Influence of timing cuts in Testbeam data and Simulation

Dissertation

zur Erlangung des Doktorgrades

des Department Physik

der Universität Hamburg

vorgelegt von

ELDWAN BRIANNE

aus Saint-Malo, Frankreich

Hamburg

2017

Gutachter/in der Dissertation:	Prof. Dr. Erika Garutti
	Dr. Katja Krüger
Gutachter/in der Disputation:	Dr. Jenny List
	????
Datum der Disputation:	????
Vorsitzender des Prüfungsausschusses:	????
Vorsitzender des Promotionsausschusses:	????
Dekan des Fachbereichs Physik:	????

Abstract

Etiam pede massa, dapibus vitae, rhoncus in, placerat posuere, odio. Vestibulum luctus commodo lacus. Morbi lacus dui, tempor sed, euismod eget, condimentum at, tortor. Phasellus aliquet odio ac lacus tempor faucibus. Praesent sed sem. Praesent iaculis. Cras rhoncus tellus sed justo ullamcorper sagittis. Donec quis orci. Sed ut tortor quis tellus euismod tincidunt. Suspendisse congue nisl eu elit. Aliquam tortor diam, tempus id, tristique eget, sodales vel, nulla. Praesent tellus mi, condimentum sed, viverra at, consectetuer quis, lectus. In auctor vehicula orci. Sed pede sapien, euismod in, suscipit in, pharetra placerat, metus. Vivamus commodo dui non odio. Donec et felis.

Zusammenfassung

Donec et nisl id sapien blandit mattis. Aenean dictum odio sit amet risus. Morbi purus. Nulla a est sit amet purus venenatis iaculis. Vivamus viverra purus vel magna. Donec in justo sed odio malesuada dapibus. Nunc ultrices aliquam nunc. Vivamus facilisis pellentesque velit. Nulla nunc velit, vulputate dapibus, vulputate id, mattis ac, justo. Nam mattis elit dapibus purus. Quisque enim risus, congue non, elementum ut, mattis quis, sem. Quisque elit.

Contents

Introduction

Chapter 1

ILD detector simulation studies

Simulation of detector response is an essential part in high energy physics experiment. In Early stage of a project, simulations are done in order to explore and understand the possibilities of a detector design as well as its limitations. Simulation can be use as a way to determine requirements of an experiment to reach certain goals. During data-taking and afterwards, simulations are used model physics processes to compare the expected value from theory to a measured value for various processes.

In this chapter, software tools will be briefly introduced. The ILCSOFT framework used for this analysis will be described in **??**. The chain starts by the simulation of single kaons (K^0_L) interaction with the ILD detector model based on GEANT 4. Then simulated events undergoes the full chain reconstruction as explained in **??**. The procedure of the analysis (based on Marlin) and its conclusions will be presented in **??**. Finally, a benchmark of a fast simulation software (SGV) against the ILD full simulation will be described in **??**, particularly focussing on particle-flow performance aspects.

1.1 Simulation and software framework

1.1.1 ILCSOFT software framework

Various tools developed by the Linear Collider community is regrouped in a common software framework called ILCSOFT [1]. It provides a complete framework that can be used for Monte-Carlo studies and experiments. As an example, physics studies, ILD detector optimisation and performance for the ILC are performed under the ILCSOFT framework.

Most of the tools in the framework use an Event Data Model (EMD) named Linear Collider I/O (LCIO) which provides a reliable and performant solution for simulation and analysis studies [5]. With this tool, various detector concepts and analysis can be shared.

The ILCSOFT framework provides a modular C++ framework named MARLIN for reconstruction and analysis of physics events [4]. MARLIN uses LCIO seamlessly and is configured using XML steering files. MARLIN enables users to develop custom modules for their own and run it along other already existing modules.

The reconstruction and analysis tools used in this analysis are mostly part of ILCSOFT. For this thesis, ILCSOFT v01-17-11 was used for simulation, reconstruction and analysis.

1.1.2 ILD Detector Simulation

The following analysis is using one of the generic ILD detector model (ILD_o1_v05) as describe in **??** within the MOKKA framework. Many other models are also considered for ILD as shown in Table 1.1. MOKKA is a front-end to GEANT 4 and provides a realistic geometry of the ILD detector. The MOKKA version used is v08-05 and the GEANT 4 version is 10.01. The simulation is performed by simply using the particle gun provided in GEANT 4 to shoot particles (π^-or K^0_L) in different regions of the detector by randomly variating the angles θ and ϕ of the gun. To model hadronic showers, the QGSP_BERT physics list was

Table 1.1 – Considered ILD detector options.

Option	ECAL Technology	HCAL Technology
ILD_o1_v05	SiW-ECAL	AHCAL
ILD_o2_v05	SiW-ECAL	SDHCAL
ILD_o3_v05	Sc-ECAL	AHCAL

used. The output of the simulation provides a lcio file containing collections of the tracking hits and simulated calorimeter hits. This file is then reconstructed within MARLIN.

1.2 Reconstruction chain

The reconstruction is done on simulated data in order to implement detector effects. For example, calorimeter hits need to be digitized by implementing threshold and readout effects.

1.2.1 Tracking

The tracking reconstruction is performed on each individual tracking detector. Track segments are identified by pattern recognition algorithms.

Track fitting is performed using the track segments with an inversed Kalman filter to identify trajectories of charged particles. Each tracks contains origin, direction, charge and momentum of the particle [3].

1.2.2 Calorimeter digitization

The calorimeter digitisation is performed on simulated calorimeter hits as part of ILD-CaloDigi processor [2]. It takes account for threshold effect from the electronics, sampling fraction of the calorimeter and the readout technology used. In the considered model of ILD, the SiW-ECAL and AHCAL are used.

In both cases, it uses a silicon-pixel based technology. The digitisation then takes into account the finite number of pixels that can be fired as well as the statistical fluctuations related to pixel readout [6].

Concerning time, it uses a simple digitisation. For a simulated hit, all contributions are looped over and only adds contributions under a certain timing cut (default is 100 ns). This modelisation of timing is very simplified as in reality the electronics are shaping the signal with a certain shapping time and register the time of the first contribution over the threshold (default is 0.5 MIP) **??**.

1.2.3 Pandora PFA

PandoraPFA [7] is the Particle Flow algorithm used for Linear Colliders as explained in **??**. It uses as input tracks and calorimeter hits to form Particle Flow Objects (PFO). It uses a complex multi-stage process but basically, calorimeter hits are clustered and associated to tracks (if any) then the energy of a cluster can be corrected to improve the energy resolution. If the right criterium are matched, it forms a PFO which contains information about the reconstructed objects.

1.3 Influence of time cuts on hadronic showers

In this section, a study of timing cuts on hadronic shower is performed. The goal of this study is to assess the influence of timing cuts on the properties of hadronics showers as for example the width of the shower as well as the needed time resolution. The study will be divided in 2 parts, the first part assuming a perfect time resolution and the second part assuming time resolution for different cases.

1.3.1 Modification of timing window in ILDCaloDigi

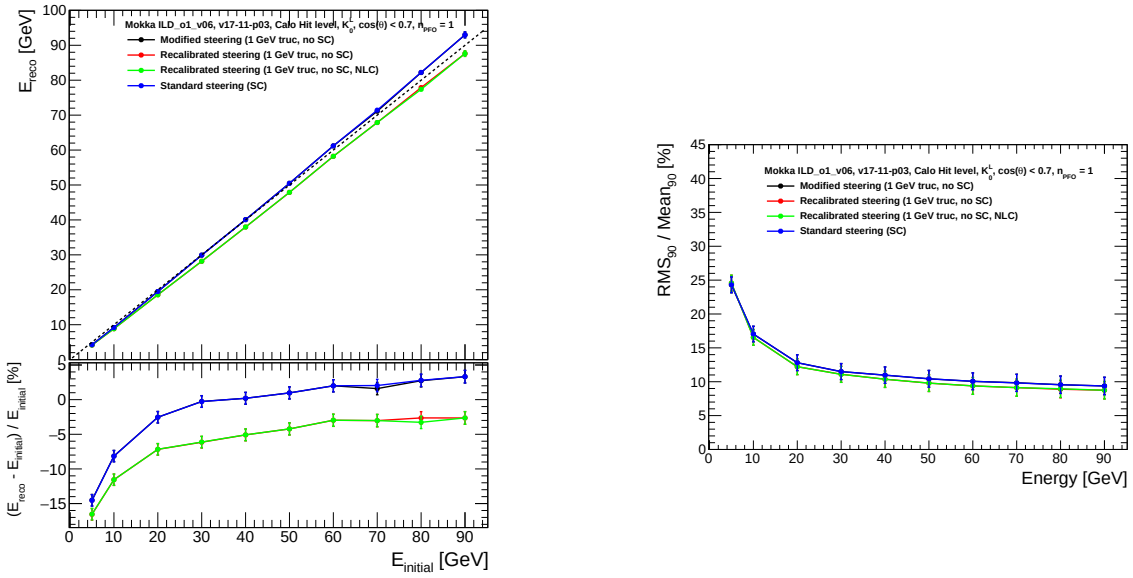
Timing of hits is registered in a very simplyfied way as explained in **??**. The modification of the time window (ranging from 1 ns to 100 ns) is performed during the reconstruction for different simulated K_L^0energies (ranging from 5 to 90 GeV).

1.3.2 Effects of calibration constants and Pandora constants

Before studying the effect of timing on hadrons showers, a check was performed on the initial provided calibration constants of the ILD detector. Several constants are used for the digitisation and reconstruction (GeV to MIP, sampling, Pandora EM/Had constants...) in order to get the correct reconstructed energy. The plots below are selecting events with only one PFO and a cos θ cut on the reconstructed particle of 0.7.

The figures 1.1a and 1.1b show the linearity and resolution curves for different sets of calibration constants at the cluster hit level i.e. looking at all the hits in a PFO cluster. Thus this enables to understand the effects of the digitisation constants in ILDDigiCalo though small clustering effects are present.

One can see that the blue and black curves are very similar due to the fact that no constants were changed in theses sets. Moreover the linearity is not perfect and varies between



(a) Cluster hit linearity curve

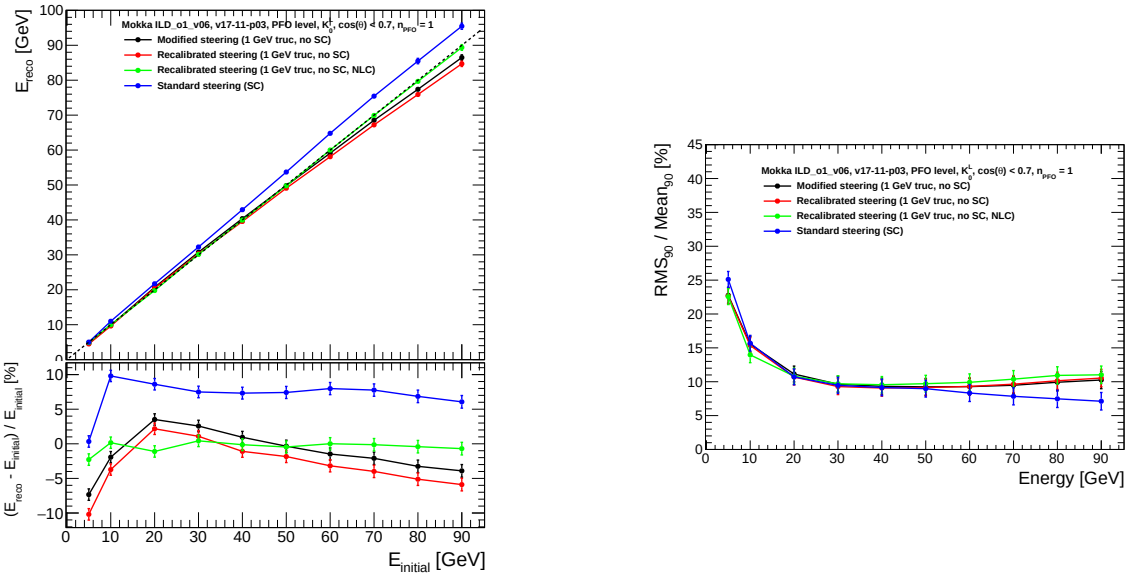
(b) Cluster hit resolution curve

Figure 1.1 – a) The top plot shows the mean reconstructed energy E_{reco} for 5 to 90 GeV K^0_L function of the simulated energy $E_{initial}$ for different constant parameters used in the reconstruction at the cluster hit level. The bottom plot shows the relative difference of the different curves to the line $x = y$. b) The plot shows the relative resolution $RMS_{90}/Mean_{90}$ for different constant parameters used in the reconstruction function of the energy. The blue curve uses the standard calibration, the black curve uses a modified set of parameters using energy truncation and no SC, the red curve uses constant parameters after recalibration and the green curve uses the same parameters as the red curve with non-linearity correction. The error bars represent statistical uncertainties.

-15% and 5% also the curve crosses the line $x = y$ which if corrected would degrade the energy resolution. The green and red lines are similar as they have the same constants (the non-linearity correction is only applied to PFOs). The linearity fluctuates between -15% and -5% but does not cross the line $x = y$.

Concerning the energy resolution all the curves are very similar and are as expected. The green and red curves are slightly better due to the improvement of the calibration constants.

Another option is to look at the PFO level as shown in figures 1.2a and 1.2b. This permits to understand the effects of the calibration constants in PandoraPFA.



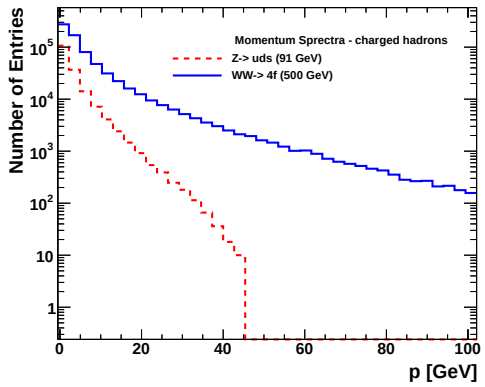
(a) PFO linearity curve

(b) PFO resolution curve

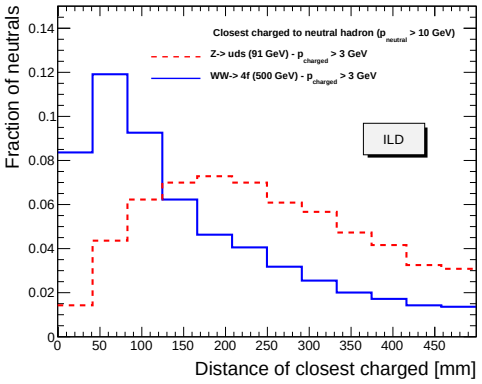
Figure 1.2 – a) The top plot shows the mean reconstructed energy E_{reco} for 5 to 90 GeV K^0_L function of the simulated energy $E_{initial}$ for different constant parameters used in the reconstruction at the PFO level. The bottom plot shows the relative difference of the different curves to the line $x = y$. b) The plot shows the relative resolution $RMS_{90}/Mean_{90}$ for different constant parameters used in the reconstruction function of the energy. The blue curve uses the standard calibration, the black curve uses a modified set of parameters using energy truncation and no SC, the red curve uses constant parameters after recalibration and the green curve uses the same parameters as the red curve with non-linearity correction. The error bars represent statistical uncertainties.

The plots show a different picture. For the linearity curve, the red and black line are quite similar and show a non-lineariy especially at high energies between -10% and 2%. The green line is nicely linear thanks to the non-linearity correction. Then the blue line is off by around 10%, this is believed to be due to the weights of the software conpensation that are not optimal for this model.

For the resolution curves, one can see a rise of the resolution at high energies for the red, green and black lines certainly due to the non-linearity. The blue curve present a bump after 50 GeV changing suddently the slope of the curve due to the over-correction of the energy.



(a) Momentum spectrum of charged hadrons.



(b) Minimal distance between a charged and neutral particle at the front face of the ECAL.

Figure 1.3 – a) Momentum distribution for charged particles in simulated $e^+e^- \to Z/\gamma \to q\bar{q}$ with $q = u,d,s$ at $\sqrt{s}$ = 91 GeV and $e^+e^- \to W^+W^- \to q\bar{q}q\bar{q}$ where q is a quark at $\sqrt{s}$ = 500 GeV . b) Distribution of distances to the closest charged track for neutral particles produced in $Z/\gamma \to q\bar{q}$ and $W^+W^- \to q\bar{q}q\bar{q}$ processes measured at the front face of the electromagnetic calorimeter in the ILD detector.

Through the linearity is not perfect over all energies, the most regarded observable is the jet energy resolution. As explained in **??**, jets are mostly composed of charged particles of around 60%. In this case, the energy of the PFO is coming from the track. Moreover neutral particles are counting in general for around 30% of the contribution in a jet.

As shown in figure 1.3a, for jets representative of heavy boson decay near production threshold, the momentum spectrum is dominant at around 10 GeV as for heavy boosted jets with a more complex event topology, the momentum distribution is still dominant to low energies but present a tail to much higher energies. Thus the non-linearity has only little effect there. It is still relevant to understand the different effects of the reconstruction at single particle level.

A complementary study was to look at the minimal distance between a charged and neutral particle for theses different physics processes. The figure 1.3b shows that for low energy jets the mean minimal distance (measured at the front face of the SiW-ECAL) between a charged and neutral hadron is around 180 mm thus in this context, showers are well

separated. But at higher energies where density is higher, typical distances of 50 mm need to be resolved. This situation can become relevant in the contribution of confusion to the jet energy resolution. In this case, the use of timing information could help to separate nearby showers and improve the pattern recognition.

1.3.3 Timing cut effects on hadronic showers in ILD detector

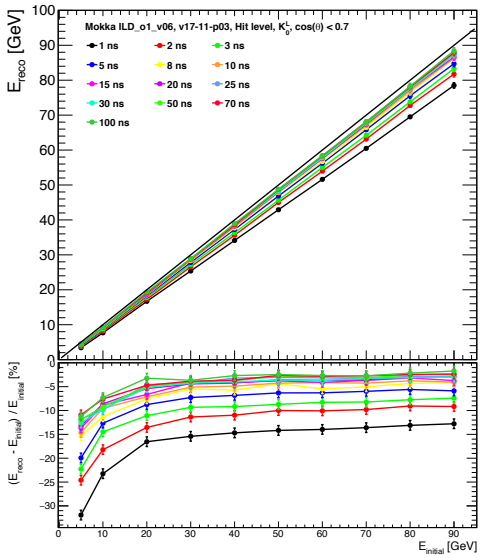
In this section, the effect of timing cuts on hadronic showers is investigated. The study was performed using the ILCSOFT framework for reconstruction and a personal MARLIN processor for analysis. In order to study the effect of timing on hadronic shower properties, the initial study was performed assuming a perfect timing resolution (i.e. the timing information is the Monte-Carlo truth). In a following step, several timing resolution were used to assess different scenarios. The smearing of the time was done by randomly sampling a normalised gaussian centered in 0 ns with a timing resolution denoted σ_t = 0.4, 1 and 8 ns.

The selection of events is fairly simple, only events in the barrel region ($cos\theta < 0.7$) are selected. All the calorimeter hits in the ECAL and HCAL are used.

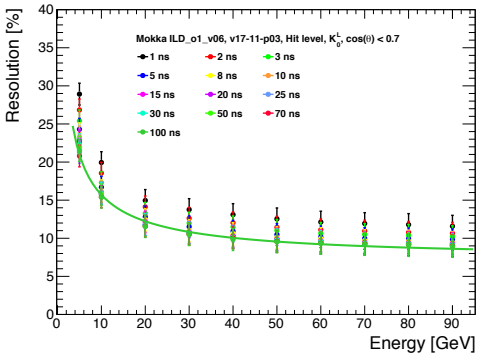
1.3.3.1 On Monte-Carlo level

The following section present results of timing cuts assuming a perfect time resolution. To avoid any effects of clustering and Pandora, the study was performed at the calorimeter hit level. Several shower observables were looked at as a function of the time cut for energies from 5 GeV to 90 GeV K^0_L. The different time cuts used are: 1, 2, 3, 5, 8, 10, 15, 20, 25, 30, 50, 70 and 100 ns.

The figures 1.4a and 1.4b show the effect of timing cut on linearity and energy resolution. The tighter the timing cut gets, the linearity slope decreases and resolution gets degraded. This effectively means that with a harder timing cut, more hits of the shower are removed but mostly only outer hits carrying only a small fraction of the total shower energy, the core



(a) Linearity curve with no time smearing.



(b) Resolution curve with no time smearing.

Figure 1.4 – a) The top plot represent the linearity curve in the ILD detector over a range of energy from 5 GeV to 90 GeV for different timing cuts assuming a perfect resolution. The bottom plot represent the relative deviation to the line $x = y$ for the different time cuts. b) The plot illustrates the relative energy resolution ($\frac{\sigma_E}{E}$) at single particle level for different timing cuts. The green line is a fit performed at 100 ns of the form $\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \bigoplus b$ where a is the stochastic term ($44.01\% \pm 3.17$) and b the constant term (7.26 ± 0.84).

of the shower mostly does not get affected by timing cut up to few nanoseconds.

The figure 1.5 shows the relative impact on the energy resolution compared to the 100 ns cut as function of timing cuts for all energies. The energy resolution is mostly not affected over a cut of around 20 ns meaning that the removed hits are not carrying a lot of energy and are part of the shower halo. Then below 20 ns, the resolution starts to degrade slowly relatively in the same way for all energies. A hard cut of 1 ns will degrade greatly the energy resolution up to around 30%.

The figure **??** shows the radial profile of a 50 GeV hadronic shower. The radial profile of the shower is filled for each hit with the distance to the main shower axis (R_i, eq.1.1)

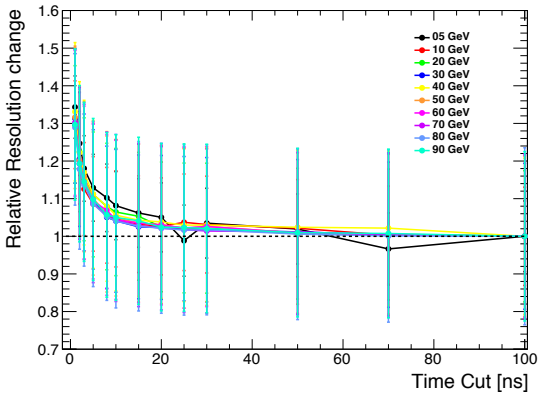


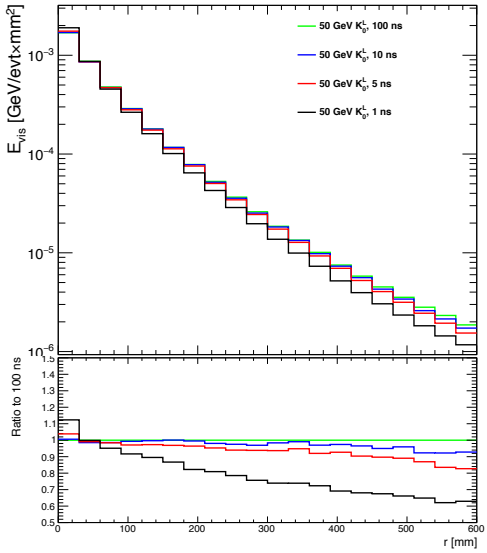
Figure 1.5 – Relative change of the energy resolution compared to 100 ns as function of the timing cut. The error bars represent the statistical uncertainty.

weighted by the hit energy (E_i).

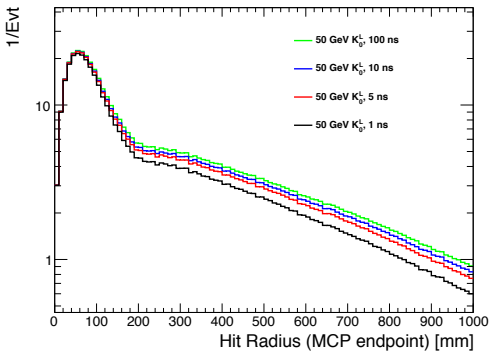
$$R_i = \sqrt{\Sigma_i (r_i - r_{cog})^2 - \|(\mathbf{r_i} - \mathbf{r_{cog}}) \cdot \mathbf{Eigen}\|^2)} \quad \text{with} \quad \mathbf{r_i} = (x_i, y_i, z_i) \quad \text{and} \quad \mathbf{r_{CoG}} = (cog_x, cog_y, cog_z) \tag{1.1}$$

The main part of the energy density is situated in the core within few centimeters. The influence of timing cuts is highly visible in the tail of the distribution (or halo of the shower) and as very little influence on the energy density deposited in core of the shower. Though an effect of increase of energy density in the two first bins of the distribution is visible. This effect is related to a displacement of the center of gravity (CoG) as function of the timing cut as outer hits of the shower are removed. This has been checked by looking at the hit radius distribution relative to a fixed reference instead of the CoG (the Monte-Carlo particle endpoint) as shown in figure 1.6b. One can observe that in this case, the timing cut removes only part of the tail and does not affect the core of the distribution.

Another aspect looked at was the influence of timing cut on the shower width <R>



(a) Radial profile.



(b) Hit radius.

Figure 1.6 – a) The top plot shows the radial profile of a 50 GeV hadronic shower overlayed for different timing cuts. The bottom plot shows the ratio of the histograms to 100 ns radial profile. b) Hit radius histograms at 50 GeV for different timing cuts.

(eq.1.2) defined as:

$$< R >= \frac{\Sigma_i E_i r_i}{\Sigma_i E_i} \tag{1.2}$$

The figures 1.7a and 1.7b show the shower width <R> for different particle energies as function of the timing cut. It shows that a tight timing cut at1 ns can reduce the shower width up to around 70%. One can observe also that the shower width at 5 and 10 GeV are behaving differently than higher than 10 GeV . This may come from the transition from the Bertini model (BERT) to the quark string gluon model (QGSP) in the physics list in this energy range. Looking at the shower width in absolute value, hadronic showers are wider at lower energies (∼180 mm for 5 GeV) certainly due to the magnetic field. Applying a timing cut removes more and more hits from the halo of the shower, thus reducing its size

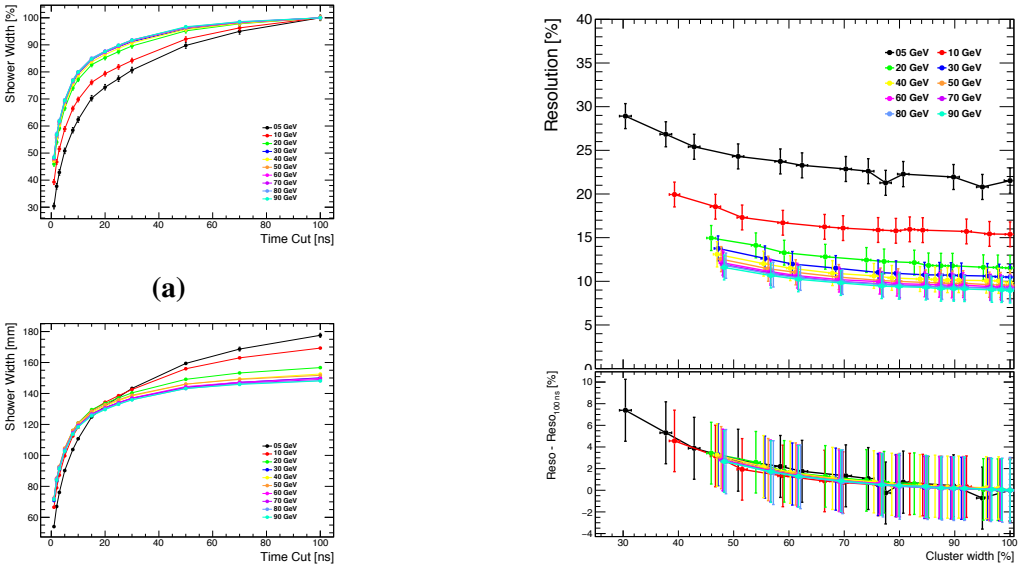


Figure 1.7 – a) The plot represents the mean of the shower width <R> as function of timing cut for different particle energies. The y-axis has been normalized to the shower width at 100 ns. The shower width decreases steadily as function of the timing cut, indicating that the shower gets narrower. b) The plots represents the absolute value of the mean shower width <R> in mm as function of the timing cut. This shows that the low energy showers are generally wider certainly due to the magnetic field. And under 20 ns, the width is very similar indicating the core of the shower is fairly similar for all energies. c) The top plot is the energy resolution as function of shower width for different particle energies. Each point represents a timing cut from 1 ns ns to 100 ns from left to right. The bottom plot is the loss of resolution compared to the gain in shower width size.

up to a point where it reaches the core of the shower at ∼20 ns timing cut where the shower is around a couple of tiles in size. At this point, the shower width behaves similarly for all energies by reducing gradually the size of the core.

It is interresting to look at the gain in the reduction of the shower width compared to the loss in energy resolution. In fact, reducing the shower width would improve the pattern recognition in Pandora. The figure 1.7c shows the resolution loss as function of the shower width. The bottom plots shows the gain in shower width is behaving in the same way for all energies. The tigher the timing cut gets, the small the shower gets as well as a loss in resolution. The main point here is that the gain in shower width is great (up to 60-70%)

Table 1.2 – Time resolution used for smearing.

Scenario	Time resolution (ns)
Testbeam	8
Ideal	1
ILC extrapolated	0.4

compared to the loss in energy resolution (up to 8%) that could be recovered in a next step after pattern recognition.

This study shows that the use of timing cuts give a great advantage in order to improve pattern recognition without degrading too much the energy resolution of a hadronic shower. However, this study assumes a perfect timing resolution which doesnt́ reflect the reality. In the next section, different time resolution were assumed based on the current knowledge on the timing resolution of the foreseen electronics.

1.3.3.2 In a realistic scenario

In this section, a similar study is performed as in section **??**. Instead it assumes realistic time resolutions based on the current electronic technology. The table 1.2 sums up the investigated time resolutions. The same selection is applied as in the previous section.

The testbeam resolution is the time resolution obtained with the current AHCAL technological prototype as explained in section **??**. The ideal time resolution is in the order of the time scale of the development of hadronic showers. And finally, the ILC extrapolated is obtained by assuming a linear extrapolation from the tesbeam time resolution with a faster slow clock of 5 MHz instead of 250 kHz (x20 faster) as explained in **??**.

Looking at the impact on linearity and energy resolution, time resolution in the order of sub-nanosecond does not affect much the linearity and resolution as shown in figures 1.8a and 1.9a. For a time resolution in the nanosecond order, the figures 1.8b and 1.9b show that a tight timing cut of 1-2 ns will start to affect linearity and resolution. However, for timing

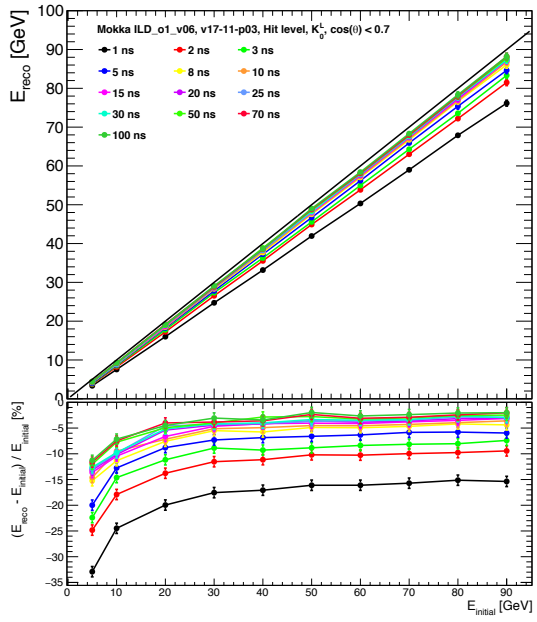
resolution higher, the linearity and resolution, as shown in figures 1.8c and 1.9c, start to get heavily degraded for timing cuts below 10-20 ns. This tells that at least a time resolution of around 1 ns is needed in order to use time information without impacting calorimeter performance too much.

Looking back again at the resolution as function of the shower width, one can notice that in the case of (sub-)nanosecond scale time resolution, the loss of resolution is minimal compared to the shower width. A timing cut of around 1-2 ns would only degrade the energy resolution of around 8-10% but would decrease the shower width by 60% as shown in figures 1.10a and 1.10b. On the other side, a time resolution of 8 ns would yield only a reduction of 40% of the shower width for the same loss in resolution corresponding to a timing cut of 10 ns as shown in figure 1.10c.

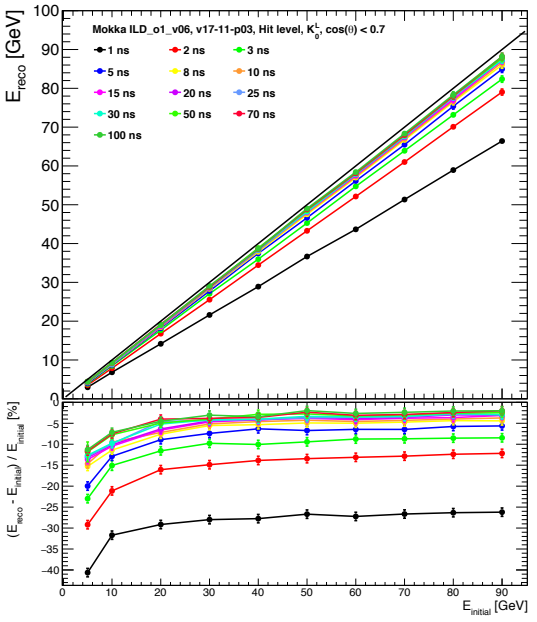
1.3.3.3 Conclusion

The study presented in this section demonstrates that timing information can be used in hadronic showers in order to improve pattern recognition and at last the jet energy resolution by reducing the confusion term. This study was performed assuming different time resolutions and shows that in order to benefit from timing information, a time resolution for the electronics in the (sub-)nanosecond scale would be ideal. Assuming a time resolution of 1 ns, a time cut around 1-2 ns would greatly reduce the shower width by about 60% and would degrade the energy resolution by about 8% as well as the linearity by decreasing the slope which could be corrected.

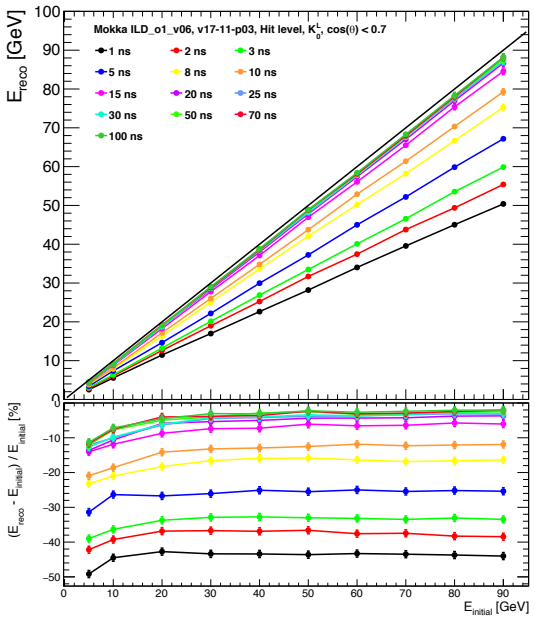
Higher time resolution would certainly help but would impact too a certain level the calorimeter energy resolution. One would need to think how to implement the use of time information in Pandora in order to improve the pattern recongnition efficiently.



(a) 0.4 ns time smearing.

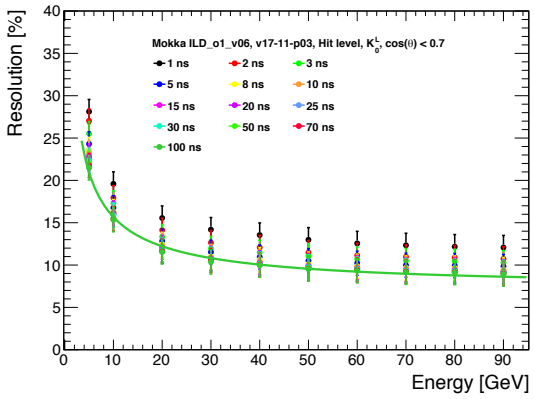


(b) 1 ns time smearing.

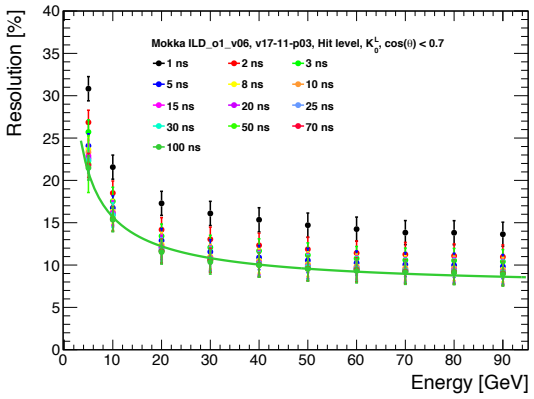


(c) 8 ns time smearing.

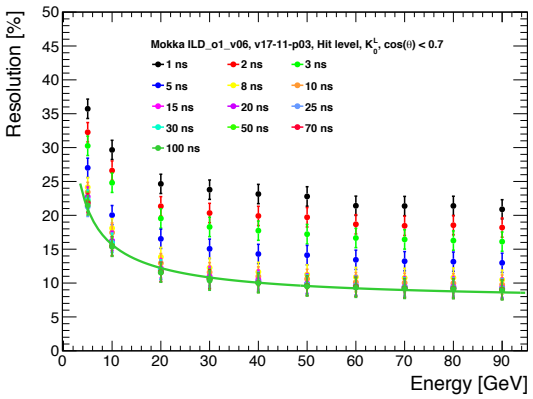
Figure 1.8 – Linearity curves for different assumed time resolutions (0.4 to 8 ns from left to right). The top plot represents the mean reconstructed energy E_{reco} for kaons from 5 to 90 GeV . The bottom plot shows the relative deviation to the line $x = y$.



(a) 0.4 ns time smearing.

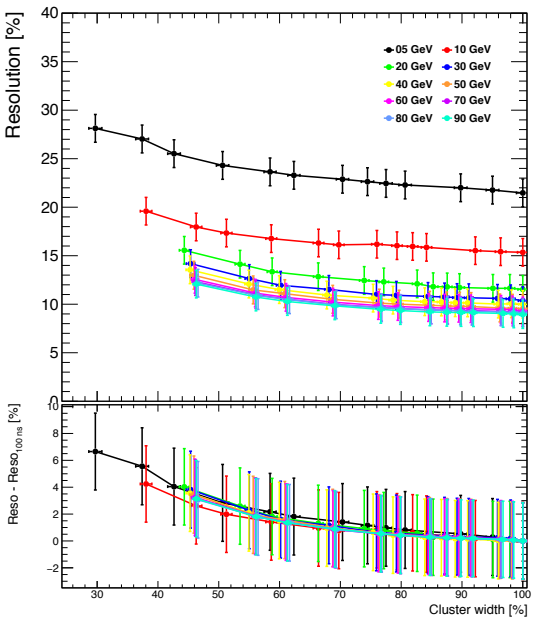


(b) 1 ns time smearing.

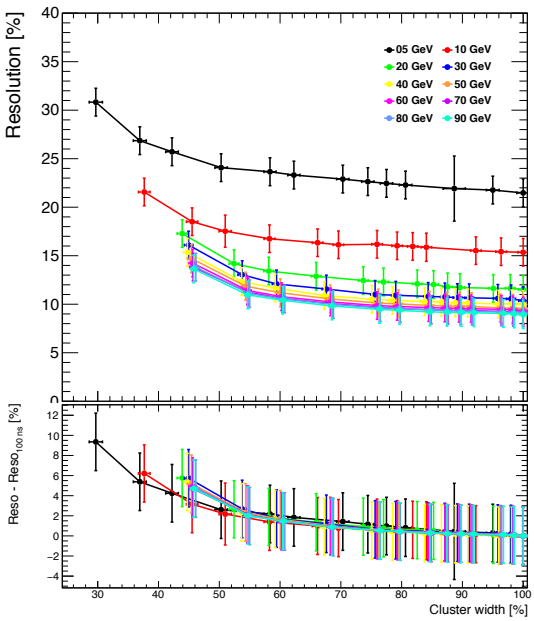


(c) 8 ns time smearing.

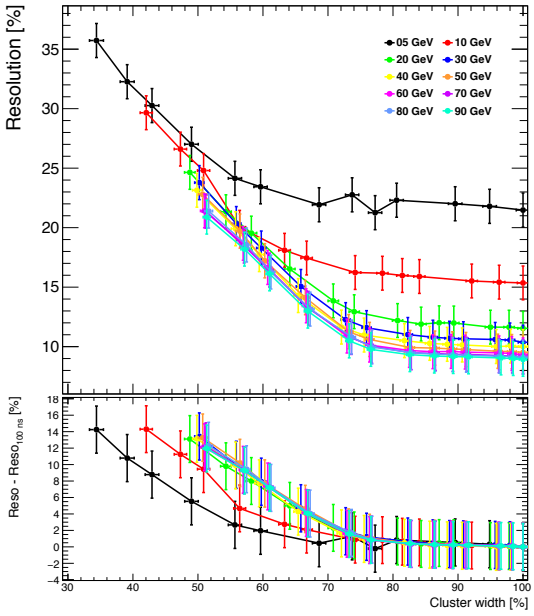
Figure 1.9 – Energy resolution curves for different assumed time resolutions (0.4 to 8 ns from left to right). The plot represents the relative energy resolution $\frac{\sigma_E}{E}$ for kaons from 5 to 90 GeV for each timing cut. The green line is a fit applied to 100 ns timing cut of the typical form $\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \bigoplus b$.



(a) 0.4 ns time smearing.



(b) 1 ns time smearing.



(c) 8 ns time smearing.

Figure 1.10 – Energy resolution as function of the shower width for different assumed time resolutions (0.4 to 8 ns from left to right). The top plot shows the relative energy resolution $\frac{\sigma_E}{E}$ for kaons from 5 to 90 GeV where each point is representing a timing cut as function of the shower width. The bottom plot shows the deviation to the nominal resolution at 100 ns as a function of the shower width.

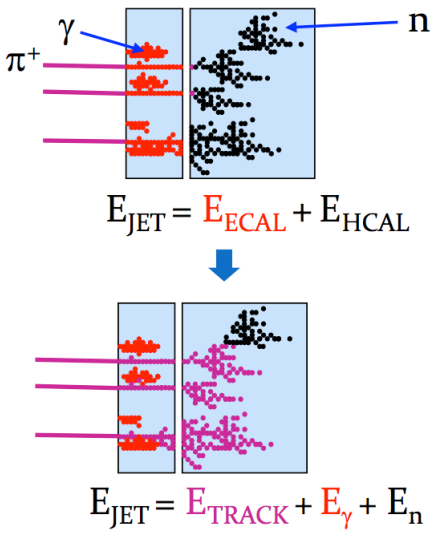


Figure 1.11 – Particle Flow Concept.

1.4 Benchmarking of fast simulation

First of all, what is Particle Flow? Particle Flow is a new approach to calorimetry in order to achieve a jet energy resolution much better than traditional calorimetry approaches (order of twice better). It has a potential to revolutionise particle detector design for future lepton collider experiment. Particle Flow has the ability of reconstructing the energy of all the individual contributions inside a jet [**?**].

Concerning simulation, Particle Flow has been implemented known as Pandora PFA in full ILD simulation. One further approach is the use of fast simulation. Why using fast simulation? In simple words, it is much faster (for example, a ttbar event takes few seconds compared to minutes in full simulation). Like that several studies could be done much faster than when using full simulation, and variation of observables much easier while keeping a precision close to the full simulation.

Currently, several fast detector simulation exist of different types or methods. SGV (Simulation a Grande Vitesse) is one of them which is developed by Mikael Berggren, it is a

fast detector simulation program using covariance matrix calculations. The status of SGV is that it performs already very well compared to full simulation and also integrates a Particle Flow Parametrisation in order to simulate Pandora PFA [**?**].

My studies will focus on the performance of SGV compared to full simulation and also the Particle Flow parametrisation performance in SGV. The motivation is to show that SGV perform as well as the full simulation (benchmark of SGV) and to improve the current Particle Flow parametrisation.

1.5 Particle Flow in SGV

Fast detector simulation exists of different types and different methods. For any of them, the goal is to be able to simulate the detector response with a time as the same order as to generate an event (order of 10 ms). For this, a four smearing vector can be done assuming global detector properties. More elaborate, a parametric simulation (like SIMDET) can be done, where the response depends on the energy of the particle and its angle. And finally, covariance matrix calculations can be done using the detector layout and the generated particles. SGV is found in this category [**?**].

1.5.1 Tracking in SGV

A track is calculated by the intersection of the helix of a particle with pseudo-layers in the tracker. From the outermost hit, an helix is parametrised and then propagated to the next inner layer and the intersection and covariance matrix are calculated, the propagation is done until the inner most layer is reached.

In SGV, there is no definition of hits. The helix is followed through the detector to find what pseudo-layers (tracker material is described by pseudo-layers) are hit by the par-

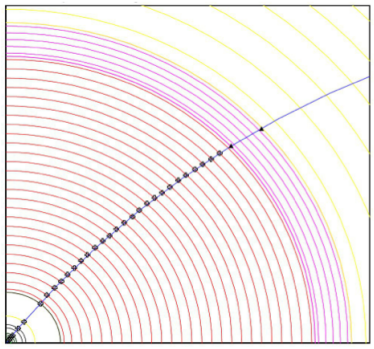


Figure 1.12 – Tracking in SGV. The tracker is represented by pseudo-layers at which intersection with a track, the covariance matrix is calculated, going from the outer part of the tracker to the inner part (inverse Kalman fitter).

ticle. The tracking is done until the intersection of the start of the innermost calorimeter is reached. At each intersected pseudo-layers, the covariance matrix of the track is calculated. The covariance matrix is translated along the trajectory and multiple scattering effects are included into calculation. In a basic view, it is what a Kalman filter is doing but not in the formal way. The track fitted is matched to the vertex and the perigee parameters are then smeared according to the calculated covariance matrix. The tracking efficiency has been parametrised from the full simulation.

1.5.2 Calorimeter Simulation

For the calorimeter simulation, the particle is extrapolated to the intersections with the calorimeters. At this point, a decision is made, that it should be detected as a minimum ionising particle, or that it should initiate an electromagnetic shower, or a hadronic shower, or that it is below the threshold... According to the decision, the detector response is simulated from different parameters (energy, type of particle, detector region...). Some parameters can be controlled by steering files, like shower can be merged if they are close to each others. To go towards more realism, the simulation of confusion between clusters can be done (Particle Flow simulation).

1.5.3 SGV Particle Flow parametrisation

Usual problems with fast simulation are already implemented in SGV: errors on detected energy, shower position and shape. However, there are association errors: Clusters might be merged, split or get associated to the wrong track. If (a part of) a neutral cluster is wrongly associated to a charged track (so then considered as a charged cluster), energy is then lost (Figure **??**). On the other hand, if (a part of) a charged cluster is not associated to any track (considered then as a neutral cluster), the energy is double counted (Figure **??**). Theses mis-associations give rise to an error on the total energy of an event and the momentum [**?**].

The study of these association errors was done by using LOI mass produced samples from full simulation using the particle flow algorithm PandoraPFA [**?**]. The most relevant observables were identified as: the cluster energy, the distance of the nearest particle of "the other type" (i.e. neutral-to-charged or charged-to-neutral), whether the particle was a hadron or not, and whether it would be detected in the barrel or end-cap calorimeters. The confusion was factorized in 4 sub-processes [**?**]:

- The probability of a cluster to split or merge (Figure 1.13).
- In case of splitting, a probability to split off/merge the **entire** cluster.
- In case of splitting but not completely, a function of the fraction of split off.

From this, probability distribution functions are derived (Figures **??** and **??**). The algorithm is applied in the Particle Flow parametrisation of SGV. So far, the results seems to be in a good agreement mostly for the neutrals but still some development is needed to get the best agreement possible between SGV and the full simulation.

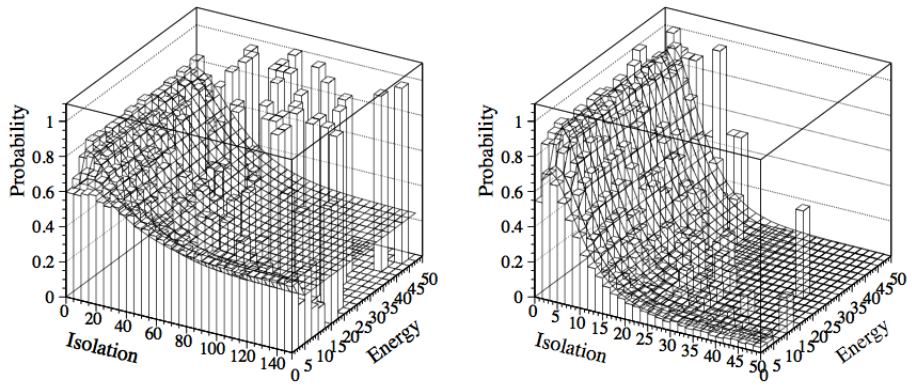


Figure 1.13 – Probability distribution of splitting in function of the cluster energy and the type of the particle.

1.5.4 Check Tracking efficiency SGV/Full simulation

Before reconstructing particles, PandoraPFA applies a selection of the tracks that can be a candidate to a reconstructed particle. In this tracking selection, PandoraPFA assumes that everything is a pion as a first guess. This is not modified after for track energy calculation.

Particles coming from V0s (decay of a particle in flight for example $\gamma \rightarrow e^+e^-$), Kinks (decay of a particle into another particle of the same charge with lower momentum giving a change in the curvature, for example $\pi \rightarrow \mu\nu$) or Prongs (decay of a τ) are identified in a first place and treated separately. Pandora PFA uses a track selection code in order to categorise them into multi parameters categories :

- CanFormPFO
- CanFormClusterlessPFO
- CannotFormPFO

To perform the categorisation, PandoraPFA applies several cuts on the track. First, it checks the number of hits of the track in the tracking chamber (TPC) and the forward tracker (the number of hits in the TPC must be between 5 and 5000 hits, in the forward tracker, a

number of expected hits is calculated based on the geometry of the forward tracker if tan λ of the track is within the acceptance of the forward tracker).

Then, Pandora checks if the track reaches the front face of the electromagnetic calorimeter. The conditions are that the radius of the outermost hit or the max z position of all hits is above a minimum radius (R^{SET}_{min}) or minimum z (Z^{ETD}_{min}), or if there is a sufficient number of hits in the TPC (11) or FTD (4). If the track has a low transverse momentum (meaning that the track may curl inside the inner tracker), that the cosine angle of the track is within the acceptance of the TPC or that the transverse momentum of the track is above a threshold (0.3*B*R^{TPC}_{outer}/2000).

If the track survive all these cuts, a quality check is performed on the tracks parameters: curvature (Ω), impact parameter (d_0), z position at the p.c.a (z_0), radius of innermost hit (r_{min}) and the track energy (E_{track}). The curvature must be different from 0, the momentum uncertainty σ_p/p must be under 15%. If the track momentum p is over 1 GeV, the transverse momentum p_T and the momentum projected on the z axis p_Z must be different of 0 GeV and finally the number of TPC or FTD hits must be over a certain value (for TPC hits, an expected number of hits is calculated based on the geometry and the track momentum and compared to the number of measured TPC hits, for FTD hits, the number must be more than 2 hits).

Once a track passes the quality checks, Pandora categorise the track on cut based differentiation. If a track has $d_0 < 50$ mm, $z_0 < 50$ mm and $r_{min} < R^{TPC}_{inner}$ + 50 mm, the track is then categorised as **CanFormPFO**.

A second criterion is used for non vertex tracks (tracks that are not matching the primary vertex), a check on $z_{min} - z_0$ (z_{min} is the z position of the first tracker hit) and the r_{min} of the track is done. If the track passes through the cuts, it is flagged also as **CanFormPFO**.

If a track (unmatched vertex track) has $d_0 < 5$ mm, $z_0 < 5$ mm, $r_{min} < R^{TPC}_{inner}$ + 50 mm and the track energy E_{track} is less than 5 GeV, the track is then categorised as **Can-**

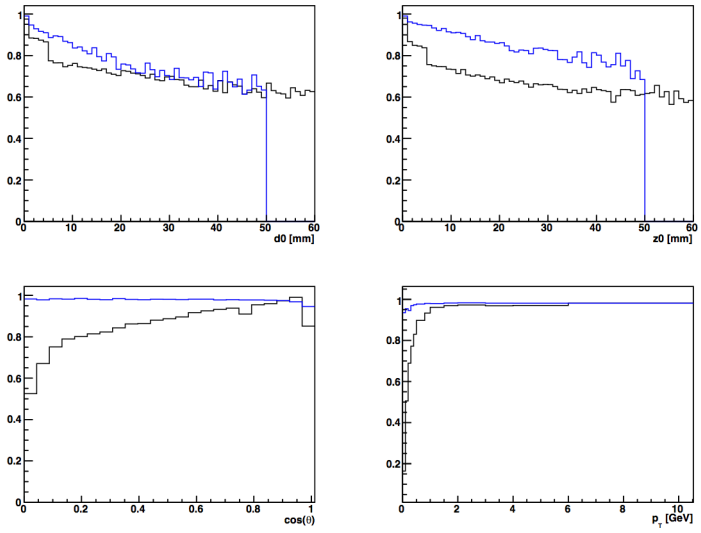


Figure 1.14 – Ratio plots for different track parameters for full simulation in black and SGV in blue.

FormClusterlessPFO. For obvious reasons, tracks matching these criteria can end up in the category **CanFormPFO**. This is then disentangled later by the track - cluster matching. For tracks that are in the first steps categorised as V0s, Kinks or Prongs are also flagged as **CanFormPFO** or **CanFormClusterlessPFO** depending on the same criteria above.

All the others tracks that do not meet the criteria are then flagged **CannotFormPFO**, theses tracks will then not form any PFO in the reconstruction.

The goal was to check if whether Pandora rejects tracks with a PFO in SGV. As in SGV, no information is provided on the state of the track at a specific point or no tracker hit information is stored, only the relevant part of the selection made by PandoraPFA was applied. After the pseudo-selection, histograms ($h_{selected}$) for each track parameter (d_0, z_0, $\cos(\theta)$ and p_T) were filled. For tracks that have the flag 'inPFO" (meaning that the track was indeed used by Pandora and formed a PFO), histograms (h_{inPFO}) with the track parameters were also filled.

To compared the performance of SGV to full simulation, a ratio was defined as the ratio of h_{inPFO} / $h_{selected}$ for the full simulation. In SGV, all tracks are by default forming a PFO,

with the selection a certain number of tracks might not pass thus with the ratio definition above, it might be over 1. In that case, for SGV, the ratio was defined as $h_{selected}$ / h_{inPFO}. We obtain the plots shown in Figure 1.14:

As expected for the full simulation, the impact parameter (d_0) and the z position at the p.c.a (z_0) ratio plots are more or less constant, as for SGV it seems to drop but the influence in general is not a problem. For the p_T plot, SGV seems to have a better ratio, but this is more or less dependant on the geometry parametrisation of SGV which is relevant for low p_T particles. The main difference appears for the $cos(\theta)$ parameter, the full simulation seems to drop very much for very low angles compared to SGV. After further investigation, theses tracks are curlers in the TPC that make through until the endcap, at this reconstruction step, theses tracks are considered by Pandora to be able to form a PFO but further a matching between clusters and track enables to get rid of most of theses curlers. By introducing a harder cut for theses curlers, basically looking at the z_{min} - z_0 (z_{min} is the z position of first hit) that should be less than half a turn of the helix. We obtain the plots shown Figure **??**: One can see that after SGV and full simulation agrees more after this harder cut. The cut seems still not enough to get rid of most of the curlers looking at the $cos(\theta)$ distribution. From this, one can say that the tracking selection in PandoraPFA is not responsible for the observed discrepancy.

1.5.5 Track Multiplicity and Correlations

A look at the number of tracks or multiplicity inside a jet was performed in order to see if SGV performs as expected in full simulation. The number of tracks were counted per jet energy bins for full simulation, SGV with and without ParticleFlow as shown on Figure 1.15.

One can see that the multiplicity of tracks are in very good agreement between SGV and full simulation. This is not the only variable that we looked at. Typically inside a jet,

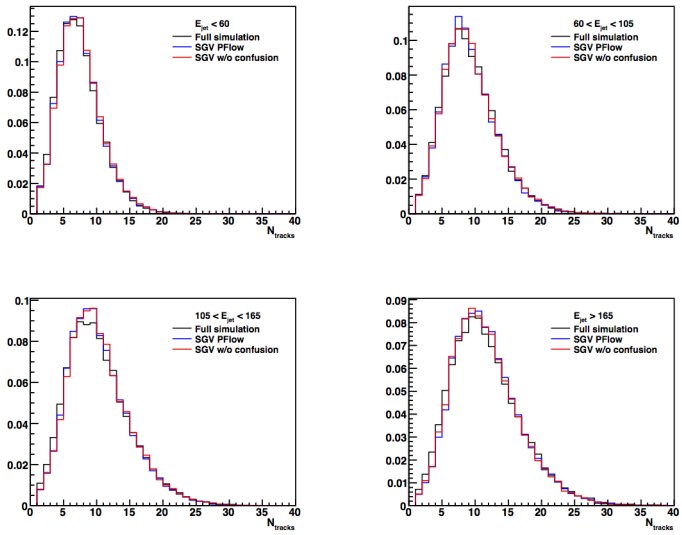


Figure 1.15 – Track Multiplicity for SGV and full simulation. The results are in a very good agreement.

60-70% is composed of charged particles and 20-30% of neutrals (neutrons and photons), thus the correlation between charged/neutral energy and track multiplicity was also looked at. The distributions shown on Figure **??** are in agreement and correspond to what one can expect.

1.6 Particle Flow studies

1.6.1 Double counted and lost energy

Association errors can happen during reconstruction as explained in section 1.5.3 because of a confusion term (the confusion is that sometimes it is very difficult to say from which particle a cluster is coming from, then a cluster is a sum of particle contributing with different weight to the cluster energy) coming from the overlap of showers into the calorimeters.

On Figure 1.16, there are mainly 2 regions, the top right corner and the bottom left cor-

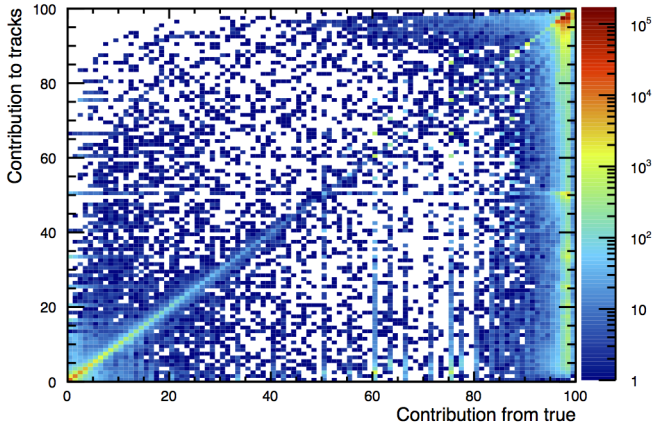


Figure 1.16 – Distribution of the weights of track-true particle relation. The contribution from true represent the contribution of a true particle (in terms of hits from this true particle compared to the total hits in the track) to a track, and the contribution to track represent the inverse relation, e.g. the number of hits from a true particle compared to the total of simulated hits from this true particle.

ner. The top right corner represents the region where there is almost no confusion between a track and the Monte-carlo particle (or true particle) associated (there is almost a one-to-one relation between a track and a true particle, e.g, one true particle is associated to a track). The bottom left corner is the region where the confusion dominates, this region shows that it is difficult to associate correctly a particle to a track, the contributions seems to come from different particles.

The goal of particle flow is to avoid confusion as much as possible. There can be different points of view concerning the treatment of the double counted and lost energy:

- At a cluster-track level, by comparing the energy of a track (momentum + assumption of π mass) to the energy of the cluster associated to the track (it would be then an energy flow point of view).
- At a jet level (clusters contained into a jet of particles), by looking at the double counted and lost energy over the jet energy.

For characterising the particle flow parametrisation performance of SGV compared to

the full simulation, a DBD sample was taken. The process chosen: $e^+e^- \rightarrow W^+W^- \rightarrow q\bar{q}q\bar{q}$ at $\sqrt{s}$ = 500 GeV, is particularly interesting in order to evaluate the separation between the two W bosons and the full hadronic decay complicating the reconstruction.

1.6.1.1 At Cluster-Track level

At a cluster-to-track point of view, the following method has been pursued. For each event, each reconstructed particle is taken, then we look at the track associated to it and calculate the track energy as $E_{track} = \sqrt{\|\overrightarrow{p}\| \cdot \|\overrightarrow{p}\| + m_\pi \cdot m_\pi}$ with m_π = 0.139 GeV and $\overrightarrow{p}$, the reconstructed particle momentum (the pion assumption is made like in PandoraPFA which in most cases is not wrong) [1.5.4]. Then we looked at the cluster associated to the reconstructed particle, and took the energy of the cluster $E_{cluster}$. A comparison between the track energy E_{track} and the cluster energy $E_{cluster}$ is done.

If $E_{track} < E_{cluster}$, the difference is counted as double counted energy (naturally the energy of the cluster should be the track energy because the resolution of the tracker is much better than the calorimeter, so if there is more energy in the cluster it means that a part of it might come from a near neutral particle or from a mis-measurement). If the opposite, then it is counted as lost energy. The double counted and lost energy is summed up for all the particles in the event.This method is done for each reconstructed particles in an event, thus for each event we have a point E_{dc} and E_l. The result is shown in Figure 1.17.

Concerning the full simulation, it seems that Pandora is compensating between the lost and double counted energy. One can think that Pandora balances both quantities event by event. For SGV, the correlation is not bad with the parametrisation but seems still to differ from the full simulation, one can see that SGV is more double counting energy than losing energy.

On the total energy (Figure 1.18), SGV without confusion seems to describe well the total energy but doesn't include the reconstruction effects widening the distribution. When

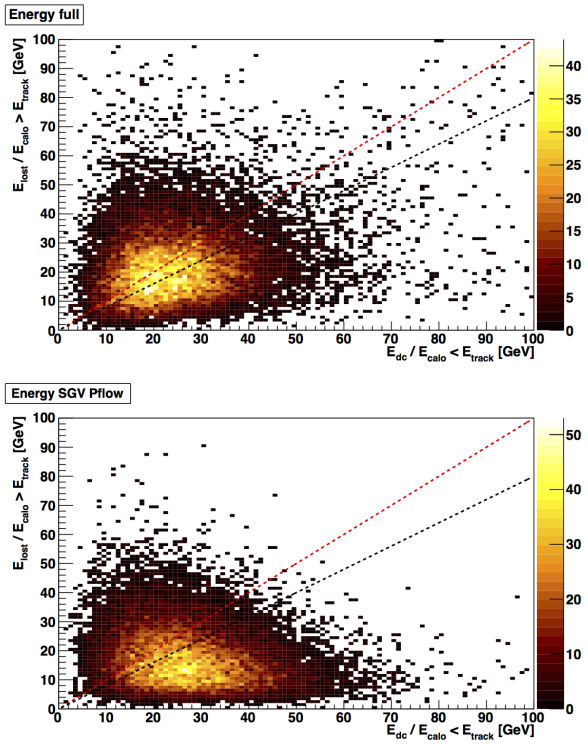


Figure 1.17 – 2D plot representing the lost energy versus the double counted energy in a full simulation of $e^+e^- \rightarrow q\bar{q}q\bar{q}$ sample at $\sqrt{s}$ = 500 GeV. Each point in this plot represent an event. The red and the black dotted line represent a correlation of 100% and 80%. One can observe that there is different correlations for SGV and the full simulation.

the confusion is turned on, a tail in the total energy appears and is not in accordance to the full simulation but the width of the distribution seems to be the same, it is then just a shift in energy. Thus, we looked at the energy distribution of neutral and charged reconstructed particles (Figure 1.19).

The neutral distribution improves in SGV with the particle flow parametrisation and agrees well with the full simulation, but for the charged distribution, one can see that the confusion doesn't change anything (as expected because only the track information matters and the cluster is not related to it) and as expected the distribution is shifted to higher energies explaining the tail observed in the total energy distribution. This gives us a hint that the parametrisation should have an effect on the charged particles which can be possible in the case that the track information is rejected and only the calorimeter information is taken into

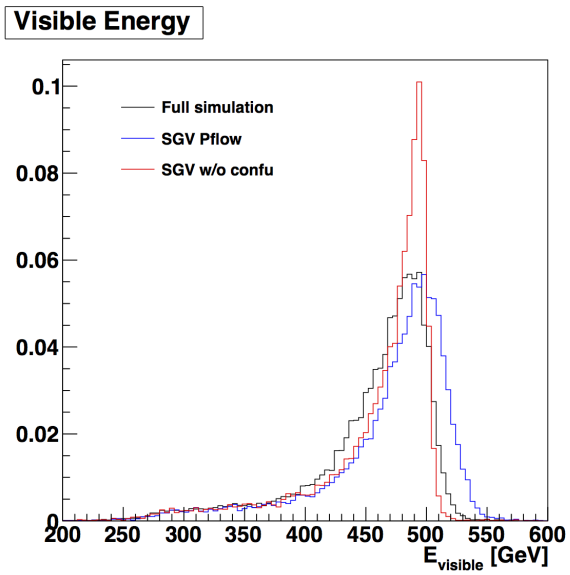


Figure 1.18 – The plot represent the total energy or all the events for the full simulation and SGV with and without the particle flow parametrisation.

account. Because of this, we decided to look at the same quantity but at the jet energy level, in order to see at what energy scale the discrepancy appears.

1.6.1.2 At Jet level

The process $e^+e^- \rightarrow q\bar{q}q\bar{q}$ at $\sqrt{s}$ = 500 GeV has a typical topology of 4 jets from the 4 primary quarks. The jets are obtained by running the Durham algorithm after the reconstruction. The Durham algorithm is a k_T-algorithm, it clusters all reconstructed particles into jets.

First, jets are classified into few energy bins: 0 to 60 GeV, 60 to 105 GeV, 105 to 165 GeV and over 165 GeV. For each event, a jet is selected by different energy E_{jet} : $<$ 60 GeV, between 60 and 105 GeV, between 105 and 165 GeV and $>$ 165 GeV (Figure 1.20). Inside each jet, each reconstructed particle is taken. Then the same method as the section above is used. Now, we calculate E_l and E_{dc} per jet. At the end, the lost energy versus the double counted energy normalised to the jet energy (Figure **??**).

One can see that for low jet energies, SGV and the full simulation seems to be quite

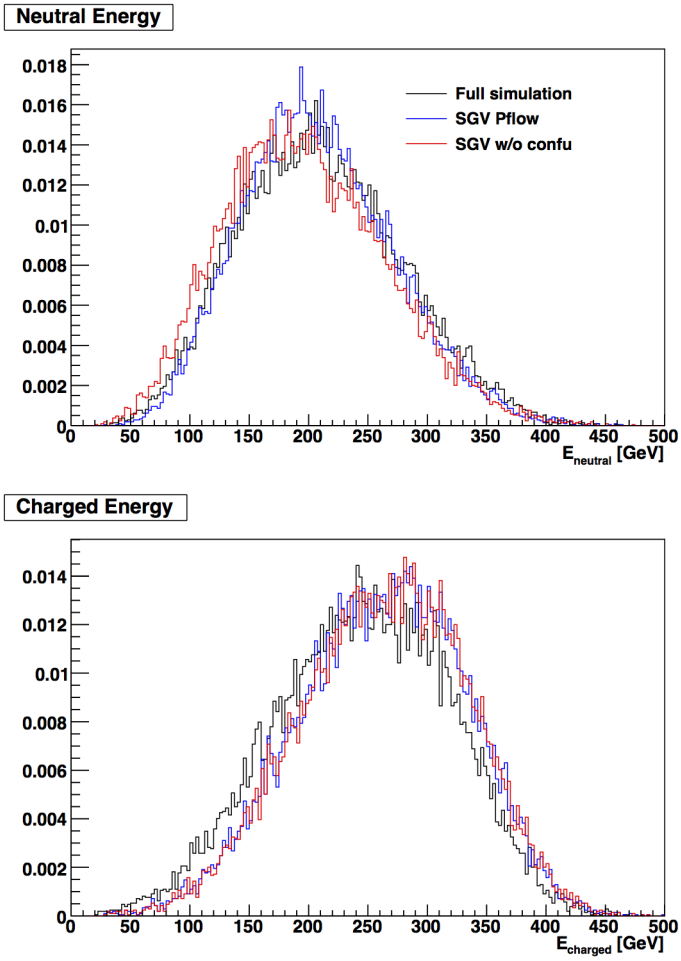


Figure 1.19 – Charged and Neutral energies distributions for SGV (with and without Particle Flow) and full simulation.

in agreement. But the higher we go in jet energy, the more both are getting different. We can mostly see that for jet energies over 165 GeV, SGV is double counting much more than the full simulation which stays with a good correlation between E_{dc} and E_l. This gives us indication that SGV is failing in regions where jet energies are high (around $>$ 165 GeV).

The reason could be that, in theses cases, the energy density in the calorimeter is so high that the association errors can be committed more easily because the confusion term of the overlapping showers is getting bigger and bigger in function of the jet energy. PandoraPFA, to solve this problem, is switching to a pure Ënergy Flowm̈ode. It is completely discarding the track information and only keeps the calorimeter information and then by reclustering,

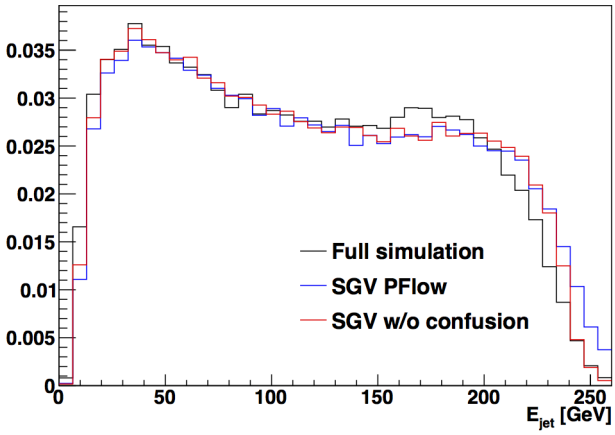


Figure 1.20 – Jet energy spectrum between full simulation and SGV. The jet energy bins were chosen in order to get the approximate same number of events in each bin.

PandoraPFA is matching the overall energy in a calorimeter region to the tracks.

In the current version of SGV, it is not implemented yet. We think that the merging and splitting probabilities should be then function of the energy density in the calorimeter region studied.

1.6.2 Energy fraction

The goal of this study was to look at the energy distribution for each jet energy bins. It can give us a clear view on the shape of the distribution and the effect of the particle flow parametrisation on it.

For the energy distribution plots on Figure **??**, one can observe that the plots of neutral and charged energy are mirror to each other, as the sum of the charged and neutral energy should be equal to the jet energy. For low energy jets, the distributions seems to be rather in a good agreement, only small discrepancies are visible. For higher energy jets, this discrepancy is getting bigger. It seems that there is more energy in the 50-70% region and much less in the 10-20% region. It seems that SGV is pulling the energy in the wrong direction.

The idea is that charged energy from the 50+% region should be transferred to the 10-20% region, meaning that charged clusters are transformed to neutral clusters. That is how the total charged energy distribution could be shifted toward lower energies by ïosingc̈harged energy and gaining neutral energy. This is consistent with the assumption that PandoraPFA is switching to an energy mode. It can occurs during this process that charged cluster are transformed to neutral cluster in order to match the E/p locally. PandoraPFA only takes care of the calorimeter information and reject the track information.

1.6.3 Occupancy and Energy density

One relative variable in the splitting and merging probabilities is the distance between a cluster of one type (hadronic or EM) and a cluster of the other type. The study of the distribution of the distance to the nearest neighbour was performed distinguishing between ECAL and HCAL (basically EM and hadronic showers). The procedure is perform such as: in each jet, a list of neutral and charged particles is filled. Each of theses particles are projected either on the barrel or the endcap respectively of the ECAL or the HCAL. The projection is done in order to calculate distances on the same geometry planes (SGV and the full simulation have a slight different geometry, the Barrel in SGV is considered as a cylinder of radius of the ECAL).

After that, we take each neutral and calculate the 3D distance: $r_i = \sqrt{x_i^2 + y_i^2 + z_i^2}$ with x_i, y_i, z_i the coordinates of the particle i intersected to the endcap or the barrel approximated to a cylinder of radius of the ECAL to each charged particles distinguishing between endcap and barrel to get rid of corner effects. The minimum distance d_{min} defined as $\min(r_i)$ for each neutral is then filled into an histogram, meaning there is an entry for each neutral particle. We obtain the plots shown on Figure **??**.

One can observe that there is a discrepancy between full simulation and SGV with Particle Flow in the ECAL. The Particle Flow parametrisation seems to have the effect that

particles are placed closer in the ECAL. On the other hand, the Particle Flow parametrisation seems to have a rather good effect in the HCAL. The idea then was to d̈isable PFAf̈or the ECAL and look at the same distributions shown of Figure **??**.

In that case, we could observe that the distributions are in rather good agreement for the ECAL and HCAL. The Particle Flow parametrisation seems to have a rather limited effect on the ECAL distribution compared to SGV without Particle Flow.

Is the parametrisation of Particle Flow in SGV useless for the ECAL? In order to check the influence of the parametrisation on the ECAL energy distribution, the charged and neutral energy were looked at again. As expected, switching off the Particle Flow parametrisation for the ECAL have rather no or limited impact on the charged energy distribution as shown on Figure 1.21.

One variable to look at would be the occupancy of the detector in the jet region or the energy density. This is could give us a clue to understand how PandoraPFA is splitting and merging in function of the energy density and then implement or correct the splitting and merging probabilities used in SGV to parametrise Particle Flow.

1.7 Conclusion

In Conclusion, the benchmark of the fast simulation SGV was done. A Particle Flow parametrisation is implemented in SGV and most of the results obtained are in agreement with the full simulation. Still the parametrisation is not perfectly correct as there is still some discrepancies between SGV and the full simulation (Total energy, distance to the nearest neighbour, correlation between E_{dc} and E_l...).

The next steps would be to look at the energy density distribution in a jet region and parametrise the splitting and merging probabilities of a cluster in function of the density and achieve an even better agreement between SGV and the full simulation.

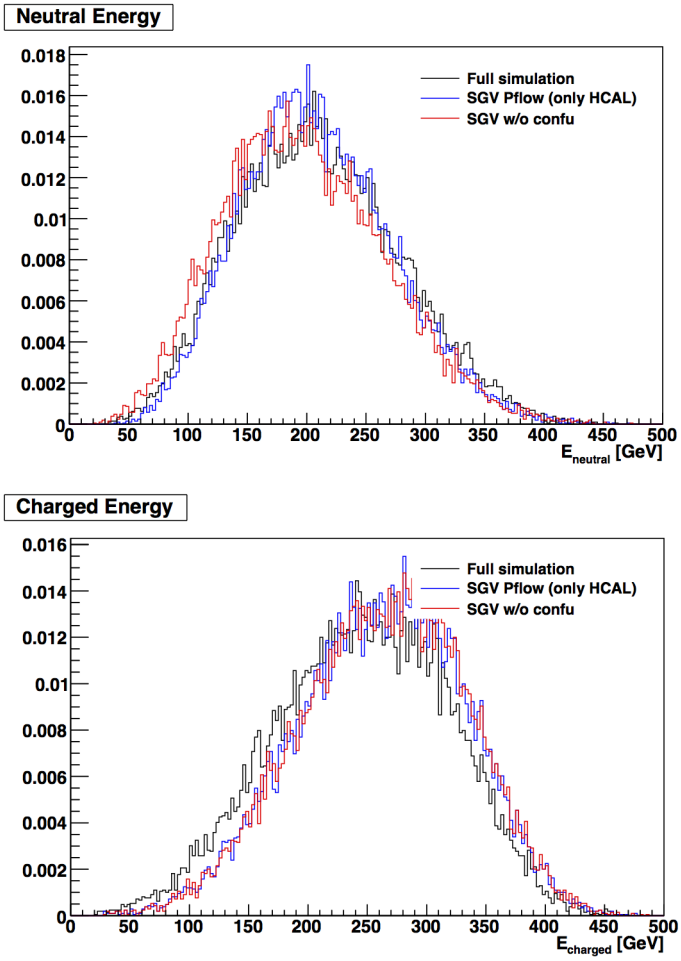


Figure 1.21 – Neutral and Charged energy distributions with ParticleFlow disabled for the ECAL.

1.8 Conclusion

References

[1] ILCSoft web portal. URL http://ilcsoft.desy.de/portal.

[2] Oskar Hartbrich Daniel Jeans. Realistic calorimeter hit digitisation in the ILDCaloDigi processor. 2015.

[3] R. Fruhwirth. Application of Kalman filtering to track and vertex fitting. *Nucl. Instrum. Meth.*, A262:444–450, 1987. doi: 10.1016/0168-9002(87)90887-4.

[4] F. Gaede. Marlin and LCCD: Software tools for the ILC. *Nucl. Instrum. Meth.*, A559: 177–180, 2006. doi: 10.1016/j.nima.2005.11.138.

[5] Frank Gaede, Ties Behnke, Norman Graf, and Tony Johnson. LCIO: A Persistency framework for linear collider simulation studies. *eConf*, C0303241:TUKT001, 2003.

[6] Oskar Hartbrich. *Scintillator Calorimeters for a Future Linear Collider Experiment*. PhD thesis, Hasylab, DESY, Hamburg, 2016.

[7] M. A. Thomson. Particle Flow Calorimetry and the PandoraPFA Algorithm. *Nucl. Instrum. Meth.*, A611:25–40, 2009. doi: 10.1016/j.nima.2009.09.009.

Acknowledgments