Selected simulations on hadron showers for AHCAL

Christian Röske

Ludwig-Maximilian-University Munich, Germany

Supervisor: Shaojun Lu

September 11, 2014

Abstract

This paper describes my work within the FLC-group, which is currently doing researches on the proposed International Linear Collider (ILC) and its system of particle detectors, the International Large Detector (ILD). More specifically, my task was to run simulations on the appearing hadron showers, while focusing explicitly on the behavior of the Analogue Hadron Calorimeter (AHCAL, which is a part of the ILD). Besides the investigation of the different types of correction the simulation software uses to converge on the reality, this also involves a comparison with real experimental test-beam data.

Contents

1. Software and Procedure

All simulations that appear in this report have been run using the pre-installed software. These include the ILC software package version 01-17 and CALICE software with its database. CALICE software again includes MOKKA, which does the actual simulation based on Geant4, as well as Marlin for the Digitization and Reconstruction. The physics list used is "QGSP_BERT"

To run a new simulation one first has to set up the basic parameters, which are the beam's particle type (i.e. pions, kaons ...), the beam's energy and the number of events one would like the simulation go through [1]. Note that a variation in the parameters may change the computation time and the size of the outputted data in a significant way. Especially with an increasing number of events (which is indispensable for high precision) and for very high beam energies, the computation time limits the work flow. So at some point, the jobs that I had run on the local machine so far, now had to be committed to the computer grid NAF.

After the primary simulation by MOKKA is finished, one can move on to the Digitization and then to the Reconstruction, which will give one the final output data. Before running the Digitization there is the possibility to activate or deactivate the different types of correction that are offered to deal with peculiarities and uncertainties of the detector. Thus it will provide a more close approximation to the real experiment.

The simulation output data, as well as the test-beam data are both stored in the same type of file (*.root) and can be evaluated with the corresponding ROOT-program, developed by CERN. This standardization considerably simplifies the comparison between simulation and reality.

Gaussian fits have been done using ROOT, further fits were done with Gnuplot.

2. Basics

2.1. ILD & AHCAL

This very short introduction will not go into detail, but should provide the reader the necessary knowledge to understand this work.

The **I**nternational **L**arge **D**etector (ILD) is a proposed detector for the **I**nternational **L**inear **C**ollider (ILC). The ILC is planned to collide electrons and positrons with a collision energy of 500 GeV, which can be further upgraded to 1000 GeV.[1]

The ILD's design will be a barrel shape. The individual parts of the detector are arranged in layers around the interaction point. They are passed through, one after another, by incoming particles and the showers they produce on their way. Every single part of

---

[1]There are even more parameters like the position of the particle generator, the beam direction and the position of binomial smearing but they have not been changed for this work.

the ILD fulfills its own individual functions. The **A**nalogue **H**adron **Cal**orimeter (AH-CAL or AHC) is one of those parts. Besides there is p.e. also the **E**lectro-**M**agnetic **C**alorimeter (EMC) which is located right inside the AHCAL, so an incoming particle will first have to go through the EMC before it enters the AHCAL. The AHCAL (ans also the EMC) specializes in hadrons, for which the particle showers usually are quiet widespread and therefor usually stretch out over more than just one detector.[1]

Every single cell (which uses **Si**licon **P**hoto-**M**ultipliers SiPM) of the consecutively ordered layers of the AHCAL can be calibrated using **M**inimum **I**onizing **P**articles (MIPs), usually muons. This is why all energy-measurements done by the detector will be exported in multiples of the energy-response of such a MIP.

2.2. Cut

After one has done a simulation or experiment and has received the root-file, with all available data attached to it, it is a good idea to first have a look at how the single shower events are perceived by the different detectors of the ILD. For that, one can have a look at the EMC and the AHCAL.
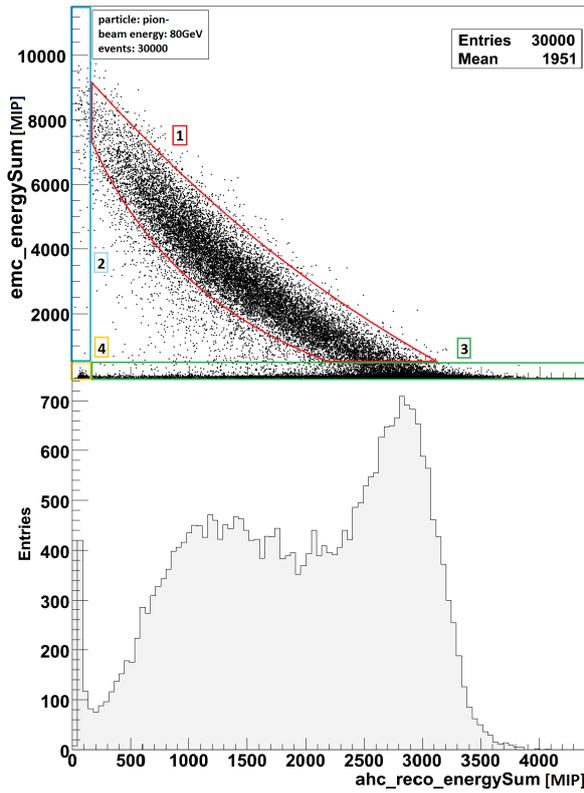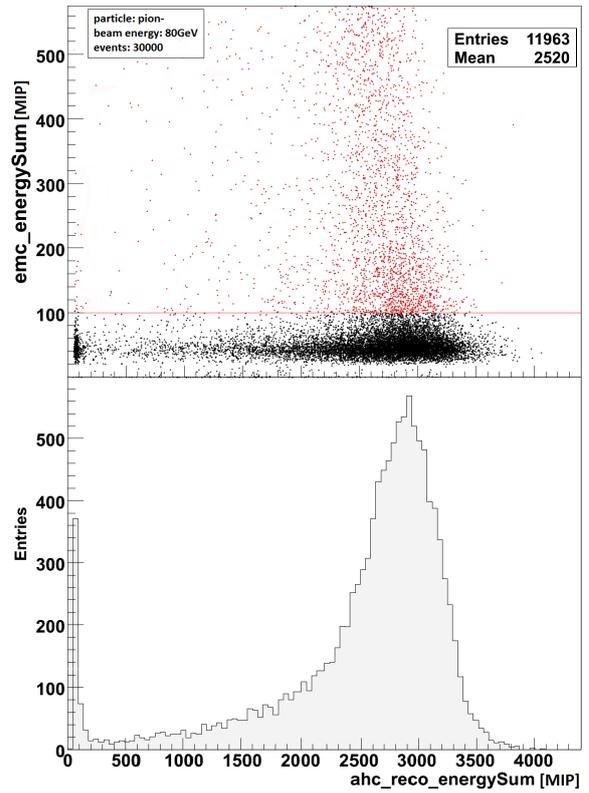

Figure 1: EMC vs. AHCAL without cut

Figure 2: EMC vs. AHCAL with cut

Figure 1 visualizes this in the following way: Each dot in the upper plot represents one single shower event and its position indicates, which amounts of energy have been mea-

sured for this event by the EMC (vertical axis) and the AHCAL respectively (horizontal axis). This plot has been taken for a simulation with 30000 events and pions[2]. The lower plot shows the summation of the events in the upper plot onto the horizontal axis. In other words, this is just the energy distribution detected by the AHCAL[3].

As described above, especially hadron showers are very wide spread and typically stretch out over more than one detector, which means that only a part of the total shower energy is noticed within one detector. This behavior is clearly visible in figure 1 as most of the dots appear in area 1 (red) (For most events the following is true: The higher the energy-response of the EMC, the lower the energy-response of the AHCAL and vice versa). Yet if we would like to obtain events, that have mainly been sensed by only one detector, so all the shower energy has been exposed in here, we have to take a look at either area 2 (blue, EMC) or area 3 (green, AHCAL)[4]. A closeup of area 3 can be found in figure 2 (upper plot).

Beneath about 100 MIP, one can see a distribution, which looks clearly different from the distribution in area 1, and this is the one that is relevant for the AHCAL. To isolate this distribution one simply cuts of all events above an EMC-energy-response of 100 MIP (red dots). For that, about two thirds of the events have to be thrown away. The lower plot of figure 2 displays the new cutted energy-distribution for the AHCAL[5]. This now is the distribution we are going to work with and on which the measurements will be taken[6].

From now on, this cut will always be applied to the raw data.

3. Simulation

3.1. Basic Corrections

As mentioned above, basic correction-options can be activated for the digitization process to improve the quality of the simulation. There are actually four of them:

1. **Noise**

   Random fluctuations that are produced by electronics (i.e. from the SiPM or the readout system)

---

[2]A plot for some other particle types as an overview can be found in the appendix, figures 14 and 15

[3]As this paper focuses on the AHCAL, the energy distribution for the EMC has been left out, yet it can be found in the appendix, figure 16.

[4]Area 4 (yellow): These are the responses for the shower-causing particle itself

[5]A plot of the none-cut and the cut distribution in one picture can be found in the appendix, figure 17

[6]There are further, more advanced methods to improve the selection of events to obtain even more precision, for the comparison between simulation and experiment the simple cut should yet be sufficient.

2. **Optical (Inter-pixel) cross-talk**

   Photons which are produced within the avalanche inside the SiPM can have an effect on nearby pixels and cause them to shoot. The effect is proportional to the SiPM's gain [2][3]

3. **Detector saturation**

   As the number of pixels is limited and the pixels themselves have got a dead time, meaning a recovery time after firing, the maximum amount of hits that can be recorded in a certain timescale is also limited. See also figure 3.
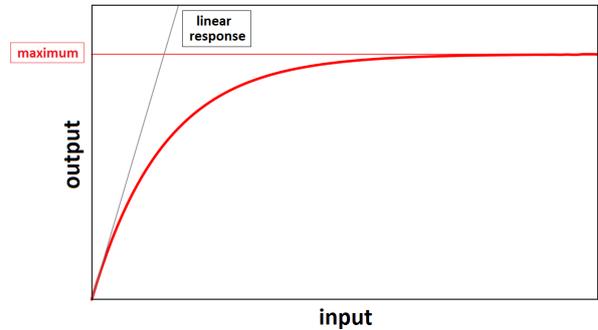
Figure 3: Detector saturation

4. **Binomial smearing**

The saturation of the detection and the optical cross-talk are the two most influential corrections, which is why the following analysis will privilege them. Therefore the different corrections are summed up into the following four collections:

1. **untouched (ganging):** This collection has only the noise correction applied to it. As the influence of the noise is very little, this stage has basically been left untouched.

2. **after cross-talk (xtalk)**: In this collection, optical cross-talk as well as noise are applied. As noise has a very little influence, this stage is dominated by the optical cross-talk.

3. **after saturation**: In this collection, noise, detector saturation and binomial smearing are applied. This stage is dominated by detector saturation.

4. **after cross-talk & after saturation (reco):** Here, all available corrections are applied and it is therefor dominated by optical cross-talk and detector saturation.

"Ganging", "xtalk" and "reco" are the names the software will use to refer to the defined collections 1,2 and 4. In figure 4 one can see how the correction of detector saturation (purple) and the correction of the optical cross-talk (green) respectively, each change the energy-distribution compared to the untouched one (black). As one can see, the cross-talk-correction shifts the distribution towards higher energies, whereas the saturation-correction shifts it towards lower energies. This behavior can easily be understood: When a detector underlies an optical cross-talk, this means that the detector will record additional, false signals (detector sees more). In contrast to that, a detector cell which is saturated cannot record any additional signals which might be there (detector sees less).
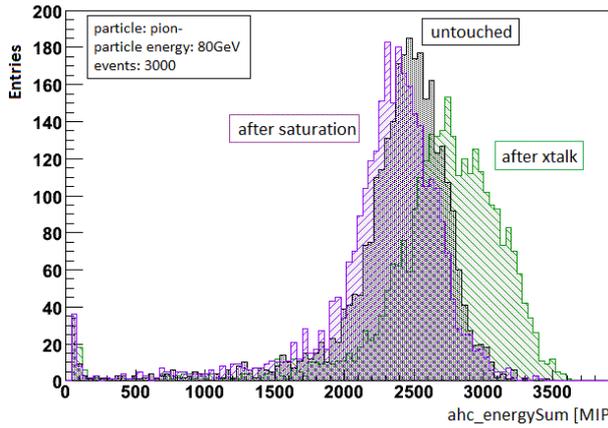
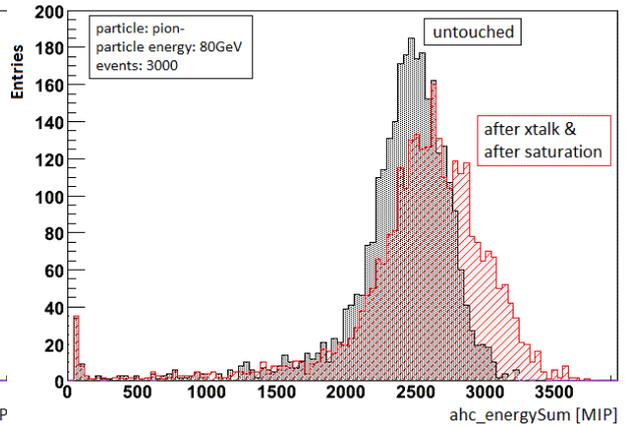Figure 4: after xtalk / after saturation

Figure 5: after xtalk & after saturation

In figure 5 one can observe how the correction of optical cross-talk and detector saturation work together(red). Except for collection 3, the number of corrections that are applied becomes bigger, therefor stages 1,2,4 should increasingly approach the reality in this order.

3.2. Their Behavior

To analyze the behavior of the different corrections in a vast size, it is only reasonable to look at them for different surroundings, like different beam energies. In figure 6 the energy-distributions for beam energies in a range from 0.5GeV to 400GeV are plotted. The general behavior is clearly visible: The higher the beam energy, the lower the amplitude, while the widths of the distributions increase.
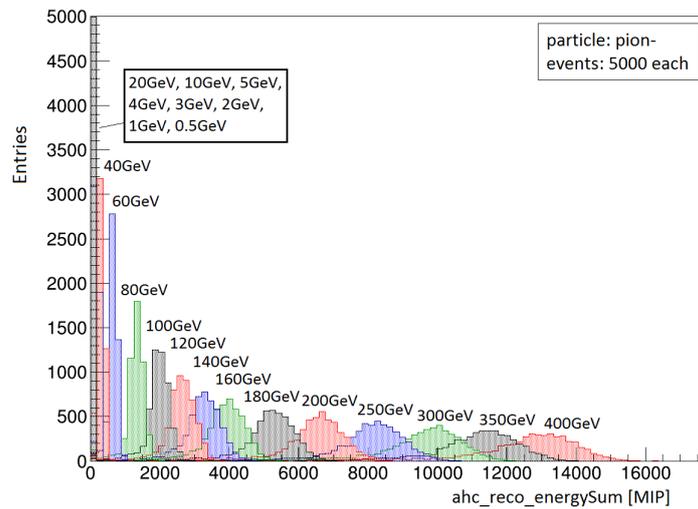

Figure 6: Various beam energies

This plot has only been done for one (after cross-talk & after saturation, collection 4) particular of the above-mentioned collections of correction, which is why it has no validity of the correlation between the different collections. Nevertheless, the described general behavior stays the same for all collections of correction.

The explicit correlation for collections 1,2,4 for varying beam energies will now be analyzed for the distribution's mean value, width(σ) and energy-resolution, assuming the energy-distributions are of a Gaussian nature.

As this part of the work only deals with the corrections the software does and which cannot directly be compared to the reality (there is p.e. no detector that underlies cross-talk but not saturation, so both effects cannot be analyzed independently, though this is possible in the simulation), this analysis is a rather qualitative one. For the particle beam, pions- have been used at an event count of 5000 each.

3.2.1. Mean value

In figure 7 one can see the mean values of the Gaussian fit, plotted for each of the collections 1,2,4 with respect to the beam energy E_π (upper plot). Error values have been adopted from the fit done by ROOT.

For collection 1 (only noise-correction), the plot follows a quite linear curve, which is expected, as an ideal detector should have a linear energy-assignment. It is also visible, that the mean values for collection 1 (black) are always the smallest, followed by the energy values for collection 4 (red). The values for collection 2 (green) are always the highest (See also figure 4 and figure 5).

(For better visualization the relative difference to the untouched is plotted below.)
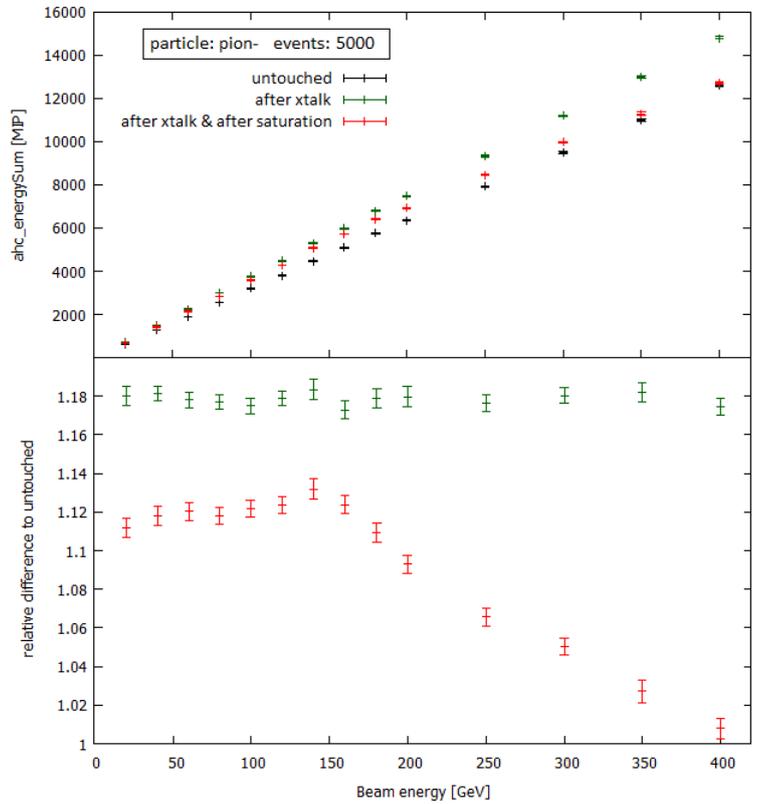

Figure 7: Mean value in comparison

This means, that for this energy range the influence of the cross-talk correction is always bigger than the influence of the correction of the detector saturation+binomial smearing.

A look at the lower plot yet reveals that the curve for collection 4 decreases substantially,

starting from a beam energy E_π of about 150GeV, while the curve for collection 2 stays rather linear in this area. For a beam energy of 400GeV the curve for collection 4 has nearly approached the curve for collection 1. As collection 4 and collection 2 basically only differ in the saturation-correction, this non-linearity has to come from there. This makes perfect scene because, especially for high energies, the detector-saturation causes that more and more detector cells are working to capacity, so they will not register any further energy that might be there.

Remember that "xtalk & saturation" is the collection where all available corrections have been applied, so the real detector should only have a rather linear energy-assignment for beam energies E_π below around 150GeV.

For a linear response, the following function can be used (a fit to the data will be made later)

$$\bar{E}(E_\pi) = A \cdot E_\pi + B$$

The energy measured by the detector is only the visible energy[7]. That is why the proportionality factor A can be used to assign a detector-response measured in MIP to a beam energy in GeV.

For every single detector cell, there is a further cut applied during the reconstruction process at about 0.5 MIP, to compensate noise. However this cut is usually not perfectly accurate as noise and actual data flow into each other. So depending on how much noise has been left or how much actual data has been thrown away, the offset, represented by B, can either be positive or negative.

---

[7]The detector consists of many consecutively ordered layers, yet only a fraction of those are active layers and can therefor be used for measuring. Additionally, there are dead areas between the detector cells, which decrease the visible energy even further.

3.2.2. Sigma

Now the focus lies on the width (σ) of the Gaussian-fitted distributions. Their devolution for different beam energies can be seen in figure 8.

The upper plot looks quiet similar to figure 7, as we again see a rather linear curve and again the order is the same: collection 1 has the lowest distribution's width, followed by collection 4 and collection 2 always has the highest distribution's width for this energy range.

The cross-talk basically increases fluctuations, so it is obvious that this leads to a higher width. As the saturation of detector cells reduces the total number of detector cells which are still able to observe a hit, this also limits the amount of fluctuations that can occur.
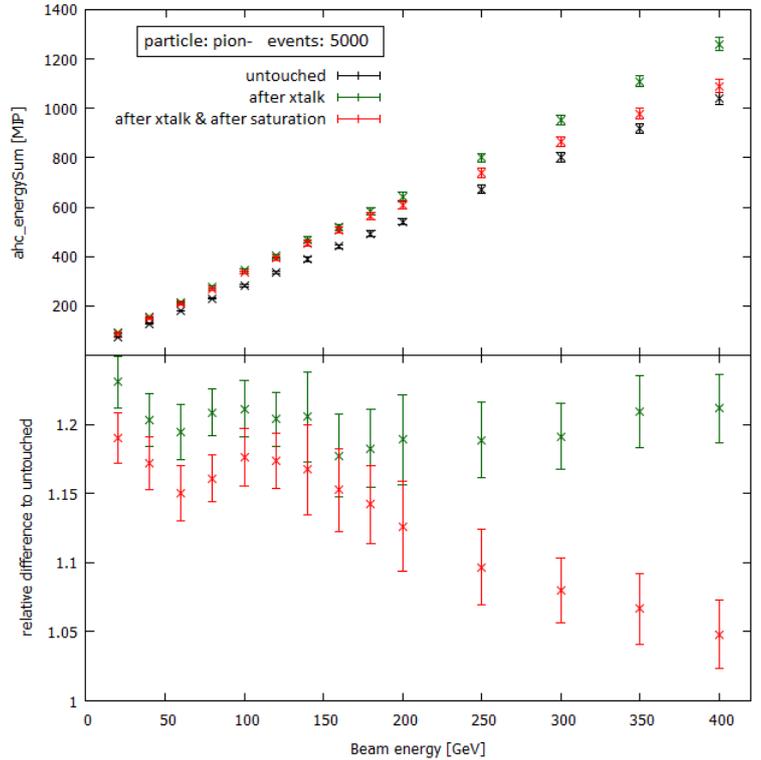

Figure 8: Distribution's width (σ) in comparison

For clarification, one could even consider an extreme case with a very high beam energy, in which all detector cells are saturated for every event. Then the detector (as it sums up over all detector cells) would only recognize that one maximum energy value for each event and therefor sigma would have reached zero.

Just like in figure 7 one can see, that the curve for collection 4 drops rapidly for beam-energies above 150GeV, which seems to be the limit for the detector. The cross-talk correction does not produce any abnormalities.

Even though the curves look linear, the function for fitting the distribution's width(σ) should be:[8]

$$\sigma(E_\pi) = \sqrt{a^2 \cdot E_\pi \cdot A^2 + b^2 \cdot (E_\pi \cdot A)^2 + (c \cdot A)^2}$$

where a represents the stochastical term, which depends on the calorimeter design and the incoming particle

             b represents calibration uncertainties, response inhomogeneities of sensitive areas and leakage effects

             c represents electronic noise and background signals[4]

---

[8]The formula for the energy resolution [4] has been rearranged for $\sigma(E_\pi)$

As a and c are quiet small, $b^2 \cdot (E_\pi \cdot A)^2$ is the dominant term and $\sigma(E_\pi)$ can be approximated with:

$$\sigma(E_\pi) \approx |b| \cdot E_\pi \cdot A$$

3.2.3. Energy resolution

Figure 9 displays the energy-resolution $\sigma/\bar{E}$ for the different collections with respect to the beam-energy.

Once again, there is a clear ranking: collection 1 achieves the best resolution, followed by collection 2 and collection 4 has got the worst energy resolution for this energy range. The more corrections applied, the worse the resolution becomes.

Abnormalities, which have been observable for $\bar{E}$ and σ at an beam-energy level above 150GeV, seem to have canceled each other out, at least none are visible.
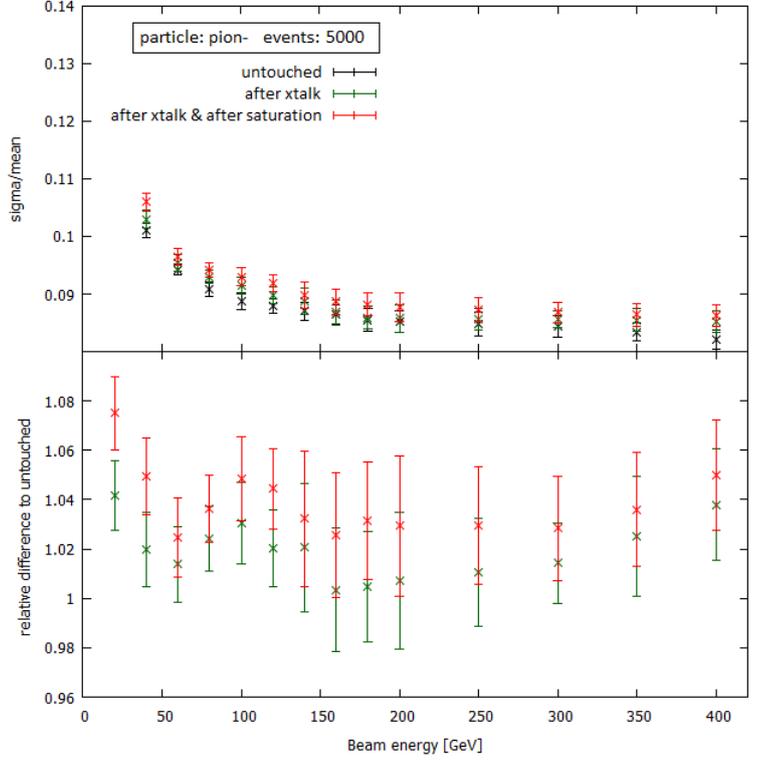
Figure 9: Energy-resolution in comparison

The energy function should follow the following function [4]

$$\frac{\sigma}{\bar{E}}(E_\pi) = \sqrt{\left(\frac{a}{\sqrt{E_\pi}}\right)^2 + b^2 + \left(\frac{c}{E_\pi}\right)^2}$$

where a,b and c are the same parameters that have been used for the distribution's width σ.

11

4. Comparison with test-beam data

For now, the examination only covered the simulation's behavior when exposed to different corrections and beam energies. Yet the quality of the simulation can only be determined in comparison with reality. For this reason this work is using test beam data taken at CERN in 2007 to match the simulation with the experiment.

The test beam provides a very high amount of events (18000-250000), while the simulation can only provide 5000 events, due to the long computation time. Therefor the measurement's errors are higher for the simulation.
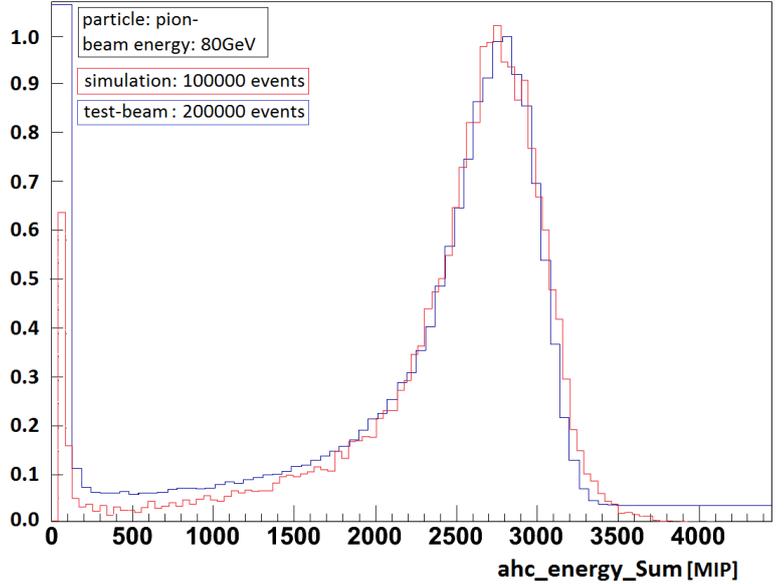
Figure 10: Simulation vs. test beam, normalized

Figure 10 exemplifies the comparison between test beam and simulation for a beam energy of 80GeV. The simulation used for the picture (with 100000 events) has only been done for this specific energy.[9]. Note that the distributions are normalized to make matching easier.

The following analysis again focuses on the mean value, sigma and the energy resolution, yet this time the beam-energy only ranges from 8GeV to 80GeV, as provided by the test beam.

Also, for the comparison with the test-beam, only collection 4, so the one with all available corrections applied to it is consulted. Furthermore, the simulations that will be used are not the exact same ones that have been used so far, as they had to be adjusted to the test beam data[10]. Finally, there have been added some more simulations to provide more data within the test beam's energy range.

Figures 11, 12 and 13, which can be found in the following sections, match simulation data with test beam data. The blue, fitted curves (gray area displays error range) are always the fits for the test beam data. The fits for the simulation data (always collection 4) are not displayed, yet their fitted parameters are. The lower plots within the mentioned figures display the relative errors due to the visualized, fitted curve (blue). The fitting-functions are the ones already written down in section "Simulation".

---

[9]This one should intentionally be used for single-cell measurement, this is also the reason for the high amount of events.

[10]The adjustment was a matter of the cut, described in section 3.2.1 , describing parameter B
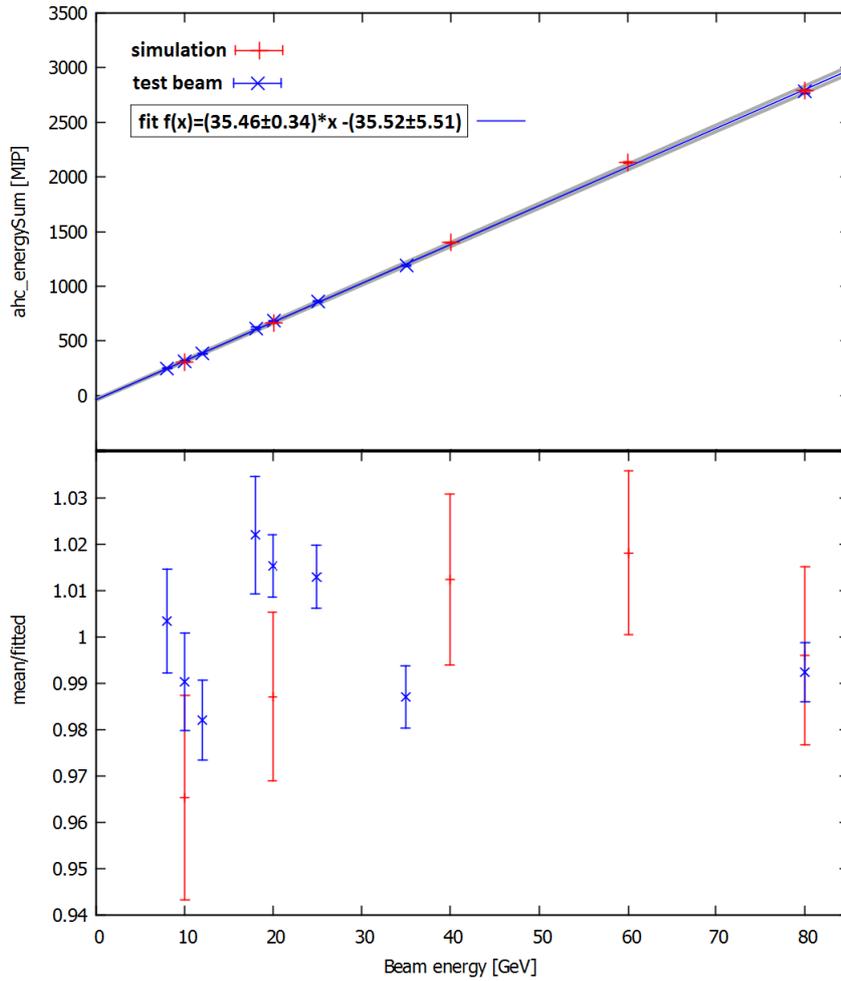
4.1. Mean value

Figure 11: Mean value, simulation vs. test beam

Fit for test beam:

$$A = (\ \ 35.46 \pm 0.34)\, \frac{\text{MIP}}{\text{GeV}}$$
$$B = (-35.52 \pm 5.51)\, \text{MIP}$$

Fit for simulation:

$$A = (\ \ 35.83 \pm 0.16)\, \frac{\text{MIP}}{\text{GeV}}$$
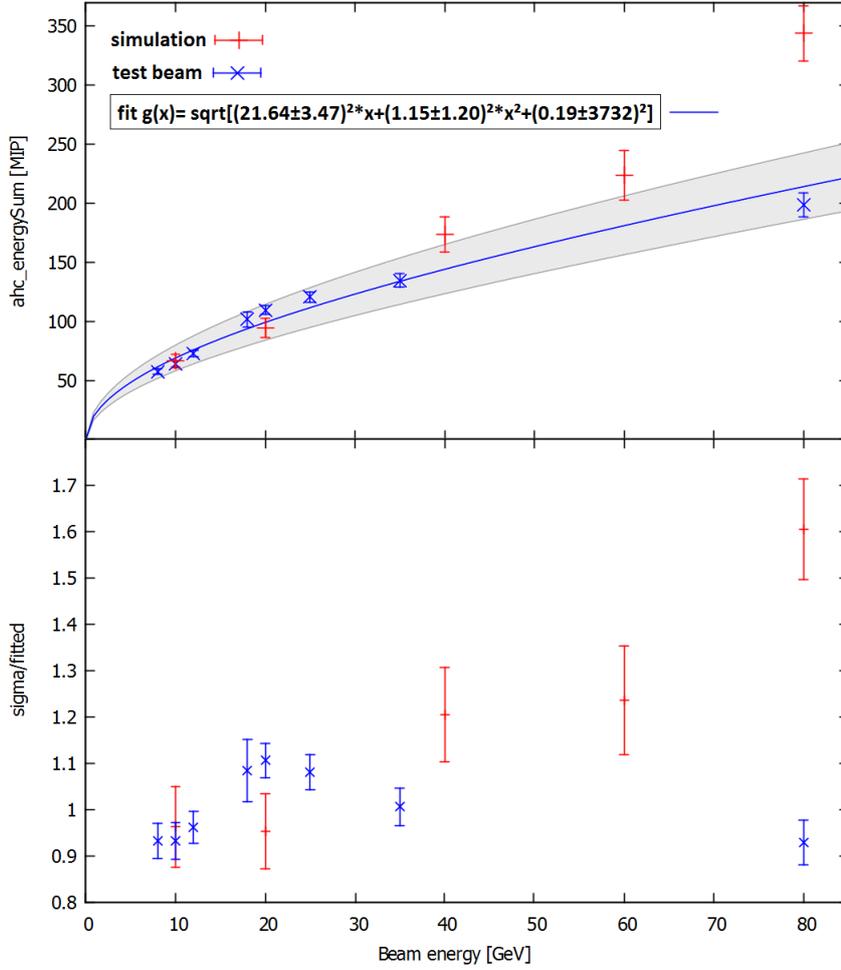$$B = (-32.49 \pm 3.51)\, \text{MIP}$$

4.2. Sigma

Figure 12: Sigma, simulation vs. test beam

Fit for test beam:

$$a = \frac{(21.64 \pm 3.47)\frac{\text{MIP}}{\sqrt{\text{GeV}}}}{A} = \frac{(21.64 \pm 3.47)\frac{\text{MIP}}{\sqrt{\text{GeV}}}}{(35.46 \pm 0.34)\frac{\text{MIP}}{\text{GeV}}} = (0.6103 \pm 0.0092)\sqrt{\text{GeV}}$$

$$b = \frac{(1.15 \pm 1.20)\frac{\text{MIP}}{\text{GeV}}}{A} = \frac{(1.15 \pm 1.20)\frac{\text{MIP}}{\text{GeV}}}{(35.46 \pm 0.34)\frac{\text{MIP}}{\text{GeV}}} = (0.0324 \pm 0.0335)$$

Fit for simulation:

$$a = \frac{(23.48 \pm 1.49)\frac{\text{MIP}}{\sqrt{\text{GeV}}}}{A} = \frac{(21.64 \pm 3.47)\frac{\text{MIP}}{\sqrt{\text{GeV}}}}{(35.83 \pm 0.16)\frac{\text{MIP}}{\text{GeV}}} = (0.6040 \pm 0.0942)\sqrt{\text{GeV}}$$

$$b = \frac{(3.08 \pm 0.81)\frac{\text{MIP}}{\text{GeV}}}{A} = \frac{(3.08 \pm 0.81)\frac{\text{MIP}}{\text{GeV}}}{(35.83 \pm 0.16)\frac{\text{MIP}}{\text{GeV}}} = (0.0860 \pm 0.0222)$$

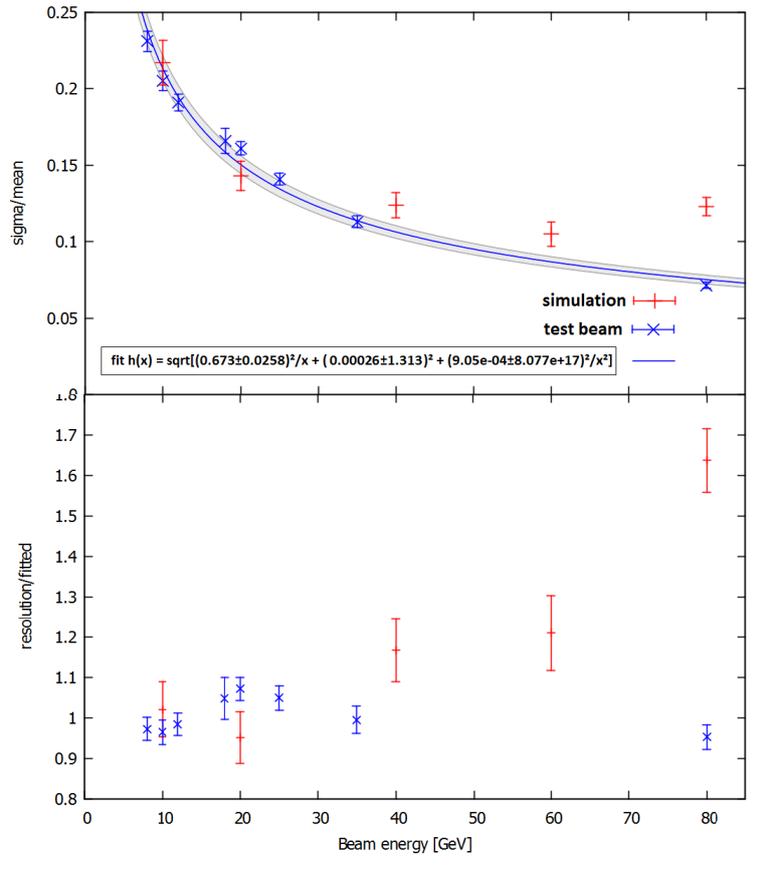c is unstable in fit.

4.3. Energy resolution

Figure 13: Energy resolution, simulation vs. test beam

Fit for test beam:

$$a = (0.673 \pm 0.026)\sqrt{\mathrm{GeV}}$$
$$b = (0.00026 \pm 1.313)$$

Fit for simulation:

$$a = (0.451 \pm 0.063)\sqrt{\mathrm{GeV}}$$
$$b = (0.0915 \pm 0.0043)$$

c is unstable in fit.

5. Conclusion

The table below lists all fitted parameters (A, B, a, b, c) , each for simulation and test beam.

	mean value		sigma		resolution	
	simulation	test beam	simulation	test beam	simulation	test beam
A $[\frac{\text{MIP}}{\text{GeV}}]$	35.83 ± 0.16	35.46 ± 0.34				
B [MIP]	-32.49 ± 3.51	-35.52 ± 5.51				
a $[\sqrt{\text{GeV}}]$			0.6040 ± 0.0942	0.6103 ± 0.0092	0.451 ± 0.063	0.673 ± 0.026
b [1]			0.0860 ± 0.0222	0.0324 ± 0.0335	0.0915 ± 0.0043	0.00026 ± 1.1313
c [GeV]			unstable	unstable	unstable	unstable

Table 1: Summary of fitted values

Values for A and B lie within the error range of one another and therefor match quiet well. The same is true for a, except not for the simulation value that has been fitted using the resolution. Taking into account the declared errors, even b can be matched very closely. Yet note that the relative errors are occasionally very high. Values for c could not be compared as the fitted values are unstable.

In general one can say that the experiment and the simulation are in a good agreement for the mean value in the observed energy range. The comparison for the distribution's width (and therefor also the energy resolution) obtains a significant difference. The sigma-value for the simulation is usually higher than the one for the test beam. A reason for this could be the use of the the physics list QGSP_BERT. Anyway, the differences can not be explained by the use of insufficient events, as even figure 10 displays this characteristic for 100000 events.

If one beliefs the simulation, we would expect the detector to fail for beam energies much higher than about 150GeV, at least the response will be highly non-linear.

References

[1] ILD Concept Group, *The International Large Detector - Letter of Intent (2010), S. 1,2*

[2] Benjamin Lutz, *Hadron Showers in a Highly Granular Calorimeter (Hamburg 2010), S. 19*

[3] L. Gallego, J. Rosado, F. Blanco and F. Arqueros *Modeling crosstalk in silicon photomultipliers (Madrid), S. 1*

[4] Nils Feege, *Low-Energetic Hadron Interactions in a Highly Granular Calorimeter (Hamburg 2011), S. 12*
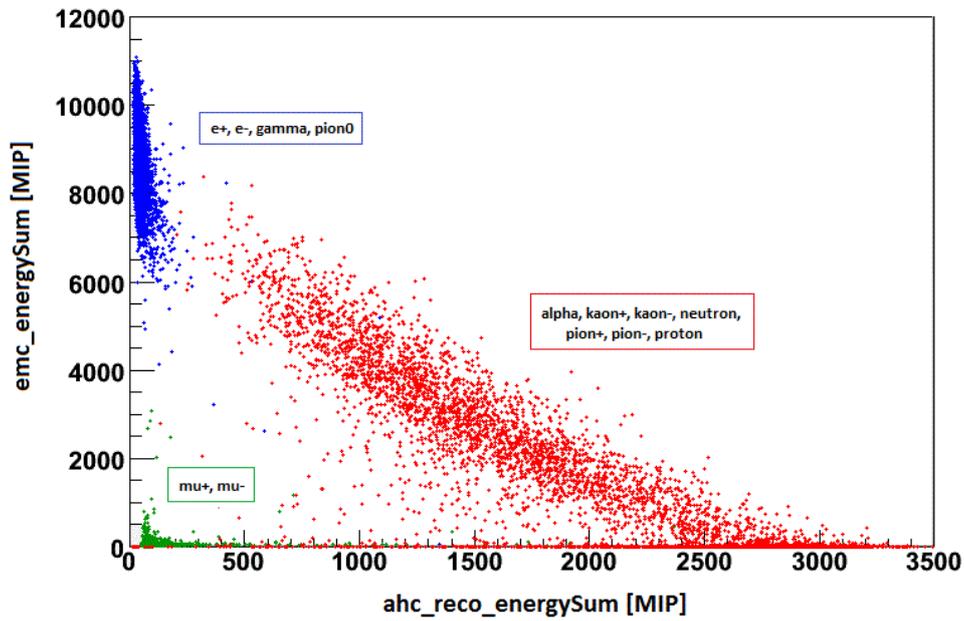
A. Appendix

Figure 14: EMC vs. AHCAL for different particle types
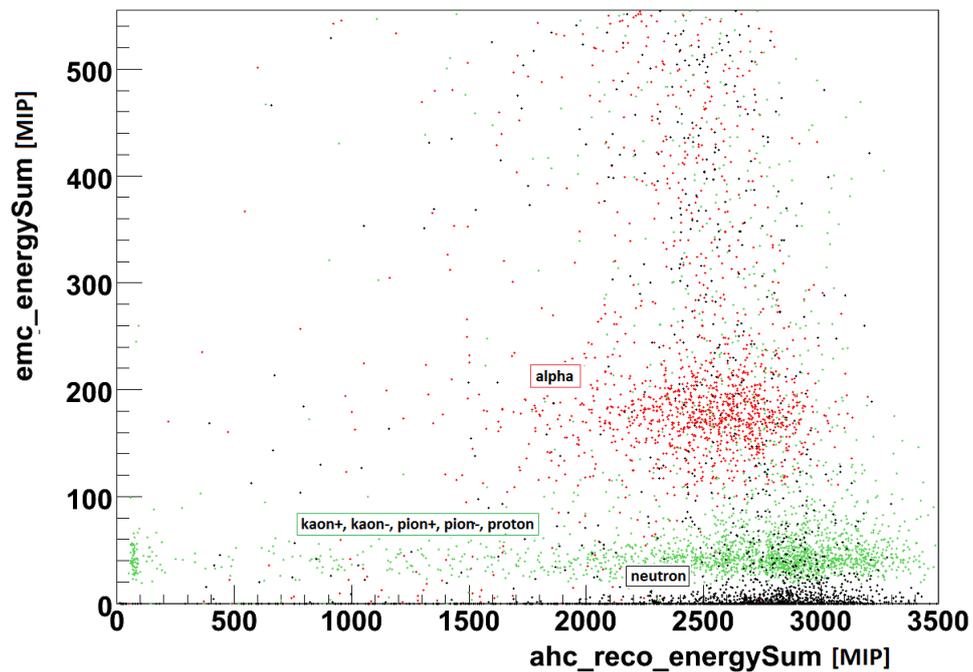

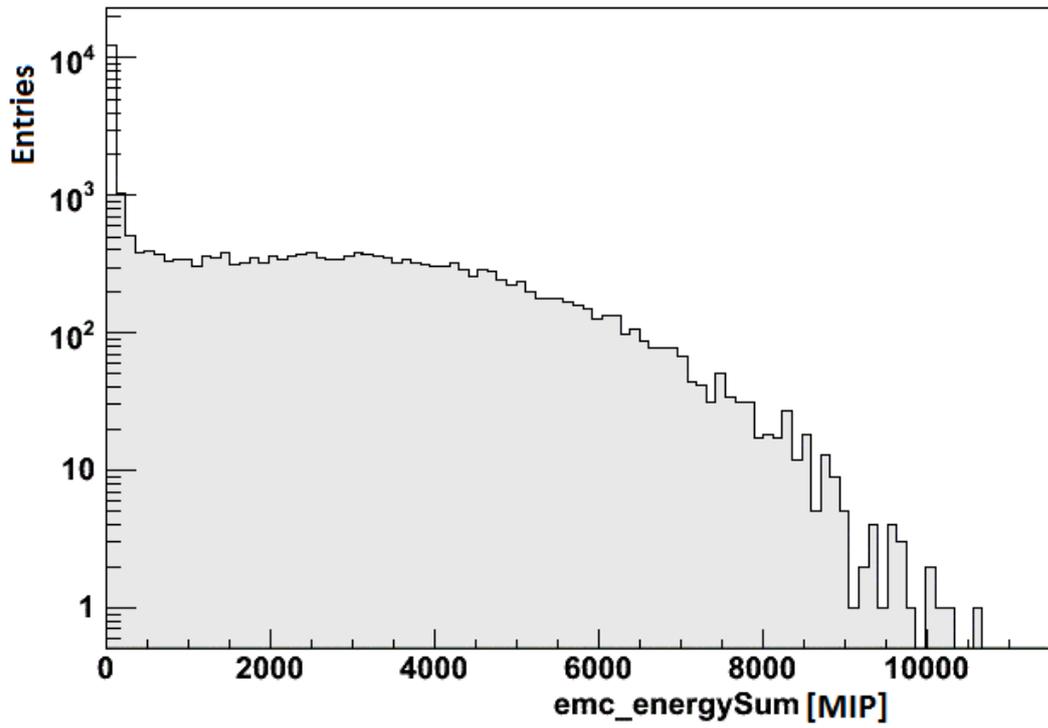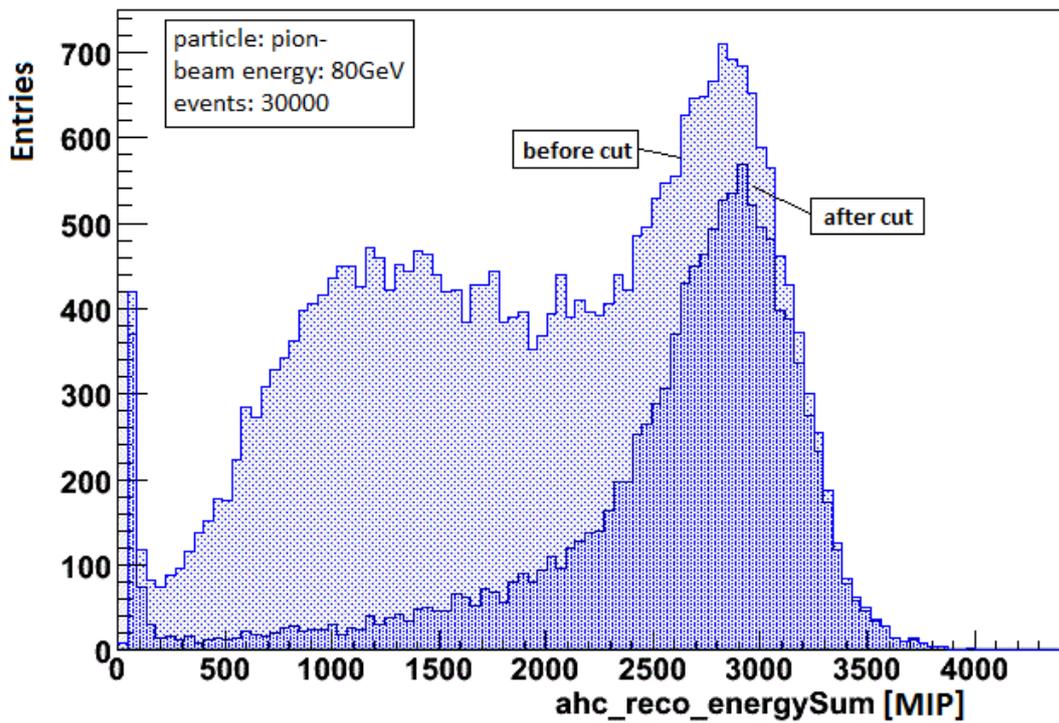Figure 15: EMC vs. AHCAL for different particle types, closeup

Figure 16: EMC

Figure 17: AHCAL: Before and after cut

19