



# Cosmic-Ray Composition Measurements Using Radio Signals

Fabrizia Canfora

# **COSMIC-RAY COMPOSITION MEASUREMENTS USING RADIO SIGNALS**

## **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 van decanen  
in het openbaar te verdedigen op  
maandag 10 mei 2021  
om 16.30 uur precies

door

**FABRIZIA CANFORA**

geboren op 25 oktober 1988  
te Napels, Italië

PROMOTOREN    prof. dr. S.J. de Jong

prof. dr. S.C.M. Bentvelsen  
*Universiteit van Amsterdam*

COPROMOTOR    dr. C.W.J.P Timmermans

MANUSCRIPTCOMMISSIE    prof. dr. J.R. Hörandel    (voorzitter)

dr. A. Castellina  
*Osservatorio Astrofisico di Torino (INAF), Italië*

prof. dr. T. Huege  
*Vrije Universiteit Brussel, België*

prof. dr. T. Peitzmann  
*Universiteit Utrecht*

dr. H. Schoorlemmer  
*Max-Planck-Institut für Kernphysik, Duitsland*

© Fabrizia Canfora 2021

**Cosmic-Ray Composition Measurements Using Radio Signals**

Thesis, Radboud University Nijmegen

viii + 141; illustrated, with bibliographic references and summary in Dutch

ISBN 978 94 6421 277 8

Cover photograph by Ognjen Dedagic

Printed by Ipskamp Printing



This work is part of the research programme of the Foundation for Fundamental Research on Matter (FOM), which is part of the Dutch Research Council (NWO).

# Contents

<b>Introduction</b>	<b>vii</b>
<b>1 Cosmic rays and Extensive air showers</b>	<b>1</b>
1.1 Ultra-high-energy cosmic rays	1
1.1.1 Energy spectrum	2
1.1.2 Origin, acceleration mechanisms and propagation	3
1.2 Extensive air showers	4
1.2.1 Longitudinal shower profile	5
1.2.2 Mass Composition	7
1.3 Radio emission from extensive air showers	8
1.3.1 Emission mechanisms and general features	8
1.3.2 Reconstruction of cosmic ray properties from the radio signal	9
1.3.3 CoREAS simulations	11
<b>2 The Pierre Auger Observatory</b>	<b>13</b>
2.1 The Fluorescence Detector	13
2.2 The Surface Detector	16
2.3 The Auger Engineering Radio Array	17
2.3.1 Antenna design	19
2.3.2 Station electronics, trigger and data acquisition	20
2.4 AugerPrime - the upgrade of the Pierre Auger Observatory	21
2.4.1 A large Radio Detector at the Pierre Auger Observatory	22
<b>3 Electric field reconstruction from the radio detector data</b>	<b>25</b>
3.1 The electric field vector	25
3.1.1 The total electric field and its uncertainty	29
3.1.2 The components of the electric field in the shower plane	29
3.2 The radio signal from a noisy measurement	32
3.2.1 Adding noise to the measurement	35
3.2.2 Toy Monte Carlo	37
3.2.3 Validation of the background subtraction	38
3.3 Energy fluence estimation in presence of noise	40
3.4 The <i>Offline</i> framework	42
3.4.1 Radio event reconstruction in <i>Offline</i>	44
3.4.2 Signal reconstruction module	45
<b>4 Calibration of the AERA stations</b>	<b>47</b>
4.1 Relative amplitude calibration	47
4.1.1 Implications for data analysis	51



4.2	Absolute amplitude calibration	52
4.2.1	Amplitude of the measured galactic emission	53
4.2.2	Model of the radio frequency sky	54
4.2.3	Results	56
4.3	Validation of the calibration	57
<b>5</b>	<b>Methods to measure cosmic-ray composition from the radio energy distribution and spectral information</b>	<b>61</b>
5.1	Simulation data set	61
5.1.1	Simulation reconstruction in <i>Offline</i>	63
5.2	Depth of the shower maximum from the radio energy distribution	63
5.2.1	Signal distribution of the geomagnetic emission	64
5.2.2	Signal distribution of the charge-excess emission	70
5.2.3	Two-dimensional signal distribution	71
5.2.4	Adding background and detector simulations	72
5.2.5	Conclusions	76
5.3	Depth of the shower maximum from the slope of the frequency spectrum	77
5.3.1	Parametrization of the spectral index	78
5.3.2	Validation on the star-pattern simulation data set	80
5.3.3	Adding background and detector simulations	83
5.3.4	Conclusions	86
5.4	Comparison and combination of the two radio $X_{\max}$ estimators	86
5.4.1	Adding background and detector simulations	88
5.4.2	Conclusions	88
<b>6</b>	<b>Mass composition measurement with AERA</b>	<b>91</b>
6.1	Data reconstruction and event selection	91
6.1.1	Station rejection	92
6.1.2	Event selection	93
6.2	Event by event reconstruction of the shower maximum	94
6.2.1	$X_{\max}$ obtained from the AERA energy distribution	94
6.2.2	$X_{\max}$ obtained from the spectral index	99
6.3	Comparison and combination of the radio shower maximum estimators	107
<b>7</b>	<b>Discussion and outlook</b>	<b>111</b>
<b>A</b>	<b>Offline Module Sequence and selection cuts</b>	<b>115</b>
A.1	RdReconstructStarshapedStationPattern	115
A.2	RdObserver	116
A.3	Quality cuts for the RD-SD(-FD) data set	117
<b>B</b>	<b>Research data management</b>	<b>119</b>
	<b>Bibliography</b>	<b>121</b>
	<b>Summary</b>	<b>129</b>
	<b>Samenvatting</b>	<b>133</b>

<b>Curriculum Vitae</b>	<b>139</b>
<b>Acknowledgements</b>	<b>141</b>



# Introduction

The discovery of cosmic rays was certainly a milestone in science. At the beginning of the 1900s scientists were intrigued by an intense ionizing radiation observed on Earth. It took several decades and different measurement strategies before it was concluded that the radiation resulted from cascades of ionising particles that are caused by mostly electrically charged high-energy particles of extra-terrestrial origin. The obvious next step was to investigate the nature of such particles, which, after more than 100 years, still remains one of the most fascinating mysteries in modern physics.

Cosmic rays are charged energetic particles generated in the outer space that strike the Earth from all directions. Their energy spectrum covers about 11 orders of magnitude in energy with great regularity. The study of the most energetic cosmic rays (*ultra-high-energy cosmic rays*) explores an energy range not yet accessible with particle colliders, providing precious information on high-energy hadronic interactions and on the most energetic phenomena in the universe. To identify the astronomical objects that can boost particles to such extreme energies, it is necessary to know the mass composition of cosmic rays. In fact, the mass, and therefore the charge, distribution will provide stringent constraints on the cosmic-ray acceleration and propagation through the galactic and extragalactic magnetic fields, as well as on the theoretical models describing the energy spectrum.

A problem with ultra-high-energy cosmic rays is that they are extremely rare: at the highest energies only one particle per square kilometer per century reaches the Earth, and therefore they can only be observed by large area detectors using the atmosphere as a calorimeter. In fact, the collision of such energetic particles with atmospheric molecules initiates a huge cascade of secondary particles, the so-called *extensive air shower*.

At present, the largest experiment dedicated to the study of ultra-high-energy cosmic rays is the Pierre Auger Observatory in Argentina. The observatory takes advantage of hybrid detection and reconstruction that combines different and complementary techniques. Among those is the detection of radio pulses induced by air showers propagating in the atmosphere with the Auger Engineering Radio Array (AERA). AERA has been designed to explore the feasibility of the radio-detection technique on large instrumented areas. By measuring air showers in coincidence with the surface and fluorescence detectors of the Pierre Auger Observatory it offers unique and optimal conditions for cross calibrations, technical developments and pathfinder studies for the next generation of radio detectors.

AERA significantly contributed to the understanding of the physics governing the radio emission in extensive air showers. Several analysis strategies have been investigated by the Auger collaboration to accurately derive the energy, the arrival direction and the mass composition of the primary cosmic ray particles using the radio signals. This work continues the effort in this direction, investigating the potential of a radio mass composition estimator that combines lateral energy distribution and spectral information from cosmic-ray induced radio frequency signals.



The thesis is organized as follows. In Chapter 1 an introduction on ultra-high-energy cosmic rays and radio emission from extensive air showers is given. The Pierre Auger Observatory is described in Chapter 2, focusing in particular on the AERA array. Chapter 3 is dedicated to the reconstruction of the radio signal from the electric field induced by cosmic rays, as measured by the radio antenna stations. A precise interpretation of the data collected by the radio stations requires a detector calibration, which is described in Chapter 4. In Chapter 5, two radio-based methods to measure the cosmic-ray mass composition are discussed. The first one exploits the correlation between the shape of the radio signal footprint on the ground and the distance to the emission region, while the second infers composition information from the frequency content of the radio signals. Both methods give an estimate for the depth of the maximum number of particle in the air shower, which is a measure for the mass composition. They have been investigated first using Monte Carlo simulations and then, in Chapter 6, applied to AERA data, resulting in a measurement of the average depth of shower maximum as a function of energy. The results are compared to measurements of the fluorescence detector. Finally, the results of this thesis are discussed in Chapter 7.

# 1 Cosmic rays and Extensive air showers

Cosmic rays are charged energetic particles originating in outer space and striking the Earth from all directions. Their discovery is attributed to Victor Hess when, in 1912, he demonstrated that the ionization of air was generated by a source entering the atmosphere from above. Hess was awarded the Nobel prize in 1936, his discovery opened a new window to observe the universe.

Later on it was found that the radiation measured on Earth originated from a cascade of secondary particles created in the collision of the primary cosmic ray with atoms in the upper atmosphere. In 1939, Pierre Auger measured secondary particles in coincidence between detectors positioned 300 m apart and confirmed the existence of extensive cosmic-ray air showers [1].

In 1962, Askar'yan made a prediction that creates a new opportunity for studying air showers: he postulated coherent radiation in the MHz range, induced by the electrons and positrons of the electromagnetic component of air showers [2]. Soon after Jelley measured the radio emission from air showers for the first time [3].

This chapter provides an introduction to the physics of cosmic rays, extensive air showers and radio-emission mechanisms.

## 1.1 Ultra-high-energy cosmic rays

The energy spectrum of cosmic rays covers more than 11 orders of magnitude between  $10^9$  eV  $\lesssim E \lesssim 10^{20}$  eV. With increasing energy the flux drops rapidly: ranging from  $10^4$  particles per square meter per second at  $10^9$  eV to less than one cosmic ray per square kilometer per century for energies larger than  $10^{20}$  eV.

Up to energies of  $\sim 10^{14}$  eV direct measurements of cosmic rays are possible with detectors on balloon flights and satellite-based experiments outside the Earth's atmosphere. This allows a detailed study of the energy and the nature of the cosmic particles. However, due to the constraints in weight and size of the detectors, direct measurements cannot be efficiently performed at higher energies.

When a cosmic ray above  $\sim 10^{15}$  eV enters in the Earth's atmosphere it interacts with air molecules and produces a shower of secondary particles detectable by large ground-based observatories. The extremely rare particles with energies above  $10^{18}$  eV are labeled as *ultra-high-energy cosmic rays* (UHECR).

Up to now, the origin and the acceleration mechanisms of ultra-high-energy cosmic rays are unknown. To be able to answer these fundamental questions it is crucial to study their energy, composition and arrival direction.

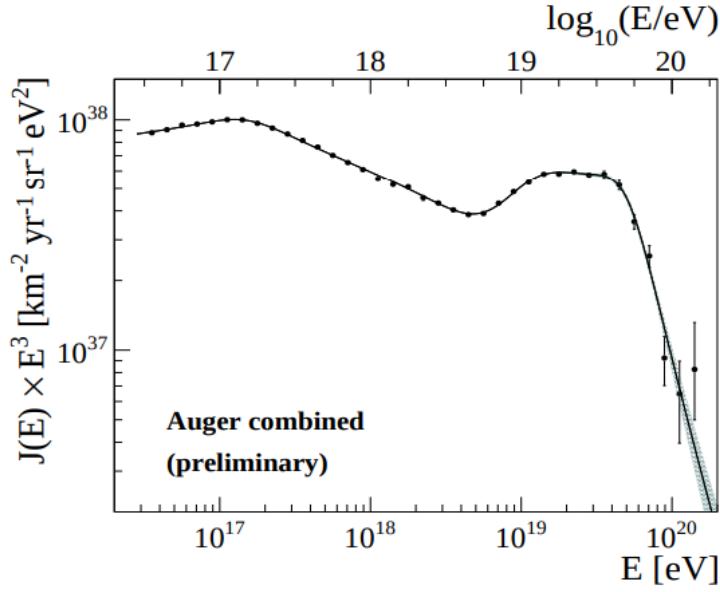


Figure 1.1: UHECR energy spectrum as measured by the Pierre Auger Observatory. The differential flux has been multiplied by  $E^3$  to enhance visibility of the features of the spectrum. The line shows a fit to the spectrum with Eq. 1.1. Figure from [5].

### 1.1.1 Energy spectrum

The energy spectrum of ultra-high-energy cosmic rays measured by the Pierre Auger Observatory is shown in Figure 1.1. The data points are obtained by combining five independent and complementary data sets. The spectrum exhibits three main features: two inflection points corresponding to the *second knee* around  $10^{17}$  eV and to the *ankle* near  $5 \times 10^{18}$  eV, and a sharp suppression for energies higher than  $\sim 5 \times 10^{19}$  eV.

A functional form made by a sequence of different power laws [4] is used to fit it:

$$J_{01234} \propto E^{-\gamma_0} \frac{1 + (E/E_{01})^{\gamma_0}}{1 + (E/E_{01})^{\gamma_1}} \frac{1 + (E/E_{12})^{\gamma_1}}{1 + (E/E_{12})^{\gamma_2}} \frac{1 + (E/E_{23})^{\gamma_2}}{1 + (E/E_{23})^{\gamma_3}} \frac{1 + (E/E_{34})^{\gamma_3}}{1 + (E/E_{34})^{\gamma_4}}. \quad (1.1)$$

The second knee is assumed to be caused by a decrease in the flux of primary irons [6, 7]. The ankle has traditionally been attributed to the transition from galactic to extragalactic origins of the cosmic rays [8]. However, it could also result from a modification of the cosmic proton flux due to a dip in the electron-positron pair production cross section [9], or from photo-disintegration of ultra-high-energy nuclei in the region surrounding cosmic-ray sources [10].

The reason for the ultra-high-energy cut-off is also still under debate: it is compatible with the predicted Greisen–Zatsepin–Kuzmin (GZK) effect [11, 12], but could also be explained, at least in part, as a limit of cosmic-ray accelerators. To distinguish the two scenarios it is necessary to study the mass composition described in Section 1.2.2.

### 1.1.2 Origin, acceleration mechanisms and propagation

Cosmic-ray sources must be the most powerful astrophysical accelerators in the universe, but they are still unknown. On their way to Earth, cosmic rays are deflected by the galactic and extragalactic magnetic fields, making it difficult to infer the location of their origin from their direction at Earth.

Various theories describe possible processes by which cosmic rays acquire their extremely high energies. One of them was proposed by Enrico Fermi in 1949: charged particles stochastically gain energy by multiple collisions in strong turbulent magnetic fields that occur in shock waves of the interstellar plasma, such as the expanding shells of supernova explosions or in jets emitted by active galactic nuclei [13]. The acceleration is acquired in multiple steps: each time the particle crosses a shock wave it gains a fraction of energy. This process is capable of reaching very high energies when the particles are trapped in the region of the shock waves. Charged particles can be trapped in a magnetic field in which they follow a circular path with a radius that increases with energy. The strength and the spatial extent of the magnetic field determine the energy reach.

It is possible to estimate an upper limit of the energy to which a source can accelerate particles by assuming that the Larmor radius  $R_L$  of a particle needs to be contained within the acceleration region of the source with magnetic field  $B$  and plasma velocity  $\beta c$  [14]:

$$\left( \frac{E}{10^{18} \text{eV}} \right) < \frac{Z\beta}{2} \left( \frac{R_L}{\text{kpc}} \right) \left( \frac{B}{\text{G}} \right), \quad (1.2)$$

where  $Z$  and  $E$  are respectively the charge number and the energy of the particle. Potential astronomical sources of ultra-high-energy cosmic rays are shown in the Hillas plot where the magnetic field is plotted versus the size of the source candidate (Figure 1.2). Large sources or strong magnetic fields are required to accelerate particles to the extreme energies at which they are observed [15].

While propagating from their sources to the observer, UHECRs experience two types of processes: interactions with cosmic backgrounds that affect their energy and their composition, but only slightly influence their direction, and interactions with cosmic magnetic fields that affect their direction and travel time, but not their energy and composition. The interaction of ultra-high-energy cosmic ray protons with the photons of the cosmic microwave background (CMB) results in the production of pions via the Delta resonance:

$$p + \gamma_{\text{CMB}} \rightarrow \Delta^+ \rightarrow \begin{aligned} & p + \pi^0 \\ & \rightarrow n + \pi^+ . \end{aligned} \quad (1.3)$$

The pions and the neutrons mostly decay generating secondary neutrinos and photons:

$$\begin{aligned} \pi^0 & \rightarrow \gamma + \gamma \\ \pi^+ & \rightarrow \mu^+ + \nu_\mu \\ n & \rightarrow p + e^- + \bar{\nu}_e . \end{aligned} \quad (1.4)$$

The pion decay is more or less immediate on cosmic distance scales, but the highest energy neutrons can travel a distance of the size of a galaxy before decaying.



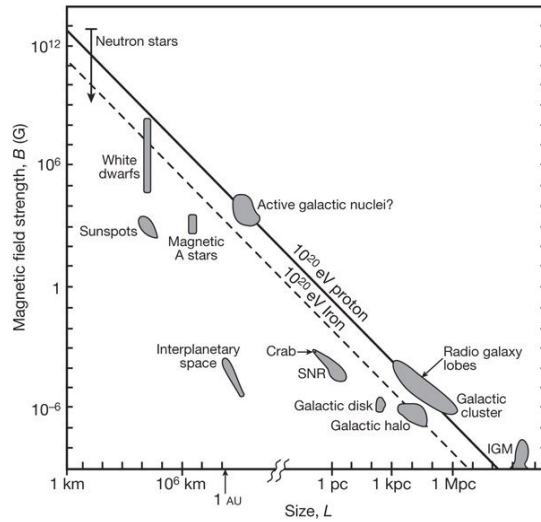


Figure 1.2: Hillas plot with several astrophysical objects that meet the requirements of size/magnetic field needed for the acceleration process. The two lines correspond to protons (solid) and iron nuclei (dashed) with  $E = 10^{20}$  eV, assuming  $\beta = 1$ . Figure from [16].

The photo-pion production processes cause a significant energy loss for primary protons with energies above  $\sim 6 \times 10^{19}$  eV and introduce a suppression in the cosmic ray proton flux beyond this energy (GZK-limit). For cosmic ray nuclei the dominant loss process at the same energies is photodisintegration with photons of the CMB and of the infrared background radiation. Moreover, both protons and heavy nuclei can further loose energy by pair production processes [17]. The energy loss introduces constraints on the potential sources of ultra-high-energy cosmic rays: due to their attenuation, the highest energy cosmic rays observed are more likely to have originated not very far away on intergalactic distance scales.

The extragalactic magnetic field deflects charged particles from their original trajectories. The magnitude of the deflection increases with the charge of the particle, meaning that iron nuclei are much more affected by magnetic fields than protons. Although the extragalactic magnetic field is not well known, it is generally believed that cosmic protons of the highest energies might therefore point back towards their sources within a few degrees. To be able to effectively differentiate between all the different candidate sources it is essential to know the composition of the cosmic rays.

## 1.2 Extensive air showers

Cosmic rays with an energy above  $\sim 10^{15}$  eV impinging on the Earth's atmosphere generate extensive air showers. The primary particle arriving at the outer edge of Earth's atmosphere interacts with nuclei of atmospheric molecules, thereby producing secondary particles which in turn interact or decay. The result is a huge particle cascade with billions of secondary particles traversing the atmosphere (see Figure 1.3(a) for an illustration).

The first interaction occurs typically between 15 and 35 km altitude and creates mainly pions, kaons and baryons forming the nucleus of the air shower [18]. The neutral pions decay almost immediately ( $\tau_{\pi^0} \simeq 8.52 \cdot 10^{-17}$  s) in two photons that start a cascade of electrons, positrons

and photons creating the *electromagnetic component* of the shower. The high-energy charged pions and kaons often interact again before they can decay. At lower energies they more often decay in muons and neutrinos, thereby creating the *muonic component* of air showers. Muons may also be created directly in the high energy interactions. The muons, having long life times and small cross-sections, will travel to the ground almost unattenuated [18]. The evolution of the particle numbers of the different shower components is shown in Figure 1.3(b).

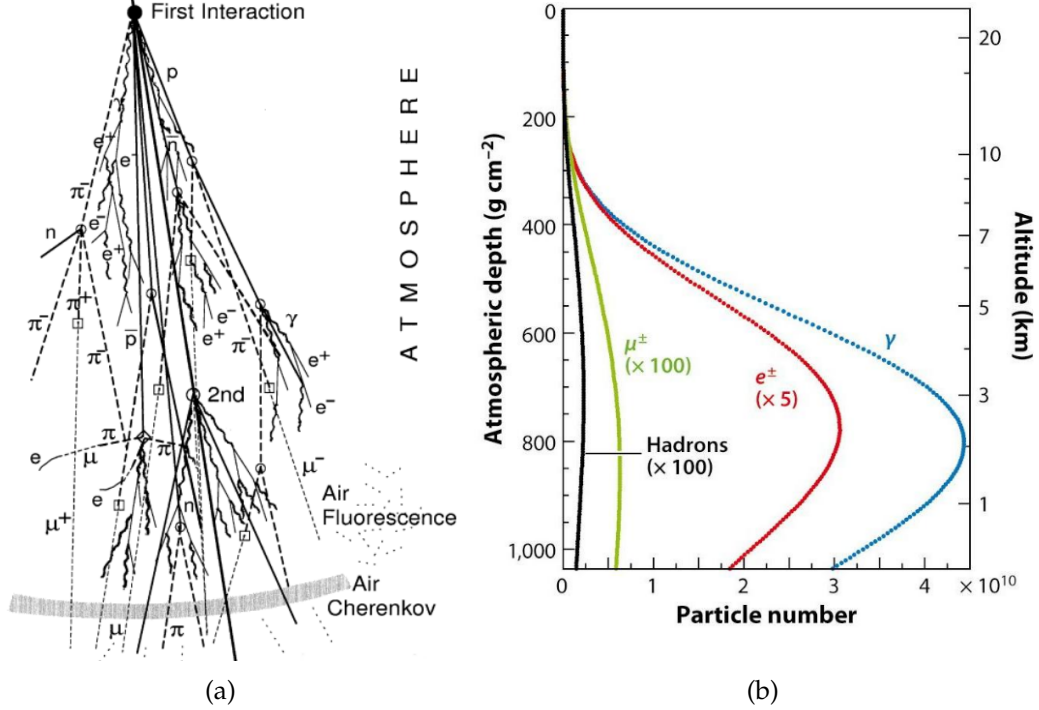


Figure 1.3: (a) Schematic view of an extensive air shower. Figure from [19]. (b) Longitudinal particle profiles (see subsection 1.2.1) of the different shower components, simulated with CORSIKA [20] for proton-induced showers of  $10^{19}$  eV. Figure from [18].

### 1.2.1 Longitudinal shower profile

The longitudinal profile is typically studied as a function of the amount of traversed atmosphere:

$$X(h) = \int_h^\infty \rho(h') dh', \quad (1.5)$$

where  $\rho$  is the density of air and the integral must be taken along the shower trajectory from the vertical height  $h$  to the top of the atmosphere. This quantity represents the column density of air at height  $h$  and is expressed in  $\text{g}/\text{cm}^2$ . The integration path depends on the zenith angle. By using the slant depth it is possible to compare air showers with different paths, or under different atmospheric conditions.

A simple model that describes the development of electromagnetic cascades in the atmosphere was developed by Heitler [21] and was later extended to the hadronic component by Matthews [22].

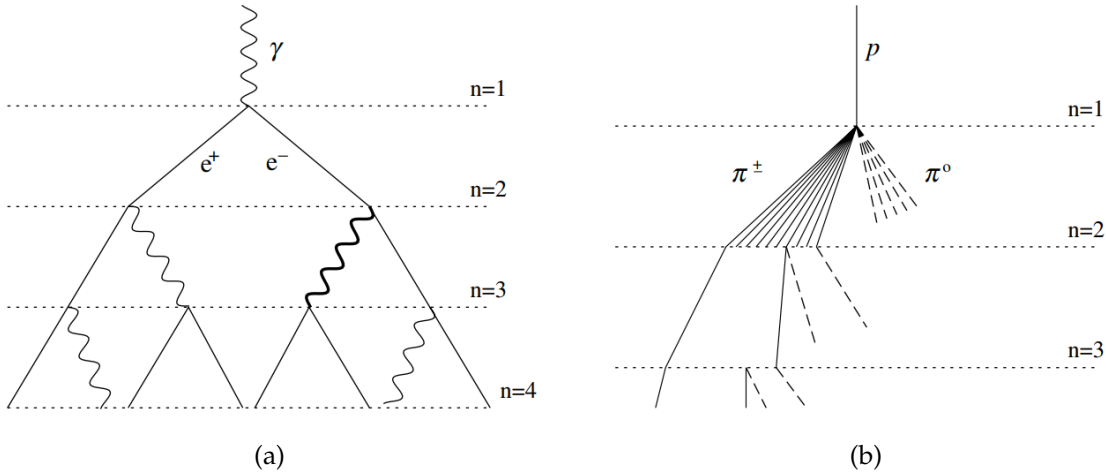


Figure 1.4: Schematic view of the Heitler-Matthews model for the development of (a) an electromagnetic shower and (b) a hadronic shower. Dashed lines indicates  $\pi^0$  which quickly decay and can initiate electromagnetic showers. Figures from [22].

According to the Heitler model, a primary electron (positron) with initial energy  $E_0$  radiates one photon in each interaction length  $\lambda_e$ , and photons convert into an electron-positron pair (Figure 1.4(a)). This process successively repeats, and at each step the energy of a particle is assumed to be equally divided between two outgoing particles. After  $n$  splitting lengths, the shower will be at the position  $X = n\lambda_e \ln 2$  in the atmosphere and there will be  $N(X) = 2^{X/\lambda_e}$  particles, each with energy  $E_X = E_0/N(X)$ . The process stops when the electron (positron) reaches a critical energy ( $E_c$ ) that is too low for pair production or bremsstrahlung. Therefore, the critical energy sets the limit on the maximum number of particles  $N_{\max} = E_0/E_c$  and it is associated to the atmospheric depth:

$$X_{\max} = \lambda_e / \ln 2 \cdot \ln(E_0/E_c). \quad (1.6)$$

Despite the simplified model, the prediction that  $X_{\max}$  linearly increases with the  $\ln(E_0)$  is qualitatively in agreement with observations and full air shower simulations that take into account all physical parameters, such as cross sections and air-densities for all individual particles in the shower [22].

The basic idea of the Heitler model can also be applied to hadronic showers: after traveling one hadronic interaction length  $\lambda_H$  the hadron produces a number of pions of which 1/3 are neutral and 2/3 are charged. Neutral pions decay and continue to feed the electromagnetic component, while charged pions interact further, until their energy drops below a critical energy and they decay into muons (Figure 1.4(b)).

Heavier nuclei with mass  $A$  and energy  $E_0$  can be described as a superposition of  $A$  nucleon showers with energy  $E_0/A$ . This is a valid assumption because the center of mass energy of the interaction between a cosmic ray and an atmospheric molecule is much bigger than the binding energy of a nucleus, therefore the nucleons in the nucleus can be considered to be independent of each other. Hence:

$$N_{\max} = A \cdot (E_0/A)/E_c \propto E_0 \quad (1.7)$$

$$X_{\max} \propto \ln(E_0/AE_c). \quad (1.8)$$

This shows that the average depth of the shower maximum  $X_{\max}$  is smaller for heavier nuclei, while  $N_{\max}$  is similar for all nuclei of the same energy.

Although the Heitler model qualitatively describes the longitudinal shower profile, a full characterization of the properties of an air shower requires a detailed Monte Carlo simulation. Creating simulations of air showers is a challenge, because the number of charged particles involved can be enormous, sometimes even exceeding  $10^{10}$ , and cross-sections and particle productions have been measured at accelerators only to a limited extent.

Another important aspect to take into account is that in nature the shower is subject to stochastic variations. The air shower development for cosmic rays with the exact same properties can vary significantly, making it difficult to determine the exact properties of the primary particle on an event-by-event basis. This is true in particular for the mass of a single cosmic ray. Consequently, the mass of an individual cosmic ray can only be determined probabilistically, but the mass composition of an ensemble of cosmic rays can be determined with some certainty.

### 1.2.2 Mass Composition

To determine the mass composition of cosmic rays it is fundamental to unveil their production and propagation mechanisms. Composition measurements allow to constrain the models for the cut-off of the energy spectrum (see Section 1.1.1). The GZK-effect would be the favorable scenario when a large fraction of protons is observed at the highest energies. An heavier composition towards the end of the spectrum would suggest that the cut-off is instead due to a limited maximum energy of the accelerators, as nuclei with a larger charge are subject to a larger acceleration in the same source.

According to the Heitler model, the depth of shower maximum,  $X_{\max}$ , is a composition sensitive parameter because it scales with  $\ln(E_0/A)$ .  $X_{\max}$  can be directly measured with the fluorescence telescopes (see Section 2.1), thus it can be used as the standard quantity for the experimental determination of the composition. The fluorescence telescopes of the Pierre Auger and Telescope Array Observatories are able to measure the depth of the shower maximum with a resolution better than  $25 \text{ g/cm}^2$  [23].

Figure 1.5 shows the evolution of the average  $X_{\max}$  and its standard deviation as a function of energy as measured by the Pierre Auger Observatory. The lines indicate the values for pure iron and proton primaries predicted by various simulation codes using different hadronic interaction models. Based on these results it seems that the observed trend is moving towards intermediate-heavy primaries at the highest energies, but the interpretation of these results strongly depends on the interaction model used. Additional information can be gained from the standard deviation of the  $X_{\max}$  distribution. A large value of  $\sigma(X_{\max})$  can originate from either a light composition, with larger shower to shower fluctuations, or a mixed composition, where the spread of the distribution increases due to the superposition of distributions with a different mean. The measurements of the Pierre Auger Observatory point to a composition quite pure and heavy at the highest energies and compatible with both a light or mixed composition around  $10^{18} \text{ eV}$  [5].



The statistics provided by the fluorescence detectors is affected by their low duty cycle (less than 15%) and the stringent cuts imposed to avoid a biased data sample. To gain additional insight new independent measurements with larger statistics are needed.

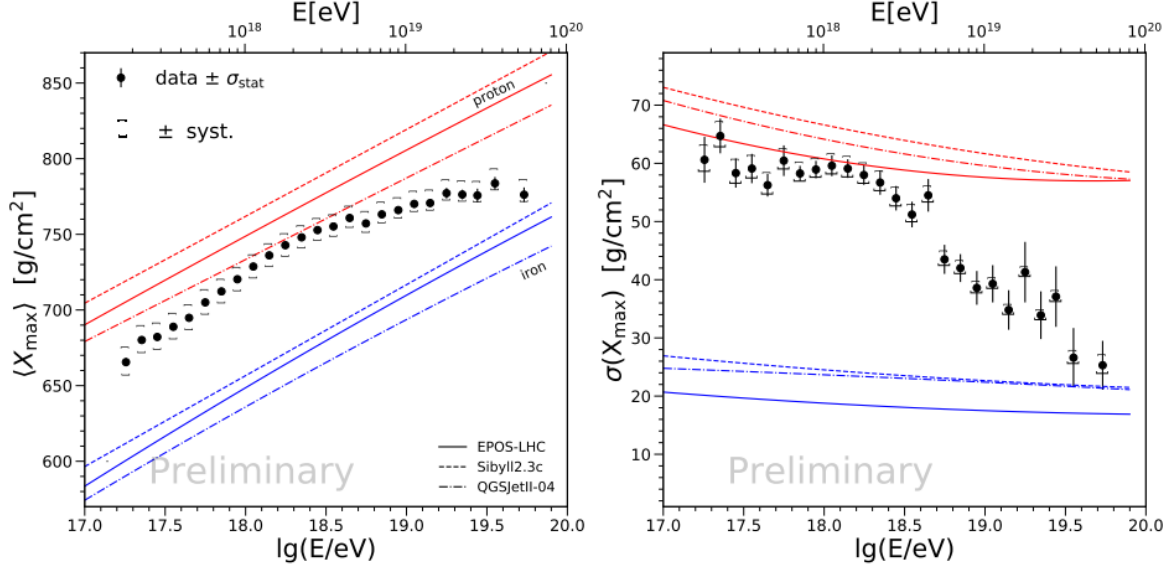


Figure 1.5: The mean (left) and the standard deviation (right) of the  $X_{\max}$  distribution as measured by the Pierre Auger Observatory as a function of energy compared to predictions for pure proton and iron compositions. Figure from [5].

### 1.3 Radio emission from extensive air showers

In the 1960s many theoretical and experimental efforts were made to understand and detect the radio emission from extensive air showers (a review of the earlier experimental results can be found in Ref. [24]). However, soon it became clear that the instruments were insufficient to reconstruct air shower parameters like energy and  $X_{\max}$  with an accuracy competitive to other cosmic-ray detection techniques. Only at the beginning of the new millennium radio detection of cosmic rays was revived, and the understanding of the radio emission has made considerable progress since then [25–29].

#### 1.3.1 Emission mechanisms and general features

Electrons and positrons in the air shower are accelerated in opposite directions by the geomagnetic field, which induces a transverse current. The transverse current varies with the electromagnetic component of the air shower: it increases until the shower reaches the shower maximum and decreases from there. The time-variation of the transverse currents leads to electromagnetic radiation polarized in the direction of the Lorentz force vector  $\vec{v} \times \vec{B}$ , where  $\vec{v}$  is the direction of movement of the shower and  $\vec{B}$  the geomagnetic field. This is called *geomagnetic emission* and its strength is proportional to the geomagnetic field and the sine of

the angle between the shower axis and the geomagnetic field direction (geomagnetic angle). The geomagnetic radiation is the dominant contribution of the radio emission from cosmic ray induced air showers.

A subdominant process, known as *charge-excess* or *Askaryan emission*, is induced by the longitudinal charge imbalance in the air shower. Electrons from atmospheric molecules are dragged along by charged particles accumulating a negative net charge excess in the shower front and leaving behind the positively charged ions. The radio emission is caused by a time-variation of the net charge-excess in the shower and is radially polarized. Its value is zero in the center at the shower axis [2, 29].

The relative fraction of charge excess emission with respect to the geomagnetic one depends on the geomagnetic angle and the shower direction. At the site of the Pierre Auger Observatory, the charge-excess emission has an average relative strength of  $0.14 \pm 0.02$  [29].

An illustration of the two emission mechanisms and their polarization directions is given in Figure 1.6. The polarization maps are shown in the shower plane perpendicular to the shower direction  $\vec{v}$ . A convenient representation of the shower plane is defined by aligning one axis with the  $\vec{v} \times \vec{B}$  vector, thus parallel to the Lorentz force, and the other one with the  $\vec{v} \times (\vec{v} \times \vec{B})$  to get an orthogonal coordinate system.

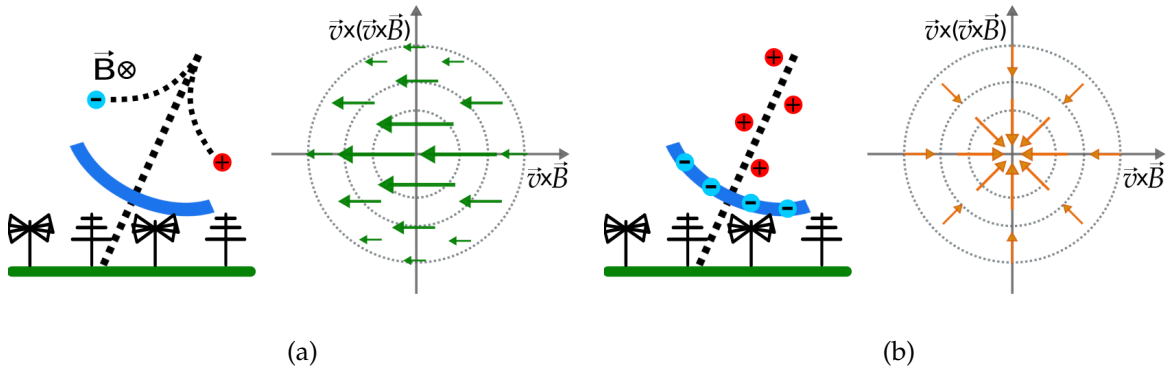


Figure 1.6: Schematic illustration of the geomagnetic (a) and the charge-excess (b) emission mechanisms and their polarization direction in the shower plane. Figure adapted from [30].

The two mechanisms interfere constructively or destructively depending on the observer location creating a radial asymmetry in the emission pattern at ground level. An example of the distribution of the total radio signal at ground is shown in Figure 1.7.

The footprint is further influenced by the atmospheric refractive index, which is larger than unity and changes with the altitude. Given that the velocity of the radio emission is governed by the refractive index of the atmosphere, signals emitted at different stages of the shower development arrive at the same time at certain observer angles. This leads to an enhancement of the pulse on a ring around the shower axis, called the *Cherenkov ring*.

### 1.3.2 Reconstruction of cosmic ray properties from the radio signal

Radio signals carry information about the nature of the primary particle. In particular, recent experimental analyses have demonstrated that the radio signal depends on the direction of

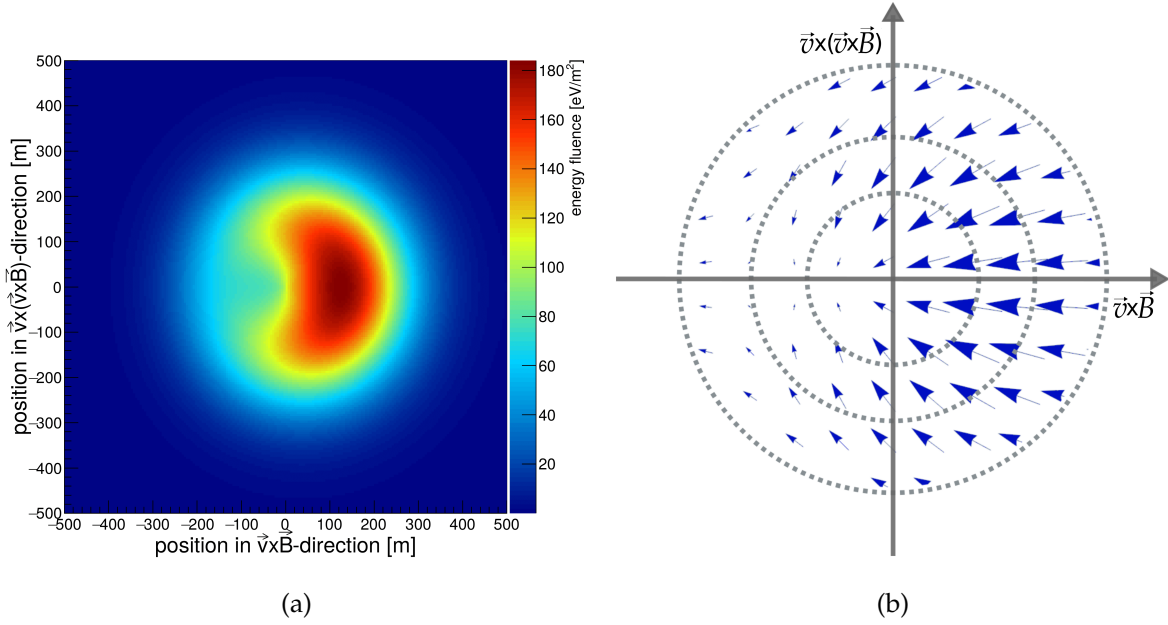


Figure 1.7: (a) Interpolation of the radio signal distribution at ground in the shower plane for a simulation of an air shower induced by a proton primary ( $E = 2.34 \text{ EeV}$ ,  $\theta = 50^\circ$ ,  $X_{\text{max}} = 772 \text{ g/cm}^2$ ). The color code indicates the signal strength. The interpolation function is described in Chapter 5. (b) Illustration of the polarization direction of the total electric field in the shower plane.

the shower axis, the energy of the electromagnetic shower component, and the distance to the shower maximum [30].

The shower direction can be reconstructed by fitting the distribution of the arrival times in several individual antennas with a wavefront model. An accuracy of  $1^\circ$  can be already achieved with a plane fit, but the LOFAR collaboration has shown that the radio shower front is better described by an hyperbola and using this model an angular resolution of  $\sim 0.1^\circ$  can be reached [31]. The hyperboloid is the expected shape created when an extended emission region is not too far away from the observer.

The amount of energy emitted in the form of radio signals by the cosmic ray induced shower is called *radiation energy*. Due to the coherent nature of the radio emission, the field strength of the radio signal is proportional to the number of electrons and positrons contributing to the emission, which is proportional to the energy of the electromagnetic shower component. As the radiation energy scales quadratically with the field strength, it scales quadratically with the energy of the electromagnetic component. This quadratic scaling has been confirmed experimentally by the Pierre Auger collaboration in [32], where the total radiation energy is determined by integrating over the footprint resulting in a cosmic-ray energy resolution of 17%.

For radio measurements  $X_{\text{max}}$  is not directly visible, but can be reconstructed from several properties of the measured signal. One of the parameters correlated to the position of the shower maximum is the size of the Cherenkov ring of the two-dimensional radio amplitude distribution [33]. By fitting the radio amplitude distribution it is possible to derive the size of the Cherenkov ring and thus estimate  $X_{\text{max}}$  [34, 35].

The radio pulse shape in the frequency domain provides an additional handle to get information on the atmospheric depth: at a given distance to the shower axis, showers initiated by

iron nuclei have on average a softer frequency spectrum than proton showers.

In Ref. [36] the hyperbolic radio wavefront reconstructed from arrival-time measurements is being used to derive  $X_{\max}$ .

This thesis will focus on the estimation of a radio  $X_{\max}$  based on amplitude and spectral information and is a continuation of the previous work described in Refs. [37–40].

### 1.3.3 CoREAS simulations

CoREAS is a Monte Carlo code for the simulation of radio emission from cosmic-ray air showers based on CORSIKA [20]. It provides a complete calculation from first principles of air shower radio emission using the endpoint formalism: particle motion is described via a series of discrete, instantaneous acceleration events, or ‘endpoints’, and each event is treated as a source of emission. This method implicitly allows for particle creation or destruction, and is suited to direct numerical implementation in either the time- or frequency-domain [41]. The radiation received at the observer position is given by the superposition of the electric fields generated in all particles tracks. The starting point for the microscopic calculation are the Liénard-Wiechert potentials that follow directly from Maxwell’s equations:

$$\begin{aligned}\phi(\vec{x}, t) &= \frac{1}{4\pi\epsilon_0} \left[ \frac{e}{(1 - n\vec{\beta} \cdot \hat{r})R} \right]_{\text{ret}} \\ \vec{A}(\vec{x}, t) &= \frac{\mu_0 c}{4\pi} \left[ \frac{e\vec{\beta}}{(1 - n\vec{\beta} \cdot \hat{r})} \right]_{\text{ret}},\end{aligned}\tag{1.9}$$

where  $R$  is the distance from the point of emission to an observer,  $\hat{r}$  is the unit vector in the direction of the observer,  $\beta$  is the velocity of the particle in units of the speed of light and  $n$  is the refractive index of the medium. The subscript “ret” denotes that the equation needs to be evaluated at the retarded time  $t' = t - nR/c$ .

The disadvantage of the microscopic calculation is that it is computationally expensive, because it requires a separate calculation for each observer position, given that the Liénard-Wiechert potential depends on  $\hat{r}$ .

Uncertainties in the simulation of the radio emission arise only from approximations made in the simulation to speed up the computation or from the simulation of the air shower itself (mainly uncertainties in hadronic interactions) and the modeling of the atmosphere.

CoREAS simulations are used in several analyses presented in this thesis.





## 2 The Pierre Auger Observatory

The Pierre Auger Observatory is the world's largest cosmic ray observatory built to investigate the origin and properties of cosmic rays with energies of  $\sim 10^{17}$  eV and above. It is located in the province of Mendoza in Argentina, on a vast high plain at an average height of 1400 m above sea level, which is the optimal altitude for the detection of showers at their maximum development at the energies of interest. The total instrumented area is approximately 3000 km<sup>2</sup>, this large detector surface is essential to measure a significant number of UHECRs. The construction of the observatory started in 2002 and was completed in 2008. The first air shower data was recorded in 2004 and since then the observatory has been operated successfully [42].

The observatory is designed as a hybrid detector measuring air showers with different complementary detection techniques. The baseline detector consists of a Surface Detector array (SD) and a Fluorescence Detector (FD), the information of both detectors is complementary and is used to reconstruct and cross-calibrate the relevant parameters of the air shower.

In order to test and calibrate new detection techniques in parallel to the established ones, the observatory was extended with a muon detector (AMIGA) and with the Auger Engineering Radio Array (AERA) to measure the radio emission from air showers. Moreover, an extensive system of atmospheric monitoring devices has been installed: detailed knowledge of the atmosphere is essential for the accurate reconstruction of air showers observed by the FD and the SD. The variations of temperature, pressure and humidity affects the longitudinal development of extensive air showers and the amount of fluorescence yield as well as the Molière radius and, hence, the lateral distribution of the electromagnetic component of the shower particles. The atmospheric state variables are also needed to determine the Rayleigh scattering of the fluorescence and Cherenkov light. The layout of the array is shown in Figure 2.1.

In this chapter the different detectors of the Pierre Auger Observatory will be described in more detail.

### 2.1 The Fluorescence Detector

The charged particles in air showers excite atmospheric nitrogen molecules, these molecules emit UV fluorescence light when returning to the ground state. The number of emitted fluorescence photons is proportional to the number of excited molecules, and therefore to the number of charged particles in the air shower.

The FD telescopes observe the longitudinal development of an extensive air shower and are used to derive the fluorescence emission as a function of atmospheric slant depth  $X$ , i.e. the column density of the air along the path of the air shower. This measurement allows a direct

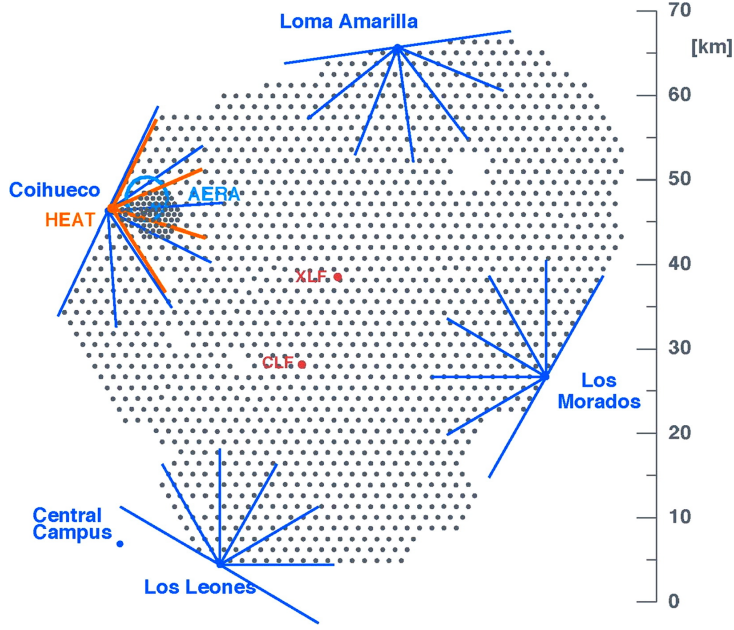


Figure 2.1: Map of the Pierre Auger Observatory. The black dots indicate the locations of the surface detectors, and the lines mark the boundaries of the fields of view of the fluorescence telescopes. AERA is located in the field of view of the Coihueco and HEAT fluorescence telescopes. The red dots mark the locations of the two laser facilities used for the atmospheric monitoring. Figure from [43].

and accurate determination of  $X_{\max}$ . Furthermore, the integral of the longitudinal profile gives the total energy dissipated electromagnetically, which is approximately 90% of the energy of the primary cosmic ray [44].

The fluorescence detector of the Pierre Auger Observatory consists of 24 telescopes distributed in four buildings located at the perimeter of the array. Each telescope has a field of view of  $30^\circ \times 30^\circ$  in azimuth and elevation, and a minimum elevation of  $1.5^\circ$  above the horizon. The telescopes face towards the interior of the array so that the combination of the telescopes at each site provides  $180^\circ$  coverage in azimuth.

As an enhancement to this baseline design, three High Elevation Auger Telescopes (HEAT) were built at the Coihueco site. These three telescopes can be tilted hydraulically by  $30^\circ$  to cover the range from  $30^\circ$  to  $60^\circ$  above horizon. The main purpose of HEAT is to observe low energy showers that generally reach their maximum of development at higher altitudes, hence, the crucial region around the shower maximum is outside the field of view of the standard FD telescopes. HEAT allows to lower the energy threshold for air shower measurements down to  $10^{17}$  eV.

A sketch of the layout of an FD telescope is shown in Figure 2.2. The light enters through a large UV-passing filter window and a Schmidt optics corrector ring designed to provide wide fields of view with limited aberrations. A segmented spherical mirror then focuses the light onto a camera made out of an array of photomultiplier tubes (PMTs). Further details are given in [44]. The detection of fluorescence light is restricted to moonless nights with good weather conditions resulting in a duty cycle of less than 15%. In the FD, air showers are detected as a sequence of triggered pixels in the camera. The PMT data are processed through a three-stage trigger system: the first level trigger compares the voltage of each PMT to a dynamic threshold adjusted in order to keep the trigger rate constant and compensate for

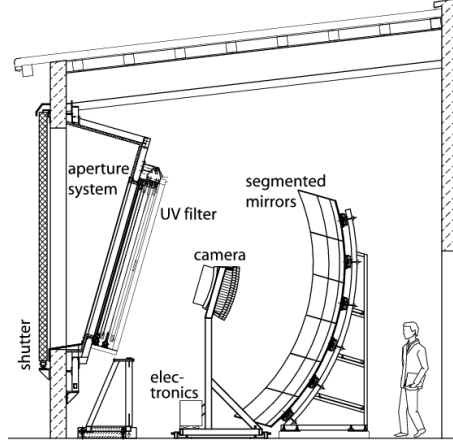


Figure 2.2: Schematic view of a fluorescence telescope of the Pierre Auger Observatory [44].

varying background light conditions, the second level trigger searches for track segments at least five pixels in length within a camera, and the third level trigger rejects lightning events and randomly triggered pixels.

An example of an event propagating through two adjacent FD telescopes is presented in Figure 2.3(a). The measured longitudinal shower profile as a function of slant depth (Figure 2.3(b)) is well described by the Gaisser-Hillas function [45, 46]:

$$f_{GH}(X) = w_{\max} \left( \frac{X - X_0}{X_{\max} - X_0} \right)^{\frac{X_{\max} - X}{\lambda}} \exp \left( \frac{X_{\max} - X}{\lambda} \right) \quad (2.1)$$

where  $\lambda$  and  $X_0$  are two shape parameters, and  $X_{\max}$  the depth where the shower reaches its maximum energy deposit  $w_{\max}$ . The resolution on  $X_{\max}$  is lower than  $20 \text{ g/cm}^2$  and the energy resolution amounts to about 8%. Combining the timing information from the telescope pixels and from the surface detector stations, the location of the impact point on ground of the air shower, the *shower core*, is estimated with a resolution of 50 m. The typical resolution for the arrival direction of cosmic rays is  $0.6^\circ$  [42, 44].

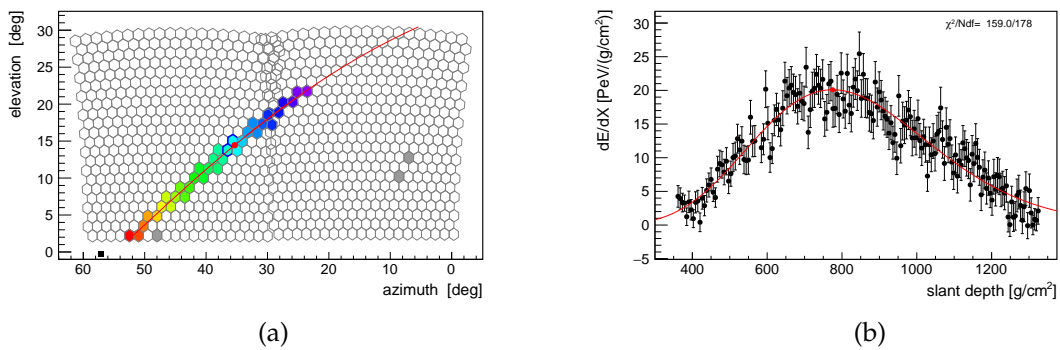


Figure 2.3: (a) Combined camera image of an air shower detected by two neighboring fluorescence telescopes. The different colors indicate the timing sequence of the triggered pixels (purple = early, red = late). The full line is the fitted shower-detector plane, the plane that includes the location of the telescope and the shower axis. (b) Energy deposit as a function of the slant depth fitted with the Gaisser-Hillas function.

A precise knowledge of the atmospheric conditions above the observatory is required for a

correct reconstruction of air showers parameters. Weather stations are installed at each FD building and at the central laser facility (CLF) and provide pressure, temperature, humidity and wind information at ground level. Furthermore, systems for the monitoring of clouds, aerosol content and lightnings are constantly operating. A description of all other atmospheric monitoring instruments of the Pierre Auger Observatory is available in Ref. [42].

Another crucial piece of information is the atmospheric profile as a function of altitude. The most-used model is the Global Data Assimilation System (GDAS), which combines several meteorological measurements with numerical weather predictions and provides the main state variables of the atmosphere every three hours [47].

### 2.2 The Surface Detector

The surface detector is an array of 1600 water Cherenkov detectors deployed on a triangular grid with a spacing of 1500 m. In total the SD covers an area of  $\sim 3000 \text{ km}^2$ .

An SD station is a cylindrical tank filled with 12 tonnes of purified water. The Cherenkov light produced by charged particles of EAS in the water is detected by three 9 inches diameter photomultipliers. Each station is also equipped with a GPS antenna, a battery and a solar panel. The signals are digitized locally and the information is transmitted wirelessly to the central data acquisition system (CDAS). A schematic illustration of an SD station can be seen in Figure 2.4.

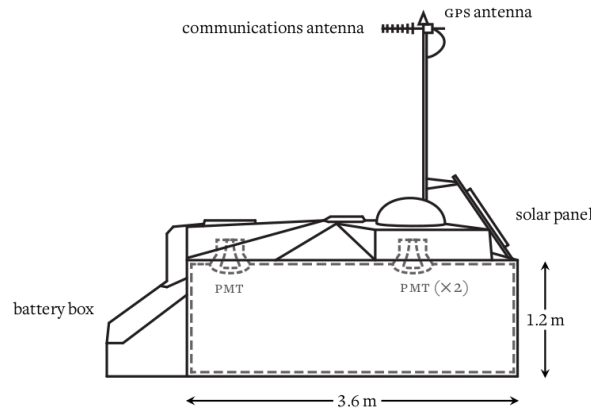


Figure 2.4: Schematic drawing of a surface detector station of the Pierre Auger Observatory. From this perspective the third PMT is directly behind the right one. Figure and caption adapted from [40].

The SD stations are sensitive to the muons and charged electromagnetic particles of the shower, as well as to high-energy photons which convert into electron-positron pairs in the water. They measure the front of the shower as it reaches the ground with an up-time of nearly 100%.

When at least three stations have triggered within  $25 \mu\text{s}$  and satisfy certain spatial conditions, the CDAS creates an event with all triggered stations.

The reconstruction of air showers from SD data is based on the arrival time and the strength of the signals in multiple detector stations. For convenience, the measured signal voltages are converted into units of vertical equivalent muon (VEM), which is the average charge deposited by a muon traversing the tank on a vertical trajectory. The signal strength in VEM as a function

of the distance  $r$  to the shower core can be parameterized by a Lateral Distribution Function (LDF), such as a modified Nishimura-Kamata-Greisen (NKG) LDF [48, 49]:

$$S(r) = S(r_{\text{opt}}) \cdot \left( \frac{r}{r_{\text{opt}}} \right)^{\beta} \left( \frac{r + r_{\text{scale}}}{r_{\text{opt}} + r_{\text{scale}}} \right)^{\beta+\gamma} \quad (2.2)$$

where  $r_{\text{scale}}$  is a scale parameter with a value of 700 m and  $r_{\text{opt}}$  is the optimum distance where signal fluctuations are minimal.  $r_{\text{opt}}$  corresponds to 1000 m for the array with a spacing of 1500 m, and 450 m for the infill array.  $\beta$ ,  $\gamma$  and  $S(r_{\text{opt}})$  are the fit parameters, in particular  $S(r_{\text{opt}})$  is proportional to the shower signal height and is used as an energy estimator for the surface detector. A typical fit of the lateral distribution function to an SD event is shown in Figure 2.5.

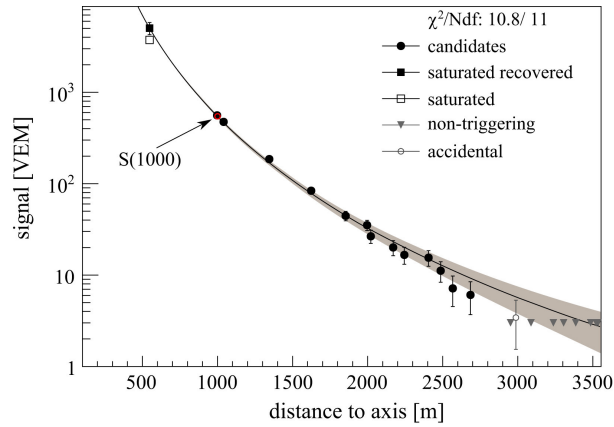


Figure 2.5: The lateral distribution of the signals as a function of the distance from the shower core. The shaded line is the resulting fit with Eq. 2.2. Figure from [42].

In addition to using the SD for the measurement of arrival direction and cosmic-ray energy, considerable effort has been made to infer the depth of shower maximum  $X_{\text{max}}$  using data from the surface detector [50–52].

AMIGA (Auger Muons and Infill for the Ground Array) is the low energy extension of the surface detector. Located in the north-west corner of the array (see Figure 2.1), it consists of an infill of 61 water Cherenkov detectors with 750 m spacing, paired with underground scintillators to directly measure the muon content of the particle showers.

The surface detector is fully efficient for the detection of air showers with energies above  $3 \times 10^{18}$  eV, for all zenith angles between  $0^\circ$  and  $60^\circ$ , and the 750 m spaced infill array is fully efficient for events with an energy above  $3 \times 10^{17}$  eV and a zenith angle below  $55^\circ$  [53].

## 2.3 The Auger Engineering Radio Array

The Auger Engineering Radio Array (AERA) detects radiowave emission from extensive air showers in the 30-80 MHz frequency range. It is located within the infill region of the surface detector and in the field of view of the Coihueco and HEAT fluorescence telescopes (Figure 2.1). AERA measures air showers with energies above  $10^{17}$  eV which allows for a coincident detection of air showers with AMIGA and HEAT.

As AERA is an engineering array, different antenna types, electronics and trigger systems have been designed and tested in the field. The deployment was done in three phases. In September 2010, 24 radio detector stations equipped with log-periodic dipole antennas (LPDA) were installed on a 144 m triangular grid (AERA-I) and since March 2011 these have been taking data. In a second stage, 100 radio detector stations equipped with a type of bow tie antenna, called *Butterfly*, were deployed with spacing of 250 m and 375 m (AERA-II). Lastly, in 2015, 29 additional radio detector stations equipped with Butterfly antennas were deployed on a 750 m triangular grid (AERA-III). In total AERA counts 153 radio detector stations distributed over an area of 17 km<sup>2</sup>.

Two different types of electronics have been developed for the AERA radio stations. One type was designed by the Karlsruhe Institute of Technology (KIT) and the Bergische Universität Wuppertal (BUW), and can handle external-triggers provided by the baseline Auger detectors. The other type was designed and assembled by the Nikhef and the Radboud University (RU) group, and focuses on self-triggering using radio pulses and a scintillator-trigger. The distribution of the electronics over the antenna stations changed multiple times. From March 2012 to November 2016, 6 LPDA and 40 Butterfly stations were equipped with the Nikhef/RU electronics, in November 2016 the 6 LPDA stations have been switch to the KIT/BUW electronic design. The layout of AERA, including all three deployment stages and the current electronics distribution, is shown in Figure 2.6.

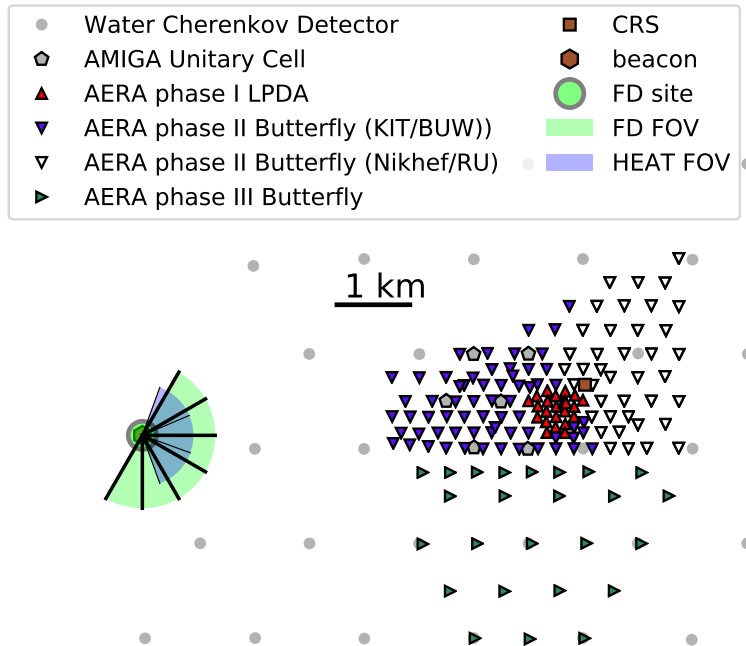


Figure 2.6: Layout of the AERA experiment. The triangles represent the radio detector stations: the orientation indicates the three stages of expansion and color the version of the electronics. The radio detector stations are surrounded by surface detector stations (gray filled circles) and underground muon counters (black pentagons). The fields of view of the Coihueco and HEAT fluorescence telescopes are also indicated. Figure adapted from [54].



### 2.3.1 Antenna design

Each radio detector station consists of two antennas perpendicularly aligned to each other. One antenna is aligned into the magnetic north-south direction and the other into the east-west direction.

The first 24 radio stations deployed in the AERA field use a log-periodic dipole antenna (LPDA). The LPDA consists of nine dipoles with different lengths designed to achieve a broadband sensitivity in the 30-80 MHz interval. The rest of the stations use Butterfly antennas, characterized by a much simpler design consisting of two triangular arms. A photo and technical drawing of the two different antenna types are shown in Figures 2.7 and 2.8.

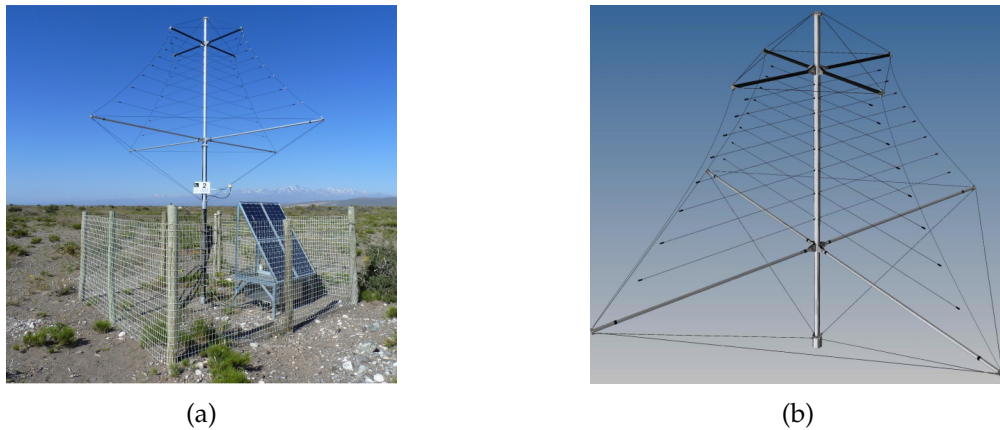


Figure 2.7: (a) Picture of an LPDA radio station in the Argentinean Pampa. The box with the electronics (and two scintillators for the Nikhef/RU design) is placed underneath the frame holding the solar panel. The station is surrounded by a fence to protect the antennas from animals. (b) Technical drawing of the LPDA antenna design. Figures and from [55].



Figure 2.8: (a) Picture of a Butterfly radio station in the Argentinean Pampa. (b) Technical drawing of the Butterfly antenna design, the triangular box hosts the electronics (and two scintillators for the Nikhef/RU design) and is used as support for the solar panel. Figures and from [55].

Each radio detector station has been deployed to work independently with its own solar power and battery system, GPS receiver for timing and is protected from animals by a fence made using non-conductive material not to influence the antenna response. All station electronics are placed inside a metallic box that protects from dust and weather influences, and shields

possible radio frequency interference (RFI). In case of an LPDA station, the box is placed underneath the support structure of the solar panel. For a Butterfly station, a triangular box is mounted on the main pole and the solar panel is placed on the box. The stations equipped with the Nikhef/RU electronics system are provided with two small plastic scintillators also positioned in the electronics box, they measure the charged particles in the air shower and are used for triggering.

The LPDA stations are connected via optical fibers to the central radio station (CRS). This wired connection was no longer feasible for AERA-II and AERA-III which have a much larger grid spacing. Instead, a 5.7 GHz commercial wireless communication system antenna is installed on top of each Butterfly antenna station, and connects to one of the four access points, two installed at the CRS and two at Coihueco.

A significant difference between the two antennas is the sensitivity to signals reflected from the ground: the LPDA is mostly insensitive to signals entering the antenna from below, while the Butterfly explicitly uses the reflected waves to enhance its signal. For this reason the Butterfly response is influenced by the electronic box and all the conductive parts below the antenna, and strongly depends on ground conditions (i.e. wet ground), which are important sources of systematic uncertainty.

A detailed documentation of the design of the AERA radio detector stations can be found in Refs. [55] and [56].

### 2.3.2 Station electronics, trigger and data acquisition

The signal received by the antenna is sent via coaxial cable to a low-noise amplifier (LNA), placed at the bottom of the LPDA antenna or close to the center of the Butterfly antenna, and then sent to the digitizer in the electronics box.

Each station is equipped with one of the two custom-made digitizer boards. In both setups an analog band-pass filter from 30 to 80 MHz is applied in the digitizer box before digitization. More details about the electronics and their corresponding trigger strategies are described below, separately for the two designs.

#### **KIT/BUW setup and external-trigger**

The KIT/BUW digitizer is equipped with a filter/amplifier which splits the signals into low-gain (+29 dB after the antenna) and high-gain channels (+49 dB). Afterwards, the signal is digitized at 180 MHz by 12 bit analog-to-digital converters (ADCs). The digitizer contains also a ring buffer to store data for up to 7.4 s, while the station waits for external-triggers from the SD or FD.

Externally triggered stations have the advantage that their measurements contain signals in coincidence with air showers, even when the radio signal might be too low to be detected. On the other hand, since only a rough estimation of the position of the radio pulse is possible, a long time trace needs to be stored. Currently a trace of 57  $\mu$ s is saved, which is much longer compared to the width of a cosmic ray pulse which is only about 100 ns long. Furthermore all active stations are read out because the position of the shower core can not be evaluated from

the trigger information. The combination of these two factors leads to large data volumes of  $\mathcal{O}(10 \text{ GB})$  per month.

The data acquisition (DAQ) is located at Coihueco and receives messages with time stamps and station lists of SD events directly from the CDAS. When the closest SD station is within 5 km from any of the externally triggered AERA station, the DAQ sends a trigger message and data from all active externally triggered stations are extracted backdated to the time of the external trigger.

In addition to receiving an external-trigger, it is also possible to trigger on radio signal, but at the moment self-triggered events from stations equipped with KIT/BUW electronics are not used for physics analysis.

In this thesis only the data from the externally triggered stations are used.

### **Nikhef/RU setup and scintillator- and self-trigger**

In the Nikhef/RU electronics two digitization channels are used for bandwidth filtered radio measurements without further amplification after the LNA, and the other two channels are dedicated to the read out of the signals from the two scintillators. The ADCs sample the analog traces with a dynamic range of 14 bits at 200 MHz. This setup does not accept external-triggers but is designed to self-trigger on any combination of the four connected channels. The scintillators are shielded from most noise sources, their signal is very clean and a simple trigger based on a single threshold above the baseline is used. An advantage of triggering using a local particle detector is that the trigger time is very close to the time of the radio pulse, thus, the signal search window can be constrained and the recorded time traces are 10  $\mu\text{s}$  long. Instead, the main limitation is the small cross section of the scintillator modules. This is overcome by including self-triggered stations that do not have a scintillator signal in the dataset. Self-triggering is quite challenging because the noise conditions at the AERA site are variable, hence, the signal threshold needs to be adjusted according to the background situation. An FPGA (field-programmable gate array) in the digitizer is programmed to implement advanced trigger algorithms and signal cleaning procedures [57].

The data from a triggered station is temporarily stored and the trigger time is sent to the DAQ. When three or more time stamps received by the DAQ fall within a coincidence window of 3  $\mu\text{s}$  an event is built with data from all triggered stations. The DAQ for the stations equipped with the Nikhef/RU electronic setup is located at the CRS. Self-triggered stations constitute an important concept for future independent large-scale radio arrays.

In addition to the triggered events, each data acquisition system requests read-outs of all active stations once every 100 s. These events are used for calibration purpose and to study the background at the AERA site, and are called *periodically triggered* data.

## **2.4 AugerPrime - the upgrade of the Pierre Auger Observatory**

A substantial upgrade of the Pierre Auger Observatory (AugerPrime) has started in 2016. The main goal of the upgrade is to improve the mass determination of primary cosmic rays with an energy above  $10^{19.5} \text{ eV}$  by measuring the ratio between the electromagnetic and the muonic

component for each event. For this purpose, all the water Cherenkov detectors, except those in the outer ring, are being equipped with a  $3.8 \text{ m}^2$ , 1 cm thick Scintillator Surface Detector (SSD). This allows the sampling of the shower front with two detectors with different responses to muons and electromagnetic particles, enabling a statistical separation of the muonic and electromagnetic components of the shower.

The SD stations will also be supplied with new electronics having better timing accuracy, and a small photomultiplier to increase the dynamic range of the signal readout.

In parallel with the surface detector changes, the upgrade will include a large array of buried muon detectors, and an extended duty cycle for operations of the fluorescence detectors [58].

### 2.4.1 A large Radio Detector at the Pierre Auger Observatory

As part of the AugerPrime upgrade, in addition to the surface scintillator detector, each SD station will be equipped with a radio antenna, forming a  $3000 \text{ km}^2$  radio array (Radio Detector, RD).

A photograph of the first AugerPrime SD station prototype installed in the field and a schematic view of the design are shown in Figure 2.9. The antenna that will be used in the

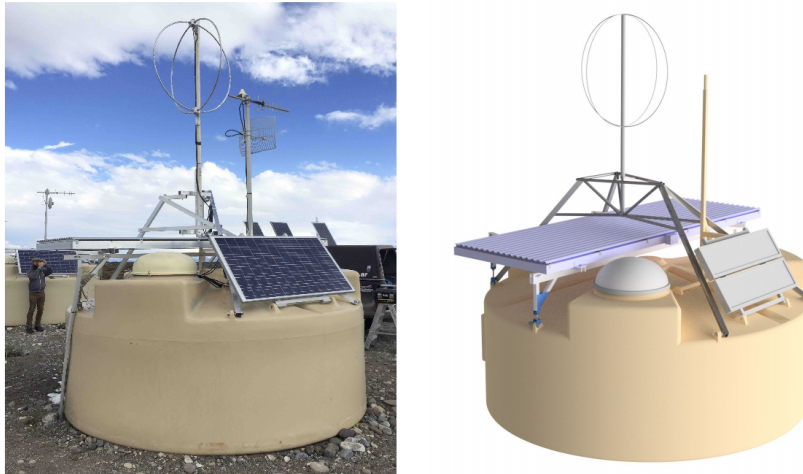


Figure 2.9: A photograph (left) and the schematic view (right) of an upgraded station of the surface detector. The upgrade includes a surface scintillator detector (Section 2.4) and a radio station equipped with SALLA antennas. Figure and caption from [59].

RD stations is a short aperiodic loaded loop antenna (SALLA). This antenna type combines a simple and robust design with ultra-wideband sensitivity and low costs for production and maintenance, essential features in the view of a large-scale radio detector [59].

The main objective of the RD is to measure air showers with zenith angles beyond  $60^\circ$  and enable the measurement of  $e/\mu$  separation for horizontal showers. Inclined showers travel a longer distance through the atmosphere, therefore when the shower reaches the detector the electromagnetic component is mostly absorbed, furthermore the cross section of the SSD becomes small for horizontal showers. Thus, only the muon signal is measured in the SD and the electron-to-muon ratio can not be measured by the combination of water Cherenkov detector and the scintillator detector. However, as there is no absorption of the emitted radio

signal in the atmosphere, horizontal air showers create a large radio footprint on the ground enabling measurement of the electromagnetic component of inclined air showers with a sparse antenna array.



## 3 Electric field reconstruction from the radio detector data

In this chapter the fundamental steps to obtain the electric field vector from the measured time series of ADC counts of the antennas will be described. The data from the different detectors of the Pierre Auger Observatory can all be processed simultaneously with the *Offline* software framework, specifically developed to handle the hybrid nature of the observatory. The general features of *Offline* and of the radio data reconstruction are described in the last section of this chapter.

### 3.1 The electric field vector

Each air shower emits a radio frequency electromagnetic field. If the emission region is far enough from the radio detector station, the electromagnetic radiation arrives approximately as a plane wavefront perpendicular to the direction of the radiation.

The three-dimensional electric field vector at the detection point can be calculated by unfolding the antenna response pattern from the voltage trace  $U(\nu)$ . The antenna response pattern is described via the *vector effective length* (VEL),  $\vec{H}$ , and it depends on the frequency,  $\nu$ , and the arrival direction,  $(\theta, \phi)$ . Thus, for each antenna arm:

$$U(\nu) = \vec{E}(\nu) \cdot \vec{H}(\nu, \theta, \phi) \quad (3.1)$$

$$\vec{E}(\nu) = E_\theta(\nu)\hat{e}_\theta + E_\phi(\nu)\hat{e}_\phi, \quad (3.2)$$

where  $\hat{e}_\theta$  and  $\hat{e}_\phi$  are the unit vectors of the spherical coordinate system shown in Figure 3.1. The component of the electric field parallel to the direction of movement of the shower  $\hat{e}_r$  is negligible.

For the Butterfly and the LPDA antenna type of the AERA array, the VEL has been simulated using a package specifically designed for simulating antennas characteristics (NEC-2 [61]). For a correct output, all parts of the radio detector stations that can influence the antenna response need to be included in the simulation, thus, when simulating the Butterfly antenna the triangular box and the communication antenna are also taken into account. The two components of the simulated VEL for the EW polarization of the Butterfly antenna are shown in Figure 3.2.

In 2015 an in-situ measurement campaign was performed to determine the VEL of Butterfly and LPDA antennas. Measurements from several directions were taken using a transmitting antenna and a signal generator carried by an octocopter drone. The octocopter was also equipped with sensors to measure its altitude, position and alignment. The overall uncertainty of the Butterfly measured VEL is 10.3% for  $|H_\phi|$  and 15.4% for  $|H_\theta|$ , while for the LPDA it is



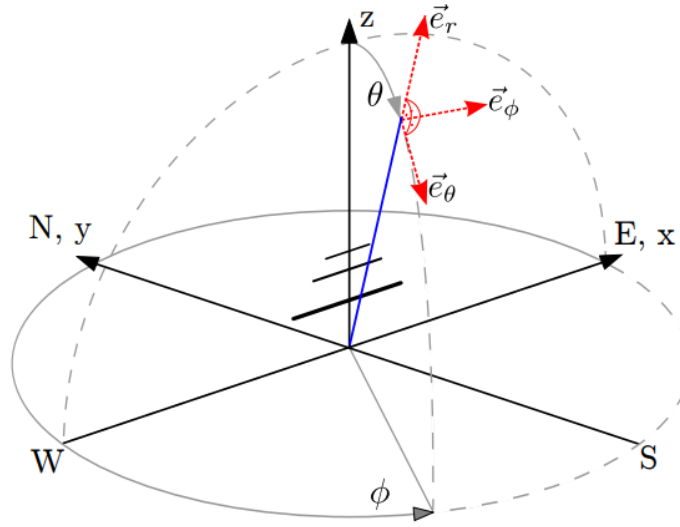


Figure 3.1: Coordinate system with the antenna placed in the center. The origin of the coordinate system is located in the  $xy$ -plane. The zenith angle  $\theta$  is zero for shower directions aligned with the  $z$ -axis, the azimuth angle  $\phi$  is measured counterclockwise from the  $x$ -axis. The blue line represents the shower direction corresponding to  $\theta$  and  $\phi$ . The units vectors of the vector effective length  $\vec{H}$  for the specified direction are shown in red. Figure from [60].

7.4% for  $|H_\phi|$  and 10.4% for  $|H_\theta|$  [62]. The two components of the measured VEL for the LPDA antenna are shown in Figure 3.3. The comparison between the output of the simulation and the measurements for the Butterfly east-west polarization is shown in Figure 3.4. A zenith and frequency dependent discrepancy between simulation and measurement is visible in both components, especially for higher frequencies, which cannot be explained within the statistical and systematic uncertainties. The reason for the discrepancy is not yet known. Figure 3.5 shows the same comparison for the LPDA. In this case the simulated VEL generally agrees with the measured one within the overall uncertainty of the measured VEL. A detailed documentation of the measurements of each VEL component and the results can be found in Refs. [62] and [63].

As the electric field is a vector perpendicular to the incoming direction, the shower direction needs to be estimated. However, its estimation changes with the unfolding of the antenna response. Hence the electric field is determined in an iterative procedure. A first approximation of the incoming direction is obtained via triangulation of the pulse arrival times on the voltage trace, and it is used to reconstruct the electric field. Next, the pulse arrival times can be determined using the electric field trace. This is used to recalculate the incoming arrival direction and, consequently, the electric field. This procedure is repeated several times until the change in reconstructed direction is below  $0.5^\circ$ , or after a maximum number of iterations is reached. Typically the direction reconstruction converges after 2-4 iterations. This approach is computationally intensive, hence it is generally replaced by the more practical solution of using directly the directional information from the SD reconstruction.

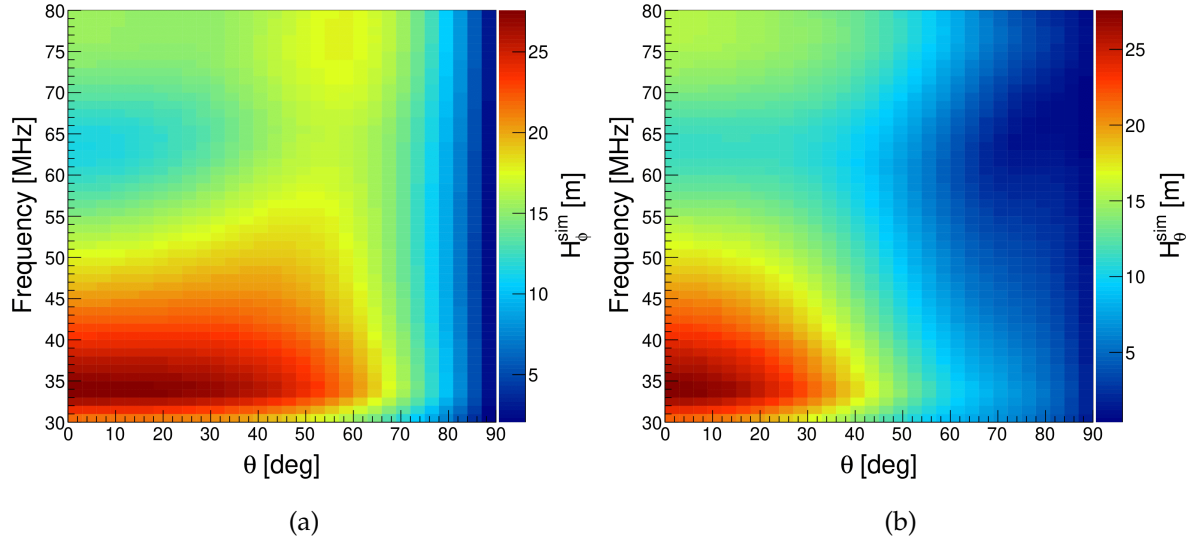


Figure 3.2: Simulated components of the VEL of the east-west Butterfly antenna as a function of the frequency and zenith angle and averaged over all azimuth angles: (a)  $H_\phi$ ; (b)  $H_\theta$ .

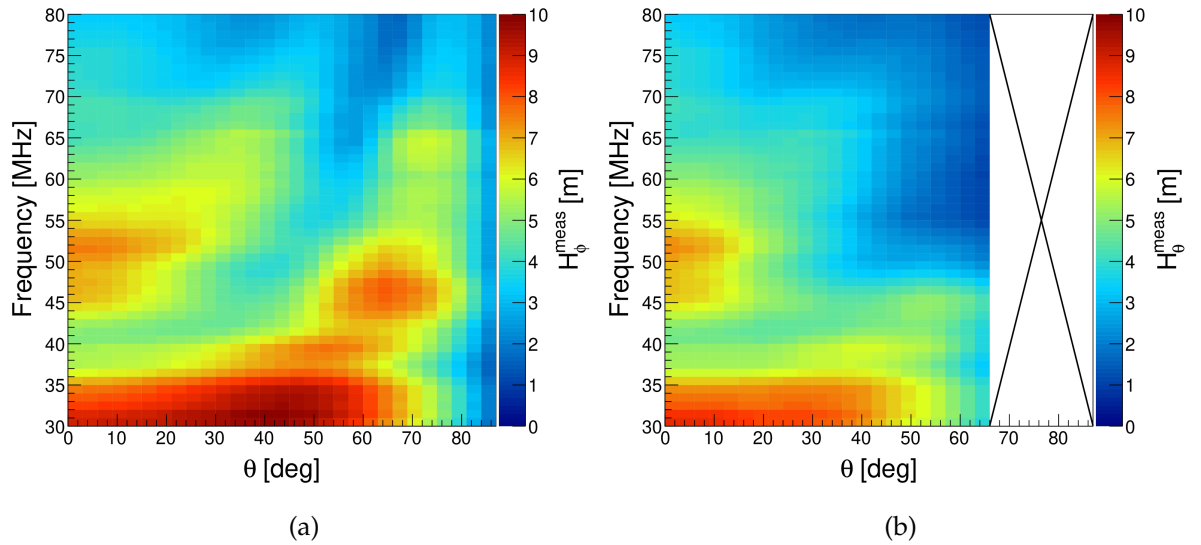


Figure 3.3: Measured components of the VEL of LPDA as a function of the frequency and zenith angle for the LPDA: (a)  $H_\phi$ ; (b)  $H_\theta$ . In the measurement of the  $H_\theta$  component the LPDA loses sensitivity for zenith angles above  $65^\circ$  and the systematic measurement uncertainty exceeds 20%. Therefore, these angles are not considered.

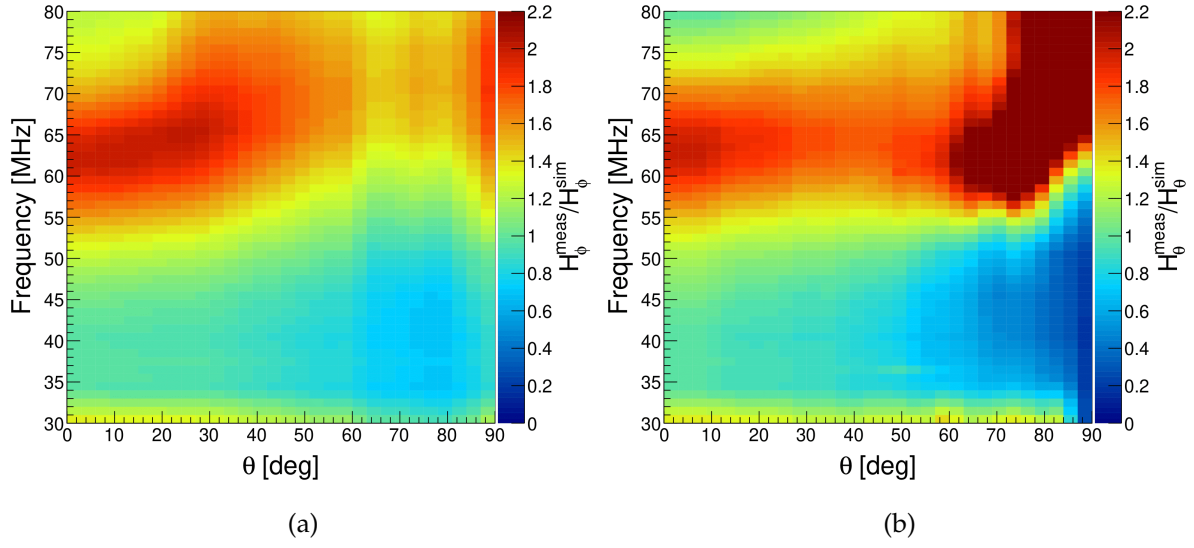


Figure 3.4: Ratio between the measured and simulated components of the VEL of the east-west Butterfly antenna as a function of the frequency and zenith angle and averaged over all azimuth angles. In (a) the  $H_\phi$  is compared; in (b)  $H_\theta$ . For  $\theta$  below  $60^\circ$  and frequencies lower than 55 MHz, measurements and simulations agree reasonably well with respect to their uncertainties of 10-15%. However, for frequencies around 60 MHz the ratio rises to values beyond 2 [63].

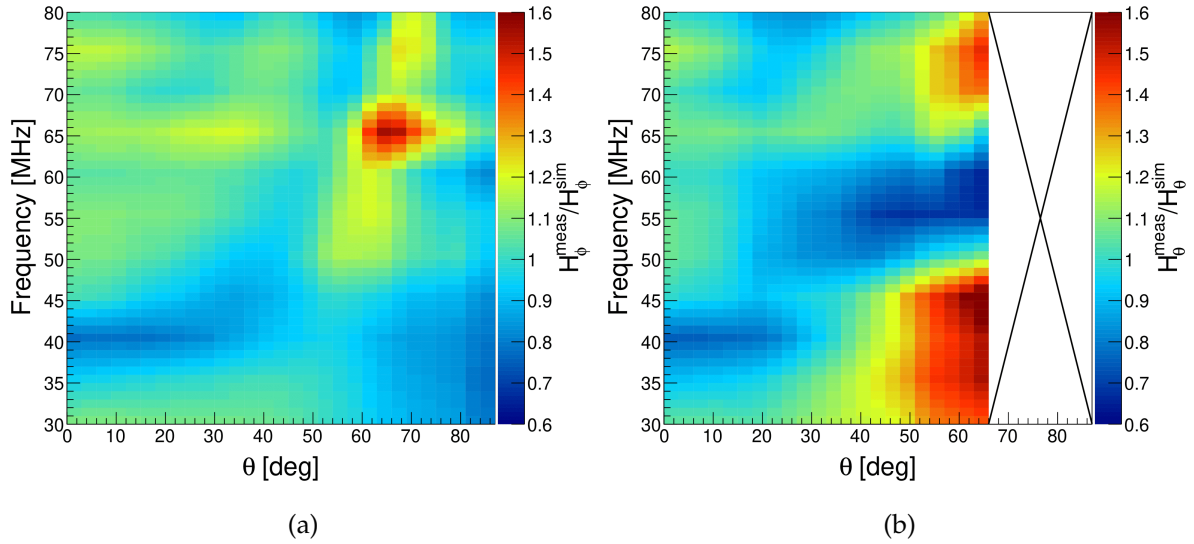


Figure 3.5: Ratio between the measured and simulated components of the VEL of LPDA as a function of the frequency and zenith angle and averaged over all azimuth angles. In (a) the  $H_\phi$  is compared; in (b)  $H_\theta$ . In the measurement of the  $|H_\theta|$  component the LPDA loses sensitivity for zenith angles above  $65^\circ$  and the systematic measurement uncertainty exceeds 20%. Therefore, these angles are not considered. In both components measurements and simulations are in reasonable agreement for zenith angles below  $50^\circ$  [62].

### 3.1.1 The total electric field and its uncertainty

The two perpendicular arms of the AERA stations are aligned in the east-west (EW) and the north-south (NS) direction, and they are assumed to measure the polarization component of the electric field in the horizontal direction in which they are aligned. As the electric field is taken to be perpendicular to the incoming direction of the radiation  $\hat{e}_r$ , the vertical component can be inferred from the two horizontal components as follows:

$$\begin{pmatrix} E_{EW} \\ E_{NS} \\ E_V \end{pmatrix} \cdot \begin{pmatrix} \cos \phi \sin \theta \\ \sin \phi \sin \theta \\ \cos \theta \end{pmatrix} = 0 \quad (3.3)$$

and

$$E_V = - (E_{EW} \cos \phi + E_{NS} \sin \phi) \tan \theta, \quad (3.4)$$

from which it follows that the uncertainty of the vertical component grows with  $\tan \theta$ , where  $\theta$  is the zenith angle.

The magnitude of the total electric field squared is:

$$\begin{aligned} E^2(\nu) &= E_{EW} E_{EW}^* + E_{NS} E_{NS}^* + E_V E_V^* \\ &= E_{EW} E_{EW}^* + E_{NS} E_{NS}^* + (E_{EW} \cos \phi + E_{NS} \sin \phi) (E_{EW}^* \cos \phi + E_{NS}^* \sin \phi) \tan^2 \theta. \end{aligned} \quad (3.5)$$

The uncertainty on the total electric field can be calculated as a function of the measured uncertainties under the assumption that the uncertainties in  $E_{NS}$  and  $E_{EW}$  are not correlated and that the uncertainty on the arrival direction is negligible. The final expression is:

$$\begin{aligned} \sigma_E^2 &= \left( \frac{2E_{EW}(1 + \cos^2 \phi \tan^2 \theta) + E_{NS} \sin 2\phi \tan^2 \theta}{2E} \right)^2 \sigma_{E_{EW}}^2 + \\ &\quad \left( \frac{2E_{NS}(1 + \sin^2 \phi \tan^2 \theta) + E_{EW} \sin 2\phi \tan^2 \theta}{2E} \right)^2 \sigma_{E_{NS}}^2. \end{aligned} \quad (3.6)$$

An illustration of this result is given in Figure 3.6, assuming that  $\frac{\sigma_{E_{NS}}}{E_{NS}} = \frac{\sigma_{E_{EW}}}{E_{EW}} = 10\%$  and that  $E_{NS} = E_{EW}$ . In this figure, the vertical axis depicts the elevation, which is  $90^\circ - \theta$ . The figure clearly shows that for horizontal showers the uncertainty becomes very large. When the elevation is below  $10^\circ$ , the uncertainty on the total electric field explodes to roughly ten times the uncertainty on the individual measurements.

### 3.1.2 The components of the electric field in the shower plane

For a more intuitive description of the radio emission from extensive air showers it is convenient to introduce a frame of reference in which the vertical axis is aligned with the shower axis ( $\vec{v}$ ), the  $x$ -axis is in the direction of the Lorentz force ( $\vec{v} \times \vec{B}$ ) and  $y$ -axis in the direction perpendicular to the shower axis and the Lorentz force ( $\vec{v} \times (\vec{v} \times \vec{B})$ ), where  $\vec{B}$  is the Earth's magnetic field. In this frame of reference the geomagnetic emission is polarized in the  $\vec{v} \times \vec{B}$

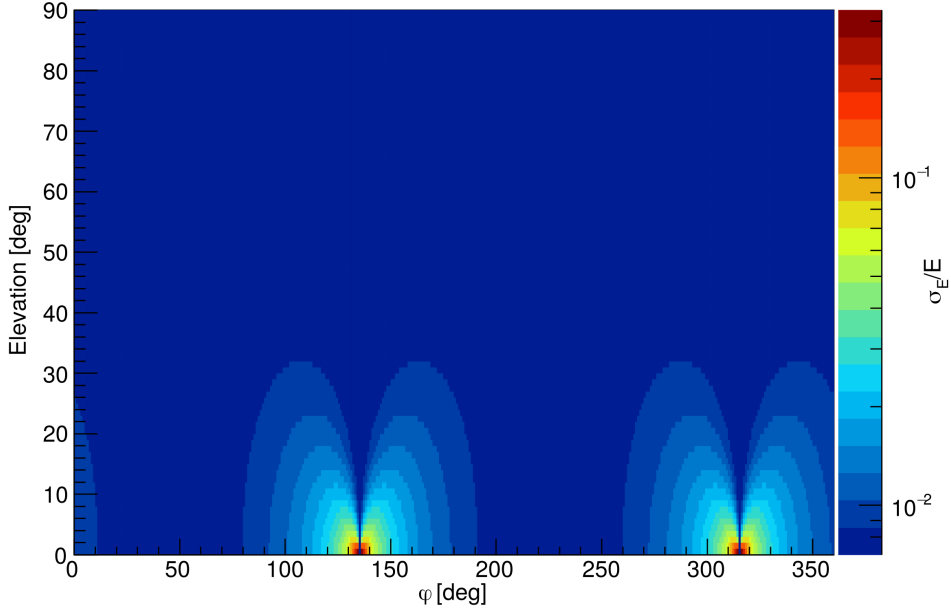


Figure 3.6: Relative uncertainty on the total electric field as a function of the elevation and  $\phi$ , assuming that  $\frac{\sigma_{E_{NS}}}{E_{NS}} = \frac{\sigma_{E_{EW}}}{E_{EW}} = 10\%$  and that  $E_{NS} = E_{EW}$ .

direction:

$$\begin{aligned} \hat{e}_{\vec{v} \times \vec{B}} &= \frac{1}{\sqrt{B^2 - (\vec{B} \cdot \hat{e}_{\vec{v}})^2}} \begin{pmatrix} \cos \phi \sin \theta \\ \sin \phi \sin \theta \\ \cos \theta \end{pmatrix} \times \begin{pmatrix} B_{EW} \\ B_{NS} \\ B_V \end{pmatrix} \\ &= \frac{1}{N_{\vec{v} \times \vec{B}}} \begin{pmatrix} B_V \sin \phi \sin \theta - B_{NS} \cos \theta \\ B_{EW} \cos \theta - B_V \cos \phi \sin \theta \\ B_{NS} \cos \phi \sin \theta - B_{EW} \sin \phi \sin \theta \end{pmatrix}, \end{aligned} \quad (3.7)$$

where

$$B = \sqrt{B_{EW}^2 + B_{NS}^2 + B_V^2} \quad (3.8)$$

and

$$N_{\vec{v} \times \vec{B}} = \sqrt{B^2 - (B_{EW} \cos \phi \sin \theta + B_{NS} \sin \phi \sin \theta + B_V \cos \theta)^2}. \quad (3.9)$$

The component of the electric field in the  $\vec{v} \times \vec{B}$  direction can be expressed as a linear combination of the components in the north-south and east-west directions:

$$\begin{aligned} E_{\vec{v} \times \vec{B}} &= \frac{E_{EW}}{N_{\vec{v} \times \vec{B}}} \left[ B_V \sin \phi \sin \theta - B_{NS} \frac{1 - \sin^2 \phi \sin^2 \theta}{\cos \theta} + B_{EW} \sin \phi \cos \phi \sin \theta \tan \theta \right] \\ &+ \frac{E_{NS}}{N_{\vec{v} \times \vec{B}}} \left[ B_{EW} \frac{1 - \cos^2 \phi \sin^2 \theta}{\cos \theta} - B_V \cos \phi \sin \theta - B_{NS} \cos \phi \sin \phi \sin \theta \tan \theta \right]. \end{aligned} \quad (3.10)$$

The uncertainty on  $E_{\vec{v} \times \vec{B}}$  is obtained again making the assumption that the uncertainties in the measurements of  $E_{NS}$  and  $E_{EW}$  are not correlated and that the uncertainty on the arrival direction is negligible. The result of the calculation, assuming that  $\frac{\sigma_{E_{NS}}}{E_{NS}} = \frac{\sigma_{E_{EW}}}{E_{EW}} = 10\%$  and

that  $E_{\text{NS}} = E_{\text{EW}}$ , is plotted in Figure 3.7. It should be noted that for zenith angles beyond  $30^\circ$  (elevations below  $60^\circ$ ), for some azimuth angles the uncertainty on the  $\vec{v} \times \vec{B}$  component of the electric field explodes to many times those of individual antenna measurements.

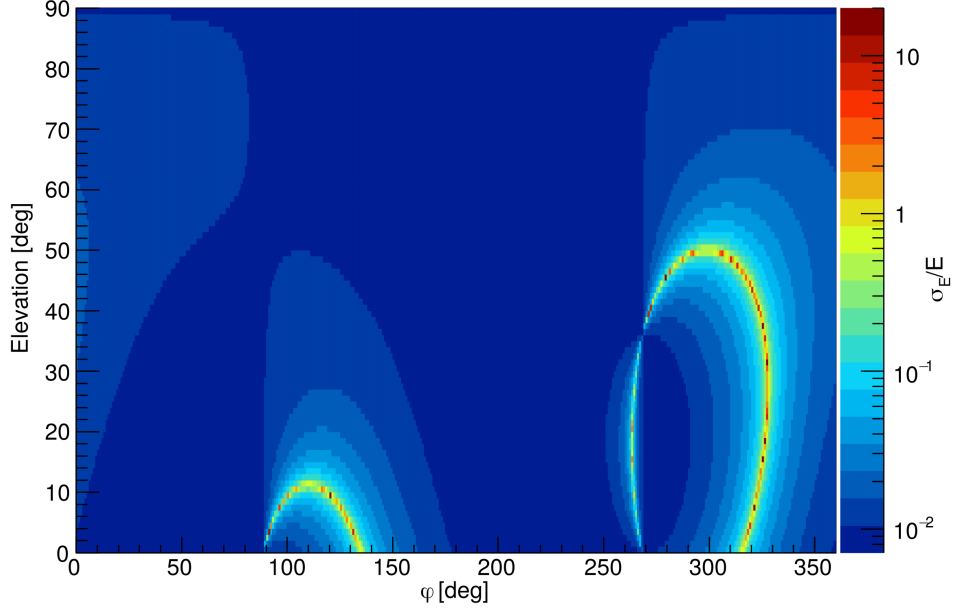


Figure 3.7: Relative uncertainty on the  $\vec{v} \times \vec{B}$  component of electric field as a function of the elevation and  $\phi$ , assuming that  $\frac{\sigma_{E_{\text{NS}}}}{E_{\text{NS}}} = \frac{\sigma_{E_{\text{EW}}}}{E_{\text{EW}}} = 10\%$  and that  $E_{\text{NS}} = E_{\text{EW}}$ .

In order to evaluate the  $\vec{v} \times (\vec{v} \times \vec{B})$  direction, the vector-relation  $\hat{e}_{\vec{v}} \times (\hat{e}_{\vec{v}} \times \vec{B}) = (\hat{e}_{\vec{v}} \cdot \vec{B})\hat{e}_{\vec{v}} - \vec{B}$  is used. Thus,

$$E_{\vec{v} \times (\vec{v} \times \vec{B})} = \frac{1}{N_{\vec{v} \times \vec{B}}} ((\hat{e}_{\vec{v}} \cdot \vec{B})\hat{e}_{\vec{v}} \cdot \vec{E} - \vec{B} \cdot \vec{E}), \quad (3.11)$$

which gives as final expression a linear combination of the components in the north-south and east-west directions:

$$E_{\vec{v} \times (\vec{v} \times \vec{B})} = \frac{E_{\text{EW}}}{N_{\vec{v} \times \vec{B}}} (B_V \cos \phi \tan \theta - B_{\text{EW}}) + \frac{E_{\text{NS}}}{N_{\vec{v} \times \vec{B}}} (B_V \sin \phi \tan \theta - B_{\text{NS}}). \quad (3.12)$$

The uncertainty on  $E_{\vec{v} \times (\vec{v} \times \vec{B})}$  for  $\frac{\sigma_{E_{\text{NS}}}}{E_{\text{NS}}} = \frac{\sigma_{E_{\text{EW}}}}{E_{\text{EW}}} = 10\%$  and  $E_{\text{NS}} = E_{\text{EW}}$  is given in Figure 3.8. It can, again, be observed that for zenith angles beyond  $30^\circ$ , for some azimuth angles the uncertainty on the  $\vec{v} \times (\vec{v} \times \vec{B})$  component of the electric field grows to many times those of individual antenna measurements.

In conclusion, a full treatment of the radio measurement uncertainties renders a rather complicated uncertainty landscape on the  $\vec{v} \times \vec{B}$  and  $\vec{v} \times (\vec{v} \times \vec{B})$  component of the electric field, depending strongly on both direction of the shower and the signal to noise ratio in the antennas. Having an additional handle on the vertical polarization of the electric field, such as a third arm to the antenna, would have reduced this complication significantly.

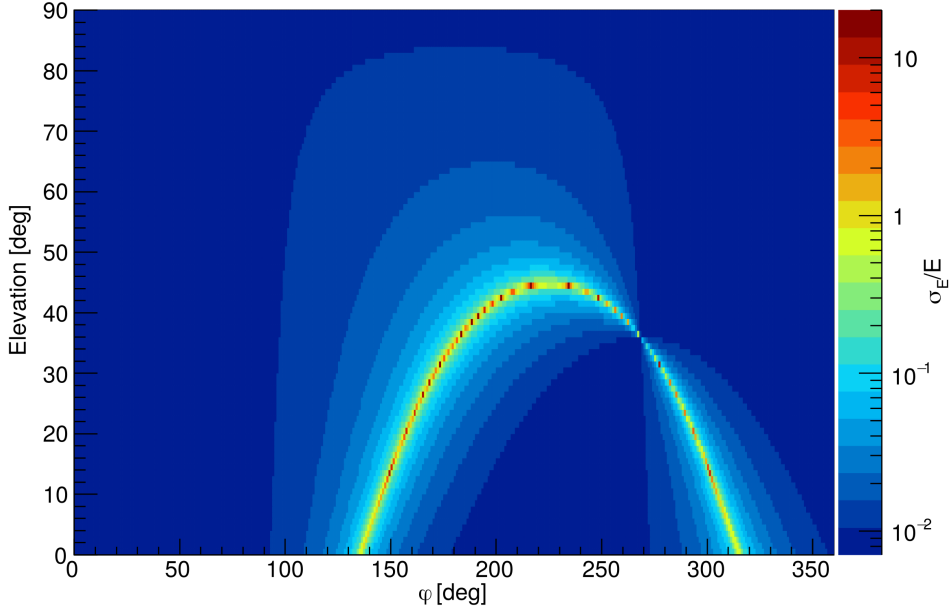


Figure 3.8: Relative uncertainty on the  $\vec{v} \times (\vec{v} \times \vec{B})$  component of electric field as a function of the elevation and  $\phi$ , assuming that  $\frac{\sigma_{E_{NS}}}{E_{NS}} = \frac{\sigma_{E_{EW}}}{E_{EW}} = 10\%$  and that  $E_{NS} = E_{EW}$ .

### 3.2 The radio signal from a noisy measurement

The measured radio signal results from the superposition of different kinds of background noise and the cosmic ray induced signal. The recorded time trace for each event is much longer than the duration of a cosmic ray induced signal, which is concentrated in a 400 ns window around the peak of the Hilbert envelope of the time trace. The parts of the time trace preceding and much after the signal window are assumed to only contain background. The expectation value of the background amplitude in the frequency domain is determined by averaging over as many background windows of 400 ns as possible, in the hypothesis that the background windows are uncorrelated.

Using the formalism defined in [37], the voltage values as a function of frequency in the signal window are given as  $M(\nu)$ , the measurement, whereas the expectation value of the background amplitude in the frequency domain is  $B(\nu)$ . The measurement squared depends on the true but unknown signal and background by vector addition as:

$$M(\nu)^2 = (S(\nu) + B(\nu) \cos \phi)^2 + B(\nu)^2 \sin^2 \phi, \quad (3.13)$$

where  $\phi$  is an unknown phase between signal and background assumed to follow a uniform distribution in the interval  $[0, 2\pi]$ .

An estimator of the true signal from the measurement can be obtained by rewriting the formula as follow:

$$S(\nu) = -B(\nu) \cos \phi \pm M(\nu) \sqrt{1 - \frac{B(\nu)^2}{M(\nu)^2} \sin^2 \phi}, \quad (3.14)$$



where  $B(\nu)$ ,  $M(\nu)$  and  $S(\nu)$  are all positive.  $1 - \left(\frac{B(\nu)}{M(\nu)}\right)^2 \sin^2 \phi$  has to be positive, therefore the phase difference between signal and background is constrained to:

$$-\frac{M(\nu)}{B(\nu)} \leq \sin \phi \leq \frac{M(\nu)}{B(\nu)}. \quad (3.15)$$

The latter is only an issue when  $M(\nu) < B(\nu)$ , i.e. there is a negative interference between signal and background, and the signal is less than twice the background.

As stated earlier  $S(\nu) \geq 0$ , therefore:

$$B(\nu) \cos \phi \leq \pm \sqrt{M(\nu)^2 - B(\nu)^2 \sin^2 \phi}. \quad (3.16)$$

When taking the  $+$ -sign, the equation above is valid for all  $\cos \phi \leq 0$  assuming the condition in Eq. 3.15 is satisfied. The only case that needs to be investigated is when  $\cos \phi > 0$ , which results in:

$$B(\nu)^2 \cos^2 \phi \leq M(\nu)^2 - B(\nu)^2 \sin^2 \phi \quad (3.17)$$

$$B(\nu)^2 \leq M(\nu)^2. \quad (3.18)$$

Therefore the  $+$ -sign is valid also for all  $\cos \phi \geq 0$ , provided  $M(\nu) \geq B(\nu)$ . This also assures that the condition in Eq. 3.15 is satisfied.

When examining the  $-$ -sign, the condition  $S(\nu) > 0$  gives:

$$B(\nu) \cos \phi \leq -\sqrt{M(\nu)^2 - B(\nu)^2 \sin^2 \phi}, \quad (3.19)$$

which is never valid for  $\cos \phi > 0$ , while for  $\cos \phi \leq 0$  and  $-\frac{M(\nu)}{B(\nu)} \leq \sin \phi \leq \frac{M(\nu)}{B(\nu)}$  and  $M(\nu) \leq B(\nu)$  this is a valid solution.

To conclude, based on the measured parameters  $M(\nu)$  and  $B(\nu)$ , two different regimes can be distinguished (see Table 3.1). The expectation value of the signal is calculated by averaging

Table 3.1: The two different regimes in which equation 3.14 is valid.

Regime	Signal	Phase
$M(\nu) \geq B(\nu)$	$S(\nu) = -B(\nu) \cos \phi + \sqrt{M(\nu)^2 - B(\nu)^2 \sin^2 \phi}$	$0 \leq \phi < 2\pi$
$M(\nu) < B(\nu)$	$S(\nu) = -B(\nu) \cos \phi + \sqrt{M(\nu)^2 - B(\nu)^2 \sin^2 \phi}$	$ \phi - \pi  \leq \arcsin\left(\frac{M(\nu)}{B(\nu)}\right)$
	$S(\nu) = -B(\nu) \cos \phi - \sqrt{M(\nu)^2 - B(\nu)^2 \sin^2 \phi}$	$ \phi - \pi  \leq \arcsin\left(\frac{M(\nu)}{B(\nu)}\right)$

the equations in Table 3.1 over all possible values of the unknown phase  $\phi$ . The two regimes will be investigated separately in the following subsections.

**Regime:**  $M(\nu) \geq B(\nu)$

When the measurement is larger than the background estimate the expectation value of signal is given by:

$$\begin{aligned} \langle S(\nu) \rangle &= \frac{1}{2\pi} \int_0^{2\pi} \left( -B(\nu) \cos \phi + M(\nu) \sqrt{1 - \frac{B(\nu)^2}{M(\nu)^2} \sin^2 \phi} \right) d\phi \\ &= \frac{M(\nu)}{2\pi} \int_0^{2\pi} \sqrt{1 - \frac{B(\nu)^2}{M(\nu)^2} \sin^2 \phi} d\phi = \frac{2M(\nu)}{\pi} \mathcal{E} \left( \frac{B(\nu)^2}{M(\nu)^2} \right) \end{aligned} \quad (3.20)$$

where  $\mathcal{E}()$  is the complete elliptic integral of the second kind. The reconstructed signal is set to the expectation value, given the measurement in the signal window and the background estimate. Therefore:

$$S_{\text{rec}}(\nu) = \frac{2M(\nu)}{\pi} \mathcal{E} \left( \frac{B(\nu)^2}{M(\nu)^2} \right). \quad (3.21)$$

The unknown phase contributes to the uncertainty on the reconstructed signal, which is estimated as the square root of the variance of the underlying probability distribution. In order to calculate this variance,  $\langle S(\nu)^2 \rangle$  needs to be evaluated:

$$\langle S(\nu)^2 \rangle = \frac{1}{2\pi} \int_0^{2\pi} \left( -B(\nu) \cos \phi + M(\nu) \sqrt{1 - \frac{B(\nu)^2}{M(\nu)^2} \sin^2 \phi} \right)^2 d\phi = M(\nu)^2. \quad (3.22)$$

Hence, the uncertainty on the signal estimator is:

$$\sigma_{\text{rec}} = \sqrt{\langle S(\nu)^2 \rangle - \langle S(\nu) \rangle^2} = \sqrt{M(\nu)^2 - \frac{4M(\nu)^2}{\pi^2} \mathcal{E}^2 \left( \frac{B(\nu)^2}{M(\nu)^2} \right)}. \quad (3.23)$$

The value of  $\sigma_{\text{rec}}$  is fairly constant. One gets  $\sigma_{\text{rec}} = 0.713 B(\nu)$  when  $M(\nu) = 2 B(\nu)$ , and  $\sigma_{\text{rec}} = 0.707 B(\nu)$  when  $M(\nu) = 10 B(\nu)$ . For all practical purposes,  $\sigma_{\text{rec}} = 0.71 B(\nu)$  is a very good estimate when  $M(\nu) > 1.5 B(\nu)$ . For  $M(\nu) = 1.1 B(\nu)$  the value increases to  $0.74 B(\nu)$ , and for  $M(\nu) = B(\nu)$  the uncertainty becomes  $\sigma_{\text{rec}} = 0.77 B(\nu)$ .

**Regime:**  $M(\nu) < B(\nu)$

In this case the expectation value for the signal will be the average over the two terms listed in Table 3.1:

$$\begin{aligned} \langle S(\nu) \rangle &= 0.5 \frac{1}{2 \arcsin \left( \frac{M(\nu)}{B(\nu)} \right)} \int_{\pi - \arcsin \left( \frac{M}{B} \right)}^{\pi + \arcsin \left( \frac{M}{B} \right)} \left( -B(\nu) \cos \phi + M(\nu) \sqrt{1 - \frac{B(\nu)^2}{M(\nu)^2} \sin^2 \phi} \right) d\phi \\ &\quad + 0.5 \frac{1}{2 \arcsin \left( \frac{M(\nu)}{B(\nu)} \right)} \int_{\pi - \arcsin \left( \frac{M}{B} \right)}^{\pi + \arcsin \left( \frac{M}{B} \right)} \left( -B(\nu) \cos \phi - M(\nu) \sqrt{1 - \frac{B(\nu)^2}{M(\nu)^2} \sin^2 \phi} \right) d\phi. \end{aligned} \quad (3.24)$$

The square-root terms nicely cancel and, again, the expectation value will be used as the reconstructed signal estimator:

$$S_{\text{rec}}(\nu) = \frac{M(\nu)}{\arcsin\left(\frac{M(\nu)}{B(\nu)}\right)}. \quad (3.25)$$

The uncertainty on the reconstructed signal due to the unknown phase is determined as the square root of the variance of the underlying probability distribution. Thus  $\langle S(\nu)^2 \rangle$  is calculated as:

$$\begin{aligned} \langle S(\nu)^2 \rangle &= 0.5 \frac{1}{2 \arcsin\left(\frac{M(\nu)}{B(\nu)}\right)} \int_{\pi - \arcsin\left(\frac{M}{B}\right)}^{\pi + \arcsin\left(\frac{M}{B}\right)} \left( -B(\nu) \cos \phi + M(\nu) \sqrt{1 - \frac{B(\nu)^2}{M(\nu)^2} \sin^2 \phi} \right)^2 d\phi \\ &\quad + 0.5 \frac{1}{2 \arcsin\left(\frac{M(\nu)}{B(\nu)}\right)} \int_{\pi - \arcsin\left(\frac{M}{B}\right)}^{\pi + \arcsin\left(\frac{M}{B}\right)} \left( -B(\nu) \cos \phi - M(\nu) \sqrt{1 - \frac{B(\nu)^2}{M^2} \sin^2 \phi} \right)^2 d\phi. \end{aligned} \quad (3.26)$$

Expanding the squares and summing the integrands:

$$\begin{aligned} \langle S(\nu)^2 \rangle &= \frac{1}{2 \arcsin\left(\frac{M(\nu)}{B(\nu)}\right)} \int_{\pi - \arcsin\left(\frac{M}{B}\right)}^{\pi + \arcsin\left(\frac{M}{B}\right)} (B(\nu)^2 \cos^2 \phi + M(\nu)^2 - B(\nu)^2 \sin^2 \phi) d\phi \\ &= \frac{1}{2 \arcsin\left(\frac{M(\nu)}{B(\nu)}\right)} \int_{\pi - \arcsin\left(\frac{M}{B}\right)}^{\pi + \arcsin\left(\frac{M}{B}\right)} (B(\nu)^2 \cos(2\phi) + M(\nu)^2) d\phi \\ &= M(\nu)^2 + \frac{\sin(2\pi + 2 \arcsin \frac{M(\nu)}{B}) - \sin(2\pi - 2 \arcsin \frac{M(\nu)}{(v)B})}{4 \arcsin\left(\frac{M(\nu)}{B(\nu)}\right)} \\ &= M(\nu)^2 + \frac{2 \sin(2 \arcsin\left(\frac{M(\nu)}{B(\nu)}\right))}{4 \arcsin\left(\frac{M(\nu)}{B(\nu)}\right)} = M(\nu)^2 + \frac{1}{\arcsin\left(\frac{M(\nu)}{B(\nu)}\right)} \frac{M(\nu)}{B(\nu)} \sqrt{1 - \frac{M(\nu)^2}{B(\nu)^2}}. \end{aligned} \quad (3.27)$$

Hence, the uncertainty on the reconstructed signal due to the unknown phase is:

$$\sigma_{\text{rec}} = \sqrt{M(\nu)^2 + \frac{1}{\arcsin\left(\frac{M(\nu)}{B(\nu)}\right)} \frac{M(\nu)}{B(\nu)} \sqrt{1 - \frac{M(\nu)^2}{B(\nu)^2}} - \left( \frac{M(\nu)}{\arcsin\left(\frac{M(\nu)}{B(\nu)}\right)} \right)^2}. \quad (3.28)$$

### 3.2.1 Adding noise to the measurement

So far, a system with an unknown signal that is added to a constant background with an unknown relative phase has been considered. The term "constant" background refers to a signal that is constant on the timescale of several microseconds (e.g. due to Galactic sources or the local environment). Here, noise is defined as a fast fluctuating variation of this background. Including these variations allows to consider additional effects, such as thermal noise in the electronics. In the following, this effect will be estimated in both regimes defined above.

**Regime:  $M(\nu) \geq B(\nu)$** 

The reconstructed signal estimator is given by Eq. 3.21. The presence of noise generates an uncertainty in the measurement  $M(\nu)$ , denoted as  $\sigma_M$ , and it also introduces an uncertainty on the knowledge of the expectation value of the background  $B(\nu)$ ,  $\sigma_B$ . Assuming  $\sigma_M$  and  $\sigma_B$  are independent, the uncertainty due to the noise can be calculated as:

$$\begin{aligned}\sigma_N^2 &= \left( \frac{dS_{\text{rec}}(\nu)}{dM(\nu)} \right)^2 \sigma_M^2 + \left( \frac{dS_{\text{rec}}(\nu)}{dB(\nu)} \right)^2 \sigma_B^2 \\ &= \left( \frac{2\mathcal{E}\left(\frac{B(\nu)^2}{M(\nu)^2}\right)}{\pi} + \frac{2M(\nu)}{\pi} \frac{\mathcal{K}\left(\frac{B(\nu)^2}{M(\nu)^2}\right) - \mathcal{E}\left(\frac{B(\nu)^2}{M(\nu)^2}\right)}{M(\nu)} \right)^2 \sigma_M^2 + \frac{4M(\nu)^2}{\pi^2} \left( \frac{\mathcal{E}\left(\frac{B(\nu)^2}{M(\nu)^2}\right) - \mathcal{K}\left(\frac{B(\nu)^2}{M(\nu)^2}\right)}{B(\nu)} \right)^2 \sigma_B^2 \\ &= \frac{4\mathcal{K}^2\left(\frac{B(\nu)^2}{M(\nu)^2}\right)}{\pi^2} \sigma_M^2 + \frac{4M(\nu)^2}{B(\nu)^2 \pi^2} \left( \mathcal{E}\left(\frac{B(\nu)^2}{M(\nu)^2}\right) - \mathcal{K}\left(\frac{B(\nu)^2}{M(\nu)^2}\right) \right)^2 \sigma_B^2,\end{aligned}\tag{3.29}$$

where  $\mathcal{K}()$  is the complete elliptic integral of the first kind. The first term is close to  $1 \sigma_M^2$  for large signals and blows up when  $M(\nu) \simeq B(\nu)$ . For  $M(\nu) = 1.1 B(\nu)$  its value is  $2.2 \sigma_M^2$ , dropping to  $1.3 \sigma_M^2$  for  $M(\nu) = 1.5 B(\nu)$  and  $1.15 \sigma_M^2$  for  $M(\nu) = 2 B(\nu)$ . Above  $M(\nu) = 3.5 B(\nu)$  its value is within 5% of  $1 \sigma_M^2$ . The second term is close to zero for large signals and is infinite for  $M(\nu) = B(\nu)$ , for  $M(\nu) = 1.1 B(\nu)$  its value is only  $0.7 \sigma_B^2$ , dropping to  $0.17 \sigma_B^2$  for  $M(\nu) = 1.5 B(\nu)$  and  $0.08 \sigma_B^2$  for  $M(\nu) = 2 B(\nu)$ . For all practical purposes this term can be neglected.

Combining the results from this section with the uncertainty due to the unknown phase derived earlier, in good approximation for  $M(\nu) > 2 B(\nu)$  the uncertainty in the reconstructed signal can be written as:

$$\sigma_{\text{rec}} = \sqrt{(0.71 B(\nu))^2 + \sigma_M^2}.\tag{3.30}$$

**Regime:  $M(\nu) < B(\nu)$** 

The reconstructed signal is given by Eq. 3.25. Again, the noise causes an uncertainty in the measurement  $M(\nu)$  and on the knowledge of the background expectation value  $B(\nu)$ , respectively  $\sigma_M$  and  $\sigma_B$ . Assuming the two uncertainties to be independent, the uncertainty introduced by the noise is calculated as:

$$\begin{aligned}\sigma_N^2 &= \left( \frac{dS_{\text{rec}}(\nu)}{dM(\nu)} \right)^2 \sigma_M^2 + \left( \frac{dS_{\text{rec}}(\nu)}{dB(\nu)} \right)^2 \sigma_B^2 \\ &= \left( \frac{1}{\arcsin\left(\frac{M(\nu)}{B(\nu)}\right)} - \frac{M(\nu)}{B(\nu) \sqrt{1 - \frac{M^2(\nu)}{B^2(\nu)}} \left( \arcsin\left(\frac{M(\nu)}{B(\nu)}\right) \right)^2} \right)^2 \sigma_M^2 + \left( \frac{M^2(\nu)}{B^2(\nu) \sqrt{1 - \frac{M^2(\nu)}{B^2(\nu)}} \left( \arcsin\left(\frac{M(\nu)}{B(\nu)}\right) \right)^2} \right)^2 \sigma_B^2.\end{aligned}\tag{3.31}$$

The first term approaches 0 for  $M(\nu) = 0$ . It is  $0.01 \sigma_M^2$  for  $M(\nu) = 0.1 B(\nu)$ , and  $0.04 \sigma_M^2$  for  $M(\nu) = 0.5 B(\nu)$ . It goes to infinite for  $M(\nu) = B(\nu)$ , reaching  $7.5 \sigma_M^2$  for  $M(\nu) = 0.99 B(\nu)$ . The second term approaches 1 for  $M(\nu) = 0$ , and also explodes for  $M(\nu) = B(\nu)$ . It is  $1.003 \sigma_B^2$  for  $M(\nu) = 0.1 B(\nu)$  and  $1.11 \sigma_B^2$  for  $M(\nu) = 0.5 B(\nu)$ . It reaches  $11.6 \sigma_B^2$  for  $M(\nu) = 0.99 B(\nu)$ . The standard uncertainty propagation used in this section, and in the previous, does not take into account that parameters are bound. This is especially relevant in the regime discussed here. This causes an overestimate of the uncertainty calculated. Given that in this regime mostly

$0 < S(\nu) < 2B(\nu)$ , the maximal uncertainty on the reconstructed value of  $S(\nu)$  is  $2B(\nu)/\sqrt{12}$ . A toy Monte Carlo is used to properly take these effects into account.

### 3.2.2 Toy Monte Carlo

A toy Monte Carlo is used in order to verify the calculations made, as well as to evaluate the effect of boundaries and limited ranges of parameters that influence the final uncertainties.

The simulation code provides a random phase between the true signal and background and creates a measurement in the signal window according to Eq. 3.13. It then applies Gaussian electronics noise to the measurement and background, while taking into account that the background is determined from several samples. In the following several standard situations will be described.

#### No electronics noise added

In this first situation the unknown phase between signal and background is the only source for uncertainties ( $\sigma_N = 0$ ). This situation allows to verify the code as well as the calculations presented above. The expectation value for the signal and its uncertainty are described respectively by Eqs. 3.21 and 3.25, and Eqs. 3.23 and 3.28. Figure 3.9 shows a comparison between the analytical calculation and the simulation of the average true signal and the standard deviation as a function of the measured signal. In this comparison the background level is set to 1. When the measurement is close to the background, there are small discrepancies between the averaged signal simulated and the result of the calculation. A similar discrepancy appears in the comparison of the calculated and simulated standard deviation: the calculated value describes the true uncertainty very well, except when the measurement is close to the background value. In that case the calculation overestimates the width of the distribution of the true signals.

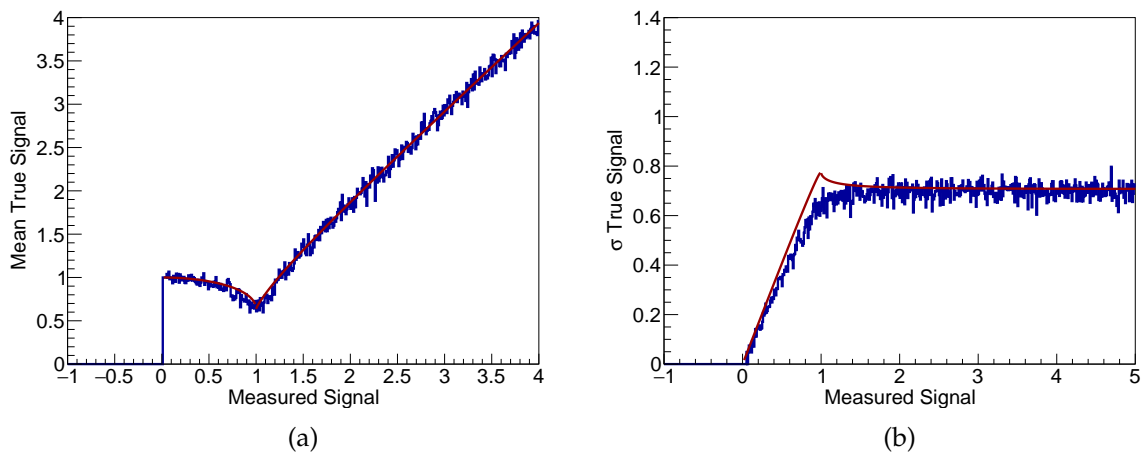


Figure 3.9: (a) Mean true signal (blue) as a function of the measurement value. The background level is set to 1. The red line shows the derived formulae Eqs. 3.21 and 3.25. (b) Standard deviation of true signal (blue) as a function of the measurement value. The background level is set to 1. The red line shows the derived formulae Eqs. 3.23 and 3.28.

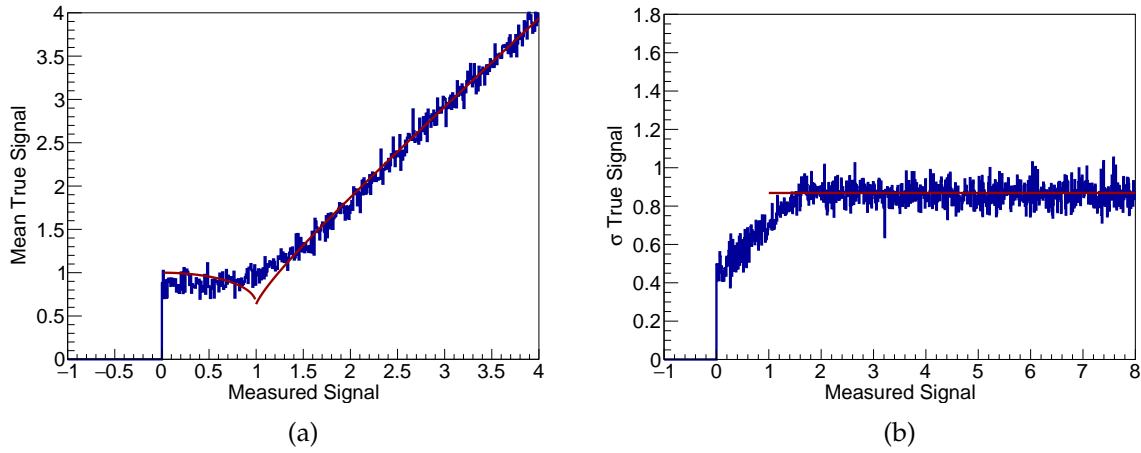


Figure 3.10: (a) Mean true signal (blue) as a function of the measurement value. The true background level is set to 1. Gaussian noise of 0.5 units is added to mimic electronics noise. The red line shows the derived formulae in 3.2.1. (b) Standard deviation of the true values (blue) as a function of the measurement value. The background level is set to 1. Gaussian noise of 0.5 units is added to mimic electronics noise. The red line shows the derived formula  $\sigma_{\text{rec}} = \sqrt{(0.71B)^2 + \sigma_M^2}$  at larger measured signals. For small measured values (below 2 times the true background) the analytical calculation does not describe the simulations and it is not shown in the plot.

### Added noise equal to 0.5 times the background. Background 49 times sampled

The situation described here is close to reality. In Figure 3.10 the average true signal as a function of the measurement is compared to the calculations. The mean true signal is close to the expectation value, but for measured values close to or below the true background, it levels out at a value of 0.88 times this background value. Contrary to the previous situation, for small measured values the uncertainty is not well described by the obtained formulae, and therefore it is not shown in the plot. The inconsistency is mainly due to the fact that the boundaries of possible signal values in reality depend on the fluctuations of the background (i.e. the noise), an effect not properly taken into account in the calculation.

In conclusion, the analytical calculation can not be used for signal values below two times the background level. Instead, the expectation value of the reconstructed signal amplitude will be estimated using the toy Monte Carlo and averaging over 100 signals with a random phase. While this is done for all signal values for the analyses in this thesis, the analysis software could be sped up considerably for measured signal values above two times the background level.

### 3.2.3 Validation of the background subtraction

The signal ( $S_{\text{rec}}(\nu)$ ) is estimated in the east-west and north-south components of the electric field in a time window of 400 ns around the peak of the Hilbert envelope of the time trace ( $M(\nu)$ ). The expectation value of the average background ( $B(\nu)$ ) is obtained by averaging over  $N$  independent background windows of 400 ns from the same electric field trace leaving at least 800 ns between the signal windows and the closest background window. The maximum number of windows considered in the average is 49, this way the procedure does not become

unnecessary time-consuming. For each frequency bin, the expectation value of the signal is defined as the average of 100 signals that satisfy Eq. 3.13, where the phase is generated from a uniform distribution in the interval  $[0, 2\pi]$ . The amplitude value of the measurement and of the expectation value of the average background are allowed to fluctuate within their own uncertainties, respectively:  $\sigma_M$  is assumed to be equal to the amplitude fluctuations in the  $N$  background windows, under the assumption that the fluctuations in the background windows are uncorrelated, and  $\sigma_B = \sigma_M / \sqrt{N}$  is the uncertainty on the expectation value of the average background amplitude. This should not be confused with the uncertainty in a single background interval, which is  $\sigma_M$ .

The background correction and the calculation of the uncertainties is verified with CoREAS simulations and real background traces selected from the AERA periodically triggered data. The amplitude of the signal reconstructed without adding background is used as reference,  $S_{\text{true}}(\nu) = M(\nu)$ , and compared to  $S_{\text{rec}}(\nu) \pm \sigma_{\text{rec}}$ . In both cases the detector simulation is included.

The stations included in the analysis are selected according to the standard definition of the signal to noise calculated as the ratio between the maximum of the Hilbert envelope of the electric field projected on the horizontal plane, and the root mean square of a  $30\mu\text{s}$  noise windows of the same electric field trace:

$$\text{SNR}_{\text{square}} = \frac{(\max \hat{E}_{\text{mag}})^2}{\text{rms}_{\text{noise}}^2}, \quad (3.32)$$

and it is required that  $\text{SNR}_{\text{square}} > 10$ . No additional quality cuts are applied. Whether the reconstructed amplitudes and the uncertainties assigned to them are correct is tested by comparing the pull distribution with the expectation of a standard Gaussian. For each station  $i$  the pull value is defined as the difference between the average reconstructed amplitude and the true one normalized by the uncertainty on the averaged reconstructed amplitude:

$$\text{pull}_i = \frac{\sum_{\nu_{\min}}^{\nu_{\max}} S_{\text{rec},i}(\nu) - \sum_{\nu_{\min}}^{\nu_{\max}} S_{\text{true},i}(\nu)}{\sqrt{\sum_{\nu_{\min}}^{\nu_{\max}} \sigma_{\text{rec},i}^2(\nu)}}. \quad (3.33)$$

Figure 3.11 shows the mean value and sigma of the Gaussian fit of the pull distribution as a function of:

$$\text{SNR}_{\text{amp}} = \frac{\sum_{\nu_{\min}}^{\nu_{\max}} S_{\text{rec}}(\nu)}{\sum_{\nu_{\min}}^{\nu_{\max}} B(\nu)}. \quad (3.34)$$

where each value displayed is composed of 500 entries. It is clear that when the reconstructed signal is below two times the measured background level the signal reconstructed overestimates the true signal. This effect is likely due to fact that a flat prior has been used in the toy Monte Carlo for the signal distribution. In reality, the signal is more likely to be small. This effect is large when the signal is of the same order as the uncertainty on the measurement. The difference between the north-south and east-west polarizations is introduced by the cut on  $\text{SNR}_{\text{square}}$ , which, even if evaluated in the horizontal plane, is dominated by the east-west component where the signal is, in general, stronger.



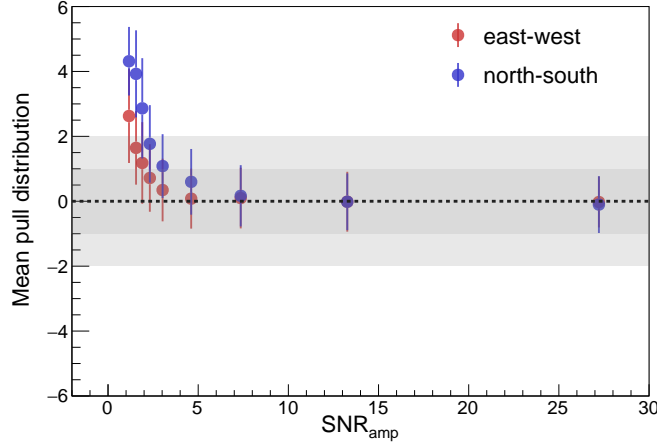


Figure 3.11: Mean value of the pull distribution of the difference between the antenna averaged reconstructed amplitude after adding real background and noise to the trace and the averaged *true* amplitude from the Monte Carlo signal trace, normalized by the reconstructed uncertainty. The error bar represents the sigma of the pull distribution. Each bin contains 500 entries.

### 3.3 Energy fluence estimation in presence of noise

The energy fluence  $f$  is the energy deposit per unit area of the electromagnetic radio pulse. It is defined as the integral (sum for a discrete sampling) of the absolute value of the Poynting vector, that, for plane waves, corresponds to the electric field amplitude squared. Therefore the energy fluence is given by:

$$f_{\text{rec}} = \epsilon_0 c \Delta t \left( \sum_{\nu_{\min}}^{\nu_{\max}} S_{\text{rec}}^2(\nu) \right), \quad (3.35)$$

where  $\epsilon_0$  is the vacuum permittivity,  $c$  is the speed of light,  $\Delta t$  is the size of the time bin and  $S_{\text{rec}}(\nu)$  is the expectation value of the electric field amplitude of the signal at each station. The energy fluence is expressed in units of  $\text{eV}/\text{m}^2$ .

However, from the study shown in Figure 3.11 it follows that for low signal stations ( $\text{SNR}_{\text{amp}} < 2$ ) the energy fluence is overestimated. Rejecting these stations significantly reduces the number of events with more than 5 signal stations and zenith angle below  $60^\circ$ . Therefore it is necessary to correct the value of the energy fluence by subtracting the background level. The energy fluence of the background,  $f_{\text{B}}$ , is calculated in a background window of 400 ns, not included in the estimation of the average background, and the expectation value of the electric field amplitude is derived following the same procedure as for the reconstructed signal. A descriptive function for the shift is evaluated using CoREAS simulations: the true energy fluence of the Monte Carlo electric field trace is compared to the corresponding reconstructed energy fluence when a real background trace is added to the electric field trace. The detector simulation is included in both cases. The correction of the energy fluence is made using the following function:

$$\hat{f} = f_{\text{rec}} + F \left( \frac{f_{\text{rec}}}{f_{\text{B}}} \right) f_{\text{B}} = f_{\text{rec}} + a \cdot \left\{ -\frac{\pi}{2} + \arctan \left[ b \left( \frac{f_{\text{rec}}}{f_{\text{B}}} - 1 \right) + \tan \left( \frac{1}{a} + \frac{\pi}{2} \right) \right] \right\} \cdot f_{\text{B}}, \quad (3.36)$$

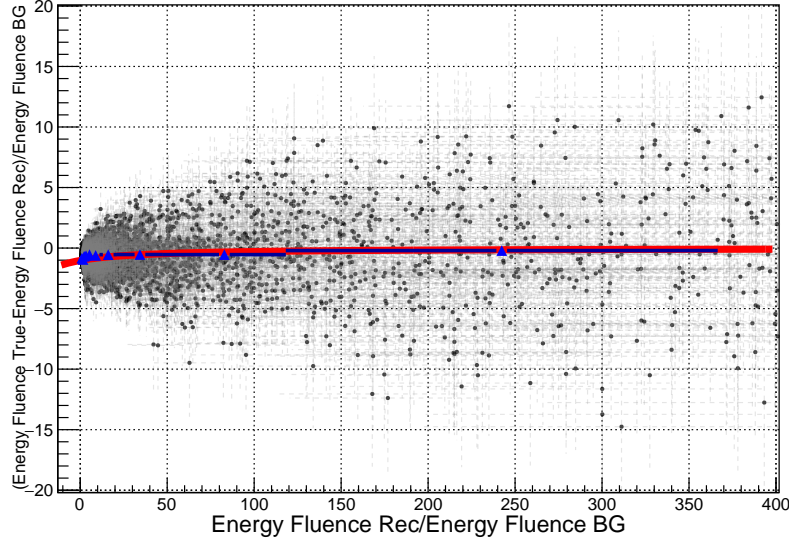


Figure 3.12: Rectification of the signal energy fluence. The red line shows the formula in Eq. 3.36, while the blue triangles are the mean values of the  $y$ -axis calculated each 500 entries.

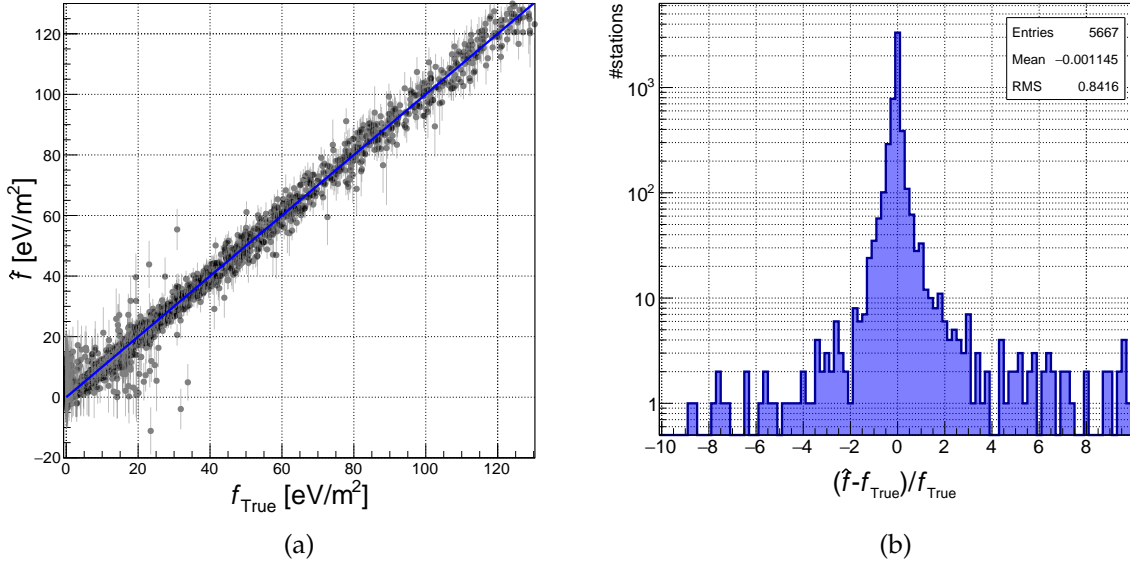


Figure 3.13: Comparison between the shifted energy fluence  $\hat{f}$  obtained with Eq. 3.36, and the *true* energy fluence value calculated using the Monte Carlo electric field trace without adding any background. The blue line marks the diagonal.

with  $a = 4$  and  $b = 0.3$ . The function  $F\left(\frac{f_{\text{rec}}}{f_{\text{B}}}\right)$ , shown in Figure 3.12, asymptotically approaches zero as the ratio between the energy fluence of the signal and the background increases. The comparison between the *shifted* and true energy fluence in the east-west polarization is shown in Figure 3.13. The stations included in the analysis fulfill the quality cut  $\text{SNR}_{\text{square}} > 10$ .

The uncertainty on the energy fluence is propagated from the uncertainty on the amplitude of the electric field. Given that the phases of the backgrounds in different frequency bins are not

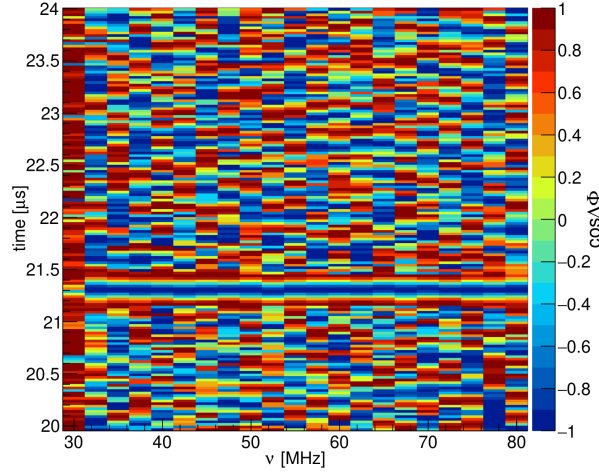


Figure 3.14: Cosine of the phase difference between two adjacent 400 ns windows from the east-west electric field trace of the Butterfly station 74 (event number: 101121.13606). The phases of the background are random, while the phases in the presence of the signal (around 21.3 μs) are fully correlated.

correlated, as is clear from Figure 3.14, the uncertainty on the energy fluence is:

$$\sigma_{f_{\text{rec}}} = \epsilon_0 c \Delta t \sqrt{\left(2 \cdot \sum_{\nu_{\text{min}}}^{\nu_{\text{max}}} S_{\text{rec}}(\nu) \sigma_{\text{rec}}\right)^2}, \quad (3.37)$$

and therefore

$$\sigma_{\hat{f}} = \sqrt{\left[\left(F\left(\frac{f_{\text{rec}}}{f_{\text{B}}}\right) \cdot \sigma_{f_{\text{B}}}\right)^2 + \sigma_{f_{\text{rec}}}^2\right]}, \quad (3.38)$$

where  $\sigma_{f_{\text{B}}}$  is also calculated using Eq. 3.37.

Figure 3.15 shows the mean and standard deviation of the pull values calculated as the difference between  $\hat{f}$  and  $f_{\text{true}}$  divided by the estimated uncertainty. Each bin contains 500 entries. To investigate possible systematic effects the frequency range has been divided in bins of 10 MHz. The trend is the same in all frequency bins: the energy fluence is slightly underestimated, and its uncertainty overestimated, for low  $f_{\text{rec}}/f_{\text{B}}$ . This can probably be improved by changing the standard signal to noise definition with one more sensitive to the power content of the signal.

### 3.4 The *Offline* framework

*Offline* is a modular software framework specifically created for analyzing data from the Pierre Auger Observatory. The framework was originally designed for analyzing fluorescence and surface detector data and simulations, and has been extended with functionality for data from the radio detectors [60].

The *Offline* framework makes a clear separation between the internal representation of the

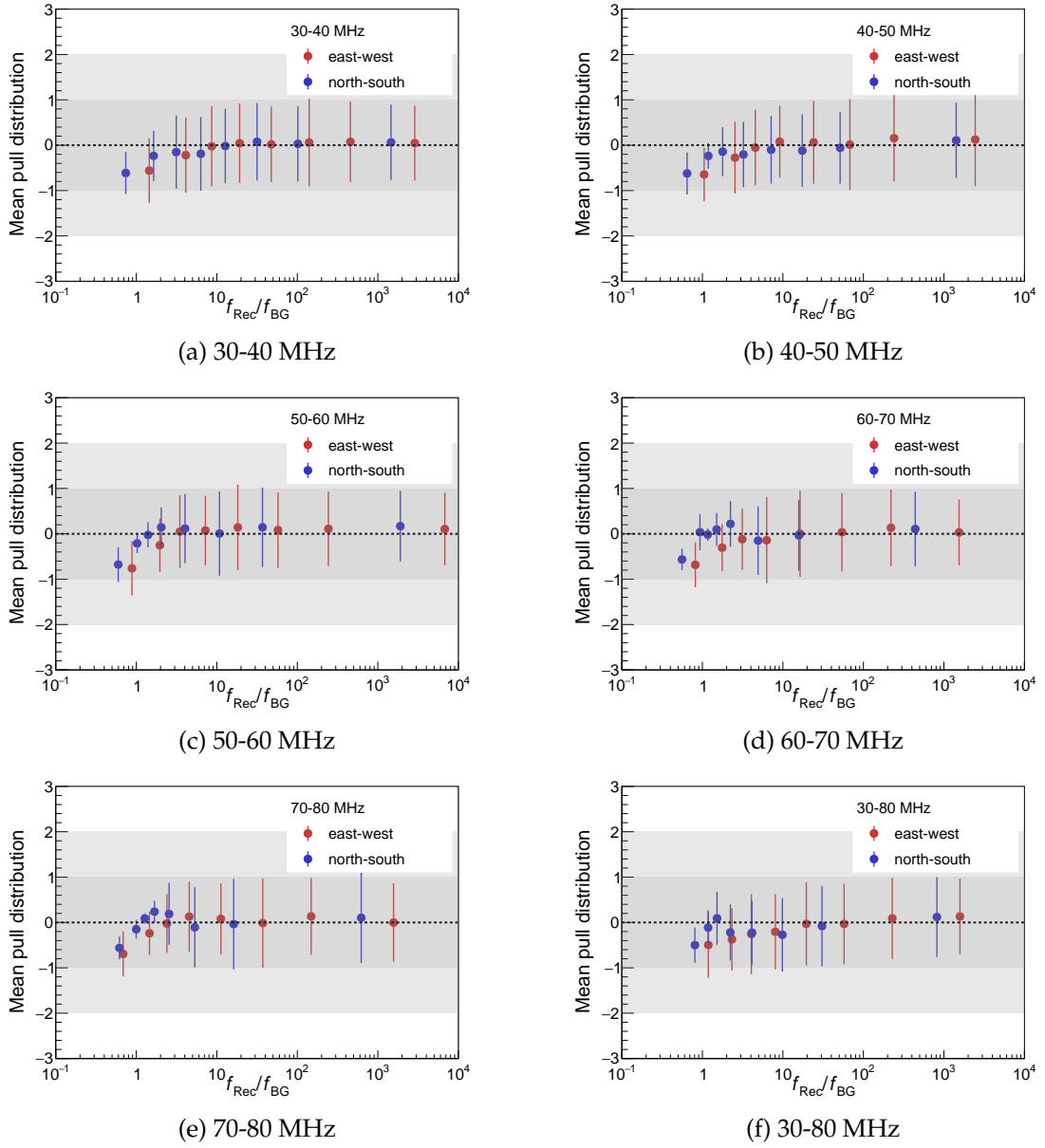


Figure 3.15: Mean of the pull distributions of the difference between  $\hat{f}$  and  $f_{\text{true}}$  divided by the estimated uncertainty in different frequency bins, and in the full 30-80 MHz range. The error bar indicates the standard deviation. Each bin contains 500 entries.

*Detector* and the *Event*. The *Detector* provides all the relevant detector information, such as the position of the stations and the hardware associated with each of them. All the data related to a specific event, such as ADC traces or a reconstructed shower axis are part of the *Event* data structure. These two independent interfaces are connected via analysis *Modules*. A sketch of the general structure of *Offline* is shown in Figure 3.16. The individual modules do not communicate directly with each other, but share data through the *Event* data structures. To perform an analysis within *Offline*, analysis modules are chained in sequence and each of them is configured individually through *xml* files.

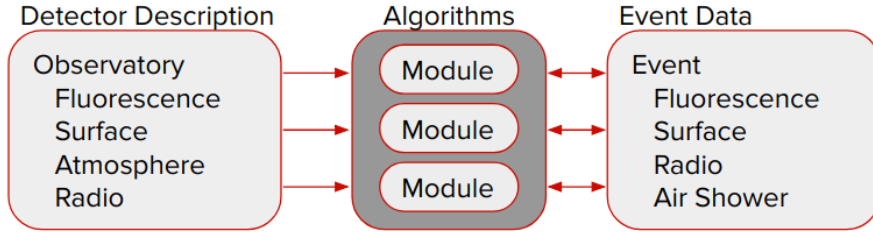


Figure 3.16: General structure of the *Offline* framework. Reconstruction tasks are performed by modules. Modules read information from the detector description and/or the event, process the information, and write the results back into the event. Figure adapted from [64].

The radio data structure is divided into *Channels* and *Stations*. The first level is defined by the individual antenna channels and contains the recorded data from each antenna in ADC counts. The *Stations* level is defined by the physical electric field reconstructed at a given detector location in the field, stored as a time-series of three-dimensional vectors. The transition between the two levels is performed by applying the characteristics of the antennas associated to each of the channels.

A time-dependent detector configuration is stored in a database that can be accessed from within *Offline*. The same is true for environmental data such as temperature and atmospheric conditions. In addition, during the reconstruction it is possible to reject the malfunctioning radio stations listed in the *bad period* database.

#### 3.4.1 Radio event reconstruction in *Offline*

The `RdObserver` sequence is the standard reconstruction for the hybrid data created by merging RD, SD and, possibly, FD and AMIGA data files. By default, the SD arrival direction is used to unfold the antenna model and reconstruct the electric field. Therefore a successful reconstructed SD event is required to begin the radio event reconstruction, while the FD event is optional in case the event contains such information.

The radio reconstruction sequence can be split into two separate main sections. The first one is dedicated to the incorporation of hardware responses as well as timing corrections (channel level), whereas the second part consists of the actual signal and directional reconstruction (station level) and eventually the estimation of energy and  $X_{\text{max}}$ . In case of radio data, the first information listed is the series of ADC counts recorded by the radio detector channels and converted into their voltage equivalent. Next, the voltage trace is corrected for the frequency-dependent response of the analog signal chain, consisting of filters, cables and the two amplification stages, to obtain the voltage induced in the antenna by the electric field pulse. A

beacon transmitter is used as reference to correct for the timing drift between the GPS clocks of each of the radio setups. It is installed at the Coihueco FD building and transmits sine waves at several defined frequencies. However, for the analysis of air showers, the beacon signals are background and therefore these frequencies are digitally suppressed in the spectrum during the analysis. In addition, specific algorithms are applied to the spectrum to filter the signal for additional RFI disturbances. Finally, the electric field vector is reconstructed, and, after the background subtraction, the signal and its uncertainty are estimated. An overview of the outcome of the different reconstruction steps is presented in Figure 3.17.

The pipeline begins differently when a CoREAS simulation is used as input file. At first each simulated radio pulse is associated to a real detector station, furthermore a sequence to perform a detector simulation is included before the reconstruction described above. It is also possible to add measured noise to the ADC traces. The noise can be chosen randomly or according to the timestamp within the periodically trigger events. *Offline* is able to handle different geometries of the antenna grid, in particular the simulations analyzed in this thesis have been reconstructed using the standard application `RdReconstructStarshapedStationPattern`, dedicated to the simulation of stations aligned in a star-shaped grid in the shower plane.

Data and simulated events presented in this work have been reconstructed and analyzed using *Offline* rev-33196.

### 3.4.2 Signal reconstruction module

The procedure to estimate the radio signal from a noisy measurement has been integrated in the *Offline* module `RdStationSignalReconstructorWithBgSubtraction`. Starting from the electric field time trace of the east-west and north-south polarization, the expectation value of the cosmic ray signal is evaluated following the procedure described in Section 3.2.3.  $S_{\text{rec,EW}}(\nu)$  and  $S_{\text{rec,NS}}(\nu)$  are then used to infer the vertical component of the electric field, the total electric field, and its components in the shower plane,  $(\vec{v}, \vec{v} \times \vec{B}, \vec{v} \times (\vec{v} \times \vec{B}))$ , using the formulae described in Section 3.1. Lastly, the energy fluence and its uncertainty are calculated, applying respectively Eq. 3.36 and Eq. 3.38, and written in the parameter storage space to be available for the following analysis steps.

The module can be used to reconstruct both simulation and data events. The background subtraction must be skipped when simulations without background are reconstructed, this is done by setting to 0 the boolean variable `BackGroundSubtraction` in the bootstrap file.

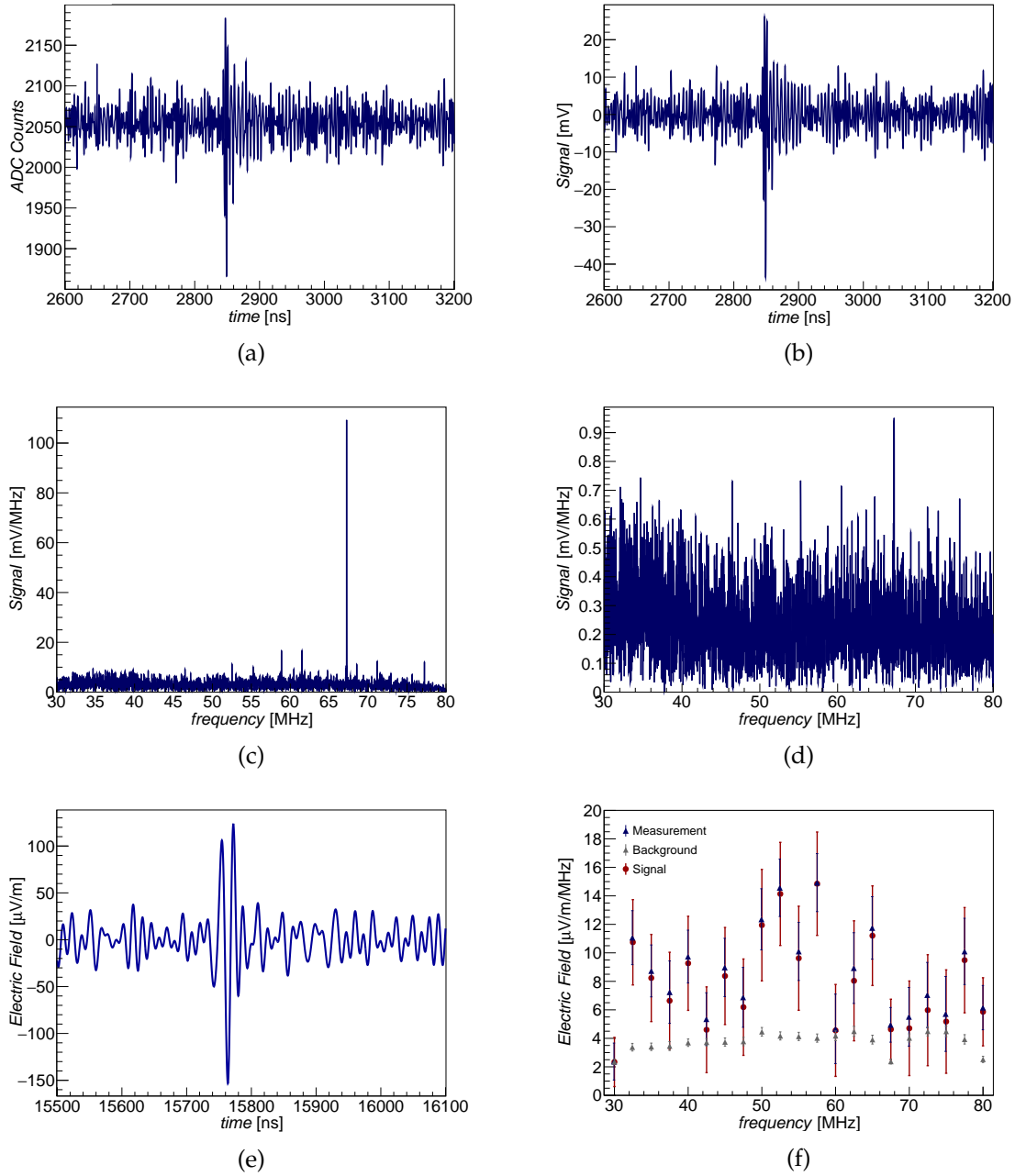


Figure 3.17: Different reconstruction steps demonstrated on an example event measured on January 4, 2016 ( $E=0.43$  EeV,  $\theta=13^\circ$  and  $\phi=354^\circ$ ). The data in the plots are from the east-west channel of the Butterfly station 54. (a) Raw ADC time trace. (b) Voltage trace after unfolding for the antenna response pattern. (c) Frequency spectrum of (b). Narrowband noise sources are visible. (d) Frequency spectrum after noise filtering using the sine-wave suppression in the time domain. (e) Electric field trace obtained by unfolding the antenna response. (f) Frequency spectrum of electric field. The amplitude values of the measurement are shown in blue, the amplitude of the expectation values of the average background in gray, the amplitude values of the signal in red.



## 4 Calibration of the AERA stations

Interpreting correctly the data collected by the AERA stations requires a good understanding of the antenna response. As described in Section 3.1, the electric field is reconstructed applying the response pattern according to the antenna type and the response of the LNA, hence, without taking variations between stations into account. These need to be evaluated for example using the Galactic radio emission. In addition, an absolute and frequency-dependent calibration is necessary in order to correct for possible deviations in the antenna response pattern.

This chapter will summarize the strategies for the relative and absolute antenna calibration and the results obtained for LPDA and Butterfly stations equipped with KIT-BUW electronics. The analyses are based on the previous work described in Refs. [37] and [40].

### 4.1 Relative amplitude calibration

Relative amplitude calibration constants are used to correct the gains of individual stations as a function of frequency. These constants are calculated under the assumption that all stations are exposed to the same radio emission from the galaxy at any given instance of time.

The received radio emission of the galaxy varies in time with a period of a local sidereal day (approximately four minutes shorter than a solar day), and therefore the time scale is binned as fraction of local sidereal time, LST. The amplitude at each individual antenna,  $A_i(\nu, t_k)$ , is expressed as a function of the frequency  $\nu$  and of the LST bins  $t_k$ .

For each station, frequency-dependent constants are calculated by averaging the ratio between the amplitude at each individual antenna and the mean amplitude of all antennas over all LST bins:

$$c_{i,\nu} = \frac{1}{m} \sum_{k=1}^m \frac{A_i(\nu, t_k)}{\overline{A}(\nu, t_k)}. \quad (4.1)$$

The constants are calculated independently for the two channels of each station type, to take into account the different exposure to the galaxy and the asymmetric antenna response of the Butterfly stations.

The data set used for the relative calibration consists of periodically triggered events from 16 December 2017 to 16 January 2018. This period has been chosen because, according to the *bad period* database, a large fraction of stations was active, and LPDAs with Nikhef electronics were already converted to KIT-BUW electronics, therefore also these stations have been calibrated. All the 86 Butterfly stations with KIT-BUW electronics active in the field are included in the calibration, while four out of the 22 LPDA stations are not present in the period chosen, namely AERA-9, AERA-17, AERA-18, AERA-22. Data of these four stations are known to have spikes in the

frequency spectrum and therefore these stations are often flagged as bad stations and rejected in the analysis.

In order to reduce the data volume per event the time series of each channel is clipped to 1024 samples ( $5.7 \mu\text{s}$ ), which correspond to 285 bins in the range of 30-80 MHz. Furthermore a Hann window is applied to the first and the last 10% of the time trace to create periodicity before applying a Fourier transformation.

Periodically triggered data provide a good measurement of the galactic background, but they also contain noise from the electronics of the stations and from sources producing radio frequency interference (RFI). Figure 4.1 shows the average amplitude for the east-west and the north-south antenna channel for all Butterfly stations. The presence of RFI makes it difficult to uncover the galaxy variation. RFI can be classified in two main classes: narrowband transmitters and broadband transient pulses. Narrowband RFI is emitted continuously by communication devices, but also the radio beacon installed by the AERA group produces narrowband lines, mainly visible in the north-south channel, while in the east-west polarization another known narrowband transmitter at 67 MHz is visible. Broadband RFI mostly originate from man-made sources. The strength of such a noise contribution depends on the distance between the station and the source, hence data affected by these sources introduce a bias in the calibration constants and need to be removed.

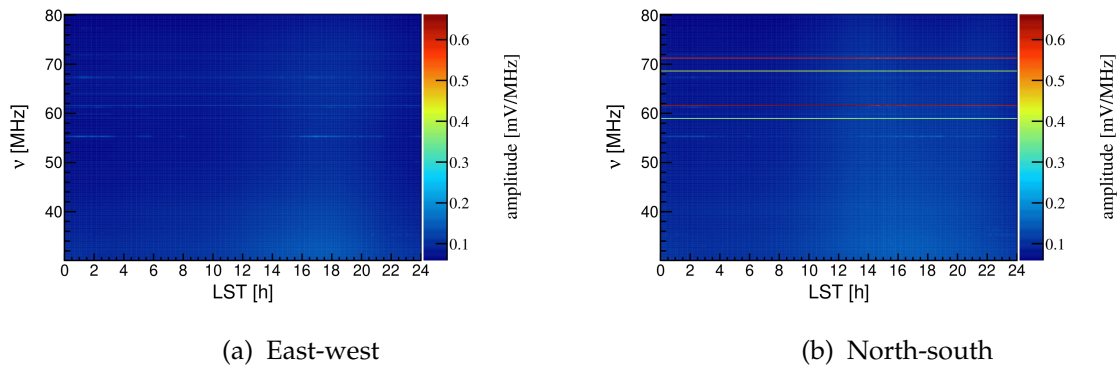


Figure 4.1: Average amplitude in the Butterfly antenna arms as a function of LST and frequency. The average is calculated using periodically triggered data recorded between 16 December 2017 and 16 January 2018. The amplitudes are averaged in bins of 10 minutes in LST. The galaxy variation is dwarfed by the presence of narrowband and broadband noise sources.

Narrowband RFI can be identified as peaks in average frequency spectra. The average spectra of the two channels of the Butterfly stations are shown in Figure 4.2. The frequency bands that need to be removed are listed in Table 4.1 for the Butterfly stations and Table 4.2 for the LPDAs.

Table 4.1: Frequency bands removed for the calculation of the relative amplitude calibration of the Butterfly stations.

EW	$\nu_{low}$ [MHz]	38.0	54.4	58.5	63.3	66.5	71.0
	$\nu_{up}$ [MHz]	40.6	56.2	62.0	64.6	69.0	72.5
NS	$\nu_{low}$ [MHz]	38.5	54.4	57.7	63.5	66.5	70.0
	$\nu_{up}$ [MHz]	41.5	56.2	62.0	64.5	69.0	72.6

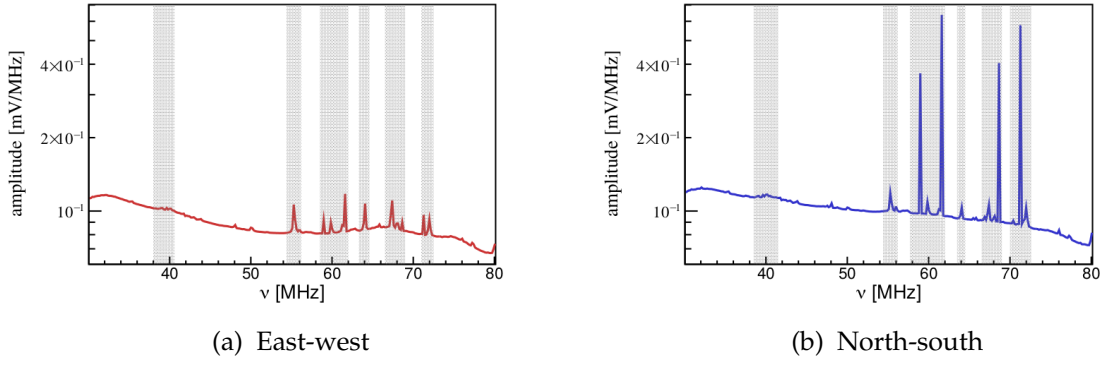


Figure 4.2: Average background frequency spectrum for both Butterfly antenna channels. The gray dashed bands indicates the narrowband RFI that needs to be removed from the calculation of the calibration constants.

Table 4.2: Frequency bands removed for the calculation of the relative amplitude calibration of the LPDA stations for both the east-west and north-south antenna channels.

$\nu_{low}$ [MHz]	38.6	54.8	58.5	60.9	63.7	66.7	70.7
$\nu_{up}$ [MHz]	40.4	56.3	60.3	62.0	64.4	69.0	72.5

Broadband RFI are excluded by setting a fixed threshold on the average spectral density. The averaged spectral density is defined as:

$$I = \frac{1}{n} \sqrt{\sum_{j=1}^n A^2(\nu_j)}, \quad (4.2)$$

where  $A(\nu_j)$  is the amplitude in frequency bin  $j$ , and  $n$  is the total number of frequency bins. The upper limit for the average spectral density is set to  $10^{-2}$  V for the Butterfly stations and to  $3 \cdot 10^{-3}$  V for the LPDAs. Figure 4.3 shows the average spectral density as measured in the east-west and the north-south antenna channels for Butterfly stations, broadband RFI are visible as clear outliers of these distributions.

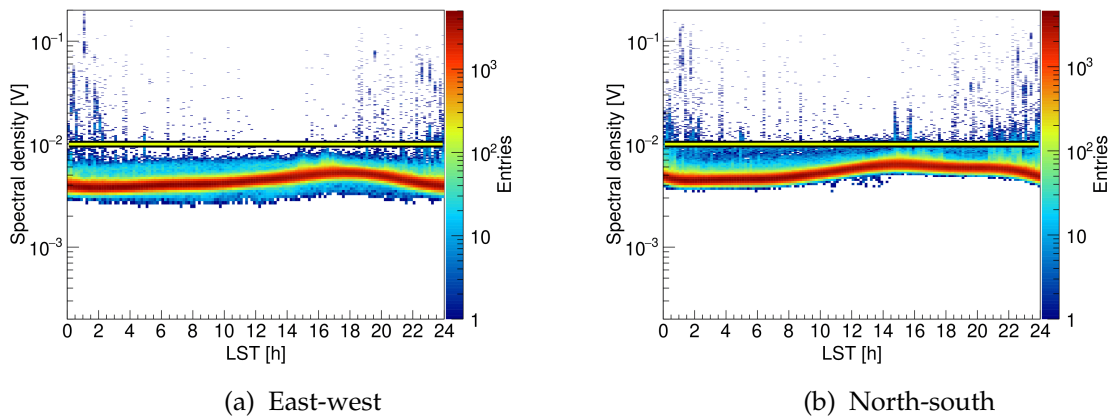


Figure 4.3: Average spectral density for the east-west and the north-south antenna channels as measured by the Butterfly stations. Broadband RFIs are outliers of these distributions and they are removed by setting a maximum threshold of  $10^{-2}$  V marked with the yellow line.

Furthermore, to prevent a bias in the calculation of the average spectrum, stations that show significant enhancements in their average spectrum are removed from the calculation of  $\bar{A}(\nu, t_k)$ . The enhancements are most likely related to hardware issues. Within the 86 Butterfly stations to be calibrated, 6 are not included in the average: AERA-28, AERA-31, AERA-37, AERA-47, AERA-120, and AERA-148. From the 18 LPDA stations in this analysis, 4 have been removed from the calculation of the average, namely AERA-1, AERA-7, AERA-11, and AERA-13.

After cleaning the spectrum, the variation due to the galaxy becomes much more pronounced, as visible in Figure 4.4. However, this technique leaves several gaps in the frequency range and, therefore, in the frequency-dependent calibration constants. The gaps are filled using a linear interpolation, which introduces an uncertainty comparable with the precision of the method itself, as was already shown in Ref. [40].

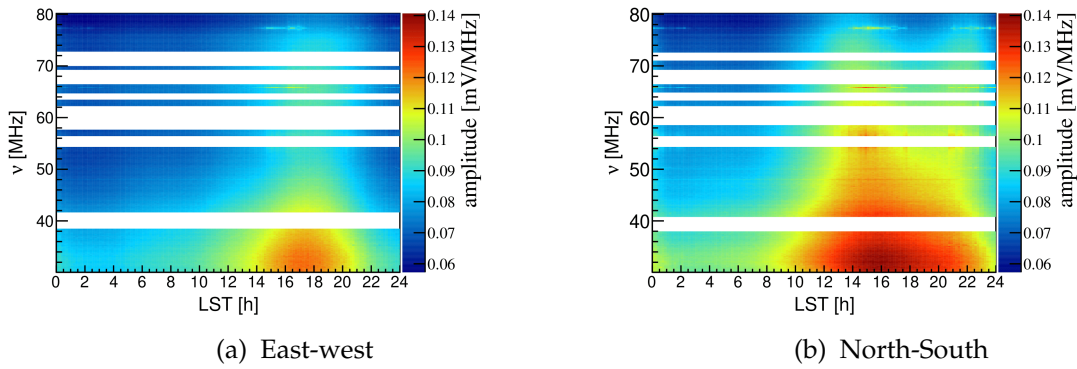


Figure 4.4: Average amplitude in both antenna arms as a function of LST and frequency. The average is calculated using periodically triggered data recorded by the Butterfly stations between 16 December 2017 and 16 January 2018. The white bands correspond to the narrowband peaks removed from the spectrum. The galaxy variation is clearly visible in both arms.

This strategy provides a gain correction factor for each individual antenna for each frequency bin in the 30-80 MHz range. The distribution of the gain factors is shown in Figure 4.5. The histogram on the left shows the calibration constants for 86 Butterfly stations, the tail above correction factors of 1.8 corresponds to the constants of station AERA-148, located at the edge of the 750 m grid and which is known to show anomalies in the frequency spectrum. Excluding this station, the  $\sigma$  of the two channels combined is 7.15%. The histogram for the LPDAs shows the calibration constants for the 18 stations included in the analysis. The high tail is created mainly by station AERA-20, while the hump below 1 corresponds to the calibration constants of station AERA-16. In this case all the stations are included in the calculation of the width of the two channels combined, which is 7%. The width of these two distributions gives an indication of the variation of the relative antenna responses averaged over the whole frequency range of interest.

In order to validate the calibration, the frequency-dependent correction factors have been applied to the data of an independent data taking period from 1 March to 31 March 2018. Once again it is important to clean the frequency spectrum from narrowband and broadband RFI before calculating the  $c_{i,\nu}$ . Figures 4.6 and 4.7 show the distribution for the validation data set respectively for Butterfly and LPDA stations. The calibration has a relevant effect on the relative antenna response: the width of the distribution changes from 8% to 4% for the Butterfly stations (again excluding the tail above 1.3 caused by station AERA-148), and from

9% to 2% for the LPDA stations. The second peak in LPDA distribution is caused by station AERA-15, while the first hump in the Butterfly distribution is caused by station AERA-50.

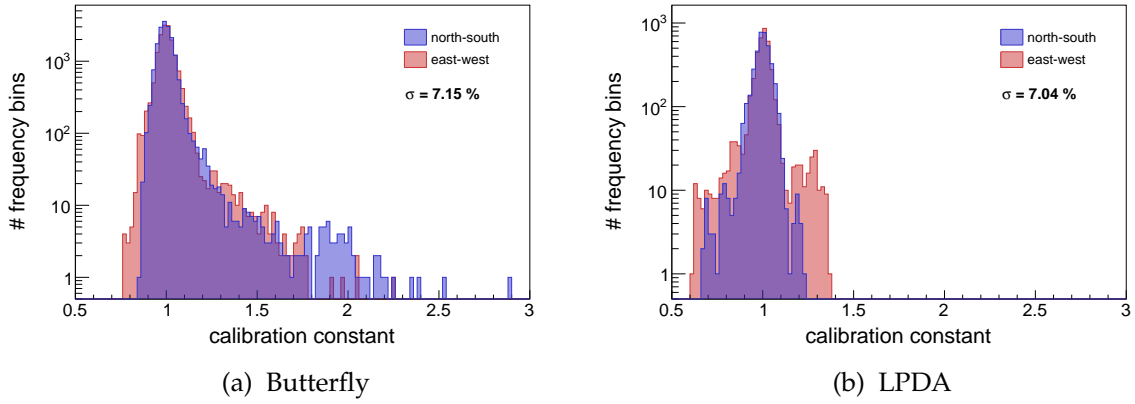


Figure 4.5: Frequency-dependent calibration constants for both antenna channels of Butterfly and LPDA stations using data obtained in the period from 16 December 2017 to 16 January 2018.

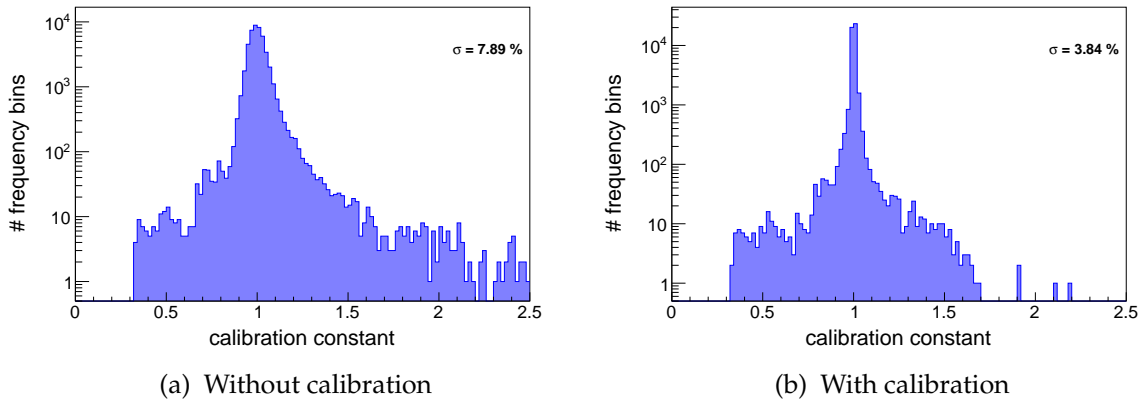


Figure 4.6: Antenna-to-antenna variation for Butterfly stations, without and with applying the calibration factors. The constants are calculated in the validation data set from 1 March and 31 March 2018.

In conclusion the relative calibration reduces the gain differences between the AERA stations and it is stable in time.

The calibration constants are generated within the *Offline* framework using the module `RdChannelGalacticConstantsGenerator`, in the example application `RdGalacticCalibration`. To apply the calibration constants to data events it is necessary to add the module `RdChannelGalacticBackgroundCalibrator` in the standard reconstruction pipeline. The calibration constants need to be calculated again after each hardware change and to be monitored in time to maintain a high quality.

#### 4.1.1 Implications for data analysis

In order to estimate the uncertainty on the electric field due to the hardware and mechanical differences between the antennas and different ground conditions for each event, the

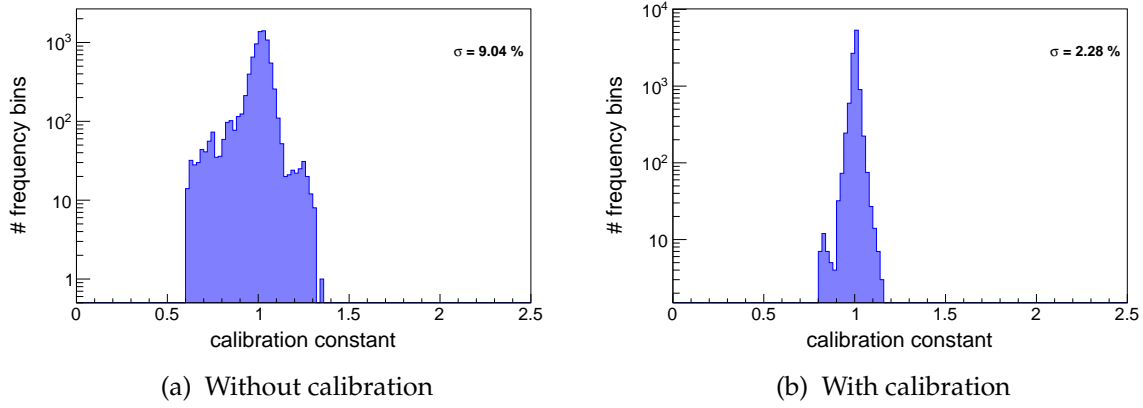


Figure 4.7: Antenna-to-antenna variation for LPDA stations without and with applying the correction factors. The constants are calculated in the validation data set from 1 March and 31 March 2018

amplitude at each antenna is compared with the average amplitude of all antennas in each frequency bin. The factor  $c_i$  is defined as:

$$c_i = \frac{1}{n} \sum_{j=v_1}^{v_n} \frac{A(v_i)}{\bar{A}_{v_i}}, \quad (4.3)$$

where  $n$  is the total number of frequency bins.

The uncertainty is determined using periodically triggered data reconstructed applying the relative calibration factors. Assuming that in a time window of 100 ns all stations are exposed to the same signal from the galaxy, the root mean square of  $c_i$  is a measurement of the antenna-to-antenna fluctuation for a single event. The average value of the root mean square of  $c_i$  over many events is used as a systematic uncertainty to the amplitude of the electric field. The uncertainty is estimated independently for Butterfly and LPDA antenna types and it is averaged over the north-south and east-west polarization. The data set used consists of seven non-consecutive days of periodically triggered data in January 2018. The results are shown in Figure 4.8 and indicate a 5% systematic uncertainty on the electric field amplitude introduced by the antenna-to-antenna fluctuations.

## 4.2 Absolute amplitude calibration

Inaccurate knowledge of the antenna response contributes to the hardware-based uncertainties in radio measurements of air showers. A precise calibration is therefore essential for a correct reconstruction of the air shower parameters. As already mentioned in Chapter 3, for each antenna type two versions of the antenna response pattern are available: the antenna responses have been simulated with the NEC-2 software, and the antenna responses have been measured in the field using radio emitters. For the Butterfly stations the two antenna patterns show a discrepancy at higher frequencies. An absolute calibration helps to compare the impact of the antenna response models on data reconstruction.

Similar to the relative calibration, it is possible to make use of the radio emission from the galaxy to perform an absolute calibration: a model of the full radio sky is propagated through

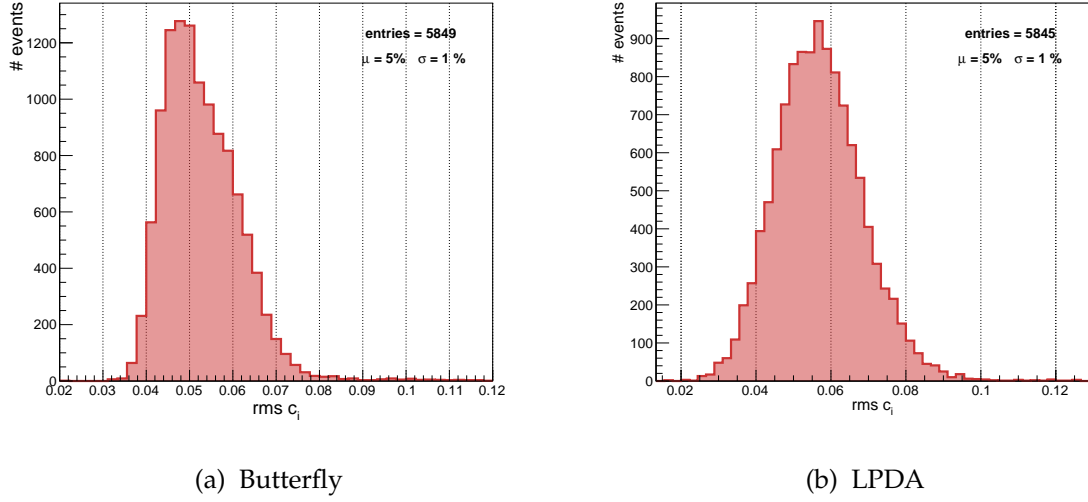


Figure 4.8: Root mean square of the station constants  $c_i$  calculated for seven non-consecutive days of periodically triggered events. The mean value of this distribution is added as systematic uncertainty to the uncertainty of the electric field amplitude.

the antenna pattern and compared to the average spectra recorded by the stations. Frequency-dependent gain correction factors are defined as:

$$g(\nu_i) = \frac{1}{m} \sum_{k=1}^m \frac{\overline{A}_{\text{galaxy}}(\nu_i, t_k)}{A_{\text{model}}(\nu_i, t_k)}, \quad (4.4)$$

where  $\overline{A}(\nu)_{\text{galaxy}}$  is the amplitude of the galactic background as measured by the AERA antennas, and  $A(\nu)_{\text{model}}$  the amplitude of a model of the radio background propagated through the antenna pattern. In each frequency bin the ratio is averaged over all  $m$  LST bins. An ideal description of the antenna pattern would lead to constants compatible with 1 within the systematic uncertainty of the model. In the next subsections the method to obtain these constants is described.

#### 4.2.1 Amplitude of the measured galactic emission

The average amplitude of the measured radio emission is the combination of the galactic background and the thermal noise from the stations electronics:

$$\overline{A}_{\text{meas}}^2(\nu) = \overline{A}_{\text{galaxy}}^2(\nu) + \overline{A}_{\text{noise}}^2(\nu), \quad (4.5)$$

where the noise term includes contributions from different parts of the signal chain: the active antenna, the coaxial cable and the digitizer.

The antenna channels of Butterfly and LPDA are calibrated separately. For both antenna types the average amplitude spectra as function of LST are calculated using periodically triggered events recorded from 16 December 2017 to 16 January 2018. The measurements are corrected for the effects of the analog chain, cleaned from RFI and the gaps left in the frequency spectrum are linearly interpolated. The average spectrum is calculated in steps of one LST-hour. An example of the measured spectrum at LST = 17 h is shown in Figure 4.9.



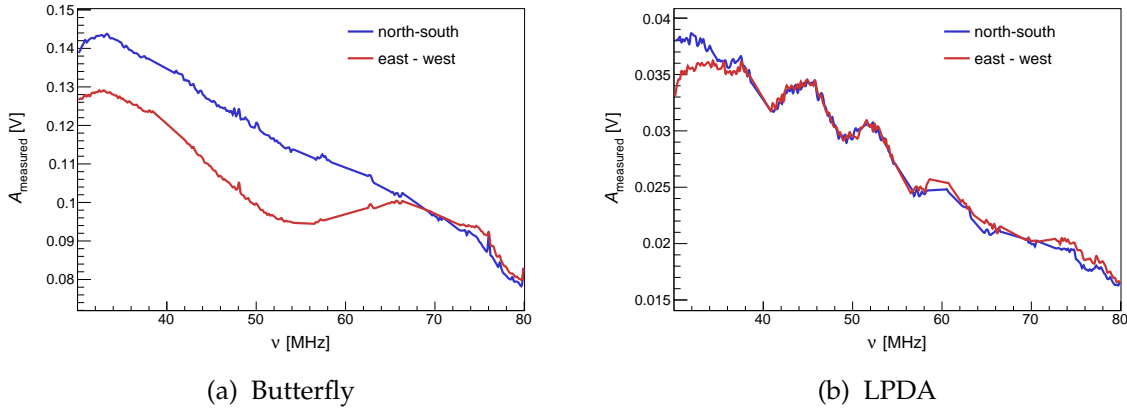


Figure 4.9: Average measured spectrum at LST 17 h with Butterfly and LPDA. The amplitude is given by the combination of the signal from galaxy and the noise from the stations.

#### 4.2.2 Model of the radio frequency sky

A model of the galactic emission is created by the LFMMap program [65]. The program produces temperature maps of the sky at a given frequency  $\nu$  as a function of right ascension  $\alpha$  and declination  $\delta$ . The sky temperature is calculated by interpolating existing sky maps and taking three components to the emission into account: the cosmic microwave background, an isotropic emission (probably due to unresolved extra-galactic sources), and anisotropic galactic emission. Using the Rayleigh-Jeans approximation of a black body radiator the sky temperature can be converted into spectral radiance at the antenna:

$$B(\nu, \alpha, \delta) = \frac{dP}{d\nu} = \frac{2k_B}{c^2} \nu^2 T_{\text{sky}}(\nu, \alpha, \delta), \quad (4.6)$$

where  $k_B$  is the Boltzmann constant and  $c$  the speed of light. The received power spectral density is given by [66]:

$$P(\nu) = \frac{1}{2} \frac{Z_0}{Z_L} \int_{\Omega} B(\nu, \alpha, \delta) |H_a(\nu, \theta, \phi)|^2 d\Omega, \quad (4.7)$$

where  $Z_0$  is the impedance of free space of  $120\pi \Omega$ ,  $Z_L$  is the load impedance of  $50 \Omega$  and

$$|H_a(\nu, \theta, \phi)|^2 = H_{a,\theta}(\nu, \theta, \phi)^2 + H_{a,\phi}(\nu, \theta, \phi)^2 \quad (4.8)$$

is the total magnitude of the VEL of each antenna arm  $a$ .

The antenna pattern  $H(\nu, \theta, \phi)$  is then mapped to the equatorial coordinate system in which the radio sky is modeled, and the integral in Eq. 4.7 is expressed as a sum over the right ascension  $\alpha$  and the declination  $\delta$ , resulting in the following expression for the received power:

$$P(\nu) = \frac{1}{2} \frac{Z_0}{Z_L} \sum_{\alpha} \sum_{\delta} B(\nu, \alpha, \delta) |H_a(\nu, \alpha, \delta)|^2 \cos \delta \Delta \delta \Delta \alpha. \quad (4.9)$$

From  $P(\nu)$ , which is in units of  $\text{WHz}^{-1}$ , it is possible to determine the amplitude in Volt per unit frequency:

$$A_{\text{model}}(\nu) = \sqrt{P(\nu) Z_L}. \quad (4.10)$$

Figure 4.10 shows the amplitude at LST = 17 h obtained using the simulated and the measured antenna response pattern of the two Butterfly antenna channels. The considerable difference in shape and gain between the two amplitude spectra is consistent with the disagreement found between the amplitudes of the VEL as reported in Section 3.1.

A comparison of these two amplitude spectra to the measured galactic emission, given in Figure 4.9(a), shows that when using the simulated antenna pattern the shapes of  $A_{\text{model}}$  and  $A_{\text{measured}}$  are most similar. Based on this result, the Butterfly simulated antenna response pattern has been used to reconstruct the data presented in the following chapters, modifying the default *Offline* configuration that includes instead the measured antenna response pattern.

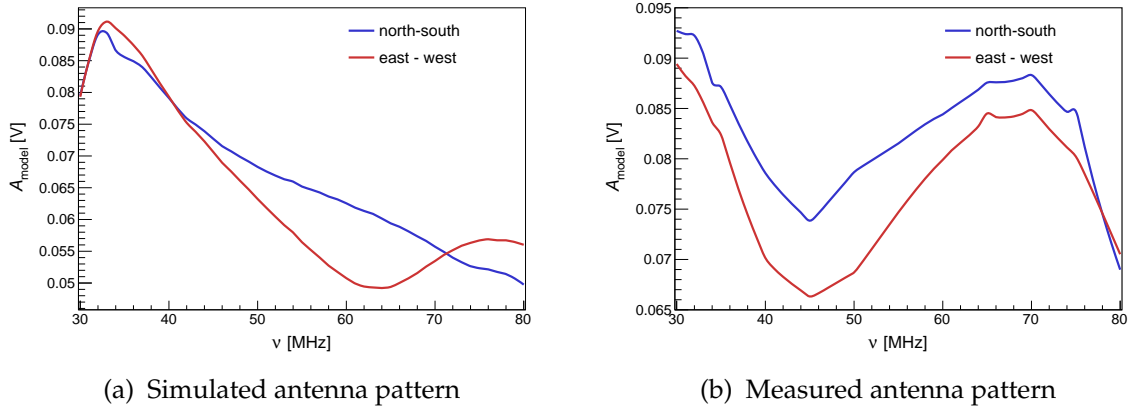


Figure 4.10: Amplitude at LST = 17 h of the radio sky model unfolded with simulated and measured antenna response pattern of the Butterfly stations.

The amplitude at LST = 17 h of the modeled sky resulting after unfolding the simulated or the measured antenna pattern of LPDA stations is shown in Figure 4.11. In this case the shape of the spectra and the gain offsets are similar. This result is in line with the study reported in Ref. [55], which indicates a mean agreement on a 5% level between the magnitudes of simulated and measured VEL. Data have been reconstructed using the LPDA measured antenna pattern.

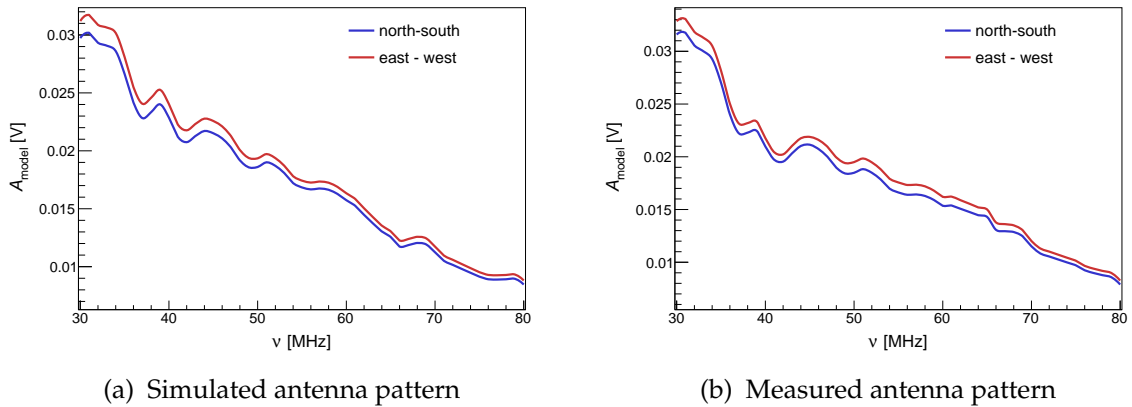


Figure 4.11: Amplitude at LST = 17 h of the radio sky model unfolded with simulated and measured antenna response pattern of the LPDA stations.

In the following the results of the absolute calibration will be shown only for the antenna patterns used in data reconstruction.

### 4.2.3 Results

In order to calculate the calibration constants it is necessary to estimate the average contribution of the noise to the measured amplitude. Furthermore, apart from the differences in the shape between the spectra, for both station types, the amplitude of the radio sky model (Figures 4.10(a) and 4.11(b)) and the measured one (Figures 4.9(a) and 4.9(b)) show significant gain offsets most likely introduced by hardware components. To compensate for this gain offset, a constant scale parameter  $C_0$  is introduced such that:

$$\bar{A}_{\text{galaxy}} = C_0 \cdot \bar{A}_{\text{model}}. \quad (4.11)$$

Thus, Eq. 4.5 can be rewritten as follows:

$$\bar{A}_{\text{meas}}^2 = \bar{A}_{\text{model}}^2 \cdot C_0^2 + \bar{A}_{\text{noise}}^2. \quad (4.12)$$

$C_0$  and  $\bar{A}_{\text{noise}}^2$  are derived using this linear relation: the offset is assumed to affect the measurements the same way and is determined with a combined north-south east-west fit, while the noise contributions are estimated independently.  $\bar{A}_{\text{meas}}^2$  and  $\bar{A}_{\text{model}}^2$  are shown in Figure 4.12 for all frequency and LST bins, and the fit results are listed in Table 4.3. Due to the presence of several emitters between 30 and 40 MHz, for the LPDA the fit is restricted to the range 40-80 MHz. This noise contamination probably originates from the electronic system or from a noise source close to the LPDA field.

$\bar{A}_{\text{noise}}^2$  represents the average contribution to the noise from the combined electronic components. To improve this analysis each step of the signal chain should be independently modeled, as was done by the LOFAR collaboration in Ref. [67].

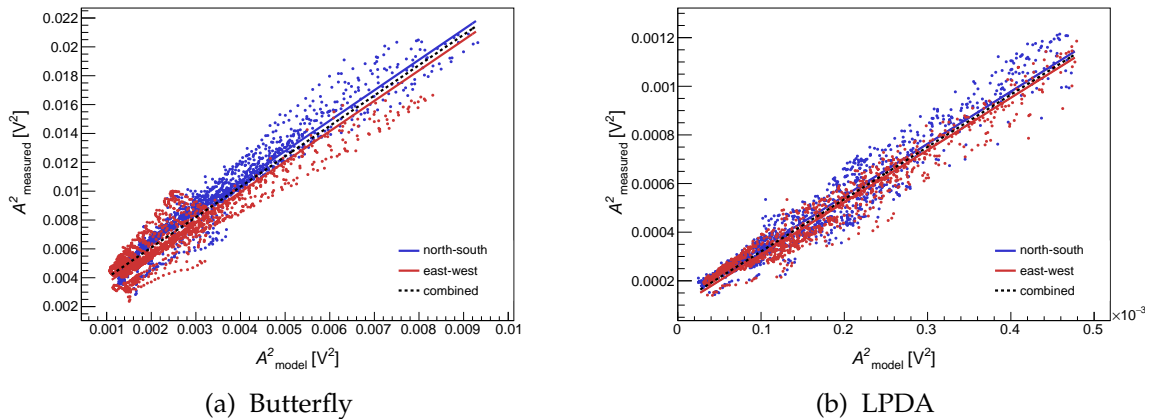


Figure 4.12: Linear fit to all data points of both station types to derive the constant scale parameter  $C_0$  that describes the gain offset between the model and the measurement, and the average noise contribution to the measured amplitude  $\bar{A}_{\text{noise}}^2$ . For the LPDA the fit is restricted to the frequency range 40-80 MHz.

Table 4.3: Fit results for Butterfly and LPDA. For the LPDA the fit is restricted to the frequency range 40-80 MHz.

	Butterfly	LPDA
$C_0$	$1.452 \pm 0.005$	$1.466 \pm 0.004$
$A_{noise,NS}$ [mV]	$47.4 \pm 0.3$	$10.88 \pm 0.09$
$A_{noise,EW}$ [mV]	$38.8 \pm 0.4$	$9.64 \pm 0.08$

Next,  $\overline{A}_{galaxy}$  can be inferred and, therefore, the frequency-dependent gain correction factors (Eq. 4.4). The results as a function of frequency for both antenna types are plotted in Figure 4.13. The strong deviation between 52 MHz and 72 MHz in the east-west channel of the Butterfly stations, correspond to the dip visible in the amplitude of the model (Figure 4.10(a)), and is most probably related to the effect of the electronic box. In the case of LPDA stations the north-south and east-west channels follow the same trend as expected.

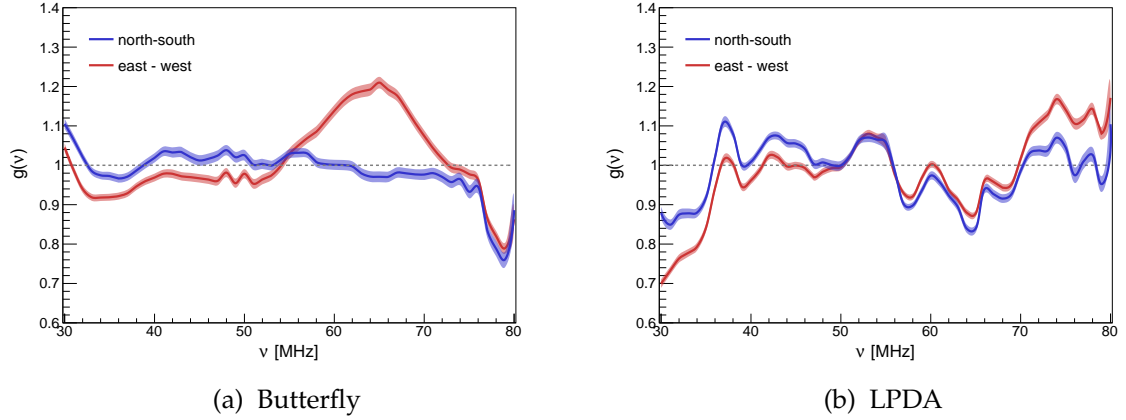


Figure 4.13: Average deviation of the galactic emission as measured by the AERA stations with respect to the modeled radio sky. The simulated antenna pattern is used for the Butterfly stations and the measured antenna pattern for the LPDAs. The shaded areas indicate one standard deviation.

Figures 4.14 and 4.15 show the directional sensitivity of the antenna pattern. The amplitude is averaged over all frequencies and plotted as a function of the LST. For both antenna types, the model of the radio sky and the measurement after the calibration are in a good agreement.

The obtained calibration constants need to be inversely applied to the data before unfolding the antenna response pattern and calculating the electric field. In *Offline* this is done by multiplying the  $g(v)$  with the relative calibration constants calculated in the previous section and adding the `RdChannelGalacticBackgroundCalibrator` module in the module sequence for the event reconstruction.

### 4.3 Validation of the calibration

Since January 2017, a Butterfly and a LPDA station are deployed at a distance of less than 30 m, hence the signals measured by the two stations are similar. In order to validate the calibration constants the deviation between the two antennas in terms of energy fluence has been investigated before and after applying the calibration constants. The data included in

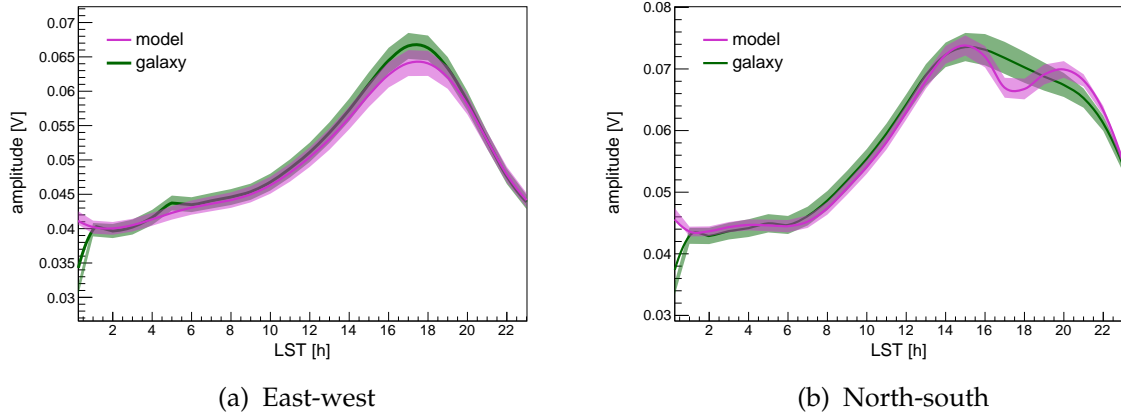


Figure 4.14: Spectral amplitude averaged over all frequencies of both the radio sky model propagated through the antenna pattern and the radio galactic emission as received by the Butterfly stations as a function of LST. The shaded areas indicate the standard errors of the average values.

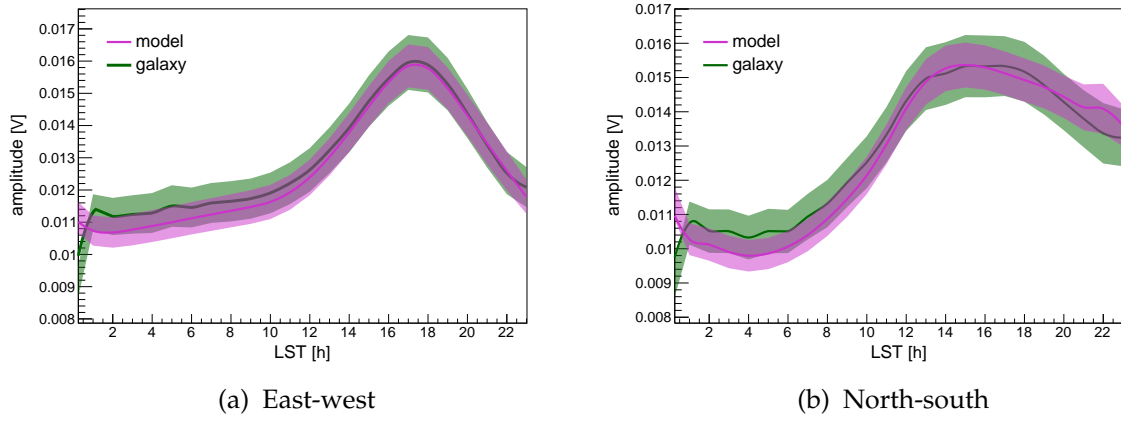


Figure 4.15: Spectral amplitude averaged over all frequencies of both the radio sky model propagated through the antenna pattern and the radio galactic emission as received by the LPDA stations as a function of LST. The shaded areas indicate the standard errors of the average values.

this analysis are externally triggered radio events recorded from January 2017 to December 2019. The data have been reconstructed twice, once without the calibration constants and once multiplying the voltage for the calibration factors before unfolding the antenna response pattern. The stations included in the analysis fulfill the quality cut  $\text{SNR}_{\text{square}} > 10$ . The result is shown in Figure 4.16: the deviation between the measurements decreases from 20% to 5% after the calibration. Although this analysis only includes two stations, it reveals the impact of the calibration on the data analysis.

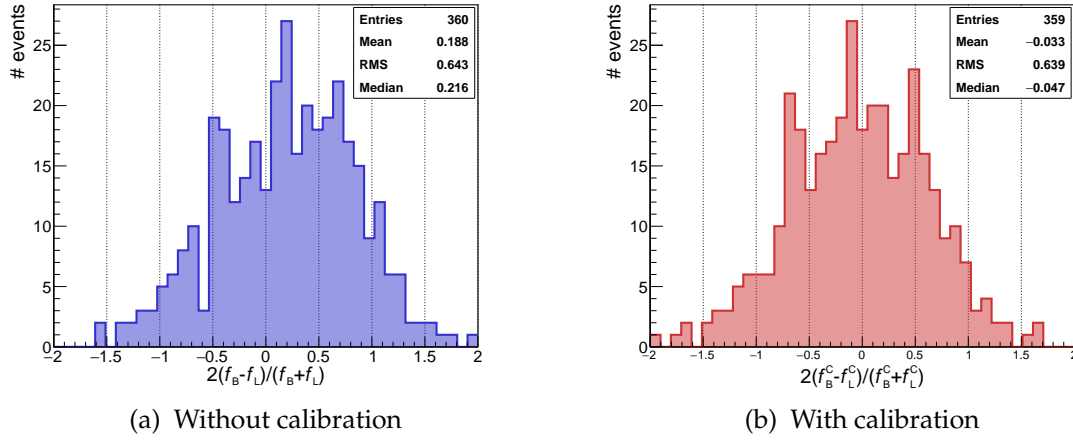


Figure 4.16: Deviations in energy fluence for two antenna stations within 30 m from each other before and after applying the relative and absolute calibration constants. The overall bias is decreased from 20% to 5%.





## 5 Methods to measure cosmic-ray composition from the radio energy distribution and spectral information

Several methods have been investigated to study the composition of cosmic rays using radio signals induced by the air shower propagation in the atmosphere. The measured radio signal is sensitive to the distance from the observer position to the emission region. The location of the emission region can be approximated by the position of the shower maximum  $X_{\max}$  [68]. Given the atmospheric depth above the detector and the zenith angle of the air shower, the distance to the shower maximum can be converted to  $X_{\max}$  in order to obtain an established mass composition parameter and to be able to compare with measurements from other detectors.

Recent experimental analyses have demonstrated that the distance from the shower core to the shower maximum ( $D_{X_{\max}}$ ) can be obtained from the shape of the radio footprint at ground [34, 35, 39] and from the spectral content of the radio signal [37, 40]. Although these studies have shown promising results, the accuracy achieved for the radio reconstructed  $X_{\max}$  is not yet at its limit. An improved parametrization of the radio energy distribution will likely increase the resolution. A complementary approach to increase the precision of the radio  $X_{\max}$  might be to combine independent reconstruction techniques to obtain a mass composition reconstruction that uses different information from the detected radio signal. Both approaches are explored in this chapter using CoREAS Monte Carlo simulations.

A parametrization of the radio energy distribution based on the analytical function presented in Ref. [35] is introduced in Section 5.2. In section Section 5.3, the sensitivity of  $X_{\max}$  to the spectral information of the radio pulse is explored. These estimates for  $X_{\max}$  are compared and combined in Section 5.4.

### 5.1 Simulation data set

The work presented in this chapter is based on CoREAS Monte Carlo simulations available within the Pierre Auger Collaboration.

A library of simulated events has been created using CoREAS 7.5602 with QGSJetII-0.4 [69] and UrQMD [70] as hadronic interaction models. The geomagnetic field is set to its average value at the site of the Pierre Auger Observatory (inclination of  $-35.9^\circ$  and a strength of 0.24 Gauss). The atmospheric model used is the monthly average atmospheric profile for October at the Pierre Auger site that is available in CORSIKA and corresponds to the yearly average at this site. The cosmic-ray energy is distributed uniformly in the logarithm of the energy between  $10^{17}$  eV to  $10^{19}$  eV. The zenith angle  $\theta$  is distributed uniformly in  $\cos\theta$  from  $0^\circ$  to  $80^\circ$  and the azimuth angle  $\varphi$  is chosen randomly. The following analysis includes only air showers with

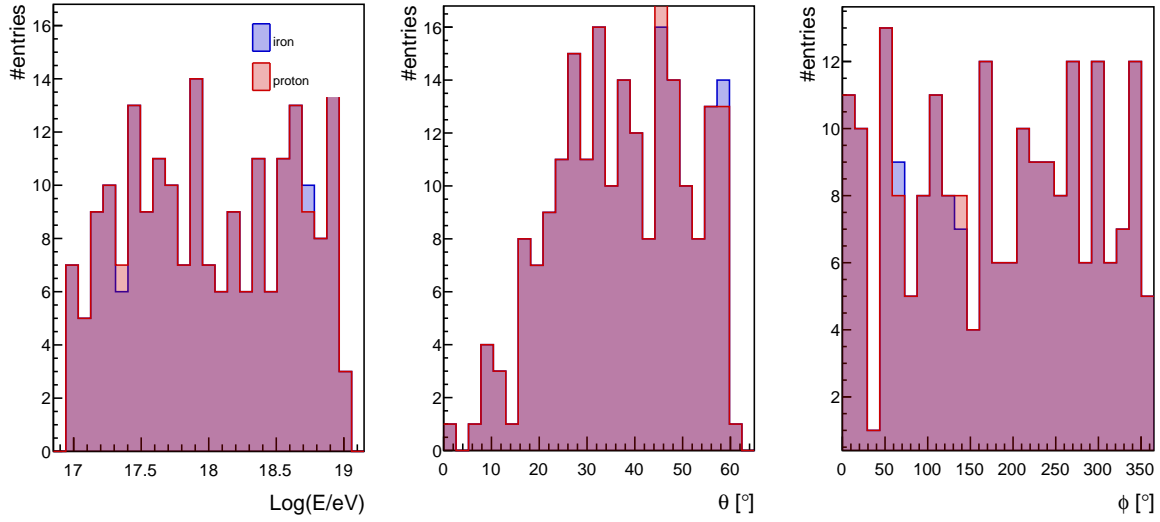


Figure 5.1: Distribution of the input parameters for the simulation data set. From left to right the energy, zenith angle  $\theta$  and azimuth  $\phi$  distributions are shown.

zenith angles below  $60^\circ$  and radio emission calculated at the altitude of AERA (1564 m a.s.l.). The simulation data set used consists of 439 air showers of which 224 have an iron primary and 215 have a proton primary. The distributions of the initial conditions are displayed in Figure 5.1.

The observer positions are aligned in a star-shaped grid in the shower reference frame<sup>1</sup> as shown in Figure 5.2. This choice of the antenna positions provides an useful sampling of

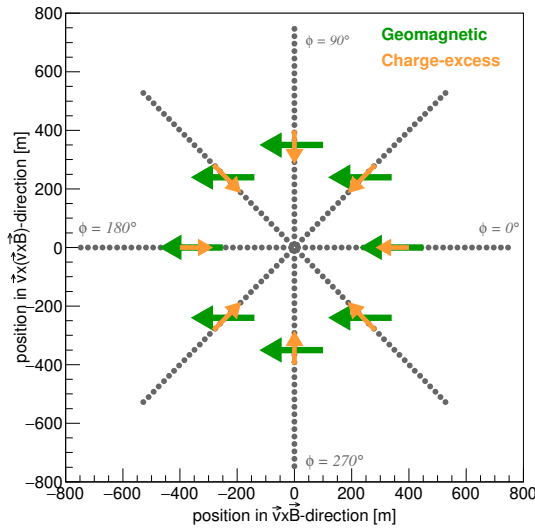


Figure 5.2: Simulated observer positions distributed in a star-shaped grid. Each dot represents a simulated observer position. The arrows indicate the polarization direction of the two emission mechanisms.  $\phi = \arctan(y/x)$  is the observer angle. Figure adapted from [35].

the energy fluence distribution at ground and allows to disentangle the contributions of the

<sup>1</sup>For this reason the simulation set will be referred to as *star-pattern* simulations.

different emission mechanisms. In fact, in CoREAS the emission mechanisms are not explicitly modeled, hence, it is not possible to separate the contribution of the geomagnetic and charge-excess emission processes to the simulated electric field. However, for the observers positioned along the  $\vec{v} \times (\vec{v} \times \vec{B})$ -axis, the polarization of the signal from the geomagnetic and charge-excess processes are orthogonal: the  $\vec{v} \times \vec{B}$  component of the electric field  $E_{\vec{v} \times \vec{B}}$  originates only from geomagnetic emission, whereas the  $\vec{v} \times (\vec{v} \times \vec{B})$  component of the electric field  $E_{\vec{v} \times (\vec{v} \times \vec{B})}$  originates only from charge-excess emission. The polarization directions of the two emission mechanisms are displayed in Figure 5.2. Thereby, the geomagnetic and charge-excess energy fluences can be calculated from the respective electric field components as follows:

$$f_{\text{geo}}(r) = f_{\vec{v} \times \vec{B}}(r, \phi = 90^\circ) = \epsilon_0 c \Delta t \sum_i E_{\vec{v} \times \vec{B}}^2(r, \phi = 90^\circ, t_i), \quad (5.1)$$

$$f_{\text{ce}}(r) = f_{\vec{v} \times (\vec{v} \times \vec{B})}(r, \phi = 90^\circ) = \epsilon_0 c \Delta t \sum_i E_{\vec{v} \times (\vec{v} \times \vec{B})}^2(r, \phi = 90^\circ, t_i). \quad (5.2)$$

### 5.1.1 Simulation reconstruction in *Offline*

The simulated events have been reconstructed in *Offline* (rev-33196<sup>2</sup>) using the standard application `RdReconstructStarshapedStationPattern`. The antennas are all simulated as LPDAs. The energy fluence and its uncertainty at each observer position have been estimated using the procedure described in Chapter 3 from the simulated electric field traces band-passed between 30.0 MHz and 80.0 MHz. The shower geometry and the core position are fixed to the true Monte Carlo value. The core position corresponds to the origin of the shower plane.

Two different cases are considered: an ideal case scenario produced without including the antenna response or any measured noise and source, and a more realistic scenario that includes the detector simulation and measured background traces. The background traces are chosen randomly from periodically triggered events recorded by LPDA stations during 2018.

## 5.2 Depth of the shower maximum from the radio energy distribution

The spatial radio signal distribution contains information about the energy and the mass of the primary cosmic ray. For a precise measurement of the shower properties a good analytic parametrization of the distribution of the radio energy fluence at ground is essential. At first, an empirical parametrization based on morphological arguments [34] has been applied to the data measured by AERA to reconstruct the cosmic-ray energy [32] and  $X_{\text{max}}$  [39]. More recently, an analytic model of the spatial signal distribution directly based on its emission processes has been formulated [35]. The energy distributions originating from the geomagnetic and the charge-excess emission mechanisms are modeled separately and the two-dimensional asymmetry in the radio signal distribution is described by the interference between the two emission mechanisms. For this reason, this analytic model is referred to as *GeoCE* parametrization. One of the parameters of the function is the radiation energy, directly

<sup>2</sup>The module `RdStationSignalReconstructorWithBgSubtraction` is the updated version from rev-33406.

related to the electromagnetic shower energy. The other parameters describe the size of the Cherenkov ring, which correlates with the distance to the emission region, and therefore with  $X_{\max}$ .

Figure 5.3 shows the sampling of the geomagnetic and charge-excess signals as a function of the position along the  $\vec{v} \times (\vec{v} \times \vec{B})$ -axis for three Monte Carlo examples with a different  $D_{X_{\max}}$ . For large distances to  $X_{\max}$ , when the shower has emitted all radiation energy when reaching the ground, the distributions are peaked in the Cherenkov cone. As the distance between the emission region and the observer becomes smaller the distributions of the energy fluence become peaked and narrow around the shower axis.

With the GeoCE parametrization obtained in Ref. [35] it is possible to model in a smooth way the transition from the broad distribution with a visible Cherenkov ring to the peaked signal distribution of the geomagnetic and the charge-excess emissions. Taking advantage from the effective sampling of the two-dimensional energy distribution available in the star-pattern simulation data set, the achieved resolution for the reconstructed  $D_{X_{\max}}$  is about 13 g/cm<sup>2</sup>. However, the sampling limitation of the AERA array introduces a technical challenge for the fitting procedure. In fact, it is difficult to model the transition from a Gaussian shaped signal distribution to a distribution peaked around the shower axis, the first produced by a fully developed shower and the second by a shower that is still developing when hitting the ground. As a consequence the multi-parameter fit does not always converge in the correct minimum. In order to avoid this situation and find a reliable estimation of the shower parameters when fitting experimental data, it has been decided to focus only on fully developed showers for which the correlation between the Cherenkov ring parameters and  $D_{X_{\max}}$  is less complex.

The following study focuses only on air showers with  $D_{X_{\max}} \geq 550$  g/cm<sup>2</sup>, whereas cases like the one shown in Figure 5.3(c) are not considered. The cut on  $D_{X_{\max}}$  reduces the usable events in the simulation data set from 439 to 141, of which 80 have an iron as a primary particle and 61 have a proton. The distributions of the initial parameters of this reduced simulation data set are shown in Figure 5.4.

Following the same strategy as described in Ref. [35], the geomagnetic and the charge-excess signal distributions have been parameterized separately. The GeoCE parametrization is also studied in the presence of background.

### 5.2.1 Signal distribution of the geomagnetic emission

The electric field vector of the geomagnetic emission is polarized in the direction of the Lorentz force  $\vec{v} \times \vec{B}$ . The strength of the emission is circular symmetric around the shower axis and is peaked in a Cherenkov cone. The energy fluence is parameterized as a function of the perpendicular distance to the shower axis  $r$ . For fully developed showers the parametrization is given by [35]:

$$f_{\text{geo}} = \frac{1}{N_{\text{geo}}} E'_{\text{geo}} \left[ \exp \left( - \left( \frac{r - R_{\text{geo}}}{\sqrt{2}\sigma_{\text{geo}}} \right)^{p(r)} \right) + \exp \left( - \left( \frac{r + R_{\text{geo}}}{\sqrt{2}\sigma_{\text{geo}}} \right)^{p(r)} \right) \right], \quad (5.3)$$

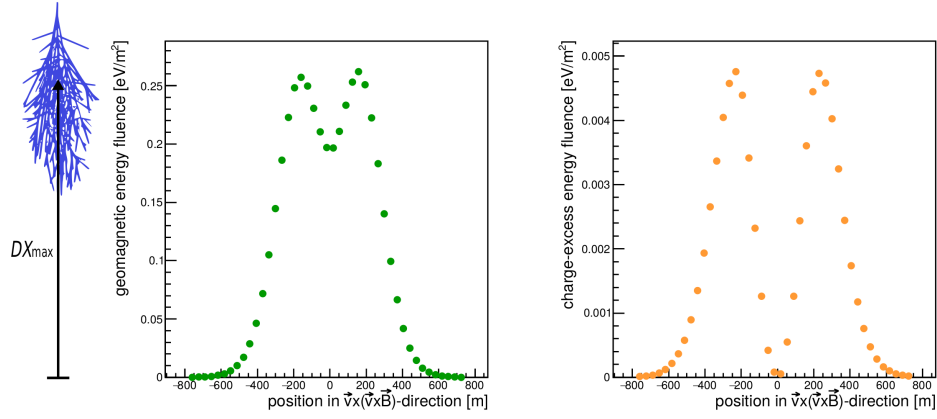
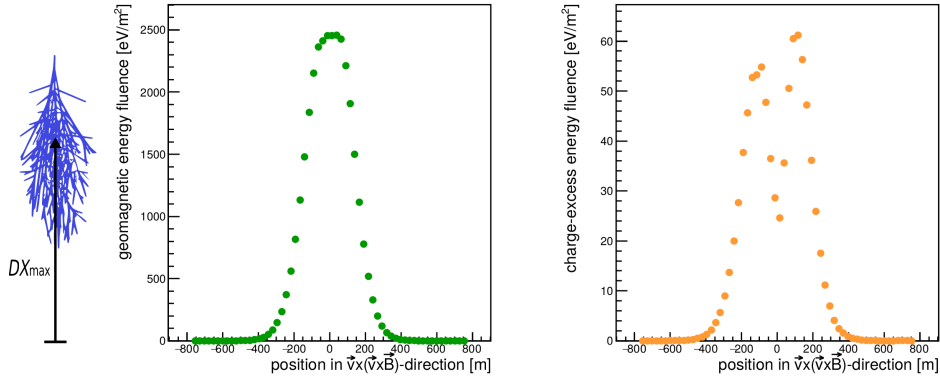
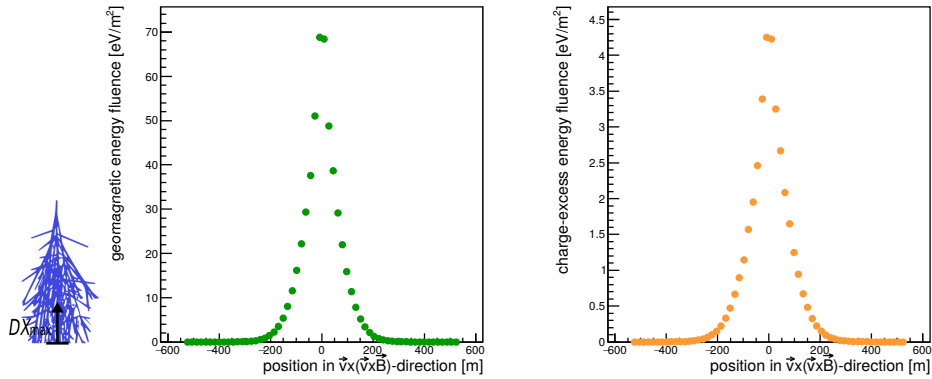

 (a)  $E = 0.1 \text{ EeV}$ ,  $(\theta, \varphi) = (58^\circ, 150^\circ)$ ,  $D_{X_{\max}} = 1060 \text{ g/cm}^2$ 

 (b)  $E = 4.5 \text{ EeV}$ ,  $(\theta, \varphi) = (46^\circ, 247^\circ)$ ,  $D_{X_{\max}} = 590 \text{ g/cm}^2$ 

 (c)  $E = 0.5 \text{ EeV}$ ,  $(\theta, \varphi) = (22^\circ, 334^\circ)$ ,  $D_{X_{\max}} = 327 \text{ g/cm}^2$ 

Figure 5.3: Sampling of the energy fluence from the geomagnetic (left) and charge-excess (right) emission along the  $\vec{v} \times (\vec{v} \times \vec{B})$ -axis for iron initiated showers with different  $D_{X_{\max}}$ . The shape of the signal distribution changes with the distance to the shower maximum. (a) For large  $D_{X_{\max}}$  both distributions are peaked in the Cherenkov cone. (b) As the distance between the emission region and the observer becomes smaller the energy distribution of the geomagnetic emission assumes more a Gaussian like shape, while the charge-excess distribution is still peaked on the Cherenkov cone. (c) For showers that hit the observer while still developing both energy distributions become peaked and narrow around the shower axis.

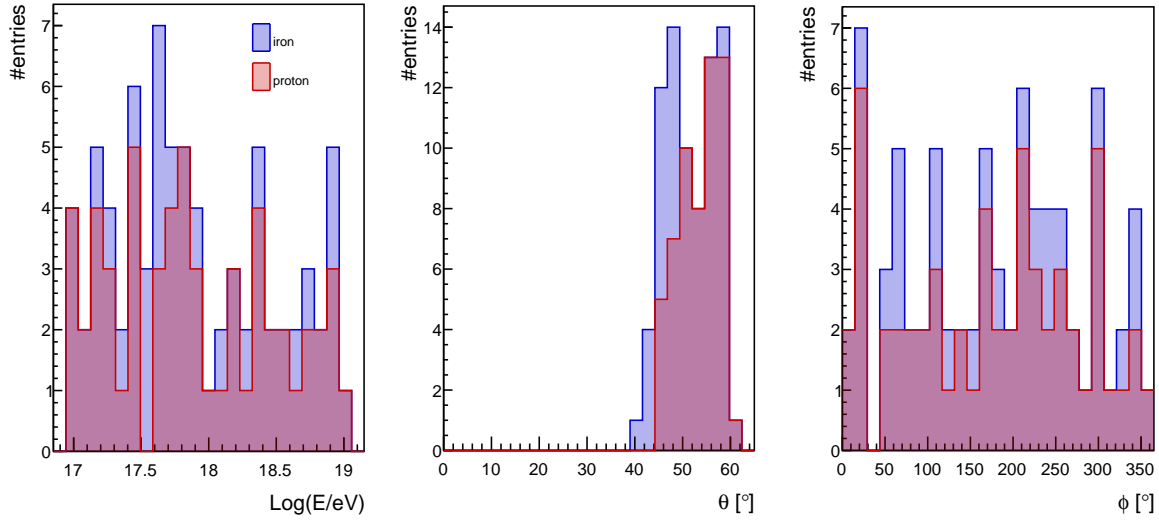


Figure 5.4: Distribution of the input parameters for the simulation data set with a Monte Carlo value of  $D_{X_{\max}} \geq 550 \text{ g/cm}^2$ . From left to right the energy, zenith angle  $\theta$  and azimuth  $\phi$  distributions are shown.

with  $R_{\text{geo}} > 0$  and  $r = \sqrt{x^2 + y^2}$ , where  $x$  and  $y$  are the coordinates in the shower plane  $\vec{v} \times \vec{B} - \vec{v} \times (\vec{v} \times \vec{B})$ . The exponent  $p(r)$  is defined as follows:

$$p(r) = \begin{cases} 2 & \text{if } r \leq r_{\text{cut}} \\ 2 \left( \frac{r}{r_{\text{cut}}} \right)^{-b/1000} & \text{if } r > r_{\text{cut}} \end{cases} \quad (5.4)$$

and it represents a small correction to an exponent of 2 needed to describe the steep falloff of the signal at large distances. In Ref. [35],  $b$  and  $r_{\text{cut}}$  are parameterized as a function of  $D_{X_{\max}}$ . In this work, to simplify the minimization procedure, the parameters are fixed to an average value in  $D_{X_{\max}}$  range under consideration:  $b = 180$  and  $r_{\text{cut}} = 220$ .  $R_{\text{geo}}$  and  $\sigma_{\text{geo}}$  can be interpreted respectively as the radius and the width of the geomagnetic Cherenkov ring,  $E'_{\text{geo}}$  does not correspond exactly to the geomagnetic radiation energy  $E_{\text{geo}}$  because of the variation of the exponent  $p(r)$ . The normalization factor  $N_{\text{geo}}$  is calculated for  $p(r) = 2$ , such that  $E'_{\text{geo}} = E_{\text{geo}}$ . The final expression is [35]:

$$N_{\text{geo}} = 2\pi\sigma \left[ \text{erf} \left( \frac{-R_{\text{geo}}}{\sqrt{2}\sigma_{\text{geo}}} \right) \sqrt{2\pi}R_{\text{geo}} + 2\sigma_{\text{geo}} \exp \left( \frac{-R_{\text{geo}}^2}{2\sigma_{\text{geo}}^2} \right) \right]. \quad (5.5)$$

Next step consists of finding the correlations between the free parameters of the function,  $R_{\text{geo}}$ ,  $\sigma_{\text{geo}}$  and  $E'_{\text{geo}}$ , and the shower parameters  $D_{X_{\max}}$  and  $E_{\text{geo}}$ . This allows to express the GeoCE parametrization of the geomagnetic emission directly as a function of the physics quantities of interest, as well as to reduce the number of free parameters. The correlations are investigated using the star-pattern CoREAS simulation data set with a Monte Carlo value of the  $D_{X_{\max}} \geq 550 \text{ g/cm}^2$ . The geomagnetic energy fluence at each antenna position along the  $\vec{v} \times (\vec{v} \times \vec{B})$ -axis is calculated using Eq. 5.1. At this stage the detector simulation has not been introduced, and no background is added to the electric field trace. The uncertainty assigned to the energy fluence is estimated adding in quadrature a constant relative uncertainty of 2.5%

and 1% absolute uncertainty for which the maximum value of the geomagnetic energy fluence is taken.  $R_{\text{geo}}$ ,  $\sigma_{\text{geo}}$  and the radiation energy  $E'_{\text{geo}}$  are fitted using a  $\chi^2$  minimization.

In Figures 5.5(a) and 5.6(a) two examples of the fit of the geomagnetic energy fluence are shown. The first example corresponds to an iron shower with a Monte Carlo value of  $D_{X_{\text{max}}}$  of 570 g/cm<sup>2</sup>: the geomagnetic distribution flattens at the shower axis and shows a Gaussian like shape. In the second example the distribution is generated by a proton with  $D_{X_{\text{max}}} = 999$  g/cm<sup>2</sup>, here  $R_{\text{geo}}$  is larger than the width  $\sigma_{\text{geo}}$  and the function becomes peaked at the Cherenkov ring. The different shape of the two distributions is due to free propagation of the electromagnetic waves and not because additional radiation energy is emitted [35, 68].

The correlation between the fit parameters and  $D_{X_{\text{max}}}$  is presented in the top panels of Figure 5.7. The dependencies of  $R_{\text{geo}}$  and  $\sigma_{\text{geo}}$  are described with a second order polynomial. In order to parameterize the ratio  $c_{\text{geo}} = E_{\text{geo}}/E'_{\text{geo}}$  Eq. 5.3 is integrated numerically varying  $D_{X_{\text{max}}}$ . The parameters of the polynomial fit functions are given in Tables 5.1 and 5.2.

Table 5.1: Polynomial fits of form  $p_0 + p_1 D_{X_{\text{max}}} + p_2 D_{X_{\text{max}}}^2$  to Cherenkov ring parameters for both emission contributions in the ideal case scenario.  $R_{\text{geo}}$ ,  $\sigma_{\text{geo}}$ ,  $k_{\text{ce}}$  and  $\sigma_{\text{ce}}$  are expressed in m and  $D_{X_{\text{max}}}$  in g/cm<sup>2</sup>.

	$p_0$	$p_1$	$p_2$
$R_{\text{geo}}$	-91.5	$3.4 \times 10^{-1}$	$-8.7 \times 10^{-5}$
$\sigma_{\text{geo}}$	93.9	$-5.0 \times 10^{-2}$	$7.3 \times 10^{-5}$
$k_{\text{ce}}$	-3.5	$9.0 \times 10^{-3}$	$-3.3 \times 10^{-6}$
$\sigma_{\text{ce}}$	-7.1	$2.8 \times 10^{-1}$	0.0 (fixed)

Table 5.2: Polynomial fits of form  $p_0 + p_1 D_{X_{\text{max}}} + p_2 D_{X_{\text{max}}}^2 + \dots$  for the  $c$  parameter for both emission contributions in the ideal case scenario.  $D_{X_{\text{max}}}$  is expressed in g/cm<sup>2</sup>.

	$p_0$	$p_1$	$p_2$	$p_3$	$p_4$
$c_{\text{geo}}$	1.0	$-2.7 \times 10^{-4}$	$4.7 \times 10^{-07}$	$-3.0 \times 10^{-10}$	$7.3 \times 10^{-14}$
$c_{\text{ce}}$	1.2	$-1.2 \times 10^{-3}$	$2.2 \times 10^{-06}$	$-1.6 \times 10^{-09}$	$4.1 \times 10^{-13}$

Having the expressions for  $c_{\text{geo}}$ ,  $R_{\text{geo}}$ , and  $\sigma_{\text{geo}}$  as a function of  $D_{X_{\text{max}}}$ , and given that [68]:

$$E_{\text{geo}} = E_{\text{rad}} \frac{1}{1 + (a/\sin \alpha)^2}, \quad (5.6)$$

where  $a$  is the relative charge-excess strength and  $\sin \alpha$  is the angle between the shower axis and the geomagnetic field, Eq. 5.3 can be expressed as a function of the total radiation energy  $E_{\text{rad}} = E_{\text{geo}} + E_{\text{ce}}$  and the distance to the shower maximum  $D_{X_{\text{max}}}$ :

$$f_{\text{geo}} = \frac{1}{N_{\text{geo}} \times c_{\text{geo}}} \frac{E_{\text{rad}}}{1 + (a/\sin \alpha)^2} \left[ \exp \left( - \left( \frac{r - R_{\text{geo}}}{\sqrt{2}\sigma_{\text{geo}}} \right)^{p(r)} \right) + \exp \left( - \left( \frac{r + R_{\text{geo}}}{\sqrt{2}\sigma_{\text{geo}}} \right)^{p(r)} \right) \right]. \quad (5.7)$$



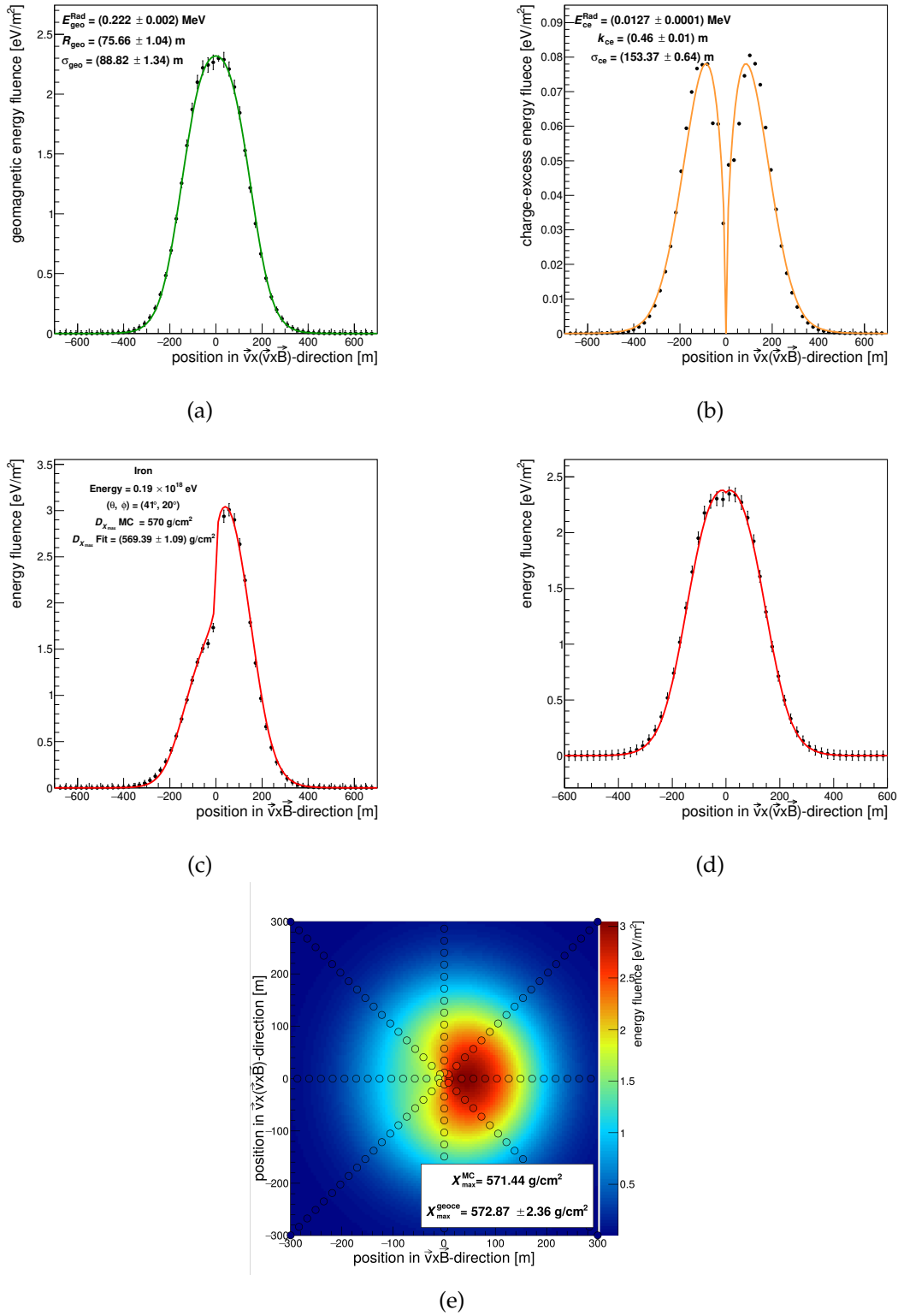


Figure 5.5: Air shower induced by an iron nucleus. Energy fluence distribution from the geomagnetic (a) and charge-excess emission (b) along the  $\vec{v} \times (\vec{v} \times \vec{B})$ -axis. The lines show the best fit using the analytic model of the geomagnetic and the charge-excess emission. (c) Total energy fluence along the  $\vec{v} \times \vec{B}$ -axis and (d) along the  $\vec{v} \times (\vec{v} \times \vec{B})$ -axis. The lines show the best fit from the interplay of the geomagnetic and the charge-excess models. (e) Two-dimensional representation of the signal distribution in the shower plane: the circles indicate the measured energy fluence, while the background is the interpolated signal strength with the best fit parameters for  $D_{X_{\text{max}}}$  and  $E_{\text{rad}}$ .

## 5.2 Depth of the shower maximum from the radio energy distribution

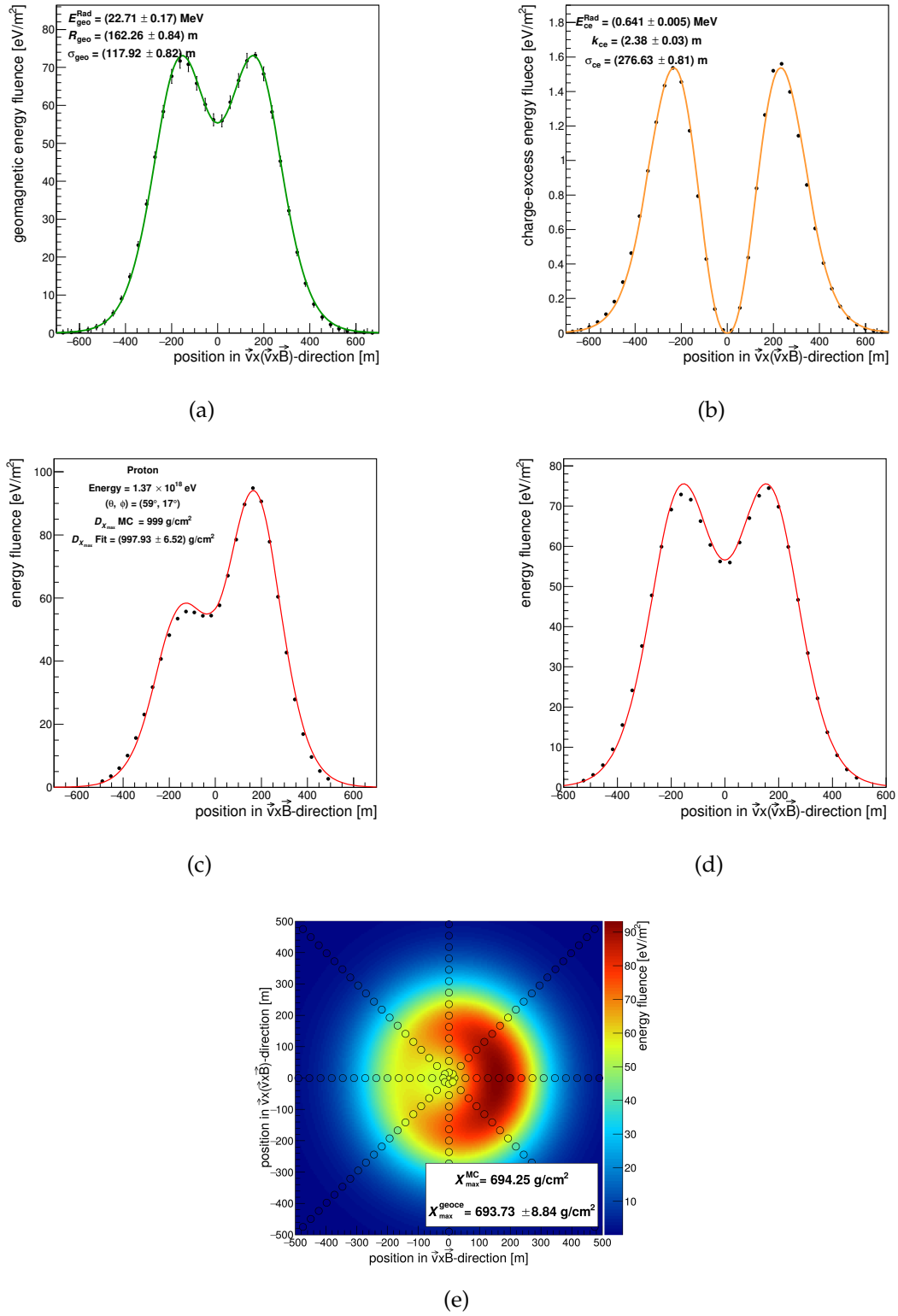


Figure 5.6: Air shower induced by a proton. Energy fluence distribution from the geomagnetic (a) and charge-excess emission (b) along the  $\vec{v} \times (\vec{v} \times \vec{B})$ -axis. The lines show the best fit using the analytic model of the geomagnetic and the charge-excess emission. (c) Total energy fluence along the  $\vec{v} \times \vec{B}$ -axis and (d) along the  $\vec{v} \times (\vec{v} \times \vec{B})$ -axis. The lines show the best fit from the interplay of the geomagnetic and the charge-excess models. (e) Two-dimensional representation of the signal distribution in the shower plane: the circles indicate the measured energy fluence, while the background is the interpolated signal strength with the best fit parameters for  $D_{X_{\text{max}}}$  and  $E_{\text{rad}}$ .

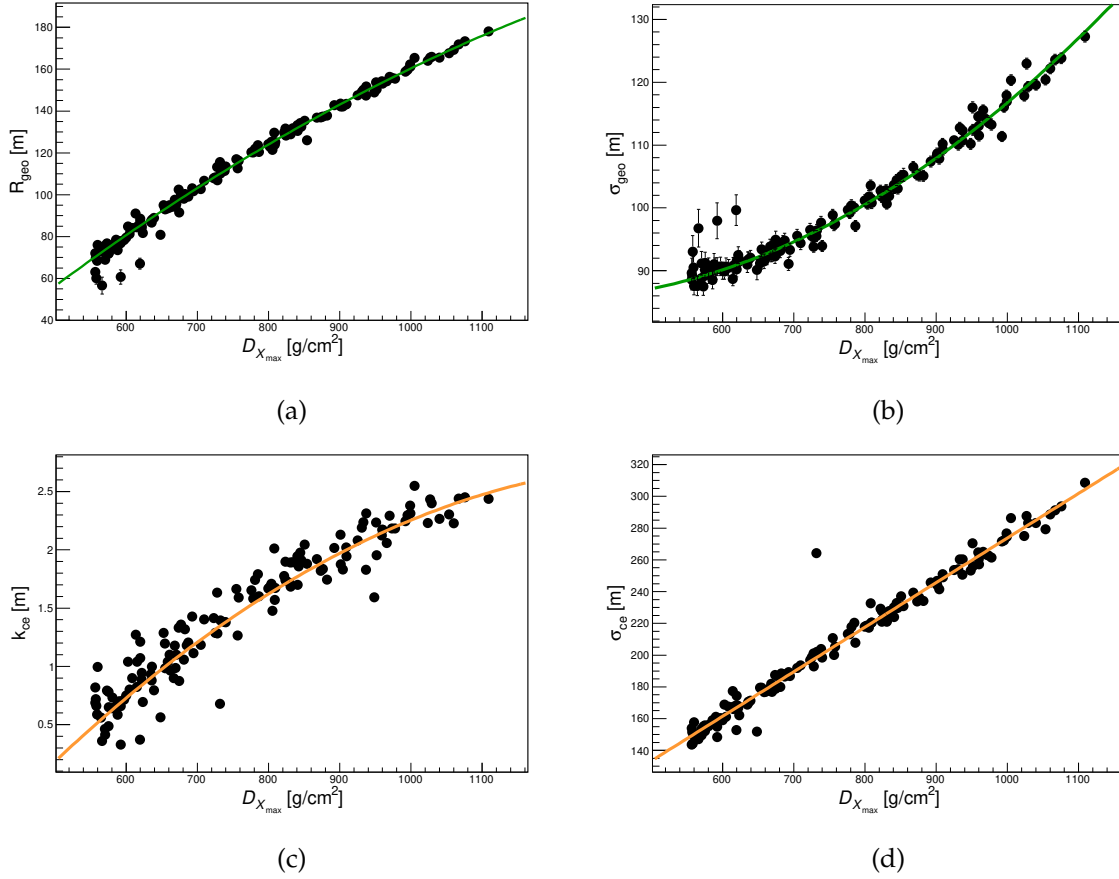


Figure 5.7: Correlation of fit parameters of the geomagnetic function ((a) and (b)) and the charge-excess function ((c) and (d)) with  $D_{X,\max}$ . The lines show the best fit of the correlation. The fit parameters are listed in Table 5.1.

### 5.2.2 Signal distribution of the charge-excess emission

The electric field vector of the charge-excess emission is radially polarized in the  $\vec{v} \times \vec{B}$ - $\vec{v} \times (\vec{v} \times \vec{B})$  plane. The strength of the emission is circular symmetric around the shower axis and is peaked in a Cherenkov cone. The energy fluence due to this emission is parameterized as [35]:

$$f_{ce} = \frac{1}{N_{ce}} E'_{ce} r^{k_{ce}} \exp \left( \frac{-r^{p(r)} (k_{ce} + 1)}{p(r) \sigma_{ce}^{p(r)}} \right) \quad (5.8)$$

with  $r = \sqrt{(x^2 + y^2)}$  and, for fully developed showers,  $k_{ce} > 0$ .  $\sigma_{ce}$  describes the width of the Cherenkov ring, while the radius of the Cherenkov ring is given by  $R_{ce} = \sigma_{ce} \sqrt{k_{ce}} / \sqrt{k_{ce} + 1}$ . The exponent  $p(r)$  is described by Eq. 5.4, and, as for the geomagnetic emission, the parameters  $b$  and  $r_{\text{cut}}$  are fixed to an average value estimated from Ref. [35]. For the charge-excess emission  $b = 315$  and  $r_{\text{cut}} = 350$ . The normalization factor  $N_{ce}$  is obtained fixing  $p(r) = 2$ , such that  $E'_{ce}$  corresponds to the radiation energy of the charge-excess emission  $E_{ce}$  when  $p(r) = 2$ . The result of the integral is [35]:

$$N_{ce} = \frac{2\pi}{k_{ce} + 1} 2^{k_{ce}} (2k_{ce} + 2)^{-0.5k_{ce}} \sigma_{ce}^{k+2} \Gamma(k_{ce}/2 + 1). \quad (5.9)$$

Following the same procedure as for the geomagnetic emission, the calibration curves between the fit parameters,  $k_{ce}$ ,  $\sigma_{ce}$  and  $E'_{ce}$ , and  $D_{X_{max}}$  are investigated using star-pattern CoREAS simulations with  $D_{X_{max}} \geq 550$  g/cm<sup>2</sup>. The charge-excess energy fluence at each antenna position along the  $\vec{v} \times (\vec{v} \times \vec{B})$ -axis is calculated using Eq. 5.2.  $E'_{ce}$ ,  $k_{ce}$  and  $\sigma_{ce}$  are fitted in a  $\chi^2$  minimization procedure.

In Figures 5.5(b) and 5.6(b) two examples of the fit on the charge-excess energy fluence are shown. The function is zero at  $r = 0$  and grows as the observer moves away from the shower until the maximum in correspondence with the Cherenkov ring is reached.

The correlations of  $k_{ce}$  and  $\sigma_{ce}$  with  $D_{X_{max}}$  are presented in the bottom panels of Figure 5.7. The best fit is a second order polynomial for  $k_{ce}$  and a first degree polynomial for  $\sigma_{ce}$ . The ratio  $c_{ce} = E_{ce}/E'_{ce}$  is studied as a function of  $D_{X_{max}}$  integrating Eq. 5.8 numerically. The results of the polynomial fit functions for the parameters of the radio energy distribution functions are given in Tables 5.1 and 5.2.

Using Eq. 5.6, Eq. 5.8 is expressed using only two parameters, the total radiation energy  $E_{rad}$ , and the distance to the shower maximum  $D_{X_{max}}$ :

$$f_{ce} = \frac{1}{N_{ce} \times c_{ce}} E_{rad} \left( 1 - \frac{1}{1 + (a/\sin \alpha)^2} \right) r^{k_{ce}} \exp \left( \frac{-r^{p(r)}(k_{ce} + 1)}{p(r)\sigma_{ce}^{p(r)}} \right), \quad (5.10)$$

where  $c_{ce}$ ,  $k_{ce}$ , and  $\sigma_{ce}$  are functions of  $D_{X_{max}}$  only.

### 5.2.3 Two-dimensional signal distribution

Assuming that the geomagnetic and the charge-excess components are in phase ( $E_{geo}(\vec{r}, t) = w(\vec{r})E_{ce}(\vec{r}, t)$ , with  $w \in \mathbb{R}$ ), the total energy fluence can be calculated as follows [68]:

$$f(\vec{r}) = f_{\vec{v} \times \vec{B}}(\vec{r}) + f_{\vec{v} \times (\vec{v} \times \vec{B})}(\vec{r}) \quad (5.11)$$

$$f_{\vec{v} \times \vec{B}}(\vec{r}) = \left( \sqrt{f_{geo}(\vec{r})} + \cos \phi \sqrt{f_{ce}(\vec{r})} \right)^2 \quad (5.12)$$

$$f_{\vec{v} \times (\vec{v} \times \vec{B})}(\vec{r}) = \sin^2 \phi f_{ce}(\vec{r}), \quad (5.13)$$

where  $\phi = \arctan(y, x)$  is the observer angle, i.e. the position of the antenna relative to  $\vec{v} \times \vec{B}$ -axis in the shower plane  $\vec{v} \times \vec{B} - \vec{v} \times (\vec{v} \times \vec{B})$ . These equations, combined with the parametrizations of the geomagnetic (Eq. 5.7) and of the charge-excess (Eq. 5.10) energy fluence distribution, provide an expression for the total energy fluence at any position with only two parameters, distance to the shower maximum  $D_{X_{max}}$  and radiation energy  $E_{rad}$ , if the shower geometry and the shower core are known. This function is now used to fit the two-dimensional radio energy distribution.

The best fits of the total energy fluence for the two example events are shown in Figures 5.5 and 5.6. Along the  $\vec{v} \times \vec{B}$ -axis (Figures 5.5(c) and 5.6(c)) the geomagnetic and charge-excess signals are polarized into opposite directions for  $r < 0$  and into the same direction for  $r > 0$ , resulting in a destructive interference on one side of the shower axis and a constructive interference on the other side of the shower axis. The asymmetry is well modeled by the interference between geomagnetic and charge-excess emission in both examples. On the  $\vec{v} \times (\vec{v} \times \vec{B})$ -axis the two

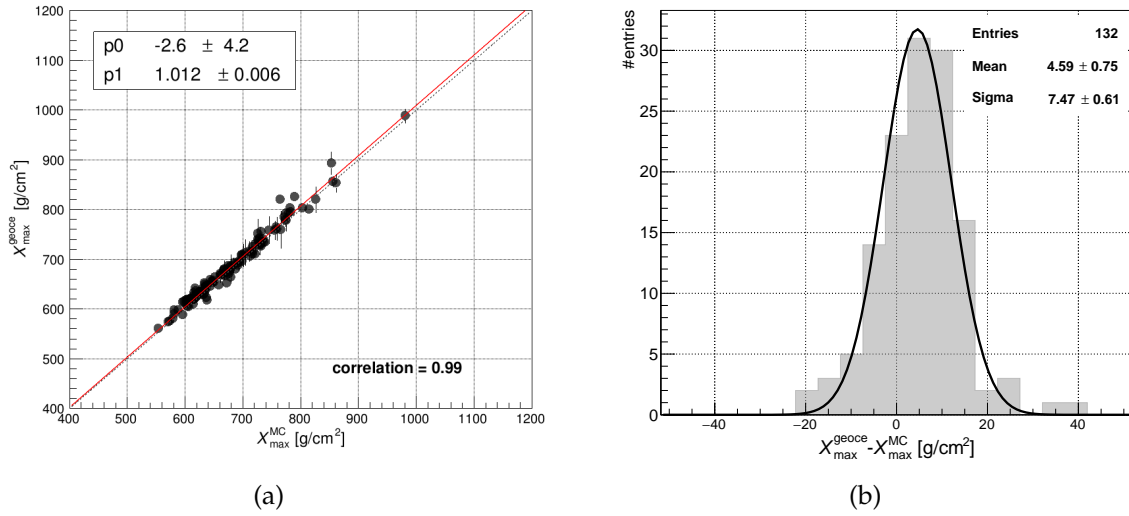


Figure 5.8: (a)  $X_{\max}$  obtained from the best fit parameter  $D_{X_{\max}}$  versus the true  $X_{\max}$  Monte Carlo value. The dashed gray line marks the diagonal. The red line shows a linear fit. (b) Distribution of the difference between reconstructed and true  $X_{\max}$ . A Gaussian fit to the distribution is shown.

signals are polarized perpendicular to each other. The bottom panels of Figures 5.5 and 5.6 show the two-dimensional map of the interpolated energy fluence. Each circle represents an antenna, the color inside the circle corresponds to the measured energy fluence, while the background shows the interpolation of the best fit function. The two-dimensional plot makes it easy to visualize that the function is a good description of the radio energy distribution in the shower plane.

The performance of the GeoCE parametrization is evaluated on the full star-pattern simulation data set, including the events with  $D_{X_{\max}} < 550 \text{ g/cm}^2$ . The fit is considered successful if the upper uncertainty limit  $D_{X_{\max}}^{\text{Fit}} + \sigma_{D_{X_{\max}}}$  is larger than  $550 \text{ g/cm}^2$ . This condition is verified for 132 events, of which 73 are air showers initiated by an iron and 59 by a proton primary. For 10 events with a Monte Carlo value of  $D_{X_{\max}} \geq 550 \text{ g/cm}^2$  and energy below  $10^{17.5} \text{ eV}$  the two-dimensional fit does not converge.

The depth of the shower maximum is inferred from the best fit value of  $D_{X_{\max}}$  and the Monte Carlo value of the zenith angle:

$$X_{\max} = \frac{X(h_{\text{AERA}})}{\cos \theta} - D_{X_{\max}}, \quad (5.14)$$

where  $X(h_{\text{AERA}})$  is the vertical atmospheric depth above the AERA site.

Figure 5.8 shows the performance of the GeoCE parametrization in reconstructing  $X_{\max}$ , which is determined with a resolution of  $7.5 \text{ g/cm}^2$ . On average  $X_{\max}$  is slightly overestimated.

#### 5.2.4 Adding background and detector simulations

The presence of background influences the measurement of the energy fluence and therefore it can influence the correlation between the parameters of the geomagnetic and the charge-excess models and the distance to  $X_{\max}$ . In order to take the effect of the background into account,

the GeoCE parametrization procedure is repeated on simulations reconstructed including the detector simulation and adding detector noise and background to the Monte Carlo electric field traces. At each observer position, the energy fluence  $f$  and its uncertainty are estimated using Eqs. 3.36 and 3.37. In addition a 2.5% uncertainty on the electric field amplitude is introduced by the signal chain of all antennas [32]. 2.5% on the amplitude results in 5% in energy fluence as  $f$  scales quadratically with the electric field amplitude. This uncertainty is added in quadrature to the signal uncertainty.

Only events with at least five stations with a positive value of the energy fluence are selected. Furthermore, for each polarization it is required that at least two stations have an energy fluence above 3 eV/m<sup>2</sup>. The events are selected independently for each polarization to maximize the number of successful fits, which results in 100 events for the geomagnetic analysis and 52 for the charge-excess analysis.

Two examples of the energy fluence from the geomagnetic and charge-excess emission fitted respectively with Eqs. 5.3 and 5.8 are shown in the top panels of Figures 5.9 and 5.10. The correlation of the fit parameters with  $D_{X_{\max}}$  are presented in Figure 5.11: the best fit is a second order polynomial for  $R_{\text{geo}}$  and  $\sigma_{\text{geo}}$ , and a first degree polynomial  $k_{\text{ce}}$  and  $\sigma_{\text{ce}}$ . The optimal parameters of the correlation functions are listed in Tables 5.3 and 5.4.

Table 5.3: Polynomial fits of form  $p_0 + p_1 D_{X_{\max}} + p_2 D_{X_{\max}}^2$  to Cherenkov ring parameters for both emission contributions in the presence of measured background.  $R_{\text{geo}}$ ,  $\sigma_{\text{geo}}$ ,  $k_{\text{ce}}$  and  $\sigma_{\text{ce}}$  are expressed in m and  $D_{X_{\max}}$  in g/cm<sup>2</sup>.

	$p_0$	$p_1$	$p_2$
$R_{\text{geo}}$	-114	$4.0 \times 10^{-1}$	$-1.2 \times 10^{-4}$
$\sigma_{\text{geo}}$	84.6	$-6.1 \times 10^{-3}$	$3.4 \times 10^{-5}$
$k_{\text{ce}}$	-1.5	$3.8 \times 10^{-3}$	0.0 ( <i>fixed</i> )
$\sigma_{\text{ce}}$	-9.1	$2.7 \times 10^{-1}$	0.0 ( <i>fixed</i> )

Table 5.4: Polynomial fits of form  $p_0 + p_1 D_{X_{\max}} + p_2 D_{X_{\max}}^2 + \dots$  for the  $c$  parameter for both emission contributions in the presence of measured background.  $D_{X_{\max}}$  is expressed in g/cm<sup>2</sup>.

	$p_0$	$p_1$	$p_2$	$p_3$	$p_4$
$c_{\text{geo}}$	1.0	$-1.8 \times 10^{-4}$	$-1.6 \times 10^{-07}$	$5.8 \times 10^{-11}$	$-2.7 \times 10^{-15}$
$c_{\text{ce}}$	1.1	$-5.9 \times 10^{-4}$	$1.2 \times 10^{-06}$	$-8.2 \times 10^{-10}$	$1.7 \times 10^{-13}$

The precision of the GeoCE parametrization is evaluated by fitting the two-dimensional energy fluence distribution with Eqs. 5.7 and 5.10 combined using Eqs. 5.11–5.13. The only free parameters of the fit are the distance to  $X_{\max}$  and the total radiation energy, given that the core position is fixed to the origin of the shower plane. The fit is considered successful if  $D_{X_{\max}} + \delta_{D_{X_{\max}}} \geq 550$  g/cm<sup>2</sup>. The list of the selection cuts and the number of showers after each cut is shown in Table 5.5. Two examples of the fit of the total energy fluence are shown in Figures 5.9 and 5.10: panels (c) and (d) show the projection of the total energy fluence on the  $\vec{v} \times \vec{B}$ -axis and the  $\vec{v} \times (\vec{v} \times \vec{B})$ -axis, panel (e) the two-dimensional representation of the fit function. The asymmetry along the  $\vec{v} \times \vec{B}$ -axis is well described by the fit function. The final resolution for  $X_{\max}$  is shown in Figure 5.12: the depth of the shower maximum can be reconstructed with a precision of 11 g/cm<sup>2</sup>, the tail on the left is caused by events with an energy below 10<sup>17.5</sup> eV.

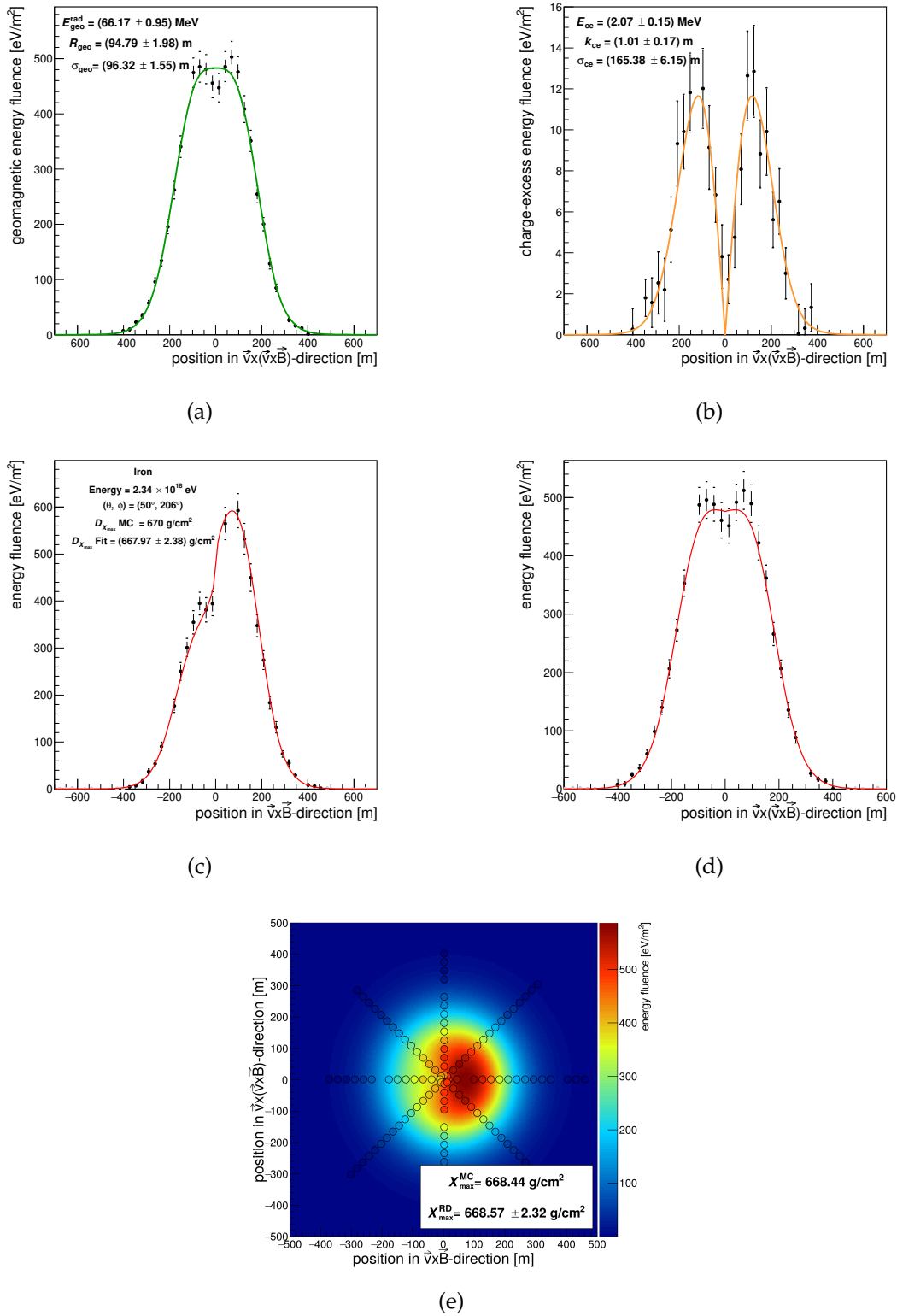


Figure 5.9: Air shower induced by an iron nucleus including noise background. The brackets show the systematic uncertainty introduced by the antenna simulation. Energy fluence distribution from the geomagnetic (a) and charge-excess emission (b) along the  $\vec{v} \times (\vec{v} \times \vec{B})$ -axis. The lines show the best fit using the analytic model of the geomagnetic and the charge-excess function. Total energy fluence along the  $\vec{v} \times \vec{B}$ -axis (c) and the  $\vec{v} \times (\vec{v} \times \vec{B})$ -axis (d). The lines show the best fit from the interplay of the two fit functions. (e) Two-dimensional representation of the signal distribution in the shower plane: the circles indicate the measured energy fluence, while the background is the interpolated signal strength using Eqs. 5.11–5.13 with the best fit parameters for  $D_{X_{\text{max}}}$  and  $E_{\text{rad}}$ .



## 5.2 Depth of the shower maximum from the radio energy distribution

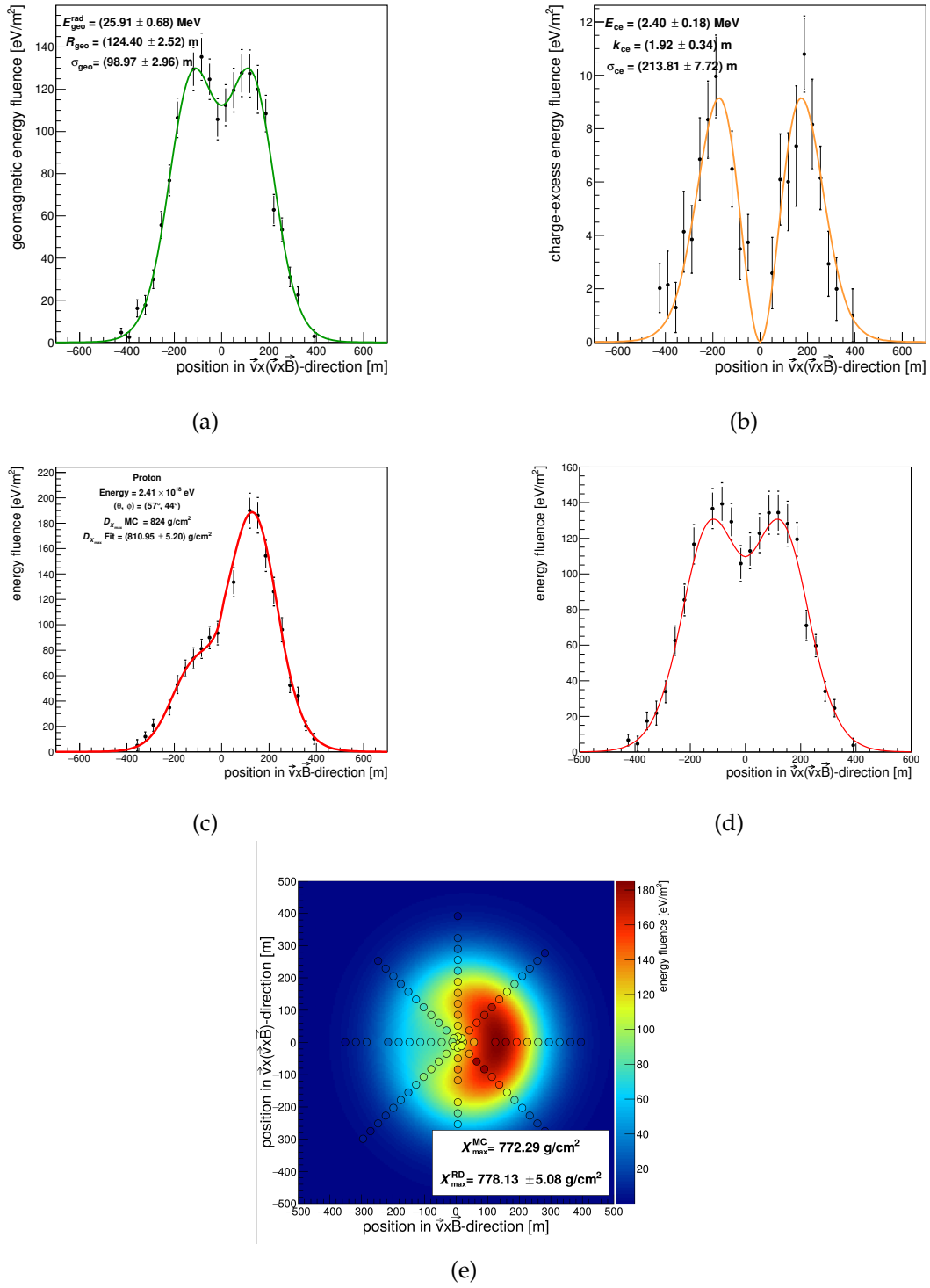


Figure 5.10: Air shower induced by a proton including measured background. The brackets show the systematic uncertainty introduced by the antenna simulation. Energy fluence distribution from the geomagnetic (a) and charge-excess emission (b) along the  $\vec{v} \times (\vec{v} \times \vec{B})$ -axis. The lines show the best fit using the analytic model of the geomagnetic and the charge-excess function. Total energy fluence along the  $\vec{v} \times \vec{B}$ -axis (c) and the  $\vec{v} \times (\vec{v} \times \vec{B})$ -axis (d). The lines show the best fit from the interplay of the geomagnetic and the charge-excess models. (e) Two-dimensional representation of the signal distribution in the shower plane: the circles indicate the measured energy fluence, while the background is the interpolated signal strength using Eqs. 5.11–5.13 with the best fit parameters for  $D_{X_{\text{max}}}$  and  $E_{\text{rad}}$ .

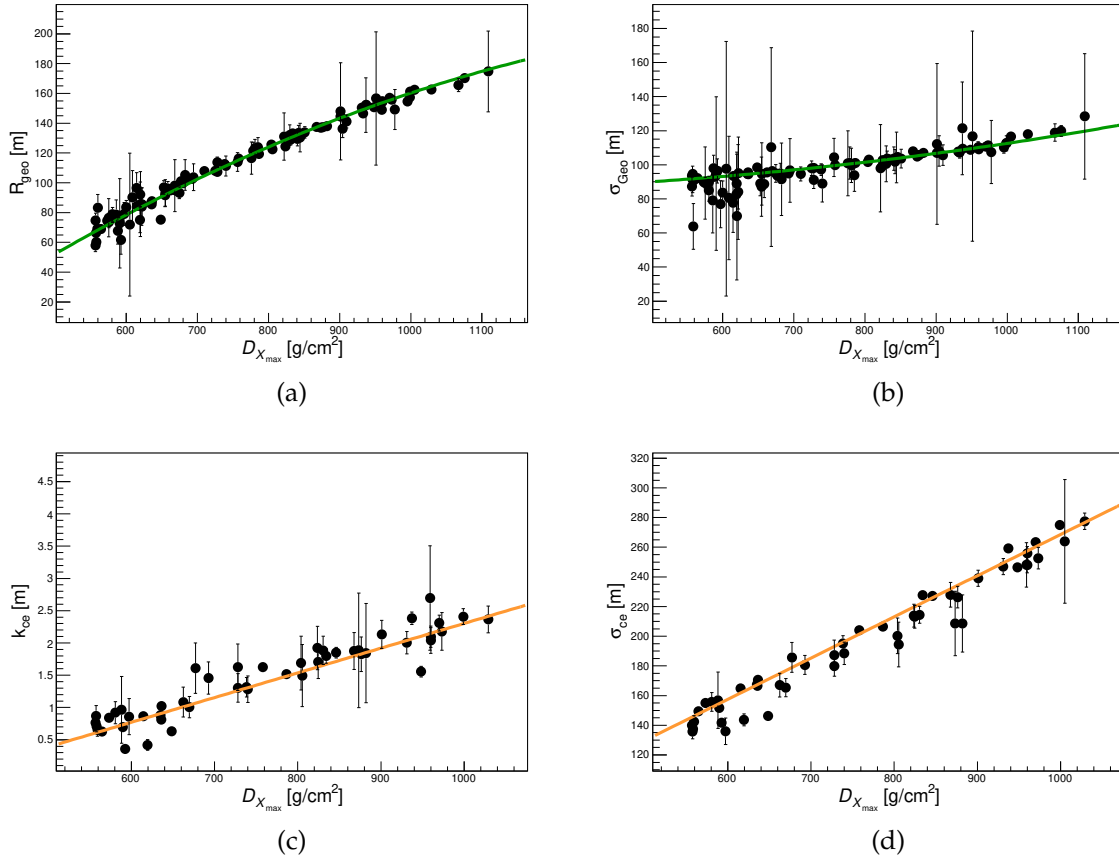


Figure 5.11: Correlation of fit parameters of the geomagnetic function ((a) and (b)) and the charge-excess function ((c) and (d)) with  $D_{X_{\max}}$  after including measured background. The lines show the analytic parametrizations of the correlation. The fit parameters are listed in Table 5.3.

Table 5.5: Number of successful fit of the radio energy distribution for the star-pattern simulation data set including measured background.

Total	439 (224 Fe, 215 p)
$D_{X_{\max}} + \delta D_{X_{\max}} \geq 550 \text{ g/cm}^2$	120 ( 64 Fe, 56 p)
$P_{\chi^2} > 1\%$	112 ( 63 Fe, 49 p)

### 5.2.5 Conclusions

In this section the GeoCE parametrization for the radio energy distribution has been introduced. The GeoCE parametrization is valid for fully developed showers that have emitted all radiation energy when reaching the observer and have  $D_{X_{\max}} \geq 550 \text{ g/cm}^2$ . The reason for introducing this restriction is that the correlation between the parameters describing the shape of the Cherenkov ring and  $D_{X_{\max}}$  becomes more complicated for showers that develop closer to the ground and needs further investigation before it can be used to fit data from a sparse array.

The GeoCE parametrization is based on the work presented in Ref. [35], with the only difference that in  $f_{\text{geo}}$  and  $f_{\text{ce}}$  the tail parameters  $r_{\text{cut}}$  and  $b$  are set to a constant value, and are not functions of  $D_{X_{\max}}$ . The simplification makes it easier to find the minima in the combined

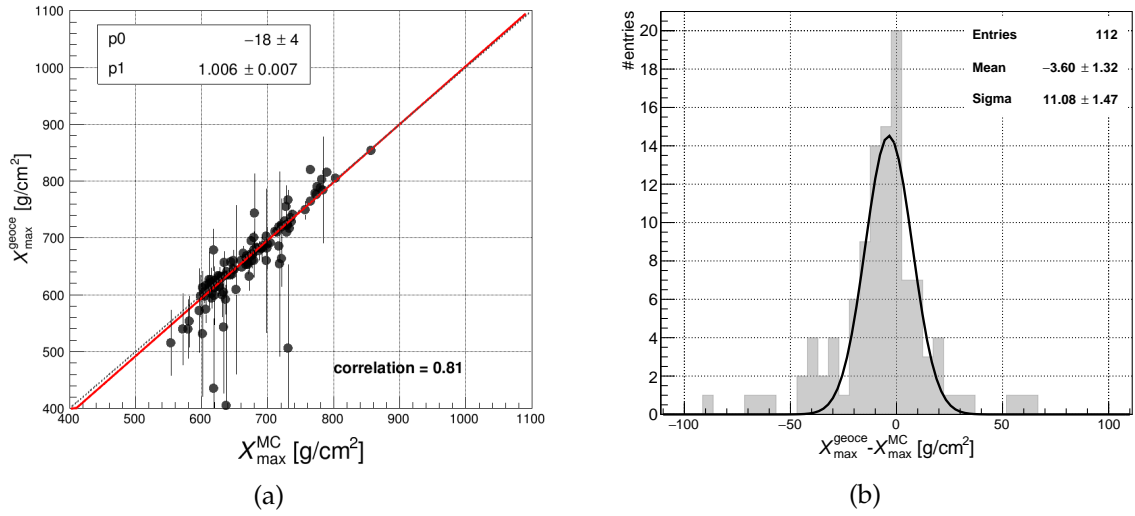


Figure 5.12: Resolution of the reconstructed  $X_{\max}$  after including measured background. The scatter plot shows  $X_{\max}$  inferred from the best fit parameter  $D_{X_{\max}}$  versus the true Monte Carlo value. The histogram shows the absolute deviation. The distribution is fitted with a Gaussian with  $\mu = -3.6$  g/cm<sup>2</sup> and  $\sigma = 11$  g/cm<sup>2</sup>.

fit, considering that for an AERA radio event the number of signal stations is limited.

The theoretical precision achieved for the reconstructed  $X_{\max}$  when adding measured noise and background is 11 g/cm<sup>2</sup>. The actual sampling of the signal provided by the AERA array and the uncertainty on the shower geometry will further affect the final resolution. A bias of -3.6 g/cm<sup>2</sup> is observed. The calibration obtained after adding detector noise and background will be used in the next chapter to fit radio events detected by the AERA array.

### 5.3 Depth of the shower maximum from the slope of the frequency spectrum

The radio pulse shape is sensitive to shower development. Generally the emitted radio signal of heavy nuclei is shorter, as the shower development is compressed into shorter times resulting in a smaller geometric path-length differences. A representation of the path-length difference is shown in Figure 5.13(a). Shorter pulses have more power in the high frequencies than longer pulses as visible for the simulated electric field traces in Figures 5.13(b) and 5.13(c). Thereby, at a given distance to the shower axis, iron-induced showers have flatter spectra than showers induced by lighter particles.

The frequency spectrum in the 30-80 MHz range can be described by an exponential function [71]:

$$S(\nu) = A \cdot 10^{b(\nu - \nu_0)}, \quad (5.15)$$

where the spectral index  $b$  represents the slope on the spectrum,  $A$  is a scale parameter, and  $\nu_0 = 60$  MHz is a constant offset added to make  $A$  proportional to the signal power in this frequency range.

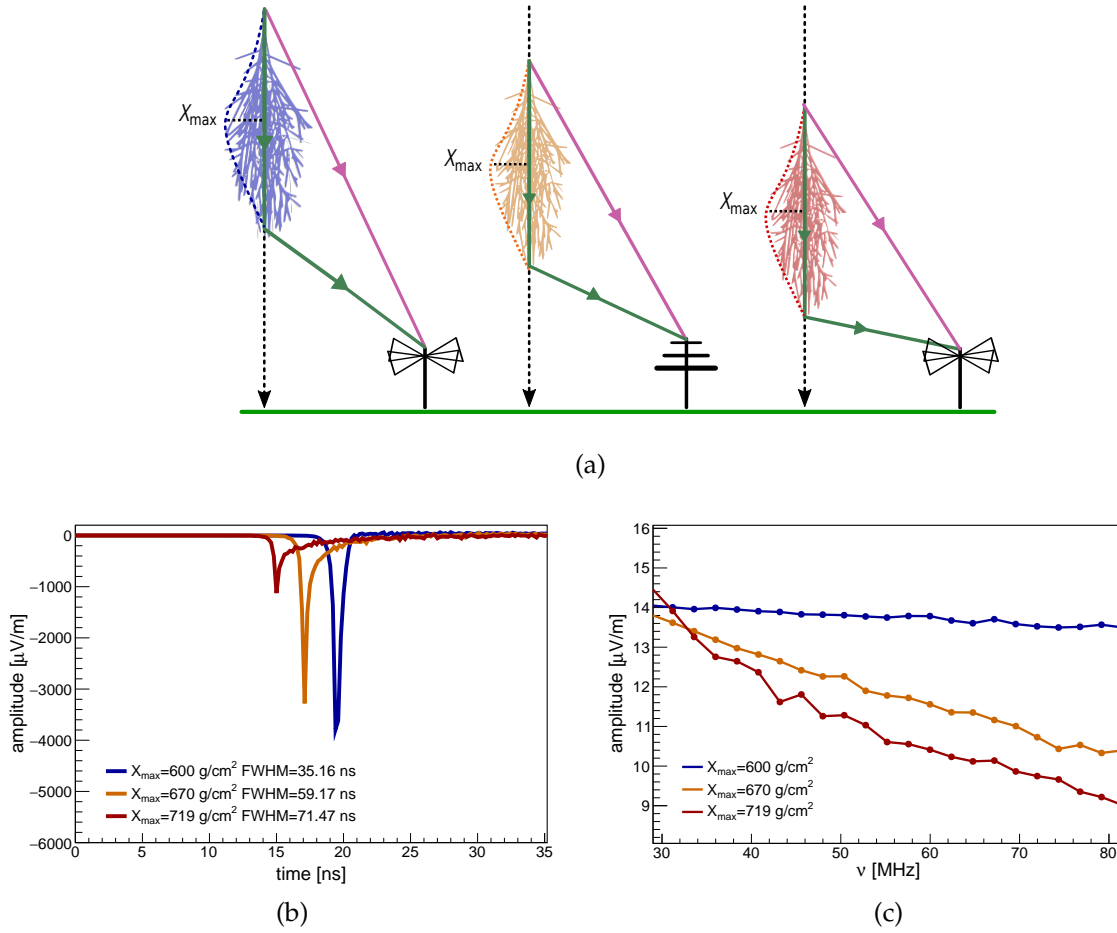


Figure 5.13: Sketch of the dependence of the spectral slope on the position of the shower maximum. The geometric path length difference is represented as the difference in length between the purple and the green line in (a). The blue shower has a shorter path length difference, which will result in a shorter pulse recorded at the observer position (b) and in a flatter spectrum (c). The electric field traces correspond to signals generated by proton-induced air showers detected by an observer positioned at 150 m from the shower axis and with an observer angle of  $90^\circ$ .

A full parametrization of the spectral index of the geomagnetic and charge-excess emission contributions has been derived in Ref. [40]. The main aspects of the model are summarized in the next section.

### 5.3.1 Parametrization of the spectral index

The spectral index of the electric field frequency spectrum along the  $\vec{v} \times \vec{B}$ -axis is described by the following equation:

$$b_{\text{model}} = \frac{1}{\nu_+ - \nu_-} \log_{10} \left[ \frac{10^{b_{\text{geo}}(\nu_+ - \nu_0)} + \cos \phi R \cdot 10^{b_{\text{ce}}(\nu_+ - \nu_0)}}{10^{b_{\text{geo}}(\nu_- - \nu_0)} + \cos \phi R \cdot 10^{b_{\text{ce}}(\nu_- - \nu_0)}} \right], \quad (5.16)$$

where  $b_{\text{geo}}$  and  $b_{\text{ce}}$  are the spectral index respectively of the geomagnetic and the charge-excess contributions,  $[\nu_-, \nu_+]$  is the frequency band,  $\phi$  is the observer angle, and  $R = A_{\text{ce}}/A_{\text{geo}}$  is the ratio between the scale parameters of both emission mechanisms.

For zenith angles up to  $60^\circ$ , the spectral index of the geomagnetic and the charge-excess contributions are parameterized as a function of the distance from the antenna station to the shower maximum,  $D_{\max}$ , and the distance from the antenna to the shower axis,  $d$ :

$$b \times 10^2 = \frac{\beta(d)}{1 + \exp(-\gamma(d) \cdot D_{\max}/\text{km})} - \beta(d), \quad (5.17)$$

where  $\beta(d)$  is a scale parameter and  $\gamma(d)$  is the exponent. The calibration functions for  $\beta(d)$  and  $\gamma(d)$  are reported in Tables 5.6 and 5.7. Table 5.8 shows the calibration used for the  $R$  factor, which also depends on  $D_{X_{\max}}$ . Furthermore,  $R$  depends inversely proportional on the geomagnetic angle  $\alpha$  between the shower propagation direction  $\vec{v}$  and the magnetic field.

Table 5.6: Polynomial fits of form  $p_0 + p_1d + p_2d^2 + p_3d^3$  to scale parameter  $\beta$  for both emission contributions [40].  $d$  is expressed in m.

	$p_0$	$p_1$	$p_2$	$p_3$
Geomagnetic	$-2.73 \times 10^{-1}$	$1.34 \times 10^{-2}$	$1.79 \times 10^{-5}$	0.0 ( <i>fixed</i> )
Charge-excess	1.12	$-8.37 \times 10^{-3}$	$1.64 \times 10^{-4}$	$-2.66 \times 10^{-7}$

Table 5.7: Fits of form  $p_0 + p_1d \cdot \exp -d/p_2$  to scale parameter  $\gamma$  for both emission contributions [40].  $d$  is expressed in m.

	$p_0$	$p_1$	$p_2$
Geomagnetic	$1.55 \times 10^{-1}$	$1.65 \times 10^{-2}$	68.3
Charge-excess	$1.91 \times 10^{-2}$	$4.43 \times 10^{-3}$	141

Table 5.8: Fit values describing the ratio  $R = A_{ce}/A_{geo}$  using a polynomial fit of form  $R \sin \alpha = p_0 + p_1D_{\max} + p_2D_{\max}^2 + p_3D_{\max}^3$ , where  $\alpha$  is the geomagnetic angle [40].  $D_{\max}$  is expressed in km.

$p_0$	$p_1$	$p_2$	$p_3$
0.439	$-7.01 \times 10^{-2}$	$6.78 \times 10^{-3}$	$-2.38 \times 10^{-4}$

Fits of the spectrum of the incoherent emission that occurs further away from the shower axis are unreliable and need to be rejected from the analysis. Furthermore, inside the Cherenkov cone the spectral slope is not sensitive to the shower development because the radio pulse becomes very short producing an almost flat spectrum. To define the region of validity of Eq. 5.17, the coherence cutoff and Cherenkov cone cutoff have been parameterized for all distances to the shower axis  $d$  as a function of  $D_{\max}$ . The final values of the parametrization constants are given in Table 5.9.

Table 5.9: Linear functions best fitting the coherence cutoff and Cherenkov cone cutoff:  $D_{\max} \times 10^{-3} = p_0 + p_1d$  [40].

	$p_0$ [km]	$p_1(\times 10^{-2})$
Cherenkov cone	0.0	7.21
Geomagnetic coherence cutoff	-5.99	3.00
Charge-excess coherent cutoff	-5.28	3.20

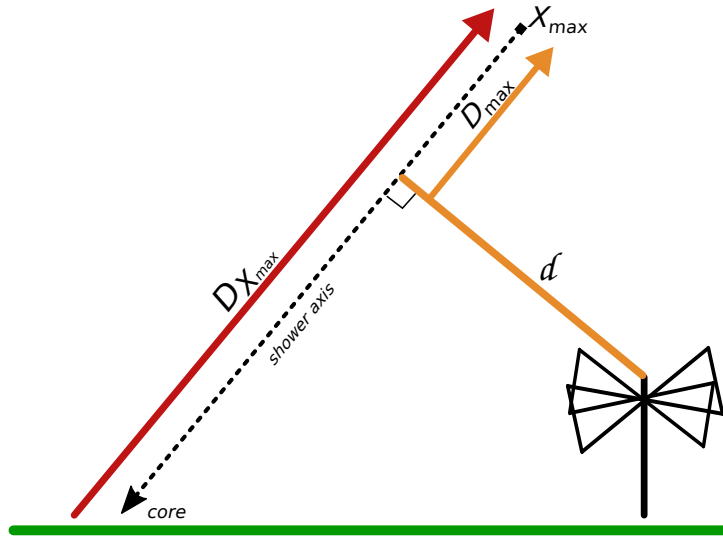


Figure 5.14: Illustration of the shower geometry and the relative position of the radio station. The perpendicular distance from the station position to the shower axis  $v$  is indicated with  $d$ . The distance from the intersection between  $d$  and  $v$  and  $X_{\max}$  is  $D_{\max}$ , i.e. the distance from the antenna station to the shower maximum projected on the shower axis.

Within an event, given the shower geometry and the core position, the distance of each station to the shower maximum is coupled via the geometrical distance from the core and the shower maximum,  $D_{X_{\max}}$ . An illustration introducing the parameters of the function is shown in Figure 5.14.

Hence, the parametrization of the spectral index in Eq. 5.16 is a function of  $d$  and  $D_{\max}(D_{X_{\max}})$ , and, when the shower geometry is known, the only free parameter of this function is  $D_{X_{\max}}$ . This means that with the spectral slope method it is possible to gain information on the shower development by using data even from a single radio station, if the spectral measurement in antenna is accurate enough.

### 5.3.2 Validation on the star-pattern simulation data set

The parametrization of the spectral index is applied to the star-pattern simulation set in order to estimate a theoretical precision of the reconstructed  $X_{\max}$ . The fitting procedure used is largely inspired by the work presented in Ref. [40].

As a first step, pure Monte Carlo electric field traces are considered. The uncertainty assigned to the amplitude of the frequency spectrum is calculated by the square sum of a constant relative uncertainty of 2.5% added to a 1% uncertainty calculated from the maximum value of the amplitude spectrum. Only stations with a distance to the shower axis between 100 m and 400 m are selected, while no restriction is applied to the observer angle.

The frequency spectrum of the  $\vec{v} \times \vec{B}$  component of the electric field is fitted with Eq. 5.15 in the 40-75 MHz range to estimate the spectral index  $b_{\text{fit}}$  and its uncertainty  $\sigma_b$  of each station. Then, the best fitting value of the geometrical distance from the shower core to shower maximum,

$D_{X_{\max}}$  is found by minimizing the following function:

$$\chi^2 = \sum_{i=0}^n \left( \frac{b_{\text{model},i}(D_{\max}(D_{X_{\max}})) - b_{\text{fit},i}}{\sigma_{b,i}} \right)^2, \quad (5.18)$$

where  $n$  is the total number of stations and  $b_{\text{model}}$  is inferred from Eq. 5.16 and Eq. 5.17. The geometry of the event, the core position and the observer angle are set to their Monte Carlo values.

Stations for which the lower uncertainty limit  $b_{\text{fit}} - \sigma_b$  is larger than the maximum value of the parameterized spectral index at its distance from the shower axis  $d$  for any value of  $D_{\max}$  and  $\phi$  are rejected from the fit. For each event a maximum of 25 stations is included in the fit because it has been observed that including more stations does not affect the fit results. When the number of stations that satisfy the aforementioned conditions is larger than 25, a random number of stations is excluded from the fit in order to keep  $1 \leq n \leq 25$ . An iterative fit procedure excludes the stations when the fit result of  $D_{\max}$  places them outside the region of validity of the parametrization defined by the coherent cutoff and the Cherenkov cone cutoff. The fit is redone until all remaining stations are within the valid region or until no station is left. The quality cuts and the number of events selected is listed in Table 5.10. For successful fits,  $D_{X_{\max}}$  is converted in  $X_{\max}$  using the following expression:

$$X_{\max} = \frac{X(h_{\text{AERA}} + D_{X_{\max}} \cos \theta)}{\cos \theta}, \quad (5.19)$$

where  $X(h)$  is the vertical atmospheric depth at height  $h$  above sea level, and  $(h_{\text{AERA}} + D_{X_{\max}} \cos \theta)$  is the geometrical height from the sea level to  $X_{\max}$ . The atmospheric depth is evaluated using the monthly average atmospheric profile for October, the same one used to generate the simulations.

Table 5.10: Number of successful fit with the spectral slope parametrization using the star-pattern simulation data set.

Total	439 (224 Fe, 215 p)
number of stations after iterative fit $\geq 1$	421 (214 Fe, 207 p)
$P_{\chi^2 > 0}$ or $n = 1$	390 (205 Fe, 185 p)

The comparison between the reconstructed and the Monte Carlo values for  $D_{X_{\max}}$  and  $X_{\max}$  is shown in Figure 5.15. The correlation factor is 0.99 for the fitted geometrical distance to  $X_{\max}$  and 0.93 when it is converted in atmospheric depth. The slope of the linear fit in the  $X_{\max}$  correlation plot deviates significantly from 1. This deviation is likely introduced by the atmospheric model. In fact, the calibration curves for the  $\beta(d)$  and the  $\gamma(d)$  functions are obtained using simulations generated with the US-standard atmospheric model while the star-pattern simulations analyzed are generated with the average atmospheric profile for October at the site of the Pierre Auger Observatory.

The resolution of  $X_{\max}$  after the slope correction is shown in Figure 5.16. The theoretical resolution achieved is 20 g/cm<sup>2</sup>. This value combines measurements from a single radio station up to 25 radio stations, if  $X_{\max}$  is derived for all events from a single (random) spectral measurement the resolution increases to 36 g/cm<sup>2</sup>.



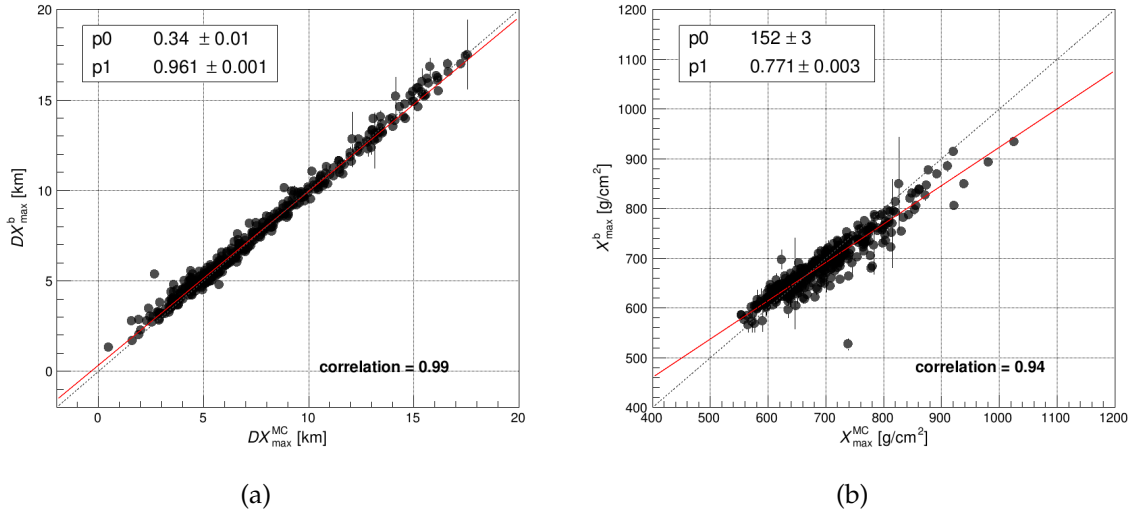


Figure 5.15: (a) The best fit parameter  $D_{X_{\max}}$  versus the true Monte Carlo value. The gray dashed line marks the diagonal. The red line shows a linear fit. (b) The  $X_{\max}$  inferred from  $D_{X_{\max}}$  versus the true Monte Carlo value. The red line shows a linear fit. The slope of the correlation is not consistent with 1. This is likely caused by the atmospheric model. More details are given in text.

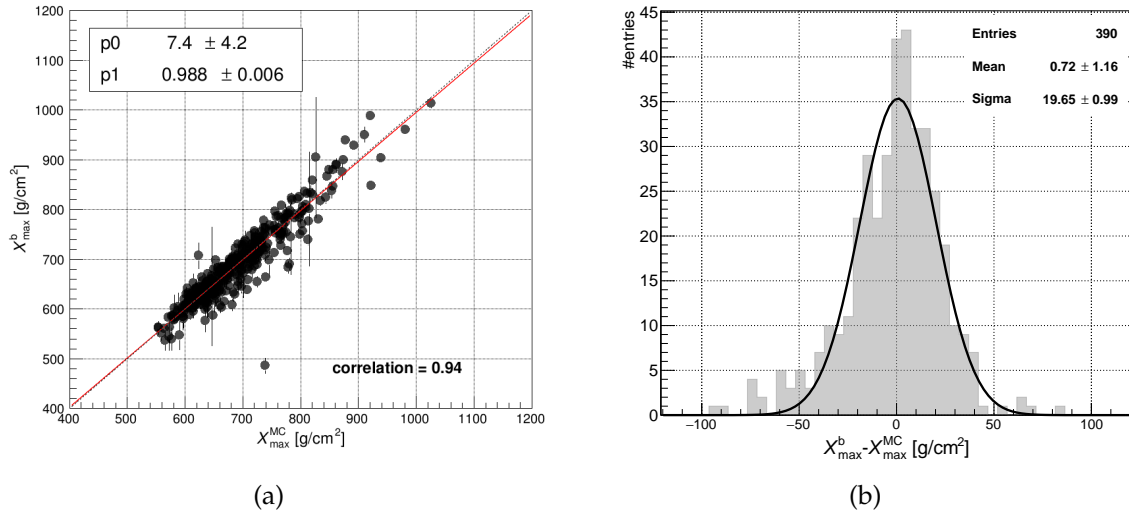


Figure 5.16: (a) The slope-corrected  $X_{\max}$  value versus the true Monte Carlo value. The gray dashed line marks the diagonal. The red line shows a linear fit. (b) Difference between the slope-corrected  $X_{\max}$  and the true Monte Carlo value, the distribution is fitted with a Gaussian with  $\mu = 0.7$  g/cm<sup>2</sup> and  $\sigma = 19.6$  g/cm<sup>2</sup>.

### 5.3.3 Adding background and detector simulations

In order to investigate the effect of the background on the  $X_{\max}$  resolution and on the rectification function for the slope correction, the analysis elaborated in the previous section is applied to the Monte Carlo simulations reconstructed including the detector simulation and adding measured noise and background to the electric field traces. The electric field amplitude and its uncertainty are estimated following the procedure described in Chapter 3. The uncertainty is added in quadrature with a 2.5% uncertainty on the amplification of the signal chain [32].

In addition to the station selection criteria described in the ideal case scenario, also stations with a signal to noise ratio  $\text{SNR}_{\text{amp}} < 5$  are rejected.  $\text{SNR}_{\text{amp}}$  is defined in Eq. 3.34. This cut is derived from Figure 3.11 that clearly shows that the amplitude of the electric field in antennas with a signal to noise below 5 is overestimated.

Following the same procedure described above, the geometrical distance to the shower maximum is inferred combining the measurements of the spectral index of all selected stations within an event. An iterative procedure discards the stations positioned outside the region of validity of the parametrization defined by the coherent cutoff and the Cherenkov cutoff. For successful fits the value of  $D_{X_{\max}}$  is converted to  $X_{\max}$  using the monthly average atmospheric profile for October at the Auger site. The number of successful fits is shown in Table 5.11.

Table 5.11: Number of successful fit with the spectral slope parametrization using the star-pattern simulation data set including background.

Total	439 (224 Fe, 215 p)
number of stations after iterative fit $\geq 1$	251 (129 Fe, 122 p)
$P_{\chi^2} > 1\%$ or $n = 1$	217 (114 Fe, 103 p)

To visualize the performance of the fit one can compare the measured spectral slope  $b_{\text{fit}}$  with the parametrization from Eq. 5.16 obtained using the Monte Carlo value of the geometrical distance to  $X_{\max}$ . An example is shown in Figure 5.17, the shower maximum of the same Monte Carlo event is determined using two and four stations. The stations are selected randomly among those that satisfy the selection criteria. The precision on the reconstructed  $X_{\max}$  increases by adding more stations to the fit.

The overall performance of the parametrization is shown in Figure 5.18. Compared to the ideal case, the spread of the points increases as the uncertainty on the spectral slope introduces more freedom in the minimization algorithm. The correlation is determined for all events, and for  $n \geq 2$  excluding the events in which only one station survives the selection criteria. The events with one station are marked in yellow in the scatter plots. The slope of linear fit in Figure 5.18(b) is close to the value obtained in the ideal case, consistent with the hypothesis that the deviation is introduced by the atmospheric model used in the parametrization. The resolution of  $X_{\max}$  after the slope correction is shown in Figure 5.19. The resolution achieved is  $32.4 \text{ g/cm}^2$ . However, events with only one station are mostly in the tail of the distribution, and it is expected that the final resolution strongly depends on the final number of stations in the  $D_{X_{\max}}$  fit. Hence, for each event of the simulation data set,  $X_{\max}$  has been determined with  $n = 1$ ,  $n = 2$  and  $n = 3$ . Figure 5.20 shows the absolute deviation between the slope-corrected  $X_{\max}^b$  and the true Monte Carlo value of the shower maximum in these three different cases. The use of single radio stations leads to an average  $X_{\max}$  resolution of  $80 \text{ g/cm}^2$ , given the geometry of the shower and the core position.

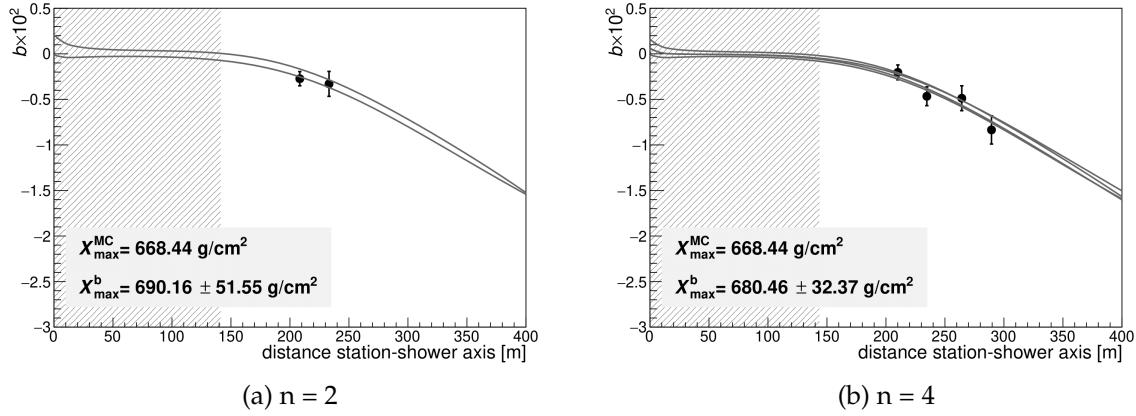


Figure 5.17: Air shower induced by an iron nucleus with initial parameters  $E = 2.34$  EeV,  $(\theta, \varphi) = (50^\circ, 206^\circ)$ ,  $X_{\max} = 668$  g/cm<sup>2</sup>. The plots show the best fit value of the spectral slope as a function of the distance from the station to the shower axis. The gray lines show the parametrization line  $b_{\text{model}}$  obtained using the Monte Carlo value of  $D_{X_{\max}}$  for each station (2 lines in (a) and 4 lines in (b)). The measured spectral slopes are consistent within the uncertainty with the Monte Carlo prediction. The dashed area show the area in which the parametrization is not valid based on the coherent cutoff and the Cherenkov cone cutoff and determined using the Monte Carlo value of  $D_{\max}$ .  $X_{\max}$  is estimated using two and four stations in the fit of the parametrization of the spectral slope. The uncertainty on the reconstructed  $X_{\max}$  becomes smaller when more stations are included in the fit.

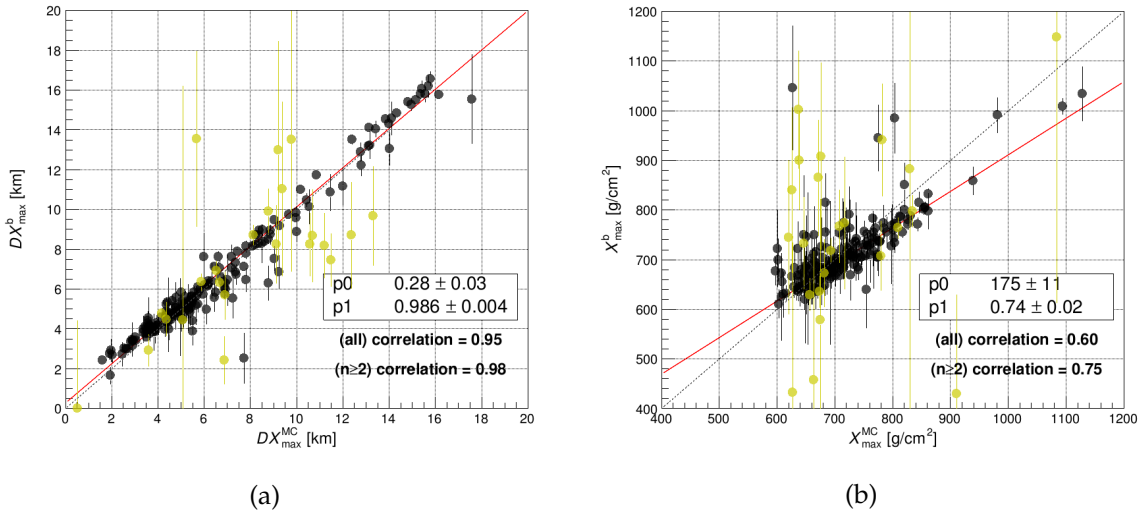


Figure 5.18: (a) The best fit parameter  $D_{X_{\max}}$  versus the true Monte Carlo value. (b) The  $X_{\max}$  inferred from the best fit parameter versus the true Monte Carlo value. The yellow points indicate the events containing a single station. The correlation is calculated for all points and for  $n \geq 2$  excluding the yellow points. The gray line marks the diagonal. The red line shows a linear fit to all points.

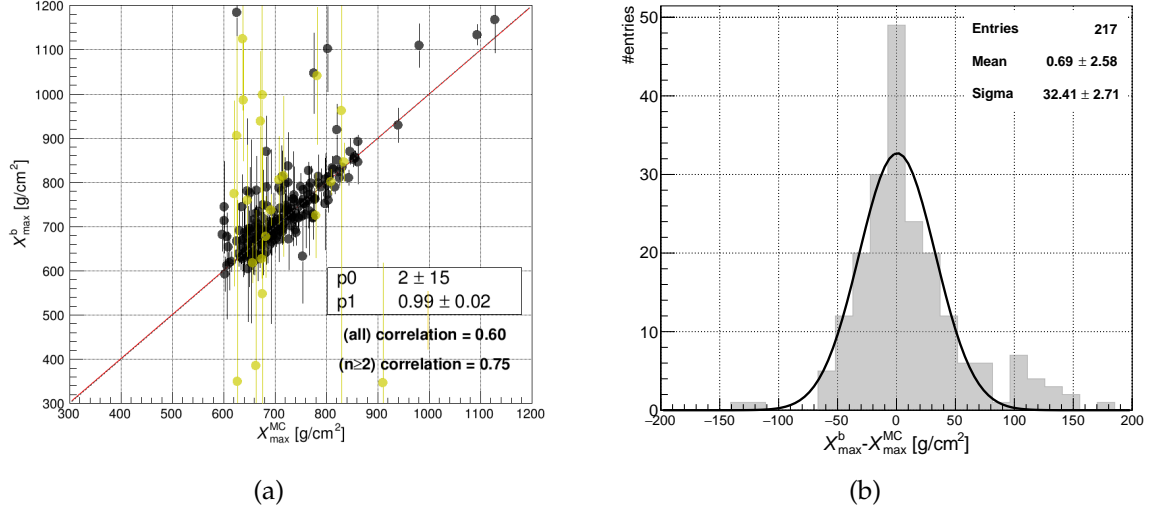


Figure 5.19: (a) The slope-corrected  $X_{\max}$  value versus the true Monte Carlo value. The yellow points indicate the events containing a single station. The correlation is calculated for all points and for  $n \geq 2$ , excluding the yellow points. The gray line marks the diagonal. The red line shows a linear fit to all points. (b) Difference between the slope-corrected  $X_{\max}$  and the true Monte Carlo value, a Gaussian fit to the distribution is drawn. Events containing a single stations are mostly in the tail of this distribution.

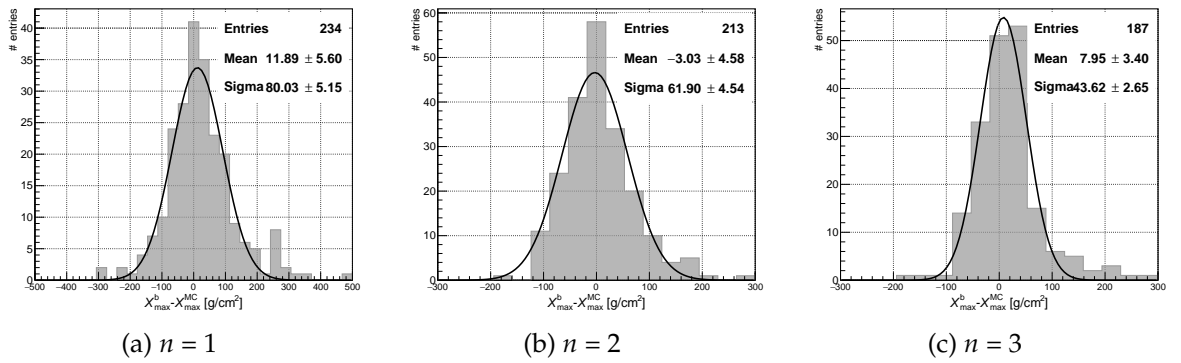


Figure 5.20: Resolution on the slope-corrected  $X_{\max}$  obtained using a different number of stations  $n$  in the fit of the parametrization of the spectral index. The precision on the reconstructed  $X_{\max}$  scales with the number of stations.

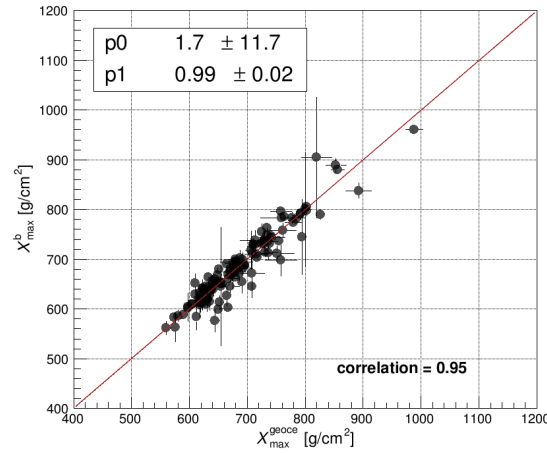


Figure 5.21: Comparison between the radio  $X_{\max}$  obtained from the fit of the energy fluence distribution with the GeoCE parametrization and the one obtained from the parametrization of the spectral slope. The dashed gray line marks the diagonal. The red line is a linear fit.

### 5.3.4 Conclusions

The features of the frequency spectrum and their correlation to the distance to the shower maximum have been extensively studied in Ref. [40], leading to a model for the spectral index as a function of the distance from the observer to the shower maximum,  $D_{\max}$ , and the distance observer-shower axis,  $d$ . In this section the spectral index parametrization has been applied to the star-pattern simulation data set. The  $X_{\max}$  resolution obtained using the measurement of a single radio station is  $36 \text{ g/cm}^2$  in the ideal case and  $80 \text{ g/cm}^2$  when measured background and the detector response are included in the simulation. Combining the measurements of several stations the precision improves to  $20 \text{ g/cm}^2$  in the ideal case scenario and  $33 \text{ g/cm}^2$  including the background. The precision that was achieved by fitting the radio energy distribution is better.

## 5.4 Comparison and combination of the two radio $X_{\max}$ estimators

Two methods to reconstruct  $X_{\max}$  using radio signals from air showers have now been introduced and are validated on the same simulation data set. The two methods use complementary information from the air shower, and, often also different signal stations. Stations located in the Cherenkov ring are excluded in the spectral index fit because their frequency spectrum is almost flat, while the same stations carry essential information for the fit of the two-dimensional energy footprint at ground.

The comparison between the two radio  $X_{\max}$  values is shown in Figure 5.21. The correlation factor is 0.95 and the linear fit is compatible with the diagonal. This result, combined with the fact that the information used in the two  $X_{\max}$  measurements is independent, allows to combine the  $X_{\max}$  measurements in a weighted average:

$$\bar{X}_{\max}^{\text{RD}} = \frac{X_{\max}^b \sigma_b^{-2} + X_{\max}^{\text{geoce}} \sigma_{\text{geoce}}^{-2}}{\sigma_b^{-2} + \sigma_{\text{geoce}}^{-2}}. \quad (5.20)$$

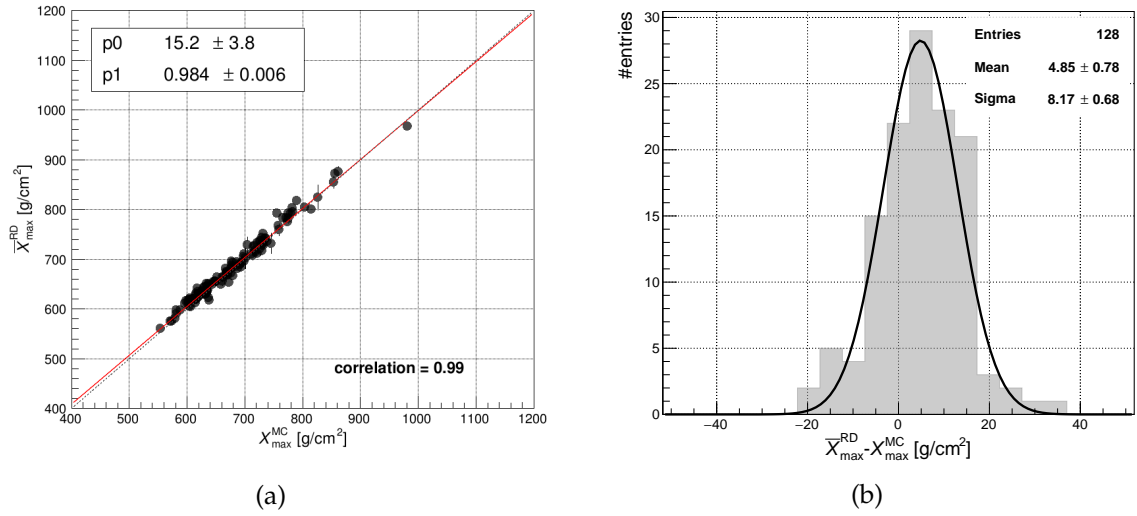


Figure 5.22: (a) The averaged  $X_{\max}$  value versus the true Monte Carlo value. The gray dashed line marks the diagonal. The red line shows a linear fit. (b) Difference between the average value  $X_{\max}$  and the true Monte Carlo value, the distribution is fitted with a Gaussian with  $\mu = 4.8 \text{ g/cm}^2$  and  $\sigma = 8.2 \text{ g/cm}^2$ .

The precision obtained by estimating  $X_{\max}$  from the energy fluence distribution is higher than the one achieved using the parametrization of the spectral slope. This is a direct consequence of the sensitivity of the fit parameters to the shower development, in other words the shape of the Cherenkov ring shows a stronger correlation to the distance to the emission region with respect to the slope of the frequency spectrum that, to first order, depends on the distance from the observer to the shower axis. On the other hand it is important to point out that the star-shaped alignment of the antennas in the shower plane significantly boosts the  $X_{\max}$  resolution in the case of the fit of the energy distribution, while it does not affect the performance of the spectral slope parametrization for which, instead, the core position needs to be known rather precisely. Thereby, in this particular case the precision of the determination of the radio shower maximum will not improve by averaging the two measurements, but it is not possible to draw a more general conclusion based on this simulation study.

The correlation plot between the radio averaged and the Monte Carlo value of  $X_{\max}$  is shown in Figure 5.22 together with the distribution of their absolute deviation. For 128 events in the star-pattern simulation data set both  $X_{\max}$  estimations can be made. The  $X_{\max}$  resolution is  $8.2 \text{ g/cm}^2$ , which is compatible with the resolution achieved using only the GeoCE parametrization of the energy fluence distribution (Figure 5.8).

Another radio  $X_{\max}$  estimator can be obtained by using the weighted average when available and the single measurement for the events that only have one reconstructed  $X_{\max}$ . This combined radio estimator for the shower maximum increases the event statistic, which is an important aspect given that the GeoCE parametrization of the energy distribution is valid only for fully developed showers with  $D_{X_{\max}} \geq 550 \text{ g/cm}^2$ . The performance of this estimator is shown in Figure 5.23. The overall precision for the combined radio  $X_{\max}$  estimator is  $17.2 \text{ g/cm}^2$ .

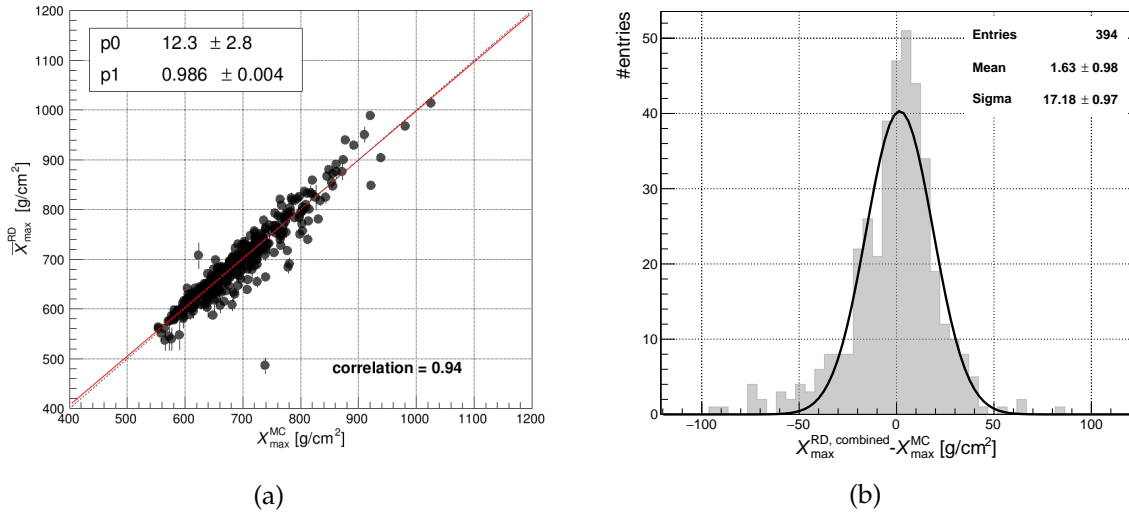


Figure 5.23: The combined radio  $X_{\max}$  estimated where the weighted average of the two estimations when possible, otherwise the available  $X_{\max}$  measurement is taken. (a) The radio  $X_{\max}$  value versus the true Monte Carlo value. The red line shows a linear fit. (b) Difference between the radio  $X_{\max}$  and the true Monte Carlo value, the distribution is fitted with a Gaussian with  $\mu = 1.6 \text{ g/cm}^2$  and  $\sigma = 17.2 \text{ g/cm}^2$ .

#### 5.4.1 Adding background and detector simulations

In the presence of background and considering the uncertainties introduced by the detector simulation, the estimation of the radio signal and the energy fluence become more complex affecting the reconstruction quality. This reflects on the number of events for which the radio shower maximum is successfully reconstructed in both methods. Only for 43 events in the simulation data set both measurements of the radio  $X_{\max}$  are available. Their comparison is presented in Figure 5.24. The yellow points mark the events for which  $X_{\max}^{\text{b}}$  is determined using a single station. The overall correlation factor is 0.5 and increases to 0.76 when these single station measurements are not considered. The fit line above the diagonal indicates that the  $X_{\max}$  estimated from the spectral slope is generally higher than the  $X_{\max}$  estimated from the energy distribution.

Also in the presence of noise, the energy distribution fit provides a better precision for  $X_{\max}$ . Hence, the weighed average of the two estimators returns a resolution of  $8.8 \text{ g/cm}^2$ , compatible with the stand-alone energy fit method, as visible from Figure 5.25.

As for the ideal case scenario, a radio  $X_{\max}$  estimator is created by combining the average of the two measurements when it is available and the single measurement when only one  $X_{\max}$  reconstruction was successful. The performance of the combined radio  $X_{\max}$  is shown in Figure 5.26. The overall  $X_{\max}$  resolution is  $25.13 \text{ g/cm}^2$ .

#### 5.4.2 Conclusions

In this section the potential of a radio mass composition observable, based on the combination the radio energy distribution on the ground and the slope of the frequency spectrum at each individual radio station has been investigated.



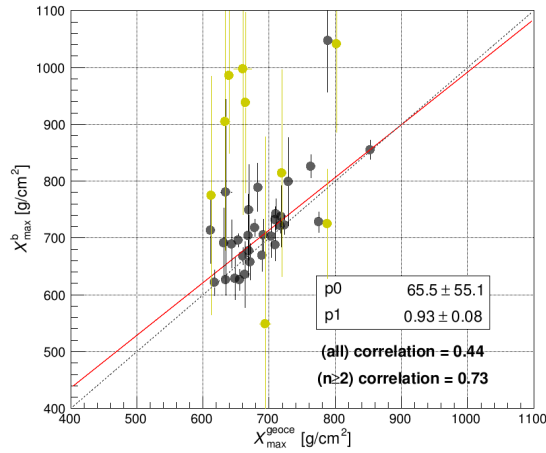


Figure 5.24: Comparison between the radio  $X_{\max}$  obtained from the fit of the energy signal distribution at the ground with the GeoCE parametrization and the one obtained from the parametrization of the spectral slope after adding background noise. The yellow points indicate the events for which  $X_{\max}^{\text{b}}$  is determined using a single station. The dashed gray line marks the diagonal. The red line is a linear fit to all points.

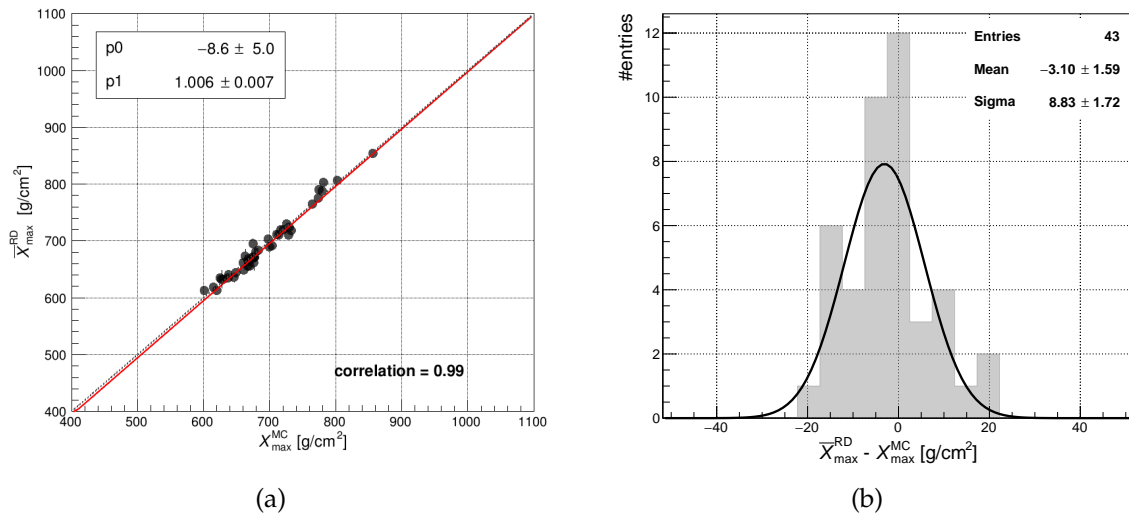


Figure 5.25: (a) The averaged radio  $X_{\max}$  versus the true Monte Carlo value. The gray dashed line marks the diagonal. The red line shows a linear fit. (b) Difference between the averaged value  $X_{\max}$  and the true Monte Carlo value, the distribution is fitted with a Gaussian with  $\mu = -3 \text{ g/cm}^2$  and  $\sigma = 8.8 \text{ g/cm}^2$ .

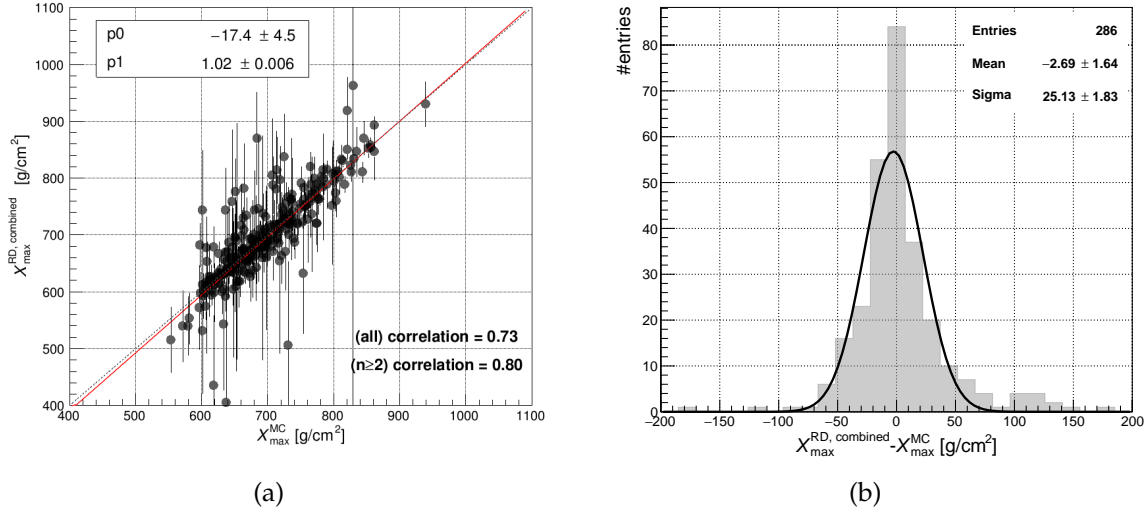


Figure 5.26: The radio  $X_{\max}$  is estimated with the weighted average of the two estimations when possible, otherwise the single measurement is taken. (a) The radio  $X_{\max}$  value versus the true Monte Carlo value. The gray dashed line marks the diagonal. The red line shows a linear fit. (b) Difference between the radio  $X_{\max}$  and the true Monte Carlo value, the distribution is fitted with a Gaussian with  $\mu = -2.7 \text{ g/cm}^2$  and  $\sigma = 25.13 \text{ g/cm}^2$ .

The radio  $X_{\max}$  measurements are independent and the obtained values are strongly correlated, therefore they can be combined in a weighted average using Eq. 5.20. Furthermore, a combined radio  $X_{\max}$  estimator is constructed by using the average  $X_{\max}$  when both methods are successful, and the single  $X_{\max}$  measurement otherwise.

The precision achieved reconstructing  $X_{\max}$  by fitting the GeoCE energy distribution is better than the one obtained using the parametrization of the spectral index. Hence, the average  $X_{\max}$  reflects the performance of the stand-alone two-dimensional fit of the radio footprint. As expected, the combined radio  $X_{\max}$  shows a small improvement with the respect to the precision obtained by the slope of the frequency spectrum, providing an  $X_{\max}$  resolution of about  $17 \text{ g/cm}^2$  in the ideal case scenario, and of about  $25 \text{ g/cm}^2$  when measured noise and background are included in the event reconstruction. These results are theoretical lower limits to the achievable resolution for experimental AERA data, given that they have been obtained using the true Monte Carlo values for the arrival direction and the core position, and that the fit of the radio energy distribution will be largely affected by sparse sampling of the AERA array.

## 6 Mass composition measurement with AERA

In this chapter, a mass composition analysis using the data measured by the AERA array is presented. Two independent estimations of the depth of the shower maximum are inferred from the radio signal using respectively energy or spectral information. The methods, detailed in Chapter 5, are applied to the SD-RD hybrid data set. As a cross-check the reconstructed radio  $X_{\text{max}}$  values are compared to measurements of the shower maximum by the fluorescence detector. Lastly, to explore the composition of the cosmic ray flux, the evolution of the average radio  $X_{\text{max}}$  is studied as a function of the primary energy.

### 6.1 Data reconstruction and event selection

The data set used for the physics analyses in this thesis runs from 02 June 2013 to 31 December 2019 and is composed only of events measured by radio antennas equipped with the KIT/BUW electronics. Measurements from Butterfly and LPDA antenna types are combined in the analyses. The arrival direction of the shower, the core impact location and the energy are inferred from the surface detector data in an hybrid event reconstruction performed within the *Offline* framework. When available, also the data from the fluorescence detector is analyzed in order to have an event by event comparison of the measurement of the radio and the fluorescence shower maximum.

Radio data has been processed using the *Offline* standard application `RdObserver` introduced in Section 3.4.1 with some modifications from the standard configuration:

- the electric field measured by Butterfly stations is unfolded using the simulated antenna response pattern, as explained in Chapter 4;
- the calibration procedure described in Chapter 4 is applied during the data reconstruction;
- the amplitude of the electric field and the energy fluence of each antenna are determined using the procedure described in Chapter 3;
- two sources of systematic uncertainty are considered to affect the measurement of the electric field amplitude: 5% uncertainty introduced by the antenna-to-antenna fluctuation, derived in Section 4.1.1, and 2.5% introduced by the signal chain of the antennas [32]. Both contributions are equal for the two antenna types under consideration. These uncertainties are added in quadrature to the signal uncertainty resulting from background subtraction, and they are also propagated into the estimation of the uncertainty of the energy fluence.

The *Offline* module sequence is given in Appendix A.

When the cosmic-ray event is successfully reconstructed all relevant information is stored in a single output file, from the shower geometry to the electric field trace and the energy fluence in all polarization directions for each radio station.

### 6.1.1 Station rejection

A radio event is considered to be a cosmic-ray candidate when it matches an SD-event in time and location. In addition, it is required that at least three radio stations have a signal to noise ratio above 10 (the signal to noise is defined in Eq. 3.32). Stations that show deficiencies in hardware or detect a significantly high ambient noise are listed in the *bad period* database and rejected in the first stage of the reconstruction. Nevertheless, some, or even all, stations of these pre-selected events are still likely to contain non cosmic-ray signals. In order to identify and reject noise pulses the selection criteria listed below are required during the event reconstruction.

A useful benchmark to distinguish between a cosmic-ray-like signal and a noise pulse is the duration of the radio pulse. Generally, noise pulses are longer than cosmic-ray signals and can last 100 ns, while the duration of typical cosmic-ray pulses is about 20 ns [72]. Therefore, one of the noise rejection algorithms is based on the pulse shape. If the magnitude of the measured electric field exceeds a certain threshold more than 3 times in a time window of 50 ns, either before or after the electric field pulse maximum, the signal is classified as a noise pulse and the station is rejected.

Background signals are also identified using the expected polarization of the electric field at each station, more precisely the polarization angle, i.e. the angle of the total electric field vector with respect to the  $\vec{v} \times \vec{B}$ -axis. The measured polarization angle is compared to the expected polarization angle inferred using the position of the station relative to the shower geometry known from the surface detector reconstruction [72]. Noise pulses are randomly polarized and their polarization angle usually does not agree with the expected polarization angle value. A station is rejected when the measured polarization angle is not included in the interval  $[\phi_{\min} - 2\sigma, \phi_{\max} + 2\sigma]$ , where  $\sigma$  is the uncertainty on the measured polarization angle, and  $\phi_{\min}$  and  $\phi_{\max}$  are the estimations for the expected polarization angle obtained taking into account the uncertainty on the core position and the relative charge excess strength. This check is also effective to reject events with a mis-reconstructed SD geometry.

Stations that are isolated from other signal stations are also rejected from the event reconstruction: if a station that shows a signal-like pulse is located more than 400 m away from any other station with a signal, or it finds only one other station within 800 m, it is very likely that the radio pulse is generated by a transient noise event and the station is rejected.

The last method to identify noise pulses is based on the pulse arrival time. The incoming direction of the air shower is reconstructed using a plane wave fit. At first only the three stations closest to the shower axis are included in this fit, then, in a loop, the next-closest station is included, until all stations have been considered. As a noise pulse will likely have a signal time incompatible with those of the other stations, including a noise station in the fit will cause a substantial decrease of the  $\chi^2$  probability, or the fit will not converge. When the  $\chi^2$  probability is lower than 5% the last added station is rejected permanently.

When three or more radio stations survive the aforementioned noise rejection cuts, the radio event is reconstructed.

### 6.1.2 Event selection

Among the reconstructed events, the ones used in the physics analyses have to fulfill the following requirements.

- The zenith angle has to be below  $55^\circ$ . This cut is imposed to guarantee a fully efficient reconstruction of the 750 m array of the SD. Additional quality cuts on the data of the surface detector from Ref. [73] are also applied. An event is selected if the signal times from the three stations with the largest signals are in agreement with the propagation of a plane shower front, and if the detector station with the highest signal is enclosed in a hexagon of six active stations.
- The reconstructed cosmic ray directions using the radio and surface detector have to agree within  $10^\circ$ . The radio arrival direction is reconstructed through a plane wave fit. This cut is not related to the angular resolution of the detectors, but it reduces the number of uncorrelated events originating from noise pulses coming from the horizon.
- Events detected during thunderstorm conditions are rejected. Ambient electric fields influence the radio emission from air showers. In particular, during thunderstorms the signal might be amplified up to an order of magnitude [74, 75]. The strength of the ambient electric field is measured by electric field mills located in weather stations which are placed within the array [76]. Events are rejected when the mean electric field calculated for periods of 30 minutes is lower than  $-150$  V/m or larger than  $50$  V/m, or when its root mean square is above  $30$  V/m, according to the study described in Ref. [77]. Also a second algorithm defined in Ref. [78] is used to distinguish between fair weather conditions and thunderstorm conditions. Events for which no weather station data is available are used in the physics analyses because the chance that they are affected by thunderstorms is reduced by the cut on the polarization of the electric field.

The complete list of cuts is reported in Appendix A. An overview of the number of selected events from the SD-RD data set is listed in Table 6.1. The distribution of the arrival directions and of the energy of the selected cosmic-ray events as measured with the surface detector are shown in Figure 6.1.

Table 6.1: Number of RD events after the selection cuts.

SD-RD successfully reconstructed	18227
selected SD-RD	6232
Thunderstorm cut	5638
(of which no Weather information	542)

### Selection of fluorescence detector events

The data available for the multi-hybrid SD-RD-FD data set stops in December 2018 because the database containing aerosol measurements was not yet up-to-date for 2019 at the time of

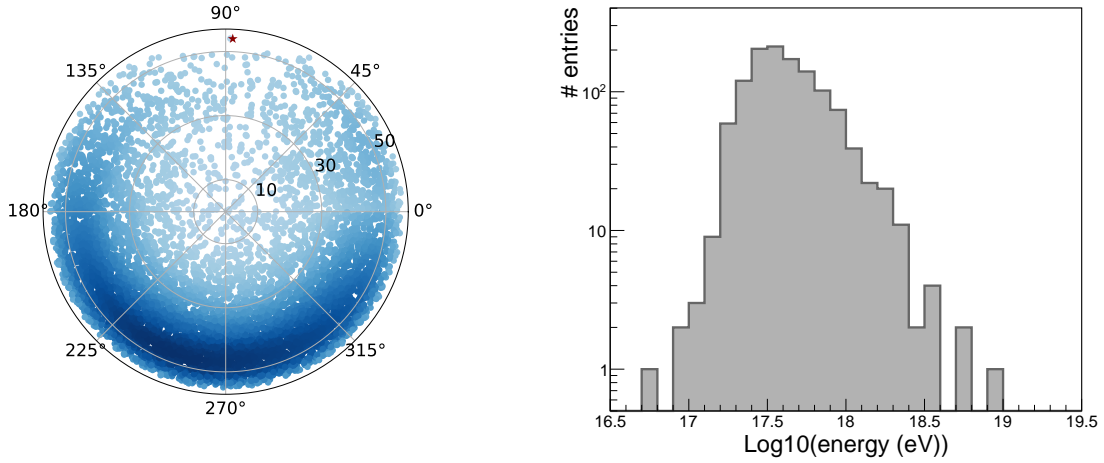


Figure 6.1: Distributions of the air shower parameters for the SD-RD selected data set after the thunderstorm rejection. Left: arrival direction from the SD reconstruction, the red star marks the direction of the magnetic field vector  $\vec{B}$ . Right: energy distribution from the SD reconstruction.

writing this thesis. Among the fluorescence cuts, fiducial cuts are normally applied to reduce the bias effects in the average value of  $X_{\max}$ . In particular,  $X_{\max}^{\text{FD}}$  has to be in the field of view of the telescopes with an upper limit of 40 g/cm<sup>2</sup> for its uncertainty. However, in this work, the FD measurement is only used for an event by event comparison and not to calculate the average value of  $X_{\max}$ , therefore, to increase the available statistics also events for which the fit of the Gaisser-Hillas function is not in the field of view of the telescopes or have an uncertainty up to 100 g/cm<sup>2</sup> are considered.

All cuts are listed in Appendix A, the number of selected events is summarized in Table 6.2. The distributions of the arrival direction, the cosmic-ray energy as measured by the SD and the  $X_{\max}$  as measured by the fluorescence detector are shown in Figure 6.2

Table 6.2: Number of SD-RD-FD events after the selection cuts.

SD-RD selected - no thunderstorm (2013-2018)	5124
have FD	1386
selected FD ( $\Delta X_{\max} \leq 100\text{g/cm}^2$ )	188
$X_{\max}$ Field of View and $\Delta X_{\max} \leq 40\text{g/cm}^2$	105

## 6.2 Event by event reconstruction of the shower maximum

The two methods to estimate the depth of the shower maximum from the radio signals are applied to the selected hybrid events.

### 6.2.1 $X_{\max}$ obtained from the AERA energy distribution

The GeoCE parametrization of showers with  $D_{X_{\max}} \geq 550\text{g/cm}^2$  obtained using CoREAS Monte Carlo simulations after including measured background traces is used to fit the energy

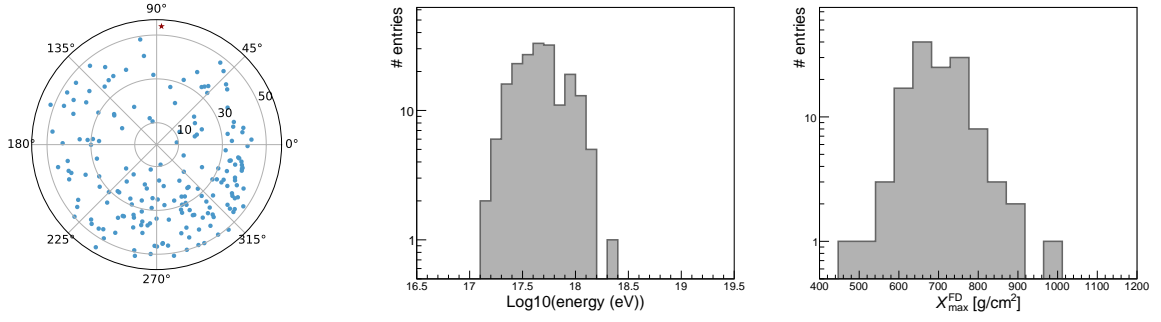


Figure 6.2: Distributions of the air shower parameters for the SD-RD-FD selected data set after thunderstorm rejection. From left to right: arrival direction from the SD reconstruction (the red star marks the direction of the magnetic field vector  $\vec{B}$ ), energy from the SD reconstruction and  $X_{\max}$  from the FD reconstruction.

fluence distribution in the shower plane. For each event, in addition to the distance to the shower maximum  $D_{X_{\max}}$  and the radiation energy  $E_{\text{rad}}$ , also the coordinates of the core position in the shower plane are considered free parameters in the fit of the radio energy distribution. Hence, only events with at least 5 signal stations are included in the analysis.

As a first check, the radio  $X_{\max}$  is compared to the measurements of the fluorescence telescopes. The cut on the number of signal stations and the fact that the GeoCE parametrization is efficient only for fully developed showers, reduces the number of events available for the comparison to 17, 7 of which have  $X_{\max}^{\text{FD}}$  in the field of view with an uncertainty lower than  $40 \text{ g/cm}^2$ . The number of events after each quality cut is listed in Table 6.3. Some examples of successful fits are shown in Figure 6.3.

Table 6.3: Number of events in the SD-RD-FD data set (2013-2018) after the quality cuts.

	SD-RD-FD	SD-RD-FD (FoV)
selected	188	105
$n \geq 5$	42	24
fit successful	29	9
$D_{X_{\max}} + \Delta D_{X_{\max}} \geq 550 \text{ g/cm}^2$	19	7
$P\chi^2 > 1\%$	19	7

The correlation coefficient between the radio and the fluorescence measurements of the shower maximum is 0.40 and increases to 0.70 for the high-quality FD selection. The scatter plot and the pull distribution are shown in Figure 6.4. The root mean square of the pull distribution is close to 1, therefore it is possible to conclude that for the few events available in the comparison the uncertainty on the radio reconstructed  $X_{\max}$  is estimated reasonably well.

### Average depth of shower maximum in AERA data

The energy fluence distribution is fitted for all selected events in the SD-RD data set. The number of events passing each quality cut is listed in Table 6.4. The SD 750 array is fully efficient from  $10^{17.5} \text{ eV}$  onward [79]. This cut is needed because the SD energy is used in the elongation rate.



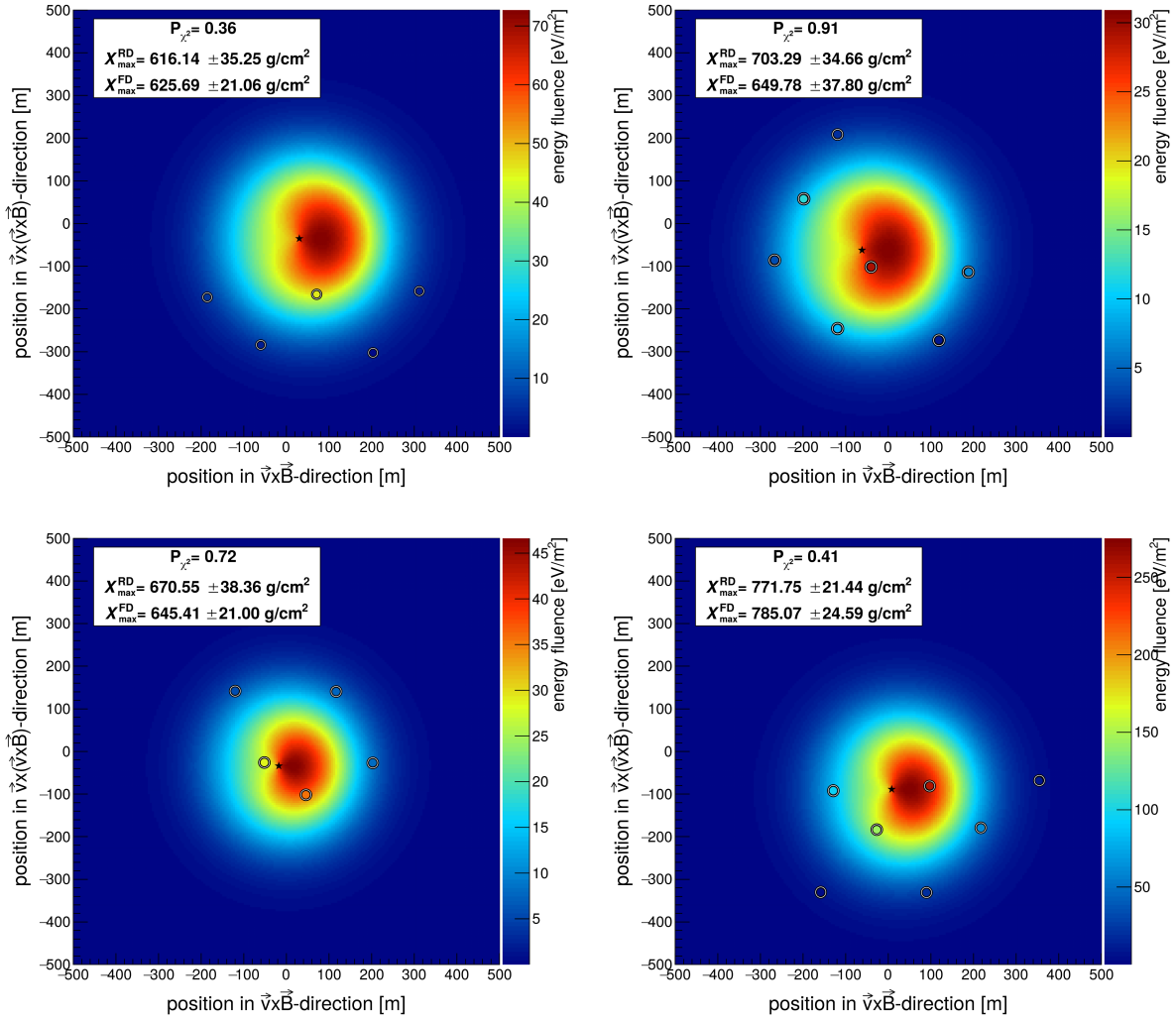


Figure 6.3: Examples of successful fits of the radio energy distribution for events for which also the FD measurement is available. The circles represent the signal stations and the color inside the circles indicate the measured energy fluence. The star marks the fitted value of the core position. The background shows the interpolation of the best fit function Eqs. 5.11–5.13).

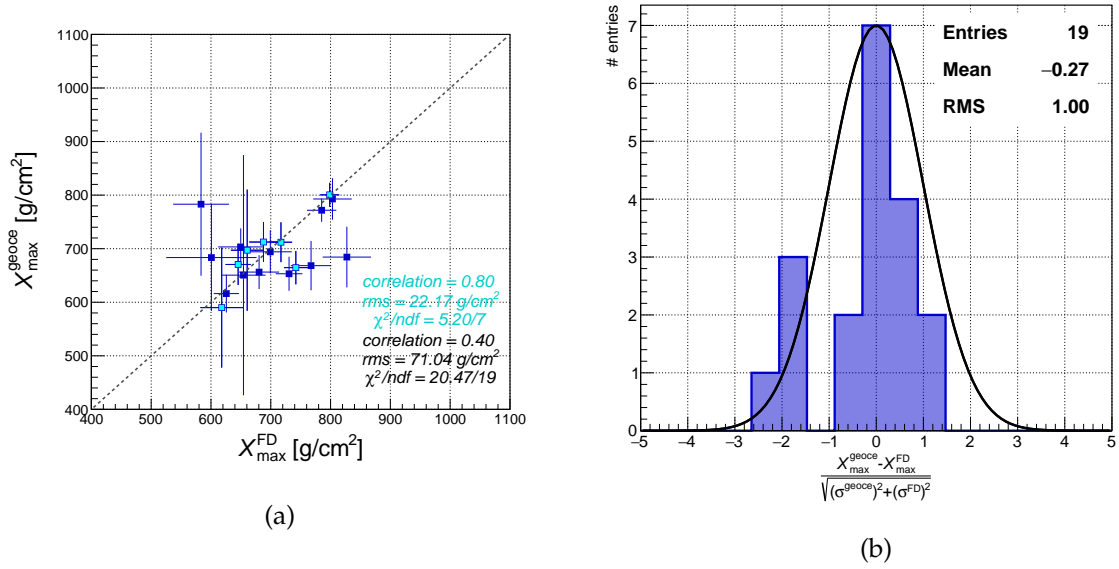


Figure 6.4: (a) Comparison between  $X_{\max}$  measured by the FD and the best fit parameters of the radio energy fluence distribution. The light blue points have the  $X_{\max}^{\text{FD}}$  in the field of view. The gray dashed line marks the diagonal. The values of the root mean square and the  $\chi^2$  are calculated with respect to the diagonal. (b) Pull distribution of the shower maximum from the fit of the radio energy distribution compared to the value measured by FD. A standard normal distribution is drawn for reference.

Table 6.4: Number of events in the SD-RD data set (2013-2019) set after the quality cuts.

selected	5638
$n \geq 5$	3863
fit successful	1331
$\text{Log}(E_{\text{SD}}/\text{eV}) > 17.5$	896
$D_{X_{\max}} + \Delta D_{X_{\max}} \geq 550 \text{ g/cm}^2$	682
$P\chi^2 > 1\%$	674

The average uncertainty of the atmospheric depth of the shower maximum as a function of the number of signal stations is shown in Figure 6.5. The uncertainty decreases as the number of stations used the fit increases.

Having a method efficient only for fully developed showers introduces a composition bias which is energy and zenith angle dependent. More specifically, the average value of  $X_{\max}$  is underestimated and the effect is stronger at lower zenith angles, at which the showers traverse less atmosphere. Therefore the distribution of  $X_{\max}^{\text{geoce}}$  is not flat in zenith angle as shown in Figure 6.6. An average correction is applied for  $\sec \theta < 1.5$ . The slight tilt for more inclined showers with  $\sec \theta \geq 1.7$  is likely introduced by differences between the atmospheric model used and the true atmosphere, as this effect is not visible in the Monte Carlo studies.

The energy dependent correction is a function of the composition and it can be evaluated using Monte Carlo simulations. However, due to the low Monte Carlo statistics available (see Table 5.5), it is not possible to accurately parameterize the uncertainty as a function of the composition in each energy bin, therefore a systematic uncertainty has been assigned, corresponding to the bias on the average value of  $X_{\max}$  in the hypothesis of a pure proton or pure iron composition. The bias is estimated using only the Monte Carlo value of  $X_{\max}$  and

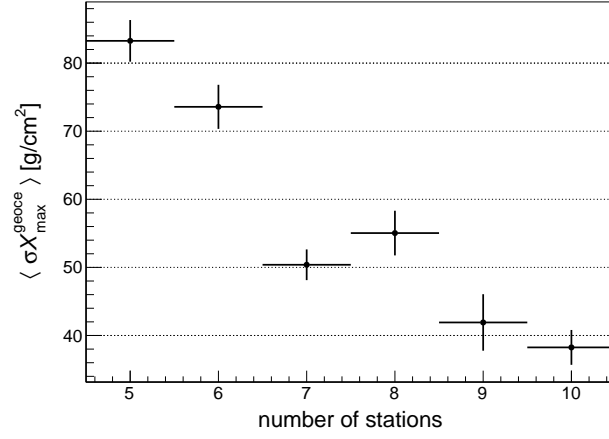


Figure 6.5: Average uncertainty on the  $X_{\text{max}}^{\text{geoce}}$  as a function of the number of signal stations used in the fit of the energy density distribution.

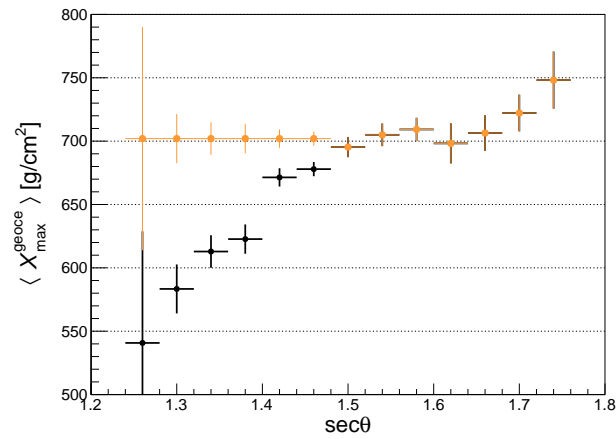


Figure 6.6: Average depth of shower maximum  $X_{\text{max}}$  as a function of the zenith angle  $\theta$ . The error bars represent the uncertainty on the average value. The measured  $X_{\text{max}}$  is shown in black: the bias for more vertical showers is introduced by the cut on fully developed showers with  $D_{X_{\text{max}}} \geq 550$  g/cm<sup>2</sup>. A correction based on average value between  $1.5 \leq \sec\theta < 1.7$  is applied to the reconstructed values and the result is shown in orange.

$D_{X_{\max}}$  and it results in an asymmetric uncertainty interval of  $[-2\%\langle X_{\max}\rangle, +3\%\langle X_{\max}\rangle]$ .

Another systematic uncertainty is introduced by the atmospheric model used in the simulations that differs from the one used in the data reconstruction. Such an effect shows up in the zenith-angle distribution as the amount of atmosphere traversed depends on this angle. From the distribution of the average value of  $X_{\max}$  as a function of the zenith angle, shown in orange in Figure 6.6, a systematic uncertainty of  $6.7 \text{ g/cm}^2$  is estimated assuming a flat distribution. The GeoCE parametrization of the energy distribution itself also creates a systematic uncertainty. Because of the limited number of events in the RD-FD data set it is difficult to determine the uncertainty of the GeoCE parametrization from the comparison with data. Instead the bias of  $5 \text{ g/cm}^2$  in  $X_{\max}^{\text{geoce}}$  resulting from the application of the GeoCE parametrization to the simulations determined in Section 5.2.4 is used. All contributions to the systematic uncertainty are listed in Table 6.5. The total systematic uncertainty is obtained by combining all contributions in quadrature.

Table 6.5: Overview of the systematic uncertainties of the average depth of shower maximum for the energy bins with at least 10 entries.

$\text{Log}(E_{\text{SD}}/\text{eV})$	17.5	17.8	18.0	18.2
Atmospheric Model	$\pm 6.7 \text{ g/cm}^2$	$\pm 6.7 \text{ g/cm}^2$	$\pm 6.7 \text{ g/cm}^2$	$\pm 6.7 \text{ g/cm}^2$
GeoCE Parametrization	$\pm 5 \text{ g/cm}^2$	$\pm 5 \text{ g/cm}^2$	$\pm 5 \text{ g/cm}^2$	$\pm 5 \text{ g/cm}^2$
$D_{X_{\max}} \geq 550 \text{ g/cm}^2$	$-13.8 \text{ g/cm}^2$ $+20.6 \text{ g/cm}^2$	$-14.2 \text{ g/cm}^2$ $+21.4 \text{ g/cm}^2$	$-14.2 \text{ g/cm}^2$ $+21.4 \text{ g/cm}^2$	$-14.2 \text{ g/cm}^2$ $+21.4 \text{ g/cm}^2$
Total	$-16.1 \text{ g/cm}^2$ $+22.3 \text{ g/cm}^2$	$-16.5 \text{ g/cm}^2$ $+22.9 \text{ g/cm}^2$	$-23.0 \text{ g/cm}^2$ $+22.9 \text{ g/cm}^2$	$-16.6 \text{ g/cm}^2$ $+22.9 \text{ g/cm}^2$

The average value of  $X_{\max}^{\text{geoce}}$  as a function of the energy reconstructed from the SD is plotted in Figure 6.7, where it is compared to measurements of shower maximum made by the fluorescence and the surface detector (see also Figure 1.5). Only bins with at least 10 events are shown in the plot, this cut removes one point with 9 events at  $10^{18.4} \text{ eV}$ , and two points with one event each at energies above  $10^{18.6} \text{ eV}$ . The average  $X_{\max}$  is weighted using the uncertainties on the reconstructed  $X_{\max}^{\text{geoce}}$  and the error bar shows the uncertainty of the weighted average. The average value of the depth of shower maximum measured from the GeoCE parametrization of the radio energy distribution is compatible with the measurements of the fluorescence and surface detector of the Pierre Auger Observatory. The average  $X_{\max}^{\text{geoce}}$  at  $10^{18.2} \text{ eV}$  is slightly lower than the Auger result. To investigate possible statistical fluctuations the same plot has been studied using 16 energy bins. Figure 6.8 indicates that the effect is not related to statistical fluctuation. It is likely to be a consequence of the cut on  $D_{X_{\max}} \geq 550 \text{ g/cm}^2$  that becomes more important at higher energies. To understand this trend and to accurately study for the composition bias, a larger Monte Carlo sample generated using the GDAS atmospheric model is needed.

### 6.2.2 $X_{\max}$ obtained from the spectral index

In the Monte Carlo studies described in Section 5.3.3, taking advantage from the large number of stations available in the star-pattern simulations, only stations with a  $\text{SNR}_{\text{amp}} \geq 5$  have been included in the fit to estimate the radio  $X_{\max}$ . This same selection applied to data events would drastically reduce the number of stations available for the fit and the number of successful fits.

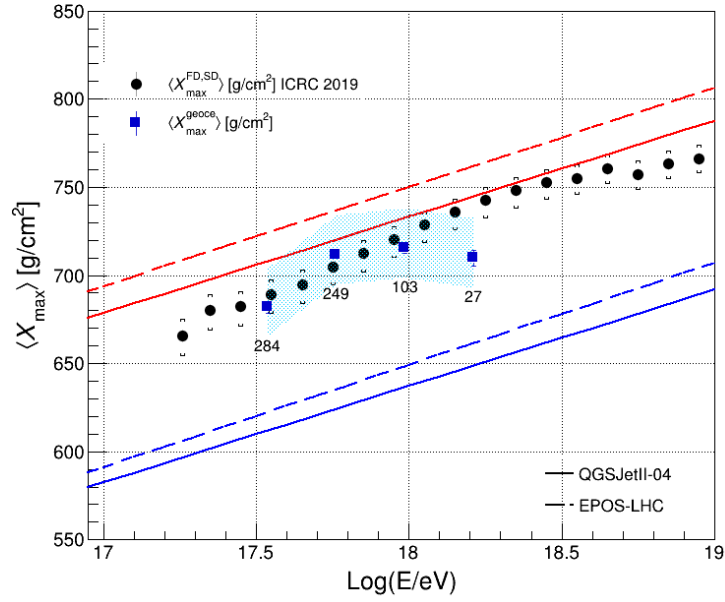


Figure 6.7: Average depth of shower maximum as measured by AERA by fitting the energy fluence distribution as a function of the SD reconstructed energy. The values are compared to the measurements from the Pierre Auger observatory (black points) [80, 81]. The error bars on the AERA data points represent the error of the weighted average. The colored band indicates the systematic uncertainty. Predictions of the depth of shower maximum of iron and proton primaries from two different interaction models are plotted as lines. The number of entries per energy bin is reported. Only points in energy bins with 10 or more entries are shown.

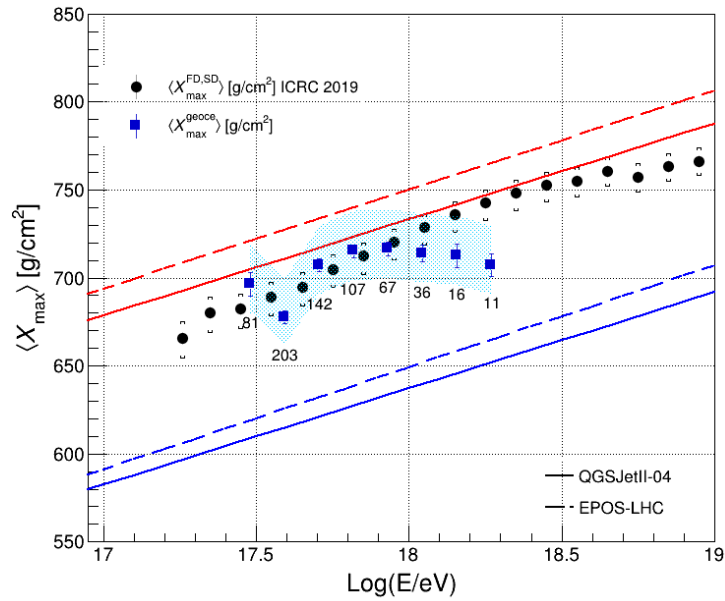


Figure 6.8: Average depth of shower maximum as measured by AERA by fitting the energy fluence distribution as a function of the SD reconstructed energy doubling the number of energy bins. Only bins with at least 10 entries are shown.

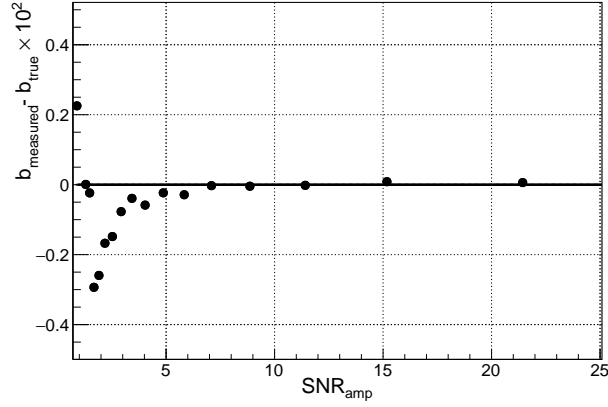


Figure 6.9: Average difference between the real spectral index  $b$  obtained in the ideal case scenario and the spectral index  $b$  measured after the addition of measured noise traces and the application of the antenna model as a function of  $\text{SNR}_{\text{amp}}$ .

However, for small SNRs a systematic bias in the slope parameter measurement is introduced. Therefore, the fit value of the spectral index is corrected according to the station  $\text{SNR}_{\text{amp}}$  to be able to include all signal stations in the data analysis.

The correction to the spectral index is evaluated using the Monte Carlo sample by comparing the spectral index obtained in the ideal case scenario with the one obtained after including measured background traces and the antenna model in the simulations. The average bias as a function of the SNR is shown in Figure 6.9. The correction is negligible for  $\text{SNR}_{\text{amp}} \geq 5$ , which is in agreement with the study shown in Figure 3.11. Figure 3.11 also indicates that the uncertainties on the amplitude of the electric field are overestimated when the signal to noise ratio is low, as the width of the pull distribution is always smaller than 1. For this reason, assuming that the model used to describe the frequency spectrum is correct, the uncertainty on the fitted value of the spectral slope is corrected such that the reduced  $\chi^2$  is 1.

When signal stations are located in the region of validity of the parametrization, their corrected value of the spectral slope  $b$  and its uncertainty are used in the iterative  $\chi^2$  minimization procedure described in the Section 5.3 to infer the geometrical distance to the shower maximum. The number of events for the SD-RD-FD data set and for the SD-RD data for which the fit is considered successful is listed in Table 6.6.

Table 6.6: Number of events after the quality cuts.

	SD-RD	SD-RD-FD (FoV)
selected	5638	105
fit successful	4122	76
$\text{Log}(E_{\text{SD}}/\text{eV}) > 17.5$ [79]	2375	not applied

An energy dependent bias is observed in the reconstructed distance to the shower maximum. This bias can be quantified making use of Monte Carlo events reconstructed including measured background traces and the antenna response model. For each event  $D_{X_{\text{max}}}$  is estimated selecting randomly one station, 2 stations, 3 stations and a number of stations between 4 and 10. The selection criteria for the stations are the same as applied in the data reconstruction, therefore also stations with SNR below 5 are included in the estimation of  $D_{X_{\text{max}}}$ . The average

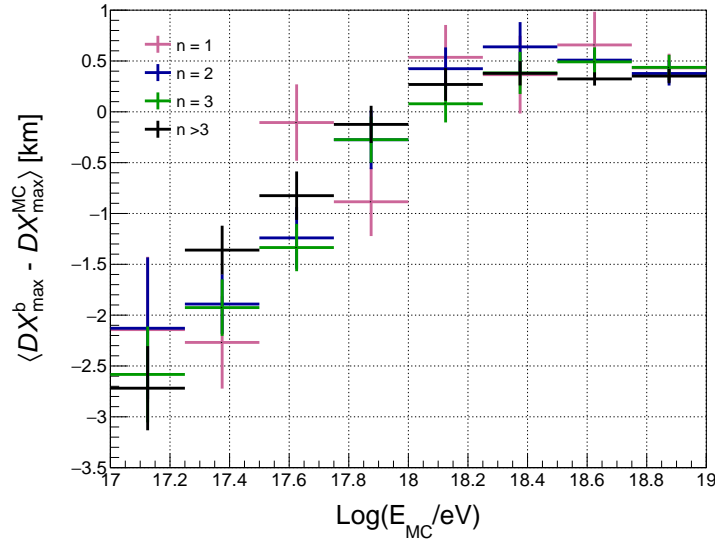


Figure 6.10: Average bias of the  $D_{X_{\max}}$  as a function of the energy and of the number of stations for simulated events obtained using one, two, three or more randomly selected signal stations.

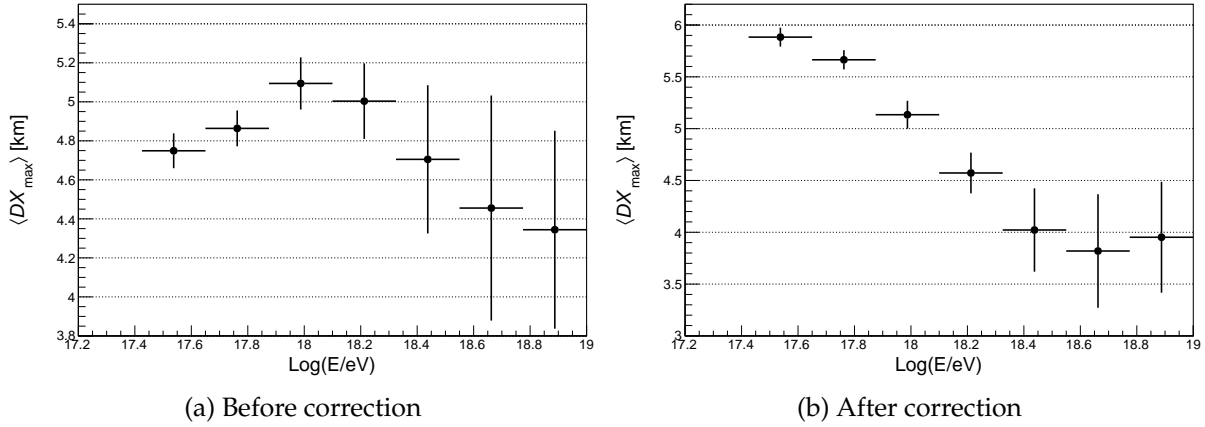


Figure 6.11: Average  $D_{X_{\max}}$  best fit value obtained from the fit of the spectral slope as a function of the energy ((a)) before and ((b)) after the bias correction estimated from simulations.

deviation as a function of energy and number of stations used in the fit is shown in Figure 6.10. The reason for this bias is not clear. An hypothesis is that for events with an energy below  $10^{18}$  eV only the upward radiation energy fluctuations of the real event distribution survive the selection criteria, while a large fraction of the other showers at these energies produces pulses which have too low power. In order to investigate this hypothesis a larger simulation data set is needed.

Using the average bias estimated from the Monte Carlo data set, it is possible to correct the fitted value of  $D_{X_{\max}}$ . The average distributions of the fitted value of  $D_{X_{\max}}$  before and after the shift are shown in Figure 6.11. The geometrical distance to the shower maximum is converted to  $X_{\max}$  using the GDAS atmospheric model.

An additional uncertainty on the reconstructed  $X_{\max}^b$  is introduced by the uncertainty of the core position, which in the fit is fixed to its value reconstructed by the SD detector. The



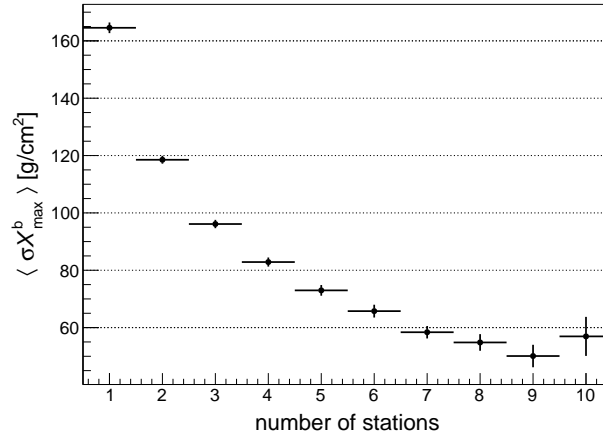


Figure 6.12: Average uncertainty on the  $X_{\max}^b$  as a function of the final number of stations for used in the  $\chi^2$  algorithm. The uncertainty decreases when more stations are available in the fit.

contribution to the uncertainty is quantified using Monte Carlo simulations. Once again, for each event  $X_{\max}^b$  has been inferred selecting randomly a single station, 2 stations, 3 stations and a number of stations between 4 and 10. Each fit is repeated several times varying the coordinate of the core positions within the correlated  $x - y$  uncertainties, both fixed to 20 m. The obtained relative variation on the reconstructed  $X_{\max}$  in 100 variations of the core position is independent from the number of stations used in the fit and corresponds to a 3% additional uncertainty on  $X_{\max}$ . This value is propagated into the uncertainty of  $X_{\max}^b$ .

The uncertainty on  $D_{X_{\max}}$ , and therefore on  $X_{\max}$ , depends on the number of stations that are included in the fit of the spectral slope parametrization. In Figure 6.12 the average uncertainty of the atmospheric depth of shower maximum is plotted as a function of the final number of stations used in the fit of the parametrization of the spectral index. As expected, the uncertainty decreases as the number of stations per event in the reconstruction increases. For  $n \geq 4$  the uncertainty on  $X_{\max}$  becomes comparable with the uncertainty obtained fitting the energy fluence distribution shown in Figure 6.5.

The next step consists of cross-checking the event-by-event reconstruction of the shower maximum obtained from the parametrization of the spectral index method to the FD measurements of  $X_{\max}$ . Some examples of events where Eq. 5.18 is minimized are plotted in Figures 6.13 and 6.14. These results confirm that a reconstruction of the shower maximum using a single station is possible, but due to the large uncertainty of the measurement of the spectral index the uncertainty of the depth of shower maximum is accordingly large.

The correlation plot between  $X_{\max}^b$  and the fluorescence measurement of the shower maximum is shown in Figure 6.15: the correlation coefficient is 0.18 for all events and increases to 0.33 when considering only events with  $n \geq 3$ . The pull distribution evidences a slight bias towards underestimation of the radio depth of shower maximum measurement. The standard deviation of the distribution is close to unity, indicating that the uncertainties have been estimated reasonably well.

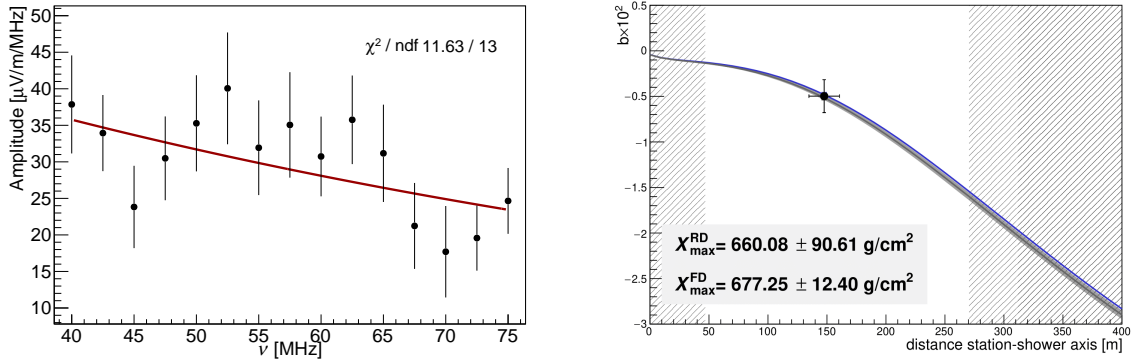


Figure 6.13: Example of an event containing one station that survived the selection procedure. On the left the fit of the electric field amplitude of the station used to estimate  $X_{\text{max}}^{\text{b}}$  is shown. On the right the spectral slope of the radio station (black point) is compared to the prediction of the parametrization obtained using  $X_{\text{max}}^{\text{FD}}$  (gray line). The parametrization obtained with the  $X_{\text{max}}^{\text{b}}$  is shown in blue. The regions of  $d$  in which the parametrization is invalid for the fitted value of  $D_{\text{max}}$ , measured from the SD shower core position, are marked with hatches.

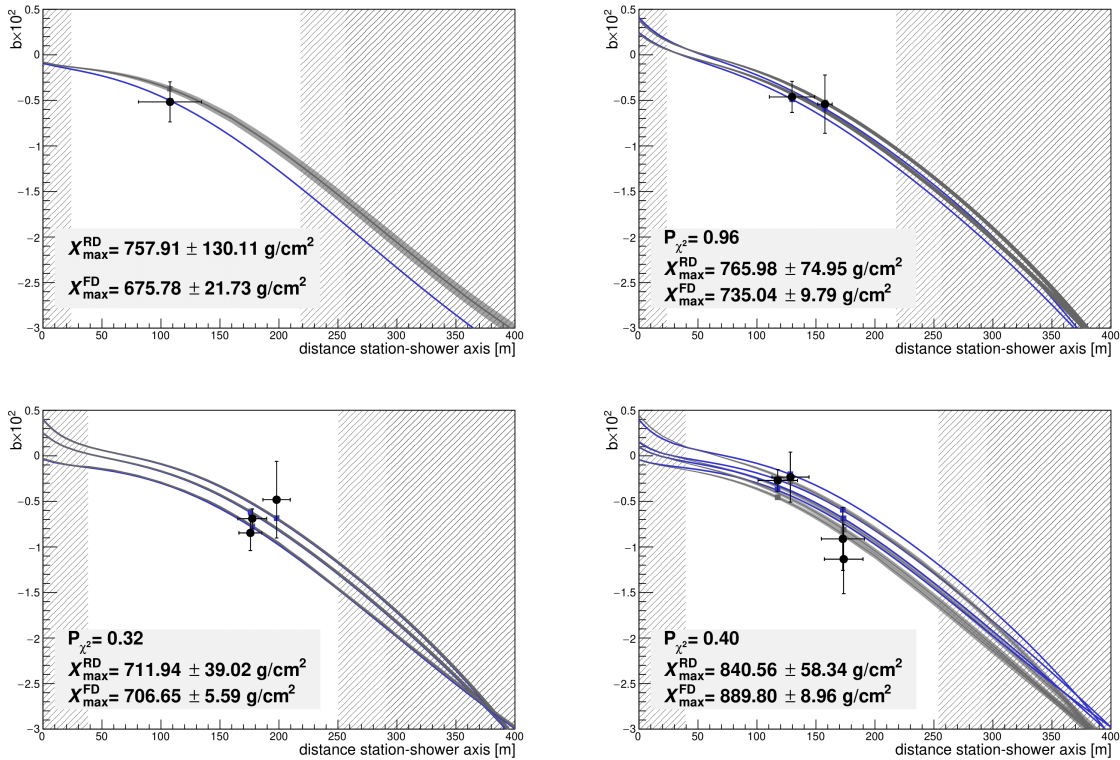


Figure 6.14: Examples of events that survived the selection procedure. The measured spectral index is plotted in black, and is compared to the parameterized spectral index as calculated from the value of  $X_{\text{max}}$  measured by FD (with the bands indicating the uncertainty) and the best fit of the parametrization to the measured index (blue line). These were calculated using the observer angle as measured from SD over a wide range of distances to the shower axis  $d$  to show their dependence. The regions of  $d$  in which the parametrization is invalid for the fitted value of  $D_{\text{max}}$ , measured from the SD shower core position, are marked with hatches.

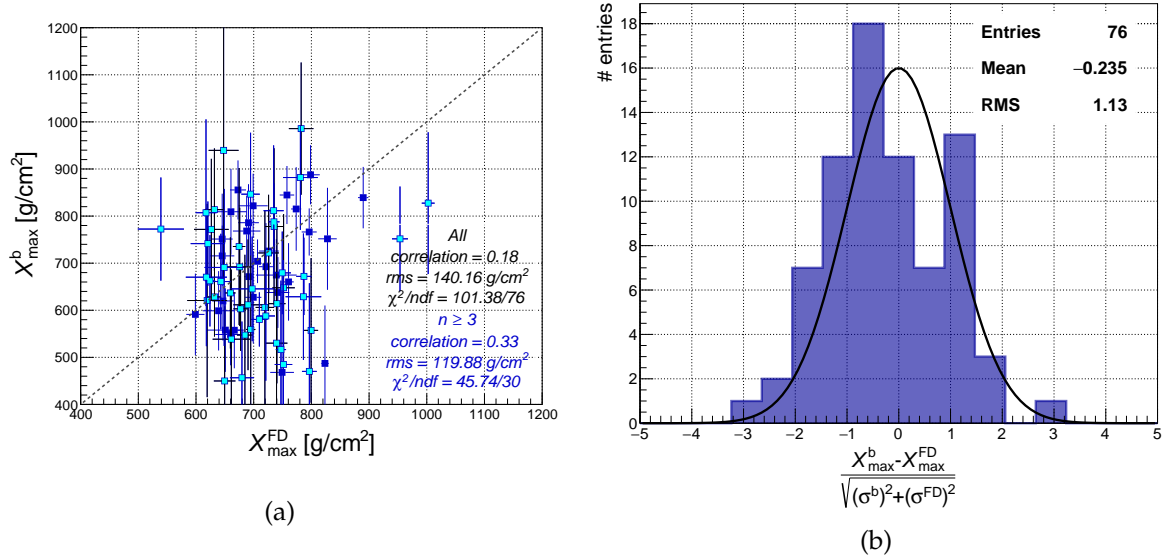


Figure 6.15: (a) Comparison between  $X_{\max}$  measured by the FD and radio  $X_{\max}$  obtained from the spectral slope method. The light blue points mark the events with one or two signal stations used in the fit to estimate  $X_{\max}^b$ . The gray dashed line marks the diagonal. The root mean square and the  $\chi^2$  are calculated with respect to the diagonal. (b) Pull distribution of the shower maximum from the parametrization of the spectral slope in comparison with the value measured by FD. A standard normal distribution is drawn for reference.

### Average depth of shower maximum in AERA data

In order to investigate the evolution of the average radio  $X_{\max}^b$  as a function of energy, the spectral slope method is applied to all selected SD-RD events. A clear dependence of  $\langle X_{\max} \rangle$  with the zenith angle is observed in Figure 6.16. This bias is in part introduced by the atmospheric model. Further investigations are needed to be able to understand it fully, a starting point would be to produce the parametrization of the spectral index on a Monte Carlo sample generated with the GDAS atmospheric model. Here, a systematic uncertainty has been assigned by assuming a flat distribution leading to an uncertainty of 31  $\text{g/cm}^2$ .

The weighted average value of  $X_{\max}^b$  as a function of energy is plotted in Figure 6.17 for the SD-RD events shown in Table 6.6. The error bars in Figure 6.17 indicate the uncertainty on the weighted average, while the contributions to the systematic uncertainty are depicted by the colored band. These contributions include the uncertainty due to the atmospheric model estimated from Figure 6.16, as well as the uncertainty on the method itself. The latter is based on the bias in the Monte Carlo analysis. The bias depends on the number of stations and its maximum is 12  $\text{g/cm}^2$  when using information from a single station. As a conservative estimate this value is used as systematic uncertainty. The two contributions to the systematic uncertainty are independent and are added quadratically.

The radio average  $X_{\max}$  obtained from the spectral information is compatible within the statistical uncertainties with the FD-SD measurements and consistent with a composition first becoming lighter and then slightly heavier again towards higher energies.

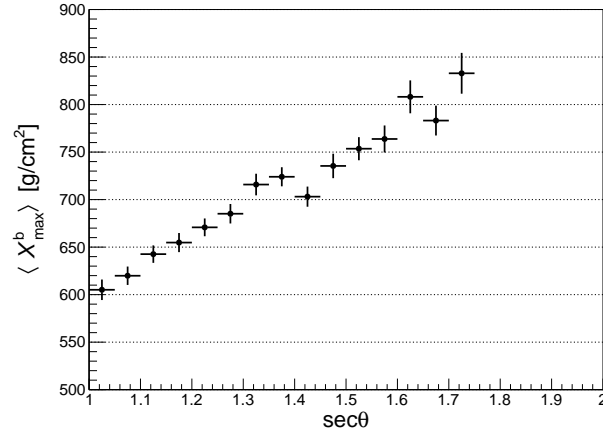


Figure 6.16: Average depth of shower maximum  $X_{\max}$  as a function of the zenith angle  $\theta$ . The error bars represent the uncertainty on the average value. A clear dependence of  $\langle X_{\max} \rangle$  with the zenith angle is observed.

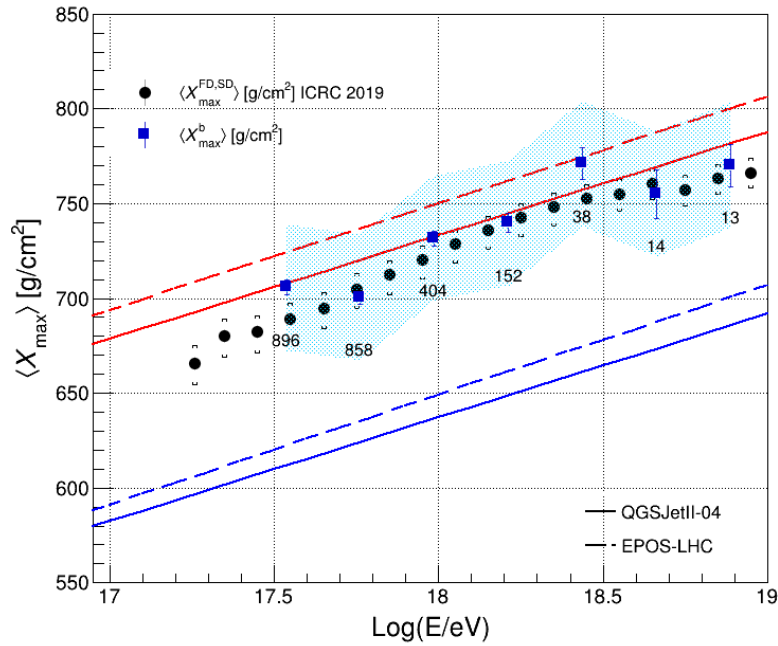


Figure 6.17: Average depth of the shower maximum as a function of the SD reconstructed energy as measured by AERA using the parametrization of the spectral slope. The band indicates the systematic uncertainty on the RD data set. The evolution of  $X_{\max}^b$  is compared to the measurements from the Pierre Auger Observatory (black points) [80, 81]. Predictions of the depth of shower maximum of iron and proton primaries from two different interaction models are plotted as lines. The number of events per energy bin is reported. Only points in energy bins with more than 10 entries are shown.

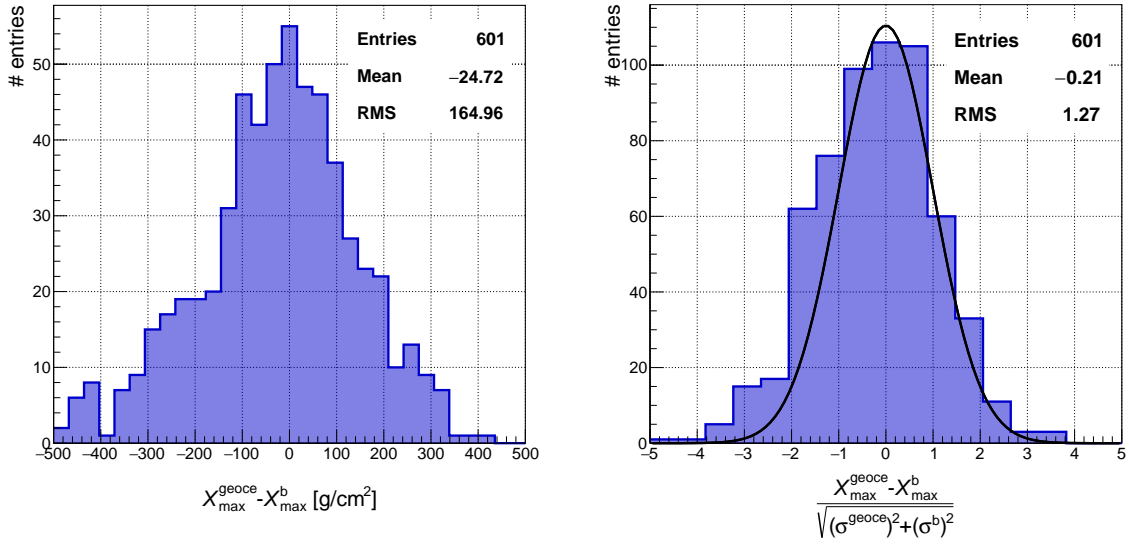


Figure 6.18: Comparison of the radio  $X_{\max}$  estimators. Left: distribution of the difference between the radio based  $X_{\max}$  values. Right: pull distribution of the reconstructed shower maximum for the fit of the energy fluence distribution in comparison with the shower maximum obtained from spectral information. The black line marks a standard normal distribution for reference.

### 6.3 Comparison and combination of the radio shower maximum estimators

In this section the two independent estimators of the depth of the shower maximum inferred using radio signals are compared and then combined in a single radio shower maximum estimator using the weighted average when both measurements are available, and otherwise considering the single successful measurement.

The comparison of the radio  $X_{\max}$  obtained with the two methods for the SD-RD data set is shown in Figure 6.18. The methods are simultaneously successful for 601 events. The events for which  $X_{\max}^b$  has been reconstructed with information from a single signal station are included in the comparison. A slight shift is observed. However, the values of the shower maximum obtained with the two methods are to a large part consistent. This is not a trivial statement as the radio information used is completely uncorrelated.

The comparison between  $X_{\max}$  measured by the FD and the combined radio  $X_{\max}$  estimator ( $\bar{X}_{\max}^{\text{RD}} \text{ XOR } X_{\max}^{\text{geoce}} \text{ XOR } X_{\max}^b$ ) is shown in Figure 6.19. The data set used for this comparison includes only events for which the reconstructed  $X_{\max}^{\text{FD}}$  is in the field of view of the telescopes. As expected, the weighted average reflects the performance of the method based on the fit of the energy fluence.

The final goal of the analysis is to investigate how the radio based cosmic-ray composition compares to the standard Auger measurements when more information from the radio signal is combined. The systematic uncertainties on the methods, respectively 5 g/cm<sup>2</sup> for the energy fit and 12 g/cm<sup>2</sup> for the spectral slope fit, are uncorrelated and result in an uncertainty of 4.6 g/cm<sup>2</sup> on the average. The uncertainties introduced by the atmospheric model are expected to be correlated. However, the parametrization of the spectral slope has been obtained with the US-standard atmospheric model, while the GeoCE parametrization of the energy footprint

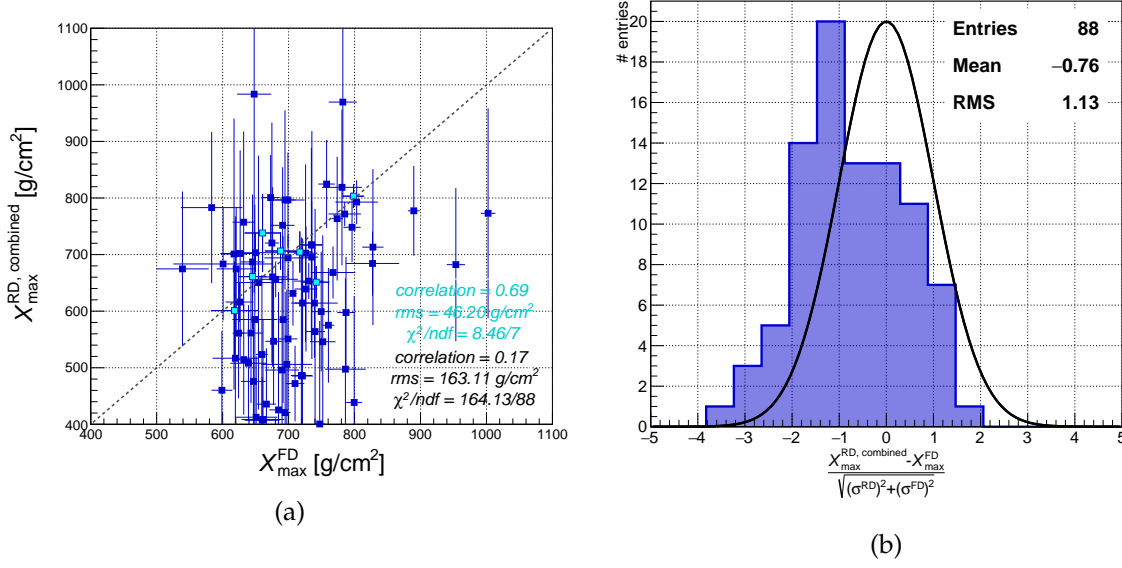


Figure 6.19: (a) Comparison between  $X_{\max}$  measured by the FD and combined radio  $X_{\max}$  ( $\bar{X}_{\max}^{\text{RD}}$  XOR  $X_{\max}^{\text{geoce}}$  XOR  $X_{\max}^{\text{b}}$ ). The light blue points mark the events for which the  $X_{\max}$  value corresponds to the weighted average of  $X_{\max}^{\text{b}}$  and  $X_{\max}^{\text{geoce}}$ . The gray dashed line marks the diagonal. The root mean square and the  $\chi^2$  are calculated with respect to the diagonal. (b) Pull distribution of the radio shower maximum obtained combined the two radio  $X_{\max}$  estimators in comparison with the value measured by FD. A standard normal distribution is drawn for reference.

was obtained with the October Malargue atmospheric model. Therefore, they are not expected to be fully correlated. To properly study the correlation one would need a simulation sample generated with the GDAS atmospheric model, the same as used in the data analysis. An empirical estimate can be obtained by exploiting the distribution of  $\langle X_{\max} \rangle$  as a function of the zenith angle shown in Figure 6.20(a). With the same approach as used for the two methods, one can assume a flat distribution. The uncertainty derived is  $14 \text{ g/cm}^2$ . The total systematic uncertainty also includes the uncertainty estimated for  $X_{\max}^{\text{geoce}}$  due to the cut on the fully developed showers that scales with  $\langle X_{\max}^{\text{geoce}} \rangle$ .

The elongation rate obtained using the weighted average of  $X_{\max}^{\text{geoce}}$  and  $X_{\max}^{\text{b}}$  is shown in Figure 6.21. A comparison of Figures 6.7 and 6.17 shows that the average values of the radio  $X_{\max}$  reflects the features of the stand-alone fit of the energy fluence distribution. The error bar indicates the error on the weighted average, while the colored band marks the systematic uncertainty.

Finally, the cosmic-ray composition can be studied with all reconstructed radio events using either the weighted average (when available) or a single radio  $X_{\max}$  estimator. The result is shown in Figure 6.22. The mean values are estimated through a weighted average to properly take into account the different uncertainties provided by the two methods. As for the average, the systematic uncertainty is calculated by considering the uncertainty on the two methods and a common uncertainty deduced from the distribution of  $\langle X_{\max}^{\text{RD,combined}} \rangle$  as a function of the zenith angle. From Figure 6.20 an uncertainty of  $18 \text{ g/cm}^2$  has been estimated.

Combining the information from the radio signal leads to a substantial boost of the number of radio events available for composition measurements. Furthermore, even though the

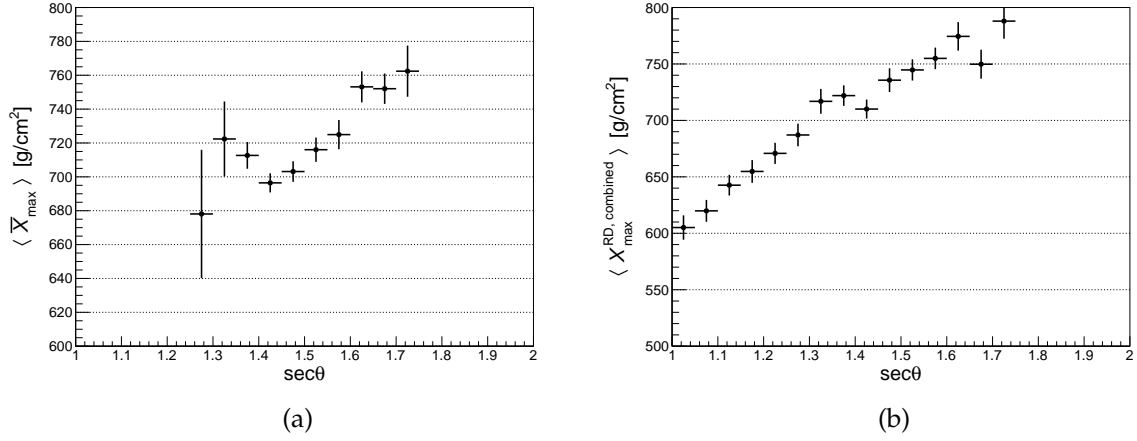


Figure 6.20: Average depth of shower maximum  $X_{\max}$  as a function of the zenith angle  $\theta$  for  $\bar{X}_{\max}^{\text{RD}}$  in (a) and for the combined radio  $X_{\max}$  estimator in (b). The error bars represent the uncertainty on the average value.

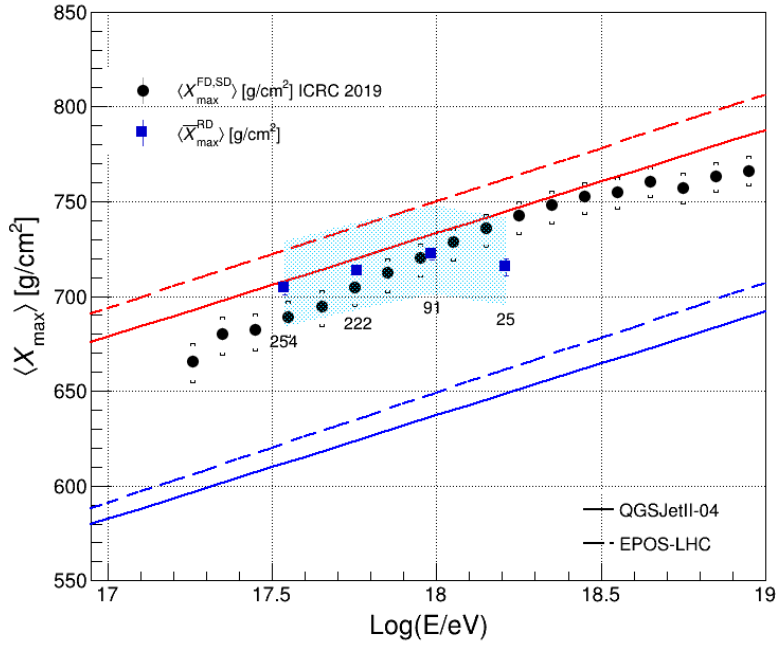


Figure 6.21: Average depth of shower maximum as a function of the SD reconstructed energy as measured by AERA combining the measurements from the spectral index and from the fit of the energy fluence distribution in a weighted average. The radio  $\langle X_{\max} \rangle$  is compared to the measurements from the Pierre Auger observatory (black points) [80, 81]. The error bars of the AERA data points represent the uncertainty propagated on the weighted average. The colored band indicates the systematic uncertainty. Predictions of the depth of shower maximum of iron and proton primaries from two different interaction models are plotted as lines. The number of entries per energy bin is reported. Only points in energy bins with 10 or more entries are shown.



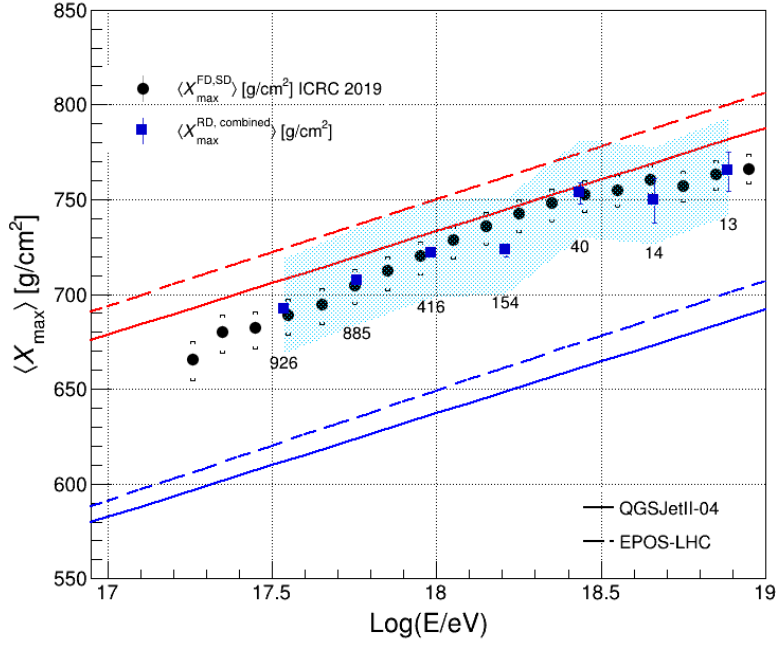


Figure 6.22: Average depth of shower maximum as a function of the SD reconstructed energy as measured by AERA combining the measured from the spectral index and the fit of the energy fluence distribution.  $X_{\text{max}}$  is estimated as the weighted average of the two methods when both methods are successful or corresponds to the single  $X_{\text{max}}$  estimation. The error bars of the AERA data points represent the uncertainty propagated on the weighted average. The colored band indicates the systematic uncertainty. The radio  $\langle X_{\text{max}} \rangle$  is compared to the measurements from the Pierre Auger observatory (black points) [80, 81]. Predictions of the depth of shower maximum of iron and proton primaries from two different interaction models are plotted as lines. The number of entries per energy bin is reported. Only points in energy bins with 10 or more entries are shown.

two methods have significantly different precision, combining them is straightforward and leads to a result consistent with the standard FD-SD composition measurements. Thus, even though the spectral slope method has a large uncertainty on individual events, the additional statistics leads to an improvement of the cosmic-ray composition measurements based on radio signals.

## 7 Discussion and outlook

Extensive air showers propagating through the atmosphere induce a short radio pulse in the MHz regime which is sensitive to all main air-shower observables: the primary cosmic-ray energy, mass, and arrival direction. In the last decade the radio detection of cosmic rays has demonstrated to be a viable alternative to the established detection techniques.

The physics of the radio emission is now sufficiently well understood to use the radio technique for physics measurements. For further progress, it is important to reduce the systematic uncertainty associated to the determination of the quantities obtained from the measured radio signal. To this avail, a new method to estimate the cosmic ray induced signal from a noisy radio measurement is described in Chapter 3. The radio measurement is given by the vector sum of the true but unknown signal and background. The expectation value of the background amplitude and its uncertainty in the frequency domain are determined from the measurement distribution over many background windows of the same length, under the hypothesis that the information in the different background windows is uncorrelated. To take into account the unknown phase between the signal and background, the expectation value of the signal is estimated by averaging over all possible phases, assuming the phases are uniformly distributed in the allowed interval.

Another important factor that affects the precision of the radio based physics measurements is the antenna calibration. In Chapter 4 two calibration techniques based on the galactic radio background are presented: a relative calibration to correct for variations between stations, and an absolute calibration to correct for deviations in the antenna angular and frequency gain pattern. The calibration has been performed on the entire LST period. Additional information can be obtained by doing it for the LST range 16h-18h for the east-west channel, and 14h-16h for the north-south channel to use dominance of the galactic signal over the (electronics) noise.

In this thesis, a new method to determine the depth of the shower maximum  $X_{\max}$ , combining independent estimators obtained using the radio signal is presented. Here the shape of the radio energy footprint at the ground and the slope of the frequency spectrum measured at each radio station are used to estimate the depth of the shower maximum. A new parametrization of the signal distribution at the ground has been devised and a more stable fit procedure has been introduced. This GeoCE parametrization is valid for fully developed showers that have emitted all radiation energy when reaching the ground. This cut limits the data set that can be treated by the GeoCE parametrization, and introduces a bias in the average  $X_{\max}$  measurement. The bias has been corrected for by using the distribution of the average value of  $X_{\max}$  as a function of zenith angle, where for large zenith angles the correction is assumed to be zero. Furthermore a systematic uncertainty has been assigned based on Monte Carlo studies. However, if the  $X_{\max}$  reconstruction from the radio signal footprint could be obtained for all measured events, this bias correction and the systematic uncertainty would be much less. The

parametrization of the spectral index of the electric field amplitude used in this work has been previously studied in Ref. [40].

First, the feasibility of the methods has been proven using CoREAS Monte Carlo simulations, and then applied to the AERA air-shower measured events. The Monte Carlo studies have been performed on star-pattern simulations and using the true values for the geometry of the shower. Thus, the results obtained in Chapter 5 show the theoretically achievable resolutions. In order to improve the interpretation of the data analysis and to have a better handle on the systematic uncertainties, the analyses should be repeated using AERA full detector simulations produced with GDAS atmospheric model. A simulated data set with such characteristics was not available at the time of writing this thesis.

The estimator of the depth shower maximum inferred from the shape of the footprint is more accurate than the one obtained from the radio pulse shape. There are two main reasons for this. Firstly, the pulse shape analysis is more susceptible to noise as it is a differential method, compared to the integral method of the radio footprint. Secondly, the pulse shape very much depends on the knowledge of the exact location of the shower core, whereas the radio footprint only marginally depends on this. Still, the two methods can be combined in a weighted average when properly taking the different resolutions into account. Furthermore, it was found that the methods are rather complementary. The fit on the energy fluence distribution can be applied to events with at least five stations above the noise threshold, and the GeoCE parametrization is efficient only for fully developed showers. While, given the geometry of the shower, it is possible to estimate  $X_{\max}$  even with a single station from spectral information. Moreover, the spectral index method can not be used for stations positioned inside and near the Cherenkov ring, while the same stations are important when fitting the energy footprint. Thus, not only the information used from the radio signal is uncorrelated, but also combining the two methods significantly increases the final number of radio reconstructed events available for compositions studies. In conclusion, despite the difference in achievable accuracy, the average calculation of the shower maximum benefits from the larger statistics as can be seen in Figure 6.22.

A more appropriate Monte Carlo study for the parametrization of spectral index and of the energy fluence distribution as a function of  $D_{X_{\max}}$ , combined with a more accurate antenna calibration, will surely improve the uncertainty of the radio reconstructed  $X_{\max}$ . It has already been proven that the radio technique provides valuable information for mass composition measurements [82–84], but the theoretical limit of these measurements are not yet reached. The radio measurements might become even more accurate than the established air-fluorescence technique. Moreover, this work proves that the combination of independent and complementary techniques is a promising approach for mass composition measurements within existing and future cosmic ray observatories.

## Inclined showers

The advantage of the radio technique is that antennas are much more affordable than particle detectors and therefore allow to instrument a much larger detection area for the same amount of money. This, combined with the near-100% duty cycle in fair weather conditions, enables to increase the number of events at the highest energies of the cosmic-ray spectrum. However, to keep instrumentation cost-effective and, at the same time, have an extensive detection area, the

---

spacing between the antennas has to be as large as possible. Likely, for inclined showers the radio footprint becomes large and can be seen at distances up to a kilometer or more from the shower axis at energies above 1 EeV [54]. The Auger Engineering Radio Array has investigated the potential of a sparse radio array. The next-generation of radio arrays, in particular with the Radio Detector of the Pierre Auger Observatory and GRAND (Giant Radio Array for Neutrino Detection) [85], will exploit the large radio footprint produced by near-horizontal air showers in order to achieve important science goals.

The two mass composition techniques described in this thesis are developed for air showers with zenith angles below  $60^\circ$ . The GeoCE parametrization of the fit function for the energy density footprints can be extended for inclined showers, but the precision on  $X_{\text{max}}$  will be affected by the larger sampling of the signal. Furthermore, considering that the radio pulse shape becomes insensitive to the position of the shower maximum after about 400 m from the shower axis [40], and given the larger antenna spacing, the spectral slope method as presented will not be applicable. However, GRAND is designing antennas to accurately measure the electric field in all three dimensions, which will decrease the uncertainty on the radio measurements in general, and therefore, the pulse shape measurements at each antenna. A detailed pulse shape analysis might contribute to a more accurate determination of  $X_{\text{max}}$ .



# A Offline Module Sequence and selection cuts

CoREAS Monte Carlo simulations and AERA data events have been reconstructed using the *Auger Offline* software (rev 33196), including the module `RdStationSignalReconstructorWithBgSubtraction` from rev-33406.

## A.1 RdReconstructStarshapedStationPattern

The events from the CoREAS Monte Carlo data set analyzed in Chapter 3 and Chapter 5 have been reconstructed using the standard application `RdReconstructStarshapedStationPattern`.

### Module Sequence for the CoREAS simulations reconstruction

The option `BackGroundSubtraction` of the module

`RdStationSignalReconstructorWithBgSubtraction` is set to 0 in case of pure Monte Carlo electric field traces, and to 1 in case background traces are added to the electric field time trace. When the boolean is set to 1 the energy fluence and its uncertainty are estimated following the procedure described in Chapter 3. The option is configurable in the *bootstrap* file.

```
<module> RdStationAssociator </module>
<module> EventGeneratorOG </module>
<!-- imitate measurement from simulation -->
<module> RdAntennaStationToChannelConverter </module>
<module> RdChannelResponseIncorporator </module>
<module> RdChannelResampler </module>
<module> RdChannelTimeSeriesClipper </module>
<module> RdChannelVoltageToADCCConverter </module>
<!-- Radio reconstruction -->
<module> RdEventInitializer </module>
<module> RdVirtualStationNoiseImporter </module> (To add noise)
<module> RdTimeJitterAdder </module>
<module> RdStationRejector </module>
<module> RdChannelADCToVoltageConverter </module>
<module> RdChannelSelector </module>
<module> RdChannelPedestalRemover </module>
<module> RdChannelResponseIncorporator </module>
<module> RdChannelTimeSeriesTaperer </module>
<module> RdChannelSineWaveSuppressor </module> (Only with noise)
<module> RdChannelUpsampler </module>
<module> RdChannelRiseTimeCalculator </module>
<module> RdAntennaChannelToStationConverter </module>
<module> RdStationSignalReconstructorWithBgSubtraction </module>
<module> RdStationEFieldVectorCalculator </module>
```

```

<module> RdStationRiseTimeCalculator          </module>
<module> RdStationPulseShapeRejector         </module> (Only with noise)
<module> RdStationPolarizationRejector       </module> (Only with noise)
<module> RecDataWriterNG                     </module>

```

### A.2 RdObserver

The *Offline* standard application `RdObserver` is used to reconstruct the SD-RD(-FD) hybrid events used in the analyses in Chapter 6. The modules used for the SD and FD events reconstruction are not listed, however the configuration can be found in the *Offline* repository.

#### Module Sequence for the radio event reconstruction

```

<!-- RD selection and reconstruction -->
<module> RdEventPreSelector          </module>
<module> RdEventInitializer          </module>
<module> RdStationRejector           </module>
<module> RdChannelADCToVoltageConverter </module>
<module> RdChannelSelector           </module>
<module> RdChannelPedestalRemover    </module>
<module> RdChannelResponseIncorporator </module>
<module> RdChannelBrokenLNAREjector  </module>
<module> RdStationPositionCorrection </module>
<module> RdChannelBeaconTimingCalibrator </module>
<module> RdChannelBeaconSuppressor   </module>
<module> RdChannelSineWaveSuppressor </module>
<module> RdStationTimingCalibrator   </module>
<module> RdStationTimeWindowConsolidator </module>
<module> RdChannelTimeSeriesTaperer  </module>
<module> RdChannelUpsampler           </module>
<module> RdChannelRiseTimeCalculator </module>
<module> RdChannelGalacticBackgroundCalibrator </module>
<module> RdAntennaChannelToStationConverter </module>
<module> RdStationSignalReconstructorWithBgSubtraction </module>
<module> RdStationEFieldVectorCalculator </module>
<module> RdStationRiseTimeCalculator </module>
<module> RdStationPulseShapeRejector </module>
<module> RdStationPolarizationRejector </module>
<module> RdClusterFinder              </module>
<loop numTimes="unbounded">
  <module> RdTopDownStationSelector </module>
  <module> RdPlaneFit              </module>
</loop>
<module> RdPlaneFit                  </module>

```

The module `RdStationSignalReconstructorWithBgSubtraction` is used with the option:

```
BackgroundSubtraction 1
```



## A.3 Quality cuts for the RD-SD(-FD) data set

### RD cuts

```
-->          minZenithRD      0
-->          maxZenithRD      55
-->          maxAngleSDRD     10
-->          contained        (boolean cut)
-->          thunderstorm      (boolean cut)
```

### SD cuts

```
-->          lightning (boolean cut) (anti-cut)
-->          minRecLevel      3
-->          maxZenithSD      55
-->          T4Trigger         2
-->          T5Trigger         2
-->          minLgEnergySD    16.5
--> badPeriodsRejectionFromFile (boolean cut)
```

### FD cuts

```
-->          isCLF            (boolean cut) (anti-cut)
-->          isXLF            (boolean cut) (anti-cut)
-->          eyeCut           111111
-->          badFDPeriodRejection (boolean cut)
--> minMeanPixelRMSMergedEyes   (17, 6, 110000)
--> minMeanPixelRMSSimpleEyes  (17, 11111)
-->          badPixels         1 (anti-cut)
-->          good10MHzCorrection (boolean cut)
-->          hasMieDatabase     (boolean cut)
-->          maxVAOD           0.1
-->          cloudCutXmaxPRD14  1
-->          hybridTankTrigger  2
-->          maxCoreTankDist    1500
-->          maxZenithFD        90
-->          minLgEnergyFD      1e-20
-->          skipSaturated      (boolean cut)
-->          minPBrass          0.9
-->          maxPBrassProtonIronDiff 0.05
-->          xMaxInFOV          0 or -100
-->          minViewAngle       20
-->          xMaxError          40 or 100
-->          maxDepthHole       100
-->          profileChi2Sigma    (3, -1.1)
-->          depthTrackLength    200
```



## B 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](https://www.ru.nl/publish/pages/868512/imapp_rdm_policy.pdf).

- The data and the CoREAS Monte Carlo simulations used in this thesis are available within the Pierre Auger Collaboration.
- The data and the CoREAS Monte Carlo simulations have been reconstructed using the Auger *Offline* software (rev 33196), including the module `RdStationSignalReconstructorWithBgSubtraction` from rev-33406.
- *Offline* externals: aerarootio v00r16, aevread v02r00p04, aires 2.8.4a, boost 1.69.0, cdas v5r4, clhep 2.4.0.4, cmake 3.14.2, cppunit 1.13.1, eigen 3.1.2, fdeventlib 4.1.10, fftw 3.3.3, pkg-config 0.27.1, xerces-c 3.1.3, ROOT framework 5.34/38.
- The reconstructed data is stored on the CNCZ computing clusters. It may be obtained upon request.
- The calibration files produced in Chapter 4 may be found in <https://gitlab.ikp.kit.edu/Radio/AERAutilities>.
- The GeoCE parametrization of the energy fluence distribution discussed in Chapter 5 may be found in <https://gitlab.ikp.kit.edu/Radio/AERAutilities>.



# Bibliography

- [1] P. Auger et al., *Extensive Cosmic-Ray Showers*, **Rev. Mod. Phys.** **11** (3-4 1939) 288.
- [2] G. A. Askar'yan, *Excess negative charge of an electron-photon shower and its coherent radio emission*, *Sov. Phys. JETP* **14** (1962) 441, URL: [http://old.inspirehep.net/record/1351286/files/e\\_014\\_02\\_0441.pdf?version=1](http://old.inspirehep.net/record/1351286/files/e_014_02_0441.pdf?version=1).
- [3] J. V. Jelley et al., *Radio Pulses from Extensive Cosmic-Ray Air Showers*, **Nature** **205** (1965).
- [4] V. Verzi for the Pierre Auger Collaboration, *Measurement of the energy spectrum of ultra-high energy cosmic rays using the Pierre Auger Observatory*, in *Proc. of the 36th ICRC*, Madison, WI, U.S.A (2019), URL: <https://pos.sissa.it/358/450/pdf>.
- [5] A. Castellina for the Pierre Auger Collaboration, *Highlights from the Pierre Auger Observatory*, in *Proc. of the 36th ICRC*, Madison, WI, U.S.A (2019), URL: <https://pos.sissa.it/358/004/pdf>.
- [6] W.D. Apel et al., KASCADE-Grande Collaboration, *Ankle-like feature in the energy spectrum of light elements of cosmic rays observed with KASCADE-Grande*, **Phys. Rev. D** **87** (8 2013) 081101, URL: <https://arxiv.org/pdf/1304.7114>.
- [7] R. U. Abbasi et al., Telescope Array Collaboration, *The Cosmic-Ray Energy Spectrum between 2 PeV and 2 EeV Observed with the TALE detector in monocular mode*, **Astrophys. J.** **865** (2018) 74, URL: <https://arxiv.org/pdf/1803.01288>.
- [8] T. K. Gaisser, T. Stanev and S. Tilav, *Cosmic ray energy spectrum from measurements of air showers*, **Front. Phys.** **8** (2013) 748, ISSN: 2095-0470, URL: <https://arxiv.org/pdf/1303.3565>.
- [9] V. Berezhinsky, *Extragalactic cosmic rays and their signatures*, **Astropart. Phys.** **53** (2014) 120, ISSN: 0927-6505, URL: <https://arxiv.org/pdf/1301.0914>.
- [10] M. Unger, G. R. Farrar and L. A. Anchordoqui, *Origin of the ankle in the ultrahigh energy cosmic ray spectrum, and of the extragalactic protons below it*, **Phys. Rev. D** **92** (2015) 123001, URL: <https://arxiv.org/pdf/1505.02153>.
- [11] K. Greisen, *End to the Cosmic-Ray Spectrum?*, **Phys. Rev. Lett.** **16** (17 1966) 748, URL: [http://personal.psu.edu/rq9/HOW/GZK\\_Cutoff\\_1.pdf](http://personal.psu.edu/rq9/HOW/GZK_Cutoff_1.pdf).
- [12] G. T. Zatsepin and V. A. Kuz'min, *Upper limit of the spectrum of cosmic rays*, *JETP Lett.* **4** (1966) 78, URL: [http://www.jetpletters.ac.ru/ps/1624/article\\_24846.pdf](http://www.jetpletters.ac.ru/ps/1624/article_24846.pdf).
- [13] E. Fermi, *On the Origin of the Cosmic Radiation*, **Phys. Rev.** **75** (8 1949) 1169, URL: [https://fermi.gsfc.nasa.gov/science/mtgs/symposia/2014/program/03\\_Hewitt.pdf](https://fermi.gsfc.nasa.gov/science/mtgs/symposia/2014/program/03_Hewitt.pdf).
- [14] A. Hillas, *The Origin of Ultrahigh-Energy Cosmic Rays*, **Ann. Rev. Astron. Astrophys.** **22** (1984) 425, URL: <http://articles.adsabs.harvard.edu/pdf/1984ARA%26A..22..425H>.

- [15] K. Kotera and A. V. Olinto, *The Astrophysics of Ultrahigh Energy Cosmic Rays*, [Ann. Rev. Astron. Astrophys. \*\*49\*\* \(2011\) 119](#), URL: <https://arxiv.org/pdf/1101.4256>.
- [16] P. Bauleo and J. Rodrigues Martino, *The dawn of the particle astronomy era in ultra-high-energy cosmic rays*, [Nature \*\*458N7240\*\* \(2009\) 847](#).
- [17] D. Harari, S. Mollerach and E. Roulet, *On the ultrahigh energy cosmic ray horizon*, [JCAP \*\*0611\*\* \(2006\) 012](#), URL: <https://arxiv.org/pdf/astro-ph/0609294>.
- [18] R. Engel, D. Heck and T. Pierog, *Extensive Air Showers and Hadronic Interactions at High Energy*, [Annu. Rev. Nucl. Part. S. \*\*61\*\* \(2011\) 467](#).
- [19] J. C. Glaser, *Absolute Energy Calibration of the Pierre Auger Observatory using Radio Emission of Extensive Air Showers*, PhD Thesis, RWTH Aachen University (2016), URL: <https://publications.rwth-aachen.de/record/686745/files/686745.pdf>.
- [20] D. Heck et al., *CORSIKA: A Monte Carlo code to simulate extensive air showers*, (1998), URL: <https://inspirehep.net/files/9bdb7ea99706d9ae48d0be29f9e19f11>.
- [21] W. Heitler, *Quantum theory of radiation*, 1954. Reprinted by Dover Publications, 1984, ISBN: 0-486-64558-4.
- [22] J. Matthews, *A Heitler model of extensive air showers*, [Astropart. Phys. \*\*22\*\* \(2005\) 387](#), ISSN: 0927-6505.
- [23] V. de Souza for the Pierre Auger Collaboration and the Telescope Array Collaboration, *Testing the agreement between the Xmax distributions measured by the Pierre Auger and Telescope Array Observatories*, in [Proc. of the 35th ICRC, Bexco, Busan, Korea \(2017\)](#).
- [24] H. R. Allan, *Progress in Elementary Particle and Cosmic Ray Physics*, vol. 10, 1971.
- [25] H. Falcke and P. Gorham, *Detecting radio emission from cosmic ray air showers and neutrinos with a digital radio telescope*, [Astroparticle Physics \*\*19\*\* \(2003\) 477](#), ISSN: 0927-6505, URL: <https://arxiv.org/pdf/astro-ph/0207226>.
- [26] D. Ardouin et al., *Radio-detection signature of high-energy cosmic rays by the CODALEMA experiment*, [Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment \*\*555\*\* \(2005\) 148](#), ISSN: 0168-9002.
- [27] H. Falcke et al., LOPES Collaboration, *Detection and imaging of atmospheric radio flashes from cosmic ray air showers*, [Nature \*\*435\*\* \(2005\) 313](#), URL: <https://arxiv.org/pdf/astro-ph/0505383>.
- [28] J. Coppens for the Pierre Auger Collaboration, *Observation of radio signals from air showers at the Pierre Auger Observatory*, [Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment \*\*604\*\* \(2009\) S41](#), ARENA 2008, ISSN: 0168-9002, URL: <http://particle.astro.ru.nl/pub/NIMA604-S41.pdf>.
- [29] A. Aab et al., Pierre Auger Collaboration, *Probing the radio emission from air showers with polarization measurements*, [Phys. Rev. D \*\*89\*\* \(5 2014\) 052002](#), URL: <https://arxiv.org/pdf/1402.3677>.
- [30] F. G. Schröder, *Radio detection of cosmic-ray air showers and high-energy neutrinos*, [Progress in Particle and Nuclear Physics \*\*93\*\* \(2017\) 1](#), ISSN: 0146-6410, URL: <https://arxiv.org/pdf/1607.08781>.

- 
- [31] A. Corstanje et al., LOFAR Collaboration, *The shape of the radio wavefront of extensive air showers as measured with LOFAR*, *Astropart. Phys.* **61** (2015) 22, URL: <https://arxiv.org/pdf/1404.3907>.
- [32] A. Aab et al., Pierre Auger Collaboration, *Energy estimation of cosmic rays with the Engineering Radio Array of the Pierre Auger Observatory*, *Phys. Rev. D* **93** (12 2016) 122005, URL: <https://arxiv.org/pdf/1508.04267>.
- [33] K. D. de Vries, A. M. van den Berg, O. Scholten and K. Werner, *Coherent Cherenkov Radiation from Cosmic-Ray-Induced Air Showers*, *Phys. Rev. Lett.* **107** (6 2011) 061101, URL: <https://arxiv.org/pdf/1107.0665>.
- [34] A. Nelles et al., LOFAR Collaboration, *A parameterization for the radio emission of air showers as predicted by CoREAS simulations and applied to LOFAR measurements*, *Astropart. Phys.* **60** (2015) 13, ISSN: 0927-6505, URL: <https://arxiv.org/pdf/1402.2872>.
- [35] C. Glaser, S. de Jong, M. Erdmann and J. R. Hörandel, *An analytic description of the radio emission of air showers based on its emission mechanisms*, *Astropart. Phys.* **104** (2019) 64, ISSN: 0927-6505, URL: <https://arxiv.org/pdf/1806.03620>.
- [36] W.D. Apel et al., LOPES Collaboration, *The wavefront of the radio signal emitted by cosmic ray air showers*, *J. Cosmol. Astropart. P.* (2014) 025, URL: <https://arxiv.org/pdf/1404.3283>.
- [37] S. Grebe, *Finger on the pulse of cosmic rays, dependence of the radio pulse shape on the air shower geometry*, PhD Thesis, Radboud University Nijmegen (2013), URL: <https://repository.ubn.ru.nl/dspace3lxmlui/handle/2066/112893>.
- [38] A. Nelles, *Radio Emission of air shower. The perspective of LOFAR and AERA.*, PhD Thesis, Radboud University Nijmegen (2014), URL: <https://repository.ubn.ru.nl/bitstream/handle/2066/129668/129668.pdf>.
- [39] J. Schulz, *Cosmic Radiation, reconstruction of cosmic-ray properties from radio emission of extensive air shower*, PhD Thesis, Radboud University Nijmegen (2016), URL: <https://repository.ubn.ru.nl/handle/2066/151781>.
- [40] S. Jansen, *Radio for the masses, cosmic ray mass composition measurements in the radio frequency domain*, PhD Thesis, Radboud University Nijmegen (2016), URL: [https://www.nikhef.nl/pub/services/biblio/theses\\_pdf/thesis\\_S\\_Jansen.pdf](https://www.nikhef.nl/pub/services/biblio/theses_pdf/thesis_S_Jansen.pdf).
- [41] C. W. James, H. Falcke, T. Huege and M. Ludwig, *General description of electromagnetic radiation processes based on instantaneous charge acceleration in "endpoints"*, *Phys. Rev. E* **84** (5 2011) 056602, URL: <https://arxiv.org/pdf/1007.4146>.
- [42] A. Aab et al., Pierre Auger Collaboration, *The Pierre Auger Cosmic Ray Observatory*, *Nucl. Instrum. Meth. A* **798** (2015) 172, ISSN: 0168-9002, URL: <https://arxiv.org/pdf/1502.01323>.
- [43] D. Martello, *The Pierre Auger Observatory: Results and open issues*, *Nucl. Instrum. Meth. A* **742** (2014) 16, 4th Roma International Conference on Astropart. Phys., ISSN: 0168-9002.
- [44] J. Abraham et al., Pierre Auger Collaboration, *The fluorescence detector of the Pierre Auger Observatory*, *Nucl. Instrum. Meth. A* **620** (2010) 227, ISSN: 0168-9002, URL: <https://arxiv.org/pdf/0907.4282>.



- [45] T. K. Gaisser and A. M. Hillas, *Reliability of the Method of Constant Intensity Cuts for Reconstructing the Average Development of Vertical Showers*, in Proc. of the International Cosmic Ray Conference (1977), URL: <http://articles.adsabs.harvard.edu/pdf/1977ICRC....8..353G>.
- [46] M. Unger et al., *Reconstruction of longitudinal profiles of ultra-high energy cosmic ray showers from fluorescence and Cherenkov light measurements*, *Nucl. Instrum. Meth. A* **588** (2008) 433, ISSN: 0168-9002, URL: <https://arxiv.org/pdf/0801.4309>.
- [47] P. Abreu et al., Pierre Auger Collaboration, *Description of atmospheric conditions at the Pierre Auger Observatory using the Global Data Assimilation System (GDAS)*, *Astropart. Phys.* **35** (2012) 591, ISSN: 0927-6505, URL: <https://arxiv.org/pdf/1201.2276>.
- [48] K. Kamata and J. Nishimura, *The Lateral and the Angular Structure Functions of Electron Showers*, *Progress of Theoretical Physics Supplement* **6** (1958) 93.
- [49] K. Greisen, *Cosmic ray showers*, *Ann. Rev. Nucl. Part. Sci.* **10** (1960) 63.
- [50] 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** (12 2017) 122003, URL: <https://arxiv.org/pdf/1710.07249>.
- [51] 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), URL: [https://www.ru.nl/publish/pages/870143/201016\\_guiseppede\\_mauro\\_tasting\\_the\\_ultra\\_high\\_energy\\_universe.pdf](https://www.ru.nl/publish/pages/870143/201016_guiseppede_mauro_tasting_the_ultra_high_energy_universe.pdf).
- [52] A. Bridgeman for the Pierre Auger Collaboration, *Shower universality reconstruction of data from the Pierre Auger Observatory and validations with hadronic interaction models*, in *Proc. of the 35th ICRC, Bexco, Busan, Korea* (2017).
- [53] 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** (2010) 29, URL: <https://arxiv.org/pdf/1111.6764>.
- [54] M. Gottowik for the Pierre Auger Collaboration, *Measurements of Inclined Air Showers with the Auger Engineering Radio Array at the Pierre Auger Observatory*, in Proc. of the 36th ICRC, Madison, WI, U.S.A (2019), URL: <https://pos.sissa.it/358/274/pdf>.
- [55] K. F. Weidenhaupt, *Antenna Calibration and Energy Measurement of Ultra-High Energy Cosmic Rays with the Auger Engineering Radio Array*, RWTH Aachen University (2014), URL: <https://publications.rwth-aachen.de/record/465387/files/5264.pdf>.
- [56] P. Abreu et al., Pierre Auger Collaboration, *Antennas for the detection of radio emission pulses from cosmic-ray induced air showers at the Pierre Auger Observatory*, *J. Instrum.* **7** (2012) P10011, URL: <https://arxiv.org/pdf/1209.3840>.
- [57] J.L. Kelley for the Pierre Auger Collaboration, *Data acquisition, triggering, and filtering at the Auger Engineering Radio Array*, *Nucl. Instrum. Meth. A* **725** (2013) 133, ISSN: 0168-9002, URL: <https://arxiv.org/pdf/1205.2104>.
- [58] A. Aab et al., Pierre Auger Collaboration, *The Pierre Auger Observatory Upgrade - Preliminary Design Report*, (2016), URL: <https://arxiv.org/pdf/1604.03637>.

- 
- [59] B. Pont for the Pierre Auger Collaboration, *A Large Radio Detector at the Pierre Auger Observatory - Measuring the Properties of Cosmic Rays up to the Highest Energies*, in Proc. of the 36th ICRC, Madison, WI, U.S.A (2019), URL: <https://pos.sissa.it/358/395/pdf>.
- [60] P. Abreu et al., Pierre Auger Collaboration, *Advanced functionality for radio analysis in the Offline software framework of the Pierre Auger Observatory*, **Nucl. Instrum. Meth. A** **635** (2011) 92, ISSN: 0168-9002, URL: <https://arxiv.org/pdf/1101.4473>.
- [61] M. Ishii and Y. Baba, *Advanced computational methods in lightning performance. The Numerical Electromagnetics Code (NEC-2)*, 2000 IEEE Power Engineering Society Winter Meeting. Conference Proceedings (Cat. No.00CH37077) **4** (2000) 2419.
- [62] A. Aab et al., Pierre Auger Collaboration, *Calibration of the logarithmic-periodic dipole antenna (LPDA) radio stations at the Pierre Auger Observatory using an octocopter*, **Journal of Instrumentation** **12** (2017) T10005, URL: <https://arxiv.org/pdf/1702.01392>.
- [63] F. Schlüter, *Unfolding the Electric Field of Radio Emission in Extensive Air Showers by Calibration of the Butterfly Antenna and Deep Learning based Noise Suppression at the Pierre Auger Observatory*, Master Thesis, RWTH Aachen (2018), URL: [https://www.institut3a.physik.rwth-aachen.de/global/show\\_document.asp?id=aaaaaaaaaetqirf](https://www.institut3a.physik.rwth-aachen.de/global/show_document.asp?id=aaaaaaaaaetqirf).
- [64] S. Argirò et al., *The offline software framework of the Pierre Auger Observatory*, **Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment** **580** (2007) 1485, ISSN: 0168-9002, URL: <https://arxiv.org/pdf/0707.1652>.
- [65] E. Polisensky, *LFmap: A Low Frequency Sky Map Generating Program*, (2007), URL: <https://www.faculty.ece.vt.edu/swe/lwa/memo/lwa0111.pdf>.
- [66] S. Fliescher, *Antenna Devices and Measurement of Radio Emission from Cosmic Ray induced Air Showers at the Pierre Auger Observatory*, PhD Thesis, RWTH Aachen University (2011), URL: <https://publications.rwth-aachen.de/record/82811/files/3998.pdf>.
- [67] K. Mulrey et al., LOFAR Collaboration, *Calibration of the LOFAR low-band antennas using the Galaxy and a model of the signal chain*, **Astroparticle Physics** **111** (2019) 1, ISSN: 0927-6505, URL: <https://arxiv.org/pdf/1903.05988>.
- [68] C. Glaser et al., *Simulation of radiation energy release in air showers*, **Journal of Cosmology and Astroparticle Physics** **2016** (2016) 024, URL: <https://arxiv.org/pdf/1606.01641>.
- [69] S. Ostapchenko, *Monte Carlo treatment of hadronic interactions in enhanced Pomeron scheme: QGSJET-II model*, **Phys. Rev. D** **83** (1 2011) 014018, URL: <https://arxiv.org/pdf/1010.1869>.
- [70] M. Bleicher et al., *Relativistic hadron-hadron collisions in the ultra-relativistic quantum molecular dynamics model*, **Journal of Physics G: Nuclear and Particle Physics** **25** (1999) 1859, URL: <https://arxiv.org/pdf/hep-ph/9909407>.

- [71] T. Huege, M. Ludwig, O. Scholten and K. de Vries, *The convergence of EAS radio emission models and a detailed comparison of REAS3 and MGMR simulations*, [Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment](#) **662** (2012) S179, 4th International workshop on Acoustic and Radio EeV Neutrino detection Activities, ISSN: 0168-9002, URL: <https://arxiv.org/pdf/1009.0346>.
- [72] C. Welling, *Identification of Radio Signals from Cosmic Ray Induced Air Showers with the Auger Engineering Radio Array*, Master Thesis, RWTH Aachen (2017), URL: [https://www.institut3a.physik.rwth-aachen.de/global/show\\_document.asp?id=aaaaaaaaarbkltf](https://www.institut3a.physik.rwth-aachen.de/global/show_document.asp?id=aaaaaaaaarbkltf).
- [73] A. Schulz for the Pierre Auger Collaboration, *The measurement of the energy spectrum of cosmic rays above  $3 \times 10^{17}$  eV with the Pierre Auger Observatory*, in Proc. of the 33th ICRC, Rio de Janeiro, Brazil (2013), URL: <https://inspirehep.net/files/1548368d85c92c23c61a9e2ade55ee95>.
- [74] P. Schellart et al., *Probing Atmospheric Electric Fields in Thunderstorms through Radio Emission from Cosmic-Ray-Induced Air Showers*, [Phys. Rev. Lett.](#) **114** (16 2015) 165001, URL: <https://arxiv.org/pdf/1504.05742>.
- [75] S. Buitink et al., LOPES Collaboration, *Amplified radio emission from cosmic ray air showers in thunderstorms*, [A&A](#) **467** (2007) 385, URL: <https://arxiv.org/pdf/astro-ph/0702432>.
- [76] J. Rautenberg for the Pierre Auger Collaboration, *Lightning Detection at the Pierre Auger Observatory*, in Proc. of the 34th ICRC, The Hague, The Netherlands (2016).
- [77] H. Schoorlemmer, *Tuning in on cosmic rays, Polarization of radio signals from air showers as a probe of emission mechanisms*, PhD Thesis, Radboud University Nijmegen (2012), URL: <https://repository.ubn.ru.nl/bitstream/handle/2066/94180/94180.pdf?sequence=1>.
- [78] S. Nehls, *Calibrated Measurements of the Radio Emission of Cosmic Ray Air Showers*, PhD Thesis, Institut für Kernphysik, Universität Karlsruhe (2008), URL: <https://inspirehep.net/files/570ee7b905299486f7df0d1dd4f4f63d>.
- [79] F. Fenu Francesco for the Pierre Auger Collaboration, *The cosmic ray energy spectrum measured using the Pierre Auger Observatory*, in Proc. of the 35th ICRC, Bexco, Busan, Korea (2017), URL: <https://pos.sissa.it/301/486/pdf>.
- [80] A. Yushkov for the Pierre Auger Collaboration, *Mass Composition of Cosmic Rays with Energies above  $10^{17.2}$  eV from the Hybrid Data of the Pierre Auger Observatory*, in Proc. of the 36th ICRC, Madison, WI, U.S.A (2019), URL: <https://pos.sissa.it/358/482/pdf>.
- [81] A. Aab et al., Pierre Auger Collaboration, *Depth of maximum of air-shower profiles at the Pierre Auger Observatory. I. Measurements at energies above  $10^{17.8}$  eV*, [Phys. Rev. D](#) **90** (12 2014) 122005, URL: <https://arxiv.org/pdf/1409.4809>.
- [82] S. Buitink et al., LOFAR Collaboration, *Corrigendum: A large light-mass component of cosmic rays at  $10^{17}$ - $10^{17.5}$  electronvolts from radio observation*, English, [Nature](#) **537** (2016) 572, ISSN: 0028-0836.
- [83] P.A. Bezyazeev et al., Tunka-Rex Collaboration, *Reconstruction of cosmic ray air showers with Tunka-Rex data using template fitting of radio pulses*, [Phys. Rev. D](#) **97** (12 2018) 122004, URL: <https://arxiv.org/pdf/1803.06862>.

- 
- [84] S. Knurenko and I. Petrov, *Mass composition of cosmic rays above 0.1 EeV by the Yakutsk array data*, *Advances in Space Research* **64** (2019) 2570, Advances in Cosmic-Ray Astrophysics and Related Areas, ISSN: 0273-1177, URL: <https://arxiv.org/pdf/1908.01508>.
- [85] J. Álvarez-Muñiz et al., GRAND Collaboration, *The Giant Radio Array for Neutrino Detection (GRAND): Science and design*, *Science China Physics, Mechanics & Astronomy* **63** (2019), ISSN: 1869-1927, URL: <https://arxiv.org/pdf/1810.09994>.



# Summary

## Hunting for the most energetic particles in nature

The most energetic elementary particles ever observed in nature are cosmic rays. Cosmic rays are atomic nuclei generated in outer space. They move through the universe at nearly the speed of light and some of them reach the Earth. The discovery of cosmic rays with *ultra-high energies* has opened a window to observe the far and unknown universe. Only the most violent phenomena in the universe can accelerate particles to such extreme energies. However, nature does not give away her secrets easily: as the cosmic rays travel through the universe magnetic forces bend their flight paths making it hard to locate their sources. Additionally, ultra-high-energy cosmic rays reaching the Earth are very rare: a detector with a size of one square kilometer would, on average, observe only one particle every hundred years. For this reason cosmic-ray observatories must have a very large detection area. The biggest experiment built so far devoted to the study of ultra-high-energy cosmic rays is the Pierre Auger Observatory in the Argentinean Pampas. It covers a surface of 3000 km<sup>2</sup>, which is equivalent to about 30 times the area of Paris. More than 500 physicists from 17 countries collaborate to maintain and upgrade the site in Argentina and collect and analyze the data.

When such a high-energy particle enters the Earth's atmosphere it collides with air molecules generating an *air shower* consisting of billions of secondary particles with a lower energy that can also be observed at the ground. Luckily, the air shower can cover a surface of tens of square kilometer, making the detection of ultra-high-energy cosmic rays on the Earth's surface possible with only a sparse array of particle detectors. In addition, the secondary particles leave behind several clues that are used to reconstruct the properties of the primary cosmic ray. For example, the charged secondary particles excite to the molecules in the atmosphere, which then emit ultraviolet light. This light is observed during dark nights by fluorescence telescopes that can reconstruct the development of the shower as it propagates through the atmosphere with high accuracy. The Pierre Auger Observatory has 24 fluorescence telescopes located at four different sites, and more than 1600 particle detectors spaced 1.5 km apart. A representation of an air shower observed simultaneously by the fluorescence telescopes and the particle detectors of the Pierre Auger Observatory is shown in Figure 1.

Electrons and positrons in the air shower, accelerated in the Earth's magnetic field, emit radio waves that carry information about the primary cosmic ray. Fluorescence telescopes are functional only in moonless clear nights, whereas radio antennas collect data 24/7, which is an important advantage since we are looking for extremely rare events. Furthermore, radio antennas are cheaper and easier to build and maintain with respect to telescopes and other particle detectors; this, in principle, allows for a large detection area while keeping instrumentation cost-effective.

From 2013, about 150 radio antenna stations spread over an area of 17 km<sup>2</sup> were installed at the site of the Pierre Auger Observatory to investigate radio detection as a viable alternative to

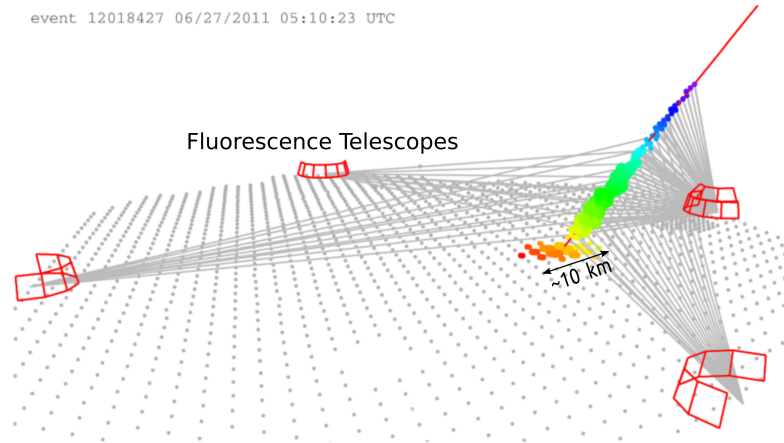


Figure 1: Representation of a cosmic-ray event viewed by the fluorescence telescopes and the particle detectors of the Pierre Auger Observatory. The telescopes detect the light emitted by the atoms in the atmosphere excited by the particles in the shower. The amount of light collected is proportional to the height along the shower direction, while the color indicates the arrival time of the signal from early blue to late in red. The gray dots indicate the particle detectors that collect information from the front of the air shower. The signal height in the particle detector stations is proportional to the colored disk, while the color itself indicates again the arrival time of the signal.



Figure 2: Radio antenna station in the Argentinean Pampas. Stations are surrounded by a fence to protect them from curious animals.

other cosmic-ray detection techniques. Figure 2 shows a picture of a radio station. The Pierre Auger Observatory offers the unique opportunity to measure the cosmic-ray properties with various independent techniques and to cross-calibrate the detector responses.

### Measuring masses with radio antennas

Several theories attempt to describe the origin and propagation of cosmic rays. In order to distinguish the correct one and solve the long-standing riddle about the origin of ultra-high-energy cosmic rays it is essential to know their mass composition. The mass composition indicates the frequency with which different particles occur and the relative abundance of



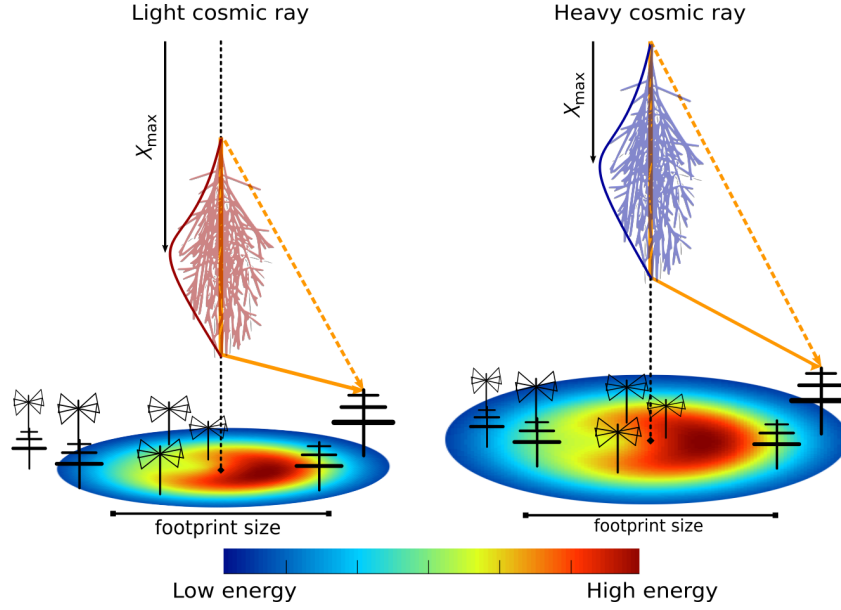


Figure 3: Illustration of the relation between the size of the footprint of the radio-energy distribution and the  $X_{\max}$  of the air shower. Lighter cosmic rays interact deeper in the atmosphere producing a narrower radio footprint. Furthermore the projected length of the radio signal is typically larger for light cosmic rays than for heavy cosmic rays. The length of the radio signal corresponds to the difference between the time it takes for a radio signal to travel from the start of the emission region to the antenna (orange dashed line), and the time it takes for the signal to reach the antenna from the end of the emission region plus the total time in which the air shower emits radiation (solid orange line).

elements. In this thesis we describe a new method to measure the mass composition of the most energetic cosmic rays using radio signals.

A heavy cosmic ray interacts higher up in the atmosphere than a lighter particle. The interaction with the atmospheric molecules initiates an air shower in which the number of secondary particles increases until it reaches a maximum and then slowly starts to diminish. The atmospheric depth  $X_{\max}$  at which the air shower reaches the maximum number of particles is related to the type of primary cosmic ray. The radio-signal-based  $X_{\max}$  is obtained by combining two independent methods. The first method exploits the relation between the shape of the radio-energy distribution at the ground and the type of primary particle. The basic idea is shown in Figure 3: a light particle penetrates deeper in the atmosphere and produces a narrower energy footprint compared to a heavier primary cosmic ray. The second method is based on the frequency content of the radio signals. As shown in Figure 4, lighter particles produce a longer radio signal in time that corresponds to a steeper frequency spectrum. Hence,  $X_{\max}$  is estimated from the steepness of the frequency spectrum measured at each radio station.

To determine the cosmic-ray composition as a function of the energy of the primary particle, we use the average value of  $X_{\max}$  of all events with a similar energy. Fluorescence telescopes can measure  $X_{\max}$  very precisely as they observe the shower profile growing and dying out in the atmosphere. In Figure 5 the radio-based averaged  $X_{\max}$  is therefore compared to the measurements of the fluorescence telescopes. In order to interpret the result, the measurements are compared with the predictions of the average  $X_{\max}$  obtained from a pure light (red) or heavy composition (blue). Our points indicate that towards higher energies the cosmic-ray composition first becomes lighter and then slightly heavier again. By comparing



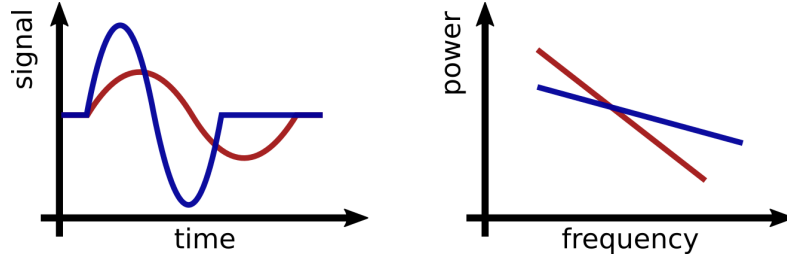


Figure 4: A long pulse from a light cosmic ray results in a steeper frequency spectrum than the spectrum of a short pulse. The color coding corresponds to the showers shown in Figure 3.

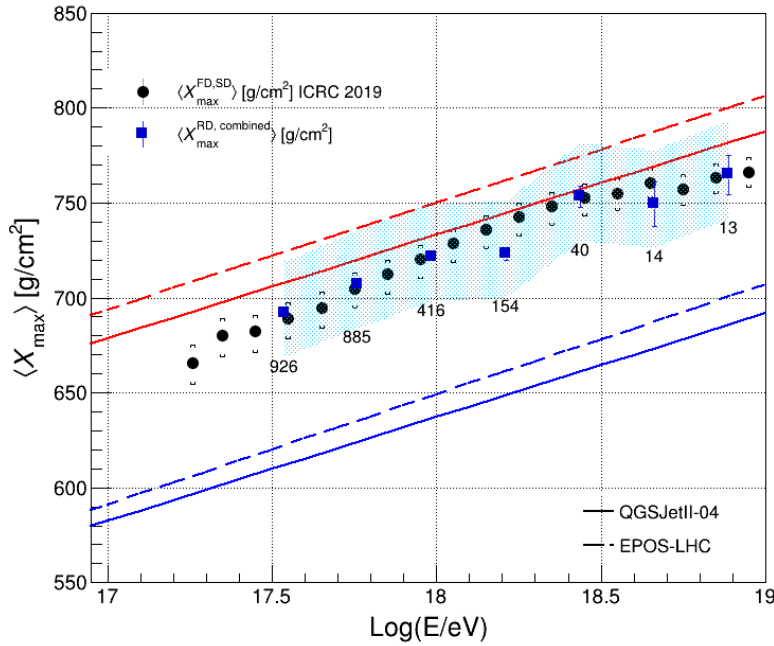


Figure 5: Average  $X_{\max}$  as a function of the energy of the cosmic rays. The blue points are obtained using radio signals, while the black points are the measurements from the fluorescence telescopes. Theoretical predictions for light (red) and heavy (blue) cosmic rays are shown for two different hadronic interaction models, QGSJetII-04 (solid line) and EPOS-LHC (dashed line).

the radio and the fluorescence technique it can be concluded that it is possible to obtain reliable composition information using radio signals from air showers with a sparse radio array. The plan for the near future is to equip each of the Pierre Auger Observatory particle detector stations with a radio antenna, resulting in a 3000 km<sup>2</sup> radio array. Another experiment, the Giant Radio Array for Neutrino Detection (GRAND), is proposed as a huge-scale detector with 200 000 low-cost antennas spread out over nearly 200 000 km<sup>2</sup> at different locations around the world. These ambitious experiments will open a new era for the radio detection of cosmic rays, which will certainly contribute to unveiling some of the most energetic mysteries of the universe.

# Samenvatting

## Op deeltjes met de allerhoogste energie jagen

Kosmische straling bevat de meest energetische elementaire deeltjes ooit gemeten. Kosmische straling bestaat uit atoomkernen die in de ruimte ontstaan zijn. Ze bewegen zich door het heelal met bijna de lichtsnelheid en enkelen treffen ook de aarde op hun pad. De ontdekking van kosmische straling met *ultrahoge energie* geeft ons een nieuwe manier om het verre en onbekende universum te ontdekken.

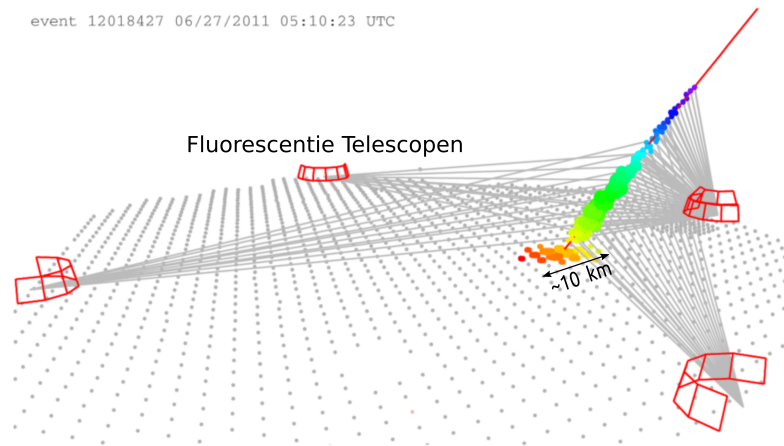
Alleen de meest heftige natuurverschijnselen in het universum kunnen deeltjes tot zulke extreem hoge energieën versnellen. Helaas geeft de natuur haar geheimen niet zo makkelijk prijs: terwijl de kosmische deeltjes door de ruimte vliegen worden ze door magnetische velden afgebogen. Dit maakt het moeilijk om hun oorsprong terug te vinden.

Daar komt nog bij dat ultrahoog-energetische kosmische straling maar heel zelden door de aarde wordt gevangen: om gemiddeld één deeltje te vangen in één vierkante kilometer moet je honderd jaar meten. Observatoria van ultrahoog-energetische kosmische straling moeten dus een erg groot oppervlakte hebben om genoeg deeltjes te detecteren. Het grootste experiment om kosmische straling met ultrahoge energie te bestuderen is het Pierre Auger Observatorium in de Argentijnse Pampa. Met een oppervlakte van 3000 km<sup>2</sup> is het ongeveer 30 keer zo groot als de oppervlakte van Parijs. Meer dan 500 natuurkundigen uit 17 landen werken samen om het te onderhouden en verbeteren en voor het verzamelen en analyseren van de meetgegevens.

Als een deeltje met hoge energie de aard atmosfeer binnenkomt, botst het met moleculen in de lucht waardoor een stortbui van deeltjes met lagere energieën ontstaat die op de grond gemeten kunnen worden. Deze *deeltjeslawine* kan een oppervlakte van enkele tientallen vierkante kilometer bestrijken, waardoor ver uit elkaar staande deeltjesdetectoren genoeg kunnen zijn om ultrahoog-energetische kosmische straling te detecteren.

Met de verschillende aanwijzingen die deze secundaire deeltjes geven kunnen de eigenschappen van het eerste deeltje goed bepaald worden. Bijvoorbeeld: de geladen secundaire deeltjes geven wat energie aan de moleculen in de atmosfeer die daardoor ultraviolet oplichten. Dit licht kan in donkere nachten gezien worden door speciale fluorescentie telescopen waardoor de deeltjeslawine goed in kaart kan worden gebracht. Het Pierre Auger Observatorium heeft 24 van die fluorescentie telescopen op vier verschillende plekken en ook meer dan 1600 deeltjesdetectoren op de grond met een onderlinge afstand van 1.5 km. Een schets van een deeltjeslawine die tegelijkertijd gemeten is met fluorescentie telescopen en de deeltjesdetectoren op de grond van het Pierre Auger Observatorium is te zien in Figuur 1.

Elektronen en positronen in de deeltjeslawine die in het magneetveld van de aarde versneld worden, zenden daarbij radiogolven uit met informatie over het eerste deeltje. Fluorescentie telescopen kunnen alleen meten bij maanloze en heldere nachten, terwijl radio antennes 24 uur



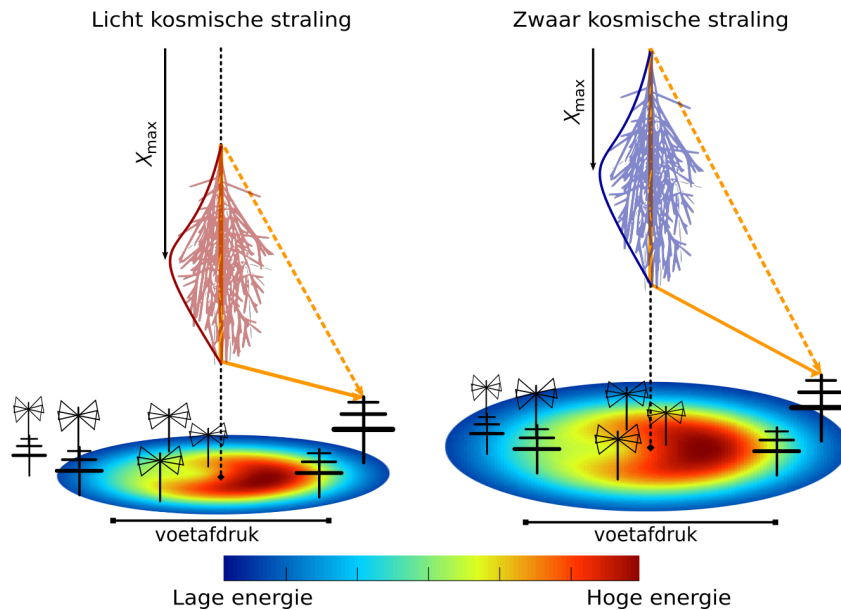
Figuur 1: Weergave van een deeltjeslawine, geïnitieerd door kosmische straling, gezien door de fluorescentie telescopen en de deeltjesdetectoren van het Pierre Auger Observatorium. De telescopen detecteren het licht dat wordt uitgezonden door atomen in de atmosfeer die aangeslagen zijn door de stortbui van deeltjes. De hoeveelheid gemeten licht is evenredig met het aantal deeltjes in de deeltjeslawine gemeten langs de richting van het oorspronkelijke deeltje. De kleur geeft de starttijd van het signaal aan: blauw is vroeg en rood is laat. De grijze punten zijn de deeltjesdetectoren op de grond die informatie verzamelen over het front van de deeltjeslawine. De hoogte van het signaal is evenredig aan de grootte van de gekleurde schijf en de kleur geeft net als hierboven de starttijd aan.



Figuur 2: Radio antenne station in de Argentijnse Pampa. Het hek dient ter bescherming tegen nieuwsgierige dieren.

per dag data kunnen registreren. Dit is een belangrijk voordeel aangezien het gaat om extreem zeldzame gebeurtenissen bij ultrahoge energieën. Daar komt nog bij dat radio antennes goedkoper en makkelijker te maken en onderhouden zijn in vergelijking met telescopen en andere deeltjesdetectoren, waardoor het eenvoudiger wordt om grote oppervlakken te bestrijken tegen een gunstige prijs.

In 2013 zijn ongeveer 150 radio antennes bij het Pierre Auger Observatorium geplaatst over een oppervlakte van  $17 \text{ km}^2$  om te onderzoeken of radio detectie een goed alternatief is voor andere detectietechnieken. Figuur 2 toont een afbeelding van een radio station. Het Pierre Auger Observatorium maakt het mogelijk om op een unieke wijze kosmische straling te meten met verschillende op zichzelf staande technieken en deze onderling te ijken.



Figuur 3: Impressie van de relatie tussen de grootte van de voetafdruk van de radio energie en  $X_{\text{max}}$ . Lichtere kosmische deeltjes dringen dieper door in de atmosfeer waardoor een smallere radio voetafdruk ontstaat. Ook is de lengte van het geprojecteerde radio signaal typisch groter voor lichte kosmische straling dan voor zware. De lengte van het radio signaal komt overeen met het verschil tussen de tijd die het radio signaal erover doet om van het begin van het emissie gebied naar de antenne te komen (oranje gestreepte lijn) en de tijd dat het signaal erover doet om bij de antenne te komen vanaf het eind van het emissie gebied, plus de totale tijd waarin de deeltjeslawine radiogolven uitzendt (oranje doorgetrokken streep).

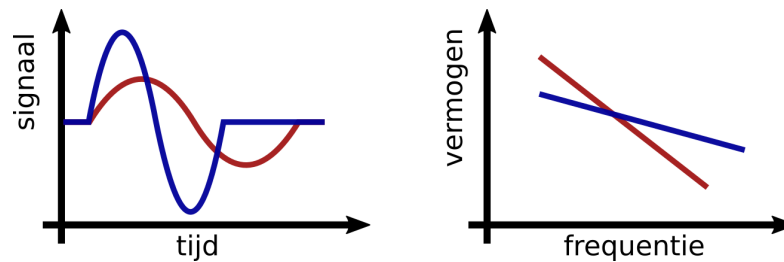
### Massa meten met radio antennes

Er zijn verschillende theorieën die de oorsprong en voortplanting door de ruimte van kosmische straling proberen te verklaren. Om te bepalen welke theorie de juiste is, en het raadsel van ultrahoog-energetische kosmische straling op te lossen, is het cruciaal om de massasamenstelling te weten te komen. De massasamenstelling geeft aan hoe vaak verschillende soorten deeltjes relatief voorkomen. In dit proefschrift is een nieuwe methode beschreven om de massasamenstelling van kosmische straling met de hoogste energie vast te stellen met behulp van radio signalen.

Een zwaar kosmisch deeltjes botst gemiddeld eerder en dus hoger in de atmosfeer op een luchtmolecuul dan een licht kosmische deeltje. Deze eerste botsing veroorzaakt een deeltjeslawine waarbij het aantal secundaire deeltjes toeneemt tot een bepaald maximum en dan langzaam afneemt. De diepte in de atmosfeer  $X_{\text{max}}$  waar zich de meeste deeltjes bevinden, correleert met het type primaire deeltje.

De op het radio signaal gebaseerde  $X_{\text{max}}$  is verkregen door het combineren van twee onafhankelijke methodes. Het eerste maakt gebruik van de relatie tussen de vorm van de energie van het radio signaal op de grond en het primaire deeltje. Dit is te zien in Figuur 3: een licht deeltje dring dieper door in de atmosfeer en produceert een smallere energie voetafdruk in vergelijking tot een zwaar primair kosmisch deeltje.

De tweede methode is gebaseerd op de frequentie inhoud van het radio signaal. Zoals te zien in Figuur 4: lichtere deeltjes produceren een langer radio signaal wat zich vertaalt naar een

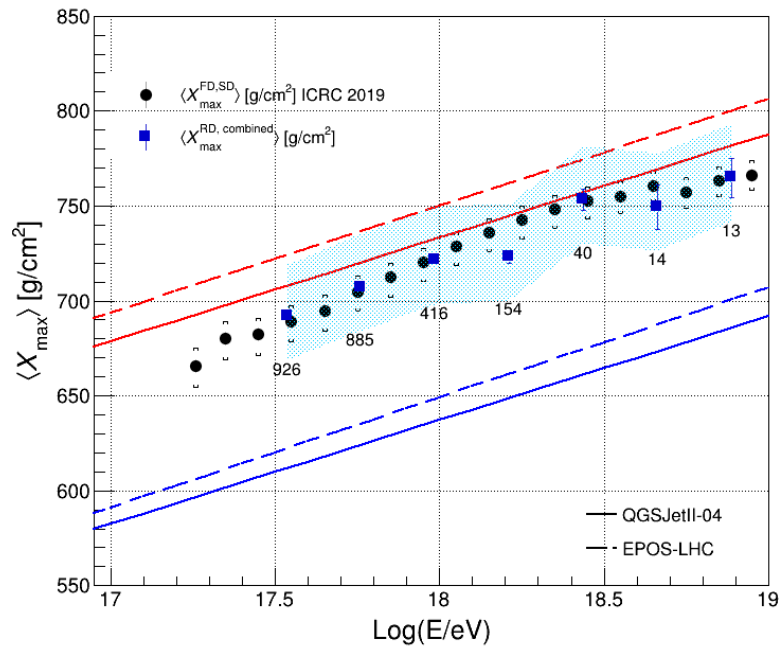


Figuur 4: Een lang signaal afkomstig van een licht kosmisch deeltjes resulteert in een steiler frequentie spectrum dan een kort signaal. De verschillende kleuren komen overeen met de deeltjeslawines in Figuur 3.

steiler frequentie spectrum.  $X_{\max}$  kan dus bepaald worden aan de hand van de helling van het frequentie spectrum dat gemeten wordt in elk radio station.

Om de gemiddelde massacompositie van kosmische straling als een functie van de energie van het primaire deeltje te bepalen, gebruiken we de gemiddelde  $X_{\max}$  van alle gemeten deeltjes met een vergelijkbare energie. Fluorescentie telescopen kunnen  $X_{\max}$  heel precies meten omdat zij letterlijk de deeltjeslawine zien groeien en uitdoven in de atmosfeer. In Figuur 5 is dus  $X_{\max}$  zoals gehaald uit het radio signaal vergeleken met de meting van de fluorescentie telescopen. Om de waarde van  $X_{\max}$  te vertalen naar een massacompositie wordt de gemeten  $X_{\max}$  vergeleken met de gemiddelde waarde van een zuiver lichte (rood) en zware (blauw) compositie. De gemeten punten van onze methode geven aan dat bij hogere energieën de massacompositie steeds lichter wordt.

Uit de vergelijking tussen de radio en fluorescentie techniek kan er geconcludeerd worden dat het mogelijk is om betrouwbare compositie informatie te halen uit radio signalen afkomstig van deeltjeslawines geïnitieerd door kosmische straling. Het plan voor de toekomst is om radio antennes over grote oppervlakken te verspreiden om ultrahoog-energetische kosmische deeltjes te vangen. Het Pierre Auger Observatorium gaat elke deeltjesdetector op de grond uitrusten met een radio antenne om zo een 3000 km<sup>2</sup> radio detector creëren. Een ander experiment, het Giant Radio Array for Neutrino Detection (GRAND), wil een grootschalige detector met 200 000 goedkope radio antennes over een oppervlakte van 200 000 km<sup>2</sup> maken op verschillende locaties. Deze ambitieuze experimenten zullen een nieuw tijdperk inluiden voor radio detectie van kosmische straling, welke zeker zullen bijdragen aan het ontrafelen van de meest energetische mysteries van ons universum.



Figuur 5: Gemiddelde  $X_{\max}$  als een functie van de energie van de kosmische straling. De blauwe punten zijn verkregen uit de radio signalen en de zwarte punten zijn de metingen van de fluorescentie telescopen. Theoretische voorspellingen voor lichte (rood) en zware (blauw) kosmische straling zijn weergegeven voor twee verschillende modellen, QGSJetII-04 (doorgetrokken lijn) en EPOS-LHC (gestipelde lijn).



# Curriculum Vitae

Fabrizia Canfora was born on the 25th of October 1988 in Naples, Italy. She studied Physics at the University of Naples Federico II. In June 2014, she completed her Master's degree *cum laude* in Subnuclear and Astroparticle Physics. During her master's project she tested the performance of the first prototype of a new photodetector (Vacuum Silicon PhotoMultiplier Tube, VSiPMT) developed at the University of Naples and produced by Hamamatsu Photonics. She also investigated the applicability of the VSiPMT to readout of the KM3NeT optical module using Geant4 simulations.

In 2011, she was selected for a summer project on gravitational waves within the LIGO experiment, and spent three months at the University of Mississippi, where she worked on an algorithm to detect noise and environmental disturbances in data collected by the LIGO interferometer.

After her master diploma, she started a research fellowship at the University of Naples and joined the MU-RAY project to work on volcano radiography with cosmic-ray muons.

In September 2015, Fabrizia Canfora started as a graduate student at the department of Experimental High Energy Physics at the Radboud University Nijmegen under the supervision of Dr. Charles Timmermans and Prof. Sijbrand de Jong. Her work focused on cosmic-ray composition measurements with the AERA radio detector of the Pierre Auger Observatory. The results of her PhD research are reported in this thesis.





# Acknowledgements

Many people helped and supported me along the way on this incredible journey.

First and foremost, I am extremely grateful to my supervisors, Dr. Charles Timmermans and Prof. Sijbrand de Jong, who have helped and taught me a lot during these years. I want to thank both of you for the fruitful scientific discussions and also for the support you gave me every time I got stuck or frustrated about a ‘bad’ result. Your honest enthusiasm and passion always renewed my curiosity and motivation, I consider myself very lucky to have had such great supervisors.

I would also like to thank my second promoter Prof. Bentvelsen, and the members of the manuscript committee: Prof. Hörandel, Dr. Castellina, Prof. Huege, Prof. Peitzmann and Dr. Schoorlemmer for taking the time to carefully review this thesis.

Thank you to all the people who joined the Dutch Auger meetings and the AERA calls and workshops: I have always received important inputs and valid suggestions from our discussions and collaboration. I would like to thank in particular Christian Glaser, who provided me with the CoREAS simulation sample used in my analysis, and the PhD students and postdocs of the AERA group who have helped me with my complicated relationship with *Offline*, and with whom I enjoyed many beers (and kölsch) around the world.

Thanks to Mart for translating my summary in Dutch, and for taking the tough decision to ‘deforest’ our office.

A special acknowledgment goes to the secretaries of the HEP department, Annelies, Gemma, Marjo and Gertie, and to Ed from Nikhef for the enormous help with the bureaucracy and many other things.

Getting to the end of this journey required more than academic support. Thanks, bedankt, grazie, gracias, спасибо to my friends and fellow PhD students from the HEP department and Nikhef. We discussed our work-related problems, giving each other feedback and support, but we also had a lot of fun together during our dinners, happy hours on the beach, and our long long (four days long) walks. In these intense years we built a deep and meaningful friendship and I am very happy our paths crossed. Thank you for making this journey so memorable!

A big thank you is for my *good old* friends from Naples, many of them are now also expats like me. Whether for a drink in Naples, a weekend in the Netherlands, a British pie in Oxford, or a wedding in Albania, I always feel *home* when we are together. A special mention is for my *BBC*: our chat history is a perfect diary of the ups and downs of PhD life, thank you guys for always being there.

Finally, I thank my family for supporting me in whatever I am up to, whenever I need it and wherever I am. Thank you for teaching me to face challenges and to take some risks in order to get somewhere, and that hard work will eventually be rewarded. Thank you for celebrating with me every important moment. Grazie famiglia, tutto è più facile con voi al mio fianco.