Clustering algorithm test for the digital electromagnetic shower

Yoshinari Mikami*

University of Birmingham, Edgbaston Birmingham, UK*

On behalf of the CALICE-UK digital ECAL team

5th Nov. 2011*
(Internal note)

Abstract

We tested an algorithm development for the clustering of the electromagnetic shower with the digital ECAL device. This study based on the Geant4 simulation for the digital ECAL. We investigated basic things such as the cluster size variation, the efficiency, the resolution and the linearity just as the first practical test.

∗ This note was summarized for the study at University of Birmingham during 2006 to 2008. Contact e-mail address now at ymikami@cern.ch

Introduction

We study an algorithm development of the clustering for the electromagnetic shower with the digital ECAL (Electromagnetic CAlorimeter) device. This digital ECAL measures the electromagnetic shower energy just by counting the numbers of secondary particles after the shower. The numbers of secondary particles are proportional to the energy of the incoming particle, therefore, we can measure the incident energy if we could obtain the precise number of secondary particles. This method requires a small cell size because we have to detect the secondary particles one by one. In order to obtain a proper energy measurement, we have to collect secondary particles properly. For this purpose, we tested an algorithm for the clustering of secondary particles after the electromagnetic shower. This clustering enables us to measure the energy of the incident particle. Since the electromagnetic shower is very high dense (e.g. roughly the order of 100 cell hits per 1 GeV incident electron), the clustering method is summing all hits inside a cylinder simply rather than searching the neighbour hits one by one. The objective of this note is a basic test that we can achieve an enough high efficiency for the lower energy of the 1 GeV electron.

Geometry

We use the ILC (International Linear Collider) detector model for this study. The ILC ECAL is a sandwich structure between silicon sensors and tungsten radiators. We modified the geometry from the conventional analogue ECAL to the digital ECAL. The modified software is the ILC detector model MOKKA [1] version 6. The only difference except for the Si cell size is the Si sensitive thickness. In the conventional analogue ECAL, the sensitive Si thickness was for full volume of 500 μm thickness. On the other hand, in the digital ECAL case, only the epitaxial layer's 15 μm was used for the Si sensitive thickness. The rest of 485 μm silicon thickness was implemented as non-sensitive silicon volume. Except for this Si sensitive thickness modification and for the cell size, all other geometry were kept as the same between the conventional ECAL case and the digital ECAL case [2]. (i.e. The W thickness, the plastic board thickness and total Si thickness were not changed at all.) In the ILC detector model, the ECAL has 30 layers, then the tungsten thickness are different between the inner layers and the outer layers. The tungsten thickness is 2.1 mm for the inner 20 layers and 4.2 mm for the outer 10 layers. This difference of the W thickness need be compensated when we counts the number of secondary particles. A weighted number which just doubled for outer layers works well. [2] The weighted number was used in this note. The cell size is 50 μm x 50 μm cell for this study. We used the Geant4 response only. It means that we did not add any realistic issues such as the charge diffusion, the dead area and the noise so on. Although we did not include these sensor level realistic issues in this note's simple study, they are discussed in elsewhere. [3] [4] [5] [6]

Clustering algorithm

We developed the electromagnetic shower clustering algorithm for the digital ECAL device. The basic algorithm steps are as follows.

- Finding a grouping of hits within inner layers.
- Opening a hemisphere.
- Obtain the center of gravity within the hemisphere.
- Obtained the direction of the tower. (From two centers of gravity)
- Adding all hits within a cylinder tower.

At first, we search a grouping of hits within inner 7 layers. Several different requirements for the number of hits within the grouping were shown in Fig 1. This figure is plotting the efficiency as a function of the radius of the grouping circle. This efficiency is defined as that the events number which has the grouping is divided by the all events number which is generated. The grouping was searched in the same layer only. We used 1 GeV single electron events. From this study, the requirement of 4 hits within a circle of 5.0 mm radius is a reasonable working assumption criteria. As a comparison, the case of using inner 3 layers was plotted in Fig 2. At this moment, no threshold for the cell energy deposit was applied. The cell hit energy deposit for the 1 GeV single electron events are shown in Fig 3. The Minimum Ionization Particle (MIP) peak is around 3.6 keV. We tested several thresholds and plotted the efficiency as a function of the incoming single electron energy in Fig 4. The half MIP value 1.8 keV is applied as the threshold for the cell energy deposit as it is indicated by the line and the arrow in Fig 3. In second, we open a hemisphere in order to obtain the direction of the cluster tower. Then we can obtain the center of gravity of all hits inside the hemisphere. Thus, we can decide the direction from the center of gravity of the initial grouping toward the center of gravity of all hits inside the hemisphere. At last, we add all hits inside a cylinder which has a radius of the Moliere radius order. Typical shower spread distributions are shown for the 1 GeV single electrons in Fig 5 and for the 5 GeV single electrons in Fig 6. These single electrons were injected from the interaction point of the ILC detector without magnetic fields. In these distributions, no threshold was applied. We plot efficiencies as a function of the cylinder radius in Fig 7. The efficiency is defined as the number of cell hits inside the cylinder divided by the number of all cell hits in the event. Both numbers of cell hits are after the 1.8 keV threshold for the cell hits energy deposit in this efficiency calculation.

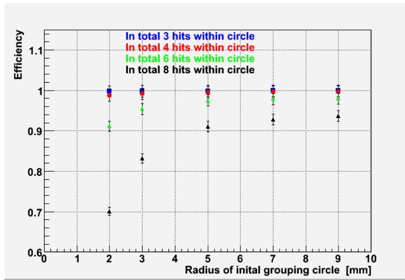


Figure 1: Efficiency vs. Radius of grouping circle. Within the inner 7 layers. Different requirements for the number of hits inside the circle as 3 hits (blue), 4 hits (red), 6 hits (green) and 8 hits (black). With the 1 GeV single electrons.

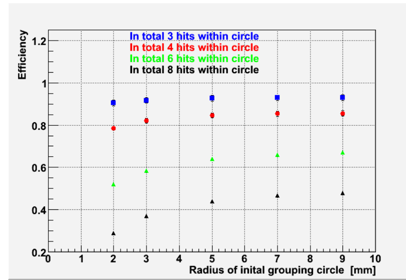


Figure 2: Efficiency vs. Radius of grouping circle. Within the inner 3 layers. Different requirement for the number of hits inside the circle as 3 hits (blue), 4hits (red), 6 hits (green) and 8 hits (black). With the 1 GeV single electrons.

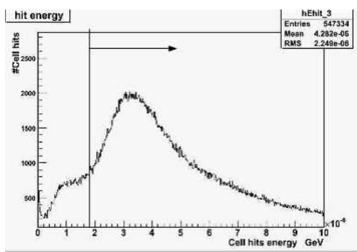


Figure 3: Cell energy deposit for the 1 GeV single electrons

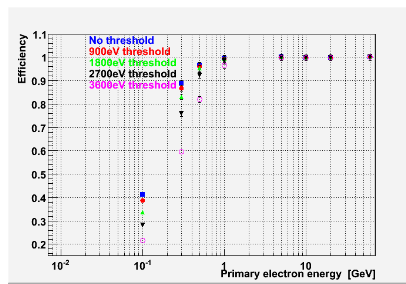


Figure 4: Efficiency vs. Incoming electron energy. Thresholds for the cell energy deposit were applied by 0 keV (blue), 0.9 keV (red), 1.8 keV (green), 2.7 keV (black) and 3.6 keV (purple). With the 1 GeV single electrons.

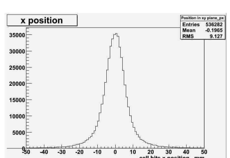


Figure 5: Shower spread of the 1 GeV single electrons

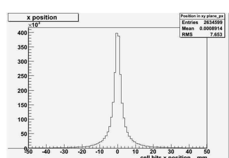


Figure 6: Shower spread of the 5 GeV single electrons

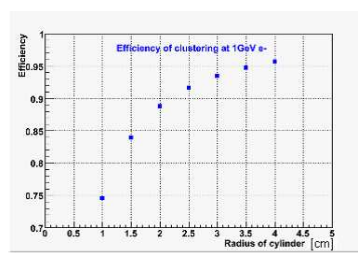


Figure 7: Efficiency vs. Radius of the cylinder. With the 1 GeV single electrons.

Energy resolution

The energy resolution as a function of the cylinder radius is plotted in Fig 8 for the 1 GeV single electrons. We use the 2.0 cm radius for the cylinder radius at this moment. The energy resolution as a function of the opening hemisphere radius is plotted in Fig 9 for the 1 GeV single electrons as well. We use the 3.0 cm radius for the opening hemisphere. The clustering algorithm steps are eventually as follows.

- Applying the 1.8 keV threshold for all cell hit energy deposit.
- Finding a grouping of 4 hits inside 5.0 mm circle within inner 7 layers.
- Obtain the center of gravity within the hemisphere of 3.0 cm radius.
- Obtained the direction of the tower. (From two centers of gravity)
- Adding all hits within the cylinder tower of 2.0 cm radius.

We plot the energy resolutions as a function of the incoming energy in Fig 10. Where the blue mark is using this note's clustering algorithm. On the other hand, the red mark is using all cell hits, which means the perfect clustering case as a comparison. As described there, this note's clustering is not so bad resolution at least up to the 50 GeV single electron energy scale. (More higher energy was not investigated at this study.) However, this is only for the single electron case. We have to care more when it is a real physics event case. They are not studied in this note. Distributions for the weighted number of cell hits per event are shown for the case of all cell hits (the perfect ideal clustering case) in Fig 11 and for the case of this clustering applied in Fig 12. These are 1 GeV single electrons which are injected just in front of the ECAL device without magnetic fields. As a comparison, distributions with 4T magnetic fields are shown for the case of all cell hits (the perfect ideal clustering case) in Fig 13 and for the case of this clustering applied in Fig 14. In order to avoid a double counting of clusters, only one cluster is used in each single event as a temporary method. The effect of 4T magnetic fields is roughly around 10% reduction for the energy resolution at the 1 GeV single electrons.

Summary

We tested an algorithm for the clustering of the electromagnetic shower with the digital ECAL device. This study based on a simple Geant4 simulation only. Some basic things were investigated such as the clustering size variation, the efficiency, the resolution and so on. These are not yet finalized for the optimization. For example, an use of a cone angle instead of the fixed hemisphere will be more flexible and it can suppress backgrounds better. We tested as a baseline of the clustering algorithm. This study can be compared with future studies which will include some realistic issues such as the charge diffusion, the digitization, the noise and the dead area and so on.

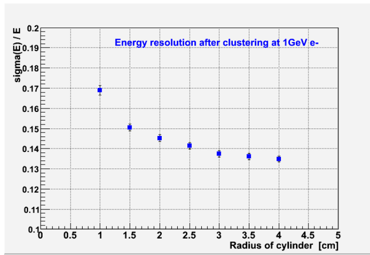


Figure 8: Energy resolution vs. Radius of the cylinder. With the 1 GeV single electrons.

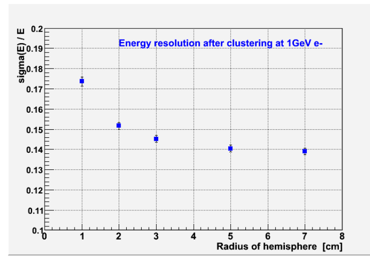


Figure 9: Energy resolution vs. Radius of the hemisphere. With the 1 GeV single electrons.

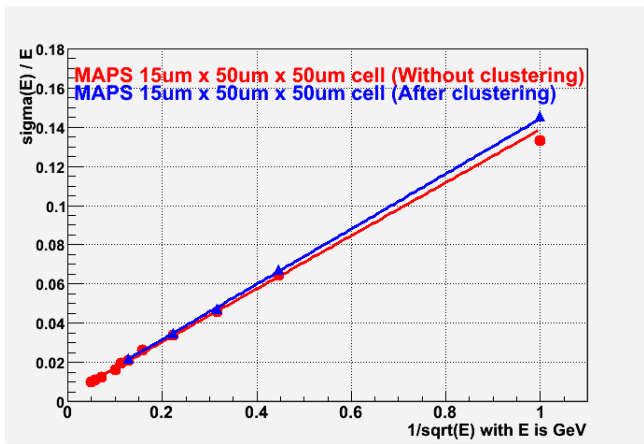


Figure 10: Energy resolution vs. $1/\sqrt{E}$ (The E is with the GeV unit for the energy of the incoming electrons). After the clustering applied (blue) and all hits without the clustering (red).

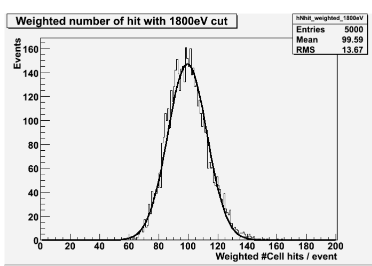


Figure 11: The weighted number of cell hits per event. Without magnetic fields. All cell hits without the clustering. With the 1 GeV single electrons.

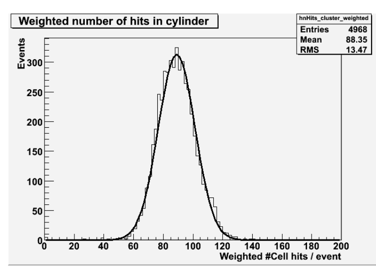
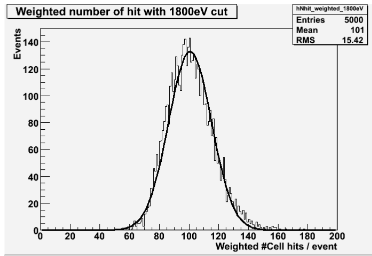


Figure 12: The weighted number of cell hits per event. Without magnetic fields. After the clustering applied. With the 1 GeV single electrons.



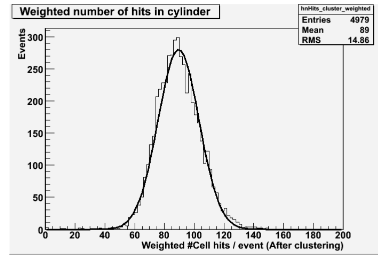


Figure 13: The weighted number of cell hits per event. With 4T magnetic fields. All cell hits without the clustering. With the 1 GeV single electrons.

Figure 14: The weighted number of cell hits per event. With 4T magnetic fields. After the clustering applied. With the 1 GeV single electrons.

References

[1] http://www-zeuthen.desy.de/linear_collider

[2] Geant4 simulation for the digital electromagnetic calorimeter, Y.Mikami, N. Watson, CALICE-UK digital ECAL internal note, 2011

[3] A MAPS-based Digital Electromagnetic Calorimeter for the ILC: J.A. Ballin et al., arXiv:0709.1346v1, LCWS07 proceeding

[4] A digital ECAL based on MAPS: J.A.Ballin et al., arXive:0901.4457v1, LCWS08 proceeding

[5] Monolithic Active Pixel Sensors (MAPS) in a Quadruple Well Technology for Nearly 100% Fill Factor and Full CMOS Pixels, J.A. Ballin et al., Sensors 2008, 8(9), 5336-5351

[6] Design and performance of a CMOS study sensor for a binary readout electromagnetic calorimeter; J.A.Ballin et al., 2011 JINST 6 P05009