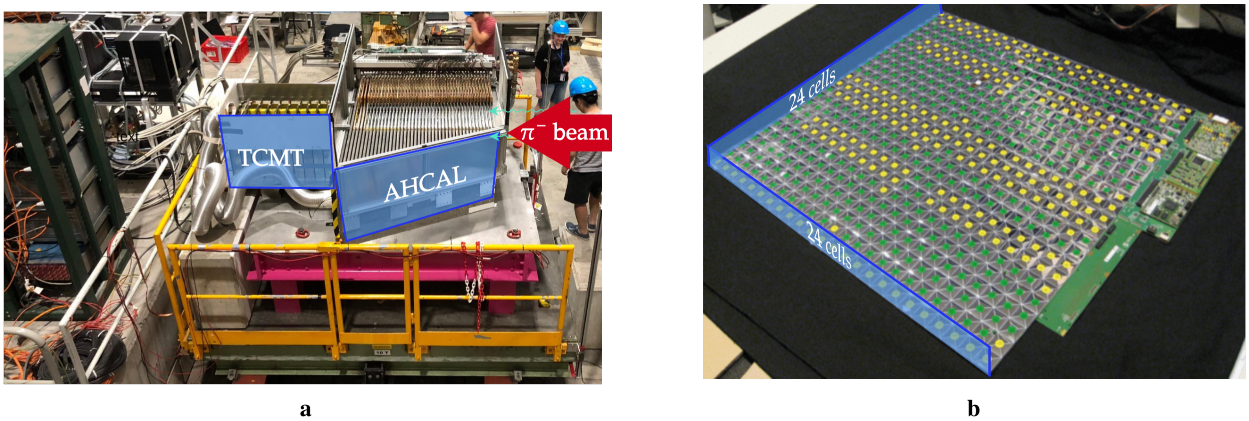


Figure 1: Pictures showing the CALICE AHCAL at testbeam. Fig. 1a shows the detector setup for a testbeam performed in June 2018 at the Super Proton Synchrotron (SPS) at CERN, Geneva [10]. Fig. 1b shows the individual cells of the calorimeter wrapped in foil to improve photon sensitivity.

The event coordinate system is changed for this study to reflect the shower development. The energy-weighted mean spatial position of the hadron shower in spatial coordinates is defined as a vector called 'centre-of-gravity' ($\mathrm{CoG} = [\mathrm{CoG}_I, \mathrm{CoG}_J, \mathrm{CoG}_K]$). Shower coordinates are converted to a cylindrical coordinate system, relative to the shower development axis, approximated by CoG_I and CoG_J, and the shower starting depth, K_S. These coordinates act as the origin and are all measured in cell units. K_S is calculated using a algorithm described in Ref. [11]. The transformation results in three new spatial coordinates: a hit radius, $R_\mathrm{hit} = \sqrt{(I_\mathrm{hit} - \mathrm{CoG}_I)^2 + (J_\mathrm{hit} - \mathrm{CoG}_J)^2}$, measured in cell units; a hit azimuthal angle, $\theta_\mathrm{hit} = \mathrm{arctan2}\,(J_\mathrm{hit} - \mathrm{CoG}_J, I_\mathrm{hit} - \mathrm{CoG}_I)$, where $\mathrm{arctan2}$ is the 2-argument arctangent, measured in radians, and a shower-start normalised depth, $K_\mathrm{hit} - K_S$, measured in layer indices. This coordinate system is advantageous to an SC algorithm because the hadron shower is represented independently of the lateral position of the hadron shower and the depth at which the hadron shower starts, and performs well for square cells of equal transverse size as in the AHCAL. Furthermore, these 'natural' spatial coordinates describe the lateral and longitudinal development of hadron showers more effectively than the raw event readout coordinates. They can be readily obtained from a well-separated single hadron shower. A visual representation of these coordinates is shown in Fig. 2 for reference.

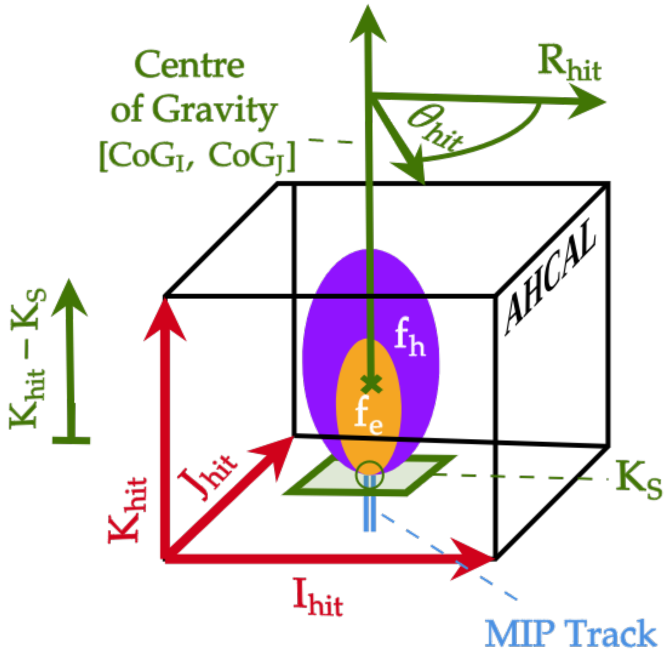


Figure 2: Diagram illustrating the modified event coordinate system. The red axes indicate the readout of the event from the cells. The green axes indicate the transformed spatial co-ordinates, as defined from the lateral centre-of-gravity, indicated by CoG$_I$ and CoG$_J$ and the shower starting position, indicated by the green plane labelled K_S.

2.1 Neural Network SC Method

SC models are typically trained indirectly since the $\frac{e}{h}$ fraction is unknown in a hadron shower event a priori. The resolution of a hadron calorimeter is described according to Eq. 1:

$$R = \frac{\sigma_E}{E} = \frac{a \cdot \sqrt{1\,\text{GeV}}}{\sqrt{E_\text{particle}}} \oplus b, \tag{1}$$

where E and σ_E are the mean and standard deviation of the response of the calorimeter to a hadron of E_particle, a describes the combined sampling and stochastic fluctuations experienced by the calorimeter, b the quality of detector calibration, non-uniformities in signal collection, imperfections in calorimeter construction etc. and $\oplus$ addition in quadrature. This equation is valid under the assumption of a normally distributed response (i.e. full shower containment, negligible electronics noise). Since reductions in σ_E imply compensation due to a smaller a, χ^2 minimisation of the calorimeter response to the known particle energy may be used to optimise SC algorithms. However, the lack of available high-statistics training samples for SC at finely binned particle energies tends to result in undesirable network biases that limit the algorithms' general applicability. In particular, two failure modes have been observed [6]: the 'classification' of the hadron showers by calorimeter response and bias to the training sample's upper and lower particle energy bins.

In particular, two failure modes have been observed [6]: the 'classification' of the hadron showers by calorimeter response and bias to the training sample's upper and lower particle energy bins.

A neural network was, therefore, designed to overcome the limitations of energy biasing. The proposed model was designed to use k nearest-neighbour (k-NN) clustering in the event coordinates defined in Sec. 2 to obtain a local estimate of the energy density in space and time. A k-NN cluster consists of the k nearest points in the event in space, energy and, optionally, time in terms of the square Euclidean distance, $\mathrm{d}s^2 = \mathrm{d}R_\text{hit}^2 + \mathrm{d}\theta_\text{hit}^2 + \mathrm{d}(K_\text{hit} - K_\text{S})^2 + \mathrm{d}\log{E_\text{hit}}^2 + \mathrm{d}\,\mathrm{arcsinh}\,{T_\text{hit}}^2$, where R_hit and $K_\text{hit} - K_\text{S}$ are in units of cells, θ_hit is in units of radians, E_hit

is in calibrated MIP units and $T_{\rm hit}$ is in units of nanoseconds. $E_{\rm hit}$ and $T_{\rm hit}$ are transformed to reduce the skewness of these variables. In practice, k-NN clustering is implemented by calculating the negative square distance matrix $-D_{ij}^2 = -|x_i|^2 - |x_j|^2 + 2 \cdot \langle x_i, x_j \rangle$ where x_i and x_j are co-ordinates of individual hits in space and time with indices i and j, $|x_{i,j}|^2$ is the absolute square of x_i or x_j, and $\langle x_i, x_j \rangle$ is the inner product of x_i and x_j. The columns of the D matrix are ranked based on their proximity to zero. The top k elements are then selected for each column, representing the k-nearest neighbours for each data point or coordinate, and the vectors between the seed cell and the other cells of the cluster are calculated. Each cluster is then treated independently by the neural network. With context to highly-granular calorimetry, this gives the local energy density surrounding a particular active cell during an event. The neural network was designed based on a single EdgeConv operator, introduced in the DGCNN graph neural network model [12]. The value of k was optimised using a hyperparameter scan to 20 cells using the Optuna hyperparameter optimisation tool [13], as shown in Table 2. The most critical aspect of the neural network design is that each cluster is operated independently of all others. This choice means that the capacity for the neural network to learn biased features of the training data, such as overall shower shape and energy, can be reduced compared to the case where the entire shower is presented as an input. This is because the network is subjected only to the local distributions of individual clusters of active cells. Therefore, the neural network is guided to infer the appropriate attenuation or enhancement of the calorimeter response from the energy distribution local to each active cell. The idea is summarised in Fig. 3.

The neural network architecture consists of five main stages:

- *Input*: The neural network is provided two inputs. The first is the hadron shower event in natural/transformed coordinates ($[R_{\rm hit}, \theta_{\rm hit}, K_{\rm hit} - K_S, \log E_{\rm hit}, \operatorname{arcsinh} t_{\rm hit}]$), where $\operatorname{arcsinh} t_{\rm hit}$ is optional. The second is the original cell energy, $E_{\rm hit}$, which is used to inform the neural network of the output scale of the compensated energy;
- *k-NN clustering*: As a pre-processing step, the neural network clusters the input according to the k-nearest neighbours. Their positions and vectors to their positions are calculated;
- *Addition of Dimensions*: Dimensions are added to each cluster using a module consisting of three sequential 2D fully connected convolutional layers 12, 24, and 48 channels, each using leaky ReLU activation and instance normalisation. Each new dimension is calculated using information from the inputs.
- *Processing*: Each cluster is passed through a deep processing layer consisting of 3 sequential 2D fully connected layers of 48 channels, each using leaky ReLU activation and instance normalisation;
- *Aggregation*: The maximum, mean and variance of the cluster dimension k are used as activation values for the cluster. These are concatenated with the cell energies of the event for each active cell;
- *Output*: The final layers of the network are five dense layers, with 1024, 512, 256 and 128 channels and leaky ReLU activation, with an output layer with ReLU activation such that the final output is positive. All dense layers, excluding the final layer, include dropout with probability $p_{\rm dropout}$. The neural network's final output is a single value for each active cell: the compensated hit energy, $\widehat{E}_{\rm hit}$. The sum of these outputs yields the total compensated response, $\widehat{E}_{\rm sum} = \sum^{\rm event} \widehat{E}_{\rm hit}$, where $\widehat{E}_{\rm hit} \in [0, \infty]$ is the compensated cell energy.

A diagram representing the proposed neural network architecture is shown in Fig. 4.

2.2 Control SC Method

The neural network was compared to the standard CALICE software compensation method called *'local software compensation'*, abbreviated hereafter as the control method and based on [14], is described as follows.

The $E_{\rm hit}$ distribution is binned in deciles (i.e. a 10% probability for a given $E_{\rm hit}$ to be found in any one of the bins). For each bin, an appropriate function is used for weighting. A function approximator in the form of a second-order Chebyshev polynomial of the first kind, ω_b, is defined as a function of the total calorimeter response, $E_{\rm sum}$, scaled using a factor, S, such that $E_{\rm sum}/S \in [0, 1]$ for the typical range of hadron shower energies of AHCAL ($S = 150\,{\rm GeV}$). ω_b has three free parameters, $\alpha_{\rm b}$, $\beta_{\rm b}$ and $\gamma_{\rm b}$, shown in Eq. 2:

$$\omega_b(E_{\rm sum}; S, \alpha_{\rm b}, \beta_{\rm b}, \gamma_{\rm b}) = \alpha_{\rm b} + \beta_{\rm b} \cdot \left(\frac{E_{\rm sum}}{S}\right) + \gamma_{\rm b} \cdot \left(2\left(\frac{E_{\rm sum}}{S}\right)^2 - 1\right) \tag{2}$$

For each bin, the corresponding weight is calculated. Finally, the energy of each active cell within the ranges defined by bin b is scaled by $\omega_{\rm b}$.

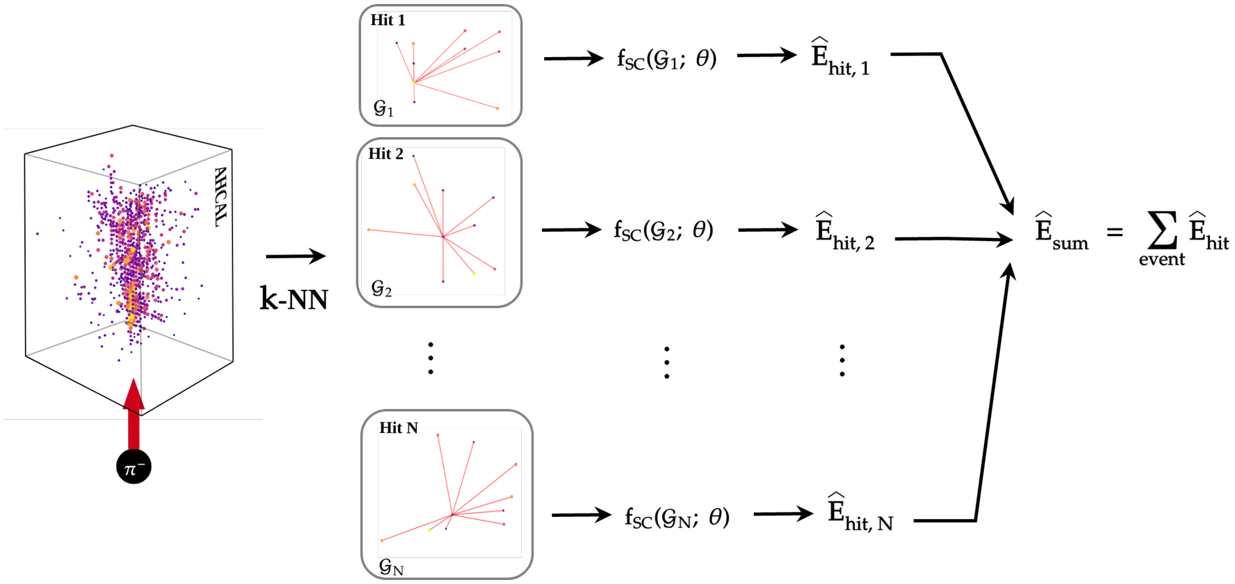

AHCAL
π^-
k-NN
Hit 1
$\mathcal{G}_1$
Hit 2
$\mathcal{G}_2$
Hit N
$\mathcal{G}_N$
$f_{SC}(\mathcal{G}_1;\ \theta)$
$f_{SC}(\mathcal{G}_2;\ \theta)$
$f_{SC}(\mathcal{G}_N;\ \theta)$
$\widehat{E}_{hit,1}$
$\widehat{E}_{hit,2}$
$\widehat{E}_{hit,N}$
$\widehat{E}_{sum} = \sum_{event} \widehat{E}_{hit}$

Figure 3: Illustration of a method for software compensation by which biasing may be reduced. First, a hadron shower of N measured with AHCAL, indicated by the event display on the left, is decomposed into a series of k-NN cluster graphs, indicated by the vertices and red lines, indicating edges between them, denoted $\mathcal{G}$. At this stage, each active cell is now represented as a local neighbourhood graph, $\mathcal{G}_i$, where i is the index of the active cell. This diagram shows the case of 9 nearest neighbours for illustration. Next, for each cluster, a SC model, $f_{\rm SC}(\mathcal{G}_i;\theta)$, is applied to each graph, where θ is the vector of the model's free parameters, producing an attenuated or enhanced calorimeter response to the kernel cell of $\mathcal{G}_i$, $\widehat{E}_{{\rm hit},i}$. The sum of the individually weighted active cells is then the compensated calorimeter response, $\widehat{E}_{\rm sum}$.

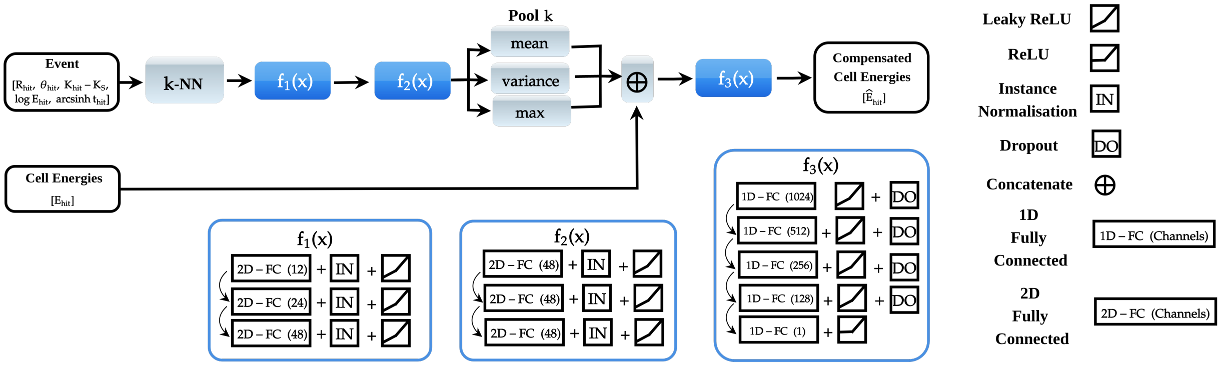


Figure 4: Flowchart describing the proposed neural network for software compensation studied in this paper. The black, blue and grey colour coding indicates inputs and outputs, convolutional operations and general operations, respectively. Additional operations are specified on the right of the figure.

$$\widehat{E}_{\mathrm{sum}} = \sum_{\mathrm{b}}^{\mathrm{bins}} \omega_{\mathrm{b}} \cdot E_{\mathrm{sum,b}} \tag{3}$$

The idea underlying this method is that higher hit energy bins attenuate the energy, as these are more likely to belong to an EM fraction and enhance the energy of low energy bins, which are more likely to belong to the HAD fraction. An example of the ten bins selected for the study is in Fig. 5.

2.3 Datasets and Training

Both the neural network model defined in Section 2.1 and the control model defined in Section 2.2 were trained and validated using experimental data from a CALICE test beam study at the Super Proton Synchrotron at CERN in 2018, as well as a simulated dataset thereof [10]. Each case was studied separately. Both simulation and experimental data were used for training and evaluation, respectively. The showers were produced from π^- hadron shower events observed with the AHCAL detector. The simulation of the particle showers was achieved using `Geant4` [15], with a full detector simulation developed using `DD4hep` [16]. Additional effects, such as digitisation of the analogue signal and reconstruction of the detector variables, were achieved for both simulation and data using `CALICESoft` [17]. Timing information from experimental data is not studied due to comparatively poor timing resolution arising from chip occupancy effects [18]. A MIP-to-GeV calibration factor of $37.3\,\mathrm{MIP/GeV}$ was used [19]. The statistics of the training, validation and test datasets are shown in Table 1.

The following selection criteria were applied:

- events were required to be identified using the standard CALICE particle identification algorithm [20] as being a single particle and having less than a $0.5\,\%$ probability of being a muon to exclude non-showering, ’punch-through’ pions;
- the 38^{th} layer of the AHCAL was ganged and required special treatment beyond the scope of this paper. Therefore, energy deposits were considered up to the 38^{th} layer of the calorimeter;
- events were selected to have a track position with a corresponding position inside the 24×24 cell AHCAL front-face and a shower starting layer within layers 1-4 of the AHCAL calorimeter. This choice was made to reduce the effect of longitudinal and lateral leakage on the experiment. These cuts were supplemented by an additional cut using the TCMT detector to only measure detector resolution and linearity. This criterion requires the TCMT to measure a total deposited energy of less than $25\,\mathrm{MIP}$ ($E_{\mathrm{sum}}^{\mathrm{TCMT}} < 25\,\mathrm{MIP}$).

The cuts applied, not including the TCMT cut, remove around two-thirds of the original sample. For the measurement of resolution, Eq. 1 cannot be used if the AHCAL experiences longitudinal shower leakage since the distributions exhibit a skewed ’leakage tail’ that deviates from the expected Gaussian response distribution. To resolve this, the TCMT is employed to tag and cut events likely to have a fraction of leakage energy.

Examples of the effect of the applied TCMT cut are shown in Fig. 6. Fig. 6a indicate the cut has practically no effect on $10\,\mathrm{GeV}$ hadron showers, while Fig. 6b show that the cut significantly reduces the leakage tail of the response distribution of $80\,\mathrm{GeV}$ hadron showers, resulting in a more Gaussian distribution at energies where leakage is observed. This means that the cut can be used to evaluate the resolution of the AHCAL alone. Fig. 6a and Fig. 6b indicate that the energy distribution of simulation is similar to experimental data and that the cut has a similar effect on both. This means the cut can be applied without modification in both cases.

As a caveat, this method is sub-optimal based on the considerable longitudinal leakage experienced by the AHCAL and will bias the resolution measurement to more compact showers. Therefore, example distributions illustrating the performance of the software compensation without the TCMT cut are supplied for reference in the results shown in Section 3, to illustrate the performance of each method without this bias.

The training and validation dataset consisted of simulated showers induced by π^- hadrons with E_{particle} in the range 10-80 GeV, in increasing steps of $10\,\mathrm{GeV}$. By contrast, the test sample contained showers induced by π^- hadrons with E_{particle} in the range 10-120 GeV, in increasing steps of $5\,\mathrm{GeV}$. The finer granularity tests the hypothesis that the neural network is unbiased to the particular particle energies used for training. Energies higher than the training range are included to test the generalisation capacity of each compensation method. The whole range of energies is used for training and testing for experimental data.

Two independent networks based on the model defined in Section 2.1 were trained on the training dataset: one without timing information and one with timing information. The proposed compensation networks were developed in `PyTorch` [21] and trained using the `PyTorch Lightning` research framework [22] on an NVidia V100 GPU. The ADAM

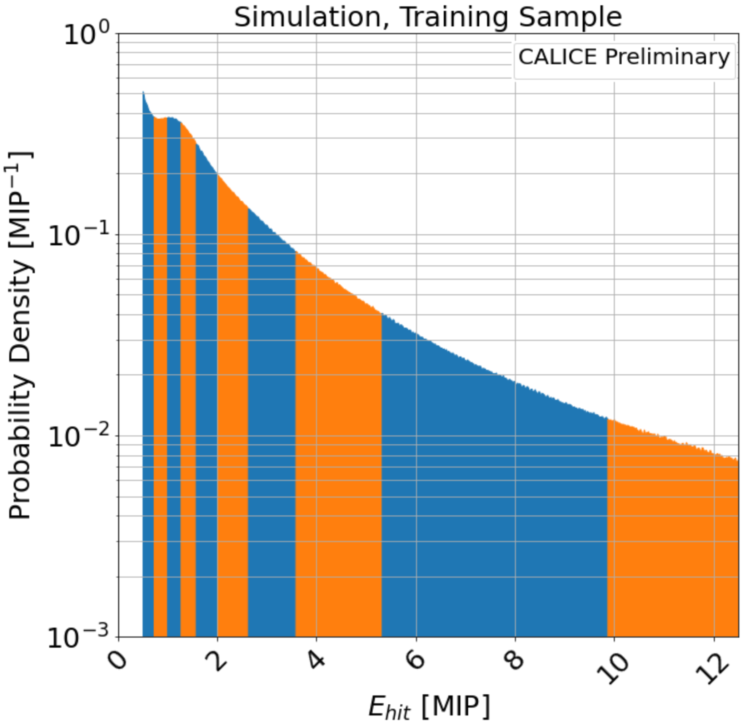
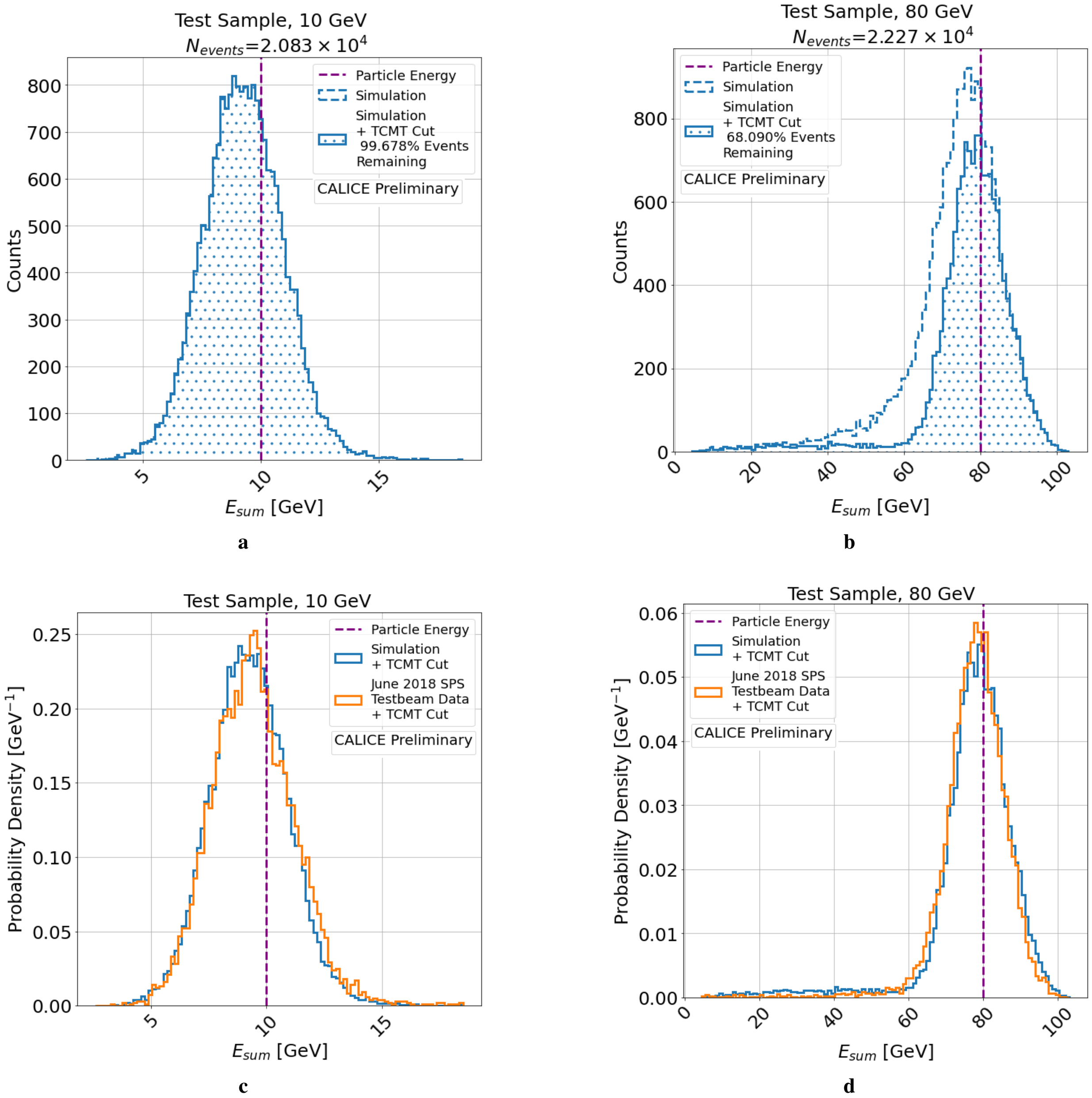


Figure 5: Histogram showing the ten decile bin ranges of the E_{hit} distribution, shown in alternating blue and orange, are shown for the training sample. Each bin corresponds to a total cumulative probability of 10% (deciles). The bin ranges extracted from the training sample discussed in Section 2.3 are shown in Appendix Table 5.

Table 1: Table of events used for training SC models after all cuts except the TCMT cut (shown separately), split into simulation and data and by the testing, training and validation samples and by data and simulation. Hyphens indicate 0 events.

Type	June 2018 SPS Testbeam Data				Simulation			
Sample	Test	Test	Training	Validation	Test	Test	Training	Validation
		+ TCMT Cut				+ TCMT Cut		
E_{particle} [GeV]								
10	6472	6460	51773	6472	20826	20759	18719	2080
15	-	-	-	-	21969	21685	-	-
20	9439	9233	75512	9439	23425	22808	21428	2381
25	-	-	-	-	25193	24124	-	-
30	-	-	-	-	24031	22491	21901	2434
35	-	-	-	-	24154	22065	-	-
40	10384	9378	83064	10383	24195	21513	23552	2617
45	-	-	-	-	23122	19981	-	-
50	-	-	-	-	27337	22889	24737	2749
55	-	-	-	-	19636	16009	-	-
60	13223	10684	105782	13223	22503	17728	24479	2720
65	-	-	-	-	25584	19374	-	-
70	-	-	-	-	18951	13889	24864	2763
75	-	-	-	-	15827	11204	-	-
80	11666	8298	93325	11666	22272	15165	25308	2813
85	-	-	-	-	22577	14875	-	-
90	-	-	-	-	26210	16618	-	-
95	-	-	-	-	20605	12475	-	-
100	-	-	-	-	17706	10385	-	-
105	-	-	-	-	17410	9873	-	-
110	-	-	-	-	16885	9161	-	-
115	-	-	-	-	18706	9820	-	-
120	10713	5829	85701	10713	18192	9239 -	-	
Total Events	61897	49882	495157	61896	497316	395131	184988	20557



a

b

c

d

Figure 6: Fig. 6a and Fig. 6b show the reconstructed AHCAL energy distributions of simulation for 10 GeV and 80 GeV π^- hadron showers. The dashed line and the solid line filled with dots indicate the distribution before and after the applied TCMT cut. Fig. 6c and Fig. 6d show response distributions of the test samples of simulation and June 2018 Testbeam data before and after the cut was applied, for 10 GeV and 80 GeV π^- hadron showers.

optimiser was used to improve the convergence rate for ten epochs, with early stopping applied. The hyperparameters used for training are shown in Table 2. Hyperparameter optimisation was achieved by optimising the networks many times with different starting parameters for k, learning rate, and dropout probability. Varying β_1 and β_2 resulted in large fluctuations in performance and were therefore held at nominal values. The hyperparameter search program `Optuna` optimised the hyperparameters for 50 trials, with 10 epochs per trial. To speed up convergence, trials were rejected using 'median pruning', where a trial is pruned if its best intermediate result is worse than the median of intermediate results of previous trials at the same epoch.

Table 2: Table of hyperparameters used to train the neural network. In this table, β_1 and β_2 are the ADAM momentum parameters, p_{dropout} is the dropout probability, and k is the number of nearest-neighbours per cluster. The parameters were informed by a hyperparameter scan using `Optuna` [13].

Parameter	Value
Learning Rate	9×10^{-5}
Batch Size	32
β_1	0.9
β_2	0.999
p_{dropout}	0.15
k	20

The control method was also trained using the training dataset, using the `MIGRAD` algorithm of the `Minuit` minimisation program [23]. Weights were initialised such that the compensation algorithm acted as the identity operator ($\alpha_{\text{b}} = 1$, $\beta_{\text{b}} = 0$, $\gamma_{\text{b}} = 0$).

The loss was chosen to be the χ^2 goodness-of-fit of the compensated energy to the known particle energy of the hadron shower:

$$\mathcal{L}(\widehat{E}_{\text{sum}}; E_{\text{particle}}) = \frac{\left(\widehat{E}_{\text{sum}} - E_{\text{particle}}\right)^2}{E_{\text{particle}} \cdot (1\,\text{GeV})} \tag{4}$$

The denominator in the loss arises from the uncertainty on the Poisson-distributed sampling quanta measured by the calorimeter, $\sigma_E = a \cdot \sqrt{E_{\text{particle}}}$. The dummy constant of 1 GeV in the denominator is formally included to make the loss unitless and merely acts to scale the loss. The mean loss was used for both implementations to optimise the control and network methods. For the network methods, the epoch with the smallest mean loss of the validation sample was chosen for further study. The control method was minimised with the MIGRAD algorithm until the mean training loss reached convergence.

The training process involved utilising both simulation and data training samples from Table 1 to train separate models for both cases, which were subsequently evaluated on their respective test samples. Consequently, the model's performance was assessed independently for scenarios in which it was trained with simulation or experimental data.

3 Results

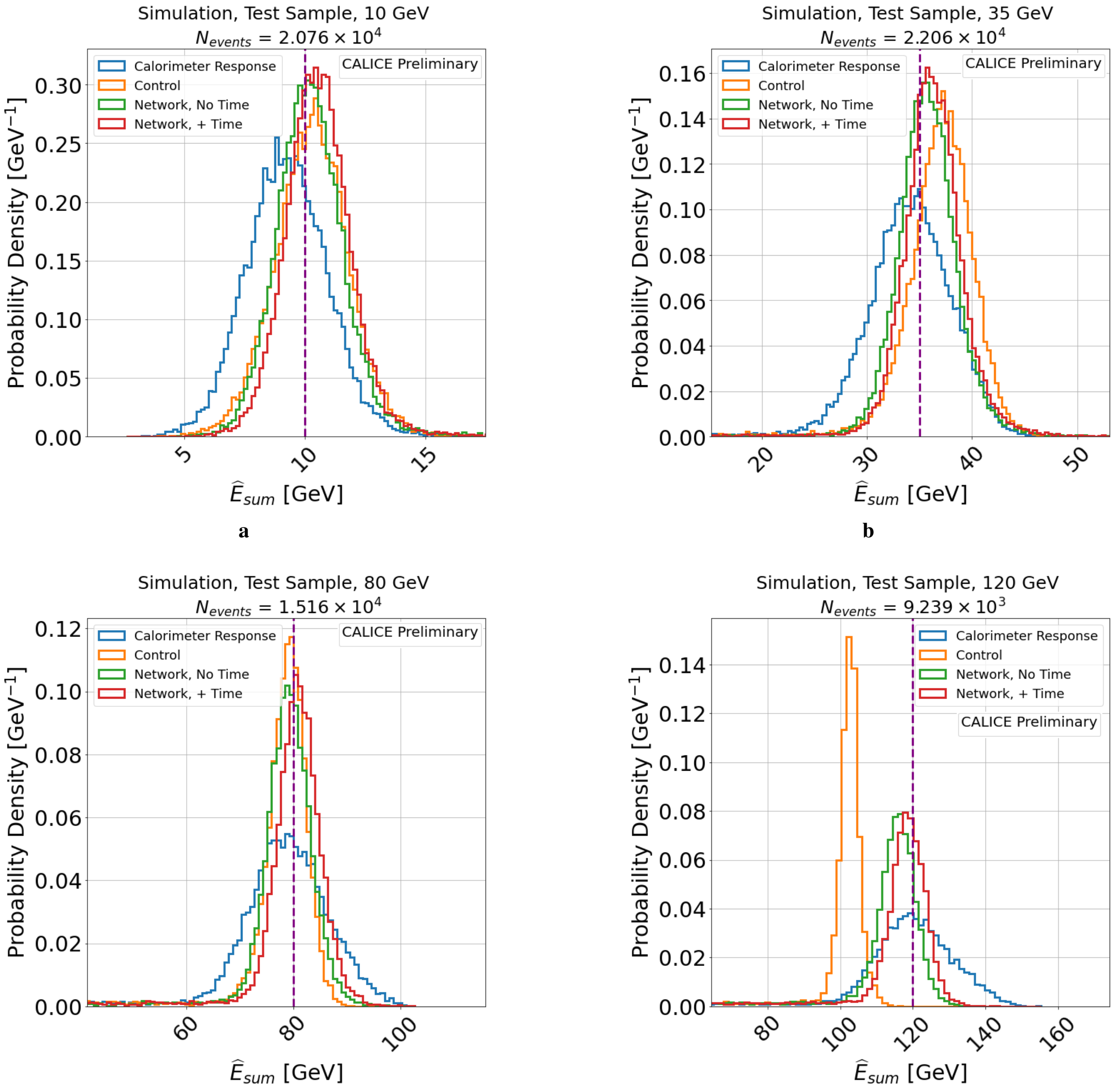
Each trained model was applied to the test sample. The effect of compensation was then analysed for each method.

3.1 Example Response Distributions

The normalised energy response distributions for the simulation and 2018 June Testbeam test samples are shown in Figs. 7a-7d and Figs. 9a-9d, with the TCMT cut applied. In simulation, particle energies of 10 GeV, 35 GeV, 80 GeV and 120 GeV are shown. For the experimental data, particle energies of 10 GeV, 40 GeV, 80 GeV and 120 GeV are shown. These samples are used for measurement of the resolution. The corresponding distributions without the TCMT cut applied are shown for the same particle energies in simulation and experimental data in Figs. 8a-8d and Figs. 10a-10d, respectively. The uncompensated sample and each sample after compensation are shown in each plot. The Freedman-Diaconis rule was applied to each sample to determine the bin width [24]. The Freedman-Diaconis rule is a commonly used binning rule that approximately minimises the integral of the square difference between a histogram and a probability density function.

Simulation Fig. 7a shows that the neural network methods outperform the control method for the 10 GeV sample, indicated by the lower spread of the response than for the control method. Furthermore, including timing information results in superior energy resolution, which is expected to play a more significant role in compensation at this energy scale due to a larger HAD fraction on average than at higher energies in the training sample, since the EM fraction increases on average with E_{particle} [3].

Fig. 7b shows the 35 GeV testing sample, which demonstrates that the neural network methods produce a more linear response than the control method and are therefore able to interpolate to samples between training energies.



a

b

c

d

Figure 7: Example normalised histograms showing the calorimeter response before and after compensation applied to the simulated test dataset of Table 1, with the TCMT cut applied. Samples of 10 GeV, 40 GeV, 80 GeV and 120 GeV hadron shower energies are shown. Blue lines indicate intrinsic calorimeter response, while orange, green and red lines indicate the control, network without and network with time compensation methods, respectively. E_{particle} is indicated as a dashed purple line. The number of events shown in the title indicate the corresponding sample sizes from Table 1.

Fig. 7c and Fig. 7d show the 80 GeV and 120 GeV samples. The control method outperforms the neural network methods for the 80 GeV sample. However, by examination of the 120 GeV sample, it becomes apparent that this result is due to the control method biasing to the highest energy sample of the training dataset. This statement is justified by the artificial attenuation of the response by the control method, resulting in a highly non-linear compensated response. By contrast, the neural network methods preserve the linearity of response beyond the training range. Therefore, it is demonstrated that the neural network model can extrapolate the compensation to higher particle energies without further training.

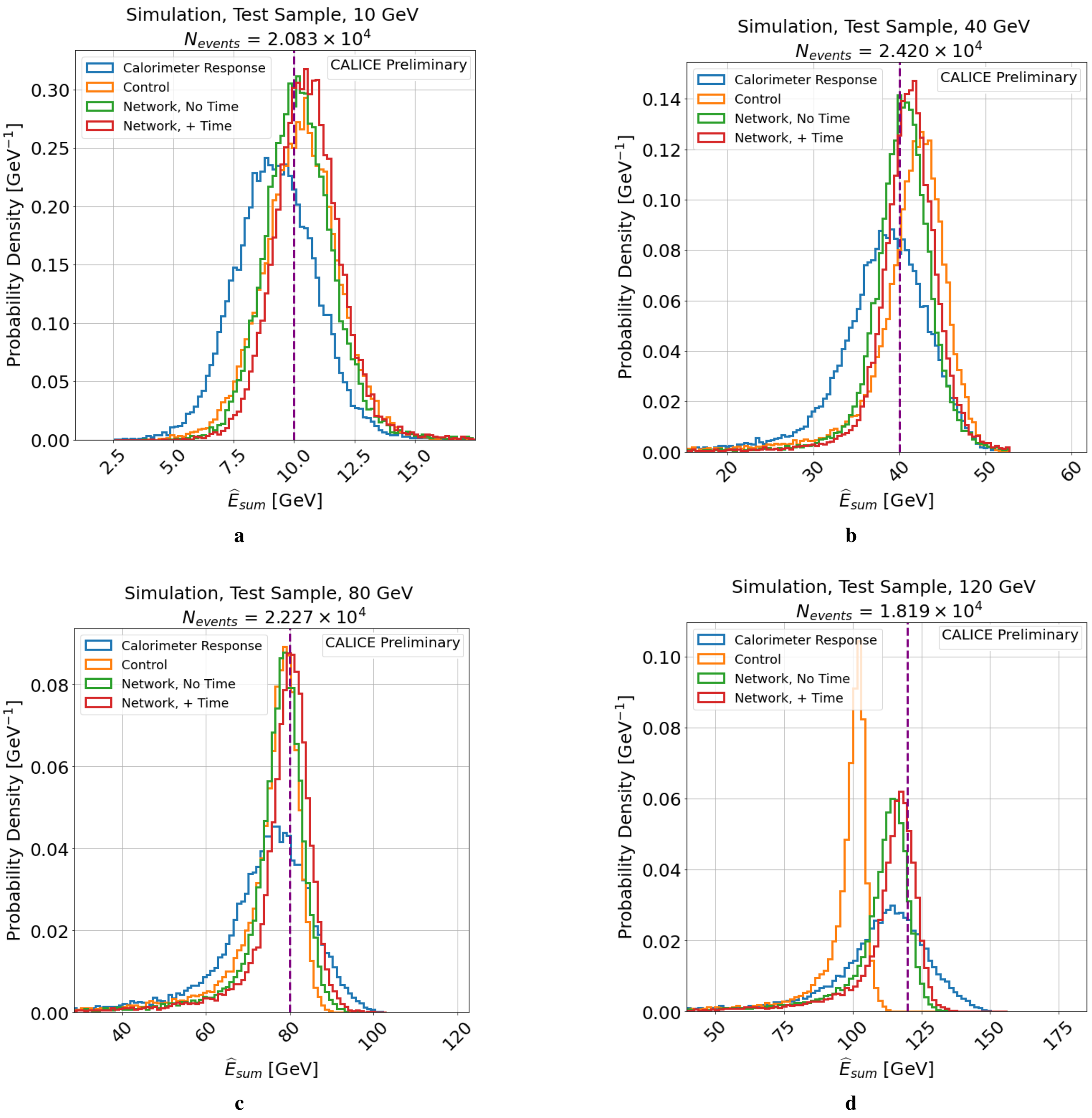
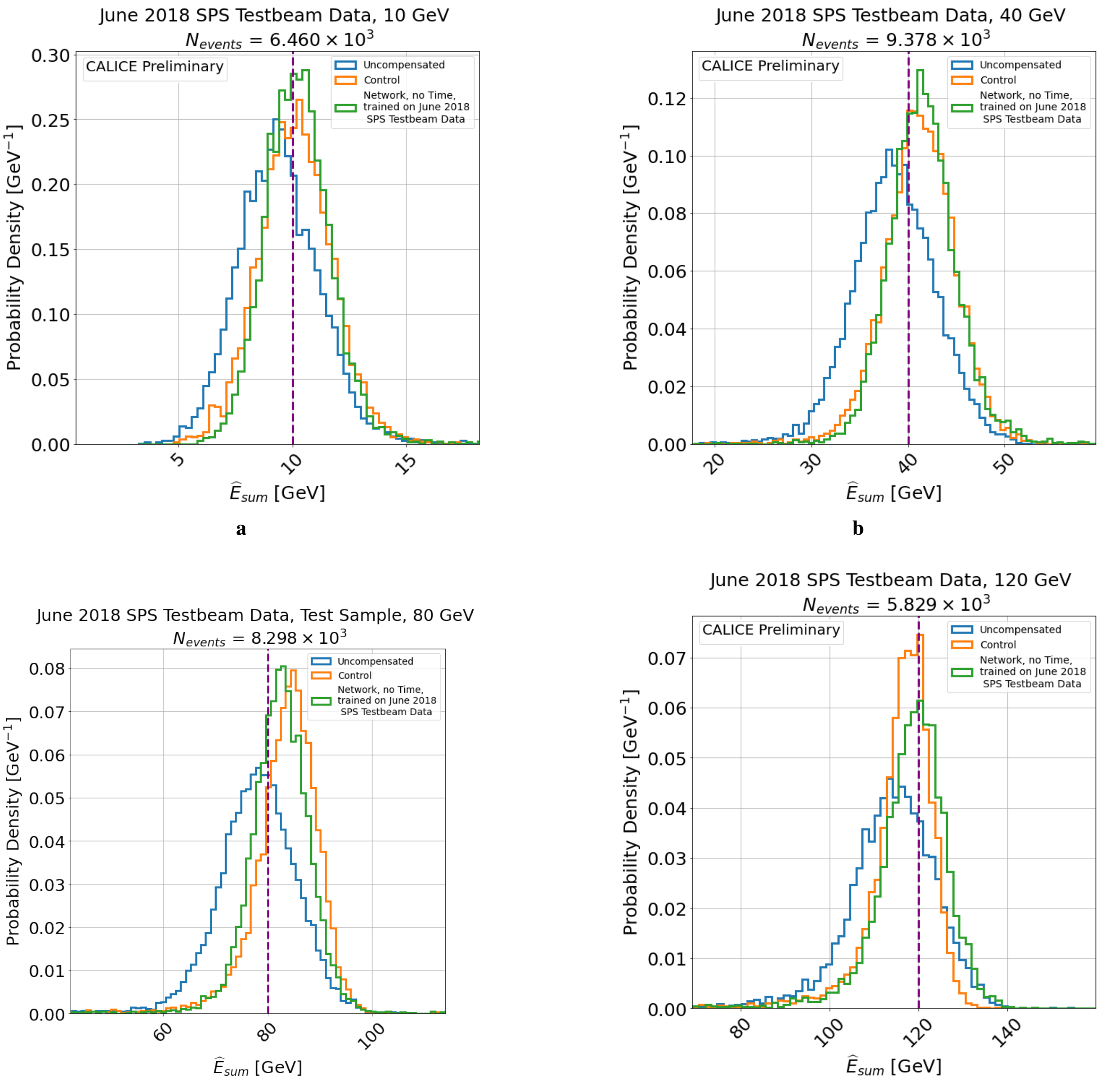


Figure 8: Example normalised histograms showing the calorimeter response before and after compensation applied to the simulated test dataset of Table 1, without the TCMT cut applied. Else, as in Fig. 7.



a

b

c

d

Figure 9: Example normalised histograms showing the calorimeter response before and after compensation applied to the 2018 June Testbeam test dataset of Table 1, with the TCMT cut applied. Samples of 10 GeV, 40 GeV, 80 GeV and 120 GeV hadron shower energies are shown. Else, as in Fig. 7.

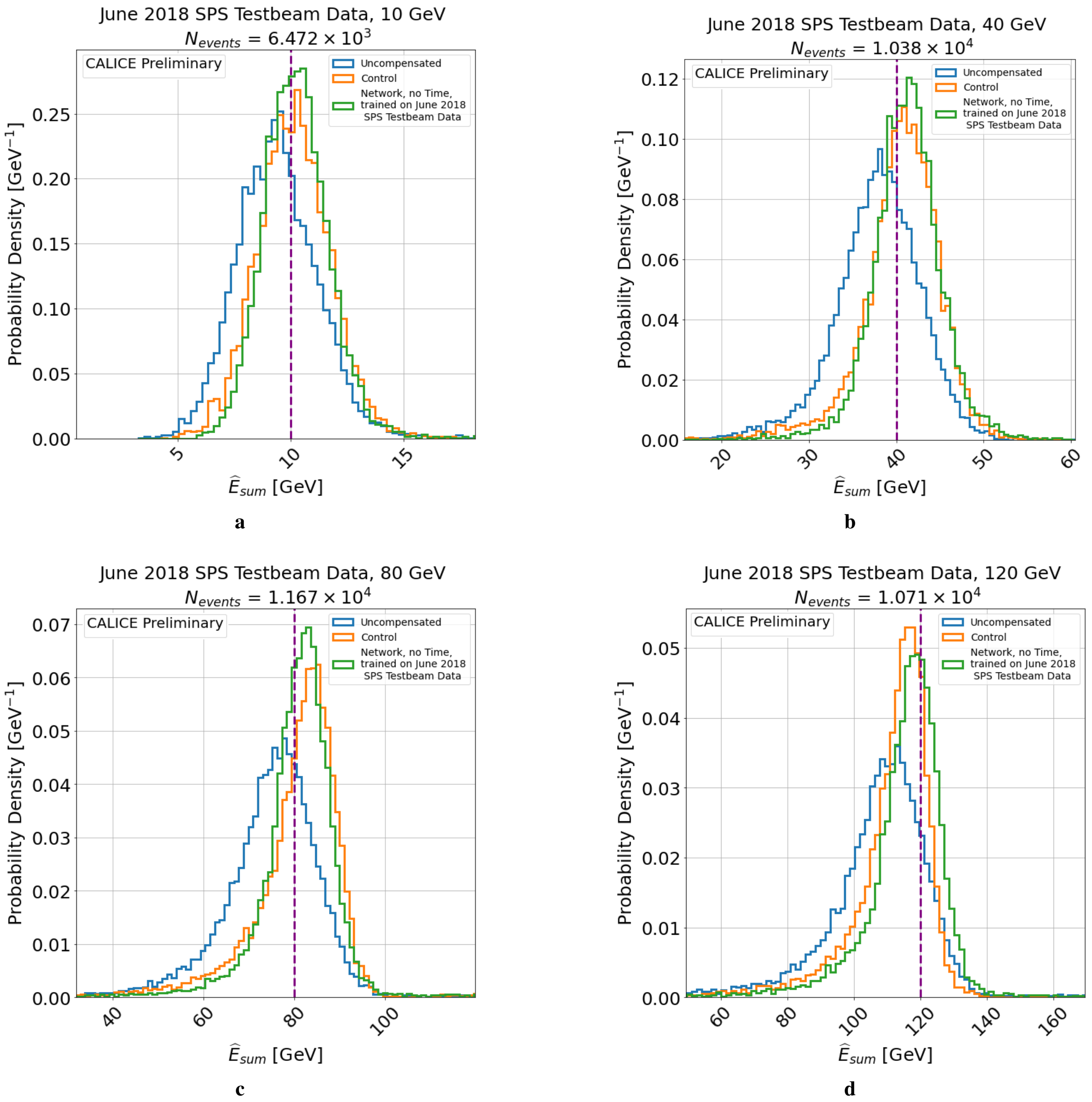


Figure 10: Example normalised histograms showing the calorimeter response before and after compensation applied to the 2018 June Testbeam test dataset of Table 1, without the TCMT cut applied. Else, as in Fig. 7 and the same selected energy samples as in Fig. 9.

Similar conclusions can be drawn for Figs. 8a-8d, indicating that the bias from the TCMT cut does not significantly influence the outcome of the experiment.

2018 Testbeam Data As in Fig. 7, the neural network method produces superior resolution than the control in Fig. 9a-Fig. 9c except for the $120\,\mathrm{GeV}$ sample shown in Fig. 9d. This observation can be attributed to the same energy biasing observed for the $80\,\mathrm{GeV}$ sample in simulation in Fig. 7c, as these are both the maximum energy bins of the training dataset in both cases.

Again, similar conclusions can be drawn for Figs. 10a-10d, once again indicating that the bias from the TCMT cut does not significantly influence the outcome of the experiment.

3.2 Resolution and Linearity of Response

The energy response distributions for each particle energy in the testing dataset, with the TCMT cut applied, were fitted with a normal distribution in the range of ± 2 standard deviations from their mean. The location and scale parameters of the fit, μ and σ, were used to estimate E and σ_E of Eq. 1 and used to study resolution (σ/μ vs. E_{particle}) and linearity of response (μ/E_{particle} vs. E_{particle}).

Simulation Fig. 11 and Table 3 show the measured resolution, fitted with Eq. 1, and the corresponding fit values, respectively, for the simulation.

The top subplot of Fig. 11 indicates the neural network methods offer improved linearity of response compared to the control, which overestimates the hadron shower energy by up to $5\,\%$ compared to 2-3 % for the network methods for most of the training range of particle energies. Moreover, the network and control methods are demonstrated to interpolate within the training range. However, the control method fails to reconstruct the particle energy entirely for particle energies greater than $80\,\text{GeV}$.

The middle and bottom subfigure of Fig. 11 demonstrates that for values of E_{particle} up to around $60\,\text{GeV}$, the neural network methods produce superior compensation, indicated by the smaller value of the compensated response to the intrinsic response. Beyond this range, the resolution produced by the control method diverges from the model of Eq. 1. For this reason, the fit to this method was only performed for E_{particle} in the range 10-60 GeV. By contrast, the uncompensated and network methods show good agreement with the expectations of Eq. 1 and were fitted over the entire range. Table 3 shows the fitted parameters indicated by the dashed coloured lines of the middle subplot of Fig. 11. The uncompensated stochastic resolution for simulated π^- hadron showers in AHCAL is in agreement within 1-2 % with the $a = 51.7 \pm 0.97\,\%$ obtained in the study of [19]. The neural network solutions improve the calorimeter's stochastic resolution, a, by comparison by around $3\,\%$ without timing information and a further $3\,\%$ with timing information, compared to the control method. This result demonstrates the improvement in SC performance that can be obtained from including spatiotemporal energy density information. This result agrees with a similar study on the additional benefit of using timing information for software compensation with AHCAL, which observed a 3-4 % improvement in energy resolution using timing information than energy density information only.

June 2018 Testbeam Data Fig. 12 and Table 4 show the measured resolution, fitted with Eq. 1, and the corresponding fit values, respectively, for the data.

As for simulation, the uncompensated stochastic resolution for data π^- hadron showers in AHCAL is in agreement within 1-2 % with the stochastic resolution term $a = 57.70 \pm 1.06\,\%$ obtained in [19]. Furthermore, the machine learning methods outperform the control method in resolution, resulting in an improvement of the intrinsic stochastic resolution term of $9.3\,\%$ and $12.2\,\%$, outperforming the control method in both cases. The neural networks also reduce the constant resolution term by $2\,\%$, indicating the neural networks perform some detector calibration and SC. A slightly superior linearity of response overall is observed compared to the control method and less than for simulation.

3.3 Correlations with Spatial and Temporal Information

The spatial and energy-temporal correlations of the hadron shower weighting are analysed to study and compare the neutral network SC methods to the control SC method.

Simulation The results for the spatial correlations ($R_{\text{hit}}, K_{\text{hit}} - K_S$) and energy-temporal correlations ($t_{\text{hit}}, E_{\text{hit}}$) are shown in the left and right columns of Fig. 13, respectively to the test sample. The colour axes indicate the percentage change of the energy due to the SC algorithm as a function of these variables. In this example, the tail-catcher cut was not applied.

Table 3: Table of fitted parameters of Eq. 1 to the training range of energies of simulation shown as dashed lines in Fig. 11, except for the control method, which was fitted up to $60\,\text{GeV}$ due to the effect of energy biasing.

	a [%]	b [%]	χ^2/NDF
Uncompensated	49.5 ± 0.4	7.1 ± 0.1	4.6
Control	43.4 ± 0.1	0.0 ± 2.9	14.3
Network, No Time	40.2 ± 0.2	2.2 ± 0.1	0.9
Network, + Time	37.3 ± 0.2	2.4 ± 0.1	1.4

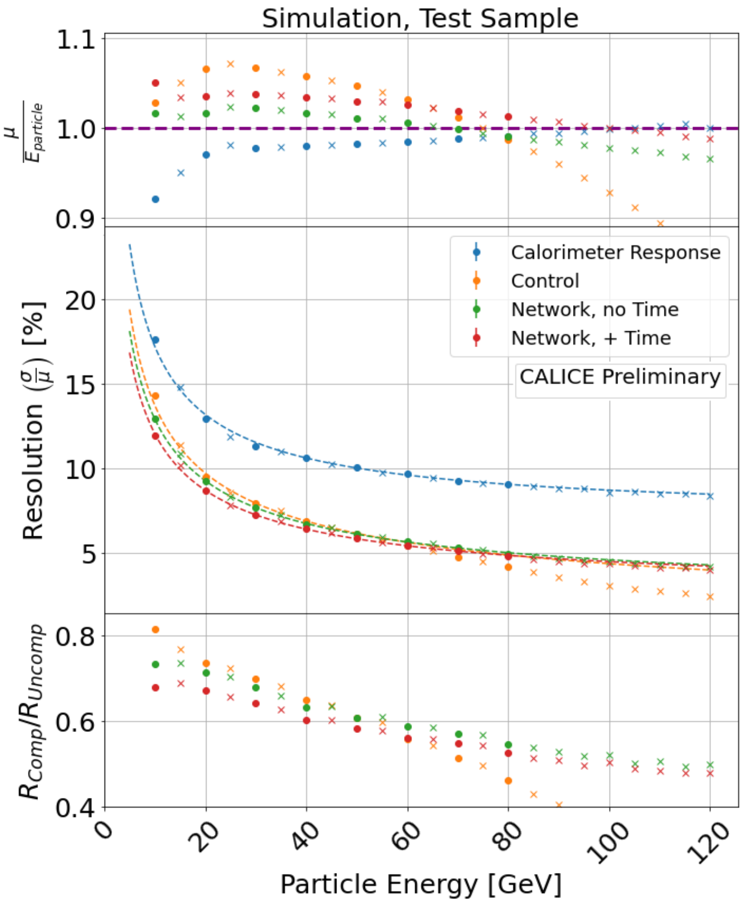


Figure 11: AHCAL linearity of response and resolution using all methods under test applied to the test dataset of simulation. Blue indicates intrinsic calorimeter response, while orange, green and red indicate the control, network without and network with time compensation methods. Circles indicate energies used for both training and testing, and cross markers indicate energies used for testing only. The top subplot shows the ratio of fitted μ to E_{particle}, where the dashed purple line indicates $\mu = E_{\text{particle}}$ The middle subplot shows the fitted σ/μ, where the dashed lines indicate fits of Eq. 1. The bottom subplot indicates the ratio of the resolution of each compensation method to the intrinsic response.

Table 4: Table of fitted parameters of Eq. 1 to the training range of energies of 2018 CALICE Testbeam data shown as dashed lines in Fig. 11. The whole range of available energies was used to fit.

	a [%]	b [%]	χ^2/NDF
Calorimeter Response	56.1 ± 0.7	6.1 ± 0.1	10.1
Control	51.5 ± 0.42	1.0 ± 0.3	38.9
Network, No Time	41.9 ± 0.5	4.0 ± 0.1	6.5

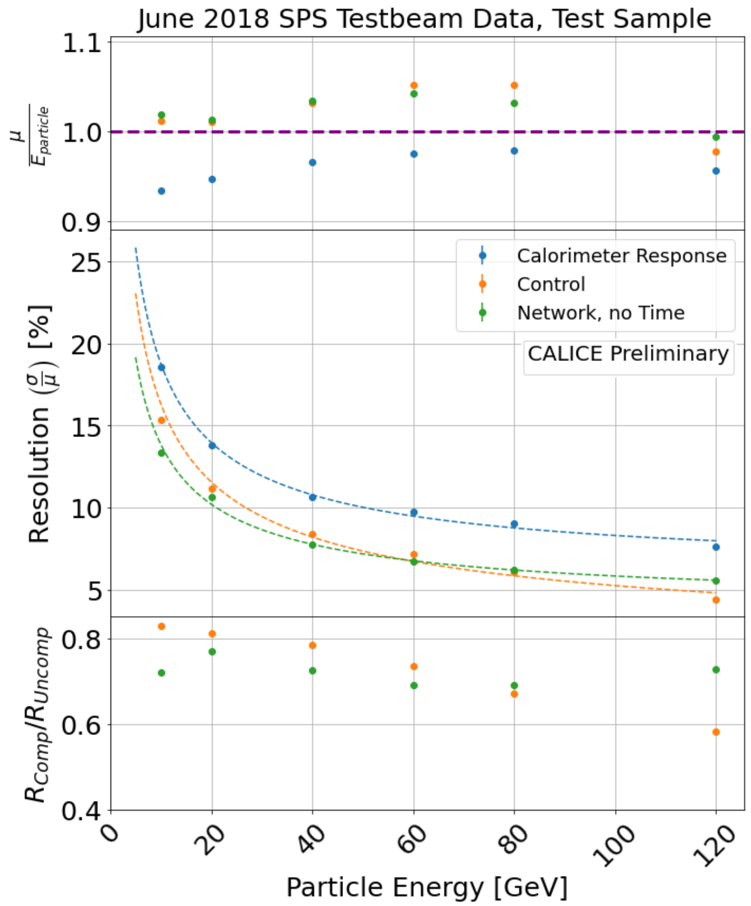


Figure 12: AHCAL linearity of response and resolution using all methods under test applied to the test dataset of 2018 CALICE Testbeam Data. Else, as in Fig. 11.

Fig. 13a demonstrates that the control method shows only a weak dependence on lateral and longitudinal development of the shower, with attenuation occurring only within $R_{\text{hit}} \lesssim 1\,\rho_M$ (the EM fraction) and enhancement beyond, with minor variation, as expected. By contrast, the neural network methods attenuate and enhance the active cell energy with much stronger spatial dependence, indicated by the broadening of the weighting profile with longitudinal shower development. Two additional effects are observed for the network methods, shown in Fig. 13c and Fig. 13e: a tendency to enhance E_{hit} in the region above the white dashed line, and to attenuate E_{hit} where $R_{\text{hit}} \lesssim 1\rho_M$ (close to the lateral shower core) and $K_S < 0$ (before the shower start). These effects are not present in Fig. 13a and must therefore be a consequence of including spatial information in the models. These results suggest the network models have learned leakage correction and to remove the energy deposited by minimum ionisation of the π^- particle before showering. This result demonstrates an improved capacity of the proposed model to learn the physical properties of the hadron shower and detector compared to the control method.

Fig. 13b, Fig. 13d and Fig. 13f demonstrate that all methods are observed to attenuate active cell energies above 5 MIP and enhance below that threshold, which is expected of all SC algorithms. The binned structure of the weighting of the control method is visible in Fig. 13b. By contrast, the neural network methods in Fig. 13d and Fig. 13f indicate a continuous weighting function has been learned. Furthermore, Fig. 13f indicates that the model with timing information enhances the threshold for energy deposited in the order of several ns to several tens of ns. A reduction in the threshold is observed after around 100 ns. These observations are consistent with the timescales of the two main neutron energy-depositing processes discussed in Section 2. Comparison of Fig. 13f and Figs. 13b- 13d indicate that this effect must be due to the inclusion of timing information since no such effect is observed in the control or method without timing information.

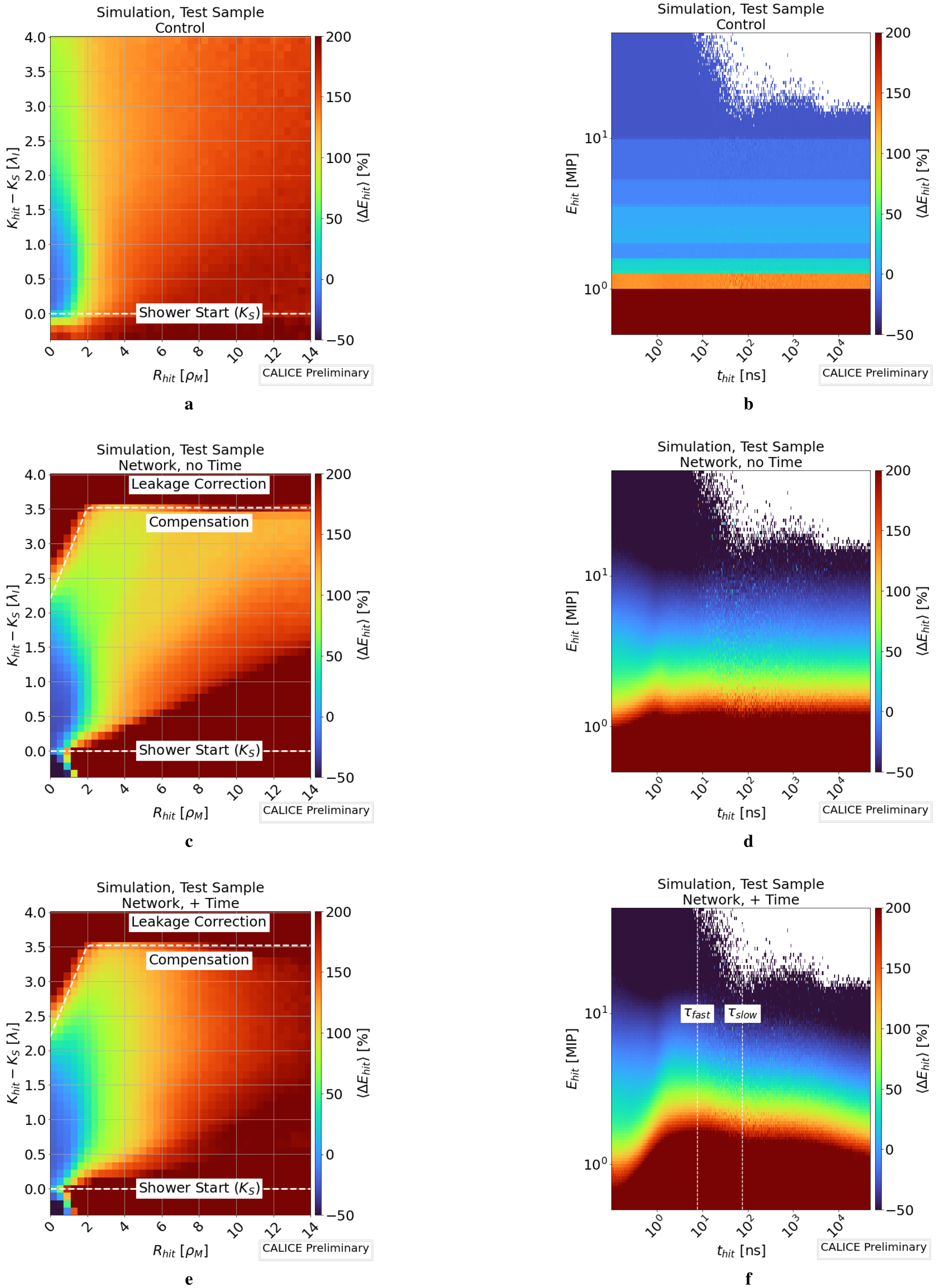
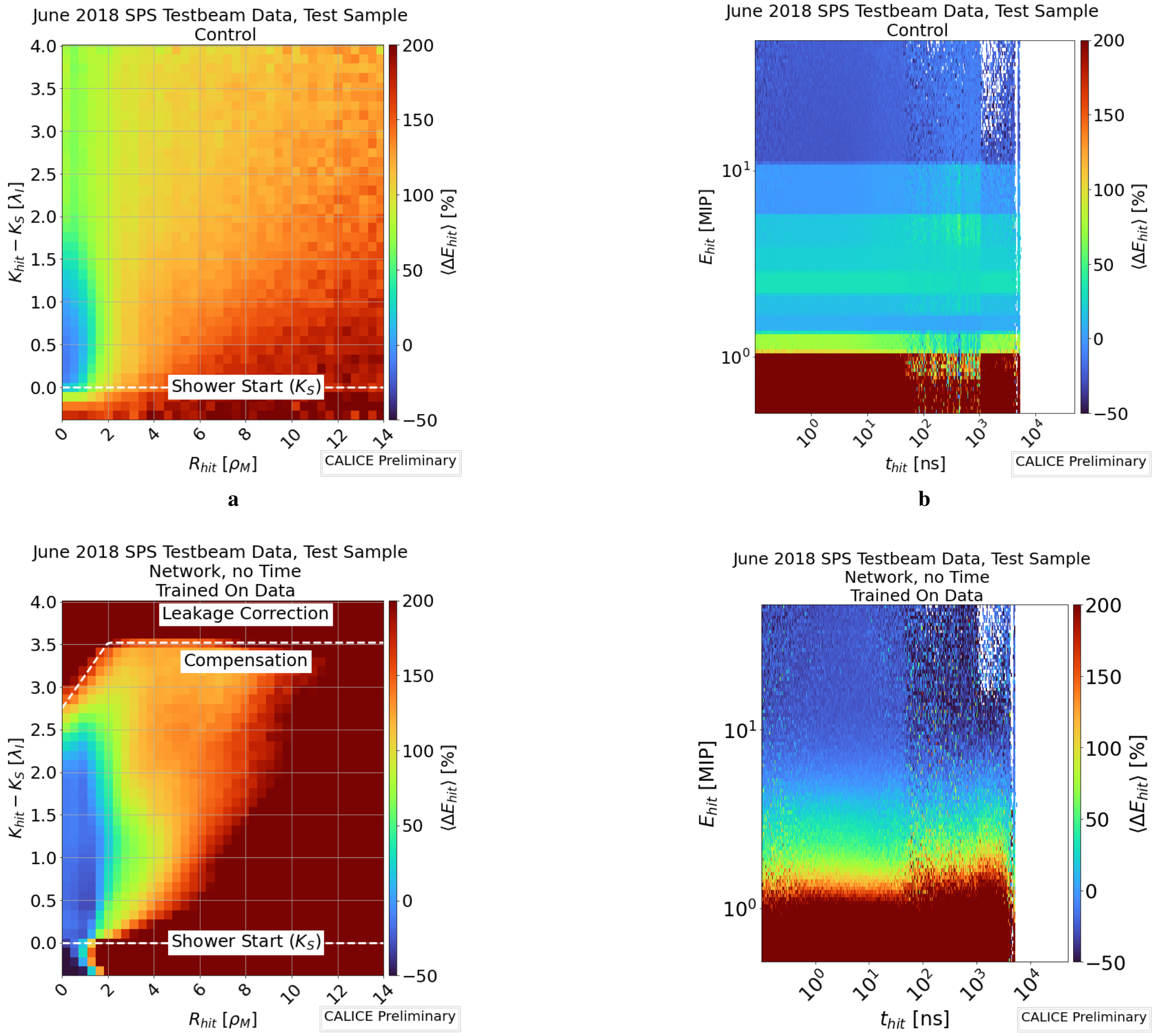


Figure 13: Average percentage change in active cell energy (E_{hit}) as a result of compensation as a function of R_{hit} and $K_{\text{hit}} - K_S$ (left column, presented in units of Moliere radius from the lateral center-of-gravity, $\rho_M = 24.9\,\text{mm}$, and nuclear interaction length from the shower start, $\lambda_I = 237.1\,\text{mm}$, respectively), and E_{hit} and t_{hit} (right column, presented in units of MIP and ns, respectively) for simulation. Each row indicates the control and network methods without and with timing information in that order. The colour axis indicates the percentage change, where blue regions indicate where the energy has been attenuated, and green through red shows where the energy has been enhanced. White space indicates no data available. Regions of interest are labelled accordingly for reference. The values of τ_{slow} and τ_{fast} were taken from [4].



a

b

c

d

Figure 14: Average percentage change in active cell energy (E_{hit}) as a result of compensation as a function of R_{hit} and $K_{\mathrm{hit}} - K_S$ as in Fig. 13 for CALICE 2018 Testbeam Data. Else, as in Fig. 13.

CALICE 2018 Testbeam Data The spatial and energy-temporal correlations in Fig. 14 show the same information for data as in Figs. 13a-13d. The conclusions to this figure are the same as in Fig. 13 for the control method and the neural network model trained only with spatial information.

4 Conclusion

A neural network method for performing software compensation was devised, trained, and tested on simulation and 2018 June Testbeam data for the AHCAL calorimeter. The model used a local energy density estimate to overcome biasing effects on particle energies.

The neural network model was trained with and without timing information with $100\,\mathrm{ps}$ timing resolution and is compared to a control method after accounting for the effect of leakage compensation learned by the networks. The neural networks yielded superior overall compensation and linearity of response to the control method when trained on simulation, resulting in calorimeter resolutions of $40.2\,\%/\sqrt{E_{\mathrm{particle}}} \oplus 2.2\,\%$ and $37.2\,\%/\sqrt{E_{\mathrm{particle}}} \oplus 2.4\,\%$. This corresponded to an absolute improvement of stochastic resolution, a, by $9.3\,\%$ and $12.2\,\%$, or a relative improvement of $19\,\%$ and $25\,\%$ respectively compared to the uncompensated a. The absolute value of the constant resolution term, b, was also found to reduce by around $5\,\%$ or a relative improvement of $70\,\%$ compared to the uncompensated b in both models. This result indicates that the model was capable of detector calibration. Both methods obtained a linearity of response within around 2-3 $\%$ of the particle energy. This result should be interpreted with the caveat of the bias caused by the TCMT cut. Nonetheless, improved performance was observed compared to the control using the neural network method without the TCMT cut.

The network without timing information trained on CALICE 2018 SPS testbeam data achieved a comparable resolution of $41.9\,\%/\sqrt{E_{\mathrm{particle}}} \oplus 4.0\,\%$. This corresponded to an absolute improvement in a by $14.2\,\%$ or a relative improvement of $25\,\%$. Additionally, this corresponded to an absolute improvement in b by around $2\,\%$ or a relative improvement of around $35\,\%$. This result indicates that the model can be trained with limited experimental data to a similar level as simulation. Additionally, the control method was observed to bias to the training range of particle energies. In contrast, the neural network method was demonstrated to both interpolate and extrapolate compensation to energies not used for training.

The networks' learned method of applying SC agreed with expectations: the attenuation of high-energy (EM) deposits and the enhancement of low-energy deposits. However, the network method was found to apply SC differently depending on the stage of the shower development, both in space and in time, the former of which included an effect consistent with longitudinal leakage correction and the latter of which was found to agree with expectations of a bi-exponential time distribution for energy deposits in a steel-scintillator calorimeter expected from Ref. [4]. Similar behaviour was observed in the independent neural network applied to data.

In summary, this study indicates that superior resolution can be obtained in highly granular calorimeters using spatiotemporal event information and neural networks and that careful model design can overcome the limitations of previous data-driven compensation techniques by reducing energy biasing. The validation of the method as part of a full Particle Flow analysis represents a promising future study.

Appendix

Table 5: Bin ranges and weights obtained for the control method described in Section 2.2. Table 5a shows the values obtained for simulation. Table 5b shows the values obtained for CALICE June 2018 SPS Testbeam data.

a

	Simulation		
Bin Range [MIP]	α_{b}	β_{b}	γ_{b}
0.500 - 0.735	-1.120	13.000	-5.499
0.735 - 1.002	-0.372	5.134	-4.727
1.002 - 1.272	-0.350	2.523	-3.054
1.272 - 1.585	-0.460	1.999	-1.587
1.585 - 2.013	-0.290	1.634	-1.012
2.013 - 2.631	0.022	1.275	-0.949
2.631 - 3.584	0.218	1.138	-0.632
3.584 - 5.328	0.387	0.658	-0.457
5.328 - 9.881	0.579	0.259	-0.281
9.881 - ∞	0.788	-0.057	0.043

b

	June 2018 SPS Testbeam Data		
Bin Range [MIP]	α_{b}	β_{b}	γ_{b}
0.500 - 0.770	-9.252	17.803	-16.047
0.770 - 1.059	-8.529	14.738	-11.256
1.059 - 1.351	-0.046	1.902	-1.797
1.351 - 1.698	1.283	-0.379	0.057
1.698 - 2.179	2.104	-1.325	0.720
2.179 - 2.875	1.025	0.384	-0.235
2.875 - 3.957	1.271	0.001	0.104
3.957 - 5.847	1.325	0.013	0.304
5.847 - 10.930	0.814	0.426	0.021
10.930 - ∞	0.148	1.032	-0.243

Table 6: Table of μ and σ from the Gaussian fits performed on the SC models trained on simulation. Tables 6a and 6b show the μ, σ and their errors as a function of particle energy for each studied method applied to the testing dataset. CR, CTRL, NN,-Time and NN,+Time are abbreviations of: 'intrinsic calorimeter response', 'control method', 'neural network, without time' and with 'neural network, with time', respectively.

a

	Simulation							
	μ				dμ			
	CR	CTRL	NN,-Time	NN,+Time	CR	CNTRL	NN,-Time	NN,+ Time
E_{particle} [GeV]								
10	9.214	10.281	10.157	10.502	0.013	0.012	0.010	0.009
15	14.257	15.754	15.186	15.512	0.016	0.014	0.012	0.012
20	19.401	21.307	20.326	20.692	0.019	0.015	0.013	0.013
25	24.516	26.796	25.569	25.954	0.021	0.016	0.015	0.014
30	29.330	31.990	30.661	31.136	0.024	0.018	0.017	0.016
35	34.230	37.160	35.688	36.247	0.027	0.020	0.018	0.018
40	39.199	42.311	40.656	41.368	0.030	0.021	0.019	0.019
45	44.148	47.374	45.651	46.468	0.034	0.023	0.022	0.021
50	49.119	52.316	50.526	51.466	0.034	0.022	0.021	0.021
55	54.066	57.157	55.555	56.583	0.043	0.028	0.027	0.026
60	59.033	61.864	60.352	61.535	0.045	0.027	0.026	0.026
65	64.099	66.412	65.148	66.438	0.045	0.026	0.026	0.026
70	69.145	70.838	69.905	71.315	0.057	0.031	0.032	0.032
75	74.178	74.941	74.578	76.158	0.066	0.035	0.037	0.037
80	79.235	78.972	79.270	80.992	0.061	0.029	0.033	0.032
85	84.506	82.781	83.912	85.779	0.064	0.029	0.034	0.033
90	89.475	86.410	88.615	90.592	0.063	0.027	0.033	0.032
95	94.663	89.725	93.157	95.238	0.078	0.031	0.039	0.039
100	99.804	92.822	97.723	99.969	0.088	0.034	0.044	0.044
105	104.983	95.722	102.337	104.685	0.095	0.032	0.047	0.047
110	110.224	98.338	106.931	109.416	0.103	0.031	0.050	0.049
115	115.567	100.585	111.316	113.949	0.106	0.028	0.050	0.049
120	120.000	102.513	115.874	118.618	0.010	0.028	0.053	0.052

b

	Simulation							
	σ				dσ			
	CR	CTRL	NN,-Time	NN,+Time	CR	CNTRL	NN,-Time	NN,+ Time
E_{particle} [GeV]								
10	1.625	1.474	1.311	1.255	0.012	0.011	0.009	0.008
15	2.112	1.792	1.651	1.579	0.015	0.012	0.010	0.010
20	2.512	2.027	1.873	1.795	0.018	0.013	0.011	0.011
25	2.905	2.294	2.128	2.016	0.019	0.013	0.013	0.012
30	3.320	2.527	2.351	2.258	0.021	0.015	0.014	0.013
35	3.754	2.779	2.576	2.487	0.023	0.016	0.015	0.015
40	4.151	2.905	2.721	2.642	0.025	0.017	0.016	0.015
45	4.518	3.081	2.961	2.858	0.027	0.018	0.018	0.018
50	4.946	3.192	3.092	3.023	0.027	0.018	0.017	0.017
55	5.253	3.316	3.296	3.182	0.033	0.022	0.023	0.022
60	5.701	3.335	3.416	3.328	0.034	0.021	0.022	0.022
65	6.059	3.406	3.598	3.510	0.034	0.021	0.022	0.022
70	6.406	3.367	3.695	3.628	0.044	0.023	0.027	0.027
75	6.736	3.376	3.840	3.751	0.051	0.027	0.032	0.032
80	7.174	3.297	3.917	3.862	0.046	0.022	0.027	0.027
85	7.540	3.167	4.021	3.925	0.048	0.021	0.029	0.029
90	7.843	3.078	4.111	4.042	0.045	0.020	0.028	0.028
95	8.297	2.943	4.237	4.141	0.059	0.021	0.033	0.034
100	8.553	2.803	4.354	4.318	0.062	0.023	0.036	0.037
105	9.049	2.741	4.425	4.422	0.071	0.023	0.039	0.039
110	9.341	2.664	4.588	4.488	0.078	0.024	0.044	0.041
115	9.832	2.603	4.666	4.643	0.081	0.026	0.040	0.041
120	10.000	2.493	4.813	4.742	0.027	0.028	0.045	0.044

Table 7: Table of μ and σ from the Gaussian fits performed on the SC models trained on data. Tables 7a, 7b μ, σ and their errors as a function of particle energy for each studied method applied to the testing dataset. Else, as in Table 6. 'Sim' and 'Data' indicate models trained on simulation and 2018 Testbeam data, resepctively.

a

	June 2018 SPS Testbeam Data							
	μ				dμ			
	CR	CTRL	NN,-Time	NN,-Time	CR	CNTRL	NN,-Time	NN,-Time
E_{particle} [GeV]		(Data)	(Data)	(Sim)		(Data)	(Data)	(Sim)
10	9.334	10.113	10.190	9.949	0.025	0.022	0.018	0.019
20	18.947	20.213	20.244	19.048	0.031	0.026	0.025	0.024
40	38.623	41.278	41.355	38.301	0.047	0.038	0.036	0.033
60	58.501	63.090	62.503	57.331	0.060	0.047	0.043	0.038
80	78.246	84.081	82.514	75.260	0.081	0.062	0.058	0.049
120	114.690	117.347	119.229	107.497	0.121	0.080	0.092	0.077

b

	June 2018 SPS Testbeam Data							
	σ				dσ			
	CR	CTRL	NN,-Time	NN,-Time	CR	CNTRL	NN,-Time	NN,-Time
E_{particle} [GeV]		(Data)	(Data)	(Sim)		(Data)	(Data)	(Sim)
10	1.733	1.554	1.363	1.407	0.023	0.019	0.016	0.016
20	2.614	2.260	2.151	2.094	0.029	0.023	0.022	0.021
40	4.124	3.456	3.198	3.012	0.042	0.033	0.031	0.029
60	5.702	4.526	4.202	3.750	0.050	0.039	0.036	0.031
80	7.058	5.093	5.144	4.381	0.065	0.046	0.048	0.040
120	8.734	5.203	6.611	5.532	0.096	0.062	0.077	0.059

Acknowledgments

We would like to thank the technicians and the engineers who contributed to the design and construction of the CALICE AHCAL prototype detector. We also gratefully acknowledge the CERN management for its support and hospitality and its accelerator staff for the reliable and efficient operation of the test beam. The authors acknowledge the support from the BMBF via the High-D consortium. This work is supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy, EXC 2121, Quantum Universe (390833306).

References

[1] M. Thomson, "Particle Flow Calorimetry and the PandoraPFA Algorithm" NIM A A247, pp. 25–40, November 2009, 0907.3577

[2] F. Sefkow et al. "A highly granular SiPM-on-tile calorimeter prototype" J. Phys.: Conf. Ser. 1162 012012, January 2019, 10.1088/1742-6596/1162/1/012012

[3] R. Wigmans, "Calorimetry: Energy Measurement in Particle Physics". International Series of 481 Monographs on Physics, Oxford University Press, Second Edition

[4] C. Adloff, J.-J.Blaising, M, Chefdeville et. al. "The Time Structure of Hadronic Showers in highly granular Calorimeters with Tungsten and Steel Absorbers", J. Inst. Vol. 9 pp.7022, July 2018, 1404.6454

[5] M. Aleksa et al. "Calorimeters for the FCC-hh", December 2019. 1912.09962

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

[7] The CALICE Collaboration, "Software Compensation for Hadronic Showers in the CALICE AHCAL and Tail Catcher with Cluster-based Methods", CALICE Analysis Note CAN-021, April 2010

[8] C. Adloff et al. "Construction and performance of a silicon photomultiplier/extruded scintillator tail-catcher and muon-tracker", J. Inst. Vol 7, April 2012, 10.1088/1748-0221/7/04/P04015

[9] The CALICE Collaboration, ILD Concept Group, "Calibration of the scintillator hadron calorimeter of ILD", CALICE Analysis Note CAN-018, October 2009

[10] The CALICE Collaboration, "Design, Construction and Commissioning of a Technological Prototype of a Highly Granular SiPM-on-tile Scintillator-Steel Hadronic Calorimeter", J. Inst. submitted for publication, September 2022, 2209.15327

[11] The CMS Collaboration and CALICE Collaboration, "Performance of the CMS High Granularity Calorimeter prototype to charged pion beams of 20−300 GeV/c", J. Inst. Vol 18, August 2023, 10.1088/1748-0221/18/08/P08014

[12] Y. Wang, Y.Sun, Z.Liu et al. "Dynamic Graph CNN for Learning on Point Clouds", CoRR, June 2019, 1801.07829

[13] T. Akiba, S. Sano, T. Yanase et. al., "Optuna: A Next-generation Hyperparameter Optimization Framework", KDD '19, July 2019, 1907.10902

[14] J. Respond, L. Xia, J. Apostolakis et. al. "Hadronic Energy Resolution of a Combined High Granularity Scintillator Calorimeter System" J. Inst. Vol. 13 pp 12022, September 2018, 1809.03909

[15] S. Agostinelli, J. Allison, K. Amako et. al. "Geant4—a simulation toolkit" NIM A 13 pp 250–303 Vol. 506, July 2003

[16] M Petrič et. al., "Detector Simulations with DD4hep", J. Phys.: Conf. Ser., Vol. 898, pp 042015, February 2017

[17] The CALICE Collaboration, "CALICESoft" [Online] Available: https://stash.desy.de/projects/CALICE

[18] L. Emberger, "Precision Timing in Highly Granular Calorimeters and Applications in Long Baseline Neutrino and Lepton Collider Experiments", PhD Thesis, Technische Universität München, August 2022

[19] O. Pinto. "Shower Shapes in a Highly Granular SiPM-on-Tile Analog Hadron Calorimeter", PhD Thesis. University of Hamburg, August 2022

[20] V. Bocharnikov "Particle identification methods for the CALICE highly granular SiPM-on tile calorimeter". Verhandlungen der Deutschen Physikalischen Gesellschaft, (Aache2019issue), pp 1, February 2019

[21] A. Paszke et. al., "Automatic differentiation in PyTorch", NIPS 2017 Workshop Autodiff, October 2017

[22] W. Falcon et. al., "PyTorch Lightning" [Online] Available: https://www.pytorchlightning.ai/

[23] F. James and M. Winkler, "MINUIT User's Guide", June 2004

[24] D. Freedman and P. Diaconis, "On the histogram as a density estimator:L2 theory", Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete, Vol. 57, pp 453–476, December 1981, BF01025868