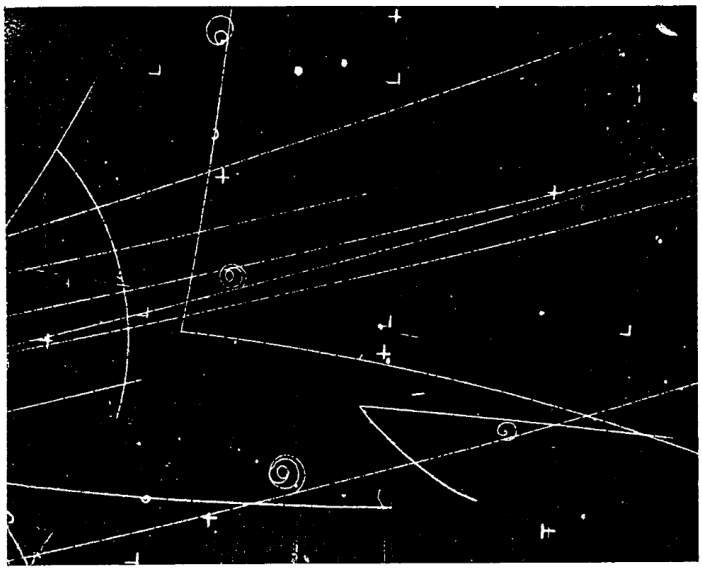
Utrecht University

Experimental Physics

Clusterization and background estimation of a direct photon signal in pp collisions at $\sqrt{s}$=13.6 TeV in ALICE

Master Thesis

Berend van Beuzekom



Supervisors:

Dr. M. Verweij
GRASP

Dr. A. Grelli
GRASP

16-12-2024

Abstract

This thesis presents a method to estimate the background contribution in a direct photon signal in proton-proton collisions at $\sqrt{s} = 13.6$ TeV in ALICE. This work serves as an initial step towards a comprehensive jet quenching study using Run 3 data. To estimate the background contribution, a new EMCAL clusterizer was implemented and advanced data selection techniques were applied. This thesis introduces and validates a suitable clusterizer for direct photon measurements. Additionally, it provides data selection criteria for isolating the direct photon signal. The estimated background contribution was found to be underestimated. Which is likely due to various factors associated with the used Monte Carlo dataset. However, the specific sources and magnitudes of these errors remain unclear and can be further investigated.

Figure from the title page is adapted from Ref. [1]

Contents

1 Introduction **1**

2 Theoretical background **2**
2.1 The standard model of particle physics 2
2.2 Quark-Gluon Plasma 4

3 ALICE **5**
3.1 LHC 5
3.2 ALICE 6
3.3 O2Physics 6

4 Photon Detection **7**
4.1 Detector design 7
4.2 Clusterization 9

5 Data **12**
5.1 Event selection 13
5.2 Track Selection 13
5.3 Cluster selection 14

6 Method **16**
6.1 Finding a Suitable clusterizer 16
6.1.1 Reconstruction method of the invariant mass of the π^0 17
6.1.2 Calculation method of the NLM 18
6.2 Direct photon measurement 18
6.2.1 Perpendicular cone method 19
6.3 Background analysis 22
6.3.1 ABCD method 22
6.3.2 Background normalization 24

7 Results **25**
7.1 Clusterizer 25
7.1.1 Reconstruction results of the invariant mass of the π^0 25
7.1.2 NLM results 26
7.1.3 Cluster shape and Cluster size evaluation 27
7.2 Background estimation 29
7.2.1 General purpose MC 29
7.2.2 Jet-jet MC 31

8 Discussion and Conlusion **32**
8.1 Clusterizer discussion 33
8.2 Background estimation discussion 34

9 Outlook **34**

1 Introduction

Under extreme temperature and/or densities, a matter of deconfined quarks and gluons, known as the quark-gluon plasma (QGP), can be formed [2]. In this state of matter, quarks are no longer confined within hadrons. It is hypothesized that the early universe, a few microseconds after the Big Bang, existed of this plasma. One method to study this plasma is through relativistic heavy-ion collisions at the Large Hadron Collider (LHC). In these collisions the energy transfer is sufficient to create the temperatures required for the QGP [3].

Due to the short lifetime of the QGP, the plasma can not be directly observed [2]. However, it can be studied indirectly using probes. A hard probe produced in the hard scattering process can interact with the plasma. Which leads to an energy loss of the parton. This energy loss is reflected in the energy/momentum of the resulting jet. This phenomenon is known as "jet quenching" [4]. The total energy loss of the initial parton can be determined if its initial energy is known. This can be achieved by requiring that a photon is produced alongside the hard probe in the initial scattering event. As A photon is color neutral and thus does not interact with the QGP [5]. Additionally, proton-proton collisions are essential for studying the QGP, as they provide a baseline measurement.

These studies have been conducted before. For example with the CMS experiment and run 1 data [5]. And with a higher integrated luminosity for run 2 data [6]. In the latter, the observable $x_{j\gamma} = P_{\mathrm{T}}^{\mathrm{jet}}/P_{\mathrm{T}}^{\gamma}$ ($P_{\mathrm{T}}^{\mathrm{jet}}$ and P_{T}^{γ} denote the transverse momentum of the jet and photon) was observed to have a depletion $x_{j\gamma} \approx 0.6$ compared to the proton-proton data. This depletion was observed for a centrality of $0-10\%$ and a $P_{\mathrm{T}}^{\mathrm{jet}} > 30$ GeV. However, this is at the edge of the statistical limit. To study if there is a further enhancement of $x_{j\gamma} < 0.6$, higher statistics are needed.

As the Run 3 data will significantly improve the available statistics, it becomes feasible to investigate the jet-quenching phenomenon with this data. This thesis provides a first step towards such an analysis, focusing on estimating the background contribution in an isolated direct photon signal from proton-proton data collected during Run 3 with the ALICE experiment at $\sqrt{s} = 13.6\ GeV$. This analysis presents several challenges. For instance, the Run 3 data uses a new analysis framework, requiring the implementation of an EMCAL (a subdetector of ALICE) clusterizer algorithm suited for direct photon measurements. Additionally, the combined cross-section of the required hard scattering processes is very low. Therefore some advanced data selections have to be applied.

This thesis begins in Sec. 2 with a brief theoretical overview of particle physics, the QGP, and jet quenching. Subsequently, in Sec. 3, an overview of the ALICE experiment and an explanation of the analysis framework used in this thesis will be provided. The photon signal will be obtained using the EMCAL detector. In Sec. 4 a detailed explanation of the EMCAL and the concept of clusterization will be presented. In Sec. 5 the used datasets will be described, along with the event, track, and cluster selection criteria. The method section (Sec. 6) is divided into three main parts. First, in Sec. 6.1 the implementation of a suitable clusterizer for direct photon detection will be explained. As well as its validation procedure.

Next, in Sec. 6.2, the process for obtaining the direct photon signal will be outlined. Finally, Sec. 6.3 will describe the estimation of the background contribution in the direct photon signal. The results section (Sec. 7) is divided into two main subsections. Section 7.1 will present the results of the clusterizer validation, and Sec. 7.2 will discuss the outcomes of the background estimation. Lastly, Sec. 8 will provide a final conclusion, and in Sec. 9 some further research prospects will be given.

2 Theoretical background

2.1 The standard model of particle physics

Particle physics is interested in the universe on its smallest scales. It studies the interactions between elementary particles, which are the fundamental building blocks of the universe. The current model to understanding these particles and their interactions is captured in the Standard Model of Particle Physics. This model provides a mathematical description of elementary particles and the forces (or interactions) mediated by force carriers, which are themselves particles. The four fundamental forces of nature are gravity, electromagnetism, the weak force, and the strong force. Gravity is very weak and therefore negligible in particle interactions. Electromagnetism governs the interactions between particles that possess electric charge, while the strong force is responsible for interactions between particles carrying color charge. The weak force acts on all particles and is responsible for β-decay in radioactive isotopes and nuclear fusion reactions.

The elementary particles of the Standard Model are shown in Fig. 1 [7]. These particles are categorized into two main groups: fermions and bosons. Fermions are particles with half-integer spin, whereas bosons have an integer spin. The fermions can be further divided into two groups: quarks and leptons. Quarks are the only particles that carry a color charge, making them the only particles that interact via the strong force. Leptons are classified into three generations. The first generation consists of the electron (e^-) and its corresponding electron neutrino (ν_e). The second and third generations consists of the muon (μ^-) with the muon neutrino (ν_μ), and the tau (τ^-) with the tau neutrino (ν_τ), respectively. The three generations have the same properties and only differ in mass. Quarks and leptons carry an electric charge. Except neutrinos as they are electrically neutral. Additionally, all quarks and leptons interact via the weak force. Bosons are divided into two categories: gauge bosons and scalar bosons. The gauge bosons act as the force carriers. The gluon (g) mediates the strong force, the photon (γ) mediates the electromagnetic force, and the Z and W bosons mediate the weak force [8]. The only scalar boson in the Standard Model is the Higgs boson (H). Which was only recently discovered in 2012 [9]. Interaction with the Higgs field gives particles their mass. Without the Higgs-field, all particles would be massless.

Feynman diagrams are a crucial tool in particle physics. These diagrams represent the interactions between particles and they are far more than simple drawings. From Quantum Field Theory (QFT) it is possible to derive rules for the interaction vertices and particles

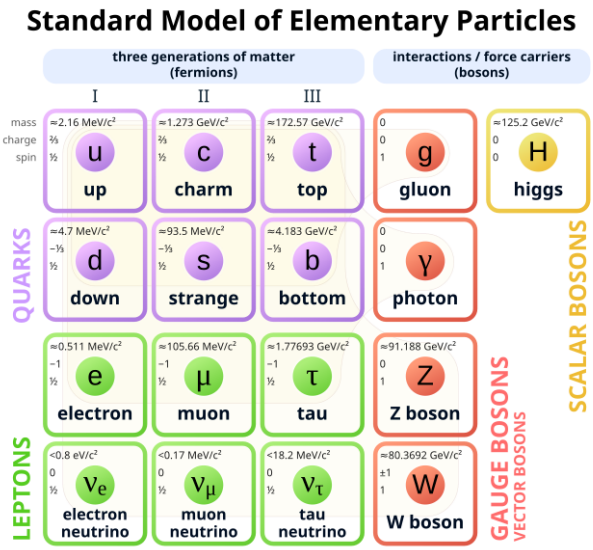


Figure 1: Standard model of elementary particles (adapted from Ref. [7])

which are part of the Feynman diagram. Once a Feynman diagram is constructed, it provides a straightforward method for calculating the matrix element. Which is the operator which transitions between initial and final quantum states. This eliminates the need to derive the matrix element directly from QFT for each interaction. An example of a Feynman diagram for the $e^-e^- \rightarrow e^-e^-$ scattering process is shown in Fig. 2. The diagram should be interpreted from left to right, with the arrows indicating the initial and final state particles, while the curly line represents the force mediator, which in this case is the photon. The Feynman diagram on the left illustrates a first-order process, while the one on the right shows a higher-order process. In principle, there are infinitely many Feynman diagrams for to the same process. However, the contribution of each diagram to the matrix element decreases exponentially with the number of interaction vertices. For example in Fig 2, the matrix element of a Feynman diagram with two interaction vertices is proportional to α^2, where α represents the intrinsic strength of the interaction. For diagrams with four interaction vertices, the matrix element is proportional to α^4. Since $\alpha \approx 1/137$ for electromagnetic interactions, higher-order diagrams can often be neglected as their contributions reduce exponentially. For the strong force, $\alpha \approx 1$, which makes analytical calculations nearly impossible due to the infinitely many contributions [8]. However, computational techniques such as lattice QCD have shown promising results [10].

There is strong experimental evidence for the existence of quarks. However, they have never been observed as free particles. Instead, quarks are always confined within bound states known as hadrons. Hadrons can be categorized into three groups: mesons, baryons, and anti-baryons. A meson consists of a quark and an antiquark, such as the $\pi^+(u\bar{d})$ meson, which consists of an up quark and an anti-down quark. Baryons are composed of three quarks, as for example the proton (uud), while anti-baryons consist of three antiquarks, such as the anti-proton $(\bar{u}\bar{u}\bar{d})$. The reason why quarks are not observed as free particles is due to color confinement. Which implies that only color-neutral particles can exist as free particles.

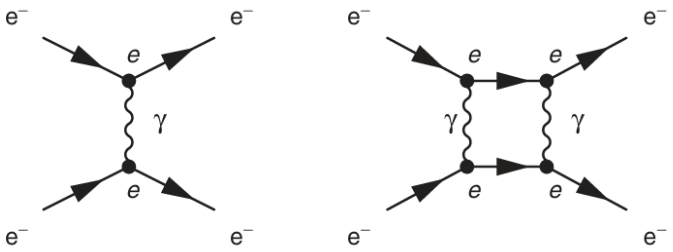


Figure 2: Two Feynman diagrams for a $e^-e^- \rightarrow e^-e$ scattering. Adapted from Ref. [8]

As a result, quarks are always bound into color-neutral states like mesons and baryons [8].

2.2 Quark-Gluon Plasma

The theory of Quantum Chromodynamics (QCD) describes the strong interaction and is a fundamental part of the Standard Model of particle physics. QCD predicts that under extreme energy and density conditions, a unique state of matter can exist in which quarks and gluons are deconfined. This state of matter is known as the Quark-Gluon Plasma (QGP) [11]. The early universe, a few microseconds after the Big Bang, is expected to have been composed of QGP. Additionally, it is hypothesized that the cores of neutron stars may consist of this state of matter [12]. Investigating the QGP not only provides a validation of QCD but is also very exciting, as it offers insights into the most extreme systems observed in the universe.

At the Large Hadron Collider (LHC) at CERN the QGP is created by colliding heavy ions, such as lead, at nearly the speed of light. These collisions generate energy levels sufficient to produce the extreme temperatures required for the QGP to exist [3]. However, the plasma is extremely short-lived as it transitions into a color-neutral hadron gas within approximately 10^{-22} s [13]. Therefore, no classical methods exist which can directly study the QGP. But, state-of-the-art detectors at the LHC enable the measurement of the energy, momentum, and flight paths of the resulting hadronised particles with high precision. Which allows for a indirect study of the plasma.

To study the plasma a probe is needed that travels through the plasma before reaching the detector. In a heavy-ion collision, the QGP forms after approximately 1 fm/c. During these collisions, hard parton scattering can occur earlier, at around 0.0 fm/c [2]. The color charged partons can interact with the QGP [14]. Which induces an energy loss on the resulting jet. This phenomenon is known as "jet quenching" [4]. To determine the total amount of energy loss of the color charged particle in the plasma. It is needed to determine its initial energy. This can be done by studying processes which have a photon and a quark (which will result in a jet) in their final state. The photon does not carry a color charge and will not have a interaction with the QGP [5]. Therefore, by measuring the photon energy, the initial energy of the quark can be determined. A photon produced in the hard scattering is often refered to as a direct photon. There are two main processes which have a quark and a photon in

there final state: quark-antiquark annihilation ($q\bar{q} \to q\gamma$, $q\bar{q} \to \gamma\gamma$) and Compton scattering ($qg \to q\gamma$). In higher-order processes, photons can also be created through bremsstrahlung and fragmentation [15, 16].

Figure 3 shows a schematic illustration of the probing of the QGP. On the left side of the figure, a Feynman diagram is drawn of the Compton scattering process ($qg \to q\gamma$). In a lead-lead collision, the quark interacts with the QGP. After the interaction with the plasma the quark hadronises in a jet which reaches the detector. The photon does not interact with the plasma and reaches the detecter after it has traversed the QGP. For comparison, the bottom of the sketch shows the same process in a proton-proton collision. Due to the lower energy transfer in proton-proton collisions, the QGP is not formed and can thus serve as a baseline measurement.

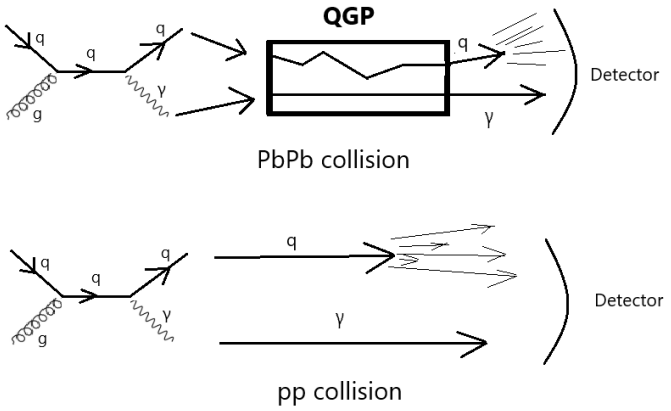


Figure 3: Schematic illustration of the usage of a probe for the QGP

3 ALICE

3.1 LHC

The Large Hadron Collider (LHC) is located underground at CERN in Geneva and is the world's most powerful particle accelerator. It is currently capable of colliding protons at a center-of-mass energy ($\sqrt{s}$) of 13.6 TeV. To facilitate these collisions, the accelerator is injected with small packets of gas consisting of protons or heavy ions. These packets, referred to as "bunches," are accelerated and cross each other inside a detector. During each "bunch crossing," depending on the density of the particles, there is a probability of having 0, 1, or multiple collisions.

The LHC was first operated in 2008 and 2009. In the following years, the center-of-mass energy and bunch intensity were increased to 8 TeV and $1.6 * 10^{11}$ (p per bunch). By 2012,

the total integrated luminosity of the proton-proton data reached 30 fb^{-1}. In 2013, the LHC was shut down for two years, marking the end of Run 1. The second data-taking period, known as Run 2, took place from 2015 to 2018, with the center-of-mass energy increased to 13 TeV. During Run 2, the total integrated luminosity for proton-proton collisions reached 165 fb^{-1}, with peak luminosities up to twice the design value [17] [18]. In 2019, the LHC was shut down for a second time. In 2022 Run 3 started and is expected to significantly increase the available statistics.

3.2 ALICE

The ALICE (A Large Ion Collider Experiment) is situated at the LHC. Its primary aim is to study quark-gluon plasma. Due to its design, ALICE can comprehensively analyze electrons, photons, muons, and charged hadrons produced in collisions. The experiment involves heavy nuclei (Pb-Pb) and lighter ions. Proton-proton collisions are used to provide reference data for the heavy-ion studies [19].

Globally, ALICE consists of four main types of detectors: tracking detectors, particle identification detectors, electromagnetic calorimeters, and a muon spectrometer. A schematic overview of the detectors can be seen in figure 4 adapted from ref [20]. The key detectors are described in the this section. The tracking detectors include the Inner Tracking System (ITS) and the Time-Projection Chamber (TPC). Their primary task is to reconstruct charged hardrons. The Time-Of-Flight (TOF) array measures the speed of particles and plays a crucial role in particle identification (PID). The TOF covers a large phase space and is optimized for particles with intermediate momenta. The High Momentum Particle Identification Detector (HMPID) enhances the hadron identification capability at higher momentum ranges. Photons, as well as neutral mesons, are measured by PHOS (PHOton Spectrometer), an electromagnetic calorimeter. PHOS is optimized to capture correlations between thermal and direct photons, as well as the pair photon decay of neutral mesons. The ElectroMagnetic Calorimeter (EMCal), another electromagnetic calorimeter, is optimized to measure electrons from heavy-flavor decays and the spectra of direct photons and neutral mesons. Compared to PHOS, the EMCal has a larger acceptance but lower energy and position resolution. A more detailed explanation of the EMCal can be found in Sec. 4, as it will be the main focus of this thesis. Finally, ALICE has a muon spectrometer that measures quarkonium resonances and is used for muon identification [19] [21].

3.3 O2Physics

The computing framework for ALICE Run 3 is called ALICE O2. In Run 3, data collection for ALICE happens in a continuous mode, where each unit of information corresponds to a 10 ms time window. Due to this continuous data-taking, tasks such as vertex-to-track matching are no longer unambiguous. As a result, collisions and tracks for example, are stored as a collection of tables linked by indices. This process leads to the reconstruction of data in the Analysis Object Data (AOD) format, which consists of a set of tables stored as ROOT

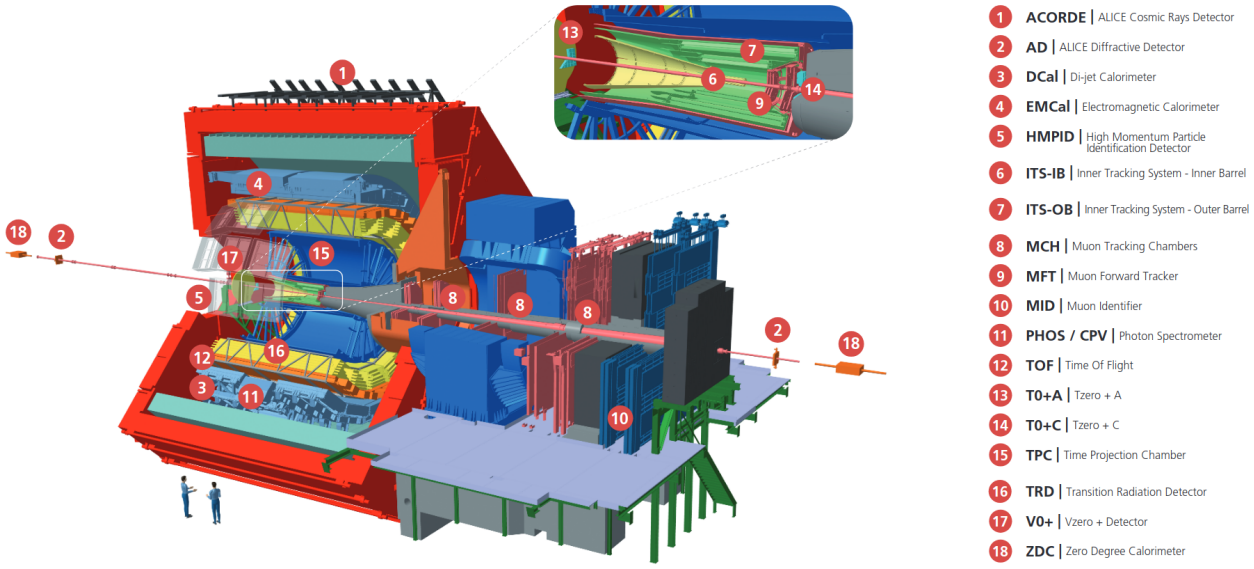


Figure 4: Schematic overview of the ALICE experiment. Adapted from [20]

trees. The source files for data analysis tasks follow a C++ structure and are stored in the O2Physics framework. These tasks are created by users of the framework, including those contributing to this thesis [22].

4 Photon Detection

A photon must interact with the material of the electromagnetic calorimeter (EMCAL) to be detected. At low energies ($E_\gamma < 1$ MeV), the dominant interaction is the photoelectric effect. Around $E_\gamma \sim 1$ MeV, Compton scattering becomes dominant, and at $E_\gamma > 10$ MeV, e^+e^- pair production is the primary interaction. The EMCAL can only measure energies above 16 MeV. Therefore, I will only consider e^+e^- pair production. When a high-energy photon strikes the material of the detector, it will produce an e^+e^- pair, which in turn radiates a bremsstrahlung photon. This photon will again produce an e^+e^- pair, leading to a cascade of particles known as an electromagnetic shower. A visualization of an electromagnetic shower is shown in Fig. 5, taken from Ref. [8]. In this figure, X_0 represents the radiation length, which is the average distance over which an electron emits a bremsstrahlung photon [8].

4.1 Detector design

The EMCAL is composed of modules. Each module consists of four (2 x 2) towers (also called cells). Each tower is made up of 76 alternating layers of natural lead and 77 layers of scintillator, with an effective radiation length $X_0 = 12.3$ mm. The lead layers serve as the absorber. When excited by radiation, the photons emitted by the scintillator layers are captured by Wavelength Shifting (WLS) fibers, allowing to track the evolution of the electromagnetic shower. Each tower covers an area of $\Delta\eta \times \Delta\phi \approx 0.0143 \times 0.0143$. Supermodules

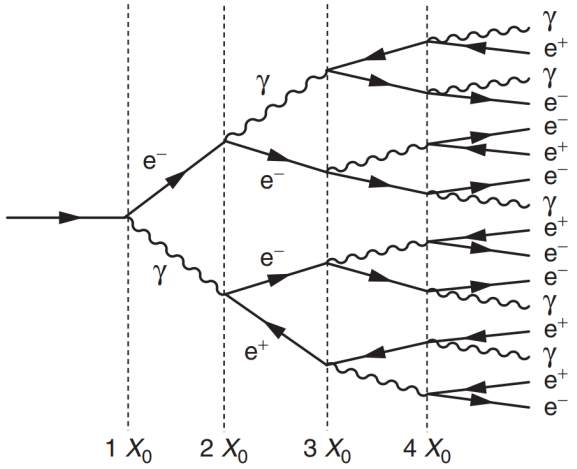


Figure 5: Schematic overview of the development of an electromagnetic shower. Adapted from Ref. [8]

.

are made up of modules and come in three different sizes: full-size (12×24 modules), 2/3-size (12×16 modules), and 1/3-size (4×24 modules). The EMCAL spans two separate regions in azimuth, formally consisting of two detectors known as the EMCal and the DCal. However, EMCAL often refers to both, as they only differ in coverage. The EMCal consists of 9 full-size and 2 1/3-size supermodules, covering $80° < \phi < 187°$, $|\eta| < 0.7$. The DCal consists of 6 2/3-size and 2 1/3-size supermodules, covering $260° < \phi < 320°$, $0.22 < |\eta| < 0.7$ and $320° < \phi < 327°$, $|\eta| < 0.7$. The DCal detector is positioned around the PHOS detector; therefore, the DCal has a gap in its pseudorapidity coverage [21]. A schematic overview of the EMCal and the DCal can be seen in Fig. 6, taken from Ref. [21].

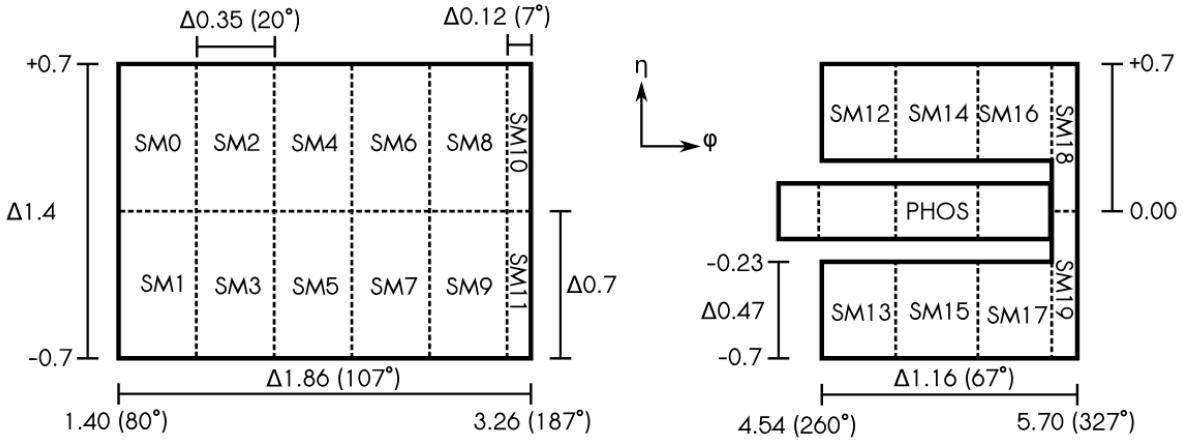


Figure 6: Schematic view of the EMCal and the DCal. With the DCal placed around the PHOS detector. Adapted from Ref. [21]

.

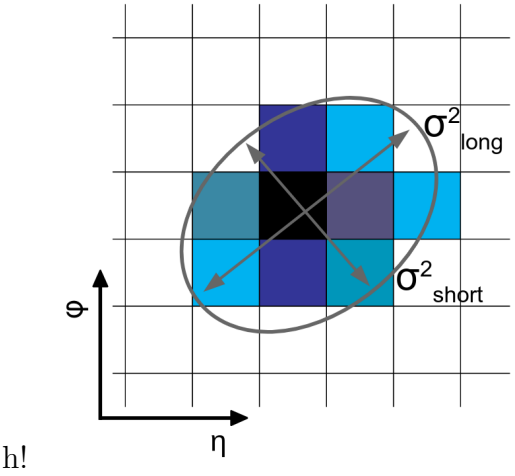
4.2 Clusterization

When a particle hits the detector, it produces an electromagnetic (or hadronic) shower that can spread its energy to neighboring cells. A collection of adjacent cells with energy above the noise threshold is called a cluster. The characteristics of a cluster depend on various factors such as the clustering algorithm, particle density, particle energy, and event type. Clusters can also include contributions from multiple particles. Photons and electrons deposit their full energy in the detector, so the cluster energy roughly corresponds to the particle's energy. Mesons like the π^0 and η, which primarily decay into two photons, have their energy measured as the sum of two separate clusters, or, if the angle between the photons is small enough, as a single merged cluster [21].

Although the EMCAL is designed to measure the energy of photons and electrons, hadrons (often charged hadrons) can also deposit energy in the EMCAL. This occurs through ionization or the production of a hadronic shower. To distinguish the signal from charged and neutral particles, charged-particle-track matching is required. Charged particles are reconstructed using the ITS (Inner Tracking System) and TPC (Time-Projection Chamber). The tracks are then extrapolated to the EMCAL, while accounting for the energy loss as the particle as it travels through the detector material. The track-matching algorithm works as follows: each track is stepwise extrapolated to the average depth of the cluster energy deposition. Then, any track within an angular distance of 0.2 radians is further extrapolated to the precise radial distance using a smaller step size. With the energy of these matched it is possible to calculate their energy contribution to a cluster [21].

The characteristics of a cluster are defined by several variables. First, the energy of a cluster, E, is the sum of the energy of all its cells. The size of the cluster is determined by the number of cells in the cluster, $NCells$. The location of the cluster is described by two variables: η (pseudorapidity) and φ (azimuth). The energy distribution within the cluster can be characterized by the number of local maxima (NLM). Finally, the shape of the cluster is defined by its elliptical axes: σ_{Long} (M02) and σ_{Short} (M20). If M20 $\approx$ M02, the cluster will have a circular shape, but if $\sigma_{\text{Long}} < \sigma_{\text{Short}}$, the cluster will have a more elliptical shape. A schematic view of a cluster is shown in Fig. 7, adapted from Ref. [21]. In the figure, a cluster is shown in the $\eta - \phi$ plane, with one local maximum and a slightly elliptical shape. The colors indicate the energy deposited in each cell, with darker shades representing higher energy.

Currently (at the time of writing), there is one clustering algorithm included in the ALICE Run 3 O2Physics framework, namely the kV3 algorithm. The algorithm begins by searching for the highest-energy cell in a region, referred to as the seed cell, which must have a minimum energy of 500 MeV. Therefore, if a region lacks a cell with an energy of $\text{E}_{\text{cell}} > 500$ MeV, no cluster will be formed. If a seed cell is found, the algorithm proceeds to evaluate its adjacent cells. An adjacent cell is added to the cluster if its energy exceeds the aggregation or noise threshold ($\text{E}_{\text{agg}} = 100$ MeV). However, if an adjacent cell has more energy than its previous neighbor, it will not be added to the cluster, unless the difference is small enough to be explained by noise fluctuations. This is known as the gradient cut, which



h!

Figure 7: Schematic view of a cluster in the $\eta - \phi$ plane with one local maximum and a slightly elliptical shape. Adapted from Ref. [21]

.

is set to 30 MeV. The above values of the seed cell, noise threshold and gradient cut are the default values of the algorithm but can be adjusted [23]. A one-dimensional visualization of the behavior of the kV3 algorithm is shown in Fig. 8, adapted from Ref. [21]. The figure shows that the algorithm reconstructed two clusters while excluding a cell that falls below the aggregation threshold.

The kV3 clusterizer described above is particularly effective at reconstructing mesons that primarily decay into two photons, such as the π^0 and η mesons. The mass of the meson can be calculated using Einstein's mass-energy equation, as shown in Eq. 1. In this equation, the mass depends on the energy (E_γ) and momentum ($\vec{p}_\gamma$) of the two photons. By applying cosine similarity ($\vec{A} \cdot \vec{B} = ||\vec{A}||||\vec{B}|| \cos(\theta)$) and that the photon is massless, Eq. 1 can be rewritten as Eq. 2, where θ is the angle between the photons in the lab frame. Thus, the mass of the π^0 (or η) depends solely on the energy of the clusters and their angle. For high-energy mesons, the angle can become small enough that the clusters may overlap. However, due to the feature of the kV1 clusterizer, which prevents it from adding adjacent cells if a previous neighbor has lower energy, the clusterizer can still distinguish the clusters and reconstruct the mass of the mesons—until the angle between the photons becomes small enough that both fall into a single cell.

$$M_{\pi^0}^2 = (E_{\gamma 1} + E_{\gamma 2})^2 - (\vec{p}_{\gamma 1} - \vec{p}_{\gamma 2})^2 \tag{1}$$

$$M_{\pi^0}^2 = 2E_{\gamma 1}E_{\gamma 2}(1 - \cos(\theta)) \tag{2}$$

For the reconstruction of direct photons, the kV3 algorithm is less suitable because it becomes very challenging to distinguish background photons (primarily from neutral meson decays) from direct photons. A 2-dimensional sketch of a direct photon signal is shown in Fig. 9(a), where red indicates high energy, green indicates intermediate energy, and blue

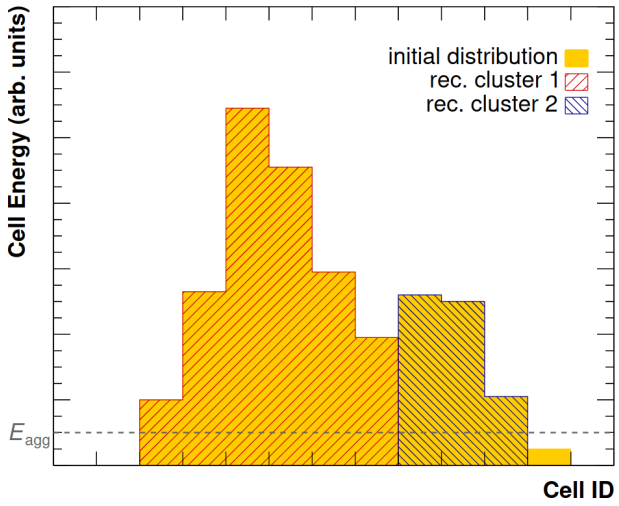
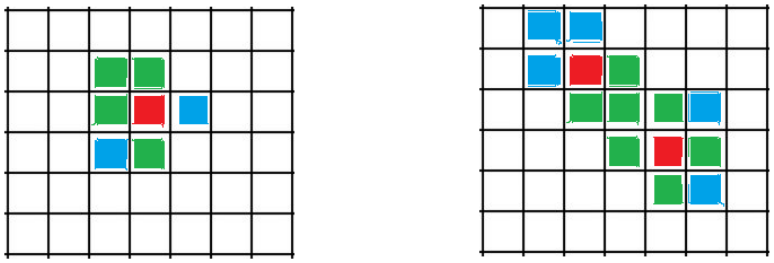


Figure 8: One-dimensional visualization of the behaviour of the kV3 algorithm. Adapted from Ref. [21]

.



(a) (b)

Figure 9: 2-dimensional sketch of a direct photon signal (a) and a neutral meson decay signal (b). Red indicates high energy, green indicates intermediate energy, and blue indicates low energy.

indicates low energy. As sketched, the direct photon produces relatively small and circular clusters, while a neutral meson decay signal produces relatively large elliptical clusters, as shown in Fig. 9(b). The kV3Default algorithm clusters the neutral meson decay signal (Fig. 9(b)) into two separate clusters, resulting in two relatively small and circular clusters. Which makes the signal indistinguishable from a direct photon signal.

In the ALICE run 2 framework there was a dedicated clusterizer to distinguish direct photons from decay photons. This is called the kV1 clusterizer. However, as of the time of writing, this clusterizer has not yet been implemented in the Run 3 framework. The kV1 clusterizer works similarly to the kV3 clusterizer but is simpler. It continues adding cells to the cluster as long as the added cell has an energy higher than the aggregation threshold (E_{agg}), without the constraint that the added cell must have a lower energy than its previous neighbor. In Fig. 10, taken from Ref. [21], a one-dimensional visualization of the kV1 algo-

rithm's behavior is shown. It can be observed that the cluster has two local maxima, and in a two-dimensional plane, the cluster takes on a more elliptical shape as sketched in Fig 9(b).

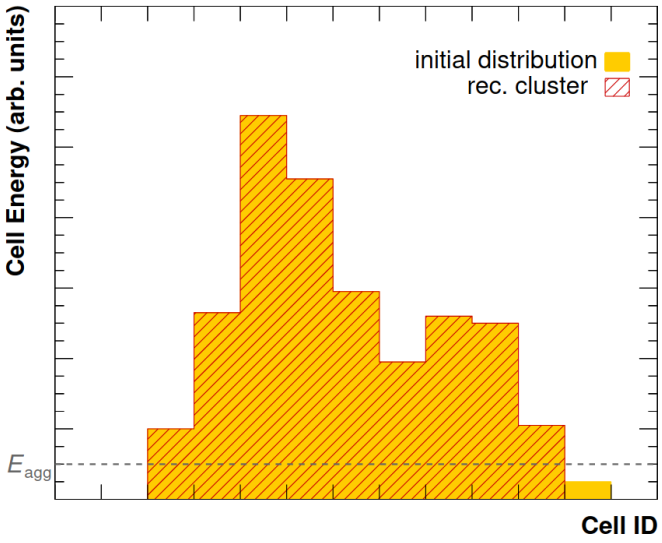


Figure 10: One-dimensional visualization of the behaviour of the kV1 algorithm. Adapted from Ref. [21]

.

5 Data

In this thesis, three different datasets are used: one is a Real Dataset (RDS), and the other two are Monte Carlo Datasets (MCD). The clusterizer analysis relies only on the RDS, while the background analysis uses both MCD's. The characteristics and functionalities of these datasets are described in the paragraphs below.

The RDS used in this thesis is LHC22o_pass7_minBias_small. This is a minimum bias dataset consisting of approximately 3×10^9 proton-proton (pp) collisions at $\sqrt{s} = 13.6$ TeV. The data was collected during the LHC22o period and includes run 526641, with the latest reconstruction (pass7) applied. This dataset will be used for all analysis's presented in this Thesis.

The first MCD used is a general purpose minimum bias Monce Carlo dataset. Consisting of approximately 7×10^7 proton-proton collisions at $\sqrt{s} = 13.6\ TeV$. The data is anchored to the LHC22o_pass6 reconstruction with the following run numbers: 526641, 526964, 527041, 527057, 527109, 527240, 527850, 527871, 527895, 527899, 528292, 528461, 528531. This dataset is going to be used for the estimation of the background contribution in a direct photon signal. As further outlined in Sec. 6.3.

The second MCD used is a jet-jet Monte Carlo called: LHC24g4. It is anchored to the LHC22o_pass7 reconstruction with the following runs: 526641, 526964, 527041, 527057, 527109, 527240, 527850, 527871, 527895, 527899, 528292, 528461, 528531. This dataset contains approximately 400×10^6 proton-proton (pp) collisions at $\sqrt{s} = 13.6$ TeV, which were generated with PYTHIA8 Monash. In the simulation, the QCD jet production process `HardQCD:all` was applied, enabling all hard QCD $2 \rightarrow 2$ processes [24]. The simulation used a $\hat{P}_{\mathrm{T,Min}}$ of 5 GeV and a $\hat{P}_{\mathrm{T,Max}}$ of 600 GeV, restricting the transverse momentum of the outgoing partons in the hard scattering to $5 \leq P_T \leq 600$ GeV. Furthermore, the data is oversampled with a weight power of 4, meaning each event must be corrected by the weight defined in Eq. 3 to represent realistic data. In Eq. 3, w represents the weight factor, $P_{\mathrm{T,ref}}$ is the transverse momentum of the outgoing partons from the hard scattering for a specific event, and n denotes the weight power. This dataset is also going to be used for the estimation of the background contribution in a direct photon signal. Were it is expected to increase the statistics at higher energies compared to the first MCD. As further outlined in Sec. 6.3

$$w = (\frac{\hat{P}_T}{P_{T,ref}})^n \tag{3}$$

5.1 Event selection

Not every collision (event) has the same quality, so certain data selections are applied to reject pileup, beam-gas interactions, and low-quality collisions. Pileup can occur when multiple collisions—either from the same bunch or from subsequent bunches—fall within the same readout time window of the detectors. Beam-gas interactions occur when a beam particles interacts with a residual gas atom inside the beam pipe.

The rejection of beam-gas and pileup events is achieved by using the "`sel8`" flag in the O2Physics framework. Sel8 is an event selection based on collision time by using the FT0A and FT0C subdetectors, which are components of the FIT (Fast Interaction Trigger) detector. The FT0A and FT0C are positioned on opposite sides of the interaction point, enabling them to precisely measure the primary collision time and identify beam-gas events, which are characterized by large deviations from the collision time. Pileup can be rejected by the FIT detector as well due to its precise measurement of the primary collision time [25]. To further reduce pileup, only events with a single collision per bunch crossing are selected. Finally, each event is required to have a primary vertex within $|Z_{vtx}| < 10$ cm from the nominal interaction point.

5.2 Track Selection

In addition to event selection, track selection is necessary to ensure high-quality tracks and reduce noise. Track selection in the O2Physics framework is achieved using global tracks, which must meet specific detector criteria to ensure its quality, as shown in Table 1. Each track must have crossed at least 70 rows in the TPC (Time Projection Chamber) to provide

Table 1: Track selection criterion for global tracks. Adapted from Ref. [22]

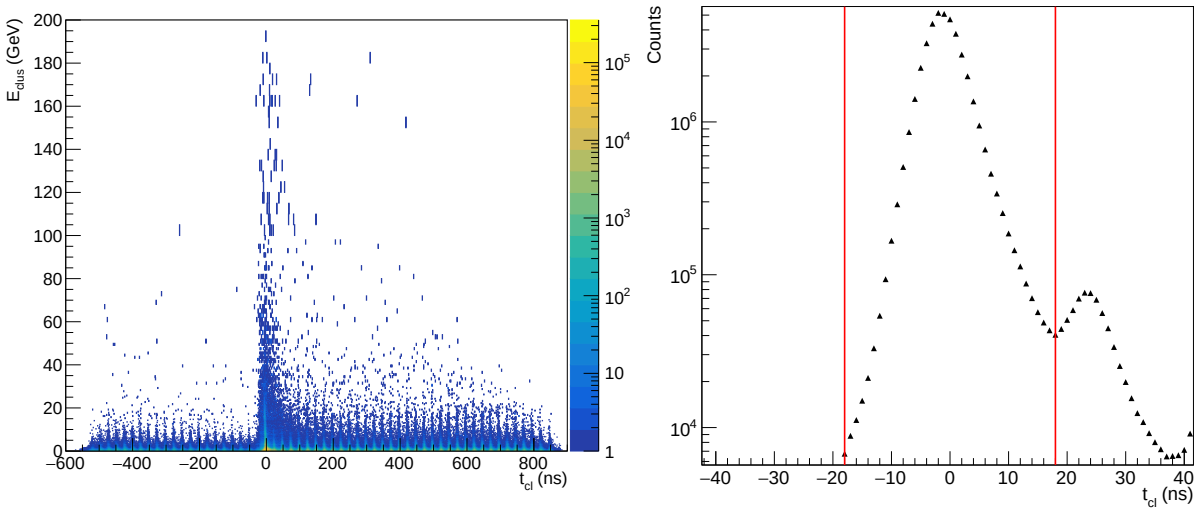
Min number of crossed rows TPC	70
Min ratio of crossed rows over findable cluster TPC	0.8
Max chi2 per cluster TPC	4.0
Max chi2 per cluster ITS	36.0
Require TPC refit	True
Require ITS refit	True
Max DCA to vertex z	2.0
Max DCA to vertex xy	$0.0105 * 0.035/p_T^{1.1}$
Cluster requirement ITS	at least one hit in SPD in one of the 3 innermost ITS layers
P_t range	0.1 - 1e10
η range	-0.8 - 0.8

sufficient data points for track reconstruction. The ratio of crossed rows to findable clusters in the TPC must be at least 0.8. The χ^2 per cluster in the TPC and ITS (Inner Tracking System) must be less than 4.0 and 36.0, respectively, to ensure realistic track reconstruction. Both the TPC and ITS require refitting, meaning each track is reconstructed both from its outermost data point relative to the vertex and from its innermost data point. The maximum distance of closest approach (DCA) of the track to the vertex along the z-axis (beam axis) must be within 2.0 cm, and the maximum DCA to the vertex in the xy-plane is $0.0105 + 0.35/p_{\mathrm{T}}^{1.1}$ cm, where p_{T} is the transverse momentum. The difference in DCA requirements between the z-axis and xy-plane is explained by the detector resolution in each plane. Additionally, each track must have at least one hit in one of the three innermost layers of the SPD (Silicon Pixel Detector). Which is a subdetector of the ITS and the closest to the interaction point. Finally, each track must have a transverse momentum in the range of 0.1–10 GeV and a pseudorapidity of $|\eta| < 0.8$ [22].

5.3 Cluster selection

Cluster selection is essential for ensuring high-quality clusters in the EMCAL. The EMCAL has a wide time integration window of 1.5 μs, while the separation between consecutive bunches is on the order of nanoseconds [21]. This wide integration window leads to pileup, as multiple collisions are recorded within each event. To select clusters associated with the primary collision, a timing cut is applied based on the arrival time of the leading cell in each cluster (referred to as the cluster time). To study the bunch spacing, the `emcclustermonitor` task in the O2Physics framework was used on the LHC22o_pass6_minBias_small dataset. This task monitors various cluster properties, including the cluster time.

In Fig. 11(a), the cluster energy ($\mathrm{E_{clus}}$) is plotted against the cluster time ($\mathrm{t_{cl}}$). The distribution of cluster time extends from approximately -600 ns to 800 ns, corresponding to



(a) (b)

Figure 11: The cluster energy (E_{clus}) plotted against the cluster time (t_{cl}) (a) and its integral with the bunch spacing marked by vertical red lines (b).

the 1.5 μs time integration window. The primary collision is centered around 0 ns, where the majority of the signal is observed. Notably, the data centered around 0 ns is independent of the cluster energy. In Fig. 11(b) the time distribution for all clusters is shown. Which is the projection of Fig. 11a onto the time axis. This figure uses a logarithmic scale for the y-axis and is zoomed in around 0 ns for clarity. The bunch spacing is clearly visible as Gaussian peaks in this graph. The peaks appear slightly asymmetrical due to their overlap, indicating that consecutive bunches do not have a clean separation. The red lines at $|t_{cl}| = 18$ ns mark the best separation between bunches and will be the cluster time cut used in the remainder of this thesis.

As described in Sec. 4, hadrons can deposit a fraction of their energy in the EMCAL. Most of these hadrons deposit the minimum ionizing particle energy, with the exception of slow neutrons. When a neutron interacts with the EMCAL, it produces a highly energetic signal where most of the energy is concentrated in a single cell, unlike the more distributed signal of an electromagnetic shower, which spreads across multiple cells. A cluster with this concentrated energy deposition is classified as an "exotic cluster." The exoticity of a cluster is defined by the parameter F_+, as shown in Eq. 4, where E_{cell}^{max} is the energy of the most energetic cell in the cluster, and E_+ is the sum of the energy of the four adjacent cells [21]. A cluster is considered exotic as F_+ approaches 1. To reduce the number of exotic clusters in the dataset, only clusters with $F_+ < 0.97$ are included in the remainder of this article. Additionally, for further noise reduction, only clusters with an energy of $E_{clus} > 0.7$ GeV are used.

$$F_+ = 1 - \frac{E_+}{E_{cell}^{max}} \tag{4}$$

6 Method

This chapter is divided into three main sections. The first section describes the method for finding a suitable clusterizer for measuring direct photons (Sec. 6.1). The second objective of this thesis is to estimate the background contribution in the direct photon signal. However, before the background can be estimated, a method for identifying direct photons must first be constructed. This method is described in the second section (Sec. 6.2) and mostly involves data selections and transformations to improve the signal purity. Finally, the third section (Sec. 6.3) of this chapter describes the method for estimating the background contribution in the signal obtained from the second section.

6.1 Finding a Suitable clusterizer

In the ALICE Run 2 framework, a dedicated clusterizer was implemented to distinguish direct photons from photons originating from neutral meson decays, as mentioned in Sec. 4. However, at the time of writing, this clusterizer has not yet been implemented in the ALICE Run 3 framework. The currently implemented clusterizer, known as the kV3 clusterizer, is suitable for detecting photons from neutral meson decays. The kV3 clusterizer starts by identifying the cell with the highest energy (the seed cell). Subsequently, it adds adjacent cells to the cluster if each adjacent cell has lower energy than its previous neighbor. An exception is made if an adjacent cell has only a slightly higher energy than the previous one, in which case it is still included in the cluster to account for noise and detector resolution variations. This criterion is known as the "gradient cut." If there are multiple local maxima within a region of triggered cells, the kV3 clusterizer divides the area into multiple clusters. This feature makes it suited for detecting neutral meson decay photons, as these particles decay at small angles for higher energies, leading to adjacent energy deposits and clusters. In Fig. 8 a one-dimensional visualization of the kV3 clusterizer can be seen, and Sec. 4 includes a more detailed explanation of the clusterizer algorithm and the measurement of neutral meson decay photons.

Rather than implementing a new clusterizer suited for direct photon detection in the ALICE Run 3 framework, it is possible to modify the existing kV3 algorithm to mimic the behavior of the clusterizer which was dedicated for direct photon detection. The kV3 algorithm is implemented in such a way that by deactivating the gradient cut, the algorithm will continue adding adjacent cells into a cluster independent of the energy differences between them, except when a cell's energy falls below the aggregation threshold. Therefore, if a region of triggered cells contains multiple local maxima, it will now be grouped into a single cluster, and thus merging adjacent clusters originating from neutral meson decays. Cluster variables like the number of local maxima (NLM) and the longitudinal cluster axis, σ_{Long}, can distinguish direct photons from photons from neutral meson decays. Since, merged clusters from neutral meson decay photons tend to have a larger σ_{Long} axis (indicating a more elliptical shape) and multiple local maxima.

To evaluate whether the adjusted kV3 clusterizer performs as hypothesized above, a series

of checks are conducted. First, the invariant mass of the π^0 meson will be reconstructed using cluster pairs. The adjusted kV3 algorithm is expected to be unsuccessful in reconstructing the π^0 mass due to the anticipated merging of adjacent clusters. Additionally, the number of local maxima will be analyzed for both clusterizers, where the adjusted algorithm is expected to show a higher fraction of clusters with multiple local maxima. Finally, the σ_{long} axis of both clusterizers will be compared; the adjusted kV3 algorithm is expected to have a larger fraction of clusters with an increased σ_{long}. The reconstruction of the π^0 mass will be further outlined below. As well as the calculation of the number of local maxima which, at the time of writing, is not yet a build in function of the ALICE run 3 framework.

6.1.1 Reconstruction method of the invariant mass of the π^0

The reconstruction of the invariant mass of the π^0 meson is performed using the existing `emcalPi0EnergyScaleCalib.cxx` task within the O2Physics framework. This task operates as follows: First, it selects only unambiguous collisions, meaning it maintains only bunch crossings with a single collision event. For every cluster in the EMCAL, the energy (E_{Clust}), pseudorapidity (η), and azimuthal angle (φ) are retrieved. From these three variables, the Lorentz vector of the photon that triggered the cluster can be constructed. The pseudorapidity, η, is defined by Eq. 5 [8], where θ is the polar angle in the spherical coordinate system. By rewriting Eq. 5 to Eq. 6, the x, y, and z components of the momentum (P_{x}, P_{y}, and P_{z}) can be calculated using conversion to the Cartesian coordinate system, as shown in Eqs. 7, 8, and 9.

$$\eta \equiv -\ln(\tan\frac{\theta}{2}) \tag{5}$$

$$\theta = 2\tan^{-1}(e^{-\eta}) \tag{6}$$

$$P_{\text{x}} = \text{E}_{\text{Clust}}\sin(\theta)\cos(\varphi) \tag{7}$$

$$P_{\text{y}} = \text{E}_{\text{Clust}}\sin(\theta)\sin(\varphi) \tag{8}$$

$$P_{\text{z}} = \text{E}_{\text{Clust}}\cos(\theta) \tag{9}$$

Next, the Lorentz vector of each hypothetical meson is calculated by summing the Lorentz vectors of pairs of photons, with each possible cluster pair in a collision considered once (without double counting). After obtaining the Lorentz vectors of the hypothetical mesons, the mass of each meson is calculated using Einstein's mass-energy equation. The mass is then stored as a function of the transverse momentum, $P_{\text{T}} = \sqrt{P_{\text{x}}^2 + P_{\text{y}}^2}$. This method produces a signal peak at the invariant mass of the π^0 meson, along with a background from paired clusters not originating from a π^0 decay.

6.1.2 Calculation method of the NLM

At the time of writing, there is no built-in function in the O2Physics framework to calculate the NLM (Number of Local Maxima). Therefore, an algorithm was developed to perform this calculation. The procedure is as follows: First, the algorithm retrieves the amplitude (proportional to energy) and cell ID of all cells in the cluster. The cell ID is then used to determine the supermodule, row, and column index of each cell. The algorithm then iterates through each cell in the cluster, checking for neighboring cells with a higher amplitude. If a neighboring cell with a higher amplitude is found, the algorithm moves to that cell and repeats the search for a neighbor with an even higher amplitude. This process continues until a local maximum (the cell with the highest amplitude in that region) is found.

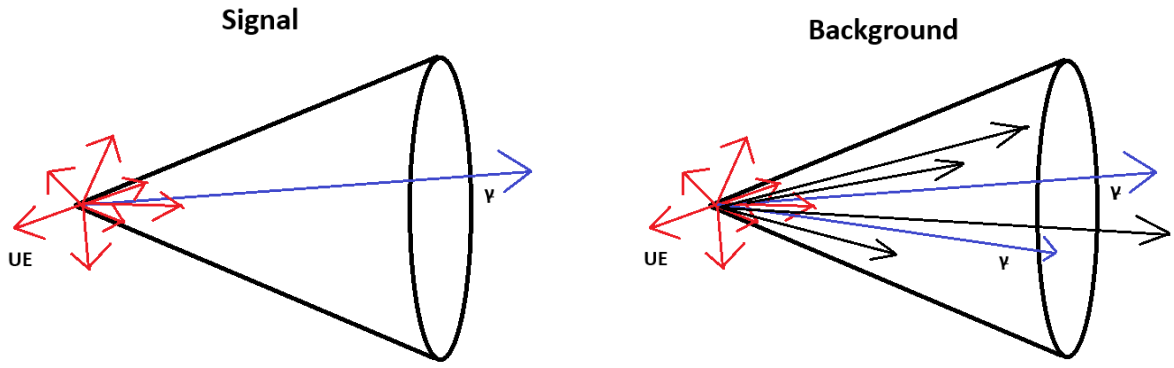
If a cluster has only one local maximum, any starting cell in the cluster will eventually lead to the highest-amplitude cell. However, if a cluster has multiple local maxima, different iterations lead to different cells. Thus, the number of distinct endpoints of the algorithm determines the NLM. Finally, if a cluster spans multiple supermodules, this algorithm incorrectly assigns multiple local maxima due to changes in row and column indices across supermodule boundaries. Therefore, if a cluster falls in multiple supermodules, it is excluded from further processing.

6.2 Direct photon measurement

Due to the high background of photons from neutral meson decays and the relatively small contribution from fragmentation photons, data cuts are necessary to increase the purity of the direct photon signal. These cuts include track matching, isolation, and cluster criteria. The same criteria used in the thesis "Probing the Initial State of Heavy-Ion Collisions with Isolated Prompt Photons" [26] are applied here and will be further outlined below.

First, a track-matching criterion will be applied. Track matching to the EMCAL clusters is performed in the data reconstruction. The process of the track-matching algorithm is further explained in Sec. 4. By associating tracks and their transverse momentum with a cluster, the cluster can be identified as being triggered by a photon rather than by an electron/positron or due to a hadronic contribution. The matched track is required to lie within a radial distance of $|\Delta\varphi| < 0.05$ in azimuth and $|\Delta\eta| < 0.05$ in pseudorapidity relative to the cluster. Additionally, if $\frac{\mathrm{E_{Clust}}}{P_{\mathrm{T,track}}} < 1.75$ (where $\mathrm{E_{Clust}}$ is the energy of the cluster and $P_{\mathrm{T,track}}$ is the transverse momentum of the track), the track's contribution to the cluster is considered too large, and the cluster will be rejected.

The cluster of the candidate photon must fulfill the following requirements. Only clusters with $NLM < 2$ are considered, as clusters with multiple NLM most likely originated from photons produced in neutral meson decays. This will, however, not filter all photons from neutral mesons. Direct photons produce narrow clusters. Therefore a criterion can be defined that the longitudinal axis of the clusters should satisfy $\sigma_{\mathrm{Long}} < 0.3$. This criterion helps reduce the background from neutral meson decay photons, which tend to produce larger clus-



(a) (b)

Figure 12: Sketch of the background in the vicinity of a direct photon (a) and of the 2 photons from a neutral meson decay (b)

ters. To further minimize contributions from exotic clusters, a lower threshold of $0.1 < \sigma_{\text{Long}}$ is also applied. Additionally, the cluster is required to contain at least two cells ($N_{\text{cells}}^{\text{clust}} \geq 2$).

Direct photons originating from Compton scattering and quark-antiquark annihilation are produced during hard scattering processes. These photons have no hadronic activity in their vicinity, apart from contributions from the underlying event (UE). The UE includes processes such as beam remnant fragmentation, initial- and final-state radiation, and multiple parton interactions [27]. In contrast, photons originating from neutral meson decays are embedded within particle jets, resulting in a high probability of being surrounded by jet fragments. Therefore, these photons are associated with high hadronic activity in their vicinity. Which makes direct photons hadronically isolated in comparison with photons from a neutral meson decay. This is illustrated schematically in Fig. 12 (a). In this sketch, red arrows represent the UE, the blue arrow denotes the direct photon, and the cone indicates the photon's vicinity. For comparison, Fig. 12 (b) provides a schematic sketch of two photons originating from the decay of a neutral meson within a particle jet. Here, the red arrows again represent the UE, the blue arrows denotes the photons, and the black arrows indicate other jet particles.

6.2.1 Perpendicular cone method

To quantify the degree of isolation of a photon, the underlying event (UE) density, which varies for each event, must first be determined. This is achieved using the perpendicular cone method. Specifically, a cone of radius R is drawn around the candidate photon, as illustrated in Fig. 12. The cone is then rotated by $\Delta\varphi = \pm\pi/2$ (or $\pm 90°$). The transverse momentum track density within these two rotated cones is calculated using the area of the cone and the sum of tracks that fall within the cone. In principle, the area of a cone is given by πR^2. However, if the cone extends beyond the acceptance boundary of the EMCAL ($|\eta| < 0.9$), the effective cone area must be adjusted using a correction factor, κ.

Figure 13 shows a sketch of a circle that extends beyond the acceptance boundary of the EMCAL. The boundary is indicated by the horizontal line separating regions B and C. Here, R denotes the radius of the circle, L is the shortest distance from the EMCAL boundary to the circle's edge, η represents the circle's center, and A, B, and C correspond to distinct areas within the sketch. The correction factor, κ, represents the fraction of the circle that lies within the EMCAL acceptance. To calculated this fraction, the area of region C must be determined. This calculation begins with the evaluation of θ, A, and B using Eqs. 10, 12, and 11 and basic trigonometric relationships. The area of C is obtained using Eq. 13, which uses the areas of A, B, and the area of a circle. Finally, the correction factor κ is defined by Eq. 14 as the fraction of the circle's area that lies within the EMCAL acceptance. Note that if the area of C equals zero, the entire circle lies within the EMCAL acceptance, resulting in $\kappa = 1$, which indicates that no correction is required.

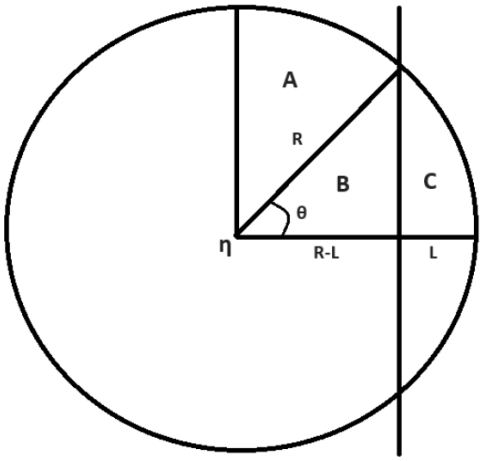


Figure 13: Sketch of a circle which falls over the acceptance boundary of the EMCAL

$$\theta = \cos^{-1}(\frac{R-L}{R}) \tag{10}$$

$$B = \frac{1}{2}(R-L)R\sin(\theta) \tag{11}$$

$$A = \int_0^{\frac{\pi}{2}-\theta}\int_0^R rdrd\theta = (\frac{\pi}{2}-\theta)\frac{1}{2}R^2 \tag{12}$$

$$C = \frac{1}{4}\pi R^2 - (A+B) \tag{13}$$

$$\kappa = \frac{\pi R^2 - 2C}{\pi R^2} \tag{14}$$

Figure 14 shows a graph of the correction factor (or area fraction) κ as a function of $|\eta|$ for different circle radii. The correction factor κ equals 1 when the entire circle lies within the EMCAL acceptance. As the circle intersects the acceptance boundary, κ decreases. Until the center of the circle crosses the boundary and κ reaches its minimum value of 0.5.

With the correction factor κ determined, the density of the UE can be calculated using Eq. 15. Here, ρ_{perp} represents the density, $P_{\text{T}}^{\text{track}}$ denotes the transverse momentum of a track, R is the cone radius, and R_i is the radial distance in the $\eta - \varphi$ plane between the track and the candidate photon. Once the UE density is determined, it is compared to the hadronic activity near the candidate photon. This is done by using the variable $P_{\text{T}}^{\text{iso}}$ defined in Eq. 16. A negative or near-zero value of $P_{\text{T}}^{\text{iso}}$ indicates that the photon is isolated relative to the average background, while a larger positive value suggests significant hadronic activity in its vicinity. In this analysis, a cone radius of $R = 0.4$ was used. A photon is considered a direct photon candidate if $P_{\text{T}}^{\text{iso}} < 1.5$. However, if $P_{\text{T}}^{\text{iso}} > 1.5$, the photon is still further analyzed for background studies. Note that κ cancels out in Eq. 16 due to its inclusion in Eq. 15. Therefore, κ was not explicitly calculated in this analysis. However, for studies focusing on UE density or when an alternative method is used to calculate the UE density. then the inclusion of κ in the analysis would be necessary. Lastly, All the data cuts discussed above, as well as those described in Sec. 5, are summarized in Table 2 for clarification.

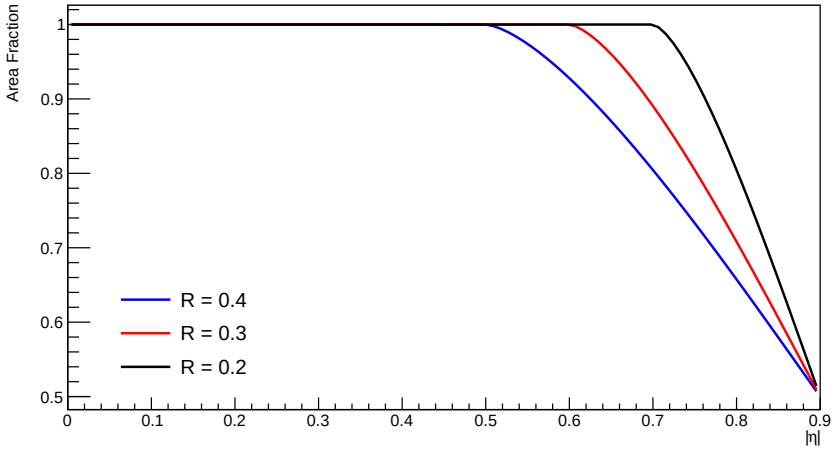


Figure 14: Graph of the area fraction (correction factor) κ as a function of $|\eta|$ for different radii R

$$\rho_{\text{perp}} = \frac{\sum_{R_i<R} P_{\text{T,i}}^{\text{track}}}{2\kappa\pi R^2} \text{ with } R_i = \sqrt{(\varphi_i - (\varphi_\gamma \pm \pi/2))^2 + (\eta_i - \eta_\gamma)^2} \tag{15}$$

Table 2: Summarization of all the cluster selections and photon identification criteria

Cluster variable	Selection criterion
Minimum cluster energy	$\mathrm{E_{Clust}} > 0.7\ GeV$
Cluster time	$\lvert t_{\mathrm{clust}}\rvert < 18\ ns$
Exoticity	$F_+ < 0.97$
Minimum number of cells per cluster	$N_{\mathrm{clust}}^{\mathrm{cells}} \geq 2$
Number of local maxima (NLM)	$\mathrm{NLM} < 2$
Longitudinal cluster axis	$0.1 < \sigma_{\mathrm{Long}} < 0.3$
Cluster track matching radial distance	$\lvert\Delta\eta\rvert \leq 0.05$, $\lvert\Delta\varphi\rvert \leq 0.05$
Track contribution to cluster	$\frac{\mathrm{E_{Clust}}}{P_{\mathrm{T}}^{\mathrm{track}}} > 1.75$

$$P_{\mathrm{T}}^{\mathrm{iso}} = \sum_{R_i<R} P_{\mathrm{T,i}}^{\mathrm{track}} - \rho_{\mathrm{perp}}\kappa\pi R^2 \tag{16}$$

6.3 Background analysis

The objective of the background analysis is to estimate the background (dominated by π^0 decays) present in the direct photon signal. The data selections described in Sec. 6.2 define the signal region where the background must be estimated. However, to perform a background study, it is also necessary to define a region dominated by background events. As will be explained in Sec. 6.3.2. This region is identified using the ABCD method, which is outlined below in Sec. 6.3.1. Once the background dominated region is defined, the Monte Carlo Datasets (MCD's) are normalized to the Real Dataset (RDS) in this region. The resulting normalization factor is then applied to the MCDs in the signal region to estimate the background contribution.

6.3.1 ABCD method

A commonly used data-driven approach for background estimation at the LHC is the ABCD method. The procedure is described in the following paragraphs and follows the explanation provided in Ref. [28]. This method divides the data into a signal region (A) and three control regions (B, C, and D), where the background is the dominant contribution. Two independent selection criteria are required to define the signal region. In this analysis, the isolation observable ($P_{\mathrm{T}}^{\mathrm{iso}}$) and the shower shape ($\sigma_{\mathrm{Long}}$) are used for this purpose. By inverting these criteria, the three control regions are constructed. The definitions of these regions are summarized in Tab. 3, where N_{A} represents the yield in region A, and analogous definitions apply for regions B, C, and D. It is important to note that the signal region (A) includes isolated and narrow clusters, which are expected to contain the signal. The signal and control regions are also illustrated in a two-dimensional data plane, as shown in Fig. 15. In this plane, the data distribution for the shower shape (on the x-axis) and the isolation criterion (on the y-axis)

are presented for clusters within the energy range of $4 < E_{\text{cluster}} < 100$ GeV.

Table 3: Definition of the signal and control regions for the ABCD method

Region	**Isolation ($P_{\text{T}}^{\text{iso}}$)**	**Shower shape (σ_{Long})**	**Yield**
A	$P_{\text{T}}^{\text{iso}} \leq 1.5 GeV/c$	$0.1 < \sigma_{\text{Long}} < 0.3$	N_{A}
B	$P_{\text{T}}^{\text{iso}} \leq 1.5 GeV/c$	$0.4 < \sigma_{\text{Long}} < 2.0$	N_{B}
C	$P_{\text{T}}^{\text{iso}} \geq 4.0 GeV/c$	$0.1 < \sigma_{\text{Long}} < 0.3$	N_{C}
D	$P_{\text{T}}^{\text{iso}} \geq 4.0 GeV/c$	$0.4 < \sigma_{\text{Long}} < 2.0$	N_{D}

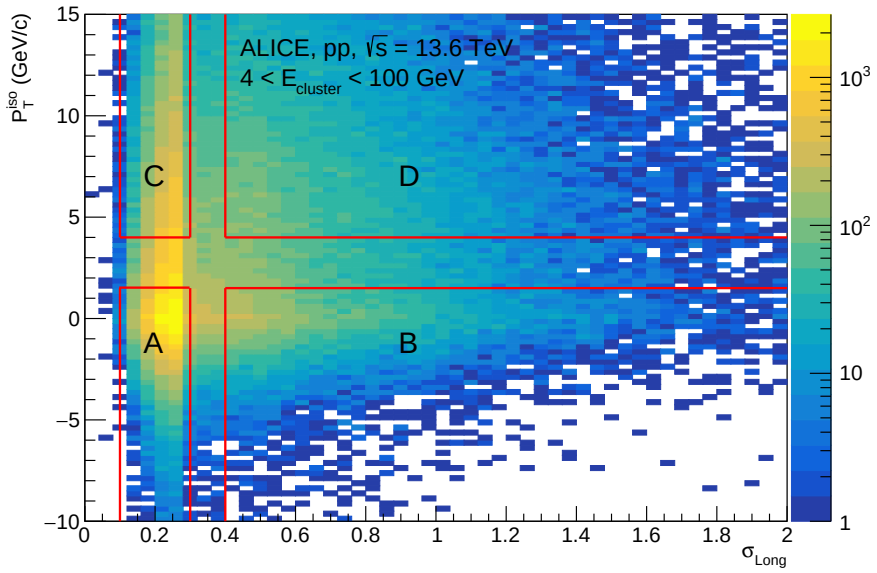


Figure 15: 2 dimensional data plane of the shower shape (σ_{Long} and isolation criterion $P_{\text{T}}^{\text{iso}}$ with the defined signal region (A) and control regions (B, C and D) for the ABCD-method.

The objective of the ABCD method is to estimate the background ($N_{\text{A}}^{\text{bkg}}$) in the signal region A. Once this background is known, the purity of the signal can be easily calculated using $P = \frac{N_{\text{A}} - N_{\text{A}}^{\text{bkg}}}{N_{\text{A}}}$, where P denotes the purity. The ABCD method relies on the assumption that the observables in the ABCD plane are uncorrelated. Specifically, in this case, it is assumed that the shower shape and the isolation criterion ($P_{\text{T}}^{\text{iso}}$) are uncorrelated. Thus, the relative difference in background yield when changing between the isolated ($P_{\text{T}}^{\text{iso}} \leq 1.5$ GeV/c) and anti-isolated ($P_{\text{T}}^{\text{iso}} \geq 4.0$ GeV/c) regions should be independent of the cluster shape. This implies that the ratio of the background yields between regions B and D and between regions A and C should be equal. This relationship is captured in Eq. 17, where $N_{\text{B}}^{\text{bkg}}$ denotes the background yield in region B, with analogous definitions for regions A, C, and D.

$$\frac{N_{\mathrm{B}}^{\mathrm{bkg}}}{N_{\mathrm{D}}^{\mathrm{bkg}}} = \frac{N_{\mathrm{A}}^{\mathrm{bkg}}}{N_{\mathrm{C}}^{\mathrm{bkg}}} \tag{17}$$

The second assumption made is that the signal contribution in the control regions B, C, and D is negligible. Thus, Eq. 17 can be rewritten as Eq. 18, providing an estimate for the background yield in the signal region A and, consequently, the purity of the signal. However, these assumptions cannot be made blindly and should be tested before being used. Reference [26] shows that the assumption of negligible signal yield in control regions B, C, and D is valid. However, it also shows that there is a correlation between the shower shape (σ_{Long}) and the isolation criterion ($P_{\mathrm{T}}^{\mathrm{iso}}$). Therefore, a correction factor (α) should be applied to Eq. 18. This correction factor can be calculated using Monte Carlo data, which introduces a new assumption: that the correlation between the shower shape (σ_{Long}) and the isolation criterion ($P_{\mathrm{T}}^{\mathrm{iso}}$) is correctly captured by the Monte Carlo simulations. This assumption is also tested in Ref. [26] and shows that the correction factor can only be estimated with limited accuracy. The accuracy of this parameter becomes increasingly important for low P_{T} ranges, where a small fluctuation in the correction factor results in a large fluctuation in the purity.

$$N_{\mathrm{A}}^{\mathrm{bkg}} = \frac{N_{\mathrm{B}} N_{\mathrm{C}}}{N_{\mathrm{D}}} \tag{18}$$

6.3.2 Background normalization

The Monte Carlo Datasets (MCDs) used in this analysis do not contain a direct photon signal. Therefore, it might seem straightforward to present the Monte Carlo results in the signal region (region A from the ABCD method) as the background contribution estimate. However, while Monte Carlo simulations have an excellent performance, they do not fully represent nature. To minimize bias in the background contribution estimate, it is necessary to carefully compare the Monte Carlo data with the RDS. This comparison is performed by evaluating the shape of the $P_{\mathrm{T}}^{\mathrm{iso}}$ variable for both datasets. If the shapes are compatible, the Monte Carlo data can be normalized to the observed data. The initial comparison will be performed using the general purpose Monte Carlo dataset, which is expected to be compatible with the ODS. Thereafter, the same will be done for the jet-jet Monte Carlo dataset to try to improve the statistics for higher cluster energies. However, due to its $\hat{P}_{\mathrm{T}} = 5$ GeV and the still relatively low statistics of the RDS, this dataset may be less compatible.

Since the Monte Carlo dataset does not include a direct photon signal, a region in the real data must be defined where the signal contribution is negligible. For this region, region D from the ABCD method is chosen as this region is expected to contain the least amount of signal. This region is specified in the two-dimensional $P_{\mathrm{T}}^{\mathrm{iso}}$ and σ_{Long} data plane as $0.4 < \sigma_{\mathrm{Long}} < 2.0$ and $P_{\mathrm{T}}^{\mathrm{iso}} \geq 4.0$ GeV. The normalization factor, α, is calculated using Eq. 19, where N_{data} and N_{MC} represent the yield in the previous specified region for real and Monte Carlo data. To estimate the uncertainty in α, region D is divided into three subregions. Region 1 is located centrally within region D, while regions 2 and 3 are positioned on either side of region 1. Regions 2 and 3 overlap partially with region 1 but also include

data points outside of it. The normalization factor is calculated independently for each of the three subregions, with the variation among these values providing an estimate of the uncertainty in α. Once calculated, α is used to scale the Monte Carlo data in the signal region. Providing an estimate of the background contribution in the direct photon signal.

$$\alpha = \frac{N_{\text{data}}}{N_{\text{MC}}} \tag{19}$$

7 Results

This chapter is divided into two main sections. First, the results of mimicking the kV1 clusterizer and its comparison to the behavior of the kV3 clusterizer is shown in Sec. 7.1. Secondly, the results of the background estimation for the direct photon signal, obtained using the method described in Sec. 6.2, are presented in Sec. 7.2.

7.1 Clusterizer

In this section, the results of mimicking the kV1 clusterizer are presented. The results of the kV3 clusterizer and the adjusted kV3 clusterizer are compared using the method described in Sec. 4. The parameters used in the kV3Default clusterizer to mimic the kV1 clusterizer are the following: seed energy $E_{Seed} = 500$ MeV, aggregation threshold $E_{\text{agg}} = 100$ MeV, and gradient cut `doGradientCut = false`. This clusterizer is defined in the EMCALClusters.h file within the O2Physics framework under the name kV3NoSplit. First, the results of the reconstruction of the invariant mass of the π^0 meson will be presented. Next, the results for the number of local maxima (NLM) will be discussed. Finally, the cluster shape and cluster size variables will be evaluated.

7.1.1 Reconstruction results of the invariant mass of the π^0

In Fig. 16 (a), the paired photon mass distribution for a transverse momentum range of $2 < P_{\text{T,meson}} < 3$ GeV/c for the hypothetical meson is presented. The kV3Default clusterizer is indicated by black markers, while the kV3NoSplit clusterizer is indicated by red markers. Both clusterizers show a clear peak at the mass of the π^0 meson ($M = 0.135$ GeV) [29], indicating that both clusterizers are able to reconstruct the π^0 meson mass within this momentum range. In Fig. 16 (b), the mass distribution for the range $7 < P_{\text{T,meson}} < 9$ GeV/c is shown. While both clusterizers are still able to reconstruct the π^0 mass, the kV3NoSplit clusterizer has a slightly reduced performance compared to the kV3Default clusterizer. This behavior becomes more visible in Fig. 16 (c), where the mass distribution is shown for $9 < P_{\text{T,meson}} < 11$ GeV/c. In this case, the kV3NoSplit clusterizer performs significantly worse than the kV3Default clusterizer in reconstructing the π^0 mass. Finally, Fig. 16 (d) shows the mass distribution for the range $11 < P_{\text{T,meson}} < 13$ GeV/c. Here, the π^0 mass peak

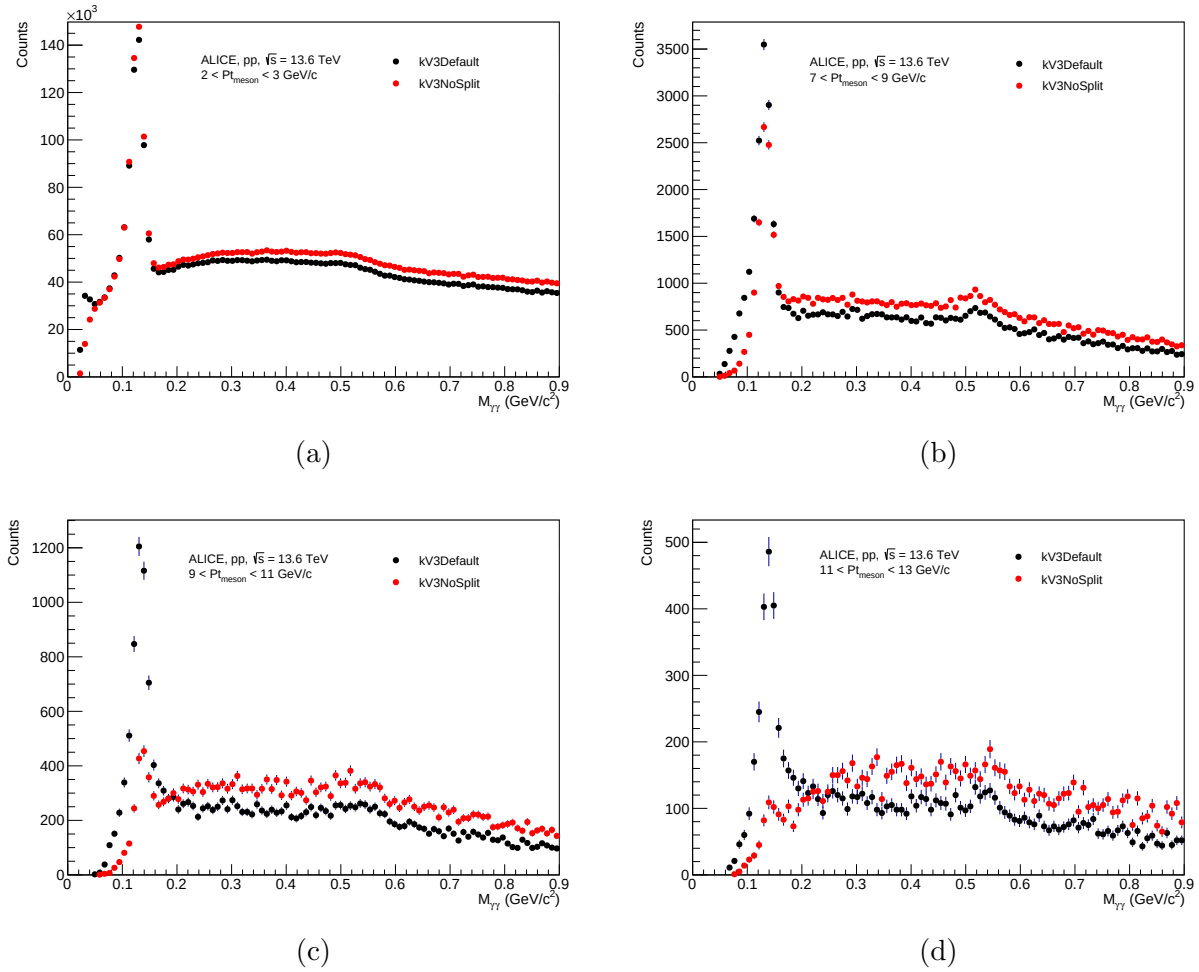
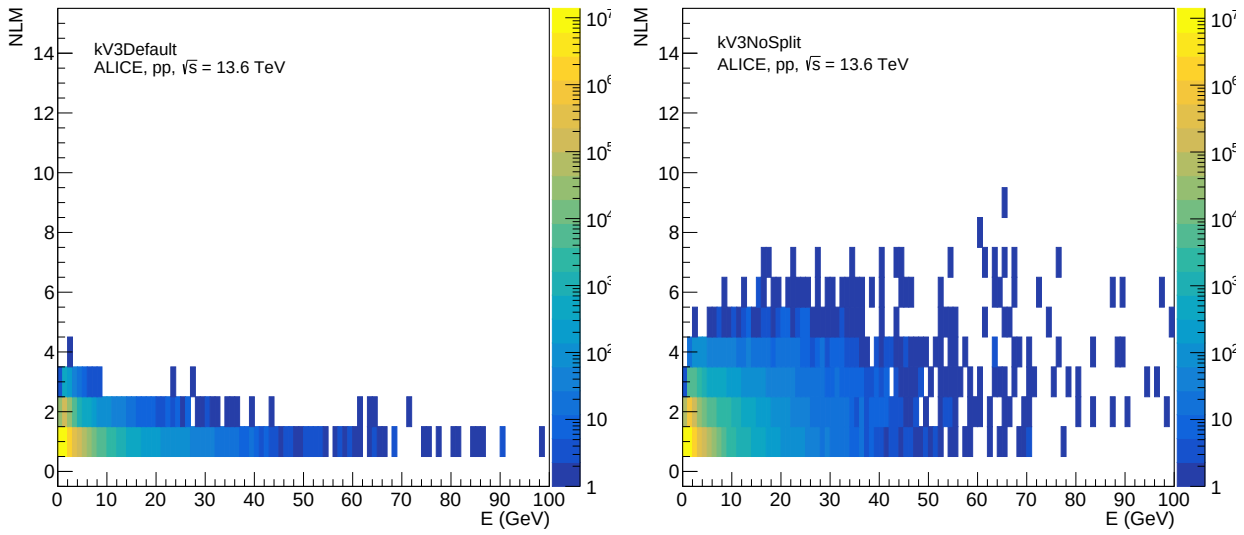


Figure 16: Paired photon mass distributions showing the π^0 mass peak for the kV3Default clusterizer (black) and the kV3NoSplit clusterizer (red). The distributions are presented for transverse momentum (P_T) ranges of $2 < P_{\mathrm{T,meson}} < 3$ GeV/c (a), $7 < P_{\mathrm{T,meson}} < 9$ GeV/c (b), $9 < P_{\mathrm{T,meson}} < 11$ GeV/c (c), and $11 < P_{\mathrm{T,meson}} < 13$ GeV/c (d).

is absent for the kV3NoSplit clusterizer, while the kV3Default clusterizer is still able reconstruct the π^0 mass. This indicates that at higher $P_{\mathrm{T,meson}}$ values, the π^0 mass reconstruction algorithm for the kV3NoSplit clusterizer is not able to pair clusters originating from photons produced in π^0 decays. Thus, clusters from π^0 decays begin to merge at lower $P_{\mathrm{T,meson}}$ for the kV3NoSplit clusterizer compared to the kV3Default clusterizer.

7.1.2 NLM results

The results of the NLM (Number of Local Maxima) calculation are presented in Fig. 17 (a) and (b), where the NLM is plotted as a function of cluster energy. Figure 17 (a) shows the results for the kV3Default clusterizer, while Fig. 17 (b) shows the results for the kV3NoSplit clusterizer. The kV3NoSplit clusterizer has a wider spread in NLM, indicating a higher



(a) (b)

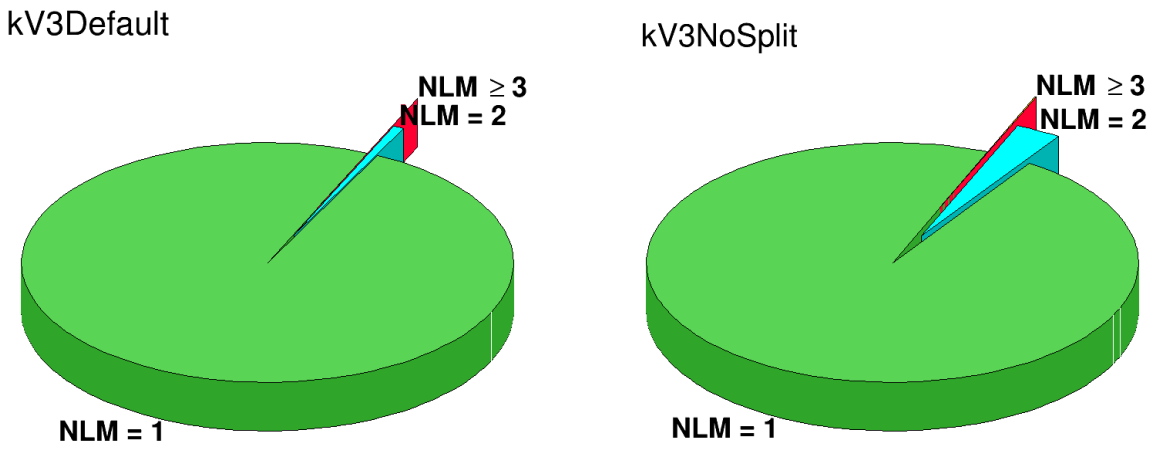
Figure 17: NLM plotted as a function of cluster energy for the kV3Default clusterizer (a) and the kV3NoSplit clusterizer (b)

probability of cluster merging for events such as neutral meson decays. The presence of clusters with multiple NLM in the kV3Default clusterizer can be explained by the gradient cut ($E_{\mathrm{Grad}} = 30$ MeV), which allows small variations due to noise and detector resolution within the cluster. This effect is discussed in more detail in Sec. 4.

The fractions of $NLM = 1$, $NLM = 2$, and $NLM \geq 3$ are illustrated in the pie charts in Fig. 18 (a) and (b). Figure 18 (a) represents the fractions for the kV3Default clusterizer, while Fig. 18 (b) shows the corresponding results for the kV3NoSplit clusterizer. For both clusterizers, the $NLM = 1$ fraction is the largest, indicating that most clusters have a single local maximum. However, the $NLM = 2$ fraction is significantly higher for the kV3NoSplit clusterizer compared to the kV3Default clusterizer, suggesting that the kV3NoSplit clusterizer has a greater probability for cluster merging. The $NLM \geq 3$ fraction is extremely small for both clusterizers, indicating that clusters with three or more local maxima are rare.

7.1.3 Cluster shape and Cluster size evaluation

The cluster shape is characterized by its longitudinal axis (σ_{Long}). Figure 19 shows the normalized distribution of σ_{Long} for various cluster energy ranges. In Fig. 19(a), the distribution is shown for the kV3Default clusterizer, while Fig. 19(b) shows the distribution for the kV3NoSplit clusterizer. For clusters with energies $0 < \mathrm{E}_{\mathrm{cluster}} < 5$ GeV (represented by black markers), both clusterizers exhibit similar behavior. Both distributions have a peak at $\sigma_{\mathrm{Long}} \approx 0.3$, indicating that most clusters fall within this region. Additionally, a secondary peak at $\sigma_{\mathrm{Long}} \approx 0.6$ is observed, which is due to to clusters consisting of two to four cells



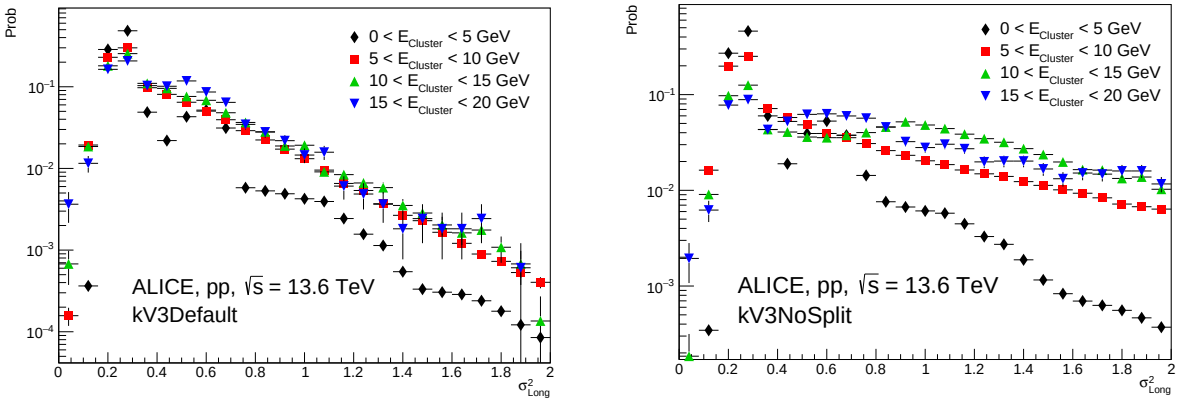
(a) (b)

Figure 18: The fractions of $NLM = 1$, $NLM = 2$, and $NLM \geq 3$ illustrated in the pie charts for the kV3Default clusterizer (a) and the kV3NoSplit clusterizer (b)

arranged in either a longitudinal or L-shaped configuration.

Clusters with energies $5 < \mathrm{E_{cluster}} < 10$ GeV (represented by red markers) also have a peak at $\sigma_{\mathrm{Long}} \approx 0.3$ for both the kV3Default and kV3NoSplit clusterizers. However, the kV3NoSplit clusterizer shows a higher probability of forming broader clusters. For clusters in the energy range $10 < \mathrm{E_{cluster}} < 15$ GeV (represented by green markers), the kV3Default clusterizer maintains behavior similar to that for energies of $5 < \mathrm{E_{cluster}} < 10$ GeV. In contrast, the kV3NoSplit clusterizer shows a broad secondary peak at $\sigma_{\mathrm{Long}} \approx 1.0$. This peak originates from the merging of clusters produced by π^0 decays, resulting in broader clusters. Clusters with energies $15 < \mathrm{E_{cluster}} < 20$ GeV have a narrower peak that shifts to $\sigma_{\mathrm{Long}} \approx 0.6$, again due to photons from π^0 decays. At higher energies, the angle between photons from π^0 decays decreases, resulting in merged clusters with larger overlap regions and thus reducing their overall size. However, the kV3Default clusterizer does not exhibit this behavior for $15 < \mathrm{E_{cluster}} < 20$ GeV. Instead, its distribution remains similar with what was observed for lower energy clusters.

The size of a cluster is determined by the number of cells it contains. The number of cells as a function of the cluster energy is shown in Fig. 20. With Fig. 20(a) presenting the results for the kV3Default clusterizer and Fig. 20(b) for the kV3NoSplit clusterizer. For both clusterizers, the majority of clusters have a relatively small size, with $NCells \approx 2$ to 5. However, the kV3NoSplit clusterizer has a much broader distribution of cluster sizes. And thus has a higher probability of finding a larger cluster compared to the kV3Default clusterizer. This behavior is expected, as the kV3NoSplit clusterizer continues adding cells to a cluster until the cell energy falls below the aggregation threshold.



(a) (b)

Figure 19: Normalized distribution of the longitudinal cluster shape (σ_{Long}) axis for different energy ranges for the the kV3Default clusterizer (a) and the kV3Nosplit clusterizer (b)

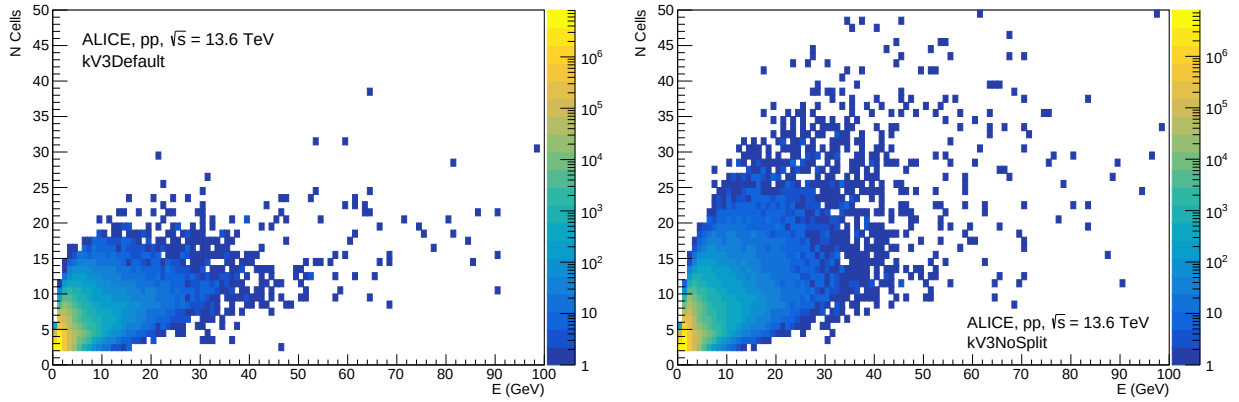
In conclusion, the kV3NoSplit clusterizer shows cluster merging for photons from neutral meson decays at significantly lower cluster energies compared to the kV3Default clusterizer. This behavior is visible in Figs. 16 and 19, where no merging of clusters is observed for the kV3Default clusterizer. As those clusters would only merge at very high energies when both photons fall in the same cell. Additionally, Figs. 17 and 18 demonstrate that a larger fraction of clusters exhibit two or more local maxima when using the kV3NoSplit clusterizer. Finally, Fig. 20 indicates an increase in cluster size for some clusters with the kV3NoSplit clusterizer. These findings are promising, as they suggest that clusters originating from neutral meson decays can be distinguished from direct photons by analysing the clustershape, NLM, and NCells variables from the kV3NoSplit clusterizer. Therefore, the kV3NoSplit clusterizer will be used for the remainder of this article and the background analysis.

7.2 Background estimation

In this section, the results of estimating the background contribution in the direct photon signal are presented. The direct photon signal was obtained by using the method outlined in Sec. 6.2, which closely follows the approach used in Ref. [26]. And by using the kV3NoSplit clusterizer. Firstly, the background estimation will be shown for the general-purpose Monte Carlo simulation. Secondly, the results obtained from the jet-jet Monte Carlo simulation will be presented.

7.2.1 General purpose MC

In Fig. 21, the normalized distribution of $P_{\mathrm{T}}^{\mathrm{iso}}$ is shown for an energy range of $2 < E_{\mathrm{cluster}} < 10$ GeV and a longitudinal cluster axis of $0.4 < \sigma_{\mathrm{Long}} < 2.0$. The black markers in the



(a) (b)

Figure 20: The number of cells in a cluster (N Cells) as a function of the energy (E) of a cluster for the kV3Default clusterizer (a) and the kV3NoSplit clusterizer (b)

graph represent the real data points, while the red markers correspond to the general purpose Monte Carlo simulation. As expected, the shape of the Monte Carlo distribution closely resembles the experimental data. The background-dominated D region, as defined by the ABCD method, is in the range of $4 < P_{\mathrm{T}}^{\mathrm{iso}} < 15$ GeV. Within this region, three subregions are defined, as also visualized in Fig. 21. Region 1, located in the middle, spans $7 < P_{\mathrm{T}}^{\mathrm{iso}} < 12$ GeV and is visualized between the red vertical lines. Regions 2 and 3, on either side of Region 1, are defined as $4 < P_{\mathrm{T}}^{\mathrm{iso}} < 9$ GeV and $10 < P_{\mathrm{T}}^{\mathrm{iso}} < 15$ GeV, respectively. And are indicated between the black dotted lines. For these three regions, the normalization factor is calculated using Eq. 19. The differences in the normalization factors between these regions provide an estimate of the uncertainty of the normalization method.

In Fig. 22, the background estimation is shown for a direct photon signal within the energy range of $2 < \mathrm{E}_{\mathrm{cluster}} < 10$ GeV and a longitudinal cluster axis of $0.1 < \sigma_{\mathrm{Long}} < 0.3$. The red markers represent the re-normalized results from the general-purpose Monte Carlo simulation, obtained using the normalization method described above. The black markers represent the the experimental data. The blue band indicates the uncertainty of the normalization method, while the red band indicates the statistical uncertainty. Since the normalization uncertainty comes from a scaling factor, the error appears smaller in regions with low data yield, such as $-10 < P_{\mathrm{T}}^{\mathrm{iso}} < -5$. In contrast, the statistical uncertainty is relatively larger in the low-yield regions. In the signal region ($P_{\mathrm{T}}^{\mathrm{iso}} < 1.5$ GeV), the experimental data has a higher yield compared to the background estimation. However, in the background region ($P_{\mathrm{T}}^{\mathrm{iso}} > 4.0$ GeV), the re-normalized Monte Carlo results and the experimental data begin to coincide. This behavior can be explained by the presence of the direct photon signal in the experimental data, although other factors may play a role as well. This will be further outlined in Sec. 8.

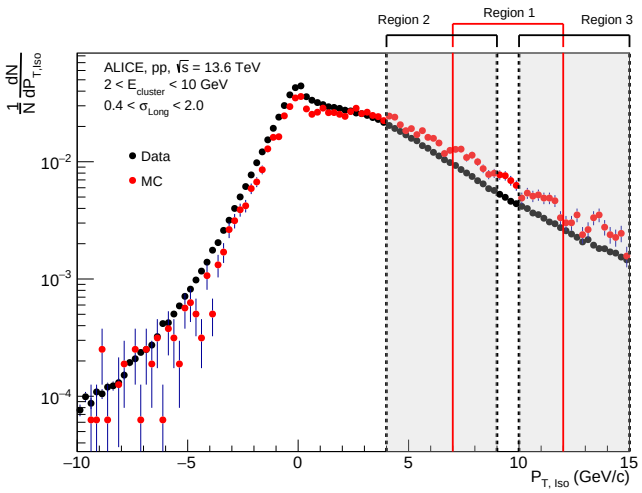


Figure 21: Normalized $P_{\mathrm{T}}^{\mathrm{iso}}$ distribution for experimental data (black markers) and general purpose Monte Carlo data (red markers) within the range $2 < \mathrm{E}_{\mathrm{cluster}} < 10$ GeV and $0.4 < \sigma_{\mathrm{Long}} < 2.0$. In the background-dominated (D) region, three subregions are defined for the re-normalization method.

7.2.2 Jet-jet MC

To improve the statistics of higher-energy clusters, the jet-jet Monte Carlo dataset was used. In Fig. 23, the normalized distribution of $P_{\mathrm{T}}^{\mathrm{iso}}$ is shown for an energy range of $5 < \mathrm{E}_{\mathrm{cluster}} < 20$ GeV and $0.4 < \sigma_{\mathrm{Long}} < 2.0$. The experimental data is represented by the black markers, while the jet-jet Monte Carlo data is represented by the red markers. It is clear that the jet-jet Monte Carlo data follows a different distribution compared to the experimental data. Therefore, re-normalizing the jet-jet Monte Carlo to match the experimental data would not be appropriate.

The difference between the jet-jet Monte Carlo distribution and the experimental data can be partially explained by the $\hat{P}_{\mathrm{T}} = 5$ GeV constraint. This constraint sets a minimum transverse momentum of 5 GeV for outgoing partons from the hard scattering. Which causes the jet-jet Monte Carlo to lack data at lower energies, which are present in the experimental data. As nature does not impose a constraint on $\hat{P}_{\mathrm{T}}$. Another factor is the matching of highly energetic tracks to incorrect events in the jet-jet production process. Since each collision is weighted with a weight factor, a mismatch can build up into a significant error. A solution to this issue is the implementation of a minimum bias (MB) gap. With this method one or more minimum bias events after each jet-jet event are generated. If a highly energetic track is matched to the wrong event, the event is now classified as a minimum bias event and thus no longer affects the jet-jet signal. This should significantly reduce errors associated with assigning incorrect weights to highly energetic tracks.

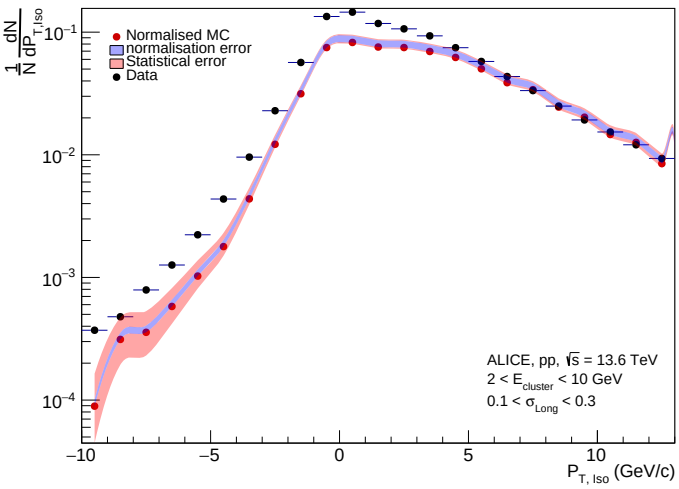


Figure 22: Background estimation for a direct photon signal with an energy range of $2 < \mathrm{E_{cluster}} < 10$ GeV and $0.1 < \sigma_{\mathrm{Long}} < 0.3$ GeV. The re-normalized general purpose Monte Carlo data is represented by the red markers and the experimental data by the black markers. The blue band indicates the uncertainty of the re-normalization method, while the red band indicates the statistical uncertainty.

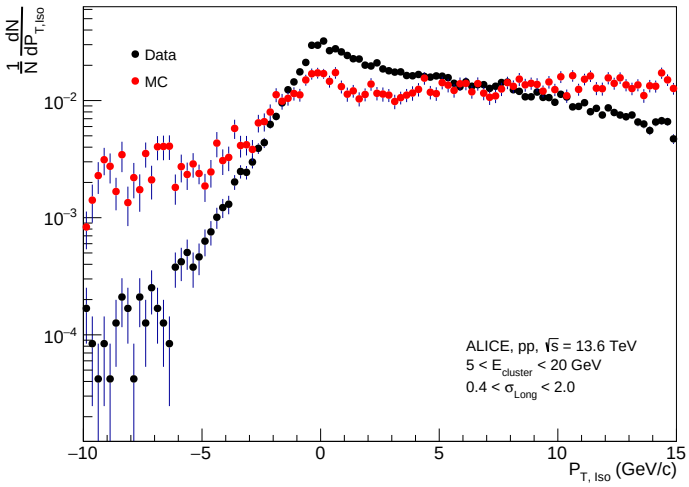


Figure 23: Normalized $P_{\mathrm{T}}^{\mathrm{iso}}$ distribution for experimental data (black markers) and jet-jet Monte Carlo data (red markers) within the range $5 < \mathrm{E_{cluster}} < 20$ GeV and $0.4 < \sigma_{\mathrm{Long}} < 2.0$.

8 Discussion and Conlusion

This chapter provides a summary, conclusion, and discussion of the results presented in Sec. 7. The chapter is divided into two sections for readability. The first section (Sec. 8.1) presents the final conclusions regarding the clusterizer results, while the second section (Sec. 8.2) presents the final conclusions for the background estimation.

8.1 Clusterizer discussion

In Sec. 7.1, the results of the clusterizer analysis were presented. The invariant mass of the π^0 meson was reconstructed using the kV3Default and kV3NoSplit clusterizers. The kV3NoSplit clusterizer has a reduced performance in the range $7 < P_{\mathrm{T,meson}} < 11$ GeV/c and fails to reconstruct the mass in the range $11 < P_{\mathrm{T,meson}} < 13$ GeV/c. This behavior suggests that clusters originating from π^0 decays begin to merge at significantly lower energies for the kV3NoSplit clusterizer compared to the kV3Default clusterizer. As the clusters for the kV3Default clusterizer only start to merge at very high energies, where the angle between the decay photons becomes sufficiently small for both photons to fall into the same or adjacent cells of the EMCAL.

Cluster variables were compared between the two clusterizers. The kV3NoSplit clusterizer shows a broader distribution of local maxima, reaching up to approximately 8, while the kV3Default clusterizer shows a more narrower distribution, with a maximum of about 3. The multiple local maxima in the kV3Default clusterizer can be explained by the gradient cut, as outlined in Secs. 7.1 and 6.1. The fraction of clusters with $NLM = 2$ was shown to be significantly larger for the kV3NoSplit clusterizer compared to the kV3Default clusterizer. Similarly, the fraction of clusters with $NLM \geq 3$ is also higher for the kV3NoSplit clusterizer. However, the $NLM \geq 3$ fraction is negligible relative to $NLM = 1$ and $NLM = 2$ fractions.

The kV3NoSplit clusterizer shows higher values for larger σ_{Long} compared to the kV3Default clusterizer, indicating a higher probability of forming larger clusters. This is further argumented by the fact that the kV3NoSplit clusterizer has a broader distribution in the number of cells per cluster compared to the kV3Default clusterizer. Which also indicates a that the kV3NoSplit clusterizer has a higher probability to form larger clusters.

To summarize: The kV3NoSplit clusterizer is unable to accurately reconstruct the invariant mass of the π^0 meson at relatively low energies. Additionally, the kV3NoSplit clusterizer tends to produce larger clusters and shows a higher probability of forming clusters with multiple local maxima. This behavior is consistent with the expected performance of the kV1 clusterizer. Therefore, it can be concluded that by modifying the kV3Default clusterizer, one can obtain a clusterizer that mimics the behavior of the kV1 clusterizer. However, its functionality for direct photon detection in lead-lead collisions may be limited. This limitation comes from the increased UE density and the simplistic design of the clusterizer. Since the kV3NoSplit clusterizer has no other terminating point than when a cell's energy falls below the aggregation threshold. The resulting clusters can become excessively large in a lead-lead collision. This can lead to inaccuracies in photon energy and the number of local maxima (NLM) in a cluster. Thus, additional constraints need to be implemented to make the kV3NoSplit clusterizer suitable for lead-lead collisions. Some possible constraints are further outlined in Sec. 9.

8.2 Background estimation discussion

In Sec. 7.2, the results of the background estimation were presented. From this section it can be concluded that the distributions of the experimental data and the general-purpose Monte Carlo simulation are similar in shape. Which makes the background estimation using the general-purpose Monte Carlo simulation a reasonable approach. In the background estimation the experimental data showed a higher yield in the signal region compared to the background estimation. The observed difference is approximately 30% while from previous studies the signal contribution is expected to be approximately 5% [26]. Therefore the difference between the experimental and Monte Carlo data is too large to be only attributed to a direct photon signal. To fully understand the underestimation of the background contribution investigation of the event generation and detector simulation is necessary. Furthermore, contributions from other sources, such as fragmentation photons, might be underestimated in the Monte Carlo production. Furthermore, the general-purpose Monte Carlo simulation was anchored to the apass6 reconstruction, while the experimental data was anchored to the apass7 reconstruction. Although no significant changes were made to the EMCAL some tracking updates could introduce errors in the $P_{\mathrm{T}}^{\mathrm{iso}}$ variable. However, these errors are very challenging to estimate.

A second background estimation was attempted using a jet-jet Monte Carlo production to improve statistics at higher cluster energies. However, the shapes of the distributions did not align. This is probably due to the absence of a minimum bias (MB) gap in the jet-jet simulation, as discussed in Sec. 7.2.2. Furthermore, the $\hat{P}_{\mathrm{T}} > 5$ GeV/c restriction in the jet-jet production introduces inconsistencies with the experimental data at low energies.

in conclusion, the background estimation underestimates the background contribution in the direct photon signal. This can be due to a number of factors, including the event generation and detector simulation processes in the Monte Carlo data. Furthermore, contributions from fragmentation photons may be underestimated. The Different reconstruction of the experimental and Monte Carlo data could also contribute to the underestimation. Additionally, inconsistencies between the experimental data and the jet-jet Monte Carlo production prevent a reliable study of higher-energy clusters.

9 Outlook

As concluded in Sec. 8.1, the clusterizer developed in this thesis is well-suited for the measurement of direct photons in proton-proton collisions. However, problems may arise when the clusterizer is used in high (UE) density environments, such as lead-lead collisions. The fact that the clusterizer has no other terminating point than when a cell has an energy below the aggregation threshold can lead to excessively large clusters. Therefore, additional constraints must be implemented in the clusterizer to increase its reliability in lead-lead like environments.

A possible solution to this problem is to subtract an average background energy from the EMCAL cells. To estimate this background energy, the EMCAL signal first has to be clus-

terised. After clusterization, some EMCAL cells will not be part of any cluster. The average energy or amplitude of these unclustered cells can give an estimate of the background energy in the EMCAL signal. By only using cells that do not belong to a cluster the background estimation avoids pollution from high-energy signals. Once the average background energy is determined, the clusters can be reevaluated by subtracting the background energy from each cell within a cluster.

Another solution would be to impose a size limitation on the EMCAL clusters. For example restricting clusters to have a maximum size of 5×5 cells. This approach would effectively solve the problem of excessively large clusters. Furthermore, direct photons tend to produce narrow clusters, making their signals detectable within a 5×5 cell region. However, distinguishing the direct photon signal from background photons may become more challenging, as photons from neutral meson decays tend to produce much broader clusters. But it would be worth investigating whether an optimal maximum cluster size exists. Both of the proposed solutions, or a combination of them both, could be further investigated with a Monte Carlo study. In which it is also possible to estimate some systematical errors which come along with these adjustments to the clusterizer.

In Sec. 8.2, it was concluded that background estimation for direct a photon signal at higher cluster energies could not be performed due to inconsistencies between the experimental data and the jet-jet Monte Carlo simulations. As discussed in Sec. 7.2.2, these inconsistencies are due to the absence of a minimum-bias (MB) gap in the jet-jet production. Therefore, the first step toward improving this analysis would be to generate a Monte Carlo dataset that includes an MB gap. Furthermore, higher statistics in the experimental data would further improve the analysis. Because the jet-jet Monte Carlo simulations are constrained to higher energy ranges due to the $\hat{P}_{\mathrm{T}} > 5$ GeV/c requirement, which limits their usage at lower energies.

Further research could focus on a complete direct photon study which includes purity and efficiency estimates. The same isolation criteria used in this thesis could be applied to extract a direct photon signal, though an optimization study of these variables would be an interesting topic. A complete direct photon study would require a photon-jet Monte Carlo simulation to estimate the correction factor for the ABCD method, which captures the correlation between the longitudinal cluster axis (σ_{Long}) and the $P_{\mathrm{T}}^{\mathrm{iso}}$ variable, as discussed in Sec. 6.3.1. Additionally, a skimmed experimental dataset consisting of candidate photons would be preferred, as it would significantly reduce the dataset size. This thesis can serve as a gateway for a complete direct photon analysis in proton-proton collisions. As it provides a suitable clusterizer for direct photon measurements and a method for calculating the number of local maxima. Which is an important variable for distinguishing signal from background. For lead-lead collisions, further adjustments to the clusterizer, as outlined above, would be necessary. Nevertheless, this thesis is a first step towards such an analyses.

References

[1] D. A. Glaser, Noble Lecture, December **12** (1960).

[2] R. Rougemont, J. Grefa, M. Hippert, J. Noronha, J. Noronha-Hostler, I. Portillo, and C. Ratti, Progress in Particle and Nuclear Physics **135**, 104093 (2024).

[3] A. collaboration et al., European Physical Journal C **84** (2024).

[4] M. Gyulassy and M. Plumer (1990).

[5] C. Collaboration, Physics Letters B **710**, 256 (2012).

[6] A. M. Sirunyan, A. Tumasyan, W. Adam, F. Ambrogi, E. Asilar, T. Bergauer, J. Brandstetter, E. Brondolin, M. Dragicevic, J. Erö, et al., Physics Letters B **785**, 14 (2018).

[7] *Wikipedia, standard model of elementary particles* (2024), accessed on November 27, 2024, URL `https://en.wikipedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg`.

[8] M. Thomson, *Modern particle physics* (Cambridge University Press, 2013).

[9] A. Collaboration et al., arXiv preprint arXiv:1207.7214 (2012).

[10] R. Gupta, arXiv preprint hep-lat/9807028 (1998).

[11] F. Karsch, Nuclear Physics A **590**, 367 (1995).

[12] E. Annala, T. Gorda, A. Kurkela, J. Nättilä, and A. Vuorinen, Nature Physics **16**, 907 (2020).

[13] P. Braun-Munzinger and J. Stachel, Nature **448**, 302 (2007).

[14] J. D. Bjorken, Tech. Rep. (1982).

[15] J. D. Bjorken, Physical review D **27**, 140 (1983).

[16] D. A. Appell, Physical Review D **33**, 717 (1986).

[17] J. Wenninger, PoS p. 001 (2019).

[18] A. Tumasyan, W. Adam, J. Andrejkovic, T. Bergauer, S. Chatterjee, K. Damanakis, M. Dragicevic, A. Del Valle, P. Hussain, M. Jeitler, et al., Journal of Instrumentation **19** (2024).

[19] K. Aamodt, A. A. Quintana, R. Achenbach, S. Acounis, D. Adamová, C. Adler, M. Aggarwal, F. Agnese, G. A. Rinella, Z. Ahammed, et al., Journal of Instrumentation **3**, S08002 (2008).

[20] A. Tauro, Tech. Rep. (2017).

[21] S. Acharya, D. Adamová, A. Adler, G. A. Rinella, M. Agnello, N. Agrawal, Z. Ahammed, S. Ahmad, S. Ahn, I. Ahuja, et al., Journal of instrumentation **18**, P08007 (2023).

[22] AliceO2Group, *O2 analysis framework documentation* (2020), accessed on October 3, 2024, URL `https://aliceo2group.github.io/analysis-framework/docs/`.

[23] AliceO2Group, *O2 analysis source code* (2020), accessed on October 9, 2024, URL `https://github.com/AliceO2Group/AliceO2`.

[24] *Pythia8 manual* (2024), accessed on November 6, 2024, URL `https://pythia.org/latest-manual/Welcome.html`.

[25] A. Maevskaya and A. collaboration, Physics of Atomic Nuclei **84**, 579 (2021).

[26] F. Jonas, Tech. Rep. (2023).

[27] M.-J. Tsai, Tech. Rep. (2016).

[28] W. Buttinger (2018).

[29] S. N. et al, Phys. Rev. D 110, 030001 (2024).