

# **Estimating Mass Composition with AugerPrime**

**Bestimmung der Massenzusammensetzung mit AugerPrime**

Master's Thesis of

Tobias Schulz

at the KIT Department of Physics  
Institute for Astroparticle Physics (IAP)

Reviewer: Prof. Dr. Ralph Engel  
Second reviewer: Prof. Dr. Guido Drexlin  
Advisor: Dr. David Schmidt  
Second advisor: Dr. Markus Roth

13. Juli 2020 – 12. Juli 2021

Karlsruher Institut für Technologie  
Fakultät für Physik  
76128 Karlsruhe

## **Declaration of independent writing**

I declare that I have developed and written the enclosed thesis completely on my own, and have not used sources or other means without properly referencing them in the text. I also declare that the rules of good scientific practice of the Karlsruhe Institute of Technology (KIT) have been followed.

**Karlsruhe, the 08. July 2021**

---

Tobias Schulz

---

Prof. Dr. Ralph Engel



# Abstract

Cosmic rays are still - almost one hundred years after their discovery - actively studied. Although the knowledge about cosmic rays physics has improved over this time, a lot of open questions are still left unanswered. Over the last decades, new technological advancements made it possible to study particle physics at high energies. However, the energies of the cosmic particles measured on earth exceed the limits of human-made accelerators by several orders of magnitude. A lot of effort has been put into developing models to explain, what objects and mechanisms can accelerate cosmic rays to such ultra high energies. Understanding the energy spectrum, its distinct features and the mass composition of cosmic rays, is of key importance to verify, discard or improve these theories and identify the sources of cosmic rays.

The number of cosmic rays at the highest energies is very low. Therefore, large, ground-based detector setups are needed to measure as many of these events as possible. The largest experiment for the measurement of these ultra-high-energy cosmic rays is the Pierre Auger Observatory, located in Argentina. About 1600 ground-based Water Cherenkov Detectors, covering an area of more than 3000 km<sup>2</sup> and 27 Fluorescence Detectors at four sites, overlooking the array from the edges, are measuring in a complementary setup the development of air showers in the atmosphere and their lateral distribution on the ground. The measurement of the longitudinal development of air showers by the fluorescence telescopes allows for estimations of the mass composition of the primary cosmic rays. However, the duty cycle is limited to 15% and the statistics of the highest energy events are limited. The Surface Detector has an operating uptime of 100% and can therefore record bigger amounts of events. With the current upgrade, called AugerPrime, an additional Scintillator Surface Detector is deployed on top of the existing ground detectors, providing additional information regarding the air shower composition.

The work presented in this thesis covers the improvement of the current method of reconstructing the lateral distributions of air showers measured with the Scintillator Surface Detector. The shower geometry reconstructed from measurements of the Water Cherenkov Detectors is commonly used in subsequent steps of event reconstruction and subordinate detectors. In this thesis, a method of propagating uncertainties of the shower core position into the measured signal is developed. Certain features of the individual detector responses to the muonic component of air showers are analyzed and discussed. A Principal Component Analysis, based on previous studies, is employed and modified to estimate the mass composition by using the combined measurements of both surface detector types. As a last step, these tools are used to give a first estimation of the mass composition of cosmic rays, using almost two years of data, measured with the preproduction array of the Pierre Auger Observatory.



# Zusammenfassung

Die Entdeckung der kosmischen Strahlung liegt nun bald 100 Jahre zurück und obwohl sich seitdem das Wissen über die kosmische Strahlung vergrößert hat, bleiben weiterhin viele Fragen unbeantwortet. Trotz des technologischen Fortschritts in den letzten Jahrzehnten, welcher die Untersuchung der Teilchenphysik bei hohen Energien ermöglicht hat, erreichen die von Menschen gebauten Beschleuniger jedoch nur einen Bruchteil der Energie der kosmischen Strahlung. Um diese extremen Energien erreichen zu können, wurden bereits verschiedene theoretische Modelle entwickelt, welche wesentlich auf dem Energiespektrum mit seinen Eigenschaften und der Massenzusammensetzung der kosmischen Strahlung aufbauen.

Die Häufigkeit der kosmischen Strahlung bei höchsten Energien ist sehr gering und große, bodengebundene Experimente werden benötigt, um möglichst viele Ereignisse zu messen. Das größte dieser Experimente ist das Pierre Auger Observatorium in Argentinien. Es besteht aus über 1600 Wasser-Cherenkov-Detektoren auf einer Fläche von mehr als 3000 km<sup>2</sup> und 27 Fluoreszenz-Teleskopen, die an vier Standorten das Feld des Oberflächendetektors überblicken. Die Messung der longitudinalen Entwicklung der Luftschaer durch die Fluoreszenz-Teleskope ermöglicht eine Abschätzung der Massenzusammensetzung der kosmischen Strahlung. Die Fluoreszenz-Teleskope messen jedoch nur 15% der Zeit, wodurch die Anzahl der detektierten Ereignisse beschränkt ist. Im Vergleich dazu misst der Oberflächendetektor dauerhaft und detektiert damit eine größere Anzahl an Ereignissen. Im Zuge der Erweiterung AugerPrime wird ein zusätzlicher Plastik-Szintillationsdetektor über den existierenden Wasser-Cherenkov-Detektoren installiert. Der zusätzliche Detektor liefert Daten für die Schauer-Zusammensetzung.

Diese Arbeit befasst sich unter anderem mit der Verbesserung der derzeitigen Rekonstruktion der Lateralverteilungen von Luftschauern, welche mit den Plastik-Szintillatoren gemessen wurden. Die Schauer-Geometrie, durch Messungen mit Wasser-Cherenkov-Detektoren rekonstruiert, wird in den nachfolgenden Schritten der Ereignisrekonstruktion von Messungen mit untergeordneten Detektoren verwendet. Im Zuge dieser Arbeit wird eine Methode vorgestellt, die Unsicherheiten der Position des Schauerzentrums auf die gemessenen Signale zu übertragen. Desweiteren soll das Ansprechen der verschiedenen Detektoren auf die myonische Luftschaerkomponente untersucht werden. Aufbauend auf vorausgehenden Arbeiten, wird eine Hauptkomponentenanalyse durchgeführt, um mittels der kombinierten Messungen von beiden Oberflächendetektoren eine Formel für die Massenzusammensetzung aufzustellen. Im letzten Teil der Arbeit werden diese Ergebnisse genutzt, um von den gesammelten Daten aus zwei Jahren Messung mit dem Preproduction Array eine erste Abschätzung der Massenzusammensetzung der kosmischen Strahlung zu geben.



# Contents

<b>Abstract</b>	<b>i</b>
<b>Zusammenfassung</b>	<b>iii</b>
<b>1. Introduction</b>	<b>1</b>
<b>2. Cosmic Rays and their Energy Spectrum</b>	<b>3</b>
2.1. Extensive Air Showers . . . . .	3
2.2. The Flux of Cosmic Rays . . . . .	4
2.3. Mass Composition with Measurements of the Shower Depth . . . . .	7
<b>3. The Pierre Auger Observatory and its Upgrade</b>	<b>9</b>
3.1. Water Cherenkov Detectors . . . . .	9
3.2. Fluorescence Detector . . . . .	10
3.3. Current Results and Open Questions . . . . .	11
3.4. The Upgrade: AugerPrime . . . . .	12
<b>4. Surface Detector Reconstruction</b>	<b>17</b>
4.1. Reconstructing Events with the WCD . . . . .	17
4.2. Reconstructing Events with the SSD . . . . .	24
4.3. Propagating Core Uncertainties in Subordinate Detector Reconstructions	25
4.3.1. Shower Size Resolution of the SSD Reconstruction . . . . .	25
4.3.2. Propagating the Core Uncertainty to the SSD Reconstruction . . . . .	27
4.3.3. Comparing Current and Propagated Uncertainty Model . . . . .	30
4.4. Quality Cuts . . . . .	35
4.4.1. Large Single Station Signals . . . . .	35
4.4.2. Zero Signal Stations . . . . .	38
<b>5. Detector Responses to the Muonic Shower Component</b>	<b>41</b>
5.1. Ideal Signal Model . . . . .	41
5.1.1. Projected Detector Area . . . . .	42
5.1.2. Average Tracklength . . . . .	43
5.2. Idealized Detector Response with a Constant Injection Disk . . . . .	44
5.3. Angular Distributions of Shower Particles . . . . .	45
5.4. Muon Energy Spectrum . . . . .	46
5.5. A Brief Look at the Electromagnetic Component . . . . .	48

<b>6. Principal Component Analysis</b>	<b>59</b>
6.1. Formulation . . . . .	59
6.1.1. Attenuation Correction of the Shower Size . . . . .	59
6.1.2. Basic Formulae . . . . .	60
6.2. Input Parameters and Training . . . . .	62
6.3. Evaluating Mass Estimation . . . . .	64
6.4. Energy Estimation . . . . .	67
<b>7. A First Look at Mass Composition</b>	<b>71</b>
7.1. Event Selection . . . . .	71
7.2. Current limitations . . . . .	72
7.3. Comparison with current SD Energy Reconstruction . . . . .	75
7.4. First Estimation of the Logarithmic Mass . . . . .	77
7.5. Possible Further Improvements . . . . .	79
<b>8. Conclusion</b>	<b>83</b>
<b>A. Appendix</b>	<b>85</b>
A.1. Primary and Model Comparison for Propagated Core Uncertainties . . . . .	85
A.2. Detector Signals from the Electromagnetic Component . . . . .	89
A.3. Extension of the PCA for the Energy Estimation . . . . .	91
A.4. Histograms of the Mass Estimation . . . . .	92
A.5. Mass Estimation for Different Hadronic Interaction Models . . . . .	93
<b>Bibliography</b>	<b>99</b>

# 1. Introduction

At the beginning of the 20th century, newly discovered physical phenomena were of rising interest for scientists. One of these discoveries was the radioactivity, examined and described first by Becquerel. The phenomenon of charged objects losing their charge, while being exposed to air, was investigated by different scientists.

In 1912, Victor Hess, inspired by the results of Domenico Pacini, who showed that the radiation strength decreases under water, proofed that the radiation increases with height [1]. He came to the assumption, that particles from extraterrestrial sources enter our atmosphere, ionizing the molecules and received the Nobel Price for his discovery of cosmic rays in 1936.

Many new experiments and research, regarding properties and origin of these cosmic rays, were conducted. Pierre Auger was the first to measure lateral distributions of extensive air showers, that evolve due to interactions of the primary cosmic ray particle with the atmosphere, causing an avalanche of further particle interactions [2]. The measurement of air showers made it possible to discover cosmic rays with energies up to  $10^{20}$  eV, which surpass the maximum energy of current man-made particle accelerators by seven magnitudes.

Until now there are still many unanswered questions regarding ultra high energy cosmic rays. It remains unclear, which possible acceleration mechanisms can bring them to such high energies.

Understanding and measuring the cosmic ray spectrum and its features is therefore of great importance. In 1966 Greisen, as well as Zatsepin and Kuzmin predicted a significant energy loss at energies above  $\approx 5 \times 10^{19}$  eV for proton primary cosmic rays, called the *GZK cut-off* [3, 4]. The GZK cut-off postulates, that these high energy protons interact with the photons of the cosmic microwave background, and lose energy in the process. At the highest energies, a suppression of the flux of cosmic rays could be proven by the Pierre Auger Observatory [5]. It is still unclear though, whether the suppression is due to the GZK cut-off or other phenomena.

In an upgrade of the Pierre Auger Observatory, new Scintillator Surface Detectors are installed on top of the existing Water Cherenkov Detectors. Different shower properties can be studied with both detectors being more or less sensitive on different components of a shower.

The main goal of this thesis, is to give a first look at the mass composition, reconstructed by only using the surface detector of the Pierre Auger Observatory. First, ultra high energy cosmic rays and air showers are introduced and explained in Chapter 2. In Chapter 3 a brief

overview of the Pierre Auger Observatory is given, as well as a description of the various detectors and the currently deployed upgrade AugerPrime. The beginning of Chapter 4 features a description of the current reconstruction of air shower events with the Offline framework used within the Auger Collaboration. As part of the work of this thesis, the reconstruction with the Scintillator Surface Detector is updated in the second part of the Chapter. This is done in order to improve the datasets, that are later used for reconstructing the mass composition. In Chapter 5 an analysis on the different detector responses to the air shower components is given, to gain more insight in the behaviour of the detector signals. A principal component analysis is trained and performed in Chapter 6 to receive equations for energy and mass estimation of the primary cosmic rays. These equations are then used in Chapter 7, to get a first look at the mass composition, reconstructed purely with surface detector measurements and an outlook for future improvements of the reconstruction of mass composition with the surface detector is given. The last Chapter features a brief summary of the thesis results.

## 2. Cosmic Rays and their Energy Spectrum

*Cosmic rays* (CR) are relativistic, charged particles that have an extraterrestrial origin and arrive at the Earth with a rate of 1000 particles per square meter per second. They can reach energies up to  $10^{20}$  eV, thus being the most energetic particles known so far. Cosmic rays above energies of  $10^{18}$  eV are also referred to as *ultra high energy cosmic rays* (UHECR). While various sources for cosmic rays are known, it remains unclear, what mechanisms can create those UHECRs. First, a brief introduction to extensive air showers and their components, based on the Heitler Model, is given. In the subsequent Section, the energy spectrum of cosmic rays with some of its features is explained. The last Section features a brief summary of estimating the mass composition of cosmic rays with the shower depth.

### 2.1. Extensive Air Showers

When entering Earth's atmosphere, the primary CRs interact with the nuclei in the air, starting an avalanche of secondary particles that also lead to subsequent interactions. An *extensive air shower* (EAS) is forming, that can be divided in three components (see Fig. 2.1), as described by the shower model by Heitler and Matthews [6, 7], which are briefly discussed below.

The core of the shower consists of the **hadronic component**. Pions, kaons, as well as neutrons and a few heavier nuclei are created in the early stages of the shower development. These particles amount to less than 1% of the hadronic component of the shower and rarely reach the ground due to decay processes. The decay of the neutral pions leads to electromagnetic sub showers

$$\pi^0 \rightarrow \gamma + \gamma. \quad (2.1)$$

The **electromagnetic component**, contributes to the majority of the secondary particles of an EAS with approximately 98%. It consists of electrons, positrons and photons that are self inducing further pair-production and bremsstrahlung processes. With each of these processes two new particles are created, leading to a total of  $N = 2^n$  particles after  $n$  steps. As soon as the individual energy of the particles drops below the critical energy of  $E_c = 85$  MeV no further particles are created and the shower starts to cease. The maximum number of particles is therefore directly proportional to the energy  $E_0$  of the primary particle  $N_{\max} \propto E_0$ .

The third component of the EAS is the **muonic component**, which includes muons, anti muons and neutrinos. They are created through the decay of charged pions and kaons and it accounts to less than 2% of the air shower. The majority of these particles reach the ground without further interaction.

The detection of these air showers can be achieved by either directly measuring energy

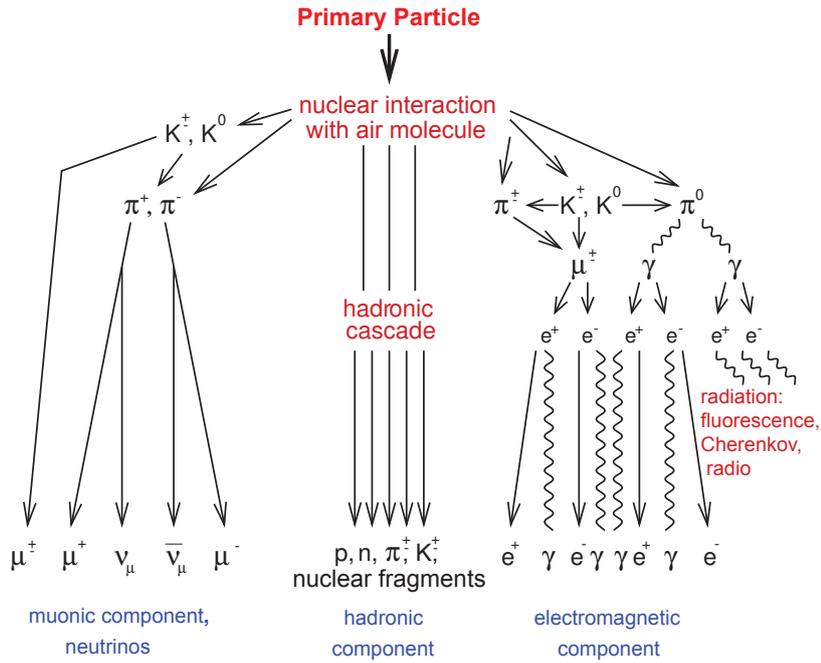


Figure 2.1.: Depiction of EAS

deposits of particles on the ground with detectors or through the emission of fluorescence light of excited nitrogen particles due to the shower passing through the atmosphere.

## 2.2. The Flux of Cosmic Rays

The flux, also referred to as energy spectrum, of the cosmic rays covers a wide range of energies from  $10^{10}$  eV up to  $10^{20}$  eV, also shown in Fig. 2.2. With increasing energy of the CRs, the flux is decreasing and at the highest energies, only one particle (or less) per square kilometer per year can be measured. In the CR spectrum various features are visible that can be related to the nature and distribution of their sources. The differential flux can be described by a decreasing broken power law with different values of the spectral index  $\gamma$  to account for the changing steepness

$$\frac{dN^4}{dE dA dt d\Omega} \propto E^{-\gamma}. \quad (2.2)$$

One of the three most evident features of the spectrum is the steepening of the flux at energies of  $3 \times 10^{15}$  eV, the so called *knee*. At that point  $\gamma$  changes from  $\approx 2.7$  to  $\approx 3.1$ . This mostly gets attributed to galactic *supernova remnants* (SNR) reaching the maximum of their acceleration energy for light nuclei as protons. The *Karlsruhe Shower Core and Array Detector* experiment KASCADE could confirm this by observing a change from heavy to light elements in that region as well [9]. A fainter, *second knee* is visible for heavy particles, such as iron nuclei, which could be accounted to the maximum acceleration energy of SNRs [10].

The next distinct feature of the spectrum is the *ankle* at an energy of around  $5 \times 10^{18}$  eV,

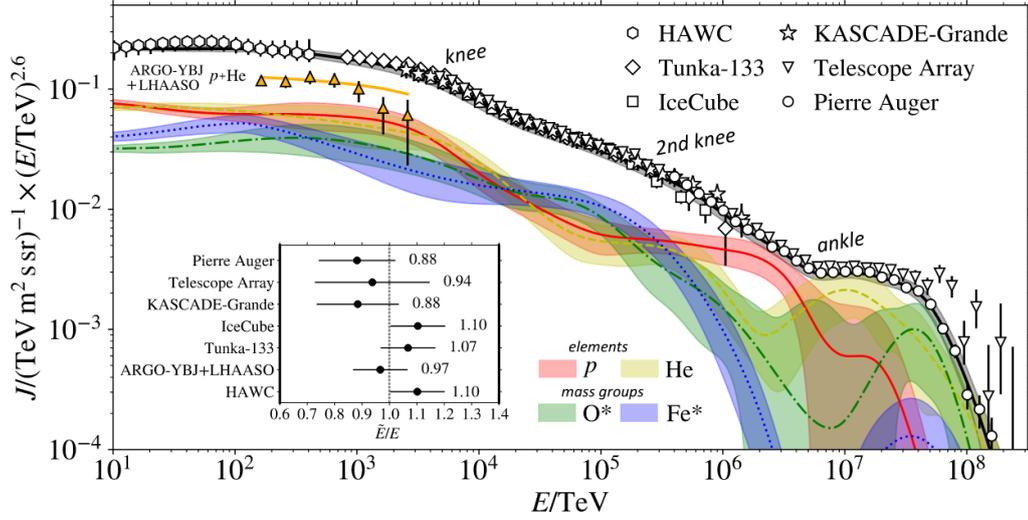


Figure 2.2.: Energy spectrum of cosmic rays. The knee is visible at an energy of approximately  $3 \times 10^{15}$  eV. At  $5 \times 10^{18}$  eV, the flux is hardening and at around  $5 \times 10^{19}$  eV the GZK cut-off becomes visible. Figure taken from [8].

where the flux is hardening. One prominent explanation is the change from galactic to extragalactic accelerators and a pile-up effect due to electron-positron pair production by the interaction of protons with photons of the *cosmic microwave background* (CMB) [11]. An absence of any directional origin in the galactic plane of the CRs at this energy also supports the change to extragalactic sources.

The third prominent feature of the spectrum is the GZK cut-off. At energies above  $5 \times 10^{19}$  eV a strong suppression of the energy spectrum becomes visible. The reason for this suppression of the spectrum is still unclear. Various theories about the different features of the energy spectrum have been developed. While some of these describe the flux suppression at the end of the spectrum, others provide explanations about the ankle. In the following, three popular theories and their prediction about the spectrum features are presented.

### Proton-dominance scenario

In the proton-dominance scenario it is assumed, that extra-galactic protons are the dominant component at the highest energies [11, 12]. A first theoretical prediction about the suppression at the end of the energy spectrum was done in 1966 by Greisen, Zatespin and Kuzmin. They suggested, that the interaction of protons and the CMB photons would lead to a  $\Delta$ -resonance at energies above  $5 \times 10^{19}$  eV

$$\gamma + p \rightarrow \Delta^+ \rightarrow p + \pi^0. \quad (2.3)$$

The energy loss caused by this reaction and the associated mean free path would then lead to a GZK horizon at around 100 Mpc. This horizon would suppress the observed proton flux at ultra high energies on earth [3, 4]. At the region of the ankle up to energies of  $1 \times 10^{19}$  eV, the interaction of protons with the CMB results in pair production, leading to

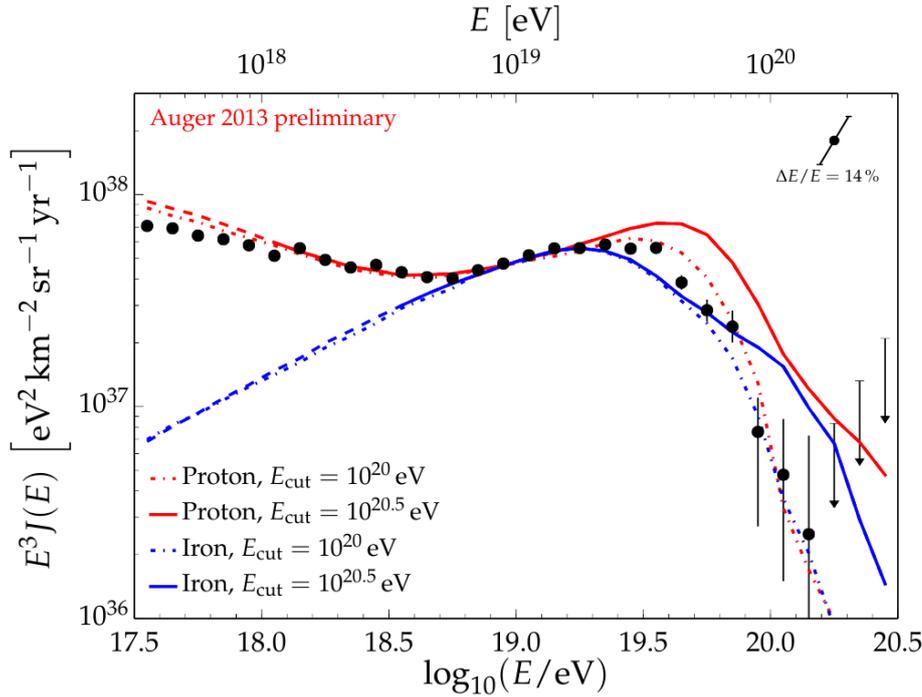


Figure 2.3.: The dip model with either iron (blue) or proton (red) primaries as dominant CRs is compared to the measured spectrum. Figure taken from [13].

a characteristic dip in the spectrum. Due to a pile-up effect in the energy range between the two processes, a bump is forming [13]. Two idealized predictions of homogeneously distributed sources, with either proton or iron primaries, are compared to the spectrum in Fig. 2.3.

### Maximum-rigidity scenario

Another reason for the flux suppression would be that the extragalactic accelerators reach their maximum energy [14]. In this maximum-rigidity scenario, it is assumed, that the maximum rigidity of all particles is of the same size  $R \sim E/Z$ . The maximum energy, a particle can be accelerated to, is thus proportional to its charge. Protons would be the first component to reach their acceleration maximum. With increasing energy heavier particles reach their acceleration maximum, up to iron. This would result in a superposition of mass dependent cut-offs at the end of the spectrum. In Fig. 2.4a the single particle spectra can be seen, as well as the sum of all spectra combined, as an attempt to explain the observed flux. At the highest energies, iron CRs would be the dominantly observed particles.

### Photo-disintegration scenario

Similar to the maximum-rigidity scenario, the photo-disintegration scenario assumes the spectrum to be a superposition of different energy spectra of different primary particles. The CR sources accelerate heavy nuclei to an energy that exceeds the threshold of photo-disintegration via interactions with photons from the CMB [15]. Lighter particles are

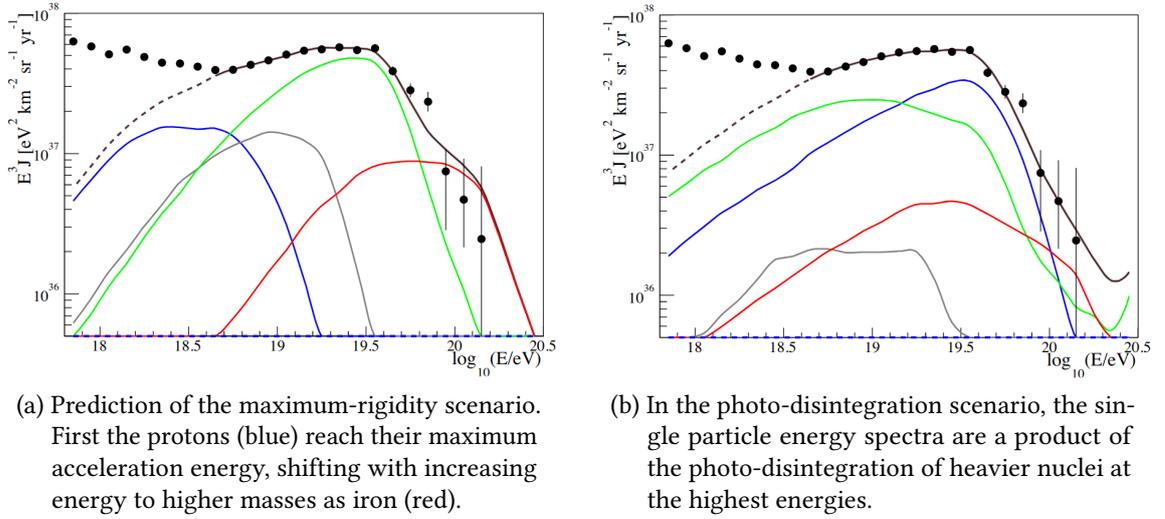


Figure 2.4.: Both figures show the prediction of either the maximum-rigidity scenario or the photo-disintegration scenario, compared to the observed energy spectrum. The proton flux is shown as blue curve. Helium is depicted in grey, Nitrogen in green and iron in red. Both Figures are taken from [13].

created as fragments of the heavier elements as a result of this process. The energies, at which the lightest particles arrive would be shifted by the daughter to parent mass ratio. At the end of the spectrum, the particles would then be related to the particles around the ankle of the spectrum. Fig. 2.4b shows a possible configuration of different particle energy spectra combined to an all-particle flux.

## 2.3. Mass Composition with Measurements of the Shower Depth

All the previously presented models are highly dependent on the charge and mass of the cosmic rays. It is therefore of great interest, to give good estimations of the mass of the cosmic rays at the highest energies. Due to the extremely low flux of UHECRs, direct detection methods, as with balloon or satellite experiments, are not viable. Ground based experiments make use of detecting the cosmic rays indirectly by the extensive air showers. As explained previously in Section 2.1, the number of electromagnetic particles of the air shower is proportional to the primary energy. The development of the air shower can be described as a function of traversed air mass, which is referred to as the *slant depth*  $X$ . It can be obtained by integrating the density of air along the arrival direction of the air shower through the atmosphere

$$X(z) = \int_z^\infty \rho(\mathbf{r}(z')) dz' \quad (2.4)$$

where  $\rho(\mathbf{r}(z))$  is the density of air at the longitudinal coordinate  $z$  of the shower axis [16]. The point, where the shower reaches its maximum is known as the *shower maximum*

$X_{\max}$ . Light particles, as for example protons, have a small cross section and are therefore interacting deep in the atmosphere. The shower reaches  $X_{\max}$  quite deep in the atmosphere. However, heavier particles interact higher up in the atmosphere due to their larger cross section. The maximum of the shower is thus reached earlier for heavy particles than for light particles. Because of this relationship,  $X_{\max}$  can be related to the logarithmic mass of the primary particle. However, the primary mass is not measurable on an event-by-event level, due to hadronic fluctuations in the shower cascade, and must be determined statistically from a distribution of shower maxima of an ensemble of air showers [16, 17].

## 3. The Pierre Auger Observatory and its Upgrade

The *Pierre Auger Observatory* is the worlds largest hybrid detector array for measuring UHECRs, located near Malargüe in Argentina [18]. It is named after the French physicist Pierre Auger, who discovered extensive air showers in 1939. It includes the *Surface Detector* (SD) [19], as well as the *Fluorescence Detector* (FD) [20]. The former includes an array of 1660 *Water Cherenkov Detectors* (WCDs), distributed over an area of 3000 km<sup>2</sup> to measure shower particles that reach the ground. The latter consists of 24 telescopes overlooking the SD array at four different locations to measure fluorescent light, emitted by the particles of the evolving shower. In 2008, three additional *High Elevation Auger Telescopes* (HEAT) were installed to detect showers with lower energy. A schematic depiction of the SD array and the FDs is given in Fig. 3.1.

### 3.1. Water Cherenkov Detectors

The SD array can sample the lateral distributions of EAS at the ground to obtain an estimate of the shower size. The 1660 WCDs are deployed in a hexagonal grid with an equidistant spacing of 1500 m (SD-1500) between each station. The SD array reaches it full efficiency above  $3 \times 10^{18}$  eV [21]. The WCDs are circular tanks with a radius of 1.8 m and are filled to a height of 1.2 m with purified water, making a total active detector volume of 12 m<sup>3</sup> [13]. A photograph of a station and a schematic can be seen in Fig. 3.2. Inside the stations, there are three *XP1805* 9-inch *photomultiplier tubes* (PMT), pointing inwards to measure Cherenkov photons, produced by the traversing particles of the EAS. Relativistic, charged particles produce Cherenkov radiation, when traveling through an electrically polarizable medium, like water. Electrons and muons lose their energy differently in the WCD due to their cross section. Electrons usually deposit their whole energy in the first few centimeters traversing the water. Muons on the other hand have a much larger mass, compared to electrons, and traverse the complete tank, emitting photons along their whole way. The PMTs have one low gain channel output at the anode and one amplified high gain channel output at the last dynode, enabling to cover a higher dynamic range to measure closer to the shower core. The analog pulses are read out in the *Unified Board* (UB), which is placed on top of the WCD and are sampled with two *flash analog to digital converters* (FADC) of 10 bit at a frequency of 40 MHz [13] The output is known as the FADC, or time trace, which is a block of 768 time bins. By integrating over the time trace and normalizing it by the value of a calibrated VEM charge, a signal can be acquired. A GPS antenna allows for communication between the detector and the *central data acquisition center* (CDAS) as well as data transmission. In order to extend the possible observation energy downwards,

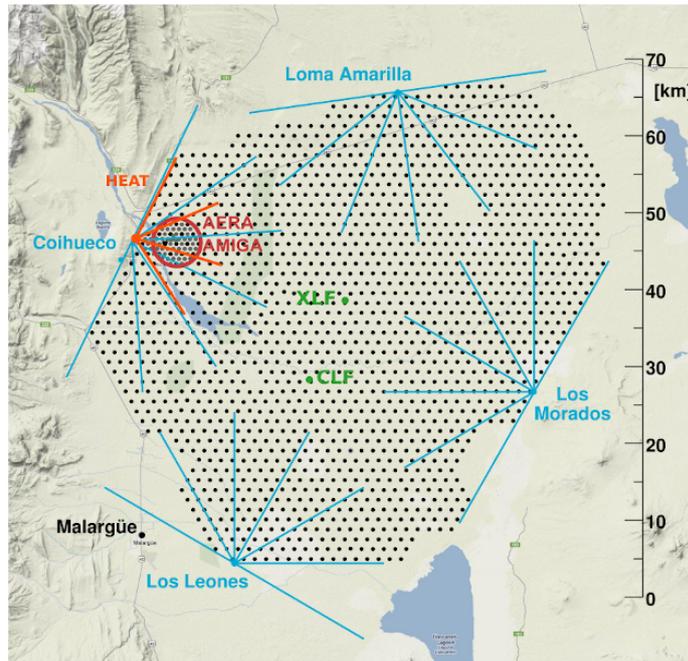


Figure 3.1.: Schematic depiction of the SD array and the FD buildings. Each black dot represents one WCD station with the spacing of 1500 m in between. The array with the denser spacing of 750 m is also visible. The field of view of the fluorescence telescopes at the 4 FD locations is marked with blue lines, as well as the view of the HEAT extension at Coihueco in red. The Central and Extreme Laser Facility for FD calibration are marked with green dots.

an additional array, called the *Infill array* (also SD-750), was built in 2011. It consists of 71 stations, with a spacing of 750 m between each WCD, lowering the energy threshold to  $3 \times 10^{17}$  eV. An additional hexagonal array with a reduced spacing of 433 m has been deployed in 2013 to further lower the energy threshold.

### 3.2. Fluorescence Detector

While the SD measures the lateral distribution of a shower, the FD observes the longitudinal development of an EAS. The 24 telescopes are installed in 4 buildings, housing 6 telescopes each and are located on the perimeter of the SD array. Fig. 3.3a shows a picture of one of the buildings. With a  $30^\circ$  by  $30^\circ$  *field of view* (FoV) of each individual telescope, each building can cover a total azimuthal FoV of  $180^\circ$ . At one site, additional three telescopes, named HEAT, are installed. These telescopes can be tilted upwards to extend the zenith FoV up to  $60^\circ$  for measuring showers from lower energy primaries, that develop at a higher altitude [23]. Particles of EAS excite nitrogen molecules in the atmosphere, that emit isotropically fluorescence radiation at the range of 300 to 400 nm. The photons from the radiation that enter the telescope bay are reflected by a segmented mirror of  $13 \text{ m}^2$  onto a camera, consisting of  $22 \times 20$  hexagonal PMTs. A schematic of a telescope is given in Fig. 3.3b. Since the light intensity of the fluorescence radiation is directly proportional to

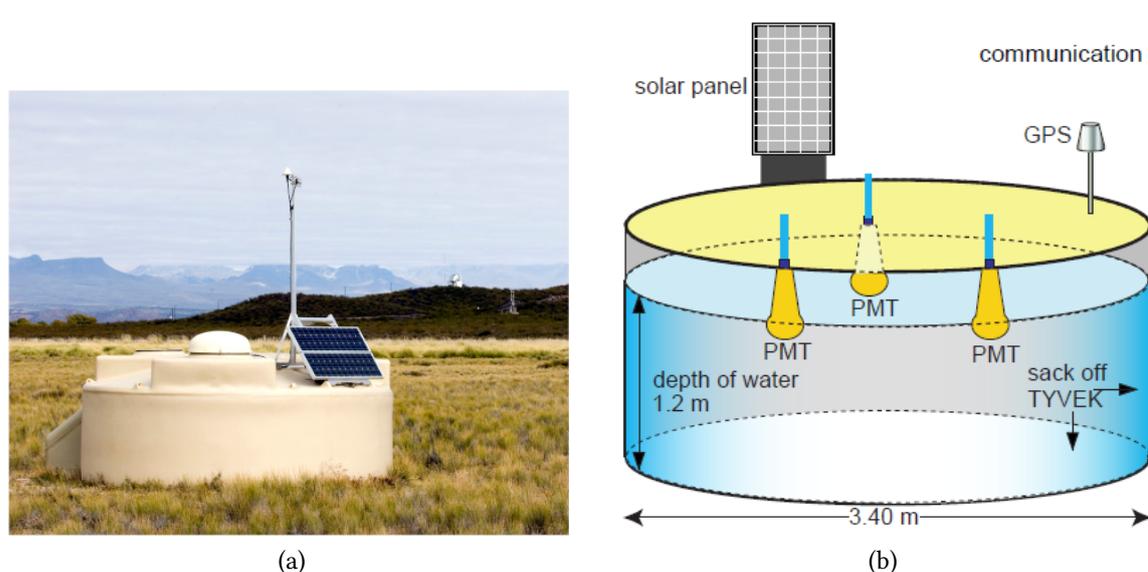


Figure 3.2.: (a) Picture of a WCD and (b) schematic view of a WCD with its components. Both pictures are taken from [22].

the number of particles in the shower, the maximum shower depth as well as the energy of the primary particle can be estimated. The duty cycle of the FD is around 15% of the SD measuring time. This comes due to the high sensitivity of the cameras, that limits their working time to clear and moonless nights with appropriate weather conditions. At the center of the SD array the *Central Laser Facility* (CLF) and the *Extreme Laser Facility* (XLF) are located. By firing a laser with known intensity and energy, the FD telescopes can be calibrated with the upwards directed tracks that are produced.

### 3.3. Current Results and Open Questions

Since the start of data collection in 2004, the Pierre Auger Observatory could provide a range of important fundamental results [25]. The flux suppression of the spectrum at energies above  $5 \times 10^{19}$  eV could be confirmed with a significance of over  $20 \sigma$  [24, 26]. The spectrum with cumulative data of 14 years of measurements can be seen in Fig. 3.4. Yet it remains unclear, whether the suppression at the highest energies arises from galactic accelerators, reaching the maximum acceleration potential or from propagation effects. The depth of the shower maximum has also been measured with Auger, combining 9 years of FD data with 5 years of data taken with HEAT [27]. The energy evolution of  $X_{\max}$ , as well as the first two central moments of the distributions are shown in Fig. 3.5. Up to energies of around  $\lg(E/\text{eV}) = 18.3$ , the observations are not in agreement with air shower simulations for a constant mass composition, but show a trend towards lighter elements [27]. The distribution then tends towards heavier nuclei for higher energies. Auger has also found a large-scale anisotropy of around 7% in the arrival direction above the ankle for energies above  $8 \times 10^{18}$  eV at a significance level of over  $5 \sigma$ . This results supports the

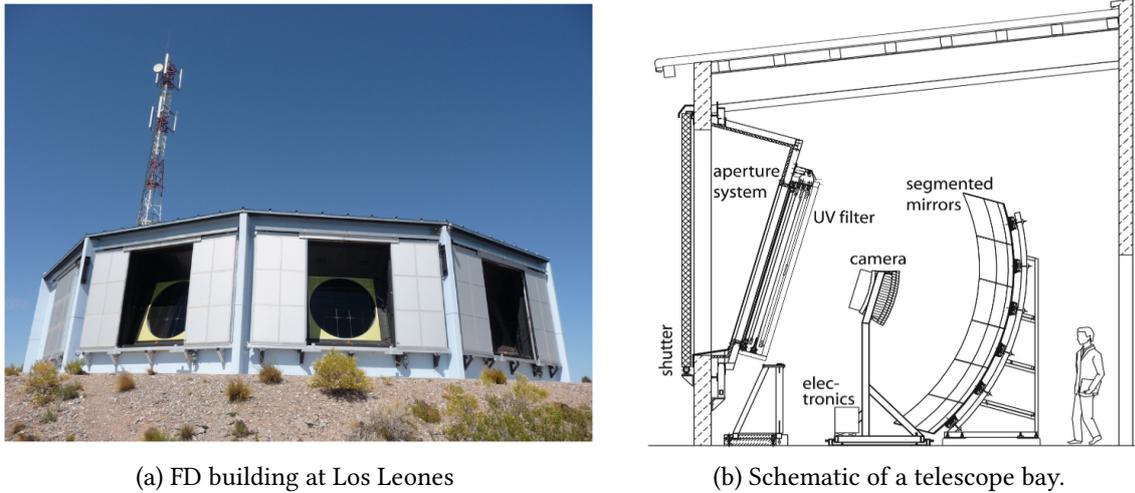


Figure 3.3.: Telescope buildings as in Fig. 3.3a are stationed at four borders of the array. Each building houses 6 telescope bays. Both images are taken from [18].

hypothesis of an extragalactic origin of UHECRs [28]. A third result is an observed muon deficit at the energy range around  $10^{19}$  eV. Depending on the model, the observed number of muons is around a factor of 2 higher than the prediction of simulations [29]. For more in-depth studies and possible discoveries of new physics, an upgrade of the observatory is currently under construction.

### 3.4. The Upgrade: AugerPrime

Some of the previously discussed open questions and results are limited to the duty cycle of 15% of the FD. This uptime rate is not sufficient enough to provide enough data for more detailed studies at the highest energies, as mass composition. An upgrade to the SD allows to retrieve results for mass composition, with a duty cycle of 100%, that is independent of the FD. One of the main ideas of the AugerPrime upgrade is the disentanglement of different shower components, mainly the muonic and electromagnetic component. By using an additional detector to the WCD, the different responses of both detectors to the shower components can be used to discriminate between these, thus enhancing the ability to reconstruct for mass composition at the highest energies. In the following, some main components of the AugerPrime upgrade are summed up briefly.

#### Scintillator Surface Detectors

A *Scintillator Surface Detector* (SSD) will be mounted on top of the WCD. A model of the setup is shown in figure 3.6. It includes two planes, consisting of 48 plastic scintillator bars in total. The scintillator bars have a length of 1.6 m, width of 5 cm, thickness of 1 cm, and are placed in a box to shield them from varying weather conditions and block external light [13]. The bars themselves are connected via wavelength shifting fibers to

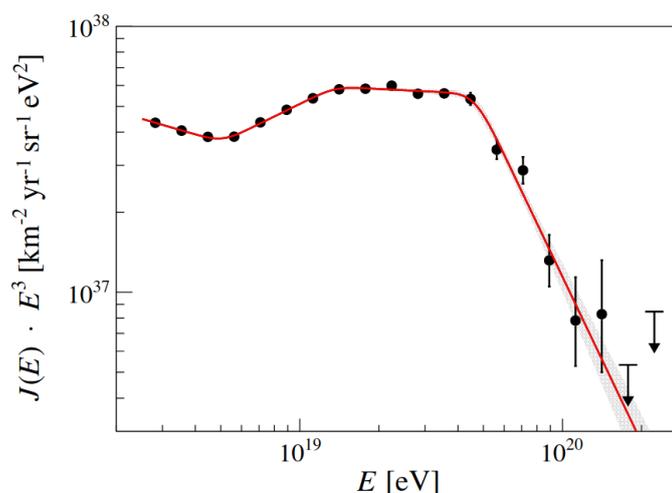


Figure 3.4.: Combined energy spectrum with 14 years of data taking. Image taken from [24].

a photomultiplier tube. With the SSD mounted on top of the WCD, it receives signals from muons and even low energy electromagnetic particles [30]. The SSD has a higher sensitivity to the electromagnetic component compared to the muonic component. While the electromagnetic particles only penetrate a few centimeters into the WCD, leaving a faint signal, the muons traverse the detector fully, producing a larger signal. However, the vast majority of both components traverse the SSD completely, producing equal signals. The signal measured with the SSD is given in *Minimum Ionizing Particles* (MIP). 1 MIP corresponds to one particle traveling vertically through the SSD. Since the SSD is mounted on top of the WCD, both detectors measure the same portion of the air shower. With the complementary measurement of the detectors, it is possible to improve the accuracy on the reconstructed composition of extensive air showers. The SSD operates in a subsidiary mode to the WCD, meaning it is triggered by the WCD. The first twelve SSD prototypes were deployed during September 2016 in the field as part of the *Engineering Array* (EA). At March 2019 the deployment of SSD started and increased over the course of the year.

### Small Photo-Multiplier Tube

Additionally to the three PMTs in the WCD, a fourth, small PMT is added in the center of the tank as well to extend the dynamic range. With the smaller active surface, the light collection is reduced and larger signals can be detected. The dynamic range will thus match the saturation limits of the SSD and allows to measure high energy showers up to distances of 200 m to the shower axis [32].

### Upgraded Electronics

The Unified Board gets exchanged with the *Upgraded Unified Board* (UUB). Currently only 6 channels are available for data acquisition. These are used by each of the high and low gain channels of the three WCD PMTs. More channels are needed with the new SSD and

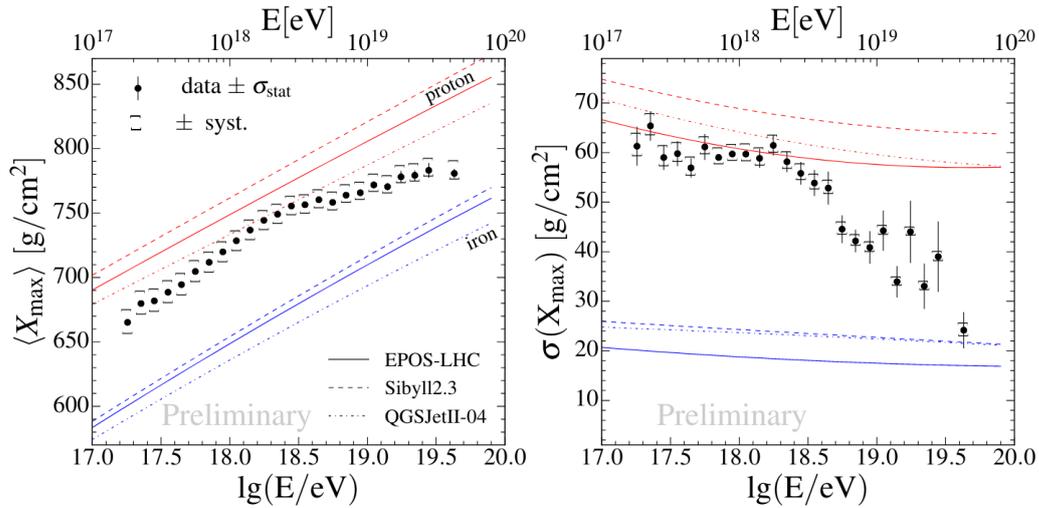


Figure 3.5.: The mean (left) and standard deviation (right) of the depth of the shower maximum as a function of energy. The red and blue lines indicate predictions for proton and iron primaries. Image taken from [27].

the small PMT. The UUB makes use of 10 channels and also the sampling frequency is increased from 40 MHz up to 120 MHz with a 12 bit FADC [33].

### Underground Muon Detector

In the SD-750 array, muon detectors will be deployed. These detectors will be buried 2.3 m below the ground, at the location of 61 WCD stations in the array. The detectors are referred to as *Auger Muon Detectors for the Infill Ground Array* (AMIGA). AMIGA will be most sensitive to the muonic shower component, because of its depth in the ground. Therefore, a direct calibration of the combined WCD and SSD reconstruction of the muon content is possible [34, 35]. Furthermore, the composition sensitivity of the energy spectrum in the region of the ankle will be improved.

### Radio Upgrade

In addition to the SSDs, a small radio antenna will be deployed to each of the stations in the array. Air showers have a large footprint of radio emission above zenith angles of  $60^\circ$ . While the SSD and WCD detectors together are sensitive to vertical air showers, the combination of WCD and radio antennas will extend the measurements for more inclined air showers [36]. Further measurements of the mass composition at large zenith angles are thus possible.

### Extended FD Uptime

The FD operation time can be extended to measure later in the morning and earlier in the evening. Lowering the high voltage as another operation mode reduces the PMT gains and



Figure 3.6.: Photography of an SSD on top of the WCD. Image taken from [31].

irreversible deterioration of the PMT sensitivity is avoided. The duty cycle can therefore be expanded by 50%, increasing the uptime from 15% to 20% [13].



## 4. Surface Detector Reconstruction

One goal of this thesis is the estimation of mass composition of the cosmic rays at the highest energies, using only the surface detector. Therefore, understanding the reconstruction procedure of the WCD and SSD is the basis for the consecutive work and analysis. In the first two sections a short summary of the current SD reconstruction procedure used at Auger, which is extensively explained in [31, 37, 38, 39, 40] is given. The subsequent sections focus on possible improvements to the SSD reconstruction. In Section 4.3 the resolution of the SSD reconstruction is evaluated and a possible improvement, by propagating core uncertainties, is developed. Afterwards, some events where the reconstruction of the shower size is still deviating strongly from the prediction are examined in Section 4.4.

### 4.1. Reconstructing Events with the WCD

Reconstructing air shower events at Auger with the SD is performed in multiple steps, starting from station-level hardware triggers up to the reconstruction of the arrival direction and energy of the primary particles. While for the WCD the parameterizations are done with a data-driven approach, the SSD reconstruction is currently derived from simulation-based data. With the deployment and data acquisition of more SSD for the AugerPrime upgrade, a data-driven reconstruction for the SSDs will be available in the near future as well. The following steps are summaries of [37, 38, 39].

#### Triggers

Different hard- and software triggers are used to identify air shower events and distinguish them from background, produced by low energy showers. The trigger chain consists of single station level triggers, central data station triggers, physics event selection and quality triggers. The lowest trigger level is the hardware T1 trigger, that operates on single station level. Either one of the following two criteria must be valid in order to satisfy the T1 trigger:

1. The *Threshold trigger* (Thr1) requires all three PMTs to measure a signal above 1.75 VEM in coincidence. The trigger is relevant to detect muons in highly inclined showers.
2. The *time-over-threshold trigger* (ToT) requires at least 2 PMTs to measure a signal above 0.2 VEM for at least 13 FADC bins in a time window of 120 bins. This trigger is used for the discrimination of background at low energy events with a zenith angle below  $60^\circ$ .

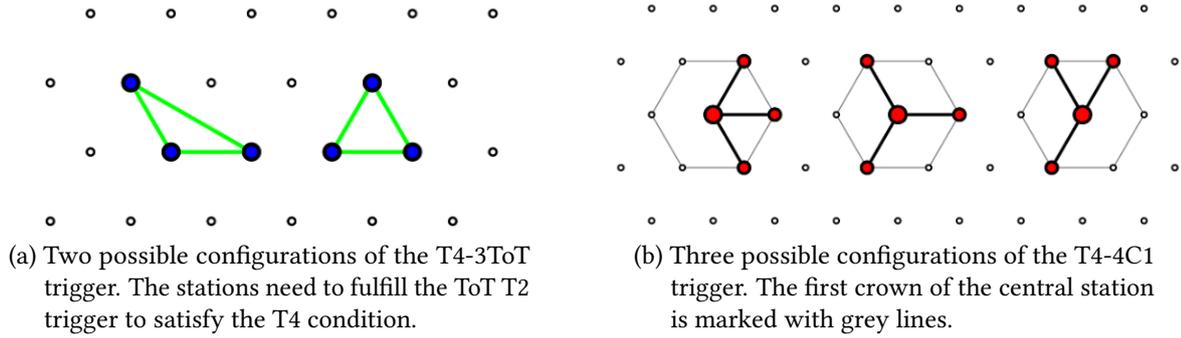


Figure 4.1.: Schematic depiction of the different T4 trigger cases from [38].

The next trigger is the T2 trigger, which is actually a stricter version of the T1 trigger. The Thr1 condition is raised to have all three PMTs measure at least a signal above 3.2 VEM, while the ToT condition stays the same. The information is sent to the *central data acquisition system* (CDAS), if either of the T2 triggers are satisfied. There, the T3 trigger searches for time coincidences of the signals that passed the T2 triggers. To satisfy the T3 trigger, one of the two possible station patterns must be fulfilled:

1. 3-fold condition (T3-3ToT): 3 neighboring stations with a coincidence of ToT triggers. In addition, the stations have to be part of the first two crowns of the station with the largest signal. A crown describes all neighboring stations with the same distance to a given central station.
2. 4-fold condition (T3-4T2): 4 stations with any T2 trigger in coincidence.

An event can be as well promoted to the T3 level, if an FD trigger is fulfilled additionally. Once an event passes the T3 trigger it is saved for later analysis.

On the next level, the T4 trigger is a physics selection, that uses stricter T3 criteria to filter out accidentally triggered stations or lightning events. The T4 trigger can be satisfied in the following ways:

1. T4-3ToT: 3 nearby stations, that fulfill the ToT T2 trigger and form an equilateral or isosceles triangle. A depiction of this is given in Fig. 4.1a.
2. T4-4C1: From 4 stations, that satisfy one of the T2 trigger conditions, 3 must lie within the first crown of the fourth station. Possible patterns can be seen in Fig. 4.1b.

The last trigger is a quality trigger, that can be prior or post reconstruction. The 6T5 trigger requires that the station with the largest signal - also called the hottest station - is surrounded by 6 functioning, but not necessarily triggered, stations. This requirement ensures, that no events are selected, where the shower might lie mainly outside of the array. For anisotropy studies the T5 trigger can be relaxed to a 5T5 trigger in order to increase the amount of events.

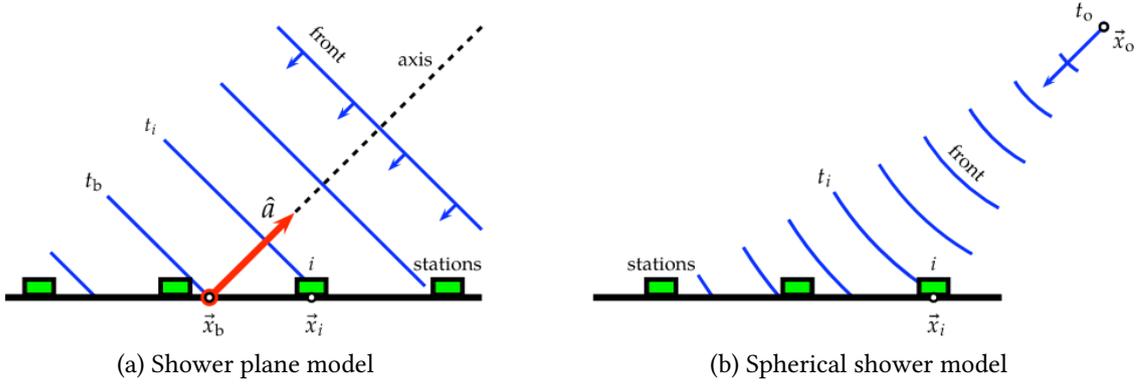


Figure 4.2.: The shower plane model gives a first approximation of the shower axis and geometry. If enough stations are triggered, the spherical shower model can be used. [37]

## Geometry

The main goal of the SD reconstruction is the estimation of the primary energy and its arrival direction. The latter is done by reconstructing the shower geometry, which means to determine the position where the center of the shower landed and the timing of the signals. As a first approximation of the shower front a plane is used. The shower plane travels along the shower axis  $-\hat{a}$  and lands at the barycenter  $\vec{x}_b$  at the time  $t_b$ . The shower front is assumed to move at the speed of light  $c$  and passes through a point on the ground  $\vec{x}$  at the time  $t_{\text{sh}}(\vec{x})$

$$ct_{\text{sh}}(\vec{x}) = ct_b - \hat{a}(\vec{x} - \vec{x}_b). \quad (4.1)$$

The barycenter  $\vec{x}_b$  is first approximated by using the signal-weighted center-of-mass of a station triangle with the highest sum of signals. Eq. (4.1) can now be solved to get an estimate for the shower axis  $\hat{a}$  and the arrival times of the shower plane at any given point  $\vec{x}$ . A visualization of the shower plane can be seen in Fig. 4.2a. A more precise estimation can be done with a spherical approximation of the shower front. At the starting time  $t_o$  and starting point  $\vec{x}_o$  a sphere is concentrically inflating with the speed of light  $c$  and the arrival time  $t_{\text{sh}}$  of such a shower front at a certain point  $\vec{x}$  is

$$ct_{\text{sh}}(\vec{x}) = ct_o + |\vec{x} - \vec{x}_o|. \quad (4.2)$$

The virtual spherical origin can be related to the shower axis  $\hat{a}$  by:

$$\vec{x}_o = \vec{x}_c + R_o \hat{a}, \quad (4.3)$$

with the radius  $R_o$  of the sphere at the impact point  $\vec{x}_c$ . A schematic of the spherical model is given in Fig. 4.2b.

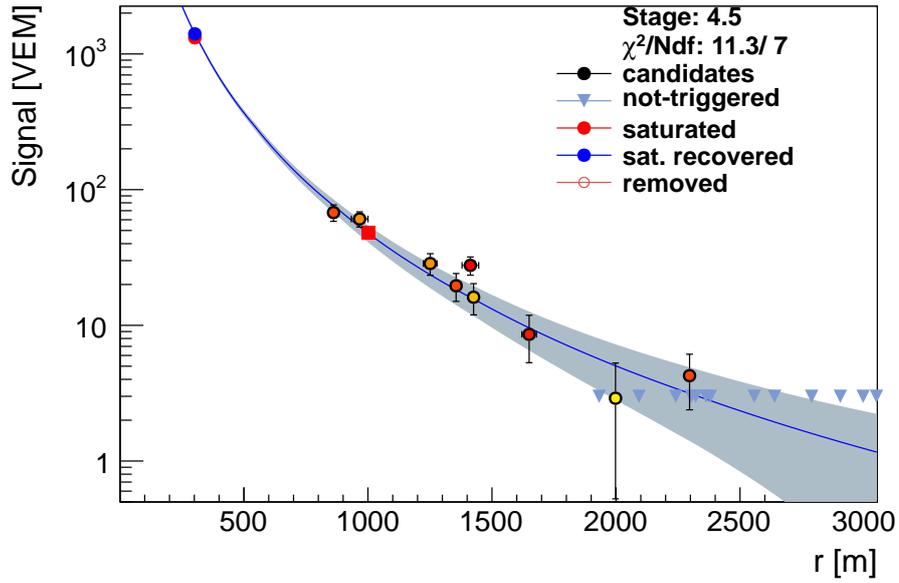


Figure 4.3.: Example reconstruction from Offline of the lateral distribution of a shower event with differently treated stations.

### Fit of the Lateral Distribution

The density of particles at ground (and therefore the measured signal) as a function of the shower distance is modeled by the *Lateral Distribution Function* (LDF), which is useful to reconstruct the energy and composition of the primary particle (see figure 4.3). The LDF used throughout this thesis is a modified *Nishimura-Kamata-Greisen* (NKG) type function

$$S(r) = S(r_{\text{opt}}) f_{\text{NKG}}(r), \quad (4.4)$$

$$f_{\text{NKG}}(r) = \left( \frac{r}{r_{\text{opt}}} \right)^{\beta} \left( \frac{r + r_{\text{scale}}}{r_{\text{opt}} + r_{\text{scale}}} \right)^{\gamma}. \quad (4.5)$$

$\beta$  and  $\gamma$  are parameters of the LDF that differ for different shower events and  $S(r_{\text{opt}})$  is the *shower size* at the optimal distance  $r_{\text{opt}}$ .  $r_{\text{opt}}$  describes the optimal distance to the shower axis, where the uncertainty of the LDF varies least. This distance has been found to be dependent on the spacing between the WCDs and is for the SD-1500 array approximately 1000 m ( $r_{\text{opt}} = 1000$  m) [41]. The scaling distance  $r_{\text{scale}}$  is set to 700 m. The parameters  $\beta$  and  $\gamma$  are dependent on zenith angle  $\theta$  and  $S(1000)$  and are fit together in a multi-event fit [42].

To fit the LDF parameters for a certain event, the principle of a *maximum likelihood* is used. With given uncertainties on the measured signals, the parameters of the LDF are adjusted to maximize the likelihood of the observed data for the given model. The following description is a summary of the actual functions used in the *LDFfinder module* of the software Offline used by the Pierre Auger Observatory [38]. The maximum likelihood method allows for the inclusion of more than only triggered stations. Thus stations can be distinguished by the signal they measure:

1. *Non-saturated stations*  $f_G$ , treated with Gaussian statistics, if the signal is above a certain threshold.

If the signal of these stations is too low, they will be treated as:

2. *Low signal stations*  $f_P$ , including Poissonian statistics.

Two more kinds of stations can be distinguished:

3. *Saturated stations*  $f_{\text{sat}}$ , that received signals above the linear range of the PMTs.
4. *Zero-signal stations*  $f_{\text{zero}}$ , that did not trigger during an event.

With  $S_i$  as the individual station signals and  $\mu_i$  as the expected signals from the LDF model, the likelihood function is therefore

$$L = \prod_i f_G(S_i, \mu_i) \times \prod_i f_P(S_i, \mu_i) \times \prod_i f_{\text{sat}}(S_i, \mu_i) \times \prod_i f_{\text{zero}}(S_i, \mu_i). \quad (4.6)$$

For easier and faster calculation the log likelihood method is used. By using the natural logarithm of the functions the products change to sums and all terms relating to each measured signal can be summed to obtain the likelihood

$$L = \sum_i \ln f_G(S_i, \mu_i) + \sum_i \ln f_P(S_i, \mu_i) + \sum_i \ln f_{\text{sat}}(S_i, \mu_i) + \sum_i \ln f_{\text{zero}}(S_i, \mu_i). \quad (4.7)$$

The likelihood factors of the non-saturated stations and low signal stations look as follows. The non-saturated station signals are described with a Gaussian distribution

$$f_G(S_i, \mu_i) = \frac{1}{\sqrt{2\pi}\sigma_i^2} \exp\left(-\frac{(S_i - \mu_i)^2}{2\sigma_i^2}\right), \quad (4.8)$$

$$\ln f_G(S_i, \mu_i) = -\frac{1}{2} \ln(2\pi\sigma_i^2) - \frac{(S_i - \mu_i)^2}{2\sigma_i^2}. \quad (4.9)$$

A continuous Poisson distribution is used for low signals

$$f_P(S_i, \mu_i) = \frac{\mu_i^{S_i}}{S_i!} e^{-\mu_i}, \quad (4.10)$$

$$\ln f_P(S_i, \mu_i) = S_i \ln \mu_i - \mu_i - \ln \Gamma(S_i + 1). \quad (4.11)$$

To distinguish between low and non-saturated signals, a conversion of the signals to their corresponding particle numbers needs to be performed. To convert the signals measured by a WCD to a corresponding particle number, the uncertainty on the signals themselves has to be determined. By using doublet stations, which are stations separated by 11 m, the signal variance  $\sigma_S$  can be obtained. Even though the stations are not measuring the same particles, they measure the same approximate location in the shower detector plane [43,

44, 45, 46]. With the difference in detector signals, the signal uncertainty can be found. The following parameterization of the uncertainty is used, where  $\sigma_S$  is dependent on the zenith angle  $\theta$  of the incoming particles

$$\sigma_S = P(\theta)\sqrt{S}, \quad (4.12)$$

$$P(\theta) = a + b \sec \theta. \quad (4.13)$$

The parameters have been experimentally determined to be  $a = 0.32$  and  $b = 0.42$  [44, 45]. In the course of an update on the signal uncertainty model a new model is used [46]

$$\frac{\sigma_S}{\sqrt{S}} = 0.865(1 + 0.593(\sec \theta - 1.22)). \quad (4.14)$$

When converting from signal to the particle number, the different behavior of muons and electromagnetic particles in the water tanks has to be taken into account. The muonic component is dominant for high zenith angles and therefore the conversion is dominated by the number of muons passing through the detector. With the signal measured in VEM, only a muon passing through the tank completely vertically will contribute 1 VEM. For smaller zenith angles, the electromagnetic component becomes important. Since the electromagnetic particles will not cross the whole tank, more particles of that component are required to leave a combined signal of 1 VEM in the tank. A Poisson factor  $\rho$ , which is dependent on the signal uncertainty, is estimated to convert the signals of the WCD to particle numbers [39]

$$n = \rho S, \quad (4.15)$$

$$\rho = \max\left(1, \frac{1}{f(\theta)^2}\right), \quad (4.16)$$

$$f(\theta) = 0.865(1 + 0.593(\sec \theta - 1.22)). \quad (4.17)$$

The factor is shown in Fig. 4.4 as function of zenith angle. With Eq. (4.15), the signal is converted to a corresponding particle number, and a distinction between low and high signal stations can be made. If the converted particle number of a station is below a threshold of 15, the signal contributes to the low signal term in the likelihood that uses the Poisson distribution.

Combined with the position of the shower core and the arrival direction, the LDF from Eq. (4.4) has a total of 8 free parameters, where 5 are attributed to the geometry and the remaining 3 parameters come from the  $S(1000)$  and the slopes of  $\beta$  and  $\gamma$ . Currently, the fit procedure implemented in Offline consists of several fit stages, where some of the parameters are fixed to previously parameterized values to reduce the number of free parameters. At the end of the reconstruction procedure, one gets the reconstructed shower size  $S(1000)$  of the WCD measurements which can be used as an estimator for the energy of primary particle.

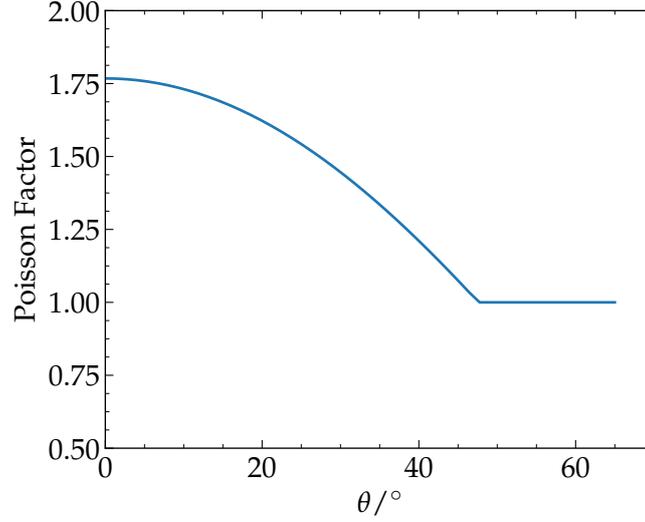


Figure 4.4.: The Poisson factor  $\rho$  to convert the signal to particle number. At low zenith angles electromagnetic particles become more important and more particles contribute to a given signal.

### Energy Estimation

An estimation of the energy of the primary particle can be done with the shower size  $S(1000)$ . However, the shower size is dependent on the zenith angle  $\theta$  of the primary particle. This is due to the attenuation of the electromagnetic component in the atmosphere. To account for this effect and get a zenith independent estimator of the shower size, a correction is applied to  $S(1000)$

$$S_{38} = \frac{S(1000)}{f_{\text{Att}}(\theta)}. \quad (4.18)$$

The attenuation function  $f_{\text{Att}}(\theta)$  is a third order polynomial of the form

$$f_{\text{Att}}(\theta) = 1 + ax + bx^2 + cx^3, \quad (4.19)$$

with the variable  $x = \cos^2 \theta - \cos^2(38^\circ)$ . The free parameters  $a$ ,  $b$  and  $c$  are obtained with a fit of the intensities of the events at different zenith angles and an isotropic flux in  $\cos^2 \theta$  is assumed. The new variable  $S_{38}$  can be interpreted as the theoretical shower size of an event, if the shower arrived at a zenith angle of  $38^\circ$ . This method is also referred to as the *Constant Intensity Cut* (CIC). With the addition of weather and atmospheric corrections to  $S_{38}$  the energy can be estimated by

$$E_{\text{SD}} = A \left( \frac{S_{38}}{\text{VEM}} \right)^B \text{ EeV}. \quad (4.20)$$

The free parameters  $A$  and  $B$  are obtained through calibration of  $S_{38}$  with high-quality events that include measurements of SD and FD.

## 4.2. Reconstructing Events with the SSD

For the SSD, a reconstruction of  $S(1000)$  can be done, similar to the reconstruction procedure with the WCD and will be summarized here shortly. For a detailed description of the SSD reconstruction see [31] and [40].

Similar to the WCD, an LDF of the form shown in Eq. (4.4) and Eq. (4.5) is chosen. For initial  $\beta$  and  $\gamma$  parameters, a parameterization of the form

$$\beta(S_{1000}, \theta) = a_\beta + b_\beta s + (c_\beta + d_\beta s) \sec \theta + (e_\beta + f_\beta s) \sec^2 \theta, \quad (4.21)$$

$$\gamma(S_{1000}, \theta) = a_\gamma + b_\gamma s + (c_\gamma + d_\gamma s) \sec \theta + (e_\gamma + f_\gamma s) \sec^2 \theta, \quad (4.22)$$

is chosen with  $s = \lg S_{1000}$ . The log-likelihood can be written up explicitly with the error function as

$$L = -\frac{1}{2} \ln(2\pi\sigma_i^2) - \frac{1}{2} \left( \frac{S_i - \hat{S}_i}{\sigma_i} \right)^2 - \ln \left[ \frac{1}{2} \left( 1 + \operatorname{erf} \left( \frac{\hat{S}_i}{\sqrt{2}\sigma_i} \right) \right) \right], \quad (4.23)$$

where  $S_i$  denotes the observed signal,  $\hat{S}_i$  the expectation and  $\sigma_i$  the signal uncertainty. Additionally, a distance cut is introduced, so that 95% of the SSDs below the cut distance have at least a signal of 1 MIP. Using the shower core position derived during the WCD LDF fit as well as  $\beta$  and  $\gamma$  from the parameterization as initial guesses, a multi-event fit is performed similar to the WCD, fixing and fitting the shape parameters and shower size successively. The SSD operates in a subordinary mode to the WCD, which means, that only if a WCD is triggered and used as a candidate station the associated SSD will be used in the fit as well. Therefore, an SSD signal can be as low as zero as well. To account for this, a distance cut is introduced to keep all stations below a certain distance, so that 95% of the stations below that distance have a signal of at least 1 MIP [31].

A model for the uncertainty of the SSD signals has been developed in [47] and improved in [31]. Contrary to the WCD doublets, which are separated by 11 m, for the SSD uncertainty model so called *pseudo doublets* have been chosen. The model was parameterized on simulations and due to thinning algorithms, the doublet pair has been chosen to be located 1000 m from the shower core and at azimuthal angles of  $90^\circ$  and  $270^\circ$  in the shower detector plane. This selection as pseudo doublet stations ensures that the particles, that arrived at the SSDs, traversed the same amount of atmosphere. Therefore, biases due to attenuation effects are reduced. Similar to the WCD, the functional form of the uncertainty model has been determined

$$\frac{\sigma_S}{\sqrt{S}} = (1.449 \pm 0.001)(1 + (0.175 \pm 0.002)(\sec \theta - \sec 35^\circ)). \quad (4.24)$$

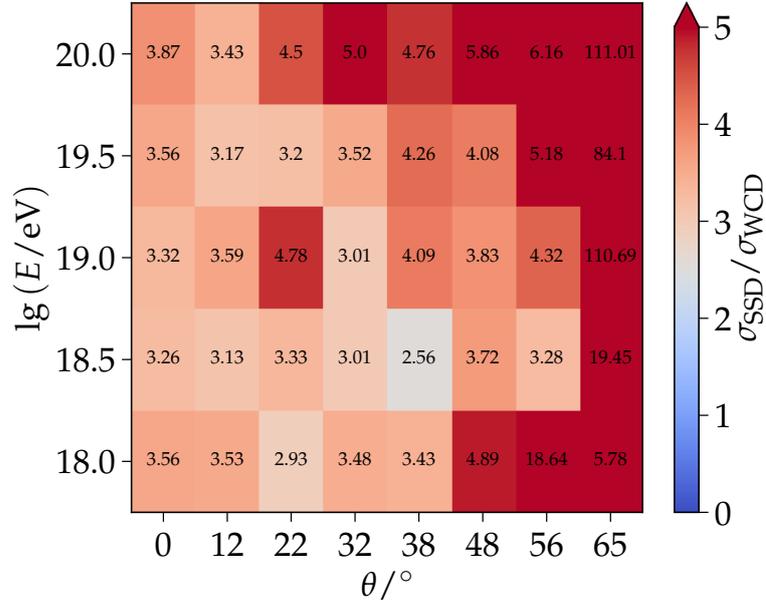


Figure 4.5.: The ratio of the SSD and WCD resolutions  $\sigma_i$  in bins of primary energies and zenith angles. At zenith angles above  $\theta/^\circ = 65$  and energies below  $\lg(E/\text{eV}) = 18.5$  the array is not at full efficiency. Although those regions will be excluded, the resolution of the SSD is still 3 to 4 times worse than the resolution of the WCD.

Table 4.1.: Fixed library of CORSIKA simulations with different combination of parameters. For each combination, 120 events were reconstructed.

primary	proton
hadronic interaction model	EPOS-LHC
$\lg(E/\text{eV})$	18.5, 19.0, 19.5, 20.0
$\theta/^\circ$	0, 12, 22, 32, 38, 48, 56, (65)

## 4.3. Propagating Core Uncertainties in Subordinate Detector Reconstructions

### 4.3.1. Shower Size Resolution of the SSD Reconstruction

The reconstruction of the LDF with the signal measurements of the WCD and SSD results in differing shower sizes  $S(1000)$  for both detectors, since the WCD and SSD are sensitive to different shower components, as discussed in the previous Chapter. The reconstructed shower size  $S_{1000}^{\text{rec}}$  can be compared to the Monte Carlo value  $S_{1000}^{\text{MC}}$  and the bias  $\langle S_{1000}^{\text{rec}}/S_{1000}^{\text{MC}} - 1 \rangle$ . Also the resolution  $\sigma(S_{1000}^{\text{rec}}/S_{1000}^{\text{MC}})$  can be calculated. For the following analysis a CORSIKA shower library with fixed energies and zenith angles was used. The parameters of this library are listed in Table 4.1. For each combination of parameters, 120 showers were simulated, making a total of 3360 unique air showers. To get an impression

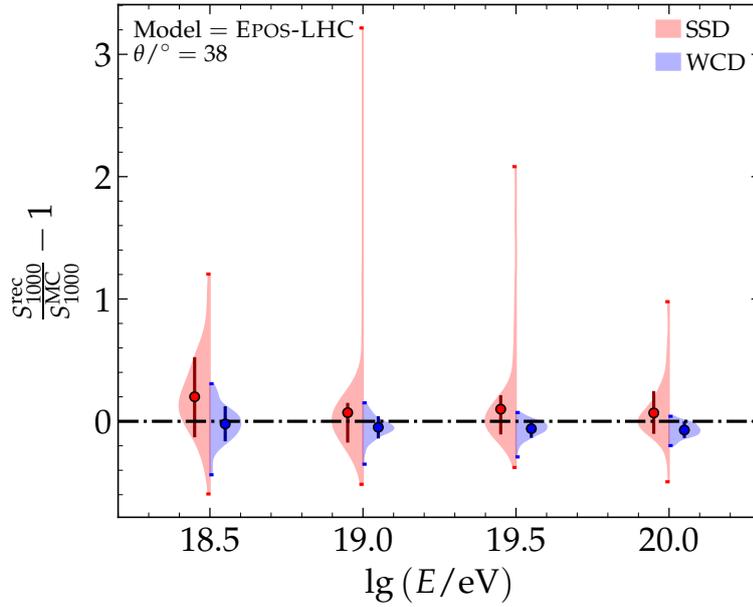


Figure 4.6.: The fraction of the SSD and WCD resolutions  $\sigma_i$  in bins of primary energies and zenith angles. At zenith angles above  $\theta/^\circ = 56$  and energies below  $\lg(E/eV) = 18.5$  the array is not at full efficiency. Although those regions will be excluded, the resolution of the SSD is still 2 to 4 times worse than the resolution of the WCD.

of the resolution of the SSD, the ratio of SSD to WCD resolution is shown in Fig. 4.5. It is evident that the resolution of the SSD is 2 to 4 times higher than the resolution of the WCD. For a more detailed picture, the distributions of reconstructed to Monte Carlo shower size for the WCD and SSD at a zenith angle of  $\theta/^\circ = 38$  are shown in Fig. 4.6. For the WCD the reconstructed shower size has an average resolution of 8% and a bias of -6%. The reconstructed shower size of the SSD shows incredibly big outliers with predictions more than 2 to 4 times as big as the true signals. When reconstructing the LDF of the SSD, the distance of stations to the shower axis in the shower plane is used. The core position is fitted during the LDF fit for the WCD and fixed for the subsequent SSD reconstruction [31]. Since the SSDs are mounted on top of the WCDs, the reconstructed relative WCD distances are also used for the SSDs and thus the SSD reconstruction is dependent on the uncertainty of the core. In Fig. 4.7 an example of an SSD LDF reconstruction, where the uncertainties in the core have not been taken into account is given. Using a simulated event in Offline [48] gives access to the true Monte Carlo distances  $r_{MC}$  to the shower axis. Slight differences from the reconstructed core position to its true location give rise to deviating station distances  $r_{rec}$ . The uncertainty on the core position translates to an uncertainty in the station distance  $r$  in the LDF which is currently not taken into account. The model used to describe signal uncertainties parameterizes only the sampling fluctuations, which can be described by underlying Poisson statistics<sup>1</sup>. Due to the nature of Poissonian fluctuations,

<sup>1</sup>The sampling fluctuations measured by WCDs were derived from coincident measurements of two WCD stations placed approximately 10 m apart. These stations, although placed at nearly identical locations in the shower-detector plane, measure entirely different particle samples.

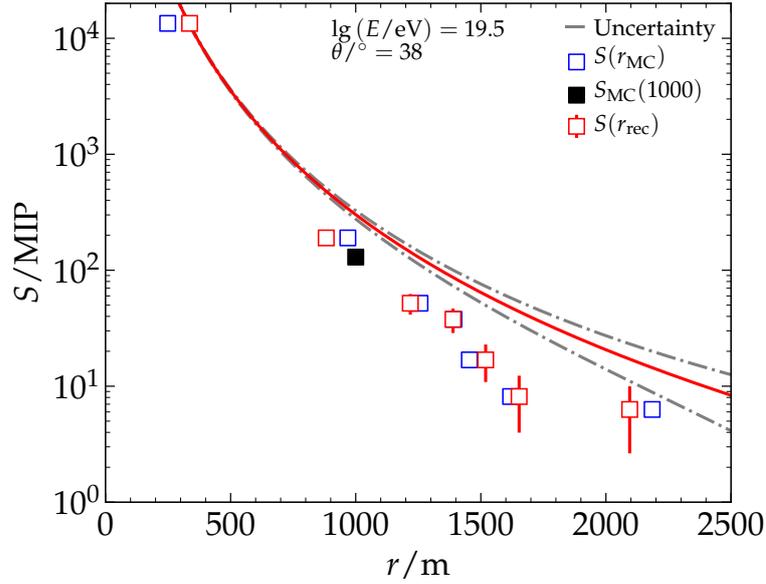


Figure 4.7.: An example reconstruction of a MC generated event without propagation of core uncertainties. The blue markers denote the MC distances of the stations, the red markers the reconstructed distances. Using a dense station ring at 1000 m, the shower size  $S_{MC}(1000)$  is obtained as the average and shown here as a black marker. The gray error bands show the uncertainty of the signal due to sampling fluctuations. The uncertainty on the signal for the hottest station is extremely small and therefore makes this station dominate the fit.

the sampling uncertainties decrease relative to the total signal for higher values. Thus, the stations close to the core will have the highest weight in the LDF reconstruction due to their small uncertainty. In the case of the SSD reconstruction, the LDF rises very steeply closer to the shower core and hence the signal increases faster with smaller values of  $r$ . Thus, an incorrectly reconstructed distance  $r_{rec}$  close to the core can result in a variation of the LDF signal prediction of several hundred MIP and a resulting misreconstructed LDF.

#### 4.3.2. Propagating the Core Uncertainty to the SSD Reconstruction

Considering all quantities relative to the true core position, it can be assumed that the reconstructed core position  $(x_r, y_r)$ , is a two-dimensional normal distribution, i.e.  $x_r$  and  $y_r$  are independent and distributed according to the normal distribution as  $\mathcal{N}(x_r; 0, \sigma_c)$  and  $\mathcal{N}(y_r; 0, \sigma_c)$ , respectively with the core uncertainty  $\sigma_c$ . The true distance  $r$  of a station located at coordinates  $(x, y)$  from the core is given by  $r^2 = x^2 + y^2$ . Since only the reconstructed core position can be accessed, the reconstructed distance  $r_r$  is given by  $r_r^2 = (x - x_r)^2 + (y - y_r)^2$ . When  $r \gg \sigma_c$ , the reconstructed distance follows a normal distribution  $\mathcal{N}(r_r; r, \sigma_c)$  and the uncertainty of the distance,  $\delta r \sim r - r_r$ , is thus equal to  $\sigma_c$ . For stations close to the core, the distribution of  $r_r$  changes from the normal to the

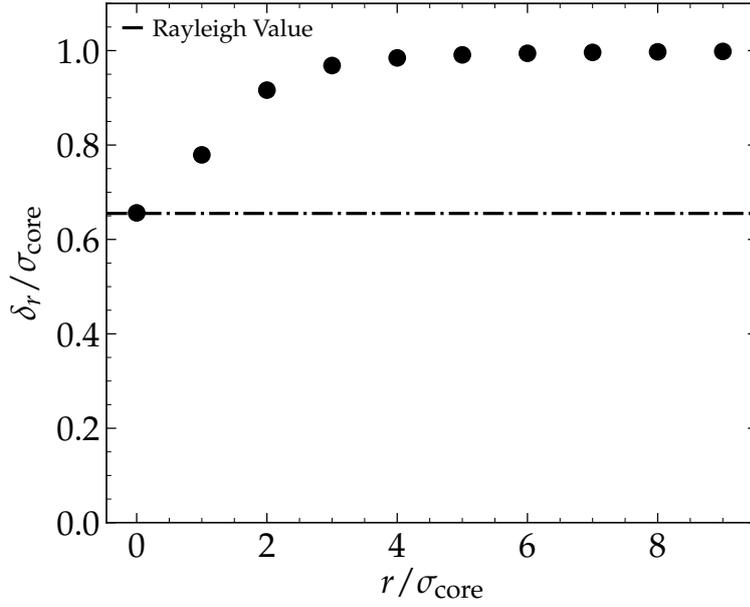


Figure 4.8.: Evolution of the station position uncertainty over the distance to the shower axis. Both are evaluated relative to the core uncertainty. The relative positional uncertainty  $\delta r/\sigma_c$  is equal to the Rayleigh number  $\rho \approx 0.655$  at the position of the core. It approaches the full core uncertainty at a distance of around double the core uncertainty.

Rayleigh distribution

$$\mathcal{R}(r; \sigma_c) = \frac{r}{\sigma_c^2} \exp(-r^2/2\sigma_c^2), \quad \text{defined for } r \geq 0. \quad (4.25)$$

The standard deviation of this distribution is smaller, namely  $\delta r = \rho \sigma_c$ , where  $\rho = \sqrt{2 - \pi/2} \approx 0.655$  is the Rayleigh number. For stations at distances of  $r \ll 2\sigma_c$  to the core, the uncertainty  $\delta r$  thus starts at the Rayleigh value  $\rho\sigma_c$  and quickly approaches the value  $\sigma_c$  at larger distances, as can be seen in Fig. 4.8.

The reconstructed core uncertainty  $\sigma_c$  for the 1500 m grid reaches a maximum value of about 100 m. For stations at a distance to the shower axis of approximately 200 m, the core resolution, relative to the distance itself, is not negligible. The uncertainty of the core position  $\sigma_c$  is dependent on zenith angle and shower size and has been parameterized as part of the analysis performed in the context of [37], although the parameterization itself is not included in this reference. The parameterization is given by

$$\sigma_c = a + b \lg\left(\frac{S_{1000}}{\text{VEM}}\right) + c \lg^2\left(\frac{S_{1000}}{\text{VEM}}\right), \quad (4.26)$$

where the parameters  $a$ ,  $b$  and  $c$  are given as

$$a = (112.71 \pm 10) + (30.85 \pm 10) \sin^2 \theta, \quad (4.27)$$

$$b = (-58.08 \pm 11) - (12.43 \pm 10) \sin^2 \theta, \quad (4.28)$$

$$c = (10.28 \pm 3) + (11.96 \pm 7) \sin^2 \theta. \quad (4.29)$$

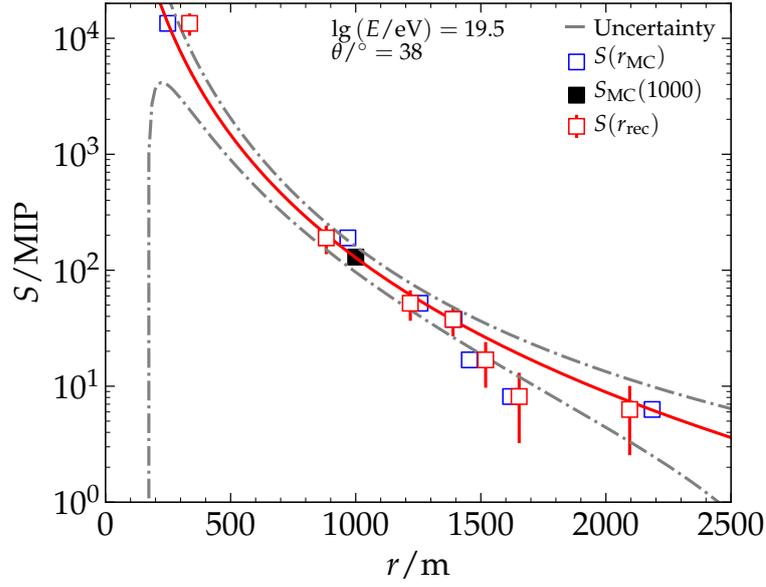


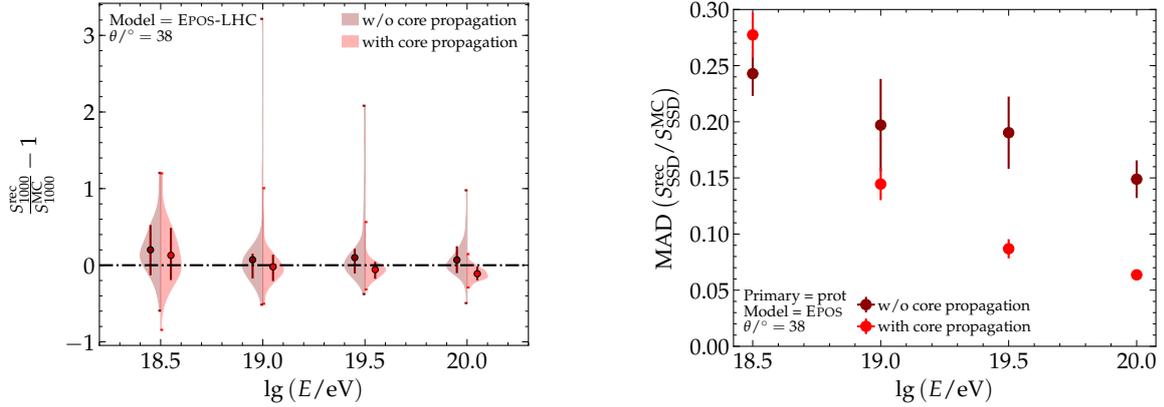
Figure 4.9.: Refitted event from Fig. 4.7, where uncertainties in the reconstructed core have been propagated. The blue and red markers respectively denote the MC and reconstructed distances of the stations from the shower axis. Using a dense station ring at 1000 m, the shower size  $S_{MC}(1000)$  is obtained as an average and shown here as a black marker. The gray error bands show the uncertainty of the signal calculated by adding the sampling fluctuations and the propagated core uncertainties in quadrature. The relative increase in the uncertainty for stations close to the shower core reduces their impact on the fit.

Horizontal errors are not treated in the LDF fit in `Offline`. The core uncertainty has to be propagated to the signal in the LDF using Gaussian error propagation. Since the core resolution is independent of the SSD signal, it can be added in quadrature with the current uncertainty model  $\sigma_{SSD}$  of sampling fluctuations as

$$\sigma_{\text{mod}}^2 \Big|_{r > 200 \text{ m}} = \sigma_{SSD}^2 + \left( \sigma_c \frac{\partial S}{\partial r} \right)^2, \quad (4.30)$$

where  $S$  is the current SSD LDF function. This propagation of core uncertainties has been implemented in `Offline` in the `ScintillatorLDFFinder` module and can be switched on if desired via the configuration file<sup>2</sup>. An example reconstruction of the same event as in Fig. 4.7 is shown in Fig. 4.9, where an improvement of the fit is clearly visible for the example used. The propagated core resolution gives rise to significantly higher uncertainties at distances below 1500 m. Very close to the core, the uncertainty rapidly increases when compared to the uncertainty from the sampling fluctuations. This effectively introduces a distance from the shower axis below which SSD measurements no longer meaningfully contribute to the likelihood as a result of the core uncertainty.

<sup>2</sup>Modules/SdReconstruction/ScintillatorLDFFinderKG/ScintillatorLDFFinderKG.xml.in



(a) Comparisons of the distributions of the relative error in the reconstructed shower size  $S_{1000}$  between cases where core uncertainties were and were not propagated. The square dots at the end of the distributions mark the most extreme outliers.

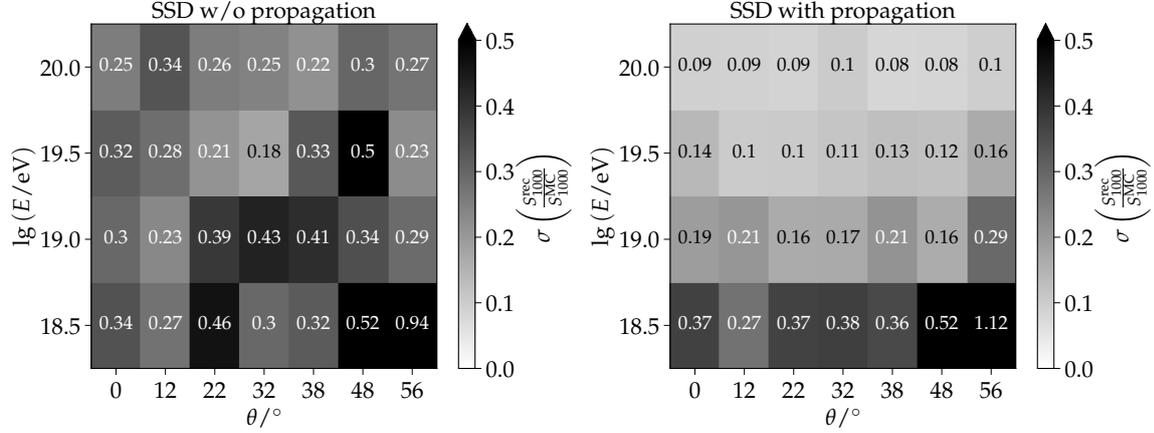
(b) MAD of the relative error in  $S_{1000}$  for the SSD reconstruction with and without the propagation of core uncertainties for a specific zenith angle of  $\theta/^\circ = 38$ .

Figure 4.10.: Assessment of the errors in reconstructed shower sizes as a function of energy for one fixed zenith angle of  $\theta/^\circ = 38$ .

### 4.3.3. Comparing Current and Propagated Uncertainty Model

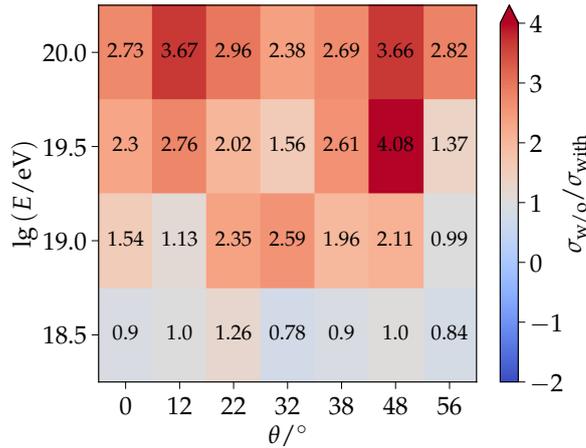
Sample events are reconstructed with and without the propagation of core uncertainties, to gauge the impact of taking the core uncertainty into account during the reconstruction. The following analysis is done using the fixed CORSIKA simulation library from Table 4.1. The reconstructed shower size  $S_{1000}^{\text{rec}}$  is then compared to the Monte Carlo value  $S_{1000}^{\text{MC}}$  to calculate the bias  $B = \langle S_{1000}^{\text{rec}} / S_{1000}^{\text{MC}} - 1 \rangle$ , the standard deviation  $\sigma(S_{1000}^{\text{rec}} / S_{1000}^{\text{MC}})$  and the *mean absolute deviation* (MAD). In Fig. 4.10, the distribution and the average from the results of reconstructions with and without the propagation of core uncertainties are shown for one specific zenith angle of  $\theta/^\circ = 38$ . The MAD is also shown for the same events. The reconstructions in which the core uncertainty is not propagated show very large outliers with some reconstructed shower sizes 4 times larger than the Monte Carlo values. When propagating the core uncertainties, these extreme outliers decrease significantly and the resolution for the bulk of events also significantly improves. At an energy of  $\lg(E/\text{eV}) = 19.5$ , the MAD improves about a factor 2. The energy bin of  $\lg(E/\text{eV}) = 18.5$  is at the lower limit of full efficiency for the SD array, where most events will have only three triggered stations. At this energy, no meaningful improvement can be observed. Reconstructing the events using the current uncertainty model results in a standard deviation for the SSD ranging from 20% up to 50% for energies at or above  $\lg(E/\text{eV}) = 18.5$ , as pictured in Fig. 4.11a. In Fig. 4.11b, the standard deviations of  $S_{1000}^{\text{rec}}$  using the uncertainty model with propagated core uncertainties are shown. Important to note is that the standard deviation for the sample without propagation of core uncertainties is influenced by the presence of significant outliers, which arise when an SSD is exceptionally close to the shower axis. The MAD is calculated as well, since the standard deviation is susceptible to those outliers. The MAD provides a measure of how the resolution improves considering

the bulk of events, as it is less influenced by significant outliers. In Fig. 4.12, the MAD for reconstructions with and without core uncertainty propagation, as well as the ratios between the two cases, are shown. As for the bias, a decrease of the absolute value of the bias may be observed in Fig. 4.13 for reconstructions with the propagated core uncertainties. However, there is still a bias in the reconstructed values, which increases in magnitude with energy. More, detailed plots for iron primaries as well as the different model QGSJET-II.04 is given in the appendix Appendix A.1.



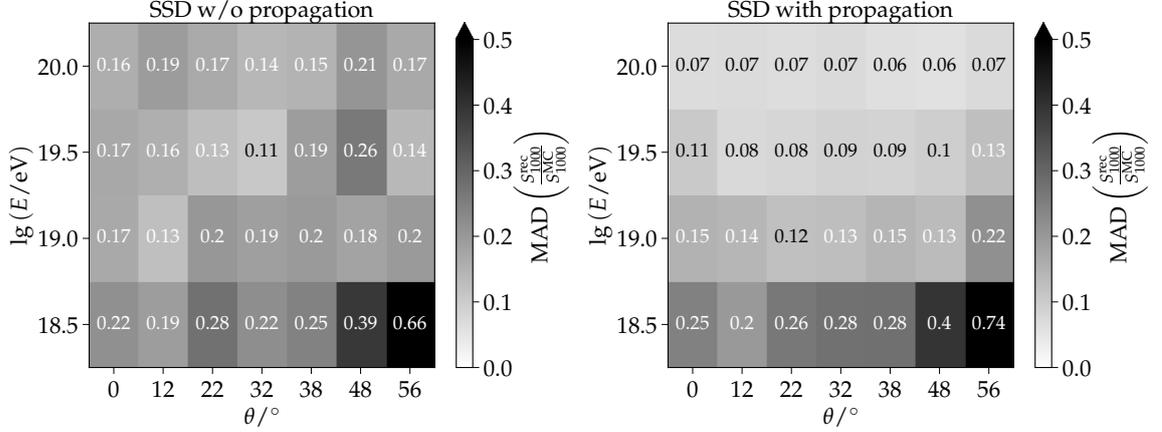
(a) Resolution of the reconstructed SSD shower size  $S_{1000}$  as estimated with the standard deviation of relative errors in the case *without* propagation of core uncertainties.

(b) Resolution of the reconstructed SSD shower size  $S_{1000}$  as estimated with the standard deviation of relative errors in the case *with* propagation of core uncertainties.



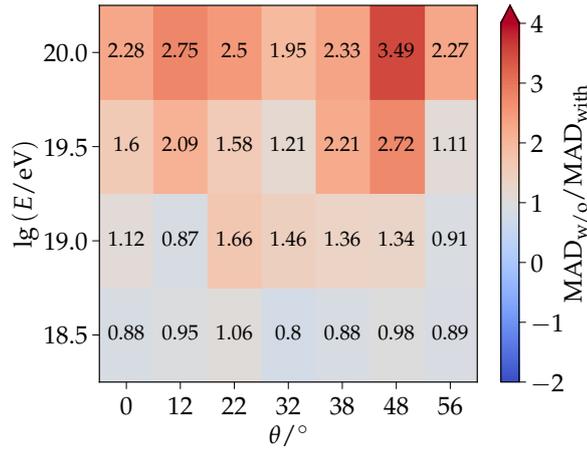
(c) Ratio of the above estimates of the resolution between reconstructions without and with the propagation of core uncertainties.

Figure 4.11.: Comparison of the resolutions (calculated using the standard deviation of relative errors) of the reconstructed SSD shower size with and without propagation of core uncertainties for different zenith angles and energy bins. With the reduction of extreme outliers,  $\sigma$  of the SSD above an energy of  $\lg(E/eV) = 18.5$  is improving up to a factor of around 2 to 3. At the energy of  $\lg(E/eV) = 18.5$ , the SD array is at the lower limit of full efficiency, where the SSD is not expected to perform well.



(a) Resolution of the reconstructed SSD shower size  $S_{1000}$  as estimated with the mean absolute deviation of relative errors in the case *without* propagation of core uncertainties.

(b) Resolution of the reconstructed SSD shower size  $S_{1000}$  as estimated with the mean absolute deviation (MAD) of relative errors in the case *with* propagation of core uncertainties.



(c) Ratio of the above estimates of the resolution using the mean absolute deviation (MAD) between reconstructions without and with the propagation of core uncertainties.

Figure 4.12.: Comparison of the resolutions (calculated using the mean absolute deviation of relative errors) of the reconstructed SSD shower size with and without propagation of core uncertainties for different zenith angles and energy bins. The mean absolute deviation is less influenced by large outliers than the standard deviation and thus serves as a good estimate of the resolution for the bulk of events. The values are smaller than for the standard deviation, but still show a sizable improvement with the propagation of core uncertainties.

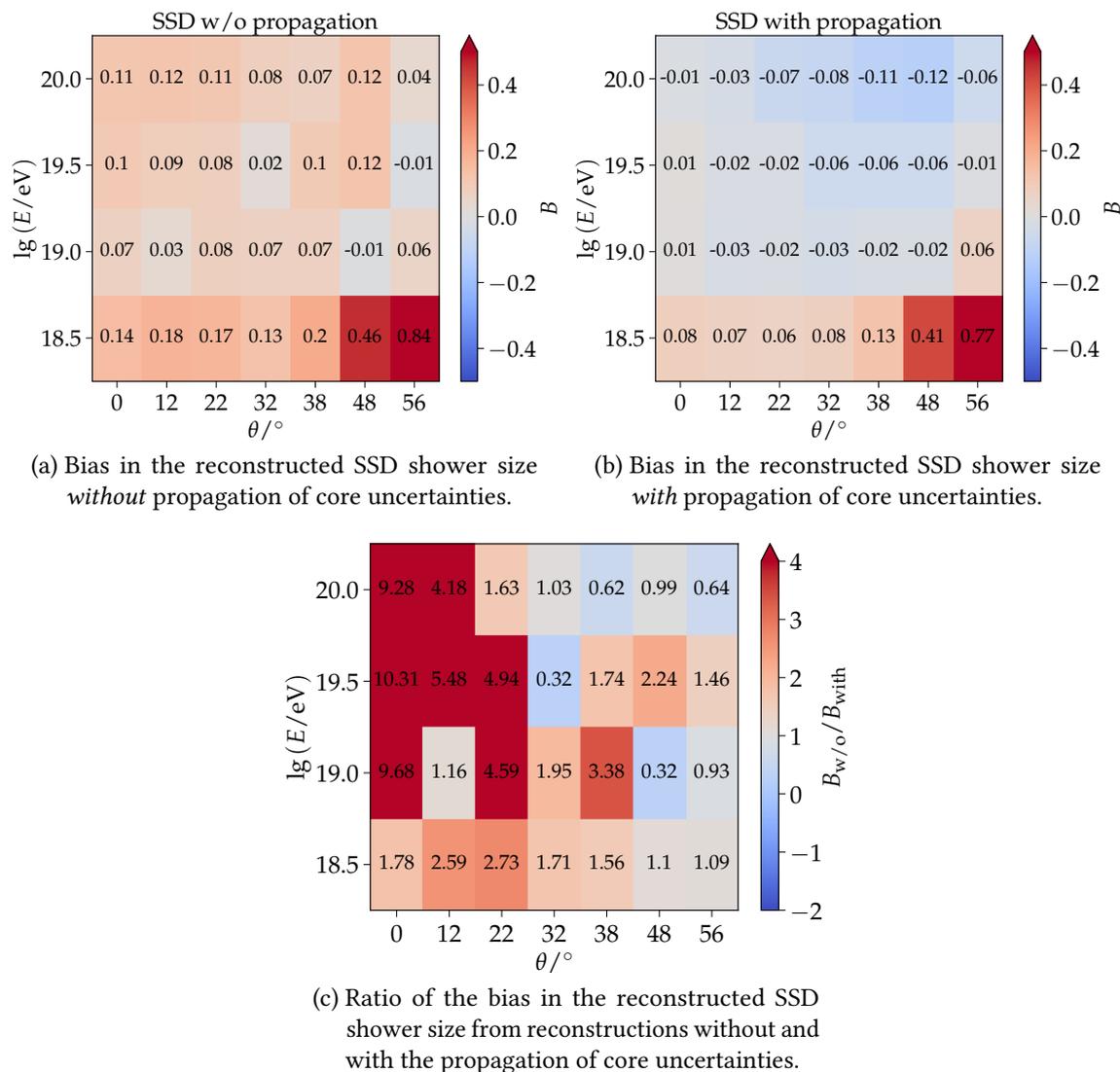


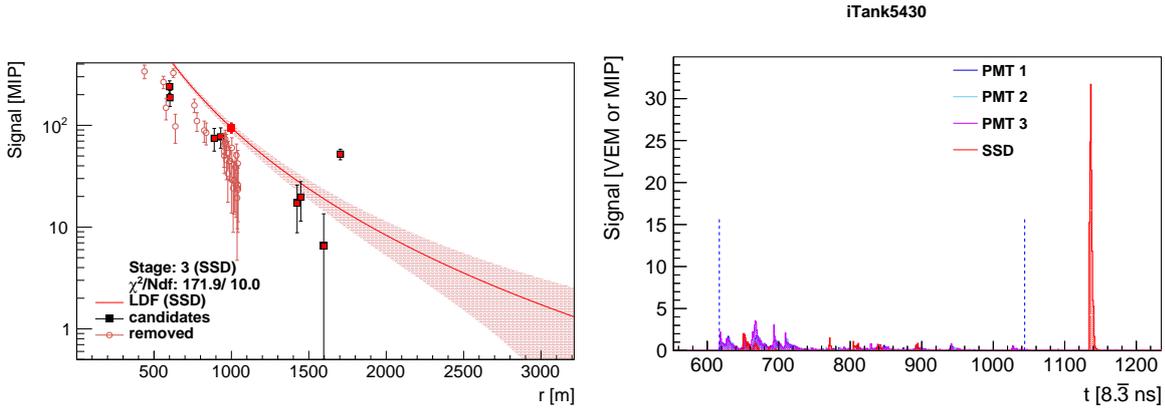
Figure 4.13.: Assessment of the bias in the reconstructed SSD shower size with and without the propagation of core uncertainties. The propagation of the core uncertainties results in a smaller, negative bias.

## 4.4. Quality Cuts

Propagating the core uncertainties from the previous subsection already improved the SSD reconstruction by a significant amount and removed a lot of extreme outliers. However, in a few remaining cases the shower size is still badly reconstructed and possible reasons shall be investigated separately. These failing reconstructions can be split up into two groups. In the first group, the shower size is estimated too large, compared to its Monte Carlo value. In the second group it is estimated too low. While both cases can have various reasons to misreconstruct  $S(1000)$ , one specific reason for each case will be discussed. For the overestimated shower size, single stations with a larger signal than the LDF prediction, can lead to a misreconstruction of the LDF. In the case of an underestimated shower size, stations with no signal are found to be causing the problem.

### 4.4.1. Large Single Station Signals

Propagating the core uncertainties into the LDF fit affected mainly the hottest station. Close to the shower core, the core uncertainties are dominating over the signal uncertainty. While previously the LDF fit was strongly dependent on the hottest station, with the larger uncertainty also stations further away from the core contribute more significantly to the fit. However, this makes the fit also more susceptible to stations that deviate from the predicted LDF. An example LDF reconstruction is shown in Fig. 4.14a. The shower size in this example is reconstructed too high, therefore most of the points are below the predicted LDF. Only a single station with a signal that is about  $5\sigma$  larger than the prediction lies above the LDF. It is apparent that the uncertainty of this particular station is extremely small compared to the other stations. Fig. 4.14b shows the WCD and SSD traces of the station. At the end of the integration window, the SSD trace shows a very large spike, compared to the remaining traces. Although this spike is outside of the range of the WCD integration window - which is marked with the blue dotted lines - it is still inside the SSD integration window, that can be larger than the WCD window. In 1985, John Linsley already observed as well very late signals in the detector traces of the experiment at the Volcano Ranch. These *sub-luminal pulses* were suggested to be related to heavy particles, as protons or neutrons, that originate from the hadronic shower component. They are traveling with a velocity less than the speed of light and thus arrive at a later time than the main portion of the shower [49]. For each of the 3 resimulated events with an observed large late signal in the SSD traces with Offline, a single injected proton could be found in the simulation files. This proton could contribute mainly to the large signal spike. However this should be treated very carefully, since this is only a representation of simulated data and not an observation of real events measured with Auger. The injected protons could also be caused due to a bug during the resampling procedure. Also, the observed protons had a very low energy and were stopped in the WCD. A light yield is estimated proportional to the energy loss in the detector and thus an increased signal, due to those late protons, is visible in the simulations. Whether this reflects observations in the data sets has yet to be determined. It should be noted, that these late traces could not only originate from late, heavy particles but might also be induced by accidental muons. For now, it is not yet known, whether these late SSD traces can be found in the Auger



(a) LDF with a single station that possesses an anomalously large signal, compared to the LDF prediction.

(b) The SSD trace of the station with the large signal shows a signal spike at the very end of the SSD integration window. The WCD integration window is depicted with the blue dotted lines.

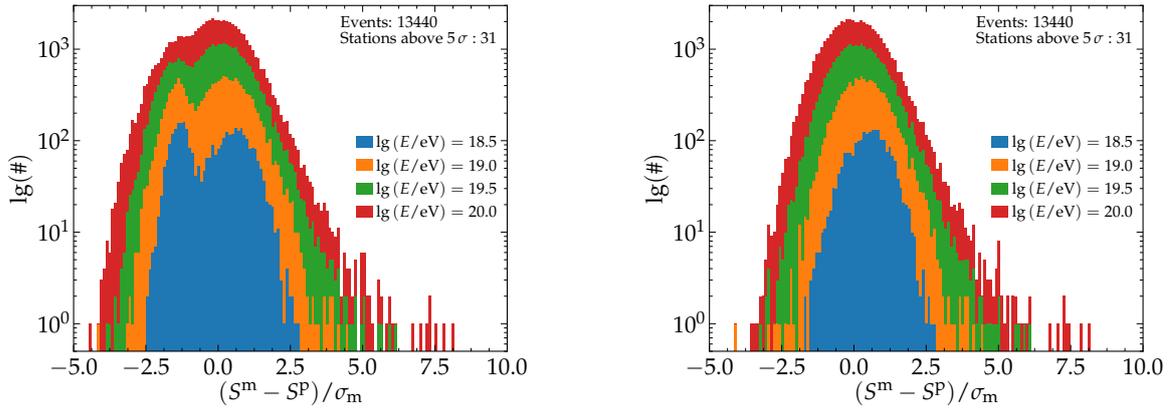
Figure 4.14.: Example LDF and station trace of an event with overestimated shower size. In the LDF it can be seen, that one station with very large signal and a small signal uncertainty causes the LDF to misfit the shower size. A very large late signal in the station trace can be attributed for the large station signal.

data, but it might open up the possibility to study to some extent the hadronic shower component. Nevertheless, some studies on these late arriving protons in simulated data can be done.

As dataset, an extended version of the fixed library of CORSIKA simulations from Section 4.3 was used, including now iron as additional primary particle and QGSJET-II.04 as additional model. With 120 simulated showers in 4 energy and 7 zenith bins, as well as 2 models and primaries, this makes a total of 13440 simulated showers. For these events it is of interest to find those, that have a station with a large deviation from the LDF prediction. As a measurement for this distance the predicted signal  $S^p$  from the LDF is subtracted from the measured signal  $S^m$  and divided by the signal uncertainty of the measured signal  $\sigma_m$

$$\frac{S^m - S^p}{\sigma_m}.$$

This variable gives the deviation of the measured signal from the LDF in terms of the uncertainty  $\sigma$ . From the total of 13440 simulated showers and 94948 stations in these events, 31 stations were found to deviate more than  $5\sigma$  from the LDF prediction. A histogram of all the stations, binned in energy can be seen in Fig. 4.15a. If the LDF reconstruction would be correct, one would expect a peak of the histogram around  $0\sigma$ . However a second peak around  $-1\sigma$  in all energy bins is visible. This second peak originates from stations, that have a signal of 0 MIP and are still included in the reconstruction procedure and will be described in Section 4.4.2. For now, a cut, requiring all stations to have at least a signal of 1 MIP, is applied to the dataset and the new histogram can be seen in Fig. 4.15b. From the 31 stations with a deviation above  $5\sigma$  from the LDF, 26 stations originate from events with



(a) Deviation of the predicted station signal from the LDF. The left peak around  $1\sigma$  comes from stations with no signal. These stations will be described in Section 4.4.2.

(b) A signal cut, accepting only stations with a signal above  $1\text{ MIP}$  removes the second peak.

Figure 4.15.: Distribution of the signal deviation of SSD signals from the LDF. Stations with a deviation above  $5\sigma$  from the LDF are selected as candidates with a large, late signal in the trace.

an energy of  $\lg(E/\text{eV}) = 20$  and the remaining 5 from events with  $\lg(E/\text{eV}) = 19.5$ . Two example traces of these stations can be seen in Fig. 4.16 and get split up in two categories. The first category includes late traces with sharp, distinct peaks, as seen in Fig. 4.16a. This peak could originate from either a late, heavy particle or an accidental muon, crossing the SSD. The traces of the second category show very broadly distributed signal peaks, as seen in Fig. 4.16b. A possible reason for this could be, that the late, heavy particle induces a small subshower in the SSD, which would lead to such a broad distribution. However, since the distribution of the signal deviation is rather continuous (see Fig. 4.15), a selection purely on this quantity is not possible. As some first attempt to filter late trace contributions, the integration window of the SSD gets split into two halves. First, the total signal in the complete integration window is calculated. Then a signal of only the first half of the integration window is calculated. Since large, late signals shall be filtered from the data set, it can be assumed, that these contribute mainly to the total SSD signal. If the signal of the first half of the integration window is contributing less than 50% to the total signal, the SSD integration window gets reduced to the WCD integration window, which is by default smaller, or as large as the SSD integration window. As for the example trace in Fig. 4.14b, the peak of the SSD trace lies outside of the WCD integration window (blue dashed lines) and thus the SSD integration window gets set to the smaller WCD window. In some cases however, a large signal spike might still be included in the newly set integration window. Again, the reduced integration window gets split into two halves and the signal of the first half, gets compared to the signal of the reduced integration window. If the signal of the first half still contributes less than 30% to the total signal, the station is removed from the fit. From the total of 13440 events and 94948 stations, 7.4% of the stations had signal of the first half contributing less than 50% to the total signal and their integration window got reduced to the WCD integration window. 1.7%

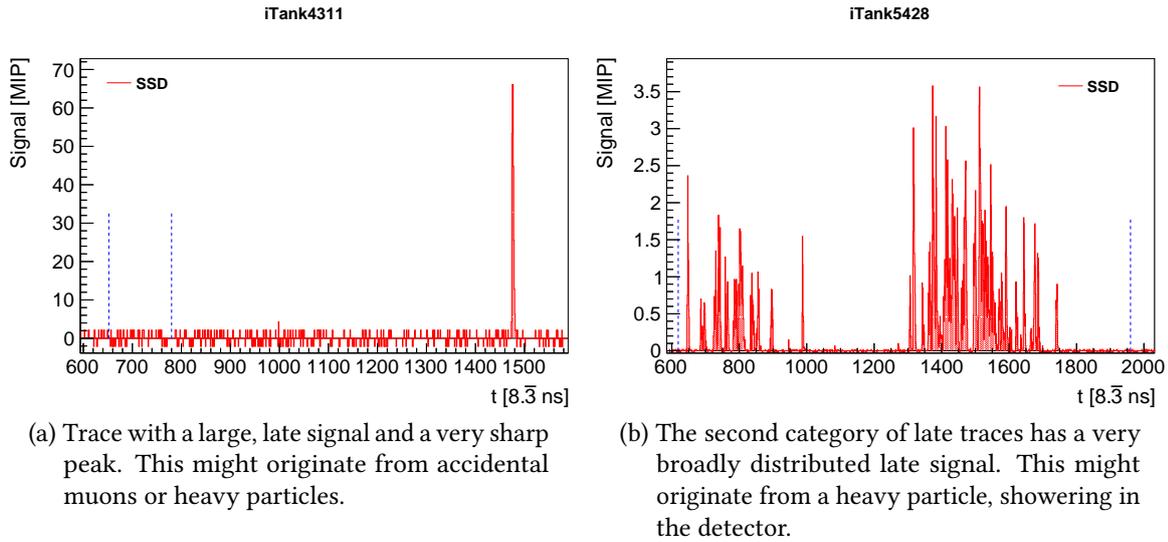
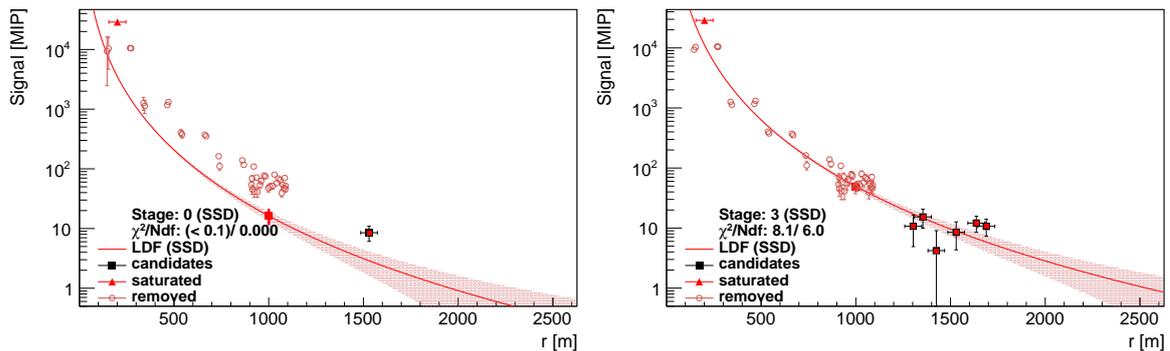


Figure 4.16.: From the 31 stations that deviate more than  $5\sigma$  from the LDF, two different types of late traces can be observed.

of the stations were removed from the fit, due to the second criterium, rejecting stations that have a signal in the first half, contributing less than 30% to the total signal. This treatment however, is only some very first approach to filter out late signals and should not be considered as a reliable method for finding late signals. The filtered stations are not necessarily an indicator for a late arriving, heavy particle. A more detailed analysis on trace measurements from simulations and data would be needed to gain more insight, if indeed late arriving, heavy particles could be observed in the data.

#### 4.4.2. Zero Signal Stations

The second group of failing reconstructions contains underestimated shower size predictions. An example of an underestimated LDF can be seen in Fig. 4.17a. Although only one station seems to be present, the LDF is not fitted to this station. However, taking a look at the list of stations that are used during the fit procedure, more than one station seem to be present. The signal of these stations is in fact 0 MIP and thus they are not visible in the LDF plot. As mentioned earlier, in Fig. 4.15a these zero signal stations are visible as a second peak, when calculating the signal deviation from the LDF prediction for each individual station. Before the station signals of the SSD are fitted to the lateral distribution in Offline, the signals are calculated in the SDCALIBRATOR module by integrating over the trace, measured by the detector. If a signal is found in the station trace, the start and stop times of the integration window are set. Due to a bug, it might occur that no signal in the station is found - although a signal is present in the trace - and the integration window gets set to zero. Integrating over an integration window of the size of only one bin results in a zero signal estimation for that station, which is later used during the fit procedure of the LDF. By forcing to use at a minimum the WCD integration window, even if no signal seems to be found, this can be bypassed and a signal will be calculated from the SSD trace.



(a) In this LDF fit, stations with zero signal are not visible but still used. This leads to an underestimated shower size.

(b) By enforcing to use the WCD start and stop times as a minimal integration window, the stations are properly used in the fit and the reconstructed shower size estimation improves.

Figure 4.17.: Due to a bug in Offline, some stations are reported with zero signal and no integration window is set for the calculation of the signal, although a signal in the trace of these stations is present.

Fig. 4.17b shows the same LDF as Fig. 4.17a, but with the WCD start and stop times as minimum forced integration window for the SSDs. The previously abundant zero signal stations are now included with an appropriately calculated signal and the shower size estimation improved.



## 5. Detector Responses to the Muonic Shower Component

Understanding the different responses of the two detector systems to the muonic and electromagnetic shower component is of interest. The contribution of these shower components to the measured signal varies with different primaries. In the following sections, the results of an analysis, based on simulations for detector responses to the muonic shower component, are described. In Section 5.1, an ideal signal model of the detector response to the muonic component is described and verified in Section 5.2 with a simple simulation method. Section 5.3 and Section 5.4 focus on two reasons why the observed signal responses deviate from the predictions. In the last Section of this Chapter, the detector response to the electromagnetic component is briefly discussed.

### 5.1. Ideal Signal Model

Muons are more likely to traverse the WCD and SSD completely, compared to electrons. Thus it is more convenient to find a simplified description of their signal in both detectors. For the signal of the muonic component in the WCD and SSD, the following variables are used

$$S_{\text{WCD}}^{\mu^\pm} \quad \text{and} \quad S_{\text{SSD}}^{\mu^\pm}.$$

Each signal can be split up and parameterized by multiple components. A particle has to hit the detector to create a signal in the WCD or SSD. Every hit of an individual particle deposits some of its energy by the excitation or ionization of the atoms, according to the *Bethe-Bloch equation*. Thus, the total signal  $S^{\mu^\pm}$  can be split up as

$$S_{\text{SSD}}^{\mu^\pm} = N_{\text{SSD}}(\theta) \tilde{S}_{\text{SSD}}(\theta), \quad (5.1)$$

$$S_{\text{WCD}}^{\mu^\pm} = N_{\text{WCD}}(\theta) \tilde{S}_{\text{WCD}}(\theta), \quad (5.2)$$

with  $N_i$  giving the number of particles that hit the detector and  $\tilde{S}_i$  as the fraction of the total signal per particle deposited in the detector. The index  $i$  marks the respective detector SSD or WCD.  $N_i$  can be further split up as

$$N_i(\theta) = n(\theta) A_i(\theta). \quad (5.3)$$

The number of particle hits is the product of a general particle density  $n$ , in a certain area on the ground and the projected area  $A_i$  of the detector shape. Furthermore, the signal

per particle  $\tilde{S}_i$  is directly proportional to the number of photoelectrons created along the tracklength  $x$  of a particle traversing the detector and can be expressed as:

$$\tilde{S}_i(\theta) = k x(\theta), \quad (5.4)$$

with the arbitrary scaling factor  $k$ . It should be noted however, that this equation holds only, if the kinetic energy  $T$  of the particle is large enough, to create cherenkov photons along the full track. With this further splitting, Eq. (5.1) and Eq. (5.2) can then be rewritten in the following way

$$S_{\text{SSD}}^{\mu^\pm} = n(\theta) A_{\text{SSD}}(\theta) k_{\text{SSD}} x_{\text{SSD}}(\theta), \quad (5.5)$$

$$S_{\text{WCD}}^{\mu^\pm} = n(\theta) A_{\text{WCD}}(\theta) k_{\text{WCD}} x_{\text{WCD}}(\theta). \quad (5.6)$$

Assuming, that the detectors will be hit by a constant particle density, the total signal of a detector should not change with it's orientation relevant to the particle flux, since the signal is in first order proportional to the detector volume, which does not change.

### 5.1.1. Projected Detector Area

The projected detector area  $A$  for the SSD and WCD in one certain direction can be calculated in the following way.

The **SSD** can be in first order assumed as two flat rectangles with a length  $l_{\text{SSD}}$  of 1.6 m and a width  $w_{\text{SSD}}$  of 1.2 m, making a total area of

$$A_{\text{SSD}}^{\text{top}} = 2 l_{\text{SSD}} w_{\text{SSD}} = 3.84 \text{ m}^2. \quad (5.7)$$

Projecting this area into the shower plane, the size of it decreases with increasing shower zenith angle  $\theta_{\text{Shower}}$ . The sides of the SSD increase the area with growing  $\theta_{\text{Shower}}$ , because the SSD is not completely flat but has a finite thickness  $t_{\text{SSD}}$  of 0.01 m

$$A_{\text{SSD}}^{\text{side}} = 2 l_{\text{SSD}} t_{\text{SSD}} = 0.032 \text{ m}^2. \quad (5.8)$$

The fully projected area of the SSD in one direction is therefore:

$$A_{\text{SSD}} = A_{\text{SSD}}^{\text{top}} \cos \theta_{\text{Shower}} + A_{\text{SSD}}^{\text{side}} \sin \theta_{\text{Shower}}. \quad (5.9)$$

The top area of the **WCD** can be given with the equation for a circle with radius  $r = 1.8$  m

$$A_{\text{WCD}}^{\text{top}} = \pi r^2 \approx 10.18 \text{ m}^2. \quad (5.10)$$

The side area can be written as rectangle with height  $h = 1.2$  m and width  $w = 3.6$  m:

$$A_{\text{WCD}}^{\text{side}} = h_{\text{WCD}} w_{\text{SSD}} = 4.32 \text{ m}^2. \quad (5.11)$$

As for the SSD, the WCD top area projected into the shower plane decreases with shower zenith angle and the side area increases

$$A_{\text{WCD}} = A_{\text{WCD}}^{\text{top}} \cos \theta_{\text{Shower}} + A_{\text{WCD}}^{\text{side}} \sin \theta_{\text{Shower}}. \quad (5.12)$$

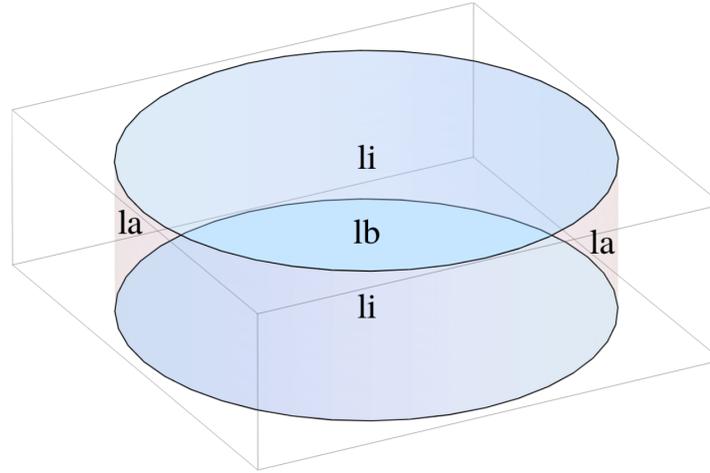


Figure 5.1.: Projection of a WCD at a zenith angle of  $50^\circ$ . Three cases of a particle crossing the detector can be seen: Full traversal of the detector (lb) or corner clipping (li and la). Image taken from [50].

### 5.1.2. Average Tracklength

A particle passing through the **SSD** exactly vertically will traverse 0.01 m of detector material. With increasing zenith angle of the particle, the tracklength  $x$  will increase:

$$x(\theta) = t_{\text{SSD}} \sec \theta. \quad (5.13)$$

There is also the possibility of the so called *corner clipping*, which means, that the particles will enter the detector at the side and not from the top and thus have a reduced tracklength. Since the SSDs thickness is only 0.01 m, this case will be neglected.

If a particle passes through the **WCD** exactly vertically, it will traverse 1.2 m of water. Since the WCD has a rather huge thickness of 1.2 m compared to the SSD, the corner clipping gives significant contributions to the average tracklength of a particle. All the possible tracks of muons through the detector from Fig. 5.1 can be categorized into the following cases

- the particle enters the tank at the top and leaves at the bottom (lb).
- the particle enters the tank at the top and leaves at the side (li). Equivalent case: the particle enters at the side and leaves at the bottom.
- the particle enters and leaves the tank at the side (la).

Each case possesses different tracklengths, dependent on the zenith angle of the ingoing particle. The average tracklength can be calculated according to [50]

$$x(\theta) = \frac{2rh}{|\cos \theta| + \frac{4h}{\pi} \sin \theta}, \quad (5.14)$$

with  $h = 1/3$ .

## 5.2. Idealized Detector Response with a Constant Injection Disk

With the necessary equations for detector dependent tracklength  $x$  and projected area  $A$ , their predictions can be tested with an idealized detector response simulation. Using the Offline simulation framework, an injection procedure for muons to a WCD and SSD setup is performed and explained in detail in this section.

The idealized simulation is done with the following assumptions:

- the particles hit the detector with equal zenith angles.
- the muons have one discrete energy of 1 GeV.

At first, an injection disc above a simulated detector station, consisting of a WCD with an SSD mounted on top, is created. The height of the disc is chosen large enough, to rotate it freely around a sphere, containing the detector station. Therefore, the injection disc never intersects with the station. The radius of the disc is chosen large enough, to always contain the full projected area of the detector station when rotating it along the sphere. On the disc, muons are placed uniformly distributed with a constant particle density  $n$ . These muons are then injected along the given shower direction of zenith angle  $\theta_{\text{shower}}$  to the WCD and SSD. One simulation run will give an output of PMT charges, that can then be converted to a detector signal and the information of each individual particle if it hit the detector or not. Each run is then repeated multiple times for different zenith angles. A visualization of the injection disc and the detector station, as well as a projected area image of the WCD and SSD is shown in Fig. 5.2. The projected area is acquired by coloring each original position of the injected particles on the disc if they hit the SSD or WCD. With the information of particle hits, the projected detector area can be acquired

$$A_i = \frac{N_i}{N_{\text{tot}}} A_{\text{injection disk}}, \quad (5.15)$$

with  $N_i$  as the number of particles that hit detector  $i$ , the total number of injected particles  $N_{\text{tot}}$  and the area of the injection disk  $A_{\text{injection disk}}$ . The resulting signal for each detector can be divided by the particle hits to receive the signal per particle  $\tilde{S}$ . A comparison of the idealized detector response simulation and the values, retrieved from Eqs. (5.9) and (5.12), Eq. (5.13) and Eq. (5.14) is shown in Fig. 5.3. The tracklength Eq. (5.13) and Eq. (5.14) are scaled according to Eq. (5.4) with a factor  $k_i$  to match the signal per particle  $\tilde{S}$ . For the SSD the factor takes the value  $k = 100$  and for the WCD  $k = 0.78$ . While  $\tilde{S}$  for the SSD follows a simple linear  $\sec \theta$  function, the WCD shape shows some distinct features, that can be explained with the different track lengths. From the three possible cases of the particle entering and exiting the tank, the corner clipping effect decreases the average tracklength at first with rising zenith angle. With a further increase in  $\sec \theta$ , the tracklength of the corner clipping particles increases, giving a rise of the total average tracklength.

### Comparison with Simulated Shower Data

With the data of the simulation library, used in Chapter 4, a comparison can be done. Muonic signal and detector particle hits can be read out from the Offline ADSTs and

compared to the ideal detector response simulation. A direct comparison of the detector hits should be treated carefully, since the number of particles that hit the detector  $N$  is dependent on detector area  $A$  and particle density  $n$ . The area  $A$  is independent from shower simulation and idealized detector response simulation. The underlying particle density  $n$  though, is chosen as a constant for the idealized simulation, while it is unknown for the shower simulation. Therefore, the shape of the particle hits can deviate as seen in Fig. 5.4a. A better indicator would be a comparison of the ratio of particle hits from both simulations, since this is a particle density independent value (see Fig. 5.4b). The ratio of both, idealized and shower simulations shows a decreasing behavior with  $\sec \theta$ . The signal per particle  $\tilde{S}$  however can be directly compared. Fig. 5.5a and Fig. 5.5b show the direct comparison and the deviations of the signal per particle between the idealized and the shower simulation. The signal per particle of the shower simulations is shifted to higher values for the SSD. The steepness of the slope is declining for higher zenith angles, which can be seen in Fig. 5.5b with the decreasing values for the SSD. The measured signal per particle is therefore higher than one would expect it. This vertical offset can be attributed due to miscalibration of the SSD signal, since it would result in just a vertical shift of the measured signal. The declining steepness however, can not be explained with a miscalibration. A possible effect could be due to the angular distribution of arriving shower particles, which will be analyzed in more detail in the next section. The signal per particle for the WCD shows as well deviations in the shape of the function. While in the idealized simulation, the signal decreases at first for rising zenith angles, the shower simulation data follows a linear function in  $\sec \theta$ . An analysis of the divergence from the shape of  $\tilde{S}$  for the WCD will be performed in the following sections.

### 5.3. Angular Distributions of Shower Particles

In a real event, as well as for the shower simulations in Offline, the angular distribution of the shower particles is not exactly equal to the zenith angle of the shower axis  $\theta_{\text{Shower}}$ . With the particles arriving at different zenith angles in the detector, their tracklength and thus the signal deposited is varying as well. The zenith angle distribution of the individual particles is formed by geometrical and physical effects. The muonic component is mainly created through charged kaon and pion decay. In an EAS with a primary energy of  $\lg(E/\text{eV}) = 20$ , the pions and kaons have a high kinetic energy and the resulting muons are boosted extremely along the direction of the shower axis, making a narrow distribution of zenith angles for individual particles as a physical effect. As geometrical effect the position of the detector stations as well as the general shower geometry can be attributed. A visualization of the shower geometry along with an example dense ring of stations is given in Fig. 5.6. Particles of an event that hit stations in the downstream region, arrive at a higher zenith angle than the shower zenith. Also these particles have to pass through a higher amount of atmosphere. Particles arriving at stations in the upstream region will hit the detectors at a lower zenith angle and pass through less atmosphere. Fig. 5.7a shows the direct comparison of the individual zenith angles of the particles to the zenith angle of the shower axis. The distributions were received by calculating for each individual particle the arrival angle at the detector with respect to the ground normal. To avoid unwanted

geometric features of the dense station ring, only the stations at  $90^\circ$  and  $270^\circ$  were used, since particles arriving at those stations have crossed the same amount of atmosphere. At low shower zenith angles, the particle angles deviate on average around  $20^\circ$  for the SSD and WCD for the muonic component. The average deviation gets smaller with increasing shower angle. The shower has to pass through more atmosphere, so that particles with too low energy will not reach the detector stations anymore. Therefore, more boosted particles will remain, that deviate less from the original shower axis. At large shower zenith angles, the averages of the angular distributions shift to the values of the shower angle. For comparison, the individual arrival angles and their average of the electrons and positrons from the electromagnetic shower component is shown in Fig. 5.7b. The distribution of the individual angles is wider than for the muons. This can be attributed to the creation of low energetic electrons and positrons through pair production and the general lower energy of the particles compared to the muonic component. The relatively high average deviation of up to around  $20^\circ$  at low shower angles for muons is not only influencing the average tracklength and signal of the particles deposited in the detector. The projected area of the detectors is smeared out as well. The consequences of the angular distribution of the electromagnetic component will be discussed in Section 5.5.

Using the tracklength Eqs. (5.13) and (5.14), a modified prediction of  $\tilde{S}$  can be acquired by using the individual particle zenith angles. The different resulting tracklengths for each particle are then scaled with the factor  $k_i$  to get  $\tilde{S}$  for the individual particles. Averaging over all individual  $\tilde{S}$  of a certain shower zenith angle gives the signal per particle from the angular prediction, shown in Fig. 5.8a. For a better comparison of differences between the angular prediction and the shower data, the deviation between both is shown in Fig. 5.8b. By calculating the fraction of  $\tilde{S}$  from angular prediction and shower simulation data, the differences in the shape can be seen more easily. If the shape is similar, a constant behavior of the fraction is expected for different shower zenith angles. For the SSD, the recalculated signal per particle from the angular distributions is following the same behavior as the data from shower simulations. There is still a vertical offset, but this can be attributed due to miscalibration of the detector signal. However, the slope of the shower simulation data is still deviating from the recalculated  $\tilde{S}$  of the WCD and needs to be explained with different effects.

### 5.4. Muon Energy Spectrum

So far the deviation of the observed signal per particle in the shower simulation data from the idealized simulation could be explained for the SSD with the angular distribution of arriving particles. For the WCD this distribution can not describe the difference sufficiently. In the idealized model, a discrete, sufficiently high enough, muon energy is used to retrieve  $\tilde{S}$ . With this energy, the muon will traverse the tank completely and produce cherenkov photons along its full track. For real events or the shower simulation however, the muons arriving at the detector station follow an energy spectrum and not a discrete energy. If the energy of the muon is too low, it will not produce any cherenkov photons. To verify if this is the case for the shower simulation, a modification of the GEANT4 module of Offline is applied.

When traversing the WCD or SSD, the particles produce cherenkov or scintillation photons, that are detected by the PMTs. Each photon is starting an avalanche of secondary photo electrons in the PMT to form an electric signal. The number of photo electrons is directly proportional to the measured signal and thus can be chosen as an estimator for  $S$  and  $\tilde{S}$ . Also, for each particle its individual tracklength through the detector is determined and saved. The ratio of number of photo electrons  $N_{PE}$  produced per tracklength  $x$  is visualized in Fig. 5.9a for the WCD and SSD. For the SSD, a large number of photo electrons is produced, that is proportional to the tracklength of the crossing particle. However, the WCD does not show the same behavior. The number of photo electrons produced in the WCD increases linearly with increasing shower zenith angle, instead of being constant. The kinetic energy of the muon must be above the cherenkov threshold, to produce cherenkov light in the WCD. The velocity of the particle needs to be larger than the phase velocity of light in the medium with refraction index  $n$ , to produce cherenkov radiation

$$\beta > \frac{1}{n}. \quad (5.16)$$

This can be turned into a condition of the minimum particle energy to produce cherenkov radiation with the help of relativistic kinematics

$$E = \frac{nm}{\sqrt{n^2 - 1}}. \quad (5.17)$$

With the muon mass of  $m = 105$  eV and a refraction index of  $n = 1.3$  the equation gives an energy of approximately 159 MeV. The kinetic energy of a muon to produce cherenkov radiation is therefore

$$E_{ch} \approx 54 \text{ MeV}. \quad (5.18)$$

If the energy of the muon drops below this threshold while crossing the tank, no more cherenkov photons are produced (see Fig. 5.9b). When crossing the tank, the muons also lose energy according to the Bethe-Bloch formula. In the WCD, the muons lose approximately 2 MeV per centimeter, which gives for a vertical crossing muon

$$E_{BB} \approx 2 \text{ MeV/cm} \cdot 120 \text{ cm} = 240 \text{ MeV}. \quad (5.19)$$

A vertical muon needs a minimum threshold energy of approximately

$$E_{th} = E_{ch} + E_{bb} \approx 294 \text{ MeV}, \quad (5.20)$$

to produce photo electrons along its full tracklength through the detector. Between the cherenkov threshold of  $\lg(E/\text{eV}) \approx 7.7$  and the energy threshold of  $\lg(E/\text{eV}) \approx 8.5$  the muons will produce cherenkov photons only partially along their tracklength, in the case of traversing the tank vertically. Since the average tracklength of the muons is changing as a function of zenith angle, the energy threshold will increase as well.

In Fig. 5.10 the energy distribution of the muons that hit the detector, as well as the cherenkov and energy threshold, are visualized. For muons, the energy distribution is

changing with  $\sec \theta$  of the shower. With increasing zenith angle, the shower has to pass more atmosphere and low energetic particles are more likely to not make it to the ground. Therefore, the energy spectrum of the muons is shifting to higher energies for an increasing shower zenith angle. In the case of a shower zenith angle of  $0^\circ$ , 31% of the muons on the ground are producing photons only along a fraction of their tracklength. At  $56^\circ$ , the distribution is shifted upwards and the fraction of muons only partially producing photons sinks to 9%. With 31% of the muons producing a reduced number of photons as expected at  $0^\circ$ , the number of photo electrons per tracklength is reduced as well. In Fig. 5.11a, the number of photo electrons produced per tracklength from Fig. 5.9a is replotted again with an energy cut at the threshold energy of  $\lg(E/eV) = 8.5$ . With the threshold cut, the overall average number of photo electrons produced per tracklength is not only increasing for the WCD, but also shifting to a more constant trend. For large shower zenith angles, there is still a slight increase visible in the number of photo electrons produced. The threshold cut is chosen to match the minimum energy for a full tank traversal for vertical muons. Muons, that arrive at the tank at a different zenith angle can have a larger tracklength and can not traverse the tank completely, resulting in a still slightly reduced number of photo electrons per tracklength. Fig. 5.11 shows the deviation of the signal per particle, derived from the number of photo electrons created and cut at the energy threshold of 8.5, from the tracklength of the particle through the detector. The WCD and SSD show now both a constant proportional behavior between signal per particle and tracklength. The features in the signal per particle distributions for the SSD and WCD can be explained with the combination from the effects of the angular distribution of arriving particles and the muon energy spectrum. With those effects changing the shape of  $\tilde{S}$  the ratio of SSD to WCD signal is not a constant as one would expect, but decreasing with  $\sec \theta$  of the shower.

### 5.5. A Brief Look at the Electromagnetic Component

As for the muonic component, an analysis for the electromagnetic component can be performed. Here, only a few core features, as the signal per particle and signal per tracklength, are discussed.

#### Energy Loss and Penetration Depth of Electrons and Positrons

The electromagnetic shower component can be split up into two components: the electron and positron component, where both particles have similar interactions with matter, and the photon component. The electromagnetic component consists mainly out of photons and only to a small fraction out of electrons and positrons. On the one hand, most of the photons do not interact with the detector material in the WCD and SSD and traverse both detectors without depositing any signal. Electrons and positrons, on the other hand, will deposit some amount of their energy in the SSD and all of their remaining energy in the WCD. When traversing matter, electrons and positrons lose energy due to different physical effects, depending on their energy. The *Berger-Seltzer formula* describes the

energy loss of electrons and positrons due to ionization in the material [51]

$$\frac{dE}{dx} = 2\pi r_e^2 m c^2 n_{el} \frac{1}{\beta^2} \left[ \ln \frac{2(\gamma + 1)}{(I/mc^2)^2} + F^\pm(\tau, \tau_{up}) - \delta \right]. \quad (5.21)$$

It follows a similar form and shape as the Bethe-Bloch formula. At the critical energy of around 78 MeV in water [52], the dominating process of energy loss for electrons and positrons becomes bremsstrahlung. In Fig. 5.12, the individual energy losses of electrons due to ionization and bremsstrahlung, as well as the total energy loss are shown. The values used are extracted from the tabulated values of the *ESTAR* database of the *National Institute of Standards and Technology* [53]. From the energy losses of the particles, a penetration depth for electrons and positrons in the WCD can be estimated. The *continuous slowing down approximation* (CSDA) range, that is an approximation to the average track length of a charged particle in a medium, can be found in the *ESTAR* database as well and is shown in Fig. 5.13a. The energy spectrum of the electrons and positrons is visualized in Fig. 5.13b. It is lower than the muonic spectrum and there is no significant dependency in zenith angle visible. From the CSDA range, around 65% of the particles penetrate the WCD up to 13 cm. Since their energy is too low, almost no electron or positron is able to traverse the full tank vertically. The number of photo electrons produced per theoretical track length should therefore be constant for different zenith angles. Fig. 5.14 shows, that this assumption becomes evident for the WCD.

### Electromagnetic Signal per Particle

With almost no particle of the electromagnetic component crossing the full WCD, but stopping in the first few centimeters of the water, the signal per particle  $\tilde{S}$  should be also constant over  $\sec \theta$  for the WCD. The SSD signal per particle is mostly dependent on  $\sec \theta$ , since the particles have sufficient energy to traverse through the full SSD. The  $\sec \theta$  dependency is altered according to the average distribution of the individual particle arrival angles in the detector. With the wider angular distribution and increased average arrival zenith angle, that can be seen in Fig. 5.7b, the signal per particle increases not as steep with  $\sec \theta$  for the SSD. A zoomed in figure of the individual electromagnetic  $\tilde{S}$  components is pictured in Fig. 5.15a.  $\tilde{S}$  of the electrons and positrons for the SSD is outside of the range of the plot and is not discussed further in this section, but can be seen in the Appendix A.2 along with further graphs regarding particle hits and signal contributions of the individual components. The signal per particle of the WCD for electrons and positrons is extremely small compared to  $\tilde{S}$  of the muonic component, since the particles do not pass the full detector as most of the muons, resulting in less Cherenkov photon production. The zenith angle independence due to the electrons and positrons only penetrating the first few centimeters of the water can be seen. However, a rise in  $\tilde{S}$  is visible for the larger zenith angles. Since the penetration depth is dependent on the energy of the incoming particles, also the number of photo electrons created increases with higher energy. For 20 different showers, the average energy of all individual electrons and positrons is shown in Fig. 5.15b. For a zenith angle of  $56^\circ$  a slight increase of the average energy is visible. This could be a possible second order factor of the increase in the signal per particle for the electrons and positrons in the WCD. A more in depth analysis of the electromagnetic

## *5. Detector Responses to the Muonic Shower Component*

---

signal and the relations of the individual components of electrons, positrons and photons would be of further interest, but exceeds the scope of this work and should be examined in future works.

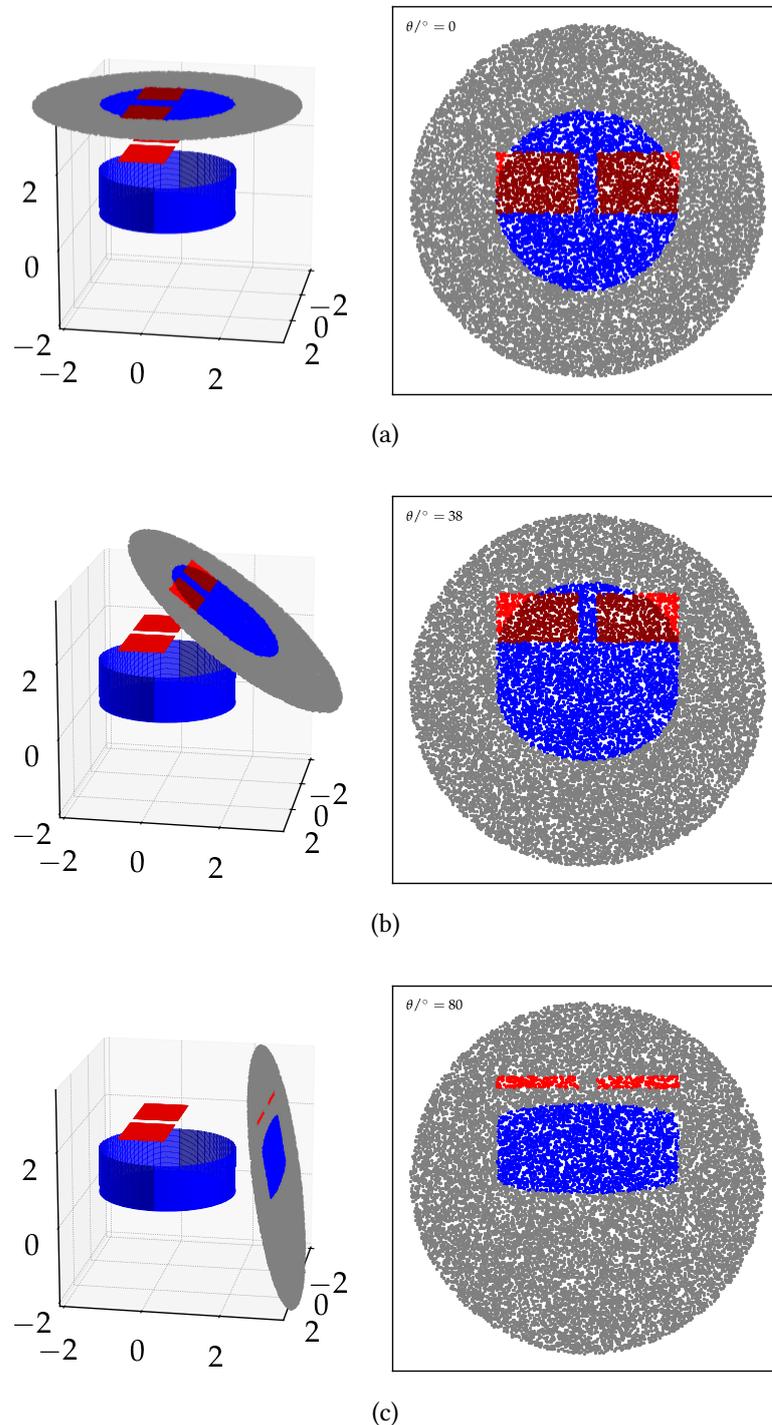


Figure 5.2.: Visualization of the injection disk and the detector station (left) and the resulting image of the projected areas of the station projected onto the disk (right). (a), (b) and (c) show different zenith angles. The grey points are points, that neither intersect with an SSD nor a WCD. Blue and red points are particles that intersect exclusively SSD or WCD and dark red points are combined SSD and WCD hits.

## 5. Detector Responses to the Muonic Shower Component

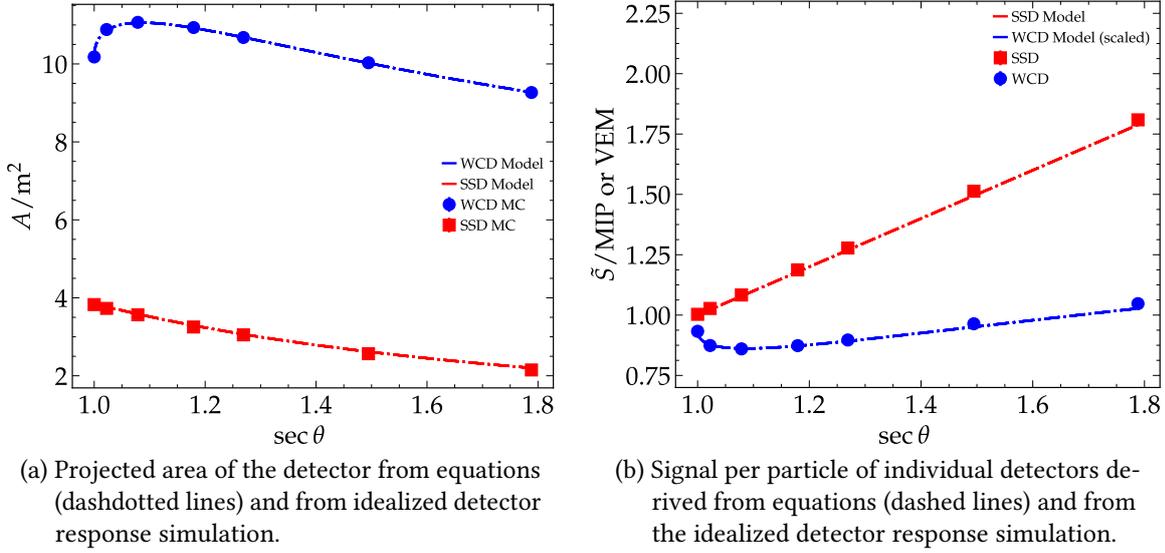


Figure 5.3.: Comparison of equation derived and simulation derived projected area (left) and signal per particle (right). The simulations match the predictions of the equations extremely well.

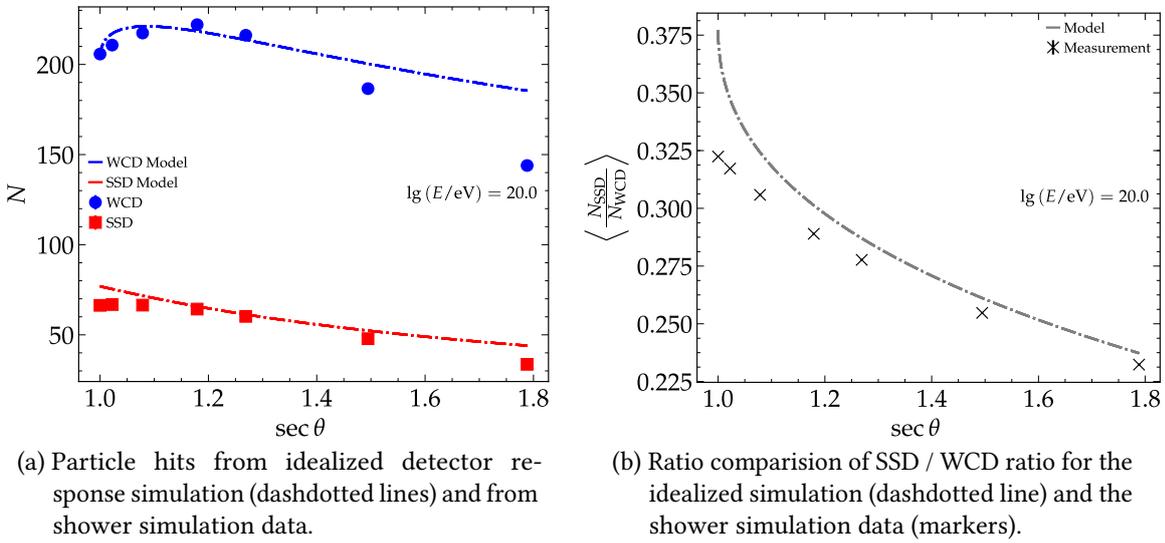


Figure 5.4.: Comparison of number of particle hits (left) and the ratio of SSD / WCD particle hits (right). The shown difference should be treated carefully, since, for particle hits, a direct comparison is not possible. The underlying particle density for the idealized simulation is constant, while the density for the shower simulation is unknown and changes the shape of the distribution. The ratio of SSD to WCD hits is independent of the underlying particle density.

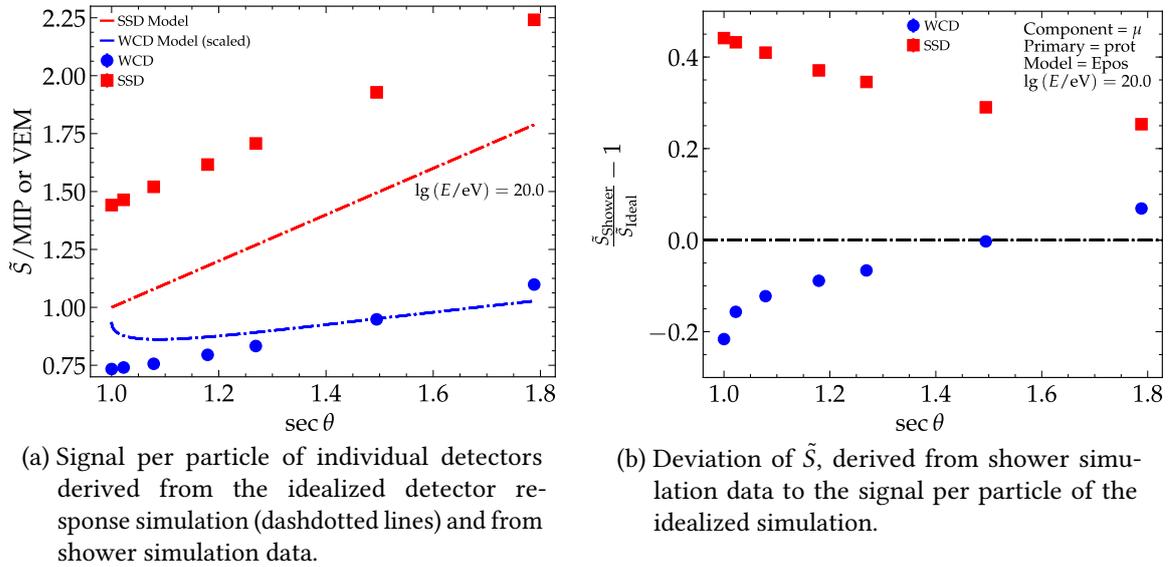


Figure 5.5.: The predicted signal per particle  $\tilde{S}$  from the model differs from the observed shower simulation data. The constantly larger  $\tilde{S}$  that is observed in the shower simulation data can be attributed to some miscalibration of the signal. The right figure shows however also a declining slope of  $\tilde{S}$  over  $\sec \theta$  which originates from other effects.

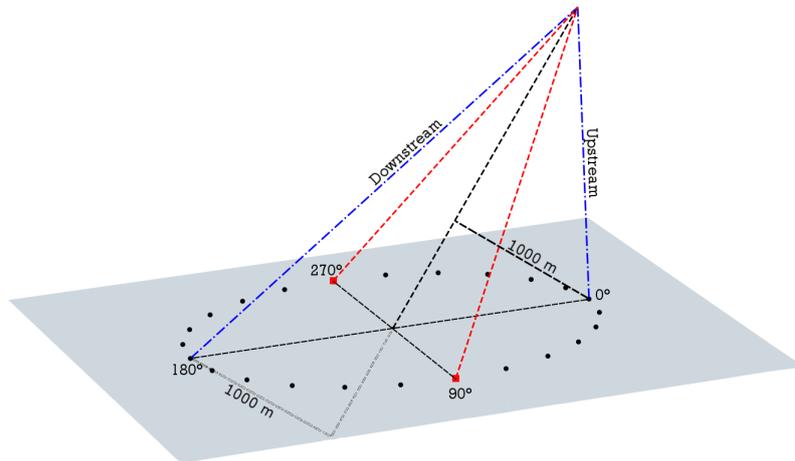


Figure 5.6.: Visualization of a dense station ring in the shower detector plane projected on the ground. All stations of the ring have a distance of 1000 m to the shower axis in this plane. The blue dashed lines represent the upstream and downstream directions. Figure taken from [47].

## 5. Detector Responses to the Muonic Shower Component

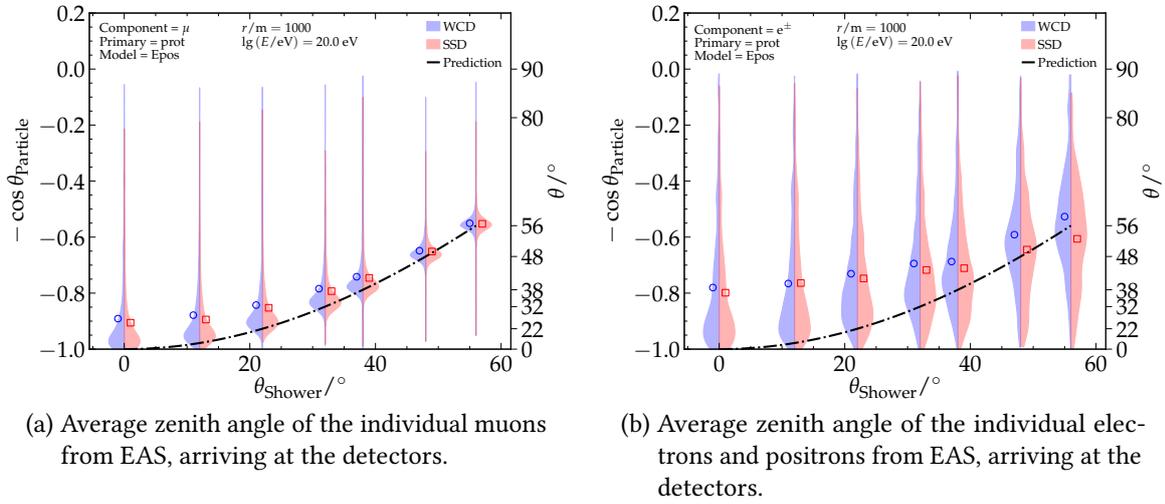


Figure 5.7.: At small shower angles, the particles deviate more from the axis than at large shower angles.

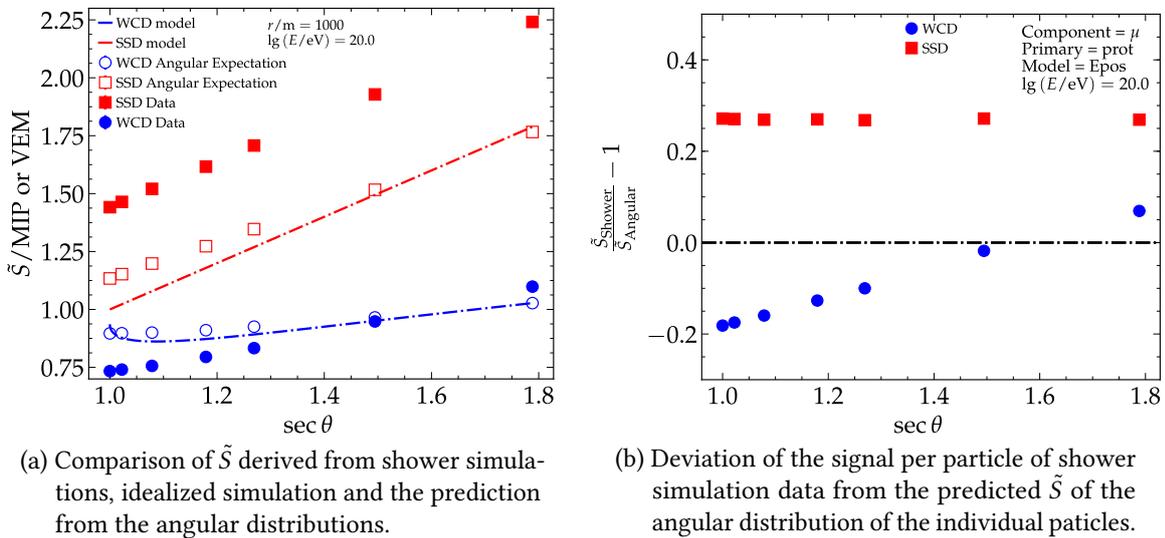


Figure 5.8.: The shape of the signal per particle, derived from angular distributions, for the SSD matches with the shower simulation data. The shape for the WCD signal per particle changed, but is still deviating from the shower simulation data.

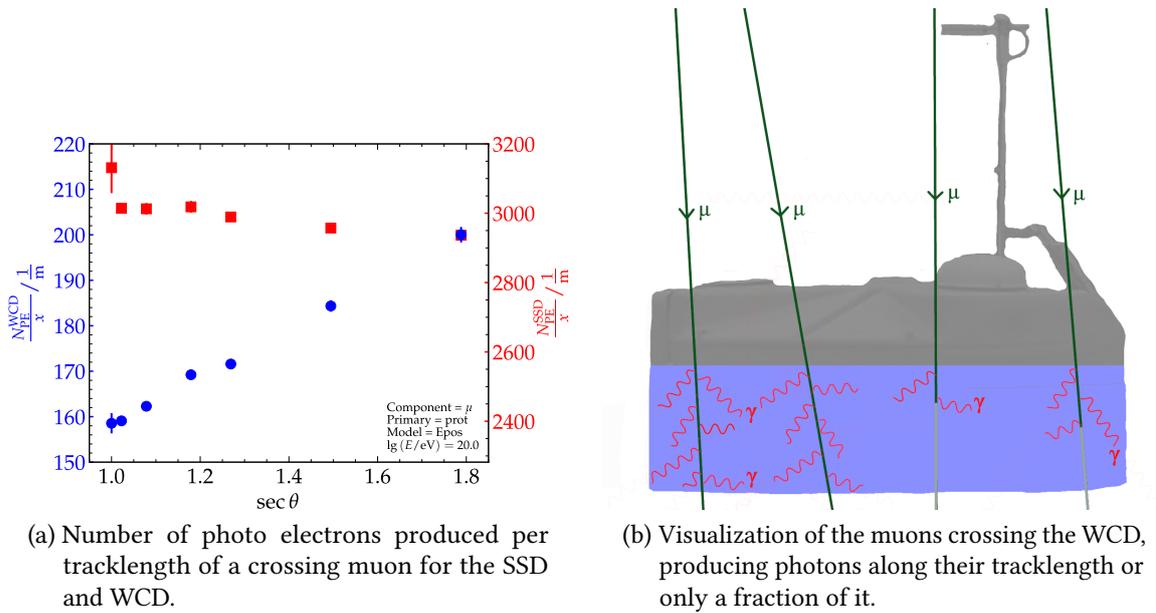


Figure 5.9.: The number of photo electrons produced per tracklength does not change for the SSD for different shower zenith angles. For the WCD however, with increasing zenith angle more photo electrons are produced. The muons that cross the WCD produce cherenkov photons along their track. If the muon energy is below the cherenkov threshold, no photons are produced, reducing the overall signal.

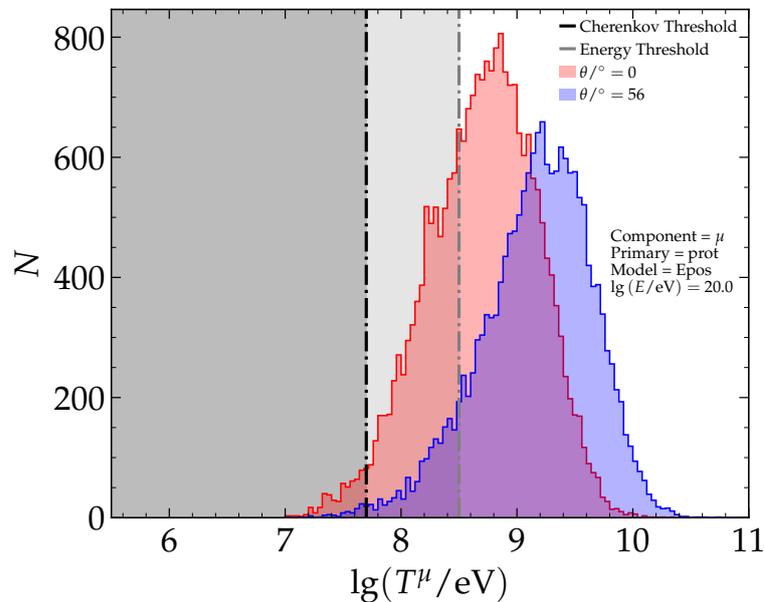
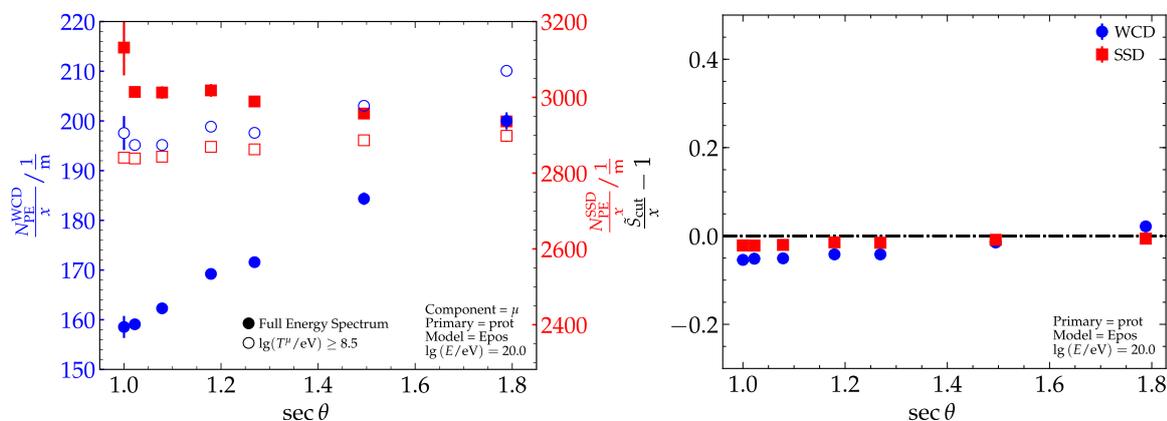


Figure 5.10.: Energy distribution for muons at ground with a shower zenith angle of  $0^\circ$  and  $56^\circ$ . Particles below the lower cherenkov threshold do not produce any photons. In the region between cherenkov and energy threshold, the particles produce photons only along a fraction of their tracklength.

## 5. Detector Responses to the Muonic Shower Component



(a) Applying a cut at the energy threshold of cherenkov photon production along the full tracklength increases the number of photo electrons per tracklength for the WCD as well as flattening their distribution at small shower zenith angles.

(b) Deviation of the signal per particle derived from photo electrons produced with a cut at the energy threshold and the tracklength of the particles.

Figure 5.11.: If an energy cut is applied, so that the muons have sufficient energy to produce cherenkov radiation along their full tracklength, the photo electron production is constant proportional to the respective tracklength.

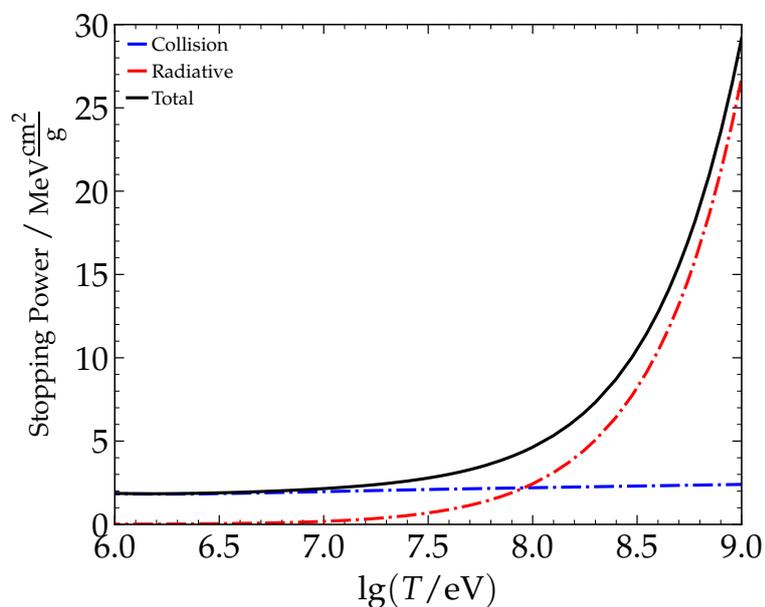


Figure 5.12.: Contributions of ionization and bremsstrahlung losses to the total energy loss for electrons in water. At the critical energy of around 78 MeV the bremsstrahlung becomes the dominant process.

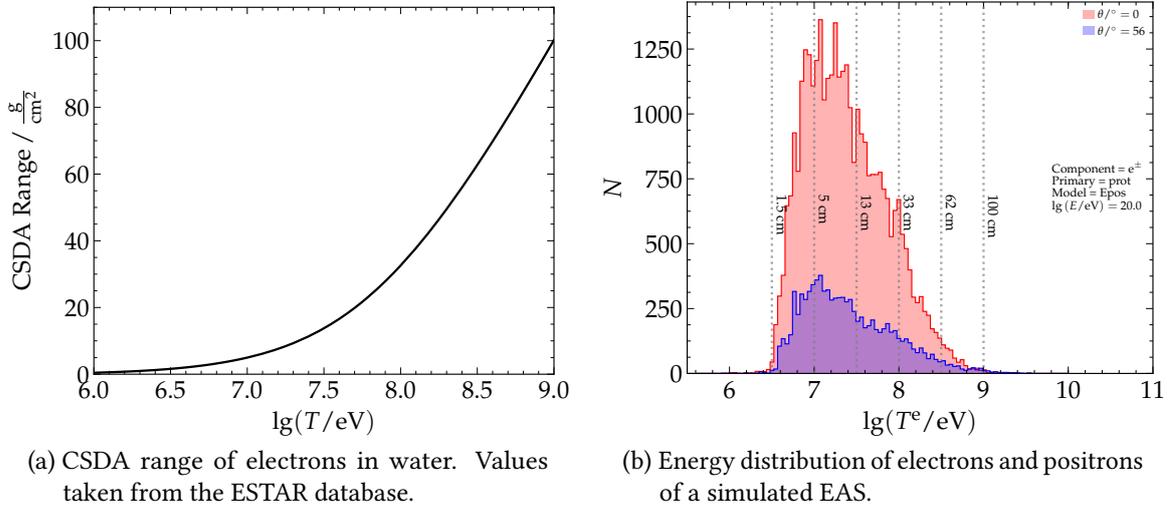


Figure 5.13.: The CSDA range of the electrons in water is included as dotted lines in the energy histogram. Around 65% of the electrons and positrons can penetrate the tank between 0 and 13 cm.

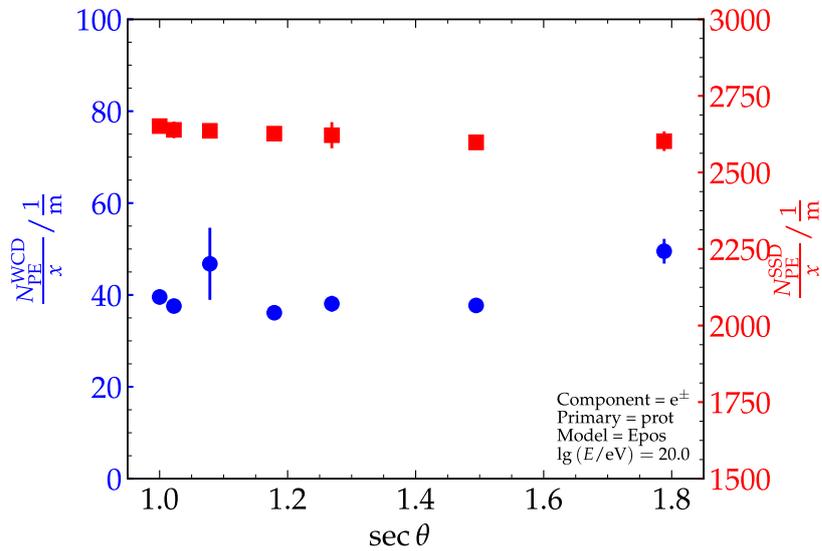


Figure 5.14.: The number of photo electrons created per theoretical tracklength shows for both detectors a constant behavior over different zenith angles. For the WCD, the electrons and positrons do not traverse the full tank, but will deposit all of their energy in the first few centimeters of the WCD.

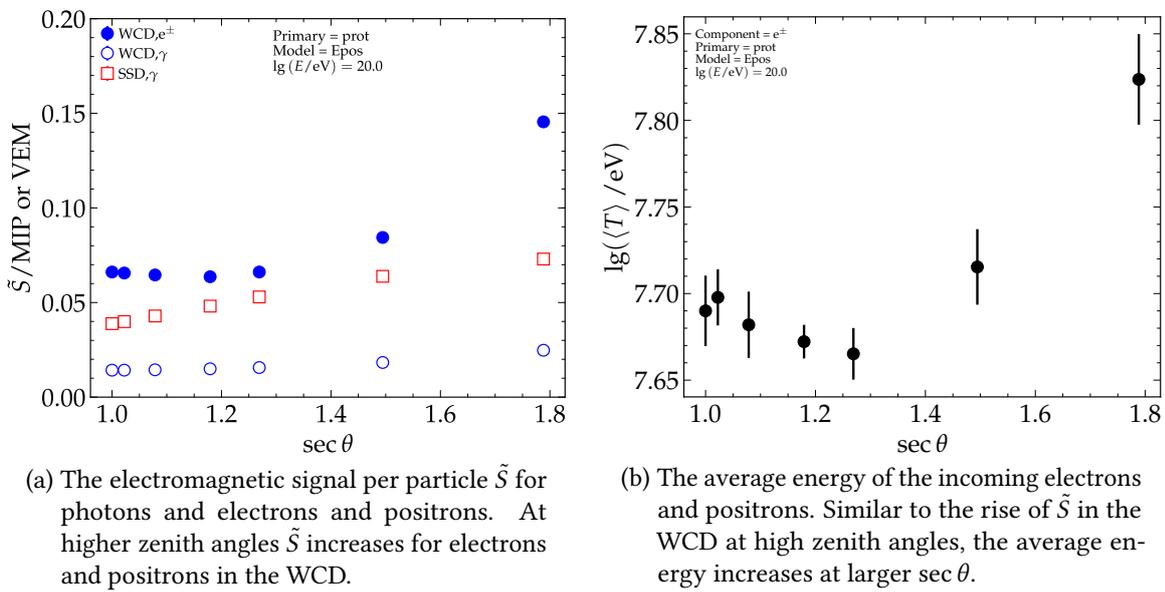


Figure 5.15.: The signal per particle in the WCD is dependent on the traversed tracklength of the particles. This tracklength is also proportional to the energy of the traversing electrons and protons. Therefore, an increase of the average energy of the incoming particles, as seen at the right figure, can lead to an increase of the signal per particle.

## 6. Principal Component Analysis

Estimating mass composition of CRs is of great interest, as described in the beginning of this thesis. With a *Principal Component Analysis* (PCA) it is possible to set up equations for estimating not only the mass composition but also obtaining an energy estimation of the primary particle. As seen in the previous Chapter, the sensitivity of the WCD and SSD varies for different shower components and thus provides two separate estimations of a shower size. The PCA is a method to linearly transform possibly correlated variables to a new subspace, with new, uncorrelated variables that inherit a maximized variance. By doing so, a  $d$ -dimensional dataset can be reduced to a  $k$ -dimensional dataset (with  $k < d$ ) with minimized information loss. In Section 6.1 some prerequisites as well as necessary basic equations and the procedure of the PCA are explained. The input parameters and training of the PCA is done in Section 6.2. In following Section 6.3 and Section 6.4, the PCA is trained and tested with a Monte Carlo dataset.

### 6.1. Formulation

#### 6.1.1. Attenuation Correction of the Shower Size

As seen in Chapter 5 the detector response of the WCD and SSD depends on the different shower components. For increasing zenith angles, the shower has to traverse more atmosphere and at around  $60^\circ$  the shower has passed twice the amount of atmosphere than for a vertical event. Especially for the electromagnetic component the attenuation increases with  $\sec \theta$  and thus the measured shower sizes  $S(1000)$  are dependent on zenith angle. In order to get a zenith-independent estimator of  $S(1000)$  an attenuation correction similar to the CIC in Section 4.1 can be done. The average shower size  $S(1000)$  for all energies is compared to the shower size at a certain reference angle  $\theta = 38^\circ$

$$S_{38} = \frac{S(1000)}{f_{\text{Att}}(\theta)}. \quad (6.1)$$

For the attenuation function the same third order polynomial as for the CIC is chosen

$$f_{\text{Att}} = 1 + ax + bx^2 + cx^3 \quad (6.2)$$

with  $x = \cos^2 \theta - \cos^2(38^\circ)$ . The fit and correction are done for the WCD and SSD separately and the resulting fits are shown in Fig. 6.1. The parameters of the fit are listed in Table 6.1. The zenith-independent shower size parameters  $S_{38}^{\text{WCD}}$  and  $S_{38}^{\text{SSD}}$  can now be used as input parameters for the following principal component analysis.

## 6. Principal Component Analysis

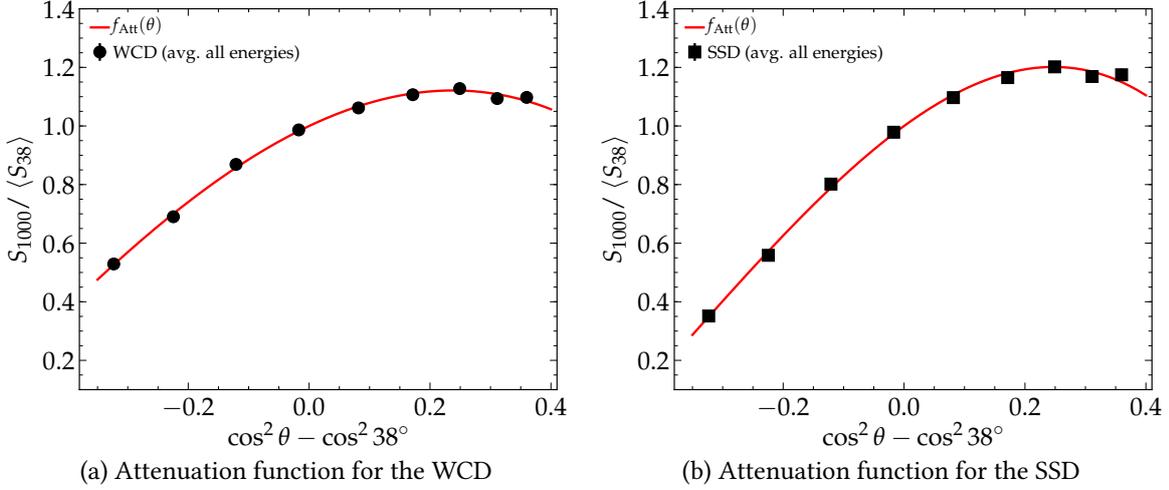


Figure 6.1.: Fit of the attenuation correction for the WCD and SSD. The shower sizes of all energies are normed to the shower size at a zenith angle of  $38^\circ$ .

Table 6.1.: Parameters of the fit of the attenuation function  $f_{\text{Att}}$  for the WCD and SSD

Detector	$a$	$b$	$c$	$\chi^2$
WCD	$0.9746 \pm 0.0021$	$-1.7686 \pm 0.0034$	$-0.7780 \pm 0.2368$	7.27
SSD	$1.5027 \pm 0.0018$	$-2.2615 \pm 0.0034$	$-2.1014 \pm 0.2036$	11.77

### 6.1.2. Basic Formulae

The subsequent calculations are following the description from [31]. When doing a principal component analysis, one is interested in finding a new subset of variables that maximize the variance for a given dataset. As input dataset, the shower size estimation  $S_{38}$  of the WCD and SSD, as well as the energy  $E$  and mass  $A$  of the primary particle is used. Since PCA is a linear method, it is important to have linear input variables as well. Therefore, the input variables are transformed to log-space and the resulting input vector has the following form

$$\mathbf{x}^T = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} s_{\text{WCD}} \\ s_{\text{SSD}} \\ \mathcal{E} \\ a \end{pmatrix}, \quad (6.3)$$

with the variables defined as follows:

$$\begin{aligned} s_{\text{WCD}} &= \lg S_{38}^{\text{WCD}}, \\ s_{\text{SSD}} &= \lg S_{38}^{\text{SSD}}, \\ \mathcal{E} &= \lg E, \\ a &= \ln A. \end{aligned}$$

The variables used as input for the PCA are simulation derived Monte Carlo values. As first step, the input variables have to be standardized, so that each variable contributes equally to the PCA. Therefore, the mean  $\bar{x}_i$  and standard deviation  $\sigma(x_i)$  are determined and then applied to each respective variable in the following way

$$x_{i, \text{norm}} = \frac{x_i - \bar{x}_i}{\sigma(x_i)}. \quad (6.4)$$

A covariance matrix  $\Sigma$  can be set up and calculated with the normalized input data. By calculating the eigenvectors of the covariance matrix, one finds exactly the desired new variables of the dataset, also called the *Principal Components* (PC), as well as an indicator for the amount of variance of the PCs, that are the eigenvalues. To obtain the eigenvectors and eigenvalues of the covariance matrix the following matrix equation is used

$$\Sigma \mathbf{v} = \lambda \mathbf{v} \quad (6.5)$$

with  $\mathbf{v}$  being the eigenvector and  $\lambda$  the eigenvalue. The PCs are then sorted by their eigenvalues, where a large eigenvalue means a large variance of that PC, that yields the most information in the transformed subset. To get a more understandable indicator of the information contained in the new PCs, the explained variance can be calculated from the eigenvectors

$$\text{var}_{\text{exp}}(\mathbf{v}_i) = \frac{\lambda_i}{\sum_{j=1}^4 \lambda_j}. \quad (6.6)$$

By linear combination, the eigenvectors can then be related back to the original input vector

$$\mathbf{x} = p\mathbf{v}_1 + q\mathbf{v}_2 + r\mathbf{v}_3 + s\mathbf{v}_4. \quad (6.7)$$

If the first two principal components inherit enough of the variance, this equation can be approximated by

$$\mathbf{x} \approx p\mathbf{v}_1 + q\mathbf{v}_2 \quad (6.8)$$

and equations for the mass and energy estimations can be set up. One can now rewrite Eq. (6.8) explicitly as follows

$$\begin{pmatrix} s_{\text{WCD}} \\ s_{\text{SSD}} \end{pmatrix} \approx p \begin{pmatrix} v_{11} \\ v_{12} \end{pmatrix} + q \begin{pmatrix} v_{21} \\ v_{22} \end{pmatrix} \quad (6.9)$$

and

$$\begin{pmatrix} \mathcal{E} \\ a \end{pmatrix} \approx p \begin{pmatrix} v_{13} \\ v_{14} \end{pmatrix} + q \begin{pmatrix} v_{23} \\ v_{24} \end{pmatrix}. \quad (6.10)$$

By rewriting Eq. (6.9), one receives for  $p$  and  $q$

$$p = \frac{1}{D} (v_{22}s_{\text{WCD}} - v_{21}s_{\text{SSD}}) \quad (6.11)$$

$$q = \frac{1}{D} (v_{22}s_{\text{SSD}} - v_{12}s_{\text{WCD}}) \quad (6.12)$$

Table 6.2.: Continuous Napoli library of CORSIKA simulations with different combination of parameters. For each combination, 15000 events were reconstructed.

primary	proton, helium, oxygen, iron
hadronic interaction model	EPOS-LHC, QGSJET-II.04
$\lg(E/\text{eV})$	18.5-20.0
$\theta/^\circ$	0 - 65

with  $D = (v_{11}v_{22} - v_{12}v_{21})$ . By inserting the derived equations  $p$  and  $q$  into Eq. (6.10) we get the equations for estimating the logarithm of energy

$$\mathcal{E} = \frac{1}{D} [(v_{13}v_{22} - v_{23}v_{12})s_{\text{WCD}} + (v_{23}v_{11} - v_{13}v_{21})s_{\text{SSD}}] \quad (6.13)$$

and mass

$$a = \frac{1}{D} [(v_{14}v_{22} - v_{24}v_{12})s_{\text{WCD}} + (v_{24}v_{11} - v_{14}v_{21})s_{\text{SSD}}]. \quad (6.14)$$

Important to note is, that the input variables were normed, following Eq. (6.4). By taking this into account, Eq. (6.13) and Eq. (6.14) can be rewritten as

$$\lg E = \frac{1}{D} \left[ (v_{13}v_{22} - v_{23}v_{12}) \frac{s_{\text{WCD}} - \overline{s_{\text{WCD}}}}{\sigma(s_{\text{WCD}})} + (v_{23}v_{11} - v_{13}v_{21}) \frac{s_{\text{SSD}} - \overline{s_{\text{SSD}}}}{\sigma(s_{\text{SSD}})} \right] \sigma(\mathcal{E}) + \overline{\mathcal{E}} \quad (6.15)$$

and

$$\ln A = \frac{1}{D} \left[ (v_{14}v_{22} - v_{24}v_{12}) \frac{s_{\text{WCD}} - \overline{s_{\text{WCD}}}}{\sigma(s_{\text{WCD}})} + (v_{24}v_{11} - v_{14}v_{21}) \frac{s_{\text{SSD}} - \overline{s_{\text{SSD}}}}{\sigma(s_{\text{SSD}})} \right] \sigma(a) + \overline{a} \quad (6.16)$$

Since the dataset, used for training the PCA is fixed, the mean  $\overline{x_i}$  and standard deviation  $\sigma(x_i)$  of the input variables can be treated as constants, that have to be determined yet. The preceding equations therefore simplify to following linear expressions

$$\lg E = A_1 s_{\text{WCD}} + B_1 s_{\text{SSD}} + C_1 \quad (6.17)$$

$$\ln A = A_2 s_{\text{WCD}} + B_2 s_{\text{SSD}} + C_2, \quad (6.18)$$

with the  $A$ ,  $B$  and  $C$  being constants determined by  $v_i$  of the principal components, as well as the mean and standard deviations of the input variables.

## 6.2. Input Parameters and Training

Using the new changes and smaller modifications, applied to the SD reconstruction procedure, that were introduced in Section 4.3 and Section 4.4, a new set of simulations was done. Instead of the previously fixed library of shower simulations, the continuous *Napoli* library was used. This library contains events for 4 different primary particles (proton,

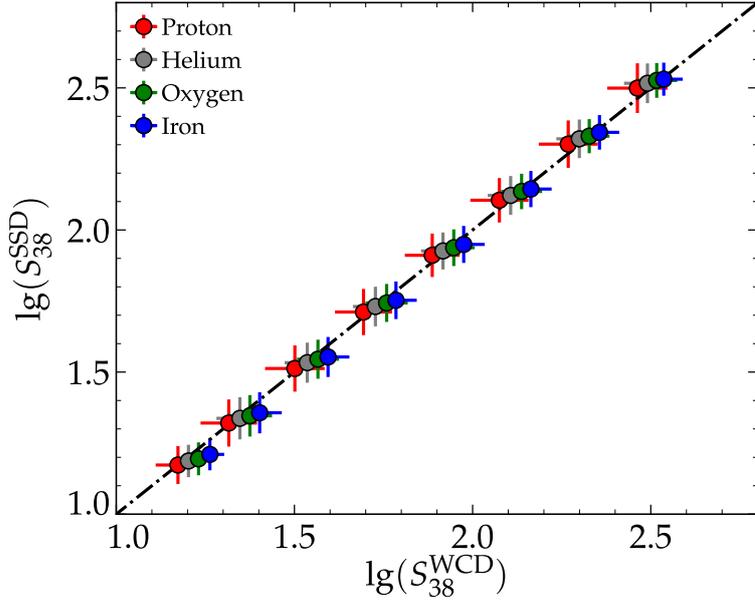


Figure 6.2.: Input variables  $\lg(S_{38}^{\text{WCD}})$  and  $\lg(S_{38}^{\text{SSD}})$  for the different primaries. The dash-dotted line represents the energy axis, ranging from  $\lg(E/\text{eV}) = 18.5$  at the leftmost datapoints to  $\lg(E/\text{eV}) = 19.9$  at the rightmost datapoints. At the lower energies, the WCD signal is slightly larger than the SSD signal. At the highest energies, this relation shifts around and the SSD signals become slightly larger than the WCD signals.

helium, oxygen and iron), as well as a continuous distribution of events over  $\sin^2 \theta$  ranging from 0 to 65 degrees. Over the energy range from  $\lg(E/\text{eV}) = 18.5$  to  $\lg(E/\text{eV}) = 20.0$ , 15000 showers for each primary are simulated, summing up to a total of 60000 events for each model. An overview of the simulated library is given in Table 6.2. The following PCA training and parameterizations were performed on the EPOS-LHC simulations. With the attenuation correction from the setup in Section 6.1,  $S_{38}$  of the WCD and SSD is calculated and converted into logspace. Fig. 6.2 shows the average of these two values, binned for the different primaries and energies. Although the distributions are spread out very wide, the average of the primaries in each bin is separated from each other. In the separate energy bins, the WCD and SSD signal increases with the mass of the primaries. Compared to the SSD signals, the WCD signals increase more significantly for different primaries. This can be attributed to the detector responses to the different shower components. The WCD is more sensitive to the muonic component compared to the SSD. In iron induced showers, more muons are produced, thus having a larger contribution to the WCD detector signal. The *merit factor* (MF) can be used to quantify the mass separation between the proton and iron primaries

$$\text{MF} = \frac{|\langle S_p \rangle - \langle S_{\text{Fe}} \rangle|}{\sqrt{\sigma^2(S_p) + \sigma^2(S_{\text{Fe}})}}. \quad (6.19)$$

This is defined as the absolute difference of the mean of the proton or iron measurements  $S_i$  divided by the squareroot of the sum of the standard deviations. The higher the merit factor,

the better are both measurements separated. From Fig. 6.2 it can be assumed, that the merit factor for the WCD measurements will be larger, than for the SSD measurements, because the points are stronger separated in the direction of  $\lg(S_{38}^{\text{WCD}})$ . The dash-dotted line in Fig. 6.2 can be seen as an energy axis. At the lowest energies, the points are below this axis, meaning that the WCD measurement is larger than the SSD measurement. The WCD is more sensitive to the muonic shower component, while the SSD is sensitive to both, the muonic and electromagnetic shower component. The attenuation of the electromagnetic component reduces and the signal contribution in the SSD increases with rising energy. With the values of  $\lg(S_{38})$ , as well as primary mass and energy at hand, the PCA can be trained. The implementation of the PCA was done with the Python package Scikit-learn [54] and the decomposition of the covariance matrix resulted in the following eigenvectors

$$\mathbf{v}_1 = \begin{pmatrix} -0.5776 \\ -0.5774 \\ -0.5764 \\ -0.0267 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} -0.0261 \\ 0.0194 \\ 0.0530 \\ -0.9981 \end{pmatrix}, \mathbf{v}_3 = \begin{pmatrix} -0.2621 \\ -0.5383 \\ 0.8000 \\ 0.0388 \end{pmatrix}, \mathbf{v}_4 = \begin{pmatrix} -0.7727 \\ 0.6135 \\ 0.1577 \\ 0.0405 \end{pmatrix}, \quad (6.20)$$

with the explained variances

$$\text{var}_{\text{exp}}(\mathbf{v}_1) = 74.765\%, \quad (6.21)$$

$$\text{var}_{\text{exp}}(\mathbf{v}_2) = 25.042\%, \quad (6.22)$$

$$\text{var}_{\text{exp}}(\mathbf{v}_3) = 0.135\%, \quad (6.23)$$

$$\text{var}_{\text{exp}}(\mathbf{v}_4) = 0.057\%. \quad (6.24)$$

Using Eq. (6.15) and Eq. (6.16) to get an estimate for the mass and energy of the primary particle, only the first two eigenvectors are used. Summing up the explained variances of both eigenvectors gives 99.807%, meaning that using the first two eigenvectors will contain 99.807% of the information and the remaining two eigenvectors can be neglected. Inserting the values of  $\mathbf{v}_1$  and  $\mathbf{v}_2$  into the equations Eq. (6.15) and Eq. (6.16) gives

$$\lg E_{\text{PCA}} = -0.7684s_{\text{WCD}} + 1.7546s_{\text{SSD}} + 17.3971, \quad (6.25)$$

$$\ln A_{\text{PCA}} = 79.8865s_{\text{WCD}} - 77.3291s_{\text{SSD}} - 2.9630. \quad (6.26)$$

The first two parameters of  $\ln A_{\text{PCA}}$  are very similar to each other with different signs. Small errors in  $s_{\text{WCD}}$  and  $s_{\text{SSD}}$  result thus in very large propagated errors.

### 6.3. Evaluating Mass Estimation

The quality of the mass estimation with the trained PCA and Eq. (6.26) can now be evaluated. Fig. 6.3a shows the average estimated  $\ln A$  for the different primaries, binned in energy. The errorbars indicate the standard deviation  $\sigma$  of the values and the dashdotted lines represent the true  $\ln A$  values of the primaries. For the different primaries a small, artificial shift in their energy bin is added, for a better readability of the plots. The absolute biases and resolutions for primary are shown in Fig. 6.3b. Again, an artificial shift of the energy bin is done for the different primaries. All primaries are reconstructed with

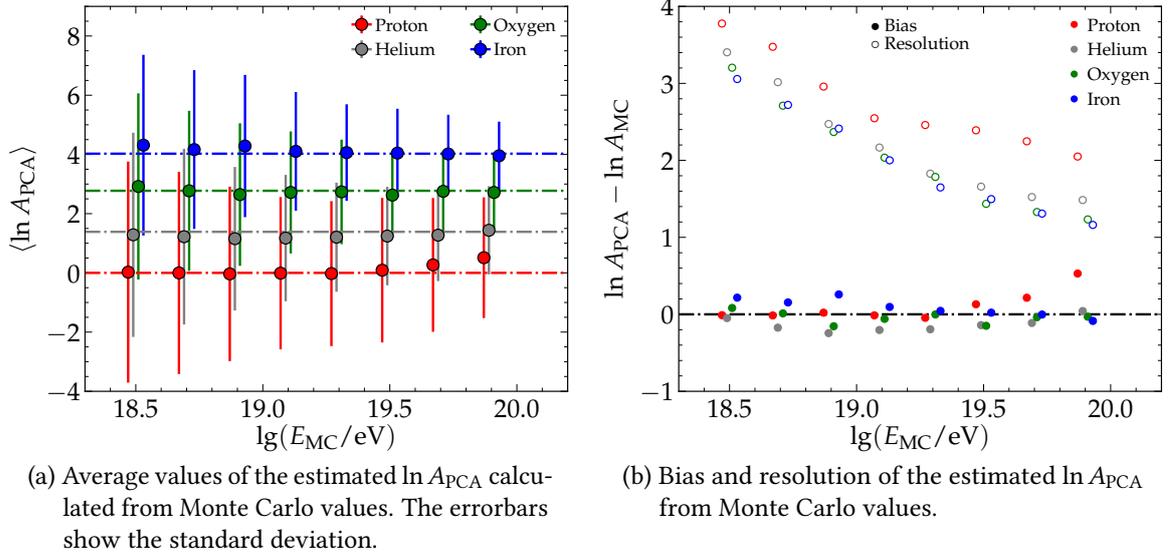


Figure 6.3.: Reconstructing mass, using Monte Carlo values, gives almost unbiased estimates for  $\ln A$ . In both plots a small artificial shift in energy for each primary is added for better readability.

almost no average, absolute bias over the range of energies. The exception is the proton primary in the highest energy bin of  $\lg(E/eV) = 19.9$ , with a fluctuation of the average, absolute bias of approximately 0.5. The resolutions decrease with increasing energy and the resolution of the proton primary is between 7% and 75% larger than the resolutions of the other primaries. At an energy of  $\lg(E/eV) = 19.5$  the merit factor is approximately 1.39, using Eq. (6.19). For smaller energies, the factor is decreasing down to 0.88 and increasing for higher energies up to 1.45. In Fig. 6.4 the distribution of the reconstructed masses for proton and iron is plotted for the energy of  $\lg(E/eV) = 19.5$ . Although the average  $\ln A_{PCA}$  estimate is almost unbiased, there can be mass estimations, that range from  $-20$  up to  $+20$  for  $\ln A_{PCA}$ . The long tails of these distributions could be caused by the combination of different zenith angles. The PCA was trained on  $S_{38}$  under the assumption, that this attenuation correction will account for the different zenith angles. In Fig. 6.5 the average  $\ln A_{PCA}$  and the respective biases and resolutions are shown for different bins in  $\sin^2 \theta$ . Contrary to the results over different energies from Fig. 6.3, there are stronger biases visible over the range of  $\sin^2 \theta$ . The input parameters  $S_{38}^{WCD}$  and  $S_{38}^{SSD}$  are split up in different bins of zenith angles.

### Extension of the PCA

Iron and proton primaries develop at different depths in the shower atmosphere. The shower has to traverse through more or less atmosphere, depending on its zenith angle. This results in a change of the shape of the LDF and the mass dependent relation between both input parameters changes for different zenith angles. The PCA and the resulting Eq. (6.18) are thus extended by zenith dependent factors. The input dataset of the PCA

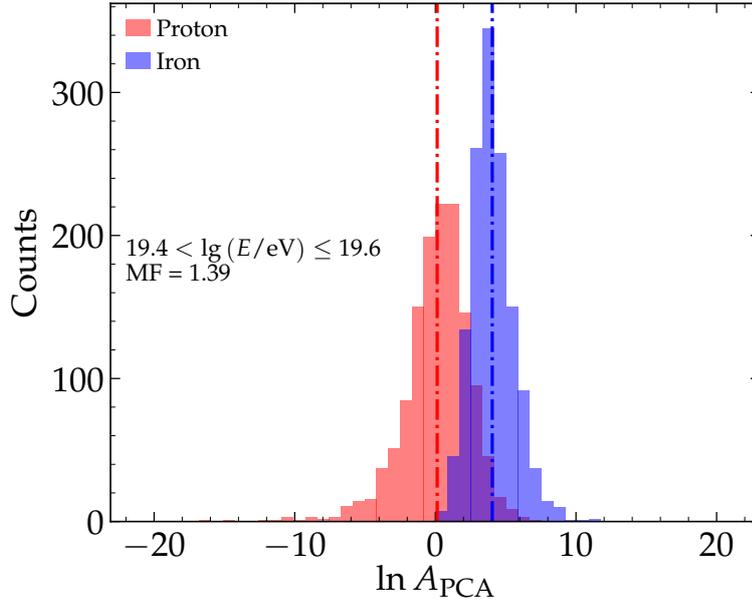


Figure 6.4.: The distributions of  $\ln A_{\text{PCA}}$  have a large spread. However, the average of the distributions gives a rather unbiased estimate of the logarithmic mass, which is depicted with dashdotted lines.

is split up in equal bins of  $\sec \theta$ , ranging from 1 to 1.78, which is approximately equal to the zenith angle range of  $0^\circ$  to  $56^\circ$ . Then, an individual training of the PCA on each of the binned input parameters is performed. As a result, multiple parameters of  $A_2$ ,  $B_2$  and  $C_2$  of Eq. (6.18) are obtained. In Fig. 6.6 the different parameters, binned over  $\sec \theta$  are shown. The  $A_2$  and  $B_2$  parameters show a strong dependency on  $\sec \theta$ . Both parameters are mirrored to each other. This means, if the value of  $A_2$  is decreasing,  $B_2$  is increasing and vice versa. The  $C_2$  parameter exhibits a decreasing trend with  $\sec \theta$ . For each parameter a fit is performed, using a fourth-order polynomial of the form

$$f(\theta) = f_1 + f_2 \sec \theta + f_3 \sec^2 \theta + f_4 \sec^3 \theta + f_5 \sec^4 \theta. \quad (6.27)$$

The fitted curves are depicted in Fig. 6.6 as red lines and the parameters of the function are listed in Table 6.3. It is evident, that the uncertainties of the individual parameters  $f_i$  are very large. This is caused by strong correlations between these parameters, leading to a large variability of their values. Eq. (6.18) can be written up as zenith dependent function with the newly parameterized  $A_2$ ,  $B_2$  and  $C_2$  as

$$\ln A_{\text{PCA}}(\theta) = A_2(\theta)s_{\text{WCD}} + B_2(\theta)s_{\text{SSD}} + C_2(\theta). \quad (6.28)$$

The logarithmic mass is reconstructed again with the zenith dependent Eq. (6.28).  $\ln A$  is plotted over  $\sin^2 \theta$  in Fig. 6.7, Compared to Fig. 6.5a, the biases improved with the new parameterization. The distributions of the logarithmic mass over binned energy did not change significantly with the re-parameterization.

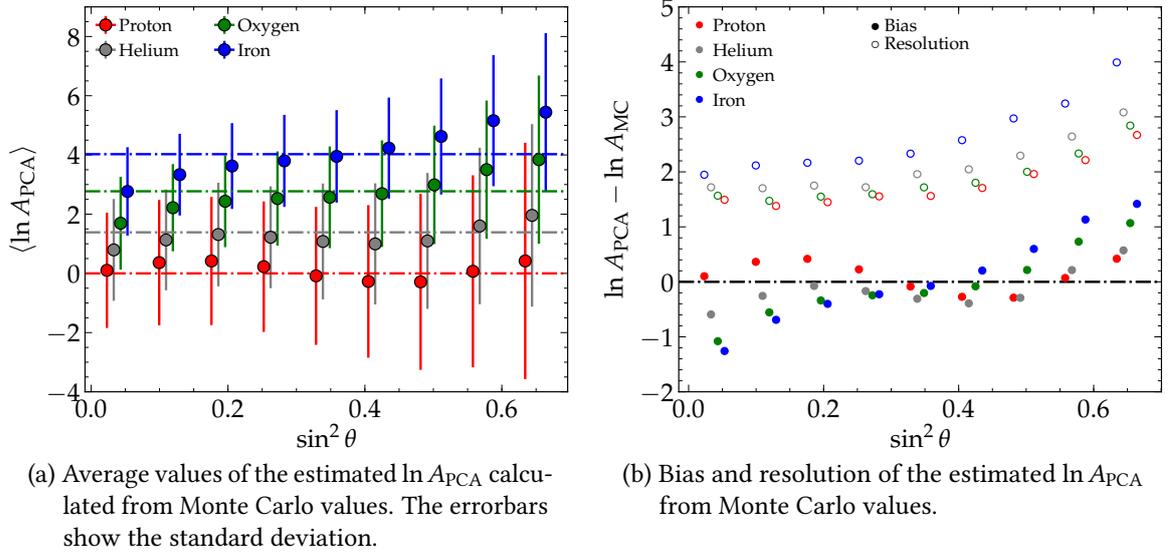


Figure 6.5.: While plotted over energy, the average of  $\ln A_{PCA}$  is unbiased, for different zenith angles, there is a larger absolute bias for the different primaries. In both plots a small artificial shift in energy for each primary is added for better readability.

## 6.4. Energy Estimation

The energy estimation can be evaluated similar to the mass estimation and Eq. (6.25) is re-written as zenith dependent equation:

$$\lg E_{PCA}(\theta) = A_1(\theta)s_{WCD} + B_1(\theta)s_{SSD} + C_1(\theta). \quad (6.29)$$

The parameters  $A_1$ ,  $B_1$  and  $C_1$  are parameterized with the same fourth-order polynomial from Eq. (6.27). The fits of the individual parameters as well as the energy distribution, binned over  $\sin^2 \theta$ , are given in Appendix A.3. The parameters of the fit are shown in Table 6.4. Due to the strong correlations between the parameters  $f_i$ , the uncertainties of these values are very large.

Fig. 6.8a shows the relative bias of the PCA energy estimation compared to the Monte Carlo values. Because Eq. (6.29) is given as logarithmic energy, this value has to be calculated to the energy in eV by

$$E_{PCA} = 10^{\lg E_{PCA}}. \quad (6.30)$$

The energy estimate of the different primaries is almost unbiased over the full energy range for all different primaries. The resolution of the energy reconstruction is ranging from approximately 10% at the highest energies up to 20% going down to lower energies. The PCA energy estimate can be compared directly to the estimate of the SD energy reconstruction, which is given as Eq. (4.20) in Section 4.1. In Fig. 6.8 the relative bias of the of the SD energy reconstruction is shown. On the one hand, the SD reconstruction has a smaller resolution, compared to the PCA, but on the other hand, there are non-neglectible,

Table 6.3.: Fit values of the zenith dependent  $A_2$ ,  $B_2$  and  $C_2$  parameters of  $\ln A_{\text{PCA}}$ .

	$A_2(\theta)$	$B_2(\theta)$	$C_2(\theta)$
$f_1$	$1\ 551.5025 \pm 1\ 120.7625$	$-1\ 592.6674 \pm 1\ 132.0843$	$310.7777 \pm 103.144$
$f_2$	$-3\ 204.1625 \pm 3\ 199.1383$	$3\ 331.6499 \pm 3\ 230.8652$	$-916.0098 \pm 302.3932$
$f_3$	$2\ 607.2681 \pm 3\ 396.4912$	$-2\ 746.9431 \pm 3\ 429.3934$	$997.483 \pm 330.122$
$f_4$	$-962.0653 \pm 1\ 590.4678$	$1\ 029.6054 \pm 1\ 605.4042$	$-478.3822 \pm 159.0162$
$f_5$	$137.8336 \pm 277.2896$	$-149.8942 \pm 279.7887$	$84.9439 \pm 28.5098$
$\chi^2$	3.77	3.41	14.78

Table 6.4.: Fit values of the zenith dependent  $A_1$ ,  $B_1$  and  $C_1$  parameters of  $\lg E_{\text{PCA}}$ .

	$A_1(\theta)$	$B_1(\theta)$	$C_1(\theta)$
$f_1$	$-195.8912 \pm 53.6416$	$196.1604 \pm 52.7383$	$9.5821 \pm 2.9262$
$f_2$	$501.5611 \pm 151.0359$	$-499.3723 \pm 148.3163$	$22.5771 \pm 8.4836$
$f_3$	$-485.4567 \pm 158.4074$	$482.9945 \pm 155.3753$	$-24.2178 \pm 9.1549$
$f_4$	$210.7052 \pm 73.3789$	$-209.4748 \pm 71.8941$	$11.3804 \pm 4.3579$
$f_5$	$-34.6235 \pm 12.6724$	$34.3905 \pm 12.4028$	$-1.969 \pm 0.7721$
$\chi^2$	0.6	0.59	20.3

mass-dependent biases for the SD energy reconstruction. The mass dependent bias of the SD energy reconstruction originates from the form of Eq. (4.20). In this equation, only the WCD detector signal is used to calculate the energy. The estimated shower size  $S(1000)$  from either detector is dependent on the shower components as shown in Chapter 5. With the different primaries developing at different points in the atmosphere the shower components are more or less attenuated and the signal is thus dependent on the primary as well. With only one mass dependent estimation of the shower size, this translates directly into the energy estimation as a mass dependent bias. By including the SSD measurement via the PCA, both detectors are now used for the energy estimation. The two correlated measurements allow then for a mass-independent estimate of the energy of the primary particle.

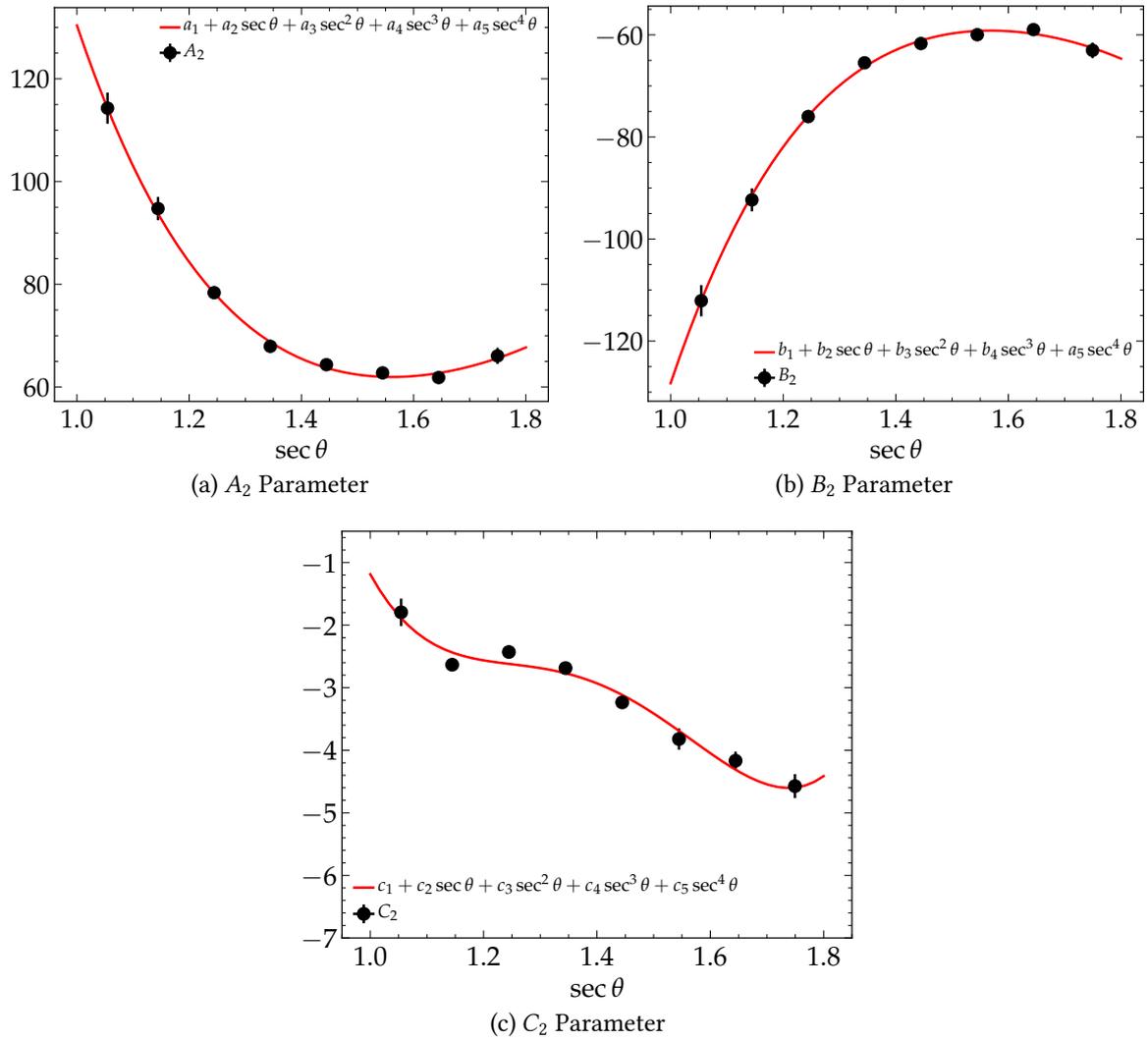


Figure 6.6.: The PCA is trained on different bins in  $\sec \theta$  and the resulting  $A_2$ ,  $B_2$  and  $C_2$  parameters are fitted to a fourth-order polynomial.

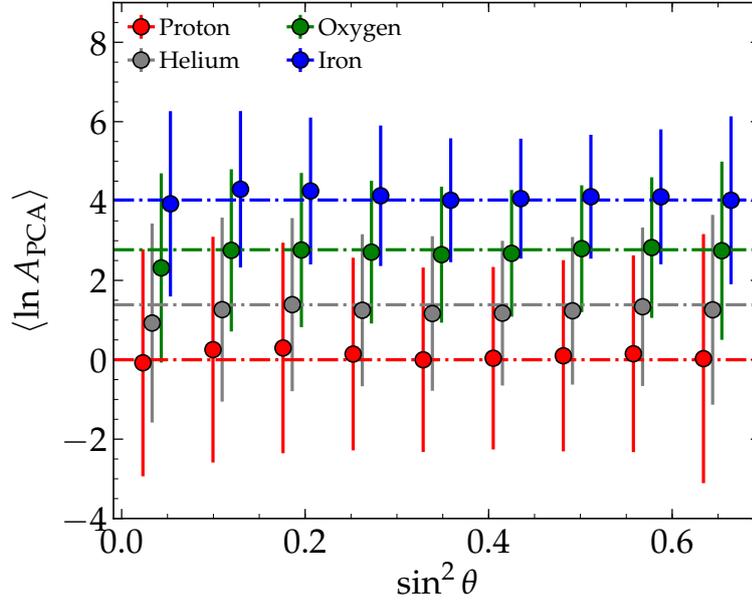
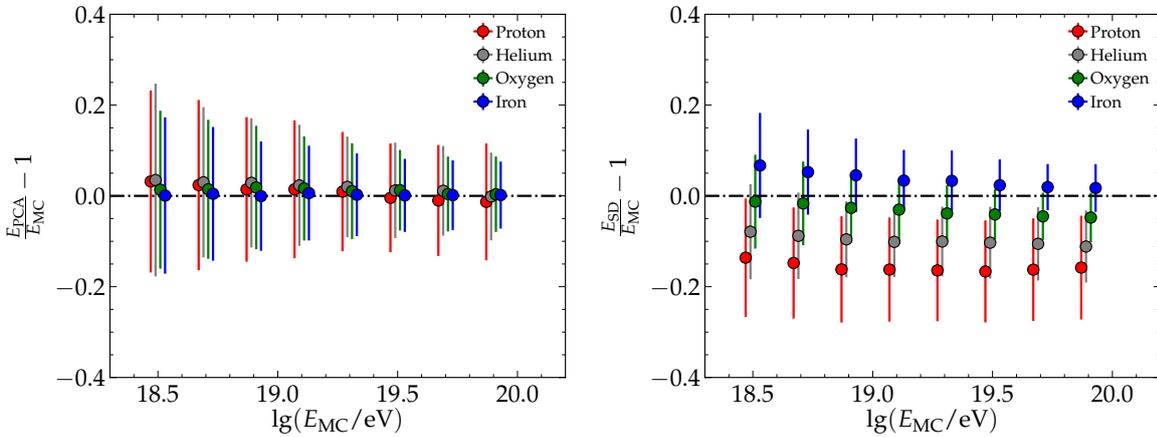


Figure 6.7.: The reconstruction of  $\ln A$  is re-evaluated with the zenith dependent function Eq. (6.28). The previous biases over  $\sin^2 \theta$  (see Fig. 6.5a) are now reduced.



(a) The energy estimation for different primaries is almost unbiased over the full energy range. The PCA can therefore give a good approximation.

(b) Using Eq. (4.20) as the SD energy reconstruction results in a mass dependent bias.

Figure 6.8.: Comparison between the PCA energy estimation and the SD energy estimation. For both plots an artificial, small shift of the Monte Carlo energy is done for a better readability.

## 7. A First Look at Mass Composition

With the PCA trained and validated with Monte Carlo values, the retrieved PCA parameters are applied to the data measured with Auger. The selection of events, based on different quality cuts, is presented in Section 7.1. Before applying the equations to the data and having a look at the mass composition, some limitations of the PCA predictions are given in Section 7.2. In Section 7.3, the energy estimation of the PCA is compared to the current energy reconstruction with the SD. In the last Section, the mass composition is examined.

### 7.1. Event Selection

In September 2016, 12 prototype SSD units with a UUB were deployed in the EA and have been taking data since then [55]. During March 2019, 77 SSDs have been deployed with the non-upgraded electronics board in the so-called *SSD Pre-Production Array* (PPA). In Fig. 7.1, an overview of the stations in the PPA and EA is shown. The 77 stations of the pre-production array are marked in blue, the engineering array is marked in yellow. Results of the performance and calibration of the upgraded stations have been discussed in [31]. For the dataset, events measured within the PPA between March 17, 2019 and May 31, 2021 are selected. Selection cuts are chosen to filter out accidental events or events out of the range of full efficiency of the array.

- **Bad period cut**

The bad period cut is a combination of two exclusion criteria. The first criteria is used to filter out lightning events. Usually, if a T2 trigger condition is fulfilled, the information is send to CDAS. Lightning events can cause a high rate of T2 triggers which results in higher T3 trigger rates, that can not all be processed in time and thus cause a large T3 error rate. The second exclusion criteria is, if the time between 2 events does not meet the expectations [56, 57].

- **6T5 cut**

The event must have passed the 6T5 trigger criterium, which ensures that the shower core has not landed outside of the array (see Section 4.1).

- **Energy cut**

At the energy of  $\lg(E/eV) = 18.5$  the array starts to operate at full efficiency. An energy cut at  $\lg(E/eV) = 18.0$  is set, for the case that the SD energy estimate predicts a lower value than the PCA energy estimate.

- **Zenith cut**

The zenith range of full efficiency for SSDs is between  $0^\circ$  and  $56^\circ$  and therefore chosen as required range.

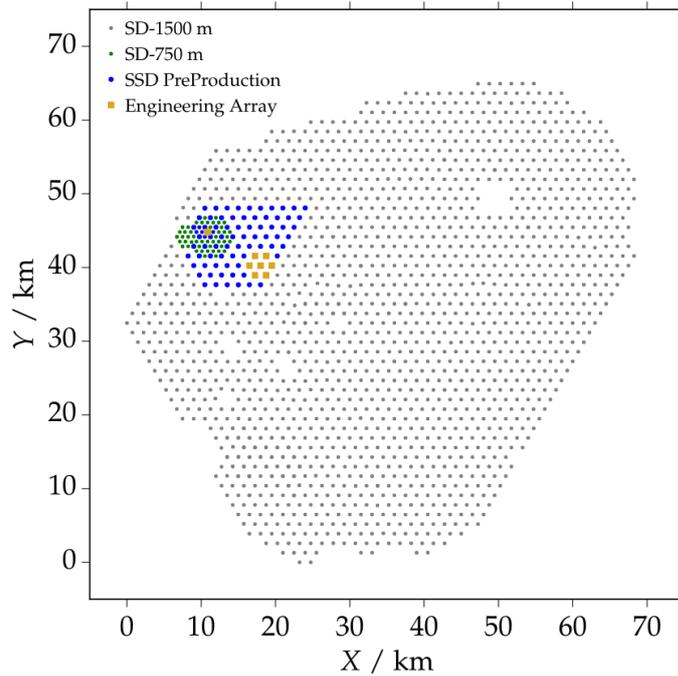


Figure 7.1.: Map of the SD array. The blue markers show the position of the seventy-seven stations of the PPA. The yellow squares show the 12 EA stations.

- **Hottest station cut**

As next requirement, the hottest station (the station closest to the shower core), is required to have a working SSD and uses the UB electronics.

- **Station multiplicity cut**

The SSDs are triggered as subsidiary detectors to the WCD. However, it is not necessary that the triggered SSD has also measured a signal. To ensure that the quality of the SSD LDF fit increases, a minimum station multiplicity of 3 SSDs is required.

During the period between March 17, 2019 and May 31, 2021, a total of 51915 events were recorded. Overall, 7191 events passed the selection cuts and are used in the downstream analysis. This number will reduce one more time, because the current energy cut is set at a lower energy of  $\lg(E/\text{eV}) = 18.0$ . The energy cut will be increased to  $\lg(E/\text{eV}) = 18.5$  depending on the chosen energy reconstruction in Section 7.3. An overview of the events after the selection cuts is given in Table 7.1.

## 7.2. Current limitations

Prior to the application of the PCA to measured data with Auger, the systematic errors are of high interest. When estimating mass and energy with the PCA derived Eqs. (6.15) and (6.16), the main source of systematic errors arise from the reconstruction of the shower size  $S(1000)$  with the WCD and SSD. For the WCD, previous studies have shown a

Table 7.1.: Number of events reconstructed with the PPA with applying the different selection cuts.

Selection cut	Number of events	$\Delta$	%
Total	51 915	0	100.00
Bad period cut	51 906	9	99.98
6T5 Cut	43 024	8 882	82.87
Energy cut	31 374	11 650	60.43
Zenith cut	27 048	4 326	52.10
Hottest station cut	9 754	17 294	18.79
Station multiplicity cut	7 191	2 563	13.85

systematic uncertainty on the reconstructed shower size of 5% [58]. However, the WCD LDF is data derived and thus, the reconstructed shower sizes differ for simulations and data. The LDF of the SSD is derived from simulations and therefore it is not known yet, how the simulated and reconstructed shower sizes differ. Fig. 7.2 shows the bias and the resolution of reconstructed shower sizes, using the WCD and SSD measurements from the previous simulations. For increasing energy, the bias of the WCD is linearly increasing from  $-4\%$  at  $\lg(E/\text{eV}) = 18.5$  up to approximately  $12\%$  in the highest energy bin of  $\lg(E/\text{eV}) = 19.9$ . The SSD has a non-constant negative bias of approximately  $5\%$ , which could arise from different biases at different zenith angles. It is smaller than the bias of the WCD. This can be attributed to the different LDFs and how they were derived. Since the SSD LDF is simulation derived, the bias is smaller, when the shower size is reconstructed from simulations. The WCD LDF is data derived and thus has a larger bias in the shower size reconstruction, due to differences between shower simulations and real air showers. One main difference is the muon deficit, mentioned in Section 3.3. To get an estimation of systematic uncertainties for the logarithmic mass, the Monte Carlo values, that were used in Sections 6.3 and 6.4, are artificially biased. A constant bias of  $5\%$  for the reconstructed shower size from the WCD and SSD is used, for the sake of simplicity. In Fig. 7.3 the reconstructed logarithmic mass for either a positive bias of the WCD, or the SSD, can be seen. A positive WCD bias of  $5\%$  results in an absolute positive bias of  $1.69$  in  $\ln A$ . For a positive SSD bias of  $5\%$ , the logarithmic mass has an absolute negative bias of  $-1.64$  and for negative biases of the WCD and SSD, the signs of the resulting absolute biases change accordingly. Thus, no significant bias will be visible, if both detector measurements are biased in the same direction. The total bias of  $\ln A$  is shown as a function of the bias of the input variables in Fig. 7.4. However, if the bias of the WCD and SSD is anticorrelated, the bias of the logarithmic mass increases further. To get a maximum estimation of the bias of  $\ln A$  a bias of  $+5\%$  for the WCD and a bias of  $-5\%$  for the SSD and vice versa is assumed. This gives a systematic bias of

$$\sigma_{\text{sys}}(S(1000)) = \pm 3.57. \quad (7.1)$$

Another bias can be caused by the choice of the hadronic interaction model, used for the air shower simulations. The PCA so far has been parameterized with the use of EPOS-LHC.

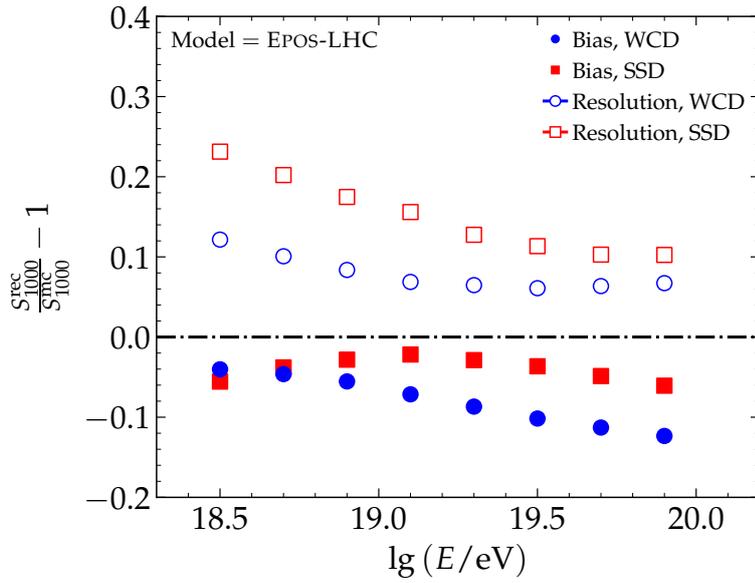


Figure 7.2.: The reconstructed shower size from WCD measurements has an increasing negative bias with logarithmic energy. The bias of the SSD is smaller than for the WCD, since it uses an simulation derived LDF, which is here applied to simulations again. However, the resolution of the SSD is for low energies approximately twice as large as for the WCD.

Using QGSJET-II.04 the shower size of the SSD and WCD is reconstructed and the previous steps to retrieve  $S_{38}$  are performed similar to the steps described in the previous Chapter. Detailed plots and values can be seen in Appendix A.5. The QGSJET-II.04 values are then used in combination with the PCA parameterization done with EPOS-LHC to retrieve  $\ln A_{PCA}$  resulting in a small bias of

$$\sigma_{\text{sys}}(\text{model}) = \pm 0.25. \quad (7.2)$$

The total systematic uncertainty is calculated by summing up both systematic errors in quadrature, assuming no correlation between the two quantities

$$\sigma_{\text{sys}}(\text{tot}) = \sqrt{\sigma_{\text{sys}}(\text{model})^2 + \sigma_{\text{sys}}(S(1000))^2} \approx 3.58. \quad (7.3)$$

The resolution of the estimated logarithmic mass is dependent on the resolution of the input parameters as well. In Fig. 7.5 the resolution of  $\ln A$  as a function of the WCD and SSD resolution of  $S_{38}$  is shown for a single energy bin of  $\lg(E/eV) = 19.5$ . Additionally the current resolution is marked in the figure with a yellow marker. This resolution is calculated from a WCD resolution of 6% and an SSD resolution of 11%, which gives a resolution of the logarithmic mass of

$$\sigma(\ln A_{PCA} - \ln A_{MC}) \approx 5.22. \quad (7.4)$$

If the resolution of the SSD could be improved to match the resolution of 6% of the WCD this would give

$$\sigma(\ln A_{PCA} - \ln A_{MC}) \approx 3.16, \quad (7.5)$$

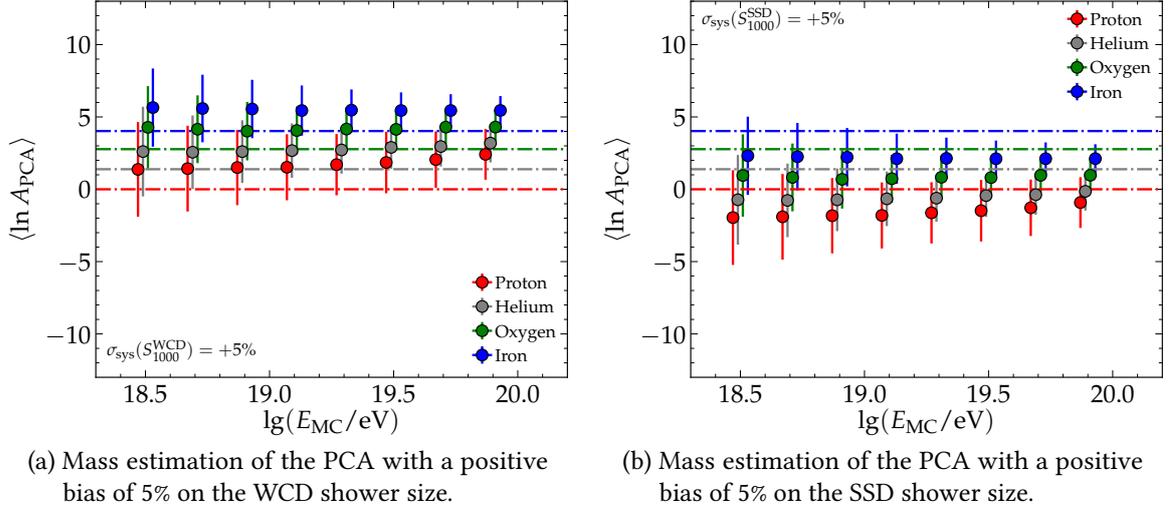


Figure 7.3.: A positive bias on either WCD or SSD measurement results in a positive or negative bias of the reconstructed logarithmic mass. For a positive WCD bias,  $\ln A$  is also positively biased. For a positive SSD bias however,  $\ln A$  is negatively biased. Adding the same positive bias to the SSD and WCD together, the resulting logarithmic mass would be almost unbiased again.

which would be an improvement of approximately 65%. Similar to the systematic uncertainties of the mass estimation, the uncertainties for the energy estimation can be calculated. The estimated energy from the PCA has a systematic uncertainty of approximately 18%, if a combined systematic bias of +5% in the SSD and  $-5\%$  in the WCD is assumed.

### 7.3. Comparison with current SD Energy Reconstruction

With a better understanding of the systematic errors of the PCA, the energy estimation can be compared to the SD energy estimation. For the energy reconstruction the 7191 events from Section 7.1 are used. The energy of the events is reconstructed in two ways. First, the SD energy reconstruction as described in Section 4.1 is performed. Then, the second estimation is done, using the PCA energy reconstruction from Eq. (6.29). Both reconstructions are compared by plotting the SD energy estimate of each event against its PCA energy estimate. The resulting plot is shown in Fig. 7.6a. A wide variation between the estimated energies of both reconstructions can be observed. In addition to the reconstructed energy, the logarithmic mass is estimated using Eq. (6.28) and added as color code in Fig. 7.6a. Events, that are reconstructed with a negative  $\ln A$ , are shown in red and events, exceeding the logarithmic mass of iron of 4.025 are shown in blue. If  $\ln A$  is reconstructed with even more extreme positive or negative values, at a value of  $\pm 14$  for  $\ln A$ , the color code is chosen to be dark red or dark blue. All events that are within the limits of  $0 \leq \ln A_{\text{PCA}} \leq 4.25$  are plotted with black markers. As seen in Section 6.3, the distributions of  $\ln A$  exceed the mass estimations beyond 0 and 4.025. It is therefore important to note, that these mass bins are just for illustrative purposes and shall not

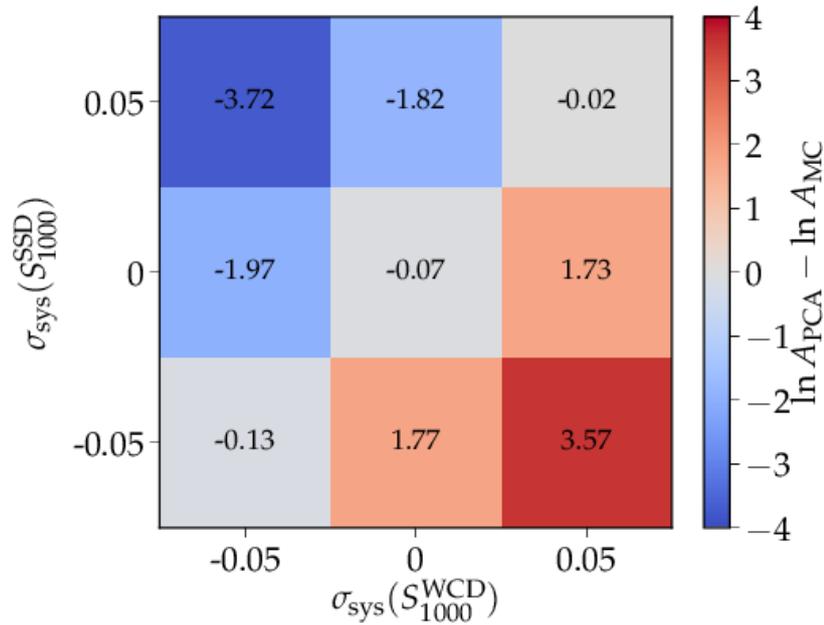


Figure 7.4.: Systematic bias of the logarithmic mass as function of the systematic biases of the reconstructed  $S(1000)$  of the WCD and SSD. The systematic bias in  $\ln A$  remains low, if both input parameters are biased in the same direction. However, if they are biased in different directions, the total bias increases.

imply any statement about the mass composition. However, some assumptions can be drawn from the plot. If the PCA estimated energy is greater than the SD energy estimation, the event will be more likely reconstructed with a negative mass. If the energy estimation from the PCA is smaller than the SD energy estimation, the event will be more likely reconstructed with a logarithmic mass, greater than 4.025. This effect is caused due the rotation of the axes of the input parameters during the PCA. For larger differences between the PCA and SD energy estimate, the mass estimation will predict as well larger or smaller masses. Since the mass composition shall be evaluated as a function of energy, it is of interest, how the average energy bins of the PCA and SD estimation are related to each other. A binning with  $\Delta \lg(E/\text{eV}) = 0.1$  from  $10^{18.5}$  eV to  $10^{19.5}$  eV is chosen. The last energy bin includes all energies greater than  $\lg(E/\text{eV}) = 19.5$ , because of the limited number of events at highest energies. The average estimated energy, using the SD and the PCA reconstruction, is calculated for every bin in two different ways. First, the average SD and PCA energy is calculated by binning over the SD estimated energy. In the lower bins, the average SD energy should match with the chosen bin centers, while the average PCA energy estimate can differ from this prediction. Then, the average SD and PCA energy is calculated again, but now binned over the PCA estimated energy. The average PCA energy should now match with the lower bin centers, while the SD energy estimate can differ. In Fig. 7.6b, the resulting plot is shown. The average energy estimates, binned over the reconstructed SD energy are plotted with blue markers, the average estimates, binned over the PCA energy are plotted with red markers. If the binning is done with the SD energy estimation, the average of the estimated energies, using the PCA or SD

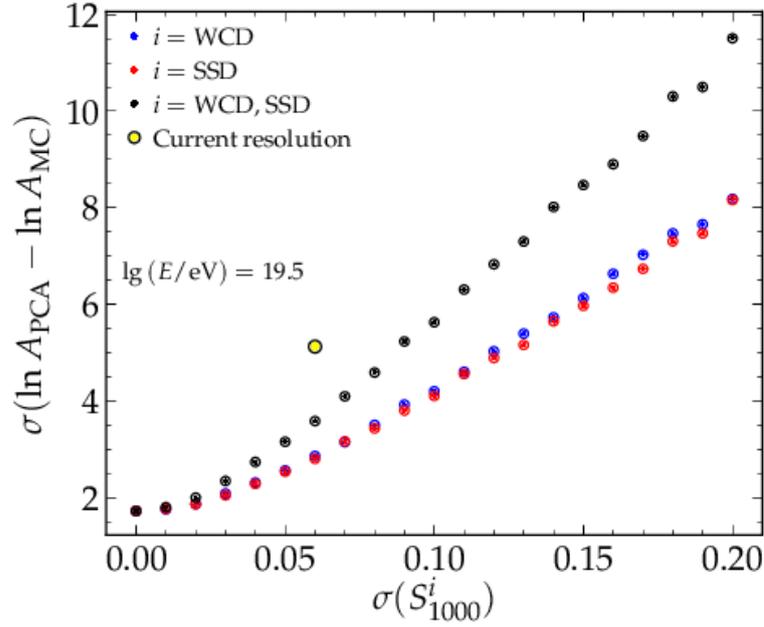


Figure 7.5.: The resolution of  $\ln A$  increases with the resolution of the WCD and SSD. The current resolution is calculated from a 6% WCD resolution and an 11% SSD resolution.

reconstruction is almost of the same size. However, if the binning is done with the PCA energy estimation, the average SD energy estimation is smaller than the average PCA estimation. As previously mentioned, a PCA energy, reconstructed larger than the SD energy, results in a logarithmic mass, that is estimated more likely with negative values. To improve the PCA energy estimations the energy cut at  $\lg(E/eV) = 18.5$  of the SD energy is raised up to  $\lg(E/eV) = 19.0$ . With this higher cut, the spread of the PCA energy prediction is reduced.

## 7.4. First Estimation of the Logarithmic Mass

The logarithmic mass of each event from the previously introduced dataset is reconstructed, using Eq. (6.28). The estimation and binning in energy can be done by either using the SD energy reconstruction, or the PCA. In Section 7.3 it was shown, that a binning over the SD energy, will result in similar average energies, estimated by SD and PCA. A binning over the PCA energy however, results in higher average PCA energy estimates than the SD energy estimates. At first, the SD energy estimate is chosen for the binning. The energy cut of  $\lg(E/eV) = 18.5$  is applied to the SD estimated energy and the previously 7191 events get further reduced to a total of 958, which is 1.85% from the initial number of recorded events. Fig. 7.7 shows the estimated mass composition as a function of SD energy. The mean of each energy bin is chosen for the position on the energy axis. At each point, the number of events contained in that bin are given and the square brackets denote the systematic uncertainty. The errorbars of the average  $\ln A$  are the standard error. Simulation predictions for proton and iron composition are shown as red and blue

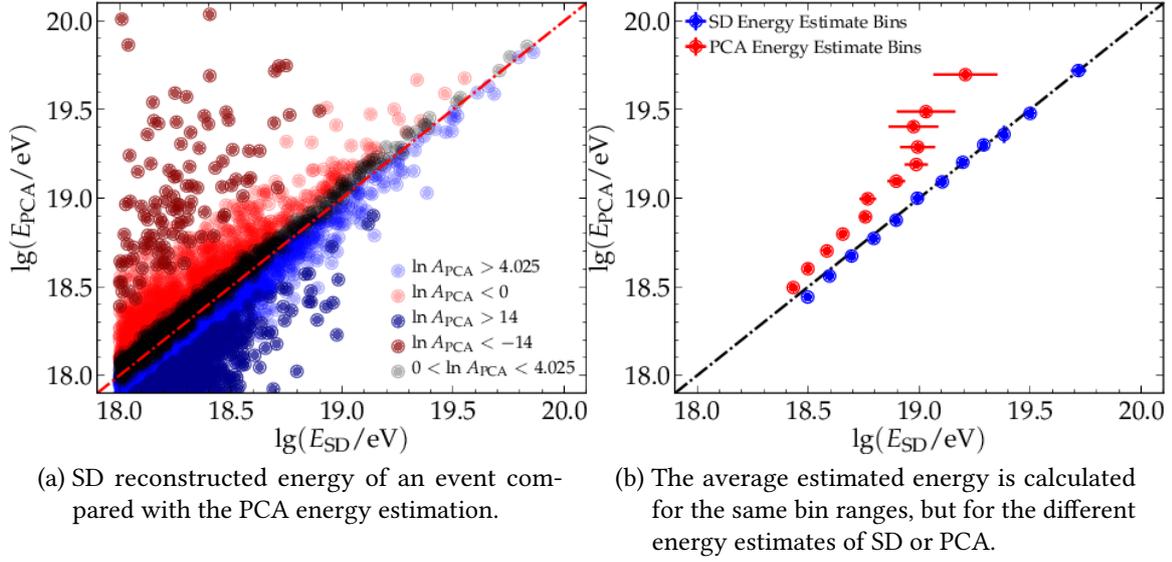


Figure 7.6.: Events with a PCA energy, estimated larger than the SD energy, tend to reconstruct negative masses. If the PCA energy is estimated smaller than the SD energy, PCA tends to reconstruct a mass larger than  $\ln A = 4.025$ . Choosing a binning over the PCA energy estimate, results in energies estimated larger than the average SD estimations.

dashdotted lines. Additionally, the logarithmic mass estimation from  $X_{\max}$  measurements is shown as brown points, using the  $X_{\max}$  data from [27]. From [59], an estimation of  $\ln A$  with the shower maximum can be calculated, using

$$\langle \ln A \rangle = \frac{\langle X_{\max} \rangle - \langle X_{\max} \rangle_p}{f_E}, \quad (7.6)$$

with  $\langle X_{\max} \rangle_p$  as the mean depth at the maximum of proton showers and the energy dependent parameter  $f_E$ :

$$f_E = \xi - \frac{D}{\ln 10} + \delta \lg \left( \frac{E}{E_0} \right). \quad (7.7)$$

The parameters  $X_0$ ,  $D$ ,  $\xi$  and  $\delta$  are dependent on the used model [59]. For this thesis, the parameters of EPOS-LHC were chosen from [60]. The average of the  $\ln A_{PCA}$  distributions is fluctuating around the iron line. This does not match with the  $X_{\max}$  measurements. The mass estimation of the PCA is dependent on the relative bias between the WCD and SSD shower size, as shown in Section 7.2. In addition, the training of the PCA was done on a simulated dataset. Therefore, a bias between the simulated and real, measured shower size complicates an interpretation of the results additionally. With the large systematic uncertainties, it is not possible to draw any conclusions for now from the logarithmic mass, estimated with the PCA. Fig. 7.8 shows the distribution of the estimated logarithmic mass in one SD energy bin of  $\lg(E/\text{eV}) = 19.0 \pm 0.05$ . As expected, the distribution has a very large spread, reaching to values of almost  $\pm 40$  for the logarithmic mass. Further

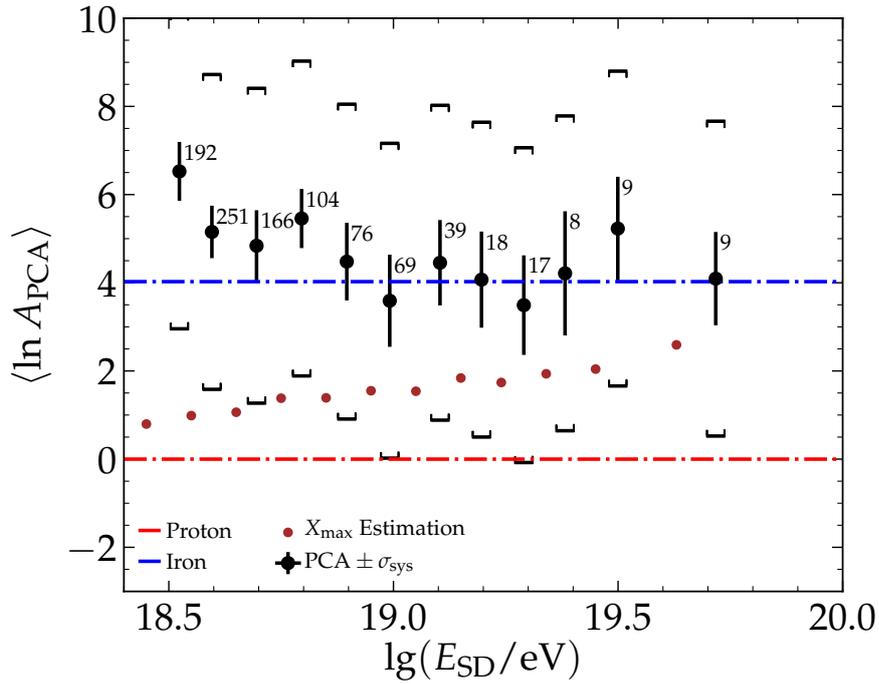


Figure 7.7.: Estimation of the mass composition as a function of logarithmic SD energy. The numbers denote the amount of data points in each bin. The estimations from  $X_{\max}$  were computed, using the data from [27] with the parameterizations from [59, 60].

histograms for different energy bins are found in Appendix A.4. The procedure is then repeated again with a binning over the PCA energy estimate. An increased energy cut of  $\lg(E/eV) = 19.0$  for the SD energy estimation is chosen, to reduce the spread of the PCA energy distribution. The resulting distribution of  $\ln A_{\text{PCA}}$  is shown in Fig. 7.9. Due to the stricter energy cut, the number of events is further reduced. The estimates differ, compared to the mass estimate with a binning over the SD energy. This can be accounted due to a differing selection of events, depending on the choice of the energy reconstruction. The use of timing information, or other variables, as well as a better knowledge about the WCD and SSD biases would be needed, to improve the resolution of  $\ln A_{\text{PCA}}$ .

## 7.5. Possible Further Improvements

As seen in the previous section, the distribution of  $\ln A_{\text{PCA}}$  and  $\lg E_{\text{PCA}}$  is very wide. The  $A$  and  $B$  parameters of Eqs. (6.28) and (6.29) are very similar to each other with inverted signs. A small error in either the SSD or WCD estimation of the shower size will thus result in a very broad distribution of estimations of logarithmic mass or energy. A prediction of the mass composition at an individual event level is thus not possible at the current state. It is of importance to understand and know the biases between the simulated and measured data. Further extensions to the PCA might lead to improved resolutions. The PCA could improve, by including more parameters to the training data, as for example

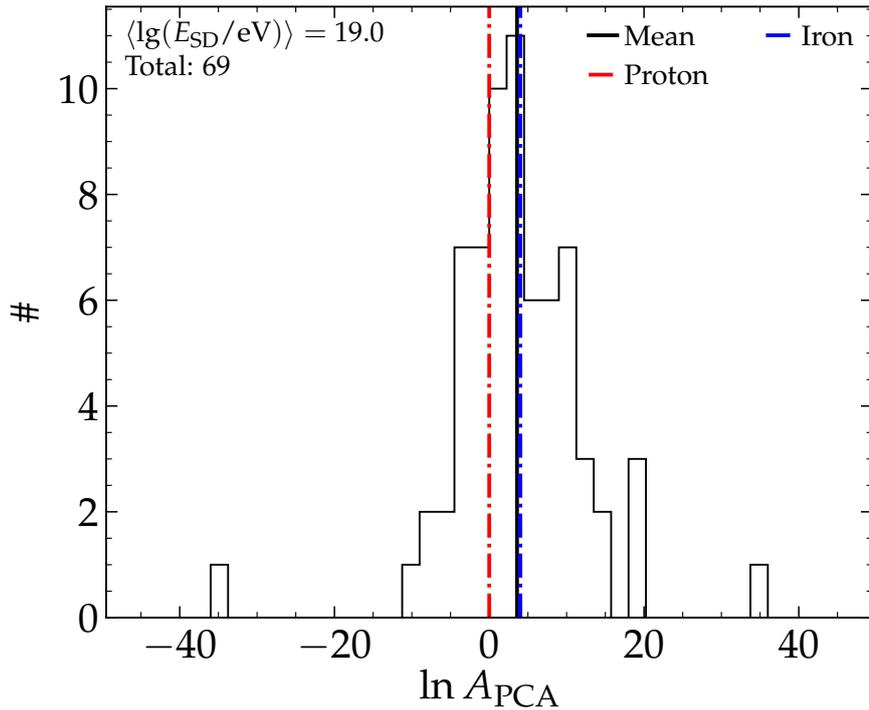


Figure 7.8.: Distribution of  $\ln A_{PCA}$  at a reconstructed SD energy of  $\lg(E/\text{eV}) = 19.0 \pm 0.05$ . The proton and iron prediction are given as red and blue, dashdotted lines and the average of the distribution is shown as black line.

timing information of the station signals. The so-called *risetime*  $t_{1/2}$  could be used as such timing information. It is defined as the time of a station signal to increase from 10% to 50% of the final magnitude of the integrated signal. This parameter has already been used to try to discriminate between late and early developing showers in the atmosphere [61, 62, 63]. An event-level risetime from WCD measurements, that has been proposed in [64], could be used as a further input parameter. Fig. 7.10 shows the average event-level risetime over  $\lg(S_{38})$ , binned in logarithmic energy, at a zenith angle of approximately  $38^\circ$ . As dataset, the Monte Carlo dataset from Section 6.2 was used. The average risetime is different for the varying primaries, thus it could be well suited as additional parameter for the PCA. Similar to this, an event-level risetime for the SSD could be calculated and evaluated in further studies. Currently, the WCD and SSD measurements are used for separate reconstructions of  $S(1000)$ . A ratio of the individual WCD and SSD station signals can directly account for the differing detector responses to the shower components. Using data of the simulation library from Section 6.2, the ratio of the individual SSD and WCD stations is calculated. The ratio as a function of distance for all energies from  $\lg(E/\text{eV}) = 18.5$  to  $20.0$  is shown in Fig. 7.11a. In Fig. 7.11b all stations up to a distance of 1200 m to the shower axis are used for the ratios. Both figures are binned in equal ranges of  $\sec \theta$ . At larger zenith angles, the ratio of SSD to WCD gets smaller which is due to the attenuation of the electromagnetic component. The ratio decreases slightly with increasing distance to the shower axis and for larger energies, the ratio increases.  $S(1000)$  of the WCD as well as the ratio of SSD to WCD station signals could be used as PCA input variables, instead of the estimated

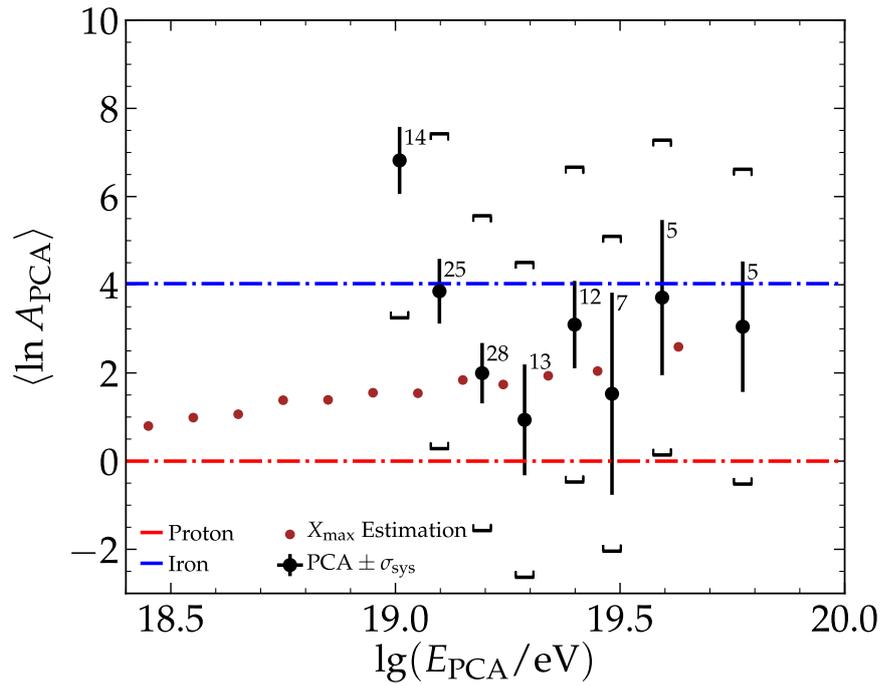


Figure 7.9.: Estimation of the mass composition as a function of logarithmic PCA energy. The numbers denote the amount of data points in each bin. The estimations from  $X_{\max}$  were computed, using the data from [27] with the parameterizations from [59, 60].

shower size of the WCD and SSD as separate inputs. Other approaches, as for example air shower universality [65, 66], or the use of deep neural networks [67], that cover more variables with respect to different shower properties, can give as well promising results in further studies with AugerPrime [68, 69, 70, 71].

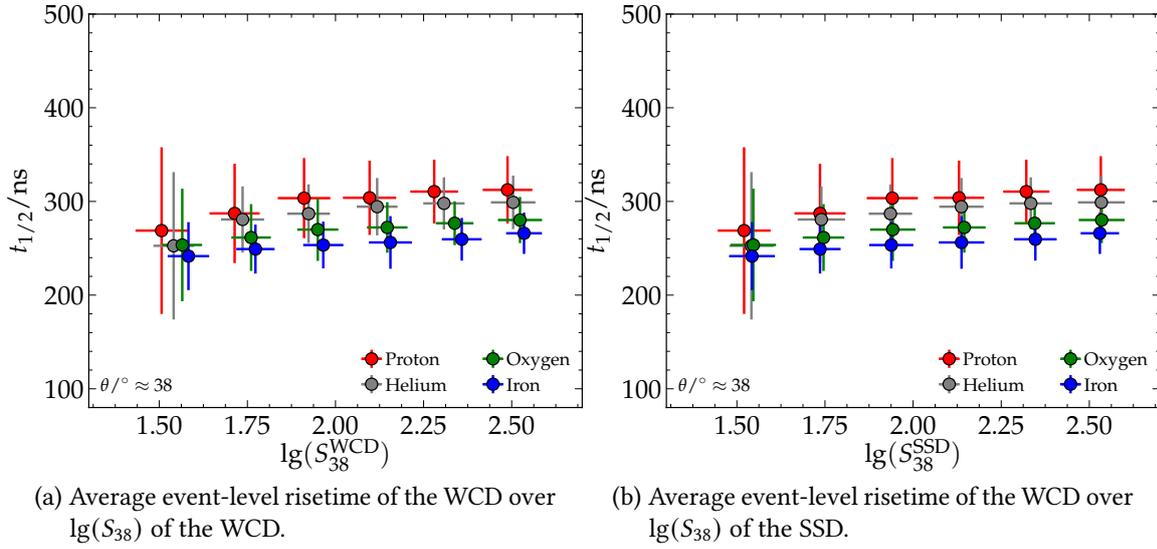


Figure 7.10.: For larger shower sizes the average risetime is increasing slightly. The average risetime varies for different primaries and might thus be a good additional input parameter for the PCA.

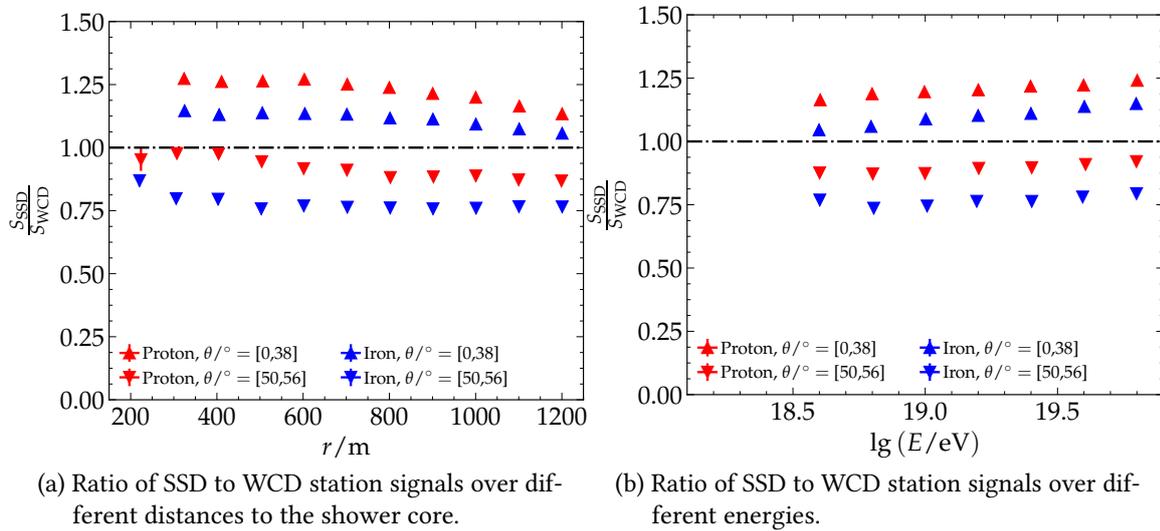


Figure 7.11.: The average ratio of SSD to WCD station signals over energy and distances to the shower core. The zenith ranges are chosen to be equal bins in  $\sec \theta$ . The ratio for different energies and distances results in rather stable values and a good discrimination between different primaries.

## 8. Conclusion

Numerous open questions remain in the field of UHECR physics, such as the cause of the flux oppression at the end of the energy spectrum or the mass composition at the highest energies. Previous studies at the Pierre Auger Observatory have shown a development of the mass composition towards lighter nuclei, which is shifting to a composition towards heavier elements for the highest energies, using the longitudinal information of air showers to estimate the depth of the shower maximum. However, these measurements are heavily limited by the duty cycles of the fluorescence detectors, which can only operate on clear, moonless nights. The surface detector can measure independently with a duty cycle of 100% and can thus acquire a large portion of data in a much smaller time span. In the upgrade of the Pierre Auger Observatory, new SSDs are currently deployed to measure in combination with the WCD detectors the lateral distributions of the air showers on the ground. Both detectors have different sensitivities to individual air shower components, which show different features for different primary particles. It has been shown, that the estimation of mass composition becomes possible, by combining the measurements of the WCD and SSD and applying a principal component analysis.

The main goal of this thesis was to elaborate the method of the PCA by training it on a large Monte Carlo data set and taking further variables into consideration. In a first step of this thesis, the reconstruction of the shower size, estimated by fitting lateral distribution functions to the SSD measurements, has been revisited and modified. The SSD, operating as subordinate detector to the WCD, uses the geometry information, as for example the position of the shower core, that was reconstructed during the fit procedure of the WCD. As it has been shown in this thesis, the uncertainty on the reconstructed position of the shower core translates into uncertainties in the position of each station, relative to the true location of the core. The uncertainties of the core position may therefore influence the results of shower reconstructions of other detector measurements. At distances greater than around 200 m from the shower axis, a value that depends on shower size and zenith angle, the uncertainties of distances of individual stations can be approximated with a Gaussian distribution with a standard deviation equal to the core resolution. Under this approximation, the core uncertainty can be propagated into an uncertainty in the signal of a subordinate detector via the estimated LDF slope. This uncertainty may then be added in quadrature with the sampling fluctuations.

A few single outliers, that had way too large predicted SSD shower size have been investigated as well. For the simulations, these have been identified to originate from late, large contributions in the signal traces of the SSD, that come from single injected hadrons, such as protons. Other studies on such sub-luminal pulses have been already done by John Linsley with data from the Volcano Ranch. However, it could not be confirmed, whether these late signals are just an artifact of the shower simulations, or are to be found in the measured data of the Pierre Auger Observatory as well, which would open up a window

for various studies of the hadronic shower component.

In a second step, the signal responses of the WCD and SSD have been analyzed in order to gain a better understanding of the signal of the muonic shower component in the detectors. An idealized model of the muonic signal in the WCD and SSD has been developed and validated with the simulation of a combined detector setup and its response to a flux of muons under different zenith angles. It could be shown, that the average arrival direction of muons differs from the actual reconstructed shower angle, resulting in a deviation of the estimated signal deposited per particle in the detector. Another deviation of the estimated signal per particle is only visible in the WCD. This is caused by the zenith dependent energy spectrum of the muons, leading to an observed signal per particle, which is smaller than the predictions at low zenith angles.

A previous implementation of the PCA, to estimate energy and mass with the combination of WCD and SSD measurements, has been revisited and elaborated during the course of this thesis. It has been shown, that the PCA gives almost unbiased estimates of the logarithmic mass and energy as a function of energy, but at the cost of large uncertainties, which make an event-level estimation of the mass impossible at the current state. Biases of  $\ln A$  and  $\lg E$  over  $\sin^2 \theta$  were studied and an extension of the linear equations from the PCA by adding the zenith angle has been investigated. This reduced the biases of  $\ln A$  over  $\sin^2 \theta$  and improved the mass resolution slightly.

In the last step of this thesis, a first estimate of the mass composition, using almost 2 years of data from the PPA, has been given. However, large systematic errors and the limited resolution do not yet allow currently for any meaningful interpretation of the results. More work towards better refined and additional input parameters is needed, to further increase the PCA resolution. Minimizing the bias of the parameters, used for the mass estimation, can significantly decrease the systematic uncertainties.

# A. Appendix

## A.1. Primary and Model Comparison for Propagated Core Uncertainties

Using the full fixed library, given in Table A.1, the propagation of core uncertainties is applied as well to the reconstruction. The resulting biases and resolutions, using the MAD and standard deviation, are given in Figs. A.1 to A.3. There are some differences in the bias between the reconstruction of proton and iron showers. In the lowest energy bin of  $\lg(E/eV) = 18.5$ , the bias of reconstructed iron showers is up to a factor of 2 smaller than for proton showers. Between the different models, QGSJET-II.04 has in the lowest energy bins a larger bias than EPOS-LHC. In the lower zenith range, no notable differences between the models and primaries for the MAD and  $\sigma$  can be seen.

Table A.1.: Fixed library of CORSIKA simulations with different combination of parameters. For each combination, 120 events were reconstructed.

primary	proton, iron
hadronic interaction model	EPOS-LHC, QGSJET-II.04
$\lg(E/eV)$	18.5, 19.0, 19.5, 20.0
$\theta/^\circ$	0, 12, 22, 32, 38, 48, 56, (65)

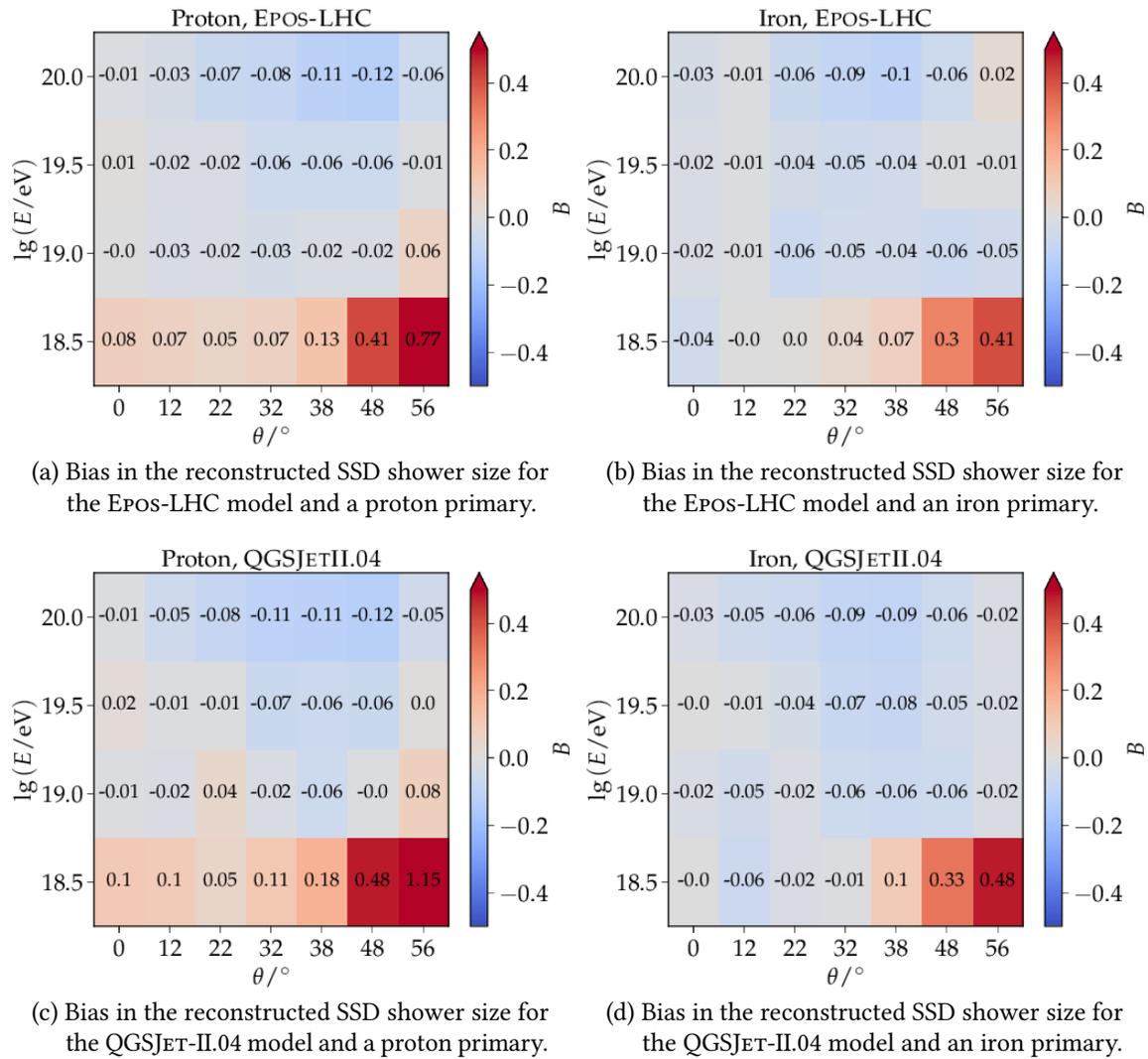


Figure A.1.: Assessment of the bias in the reconstructed SSD shower size with the propagation of core uncertainties for different primaries and models.

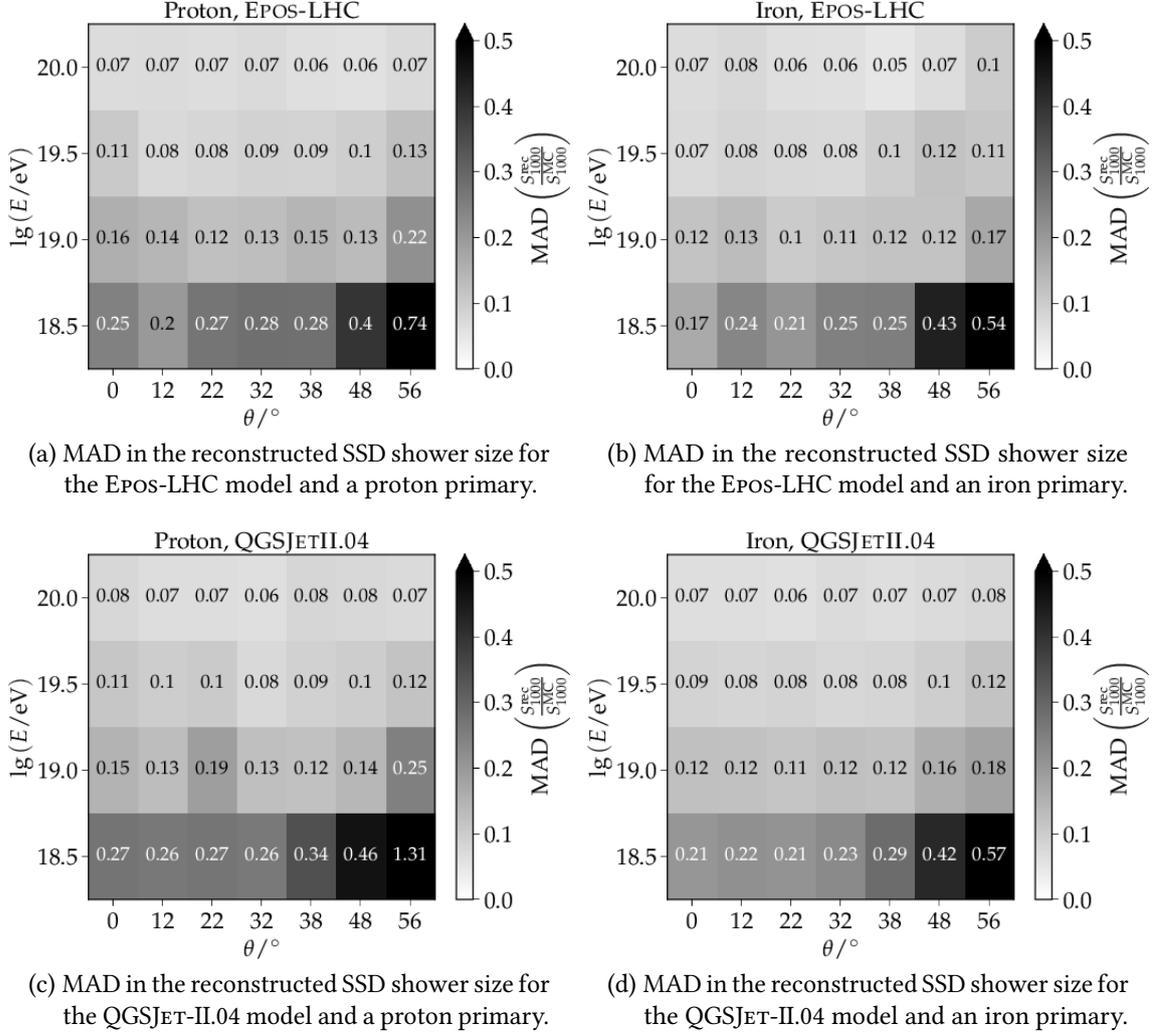


Figure A.2.: Comparison of the resolutions (calculated using the mean absolute deviation of relative errors) of the reconstructed SSD shower size with propagation of core uncertainties for different zenith angles and energy bins. The mean absolute deviation is less influenced by large outliers than the standard deviation and thus serves as a good estimate of the resolution for the bulk of events. The values are smaller than for the standard deviation, but still show a sizable improvement with the propagation of core uncertainties.

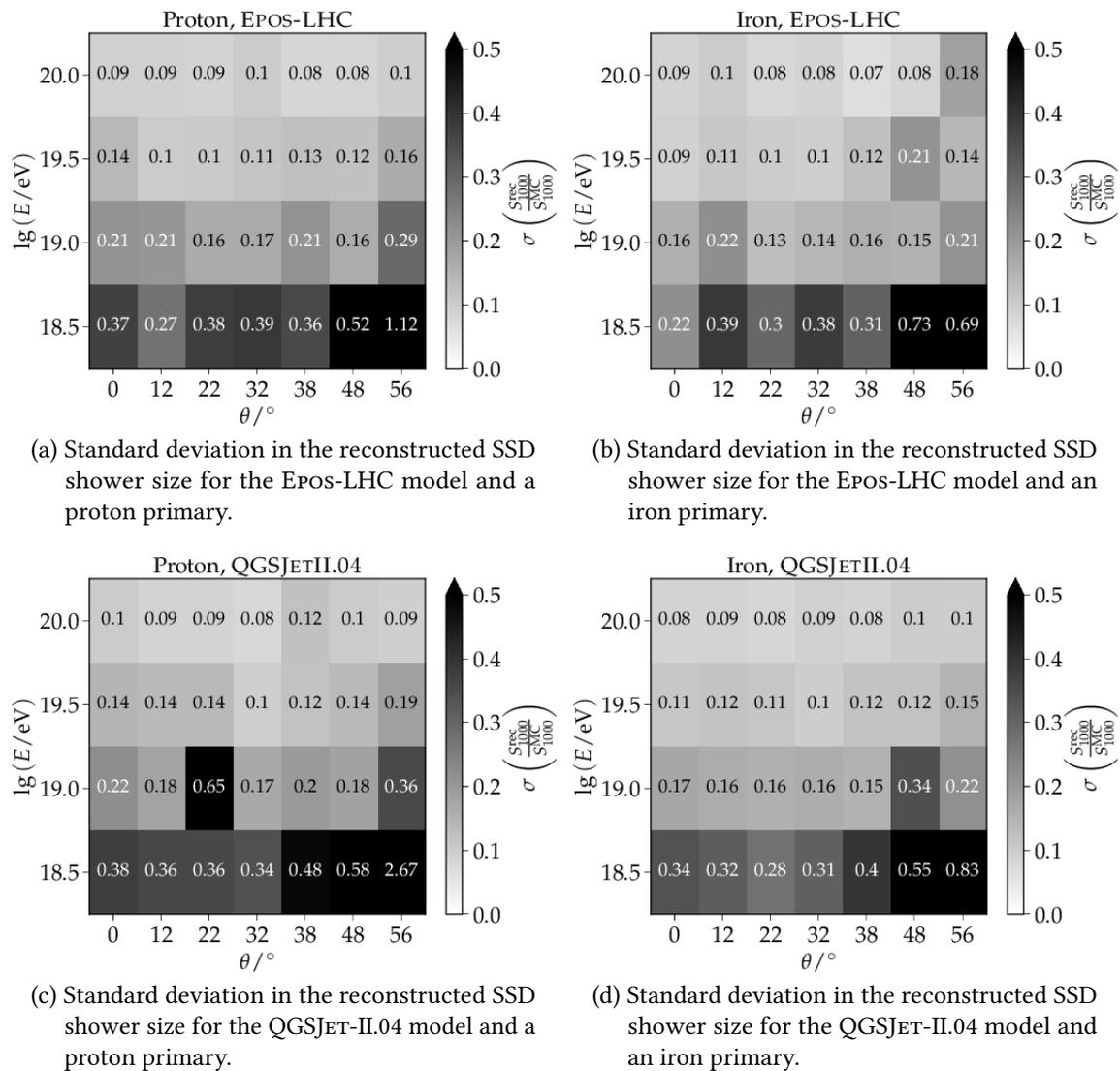


Figure A.3.: Comparison of the resolutions (calculated using the standard deviation of relative errors) of the reconstructed SSD shower size with propagation of core uncertainties for different zenith angles and energy bins. At the energy of  $\lg(E/\text{eV}) = 18.5$ , the SD array is at the lower limit of full efficiency, where the SSD is not expected to perform well.

## A.2. Detector Signals from the Electromagnetic Component

The electromagnetic shower component, consisting of electrons, positrons and photons, shows no significant dependency on the signal per theoretical tracklength over different zenith angles. This can be attributed due to the kinetic energy of the arriving particles, as previously discussed in Section 5.5. For the SSD, the energy of electrons and positrons is large enough to traverse the full SSD. However, for the WCD, the energy is not large enough and the particles will deposit their full energy in the detector. The signal per particle  $\tilde{S}$  in the SSD is only dependent on the zenith angle of the arriving shower. In Fig. A.4, a zoomed out version of Fig. 5.15a for  $\tilde{S}$  over different zenith angles can be seen. At lower shower zenith angles, the average arrival angle of individual particles in the detectors is larger and therefore, the increase of  $\tilde{S}$  over  $\sec \theta$  is lower than expected, as shown in Section 5.3. The photons have a relatively small cross section compared to the electrons and positrons and deposit almost no energy in the detectors, leading to a smaller signal per particle. The numbers of photons arriving at the detectors exceed the number of electrons and positrons by almost a factor of 1000. Fig. A.5a shows the number of electrons, positrons and photons, traversing the SSD and WCD. Due to the geometry and size of the detectors, more particles traverse the WCD than the SSD, similar to Fig. 5.4 shown in Section 5.2. In Fig. A.5b the total signal, as well as the contributions of electrons, positrons and photons to the signal in the detectors is shown. Although their signal per particle is very low, the contribution of the photons to the total electromagnetic signal, measured in the WCD, exceeds the contribution of the electrons and positrons by over a factor 10, due to the large number of particles. As for the SSD, electrons and positrons have only a slightly larger contribution to the total signal than the photons.

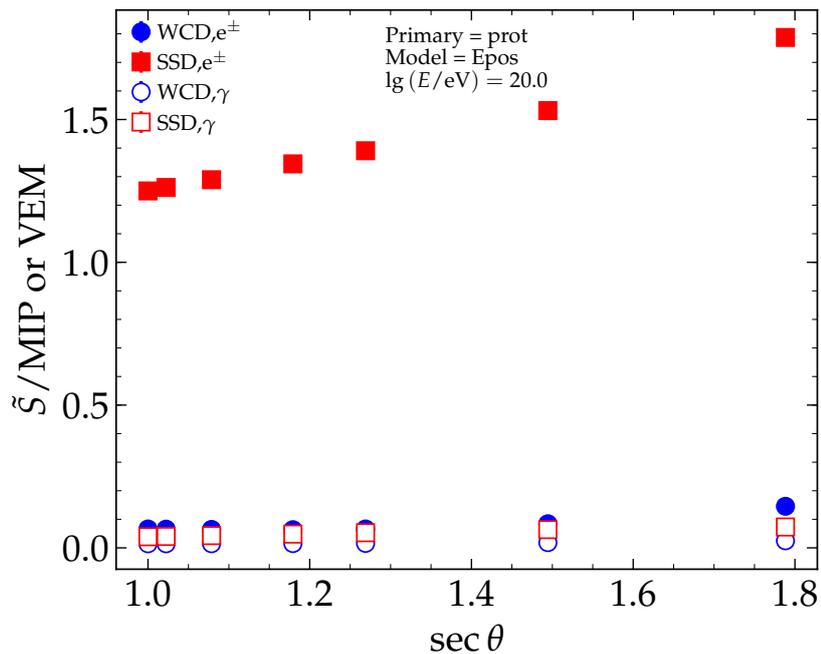
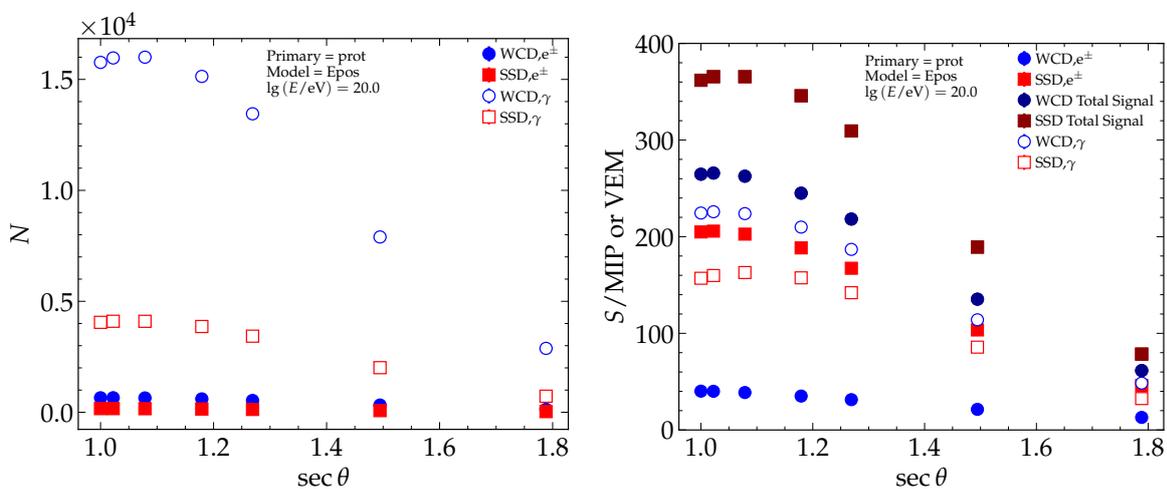


Figure A.4.: The electromagnetic signal per particle  $\tilde{S}$  for photons and electrons and positrons. Electrons and positrons have the largest  $\tilde{S}$  in the SSD, since they traverse the detector completely, contrary to the WCD, where they will deposit all their remaining energy. Photons are rarely interacting in both detectors due to their small cross section and thus have a very small  $\tilde{S}$ .



(a) The particle hits of electrons and positrons (filled markers) and photons (hollow markers) are decreasing over  $\text{sec } \theta$  due to the detector geometry.

(b) Total signal and the individual contributions to it of electrons, positrons and photons for the WCD and SSD.

Figure A.5.: The number of photons, arriving at the detectors, is up to a factor of 1000 larger than electrons and positrons. Although photons rarely interact in the detectors, due to their large number they still contribute the most to the electromagnetic signal in the WCD.

### A.3. Extension of the PCA for the Energy Estimation

As for  $\ln A_{\text{PCA}}$ , the parameters  $A_1$ ,  $B_1$  and  $C_1$  of Eq. (6.17) are written up as fourth order polynomials

$$f(\theta) = f_1 + f_2 \sec \theta + f_3 \sec^2 \theta + f_4 \sec^3 \theta + f_5 \sec^4 \theta. \quad (\text{A.1})$$

The input parameters of the PCA are split up in equal bins of  $\sec \theta$  and values for  $A_1$ ,  $B_1$  and  $C_1$  for different  $\sec \theta$  values are retrieved. In Fig. A.6 the different parameters, as well as a fitted curve are shown for the  $\sec \theta$  bins. Similar to the parameters of  $\ln A_{\text{PCA}}$  the first two parameters are very similar to each other with a different sign.

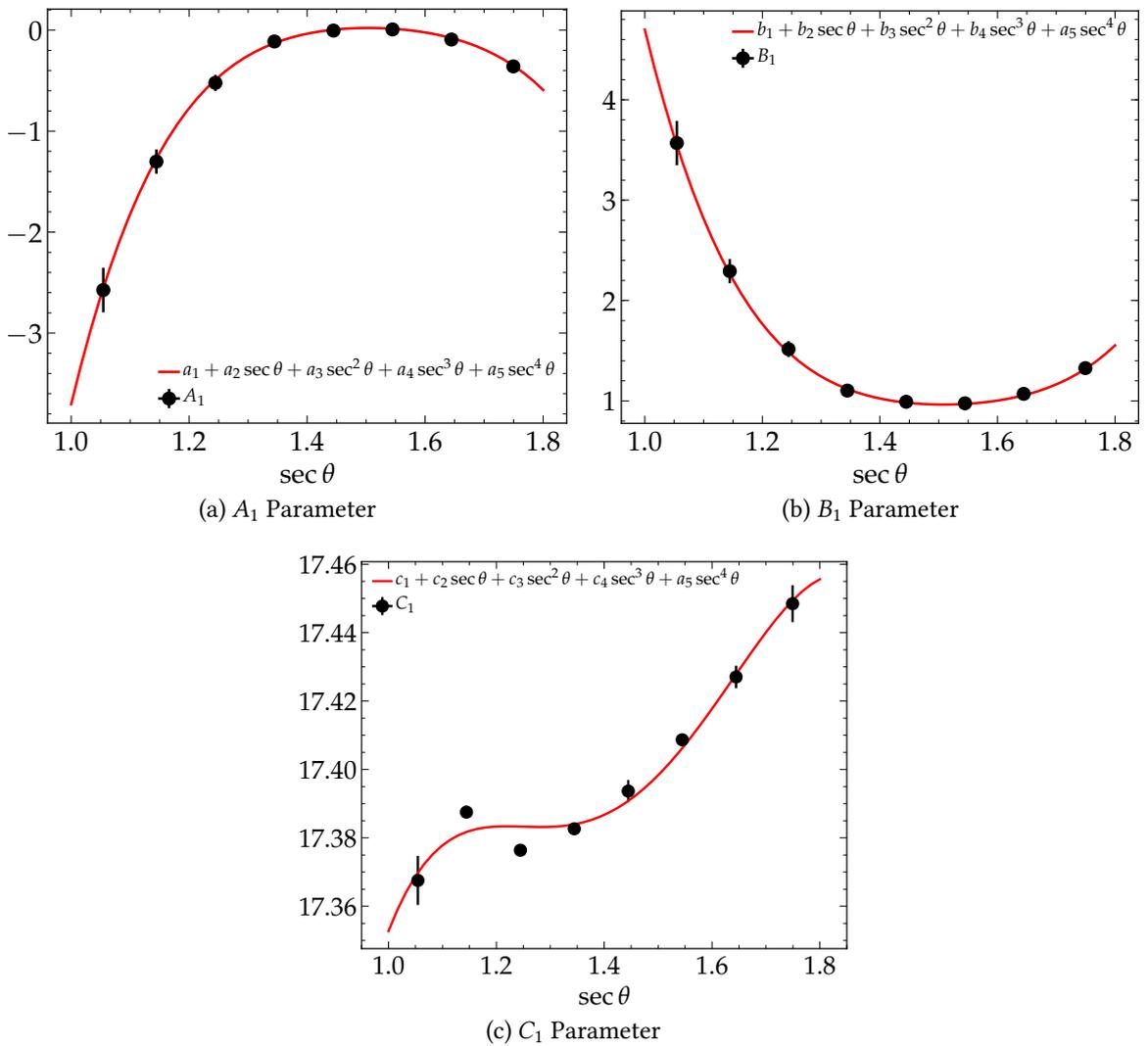


Figure A.6.: The PCA is trained on different bins in  $\sec \theta$  and the resulting  $A_1$ ,  $B_1$  and  $C_1$  parameters are fitted to a fourth-order polynomial.

## A.4. Histograms of the Mass Estimation

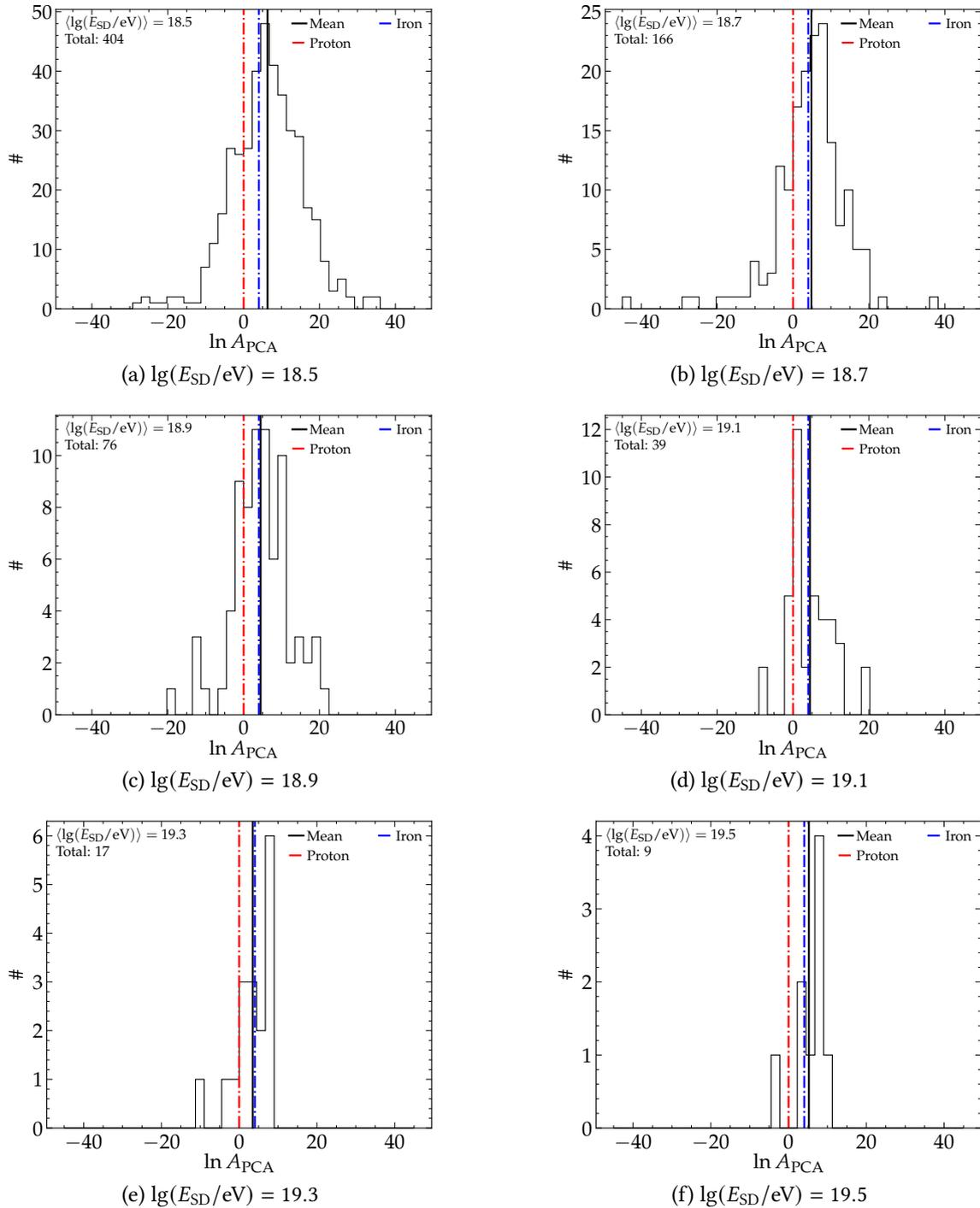


Figure A.7.: Distribution on  $\ln A_{PCA}$  for different energy bins from Fig. 7.7. Proton and iron prediction are given as red and blue dashdotted lines.

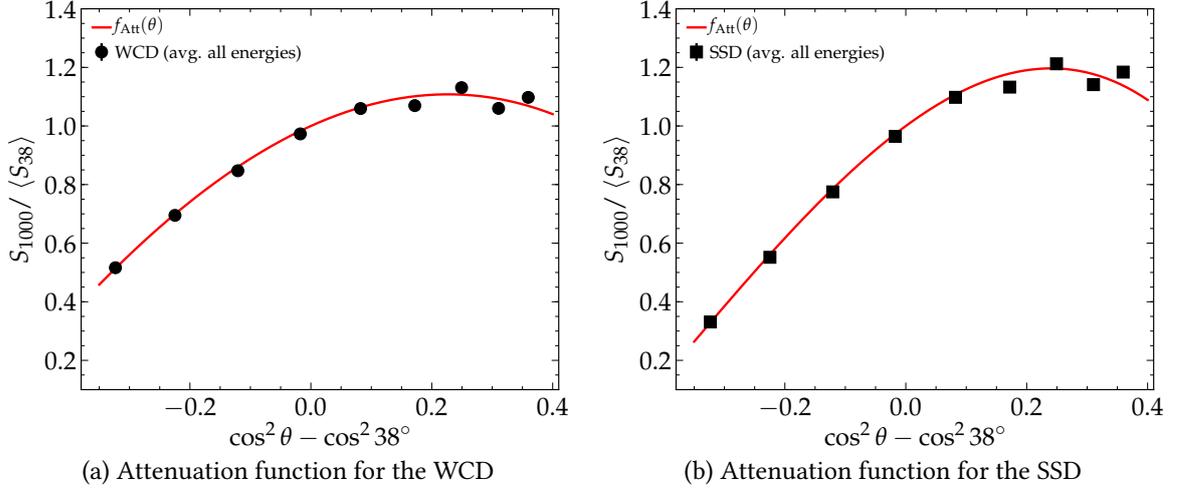


Figure A.8.: Fit of the attenuation correction for the WCD and SSD. The shower sizes of all energies are normed to the shower size at a zenith angle of  $38^\circ$ .

Table A.2.: Parameters of the fit of the attenuation function  $f_{\text{Att}}$  for the WCD and SSD

Detector	$a$	$b$	$c$	$\chi^2$
WCD	$0.9302 \pm 0.0016$	$-1.9070 \pm 0.0026$	$-0.4148 \pm 0.1885$	21.51
SSD	$1.5156 \pm 0.0014$	$-2.4035 \pm 0.0027$	$-2.0735 \pm 0.1600$	30.62

## A.5. Mass Estimation for Different Hadronic Interaction Models

The mass estimation in Chapter 7 was done, using a PCA trained on a data set that made use of simulations performed with EPOS-LHC. For the different hadronic interaction model QGSJET-II.04 the full training of the PCA, as described in Chapter 6, is repeated.

### Attenuation Correction of $S(1000)$

As first step, the attenuation correction of  $S(1000)$  is performed for the WCD and SSD, using Eq. (6.2) as attenuation function. In Fig. A.8, the resulting fits are shown and the fit parameters are given in Table A.2. For both models, the fit parameters show only small differences. The estimated shower size is then converted to  $S_{38}$ , using the attenuation correction.

### PCA Training

In the next step, the QGSJET-II.04 data set is binned in  $\sec \theta$  and the PCA is trained each bin, using  $S_{38}$  of the WCD and SSD, as well as the logarithmic shower energy and primary mass as input parameters. The resulting six parameters  $A_{1,2}$ ,  $B_{1,2}$  and  $C_{1,2}$  of the equations

Table A.3.: Fit values of the zenith dependent  $A_1$ ,  $B_1$  and  $C_1$  parameters of  $\lg E_{\text{PCA}}$ .

	$A_1(\theta)$		$B_1(\theta)$		$C_1(\theta)$	
$f_1$	-181.6144	$\pm 51.2579$	179.8491	$\pm 50.8242$	13.29	$\pm 2.8069$
$f_2$	477.7149	$\pm 146.3077$	-469.6408	$\pm 144.9893$	11.4831	$\pm 8.1286$
$f_3$	-475.611	$\pm 155.5636$	466.809	$\pm 154.0735$	-11.8879	$\pm 8.768$
$f_4$	212.6499	$\pm 73.0543$	-208.4082	$\pm 72.3117$	5.4132	$\pm 4.1753$
$f_5$	-36.0192	$\pm 12.7896$	35.2555	$\pm 12.6519$	-0.9088	$\pm 0.7407$
$\chi^2$	3.54		2.17		17.0	

Table A.4.: Fit values of the zenith dependent  $A_2$ ,  $B_2$  and  $C_2$  parameters of  $\ln A_{\text{PCA}}$ .

	$A_2(\theta)$		$B_2(\theta)$		$C_2(\theta)$	
$f_1$	3 302.1061	$\pm 1 309.9795$	-3 171.8849	$\pm 1 270.1232$	8.3066	$\pm 98.0208$
$f_2$	-8 737.0412	$\pm 3 801.3544$	8 351.108	$\pm 3 683.3545$	-4.5062	$\pm 285.2691$
$f_3$	9 002.4676	$\pm 4 099.9909$	-8 570.2464	$\pm 3 970.8575$	-23.1572	$\pm 308.9855$
$f_4$	-4 187.7266	$\pm 1 948.9102$	3 974.8343	$\pm 1 886.9287$	24.6071	$\pm 147.6099$
$f_5$	738.3757	$\pm 344.6617$	-699.4187	$\pm 333.6401$	-6.9833	$\pm 26.2438$
$\chi^2$	3.37		5.17		19.13	

for  $\lg E$  and  $\ln A$  are then fitted to a fourth-order polynomial as in Eq. (6.27). The fit parameters  $f_i$  for both equations are listed in Tables A.3 and A.4 and Fig. A.9 shows the fits of all parameters. All parameters are strongly correlated and thus have a large error. A systematic error is estimated by adding a bias of 5% to the SSD and WCD shower size. The resulting bias is of the same order as the bias for EPOS-LHC

$$\sigma_{\text{sys}}(\text{tot}) = \pm 3.36. \quad (\text{A.2})$$

## Mass Estimation

The mass estimation of the data set from Chapter 7 is repeated, using Eqs. (6.28) and (6.29) with the parameters from Table A.4. In Fig. A.10 the estimated logarithmic mass is shown as function of the estimated energy, using either the SD or the PCA estimate. The results are similar to the values obtained using a PCA trained on EPOS-LHC data. At the lowest energy bin of  $\lg(E/\text{eV}) = 18.5$  for the binning in SD energy, the mass estimate is deviating the most with a difference of  $\delta \ln A_{\text{PCA}} = 0.38$  between both models.

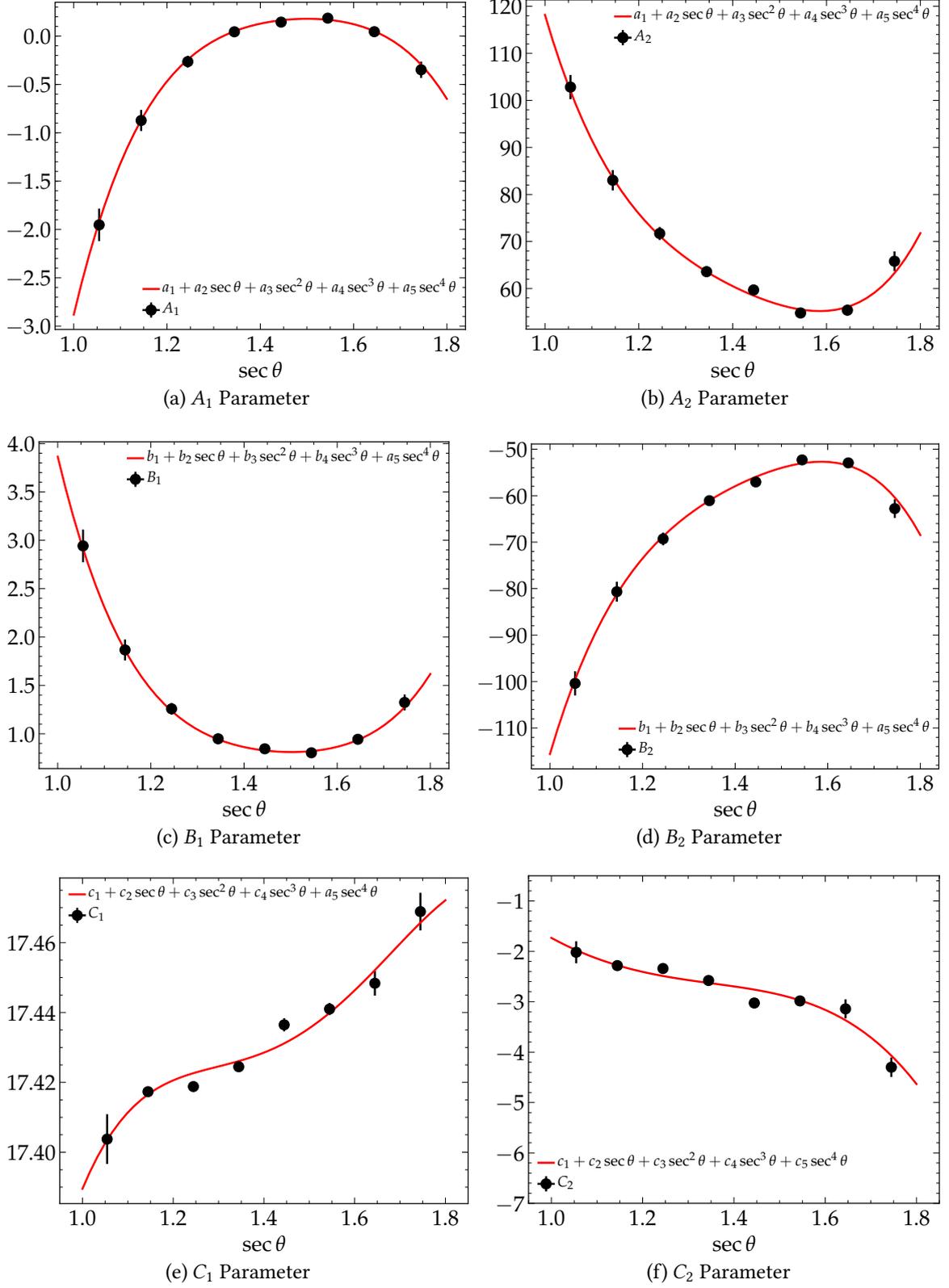


Figure A.9.: The PCA is trained on different bins in  $\sec \theta$  and the resulting  $A_{1,2}$ ,  $B_{1,2}$  and  $C_{1,2}$  parameters are fitted to a fourth-order polynomial.

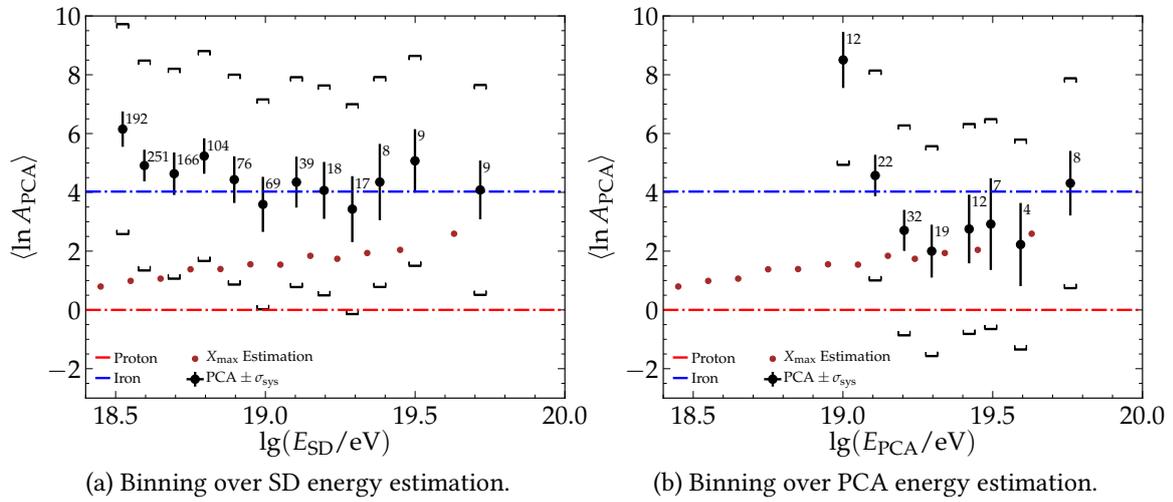


Figure A.10.: Estimation of the mass composition as a function of the estimated logarithmic energy. The numbers denote the amount of data points in each bin. The estimations from  $X_{\text{max}}$  were computed, using the data from [27] with the parameterizations from [59, 60].

# Acknowledgements

First, I would like to express my gratitude towards Prof. Ralph Engel for enabling me to write my thesis at IAP and reviewing it, as well as providing me with valuable, additional input during the last weeks of this work. I also would like to thank Prof. Guido Drexlin for acting as the second reviewer and showing interest in this thesis.

My biggest thank goes to my two advisors Dr. David Schmidt and Dr. Markus Roth, who have advised me for the entire year of the thesis and always provided useful feedback, as well as proofread this thesis. Additional thank goes to Dr. Darko Veberič and again Dr. David Schmidt for giving me a great help in learning to work with the `Offline` framework. I would like to thank Dr. Doris Wochele, for providing me useful information about the computer cluster at the IAP, as well as providing me with nice statistics to my simulations. My sincere gratitude to all of my colleagues at Campus North from the Pierre Auger Collaboration for all the support.

Great thanks goes to Fabian Kreß and Alexander Böhmländer for additional proofreading of this thesis and providing me with useful tips for `LaTeX`.

Finally, I am very grateful to my family who supported me during all this time.



# Bibliography

- [1] V. F. Hess. “Über Beobachtungen der durchdringenden Strahlung bei sieben Freiballonfahrten”. In: *Physikalische Zeitschrift* 13 (Nov. 1912), pp. 1084–1091.
- [2] P. Auger et al. “Extensive cosmic ray showers”. In: *Rev. Mod. Phys.* 11 (1939), pp. 288–291. DOI: 10.1103/RevModPhys.11.288.
- [3] K. Greisen. “End to the cosmic ray spectrum?” In: *Phys. Rev. Lett.* 16 (1966), pp. 748–750. DOI: 10.1103/PhysRevLett.16.748.
- [4] G. T. Zatsepin and V. A. Kuzmin. “Upper limit of the spectrum of cosmic rays”. In: *JETP Lett.* 4 (1966). [Pisma Zh. Eksp. Teor. Fiz.4,114(1966)], pp. 78–80.
- [5] J. Abraham et al. “Measurement of the energy spectrum of cosmic rays above  $10^{18}$  eV using the Pierre Auger Observatory”. In: *Phys. Lett. B* 685 (2010), pp. 239–246. DOI: 10.1016/j.physletb.2010.02.013. arXiv: 1002.1975 [astro-ph.HE].
- [6] W. Heitler. “The quantum theory of radiation”. In: The international series of monographs on physics (1954).
- [7] J. Matthews. “A Heitler model of extensive air showers”. In: *Astropart. Phys.* 22 (2005), pp. 387–397. DOI: 10.1016/j.astropartphys.2004.09.003.
- [8] F. G. Schröder et al. “High-Energy Galactic Cosmic Rays (Astro2020 Science White Paper)”. In: *Bull. Am. Astron. Soc.* 51 (Mar. 2019), p. 131. arXiv: 1903.07713 [astro-ph.HE].
- [9] T. Antoni et al. “KASCADE measurements of energy spectra for elemental groups of cosmic rays: Results and open problems”. In: *Astropart. Phys.* 24 (2005), pp. 1–25. DOI: 10.1016/j.astropartphys.2005.04.001. arXiv: astro-ph/0505413.
- [10] W. D. Apel et al. “Kneelike structure in the spectrum of the heavy component of cosmic rays observed with KASCADE-Grande”. In: *Phys. Rev. Lett.* 107 (2011), p. 171104. DOI: 10.1103/PhysRevLett.107.171104. arXiv: 1107.5885 [astro-ph.HE].
- [11] V. Berezhinsky, A. Z. Gazizov, and S. I. Grigorieva. “On astrophysical solution to ultrahigh-energy cosmic rays”. In: *Phys. Rev. D* 74 (2006), p. 043005. DOI: 10.1103/PhysRevD.74.043005. arXiv: hep-ph/0204357.
- [12] V. Berezhinsky, A. Z. Gazizov, and S. I. Grigorieva. “Dip in UHECR spectrum as signature of proton interaction with CMB”. In: *Phys. Lett. B* 612 (2005), pp. 147–153. DOI: 10.1016/j.physletb.2005.02.058. arXiv: astro-ph/0502550.
- [13] A. Aab et al. “The Pierre Auger Observatory Upgrade - Preliminary Design Report”. In: (2016). arXiv: 1604.03637 [astro-ph.IM].
- [14] D. Allard. “Extragalactic propagation of ultrahigh energy cosmic-rays”. In: *Astropart. Phys.* 39–40 (2012), pp. 33–43. DOI: 10.1016/j.astropartphys.2011.10.011. arXiv: 1111.3290 [astro-ph.HE].

- [15] D. Hooper and A. M. Taylor. “On The Heavy Chemical Composition of the Ultra-High Energy Cosmic Rays”. In: *Astropart. Phys.* 33 (2010), pp. 151–159. DOI: 10.1016/j.astropartphys.2010.01.003. arXiv: 0910.1842 [astro-ph.HE].
- [16] A. Aab et al. “Depth of maximum of air-shower profiles at the Pierre Auger Observatory. I. Measurements at energies above  $10^{17.8}$  eV”. In: *Phys. Rev. D* 90.12 (2014), p. 122005. DOI: 10.1103/PhysRevD.90.122005. arXiv: 1409.4809 [astro-ph.HE].
- [17] A. Aab et al. “Depth of maximum of air-shower profiles at the Pierre Auger Observatory. II. Composition implications”. In: *Phys. Rev. D* 90.12 (2014), p. 122006. DOI: 10.1103/PhysRevD.90.122006. arXiv: 1409.5083 [astro-ph.HE].
- [18] A. Aab et al. “The Pierre Auger Cosmic Ray Observatory”. In: *Nucl. Instrum. Meth. A* 798 (2015), pp. 172–213. DOI: 10.1016/j.nima.2015.06.058. arXiv: 1502.01323 [astro-ph.IM].
- [19] I. Allekotte et al. “The Surface Detector System of the Pierre Auger Observatory”. In: *Nucl. Instrum. Meth. A* 586 (2008), pp. 409–420. DOI: 10.1016/j.nima.2007.12.016. arXiv: 0712.2832 [astro-ph].
- [20] J. Abraham et al. “The Fluorescence Detector of the Pierre Auger Observatory”. In: *Nucl. Instrum. Meth. A* 620 (2010), pp. 227–251. DOI: 10.1016/j.nima.2010.04.023. arXiv: 0907.4282 [astro-ph.IM].
- [21] J. Abraham et al. “Trigger and Aperture of the Surface Detector Array of the Pierre Auger Observatory”. In: *Nucl. Instrum. Meth. A* 613 (2010), pp. 29–39. DOI: 10.1016/j.nima.2009.11.018. arXiv: 1111.6764 [astro-ph.IM].
- [22] B. G. Keilhauer. “Investigation of atmospheric effects on the development of extensive air showers and their detection with the Pierre Auger Observatory”. 51.04.02; LK 01. PhD thesis. 2004. DOI: 10.5445/IR/4512003.
- [23] H. O. Klages. “HEAT – Enhancement Telescopes for the Pierre Auger Southern Observatory”. In: *30th International Cosmic Ray Conference*. July 2007.
- [24] A. Aab et al. “Measurement of the cosmic-ray energy spectrum above  $2.5 \times 10^{18}$  eV using the Pierre Auger Observatory”. In: *Phys. Rev. D* 102.6 (2020), p. 062005. DOI: 10.1103/PhysRevD.102.062005. arXiv: 2008.06486 [astro-ph.HE].
- [25] A. Castellina. “Highlights from the Pierre Auger Observatory”. In: *PoS ICRC2019* (2021), p. 004. DOI: 10.22323/1.358.0004. arXiv: 1909.10791 [astro-ph.HE].
- [26] I. Valino. “The flux of ultra-high energy cosmic rays after ten years of operation of the Pierre Auger Observatory”. In: *PoS ICRC2015* (2016), p. 271. DOI: 10.22323/1.236.0271.
- [27] J. Bellido. “Depth of maximum of air-shower profiles at the Pierre Auger Observatory: Measurements above  $10^{17.2}$  eV and Composition Implications”. In: *PoS ICRC2017* (2018). Ed. by Darko Veberic, p. 506. DOI: 10.22323/1.301.0506.
- [28] A. Aab et al. “Observation of a Large-scale Anisotropy in the Arrival Directions of Cosmic Rays above  $8 \times 10^{18}$  eV”. In: *Science* 357.6537 (2017), pp. 1266–1270. DOI: 10.1126/science.aan4338. arXiv: 1709.07321 [astro-ph.HE].

- 
- [29] A. Aab et al. “Muons in Air Showers at the Pierre Auger Observatory: Mean Number in Highly Inclined Events”. In: *Phys. Rev. D* 91.3 (2015). [Erratum: *Phys.Rev.D* 91, 059901 (2015)], p. 032003. DOI: 10.1103/PhysRevD.91.032003. arXiv: 1408.1421 [astro-ph.HE].
- [30] R. Šmída. “Scintillator detectors of AugerPrime Scintillator detectors of AugerPrime”. In: *The Pierre Auger Observatory: Contributions to the 35th International Cosmic Ray Conference (ICRC 2017)*. 2017, pp. 169–176. URL: [http://inspirehep.net/record/1618433/files/1617990\\_169-176.pdf](http://inspirehep.net/record/1618433/files/1617990_169-176.pdf).
- [31] A. Taboada. “Analysis of the First Data of the AugerPrime Detector Upgrade”. GAP-2020-002. PhD thesis. KIT, 2020.
- [32] A. Castellina. “The dynamic range of the AugerPrime Surface Detector: technical solution and physics reach”. In: *PoS ICRC2017* (2018). Ed. by Darko Veberic, p. 397. DOI: 10.22323/1.301.0397.
- [33] D. F. Nitz. “New Electronics for the Surface Detectors of the Pierre Auger Observatory”. In: *PoS ICRC2019* (2021), p. 370. DOI: 10.22323/1.358.0370.
- [34] F. Suarez. “The AMIGA muon detectors of the Pierre Auger Observatory: overview and status”. In: *33rd International Cosmic Ray Conference*. 2013.
- [35] A. M. Botti. “The AMIGA underground muon detector of the Pierre Auger Observatory - performance and event reconstruction”. In: *PoS ICRC2019* (2021), p. 202. DOI: 10.22323/1.358.0202.
- [36] B. Pont. “A Large Radio Detector at the Pierre Auger Observatory - Measuring the Properties of Cosmic Rays up to the Highest Energies”. In: *PoS ICRC2019* (2021), p. 395. DOI: 10.22323/1.358.0395.
- [37] A. Aab et al. “Reconstruction of Events Recorded with the Surface Detector of the Pierre Auger Observatory”. In: *JINST* 15.10 (2020), P10021. DOI: 10.1088/1748-0221/15/10/P10021. arXiv: 2007.09035 [astro-ph.IM].
- [38] D. Veberič and M. Roth. “Offline Reference Manual: SD Reconstruction”. Auger internal note GAP-2005-035. 2005.
- [39] A. Schulz. “Measurement of the Energy Spectrum and Mass Composition of Ultra-high Energy Cosmic Rays”. PhD thesis. University Karlsruhe, 2016.
- [40] D. Schmidt. “Sensitivity of AugerPrime to the masses of ultra-high-energy cosmic rays”. GAP-2019-030. PhD thesis. KIT, 2019.
- [41] D. Newton, J. Knapp, and A. A. Watson. “The Optimum Distance at which to Determine the Size of a Giant Air Shower”. In: *Astropart. Phys.* 26 (2007), pp. 414–419. DOI: 10.1016/j.astropartphys.2006.08.003. arXiv: astro-ph/0608118 [astro-ph].
- [42] P. Billoir, A. Letessier-Selvon, and C. Macolino. “Multi event fit for the LDF”. Auger collaboration meeting. 2011.
- [43] A. Castellina. “Accuracy of signal measurements in the Auger Surface Detector”. Auger internal note GAP-2003-031. 2003.

- [44] M. Ave et al. “The accuracy of signal measurement with the water Cherenkov detectors of the Pierre Auger Observatory”. In: *Nucl. Instrum. Meth. A* 578 (2007), pp. 180–184. DOI: 10.1016/j.nima.2007.05.150.
- [45] R. Hiller and M. Roth. “An update on the signal accuracy using the infill array”. Auger internal note GAP-2012-012. 2012.
- [46] L. M. Bueno, P. Billoir, and I. C. Mariş. “Signal variance for the TOTd and MoPS triggers”. Auger internal note GAP-2014-035. 2014.
- [47] T. Schulz. “Enhanced UHECR Event Reconstruction by means of Sampling Lateral Distributions with Multiple Surface Sub-Detectors”. Bachelor’s thesis. KIT, 2017.
- [48] S. Argiro et al. “The Offline Software Framework of the Pierre Auger Observatory”. In: *Nucl. Instrum. Meth. A* 580 (2007), pp. 1485–1496. DOI: 10.1016/j.nima.2007.07.010. arXiv: 0707.1652 [astro-ph].
- [49] J. Linsley. “Subluminal pulses from cosmic ray air showers”. In: *J. Phys. G* 10 (1984), pp. L191–L195. DOI: 10.1088/0305-4616/10/8/005.
- [50] B. Kégl and D. Veberič. “Single Muon Response: Tracklength”. Auger internal note GAP-2009-043. 2009.
- [51] S. M. Seltzer and M. J. Berger. “Bremsstrahlung spectra from electron interactions with screened atomic nuclei and orbital electrons”. In: *Nucl. Instr. Meth. B* 12 95 (1985).
- [52] P. A. Zyla et al. “Review of Particle Physics”. In: *PTEP* 2020.8 (2020), p. 083C01. DOI: 10.1093/ptep/ptaa104.
- [53] M. J. Berger et al. *ESTAR, PSTAR, and ASTAR: Computer Programs for Calculating Stopping-Power and Range Tables for Electrons, Protons, and Helium Ions (version 1.2.3)*. 2005. URL: <http://physics.nist.gov/Star>.
- [54] F. Pedregosa et al. “Scikit-learn: Machine Learning in Python”. In: *J. Machine Learning Res.* 12 (2011), pp. 2825–2830. arXiv: 1201.0490 [cs.LG].
- [55] Z. Zong. “First results from the AugerPrime engineering array”. In: *PoS ICRC2017* (2018), p. 449. DOI: 10.22323/1.301.0449.
- [56] C. Bonifazi and A. Letessier-Selvon. “Eventselection using the T5 time distribution”. Auger internal note GAP-2006-042. 2006.
- [57] C. Bonifazi and P. L. Ghia. “Selection of data periods and calculation of the SD geometrical acceptance”. Auger internal note GAP-2006-101. 2006.
- [58] D. Mockler. “Measurement of the Cosmic Ray Spectrum with the Pierre Auger Observatory”. PhD thesis. Karlsruhe Institut für Technologie (KIT), 2019. 173 pp. DOI: 10.5445/IR/1000089486.
- [59] P. Abreu et al. “Interpretation of the Depths of Maximum of Extensive Air Showers Measured by the Pierre Auger Observatory”. In: *JCAP* 02 (2013), p. 026. DOI: 10.1088/1475-7516/2013/02/026. arXiv: 1301.6637 [astro-ph.HE].
- [60] S. Petrera. “ $X_{\max}$  parameterizations for post-LHC hadronic models”. Auger internal note GAP-2020-058. 2020.

- 
- [61] M. Ave et al. “Mass composition of cosmic rays in the range  $2 \times 10^{17}$  eV -  $3 \times 10^{18}$  eV measured with Haverah Park array”. In: *Astropart. Phys.* 19 (2003), pp. 61–75. DOI: 10.1016/S0927-6505(02)00189-5. arXiv: astro-ph/0203150.
- [62] B. Smith, C. Wileman, and A. Watson. “Can the risetime be used to infer the position of shower maximum”. Auger internal note GAP-2007-092. 2007.
- [63] J. Abraham et al. “Upper limit on the cosmic-ray photon flux above  $10^{19}$  eV using the surface detector of the Pierre Auger Observatory”. In: *Astropart. Phys.* 29 (2008), pp. 243–256. DOI: 10.1016/j.astropartphys.2008.01.003. arXiv: 0712.1147 [astro-ph].
- [64] A. Aab et al. “Inferences on mass composition and tests of hadronic interactions from 0.3 to 100 EeV using the water-Cherenkov detectors of the Pierre Auger Observatory”. In: *Phys. Rev. D* 96.12 (2017), p. 122003. DOI: 10.1103/PhysRevD.96.122003. arXiv: 1710.07249 [astro-ph.HE].
- [65] D. Gora et al. “Universal lateral distribution of energy deposit in air showers and its application to shower reconstruction”. In: *Astropart. Phys.* 24 (2006), pp. 484–494. DOI: 10.1016/j.astropartphys.2005.09.007. arXiv: astro-ph/0505371.
- [66] P. Lipari. “The Concepts of ‘Age’ and ‘Universality’ in Cosmic Ray Showers”. In: *Phys. Rev. D* 79 (2009), p. 063001. DOI: 10.1103/PhysRevD.79.063001. arXiv: 0809.0190 [astro-ph].
- [67] A. Aab et al. *Deep-Learning based Reconstruction of the Shower Maximum  $X_{\max}$  using the Water-Cherenkov Detectors of the Pierre Auger Observatory*. Jan. 2021. arXiv: 2101.02946 [astro-ph.IM].
- [68] A. Bridgeman. “Shower universality reconstruction of data from the Pierre Auger Observatory and validations with hadronic interaction models: Shower universality with Auger data”. In: *PoS ICRC2017* (2018). Ed. by Darko Veberic, p. 323. DOI: 10.22323/1.301.0323.
- [69] A. Bridgeman. “Determining the Mass Composition of Ultra-high Energy Cosmic Rays Using Air Shower Universality”. 51.03.03; LK 01. PhD thesis. Karlsruhe Institut für Technologie (KIT), 2018. 217 pp. DOI: 10.5445/IR/1000085504.
- [70] G. A. Anastasi. “Reconstruction of events from the surface detector of the Pierre Auger Observatory using air shower universality”. PhD thesis. SISSA, Trieste, 2019.
- [71] M. Erdmann and J. Glombitza. “Deep Neural Network for the Reconstruction of the Shower Maximum  $X_{\max}$  using the Water-Cherenkov Detectors of the Pierre Auger Observatory”. Auger internal note GAP-2020-005. 2020.