Something something something physics

Steven Green
of Emmanuel College

A dissertation submitted to the University of Cambridge
for the degree of Doctor of Philosophy

Abstract

This thesis describes the optimisation of the calorimeter design for collider experiments at the future Compact LInear Collider (CLIC) and the International Linear Collider (ILC). The detector design of these experiments is built around high-granularity Particle Flow Calorimetry that, in contrast to traditional calorimetry, uses the energy measurements for charged particles from the tracking detectors. This can only be realised if calorimetric energy deposits from charged particles can be separated from those of neutral particles. This is made possible with fine granularity calorimeters and sophisticated pattern recognition software, which is provided by the PandoraPFA algorithm. This thesis presents results on Particle Flow calorimetry performance for a number of detector configurations. To obtain these results a new calibration procedure was developed and applied to the detector simulation and reconstruction to ensure optimal performance was achieved for each detector configuration considered.

This thesis also describes the development of a software compensation technique that vastly improves the intrinsic energy resolution of a Particle Flow Calorimetry detector. This technique is implemented within the PandoraPFA framework and demonstrates the gains that can be made by fully exploiting the information provided by the fine granularity calorimeters envisaged at a future linear collider.

A study of the sensitivity of the CLIC experiment to anomalous gauge couplings that effect vector boson scattering processes is presented. These anomalous couplings provide insight into possible beyond standard model physics. This study, which utilises the excellent jet energy resolution from Particle Flow Calorimetry, was performed at centre-of-mass energies of 1.4 TeV and 3 TeV with integrated lumi-

nosities of 1.5ab^{-1} and 2ab^{-1} respectively. The precision achievable at CLIC is shown to be approximately one to two orders of magnitude better than that currently offered by the LHC.

Finally, a study into various technology options for the CLIC vertex detector is described.

Declaration

This dissertation is the result of my own work, except where explicit reference is made to the work of others, and has not been submitted for another qualification to this or any other university. This dissertation does not exceed the word limit for the respective Degree Committee.

Andy Buckley

Acknowledgements

Of the many people who deserve thanks, some are particularly prominent, such as my supervisor. . .

Preface

This thesis describes my research on various aspects of the LHCb particle physics program, centred around the LHCb detector and LHC accelerator at CERN in Geneva.

For this example, I'll just mention Chapter **??** and Chapter 1.

x

Contents

"Writing in English is the most ingenious torture ever devised for sins committed in previous lives."
— James Joyce

Chapter 1

Energy Estimators

"There, sir! that is the perfection of vessels!"
 — Jules Verne, 1828–1905

1.1 Calibration

1.1.1 Calibration in the particle flow paradigm

TODO: Trim down.

In any experiment, calibration is essential for ensuring reliability in measured quantities and the linear collider will be no exception to this. At the linear collider there will be several measured quantities each of which will be converted into a measure of the energy deposited in a given region of the detector. These fall into two distinct classes (i) calorimeter energy deposits and (ii) track deposits. The focus of this section will be on the energy deposits in the calorimeter and the procedure developed to ensure that they are reliable.

Calorimeter energy deposits are an essential building blocks for the application of the particle flow paradigm. The separation of energy deposits from charged and neutral particles in the calorimeters is crucial to achieve the full potential of particle flow and this is only possible with accurate energy estimators for those energy deposits. A robust calibration scheme has been developed and will be discussed in the following chapter.

The other crucial energy deposit used in particle flow calorimetry are track energy deposits. These are also crucial to physics performance, however, in the particle flow paradigm these energy deposits are topologically related to the energy of the reconstructed particle. A spatial helix fit is applied to the track energy deposits which when combined with knowledge of the magnetic filed yields the momentum of the particle producing the track. Therefore, there is no direct relationship between the energy deposited by the monte-carlo particle in the active medium and the energy of the reconstructed energy. Therefore, precise calibration of the energy deposited by a charged particle track is less crucial than for calorimeter energy deposits. For this reason the focus of this chapter is on the calibration of calorimeter energy deposits.

Calibration of the linear collider simulation extends beyond the raw calorimeter hits and into the particle flow algorithm itself. The fine calorimeter granularity required for particle flow calorimetry yields excellent physical separation of hadronic and electromagnetic showers. Thanks to sophisticated particle identification occurring within PandoraPFA it is possible to distinguish hadronic and electromagnetic showers, which allows for distinct treatments of the hadronic and electromagnetic energy estimators. This distinction can be used to produce a response from the calorimeter that is compensating despite the intrinsic response being non-compensating. A compensating calorimeter would give significant improvements to the energy resolution for the detector.

Two treatments designed to achieve a compensating response from the calorimeters are discussed. The both involve rescaling the energies, the first rescaling is applied using a series of fixed energy independent constants, while the second uses the energy density of the calorimeter hits in the shower to determine the energy rescaling factor.

Details regarding how the detector performance metrics used in this chapter are calculated can be found in section **??**.

1.1.2 Calibration and detector optimisation

Optimising the detector at a future linear collider will be crucial to exploit the full physics potential available to it. An extensive optimisation of the calorimeters was performed and the results can be found in chapter **??**. For each detector model considered in this study the calibration procedure outlined in this section was applied to ensure optimal performance was achieved. This made unbiased comparison between detector

models performance possible and ensured reliability in the conclusions drawn from this study.

1.1.3 Calibration Goals

The calibration procedure aims to accurately set parameters related to four aspects of the reconstruction, which are:

1. **Digitisation of calorimeter hits**. Digitisation in this sense is the estimation of the energy deposited within a calorimeter cell, the active and absorber layers, based on the signal measured in the sensitive region of the cell, the active layer.

2. **Minimum ionising particle (MIP) scale setting in the digitisation processor and PandoraPFA**. The MIP scale has to be set in the digitiser as it simulates the response of the readout technology including a maximum readout value, which is in units of MIPs. The digitiser also applies a minimum active layer cell energy threshold, in units of MIPs, that has to be passed for creation of a calorimeter hit to occur. This is designed to veto noise that would be present in a real detector. PandoraPFA uses the MIP scale to place further threshold cuts on the cell energy that must be exceeded before a calorimeter cell is used in the reconstruction. Both of these thresholds are designed to veto noise that would be present in a real detector. While no noise is applied to the simulation these cuts are still applied to better represent the performance of a real detector.

3. **Electromagnetic and hadronic scale setting in PandoraPFA**. As discussed in chapter CALORIMETER CHAPTER, the response of a calorimeter to electromagnetic and hadronic showers is different due to the fundamentally different mechanisms governing their propagation. One key difference between the two is the presence of an invisible energy component in the hadronic shower. Therefore, the reported energy from a calorimeter to a hadronic showers will be lower than that of an electromagnetic shower. To account for effects such as this and to account for energy losses due to the application of noise vetoing cuts in PandoraPFA the PFO energies are rescaled by PandoraPFA depending on whether the PFO has showered electromagnetically or hadronically. Determination of these scaling factors is the setting of the electromagnetic and hadronic energy scales.

4. **Retraining photon likelihood data**. The PandoraPFA algorithm uses likelihood data to determine whether a reconstructed object is a photon. This likelihood data has to be retrained for every new detector model considered.

Each of these aspects needs separately addressing for each new detector model considered. The ordering of each of these calibration steps also had to be taken into consideration as it is possible to get interference between the different stages if applied in the wrong order.

1.1.4 Digitisation

Calibration of the digitisation of the calorimeter hits involves accurately estimating the energy deposited in a calorimeter cell, in both the active and absorber layers, based on the energy deposited in the sensitive element of the calorimeter, the active layer. The relationship between the energy deposited in the active layers and the absorber layers of a calorimeter is linear as the energy deposited in both layers are proportional to the number of charged tracks passing through them. The ratio of the calorimeter cell energy to the energy deposited in the active layers is hereby called the digitisation constant. . The digitisation constants depend upon several aspects of the detector including the material properties of the calorimeters, the magnetic field strength and energy losses occurring within the gaps between the active and absorber layers. To account the effect of instrumented read out technology that would exist in a completed calorimeter some material is included in the simulation around the active and absorber layer. In comparison to the absorber layer this extra material adds little to the detector, but energy losses here are accounted for in digitisation.

ECal Digitisation

The procedure for determining the digitisation constants in the ECal involved the simulation of single γ events at $E_{MC} = 10$ GeV. γ events are ideal for calibration of the ECal as in the particle flow paradigm γ energy measurements are made primarily within the ECal. Also at this energy γs are largely contained within the ECal, as shown in figure 1.1, making them ideal for isolating the ECal digitisation calibration from that of the HCal digitisation calibration.
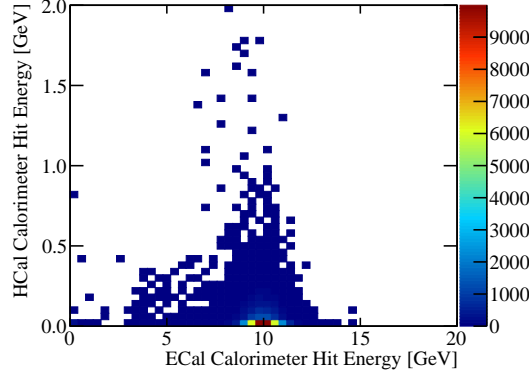
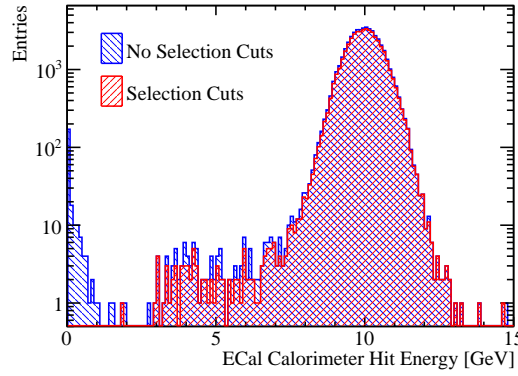**Figure 1.1:** Sum of calorimeter hit energies in ECal and HCal for 10 GeV γ events.

Figure 1.2: Sum of the ECal calorimeter hit energies for 10 GeV γ events with and without the selection cuts.

Events are only considered in this analysis if they are confined to the ECal and to that extent cuts are applied ensuring that the sum of any reconstructed energy found outside the ECal is less than 1% of E_{MC} and that the $\cos(\theta) < 0.95$ where θ is the polar angle of the γ. γ conversions are also vetoed from this event sample at this stage by requiring the reconstruction to give a single photon PFO. The impact of these cuts on the sum of ECal calorimeter hit energies for the $E_{MC} = 10$ GeV γ events is shown in figure 1.2.

The calibration procedure is iterative and begins with the simulated single γ events using a trial calibration, with digitisation constant in the ECal α^0_{ECal}, which may not be ideal. Then the distribution of the sum of calorimeter hit energies within the ECal is produced for events passing the selection cuts using, as shown in figure 1.2. For an ideal calorimeter this distribution should be Gaussian, as was described in section
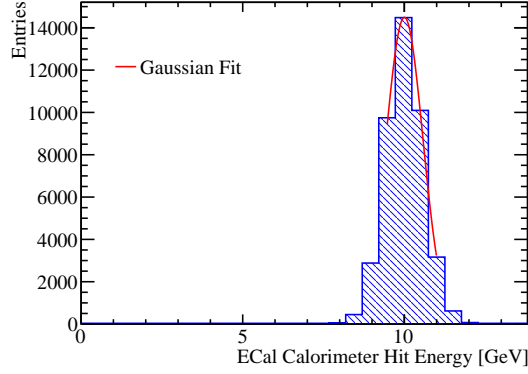
Figure 1.3: Gaussian fit to sum of the ECal calorimeter hit energies for 10 GeV γ events with selection cuts.

CALORIMETER CHAPTER. Therefore, a Gaussian fit is applied to this distribution and the mean, E_{Fit}, extracted. In an attempt to remove the effect of any outliers in this distribution, the fit is applied to the range of data with the smallest root mean square that contains at least 90 % of the data. An example of such a fit is shown in figure **??**. In the case of perfect calibration the mean of this fit would be equal E_{MC} and it is assumed that any deviation between the two is due to inaccurate calibration. To correct for any such deviation the digitisation constant from the trial calibration, α_{ECal}^{0}, is rescaled by the ratio of the E_{MC} to E_{Fit}.

$$\alpha_{\text{ECal}}^{0} \to \alpha_{\text{ECal}} = \alpha_{\text{ECal}}^{0} \times \frac{E_{MC}}{E_{Fit}} \tag{1.1}$$

This procedure is then repeated until the E_{Fit} falls within a chosen tolerance of E_{MC}. The tolerance that was applied here is that $|E_{\text{Fit}} - E_{\text{MC}}| < E_{\text{MC}} \times 5\%$. The binning for the fitted histogram is chosen such that the bin width is equal to the desired tolerance on E_{Fit} e.g. $E_{\text{MC}} \times 5\% = 0.5$ GeV. This tolerance is somewhat large, however, it is tight enough to ensure successful application of PFA. It should also be emphasised that the PFO energies used in downstream analyses have the electromagnetic and hadronic energy scale corrections applied and these are calibrated to a much tighter accuracy.
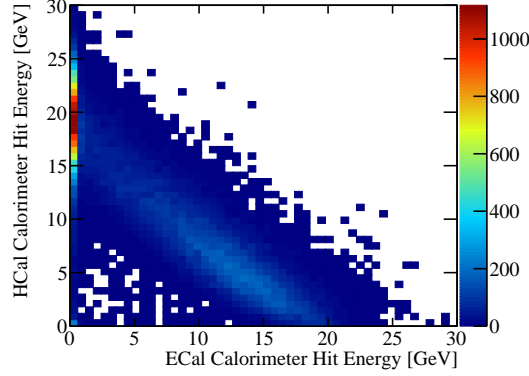
Figure 1.4: Sum of calorimeter hit energies in ECal and HCal for 20 GeV K_L^0 events.

HCal Digitisation

The calibration for the digitisation in the HCal proceeds in a similar manor to that described for the ECal with a few key differences. This calibration uses K_L^0 events at $E_{MC} = 20$ GeV as these neutral hadrons will deposit the bulk of their energy in the HCal. The higher energy is used to create larger particle showers and sample deeper into the calorimeters.

As the K_L^0 samples have to pass through the ECal before arriving at the HCal and as the ECal contains $\approx 1\lambda_I$, some of the particles begin showering in the ECal, as shown by figure 1.4. These events are unsuitable for calibration as rescaling α_{HCal}^0 does not lead to a linear rescaling in the mean of E_{Fit}. This issue is resolve by applying selection cuts to select events that deposit the bulk of their energy in the HCal.

Events are only considered in this analysis if a single neutral hadron PFO is reconstruction, the sum of any reconstructed energy found outside the HCal is less than 5% of E_{MC} and the last layer of the HCal where energy is deposited is in the first 90% of the HCal. The cut on the last HCal layer where energy is deposited is applied to veto events that shower late in the HCal and deposit a significant amount of energy in the uninstrumented coil region of the detector. The impact of these cuts on the sum of HCal calorimeter hit energies for the $E_{MC} = 20$ GeV K_L^0 events is shown in figure 1.5.

There are two HCal digitisation constants used in the detector simulation, one applied for the HCal barrel and another for the HCal EndCap. This is to account for differences in hadronic shower dynamics that exist for a variety of reasons such as differing magnetic field configurations in the Barrel and EndCap. Both parameters
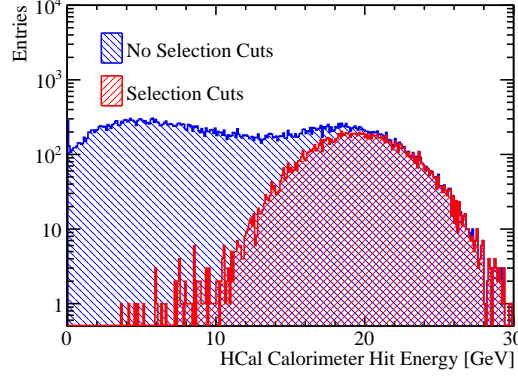
Figure 1.5: Sum of the HCal calorimeter hit energies for a 20 GeV K_L^0 events with and without the selection cuts.

are calibrated in the same manor, but have different cuts on θ, the polar angle of the K_L^0. For the Barrel region of the HCal $0.2 < \cos(\theta) < 0.6$, while for the EndCap $0.8 < \cos(\theta) < 0.9$. These angular cuts are conservative to account for the transverse profile of the hadronic showers and ensure that they are confined to the relevant sub-detector.

Using these cuts the calibration procedure for the digitisation of the HCal Barrel and EndCap proceeds in the same manor as was described for the ECal, the details of which can be found in section 1.1.4, and examples of the Gaussian fits applied to the sum of the calorimeter hit energies in the HCal Barrel and EndCap can be found in figure 1.6.

A noteworthy difference to the ECal digitisation procedure is that the target reconstructed energy for the K_L^0 samples is the kinetic energy as opposed to the total energy. This decision was made as the majority of the neutral hadrons appearing in jets are neutrons and their accessible energy, what they can deposit in the detector, is their kinetic energy and not their rest mass energy. This is the case as neutrons will come to a halt rather than decaying in the detector. Therefore, calibrating to the kinetic energy should give the best overall performance for jet reconstruction.

HCal Ring Digitisation

The HCal ring also has an independent digitisation constant for the same reasons that the Barrel and EndCap constants differ. The procedure used to calibrate this constant
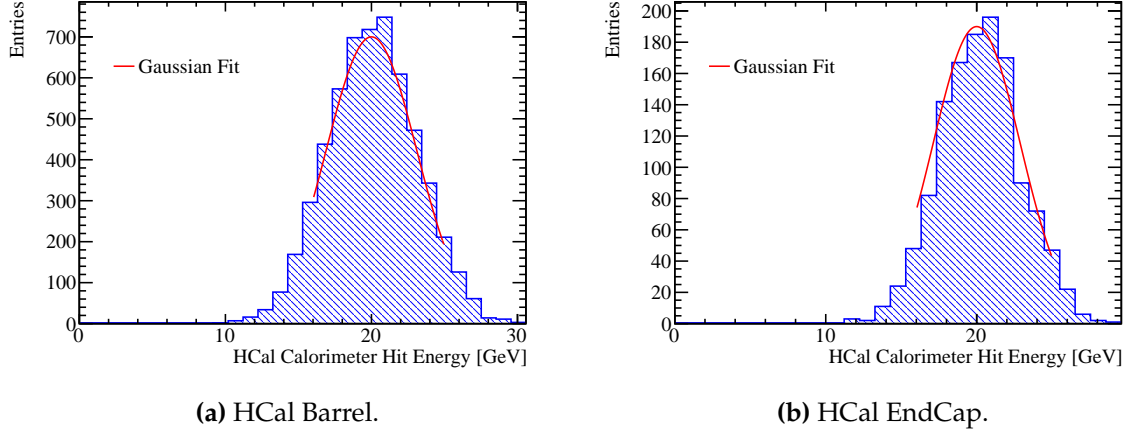
(a) HCal Barrel.

(b) HCal EndCap.

Figure 1.6: Gaussian fit to sum of the HCal calorimeter hit energies for 20 GeV K_L^0 events with selection cuts.

has to differ from that presented in section 1.1.4 as it is unfeasible, due to the depth of the ring, to produce events that are wholly contained within it. Fortunately, the size of the HCal ring means that it will play a minimal role in the reconstruction, so precise calibration is not crucial. To ensure that the calibration is approximately correct for the HCal Ring, $\alpha_{\text{HCal Ring}}$ is assumed to equal $\alpha_{\text{HCal EndCap}}$ multiplied by several factors designed to accounts for changes in the active layer thickness, absorber layer thickness and the MIP response between the HCal EndCap and Ring. In detail:

$$\alpha_{\text{HCal Ring}} = \alpha_{\text{HCal EndCap}} \times \frac{\langle \cos(\theta_{\text{EndCap}}) \rangle}{\langle \cos(\theta_{\text{Ring}}) \rangle} \times \frac{P_{\text{EndCap}}}{P_{\text{Ring}}} \times \frac{L_{\text{EndCap}}^{Absorber}}{L_{\text{Ring}}^{Absorber}} \times \frac{L_{\text{Ring}}^{Active}}{L_{\text{EndCap}}^{Active}} \quad (1.2)$$

where θ is the incident angle of the incoming particle to the calorimeter cells determined using the 20 GeV K_L^0 events, L^{Active} is the active layer thickness, $L^{Absorber}$ is the absorber layer thickness and P is the position of the MIP peak in the distribution of active layer cell energies, corrected so that the particles appear normally incident, using 10 GeV μ^- events. Details on how P is determined can be found in section 1.1.5.

1.1.5 MIP Scale Setting

The response of the various sub-detectors to a MIP has to be determined for both the digitisation processor and for PandoraPFA as both apply cuts in units of MIP response. The digitiser applies cuts related to the electronic readout range of the various active layer technology options and applies thresholds on the minimum active layer energy for the creation of calorimeter hits, while PandoraPFA applies cuts that would veto noise that would be present in a real detector. Both these MIP responses, while intrinsically linked, have to be calculated separately as the digitiser requires the MIP peak definition from the active layer cell energies while, PandoraPFA requires the definition from the full cell, active and absorber layer, energies. In these studies a MIP was defined as a 10 GeV μ^- [1] and there were no selection cuts applied to this sample.

For the digitiser the MIP scale was defined as the, non-zero, peak in the distribution of the active layer calorimeter cell energies for normally incident μ^- as shown in figure 1.7. This distribution was produced using a sample of μ^- events that are spatially isotropic about the impact point. A direction correction factor, $\cos(\theta)$ where θ is the indecent angle of the incoming μ^- to the calorimeter cell, was applied to the active layer cell energies to generate the effect of having normally incident μ^-.

In the digitiser processor a single value for the MIP peak was required for the HCal and that was taken as the MIP peak position for the HCal Barrel. The MIP peaks were separately calculated for the HCal EndCap and Ring for the purposes of the HCal Ring digitisation described in section 1.1.4. The realistic digitisation features present in the simulation of the ECal and HCal were not present for the muon chamber digitisation and so no MIP peak setting for that digitisation stage was required.

A similar procedure was employed for calculation of the MIP peak in PandoraPFA. One important difference was the distribution used for setting the MIP scale in PandoraPFA is the distribution of calorimeter cell energies, i.e. the energy in the active and absorber layers of a cell, and not just the active layer energies. Examples of the distributions used to set the MIP scale in PandoraPFA can be found in figure 1.8. There are few populated low calorimeter cell energy bins due to cuts applied in the digitiser on the minimum active layer energy required to make a calorimeter hit. The double peak structure in the ECal calorimeter hit energy distribution is present due to the doubling of the thickness of the ECal absorber material, from 2.1 mm to 4.2 mm tungsten, in the ILD detector model, which occurs for the back the first 10 layers of
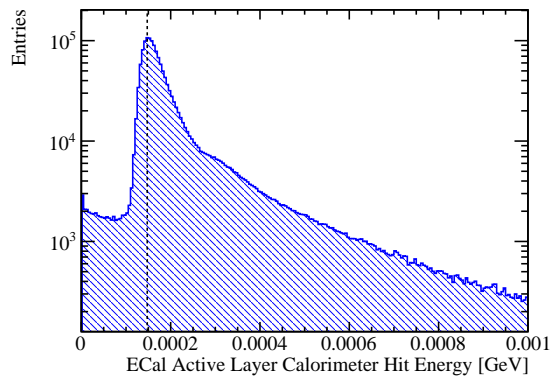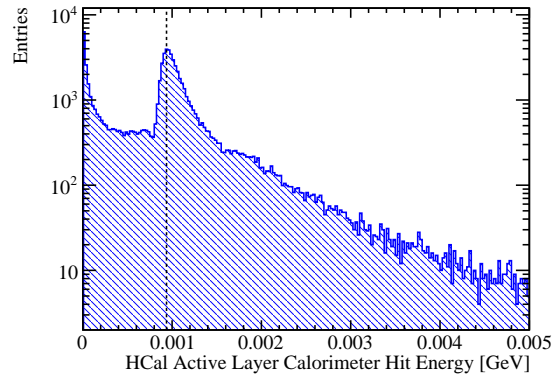
(a) ECal.
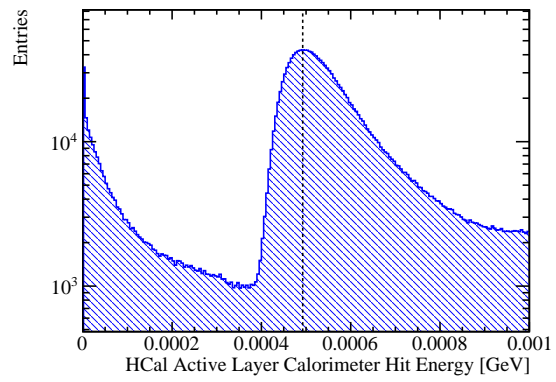
(b) HCal Barrel.
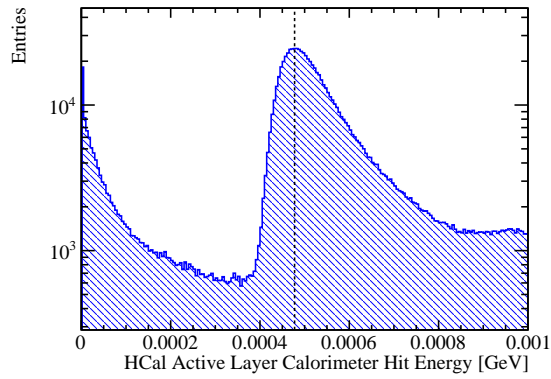
(c) HCal EndCap.

(d) HCal Ring.

Figure 1.7: The active layer calorimeter cell energy distributions for various sub-detectors for 10 GeV μ^- events. The MIP peak is represented by the black dotted line.
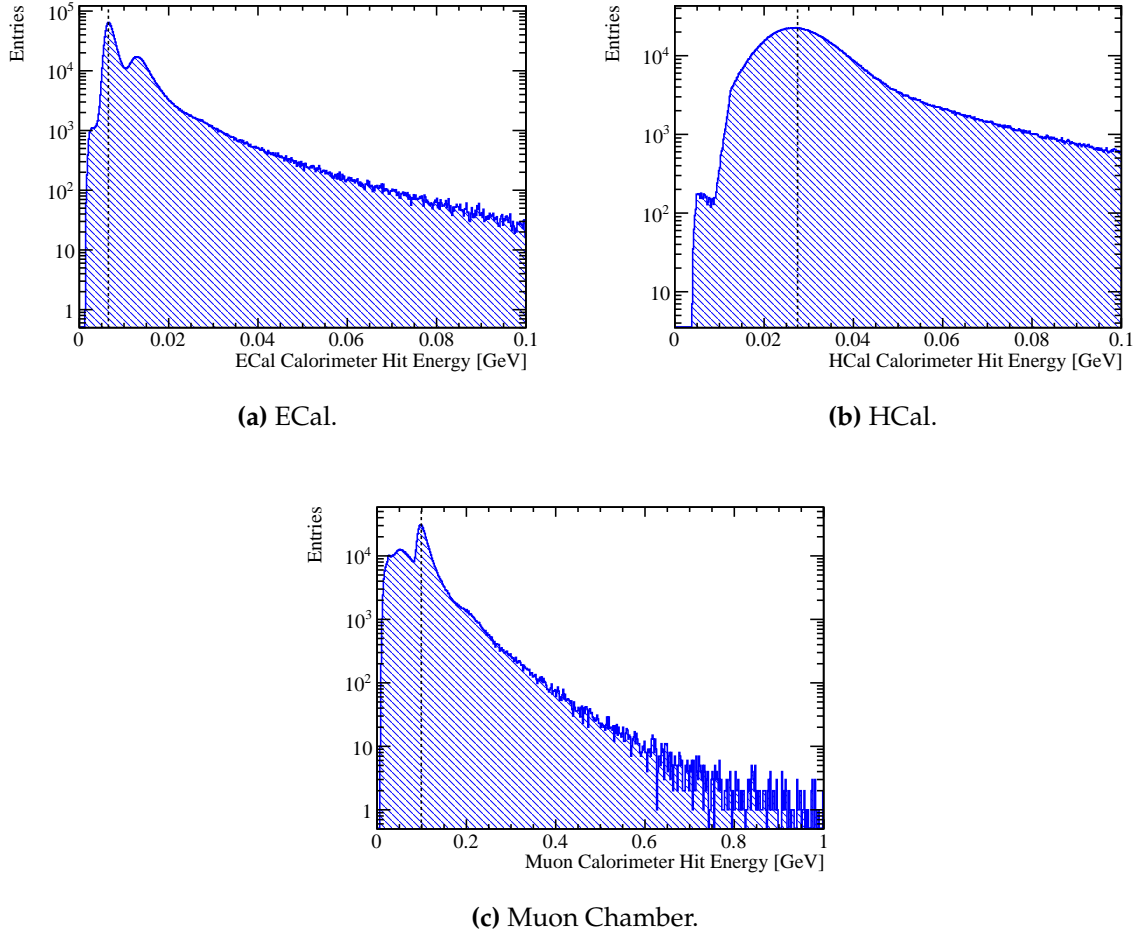
(a) ECal.

(b) HCal.

(c) Muon Chamber.

Figure 1.8: The calorimeter cell energy distributions for various sub-detectors for 10 GeV μ^- events. The MIP peak is represented by the black dotted line.

the 30 layer ECal. Two differences between the MIP scale setting in the digitiser and PandoraPFA worthy of note are that for the PandoraPFA MIP scale setting the HCal sub-detectors, the Barrel, EndCal and Ring, are all combined into a single cell energy distribution and PandoraPFA requires the MIP scale to be set in the muon chamber, which required a muon chamber cell energy distribution to be created.

1.1.6 Electromagnetic and hadronic scale setting

The electromagnetic and hadronic scales have to be independently set in the simulation to account for the different mechanisms governing the propagation of electromagnetic and hadronic showers. As referenced earlier, one crucial difference the setting of these
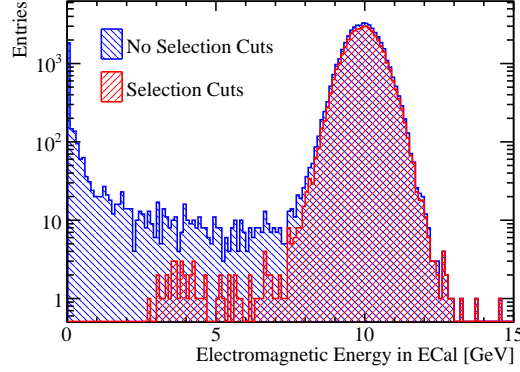
Figure 1.9: Sum of the electromagnetic energy measured in the ECal for 10 GeV γ events with and without the selection cuts.

scales accounts for is the invisible energy component of hadronic showers, however, in this simulation it also accounts for the effect of energy loss due to the threshold cuts that are applied in the reconstruction. The setting of the scales involves tuning four parameters in PandoraPFA, which correspond to the scaling factors that are applied to PFO energies arising from electromagnetic and hadronic showering particles in the ECal and HCal.

Electromagnetic scale setting

The electromagnetic scale in the ECal, β_{ECal}^{EM}, is set in the detector using γ events at $E_{MC} = 10$ GeV. γ events are ideal for the setting of the electromagnetic scale as they procedure electromagnetic showers and are primarily confined to the ECal at the energy considered as was shown in figure 1.1.

Cuts are applied to ensure that only events where the bulk of the energy is deposited within the ECal are applied. The cuts require a single γ be reconstructed, excluding conversions, and that less than 1% of the reconstructed energy is outside the ECal. The impact of these cuts on the electromagnetic energy measured in the ECal for these 10 GeV γ events is shown in figure 1.9.

The fitting procedure proceeds in a similar manor to that described in section 1.1.4 whereby a trial calibration for the electromagnetic energy scale in the ECal, β_{ECal}^{EM0}, is assumed and the single γ events simulated. The distribution of the PFO electromagnetic energy in the ECal is produced and a Gaussian fit applied to the range
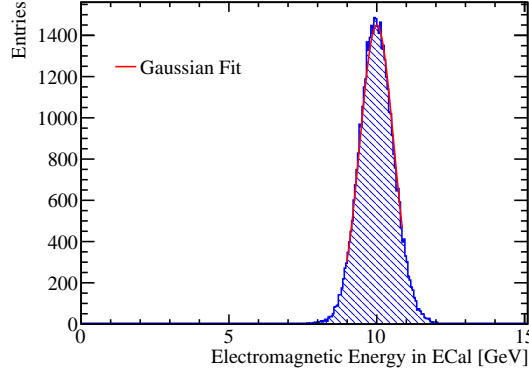
Figure 1.10: Gaussian fit to sum of the electromagnetic energy deposited in the ECal for 10 GeV γ events with selection cuts.

of data with the smallest root mean square that contains at least 90 % of the data. The mean of this fit, E_{Fit}, is then used to scale β_{ECal}^{EM0} in the following way

$$\beta_{ECal}^{EM0} \rightarrow \beta_{ECal}^{EM} = \beta_{ECal}^{EM0} \times \frac{E_{MC}}{E_{Fit}} \tag{1.3}$$

An example distribution and fit used for calibration of the nominal ILD detector model can be found in figure 1.10. The procedure is then repeated using the updated β_{ECal}^{EM} until E_{Fit} falls within a chosen tolerance, which in this case is that $|E_{\text{Fit}} - E_{\text{MC}}| < E_{\text{MC}} \times 0.5\%$. The binning for the fitted histogram is chosen such that the bin width is equal to the desired tolerance on E_{Fit} e.g. $E_{\text{MC}} \times 0.5\% = 0.05$ GeV. This tolerance is tighter than that applied for the digitisation as it is the PFO energies that are used in downstream analyses and therefore require high precision.

The electromagnetic scale in the HCal, β_{HCal}^{EM}, is chosen to be equal to the hadronic scale in the HCal, β_{HCal}^{Had}. The details of the determination of β_{HCal}^{Had} can be found in section 1.1.6. For ILC like energies, β_{HCal}^{EM} is not a critical parameter in the reconstruction as photons are largely contained within the ECal meaning little to no electromagnetic energy is measured in the HCal.
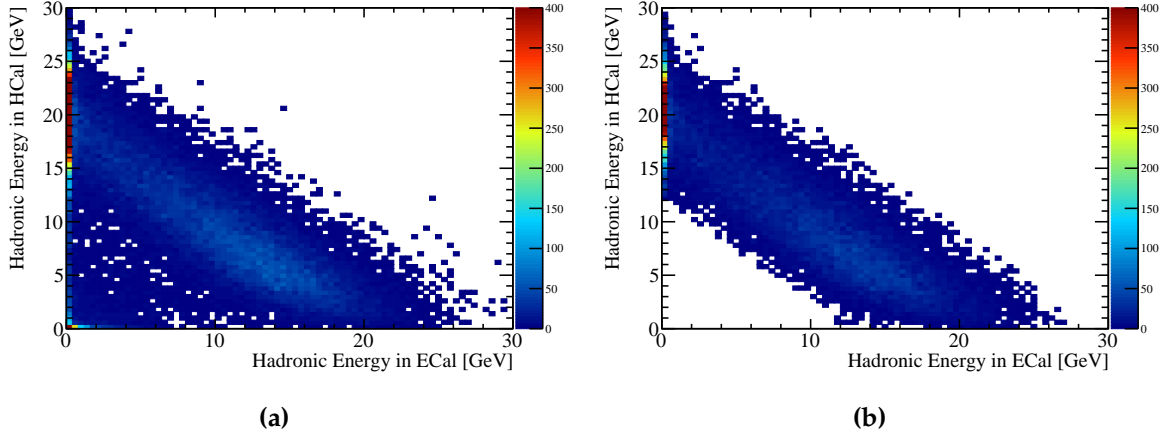
Figure 1.11: The distribution of hadronic energy measured in the ECal and HCal for 20 GeV K_L^0 events (a) without selection cuts and (b) with selection cuts.

Hadronic scale setting

β_{ECal}^{Had} is important to detector performance as a non-negligible amount of hadronic energy is recorded in the ECal. However, as the ECal contains $\approx 1\lambda_I$, the hadronic scale in the ECal cannot be independently as it is unfeasible to create a large sample of 20 GeV K_L^0 events that are fully contained within it. Therefore, the hadronic scale is set in the ECal and HCal simultaneously.

Cuts are applied to select K_L^0 that are appropriate to use for the hadronic scale calibration. The last layer in which energy is deposited in the HCal must not occur in the back 10 % of the HCal to ensure that the event does not suffer from leakage of energy out of the calorimeter. A single neutral hadron must be reconstructed to veto events with reconstruction failures. Finally, the total hadronic energy measured in ECal and HCal, $E_{ECal}^{Had} + E_{HCal}^{Had}$, must fall within 3 σ of the desired hadronic energy distribution, $E_{ECal}^{Had} + E_{HCal}^{Had} = 20\text{GeV} - m_{K_L^0} = E_K$. σ is defined to be $55\% \times \sqrt{E} = 2.46\text{GeV}$ for 20 GeV K_L^0. This definition of sigma is the nominal energy resolution for neutral hadrons in the ILD HCal. This cut ensures that when fitting the two dimensional distribution of hadronic energy measured in the ECal and HCal the outliers do not skew the fit. Once again the target reconstructed energy for this sample is the kinetic energy and not the total energy of the K_L^0 for the reasons outlines in section 1.1.4. The impact of cuts are illustrated in figure 1.11.
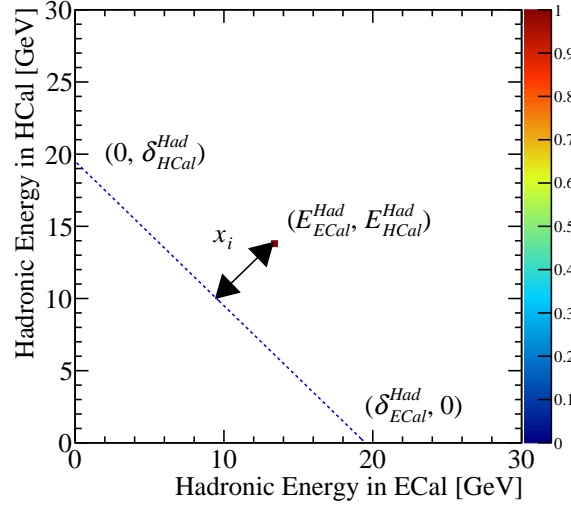
Figure 1.12: An example showing the definition of x_i, the variable used for the calculation of χ^2 for the setting of the hadronic energy scale. For an event that has been measured with hadronic energy E_{ECal}^{Had} in the ECal and E_{HCal}^{Had} in the HCal, the geometric interpretation of x_i is shown. The blue dotted line is defined as $E_{HCal}^{Had} = \delta_{HCal}^{Had} - E_{ECal}^{Had} \frac{\delta_{HCal}^{Had}}{\delta_{ECal}^{Had}}$.

The calibration procedure is iterative and begins by assuming trial values, β_{ECal}^{Had0} and β_{ECal}^{Had0}, for the hadronic scale calibration factors β_{ECal}^{Had} and β_{ECal}^{Had}. The 20 GeV K_L^0 events are then simulated and reconstructed. Following this a linear fit to the distribution of E_{ECal}^{Had} against E_{HCal}^{Had} for 20 GeV K_L^0 events passing the selection cuts is applied. The fit is performed by minimising χ^2, which is defined as

$$\chi^2(\delta_{ECal}^{Had}, \delta_{HCal}^{Had}) = \sum_i \frac{x_i}{\sigma_{x_i}} \tag{1.4}$$

where x_i is the perpendicular distance from E_{ECal}^{Had} and E_{HCal}^{Had} for event i to the line $E_{HCal}^{Had} = \delta_{HCal}^{Had} - E_{ECal}^{Had} \frac{\delta_{HCal}^{Had}}{\delta_{ECal}^{Had}}$. The definition of x_i is given in equation 1.5, but best illustrated by considering figure 1.12. σ_{x_i} is the uncertainty on x_i, which is calculated by propagating the uncertainties on E_{ECal}^{Had} and E_{HCal}^{Had}, which are assumed to be $\sigma_{E_{E/HCal}^{Had}} = 55\% \times \sqrt{E_{E/HCal}^{Had}}$, into the expression for x_i. The result of this propagation of errors is given in equation 1.6. The sum runs over all events, i, passing the selection cuts.

$$x_i = \frac{E_{HCal}^{Had} \delta_{ECal}^{Had} + E_{ECal}^{Had} \delta_{HCal}^{Had} - \delta_{ECal}^{Had} \delta_{HCal}^{Had}}{\sqrt{(\delta_{ECal}^{Had})^2 + (\delta_{HCal}^{Had})^2}}$$ (1.5)

$$\sigma_i = \frac{(\sigma_{E_{HCal}^{Had}} \delta_{ECal}^{Had})^2 + (\sigma_{E_{ECal}^{Had}} \delta_{HCal}^{Had})^2}{\sqrt{(\delta_{ECal}^{Had})^2 + (\delta_{HCal}^{Had})^2}}$$ (1.6)

The minimisation steps across a range of δ_{ECal}^{Had} and δ_{HCal}^{Had} centred about the ideal value of $20\,\text{GeV} - m_{K_L^0}$ in search for the minimum χ^2. Once the minima in χ^2 is found the trial calibration factors β_{ECal}^{Had0} and β_{ECal}^{Had0} are rescaled to correct for any deviation from the desired fit as follows

$$\beta_{ECal}^{Had0} \rightarrow \beta_{ECal}^{Had} = \beta_{ECal}^{Had0} \times \frac{E_K}{\Delta_{ECal}^{Had}}$$ (1.7)

$$\beta_{HCal}^{Had0} \rightarrow \beta_{HCal}^{Had} = \beta_{HCal}^{Had0} \times \frac{E_K}{\Delta_{HCal}^{Had}}$$ (1.8)

where Δ_{ECal}^{Had} and Δ_{ECal}^{Had} are the values of δ_{ECal}^{Had} and δ_{ECal}^{Had} giving the minimum χ^2. The step sizes used for minimising χ^2 with respect to δ_{ECal}^{Had} and δ_{ECal}^{Had} is chosen such that a single step corresponds to the target final tolerance on δ^{Had} i.e. $|\delta_{E/HCal}^{Had} - E_{\text{MC}}| < E_{\text{MC}} \times 0.5\% \approx 0.1\text{GeV}$.

This procedure is then repeated until Δ_{ECal}^{Had} and Δ_{ECal}^{Had} both fall within a given tolerance, which in this case it taken to be $|\Delta_{E/HCal}^{Had} - E_{\text{MC}}| < E_{\text{MC}} \times 0.5\% \approx 0.1\text{GeV}$

1.1.7 Calibration step ordering

The calibration procedure has to be run in the following order so that the building blocks used for each stage of the calibration procedure can be assumed to be correctly calibrated:

- MIP Scale setting in the digitiser as described in section 1.1.5.

- Calibration of digitisation of calorimeter hits in the ECal and HCal as described in section 1.1.4.

- MIP Scale setting in PandoraPFA as described in section 1.1.5.

- Electromagnetic and hadronic scale settings in PandoraPFA as described in section 1.1.6.

1.1.8 Retraining photon likelihood data

PandoraPFA uses likelihood data in the identification of γs. Data related to the topology and energy of electromagnetic showers and the wider event environment is used to determine whether a given shower is likely to be a γ. The likelihood data is trained using off-shell mass Z boson (Z') events at 500 GeV that decay into light quarks (u, d, s). To ensure correct calibration for these simulations it is necessary to retrain this data when varying the ECal. As photons are contained largely within the ECal at the energies being considered and the likelihood data only uses measurements made in the ECal, it is only necessary to retrain this data when varying the ECal.

As this data uses post digitisation hits it is important to ensure that a well calibrated detector is used to retrain this information. To ensure accurate calibration it is necessary to run the calibration procedure, as described in section 1.1.7, before retaining the data. However, as the reconstruction uses likelihood data the calibration procedure has to be run twice. Initially the calibration procedure is performed where PandoraPFA is run without the inclusion of this likelihood data, using PandoraSettingsMuon.xml. Then the likelihood data is retrained using the first set of calibration factors and then this likelihood data is used in the second run of the calibration procedure.

In the optimisation studies presented in chapter OPTIMISATION STUDIES this procedure was followed whenever the ECal was modified so that optimal detector performance was achieved.

1.2 Timing Cuts

The ILC and CLIC will operate using a trigger-less readout approach whereby the recorded data for each sub-detector is readout between bunches. The train structure

for the ILC and CLIC at maximum operating energy is shown in table 1.1. Event selection proceeds through the application of a software trigger. This involves the identification of hard interactions, prior to full event reconstruction, and only putting data into the event reconstruction if it is measured within a chosen time window about this interaction. Timing cuts placed on the calorimeter hits are corrected for straight time-of-flight to the IP. This ensures that the amount of time particle showers have to develop in the calorimeters is independent of their position in the calorimeters. As the size of the time window around the identified hard interaction changes the amount of time particle showers have to develop within the calorimeter, the choice of time window will affect the performance of the detector.

	ILC 500 GeV	CLIC 3 TeV
Electrons per bunch	2.0	0.37
Bunches per train	2810	312
Train repetition rate [Hz]	5	50
Bunch separation [ns]	308	0.5

Table 1.1: The train structure for 500 GeV ILC and 3 TeV CLIC. CITE

For all detector models considered in this analysis the calibration procedure was reapplied. This means that the mean of the reconstructed energy distribution is invariant to changes in the calorimeter timing window as the calibration procedure fixes the relevant distributions to the MC energy of the calibration sample.

The energy resolution for 100 GeV γ and 50 GeV K_L^0 events using the nominal ILD detector model as a function of the timing window used for the calorimeter hits is shown in figure 1.13. The timing cuts make little difference in the case of the γ events as they produce electromagnetic showers that propagate rapidly. However, there is a significant decrease in the energy resolution for the neutral hadrons as these showers develop much more slowly. Truncating the measurement of the hadronic showers by having a small time window leads to a reduction in sampling of the shower, as those hits passing the time window are no longer counted, and a broadening of the reconstructed energy distribution.

The degradation in neutral hadron energy resolution with decreasing calorimeter time window affects the jet energy resolution, which can be seen in figure **??**. The sole exception to this is the 250 GeV jets for the 100 ns time window whereby the
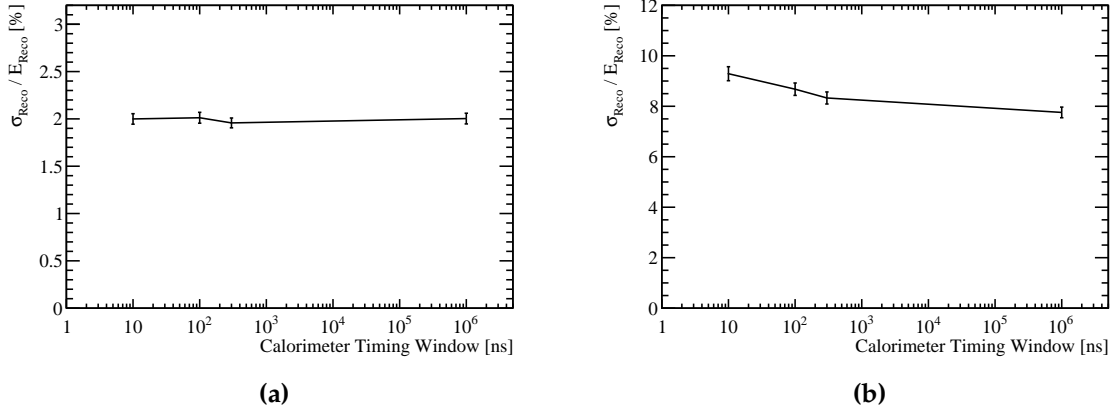
Figure 1.13: The energy resolution as a function of calorimeter timing window for (a) 100 GeV γ events and (b) 50 GeV K_L^0 events for the nominal ILD detector model.
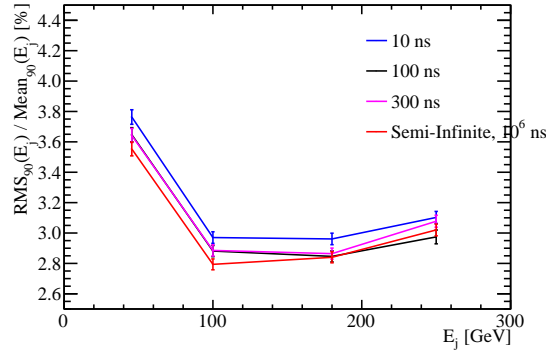

Figure 1.14: The jet energy resolution as a function of jet energy for various calorimeter timing cuts. The nominal ILD detector model was used for this study.

jet energy resolution is slightly better than both the 300 ns and semi-infinite time window. As the magnitude of the changes to the jet energy resolution when varying the time window size are small in comparison to the absolute jet energy resolutions, this exception will most likely be due to a fluctuation in either the event sample used or in the reapplication of the calibration procedure.

The time window applied on the calorimeter hits, which are apart of the software trigger that will be used at the linear collider, does affect both the neutral hadron and jet energy resolutions with a larger timing window leading to better resolutions. When reducing the time window to the order of 10 ns a large number of hadronic showers will not have fully developed in the calorimeters in this time, which leads to a poor sampling of the hadronic shower. However, from 100 ns onwards there is relatively

little gain to be made with further increases to the time window, as the majority of hadronic showers will have fully developed by this time.

For the optimisation studies presented in section OPT STUDIES a 100 ns timing window was applied to all models considered. As the choice of timing window has yet to be finalised for the linear collider this value was chosen as it represents something that can be achieved using the readout technology options presently available CITE. Furthermore, it adds further realism to the detector simulation in comparison to omitting this effects of the calorimeter time window. The categorisation of changes to the detector performance when varying the calorimeter timing window presented here can be used to discern the impact of changing the timing window used for the optimisation studies at a later date if so desired.

1.3 HCal Cell Truncation

A powerful tool that was used by PandoraPFA in performing the reconstruction was the application of a truncation of the maximum amount of energy that can be recorded in a calorimeter cell in the HCal. The purpose of this truncation is to eliminate the effect of spuriously high energy calorimeter cells that would skew the reconstruction. The origin of these high energy cells is twofold: showering particles may be moving within the plane of the active material, which can lead to an overestimation of the deposited energy if the shower is not sufficiently uniform across a calorimeter cell, and Landau fluctuations [4], which originate from high energy knock-on electrons appearing within particle showers [1]. These effects are only relevant in hadronic showers as electromagnetic showers begin showering almost immediately within the ECal, meaning they are unlikely to be directed within the plane of the active material of the calorimeter, and the Landau distribution describes the energy loss for charged particles meaning showering photons do not have Landau fluctuations. While e^- and e^+ do have from Landau fluctuations their energy, in the particle flow paradigm, arises from the curvature of the track and not the calorimeter hits, so correcting calorimeter hits to correct for Landau fluctuations is not key. The truncation applied to counter these effects is only applied in the HCal as, at the energies being considered, it primarily measures hadronic showers.

A great deal of care has to be given to choice of the truncation applied to the HCal cell energy as to not truncate cell energies that are from typical hadronic shower
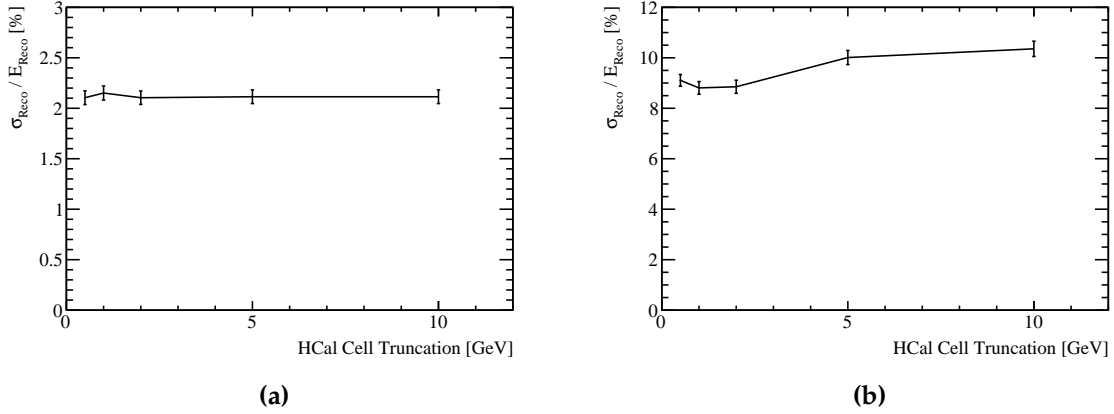
(a)   **(b)**

Figure 1.15: The energy resolution as a function of HCal cell truncation for (a) 100 GeV γ events and (b) 50 GeV K_L^0 events for the nominal ILD detector model.

development i.e. particle showers without spuriously high energy calorimeter cells. This can be illustrated by considering the single particle energy resolutions as a function of the HCal cell truncation that are shown in figure 1.15. The photon energy resolution is invariant to changes in the cell truncation as no photon energy is recorded in the HCal, while the energy resolution for the neutral hadrons improves has a minima around 1 GeV. This indicates that a 1 GeV truncation is sufficient for dealing with the Landau fluctuations and showering particles travelling within the active material of the calorimeter cell. For cell truncations greater than 1 GeV the resolution degrades as the spuriously high energy cells are not accounted, while truncations below 1 GeV truncate the calorimeter cells from typical hadronic shower development leading to inaccurate sampling of the shower and a degradation in the energy resolution.

Once again, this effect propagates into the jet energy resolutions as shown by figure 1.16. The trends in this plot are complex as the optimal cell truncation varies with the jet energy. At low energies a 0.5 GeV truncation gives the best performance, however, when the jet energies reach $\approx$ 180 GeV a 1-2 GeV truncation giving the best performance. This is to be expected based on the Landau fluctuations. The Landau distribution is essentially a Gaussian with a high energy tail and as the jet energy increases the mean of the Gaussian increases, which changes the position of the high energy tail. Therefore, the definition of the cells requiring truncating changes as a function of jet energy and a procedure with more degrees of freedom than a single truncation value is required to properly account for this. As particle showers grow in size with increasing incident particle energy it is expected that the problem of particles
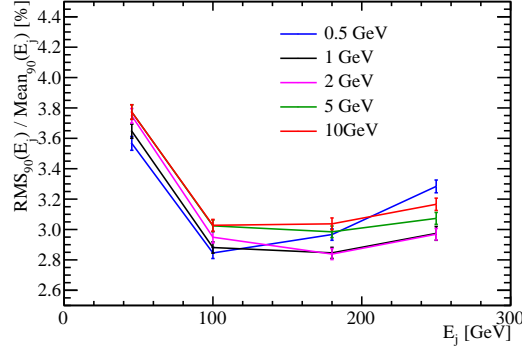
Figure 1.16: The jet energy resolution as a function of jet energy for various hadronic cell truncations. The results shown use the nominal ILD detector model.

travelling within the active material plane, and not uniformly across the rest of the cell, should not change significantly with changes to the jet energy. When examining the breakdown of the jet energy resolutions into the intrinsic energy resolution and the confusion it was noted that the cell truncation improved both the intrinsic energy resolution and the confusion. This indicates that the pattern recognition performed by PandoraPFA benefits from the absence of these spuriously high energy cells, which if not omitted skew energy comparisons made in the reconstruction. Any skewed energy comparisons in the reconstruction leads to inaccurate association of calorimeter hits to charged particle tracks, which in turn causes double counting and/or omission of energy deposits in the calorimeters leading to a degradation of the energy resolution.

While it is challenging to determine the optimal performance for a given detector model it is clear that applying a truncation can produce significant improvement in detector performance. Therefore, for the optimisation studies presented in section OPT STUDIES, the performance of each detector model is determined using a range of HCal cell truncations and the optimal resolutions quoted. The HCal cell truncations considered in the optimisation were 0.5, 0.75, 1, 1.5, 2, 5, 10, and 10^6 GeV (semi-infinite). It was found that for all studies apart from the HCal cell size, the optimal performance was achieved using a 1 GeV truncation. For the HCal cell size study the optimal performance for the 10, 20, 30, 40, 50 and 100 mm HCal cell size detector models was achieved using a 0.5, 0.75, 1, 1.5, 2 and 5 GeV truncation. This optimisation has a significant impact on detector optimisation, which can be seen by comparing the jet energy resolutions using the optimised cell truncation and a uniform 1 GeV truncation, as found in 1.17. Without this optimisation of cell truncation the significance of the
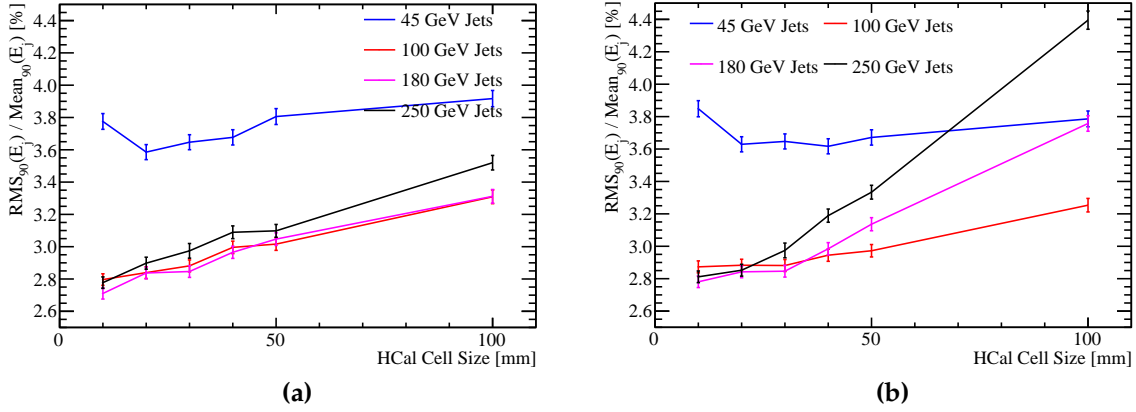
Figure 1.17: The jet energy resolution as a function of HCal cell size using (a) an optimised HCal cell truncation and (b) a 1 GeV truncation.

HCal cell size is overinflated and could have led to a misinformed detector design choice.

1.4 Software Compensation

As discussed in chapter CALORIMETERS CHAPTER, the response of a calorimeter to a hadronic shower is different to that of an electromagnetic showers. The particle shower produced by a hadron when passing through a calorimeter has two components; an electromagnetic shower core, which originates from the production and decay of π^0, and a hadronic shower component originating from all other interacting and decaying hadrons. Hadronic showers have an "invisible" energy component, due to various factors such as neutrons stopping within the calorimeter and nuclear binding energy losses. This component leads to a reduced response from a calorimeter to a hadronic shower in comparison to an electromagnetic shower of the same initial energy. An event display showing the high energy density electromagnetic core of a hadronic cluster for a 500 GeV Z→uds di-jet event can be found in figure 1.18. This different calorimetric response will lead to a degradation in the energy resolution of a detector if not properly compensated for.

This compensation can be applied either at the hardware level, whereby a calorimeter is made to be intrinsically compensating, or at the software level, whereby hadronic showers are identified and their energy estimators modified.

(a) **(b)**
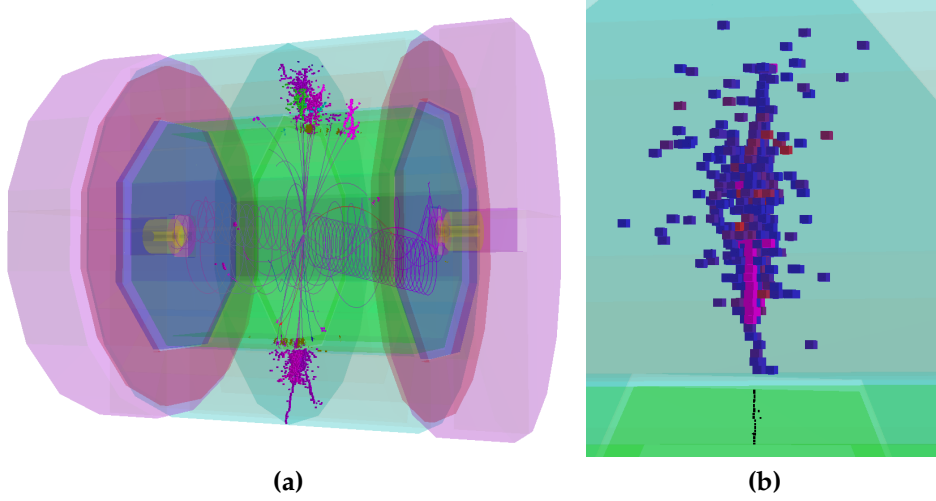
Figure 1.18: An event display for a 500 GeV Z→uds di-jet event reconstructed using the nominal ILD detector. (a) shows the full event environment. (b) shows a single hadronic cluster from the same event where shading indicates the energy density in the HCal. High energy density cells are coloured red, while lower energy density cells are coloured blue. All ECal hits are shaded black. The high energy density electromagnetic core of the selected hadronic cluster is clearly visible.

An example of hardware compensation would be the ZEUS calorimeter [3] that was constructed using uranium as the absorber material. In response to neutral hadrons the uranium underwent fission producing extra energy, which raises the hadronic response of the calorimeter. The amount of uranium was carefully chosen to achieve a fully compensating calorimeter response i.e. identical calorimeter response to electromagnetic and hadronic showers. While hardware compensation is possible for the linear collider calorimeters, restrictions on calorimeter construction and the use of a large amount of radioactive material are highly undesirable.

The high granularity calorimeters and sophisticated pattern recognition software used at the linear collider give excellent resolution on individual particle showers. This resolution means that software compensation can be applied at the linear collider to a greater level of detail that has ever been considered for previous experiments.

1.4.1 Application

The software compensation technique applied in this study involves reweighing HCal hits based on their energy density and the energy of the cluster those hits

belong to. This is employed in the PandoraPFA algorithm in the form of an energy correction function, which in effect means whenever the energy of a cluster of hits is considered in PandoraPFA the software compensated energy is used. Applying software compensation in this way benefits the detector energy resolution in two ways; firstly the intrinsic energy resolution of the detector improves and secondly the confusion from incorrect association of track to calorimeter hits is reduced. Lowering the confusion benefits the energy resolution as it decreases the number calorimeter hits where the energy of the hit is effectively double counted, if charged particle hits are not associated to a track, or not counted at all, if neutral hadron hits are associated to a track.

Software compensation is applied to clusters of calorimeter hits, as opposed to being applied directly to PFOs, so that the more precise energy estimators can be used during the reconstruction. For a cluster of calorimeter hits, with an initial energy of E_{Raw} calculated by summing the calorimeter hit energies, the software compensated cluster energy, $E_{SoftComp}$, is given by

$$E_{SoftComp} = E_{ECal} + \sum_i E_i \times \omega_i(E_{\text{Raw}}, \rho_i) + E_{\text{Muon Chamber}} \tag{1.9}$$

where E_{ECal} is the sum of the calorimeter hit energies measured in the ECal, E_i and ρ_i are the energy and energy density of HCal hit i respectively, ω_i is the software compensation weight applied to hit i, $E_{\text{Muon Chamber}}$ is the cluster energy recorded in the muon chamber and the sum runs over all hits, i, in the HCal. The weight function $\omega_i(E_{\text{Raw}}, \rho_i)$ is defined as

$$\omega_i(E_{\text{Raw}}, \rho_i) = p_1(E_{\text{Raw}}) \times exp(p_2(E_{\text{Raw}}) \times \rho_i) + p_3(E_{\text{Raw}}) \tag{1.10}$$

$$p_1 = p_{11} + p_{12} \times E_{\text{Raw}} + p_{13} \times E_{\text{Raw}}^2 \tag{1.11}$$

$$p_2 = p_{21} + p_{22} \times E_{\text{Raw}} + p_{23} \times E_{\text{Raw}}^2 \tag{1.12}$$

$$p_3 = \frac{p_{31}}{p_{32} + exp(p_{33} \times E_{\text{Raw}})} \tag{1.13}$$

where ρ_i is the energy density of the cell for hit i and p_{ij} are trained parameters. The parameters p_{ij} are determined by performing a χ^2 fit of $E_{SoftComp}$ to the MC energy
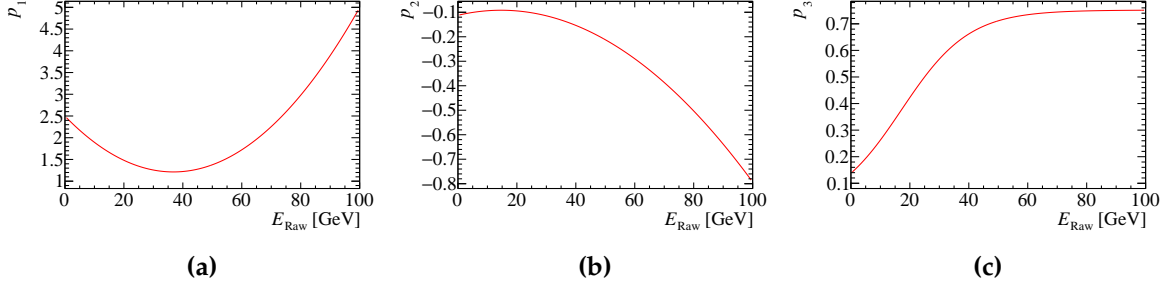
Figure 1.19
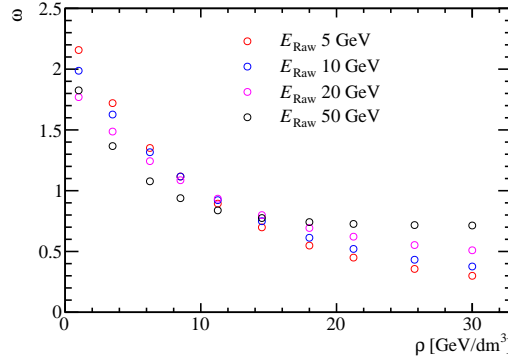
Figure 1.20: The software compensation weight applied to a calorimeter hit as a function of calorimeter hit energy density for various cluster energies.

for samples of K_L^0 ranging from 10 to 100 GeV in steps of 10 GeV. Using the fitted parameters, p_1, p_2 and p_3 as a function of E_{Raw} and $\omega(E_{\text{Raw}}, \rho)$ as a function of ρ for various E_{Raw} are shown in figure 1.19 and 1.20 respectively.

As software compensation only modifies the energy of HCal hits there is freedom to apply further energy corrections to the ECal hits. Application of the Clean Clusters energy correction logic, described in section **??**, to the ECal hits alongside software compensation gave further improvements to the jet energy resolution. Therefore, as standard the application of software compensation within PandoraPFA implicitly involves the application of the Clean Clusters logic to the ECal hits.

As software compensation is trained using a maximum K_L^0 energy of 100 GeV, software compensation is not applied to clusters where $E_{\text{Raw}} > 100$ GeV. While it would be possible to modify the energy range of the training sample to go to higher energies, hadronic clusters with energy greater than 100 GeV will be rare at the ILC as the maximum running energy of 500 GeV.

1.4.2 Results

Legacy Energy Corrections

Before examining the impact of software compensation on detector performance is it necessary to address the 'legacy' energy corrections that are used as default in PandoraPFA. There are three energy correction that PandoraPFA has used prior to the development of software compensation, which are:

- **HCal cell truncation**, the details of which can be found in section 1.3.

- **Clean Clusters**. This algorithm checks to see whether the energy measured within a calorimeter hit is anomalously high. This is defined as a cell where the energy in that cell is more than 10% of the energy of the cluster that the cell has been associated to. If a cell is deemed to have anomalously high energy and this energy is above a threshold, 0.5 GeV, the cell energy used by PandoraPFA is modified. The updated cell energy is taken as the average cell energy in the calorimeter layers immediately before and after the layer containing the high energy cell.

- **Scale Hot Hadrons**. This algorithm calculates the average number of MIP equivalent particles passing through each calorimeter cell in a cluster. If this number is larger than a given value, default 15 MIPs per hit, the cluster energy is rescaled to give a lower average number of MIPs per hit, default low value is 5 MIPs per hit.

Each of these energy corrections help to deal with the effects of spuriously high energy cells the origin of which is described in section 1.3. However, the algorithms are simplistic and software compensation is expected to give far better results than these 'legacy' options.

The optimisation studies presented in section OPTIMISATION STUDIES use all three of these legacy options simultaneously, which was the default behaviour for PandoraPFA when the studies were undertaken. The new default behaviour in PandoraPFA is to use software compensation.
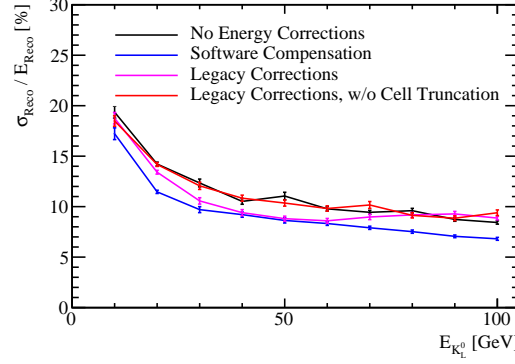
Figure 1.21: The energy resolution as a function of the MC energy for single K_L^0 events using various energy correction settings. The detector model used was the nominal ILD detector model.

Energy Resolution

?? The energy resolution as a function of the MC energy for single K_L^0 events is shown in figure **??** using various energy correction settings.

When comparing the energy resolution given by software compensation to that obtained using no energy corrections it can be seen that software compensation offers a gain of $\approx 2\%$ in energy resolution across all energies considered in comparison. The uniformity of this improvement is encouraging, indicating software compensation has been successfully trained across this energy range.

Comparing the performance of software compensation to the legacy corrections it can be seen that software compensation gives a better energy resolution across almost the entire range of energies considered. The only exception to this is around $E_{K_L^0} \approx 50$ where the performance of software compensation and the legacy corrections are comparable. By removing the cell truncation from the legacy options it is clear that the changes in energy resolution when using the legacy options are being driven by the cell truncation. This makes clear the trend observed using the legacy corrections as at low K_L^0 energies very few cells are affected by the truncation so the performance is comparable to not using any energy corrections. While at high K_L^0 the truncation is too aggressive and removes energy from cells that are not spuriously high leading to a worsening energy resolution. Between these two extremes, $E_{K_L^0} \approx 50$, the truncation works ideally and improvement in energy resolution using the legacy corrections is the largest.
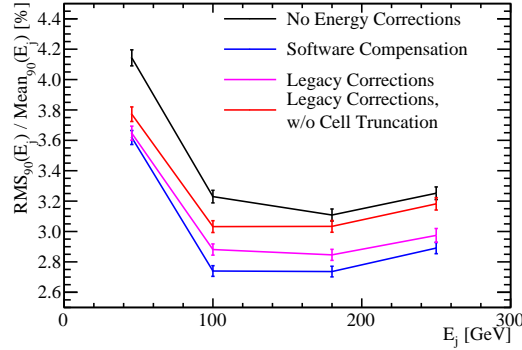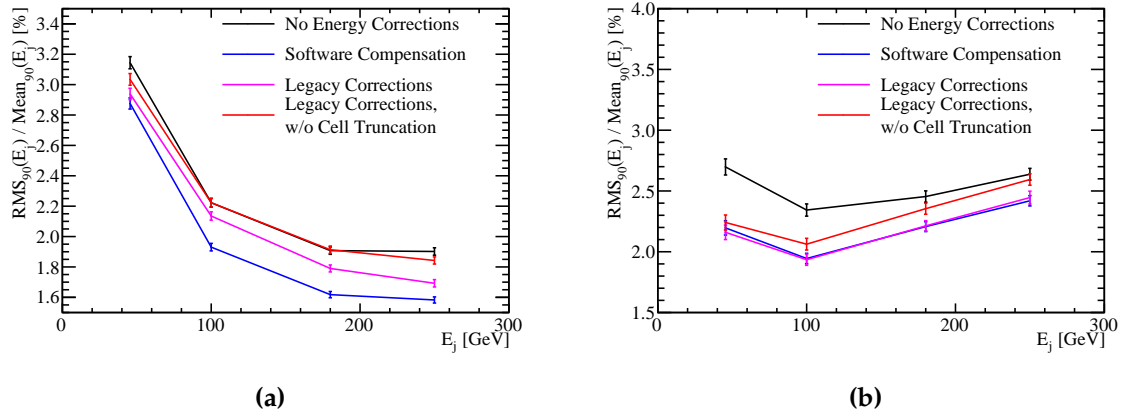
Figure 1.22: The jet energy resolution as a function of the jet energy for a variety of different energy correction options. These results were produced for the nominal ILD detector model.

Jet Energy Resolution

The improvements in the intrinsic energy resolution of the detector that have been observed when using software compensation will propagate into the reconstruction of jets. These effects are illustrated by examining the jet energy resolution as a function of jet energy, which is shown in figure 1.22. Again it is clear that software compensation is extremely beneficial to the detector performance.

Software compensation gives a significant reduction in the jet energy resolution in comparison to using no energy corrections. It also reduces the jet energy resolution in comparison to using the legacy corrections.

Further light can be shed on these trends by examining the contribution to the jet energy resolutions from the intrinsic energy resolution and the pattern recognition confusion, which are shown in figure 1.23. The intrinsic energy resolution contribution that software compensation is significantly better than all other energy corrections options, which is to be expected from the energy resolution studies presented in section **??**. Unlike the single particle study there is no jet energy where the cell truncation matches the performance obtained using software compensation. This is due to the fact that the energy resolution when using the cell truncation is only comparable to the energy resolution using software compensation for a narrow range of hadronic cluster energies. As the jet contains a broad spectrum of hadronic cluster energies the performance obtained when using the cell truncation will always be worse than when using software compensation. When comparing the jet energy resolution for the

legacy corrections is again apparent that the term driving the jet energy resolution is the cell truncation.

The change in the confusion contribution to the jet energy resolution when using software compensation and the legacy corrections is almost identical. This indicates that the improvement seen in the jet energy resolution in figure 1.22 when using software compensation instead of the legacy corrections is being driven by improvements to the intrinsic energy resolution.

It was observed that at low jet energies the Clean Clusters and Scale Hot Hadrons energy corrections are beneficial at reducing the confusion contribution and the cell truncation is largely redundant while for high energy jets the reverse is true. As the use of the Clean Clusters and Scale Hot Hadrons energy corrections does not change the intrinsic energy resolution of the detector it is apparently that these energy corrections are purposed to account for failures in the pattern recognition that occurs largely at low jet energies. On the other had the cell truncation and software compensation techniques aim to improve the energy resolution of the hadronic clusters, which has a further knock-on effect of improving the track cluster associations made in the pattern recognition. These corrections work across all energy ranges, but have a greater impact at high energies.

By extracting the Clean Clusters logic, which is the driving term reducing the confusion contribution to the jet energy resolution for low energy jets, and embedding it within the software compensation technique it is possible to achieve exceptional jet energy resolution that will extend the physics reach of the linear collider detector.

Figure 1.23: The contributions to the jet energy resolution as a function of the jet energy for a variety of different energy correction options. (a) is the intrinsic energy resolution of the detector and (b) is the total confusion term. The quadrature sum of both yields the standard reconstruction performance. These results were produced for the nominal ILD detector model.

Colophon

This thesis was made in LaTeX 2_ε using the "hepthesis" class [2].

Bibliography

[1] H. Bichsel, Donald E. Groom, and S. R. Klein. Passage of particles through matter. 2004.

[2] Andy Buckley. The hepthesis LaTeX class.

[3] M. Derrick, D. Gacek, N. Hill, B. Musgrave, R. Noland, E. Petereit, J. Repond, R. Stanek, and K. Sugano. Design and construction of the ZEUS barrel calorimeter. *Nucl. Instrum. Meth.*, A309:77–100, 1991.

[4] L. Landau. On the energy loss of fast particles by ionization. *J. Phys.(USSR)*, 8:201–205, 1944.

List of figures

List of tables