Bachelor's Thesis

Schnelle Simulationsmethoden für elektromagnetische Schauer mit dem hochgranularen CALICE AHCAL Prototyp

Fast Simulation Methods for Electromagnetic Showers with the Highly Granular CALICE AHCAL Prototype

prepared by

Victoria Swoboda

from Gehrden

at the II. Physikalischen Institut

Abstract

In dieser Bachelorarbeit wird die Reduktion von Teststrahldaten eines hochgranularen Kalorimeters mit Hilfe der diskreten Kosinustransformation untersucht. Ziel ist es, eine datenbasierte, schnelle Teilchenschauersimulation zu implementieren, die im Vergleich zu Vollsimulationen eine stark reduzierte Rechenzeit und einen geringeren Bedarf an Rechenressourcen aufweist. Für die schnelle Simulation werden Kerndichteschätzer für die Erzeugung simulierter Ereignisse verwenden. Der verwendete Teststrahldatensatz umfasst Daten von elektromagnetischen Schauern, ausgelöst von Elektronen verschiedener Energien, und wurde 2018 am CERN mit dem hochgranularen Analogen Hadronkalorimeter der CALICE-Kollaboration aufgezeichnet.

Für die Untersuchung werden die Hitenergien mittels der diskreten Kosinustransformation in Koeffizienten von Kosinusschwingungen transformiert. Die Koeffizienten der Schwingungen, von denen angenommen wird, dass sie ausschließlich zum Rauschen beitragen, werden durch zufällige, gaußverteilte Werte ersetzt. Nach Anwendung der Inversen der diskreten Kosinustransformation werden die Verteilungen der kinematischen Variablen zwischen dem Originaldatensatz und dem, welcher zufällig erzeugtes Rauschen beinhaltet, verglichen. Die Untersuchung zeigt, dass eine gute Übereinstimmung erzielt werden kann, wenn die Parameter der Gauß-Verteilungen sowie die generierten Koeffizienten sorgfältig ausgewählt werden. Weitere Untersuchungen könnten zu einer noch stärkeren Komprimierung und damit auch zu einer stärkeren Verringerung des Rechenaufwands führen.

Abstract

In this Bachelor thesis, the reduction of highly granular calorimeter test beam data is investigated by using the discrete cosine transformation. The aim is to implement a data-driven fast particle shower simulation with greatly reduced computing-time and easier requirements for computational resources compared to full simulations. The fast simulation will then use kernel density estimators for the generation of simulated events. The test beam dataset this investigation is based on comprises electromagnetic shower data, initiated by electrons of various energies. The dataset has been recorded in 2018 at CERN with the highly granular Analogue Hadron Calorimeter of the CALICE Collaboration.

For the investigation, hit energies have been transformed via the discrete cosine transformation into coefficients of cosine nodes. The coefficients of nodes considered to be only contributing to noise have been replaced by random values that are Gaussian distributed. After applying the inverse discrete cosine transformation, distributions of kinematic variables have been compared between the original dataset and the one based on randomly generated noise. The investigation showed that good agreement can be achieved if the parameters of the Gaussian distributions as well as which coefficients are being replaced are chosen carefully. Further investigations could lead to even greater compression, and thus greater reduction of computational resources too.

Contents

1. Introduction

The smallest building blocks of our known universe are the elementary particles, which are currently described by the Standard Model of particle physics [1–7]. The Standard Model (SM) predicts a variety of elementary spin-$\frac{1}{2}$ particles, so-called fermions, that form matter as well as integer spin particles, the bosons, that mediate forces between fermions. It is a well established theory, as its predictions have been verified over and over again by using particle accelerators such as the Large Hadron Collider (LHC) [8] in Geneva. Experiments at the LHC, but elsewhere too, continuously search for evidence that supports the SM, but also for hints of theories beyond known ones. The last major confirmation of the SM's predictions was the latest discovery of an elementary particle, the Higgs boson, by the CMS and ATLAS Collaborations [9, 10] in 2012.

The detection of elementary particles requires different detection systems such as trackers, calorimeters or muon chambers. Since most elementary particles are unstable and decay before they can be directly detected, their properties must be inferred indirectly from measurements of their decay products. Calorimeters in particular are used to measure the energy of showers initiated by elementary particles or their decay products. Calorimetry is an active field of research investigated by different projects around the world. For example, the CALICE Collaboration [11] concentrates on the research and development of highly granular calorimeters for the future International Linear Collider. The Analogue Hadron Calorimeter Technological Prototype (AHCAL) [12, 13] is one of the many detector prototypes of CALICE.

In order to improve the energy resolution of future calorimeters, test beam campaigns for performance testing as well as particle shower simulations are of great importance. To make such simulations more efficient, it is crucial to minimise computational resources and time significantly, with a minimum of information loss. A method that offers such resource-saving alternatives are data-based fast simulations, in conjunction with (loss-free) data compression. This thesis investigates the possibility of compressing a dataset recorded at CERN using the discrete cosine transformation and how this affects the shape

of particle showers. To do so, electron test beam data of the AHCAL is used.

The thesis is structured as follows. Chapter 2 presents an overview of the Standard Model, as well as the theory behind particle showers and different calorimeter types. The energy resolution for such showers will be explained too. Chapter 3 introduces the CALICE Collaboration and their test beam campaigns, as well as the structure of the AHCAL prototype. In Chapter 4, the fast particle shower simulation method, the kernel density estimation, and data reduction using the discrete cosine transformation are explained, and the results are presented and discussed. In the end, a conclusion and an outlook are given in Chapter 5.

2. Theoretical Background

2.1. Theory of the Standard Model

The Standard Model (SM) is a widely accepted theory which describes physics accurately on the scale of the smallest known particles [1–7, 14]. The model contains a set of elementary particles and their antiparticles, the latter differing in the sign of their electric charges, as well as three of the four fundamental forces: the electromagnetic force, the weak nuclear force and the strong nuclear force. The fourth fundamental force, the gravitational one, is not described by the SM.

The elementary particles can be classified through their spin into those with spin $\frac{1}{2}$, called fermions, and those with an integer spin, called bosons. Fermions make up all matter, while bosons are force carrying particles. The particle content of the SM is illustrated in Figure 2.1.

Standard Model of Elementary Particles
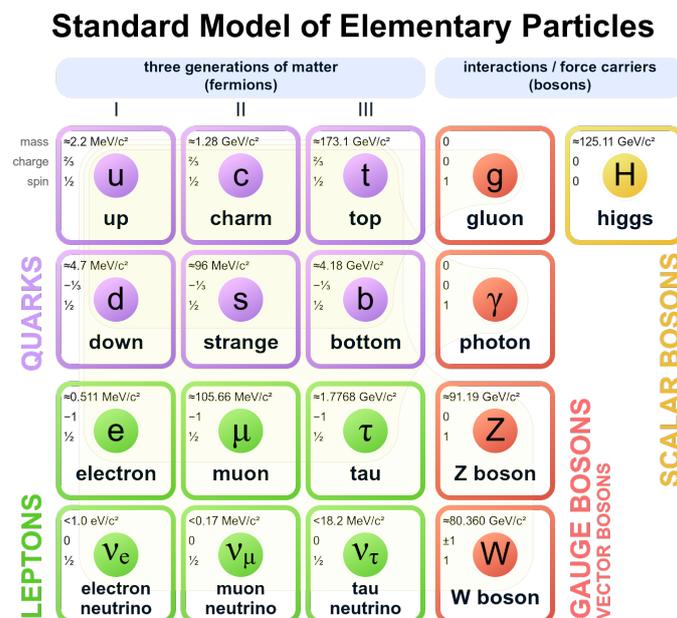
Figure 2.1.: The fundamental particles of the Standard Model of particle physics [15].

2. Theoretical Background

Fermions can further be divided into quarks and leptons. The six quarks are the only colour-charge-carrying fermions and thus subject to the strong force. Possible colour charges for quarks are red, blue, and green, and for antiquarks anti-red, anti-blue and anti-green. Since only colourless objects exist in nature, quarks are prevented from appearing isolated. This property is called confinement and leads to quarks only occurring as hadrons, which are particles composed of multiple quarks [16].

Furthermore, quarks can be divided into up-type and down-type quarks. Up-type quarks such as up (u), charm (c) and top (t) carry an electric charge of $+\frac{2}{3}$ in units of the elementary charge, while down-type quarks like down (d), strange (s) and bottom (b) carry a charge of $-\frac{1}{3}$. All the left-handed quarks and right-handed antiquarks, referring to their chirality, carry weak isospin, which is the charge of the weak interaction. While the up-type quarks have an isospin of $1/2$, the down-type quarks carry an isospin of $-1/2$. It is the same for right-handed antiquarks, but with changed sign, left-handed antiquarks and right-handed quarks are all considered to have a weak isospin of zero. Quarks are also categorised in generations, where the first generation includes the lightest quarks, up and down, the second the heavier ones with charm and strange, and the third the heaviest with top and bottom.

A similar division can be made for the leptons, which include electrons e^-, muons μ^-, and tauons τ^-, all carrying a charge of -1, and the electrically neutral electron neutrino ν_e, the muon neutrino ν_μ and the tau neutrino ν_τ. The electron and the electron neutrino form the first generation, the muon and the muon neutrino the second and the tau lepton and its corresponding neutrino the third. The generations are again sorted from the lightest to the heaviest. Leptons that are left-handed in terms of their chirality also carry weak isospin. Here, the neutrinos have a weak isospin of $1/2$ and the charged leptons of $-1/2$. Again, it is the same for their antiparticles, but with opposite sign. Right-handed leptons and left-handed antileptons have an weak isospin of zero.

Elementary particles with a spin of 1 are called gauge bosons. They are force-carrying particles and act as mediators between fermions. The exchange particle of the electromagnetic force is the photon γ. It couples to electrically charged particles, which means it does not couple to itself. As a virtual particle, it exchanges four-momentum between two real particles. The quantum field theory describing the electromagnetic force is Quantum Electrodynamics.

4

The gauge bosons of the weak force are the electrically neutral Z- and the W$^\pm$-bosons, which have an electric charge of ±1 and masses of $m_\mathrm{W} = 80.376(33)\,\mathrm{GeV}$ [17] for the W- and $m_\mathrm{Z} = 91.1876(21)\,\mathrm{GeV}$ [17] for the Z-Boson. Both couple to all leptons, since all leptons carry the weak isospin, the charge of the weak interaction. The W$^\pm$-boson itself carries an isospin of ±1, while the Z-Boson has a isospin of zero. The weak isospin can be described as a doublet of either a charged lepton and a neutrino or two quarks. By emitting a W$^\pm$-boson, one fermion changes to the other part of its doublet, with the electric charge always being conserved. While only specific doublets for leptons are possible, as in Ref. [18]

$$\begin{pmatrix} \nu_e \\ e^- \end{pmatrix}, \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}, \begin{pmatrix} \tau_e \\ \tau^- \end{pmatrix} \quad , \tag{2.1}$$

mixing between generations is possible for quarks as long as an up-type quark is combined with a down-type quark. The weak and the electromagnetic force have already been unified into the electroweak interaction [1, 2].

The gluon represents the strong nuclear force, which is described by Quantum Chromodynamics [4, 14]. It couples to colour charged particles like quarks. The gluon itself carries a colour and an anti-colour too, which is why it couples to itself. By emitting a gluon, the colour charge of a quark or a gluon changes. Like the photon, the gluon is a massless particle.

The last elementary particle of the SM is a spin-0 particle, the Higgs boson [5–7]. The Higgs boson can be understood as an excitation in the Higgs field, which has a non-vanishing vacuum expectation value, unlike other fundamental quantum fields. The interaction of the Higgs field with other massive, elementary particles causes them to obtain mass. The Higgs boson was discovered in 2012 by the ATLAS and the CMS Collaborations at CERN [9, 10].

2.2. Electromagnetic Showers

There are two main processes contributing to the development of electromagnetic showers at high energies: bremsstrahlung and pair production [19]. Bremsstrahlung is relevant for electrons and positrons passing through matter. When being slowed down by the electric potential of nuclei of the medium of the calorimeter, the electron/positron will emit a photon, which is considered to be bremsstrahlung. Tau leptons cannot initiate

an electromagnetic shower, since they will decay before reaching the calorimeter. Muons usually do not start a shower either, since they do not radiate off enough bremsstrahlung due to their mass [19].

The relevant process for photons, on the other hand, is pair production, where the photon decays into an electron-positron pair, which will emit bremsstrahlung again. These emitted photons will proceed with pair production, and so on. In this way, the electromagnetic shower will evolve until the average energy of a particle in the shower reaches the critical energy E_c. As long as an electron has an energy above E_c, it will lose its energy mainly through bremsstrahlung. As soon as the energy falls below the critical energy, the energy loss through ionisation will dominate and the electromagnetic shower will stop soon afterwards.

An important characteristic of electromagnetic showers is the radiation length X_0 [17]. It is defined as the distance an electron has to travel to be left with a fraction of $1/e$ of its initial energy. Similarly, it can be approximated as the average distance after which the electron will emit bremsstrahlung. Analogously, the electromagnetic absorption length [17, 19],

$$\lambda \approx \frac{9}{7} X_0 \quad , \tag{2.2}$$

is the average distance after which a photon undergoes pair production at high energies. Both radiation and absorption length are used to parametrise the total length of an electromagnetic shower. The schematic development of a simplified electromagnetic shower is shown in Figure 2.2, where $X_0 = \lambda$ has been assumed for simplicity.
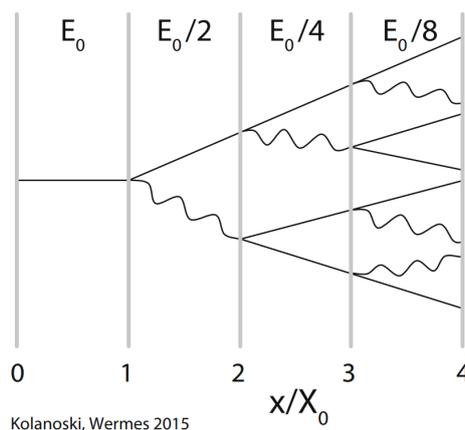
Figure 2.2.: Simplified model of the development of an electromagnetic shower, where E_0 is the energy of the initial particle and x/X_0 the distance to the shower start in units of the radiation length [19].

The approximate longitudinal length in units of absorption lengths, in which 98% of the total shower energy is included, is [20]

$$t_{98\%} \approx (t_\text{max} + 13.6) \pm 2.0 \quad . \tag{2.3}$$

Here, t_max is the position of the shower maximum, which is the point where the energy deposition is maximal. The distance to the shower maximum (given in units of absorption lengths) is determined by [17]

$$t_\text{max} = \ln\frac{E_0}{E_\text{c}} + \begin{cases} -0.5 & , \quad \text{electrons} \\ +0.5 & , \quad \text{photons} \end{cases} , \tag{2.4}$$

where E_0 is the energy of the initial particle. Therefore, the shower length increases with higher energy E_0, but only logarithmically, which means that calorimeters, with fixed lengths can be used for a large range of beam energies. The energy distribution in longitudinal direction is then described by the empirical formula [21]

$$\frac{dE}{dt} = E_0\frac{b^a}{\Gamma(a)}t^{a-1}e^{-bt} \quad , \tag{2.5}$$

where a and b are parameters depending on the initial energy and the atomic number of the material used, and $\Gamma(a)$ is the gamma function defined as

$$\Gamma(x) = \int_0^\infty y^{x-1}e^{-y}dy \quad . \tag{2.6}$$

Using Equation (2.5) the shower maximum is determined to be most probably at

$$t_\text{max} = \frac{a-1}{b} \quad . \tag{2.7}$$

The longitudinal shape of an electromagnetic shower described by Equation (2.5) can be seen as a fit in Figure 2.3, where the average energy deposit is plotted against the hit position in units of the radiation length. The data for the fit are taken from a shower initiated by a $10\,\text{GeV}$ positron beam. The fit parameters are listed in the legend. Here, a is represented as α and b as $1/\beta$. It can be seen that the energy of the shower, which is detected in the medium, rises rapidly after just a few radiation lengths and then falls of more slowly. Almost no energy is measured after 20 radiation lengths, which indicates that the shower only developed in the front of the material.
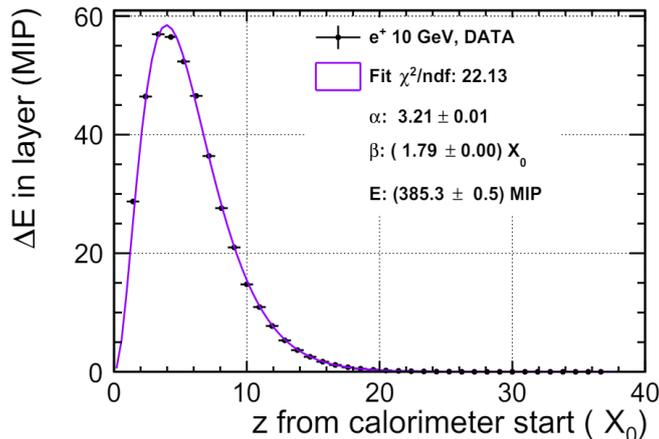
Figure 2.3.: Data and fit of the average energy deposit ΔE depending on the hit position z in units of the radiation length using Equation (2.5) for data of electromagnetic showers initiated by positrons with $10\,\text{GeV}$ beam energy [22].

Similar to $t_{98\%}$, a cylinder with the Molière radius [17]

$$R_{\text{M}} = \frac{E_{\text{s}}}{E_{\text{c}}} X_0 \quad \text{where } E_{\text{s}} = 21.2\,\text{MeV} \tag{2.8}$$

around the shower axis contains approximately 90% of the shower energy in the transverse plane. The proportionality of the ratio between the Molière radius and the radiation length is [19]

$$\frac{R_{\text{M}}}{X_0} \propto \frac{1}{E_{\text{c}}} \propto Z \quad . \tag{2.9}$$

Here, Z is the atomic number of the medium. This implies that the shower is slimmer in relation to its length in media with high values for the atomic number Z.

The radial distribution can be parametrised by [22]

$$\frac{\Delta E}{\Delta S}(r) = \frac{E}{2\pi} \cdot \left\{ f_{\text{c}} \cdot \frac{\exp\left(\frac{-r}{\beta_{\text{c}}}\right)}{\beta_{\text{c}}^2} + (1 - f_{\text{c}}) \cdot \frac{\exp\left(\frac{-r}{\beta_{\text{h}}}\right)}{\beta_{\text{h}}^2} \right\} \quad , \tag{2.10}$$

which is divided into a core component, given by the first summand, describing the inner electromagnetically dominated region, and a halo component, given by the second summand, describing the outer part. f_{c} is the fraction of the total energy deposited in the core, and ΔS is the area of a ring with a thickness of Δr, where r is the distance from the shower axis. β_{c} and β_{h} are fit parameters describing the core or halo part, respectively.

The energy deposit in radial direction depending on the radial distance to the shower axis of an electromagnetic shower started by positrons with $10\,\mathrm{GeV}$ beam energy can be seen in Figure 2.4. Figure 2.4 also shows a fit using Equation (2.10) and fit parameters, which are explained in [22]. It is recognisable that the energy deposit falls continuously with the distance to the shower axis, while the fall is more rapid closer to the centre.
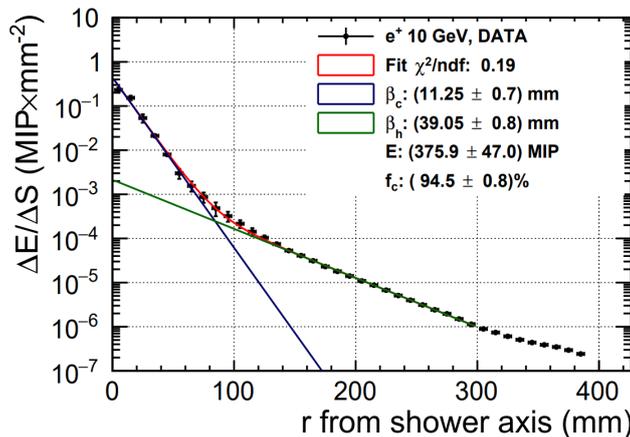
Figure 2.4.: Data and fit of the average energy deposit ΔE (scaled with ΔS, which is a ring of width Δr) depending on the hit position r as distance from the shower axis for data of electromagnetic showers initiated by electrons with $10\,\mathrm{GeV}$ beam energy [22].

2.3. Hadronic Showers

High energy hadrons passing through dense matter will develop hadronic showers. Unlike electromagnetic showers, there are many processes involved, which make the hadronic shower development more complicated and lead to greater fluctuations in their deposited energy. The first phase of the shower development is a high energy cascade, where a hadron collides inelastically with a nucleus of the medium. Through the inelastic scattering, new particles are formed. The hadronic cascade continues when the new particles scatter inelastically as well. However, some of these newly formed particles will be neutral pions, which are most likely to decay into two photons. Those photons will start an electromagnetic subshower which will spread independently of the hadronic shower.

A more detailed description of the initial inelastic scattering is that the incoming hadron interacts with only one of the nucleons of the nucleus. This results in newly formed particles with high energy inside the nucleus, starting an intranuclear cascade by interacting inelastically themselves. Over time, the particles will either have less energy than needed

to continue the inelastic scattering or will escape the nucleus, which leads to highly exited nuclei emitting nucleons and light nucleus fragments. This process is called spallation and takes place about 10^{-22} s after the initial collision of the hadron with the nucleon [19]. The emitted particles have an energy of the order of 100 MeV [19]. The next phase takes place about 10^{-18} s after the first collision [19]. The still exited nucleus emits further nucleons and light nucleus fragments, which is called evaporation, but now with just a few MeV. During evaporation, a nucleus may undergo nuclear fission as well [19]. The schematic development of the hadronic shower can be seen in Figure 2.5.
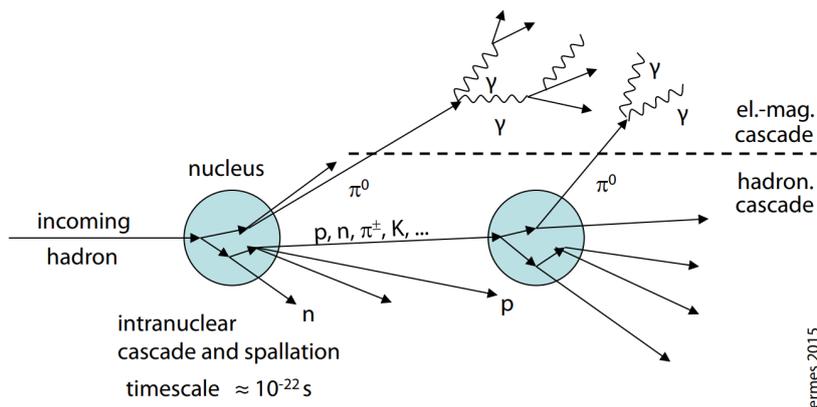
Figure 2.5.: Model of the development of a hadronic shower. The incoming hadron will first scatter inelastically with one nucleon of the nucleus, starting an intranuclear cascade. Eventually, nucleons will escape the nucleus, forming new hadrons, which will again scatter inelastically with other nuclei. Neutral pions are produced too, which start electromagnetic showers by emitting two photons each [19].

As already mentioned, hadronic showers also contain electromagnetic subshowers, mostly initiated through neutral pion decays. The mean fraction of energy of a hadronic shower which belongs to electromagnetic subshowers, $\langle f_{\text{EM}} \rangle$, is described through

$$\langle f_{\text{EM}} \rangle \approx 1 - \left(\frac{E}{\tilde{E}} \right)^{k-1} \quad , \tag{2.11}$$

where $k \approx 0.82$ and $\tilde{E}$ is the average energy necessary to produce a neutral pion [19, 23]. The value of $\tilde{E}$ depends on the medium in which the hadronic shower develops. Since electromagnetic shower development only involves two processes, their fluctuations are smaller than those of hadronic shower development, which results in broader energy distributions for hadronic showers. According to Equation (2.11), the relative energy of the electromagnetic part of the hadronic shower increases with beam energy, thus fluctuations in the hadronic shower decrease with higher energy. However, the actual

value of f_{EM} fluctuates strongly. The strong fluctuations are recognisable in Figure 2.6, where the electromagnetic fraction of a shower per event is plotted for data of pions with 150 GeV beam energy. The dependence of the calorimeter response and therefore of the energy resolution on the electromagnetic fraction will be discussed in Section 2.4.2.
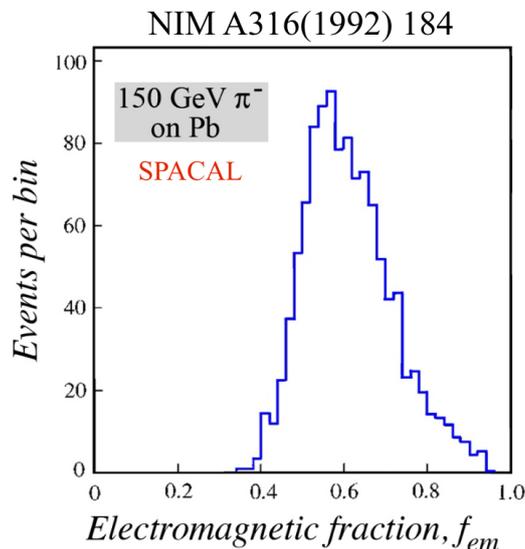
Figure 2.6.: Distribution of the electromagnetic fraction per event for showers initiated by negative pions with 150 GeV beam energy [24].

Similar to the radiation length of electromagnetic showers, the nuclear interaction length can be approximated by [19]

$$\lambda_{\mathrm{int}} \approx 35\,\mathrm{g\,cm}^{-2}\frac{A^{1/3}}{\rho} \quad , \tag{2.12}$$

which parametrises the total length of a hadronic shower. Here, A is the atomic weight and ρ the density of the medium. The ratio of the nuclear interaction length to the radiation length can be parametrised as [19]

$$\frac{\lambda_{\mathrm{int}}}{X_0} \approx 0.37Z \quad . \tag{2.13}$$

Here, a linear dependence on the atomic number Z of the medium can be observed. Since materials made from elements with higher atomic numbers tend to be more dense, the interaction length is significantly larger in dense materials than the radiation length.

The hadronic shower length in longitudinal direction which includes 95% of the total

shower energy on average can be approximated by [20]

$$t_{95\%} \approx t_{max} + 2.5\lambda_{int} \left(\frac{E}{\text{GeV}}\right)^{0.13} \quad , \tag{2.14}$$

where E is the initial energy and t_{max} the length where the hadronic shower reaches its maximum, approximated by [20]

$$t_{max} \approx 0.2 \ln \frac{E}{\text{GeV}} + 0.7 \quad . \tag{2.15}$$

Both quantities are shown in units of the interaction length and were tested to be valid for energies between a few GeV to multiple 100 GeV [20]. The longitudinal energy distribution can be described through [25, 26]

$$\frac{dE}{dz} = E_0 \cdot \left\{\frac{f_{EM}}{\Gamma(\alpha_s)} \cdot \left(\frac{z_s}{\beta_s}\right)^{\alpha_s - 1} \cdot \frac{\exp\left(-z_s/\beta_s\right)}{\beta_s} + \frac{1 - f_{EM}}{\Gamma(\alpha_l)} \cdot \left(\frac{z_l}{\beta_l}\right)^{\alpha_l - 1} \cdot \frac{\exp\left(-z_l/\beta_l\right)}{\beta_l}\right\} \quad , \tag{2.16}$$

where the first term, the "short" component, describes the electromagnetic part of the shower and the second term, the "long" component, represents the pure hadronic part. Moreover, α_s, β_s, α_l and β_l are parameters describing the shower shape (depending on the energy of the shower) for the "short" (index s) and for the "long" component (index l). z_s is the depth in units of the radiation length for the "short" component and z_l in units of the interaction length for the "long" component. A fit using Equation (2.16) can be seen in Figure 2.7 (red), where the average energy deposit is plotted against the hit position in units of the interaction length. The data are taken from a shower initiated by a 200 GeV pion beam. The "short" (blue) and the "long" (green) component are also plotted individually. The fit parameters are shown in the legend.
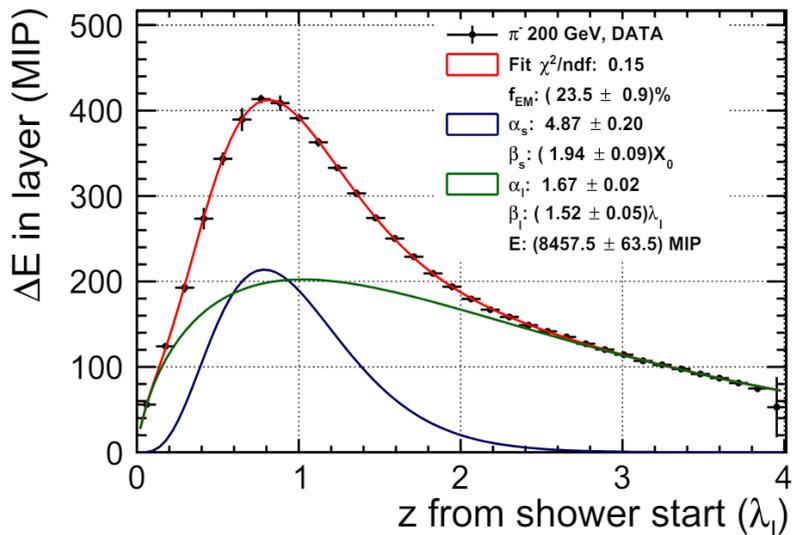
Figure 2.7.: Data and fit of the average energy deposit depending on the hit position z in units of the interaction length using Equation (2.16) for data of hadronic showers initiated by pions with $200\,\text{GeV}$ beam energy [22].

In the transversal direction, a radius of approximately the size of the interaction length [19],

$$R_{95\%} \approx \lambda_{\text{int}} \quad , \tag{2.17}$$

can be used to draw a cylinder around the shower, which contains about 95% of the total energy. Therefore, hadronic showers are larger in the transverse plane as well, compared to electromagnetic showers, since the Molière radius is proportional to the radiation length. The radial difference between electromagnetic and hadronic showers can be seen in the mean radius distribution in Figure 2.8, where the number of events is plotted against the event radius, which is the average distance of the hit to the centre of gravity. The plot shows data from showers initiated by electrons or pions at a beam energy of $60\,\text{GeV}$.
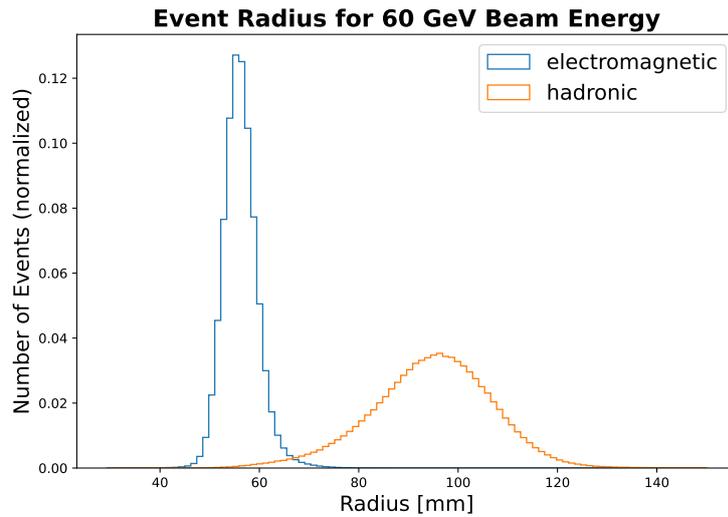
Figure 2.8.: Distribution of the mean radius with respect to the centre of gravity for electromagnetic and hadronic showers initiated by electrons and pions with a beam energy of 60 GeV, respectively, based on data from the AHCAL test beam run of 2018.

2.4. Calorimeters in Particle Physics

Since many particles cannot be detected directly due to their short lifetime, calorimeters are crucial in particle detection. In this way, the energy of the decay products of the initial particles can be measured, which enables physicists to reconstruct which particle initiated the shower. There are two main categories of calorimeters: sampling calorimeters and homogenous calorimeters.

In general, a sampling calorimeter consists of alternating active and passive materials. The shower development takes place in the passive material while in the active medium, measurable signals, like charged ions, Cherenkov- or scintillator light are generated. In some calorimeters, the active and the passive medium are combined into just one material. These calorimeters are called homogenous and are able to take data at any point in the calorimeter [19].

2.4.1. Electromagnetic Calorimeters

Electromagnetic calorimeters can measure the energy of an electromagnetic shower, which can be initiated by an electron, a positron, or a photon. For electromagnetic calorimeters, both homogeneous and sampling calorimeters can be used. While sampling calorimeters can be better sectioned, which results in a better position determination and particle identification, homogeneous calorimeters have a better energy resolution. The energy resolution for calorimeters in general can be parametrised by [19]

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c \quad , \tag{2.18}$$

where a, b and c are free parameters and where σ_E is the standard deviation of the energy E. The symbol $\oplus$ indicates addition in quadrature. The term with the parameter a describes stochastic fluctuations. Since the distribution of the number of shower particles N follows the Poisson distribution, the standard deviation is described by $\sqrt{N}$. The number of particles is then proportional to the beam energy, thus the standard deviation is proportional to $\sqrt{E}$. The parameter b describes the effect of electronic noise which is independent of the beam energy. Lastly, the term with c characterises imperfections of the calorimeter, which results in a linear dependence on energy, since more energy will remain undetected if more energy is put into the particle beam. The energy resolution in homogeneous calorimeters also depends on the active material used. Since sampling calorimeters have sections where it is not possible to measure energy deposits, the energy resolution is on average worse than that of a homogenous calorimeter. For example, the liquid argon calorimeter, a sampling calorimeter, from the H1 detector at HERA at DESY had an energy resolution of [27]

$$\frac{\sigma_E}{E} = \frac{11\%}{\sqrt{E/\text{GeV}}} \oplus \frac{150\,\text{MeV}}{E} \oplus 0.6\% \quad , \tag{2.19}$$

whereas a homogeneous calorimeter such as the $PbWO_4$ calorimeter of the CMS detector has an energy resolution of [28]

$$\frac{\sigma_E}{E} = \frac{2.8\%}{\sqrt{E/\text{GeV}}} \oplus \frac{120\,\text{MeV}}{E} \oplus 0.3\% \quad . \tag{2.20}$$

2.4.2. Hadronic Calorimeters

Hadronic calorimeters measure the energy of showers initiated by hadrons. One of the first characteristics of hadronic calorimeters is that they are almost always sampling calorimeters. In homogeneous calorimeters, the material used is very dense, such that the shower can develop, since it acts as a passive medium as well. As mentioned in Chapter 2.3 the hadronic interaction length is much larger than the radiation length, especially in dense materials, which would result in very large calorimeters if homogeneous ones were to be used. Therefore, usually homogeneous calorimeters are not used for hadronic showers, but sampling calorimeters instead [19].

A suitable passive medium for hadronic sampling calorimeters is iron. It has a density of $\rho = 7.87\,\mathrm{g\,cm^{-3}}$ and a comparatively small ratio of $\lambda_{\mathrm{int}}/X_0 = 9.5$, which indicates that electromagnetic and hadronic showers develop similarly in iron. This is necessary to keep the calorimeter compact and costs low [19]. For active media, typically scintillators or liquid argon are used. Liquid argon creates signals through ionisation, while scintillators will be excited by the energy of the particles entering the material, which is then emitted again as photons [19].

The energy resolution of a hadronic calorimeter can also be described via Equation (2.18). Another parametrisation is [29]

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E/\mathrm{GeV}}} + b\left(\frac{e}{h} - 1\right) \quad , \tag{2.21}$$

where σ_E is the standard deviation of the total energy E, a and b are real parameters and e/h is the ratio of the signal efficiency of the electromagnetic part of the shower (e) and the hadronic part (h), which is an intrinsic property of the calorimeter. The signal for a hadronic shower $S(\pi)$ and an electromagnetic shower $S(e)$ can be described via [19]

$$\frac{e}{\pi} := \frac{S(e)}{S(\pi)} = \frac{e/h}{1 - f_{\mathrm{em}}\left(1 - \frac{e}{h}\right)} \quad , \tag{2.22}$$

where it can be seen that the calorimeter response is directly dependent on the electromagnetic fraction of the shower energy. This implies that, if the electromagnetic fraction fluctuates strongly, fluctuations will occur in the electromagnetic and hadronic signals too, which ultimately has an impact on the energy resolution.

Equation (2.22) also shows that for $e/h = 1$, $e/\pi = 1$ follows. Therefore, a calorimeter with $e/h = 1$ generates the same signal for electrons and hadrons. Such calorimeters are called compensating calorimeters, and their energy resolution is given by the first term of Equation (2.21). Compensation can be achieved by adjustments in the choice of material of the active and passive media, as well as their thickness and the general construction of the calorimeter. Another way to accomplish compensation are software corrections [19].

The origin of non-compensating calorimeters having a ratio of $e/h \neq 1$ is non-measurable shower energy. Due to pion and kaon decays, neutrinos and muons are produced, which will escape the calorimeter. A fragment of neutrons, produced during nuclear reactions, does not leave signals in the calorimeter either. Consequently, the non-measurable energy fraction is part of the pure hadronic component of a hadronic shower, and that leads to a fraction of $e/h > 1$ [29]. Therefore, the energy resolution for non-compensating calorimeters is worse than for compensating ones. In compensating calorimeters, materials such as uranium are used, which are able to store the energy of the neutrons that would otherwise escape in non-compensating calorimeters [30].

Energy resolutions of non-compensating calorimeters are typically $\frac{50\%}{\sqrt{E/\mathrm{GeV}}}$, while the resolution of compensating calorimeters is better. For example, the energy resolution of a Pb/scintillator sandwich calorimeter is $\frac{\sigma_E}{E} \approx \frac{45\%}{\sqrt{E/\mathrm{GeV}}} + 1\%$, for a Pb/scintillator spaghetti calorimeter, which has scintillator fibres surrounded by the passive medium, it is $\frac{\sigma_E}{E} \approx \frac{28\%}{\sqrt{E/\mathrm{GeV}}} + 2.5\%$ and for a U/scintillator calorimeter it is $\frac{\sigma_E}{E} \approx \frac{35\%}{\sqrt{E/\mathrm{GeV}}} + 1\%$ [19].

The Analogue Hadron Calorimeter (AHCAL), which is a prototype for a hadronic calorimeter and the focus of this thesis, has an energy resolution for electromagnetic showers of [31]

$$\frac{\sigma_E}{E} = \frac{(21.9 \pm 1.4)\%}{\sqrt{E/\mathrm{GeV}}} \oplus \frac{58\,\mathrm{MeV}}{E} \oplus (1.0 \pm 1.0)\% \tag{2.23}$$

and for hadronic showers of [32]

$$\frac{\sigma_E}{E} = \frac{(57.6 \pm 0.4)\%}{\sqrt{E/\mathrm{GeV}}} \oplus \frac{0.18\,\mathrm{GeV}}{E} \oplus (1.6 \pm 0.3)\% \quad . \tag{2.24}$$

3. The CALICE Collaboration and the AHCAL Prototype

The CALICE Collaboration ("Calorimeter for Linear Collider Experiment") is an international collaboration concentrating on the research and development of highly granular calorimeters, including electromagnetic and hadronic calorimeters, for a future linear electron-positron collider, the International Linear Collider (ILC) [11, 12, 33]. In test beam programs, different groups of the collaboration test different calorimeter prototypes as well as a tail catcher and muon tracker [11]. This thesis focuses on a prototype for a hadronic calorimeter called the Analogue Hadron Calorimeter (AHCAL) which is described in Chapter 3.1. Two test beam campaigns for the AHCAL have already taken place in 2018 and 2022, respectively. The test beam run of 2022 is not part of this thesis.

3.1. The AHCAL Technological Prototype

The Analogue Hadron Calorimeter (AHCAL), shown in Figure 3.1 together with a tail-catcher, is a prototype from the CALICE Collaboration for a high granularity sampling calorimeter with steel as the passive and scintillator tiles as the active medium. The tiles are read out individually via silicon photomultipliers (SiPMs) directly attached to them. The information about the position of a particle in the calorimeter is provided, as well as information about the time-of-arrival of the particle on the scintillator tiles, with a resolution of the order of $1\,\text{ns}$. The prototype was built in 2017 [12].

Figure 3.1.: The AHCAL prototype with an additional tail-catcher [13].

The active parts of the AHCAL comprise 3 mm thick injection-moulded polystyrene scintillator tiles, each with an area of $3 \times 3 \, cm^2$, covered in reflective foil (to reduce optical cross talk between tiles). A tile and a foil covered tile can be seen in Figure 3.2. The SiPM is placed in a dimple in the centre of the tile. A base unit of the AHCAL (HBU, which stands for HCAL Base Unit) is $36 \times 36 \, cm^2$ wide and contains 144 tiles. An active layer of the size of $72 \times 72 \, cm^2$ is composed of four HBUs, and thus encompasses 24×24 tiles. One active layer module also shares a common set of interface boards, connected to power supplies and to the data acquisition system. An active layer from the top and from the bottom view is shown in Figure 3.3. The active layers are then arranged in a wedge shaped absorber structure. In total, 38 active layers are built into the AHCAL.
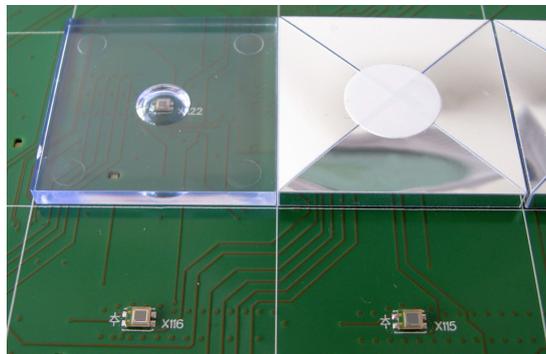
Figure 3.2.: Scintillator tile with one SiPM (left) and a foil covered scintillator tile (right) [13].

(a) *(b)*
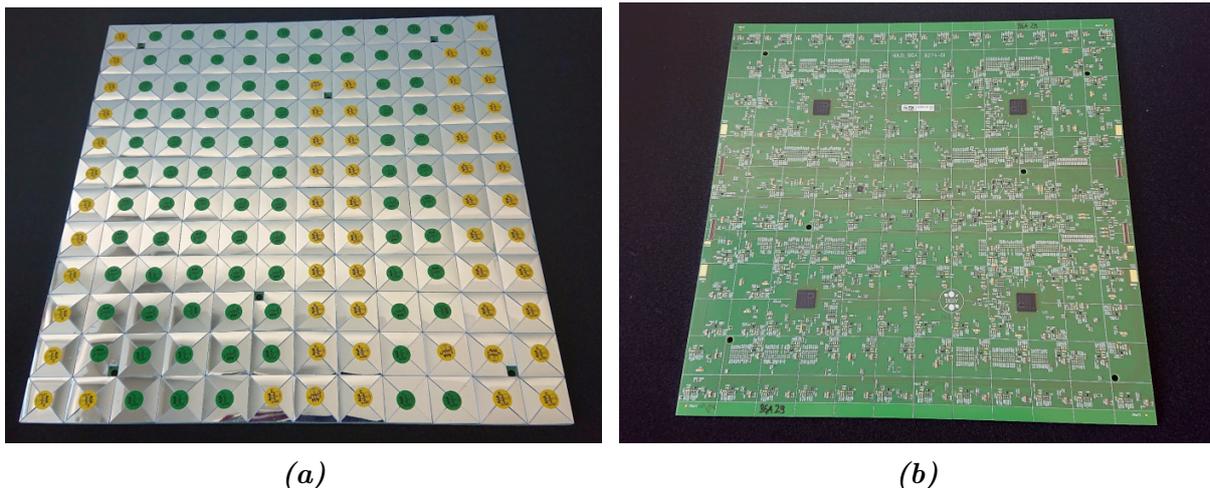
Figure 3.3.: HBU with scintillator tiles seen from the top (a) and from the bottom (b) [12].

An SiPM is a matrix consisting of usually 1000 pixels per $1\,\mathrm{mm}^2$, where each pixel is composed of a photodiode and a quenching resistor, which are operated just above break-down voltage [34]. SiPMs are based on single-photon avalanche diodes (SPADs). SPADs are composed of a p^+-i-n^+ photodiode and an additional slightly p-doped layer, whose general structure can be seen in Figure 3.4. Between the anode at the p^+ region and the cathode at the n^+ region, a voltage is applied. A photon entering the active material will be absorbed in the i-layer, which is the intrinsic region, where it induces the production of an electron-hole pair. The electron then moves in the direction of the cathode, due to the external field, and the positive hole to the anode and the p^+ layer. Once the electron reaches the p-doped layer, where the electric field suddenly increases, the electron accelerates, which enables it to create new electron-hole pairs. These newly formed pairs will continue with the electron-hole production as well, forming a cascade. This is why the p-doped and the following n^+ layer are referred to as the multiplication region [35]. The electric signal from the electrons at the n^+ layer and the hole at the p^+ layer can then be measured.

An electric signal within an SPAD could also be induced due to thermal electrons, which is considered to be thermal noise that forms the majority of the noise in a SiPM [34]. The rate of electrical signals created due to noise is called the dark count rate. The dark count rate can be minimised by operating the SiPMs at low temperatures [36].
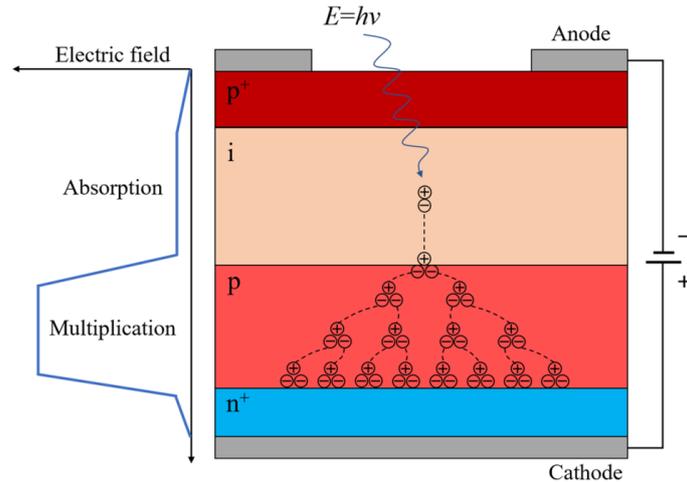
Figure 3.4.: A schematic model of an SPAD, which consists of a highly p-doped layer (p$^+$) followed by the intrinsic (i) and an additional slightly p-doped layer (p) connecting to an n-doped layer (n$^+$). At the p$^+$ layer and the cathode, a voltage is applied. The electromagnetic field strength within the layers is shown on the left-hand side [35].

The absorber material of the AHCAL is non-magnetic stainless steel whose ratio between the interaction and radiation length is comparatively small, as already mentioned in Section 2.4.2. One absorber layer has a thickness of 1.7 cm which corresponds to one radiation length or 0.1 interaction lengths. The whole absorber consists of 44 absorber layers with gaps for active layers in between. Combined with active layers, it measures a depth of 4.4 interaction lengths [12].

3.2. The Test Beam Run in 2018

The test beam run in 2018 was performed at the H2 beam line at the CERN Super Proton Synchrotron beam test facility and was divided into three periods. The first period was during the first two weeks of May. The prototype, including 38 active layers placed in the first 38 gaps of the absorber structure, was installed on a platform adjustable in the x-y-plane which was placed orthogonal to the beam axis. The event recording relied on the matching triggering of two external scintillators in the beam line. Data were taken for electrons with beam energies of 10 GeV, 20 GeV, 30 GeV, 40 GeV, 50 GeV, 60 GeV, 70 GeV, 80 GeV, 90 GeV and 100 GeV, for muons for energies between 40 GeV and 120 GeV and for negative pions in the range between 10 GeV and 160 GeV [12].

The second run took place during one week in June. The 38th active layer, which had

scintillator tiles of the size $3 \times 3 \, \text{cm}^2$, was replaced with a layer comprising $6 \times 6 \, \text{cm}^2$ tiles. Furthermore, the previous 38th layer was placed in the 41st gap of the absorber structure. This time, the AHCAL was tested together with a pre-shower layer in front and a tail catcher in the rear of the calorimeter prototype. The pre-shower layer was one HBU. The tail catcher consisted of 12 HBUs, which had a different tile geometry than those used in the prototype, alternating with $7.4 \, \text{mm}$ thick steel as absorber. Everything was placed on the adjustable platform. Data were taken for positrons with beam energies in the range between $10 \, \text{GeV}$ and $100 \, \text{GeV}$ in steps of $10 \, \text{GeV}$, for muons at a beam energy of $40 \, \text{GeV}$, and for negative pions for $10 \, \text{GeV}$, $20 \, \text{GeV}$, $30 \, \text{GeV}$, $40 \, \text{GeV}$, $60 \, \text{GeV}$, $80 \, \text{GeV}$, $120 \, \text{GeV}$, $160 \, \text{GeV}$ and $200 \, \text{GeV}$ [12].

The third period of the 2018 test beam campaign was in October. The AHCAL, with 39 layers, which includes 38 layers with $3 \times 3 \, \text{cm}^2$ tiles and one with $6 \times 6 \, \text{cm}^2$ tiles, was tested together with the prototype for the silicon part of the CMS High-Granularity Calorimeter (HGCAL), which was placed in front of the AHCAL. The system of both calorimeter prototypes was installed on a fixed platform, since it was too large to be mounted on the movable one. Due to the depth of the HGCAL being five interaction lengths, only tails of showers and muons could reach the AHCAL. In total, about 93 million events were recorded in 2018 [12].

4. Electromagnetic Shower Simulation with the Discrete Cosine Transformation

Simulations are an essential tool in modern high energy particle physics. Not only are simulations used to verify whether detectors function as expected or not, but they also allow for probing the SM and its interactions, as well as putting its theories to the test. Unfortunately these simulations are very time- and computing-resource-consuming due to large amounts of data that have to be simulated. Therefore, it is necessary to find alternatives that offer simpler simulation methods without losing important information about particle showers. Since electromagnetic subshowers are a crucial part of hadronic showers and have a significant impact on the energy resolution of hadronic calorimeters, their development in hadronic calorimeters, in particular highly granular ones such as the AHCAL test beam setup, needs to be simulated as well. Therefore, this thesis concentrates on the data reduction of electromagnetic showers in the AHCAL using the discrete cosine transformation in order to enable fast shower simulation methods to achieve good agreement with data within reasonable time and with high efficiency.

Section 4.1 describes how particle showers can be simulated via kernel density estimators. In Section 4.2 the discrete cosine transformation will be explained as well as how it will be used to reduce data. Section 4.3 addresses centre-of-gravity cuts simplifying the computation of the hit energy positions relative to the centre of gravity. Section 4.4 shows properties of the nodes and how these are used in the data reduction. Sections 4.5 and 4.6 present distributions of kinematic variables for several methods of random node generation. Finally, a comparison of hit energy data of a single event before and after the discrete cosine transformation is presented in Section 4.7.

4.1. Shower Simulation using Kernel Density Estimators

The kernel density estimation [37] is a method of describing distributions of data. It relies on estimating and approximating probability density functions (PDFs), which is particularly useful when working with very large amounts of data of which underlying PDFs are not known or have no analytical form. In one dimension, PDFs can, for instance, be approximated by the following kernel density estimator (KDE) [37]

$$f(x) = \frac{1}{nh} \sum_{i=1}^{n} \text{K}\left(\frac{x - x_i}{h}\right), \tag{4.1}$$

where x can be any physical variable of which n measurements $(x_1, x_2, ..., x_n)$ have been taken. The kernel K is a non-negative density function describing individual data points well enough. For example, a Gaussian distribution

$$\text{K}(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) \tag{4.2}$$

can be chosen. The free parameter h is the so-called bandwidth. It determines the smoothness of the estimated PDF. If h is chosen too small, the curve of the KDE will show every data point as a peak, in contrast to when h is chosen too big, the curve becomes flat, which results in the desired structure of the PDF being lost.

The KDE simulates the PDFs of the hit energy well, which can be seen in Figure 4.1a and 4.1b, where the distributions of the energy in layers 1 (4.1a) and 20 (4.1b) are shown for actual data of an electron beam with $60\,\text{GeV}$ beam energy and for the simulated energies using KDEs with a bandwidth of $h = 0.01$. For both plots, the simulation differs from the original data only in the form of small fluctuations. It was shown that global kinematic variables for hadronic showers are simulated well by KDEs [38], therefore it is assumed that the KDEs are able to simulate electromagnetic showers as well.

Even though some hits have been measured in layer 20, for most events no energy has been detected in this layer, which is indicated by the peak at $0\,\text{MIP}$. Therefore, an event usually comprises many hit energies per tile equal to zero. These values require a great amount of storage and are not crucial in describing the shower accurately, which complicates the simulation via KDEs a lot, since the large amount of input data results in very long computing-time, which is unpractical for a fast simulation. To make the simulation

more efficient, the simulation could be sped up by minimising the large amount of zero energy values simulated by the KDEs. One possible way to do that is to use transformations that transform the data into more manageable values, or even into a dataset of smaller size. Such transformations, ideally, are information-loss-free. If not, one must find ways to keep information loss as small as possible. How exactly the data reduction has been done in the context of this thesis is described in the following Sections.

(a) *(b)*
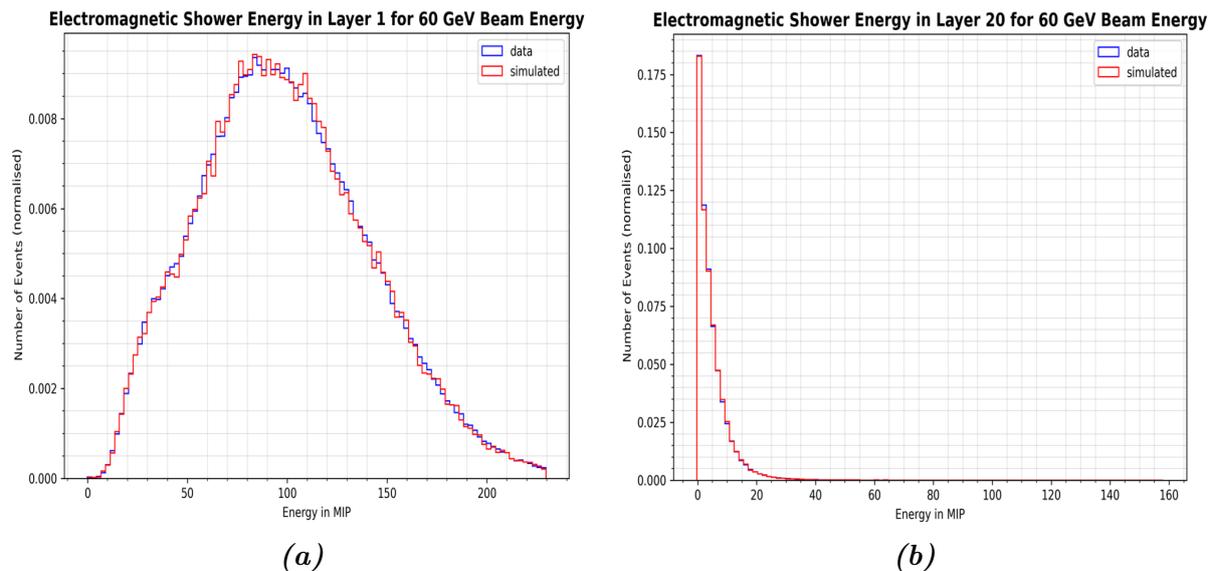
Figure 4.1.: Distribution of the energy deposited in layers 1 (a) and 20 (b) for data and simulated data with KDEs.

4.2. The Discrete Cosine Transformation

The discrete cosine transformation (DCT) is similar to the discrete Fourier transformation (DFT). It is a mathematical series used to describe the shape of a distribution via superposition of cosine nodes. The DCT can be derived from the DFT, which is described through

$$\tilde{X}_k = \sum_{n=0}^{N-1} x_n \cdot e^{-i\frac{2\pi k n}{N}} \quad , \tag{4.3}$$

where $\{\tilde{X}_k\} = \tilde{X}_0, \tilde{X}_1, \cdots, \tilde{X}_{N-1}$ and $\{x_n\} = x_0, x_1, \cdots, x_{N-1}$. The sequence $\{x_n\}$ will be mirrored and doubled, such that

$$\tilde{x}_n = \begin{cases} x_n, & 0 \leq n < N \\ x_{2N-1-n}, & N \leq n < 2N - 1 \end{cases} . \tag{4.4}$$

Described via the DFT, one then finds that

$$\tilde{X}_k = \sum_{n=0}^{N-1} x_n \cdot e^{-i\frac{\pi kn}{N}} + \sum_{n=N}^{2N-1} x_{2N-1-n} \cdot e^{-i\frac{\pi kn}{N}} \quad , \tag{4.5}$$

and using the index shift $(2N - 1 - n) \mapsto n$ in the second sum, it follows that

$$\tilde{X}_k = \sum_{n=0}^{N-1} x_n \cdot e^{-i\frac{2\pi kn}{2N}} + \sum_{n=0}^{N-1} x_n \cdot e^{-i\frac{2\pi k}{2N}\cdot(2N-1-n)} \quad . \tag{4.6}$$

It is then possible to rearrange Equation (4.6), using $\cos(x) = \frac{1}{2}\left(e^{ix} + e^{-ix}\right)$, into

$$\tilde{X}_k = e^{i\frac{\pi(k/2)}{N}} 2 \cdot \sum_{n=0}^{N-1} x_n \cos\left(\pi\frac{k}{N} \cdot \left(n + \frac{1}{2}\right)\right) \quad . \tag{4.7}$$

The discrete cosine transform is then defined as

$$X_k = \sum_{n=0}^{N-1} x_n \cos\left[\frac{\pi}{N}\left(n + \frac{1}{2}\right)k\right] \quad . \tag{4.8}$$

Here, x_n represents the data points, N is the number of data points and $k = 0, 1, \dots, N-1$ is the index of the coefficients X_k. The DCT coefficients quantify how strongly a specific cosine node is represented in the given distribution of data, and will henceforth be referred to as "nodes". This concept can be visualised in Figure 4.2. The cosine nodes shown in Figure 4.2b contribute differently to a probability density function (PDF), such as the blue curve in Figure 4.2a. The cosine nodes from 4.2b are also plotted together with the PDF in Figure 4.2a, which already allows to estimate that in this example, node 2 has the highest contribution, followed by node 1 and node 3.

A visualisation of the cosine nodes in two dimensions can be seen in Figure 4.3. Here, each square represents a specific two-dimensional standing cosine wave with the dark shades indicating minima and the light regions indicating maxima.
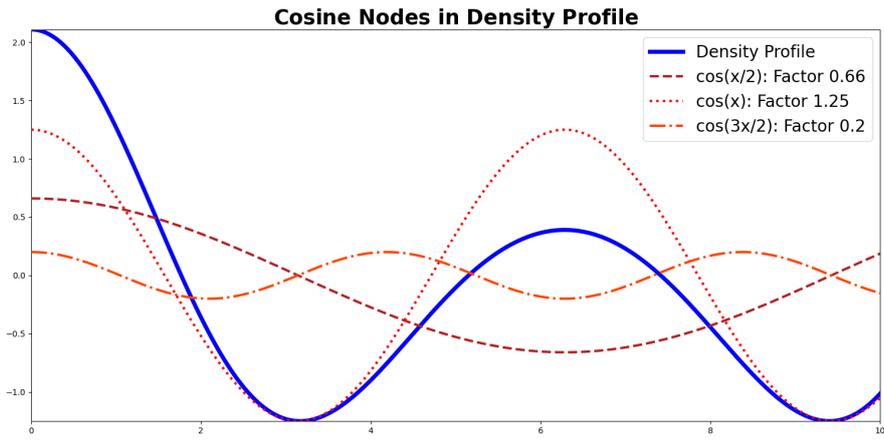
(a)
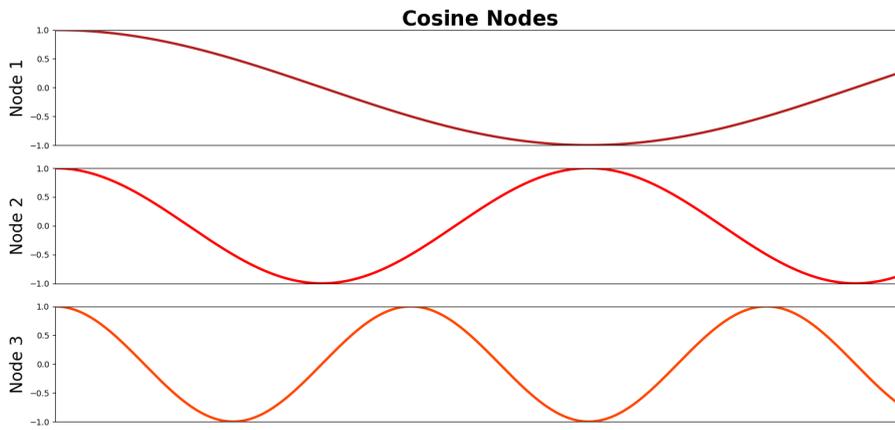
(b)

Figure 4.2.: Visualisation of cosine nodes in one dimension for an arbitrary variable.
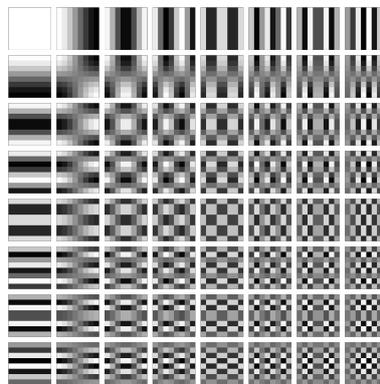
Figure 4.3.: Visualisation of cosine nodes in two dimensions, where the light shade represents maxima and the dark minima [39].

4. Electromagnetic Shower Simulation with the Discrete Cosine Transformation

The DCT is widely used for image compression. In order to use the DCT for data reduction for particle showers, the hit energies of a shower are inserted as the data points x_i. The resulting coefficients X_k describe the contribution of the cosine nodes to the shape of the hit energy distribution in the calorimeter. For the investigation of this thesis, cosine nodes with coefficients close to zero can be seen as noise and are generated randomly with the help of Gaussian distributions. Nodes with larger coefficients, on the other hand, would then be simulated using KDEs. The more nodes are generated randomly, the less nodes need to be simulated later on. After the generation and simulation of nodes, the nodes can be transformed back using the inverse DCT [40]

$$x_n = \frac{2}{N} \left(\frac{1}{2} X_0 + \sum_{k=1}^{N-1} X_k \cos \left[\frac{\pi}{N} \left(n + \frac{1}{2} \right) k \right] \right) \quad , \tag{4.9}$$

to obtain hit energies again.

The initial hit energy dataset contains one value for each tile of the calorimeter per event, which can be characterised by their position in the calorimeter in x, y and z direction. Here, z is the beam direction and x and y form the transverse plane. Because of this three-dimensional arrangement of the hit energy data, it is necessary to use the three-dimensional DCT and its corresponding inverse. One can obtain the three-dimensional transformations by multiplying three one-dimensional transformations. This way, the three-dimensional DCT results in the expression:

$$X_{k_1,k_2,k_3} = \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} \sum_{n_3=0}^{N_3-1} x_{n_1,n_2,n_3} \times$$
$$\cos \left[\frac{\pi}{N_1} \left(n_1 + \frac{1}{2} \right) k_1 \right] \cos \left[\frac{\pi}{N_2} \left(n_2 + \frac{1}{2} \right) k_2 \right] \cos \left[\frac{\pi}{N_3} \left(n_3 + \frac{1}{2} \right) k_3 \right] \tag{4.10}$$

and the three-dimensional inverse DCT in

$$x_{n_1,n_2,n_3} = \frac{8}{N_1 N_2 N_3} \epsilon_1 \epsilon_2 \epsilon_3 \sum_{k_1=0}^{N_1-1} \sum_{k_2=0}^{N_2-1} \sum_{k_3=0}^{N_3-1} X_{k_1,k_2,k_3} \times$$
$$\cos \left[\frac{\pi}{N_1} \left(n_1 + \frac{1}{2} \right) k_1 \right] \cos \left[\frac{\pi}{N_2} \left(n_2 + \frac{1}{2} \right) k_2 \right] \cos \left[\frac{\pi}{N_3} \left(n_3 + \frac{1}{2} \right) k_3 \right] \quad . \tag{4.11}$$

Here, the indices indicate the three spacial dimensions. The factors ϵ_i with $i \in \{1, 2, 3\}$ equal $1/2$ for $k_i = 0$, else they equal 1.

4.3. Shower Coordinates Relative to the Centre-of-Gravity

In order to be able to describe showers with varying shower starts and centres of gravity, it is necessary to reformulate the given hit energy data relative to these two variables. For this, one must subtract the shower start from the z- and the CoG in the transverse plane from the x- and y-coordinates of a hit. The CoG in x-direction is defined as

$$x_{\mathrm{CoG}} = \frac{1}{E_{\mathrm{tot}}} \sum_{\mathrm{hits}} E_{\mathrm{hit}} \cdot x_{\mathrm{hit}} \quad , \tag{4.12}$$

with analogous definitions for y_{CoG} and z_{CoG}. Here, E_{tot} is the total energy of the shower, E_{hit} the hit energy and x_{hit} the tile number of the hit (or the layer number in z-direction). The distributions of the CoG in the transverse plane for showers initiated by electrons for beam energies of $10\,\mathrm{GeV}$, $50\,\mathrm{GeV}$ and $80\,\mathrm{GeV}$ are shown as the two-dimensional histograms in Figure 4.4. The axes show the tile number in x- or y- direction and the colour scale, which varies for each plot, indicates the probability of the CoG to lie in a specific tile.

(a)

(b)

(c)
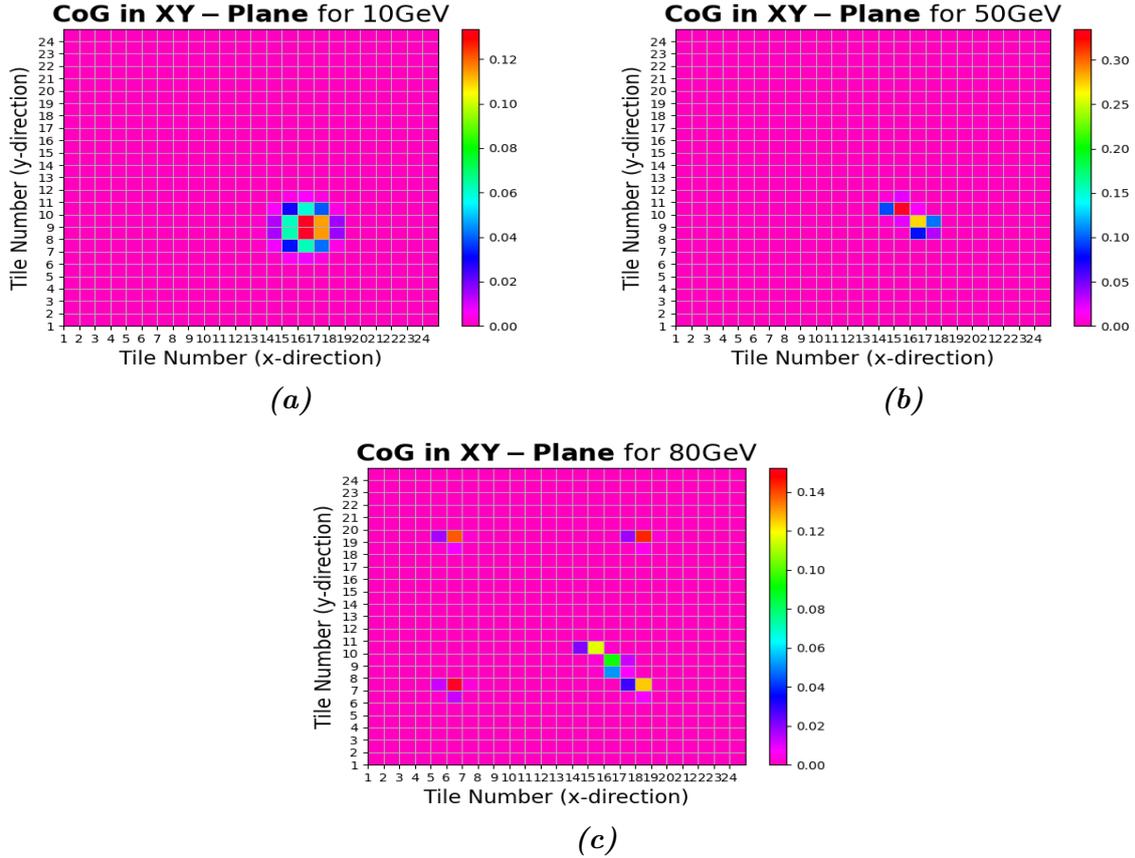
Figure 4.4.: Centre of Gravity in the transverse plane for electromagnetic showers initiated by electrons with beam energies of $10\,\mathrm{GeV}$ (a), $50\,\mathrm{GeV}$ (b) and $80\,\mathrm{GeV}$ (c).

31

In the plots for $10\,\mathrm{GeV}$ and $50\,\mathrm{GeV}$ the probability is the highest for the CoG to be in the lower right part of the calorimeter. It is noticeable that the plot for $80\,\mathrm{GeV}$ electrons shows four distinct clusters with non-zero CoGs. This is likely due to movement of the test beam during the test beam run. Additional plots of the CoG for the energies $20\,\mathrm{GeV}$, $30\,\mathrm{GeV}$, $40\,\mathrm{GeV}$ and $60\,\mathrm{GeV}$ can be seen in the Appendix in Figure A.1.

New positions for the hit energies relative to the CoG will be computed using

$$x_\mathrm{rel} = x_\mathrm{hit} - x_\mathrm{CoG} \quad \text{and} \tag{4.13}$$

$$y_\mathrm{rel} = y_\mathrm{hit} - y_\mathrm{CoG} \quad . \tag{4.14}$$

Here, x_rel, y_rel indicate the relative position, x_hit, y_hit the actual hit position and x_CoG, y_CoG the CoG in x/y-direction. The new positions will be referred to as "relative tiles". For the z-direction, the position of the shower start will be subtracted from the hit position:

$$z_\mathrm{rel} = z_\mathrm{hit} - z_\mathrm{ss} \quad . \tag{4.15}$$

Here, z_rel is the relative position, z_hit the hit position and z_ss the layer of the shower start in z-direction. The redefinition of the hit energy positions results in a number of relative tiles differing from the actual number of tiles of the calorimeter prototype. The number of relative tiles is then limited to 23 relative tiles in the x-direction and 23 tiles in the y-direction (11 left of the CoG, 11 right of it, and the CoG tile itself).

For the following investigation, only data from the $50\,\mathrm{GeV}$ electron beam and events around the most probable CoG in the transverse plane, which is within a 4×4 tile block, are considered. This includes the tiles 14 to 17 in the x-direction and 8 to 11 in the y-direction.

4.4. Properties of DCT Nodes

The purpose of data reduction is to improve the computing-time of KDEs while minimising information loss at the same time. With a large enough number of nodes being generated as random noise instead of, based on their actual distributions, being obtained through KDEs, this goal can be achieved. Therefore, it is necessary to investigate which nodes are suitable for random noise generation.

After applying the DCT, each event still has the same dimensions as before the DCT,

which means that the total number of nodes equals the number of tiles. Thus, it is still possible to assign an x, y and z value to every node, acting as a "position". In Figure 4.5a, the mean absolute nodes per tile are shown for all x- and y-values and for $z = 0$. It is evident that a significant number of nodes possesses a rather small value. These nodes are generally assumed to be fitting candidates for random noise generation. For higher z-values, even more nodes feature rather small mean values, which can be seen in Figure 4.5b where the means of absolute nodes are shown for $z = 10$.

It is important to only consider means of absolute nodes and not of these of the signed values of nodes, since the distributions of nodes often exhibit a double peak structure, as shown in Figure 4.6, where the distribution for the node with $x = 4$, $y = 22$ and $z = 0$ is shown. This double peak structure would yield a mean value near zero, even though the probability for the node having a larger value is much larger too. Hence, only absolute nodes are randomly generated, with their signs being separately determined.

(a)

(b)
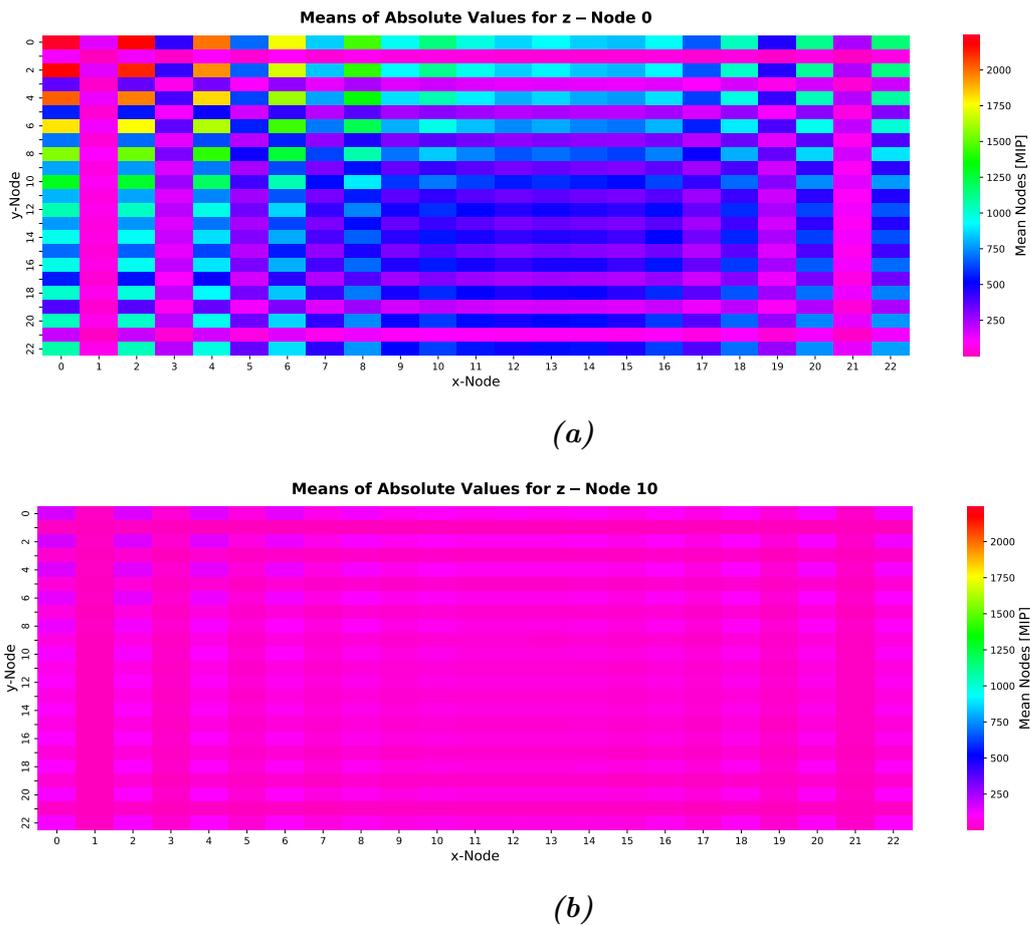
Figure 4.5.: Mean absolute nodes per tile in x- and y-direction for $z = 0$ (a) and $z = 10$ (b).
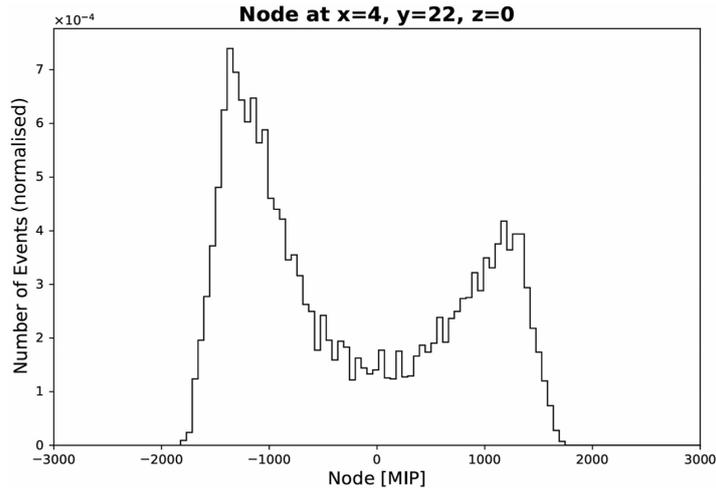
Figure 4.6.: Node distribution of the node at $x = 4$, $y = 22$ and $z = 0$. The node takes on large absolute values on an event-to-event basis, which average to zero, though, for many events.

To begin, all nodes with an absolute value less than or equal to a certain threshold are generated as random noise. If the data with randomly generated noise still matches the original well, the threshold can be increased until acceptable agreement is not given anymore. This investigation is discussed in Section 4.5. An initial threshold of 10 MIP has been chosen. This means all nodes within $(-10, 10)$ MIP are generated randomly using appropriate Gaussian distributions. To elaborate on this, Figure 4.7 shows the means of the absolute nodes sorted by their magnitude. This plot shows that only approximately 3000 nodes have a mean absolute value of ≤ 10 MIP, which is a small fraction out of the total ~20000 nodes.
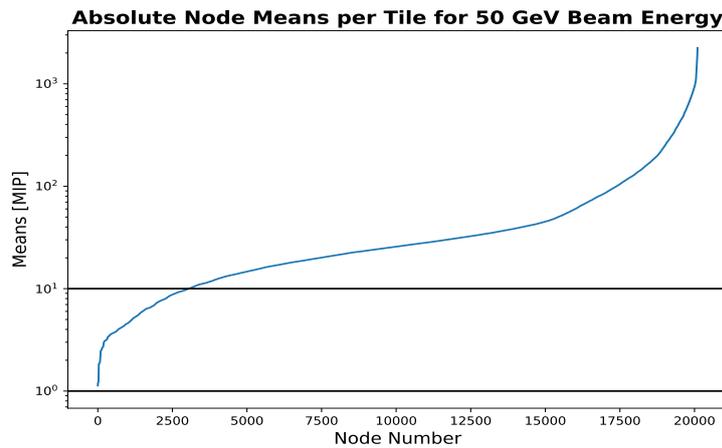
Figure 4.7.: Sorted mean absolute node values. Horizontal lines for orientation are shown for 1 MIP and 10 MIP.

In Figure 4.8, the distribution of the actual number of nodes within $(-10, 10)$ MIP per event is shown, where it becomes clear that an event has on average approximately 7000 nodes that lie in the given interval.

Number of Events per Number of Nodes within (-10,10) for 50 GeV
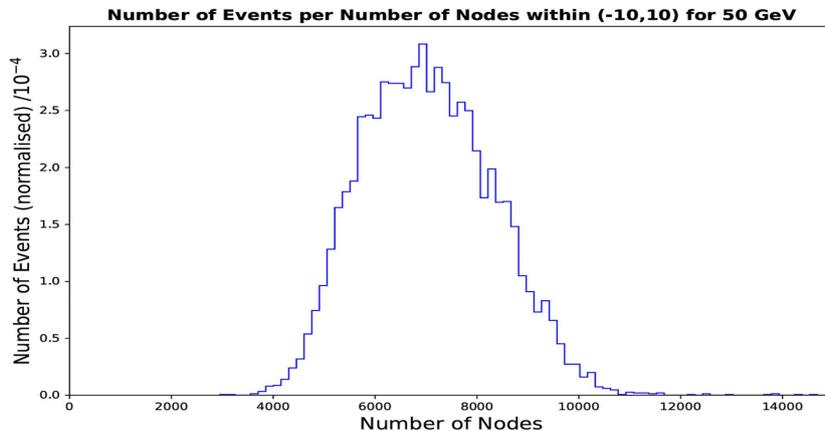
Figure 4.8.: Distribution of the number of nodes within $(-10, 10)$ MIP per event.

4.5. Generating DCT Nodes Randomly and Kinematic Shower Variables

As already mentioned in Section 4.4, a possible approach to decrease the amount of hit energy data is to generate all of those node amplitudes randomly that are smaller than a chosen threshold by using appropriate Gaussian distributions. These nodes would be small enough to be considered noise. Nodes with larger values are taken directly from data. This allows to investigate which nodes can be simulated in a random fashion. As a first attempt, all node amplitudes between $(-10, 10)$ MIP were generated randomly.

The impact of the noise generation can be seen in Figure 4.9 where distributions of the original data and with randomly generated noise can be seen for the total energy in Figure 4.9a, the energy-weighted mean shower radius in Figure 4.9b and the CoG in z-direction in Figure 4.4a. The total energy of a shower is defined as the sum over all hit energies and the energy-weighted mean shower radius $\bar{r}$ is described through

$$\bar{r} = \frac{1}{E_{\text{tot}}} \sum_{\text{hits}} r_{\text{hit}} \cdot E_{\text{hit}} \quad , \tag{4.16}$$

where r_{hit} is the radial distance of the hit to the CoG in millimetres. Here, it was assumed that the hit and the CoG are located in the centre of a tile. Furthermore, the distributions

of the shower moments in z-direction can be seen in Figure 4.9d for the variance, the skewness in Figure 4.9e and the kurtosis in Figure 4.9f. The variance μ_2 in z-direction is defined as

$$\mu_2 = \frac{1}{E_{\text{tot}}} \sum_{\text{hits}} (z_{\text{hit}} - z_{\text{CoG}})^2 \quad , \tag{4.17}$$

the skewness μ_3 as

$$\mu_3 = \frac{1}{E_{\text{tot}}} \sum_{\text{hits}} \left(\frac{z_{\text{hit}} - z_{\text{CoG}}}{\sigma} \right)^3 \quad , \tag{4.18}$$

and the kurtosis μ_4 as

$$\mu_4 = \frac{1}{E_{\text{tot}}} \sum_{\text{hits}} \left(\frac{z_{\text{hit}} - z_{\text{CoG}}}{\sigma} \right)^4 \quad . \tag{4.19}$$

Here, z_{hit} is the layer number of the hit, z_{CoG} the position of CoG in z-direction and σ the standard deviation. These moments can be analogously computed for the x- and y-direction.

The noise generation is divided into two regions. All nodes with a value between $-1\,\text{MIP}$ and $1\,\text{MIP}$ are generated using a Gaussian distribution with a mean value of $-1.03 \cdot 10^{-4}\,\text{MIP}$ and a standard deviation of $0.571\,\text{MIP}$. For nodes within $(-10, -1] \cup [1, 10)\,\text{MIP}$, the Gaussian distribution has a mean of $2 \cdot 10^{-6}\,\text{MIP}$ and a standard deviation of $5.95\,\text{MIP}$.

The mean total energy from Figure 4.9a is ~2250 MIP, and the majority of events has a total energy between 2000 MIP and 2500 MIP. The data before and after the DCT is nearly symmetrically spread around the mean, although events with a total energy below 2000 MIP appear as well, whereas almost no events with total energy above 2750 MIP exist. Both curves for the mean event radius from Figure 4.9b feature a distribution with two peaks, where the first one is located at 15 mm and the second one between 30 mm and 40 mm. The CoG in z-direction in Figure 4.9c has a peak around the sixth and seventh layer for both distributions. All shower moments feature a peak structure as well. The mean variance lies between 0.3 MIP and 0.35 MIP, the mean skewness between 5.0 and 6.0 and the mean kurtosis between 150 and 200. In all plots, the general shape of the distributions with randomly generated values matches the shape of the data distributions. Differences only appear in the form of small fluctuations. This can also be seen in the shower moment distributions and CoG in the transverse plane in Figures A.2 and A.5 in the Appendix.

(a)

(b)

(c)

(d)

(e)

(f)
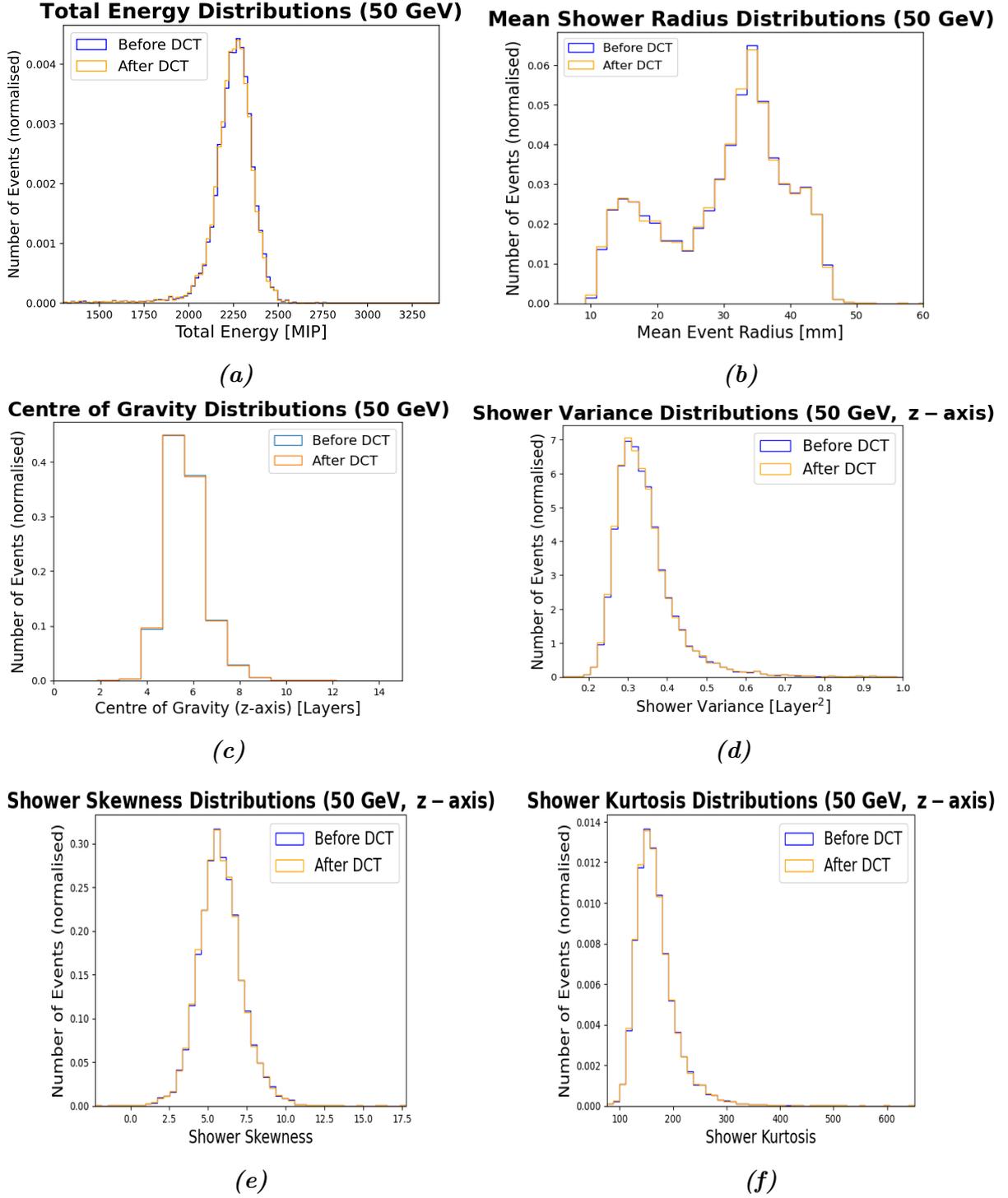
Figure 4.9.: Distributions of the total energy (a), the energy-weighted mean shower radius (b), the CoG in z-direction (c), the variance (d), the skewness (e) and the kurtosis (f). The shower moments describe the shower in the z-direction. The original data is shown in blue and the data with generated noise within $(-10, 10)$ MIP in orange.

4. Electromagnetic Shower Simulation with the Discrete Cosine Transformation

As shown in Section 4.4, only about 7000 nodes out of in total approximately 20000 nodes per event are generated randomly. To increase the number of randomly generated nodes, the threshold can be increased. Hence, the noise was now generated with three different Gaussian distributions. The first two distributions have the same parameters and are used for the same intervals as before. The third generates nodes within $(-20, -10] \cup [10, 20)$ MIP with a mean of 15 MIP and a standard deviation of 18 MIP. This time, the parameters for the third Gaussian distribution were estimated using Figure 4.7 for the mean and from Figure 4.10 for the standard deviation, which shows the standard deviations of all nodes sorted from smallest to largest mean node value.
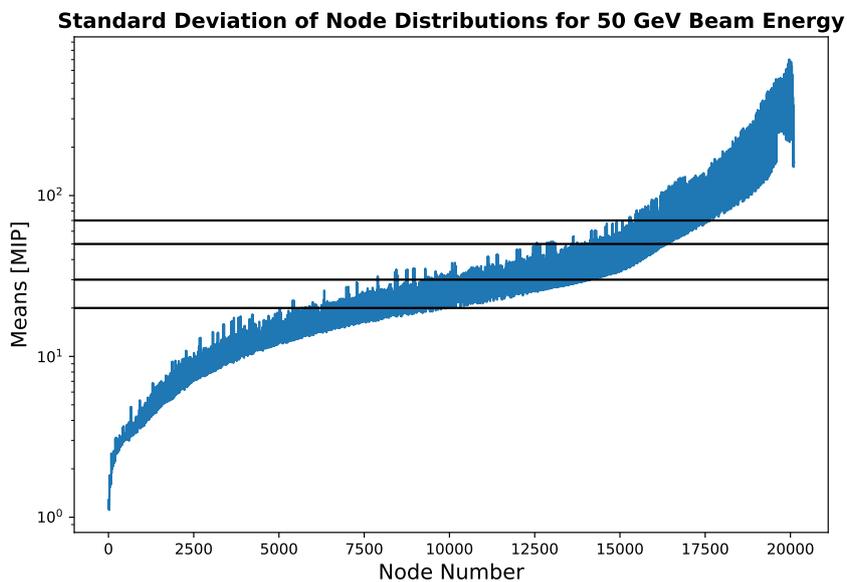
Figure 4.10.: Standard deviations of the node distributions, sorted from smallest to largest mean node value. Horizontal lines for orientation are shown for $20, 30, 50$ and 70 MIP.

Figure 4.11 also shows the distributions of the total energy (Figure 4.11a), the energy weighted shower radius (Figure 4.11b), the CoG in z-direction (Figure 4.11c), the variance in z-direction (Figure 4.11d) as well as the skewness (Figure 4.11e) and the kurtosis (Figure 4.11f) for generated noise within $(-20, 20)$ MIP. The overall shape seems to be conserved, but this time the curves with generated nodes are shifted to the right for the total energy and the shower radius as well as for all shower moments in z-direction. The CoG did not shift, but the fluctuations increased in comparison to the plot for noise between -10 MIP and 10 MIP in Figure 4.9c. Plots for the shower moments and the CoGs in the transverse plane can be seen in Figures A.2 and A.6 in the Appendix.
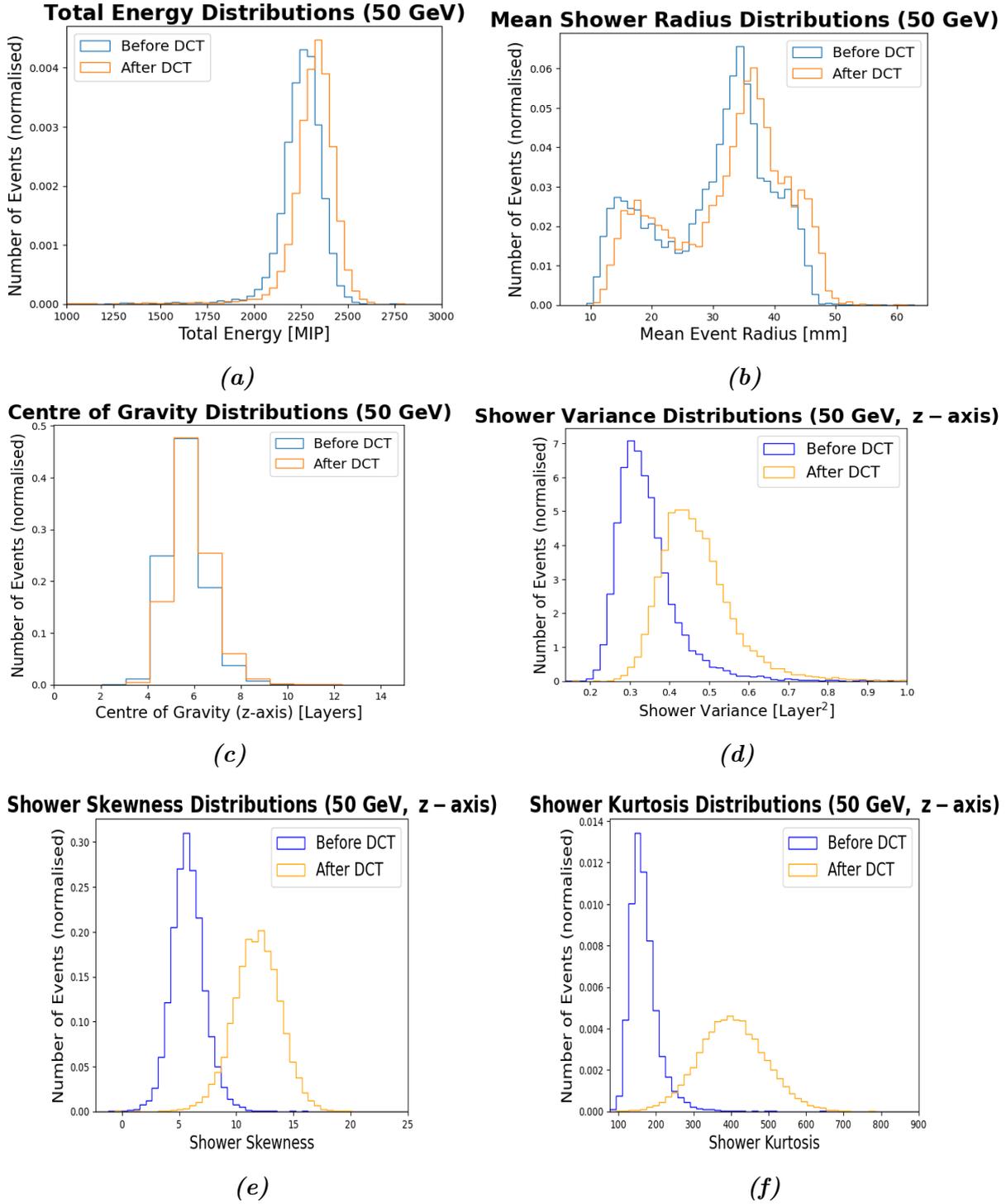
Figure 4.11.: Distributions of the total energy (a), the energy weighted mean shower radius (b), the CoG in z-direction (c), the variance (d), the skewness (e) and the kurtosis (f). The shower moments describe the shower in the z-direction. The original data is shown in blue and the data with generated noise within $(-20, 20)$ MIP in orange.

4. Electromagnetic Shower Simulation with the Discrete Cosine Transformation

In order to investigate at which threshold the random node generation will not fit the data anymore, one more set of plots was made using this method. This time, all nodes within $(-30, 30)$ MIP were considered to be noise. To do that, again the same Gaussian distributions were used for the nodes within $(-10, 10)$ MIP as previously. For the interval $(-30, -10] \cup [10, 30)$ MIP the Gaussian distribution now used a mean of 18 MIP and a standard deviation of 20 MIP. In Figure 4.12 the distributions of the kinematic variables as well as the shower moments in z-direction are shown for data and for data with simulated noise within $(-30, 30)$ MIP.

This time, almost all distributions are shifted to the right to an amount such that data and the curves with generated noise have only a small or almost none overlap. The shape of the curves of data with generated noise differs more from the original data for the total energy, the shower radius, the CoG and the variance. The two peaks from the mean radius have become only one peak with a wider distribution. The distributions of the total energy and the CoG in z-direction are wider and have a smaller peak than the original data. In contrast, the distribution of the variance in z-direction has a higher peak compared to the actual data. The shape of the skewness and the kurtosis in z-direction seem to have improved from the plots with noise between -20 MIP and 20 MIP in Figure 4.11, although here the curves are clearly shifted as well. In general, it seems that the limit up to which threshold nodes can be generated randomly and still show acceptable results lies between 20 MIP and 30 MIP.
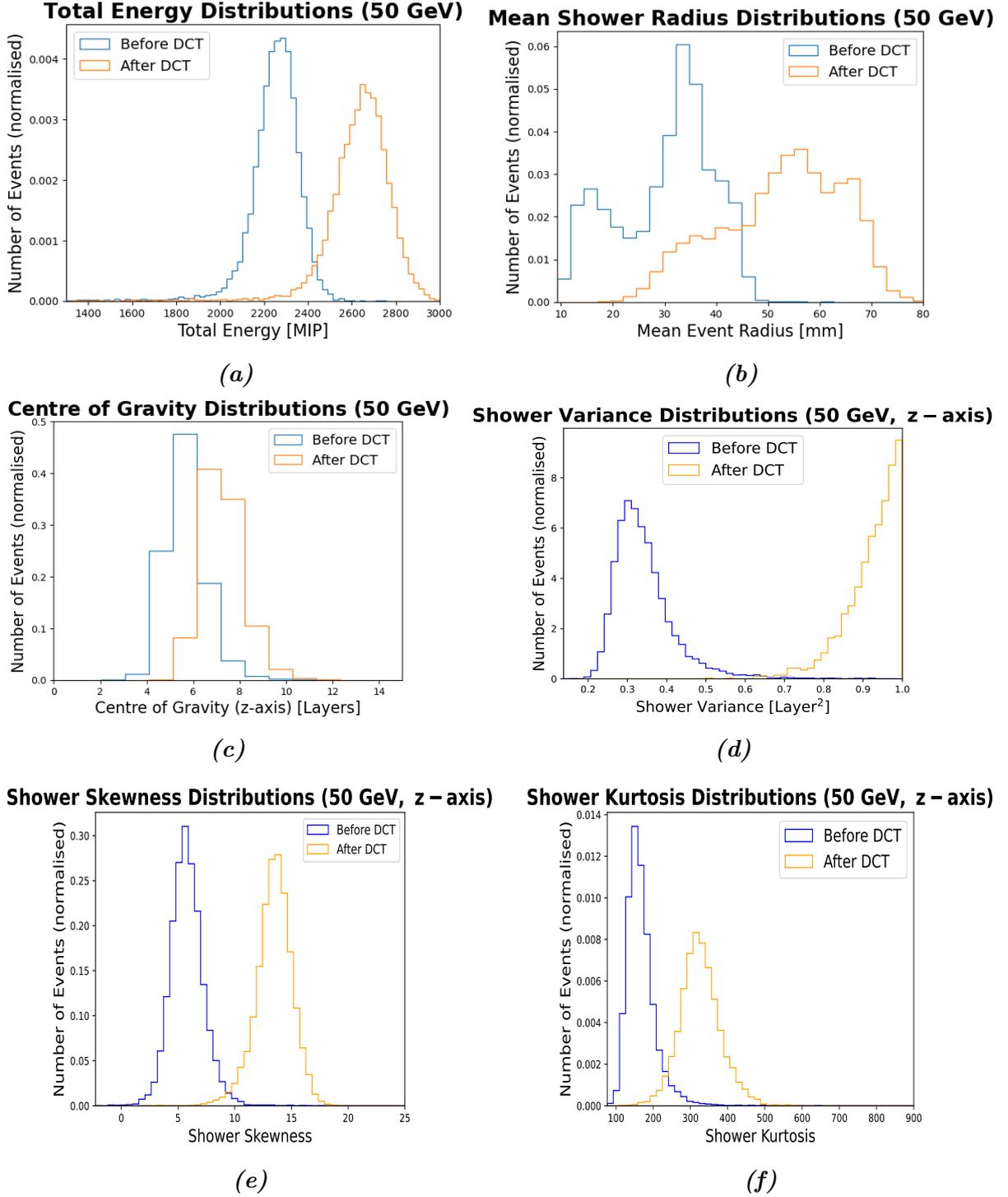
Figure 4.12.: Distributions of the total energy (a), the energy weighted mean shower radius (b), the CoG in z-direction (c), the variance (d), the skewness (e) and the kurtosis (f). The shower moments describe the shower in the z-direction. The original data is shown in blue and the data with generated noise within $(-30, 30)$ MIP in orange.

4.6. Generating Odd Nodes randomly

Another way of choosing nodes for the random generation becomes apparent by inspecting Figure 4.5a. It is noticeable that the nodes where x and y are both odd have a much smaller mean value than those where one or especially both coordinates are even. Thus, the next method is to generate all nodes where x and y are odd, henceforth referred to as "odd nodes", up to ± 30 MIP. This was the threshold where the previous node generation method stopped to yield results that matched data distributions of the original dataset. The parameters for the Gaussian distributions are chosen to be the same as for the node generation between $(-30, 30)$ MIP for even and odd nodes. This time the nodes are not checked individually, though, whether their value lies within the required interval. Instead, all nodes, identified by their x-, y- and z-value, whose mean absolute value is within $[0, 30)$ MIP will be generated randomly via appropriate Gaussian distributions. This leaves the number of nodes that are generated randomly unchanged from event to event, which was not the case previously. Furthermore, the constant number of randomly generated nodes is also important because KDEs also require a constant number of input values.

The distributions of the kinematic variables are shown in Figure 4.13. The plots for the shower moments and the CoGs in the transverse plane are shown in the Appendix in Figures A.8 and A.12.

The kinematic distributions of the data after the DCT are all shifted to the right-hand side. For the total energy, the skewness and kurtosis in z-direction, the curves have smaller maxima and are all wider than the original PDFs. The distribution of the variance in z-direction is much wider than the curve from before the DCT. The shape of the mean shower radius after the DCT differs a lot from the actual data, since it now only shows one peak, which is also larger than the original ones. Therefore, the random generation of only odd nodes does not seem to work better than the random generation for all nodes within $(-30, 30)$ MIP.

(a)

(b)

(c)

(d)

(e)

(f)
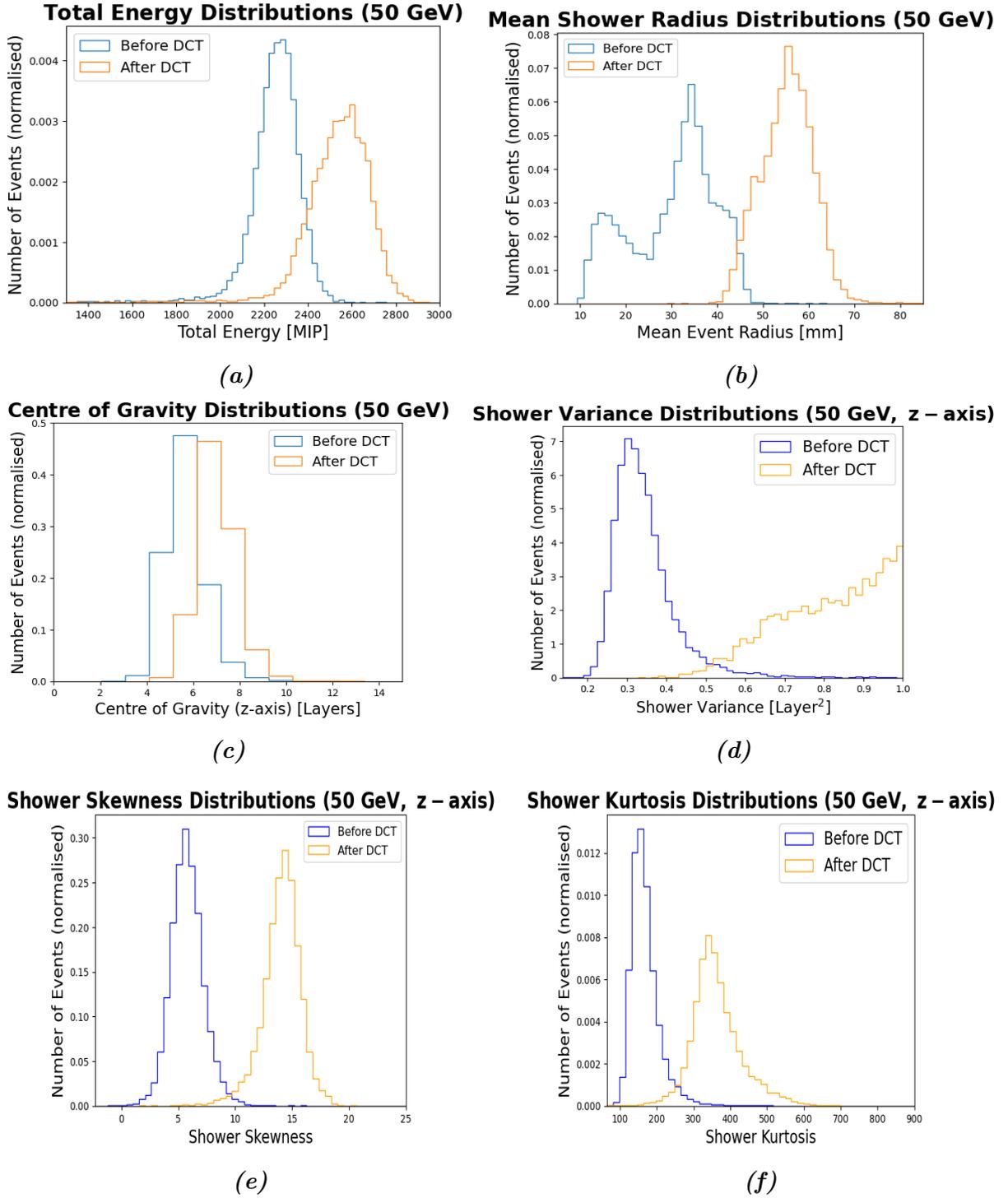
Figure 4.13.: Distributions of the total energy (a), the energy weighted mean shower radius (b), the CoG in z-direction (c), the variance (d), the skewness (e) and the kurtosis (f). The shower moments describe the shower in the z-direction. The original data is shown in blue and the data with generated noise within $(-30, 30)$ MIP for nodes with odd x- and y-values in orange.

Another method is to generate single rows or columns from Figure 4.5a randomly. To start off, only nodes where $(x = 21 \land y \text{ odd})$ or $(x \text{ odd } \land y = 21)$ are generated randomly. This time though, only one Gaussian distribution is used for all nodes which fulfil the given requirements. The mean and the standard deviation are chosen based on Figures 4.14 and 4.15. The former shows the mean value of the absolute nodes where $(x = 21 \land y \text{ odd})$ or $(x \text{ odd } \land y = 21)$ sorted from lowest to highest. The latter depicts values of the standard deviation plotted against the node number. Based on these plots, 30 MIP for the mean and 15 MIP for the standard deviation have been deemed as appropriate parameters for the Gaussian noise distribution.

The distributions of the kinematic variables for data before and after the DCT can be seen in Figure 4.16. The plots for the shower moments and the CoG in the transverse plane are shown in the Appendix in Figures A.9 and A.13.

The distribution of the CoG in z-direction after the DCT seems to match the original data well. Only small fluctuations are visible. The distribution of the variance is only slightly shifted to the right, whereas the shape of the curve has not changed significantly. The other shower moments (skewness and kurtosis), feature a much smaller peak and a wider distribution than the original data. Both curves are shifted even more. The total energy after the DCT deviates, but not strongly, from data on the left-hand side of the distribution, as the maximum peak is smaller than before the DCT. However, on the right-hand side, a small but noticeable second peak has formed. The same description fits the mean radius, where the additional peak is even more pronounced.
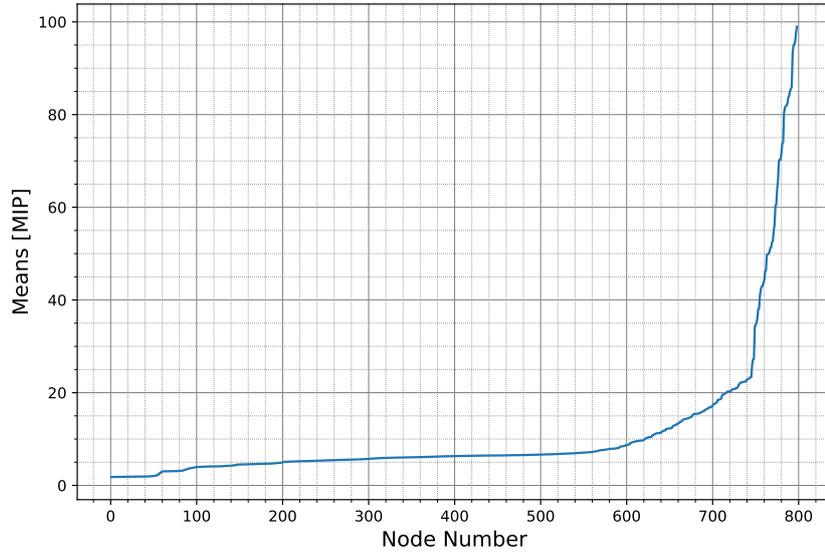
Figure 4.14.: Mean absolute values of all nodes that satisfy $(x = 21 \land y$ odd$)$ or $(x$ odd $\land y = 21)$.
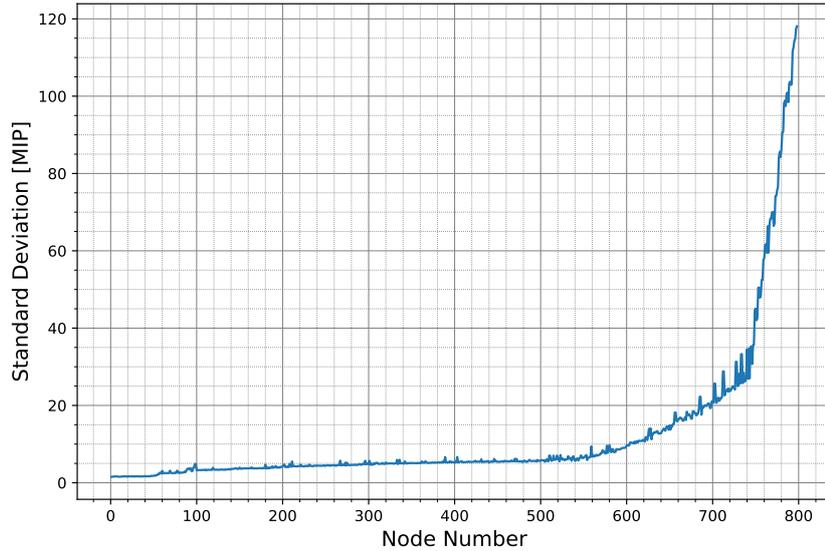
Figure 4.15.: Standard deviations of all nodes that satisfy $(x = 21 \land y$ odd$)$ or $(x$ odd $\land y = 21)$.

(a)

(b)

(c)

(d)

(e)

(f)
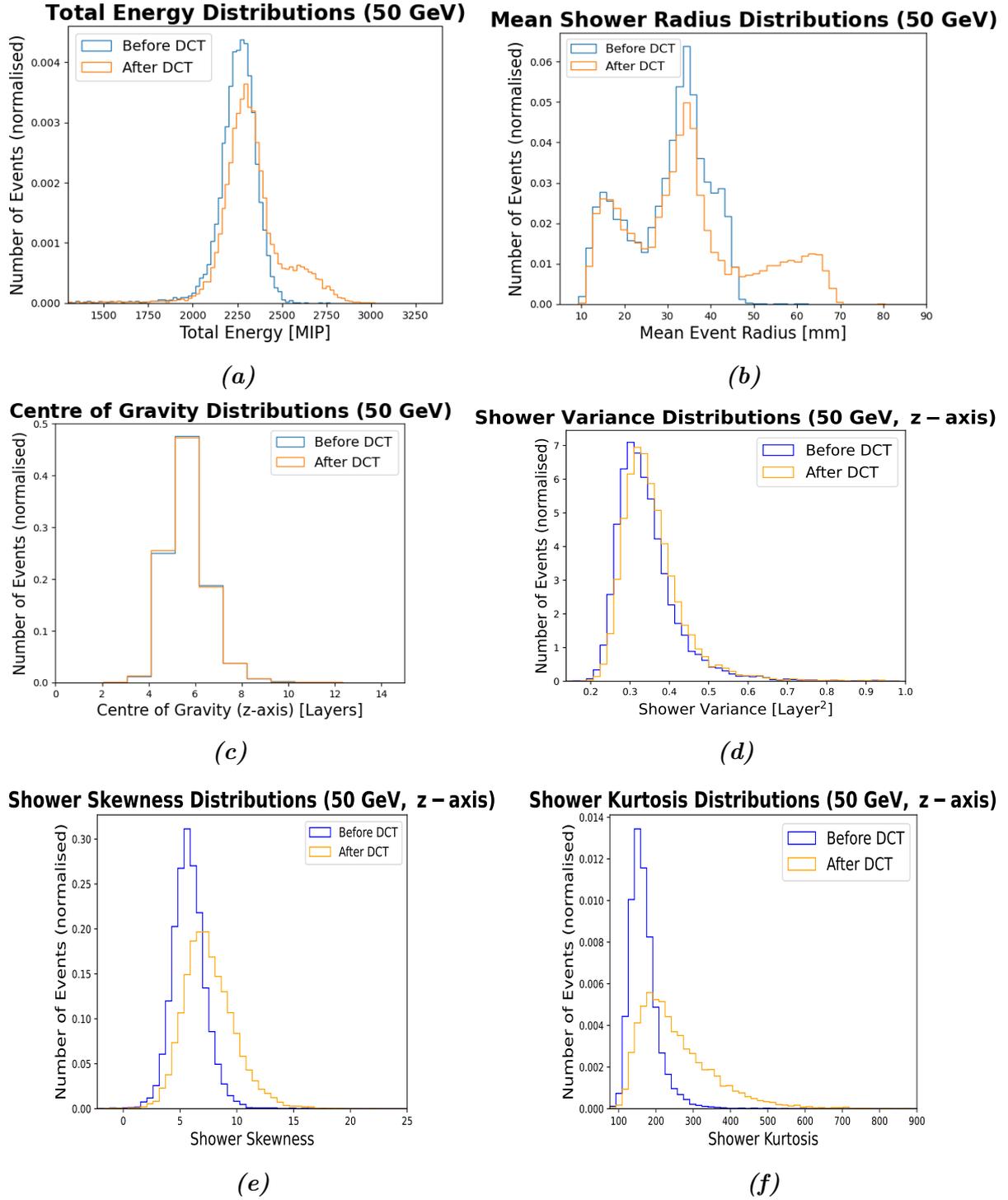
Figure 4.16.: Distributions of the total energy (a), the energy weighted mean shower radius (b), the CoG in *z*-direction (c), the variance (d), the skewness (e) and the kurtosis (f). The shower moments describe the shower in the *z*-direction. The original data is shown in blue and the data with generated noise (one Gaussian distribution) for nodes with ($x = 21 \wedge y$ odd) or (x odd $\wedge y = 21$) in orange.

To obtain better results, in particular for the total energy and the energy-weighted mean shower radius, noise was generated randomly for the same node configuration as previous, but with three Gaussian distributions. Again, all nodes between $-10\,\text{MIP}$ and $10\,\text{MIP}$ are generated with Gaussian distributions with the same parameters as before. For all nodes with a higher absolute value than $10\,\text{MIP}$, a mean of $30\,\text{MIP}$ and a standard deviation of $30\,\text{MIP}$ have been chosen. The distributions of the kinematic variables can be seen in Figure 4.17. The shower moments and CoGs in x- and y-direction are shown in the Appendix in Figures A.11 and A.14.

The distributions after the DCT for the shower moments in z-direction have improved compared to the distributions where only one Gaussian distribution has been used. The shapes of the curves match the data and only differ in small fluctuations. The CoG in z-direction still exhibits the same shape as the data, but the fluctuations seem to be a little stronger than in the previous figure. The distributions of the total energy and the energy-weighted mean shower radius did not change noticeably. The problem of the additional peaks in the distribution of the total energy and the energy-weighted mean shower radius still remains.

The last method that has been investigated was to set all nodes with ($x = 21 \wedge y$ odd) or (x odd $\wedge y = 21$) to zero. The distributions of the kinematic variables for this method are shown in Figure 4.18. The shower moments and the CoG in x- and y-direction can be seen in the Appendix in Figures A.9 and A.15. The curves of the data after the DCT did not change noticeably. Only the fluctuations of the shower variance and skewness in z-direction seem to have decreased slightly. Moreover, replacing the random noise with zeros did not improve the results and the additional peak in the total energy and the mean shower radius PDFs still persists.

It turns out that the random noise generation yields better results for specific odd rows or columns in Figure 4.5 than for all odd nodes up to a threshold of $30\,\text{MIP}$. Using three Gaussian distributions for random noise generation also improves the results too, compared to only one Gaussian function. Furthermore, setting the same nodes to zeros works just as well. However, for the last three configurations, an additional peak became visible in the distributions of the total energy and the mean shower radius, respectively. Hence, if the origin of these peaks was to be investigated further, one could possibly improve the method significantly for which nodes that satisfy ($x = 21 \wedge y$ odd) or (x odd $\wedge y = 21$) were either replaced by zero or a random value.

(a)

(b)

(c)

(d)

(e)

(f)
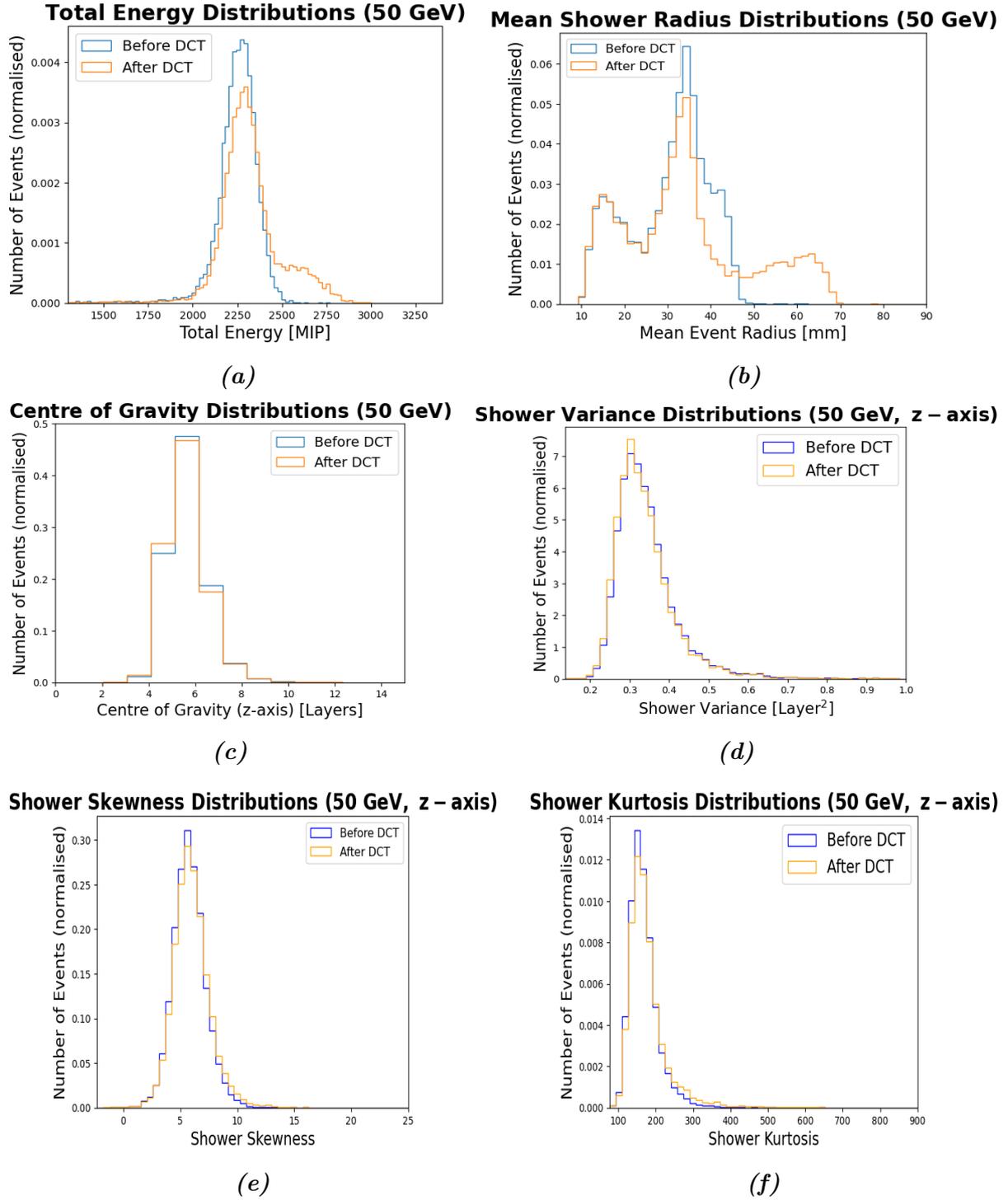
Figure 4.17.: Distributions of the total energy (a), the energy weighted mean shower radius (b), the CoG in z-direction (c), the variance (d), the skewness (e) and the kurtosis (f). The shower moments describe the shower in the z-direction. The original data is shown in blue and the data with generated noise (three Gaussian distributions) for nodes with ($x = 21 \wedge y$ odd) or (x odd $\wedge y = 21$) in orange.

(a)

(b)

(c)

(d)

(e)

(f)
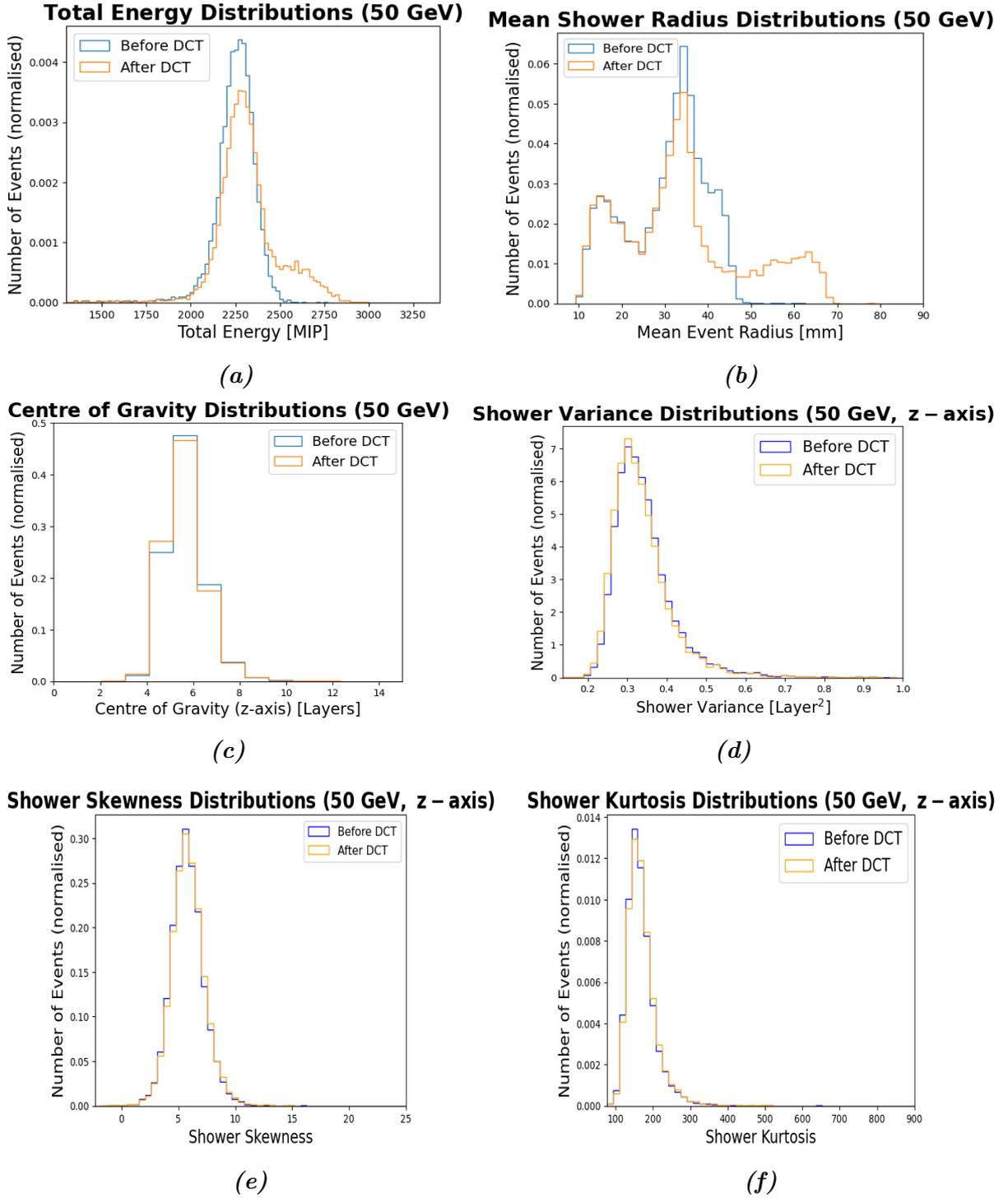
Figure 4.18.: Distributions of the total energy (a), the energy weighted mean shower radius (b), the CoG in z-direction (c), the variance (d), the skewness (e) and the kurtosis (f). The shower moments describe the shower in the z-direction. The original data is shown in blue and the data where all nodes with ($x = 21 \land y$ odd) or (x odd $\land y = 21$) were set to zero in orange.

49

4.7. Single event comparison

In order to analyse where the additional peaks in the total energy and the mean shower radius distributions originate from, and to investigate how the DCT changes a single event, a detailed comparison was made for a single event. For this, only one event was transformed and transformed back with the DCT and compared with its original. For the noise generation one Gaussian distribution was chosen which was also used in the random node generation with one Gaussian function from the previous section. Again, all nodes which fulfil the requirement ($x = 21 \wedge y$ odd) or (x odd $\wedge y = 21$) were generated randomly.

For this comparison, two kinds of plots were made. The first one shows the difference of the hit energies before and after the DCT as a two-dimensional histogram for each tile in the x- and y- direction. One plot has been made for each layer in the z-direction, where the colour bar range varies for each layer. This can be seen in Figures 4.19 and 4.20, where the hit energy differences for layers 6 and 7 are shown. Layer 6 and 7 are the most likely positions of the CoG in z-direction. The hit energy differs around the CoG in the transverse plane, which is in the centre of the plot, for layer 6. For layer 7 the biggest differences are located more in the upper and left parts of the CoG, whereas the differences all lie in the range of $-0.1\,\mathrm{MIP}$ and $0.3\,\mathrm{MIP}$.
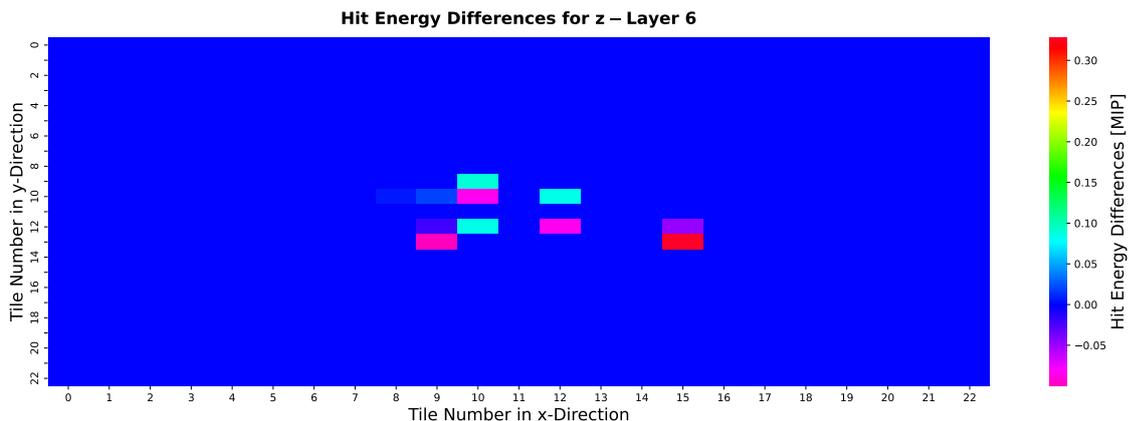
Figure 4.19.: Difference of the hit energies in layer 6 before and after the DCT, where all nodes that satisfy ($x = 21 \wedge y$ odd) or (x odd $\wedge y = 21$) were generated randomly.
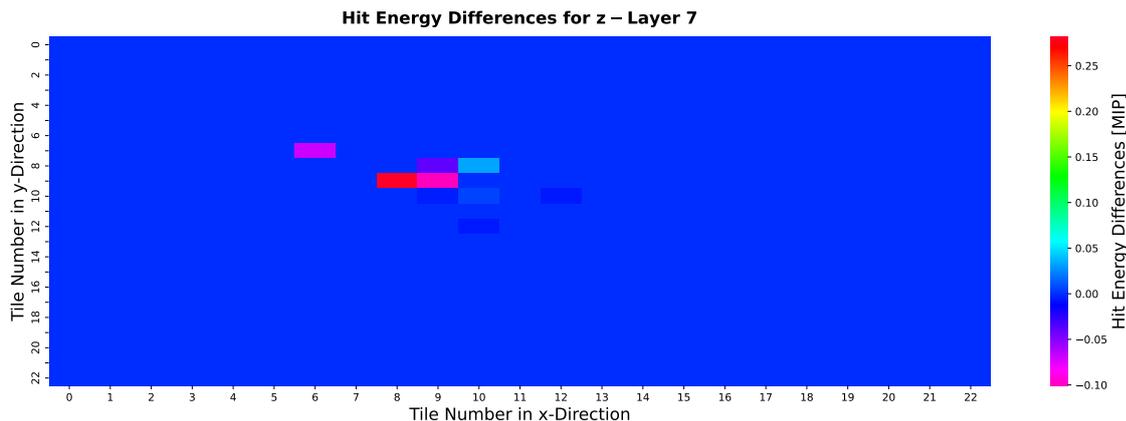
Figure 4.20.: Difference of the hit energies in layer 7 before and after the DCT, where all nodes that satisfy $(x = 21 \wedge y \text{ odd})$ or $(x \text{ odd} \wedge y = 21)$ were generated randomly.

The second kind is a two-dimensional histogram of the differences of the energy-weighted hit radii, which are the distances of the hits to the CoG multiplied by the corresponding hit energy. This plot is interesting, since the mean shower radius distributions in the previous Section showed an additional peak as well. Again, one plot represents all radii for all tiles in one layer, and the colour bar varies for each plot. In Figure 4.21 the hit radius differences for layer 6 and in Figure 4.22 for layer 7 are shown. The distributions of the radius differences look similar to the hit energy differences for both plots, whereas the differences lie within the range of $-15\,\mathrm{MIP \cdot mm}$ and $45\,\mathrm{MIP \cdot mm}$.
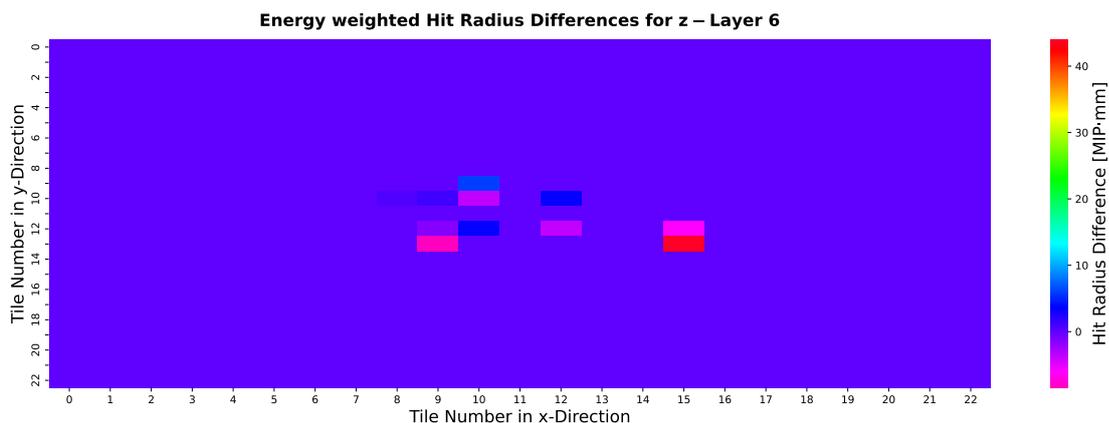
Figure 4.21.: Difference of the energy weighted hit radius in layer 6 before and after the DCT, where all nodes that satisfy $(x = 21 \wedge y \text{ odd})$ or $(x \text{ odd} \wedge y = 21)$ were generated randomly.
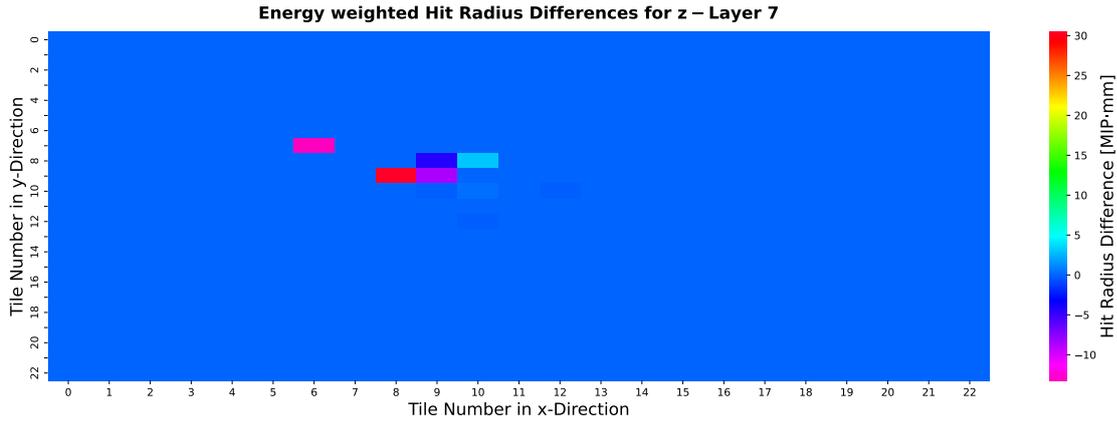
Figure 4.22.: Difference of the energy weighted hit radius in layer 7 before and after the DCT, where all nodes that satisfy ($x = 21 \wedge y$ odd) or (x odd $\wedge y = 21$) were generated randomly.

Since this comparison was only done for one event, the differences for the hit energy and the hit radius cannot be used to state a conclusion for electron showers in general. Moreover, since multithreading has been used for the computations of the transformations, it is currently not possible to say how much the shown event does actually contribute to the formation of the additional peaks. If it was possible to deduce which events in the original dataset contribute to the additional peak after back transformation, a one-event comparison for these specific events could be performed and the origin of the additional peaks analysed. However, in order to do that, it would be necessary to track the events during both transformations, which would require to rewrite programs such that they do not use multithreading, but transform events one by one. This would, however, greatly increase the computational time too. A further analysis of single events is thus outside the scope of this thesis.

5. Conclusion

Data reduction methods for a data-based fast particle shower simulation have been investigated in this thesis for electromagnetic showers. The investigation is based on a dataset recorded by the AHCAL group of the CALICE Collaboration with their highly granular technological detector prototype. For the data comparison the discrete cosine transformation has been utilised. The DCT is able to translate the shower shape into cosine nodes of various contributions. By considering highly frequent nodes with small values as noise and generating them randomly, following Gaussian distributions, the amount of data per event could be reduced.

First off, nodes have been generated randomly if their values lie within $(-10, 10)$ MIP by using two Gaussian distributions, one for nodes within $(-1, 1)$ MIP and the other for $(-10, -1] \cup [1, 10)$ MIP. Distributions of kinematic shower variables with generated noise featured good agreement with the original data. Differences only appeared as small fluctuations. By increasing the threshold to a larger value and introducing a third Gaussian distribution for the interval $(-20, -10] \cup [10, 20)$ MIP, the distributions with generated noise started to show signs of disagreement while mostly keeping their original shape. The next increase of the threshold to 30 MIP rendered the curves ultimately incompatible with the original distributions. Here again, three Gaussian distributions for random noise generation were used. Therefore, it can be concluded that the limit where less important nodes can be generated randomly lies between 20 MIP and 30 MIP.

The second method focused on generating only odd nodes randomly, which originated from the observation that these nodes tend to have smaller values than even nodes. The first attempt tried to generate all odd nodes randomly whose values lie within $(-30, 30)$ MIP, which was the threshold where the random generation stopped to match the data. However, this method did not improve the distributions any further. Therefore, the following idea was to generate all nodes that fulfil ($x = 21 \wedge y$ odd) or (x odd $\wedge y = 21$) since nodes with a higher node number tend to have smaller values. This was done with only one Gaussian distribution, which still resulted in slightly shifted curves and also altered

shapes of distributions of the kinematic variables compared to the original data. The most apparent differences were additional peaks in the distributions of the total energy and the energy-weighted mean shower radius. Each showed one additional peak on the right-hand side of the main distributions. In the next attempt, the same node configuration, but with three Gaussian distributions, was generated randomly. This time, the curves only differed in small fluctuations and matched the data well, except for the additional peaks in the total energy and mean radius distribution, which did not vanish. Lastly, all nodes, again with the same node configuration, were set to zero. The distributions of the kinematic variables showed slightly smaller fluctuations than the previous set of plots, but the problem of the additional peaks in the total energy and the mean shower radius still remained.

Furthermore, a single-event comparison of the data before and after applying the DCT was made. However, since multithreading has been used for both transformations, it is not possible to say to which part of the distributions the analysed event contributed. For an investigation of specific events and their contributions to the previously mentioned distributions, such as the additional peaks in the total energy and the mean shower radius, and how the DCT affects them, an implementation of the DCT without multithreading, but with individual event transformations is required.

To conclude, it is possible to use the DCT to reduce the hit energy data while obtaining good agreement with the original data, but only to a certain extent. All of this does, however, not compress the data set enough to achieve significant improvement in computing-time and computational resources of shower simulations. To achieve better agreement with the data, while generating more nodes randomly, a more detailed analysis of suitable Gaussian distributions or other possibly more fitting distributions could be done.

Calorimetry and thus particle shower simulations are, and will be, a crucial part of modern particle physics. Therefore, the search for possibilities to improve such simulations goes on, which will also contribute significantly to experiments at future particle accelerators such as the International Linear Collider, which will open many possibilities for the search for evidence for theories beyond the Standard Model.

A. Additional Plots

A.1. Centre-of-Gravity Distributions
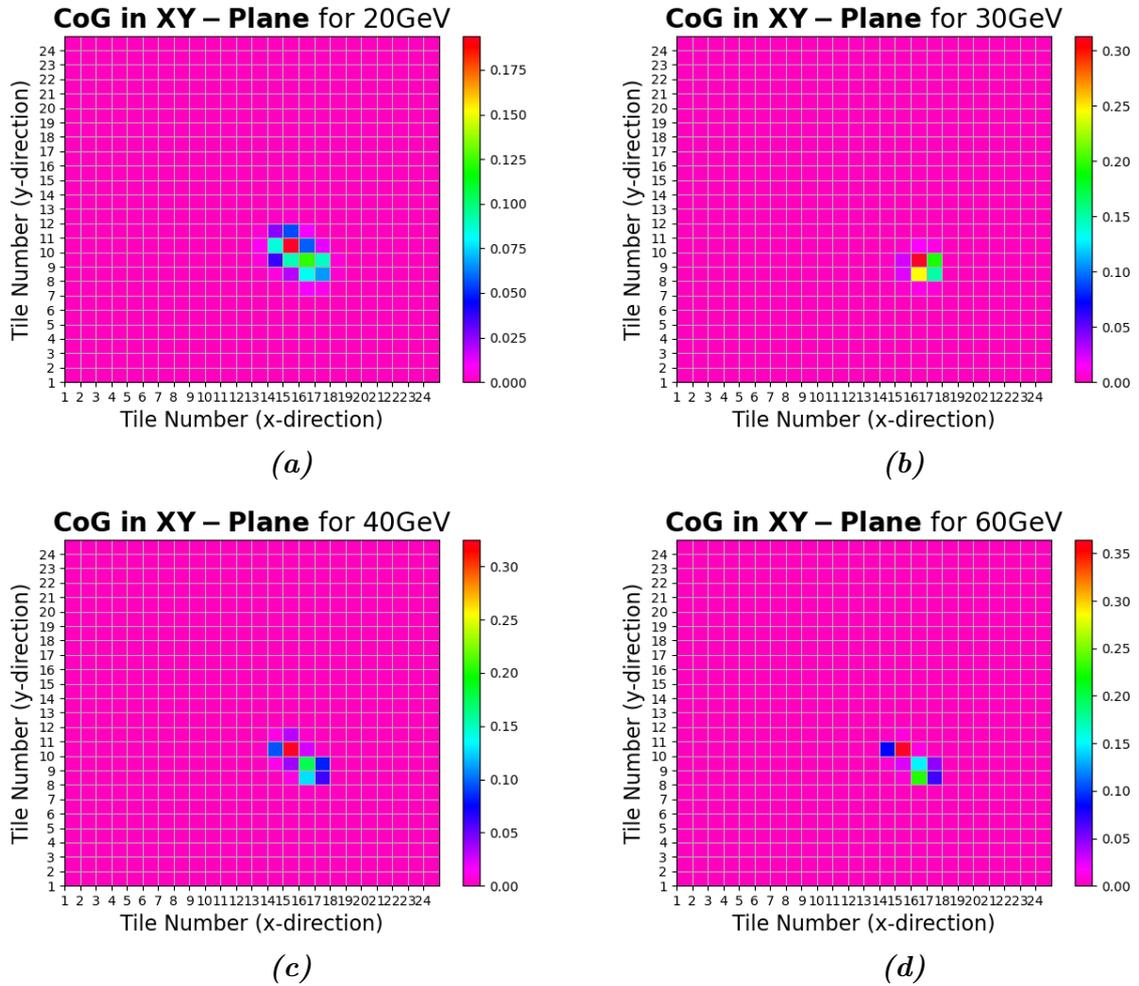
Figure A.1.: Centre of Gravity in the transverse plane for electromagnetic showers initiated by electrons with beam energies of 20 GeV (a), 30 GeV (b), 40 GeV (c) and 60 GeV (d).

A.2. Generating DCT Nodes Randomly and Kinematic Shower Variables

Figure A.2.: Distributions of the shower moments in x- (left) and in y-direction (right), for the variance (a), (b), the skewness (c), (d) and the kurtosis (e), (f). The original data is shown in blue and the data with generated noise within $(-10, 10)$ MIP in orange.

Figure A.3.: Distributions of the shower moments in x- (left) and the CoG in y-direction (right), for the variance (a), (b), the skewness (c), (d) and the kurtosis (e), (f). The original data is shown in blue and the data with generated noise within $(-20, 20)$ MIP in orange.

(a)

(b)

(c)

(d)
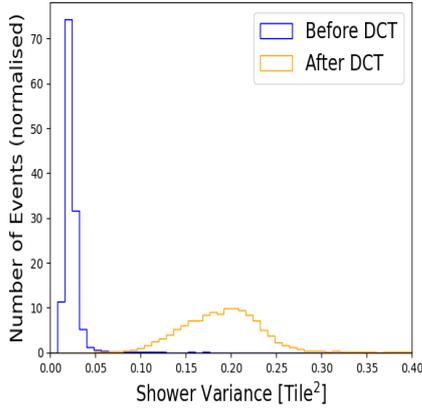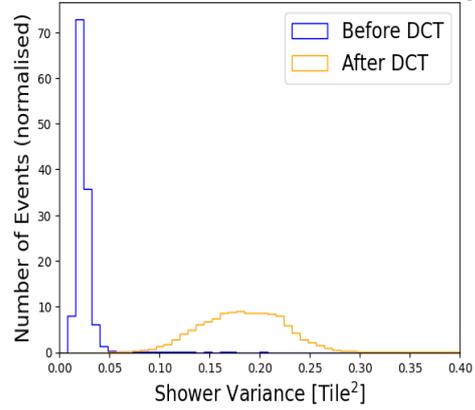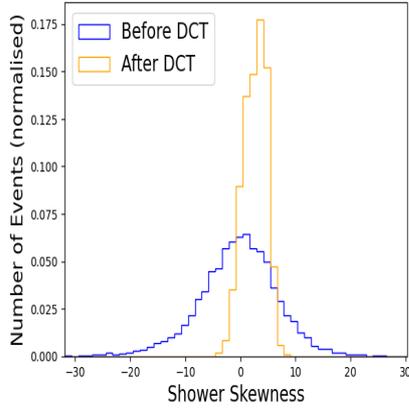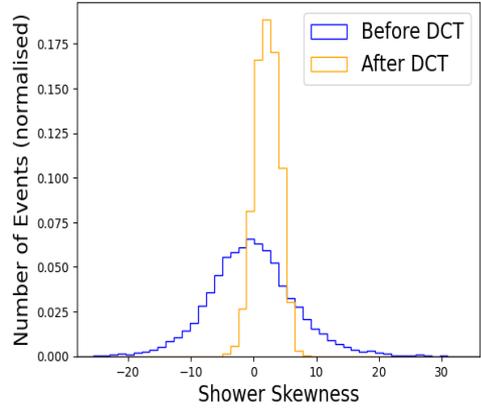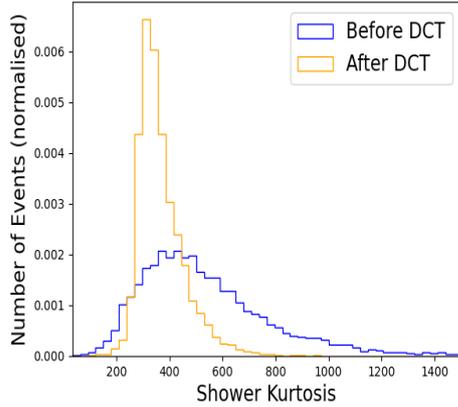
(e)
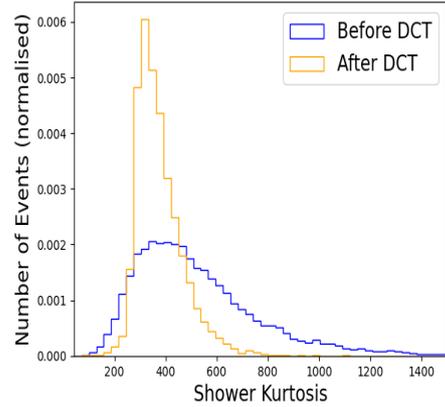
(f)

Figure A.4.: Distributions of the shower moments in x- (left) and in y-direction (right), for the variance (a), (b), the skewness (c), (d) and the kurtosis (e), (f). The original data is shown in blue and the data with generated noise within $(-30, 30)$ MIP in orange.

(a) (b)
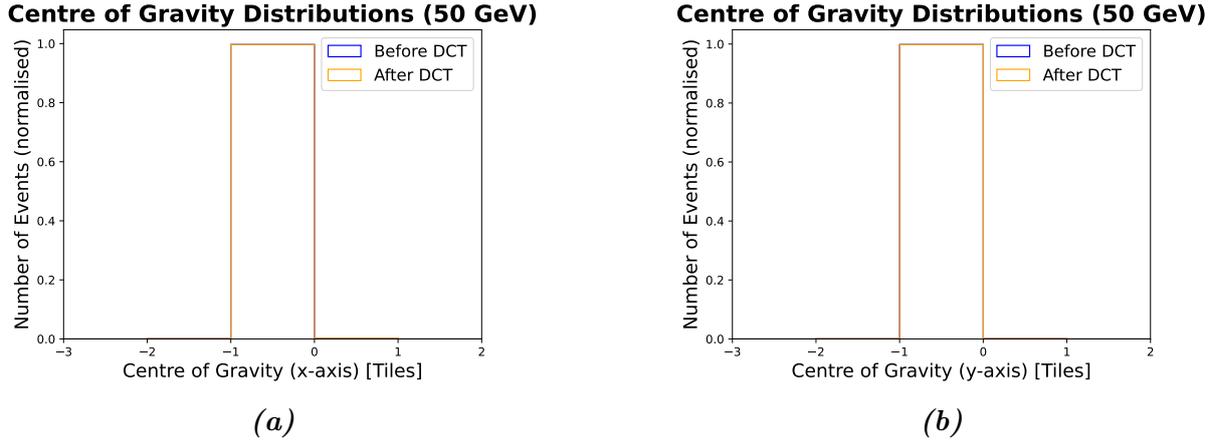
Figure A.5.: Distribution of the CoG in the transverse plane for electromagnetic showers initiated by electrons with a beam energy of 50 GeV for data (blue) and data with generated noise between $(-10, 10)$ MIP (orange).

(a) (b)
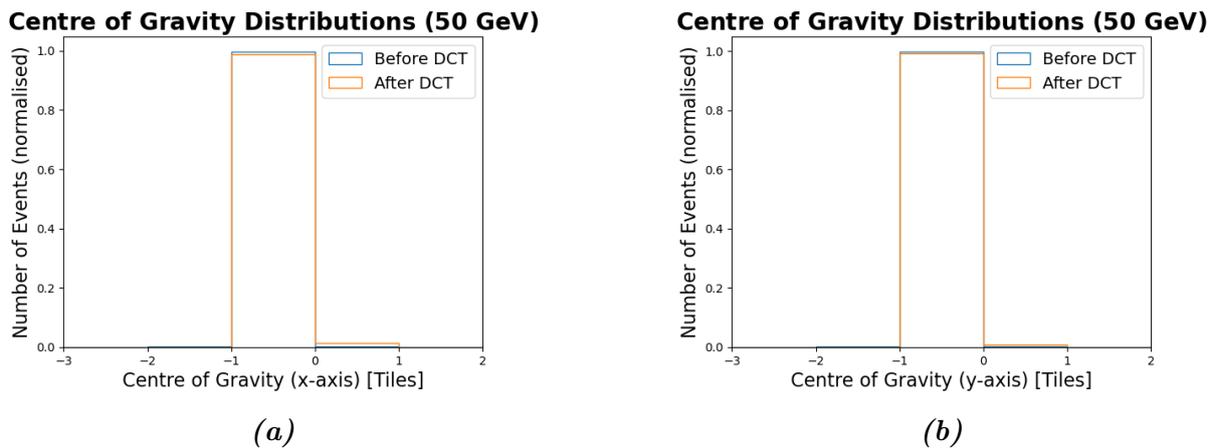
Figure A.6.: Distribution of the CoG in the transverse plane for electromagnetic showers initiated by electrons with a beam energy of 50 GeV for data (blue) and data with generated noise between $(-20, 20)$ MIP (orange).

Centre of Gravity Distributions (50 GeV)

(a)

Centre of Gravity Distributions (50 GeV)

(b)

Figure A.7.: Distribution of the CoG in the transverse plane for electromagnetic showers initiated by electrons with a beam energy of $50\,\text{GeV}$ for data (blue) and data with generated noise between $(-20, 20)\,\text{MIP}$ (orange).

A.3. Generating odd Nodes randomly
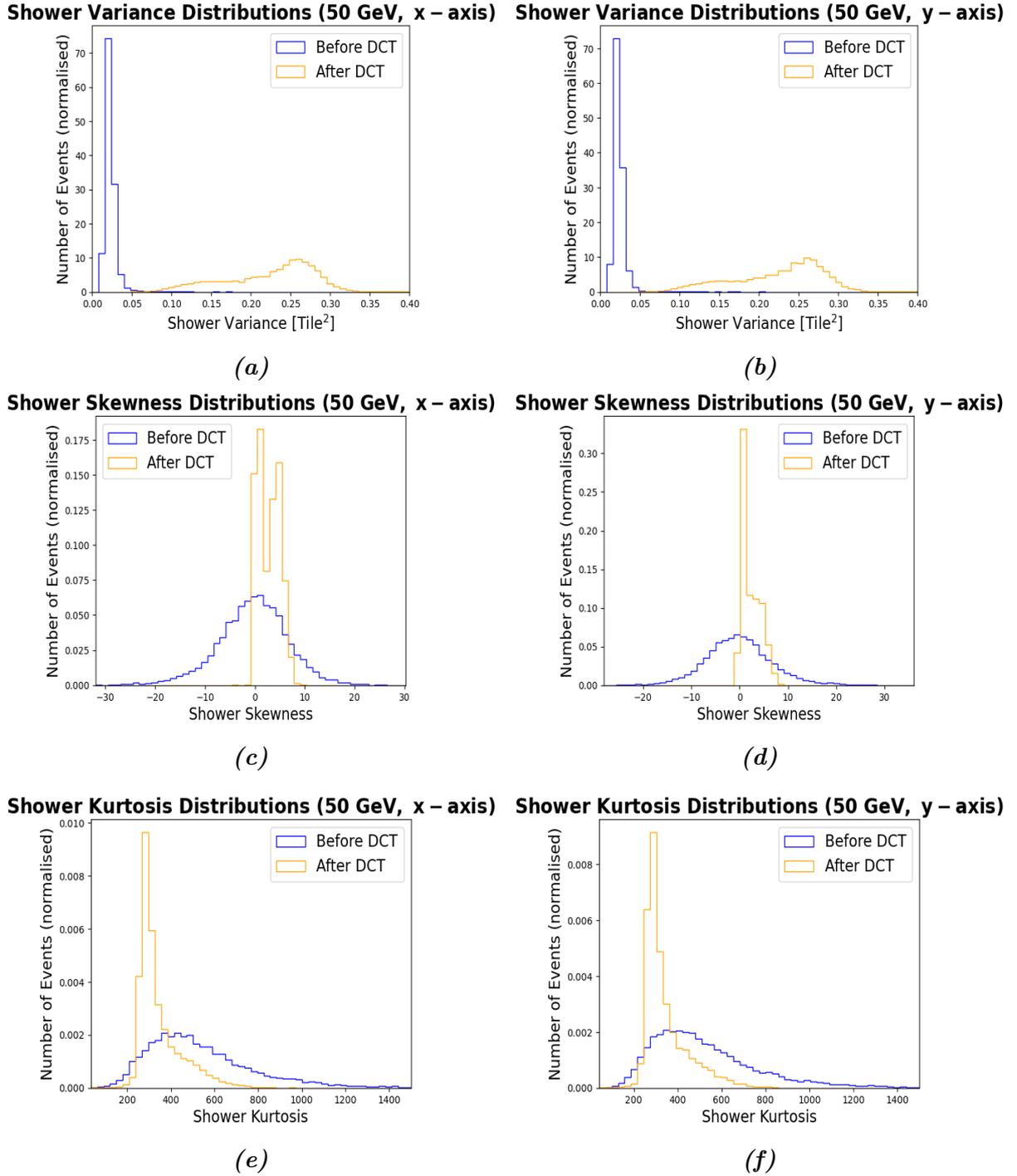
Figure A.8.: Distributions of the shower moments in x- (left) and in y-direction (right), for the variance (a), (b), the skewness (c), (d) and the kurtosis (e), (f). The original data is shown in blue and the data with generated noise for all odd nodes within $(-30, 30)$ MIP in orange.

Figure A.9.: Distributions of the shower moments in x- (left) and in y-direction (right), for the variance (a), (b), the skewness (c), (d) and the kurtosis (e), (f). The original data is shown in blue and the data with generated noise (one Gaussian distribution) for nodes with ($x = 21 \wedge y$ odd) or (x odd $\wedge y = 21$) in orange.

(a)

(b)

(c)

(d)

(e)

(f)
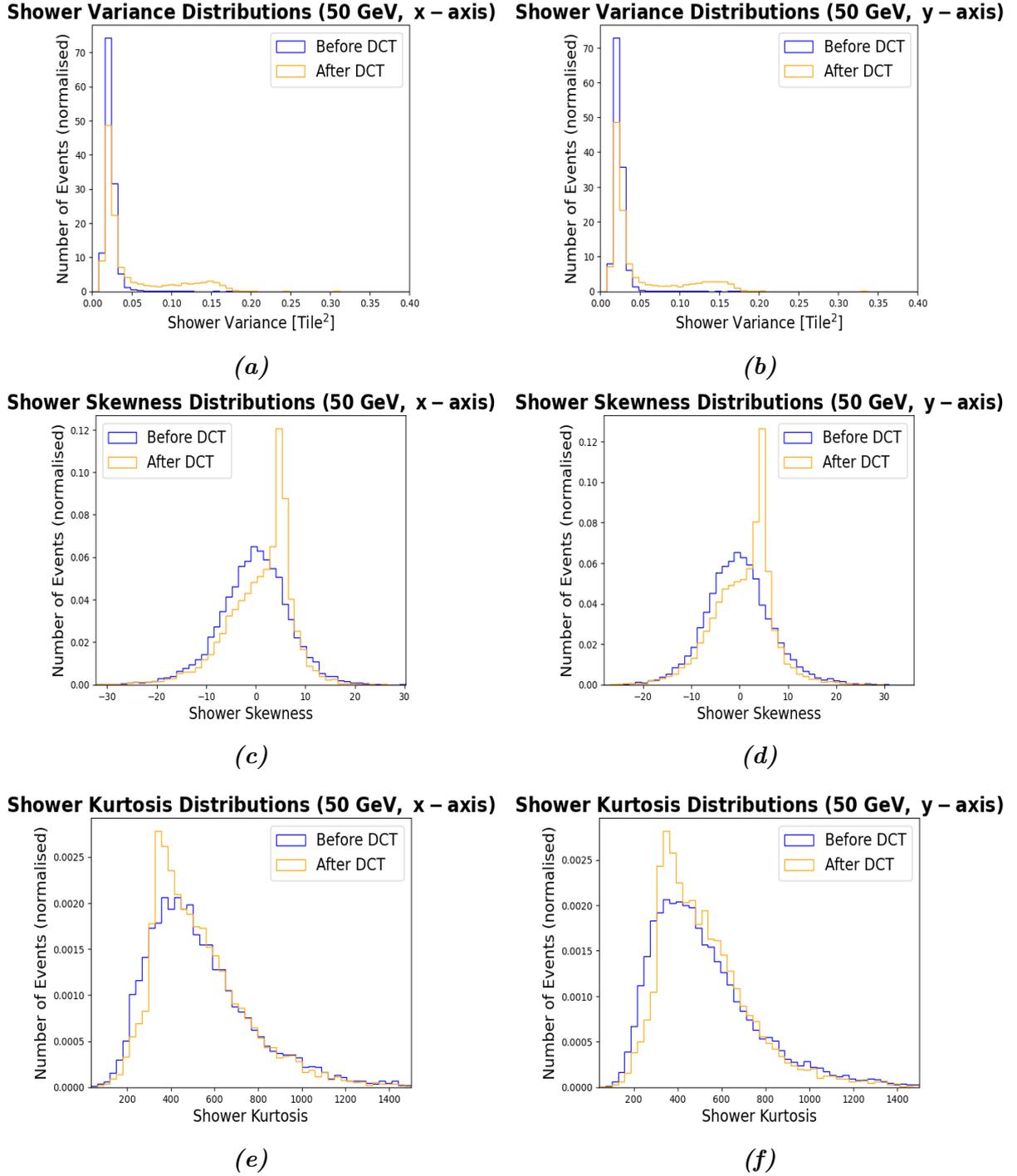
Figure A.10.: Distributions of the shower moments in x- (left) and in y-direction (right), for the variance (a), (b), the skewness (c), (d) and the kurtosis (e), (f). The original data is shown in blue and the data with generated noise (three Gaussian distributions) for nodes with ($x = 21 \land y$ odd) or (x odd $\land y = 21$) in orange.

A. Additional Plots

Shower Variance Distributions (50 GeV, x − axis) (a)

Shower Variance Distributions (50 GeV, y − axis) (b)

Shower Skewness Distributions (50 GeV, x − axis) (c)

Shower Skewness Distributions (50 GeV, y − axis) (d)

Shower Kurtosis Distributions (50 GeV, x − axis) (e)

Shower Kurtosis Distributions (50 GeV, y − axis) (f)
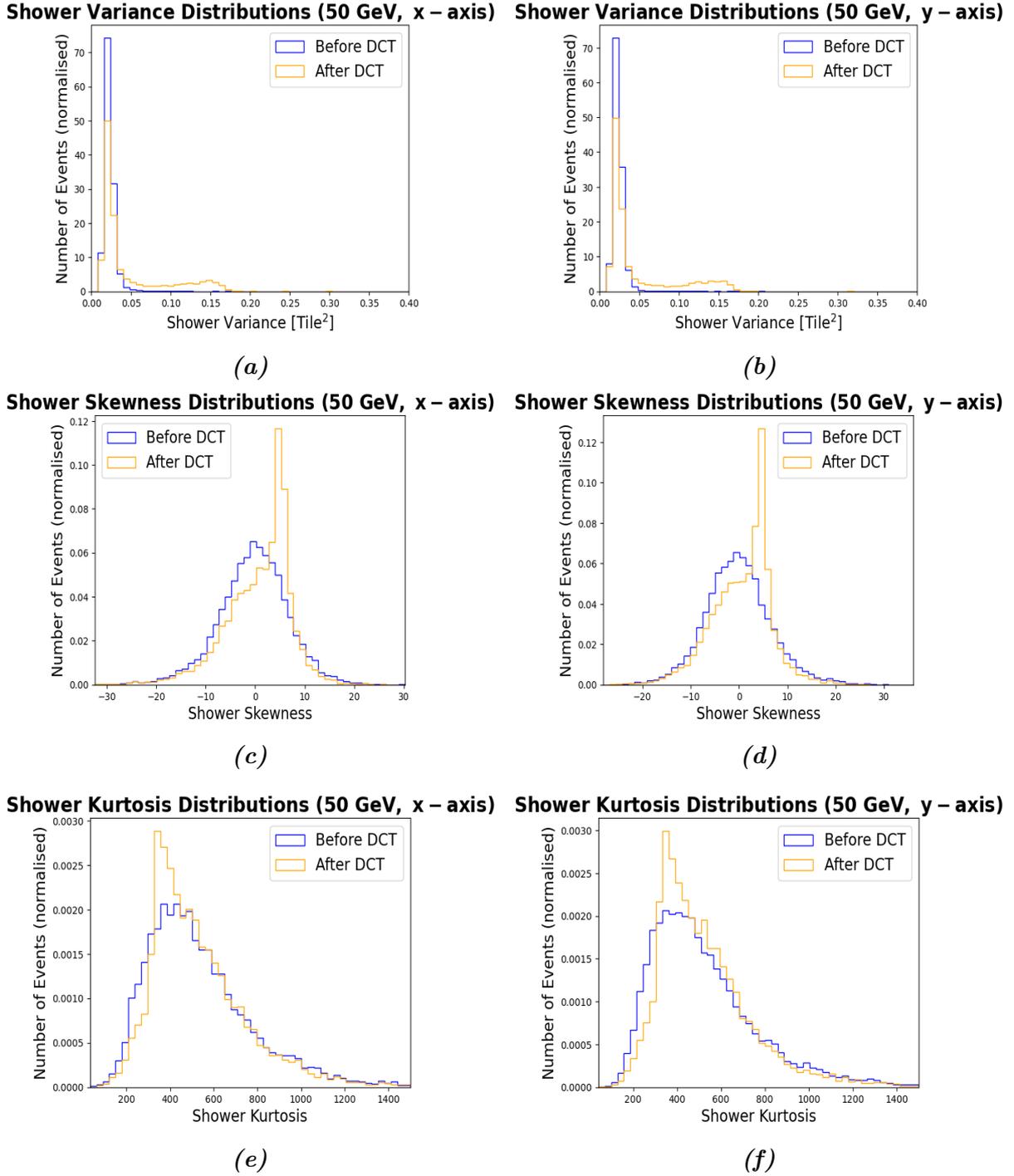
Figure A.11.: Distributions of the shower moments in x- (left) and in y-direction (right), for the variance (a), (b), the skewness (c), (d) and the kurtosis (e), (f). The original data is shown in blue and the data where all nodes with $(x = 21 \wedge y \text{ odd})$ or $(x \text{ odd} \wedge y = 21)$ were set to zero in orange

(a)   *(b)*
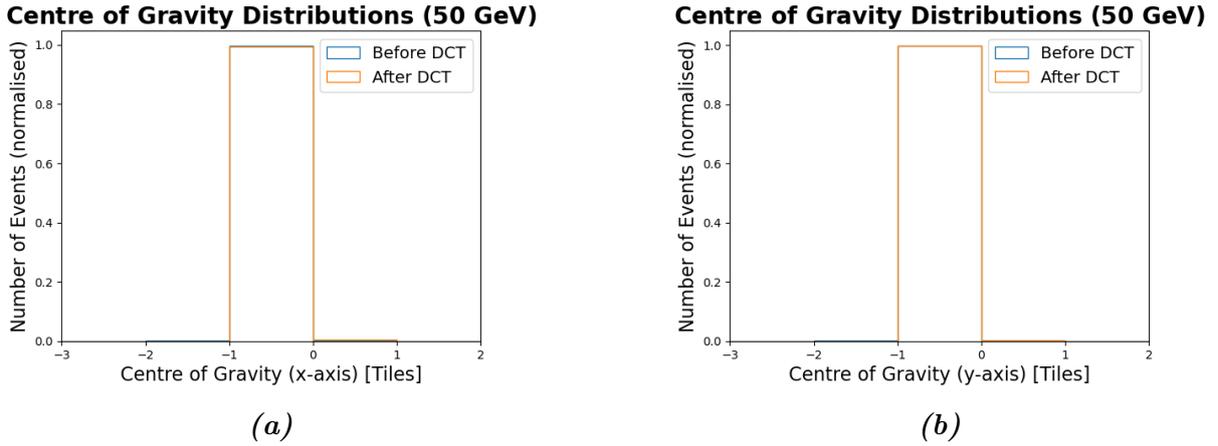
Figure A.12.: Distribution of the CoG in the transverse plane for electromagnetic show-
ers initiated by electrons with a beam energy of 50 GeV for data (blue)
and data with generated noise for all odd nodes between $(-30, 30)$ (or-
ange).

(a)   *(b)*

Figure A.13.: Distribution of the CoG in the transverse plane for electromagnetic show-
ers initiated by electrons with a beam energy of 50 GeV for data (blue)
and data with generated noise (one Gaussian distribution) for nodes with
$(x = 21 \wedge y$ odd) or $(x$ odd $\wedge y = 21$ (orange).

(a)

(b)

Figure A.14.: Distribution of the CoG in the transverse plane for electromagnetic showers initiated by electrons with a beam energy of 50 GeV for data (blue) and data with generated noise (three Gaussian distributions) for nodes with $(x = 21 \wedge y$ odd$)$ or $(x$ odd $\wedge y = 21)$ (orange).

(a)

(b)

Figure A.15.: Distribution of the CoG in the transverse plane for electromagnetic showers initiated by electrons with a beam energy of 50 GeV for data (blue) and data where all nodes with $(x = 21 \wedge y$ odd$)$ or $(x$ odd $\wedge y = 21)$ were set to zero (orange).

Bibliography

[1] S. L. Glashow. "Partial-symmetries of weak interactions". In: *Nucl. Phys.* 22.4 (1961), pp. 579–588.

[2] S. Weinberg. "A Model of Leptons". In: *Phys. Rev. Lett.* 19 (21 Nov. 1967), pp. 1264–1266.

[3] A. Salam. "Weak and Electromagnetic Interactions". In: *Conf. Proc. C* 680519 (1968), pp. 367–377.

[4] M. Gell-Mann. "A schematic model of baryons and mesons". In: *Phys. Lett.* 8.3 (1964), pp. 214–215.

[5] P. Higgs. "Broken symmetries, massless particles and gauge fields". In: *Phys. Lett.* 12.2 (1964), pp. 132–133.

[6] P. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Phys. Rev. Lett.* 13 (16 Oct. 1964), pp. 508–509.

[7] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". In: *Phys. Rev. Lett.* 13 (9 Aug. 1964), pp. 321–323.

[8] `https://home.cern/science/accelerators/large-hadron-collider`. [Retrieved: 12.08.2024].

[9] ATLAS Collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Physics Letters B* 716.1 (2012), pp. 1–29.

[10] CMS Collaboration. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Physics Letters B* 716.1 (2012), pp. 30–61.

[11] `https://twiki.cern.ch/twiki/bin/view/CALICE/CaliceOutline`. [Retrieved: 18.08.2024].

[12] The CALICE Collaboration. "Design, construction and commissioning of a technological prototype of a highly granular SiPM-on-tile scintillator-steel hadronic calorimeter". In: *JINST* 18.11 (Nov. 2023), P11018.

Bibliography

[13]   A. Laudrain and on behalf of the CALICE Collaboration. "The CALICE AHCAL: a highly granular SiPM-on-tile hadron calorimeter prototype". In: *J. Phys.: Conf. Ser.* 2374.1 (Nov. 2022).

[14]   H. Fritzsch, M. Gell-Mann, and H. Leutwyler. "Advantages of the color octet gluon picture". In: *Phys. Lett. B* 47.4 (1973), pp. 365–368.

[15]   `https://commons.wikimedia.org/w/index.php?curid=4286964`. [Retrieved: 18.08.2024].

[16]   K. G. Wilson. "Confinement of quarks". In: *Phys. Rev. D* 10 (8 Oct. 1974), pp. 2445–2459.

[17]   R. L. Workman et al. "Review of Particle Physics". In: *PTEP* 2022 (2022), p. 083C01.

[18]   M. Thomson. *Modern Particle Physics*. Cambridge University Press, 2013.

[19]   H. Kolanoski and N. Wermes. *Teilchendetektoren*. Springer Spektrum, 2016.

[20]   C. W. Fabjan. "Calorimetry in High-Energy Physics". In: *NATO Sci. Ser. B* 128 (1985), pp. 281–336.

[21]   E. Longo and I. Sestili. "Monte Carlo calculation of photon-initiated electromagnetic showers in lead glass". In: *Nuclear Instruments and Methods* 128.2 (1975), pp. 283–307.

[22]   O. L. Pinto. "Shower Shapes in a Highly Granular SiPM-on-Tile Analog Hadron Calorimeter". `https://ediss.sub.uni-hamburg.de/handle/ediss/9855`. PhD dissertation. (2022).

[23]   T. Gabriel et al. "Energy dependence of hadronic activity". In: *Nucl. Instrum. Methods Phys. Res. A* 338.2 (1994), pp. 336–347.

[24]   S. Lee. "On the limits of the hadronic energy resolution of calorimeters". In: *J. Phys. Conf. Ser.* 1162.1 (Jan. 2019), p. 012043.

[25]   G. Eigen et al. "Hadron shower decomposition in the highly granular CALICE analogue hadron calorimeter". In: *J. Instrum.* 11.06 (June 2016), P06013.

[26]   R. Bock, T. Hansl-Kozanecka, and T. Shah. "Parametrization of the longitudinal development of hadronic showers in sampling calorimeters". In: *Nucl. Instrum. Methods Phys. Res.* 186.3 (1981), pp. 533–539.

[27]   B. Andrieu et al. "The H 1 liquid argon calorimeter system". In: *Nucl. Instrum. Methods Phys. Res. A* 336.3 (1993), pp. 460–498. ISSN: 0168-9002.

[28]   The CMS Collaboration. "The CMS experiment at the CERN LHC". In: *JINST* 3.08 (Aug. 2008), S08004.

[29] R. Wigmans. "On the energy resolution of uranium and other hadron calorimeters". In: *Nucl. Instrum. Methods Phys. Res. A* 259.3 (1987), pp. 389–429.

[30] C. Fabjan et al. "Iron liquid-argon and uranium liquid-argon calorimeters for hadron energy measurement". In: *Nuclear Instruments and Methods* 141.1 (1977), pp. 61–80.

[31] The CALICE Collaboration. "Electromagnetic response of a highly granular hadronic calorimeter". In: *JINST* 6.04 (Apr. 2011).

[32] The CALICE Collaboration. "Hadronic energy resolution of a highly granular scintillator-steel hadron calorimeter using software compensation techniques". In: *JINST* 7.09 (Sept. 2012).

[33] `https://linearcollider.org/`. [Retrieved: 28.07.2024].

[34] P. Eckert et al. "Characterisation studies of silicon photomultipliers". In: *Nucl. Instrum. Methods Phys. Res. A* 620.2 (2010), pp. 217–226.

[35] B. Wang and J. Mu. "High-speed Si-Ge avalanche photodiodes". In: *PhotoniX* 3 (Mar. 2022), p. 8.

[36] N. Anfimov et al. "Study of silicon photomultiplier performance at different temperatures". In: *Nucl. Instrum. Methods Phys. Res. A* 997 (2021), p. 165162.

[37] D. Coomans et al. "Potential methods in pattern recognition: Part 1. Classification aspects of the supervised method ALLOC". In: *Anal. Chim. Acta* 133.3 (1981), pp. 215–224.

[38] A. Wilhahn. "Investigation of Fast Hadron Shower Simulation Methods with the CALICE AHCAL Prototype". Reference number: II.Physik-UniGö-MSc-2022/03. Master's thesis. (2022).

[39] `https://commons.wikimedia.org/w/index.php?curid=10414002`. [Retrieved: 29.07.2024].

[40] N. Ahmed, T. Natarajan, and K. Rao. "Discrete Cosine Transform". In: *IEEE Trans. Comput.* C-23.1 (1974), pp. 90–93.

Acknowledgements

The work on this bachelor's thesis in the II. Physikalisches Institut was a great experience, which enabled me to get an insight in the work and research of particle physicists. As it would not have been possible to write this bachelor's thesis without the support of many amazing people, I would like to thank them:

First of all, I would like to thank Prof. Dr. Stan Lai, who is very dedicated to his Bachelor students, for giving me the opportunity to write my Bachelor's thesis in his research group. Especially the weekly research chats helped a lot with the progress of this thesis. Additionally, I would like to thank PD Dr. Ralf Bernhard for being my second referee.

A big thank you goes to André Wilhahn for the support through the whole thesis and for answering all my questions, even the stupid ones.

Also thanks to Noah for proofreading my thesis.

Last but not least, I would like to thank my parents, who always supported me.

Erklärung nach §13(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 nichtbestandenen Prüfung an dieser oder einer anderen Hochschule eingereicht wurde.

Göttingen, den 25. Oktober 2024

(Victoria Swoboda)