CALICE Analysis Note CAN-xxx
June 2, 2010

PandoraPFA tests using overlaid charged pion test beam data

The CALICE Collaboration [1]

Abstract

CALICE 2007 test beam data were used to test the PandoraPFA program. The program capability to recover a neutral hadron energy in the vicinity of a charged hadron was studied. The impact of overlapping of two hadron showers on energy resolution was investigated. The dependence of the confusion error on the distance between a 10 GeV neutral hadron and a charged pion was derived for pion energies of 10 and 30 GeV which are representative of a 100 GeV jet. The comparison of these test beam data results with Monte Carlo simulation was done for LHEP and QGSP_BERT physics lists. The probability of correct recovering of neutral hadron energy has been calculated. The dependence of the confusion error on the neutral hadron energy was also studied. The confusion error averaged over 100 GeV jet fragment energies has been estimated for the case of one neutral and one charged hadron.

[1]Corresponding author: Oleg Markin; `markin@itep.ru`

Contents

1 Introduction

The Particle Flow Analysis (PFA) algorithm was implemented in the PandoraPFA program [1] as a part of the software for a future Linear Collider [2] and was tested using Monte Carlo (MC) simulated jets. The capability of the PandoraPFA program to recover neutral hadron energy in the vicinity of a charged hadron is of crucial importance because the recovery mistake would degrade the energy resolution. This occurs in case the reconstructing program mixes up hits from showers created by charged and neutral hadrons as a result of shower overlapping. One more factor that degrades the energy resolution is an overlapping of a neutral hadron shower and a photon shower. However to resolve this confusion, contrary to the case of two hadron showers, one can use energy profiles of electromagnetic showers. In case of two hadron showers overlap the task for PFA becomes more complicated because the energy profiles are useless and only topological and energy criteria can help to disentangle showers.

The impact of the overlapping of showers on energy resolution for Monte Carlo simulated jets is shown in [1]. However it is known that different available physics lists give noticeably different predictions for hadron shower shapes, that might be important for resolving the overlapped showers. Moreover, the real detector could have not as good performance as its idealized MC model. The main goal of this study is to check Pandora PFA program performance using real test beam data and to compare the result with MC predictions.

To investigate this issue we use test beam data collected at CERN in 2007 by the CALICE detector prototype. This prototype allows to reconstruct a hadron shower shape with unprecedented accuracy. It consists of $\sim 1\ \lambda_I$ electromagnetic calorimeter (ECAL), $\sim 4.5\ \lambda_I$ hadronic calorimeter (HCAL) and $\sim 5\ \lambda_I$ tail catcher and muon tracker (TCMT). The detailed description of the complete CALICE setup and first results on hadronic shower reconstruction and analysis can be found in [4, 5]. The CALICE calorimeter prototypes are very similar to one of the ILD detector technologies proposed for the future ILC [7] and previously known as LDC [8]. The size of tiles of the prototype HCAL in the zone close to the beam line coincides with that of LDC. Even though the prototype has a little bit different thickness of layers and smaller number of layers in HCAL, than LDC, for energies up to 30 GeV it is capable to reconstruct the structure of hadron showers with almost the same accuracy as LDC.

The first procedure of shower overlapping at CALICE structure was described in [3]. The results concerning cluster overlapping were presented in [4]. In our study we present more detailed analysis including (i) energies of hadrons representative of a 100 GeV jet, (ii) wider range of distances between hadron shower axes, and (iii) direct comparison between hadron shower energies recovered by PandoraPFA and measured in the calorimeter prototype. To present results, we use (i) the difference between the neutral hadron shower energy recovered by PandoraPFA and measured by the prototype and (ii) the probability of correct recovering of the energy of a neutral hadron. The latter value is called "efficiency" in [4].

The test beam data provide us with pion showers reconstructed in the calorimeter pro-

Table 1: List of data runs used for analysis

Run #	330777	330850	330849	330848	330797	330796	331298	331339	331337
Particle	π^-	π^-	π^-	π^-	π^-	π^-	π^+	π^+	π^+
GeV	10	10	12	15	20	25	30	40	50

totype. To study the confusion error of overlapped hadron shower recovering, we pass to PandoraPFA two showers: one charged pion shower as it is and one emulating a neutral hadron (e.g. K_L^0 or neutron) shower. Before this, the 4-step preparation procedure was applied:

- single pion events were selected from test beam data (see section 2);
- primary track for every event was found and all hit coordinates were shifted to zero position in XY-plane in CALICE coordinate system (beam direction is along Z axis, y is vertical and x horizontal);
- one of the two events was shifted w.r.t. other by specified distance, keeping only hits behind the shower start (see subsection 3.2);
- both events were mapped onto the LDC geometry, summing up hit energies in each LDC calorimeter cell (see subsection 3.3).

To confront test beam data with MC, GEANT4 simulation for two physics lists, LHEP and QGSP_BERT, was performed using beam profiles corresponding to the data runs. For simulation and digitization the standard CALICE software packages were used: Mokka v07-00 (detector model TBCern07_p0709, range cut 50 micron and time cut 150 ns), calice_sim (as of December 2009). The MIP to GeV visible conversion factors for MC digitization set in official software were 0.000147 GeV/MIP for ECAL and 0.000816 GeV/MIP for HCAL.

2 Event preparation

The pion test beam includes also as a background protons, electrons/positrons and muons. Since for our study we use positive and negative pion induced showers, we have to select only pion events. The pion events were selected from several CERN 2007 test beam runs with energies from 10 to 50 GeV and without HCAL rotation (see Table 1). For calibration and reconstruction the official CALICE software packages were used (as of December 2009). The conversion coefficients from visible to deposited energy for different subdetectors were taken from preliminary CALICE results [4,6] and are listed in Table 2.

First of all, we use PureBeamTrigger flag to exclude pedestal and calibration events. Then the information from the Čerenkov trigger is taken into account to exclude electrons in case of negative pion beam and protons in case of positive pion beam. Besides these triggers the

Table 2: Conversion coefficients from visible to deposited energy

Subdetector	$\frac{GeV}{MIP} \times$ Sampling Factor
ECAL 1	0.00376
ECAL 2	0.00752
ECAL 3	0.01128
HCAL	0.02653
TCMT 1	0.02653
TCMT 2	0.13267

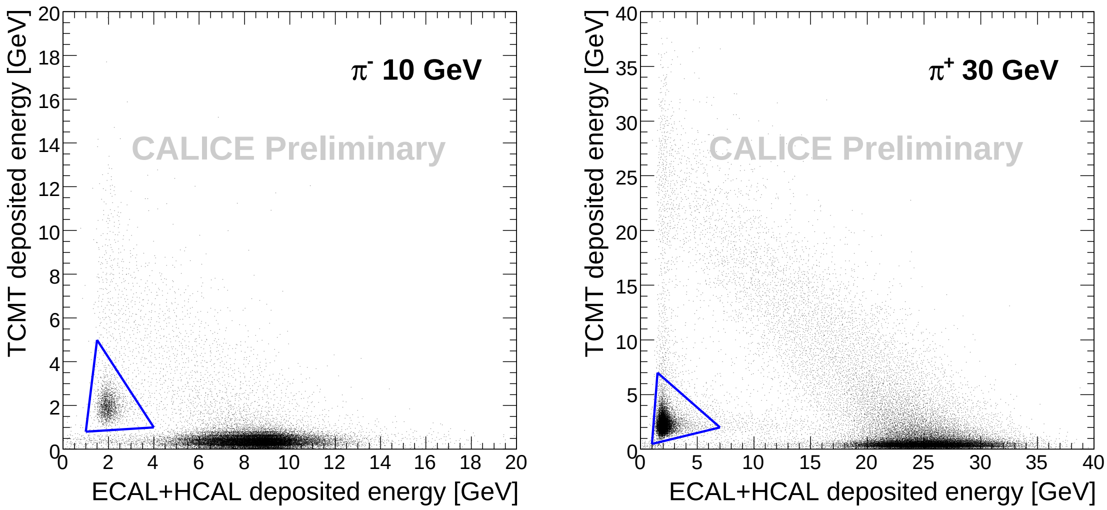

Figure 1: *Energy deposited in ECAL+HCAL versus energy deposited in TCMT for 10-GeV (left) and 30-GeV (right) hadron test beams. Triangles correspond to muon selection conditions.*

additional selection procedure is used to purify the sample by excluding muons, electrons and trash events including multiparticle ones. During selection procedure the 0.5-MIP threshold is applied to all cell signals to reject noise.

To identify muons the TCMT information is invoked. In Fig. 1 the 2D distributions are shown where the energy deposited in ECAL+HCAL is plotted versus the energy deposited in TCMT, for muons both depositions being approximately equal. The events that fall inside the highlighted triangle in the lower left corner of the plot we consider as muons. As follows from the left and right plots in Fig. 1 for 10 and 30 GeV respectively, the signal from muons almost does not interfere with that from hadrons and this approach allows to select muons from hadron beams down to 10 GeV in the complete CALICE setup. For the runs used, the muon containment was ~7% in 10-GeV pion beam and ~30% in 30-GeV pion beam.

After selecting muons, we reject events with too low overall energy deposition, less than $E_{beam} - 3.6 \cdot \sqrt{E_{beam}}$, where E_{beam} is the beam energy (in fact this condition works only for

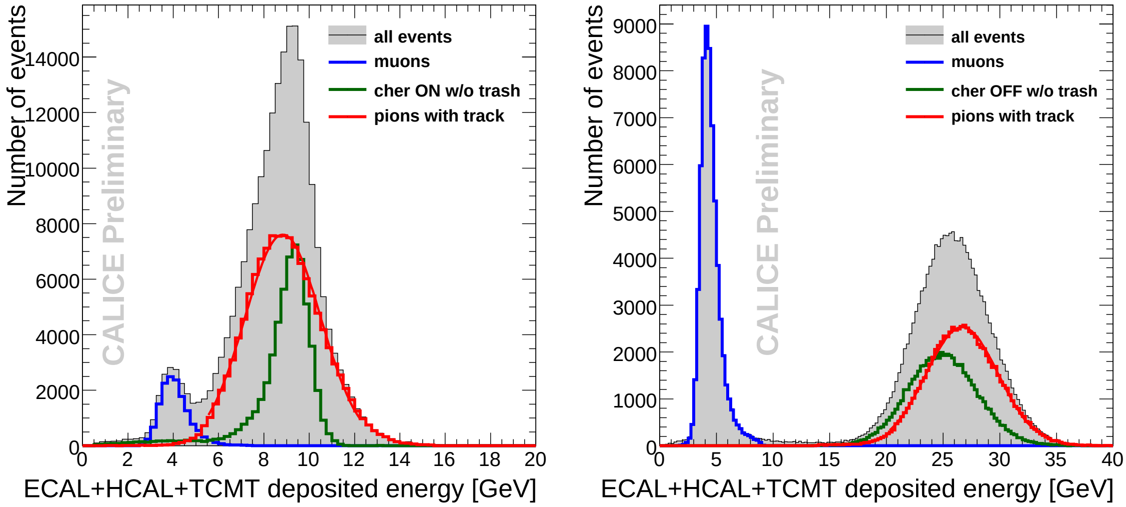

Figure 2: *Deposited energy distributions before (shaded) and after selections for 10-GeV negative pion (left) and 30-GeV positive pion (right) beams. Red solid lines show Gaussian fits to resulting pion sample distributions.*

beam energies higher than 12 GeV). The events with no hits in ECAL are also excluded from the analysis. The fraction of trash events was less than 3% (for 10 GeV) and less than 2% (for 30 GeV).

The following events are considered to be multiparticle: (i) with deposited energy higher than $1.5 \cdot E_{beam}$ (note that the mean value of energy distribution is estimated to be about $0.86 \cdot E_{beam}$ for CALICE prototype due to π/e ratio); (ii) with shower start in HCAL while having more than 50 hits in ECAL; (iii) with several parallel primary track candidates found by the PrimaryTrackFinder processor. The fraction of multiparticle events does not exceed 0.2%.

The deposited energy distributions before and after selections are shown on Fig. 2. The green histogram corresponds to an electron admixture to the negative pion beam identified by Čerenkov trigger on the left plot and to a proton admixture to the positive pion beam on the right one. The resulting selected pion sample is well fitted by Gaussian distribution. Trash and multiparticle contributions are not plotted here because of their low impact comparing to shown components.

The same selection procedure was applied to MC samples after digitization. To find a shower starting layer and a primary track we used algorithms included in the PrimaryTrackFinder processor from CALICE software. The primary track information is necessary for this analysis so there would be a possibility to shift any single shower independently as a whole and then displace them on the a priori specified distance between their axes before overlapping. There were ∼12% and ∼7% of events without any primary track data for 10 GeV and 30 GeV runs respectively.

3 Mixing of showers

3.1 Selection of showers

The selected pion events were overlaid and passed to PandoraPFA after an energy selection procedure. This procedure is necessary because the HCAL of the detector prototype is not deep enough to contain every hadron shower and some showers partially leak to the TCMT, where it is impossible to determine shower hit positions. Since hit positions are necessary for PandoraPFA to work, we selected for the analysis only events with overwhelming containment in the ECAL plus HCAL (more than 95% of deposited energy), where the fine granularity allows us to retrieve the shower hit positions. At energies from 10 GeV to 30 GeV showers are not so long; their energy density drops e times after near 15 HCAL layers from the shower start. For this reason such a selection does not restrict physics too much. It means that we do not use showers which start in far HCAL layers and it is worth to note that such showers will be better separated due to the magnetic field in a future detector and hence the confusion for them will be smaller.

3.2 Shifting of showers

To understand to what extent PandoraPFA confuses hits in hadron showers we take first only two particles: two selected pion events. Because of beam smearing, it is necessary to identify a beam particle entrance point (primary track coordinates) for each event to get its shower axis position. This information was extracted from drift chambers (DC) data which is available for ∼30% of events in the analyzed runs. For other events the algorithm of primary track finding (from the PrimaryTrackFinder processor) was used, which accuracy is worse than that of DC but does not exceed 0.5 cm (half of the ECAL cell transversal size). The results got by both approaches are in very good agreement with each other. After extracting primary track coordinates we shift all event hits placing the shower axis at the (0,0) position in the XY-plane in CALICE coordinate system. Then we took two events from different runs and for one of them moved all hits by a certain transverse distance from their original position. This distance varied from 5 cm to 30 cm. For this displaced event we kept only hits behind the shower start, i.e. proper shower hits without incident track, and used this shower as an imitation of the neutral hadron shower. This imitated shower represents the main subject of our study. In what follows we will call the energy of this imitated hadron shower the *neutral hadron shower energy*.

The top row of Fig. 3 shows the energy distributions for the 30 GeV charged (left), 10 GeV charged (middle), and 10 GeV "neutral"(right) hadron events after our preparation procedure. These energies were chosen for the following analysis as being typical for 100-GeV jet (see section 6).

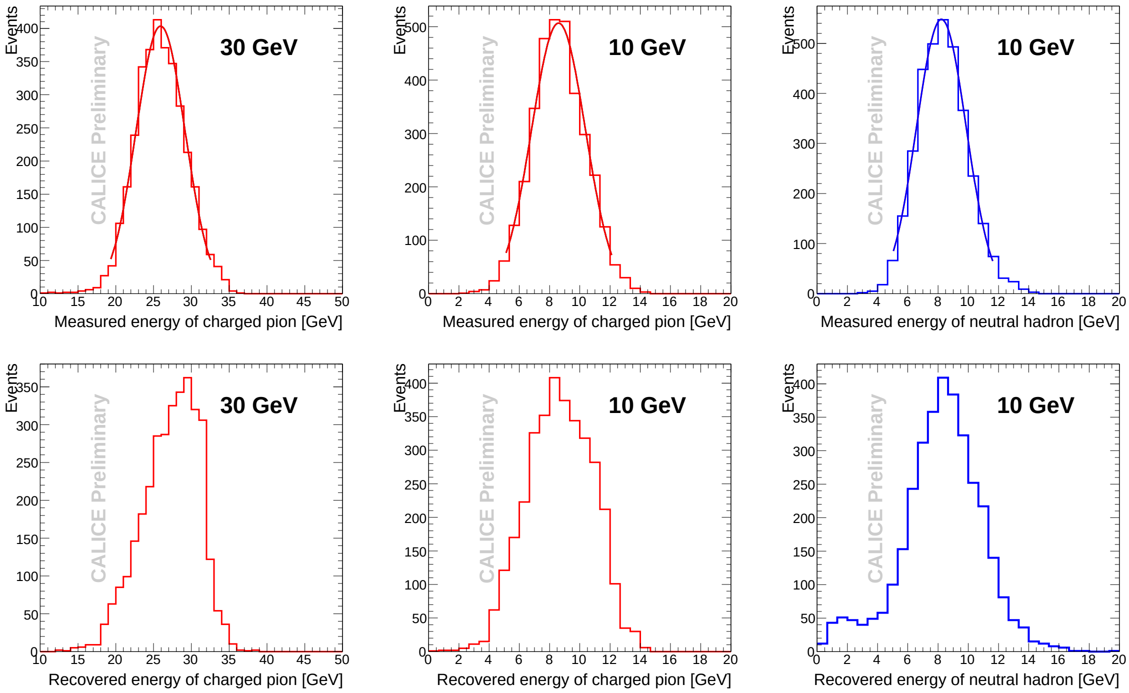

Figure 3: *Top row: energy distributions for 30 GeV charged (left), 10 GeV charged (middle) and 10 GeV "neutral" (right) hadron events prepared from data runs for mixing of two showers. Solid lines correspond to Gaussian fit. Bottom row: the same energy distributions after their disantangling by PandoraPFA for the case of 15 cm distance between corresponding charged and "neutral" hadrons.*

3.3 Mapping of shower hits

This event with two showers was mapped to the LDC detector structure and written to an input LCIO file for Pandora. The LDC detector is an octahedral barrel with two endcaps (see [8]). The CALICE prototype hits were put in the top octant of the barrel, layer by layer. Thus the CALICE beam became directed vertically up along the Y axis in the LDC geometry. Together with hits, the energy of the two showers measured by the prototype was passed to PandoraPFA for comparison with the recovered shower energy. After the official CALICE reconstruction procedure the 0.5-MIP cut was applied to each hit before mixing. To make the energy comparison fair, one needs the equality between the sum of the first and the second shower energy measured by the prototype and the full energy written in the two hadron event. For this reason signals from two hits in the same tile were simply added in the process of shower merging. The possibility that the sum of two signals below 0.5 MIP threshold exceeds the threshold after shower merging is ignored. We have chosen such an approach to reveal proper PandoraPFA confusion, avoiding double counting the noise at the price of loss for some number of subliminal hits. At large distance between showers the probability to find two hits from different showers in one cell naturally vanishes.

Mapping the CALICE prototype events to LDC leads to a slight distortion of shower shapes. There are only 29 LDC ECAL layers instead of 30 prototype ECAL layers that entails a loss of a few hits. The different absorber layer thickness layout in the prototype ECAL and LDC ECAL leads to distortion of the ECAL shower longitudinal profile; however for hadron cluster reconstruction PandoraPFA uses this profile very little and indirectly. As for HCAL, the local shift of hit position takes place on the border between the 3×3 cm and 6×6 cm cells. However due to our beam smearing correction a hit from a 6×6 prototype HCAL tile was put in one of the four corresponding 3×3 tiles of the LDC HCAL practically randomly. The CALICE HCAL has about 32 mm thick layers while the LDC HCAL has 26 mm thick layers, the absorber thickness being equal in both calorimeters. This means that after mapping, the shower is shorter than in the calorimeter prototype and it is still a little bit wider than in reality. All these distortions do not change the shower shape considerably. They only slightly affect the shower topology in a way which complicates the task for PandoraPFA to resolve two showers, making our conclusions about PandoraPFA performance rather conservative.

4 PandoraPFA processor adjustment

The PandoraPFA processor has been adjusted and simplified for our purposes. Modifications we had to make in PandoraPFA are not concerned with its algorithm and only affect the way charged hadron track characteristics are introduced in the program. To calculate the energy of the track and its entrance point position and direction, PandoraPFA reads TPC hits and makes a helix fit. We just give to the program a straight track which intersects the calorimeter barrel inner surface at zero XZ position with normal incidence and has definite energy. Of course any subsequent calculation of a distance between the track prolongation and shower hits or clusters was replaced by a calculation of the distance from the prolongation of our straight track. In the presence of a magnetic field, even though the hadron shower gets little smearing, its end appears to be further from the jet axis than the shower beginning, proportionally to the squared distance from the interaction point. Thus the magnetic field makes it easier for PandoraPFA to separate showers. Therefore our analysis gives a conservative estimate for the PandoraPFA performance.

We do not use some methods of PandoraPFA like kink track cluster association, primary photon recovering and multi-track cluster association splitting because we have no such situations in our study. Besides after the program ends we consider that a hit belongs to the neutral hadron shower if it has not been attached by the program to the charged hadron shower. We also retain the recovered energy for all output clusters and can potentially use more elaborate algorithm for estimating the neutral hadron energy.

5 Recovering of showers

Due to shower overlapping in the calorimeter, the energy recovered by PandoraPFA for each of two showers is not accurate. The resulting energy distributions after their recovering by PandoraPFA are shown in the bottom row of Fig. 3. The overlapping is considerable in cases where the charged showering particle appears to be close to the neutral one. Since we know the neutral hadron energy from calorimetric measurements, we can compare it with the energy recovered by PandoraPFA and get the confusion error.

As soon as PandoraPFA tries to pull the charged hadron shower energy up to the TPC track energy, the neutral hadron shower hits are often attributed to the charged one. For this reason it is natural to expect that the difference between the neutral hadron energy recovered by PandoraPFA and its energy measured in the calorimeter prototype to be negative.

The maximum confusion takes place between the high energy charged hadron and the small energy neutral one (see bottom left plot in Fig. 4). In such a case that high energy is given to the program as a reference point. So PandoraPFA is to consider a significant number of neutral shower hits as charged shower hits while keeping the difference between the TPC track energy and the energy of charged hadron within the measurement error. Therefore for a large number of events PandoraPFA recovers the charged hadron energy higher than it actually was in the calorimeter, adding to it a significant part of the neutral hadron energy. That mostly happens for events, in which due to intrinsic shower fluctuations the difference between the measured charged hadron energy and the beam energy is comparable with the neutral hadron energy and PandoraPFA tries to cover up this difference at the expense of the neutral hadron energy. On the other hand, PandoraPFA can not add to the charged shower the energy much larger than the typical measurement error. Both effects together give a peak around -6 GeV for the 30 GeV charged and the 10 GeV neutral hadrons (see Fig. 4).

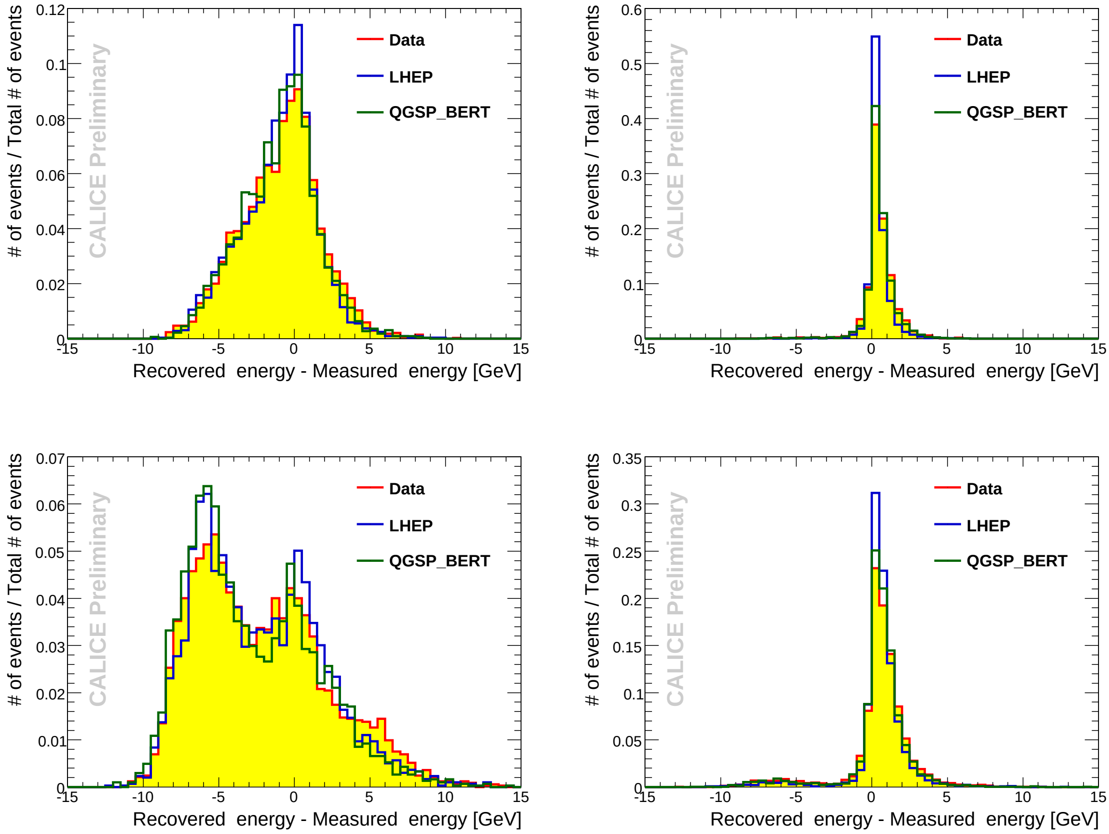

Figure 4: *Difference between the recovered energy and the measured energy for the 10 GeV neutral hadron at 5 cm (left) and at 30 cm (right) from the 10 GeV (top) and 30 GeV (bottom) charged hadrons. Data (red, yellow shaded) is compared to MC predictions for LHEP (blue) and QGSP_BERT (green) physics lists.*

The shoulder on the right slope of the zero difference peak appears due to the fact that the sufficiently large size of the 30 GeV charged hadron shower gives PandoraPFA an opportunity to associate many hits of this shower with the neutral hadron and to recover its energy even higher than it actually was. At large distances this confusion vanishes and the mean value of the difference becomes positive since we include into the recovered neutral hadron shower all hits which PandoraPFA has not attached to the charged hadron shower, among them there are isolated hits and small clusters which actually belong to the charged hadron shower but were not recognized. For 10 GeV charged hadron, the neutral hadron energy reconstruction is considerably better (see top plots in Fig. 4).

The confusion between showers depends not only on the radial distance between showers, but also on the longitudinal difference between shower starting positions. Fig. 5 illustrates the behavior of the difference between the recovered energy and the measured energy for showers starting in different subdetectors. The worst case is when the charged shower starts earlier then the neutral one. In this case some hits of the neutral shower fall into the charged shower cone. In the case where the charged shower starts after the neutral one, the confusion is almost negligible.

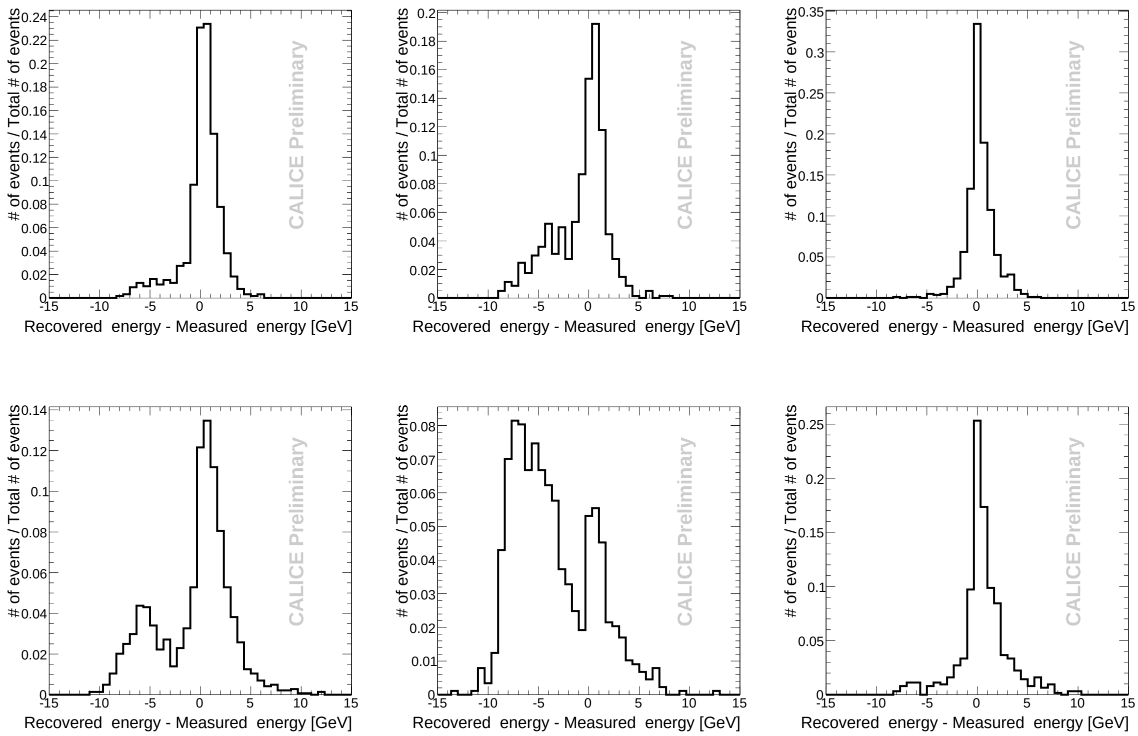

Figure 5: *Difference between the recovered energy and the measured energy of the 10-GeV neutral hadron at 15 cm from 10 GeV (top) and 30 GeV (bottom) charged hadrons for different shower start positions: in ECAL for both hadrons (left), charged in ECAL and neutral in HCAL (middle), charged in HCAL and neutral in ECAL (right).*

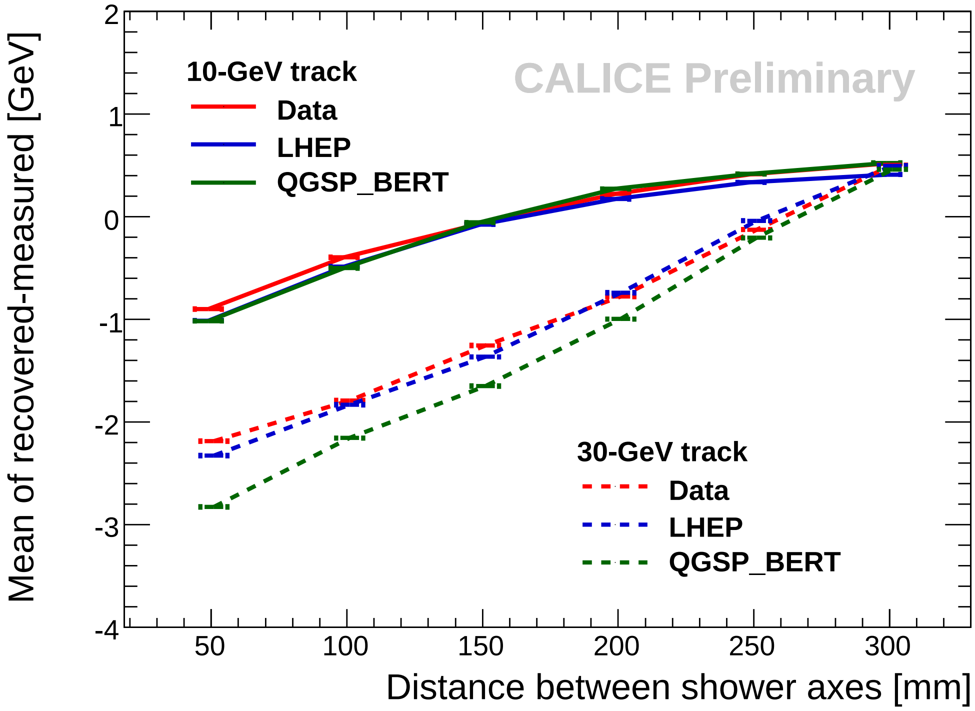

Figure 6: *Mean difference between the recovered energy and the measured energy for the 10 GeV neutral hadron vs. the distance from the 10 GeV (continuous line) charged hadron and the 30 GeV (dashed line) charged hadron.*

Using the histograms shown in Fig. 4 one can extract the mean value of the difference between recovered energy and measured energy of neutral hadron. The second characteristic used to estimate the confusion error is the RMS value. However, to avoid the over-emphasizing of the distribution tails, the RMS_{90} value is usually used, which means the root mean square deviation of the recovered energy from the energy measured in the calorimeter prototype and is calculated for 90% of events (see e.g. [1]).

The mean value of the neutral hadron energy measured in the calorimeter exceeds the mean energy recovered by PandoraPFA at small distances between particles where shower overlapping is considerable (see Fig. 6). At large distance where confusion vanishes, the mean measured energy of the neutral hadron becomes even smaller than its mean recovered energy. That happens because we include in the neutral hadron energy the energy of isolated hits and small clusters which in fact belong to the charged hadron shower but could not be associated with it because of remoteness.

The half of the transversal shower size is usually called shower radius. The radius of showers does not affect the situation appreciably whilst the showers are so close to each other that PandoraPFA tends to attach the small neutral particle shower to the big charged one. As the distance between hadrons grows, the shower radius naturally comes to the first place in the shower confusion (see Fig. 7). For this reason LHEP based simulation which gives smaller values for shower radii, predicts smaller confusion for the

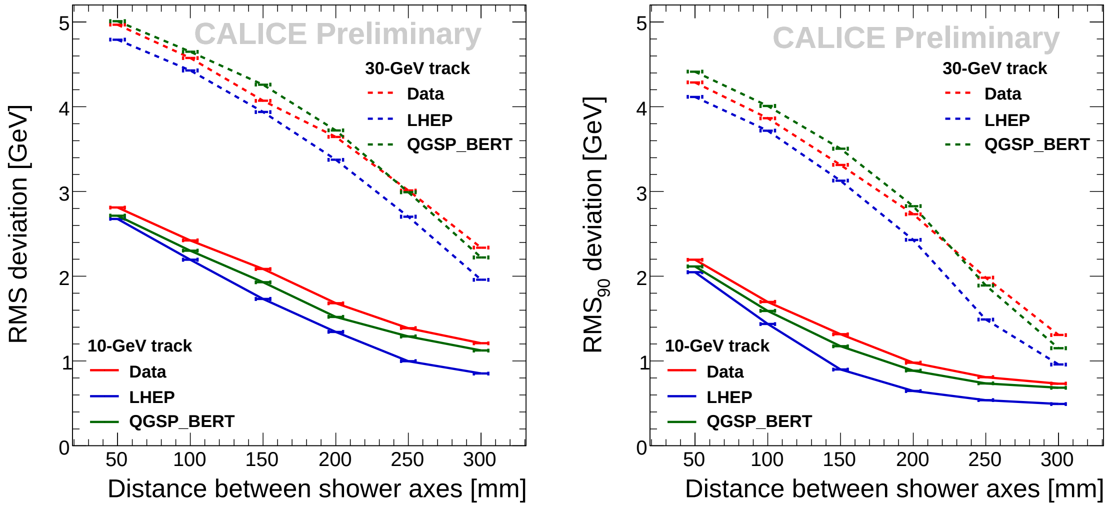

Figure 7: *RMS (left) and RMS_{90} (right) deviations of the recovered energy of the neutral 10 GeV hadron from its measured energy vs. the distance from the charged 10 GeV (continuous line) and the 30 GeV (dashed line) hadron for beam data (red) and for Monte Carlo simulated data, for both LHEP (blue) and QGSP_BERT (green) physics lists.*

distances larger than 10 cm and 20 cm for 10 GeV and 30 GeV charged hadron respectively.

Fig. 8 shows the probability of recovering of the neutral hadron energy within 2 and 3 standard deviations from its real energy at different distances from 10 GeV and 30 GeV charged hadrons. For the 10 GeV neutral hadron we take the standard deviation equal to $0.6\sqrt{10 \times 0.86 \times 0.97}$ GeV, here 0.86 is the π/e ratio and the coefficient 0.97 takes account of track fragment loss for the imitated neutral shower. The latter coefficient is just the approximate ratio of the mean value of the right histogram in Fig. 3 to the mean value of the middle one.

If the charged hadron is situated in the vicinity of a neutral one with similar or exceeding energy, the confusion is not so large. The RMS_{90} deviation of the recovered neutral hadron energy from its measured energy almost does not depend on the neutral hadron energy (see left plot in Fig. 9). The relative confusion is large for small neutral hadron energy. This results in a smaller probability of neutral hadron energy recovering for small neutral hadron energy (see right plot in Fig. 9).

6 Confusion error for 100 GeV jets

So far the confusion for the case of 10 and 30 GeV charged pions has been discussed. To estimate the confusion error for the neutral hadrons in a 100 GeV jet, we have to take into account the number, energy and position for possible charged jet fragments. Note that in the following we discuss the confusion error for only *one charged* and *one neutral*

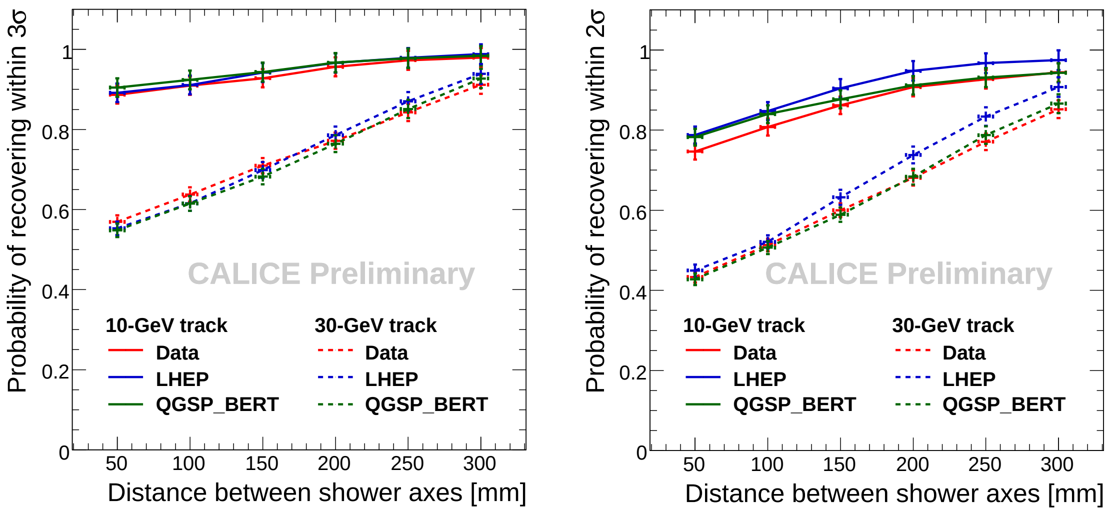

Figure 8: *Probability of the neutral 10 GeV hadron energy recovering within 3 (left) and 2 (right) standard deviations from its real energy vs. the distance from the charged 10 GeV (continuous line) and the 30 GeV (dashed line) hadron for beam data (red) and for Monte Carlo simulated data, for both LHEP (blue) and QGSP_BERT (green) physics lists.*

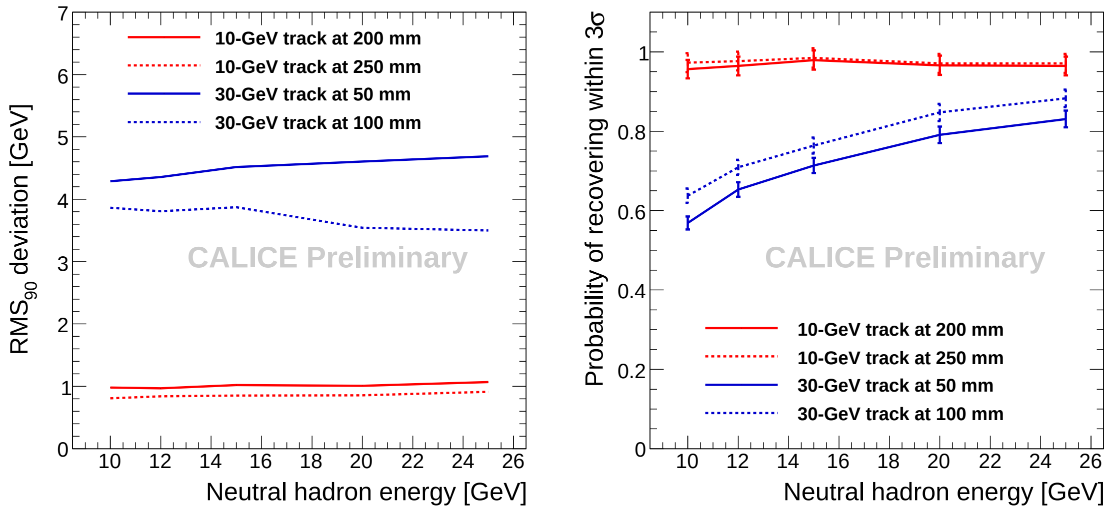

Figure 9: *RMS_{90} deviation of the recovered energy of the 10 GeV neutral hadron from its measured energy (left) and the probability of the neutral hadron energy recovering within 3 standard deviations (right) vs. the neutral hadron energy in the vicinity of the 10 GeV charged hadron (red lines) and the 30 GeV charged hadron (blue lines) for typical distances from the neutral hadron in a 4T magnetic field (see section 6).*

jet fragments.

Actually there is some probability to find the different fragments in the jet, with energies from the pion mass up to the jet energy. This probability eventually results in the confusion error for reconstructed energy of the whole jet because it makes insignificant the contribution to the confusion error from those charged fragments which are unlikely to appear. The energies of 100 GeV jet fragments are distributed approximately according to the fragmentation function [9]:

$$D(z) = (\alpha + 1)\frac{(1-z)^{\alpha}}{z}. \quad (1)$$

It is reasonable to take $\alpha = 3$. This gives an average multiplicity $\langle n \rangle = 19$; although 90% of the jet energy is carried by 10 particles [10]. Since high energy fragments have small probability to exist, the typical energy of a jet fragment is about 10 GeV. This is the reason to take 10 GeV for the neutral hadron energy in our study.

In the full-size experiment, the magnetic field deflects charged hadrons from a jet axis to distances dependent on a charged hadron momentum. For example at 190 cm from interaction point in a 4 T magnetic field this distance is equal to about 7 cm for the 30 GeV charged hadron and about 22 cm for the 10 GeV one; 2 GeV charged hadrons do not reach the calorimeter barrel at all. In addition, there is some deflection from the jet axis caused by exponentially suppressed fluctuations of the hadron transverse momentum [11]. The mean value of this deviation for the 10 GeV hadron is 3 times smaller than the deflection by the magnetic field. For a neutral hadron the deviation from the jet axis is defined only by its transverse momentum so that the overwhelming majority of neutral hadrons move in a pretty narrow cone near the jet axis.

Thus, for the simplest approximation let us consider that the distance between a neutral hadron and a charged one is approximately equal to the deflection of the latter in a 4T field. Then the distance between the neutral and charged hadrons and hence the difference between measured and recovered energies for the neutral one depends on the charged hadron energy and on the shower starting point. Although about 60% of showers start in the ECAL, for our estimation of the confusion error we take the distance between the hadrons at the very beginning of the HCAL (190 cm from the interaction point). Such a choice takes into account the fact that an angle between the charged hadron shower direction and the jet axis results in larger remoteness of distant shower hits from the axis that leads to a smaller confusion.

The mean difference between the neutral hadron shower energy recovered by PandoraPFA and its energy measured in the calorimeter is shown in Fig. 10 for different charged hadron energies. Since we know this difference only for charged hadron energies available at the test beam runs we have to fit it. Integrating this dependence with the probability to find definite charged hadron energy in the jet we get an estimation for the mean difference averaged over the jet fragment energies. We integrated the fitting function multiplied by the normalized fragmentation function $D(z)/\langle n \rangle$, the integration limits being determined from the range of possible fragment energies (from 2 GeV to 100 GeV). As a result the

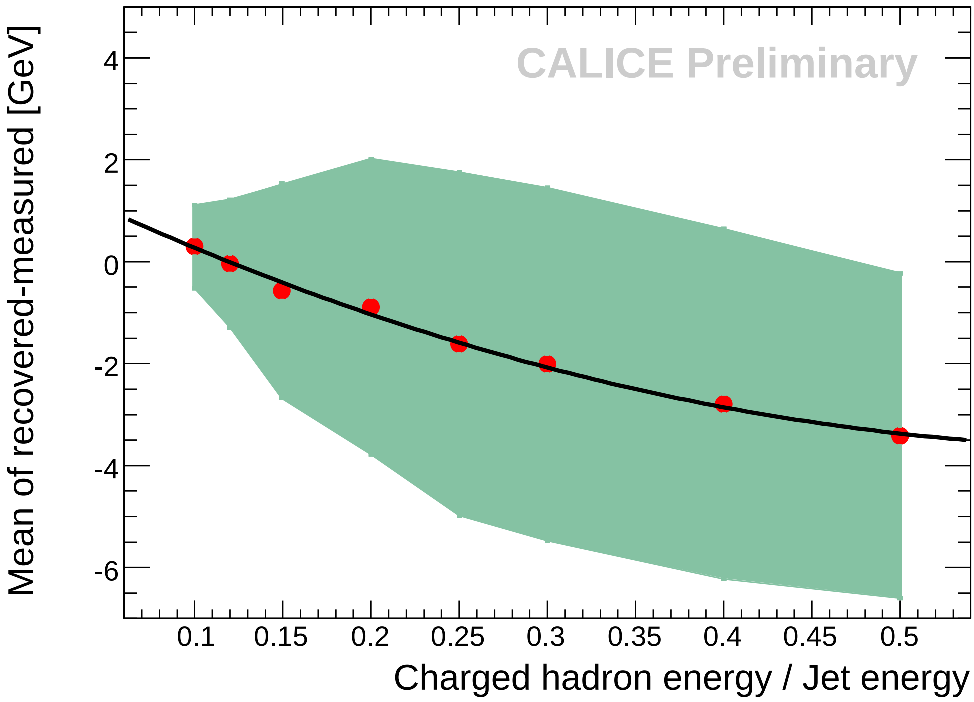

Figure 10: *Mean difference between the recovered energy and the measured energy for the 10 GeV neutral hadron vs. the charge hadron energy in the 100 GeV jet. Shaded is the region inside ±1 RMS_{90} deviation from the mean value. Solid line is the second-order polynomial fit.*

value of the mean difference between recovered and measured energy of 10-GeV neutral hadron in 100-GeV jet was estimated to be about 0.14 GeV. The result is positive since in our study we have included into the recovered neutral hadron shower all hits which PandoraPFA has not attached to the charged hadron shower. In that number there are some isolated hits and small clusters which actually belong to the charged hadron shower but were not identified correctly because of their remoteness.

We performed similar procedure for the squared RMS_{90} deviation of the recovered energy of the neutral hadron from its measured energy. The resulting value of RMS_{90} is 1.25 GeV. This integral is a coarse estimate for the neutral hadron energy recovery error, averaged over possible jet fragment energies. Due to the $D(z)$ form, the contribution to the integral from jet fragments with high energy is small. For this reason the average confusion error is close to the case of a charged hadron with medium (10 GeV) energy (see Fig. 7). The main part of the integral is due to small energy fragments. For them the remoteness from the shower axis is basically defined by the magnetic field, not by the transverse momentum. In other words, our approximation is self consistent. The confusion error of 1.25 GeV is smaller than the stochastic error of the 10-GeV hadron measurement in the calorimeter of about $0.6 \times \sqrt{10}$ GeV [4, 5]. Therefore the total error will not be much larger than the stochastic term.

The above estimation just illustrates the level of contribution to the confusion error from different charged jet fragments. In the real jet, due to non-zero transverse momentum, the distance between showers can be in some range around the distance used for our approximation. Besides one should more carefully take into account the fact that the distance between showers varies along the shower. As a result of these two factors, in reality the distances between neutral and charged jet fragments differ from those used in our estimation. However the dependence of confusion error on distance is rather flat around the approximate distances, see Fig. 7. Therefore one can expect that a more accurate calculation would not change too much the estimation of the confusion error obtained above.

7 Summary

To test the PandoraPFA algorithm, we have mapped pairs of CALICE test beam events shifted by the definite distances from each other onto the LDC detector geometry. Then we modified the treatment of tracks in the PandoraPFA processor for the case of straight tracks. In this study we have investigated the hadron energy range for a 100 GeV jet. For jet fragment energies from 10 GeV to 50 GeV we estimated the confusion error for the recovered neutral hadron energy caused by the overlapping of showers. We have confronted our result for test beam data with the result of Monte Carlo simulations for LHEP and QGSP_BERT physics lists. The results for the data and MC are in a good agreement. This fact together with the successful PandoraPFA performance for simulated jets [1] allows us to consider the PandoraPFA program as a good reconstruction tool for a full-size experiment.

The agreement between the PandoraPFA performance achieved with real calorimeter prototype data and with the MC simulation demonstrates that the extrapolation to the complete detector is reliable. No hidden imperfections in the real data (non perfect calibration, non uniformity of tile response, cross talk between tiles, noise, etc.) which could deteriorate the PFA performance were found. In particular, this conclusion is in agreement with the results of tiles non-uniformity impact study performed in [12]. We point out that in our study LHEP physics list gives worse predictions for test beam data than the QGSP_BERT one.

All the way we were trying not to simplify the task for PandoraPFA to disentangle showers in order to get a conservative estimate for the confusion error. In particular we assign all isolated hits to the neutral hadron shower and also we underestimate the separation of shower ends caused by the magnetic field.

We also have evaluated the averaged over possible jet fragment energies RMS_{90} deviation for the recovered 10 GeV neutral hadron energy from its energy measured in the calorimeter prototype. The large confusion error which occurs between a high energy charged hadron and a low energy neutral one is suppressed by the low probability to find the energetic hadron in the jet.

References

[1] M.A. Thomson, *Particle Flow Calorimetry and the PandoraPFA Algorithm*, NIM A611 (2009) 25-40.

[2] http://ilcsoft.desy.de/portal

[3] V. Morgunov, A. Raspereza *Novel 3-D clustering algorithm and two particle separation with tile HCAL*, Proceedings of International Conference on Linear Colliders (LCWS 04), Paris, France, 19-24 Apr 2004. Published in "Paris 2004, Linear colliders, vol. 1" 431-436; e-Print: physics/0412108.

[4] E. Garutti, B. Lutz, J. Samson, O. Wendt, *Preliminary results from hadron shower data with the CALICE tile AHCAL prototype*, CALICE Analysis Note CAN-011 (2008).

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

[6] The CALICE Collaboration, *Response of the CALICE Si-W ECAL Physics Prototype to Electrons*, CALICE Analysis Note CAN-008 (2008).

[7] The ILD concept group, *International Large Detector. Letter of Intent*, http://www.ilcild.org/documents/ild-letter-of-intent/LOI%20Feb2010.pdf/view

[8] The LDC group, *Detector Outline Document for the Large Detector Concept*, http://www.ilcild.org/documents/ldc/outlinedoc/view

[9] D. Green, *Dijet spectroscopy at high luminosity*, Fermilab Report Fermilab-Conf-90/151, 1990.

[10] O. Lobban, A. Sriharan, R. Wigmans, *On the energy measurement of hadron jets*, NIM A495 (2002) 107-120

[11] I.G. Knowles and G.D. Lafferty, *Hadronization in Z^0 decay*, J. Phys. G23 (1997)731

[12] F. Sefkow, A. Lucaci-Timoce, *Mokka Studies of AHCAL Tiles Gaps and Non-uniformities*, http://ilcagenda.linearcollider.org/getFile.py/access?contribId=265&sessionId=20&resId=1&materialId=slides&confId=4175