UNIVERSITEIT UTRECHT

SIMULATION ANALYSIS OF NEUTRAL PIONS FOR THE FORWARD EM CALORIMETER

A short study of the FoCal-EM detector for the CERN ALICE experiment, using simulated data from single-pion events and $PbPb$ collisions

Amber de Bruijn
January 2015

Simulation analysis of neutral pions for the Forward EM Calorimeter

Thesis prepared as partial fulfilment of the requirements for the degree of Bachelor of Science

BY AMBER DE BRUIJN

UNDER THE SUPERVISION OF DR. IR. MARCO VAN LEEUWEN

UTRECHT UNIVERSITY
FACULTY OF SCIENCE

Department of Physics & Astronomy
Institute for Subatomic Physics

The aim of this project was to study how effectively neutral pions of high momenta and rapidities (η) can be identified with the Forward Calorimeter (ALICE-CERN). Monte Carlo simulations of single-pion and PbPb embedded data were made, considering detector geometry and interaction probabilities. The simulated data was analyzed using energy resolution graphs, invariant mass distributions and efficiency. An improvement of energy division by High Granularity Layer indication would benefit pion identification. The clustering algorithm could be improved for selection of second clusters. For embedded data, the efficiency plateau mean is 60-75% for 3.0<η<3.5 and 55% in the 3.5<η<4.0 range (vs 80-85% for single-particle data). Invariant mass distributions show a pronounced pion peak for all momenta and rapidities. Judging from that and the energy resolution, neutral pions can be identified with reasonable effectivity at high momenta and rapidity.

CONTENTS

1
INTRODUCTION

We want to understand our universe from its smallest building blocks to its most massive astronomical objects. To achieve this, we need to know more about the fundamental relations between particles and the building blocks of particles themselves. Important experimental tools that physicists use to study the constituents and fundamental laws of matter are particle accelerators. In accelerators, particles such as protons and heavy ions are accelerated to high energies and then collided with each other or with stationary targets.

High-energy particle collisions provide us with information about the composition of the particles - and this gives us insight into the smallest constituents of matter, called quarks and gluons. Just after our universe came into existence through the big bang, it is thought to be exclusively composed of a particle soup called the Quark-Gluon Plasma. This state of matter can be studied using accelerators, but only if the machine is able to produce particle collisions of sufficiently high energies.

The accelerator that can reach the highest particle beam energies on the planet today is the Large Hadron Collider, or LHC. It was built at CERN, a research collaboration on the Swiss-French border. The LHC has reported a record particle beam energy of 4 TeV[1], and strives to be capable of producing 6.5 TeV beam energies in 2015[2]. The data from the particle beam collisions in the LHC will be collected in seven different experiments. One of these is called ALICE, and within this experiment the collaboration strives to study the Quark-Gluon Plasma.

There are 18 subdetectors in the experimental hall of ALICE[3] (see Figure # in the FoCal Detector section). During the LHC shutdown period of 2018-2019, a number of ALICE upgrades are scheduled[4]. One of these upgrades is an advanced electromagnetic Forward Calorimeter (FoCal). The object of this thesis is to discuss the performance of the FoCal on one of the possible locations, as far as can be determined by using simulated data.

The FoCal Group Utrecht is concerned with building and improving the FoCal detector. They are optimizing the FoCal performance, within function, locality, financial, and time constraints. The most important purpose of the FoCal is the measurement of direct photons at small angles from the collision point. With this, and the available space at ALICE, in mind, there is a proposition to place the FoCal at either 3.6 or 8 meters from the collision point[5].

This thesis details my Bachelor's research project. The aim of this project was to study how effectively we can identify neutral pion in PbPb events, at high momenta and rapidities. I have studied the FoCal efficiency, energy resolution and space resolution, at the 3.6 meters site, using Monte Carlo simulation data that was provided to me. I will provide some necessary theory first (e.g. about relevant particle theories and simulation methods). A description of my experiment and its results; a discussion of these results; and an overall recommendation will follow in subsequent chapters.

2
PARTICLE PHYSICS

This section contains a short description of the particle physics theory that is relevant for the thesis.

ABOUT ELEMENTARY PARTICLES

All matter that we see around us is made up of elementary particles. Molecules are made of atoms, and atoms are made of neutrons, protons and electrons. While the electron is considered an elementary particle, the neutron and proton are, in turn, built up of particles – called quarks.

The quarks, gluons, electrons, and a few other particles make up a set of elementary building blocks. They might form hadrons, composed of two, three, or even more quarks, and bound together by gluons. The leptons, such as electrons and neutrinos, are produced in interactions and through decays.

There are four fundamental forces; the strong force, carried by gluons; the weak force, electromagnetic force, and the gravitational force. The main difference between quarks and leptons is that quarks are bound together by the strong-force carrying gluons, while leptons only interact via the electromagnetic and weak forces. Also, we usually only observe quarks in bound states, while we can observe leptons individually. The best theory that there currently is about the world of subatomic physics is that of the Standard Model. However, this theory does not give an explanation for everything (i.e. for dark matter and the ratio of matter to antimatter today).

THE BIG BANG AND A PARTICLE SOUP

Shortly after our universe came into existence - that is, for a few millionths of a second after the big bang - the universe consisted of an extremely hot and dense particle soup. This energetic soup, called the "Quark-Gluon Plasma" or QGP, contained the elementary quarks, and the gluons that can "glue" the quarks together[6]. Many particles, like neutrons and protons, are made of bound quark constructions. Aside from the main 'valence' quarks, bound states also contain a strong force field, which consists of virtual photons and quark-anti-quark pairs. Collisions of gluons or quarks can produce leptons and 'interaction particles' like the famous Higgs boson. A free quark cannot usually be observed. In the Quark-Gluon Plasma, however, these fundamental particles were supposedly free to move on their own, at near-light speeds.

To recreate the state of our universe just after the big bang, we use particle accelerators. These are able to propel particles, like ions, forth using magnetic fields. The heavy ions that are used to produce a QCP consist of hundreds of protons and neutrons, and they are smashed into one another at energies of a few teraelectronvolts (TeV or 10^{12} eV). These energies are so high and concentrated that the particles within the ions melt together to form a quark gluon soup.

In the QGP, matter melts into scattered semi-free quarks and gluons. When the plasma cools, almost instantly, the quarks and gluons recombine into ordinary matter that blasts away in all directions. Because there are so many scattered particles, examining properties of the QGP can be difficult. However, the ALICE collaboration at CERN continues studying, and continually innovates the study of, the Quark-Gluon Plasma. Knowledge about this state of matter gives us knowledge about the first moments of our universe, and a better understanding of particle physics in general.

3
THE NEUTRAL PION AND SIMULATIONS

A brief overview of relevant theory will be provided in this section. More specifically, the neutral pion, kinematic variables, simulation techniques and the computer software that was used for this project will be discussed.

THE NEUTRAL PION

The previous section introduced hadrons as the class of quark-composite particles. Hadrons themselves can be divided into three (or more) subclasses: baryons, particles composed of three quarks, and mesons, particles composed of a quark and antiquark. An exotic meson (consisting of four quarks) has been discovered, making up the third subclass, and research is still being done on possible exotic baryons. This research focusses on the neutral pion, which is a meson. All mesons are unstable. The π^0 particles are amply produced in pp or PbPb collisions. A neutral pion usually decays into two photons, but also has other decay modes:

$\pi^0 \rightarrow \gamma + \gamma,$ 98.823%

$\pi^0 \rightarrow \gamma + e^- + e^+,$ 1.174%

Other decay modes only occur <0.003% of the time, and they are not relevant to this project. One of the main goals of the FoCal detector, on which more information is provided in the next section, is to identify the photons from the neutral pion.

Electromagnetic showering

When photons and electrons interact with matter, they produce an electromagnetic shower. Electrons produce secondary photons via Bremsstrahlung, while photons convert to electron-positron pairs. If the secondary particles are of high enough energy, this can happen numerous times. Up to the point where energies are no longer sufficient, the number of particles would always be increasing and the energy per particle always decreasing.

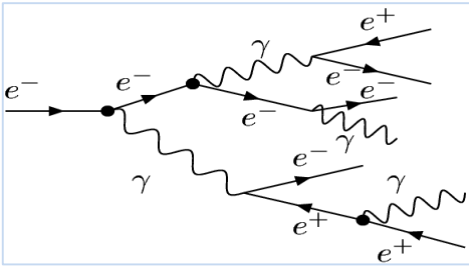

Figure 1: Electromagnetic showering

Higher energy particles have smaller opening angles between their secondary particles. Usually, a shower takes the shape of a cone. Higher energy particles therefore produce deeper 'shower cones'. The shower cannot produce more particles when the energy of the individual particles is not high enough to provide mass-energy to an electron or positron.

The so-called radiation length X_0 gives a measure of $\frac{7}{9}$ of the distance that a fraction of $\left(1-\frac{1}{e}\right)$ of the photons has converted into an electron-positron pair. This is also the average distance an electron of high energy needs to travel to have only $1/e$ of its energy left. The radiation length depends on atomic number, mass number, and absorber density (and therefore is a characteristic number of a material). For example, the radiation length of tungsten is approximately 3.5 mm. The FoCal incorporates radiation lengths in its prototype design.

Kinematic variables

This is a short introduction of the concepts of invariant mass, energy, rapidity, and (transverse) momentum. The main task for the FoCal detector is the measurement of photons produced in beam collisions. When this data is collected, the photons are reconstructed to find the original pion energy. Since the purpose of the FoCal is to measure photons at high momenta and high rapidity particles, it is important to have a basic understanding of what these variables signify.

Invariant mass

The invariant mass, or rest mass, of a particle, is a Lorentz invariant characteristic of that particle. In its rest frame of reference, its energy is simply $E = m_0c^2$, where m_0 is the rest mass. In other frames, the mass of a particle is given by:

$$m_0^2c^4 = E^2 - |\boldsymbol{p}|^2c^2 \qquad [1]$$

Experimental physicists commonly use natural units, assuming $c = 1$. The mass of a particle is then given by:

$$m_0^2 = E^2 - |\boldsymbol{p}|^2 \qquad [2]$$

Energy and momentum are conserved during a decay. When a π^0 decays, the invariant mass of two decay photons is therefore equal to the pion rest mass. This means that if photons can be accurately reconstructed to a same pion, then they can be identified as decay photons rather than direct photons.

Photons are massless and therefore the magnitude of their momentum is equal to E. Their four-momentum vector becomes $(E, |\boldsymbol{p}|) \rightarrow (E, E\hat{r})$, with $\hat{r}$ in the direction from the collision point to the photon's measured location. For two photons, the summed four-vector is $(E_1 + E_2,\ E_2\hat{r}_1 + E_2\hat{r}_2)$. Applying this to Equation [2] yields:

$$m_0^2 = (E_1 + E_2)^2 - |E_2\hat{r}_1 + E_2\hat{r}_2|^2$$

$$= 2E_1E_2 - 2E_1E_2(\hat{r}_1 \cdot \hat{r}_2) \qquad [3]$$

$$m_0 = \sqrt{2E_1E_2(1 - \cos(\phi))} \qquad [4]$$

Here, ϕ is the opening angle between the two photons.

Transverse momentum

The momentum of every particle produced in a collision can be decomposed into two components. One is parallel to the beams (the z-direction), and also called longitudinal component p_L , and one is perpendicular to the beams (in the x-y plane), also called the transverse momentum p_T (see Figure 2).

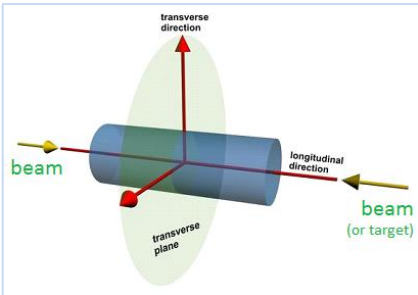

Figure 2: Axis choices: the beam direction compared with the transverse direction.

Longitudinal angles (often denoted θ) can be described with rapidity and pseudorapidity. Rapidity y and pseudorapidity η depend on velocity or longitudinal angle, and are described in the next paragraph.

Rapidity and pseudorapidity

The rapidity is a measure of the rate of motion, and can be calculated from the energy and momentum of a particle. Relativity theories often describe or use the Lorentz factor γ, which relates to the rapidity y as follows:

$$\gamma = \frac{1}{\sqrt{1 - v^2/c^2}} = \cosh y \qquad [5]$$

From this equality, an expression of the rapidity can be derived:

$$y = \frac{1}{2}\ln\left(\frac{E + p_L c}{E - p_L c}\right) \qquad [6]$$

Note that p_L is the longitudinal momentum. For highly relativistic particles, where $|\boldsymbol{p}|c \gg mc^2$, the rapidity is almost equal to pseudorapidity η. This is convenient, because pseudorapidity is much easier to measure. It is defined as:

$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right) \qquad [7]$$

Figure 3 shows approximate pseudorapidities for certain acute angles. Note that the smaller the pseudorapidity in a certain pseudorapidity interval, the greater the angle coverage. I.e., for angles between $80° < \theta < 85°$, the pseudorapidity is $0.18 > \eta > 0.09$, so that $\frac{\Delta\theta}{\Delta\eta} \cong 57$. For $5° < \theta < 10°$, however, the rapidity range is $3.13 > \eta > 2.44$, and $\frac{\Delta\theta}{\Delta\eta} \cong 7$. It is clear that intervals at high rapidities represent smaller angle ranges.

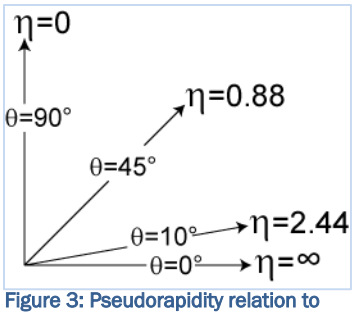

Figure 3: Pseudorapidity relation to angle made with beam direction

For the FoCal, specifically, the pseudorapidity range that is covered depends on the distance of the detector from the interaction point and the inner and outer radii. The minimum and maximum pseudorapidities are simply obtained by filling in the minimum and maximum longitudinal angles, respectively, into the equation above.

The differential cross section (which is used to express the 'likelihood' of interactions and scattering between i.e. photons or electrons and detector material) is roughly the same for every rapidity region that the FoCal covers. Because ever greater rapidities represent ever smaller detector incident areas, the particle density at smaller θ angles is much bigger than the density at large angles.

Additionally, when the pion energy E_{π^0} is large, its decay products' opening angles are small. At high rapidities, energies are higher for the same p_T, so resolving two separate electron showers becomes more difficult.

SIMULATION TECHNIQUES

Monte Carlo simulations

The simulated data that have been used for this project have been produced using a Monte Carlo method. Pions were generated with random energies between $0 - 400$ GeV. Maintaining certain constrictions within the simulations – such as: particles should interact with materials as they normally would – ensures that the final data are as representative as possible. During a beam collision, many different particles are created, and there are therefore many coupled degrees of freedom in the system. A Monte Carlo simulation produces a random sampling of the expected particles.

After the generation of the particles, they are transported through every geometric feature of the detector. Probabilities of the occurrence of all relevant physical processes are calculated for every step of the simulation. The physical processes are then, according to their probabilities, randomly applied to the generated particles. The main limitation of this method is the relatively long time it takes to run the full simulation.

Embedded data

An ‘embedded’ simulation consists of data of a certain type, with data of another type 'embedded’ into it. In this experiment, a full PbPb collision event is simulated, and data of a single pion is added to this data. If data is embedded, this means that a known quantity or event is added to a pile of unknowns. The full PbPb collision data incorporates not only π^0 particles, but also a lot of other particles – like direct photons, protons, et cetera. We measure the effect that the PbPb background has on π^0 reconstruction.

4
FoCal Detector

In this section, the Forward Calorimeter – FoCal – will be introduced. Calorimeters are energy detectors, and can be used to detect electromagnetic or hadronic particles. The electromagnetic FoCal will help to identify photons from heavy ion collisions at ALICE, CERN. The FoCal upgrade in ALICE is not definitive, and therefore more research needs to be done about the capabilities and possibilities of the detector. When π^0 particles decay into two photons ($\gamma\gamma$), the FoCal should be able to accurately detect these photons, even at small opening angles. This is necessary to identify the pions and, later, subtract their energy signature from the total signal to gain information about the direct photons.

Prototype FoCal

The primary goal of the FoCal is to resolve π^0's decay photons. For this, it must both be able to absorb significant energy quantities, and to resolve proximate photons correctly. In order to detect the particles (to receive energy signals), the detector is built up of layers, placed behind one another along the beam axis. To optimize photon shower separation, a small electron shower size is necessary. Therefore, the absorber material between the detection layers is tungsten (or wolfram), which has a small Molière radius and radiation length. The Molière radius is a material constant which quantifies the scale of transverse direction of the electromagnetic showers produced by high energy photons or electrons. The active (detection) layers are mostly Low Granularity Layers (LGLs), with a few High Granularity Layers (HGLs) in between.

High granularity will increase the resolving power of the FoCal. Whereas the LGLs consist of silicon sensor pads of ≈ 1 cm^2, the HGLs have pixel sensors of $100\times100\mu$m^2, summed to 1×1mm^2 macro pixels. The HGLs will most likely use monolithic active pixel sensors (MAPS). The granularity will be higher in the final setup, but memory limitations restrict the granularity in the simulations. The LGL

segments will be longitudinally summed and read out independently. In this project, the FoCal is set up with a total 20 layers; 4 LGLs make up segment 0, the first HGL is segment 1, the next 4 LGLs form segment 2, another HGL makes up segment 3, and finally, 2×5 LGLs make up segments 4 and 5. This setup is shown in Figure 4.

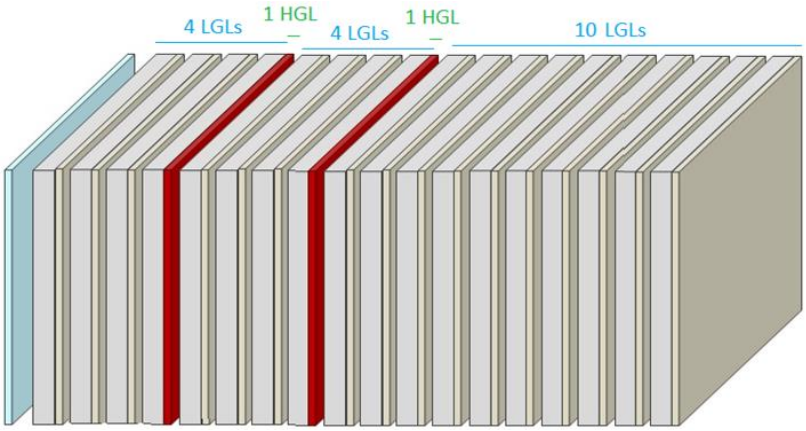

Figure 4: FoCal prototype as used in this project. Particles come into the detector from left to right. Readout points have not been depicted.

At 3.6 meters from the interaction point, the rapidity range that is covered is roughly between $2.2 < \eta < 4.5$. The rapidity is constrained both by the size of the detector (radius of about 80 cm) and by a hole of approximately 8 cm around the beam pipe. This is schematically displayed in Figure 5.

As this project is based on simulated data, the actual behaviour of particles in the detector has not been studied. However, there is substantial knowledge of particle-matter interactions available, and the prototype of the FoCal is well-defined in terms of geometry and detector materials. The Monte Carlo simulation of both single pion and PbPb collision events can therefore be considered representative for 'real' data.

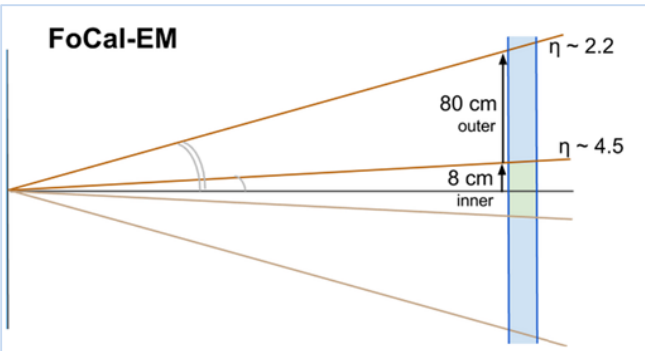

Figure 5: Schematic view of pathway from interaction point to FoCal.

5
PROJECT DESCRIPTION

The aim of this project was to study how effectively we can identify π^0 particles in PbPb events. The analysis that is required for this study consists of a few research questions. Monte Carlo methods were used to generate single-particle and embedded data.

The ALICE experiment already has good detection methods to measure photons at mid p_T, and the FoCal will be used as a way to identify photons at high momenta. It is therefore interesting to see whether the FoCal performs well in high momenta ranges of particles from PbPb collisions. The measurement of particles at large rapidities over a large range of momenta is one of the main observables of the FoCal detector, and will therefore be studied in this simulation analysis of embedded data.

CHOICE OF FOCAL SITE FOR SIMULATION

If the FoCal is implemented into ALICE, the placement of the detector is likely to be 8 meters from the interaction point (or IP) rather than 3.6 meters. However, the simulation data has not yet been generated for the 8 meters site. Also, the main question of this project is how effectively we can identify π^0 particles in PbPb events. If the pions are not discernible at all at the 3.6 meters point, it is not likely that they could be reconstructed at the 8 meters point either. The photon pair production $\gamma \rightarrow e^- e^+$ happens only inside the detector, as do the electron and positron to photon reactions. For a great part, the differences between the sites (when considering electromagnetic shower measurements) are limited to interactions with matter between the IP and the detector, and the rapidity coverage. Earlier studies have shown that the pion energy and mass resolution are similar at the two positions, despite the rapidity coverage. We expect that the results presented here for the 3.6 meters geometry will be close to what one would obtain for the 8 meters position.

OUTLINE OF THE PROJECT

In this project, I will use simulation data of the 3.6 meters site. To determine whether the FoCal is able to identify pions even at high rapidities, most of the results have been split into two or three rapidity regions. The first section of the results contains a short study of the energy resolution at different rapidities and for the two separate clusters that the clustering algorithm finds. The second section of the results concerns invariant mass distributions. These have been made from both datasets. The profile of the distributions, their mean values and mass peak location can be compared between datasets, rapidities and momenta. Finally, the third results section will combine this information to explain efficiency results. All results consists of data plots in combination with a discussion of their possible interpretations.

SIMULATION GENERATION

The Monte Carlo simulated data that was used has been generated by Ing. Čeněk Zach, from the Czech Technical University. The simulation generation was done using the parameters and geometry of the 'normal' FoCal. The generated hits were digitized while considering the eventual FoCal design. That is, the LGLs consist of pads that give an output signal proportional to the deposited energy in the volume.

CLUSTERING ALGORITHM

Note that much of the information about the FoCal workings and about the signal clustering originates from the FoCal Letter of Intent by the ALICE FoCal Collaboration[7]. The clustering method, as outlined below, has been incorporated in the simulation.

In order to accurately cluster the incoming signals, it must be considered that there are some hits in the diffuse tail of the photon shower - hits that are too distant to be immediately obvious - that must also be incorporated in the total signal cluster. If this is not done properly, pions of too low energies will be reconstructed. These diffuse pad signals of low energies must be collected and bunched into clusters, but the central accumulation of hits (of relatively high energies) must also be resolved into two separate showers. The clustering of the generated data is done using a clustering algorithm. This algorithm is the same for LGLs and HGLs.

For each readout segment, cluster 'seeds' are sought from an energy-sorted list of segment digits. The digits must be above a certain minimum energy (SeedThreshold, see Table 1 and 2) to become a seed. Digits that are within a certain radius (MaxRing, see tables) of a higher energy seed are included in the same seed. If the distance between seeds is smaller than a certain separation distance (MinRing), they are merged. Clusters are created, merged, and split based on weights that the seeds assign to the nearby digits. These weights are calculated using a shower shape "weighting function". The shower shape is approximated by a

Cauchy-Lorentz function with a squared exponential tail, and depends mainly on the distance travelled into the detector. The shower shape is calculated using the seed energy and the distance of a digit from the seed. The shower shape parameters are calculated for each segment separately. This is because the segment number indicates the depth into the detector. It also helps to make sure that it remains possible to examine each segment of the detector separately.

In this experiment, parameters were set as in the following tables.

Segment	0	1	2	3	4	5
MinRadius (rings)	1	2	1	2	1	1
MaxRadius (cm)	5	4	4	4	5	5
SeedThreshold (keV)	0	4	4	4	2000	2000
ClusterThreshold (keV)	5000	7	10	10	15000	8500
Weight 1	0.248	0.230	0.526	0.380	1.259	2.164
Weight 2	0.5	0.5	0.5	0.5	0.5	0.5

Table 1: Single-particle pion parameters

Segment	0	1	2	3	4	5
MinRadius (rings)	1	2	1	2	1	1
MaxRadius (cm)	1.5	1.0	1.5	1.0	1.5	1.5
SeedThreshold (keV)	0	4	0	4	2000	2000
ClusterThreshold (keV)	5000	7	10000	10	15000	8500
Weight 1	0.481	0.230	0.526	0.380	1.259	2.164
Weight 2	0.5	0.5	0.5	0.5	0.5	0.5

Table 2: Embedded data parameters

Note that the MinRadius, Weight 1, Weight 2, and the SeedThreshold and ClusterThreshold except in segment 2, are identical between the datasets. The MaxRadius is lower for the embedded data compared to the single-particle data, because it is useful to try to cut some digits that are less likely to be caused by the photon shower. The same reasoning goes for the SeedThreshold and ClusterThreshold, of which the embedding data figures are equal to or greater than those of the single pion event data.

This concludes the cluster finding part of the processing of the generated data. The clusters from the different segments are combined. First, the LGL segments (0, 2, 4, 5) and the HGL segments (1, 3) are combined separately. If a HGL segment indicates that a single cluster on the LGL segments actually consists of two separate clusters, the LGL cluster energy is divided over the two clusters. The energy of these two clusters are partitioned according to the relative cluster energies in the HGL segment. The total energy is retrieved when summing the energies of the matched clusters over all segments. For different analysis purposes, different parameters and algorithms may be used.

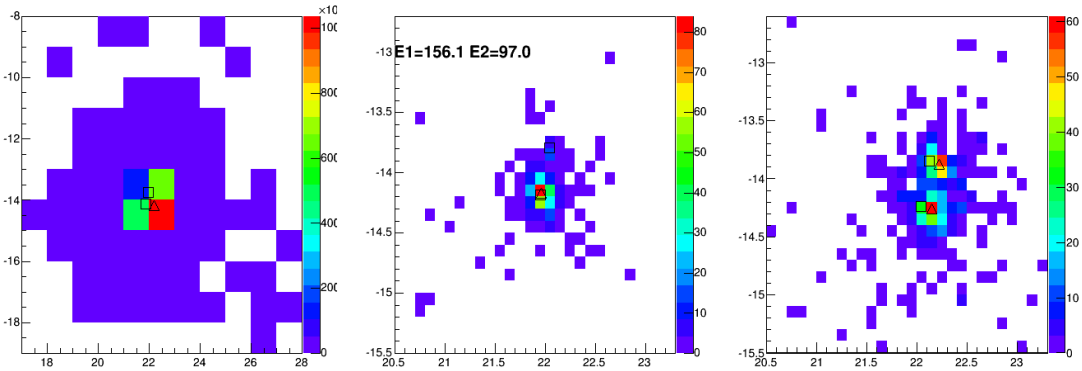

Figure 7: Event display of single-particle simulation. Left: summation of LGL signals. Middle: first HGL. Right: second HGL. X and Y axes are noted in centimeter units. LGLs scaled by keV; HGLs scaled by no. of hits.

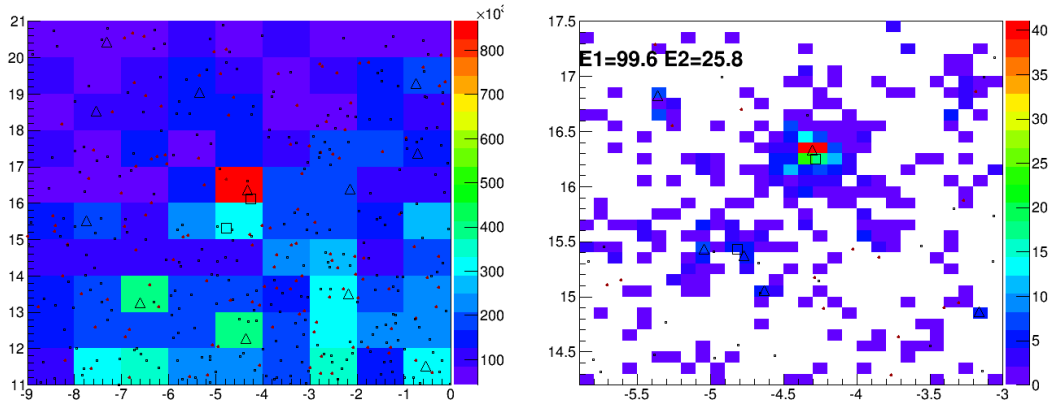

Figure 6: Event display of embedded simulation. Left: summation of LGL signals. Right: second HGL. X and Y axes are noted in centimeter units. LGLs scaled by keV; HGL scaled by no. of hits.

DETECTOR HITS DISPLAY

Figure 6 and 7 show 'event displays' of the single-particle and embedded data, respectively, showing deposited energy in the FoCal. The event displays show found γ clusters, represented by the triangles (△). Two squares (□) represent the real position of the photons. In the embedded data event, a lot of background and/or surrounding energy hits are visible.

In Figure 6, the first graph represents the measured energy intensities per pad, as summed over the LGLs longitudinally (so over 18 low granularity layers). The second graph presents the measured energy intensities of the first high granularity layer, which is the 5th layer of the detector, and the third graph presents the measurements of the second HGL - the 10th layer. The first HGL energy depiction is naturally more compact than the second HGL image. These images are of a random event, of which we can see the first cluster energy is $E_1 = 251.7$ GeV and the second is $E_2 = 104.8$ GeV. The LGL would not (easily) be able to resolve the clusters, as the particles are quite energetic and the clusters are fairly close together. The HGLs can make a distinction and therefore improve the overall resolution efficiency dramatically.

Distinguish photons using shower profile

Electromagnetic showers from photons look different from the signatures of other particles. This is mainly because photons will characteristically deposit (almost) all of their energy into the detector, while hadrons will usually only lose a small part of their energy. Because the hadrons lose only little energy, this energy might not be counted as a seed, or even a digit. The hadrons will not make a wide shower trail. Cuts have been made on the width and energy deposit distribution, to ensure that photons are identified with greater accuracy. The most important variables that quantify the shower profile are the lateral width of the shower and longitudinal distribution of the deposited energy.

When the photons have been identified, neutral mesons can be reconstructed. We want to eventually be able to accurately identify direct photons. Therefore, we need to make sure that we can identify the pion decay photons as accurately as possible - to cut them out and gather good direct photon data later.

Saving and using simulated data

The output for the single-particle simulation data comes in the form of 10,000 events containing single pions, and has been stored in a ROOT TTree. The output for the embedded simulation data comes in the form of 10,000 events containing a full lead-lead (PbPb) collision simulation with a single embedded pion. The analysis results of the simulation data are most easily illustrated using histograms.

Note that cuts have been made on both datasets. Where the clustering algorithm was not able to find a cluster's energy or position, the associated pion was not used in the energy resolution or invariant mass distributions. The efficiency is defined as $\frac{reconstructed\ events}{generated\ events}$. In the 'generated events', the whole dataset is used, while in the 'reconstructed events', I have made the mentioned data selections. Additionally, the 'peak part' of the invariant mass distribution is selected as indication of good pion identification.

6
FoCal Energy Resolution:

Results & Discussion

The energy and space resolution plots in this section provide an indication of the quality of the data and the effectivity of the clustering algorithm. Effects of rapidity and different cluster selections have been analysed for both single-particle and embedded simulation data.

Pion cluster analyses

Figure 8 and 9 show the energy resolution of the single-particle and embedded simulation data, respectively. The top three diagrams display the cluster energy of the first found cluster, for rapidity regions $2.5 < \eta < 3.0$, $3.0 < \eta < 3.5$ and $3.5 < \eta < 4.0$, so over most of the rapidity range. For no-background, perfectly resolved cluster data, the output would look like a "$y = x$" diagram, i.e. a straight 45° line. The above also goes for the lower three diagrams, which represent the cluster energy of the second found cluster. Note that the data points do not extend as far upwards on the bottom graphs. The two simulated decay photons of the pion are ordered by energy, so the particle 1 is always of higher energy than particle 2. The found cluster energies should correspond to this division.

As expected, the single-particle figure shows good energy resolutions. The deviations from the $y = x$ line are not big, not even for the highest rapidities and especially not for the first particle clusters. What is noteworthy in the bottom three graphs is that clusters of too low energies have been selected. This happens especially in the lower particle energy region and only for these (second) clusters. As there is no background in the single-particle simulation data, lower energy resolutions are caused by the clustering algorithm. The more energetic cluster is selected first, after which a second cluster is selected from a distance-sorted list. Of course, the algorithm also uses cluster energy values in this selection. However, it is clear from the second particle resolution that the algorithm might be improved.

A possible solution for this would be to make sure that even distant seeds are selected. This might cause clusters of too high energies to be selected. Further experimenting with the clustering algorithm would be beneficial for the energy resolution and efficiency.

When comparing the single-particle graphs of Figure 8 with the embedded graphs of Figure 9, we can see that the data points are much more dispersed in the last. This is due to background signals, and expected. The dispersion is most evident in the second particles, and at higher energies. For the second particles it is more visible at lower rapidities. The same reasoning about the clustering algorithm as in the previous paragraph on single-particle data can be applied here.

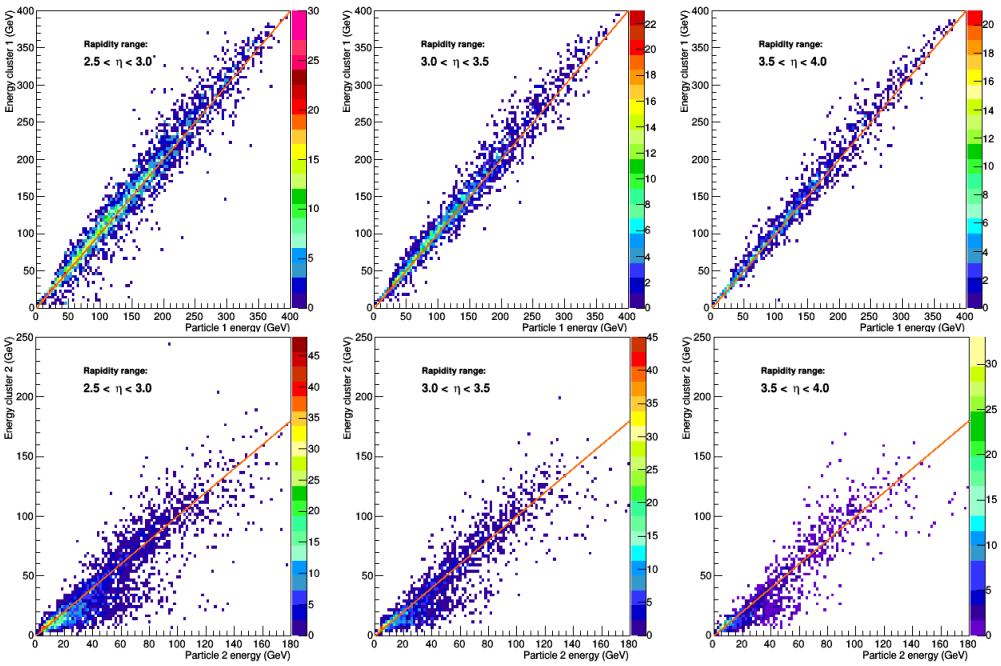

Figure 8: Energy resolution graphs of the single-particle simulation. The top three graphs show particle 1 vs. cluster 1 energies. The lower three graphs show particle 2 vs. cluster 2 energies. From left to right the rapidity increases.

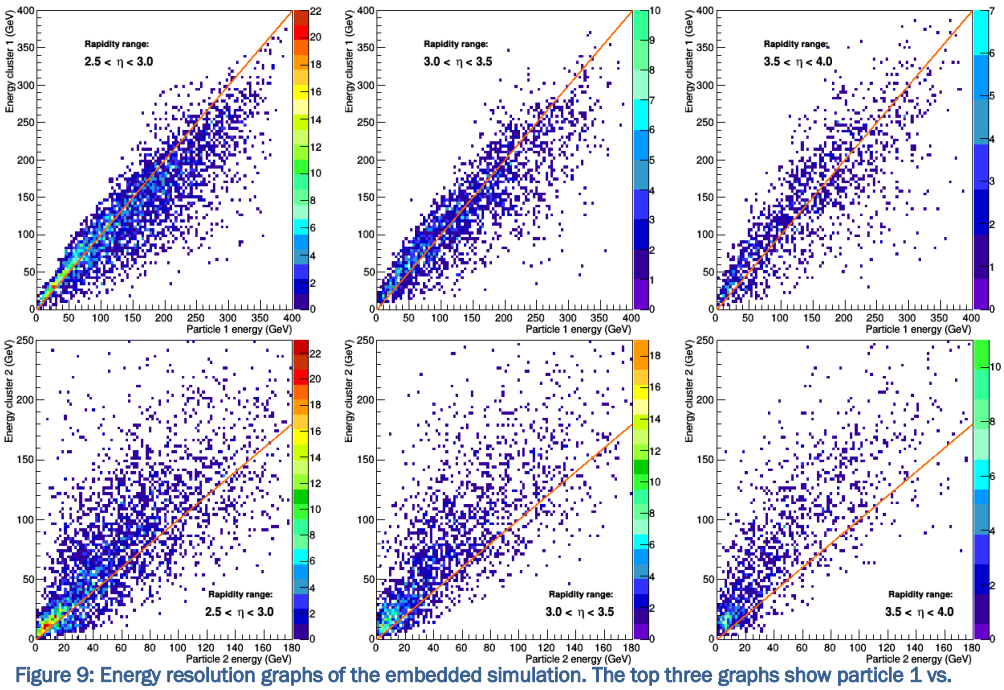

Figure 9: Energy resolution graphs of the embedded simulation. The top three graphs show particle 1 vs. cluster 1 energies. The lower three graphs show particle 2 vs. cluster 2 energies. From left to right the rapidity increases.

In the embedded graphs, it is interesting to see that the second clusters have a distribution that has shifted upwards (above the $y = x$ line), while the first clusters of low rapidities are represented by many data points that lie under the $y = x$ line. Also, the second found clusters' resolution deteriorates significantly at higher energies (the distribution becomes wider). Even the first clusters' energy distributions become wider at higher energies.

The clusters found for the first photon seem to have been chosen at too low energies. As this downwards shift is most visible at high energies, is it possible that the clusters were too energetic to be resolvable (by the low granularity layers). If the high granularity layers were needed to split clusters, and the splitting was not done well, the first particles might have less energetic clusters than expected, while the second particles might have more energetic clusters. If this is the reason for bad energy resolutions, a graph that sets out the generated energy of the π^0 to the combined energy of clusters 1 and 2 should have a better resolution. Figure 10 shows just that for the embedded data.

The left graph of Figure 10, that contains particles of low rapidities, has a clear $y = x$ distribution. Particles that deviate usually do so towards lower cluster energies. The middle and right graphs of Figure 10 also show a narrower energy distribution compared to the graphs in Figure 8 and 9. This indicates that the bad cluster resolutions of Figure 8 and 9 are likely caused by a bad energy division, as done by the clustering algorithm using the high granularity layers.

It looks as though the distributions in Figure 10 shift upwards with increasing rapidity. The data points in the low rapidity graph are neatly located along or just above the 45° line. The data points in the middle graph seem to have shifted to just above this, and the distribution in the highest rapidity graph is steeper still. For low rapidities, the electromagnetic showers of two photons are further apart at the same momenta. Also, the surrounding cluster density is lower. Distant shower hits might not be added to a cluster, accounting for the few data points below the $y = x$ line. As the rapidity increases, cluster density increases and photons have smaller opening angles at the same momenta. There are more surrounding clusters (and cluster energies) available and therefore there is a higher chance of selecting the right clusters. The cluster energy is generally higher in all three rapidity ranges (and especially for the mid and high rapidity data) because of the background signal that is added to the clusters. For the high rapidity graph, the data points are more dispersed at all pion energies. In these regions with high cluster densities, it is more difficult to discriminate between clusters and to assign hits to the right clusters.

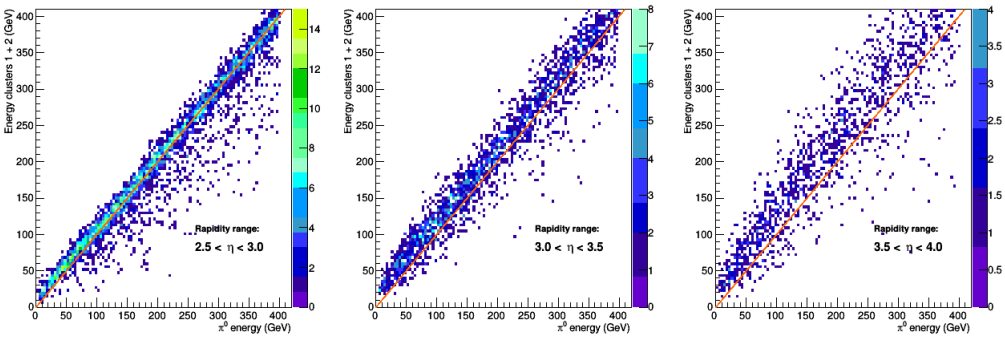

Figure 10: Summed (embedded data) energy resolution plots, showing generated pion energy vs. the clusters 1+2 energy. Sorted by rapidity.

Though the high rapidity graph has a wider distribution, it, too, has a clear, roughly 45°, line in the data distribution (in contrast to the 2nd particle embedded graphs in Figure 9). Clearly, for the most part, the resolution problems in embedded (PbPb) data come from cluster splitting. Other than that, changes to the algorithm might improve second cluster selection. High energy particles are more difficult to resolve, as they have smaller opening angles. High rapidity particles are more difficult to resolve, as they might be indiscernible due to high cluster density. The remaining resolution issues are not unexpected: there is background signal in the embedded data and this will disturb the pion reconstruction efficiency.

Relative energy resolution

It is also interesting to take a look at the usual deviation of the added cluster energies from the pion energy. For this, see Figure 11 and 12 of the single-particle and embedded events respectively. The y axis shows a "ΔE"; formulated as:

$$\Delta E = \frac{Cluster\ energy\ \gamma_1 + Cluster\ energy\ \gamma_2 - \pi^0\ energy}{\pi^0\ energy}.$$

Herein, the cluster energy of "γ_1" and "γ_2" are named as they are for convenience. Keep in mind that π^0 decay occasionally produces an electron-positron pair (as in $\pi^0 \rightarrow \gamma + e^+ + e^-$) rather than two photons (as in $\pi^0 \rightarrow \gamma_1 + \gamma_2$), so that this might actually be a cluster energy of $\gamma + e^-$ or something similar. In any case, the ΔE depends on the first two cluster energies. For a perfectly resolved dataset without detector effects, ΔE would simply be 0.

Any data points below $\Delta E = 0$ indicate a higher generated π^0 energy than the combined energy of clusters 1 and 2 (which we have seen in the previous subsection). These $\Delta E < 0$ data points are most common in the left (rapidity range $2.5 < \eta < 3.0$) graph of the single-particle figure. These fluctuations go down to around $\Delta E = -0.2$. The effect is less apparent in the data with higher rapidities. The clustering algorithm seems to work better for particles of high rapidities and energies. As noted before, this is likely due to large opening angles for particles of low energies, making cluster recognition and splitting more difficult.

At the lowest pion energies, there is a data 'tail' pointing towards $\Delta E = 1$. This is exceedingly apparent in the embedded data plots. In single-particle data, this tail does not get noticeably more or less pronounced in different rapidity regions. As the purpose of the FoCal is not to measure photons at low energies, we might consider cutting out data of under 50 GeV.

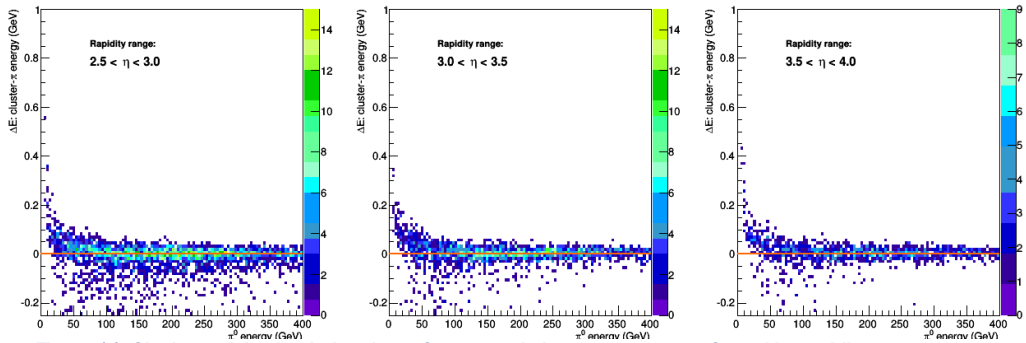

Figure 11: Single-particle simulation data of generated pion energy vs. ΔE. Sorted by rapidity.

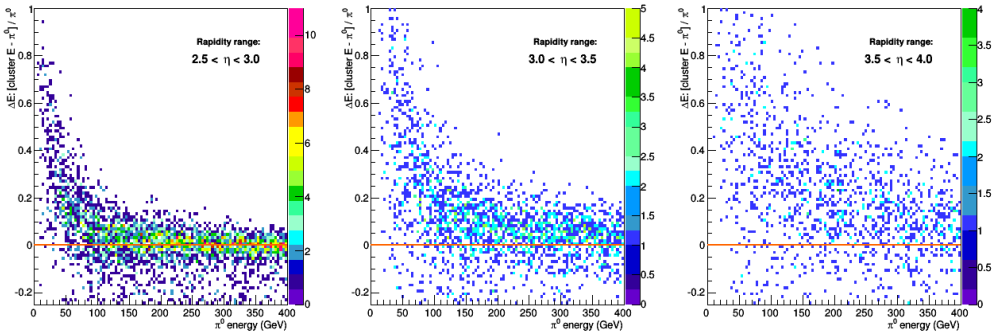

Figure 12: Embedded simulation data of generated pion energy vs. ΔE. Sorted by rapidity.

The embedded data graphs have much wider data distributions. In the left and middle graphs, a similar distribution to that of the single-particle data is discernible. The energy distribution in the right graph is very widespread. In all three graphs, many of the data points lie above the $\Delta E = 0$ line, meaning that the two cluster energies add up to a value higher than the generated pion energy. This is consistent with the expected effect of the background in embedded data. The high rapidity graph of Figure 12 shows a broad distribution, consistent with the distribution in the high rapidity graph of Figure 10.

The low energy upward tail is still visible in the three embedded graphs. The embedded data graphs do seem to show a degeneration of energy resolution in higher rapidity regions. In high (pseudo)rapidity regions, there is a higher particle density, and the resulting resolution difficulty is clearly visible. Though the effects of the PbPb collision background signal are apparent, the influence of this background on the identification of the π^0 is not clear from the plots in this section. This becomes clearer when the pions are actually reconstructed and displayed in an invariant mass distribution.

7

RECONSTRUCTION OF π^0 INVARIANT MASS:

RESULTS & DISCUSSION

All figures in this section have been generated from both the single-particle and the embedded data. The simulation data covers rapidities up to $\eta = 4.0$. Most histograms are shown for both 3.0-3.5 and 3.5-4.0 ranges, so that we might explore possible data degeneration at high rapidities. This degeneration might occur, especially in the embedded dataset, because the particle density is higher in the detector.

INVARIANT MASS DISTRIBUTIONS

The two invariant mass diagrams below have been generated from the single-particle data (Figure 13) and the PbPb embedded data (Figure 14). The first graph has a relatively sharp peak around the pion mass, with the left side of the peak less concave than the right side. The peak of the embedded graph has shifted to the right. The graph also shows significantly more background over the whole mass axis, and more so to the right of the peak compared to the left.

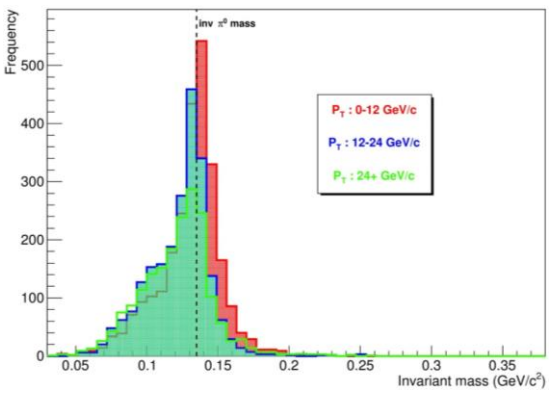

Figure 13: Single-particle simulation data. Figure generated for rapidities 2.5-4.0 and for three transverse momentum selections

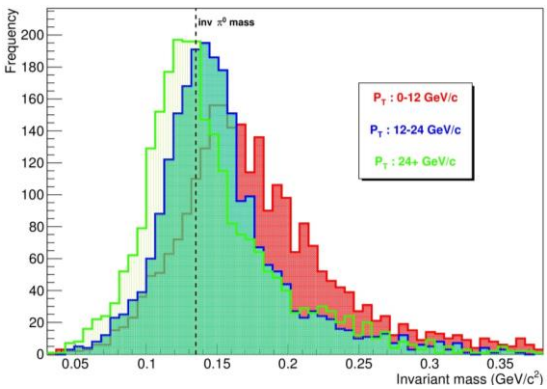

Figure 14: Embedded PbPb simulation data. Generated for rapidities 2.5-4.0 and for three transverse momentum selections

For both datasets, the red graph ($0 < p_T < 12$ GeV/c) is peaked to the right of the dashed π^0 mass line, while the green graph ($p_T > 24$ GeV/c) peaks to the left. The blue graph appears to have a good correspondence with the pion mass. For the embedded data, the lower momentum graph seems to have a greater deviation from the π^0 mass line than the high momenta graph.

If a particle's divergence from the forwards direction is greater, its distance travelled through the detector and beam pipe material also increases. Pions of high momenta that have yet to decay might be stopped prematurely. Photons interact with the material and can convert before the detector, possibly producing more than one shower in the FoCal. Also, since about 90% of proton-nucleus collisions produce pions, the number of background signals will increase tremendously if the pathlength through heavy materials increases. This is hypothesized to be most evident in the PbPb embedded data, where there are also background protons. This corresponds with the findings in the previous chapter.

Particles at high rapidities have higher energies at the same momenta as particles at low rapidities. They are also more difficult to resolve because of high cluster densities at high rapidities. In the embedded dataset, this means that there is a lot of background per unit detector area in these regions. In both datasets, clusters are more difficult to resolve reliably when they are situated close together.

Figure 15 consists of six graphs with data from the single-particle simulations, and can be used to compare not only particles of different momenta, but also of different rapidities. Figure 16 consists of graphs made from the embedded simulation data. The top histograms from each figure use data with rapidities ranging from $3.0 - 3.5$, while the lower histograms are in the $3.5 - 4.0$ range. The embedded data graphs have relatively great peak deviations from the pion invariant mass at low momenta.

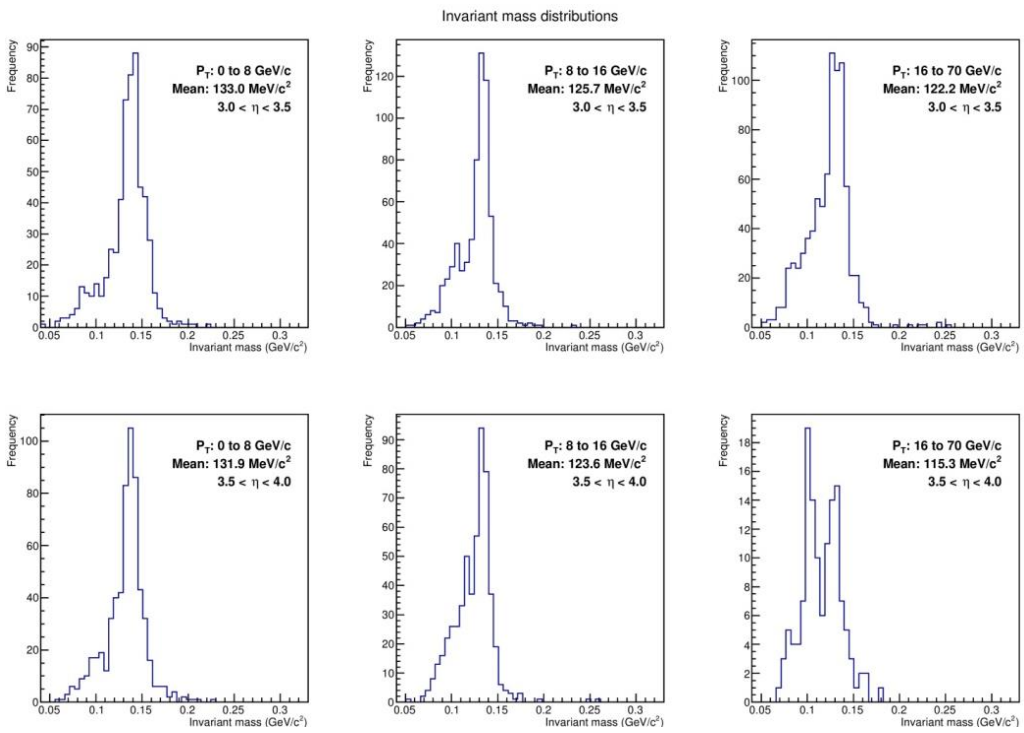

Figure 15: Single-particle data plots of the invariant mass distribution. Sorted per momentum and rapidity range.

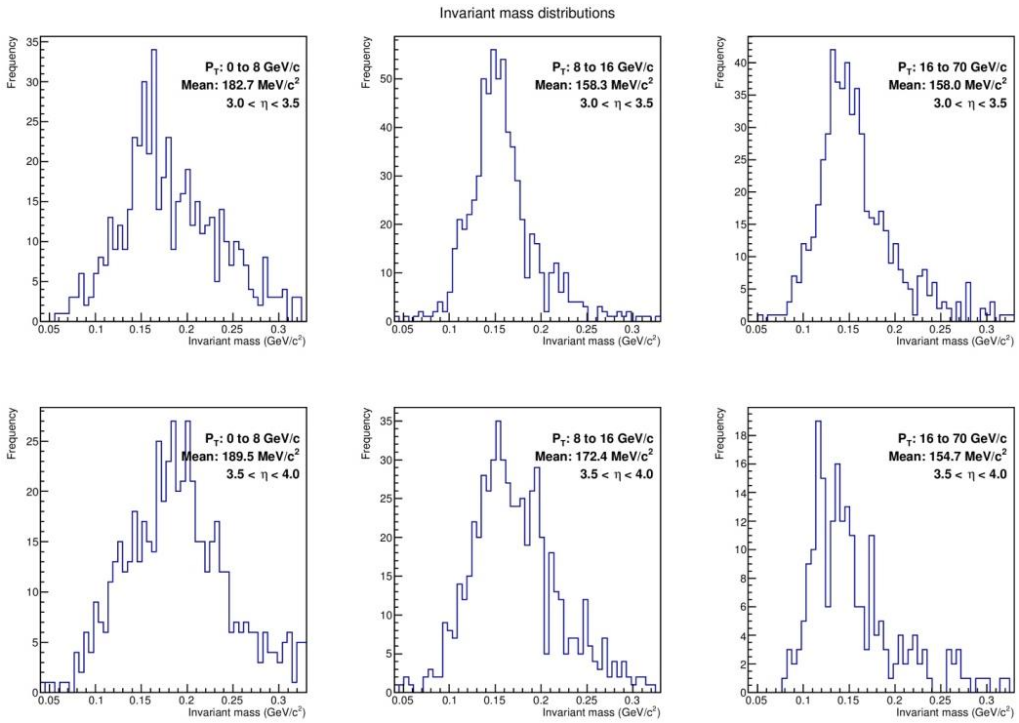

Figure 16: Embedded simulation data plots of the invariant mass distribution. Sorted per momentum and rapidity range.

The single-particle data graphs show small differences between the rapidity ranges. For the $0 < p_T < 8$ GeV/c graphs of single-particle data, there is a $0.8 \pm 1.0\%$ mean decrease in the higher rapidity region compared to the lower region. For the $8 < p_T < 16$ GeV/c graphs, there is a $1.7 \pm 0.9\%$ mean decrease. Neither decrease seems to be substantial or significant. The $p_T > 16$ GeV/c graphs of single-particle data show a $5.6 \pm 1.6\%$ mean invariant mass decrease. Also, a mass peak split is visible in the high rapidity graph. The higher mass peak is roughly located at the pion mass, while the lower mass peak is centred at just over 100 MeV/c^2. The split is only visible in high rapidity and high momenta data. For high rapidities, energies are higher at the same momenta compared to lower rapidities. For the high rapidity, high p_T graph, the main reason for the lower-mass reconstructions is therefore that the clusters are too close together, and could not be resolved adequately. If the HGLs were used to split clusters, the cluster energies might not have been divided properly – a problem discussed in the previous section.

The embedded data graphs in Figure 16 have a much broader mass distribution. For high rapidities, the low and mid momentum graphs show an increase in higher-mass reconstructions, so clusters with more background-energy contributions. The bottom right graph shows a split in the mass peak, as in the single-particle data, this time around 110 MeV/c^2. The single-particle graphs all had a very steep right-hand slope of the mass peak, but none of these graphs do.

In the PbPb embedded data graphs, the invariant mass mean shifts upwards (from low to high rapidities) two times out of three. The only mean decrease occurs in the high momenta graphs. However, the number of reconstructions in the high rapidity,

high p_T graph is relatively small (about $\frac{1}{3}$ of the number of reconstructions in the other graphs), so its statistical accuracy might be disputed.

In the embedded $0 < p_T < 8$ GeV/c graphs, the peak seems to shift to the right, while the background signals essentially stay the same. A different view of this might be that the pion peak becomes less distinguished, while the right-hand background increases. The mean mass shifts upward by $3.7 \pm 1.9\%$. The middle two graphs ($8 < p_T < 16$ GeV/c) show an increase in higher-mass reconstructions as well; the mean in these graphs shifts from around 158 to 172 MeV/c^2; an $8.9 \pm 1.6\%$ increase. The upward mass shifts are not unexpected: clusters are created with both photon shower hits and background energy signals. When the cluster energies go up, higher-mass reconstructions are made. For low-energy photons, with large opening angles, the amount of added background signals per cluster is higher than for high-energy photons, of which the clusters may overlap. In the previous sections, a problem with the energy resolution was the energy division over clusters by the HGLs. Any discrepancy between photon energy and measured cluster energy will influence the reconstructed pion invariant mass.

The last two graphs show a decrease of $2.1 \pm 2.4\%$, but, as indicated above, this number is less reliable. The extra lower-mass peak of the high rapidity, high p_T graph is located close to the mass extra peak in the single-particle data. This may indicate that the peak is caused by the same effects as that of the single-particle data. Whether this is true and the clusterfinder can be improved to reduce these effects is a topic for future research.

8
FoCal Efficiency:

Results & Discussion

One of the ways to quantify detector performance is by examining efficiencies. In the 'Project description' section, some of the applied data cuts have been discussed. In order to evaluate how effectively we can identify π^0 particles in PbPb events, I have made an invariant mass selection that is most likely to contain the well-reconstructed pions. For other invariant mass selections, the efficiency may go up or down; the current selection may not be useful for all other research projects. The efficiency of a certain dataset is obtained by dividing the cut data by the unaltered data ($\frac{reconstructed\ events}{generated\ events}$). In this project, the efficiency provides a measure for how much of the data is useful for identifying pions, while considering the geometry of the detector and using the current clustering algorithm.

Efficiency graphing

The figures below (Figure 17 and 18) each consist of 4 diagrams displaying FoCal efficiency. In Figure 17, the top left diagram shows efficiency of single-particle data in the $3.0 < \eta < 3.5$ rapidity range per pion energy, while the bottom left diagram shows the same in the $3.5 < \eta < 4.0$ rapidity range. The right two diagrams show the same rapidity ranges per transverse pion momentum value. The same order applies to Figure 18, only these diagrams have been generated from the embedded simulation dataset.

A good invariant mass selection for the efficiency is one where most of the selected data is part of the pion mass peak, and preferably only a marginal amount of background signal is included. In the single-particle data, a reconstructed invariant mass selection of $77.4 < m_{inv} < 186.3$ MeV/c^2 has been made. In the embedded data, a selection of $92.6 < m_{inv} < 201.5$ MeV/c^2 was made. The mass ranges are equal in breadth but shifted upwards for the PbPb data.

The dashed vertical pink lines in the plots compare pion energies with pion momenta, at the mean rapidity of the selected rapidity range. I.e., in the $3.0 < \eta < 3.5$ band, a rapidity of $\eta = 3.25$ is used to calculate equivalent momentum values of a certain selected energy. The horizontal purple top line indicates the Efficiency $= 1$ (or 100%) line. The dashed lines roughly give the boundaries of an efficiency 'plateau'. This is a region where the efficiency is particularly unaffected by energy or momentum fluctuations. Note that this does not exclude the usage of the rest of the data - it simply indicates a stable efficiency region.

In the single-particle data plots, this plateau seems to correspond to an efficiency around 80-85% for both rapidity regions, which is decent. The plateau shifts to lower energies and momenta in the higher rapidity band compared to the lower rapidity band. This makes sense, as the particle density in higher rapidity regions is higher. Since high momenta usually make for small angles, these particles will be more difficult to resolve. The efficiency of the embedded data goes down for the high rapidities. This is likely due to resolution issues; not only are the clusters closer together, but the background also makes for more energetic clusters. This causes the clusterfinder to resolve the clusters less accurately, resulting in less accurately reconstructed pions. Because the mass selection removes many inaccurately reconstructed pions for the 'reconstructed data', the efficiency is lower for higher rapidities.

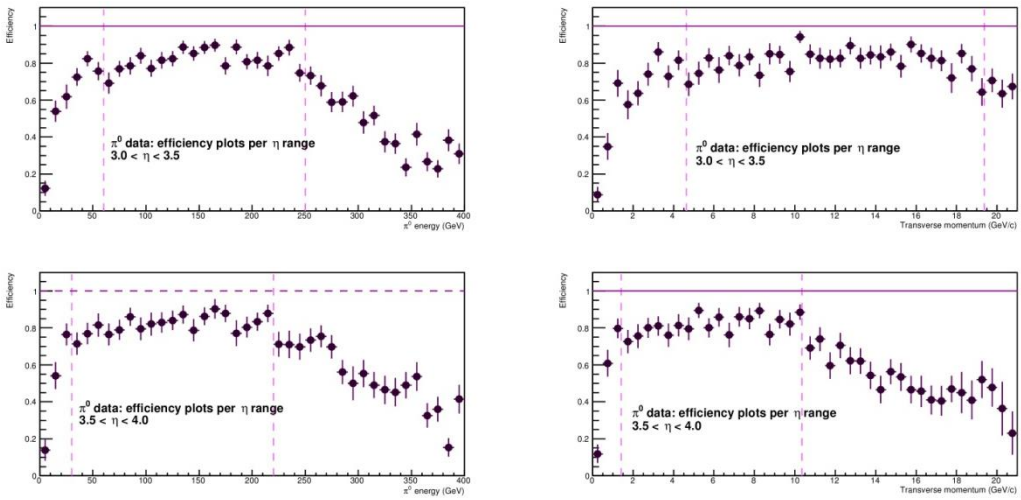

Figure 18: Efficiency data generated from single-particle simulations.

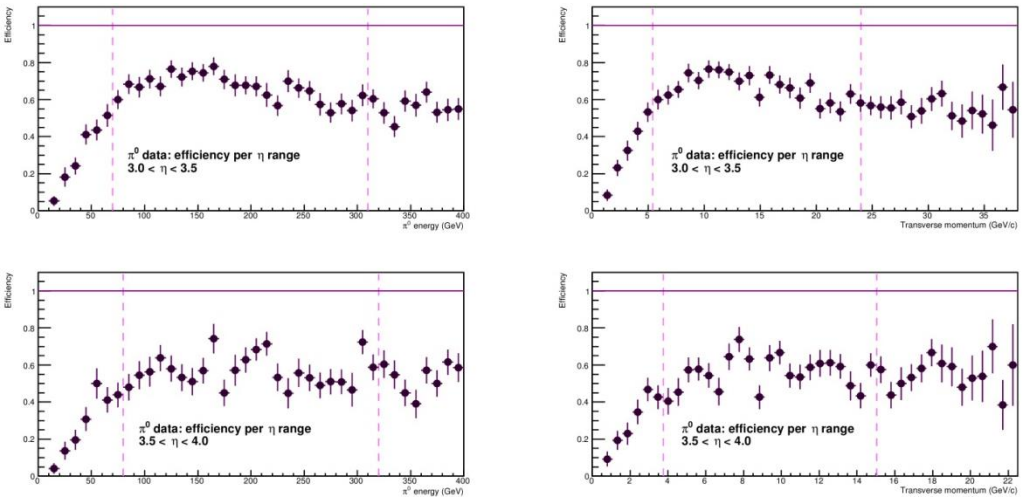

Figure 17: Efficiency data generated from PbPb embedded simulations.

The same does not apply to the plateau in the embedded simulation. There is a lot of background in this dataset. The clusters that are selected to reconstruct pions with are generally of higher summed energy (so higher invariant mass reconstructions) than the clusters of the single-particle simulation. An inclusion of background signals is unavoidable in the reconstructions. From the graphs in Figure 18, it is apparent that the efficiency in the higher rapidity region is more variable over the energy and momentum axis compared to the efficiency in the lower rapidity region.

For low energies (meaning the energies before the plateau), the efficiency is low. This is because, as discussed in previous sections, clusters are more widely dispersed. Distant hits may not be added to a cluster and the wrong second clusters might be selected if they happen to be closer to a first cluster.

The efficiency at high energies (after the plateau) does not slope downwards as fast as the single-particle efficiency. This is likely because of the invariant mass selection that was made for the embedded graphs. There are more higher-mass pion reconstructions in the high rapidity region (because of the background, see previous section), and these reconstructions have partially been included in the mass selection. From a certain energy onwards, the efficiency of the embedded data is actually higher than that of the single-particle data, illustrating the deceptiveness of the efficiency at high energies. The efficiency plateau in the embedded simulation data does not shift to the left in the high rapidity region, as it did in the single-particle simulation, but slightly to the right. This is mostly due to the reasons as described above. A possible additional reason is that high rapidity particles travel through less detector and beam pipe material, thereby improving the efficiency.

Within the plateau region, the efficiency of the embedded data in the $3.0 < \eta < 3.5$ range is about 60-75%. That in the $3.5 < \eta < 4.0$ range is about 55%, but fluctuates from under 45 to over 65%. As there are fewer events in the higher rapidity range, these deviations are likely to be caused by statistical fluctuations. By generating more data, these variations can be reduced.

Whether the high p_T, high rapidity efficiency is high enough depends on the requirements of a particular study. In the previous section, the invariant mass distributions displayed a pronounced – though broad – pion mass peak for all momenta and rapidity regions. The energy resolution section provided a few possibilities for further research and possible improvements that can be made to e.g. the clustering algorithm. It is unclear how much this will improve the efficiency. If the clustering algorithm is improved, and judging from the energy resolution and invariant mass analyses, I believe that neutral pions can be identified with reasonable effectivity at high momenta and rapidity.

9
Overall Evaluation

The aim of this project was to study how effectively we can identify neutral pions in PbPb events. Monte Carlo simulation data were used for this purpose. A great part of the resolution problems in embedded data come from energy divisions over separate clusters by the HGLs. A revision of the clustering algorithm could improve this. The method of second cluster selection may also be improved. There are more resolution difficulties with particles of high or low energy or rapidity than there are with particles of mid energy or rapidity.

High energy particles are more difficult to resolve, as they have smaller photon opening angles. At high rapidities, there are more surrounding clusters (and cluster energies) available, and this gives a higher chance of selecting clusters with the right energies. However, different clusters are not as easily discernible, and hits are easily assigned to neighbouring clusters. For low energies, the electromagnetic showers of two photons are further apart. For low rapidities, surrounding cluster densities are lower. Distant shower hits might not be added to a cluster, and a wrong cluster might be selected as second cluster if it is near the first cluster. In the embedded simulation, and especially for high rapidities, cluster energies become higher due to background signals – resulting in higher-mass pion reconstructions.

For the efficiency analysis, I have made an invariant mass selection that is most likely to contain the well-reconstructed pions. This selection may not be useful for all other research projects. An inclusion of background signals is unavoidable in the efficiency results, and this falsely improves efficiencies at the highest energies. Within the plateau region, the efficiency of the embedded data in the $3.0 < \eta < 3.5$ range is about 60-75%. That in the high rapidity range is about 55%, but fluctuates more because of a lack of data. Whether the high p_T, high rapidity efficiency is high enough is difficult to conclude. The invariant mass distributions displayed a pronounced – though broad – pion mass peak for all momenta and rapidity regions, and the energy resolution section showed adequate resolution results. Already, we can identify many of the pions. With the suggested improvements and further research and data generation, the effectivity will further improve.

REFERENCES

[1] CERN Press Office (2012). *LHC to run at 4 TeV per beam in 2012* and *The accelerator complex*. CERN Press Office, Meyrin, Switzerland. Accessed 23 dec 2014: http://press.web.cern.ch/press-releases/2012/02/lhc-run-4-tev-beam-2012 and http://home.web.cern.ch/about/accelerators.
[2] O'Luanaigh, C. (2014). *CERN's Large Hadron Collider gears up for run 2*. CERN Communications group, Meyrin, Switzerland. Accessed 23 dec 2014: http://home.web.cern.ch/about/updates/2014/12/cerns-large-hadron-collider-gears-run-2.
[3] ALICE Public content (2014). *The ALICE subdetectors*. Accessed 23 dec 2014: http://alipub.web.cern.ch/content/alice-subdetectors.
[4] Riegler, Dr. W. (2013). *ALICE looks to the future*. CERN Courier, IOP Publishing, UK. Accessed 23 dec 2014: http://cerncourier.com/cws/article/cern/53084.
[5] The ALICE FoCal Collaboration (2013). Letter of Intent: A Forward Calorimeter (FoCal) for the ALICE experiment. EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH.
[6] CERN Press Office. *Heavy-ion collisions, the quark-gluon plasma, and quantum chromodynamics*. Accessed 22 dec 2014: http://press.web.cern.ch/backgrounders/heavy-ion-collisions-quark-gluon-plasma-and-quantum-chromodynamics.