

# Comparison of Particle Flow Algorithms Using $e^+e^- \rightarrow ZHH$ events at 500 GeV Centre of Mass

*M. Faucci Giannelli, M.G. Green, F. Salvatore*

Royal Holloway University of London

## 1. Introduction

One of the requirements in the R&D for the International Linear Collider detectors is the development and test of a Particle Flow Algorithm (PFA) capable of high performances. In this work a comparison between two available algorithms is presented, using  $e^+e^- \rightarrow ZHH$  events at 500 GeV Centre of Mass (CM). This channel has been chosen from the list proposed by M. Battaglia et al [1] because it can test simultaneously almost all the sub-detectors. Therefore both tracking and calorimetry will affect the performances of the particle flow algorithms.

Since the aim of the Calice collaboration is to study in detail the calorimetry for ILC, in the analysis presented in this note the tracking has always been done using the same package. In this study, the calorimetric differences in the PFAs will be compared, focusing especially on the results from the clustering algorithms. The two clustering software and particle flow algorithms that have been compared are: Pandora [2] by M. Thomson and the combination of TrackwiseClustering and Wolf [3] by A. Raspereza.

An overview of the software used in the analysis is presented in section 2. In section 3 a calibration study is presented to confirm the correctness of the values used in the calorimeters digitization. The comparison between the two particle flow algorithms is presented in detail in section 4 and the problems arising from the detector acceptance and particles escaping detection are shown in section 5. In the last section the comparison between two detector models, LDC00Sc and LDC01Sc, is presented.

## 2. Event generation and detector reconstruction

The events were generated using the Pandora Pythia event generator v3.3 [4], including effects of initial and final state radiation, and bremsstrahlung. The mass of the Higgs was set to 120 GeV and its decay was forced to  $\bar{b}b$ . The energy of each beam was 250 GeV for a total CM energy of 500 GeV. The electron beam was simulated with 80% polarization. For the detector simulation Mokka v06-00 was used [5].

The event reconstruction was performed using Marlin 09-04 [6], which includes v1.21 of LEPTrack processor for track reconstruction [7]. V1.5 of the TrackwiseClustering processor, v1.6 of the Wolf processor and the pre-release version of PandoraPFA were used for the clustering and the particle flow.

The two detector models studied were LDC00Sc and LDC01Sc. The main differences between the two detectors are the smaller radius of the TPC and a reduced number of sensitive layers in the electromagnetic calorimeter in LDC01Sc with respect to LDC00Sc. More details on the two detectors can be found in [8].

### 3. Calibration

Due to the small dimension of the silicon pads in ECal and of the scintillator tiles in HCal, the amount of energy deposited in the calorimeters depends on the number of steps performed in the detector simulation. This value has been changed since the calibration was originally done using Mokka v5.4 [9] thus there was some concern on the values of the calibrations used in the digitization of the calorimeters and these were studied using single pion and single electron events before performing the reconstruction of ZHH events. These single particle events were generated with the same version of Mokka used to perform the detector reconstruction of the ZHH events. The results obtained are summarized in the following:

- ECal:

The electromagnetic calorimeter is divided into two parts (30+10 layers for LDC00Sc, 20+10 for LDC01Sc) thus two values are needed for the calibration. The total radiation length in the two cases is anyhow the same. The calibration values used are presented in table 1.

- Hcal:

Only one value is needed for the calibration of the scintillator hadronic calorimeter. This value is independent of the detector model since there are no differences in this detector between LDC00Sc and LDC01Sc.

**Table 1 Value used in the digitization of calorimeters in the reconstruction. The electromagnetic calorimeter (ECal) is divided into two regions with different thickness of passive material.**

	ECal 1	ECal 2	HCal
LDC00Sc	27.3	72.4	26
LDC01Sc	40.4	71.5	26

The plots in figure 1 show the reconstructed energy in the Ecal obtained using 10 GeV electrons in LDC01Sc, with particle properties reconstructed using the two different PFAs introduced above. There are almost no differences between Wolf and Pandora and in both cases the peak is centred at 10 GeV. To select pions that only interacted in the hadronic calorimeter, particles that release less than 0.52 GeV in the Ecal were used in this analysis. In figure 2 the cluster energy of 10 GeV pions is plotted for both PFAs in LDC01Sc. In this case there are differences in particular the mean value for Wolf is high. This difference must be due to the PFAs since the calibration is the same; this value has been optimized for Pandora. Similar results were obtained for LDC00Sc.

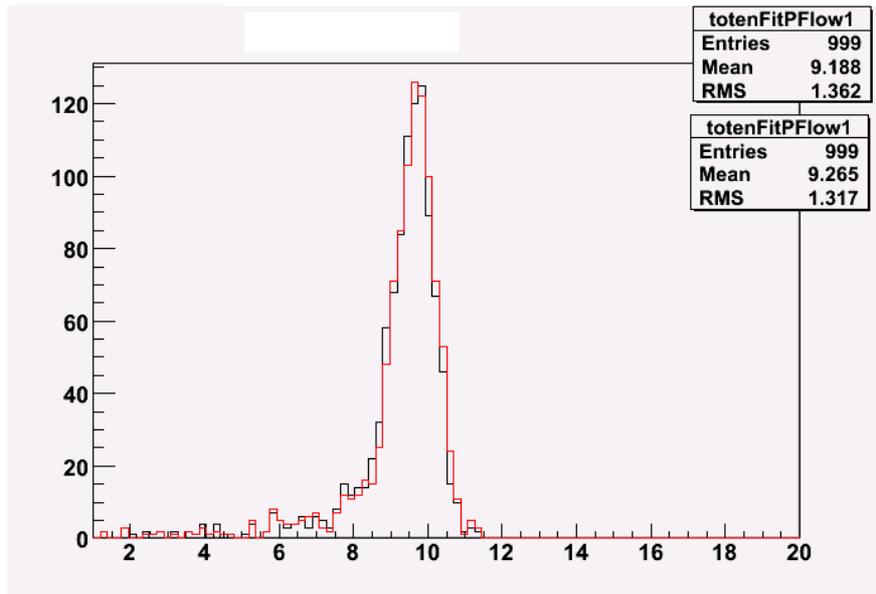


Figure 1 Reconstructed ECal energy for 10 GeV single electrons using Wolf (red) and Pandora (black) in LDC01Sc

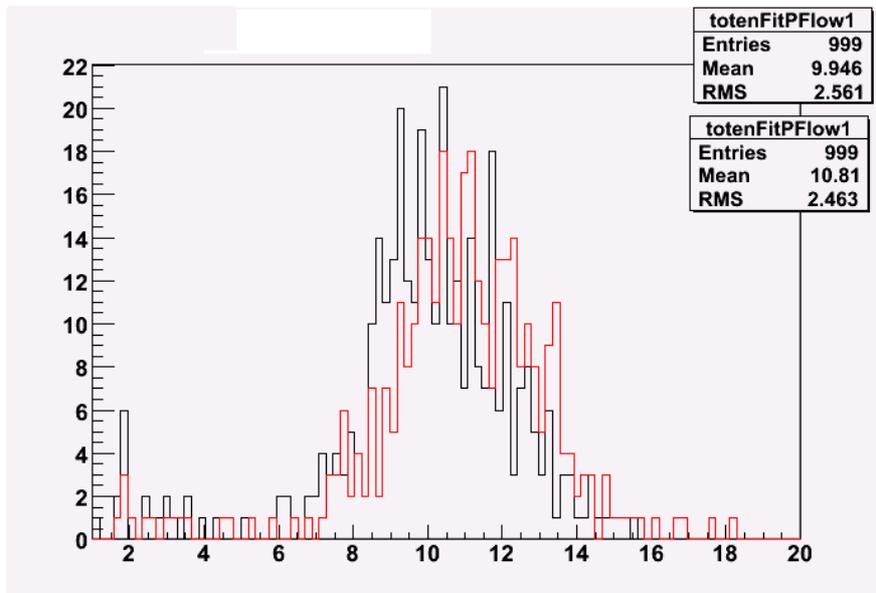


Figure 2 Reconstructed HCal energy for 10 GeV single pions using Wolf (red) and Pandora (black) in LDC01Sc.

#### 4. Analysis of ZHH events

For this analysis we concentrated on a sample of ZHH in which the Z decays into electrons or muons in order to have good tagging of the Z so that we can better control the reconstruction of the Higgs.

After the particle flow algorithm reconstructs all particles, the two particles that better reconstruct the Z are selected while the others are clustered to form the jets. Since at present there is no muon identification, pion mis-identification is much higher in these events than in the electron case, resulting in a higher combinatorial in the Z selection. A wrong selection affects jet reconstruction and therefore the Higgs mass; therefore we have looked separated at Z decays into electrons and muons.

The particles remaining in the event after the Z selection are forced to form 4 jets using the Durham algorithm [10]. The jets are paired in the three possible ways and the combination that minimizes the quantity

$$D^2 = (M_{ij} - M_H)^2 + (M_{kl} - M_H)^2 \quad (1)$$

is selected. In (1)  $M_H$  is the generated mass of the Higgs, while the indexes represent the jet combinations.

## 5. Comparison of PFA algorithms in the case $Z \rightarrow e^+e^-$

Figure 3 shows the mass of the two reconstructed Higgs obtained using the LDC01Sc model. There is a clear difference between the two PFAs and neither reconstructs the Higgs correctly: Pandora reconstructs a low mass while Wolf reconstructs higher mass. However this plot alone is not enough to judge the reconstruction since every event enter the plot twice. Therefore the information regarding the correlation between the two is lost. The correlation, although very small, may be avoided studying variables that include both reconstructed masses such as those presented in figure 4 and 5. The first is the distribution of (1), which for correctly identified jets from Higgs decay should peak at zero. The latter plot is the distribution of the variable

$$D = M_{ij} + M_{kl} - 2M_H. \quad (2)$$

This distribution should be symmetrical around zero for a good reconstruction. The width of the distribution is another parameter to consider and reflects how well the two Higgses are reconstructed; a small width means the majority of events are well reconstructed while a large sigma means many events are poorly reconstructed.

At this point of the analysis, the results obtained by the two algorithms are comparable: Wolf is reconstructing more events correctly (see the peak at zero for Wolf in figure 5) but Pandora has a smaller width in the D distribution (see the smaller RMS in the same plot for Pandora).

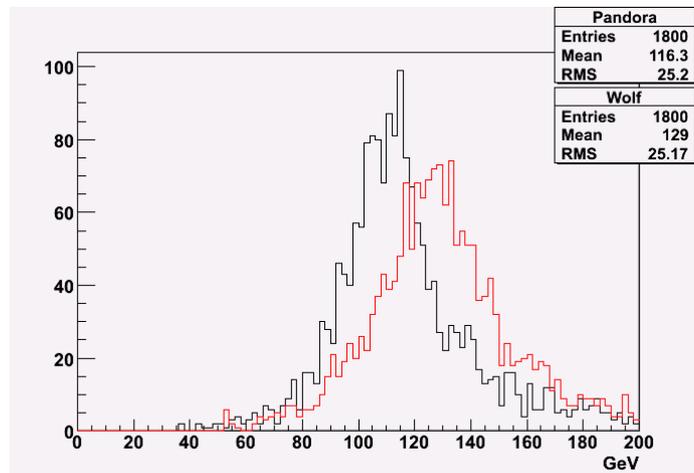


Figure 3 Reconstructed Higgs mass in LDC01Sc using Wolf (red) and Pandora (black). The generated Higgs mass is 120GeV.

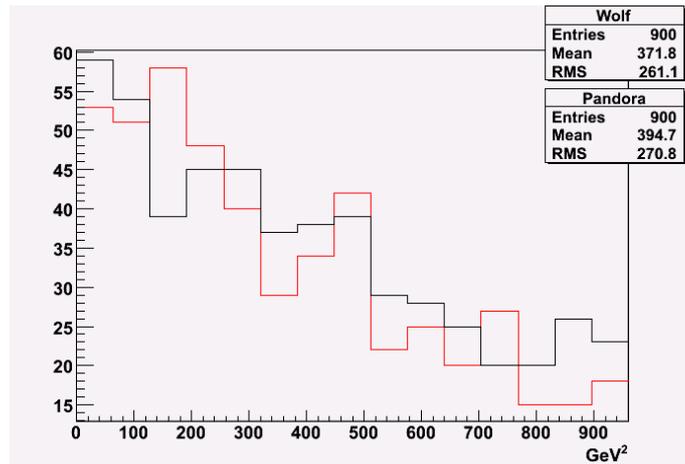


Figure 4 Distribution of  $D^2$  (as defined in (1)) for Wolf (red) and Pandora (black) in LDC01Sc.

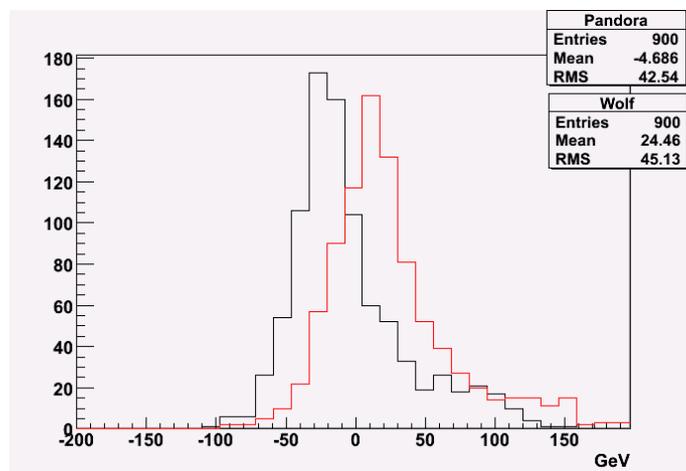


Figure 5 Distribution of  $D$  (as defined in (2)) for Wolf (red) and Pandora (black) in LDC01Sc.

These results are obtained using only reconstructed particles, thus not taking into account particles escaping in the beam pipe, particles not reconstructed by the tracking algorithm at small angle and neutrinos. Since all these factors should increase the mass of the Higgs, it is expected that the result obtained with Pandora would be better compared to the one obtained with Wolf. A detailed analysis on this topic is presented in the next section.

A more detailed analysis of the reconstruction in the calorimeters has been carried out. Two variables useful to compare the PFA in this case are the number of neutral particles reconstructed and their total energy. These have been obtained independently for the electromagnetic and the hadronic calorimeters. In figure 6 the total number of photons and their total energy is shown. Comparing these two distributions for Pandora and Wolf it can be noted that, while the number of reconstructed photons is almost the same, the total energy is very different between the two algorithms. In figure 7 the same plots are presented for the hadronic calorimeter. In this case there is a different effect; the number of neutral hadrons is quite different but the total energy is almost identical. The later result is unexpected due to the difference observed in the single pion reconstruction. One possible interpretation of these data is that the way the energy is clustered in the hadronic calorimeter is in fact not so relevant.

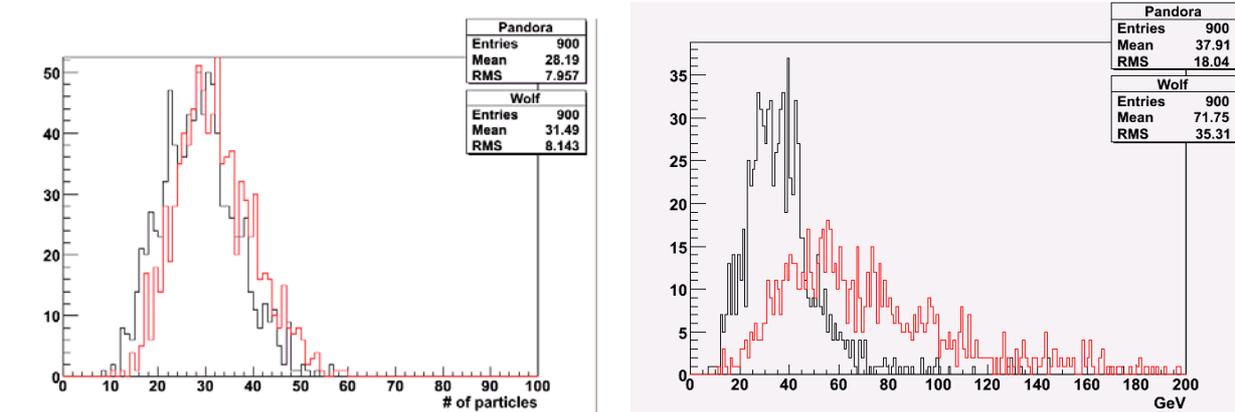


Figure 6 Distribution of the number of reconstructed photons and their total energy for Wolf (red) and Pandora (black).

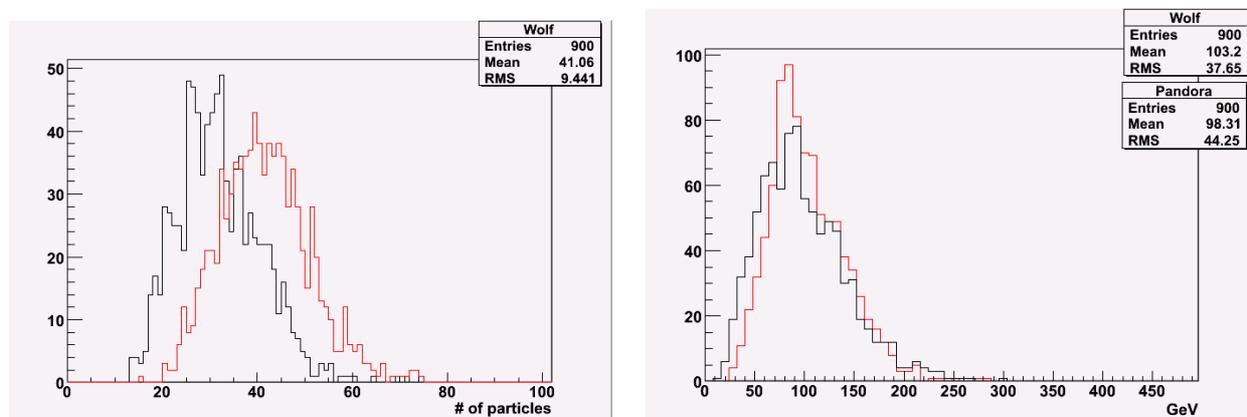


Figure 7 Distribution of the number of reconstructed neutral hadrons and their total energy for Wolf (red) and Pandora (black).

For completeness, the same plots for charged particles are presented in figure 8. The difference is small and is probably due to different attempt in both algorithms to recover kinks in the tracking system such as those occurring because of bremsstrahlung, K decays and V0s.

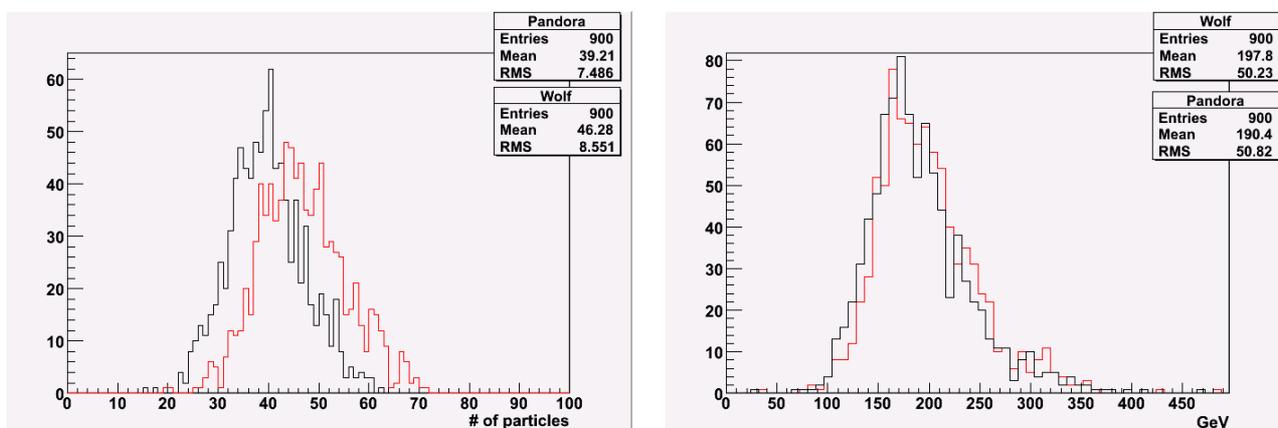


Figure 8 Distribution of the number of reconstructed charge particles and their total energy for Wolf (red) and Pandora (black).

## 6. Effect of invisible particles to the Higgs mass peak

In the previous section the analysis was carried out using only reconstructed particles from the particle flow algorithms; in this section the results of the same analysis will be presented including all ‘invisible’ particles. Invisible particles are defined as all those particles that are not reconstructed in the event: neutrinos, particles in beam pipe or at very low angles in the TPC, that are difficult to reconstruct by the software. It is possible to recover these particles using the true Monte Carlo (MC) information and calculate their contribution to the Higgs mass.

In order to define the detector acceptance, the cosine of the longitudinal angle  $\theta$  of all charged and neutral particles from the events is plotted. In the  $\theta$  region of the detector where the efficiency is very low the particles are added to the event using the true MC information. In figure 9 the distribution of  $\theta$  for all charged particles in LDC01Sc is presented. There is an unexpected peak at zero for events reconstructed using Wolf that is not present when using Pandora: the reason for this difference is not clear and needs further investigation. In figure 10 the same plot is presented for neutral particles. In this case it is more difficult to identify the correct value of  $\theta$  that defines the detector acceptance, since the distribution is not flat. To avoid overestimation of beam pipe particle contribution, a very high cut is applied. For both distributions, the end point is identical for the two particle flow. Figure 11 shows the end point of  $\theta$  distributions of neutral and charged particles for LDC00Sc detector. In table 2 are presented the values used for LDC00Sc and LDC01Sc.

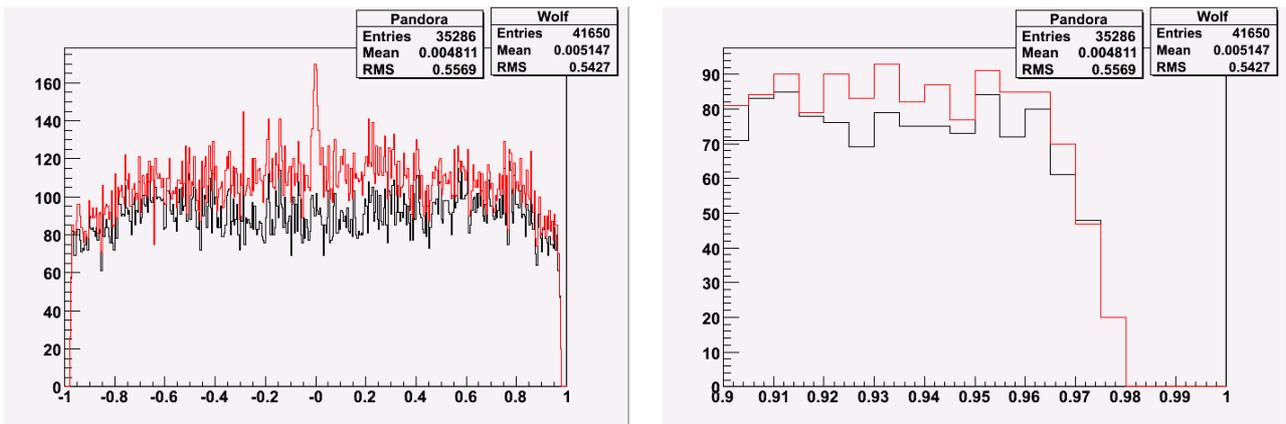


Figure 9 Distribution of charged particles in LDC01Sc detector after reconstruction using Wolf (red) and Pandora (black). On the right is shown a zoom of the same distribution.

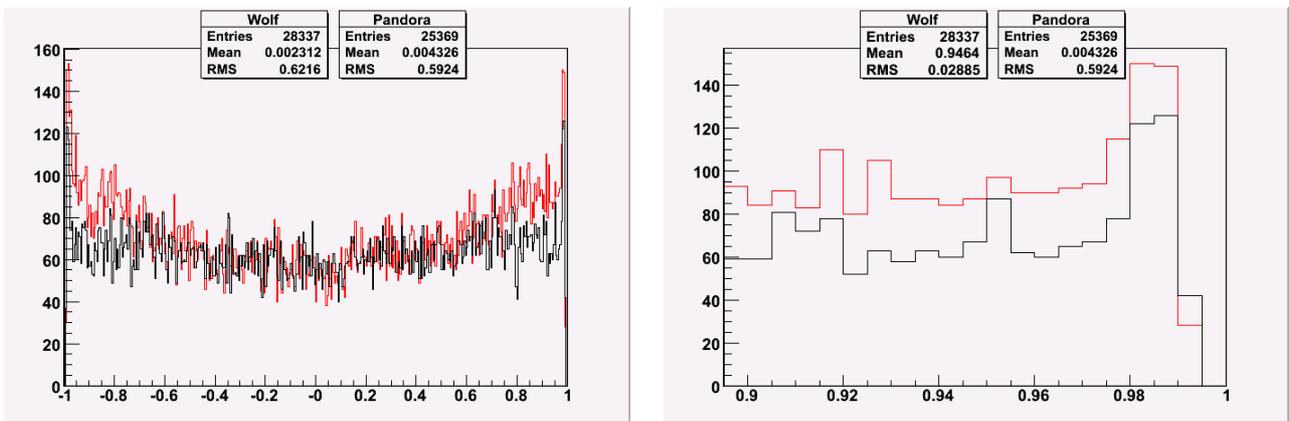


Figure 10 Distribution of neutral particles in LDC01Sc detector after reconstruction using Wolf (red) and Pandora (black). On the right is shown a zoom of the same distribution.

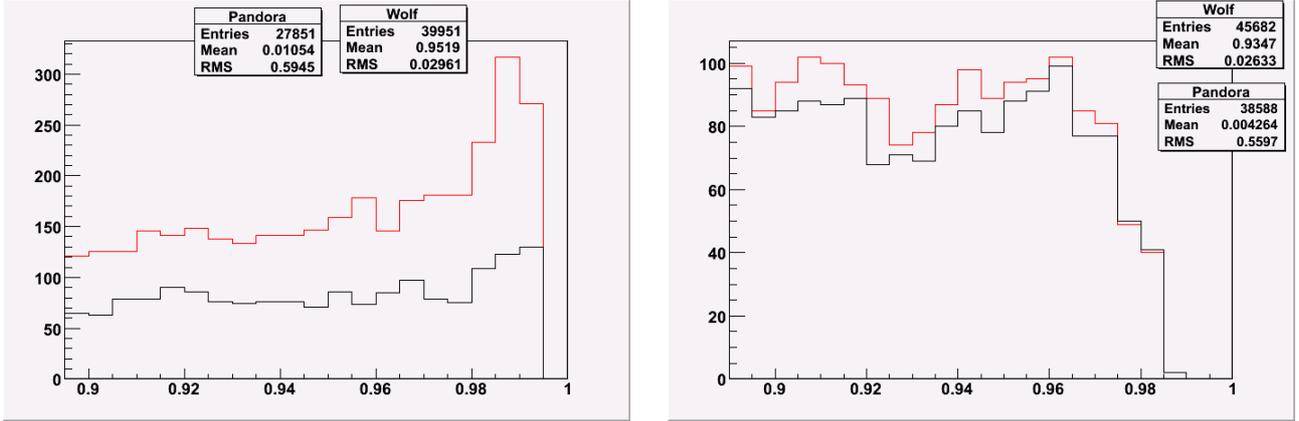


Figure 11 Detail of the distribution of neutral particles (left) and charged particles (right) in LDC00Sc after reconstruction using Wolf (red) and Pandora (black).

Table 2 Limits of reconstruction for charged and neutral particles in LDC00Sc and LDC01Sc.

Maximum angle of reconstruction for	Charged particles	Neutral particles
LDC00Sc	0.98	0.997
LDC01Sc	0.975	0.99

In figure 12 the Higgs masses are plotted for Pandora and Wolf with and without the contribution of invisible particles. The contribution of invisible particles to the mass distribution is 6 GeV, with the neutrinos alone contributing 3 GeV [11]. It is clear from the plot that, adding the information from ‘invisible’ particles, Pandora correctly reconstructs this variable, since the peak can be found at the correct place and the sigma of the distribution is smaller than the one obtained from Wolf. It can be concluded that the discrepancies seen in Pandora in the previous section in the Higgs mass distribution are due the presence of invisible particles.

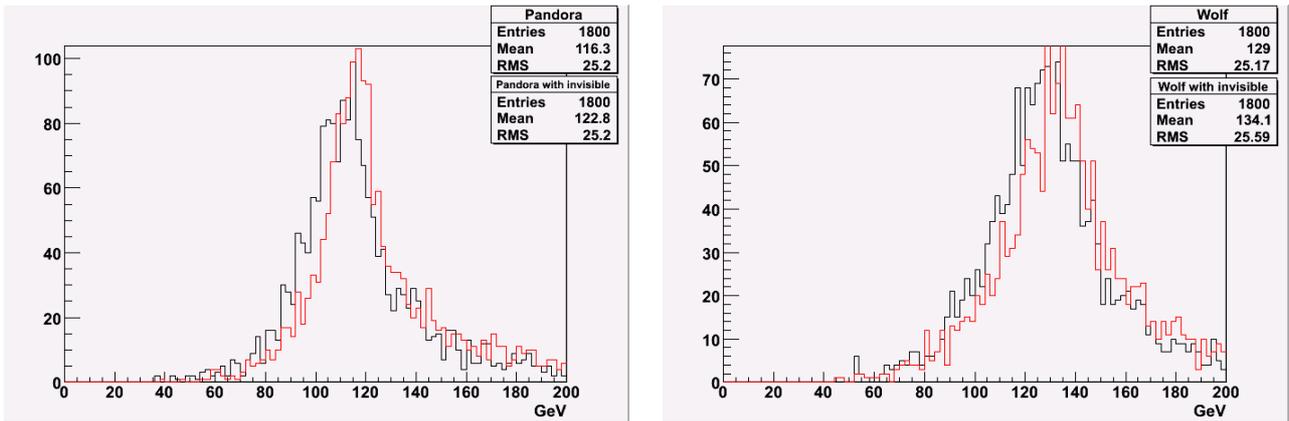


Figure 12 Mass distribution of the two Higgs in LDC01Sc for Pandora (left) and Wolf (right). In black the distribution using only reconstructed particles, in red adding the invisible particles.

The better results obtained with Pandora are visible when comparing the distributions for  $D$  and  $D^2$  presented in figure 13 and 14. The distribution of  $D^2$ , when considering the ‘invisible’ particles, peaks to zero for Pandora while it is moved far from zero for Wolf. The same improvement is visible in the distribution of  $D$ . For Pandora the peak is now well centred to zero.

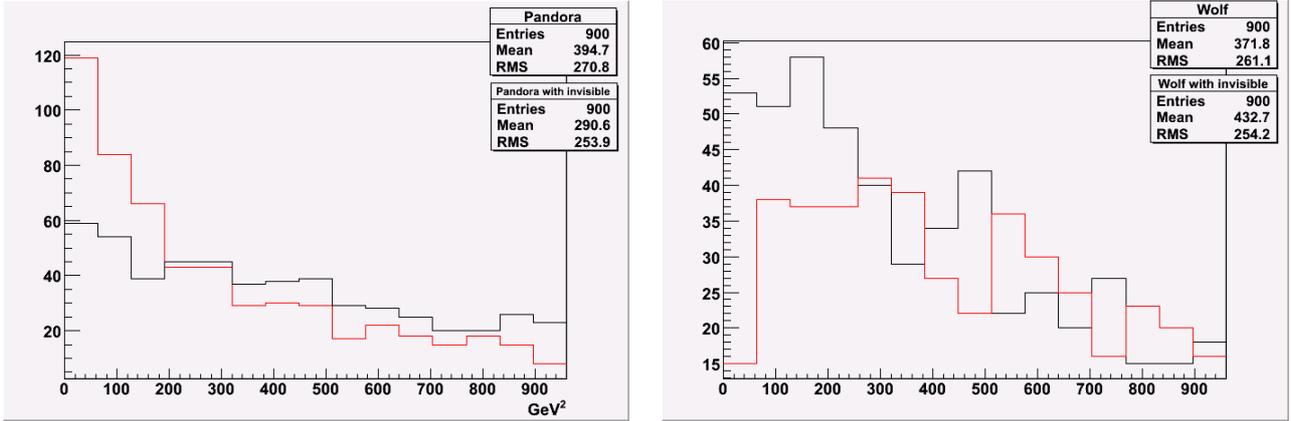


Figure 13 Distribution of  $D^2$  for Pandora (left) and Wolf (right) in LDC01Sc: in black the distribution using only reconstructed particles, in red adding the invisible particles.

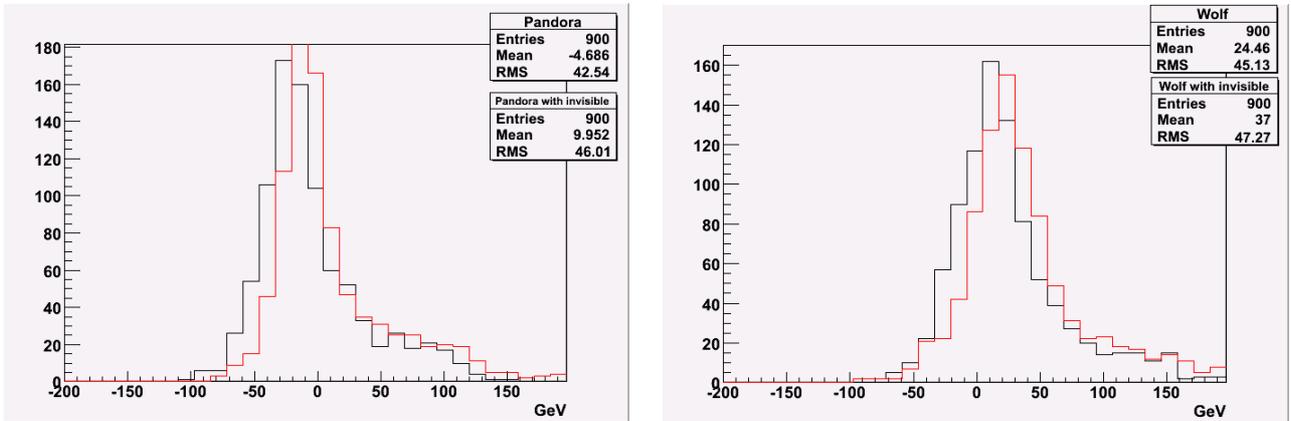


Figure 14 Distribution of  $D$  for Pandora (left) and Wolf (right) in LDC01Sc: in black the distribution using only reconstructed particles, in red adding the invisible particles.

Since the contribution of invisible particles is now known, for future analyses the value of the theoretical mass of the Higgs will be reduced by 6 GeV in all  $D$  and  $D^2$  distributions.

## 7. Comparison of PFA algorithms in the case $Z \rightarrow \mu^+ \mu^-$

To compare the analysis in this channel to the results obtained for electrons, a cut on the reconstructed  $Z$  mass is applied: only those events with a reconstructed  $Z$  mass between 80 and 100 GeV are compared. This should reduce the differences caused by the possible wrong selection of the two muons because of pion mis-identification. The indication from the previous section is used in that the generated value of the Higgs mass used to obtain the  $D$  and  $D^2$  distributions is taken to be 114 instead of 120 GeV. This value has not been changed in the minimization procedure to combine the jets in bosons.

In figure 15 the mass plots for Pandora and Wolf in LDC01Sc are presented with electrons rescaled to take into account a small difference in the size of the samples. The mean value for muons is lower than for electrons, probably due to the fact that the cut on  $Z$  mass is too loose to effectively solve the problem caused by pion mis-identification. In figure 16 and 17, Pandora and Wolf are compared examining the distribution of  $D$  and  $D^2$ . It's clear that, using 114 GeV instead of 120 GeV for the Higgs mass, Pandora reconstructs the Higgs mass better than Wolf. The muons in Pandora better peak at zero in the  $D^2$  plot even if electrons are better centred on zero in the  $D$  distribution. In Wolf the differences are smaller, the  $D^2$  plot has the same shape for both channel

while muons are closer to zero than electrons. However, these differences are small and may be attributed to inefficiencies in the particle flow such as the missing tag of the muons or the missing electron-photon association for bremsstrahlung in Pandora. Thus these differences will be not analysed until new software will be available.

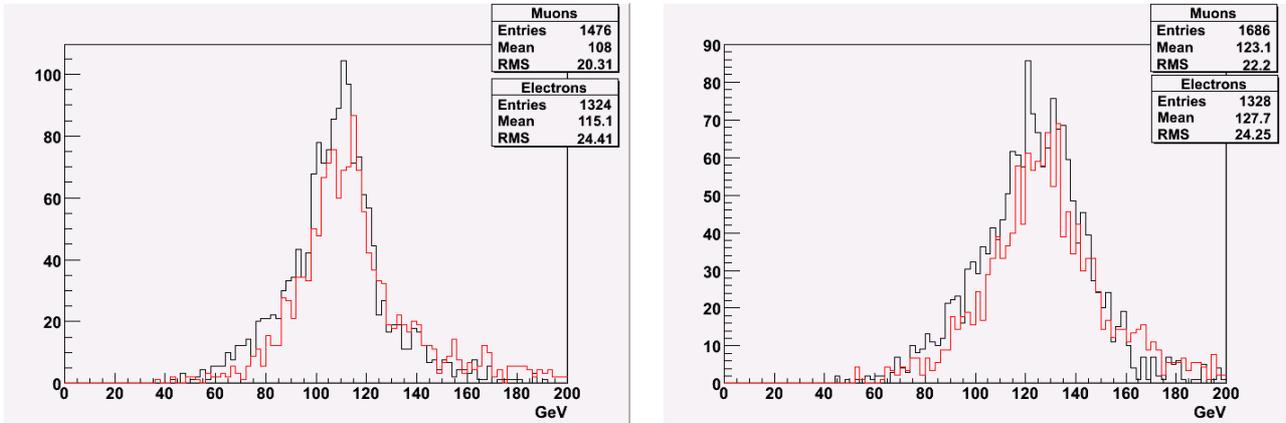


Figure 15 Mass distribution of the two Higgs in LDC01Sc for Pandora (left) and Wolf (right). In black the distribution using the muon channel, in red the electron channel.

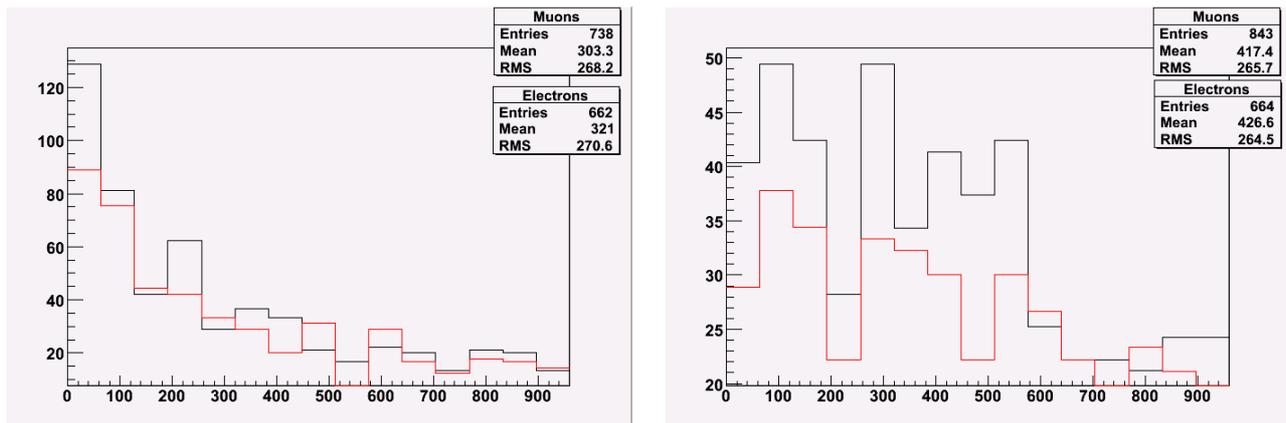


Figure 16 Distribution of  $D^2$  for Pandora (left) and Wolf (right) in LDC01Sc. In black the distribution using the muon channel, in red the electron channel.

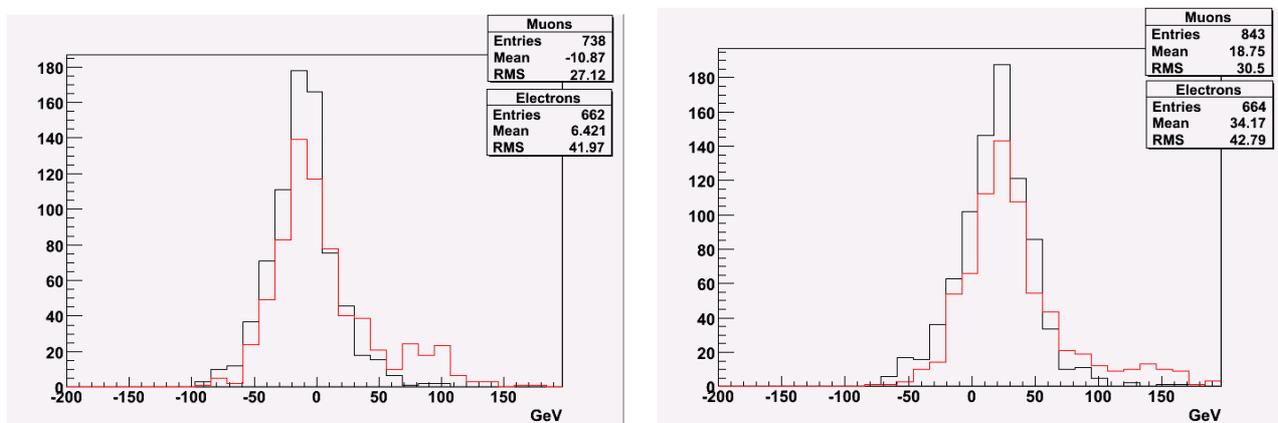


Figure 17 Distribution of  $D$  for Pandora (left) and Wolf (right) in LDC01Sc. In black the distribution using the muon channel, in red the electron channel.

## 8. Comparison between detector models

With the information obtained with the analysis presented so far it is possible to look at some differences between detector models using the same variables,  $D$  and  $D^2$ , used to compare the two PFAs. To perform this comparison Pandora has been used, with 114 GeV for the generated Higgs mass in the definition of  $D$  and  $D^2$ , as in the previous section. Only events with a reconstructed  $Z$  mass between 80 and 100 GeV are selected.

Figure 18 shows the reconstructed Higgs mass distributions for both the muon and electron decay channel of the  $Z$  boson in both LDC00Sc and LDC01Sc. In figure 19 and 20 the distributions of  $D$  and  $D^2$  are plotted for both channels.

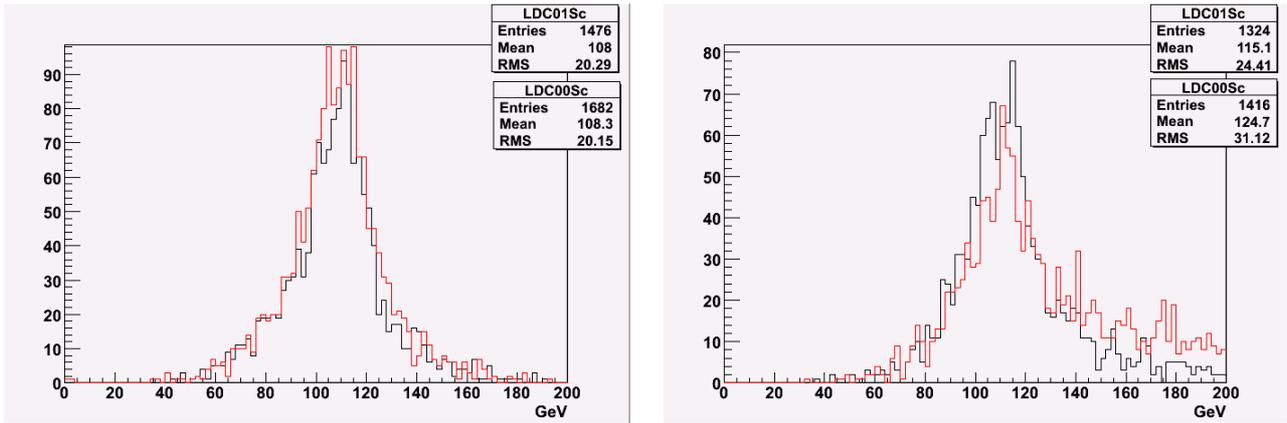


Figure 18 Higgs mass distribution for  $Z \rightarrow \mu\mu$  (left) and  $Z \rightarrow e^+e^-$  (right). In each plot the result for LDC00Sc (red) and LDC01Sc (black) is plotted.

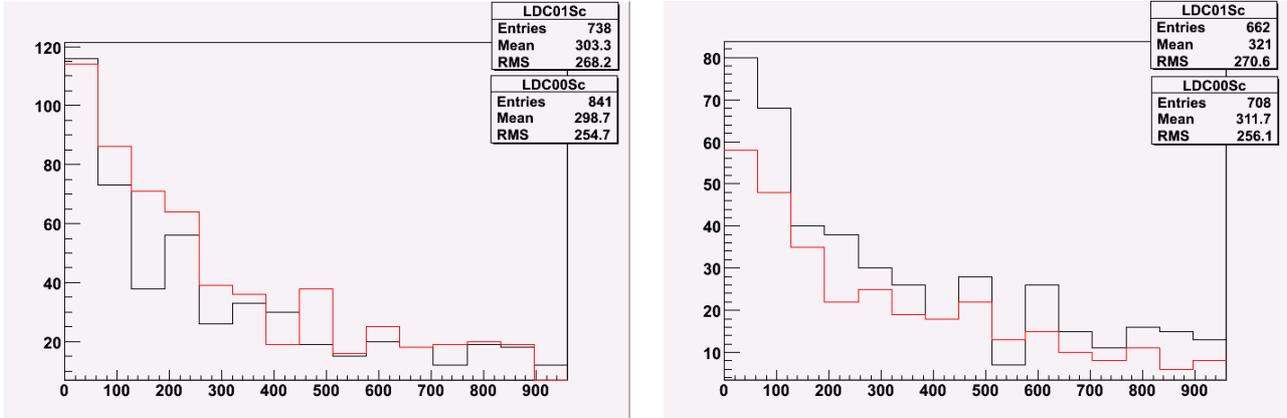


Figure 19 Distribution of  $D^2$  for  $Z \rightarrow \mu\mu$  (left) and  $Z \rightarrow e^+e^-$  (right). In each plot the result for LDC00Sc (red) and LDC01Sc (black) is plotted.

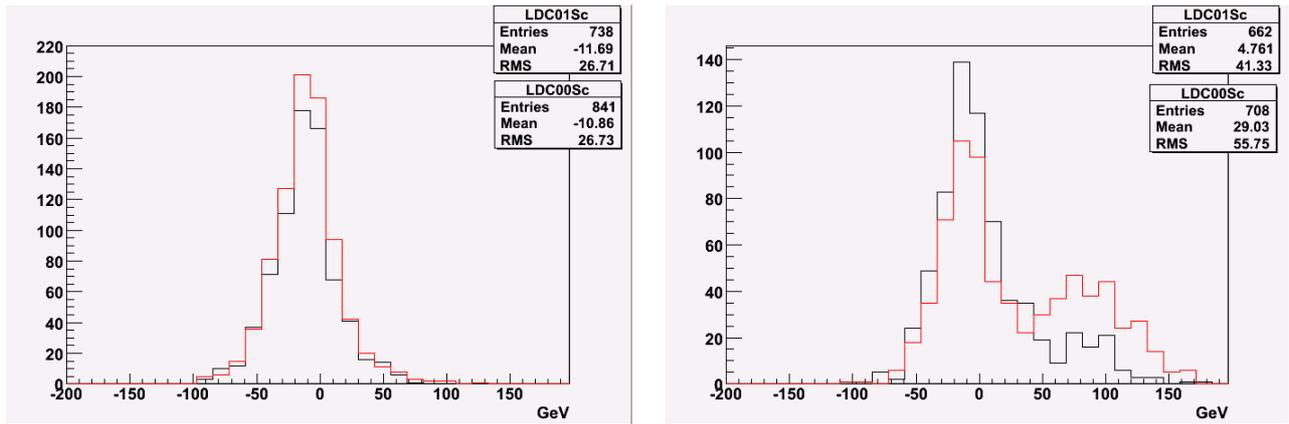


Figure 20 Distribution of  $D$  for  $Z \rightarrow \mu$ ons (left) and  $Z \rightarrow e$ lectrons (right). In each plot the result for LDC00Sc (red) and LDC01Sc (black) is plotted.

In the muon channel the differences are very small. In the electron channel some differences are present, possibly because bremsstrahlung in electrons is not taken into account in the particle flow algorithm. In fact, due to the extra material in LDC00Sc, more bremsstrahlung photons are emitted; if they are not correctly associated to the electrons, they are added to a jet causing the long high energy tail visible in the mass spectrum.

## 9. Conclusion

The analysis performed on the two available particle flow algorithms using the ZHH channel indicates that the Pandora algorithm better reconstructs the mass of the two Higgs particles. Before starting the analysis, the values of the calibration for the calorimeters digitization have been confirmed with a preliminary study using single particles. Small differences have been observed between Wolf and Pandora.

The comparison between PFAs has been presented for the cases in which the  $Z$  decays into electrons; detailed comparison for reconstructed neutral particles has shown that the biggest difference is in the photon reconstruction, with Pandora reconstructing photons with a smaller energy than Wolf. The effect of invisible particles, such as neutrinos and particles in the beam pipe, has been considered. The inclusion of those particles in the PFA, using the true information from MC, increases the Higgs mass by about 6 GeV for both PFA. After taking this contribution into account, Pandora gives a better reconstruction of the Higgs mass than Wolf.

The case of  $Z$  decays into muons has also been analyzed. Differences between muon and electron channels are present but small and are probably due to lack of muon identification and to bremsstrahlung.

In the second part of the analysis, Pandora has been used to compare the two different LDC models, LDC00Sc and LDC01Sc. The result indicates that the two detectors can reconstruct the signal in a very similar way.

## Bibliography

- [1] M. Battaglia et al. “Physics Benchmarks for the ILC Detector”, 2005 International Linear Collider Workshop, Stanford, USA.
- [2] M. Thomson, “Particle Flow Calorimetry at ILC”, arXiv:physics/0607261.
- [3] A. Raspereza, “Modular Implementation of Particle Flow Algorithm with Minimized Dependence on the Detector Geometry”, arXiv.org:physics/0601069.
- [4] [http://www-sldnt.slac.stanford.edu/nld/new/Docs/Generators/PANDORA\\_PYTHIA.htm](http://www-sldnt.slac.stanford.edu/nld/new/Docs/Generators/PANDORA_PYTHIA.htm)
- [5] <http://polywww.in2p3.fr:8081/MOKKA>
- [6] <http://ilcsoft.desy.de/marlin/>
- [7] See the tracking section on [http://ilcsoft.desy.de/MarlinReco/v00-02/manual\\_html/manual.html](http://ilcsoft.desy.de/MarlinReco/v00-02/manual_html/manual.html)
- [8] <http://polywww.in2p3.fr:8081/MOKKA/detector-models/ldc/ldc-models>
- [9] Predrag Krstonosic, private communication.
- [10] Stirling W J et al, 1991, J. Phys. G: Nucl. Part. Phys. 17 1567.
- [11] M. Fauci Giannelli, “Comparison between PFlow algorithms using ZHH channel”, presentation at Manchester Calice-UK Meeting.  
<http://www.hep.ph.ic.ac.uk/calice/generalMeetings/061103manchester/faucigiannelli.ppt>.