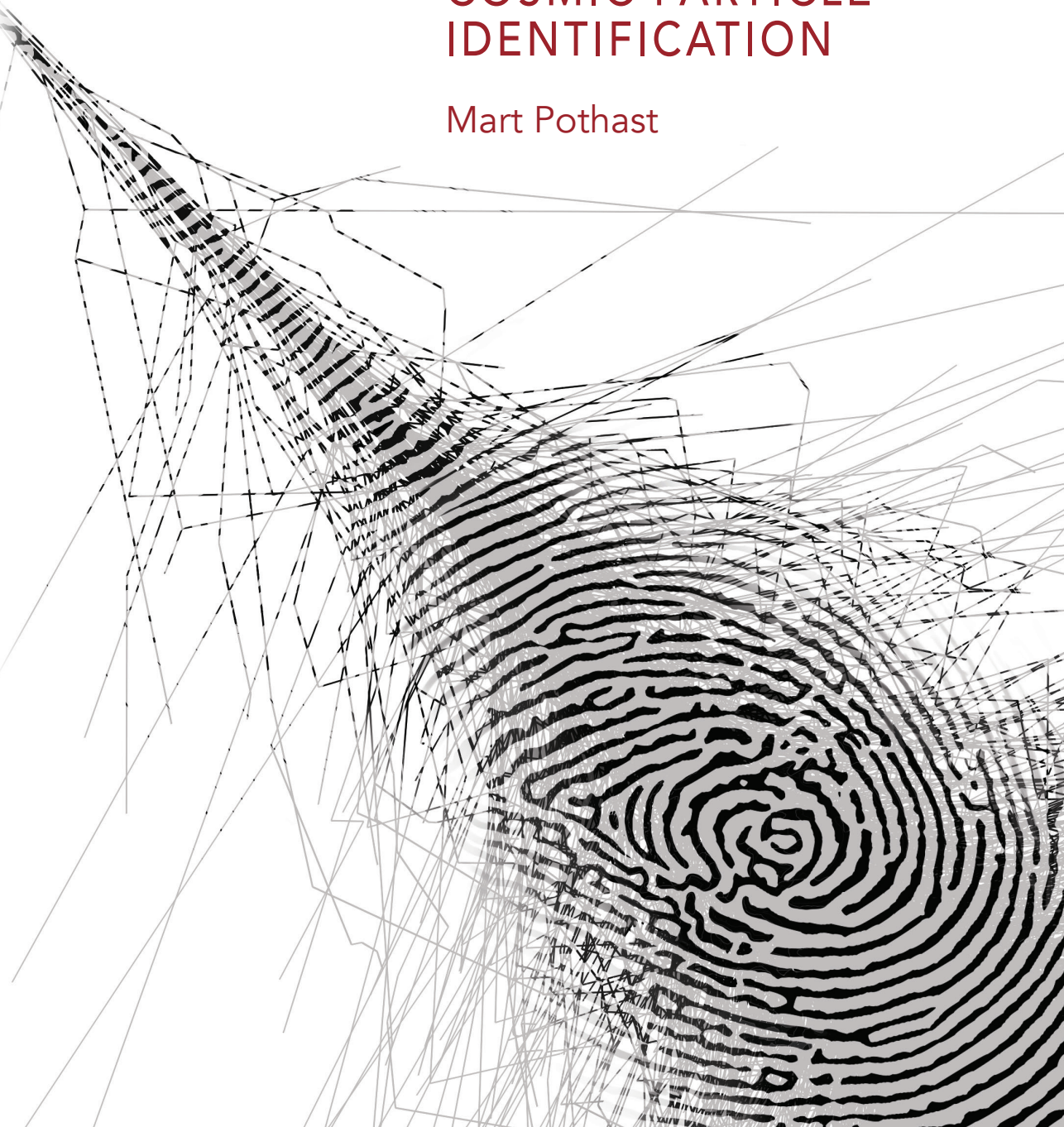


ULTRA-HIGH-ENERGY COSMIC PARTICLE IDENTIFICATION

Mart Pothast



Ultra-high-energy cosmic particle identification

Mart Pothast

© Mart Pothast 2023

Ultra-high-energy cosmic particle identification

Thesis, Radboud University Nijmegen

ISBN: 978-90-831804-9-6

Cover design by © Elma Hogeboom 2023 for GreenThesis

Printed on 100% recycled paper. A tree has been planted for every copy of this thesis.



This work originates as part of the research programme of the Foundation for Fundamental Research on Matter (FOM), and falls as of 1 April 2017, under the responsibility of Foundation for Nederlandse Wetenschappelijk Onderzoek Instituten (NWO-I), which is part of the Dutch Research Council (NWO).

Ultra-high-energy cosmic particle identification

Proefschrift ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,
volgens besluit van het college voor promoties
in het openbaar te verdedigen op

vrijdag 15 september 2023
om 10.30 uur precies

door

Mart Ruben Pothast

geboren op 4 februari 1991
te Rheden

Promotoren:

Prof. dr. S.J. de Jong

Prof. dr. C.W.J.P. Timmermans

Manuscriptcommissie:

Prof. dr. M. Haverkorn van Rijsewijk

Prof. dr. M. Erdmann

RWTH Aachen, Duitsland

Prof. dr. S.C.M. Bentvelsen

Universiteit van Amsterdam

Prof. dr. R.J.M. Snellings

Universiteit Utrecht

Dr. D. Veberič

Karlsruher Institut für Technologie, Duitsland

Contents

Introduction	1
1 Ultra-high-energy cosmic rays	3
1.1 Cosmic rays	3
1.2 Extensive air showers	7
1.3 Universality	9
1.4 Recent results on UHECRs	13
2 The Pierre Auger Observatory	17
2.1 Surface Detector	18
2.2 Fluorescence Detector	20
2.3 AugerPrime	21
3 Signals in water-Cherenkov and scintillator detectors	23
3.1 Matrix formalism	23
3.2 The detector response	24
3.3 The energy spectrum of secondary particles	29
3.4 The signal ratios	32
3.5 Conclusion	34
4 Time distribution of signal in surface particle detectors	39
4.1 Muon production depth	39
4.2 Derivation of the arrival time distribution of muons	40
4.3 Detector energy response to muons	46
4.4 Detector time response	47
5 Time trace templates	49
5.1 Simulations	49
5.2 Total signal model	50
5.3 Time-dependent signal	55
5.4 Uncertainty of signal in single time bins	60
5.5 Fit procedure	65
6 Depth of shower maximum with the SD	69
6.1 Data selection	69
6.2 Start time of the signal	72
6.3 Validation using MC simulations	74
6.4 Calibration on hybrid data	78
6.5 Surface detector only data	79
7 Conclusion and outlook	91

Bibliography	100
A Additional plots and figures	101
B Additional tables	107
C Research data management	109
Summary	111
Samenvatting	113
Acknowledgements	115
Curriculum vitae	117

Introduction

Astroparticle physics is the field where physics on the smallest scales—what are the building blocks of matter?—and on the largest scales—can we understand the universe?—meet. It started early 20th century with the discovery of cosmic rays as background radiation coming from space. Today, astroparticle physics is more relevant than ever if we want to understand the origin of the universe, what are black holes, what is dark matter and how do neutrinos fit into the standard model of particle physics. The backbone of astroparticle physics is the one that concerns ultra-high-energy cosmic rays (UHECRs), the topic of this thesis. UHECRs are the highest energy particles coming from space making them tantalizing objects for studying the universe. Understanding the origin and nature of the highest energy cosmic rays will not only help us understand their astrophysical sources—whether they are supermassive black holes, exploding stars or even decaying dark matter—but will also give us the opportunity to study particle physics interactions at energies that are well beyond what current human-made accelerators (and those currently under design) are capable of.

Finding the sources of ultra-high-energy cosmic rays and how they can accelerate particles up to enormous energies requires to first identify the kind of particles that make up these cosmic rays. Because cosmic rays are charged they are deflected by magnetic fields in and outside our galaxy, which makes backtracking the arrival direction of cosmic rays to their sources difficult. However, the deflection is minor if the charge of these particles is sufficiently low and their energy is ultra high. The particle type determines both the mass and the charge of the particles and a measurement of the particle type is generally referred to as the *mass composition*.

Ultra-high-energy cosmic rays can be detected through the extensive particle cascades (*air showers*) they create in the atmosphere. The largest observatory of cosmic-ray induced air showers is the Pierre Auger Observatory. It is a hybrid detector consisting of a Fluorescence Detector (FD), capable of measuring the longitudinal development of the shower in the atmosphere, and a Surface Detector array (SD) that measures the particle content of the shower footprint on the ground. The fluorescence method currently provides the best estimate of the mass composition through the measurement of the depth at which the energy deposit of the shower reaches its maximum, X_{max} . This measurement is, however, limited by statistics at the highest energies because the FD operates only on clear and dark nights. The SD, on the other hand, operates almost continuously but was never intended to be used to measure the mass composition. The Pierre Auger Observatory is undergoing an upgrade that intends to provide a measurement of the number of muons in air showers with the SD, which allows to discriminate between light and heavy primary cosmic rays.

In the meantime, it is worthwhile to study the currently available data of the SD and investigate the option of obtaining the mass composition via the timing information given by the current detectors. This is the primary goal of this thesis. Its content is as follows: [Chapter 1](#) gives additional details on the field of cosmic rays and extensive air showers. [Chapter 2](#) introduces the Pierre Auger Observatory. [Chapter 3](#)

describes how the different response of the current and new surface detectors allows the discrimination between muons and electrons. In [Chapter 4](#) an analytical model of the arrival time of particles on the ground is discussed and how this connects to the longitudinal development of the air shower. With the knowledge of the previous chapters, in [Chapter 5](#) a method is presented to create templates from simulations that describe the signal as a function of time. Furthermore, it is shown how to infer X_{max} from the SD using a template fit. [Chapter 6](#) shows the validation of the template fit procedure using Monte Carlo simulations. A calibration with hybrid data (both measured with the FD and SD) is performed and finally an extension of X_{max} measurement with the SD at the highest energies is presented. The conclusions are discussed in [Chapter 7](#).

1 Ultra-high-energy cosmic rays

Ultra-high-energy cosmic rays are generally regarded as the charged particles with energy above 10^{18} eV that come from space. These particles are truly the highest energy particles that we know of and their maximum energy is larger than 10^{20} eV. However, the origin of UHECRs—let alone how they acquire their enormous energies—is still elusive. In this chapter the research field of UHECRs is introduced.

1.1 Cosmic rays

A brief history

Cosmic rays were discovered by a combination of studies that tried to understand the ever present background radiation measured with electroscopes in the early 1900s. Most famous are the balloon flights by Victor Hess that showed that the radiation increased at higher altitudes [1], but there was more evidence that proved that this radiation came from the cosmos and not from the earth or the sun (see [2] for an elaborate historic overview). Later it was shown that the bulk of this radiation coming from above consists mostly of (positively) charged particles, but the name cosmic rays (a misnomer) has stuck.

Subsequent technological improvements in detecting high energy radiation from cosmic rays (like cloud chambers [3], photographic emulsion [4], Geiger counters [5]) sparked the detection of new particles like the positron, muon and pion and led to many Nobel Prizes. It also meant the birth of the field of particle physics before human-made accelerators were available. When faster electronics allowed to precisely measure coincidences in time between Geiger counters, it was French physicist Pierre Auger who reported coincidence rates of far apart placed detectors and concluded that these must stem from very-high-energy cosmic rays creating cascades of particles in the atmosphere [6]. Bruno Rossi reported on these coincidences even earlier [7] and made early measurements and calculations of particle showers in other media [8]. Later, after the Second World War, larger observatories capable of measuring extensive air showers saw light and allowed to establish the cosmic-ray energy spectrum from 10^{14} eV up to 10^{20} eV.

One of the first ‘large’ observatories was build by Linsley at the Volcano Ranch site in New Mexico and consisted of particle detectors on the ground to detect air showers induced by UHECRs [9, 10]. Other notable observatories using particle detectors were AGASA [11], Haverah Park [12] and Yakutsk [13]. A new technique for detecting air showers was developed in the 1960s. It uses the fluorescence light emitted by excited molecules in the atmosphere and was first successfully implemented in the Fly’s Eye project [14] that later evolved into HiRes [15]. At present, there are two very large observatories of UHECRs: the Pierre Auger Observatory and Telescope Array (TA). The Pierre Auger Observatory is the subject of this thesis and is discussed in detail in [Chapter 2](#). Telescope Array [16] is located in the Utah desert and covers the Northern

Hemisphere. It is also a hybrid detector with both particle detectors and fluorescence telescopes, but it is somewhat smaller than the Pierre Auger Observatory.

Energy spectrum

The energy spectrum of cosmic rays (shown in Fig. 1.1) is established over many decades in energy and encompasses an even larger spread in flux. For low energies the flux is high enough such that cosmic rays can be directly detected by relatively small space-based instruments like AMS [17] and PAMELA [18]. These instruments can measure the energy spectrum and the composition of cosmic rays up to several TeV very accurately. At higher energies the flux becomes too low for direct detection to be feasible, but these particles can still be studied indirectly via the air showers they create in the atmosphere. Because these air showers can be extended over several kilometers at the highest energies, it is still possible to detect and analyze these particles using sparse arrays of detectors on the ground. The energy spectrum of cosmic rays is at a first glance a quite featureless power-law spectrum, but subtle differences are apparent such as the so-called *knee* at $\text{PeV} = 10^{15} \text{ eV}$ energies, the *ankle* around an energy of $\text{EeV} = 10^{18} \text{ eV}$ and a steep suppression above $E \approx 4 \times 10^{19} \text{ eV}$.

Sources, acceleration and propagation

Rather embarrassingly, even after more than 100 years of cosmic-ray research, the problem of the cosmic-ray origin—where do these particles come from and how do they get their energy?—is still not solved. Cosmic rays below the knee are thought to originate from our own galaxy and supernova remnants are good candidates but firm proof is still sought-after [27, 28].

In the ultra-high-energy regime cosmic rays are thought to come from outside our own galaxy. In general, two types of origins are possible: Top down and Bottom up. Top down refers to UHECRs as decay products of super heavy particles that originated in the early universe, while Bottom up means that ordinary particles are accelerated at some astrophysical source up to ultra-high energies. The Top down origin is more and more constrained by lack of detection of ultra-high-energy photons [29]. For the Bottom up origin several extragalactic astrophysical source classes can be considered like: Active Galactic Nuclei (AGN) [30], Starburst galaxies [31], Gamma-ray Bursts (GRBs) [32], magnetars [33] and Tidal eruption events (TDEs) [34].

How do cosmic rays acquire their, sometimes, enormous energy? The mechanism of accelerating particles by astronomical objects was first considered by Enrico Fermi. From scattering on magnetic fields particles can gain small amounts of energy, this is called Second order Fermi acceleration. First order Fermi acceleration refers to particles gaining small amounts of energy every time they cross a shock front as, for example, can be formed from an exploding supernova. Other acceleration mechanisms are acceleration in electric fields [35] and the ‘espresso’ mechanism [30].

In any case, the sources of cosmic rays must be powerful enough and, to accelerate particles efficiently, they must be large enough and have sufficiently strong magnetic fields to confine these particles for ample amount of time to allow them to be accelerated to ultra-high energies. The Hillas criterion allows to select possible source candidates based on their size and magnetic field strength [36]. This criterion states that particles will be confined to the source for as long as their Larmor radius is smaller

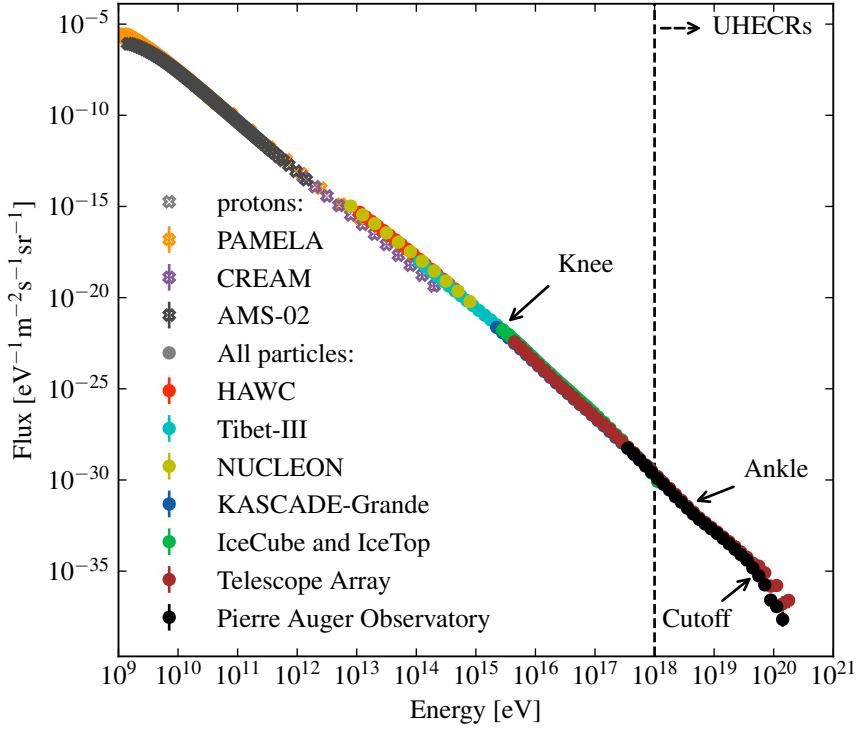


Figure 1.1: The cosmic ray energy spectrum from $10^9 \text{ eV} = 1 \text{ GeV}$ up to 10^{21} eV as measured by a selection of recent observatories: AMS-02 [17], PAMELA [18], CREAM [19], IceCube and IceTop [20], KASCADE-Grande [21], NUCLEON [22], HAWC [23], Tibet-III [24], Pierre Auger Observatory [25], Telescope Array [26]. At lower energies the pure proton energy spectrum is shown.

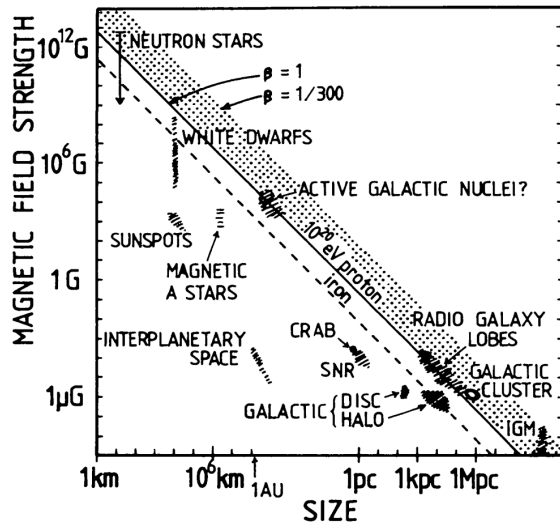


Figure 1.2: Original Hillas criterion plot [36]. Several potential source classes are shown and whether they can possibly accelerate cosmic rays up to $E = 10^{20}$ eV based on their size and magnetic field strength. Sources to the upper right of the diagonal satisfy the criterion. β indicates the speed of the outflow.

than the size of the source. The corresponding Hillas plot with some potential sources is shown in Fig. 1.2.

The origin of cosmic rays is obscured because these particles undergo interactions on the way from their sources and because charged particles are deflected by magnetic fields in and outside the galaxy. Therefore, cosmic rays cannot be easily traced back to where they come from. Several magnetic field components contribute to the shift in arrival direction of UHECRs (see Fig. 4.3 in [37]). The galactic magnetic field (GMF) can be modeled from observations and deflections for $E = 10^{20}$ eV protons are of the order of a few degrees, where the exact magnitude depends on the position on the sky. The magnetic fields in between galaxies are not very well known, but considering a uniform random magnetic field with a strength of the order of nG, a 10^{20} eV proton would be deflected by less than a degree over 10 Mpc.

At the highest energies, the propagation of cosmic rays through the universe is mostly affected by their interaction with the cosmic microwave background (CMB). After the discovery of the CMB; Greisen, Zatsepin and Kuz'min [38, 39] predicted a hard cutoff in the energy spectrum of cosmic rays (*GZK cutoff*) due to photopion production. Depending on the composition of the primary cosmic rays the cutoff energy is about $E \sim 5 \times 10^{19}$ eV. For heavier nuclei photodisintegration and pair production are the most important energy loss processes. The attenuation length for cosmic rays with energies above the GZK cutoff are below 1000 Mpc, meaning that the highest energy particles must come from relatively nearby sources (at cosmic scales).

One of the most important open questions in the ultra-high energy regime is the origin of the flux suppression (cutoff) above $E \approx 4 \times 10^{19}$ eV observed by the Pierre Auger Observatory and the Telescope Array (see Section 1.4). Possible scenarios to

CHAPTER 1. ULTRA-HIGH-ENERGY COSMIC RAYS

explain the cutoff can be roughly classified as follows:

Maximum-rigidity scenario Particles at the sources are only accelerated up to some maximum rigidity ($\mathcal{R} = E/Z$). In other words: the sources run out of steam and the observed suppression is due to an energy cutoff of the source spectrum. An implication of this scenario is that the mass composition becomes heavier at increasing energies, because particles with higher charge (and thus mass) have lower rigidity and will be accelerated to a higher energy at the sources with respect to particles with a lower charge.

Photo-disintegration scenario The flux suppression is caused by propagation effects, mainly from photo-disintegration of nuclei. In this scenario light elements reaching Earth are the result of fragmentation of heavier nuclei that are produced in the sources. In this scenario the maximum energy of particles at the sources is larger than the energy at which energy losses start to dominate.

Proton-dominance scenario This classical model assumes almost all UHECRs are protons and the suppression is the result of photo-pion production (GZK cutoff). In this scenario the ankle is the result of e^+e^- production [40].

Of course, reality can be more complex with different sources contributing different primary particle types and having different energy cutoffs.

1.2 Extensive air showers

The flux of UHECRs is so low (~ 1 per km^2 per century) that direct detection is quite impractical. They can only be indirectly detected through the extensive air showers they produce in the Earth's atmosphere. When a highly energetic particle interacts with molecules in the atmosphere it creates secondary particles that can in turn also interact or decay etcetera, creating a cascade of many particles. This cascade of particles becomes very extended and can be detected at the surface of the Earth.

Generally, three components are identified in an extensive air shower: the hadronic, electromagnetic (photons, electrons and positrons) and muonic component. The first interactions mostly produce hadrons but these can decay to muons (mostly from charged pions) and to electromagnetic particles (mostly from neutral pions).

Longitudinal development

The electromagnetic cascade has a characteristic longitudinal profile generally described with respect to the amount of atmosphere that is traversed, also called the slant depth:

$$X(h) = \int_h^\infty \rho(\vec{r}(z)) dz, \quad (1.1)$$

where $\rho(\vec{r}(z))$ is the density of the atmosphere at a position $\vec{r}(z)$ so that the z component points along the shower axis.

For a (toy model) pure electromagnetic shower with only photons, electrons and positrons that undergo either pair production (photons) or bremsstrahlung (electrons and positrons) it can be shown that the number of particles increases up to a maximum point called X_{\max} . This point is reached when the average energy of the particles drops below the critical energy for pair production and bremsstrahlung [41].

The value of X_{\max} increases with primary energy and depends on the mass of the primary particle for hadronic showers. Commonly, the mean of the depth of shower maximum $\langle X_{\max} \rangle$ and its standard deviation $\sigma(X_{\max})$ are used to probe the mass composition of UHECRs. The change of $\langle X_{\max} \rangle$ with energy is called the *elongation rate*. A change in the elongation rate can point to a change of the mass composition.

Air showers induced by hadrons produce mostly neutral and charged pions, the former decay into photons while the latter decay to muons when their energy drops below a critical energy. Showers induced by nuclei with atomic mass A can be treated as a superposition of A proton induced showers. The total energy is effectively distributed over the nucleons thereby decreasing the value of $X_{\max} \sim -\ln A$. Because more energy goes into charged pions, showers induced by heavier primaries have more muons at the same total energy [42].

The longitudinal profile is often described with an empirical formula (which is inspired by electromagnetic cascade theory) that was suggested by Gaisser and Hillas [43]:

$$f_{\text{GH}}(X) = \left(\frac{dE}{dX} \right)_{\max} \left(\frac{X - X_0}{X_{\max} - X_0} \right)^{\frac{X_{\max} - X_0}{\lambda}} \exp \left(-\frac{X - X_{\max}}{\lambda} \right). \quad (1.2)$$

This formula gives the energy deposited at slant depth X which peaks at X_{\max} and has shape parameters λ and X_0 .

Lateral development

Extensive air showers induced by UHECRs have a projected size on the ground of several kilometers in width and length. The lateral distribution is generally regarded with respect to the distance to the shower core in the shower plane—denoted as r . For non-vertical showers there is an asymmetry in the azimuth angle in the shower plane—denoted with ψ —due to the fact that the amount of atmosphere traversed depends on this angle. For electromagnetic cascades the lateral distribution can be calculated as was done by Kamata and Nishimura [44]. Their result was parametrized by Greisen [45] into what is now commonly referred to as the NKG formula that describes the electron density as a function of the distance to the shower core

$$\rho(r) = C \left(\frac{r}{r_M} \right)^{s-2} \left(1 + \frac{r}{r_M} \right)^{s-9/2}. \quad (1.3)$$

C is the normalization that depends on the total number of electrons, s depends on the shower age (the depth with respect to X_{\max}) and r_M is the Molière radius which is the radius that contains about 90% of the energy of the shower and depends on the atmospheric depth.

Muons have larger transverse momentum and are less hampered by the atmosphere and so they make it out to farther distances. For hadronic showers often a modified

version of the NKG function is used to describe the lateral density [46] (see also Eq. (2.1)). Such a function to describe the lateral distribution is commonly called a Lateral Distribution Function or LDF.

Simulations

The properties of UHECRs (type, energy, direction) have to be inferred indirectly from the induced extensive air showers. To understand the properties of extensive air showers and interpret the data requires simulation of all physics involved from the primary particle injected in the atmosphere up to the detector response. Monte Carlo (MC) methods are used to accurately predict the shower-to-shower fluctuations that arise through the probabilistic nature of physics at the quantum scale.

At ultra-high energies the physics of hadronic particle interactions have to be extrapolated from accelerator data. Consequently, there are different hadronic interaction models that each have a different phenomenological implementation of hadronic particle interactions. Recent models that are tuned to LHC data are: EPOS-LHC [47], QGSJET-II.04 [48] and SIBYLL 2.3c (most recent version 2.3d [49]), but note that they have significantly different predictions at ultra-high energies. On the low energy side ($E \lesssim 200$ GeV) different models like GHEISHA [50], FLUKA [51] and UrQMD [52] are used that are more in agreement with each other and all describe the available accelerator data. The high energy and low energy models are generally patched together in an air shower simulations package that also takes care of propagation of these particle in the atmosphere. Common simulation packages for cosmic-ray induced air showers are CORSIKA [53] and AIRES [54].

The number of particles in the shower grows roughly linear with energy. At ultra-high energies it becomes infeasible to follow every particle produced. To overcome this, a thinning algorithm is used that assigns a weight to particles below a chosen energy cutoff such that one particle represents many particles and the other particles are discarded from those that are tracked.

For the Pierre Auger Observatory a full description of the measured signal is obtained using a complete simulation chain as implemented in the Offline package [55]. In short: the particle output of simulations is injected into the detector configuration and the detector response is simulated with the GEANT4 package [56].

1.3 Universality

An extensive air shower initiated by a primary with large energy (in this case with $E \gtrsim 10^{18}$ eV) consists of many interactions and particles being created in the atmosphere. It is not very surprising that features of individual particle interactions are averaged out and that the air shower can be described by only a few parameters, much like a thermodynamic system. This is called *air shower universality*.

Electromagnetic showers

Universality in electromagnetic (EM) showers is well established theoretically. Considering a pure electromagnetic shower that consist only of photons, electrons and positrons that undergo either bremsstrahlung or pair production it is possible to calculate the energy distribution of secondary electrons and photons and it can be shown

that these are only a function of the shower age, a feature that was already described by Rossi and Greisen in 1941 [57]. The shower age is sometimes defined differently in the literature (see [58] for a discussion), in this case it is sufficient to realize that the shower age depends on the distance to X_{\max} . A shower that is still growing ($X < X_{\max}$) is termed ‘young’ while a shower past X_{\max} is ‘old’. A consequence of this universality with respect to the shower age is that the longitudinal profile (as a function of the amount of matter traversed) can be described analytically and depends only on the distance to X_{\max} . An electromagnetic shower can thus be described with only two parameters: the primary energy—which determines the total number of particles—and (the distance to) X_{\max} .^{*}

Hadronic showers

Air showers initiated by protons or nuclei are complicated by hadronic interactions in the atmosphere. Mostly neutral and charged pions are produced in the first interactions. Neutral pions decay almost immediately to photons and initiate electromagnetic shower components while charged pions continue to interact and eventually decay to muons. Monte Carlo simulations show that the secondary electron and positron energy spectra of hadron induced showers still behave universally with respect to the shower age [59–61]. The obtained spectra are very similar to the analytical calculations as done for electromagnetic showers by Rossi and Greisen in 1941 as shown in [58].

The hadronic component in air showers is less well understood as can be judged from the different predictions by hadronic interaction models on X_{\max} (see Fig. 1.6). At large distances from the core (as in observatories of UHECRs) or at very old shower age the hadronic component is probed by muons that make it to the surface.

In [62] it was shown that the signal in water-Cherenkov detectors (WCDs) at 1000 m from the core can be well described with respect to the distance of that detector to X_{\max} :

$$\Delta X = X(h_{\text{ground}} + h_{\text{proj}}) - X_{\max}, \quad (1.4)$$

with $h_{\text{proj}} = r \sin \theta \cos \psi$, where θ is the zenith angle of the shower, r and ψ are, respectively, the distance to the core and the azimuth angle in the shower plane. See Fig. 1.3 for a sketch of a shower and the definition of ΔX .

A remarkable degree of universality for the muon component with respect to ΔX [†] is found when correcting the normalization with a new parameter N^μ , which is the total number of muons in the shower. N^μ is essentially used as a proxy for the number of hadrons produced in the air shower. In this thesis the relative number of muons is used, denoted as R^μ . It is defined as the number of muons produced in an air shower with respect to the average number of muons of a similar (same energy, θ and X_{\max})

^{*}Assuming that the atmosphere is known and uniform in the lateral direction.

[†]Note that for muons the universality is expected with respect to the muon production depth and it is perhaps not the traversed atmosphere that is important but the actual length from the detector to the production point. See also Chapter 4.

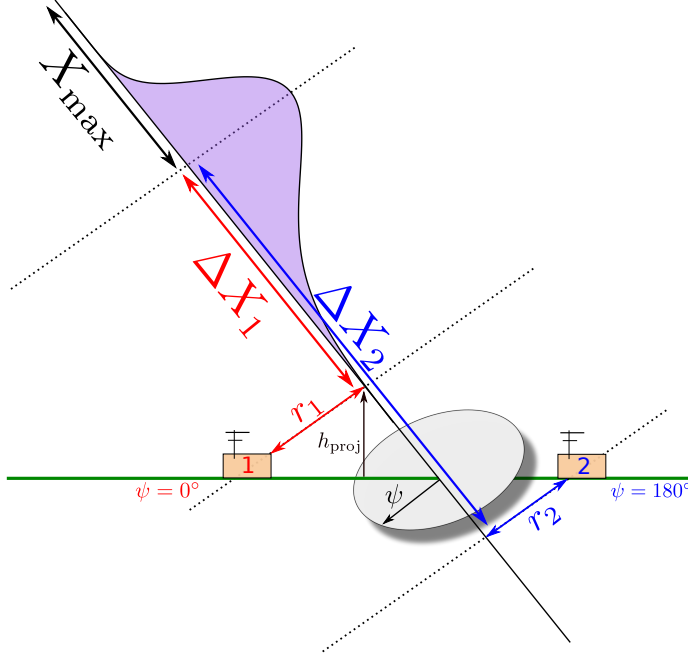


Figure 1.3: Sketch of a non-vertical shower. The electromagnetic longitudinal profile is shown with the position of X_{\max} and the definition of ΔX for two different stations on the ground (by Eq. (1.4)) is shown.

proton induced shower simulated with EPOS-LHC[‡]

$$R^\mu = \frac{N^\mu}{N_p^\mu(E, \theta, X_{\max})}, \quad (1.5)$$

with $N_p^\mu(E, \theta, X_{\max})$ the expected average number of muons for a proton shower with energy E , θ and X_{\max} . The numerator and denominator in Eq. (1.5) do not need to be the number of muons, but can also be a proxy for it. For example in Chapter 5 R^μ is defined as the ratio of the signal from muons in WCDs at 1000 m from the shower core with respect to what would be expected on average for that signal from proton showers with the same energy, θ and X_{\max} .

Particle components of hadronic air showers

In general the secondary particles created in a hadronic air shower are divided into similarly behaving components. The two dominant components at large distances from the core are electromagnetic particles (consisting of photon, electrons and positrons) and muons and anti-muons. Naturally, both components are fed by initial hadronic interactions.

[‡]Beware that different definitions of R^μ are around. For example in [63] R^μ denotes the number of muons with respect to proton simulated showers at $E = 10^{19}$ eV. In this thesis R^μ for iron induced showers simulated with EPOS-LHC have $R^\mu \approx 1.3$.

The EM component still behaves universally because there are many interactions between the initial (high energy) pion decay and the particles on the ground. However, parts of the EM component in hadronic air showers scale with the number of muons and thus with R^μ . Electromagnetic particles that are the result of muon decay obviously behave similarly to their parent muons and are considered as a third component. A fourth component was introduced in [64]. This fourth component is the result of low energy hadrons (mostly π^0 s) produced in the shower core with relatively large transverse momentum that create a sub-shower of electromagnetic particles close to the detector, this component is sometimes called the *jet* component. Because of its close relation to hadronic interactions—there are not many interactions since the last hadronic interaction—this component is not universal with respect to the primary energy and ΔX like in a pure electromagnetic shower. This jet component grows with the number of hadrons produced in the shower, where R^μ is used as a proxy for the number of hadrons.

In CORSIKA simulations, particles belonging to the jet or hadron EM component are identified by their hadronic generation and particle weight. Their particle weight is smaller with respect to the pure EM component because of their close relation to hadrons which retain a higher weight in CORSIKA (see [64] for details).

Summarized, the four components—that are the largest in terms of signal in WCDs—are:

- μ The muon component which consist of all muons and anti-muons.
- $e\gamma$ The pure electromagnetic component that scales only with the energy and distance to X_{\max} .
- $e\gamma(\mu)$ Photons, electrons and positrons that are the result of muon decay so that they trace the muon component. This component is also known as the *EM halo*.
- $e\gamma(\text{had})$ EM component from low energy hadrons, or jet component.

In practice, the very few hadrons that create any signal in the WCD are assigned to the fourth component.

Parametrization

The idea of universality is that the four components above can be described with only three parameters E , ΔX and R^μ , although each component behaves differently. Of course, the signal from each component depends on the location of the detector with respect to the shower core. Other effects of the ground plane and the shape of the detector are also imaginable[§]. A parametrization of the signal in a detector will thus look like:

$$S(E, X_{\max}, R^\mu, \hat{a}, \vec{r}_{\text{core}}) = \sum_{i=\text{comp}} f_i(E, X_{\max}, R^\mu, \hat{a}, \vec{r}_{\text{core}}), \quad (1.6)$$

[§]For example, for a non-spherical detector—like the WCD—the direction of incoming particles influences the signal.

CHAPTER 1. ULTRA-HIGH-ENERGY COSMIC RAYS

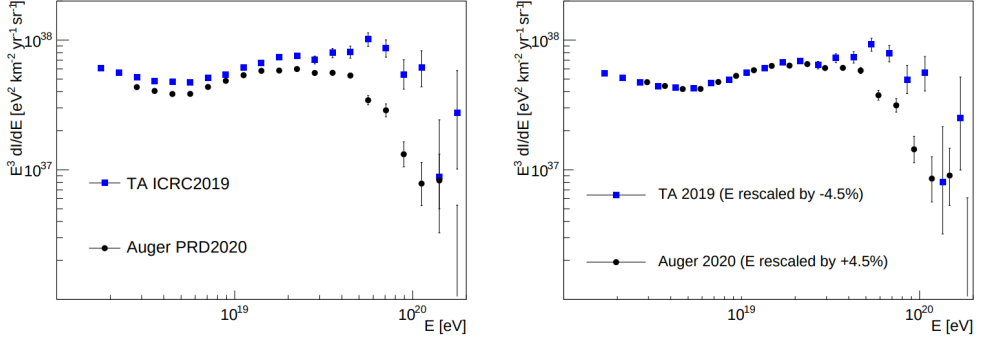


Figure 1.4: Energy spectrum of UHECRs as measured at the Pierre Auger Observatory and Telescope Array on the left. On the right when rescaling the energy by $\pm 4.5\%$ [66].

where \hat{a} is the shower axis and \vec{r}_{core} the position of the core with respect to the observation plane. $f_i(E, X_{\text{max}}, R^\mu, \hat{a}, \vec{r}_{\text{core}})$ can be any arbitrary function (or combination of functions) that describes the component i well. A specific parametrization will be described in [Chapter 5](#).

Universality in the arrival time

In [Chapter 4](#), a model will be discussed that illustrates how the arrival time of muons is related to the longitudinal development. The arrival time of the electromagnetic component also behaves universally [65]. In that reference the time distribution of the signal in WCDs is parametrized as a log-normal distribution. A similar procedure will be applied in [Chapter 5](#).

1.4 Recent results on UHECRs

A selection of recent results by the currently operating observatories for UHECRs are presented below.

Energy spectrum

The fluxes of UHECRs (multiplied by E^3) as a function of energy as measured by the Pierre Auger Observatory and Telescope Array are shown in [Fig. 1.4](#). The discrepancy of the measurements are striking but are compatible within the systematic uncertainties on the energy scale. After rescaling the energy by $\pm 4.5\%$ there is agreement up to the flux suppression (right panel of [Fig. 1.4](#)). Work is underway to study whether the remaining differences at high energy are the result of a different reconstruction, an energy dependent difference in the calibration of the energy scale or if there is a different energy spectrum of UHECRs in the northern and southern hemisphere [66].

A new feature that was discovered using the Pierre Auger Observatory is the *instep*, which is a flattening of the energy spectrum just before the cutoff [67].

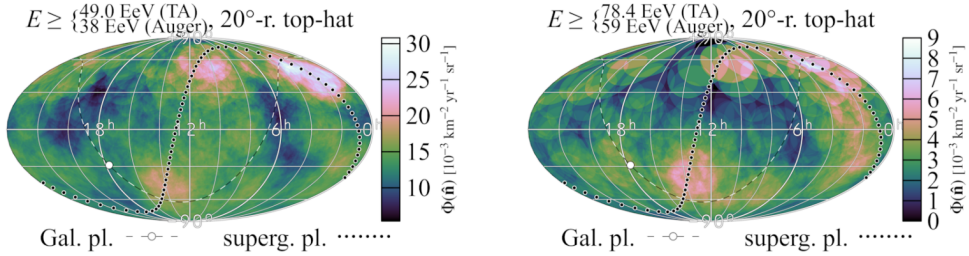


Figure 1.5: Arrival direction of UHECRs from the combined data set of the Pierre Auger Observatory and Telescope Array in equatorial coordinates for two energy cutoffs [70]. For each experiment a different energy cutoff is chosen so that the energy spectrum in the common declination band matches [66]. The arrival direction is smeared with a 20° top-hat.

Arrival directions

Even though cosmic rays are charged and do not point directly back to their sources, the distribution of arrival directions can point to large, intermediate or small scale anisotropies that can reveal potential astrophysical sources. The arrival direction of UHECRs with energies above 8 EeV revealed a dipole structure pointing away from the galactic centre, indicating that UHECRs do not come from our own galaxy [68]. There is additional evidence of a deviation of isotropy at energies above 32 EeV [69].

Figure 1.5 shows the combined arrival direction of UHECRs as detected by the Pierre Auger Observatory and Telescope Array for two energy cutoffs. There is no evidence of individual sources of UHECRs, but there is a correlation of the arrival direction with galaxy catalogs, of which the correlation with nearby starburst galaxies has the highest significance [70].

Depth of shower maximum

The mass composition is best probed by the depth of shower maximum. Figure 1.6 shows the mean and standard deviation of X_{\max} as a function of energy as measured using the Fluorescence Detector of the Pierre Auger Observatory. The data (when compared to MC simulations) suggest a relatively light composition around the ankle that becomes progressively heavier with increasing energies. In other words, there is a break in the elongation rate. Also the fluctuations of X_{\max} decrease above $\log_{10}(E/\text{eV}) = 18.5$ indicating a more pure and heavier composition. Figure 1.7 shows the mean depth of shower maximum as a function of energy as measured at Telescope Array. It is not trivial to compare the TA results directly with the Pierre Auger Observatory measurement because the latter reports X_{\max} corrected for detector effects (the ‘true’ X_{\max} distribution) while TA compares the measured value to simulations that underwent a full reconstruction procedure.

Number of muons

The number of muons in air showers can be directly measured using highly inclined showers ($\theta > 60^\circ$) where the electromagnetic part of the showers is absorbed in the atmosphere. Hybrid inclined events measured by the Pierre Auger Observatory showed

CHAPTER 1. ULTRA-HIGH-ENERGY COSMIC RAYS

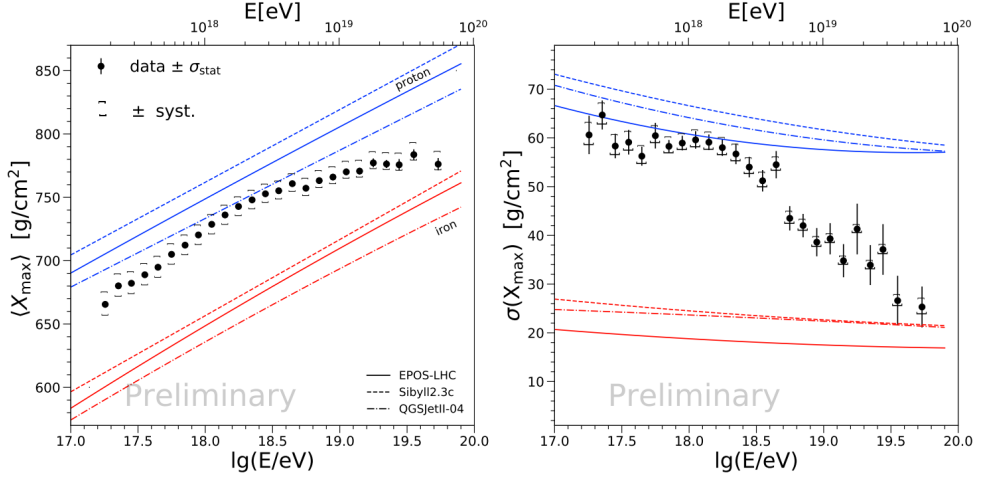


Figure 1.6: Depth of shower maximum from hybrid events detected by the Pierre Auger Observatory as a function of energy [71]. On the left the mean and on the right the standard deviation of the distribution in bins of $\Delta \log_{10}(E/\text{eV}) = 0.1$, where the last bin is integrated to include all events above $\log_{10}(E/\text{eV}) = 19.6$.

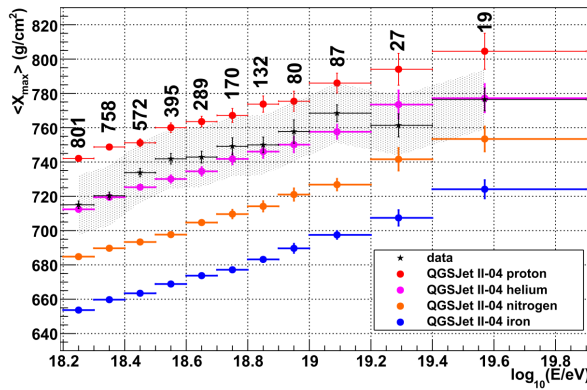


Figure 1.7: Mean depth of shower maximum from hybrid events as reported by the Telescope Array Collaboration as a function of energy [72].

that the number of muons measured is incompatible with the mass composition from the X_{\max} measurement when compared to simulations. For a given mass composition (for example set by the X_{\max} measurements), the number of muons in simulations is underestimated by about 30%. This muon deficit in the simulations is sometimes referred to as the ‘muon puzzle’.

Recently, a meta-analysis was performed that combined muon measurements of different experiments [73]. To be able to compare different experiments the z -scale was introduced:

$$z = \frac{\ln(N_{\mu}^{\text{det}}) - \ln(N_{\mu_{\text{p}}}^{\text{det}})}{\ln(N_{\mu_{\text{Fe}}}^{\text{det}}) - \ln(N_{\mu_{\text{p}}}^{\text{det}})}, \quad (1.7)$$

where N_{μ}^{det} is any estimate of the muon density as measured by different experiments and $N_{\mu_{\text{p}}}^{\text{det}}$ and $N_{\mu_{\text{Fe}}}^{\text{det}}$ are the simulated muon density estimates for proton and iron showers respectively (which depend on the hadronic interaction model). The z -scale as a function of energy is shown in Fig. 1.8 for two different hadronic interaction models. In [73] an increase of z per decade of energy of 0.22 (0.35) for EPOS-LHC (QGSJET-II.04) is reported with an 8σ significance. This indicates that the excess of muons in data increases with energy.

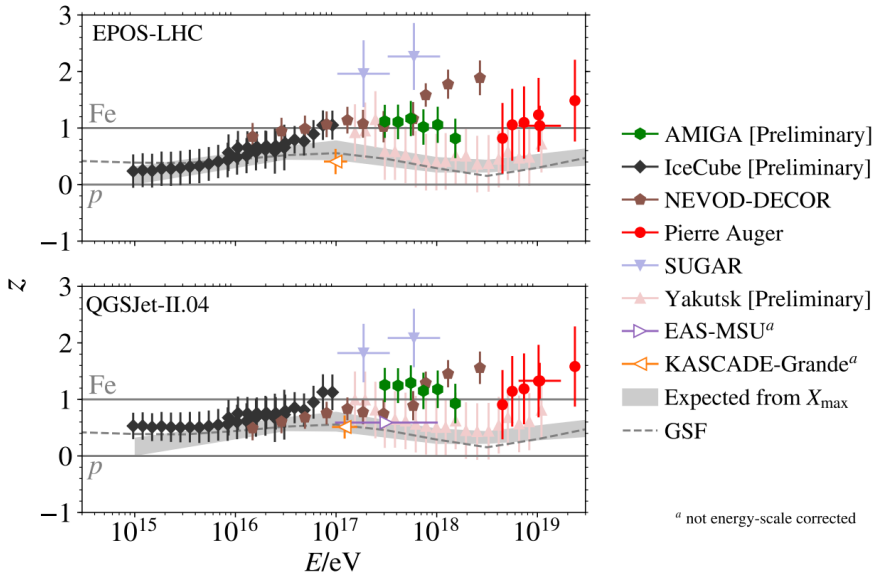


Figure 1.8: Muon density proxy (converted into z -scale, see text) from different experiments as a function of energy [73]. Shown are two hadronic interaction models. GSF is the *Global Spline Fit* model to high precision cosmic ray data that gives a prediction for the mass composition over all energies [74].

2 The Pierre Auger Observatory

The Pierre Auger Observatory is currently the largest observatory designed to detect cosmic rays [75]. It is located on a high plane in Argentina at 1400 m a.s.l. and covers about 3000 km². Its main part that focuses on detecting ultra-high-energy cosmic rays consist of 24 fluorescence telescopes shared over 4 sites (together called the Fluorescence Detector) and 1600 water-Cherenkov detectors (together called the Surface Detector array) spaced 1500 m apart. A small part of the array is filled with additional detectors on the ground spaced 750 m and 433 m and overlooked by three extra fluorescence telescopes (HEAT) pointing at higher elevations to detect lower energy showers. The Pierre Auger Observatory also accommodates a radio engineering array (AERA) for detecting air showers using radio waves. Several monitoring and calibration infrastructures are deployed at the Pierre Auger Observatory site, including laser facilities for calibration of the fluorescence telescopes (CLF and XLF), a balloon launch facility (BLF) for atmospheric monitoring and LIDAR and cloud cameras at every fluorescence telescope site. Figure 2.1 shows the observatory projected on the map.

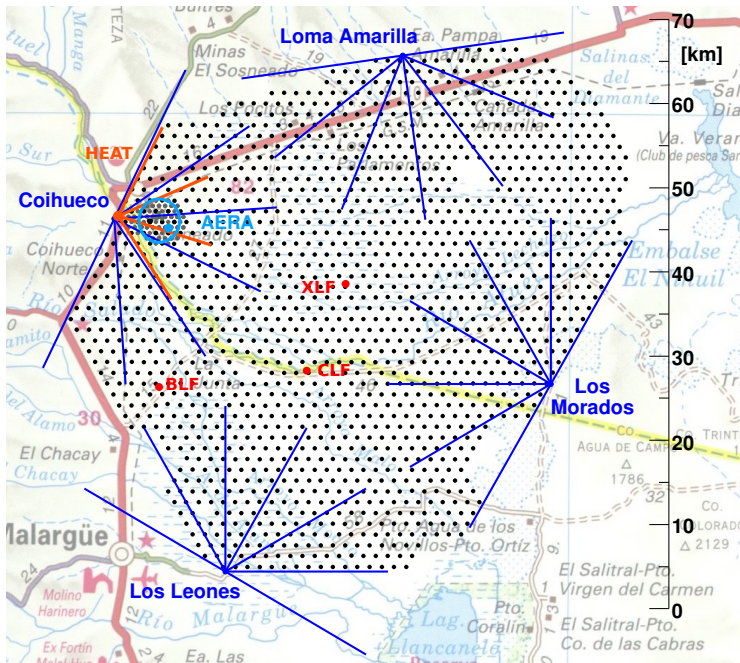


Figure 2.1: Map showing the Pierre Auger Observatory and its components [37]. Each dot represents a surface detector station and the blue and red lines indicate the line of sight of each fluorescence telescope.

2.1 Surface Detector

The Surface Detector array (SD) is composed of 1600 water-Cherenkov detectors (WCDs) on a triangular grid separated by 1500m (the SD is described in detail in [76]). Each surface detector station (see Fig. 2.2) consists of a cylindrical tank with a 1.8m radius and a height of 1.2m that is filled with purified water. Because of the depth of the water it serves as a calorimeter for electromagnetic particles in the tail of the shower, but high energy muons ($E_k \gtrsim 350$ MeV, see Chapter 3) traverse the whole tank and are dominant in terms of signal, especially far away from the shower core. The WCD has three photomultipliers that record the Cherenkov light produced by charged particles that travel at a speed greater than the speed of light in water. To extend the dynamic range each PMT sends two signals to the electronics. One *low gain* from the anode and one *high gain* from the dynode, that is amplified and inverted. The signal is digitized with a 40 MHz flash analog-to-digital converter (FADC) such that each FADC bin corresponds to a time window of 25 ns.

To provide a common reference level between detectors the signal is calibrated to the signal deposited by a vertically throughgoing muon into a unit commonly known as a VEM (vertical-equivalent muon). A calibration procedure is performed onboard the tanks to determine the value of 1 VEM in electronics units. The full procedure is described in [77]. In short: a histogram of the signal (in electronics units) from omni-directional atmospheric muons is collected every 60 s, the peak of the histogram corresponds to the value of 1.03 VEM for each PMT, where the value 1.03 is measured on a reference tank with a muon telescope. Note that there are actually two definitions of the VEM (and also two different histograms are collected) one refers to a VEM *peak* which is the pulse height of the signal and the other is the VEM *charge* which is the integrated signal of one vertical muon. The VEM peak is used to set the trigger threshold.

The SD trigger for showers of UHECRs has a hierarchical structure (described in detail in [78]). Single tank triggers (T1 and T2) are processed locally (to not strain the limited communication bandwidth) and sent to the central data acquisition system (CDAS) that checks for spatial and temporal correlation and decides to send a T3 trigger that starts the data acquisition. The T4 trigger selects real showers by requiring that the timing of the signal can be fitted by a plane shower front. Finally, there is a T5 trigger that places constraints on whether neighboring stations in the hexagon with the station with the largest signal (*hottest station*) were active. The highest quality events have a 6T5 trigger which means that all stations in the hexagon were taking data properly.

The data that is eventually stored for each station consist of two FADC traces (high and low gain) for each PMT that have a length of 768 bins and a GPS based time stamp indicating the time of the start of the signal. A signal finding algorithm (described in detail in [79]) determines which segments of the trace contain signal and which background noise. The baseline is determined from flat segments (that are assumed to be only noise) before and after signal segments. The start time is the time corresponding to the bin of the first signal segment.



Figure 2.2: On the left an upgraded SD station in the field. On the right the Los Morados FD station (credit: Steven Saffi).

SD event reconstruction

There are currently two reconstruction frameworks in use: the *Herald* and the *Observer*. In this thesis the *Observer* reconstruction will be used. The standard SD reconstruction aims to provide the incoming direction of the UHECR and (a proxy for) its energy. The direction can be obtained from the arrival times of the signal at each individual SD station and the energy is extracted using the total signal at 1000 m from the shower core in the shower plane, designated as $S(1000)$ which, for the spacing of 1500 m gives an optimal proxy for the energy [80]. $S(1000)$ is estimated from a fit to the lateral distribution of the total signal. A lateral distribution function similar to the NKG function is used:

$$S(r) = S(1000) \left(\frac{r}{1000} \right)^\beta \left(\frac{r + r_1}{1000 + r_1} \right)^{\beta + \gamma}, \quad (2.1)$$

with $r_1 = 700$ m is fixed. β and γ are parametrised as a function of the zenith angle and $S(1000)$ in a data-driven way. See Fig. 2.3 (left) for an example event reconstructed with Eq. (2.1) and alternatively with a log-log parabola (as used in the Herald reconstruction).

The reconstruction procedure is described in detail in [79]. In summary: first the direction of the shower axis is estimated by fitting a plane front to the arrival times, then the shower core and $S(1000)$ are determined by fitting the LDF to the total signal and finally the axis estimate is improved by fitting a spherical front. The LDF fit is a likelihood fit that also includes non-triggered stations. The uncertainty on the total signal is derived from twin-stations (stations in the array that are separated by 11 m). It is parametrized as a function of \sqrt{S} and the zenith angle θ under the assumption that the largest contribution comes from the fluctuation of the number of particles measured. The signal measured in VEM is effectively converted into the number of expected particles. The Poisson factor is

$$f_P(\theta) = 0.34 + 0.46 \sec \theta, \quad (2.2)$$

so that $\sigma = f_P(\theta)\sqrt{S}$. For large signals ($S \geq 20$ VEM) the likelihood is a Gaussian

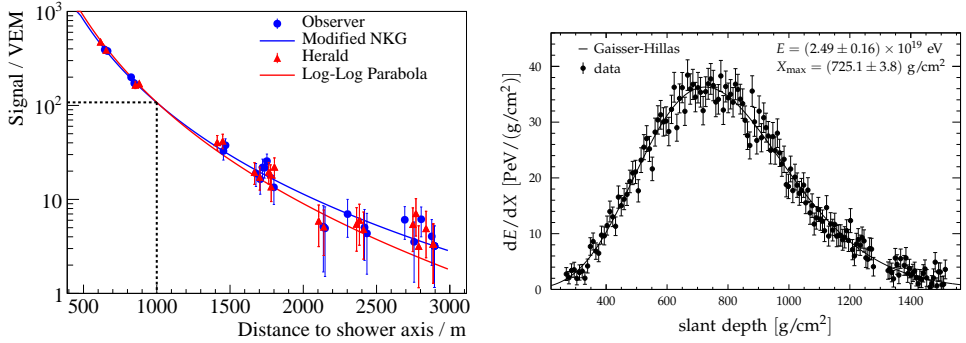


Figure 2.3: On the left: example event detected with the SD with $\theta = 55.2^\circ$ and an energy of 38.7 EeV. The different points are the same event but reconstructed with the Observer or Herald reconstruction [79]. On the right: Another example event detected by the FD. The points are fitted with a Gaisser-Hillas function [82].

with width σ and for small signals ($S < 20$ VEM) it is a scaled Poisson distribution where the number of effective particles is $n = S/f_P^2$.

Energy calibration

The energy calibration is described in [67]. $S(1000)$ is the SD estimator for the energy. However, it depends on the zenith angle and on the mass composition. Before calibrating with the FD energy (see next section) $S(1000)$ is converted into $S(38)$ via the Constant Intensity Cut (CIC) method [81]. The CIC method assumes that the cosmic-ray flux, above a certain threshold, is isotropic so that the probability of detecting a shower with a certain energy is independent of the zenith angle. This allows to correct for the attenuation of $S(1000)$ as a function of θ in a data-driven way. $S(38)$ is calibrated with the energy measurement of the FD via a power law relation. The energy resolution of the SD energy estimator is about 12% at the highest energies. The systematic uncertainty on the energy scale is about 14% [67].

2.2 Fluorescence Detector

The Fluorescence Detector (FD) consists of 4 sites (Los Leones, Los Morados, Coihueco and Loma Amarilla) with each 6 telescopes covering a field of view of $30^\circ \times 30^\circ$. At the Coihueco site there are 3 additional telescopes that point at high elevations to detect lower energy showers. The Los Morados FD station is shown in Fig. 2.2. The fluorescence telescopes allow to observe the longitudinal development of cosmic-ray induced air showers in the atmosphere. Charged particles (mainly electrons and positrons) created in the air shower can excite nitrogen molecules in the atmosphere. These excited molecules emit light in the UV range which can be detected. The amount of measured fluorescence light is proportional to the number of charged particles in the air shower.

The FD is described in detail in [83]. The reconstruction of FD events gives the direction, energy and X_{\max} of the shower. For certain geometries it can be difficult to estimate the shower core and direction with only the FD and so it is important

that the Pierre Auger Observatory is a hybrid detector. Including even only one SD station in the reconstruction significantly improves the FD reconstruction. The fluorescence light as measured by the FD telescopes as a function of time is converted into the energy deposited by the shower as a function of depth. The calorimetric energy is estimated by fitting a Gaisser-Hillas type function (Eq. (1.2)) which also determines X_{max} . The total energy is then the calorimetric energy plus the *invisible energy* that is lost by neutrinos and high energy muons [84]. The resolution on the energy measured by the FD is about 8% and the systematic uncertainty on the energy scale is 9% [85].

Monitoring of atmospheric conditions is key in correctly reconstructing the longitudinal profile. Several instruments at the Pierre Auger Observatory are available to detect clouds, measure the aerosol content of the atmosphere and measure the pressure and temperature on the ground. The Global Data Assimilation System (GDAS) is a global model for the atmosphere at different altitudes [86]. It was verified by balloon measurements at the Pierre Auger Observatory [87].

The fluorescence telescopes can only operate on dark clear nights which makes the duty cycle of FD about 10-15%.

2.3 AugerPrime

Recent results on UHECRs (see Section 1.4) paint a picture that is not easy to interpret from a phenomenological point of view. The predicted flux suppression has been firmly established by the Pierre Auger Observatory and Telescope Array—albeit at slightly different energy—but its nature is not clear, especially in view of the mass composition measurements. There appears to be a light component around the ankle—that is likely to originate from extragalactic sources because of the measurement of the dipole in arrival directions—which transitions into a heavier component above $E = 10^{19}$ eV. On the other hand there are also indications of anisotropy at higher energies which points to the presence of a fraction of light particles, although definitive measurements of the mass composition are absent. To further complicate things, it is clear that hadronic interactions are not well understood from the muon measurements. It is therefore necessary to probe the mass composition (and/or hadronic interactions) at energies beyond what is currently available using more detailed information from the measurements of air showers at the Pierre Auger Observatory.

Because of the steeply falling energy spectrum a measurement of the mass composition at energies around 10^{20} eV with the FD is not possible within a reasonable amount of time. To have enough statistics at the highest energies a measurement of the primary mass using detectors that have a duty cycle of nearly 100% is necessary. Currently, the Pierre Auger Observatory is undergoing an upgrade (‘AugerPrime’) that enhances the mass composition capabilities using the SD [88]. One way this is achieved is by adding Scintillator Surface Detectors (SSDs) on top of the current WCDs. Another is the addition of radio antennas to enhance the mass composition measurement abilities for horizontal showers. Furthermore, new electronics will improve the timing resolution by increasing the sampling frequency from 40 MHz to 120 MHz and the addition of a small PMT extends the dynamic range of the WCD and allows for preciser measurements closer to the shower core.

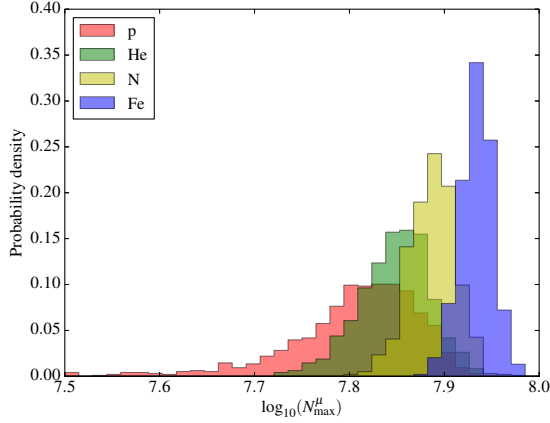


Figure 2.4: Number of muons at the depth where the number of muons is maximum. For showers with a fixed energy of 10 EeV and $\theta = 38^\circ$ simulated with EPOS-LHC as the hadronic interaction model [88].

Scintillator Surface Detector

The SSDs are made of 10 mm thick scintillator bars that together have a total exposed area of 3.78 m^2 . It is placed on top of the WCD tank (see Fig. 2.2 left panel). The scintillator light is read out through 1 PMT via wavelength-shifting fibers (more details are in [89]). The SSD is operated in slave-mode to the WCD, meaning that it is read out only if the WCD is triggered.

The addition of the SSD complements the WCD because it has a different response to electromagnetic particles. Because of its relative thinness the response to (low energy) electrons and positrons is very similar to the response to high energy muons. While the WCD has a different response to electromagnetic particles and muons. The SSD signal is calibrated—in a similar way as the WCD—to the signal of that of a vertical Minimum Ionizing Particle (MIP) which corresponds to a deposited energy of about 2 MeV. The difference in relative responses to muons and electromagnetic particles between the SSD and WCD can be used to infer the muonic and electromagnetic sizes of the air shower via a matrix inversion technique that is further detailed in Chapter 3. The number of muons in a shower (for a given energy) can be used to probe the mass composition (see Fig. 2.4).

Radio Detector

The effective area of the SSD drops rapidly with increasing zenith angle. As such it is only reasonably efficient for showers below $\sim 45^\circ$. To enable mass composition studies for horizontal showers ($\theta > 60^\circ$) each SD station is equipped with a radio antenna [90]. At large inclinations the radio emission from air showers covers a large area on the ground that can be detected by multiple antenna stations [91]. The radio emission originates from the electromagnetic components in the shower (electrons and positrons). For inclined showers, the electromagnetic particles do not reach the ground so that the WCD detects only muons. Again, the ratio of the muon to electromagnetic component can be used to identify the mass of the primary.

3 Signals in water-Cherenkov and scintillator detectors

The goal of the addition of Surface Scintillator Detectors (SSDs) on top of the water-Cherenkov Detectors (WCDs) of the Pierre Auger surface detectors is to improve the decomposition of the signal from extensive air showers into a muonic and an electromagnetic component. What makes this signal separation possible is that the detector response for muons and electromagnetic particles (photons, electrons and positrons) is different for each detector type.

3.1 Matrix formalism

Let us start by writing out the signal in both detectors as the sum of the particle fluxes (in units of m^{-2}) for electromagnetic particles (\mathcal{F}^{em}) and muons (\mathcal{F}^μ) and ignore other particles. The particle flux should be multiplied by the projected surfaces of each detector $A_{\text{det}}(\chi)$, with χ the angle with respect to the zenith of the incoming particles*, and a number a - d representing the detector responses. In matrix form this reads:

$$\begin{pmatrix} S_{\text{WCD}} \\ S_{\text{SSD}} \end{pmatrix} = \begin{pmatrix} \mathcal{A}_{\text{WCD}}(\chi) & 0 \\ 0 & \mathcal{A}_{\text{SSD}}(\chi) \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \mathcal{F}^{\text{em}} \\ \mathcal{F}^\mu \end{pmatrix}. \quad (3.1)$$

To reduce the number of parameters it is convenient to decompose the signal into the signal from muons and electromagnetic particles in the WCD so that:

$$\begin{pmatrix} S_{\text{WCD}} \\ S_{\text{SSD}} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ \alpha & \beta \end{pmatrix} \begin{pmatrix} S_{\text{WCD}}^{\text{em}} \\ S_{\text{WCD}}^\mu \end{pmatrix} \quad (3.2)$$

and

$$\alpha = \frac{S_{\text{SSD}}^{\text{em}}}{S_{\text{WCD}}^{\text{em}}} = \frac{\mathcal{A}_{\text{SSD}}(\chi)}{\mathcal{A}_{\text{WCD}}(\chi)} \frac{c}{a} \text{ MIP/VEM}, \quad (3.3)$$

$$\beta = \frac{S_{\text{SSD}}^\mu}{S_{\text{WCD}}^\mu} = \frac{\mathcal{A}_{\text{SSD}}(\chi)}{\mathcal{A}_{\text{WCD}}(\chi)} \frac{d}{b} \text{ MIP/VEM}. \quad (3.4)$$

To decompose the signal into a muonic and electromagnetic component one can invert Eq. (3.2) to get:

$$\begin{pmatrix} S_{\text{WCD}}^{\text{em}} \\ S_{\text{WCD}}^\mu \end{pmatrix} = \frac{1}{\beta - \alpha} \begin{pmatrix} \beta & -1 \\ -\alpha & 1 \end{pmatrix} \begin{pmatrix} S_{\text{WCD}} \\ S_{\text{SSD}} \end{pmatrix} \quad (3.5)$$

* χ is used as the angle of the incoming particles, which for showers starting at infinity coincides with the incoming zenith angle of the shower usually denoted by θ .

PMT1	PMT2	PMT3	PMT5
1606.17	1606.42	1606.81	152.37

Table 3.1: Numbers to convert FADC counts to VEM or MIP charge. Taken from detector simulations with `Offline` revision 33544 with upgraded electronics.

and one needs to know the values for α and β . For now these should be inferred from simulations as was done in [88, 92].[†] Unfortunately, physical intuition for these parameters can be obscured. In fact, the parameters α and β are not only determined by the detector configuration but also by the energy and incoming angle of the particles.

In the rest of this chapter the values that should go into the matrices Eqs. (3.1) and (3.2) and what physics determines these values will be discussed.

3.2 The detector response

In the following it was assumed that both detectors are properly calibrated, meaning that high energy muons going vertically through the SSD and WCD give a signal that is denoted as 1 MIP and 1 VEM respectively.

To determine the detector response, single particles hitting an upgraded Pierre Auger surface detector station were simulated using `Offline` revision 33544 with the module `G4StationSimulator0G`. The numbers for converting FADC counts to VEM and MIP charges are shown in Table 3.1.

Energy response

Figure 3.1 shows the average signal deposited as a function of the incoming particle kinetic energy for particles going vertically through the SSD and the WCD. Particles were injected at a position with $(x, y) = (0.99, -0.36)$ meters with respect to the center of the tank so that the SSD was hit in the middle of one of the scintillator bars. Also simulations were done without the SSD. Including the SSD has a small effect on the WCD signal for very low energy particles.

The WCD response can be summarized as follows:

- Muons above a certain threshold ($E_{\text{VEM}} \approx 350 \text{ MeV}$ kinetic energy) give a signal of 1 VEM. Below this threshold muons deposit all of their energy in the water proportional to approximately E^2 . One may notice from Fig. 3.1 that the dependence is steeper than E^2 below 100 MeV, but that there are also some muons that decay in (or right above) the tank and the electron with average energy $\approx 45 \text{ MeV}$ gives a signal of about 0.1 VEM.
- Electrons (positrons are assumed to behave similarly) and photons are stopped in the water and consequently give a signal linearly depending on their kinetic energy. Note that at low energies some particles do not make it into the tank and at high kinetic energy some energy is leaked out of the tank.

[†]In [92] the values $a = \frac{\alpha}{\alpha+1}$ and $b = \frac{\beta}{\beta+1}$ are used.

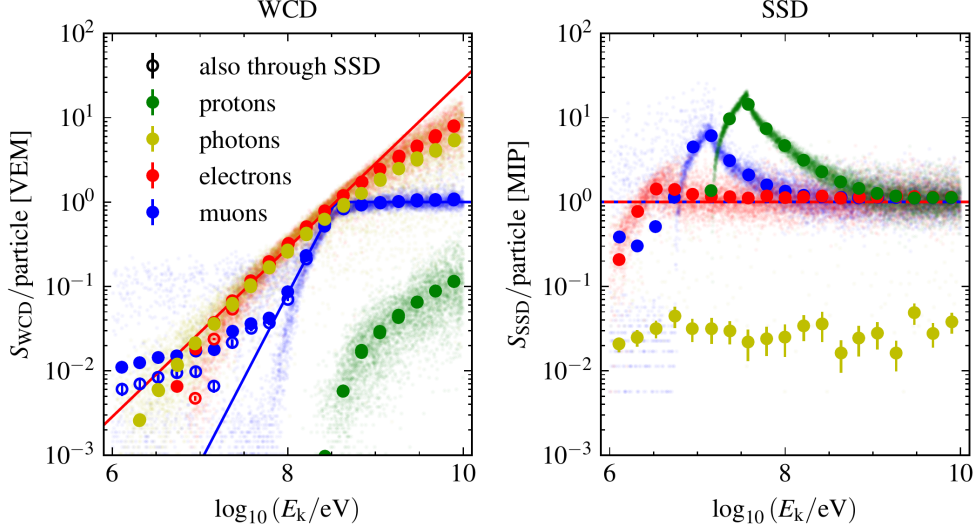


Figure 3.1: Calibrated signal versus kinetic energy for different particle species simulated going vertically through the WCD (left) and SSD (right). The cloud of small points are individual simulations while the large points are the average in equally sized bins.

The SSD response is much simpler: it is a thin scintillator detector and all charged particles above a certain energy create the same signal. At low kinetic energies the energy loss goes like $1/v^2$ (with v the velocity of the particles) from the Bethe-Bloch formula, hence the enhancement of the signal in that region. This is ignored at this point since the number of muons at this energy is expected to be negligible and the effect for electrons is small. Also notice that a few percent of the photons convert to an electron-positron pair and give a signal in the SSD. For completeness the signal for protons is also shown and one may appreciate that the signal in the SSD per particle can be quite substantial leading to large spikes in the signal trace. However, protons (and all other particles) are not taken into account in the following for simplicity and because they only make up a small fraction of the total number of particles.

The mean of the signal per particle is slightly larger than 1 VEM or MIP, this is because the calibration is based on the *muon hump* (i.e. mode) of the calibration histogram. Because of the skewed distribution of the signal (which should follow a Landau distribution) the mean is larger than the mode. This has a much larger effect on the SSD signal (the difference is a factor 1.13) while for the WCD it is negligible which can be seen in [Figure 3.2](#).

Tracklength

To study the effect of the incoming direction, particles were simulated with a random position sampled on a disk above the station and this disk was rotated from 0 to 60 degrees with respect to the z axis. In this way all possible *tracklengths* of particles crossing the detector are acquired (up to 60°). The energy of the particles was fixed to a kinetic energy of 2 GeV for muons (so that they are all through-going) and 100 MeV

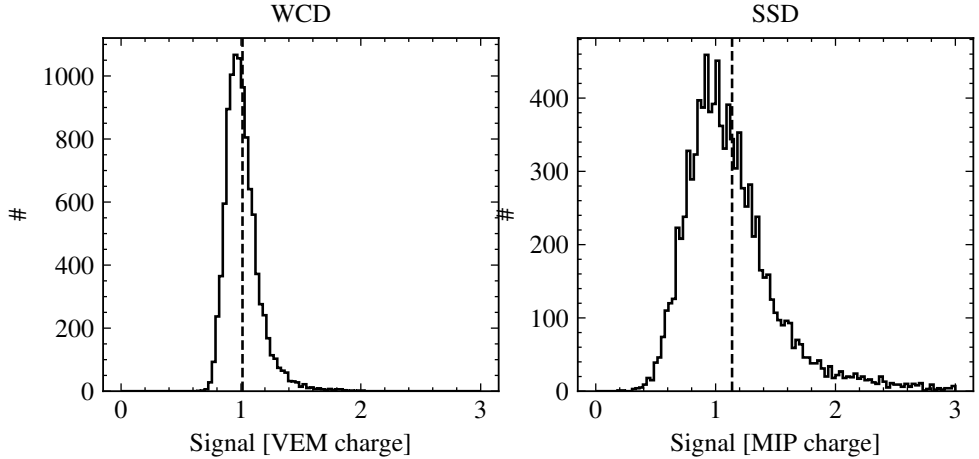


Figure 3.2: Simulated signal distribution in the WCD (left) and the SSD (right) for vertical muons of $E_k = 2\text{ GeV}$. The dashed vertical line shows the mean of the distribution.

for electrons and photons (so that they are mostly stopped).

Figure 3.3 shows the average signal versus the incoming zenith angle of the particles. The average tracklength in the WCD is simply the volume of the tank divided by the projected area under an angle χ . The normalized average signal in the WCD as a function of χ then reads:

$$\langle S_{\text{WCD}/\text{muon}} \rangle(\chi) = \langle L_{\text{WCD}} \rangle(\chi)/H = \frac{1}{|\cos \chi| + \frac{2H}{\pi R} \sin \chi} \text{VEM}, \quad (3.6)$$

where $H = 1.2\text{ m}$ and $R = 1.8\text{ m}$ are, respectively, the height and the radius of the tank. The fluctuations of the tracklength in the WCD increase rapidly for $\chi > 0$, and the mean is actually a poor representation of the distribution. For more details on the distributions of tracklengths in the WCD see [93].

The normalized average signal in the SSD —approximating it as an infinitely large and thin plane— is simply:

$$\langle S_{\text{SSD}/\text{particle}} \rangle(\chi) = \langle L_{\text{SSD}} \rangle(\chi)/d = \sec \chi \text{ MIP}, \quad (3.7)$$

with $d = 10\text{ mm}$ the thickness of the scintillator bars. In the following a simpler notation is adopted such that $L_{\text{det}}(\chi)$ should be read as the average signal per muon due the average tracklength in units of VEM or MIP.

A toy model for the detector response

A simple toy model for the detector response in the WCD and SSD can now be constructed. The average signal in the WCD and SSD for muons and electromagnetic particles is simply the integral over the energy distribution of particles and the detector response.

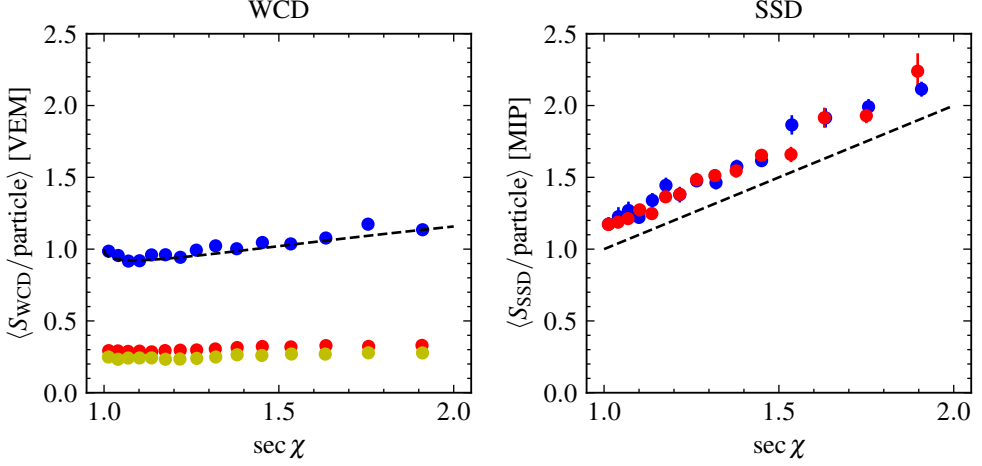


Figure 3.3: Average calibrated signal versus $\sec \chi$, where χ is the angle with respect to the zenith of the incoming particle. The color coding is the same as in Fig. 3.1. The dashed lines show the theoretical expectation for the mean (see text). Notice again the offset of 1.13 for the mean of the SSD signal as noted in the text.

For the muon part:

$$\begin{aligned} S_{\text{WCD}}^{\mu} &= \mathcal{A}_{\text{WCD}}(\chi) \int dE \frac{d\mathcal{F}^{\mu}}{dE} \mathcal{R}_{\text{WCD}}^{\mu}(E, \chi), \\ S_{\text{SSD}}^{\mu} &= \mathcal{A}_{\text{SSD}}(\chi) \int dE \frac{d\mathcal{F}^{\mu}}{dE} \mathcal{R}_{\text{SSD}}^{\mu}(E, \chi), \end{aligned} \quad (3.8)$$

and for the electromagnetic part:

$$\begin{aligned} S_{\text{WCD}}^{\text{em}} &= \mathcal{A}_{\text{WCD}}(\chi) \left[\int dE \frac{d\mathcal{F}^e}{dE} \mathcal{R}_{\text{WCD}}^e(E, \chi) + \int dE \frac{d\mathcal{F}^{\gamma}}{dE} \mathcal{R}_{\text{WCD}}^{\gamma}(E, \chi) \right], \\ S_{\text{SSD}}^{\text{em}} &= \mathcal{A}_{\text{SSD}}(\chi) \left[\int dE \frac{d\mathcal{F}^e}{dE} \mathcal{R}_{\text{SSD}}^e(E, \chi) + \int dE \frac{d\mathcal{F}^{\gamma}}{dE} \mathcal{R}_{\text{SSD}}^{\gamma}(E, \chi) \right]. \end{aligned} \quad (3.9)$$

Here $\frac{d\mathcal{F}^{\text{comp}}}{dE}$ denotes the differential particle flux. Note that \mathcal{F}^e also includes positrons. $\mathcal{R}_{\text{det}}^{\text{comp}}$ denotes the detector response that explicitly depends on the energy and incoming angle of the particles. The detector responses in VEM and MIP, considering Section 3.2, can be approximated as:

$$\begin{aligned} \mathcal{R}_{\text{WCD}}^{\mu} &= \min \left[\left(\frac{E}{E_{\text{VEM}}} \right)^2, L_{\text{WCD}}(\chi) \right], \\ \mathcal{R}_{\text{WCD}}^e &= \mathcal{R}_{\text{WCD}}^{\gamma} = \frac{E}{E_{\text{VEM}}}, \\ \mathcal{R}_{\text{SSD}}^{\mu} &= \mathcal{R}_{\text{SSD}}^e = L_{\text{SSD}}(\chi), \\ \mathcal{R}_{\text{SSD}}^{\gamma} &= \epsilon_{\text{SSD}}^{\gamma}, \end{aligned} \quad (3.10)$$

where $\epsilon_{\text{SSD}}^\gamma$ is the fraction of photons that convert to electron-positron pairs and create a signal in the SSD, which is assumed to be independent of energy. Combining Eqs. (3.8) and (3.9) and Eq. (3.10) then gives for the muon part:

$$S_{\text{WCD}}^\mu = \mathcal{A}_{\text{WCD}}(\chi) \int dE \frac{d\mathcal{F}^\mu}{dE} \min \left[\left(\frac{E}{E_{\text{VEM}}} \right)^2, L_{\text{WCD}}(\chi) \right], \quad (3.11)$$

$$S_{\text{SSD}}^\mu = \mathcal{A}_{\text{SSD}}(\chi) L_{\text{SSD}}(\chi) \mathcal{F}^\mu$$

and for the electromagnetic part:

$$S_{\text{WCD}}^{\text{em}} = \mathcal{A}_{\text{WCD}}(\chi) \left[\frac{\langle E^e \rangle}{E_{\text{VEM}}} \mathcal{F}^e + \frac{\langle E^\gamma \rangle}{E_{\text{VEM}}} \mathcal{F}^\gamma \right], \quad (3.12)$$

$$S_{\text{SSD}}^{\text{em}} = \mathcal{A}_{\text{SSD}}(\chi) L_{\text{SSD}}(\chi) [\mathcal{F}^e + \epsilon_{\text{SSD}}^\gamma \mathcal{F}^\gamma].$$

The values that go in the matrix in Eq. (3.1) are thus given by:

$$a = \frac{\langle E^e \rangle}{E_{\text{VEM}}} f^e + \frac{\langle E^\gamma \rangle}{E_{\text{VEM}}} f^\gamma, \quad (3.13)$$

$$b = \frac{\langle (E_1^\mu)^2 \rangle}{E_{\text{VEM}}^2} + f_2^\mu L_{\text{WCD}}(\chi), \quad (3.14)$$

$$c = L_{\text{SSD}}(\chi) [f^e + \epsilon_{\text{SSD}}^\gamma f^\gamma], \quad (3.15)$$

$$d = L_{\text{SSD}}(\chi), \quad (3.16)$$

where f^e and f^γ denote the fraction of electrons/positrons and photons respectively over the total number of electromagnetic particles. The integral over the energy spectra was evaluated for muons so that

$$\langle (E_1^\mu)^2 \rangle = \int_{E_{\text{CH}}}^{E_{\text{VEM}}} dE \left(\frac{E}{E_{\text{VEM}}} \right)^2 \frac{d\mathcal{F}^\mu}{dE} \bigg/ \mathcal{F}^\mu, \quad (3.17)$$

$$f_2^\mu = \int_{E_{\text{VEM}}}^{\infty} dE \frac{d\mathcal{F}^\mu}{dE} \bigg/ \mathcal{F}^\mu, \quad (3.18)$$

where $E_{\text{CH}} = 54 \text{ MeV}$ is the Cherenkov threshold in water and \mathcal{F}^μ is the total muon flux.

The signal ratios α and β can now be calculated. The ratio of the electromagnetic signal in the SSD over the WCD is

$$\alpha = \frac{S_{\text{SSD}}^{\text{em}}}{S_{\text{WCD}}^{\text{em}}} = \frac{\mathcal{A}_{\text{SSD}}(\chi)}{\mathcal{A}_{\text{WCD}}(\chi)} \frac{c}{a} = \mathcal{G} L_{\text{WCD}}(\chi) \frac{1 + r^\gamma \epsilon_{\text{SSD}}^\gamma}{\langle E^e \rangle / E_{\text{VEM}} + r^\gamma \langle E^\gamma \rangle / E_{\text{VEM}}}, \quad (3.19)$$

with $r_\gamma = f^\gamma / f^e$. And the signal ratio for muons:

$$\beta = \frac{S_{\text{SSD}}^\mu}{S_{\text{WCD}}^\mu} = \frac{\mathcal{A}_{\text{SSD}}(\chi)}{\mathcal{A}_{\text{WCD}}(\chi)} \frac{d}{b} = \mathcal{G} \frac{L_{\text{WCD}}(\chi)}{\langle (E_1^\mu)^2 \rangle / E_{\text{VEM}}^2 + f_2^\mu L_{\text{WCD}}(\chi)}. \quad (3.20)$$

CHAPTER 3. SIGNALS IN WATER-CHERENKOV AND SCINTILLATOR DETECTORS

The geometrical factor \mathcal{G} is simply the surface of the SSD divided by the surface of the WCD as seen from $\chi = 0^\circ$:

$$\mathcal{G} = \frac{A_{\text{SSD}}(\chi)L_{\text{SSD}}(\chi)}{A_{\text{WCD}}(\chi)L_{\text{WCD}}(\chi)} = 0.378 \text{ m}^2 \text{ MIP/VEM}. \quad (3.21)$$

This is useful because on average the χ dependence of the projected surface and the average signal due to the tracklength cancels, with the downside that this leaves a factor of $L_{\text{WCD}}(\chi)$ because the electromagnetic signal and the low energy muon signal in the WCD do not depend on χ , because these particles are stopped in the tank. Note that $L_{\text{WCD}}(\chi)$ is relatively constant from [Fig. 3.3](#).

3.3 The energy spectrum of secondary particles

In the above it was shown that, to first order, the signal in the WCD depends on the energy spectrum ($\frac{dF^{\text{comp}}}{dE}$) of the particles, while that in the SSD does not and particles are counted. This is a consequence of the thick water tank in which particles can be completely stopped and the thin scintillator bars which most particles above a certain threshold pass through.

With the simplified detector response from the previous section there are thus three parameters for the electromagnetic signal that depend on the physics of the air shower: the average photon energy $\langle E^\gamma \rangle$, the average electron energy $\langle E^e \rangle$ and the number ratio of photons to electrons r^γ . There are two parameters for the muon signal: the second moment of the energy below E_{VEM} denoted as $\langle (E_1^\mu)^2 \rangle$ and the fraction of muons above E_{VEM} with respect to the total number of muons: f_2^μ .

The energy spectrum of particles at ground was determined from simulated air showers with CORSIKA 7.5600. The log of the kinetic energy of the different particle types was binned with respect to the log of the distance to the shower core r . An example of such an energy spectrum is shown in [Fig. 3.4](#). One important thing should be mentioned: by default muons are not followed anymore below 100 MeV in CORSIKA (for example in the Napoli library used in [Chapter 5](#) this is the case). In the simulations shown here the muon energy cutoff was set to 10 MeV, for hadrons at 50 MeV and for electrons/positron and photons at 0.5 MeV.

[Figure 3.5](#) shows the normalized energy spectrum so that one can appreciate the changes when moving farther away from the shower core and changing the zenith angle. For muons going farther away from the core or towards a more vertical shower leads to a softening of the spectrum. The energy spectrum of muons can be reasonably well modeled by considering the path length traveled by the muons and the fact that lower energy muons have larger transverse momentum as will be discussed in [Chapter 4](#). For electrons and photons the effect of the distance to the core and the zenith angle is more subtle.

Parameters for the toy model

With the energy spectra from CORSIKA simulations we can now determine the parameters for calculating α and β . For example from the spectrum shown in [Fig. 3.4](#)

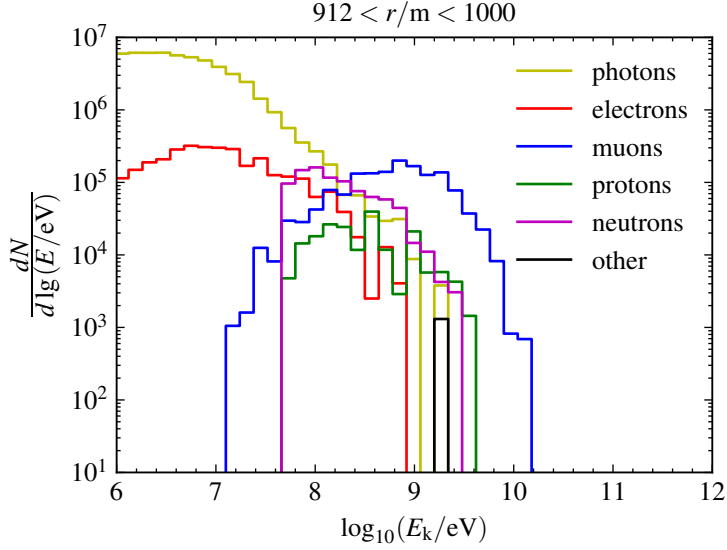


Figure 3.4: Energy spectrum of particles at ground, 1000 m from the core, for a simulated proton shower with EPOS-LHC with a primary energy of 10 EeV and $\theta = 0^\circ$.

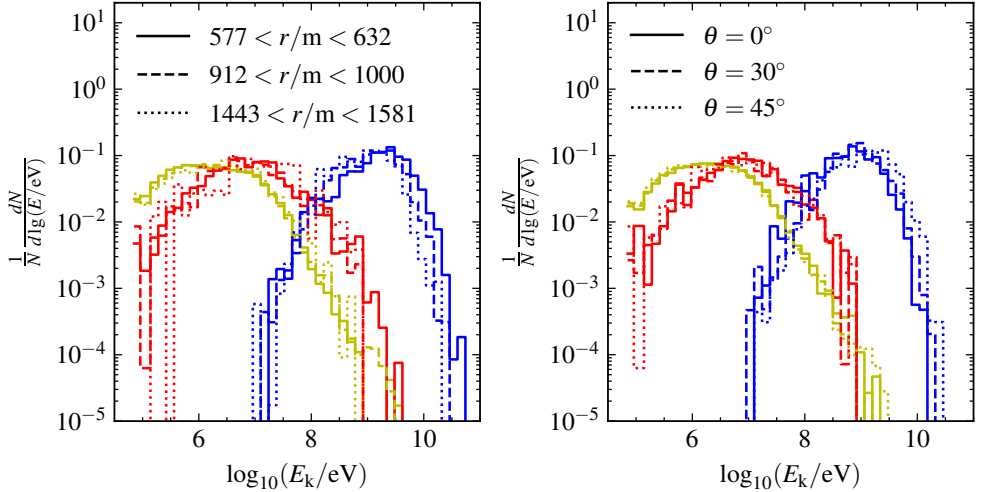


Figure 3.5: Normalized energy spectrum for different distances to the shower core (left) and different zenith angles (right). Proton showers with EPOS-LHC and a primary energy of 10 EeV. On the left $\theta = 0^\circ$ and on the right $912 < r/m < 1000$. The color coding is the same as in Fig. 3.4.

we get for the electromagnetic part

$$\langle E^e \rangle \approx 25 \text{ MeV}, \quad (3.22)$$

$$\langle E^\gamma \rangle \approx 7 \text{ MeV} \quad (3.23)$$

$$r^\gamma \approx 25 \quad (3.24)$$

and for the muonic part:

$$\frac{\langle (E_1^\mu)^2 \rangle}{E_{\text{VEM}}^2} \approx 0.11, \quad (3.25)$$

$$f_2^\mu \approx 0.7, \quad (3.26)$$

with $E_{\text{VEM}} = 350 \text{ MeV}$. Where all energies correspond to the kinetic energy. The five parameters all depend on the distance to the shower core as can be seen in Figs. 3.6 and 3.7. The number of photons diminishes at larger distances while the average energy of photons and electrons grows slightly. For muons the spectrum becomes softer at large r and thus $\langle (E_1^\mu)^2 \rangle$ grows while f_2^μ declines.

For muons there is also a clear dependence on the zenith angle of the shower (Fig. 3.8). More inclined showers traverse more atmosphere in which low energy muons are stopped and only high energy muons survive, leading to a harder spectrum and thus smaller $\langle (E_1^\mu)^2 \rangle$ and larger f_2^μ . The parameters of the electromagnetic part are not substantially affected by the zenith angle up to 45° (Fig. 3.9).

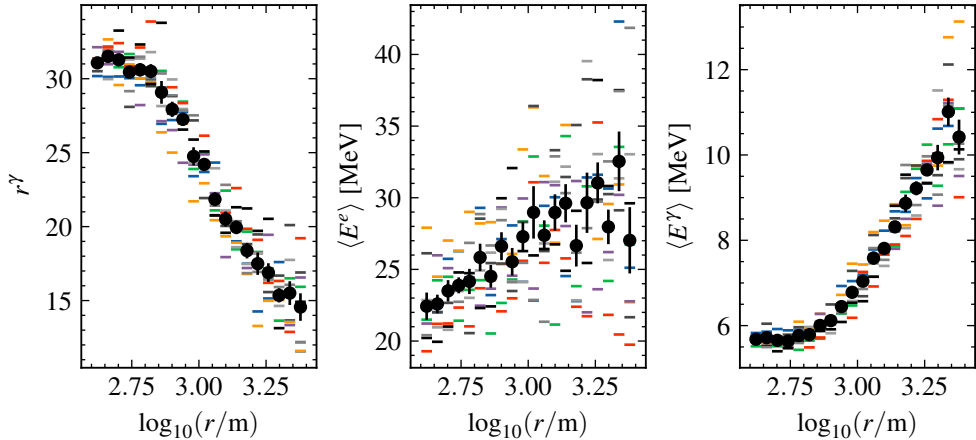


Figure 3.6: The number ratio of photons to electrons (left), average energy of electrons (middle) and photons (right) versus distance to the shower core. Vertical proton showers with $E = 10 \text{ EeV}$. The black error bars show the mean and standard error of the mean of 10 showers (shown in colors).

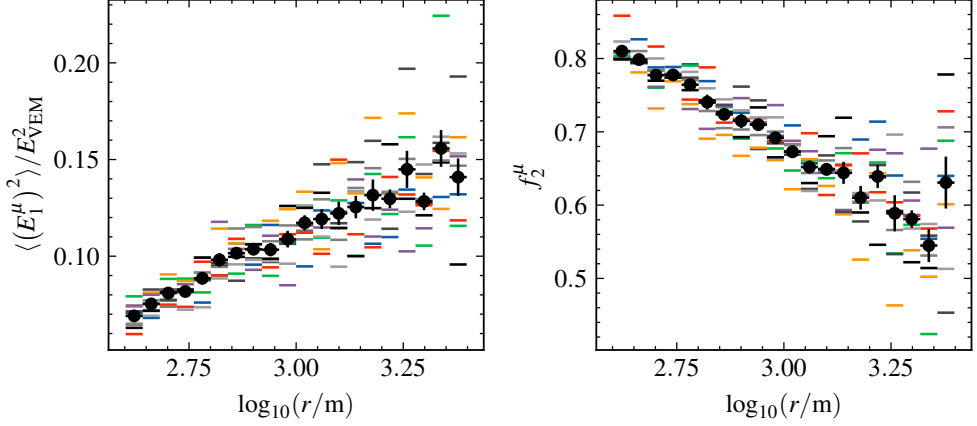


Figure 3.7: The second moment of the muon energy spectrum below $E_{\text{VEM}} = 350$ MeV (left) and the fraction of muons above E_{VEM} (right) as appear in Eq. (3.20). Same showers as in Fig. 3.6. Notice that because the spectrum becomes softer for larger distances the energy below E_{VEM} (left) grows with r and the fraction of muons above E_{VEM} (right) goes down with increasing r .

3.4 The signal ratios

The parameters as determined from the energy spectrum can now be used to calculate the signal ratios α and β with Eqs. (3.19) and (3.20) to determine their dependence on the distance to the shower core and the zenith angle. Values for the detector response are chosen as $E_{\text{VEM}} = 350$ MeV (best fit from Fig. 3.1) and $\epsilon_{\text{SSD}}^\gamma = 0.03$ and a factor of 1.13 is used to convert the mode of the SSD signal to the mean. The average signal due to the tracklength in the WCD that appears in Eqs. (3.19) and (3.20) was set to 1, since it is not clear what the exact value for χ should be and the difference of the mean signal up to 55° is within 10%. The resulting α and β as a function of r and θ are shown in Figures 3.10 and 3.11 respectively.

Comparing to simulated signals

Because α and β are simply the ratios of the signal in the SSD and the WCD for the electromagnetic and muon component respectively, the toy model from the above section can be compared with full detector simulated signals of the same CORSIKA simulations. Offline revision 33544 with GEANT4 was used to simulate the upgraded stations (with UUB and SSD)[‡] with 36 dense stations at every 100 meter. The average signal from those 36 stations was calculated—a non-triggered station was treated as having no signal—and α and β are calculated from these averages. Figure 3.12 shows the comparison as a function of r for EPOS-LHC proton showers with $\theta = 0^\circ$ to mitigate the effect of the tracklength.

A few remarks should be made on this comparison. Far away from the core the signal is very prone to fluctuations (even with 36 dense stations) and this very

[‡]Offline application SdSimulationReconstructionUpgrade.

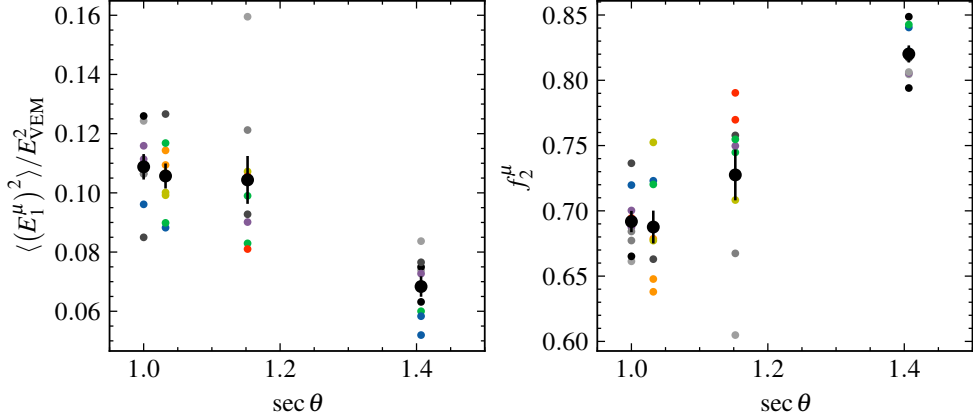


Figure 3.8: The second moment of the muon energy spectrum below $E_{\text{VEM}} = 350$ MeV (left) and the fraction of muons above E_{VEM} (right) as appear in Eq. (3.20) versus the zenith angle θ at a distance between $912 < r/m < 1000$. Proton induced showers with $E = 10$ EeV. The colored dots show individual showers and the black markers with error bars the mean of those showers.

much influences the ratio of the SSD to the WCD signal because the SSD is almost three times smaller. The tracklength will also play a bigger role at larger distances because the incoming particle angle is different from the zenith angle of the shower. Also the fluctuations of the tracklength in the WCD increase rapidly [93]. Finally, a trigger effect is present since the SSD runs in slave mode to the WCD and significant electromagnetic signal may be more easily missed by the WCD (which is biased to a large muon signal). Closer to the core low-gain saturation plays a role. The window to compare the toy model is then approximately between $600 < r/m < 1400$ at 10 EeV.

It is clear from Figs. 3.12 and 3.13 that, even in this distance range, there is an offset for both α and β compared to the toy model. The SSD signal from the toy model (both muon and electromagnetic) should be multiplied by a factor ~ 1.2 , or the WCD divided by the same number, to reproduce the true-MC simulated signal. The rudimentary detector response that is used in the toy model could be responsible. The SSD response is not independent of energy as was assumed: low energy particles will create a signal larger than 1 MIP (see Fig. 3.1). Furthermore, setting $E_{\text{VEM}} = 400$ MeV brings the toy model values for α close to the simulated signal, but β is not much affected. Changing E_{CH} has no significant effect.

The relation of α and β with respect to the distance to the core up to $r \approx 1200$ m is quite well reproduced (for larger r a trigger effect is suspected which biases the WCD to larger signals). The decline of α with respect to r can now be understood from the increase of the average energy of electrons and photons (see Fig. 3.6), generating a larger WCD signal, while the SSD signal is not affected. The increase of β as a function of r is caused by the increase of the ratio of low to high energy muons (see Fig. 3.7), which do not traverse the whole tank creating a smaller WCD signal while the SSD signal is again not affected.

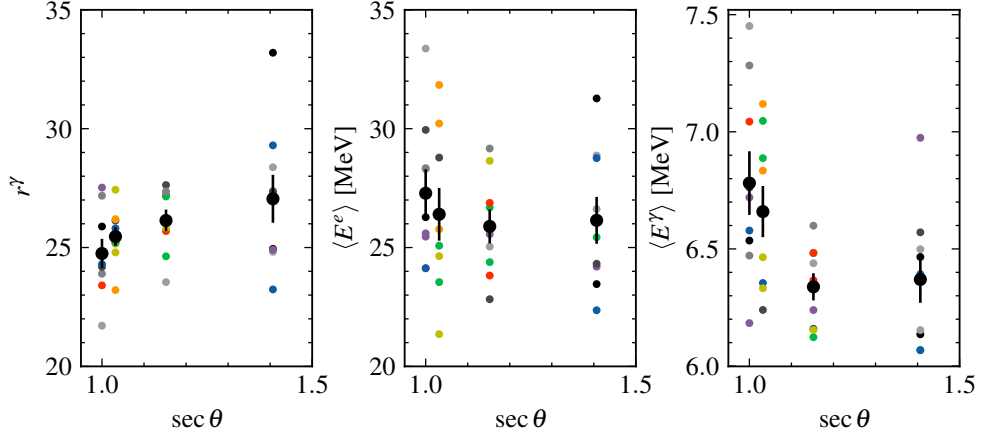


Figure 3.9: The ratio of photons to electrons (left), average energy of electrons (middle) and photons (right) versus the zenith angle θ at a distance between $912 < r/m < 1000$. Proton induced showers with $E = 10$ EeV.

Influence of primary mass and hadronic interaction model

To use matrix inversion (Eq. (3.5)) to infer the muon component of extensive air showers and subsequently use that for a mass composition or hadronic physics study it is necessary that the parameters α and β as discussed in this note are relatively independent of the cosmic-ray primary and the hadronic interaction model used. Proton and iron induced showers simulated with EPOS-LHC and QGSJET-II.04 have been compared. The average values for α and β (calculated from Offline simulated showers, not the toy model) from 10 vertical showers at 1000 m (36 stations) are shown in Fig. 3.14. The 5 parameters from the energy spectrum are shown in Fig. 3.15. Within the limited statistics no significant differences can be seen.

3.5 Conclusion

The goal of this chapter was to understand the detector responses of the WCD and SSD to muons and electromagnetic particles. The parameters that determine the relative signal between the two detectors contributed by these two particle components are essential if one wants to determine the number of muons in extensive air showers with the upgraded surface detector stations. It was shown that these parameters are a convolution of the detector response and the energy spectrum of the particles on the ground. The electromagnetic signal ratio is generated by the average energy of photons and electrons and the number ratio between the two. The muonic part is determined by the energy of muons that create less than a VEM in the WCD and the number of muons that do create a VEM. These 5 parameters turn out to depend on the inclination of the shower and on the distance to shower core.

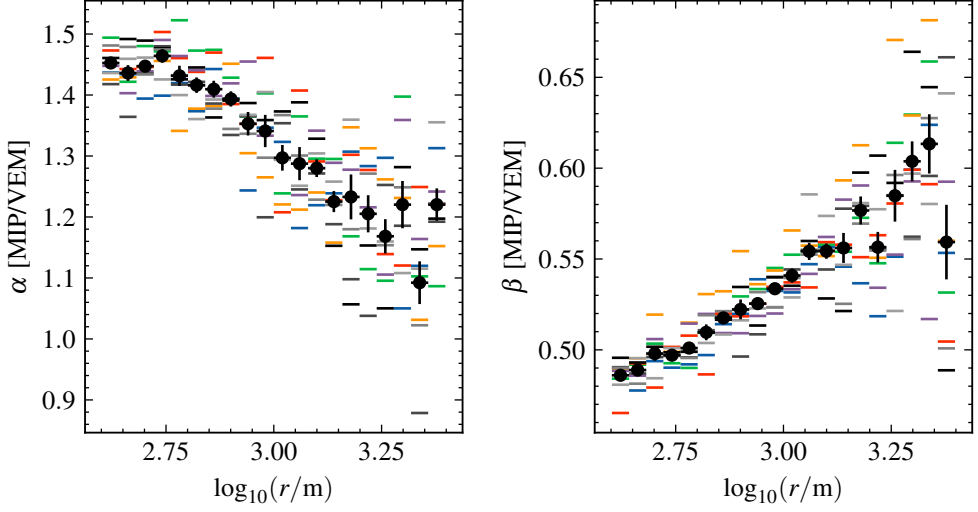


Figure 3.10: The signal ratios as calculated by the toy model α (left) and β (right) versus the distance to the shower core for a vertical proton shower. Calculated with Eqs. (3.19) and (3.20) using the values shown in Figs. 3.6 and 3.7.

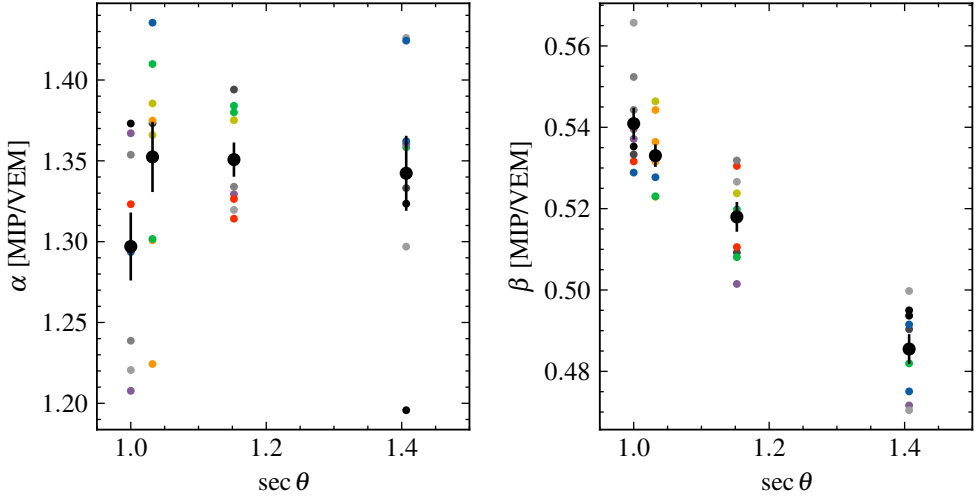


Figure 3.11: The signal ratios as calculated by the toy model α (left) and β (right) versus the zenith angle θ at $912 < r/m < 1000$. Calculated by Eqs. (3.19) and (3.20) using the values shown in Figs. 3.8 and 3.9.

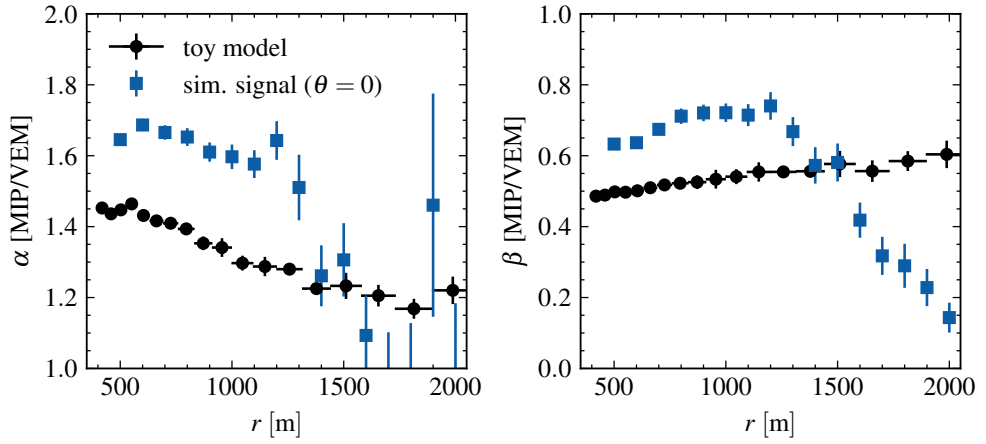


Figure 3.12: Comparison of α and β from the toy model and from the simulated signal versus r for the same 10 vertical proton showers with a primary energy of 10 EeV. The average signal over 36 dense stations is shown. For $r \gtrsim 1200$ m a trigger effect is apparent as explained in the text.

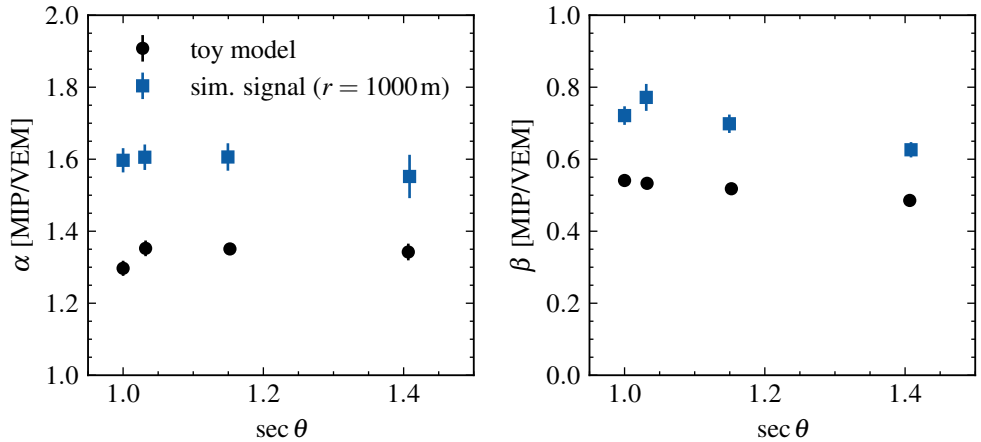


Figure 3.13: Comparison of α and β from the toy model and from the simulated signal at approximately 1000 m versus the zenith angle.

CHAPTER 3. SIGNALS IN WATER-CHERENKOV AND SCINTILLATOR DETECTORS

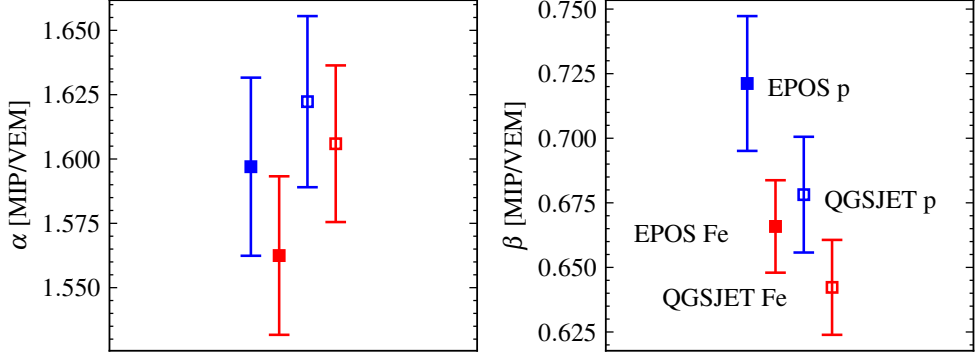


Figure 3.14: α and β comparison between primaries (blue=proton and red=iron) and hadronic interaction models (filled=EPOS-LHC and empty=QGSJET-II.04) at $r = 1000$ m and $\theta = 0$.

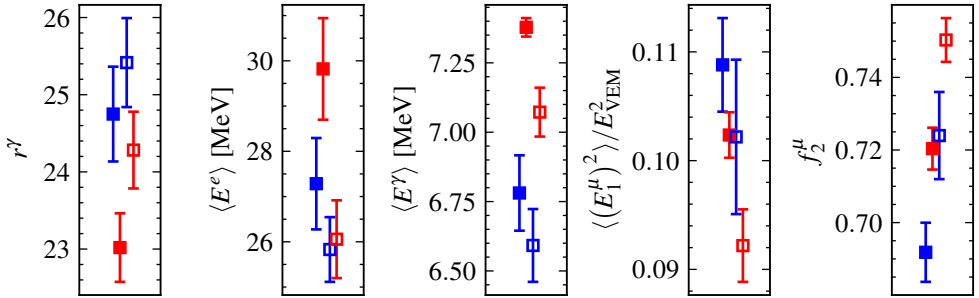


Figure 3.15: The 5 parameters that determine α and β via the toy model in Eqs. (3.19) and (3.20). The markers have the same labeling as in Fig. 3.14.

4 Time distribution of signal in surface particle detectors

In surface detector arrays for detecting cosmic rays the arrival time of secondary particles is routinely used to infer the direction of the primary cosmic ray. This is commonly achieved by fitting a plane or curved front to the start time of the signal, which either assumes that all secondary particles come from infinity or from a single point. Neither is very close to the reality of multiple particles being created over an extended length near the shower axis and subsequently also interacting with the atmosphere. Nevertheless, the approximation of a curved shower front is usually sufficient to estimate the arrival direction within a few degrees for UHECR induced showers [79].

Because particles in an air shower are not created in a single point, there is considerable spread in the arrival times of particles as was first systematically studied by Linsley in 1961 [10]. The shower front has a thickness of a few 100 ns up to several μ s depending on the zenith angle, energy and nature of the primary cosmic ray. It was shown by Watson & Wilson in 1974 that the thickness of the shower front can also be used to infer the mass composition [94]. Their method relied on the so called *rise time*, which is the time between when the signal reaches 10% and 50% of its total signal. A similar approach was recently used by the Pierre Auger Observatory to infer the mass composition and study hadronic interactions [95].

To understand the thickness of the shower front and exploit all information available in particle detectors on the ground it is necessary to understand the full arrival time distribution of particles. There have already been several (semi-)empirical attempts to describe the functional form of the time distribution of the signal in WCDs [65, 96–98]. However, it is possible to directly describe the arrival time of particles as a function of the geometry and longitudinal development of the air shower under the assumption that these particles travel in straight lines, which is a good approximation for muons.

4.1 Muon production depth

Because muons travel approximately unhindered by the atmosphere it is possible to point them back to their point of production by simple geometry (as will be shown in the next subsection). Already in 1962 Beer et al. [99] performed a measurement of the muon production depth (MPD) using layered spark counters which allowed the reconstruction of the direction of the ‘mu-mesons’.

More recently, the relation of the MPD and the arrival times of muons has been studied [100–102]. The Pierre Auger Observatory has measured the maximum of the muon production depth, X_{\max}^{μ} , in air showers induced by $E > 2 \times 10^{19}$ eV cosmic rays shown in Fig. 4.1 [103]. The mass composition as inferred from X_{\max}^{μ} is in tension with the mass composition as inferred from X_{\max} considering the current hadronic interaction models.

However, X_{\max}^{μ} is a ‘tricky quantity’ [104]. If one tries to measure X_{\max}^{μ} (or study it from MC simulations) from the muons that reach the ground, what one acquires is what is sometimes referred to as the *apparent* MPD. The apparent MPD is deformed by muons that have decayed before reaching the ground and thus are not measured (or not saved in simulations). The X_{\max}^{μ} as inferred from the apparent MPD is closer to the ground than if it would be inferred from the *true* MPD, which can only be acquired by performing simulations and storing the position of the creation of every muon. What is cumbersome is that the apparent MPD depends on the minimum energy of the muons the detector is sensitive to and—because the probability of muon decay depends on the length of its traveled path—on the inclination of the shower and on the position of the detector with respect to the shower axis.

4.2 Derivation of the arrival time distribution of muons

It is interesting to follow the derivation of the arrival time distribution of muons as was developed in [100, 101], because it gives some intuition on how the arrival time is related to the longitudinal profile of a shower.

When assuming that the particles are created on the shower axis and travel in straight lines with the speed of light it is straightforward to convert their arrival time on the ground with respect to a plane front to their point of creation z :

$$ct = ct_1 - ct_{\text{pf}} = l - (z - \Delta) = \sqrt{r^2 + (z - \Delta)^2} - (z - \Delta), \quad (4.1)$$

see Fig. 4.2 for a sketch of the situation. The correction to z is $\Delta = r \tan \theta \cos \psi$. Equation (4.1) can be inverted to obtain

$$z(t) = \frac{1}{2} \left(\frac{r^2}{ct} - ct \right) + \Delta, \quad (4.2)$$

where t now refers to the time with respect to the plane front and this convention will be adopted in the following.

Using Eq. (4.2) it is possible to convert the arrival time to a point in atmosphere, which can be used to determine the muon production depth as was done in [103]. This can also be turned around: assuming a function for the distribution of the number of particles created at a certain depth, this function can be converted into an arrival time distribution on the ground.

Let us assume the number of muons created as a function of depth to follow a Gaisser-Hillas profile (but see for example [105] for a different function to describe the MPD profile):

$$\frac{dN}{dX} = N_{\max} \left(\frac{X - X_0}{X_{\max}^{\mu} - X_0} \right)^{\frac{X_{\max}^{\mu} - X_0}{\lambda}} \exp \left(-\frac{X - X_{\max}^{\mu}}{\lambda} \right). \quad (4.3)$$

Given a simple isothermal atmosphere the atmospheric depth X can be written as a

CHAPTER 4. TIME DISTRIBUTION OF SIGNAL IN SURFACE PARTICLE DETECTORS

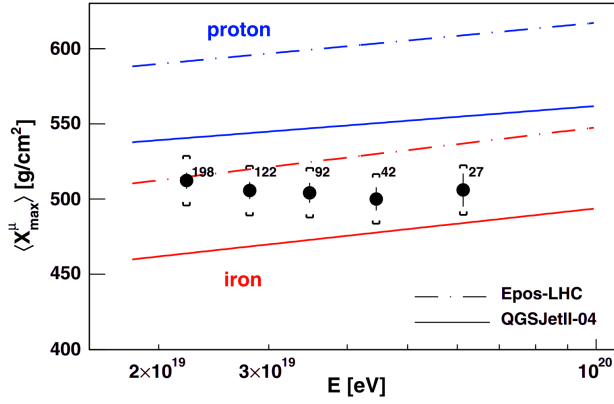


Figure 4.1: Mean X_{\max}^{μ} as a function of energy as measured by the Pierre Auger Observatory. Numbers show the number of events in each bin. Figure adapted from [103].

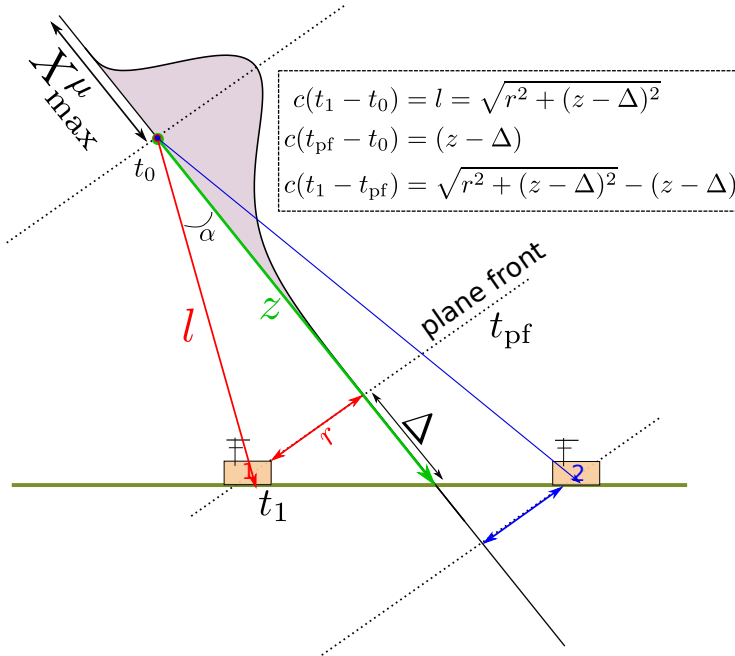


Figure 4.2: Sketch of the muonic part of a non-vertical shower. Shown is the muon production distribution and how the height of production is related to the arrival time at the surface.

function of time:

$$X(t) = \frac{X_{\text{vg}}}{\cos \theta} \exp \left(-\frac{h}{h_s} \right) \quad (4.4)$$

$$= \frac{X_{\text{vg}}}{\cos \theta} \exp \left[-\frac{\cos \theta}{h_s} \left(\frac{1}{2} \left(\frac{r^2}{ct} - ct \right) + \Delta \right) \right], \quad (4.5)$$

where $h = z \cos \theta$ is the height above sea level, $X_{\text{vg}} \approx 1040 \text{ g/cm}^2$ is the vertical atmospheric depth at sea level and $h_s \approx 8 \text{ km}$ is the atmospheric scale height.

One may already observe that the arrival time distribution on the ground can now be written as:

$$\frac{dN}{dt} = \frac{dN}{dX} \frac{dX}{dt} \quad (4.6)$$

and that this gives a function with a double exponential in t as was empirically found in [96].

But also the lateral distribution of muons influences the arrival time distribution, because this determines at which location particles from a particular production depth z end up on the ground. For muons it turns out that the lateral distribution can be described by assuming a simple energy and transverse momentum distribution [100]. The normalized energy distribution of muons is given by

$$\frac{dN}{dE} = N \frac{\gamma - 1}{m} \left(\frac{E}{m} \right)^{-\gamma} \left(1 - \frac{\rho a l}{E} \right)^\kappa, \quad (4.7)$$

which assumes a power-law distribution in energy at production and the last term arises from energy loss by ionization and the probability of muon decay that scales with the path length l . In accordance with [100]: $\kappa = \frac{m}{\rho a c \tau} \approx 0.8$, with τ the muon mean lifetime, ρ is the atmospheric density and a the ionization energy loss for muons, $\rho a \approx 0.0002 \text{ GeV/m}$, and m is the mass of the muon. The final muon energy is given by $E_f = E - \rho a l$.

The transverse momentum distribution is well described by

$$\frac{dN}{dp_T} = \frac{p_T}{Q^2} e^{-p_T/Q}, \quad (4.8)$$

with $Q \approx 0.17 \text{ GeV}$ [100].

It is assumed that Eqs. (4.3), (4.7) and (4.8) factorize:

$$\frac{d^3 N}{dX dE dp_T} = \frac{dN}{dX} \frac{dN}{dE} \frac{dN}{dp_T}. \quad (4.9)$$

Then, for high energies $p_T \simeq E \sin \alpha$, where α is the angle the muon makes with respect to the shower axis (see Fig. 4.2), and changing variables gives

$$\frac{d^3 N}{dX dE d \sin \alpha} = \frac{dN}{dX} \frac{dN}{dE} \frac{dN}{dp_T} E. \quad (4.10)$$

CHAPTER 4. TIME DISTRIBUTION OF SIGNAL IN SURFACE PARTICLE DETECTORS

To get the number of muons on the ground the integral over the initial energy E is evaluated but only for possible final muon energies larger than the muon mass $E_f = E - \rho al \geq m$:

$$\frac{d^2 N}{dX d \sin \alpha} = \frac{dN}{dX} \int_{m+\rho al}^{\infty} \frac{dN}{dE} \frac{dN}{dp_T} E dE \quad (4.11)$$

$$= \frac{dN}{dX} (\gamma - 1) \left(\frac{m}{Q} \right)^{\gamma-1} (\sin \alpha)^{\gamma-2} I_0(l, \sin \alpha). \quad (4.12)$$

The integral $I_0(l, \sin \alpha)$ is the result of integrating over energy and changing variables to $x = \frac{E \sin \alpha}{Q}$:

$$I_i(l, \sin \alpha) = \int_{x_0+y}^{\infty} x^{2-\gamma+i} \exp(-x) \left(1 - \frac{x_0}{x} \right)^{\kappa} dx, \quad (4.13)$$

with $x_0 = l \frac{\rho a \sin \alpha}{Q} = \frac{\rho ar}{Q}$ and $y = \frac{m \sin \alpha}{Q}$. Equation (4.13) is only moderately dependent on time—it marginally makes the distribution smaller. The angle between the muons and the shower axis is also time dependent because $\sin \alpha = r/l$. Now, changing coordinates to r and t so that the time distribution for a given r is

$$\frac{d^2 n}{dr dt} = \frac{d^2 N}{dX d \sin \alpha} |J|, \quad (4.14)$$

with

$$J = \left| \frac{dX}{dt} \frac{d \sin \alpha}{dr} - \frac{dX}{dr} \frac{d \sin \alpha}{dt} \right| = \frac{X(t)}{h_s t} \left(\frac{r^2 - c^2 t^2}{r^2 + c^2 t^2} \right)^2. \quad (4.15)$$

Writing down the full expression as a function of t (and ignoring the normalization):

$$\begin{aligned} \frac{d^2 n}{dr dt} \sim I_0(l(t), \sin \alpha(t)) & \frac{(r^2 - c^2 t^2)^2}{h_s t (r^2 + c^2 t^2)^2} \left(\frac{2ctr}{r^2 + c^2 t^2} \right)^{\gamma-2} \\ & \times \exp \left[\frac{-z(t) \cos \theta}{h_s} \left(1 + \frac{X_{\max}^{\mu}}{\lambda} \right) - \frac{X_{\text{vg}}}{\lambda \cos \theta} e^{-z(t) \cos \theta / h_s} \right] \end{aligned} \quad (4.16)$$

In this equation the Gaisser-Hillas formula is inserted with $X_0 = 0$ for brevity.* Equation (4.16) is essentially the same as Eq. (9) in [101] but now fully written down as a function of t .

Figure 4.3 shows the effect of changing X_{\max}^{μ} on the time distribution. As expected, for showers closer to the ground the time distribution shifts to the right and becomes wider. Figure 4.4 shows Eq. (4.16) compared to a simulated shower at different positions on the ground. Only X_{\max}^{μ} and λ were tuned, globally, for the entire shower.

In truth, the muons propagate with a velocity below the speed of light and there is an additional kinematic time-delay (which depends on the energy of the muons). This kinematical delay can be estimated from the energy spectrum and the total arrival time distribution is then the convolution of the geometrical time-delay Eq. (4.16) and this additional kinematical time-delay. See [100] for details.

*In the application here X_0 determines the earliest possible arrival time of muons, for $X_0 = 0$, particles can come from infinity but the probability quickly becomes very small. For reasonable values of X_0 , it shapes the time distribution similarly as λ .

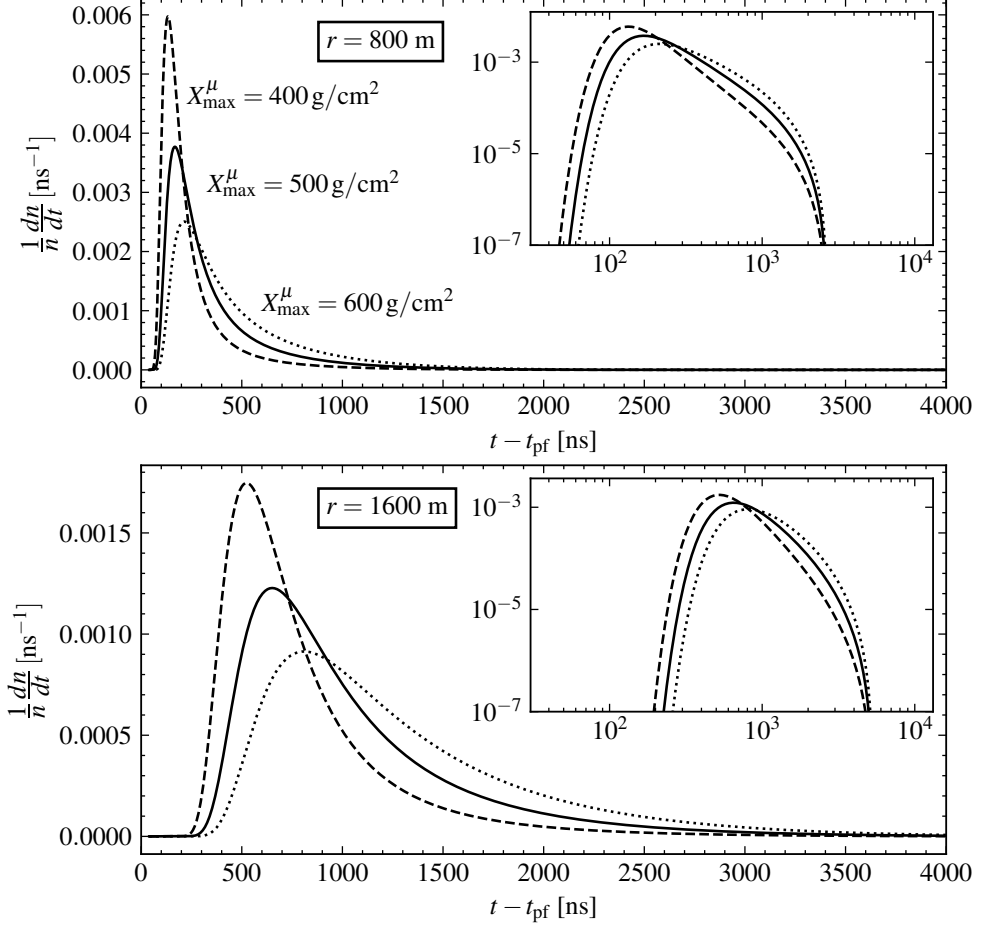


Figure 4.3: Muon time model for different values of X_{\max}^{μ} for $\theta = 30^\circ$ and $\psi = 90^\circ$. The top shows $r = 800$ m and the bottom $r = 1600$ m. The inset shows the same data but now with both axes converted to log scale.

CHAPTER 4. TIME DISTRIBUTION OF SIGNAL IN SURFACE PARTICLE DETECTORS

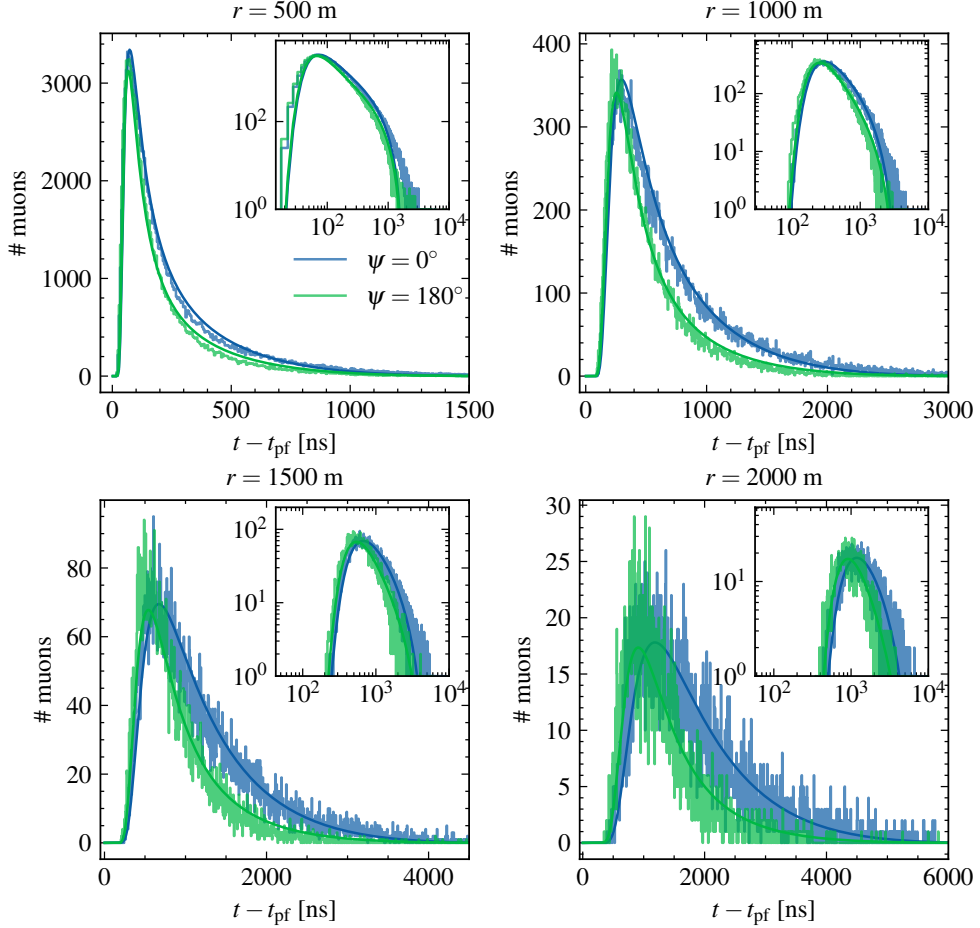


Figure 4.4: Equation (4.16) (smooth line) compared to the arrival times of muons in sampling areas at r and ψ . The simulated shower is a non-thinned (so all muons are followed) proton simulation with $\log_{10}(E/\text{eV}) = 18.5$ and a zenith angle $\theta = 30^\circ$. The CORSIKA version used was 7.40005. The values for $X_{\text{max}}^\mu = 516 \text{ g/cm}^2$ and $\lambda = 78 \text{ g/cm}^2$ were obtained as best fit to the time distributions of all sampling areas together. All other variables were fixed as explained in the text. The inset shows the same data but with both axes transformed to log scale.

4.3 Detector energy response to muons

The response of the Pierre Auger Observatory water-Cherenkov detectors (WCDs) to high energy muons is straightforward if the energy of muons is high enough such that it traverses the whole tank[†]. If this is the case a vertical muon will deposit an energy of ≈ 240 MeV and the signal is calibrated to this value and is designated in units of VEM (see [Chapter 2](#)). If the muon energy is below some threshold E_{VEM} such that it does not traverse the whole tank anymore, the signal is smaller than 1 VEM. As was shown in [Chapter 3](#), the detector response of muons with $E_f < E_{\text{VEM}}$ goes like $S \sim \left(\frac{E_k}{E_{\text{VEM}}}\right)^2$ and the value of $E_{\text{VEM}} \approx 350$ MeV in kinetic energy (see [Section 3.2](#)). In the following, for convenience, E_{VEM} denotes the total energy which is then $350 \text{ MeV} + m$.

The signal arrival time distribution is modified by this detector response. Starting with [Eq. \(4.12\)](#) and including this detector response, the signal distribution is:

$$\frac{d^2 S}{dX d\sin\alpha} = \frac{dN}{dX} \int_{m+\rho al}^{\infty} \frac{dN}{dE} \frac{dN}{dp_T} E \min \left[\left(\frac{E - \rho al}{E_{\text{VEM}}} \right)^2, 1 \right] dE \quad (4.17)$$

$$= \frac{dN}{dX} (\gamma - 1) \left(\frac{m}{Q} \right)^{\gamma-1} (\sin\alpha)^{\gamma-2} I_{\text{WCD}}(l, \sin\alpha). \quad (4.18)$$

The integral over E can again be written down in the dimensionless variable $x = \frac{E \sin\alpha}{Q}$, but is now made up of two terms, one for muons with $E_f < E_{\text{VEM}}$ and one for $E_f \geq E_{\text{VEM}}$:

$$I_{\text{WCD}}(l, \sin\alpha) = I_{E_f < E_{\text{VEM}}}(l, \sin\alpha) + I_{E_f \geq E_{\text{VEM}}}(l, \sin\alpha), \quad (4.19)$$

where the $(E - \rho al)$ term splits the first integral in three pieces which can be written down in terms of [Eq. \(4.13\)](#) with $i = 0, 1, 2$:

$$\begin{aligned} I_{E_f < E_{\text{VEM}}}(l, \sin\alpha) &= \left(\frac{Q}{E_{\text{VEM}} \sin\alpha} \right)^2 I_2(l, \sin\alpha)|_{x_0}^{x_1} \\ &\quad - \frac{2\rho al Q}{E_{\text{VEM}}^2 \sin\alpha} I_1(l, \sin\alpha)|_{x_0}^{x_1} \\ &\quad + \left(\frac{\rho al}{E_{\text{VEM}}} \right)^2 I_0(l, \sin\alpha)|_{x_0}^{x_1} \end{aligned} \quad (4.20)$$

and for the high energy part:

$$I_{E_f > E_{\text{VEM}}} = I_0(l, \sin\alpha)|_{x_1}^{\infty}. \quad (4.21)$$

The limits of integration are given by $x_0 = \frac{\rho ar}{Q} + \frac{m \sin\alpha}{Q}$ (same as before) and $x_1 = \frac{\rho ar}{Q} + \frac{E_{\text{VEM}} \sin\alpha}{Q}$. The integrals can be evaluated numerically. With $E_{\text{VEM}} = m$ the integral reduces to [Eq. \(4.13\)](#) without any detector effects. The difference between with and without this detector effect is shown in [Fig. 4.5](#). Notice how increasing E_{VEM} makes the time distribution more narrow. This is because low energy particles that arrive on average later, are suppressed by the $(E/E_{\text{VEM}})^2$ term.

[†]Also, the incoming direction of the muon influences the signal as was discussed in [Section 3.2.2](#).

4.4 Detector time response

Another effect of the detector is the ‘smearing’ of the arrival time. Due to various effects in the WCD—such as the Cherenkov light being scattered multiple times inside the tank before reaching the PMTs—the signal of one particle is not contained in a single time bin, but has an exponential tail.

The detector time response was studied by simulating down-going muons with a kinetic energy of 2 GeV injected 2 m above ground (0.8 m above the water). Because also the effect of the SSD was studied the muons were injected with $(x, y) = (0.99, -0.36)$ meters with respect to the center of the tank so that the SSD was hit in the middle of one of the scintillator bars.

The average signal from 10000 injected muons as a function of time after the time of injection is shown on the left in Fig. 4.6 for the WCD (averaged over 3 PMTs) and on the right the same for the SSD. Simulations with the new electronics (UUB) overlap with those of the old electronics (UB). No differences between the 3 PMTs in the WCD were observed, although an effect for muons coming from the side would be expected, but this was not studied. Removing the SSD from the tank did not change the time response of the WCD. Furthermore, electrons (positrons and photons behave the same) with an energy of 30 MeV produce an almost identical time response as muons.

The tail of the WCD response ($t \gtrsim 50$ ns after injecting the particles) is well described by an exponential with a decay time $\tau \approx 60$ ns. The SSD response is much faster, as can be expected, but there is some spillover into neighbouring bins that is approximately Gaussian with a width ≈ 18 ns as determined from these simulations.

It is known that the time response of the WCD can vary between individual stations and that there is an effect on the mean time response with the aging of the tank. It is believed that the liner-reflectivity of the inside of the tank changes over time such that the number of reflections of the Cherenkov light in the water is decreased. This makes the time response to single muons smaller for data that is recorded more recently. The deterioration of the tanks was also seen as an effect on a neural networks performance to estimate X_{\max} from the time traces [106].

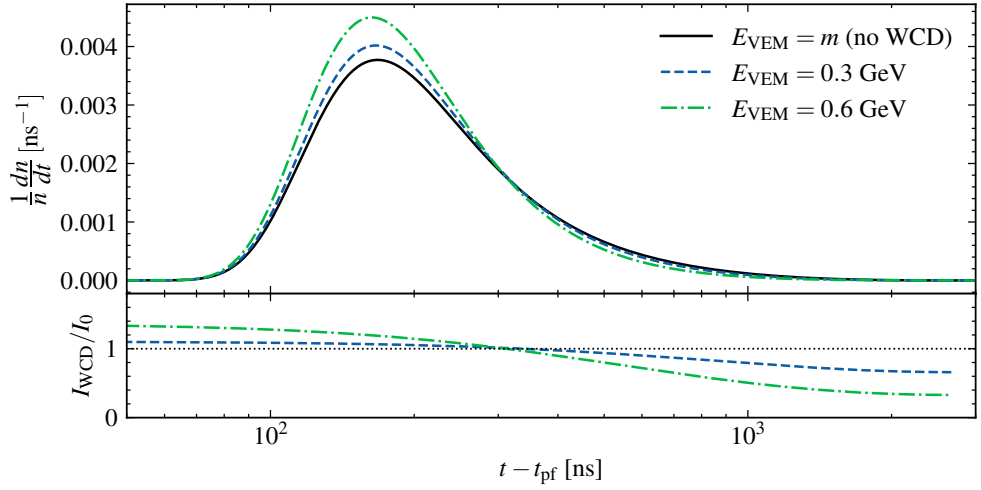


Figure 4.5: The muon time model with and without the WCD detector response as discussed in the text. The shower parameters are the same as in Fig. 4.3 and the black solid line shows the time distribution at $r = 800$ m calculated with Eq. (4.16). The dashed and dashed-dotted lines show two different values of E_{VEM} when including the detector response from the WCD as in Eq. (4.18). The bottom panel shows the dashed and dashed-dotted line in the top figure divided by the black solid line and is effectively Eq. (4.19) divided by Eq. (4.13) as a function of time.

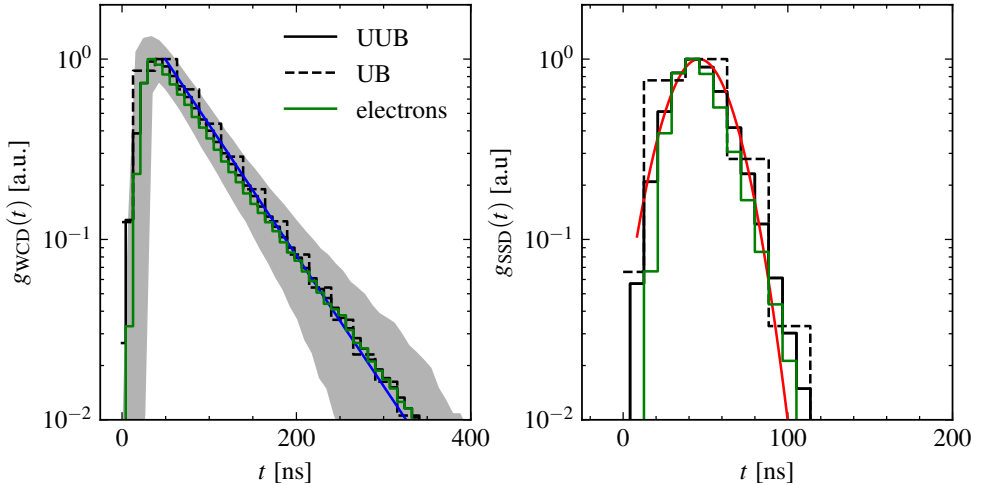


Figure 4.6: Response to vertical muons of 2 GeV in the WCD (left) and SSD (right), the vertical scale is set to have the mean maxima at 1. The solid line is the mean with new electronics (UUB) and dashed with old (UB). The shaded band denotes the 16 – 84% interval. The green line shows the response to 30 MeV electrons. The blue line on the left shows a falling exponential with $\tau = 60$ ns and the red line on the right shows a Gaussian distribution with $\mu = 45$ ns and $\sigma = 18$ ns.

5 Time trace templates

The steep cutoff in the energy spectrum of ultra-high-energy cosmic rays results in only a very limited number of events that have been observed at the highest energies. In combination with the limited duty-cycle of the Fluorescence Detector this means that there are virtually no X_{\max} measurements—currently the best estimator of the mass of the primary cosmic ray—above 60 EeV. This is also the motivation behind the current upgrade with scintillator surface detectors: to study the mass composition at the highest energies with the surface detector (as explained in [Chapter 3](#)). In the meantime we can use additional information from the water-Cherenkov detector (WCD) arrival time of the signal to infer the mass composition and the particle type for energies up to about 100 EeV.

In [Chapter 4](#) a theoretical framework that connects the arrival time of particles on the ground to the longitudinal development of the shower was described. In this framework, the arrival time can be used to determine the longitudinal shower development which in turn allows to probe the mass composition and/or infer hadronic physics at ultra-high energies. Also in earlier studies, several aspects of the timing information were used as a probe for mass composition and/or hadronic physics:

- Using the start time of the signal [[107](#), [108](#)].
- Using the rise time of the signal. As early as 1974 from the Haverah Park array [[94](#)] and more recently with the Pierre Auger Observatory [[95](#), [109](#)].
- With shower universality [[97](#), [98](#)].
- And very recently with Deep Neural Networks [[106](#)] (although it is not clear how important the timing information is to the neural network).

The signal time shape was also used to find photons at ultra-high energies [[96](#)].

To be able to utilize as much information as possible in a comprehensible procedure a template fit on the full time traces is desired. In this chapter, air shower universality is used to create these *time trace templates* for the signal from the WCDs as a function of time and how one can use these templates to infer X_{\max} and the number of muons.

5.1 Simulations

In this work end-to-end extensive air shower simulations were used to determine the signal recorded by the upgraded Pierre Auger Observatory (WCDs and SSDs).

Air showers simulated with CORSIKA 7.5600 were provided by the Napoli-Praha library [[110](#)]. This library contains showers with a flat distribution in $\log_{10}(E/\text{eV})$ between 18.5 and 20 and a zenith angle distribution that is flat in $\cos^2 \theta$ with $0^\circ < \theta < 65^\circ$. About 9000 proton and 8000 iron induced showers simulated with the hadronic interaction model EPOS-LHC were used from this simulation library. The simulated air showers underwent a full surface detector simulation with Geant4 to simulate the

upgraded detector response with UUB and SSD and a standard SD reconstruction with the `Offline` framework (git hash `b8dfbd6c9`).

Additional stations were added to the regular 1500 m-grid at distance to the shower core in the shower plane $r/\text{m} \in \{500, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000, 2500\}$. At each distance 10 dense station were distributed on a ring with angle in the shower plane $\psi/^\circ \in \{0, 36, 72, 108, 144, 180, 216, 252, 288, 324\}$.

5.2 Total signal model

As explained in [Section 1.3](#) extensive air showers are on average very similar—*universal*—when described with respect to the amount of atmosphere traversed. It is then possible to describe the secondary particle spectrum at the surface as a function of the distance to X_{max} which in turn determines the signal as recorded by particle detectors at the surface.

In accordance with [\[97, 111–113\]](#) the total signal in a WCD at location in the shower plane \vec{r} , with distance to shower maximum ΔX (see [Section 1.3](#) for a more thorough definition of ΔX) from a cosmic-ray induced air shower with primary energy E is parametrized as

$$\hat{S}(\vec{r}, \Delta X, E) = \sum_k \hat{S}_k(E, \Delta X, r = 1000 \text{ m}) \rho_k(\vec{r}, \Delta X). \quad (5.1)$$

Here, $S_k(E, \Delta X, r = 1000 \text{ m})$ is the signal at 1000 m—a distance that is optimized for the Pierre Auger Observatory as the shower size [\[80\]](#)—for component $k \in \{\mu, e\gamma, e\gamma(\mu), e\gamma(\text{had})\}$ and ρ_k is designated as the *signal density* for each component k defined as

$$\rho_k(\vec{r}, \Delta X) = \frac{S_k(\vec{r}, \Delta X, E)}{\hat{S}_k(E, \Delta X, r = 1000 \text{ m})}, \quad (5.2)$$

where the numerator is the actual signal recorded by a WCD and the denominator is the model prediction (denoted by a hat) of the shower size for component k .

As a reference model for the determination of the total signal, proton showers simulated with the EPOS-LHC hadronic interaction model were used.

Shower size

First, the shower size for each component was parametrized using the reference model. The following function was used [\[111\]](#)

$$\begin{aligned} \hat{S}_k(E, \Delta X, r = 1000 \text{ m}) = & S_k^{\text{ref}} \left(\frac{E}{10^{19} \text{ eV}} \right)^{\gamma_k} \\ & \times \left(\frac{\Delta X - \Delta X_{0,k}}{\Delta X_{\text{ref}} - \Delta X_{0,k}} \right)^{\frac{\Delta X_{\text{max},k} - \Delta X_{0,k}}{\lambda_k}} e^{-\frac{\Delta X - \Delta X_{\text{ref}}}{\lambda_k}}, \end{aligned} \quad (5.3)$$

the first part contains the scaling with the primary energy E while the latter part is a modified Gaisser-Hillas function that describes the longitudinal development of the

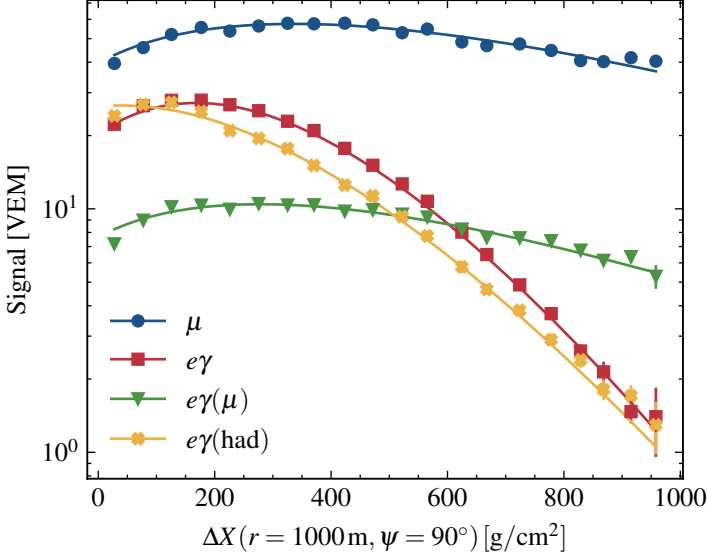


Figure 5.1: Average signal for the four components at 1000 m (shower size) for EPOS-LHC proton showers at $19.4 < \log_{10}(E/\text{eV}) < 19.5$ and $0^\circ < \theta < 53^\circ$. The points show the mean in bins with a size of 50 g/cm^2 and the error bars are the standard error of the mean. Only showers with ΔX are shown. The lines show a fit with Eq. (5.3).

shower with respect to the distance to shower maximum ΔX . S_k^{ref} is the signal of a proton shower with $E = 10^{19} \text{ eV}$ at $r = 1000 \text{ m}$ at $\Delta X = \Delta X_{\text{ref}}$. The parameters ΔX_{max} , ΔX_0 and λ determine the shape of the Gaisser-Hillas function.

The average signal from 10 dense station at 1000 m was determined for bins in $\log_{10}(E/\text{eV})$ with size 0.1 and bins in $\Delta X(r = 1000 \text{ m}, \psi = 90^\circ)$ with a size of 50 g/cm^2 . If X_{max} is very close to the ground, universality is not expected [58] and so only events with $\Delta X(r = 1000 \text{ m}, \psi = 90^\circ) > 10 \text{ g/cm}^2$ were used.

The following parameters were fixed to reduce correlations between the fitted parameters: $\Delta X_{\text{ref}} = 400 \text{ g/cm}^2$, $\Delta X_{0,\mu} = \Delta X_{0,e\gamma(\mu)} = -250 \text{ g/cm}^2$, $\Delta X_{0,e\gamma} = -500 \text{ g/cm}^2$ and $\Delta X_{0,e\gamma(\text{had})} = -800 \text{ g/cm}^2$ [111]. The other parameters were determined from a least-squares fit. As in [111] the λ parameter is allowed to depend logarithmically on the primary energy for the $e\gamma$ and $e\gamma(\text{had})$ components:

$$\lambda(E) = \lambda_0 + f_\lambda \log_{10} \left(\frac{E}{10 \text{ EeV}} \right). \quad (5.4)$$

It was established that Eq. (5.3) does not describe the θ dependency perfectly (see also [111]). Below $\approx 53^\circ$ the deviations are within $\sim 10\%$ and can be corrected for with a third-degree polynomial (see Fig. A.1).

Now with the reference model in place it is possible to define the relative number of muons as:

$$R^\mu = \frac{S_\mu(r = 1000 \text{ m})}{\widehat{S}_\mu(E, \Delta X, r = 1000 \text{ m}, \text{EPOS-LHC, proton})}, \quad (5.5)$$

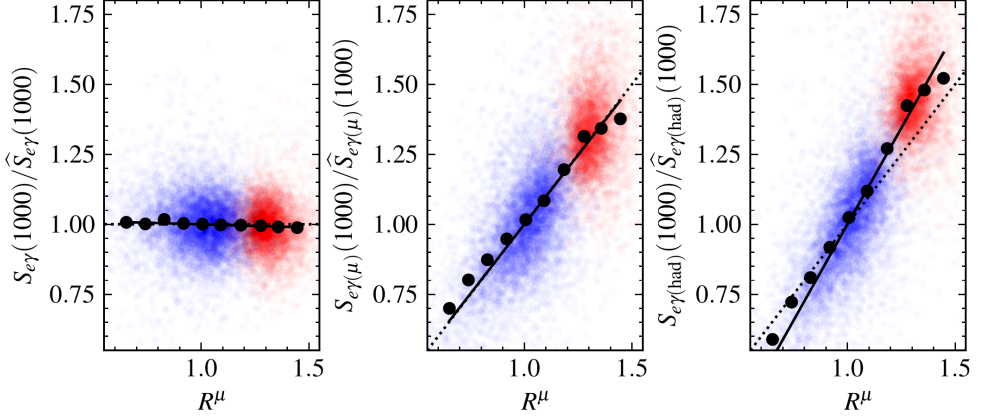


Figure 5.2: The ratio of the shower size signal for the $e\gamma$, $e\gamma(\mu)$ and $e\gamma(\text{had})$ components divided by the reference model as a function of the relative muon content R^μ (the pure muon component is not shown because the ratio is 1:1 by construction). The large round markers are the mean in equally spaced bins containing both proton and iron showers to which the solid black line is fitted with a slope α_k . The cloud of points show individual showers, proton in blue and iron in red. The dotted line is the 1:1 line except for the $e\gamma$ component where it is a horizontal line at 1.

where the numerator is the muon shower size for a single event^{*} and the denominator is the muon shower size of the reference model.

As explained in [Section 1.3](#) the relative number of muons is used as a proxy for the hadronic content of the air shower. EPOS-LHC iron induced showers have $R^\mu \approx 1.3$. The ratio of the shower size signal with respect to the reference model for the three electromagnetic components scales with R^μ which can be approximated with a straight line as can be seen from [Fig. 5.2](#):

$$\frac{S_k(r = 1000 \text{ m})}{\widehat{S}_k(r = 1000 \text{ m}, \text{EPOS-LHC, proton})} = 1 + \alpha_k(R^\mu - 1). \quad (5.6)$$

Such that [Eq. \(5.1\)](#) is modified to:

$$\widehat{S}(\vec{r}, \Delta X, E, R^\mu) = \sum_k (1 + \alpha_k(R^\mu - 1)) \widehat{S}_k(E, \Delta X, r = 1000 \text{ m}) \rho_k(\vec{r}, \Delta X), \quad (5.7)$$

with $\alpha_k = 1, -0.023, 0.99, 1.37$ for, respectively, the μ , $e\gamma$, $e\gamma(\mu)$ and $e\gamma(\text{had})$ components as determined from the straight-line fit as shown in [Fig. 5.2](#).

There is some non-universality for more inclined showers. The mean bias as a function of $\sec\theta$ behaves different for iron and proton. This is mostly seen in the $e\gamma(\text{had})$ component (see [Fig. A.2](#)). The effect on the total signal is small and within 5% for showers with $\theta < 53^\circ$ as can be seen from [Fig. 5.3](#) (left). This was also observed in [\[112\]](#). There is also a small energy dependent bias for iron showers as can be seen in [Fig. 5.3](#) (right), this artificially introduces an energy dependence of R^μ .

^{*}In simulations this is acquired by taking the average muon signal over 10 stations at 1000 m.

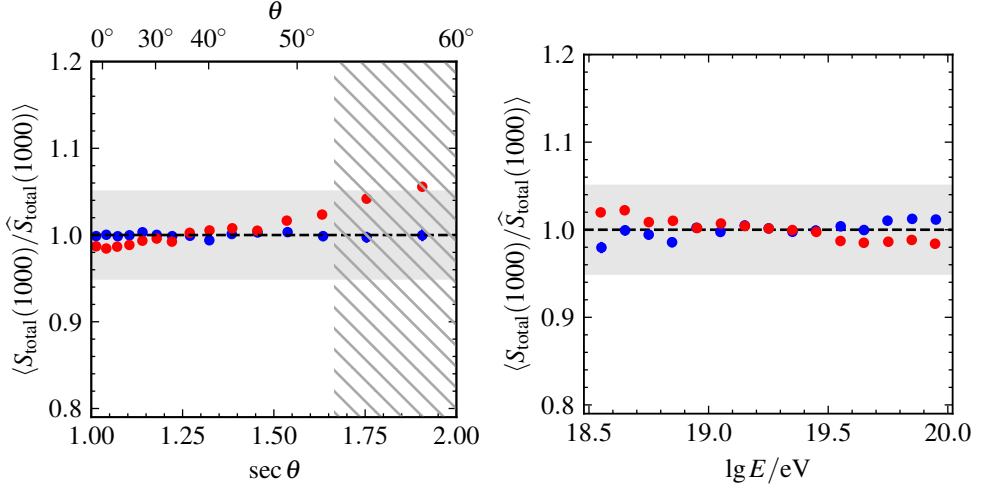


Figure 5.3: Mean bias of the total shower size signal. On the left: as a function of θ . On the right as a function of $\log_{10}(E/\text{eV})$. $\theta > 53^\circ$ is crossed out to indicate there is additional non-universality and these showers are not used. The shaded band denotes $\pm 5\%$ deviation.

Signal density

For all stations in the shower library the signal density ρ_k for each of the four components was calculated following Eq. (5.2) with the reference model as described in Section 5.2.1. Only showers with $\theta < 53^\circ$ were used because the model for the shower size does not perform well for more inclined showers as discussed in Section 5.2.1. Also, only showers with $\log_{10}(E/\text{eV}) > 19$ were used, because including lower energy showers deteriorated the best-fit model significantly. Furthermore, only stations with $450 < r/\text{m} < 2100$ are included in order to remove stations that are likely to be saturated at low r and stations with very small signals at large r . Stations with $\Delta X < 10 \text{ g/cm}^2$ are also removed. The model was fitted only on proton showers. To ensure that for small signals the model is not biased by the trigger an iterative approach was used:

1. fit on all stations
2. remove stations that have a predicted total signal $< 5 \text{ VEM}$ (above this threshold the trigger probability is $> 80\%$ [78]) from the previous fit and the shower size model from Section 5.2.1 and refit the model on the remaining stations.

After repeating step 2 a couple of times the number of stations surviving the cut always converged.

The signal density should be universal with respect to the primary energy and R^μ [111, 114]. It depends on the distance to the shower core r (which can be a usual lateral-density function) and it is expected there is a universal behaviour with respect to ΔX . The distance to X_{max} already contains a factor that depends on the asymmetry in the shower plane (ψ angle) for non-vertical showers, but there can be another geometrical dependency that scales with $\sin \theta \cos \psi$ (for muons see [101]).

Also, an additional term that depends on θ can be expected, for example for muons such a term would originate from the tracklength of the particles in the WCD.

Because it is not trivial how the exact signal forms from the before mentioned parameters a best-fit model for $\log \rho_k$ was searched for as a polynomial of the (combination) of $x_i = (\log r, \Delta X, \cos \theta, \sin \theta \cos \psi)$. Each input feature i is scaled with its mean μ and its standard deviation σ as: $\hat{x}_i = (x_i - \mu_i)/\sigma_i$.

A grid search with cross validation was done for polynomials of degree 2, 3 and 4 using ridge regression to reduce overfitting [115]. The regularization hyperparameter was determined on a log grid from 10^{-2} to 10^3 using a 5-fold cross validation [116]. The smallest mean least square was found for a 4-degree polynomial with regularization hyperparameters 88, 379, 88, 233 for, respectively, the μ , $e\gamma$, $e\gamma(\mu)$ and $e\gamma(\text{had})$ components. In total there are 70 fit parameters for each component.

In Fig. 5.4 (left) an example of the model for the total signal—that is comprised of the shower size and the signal density as in Eq. (5.7)—is shown for each of the four components. The total signal is accurately described with a resolution $\sigma_S \approx 15\%$ for a total signal of 30 VEM (see Fig. 5.5)—comparable to the total signal uncertainty model from data as was described in Section 2.1.1. At signals < 10 VEM the model starts to deviate because of the trigger. Figure 5.4 (right) shows the deviation of the true total signal with respect to the best-fit model as a function of r . For iron induced showers the signal is slightly underestimated at larger distances, meaning that these showers have a slightly less steep LDF which can be understood from the larger transverse momentum from nuclei-air interactions with respect to proton-air interactions.

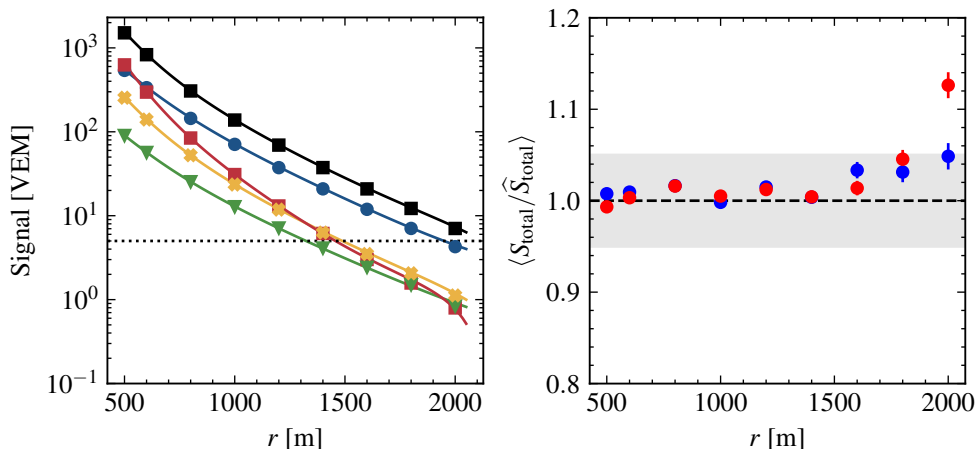


Figure 5.4: On the left: total signal for the four components as a function of distance to the shower core. Markers show the average of 450 EPOS-LHC proton showers with $35^\circ < \theta < 40^\circ$ and $19.5 < \log_{10}(E/\text{eV}) < 19.6$. The lines show the corresponding model that is described in the text. From top to bottom at 1000m: total = black, μ = blue, $e\gamma$ = red, $e\gamma(\text{had})$ = yellow and $e\gamma(\mu)$ = green. On the right: Mean bias of the total signal as a function of distance to the shower core for the same subset of showers shown on the left in blue (proton) and iron in red.

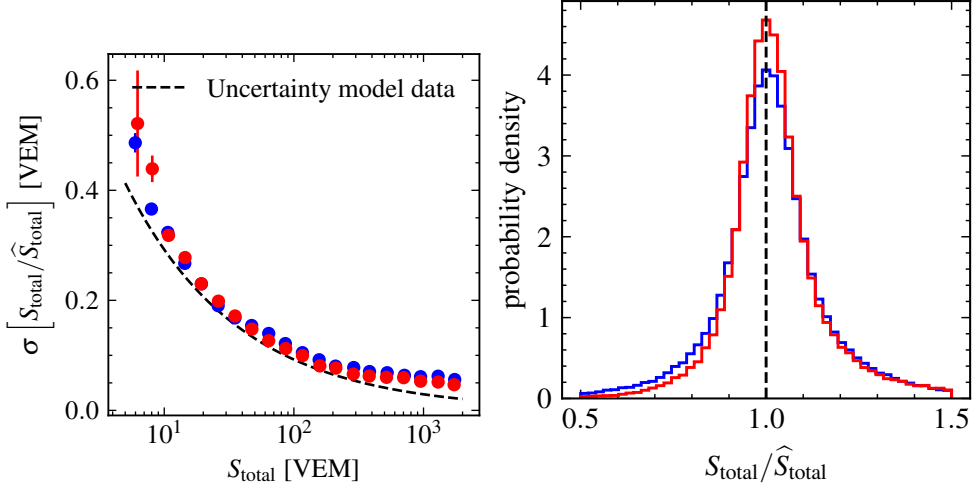


Figure 5.5: Resolution of the model. On the left as a function of the total signal. Proton (blue) and iron (red) are compared to the uncertainty model from data Eq. (2.2) (dashed line) for showers $35^\circ < \theta < 40^\circ$ and $19.5 < \log_{10}(E/\text{eV}) < 20$. The error bars are the standard deviation of the mean bootstrapped 1000 times. On the right the distribution of the deviation of the total signal with respect to the model for all stations from 450 m to 2550 m.

5.3 Time-dependent signal

As described in Chapter 4 the arrival time of particles on the ground is related to the longitudinal development of an extensive air shower. This section describes how the arrival time distribution is determined from MC simulations.

Start time

The most straightforward method to study the longitudinal development of air showers through the arrival time is through the use of the radius of curvature determined from a fit to the start times. Assuming that all particles originate from a single point, and that these particles all travel with the speed of light, the shower front can be (crudely) approximated as an expanding sphere with a radius of curvature R_c (see Fig. 5.6). The radius of curvature is related to the depth of the first interaction and the inclination of the shower. It was shown that R_c correlates with the distance to X_{max} , as measured by the FD [107].

An improvement on R_c was made in [108], where the shower front was approximated as a catenary-function. Also the time variance model of the start time was scrutinized and the correlation with X_{max} , measured by the FD, was improved.

Time distribution

In Chapter 4 an analytical function was found that describes the muon arrival time distribution. For the electromagnetic components this function does not work very

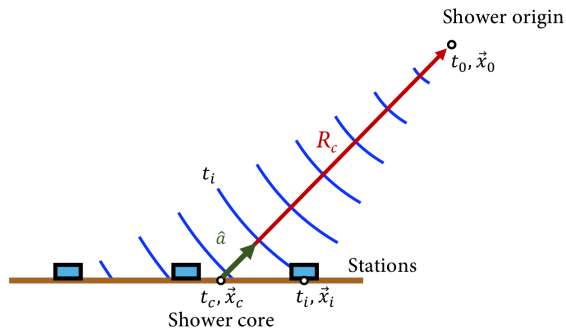


Figure 5.6: Illustration of a spherical shower front with the radius of curvature R_c as the distance from the shower origin to the shower core. Figure taken from [108].

well, because these particles are scattered to a much greater extent than muons. In previous works it was shown that a lognormal distribution well describes the signal as a function of time in the WCDs [65, 97, 98]. A downside of the lognormal function is that there is no physical intuition of what determines the shape parameters. Also, because no particle moves faster than the speed of light, there is an inherent start time before which no particles can possibly arrive. Thus, an additional parameter t_0 has to be added to the lognormal distribution. The following function describes the probability density of the arriving signal:

$$f(t) = \frac{1}{s\sqrt{2\pi}(t - t_0)} \exp\left(-\frac{(\ln(t - t_0) - m)^2}{2s^2}\right), \quad (5.8)$$

with t and t_0 in units of nanoseconds. To reduce the number of parameters t_0 is set to the arrival time model from [108]. The remaining shape parameters m and s have to be parametrized from simulations.

Equation (5.8) is not parametrized directly on the signal from the detectors. Instead, first a convolution between Eq. (5.8) and the detector time response as set out in Section 4.4 is made. The probability density of the signal as a function of time is thus:

$$h(t) = \frac{1}{|S|} \frac{dS}{dt} = f(t) \otimes g(t), \quad (5.9)$$

where $|S|$ is the total integrated signal and $g(t)$ is described by a falling exponential[†] as described in Section 4.4

$$g(t) = \Theta(t)e^{-t/\tau}, \quad (5.10)$$

with $\Theta(t)$ the Heaviside step function. Including this detector time response is important to be able to tune simulations to data (due to the detector aging, see Section 4.4) and also for estimating the uncertainty on the signal as a function of time as will be explained in the next section. The value for $\tau = 61$ ns in simulations and for data

[†]The rising part of the response is ignored since it approximately only leads to a shift in time.

CHAPTER 5. TIME TRACE TEMPLATES

can be acquired from the monitoring data that stores this values for each PMT. The average over all (working) PMTs is used.

Chapter 4 provides some intuition of what determines the shape of the arrival time of the signal: it depends largely on the distance to the shower core and on the distance to X_{\max} . But additional effects can be expected as a function of θ and ψ for example due to the tank geometry.

The signal traces of the stations from all showers were placed in bins of ΔX with size 50 g/cm^2 for each distance to the core with dense stations simulated. All proton and iron showers with $19 < \log_{10}(E/\text{eV}) < 20$ were averaged together. First, the average trace using all 3 PMTs was taken for each station. Then, all traces were synchronized with respect to the predicted t_0 from the arrival time model. The mean and standard deviation in each time bin of the synchronized traces was calculated for all dense stations in the corresponding ΔX bin. For each of these traces the lognormal shape parameters were determined by a least-square fit with Eq. (5.9). The uncertainty on the mean in each time bin is given by the standard error on the mean. To make sure that the fit converges only time bins are used that are below the time that includes 95% of the total signal. An example is given in Fig. 5.7. One may observe that due to fluctuations of the start time the first few bins are not very well reproduced.

For each distance to the core a cubic polynomial was fitted for m and a straight line for s as a function of $\Delta X/(400 \text{ g/cm}^2) - 1$. A selection of distances to the core is shown in Figs. 5.8 and 5.9. There are small differences between proton and iron especially for the s parameter and at larger distances to the core as can be seen in Figs. A.3 and A.4. For each distance to the core a lookup table was created with fit values from the polynomial as a function of ΔX . The parameters of the polynomials are linearly interpolated in $\ln r$.

Total model

The total model of the signal at a time t is the combination of the shower size, the signal density and the time dependent part. For completeness, the predicted signal in a time bin t_i is given by

$$S(t_i, r, \psi, E, \theta, X_{\max}, R^\mu) = \sum_k (1 + \alpha_k(R^\mu - 1)) S_k(E, \theta, X_{\max}, r = 1000 \text{ m}) \\ \times \rho_k(r, \psi, \theta, X_{\max}) \times h_k(t_i, r, \psi, X_{\max}) \Delta t. \quad (5.11)$$

On the last line the model is multiplied by the bin width Δt . This is a fair approximation of the integrated signal in the bin for a bin width of 25 ns or below. Also, the atmospheric slant depth at the detectors are an input parameter.

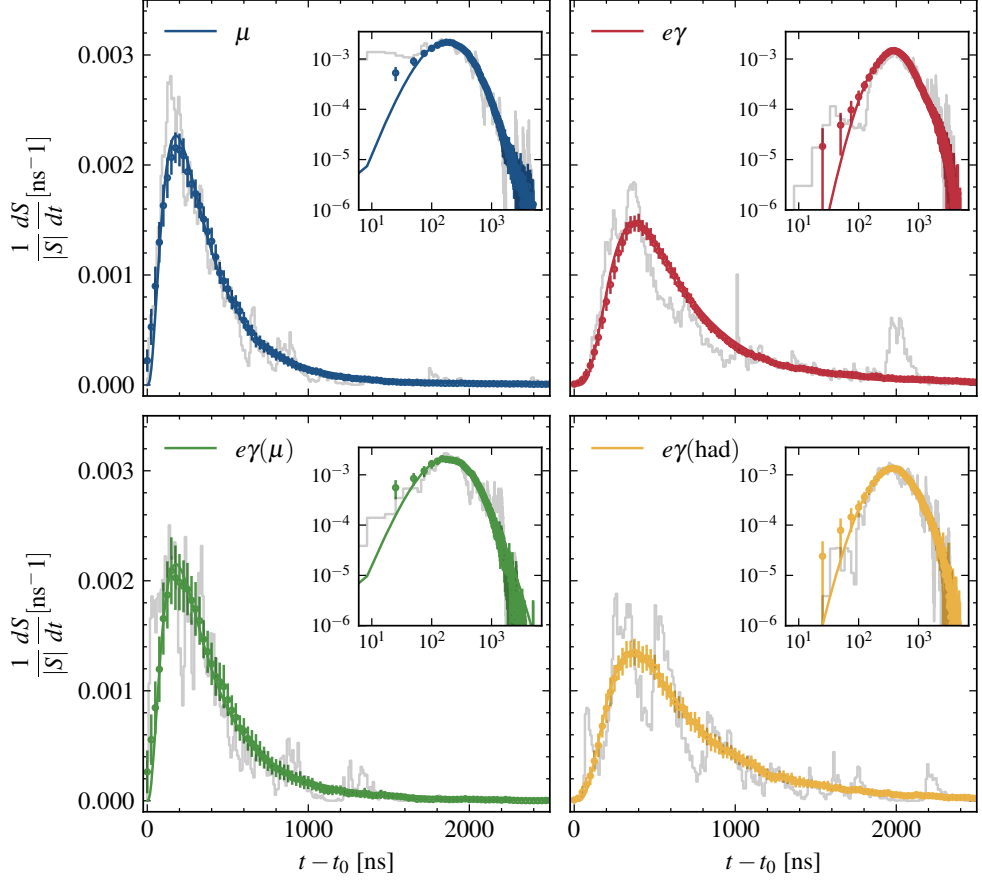


Figure 5.7: Mean traces for the 4 signal components at $21^\circ < \theta < 30^\circ$, $175 < \Delta X/(\text{g}/\text{cm}^2) < 200$, $r = 800 \text{ m}$ and $\psi = 0^\circ$. The error bars are blown up by a factor 2 and only shown every 25 ns for clarity. The colored line shows the lognormal fit and the gray line shows one of many contributing traces in this bin. The inset shows the same data on a log-log scale.

CHAPTER 5. TIME TRACE TEMPLATES

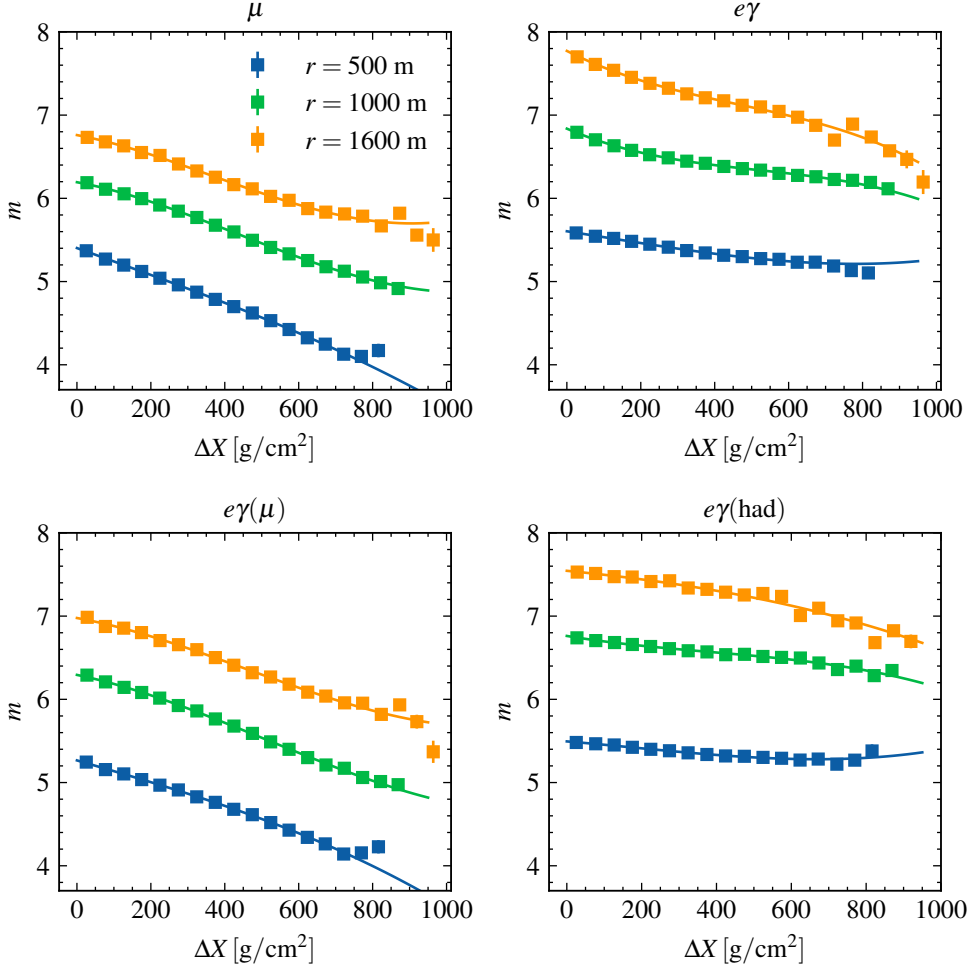


Figure 5.8: The m parameter from the lognormal fit to the average WCD traces from both proton and iron induced showers as a function of ΔX . A selection of distances to the core is shown. The energy of the showers is between $19 < \log_{10}(E/\text{eV}) < 20$. The full line shows the parametrization as described in the text.

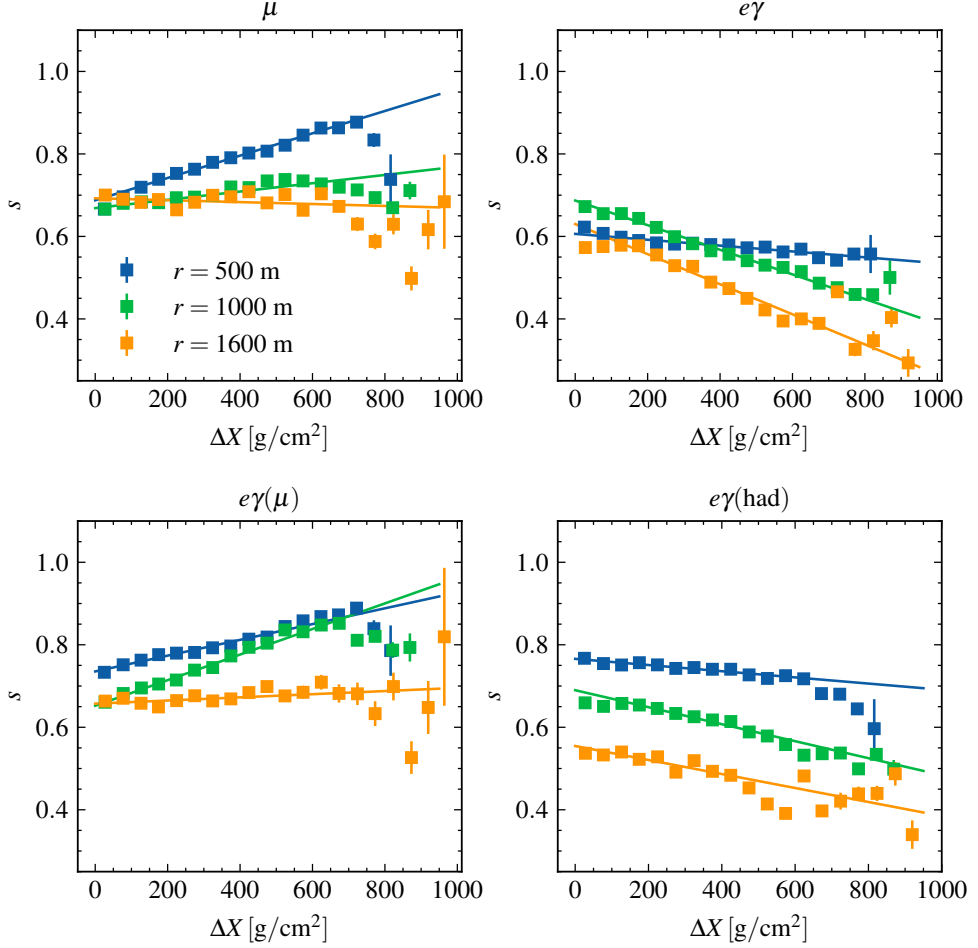


Figure 5.9: The s parameter from the lognormal fit to the average WCD traces from both proton and iron induced showers as a function of ΔX . A selection of distances to the core is shown. The energy of the showers is between $19 < \log_{10}(E/\text{eV}) < 20$. The full line shows the parametrization as described in the text.

5.4 Uncertainty of signal in single time bins

To compare the model, as described in the previous section, to data another piece of the puzzle is the correct probability distribution of the expected signal in a time window as from Eq. (5.11) with respect to the measured signal in that time window. In the Gaussian limit this amounts to knowing the uncertainty on the measured signal. However, for small signals a Poisson distribution will be more appropriate. Moreover, the detector response to different kinds of particles and the smearing of the arrival time complicates the likelihood.

Scaled Poisson distribution

It is clear that there are different contributions to the uncertainty on the signal. In most cases the dominant contribution to the uncertainty is that of the limited number of particles that are sampled in a small time window. Other than that, there are several detector and calibration contributions such as the baseline fluctuations and the number of photoelectrons measured.

Let us assume a very simple model that mimics the WCDs: every muon creates the same signal (1 VEM integrated signal is approximately 240 MeV deposited energy) and every other particle (mostly electrons, positrons and photons) creates about 1/25 VEM, corresponding to electrons/photons with a kinetic energy of about 10 MeV [61]. Ignoring any other detector influences, the probability distribution of measuring a signal in a specified time bin i is given by a scaled continuous Poisson distribution:

$$P(S_i) = \text{Poisson}(S_i/f_P^2)/f_P^2 = \frac{e^{\mu_i/f_P^2} (\mu_i/f_P^2)^{S_i/f_P^2}}{f_P^2 \Gamma(S_i/f_P^2 + 1)}, \quad (5.12)$$

which is nothing more than converting the signal into the expected number of particles with a scale factor $f_P^2 = S/n$ (see also Eq. (2.2)). The factorial is replaced by the gamma function $\Gamma(x)$. The expected signal is

$$\mu_i = (n_i^\mu + n_i^{\text{em}}/25) \text{ VEM}, \quad (5.13)$$

where n_i^μ is the expected number of muons at time bin i and n_i^{em} the expected number of electrons, positrons and photons at bin i . Without any uncertainty on the calibration to VEM, the variance is simply

$$\sigma_i^2 = (n_i^\mu + n_i^{\text{em}}/25^2) \text{ VEM}^2 \quad (5.14)$$

and the Poisson scale factor is

$$f_P = \sigma_i / \sqrt{\mu_i}, \quad (5.15)$$

which is the same as in Eq. (2.2) and in this toy model just a function of the fraction of muons with respect to other particles and the average signal from electromagnetic particles.

Figure 5.10 shows the distribution of 360 WCD traces (averaged over all 3 PMTs) from a single simulated vertical iron induced air shower at 1000 m from the core reconstructed with the `Offline` application `SdSimulationReconstructionUpgrade`. All traces were synchronised with respect to the plane time front. From these synchronised traces the mean and the variance of the signal were determined in each bin in order to calculate the scale factor that goes into the Poisson distribution. One may observe from the right panel that a scaled Poisson distribution fits the signal distribution quite well at several times selected from the left figure.

Similarly to the uncertainty on the total signal, the signal in individual time bins can thus be described by scaling the signal with a Poisson factor. But there is another subtlety that is addressed in the next part.

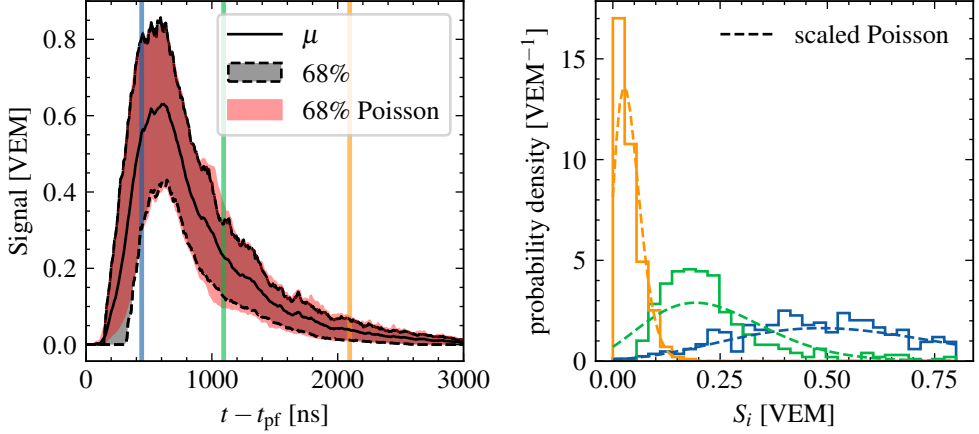


Figure 5.10: On the left: the mean and 68% interval of traces at 1000 m. Overlaid is the scaled Poisson 68% interval for corresponding mean and variance. On the right: the distribution of the signal at specified bins shown on the left in the same colour. Dashed lines indicate scaled Poisson pdfs with the acquired mean and variance (not fitted).

Correlation between time bins

The signal of one particle entering the detector is not contained in a single time bin because of various effects such as the Cherenkov light bouncing around in the tank before reaching the PMTs (see [Section 4.4](#)). This means that the signal in neighbouring bins are not independent and cannot be treated as such in a likelihood calculation.

The time response of the WCD can be approximately described by an exponential decay with a decay constant $\tau \approx 60$ ns as discussed. Additionally, there is an exponential rise which is replaced by a simple offset of $t_0 \approx 30$ ns before any signal arises (which is not important for the correlation). The signal in individual time bins can then be written as a vector \vec{S} that is the result of a matrix multiplication between the convolution matrix \mathbf{A} and the number of particles arriving at a time corresponding to a single time bin:

$$\vec{S} = \mathbf{A}\vec{n}. \quad (5.16)$$

The convolution matrix is the matrix that disperses the signal out over time:

$$A_{ij} = \begin{cases} \frac{\Delta t}{\tau} e^{-(t_i - t_j)/\tau} \text{ VEM charge,} & \text{if } t_i \geq t_j > t_0 \\ 0, & \text{otherwise,} \end{cases} \quad (5.17)$$

where the term $\frac{\Delta t}{\tau}$ arises because the units are in VEM charge and Δt is the bin size. The covariance Σ_{ij} between signals in bin i and j is the autocovariance of the vector \vec{S} which, by uncertainty propagation, can be shown to be

$$\Sigma_{ij} = \mathbf{A}\Sigma_n\mathbf{A}^T, \quad (5.18)$$

where Σ_n is the covariance of the particles arriving in a time bin. Observe that if the arrival times are uncorrelated so that $\Sigma_n = \mathbb{1}\sigma_n^2$, the covariance matrix of the signal—Eq. (5.18)—is *not* diagonal and the signal in the bins are correlated.

The total signal is the sum of the trace in units of VEM charge: $S_{\text{tot}} = \sum_i S_i$. The variance of the total signal is then the sum of the diagonals of the covariance matrix Σ_n (without the detector convolution) which is the same as the sum of all elements of Σ_{ij} (with the detector convolution). This can be exploited to calibrate the uncertainty in individual time bins from simulations with the uncertainty on the total signal that is determined from data (see Eq. (2.2)).

The reason that the time distribution in Eq. (5.9) was explicitly described including the convolution with the detector response is now understood: it is needed for a proper uncertainty calculation.

Tracklength influence on uncertainty

Additionally, the variation on the length of the track created by muons in the WCD is an important factor that determines the total uncertainty [93]. The mean $\bar{L}(\chi)$ and 68% interval around the median (for simplicity called $\sigma_L^2(\chi)$) on the signal normalized tracklength are found from [93] for incoming angle χ with respect to the top of the tank—which is approximated as the zenith angle of the shower: $\chi \approx \theta$. $\bar{L}(\theta = 38^\circ) \approx 0.95$ VEM and $\sigma_L(\theta = 38^\circ) \approx 0.42$ VEM. They modify the variance for the muon component as

$$\sigma_{S_\mu}^2 = \sigma_L^2(\theta) \left(\frac{S_\mu}{\bar{L}(\theta)} \right)^2 + \bar{L}(\theta) S_\mu. \quad (5.19)$$

The first term is from the tracklength variation and the second term is from the Poisson counting process.

Likelihood

In practice the variance (without any detector effects) is calculated for each component and summed as the total as in Eq. (5.14) using Eq. (5.19) for muons and $f_{P,\text{em}}^2 = 1/25$ for all other (EM) components. The covariance matrix is calculated using Eq. (5.18). The variance of the total signal is calculated as described in Section 2.1.1 and the covariance matrix is calibrated such that the sum is equal to the variance on the total signal. In this way the covariance matrix is approximately scaled up by a factor 1.1, meaning that the model uncertainty is underestimated with respect to data. See Fig. 5.11 for an example trace model and the corresponding covariance matrix. The uncertainty is dominated by the muon component.

It is now possible to use the covariance matrix in a χ^2 fit, but that is only correct for large signals. Moreover, inverting the covariance matrix is time consuming and can be numerically unstable and lead to fit instabilities. It is simpler (and more correct for small signals which naturally occur in large parts of the trace) to use a Poisson likelihood. The Poisson scale factor can be calculated for each time bin from the total variance in each time bin—the diagonal of the (calibrated) covariance matrix—and the expected signal as in Eq. (5.15). In this way, bin-to-bin correlations are not taken into account, but the uncertainty on each bin is correct and not overestimated when for example the total signal variance is used per bin as in [97, 98].

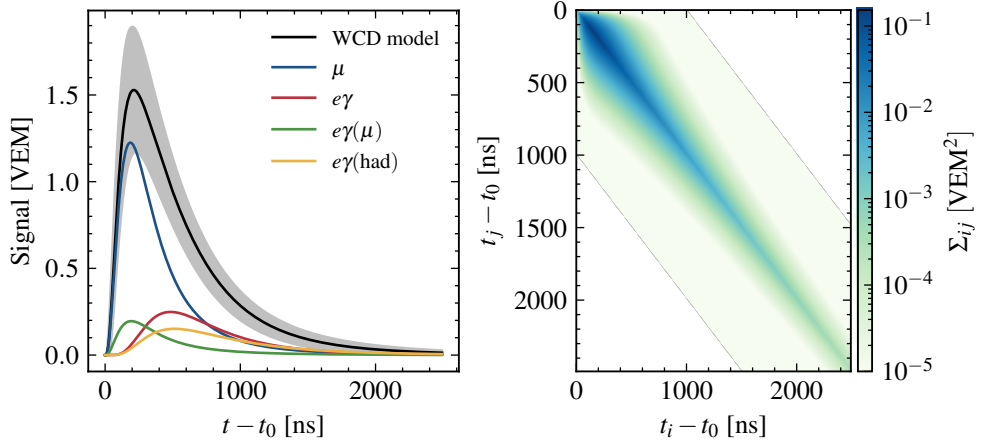


Figure 5.11: On the left: an example of the trace model at $r = 1000$ m, $\psi = 90^\circ$, $\log_{10}(E/\text{eV}) = 19.5$, $\theta = 38^\circ$, $X_{\text{max}} = 750 \text{ g/cm}^2$ and $R^\mu = 1$. The dark shaded band shows the 68% Poisson confidence level from the uncertainty model discussed in the text (without the factor of ≈ 1.1). On the right: covariance matrix of the total signal from the example on the left.

The Poisson deviance, defined as negative two times the Poisson log-likelihood for the model minus the same log-likelihood with $\lambda = n$, so that the normalization and the factorial cancel out. The Poisson deviance for the model on a single trace is calculated as:

$$\begin{aligned} \mathcal{D}_{\text{trace}} &= -2 (\ln \mathcal{L}_P(\lambda, n) - \ln \mathcal{L}_P(n, n)) \\ &= 2 \sum_i \lambda_i - n_i + n_i \ln \frac{n_i}{\lambda_i}, \end{aligned} \quad (5.20)$$

where the sum is over all bins in the trace i and $\lambda_i = \mu_i / f_{P,i}^2$ is the expected number of particles with μ_i given by Eq. (5.11) and $n_i = S_i / f_{P,i}^2$ is the signal converted into the number of particles, where S_i is the mean signal over 3 PMTs in units of VEM charge.

Fig. 5.12 (left) shows a mock trace created from the model by sampling the expected number of particles for each component from a Poisson distribution and for muons also the tracklength from the distribution from [93]. Also the mean and 68% interval around the median from 1000 mock traces are shown and compared with the uncertainty model described above. In Fig. 5.12 (right) the Poisson deviance calculated on 1000 mock traces for two cutoffs on the expected number of counts λ is shown. Observe that for $\lambda \geq 1$ the distribution peaks close to the number of degrees of freedom, but the distribution is nowhere near a χ^2 distribution with that number of degrees of freedom.

The Poisson deviance for an event is the sum over all trace deviances, j , in that event:

$$\mathcal{D}_{\text{event}} = \sum_j \mathcal{D}_{\text{trace},j}. \quad (5.21)$$

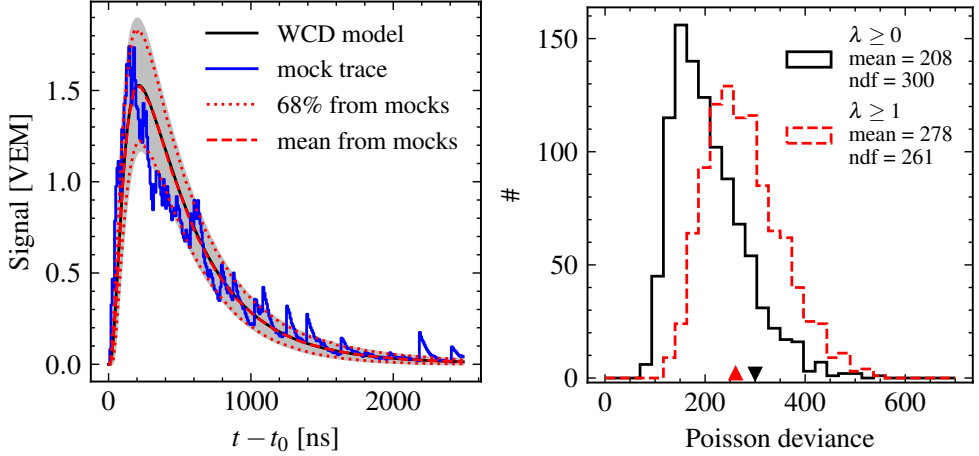


Figure 5.12: Left: example of a mock trace sampled from the model trace shown in Fig. 5.11. The 68% interval around the median from 1000 samples is shown as red dotted lines, while the model 68% interval as explained in the text is shown as a gray band. The small discrepancy comes from the calibration with the total signal uncertainty which is not included in the sampling. Right: the Poisson deviance distribution as calculated on 1000 mock traces with respect to the model shown on the left. On all bins with $\lambda \geq 0$ (black) and only for bins that have $\lambda \geq 1$ (red). The (upside down) triangle denotes the number of degrees of freedom for $\lambda \geq 0$ ($\lambda \geq 1$).

5.5 Fit procedure

The time and signal distribution on the ground produced by an extensive air shower in this framework is described by the following parameters: core location, shower axis, E , X_{\max} and R^μ . Because there is limited information from the surface detectors and there can be correlations between parameters it is difficult to determine all parameters at the same time. There already is a good reconstruction of the core location, the shower axis and the energy of the air shower from the standard reconstruction [79]. The emphasis of this thesis is on the determination of X_{\max} and R^μ and so the standard reconstruction is used for the other parameters.

The arrival time of the signal can be difficult to estimate and is per definition biased towards later times for small signals. Any significant deviation of the start time would mean that the time-trace-template is not in the correct place and the fit is bound to fail. On the other hand, the start time does hold valuable information on X_{\max} . To overcome this issue, when fitting the time template shape, the start time is treated as a nuisance parameter. Meaning that the template is allowed to be shifted to the left and the right within the uncertainty on the start time (as from [108]).

The following fit procedure turned out to work well, although it is very well possible that things can be improved. For example, it is known that the core location from the standard reconstruction is biased for non-vertical showers.

1. Get the core location, shower axis and energy estimator from the standard SD reconstruction.

2. Fit the start time model from [108] to find a new shower axis and get an estimate of X_{\max} .
3. Fit the total signal with the model as described in Section 5.2 with the core and the energy fixed to the standard reconstruction from step 1. Find the best-fit R^μ .
4. At this point remove stations with $\hat{S} < 5$ VEM. To continue, there should be three or more stations for the time-trace-template fit, else the fit is deemed to have failed.
5. Calculate the uncertainty for each time bin and each station from the model as explained in Section 5.4 with the parameters determined in the previous steps.
6. Fit the time-trace-template on the remaining stations for X_{\max} , but constrain the start time with a normal distribution where μ and σ are from the start time model from [108]. Only time bins are used that are below the time that includes 95% of the predicted total signal (from the model).
7. Repeat step 3 and fit R^μ with the improved X_{\max} .
8. Repeat steps 4 to 6 and fit X_{\max} with the improved R^μ , the start time is now fixed to the value as found in step 6.

Because the model was build only for stations with distances to shower core between 500 and 2000 m only these stations are used in steps 3 to 8, but for step 2 all stations are used.

This procedure allows us to determine R^μ and X_{\max} while minimizing the correlation between their residuals. In the following chapter the focus is on determining X_{\max} . Figure 5.13 shows an overview of an SD event from the hybrid data set presented in the next chapter with the templates fitted with the procedure discussed above.

CHAPTER 5. TIME TRACE TEMPLATES

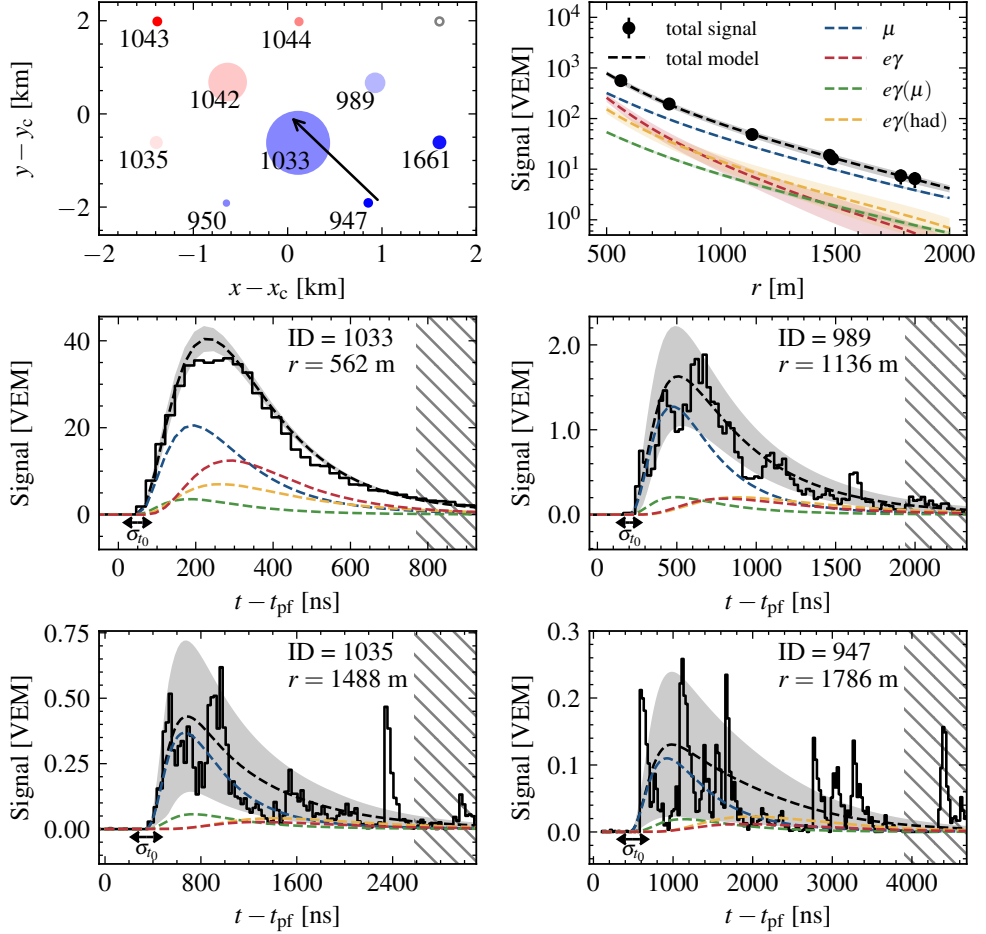


Figure 5.13: Hybrid event with SD ID 5184645. The event has $\log_{10}(E/\text{eV}) = 19.22$, $\theta = 35^\circ$ and $X_{\text{max}} = 761 \pm 10 \text{ g/cm}^2$ as measured by the FD and $X_{\text{max}}(\text{SD}) = 729 \pm 29 \text{ g/cm}^2$ (without any of the corrections as will be discussed in the next chapter). The top left panel shows the SD event on the map where the arrow points to the core from the direction of the incoming shower. The size of the points reflect the total signal and the color the start time of the signal (blue is early, red is late). The top right panel shows the total signal as a function of the distance to the shower core in the shower plane along with the total signal as predicted by the templates for the four components as dashed lines. The shaded band denotes the difference between $\psi = 0^\circ$ and 180° . In the bottom four panels a selection of time traces from different stations is shown. The solid black line is the averaged signal over the three PMTs and the templates are again given as dashed lines. The shaded band is the uncertainty on the signal as described in Section 5.4. σ_{t_0} denotes the uncertainty on the start time with the double-headed arrow. The crossed out area is not used in the fit because it is outside the time that includes 95% of the predicted signal.

6 Depth of shower maximum with the Surface Detector

In the previous chapter a method was presented to estimate X_{\max} using surface detectors. This $X_{\max}(\text{TTT})$ is estimated by fitting time trace templates (TTT) to the signal traces collected by the water-Cherenkov detectors (WCDs). In this chapter, first, some data selection criteria are discussed. Secondly, it is important to address the differences of the parametrization of the start time in MC and data. Thirdly, validation of the method on MC simulations is presented. Finally, the method is applied to real data. After, a calibration with hybrid data, the method is applied to the full SD data set.

6.1 Data selection

The data set from the Pierre Auger Observatory used in this work contains all Observer reconstructed events from data set presented at the ICRC 2019 [117]. This data set includes events from 01-01-2004 to 31-08-2018 reconstructed with Offline v3r99. All quality cuts with their efficiencies are summarized in [Table 6.1](#).

SD data

All events are subject to some basic quality cuts. Only events are selected that fulfill the most stringent SD trigger condition: 6T5, which ensures that all 6 SD stations surrounding the hottest station are functioning (see [Chapter 2](#)). Also, events that are taken during periods when there were known problems with the data acquisition or when there was lightning close to the array are removed. Only showers with $\theta < 53^\circ$ and $\log_{10}(E/\text{eV}) > 19$ are selected in accordance with the model that was built in [Chapter 5](#).

For showers with an X_{\max} below or close to the ground the universality principle is not expected to hold and these showers were not used in the model building process. For FD measurements these showers are cut away using the field-of-view selection (see [Section 6.1.2](#)). With the SD it is not possible to make such a cut (but see [112] for an attempt). It is also not possible to cross-calibrate such showers with the FD, because they are simply cut away. Therefore, it was decided to not use showers with $\sec\theta < 1.1$ ($\theta < 24^\circ$) which are often very close to the ground and thus there are almost no FD measured showers available for calibration.

The atmosphere model is taken from GDAS [86]. For some events the GDAS database is not available and these events are not used.

Additional quality cuts are applied to make sure the template fit is successful. For a few events there are less than 3 triggered stations with an expected signal larger than 5 VEM at distances between 500 m to 2000 m, the fit does not have enough information to reconstruct X_{\max} reliably and these events are discarded. Also, if the MINUIT algorithm did not converge the event is not used.

Hybrid data

Hybrid events are simultaneously measured by the FD and the SD. In this work only so called *Golden* events are selected which can be reconstructed by both the FD and SD. These Golden events will be used to calibrate X_{\max} as determined by the time-trace-template fit to the one measured by the fluorescence telescopes. All hybrid events undergo the same basic SD cuts as described in the previous section. Additionally, there are specific quality cuts for the FD measurement as described in detail in [82]. The most rigorous selection makes sure that the X_{\max} distribution is not biased because of the field of view of the telescopes. To have a proper measurement of X_{\max} it is requested that the reconstructed value of X_{\max} is inside the field of view of the telescopes. To ensure that this cut does not lead to biased moments of X_{\max} (due to cutting away very deep or shallow showers) the fiducial field of view is chosen such that all probable values of X_{\max} can be measured (see [82] for details). This cut effectively removes all events with $\theta < 24^\circ$.

MC data

The simulation and reconstruction of air showers with the Offline framework changes (slightly) from time to time. The time trace templates were constructed using a different version than the data set discussed above. To validate the reconstruction of X_{\max} with the template-fit procedure, showers simulated and reconstructed using the Offline version also used for the data reconstruction were used. Showers simulated with EPOS-LHC from the same library (Napoli-Praha) were selected, consisting of 8000 proton and 8000 iron showers with energy $19 < \log_{10}(E/\text{eV}) < 20$ and $\theta < 53^\circ$. Note that these are in fact 1600 showers per primary, but thrown five times at a random position on the array.

These simulated showers were subject to the same quality cuts as all SD events as listed in Table 6.1. Both iron and proton primaries are shown together, because no large differences between primaries were observed for any of the cuts. The largest difference was found for the cut on the hottest station r which removes 4% more proton showers than iron showers, which is slightly more than can be expected from statistical fluctuations.

Cut on position of hottest station

The model described in Chapter 5 was optimized for signals obtained from detectors located at distances to the shower core between 500 m to 2000 m as measured in the shower plane. This was motivated by the fact that stations with $r < 500$ m are likely to have a signal that is large enough to saturate the low-gain channel for showers with $E > 10 \text{ EeV}$ and thus cannot be (easily) used in the template fit. Another issue that was found for events that have their hottest station (the one closest to the core) with $r < 500$ m is that these events have a poorer core reconstruction, either due to losing information from the low gain saturated station or because of the LDF used by the standard reconstruction fails to properly describe signal sizes from stations closest to the core [79]. If the core position is biased then the core time and hence also the time with respect to the plane front is incorrect and this can have a large influence determining the best-fit template for a station time trace. As can be seen in Fig. 6.1

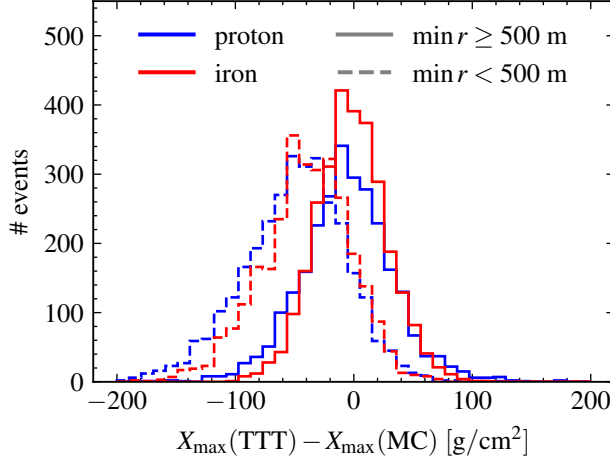


Figure 6.1: Distribution of the difference between X_{\max} as determined by the template fit and the MC value. The dashed line are showers where the hottest station is closer to the core than 500 m.

the reconstructed X_{\max} from the time-template fit is biased for events with a station for which $r < 500$ m. A downside of removing events with a station with $r < 500$ m is that this cuts away more events that are inclined. Nevertheless, events with a station with $r < 500$ m are removed in the following, which removes about $\sim 50\%$ of all events. In addition, the maximum allowed distance for the hottest station is set at 1000 m. The latter cut only removes a few events in which the station closest to the reconstructed core was not working, these events are not well reconstructed by the SD.

Table 6.1: Quality cuts applied to the data sets as described in the text. For hybrid and SD events the selection is made on the SD reconstructed energy, but for MC simulated events the selection is made on the MC energy.

	SD		Hybrid		MC	
	#	%	#	%	#	%
Basic SD cuts, $\log_{10}(E/\text{eV}) > 19$ and $24^\circ < \theta < 53^\circ$	11 232	100.0	1641	100.0	11 595	100.0
FD quality cuts	-	-	521	31.7	-	-
GDAS atmosphere	11 074	98.6	519	31.6	-	-
Number of stations in fit > 3	10 888	96.9	517	31.5	11 567	99.8
Fit successful	10 878	96.8	516	31.4	11 566	99.7
$\min r > 500$ m	5120	45.6	245	14.9	5517	47.6
Total	5120	45.6	245	14.9	5517	47.6

6.2 Start time of the signal

The curvature of the shower front that is determined by the start time of the signal is correlated with X_{\max} in data as was shown in [107, 108]. However, the relation between the radius of curvature and X_{\max} in MC simulations is different than what is found in data. In concordance with [107, 108], the radius of curvature R_c is converted into an atmospheric depth

$$DX_{R_c} = \frac{X(h_{\text{ground}}) - X(h_{\text{ground}} + R_c \cos \theta)}{\cos \theta} \quad (6.1)$$

and DX_{R_c} is calibrated with DX_{\max} with the following function

$$DX_{\max} = a + b \left(DX_{R_c} - \overline{DX_{R_c}} + c(\log_{10}(E/\text{eV}) - \overline{\log_{10}(E/\text{eV})}) \right), \quad (6.2)$$

where $\overline{DX_{R_c}} = 855.4 \text{ g/cm}^2$ and $\overline{\log_{10}(E/\text{eV})} = 19.15$ are the average values from the hybrid data set used in [108]. The distance to X_{\max} is defined as:

$$DX_{\max} = X_{\text{vg}} \sec \theta - X_{\max} \quad (6.3)$$

with X_{vg} the vertical slant depth at the surface. The fit parameters a, b, c are determined by an orthogonal distance regression. The uncertainties on DX_{R_c} are given in [108]. The uncertainties on DX_{\max} for MC simulations are determined by the bin size $\sigma_{DX_{\max}} = 2.5 \text{ g/cm}^2$. For hybrid data $\sigma_{DX_{\max}}$ is either taken as the statistical uncertainty obtained from the Gaisser-Hillas fit to the longitudinal profile or the parametrization from [82] as a function of energy whichever is larger. To make a fair comparison to hybrid data only simulations that have an X_{\max} above ground level are used. Fitted values for a, b, c are given in Table 6.2 for hybrid data and the simulations used in this chapter.*

The relation between DX_{R_c} and the distance to X_{\max} is shown in Fig. 6.2 for different simulations sets and for hybrid data. It is clear that data is not described by any of the simulations. It is not straightforward to identify the exact reason for this. Nonetheless, it is also not surprising for several reasons:

- The start time is determined by the first arriving particle which is most likely a muon. It is thus related to the muon production depth. There is some evidence that the relation between X_{\max}^{μ} and X_{\max} is also not described by any of the simulations (see Chapter 4).
- The total number of muons is higher in data than in simulations (using the mass composition derived from FD measurements) [63]. An increase in the number of muons will also move the mean arrival time forward at low particle multiplicities (far away from the core) because the WCD is most sensitive to muons. This increases the radius of curvature and changes the relation between DX_{R_c} and DX_{\max} .

*The showers shown in this section are subject to the same quality cuts as reported in Table 6.1 except that showers with $\theta < 24^\circ$ are also included and no cut is made on the number of station or if the template fit is successful. It was verified that including these cuts did not significantly change any of the best-fit values reported in Table 6.2.

CHAPTER 6. DEPTH OF SHOWER MAXIMUM WITH THE SD

Table 6.2: Best-fit parameter values of the calibration function Eq. (6.2) for DX_{\max} as a function of DX_{R_c} . Other simulated shower sets are shown in Table B.1.

	a [g/cm ²]	b [-]	c [g/cm ²]	χ^2/ndf [-]
Hybrid data set	460 ± 6	1.060 ± 0.040	-147 ± 23	286/247
UB v3r99p2a EPOS-LHC iron	526 ± 1	1.200 ± 0.005	-132 ± 2	5771/4835
UB v3r99p2a EPOS-LHC proton	533 ± 2	1.287 ± 0.006	-138 ± 2	7836/4642

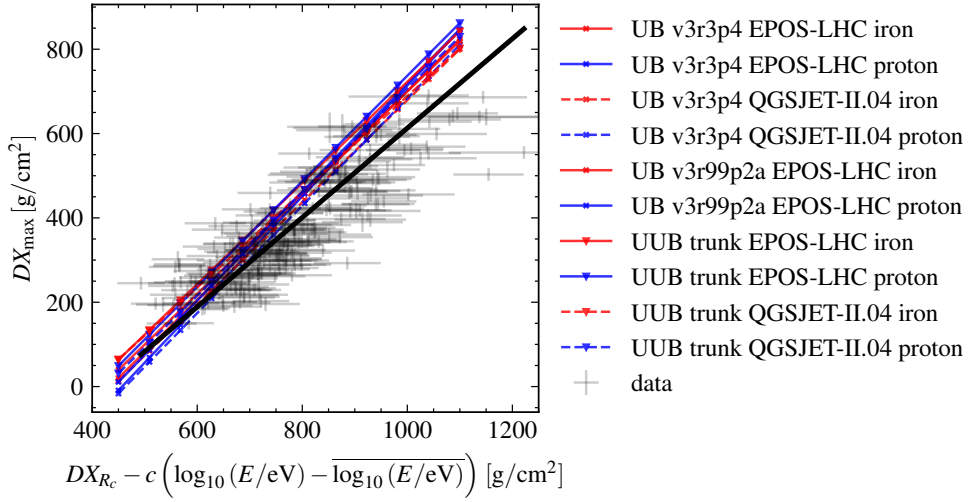


Figure 6.2: Correlation of DX_{R_c} and the distance to X_{\max} for data in black. The colored lines show Eq. (6.2) with values a, b, c as given in Table B.1.

- In simulations a thinning algorithm is used. Although it was shown that the shape of the arrival time distribution is not much altered [96] it is conceivable that the arrival time in individual stations is biased due to the unthinning procedure.
- Technical details. For example, the determination of the start time might be slightly different in simulations than in data. Also, it was noticed that the radius of curvature is different in simulations with upgraded electronics (UUB). This was expected since the sampling frequency is three times faster with respect to the old electronics which makes the start time of the UUB more precise but also on average $8.33\overline{3}$ ns later.

In the time-trace-template fit procedure (Section 5.5), the start time is treated differently in MC and in data. In simulations the average of the parametrization (a, b, c) between proton and iron induced showers using the EPOS-LHC interaction model and reconstructed with Offline v3r99p2a, is used. In data the calibration on hybrid data is used as listed in Table 6.2.

6.3 Validation using MC simulations

The time-trace-template fit is validated on proton and iron induced simulated showers with the EPOS-LHC interaction model reconstructed with [Offline](#) version v3r99p2a. Only showers that survive the quality cuts as described in [Table 6.1](#) are used.

Investigating the bias on the distance to X_{\max}

Showers for which all stations are positioned more than 500 m from the core still have a small bias of a few g/cm^2 when comparing $X_{\max}(\text{TTT})$ to the true MC value. It was found that those showers with X_{\max} values close to the surface have the largest bias as was anticipated because universality does not properly describe these showers. Nevertheless, the bias appears to be well described as a function of DX_{\max} ([Eq. \(6.3\)](#)). The bias is largely universal for the proton and iron induced showers used here. A smaller additional effect of the primary energy is also present. The mean bias of the subset of showers from proton and iron together (numbers are roughly equal) is calculated in bins of the primary energy ($\Delta \log_{10}(E/\text{eV}) = 0.1$) and DX_{\max} . For DX_{\max} the bins range from $0 \text{ g}/\text{cm}^2$ to $750 \text{ g}/\text{cm}^2$ with a width of $50 \text{ g}/\text{cm}^2$ and an overflow bin at the lower and upper edge. The 2D profile for the bias $\mu = \langle DX_{\max}(\text{TTT}) - DX_{\max}(\text{MC}) \rangle$ is fitted using the following function

$$\mu = p_0 + \begin{cases} p_1 y + (p_2 + p_3 y)x + p_4 x^2 + p_5 x^3 & \text{for } DX_{\max} \leq 600 \text{ g}/\text{cm}^2 \\ 0 & \text{otherwise} \end{cases}, \quad (6.4)$$

where $x = DX_{\max}/(600 \text{ g}/\text{cm}^2) - 1$ and $y = \log_{10}(E/\text{eV}) - 19.5$. For $DX_{\max} > 600 \text{ g}/\text{cm}^2$ the bias is independent of energy and DX_{\max} so the correction is set to a constant value by hand to mitigate an, otherwise, artificial correction from the cubic term. The best-fit values for p_i are given in [Table 6.3](#). The bias and the surface obtained from [Eq. \(6.4\)](#) as a function of DX_{\max} and $\log_{10}(E/\text{eV})$ is shown in [Figure 6.3](#). [Figure 6.4](#) shows the bias as a function of DX_{\max} for two energy intervals. Note that the bias on DX_{\max} is just the opposite of the bias on X_{\max} . Therefore, showers close to the ground (small DX_{\max}) are estimated with a too small value for X_{\max} .

The bias for showers with $DX_{\max} < 50 \text{ g}/\text{cm}^2$ is more than $50 \text{ g}/\text{cm}^2$ ($\approx 5\%$ in X_{\max}). The average correction as a function of energy ranges from $-10 \text{ g}/\text{cm}^2$ to $10 \text{ g}/\text{cm}^2$. The bias on X_{\max} as a function of $X_{\max}(\text{MC})$ before and after the correction is shown in [Fig. 6.5](#). Small ($< 2 \text{ g}/\text{cm}^2$) differences between proton and iron induced showers remain.

Table 6.3: Best-fit values for the free parameters in [Eq. \(6.4\)](#) as determined from the 2D distribution of a mix of proton and iron induced showers as shown in [Fig. 6.3](#). The full covariance matrix is reported in [Table B.2](#).

p_0	p_1	p_2	p_3	p_4	p_5	χ^2/ndf
3.5 ± 0.7	-3.6 ± 2.3	-30 ± 8	57 ± 5	-171 ± 21	-200 ± 16	172/130

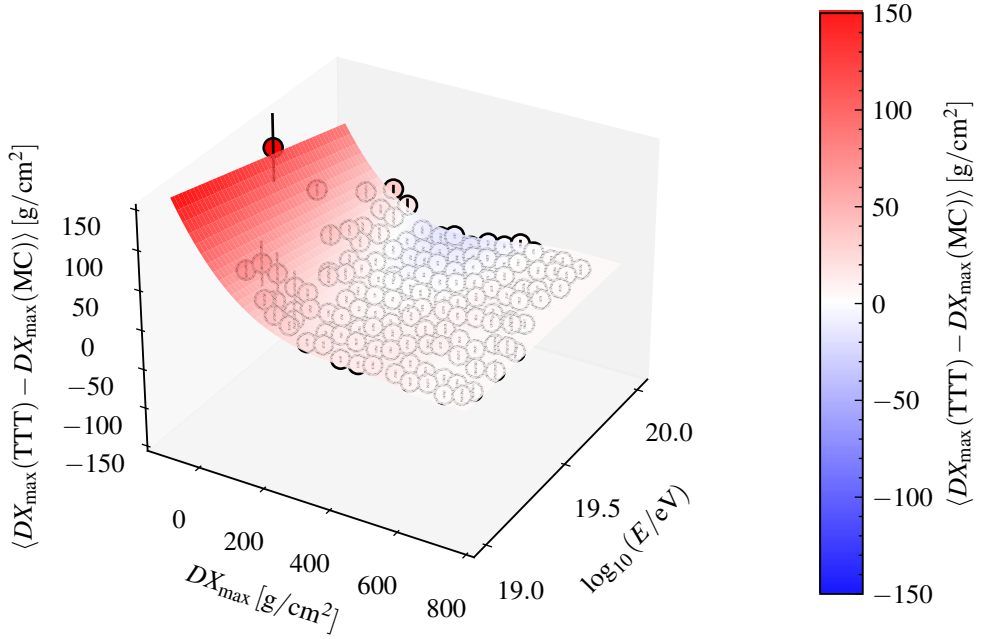


Figure 6.3: The bias on DX_{\max} for a mix of proton and iron induced showers as described in the text.

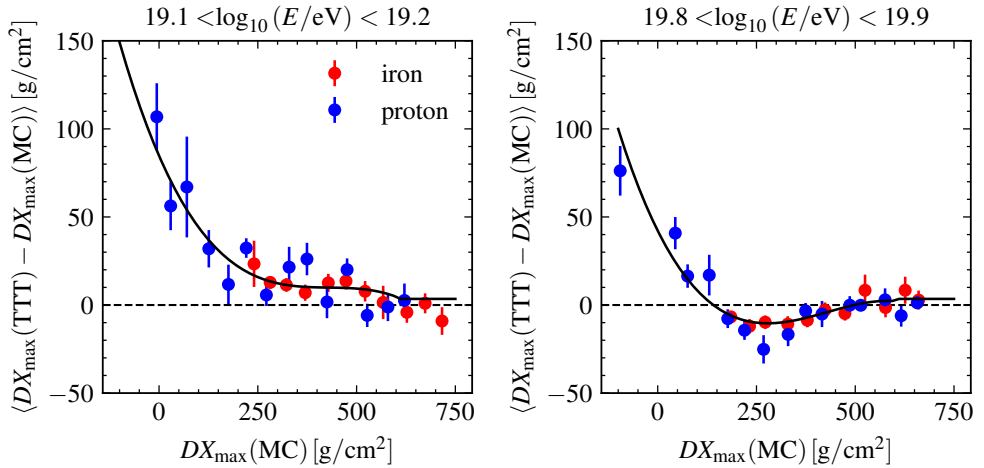


Figure 6.4: The bias on DX_{\max} for two energy intervals as a function of $DX_{\max}(\text{MC})$. The black line is Eq. (6.4) with the parameters given by the fit to the 2D distribution in Fig. 6.3 which is a mix of proton and iron induced showers (shown here separately).

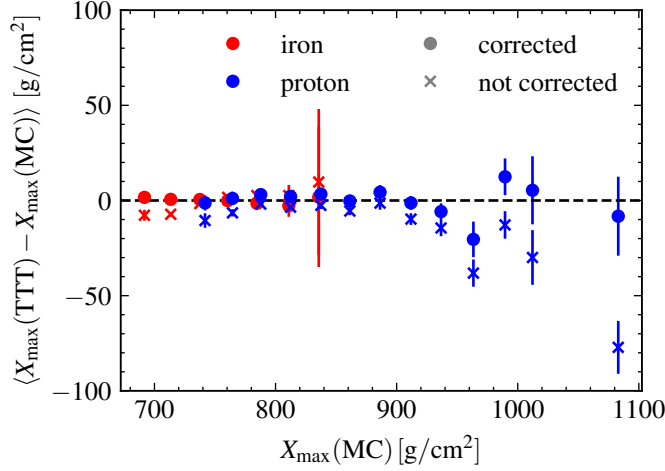


Figure 6.5: The bias on X_{\max} as a function of $X_{\max}(\text{MC})$ before and after applying the correction from Eq. (6.4).

X_{\max} resolution

The distribution of the difference between the reconstructed X_{\max} (with and without the correction described before) and the MC-true X_{\max} is shown on the left in Fig. 6.6. On the same figure on the right the resolution of X_{\max} is shown as a function of energy. The resolution (the width of the distribution) is much narrower for showers initiated by iron nuclei than by protons which can be understood from the increased shower-to-shower fluctuations in proton induced showers that lead to more variation in the particle distribution on the ground. The resolution improves slightly with energy from 35(45) g/cm² at $E = 10^{19}$ eV up to 25(35) g/cm² at $E = 10^{20}$ eV for iron (proton) induced showers. Seemingly paradoxical, the resolution becomes slightly worse after correcting for the bias from DX_{\max} . This is mainly caused by relatively vertical proton induced showers that happen to have an overestimated value of X_{\max} (contrary to the average underestimation of X_{\max} for these showers). The bias correction moves these events even further into the tail of the distribution, increasing the resolution slightly. As will be shown in the next section the bias correction is essential for correctly determining the first and second moment of X_{\max} .

Validation of the moments of X_{\max}

The first two moments of the distribution of X_{\max} —mean and standard deviation—are commonly reported to study the mass composition. The validation of the two moments as reconstructed with the time-trace-template method on the subset of MC simulations is shown in Fig. 6.7. The standard deviation of the true X_{\max} distribution is estimated by subtracting the resolution of the method (as shown in Fig. 6.6) in quadrature. Note, that the resolution is different for proton and iron. Both moments are well reconstructed for MC events as a function of energy after including the bias correction.

CHAPTER 6. DEPTH OF SHOWER MAXIMUM WITH THE SD

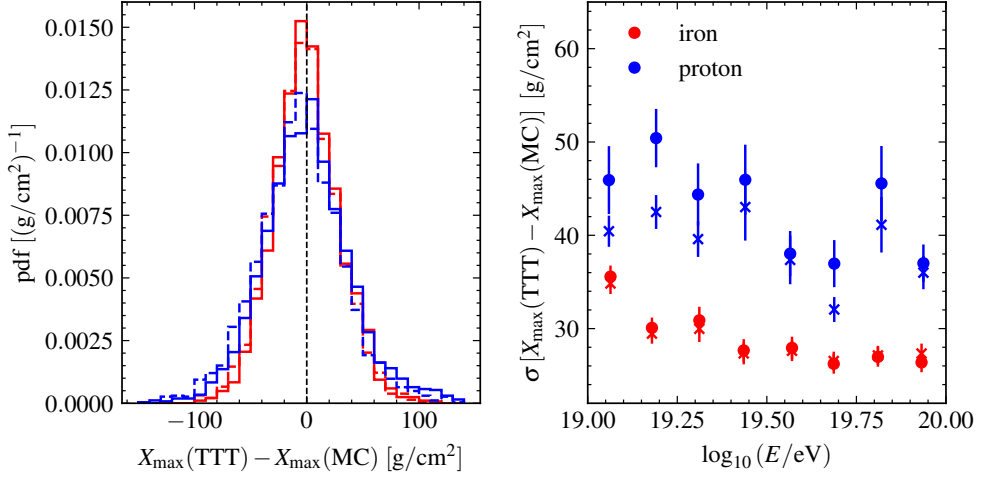


Figure 6.6: On the left: distribution of the difference between the reconstructed X_{\max} (dashed line: without correction, solid line: with correction) and the MC value. On the right: the resolution—standard deviation—of the difference as a function of the primary energy. Crosses: without correction. Solid circles: with correction.

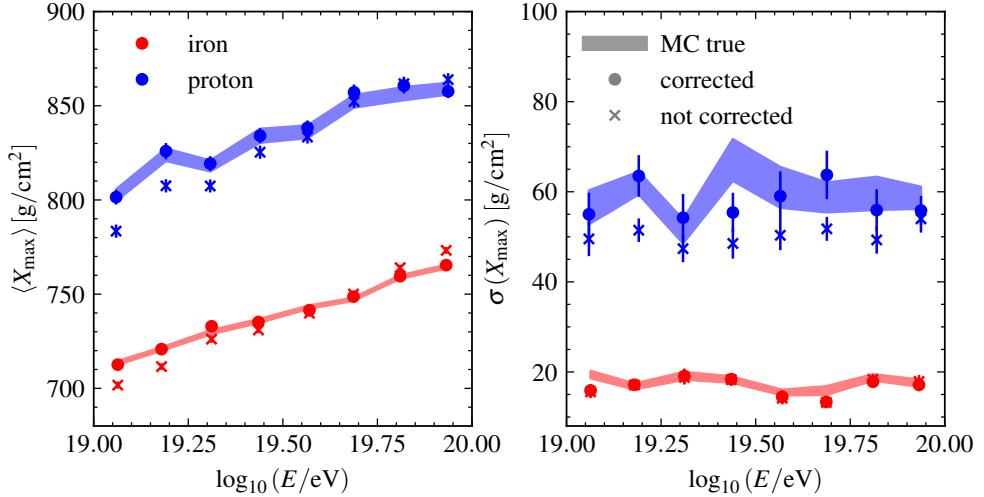


Figure 6.7: On the left the mean of the X_{\max} distribution as a function of energy. On the right the standard deviation of the X_{\max} distribution as a function of energy. $\sigma(X_{\max})$ is reconstructed by subtracting the resolution as shown in Fig. 6.6. The shaded area denotes the moment obtained from the true MC X_{\max} , where the band denotes the statistical uncertainty from the limited number of showers. The solid circles are the reconstructed X_{\max} moments with bias correction and the crosses are without the bias correction.

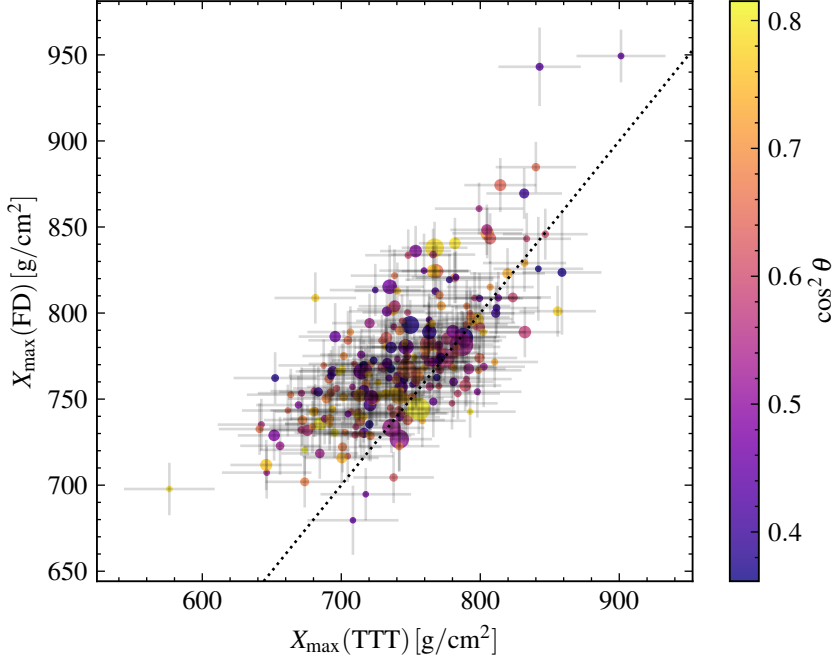


Figure 6.8: Correlation between $X_{\max}(\text{TTT})$ as determined by the time-trace-template fit and X_{\max} as measured by the FD. The size of the points scale with the energy and the color indicates the zenith angle. The error bars on $X_{\max}(\text{TTT})$ are calculated with Eq. (6.5).

6.4 Calibration on hybrid data

The hybrid data set can be used to calibrate $X_{\max}(\text{TTT})$ to the X_{\max} value as measured by the FD. This correction is necessary because none of the simulations describe the data (as seen in Section 6.2), in addition, systematic uncertainties that arise from different hadronic interaction models are removed.

For convenience from now on $X_{\max}(\text{TTT})$ refers to the bias corrected value from this analysis.

Calibration

The correlation between $X_{\max}(\text{TTT})$ and $X_{\max}(\text{FD})$ is shown in Fig. 6.8. The Pearson correlation coefficient is 0.69. On average $X_{\max}(\text{TTT})$ is $28.3 \pm 2.8 \text{ g/cm}^2$ below $X_{\max}(\text{FD})$. There is no evidence for an energy or zenith angle dependent bias as can be judged from Fig. 6.9. Therefore, a constant offset is used so that $X_{\max}(\text{SD}) = X_{\max}(\text{TTT}) + 28.3 \text{ g/cm}^2$.

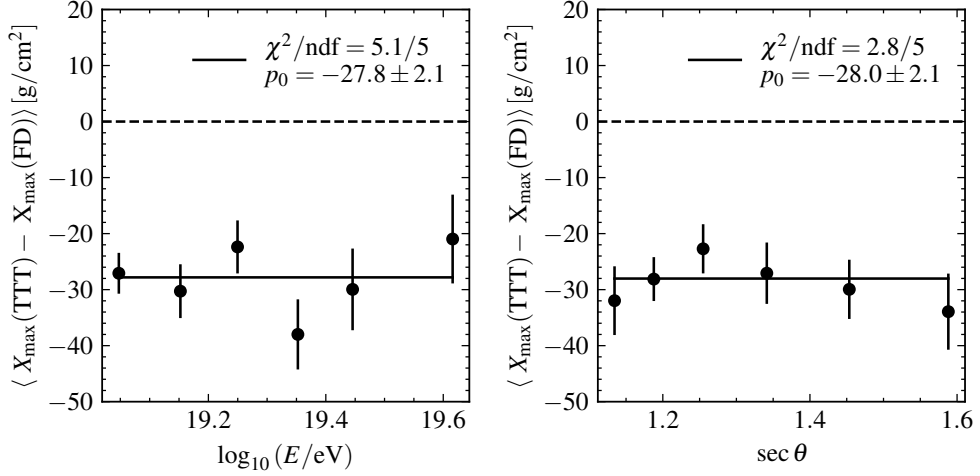


Figure 6.9: Bias as a function of energy (left) and zenith angle (right). p_0 is the best-fit for a constant line.

Resolution from hybrid data

The standard deviation of $X_{\max}(\text{SD}) - X_{\max}(\text{FD})$ as a function of energy is shown in Fig. 6.10 (left). It is described with the following function

$$\sigma [X_{\max}(\text{SD}) - X_{\max}(\text{FD})] (E) = q_0 + q_1 \frac{10 \text{ EeV}}{E}. \quad (6.5)$$

The coefficients are determined to be $q_0 = 25 \pm 4 \text{ g/cm}^2$ and $q_1 = 11 \pm 6 \text{ g/cm}^2$ with a correlation between q_0 and q_1 of $\rho = -0.95$.

The $X_{\max}(\text{FD})$ resolution is taken from [82]. It is approximately constant in this energy range $\sigma(X_{\max}(\text{FD})) \sim 15 \text{ g/cm}^2$. The FD resolution is subtracted in quadrature to get the resolution on $X_{\max}(\text{SD})$. Figure 6.10 (right) shows the pull distribution of $(X_{\max}(\text{SD}) - X_{\max}(\text{FD}))/\sigma$. The pull distribution is well described by a standard normal distribution: the deviance $\mathcal{D} = 251$ for 245 data points.

6.5 Surface detector only data

The method is now applied to the full SD data set. The focus is on determining the first and second moment of the X_{\max} distribution as a function of energy.

Zenith dependent bias

No bias as a function of the zenith angle was found when calibrating $X_{\max}(\text{SD})$ on hybrid events. However, when using the full SD data set (even though very vertical events are already removed) the mean of $X_{\max}(\text{SD})$ is only flat as a function of zenith angle for $\sec \theta \gtrsim 1.25$ ($\theta = 36^\circ$) as can be seen in Fig. 6.11 (left). The difference between vertical ($24^\circ < \theta < 36^\circ$) and horizontal showers ($36^\circ < \theta < 53^\circ$) is at most 20 g/cm^2 . Below $\sec \theta = 1.25$ the shape is well described by a straight line.

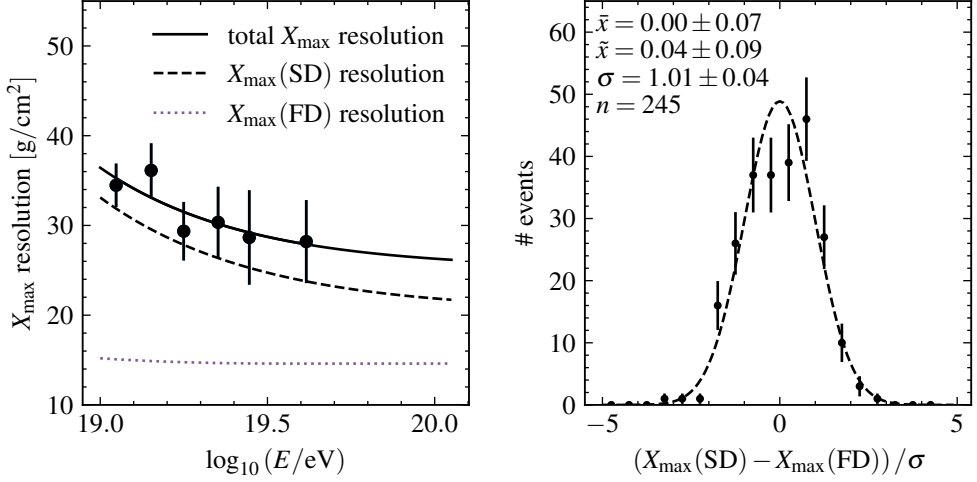


Figure 6.10: On the left: X_{\max} resolution as a function of energy. The X_{\max} (FD) resolution is subtracted in quadrature to get the X_{\max} (SD) resolution. On the right: pull distribution of the difference between X_{\max} (SD) and X_{\max} (FD). σ is the uncertainty on X_{\max} (SD) and X_{\max} (FD) added in quadrature. The dashed line shows the standard normal distribution. The numbers in the corner show from top to bottom: the mean, the median, the standard deviation and the number of events.

A broken line with a break at $\sec \theta = 1.25$ is fitted to the data points in Fig. 6.11 (left). The best-fit slope for $\sec \theta < 1.25$ is found to be $133 \pm 9 \text{ g/cm}^2$ with a $\chi^2/\text{ndf} = 14.5/14$. The position of the break and the slope are correlated. Two other choices for the break are shown in Fig. 6.11 (left). A break at $\sec \theta = 1.2(1.3)$ gives a best-fit slope $209(94) \text{ g/cm}^2$ with a $\chi^2/\text{ndf} = 23(20)/14$.

Vertical events ($\sec \theta < 1.25$) are then shifted by the slope of this line up to where $\langle X_{\max}(\text{SD}) \rangle$ is flat. The improved $X_{\max}(\text{SD})$ is:

$$X_{\max}(\text{SD}) = \begin{cases} X_{\max}(\text{SD})' + 133 \text{ g/cm}^2 (\sec \theta - 1.25) & \text{if } \sec \theta < 1.25 \\ X_{\max}(\text{SD})' & \text{if } \sec \theta \geq 1.25. \end{cases} \quad (6.6)$$

The average correction over the whole data set is about 6 g/cm^2 and constant as a function of energy. Changing the position of the break in $\sec \theta$ and the corresponding best-fit slope changes the average correction by less than $\pm 1 \text{ g/cm}^2$.

Also $\sigma(X_{\max}(\text{SD}))$ is affected as can be seen in Fig. 6.11 (right). Again, the distribution appears flat for $\sec \theta > 1.25$ ($\chi^2/\text{ndf} = 9.3/8$). In the case of $\sigma(X_{\max}(\text{SD}))$ the zenith angle effect can be an artefact of the method, but may also be the result of a zenith angle dependency of the resolution parametrized on hybrid data. No such a dependency could be found with the limited number of hybrid events used for the parametrization. Interestingly, in MC simulations the resolution decreases for more inclined events which would cause the opposite behaviour as found here in data. For $\sigma(X_{\max}(\text{SD}))$ no correction is made, as was done for the mean, but only events with $\sec \theta > 1.25$ are used. This leaves 2731 out of the total of 5210 events.

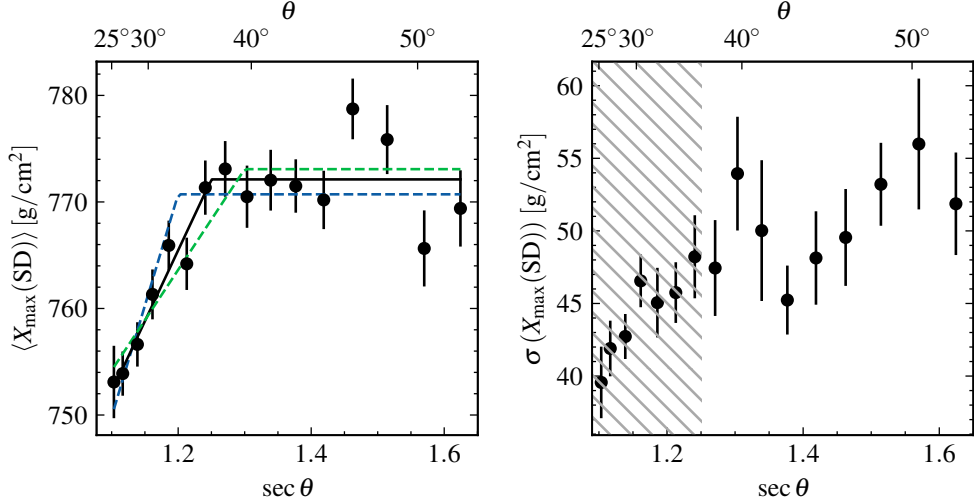


Figure 6.11: (Left) Mean $X_{\max}(\text{SD})$ as a function of zenith angle. The black solid line shows the best-fit broken line with a break at $\sec \theta = 1.25$ and the two dashed lines show the best-fit lines with a break at $\sec \theta = 1.2$ (blue) and 1.3 (green). See text for more details. (Right) Second moment of $X_{\max}(\text{SD})$ as a function of zenith angle. Here $\sigma(X_{\max}(\text{SD}))$ is the (raw) standard deviation calculated from the measured distribution of X_{\max} , meaning: without subtracting the resolution in quadrature.

Systematic uncertainty

First moment – The systematic uncertainty on the mean of X_{\max} is shown in Fig. 6.12. It is composed of the following components:

- Due to the calibration with the FD the largest systematic on the first moment is from the X_{\max} scale measured with FD which is given in [82]. It is energy dependent, but mostly flat in the energy range considered here, and slightly asymmetric. Its value is approximately 10 g/cm^2 in both directions. The FD systematic uncertainty also includes the uncertainty on the knowledge of the atmosphere. Additionally, the calibration on hybrid events has a statistical uncertainty on the constant offset with respect to $X_{\max}(\text{FD})$ which is 2.8 g/cm^2 (see Section 6.4). These systematics are combined and shown as the purple dashed line in Fig. 6.12.
- The zenith angle dependence discussed above creates an average correction of 6 g/cm^2 . The systematic uncertainty is taken as half of this correction in both directions: $\pm 3 \text{ g/cm}^2$. The uncertainty on the slope and position of the break of the correction function give an additional 1 g/cm^2 in both directions. Also a residual yearly diminishing of $\langle X_{\max}(\text{SD}) \rangle$ was found that is probably the result of the aging of the tanks that could not be fully corrected for by using the detector time response based on the monitoring data (see Section 4.4). The average X_{\max} as a function of the year is shown in Fig. 6.13. The maximal difference from the mean is about 3 g/cm^2 in both directions which is taken as

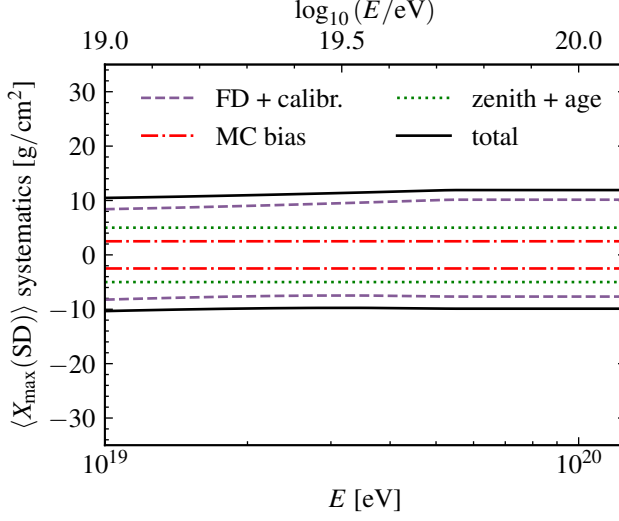


Figure 6.12: Summary of the systematic uncertainties on the mean of $X_{\max}(\text{SD})$ as discussed in the text.

an additional systematic uncertainty. The uncertainty from the zenith angle and the age of the observatory are combined quadratically and shown as the dotted green line in Fig. 6.12.

- The bias correction from MC simulations as in Section 6.3 has some statistical uncertainties that are propagated into $X_{\max}(\text{SD})$ (see Table B.2 for the covariance matrix of the fit parameters). The propagation also includes the uncertainty on the energy scale which is 14% [67]. The average correction that is added to $X_{\max}(\text{TTT})$ as a function of energy is shown in Fig. 6.14, the uncertainty is shown as the filled band. The uncertainty on the correction is included as a 1.5 g/cm^2 systematic uncertainty in both directions. It is constant as a function of energy. An additional 2 g/cm^2 uncertainty was added to account for the small difference between proton and iron after correcting for the bias (shown as the red dashed-dotted line in Fig. 6.12).

All contributions are added in quadrature to get the total systematic uncertainty, which is approximately 13 g/cm^2 in both directions. Additional cross checks are shown in Fig. A.5. Noteworthy, no seasonal or diurnal variation of $X_{\max}(\text{SD})$ were found.

Second moment – The systematic uncertainty on $\sigma(X_{\max}(\text{SD}))$ is shown in Fig. 6.15. The following components are considered:

- The second moment is not directly calibrated with FD measurements. However, the largest contribution to the systematic uncertainty is from the parametrized resolution (Eq. (6.5)). The uncertainty on q_0 and q_1 (including the correlation) are propagated with the uncertainty on the energy scale. The propagated uncertainty is shown in Fig. 6.15 as the purple dashed line. The correlation be-

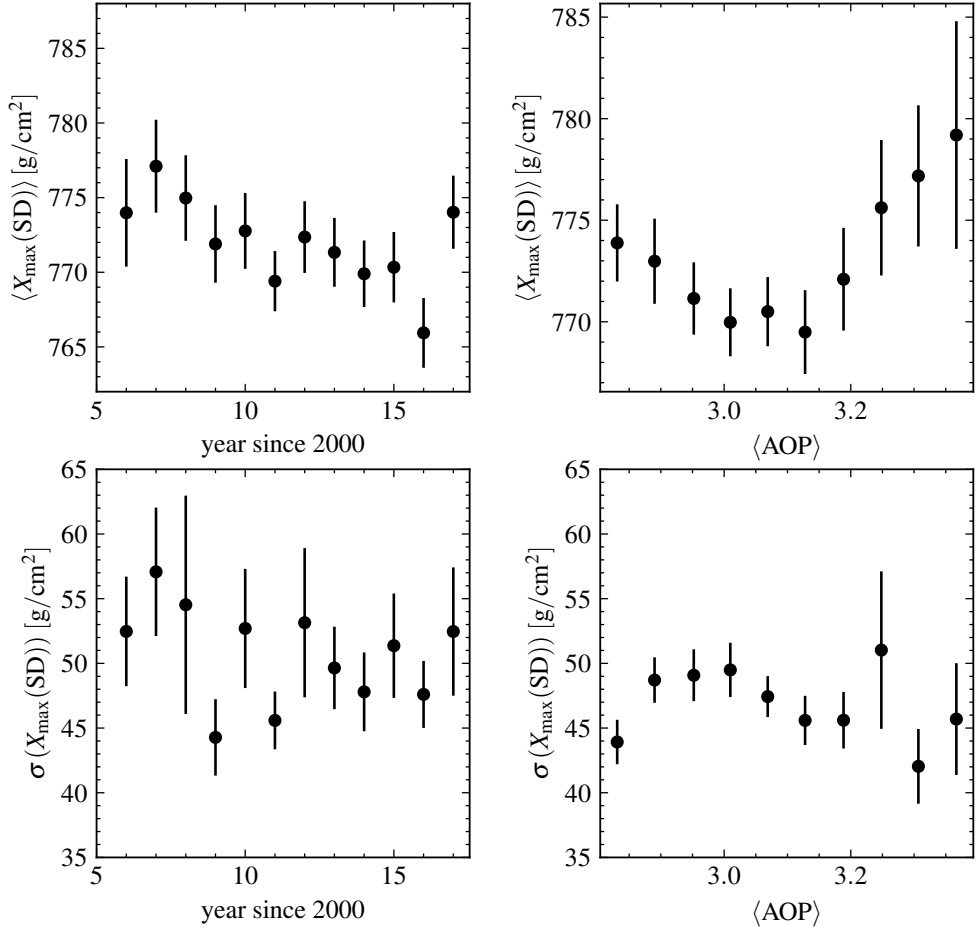


Figure 6.13: Top left: annual decline of $\langle X_{\max}(\text{SD}) \rangle$. Top right: $\langle X_{\max}(\text{SD}) \rangle$ as a function of the average Area-Over-Peak in the SD stations. Bottom left: $\sigma(X_{\max}(\text{SD}))$ as a function of the year. Bottom right: $\sigma(X_{\max}(\text{SD}))$ as a function of the average Area-Over-Peak. For $\sigma(X_{\max}(\text{SD}))$ only events with $\sec \theta > 1.25$ are used.

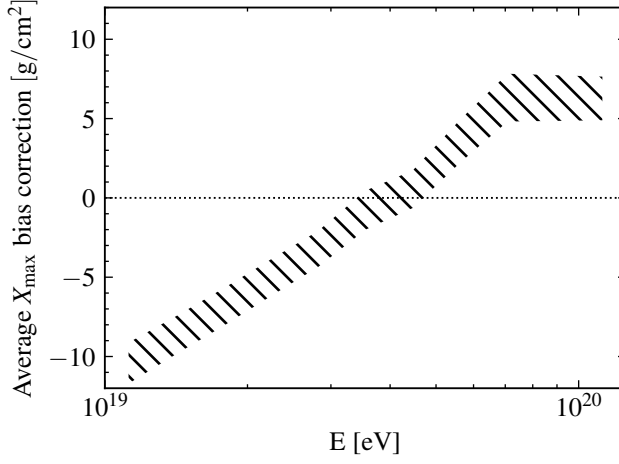


Figure 6.14: Average correction to $X_{\max}(\text{TTT})$ from the MC bias as described in [Section 6.3](#) on the SD dataset. The band shows the uncertainty from the fit parameters as in [Table 6.3](#).

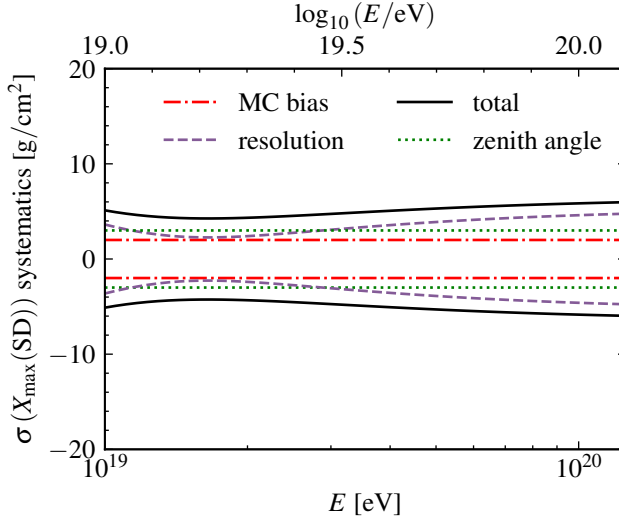


Figure 6.15: Summary of the systematic uncertainties on $\sigma(X_{\max}(\text{SD}))$ as discussed in the text.

CHAPTER 6. DEPTH OF SHOWER MAXIMUM WITH THE SD

tween q_0 and q_1 makes that the uncertainty is smallest at $E \approx 2 \times 10^{19}$ eV. For $E \gtrsim 5 \times 10^{19}$ eV the resolution is extrapolated and so the uncertainty increases.

- Because of the limited number of simulated events the uncertainty on the MC value for $\sigma(X_{\max})$ is not negligible as can be seen from Fig. 6.7. Hence, an additional $\pm 2 \text{ g/cm}^2$ was added to the systematic uncertainty (red dashed-dotted line in Fig. 6.15).
- A possible further increase of $\sigma(X_{\max}(\text{SD}))$ for higher zenith angles is apparent from Fig. 6.11. Accordingly, an additional systematic uncertainty of $\pm 3 \text{ g/cm}^2$ is added.

All contributions are added in quadrature to get the total systematic uncertainty. The total systematic uncertainty is approximately 5 g/cm^2 in both directions. No relation with respect to the age of the SD stations is observed for $\sigma(X_{\max}(\text{SD}))$ as can be seen in Fig. 6.13 (bottom left), although a small effect with respect to the Area-Over-Peak cannot be ruled out (bottom right of that figure). Additional cross checks are shown in Fig. A.6.

Results

First moment – The mean of X_{\max} as a function of energy is shown in Fig. 6.16. The bin width is $\log_{10}(E/\text{eV}) = 0.1$, above $\log_{10}(E/\text{eV}) = 19.9$ all events are included in the last bin that contains 12 events. The highest energy event is at $E = 1.33 \pm 0.11 \times 10^{20}$ eV and has $X_{\max} = 798 \pm 21 \text{ g/cm}^2$ (the uncertainty is from the expected resolution at that energy). The measurement with the FD is also shown [71]. Note, that these FD points come from the same data period and have the same reconstruction, but contain more events than those that were used for the calibration because of less strict cuts on the SD reconstruction. Therefore, the correlation between the FD points and SD points shown here is reduced with respect to what can be expected from the calibration procedure. It was verified that the hybrid events used for the calibration are compatible with the FD points shown in the figure.

Second moment – The standard deviation of the distribution of X_{\max} as a function of energy is shown in Fig. 6.17. The standard deviation of the true distribution of X_{\max} is estimated by subtracting the resolution on $X_{\max}(\text{SD})$ from Eq. (6.5) in quadrature. Only showers with $\sec \theta > 1.25$ are used for the second moment, because of the zenith angle dependency as was discussed above. Because of the diminished number of events available, above $\log_{10}(E/\text{eV}) = 19.7$ an integrated bin is used, it contains 47 events. The distribution of $X_{\max}(\text{SD})$ for each energy bin is shown in Fig. 6.18.

Statistical uncertainties on $\langle X_{\max} \rangle$ and $\sigma(X_{\max})$ are calculated by a bootstrap method. In each energy bin the X_{\max} values are randomly sampled with replacement with the size of the sample the same as the number of events in that bin. On each sample the mean and standard deviation are calculated and this is repeated 10000 times to determine the 1σ uncertainty on $\langle X_{\max} \rangle$ and $\sigma(X_{\max})$. If the uncertainty on $\langle X_{\max} \rangle$ is smaller than what can be expected from the resolution, the latter, divided by the square root of the number of events in that bin, is used instead—this is only the case for the energy bin $\log_{10}(E/\text{eV}) > 19.9$.

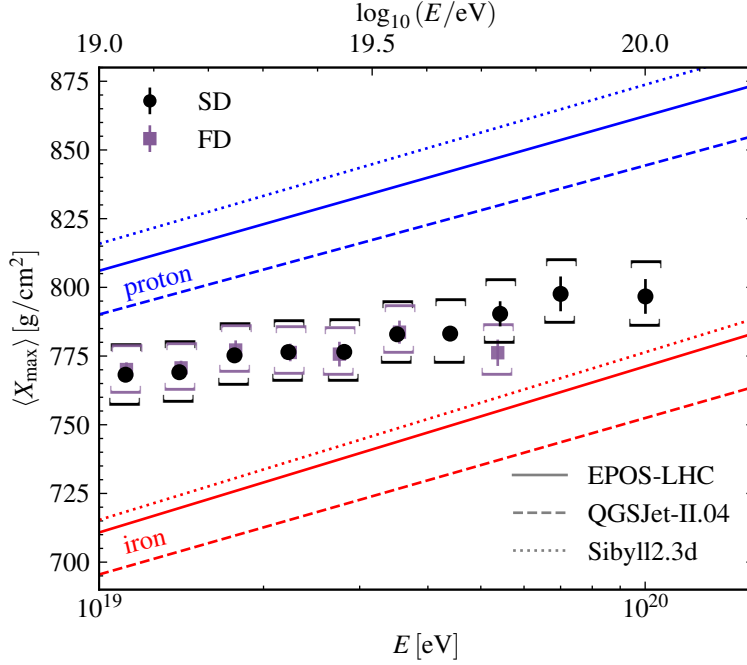


Figure 6.16: First moment of X_{\max} as a function of energy. $X_{\max}(\text{SD})$ of the full SD data (black circles) set is compared to the measurement with the FD (purple squares) [71]. The error bars indicate the statistical uncertainty on the mean and the systematic uncertainty is indicated by brackets.

Discussion

The first moment of $X_{\max}(\text{SD})$ as a function of energy is compatible, within systematics, with FD measurements. The final integrated bin of the FD measurements is the most disparate but still bracketed by the independent systematic uncertainties.[†]

The second moment of $X_{\max}(\text{SD})$ follows the same trend as the FD measurements towards less fluctuations at higher energies as can be seen in Fig. 6.17. At medium energies the points are well in agreement, but there is some tension at low and high energies. At the lowest energy bin $\sigma(X_{\max}(\text{SD}))$ is significantly larger than the FD measurement while at higher energies it is lower.

Other estimates of X_{\max} with the surface detector are available (see list at Chapter 5 on page 49). Currently, the only published (and widely accepted) method is the so called *Delta method*, which utilizes the rise time of the signal in the surface detectors [95, 118]. The Delta method is also calibrated to $X_{\max}(\text{FD})$. In Fig. 6.19 $\langle X_{\max} \rangle$ as obtained by the Delta method is compared to $\langle X_{\max} \rangle$ as presented in this work. There is excellent agreement between the two methods: all points are within

[†]The systematic uncertainty of the FD and SD points as shown in Fig. 6.16 are not independent, because the SD inherits the systematic uncertainty from the FD through the calibration procedure. The independent systematic uncertainty on $\langle X_{\max}(\text{SD}) \rangle$ is about 6 g/cm^2 as can be judged from Fig. 6.12.

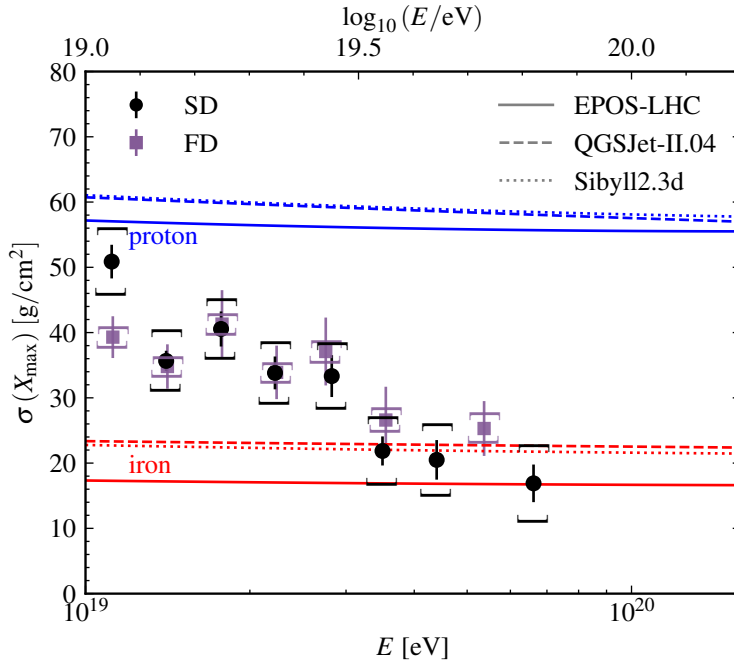


Figure 6.17: Second moment of X_{\max} as a function of energy. Only showers with $\sec \theta > 1.25$ are used for SD. The FD data is from [71]. The error bars denote the statistical uncertainty and the brackets the systematic uncertainty.

$4 \text{ g}/\text{cm}^2$ of each other which is covered by the independent systematic uncertainties of both methods.

The first and second moment of the X_{\max} distribution can be used to infer the mass composition of UHECRs. To achieve this, X_{\max} measurements have to be compared to simulations introducing model dependencies. Three recent hadronic interaction models are shown in the accompanying plots to compare $\langle X_{\max} \rangle$ and $\sigma(X_{\max})$: EPOS-LHC, QGSJET-II.04 and SIBYLL 2.3d. Differences between these models range from $10 \text{ g}/\text{cm}^2$ to $30 \text{ g}/\text{cm}^2$ for the mean of X_{\max} .

A change in the mass composition should show up as change in the slope of $\langle X_{\max} \rangle$ as a function of $\log E$ —what is usually known as the elongation rate:

$$D_{10} = \frac{d\langle X_{\max} \rangle}{d \log_{10}(E/\text{eV})}. \quad (6.7)$$

A constant mass composition has an elongation rate of $D_{10} \approx 60 \text{ g}/\text{cm}^2/\text{decade}$. $\langle X_{\max} \rangle$ as determined in this work at $\log_{10}(E/\text{eV}) > 19$ can be well described by a single straight line with an elongation rate of $D_{10} = 29.7 \pm 3.0 \text{ g}/\text{cm}^2/\text{decade}$ with a $\chi^2/\text{nfd} = 6/8$ as shown in Fig. 6.20. This is compatible with the elongation rate measured by FD above $E = 2 \times 10^{18} \text{ eV}$ which is $26 \pm 2 \text{ g}/\text{cm}^2/\text{decade}$ [71].

The fluctuations of X_{\max} are the result of a combination of the shower-to-shower fluctuations of individual primaries and the mixing of these primaries (if multiple are present). Proton induced showers have the largest shower-to-shower fluctuations, but

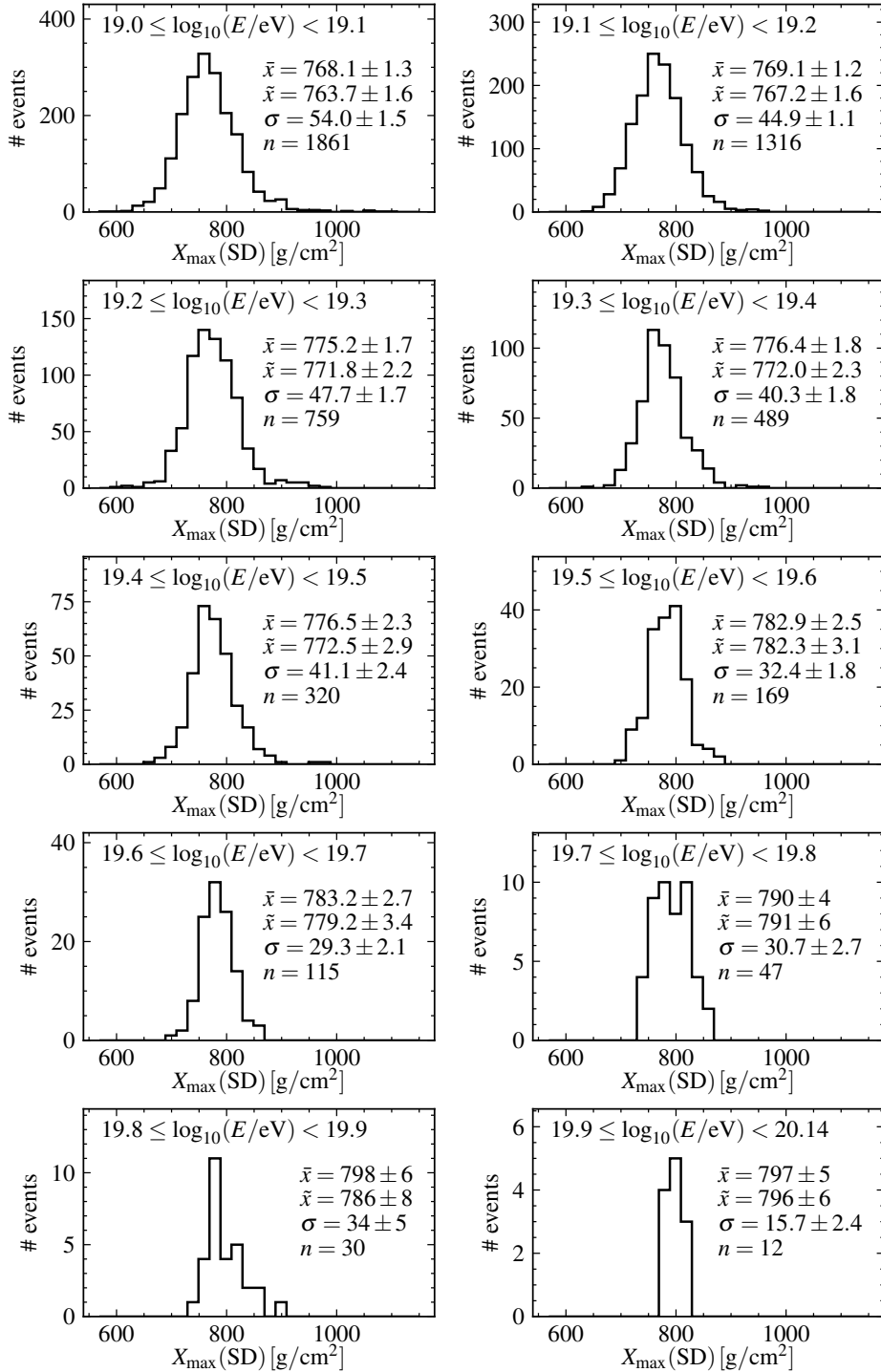


Figure 6.18: $X_{\max}(\text{SD})$ distribution in different energy bins. The bin size in X_{\max} is $20 \text{ g}/\text{cm}^2$. In each figure the mean, median, standard deviation and total number of events are shown.

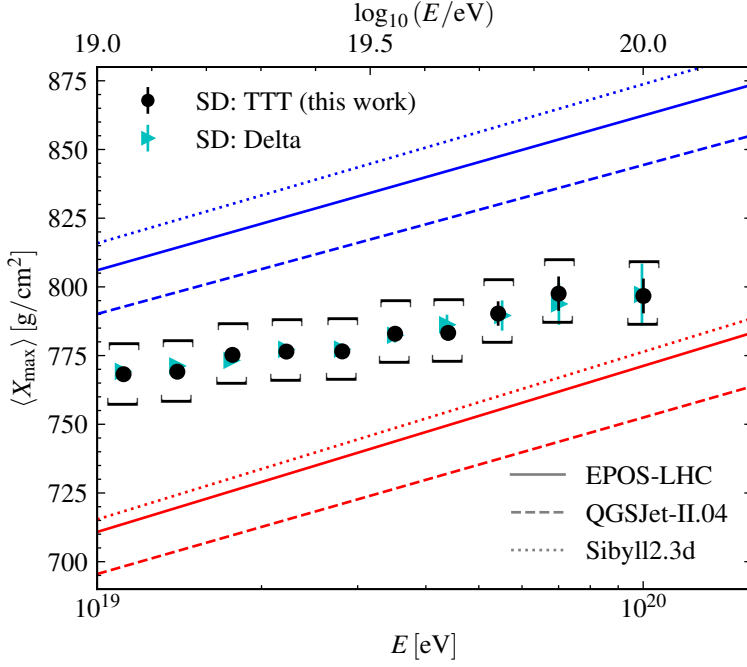


Figure 6.19: Mean X_{\max} as measured with the SD as described in this work compared to X_{\max} as determined by the Delta method [118]. The error bars for the Delta method show the statistical uncertainty only. The independent systematic uncertainty (without the FD scale) of the Delta method is about 7 g/cm^2 [95].

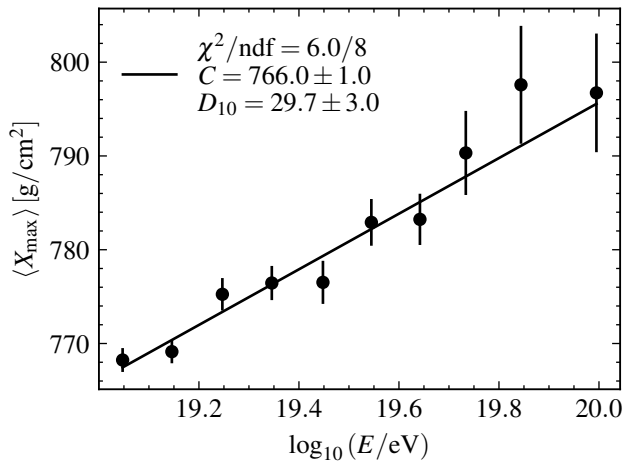


Figure 6.20: $\langle X_{\max} \rangle$ as a function of the logarithm of the energy (same bins used as in Fig. 6.16) with a straight line fit of the form: $\langle X_{\max} \rangle = C + D_{10}(\log_{10}(E/\text{eV}) - 19)$.

for heavier primaries these fluctuations decrease non-linearly so that $\sigma(X_{\max})$ for a pure nitrogen composition will be only $\approx 5 \text{ g/cm}^2$ above iron. The interpretation of $\sigma(X_{\max})$ is complicated by a possibly mixed composition, which would increase the observed fluctuations with respect to a pure composition. Regarding the measurement of $\sigma(X_{\max}(\text{SD}))$ (Fig. 6.17), it appears the fluctuations above $E \approx 40 \text{ EeV}$ are compatible with a pure iron composition. Still, a pure nitrogen composition is also compatible within uncertainties which would be better in line with the $\langle X_{\max} \rangle$ results.

Using predictions by hadronic interaction models it is possible to convert $\langle X_{\max} \rangle$ into a mean atomic mass using:

$$\langle \ln(A) \rangle = \frac{\langle X_{\max} \rangle_{\text{data}} - \langle X_{\max} \rangle_{\text{p}}}{\langle X_{\max} \rangle_{\text{Fe}} - \langle X_{\max} \rangle_{\text{p}}} \ln(56), \quad (6.8)$$

where $\langle X_{\max} \rangle_i$ depends on the energy and the hadronic interaction model. In Fig. 6.21 $\langle X_{\max}(\text{SD}) \rangle$ is converted into $\langle \ln(A) \rangle$ for the three hadronic interaction models. The difference in the energy dependence of X_{\max} between the models is negligible so their differences appear as a constant scaling. The conversion into $\langle \ln(A) \rangle$ suggests an increasingly heavier mass composition up to $E \approx 40 \text{ EeV}$ and a more or less constant atomic mass at higher energies. Note that this break appears at the energy where the flux starts to be heavily suppressed ($E_{\text{sup}} = 46 \pm 9 \text{ EeV}$ [67]).

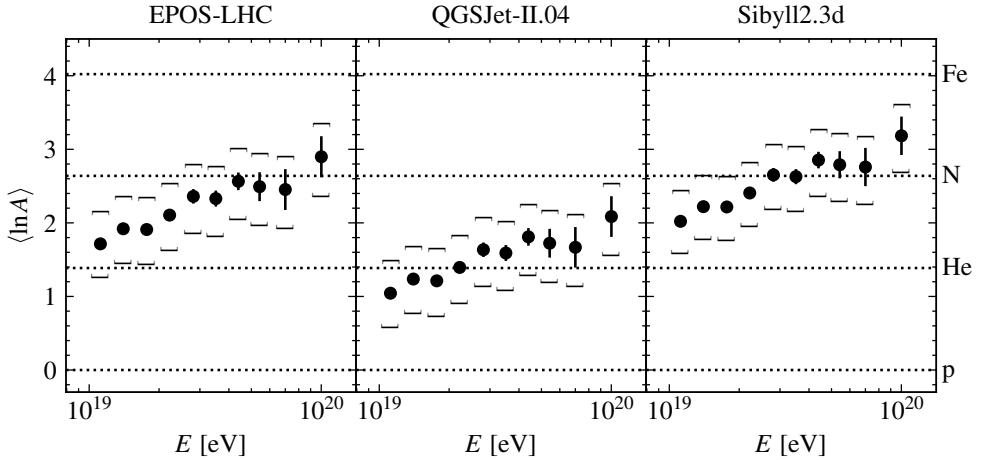


Figure 6.21: $\langle X_{\max} \rangle$ as measured with the SD converted into the mean logarithm of the atomic mass for three different hadronic interaction models. The error bars (brackets) indicate the statistical (systematical) uncertainties propagated from the X_{\max} measurement.

7 Conclusion and outlook

The study of ultra-high-energy cosmic rays is a rich and complex field with many (astro)physics puzzles still to be solved. Whether it is determining the astrophysical origin of UHECRs, studying their interactions in the universe or in the atmosphere, to move forward cosmic particle identification is key. Knowledge of the mass composition of UHECRs will reveal which astrophysical scenarios are viable to accelerate particles up to extreme energies and how these particles make it to Earth. Furthermore, it also determines the secondary photon and neutrino flux created by UHECRs in the universe. On the other hand, primary particle identification is also critical to use cosmic-ray induced air showers as a probe for (hadronic) particle physics at ultra-high energies. For the purpose of interpretation, the mass composition and hadronic physics are inherently intertwined, because to determine the primary mass means to rely on the extrapolation of hadronic physics as measured by accelerator experiments.

At the highest energies, the study of ultra-high-energy cosmic rays is complicated by their extremely low flux, requiring large observatories with many (cost-effective) detector units. In this thesis, the Pierre Auger Observatory is used, the largest available observatory. Already a considerable number of ultra-high-energy events have been detected with its surface detector array, but its water-Cherenkov detectors were never thought to provide enough information to determine the mass composition. However, information on the longitudinal development of the shower is fundamentally connected to the arrival times of particles at the surface as was demonstrated in [Chapter 4](#) and this can be used to determine X_{\max} as was shown in this thesis. An upgrade programme to improve mass composition with the SD by including a thin scintillator detector and radio frequency detection units is underway ([Section 2.3](#)).

Shower max measurement with the SD

In order to use all information available from the WCDs, X_{\max} was determined through a parametric template fit on each event. The well established concept of air shower universality was used to create a model of the time dependent signal in WCDs, described in [Chapter 5](#). The universality paradigm states that any shower, regardless of the hadronic interaction model or primary mass, can be approximated as a function of the primary energy, X_{\max} and the number of muons in the shower. The signal model was built using simulated air showers, with the contemporary hadronic interaction model EPOS-LHC. The simulated showers were subject to a full detector simulation of the Pierre Auger Observatory detectors.

To compare the model to data, it is critical to understand the expected fluctuations of the detected signal as a function of time. It was shown that these can be described by a Poisson distribution, but that the uncertainty on the signal is modified by the detector time response which introduces bin-to-bin correlations. The detector response (as is determined by continuous calibration) has a significant impact on the shape of the time distribution. Aging of the WCDs influences the time response to single particles, an effect that is not currently implemented in the simulation frame-

work. Hence, the effect of the detector was included in the time trace templates allowing to correctly tune the expected time distribution for individual stations.

It was shown that the first and second moment of the X_{\max} distribution can be well reconstructed by the template-fit procedure on MC simulated air showers, but that a correction for close to the ground showers is necessary. After this correction an event-by-event comparison of showers simultaneously measured by the FD and SD revealed that there is an offset of 28 g/cm^2 between X_{\max} measured with the FD and SD, a feature that was also found in [106]. This means that X_{\max} measured by the SD is systematically higher up in the atmosphere, an effect that could be introduced by a problem in understanding the muon component in simulations. A large fraction of the signal is created by muons and a change in the arrival time distribution could be due to a different relation between X_{\max} and X_{\max}^{μ} in simulations compared to reality. Indeed, X_{\max}^{μ} as measured—albeit in a restricted phase space—by the Pierre Auger Observatory is higher up in the atmosphere than what would be expected from the mass composition as measured via X_{\max} [103]. The resolution on $X_{\max}(\text{SD})$ is between 20 and 25 g/cm^2 for showers with $\log_{10}(E/\text{eV}) > 19.5$. Because of the calibration the systematic uncertainty is largely inherited from the X_{\max} scale set by the FD. It is approximately 13 g/cm^2 for the mean and 6 g/cm^2 for the standard deviation.

After the calibration with fluorescence detector data, the method was applied to the high statistics data set using only the surface detector array. A discrepancy was found for very vertical showers, which are removed when calibrating against the FD data, making it impossible to correctly calibrate vertical showers. $X_{\max}(\text{SD})$ measurements from showers with $\sec \theta < 1.25$ are shifted to the mean value of $X_{\max}(\text{SD})$ for showers with $\sec \theta \geq 1.25$ for which $\langle X_{\max}(\text{SD}) \rangle$ was found to be constant as a function of zenith angle. After this shift, the mean X_{\max} as a function of energy as measured by SD is fully compatible with FD measurements and extends the measurement of the elongation rate up to $E = 10^{20} \text{ eV}$ (Fig. 6.16). The elongation rate from 10^{19} eV to 10^{20} eV is well described by a single slope of $30 \text{ g/cm}^2/\text{decade}$. The second moment of the X_{\max} distribution can also be measured by the SD, but only for showers with $\sec \theta > 1.25$. The measurement of $\sigma(X_{\max}(\text{SD}))$ confirms earlier measurements that the fluctuations decrease with increasing energy (Fig. 6.17).

All in all, the mass composition as measured with the SD in this thesis points to an increasingly heavier composition as a function of energy up to 10^{20} eV , although a more constant mean atomic mass in the region of the flux suppression—as is apparent from Fig. 6.21—cannot be ruled out. The fluctuations of X_{\max} imply an almost pure composition above the flux suppression, leaving little room for a substantial proton fraction.

Outlook

The focus of this thesis has been on measuring X_{\max} with surface detectors. In the same way, using the universality paradigm, it is possible to determine the relative number of muons R^{μ} , another excellent mass sensitive parameter. For hybrid events, which have an independent measurement of the energy and X_{\max} , this is especially straightforward [64, 112, 113]. With only SD data this is more cumbersome because the energy and R^{μ} (and also to a lesser degree X_{\max}) are degenerate, but one could

CHAPTER 7. CONCLUSION AND OUTLOOK

exploit the difference in arrival time between muons (early) and EM particles (late) to break the degeneracy. In this work, R^μ is fitted with the energy fixed to the standard SD energy estimator, which will be biased for light or heavy primaries and therefore, so will R^μ . In that sense, R^μ is treated in this thesis as a nuisance parameter and only X_{\max} has meaningful mass composition information.

The addition of SSDs will break the degeneracy between the energy and R^μ , because the response to the electromagnetic part is different in the WCD [64, 88]. Using the SSDs in a similar framework as presented here involves incorporating the information presented in Chapter 3 into the model from Chapter 5 to arrive at a model for the SSD signal. Because the SSD is smaller than the WCD, perhaps not much information can be gained from the timing, but in any case the total signal should be useful. Having an unbiased estimator for R^μ , the energy and X_{\max} will significantly improve the event-by-event identification of the primary mass.

Additional improvements in the template fit should focus on events with a station close to the core. These showers were not used in this thesis, because they have been shown to have a large bias in reconstructing X_{\max} . Including these will enhance the statistics by a factor two. Stations close to the core are likely to be saturated and thus provide less information. This will be much improved with the addition of a small PMT as is being installed as part of the AugerPrime upgrade [88].

Faster timing resolution is also part of the upgrade. This enhances the reconstruction of the shower front, but would not necessarily improve the template fit presented here, because spiky features are smoothed out by the template in any case. Progress on the start time fit can probably be made on the modeling side, because the shower front is known to be more complicated than simply spherical. Additionally, the fluctuations on the start time are dominated by the number of particles collected and not by the timing resolution [108].

It would be interesting to disentangle X_{\max}^μ from X_{\max} in the template fit to address the nature of the offset with $X_{\max}(\text{FD})$. Although this might be difficult with the limited available information from only the WCD time traces.

Independent evidence for the nature and origin of UHECRs can come from *cosmogenic* photons and neutrinos produced when UHECRs interact with the CMB. Current limits set using the Pierre Auger Observatory are already constraining several astrophysics scenarios [119, 120]. The AugerPrime upgrade also enhances the sensitivity to ultra-high-energy photons and neutrinos. In the future, the GRAND (Giant Radio Array for Neutrino Detection) observatory is designed to detect exactly these cosmogenic neutrinos and will also provide unprecedented UHECR statistics [121].

Bibliography

- [1] V. F. Hess, Über Beobachtungen der durchdringenden Strahlung bei sieben Freiballonfahrten, *Phys. Z.* **13**, 1084 (1912) (cit. on p. 3).
- [2] M. Walter and A. W. Wolfendale, Early history of cosmic particle physics, en, *EPJ H* **37**, 323 (2012) (cit. on p. 3).
- [3] C. T. R. Wilson, On an Expansion Apparatus for Making Visible the Tracks of Ionising Particles in Gases and Some Results Obtained by Its Use, *Proc. Roy. Soc. Lond. A* **87**, 277 (1912) (cit. on p. 3).
- [4] C. F. Powell, Applications of the Photographic Method to Problems in Nuclear Physics. I. (a) The Determination of the Energy of Homogeneous Groups of α -Particles and Protons (b) The Determination of the Energy of Fast Neutrons, *Proc. Math. Phys. Eng. Sci.* **181**, 344 (1943) (cit. on p. 3).
- [5] H. Geiger and W. Müller, Elektronenzählrohr zur messung schwächster aktivitäten, *Naturwissenschaften* **16**, 617 (1928) (cit. on p. 3).
- [6] P. Auger, P. Ehrenfest, R. Maze, J. Daudin, and R. A. Fréon, Extensive cosmic-ray showers, *Rev. Mod. Phys.* **11**, 288 (1939) (cit. on p. 3).
- [7] B. Rossi, Supplemento a la Ric. Sci. (1934) (cit. on p. 3).
- [8] B. Rossi, Über die eigenschaften der durchdringenden korpuskularstrahlung im meeresniveau, *Zeitschrift für Physik* **82**, 151 (1933) (cit. on p. 3).
- [9] J. Linsley, Evidence for a primary cosmic-ray particle with energy 10^{20} ev, *Phys. Rev. Lett.* **10**, 146 (1963) (cit. on p. 3).
- [10] J. Linsley and L. Scarsi, Arrival Times of Air Shower Particles at Large Distances from the Axis, *Phys. Rev.* **128**, 2384 (1962) (cit. on pp. 3, 39).
- [11] N. Chiba et al., Akeno giant air shower array (agasa) covering 100 km² area, *Nucl. Instrum. Methods Phys. Res., Sect. A* **311**, 338 (1992) (cit. on p. 3).
- [12] M. A. Lawrence, R. J. O. Reid, and A. A. Watson, The cosmic ray energy spectrum above 4×10^{17} eV as measured by the Haverah Park array, *J. Phys. G: Nucl. Part. Phys.* **17**, 733 (1991) (cit. on p. 3).
- [13] V. Artamonov, B. Afanasyev, A. Glushkov, et al., Status and prospects of the yakutsk complex array of eas, *Bull. Russ. Acad. Sci. Phys. Ser* **58**, 98 (1994) (cit. on p. 3).
- [14] R. M. Baltrusaitis et al., The Utah Fly's Eye detector, *Nucl. Instrum. Meth. A* **240**, 410 (1985) (cit. on p. 3).
- [15] R. U. Abbasi et al. (HiRes Collaboration), Measurement of the flux of ultrahigh energy cosmic rays from monocular observations by the High Resolution Fly's Eye experiment, *Phys. Rev. Lett.* **92**, 151101 (2004) (cit. on p. 3).
- [16] H. Kawai et al. (Telescope Array Collaboration), Telescope array experiment, *Nucl. Phys. B Proc. Suppl.* **175-176**, edited by K. S. Cheng et al., 221 (2008) (cit. on p. 3).
- [17] M. Aguilar et al., Precision Measurement of the Proton Flux in Primary Cosmic Rays from Rigidity 1 GV to 1.8 TV with the Alpha Magnetic Spectrometer on the International Space Station, *Phys. Rev. Lett.* **114**, 171103, 171103 (2015) (cit. on pp. 4, 5).
- [18] O. Adriani et al., Measurements of cosmic-ray proton and helium spectra with the PAMELA calorimeter, *Adv. Space Res.* **51**, 219 (2013) (cit. on pp. 4, 5).
- [19] Y. S. Yoon et al., Proton and Helium Spectra from the CREAM-III Flight, *The Astrophys. J.* **839**, 5, 5 (2017) (cit. on p. 5).

- [20] K. Rawlins and IceCube Collaboration, Latest Results on Cosmic Ray Spectrum and Composition from Three Years of IceTop and IceCube, in 34th International Cosmic Ray Conference (ICRC2015), Vol. 34 (July 2015), p. 334 (cit. on p. 5).
- [21] S. Schoo et al., The energy spectrum of cosmic rays in the range from 10^{14} to 10^{18} eV, in 34th International Cosmic Ray Conference (ICRC2015), Vol. 34 (July 2015), p. 263 (cit. on p. 5).
- [22] V. Grebenyuk et al., Energy spectra of abundant cosmic-ray nuclei in the NUCLEON experiment, *Adv. Space Res.* **64**, 2546 (2019) (cit. on p. 5).
- [23] R. Alfaro et al. (HAWC), All-particle cosmic ray energy spectrum measured by the HAWC experiment from 10 to 500 TeV, *Phys. Rev. D* **96**, 122001 (2017) (cit. on p. 5).
- [24] M. Amenomori et al. (TIBET III), The All-particle spectrum of primary cosmic rays in the wide energy range from 10^{14} eV to 10^{17} eV observed with the Tibet-III air-shower array, *Astrophys. J.* **678**, 1165 (2008) (cit. on p. 5).
- [25] The Pierre Auger Observatory: Contributions to the 34th International Cosmic Ray Conference (ICRC 2015) (Sept. 2015) (cit. on p. 5).
- [26] D. Ivanov, TA Spectrum Summary, in 34th International Cosmic Ray Conference (ICRC2015), Vol. 34 (July 2015), p. 349 (cit. on p. 5).
- [27] V. L. Ginzburg and S. I. Syrovatskii, The Origin of Cosmic Rays (1964) (cit. on p. 4).
- [28] M. Ackermann et al. (Fermi-LAT), Detection of the Characteristic Pion-Decay Signature in Supernova Remnants, *Science* **339**, 807 (2013) (cit. on p. 4).
- [29] E. Alcantara, L. A. Anchordoqui, and J. F. Soriano, Hunting for superheavy dark matter with the highest-energy cosmic rays, *Phys. Rev. D* **99**, 103016 (2019) (cit. on p. 4).
- [30] D. Caprioli, “Espresso” Acceleration of Ultra-high-energy Cosmic Rays, *Astrophys. J. Lett.* **811**, L38 (2015) (cit. on p. 4).
- [31] L. A. Anchordoqui, G. E. Romero, and J. A. Combi, Heavy nuclei at the end of the cosmic ray spectrum?, *Phys. Rev. D* **60**, 103001 (1999) (cit. on p. 4).
- [32] K. Murase, K. Ioka, S. Nagataki, and T. Nakamura, High-energy neutrinos and cosmic rays from low-luminosity gamma-ray bursts?, *The Astrophys. J.* **651**, L5 (2006) (cit. on p. 4).
- [33] K. Fang, K. Kotera, and A. V. Olinto, Newly born pulsars as sources of ultrahigh energy cosmic rays, *750*, 118 (2012) (cit. on p. 4).
- [34] B. T. Zhang, K. Murase, F. Oikonomou, and Z. Li, High-energy cosmic ray nuclei from tidal disruption events: Origin, survival, and implications, *Phys. Rev. D* **96**, [Addendum: *Phys.Rev.D* 96, 069902 (2017)], 063007 (2017) (cit. on p. 4).
- [35] P. Goldreich and W. H. Julian, Pulsar Electrodynamics, *The Astrophys. J.* **157**, 869 (1969) (cit. on p. 4).
- [36] A. Hillas, The Origin of Ultra-High-Energy Cosmic Rays, *Ann. Rev. Astron. Astrophysics* **22**, 425 (1984) (cit. on pp. 4, 6).
- [37] A. Coleman et al., Ultra-High-Energy Cosmic Rays: The Intersection of the Cosmic and Energy Frontiers, (2022) (cit. on pp. 6, 17).
- [38] K. Greisen, End to the cosmic-ray spectrum?, *Phys. Rev. Lett.* **16**, 748 (1966) (cit. on p. 6).
- [39] G. T. Zatsepin and V. A. Kuzmin, Upper limit of the spectrum of cosmic rays, *JETP Lett.* **4**, 78 (1966) (cit. on p. 6).
- [40] V. Berezhinsky, A. Z. Gazizov, and S. I. Grigorieva, On astrophysical solution to ultrahigh-energy cosmic rays, *Phys. Rev. D* **74**, 043005 (2006) (cit. on p. 7).
- [41] W. Heitler, The quantum theory of radiation. English, Int. Ser. Monogr. Phys. (Oxford University Press, Oxford, 1936) (cit. on p. 8).
- [42] J. Matthews, A heitler model of extensive air showers, *Astropart. Phys.* **22**, 387 (2005) (cit. on p. 8).
- [43] T. K. Gaisser and A. M. Hillas, Reliability of the Method of Constant Intensity Cuts for Reconstructing the Average Development of Vertical Showers, in International Cosmic Ray Conference, Vol. 8 (Jan. 1977), p. 353 (cit. on p. 8).

BIBLIOGRAPHY

- [44] K. Kamata and J. Nishimura, The Lateral and the Angular Structure Functions of Electron Showers, *Progress of Theoretical Physics Supplement* **6**, 93 (1958) (cit. on p. 8).
- [45] K. Greisen, Cosmic Ray Showers, *Annu. Rev. Nucl. Sci.* **10**, 63 (1960) (cit. on p. 8).
- [46] M. Nagano and A. A. Watson, Observations and implications of the ultrahigh-energy cosmic rays, *Rev. Mod. Phys.* **72**, 689 (2000) (cit. on p. 9).
- [47] T. Pierog, I. Karpenko, J. M. Katzy, E. Yatsenko, and K. Werner, EPOS LHC: Test of collective hadronization with data measured at the CERN Large Hadron Collider, *Phys. Rev. C* **92**, 034906 (2015) (cit. on p. 9).
- [48] S. Ostapchenko, Monte Carlo treatment of hadronic interactions in enhanced Pomeron scheme: QGSJET-II model, *Phys. Rev. D* **83**, 014018 (2011) (cit. on p. 9).
- [49] F. Riehn, R. Engel, A. Fedynitch, T. K. Gaisser, and T. Stanev, Hadronic interaction model Sibyll 2.3d and extensive air showers, *Phys. Rev. D* **102**, 063002 (2020) (cit. on p. 9).
- [50] Z. Jakubowski and M. Kobel, A verified upgrade of the GHEISHA-6/7 simulation of particle interactions, *Nucl. Instrum. Meth. A* **297**, 60 (1990) (cit. on p. 9).
- [51] T. Böhlen et al., The FLUKA Code: Developments and Challenges for High Energy and Medical Applications, *Nucl. Data Sheets* **120**, 211 (2014) (cit. on p. 9).
- [52] M. Bleicher et al., Relativistic hadron-hadron collisions in the ultra-relativistic quantum molecular dynamics model, *Journal of Physics G Nuclear Physics* **25**, 1859 (1999) (cit. on p. 9).
- [53] D. Heck, J. Knapp, J. N. Capdevielle, G. Schatz, and T. Thouw, CORSIKA: a Monte Carlo code to simulate extensive air showers. (1998) (cit. on p. 9).
- [54] S. J. Sciutto, The AIREX system for air shower simulations. An update. In International Cosmic Ray Conference, Vol. 1 (Jan. 2001), p. 237 (cit. on p. 9).
- [55] S. Argiro et al., The Offline Software Framework of the Pierre Auger Observatory, *Nucl. Instrum. Meth. A* **580**, 1485 (2007) (cit. on p. 9).
- [56] S. Agostinelli et al., Geant4—a simulation toolkit, *Nucl. Instrum. Meth. A* **506**, 250 (2003) (cit. on p. 9).
- [57] B. Rossi and K. Greisen, Cosmic-ray theory, *Rev. Mod. Phys.* **13**, 240 (1941) (cit. on p. 10).
- [58] P. Lipari, Concepts of “age” and “universality” in cosmic ray showers, en, *Phys. Rev. D* **79**, 063001 (2009) (cit. on pp. 10, 51).
- [59] M. Giller, A. Kacperczyk, J. Malinowski, W. Tkaczyk, and G. Wieczorek, Similarity of extensive air showers with respect to the shower age, *J. Phys. G: Nucl. Part. Phys.* **31**, 947 (2005) (cit. on p. 10).
- [60] F. Nerling, J. Bluemer, R. Engel, and M. Risse, Universality of electron distributions in high-energy air showers: Description of Cherenkov light production, *Astropart. Phys.* **24**, 421 (2006) (cit. on p. 10).
- [61] S. Lafebre et al., Universality of electron-positron distributions in extensive air showers, *Astropart. Phys.* **31**, 243 (2009) (cit. on pp. 10, 61).
- [62] F. Schmidt, M. Ave, L. Cazon, and A. S. Chou, A Model-Independent Method of Determining Energy Scale and Muon Number in Cosmic Ray Surface Detectors, *Astropart. Phys.* **29**, 355 (2008) (cit. on p. 10).
- [63] A. Aab et al. (Pierre Auger Collaboration), Muons in Air Showers at the Pierre Auger Observatory: Mean Number in Highly Inclined Events, *Phys. Rev. D* **91**, [Erratum: *Phys. Rev. D* **91**, 059901 (2015)], 032003 (2015) (cit. on pp. 11, 72).
- [64] M. Ave, R. Engel, M. Roth, and A. Schulz, A generalized description of the signal size in extensive air shower detectors and its applications, *Astropart. Phys.* **87**, 23 (2017) (cit. on pp. 12, 92, 93).
- [65] M. Ave, M. Roth, and A. Schulz, A generalized description of the time dependent signals in extensive air shower detectors and its applications, *Astropart. Phys.* **88**, 46 (2017) (cit. on pp. 13, 39, 56).

- [66] R. Abbasi et al. (Telescope Array, Pierre Auger Collaboration), Joint analysis of the energy spectrum of ultra-high-energy cosmic rays as measured at the Pierre Auger Observatory and the Telescope Array, [PoS ICRC2021, 337 \(2021\)](#) (cit. on pp. 13, 14).
- [67] A. Aab et al. (Pierre Auger Collaboration), Measurement of the cosmic-ray energy spectrum above 2.5×10^{18} eV using the Pierre Auger Observatory, [Phys. Rev. D **102**, 062005 \(2020\)](#) (cit. on pp. 13, 20, 82, 90).
- [68] A. Aab et al. (Pierre Auger Collaboration), Observation of a Large-scale Anisotropy in the Arrival Directions of Cosmic Rays above 8×10^{18} eV, [Science **357**, 1266 \(2017\)](#) (cit. on p. 14).
- [69] P. Abreu et al. (Pierre Auger Collaboration), Arrival Directions of Cosmic Rays above 32 EeV from Phase One of the Pierre Auger Observatory, [Astrophys. J. **935**, 170 \(2022\)](#) (cit. on p. 14).
- [70] A. di Matteo et al. (Telescope Array, Pierre Auger Collaboration), UHECR arrival directions in the latest data from the original Auger and TA surface detectors and nearby galaxies, [PoS ICRC2021, 308 \(2021\)](#) (cit. on p. 14).
- [71] A. Yushkov (Pierre Auger Collaboration), Mass Composition of Cosmic Rays with Energies above $10^{17.2}$ eV from the Hybrid Data of the Pierre Auger Observatory, [PoS ICRC2019, 482 \(2020\)](#) (cit. on pp. 15, 85–87).
- [72] R. U. Abbasi et al. (Telescope Array Collaboration), Depth of Ultra High Energy Cosmic Ray Induced Air Shower Maxima Measured by the Telescope Array Black Rock and Long Ridge FADC Fluorescence Detectors and Surface Array in Hybrid Mode, [Astrophys. J. **858**, 76 \(2018\)](#) (cit. on p. 15).
- [73] H. P. Dembinski et al., Report on Tests and Measurements of Hadronic Interaction Properties with Air Showers, in [EPJ Web Conf. Vol. 210 \(Oct. 2019\)](#) (cit. on p. 16).
- [74] H. P. Dembinski et al., Data-driven model of the cosmic-ray flux and mass composition from 10 GeV to 10^{11} GeV, [PoS ICRC2017, 533 \(2018\)](#) (cit. on p. 16).
- [75] A. Aab et al. (Pierre Auger Collaboration), The Pierre Auger Cosmic Ray Observatory, [Nucl. Instrum. Methods Phys. Res., Sect. A **798**, 172 \(2015\)](#) (cit. on p. 17).
- [76] I. Allekotte et al., The surface detector system of the Pierre Auger Observatory, [Nucl. Instrum. Meth. A **586**, 409 \(2008\)](#) (cit. on p. 18).
- [77] X. Bertou et al. (Pierre Auger Collaboration), Calibration of the surface array of the Pierre Auger Observatory, [Nucl. Instrum. Meth. A **568**, 839 \(2006\)](#) (cit. on p. 18).
- [78] J. Abraham et al. (Pierre Auger Collaboration), Trigger and Aperture of the Surface Detector Array of the Pierre Auger Observatory, [Nucl. Instrum. Meth. A **613**, 29 \(2010\)](#) (cit. on pp. 18, 53).
- [79] A. Aab et al. (Pierre Auger Collaboration), Reconstruction of events recorded with the surface detector of the Pierre Auger Observatory, [J. Instrum. **15**, P10021 \(2020\)](#) (cit. on pp. 18–20, 39, 65, 70).
- [80] D. Newton, J. Knapp, and A. A. Watson, The Optimum Distance at which to Determine the Size of a Giant Air Shower, [Astropart. Phys. **26**, 414 \(2007\)](#) (cit. on pp. 19, 50).
- [81] J. Hersil, I. Escobar, D. Scott, G. Clark, and S. Olbert, Observations of extensive air showers near the maximum of their longitudinal development, [Phys. Rev. Lett. **6**, 22 \(1961\)](#) (cit. on p. 20).
- [82] A. Aab et al. (Pierre Auger Collaboration), Depth of Maximum of Air-Shower Profiles at the Pierre Auger Observatory: Measurements at Energies above $10^{17.8}$ eV, [Phys. Rev. D **90**, 122005 \(2014\)](#) (cit. on pp. 20, 70, 72, 79, 81).
- [83] J. Abraham et al., The fluorescence detector of the Pierre Auger Observatory, [Nucl. Instrum. Meth. A **620**, 227 \(2010\)](#) (cit. on p. 20).
- [84] A. Aab et al. (Pierre Auger Collaboration), Data-driven estimation of the invisible energy of cosmic ray showers with the Pierre Auger Observatory, [Phys. Rev. D **100**, 082003 \(2019\)](#) (cit. on p. 21).
- [85] V. Verzi (Pierre Auger Collaboration), The Energy Scale of the Pierre Auger Observatory, in 33rd International Cosmic Ray Conference (2013), p. 0928 (cit. on p. 21).

BIBLIOGRAPHY

- [86] NOAA Air Resources Laboratory, Global Data Assimilation System, <https://www.ncei.noaa.gov/products/weather-climate-models/global-data-assimilation> (cit. on pp. 21, 69).
- [87] P. Abreu et al., Description of atmospheric conditions at the pierre auger observatory using the global data assimilation system (gdas), *Astropart. Phys.* **35**, 591 (2012) (cit. on p. 21).
- [88] A. Aab et al. (Pierre Auger Collaboration), The Pierre Auger Observatory Upgrade - Preliminary Design Report, (2016) (cit. on pp. 21, 22, 24, 93).
- [89] G. Cataldi et al. (Pierre Auger Collaboration), The upgrade of the Pierre Auger Observatory with the Scintillator Surface Detector, *PoS ICRC2021*, 251 (2021) (cit. on p. 22).
- [90] B. Pont (Pierre Auger Collaboration), A Large Radio Detector at the Pierre Auger Observatory - Measuring the Properties of Cosmic Rays up to the Highest Energies, *PoS ICRC2019*, 395 (2021) (cit. on p. 22).
- [91] M. Gottowik (Pierre Auger Collaboration), Measurements of Inclined Air Showers with the Auger Engineering Radio Array at the Pierre Auger Observatory, *PoS ICRC2019*, 274 (2021) (cit. on p. 22).
- [92] D. Schmidt, Sensitivity of AugerPrime to the masses of ultra-high-energy cosmic rays, PhD thesis (Karlsruher Institut für Technologie (KIT) and the Universidad Nacional de San Martín (UNSAM), 2018) (cit. on p. 24).
- [93] B. Kégl and D. Veberič, Single muon response: tracklength, Pierre Auger Collaboration Internal Note - GAP2009-043 (2009) (cit. on pp. 26, 33, 63, 64).
- [94] A. A. Watson and J. G. Wilson, Fluctuation studies of large air showers: the composition of primary cosmic ray particles of energy $E \sim 10^{18}$ eV, en, *J. Phys. A: Math. Nucl. Gen.* **7**, 1199 (1974) (cit. on pp. 39, 49).
- [95] A. Aab et al. (Pierre Auger Collaboration), 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, *Phys. Rev. D* **96**, 122003 (2017) (cit. on pp. 39, 49, 86, 89).
- [96] A. Aab, Air shower signal shape analysis, a new method to separate ultra-high-energy photons from cosmic rays, PhD thesis (Radboud University Nijmegen, 2020) (cit. on pp. 39, 42, 49, 73).
- [97] A. Schulz, Measurement of the energy spectrum and mass composition of ultra-high energy cosmic rays, PhD thesis (Karlsruhe Institute of Technology, Germany, 2016) (cit. on pp. 39, 49, 50, 56, 63).
- [98] D. Maurel, Mass composition of ultra-high energy cosmic rays based on air shower universality, PhD thesis (Karlsruhe Institute of Technology, Germany, 2015) (cit. on pp. 39, 49, 56, 63).
- [99] J. F. de Beer, T. E. Cranshaw, and A. G. Parham, Observations on extensive air showers ix. an experimental investigation into the mu-meson component by means of triggered spark counters, *The Philos. Mag.: A Journal of Theoretical Experimental and Applied Physics* **7**, 499 (1962) (cit. on p. 39).
- [100] L. Cazon, R. A. Vazquez, A. A. Watson, and E. Zas, Time structure of muonic showers, *Astropart. Phys.* **21**, 71 (2004) (cit. on pp. 39, 40, 42, 43).
- [101] L. Cazon, R. A. Vazquez, and E. Zas, Depth development of extensive air showers from muon time distributions, *Astropart. Phys.* **23**, 393 (2005) (cit. on pp. 39, 40, 43, 53).
- [102] L. Cazon, R. Conceicao, M. Pimenta, and E. Santos, A model for the transport of muons in extensive air showers, *Astropart. Phys.* **36**, 211 (2012) (cit. on p. 39).
- [103] A. Aab et al. (Pierre Auger Collaboration), Muons in Air Showers at the Pierre Auger Observatory: Measurement of Atmospheric Production Depth, *Phys. Rev. D* **90**, [Addendum: *Phys.Rev.D* 90, 039904 (2014), Erratum: *Phys.Rev.D* 92, 019903 (2015)], 012012 (2014) (cit. on pp. 39–41, 92).
- [104] L. Cazon, private communication, 2021 (cit. on p. 40).
- [105] T. K. Gaisser and S. Verpoest, Profiles of energetic muons in the atmosphere, *Astropart. Phys.* **133**, 102630 (2021) (cit. on p. 40).

- [106] A. Aab et al. (Pierre Auger Collaboration), Deep-learning based reconstruction of the shower maximum X_{max} using the water-Cherenkov detectors of the Pierre Auger Observatory, *JINST* **16**, P07019 (2021) (cit. on pp. 47, 49, 92).
- [107] G. van Aar, On the nature and origin of ultra-high-energy cosmic rays, PhD thesis (Radboud University Nijmegen, 2016) (cit. on pp. 49, 55, 72).
- [108] G. de Mauro, Tasting the ultra-high-energy universe. weighing in on mass composition measurement of ultra-high-energy cosmic rays, PhD thesis (Radboud University Nijmegen, 2020) (cit. on pp. 49, 55, 56, 65, 66, 72, 93).
- [109] A. Aab et al. (Pierre Auger Collaboration), Azimuthal asymmetry in the risetime of the surface detector signals of the Pierre Auger Observatory, *Phys. Rev. D* **93**, 072006 (2016) (cit. on p. 49).
- [110] R. Colalillo, F. Guarino, and A. Yushkov, Napoli + Praha library: a brief guide to the summary trees, including the correct way to calculate components of signals in SD stations, Pierre Auger Collaboration Internal Note - GAP2019-069 (2019) (cit. on p. 49).
- [111] M. Ave et al., Extensive Air Shower Universality of Ground Particle Distributions, *Astropart. Phys.* **87**, 23 (2017) (cit. on pp. 50, 51, 53).
- [112] A. Bridgeman, Determining the mass composition of ultra-high energy cosmic rays using air shower universality, PhD thesis (Karlsruhe Institute of Technology, Germany, 2018) (cit. on pp. 50, 52, 69, 92).
- [113] M. Stadelmaier, On air-shower universality and the mass composition of ultra-high-energy cosmic rays, PhD thesis (Karlsruhe Institute of Technology, Germany, 2022) (cit. on pp. 50, 92).
- [114] H. Dembinski, P. Billoir, O. Deligny, and T. Hebbeker, A phenomenological model of the muon density profile on the ground of very inclined air showers, *Astropart. Phys.* **34**, 128 (2010) (cit. on p. 53).
- [115] D. E. Hilt and D. W. Seegrist, Ridge, a computer program for calculating ridge regression estimates, Vol. no.236, <https://www.biodiversitylibrary.org/bibliography/68934> (Upper Darby, Pa, Dept. of Agriculture, Forest Service, Northeastern Forest Experiment Station, 1977) (cit. on p. 54).
- [116] F. Pedregosa et al., Scikit-learn: machine learning in Python, *Journal of Machine Learning Research* **12**, 2825 (2011) (cit. on p. 54).
- [117] The Pierre Auger Observatory: Contributions to the 36th International Cosmic Ray Conference (ICRC 2019): Madison, Wisconsin, USA, July 24- August 1, 2019 (Sept. 2019) (cit. on p. 69).
- [118] C. J. Todero Peixoto (Pierre Auger Collaboration), Estimating the Depth of Shower Maximum using the Surface Detectors of the Pierre Auger Observatory, *PoS ICRC2019*, 440 (2020) (cit. on pp. 86, 89).
- [119] A. Aab et al. (Pierre Auger), Probing the origin of ultra-high-energy cosmic rays with neutrinos in the EeV energy range using the Pierre Auger Observatory, *JCAP* **10**, 022 (2019) (cit. on p. 93).
- [120] A. Aab et al. (Pierre Auger), Search for photons with energies above 10^{18} eV using the hybrid detector of the Pierre Auger Observatory, *JCAP* **04**, [Erratum: *JCAP* **09**, E02 (2020)], 009 (2017) (cit. on p. 93).
- [121] J. Álvarez-Muñiz et al. (GRAND), The Giant Radio Array for Neutrino Detection (GRAND): Science and Design, *Sci. China Phys. Mech. Astron.* **63**, 219501 (2020) (cit. on p. 93).

A Additional plots and figures

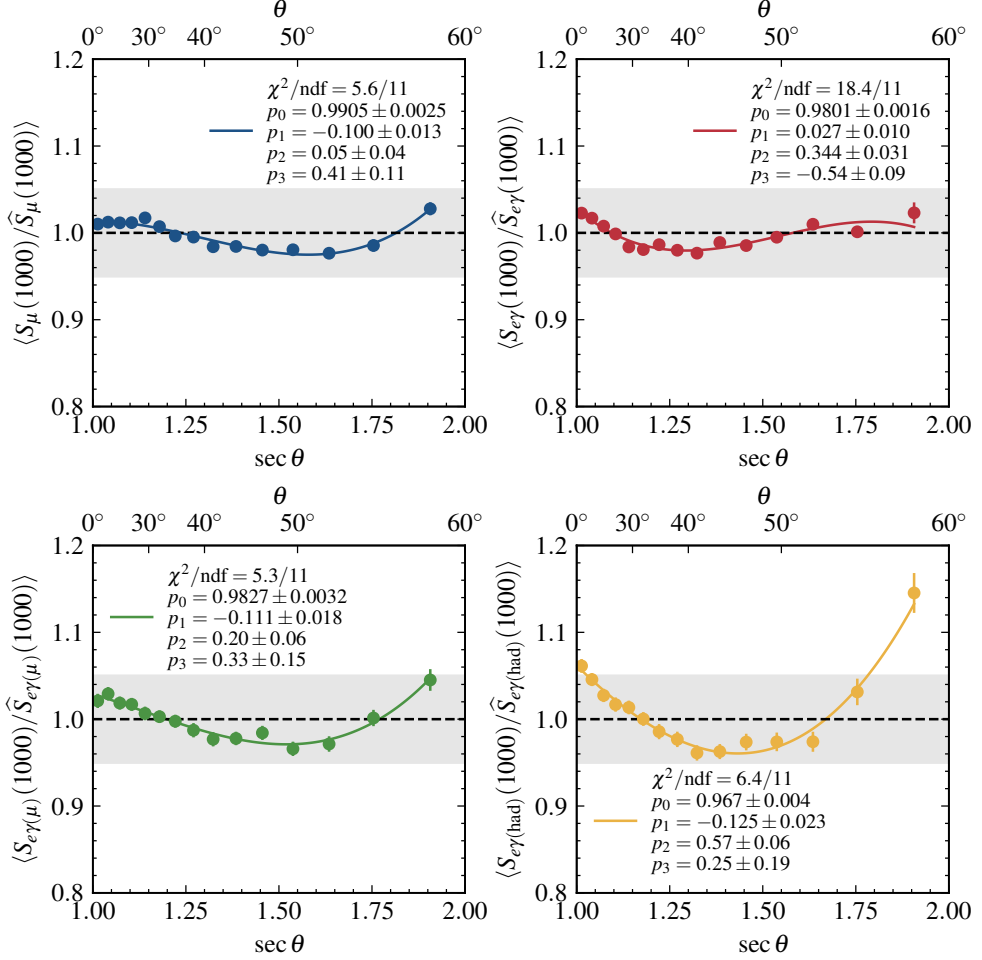


Figure A.1: Mean ratio of the shower size signal for the four universality components divided by the reference model as a function of zenith angle. Only EPOS-LHC proton showers are shown. The line is a fit of a 3rd degree polynomial in $x = \sec \theta - 1.3$ and the best fit values are given in the figure.

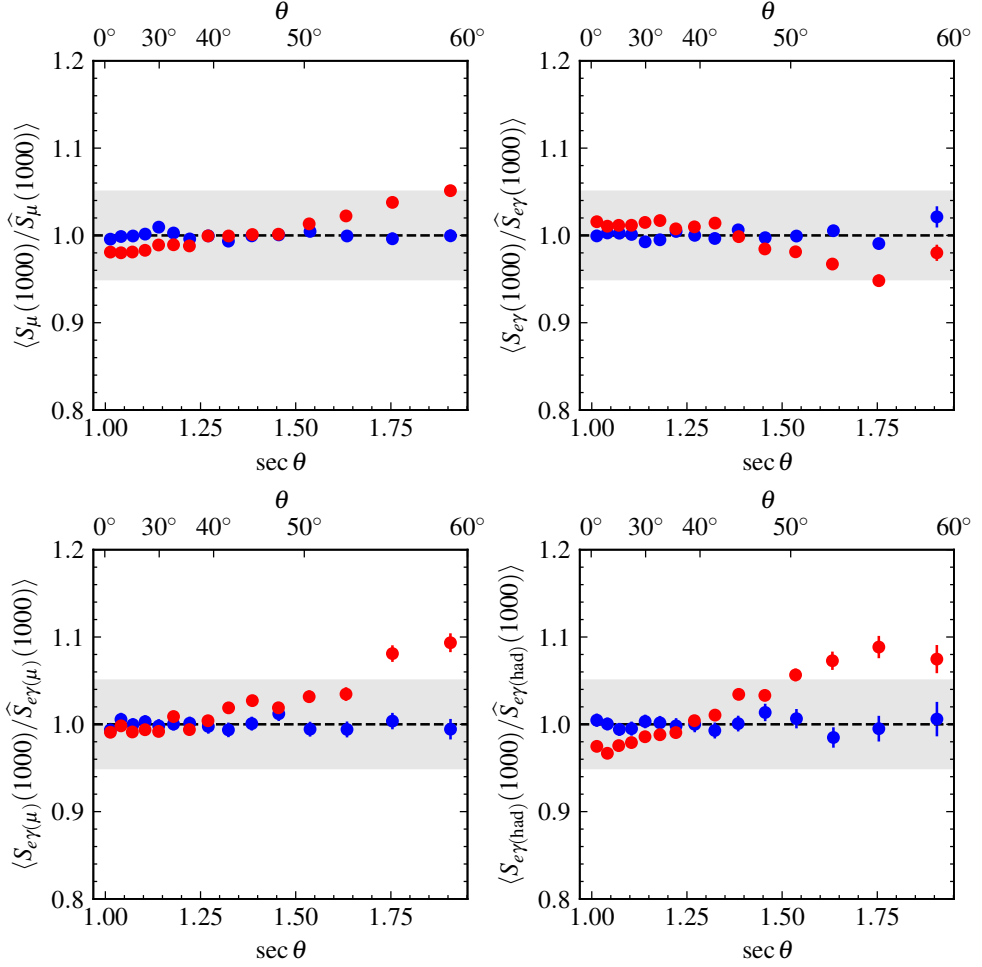


Figure A.2: Mean ratio of the shower size signal for the four universality components divided by the reference model including the θ correction from Fig. A.1. Notice how there are some additional differences between proton (blue) and iron (red) induced showers.

APPENDIX A. ADDITIONAL PLOTS AND FIGURES

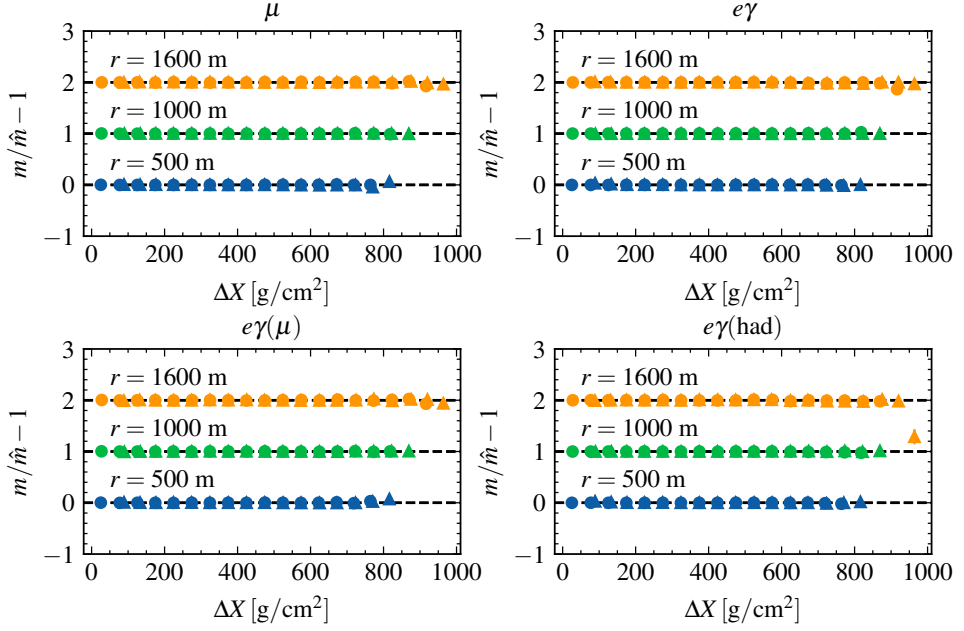


Figure A.3: Deviation of the m parameter for the lognormal shape for proton (round marker) and iron (triangle) with respect to the parametrization (\hat{m}) from [Section 5.3.2](#). The points for $r = 1000(1600)$ m are shifted upwards by 1(2) on the y-axis. Some outliers at high/low ΔX are due to small statistics.

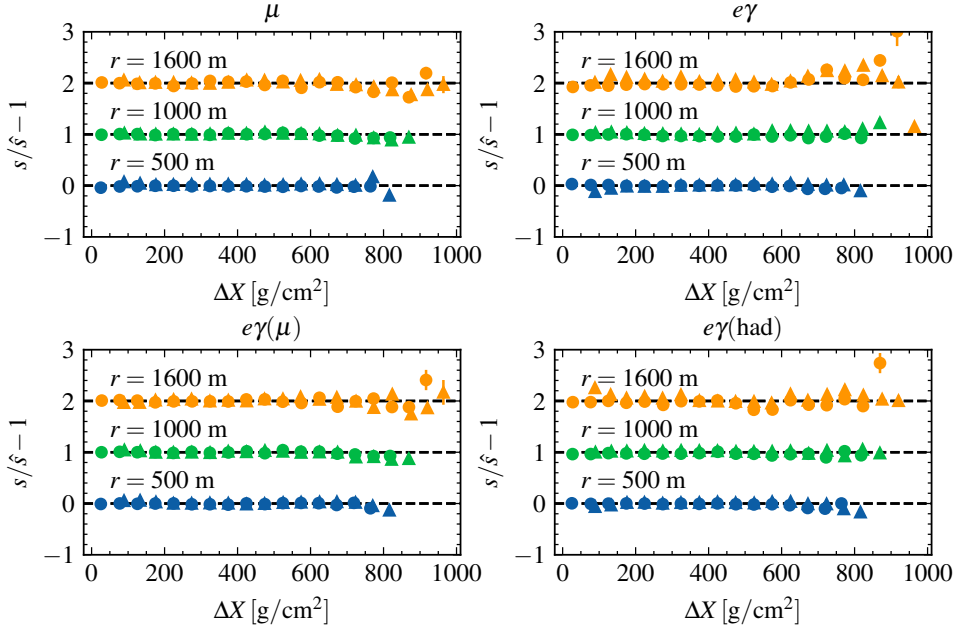


Figure A.4: Deviation of the s parameter similar to [Fig. A.3](#).

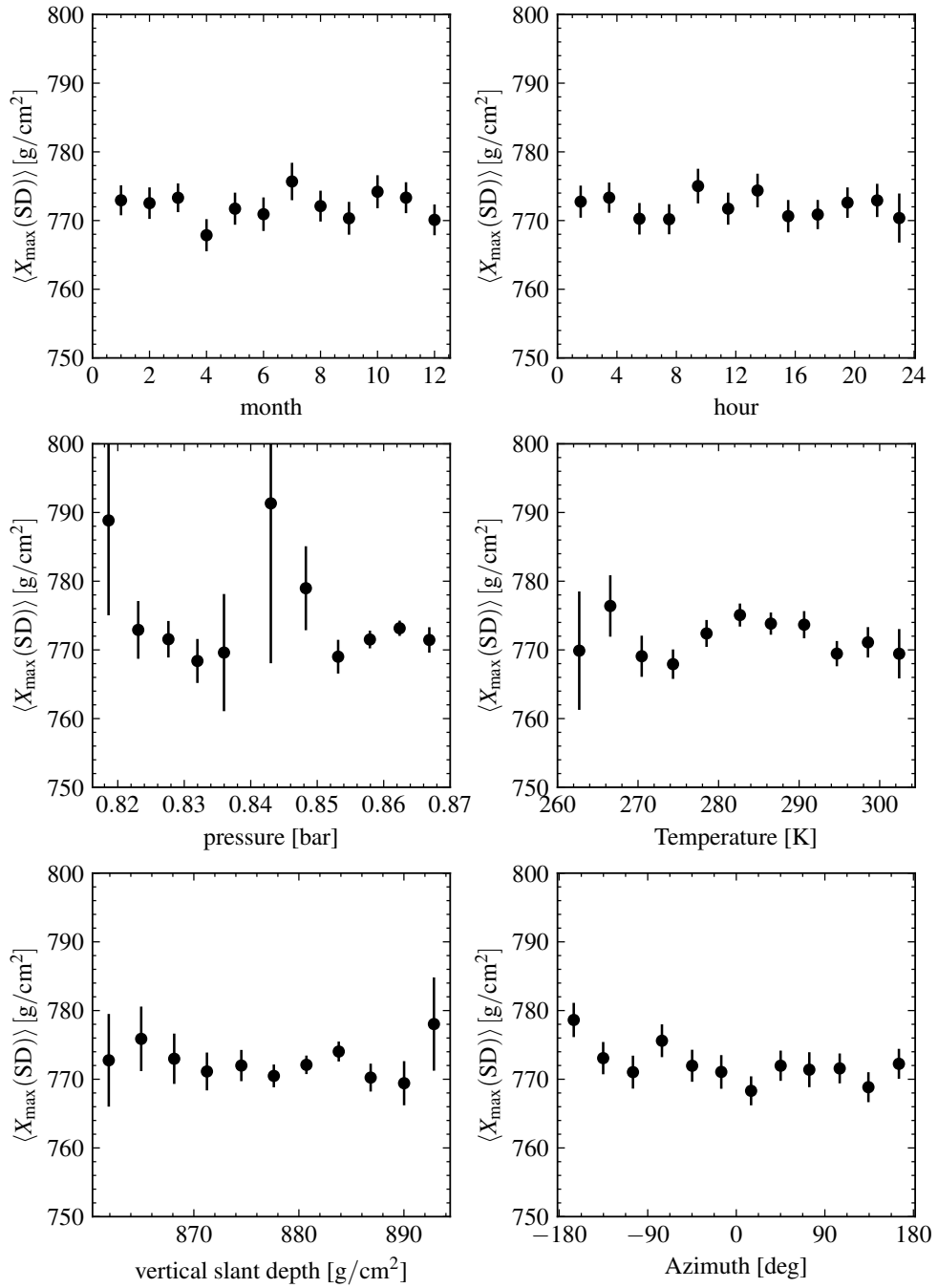


Figure A.5: Mean X_{\max} as a function of a selection of variables that could point to additional systematics.

APPENDIX A. ADDITIONAL PLOTS AND FIGURES

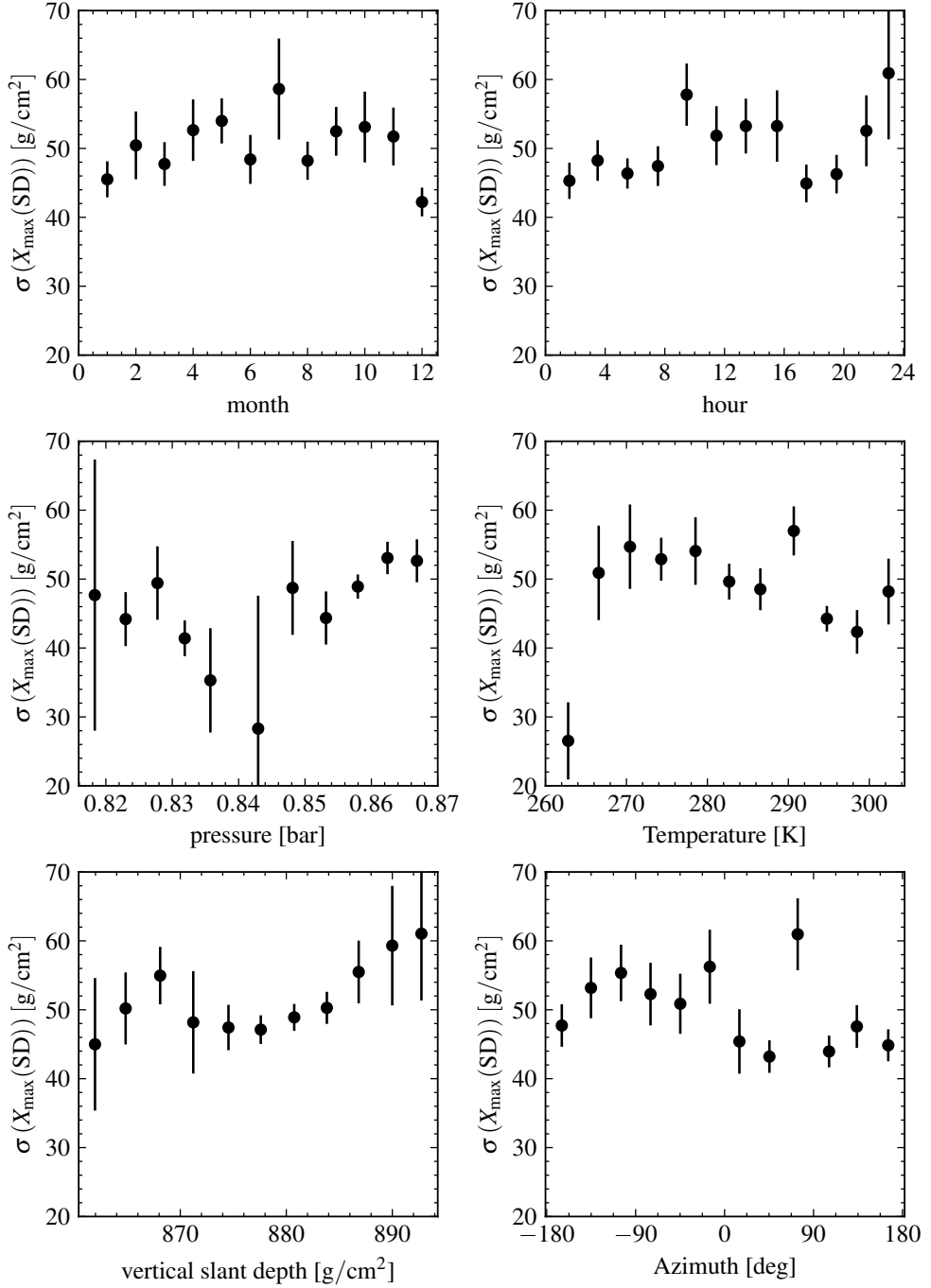


Figure A.6: Fluctuations of X_{\max} as a function of a selection of variables that could point to additional systematics. Only showers with $\sec \theta > 1.25$ are shown.

B Additional tables

Table B.1: Best-fit parameter values of the calibration function Eq. (6.2) for DX_{\max} as a function of DX_{R_c} . Extension of Table 6.2.

	a [g/cm ²]	b [-]	c [g/cm ²]	χ^2/ndf [-]
Hybrid data	460 ± 6	1.060 ± 0.040	-147 ± 23	286/247
UB v3r3p4 EPOS-LHC Fe	522 ± 2	1.238 ± 0.007	-129 ± 3	3057/2945
UB v3r3p4 EPOS-LHC p	524 ± 2	1.314 ± 0.008	-133 ± 3	4197/2788
UB v3r3p4 QGSJET-II.04 Fe	504 ± 2	1.211 ± 0.007	-128 ± 3	3138/2720
UB v3r3p4 QGSJET-II.04 p	499 ± 2	1.273 ± 0.008	-123 ± 3	4333/2768
UB v3r3p4 SIBYL2.3c Fe	498 ± 2	1.214 ± 0.007	-132 ± 3	3557/3106
UB v3r3p4 SIBYL2.3c p	515 ± 2	1.435 ± 0.009	-127 ± 3	6932/2894
UUB trunk EPOS-LHC Fe	549 ± 2	1.199 ± 0.009	-135 ± 3	3039/1872
UUB trunk EPOS-LHC p	556 ± 3	1.251 ± 0.008	-147 ± 3	5556/2188
UUB trunk QGSJET-II.04 Fe	525 ± 3	1.130 ± 0.010	-133 ± 5	1464/763
UUB trunk QGSJET-II.04 p	529 ± 4	1.230 ± 0.010	-125 ± 5	1940/742

Table B.2: Covariance matrix for the best fit values for Eq. (6.4).

	p_0	p_1	p_2	p_3	p_4	p_5
p_0	0.50	-0.02	4.02	0.02	8.58	5.25
p_1	-0.02	5.31	1.00	10.82	2.77	1.28
p_2	4.02	1.00	58.35	1.68	152.59	103.73
p_3	0.02	10.82	1.68	28.01	2.32	-2.28
p_4	8.58	2.77	152.59	2.32	437.65	319.46
p_5	5.25	1.28	103.73	-2.28	319.46	249.18

C Research data management

This thesis research has been carried out under the institute research data management policy of the Institute for Mathematics, Astrophysics and Particle Physics as documented in https://www.ru.nl/publish/pages/868512/imapp_rdm_policy.pdf (accessed on 3 November 2022).

- The data, CORSIKA simulations and software are available within the Pierre Auger Collaboration.
- The simulations used in [Chapters 3](#) and [4](#) can be found on the CNCZ computing clusters at `/vol/augerprime/users/mpothast/` and are available upon request. The detector simulation and reconstruction is done with [Offline](#) r33544.
- The CORSIKA simulations used in [Chapter 5](#) are from the Napoli-Praha library available for members from the Pierre Auger Collaboration on iRODS at `/pauger/Simulations/libraries/rcola/Corsika/EPOS_LHC/`. The detector simulation and reconstruction are done with [Offline](#) git hash `b8dfbd6c9`. The processed simulations are stored on the CNCZ computing clusters at `/vol/augerprime/users/mpothast/` and can be obtained upon request.
- The simulations, including detector and reconstruction, used in [Chapter 6](#) are from the Napoli-Praha library available for members from the Pierre Auger Collaboration on iRODS at `/pauger/Simulations/libraries/MCTask/Offline_v3r99p2a/IdealMC_CORSIKA/Hybrid_NapoliCORSIKA/`. Reconstructed data and simulations with the template fit are on the CNCZ computer cluster at `/vol/augerprime/users/mpothast` and are available upon request.
- All analysis codes can be found at <https://gitlab.iap.kit.edu/auger-observatory/sandboxes/pothast>.

Summary

This thesis is about *ultra-high-energy cosmic rays*: particles from space that come to Earth with the highest energy ever measured. With their high energy they can tell us a lot about physics on the largest scales—the universe—and on the smallest scales—subatomic particle physics. Unfortunately, the astrophysical sources of ultra-high-energy cosmic rays and how they acquire their ultra-high energy are still unknown.

Ultra-high-energy cosmic rays—actually particles, probably atomic nuclei from hydrogen to iron—collide with the atmosphere with so much energy that new particles are created which in turn also create new particles. This process continues and creates a shower of billions of secondary particles that can be measured from the ground. The Pierre Auger Observatory is the largest project to catch these air showers. Specialized fluorescence detectors are used to detect the light created when electrons move through the atmosphere and allow to reconstruct the direction of the primary cosmic particle, its energy and an estimate of the mass—the latter determines the particle type. Another approach uses water tanks to catch the secondary particles. If these secondary high-energy particles pass through water, a flash of light is created that can be recorded. The advantage of the method with water tanks is that particles can be detected all the time while the fluorescence method only works on dark nights (only 10% of the time). This is important, because cosmic rays with the highest energy are very rare.

In this thesis, a method has been developed that improves the identification of the primary cosmic particle type with the water detectors. This method uses the time dependent signal to determine the development of the shower in the atmosphere. In particular, the point in the atmosphere where the number of particles is greatest, X_{\max} , can be determined and has been verified with the measurement of X_{\max} with the fluorescence method. The value of X_{\max} is determined by the mass of the primary particle (higher is lighter). In this thesis the average mass of ultra-high-energy cosmic rays has been determined at higher energies than is possible with only the fluorescence telescopes. The mean value of X_{\max} shows that the mean mass becomes increasingly heavier as a function of energy, while the fluctuations of X_{\max} show that there is also less variation in the particle type as a function of energy.

An accurate measurement of X_{\max} with the water detectors is a step in the right direction for cosmic particle identification of individual measurements. In the near future it will be possible to simultaneously measure the number of muons at the Pierre Auger Observatory using the new scintillator detectors. The combination of the number muons and X_{\max} provides enough information to separate heavy from light particles to improve our understanding of the astrophysical sources that shoot these extreme particles at us.

Samenvatting

Dit proefschrift gaat over *ultra-hoogenergetische kosmische straling*: deeltjes uit de ruimte die met de allerhoogste energie ooit gemeten op de aarde afkomen. Met hun hoge energie kunnen ze ons veel vertellen over het grote—het universum—en het kleine—subatomaire deeltjesfysica. Helaas zijn de astrofysische bronnen van ultra-hoogenergetische kosmische straling en hoe ze hun ultrahoge energie krijgen nog onbekend.

Ultra-hoogenergetische kosmische straling—eigenlijk dus deeltjes, waarschijnlijk atoomkernen van waterstof tot ijzer—botsen op de atmosfeer met zoveel energie dat daarbij nieuwe deeltjes ontstaan, die op hun beurt ook weer nieuwe deeltjes creëren. Dit proces gaat door zodat een stortbui van miljarden secundaire deeltjes ontstaat die gemeten kan worden vanaf de grond. Het Pierre Auger Observatorium is het grootste project dat deze stortbuien van deeltjes vangt. Met gespecialiseerde fluorescentie telescopen kan het licht gedetecteerd worden dat ontstaat als elektronen (gecreëerd in de stortbui) door de atmosfeer bewegen. Daardoor kan worden teruggerekend waar het eerste deeltje vandaan komt, met welke energie en hoe zwaar het ongeveer was—dat laatste bepaalt het type deeltje. Een andere manier die gebruikt wordt, is de secundaire deeltjes uit de stortbui vangen met waterbakken. Als deze deeltjes met hoge energie door water heen gaan ontstaat er een lichtflits die gemeten kan worden. Het voordeel van de waterbakken is dat met deze techniek altijd deeltjes gedetecteerd kunnen worden, terwijl de fluorescentie methode alleen op donkere nachten werkt (maar 10% van de tijd). Dit is belangrijk, want kosmische straling met de allerhoogste energie is erg zeldzaam.

In dit proefschrift is een methode ontwikkeld waarmee met de water detectoren het primaire kosmische deeltje beter geïdentificeerd kan worden. Deze methode gebruikt het tijdsafhankelijke signaal om de ontwikkeling van de stortbui in de atmosfeer te bepalen. Met name het punt in de atmosfeer waar het aantal deeltjes het grootst is, X_{\max} , kan bepaald worden en is geverifieerd met de meting van X_{\max} met de fluorescentie methode. De waarde van X_{\max} is een maat voor de massa van het primaire deeltje (hoger is lichter). In dit proefschrift is de gemiddelde massa van ultra-hoogenergetische straling bepaald bij hogere energieën dan mogelijk was met alleen de fluorescentie telescopen. De gemiddelde waarde van X_{\max} laat zien dat de massa gemiddeld steeds zwaarder wordt als functie van de energie, terwijl de fluctuaties van X_{\max} laten zien dat er ook steeds minder variatie in het type deeltje is bij hogere energieën.

Het accuraat meten van X_{\max} met de water detectoren is een goede stap in de richting van het identificeren van het type kosmische deeltje voor individuele metingen. In de nabije toekomst zal het ook mogelijk zijn om het aantal muonen te meten door de extra scintillator detectoren van het Pierre Auger Observatorium. De combinatie van het aantal muonen en X_{\max} levert genoeg informatie op om zware van lichte deeltjes te scheiden en zo een beter begrip te krijgen van de astrofysische bronnen die deze extreme deeltjes op ons afschieten.

Acknowledgements

A lot of people helped me in the process of making this thesis, even if it was simply being good company at conference dinners, having bad (or surprisingly good) pizza together, drinking strange beers, mountainbiking with or without functioning brakes or just listening to my regular complaints. Thank you all (you know who you are). But let me name a few:

Special thanks to my supervisors, Charles and Sijbrand. Charles, for our shared sense of (sometimes dark) humor and for your keen eye for reviewing the many figures I've showed you over the years. Sijbrand, thank you for our inspirational, educational and motivational talks, especially at times when things sometimes looked too dark.

Thanks to everyone in the Pierre Auger Collaboration for listening to my, always ongoing, 'work in progress' and providing valuable comments. I want to acknowledge Max Stadelmaier for working out the model of the arrival times of muons together, having interesting discussions on universality and for providing some much-needed motivation—all of this via Zoom during the dark days of lockdown.

A large part of this PhD research was done during Covid times. This has made my time at the departement of High Energy Physics less than it could have been. Still, I would like to thank everyone who joined for an early lunch (12 o'clock sharp) to complain about the weather or the daily soup. Many thanks to Cristina for always asking me how I was doing two times: the first time to get a socially-accepted answer, but the second time to get a real answer and start a genuine conversation.

Of course, there were Zoom meetings to connect with people. Thanks to everyone in the NL-Augger and HEP-CR group to hear about my surface detector troubles. Special thanks to Harm for reviving the HEP-CR group and bringing new energy (and people!) to our group.

Our PhD office changed quite a bit since I started. I was very kindly welcomed into (what I thought was) the 'Italian Office' by Giuseppe and Fabrizia. Thank you both for putting up with my naive enthusiasm while you were struggling to finish your thesis. I fondly remember some good (Italian) drinks and dinners, also with everyone from the other offices. Later in *the* office, some people faded in and out, but Daniil deserves mentioning for helping me cut away the (dead) office jungle (sorry Harm, it had to be done). Then Mohit came and started bringing plants into the office again! But seriously, thank you for living up the place. In the end the office was packed to the max again with Anthony and Mohamed: thanks for allowing me enough room on the whiteboard.

Away from the office; Daan, Kunst, Pascal and Tim deserve to be mentioned. Not only for coining the phrase 'mijn werkstukje', but for being friends for half our lives. That also applies to Mara and thank you for supporting my physics studies since the beginning with large bottles of beer.

My parents, Bert and Marja, thank you for everything and specifically for giving me a place to stay when needed. To my sisters, Eef and Sara, thank you for having someone to share with in our happy family. Finally, thank you Kim, and remember: there are always expectations, but the next bike ride is always better than the last.

Curriculum vitae

Mart Pothast was born on 4th of February 1991 in Rheden. He spend most of his youth in Huizen, where he received his high school diploma at SG Huizermaat in 2009. After initially pursuing a degree in medicine, he switched to physics & astronomy at the University of Amsterdam in 2013. He received his master's degree in 2018 following the GRAPPA track.

Mart's master thesis, titled 'Gamma Rays as Probes of Cosmic-Ray Diffusion Throughout the Galaxy' and supervised by Dr. Weniger, was awarded the Pieter Zeemanprijs for best physics master's thesis from the University of Amsterdam and Vrije Universiteit Amsterdam.

In December 2018, Mart began his doctoral research at the department of High Energy Physics at Radboud University Nijmegen under the supervision of prof. dr. de Jong and prof. dr. Timmermans. The findings from his research are presented in this thesis. Additionally, during this time, Mart also taught tutorial classes for physics courses at the bachelor's and master's levels.

