

Master's Thesis

Untersuchung schneller Simulationsalgorithmen für den CALICE AHCAL Prototyp einschließlich Methoden zur Anordnung und Gruppierung von Kalorimeterzellen

Investigations of Fast Simulation Algorithms for the CALICE AHCAL Prototype including Methods for Calorimeter Cell Ordering and Grouping

prepared by

Zobeyer Ghafoor

from Göttingen

at the II. Physikalischen Institut

Thesis number: II.Physik-UniGö-MSc-2025/01
Thesis period: 1st April 2024 until 31st March 2025
First referee: Prof. Dr. Stan Lai
Second referee: apl. Prof. Dr. Jörn Große-Knetter

Abstract

This thesis investigates a data-driven fast simulation algorithm for pion showers in a highly granular calorimeter. The dataset has been recorded in a test beam campaign at CERN in 2018 with the AHCAL Technological Prototype of the CALICE Collaboration. The AHCAL was exposed to electron, muon and pion beams of different energies during the test beam run. For this thesis, the entire pion shower dataset has been used.

The algorithm is based on ordering the calorimeter cells of the AHCAL in ascending order based on their distance to the centre of gravity of the pion shower. This assigns a higher importance during simulation to tiles in proximity of the centres of gravity, where on average more energy is deposited. The data-driven simulation is conducted with Kernel Density Estimators on hit energy level. As Kernel Density Estimators face computational limitations, the entire number of calorimeter tiles has not been used in the simulation. Instead, certain tiles and layers have been excluded for the simulation, in order to reduce input data, while keeping the information loss as minimal as possible. To account for the missing energy, Gaussian noise is added to the missing tiles and layers afterwards. The performance of this method has been evaluated based on shower shape variables that characterise a hadron shower. The results showed good agreement between the distributions of the full datasets and simulation.

Zusammenfassung

Diese Arbeit untersucht einen datenbasierten Algorithmus für eine schnelle Simulation von Hadronenschauern in einem hochgranularen Kalorimeter. Die Daten wurden während einer Teststrahlkampagne am CERN im Jahr 2018 mit dem AHCAL der CALICE Kollaboration aufgenommen. Während diese Kampagne wurde der AHCAL Elektronen-, Myonen- und Pionenstrahlen mit verschiedenen Energien ausgesetzt. Für diese Arbeit wurden die Pionenstrahldaten genutzt.

Der Ansatz basiert auf der Sortierung der Kalorimeterzellen in aufsteigender Reihenfolge nach ihrer Entfernung zum Schauerschwerpunkt. Dadurch erhalten Kacheln in unmittelbarer Nähe des Schwerpunkts ein höheres Gewicht, da hier im Mittel mehr Energie deponiert wird. Die datenbasierte Simulation erfolgt mithilfe von Kernelichteschätzern auf Kachelniveau. Aufgrund von technischen Einschränkungen der Kernelichteschätzer konnten jedoch nicht alle Kalorimeterkacheln in der Simulation betrachtet werden. Daher wurden in der Simulation verschiedene Kacheln und Detektorschichten ausgeschlossen, was zu einem geringeren Energiegehalt in der Simulation führt. Um die fehlende Energie auszugleichen, wird für die nicht berücksichtigten Kacheln und Schichten zusätzlich gaußsches Rauschen hinzugefügt.

Die Leistungsfähigkeit dieser Methode wurde anhand von Schauerformvariablen bewertet, die die Eigenschaften von Hadronenschauern charakterisieren. Die Ergebnisse zeigen gute Übereinstimmung zwischen den Verteilungen der Datensätze und der Simulation.

Contents

1. Introduction	1
2. Standard Model of Particle Physics	3
3. Energy Loss of Charged Particles in Matter	7
4. Particle Showers in Calorimeters	11
4.1. Electromagnetic (EM) Showers	11
4.1.1. Physics of Electromagnetic Showers	11
4.1.2. Longitudinal and Radial Shower Profiles of Electromagnetic Showers	13
4.2. Hadronic Showers	17
4.2.1. Physics of Hadronic Showers	17
4.2.2. Longitudinal and Radial Shower Profiles of Hadronic Showers . . .	21
4.3. Construction of Calorimeters	24
4.3.1. Homogeneous and Sampling Calorimeters	24
4.3.2. Energy Resolution of Calorimeters	25
4.3.3. Software Compensation for Calorimeters	25
5. CALICE Collaboration/DRD6 and the AHCAL Prototype	29
5.1. Possible High Energy Experiments in the Post-LHC Era	29
5.2. The CALICE and DRD6 Collaborations	31
5.3. The AHCAL Prototype	31
5.3.1. Silicon-Photomultipliers (SiPMs)	33
5.3.2. Test Beam Campaign	33
6. Fast Simulation for Hadronic Showers	37
6.1. Kernel Density Estimators (KDEs)	37
6.2. Alternative Fast Simulation Algorithms for Particle Showers	42
7. Cell-Ordering Algorithm	45
7.1. Event Selection	45

Contents

7.2. Cell-Ordering based on Centres of Gravity	46
7.3. Simulation of the Centres of Gravity with KDEs	48
7.4. Simulation of Energies with KDEs	49
7.4.1. Energy Distribution per Tile	50
7.4.2. Longitudinal Energy Deposition	52
8. Shower Shape Variables	55
8.1. Definition of Shower Shape Variables	55
8.2. Gaussian Noise for Neglected Calorimeter Regions	56
8.3. Distributions of Shower Shape Variables	59
9. Results at Different Energies	69
10. Conclusion and Outlook	79
A. Centres of Gravity	83
A.1. Centres of Gravity - 3D	83
A.2. Distributions of the Centres of Gravity	86
B. Shower Shape Variables	91
B.1. Other Compositions - 40 GeV Pions	91
B.2. Other Variables for Other Pion Energies	98

1. Introduction

The Standard Model (SM) of particle physics [1-4] serves as the foundational framework for understanding the universe, providing insight into the physics of fundamental particles and forces of nature. For centuries, the quest to uncover the underlying structure of the universe has been a central driving force in physics, leading to the formulation of the SM, which remains the most successful and comprehensive theory in the field until this day.

Major research facilities worldwide, including CERN and its Large Hadron Collider (LHC) [5] in Geneva, and the development of sophisticated detectors, and numerous experimental discoveries have greatly advanced our understanding of particle physics and the universe.

General-purpose detectors, such as the ATLAS [6] and CMS [7] detectors at the LHC, are built on the same principle: particles are accelerated and via their collisions, other particles are produced. Most of these decay before reaching the detector and the decay products are detected via the interaction with the experimental apparatus, where different components measure the properties of the particles, such as charge, momentum or energy. These are used to reconstruct the original particle. One of the most crucial components is the calorimeter, responsible for measuring the energy deposited by particles as they pass through its layers. Within calorimeters, high-energy particle cascades develop, known as showers, through the interaction with the detector material. Particle showers develop when a particle interacts with the detector material, deposits energy, and creates secondary particles, which in turn also deposit energy leading to the cascade of particles. This energy deposition can be read out via electrical signals, providing valuable information about the nature of the particle and its interactions.

Experimental research aims to put theoretical models to the test, and the SM has proven to be remarkably successful in explaining a wide range of physical phenomena. However, it also remains incomplete. Current research tries to extend the SM by searching for new particles and forces that may offer insights into unresolved questions, such as what dark matter is [8], why neutrinos have masses [9], and the unification of all fundamental forces, including gravity [10, 11]. To address these challenges, more advanced detectors and, consequently, more sophisticated calorimeters are required. The CALICE Collaboration,

1. Introduction

now integrated into the six taskforce of the Detector Research and Development (DRD6) Collaboration, is developing highly granular calorimeters. This thesis focuses on the AHCAL prototype [12], which is a highly granular calorimeter prototype developed by the CALICE collaboration.

In order to verify or falsify extensions to the SM, it is necessary to simulate the interactions of particles with the detector material. This allows the hypothesis to be accepted or rejected, thereby providing insight into physics beyond the SM. The simulation process of particle showers has the potential to require a significant amount of computational resources, which could make it challenging to manage. Thus, this Master's thesis focuses on a fast simulation algorithm for particle showers through a data-driven method based on Kernel Density Estimators (KDEs) [13].

Chapter 2 provides an introduction to the SM of particle physics, followed by Chapter 3, which briefly summarises the theory of energy loss of charged particles in matter. Chapter 4 offers a brief overview of the theory behind particle showers and the composition of calorimeters. Chapter 5 introduces the CALICE Collaboration and DRD6 and describes the AHCAL prototype, while also giving an outlook on future high-energy colliders. The fast simulation using KDEs is discussed in Chapter 6, while Chapter 7 presents the cell-ordering algorithm for the fast simulation based on centres of gravity (CoGs). The performance of this algorithm is determined by the distributions of different shower shape variables, as discussed in Chapter 8. Lastly, Chapter 9 extends the investigation to datasets of varying pion energies. Finally, a summary and outlook are given in Chapter 10.

2. Standard Model of Particle Physics

The SM of particle physics is a well-established theoretical framework that consists of the fundamental particles of nature and their interactions; it describes the elementary particles that constitute matter and the force-carrier particles that mediate their interactions. These are shown in Figure [2.1](#). Of all four fundamental forces, the strong, weak or electromagnetic forces are incorporated into the SM. The known particles are categorised into two classes based on their fundamental properties: fermions, which have half-integer spin ($1/2$), and bosons, which have integer spin (1 or 0). Fermions are the building blocks of matter and are further divided into two groups: quarks and leptons. Moreover, each particle has a corresponding anti-particle with opposite charge [\[14, 15\]](#).

The quarks are classified by their electric charge and arranged into three generations. The up, charm, and top quark each have an electric charge of $+2/3$ (in units of the elementary charge e), while the down, strange, and bottom quark have a charge of $-1/3$. Each generation contains one up-type and one down-type quark: the up and down quarks form the first generation, charm and strange form the second, and top and bottom complete the third. Of the three quark generations, only the first is stable. The heavier quarks in the second and third generations undergo decay, as transitioning into lighter quarks is energetically allowed [\[14, 15\]](#).

Leptons are similarly divided into three generations. The electron (e^-), muon (μ^-) and tau lepton (τ^-), each carry an electric charge of -1 . They are associated with neutrinos, which are electrically neutral. Each charged lepton forms a generation with its corresponding neutrino (ν_e , ν_μ and ν_τ), in a manner analogous to the quark sector [\[14, 15\]](#).

The force-carrying bosons (spin-1 particles) mediate three of the four known fundamental forces and the interactions between fermions are described by the exchange of gauge bosons. The electromagnetic force, which acts on all particles with electric charge, is mediated by the electrically neutral photon (γ) [\[14, 15\]](#).

The gluon (g) is the mediator of the strong nuclear force, binding quarks together within

2. Standard Model of Particle Physics

hadrons. Quarks are the only fermions that possess colour charge (red, green, or blue), and thus, can interact via the strong force. Anti-quarks carry anti-colour.

		three generations of matter (fermions)			interactions / forces (bosons)	
		I	II	III		
QUARKS	mass	$\simeq 2.16 \text{ MeV}$	$\simeq 1.27 \text{ GeV}$	$\simeq 172.57 \text{ GeV}$	0	$\simeq 125.20 \text{ GeV}$
	charge	$+2/3$	$+2/3$	$+2/3$	0	0
	spin	$1/2$	$1/2$	$1/2$	1	0
		u up	c charm	t top	g gluon	H Higgs
		$\simeq 4.70 \text{ MeV}$	$\simeq 93.5 \text{ MeV}$	$\simeq 4.18 \text{ GeV}$	0	
		$-1/3$	$-1/3$	$-1/3$	0	
		$1/2$	$1/2$	$1/2$	1	
		d down	s strange	b bottom	γ photon	
		$\simeq 0.511 \text{ MeV}$	$\simeq 105.66 \text{ MeV}$	$\simeq 1.78 \text{ GeV}$	$\simeq 80.37 \text{ GeV}$	
		-1	-1	-1	± 1	
		$1/2$	$1/2$	$1/2$	1	
		e electron	μ muon	τ tau	W W boson	
LEPTONS		$< 1.0 \text{ eV}$	$< 0.17 \text{ eV}$	$< 18.2 \text{ MeV}$	$\simeq 91.19 \text{ GeV}$	
		0	0	0	0	
		$1/2$	$1/2$	$1/2$	1	
		ν_e electron neutrino	ν_μ muon neutrino	ν_τ tau neutrino	Z Z boson	

Figure 2.1.: The particles of the SM of particle physics, taken from Ref. [16] and modified. The mass, electric charge, and spin values are taken from Ref. [17].

Quarks combine in groups to form particles called hadrons. Hadrons are further classified into baryons, such as protons and neutrons, which are made of three quarks, and mesons, which consist of a quark-antiquark pair. Free quarks have never been observed, a phenomenon explained by colour confinement, which states that particles with a non-zero colour charge cannot exist independently. Instead, quarks are always confined to colour-neutral states. Hadrons are held together by gluon exchange between quarks. Gluons are also carriers of colour charge, enabling them to interact with each other through the strong force. This interaction creates a confined colour field between the quarks, forming a tube-like structure. Unlike the electromagnetic field, which spreads out in space, the colour field remains concentrated. This confinement results in a potential that scales linearly with the distance between the quarks, preventing their separation and ensuring they only exist in bound states. Hadronisation then occurs when quarks are pulled apart,

and the energy stored in the colour field between them becomes large enough to create new quark-antiquark pairs. These new quarks combine with the original quarks to form colour-neutral bound states known as hadrons. This process leads to the production of multiple hadrons, which manifest as collections of particles called jets that can be detected in experiments [18, 19].

All elementary particles in the SM possess the weak charge, allowing them to interact through the weak force, which is mediated by the W^\pm and Z bosons. The W^\pm is the mediator for flavour changing charged processes. Transitions between up- and down-type quarks and vice versa are therefore allowed as long as a W^\pm is exchanged between them. However, the transition between quark generations is suppressed and transitions within one generation are more favoured. This mechanism is described by the Cabibbo-Mechanism [20] and the Cabibbo-Kobayashi-Maskawa (CKM) [21] matrix. Furthermore, flavour-changing neutral currents are not a feature of the SM, at leading order. Consequently, flavour change involving the Z boson is prohibited, at leading order. In addition, due to the conservation of lepton number per generation in the SM, only leptons within the same generation couple to each other via the emission or absorption of a W^\pm [14, 15, 22].

The final piece of the SM was the discovery of the Higgs boson (spin-0 particle) in 2012 by the ATLAS and CMS Collaborations [23, 24], a particle whose existence had already been predicted in 1964 [25-27]. The Higgs boson is central to the Higgs mechanism, which explains how particles acquire mass. The Higgs potential,

$$V(\phi) = \mu^2(\phi^\dagger\phi) + \lambda(\phi^\dagger\phi)^2, \quad (2.1)$$

with $\mu^2 < 0$, results in a degenerate vacuum state with a non-zero vacuum expectation value [25-28]. This leads to *spontaneous symmetry breaking*, and in combination with the unification of the electromagnetic and weak force, particles interact with the Higgs field and acquire mass. Without this mechanism, particles would remain massless, resolving a key discrepancy between theoretical predictions and observed non-zero particle masses [25-28].

In quantum field theory, the underlying gauge symmetry of the SM is described by three groups:

$$\text{SU}(3)_C \times \text{SU}(2)_L \times \text{U}(1)_Y, \quad (2.2)$$

where the $\text{SU}(3)$ group governs the strong interaction between quarks via gluons and $\text{SU}(2)_L \times \text{U}(1)_Y$ governs the electroweak interaction - the unification of the weak interaction with the $\text{SU}(2)_L$ and the $\text{U}(1)_Y$ group that accounts for the electromagnetic interaction. The indices represent the charges and symmetries associated with each group

2. *Standard Model of Particle Physics*

[1-4].

The SM provides a comprehensive description of the current state of knowledge in the field of particle physics. However, it is important to acknowledge that the SM is not capable of addressing all research questions, such as the nature of dark matter [8] and the observed imbalance between matter and antimatter in the universe [29, 30]. Current research efforts focus on developing novel theories or extending the SM to resolve these issues.

3. Energy Loss of Charged Particles in Matter

The detection of particles traversing a medium relies on their interactions with the material, where the probability of an interaction occurring is quantified by the cross-section σ , representing the effective area of a target atom or particle that a traversing particle encounters during propagation. The utilisation of diverse processes between the particle and the medium facilitates the identification of the specific particle type and its associated properties [31, 32].

When an electrically charged particle traverses a medium and interacts with it, the material atoms can be ionised or excited. The interaction between the traversing charged particle and the electrons and nucleus of the atoms leads to a signal that can be measured and used for further analysis. During the ionisation and excitation processes, energy is lost, leaving the charged particle with a reduced energy level after the interaction. If the particle is heavier than an electron (or positron), the average energy loss per unit length is characterised via the Bethe-Bloch equation [31, 32]

$$-\left\langle \frac{dE}{dx} \right\rangle = K \frac{Z}{A} \rho \frac{z^2}{\beta^2} \left[\frac{1}{2} \ln \left(\frac{2m_e c^2 \beta^2 \gamma^2 \epsilon_{\max}}{I^2} \right) - \beta^2 - \frac{\delta(\beta\gamma)}{2} \right], \quad (3.1)$$

which was proposed by Hans Bethe in 1930 [33, 34] and Felix Bloch in 1933 [35, 36]. The individual components are as follows [31]:

- $K = 4\pi N_A r_e^2 m_e c^2 = 0.307 \text{ MeV cm}^2 \text{ mol}^{-1}$ is a constant. N_A is the Avogadro constant, r_e the classical electron radius, m_e the electron mass and c the speed of light in vacuum.
- z is the electric charge of the propagating particle and $\beta = v/c$ its relativistic velocity.
- Z is the atomic number, A the atomic mass, and ρ is the density of the medium.
- I is the average energy that is needed to ionise the medium.

3. Energy Loss of Charged Particles in Matter

- $\epsilon_{\max} = 2m_e c^2 \beta^2 \gamma^2 / (1 + 2\gamma m_e / M + (m_e / M)^2)$ is defined as the maximal energy that is deposited via a central collision between the particle and a shell electron.
- $\delta(\beta\gamma)$ describes a density correction that appears for high energies.

The equation above describes the stopping power of a medium on particles traversing it and is shown in Figure 3.1. Below $\beta\gamma \approx 0.05$, the effects of shell corrections must be

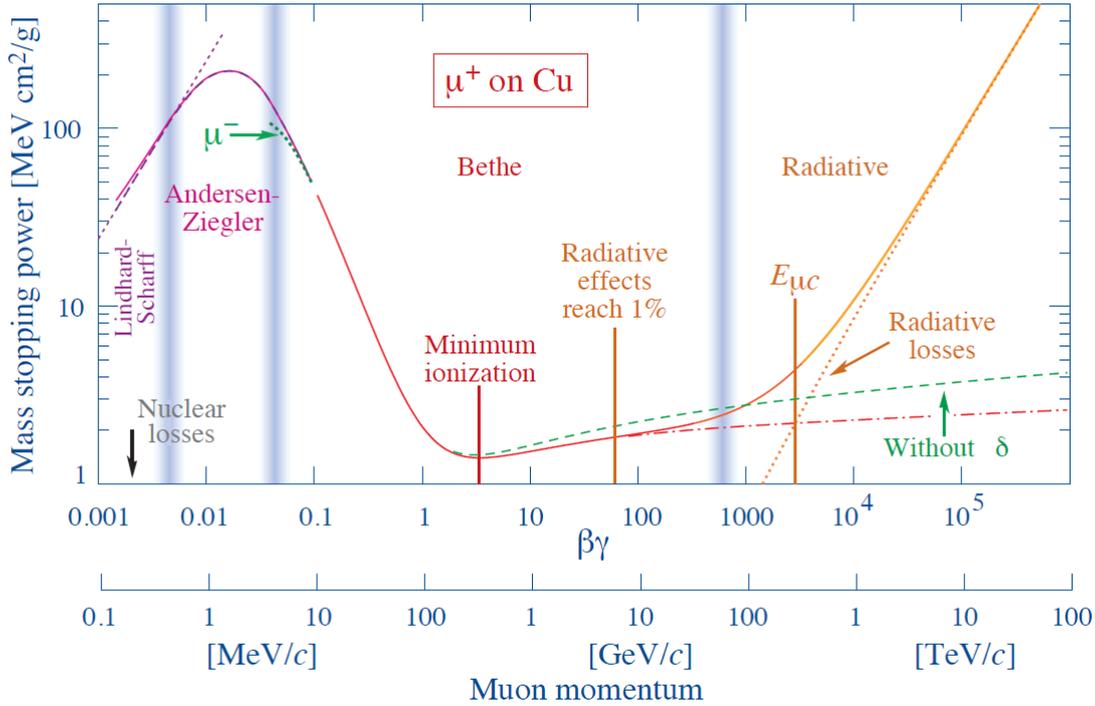


Figure 3.1.: The Bethe-Bloch equation (3.1) for μ^+ in copper as a function of $\beta\gamma$ [17].

taken into account, resulting in a maximum in the energy loss distribution. Subsequently after the maximum, the $1/\beta^2$ -term of Equation (3.1) dominates, resulting in a decrease of energy deposition until a minimum is reached at intermediate energies around $\beta\gamma \approx 3-4$. A particle falling in this region is described as a minimum ionising particle (MIP), as the energy loss of the particle is reduced to a minimum. In the case of particles with energies exceeding the minimum, the energy deposition increases in accordance with the logarithmic term in the Bethe-Bloch equation until the value of $\beta\gamma$ approaches approximately 1000. Thereafter, radiation effects begin to dominate. The density correction $\delta(\beta\gamma)$ of the Bethe-Bloch equation dominates at larger energies and limits the energy loss of particles. This correction originates from the shielding of the transversal component of the electrical field of the propagating particle. As the particle's energy increases, its electric field flattens and extends, leading to the polarisation of the medium. This results

in the subsequent limitation of the electric field extension. Consequently, the impact of the logarithmic-term is diminished, leading to a reduction in energy loss [31, 32].

Electrons and positrons must be considered individually due to their differences to heavier particles in interactions with a medium, and are discussed in Chapter 4.1.

4. Particle Showers in Calorimeters

The various components of particle detectors exploit a range of particle properties to facilitate detection. To quantify the energy deposition of particles, calorimeters are utilised to investigate so-called particle showers. The theory of shower processes and development distinguishes between electromagnetic (EM) and hadronic showers due to the different nature of electromagnetically and hadronically interacting particles with material. Chapter [4.1](#) explores the characteristics of EM showers, while Chapter [4.2](#) focuses on the properties of hadronic showers.

4.1. Electromagnetic (EM) Showers

4.1.1. Physics of Electromagnetic Showers

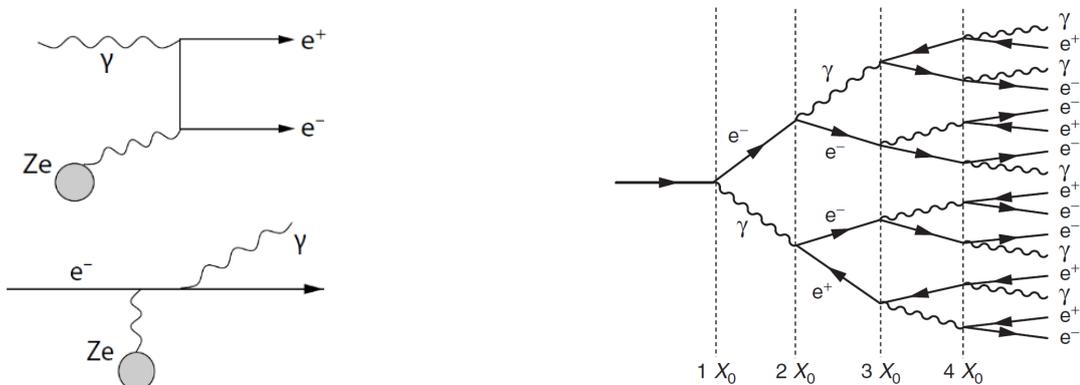
EM showers are created via the interaction of high-energetic electrons, positrons and photons with matter. These particles deposit their energy within the material, resulting in the loss of energy of the initial particle. The processes that drive EM shower creation at high energies at leading order are: pair production and Bremsstrahlung. Pair production is conversion of a photon into an electron-positron pair. Bremsstrahlung, on the other hand, is the emission of photons in the field of a nucleus by charged leptons. These processes are depicted in Figure [4.1\(a\)](#), and it can be assumed that energy loss per unit length via ionisation is negligible as long as the energy per particle is above the critical energy, E_{crit} . The energy loss associated with bremsstrahlung is directly proportional to the inverse square of the particle's mass and scales linearly with the energy,

$$\left(\frac{dE}{dx}\right)_{\text{Bremsstrahlung}} \propto Z^2 \frac{E}{M^2}, \quad (4.1)$$

with Z being the material's atomic number and M the incoming particle's mass [\[31\]](#). Therefore, the energy loss due to the bremsstrahlung process depends heavily on the material, the particle's energy and mass. As the mass of the particle decreases, the energy loss increases. Additionally, greater energy deposition occurs in denser materials.

4. Particle Showers in Calorimeters

Moreover, the energy loss resulting from multi-scattering processes, i.e. Bhabha- [37] and Møller-scattering [38], is disregarded in this simplified model [15, 31, 39].



(a) The processes pair production (top) and bremsstrahlung (bottom) [31].

(b) The simplified model of the development of an EM shower [14].

Figure 4.1.: (a) The two processes that mainly drive EM creation in the simplified model of an EM shower, and (b) a sketch of an EM shower in the calorimeter.

The combination of pair production and bremsstrahlung gives rise to EM showers, leaving a cascade in the calorimeter as shown in Figure 4.1(b). Bremsstrahlung dominates in the high-energy-regime, with pair production also being dominant in the high MeV- and GeV-region. The characteristic length of a shower is the radiation length X_0 , which is inversely proportional to the square of the atomic number of the material (Ref. [31]):

$$X_0 \propto \frac{1}{Z^2}. \quad (4.2)$$

Additionally, the initial energy E_0 of the particle initiating the EM shower plays a crucial role in describing its development. Once a particle has traversed a distance x , the remaining energy of the particle is given by

$$E(x) = E_0 \cdot e^{-\frac{x}{X_0}}, \quad (4.3)$$

with E_0 being the initial energy, and the radiation length X_0 being the mean free path of a particle in an EM calorimeter [31]. Thus, on average, a particle only possesses $1/e$ of the initial energy after a path length of $x = X_0$. The number of particles, N , doubles approximately after every radiation length and the total number of particles in the shower after traversing $n \in \mathbb{N}$ radiation lengths is given by (Ref. [31])

$$N \approx 2^n. \quad (4.4)$$

The mean energy of a particle after n radiation lengths can therefore be determined to be (Ref. [31])

$$\langle E \rangle = \frac{E_0}{2^n}. \quad (4.5)$$

The development of a EM shower stops when the particle's energy has fallen below a threshold of

$$E_{\text{crit}} = \frac{E_0}{2^{n_{\text{total}}}} \quad (4.6)$$

after n_{total} radiation lengths [31]. Upon reaching this threshold, ionisation becomes the dominant mechanism for energy loss. The total number of particles after n_{total} radiation lengths is then given by the fraction of the initial energy to critical energy (Ref. [31]):

$$n_{\text{total}} = \frac{\ln(E_0/E_{\text{crit}})}{\ln(2)}. \quad (4.7)$$

Hence, the total length s_{total} of an EM shower can be determined to be

$$s_{\text{total}} = n_{\text{total}} \cdot X_0 = \frac{\ln(E_0/E_{\text{crit}})}{\ln(2)} \cdot X_0, \quad (4.8)$$

which states that the longitudinal size of an EM shower increases logarithmically with the initial energy, leading to only a logarithmic increase in size of an EM calorimeter as the energy increases [31]. It can be concluded that the use of a thick EM calorimeter is not necessary at high particle energies.

4.1.2. Longitudinal and Radial Shower Profiles of Electromagnetic Showers

The characteristics of an EM shower are described by its longitudinal and radial shower profiles. The longitudinal shower profile is defined as the energy deposition along the shower axis. It is characterised by the gradual build-up of energy deposition until it has reached its maximum in the calorimeter. Thereafter, the energy deposition falls quickly in the last regions of the calorimeter. The radial shower shape of EM showers describes the lateral spread of energy loss of electrons, positrons and photons in the calorimeter. Thus, the study of EM shower development in the longitudinal and radial direction is crucial for understanding their formation and development.

Longitudinal Shower Profile

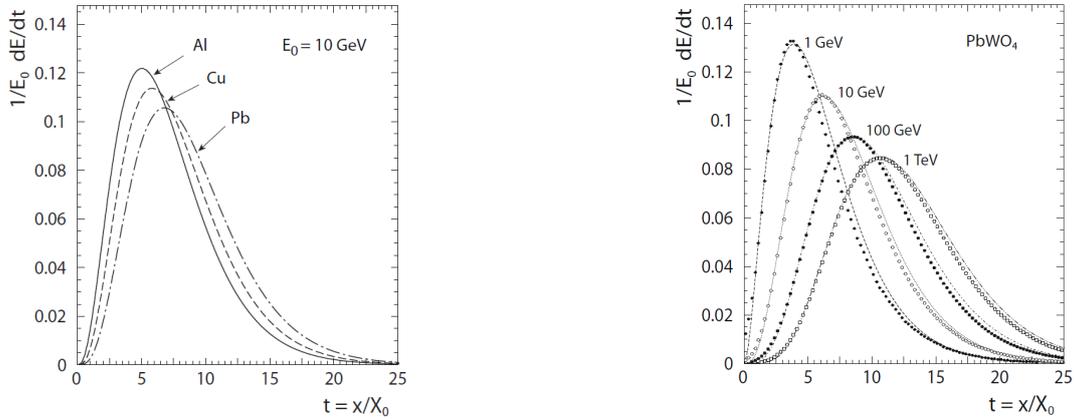
The longitudinal shower development of an EM shower can be described via

$$\frac{dE}{dn} = E_0 \frac{b^a}{\Gamma(a)} n^{a-1} e^{-bn}, \quad (4.9)$$

which was proposed by Longo and Sestili in 1975 [40], with the parameters a and b , the initial energy E_0 and the material's atomic number Z . $\Gamma(a) = \int_0^{\infty} t^{a-1} e^{-t} dt$ [41] describes the gamma function. The maximum of this function lies at (Ref. [31])

$$n_{\max} = \frac{a - 1}{b}. \quad (4.10)$$

Longitudinal shower profiles for different materials are shown in Figure 4.2(a), with Equation (4.9) fitted to the simulation. The peaks of the simulated longitudinal energy distributions are observed at greater penetration depths for materials with larger atomic numbers. Furthermore, as indicated by Equation (4.2), the radiation length X_0 is smaller for denser materials. In addition, differences in the critical energy of the shown materials lead to different shower developments.



(a) The longitudinal energy distribution for EM showers for different materials [31]. The showers are calculated from 10 GeV electrons.

(b) The longitudinal energy distribution for EM showers for different energies, detected in PbWO₄ [31].

Figure 4.2.: The longitudinal energy distributions of EM showers. Here, t denotes the number of radiation lengths.

Table 4.1 lists the parameters for the development of EM showers for four materials. The larger the atomic number of the material is, the smaller the radiation length X_0 and critical energy E_{crit} of a particle traversing the medium. The parameter b from Equation

(4.10) decreases likewise in denser material and is responsible for the decrease of the energy deposition after the maximum energy deposition is reached. Furthermore, as the material density increases, the maximum energy deposition occurs deeper within the calorimeter, here shown for 100 GeV, and it goes along with a decreasing value for the parameter b as n_{\max} increases. Table 4.1 also shows the number of radiation lengths $n^{98\%}$ for each material, in which 98% of the initial energy is deposited [31]. Also, a more energetic beam leads to more energy deposited in the outer region of the calorimeter, as shown in Figure 4.2(b).

Table 4.1.: Properties of EM showers in different materials [31].

Material	Z	X_0 (mm)	E_{crit} (MeV)	b	n_{\max} (100 GeV)	$n^{98\%}$ (100 GeV)
H ₂ O	1, 8	361	78.6	0.63	6.6	17.3
Al	13	89	42.7	0.58	7.3	18.8
Fe	26	17.6	21.7	0.53	7.9	20.6
Pb	82	5.6	7.4	0.50	9.0	22.7

Radial Shower Profile

The particles produced via the two dominant processes in the calorimeter are mostly emitted in forward direction for high energies. The angle θ at which the particles are emitted increases inversely with energy (Ref. [31]):

$$\theta \propto \frac{1}{E}. \quad (4.11)$$

Therefore, the lateral expansion of the shower is mainly dominated by Compton scattering of low-energy photons and multiple scattering of low-energy electrons/positrons. On average, approximately 90% of the energy is deposited in a cone with radius

$$\rho_M = \frac{21.2 \text{ MeV}}{E_{\text{crit}}} \cdot X_0, \quad (4.12)$$

also known as Molière radius, which depends on the critical energy and the radiation length of the material [31]. Approximately 90% of the energy is deposited within a cylinder with a radius of one Molière radius ρ_M , while 95% of the deposited energy is deposited in a cylinder of twice the radius.

The combination of Equations (4.2) and (4.12) results in an increase in the ratio R_M/X_0 with the atomic number of the detector material, and thus, resulting in narrower EM showers in low- Z materials than for materials with a large atomic number. However, the Molière radius itself becomes smaller for heavy detector materials. Table 4.2 shows the

4. Particle Showers in Calorimeters

Molière radius and the ratio R_M/X_0 for four different materials [31]. Figure 4.3 shows the radial shower shape for a brass-fiber dual-readout calorimeter, measured separately with Čerenkov and the scintillation signals. From Figure 4.3(a), it can be determined that the farther one is from the shower axis, the less energy will be deposited. Consequently, the majority of the energy will be deposited in the vicinity of the shower axis, with the outer regions of the calorimeter layers exhibiting almost no deposited energy. Figure 4.3(b) illustrates the percentage of the deposited energy as a function of the cylinder radius. The cylinder is drawn around the shower axis. It can be observed that, as the radius increases, the energy deposited in this region also increases. However, the gradients of the curves decrease, also indicating a reduction in energy deposition for larger cylinder radii.

Table 4.2.: The Molière radius and the ratio R_M/X_0 for four different materials [31].

Material	Z	R_M (mm)	$\frac{R_M}{X_0}$
H ₂ O	1, 8	83	0.23
Al	13	45	0.51
Fe	26	18	1.02
Pb	82	16	2.86

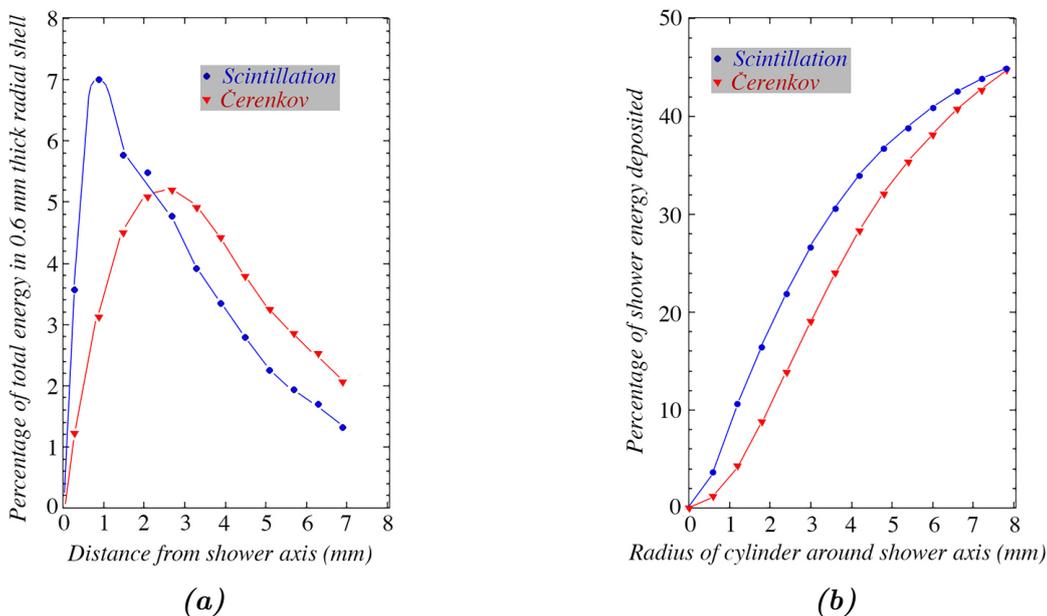


Figure 4.3.: The radial development of EM showers in a brass-fiber dual-readout calorimeter, measured separately with the Čerenkov and the scintillation signals [42]. Deposited energy as a function of the (a) distance to the shower axis in mm and (b) radius of the cylinder around the shower axis in mm.

4.2. Hadronic Showers

4.2.1. Physics of Hadronic Showers

Particle showers generated by hadrons differ from EM showers in that the underlying processes are more complex. The two processes that drive the expansion of hadronic showers are shown in Figure 4.4.

A high-energetic incoming hadron initially undergoes an inelastic interaction with a nucleus, resulting in the generation of high-energetic particles within the nucleus. An intra-nuclear cascade of particles, such as neutral and charged pions, kaons, protons, and other hadrons, occurs. Subsequently, these particles are able to escape the nucleus if they possess sufficient energy. Thereafter, they undergo inelastic scattering with other nuclei in a similar manner, which results in the formation of further secondary particles. This process gives rise to the formation of a cascade of particles within the hadronic calorimeter. In these interactions, only one nucleon within the nucleus is involved, while the remaining nucleons remain unaffected. As a consequence, the nucleus will become excited. The process known as **spallation** involves the emission of energy through the release of highly energetic fragments, including protons, neutrons, kaons, and pions. Neutral pions decay into two photons after $\sim 10^{-16}$ s, which triggers EM showers within the larger hadronic shower. Spallation begins approximately 10^{-22} s seconds after the initial collision and continues until the threshold for inelastic scattering is reached or the particles produced exit the nucleus, which leaves the nucleus in an excited state. The emitted spallation particles typically have kinetic energies around 100 MeV. Additionally, hadrons produced within the nucleus may scatter off spectator nucleons, further exciting the nucleus to an even higher energetic state [31, 32, 39].

Approximately 10^{-18} seconds after spallation, the excited nucleus releases its energy by emitting low-energy nucleons and fragments, such as protons, neutrons, α -particles, and photons. This phenomenon is known as **evaporation**. Additionally, the nucleus may also undergo nuclear fission, which leads to the formation of two or more lighter fragments, often with additional neutron or photon emission. The energy spectrum of particles generated during evaporation is significantly lower than that of the spallation process, typically falling within the MeV range [31].

A part of the energy is absorbed and required for breaking up nuclei. This is referred to as *invisible energy*, as this type of energy is not available for other processes and not assigned to the energy of the shower. It is estimated that up to 40% of total non-EM energy remains undetected. Furthermore, the neutrinos in these processes escape detection, and consequently, their energy is not measured either [31, 32, 39].

with the signals $S(e)$ and $S(\pi)$ for the EM and hadronic term and ϵ_{EM} and ϵ_{had} for the respective signal efficiencies [31]. Figure 4.5(a) shows that the signal of the hadronic component of the shower varies significantly, spreading over a wide range, while the EM component features a distinct peak. The assumption of equal energy for electrons and pions is used to compare the calorimeter response for electrons and pions as Equation (4.15) corresponds to the ratio of response of a calorimeter to EM showers to the hadronic part, also written as e/π . The fraction $\epsilon_{EM}/\epsilon_{had}$ in the equation above, denoted as e/h , describes the ratio of the calorimeter's EM to hadronic efficiency. In general, a calorimeter's response is characterised by its efficiency in converting deposited energy into a measurable signal [31, 32, 43].

Typically, the response to hadron showers is lower than that for EM showers, meaning $e/h > 1$. This discrepancy arises primarily from significant fluctuations in energy deposition and mainly from the presence of invisible energy in hadronic showers. The entire energy of photons emitted from π^0 -decay is deposited and is measurable, whereas a substantial fraction of the total energy from the non-EM shower remains undetected. Calorimeters with such ratios are called non-compensating calorimeters. On the other hand, if $e/h = 1$, the calorimeter is referred to as compensating [31, 32, 43].

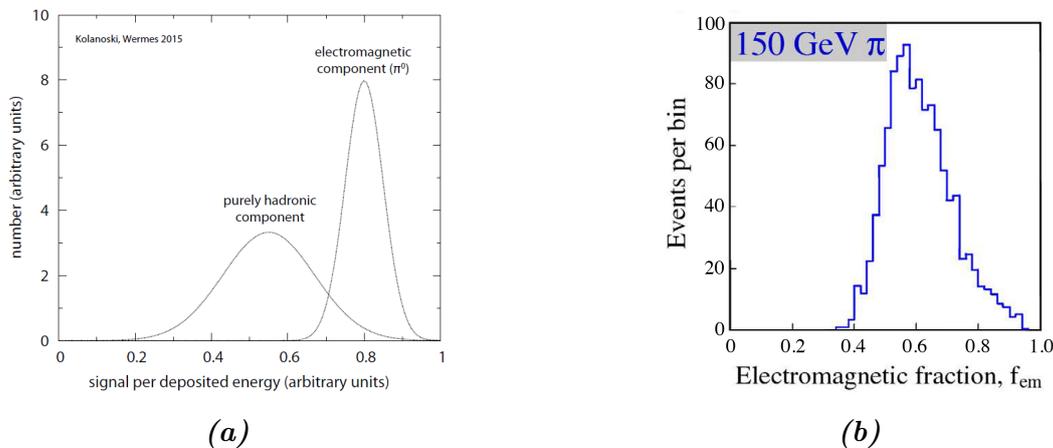


Figure 4.5.: (a) The calorimeter response for the EM component in hadronic showers (mainly driven by neutral pion decay) and the pure hadronic shower from Equation (4.14) and (4.15) [31]. (b) The EM fraction f_{EM} generated by 150 GeV pions, measured using a lead-based calorimeter [44].

Equation (4.15) highlights the critical role of f_{EM} in determining the behaviour of hadronic signals. Since a hadronic shower can produce a countless number of neutral pions as long as energy is conserved, the value of f_{EM} is not fixed and can vary continuously between zero and one, and since the EM fraction f_{EM} depends on the energy, the value of e/π too. The typical distribution of the electromagnetic signal fraction π is shown in Figure

4. Particle Showers in Calorimeters

4.5(b). The measurement was performed using 150 GeV pions in a lead (Pb)-calorimeter. The f_{EM} distribution lacks a distinct, well-defined peak and is instead widely spread out. This broad distribution leads to unbounded behaviour in f_{EM} , causing fluctuations in the signal ratio, which leads to $e/\pi > 1$ [31, 32, 43, 44].

However, an average value of f_{EM} can be calculated. The fraction f_{EM} rises in accordance with the photon multiplicity:

$$\langle f_{EM} \rangle \approx 1 - \left(\frac{E}{E_0} \right)^{k-1}, \quad (4.16)$$

with the energy E_0 representing the average energy required to produce a hadron, which is material-dependent [31]. The parameter k accounts for the particle multiplicity and is approximated to $k \approx 0.82$. Therefore, for larger energies E , the average EM fraction in hadronic showers increases [31]. Figure 4.6(a) illustrates the development of $\langle f_{EM} \rangle$ as a function of pion energy in two different calorimeters. The increase of the value of $\langle f_{EM} \rangle$ with the pion energy is clearly observable [31, 44]. Figure 4.6(b) shows the calorimeter response e/π as a function of the energy and e/h with $E_0 = 1$ GeV and $k = 0.82$. It is clearly visible that, at larger energies, the response e/π approaches unity asymptotically and is independent of the energy and the ratio of response between EM showers and non-EM showers due to the increase of the EM fraction in a hadronic shower [31, 43].

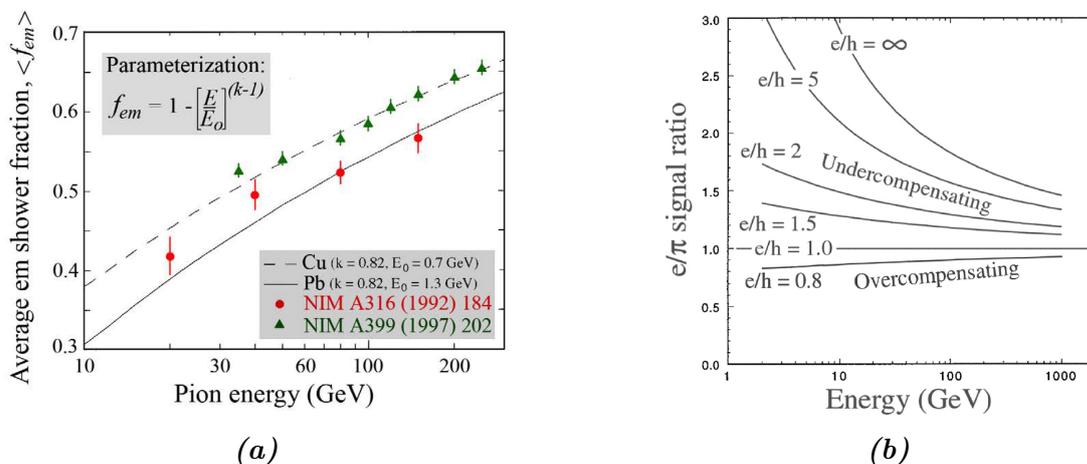


Figure 4.6.: (a) The average EM fraction $\langle f_{EM} \rangle$ as a function of the pion energy using a copper- and lead-calorimeter. The data is fitted with Equation (4.16) [44]. (b) The ratio of calorimeter response e/π as a function of energy and the response-ratio e/h for $E = 1$ GeV and $k = 0.82$ [43]. With increasing energy, e/π approaches one.

4.2.2. Longitudinal and Radial Shower Profiles of Hadronic Showers

Similar to EM showers, the size of hadronic showers is determined by their longitudinal and lateral extent. As mentioned in the previous chapter, EM showers develop within hadronic showers as a result of neutral pion decay into two photons. Therefore, the contribution of EM showers must be carefully considered when analysing the overall hadronic shower development.

Longitudinal Shower Profile

The characteristic length of hadronic showers can be described via one parameter, the nuclear absorption length

$$\lambda_a \approx 35 \frac{g}{\text{cm}^2} \frac{A^{\frac{1}{3}}}{\rho}, \quad (4.17)$$

where A is the atomic weight of the detector material and ρ its density [31]. The remaining energy $E(x)$ of a particle after traversing a distance x can be described analogously to Equation (4.3), with the radiation length X_0 for an EM shower replaced by the nuclear absorption length λ_a of a hadronic shower.

Table 4.3 shows the nuclear absorption length and radiation length for some materials, along with the material's properties. The ratio of nuclear absorption length to radiation length increases with larger atomic weight and density and is approximately proportional to the atomic number of the material (Ref. [31]):

$$\frac{\lambda_a}{X_0} \approx 0.37Z. \quad (4.18)$$

This leads to the conclusion that hadronic showers are larger than EM showers, and thus, hadronic calorimeters are, on average, larger than EM calorimeters. Figure 4.7 compares the size of an EM and a hadronic shower in a simulation for a 100 GeV electron and pion in an iron block. The figure demonstrates the significant size difference between the two shower types. EM showers are drastically shorter than showers initiated by hadrons. The EM shower in this simulation only covers approximately one-third of the iron block, whereas energy is deposited even at the end of the iron block for the hadronic shower. A notable parallel between hadronic and EM showers is that their sizes increase with the logarithm of energy, as stated in Equation (4.8) [31].

4. Particle Showers in Calorimeters

Table 4.3.: Properties of materials for hadronic showers [31].

Material	Z	A (g/mol)	ρ (g/cm ³)	X_0 (cm)	λ_a (cm)	λ_a/X_0
H ₂ O	1, 8	18	1.00	36.1	83.3	2.3
Air	7, 8	14.3	1.205×10^{-3}	$3.0 \cdot 10^4$	7.5×10^4	2.5
Al	13	27	2.70	8.9	39.8	4.5
Fe	26	56	7.87	1.76	16.8	9.5
Cu	29	64	8.96	1.43	15.3	10.7
W	74	184	19.30	0.35	9.9	28.3
Pb	82	207	11.35	0.56	17.6	31.4
U	92	238	18.95	0.32	11.0	34.4

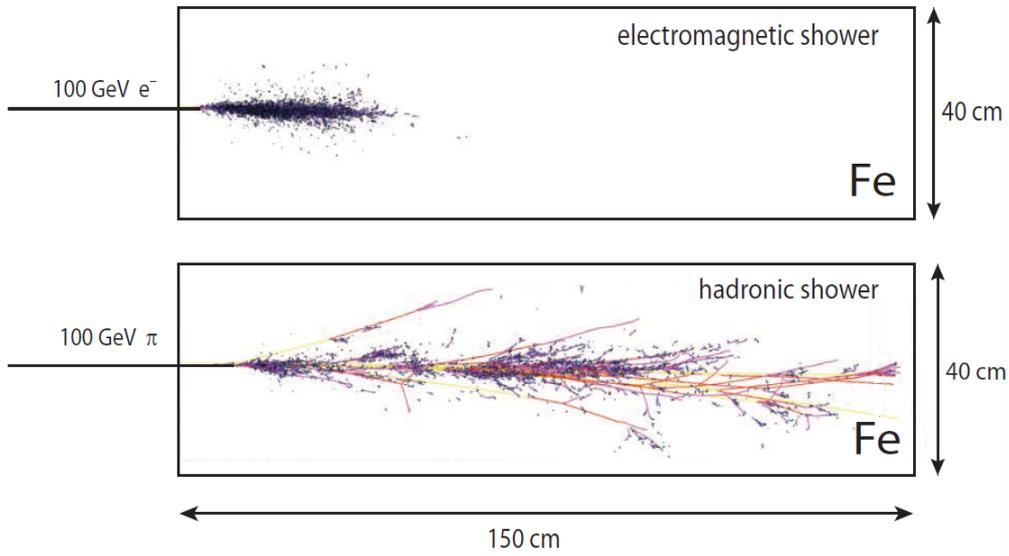


Figure 4.7.: The size of an EM and hadronic shower [31]. The simulation was conducted using a 100 GeV electron and pion in an iron block.

The change in energy along the shower axis can be parametrised as

$$\frac{dE}{dz} = E_0 \cdot \left[\frac{f_{\text{EM}}}{\Gamma(\alpha_s)} \cdot \left(\frac{z}{\beta_s} \right)^{\alpha_s-1} \cdot \frac{e^{-\frac{z}{\beta_s}}}{\beta_s} + \frac{f_{\text{had}}}{\Gamma(\alpha_l)} \cdot \left(\frac{z}{\beta_l} \right)^{\alpha_l-1} \cdot \frac{e^{-\frac{z}{\beta_l}}}{\beta_l} \right], \quad (4.19)$$

with the determination of the slope via α_i and β_i [45]. The first term accounts for the shorter EM shower, while the second term accounts for the larger hadronic shower. Figure 4.8 shows the longitudinal energy distribution of hadronic showers for data using 80 GeV pion and proton beams and the fitted Equation (4.19) with different values for the parameters. The two plots demonstrate that the primary energy loss occurs at approximately one nuclear absorption length. The tail of the distribution is dominated by the hadronic term in the energy distribution. In contrast, the EM sub-shower, being the shorter com-

ponent, demonstrates no deposited energy after approximately two nuclear absorption lengths. This observation clearly indicates the different sizes of the shower types.

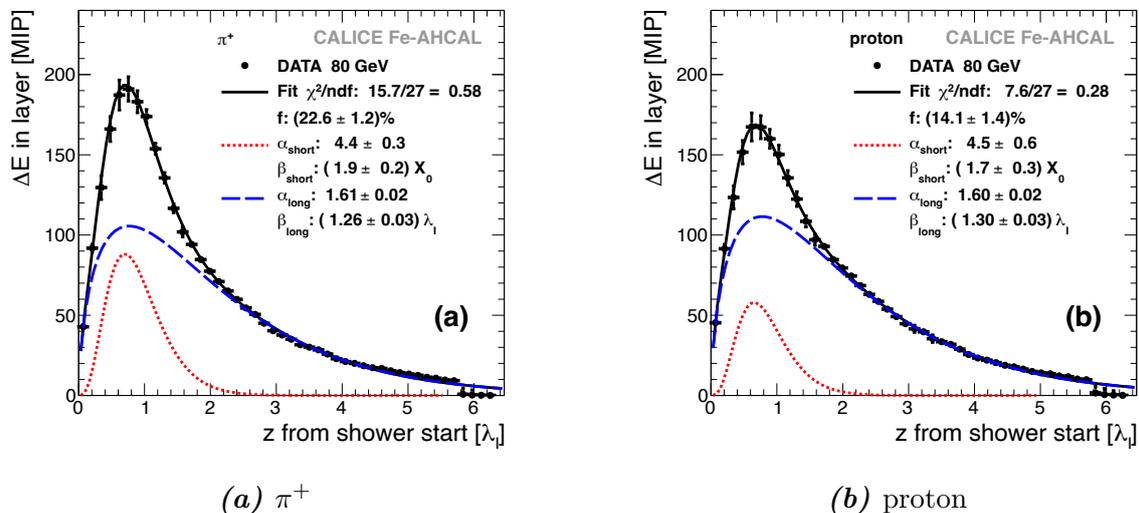


Figure 4.8.: Longitudinal energy distributions (measured in MIPs) of hadronic showers, described by Equation (4.19), shown as a function of shower depth (in units of nuclear absorption lengths, λ_l) [45]. The showers are initiated by 80 GeV (a) π^+ and (b) protons. Additionally, the fitted data points are shown, along with separate plots of the two components from Equation (4.19).

Radial Shower Profile

The nuclear absorption length is a measure not only of the longitudinal size of hadronic showers, but also for the radial width of the shower. Deposited energy spreads in two distinct patterns: a dense central region formed by electromagnetic sub-showers from neutral pion decays and a more diffuse outer region mainly influenced by neutrons. The majority, about 95 %, remains confined within a cylindrical volume with radius (Ref. [31])

$$R_{M|95\%} \approx \lambda_a. \quad (4.20)$$

Similar to Equation (4.19), the radial parameterisation of the energy can be divided into a core and halo part (Ref. [45]):

$$\frac{\Delta E}{\Delta S} = A_{\text{core}} \exp\left(-\frac{r}{\beta_{\text{core}}}\right) + A_{\text{halo}} \exp\left(-\frac{r}{\beta_{\text{halo}}}\right). \quad (4.21)$$

It describes the energy density as a function of the radius r measured from the shower axis. ΔS represents the area of a ring with radius r and width Δr around the shower

4. Particle Showers in Calorimeters

axis. The coefficients A_{core} and A_{halo} are scaling factors of the EM and non-EM part of an entire hadronic shower, while the parameters β_{core} and β_{halo} determine the slope of the function. Equation (4.21) is shown in Figure 4.9. The data has been obtained from 30 GeV pions and protons and has been fitted according Equation (4.21). The first term in the equation describes the EM subshower, while the second term represents the hadron-induced shower. The EM subshower curve declines more steeply compared to the hadronic shower, suggesting that EM subshowers are more confined radially than hadronic showers, which can also be seen in Figure 4.7 [45].

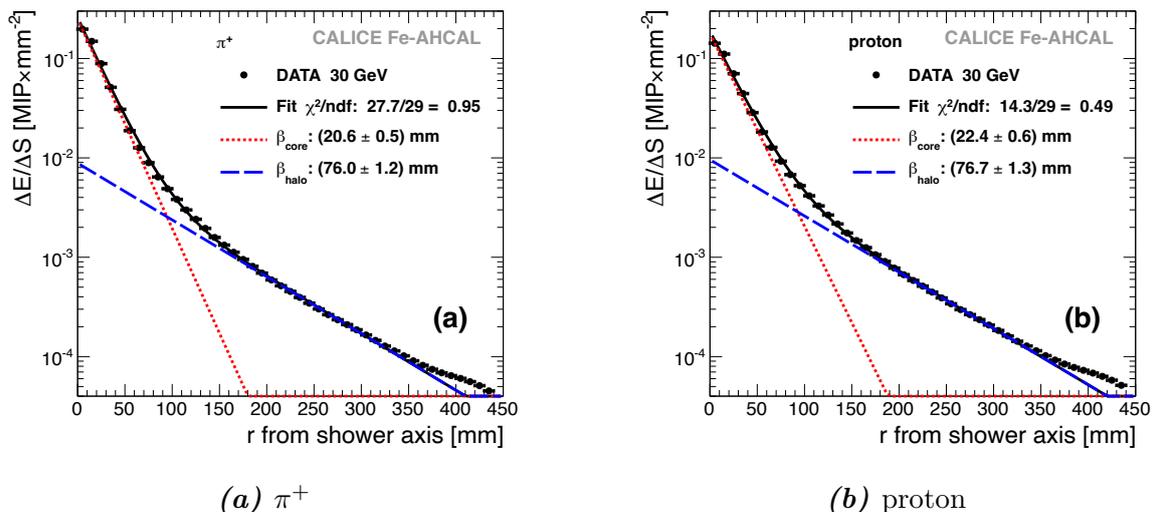


Figure 4.9.: Radial energy density distributions (measured in MIP mm^{-2}) of hadronic showers, described by Equation (4.21), shown as a radius from shower axis (in mm) [45]. The showers are initiated by 30 GeV (a) π^+ and (b) protons. Additionally, the fitted data points are displayed, along with separate plots of the two components from Equation (4.21).

4.3. Construction of Calorimeters

4.3.1. Homogeneous and Sampling Calorimeters

In calorimeters, a distinction is made between a **homogeneous calorimeter**, where a single medium is responsible for both shower development and signal registration, and a **sampling calorimeter**, which uses two separate media, thereby decoupling the shower development from the signal recording process. A sampling calorimeter typically contains two types of layers: a **passive medium**, in which the shower develops, and an **active medium**, which records the resulting electric signals. These passive and active layers are

usually arranged alternately in the sampling calorimeter. Homogeneous calorimeters have the advantage of delivering a better energy resolution than sampling calorimeters because only a fraction of the deposited energy is measured through the active layers in a sampling calorimeter, leading to sampling fluctuations, whereas the whole calorimeter volume of homogeneous calorimeters can be used for readout. However, sampling calorimeters are more compact, have a better particle identification due to the separation of shower development and readout, and are more cost-effective [31, 39, 46]. A detailed description of the construction of the AHCAL prototype is provided in Chapter 5.

4.3.2. Energy Resolution of Calorimeters

The main criteria for evaluating the performance of a calorimeter is its energy resolution,

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c, \quad (4.22)$$

where \oplus means addition in quadrature [31]. Stochastic fluctuations of the total number of particles dominate the resolution for low energies and are considered in the first term. Assuming that the total number of particles N_{total} follows a Poisson distribution and thus possesses a standard deviation of $\sqrt{N_{\text{total}}}$, the resolution yields

$$\frac{\sigma_E}{E} \propto \frac{\sqrt{N_{\text{total}}}}{N_{\text{total}}} \propto \frac{1}{\sqrt{N_{\text{total}}}} \propto \frac{1}{\sqrt{E}}, \quad (4.23)$$

since $N_{\text{total}} \propto E$ [31]. Hence, this results in an improvement in resolution for higher energies [31]. The second term, represented by the factor b , accounts for electronic and thermal noise or radioactivity. The constant term, c , reflects mechanical and electronic irregularities of the calorimeter, as well as fluctuations caused by leakage. Figure 4.10 shows the relative energy resolution as a function of energy. Furthermore, the three terms of the resolution are shown separately. At low energies, the stochastic and noise terms dominate, whereas at high energies, the constant term is the primary factor in the relative energy resolution, particularly in the tail of the distribution [31]. The falling energy resolution makes calorimetry particularly attractive for high-energy physics experiments, because the higher the energy, the better the energy resolution of the calorimeter.

4.3.3. Software Compensation for Calorimeters

The resolution of calorimeters can be enhanced by converting a non-compensating calorimeter into a compensating one, ensuring an equal response to both EM and hadronic show-

4. Particle Showers in Calorimeters

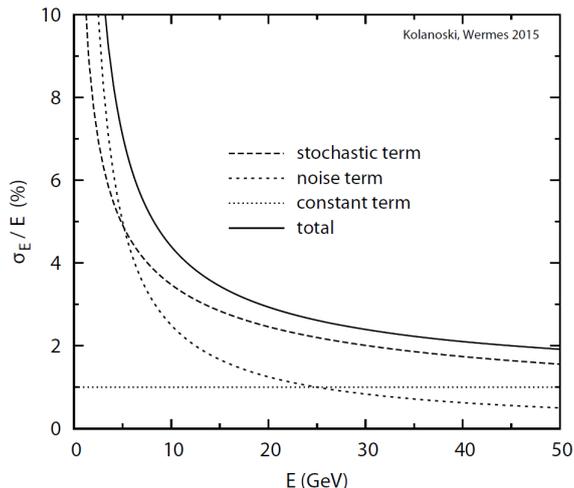


Figure 4.10.: The relative energy resolution, σ_E/E , and its three terms for calorimeters, expressed as a function of the energy [31]. In this case, the parameters are set to $a = 0.11 \sqrt{\text{GeV}}$, $b = 250 \text{ MeV}$, and $c = 0.01$.

ers, i.e. $e/h \rightarrow 1$. As mentioned in Chapter 4.2.1, the calorimeter's response to purely hadronic showers is generally lower than its response to electromagnetic (EM) showers, leading to $e/h > 1$. To achieve compensation, and with that the improvement of the energy resolution, the response to hadronic showers must be increased and/or the response to EM showers must be reduced. Two approaches can be considered: hardware or software compensation. For hardware compensation, various constructions with different materials can be used to achieve the compensation, e.g. the passive material could have a higher atomic number Z , as this reduces its sensitivity to the EM component. To enhance the calorimeter's response h to the hadronic component, sensitivity to neutrons or photons could be increased as well [31, 47].

However, achieving a compensating calorimeter through construction constraints requires careful material selection, often resulting in heavier and larger detectors, which in turn leads to higher costs and also worsens the EM shower resolution [31, 47]. Software compensation algorithms provide alternative methods for achieving calorimeter compensation. By correcting energy deposition, compensation can be achieved; however, it remains approximate due to shower-to-shower fluctuations. On an event-to-event basis, weights are applied to the energy deposition, ensuring that the response to EM and fully-hadronic-driven showers are equal on average. Here, an approach for the AHCAL is presented [48]. The reconstructed energy E_i of an hit i is weighted via w_{θ_i} that depends on observables θ_i , such as time of the hit:

$$E_{\text{SC},i} = E_i w(\theta_i). \quad (4.24)$$

The weight $w(\theta_i)$ has the form of

$$w(\theta_i) = a + b\theta + c\theta^2, \quad (4.25)$$

where the parameters a , b and c are determined from minimising the function

$$\chi^2 = \sum_i \frac{(E_i(a + b\theta_i + c\theta_i^2) - E_{\text{beam},i})^2}{E_{\text{beam},i}}, \quad (4.26)$$

with the beam energy E_{beam} [48]. A slightly different approach can be found in Ref. [49].

Additionally, machine learning techniques are also used to achieve compensation for a calorimeter [48, 50].

5. CALICE Collaboration/DRD6 and the AHCAL Prototype

An overview of selected possible future high-energy physics experiments is given in Chapter [5.1](#), along with an introduction to the CALICE Collaboration and the Detector Research and Development (DRD) Collaboration in Chapter [5.2](#). Lastly, Chapter [5.3](#) focuses on the structure and properties of the Analogue Hadron Calorimeter (AHCAL), whose recorded data was used in this thesis.

5.1. Possible High Energy Experiments in the Post-LHC Era

With the LHC reaching its limits, scientists are planning new high-energy experiments to explore the mysteries of the universe. The High-Luminosity upgrade of the LHC (HL-LHC) [\[51\]](#) is expected to be a major step towards answering open questions in (particle) physics. The commissioning of the HL-LHC will start around 2030. As more data is recorded, the detectors must also be upgraded to withstand the increased flux of particles, as 200 proton-proton collisions will occur on average per bunch crossing at a collision energy of $E \approx 14$ TeV at the HL-LHC [\[52\]](#). Ongoing research focuses on developing and testing detector components for future data collection, and besides the upgrade of the current biggest particle accelerator, future projects are proposed and discussed, such as:

- **Future Circular Collider (FCC):** With a circumference of around 90 km, it would be the largest (circular) particle accelerator that was ever built. The construction would be scheduled to begin in the 2030s, with data collection starting in the 2040s. Initially, it will operate as a lepton collider (electron-positron; FCC-ee). The FCC-ee is planned to be replaced in the future by a hadron collider (proton-proton; FCC-hh), designed to achieve a centre-of-mass energy of $E = 100$ TeV. The goal of the FCC is to conduct sensitivity studies of the SM, overcoming current limitations caused by the lower collision energy or lower luminosity at the LHC [\[53\]](#), [\[54\]](#). Research

5. CALICE Collaboration/DRD6 and the AHCAL Prototype

in the Higgs sector and precision measurements of extremely rare channels, such as $e^+e^- \rightarrow ZH$, will particularly benefit from the lepton collider, and it will also shine light to couplings between different particles [53–55]. The hadron collider FCC-hh focuses on the high-energy regime, providing insights into theories beyond the SM, where yet-unobserved exotic particles with high masses could be produced. Additionally, the FCC-hh could provide insight into the origin and structure of dark matter [53, 54, 56].

- **International Linear Collider (ILC):** The International Linear e^+e^- -Collider is proposed to be built in Japan, with a focus on high-precision measurements, particularly of particle masses, particle couplings, and cross-sections. Contrary to a hadron collider, where the interacting quarks within the hadrons carry only a fraction of the hadron’s momentum, the collision energy of a lepton collider is entirely available to the colliding leptons. Lepton colliders also provide cleaner initial states with well-defined energies, allowing for more precise measurements of fundamental particle properties and interactions [14]. Higgs physics, the Higgs couplings to massive bosons, and in particular the Higgs self-coupling can be studied at the ILC [57]. With a center-of-mass energy of $E = 350$ GeV, the ILC enters the mass range of top-quark pairs, providing an opportunity to investigate top-quark bound states. Moreover, the collider is designed to be upgradeable, ultimately reaching a collision energy of up to 1 TeV [58, 59]. The special feature about the ILC are the superconducting radio-frequency cavities for efficient and stable acceleration of the particle beams [58].
- **Compact Linear Collider (CLIC):** CLIC is a proposed linear e^+e^- -collider designed to achieve collision energies of up to 3 TeV, surpassing the beam energy of the ILC. A key objective of CLIC is the precise measurement of Higgs boson couplings to other particles, which is essential for testing the SM. Furthermore, its wide energy range allows searching for new physics beyond the SM [59–61]. A fundamental distinction between CLIC and the ILC lies in their acceleration mechanisms. While the ILC would utilise superconducting radio-frequency cavities, CLIC would use a two-beam acceleration technology. This enables a more compact accelerator design while achieving higher collision energies [59–61].

As the requirements for new detectors are increasing due to the higher collision energies and the increased particle flux, new detector components for future experiments must be developed and tested.

5.2. The CALICE and DRD6 Collaborations

As stated in Chapter 4, calorimeters are of great importance in particle physics experiments, as they are used to measure the energy of particles generated in high-energy collisions. This chapter introduces the CALICE and DRD6 Collaborations, and the AHCAL prototype, which is the calorimeter investigated in this analysis. Moreover, an overview of silicon photomultipliers is given.

The CALICE Collaboration focuses on the development of high-performance and highly granular calorimeters. CALICE is the abbreviation for **Calorimeter for Linear Collider Experiment**. With around 300 physicists, the collaboration is split up in three primary groups. The first group deals with EM calorimeters, the second with hadronic calorimeters and the third group with a tail catcher/muon tracker (TCMT). For this report, the work has been done in the hadronic section of the collaboration, namely with the **Analogue Hadron Calorimeter** (AHCAL) group [12].

In 2024, the CALICE Collaboration has become part of DRD6, which is the sixth task force of the **Detector Research and Development** collaboration, that was formed due to the European Committee for Future Accelerators (ECFA) roadmap [62]. The focus of this part of DRD is the development of calorimeters, while the other sections focus on gaseous, liquid and photon detectors, as well as semiconductors and particle identification. A total of 135 research institutes from 27 countries are participating in DRD6 [62].

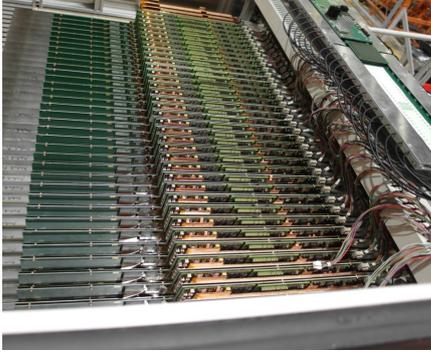
5.3. The AHCAL Prototype

The AHCAL is a sampling calorimeter, depicted in Figure 5.1(a), and consists of 38 active layers. Its active layers are made of scintillator tiles, as shown in Figure 5.1(b). Each active layer measures $72 \times 72 \text{ cm}^2$ in area, 3 mm in thickness and is composed of four HCAL Base Units (HBUs), each measuring $36 \times 36 \text{ cm}^2$. One HBU contains $12 \times 12 = 144$ scintillator tiles, with each tile having a dimension of $3 \times 3 \text{ cm}^2$. Consequently, an entire active layer is divided into $24 \times 24 = 576$ tiles, resulting in a total of $24 \times 24 \times 38 = 21888$ readout channels for the entire calorimeter. These tiles are individually read out via silicon photomultipliers (SiPMs, see Chapter 5.3.1). The SiPM *Hamamatsu MPPC* of type *S13360-1325PE* was chosen for the detector and is shown in Figure 5.1(c). Reflective foil on top of all scintillator tiles is used to reduce cross-talk between the tiles, namely the migration of photons from one SiPM cell to another during the amplification process [12, 63, 64].

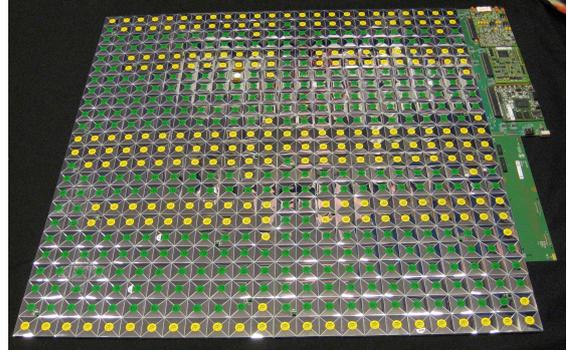
For the passive medium, non-magnetic stainless steel is used. One passive layer has a

5. CALICE Collaboration/DRD6 and the AHCAL Prototype

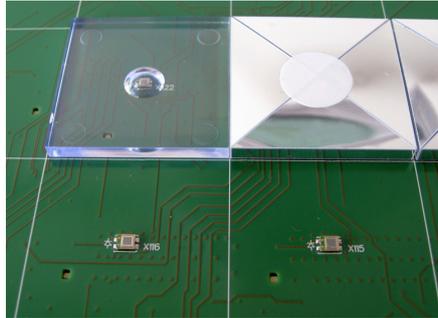
thickness of 17 mm, which corresponds to one radiation length and 0.1 nuclear absorption lengths, resulting in a total length of $4.4\lambda_a \approx 75$ cm. A passive layer is positioned between two active layers, thereby ensuring the shower development [12, 63, 64].



(a) The AHCAL prototype [63].



(b) One active layer of the AHCAL prototype [64].



(c) SiPMs (bottom) and wrapped with scintillator tiles (top left) and a reflective foil (top right) [12].

Figure 5.1.: (a) The AHCAL prototype, (b) one of the 38 active layers and (c) its SiPMs.

The energy resolution of the AHCAL prototype has been determined from showers initiated by negatively charged pions during a test beam run in 2018 (see Chapter 5.3.2). Table 5.1 summarises the energy resolution for the uncompensated calorimeter and compensated calorimeter with a local software compensation approach and a machine learning method based on a neural network [50]. The factor b in Equation (4.22) can be neglected. After applying software compensation algorithms, the energy resolution of the AHCAL improves. Local compensation weights the individual hit energies, while the machine learning approach takes the hadron shower event and the cell energies as inputs. Both methods yield comparable results, with the first term in Equation (4.22), representing stochastic fluctuations, showing a significant improvement after their application, especially with the neural network approach. Moreover, the constant term also shows an improvement after applying software compensation algorithms.

Table 5.1.: The contributions to the energy resolution in Equation (4.22) based on pion beam datasets from 2018 [50] for uncorrected and corrected energies for the AHCAL. The factor b in Equation (4.22) is negligible.

	a [%]	c [%]
uncorrected	56.1 ± 0.7	6.1 ± 0.1
local compensation	51.5 ± 0.42	1.0 ± 0.3
neural network	41.9 ± 0.5	4.0 ± 0.1

5.3.1. Silicon-Photomultipliers (SiPMs)

SiPMs are based on the principle of single photon avalanche diodes (SPADs), which are used for signal amplification in semiconductors. This is achieved by combining a highly doped p - and n -layer, p^+ and n^+ , where an extended electrical field reaches its maximum. The design of such SiPMs is shown in Figure 5.2. A high electric field is established across the p^+ - n^+ junction. This high-field region enables charge multiplication. When the silicon-based semiconductor in an SiPM is operated in reverse bias at a voltage slightly above the breakdown voltage, the device enters Geiger mode, where the signal becomes independent of the primary ionisation. When a photon, acting as the primary ionisation source, enters the material, it generates an electron-hole pair. The primary electron is accelerated by the high electric field, gaining kinetic energy and generating additional charge carriers through ionisation. The secondary charge carriers lead to the creation of additional electron-hole pairs, resulting in a multiplication process, where the number of charge carriers rapidly increases in the high-field region. The avalanche produces a current pulse that is detected as a signal, enabling the SiPM to register the presence of a single photon with high sensitivity. SiPMs are capable of achieving amplification factors/gain factors of approximately 10^6 [31]. However, signal generation can also be initiated via noise. The main contribution towards noise comes from thermal excitation of charge carriers at higher temperatures and crosstalk. Thus, to avoid noise, operating the SiPMs at lower temperatures and the use of guard rings or reflective foils is beneficial [31, 65, 66].

The SiPMs used in the AHCAL-prototype were operated at 5 V above breakdown voltage and achieve a gain of approximately $7 \cdot 10^5$. The small crosstalk of about 1 % and very low temperature sensitivity of 54 mV/K makes them suitable for high-energy-physics [12].

5.3.2. Test Beam Campaign

The data that was used for this investigation has been recorded in three different periods during a test beam campaign in 2018 at the Super Proton Synchrotron (SPS)

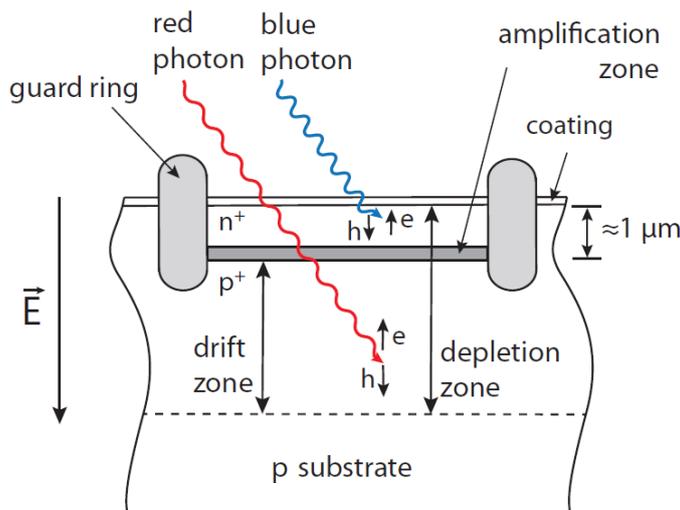


Figure 5.2.: The structure of an SiPM [31]. A photon generates an electron-hole pair, who both move towards their respective electrodes. In the high-electric-field region, the electron accelerates, creating secondary charge carriers through ionisation, which results in the creation of additional electron-hole pairs. This triggers an avalanche of charge carriers, resulting in an amplified signal. The purpose of guard rings is to mitigate crosstalk between neighbouring SiPMs.

facility of CERN. The datasets used in this thesis are based on the 38-layer setup of the AHCAL. The calorimeter was tested using muons with energies of 40 GeV and 120 GeV for calibration, as well as negative pion and electron beams with energies of 10, 15, 20, 30, 40, 50, 60, 80, 100, 120, 160, 200 GeV and 10, 20, 30, 40, 50, 60, 80, 100 GeV, respectively. For all three particle types, approximately $\mathcal{O}(10^7)$ events were recorded. The first test was conducted in May 2018, where the AHCAL prototype was mounted on a movable platform to allow the beam to hit the calorimeter at various positions in the xy -plane. A signal was recorded only when two external scintillation triggers detected a particle at the same time. The second test beam run was conducted a month later, with an additional layer of scintillator tiles of size $6 \times 6 \text{ cm}^2$ inserted between the 37th and 38th active layer. Additionally, a pre-shower detector was positioned in front of the AHCAL to determine the energy deposition before the beam enters the calorimeter, while a TCMT was placed behind it to capture the remnants of the showers and to avoid leakage, as seen in Figure 5.3. The pre-shower detector comprises a single HBU unit, whereas the tail catcher consists of twelve layers, each containing one HBU unit placed between the 7.4 mm thick absorbers. In the third test beam run in October 2018, the CMS High-Granularity Calorimeter (HGCal) [67, 68] was placed in front of the AHCAL. Therefore,

5.3. The AHCAL Prototype

only the shower tails and muons reached the AHCAL and thus, only 2.5 million events were recorded. In total, 93 million events were recorded in 2018 [12, 69, 70].



Figure 5.3.: The design of the AHCAL prototype and the TCMT for the test beam run at CERN in 2018 [50].

6. Fast Simulation for Hadronic Showers

This chapter focuses on the aspect of a data-driven fast simulation for pion showers in the AHCAL prototype. Due to the high computational power and time required for simulating particle interactions with matter, a data-driven approach is considered for fast simulation. This approach significantly reduces the need for computational resources compared to Monte Carlo simulations, such as with GEANT4 [71]. Also, fast simulations play a crucial role in tackling the challenges introduced by higher luminosity and increased detector granularity in collider experiments, which significantly extend the CPU time required for (Monte Carlo) simulations. A significant fraction of this computational effort is dedicated to calorimeter simulations. To optimally use computing resources, developing efficient methods for the simulation of particle showers is essential [72]. The dataset used in this investigation comprises the pion beam energies of 20, 40, 60, 80, 120, and 200 GeV from the 2018 test beam run.

This chapter provides an overview of the basis of the approach for the data-driven fast simulation chosen for this investigation, Kernel Density Estimators (see Chapter 6.1). Additionally, a brief summary of other data-driven fast simulation algorithms for pion showers is provided in Chapter 6.2.

6.1. Kernel Density Estimators (KDEs)

In order to simulate showers, a data-driven approach is employed in this analysis. The dataset obtained from the test beam runs serves as the basis for the fast simulation. To predict the underlying probability density function (PDF) of the dataset in the absence of an analytical function, estimators are used. This analysis uses **Kernel Density Estimators (KDEs)** to simulate a given dataset. This investigation uses KDEs for the simulation of the deposited energy for each tile in the calorimeter. A kernel function is assigned to each x_i from the dataset $\{x_1, x_2, \dots, x_N\}$, $K(x_i)$, that can be any non-negative

6. Fast Simulation for Hadronic Showers

density function. In this instance, a Gaussian distribution

$$K(x_i) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x_i^2\right) \quad (6.1)$$

with a mean μ_i of zero and a standard deviation σ_i of one has been selected as the kernel function [73]. The sum of kernel functions of all data points in the dataset is used to estimate the PDF (Ref. [13]):

$$f(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right). \quad (6.2)$$

The estimated PDF is normalised with the total number of data points in the dataset n and the bandwidth parameter $h > 0$, which exerts a significant influence on the shape of the PDF. An example is shown in Figure 6.1, where the blue crosses represent the data points. The black curve represents the true PDF underlying the blue data points. The blue, red, and green curves are estimated PDFs obtained using KDEs with varying bandwidth values specified. The blue curve uses a bandwidth of $h = 0.05$, the red curve uses $h = 0.5$, and the green curve uses $h = 2$. As previously mentioned, the bandwidth parameter controls the smoothness of the estimated PDF: smaller bandwidths (e.g. the blue curve) result in less smoothness, causing the PDF to peak at individual data points, while larger bandwidths (e.g. the green curve) flatten the distribution too much, potentially obscuring finer details of its true structure. The red curve provides a good approximation of the true PDF.

A grouping of calorimeter cells, henceforth referred to as "buckets", were introduced to reduce the total number of readout channels of the AHCAL. The geometrical categorisation of the calorimeter cells into buckets is based on radial, angular and longitudinal segmentations of the calorimeter cells. At first, the 38 active layers of the AHCAL are grouped into eight layer groups: {1-2, 3-4, 5-6, 7-8, 9-12, 13-16, 17-24, 25-38}. Eight circles with radii {25, 50, 75, 100, 150, 200, 300, 400} mm and eight angular segments $\{\frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}, \pi, \frac{5\pi}{4}, \frac{3\pi}{2}, \frac{7\pi}{4}, 2\pi\}$ are then the structure for the 64 buckets per layer group, which can be seen in Figure 6.2. Thus, the 21888 cells of the AHCAL are grouped into 512 buckets.

For the fast simulation, different bandwidth parameters were investigated to determine the best value for the PDF of energy differences in each bucket,

$$\Delta E = E_S - E_A, \quad (6.3)$$

where E_S describes a single energy in a bucket of the calorimeter and E_A the average energy of the said bucket. Figure 6.3 shows the energy distributions. For the simulation

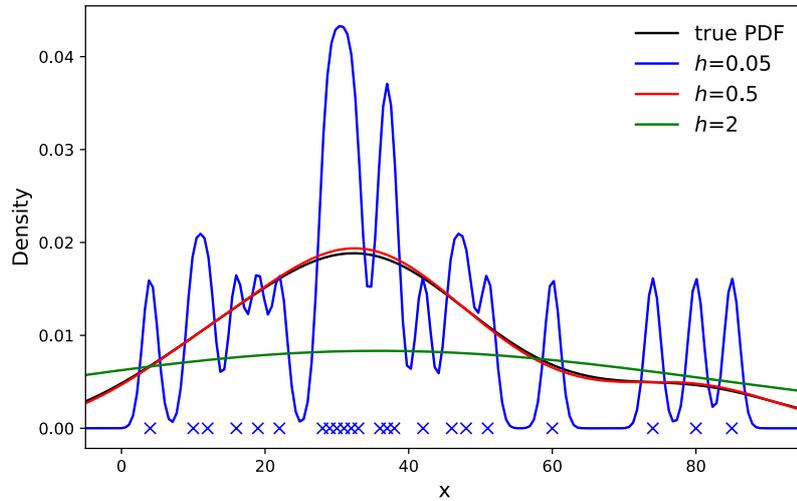


Figure 6.1.: The effect of the bandwidth h on a KDE-estimated PDF: the black curve represents the true distribution, while the blue, red, and green curves show estimates with $h_{\text{blue}} = 0.05$, $h_{\text{red}} = 0.5$, and $h_{\text{green}} = 2$.

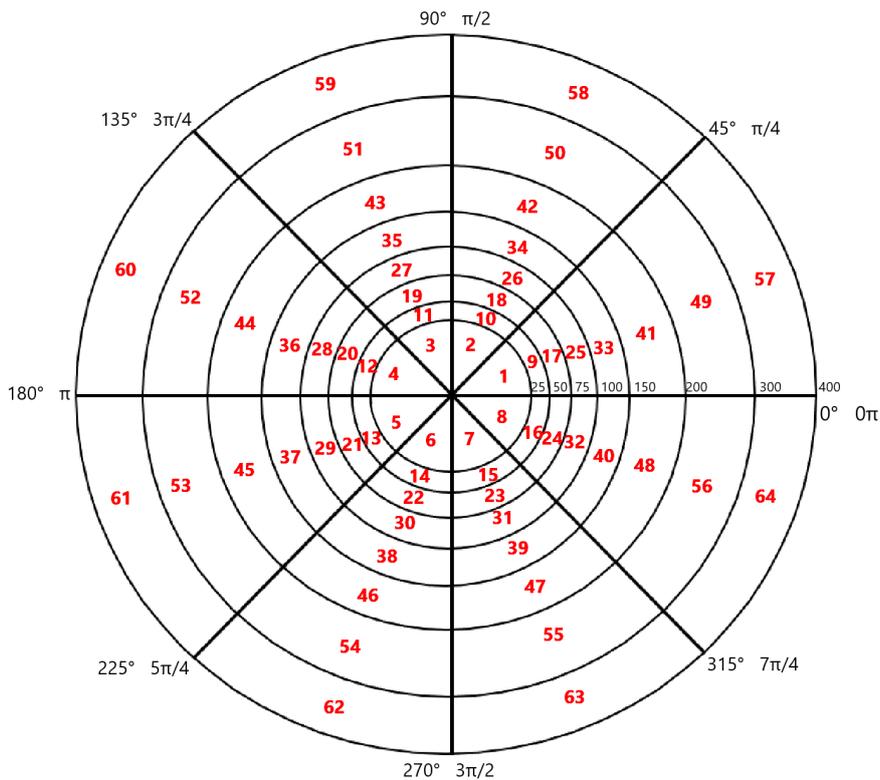


Figure 6.2.: The 64 buckets for each of the eight layer groups based on a radial and angular classification.

6. Fast Simulation for Hadronic Showers

of these energy differences with KDEs, 28 values for the bandwidth were used: $h \in \{0.001, 0.002, \dots, 0.009, 0.01, 0.02, \dots, 0.09, 0.1, 0.2, \dots, 0.09, 1\}$. Here, the distributions for bucket number 261, which corresponds to the fifth bucket in the third layer group, are shown for $h = 0.001, 0.01, 0.1$ and 1 . The fifth bucket in the third layer group corresponds to an angular bin of π and a radial bin of 25 mm, as seen in Figure 6.2, and layer five and six which are added to form the third layer group. The larger the bandwidth becomes, the worse the agreement between data and simulation will be.

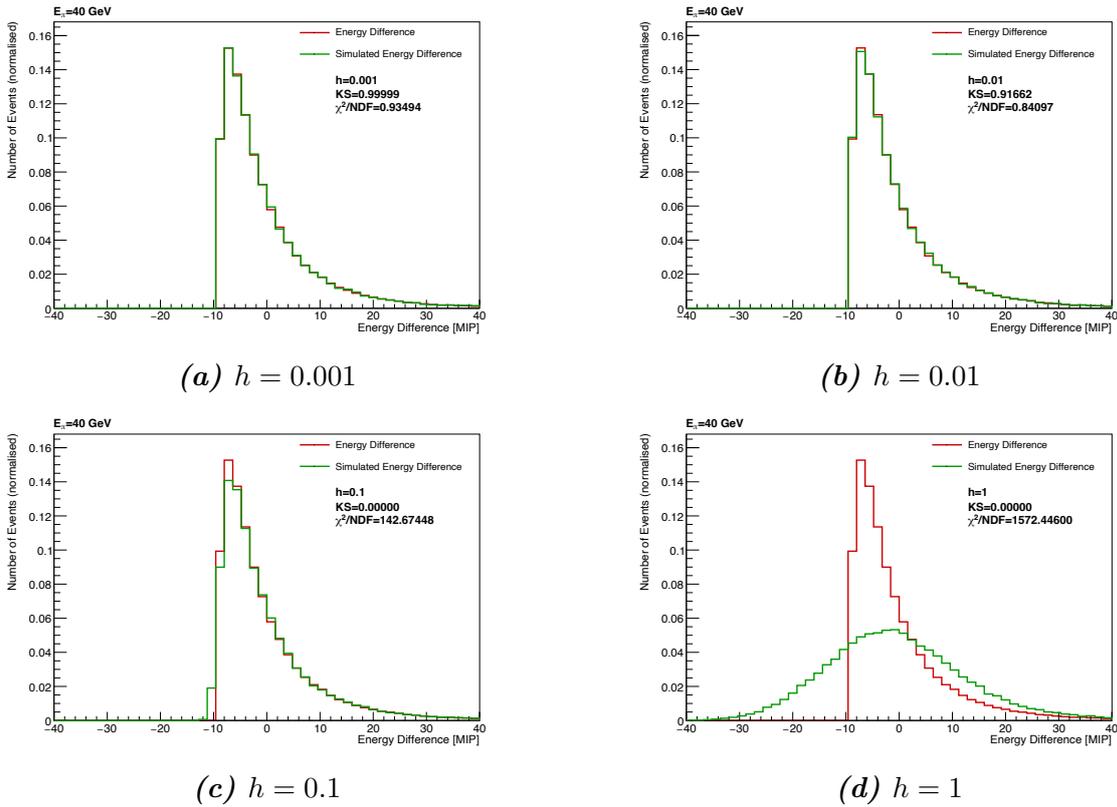


Figure 6.3.: The distributions of the energy differences ΔE in units of MIP for $h = 0.001, 0.01, 0.1$ and 1 for the fixed bucket number 261(= fifth bucket in the third layer group). The red histograms show the energy distributions obtained from data and the histograms in green correspond to the simulated energy differences. Additionally, the KS probability and χ^2/NDF values are displayed.

To determine the degree of agreement between data and simulation, two statistical tests were taken into consideration: the Kolmogorov-Smirnov (KS) test [74, 75] and the χ^2/NDF [76] test. The KS test assesses the shape compatibility between two histograms (data and simulation) in each bucket by calculating the supremum (= maximum absolute

difference [73]) between them,

$$D_n = \sup_X |F_X - F_{\text{emp}}|, \quad (6.4)$$

where F_X is the distribution obtained from simulation, representing the KDE from Equation (6.2), while F_{emp} is obtained from data [77]. The supremum of the differences between two histograms is computed across all bins. The maximum of these values is used as the KS test statistic, which is then converted into a probability,

$$P(z) = 2 \sum_{j=1}^{\infty} (-1)^{j-1} e^{-2j^2 z^2} \text{ with } z = D_n \sqrt{n}, \quad (6.5)$$

where n describes the sample size [78, 79]. A KS probability of zero indicates no agreement between the data and simulation, while a probability of one means perfect agreement.

The χ^2/NDF test between two histograms is determined as follows (Ref. [76, 80])

$$\chi^2/\text{NDF} = \frac{1}{\text{NDF}} \sum_{i=1}^{\text{number of bins per bucket}} \left(\frac{x_i^{\text{sim}} - x_i^{\text{data}}}{\sigma_i^{\text{sim}}} \right)^2. \quad (6.6)$$

It computes the sum of the squared differences between data and simulation, normalised to the squared standard deviation of the simulation, for each bin in an energy distribution of a bucket. The index i runs over the number of bins in an energy distribution of a bucket and the sum is then normalised by the number of degrees of freedom (NDF), i.e. the number of bins in the respective bucket. χ^2/NDF is always non-negative. A value of zero indicates perfect agreement between two histograms; the higher the value, the greater the disagreement.

Both tests were calculated for each of the 512 buckets for each bandwidth. The mean KS test value and mean χ^2/NDF value for each bandwidth were then determined, and the results in Figure 6.4 and 6.5 demonstrate that a value of $h = 0.01$ exhibits a good balance between over-fitting and under-fitting the data.

6. Fast Simulation for Hadronic Showers

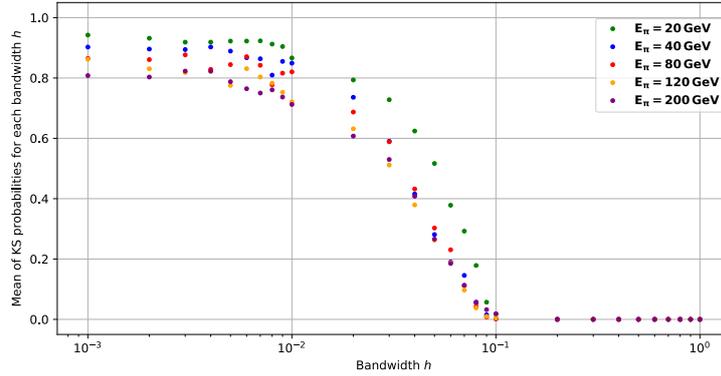


Figure 6.4.: The KS probability for 28 different bandwidth parameters for different pion energies. Both axes are presented on a logarithmic scale.

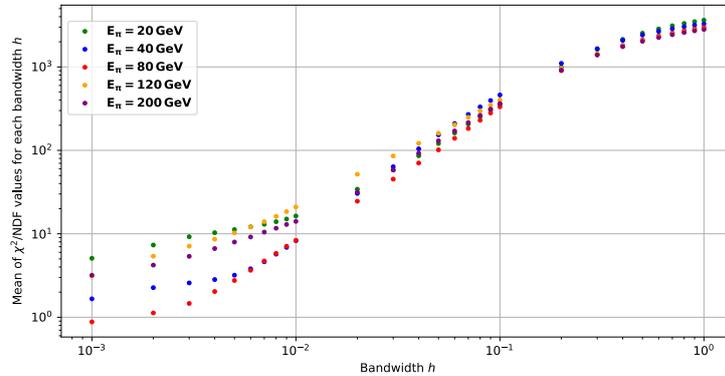


Figure 6.5.: The χ^2/NDF test for 28 different bandwidth parameters for different pion energies. Both axes are presented on a logarithmic scale.

6.2. Alternative Fast Simulation Algorithms for Particle Showers

Different algorithms have already been investigated for a fast simulation of particle showers. As previously mentioned, the goal is to reduce computational requirements for the simulation of particle showers. The **binning based on the geometrical categorisation of the calorimeter tiles** into buckets has already been described above. In this approach, the original $24 \times 24 \times 38 = 21888$ tiles of the AHCAL prototype are reduced to 512 readout channels for the fast simulation. The energy depositions in these buckets form the input for the KDEs. The main difficulty of this approach is the geometrical transformation of the hit energies per tile to the buckets. Its performance and more details on the algorithm can be found in the PhD thesis of Julian Utehs (to be published).

A different method is based on a **Principle Component Analysis (PCA)** [81], which was investigated by André Wilhahn in his Master's thesis [82]. A PCA converts a set of correlated data points into an uncorrelated one. In this approach, the distributions of the computed energy differences from data (see Equation (6.3)) are centred around zero and normalised to their standard deviations. Next, the PCA is applied by transforming the energy differences into uncorrelated principal components and obtaining their PDFs. The inverse transformation is then performed, converting the principal components back into simulated energies, which remain normalised and centred around zero. Finally, the simulated energy differences are obtained by scaling the normalised energy differences with their respective standard deviation and adding the respective mean value. For this algorithm, the energy depositions in the last seven layers of the AHCAL were summed, as less energy is deposited in the rear region of the calorimeter. On the contrary, the PCA was applied layer-by-layer to the first 32 layers. As a result, a total of 33 layers were used as input for the PCA. The simulated energy differences obtained through the PCA showed significant deviations from the distribution of energy differences in data, which is why this approach is no longer being pursued.

A third approach, that exploits the **Discrete Cosine Transformation (DCT)** [83], is currently under investigation. A dataset $\{x_0, x_1, \dots, x_{N-1}\}$ is transformed into another of same size $\{X_0, X_1, \dots, X_{N-1}\}$ via (Ref. [83])

$$X_k = \sum_{n=0}^{N-1} x_n \cos \left[\frac{\pi}{N} \left(n + \frac{1}{2} \right) k \right]. \quad (6.7)$$

Since three numbers are required to determine the position of a hit energy in the calorimeter, this approach utilises a three-dimensional DCT. However, the computation of a three-dimensional DCT has a runtime of the order $\mathcal{O}(n^6)$, where n represents the number of outputs. Thus, a Fast Cosine Transformation (FCT) [84] is required and is based on a Fast Fourier Transformation (FFT) [85, 86]. Since the formulas behind the FFT are more straightforward, it is presented here. The key idea is that the FCT/FFT separates the sum over hit energies into even and odd indices (Ref. [86]):

$$\begin{aligned} X_k &= \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i}{N} nk} \text{ with } k = 0, 1, 2, \dots, N-1 \\ &= \underbrace{\sum_{m=0}^{\frac{N}{2}-1} x_{2m} e^{-\frac{2\pi i}{N/2} mk}}_{\text{even indices } E_k} + e^{-\frac{2\pi i}{N} k} \underbrace{\sum_{m=0}^{\frac{N}{2}-1} x_{2m+1} e^{-\frac{2\pi i}{N/2} mk}}_{\text{odd indices } O_k} \\ &= E_k + e^{-\frac{2\pi i}{N} k} O_k \end{aligned} \quad (6.8)$$

6. *Fast Simulation for Hadronic Showers*

With the FCT/FFT, the runtime reduces to $\mathcal{O}(n \log(n))$ (or less, depending on the implementation details). Results show that only the even-even nodes (= even x - and y -components) have large coefficients and thus, even-even nodes dominate in the FCT. These nodes are then used in the simulation process with KDEs. The results will be shown in the PhD thesis of André Wilhahn (to be published).

7. Cell-Ordering Algorithm

A new fast simulation algorithm based on ordering calorimeter cells is proposed in this chapter. This approach is to replace the current binning method based on a geometrical classification of the calorimeter cells, as described in Chapter [6](#). This chapter starts with the event selection procedure for all datasets, followed by an explanation of the algorithm and the results obtained for 40 GeV pion showers are shown afterwards.

The general approach and its individual steps are illustrated in Figure [7.1](#). Blue boxes represent data-related steps, red boxes correspond to simulation steps, and purple boxes indicate the results. Initially, the hit energies and their corresponding positions are extracted from the dataset, and the centre of gravity (CoG) in the xy -plane is computed for each event. Next, the distance of each tile to the CoG is determined, and the tiles are arranged in ascending order based on this distance. These steps are described in more detail in Chapter [7.2](#).

For the KDE-based simulation, CoGs are simulated for each event, followed by the ordering of tiles based on their distances from the simulated CoGs, as discussed in Chapter [7.3](#). In the final step, the ordered energy values from data serve as input for the KDEs, which then produce simulated hit energies, sorted by distance. This process is elaborated upon in Chapter [7.4](#). Lastly, the simulated hit energies are assigned to the ordered tiles according to their distance from the simulated CoGs.

7.1. Event Selection

Different criteria were defined for the events to pass in order to be considered in the investigation:

- Energy deposition in the first physical layer is not considered in order to minimise uncertainties in the shower start finding algorithm [\[69\]](#). Therefore, in the following, the layers are indexed from zero to 37.
- A particle identification algorithm based on a boosted decision tree was applied for the removal of beam contamination [\[69\]](#).

7. Cell-Ordering Algorithm

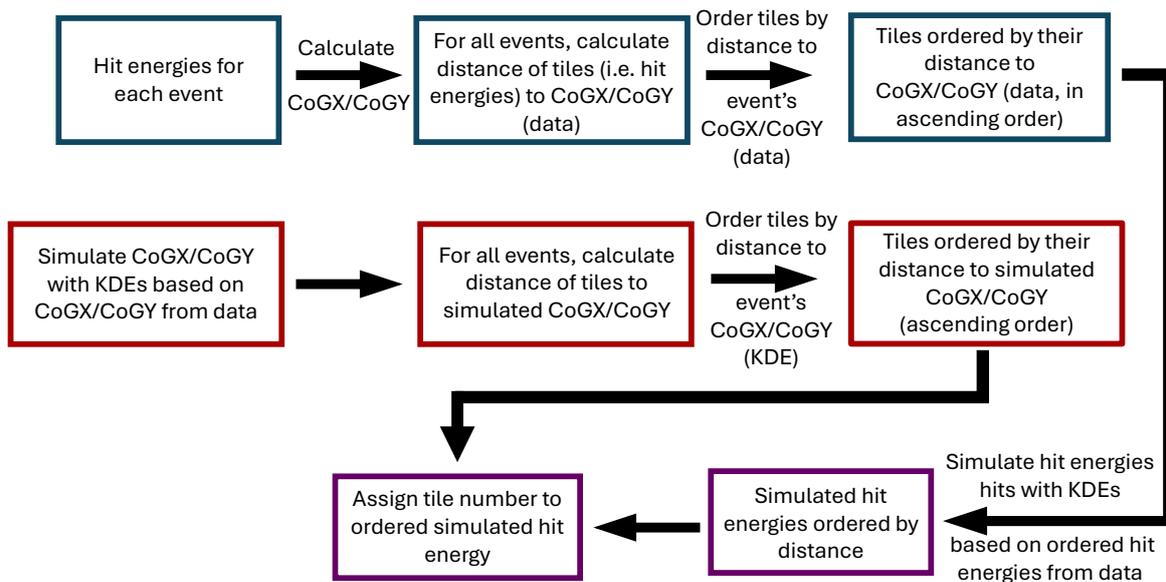


Figure 7.1.: Strategy for the cell-ordering algorithm for a fast pion shower simulation.

- A low-energy cut has been applied in order to remove muon events, which the particle identification algorithm could not filter out itself. For a given pion energy, ten percent of the most probable value of the corresponding total energy PDF were used as threshold.

7.2. Cell-Ordering based on Centres of Gravity

The approach for the fast simulation is based on the distance of the tiles to the event's centre of gravity, CoG, in the x - and y -direction (CoGX and CoGY). The CoG in a specific direction is defined as the sum over all hits of the hit energy E_{hit} times the hit position in the respective direction, normalised to the total energy E_{total} of the event:

$$\text{CoG}_i = \frac{1}{E_{\text{total}}} \sum_{\text{hits}} E_{\text{hit}} \cdot i_{\text{hit}} \text{ with } i_{\text{hit}} \in [x, y, z]. \quad (7.1)$$

The function `scipy.stats.gaussian_kde()` provided by the Python package SciPy [87] is unable to simulate all $24 \times 24 \times 38$ hit energies of the 100000 events simultaneously since many hits in an event, on average, have zero energy. Since the KDEs read the given energies as a matrix and invert it, the presence of a considerable number of zeros results in the matrix becoming singular, thereby preventing the KDE from inverting the matrix.

7.2. Cell-Ordering based on Centres of Gravity

Consequently, only the first 20 layers of the detector were considered for now. Given that the global CoGX/CoGY per event was determined, the distance calculation and ordering of the calorimeter cells were only required for a single layer per event. The distance was calculated using the Euclidean distance formula

$$d = \sqrt{(I - \text{CoGX})^2 + (J - \text{CoGY})^2}, \quad (7.2)$$

where I and J represent the tile centre position in millimetres in the x - and y -direction, respectively, with the coordinate system's origin situated at the bottom left corner of the layers [73]. A schematic view of the distance calculation algorithm is shown for a 5×5 grid in Figure 7.2. The distance from the centre of each tile to the exact position of the event's CoG, here indicated as a star, is calculated. Subsequently, the tiles are arranged in ascending order based on their distance to the event's CoG in the xy -plane. As a result, the tile corresponding to the event's CoG tile is always the first tile/bin and is labelled with the tile number zero.

23	19	18	21	24
16	7	6	10	20
11	2	0 *	4	14
12	3	1	5	15
17	9	8	13	22

Figure 7.2.: A schematic view of the distance calculation from the centre of the tiles to the event's centre of gravity tile.

Figure 7.3 shows the distances in millimetres for the first two events. The white star indicates the CoG in the xy -plane.

Subsequently, the hit energies per tile and their tile positions were extracted from data for 100000 events.

7. Cell-Ordering Algorithm

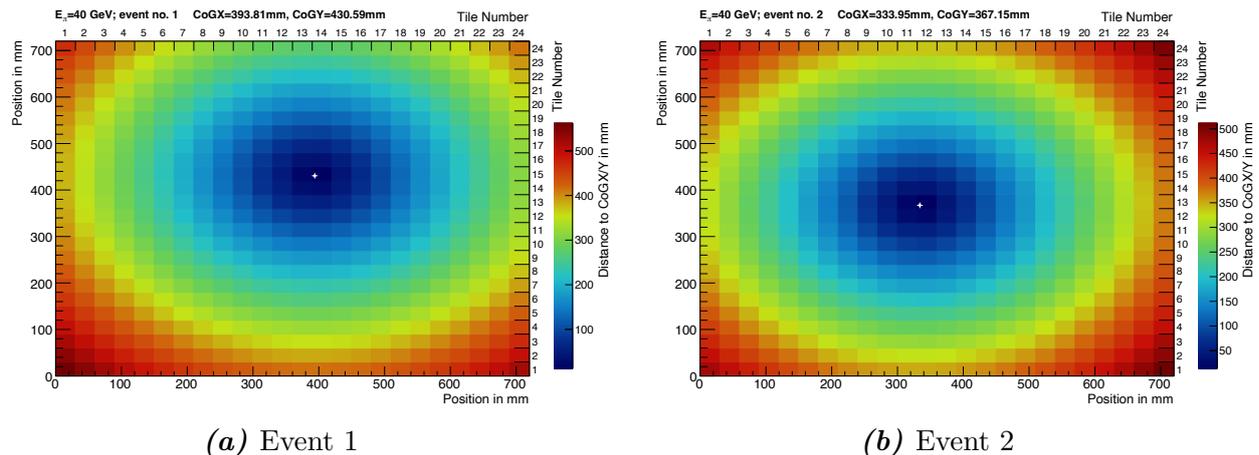


Figure 7.3.: The distance of the tiles to the event’s global CoG for the first two events. The axes show the tiles in mm and tile numbers. The white star indicates the CoG in the xy -plane

7.3. Simulation of the Centres of Gravity with KDEs

Since a tile number relative to the shower CoG in all three directions must be assigned to the simulated hit energies (see Chapter 7.4), simulating the CoGs is essential. The 24×24 tiles need to be sorted in ascending order of distance to the simulated CoG in the xy -plane. This is achieved via KDEs with the bandwidth parameter set to $h = 0.01$, in which the CoGs in all three dimensions are simulated simultaneously. For 100000 extracted events in the data, the CoG in all three dimensions were computed for each event, resulting in a total of 300000 values, which serve as the input for the CoG simulation with KDEs. From the estimated PDF, 10000 CoGs were sampled, which include three values per event (x -, y -, and z -direction). Figure 7.4 shows the CoGs in all three dimensions for data and simulation for 40 GeV pions. The distributions for the other datasets are shown in Appendix A.1. The axes show the tiles in mm and layer number. The taller a peak in the distribution, the deeper the CoG is located inside the detector. Additionally, the colour indicates the layer number. A significant fraction of the CoGs in the xy -plane is centred in the detector and only a few events have CoGs further to the edges of the calorimeter layers. Table 7.1 summarises the mean values and standard deviations of the CoGs, and it is clearly visible that the mean values and standard deviations of the simulated CoGs align well with the ones obtained from data.

The distributions of the CoGs as functions of tile number or layer number are shown in Figure 7.5, demonstrating strong agreement between data and simulation. In the x - and y -directions, the distributions exhibit the expected Gaussian behavior, whereas the CoGZ distribution follows a Landau distribution, peaking around layer nine or ten. For

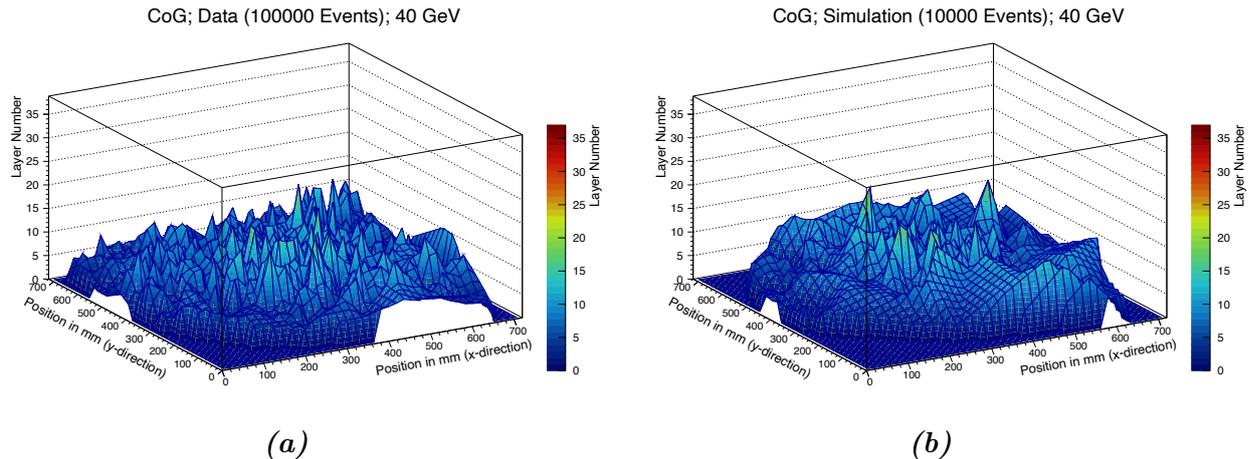


Figure 7.4.: The distributions of the CoGs in all three directions for (a) data and (b) simulation with KDEs for 40 GeV pions. The axes show the tiles in mm and layer number, which is also shown in colour. A small fraction of the CoGs are on the edges of the layers, and the higher the peak, the larger the value of the CoG in z -direction.

the five other datasets, the corresponding tables and histograms are shown in Appendix [A.2](#). Subsequently, the distances of the 24×24 tiles to the simulated CoGs in x - and y -direction are then calculated via Equation [\(7.2\)](#).

Table 7.1.: The mean values and standard deviations of the CoGs in each direction for data and the respective simulation for 40 GeV pions.

40 GeV	CoGX [mm]	CoGY [mm]	CoGZ [layer]
Data	383.77 ± 37	382.82 ± 31	9 ± 3
Simulation	383.83 ± 37	382.44 ± 29	9 ± 3
	CoGX [tile]	CoGY [tile]	CoGZ [layer]
Data	13 ± 2	13 ± 2	9 ± 3
Simulation	13 ± 2	13 ± 1	9 ± 3

7.4. Simulation of Energies with KDEs

In order to simulate hit energies, the sorted hit energies by distance to CoG from data were passed to the KDEs with a value of $h = 0.01$ as bandwidth. In a similar manner to the simulation of the CoGs (see Chapter [7.3](#)), a single multidimensional PDF was generated based on the $((24 \times 24)_{\text{tiles}} \times 20_{\text{layers}})$ hit energies per event for the estimation of the true PDF of these hit energies. From the estimated PDF, 10000 events were extracted, each comprising $(24 \times 24)_{\text{tiles}} \times 20_{\text{layers}}$ hit energies. As the hits from data

7. Cell-Ordering Algorithm

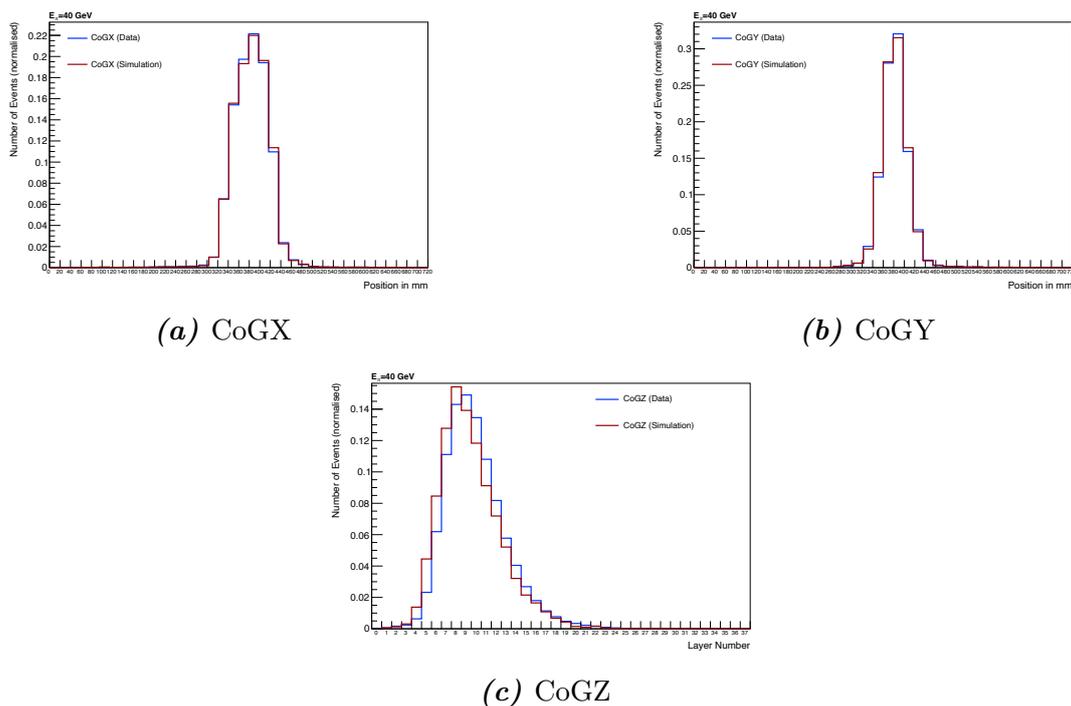


Figure 7.5.: The normalised distributions for the simultaneously simulated CoGs with KDEs.

were ordered according to their distance to the CoGX/CoGY, the energies in the 10000 simulated events are automatically sorted. Therefore, the simulated energies are assigned to the sorted tiles based. In the following, all energies are expressed in units of minimum ionising particles (MIPs). The conversion factor from GeV to MIPs for this thesis was determined to be 37.3 MIP/GeV [88]. Moreover, due to the KDE simulating hit energies smaller than zero, which originate from the extraction of the hit energies per event from the estimated PDF, a threshold was established: all tile-energies less than 0.05 MIP are set to zero.

7.4.1. Energy Distribution per Tile

Figure 7.6 shows the energy deposition per tile, ordered by distance, for data and simulation for layers zero, five and 12. This provides a more effective approach of comparison between the two, as the tiles are categorised by distance and the hit energies are simulated for each tile category. The first tile, denoted as tile zero, corresponds to the CoG tile itself, which varies between events. The second tile is thus the next closest tile to the event's CoG in the xy -plane, and so forth. The distributions exhibit strong agreement between the data and the simulation. At higher hit energies, statistical fluctuations become

more pronounced in the simulation histograms due to the limited sample size of 10000 events. Nevertheless, the data and simulation remain in good agreement over the entire energy range. Moreover, energy deposition generally decreases as the distance from the CoG-tile increases. However, as the shower develops, energy is deposited further away from the centre of the detector, affecting more tiles and leading to an increased number of registered hits in the layers deeper within the detector. Around layer number 4 or 5, tiles farther from the CoGs generally record more hits, as the maximum in energy deposition in the longitudinal direction is reached, as shown in Figure 7.7. As the shower continues to develop, the overall energy deposition decreases and even more tiles in the outer region of the layers register hits.

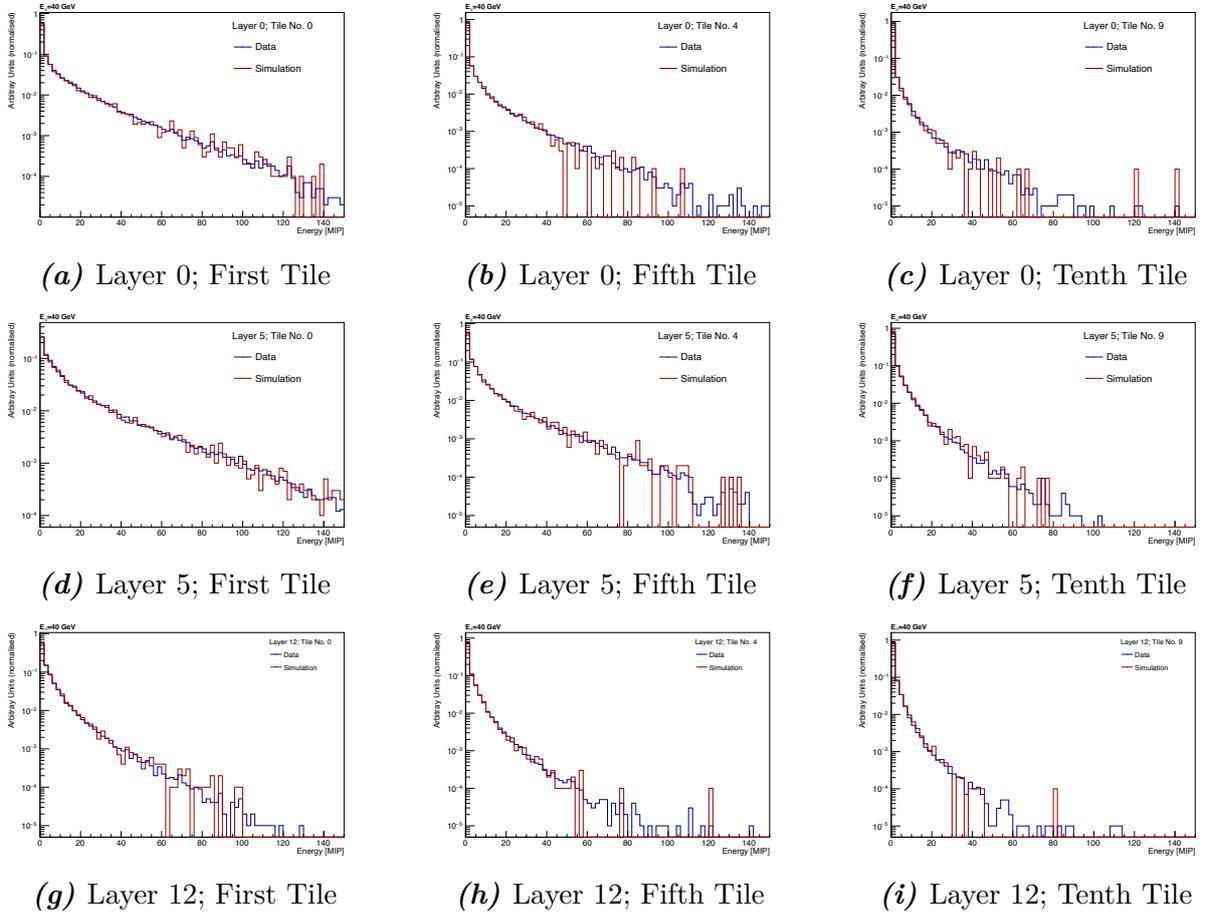


Figure 7.6.: The normalised deposited energy for the first, fifth and tenth closest tiles to the event's CoGX/CoGY in the first, sixth and 12th layer (layer number zero, five and 12, respectively). Simulated hit energies with a value of less than 0.05 MIP are set to zero. Additionally, the y -axis is plotted on a logarithmic scale.

7. Cell-Ordering Algorithm

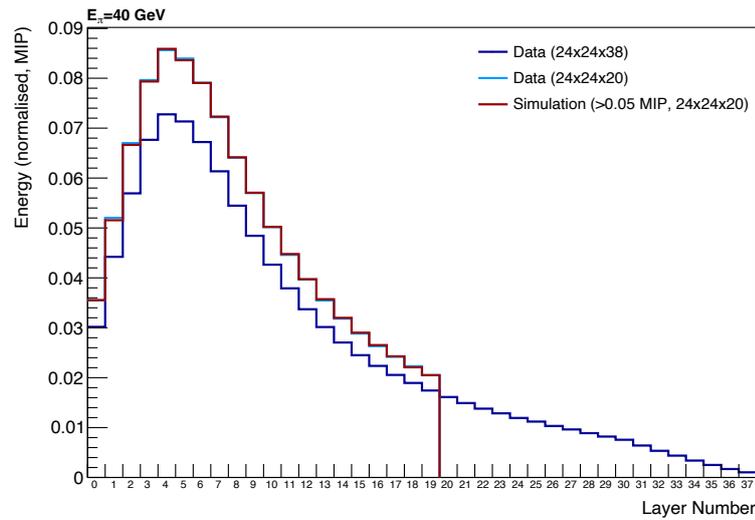


Figure 7.7.: The longitudinal energy distribution for the full dataset ($(24 \times 24)_{\text{tiles}} \times 38_{\text{layers}}$) and the reduced dataset ($(24 \times 24)_{\text{tiles}} \times 20_{\text{layers}}$) and its simulation. The distribution of the reduced dataset (light-blue) is covered by the distribution of the simulation (red).

7.4.2. Longitudinal Energy Deposition

The deposited energy per tile has been plotted for data and simulation and is shown for the first, tenth and 20th layer for an example event in Figure 7.8.

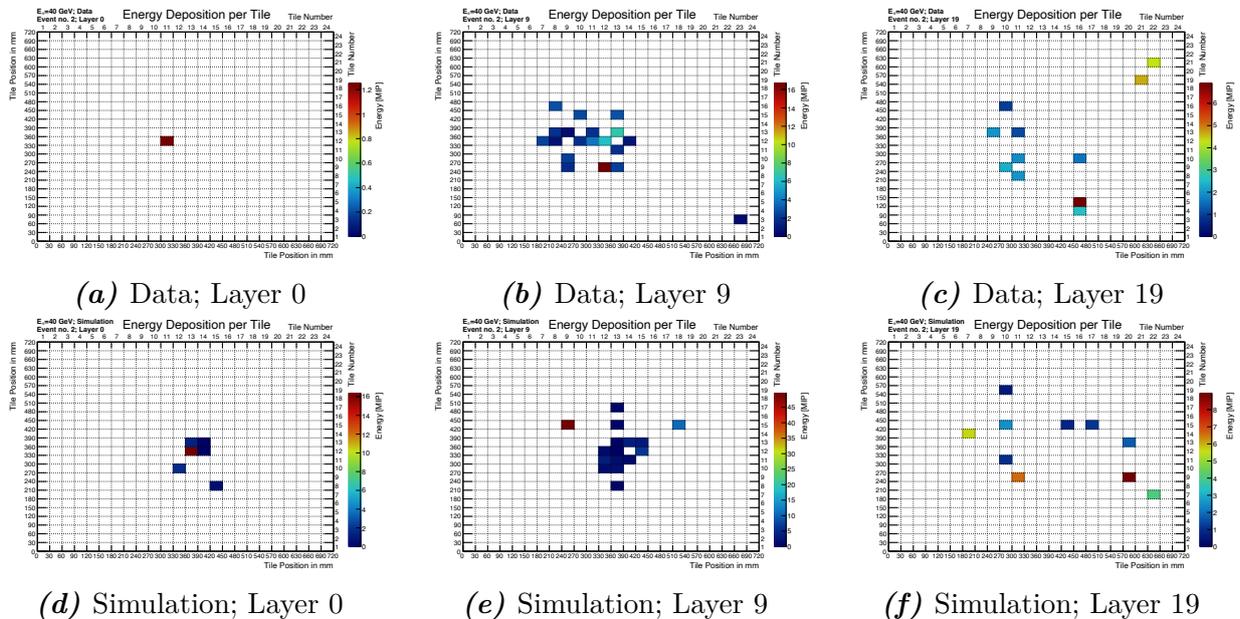


Figure 7.8.: The deposited energy of the second event for the first, tenth and 20th layer for data and simulation.

The deposited energy from data, especially in the first layers, show that the hits do not all lie in the centre of the layers, but that they are slightly shifted. This is due to a slight tilt of the AHCAL prototype during the test beam runs.

As anticipated, the simulation exhibits the expected behaviour. In the initial layers, the deposited energy is closer to the centre to the layers. There is a greater deposition of energy in the central region of the detector. Additionally, a higher number of tiles register some energy. As the pion shower progresses, the shower radius increases. This can be seen, as tiles further away from the centre of the layer also register hits, and the overall energy deposition in the layers further back decreases, indicating that the simulated shower has reached its maximum longitudinal extent.

8. Shower Shape Variables

To examine whether the simulation based on the distance-sorting algorithm also simulates a shower's kinematic behaviour correctly and not only its hit distributions, shower shape variables have been investigated. This investigation ensures that the algorithm properly captures the physical characteristics and evolution of particle showers within the calorimeter. Chapter [8.1](#) defines the variables. As the simulation has to neglect tiles or layers due to the limitation by KDEs, Chapter [8.2](#) introduces an approach based on Gaussian noise to account for the energy loss. Lastly, Chapter [8.3](#) shows the distributions of the shower shape variables.

8.1. Definition of Shower Shape Variables

The following variables are investigated:

- Total Energy: $E_{\text{total}} = \sum_{\text{hits}} E_{\text{hit}}$,
- Centre of Gravity: $\text{CoG}_i = \frac{1}{E_{\text{total}}} \sum_{\text{hits}} E_{\text{hit}} \cdot i_{\text{hit}}$ with $i \in [x, y, z]$,
- Central Fraction: $F_{\text{central}} = \frac{1}{E_{\text{total}}} \sum_{\text{hits}} E_{\text{hit}}$ if $r_{\text{hit}} < 30$ mm or $r_{\text{hit}} < 60$ mm,
- Shower Radius: $R = \frac{1}{E_{\text{total}}} \sum_{\text{hits}} E_{\text{hit}} \cdot r_{\text{hit}}$ with $r_{\text{hit}} = \sqrt{(x_{\text{hit}} - \text{CoGX})^2 + (y_{\text{hit}} - \text{CoGY})^2}$,
- Shower Variance: $\text{Var}(i) = \frac{1}{E_{\text{total}}} \sum_{\text{hits}} E_{\text{hit}} \cdot (i_{\text{hit}} - \text{CoG}_i)^2$ with $i \in [x, y, z]$,
- Shower Skew: $\text{Skew}(i) = \frac{1}{E_{\text{total}}} \sum_{\text{hits}} E_{\text{hit}} \cdot \left(\frac{i_{\text{hit}} - \text{CoG}_i}{\sigma_i} \right)^3$ with $\sigma_i = \sqrt{\text{Var}(i)}$ and $i \in [x, y, z]$,
- Shower Kurtosis: $\text{Kurt}(i) = \frac{1}{E_{\text{total}}} \sum_{\text{hits}} E_{\text{hit}} \cdot \left(\frac{i_{\text{hit}} - \text{CoG}_i}{\sigma_i} \right)^4$ with $\sigma_i = \sqrt{\text{Var}(i)}$ and $i \in [x, y, z]$,
- Number of hits per event: N_{hits} with $E_{\text{hit}} > 0$ MIP and
- Hit energies: E_{hit} with $E_{\text{hit}} > 0$ MIP.

8. Shower Shape Variables

Since KDEs cannot simulate the whole dataset, i.e. $(24 \times 24)_{\text{tiles}} \times 38_{\text{layers}}$, constraints on the number of tiles and layers had to be introduced. So far, the choice was made to exclude the 18 layers of the detector, while all tiles in the first 20 layers were included. In addition to that, for the analysis of shower shape variables, three alternative configurations were explored too. All configurations investigated are:

1. $(24 \times 24)_{\text{tiles}} \times 20_{\text{layers}}$ (composition used until now)
2. $(16 \times 16)_{\text{tiles}} \times 38_{\text{layers}}$,
3. $(20 \times 20)_{\text{tiles}} \times 25_{\text{layers}}$ and
4. $(20 \times 20)_{\text{tiles}} \times 30_{\text{layers}}$.

It is important to note that, for configurations such as $(20 \times 20)_{\text{tiles}} \times 30_{\text{layers}}$, the first 20×20 tiles, after the distance-ordering algorithm was applied, for each of the first 30 layers were selected. As a result, the chosen tiles in these layers do not necessarily form a rectangle on average.

In this chapter, the results for the composition $(20 \times 20)_{\text{tiles}} \times 30_{\text{layers}}$ are shown. For the other compositions of tiles and layers, the distributions of the shower shape variables for 40 GeV pion dataset are shown in Appendix [B.1](#).

8.2. Gaussian Noise for Neglected Calorimeter Regions

For every shower shape variable, the simulated distributions of the reduced datasets are expected to deviate from the distributions of the full dataset. An example is shown in Figure [8.1](#) for the total energy. The simulation of the reduced dataset is shown in red, whereas the distributions for the full and reduced dataset are shown in dark blue and light blue, respectively. Less tiles per layer and exclusion of the outer tiles per layer reduces the total energy per event, which is visible as a shift of the simulated distribution towards smaller total energies. The $\chi^2/\text{NDF} = 11.50$ supports the expected disagreement.

In order to achieve the desired agreement between the full dataset and simulation, an approach based on the simulation of Gaussian noise is investigated. In this approach, the energy distribution of the excluded tiles in data, as shown in Figure [7.6](#), are examined first on a linear scale, as shown in Figure [8.2](#). It is apparent from the excess of events around zero, that a significant fraction of hits is simulated with $E \simeq 0$ MIP. In order to

8.2. Gaussian Noise for Neglected Calorimeter Regions

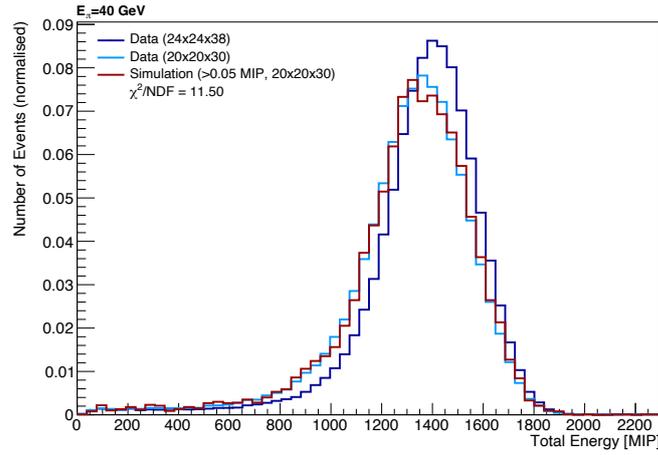


Figure 8.1.: The distribution of the total energy for the $(20 \times 20)_{\text{tiles}} \times 30_{\text{layers}}$ configuration.

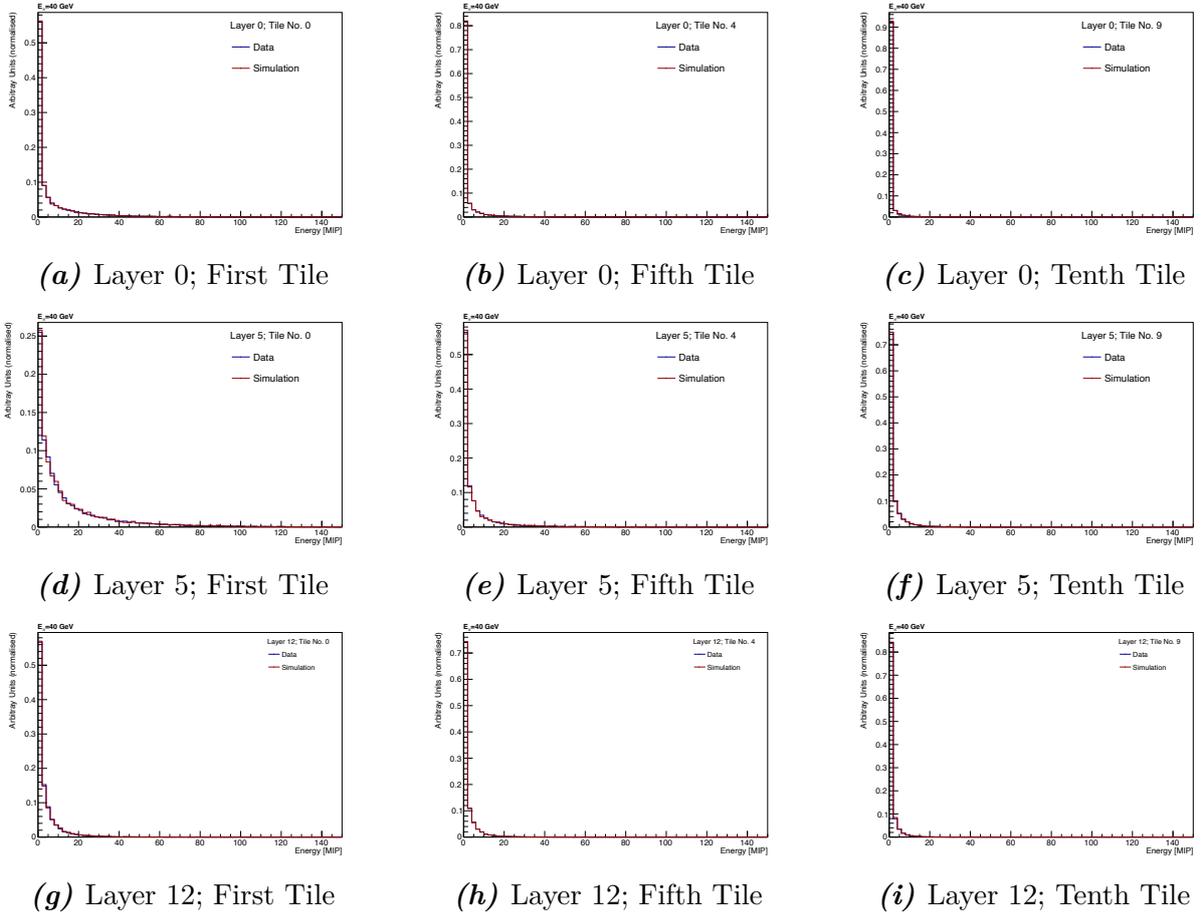


Figure 8.2.: The normalised deposited energy for the first, fifth and tenth closest tile to the event's CoGX/CoGY in the first, sixth and 13th layer (layer number zero, five, and 12, respectively). All simulated energies with a value of less than 0.05 MIP are set to zero.

8. Shower Shape Variables

find the fraction of non-zero hits, the following integral is defined:

$$f_{>0} = 1 - \int \text{bin at zero} = x \sum_{\text{bins}} y_i - y_0. \quad (8.1)$$

Here, y_i is the height of bin i and y_0 is the bin located at zero and x the bin width, which is the same for each bin. The integral $f_{>0}$ serves for the amount of energy deposited in this tile.

The integral is performed on all energy distributions of all tiles that have been neglected so far, for all four configurations. E.g. for the $(20 \times 20)_{\text{tiles}} \times 30_{\text{layers}}$ configuration, the distributions of the last $24 \times 24 - 20 \times 20 = 144$ in the first 30 layers are investigated, in data. Additionally, the integral is similarly calculated for the energy distribution for all tiles of the entire previously unconsidered last eight layers.

Next, a respective mean, μ , and standard deviation, σ , are calculated for all the distributions excluding the zero bin, and Gaussian distributions are created from these parameters. Since the chosen histogram binning for the distributions not only affects the integrals $f_{>0}$ but also the means and standard deviations, the following binning for the energy distributions was selected for the $(20 \times 20)_{\text{tiles}} \times 30_{\text{layers}}$ configuration:

- **Layer 0-29:** for the unconsidered tiles, i.e. the remaining $24 \times 24 - 20 \times 20 = 144$ tiles: 35 bins, ranging from 0 to 50 MIPs.
- **Layer 30-37:**
 - **For the first 20×20 tiles:** 75 bins, ranging from 0 to 150 MIPs, as more energy will be deposited in this region, on average.
 - **For the remaining $24 \times 24 - 20 \times 20 = 144$ tiles:** 35 bins, ranging from 0 to 50 MIPs.

An analogous binning approach was applied to the three other tile-layer compositions.

To assign simulated energy to the previously neglected tiles, 10000 random numbers (one per event) have been generated for each unconsidered tile, each number lying between zero and one. The computed integrals per tile now serve as a threshold: if a randomly generated number is less than or equal to its respective integral $f_{>0}$, another random number greater than zero is drawn from the corresponding Gaussian distribution. This is then the hit energy in the tile for that specific event. Conversely, if the random number is larger than the integral $f_{>0}$, the energy deposition in the tile remains zero for that specific event.

8.3. Distributions of Shower Shape Variables

The objective is to ensure a good agreement between the distribution of the full dataset and the distribution of the simulation for each shower shape variable, as the simulation should accurately approximate the hit energies across the entire calorimeter. However, as mentioned earlier, an exact agreement between the distributions of the full dataset and the simulation for the shower shape variables is not expected. This discrepancy arises because neglecting tiles in the simulation process excludes hit energies, which in turn affects the calculation of these variables. Therefore, hit energies based on Gaussian distributions were included into the simulation, to account for the missing energy. The expectation is that the agreement between the resulting distribution and the distribution of the full dataset for each shower shape variable will further improve. In the following plots, the full dataset is always drawn in dark blue, the reduced dataset in light blue, the simulation without Gaussian noise in dark red, and the simulation with Gaussian noise in dark green.

Total Energy

Figure [8.1](#) shows the total energy distribution per event. As expected, there is a discrepancy between the simulation and the full dataset. However, incorporating Gaussian noise into the missing tiles in the simulation significantly improves the agreement with the full dataset. The simulation with Gaussian noise shifts slightly to higher total energies, with its peak almost aligning with the full dataset. The most noticeable difference between these two histograms occurs in the energy range $[800, 1000]$ GeV, where the inclusion of Gaussian noise did not have a significant impact.

8. Shower Shape Variables

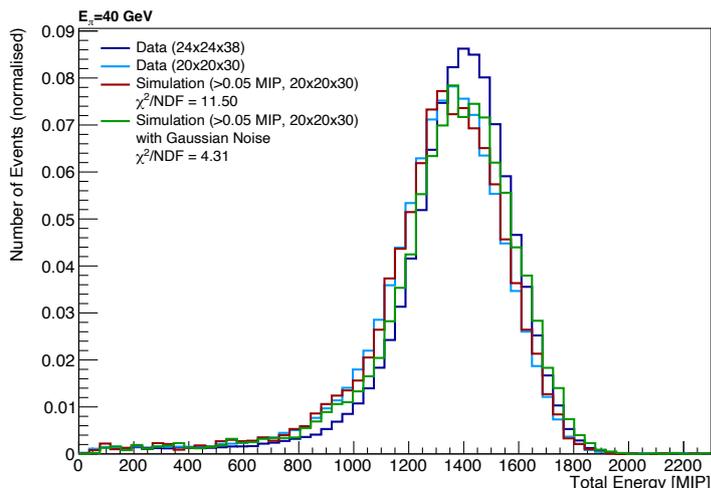


Figure 8.3.: The distribution of the total energy for the $(20 \times 20)_{\text{tiles}} \times 30_{\text{layers}}$ configuration for the full and reduced dataset and simulation with and without Gaussian noise.

Centres of Gravity

The distributions for the CoGs in the x -, y -, and z -directions are shown in Figure 8.4. Here, the CoGs are calculated directly from the hit energies and their position in the calorimeter, whereas the CoGs presented in Chapter 7.3 were simulated via KDEs. The simulation with KDEs had to be done beforehand, to order the tiles according to their distance to the CoG, as a tile number needed to be assigned to the simulated hit energies.

The means for the distribution and the mode for the z -direction, respectively, of the simulated CoGs agree well with data in all three directions and the values listed in Table 7.1. The addition of Gaussian noise slightly improves the agreement between the simulation and the full dataset, which is also reflected by the slight decrease of the χ^2/NDF values. Furthermore, in the z -direction, the simulated CoG distributions show very good agreement with data after including Gaussian noise.

8.3. Distributions of Shower Shape Variables

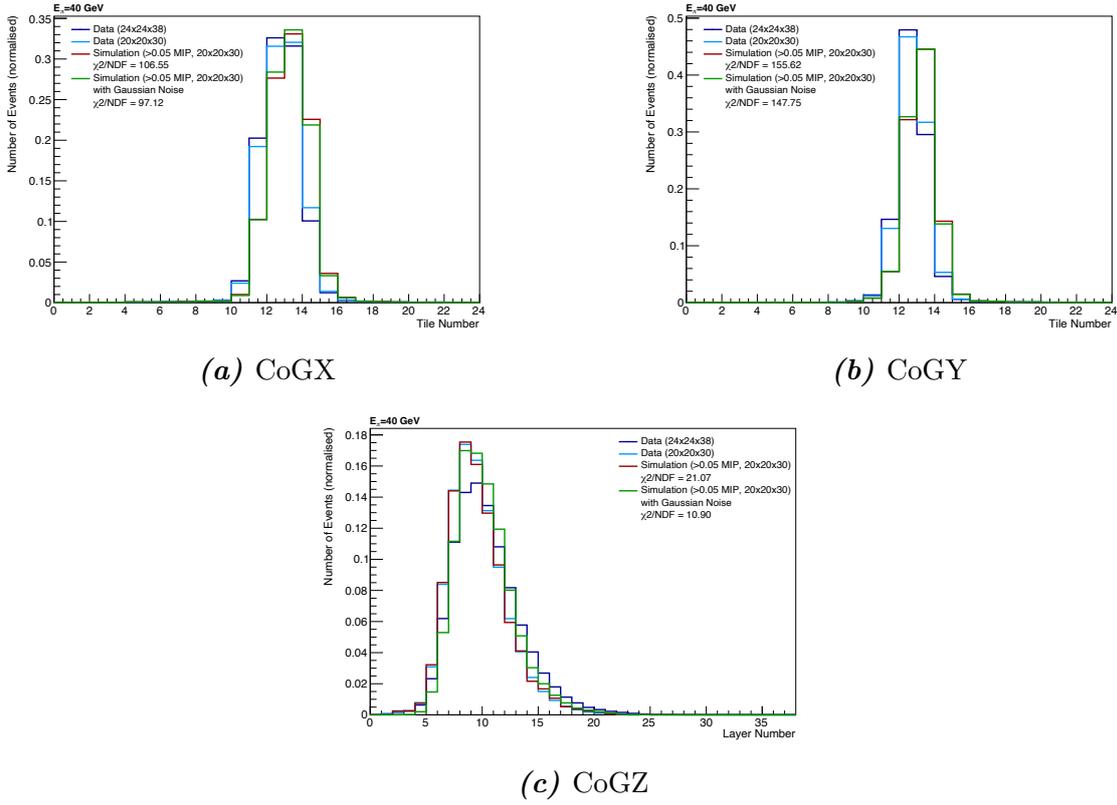


Figure 8.4.: The distribution of CoGs for the $(20 \times 20)_{\text{tiles}} \times 30_{\text{layers}}$ configuration for the full and reduced dataset and simulation with and without Gaussian noise.

Central Fraction

This variable represents the energy in a cylinder with radius of either 30 or 60 mm around the shower axis, as shown in Figure 8.5. On average, 30% of the total energy per event is contained within a 30 mm radius cylinder, while approximately 60% of the total energy is included in cylinder with a 60 mm radius. The central fraction is well modelled for both thresholds. Even without the addition of Gaussian noise, the distributions for the full dataset and simulation align closely, as reflected in the χ^2/NDF test value.

Shower Radius

Excluding outer tiles in the first 30 layers has minimal impact on the distribution of the shower radius from simulation without Gaussian noise, as it remains similar to that of the full dataset, as shown in Figure 8.6. However, there are more simulated events with a smaller radius and fewer events with a larger radius than in the distribution for the full dataset. Taking into account Gaussian noise, the simulated distribution of the shower radius closely matches that of the full dataset in the aspects of the shape and width of the

8. Shower Shape Variables

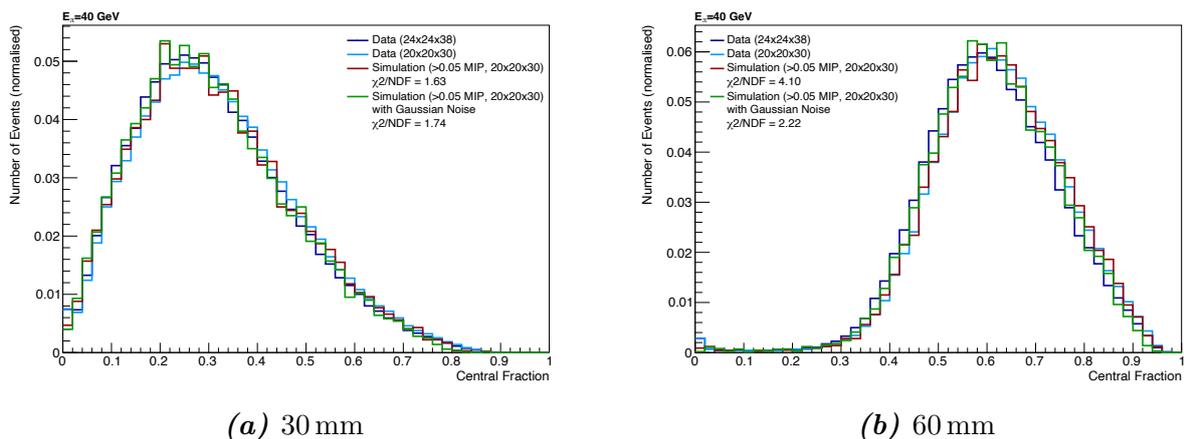


Figure 8.5.: The distribution of the central fraction for (a) $r_{\text{hit}} < 30$ mm and (b) $r_{\text{hit}} < 60$ mm for the $(20 \times 20)_{\text{tiles}} \times 30_{\text{layers}}$ configuration for the full and reduced dataset and simulation with and without Gaussian noise.

distribution and location of the mean. The entire distribution shifts toward the peak of the full dataset, which is located at approximately 75 mm. Additionally, the height of the distribution for simulation with Gaussian noise improves significantly in almost each bin, leading to much better agreement with the full dataset. This improvement also results in a substantial reduction of the χ^2/NDF value, decreasing from nearly 27 to slightly above six, highlighting a more accurate modelling of the shower radius. Nevertheless, the simulation with Gaussian noise still yields fewer events than expected with a radius greater than 85 mm.

Shower Variance

The shower variance describes how widely the energy is distributed around the shower axis along a specific direction. Figure 8.7 indicates that, in the xy -plane, the mean shower variance is approximately five, suggesting that on average, energy deposition occurs within a range slightly less than three tiles from the CoG in this plane. In the z -direction, a significant portion of the energy deposition is concentrated around nine layers from the event's CoGZ. In all three directions, the simulation without Gaussian noise is shifted towards lower variances. Adding Gaussian noise helps aligning the peak of the simulation's distribution with the full dataset. The shower variance in the x - and y -directions matches well, but in the z -direction, there are noticeable differences, especially in the tails of the distribution, as the distribution is narrower than that of the full dataset.

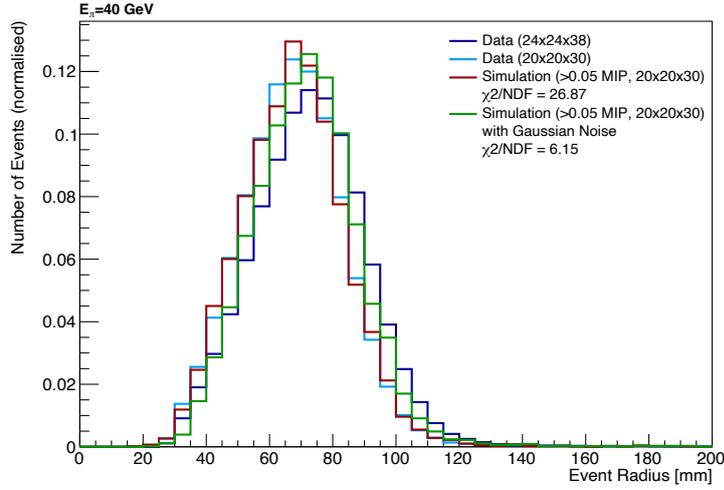


Figure 8.6.: The distribution of the shower radius for the $(20 \times 20)_{\text{tiles}} \times 30_{\text{layers}}$ configuration for the full and reduced dataset and simulation with and without Gaussian noise.

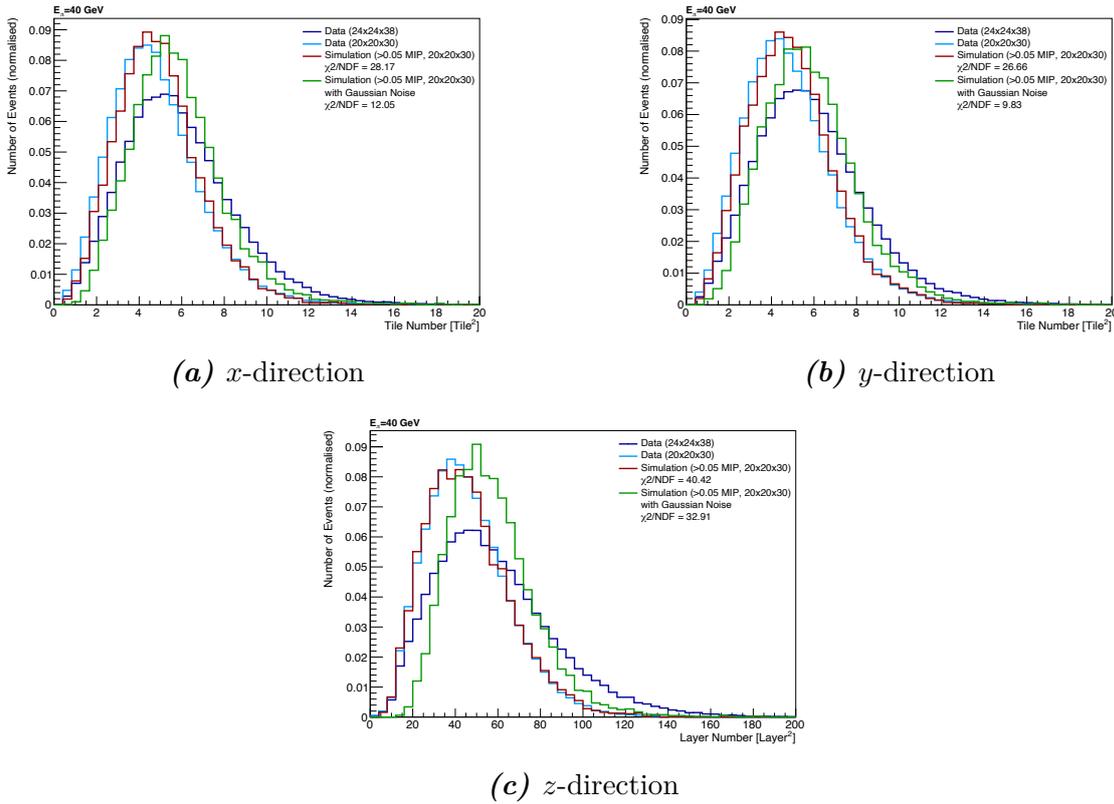


Figure 8.7.: The distribution of the shower variance in all three directions for the $(20 \times 20)_{\text{tiles}} \times 30_{\text{layers}}$ configuration for the full and reduced dataset and simulation with and without Gaussian noise.

8. Shower Shape Variables

Shower Skewness

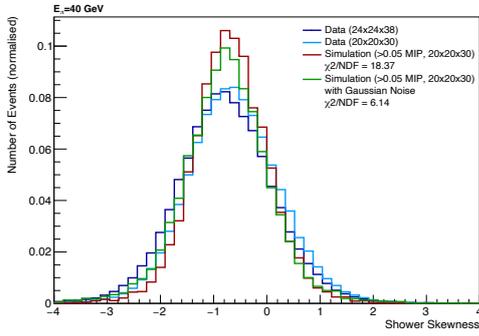
This variable describes the degree of asymmetry of the shower inside the detector. The shower skewness in the x - and y -directions is expected to be zero on average. However, the skewness exhibits predominantly negative values, with an average of approximately -1 , as shown in Figure 8.8. Therefore, the showers are mostly left-skewed. This asymmetry originates from a small beam tilt in the test beam run, where the beam was not perfectly centred on the AHCAL prototype. Conversely, the z -direction exhibits positive skewness, indicating a right-skewed shower in this direction. The distributions in all three figures show good, but not ideal agreement.

In the radial plane, incorporating Gaussian noise improves the distribution, particularly for smaller values of the shower skewness. However, the simulation exhibits a smaller fraction of events with negative skewness compared to data, while a larger fraction of events has a skewness of approximately -1 . For positive skewnesses, the simulation with Gaussian noise shows good agreement with the full dataset. In the z -direction, incorporating Gaussian noise worsens the skewness, particularly for larger values. Furthermore, the distribution of the simulation with Gaussian noise is significantly wider than that of the full dataset. In this case, the simulation without Gaussian noise exhibits better agreement with the distribution of the full dataset than the one with Gaussian noise due to noise being symmetric instead of antisymmetric.

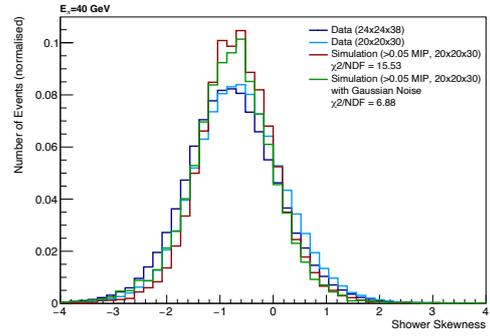
Shower Kurtosis

The kurtosis characterises the steepness and shape of the shower in the calorimeter. As shown in Figure 8.9, it takes on only positive values across all spatial directions, indicating that the showers remain peaked rather than flat. In the x - and y -directions, the distributions peak around a value of six, while in the z -direction, the maximum occurs at approximately three. With the inclusion of Gaussian noise in the previously unconsidered tiles, the distribution aligns almost perfectly with the one from the full dataset, also supported by the χ^2/NDF value: $\chi^2/\text{NDF}_x = 36.14 \rightarrow 3.02$ and $\chi^2/\text{NDF}_y = 24.65 \rightarrow 3.86$. However, in the z -direction, the agreement between the full dataset and the simulation with Gaussian noise deteriorates. The distribution is too broad compared to that of the full dataset, particularly for values starting at approximately 5.5. The distribution without Gaussian noise shows better agreement with the full dataset, although the simulation underestimates the dataset for values greater than or equal to 4.5.

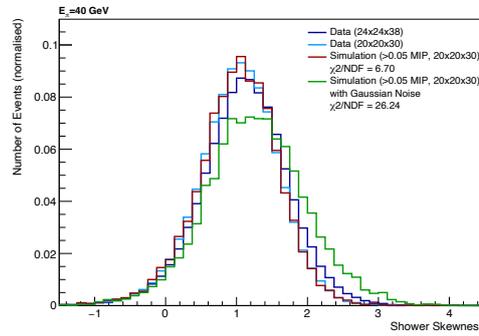
8.3. Distributions of Shower Shape Variables



(a) x -direction

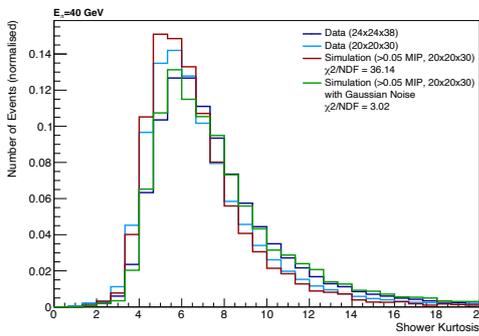


(b) y -direction

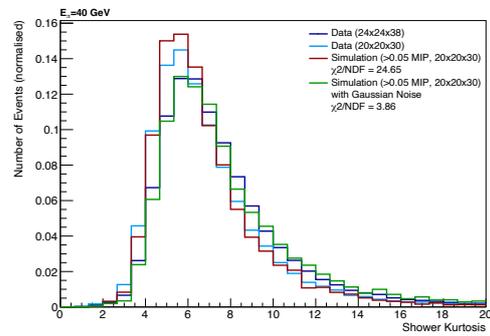


(c) z -direction

Figure 8.8.: The distribution of the shower skewness in all three directions for the $(20 \times 20)_{\text{tiles}} \times 30_{\text{layers}}$ configuration for the full and reduced dataset and simulation with and without Gaussian noise.

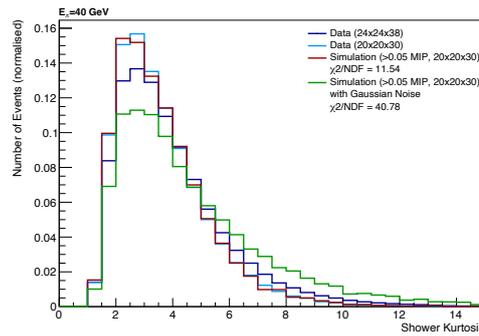


(a) x -direction



(b) y -direction

8. Shower Shape Variables



(c) z -direction

Figure 8.9.: The distribution of the shower kurtosis in all three directions for the $(20 \times 20)_{\text{tiles}} \times 30_{\text{layers}}$ configuration for the full and reduced dataset and simulation with and without Gaussian noise.

Number of Hits

Figure 8.10 presents the distributions of the number of hits per event. Both the simulation with and without Gaussian noise overestimate the number of hits in the shower. This suggests that the simulation using KDEs, generates too many hits above zero, even though a threshold of 0.05 MIP was applied to filter out low-energy depositions. The distribution with additional Gaussian noise hence exceeds the one without even more.

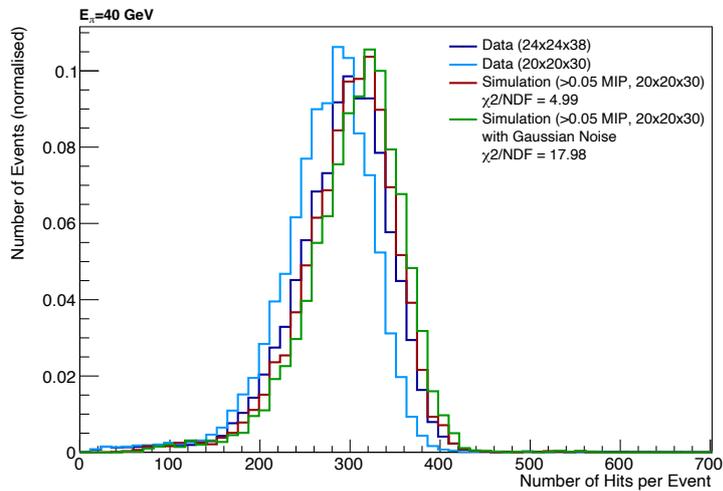


Figure 8.10.: The distribution of the number of hits for the $(20 \times 20)_{\text{tiles}} \times 30_{\text{layers}}$ configuration for the full and reduced dataset and simulation with and without Gaussian noise.

Hit Energies

For the distribution of the hit energies per event, the hit energies larger than zero were considered, as shown in Figure 8.11. This indicates that for data and simulation, most of the energies have values in the lower energetic region, such as the interval $[0, 5]$ MIPs. Including non-zero hit energies randomly generated from a Gaussian distributions does not change the agreement between the full data and simulation significantly. Furthermore, the fraction of low energetic hits in the first bin is larger for simulation than for datasets. This indicates that too many hits smaller than 0.1 MIP exist in the simulation. For larger hit energies, data and simulation are in good agreement.

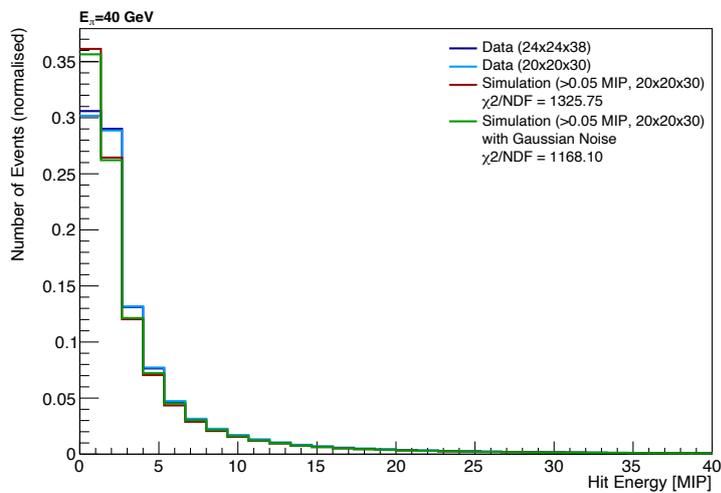


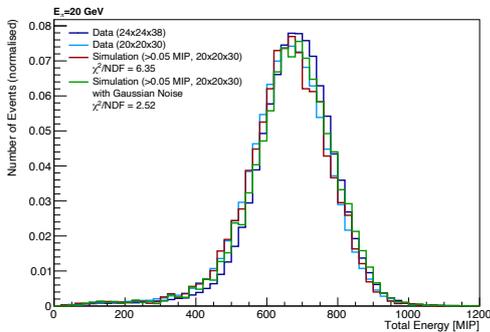
Figure 8.11.: The distribution of hit energies for the $(20 \times 20)_{\text{tiles}} \times 30_{\text{layers}}$ configuration for the full and reduced dataset and simulation with and without Gaussian noise.

9. Results at Different Energies

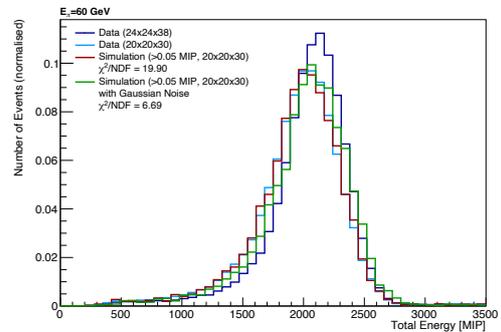
It is of considerable importance to investigate the modelling of the fast simulation algorithm for other pion beam energies. Therefore, this chapter shows the results of the shower shape variables for the pion energies of 20, 60, 80, 120 and 200 GeV for the tile-layer composition $(20 \times 20)_{\text{tiles}} \times 30_{\text{layers}}$. Also, only the distributions for the total energy, central fraction, shower radius, shower variance as well as shower skewness are shown. For the three-dimensional variables, only the results in x - and z -direction are shown, as the distributions in y -direction exhibit similar results to the ones in x -direction. These results and the distributions for the other variables for the $(20 \times 20)_{\text{tiles}} \times 30_{\text{layers}}$ configuration can be found in Appendix [B.2](#).

Total Energy

The distributions of the total energy per event are shown in Figure [9.1](#) for the other datasets. With an increase in beam energy, the total energy also increases. Additionally, the smaller the beam energy, the better the agreement is between the distribution of the full dataset and the simulation without Gaussian noise. This indicates that for lower beam energies, the cutoff at 20×20 tiles for the first 30 layers already includes a significant part of the total deposited energy. Including hit energies for neglected tiles provides a better agreement between data and simulation, as the simulation distributions are shifted towards larger energies.

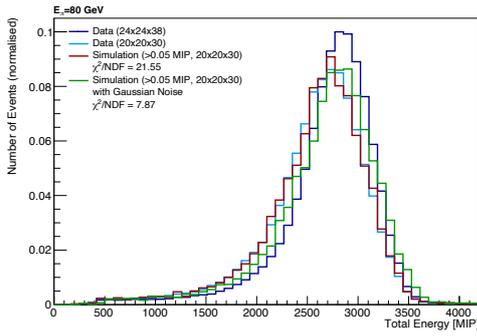


(a) 20 GeV

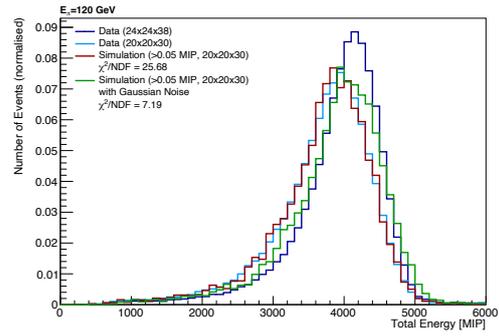


(b) 60 GeV

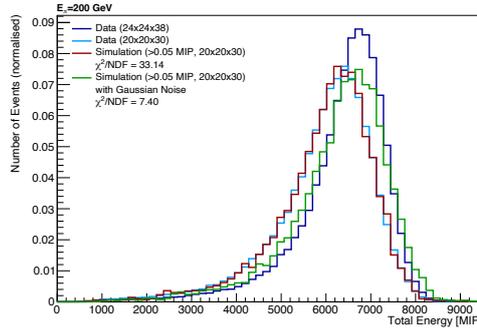
9. Results at Different Energies



(c) 80 GeV



(d) 120 GeV

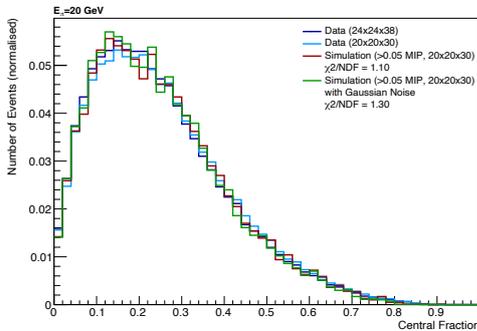


(e) 200 GeV

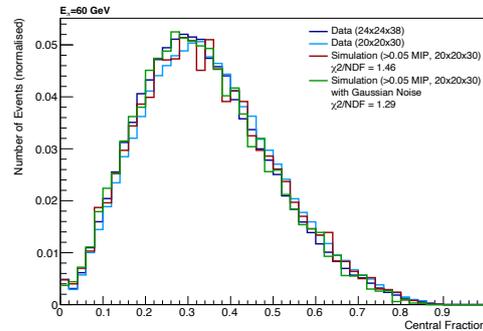
Figure 9.1.: The distribution for the total energy for the $(20 \times 20)_{\text{tiles}} \times 30_{\text{layers}}$ configuration for the full and reduced dataset and simulation with and without Gaussian noise for 20, 60, 80, 120 and 200 GeV pions.

Central Fraction

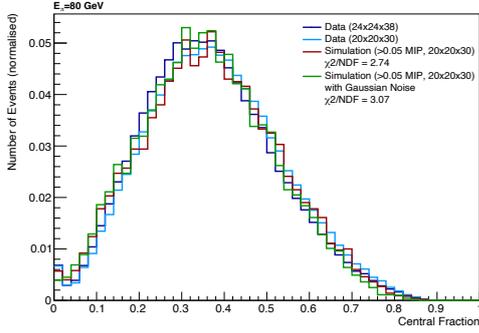
Similar to the distributions for 40 GeV pions, the central fraction exhibits a very good agreement between data and simulation (see Figure 9.2 and 9.3). There are now significant deviations visible. Moreover, for larger pion beam energies, the distributions shift to larger values of the central fraction due to higher beam energies.



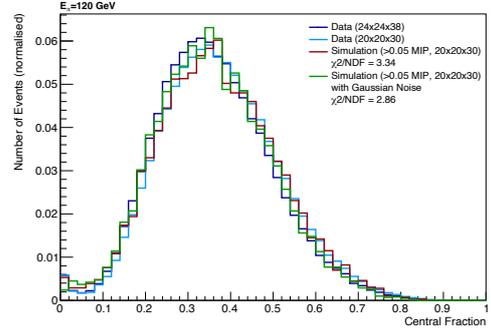
(a) $r_{\text{hit}} < 30$ mm; 20 GeV



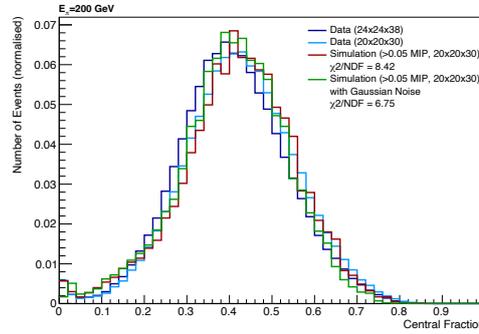
(b) $r_{\text{hit}} < 30$ mm; 60 GeV



(c) $r_{\text{hit}} < 30$ mm; 80 GeV

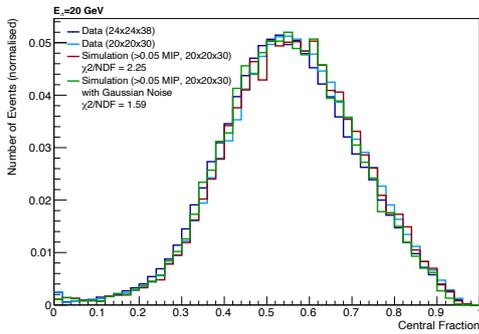


(d) $r_{\text{hit}} < 30$ mm; 120 GeV

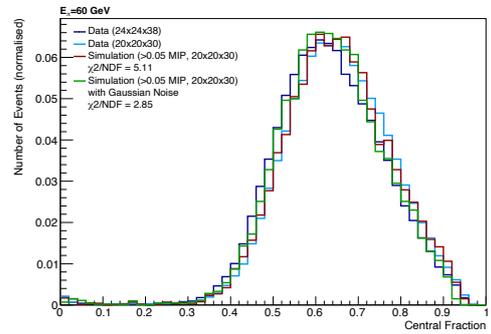


(e) $r_{\text{hit}} < 30$ mm; 200 GeV

Figure 9.2.: The distribution for the central fraction with 30 mm for the $(20 \times 20)_{\text{tiles}} \times 30_{\text{layers}}$ configuration for the full and reduced dataset and simulation with and without Gaussian noise for 20, 60, 80, 120 and 200 GeV pions.

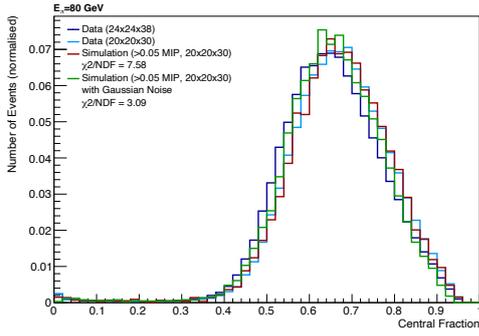


(a) $r_{\text{hit}} < 60$ mm; 20 GeV

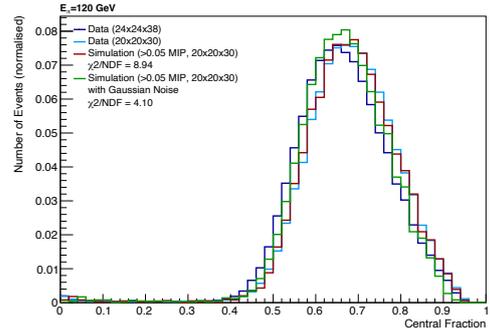


(b) $r_{\text{hit}} < 60$ mm; 60 GeV

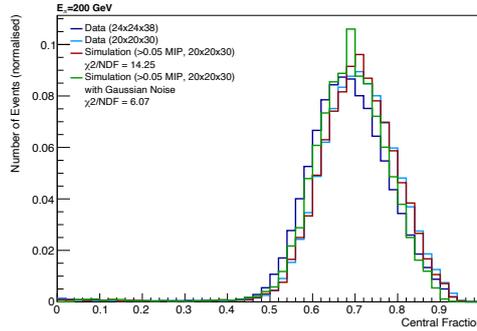
9. Results at Different Energies



(c) $r_{\text{hit}} < 60$ mm; 80 GeV



(d) $r_{\text{hit}} < 60$ mm; 120 GeV

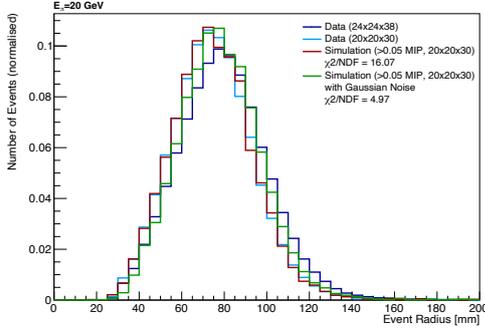


(e) $r_{\text{hit}} < 60$ mm; 200 GeV

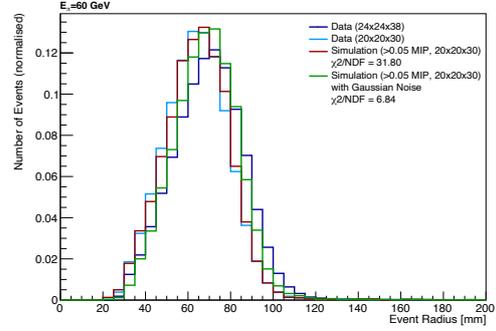
Figure 9.3.: The distribution for the central fraction with 60 mm for the $(20 \times 20)_{\text{tiles}} \times 30_{\text{layers}}$ configuration for the full and reduced dataset and simulation with and without Gaussian noise for 20, 60, 80, 120 and 200 GeV pions.

Shower Radius

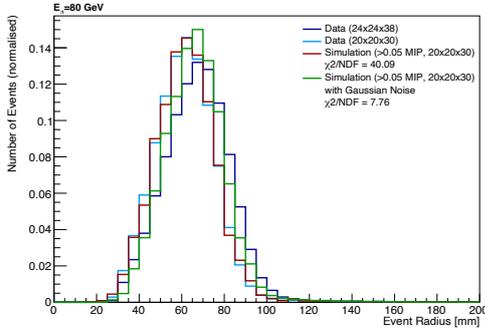
The shower radius becomes smaller as the beam energy increases, as shown in Figure 9.4. Mean values of approximately 80 mm are reached for low energetic beams, whereas the mean of the shower radius of the 200 GeV dataset is 20 mm smaller. Furthermore, the distributions are narrower for higher energies. Without Gaussian noise, the simulation on average takes on values smaller than that of the full dataset. By simulating additional energy via Gaussian noise, the distributions shift to larger radii, achieving a better agreement with the distributions of the full dataset.



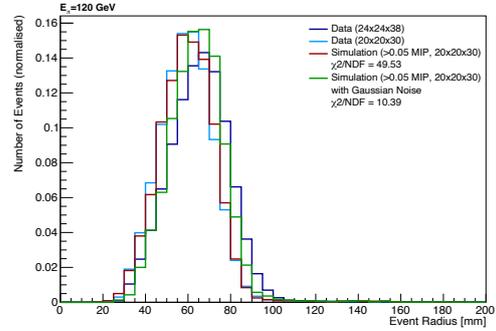
(a) 20 GeV



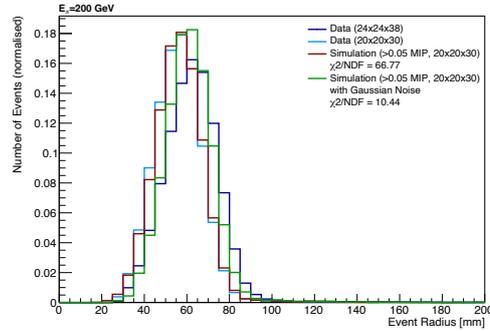
(b) 60 GeV



(c) 80 GeV



(d) 120 GeV



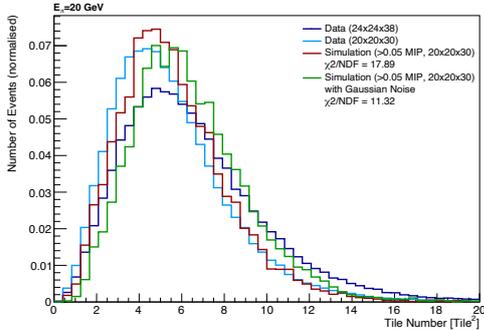
(e) 200 GeV

Figure 9.4.: The distribution for the shower radius for the $(20 \times 20)_{\text{tiles}} \times 30_{\text{layers}}$ configuration for the full and reduced dataset and simulation with and without Gaussian noise for 20, 60, 80, 120 and 200 GeV pions.

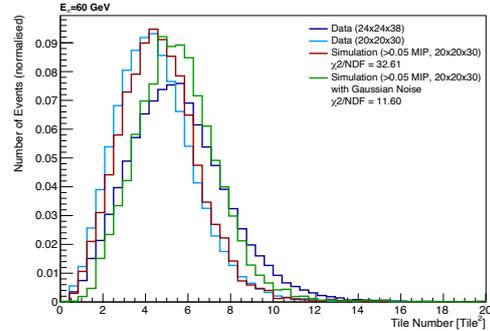
9. Results at Different Energies

Shower Variance

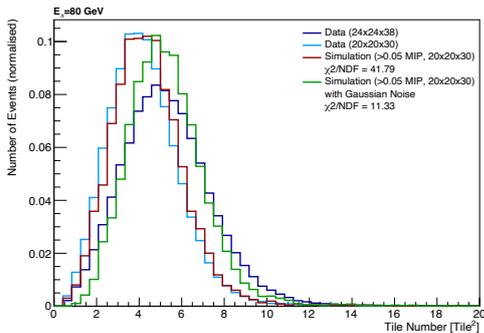
For the shower variance in x -direction, the distributions become narrower and approach a Gaussian distribution with increasing beam energy, as less events have a high variance, which can be seen in Figure 9.5 for each beam energy. The distributions for the simulation agree more with the distributions for the full dataset if Gaussian noise accounts for the missing energy in the neglected calorimeter cells. The peaks and tails of the simulated distributions align more with the peaks of the distributions for the full dataset if Gaussian noise is used. In the z -direction (Figure 9.6), the variance is significantly shifted towards smaller values in simulation than that of the full dataset. Additionally, the simulation distributions are narrower. Although the modelling improves slightly after the addition of Gaussian noise, as the maximum values for the full dataset and simulation align, the tails and overall shape do not.



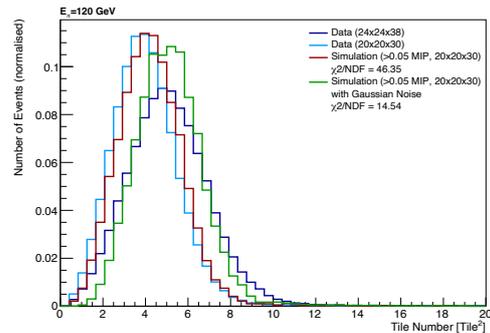
(a) 20 GeV



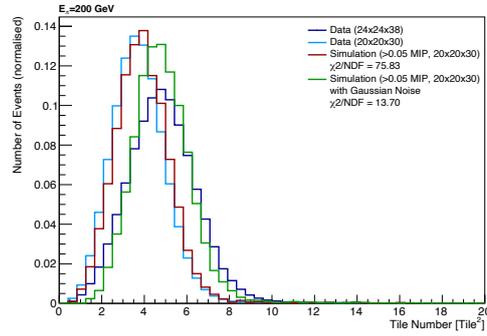
(b) 60 GeV



(c) 80 GeV

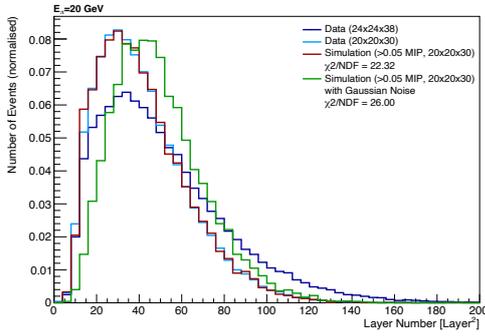


(d) 120 GeV

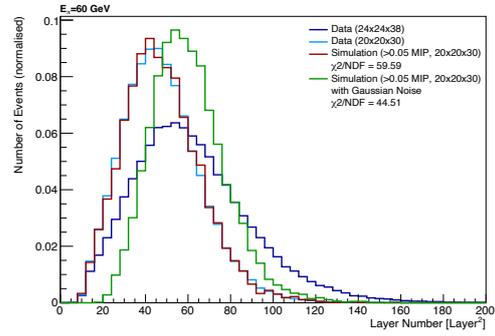


(e) 200 GeV

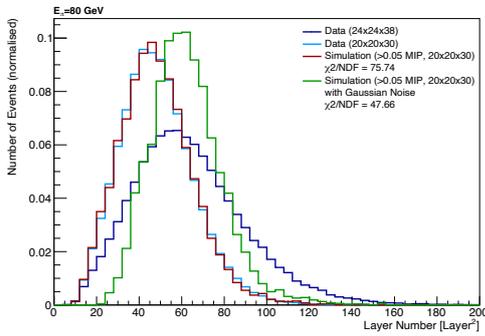
Figure 9.5.: The distribution for the shower variance in x -direction for the $(20 \times 20)_{\text{tiles}} \times 30_{\text{layers}}$ configuration for the full and reduced dataset and simulation with and without Gaussian noise for 20, 60, 80, 120 and 200 GeV pions.



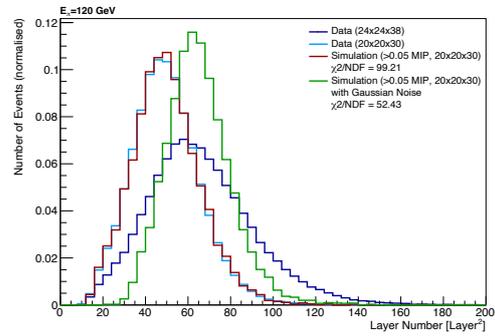
(a) 20 GeV



(b) 60 GeV

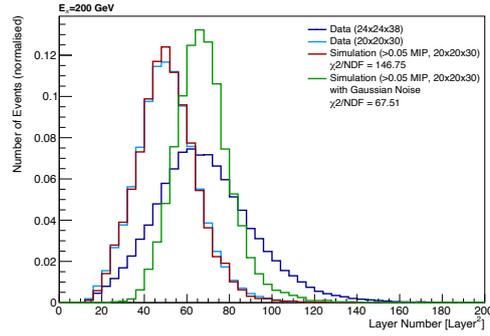


(c) 80 GeV



(d) 120 GeV

9. Results at Different Energies

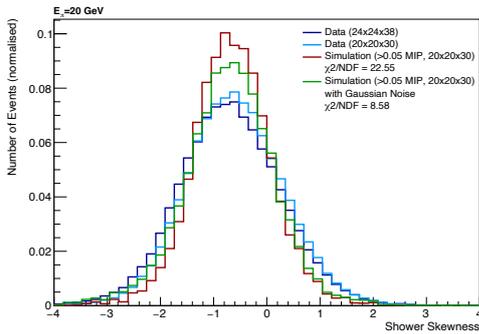


(e) 200 GeV

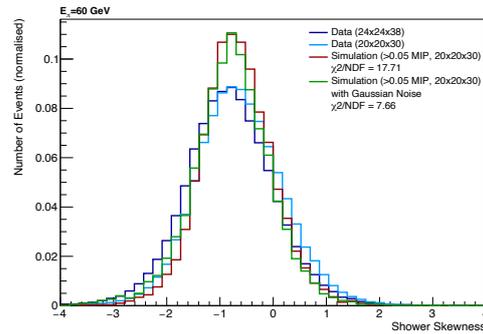
Figure 9.6.: The distribution for the shower variance in z -direction for the $(20 \times 20)_{\text{tiles}} \times 30_{\text{layers}}$ configuration for the full and reduced dataset and simulation with and without Gaussian noise for 20, 60, 80, 120 and 200 GeV pions.

Shower Skewness

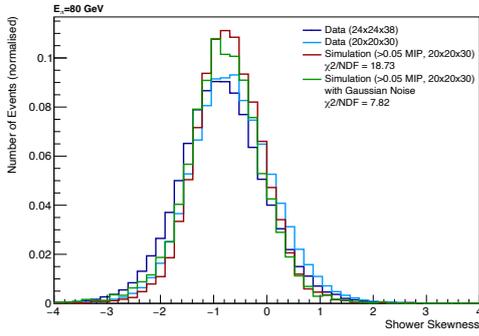
Similar values for the shower skewness are obtained for the other pion energies compared to the 40 GeV dataset: negative skewness in x -direction and positive skewness in z -direction, as seen in Figure 9.7 and 9.8. The distributions in x -direction take on a form with a more distinct peak for higher energies. The agreement between simulation and data improves significantly. However, for 200 GeV pions, the addition of Gaussian noise does not influence the form and shape of the distribution significantly, which is also represented by the χ^2/NDF value. For smaller energies, the distributions for simulation with Gaussian noise in z -direction increases the discrepancy between simulation and data. The addition of Gaussian noise leads to more events with larger values of skewness. However, this phenomena improves again with higher energy.



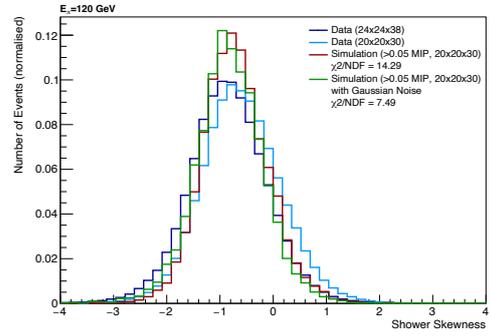
(a) 20 GeV



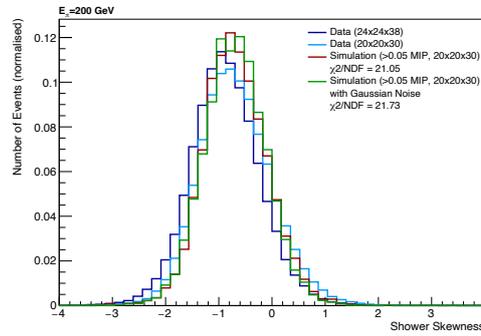
(b) 60 GeV



(c) 80 GeV

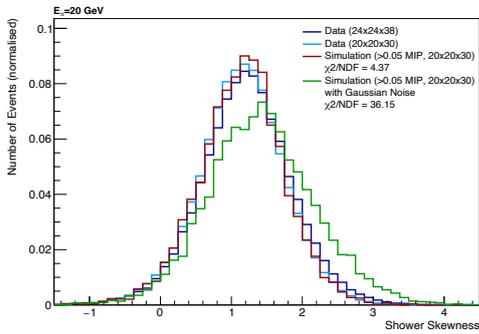


(d) 120 GeV

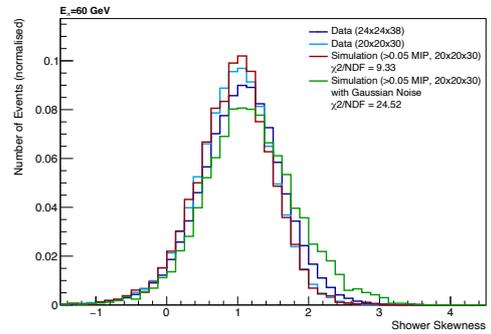


(e) 200 GeV

Figure 9.7.: The distribution for the shower skewness in x -direction for the $(20 \times 20)_{\text{tiles}} \times 30_{\text{layers}}$ configuration for the full and reduced dataset and simulation with and without Gaussian noise for 20, 60, 80, 120 and 200 GeV pions.

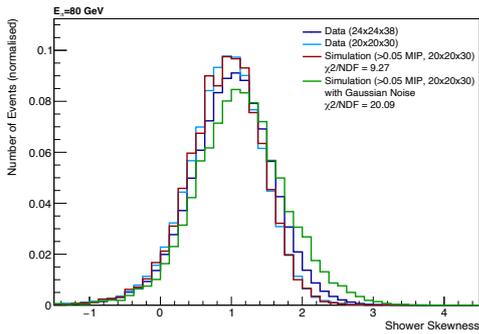


(a) 20 GeV

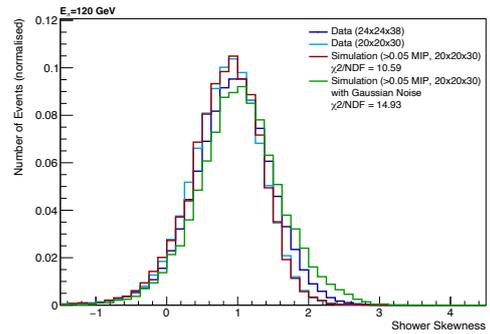


(b) 60 GeV

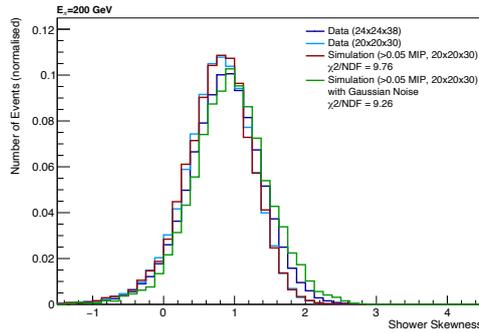
9. Results at Different Energies



(c) 80 GeV



(d) 120 GeV



(e) 200 GeV

Figure 9.8.: The distribution for the shower skewness in z -direction for the $(20 \times 20)_{\text{tiles}} \times 30_{\text{layers}}$ configuration for the full and reduced dataset and simulation with and without Gaussian noise for 20, 60, 80, 120 and 200 GeV pions.

10. Conclusion and Outlook

This thesis investigated a new algorithm for a data-driven fast simulation of hadron showers with test beam data taken with the AHCAL prototype of the CALICE Collaboration at CERN in June 2018. The calorimeter has a total of 38 active layers, with 24×24 readout channels per layer. The pion energies that were used in this thesis are: 20, 40, 60, 80, 120 and 200 GeV. From the datasets, 100000 events were used in the analysis.

The data-based simulation is carried out via KDEs to estimate the PDF of the hit energies sorted by distance to the event's CoG in the xy -plane. A kernel function is assigned to each data point, which is a Gaussian distribution with a mean of zero and a standard deviation of one. The optimal bandwidth parameter of the KDE was found to be $h = 0.01$. 10000 events are extracted from the KDEs for all pion energies.

The approach for the fast simulation is based on the ordering of the calorimeter cells based on their distance to the event's CoG in the xy -plane. This allows to put emphasis on the tiles in the proximity of the CoG, as more energy will be deposited in this region, and assigns a lower importance to cells farther away, where, on average, less energy is deposited. Moreover, this approach avoids the geometrical transformation of the calorimeter tiles, and consequently of the hit energies. The output of the KDEs is a set of sorted simulated hit energies. As the KDEs cannot simulate the whole dataset, only the hit energies within the first 20 layers were considered in the beginning. At a later stage, various configurations of tiles and layers were introduced to balance information loss due to neglecting outer tiles and layers.

To assign a tile and layer number to the simulated hit energies, the CoGs in all three spatial directions were simulated with KDEs. A good agreement between the distributions of the simulated CoGs and the distributions of the CoGs from data is observable.

From the simulated CoGs, the distances to the calorimeter cells were calculated, and the tiles were subsequently sorted in ascending order. In the last step, the ordered tiles were assigned to the ordered hit energies obtained from the simulation with KDEs. The energy distributions for the sorted tiles showed very good agreement between simulation and data. A larger energy deposition is expected in the first tiles, as the first tiles correspond to the tiles closest to the event's CoG. The larger the distance to the CoG, the less energy

10. Conclusion and Outlook

is deposited and this is reflected in the energy distributions per tile. Additionally, the hit energies per tile meet the expectation of average hadronic shower development: at first, the energy depositions takes place in a small region in the detector. As the particles propagate and deposit more energy, more tiles are affected, the shower spreads out, and an increase in energy deposition can be observed. Reaching the outer region of the detector, more tiles in the outer region of the layers are affected, while less energy will be deposited in the outer layers due to the reduction of energy deposition via the particles. This phenomenon is modelled by the simulation.

Due to the limitations of the KDEs, the whole dataset with its almost 22000 readout channels cannot be used as an input for the KDEs. Therefore, tiles and layers have been neglected, in order to reduce the total number of input values that need to be simulated. Here, the focus lied on the closest 20×20 tiles to the CoGX/CoGY for the first 30 layers of the detector. This reduced dataset was used as the base of the simulation via KDEs. The last 144 tiles in the first 30 layers and the outer 8 layers were neglected for now. As the neglect of tiles and layers leads to a reduction of the total energy deposition per event, the full dataset could not be replicated.

To account for missing energy, Gaussian noise was simulated using the energy distribution of neglected tiles. The integral of these distributions without the bin at zero was computed, and the mean and standard deviation were determined. If a random number was equal or smaller than the integral, a Gaussian-sampled non-zero value was assigned to the tile; otherwise, zero energy was assigned.

The performance of this approach was determined with shower shape variables. In most cases, the distributions for simulation with Gaussian noise were in good agreement with the distributions of the full dataset. Therefore, this approach showed good performance in replicating the shape of the hadronic showers. However, since the distance ordering focuses only on the distance to the CoG in the xy -plane, limitations were observed for the shower shape variables in the z -direction. The simulation with and without Gaussian noise cannot replicate the distributions for the full dataset in this direction. For larger beam energies, the discrepancy gradually vanishes between the full dataset and simulation. In contrast, only the shower variance in z -direction exhibits a better agreement between simulation with Gaussian noise and data at smaller pion energies.

To improve the modelling of the kinematic variables in the z -direction, correlations between hit energies from Gaussian distributions and the simulated hit energies via KDEs could be included in the simulation, which have so far been neglected. Furthermore, the hit energies from the Gaussian distributions have also been generated without internal correlations. Thus, the inclusion of correlations could result in a better agreement between

the distributions of the full dataset and the simulation. Additionally, a comparison with other fast simulation algorithms must be realised, but also with the full simulation. This will help to compare the performance of this distance-based fast simulation with other fast simulation approaches and the full simulation.

Fast simulations of hadronic showers will become increasingly vital as complexity and quantity of data will increase in the future of high-energy physics. More data could lead to answers for the open questions in particle physics. As detectors improve, such as with the development of highly granular calorimeters, and as the amount of data will grow steadily, it is becoming more difficult to process this data quantity. Therefore, the investigation and utilisation of efficient simulation methods will become increasingly important.

A. Centres of Gravity

A.1. Centres of Gravity - 3D

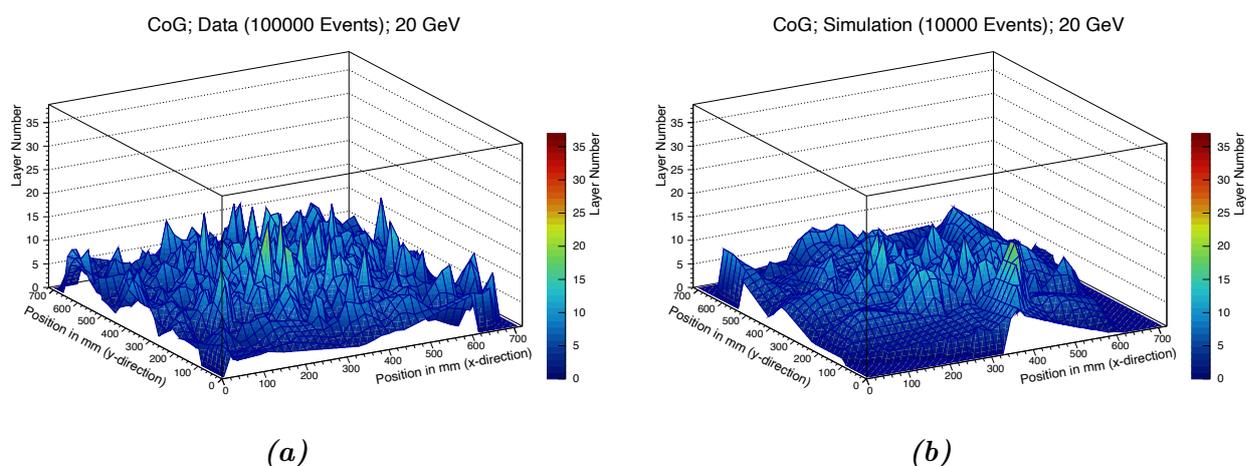


Figure A.1.: The distributions of the CoGs in all three directions for (a) data and (b) simulation with KDEs for 20 GeV pions.

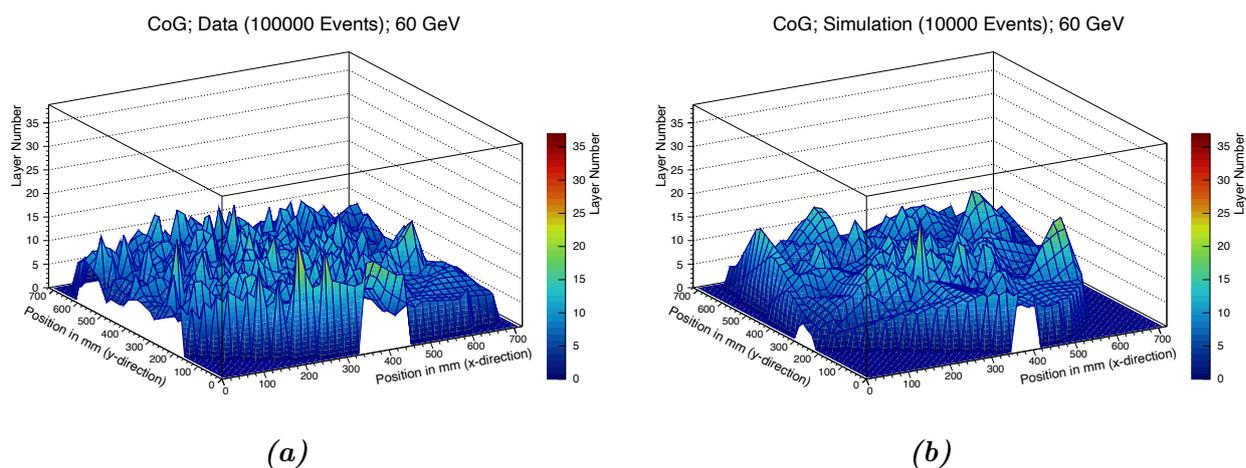


Figure A.2.: The distributions of the CoGs in all three directions for (a) data and (b) simulation with KDEs for 60 GeV pions.

A. Centres of Gravity

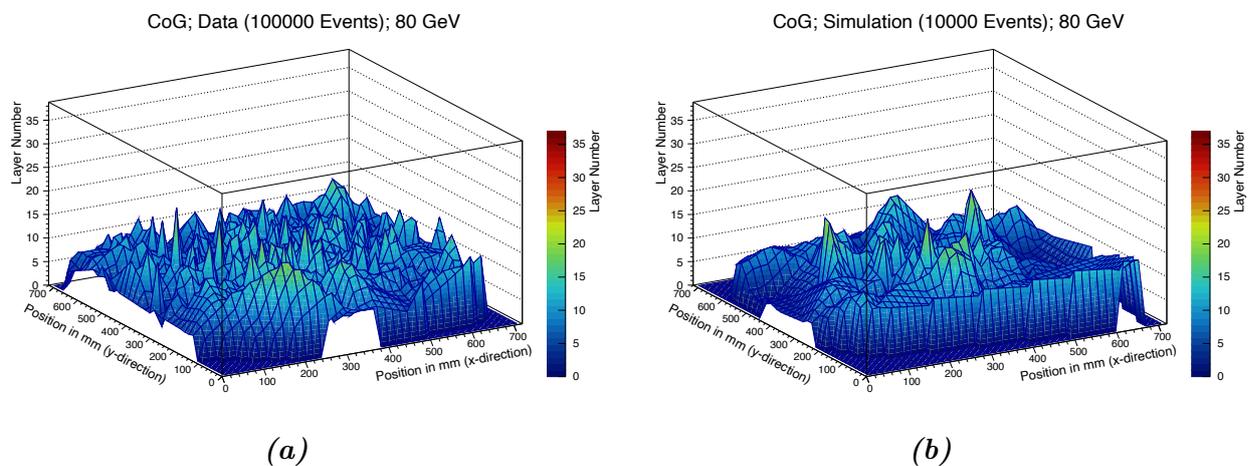


Figure A.3.: The distributions of the CoGs in all three directions for (a) data and (b) simulation with KDEs for 80 GeV pions.

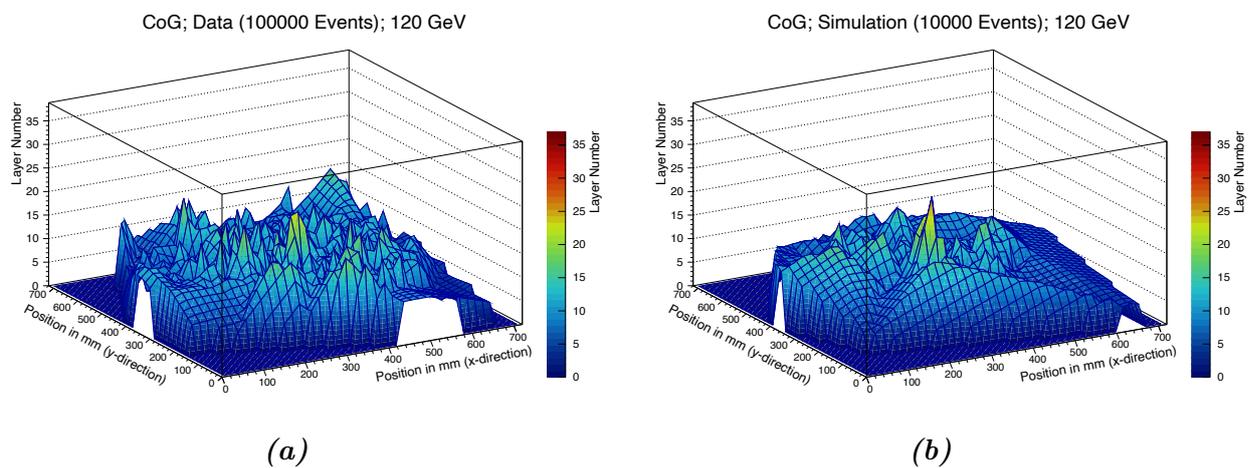


Figure A.4.: The distributions of the CoGs in all three directions for (a) data and (b) simulation with KDEs for 120 GeV pions.

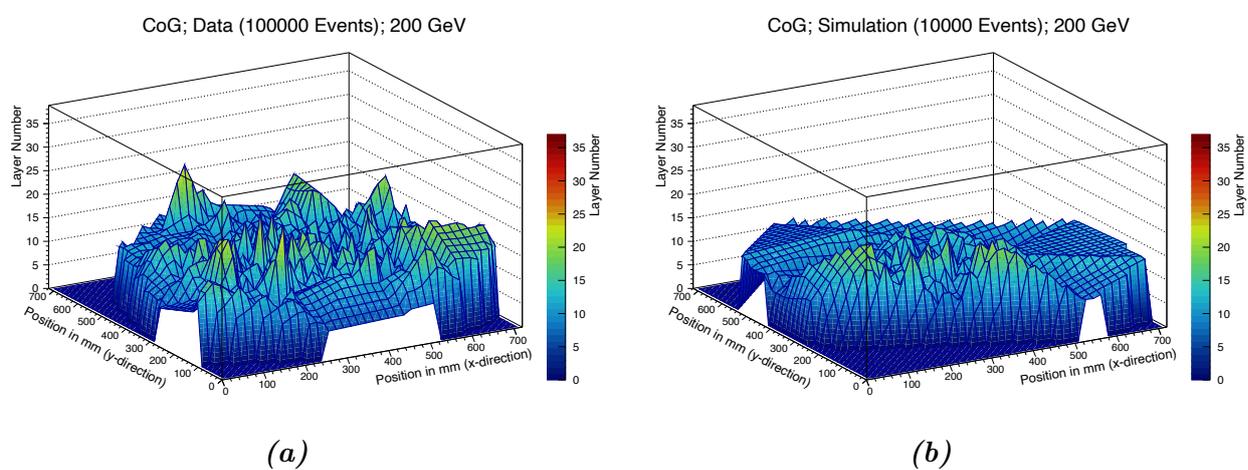
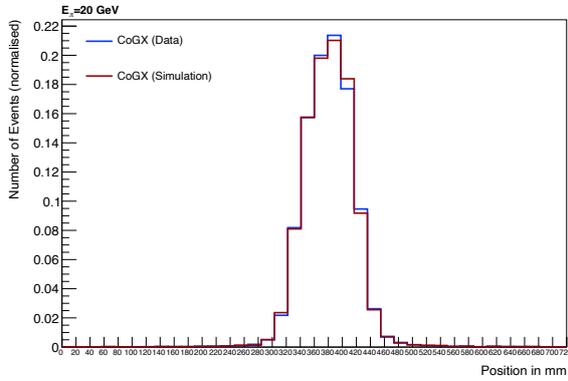


Figure A.5.: The distributions of the CoGs in all three directions for (a) data and (b) simulation with KDEs for 200 GeV pions.

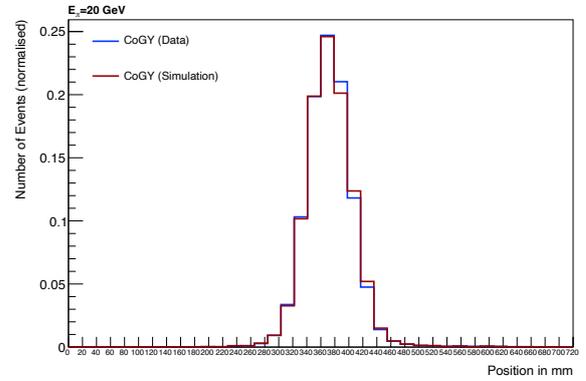
A.2. Distributions of the Centres of Gravity

Table A.1.: The mean values and standard deviations of the CoGs in each direction for 20 GeV dataset and respective simulation.

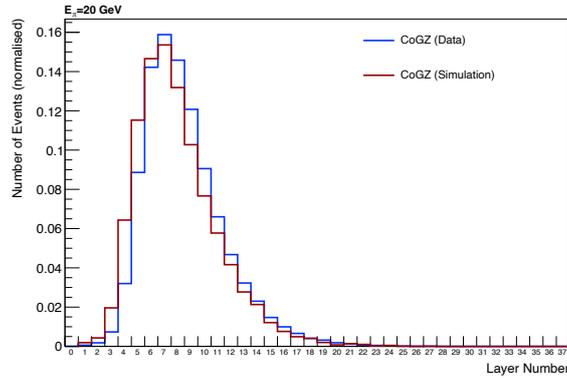
20 GeV	CoGX [mm]	CoGY [mm]	CoGZ [layer]
Data	381 ± 38	373 ± 35	8 ± 3
Simulation	382 ± 38	373 ± 36	7 ± 3
	CoGX [tile]	CoGY [tile]	CoGZ [layer]
Data	13 ± 2	13 ± 2	8 ± 3
Simulation	13 ± 2	13 ± 2	7 ± 3



(a) CoGX



(b) CoGY

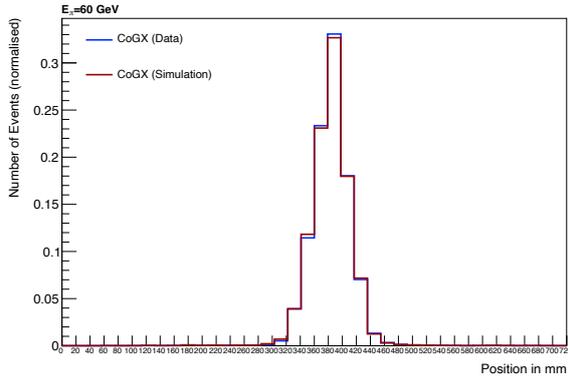


(c) CoGZ

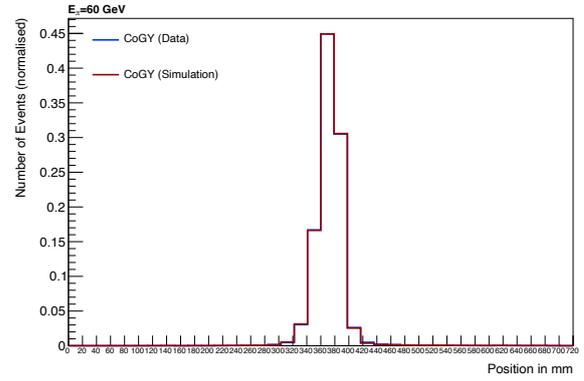
Figure A.6.: The normalised distributions for the simultaneously simulated CoGs with KDEs for 20 GeV pions.

Table A.2.: The mean values and standard deviations of the CoGs in each direction for 60 GeV dataset and respective simulation.

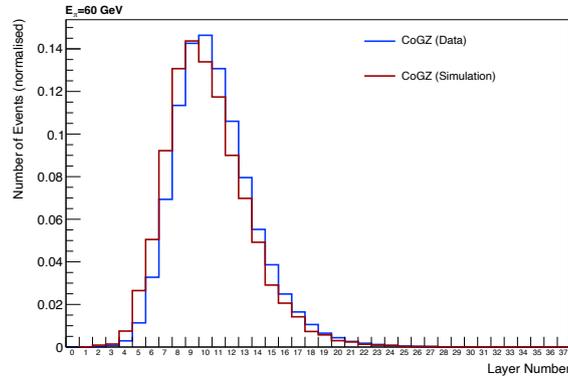
60 GeV	CoGX [mm]	CoGY [mm]	CoGZ [layer]
Data	384 ± 30	373 ± 23	10 ± 3
Simulation	384 ± 30	373 ± 23	10 ± 3
	CoGX [tile]	CoGY [tile]	CoGZ [layer]
Data	13 ± 2	13 ± 2	10 ± 3
Simulation	13 ± 1	13 ± 1	10 ± 3



(a) CoGX



(b) CoGY



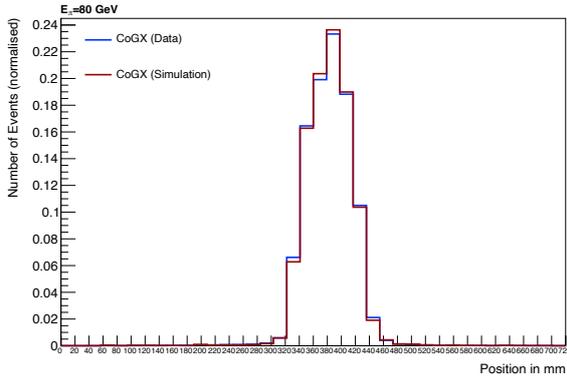
(c) CoGZ

Figure A.7.: The normalised distributions for the simultaneously simulated CoGs with KDEs 60 GeV pions.

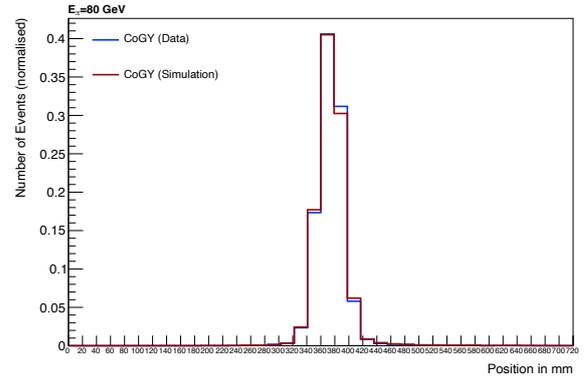
A. Centres of Gravity

Table A.3.: The mean values and standard deviations of the CoGs in each direction for 80 GeV dataset and respective simulation.

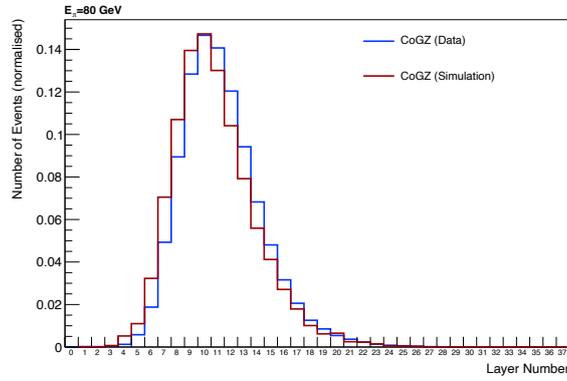
80 GeV	CoGX [mm]	CoGY [mm]	CoGZ [layer]
Data	383 ± 34	375 ± 24	10 ± 3
Simulation	383 ± 34	375 ± 25	10 ± 3
	CoGX [tile]	CoGY [tile]	CoGZ [layer]
Data	13 ± 2	13 ± 1	10 ± 3
Simulation	13 ± 2	13 ± 1	10 ± 3



(a) CoGX



(b) CoGY

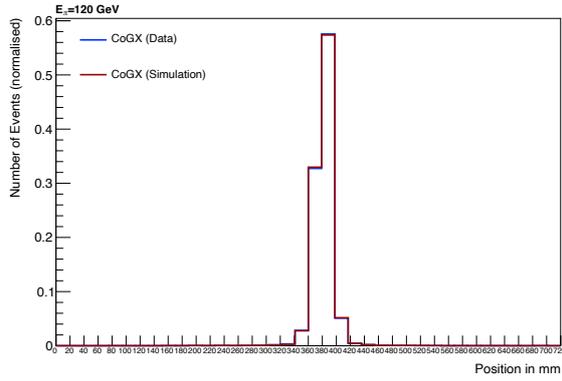


(c) CoGZ

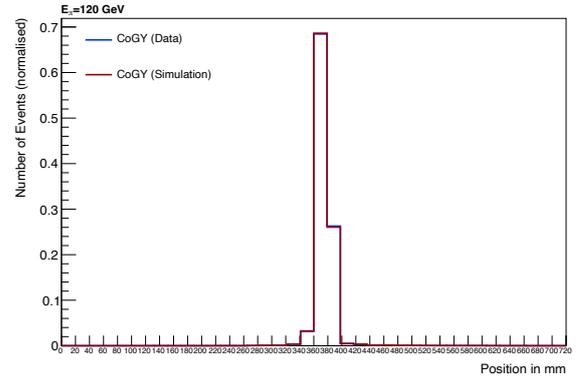
Figure A.8.: The normalised distributions for the simultaneously simulated CoGs with KDEs 80 GeV pions.

Table A.4.: The mean values and standard deviations of the CoGs in each direction for 120 GeV dataset and respective simulation.

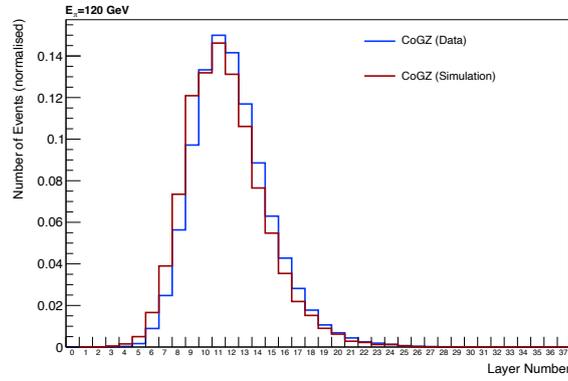
120 GeV	CoGX [mm]	CoGY [mm]	CoGZ [layer]
Data	382 ± 17	375 ± 15	11 ± 3
Simulation	382 ± 17	375 ± 16	11 ± 3
	CoGX [tile]	CoGY [tile]	CoGZ [layer]
Data	13 ± 1	13 ± 1	11 ± 3
Simulation	13 ± 1	13 ± 1	11 ± 3



(a) CoGX



(b) CoGY



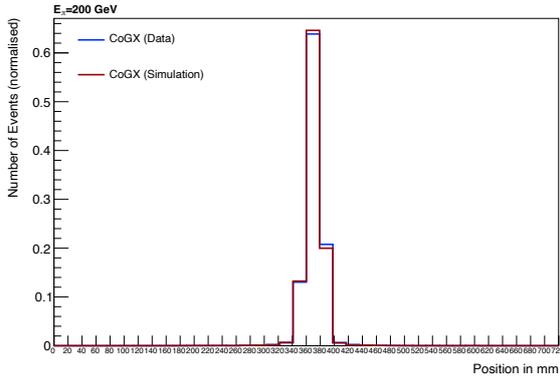
(c) CoGZ

Figure A.9.: The normalised distributions for the simultaneously simulated CoGs with KDEs 120 GeV pions.

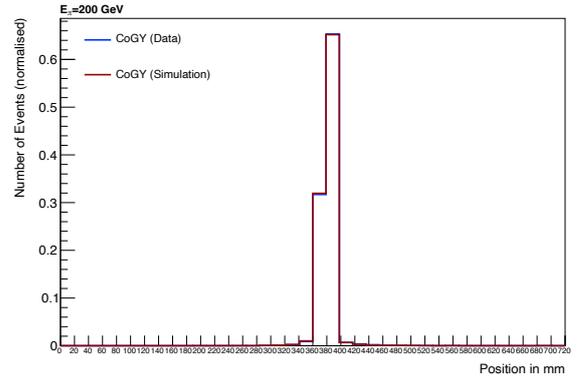
A. Centres of Gravity

Table A.5.: The mean values and standard deviations of the CoGs in each direction for 200 GeV dataset and respective simulation.

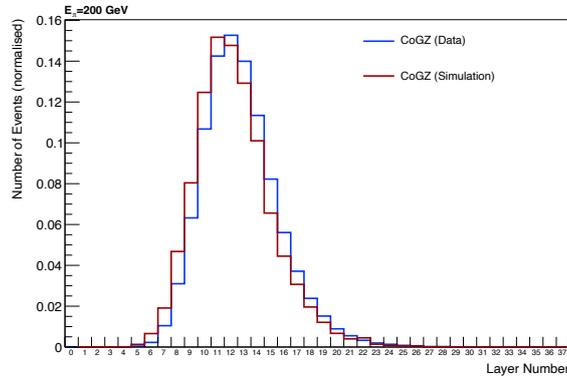
200 GeV	CoGX [mm]	CoGY [mm]	CoGZ [layer]
Data	371 ± 15	381 ± 12	12 ± 2
Simulation	370 ± 14	381 ± 12	12 ± 2
	CoGX [tile]	CoGY [tile]	CoGZ [layer]
Data	13 ± 1	13 ± 1	12 ± 2
Simulation	13 ± 1	13 ± 1	12 ± 2



(a) CoGX



(b) CoGY



(c) CoGZ

Figure A.10.: The normalised distributions for the simultaneously simulated CoGs with KDEs 200 GeV pions.

B. Shower Shape Variables

B.1. Other Compositions - 40 GeV Pions

Total Energy

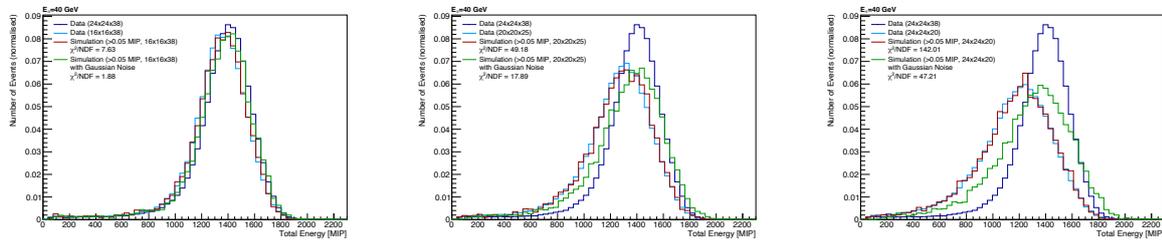


Figure B.1.: The distribution for the total energy for the other three configuration for the full and reduced dataset and simulation with and without Gaussian noise for 40 GeV pions.

Centre of Gravity

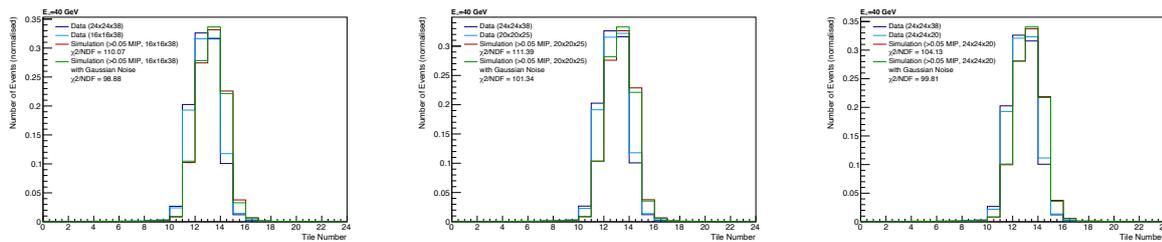


Figure B.2.: The distribution for the CoGX for the other three configuration for the full and reduced dataset and simulation with and without Gaussian noise for 40 GeV pions.

B. Shower Shape Variables

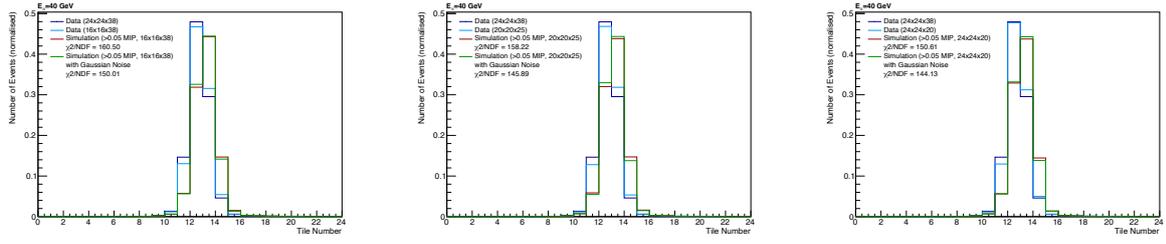


Figure B.3.: The distribution for the CoGY for the other three configuration for the full and reduced dataset and simulation with and without Gaussian noise for 40 GeV pions.

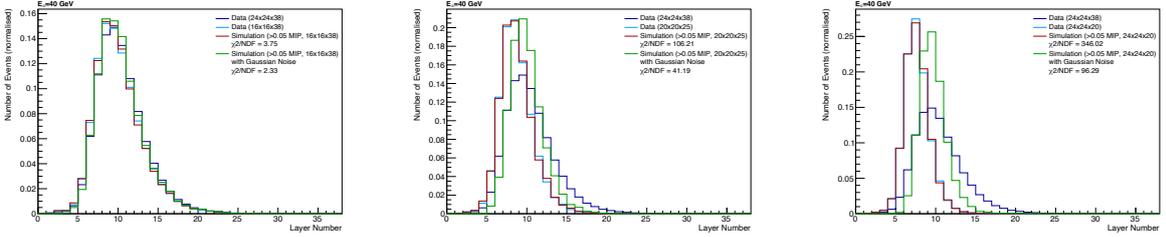


Figure B.4.: The distribution for the CoGZ for the other three configuration for the full and reduced dataset and simulation with and without Gaussian noise for 40 GeV pions.

Central Fraction

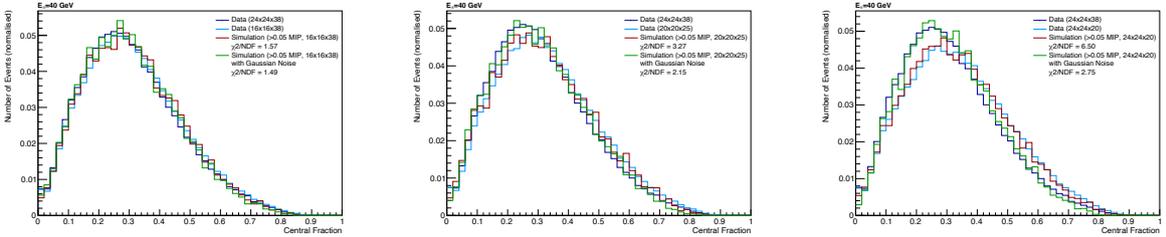


Figure B.5.: The distribution for the central fraction for $r_{\text{hit}} < 30$ mm for the other three configuration for the full and reduced dataset and simulation with and without Gaussian noise for 40 GeV pions.

B.1. Other Compositions - 40 GeV Pions

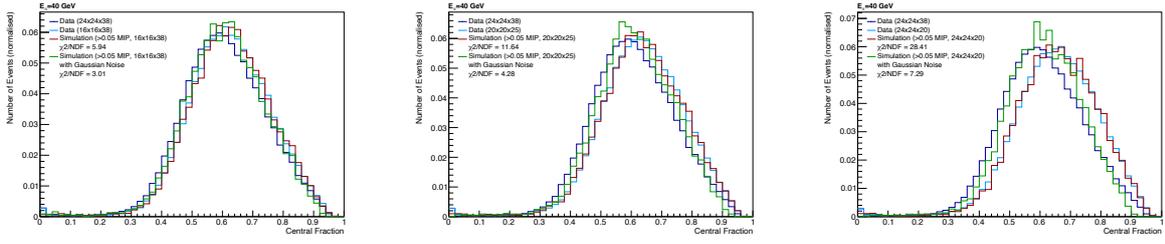


Figure B.6.: The distribution for the central fraction for $r_{\text{hit}} < 60$ mm for the other three configuration for the full and reduced dataset and simulation with and without Gaussian noise for 40 GeV pions.

Shower Radius

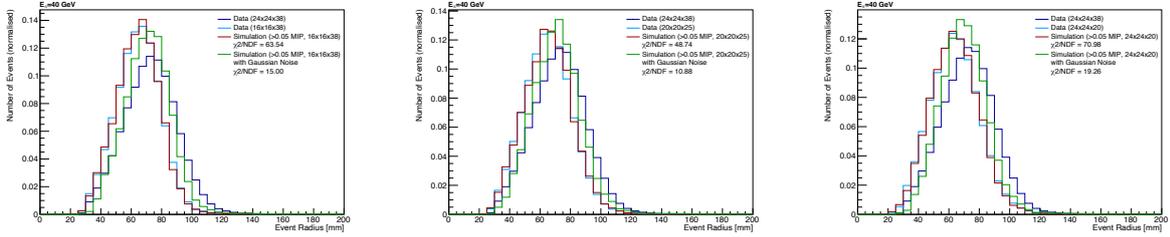


Figure B.7.: The distribution for the shower radius for the other three configuration for the full and reduced dataset and simulation with and without Gaussian noise for 40 GeV pions.

Shower Variance

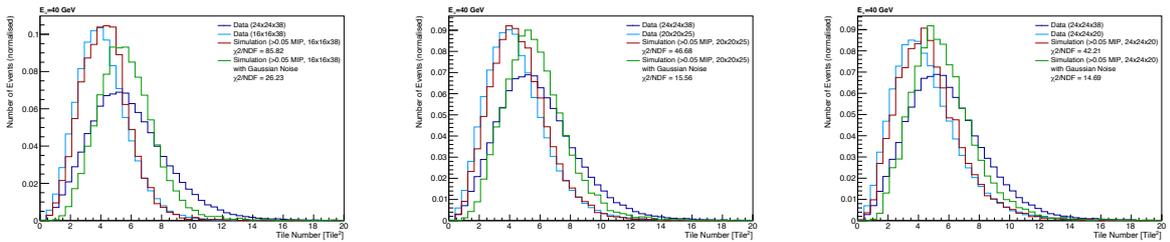


Figure B.8.: The distribution for the shower variance in x -direction for the other three configuration for the full and reduced dataset and simulation with and without Gaussian noise for 40 GeV pions.

B. Shower Shape Variables

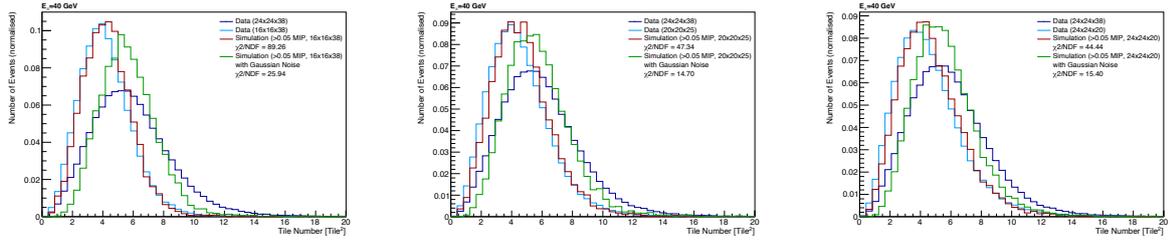


Figure B.9.: The distribution for the shower variance in y -direction for the other three configuration for the full and reduced dataset and simulation with and without Gaussian noise for 40 GeV pions.

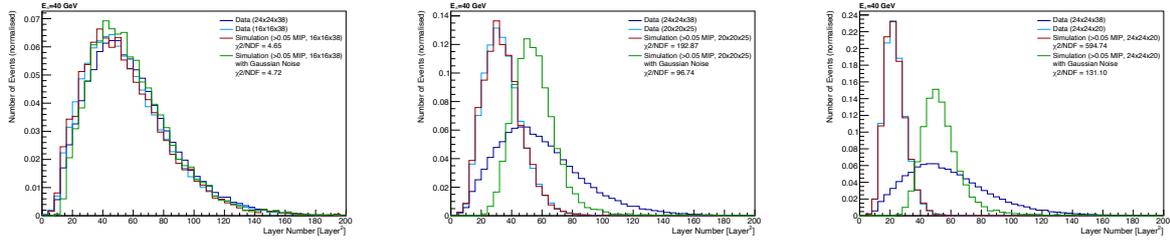


Figure B.10.: The distribution for the shower variance in z -direction for the other three configuration for the full and reduced dataset and simulation with and without Gaussian noise for 40 GeV pions.

Shower Skewness

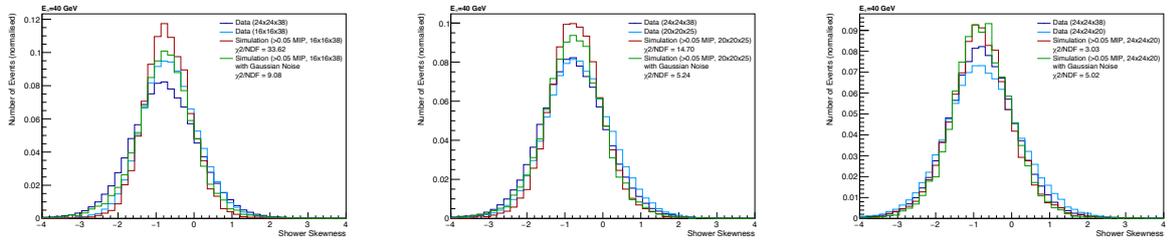


Figure B.11.: The distribution for the shower skewness in x -direction for the other three configuration for the full and reduced dataset and simulation with and without Gaussian noise for 40 GeV pions.

B.1. Other Compositions - 40 GeV Pions

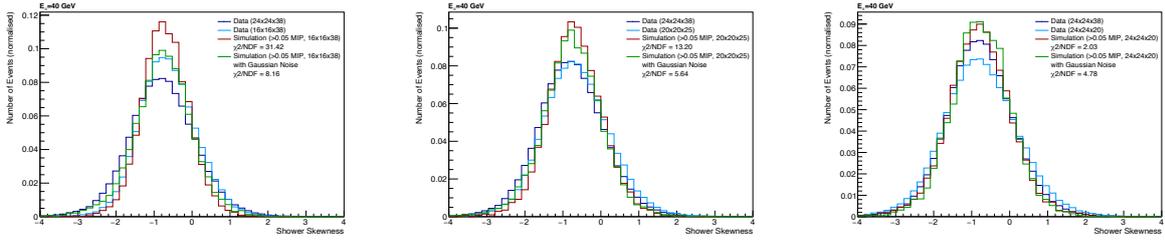


Figure B.12.: The distribution for the shower skewness in y -direction for the other three configuration for the full and reduced dataset and simulation with and without Gaussian noise for 40 GeV pions.

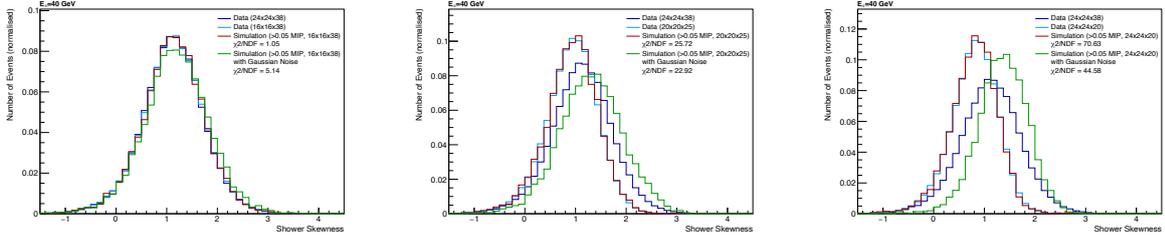


Figure B.13.: The distribution for the shower skewness in z -direction for the other three configuration for the full and reduced dataset and simulation with and without Gaussian noise for 40 GeV pions.

Shower Kurtosis

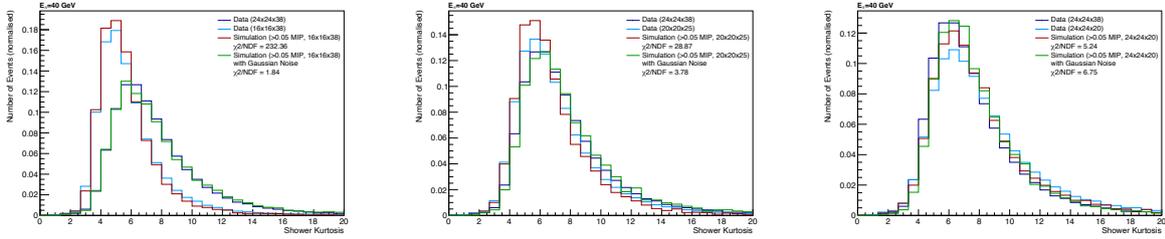


Figure B.14.: The distribution for the shower kurtosis in x -direction for the other three configuration for the full and reduced dataset and simulation with and without Gaussian noise for 40 GeV pions.

B. Shower Shape Variables

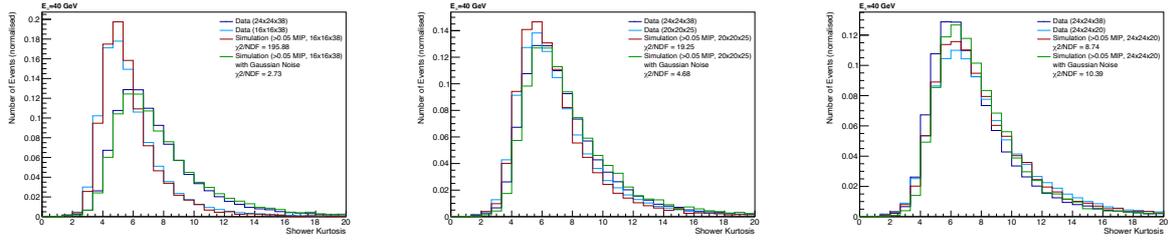


Figure B.15.: The distribution for the shower kurtosis in y -direction for the other three configuration for the full and reduced dataset and simulation with and without Gaussian noise for 40 GeV pions.

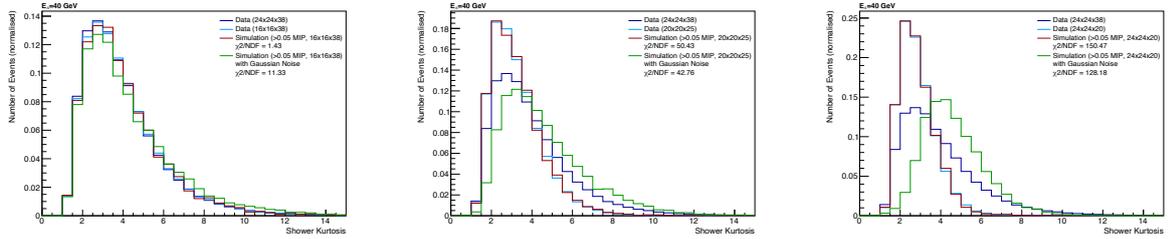


Figure B.16.: The distribution for the shower kurtosis in z -direction for the other three configuration for the full and reduced dataset and simulation with and without Gaussian noise for 40 GeV pions.

Hit Energies

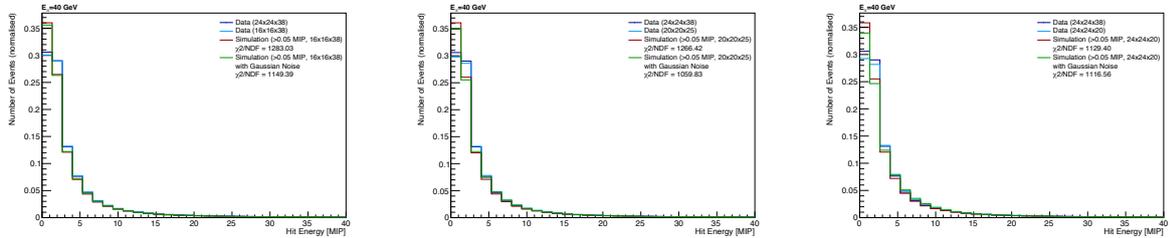


Figure B.17.: The distribution for the hit energies for the other three configuration for the full and reduced dataset and simulation with and without Gaussian noise for 40 GeV pions.

Number of Hits

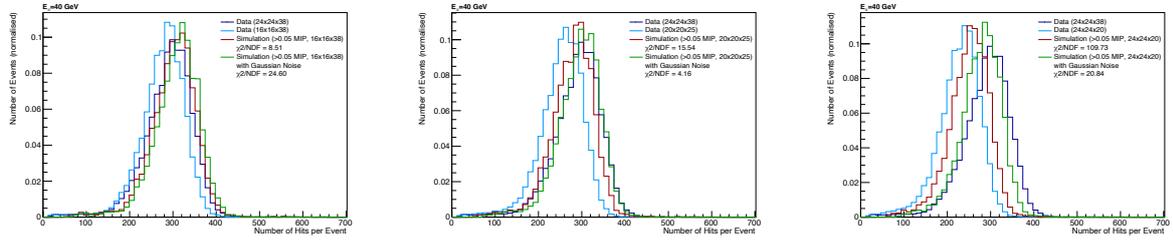


Figure B.18.: The distribution for the number of hits for the other three configuration for the full and reduced dataset and simulation with and without Gaussian noise for 40 GeV pions.

B.2. Other Variables for Other Pion Energies

Centre of Gravity

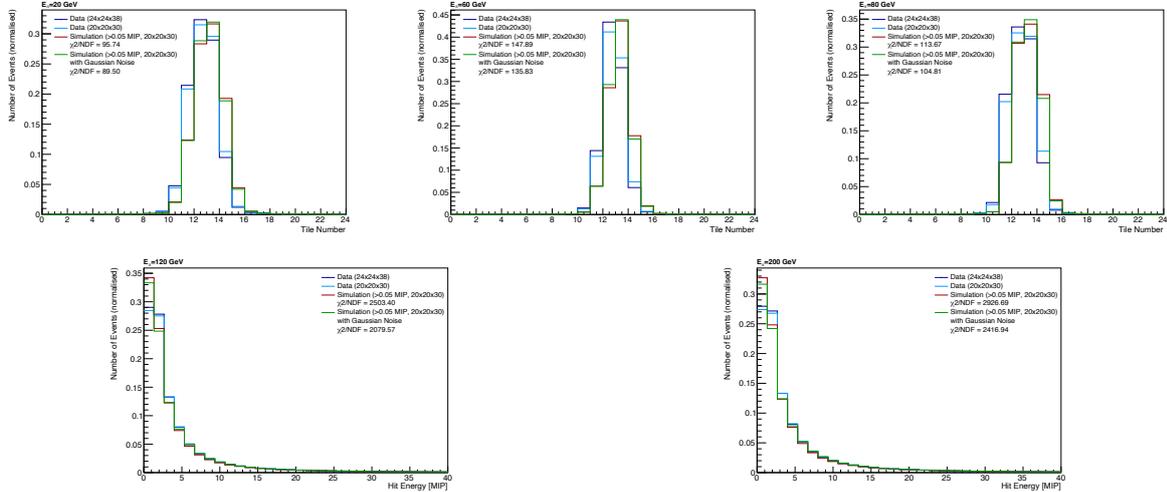


Figure B.19.: The distribution for the CoGX for the $(20 \times 20)_{\text{tiles}} \times 30_{\text{layers}}$ configuration for the full and reduced dataset and simulation with and without Gaussian noise for 20, 60, 80, 120 and 200 GeV pions.

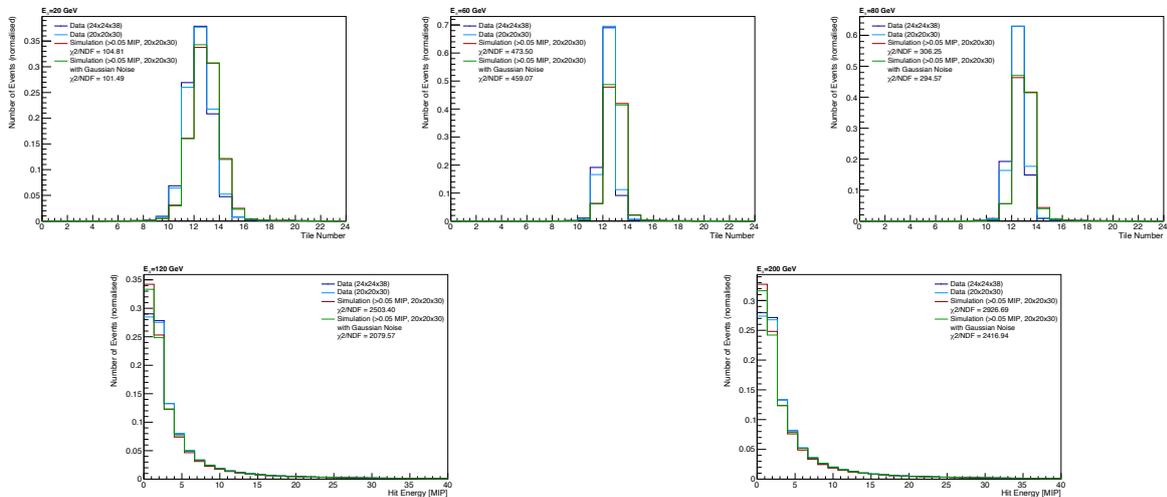


Figure B.20.: The distribution for the CoGY for the $(20 \times 20)_{\text{tiles}} \times 30_{\text{layers}}$ configuration for the full and reduced dataset and simulation with and without Gaussian noise for 20, 60, 80, 120 and 200 GeV pions.

B.2. Other Variables for Other Pion Energies

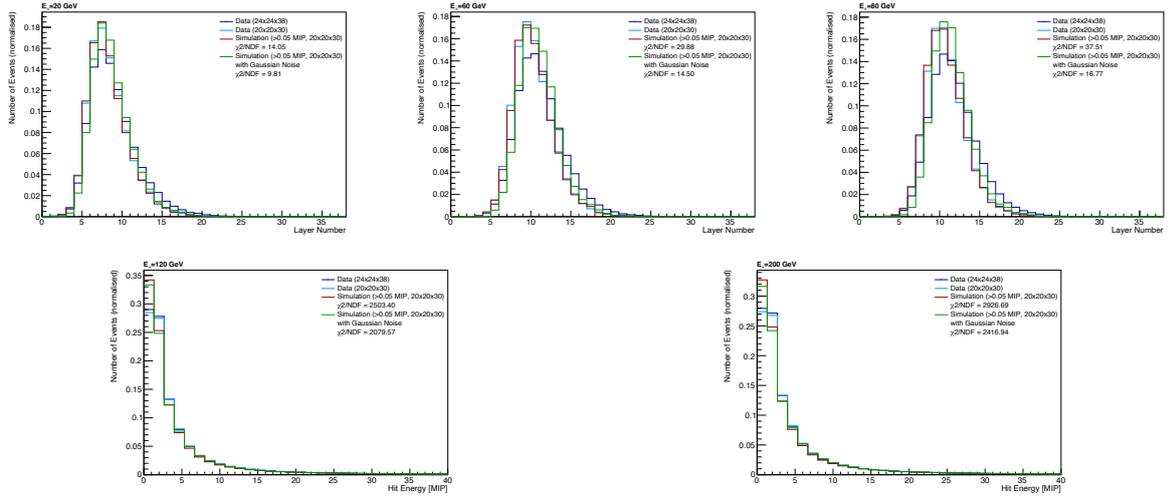


Figure B.21.: The distribution for the CoGZ for the $(20 \times 20)_{\text{tiles}} \times 30_{\text{layers}}$ configuration for the full and reduced dataset and simulation with and without Gaussian noise for 20, 60, 80, 120 and 200 GeV pions.

Shower Variance

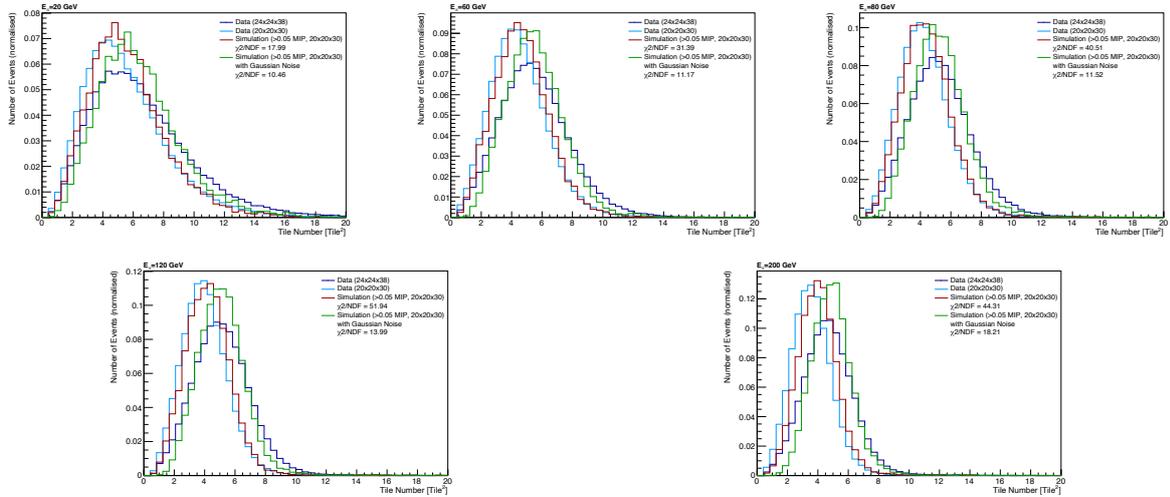


Figure B.22.: The distribution for the shower variance in y -direction for the $(20 \times 20)_{\text{tiles}} \times 30_{\text{layers}}$ configuration for the full and reduced dataset and simulation with and without Gaussian noise for 20, 60, 80, 120 and 200 GeV pions.

B. Shower Shape Variables

Shower Skewness

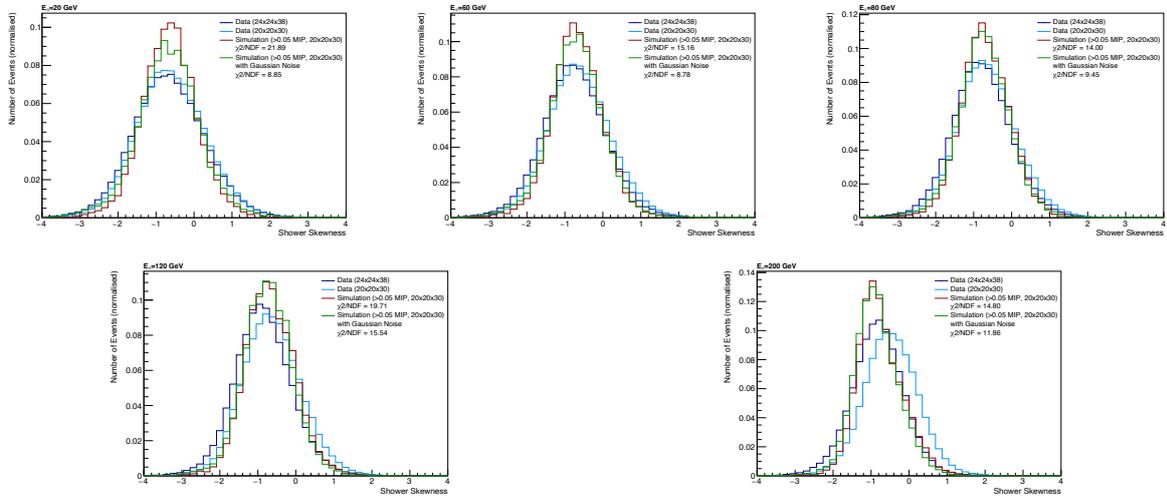


Figure B.23.: The distribution for the shower skewness in y -direction for the $(20 \times 20)_{\text{tiles}} \times 30_{\text{layers}}$ configuration for the full and reduced dataset and simulation with and without Gaussian noise for 20, 60, 80, 120 and 200 GeV pions.

Shower Kurtosis

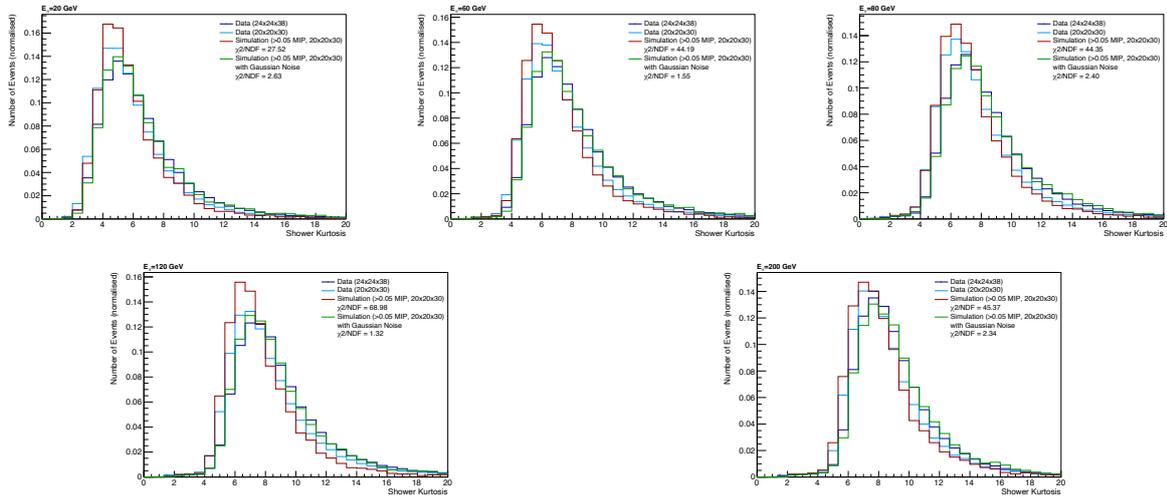


Figure B.24.: The distribution for the shower kurtosis in x -direction for the $(20 \times 20)_{\text{tiles}} \times 30_{\text{layers}}$ configuration for the full and reduced dataset and simulation with and without Gaussian noise for 20, 60, 80, 120 and 200 GeV pions.

B.2. Other Variables for Other Pion Energies

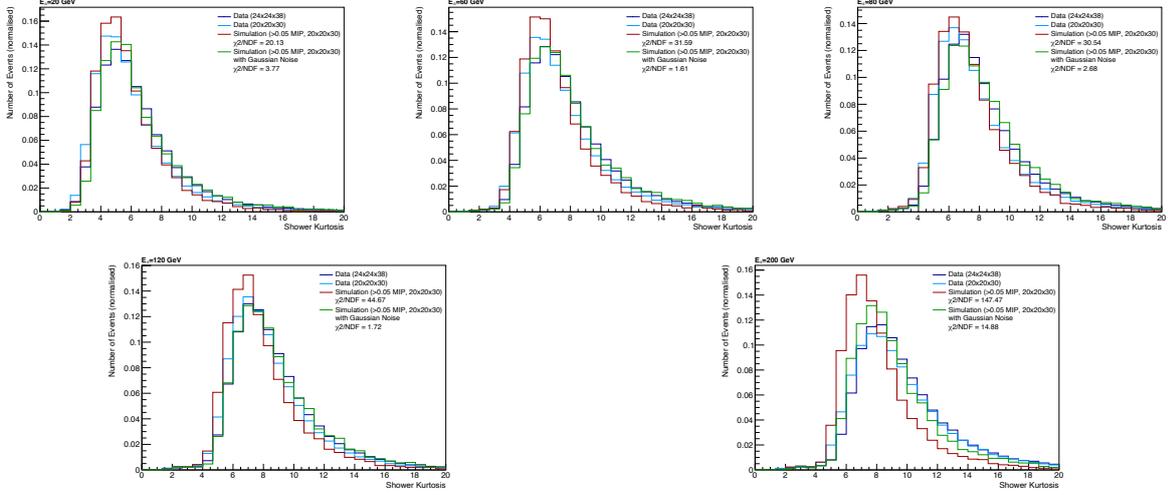


Figure B.25.: The distribution for the shower kurtosis in y -direction for the $(20 \times 20)_{\text{tiles}} \times 30_{\text{layers}}$ configuration for the full and reduced dataset and simulation with and without Gaussian noise for 20, 60, 80, 120 and 200 GeV pions.

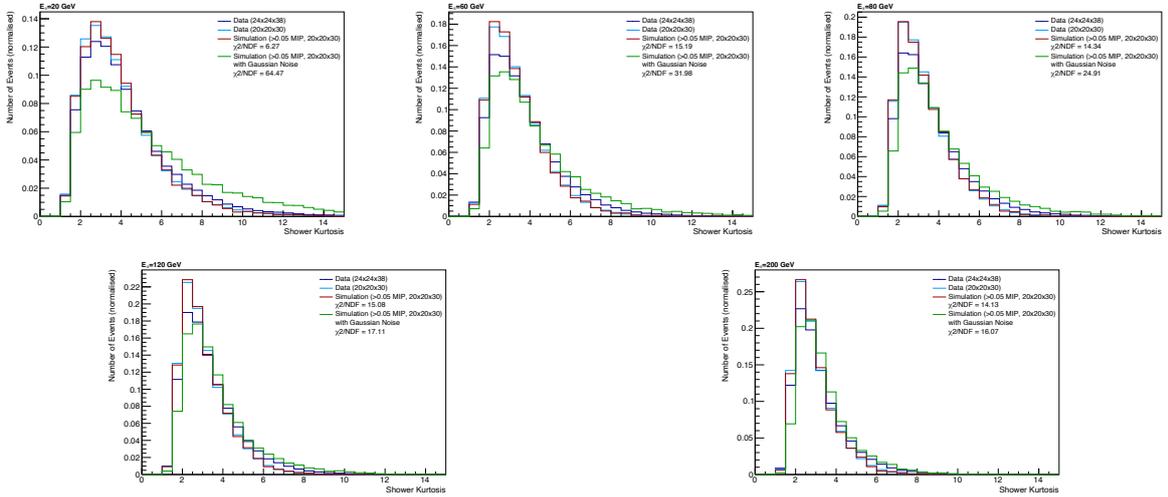


Figure B.26.: The distribution for the shower kurtosis in z -direction for the $(20 \times 20)_{\text{tiles}} \times 30_{\text{layers}}$ configuration for the full and reduced dataset and simulation with and without Gaussian noise for 20, 60, 80, 120 and 200 GeV pions.

B. Shower Shape Variables

Number of Hits

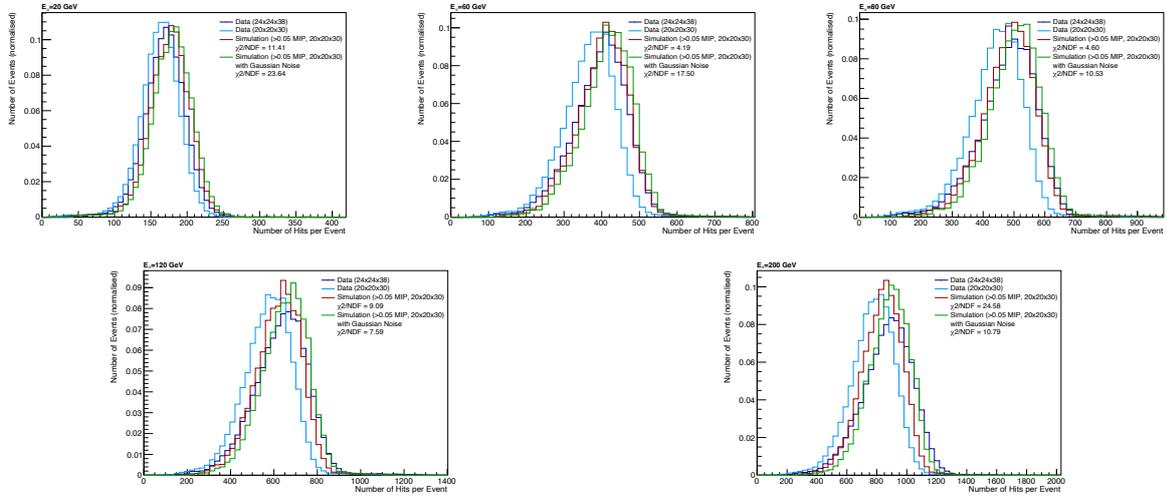


Figure B.27.: The distribution for the number of hits for the $(20 \times 20)_{\text{tiles}} \times 30_{\text{layers}}$ configuration for the full and reduced dataset and simulation with and without Gaussian noise for 20, 60, 80, 120 and 200 GeV pions.

Hit Energies

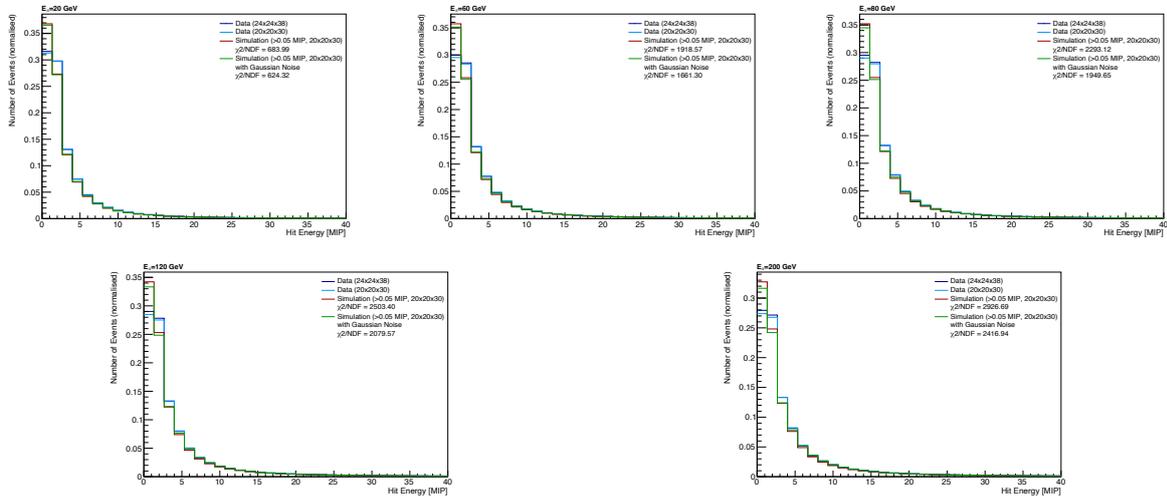


Figure B.28.: The distribution for the hit energies for the $(20 \times 20)_{\text{tiles}} \times 30_{\text{layers}}$ configuration for the full and reduced dataset and simulation with and without Gaussian noise for 20, 60, 80, 120 and 200 GeV pions.

Bibliography

- [1] S. L. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. **22**, 579 (1961)
- [2] A. Salam, J. C. Ward, *Electromagnetic and weak interactions*, Phys. Lett. **13**, 168 (1964)
- [3] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19**, 1264 (1967)
- [4] A. Salam, *Weak and Electromagnetic Interactions*, Conf. Proc. C **680519**, 367 (1968)
- [5] L. Evans, P. Bryant, *LHC Machine*, JINST **3(08)**, S08001 (2008)
- [6] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3**, S08003 (2008)
- [7] CMS Collaboration, *The CMS Experiment at the CERN LHC*, JINST **3**, S08004 (2008)
- [8] G. Bertone, D. Hooper, J. Silk, *Particle dark matter: evidence, candidates and constraints*, Phys. Rep. **405(5)**, 279 (2005)
- [9] Super-Kamiokande Collaboration, *Evidence for oscillation of atmospheric neutrinos*, Phys. Rev. Lett. **81**, 1562 (1998)
- [10] H. Georgi, S. L. Glashow, *Unity of All Elementary Particle Forces*, Phys. Rev. Lett. **32**, 438 (1974)
- [11] P. Langacker, *Grand Unified Theories and Proton Decay*, Phys. Rept. **72**, 185 (1981)
- [12] CALICE Collaboration, *Design, Construction and Commissioning of a Technological Prototype of a Highly Granular SiPM-on-tile Scintillator-Steel Hadronic Calorimeter* (2022), [arXiv:2209.15327](https://arxiv.org/abs/2209.15327)
- [13] E. Parzen, *On Estimation of a Probability Density Function and Mode*, Ann. Math. Stat. **33(3)**, 1065 (1962)

Bibliography

- [14] M. Thomson, *Modern Particle Physics*, Cambridge University Press (2013)
- [15] B. Povh, et al., *Teilchen und Kerne: Eine Einführung in die physikalischen Konzepte*, Springer Spektrum
- [16] I. Neutelings, *Standard Model* (2021), © CC BY-SA 4.0, last accessed: 16.02.2025, URL https://tikz.net/sm_particles/
- [17] Particle Data Group, *Review of Particle Physics*, Phys. Rev. D **110**(3), 030001 (2024)
- [18] M. Gell-Mann, *Symmetries of Baryons and Mesons*, Phys. Rev. **125**, 1067 (1962)
- [19] M. Gell-Mann, *A Schematic Model of Baryons and Mesons*, Phys. Lett. **8**, 214 (1964)
- [20] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, Phys. Rev. Lett. **10**, 531 (1963)
- [21] M. Kobayashi, T. Maskawa, *CP Violation in the Renormalizable Theory of Weak Interaction*, Prog. Theor. Phys. **49**, 652 (1973)
- [22] C. Amsler, *Kern- und Teilchenphysik*, vdf Hochschulverlag (2007)
- [23] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716**, 1 (2012)
- [24] CMS Collaboration, *Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC*, Phys. Lett. B **716**, 30 (2012)
- [25] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13**, 508 (1964)
- [26] F. Englert, R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13**, 321 (1964)
- [27] G. S. Guralnik, C. R. Hagen, T. W. B. Kibble, *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. **13**, 585 (1964)
- [28] T. W. B. Kibble, *Symmetry breaking in nonAbelian gauge theories*, Phys. Rev. **155**, 1554 (1967)
- [29] A. D. Sakharov, *Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe*, Pisma Zh. Eksp. Teor. Fiz. **5**, 32 (1967)
- [30] A. G. Cohen, D. B. Kaplan, A. E. Nelson, *Progress in electroweak baryogenesis*, Ann. Rev. Nucl. Part. Sci. **43**, 27 (1993)

- [31] H. Kolanoski, N. Wermes, *Particle Detectors. Fundamentals and Applications*, Oxford University Press (2022)
- [32] C. Grupen, I. Buvat, *Handbook of Particle Detection and Imaging*, volume 2, Springer (2012)
- [33] H. Bethe, *Zur Theorie des Durchgangs schneller Korpuskularstrahlen durch Materie*, Ann. Phys. **397(3)**, 325 (1930)
- [34] H. Bethe, *Bremsformel für Elektronen relativistischer Geschwindigkeit*, Z. Phys. **76(5-6)**, 293 (1932)
- [35] F. Bloch, *Bremsvermögen von Atomen mit mehreren Elektronen*, Zeitschrift für Physik **81(5-6)**, 363 (1933)
- [36] F. Bloch, *Zur Bremsung Rasch Bewegter Teilchen beim Durchgang durch Materie*, Ann. Phys. **408**, 285 (1933)
- [37] H. J. Bhabha, Proc. R. Soc. Lond. A **154(881)**, 195 (1936)
- [38] C. Møller, *Zur Theorie des Durchgangs schneller Elektronen durch Materie*, Ann. Phys. (Berlin) **406(5)**, 531 (1932)
- [39] C. Grupen, *Teilchendetektoren*, BI-Wissenschaftsverlag (1993)
- [40] E. Longo, I. Sestili, *Monte Carlo Calculation of Photon Initiated Electromagnetic Showers in Lead Glass*, Nucl. Instrum. Meth. **128**, 283 (1975), [Erratum: Nucl.Instrum.Meth. 135, 587 (1976)]
- [41] G. Merziger, G. Mühlbach, D. Wille, T. Wirth, *Formeln + Hilfen. Höhere Mathematik*, Binomi Verlag
- [42] M. Antonello, et al., *Tests of a dual-readout fiber calorimeter with SiPM light sensors*, Nucl. Instrum. Meth. A **899**, 52 (2018)
- [43] R. Wigmans, *Calorimetry: Energy Measurement in Particle Physics*, Oxford University Press (2017)
- [44] S. Lee, *On the limits of the hadronic energy resolution of calorimeters*, J. Phys: Conf. Ser. **1162**, 012043 (2019)
- [45] CALICE Collaboration, *Hadron shower decomposition in the highly granular CALICE analogue hadron calorimeter*, JINST **11(06)**, P06013 (2016)

Bibliography

- [46] U. Amaldi, *Fluctuations in Calorimetry Measurements*, Phys. Scr. **23(4A)**, 409 (1981)
- [47] C. W. Fabjan, H. Schopper, *Particle Physics Reference Library. Volume 2: Detectors for Particles and Radiation*, Springer International Publishing AG (2020)
- [48] C. Graf, *Towards Precision Time and Energy Measurements in Highly Granular Hadronic Calorimeters*, Ph.D. thesis, Technische Universität München (2020)
- [49] CALICE Collaboration, *Hadronic energy resolution of a highly granular scintillator-steel hadron calorimeter using software compensation techniques*, JINST **7**, P09017 (2012)
- [50] CALICE Collaboration, *Software Compensation for Highly Granular Calorimeters using Machine Learning*, JINST **19**, P04037 (2024)
- [51] I. Zurbano Fernandez, et al., *High-Luminosity Large Hadron Collider (HL-LHC): Technical design report 10/2020* (2020), CERN-2020-010
- [52] ATLAS Collaboration, *Technical Design Report: A High-Granularity Timing Detector for the ATLAS Phase-II Upgrade*, Technical report (2020), CERN-LHCC-2020-007, ATLAS-TDR-031, URL <https://cds.cern.ch/record/2719855>
- [53] R. G. Suarez, *The Future Circular Collider (FCC) at CERN*, PoS **DISCRETE2020-2021**, 009 (2022)
- [54] A. Abada, et al., *FCC Physics Opportunities*, Eur. Phys. J. C. **79(6)**, 474 (2019)
- [55] A. Abada, et al., *FCC-ee: The Lepton Collider. Future Circular Collider Conceptual Design Report Volume 2*, Eur. Phys. J. ST **228(2)**, 261 (2019)
- [56] A. Abada, et al., *FCC-hh: The Hadron Collider: Future Circular Collider Conceptual Design Report Volume 3*, Eur. Phys. J. ST **228(4)**, 755 (2019)
- [57] T. Barklow, et al., *ILC Operating Scenarios* (2015), [arXiv:1506.07830](https://arxiv.org/abs/1506.07830)
- [58] Linear Collider Collaboration, *The International Linear Collider. A Global Project*, Technical report (2019), URL <https://cds.cern.ch/record/2655225>
- [59] A. F. Zarnecki, *On the physics potential of ILC and CLIC*, PoS **CORFU2019**, 037 (2020)

- [60] P. Roloff, et al., *The Compact Linear e^+e^- Collider (CLIC): Physics Potential* (2018), [arXiv:1812.07986](#)
- [61] O. Brunner, et al., *The CLIC project* (2022), [arXiv:2203.09186](#)
- [62] T. Bergauer, *The European strategy and detector R&D program*, Nucl. Instrum. Meth. A **1069**, 169949 (2024)
- [63] O. Pinto, *Operation and Calibration of a Highly Granular Hadron Calorimeter with SiPM-on-Tile Read-out* (2020), [arXiv:2004.00370](#)
- [64] D. Heuchel, *Commissioning, characterisation and temperature stabilisation of a 22000 channel SiPM-on-tile hadron calorimeter system*, NIM-A **1045**, 167536 (2023)
- [65] P. Eckert, et al., *Characterisation Studies of Silicon Photomultipliers*, Nucl. Instrum. Meth. A **620**, 217 (2010)
- [66] N. Anfimov, et al., *Study of Silicon Photomultiplier Performance at Different Temperatures*, Nucl. Instrum. Meth. A **997**, 165162 (2021)
- [67] CMS CE Collaboration, *Construction and commissioning of CMS CE prototype silicon modules*, JINST **16(04)**, T04002 (2021)
- [68] CMS CE Collaboration, *The DAQ system of the 12,000 Channel CMS High Granularity Calorimeter Prototype*, JINST **16(04)**, T04001 (2021)
- [69] D. Heuchel, *Particle Flow Studies with Highly Granular Calorimeter Data*, Ph.D. thesis, Heidelberg U. (2022), [heidok.00031794](#)
- [70] F. Sefkow, F. Simon, *A highly granular SiPM-on-tile calorimeter prototype*, J. Phys. Conf. Ser. **1162(1)**, 012012 (2019)
- [71] GEANT4 Collaboration, *GEANT4. A simulation toolkit*, last accessed: 26.03.2025, URL <https://geant4.web.cern.ch>
- [72] ATLAS Collaboration, *ATLAS Software and Computing HL-LHC Roadmap*, Technical report, CERN, Geneva (2022), CERN-LHCC-2022-005, LHCC-G-182, URL <https://cds.cern.ch/record/2802918>
- [73] T. Arens, et al., *Mathematik (German Edition)*, Spektrum Akademischer Verlag
- [74] I. Chakravarti, R. Laha, J. Roy, *Handbook of Methods of Applied Statistics*, volume 1 of *Handbook of Methods of Applied Statistics*, Wiley (1967)

Bibliography

- [75] A. N. Kolmogorov, *Sulla Determinazione Empirica di una Legge di Distribuzione*, *Giornale dell'Istituto Italiano degli Attuari* **4**, 83 (1933)
- [76] K. Pearson, *On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling*, *London Edinburgh Philos. Mag. & J. Sci.* **50(302)**, 157 (1900)
- [77] *Kolmogorov-Smirnov Goodness-of-Fit Test*, last accessed: 07.03.2025, URL <https://www.itl.nist.gov/div898/handbook/eda/section3/eda35g.htm>
- [78] W. T. Eadie, *Statistical Methods in Experimental Physics*, Elsevier Science Ltd
- [79] *Kolmogorov Probability*, URL <https://root.cern/root/html602/TMath.html#TMath:KolmogorovProb>
- [80] G. Cowan, *Statistical data analysis*, Clarendon Press (1998)
- [81] K. Pearson, *On lines and planes of closest fit to systems of points in space*, *London Edinburgh Dublin Philos. Mag. & J. Sci.* **2(11)**, 559 (1901)
- [82] A. Wilhahn, *Investigation of Fast Hadron Shower Simulation Methods with the CALICE AHCAL Prototype*, Master's thesis, [II.Physik-UniGö-MSc-2022/03](#)
- [83] N. Ahmed, T. Natarajan, K. Rao, *Discrete Cosine Transform*, *IEEE Trans. Comput.* **C-23(1)**, 90 (1974)
- [84] W.-H. Chen, C. Smith, S. Fralick, *A Fast Computational Algorithm for the Discrete Cosine Transform*, *IEEE Trans. Comput.* **25(9)**, 1004 (1977)
- [85] C. F. Gauß, *Theoria interpolationis methodo nova tractata* Gauß, Carl Friedrich: Werke. Bd. 3 (1866), S. 265 - 327
- [86] J. W. Cooley, J. W. Tukey, *An algorithm for the machine calculation of complex Fourier series*, *Math. Comput.* **19(90)**, 297 (1965)
- [87] *Gaussian KDE*, last accessed: 10.03.2025, URL https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.gaussian_kde.html
- [88] O. L. Pinto, *Shower Shapes in a Highly Granular SiPM-on-Tile Analog Hadron Calorimeter*, Ph.D. thesis, U. Hamburg (2022)

Danksagung

First of all, I would like to thank Prof. Stan Lai for allowing me to work on my Master's thesis in his group. Your supervision was fantastic and I enjoyed every single second of it. It was a pleasure to work with you, not only during this Master's thesis research, but also during my Bachelor's thesis and internship. I always felt comfortable working with you. You really care about your students and you show it every day. Regardless of the problems or extra wishes, you were always there to help, which I really appreciate. Thank you, Stan, for taking me into your group and offering me the Bachelor's thesis that started it all, and for always believing in me from day one.

I would also like to thank apl. Prof. Jörn Große-Knetter for being my second referee.

A special thanks goes to Julian and especially André for the day-to-day supervision during my internship and this thesis. This thesis would not have been possible without your help. Thank you for being so patient with me and I am sorry if I ever annoyed you with my problems and especially bugs in my code :D Also a big thank you to Kira, who I sometimes annoyed with my ROOT problems. Thanks for letting me come to you with my problems sometimes!

Thanks to the group in general! The meetings we had were always very fun and entertaining. I wish you all the best for the future.

My office partner, Lena: a very, very special and big thank you. Ohhh, it was an amazing time with you in the office!!! I could not have asked for a better office mate. Thank you for always bringing so much fun to the office and making it not just any office. The amount of times we just laughed or were on the verge of ending it all, this whole year with you has been incredible and I wish you all the best for your future! May all your wishes come true and always be the positive person you have always been!!!

Thanks to all my friends and family. Without your moral support this thesis would not have been possible!

I have to mention some friends in particular that I made during this thesis: my CERN friends, or as I call them: my Swiss mausis. I had the best time with you! Thank you Christine, Laura, Zsófia, Moritz, Fabi, Alice, Luisa and Niklas. I love y'all so much.

Bibliography

With this thesis my studies at the University of Göttingen come to an end. The last people I would like to thank are probably the most important people during my studies here: Marina, Lisa and Laila. Thank you for being with me during this whole time. We met in the preliminary course before our very first semester and you stayed with me until the very last day. We have been through so much during both degrees and you have helped me so much during those times - not just physics wise. I cannot put it into words. Whatever you do in your life, I wish you all the best. Thank you so much, from the bottom of my heart. I will miss you and I love you all.

Erklärung

nach §17(9) der Prüfungsordnung für den Bachelor-Studiengang Physik und den Master-Studiengang Physik an der Universität Göttingen: Hiermit erkläre ich, dass ich diese Abschlussarbeit selbständig verfasst habe, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe und alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen wurden, als solche kenntlich gemacht habe.

Darüberhinaus erkläre ich, dass diese Abschlussarbeit nicht, auch nicht auszugsweise, im Rahmen einer nichtbestanden Prüfung an dieser oder einer anderen Hochschule eingereicht wurde.

Göttingen, den 14. Juni 2025



(Zobeyer Ghafoor)