CALICE Analysis Note CAN-035
September 29, 2011

Local and global software compensation approaches: application to test beam data

The CALICE Collaboration [1]

Abstract

In this note, we present improved results for the energy reconstruction of pion events in the CALICE AHCAL, using local and global software compensation. Both techniques are applied to the same event samples selected from the CERN 2007 test beam period, covering an energy range from 10 GeV to 80 GeV. The run and event selection procedures have been improved and are identical for both methods. Both local and global software compensation techniques show similar improvement in single particle energy resolution ranging from 12% to 25%, depending on particle energy.

[1] Corresponding authors:
Marina Chadeeva; marina.chadeyeva@itep.ru
Katja Seidel; kseidel@mpp.mpg.de

Contents

1 Energy reconstruction in the AHCAL

The energy resolution for hadrons is one of the key performance parameters for the CALICE analog hadron calorimeter. To improve the energy resolution, two methods of software compensation were developed based on local and global approaches that are described in detail in the notes CAN-015 [1] and CAN-028 [2], respectively. Both methods are based on the event-by-event analysis of the energy density spectrum obtained from the signals of individual calorimeter cells. At the same time the approaches follow different ways to construct and apply compensation factors: in the local approach there are several weights applied to the signals of individual calorimeter cells, while in the global approach one compensation factor is applied to the energy sum calculated over all calorimeter.

Both software compensation algorithms use information on the shower substructure to compensate fluctuations in the electromagnetic content of the showers and have demonstrated significant improvement of the energy resolution. In this note, the data collected during CERN 2007 test beam campaign were analyzed. In this data taking period, CALICE test beam setup was equipped by silicon-tungsten ECAL, scintillator-steel analogous HCAL and scintillator-steel tail catcher and muon tracker (TCMT). To provide a direct comparison of the different compensation strategies, we present updated results with both techniques applied to the same event samples. The run and event selection procedures have been improved and are used identically for both techniques. The reconstruction and calibration were performed with the standard CALICE calibration chain (CALICE software version v04-01) followed by dedicated event selection procedure that includes muon rejection and shower start finding (see subsection 2.2 for details).

The total deposited energy is obtained at the "electromagnetic scale" from the visible signal measured in different detector sections, multiplied with suitable calibration factors. The conversion factors from the visible signal in units of minimum-ionizing particles (MIP) to the total energy at the electromagnetic scale in units of GeV are given in table 1. The Si-W ECAL comprises three sections with different sampling fractions [3], the AHCAL has the same sampling fraction for all 38 layers [4]. The TCMT consists of two sections, the first section having the same sampling fraction as AHCAL; the absorber thickness of the second TCMT section is by factor of 5 larger than in its first section [5]. During selection procedure, the conversion factors w were used to calculate the energy deposited in the ECAL[2], while for selected events with minimum-ionizing track in ECAL factors v were applied. The conversion factor v from visible signal to deposited energy in the ECAL for non-showering hadrons was estimated using simulated muons in the ECAL and the measured muon response from test beam runs [6]. For the conversion to the hadronic energy scale, the $\frac{e}{\pi}$ factor has to be considered in addition to calculate a reconstructed energy.

[2]The electromagnetic calibration coefficients for ECAL are based on the CERN 2006 test beam data. For CERN 2007 data they might differ by several percent from indicated values. In the present analysis, the ECAL is effectively being used as a veto, because showers are required to start in the AHCAL, so a precise calibration of the ECAL at the electromagnetic scale is not essential.

Table 1: Conversion coefficients from visible signal to deposited energy, on the electromagnetic scale. In the last column, the coefficients for the conversion of visible signal to the energy loss of minimum ionizing particles in ECAL are given, which are applied for events with track in ECAL.

Subdetector	Number of subsection	Electromagnetic scale w, $\frac{GeV}{MIP}$	MIP scale v, $\frac{GeV}{MIP}$
ECAL	1	0.00376	0.002953
ECAL	2	0.00752	0.005906
ECAL	3	0.01128	0.008859
AHCAL		0.02364	-
TCMT	1	0.02364	-
TCMT	2	0.11820	-

2 Run and event selection

The software compensation techniques described in CAN-015 [1] and CAN-028 [2] were applied to the same samples selected from CERN 2007 test beam data. For this study, updated run and event selection procedures were used identically for both methods.

2.1 Run selection

For the present study, pion runs in the energy range from 10 GeV to 80 GeV with beams at normal incidence around the center of the calorimeter and without known problems were selected from the CERN 2007 data set. The following characteristics were investigated to check the run quality:

- **Longitudinal profiles:** The longitudinal profile for the full setup (ECAL + AHCAL + TCMT) helps to identify noisy or abnormal behavior of individual layers for a given run. One run with extremely noisy TCMT was rejected.
- **MIP calibration stability:** Using muons identified in each run (see section 2.2), the position of the MIP peak, which was observed to be temperature independent, was extracted. For most runs these values are within a $\pm 1\sigma$ interval around the calibration value of 1 MIP, where σ is a statistical error of most probable value estimate. One run with low muon statistics and therefore unreliable MIP peak position estimation was rejected.
- **Pedestal behavior:** The width of the visible signal distribution for pedestal events was monitored on run-to-run basis. Abnormal pedestal behavior was observed for one run taken at highest detected temperature of 29.5°C. This run was excluded from the analysis.

Table 2: List of used data runs.

run number	particle type	beam energy, GeV	run number	particle type	beam energy, GeV
330332	π^-	10	330550	π^-	45
330643	π^-	10	330559	π^-	45
330777	π^-	10	330961	π^-	45
330850	π^-	10	330391	π^-	50
330328	π^-	15	330558	π^-	50
330327	π^-	18	331335	π^+	50
330649	π^-	20	331282	π^+	60
330771	π^-	20	331333	π^+	60
330325	π^-	25	331334	π^+	60
330650	π^-	25	331556	π^-	60
331298	π^+	30	331568	π^-	60
331340	π^+	30	331655	π^-	60
330551	π^-	35	331664	π^-	60
330960	π^-	35	330392	π^-	80
330390	π^-	40	330962	π^-	80
330412	π^-	40	331280	π^+	80
330560	π^-	40	331324	π^+	80
331338	π^+	40	331554	π^-	80
331339	π^+	40	331567	π^-	80
			331654	π^-	80

After these quality assurance procedures, 39 runs listed in table 2 were selected for the software compensation study.

2.2 Event selection

The event selection procedure provides a purification of the pion samples that have an admixture of muons as well as electrons or protons, and is implemented in the HadronSelection processor [7]. To suppress noise, only cells with the visible signal $\geq$0.5 MIP were included in the analysis. Hereinafter such cells are called hits.

To identify muons, a 2D distribution of deposited energy ($E_{\text{dep}}^{\text{ECAL}} + E_{\text{dep}}^{\text{HCAL}}$) vs. $E_{\text{dep}}^{\text{TCMT}}$ is analyzed. All events that have more than 25 hits in the HCAL and more than 10 hits in the TCMT, fall inside triangle with vertexes $(0.6; 0.2)$, $(0.4 \cdot E_{\text{beam}} + 1; 1.3)$ and $(1.3; 0.4 \cdot E_{\text{beam}})$ in the 2D distribution are muon candidates (E_{beam} is a beam energy in GeV). The candidates with track found[3] are considered to be muons. The efficiency of

[3]The procedure of primary track finding is implemented in the PrimaryTrackFinder processor. It is based on the nearest neighbour criterium and uses information about the found starting layer. A normal

muon identification was checked using both simulated muon samples and muon runs from test beam data and was found to be better than 98% [7]. Figure 1 shows an example of constraints applied to 2D histogram for the most problematic case with a muon admixture to 10-GeV pion sample. The red triangle represents the muon selection conditions used. The insert in the upper right contains an example of application of the criteria to the pure muon run. Two examples of muon selection for higher energies can be found in the Appendix A.

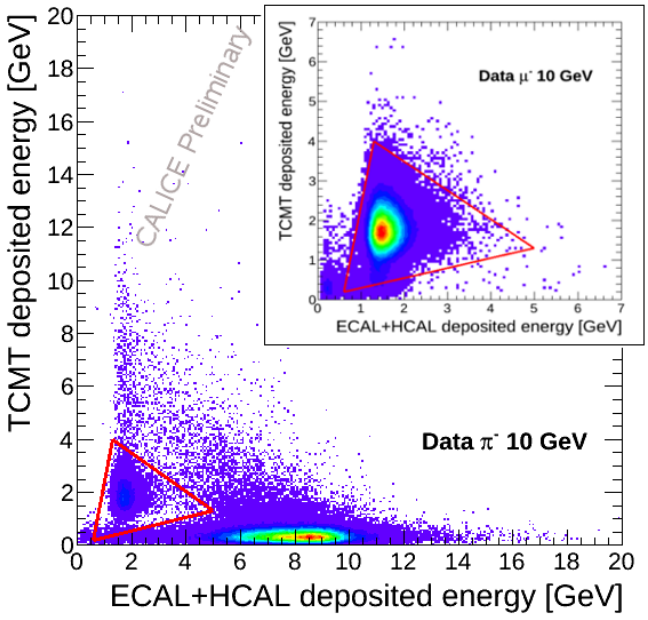

Figure 1: Constraints (red triangle) applied to select muons from 10 GeV π^- sample using the distribution of deposited energy in ECAL+HCAL versus energy deposited in TCMT. In the insert the same constraints applied to 10 GeV muon run are shown. See text for details.

The highest muon admixture of 15 and 30% was observed for 35 GeV and 30 GeV beams, respectively, while for other energies it did not exceed 7%. The purity of pion samples from muons was estimated to be better than 0.5% for all energies.

The Čerenkov counter is used to remove electrons from the π^- samples and protons from the π^+ samples. Events with the total deposited energy higher than $E_{\rm beam} + 2.4\sqrt{E_{\rm beam}}$ ($E_{\rm beam}$ in GeV) as well as events with several parallel ingoing tracks were considered as multi-particle events and are excluded from the analysis. The fraction of such events was <1% below and <2% above 50 GeV. Events with the total deposited energy lower than $0.1 \cdot E_{\rm beam}$ were also rejected, their fraction being less than 0.5%.

incidence is assumed for ingoing tracks and only tracks longer than 4 layers can be identified. Starting from hits in the ECAL front layer, the hits (one per layer) which belong to a primary track are successively found up to the starting layer. For muon candidates starting layer is assigned to be behind the HCAL. If several parallel track candidates are identified with the transversal displacement larger than 2 tile sizes such event is considered as multipartical.

To analyze the intrinsic AHCAL energy resolution for hadrons, an additional constraint is applied to the purified pion samples: a shower start (a position of the primary inelastic interaction) was required to be in the first five layers of the AHCAL. This requirement allows the effect of leakage into the TCMT to be reduced by selecting hadronic showers which are mostly contained in the AHCAL.

The following algorithm was used to find the layer of primary interaction: the moving average A_i of the visible signal in MIP in ten successive layers up to the i^{th} layer and the number of hits in the i^{th} layer N_i were analyzed on a layer-by-layer basis, starting from the first ECAL layer. If the conditions $(A_i + A_{i+1}) > (6.0 + 0.1\frac{E_{\text{beam}}}{\text{GeV}})$ MIP and $(N_i + N_{i+1}) > (3.77 + 1.44 \cdot \ln(\frac{E_{\text{beam}}}{\text{GeV}}))$ were satisfied, the i^{th} layer was considered to be the primary interaction layer. For both criteria, the sum of visible signals in two successive layers is used to distinguish between shower development and local Landau fluctuations. The application of moving average for the values A_i helps to minimize the impact of noise. The thresholds for several beam energies were estimated using MC samples with known first interaction point [8] by means of minimization of the RMS deviation of the found shower starting layer from the true one. The expressions above contain the results of threshold energy dependence fits. The energy dependence of the determined thresholds is due to the fact that fluctuations and mean values of the energy loss of minimum ionizing particle increase with energy. Tests on simulated samples have shown that the difference between the found and true primary interaction layer does not exceed one layer for 78% of the events and two layers for more than 90% of the events in the studied energy range from 10 to 80 GeV.

2.3 Systematic uncertainties

The systematic uncertainties were analyzed in detail in the study of the electromagnetic response of the AHCAL [4]. In contrast to the case for electrons, for hadrons in the energy range from 10 GeV to 80 GeV one expects negligible contributions from SiPM gain and saturation correction accuracy to the hadronic resolution. The impact of these uncertainties was studied for hadron runs by varying the corresponding gain and saturation parameters and their contributions were found to be significantly less than 1%.

The conversion from MIP to GeV was done using weights defined on the electromagnetic scale (see table 1). The weight for the AHCAL was obtained from electron and positron runs with an accuracy of 0.9% [4], which contains a wide range of different systematic contributions studied in the electromagnetic analysis. This accuracy can thus be taken as a systematic uncertainty of the reconstructed energy in the case of hadron runs. The observed difference between means determined for runs at the same beam energy is within this systematic uncertainty.

The uncertainty of the mean energy of the beam is also taken into account using equation 5.1 from [4].

3 Intrinsic AHCAL energy resolution

Two statistically independent sets of data in the energy range from 10 to 80 GeV are necessary for the determination of software compensation factors and for the application of the compensation procedure. For this purpose, all data runs were split into two sets, one with even and one with odd event numbers. All sets with equal beam energies were merged, resulting in two subsamples for each beam energy. The effect of the run splitting and merging is shown in fig. 2 and 3.

For selected pion events with a shower start in the first five layers of the HCAL, as described in section 2.2, the reconstructed energy E_{event} of an event is calculated as follows:

$$E_{event} = \sum_{k=1}^{3} v_k \cdot M_k^{\mathrm{ECAL}} + \frac{e}{\pi}\left(w^{\mathrm{HCAL}} \cdot M^{\mathrm{HCAL}} + \sum_{k=1}^{2} w_k^{\mathrm{TCMT}} \cdot M_k^{\mathrm{TCMT}}\right), \qquad (1)$$

where $\frac{e}{\pi} = 1.19$ is a scaling coefficient to take into account a different response to electrons and hadrons in the non-compensating AHCAL (the coefficient was obtained for pions by averaging the ratio of beam energy to the total energy reconstructed at electromagnetic scale over the studied energy range); M^{ECAL}, M^{HCAL}, and M^{TCMT} are sums of visible signals in each subsection.

The reconstructed energy distributions were fitted with a Gaussian in the interval of ± 2 RMS around the mean value, resulting in good fits with a $\chi^2/\mathrm{NDF} < 2$ for all but three runs. For these three runs, a $\chi^2/\mathrm{NDF} < 2$ was obtained by reducing the fit range to an interval of ± 1.8 RMS. Hereinafter, the mean and sigma of this Gaussian fit at a given beam energy are referred as a mean reconstructed energy E_{reco} and resolution σ_{reco}, respectively.

The response of the calorimeter setup to pions as a function of beam energy is shown in fig. 2a. Due to the intrinsic non-compensation of the CALICE AHCAL, the response to pions is non-linear with energy, deviating $\pm 2\%$ from a perfectly linear behavior in the studied energy range as demonstrated in fig. 2b, where relative residuals to the true beam energy are shown. As expected, the difference between the even and odd subsamples is within statistical errors.

The relative energy resolution is shown in fig. 3. The resolution for π^- events (filled markers) is in good agreement with that observed for π^+ events (open markers). The solid curve represents a fit with the following function:

$$\frac{\sigma}{E} = \frac{a}{\sqrt{E}} \oplus b \oplus \frac{c}{E}, \qquad (2)$$

where E is in GeV, and a, b and c are stochastic, constant and noise contributions, respectively. The noise term is fixed at $c = 0.18$ GeV, corresponding to the measured

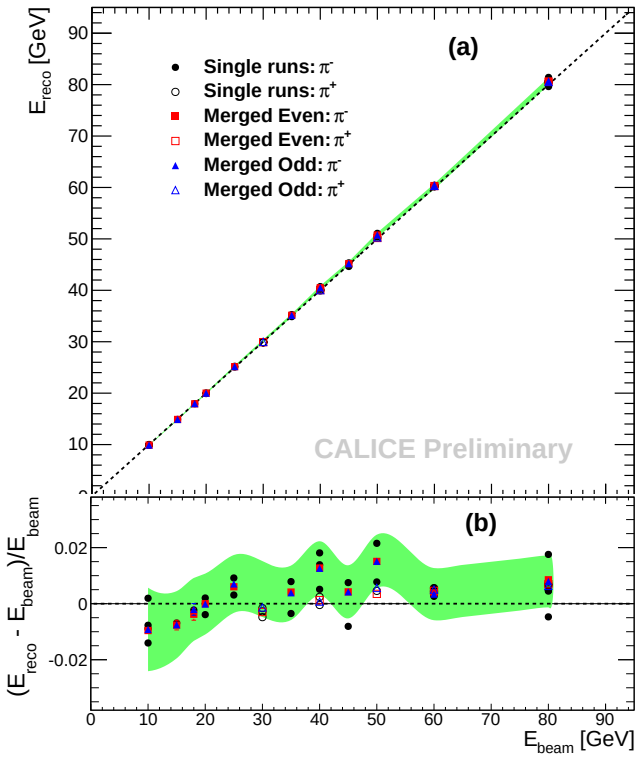

Figure 2: (a) Linearity of the CALICE AHCAL response to pions and (b) relative residuals to beam energy estimated for single runs (black circles), even (red squares) and odd (blue triangles) subsamples of merged runs. Filled and open markers show π^- and π^+ data, respectively. Dotted lines correspond to $E_{\mathrm{reco}} = E_{\mathrm{beam}}$. The green band indicates systematic uncertainties for merged even π^- subsamples.

noise contribution in the full CALICE setup taking in account contributions from ECAL (0.004 GeV), AHCAL (0.06 GeV) and TCMT (0.17 GeV). These estimates were obtained using dedicated runs without beam particles as well as using random trigger events constantly recorded during data taking [4]. Both measurements gave consistent RMS values of the noise distributions. As shown in fig. 3, the results for even and odd subsamples agree within errors.

4 Application of software compensation techniques to test beam data

4.1 Tuning of compensation techniques

Both software compensation techniques described in CAN-015 [1] and CAN-028 [2] do not require a knowledge of particle energy for the compensation to be applied. At the

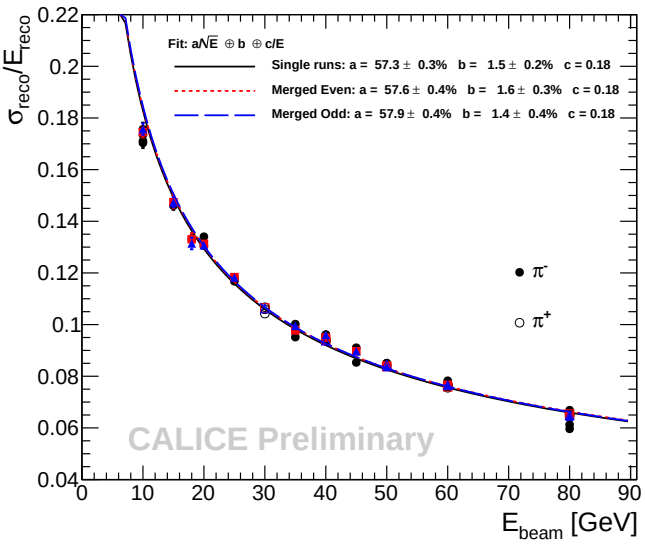

Figure 3: Intrinsic relative energy resolution of the CALICE AHCAL for data sets from single runs (black circles), even (red squares) and odd (blue triangles) subsamples of merged runs. Filled and open markers correspond to π^- and π^+, respectively. Curves represent a fit with the function from eq. 2.

same time, these techniques need test beam information to extract parameters of the energy dependence of compensation factors. The parameters for both techniques were adjusted using the training data set obtained with the selection conditions and calibration coefficients described above.

Global compensation (GC) tuning. For the global compensation method, described in detail in [2], a threshold of $e^{\mathrm{lim}} = 5$ MIP (see equation 1 in [2]) was chosen for the present study. The following compensation parameters, as used in equation 5 from [2], have been derived from the training dataset: $a_1 = 0.982 \pm 0.007$, $a_2 = (4.1 \pm 0.3) \cdot 10^{-3}$ GeV^{-1}, $a_3 = (-2.2 \pm 0.3) \cdot 10^{-5}$ GeV^{-2}. The obtained compensation factor was applied to a shower energy, i.e. to the sum of HCAL and TCMT deposited energies.

Local compensation (LC) tuning. The parameters for the local software compensation method are determined from the training dataset as described in detail in [1]. The reconstruction of the calorimeter energy is performed in a two-step procedure. In the first step the reconstructed energy is calculated without software compensation to obtain the energy-dependent compensation factors. These are then applied in the second step to obtain an improved energy measurement. In the present study, compensation is only applied to the HCAL energy, while the measured energy in the ECAL and in the TCMT are used without additional compensation procedures.

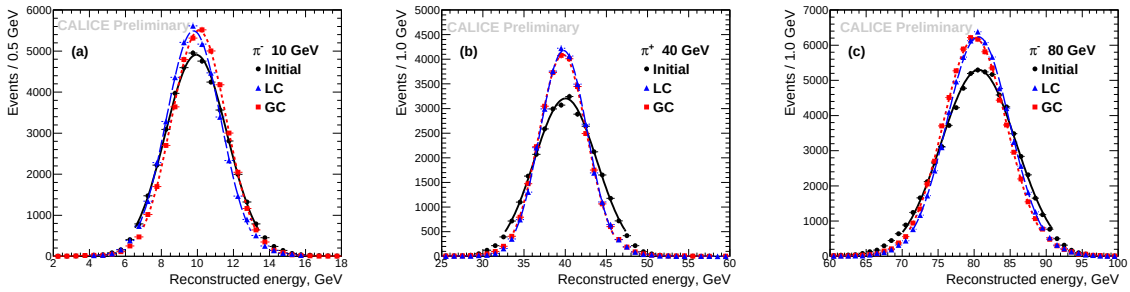

Figure 4: Reconstructed energy distributions for pions at 10 GeV (a), 40 GeV (b) and 80 GeV (c) without compensation (black circles) and after local (LC - blue triangles) and global (GC - red squares) software compensation applied. The curves show a Gaussian fit for the corresponding distributions. Statistical errors are shown.

4.2 Comparison of compensation techniques

It should be emphasized that the application of either local or global compensation does not distort the Gaussian form of initial reconstructed energy distributions. Figure 4 shows examples of energy distributions, comparing the initial (black circles) distribution with the distributions resulting from both software compensation techniques. Both techniques - global (red squares) and local (blue triangles) - lead to a narrowing of the energy distributions while retaining the Gaussian shape.

Figure 5 shows the linearity of the response of the complete CALICE setup to pions for selected events, with showers starting in the first five layers of the AHCAL (see section 2). Black circles correspond to the initial response without any compensation, while red squares and blue triangles correspond to the response after application of global or local compensation, respectively. Improved linearity after compensation is observed. In general, the response is linear within $\pm 1.5\%$ in the studied energy range.

The relative energy resolution before and after compensation is shown in fig. 6. Good agreement of the π^- and π^+ samples was observed. The energy dependence of the relative resolution is well described by eq. 2 with a fixed noise term c =0.18 GeV, which was derived from noise measurements discussed above. The fit results are given in table 3. The application of software compensation results in a decrease of the stochastic term while the constant terms before and after compensation stay constant within errors. Both compensation techniques show very similar performance, with the local software compensation providing a slightly smaller stochastic term.

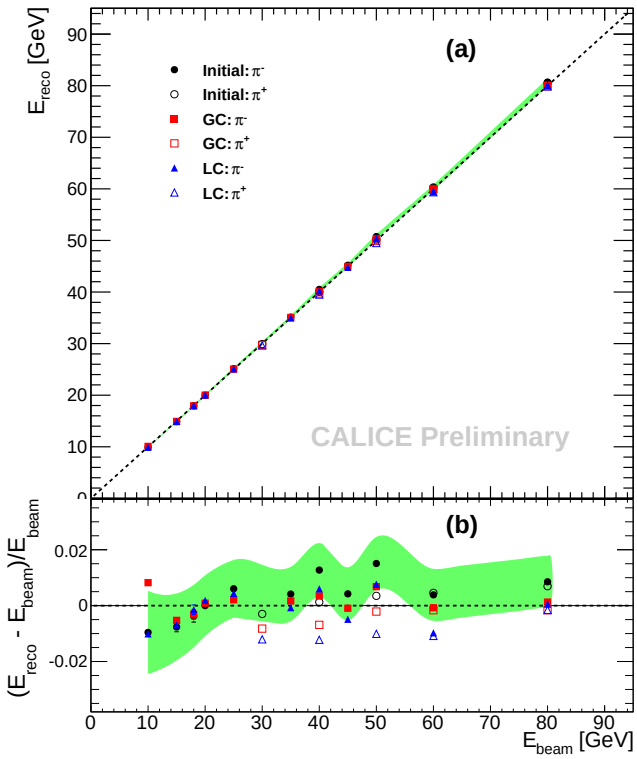

Figure 5: (a) Linearity of the CALICE AHCAL response to pions and (b) relative residuals to beam energy versus beam energy without compensation (black circles) and after local (LC - blue triangles) and global (GC - red squares) compensation. Filled and open markers indicate π^- and π^+, respectively. Dotted lines correspond to $E_{\mathrm{reco}} = E_{\mathrm{beam}}$ and green band shows systematic uncertainties for the initial π^- data sample.

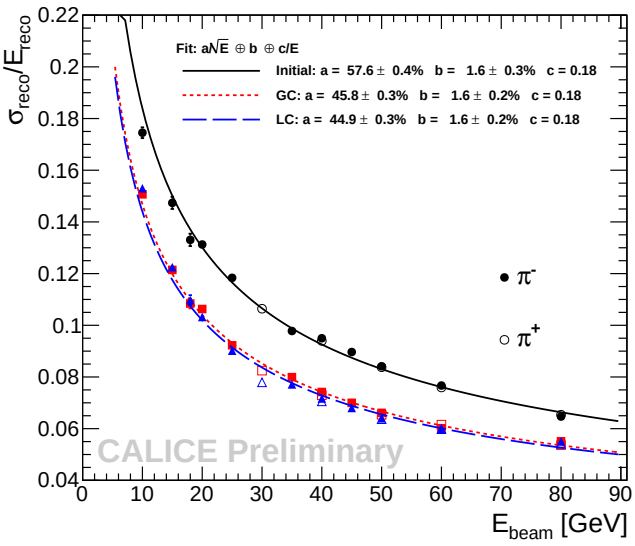

Figure 6: Relative resolution versus beam energy without compensation (black circles and solid line) and after local (LC - blue triangles and dashed line) and global (GC - red squares and dotted line) compensation. The curves show fits using eq. 2.

Table 3: Stochastic, constant and noise term contributions to hadronic energy resolution of the CALICE AHCAL obtained from a fit with eq. 2.

Resolution	a, %	b, %	c, GeV	χ^2/NDF
Initial	57.6±0.4	1.6±0.3	0.18	3.6
Local compensation	44.9±0.3	1.6±0.2	0.18	7.2
Global compensation	45.8±0.3	1.6±0.2	0.18	2.6

The relative improvement of the energy resolution can be expressed as a ratio of the resolution after software compensation σ_{SC} (local or global) and the initial resolution $\sigma_{\mathrm{initial}}$, as shown in fig. 7. An energy dependent improvement, ranging from ∼12% to ∼25% in the studied energy range, with a maximum improvement around 30 GeV, was observed. The local technique yields a slightly better improvement in the middle of the studied energy range from 25 GeV to 60 GeV. The worsening of the relative improvement at higher energies can be caused by a distortion of hit energy spectra due to increasing leakage into the TCMT. In the local compensation procedure, the shower hits from TCMT are not weighted. In the global compensation procedure, the partial absence of hits in the hit spectrum results in wrong (or shifted) estimate of the global compensation factor.

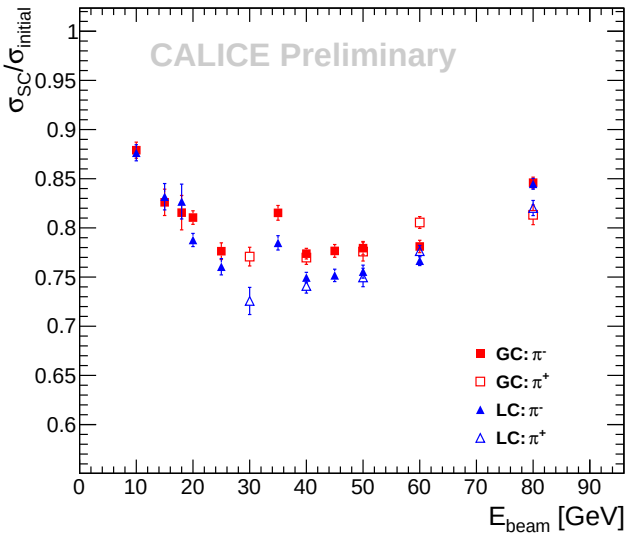

Figure 7: Energy dependence of the relative improvement in resolution for local (LC - blue triangles) and global (GC - red squares) software compensation techniques.

5 Conclusion

The intrinsic relative energy resolution of the CALICE AHCAL for hadrons has been measured to be about $57.5\%/\sqrt{E}$, with a constant term of about 1.6%. Updated procedures were used for run and event selection. All runs of the same beam energy and particle type were merged and all samples were split into two subsamples to get statistically independent data sets for the training of software compensation methods. The methods use different approaches to achieve compensation: individual hit weighting in the local approach vs. one factor for energy sum over shower hits in the global one, the local technique implying twice as many parameters as the global one. Both local and global software compensation techniques were adjusted using one of subsamples and applied to another to support the sample independence of the methods. The achieved relative improvement in hadronic energy resolution varies from 12% to 25% in the studied energy range from 10 GeV to 80 GeV, resulting in a reduction of the stochastic term down to $45\%/\sqrt{E}$. Both local and global compensation techniques give almost equal improvement in the stochastic term, with the local approach providing a 3% better relative improvement in the range from 25 GeV to 60 GeV.

References

[1] The CALICE Collaboration, *Initial Study of Hadronic Energy Resolution in the Analog HCAL and the Complete CALICE Setup*, CALICE Analysis Note CAN-015 (2009).

[2] The CALICE Collaboration, *A new approach to software compensation for the CALICE AHCAL*, CALICE Analysis Note CAN-028 (2010).

[3] The CALICE collaboration, *Design and electronics commissioning of the physics prototype of a Si-W electromagnetic calorimeter for the International Linear Collider*, 2008 JINST 3 P08001.

[4] The CALICE collaboration, *Electromagnetic response of a highly granular hadronic calorimeter*, 2011 JINST 6 P04003.

[5] The CALICE Collaboration, *Preliminary results from hadron shower data with the CALICE tile AHCAL prototype*, CALICE Analysis Note CAN-011 (2008).

[6] K. Seidel, *Si-W-Ecal treatment for HCAL only analysis and development of event selection processors*, Talk on AHCAL Main Meeting, DESY, 20.01.2011.

[7] M. Chadeeva, *HadronSelection Processor for software compensation study*, Talk on Analysis Meeting, 18.04.2011.

[8] The CALICE Collaboration, *Pion Showers in the CALICE AHCAL Prototype*, CALICE Analysis Note CAN-026 (2011).

6 Appendix A. Examples of muon selection

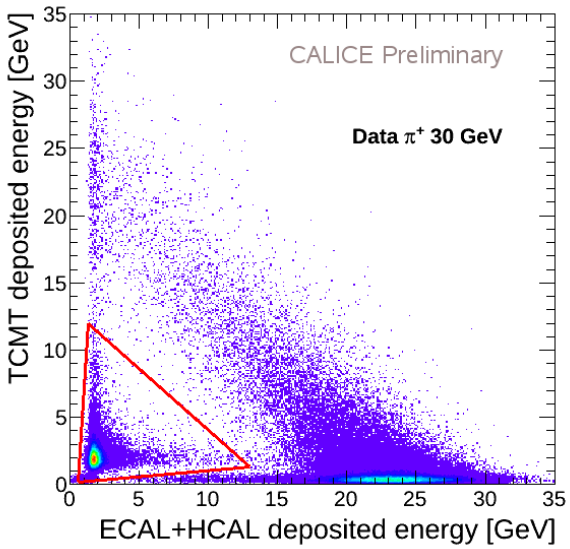

Figure 8: Constraints (red triangle) applied to select muons from 30 GeV π^+ sample using the distribution of deposited energy in ECAL+HCAL versus energy deposited in TCMT. See text for details.

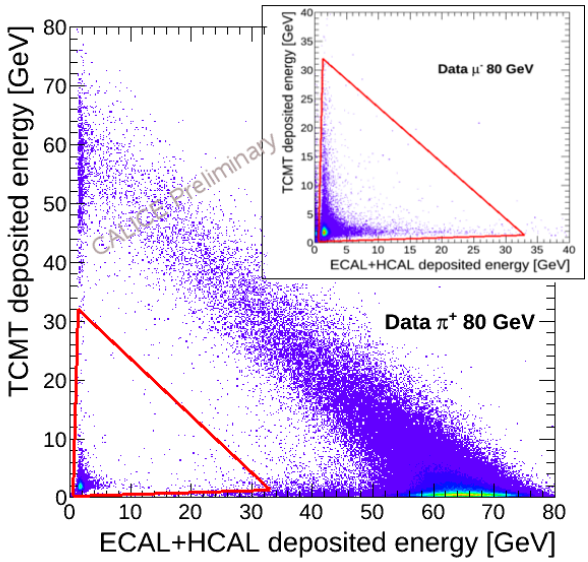

Figure 9: Constraints (red triangle) applied to select muons from 80 GeV π^- sample using the distribution of deposited energy in ECAL+HCAL versus energy deposited in TCMT. In the insert the same constraints applied to 80 GeV muon run are shown. See text for details.