Energy Resolution Parameterisation for Fast Simulation

Liang Dai,
tutored by Mikael Berggren and Jenny List

DESY summer student program 2010, FLC group

1 Introduction

The central task of a detector at a collider is to measure particles coming out of a scattering event that carry physics information. However, we are faced with different strategies for measurement of charged particles and neutral particles. For charged particles one makes use of ionisation and measure the momentum from curved track, which is left in the tracking system when a magnetic field is applied. For neutral particles one turns to calorimeter by measuring energy deposits as a result of particle shower. More importantly, calorimeter provides plentiful information because both charged and neutral particles leave signals inside it. A typical calorimetry system consists of a hadronic calorimeter (HCAL) and a much finer electromagnetic colorimeter (ECAL).

In a realistic detector design project, performance of detection, which tells you whether physics goals can be fulfilled, will be tested by computer simulation in advance. In a full simulation, a Monte Carlo (MC) event generator first generates a large number of hard scattering events, which reveal physics at a high energy scale, with decays of short-lifetime particles in the final state also simulated. Further interaction with all layers of the detector is simulated by a well-developed detector simulation package, where processes of bremsstrahlung, pair production, electromagetic and hadronic shower give rise to even more particles and complicate the final state. Then a reconstruction package rebuilds the final state by making use of information in the output of simulation which mimics a realistic colliding experiment, and finally detection capabilities are benchmarked. Currently, particle flow calorimetry and the PandoraPFA algorithm is prepared for ILC detectors [1], tested in a full simulation of calorimeter named Mokka based on GEANT4 [2].

The problem with a full simulation program lies in that it consumes too much CPU runtime if large statistics is needed. This makes further opimisation of detector very low in efficiency. Therefore fast simulation is developed, in which complicated interaction between particles and detector materials is depicted effectively by certain parameters. The task is to choose good parameterisation and to extract parameters from full simulation data. In order to achieve this, one must well understand the physics inside a detector that limits the energy resolution and hence the precision of the detector.

In this report on the summer student project, we focus on the energy resolution of calorimeter and investigate proper parameter values for different particle types. We mainly focus on photons which stop in ECAL, and both charged and neutral hadrons which trigger hadronic showers in HCAL. Our aim is to disentangle various processes that confuse reconstruction, and then properly account for them for fast simulation.

2 Parameterisation and confusion

Since detection is not ideal, one should expect some deviation in the measured energy of reconstructed particle from the true energy in MC. This deviation is not gaussian distributed because it depends on energy itself. One thing characteristic of calorimeter is that better relative precision is achieved for high energy particles. A typical dependence for calorimetry is of the form

$$\frac{\sigma_E}{E} = \frac{\alpha}{\sqrt{E}} \oplus \beta \tag{1}$$

where coefficient α accounts for intrinsic calorimetric capability due to traditional approach to clustering. Constant β, on the other hand, encompasses a variety of effects that limit reconstruction. By looking at a narrow energy interval we can obtain gaussian-like distribution and extract parameters by fitting. But it turns out that this naive approach does not lead to a smooth curve according to Eq. (1), but more complicated shape. We believe some complication related to energy is in action.

For very low energy particles, the calorimeter can fail to detect them at all. Moreover, due to the finite depth of calorimeter leakage can happen when some fraction of energy escapes deposition in calorimeter. However, an even more important effect is confusion, when clusters overlap with each other. Unlike traditional clustering, track information is used to improve clustering in particle flow calorimetry strategy; nevertheless a number of cases can lead to error from confusion. For example, a high energy neutral particle can split up when hitting calorimeter and leave two clusters from which two particles are reconstructed because of lack of track momentum to match. Occasionally one cluster is merged with another nearby cluster energy measurement is wrong. In addition, in the bordering region between barrel and encaps, with cables and supporting structures, incorrect clustering is more common at geometric edges. These effects can dramatically complicate the parameterisation on energy resolution and we want to separate these cases and give estimate of how often they can occur.

Measurement of α and β has been down using test beam where confusion is not present. To be more realistic we will use simulation output in this project and account for the effect of confusion, which must be considered in fast simulation.

3 Photons in calorimeter

Detection of photons are relatively clean, not only because a photon always stops in ECAL and is free from mixing up with hadronic showers, but also due to the fact that an electromagnetic shower is more compact than a hadronic one, which reduces confusion. We investigate simulated photons with energy $0GeV < E_\gamma < 50GeV$ produced in an electron-position collision at a beam energy of $250GeV$. We find that resolution for low energy regime $0GeV < E_\gamma < 20GeV$ agrees well with (1) , but for high energies resolution is larger than predicted by the parameterisation. As we discovered, this is due to the process in which a photon splits up as soon as it hits ECAL and leaves several clusters. High energy photons are more likely to split up as shown in Figure 1 and we underestimate their energies.

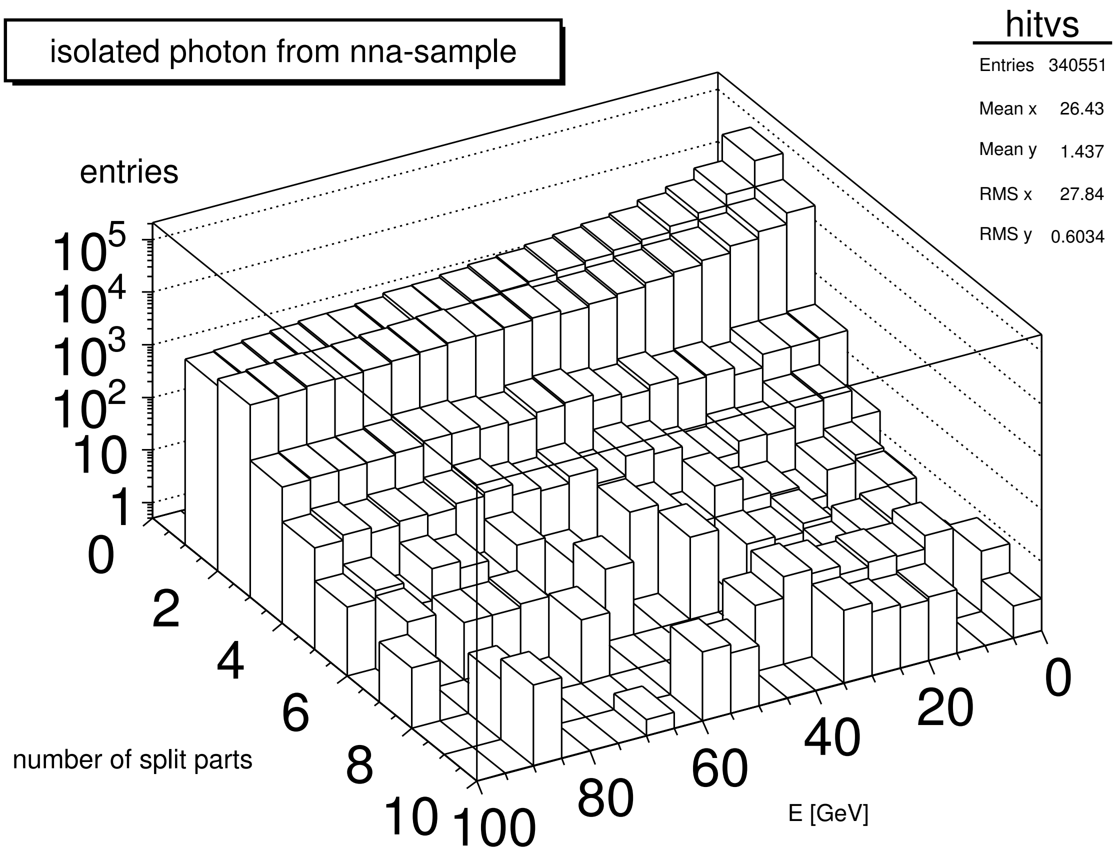

Figure 1: Isolated photons from $\nu\nu\gamma$ sample distributed over energy and number of split parts, with energy $0 < E < 100GeV$. Most split photons split into two parts and high energy photons split more often. (Note that entries are counted in logarithm scale.)

To improve high energy regime, we adopt a merging technique in which we try to combined as many clusters as possible that come from the same photon. In realistic reconstruction, lack of information that simulation can provide means we can not select clusters coming from the same photon with 100% accuracy. To solve this problem, we first define cone distance between two clusters

$$\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} \tag{2}$$

where $\Delta\phi$ is difference in azimuthal angle and $\Delta\eta$ difference in pseudorapidity, with respect to the beam axis. This cone distance measures how close to each other two ECAL clusters are and small cone distance gives good credibility that both originate from a single photon. Previous studies by our group have shown that a reasonable $\Delta R_{cut} = 0.04$ can be chosen as a cut for re-grouping parts. We start with the most energetic ECAL cluster and combine all others within the cut and thus take the combined energy as measured energy. We test this merging technique with di-neutrino plus photon sample, in which only one isolated, high energy photon is detected so that merging is convenient, shown in Figure 2. In principle, this technique can also be used in events where isolated photon signals are seen in ECAL as long as one looks for ECAL clusters. However, for multi-jet samples highly boosted di-photon from π^0 decay will worsen energy resolution.

sample	merging	position	$\alpha[GeV^{1/2}]$	β
e^+e^-	no	barrel & endcaps	0.155 ± 0.007	0.028 ± 0.004
uu	no	barrel & endcaps	0.175 ± 0.001	0.077 ± 0.001
$ccdd$	no	barrel & endcaps	0.179 ± 0.001	0.069 ± 0.001
$\nu\nu\gamma$	yes	barrel	0.164 ± 0.001	0.022 ± 0.001
$\nu\nu\gamma$	yes	endcaps	0.167 ± 0.001	0.030 ± 0.001
CALICE[4]	-	barrel & endcaps	0.166 ± 0.001	0.011 ± 0.001

Table 1: Parameter fitting for photons using different event samples. Test beam results from CALICE are presented for comparison

4 Hadrons in calorimeter

When colored final states are present in hard scattering process, hadronisation is followed in which a great number of mesons and baryons are produced, forming hadronic jets. Stable or long-lived hadrons make up of the major contents that reach calorimeter and deposit most energy inside HCAL. Those include charged pions $\pi^\pm$ (but not neutral π^0 which decays immediately into di-photon), long-lived neutral K_L^0s, protons and neutrons, which will be the focus of our study. Other particles contribute a small fraction of jet energy.

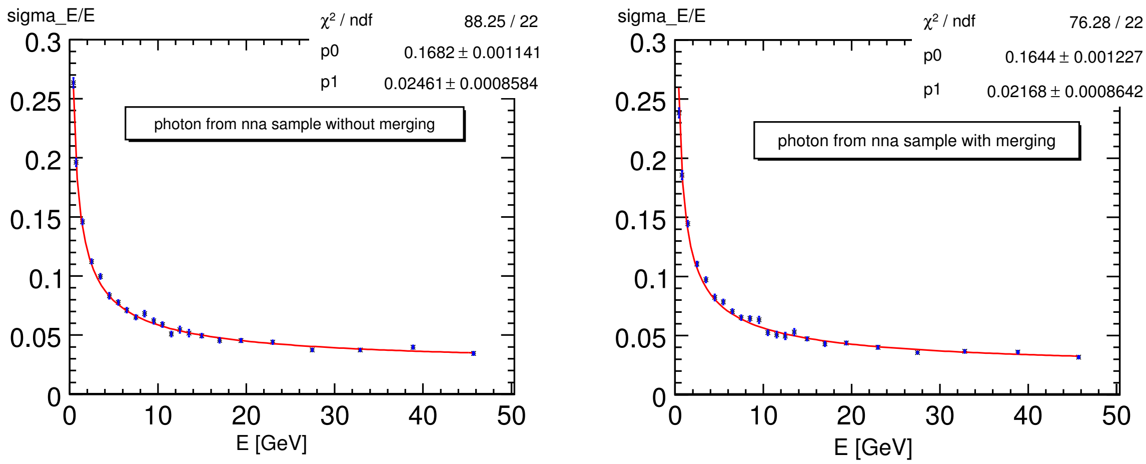

Figure 2: Comparison between results without (left) and with (right) merging technique. The results are similar showing that split-up photons do not affect energy resolution significantly for $\nu\nu\gamma$ sample.

The overall feature of hadronic calorimetry is that it is less precise and more complicated than measurement of photon, due to less compact hadronic showers in HCAL and more probability of confusion between particles within a jet. There are also distinctions between charged hadrons and neutral hadrons, because trajectories of the former can bend under the influence of external magnetic field leading to better separation, and also because their tracks provide more information for reconstruction.

To find out intrinsic performance of the calorimeter, we try to select isolated MC particles producing isolated clusters that are free from confusion. This is especially crucial for the 4-quark *ccdd* sample, where four hadronic jets give rise to high probability of confusion. In principle we can make use of simulation information combined with MC information to select isolated particles and look at their energy resolutions. One needs to look at one cluster for one reconstructed particle to see if it contains hits from more than one true particle or if cluster energy matches that of the ancester true particle. For some technique reason there is no access to multiple ancester true particles from any particular cluster, but we can still use in the data tree a weight ρ

$$\rho = \frac{number\,of\,hits\,in\,cluster\,from\,the\,true\,particle}{total\,number\,of\,hits\,in\,cluster} \tag{3}$$

associated with each MC particle, which tells you what fraction of energy in the cluster (or in the track for a charged particle) comes from it. By demanding that this weight is very close to unity we can select isolated particles with good confidence. (For charged particles this does not help in the analysis of calorimeter performance because that weight is calculated using track information.) For isolated particles we fit resolution parameters for different types of particles, as shown in Table 2.

For the rest, there are several different situations. One posssibility is that no recon-

particle	isolation condition	position	$\alpha[GeV^{1/2}]$	β
$\pi^{\pm}$	$\|\rho-1\|<5\%$	barrel & endcaps	0.480 ± 0.009	0.172 ± 0.004
$p(\bar{p})$	$\|\rho-1\|<5\%$	barrel & endcaps	0.630 ± 0.038	0.103 ± 0.026
$n(\bar{n})+K_L^0$	$\|\rho-1\|<5\%$	barrel & endcaps	0.548 ± 0.030	0.112 ± 0.023
hadrons	CALICE test beam[4]	barrel & endcaps	0.613 ± 0.001	0.023 ± 0.001

Table 2: Resolution parameters for isolated hadrons for *ccdd* four-quark sample. The fitting is conducted at energy range $3GeV < E < 20GeV$. Test beam results from CALICE are presented for comparison

structed particle exists, which is because either particle energy is so small that detector can miss it (actually, for low energy charged particle it can also happen that only tracks are seen, since their trajectories are bent so much that they never reach calorimeter; moreover, we discover that this case happens more often in joint region between barrel and endcaps), or the cluster is confused with another nearby cluster. To effectively rule out the less interesting former case, an energy cut $E_{cut} = 3GeV$ is imposed. Another phenomenon can be seen very clearly in Figure 3, as a secondary peak corresponding to a measured energy close to zero appearing on the left of the central gaussian peak. This is a result of particle split up and can be eliminated from gaussian distribution by demanding only one reconstructed particle for one true particle. It is worth noting that for neutral particles such cut cannot reduce secondary peak completely. If a neutral particle splits into two part, one of which merged with a nearby neutral cluster, the other cluster with smaller seen energy will be associated with the true particle; and since there is no track information to match, this confusion is never corrected. Another

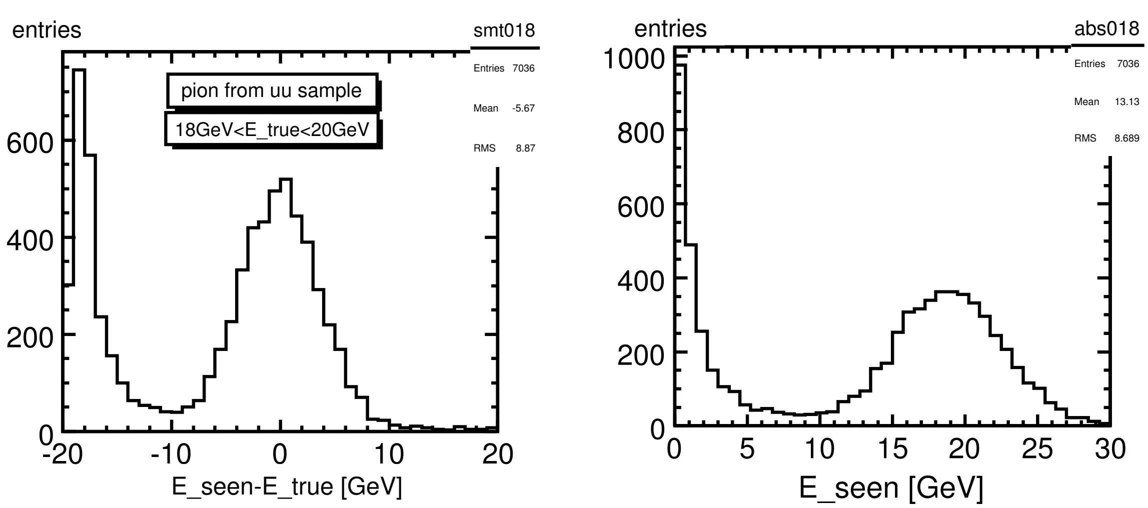

Figure 3: Secondary peak corresponding to very small measured energy, corresponding to $18GeV < E_{true} < 20GeV$ pions from the 2-quark uu sample.

possibility is that the weight is far from unity indicating that multiple clusters overlap is very probable.

A comparison between isolated sample and possibly confused sample is presented in Figure 4, in which we can clearly see that the isolated sample has better energy resolution and fits Eq. (1) much better than the confused.

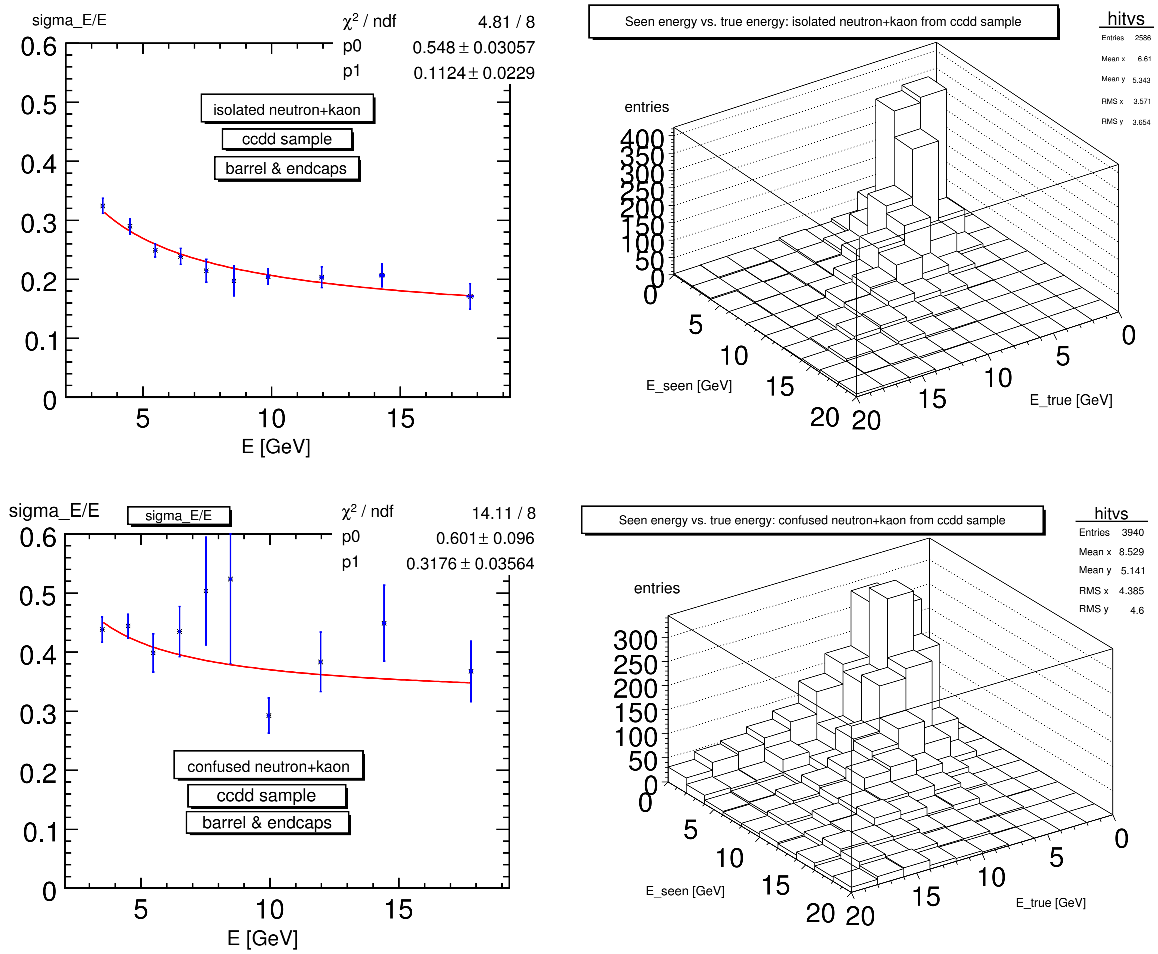

Figure 4: Comparison of energy resolution fit and measurement error between isolated (up) and confused (down) neutral hadrons (neutrons and kaons), analysing the 4-quark *ccdd* sample. Energy range $3GeV < E < 20GeV$ is considered.

Since in fast simulation particle interection with detector will not be simulated in full detail, we want to put in resolution parameters α and β for isolated MC particles as well as the probability of confusion. The probability of confusion is associated with MC particle's distance to its closest neighbor. Naturally, when two particles are close by, their showers tend to overlap. A proper definition of such distance is still needed. We no longer use cone distance as defined in Eq. (2), since for charged particle track bends in magnetic field and hence momentum direction at interaction point (IP) differs from direction of calorimeter hit. We extract from data tree for each MC particle the endpoint position, which is usually where particle shower starts in the calorimeter. A naive definition could be 3-dimensional distance

$$\Delta r = \sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2} \tag{4}$$

However, it does not take into account that particle shower in calorimeter, starting from endpoint, can extend quite long in the radial direction. We propose projected distance given by

$$\Delta r = \sqrt{\Delta x'^2 + \Delta y'^2 + \Delta z'^2} \tag{5}$$

where $x', y', z'$are the coordinates of the point where line connecting endpoint and IP intersects with inner edge of HCAL, with a radius $r_{min} = 1800mm$ for barrel and half-length $z_{min} = 2346mm$ for endcaps [3]. For a given minimum distance to the closest neighbor Δr_{min}, we can extract from full simulation data the probability for a particle to be isolated, and such probability should depend on both detection capability and reconstruction algorithm. For large Δr_{min} probability for isolation is large while for small Δr_{min} confusion is more likely to occur.

The probability for confusion as a function of Δr_{min}, which we would like to put into fast simulation, is also of great interest to be studied. One would expect confusion to be cut off at some large Δr_{min}, but it is difficult to select truely confused particle with 100 percent confidence. At first, particles with $|\rho - 1| > 5\%$ are selected and the ratio between confused number and total number are plotted. However, a tail of $20 \sim 30\%$ at large distance $\Delta r_{min} \gtrsim 600mm$ is observed and it does not decrease rapidly with increasing Δr_{min}. The question comes that how can a cluster be contaminated quite much when the closest true particle is far away? We propose two explanation for this. One is that neutral particles in hadronic shower can travel far before it interacts, thus can happen to confuse with other clusters. The other possibility is rather a false confusion. If a MC particle splits into two particles just before it hits the calorimeter, showers tend to be combined but only attributed to one of the secondary particles rather than the very first one. The first possibility is found not responsible for a persisting tail for large Δr_{min}. Given the following simple estimate, contribution from split-up particles is an order smaller than observed (shown in Figure 5)

$$(probability\, of\, confusion\, for\, \Delta r_{min} = x) \approx \frac{\pi R^2}{2\pi x} P(x) \tag{6}$$

where R is a proper estimate of confusion distance (we find typically $R \sim 200mm$ for photon and $R \sim 400mm$ for hadrons): for $\Delta r_{min} > R$ true confusion is almost impossible. And $P(x)$ is the probability density that a split-up particle has 2 clusters at a distance of x from each other. Both can be obtained from simulation data. It can be shown through geometric argument that Eq. (6) holds for large x.

To eliminate the tail at large x, we choose to cut away false confusion by hand. In Figure 6 histogram for confused neutral hadron count against Δr_{min} is shown in logarithmic scale. One can clearly see that for small x the count decreases exponentially at a larger rate than for large x. We choose to substract the false confusion according

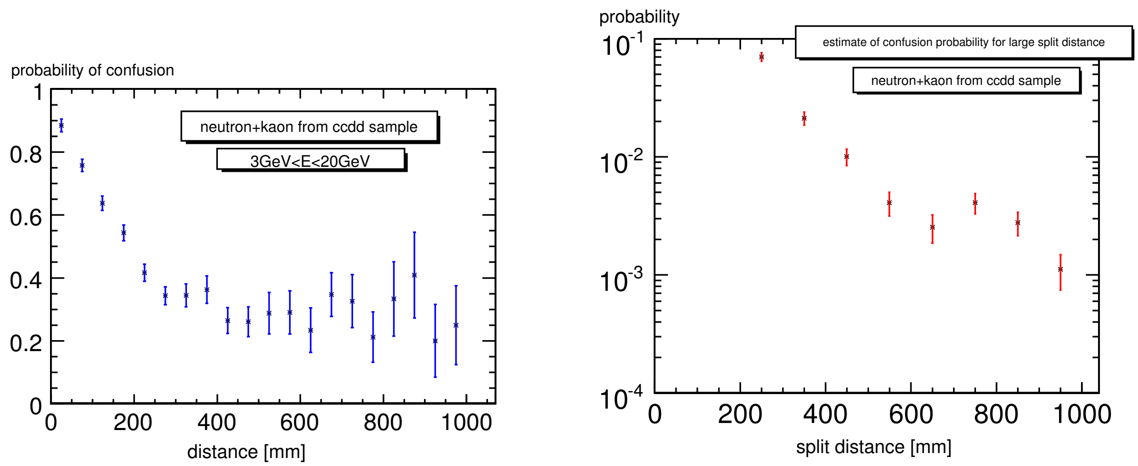

Figure 5: Probability of confusion plotted against Δr_{min} (left) and estimated contribution from large-distance split-up (right). Neutral hadrons with $3GeV < E < 20GeV$ from the 4-quark *ccdd* sample is taken for example.

to the line drawn, and then improve our result for probability of confusion. We use gaussian distribution centered at zero distance to fit the probability

$$P(x) = P_0 exp(-x^2/2\sigma^2) \tag{7}$$

obtaining parameters P_0, σ that will be put into simulation. We find from fit $P_0 = 0.8$ and $\sigma = 200mm$. However, one can see that probability distribution is not exactly gaussian. Although it does cut off at some large distance, there is a tail extended from the end of gaussian core. Note that even for zero distance confusion is not 100 percent, because x is calculated using endpoint position; clusters can be smeared out, or can be separated in the radial direction. Still, theoretical argument for the correct form of probability distribution is in need; and in order to justify the method, parameters extracted via fitting should be compared with results by other methods. It may be more reliable to modify the code for simulation output generation so that false confusion can be ruled out by looking into detailed information of particle flow.

5 Summary

In this summer student project, we have studied parameterisation of calorimetry energy resolution by analyzing full simulation output generated by Mokka. Calorimetry energy resolution has been studied previously with test beams; however, a study based on simulation output takes into account the effect of confusion which is not present in test beam study. We have parameterised energy resolution as a function of particle energy in accord with the conventional way , and will put the parameters obained into a fast simulation that is much faster than a full simulation. We have studied ECAL

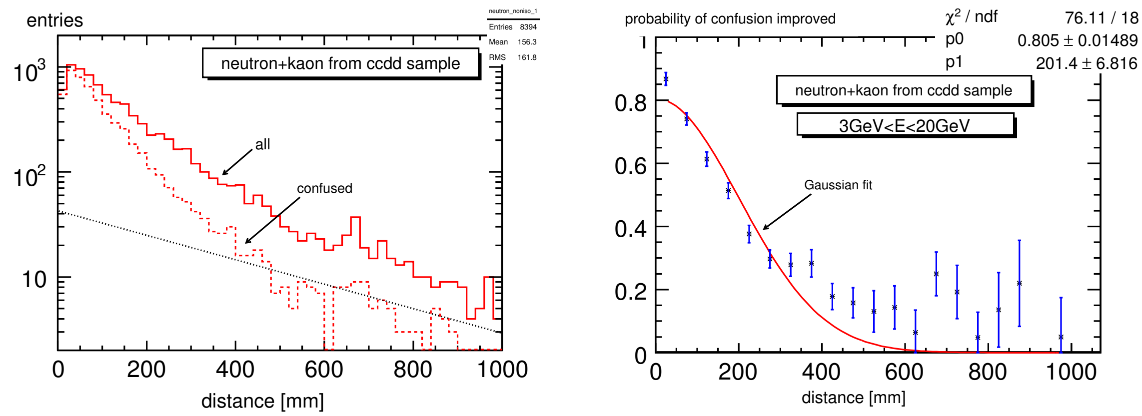

Figure 6: False confusion is substracted according to the dashed line in logarithmic scale (left). Improved probability of confusion for neutral hadrons with $3GeV < E < 20GeV$ from the 4-quark *ccdd* sample (right). One would like to compare this with Figure 5.

performance by investigating single photon energy resolution and have adopted merging technique to account for split-up photons. Results are in good agreement with test beam results from CALICE. We have also studied HCAL performance by investigating hadron energy resolution in multi-jet samples. Isolated particles have been selected and emphasis on neutral hadrons has been put, with parameters close to CALICE results extracted. Finally a tentative investigation into confusion has also been included and a parameterisation of confusion probability is proposed as a function of minimum projected distance between endpoints. Further investigation could be conducted with improvement on full simulation output so that as much as useful information is accessible.

Acknowledgement

I would like to thank my supervisers Mickel Berggren and Jenny List for their instruction during the project. I would also like to thank colleagues in FLC simulation group for helpful discussion. Finally I am grateful to organisers of the DESY Summer Student Program 2010 for their hospitality.

References

[1] M. A. Thomson, Nucl. Instrum. Meth. **A611**: 25-40, 2009, arXiv:0907.3577v1 [physics.ins-det].

[2] J-C. Brient, H. Videau (laboratoire Leprince-Ringuet), eConf C010630 (2001) E3047, arXiv:hep-ex/0202004v1.

[3] ILC Global Design Effort and World Wide Study, International Linear Collider reference design report, Volume 4: Detectors.

[4] The ILD concept group, International Large Detector, Letter of Intent, DESY 2009-87, FERMILAB-PUB-09-682-E, KEK Report 2009-6.