Tracks of hadronic showers in the Si-W ECAL physics prototype

The CALICE Collaboration [1]

Abstract

A simple track-finding algorithm was developed to study tracks left by secondary particles, emerging from hadronic interactions in a highly granular electromagnetic calorimeter prototype. The predictions of two Monte Carlo GEANT4 models are compared with experimental data, taken at FNAL in 2008, using a novel track-finding algorithm. Present Monte Carlo simulations provide a good description of the experimental data in terms of new observables, available through the detailed analysis of the secondary particles; the Monte Carlo predictions are within 20% of the data, and for many observables much closer. Main systematic effects were studied using electron and muon samples.

[1] Corresponding authors:
Roman Pöschl: poeschl@lal.in2p3.fr, Sviatoslav Bilokin: bilokin@lal.in2p3.fr

Contents

1 Introduction

The design of particle detectors at future high-energy physics experiments and, in particular, at linear colliders is oriented towards the usage of Particle Flow algorithms (PFA) for the event reconstruction. These algorithms require calorimeters with high granularity to reconstruct individual particles, aiming at the improvement of the jet energy resolution [1, 2, 3].

The primary objective of the CALICE (Calorimeter for the Linear Collider Experiment) collaboration is the development, construction and testing of highly granular hadronic and electromagnetic calorimeters for future particle physics experiments.

A detailed study of the calorimeter response to particle interactions is necessary to verify existing Monte Carlo simulation models and to build a reliable PFA. This implies the precise simulation and reconstruction of the interaction of neutral and charged hadrons and the subsequent particle cascade.

This note reports on a detailed study of hadronic interactions in the CALICE Silicon-Tungsten Electromagnetic Calorimeter (Si-W ECAL) physics prototype [4]. The Si-W ECAL was tested at Fermi National Accelerator Laboratory (FNAL) in 2008 using a beam of π^--mesons in the energy range from 2 to 10 GeV. The highly granular structure of the Si-W ECAL permits a detailed measurement of hadronic showers in terms of integral observables [5] such as cluster extensions and energy depositions. As will be shown in this note, the high granularity allows in addition for deeper studies of the interactions between the hadrons and the absorber material such as the characterisation of the interaction region and the analysis of secondaries emerging from the interaction. The tracks produced by these secondaries are reconstructed by a new simple track-finding algorithm. The resulting observables are subject to comparison of data with predictions from different GEANT4 Monte Carlo physics lists. The analysis complements studies presented in [6] and [7] for tracking in CALICE prototypes of hadronic calorimeters.

2 The Si-W ECAL physics prototype

The Si-W ECAL physics prototype has a sandwich-like structure comprising 30 layers of silicon (Si) as active material, alternated with tungsten (W) as absorber material. The active layers are made of Si wafers segmented in $1 \times 1\,\mathrm{cm}^2$ pads. As shown in Fig. 1, each wafer consists of a square of 6×6 pads and each layer is a matrix of 3×3 of these wafers resulting in an active zone of $18 \times 18\,\mathrm{cm}^2$.

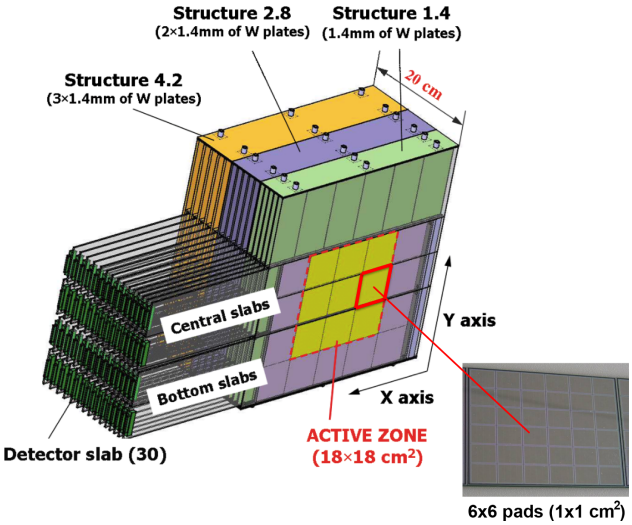


Figure 1: *A schematic view of the Si-W ECAL physics prototype.*

The Si-W ECAL is subdivided into 3 modules of 10 layers each. The W depth per layer is different in each module increasing from 1.4 mm (0.4 radiation lengths or X_0) in the first one, to 2.8 mm in the second and 4.2 mm in the last one. The total thickness corresponds to 24 X_0 or about 1 nuclear interaction length λ_I which ensures that more than half of the hadrons will have a primary interaction within the detector volume. A more detailed description of the prototype can be found in [4].

3 Data and Monte Carlo samples

3.1 Experimental setup at FNAL

The test beams were carried out at the Fermilab Test Beam Facility[2], FTBF, at FNAL in May and July 2008. The Si-W ECAL was placed in front of two other CALICE physics prototypes: an AHCAL [8] and a TailCatcher (TCMT) [9]. The beam-line comprised in addition two scintillator counters, covering an area of 10×10 cm^2, for triggering on incoming particles and two Cherenkov detectors for particle identification. The chosen coordinate system is right-handed with the z-axis pointing along the beam direction and the y-axis being vertical. The analysed data in this note comprise runs with primary π^--mesons. The energies of the primary particles are 2, 4, 6, 8 and 10 GeV.

3.2 Monte Carlo simulations

Monte Carlo simulations were carried out within the Mokka framework [10], which provides the geometry interface to GEANT4 [11]. There are several simulation models of hadronic interactions available within GEANT4, that are combined into so-called *physics lists*. Each model has its own theoretical basis valid mainly in a specific energy range of hadrons. In this analysis, two physics lists contained in GEANT4 version 9.6 p01 are compared with the experimental data:

- QGSP_BERT combines the Bertini model at energies below 9.9 GeV, with the Low Energy Parametrised model at energies above 9.9 GeV;
- FTFP_BERT has a transition from the Bertini model to the Fritiof model around a primary particle energy of 4.5 GeV.

More information about these and other physics lists can be found in [12].

[2]Fermilab Test Beam Facility web page: `http://www-ppd.fnal.gov/MTBF-w`

3.3 Event selection and preprocessing

The FNAL π^- test beam is contaminated with μ^- and e^-, in particular at lower energies where the beam is dominated by e^-. At 2 GeV the beam contains about 5% of π^- mesons and 70% of electrons. Events are triggered using the signals from the two scintillator counters upstream of the Si-W ECAL and π^- are identified by using Cherenkov counters. The response of the Si-W ECAL to charged particles was calibrated with an energetic μ^- beam [13] and the hit energy is converted into units of most probable energy depositions by minimum ionising particles (MIP).

To select π^- showers the data and simulation samples undergo the following selection steps [5][14]:

- A series of cuts is applied to reject multi-particle events caused by beam impurities or products of decays or upstream interactions of primary particles. The influence of residual multi-particle events will be discussed in Sec. 5;
- A threshold of 0.6 MIP is chosen to remove noisy hits in the Si-W ECAL;
- A hit is removed as being isolated if all the 26 pads in the surrounding cube have no signal above the noise threshold. The analysis presented in this note applies to the non-isolated hits that remain after this removal. The term hits will continue to be used.
- The total number of hits in the ECAL is required to be at least 25 to remove particles with large incident angle;
- The barycentres of the transverse coordinates $\bar{x}_{hit}$ and $\bar{y}_{hit}$ of the hits are calculated as:

$$\bar{x}_{hit} = \frac{\sum_{hits} x_{hit}\, E_{hit}}{\sum_{hits} E_{hit}} \text{ and } \bar{y}_{hit} = \frac{\sum_{hits} y_{hit}\, E_{hit}}{\sum_{hits} E_{hit}}, \tag{1}$$

  where E_{hit} is the energy of a hit in MIP units, and the sums run over all hits in the calorimeter. The event is accepted if $-50\,\text{mm} < \bar{x}_{hit} < 50\,\text{mm}$ and $-50\,\text{mm} < \bar{y}_{hit} < 50\,\text{mm}$ to reduce lateral shower leakage;

- In first approach the interaction layer i is identified with the first of three consecutive layers for which

$$E_i > E_{cut}, E_{i+1} > E_{cut} \text{ and } E_{i+2} > E_{cut}. \tag{2}$$

  This simple condition is inefficient at small energies and is complemented by using the following relative energy increase

$$\frac{E_i + E_{i+1}}{E_{i-1} + E_{i-2}} > F_{cut} \text{ and } \frac{E_{i+1} + E_{i+2}}{E_{i-1} + E_{i-2}} > F_{cut}, \tag{3}$$

  with E_i being the total energy of layer i. The variables E_{cut} and F_{cut} are free parameters with empirical values of 8 MIP and 6, respectively. It is argued in [5] and references therein that these values optimise the selection efficiency in the energy range relevant for the present study. The event is selected if $5 < i < 15$ to suppress electron contamination and to assure "long" secondary tracks after the interaction.

4 The track-finding algorithm

The main objective of the track-finding algorithm is the detection of forward-scattered tracks from the interaction between the primary pions and the absorber material in the absence of a magnetic field.

The designed algorithm has the following execution scheme:

- After the selection cuts, see Sec. 3, the interaction region is identified and singled out. The interaction region will be defined in Sec. 4.1;
- The remaining energy depositions are used for clusterisation;
- The obtained clusters are classified to select track-like clusters from residual noise;
- After classification different clusters from a single outgoing secondary particle are merged into one track.

The entire algorithm is executed on a unit grid based on the Si-W ECAL pad identifiers according to

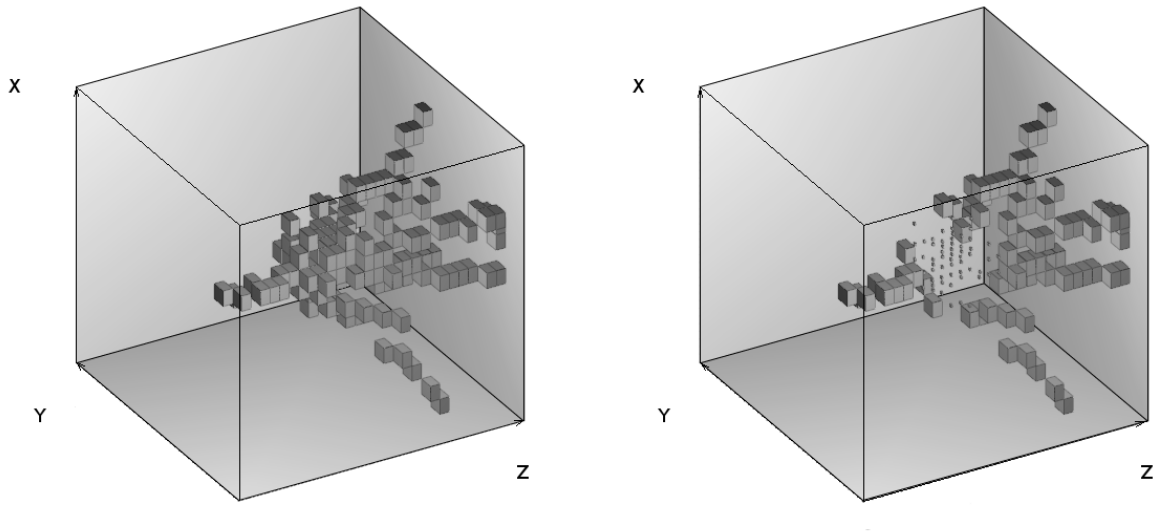
$$\vec{x} = (x, y, z) = \begin{cases} x = 0..17 \\ y = 0..17 \\ z = 0..29, \end{cases} \tag{4}$$

where pad counting starts in the bottom right pad, see Fig. 1. Distances in this grid are measured in *grid units*, g.u.

4.1 Removing the interaction region

A typical inelastic hadronic interaction in the Si-W ECAL creates a shower with an interaction region and tracks of long-lived particles emerging from it. The interaction region is created by particles with a short distance of flight in the absorber material of the Si-W ECAL, like electrons, photons and low momentum hadrons.

In the present analysis the interaction region is defined by all hits that have at least six neighbouring pads with a signal above the noise threshold. In this sense the neighbouring pads are always also part of the interaction region, which applies in particular at the border of the interaction region. The minimal value of six pads is chosen such that muon events remain unaffected meaning that a fake interaction region is found in only 1% of single muon events. On the other hand for a value of five neighbouring pads about 10% of muon events would get assigned a fake interaction region. Increasing the minimal value to seven neighbouring pads with hits reduces the fraction of events with a fake interaction still a bit but does not alter the results presented in the following.



(a) *with interaction region.* (b) *without interaction region.*

Figure 2: *Event display of a primary pion with an energy of 10 GeV simulated using the* FTFP_BERT *physics list before (a) and after removal of the interaction region (b). Smaller cubes are pads that are part of the interaction region and are not processed by the track-finding algorithm. In this event the hits in the first ten layers are classified as hits left by a primary particle.*

Figure 2a displays a simulated event after noise and isolated hits filters and Fig. 2b is the same event after removal of the interaction region. As can be seen in Fig. 2b the interaction region is the starting point for secondary tracks.

4.2 Clusterisation

During the clusterisation step the energy depositions that are not associated to the interaction region are grouped into clusters by a topological principle.

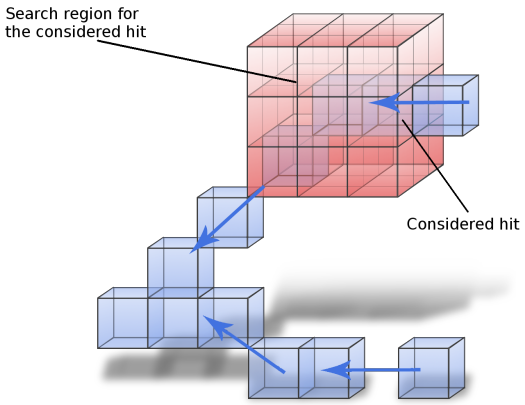


Figure 3: *Illustration of the clusterisation step. The Si-W ECAL hits are represented by blue cubes, and the search region for adjacent hits is indicated by red cubes. The blue arrows point in the direction of the clusterisation flow.*

The algorithm is described in the following. The description is supported by Fig. 3.

1. The separation of tracks improves with increasing distance from the interaction layer. Therefore, hits, isolated within one of the rear layers, seed a cluster. The search for these hits is carried out progressively starting from the last layer in the direction of decreasing z. Typically, seeding hits are found in the last layer of the detector;

2. A hit can be associated to a cluster if it was not yet joined to another cluster. This condition excludes the double counting of hits. Effects, arising from ambiguities in the assignment of hits to clusters such as the order in which clusters are created, are expected to be small, but will be addressed in future studies;

3. For the clusterisation a usual nearest neighbour clustering scheme is applied. More precisely, for each newly associated hit with coordinates (x_n, y_n, z_n) the algorithm finds nearby hits with the following conditions:

   - A neighbour hit should have a z coordinate within $[z_n - 2, z_n]$
   - The transverse coordinates of neighbouring hits is searched within ranges $[x_n - 1, x_n + 1]$ and $[y_n - 1, y_n + 1]$

   The search region for nearby hits is visualised in Fig. 3 as a "red cube" with $3 \times 3 \times 3$ pads;

4. For each newly associated hit the Steps 2 and 3 are repeated until the process reaches the first layer of the calorimeter or until no more neighbour hits are found.

The choice of this clusterisation method is motivated by a maximum correspondence between the number of clusters and the number of detected tracks.

4.3 Classification and merging

Long-lived charged secondary particles from hadronic interactions are expected to leave straight MIP-like tracks in the detector. The goal of the classification of the clusters obtained in the previous step is thus to select track-like clusters.

The classification algorithm executes the following steps:

1. Reject all clusters with 2 hits (N_{hits}) as residual noise clusters;

2. Calculate the length l of the considered cluster as the maximal distance between any pair of hits that are in the cluster;

3. A cluster is rejected if it has a length of less than $l_{cut} = 2\,\text{g.u.}$, which corresponds to the minimal length of a track-like cluster with 3 hits;

4. Compute the following observable:

$$\xi = \frac{l}{N_{hits} - 1} + \varepsilon N_{hits}, \tag{5}$$

as a measure for the eccentricity of the cluster. The first term of Eq. 5 in motivated by the linear dependence of $N_{hits} - 1$ on the cluster length l, illustrated in Fig. 4. The second term introduces a free parameter ε as an ad hoc correction to increase the efficiency for selecting clusters that do not conform to the nominal pencil-like topology, as explained below. The value of the parameter was chosen to be 0.03 after visual inspection of a few tens of events in the event display;

5. If $\xi \geq 1$ a cluster is considered as track-like. Otherwise, the cluster can be classified as e.g. two inseparable tracks.

Due to a number of effects like multiple scattering, residual detector-noise, δ-rays or the residual arbitrariness in the assignment of hits to clusters, the reconstructed tracks are in general not exactly pencil-like. The correction term εN_{hits} in the definition of ξ serves to keep a cluster as track-like even if it has large N_{hits} and its form is not strictly pencil-shaped, i.e. $l/(N_{hits} - 1) < 1$.

A deeper discussion of the effects of the ε parameter is presented in Sec. 4.4.

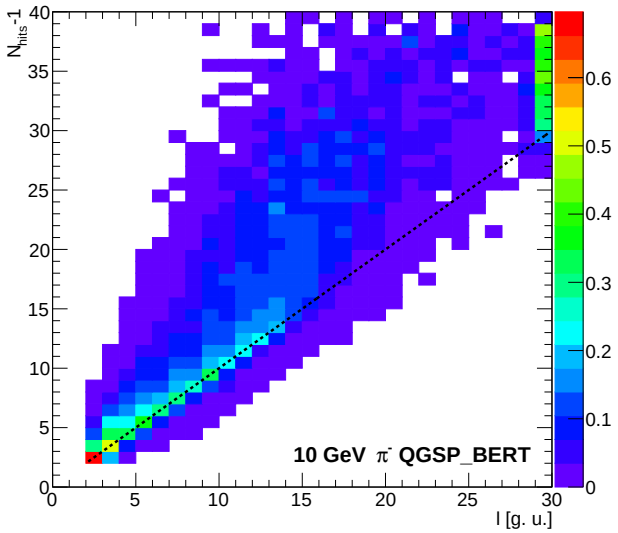


Figure 4: *Correlation between $N_{hits} - 1$ and cluster length l in g.u. at the example of simulated pions with an energy of 10 GeV using the* QGSP_BERT *physics list. To guide the eye a line for $N_{hits} - 1 = l$ is included in the figure.*

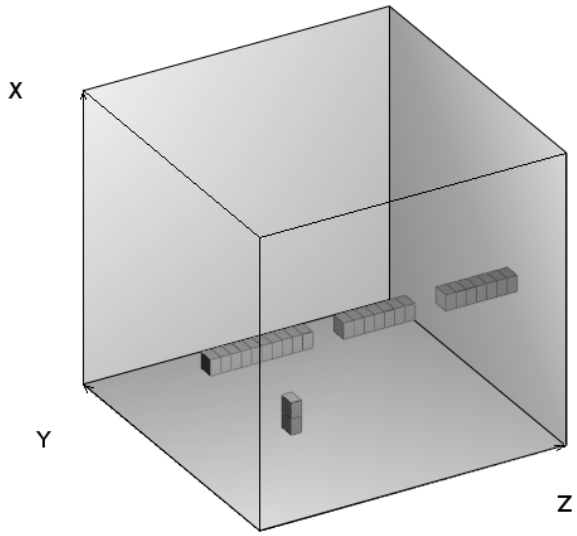


Figure 5: *An example of a segmented track in the Si-W ECAL from a single 6 GeV μ^- in* FTFP_BERT *Monte Carlo simulation.*

The next step of the classification is to detect a cluster from the primary particle. If this cluster exists and meets the conditions for a track-like cluster, it affects the track counting and merging algorithm. A cluster is classified as being produced by the primary particle if it starts in the first module of the Si-W ECAL and if it encloses a small angle with respect to the z-axis. An example of a cluster by a primary particle is visible in Fig. 2b. Clusters assigned to primary particles are discarded in the following analysis.

Different track-like clusters that correspond to track segments from a single particle, see Fig. 5, have to be merged into a single track. The merging procedure combines track-like clusters with any type of clusters using a simple cone algorithm. Tested on a sample of single, isolated muons with an energy of 6 GeV simulated with the FTFP_BERT physics list, the track-finding algorithm finds correctly only one track with 99.7% efficiency. The sample contains about 3% events with segmented tracks.

4.4 Discussion of the ε parameter

The track-finding algorithm depends on a number of parameters. The biggest sensitivity of the track-finding algorithm is expected to be introduced by the empirically defined ε parameter, see Eq. 5. Therefore its influence on the results and a further motivation of the choice of the working point in terms of the ε parameter is given in the following.

After the cut on the minimum cluster length l_{cut}, see Sec. 4.3, all clusters with 3 hits are track-like, independently of the ε value. Therefore, these clusters are not considered in this discussion.

Figure 6 shows the dependence of $\langle N_{tracks} \rangle$ on ε for different beam energies[3]. Each curve has its minimum value at $\varepsilon = 0$ that is the mean number of ideal pencil-like tracks per event and saturates at large ε, when each cluster is taken as track-like.

[3]The mean number of tracks as function of ε in data and simulation is given in Appendix A for future reference.

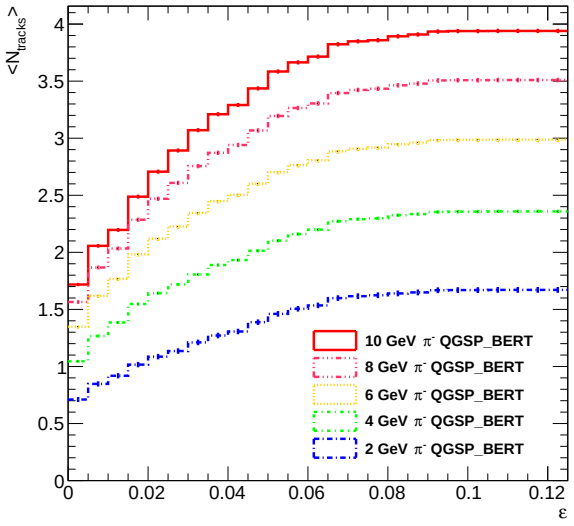
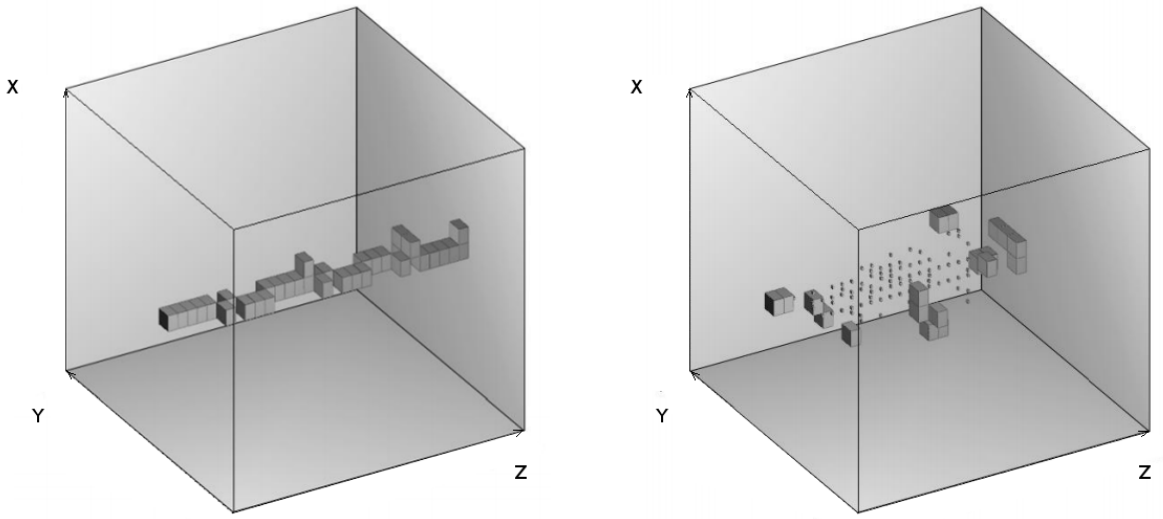


Figure 6: *Mean number of tracks found by the track-finding algorithm as a function of ε for 2, 4, 6, 8 and 10 GeV beam energy for the* QGSP_BERT *physics list simulation. Events without a detected interaction region are discarded.*

For all clusters with a number of hits larger than 3, the simulated muon and electron samples are used to define a motivated choice for the value of ε to be applied for the track finding.

A Monte Carlo sample of μ^- is used to determine a lower bound on the ε parameter. The events for this study are selected if the number of hits is larger than the number of layers in the Si-W ECAL physics prototype. After this cut the muon tracks do not have a straight pencil-like shape, but rather a line with a number of adjacent hits generated by residual detector noise or δ-rays. An example of such a track is given in Fig. 7a.

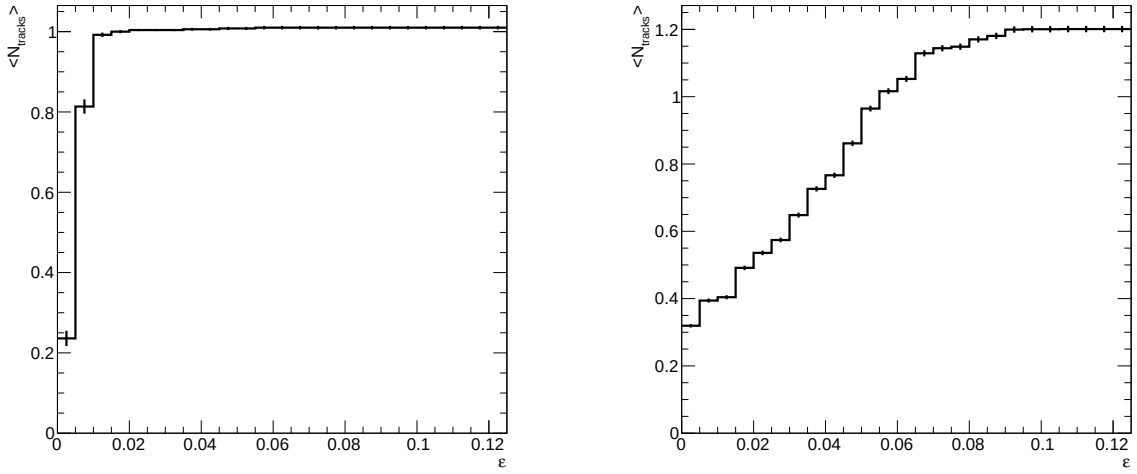


(a) (b)

Figure 7: *Event displays of simulated events showing a muon with an energy of 10 GeV after selection by the number of hits in Si-W ECAL (a) and a 6 GeV electron after removal of the interaction region (b).*

The resulting sample represents secondary tracks from hadronic interactions with noise or other additional hits from the debris of the interaction region.

A Monte Carlo sample based on electrons is used to estimate an upper bound on ε. Electrons are not expected to generate tracks but rather only an interaction region accompanied by low energetic satellite clusters, see Fig 7b. Events in this sample contain only non track-like clusters.



(a) *muon sample* (b) *electron sample*

Figure 8: *The distribution of $\langle N_{tracks}\rangle$ in simulated events as a function of ε for (a) 10 GeV μ^- after selection by the number of hits in Si-W ECAL and (b) 6 GeV electrons after removal of the interaction region .*

Figure 8 shows the dependence of $\langle N_{track}\rangle$ on ε for the muon (Fig. 8a) and electron (Fig. 8b) samples. The lower bound ε_{low} of the ε parameter is identified with the value of ε for which $\langle N_{tracks}\rangle \simeq 1$ in the muon sample. Correspondingly, the upper bound ε_{up} is identified with that value of ε for which $\langle N_{tracks}\rangle = 1$ in the electron sample. The approximate values are:

$$\varepsilon_{low} \simeq 0.015 \texttt{ and } \varepsilon_{up} \simeq 0.05. \tag{6}$$

The empirically chosen value for ε of $\varepsilon = 0.03$ lies within these bounds.

As the algorithm is a new development it will be convenient to give the reader a feeling on the sensitivity of the results with respect to the actual choice of the ε parameter For this the estimator

$$S_{\mathcal{O}} = \frac{\langle \mathcal{O}(\varepsilon_1) - \mathcal{O}(\varepsilon_2)\rangle}{\langle \mathcal{O}(\varepsilon_{nom})\rangle} \tag{7}$$

is introduced and will be evaluated in Sec. 5 where applicable.

Further dependencies of the track-finding algorithm on the value of the cone angle of the merging algorithm, initial MIP exclusion and residual noise are expected to be small but will be addressed for the final paper.

5 Results

In the following observables on the interaction region and on secondary particles as obtained in beam test data are compared with simulations based on two GEANT4 physics lists, i.e. QGSP_BERT and FTFP_BERT. According to [5] the data are contaminated after the pre-selection with 8.8% double-π events at 2 GeV and as low as 1.5% at 10 GeV. The Monte Carlo samples have thus been produced with an admixture of double-π events for the comparison with the data. When averaged results are shown, correction factors will be extracted by comparing the results for contaminated samples with those from pure samples. The correction factors will be given by the means calculated from the individual correction factors extracted from the two physics lists. The uncertainty on the correction factors will constitute the systematic error and is given by the difference between the mean corrections factor and the individual correction factors. The correction factors are between 0.93 and 1 and the uncertainties are of the order of a few %. Another source of systematic uncertainty suggested in [6] that may be caused by the uncertainty on the MIP energy scale has been studied and was found to be negligible for the results presented in the following.

5.1 Energy fraction of the interaction region

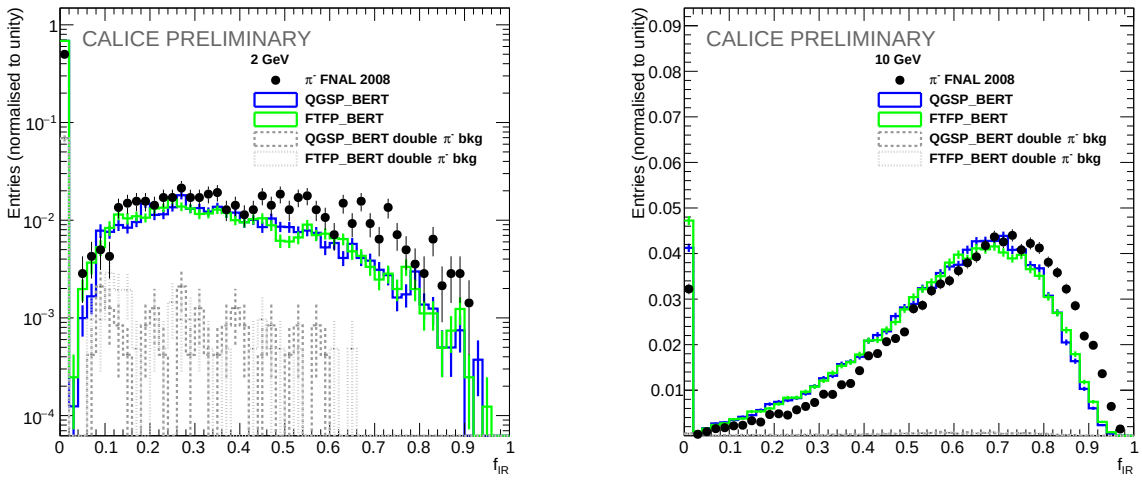
An intuitive estimator to characterise the interaction of the π^- with the absorber material is the fraction f_{IR} of energy deposited in the interaction region E_{IR} over the total energy deposited in the calorimeter $E_{tot.}$.

Hence, f_{IR} is defined as

$$f_{IR} = \frac{E_{IR}}{E_{tot.}} \tag{8}$$

Figure 9 shows comparisons of f_{IR} distributions between data and the two GEANT4 physics lists. The first bin of these histograms corresponds to the fraction of events for which no interaction region is found by the algorithm. The rest of the distribution can be briefly described by a skewed normal distribution. The mean value of f_{IR} is shifted towards larger values in data compared with the Monte Carlo simulation. Qualitatively, this observation may suggest for example a different repartition between visible and invisible energy in data and Monte Carlo.

In Fig. 10 the mean value of f_{IR} is shown as a function of the beam energy for beam energies of 2, 4, 6, 8 and 10 GeV. Events without a detected interaction region according to Sec. 4.1 are discarded. An increase of f_{IR} with increasing beam energy from 43% at 2 GeV to around 64% around is observed. Qualitatively this is expected as the electromagnetic component of the hadronic shower becomes increasingly dominant with increasing energy of the primary particle. In case of the QGSP_BERT physics list the mean value is up to 20% smaller than observed in the data. The FTFP_BERT physics list changes its behaviour above 4 GeV, i.e. at the transition between the Bertini cascade and the Fritiof model bringing the prediction closer to the data. The observed discrepancy between data and the predictions by the GEANT4 physics lists is consistent with an underestimation of the total energy deposition by the Monte Carlo models as reported in [5].



(a) (b)

Figure 9: *Comparison of f_{IR} between data and Monte Carlo simulations for two GEANT4 physics lists for energies of 2 (a) and 10 (b) GeV of the primary particle, respectively. The first bin contains events without a detected interaction region. All histograms are normalised to unity. Error bars represent statistical uncertainties only.*

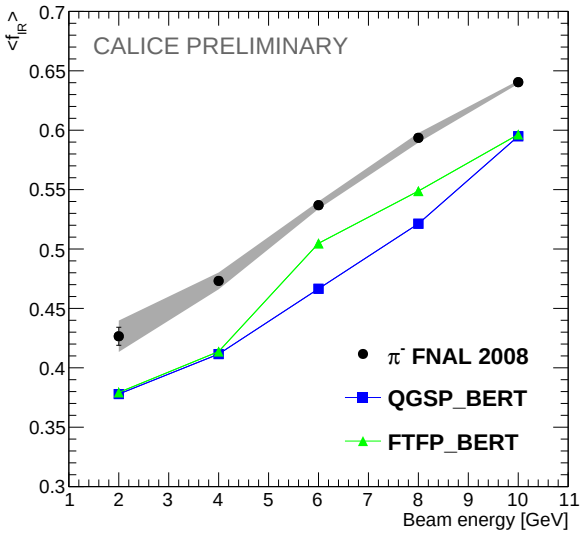


Figure 10: *Mean fraction of energy deposition in the interaction region in Si-W ECAL for data and Monte Carlo simulations for two* GEANT4 *physics lists as a function of the beam energy (2 GeV to 10 GeV). Events without a detected interaction region according to Sec. 4.1 are discarded. The error bars represent statistical uncertainties and the error band the systematic error from the correction for double* π *events.*

5.2 Lateral radius of interaction region

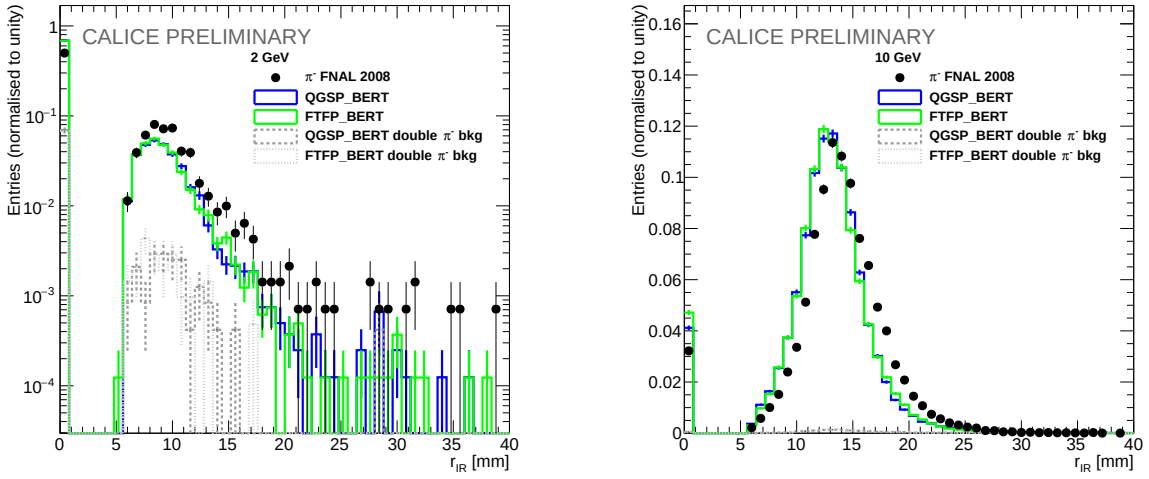
The lateral radius r_{IR} of the detected interaction region averaged over hits with respect to the lateral barycentre is a measure of the spatial extension of the interaction region. It is defined as:

$$r_{IR} = \frac{\sum\limits_{hit \in IR} \sqrt{(\bar{x}_{IR} - x_{hit})^2 + (\bar{y}_{IR} - y_{hit})^2}}{N^{IR}_{hits}}, \tag{9}$$

where the sum runs over the hits in the interaction region, here labeled by IR, and N^{IR}_{hits} is the number of hits in the interaction region. In Eq. 9 $\bar{x}_{IR}$ and $\bar{y}_{IR}$ are the transversal coordinates of the barycentre of the interaction region that in analogy with Eq. 1 are defined as:

$$\bar{x}_{IR} = \frac{\sum\limits_{hit \in IR} x_{hit}\, E_{hit}}{\sum\limits_{hit \in IR} E_{hit}} \text{ and } \bar{y}_{IR} = \frac{\sum\limits_{hit \in IR} y_{hit}\, E_{hit}}{\sum\limits_{hit \in IR} E_{hit}}, \tag{10}$$

Distributions of r_{IR} for data and the predictions by the two tested GEANT4 physics lists are displayed in Fig. 11 for energies of the primary particle of 2 GeV and 10 GeV. In both cases the measured interaction region is wider than the predictions by the GEANT4 physics lists. Figure 12 displays the dependence of the mean r_{IR}, $\langle r_{IR} \rangle$, on the beam energy for the data and the two GEANT4 physics lists. Here again, events without a detected interaction region according to Sec. 4.1 are discarded. The lateral size of the interaction region increases with increasing energy of the primary particle. For all tested energies the interaction region measured in data is constantly around 10% wider than is the case of the GEANT4 physics lists.



(a) (b)

Figure 11: *Comparison of r_{IR} distributions for data and Monte Carlo simulations for two* GEANT4 *physics lists for energies of the primary particle of 2 (a) and 10 (b) GeV, respectively. All histograms are normalised to unity. Error bars represent statistical uncertainties only.*

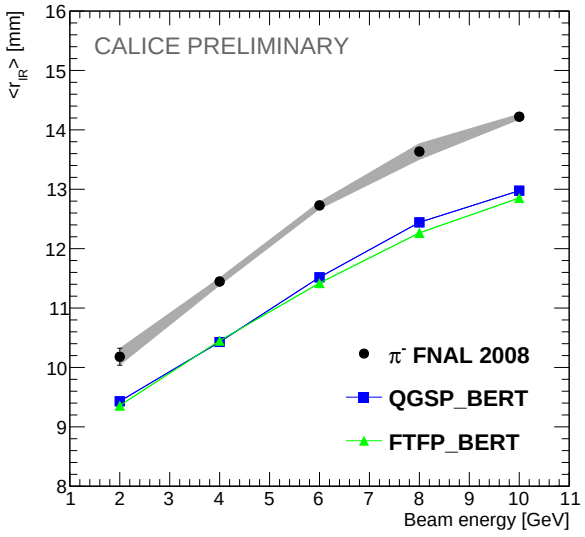


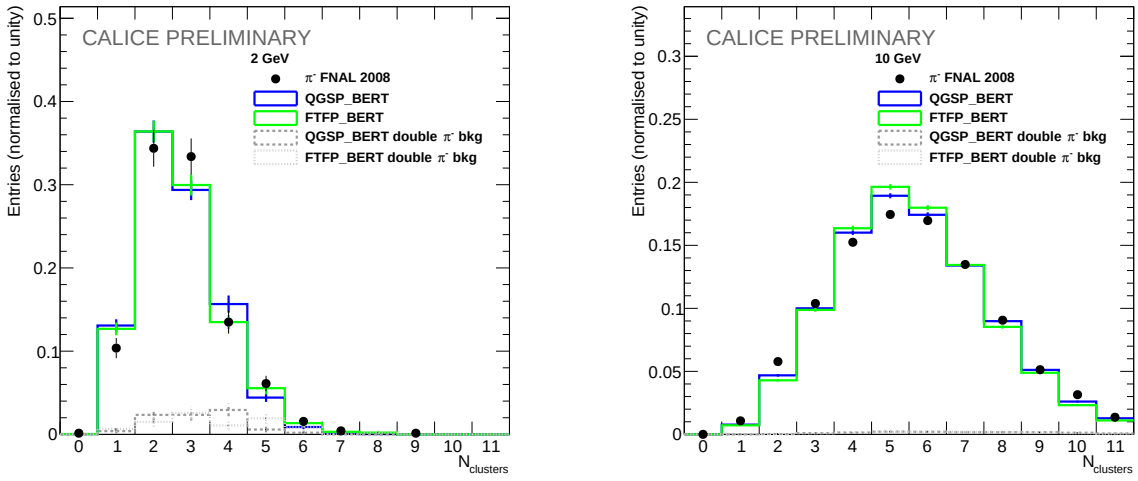
Figure 12: *Mean r_{IR} for data and Monte Carlo simulations for two* GEANT4 *physics lists as a function of the beam energy. Events without a detected interaction region according to Sec. 4.1 are discarded. Error bars represent statistical uncertainties only and the error band the systematic error from the correction for double π events.*

5.3 Number of clusters

As the final tracks are composed from segments that are given by clusters according to Sec. 4.2, it is instructive to study the total number of clusters ($N_{clusters}$) detected by the track-finding algorithm in the event. This observable is stable against details of the track-finding algorithm since it does not depend neither on the ε parameter value nor on other free parameters of the classification algorithm. Here and in all of the following events without a detected interaction region according to Sec. 4.1 are discarded. The $N_{clusters}$ distribution is given in Fig. 13 for data and Monte Carlo simulation for the two GEANT4 physics lists for energies of 2 and 10 GeV of the incoming π^--meson, respectively. The measured distributions are slightly shifted towards higher values with respect to those obtained for the two GEANT4 physics lists.

Figure 14 shows the dependence of $\langle N_{clusters} \rangle$ on different beam energies for data, FTFP_BERT and QGSP_BERT Monte Carlo simulations. At all energies the data are systematically above the Monte Carlo

predictions with deviations of up to 7%. The agreement tends to improves with increasing energy of the primary particle and is best at 10 GeV.



(a) (b)

Figure 13: *Comparison of the number of clusters found between data and Monte Carlo simulations for two* GEANT4 *physics lists for energies of the primary particle of 2 (a) and 10 (b) GeV, respectively. Events without a detected interaction region according to Sec. 4.1 are discarded. Error bars represent statistical uncertainties only.*

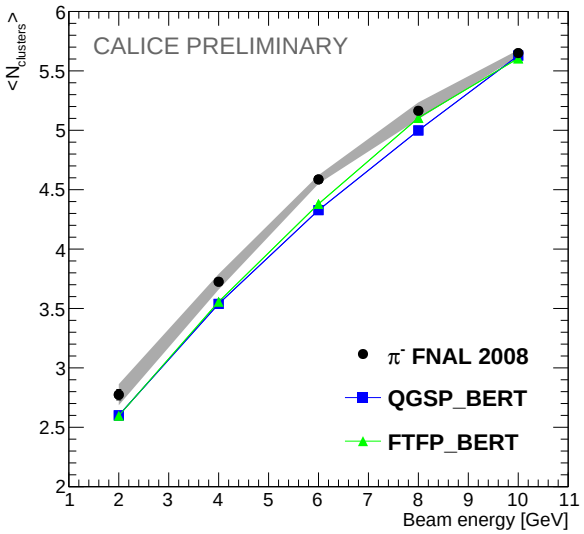


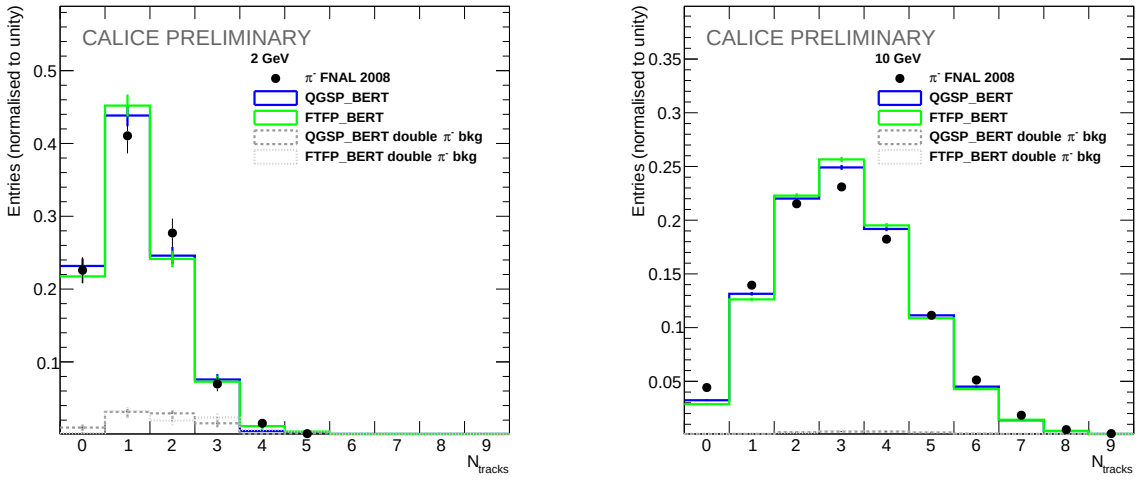
Figure 14: *Mean number of clusters in the Si-W ECAL for data and Monte Carlo simulations for two* GEANT4 *physics lists as a function of beam energy (2 GeV to 10 GeV). Events without a detected interaction region according to Sec. 4.1 are discarded. Error bars on the graph represent statistical uncertainties and the error band the systematic error from the correction for double π events.*

5.4 Number of tracks

A central result of the track-finding algorithm is the number of secondary tracks (N_{tracks}) and observables based on their properties. The N_{tracks} distributions are given in Fig. 15 for data and Monte Carlo simulations based on the two tested GEANT4 physics lists for energies of 2 and 10 GeV of the incoming π^--mesons. A remarkably good agreement between data and both physics lists can be reported, given the fact that this observable is analysed for the first time in the Si-W ECAL.

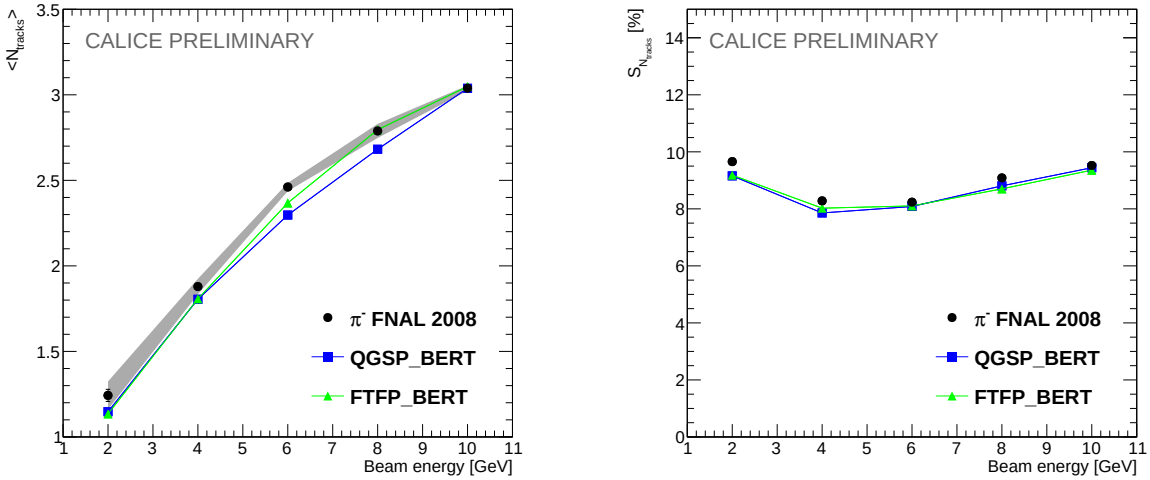
Figure 16a shows the dependence of $\langle N_{tracks} \rangle$ on the beam energy for data and the two GEANT4 physics lists. Both physics lists, presented in Fig. 16a underestimate the number of secondary tracks by 7% on average below 10 GeV and are in agreement with the data at 10 GeV.

The sensitivity to the ε parameter defined by Eq. 7 for $\mathcal{O} = N_{tracks}$, $\varepsilon_1, = 0.04$, $\varepsilon_2, = 0.02$ and $\varepsilon_{nom.} = 0.03$ is shown in Fig. 16b. Within the chosen range the number of reconstructed tracks varies by about 10% for both, data and the two GEANT4 physics lists.



(a) (b)

Figure 15: *Comparison of the number of secondary tracks between data Monte Carlo simulations for two* GEANT4 *physics lists for energies of the primary particle of 2 (a) and 10 (b) GeV. Events without a detected interaction region according to Sec. 4.1 are discarded. Error bars represent statistical errors only.*



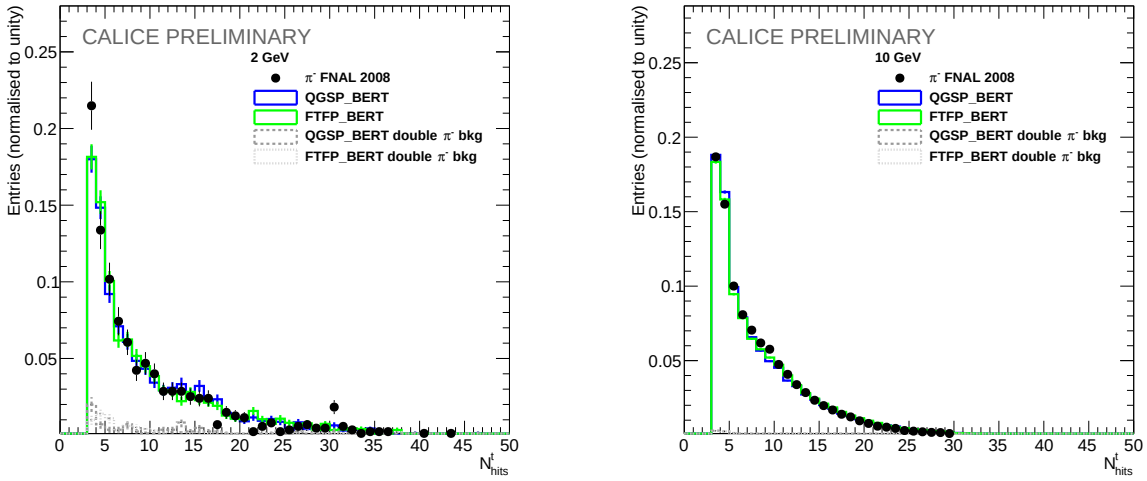
(a) (b)

Figure 16: *Mean number of secondary tracks $\langle N_{tracks} \rangle$ for $\varepsilon = 0.03$ (a) and the corresponding sensitivity according to Eq. 7 of $\langle N_{tracks} \rangle$ on the ε parameter (b) for data and Monte Carlo simulations for two* GEANT4 *physics lists as a function of the beam energy (from 2 to 10 GeV). The sensitivity, see Eq. 7, is estimated by the mean difference in N_{tracks} for $\varepsilon_1 = 0.04$ and $\varepsilon_2 = 0.02$ normalised to the result for $\varepsilon_{nom} = 0.03$. Events without a detected interaction region according to Sec. 4.1 are discarded. Error bars represent statistical errors and the error band the systematic error from the correction for double π events.*

5.5 Number of hits per track

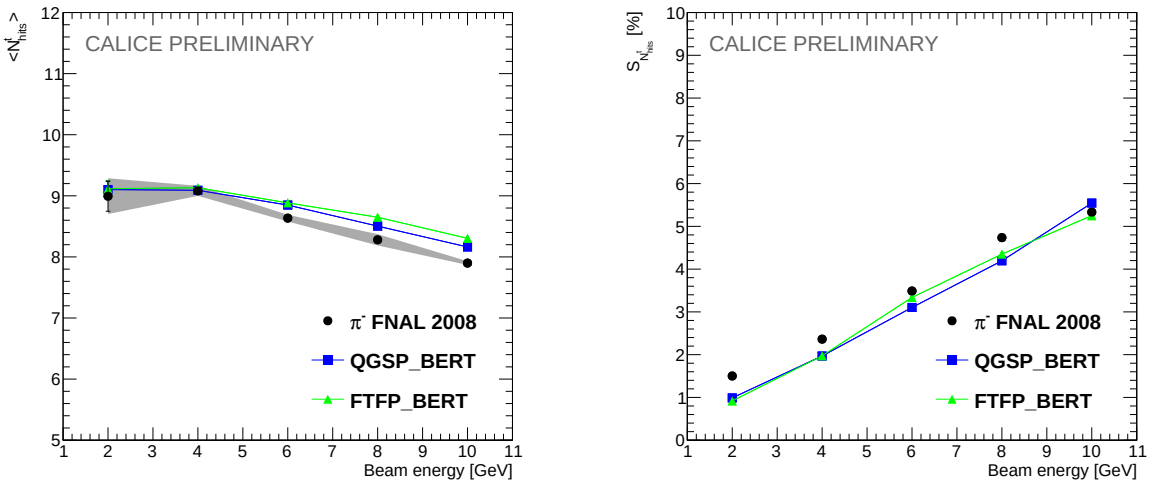
The number of hits per track N^t_{hits} is an essential feature to characterise the reconstructed tracks. The histograms of N^t_{hits} for 2 and 10 GeV beam energy are shown in Fig. 17. The distributions obtained for data and Monte Carlo agree in many bins within statistical errors and are therefore in good overall agreement with each other.

Figure 18a shows the dependence of $\langle N^t_{hits} \rangle$ on the beam energy for data, FTFP_BERT and QGSP_BERT Monte Carlo simulations. The Monte Carlo models agree with the data within 5%. For energies greater than 4 GeV both models are however systematically above the data. The sensitivity to the ε parameter as defined by Eq. 7 for $\mathcal{O} = N_{hits}$, $\varepsilon_1, = 0.04$, $\varepsilon_2, = 0.02$ and $\varepsilon_{nom.} = 0.03$ is shown in Fig. 18b. For the chosen parameter range the sensitivity increases with increasing beam energy from 1% to about 5% for both, data and the two GEANT4 physics lists.



(a) (b)

Figure 17: *A comparison of the number of hits per reconstructed track between data and Monte Carlo simulations for two* GEANT4 *physics lists or energies of the primary particle of 2 (a) and 10 (b) GeV, respectively. Events without a detected interaction region according to Sec. 4.1 are discarded. The distributions are normalised to unity. Error bars represent statistical uncertainties only.*



(a) (b)

Figure 18: *Mean number of hits per reconstructed track $\langle N^t_{hits}\rangle$ for $\varepsilon = 0.03$ (a) and the corresponding sensitivity according to Eq. 7 of $\langle N^t_{hits}\rangle$ on the ε parameter (b) for data and Monte Carlo simulations for two* GEANT4 *physics lists as a function of beam energy (from 2 to 10 GeV). Events without a detected interaction region according to Sec. 4.1 are discarded. The sensitivity, see Eq. 7, is estimated by the mean difference in N^t_{hits} for $\varepsilon_1 = 0.04$ and $\varepsilon_2 = 0.02$ normalised to the result for $\varepsilon_{nom} = 0.03$. Error bars represent statistical errors and the error band the systematic error from the correction for double π events.*

5.6 Angular distributions

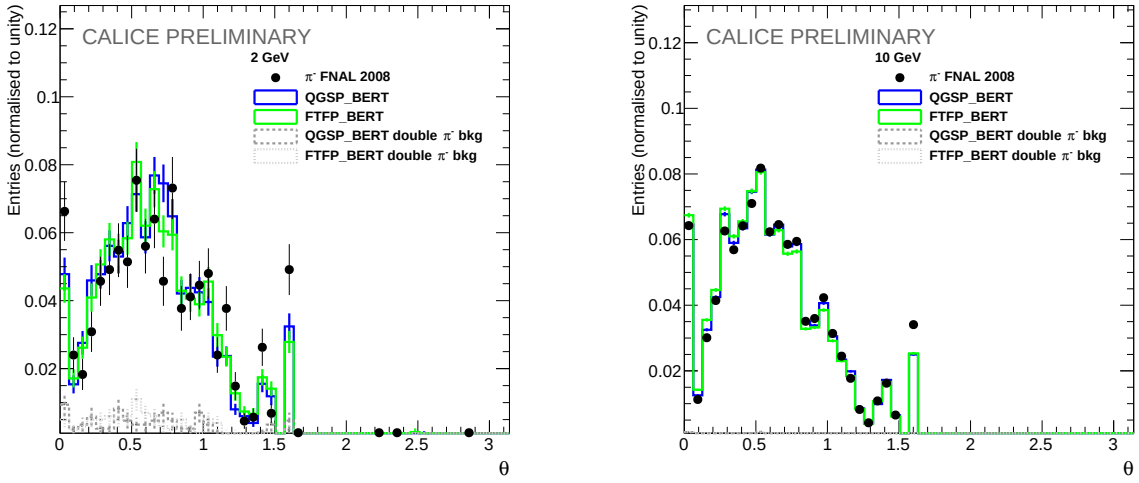
Due to the high granularity of the Si-W ECAL further tracking observables as the polar (θ) and azimuthal (ϕ) angles of secondary tracks become available. Without a magnetic field, the secondary particles from hadronic interaction undergo only multiple elastic scattering in the detector material. Therefore, the direction of the initial momentum coincides approximately with the direction of the track that is visible in the Si-W ECAL. Both angles are measured with respect to the z-axis in the right-handed coordinate frame defined in Sec. 3.1. The track direction is calculated from the position of the first and last hit of the track along the z-axis.

Figures 19 and 20 display histograms of the θ and ϕ angles, respectively, for 2 and 10 GeV data and simulations based on the FTFP_BERT an QGSP_BERT physics lists.

When corrected for the staggering of the detector layers in x-direction [4], the pad coordinates of the Si-W ECAL define a grid with a step width of about 1 cm in lateral direction. This leads to a discretisation of the measured track direction and, therefore, to the ϕ and θ distributions in Figs. 19 and 20. In particular, ϕ, angles that are a multiple of $\phi/4$ are privileged. For 2 and 10 GeV beam energy both GEANT4 physics lists produce tracks with a similar angular distribution and both reproduce the measured distributions adequately, which gives evidence that the Si-W ECAL geometry is correctly implemented into the Monte Carlo simulation.

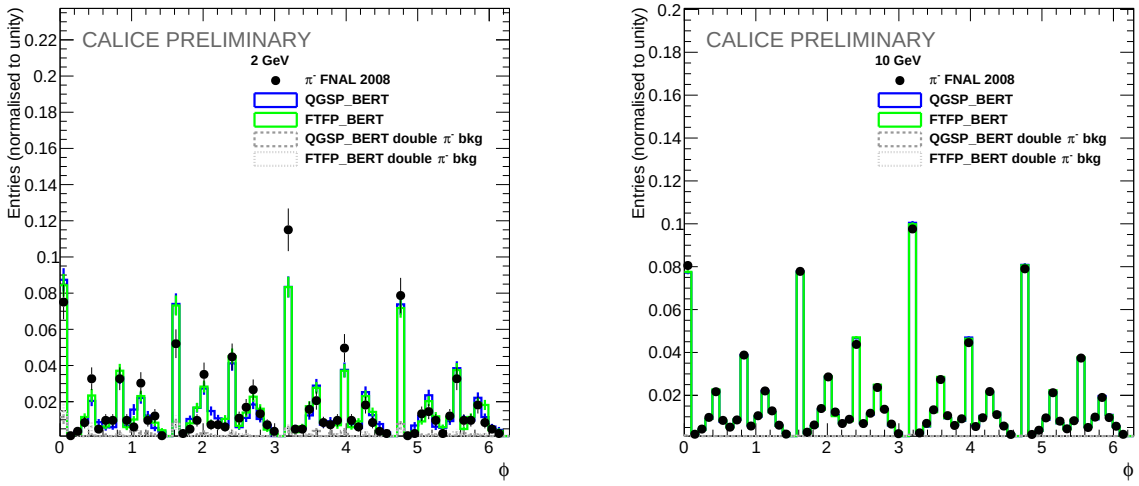
The mean θ angle, $\langle\theta\rangle$, that can be interpreted as a measure of the collimation of the secondary particles, as a function of the beam energy is shown in Fig. 21a. It can be seen that $\langle\theta\rangle$ has only a weak dependence on the beam energy but shows the tendency to decrease as expected due to the increase of the boost transferred to the secondary particles. The data are reproduced within a few % by the Monte Carlo simulations. However, in case of the QGSP_BERT physics list the collimation features a step between energies of the primary particle of 8 GeV and 10 GeV, i.e. at the transition between Bertini and LEP cascades. It seems also that the curve for FTFP_BERT physics flattens out above 4 GeV beam energy, corresponding to the transition between Bertini and Fritiof models.

The sensitivity to the ε parameter as defined by Eq. 7 for $\mathcal{O} = \theta$, $\varepsilon_1, = 0.04$, $\varepsilon_2, = 0.02$ and $\varepsilon_{nom.} = 0.03$ is shown in Fig. 21b. In both cases the sensitivity is between 1.5% and 3%.



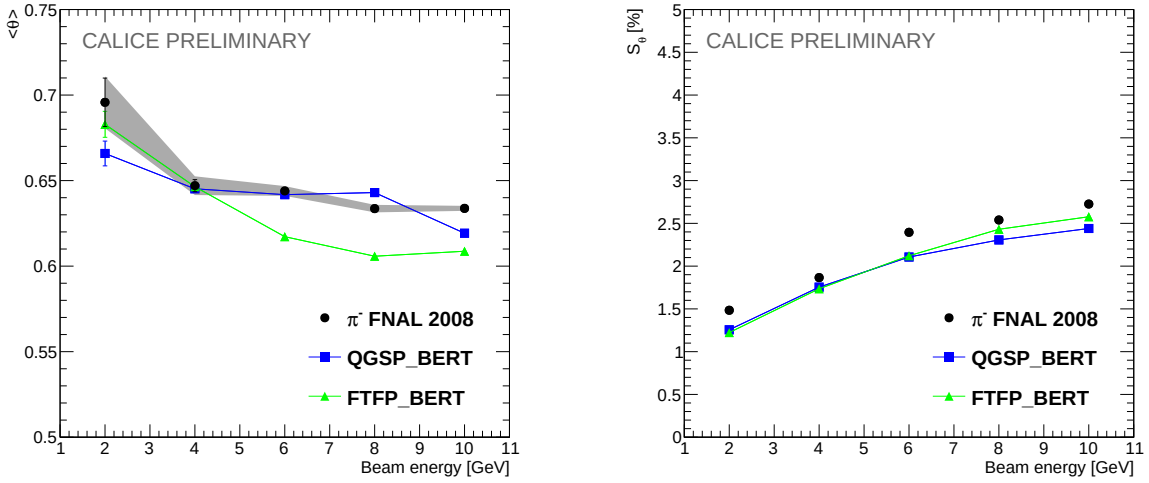
(a) (b)

Figure 19: *Comparison of the polar angle θ of secondary tracks found between data and Monte Carlo simulations for two* GEANT4 *physics lists for energies of the primary particle of 2 (a) and 10 (b) GeV, respectively. Events without a detected interaction region according to Sec. 4.1 are discarded. Error bars represent statistical uncertainties only.*



(a) (b)

Figure 20: *Comparison of the azimuthal angle ϕ of secondary tracks between data and Monte Carlo simulations for two* GEANT4 *physics lists for 2 (a) and 10 (b) GeV beam energies. Events without a detected interaction region according to Sec. 4.1 are discarded. Error bars represent statistical uncertainties only.*



(a) (b)

Figure 21: *Mean polar angle $\langle\theta\rangle$ of secondary tracks for $\varepsilon = 0.03$ (a) and the corresponding sensitivity according to Eq. 7 of $\langle\theta\rangle$ on the ε parameter (b) for data and Monte Carlo simulations for two* GEANT4 *physics lists as a function of beam energy. The sensitivity, see Eq. 7, is estimated by the mean difference in θ for $\varepsilon = 0.04$ and $\varepsilon_{nom} = 0.02$. Events without a detected interaction region according to Sec. 4.1 are discarded. Error bars represent statistical errors and the error band the systematic error from the correction for double π events.*

A further discussion on the relationship between the ε parameter, the polar angle θ and the track length l can be found in Appendix B.

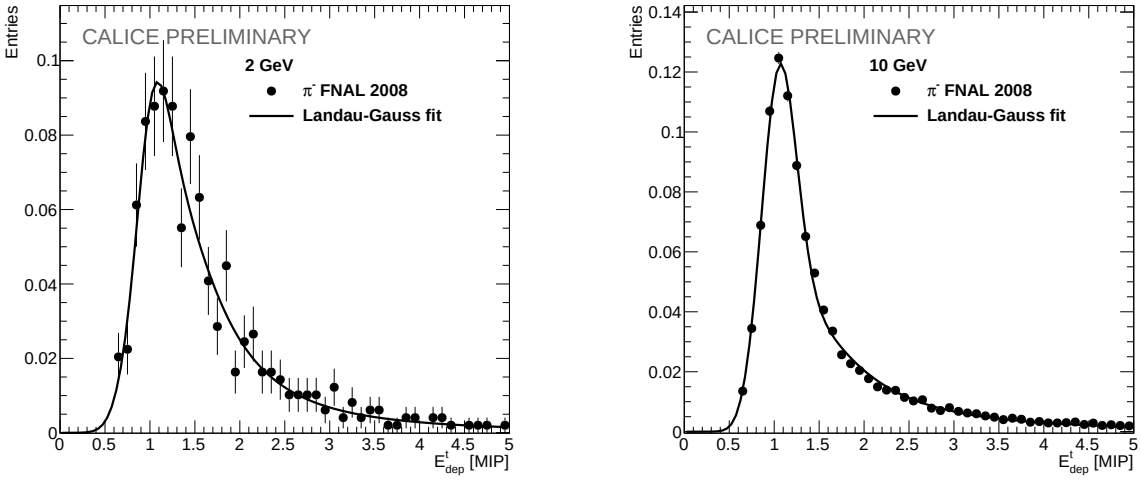
5.7 Energy deposition by secondary tracks

At energies relevant for this study, the secondaries that create sizeable tracks, cross the detector volume as minimal ionising particles. This fact may be exploited for an in-situ calibration of the detector or at least for a monitoring of the response of individual detector regions. For this purpose the following additional selection cuts are applied:

- The events are required to have more than one track and an interaction region to suppress soft inelastic scattering interaction for lower energies;
- The reconstructed tracks should have a length $l \geq 8$ and $l/N_{hits} > 0.9$ to select long pencil-like tracks;
- the reconstructed tracks should have an polar angle $\theta < 0.3$ to reduce the angular dependence of the energy depositions.

The energy deposition by secondary tracks E^t_{dep} using data for 2 and 10 GeV beam energy is displayed in Fig. 22. Both distributions peak at around 1 MIP as expected for straight MIP like tracks. Overlaid is a fit of the convolution of a Landau and a Gaussian that approximates well the measured distribution. However, as can be already inferred from Fig. 22a, the tighter selection criteria reduce considerably the statistics at 2 GeV. As a consequence, the uncertainty on the fit is large for the 2 GeV sample. The data at 2 GeV are thus discarded in the following.

Figure 23 presents the dependence of the most probable value (MPV) of the energy deposition in secondary tracks on beam energy. It can be seen that the detector response is uniform within 1-2% over the energy range of the primary particles in data and that also the energy deposition by the tracks is reproduced by the Monte Carlo simulations within 1-2%.



(a) (b)

Figure 22: *Histograms of the energy deposition in secondary tracks for the data with 2 (a) and 10 (b) GeV beam energies. The spectra are fitted by the convolution of a Landau and a Gaussian. Events without a detected interaction region according to Sec. 4.1 are discarded. Error bars represent statistical uncertainties only.*

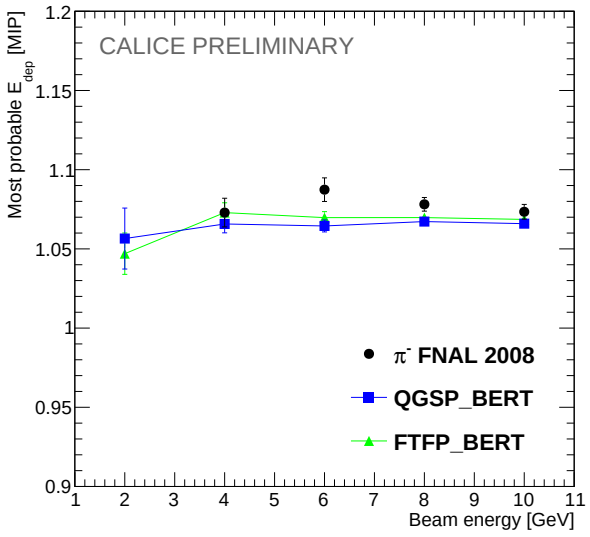


Figure 23: *MPV of the Landau fit to the E^t_{dep} distributions of the pencil-like secondary tracks as a function of the beam energy for π^- data and two Monte Carlo samples. The MPV point the of 2 GeV data sample is excluded because of the small statistics left after selection. Events without a detected interaction region according to Sec. 4.1 are discarded. Error bars represent the statistical fit uncertainty.*

This result is not trivial. It shows that the algorithm has indeed selected MIP-like secondary tracks since the MIP scale is expected to be independent of the underlying physics lists and the detector response should be largely independent of the energy of the primary particle.

6 Summary and outlook

This study reveals the outstanding potential of the CALICE Si-W ECAL physics prototype to obtain a detailed picture of the interactions of hadrons with matter. This note describes basic ideas and the application of a new simple track-finding algorithm for the Si-W ECAL. This algorithm allows for the reconstruction of tracks produced by secondary particles created in the interaction of hadrons with the absorber material, and

hence to study the interaction region of hadronic showers in the Si-W ECAL. The track-finding algorithm produces a new set of differential observables, based on reconstructed tracks of secondary particles and the interaction region of the hadronic cascades. The results are stable w.r.t. small variations of the main parameter of the track-finding algorithm.

Data recorded in test beams at FNAL in 2008 with pions as primary particles with energies between 2 and 10 GeV are compared with predictions by the physics lists QGSP_BERT and FTFP_BERT as contained in GEANT4 version 9.6 p01. The accuracy with which the simulation describes the data varies with the beam energy and the chosen physics observable. In most of the cases data and Monte Carlo agree within 10% without revealing a clear preference for one of the chosen physics lists. In this context it is worthwhile to remind that the interaction region is systematically 10% wider than s the case for the Monte Carlo simulation.

The largest source of discrepancy between data and Monte Carlo is the energy and radius of the interaction region. The measured energy deposition in the interaction region is up to 20% higher than predicted by the Monte Carlo simulation. The distributions of the number of secondary tracks and the number of hits per track for data are well described by the used physics lists. The polar angles of reconstructed tracks in the Monte Carlo simulation agree with data within 8% on average and the distribution azimuthal angles is well reproduced by the Monte Carlo simulations even in view of the non trivial detector geometry.

With respect to a more general outlook future work should aim at transferring the insights about the interaction region and the secondaries emerging from it to the optimisation of Particle Flow Algorithms.

A tighter track selection leads to clean MIP-like tracks. The detector response is stable to about 1-2% over the tested energy range with an expected good agreement with Monte Carlo simulations. This observation can be exploited in the future as a starting point for a study on the possibility of an in-situ calibration or at least a regular monitoring of the detector by means of the selected tracks.

References

[1] V. Morgunov, "Energy flow method for multi - jet effective mass reconstruction in the highly granulated TESLA calorimeter" *eConf* **C010630** (2001) E3041.

[2] J.-C. Brient and H. Videau, "The Calorimetry at the future e+ e- linear collider" *eConf* **C010630** (2001) E3047, arXiv:hep-ex/0202004 [hep-ex].

[3] M. Thomson, "Particle Flow Calorimetry and the PandoraPFA Algorithm" *Nucl.Instrum.Meth.* **A611** (2009) 25–40, arXiv:0907.3577 [physics.ins-det].

[4] **CALICE**, J. Repond *et al.*, "Design and Electronics Commissioning of the Physics Prototype of a Si-W Electromagnetic Calorimeter for the International Linear Collider" *JINST* **3** (2008) P08001, arXiv:0805.4833 [physics.ins-det].

[5] **CALICE**, B. Bilki *et al.*, "Testing hadronic interaction models using a highly granular silicon–tungsten calorimeter" *Nucl. Instrum. Meth.* **A794** (2015) 240–254, arXiv:1411.7215 [physics.ins-det].

[6] **CALICE**, C. Adloff *et al.*, "Track segments in hadronic showers in a highly granular scintillator-steel hadron calorimeter" *JINST* **8** (2013) P09001, arXiv:1305.7027 [physics.ins-det].

[7] **CALICE**, "Tracking within Hadronic Showers in the SDHCAL prototype using Hough Transform Technique" *CALICE Analysis Note CAN-047* . https://twiki.cern.ch/twiki/pub/CALICE/CaliceAnalysisNotes/CAN-047.pdf.

[8] **CALICE**, C. Adloff *et al.*, "Construction and Commissioning of the CALICE Analog Hadron Calorimeter Prototype" *JINST* **5** (2010) P05004, arXiv:1003.2662 [physics.ins-det].

[9] **CALICE**, "Construction and performance of a silicon photomultiplier/extruded scintillator tail-catcher and muon-tracker" *JINST* **7** (2012) P04015, arXiv:1201.1653 [physics.ins-det].

[10] P. Mora de Freitas and H. Videau, "Detector simulation with MOKKA / GEANT4: Present and future" in *Linear colliders. Proceedings, International Workshop on physics and experiments with future electron-positron linear colliders, LCWS 2002, Seogwipo, Jeju Island, Korea, August 26-30, 2002*, pp. 623–627. 2002. http://www-library.desy.de/cgi-bin/showprep.pl?lc-tool03-010.

[11] J. Allison *et al.*, "Geant4 developments and applications" *IEEE Trans. Nucl. Sci.* **53** (2006) 270. http://geant4.web.cern.ch/geant4.

[12] GEANT4 Collaboration, "Reference physics lists". http://geant4.cern.ch/support/proc_mod_catalog/physics_lists/referencePL.shtml.

[13] H. Li, *Higgs Recoil Mass and Cross-Section Analysis at ILC AND Calibration of the CALICE SiW ECAL Prototype*. Theses, Université Paris Sud - Paris XI, Oct., 2009. https://tel.archives-ouvertes.fr/tel-00430432.

[14] P. Doublet, *Hadrons in a highly granular SiW ECAL – Top quark production at the ILC*. Theses, Université Paris Sud - Paris XI, Oct., 2011. https://tel.archives-ouvertes.fr/tel-00657967.

A ε parameter- Number of tracks in data and Monte Carlo simulation

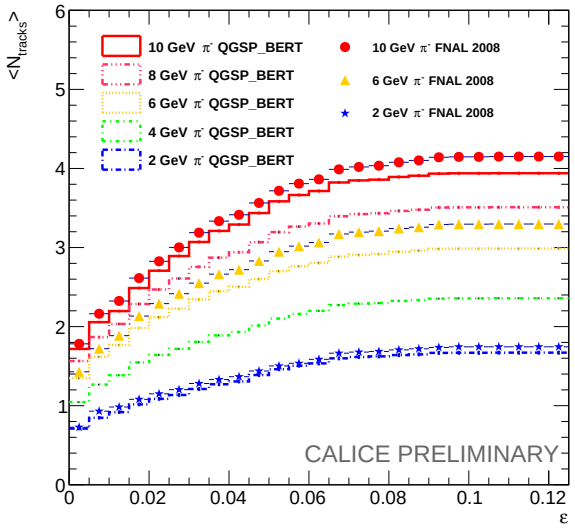


Figure 24: *Mean number of tracks found by the track-finding algorithm as a function of ε values for 2, 4, 6, 8 and 10 GeV beam energy for data (bullets) and for* QGSP_BERT *physics list simulation (lines). Events without a detected interaction region are excluded.*

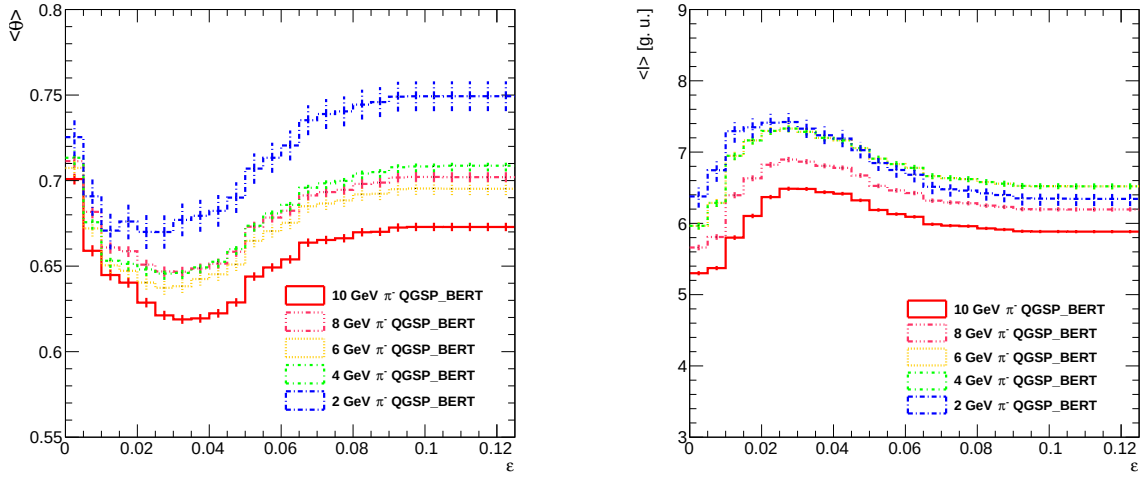
For future reference the Fig. 24 shows a comparison between data and Monte Carlo for the studies presented in Sec. 4.4.

B Polar angle and track length as a function of the ε parameter

Figure 25a presents the variation of $\langle\theta\rangle$ in QGSP_BERT simulation for different beam energies as a function of the ε parameter value. The mean polar angle saturates for large ε. The empirically chosen value of $\varepsilon = 0.03$ is close to a local minimum of the function, therefore, the $< \theta >$ observable is stable against a small variation of the ε value.

The mean track length $< l >$ as function of ε for different beam energies is shown in Fig. 25b. The function has a local maximum around $\varepsilon = 0.03$ and it saturates for large ε. These observations can be explained as follows:

- At $\varepsilon \to 0$ only small pencil-like clusters are considered as tracks. The small tracks can fit the Si-W ECAL pad volume for any direction. Therefore, $< \theta >$ is large;
- If the ε value is slightly above zero, the longer clusters, with some amount of adjacent hits become tracks. These tracks have smaller $< \theta >$ angle due to a fact that the Si-W ECAL physics prototype has 30 layers in depth and 18×18 lateral size.
- With large ε values also spherical shaped clusters are accepted as tracks. These clusters have a small amount of hits, since otherwise, they would be counted as an interaction region. The algorithm can assign any direction for these spherical clusters, resulting in an increase of $\langle\theta\rangle$ and a decrease of the mean track length.



(a) (b)

Figure 25: *Mean θ angle (a) and mean track length l (b) of the tracks found by the track-finding algorithm with different ε parameter values for 2, 4, 6, 8 and 10 GeV beam energy in* QGSP_BERT *simulation. The minimum value of $\langle\theta\rangle$ and the maximum value of the track length is around the empirically chosen value of* 0.03.