

RADIO FOR THE MASSES

*Cosmic ray mass composition measurements
—— in the radio frequency domain ——*



Stefan Jansen

Radio for the Masses

Cosmic ray mass composition measurements in the radio frequency domain

Stefan Jansen

© Stefan Jansen 2016

Radio for the Masses

– Cosmic ray mass composition measurements in the radio frequency domain
Thesis, Radboud University Nijmegen

viii + 164 pages; illustrated, with bibliographic references and summary in Dutch

ISBN 978 94 028 0073 9

Cover design by Mario Verkerk

Printed in the Netherlands by Ipskamp Drukkers, Enschede

This work is part of the research programme of the *Stichting voor Fundamenteel Onderzoek der Materie* (FOM), which is part of the *Nederlandse Organisatie voor Wetenschappelijk Onderzoek* (NWO)



Radio for the Masses

Cosmic ray mass composition measurements in the radio frequency domain

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus,
volgens besluit van het college van decanen
in het openbaar te verdedigen
op woensdag 13 april 2016 om 16.30 uur precies

door

Stefan Jansen

geboren op 18 augustus 1985
te Ludvika, Zweden

PROMOTOR Prof. dr. S.J. de Jong

COPROMOTOR Dr. C.W.J.P. Timmermans

MANUSCRIPTCOMMISSIE Prof. dr. P.C.M. Christianen (voorzitter)

Prof. dr. F. Linde

Universiteit van Amsterdam

Prof. dr. S. Buitink

Vrije Universiteit Brussel (België)

Dr. O. Martineau-Huynh

Université Pierre et Marie Curie (Parijs, Frankrijk)

Dr. J.R. Hörandel

CONTENTS

CHAPTER 1

Introduction	1
--------------	---

CHAPTER 2

Theory of cosmic rays	5
2.1 Ultra-high-energy cosmic rays	5
2.1.1 Energy spectrum	7
2.1.2 Sources, acceleration mechanisms and propagation	8
2.2 Extensive air showers	10
2.2.1 Shower development and components	11
2.2.2 Longitudinal shower profile	13
2.2.3 Primary cosmic ray composition	14
2.3 Radio emission from extensive air showers	15
2.3.1 Emission mechanisms	15
2.3.2 Modeling radio emission from air showers	16
2.3.3 Radio emission and air shower parameters	18

CHAPTER 3

Method: Composition signatures in the radio pulse shape	19
3.1 The radio pulse shape	19
3.2 The spectral index	22
3.3 Simulation data set	24
3.3.1 Simulation setup	24
3.3.2 Initial conditions	25
3.3.3 Post-processing	26
3.3.4 Limitations and coherence	28
3.4 Parameterization of the spectral index	29
3.4.1 Separate radio emission contributions	29
3.4.2 Mixed radio emission contributions	32
3.5 Parameterization performance	34
3.6 Discussion	35

CHAPTER 4

The detector: The Pierre Auger Observatory and AERA	39
4.1 The Surface Detector	40
4.2 The Fluorescence Detector	43

CONTENTS

4.3	The Auger Engineering Radio Array	45
4.3.1	Introduction	45
4.3.2	Antenna station design	46
4.3.3	Trigger and data acquisition	49
4.3.4	Electric field reconstruction	51
<hr/>		
CHAPTER 5		
	Data sets and event selection	57
5.1	Radio frequency interference at AERA	57
5.1.1	Narrowband RFI	58
5.1.2	Broadband RFI	59
5.2	Station performance monitoring	60
5.2.1	Method	60
5.2.2	Results	62
5.3	Weather monitoring	65
5.4	Event selection	66
5.4.1	Scintillator-triggered setup	67
5.4.2	Externally triggered setup	70
5.4.3	Additional selection criteria	73
5.4.4	Summary of radio data selection	77
5.4.5	Selection of fluorescence data	79
<hr/>		
CHAPTER 6		
	Calibration and cleaning of AERA data	83
6.1	Relative amplitude calibration of AERA-II stations	83
6.1.1	Method: calculation of calibration constants	84
6.1.2	Exclusion of broadband RFI	85
6.1.3	Exclusion of narrowband RFI	85
6.1.4	Interpolation	87
6.1.5	Application to AERA-II data	88
6.1.6	Integration into the software framework	92
6.2	Absolute calibration of AERA-II stations	92
6.2.1	Motivation	92
6.2.2	Method	93
6.2.3	Results	95
6.3	Suppression of narrowband RFI in the time domain	99
6.3.1	Method	101
6.3.2	Dynamic RFI identification and parameter estimation	101
6.3.3	Results	103
6.3.4	Integration into the software framework	104

CHAPTER 7

Results: Measuring shower development with AERA	105
7.1 The spectral index in AERA data	105
7.1.1 Extraction of the pulse spectrum	105
7.1.2 The spectral index and its uncertainties	107
7.2 Dependence of the spectral index on shower development	113
7.2.1 Measurements of the spectral index	114
7.2.2 Event-by-event reconstruction of shower maximum	114
7.2.3 Comparison with fluorescence measurements	118
7.2.4 Average depth of shower maximum in AERA data	121

CHAPTER 8

Discussion and outlook	131
8.1 Discussion of results	131
8.2 Outlook and recommendations	132

BIBLIOGRAPHY	135
---------------------	------------

SUMMARY	143
----------------	------------

LIST OF PUBLICATIONS	151
-----------------------------	------------

SAMENVATTING	153
---------------------	------------

CURRICULUM VITAE	161
-------------------------	------------

ACKNOWLEDGEMENTS	163
-------------------------	------------

Introduction

IN OUR DAY-TO-DAY LIVES, we are exposed to a constant torrent of microscopic particles raining down on us. This never-ending stream of electrons, neutrinos, muons and many other particles is the debris from collisions between charged particles from distant astronomical sources and molecules in Earth's atmosphere. These charged particles are what we call *cosmic rays*. The most energetic cosmic rays have energies far beyond the reach of current accelerator experiments.

Even though their existence has been known for more than a century, the nature and origin of cosmic rays of the highest energies remains elusive. The mysteries of high-energy cosmic rays can be captured in four fundamental questions:

- *What are the sources of high-energy cosmic rays?*
- *By which processes are they accelerated?*
- *How do they propagate towards Earth?*
- *What are the laws governing their interactions in Earth's atmosphere?*

The most important obstacle in conclusively answering these questions is the presence of (inter)galactic magnetic fields, which influence the trajectory of the particles on their way towards Earth. Because of the deflection in these fields, it is difficult – if not impossible – to point back measured cosmic rays towards their sources. Figure 1.1 summarizes the four fundamental questions and their interrelations.

As of yet, only few of the many proposed solutions to these four fundamental questions have been verified experimentally. The key to unravel the questions is to understand cosmic ray *composition*. If we have knowledge about the primary particle type, and therefore its charge, we might be able to create sub-sets of measured cosmic rays that are less affected by the magnetic fields, such as protons. These could direct us towards the sources of cosmic rays, which in turn will teach us about their acceleration and

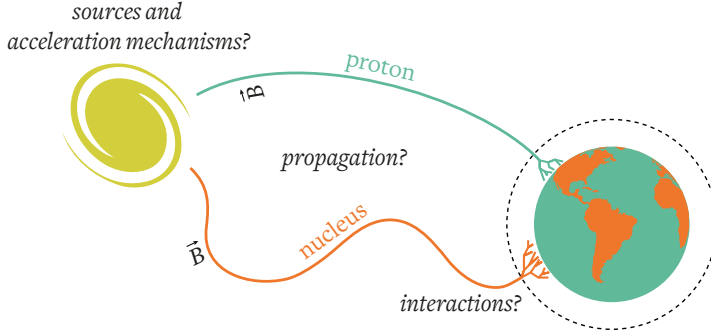


Figure 1.1: The four fundamental questions in the field of high-energy cosmic rays. Deflection by (inter)galactic magnetic fields hinders a direct observation of their sources, which might provide us with information about the acceleration mechanisms. At the same time, the physics underlying the first interactions in Earth’s atmosphere is uncertain because they occur at energies beyond the reach of collider experiments.

propagation. The goal of this thesis, therefore, is not to answer one of the fundamental questions directly, but to develop experimental techniques to be able to determine the composition of high-energy cosmic rays.

The debris from the collision of the cosmic ray with an atmospheric molecule that arrives at the earth’s surface is the tail of a cascade of interactions and decays: the *air shower*. A schematic representation of an air shower can be seen in figure 1.2. In this thesis we will focus on the radio pulses emitted by the charged particles in the air shower as they travel downward. More specifically, we will investigate which properties of the air shower and the cosmic ray we can derive from the frequency content of the radio pulse. Ultimately, we will deduce information about air shower development, from which we can infer the composition of the cosmic ray.

In chapter 2 of this thesis, we outline the general theoretical framework of cosmic rays, air showers and radio emission. The fundamental questions of high-energy cosmic rays and their possible solutions are also addressed in more depth in this chapter. After this general theory, we investigate the dependence of the frequency spectrum of the radio pulse on air shower parameters using simulations in chapter 3. This investigation has resulted in a method to extract information about air shower development through a parameterization of the spectral index.

To experimentally verify our method we have used data from the Pierre Auger Observatory in Argentina, which is described in chapter 4. The quality of this data is discussed in chapter 5, which also details selection procedures to enhance the quality of the data

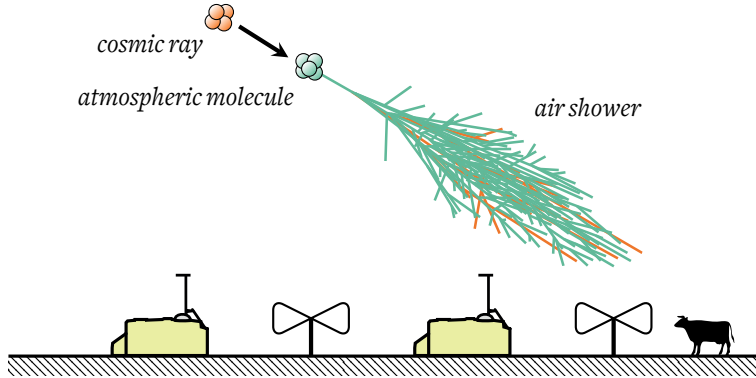


Figure 1.2: Schematic representation of a cosmic ray induced air shower above a detector plane filled with several types of detectors and a cow.

set. Chapter 6 stipulates several calibration and cleaning methods we have performed on the radio data.

We study the dependence of the measured spectral index from the calibrated data on air shower parameters in chapter 7. Here, we also test the parameterization of the spectral index we developed earlier. This has resulted in measurements of the composition of air showers using radio data, which are discussed in chapter 8. This last chapter also contains several recommendations for future improvements of radio observatories for cosmic rays.

Theory of cosmic rays

THE NATURE AND ORIGIN of cosmic rays have occupied and mystified scientists for more than one hundred years. At the dawn of the twentieth century, both the ionization of air and radioactivity were known. It was thought that decaying radioactive isotopes in the earth's crust were responsible for the ionization of air and the charges that were readily measurable by electroscopes.

It was the Austrian Victor Hess, however, who demonstrated in 1912 that when moving further from the earth's surface, the amount of ionizing radiation was increasing instead of decreasing, which would have been expected if the source of the ionizing radiation was the earth [1]. He measured this in a series of daring balloon flights, where he reached altitudes of up to five kilometers. His conclusion was that the earth was not the dominant source of the radiation, and that it must originate from above (he already excluded the Sun in a balloon flight during a solar eclipse). This discovery essentially started the field of cosmic rays, and won Hess the Nobel prize in 1936.

In the course of the 1920s and 1930s it became clear that the measured “cosmic” rays were actually part of cascades of secondary particles originating from one primary cosmic ray interacting with the particles in the atmosphere: *extensive air showers*. It was Pierre Victor Auger who confirmed their existence by measuring coincident events between two Geiger counters positioned 300 m apart [2]. Furthermore, he estimated that the primary cosmic ray energy that produces this kind of showers must be at least 10^{15} eV.

2.1 Ultra-high-energy cosmic rays

More than 75 years after Auger's energy estimate, a cosmic ray energy spectrum covering more than eleven decades in energy has been measured. The measurements range from a few GeV up to an energy of more than 10^{20} eV. The energy distribution as mea-

sured by a combination of several experiments is plotted in figure 2.1. The differential flux J of cosmic rays as a function of energy E can be described using a power law:

$$J \propto E^{-\gamma}, \quad (2.1)$$

where the spectral index γ is energy-dependent and has a value close to 3. At low energies – up to about 10^{14} eV – the flux of cosmic rays is still high enough to allow for direct detection by relatively small experiments carried to high altitudes using satellites or balloons. Examples of direct detection experiments include PAMELA [3] and BESS [4], but also the Voyager satellite [5].

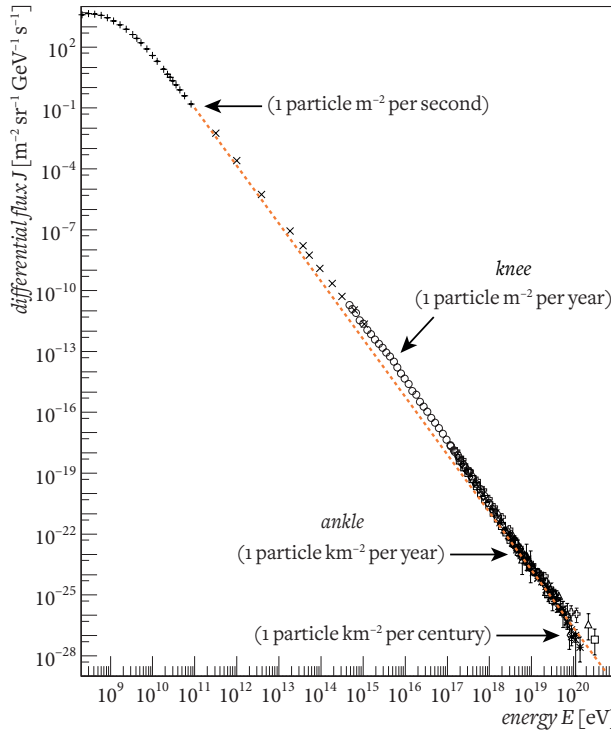


Figure 2.1: Cosmic ray energy spectrum spanning 11 orders of magnitude. The dashed line follows a power law with a constant slope. This emphasizes the main features in the spectrum, which are indicated by arrows and are discussed in section 2.1.1 (adapted from [6]).

At higher energies, the flux becomes too low to effectively detect cosmic rays with direct measurement experiments, which typically have an aperture of not much more than 1 m. Instead, we have to rely on ground-based observatories which measure the extensive air showers produced by high-energy cosmic rays. Haverah Park [7], AGASA [8], Yakutsk [9] and HiRes [10] are just some examples of cosmic ray observatories. Above

10^{18} eV we enter the realm of *ultra-high-energy cosmic rays* [11]. Cosmic rays of these energies become exceedingly rare, and at 10^{20} eV the flux has decreased to one particle per km^2 per century. This means that enormous experiments with huge apertures are needed to be able to probe the properties of cosmic rays at these energies.

Ultra-high-energy cosmic rays present an enigma to the science community. There are only few types of astronomical sources that would be able to accelerate particles to these energies, but at the same time there are limitations on the distance charged particles of these energies can travel through the universe. Yet, up to now it has been impossible to identify actual sources, and various theories exist about the mechanisms to accelerate these particles. By studying the cosmic ray properties accessible to us on earth, their energy, composition and arrival direction, we can attempt to unravel the fundamental and interconnected questions about the sources, acceleration and propagation of ultra-high-energy cosmic rays.

2.1.1 Energy spectrum

Several kinks can be observed in the otherwise smooth energy spectrum of cosmic rays. These features, which are indicated in figure 2.1, are related to the sources and acceleration of cosmic rays as well as their propagation towards Earth. For example, it has been proposed that the *knee* at $\sim 10^{15}$ eV might be the result of subsequent energy cutoffs of individual elements in galactic sources [12]. The tail of the energy distribution as measured by the Pierre Auger Observatory is displayed in figure 2.2.

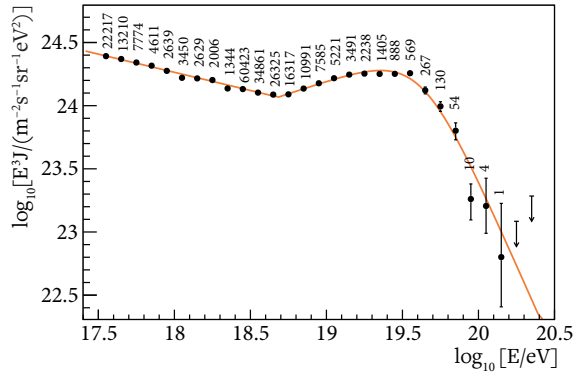


Figure 2.2: Energy spectrum at the highest energies as measured by the Pierre Auger Observatory, fitted by a flux model above the ankle energy [13]. The systematic uncertainty on the energy scale is 14 %. The numbers indicate the number of events that were used to calculate each data point.

This provides a closer look at the *ankle* region of the spectrum around $\sim 3 \times 10^{18}$ eV. The ankle is thought to represent the transition of cosmic rays from galactic sources

to primarily extragalactic sources, as the galactic magnetic field cannot effectively trap particles above this energy (see e.g. [14, 15]). Additional explanations of the knee include photodisintegration of nuclei near the sources of cosmic rays [16], and a modification of the cosmic proton flux at the location of the ankle is predicted because of a dip in the pair production cross section [17].

Another remarkable feature of the spectrum at the highest energies is the strong suppression of the flux at energies higher than $\sim 3 \times 10^{19}$ eV. The cause of this cutoff is still under debate. A probable scenario is that it is produced by a combination of the inability of the sources to accelerate cosmic rays to even higher energies, and limitations imposed by the propagation through intergalactic space (see e.g. [18]).

2.1.2 Sources, acceleration mechanisms and propagation

The primary mechanism by which cosmic rays acquire their extremely high energies is believed to be Fermi acceleration. Proposed by Enrico Fermi in 1949, this model describes how charged particles statistically gain energy by multiple collisions with the shock waves in a turbulent magnetic field [19]. This form of acceleration naturally leads to an energy spectrum following an inverse power law. Although the rate of energy gain is initially very slow, it is capable of reaching very high energies, given sufficiently large sources to contain the particles.

We can get an idea of the size of the sources by assuming that the Larmor radius R_L of a particle of charge number Z and energy E needs to be contained within the acceleration region of the source with magnetic field B and plasma velocity βc [20]:

$$\left(\frac{E}{10^{18} \text{eV}} \right) < \frac{Z\beta}{2} \left(\frac{R_L}{\text{kpc}} \right) \left(\frac{B}{\mu\text{G}} \right). \quad (2.2)$$

This equation provides an upper limit of the energy to which particles of a certain charge can be accelerated, given the size of a source and its magnetic field. For cosmic rays of the highest energies, exceedingly vast sources or extremely powerful magnetic fields are required to accelerate them. Particles with higher charge are easier to confine in the magnetic fields, so they can be accelerated to higher energies, which supports the explanation of the knee in the cosmic ray energy spectrum. With equation (2.2), it is possible to construct the *Hillas plot*: an overview of potential astronomical sources of cosmic rays, which can be seen in figure 2.3.

The energy of cosmic rays is not just limited by the ability of the sources to accelerate them. On their way towards Earth, the cosmic rays can be subjected to several different types of interactions and decays, limiting the range of particles with the highest energies. Greisen, Zatsepin and Kuzmin have predicted that cosmic ray protons would interact with photons of the cosmic microwave background (CMB) via the Δ -resonance,

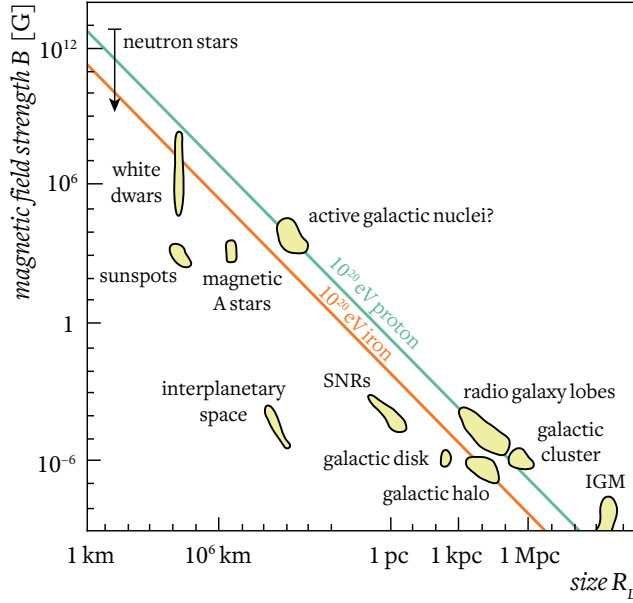
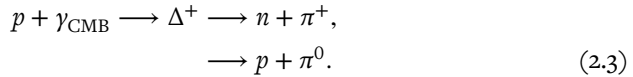


Figure 2.3: Hillas plot with several different astronomical sources of cosmic rays and diagonal lines illustrating the minimum size and magnetic field of cosmic rays of 10^{20} eV, assuming $\beta = 1$ (adapted from [20]).

producing a pion [21, 22]:



This process has a significant cross-section for cosmic ray protons with an energy of the order of 10^{20} eV, which impairs their ability to travel large distances. Several other processes may further increase the rate of attenuation of cosmic rays (see [23] for an overview). These include electron-positron pair production, interactions with the infrared background radiation, and photodisintegration in the case of heavier nuclei. Several processes and their associated attenuation lengths as a function of energy for different particles can be seen in figure 2.4. Because of these limitations it is likely that the sources of cosmic rays of the highest energies are not very far away on intergalactic distance scales.

In addition to background photons, the space between us and the cosmic ray sources is filled with galactic and intergalactic magnetic fields by which the cosmic rays are deflected. It is therefore difficult to directly point cosmic rays with known arrival directions back to their sources, depending on their charge and the strength of the magnetic fields. The Larmor radius of a particle with charge number Z and energy E in a mag-

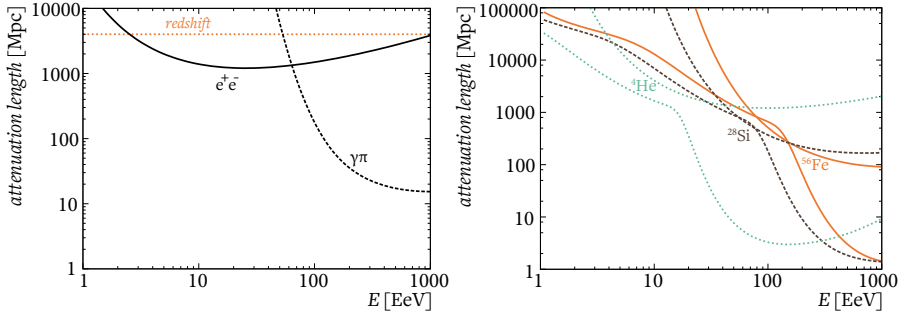


Figure 2.4: Attenuation of cosmic rays resulting from various processes for different particle types as a function of energy [23]. *Left:* proton; *right:* several specific nuclei, where the lower curves are due to photodisintegration and the upper curves are due to pair production. Note the different scale on the vertical axes.

netic field of strength B is given by $R_L = E/ZeB$, with which we can approximate the deflection angle α of a particle moving over a distance d as $\alpha \approx d/R_L$. By virtue of their charge, it is obvious that iron nuclei are much more affected by magnetic deflections than protons. Although the nature of the extragalactic magnetic fields is not exactly known, they are assumed to be structured with a typical size of about one Mpc, and their typical strength is assumed to be in the order of one nG (see e.g. [24]). It is believed that cosmic protons of the highest energies might therefore point back towards their sources.

The preceding treatment of ultra-high-energy cosmic rays illustrates that the key to understanding their origin is measuring their composition. Acceleration, attenuation and magnetic deflection all depend on the composition of the particles, so in order to differentiate between all the different candidate sources we need to distinguish between the different particles. Unfortunately, the composition is not a directly accessible parameter in indirect measurements concerning extensive air showers, and can only be inferred from shower development, as we shall see in the next section.

2.2 Extensive air showers

When a cosmic ray enters the earth's atmosphere it will collide with an atmospheric nucleus. This collision creates secondary particles which themselves collide with atmospheric nuclei or initiate a cascade of decay products which together move towards the earth's surface: the air shower. An overview of the primary interactions in the air shower and its *components* can be seen in figure 2.5. The interactions and components will be discussed in more depth in this section.

2.2.1 Shower development and components

The initial collision creates a host of hadrons, mostly pions but also kaons and baryons, and depending on the composition of the primary cosmic ray, secondary nuclear fragments. This part of the shower is called the *hadronic component*.

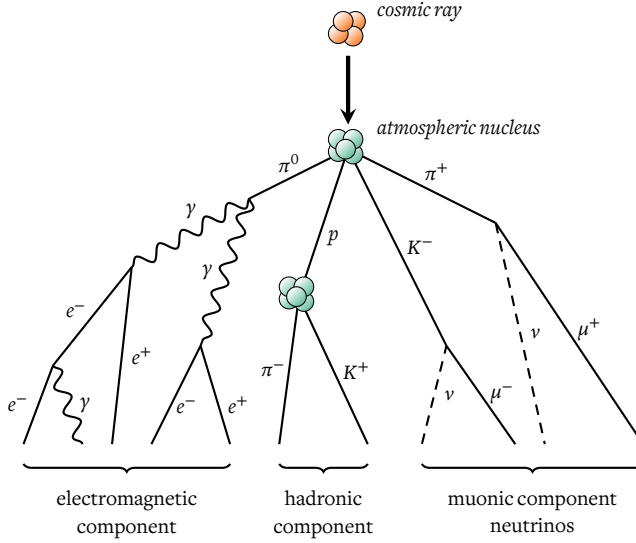


Figure 2.5: Schematic overview of the primary interactions in an extensive air shower and the division into components.

Neutral pions have a life time of just $\sim 10^{-16}$ s, and will therefore almost immediately decay into two gamma rays:

$$\pi^0 \longrightarrow \gamma\gamma. \quad (2.4)$$

These photons initiate sub-cascades of electrons, positrons and photons through pair production, bremsstrahlung, ionization and Compton scattering. This *electromagnetic component* of the air shower will continue to grow until the energy of the electrons falls below the critical value necessary for radiative processes, at which point they will mostly ionize the surrounding air and the component starts to die out. The charged pions have a longer decay time than the neutral pions, and will either interact with the atmosphere to create more secondary particles, or decay into muons:

$$\pi^\pm \longrightarrow \mu^\pm + \nu. \quad (2.5)$$

The muons have long life times and small cross-sections, and will therefore travel to the ground almost unimpeded. They constitute the *muonic component* of the shower. The remainder of the hadronic component will continue to interact and decay and feed into

the muonic and electromagnetic components of the shower. For example, the kaons can decay into pions or straight into muons:

$$\begin{aligned} K^\pm &\longrightarrow \pi^\pm + \pi^0, \\ &\longrightarrow \mu^\pm + \nu. \end{aligned} \quad (2.6)$$

Once the shower reaches ground level, the hadronic component has almost completely been converted into the other two components. Because the three different kinds of pions are created in roughly the same abundance in hadronic interactions, about one third of the available energy will be converted into the electromagnetic component at each hadronic interaction, and only a small fraction of the charged pions are converted into muons at high energy. As a consequence, the electromagnetic component carries the bulk of the available energy of the air shower. An example of the energy flow in the different shower components in a shower simulation can be seen in figure 2.6.

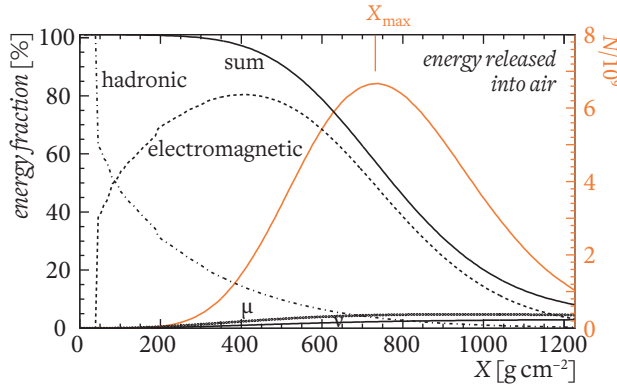


Figure 2.6: Energy flow in a 10^{19} eV proton shower simulated by the CORSIKA Monte Carlo code as a function of atmospheric depth X (see equation (2.7)). The fraction of energy stored in the different components is shown, including the sum of all the energy stored in particles, with the rest released into the air. The total number of electrons and positrons is indicated in orange and displayed on the right-hand scale (adapted from [25]).

Because of the huge number of particles involved in the interactions and the stochastic nature of the interaction processes, Monte Carlo simulations are used to study the properties of air showers, which take into account all interactions and decays. Examples of full Monte Carlo codes to study air shower characteristics are CORSIKA [26] and AIRES [27]. The energies necessary to determine the cross-sections of the first interactions of ultra-high-energy cosmic rays have not been reached by collider experiments, which means they have to be extrapolated (see e.g. [28]).

2.2.2 Longitudinal shower profile

Since the interaction lengths in the air shower are a function of the amount of traversed air, shower development is usually expressed as a function of atmospheric depth X instead of height:

$$X(h) = \int_h^{\infty} \rho(h') dh', \quad (2.7)$$

where h is the vertical height in the atmosphere, and ρ is the density, which is integrated from h to the top of the atmosphere along the axis of the shower.

Even though a full characterization of the developing air shower requires Monte Carlo simulations, a simple toy model can already help us to gain some insight into the main properties of the air shower as a function of the energy and composition of the primary cosmic ray. The Heitler model of air showers [29] considers a primary particle with energy E_0 , which splits into two secondary particles after traversing an amount of atmosphere λ , and repeats this process thereafter. At atmospheric depth X , the cascade has evolved into $N(X) = 2^{X/\lambda}$ particles, each carrying an energy of $E(X) = E_0/N(X)$. At some point, the particles will have reached the critical energy E_c at which they are unlikely to interact any further, and the maximum number of particles N_{\max} is reached. This number and its associated atmospheric depth X_{\max} are then:

$$N_{\max} = E_0/E_c, \quad (2.8)$$

$$X_{\max} = \lambda/\ln 2 \cdot \ln(E_0/E_c). \quad (2.9)$$

This demonstrates that the maximum number of particles as well as the depth of shower maximum is dependent on the energy of the primary particle. To illustrate the effect of a primary cosmic ray of mass A , we can assume a superposition of A independent particles of energy $E_A = E_0/A$. Using the same logic as above, we can again calculate the maximum number of particles and the depth of shower maximum:

$$N_{\max} = A \cdot E_A/E_c = E_0/E_c, \quad (2.10)$$

$$X_{\max} = \lambda/\ln 2 \cdot \ln(E_0/AE_c). \quad (2.11)$$

We can therefore predict that the maximum number of particles stays the same for primaries of different mass, but X_{\max} has a logarithmic dependence on the number of nucleons.

These results can be compared with the longitudinal shower profiles from full Monte Carlo simulations, such as the ones displayed in figure 2.7. These shower profiles show the same qualitative behavior as predicted by the Heitler model: the maximum number of particles is the same in all showers with the same energy, and a heavier primary cosmic ray produces a shower which generally has a smaller depth of shower maximum X_{\max} .

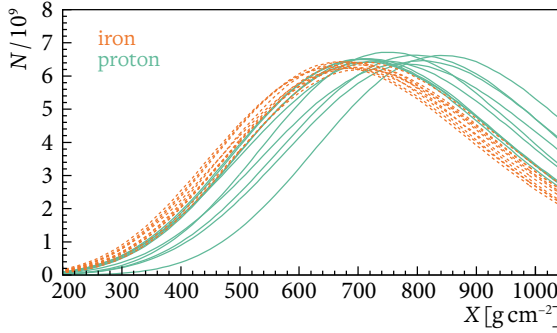


Figure 2.7: Several individual longitudinal shower profiles simulated with the CORSIKA Monte Carlo code [25]. The proton and iron primary cosmic rays with an energy of 10^{19} eV result in differently developing showers, expressed by the number of charged particles N .

The longitudinal shower profile is furthermore shaped by the depth of first interaction, which also depends on the cross-section and therefore on the composition of the primary cosmic ray. The position of shower maximum X_{\max} can be fitted to the profile using the Gaisser-Hillas function [30]:

$$N(X) = N_{\max} \left(\frac{X - X_0}{X_{\max} - X_0} \right)^{\frac{X_{\max} - X}{\lambda}} \exp \left(\frac{X_{\max} - X}{\lambda} \right), \quad (2.12)$$

where λ and X_0 are both shape parameters, with the former related to the characteristic decay length of the particles and the latter to the initial interaction point.

2.2.3 Primary cosmic ray composition

The average value of X_{\max} is a composition-sensitive parameter, and so are its shower-to-shower fluctuations, as was already visible in the simulations of figure 2.7, where the spread in X_{\max} is much larger for the proton-induced showers than the iron-induced showers. The average and the standard deviation of X_{\max} as a function of energy as measured by the Pierre Auger Observatory are plotted in figure 2.8. The lines in these plots indicate the values for pure iron and proton primaries predicted by various simulation codes using different interaction models. Based on these results it seems that the trend is moving towards heavier primaries at the highest energies, but the interpretation of these results strongly depends on the results from interaction models. The measurements of the average depth of shower maximum from the Pierre Auger Observatory are in agreement with the results the Telescope Array Project [31, 32].

In addition to the depth of shower maximum, several other parameters of the air shower are sensitive to the composition of the primary cosmic ray. One of them is the *muon*

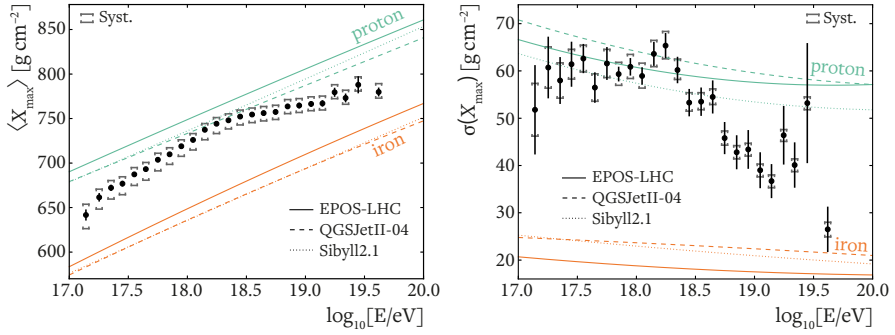


Figure 2.8: Average and standard deviation of the depth of shower maximum X_{\max} as a function of energy as measured by the Pierre Auger Observatory [33].

number. The muonic component of the shower originates in the hadronic component of the shower only, and is correlated with the mass of the primary particle. The composition of the primary cosmic ray can therefore be deduced by measuring the number of muons at ground level, either in inclined showers in which the electromagnetic component has died out or with special muon-counters (see e.g. [34, 35] and references therein).

2.3 Radio emission from extensive air showers

Charge creation and acceleration in the air shower induces radio emission, a phenomenon which follows directly from their description using classical electrodynamics. In the 1960s, it was measured for the first time [36], and at around the same time the principal processes by which the radio emission is produced were already identified.

2.3.1 Emission mechanisms

The primary process by which radio emission is produced is the *geomagnetic* mechanism, described by Kahn and Lerche in 1966 [37]. The electrons and positrons in the air shower are accelerated in opposite directions under the influence of the geomagnetic field. This acceleration is along the direction of the Lorentz force vector $\mathbf{v} \times \mathbf{B}$, where \mathbf{v} is the direction of movement of the shower and \mathbf{B} the geomagnetic field. The resulting transverse current produces radio emission which is linearly polarized in the $\mathbf{v} \times \mathbf{B}$ direction. The strength of the geomagnetic component of the radio emission during shower development depends on the change in the amount of charge carriers in the shower, and therefore scales roughly with the derivative of the total number of shower particles N .

In addition to the geomagnetic mechanism, the *charge-excess* mechanism also contri-

butes to the radio emission of air showers. The charge-excess or Askaryan mechanism was proposed as early as 1961 by Gurgan Askaryan [38]. In this process, charged particles in the air shower knock electrons from air molecules, which will travel with the particle front, leaving ionized molecules behind. This results in the build-up of a net negatively charged particle front, with a positively charged region behind it, creating a dipole along the shower axis which produces radio emission. Schematic illustrations of both emission mechanisms can be seen in figure 2.9.

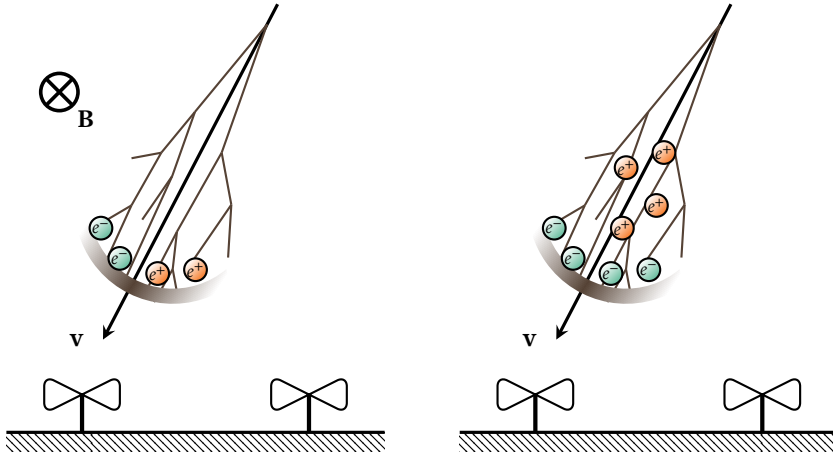


Figure 2.9: Schematic illustration of the two leading mechanisms of radio emission. *Left:* geomagnetic mechanism; *right:* charge-excess mechanism. The total measurable radio emission is a superposition of these two contributions.

Both mechanisms lead to coherent radio emission in the MHz regime, which can be measured as radio pulses which are a superposition of both effects. Measuring radio emission from extensive air showers with multiple antenna stations in coincidence is experimentally challenging, especially in the analogue era before the dawn of fast computers. The geomagnetic emission mechanism, however, was already experimentally verified in the 1960s [39]. It took until the revival in interest in radio detection of air showers in the 2000s, with experiments such as CODALEMA [40], LOPES [41] and finally AERA until the charge-excess emission mechanism was finally demonstrated by experiment [42]. The two emission mechanisms produce a characteristic polarization signature in the measured radiation by which they can be distinguished. This will be further discussed in chapter 3.

2.3.2 Modeling radio emission from air showers

In order to help to interpret the measurements of air shower induced radio pulses, several models describing the radio emission from air showers have been developed. To

simulate the radio pulses measured at the position of an observer, the radiation induced by a huge number of charge particles in the air shower needs to be considered. A macroscopic approach to this problem is to consider general features of the air shower, such as currents and net charges. This approach leads to fast computer simulations, but has the disadvantage that assumptions need to be made, for example about the charge distribution inside the shower. Examples of macroscopic models include MGMR [43] and EVA [44].

A different approach is to microscopically model the radio emission by using full Monte Carlo air shower particle simulations with a certain degree of thinning (usually particles created below a defined energy threshold are grouped into one average particle) as input for radio simulations which calculate the emission along the individual particle tracks. The final radio pulse is then given by the superposition of all emission contributions from the particles. Two examples of simulation packages in which this method is applied are COREAS (based on CORSIKA) [45] and ZHAires (based on AIRES) [46].

These microscopic models directly derive the emission from the electrodynamics of the particles involved, and are therefore independent of assumptions about the emission mechanisms in the shower. In both example models, the vector potential resulting from the charged particles is derived from the Liénard-Wiechert potentials (see e.g. [47]):

$$\Phi(\mathbf{r}, t) = \frac{1}{4\pi\epsilon_0 n^2} \left[\frac{q}{|1 - n\boldsymbol{\beta} \cdot \mathbf{n}|R} \right]_{\text{ret}}, \quad (2.13)$$

$$\mathbf{A}(\mathbf{r}, t) = n^2 \boldsymbol{\beta} \Phi(\mathbf{r}, t)/c, \quad (2.14)$$

where q is the charge of the particle with relativistic velocity $\boldsymbol{\beta} = \mathbf{v}/c$ at distance R from the observer, evaluated in retarded times. The time derivative of the vector potential gives the electric field generated by the particle.

In COREAS, the particle tracks are divided into straight segments, with kinks in which the velocity changes instantaneously. At each track, the particles are accelerated from standstill to their actual speed, and then decelerated again at the end of the track. This is the “end-point formalism” [48], in which the emission resulting from the accelerations and decelerations cancel each other, except for the momentum change of the particle at the boundary of the tracks. From each particle track, the electric field is calculated, and the radiation received at the observer position is the superposition of the fields generated in all particle tracks. This approach is different from the one used in ZHAires, in which the “zhs-formula” [49] is used to calculate the emission. Here, the vector potentials of the tracks are added at the observer position, where one time derivative will produce the total electric field.

Both microscopic models produce similar results [50], but no large-scale systematic comparisons have been performed. The different approximations used in the two mod-

els do produce different results, which can become significant under specific circumstances, for example close to the Cherenkov cone of individual particle tracks [51]. Therefore, comparisons on a case-by-case basis should be made to estimate the uncertainties in the model results (see e.g. [52] for a comparison in the UHF range specifically for the ANITA experiment). We will revisit this point at the end of chapter 3.

2.3.3 Radio emission and air shower parameters

An important attribute of the radio emission of the air shower is that it travels slower than the particle front, which traverses the atmosphere with the speed of light. The velocity of the radio emission is governed by the refractive index of the atmosphere, which is ≈ 1.003 at sea level, and drops to unity at the top of the atmosphere as the density decreases. This means that emission originating from different regions of the shower might arrive at an observer at the same time, resulting in a strongly compressed pulse. The locations where the emission contributions from the shower sum constructively form a ring around the shower axis, located at the *Cherenkov radius* (see e.g. [53]). We will look at this effect in more detail in chapter 3.

The effect of the refractive index will, in combination with the location of the radio emission regions of the shower, result in a hyperbolic radio wave front symmetric around the shower axis [54]. Using a fit of the pulse arrival times on the ground to an assumed radio wave front shape we are able to reconstruct the arrival direction of the air shower. The lateral distribution of the signal strength of the radio pulse is dependent on the location of the emission region in the shower and the air shower energy. By performing a parameterized fit of this distribution, the energy of the primary cosmic ray as well as the shower geometry can be reconstructed [55, 56].

In addition to arrival direction and cosmic ray energy, considerable effort has been made to develop methods to derive composition-sensitive information from the air shower induced radio signals. Most of these methods assume that the location of the radio emission region is related to shower development in general, and therefore sensitive to the composition of the primary cosmic ray. For example, the radio wave front shape can be used to derive where the emission originates [57, 58]. The location of the emission region also has an effect on the lateral distribution of the signal strength of the radio pulse, which can either be fitted with a full Monte Carlo simulation to derive the composition [59], or can be parameterized in a similar way as was done to derive the energy [60]. Finally, the radio pulse shape itself and its degree of compression due to effects of the variable refractive index contains information about the location of the emission region [61], as will be demonstrated in the next chapter.

Method: Composition signatures in the radio pulse shape

THE EXACT SHAPE of an air shower induced radio pulse depends strongly on the shower geometry. More specifically, on the position of an observer relative to the emission region of the shower. This makes the radio pulse shape a parameter which is sensitive to shower development, provided that the remainder of the geometry of the shower is fully determined. In [61], the geometrical dependencies of the pulse shape were scrutinized, and the feasibility of studying shower development using the pulse shape was proven. In this chapter, we will continue this effort by considering a detailed study and parameterization of the geometrical dependencies of the pulse shape and its derived quantity, the spectral index, using a large set of simulated air showers.

3.1 The radio pulse shape

The dependence of the radio pulse shape on the shower geometry can be described by the *path length difference* of radio emission traveling towards an observer from different regions of the air shower, including the propagation of the shower through the atmosphere. A schematic drawing of this effect can be seen in figure 3.1. A shower which penetrates more deeply into the atmosphere will have its emission region closer to the observer. If a shower develops lower in the atmosphere (figure 3.1, left), the difference between the path from the start of the emission region towards the observer and the path following the shower axis until the end of the emission region and then traveling towards the observer will be larger than for a shower developing higher up in the atmosphere (figure 3.1, right). Because of this larger path length difference, the radiation emitted in the shower will be smeared out over a longer period of time when recorded at the observer position, and thus result in a wider pulse.

Not just the interaction depth of the shower determines the duration of the recorded radio pulse. Any aspect of the shower geometry with respect to the observer position influences the pulse length, such as the distance to the shower axis and the zenith angle

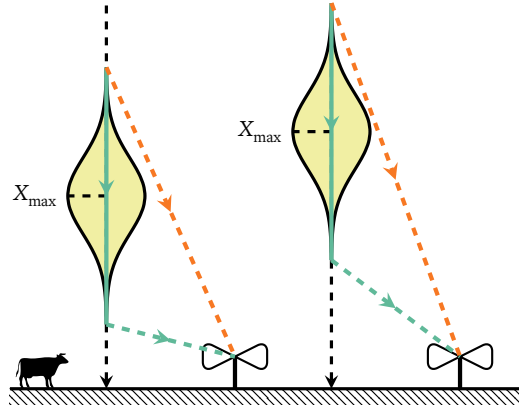


Figure 3.1: The change in *path length difference* between the orange and blue paths for two differently developing showers. The shower on the left has a larger path length difference, which will result in a longer pulse recorded at the observer position.

of the shower. In figure 3.2(a), an air shower simulated with COREAS was sampled at different distances from the shower axis. As we move away from the shower axis, the path length difference increases and therefore also the recorded pulse becomes longer.

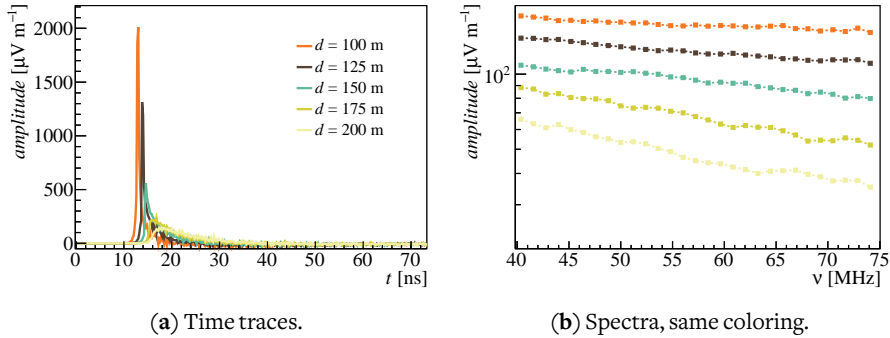


Figure 3.2: Truncated pulses and spectra resulting from a simulated proton shower of 0.47 EeV sampled at 5000 Msps at different distances d from the shower axis using COREAS.

The full pulse shape can be probed through the frequency spectrum of the pulse. In figure 3.2(b) the frequency spectra corresponding to the pulses shown in figure 3.2(a) are drawn. A short pulse results in a flat spectrum, while in a longer pulse the lower frequencies are relatively more important. The longer the pulse, the stronger the decrease in amplitude as a function of increasing frequency in the spectrum.

In addition to the size of the emission region and its location relative to the observer, the pulse length at the observer position is determined by the propagation velocity of the radiation through the atmosphere. As mentioned in section 2.3.3, the radiation propagates with the local speed of light in the atmosphere. The propagation velocity is c/n , with n the index of refraction, which is a function of atmospheric depth. Predicting the exact pulse length is therefore not trivial.

We can illustrate some of the effects that result from the variable index of refraction of the atmosphere with a simple one-dimensional toy model. This model assumes a vertical air shower which emits radiation along a line of arbitrary length L around shower maximum X_{\max} , as can be seen in figure 3.3(a). We can propagate signals from the start and end of this emission region to an observer, using a model of the atmosphere to calculate the index of refraction. In this case the US standard atmosphere [26, 62] is used. For an observer at distance d from the shower axis, the pulse length is given by the difference in arrival times $|t_2 - t_1|$ of radiation originating from points x_1 and x_2 on the shower axis, separated by emission length L :

$$|t_2 - t_1| = \frac{1}{c} |d_1 \bar{n}_1 - (d_2 \bar{n}_2 + L)|, \quad (3.1)$$

where $\bar{n}_{1,2}$ is the average refractive index of the atmosphere along the path $d_{1,2}$ from $x_{1,2}$ to the observer.

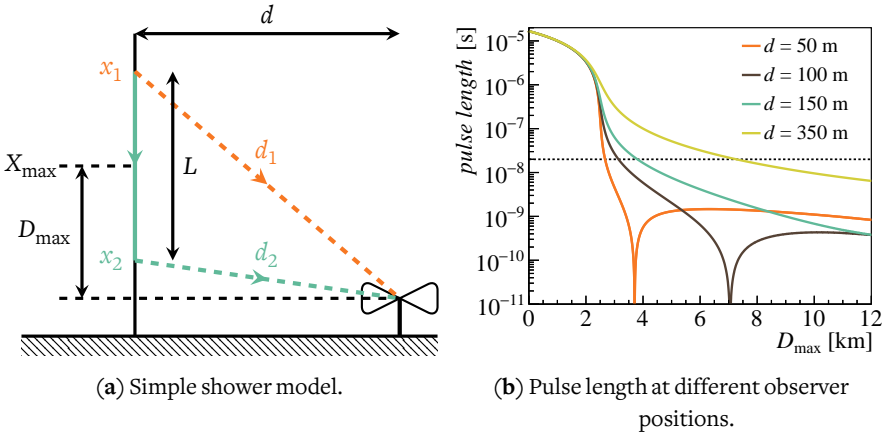


Figure 3.3: Simple one-dimensional model to estimate the pulse length for vertical showers with an emission region around X_{\max} . The distance D_{\max} is the projection onto the shower axis of the distance from X_{\max} to an observer positioned at distance d from the shower axis. The dashed line in figure 3.3(b) indicates the sensitivity region to temporal features of the AERA detector stations at 20 ns (see section 3.3.4).

For an emission length L of 5 km, we have calculated the pulse length as a function of D_{\max} , the distance from the observer to X_{\max} projected on the shower axis, for several different observer positions relative to the shower axis. The pulse lengths at four of these observer positions are plotted in figure 3.3(b). It is obvious that showers which interact close to the observer position produce longer pulses. The most striking features of these plots are the singularities for distances close to the shower axis. These are points on the Cherenkov cone, where the radiation from the start of the emission region arrives at the same instance of time as the radiation from the end of the emission region.

Although the one-dimensional emission along a line is a much too simplistic view of the air shower, this model will be helpful to explain some of the features we will see later on in the simulation study. To get a more realistic representation of the radio pulse shape resulting from a cosmic ray induced air shower, we will use a set of COREAS simulations to derive and geometrical dependencies of the pulse shape. A parameterization of these dependencies will enable us to make statements about shower development using measurements of just one radio station.

3.2 The spectral index

We will probe the radio pulse shape by means of the decrease in amplitude as a function of frequency in the spectrum of the pulse. In [63], it has been shown that the spectrum in our frequency band can be described by an exponential function. We will therefore follow the same approach as in [61], and fit the following exponential function to the data to approximate the general shape of the frequency spectrum $S(\nu)$:

$$S(\nu) = A \cdot 10^{b(\nu-\nu_0)}, \quad (3.2)$$

where b , the *spectral index*, is sensitive to the slope of the spectrum, and A is a scale parameter. A constant frequency offset ν_0 is added to make A proportional to the signal power in our frequency band, and its value is set at 60.0 MHz.

As discussed in chapter 2, the electric field recorded by one antenna station is a superposition of emissions produced by the geomagnetic mechanism and the charge-excess mechanism. They stem from different regions in the air shower [63], and therefore should be treated separately in a parameterization. In order to decompose the signal into these contributions it is insightful to move our frame of reference to the shower plane in such a way that the axes align with the \mathbf{v} , $\mathbf{v} \times \mathbf{B}$ and $\mathbf{v} \times (\mathbf{v} \times \mathbf{B})$ directions, where \mathbf{v} is the direction of movement of the shower and \mathbf{B} the geomagnetic field. The geomagnetic component will always be pointed in the $-\mathbf{v} \times \mathbf{B}$ direction, while the charge-excess component is pointed inward towards the shower core. Fields not perpendicular to the direction of motion of the air shower \mathbf{v} will not be considered. The addition of the two components forms the total field measured by the antenna, as is illustrated in figure 3.4.

We will define the *observer angle* Φ_{obs} as the position of the antenna relative to the $\mathbf{v} \times \mathbf{B}$ -axis in the $(\mathbf{v} \times \mathbf{B}, \mathbf{v} \times (\mathbf{v} \times \mathbf{B}))$ -plane.

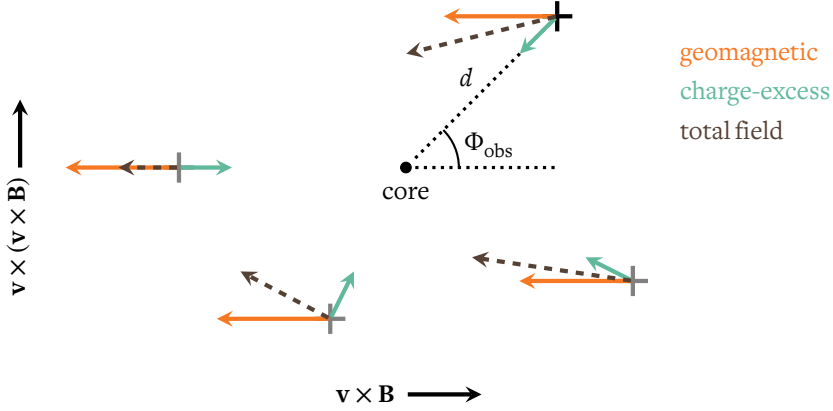


Figure 3.4: Decomposition of the signals into the different emission contributions to the total electric field at several radio stations, denoted by crosses, in the shower reference frame. The direction of movement of the shower \mathbf{v} is perpendicular to the figure plane and pointing inwards.

If we neglect the phase component of the frequency spectrum, we can approximate the shape of a total spectrum $S_T(\nu)$ with mixed contributions of the geomagnetic emission $S_G(\nu)$ and the charge-excess emission $S_C(\nu)$ as:

$$\begin{aligned} S_T(\nu) &= S_G(\nu) + f(\Phi_{\text{obs}})S_C(\nu) \\ A_T \cdot 10^{b_T(\nu-\nu_0)} &= A_G \cdot 10^{b_G(\nu-\nu_0)} + f(\Phi_{\text{obs}})A_C \cdot 10^{b_C(\nu-\nu_0)}, \end{aligned} \quad (3.3)$$

where $f(\Phi_{\text{obs}})$ describes the fraction of charge-excess emission as function of observer angle. If we consider the electric field along the $-\mathbf{v} \times \mathbf{B}$ direction, this function is given by:

$$f_{-\mathbf{v} \times \mathbf{B}}(\Phi_{\text{obs}}) = \cos \Phi_{\text{obs}}. \quad (3.4)$$

Therefore, there is maximum constructive interference at $\Phi_{\text{obs}} = 0^\circ$ and maximum destructive interference at $\Phi_{\text{obs}} = 180^\circ$. Since the exponential functions of the form of equation (3.2) are not additive, equation (3.3) is generally not valid. However, we can approximate it assuming that it is valid at the lower limit ν_- and the upper limit ν_+ of the frequency band we consider. In this case, the set of equations we solve is:

$$\begin{aligned} A_T \cdot 10^{b_T(\nu_- - \nu_0)} &= A_G \cdot 10^{b_G(\nu_- - \nu_0)} + f(\Phi_{\text{obs}})A_C \cdot 10^{b_C(\nu_- - \nu_0)} \\ A_T \cdot 10^{b_T(\nu_+ - \nu_0)} &= A_G \cdot 10^{b_G(\nu_+ - \nu_0)} + f(\Phi_{\text{obs}})A_C \cdot 10^{b_C(\nu_+ - \nu_0)}, \end{aligned} \quad (3.5)$$

from which follows that the spectral index b_T of the pulse with mixed contributions

can be approximated by:

$$b_T = \frac{1}{v_+ - v_-} \log_{10} \left[\frac{10^{b_G(v_+ - v_0)} + f(\Phi_{\text{obs}})R \cdot 10^{b_C(v_+ - v_0)}}{10^{b_G(v_- - v_0)} + f(\Phi_{\text{obs}})R \cdot 10^{b_C(v_- - v_0)}} \right], \quad (3.6)$$

where $R = A_C/A_G$ is the ratio between the scale parameters belonging to both emission mechanisms. We assume that, even though mathematically inexact, the functional description obtained is a good approximation of the true spectral behavior.

3.3 Simulation data set

3.3.1 Simulation setup

As part of an earlier study into the shape of the lateral signal power distribution in relation to the cosmic ray energy [56, 64], an extensive library of shower simulations was created in which the radio emission was analyzed using a star-shaped antenna alignment in the shower reference frame. These simulations were performed using CORSIKA 7.400 with FLUKA 2011.2b [65] and QGSJETII.04 95 [66], including the COREAS plug-in to generate the radio emission. The star shape allows us to disentangle the contributions of the different emission mechanisms and study the two-dimensional properties of the signal. This setup is therefore also well suited to study the geometric dependencies of the spectral index of the radio emission frequency spectrum.

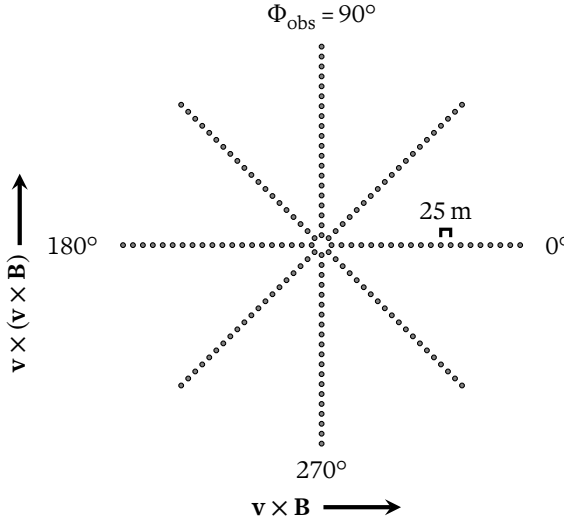


Figure 3.5: Star-shaped antenna arrangement in the shower reference frame which is used to sample the radio signals in the simulations. The distance between the antennas along the principal axes is 25 m.

The star-shaped antenna grid is centered around the shower axis and the antenna positions are projected in the shower plane along 4 principal axes on which the antennas are placed using a 25 m spacing, to a maximum distance of 500 m. An overview of the setup is drawn in figure 3.5. One axis aligns with the $\mathbf{v} \times \mathbf{B}$ direction and one with the $\mathbf{v} \times (\mathbf{v} \times \mathbf{B})$ direction. This way, lines of antennas are located where maximum constructive interference between the two emission mechanisms happens (where $\Phi_{\text{obs}} = 0^\circ$), and maximum destructive interference occurs (at $\Phi_{\text{obs}} = 180^\circ$). At the lines along $\Phi_{\text{obs}} = 90^\circ$ and $\Phi_{\text{obs}} = 270^\circ$ the two emission contributions can be completely separated.

3.3.2 Initial conditions

We will use two sets of proton simulations: one set with initial conditions closely representing the distribution of arrival directions and energies measured by AERA, and one with a set of discrete initial conditions. The non-discrete simulation set has an energy range between 1×10^{17} eV (the lower energy limit of the SD inflight array [67], see also section 4.1) and 1×10^{20} eV, and covers zenith angles θ up to 60° and the full range of azimuth ϕ . A total of 149 unique initial conditions were used, and these were simulated up to 5 times. For every condition, these produced up to 5 different showers with different values for X_{max} . For a variety of reasons, 122 simulations have failed to complete, and were not rerun because a significant part of this set consists of earlier simulations from the lateral signal power distribution study (see section 2.3.3). A total of 623 simulated showers with non-discrete initial conditions is available, and the distributions of the initial conditions are displayed in figure 3.6.

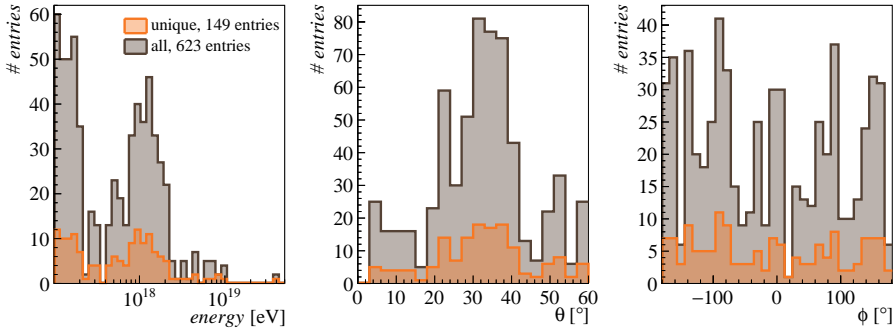


Figure 3.6: Input parameters of the non-discrete set of simulations in terms of energy, zenith angle θ and azimuth ϕ . The same unique parameters were repeated up to 5 times.

For the discrete initial conditions, air showers with azimuth $\phi = \{0^\circ, 45^\circ, 90^\circ, \dots, 315^\circ\}$ and zenith angle $\theta = \{30^\circ, 40^\circ, 50^\circ, 60^\circ\}$ were simulated at two different shower energies:

5×10^{17} eV and 1×10^{18} eV. Similar as before, these were run up to 5 times. In total 4 failed to complete. Therefore, 316 showers were added to the set, bringing the total number of simulated showers available for this analysis to 939. The distribution of all values of X_{\max} of these showers is displayed in figure 3.7.

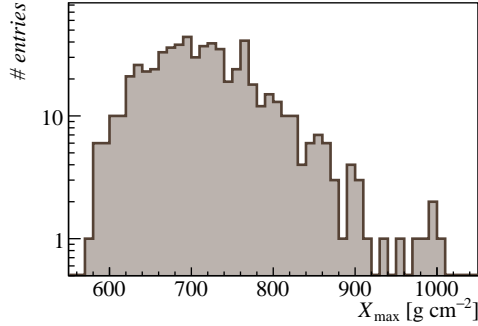
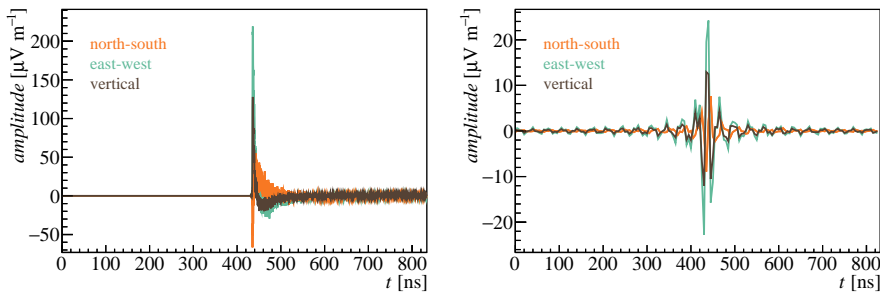


Figure 3.7: Distribution of depth of shower maximum X_{\max} of all 939 showers in the simulation data set.

3.3.3 Post-processing

The time-dependent electric field of the radio emission produced in the simulated showers is recorded at the antenna positions with a sampling rate of 5000 Msps in all three dimensions. To better represent the data recorded by the AERA antenna stations, the time traces will be bandpassed between 30.0 MHz and 80.0 MHz and downsampled to 200 Msps. The effects of these two post-processing steps are shown in figure 3.8.



(a) Original time trace at 5000 Msps.

(b) Bandpassed and downsampled time trace.

Figure 3.8: Processing of traces by bandpassing between 30.0 MHz and 80.0 MHz and downsampling to 200 Msps, before rotation to the shower reference frame. The traces were recorded at a distance $d = 175$ m at $\Phi_{\text{obs}} = 90^\circ$ of a 0.27 EeV shower. Filter delays were not taken into account.

As a final processing step we will convert the time traces from the standard CORSIKA coordinate system, which is reminiscent of the actual antenna alignment of AERA – parallel to the ground and one axis parallel to the geomagnetic field (see section 4.3.2) – to a rotated coordinate system along the \mathbf{v} , the $\mathbf{v} \times \mathbf{B}$ and the $\mathbf{v} \times (\mathbf{v} \times \mathbf{B})$ vectors. This way the traces are also aligned in the shower reference frame, enabling us to separate the different emission contributions. Examples of the Fourier transforms of the rotated traces are visible in figure 3.9. We will fit equation (3.2) only to the traces parallel to the $\mathbf{v} \times \mathbf{B}$ and $\mathbf{v} \times (\mathbf{v} \times \mathbf{B})$ directions, because along the \mathbf{v} direction the field is negligibly small, except for the close vicinity of the shower axis. To make the results compatible with the limitations posed by the narrowband sources at the AERA site (see section 6.1.3) we will fit the obtained spectra between 40.0 MHz and 75.0 MHz.

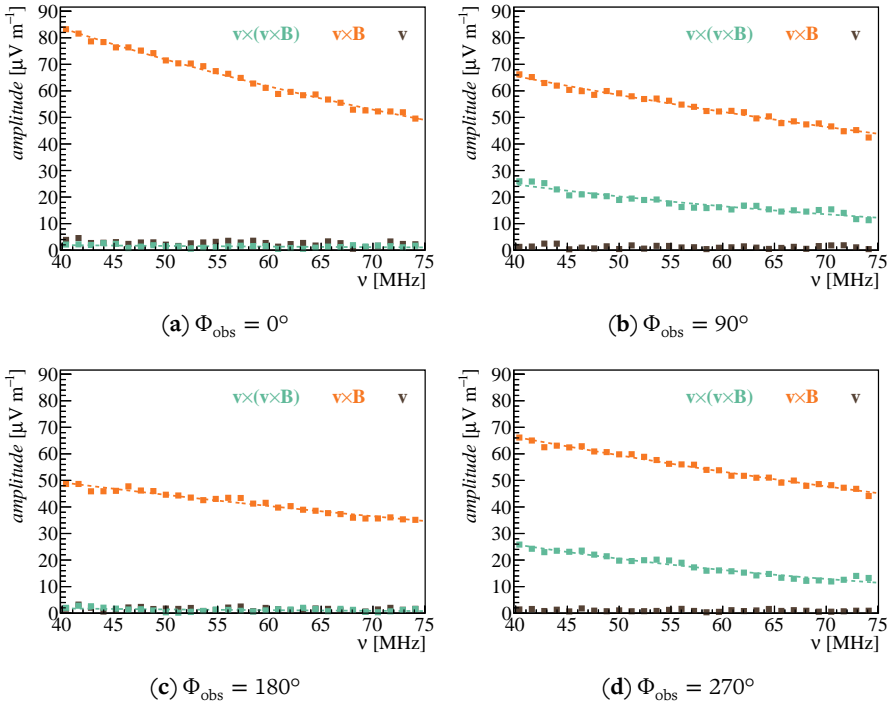


Figure 3.9: Spectra of the same simulated shower as in figure 3.8 after rotation to the shower reference frame, recorded at $d = 175$ m. The $\mathbf{v} \times (\mathbf{v} \times \mathbf{B})$ contribution vanishes at $\Phi_{\text{obs}} = 0^\circ$ and $\Phi_{\text{obs}} = 180^\circ$ as expected, while the two contributions are completely separated at $\Phi_{\text{obs}} = 90^\circ$ and $\Phi_{\text{obs}} = 270^\circ$. At $\Phi_{\text{obs}} = 0^\circ$, there is constructive interference between the geomagnetic and charge-excess emission contributions, while at $\Phi_{\text{obs}} = 180^\circ$ there is destructive interference. The spectra are fitted with equation (3.2).

3.3.4 Limitations and coherence

Since we have limited the spectra to the range between 40.0 MHz and 75.0 MHz, we are most sensitive to temporal features in the pulse of approximately 20 ns. We have marked this duration with a dashed line in figure 3.3(b). If the temporal features become much longer or much shorter than this duration, our sensitivity is diminished. Even though the numerical values of the pulse length in our one-dimensional toy model should be treated with caution, we can derive from figure 3.3(b) that showers with a deep shower maximum will have pulses which are too wide to be properly measured by antenna stations far away from the shower axis. In other words: if D_{\max} is small and the distance to the shower axis d is large, most of the signal power will be in frequencies below our bandwidth.

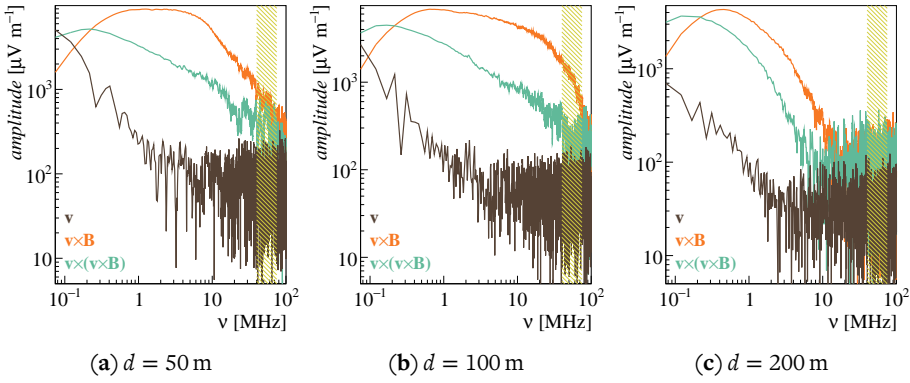


Figure 3.10: Unfiltered spectra of a 0.47 EeV vertical proton shower at $\Phi_{\text{obs}} = 90^\circ$, with the measurement bandwidth of AERA (see chapter 4) marked by the hatched area.

Effects that play a role at high frequencies are incoherence of the radio signal and unphysical effects caused by the thinning level of the simulations [46]. This is illustrated in figure 3.10, in which the unfiltered spectra at three distances from the shower axis d are plotted. Our measurement bandwidth is marked by the hatched area. It is clear that as we move away from the shower axis, the defining features of the spectrum move out of the measurable band and the incoherent and “noisy” regime will become more important. Fits of the spectrum in this incoherent regime are unreliable, and the parameter space in which it occurs has to be rejected from the analysis. Because of the complex interplay of all of the aforementioned effects we will *a posteriori* determine an incoherence cutoff as a function of d and D_{\max} from the fitted spectra of the simulated showers.

3.4 Parameterization of the spectral index

First, we will consider the spectral indices of both radio emission contributions separately. After parameterizing their main dependencies, we will attempt to combine both contributions using equation (3.6).

3.4.1 Separate radio emission contributions

At the simulated antenna positions with observer angles of 90° and 270° the geomagnetic and charge-excess components are exactly perpendicular, and thus can be separated. The average spectral indices of the two contributions measured at these antenna positions are plotted in figure 3.11, out to a geometrical distance to shower maximum D_{\max} of 14 km.

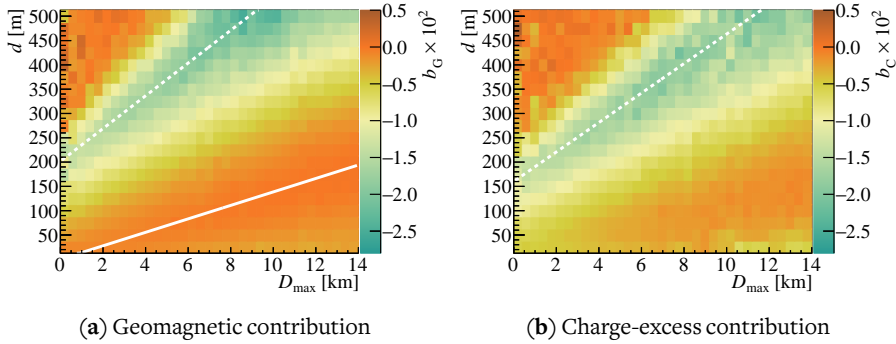


Figure 3.11: Average spectral index of the separate contributions for all distances to the shower axis d as a function of the distance from the antenna to shower maximum projected on the shower axis, D_{\max} . The dashed white line marks the cutoff beyond which incoherence is dominant, the solid white line marks the Cherenkov radius where the spectral index approaches zero.

When examining the geomagnetic contribution in figure 3.11(a), we see that the spectral index approaches zero at distances d close to the shower axis. This is the location of the Cherenkov cone, where the radio pulse becomes very short and produces an almost flat spectrum. Inside this cone, the spectral index as a function of D_{\max} is more or less flat, as was also implied by the pulse lengths of the simple model in figure 3.3(b). As this implies a lack of sensitivity of our measurement to the shower development, we will reject this region from the analysis.

We can determine the maxima of the spectral index distribution as a function of d and D_{\max} and fit these with a linear function to estimate the size of the Cherenkov cone. The resulting line is plotted as a solid white line in figure 3.11(a) and the fit values are listed in table 3.1. Interestingly, this feature is not as pronounced in the charge-excess compo-

nent, although a flattening of the D_{\max} -dependence can be observed at small distances to the shower axis. This likely points towards a difference in height of the emission regions of both contributions, with the geomagnetic emission region more or less centered around X_{\max} , and the charge-excess building up and continuously emitting along the shower axis [53].

Table 3.1: Linear functions best fitting the coherence cutoff and Cherenkov cone:

$$D_{\max} \times 10^{-3} = a_0 + a_1 d.$$

	<i>fit</i>	a_0 [km]	a_1 ($\times 10^{-2}$)
Cherenkov cone		0.0	7.21
Geomagnetic coherence cutoff		-5.99	3.00
Charge-excess coherence cutoff		-5.28	3.20

As we move away from the shower axis, the spectral index decreases until a certain minimum is reached, after which the spectral index increases again. This minimum is the cutoff between coherent and incoherent regions of the shower, and the increase in the spectral index is caused by flat “noise” spectra which start to dominate the underlying distributions, as was demonstrated in figure 3.10. In similar fashion to the Cherenkov cone, we determine the minima of the spectral index of both contributions, and fit them with a linear function to estimate the cutoff. The cutoffs of geomagnetic and charge-excess emission are plotted as dashed lines in figure 3.11, and the fit values are listed in table 3.1. The difference between the cutoffs of the two emission mechanisms is also likely to be caused by differences in the emission height.

In the remainder of this analysis, we will only consider the coherent region outside of the Cherenkov cone, limiting ourselves mostly to distances to the shower axis between 100 m and 350 m. Coincidentally, this corresponds to the typical spacing of AERA antenna stations (see section 4.3). We will now parameterize the spectral index as a function of D_{\max} for each ring of antennas in the star-shaped pattern with distance d to the shower axis. Within the region of interest defined by values in table 3.1, we fit the following function with scale parameter β and exponent γ :

$$b \times 10^2 = \frac{\beta}{1 + \exp(-\gamma \cdot D_{\max}/1\text{km})} - \beta. \quad (3.7)$$

In figure 3.12, the spectral index of both emission contributions as function of D_{\max} is plotted for two distances to the shower axis d . These plots include fits of equation (3.7) to the spectral index of each of the emission contributions.

The scale parameter β and exponent γ of the fit are plotted as a function of distance to shower axis d in figure 3.13 for both emission contributions. A strong dependence on d is visible, and we parameterize this dependence by fitting the scale parameter and

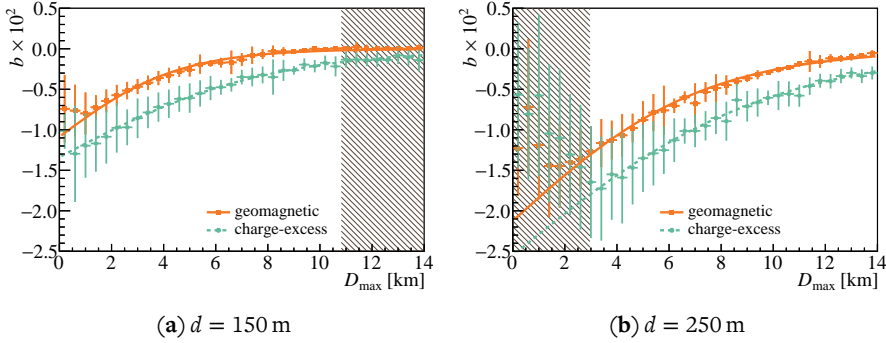


Figure 3.12: Average spectral index as function of D_{\max} at two distances from the shower axis, including a fit of equation (3.7). The hatched areas mark the rejected regions defined in table 3.1. The error bars indicate one standard deviation of the average values.

exponent, which is also drawn in figure 3.13. The dependence on b of the scale parameter β has been fitted with polynomials of the third degree. The results of this fit for each of the emission contributions are listed in table 3.2. A product of a linear function and an exponential decay function has been fitted to the exponent γ . The results are given in table 3.3.

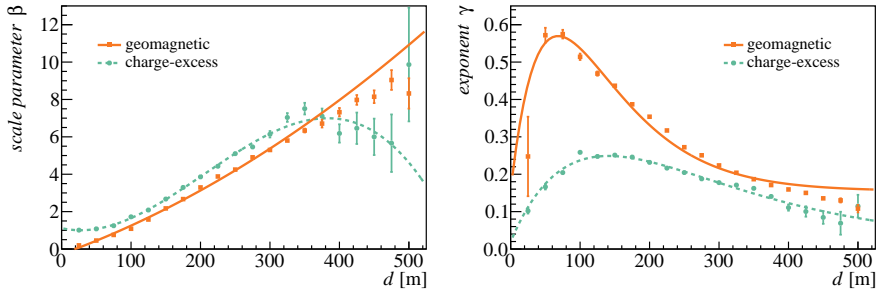


Figure 3.13: Fit parameters of equation (3.7) as function of distance to shower axis d , which themselves have been fitted by the functions described in tables 3.2 and 3.3.

With these fits, we have created a full parameterization of the spectral index as a function of d and D_{\max} of the pure geomagnetic and pure charge-excess emission contributions. It should be noted that the fit-functions are ad-hoc descriptions, and are not inspired by physics considerations.

Table 3.2: Polynomial fits of form $b_0 + b_1d + b_2d^2 + b_3d^3$ to scale parameter β for both emission contributions.

<i>contribution</i>	b_0	b_1	b_2	b_3
geomagnetic	-2.73×10^{-1}	1.34×10^{-2}	1.79×10^{-5}	0.0
charge-excess	1.12	-8.37×10^{-3}	1.64×10^{-4}	-2.66×10^{-7}

^a parameter fixed

Table 3.3: Fits of form $g_0 + g_1x \cdot \exp(-x/g_2)$ to exponent γ for both emission contributions.

<i>contribution</i>	g_0	g_1	g_2
geomagnetic	1.55×10^{-1}	1.65×10^{-2}	68.3
charge-excess	1.91×10^{-2}	4.43×10^{-3}	141

3.4.2 Mixed radio emission contributions

In order to calculate the spectral index at observer angles where we have a mixing between the two emission mechanisms, we need to parameterize the ratio of scale factors of the pure geomagnetic and charge-excess contributions $R = A_C/A_G$ of the fit to the spectrum, as was defined in equation (3.6). The fraction of charge-excess emission versus geomagnetic emission can be expressed by the charge-excess fraction a [68], defined as:

$$a \equiv \sin \alpha \frac{|E_C|}{|E_G|}, \quad (3.8)$$

where E_C is the charge-excess contribution to the total electric field and E_G is the geomagnetic contribution. The geomagnetic angle α is the angle between the shower propagation direction \mathbf{v} and the geomagnetic field \mathbf{B} . In [68], it was empirically established that a is not constant but depends on the zenith angle of the air shower, which is strongly correlated to D_{\max} because the amount of traversed atmospheric depth increases for more inclined showers. Because A_C and A_G are related to the total electric field of the two emission contributions, we can use the relation of equation (3.8) to derive the dependencies of the ratio R . The charge-excess fraction $a \approx R \cdot \sin \alpha$ is plotted as a function of D_{\max} in figure 3.14. We can parameterize the dependence on D_{\max} with a polynomial of the third degree, which is also drawn in figure 3.14. The fit parameters of are listed in table 3.4.

Now we have all ingredients to fully parameterize the spectral index of pulses in the $\mathbf{v} \times \mathbf{B}$ direction as function of d , D_{\max} and Φ_{obs} using equation (3.6) with $\nu_- = 40$ MHz and $\nu_+ = 75$ MHz. In figure 3.15, the parameterized spectral index for the most “extreme”

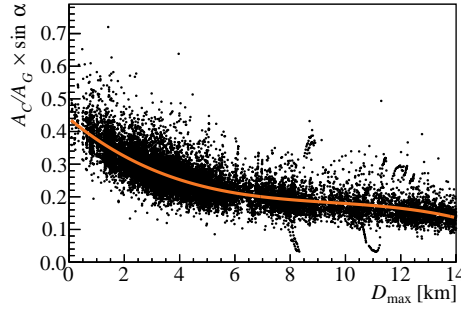


Figure 3.14: The charge-excess fraction $\alpha \approx R \cdot \sin \alpha$ with ratio $R = A_C/A_G$ as a function of D_{\max} , calculated for all observer positions where the two emission contributions are perpendicular, for all simulations. The fit is a polynomial of the third degree, the parameters of which are listed in table 3.4.

Table 3.4: Fit values describing the ratio $R = A_C/A_G$ using a polynomial fit of form $R \cdot \sin \alpha = r_0 + r_1 D_{\max} + r_2 D_{\max}^2 + r_3 D_{\max}^3$.

r_0	r_1	r_2	r_3
0.439	-7.01×10^{-2}	6.78×10^{-3}	-2.38×10^{-4}

observer angles is compared with the profiles extracted from the simulation set for two different distances to the shower axis. We have used $\alpha = 90^\circ$ in the parameterization to generate the largest amount of mixing between geomagnetic and charge-excess emission.

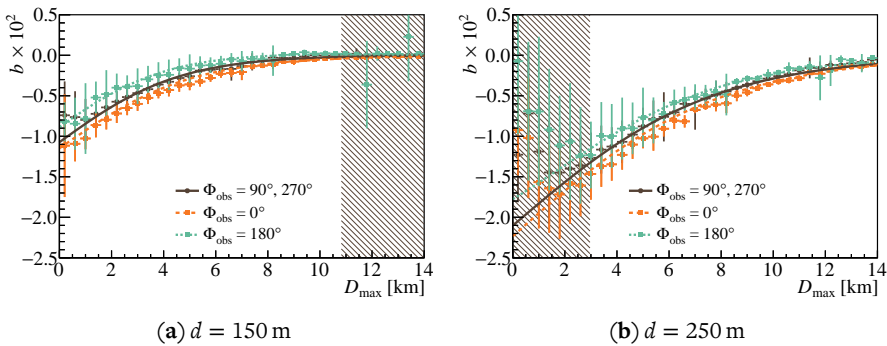


Figure 3.15: Parameterization of the spectral index of observer angles with mixed contributions from geomagnetic and charge-excess emission. The hatched areas mark the rejected regions defined in table 3.1. The error bars indicate one standard deviation of the average values.

3.5 Parameterization performance

Given a shower arrival direction, core impact location, antenna position and a spectral index we can now determine D_{\max} , and propagate back through an atmospheric model to calculate X_{\max} . We compare the actual values of X_{\max} of the Monte Carlo simulation with the ones determined from the parameterization. In COREAS, the US standard atmosphere is used, and we also use this atmospheric model to convert D_{\max} into an atmospheric depth. The distributions of the difference between the true value of X_{\max} as produced by the Monte Carlo and the reconstructed value of X_{\max} calculated with the parameterization and propagation through the atmospheric model are visible in figure 3.16.

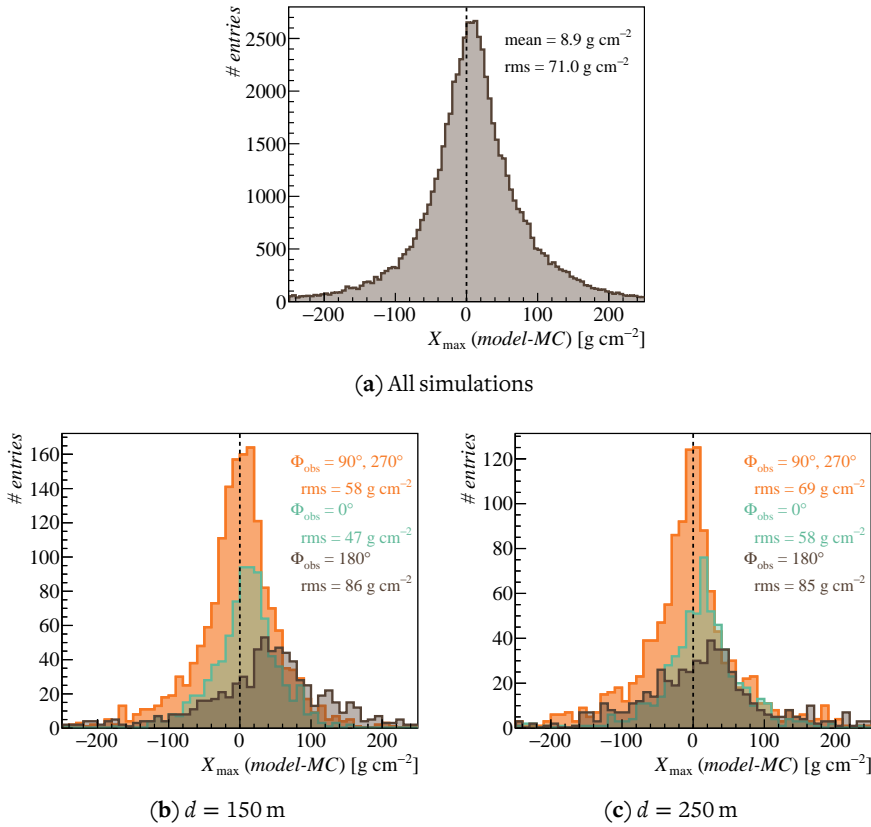


Figure 3.16: Difference between the true Monte Carlo (MC) value of X_{\max} and the reconstructed value using the parameterization ($model$) for all simulations (a) and two specific distances from the shower axis and different observer angles Φ_{obs} (b),(c).

Only simulations with spectral indices for which a fit of the parameterization converges have been used to generate these distributions, which was the case in 65 631 of the 69 409 simulated antenna stations which were inside the validity region defined in table 3.1. A complete overview of the average and rms of the difference between the true and reconstructed values of X_{\max} , for every observer angle in the subset of distances to the shower axis we are interested in, is visible in figure 3.17. For distances closer than 75 m to the shower axis and further than 400 m the method becomes unstable because of the small number of usable bins that are not affected by either incoherence or the Cherenkov cone, and therefore to which equation (3.7) could be reliably fitted.

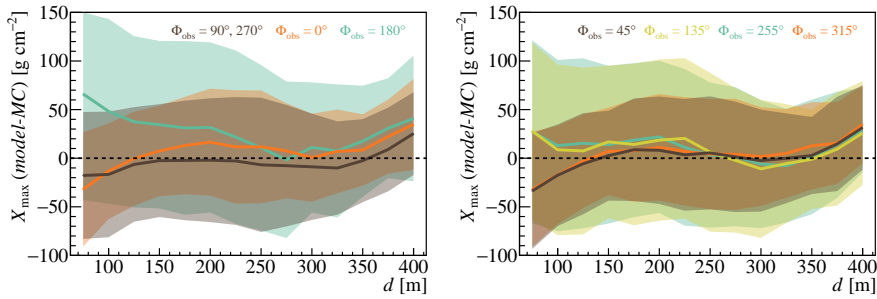


Figure 3.17: Average difference between the true Monte Carlo (MC) value of X_{\max} and the reconstructed value (*model*) from the parameterization for all simulations. The colored bands indicate the rms of the distribution.

3.6 Discussion

We have shown in this chapter that a measurement of the radio pulse and its spectral index can yield composition-sensitive information. Using the derived parameterization, we are able to determine X_{\max} with information from one radio station down to a theoretical sensitivity of about 60 g cm^{-2} in our range of interest between 100 m and 350 m if we consider a signal of purely geomagnetic origin. For observer angles where mixing between the emission contributions occurs the sensitivity is similar, except at 180° , where uncertainties are much larger. This is due to the lower signal at this angle, which results in a higher fit uncertainty. Biases of up to 30 g cm^{-2} are also visible in figure 3.17. They are also largest for the observer angles with destructive interference between the geomagnetic and charge-excess emission contributions.

The reliability and uncertainty of the parameterization based on COREAS simulations are tested by comparing it with a shower simulated with ZHAIRES. For this, we simulate a vertical $1 \times 10^{17} \text{ eV}$ proton shower with both packages. These showers are analyzed with a higher density of radio stations on the arms of the star pattern, using a separation of 5 m instead of 25 m. A comparison between the spectral index b as a function

of distance to the shower axis d and observer angle as obtained by the two models is shown in figure 3.18. The parameterized spectral indices at the extreme observer angles are also plotted in this figure. The simulations are analyzed at different elevations, namely at 1564 m in the COREAS simulation and at 1400 m in the ZHAires simulation. Therefore, even though they have a different value for X_{\max} (663.2 g cm^{-2} for COREAS and 682.1 g cm^{-2} for ZHAires), they have a similar geometrical distance to shower maximum D_{\max} (2.07 km for COREAS and 2.01 km for ZHAires). Both simulation packages use the same parameterization of the US standard atmosphere, but use a slightly different refractive index at sea level. The geomagnetic field direction has a 1.3° difference and COREAS uses a magnetic field which is 5 % stronger than the field in ZHAires.

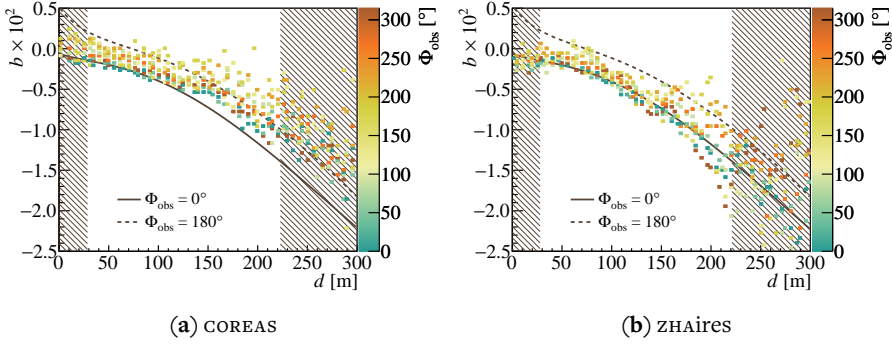


Figure 3.18: Spectral index b as a function of distance to the shower axis d and observer angle Φ_{obs} for a vertical 1×10^{17} eV proton shower simulated by COREAS and ZHAires. The parameterizations of the spectral index at observer angles with maximum constructive and destructive interference between geomagnetic and charge-excess emission are indicated by lines. The hatched regions mark the end of validity of our parameterization.

The same general trend in the spectral index as a function of distance and observer angle in both simulated showers is observed in figure 3.18. Deviations are visible close to the shower axis, where the ZHAires shower produces a much more pronounced Cherenkov effect than the COREAS shower. Further from the axis, the COREAS shower produces flatter spectra than obtained by the parameterization, while the ZHAires shower produces steeper spectra. Incoherence effects appear at relatively small distances to the shower axis in both simulations because of the low energy of the shower.

A more detailed comparison is shown in figure 3.19, where average profile histograms of the purely geomagnetic and charge-excess components calculated at observer angles of 90° and 270° are plotted. It is striking that for these particular two simulated showers, the ZHAires shower seems to follow the parameterization better than the COREAS shower. From this, we conclude that the deviations between the simulation packages

are comparable to the fluctuations between different showers produced by COREAS, and based on this comparison there is no reason to assume additional systematic uncertainties in the simulations due to the particular package that is being used. The similarity between the packages is remarkable considering the different environmental variables such as magnetic field and refractive index which were used as input for both simulations.

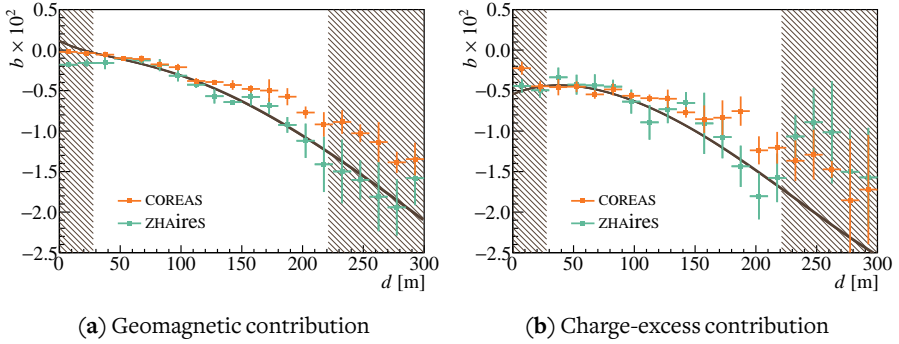


Figure 3.19: Average spectral index of the purely geomagnetic and charge-excess emission contributions in the two simulated showers. The parameterizations of the spectral index of both showers based on their respective distances to shower maximum are displayed as lines, which are almost completely overlapping. The error bars indicate one standard deviation of the average spectral index.

The detector: The Pierre Auger Observatory and AERA

THE PIERRE AUGER OBSERVATORY is located in the province of Mendoza in Argentina. The observatory was designed to study cosmic rays of the highest energies: it is sensitive to cosmic ray energies of $\sim 10^{18}$ eV and above [69]. Because of the low flux of cosmic rays in this energy range, the site covers a surface area of about 3000 km², which enables the measurement of a significant number of cosmic rays in this range. Among other things, the location was selected because of its good weather conditions and the altitude of about 1400 m above sea level, which at these energies is close to shower maximum [70].

More than 1600 water Cherenkov detectors form the *surface detector* (SD), and measure the charged particles in the air shower as they hit the detectors at ground level. The surface detector was designed to have a duty cycle of almost 100 % [71, 72], and its measurements enable us to reconstruct air shower arrival direction and energy. There is also considerable ongoing effort to derive composition sensitive parameters from the SD measurements (see e.g. [73] and [74]). On clear and moonless nights, these measurements are complemented by the *fluorescence detector* (FD), which records the development of air showers. This is used to determine the cosmic ray composition and perform a calorimetric measurement of the air shower from the deposited energy in the atmosphere. The fluorescence detector consists of a total of 27 fluorescence telescopes, distributed over four sites surrounding the array. The locations of these sites can be seen in the overview of figure 4.1. Because it requires darkness and clear atmospheric conditions, the uptime fraction of the FD is about 13 % [75].

Together, the SD and FD constitute a *hybrid detector*, in which the information of both detectors is complementary and is used to reconstruct and cross-calibrate the relevant parameters of the air shower such as the arrival direction and primary energy of the cosmic ray. An example of an event which is measured by both detector systems can be seen in figure 4.2.

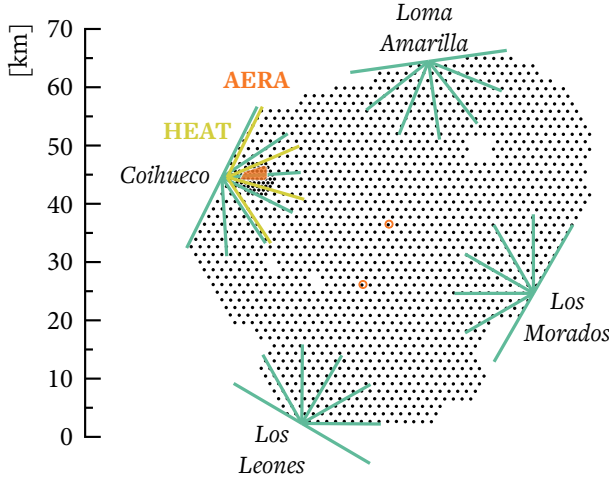


Figure 4.1: Layout of the Pierre Auger Observatory. The black dots indicate the locations of the surface detectors, and the lines demarcate the boundaries of the fields of view of the fluorescence telescopes including HEAT. AERA is located within the orange region. The orange circles mark the locations of the two laser facilities (see section 4.2). North is up.

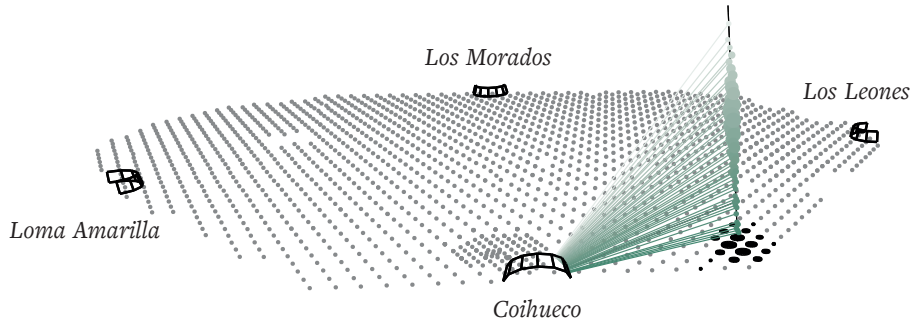


Figure 4.2: Illustration of the hybrid detector: an event is recorded by both the surface detector and the fluorescence detector (adapted from [76]).

4.1 The Surface Detector

The 1600 particle detectors of the regular surface detector array sit on a triangular grid with a spacing of 1500 m. The detectors consist of tanks with a reflective inner liner holding 12 t water [77]. Charged particles from the air shower travel at relativistic speeds, and when they traverse the water in the tank they will produce Cherenkov light. This light is collected by three photomultiplier tubes (PMTs), which look downward into the tank. A schematic image of a surface detector station can be seen in figure 4.3.

Each of the detector stations is self-powered by a solar panel, has a GPS module for timing and communicates with the central data acquisition system (CDAS) using a wireless connection.

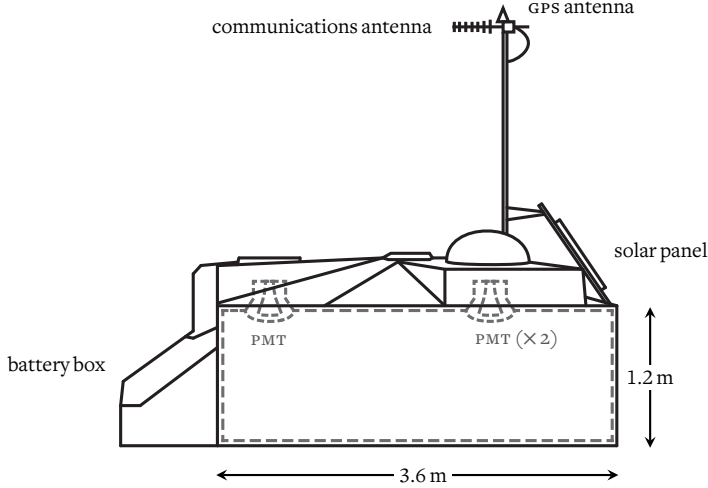


Figure 4.3: Schematic drawing of a surface detector station. Three PMTs look down into a water-filled reflective liner, one of which is directly behind the right-most PMT from this perspective.

The signal strength which is measured by the PMTs depends on the number of charged particles traversing the tank and the length of the tracks along which they emit light (which depends on the inclination of the shower). For convenience, the measured signal voltages are converted into *vertical equivalent muons* (VEMs), which are equivalent to the amount of light that would be measured coming from a muon traveling straight down through the center of the tank. When a surface detector measures a signal of 3VEM in all three PMTs it is triggered, and it sends a message to the CDAS. When at least three detectors have triggered within 25 μ s and satisfy certain spatial conditions, the CDAS retrieves the information from all triggered detectors and an event is created [78].

Air shower geometry (core impact position and shower arrival direction) and cosmic ray primary energy can be reconstructed using a combination of a signal time fit to a curved particle front and fitting a description of the lateral fall-off of the particle distribution to the measured signal strength. This is done in an iterative process, in which the timing fit provides the shower axis, core impact time and the curvature of the particle front, and the signal strength fit provides the core impact location and ultimately the shower energy. The signal strength can be parameterized by a function such as a modified Nishimura-Kamata-Greisen (NKG) lateral distribution function (see e.g. [79]), which describes the number of charged particles in the shower as a function

of distance to the shower axis r , and therefore the signal strength S measured in VEMS by a surface detector station at this location:

$$S(r) = S(r_{\text{opt}}) \cdot \left(\frac{r}{r_{\text{opt}}} \right)^{\beta} \left(\frac{r + r_{\text{scale}}}{r_{\text{opt}} + r_{\text{scale}}} \right)^{\beta+\gamma}. \quad (4.1)$$

Here, r_{opt} is the optimum distance (where signal fluctuations are minimal) and r_{scale} is a scale parameter with a value of 700 m. For the array with a spacing of 1500 m, the optimum distance was found to be 1000 m [80]. $S(1000 \text{ m})$, β and γ are fit parameters, of which $S(1000 \text{ m})$ is proportional to the primary energy and serves as an energy estimator for the surface detector. After a correction for the inclination of the shower, the proportionality of this estimator can be calibrated by the events that are measured simultaneously by the SD and FD (see e.g. [81]).

In addition to the regular surface detector array, 61 detectors have been used to create a denser grid with a 750 m spacing, covering about 24 km^2 in the north-west corner of the array (see figure 4.1) [82]. This is the *SD infill array*, which was designed to be sensitive to cosmic rays of lower energies – down to an energy of $\sim 10^{17} \text{ eV}$ – to assist with the AMIGA extension of the observatory [67]. As part of AMIGA, the surface detectors in this denser grid will be complemented by several buried scintillators to measure the muon-fraction of the shower, which is a composition-sensitive parameter. Because of the smaller detector spacing, r_{opt} was found to be 450 m for the fit of equation (4.1) to the SD infill data [83]. A typical SD infill event including a fit of the lateral distribution function can be seen in figure 4.4.

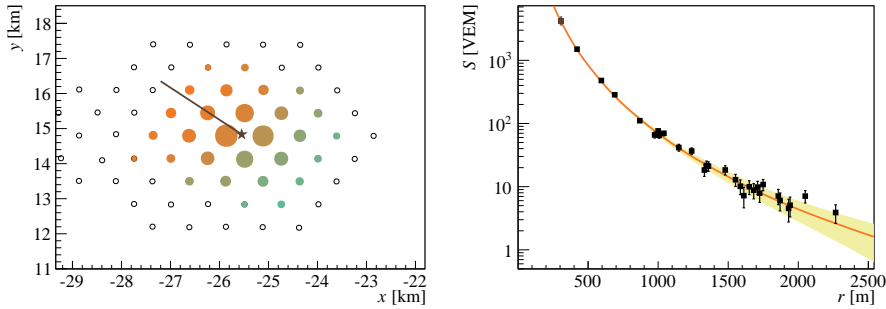


Figure 4.4: Event with SD ID 29894169 as recorded by the SD infill array. *Left:* the geometry of the event on the array, where the size of the circles are proportional to the measured signal strength and the colors indicate signal time. The star marks the core impact position and the line the direction of the shower axis. *Right:* the lateral distribution of the signal strength with a fit of equation (4.1).

4.2 The Fluorescence Detector

The charged particles in extensive air showers excite molecules in the atmosphere, mostly nitrogen and oxygen. When these molecules fall back to their ground state, they isotropically emit fluorescence light in the uv-band. 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 along its path through the atmosphere. By measuring this light the longitudinal shower profile can be determined, including the depth of shower maximum X_{\max} . It also serves as a calorimetric measurement of the energy deposited electromagnetically in the atmosphere, which accounts for $\sim 90\%$ of the primary cosmic ray energy [75].

The emitted fluorescence light is measured by telescopes, which are located at four elevated positions overlooking the array and each have a field of view of $30^\circ \times 30^\circ$. They consist of an aperture system with corrective optics and a uv filter, and a segmented mirror which focuses the light onto a camera composed of an array of PMTs, as can be seen in figure 4.5(a). At each of the four FD sites, six standard FD telescopes are housed in a dedicated building. At the Coihueco site an additional three telescopes are installed which can be tilted upwards by 29° : the *high elevation Auger telescopes* (HEAT) [84]. These telescopes are designed to be sensitive to air showers down to an energy of 10^{17} eV, which is comparable to the lower energy bound of the SD infill array, which it overlooks. Showers with low energy generally interact earlier in the atmosphere, and because they produce fewer air shower particles, their fluorescence yield is also less. Therefore, they develop mostly outside the field of view of the standard FD telescopes, but by pointing the telescopes upward the development of low energy showers can also be probed with high efficiency. A comparison between both telescope types is visible in figure 4.5.

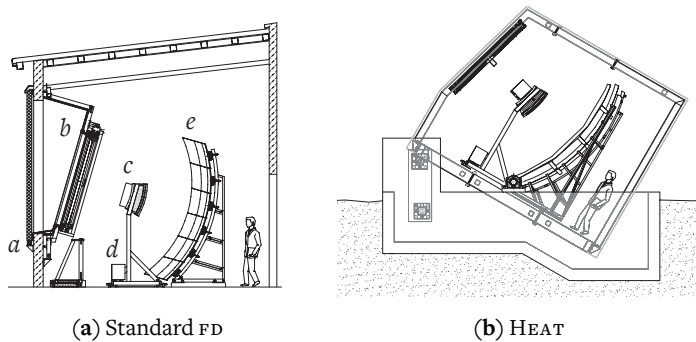


Figure 4.5: Schematic drawing of the standard FD telescope (a) and HEAT in inclined mode (b). *a*: shutter; *b*: aperture system and uv filter; *c*: camera; *d*: electronics; *e*: segmented mirror [75, 85].

The data acquisition of the FD runs completely independent of the data acquisition of the surface detector, except for triggers which are sent to the CDAS by the FD when an air shower candidate is detected. In order to reconstruct the shower profile, at least one SD station must be triggered by the same shower as measured by the FD to constrain the shower axis and core impact time [86]. The signals measured by the telescope pixels as a function of time can then be projected onto the shower axis, and a longitudinal shower profile as a function of atmospheric depth X can be reconstructed, as is visible in the example of figure 4.6. To calculate the energy deposit, the attenuation in the atmosphere needs to be known, as well as the fraction of the measured light that is emitted as Cherenkov photons by the shower particles instead of fluorescence light [87]. The depth of shower maximum X_{max} can be determined by fitting the Gaisser-Hillas function (equation (2.12)) to the longitudinal profile, which is also shown in figure 4.6.

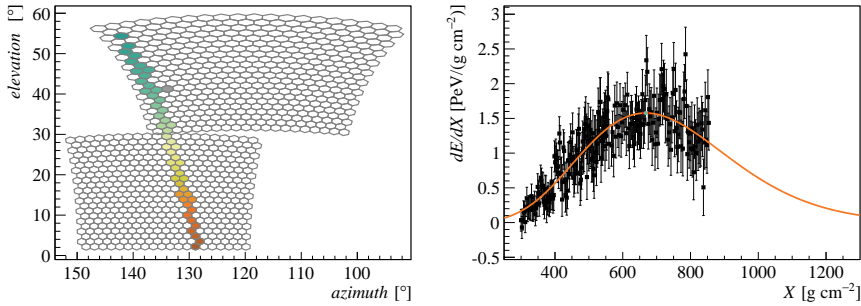


Figure 4.6: An example of an event measured by both Coihueco and HEAT (FD ID 1407_11995). *Left:* the individual pixels of the cameras, where the colors indicate the signal arrival time; *right:* the reconstructed longitudinal shower profile of the same shower, with a fit of the Gaisser-Hillas function.

To calculate the attenuation of the fluorescence light in the atmosphere, molecular scattering and scattering off aerosols needs to be considered. For molecular scattering, atmospheric profiles of temperature, density and humidity are sufficient to calculate the scattering coefficients. Local monthly profiles have been determined in extensive on-site measurement campaigns, but they can also be obtained from tables that are produced on a 3-hour basis by the *Global Data Assimilation System* (GDAS) [88]. Much more volatile than the atmospheric molecular profiles is the aerosol content of the atmosphere above the detector [89]. The aerosol content is monitored by using the FD telescopes to measure hourly laser shots by two laser facilities near the center of the array (see figure 4.1), which can be used to calculate the vertical aerosol optical depth (VAOD) [90]. Furthermore, clouds that may obscure the shower are monitored by several systems [91, 92], and an efficient algorithm rejects lightning events which might trigger the FD telescopes altogether [93].

4.3 The Auger Engineering Radio Array

4.3.1 Introduction

The first experiments intended to measure radio emission in the MHz regime at the Pierre Auger Observatory were RAUGER [94], the BLS-setup and its successor MAXIMA (see e.g. [95]). RAUGER was installed in 2006, and development at the BLS site started in 2008. Both setups successfully recorded air shower induced radio emission in coincidence with the surface detector.

These experiments have been succeeded by the Auger Engineering Radio Array (AERA), at a site which overlaps with the SD infill array and is within the field of view of both the Coihueco and HEAT fluorescence detectors. The first stage of this array, AERA-I, was deployed in September 2010, and consists of 24 radio detector stations on a 144 m grid equipped with log-periodic dipole antennas (LPDAs), which were based on the MAXIMA design [96]. After a few months commissioning, AERA-I has been taking data since March 2011. Its radio detector stations are equipped with either one of two different types of electronics. One type was developed jointly by the Karlsruhe Institute of Technology (KIT) and the Bergische Universität Wuppertal (BUW), and focuses on external triggers provided by the baseline Auger detectors. The other type was developed by the Radboud University (RU) and Nikhef, and focuses on self-triggering on radio pulses. The distribution of electronics over the antenna stations changed multiple times until March 2012, after which 6 antenna stations were equipped with RU/Nikhef electronics and 18 with KIT/BUW electronics, as can be seen in figure 4.7.

An extension to the existing array was deployed as AERA-II in early 2013, and has been taking data since April 2013. AERA-II consists of 95 radio detector stations equipped with a type of bow tie antenna called the *Butterfly*, which was originally developed for the CODALEMA experiment and adapted for AERA [97, 98]. The stations were deployed in a semi-regular grid, with 55 stations equipped with KIT/BUW electronics (mostly with a 250 m spacing), and 40 stations with RU/Nikhef electronics (mostly with a 375 m spacing). Apart from some smaller differences discussed later in this chapter, the major difference between the two types of electronics is still the focus on different trigger mechanisms. A full overview of the layout of the antenna array can be seen in figure 4.7. Five stations in the south-east corner of the array were used to test an experimental antenna type which also measures the vertical electric field component in the period from 7 November 2013 to 22 November 2014. In the remainder of this thesis, only data from the AERA-II radio detector (RD) will be used, so we will limit our discussion of the detector in the rest of this chapter to stations from AERA-II. Whenever we discuss stations in this thesis, we refer to AERA-II antenna stations, unless explicitly stated otherwise.

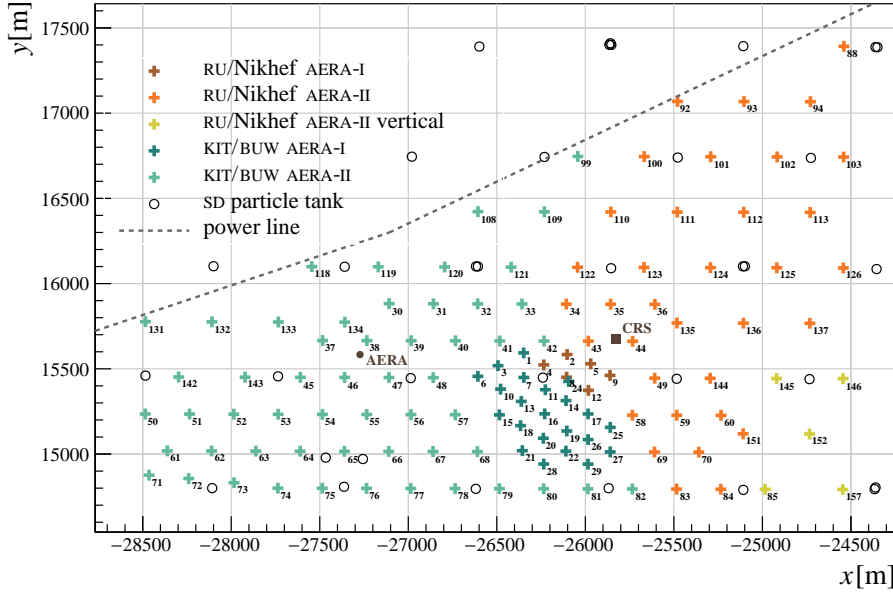


Figure 4-7: Layout of the AERA experiment. The colored crosses denote the different types of stations in the array. The stations equipped with KIT/BUW electronics use an external trigger, while the stations equipped with RU/Nikhef electronics are self-triggering on radio (AERA-I) and scintillators (AERA-II). The central radio station (CRS) and the auxiliary weather station (AERA) are also indicated, as well as the positions of the SD particle tanks in this area and the nearby power line.

4.3.2 Antenna station design

A schematic drawing of the AERA-II station design is visible in figure 4.8. The Butterfly physics antenna was designed to be sensitive between 30.0 MHz and 80.0 MHz, and measures in two polarization directions. One antenna arm is aligned parallel to the geomagnetic field (approximately 3° east of true north), with the other arm perpendicular to it. The stations are designed to operate autonomously, generating power with a solar panel attached to the electronics box, which houses the digitizer and a battery. In the stations equipped with RU/Nikhef electronics, this box also houses the scintillator(s) used for triggering. A 5.7 GHz commercial wireless communications system antenna is installed on top of the physics antenna, which communicates with one of the central data acquisition systems (DAQs). In addition, a GPS receiver is attached to the station to provide timing.

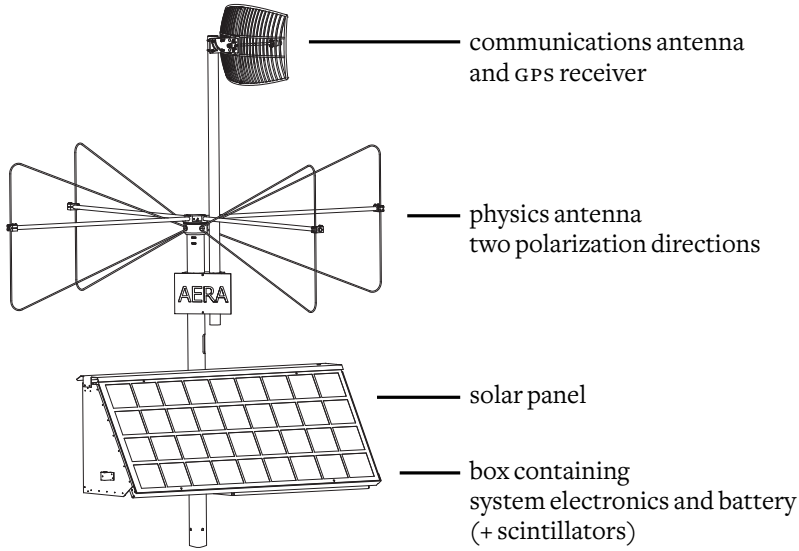


Figure 4.8: Station design of AERA-II. The stations with RU/Nikhef electronics are equipped with one or two scintillators in the electronics box.

Station electronics

The measured signal is amplified by a low-noise amplifier (LNA), located close to the center of the Butterfly physics antenna. The characteristics of the signal entering the electronics are determined by the interplay between the LNA and the antenna [96]. The frequency responses and the group delays are therefore combined in a single response pattern, which will be called the *antenna response*. It should be noted that because of the directional sensitivity of the antenna this is not just dependent on the frequency and the phase of the signal, but is also a function of direction of the incoming radiation.

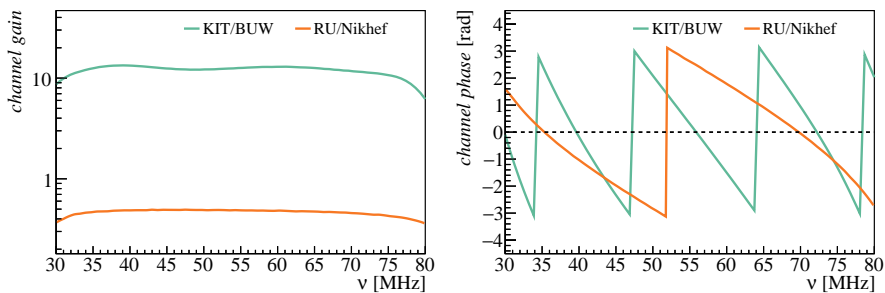


Figure 4.9: Comparison of the measured response of a high-gain channel of KIT/BUW station AERA 33 and a channel of RU/Nikhef station AERA 93.

After the LNA, the signal is fed via coaxial cables to the electronics box and the filter boards inside the digitizer, which filter the signal between 30.0 MHz and 80.0 MHz. The KIT/BUW filter/amplifiers contain two low-gain channels and two high-gain channels to increase the total dynamic range of the measurements. With respect to the low-gain channels, the high-gain channels are amplified with an additional 20 dB [99]. The two filter channels that are used for the radio measurements in the RU/Nikhef electronics use no amplification at all. In figure 4.9, a comparison of the filter characteristics as measured in the two types of station electronics can be seen. The gain difference resulting from the amplification in the KIT/BUW electronics can clearly be seen. We call the combination of the frequency-dependent delays introduced by the cables and the amplification and dispersion caused by the filter boards the *channel response*. A schematic overview of the analog electronics can be seen in figure 4.10.

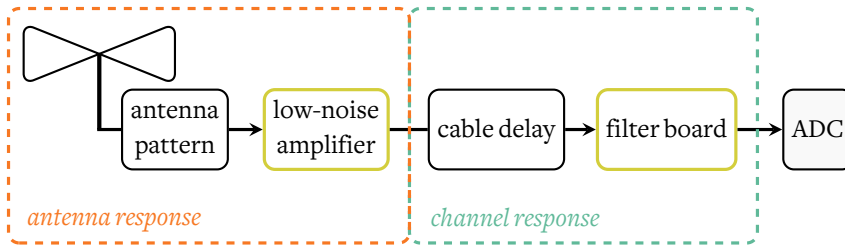


Figure 4.10: Overview of the logical analog electronics components from the antenna up to the analog-to-digital converter (ADC).

After filtering, the signal is processed by analog-to-digital converters (ADCs), four of which are present in each digitizer. In RU/Nikhef digitizers these are 14 bit ADCs, which digitize the signal at 200 Msps, while the KIT/BUW digitizers use 12 bit ADCs, which sample with 180 Msps. The digitized signals are further handled by the field-programmable gate array (FPGA) and the central processing unit (CPU) on the digitizer board, which are able to acquire and format the data and make trigger decisions.

Two channels in the RU/Nikhef digitizers are used to read out the signals from the antenna, while for the other two channels the filters are bypassed and are used to process signals from the scintillators. All AERA-II stations with RU/Nikhef electronics are equipped with two small plastic scintillators, except for the five experimental stations with vertically polarized antennas, which only have one. The scintillators measure $(45.7 \times 17.2 \times 2.4) \text{ cm}^3$, with one of them positioned on top of the battery and one below it (in case there is only one, it is positioned below the battery). The scintillators measure the charged particles in the air shower and are used for triggering, as will be discussed in the next section.

The KIT/BUW digitizers split the signals coming from the two antenna directions into high-gain and low-gain channels, which then take up all four channels of the digitizer.

These digitizers also contain a ring buffer to store data for up to 7.4 s while the station waits for the DAQ to send external trigger messages [100]. The RU/Nikhef electronics does not accept external triggers from the SD or FD and does not have such extensive data storage.

4.3.3 Trigger and data acquisition

Because of the fundamental differences between the trigger strategies of the two types of electronics, both experimental setups are connected to their own respective DAQs, which are described below. In the remainder of this thesis, we will use the term *experimental setup* to refer to either one of these systems.

Scintillator-triggered setup (RU/Nikhef)

The stations equipped with RU/Nikhef digitizers have the capability to self-trigger on any of the four connected channels. In its simplest form, a self-trigger can be just a signal crossing a threshold. The scintillator channels are shielded from most outside influences, so the signal is very clean and there is no need for a refined trigger scheme, and a single threshold is used. During the commissioning phase of AERA-II this threshold was set at 150 ADC counts below the dynamically established baseline of the signal feed.

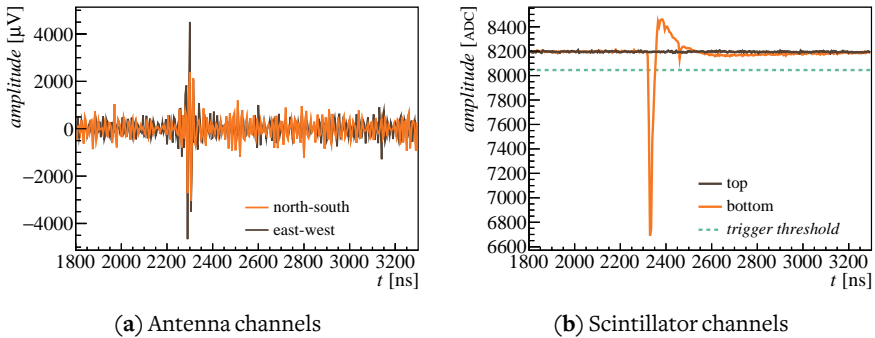


Figure 4.11: Radio traces and scintillator traces of the same station of the same event. The bottom scintillator detected a charged particle and triggered the local station, which recorded a radio pulse at roughly the same time.

For the signal coming from the radio antenna this is different. The signals are weak and inherently noisy, and can be contaminated by pulsed radio frequency interference (RFI) that is unrelated to the radio signal from air showers. For AERA-I, an elaborate trigger algorithm was developed to reject man-made noise pulses with specific characteristics. It uses a multi-step process to reject pulses with multiple secondary threshold-crossings, a feature common in broad pulses which are typical for man-made transient noise [101].

This algorithm has been retained in the AERA-II stations. An example of all four channels recorded by a station triggered by the scintillator below the battery is shown in figure 4.11. A radio pulse was recorded within a few tens of ns of the particle hitting the scintillator, and can therefore be considered to be air shower induced with high certainty.

If the FPGA and CPU decide there is a trigger on station level, the data is temporarily stored and a time stamped message is sent to the DAQ at the CRS. If three or more time stamps received by the DAQ fall within a coincidence window of 3 μ s, the DAQ requests the data from all triggered stations and builds an event. Initially, only scintillator triggers were used in the AERA-II DAQ, but since September 2013 also triggers on the radio signal are included. This has the advantage that stations that produce no scintillator trigger because of the small cross-section of the scintillator module can still be included in the event if they measure a radio pulse. In this case, a minimum of two stations triggered by scintillator is required for an event, but the total number of triggered stations still has to be at least three. A schematic overview of the DAQ of the scintillator-triggered setup is drawn in in figure 4.12.

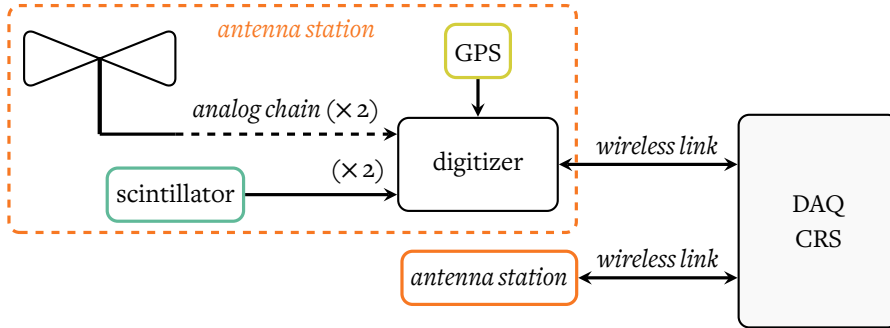


Figure 4.12: Data acquisition of the scintillator-triggered setup (equipped with RU/Nikhef digitizers). Two of many antenna stations are displayed.

After the DAQ has created an event, it is written to disk and saved for off-line analysis. Initially, the length of all recorded time traces was 1024 samples (5120 ns), but this was increased to 2048 samples (10 240 ns) after 2 October 2014 to better be able to estimate the properties of narrowband noise (see section 6.3). In addition to the self-triggered events, the DAQ requests read-outs of all active stations once every 10 s. These events can for instance be used to study detector performance or the noise situation (see section 5.2) and are called *periodically triggered* data.

Externally triggered setup (KIT/BUW)

The DAQ of the externally triggered setup is located at Coihueco and receives messages with time stamps and station lists of SD events directly from the CDAS. Usually, they arrive within a few seconds of the impact time of the shower. All SD events of which the nearest surface detector station is within 5 km of any of the AERA stations in the externally triggered setup are selected by the DAQ. Using the trigger time stamp of this nearest SD station, it is possible to perform a very rough approximation of the position in the ring buffer of a radio pulse that might be produced in the event. This approximation is then used in a trigger message from the DAQ that requests a read-out of all active stations equipped with KIT/BUW electronics. In addition to SD triggers, triggers from the FD detectors Coihueco and HEAT are also accepted by the DAQ. A schematic overview of the DAQ of the externally triggered stations is visible in figure 4.13.

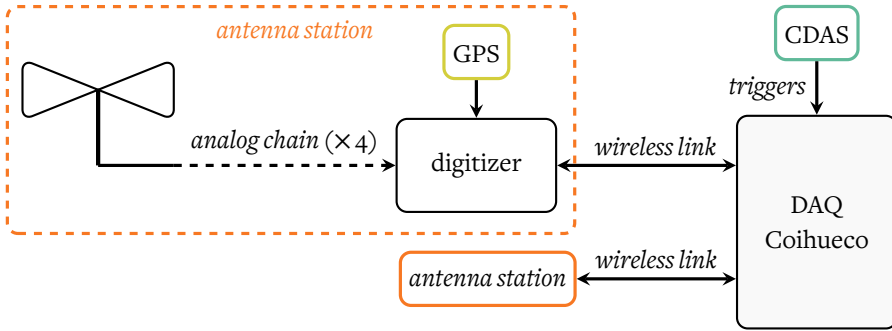


Figure 4.13: Data acquisition of the externally triggered setup (equipped with KIT/BUW digitizers). Two of many antenna stations are displayed.

The exact position of a radio pulse in the ring buffer depends strongly on shower geometry, which is unknown without performing a directional reconstruction. Therefore, the time traces that are written to disk are chosen to be very long to increase the probability that the pulse is contained by them: 10 240 samples (56 888.9 ns). In addition to external triggers, the digitizers developed by KIT/BUW also have self-triggering capabilities, but this data is currently not considered for physics analysis. The DAQ of the externally triggered setup also used to store periodically triggered data every 10 s, but to conserve power this has been decreased to once every 100 s as of March 2014.

4.3.4 Electric field reconstruction

Off-line analysis

After the data recorded by the radio stations is stored to disk by the DAQ, it has to be analyzed off-line and merged with coinciding SD and FD events. This can be done on-site, such as at the Coihueco DAQ, or at one of the home research institutes. A tool that was

specifically created for analyzing data from the Auger Observatory is the *Offline* software framework [102], which contains analysis modules for SD and FD, and has been extended with functionality for data from the radio detectors [103].

The *Offline* framework is written in C++. It makes a clear distinction between *detector description* and *event* data structures. The detector description deals with all properties related to the hardware in the field, such as the position of a particular detector station or the quantization levels of an ADC. The event data structure on the other hand holds all the data concerning a specific event, such as ADC traces or a reconstructed shower axis. Analysis modules are used to access the two interfaces, and can modify the event data structure, such as clipping time traces or filtering frequency bands. An analysis pipeline, for example to reconstruct the shower axis of an event, usually consists of a chain of modules which each perform specific analysis tasks. The modules can be configured by XML files, and detector properties are either stored in XML tables or in MySQL databases.

In order to perform a physics analysis like the one outlined in chapter 3, we need to reconstruct the three-dimensional electric field at the antenna positions from the time traces in ADC-counts from the two channels as recorded by the radio detector stations. By using the quantization levels of the ADCs we can revert to the original signal voltages, which were fed into the ADCs. We can then correct for the response of the cables and filters by applying the inverse of the channel response functions that were measured for all radio stations in the field. This is done in the frequency domain, and examples of these response functions can be seen in figure 4.9. Disregarding electronic noise, the voltages after applying the inverse channel response are the two components of the signal voltage at the readout terminals of the LNA. From this, we are able to reconstruct the full three-dimensional electric field if we know the arrival direction of the pulse and the directional sensitivity of the antennas. Before, after and in between these reconstruction steps, we can place selection cuts or perform calibration and correction operations. These will be discussed in depth in chapter 5.

Electric field reconstruction

The antenna response pattern represents the mapping of the electric field \mathbf{E} to the measured voltages $U_{1,2}$ at the readout terminals of both antenna arms, and is denoted by the *vector effective length* (VEL) $\mathbf{H}_{1,2}$ of each antenna arm:

$$U_{1,2} = \mathbf{E} \cdot \mathbf{H}_{1,2}. \quad (4.2)$$

Because we measure with only two antennas, we have one free parameter too many to solve this equation. We can, however, use the fact that the electric field parallel to the direction of movement of the shower (\mathbf{v}) is negligible [104], reducing equation (4.2) to

a two-dimensional problem. The electric field can be projected onto two components orthogonal to the shower axis: \mathbf{e}_θ and \mathbf{e}_ϕ , as is illustrated in figure 4.14.

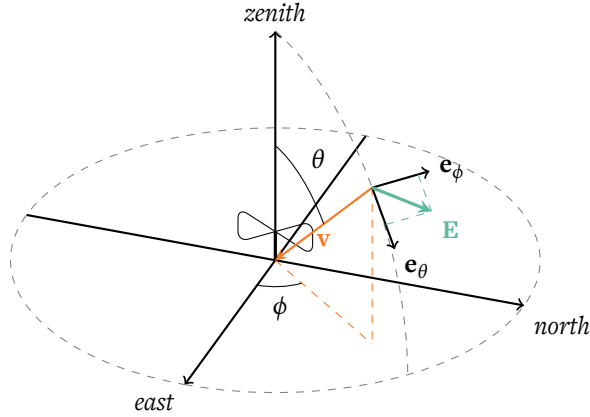


Figure 4.14: Coordinate system in which the electric field vector \mathbf{E} is projected orthogonal to the shower axis \mathbf{v} along unit vectors \mathbf{e}_θ and \mathbf{e}_ϕ .

We can rewrite equation (4.2) in this representation and write out the dot product:

$$U_1(v) = H_{1,\theta}(v, \theta, \phi)E_\theta(v) + H_{1,\phi}(v, \theta, \phi)E_\phi(v) \quad (4.3)$$

$$U_2(v) = H_{2,\theta}(v, \theta, \phi)E_\theta(v) + H_{2,\phi}(v, \theta, \phi)E_\phi(v), \quad (4.4)$$

where we evaluate the signal voltages, electric fields and the v_{EL} in the frequency domain. With these two equations, we can solve the two components of the electric field vector, given that the arrival direction of the pulse is known, and there is a complete description of the v_{EL} of both antenna arms as a function of arrival direction and frequency. The arrival direction can be obtained from a directional reconstruction using the radio signal arrival times, or from the directional reconstruction of the coinciding SD event. The v_{EL} can be produced by using computer simulations of the antenna characteristics from a detailed description of the antenna. It is also possible to measure the v_{EL} by using well-defined incoming electric fields, such as coming from calibrated emitters for a variety of directions and frequencies.

It is important to note that equations (4.3) and (4.4) can only give an exact solution of the full electric field in complete absence of noise. In the real world, a variety of noise contributions contaminate the signal, and while the air shower induced field may be contained in two components orthogonal to the shower axis, the noise is randomly polarized and can come from any direction. When either the signal strength is too low compared to the noise level, or the signal is mostly contained in either one of the two channels (for example, an inclined shower arriving parallel to one of the antenna arms),

this will lead to significant mixing and amplification of noise into the reconstructed electric field when equations (4.3) and (4.4) are applied.

We can illustrate this by a toy Monte Carlo, where we can look at how well certain parameters can be recovered from the radio signal as a function of arrival direction under the influence of noise. We will use one simulated shower from the set described in section 3.3. From this shower, we have selected 50 traces in three dimensions containing pulses with spectral indices b_{real} ranging from -1.5×10^{-2} to 0.0, which were calculated on the $\mathbf{v} \times \mathbf{B}$ component. We will randomly draw from this set. We will also assign a random arrival direction to these traces, and add Gaussian white noise with a signal-to-noise ratio SNR between 1 and 50 as defined by:

$$\text{SNR} = \sum_i^n \frac{|x_{\text{signal},i}|}{|x_{\text{noise},i}|}, \quad (4.5)$$

where $|x_{\text{signal},i}|$ is the magnitude of the signal amplitude spectrum of frequency bin i , and $|x_{\text{noise},i}|$ is the magnitude of frequency bin i of the generated noise spectrum. After the addition of noise, we can apply the simulated antenna response pattern, which describes the VEL of the Butterfly antenna, and then apply the inverse antenna response on the two signal voltages to get a three-dimensional electric field again, in both cases using the same arrival directions. We can then recover the $\mathbf{v} \times \mathbf{B}$ component of the field, calculate the pulse amplitude and fit the spectral index. We define the original pulse amplitude A_{real} before applying the antenna pattern as:

$$A_{\text{real}} = \sum_i^n |x_{\text{signal},i}|. \quad (4.6)$$

The measured pulse amplitude A_{measured} after the addition of noise and going back and forth through the simulated antenna response pattern can be calculated when we correct for the noise:

$$A_{\text{measured}} = \sum_i^n \sqrt{x_{\text{signal},i}^2 - x_{\text{noise},i}^2}. \quad (4.7)$$

After doing this five million times we can look at the average difference between the original and measured values of the pulse amplitude and the spectral index as a function of arrival direction, as can be seen in figure 4.15.

It is obvious that the average deviations in the measured parameters become very large (compare for example with the values in figure 3.15 for the spectral index) at zenith angles above 60° . This is mostly caused by events that are arriving more or less parallel to one of the antenna arms at high zenith angles, where one of the arms will have its signal completely drowned by noise, which in turn will be blown up when calculating the vertical component of the electric field. The specific shape of the pattern that is visible above $\theta = 60^\circ$ is the result of the interplay between the antenna response pattern and

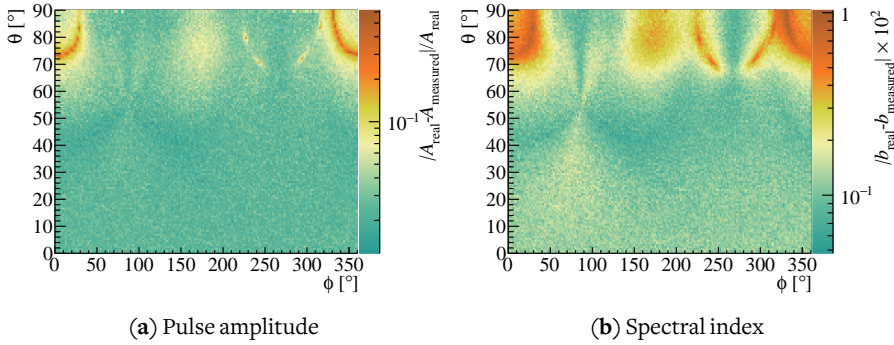


Figure 4.15: Average difference between the real and measured values of the pulse amplitude (a) and the spectral index (b) as a function of arrival direction when the antenna response pattern is applied in the described toy Monte Carlo. The pulse amplitude is displayed as a relative deviation.

the magnetic field vector \mathbf{B} that was used to determine the component of the electric field of which the spectral index is calculated.

Data sets and event selection

IN THIS CHAPTER we will discuss how we obtain a high quality data set from the data written to disk by the two data acquisition systems on which we can perform physics analyses.

We start off by discussing the sources of radio frequency interference (RFI) influencing our measurements, as this is crucial if we want to make sense of the recorded signals. Next, we will outline a method to monitor data quality, remove dysfunctional stations from the data set and flag periods with thunderstorm conditions. After this, we are able to go through data selection procedures for the scintillator-triggered and externally triggered setups to extract air shower induced radio pulses.

The data set that is used for the physics analysis in this thesis runs from 10 May 2013 to 19 November 2014 in the case of the scintillator-triggered setup. The data set that was used from the externally triggered setup runs from 22 April 2013 to 13 April 2015. The bounds of these data sets are set by the commissioning dates of the two experimental setups on the one hand, and the available volume of periodically triggered data (which is only sporadically transferred by hard drive to the home institutes) on the other.

5.1 Radio frequency interference at AERA

In addition to air shower induced radio pulses, the antenna stations constantly record a combination of ambient noise picked up by the antennas and electronic noise introduced by the system electronics. The ambient noise mostly consists of radio sources in the sky. The galactic center is the most important source in our frequency range and can be used for calibration purposes (see chapter 6). These contributions are superimposed by man-made radio frequency interference, which we can subdivide into narrowband and broadband RFI.

5.1.1 Narrowband RFI

Several sources of narrowband noise negatively influence the data recorded at the AERA site. Narrowband RFI can be identified as high-powered sources in a small spectral region, usually peaking at a single frequency, as can be seen in the frequency spectra of figure 5.1. Depending on the location of the source, it might show up in either one or both of the channels.

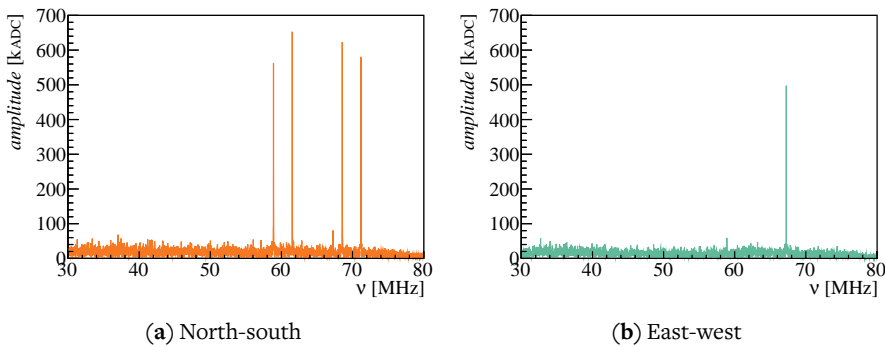


Figure 5.1: Narrowband RFI visible in both channels of station AERA 64 in an externally triggered event from August 2014.

The strongest source of narrowband RFI at the AERA site is the AERA beacon [105]. Conceived as a timing calibration source, it continuously emits sine waves with the frequencies listed in table 5.1. Its antenna is located at the Coihueco FD building, and is therefore mostly visible in the north-south channel, although this depends on the exact station location relative to the beacon.

Table 5.1: List of AERA beacon frequencies as of 14 August 2012.

Beacon frequency [MHz]	58.887	61.523	68.555	71.191
------------------------	--------	--------	--------	--------

The nature and origin of most other sources of narrowband RFI is unknown. Some have a very narrow frequency range, while others are somewhat broader. An important example of the latter is the transmitter at ~ 67 MHz in the east-west channel, which corresponds to channel 4 of the lower VHF television broadcast band. This is a frequency modulated signal, centered around a carrier frequency of 67.25 MHz [106]. Although the beacon and the ~ 67 MHz transmitter are almost always present in the recorded data, several other sources are of a more transient nature and related to human activity. In order not to bias our analysis of the frequency spectrum of air shower induced radio pulses, the contribution of narrowband RFI to the pulse needs to be suppressed, as will be explained in detail in section 6.3.

5.1.2 Broadband RFI

Broadband pulses that mimic air shower induced radio pulses are another source of contamination to our data. In radio self-trigger mode, the station level trigger rate of AERA-I was in the order of 200 Hz, while the coincidence rate with SD was only about one per day [61]. This means that, regardless of the trigger mechanism, broadband pulses are commonplace at AERA and the probability of misidentifying noise pulses as air shower pulses is high. By using very short coincidence windows and careful comparison with reconstructed arrival directions from SD, we try to limit the number of misidentified air shower induced pulses, as will be shown in section 5.4.

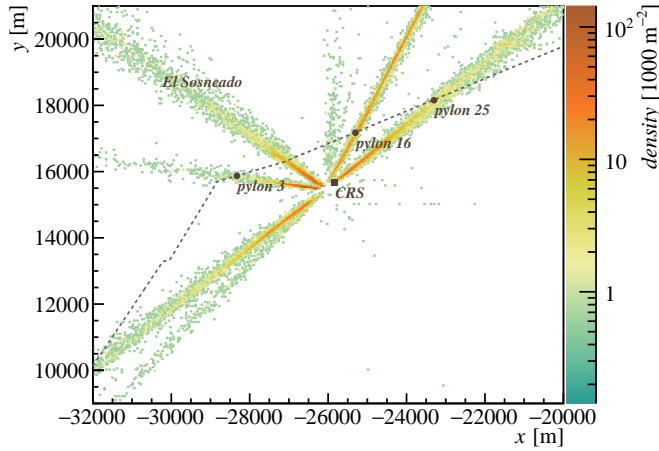


Figure 5.2: Reconstructed sources of pulsed noise of AERA-I self-triggered data from 29 August 2012 to 10 September 2012. The power line is marked by a dashed line. Because of the limited spatial distribution of the stations and timing uncertainties, the sources are reconstructed in tracks fanning out from the array.

To investigate the nature of sources of broadband pulses we can perform a directional reconstruction of events which were self-triggered on the radio signal. This is done based on the pulse time, assuming a spherical radio wave front centered at the source (see e.g. [104]). In figure 5.2, the reconstructed sources from 12 days of AERA-I self-triggered data are plotted, after selecting only events of which a plane wave arrival direction reconstruction points towards the horizon. A few pylons of the power line were identified in the field to be sparking with a hand-held oscilloscope and an antenna. This is probably because of defective isolators, and their locations are drawn in figure 5.2 as well, showing an obvious correlation with frequently reconstructed directions. In addition to the power line, several transformers in the neighboring village of El Sosneado are suspected to be sources of broadband pulses.

5.2 Station performance monitoring

5.2.1 Method

To monitor the performance of the AERA antenna stations we use the periodically triggered data, which is created every 10 or 100 seconds (see section 4.3.3). The uptime of the stations is determined by the mere presence of the stations in this data stream, while the quality of the data which the stations record is inferred from its noise level. A convenient quantity by which to monitor the noise level of the stations is the *root mean square* (rms) of the amplitude of the time trace recorded by the stations, defined as:

$$x_{\text{rms}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}, \quad (5.1)$$

where x_i is the amplitude in ADC counts of sample i , and n the number of samples in one trace. The value is corrected for the average amplitude of the trace \bar{x} . If we assume that all antennas have roughly the same gain, abnormal behavior of a station will manifest itself as a deviation of the rms from the mean rms of all stations. We use this behavior to reject badly performing stations.

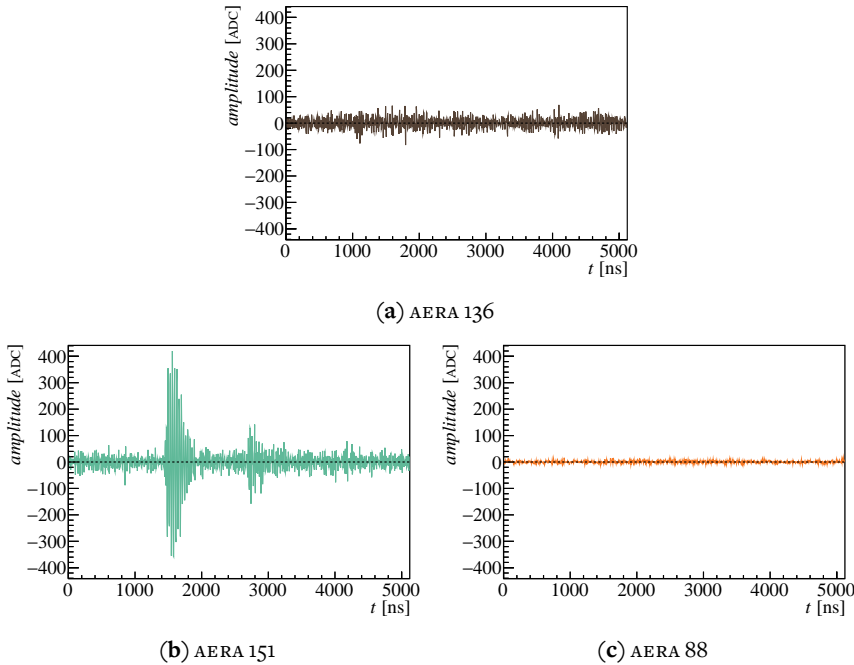


Figure 5.3: Time traces of the north-south antenna channel from event 744026 from run 537 of the scintillator-triggered setup recorded on 29 November 2013.

In figure 5.3, the time traces of several stations within one periodically triggered event are shown. The trace of AERA 136 is typical for traces recorded by correctly functioning stations, whereas AERA 151 has strong pulses and AERA 88 is quiet by comparison. The source of the pulsing behavior is not exactly known, but might result from a failure in the grounding of the electronics. The quiet traces are likely caused by a broken fuse on the filter board. The differences between these stations are also reflected in the rms, as can be seen in figure 5.4. On top of the quasi-daily galactic variation and the occasional (man-made) noise burst a clear enhancement is visible for AERA 151 caused by the pulses, while the rms of AERA 88 is nearly a flat line.

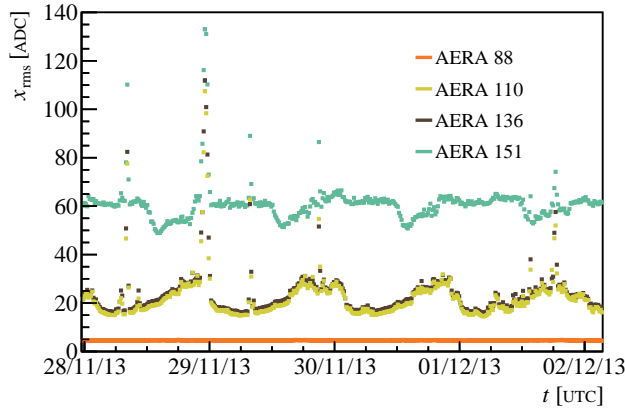


Figure 5.4: The rms averaged in 100 bins per day of the time trace in the north-south polarization channel, as a function of time for four stations from run 537. AERA 110 and AERA 136 are behaving normally.

To reduce the effects of short term bursts of noise and the galactic variation we first calculate the average and spread of x_{rms} for 100 bins a day. We then calculate a weighted average from these values for periods of two days, $\langle x_{\text{rms}} \rangle$. Episodes of pulsed man-made noise result in a large spread, so we can negate their effects if we weigh every contribution with the inverse of the spread in that bin. The time evolution of this average for all stations from the scintillator-triggered setup in a period of little more than a month is visible in figure 5.5. It is clear that the correctly functioning stations are grouped around a certain mean value, which changes over time because of variations in the background noise. Dysfunctional stations are easily identified as outliers to this distribution. A colored band marks the region of 6 standard deviations above and below the mode of the distribution of stations per time bin. Stations within this region will be considered normal, while the rest will be rejected with a time resolution of two days. Exceptions are possible because of specific noise conditions and changes to the specific ensemble of active stations at any instance of time.

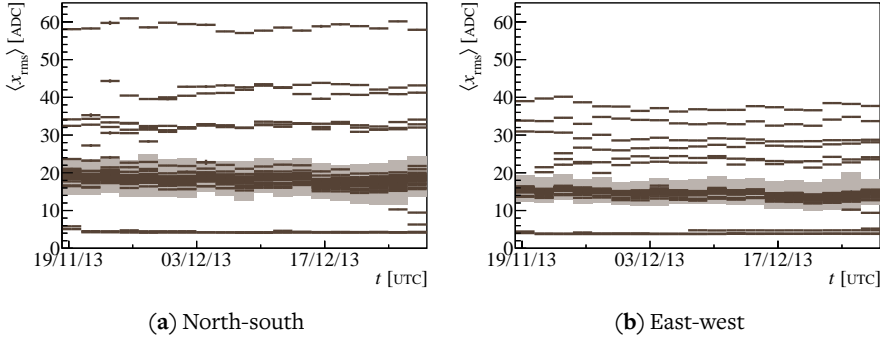


Figure 5.5: The rms of the time trace of every station of the scintillator-triggered setup averaged over a period of two days as a function of time in run 537. The shaded bands indicate 6 standard deviations above and below the mode of the distribution per time bin.

As is visible in figure 5.5, the noise levels recorded by the two antenna arms as well as their spread are quite different. This is caused by a combination of the asymmetry of the antenna sensitivity pattern, and the location of the galaxy and other noise sources relative to the antenna. Both antenna arms should therefore be treated separately in this analysis. The distance from a specific antenna station to man-made noise sources also has a strong influence on the noise level recorded by a station. Therefore, we try to limit the influence of some known narrowband sources by setting the amplitudes in their frequency bands to zero in both antenna arms before calculating $\langle x_{\text{rms}} \rangle$. These frequency bands are listed in table 5.2.

Table 5.2: Filtered frequency bands containing strong narrowband noise, such as the AERA beacon.

Filter lower bounds [MHz]	57.7	60.5	66.5	70.0
Filter upper bounds [MHz]	59.5	62.0	69.0	72.0

5.2.2 Results

After determining the rejected periods for all available periodically triggered data, we can look at the fractional uptime and the rejected periods per station. The fractional uptime is defined as the fraction of periodically triggered events in which a station is present in one full day. A station can be absent from a periodically triggered event because it is non-responsive or because it is busy processing a trigger, but it can also be completely shut down. For two example stations, the uptime as a function of time is plotted in figure 5.6.

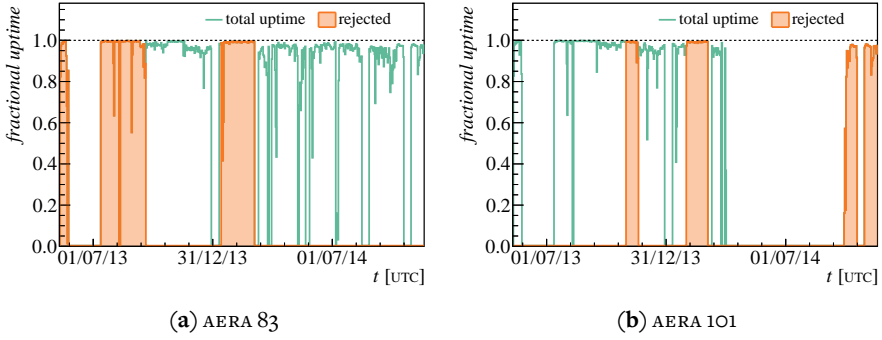


Figure 5.6: The fractional uptime as a function of time in the complete data set for two example stations of the scintillator-triggered setup, with colored regions marking the rejected periods as determined by the rms of the time trace.

Because of the different filter-amplifiers in the two experimental setups we need to treat the data from the scintillator-triggered setup separately from the externally triggered data set when determining the distribution of $\langle x_{\text{rms}} \rangle$. An overview of the total fractional uptime and rejected periods for all stations as a function of time for the complete available data set of each of the setups is plotted in figure 5.7.

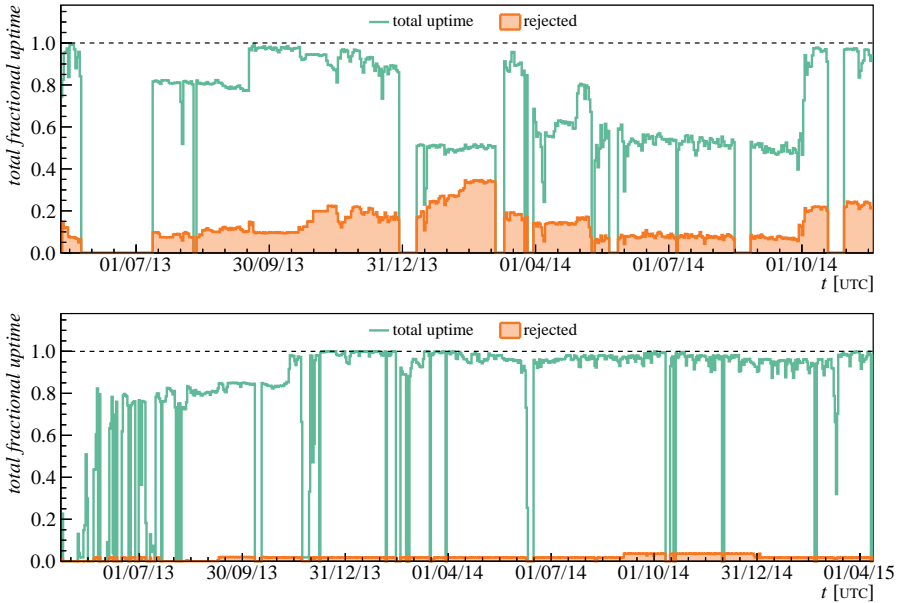


Figure 5.7: Total fractional uptime and rejection as a function of time for all stations in the scintillator-triggered (*top*) and the externally triggered (*bottom*) setups.

In figure 5.8, the fractional uptime and the rejected periods are listed for each station. The data produced by stations in periods in which they are rejected will not be used in any calibration method or physics analysis described in the rest of this thesis.

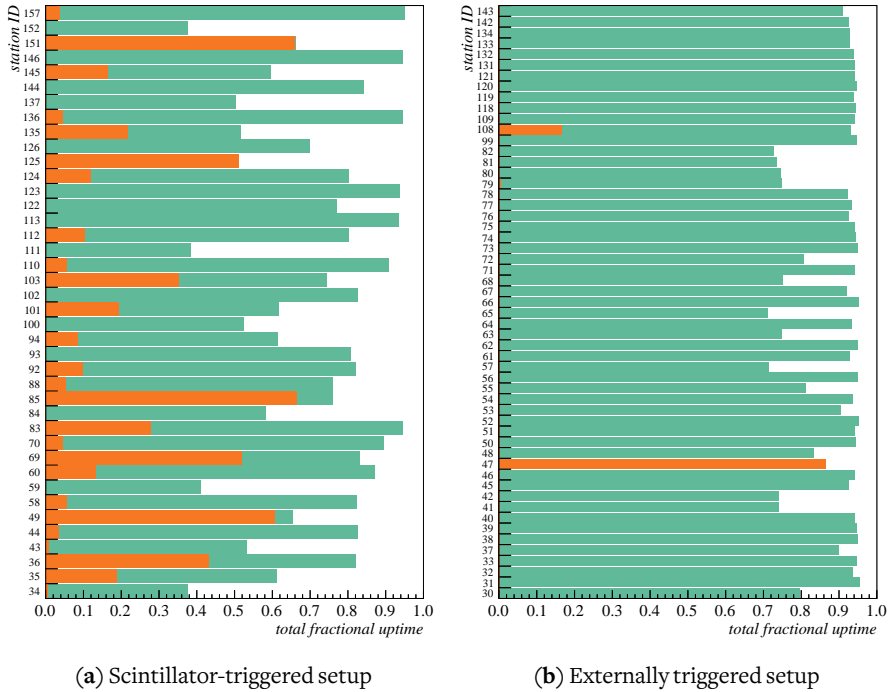


Figure 5.8: Total fractional uptime and rejection of each station in the available data set of both experimental setups. The orange bars indicate the fraction of time in which the stations were rejected.

Every now and then the self-triggered stations go through short episodes where they store saturated signals at 50 MHz, with durations of the order of one hour. These periods are also easily distinguishable by their rms, and will be flagged separately with a time resolution of 100 s. In addition to the rejected stations based on the rms of the trace, some experimental stations will be removed from the analysis. As mentioned in section 4.3.1, five stations were outfitted with an experimental antenna which is sensitive to the vertical electric field component for a period of time. Due to complications in modeling and understanding the antenna response pattern, these station cannot reliably be used in a physics analysis. The affected stations are AERA 85, AERA 145, AERA 146, AERA 152 and AERA 157.

5.3 Weather monitoring

Ambient electric fields have an influence on the radio emission from air showers [107, 108]. The strength of the ambient electric field is related to local weather conditions, and changes considerably during thunderstorms. To prevent the inclusion of events where the ambient electric field might have enhanced the air shower induced pulse we need to flag the periods where the field deviates from its nominal value. This can be done by looking at the electric fields as measured by electric field mills located in weather stations which are placed within the array [109]. One weather station is located on the roof of the CRS, and since 29 July 2014 an additional weather station has been operational in the antenna field about 1400 m due west from the CRS (see the map in figure 4.7). We will refer to the former as the CRS weather station, and to the latter as the AERA weather station. The half-hour mean of the measured electric field as a function of time during two periods is plotted in figure 5.9. A period with clear weather is plotted in figure 5.9(a), in which only small oscillations were recorded. In figure 5.9(b), a period with daily thunderstorms is plotted, with enormous variations compared to the clear period were recorded by both weather stations.

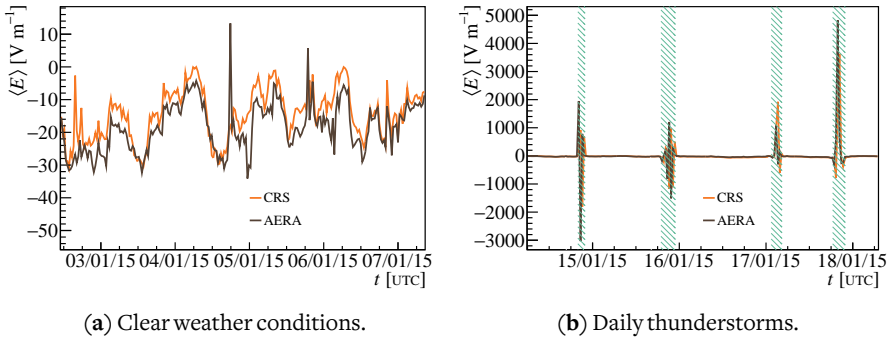


Figure 5.9: The mean electric field in bins of 30 minutes as a function of time in two periods as measured by the two weather stations. The hatched regions are periods flagged as thunderstorms.

To distinguish clear weather conditions from enhanced fields we can place limits on the mean electric field $\langle E \rangle$ and its rms E_{rms} , calculated for periods of 30 minutes [68]. The distributions of the mean value and rms of the electric field for the period from 20 April 2013 to 16 May 2015 (and thus covering our entire data taking period) are plotted in figure 5.10. From these distributions it is possible to estimate the limits within which the electric field conditions can still be considered normal. Just like in [68], we set the limits to $-150 \text{ V m}^{-1} \leq \langle E \rangle \leq 50 \text{ V m}^{-1}$ and $E_{\text{rms}} \leq 30 \text{ V m}^{-1}$. Outside of these regions, the half-hour bins will be flagged as thunderstorm periods.

From 29 July 2014 onward, when the AERA weather station became active, we will flag

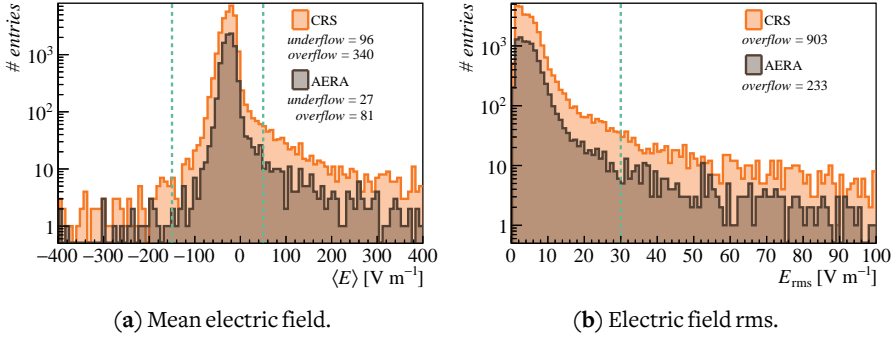


Figure 5.10: Distributions of the mean value and the rms of the electric field as measured by the two weather stations from 20 April 2013 to 16 May 2015. The cuts which separate clear from thunderstorm conditions are marked with dashed lines.

bins as thunderstorm periods whenever either one weather stations records conditions outside of the aforementioned limits. In the entire data taking period from 20 April 2013 to 16 May 2015, there was a total down fraction of the weather stations of 14.9 %. In these periods the electric field is not known. In the same period, 5.4 % of the time has been flagged as thunderstorm periods, and air shower induced radio pulses from these periods will not be used. Whenever events for which no weather information is available are used in this thesis, such as in the physics analysis of chapter 7, this will be mentioned explicitly.

5.4 Event selection

In order to perform a cosmic ray physics analysis on our data set we need to select events that contain air shower induced radio pulses. Even though the antenna stations trigger on air showers by means of scintillators or the surface detector, many traces recorded by the radio stations will contain no air shower induced pulses above the noise level. This is because the lateral signal power distribution falls off steeply for events with low zenith angles (see e.g. [110]), resulting in a small radio “footprint” in comparison to the particle front, which impacts the earth’s surface. In addition, the radio measurements are highly sensitive to pulsed broadband noise, which contaminates our events and needs to be rejected.

Because we have two data sets created by two experimental setups using different trigger mechanisms, we also need two different strategies for event selection. In both approaches, we will require that the radio event is in coincidence with an event recorded by the surface detector from which parameters such as the core impact location, shower

axis and primary energy can be reconstructed. All operations performed on the data in the selection procedure, such as the filtering of narrowband noise, only affect the data for the duration of the selection procedure. For any higher level analysis later in this thesis we will revert to the original raw recorded data, to which we will apply more elaborate calibration and filtering techniques, which will be described in chapter 6.

5.4.1 Scintillator-triggered setup

The events triggered by the scintillators first have to be compared to SD events to find coincidences. In contrast to the radio self-trigger from AERA-I , the scintillator event trigger rate is low – only about 0.05 Hz. Because the scintillator-triggered radio traces have fixed pre-trigger windows, the radio event time stamp created by the DAQ is strongly correlated with the impact time of the shower. The radio time stamps are compared with the SD core impact time as reconstructed by the CDAS v5r2 software package [111]. The raw event trigger rate of SD is about 0.2 Hz, but the rate of events of which the shower geometry can be reconstructed is about one tenth of this. In figure 5.11, the difference between the radio time stamp and the reconstructed SD core impact time within a window of $100\ \mu\text{s}$ is plotted for a data set covering 16 days. Because of the low event rate, all coincidences are contained within a few μs , and there is almost no background of uncorrelated events within the $100\ \mu\text{s}$ time window in this particular set. Accordingly, we select all radio events that are coincident with SD core impact time within $\pm 10\ \mu\text{s}$.

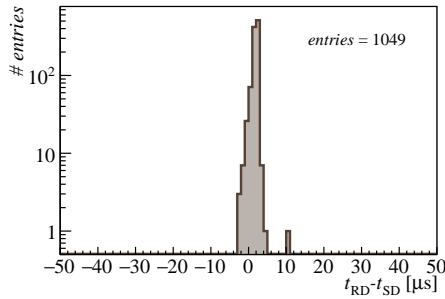


Figure 5.11: Distribution of scintillator-triggered events in coincidence with SD within a $100\ \mu\text{s}$ window in the period from 25 July 2013 to 11 August 2013. The total number of scintillator-triggered events in this period was 90 120.

Because of the difference between the radio and the particle footprint size, the selected events still mostly contain radio traces with just noise and no air shower induced radio pulses. To select stations with pulses we first need to determine a *signal window* in which to look for them. Because the scintillator trigger is located at $\sim 2325\ \text{ns}$ relative to the start point of the recorded traces, we select a window from 2000 ns to 2500 ns for the scintillator triggered stations. The distribution of the position of the radio pulse maxi-

imum and scintillator minimum calculated relative to the trace start time in this window for one and a half year of data is displayed in figure 5.12(a). For the scintillator traces, the absolute minimum was taken. For the radio traces, we have first converted electric field components to the shower reference frame using the reconstructed shower axis from SD (see also section 3.3.3). We have then determined the maximum of the *Hilbert envelope* of the electric field component parallel to the $\mathbf{v} \times \mathbf{B}$ vector, because this is where we expect most signal. The Hilbert envelope is a transform to remove the oscillations in the signal, and therefore provides a better estimation of the actual time of pulse maximum of the recorded signal. A detailed description of the Hilbert envelope can be found in [68].

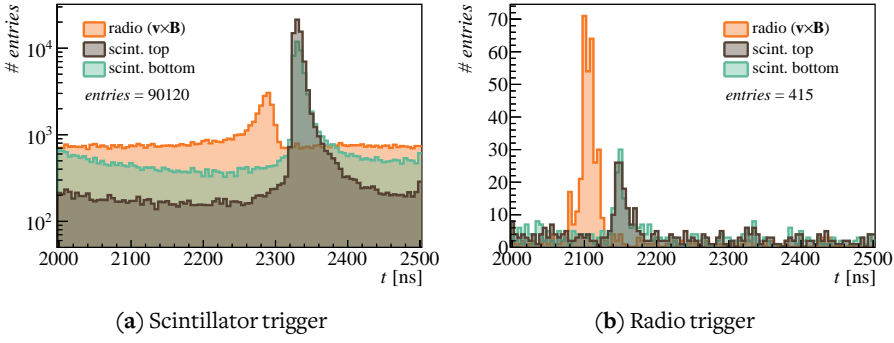


Figure 5.12: Position of the radio pulse maxima and scintillator minima relative to the trace start time, for scintillator-triggered stations (a) and radio-triggered stations (b). The histograms appear shifted because the radio signal was corrected for cable delays and the scintillator signals were not, which makes the signal shift forward. The histograms contain all radio events coincident with SD from 16 May 2013 to 28 September 2014.

From figure 5.12 it is clear that only a small fraction of the traces contains an air shower induced radio pulse, and they seem to be contained in a small time interval relative to the size of the signal window. Because of the shape of the radio shower front and differences in the arrival times of the radio pulse and the particles striking the scintillators, there might be radio pulses hidden in the noise background. These could be interesting events from an analysis point of view (for example, because they have a large distance to the shower axis), so we try to include these events by leaving the signal window at 500 ns. It should be noted that we do not correct for cable delays in the scintillator channels, which is why the particles appear to arrive later than the radio signal. In the actual air shower, the particles will usually arrive at the station before the radio signal because of the refractive index of air. This does however depend on the exact shower geometry and the distribution of the particles in the shower front, which becomes thicker at large distances from the shower axis.

A small fraction of the stations in the data set are triggered by the radio signal only, as can be seen in figure 5.12(b). The radio pulse maxima are shifted forward because the radio signals are corrected for cable delays, and the scintillator traces are not, which shifts the apparent trigger position. We see that some of the radio-triggered stations have a scintillator minimum ~ 50 ns later than the radio pulse maximum, exactly where we might expect it on the basis of the distribution of the scintillator trigger in figure 5.12(a). It turns out that these are failed scintillator triggers. By demanding a threshold crossing of 150 ADC units below the mean of the scintillator trace, we can reproduce the trigger logic and recover these stations as if they were scintillator-triggered (see section 4.3.3). This accounts for 146 triggered stations in the data set used to produce figure 5.12. The remainder of the radio triggered stations will be discarded for further analysis because they are probably mostly isolated noise pulses. By using the scintillator trigger as an indicator for an impacting air shower at the location of the station we do not need a full radio shower reconstruction to determine if a pulse is originating from an air shower. This allows for single-station radio events.

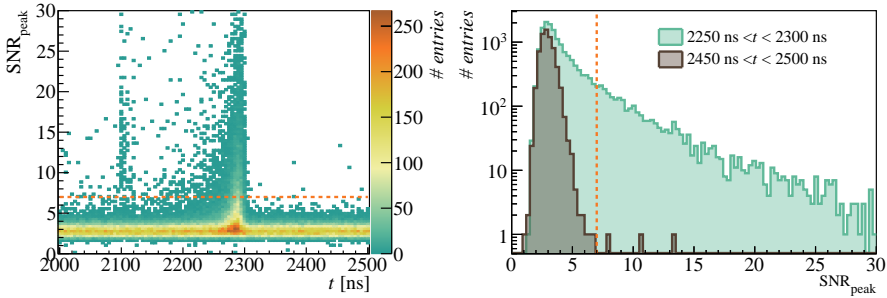


Figure 5.13: Signal-to-noise ratio as a function of maximum pulse time (*left*) of all scintillator-triggered stations and the radio-triggered stations with a scintillator peak below threshold, with histograms of SNR_{peak} in two specific time intervals (*right*). The dashed line marks the $\text{SNR}_{\text{peak}} = 7.0$ cut.

To select actual pulses in the signal window we can use the signal-to-noise ratio SNR_{peak} as a discriminator. We will define this as the maximum value of the Hilbert envelope of the electric field component parallel to the $\mathbf{v} \times \mathbf{B}$ vector $\hat{E}_{\mathbf{v} \times \mathbf{B}}$ divided by the root mean square of the noise $\text{rms}_{\text{noise}}$, calculated in a different region of the electric field trace of the $\mathbf{v} \times \mathbf{B}$ component:

$$\text{SNR}_{\text{peak}} = \frac{\max \hat{E}_{\mathbf{v} \times \mathbf{B}}}{\text{rms}_{\text{noise}}}. \quad (5.2)$$

The distribution of SNR_{peak} as a function of time of pulse maximum for the conjunct set of all scintillator triggered stations and the radio triggered stations with a scintillator pulse is plotted in figure 5.13. The signal region from the scintillator trigger can clearly be recognized in this plot, as can a small secondary peak which results from shifted ra-

dio triggered pulses. In the same figure, the distributions of SNR_{peak} in time intervals in the signal region and a more quiet region are compared. On the basis of these distributions we estimate that a cut of $\text{SNR}_{\text{peak}} = 7.0$ removes most of the noise from the data set. This does, however, not completely rid us of noise pulses which mimic air shower induced pulses, but these will be dealt with in section 5.4.3.

5.4.2 Externally triggered setup

By construction, every externally triggered event has a surface detector counterpart, so the off-line coincidence search for these events is rather trivial. As already mentioned in the previous section, about nine tenths of these events have to be discarded because the geometry of the SD event cannot be reconstructed. Externally triggered events contain read-outs of all active stations from the experimental setup. Because of the limited radio footprint size, the vast majority of the recorded traces in the event will therefore contain no air shower induced pulses above the noise level. When there is such a pulse, the time at which it occurs within the recorded trace depends on shower geometry, and is therefore very uncertain (see section 4.3.3). As a result, the time window in which we need to search for pulses is large, which increases the probability of pulsed noise to contaminate the trace. In figure 5.14, the location of the maximum of the time trace is plotted of all stations within a set of 55 externally triggered events. Above the background of stations with no measurable radio pulse, an excess is visible in a window from $26 \mu\text{s}$ to $33 \mu\text{s}$, which will be our signal window. It is clear that this window still contains a considerable number of background noise pulses. To make the distinction between noise and air shower induced pulses we will need to reconstruct the shower direction from the timing of the radio pulses and compare this with a directional reconstruction from the surface detector.

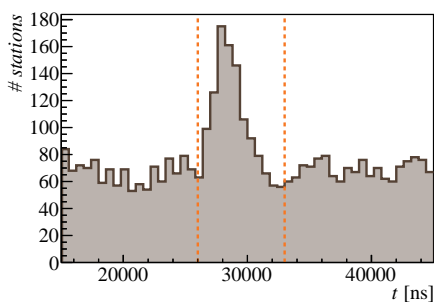


Figure 5.14: Location of the maximum of the time trace in all stations for 55 externally triggered events. The edges of the signal window are indicated with dashed lines.

Before we are able to perform a reliable directional reconstruction we first need to de-

termine criteria by which to select stations with strong enough pulses. Without this cut, stations in which the signal of the same magnitude as the ambient noise level will negatively influence the reconstruction quality. For historic reasons, we use a quadratic definition of the signal-to-noise ratio as a measure of the pulse strength:

$$\text{SNR}_{\text{quad}} = \frac{(\max \hat{E}_{\text{mag}})^2}{\text{rms}_{\text{noise}}^2}, \quad (5.3)$$

where \hat{E}_{mag} is the maximum value of the Hilbert envelope of the magnitude of the electric field projected on the horizontal plane, and $\text{rms}_{\text{noise}}$ the root mean square of the noise calculated in a different region of the same electric field trace. Because a projection of the electric field is used, the numerical values of SNR_{quad} are difficult to compare with SNR_{peak} used in the context of the scintillator-triggered data. The maximum of the time trace is calculated in the signal window defined above, and the noise rms is calculated in a separate window consisting of the first 25 μs of the trace. To derive the strength of the cut on SNR_{quad} we need to create a very pure data set by selecting only events in which we have used strong pulses ($\text{SNR}_{\text{quad}} > 20.0$) in the reconstruction and demanded a small opening angle Ω_{SD} between the reconstructed radio shower axis and the shower axis determined by SD of 7.0° . The events in this pure data set contain three or more stations with SNR_{quad} above 20.0, the minimum number for which a directional reconstruction of the air shower can successfully be performed. They also contain all other active stations in the array which have a lower signal strength, and accordingly were not used. We then perform a series of directional radio reconstructions on this data set, where we change the lower limit of SNR_{quad} , and investigate the influence of the limit on the reconstruction quality. This is quantified by the time residuals of the pulse time of the individual stations with respect to the reconstructed radio wave front. The directional reconstruction of the radio data is performed using the peak time of the Hilbert envelope and assuming a plane radio wave front.

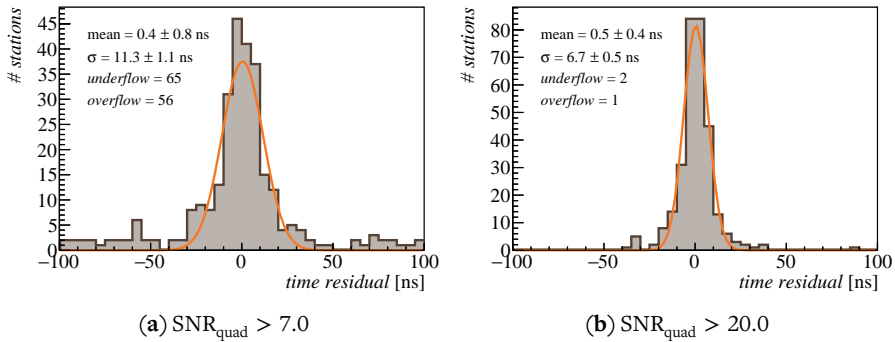


Figure 5.15: Time residuals of a plane wave fit to 55 events from the externally triggered data set from 1 April 2014 to 11 April 2014.

Two distributions of the time residuals of the pure data set derived from data recorded between 1 April 2014 and 11 April 2014 are displayed in figure 5.15. A high SNR-cut results in a distribution with a Gaussian shape (caused by timing uncertainties and deviations from the plane wave front) with very few outliers. When a lower SNR-cut is applied, the distribution has many residuals deviating from the Gaussian, which is caused by uncorrelated noise pulses wrongly identified as signal pulses and mis-reconstructed showers. At the same time, the distribution gets broader because more signal stations that are typically further away from the axis and thus deviate more from the plane shower front are included. We quantify this behavior by fitting a Gaussian to this distribution, and look at the fraction of stations with residuals outside three standard deviations σ of this fit. The resulting fractions are plotted in figure 5.16. The bulk of stations with signals unassociated to the air shower are rejected above $\text{SNR}_{\text{quad}} = 9.0$. Between 9.0 and 15.5, fluctuations in the fraction of residuals within 3σ are observed, which are caused by stations with higher unassociated signals that are rejected one by one from the events. At $\text{SNR}_{\text{quad}} > 15.5$ the fraction appears to stabilize, so we choose this as our SNR-cut.

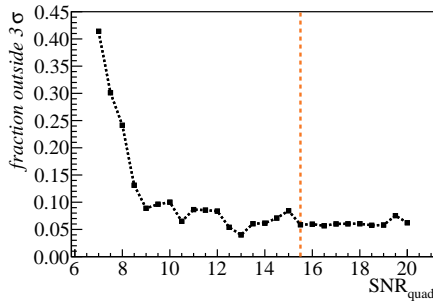


Figure 5.16: Fraction of stations with time residuals outside of 3σ of a Gaussian fit through the distribution of time residual distributions.

With this SNR-cut, we perform a full directional reconstruction of all externally triggered data to select events and stations with air shower induced radio pulses, for which we need at least three stations with a signal. We include AERA-I stations and saturated AERA-II stations in order to increase the probability to achieve a good directional reconstruction, but these will not be used in any further analysis. The analog response and the antenna gain is evaluated, using the direction from the SD reconstruction. To lower the noise level and increase the signal-to-noise ratio, we filter narrowband noise before we calculate the peak signal. We therefore reject the frequency bands from table 5.2 and add a filter ranging from 30.0 MHz to 40.0 MHz, setting the amplitudes in all frequency bins in these ranges to zero. When the distribution of the distance from the station to the SD shower axis of all signal stations in an event contains gaps of more than 1000 m, the stations farther from the SD shower axis than this gap will also be rejected. This is mostly to reject events in which the shower core is removed far from the radio array,

and therefore produce no measurable air shower induced radio signals, but might contain noise pulses. As a final step we reject stations that measure pulse trains, as these are usually a good indicator of noise contamination in the trace. These are rejected by scanning the entire trace outside the signal region on additional pulses that have an amplitude above the SNR-cut.

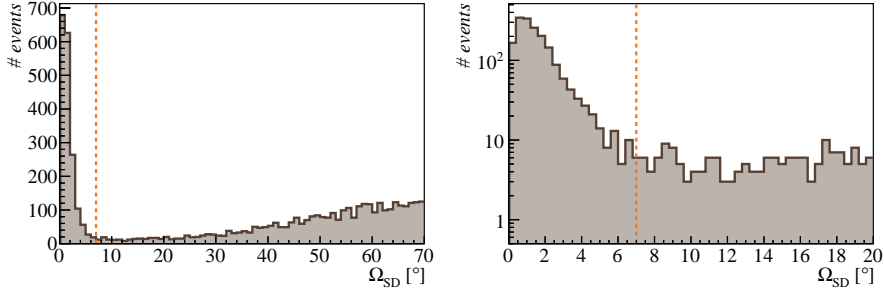


Figure 5.17: Distribution of the opening angle Ω_{SD} between the reconstructed shower axis from radio data and the axis determined by sd, calculated on all externally triggered events with three or more signal stations. *Left:* overview which shows the rising distribution at high opening angles which can be expected from uncorrelated events; *right:* close-up of the region with low opening angles. The dashed line indicates the opening angle cut of 7.0° .

Using these reconstructed events we now attempt to establish a cut on the opening angle Ω_{SD} between the direction reconstructed by a plane wave fit to the radio data and the direction reconstructed by using sd data. The resulting distribution of opening angles is plotted in figure 5.17. At small opening angles, we expect that only air shower induced radio pulses were used in the reconstruction, while at high opening angles noise pulses are fitted. Based on the distribution, we will place the opening angle cut on $\Omega_{SD} < 7.0^\circ$.

5.4.3 Additional selection criteria

Pulse window

We have now reduced the data of both experimental setups to a data set containing events with likely air shower induced radio pulses. Before we can extract these pulses for physics analysis, we need to determine the time window that contains the radio pulse. In figure 5.18, the average Hilbert envelope $\langle \hat{E}_{\mathbf{v} \times \mathbf{B}} \rangle$ of the $\mathbf{v} \times \mathbf{B}$ component of the electric field relative to the maximum value of the Hilbert envelope at t_{peak} is plotted for both data sets. The average contains all stations and events that pass the selection procedure outlined in this section, with $\text{SNR}_{\text{peak}} > 7.0$. The average pulse shapes measured by the two different experimental setups align remarkably well, apart from the

bias to higher signals for the externally triggered setup, which is caused by our selection method. Based on this average pulse, we decide that a pulse window of 400 ns around the peak value will be sufficient to contain the features of the pulse, while minimizing the amount of noise in the window.

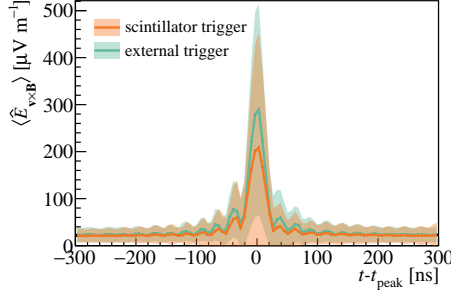


Figure 5.18: Average Hilbert envelope of the pulse of all selected stations in the complete data set relative to the time of pulse maximum, t_{peak} .

Polarization cut

An extra selection criterion we can apply to both data sets is one based on the polarization angle ϕ_p , or to be more precise the angle of the total electric field vector with the $\mathbf{v} \times \mathbf{B}$ -axis in the shower reference frame, as can be seen in figure 5.19. We can compare the measured polarization angle $\phi_{p,\text{measured}}$ with the expected polarization angle $\phi_{p,\text{expected}}$ based on the position of the station relative to the reconstructed shower geometry from SD. This selection criterion has the advantage that we will not only be able to discriminate between air shower induced pulses and noise pulses, which are randomly polarized, but can also discard events that have a mis-reconstructed SD geometry.

The direction of the total electric field vector depends on the relative amounts of geomagnetic and charge-excess emission in an event. This can be expressed by using the geomagnetic angle α and equation (3.8). If we consider the shower reference frame, the two principal components of the electric field $E_{\mathbf{v} \times \mathbf{B}}$ and $E_{\mathbf{v} \times (\mathbf{v} \times \mathbf{B})}$ can be decomposed into E_G and E_C at the position of the antenna (see also figure 5.19):

$$-E_{\mathbf{v} \times \mathbf{B}} = E_G + E_C \cdot \cos \Phi_{\text{obs}}, \quad (5.4)$$

$$-E_{\mathbf{v} \times (\mathbf{v} \times \mathbf{B})} = E_C \cdot \sin \Phi_{\text{obs}}, \quad (5.5)$$

where Φ_{obs} is the observer angle which describes the position of the antenna relative to the $\mathbf{v} \times \mathbf{B}$ -axis, as was introduced in section 3.2. Using equations (3.8), (5.4) and (5.5)

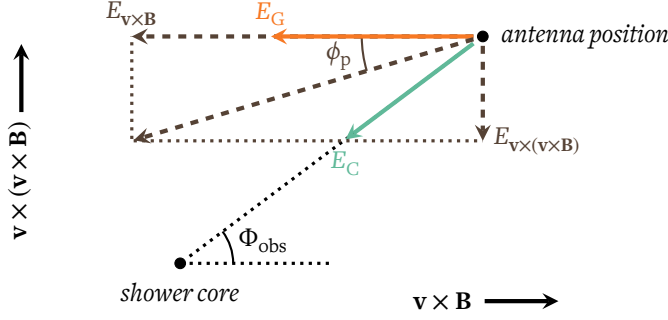


Figure 5.19: Definition of all used vector components of the electric field in the shower reference frame. The observer angle Φ_{obs} is the angle with the $\mathbf{v} \times \mathbf{B}$ -axis defined by the antenna position, while the polarization angle ϕ_p is the angle between the total electric field vector and the $\mathbf{v} \times \mathbf{B}$ -axis. The total electric field is the sum of the field resulting from the geomagnetic emission (E_G) and the field resulting from the charge-excess emission (E_C). The direction of movement of the shower \mathbf{v} is perpendicular to the figure plane and pointing inwards. Note that the relative amount of charge-excess emission is exaggerated.

it is now straightforward to express the expected polarization angle:

$$\phi_{p, \text{expected}} = \arctan \left(\frac{E_{\mathbf{v} \times (\mathbf{v} \times \mathbf{B})}}{E_{\mathbf{v} \times \mathbf{B}}} \right) = \arctan \left(\frac{\sin \Phi_{\text{obs}}}{\frac{\sin \alpha}{a} + \cos \Phi_{\text{obs}}} \right). \quad (5.6)$$

The average charge-excess fraction a at AERA was measured to be 0.14 ± 0.02 [42, 68], and we will use this value to calculate the expected polarization.

Following the approach from [68], the measured polarization angle of the electric field is calculated using the *Stokes parameters*. In the shower reference frame, the relevant Stokes parameters Q and U are defined as [112]:

$$Q = \frac{1}{n} \sum_{i=1}^n (E_{\mathbf{v} \times \mathbf{B}, i}^2 + \hat{E}_{\mathbf{v} \times \mathbf{B}, i}^2 - E_{\mathbf{v} \times (\mathbf{v} \times \mathbf{B}), i}^2 - \hat{E}_{\mathbf{v} \times (\mathbf{v} \times \mathbf{B}), i}^2), \quad (5.7)$$

$$U = \frac{2}{n} \sum_{i=1}^n (E_{\mathbf{v} \times \mathbf{B}, i} E_{\mathbf{v} \times (\mathbf{v} \times \mathbf{B}), i} + \hat{E}_{\mathbf{v} \times \mathbf{B}, i} \hat{E}_{\mathbf{v} \times (\mathbf{v} \times \mathbf{B}), i}), \quad (5.8)$$

where we sum over n time bins in the electric field traces $E_{\mathbf{v} \times \mathbf{B}}$ and $E_{\mathbf{v} \times (\mathbf{v} \times \mathbf{B})}$ and their Hilbert transforms $\hat{E}_{\mathbf{v} \times \mathbf{B}}$ and $\hat{E}_{\mathbf{v} \times (\mathbf{v} \times \mathbf{B})}$. In the presence of a noise background, the Q

parameter must be corrected to prevent biases:

$$Q \approx \frac{1}{n} \sum_{i=1}^n (E_{\mathbf{v} \times \mathbf{B}, i}^2 - \overline{B_{\mathbf{v} \times \mathbf{B}}^2} + \hat{E}_{\mathbf{v} \times \mathbf{B}}^2 - \overline{\hat{B}_{\mathbf{v} \times \mathbf{B}}^2} - E_{\mathbf{v} \times (\mathbf{v} \times \mathbf{B})}^2 + \overline{B_{\mathbf{v} \times (\mathbf{v} \times \mathbf{B})}^2} - \hat{E}_{\mathbf{v} \times (\mathbf{v} \times \mathbf{B})}^2 + \overline{\hat{B}_{\mathbf{v} \times (\mathbf{v} \times \mathbf{B})}^2}), \quad (5.9)$$

where $\overline{B_j^2}$ is the average value of B_j^2 calculated in a time window outside the signal window of the j component of the electric field trace, of which \hat{B}_j is the Hilbert transform. It is now possible to calculate the measured polarization angle:

$$\phi_{p, \text{measured}} = \frac{1}{2} \arctan \left(\frac{U}{Q} \right). \quad (5.10)$$

We define a selection criterion based on the polarization offset β , describing the difference between the expected and measured polarization angles:

$$\beta = |\phi_{p, \text{expected}} - \phi_{p, \text{measured}}|. \quad (5.11)$$

To determine the value of the cut on β , we calculate the expected and measured polarizations for a set of events. The distributions of β for all events in the combined data set from both experimental setups passing our selection with zenith angle $\theta < 60^\circ$ are plotted in figure 5.20 for two different SNR-cuts. These events have all been fully calibrated and cleaned (see chapter 6). Based on these distributions, we set the cut on the polarization offset β at 25° .

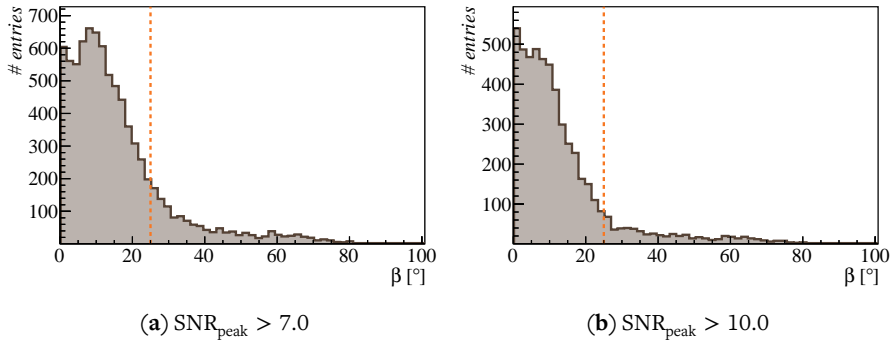


Figure 5.20: Distribution of the difference between the expected and measured polarization angle β for two different SNR-cuts. It contains all events that pass the selection criteria described in this chapter with $\theta < 60^\circ$. The distribution in (b) clearly shows the two underlying distributions on which we based our selection criterion. The dashed line indicates the selection cut of 25° .

Zenith angle cut

We will only consider events with a zenith angle θ lower than 60° . The reason for this is twofold. Firstly, the three-dimensional electric field reconstruction from the measurement with two antenna arms becomes uncertain above this angle because of the mixing of noise, as was demonstrated in section 4.3.4. This will introduce artificial noise and biases in reconstructed parameters such as the pulse power and the spectral index (see figure 4.15). Additionally, the SD reconstruction of events of the infill array also becomes uncertain at high zenith angles, which might result in a mis-reconstruction of the shower geometry and energy.

5.4.4 Summary of radio data selection

In this chapter we have described a method to reject dysfunctional stations from the data set, and outlined two approaches to reduce the data into sets that are usable for physics analysis. We can now get this machinery to work on our full available data set.

An overview of how the events were selected from the complete scintillator-triggered data set is listed in table 5.3. The final set contains radio events that are in coincidence with SD events, and have a shower geometry that can be reconstructed using the SD information, with a zenith angle reconstructed by the SD that is less than 60° . It should be noted that the complete data set also includes many experimental runs, and therefore contains a huge total number of events.

Table 5.3: Selected AERA-II events from the scintillator-triggered setup

full data set	27 323 100
SD coincidence	38 525
SD reconstruction and $\theta_{SD} < 60^\circ$	30 274

A similar overview, but then for the full externally triggered data set, is listed in table 5.4. Unlike the scintillator-triggered events, these have been compared to SD parameters immediately to determine if they contain air shower induced radio pulses. We went from a complete data set, which by construction is in coincidence with SD events, to a smaller set of which the SD shower geometry could be reconstructed. After this, we have selected events for which the radio wave plane could be reconstructed, which requires at least three stations to fulfill the SNR criterion determined in section 5.4.2. The radio geometry from these events could then be compared to the shower geometry from SD to determine the opening angle Ω_{SD} .

From here, we can reduce both data sets further. The scintillator-triggered events still contain stations that do not pass our quality criteria in the sense that they are flagged as dysfunctional or that have radio-triggers without scintillator peak. In addition, the vast

Table 5.4: Selected AERA-II events from the externally triggered setup

full data set	10 756 134
SD reconstruction	865 701
$n(\text{SNR}_{\text{quad}} > 15.5) \geq 3$	9472
$\Omega_{\text{SD}} < 7.0^\circ$	1818

majority of the scintillator-triggered data set still consists of stations without a radio pulse above the noise level. In table 5.5, these reductions are listed as total number of candidate stations for each step. We have also made additional selections based on the SD zenith angle (redundant for the scintillator-triggered data), signal-to-noise SNR_{peak} (redundant for the externally triggered data) and the polarization offset β .

Table 5.5: AERA-II candidate events and stations after the initial selection

	scintillator-triggered		externally triggered	
	<i>events</i>	<i>stations</i>	<i>events</i>	<i>stations</i>
initial selection	30 247	122 412	1818	5375
pass quality criteria	23 847	91 906	– – –	– – –
$\theta_{\text{SD}} < 60^\circ$	– – –	– – –	1215	4660
$\text{SNR}_{\text{peak}} > 7.0$	2862	3507	– – –	– – –
$\beta < 25^\circ$	2429	2904	1114	3816

The selection procedure has resulted in a total data set of 3543 selected events containing 6796 signal stations that can be used for physics analysis. The selection of the scintillator-triggered data has produced mostly single station events, while the selection of the externally triggered data has produced mostly events with three stations. This is also reflected in the ratio of stations of events selected from both experimental setups. The distribution of the arrival direction of all signal stations is plotted in figure 5.21, where a clear excess from the south is visible. This excess is expected because for these showers, the arrival direction is perpendicular to the geomagnetic field and therefore the magnitude of the $E_{\mathbf{v} \times \mathbf{B}}$ vector is the largest. Of the selected signal stations, 552 out of 6796 are recorded during thunderstorm conditions, and for 935 signal stations there is no weather station data available. It is interesting to note that of the events that fall outside the polarization cut $\beta < 25^\circ$, a much higher fraction is recorded during thunderstorm conditions: 384 out of a total of 1447 stations. This is in line with the expectation that enhanced electric fields not only change the strength of the air shower induced radio pulse, but also influence the polarization signature (see e.g. [108]).

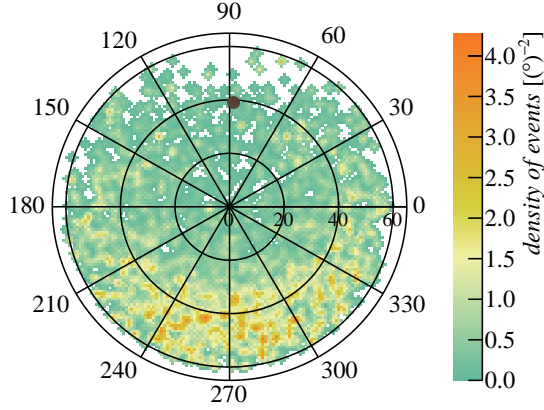


Figure 5.21: Distribution of the arrival direction of all signal stations. The large dot indicates the direction of the magnetic field vector \mathbf{B} . Azimuth $\phi = 0^\circ$ is due east.

5.4.5 Selection of fluorescence data

In order to compare parameters extracted from the radio signal, such as the spectral index, with the development of the air shower, it is important to create a subset of radio events that is in coincidence with FD measurements. To ensure a qualitatively good reconstruction of shower maximum from the fluorescence measurement we adopt several quality cuts from [76]:

- Depth of shower maximum X_{\max} must be observed by at least one of the FD telescopes.
- The uncertainty of X_{\max} determined by the fit of the Gaisser-Hillas function must be less than 40.0 g cm^{-2} .
- The minimum viewing angle α_{\min} must be larger than 20.0° to limit the contribution to the profile of Cherenkov light produced in the shower (see e.g. [113]).

In addition to these quality cuts, fiducial cuts are applied to the events to prevent biases in the X_{\max} distribution in [76]. These do not concern us because we want to study event-by-event correlations between the radio pulse and FD measurements.

One other important quality cut we have disregarded so far is the cut on the vertical aerosol optical depth (VAOD). According to [76], this should be less than 0.1 when measured from the ground to 3 km, and events without an aerosol measurement should be completely rejected. The database containing VAOD measurements is however only irregularly updated, and at the time of writing only contained data until the end of 2013.

Applying the strict quality cuts would therefore disqualify almost our entire data set for physics analysis.

To investigate the effects of the usage of the `VAOD` database on the shower profile reconstruction, we look at the reconstructed value of X_{\max} in the set for which we do have a filled `VAOD` database at our disposal. These are produced by the standard `HEAT-Coihueco` reconstruction of *Offline*. We reconstruct the shower profile one time with the Mie scattering calculated with the `VAOD` value from the database applied, $X_{\max, \text{db}}$, and one time using an average aerosol model, $X_{\max, \text{model}}$. In figure 5.22, the distribution of the differences between the two reconstructed values of X_{\max} is plotted. The plot contains all `FD` events that are coincident with scintillator-triggered radio events from runs 513 (June 2013), 519 (August 2013) and 526 (October 2013), and survive the three `FD` selection criteria listed above. No selection cuts on the radio events were applied. The resulting distribution has an rms of 6.17 g cm^{-2} , which is smaller than the typical uncertainty of X_{\max} per event. This might be a selection effect, because the showers measured by `AERA` are all relatively close to the `FD` telescopes. We conclude that in the case of `AERA` the absence of aerosol measurements does not influence the reconstruction of X_{\max} significantly, and we can also include events for which only an average aerosol model is available in our analysis.

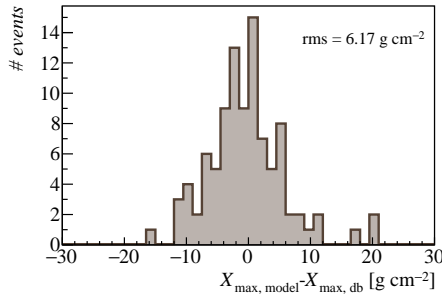


Figure 5.22: Comparison between reconstructed X_{\max} values where the aerosol measurement database was used, and where an average model was used. The width of the distribution is smaller than the typical uncertainty of X_{\max} per event.

An overview of the number of `FD` coincidences with the signal stations selected in the previous section is listed in table 5.6. The fraction of the total of selected scintillator-triggered stations that is coincident with `FD` of 7.6 % is not far from the maximum achievable uptime of the `FD` of 13 %. The fraction of the total of selected externally triggered stations with `FD` is already quite a bit lower. This is probably due to the power saving measures employed by the externally triggered stations, which shuts them down during periods of the night in Argentinian winter, which coincides with the darkest pe-

riods of the year and therefore also the periods in which most FD events are recorded. When we apply the FD quality cuts about half of the scintillator-triggered coincident stations survive. When applied to the externally triggered stations however, we see the number of stations reduced to about one fifth. This is because the externally triggered setup is located too close to the FD building to efficiently measure air showers which have X_{max} within the field of view of the telescopes. In the remainder of this thesis, we will refer to the combined data set from both experimental radio setups with coincident FD measurements of shower maximum after applying the FD selection criteria as the RD-FD data set.

Table 5.6: Number of stations coincident with FD in the selected data set and the fraction relative to the total number of selected signal stations, disregarding weather conditions. The selected set is obtained after applying the three FD selection criteria.

	all FD	<i>fraction</i>	selected FD	<i>fraction</i>
scintillator trigger	223	7.6 %	111	3.8 %
external trigger	186	4.9 %	35	0.9 %
total	409		146	

None of the 146 selected stations in the RD-FD data set are flagged as thunderstorm events by the procedure outlined in section 5.3. This was to be expected because the FD utilizes its own lightning rejection algorithm. A significant number of 34 stations in the RD-FD data set were triggered in periods when the weather station was down. Since externally triggered events are disproportionally represented in this number, we chose to retain these events in the following analysis in order not to further decrease the number of multi-station events. We assume that the lightning rejection of the FD and the selection cut made on the difference between the expected and measured polarization of the electric field are sufficient to prevent significant contamination of the data set by events recorded during thunderstorm conditions.

Calibration and cleaning of AERA data

AS OUTLINED IN SECTION 4.3.4, some hardware components of the radio detector stations are measured individually in the lab to be able to correct for their response. This includes the LNAs, the filter-amplifiers and the cables. The antenna, however, has a single universally applied response pattern, which does not take into account variations between the stations. These need to be calibrated with the detector in the field using the radio background or using test emitters. Attempts at a calibration with test emitters have been made, but so far have not led to a reliable measurement of the antenna response pattern. In this chapter, we will present two calibration techniques based on the galactic radio background: a relative calibration to correct for variations between stations, and an absolute calibration to correct for deviations in the simulated antenna gain pattern. Finally, we will introduce a method for cleaning the data by removing narrowband RFI in the time domain.

6.1 Relative amplitude calibration of AERA-II stations

In [61], a method has been developed to perform a frequency-dependent relative amplitude calibration for AERA-I stations by using the background emission from the galaxy. This calibration can be used to correct the gains of the individual stations as a function of frequency, but the group delays are outside of its scope, as we would need a pulsed calibration source for this.

We perform the same calibration for the AERA-II data set. In addition to the aforementioned study on AERA-I data, we also investigate the time evolution of the calibration constants and improve on the calibration by introducing and testing an interpolation method. The calibration was also adapted for integration in the *Offline* software framework.

6.1.1 Method: calculation of calibration constants

We assume that all stations are exposed to the same frequency-dependent radio emission from the galaxy at any given instance of time. The galactic background has a variation with a period of 23.934 h, which can be expressed in terms of *local sidereal time* (LST). We can use the periodically triggered data to study the average frequency spectra measured by the stations as a function of LST.

By comparing the spectra of an individual station i with the average of all stations as a function of LST it is possible to determine frequency-dependent calibration constants $c_{i,v}$. For this, we can simply divide the amplitude $A_i(v, t_k)$ of a single station at frequency v and LST t_k by the average amplitude of all stations $\bar{A}(v, t_k)$, and then take the average of all LST bins k :

$$c_{i,v} = \frac{1}{m} \sum_{k=1}^m \frac{A_i(v, t_k)}{\bar{A}(v, t_k)}, \quad (6.1)$$

where m is the total number of LST bins. The resulting calibration constants indicate the relative frequency-dependent gain offsets of the stations, and the inverse of them can be applied to the spectra to correct them. Because of differences in exposure to the galaxy, the asymmetry of the antenna sensitivity and influence of man-made noise sources, this must be done for both antenna channels independently. Examples of the measured average frequency spectra as a function of LST are visible in figure 6.1.

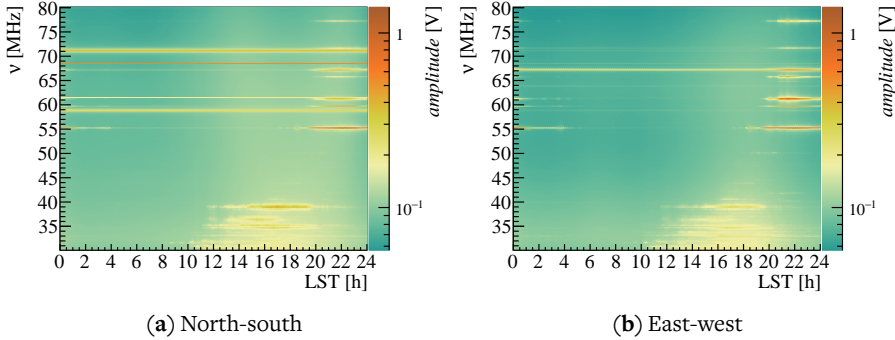


Figure 6.1: Dynamic average frequency spectrum as a function of LST in both antenna channels, for the scintillator-triggered setup in the period from 20 September 2013 to 22 October 2013 (run 526). The spectra were averaged over periods of 10 minutes in LST. Note that the variations in the spectra are dominated by narrowband noise sources.

It is difficult to distinguish the galactic variation in the plots because of various noise contributions to the signal. As was demonstrated in section 5.1, the noise sources are typically nearby. Because their intensity falls off with the distance to the source, their

contribution is different at each station. Therefore, they introduce biases in the calibration and need to be excluded.

6.1.2 Exclusion of broadband RFI

We can exclude high broadband activity by rejecting traces with an average spectral density above a certain threshold. We define the average spectral density $\langle S \rangle$ as:

$$\langle S \rangle = \frac{1}{n} \sqrt{\sum_{j=1}^n A_j^2}, \quad (6.2)$$

where A_j is the amplitude of frequency bin j , and n the total number of frequency bins. The average spectral density as a function of LST for both antenna channels is plotted in the histograms of figure 6.2. The galactic variation is clearly visible in the baseline average, with broadband pulses sticking out above it.

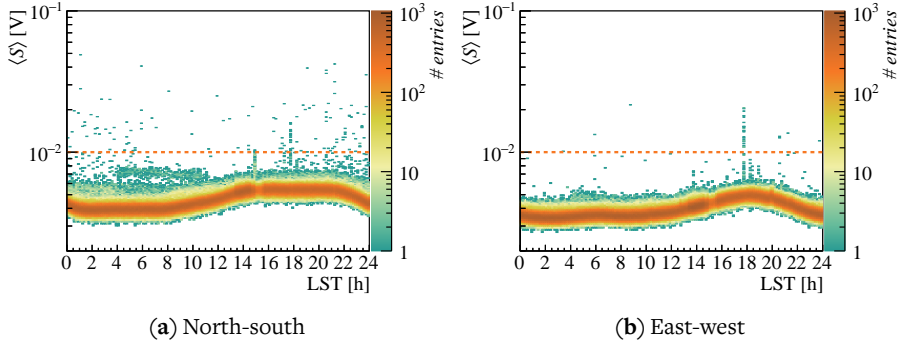


Figure 6.2: Average spectral density of all stations as a function of LST, for data from the scintillator-triggered setup of the period from 6 September 2013 to 13 September 2013. The density limit of 0.01 V is marked with a dashed line.

Instead of a dynamic maximum depending on the average of $\langle S \rangle$ as a function of LST – as was done in [61] – we will use a fixed limit of 0.01 V. This is computationally much more efficient with comparable discrimination power, but it has a possible LST-dependent bias.

6.1.3 Exclusion of narrowband RFI

Narrowband RFI can easily be recognized as horizontal lines in the dynamic frequency spectra of figure 6.1. They are either spanning the entire LST-range, or are more temporal and related to periods with human activity. The former is the case for continuous radio transmitters and the AERA beacon, while the latter are short-term bursts, mostly

in the lower frequency range. The frequency bands of these emitters need to be manually flagged and removed from the calibration data. They can be identified in the long-term average spectra as peaks on top of an otherwise smooth noise floor, as is visible in figure 6.3. This figure also illustrates the frequency bands that are set to zero.

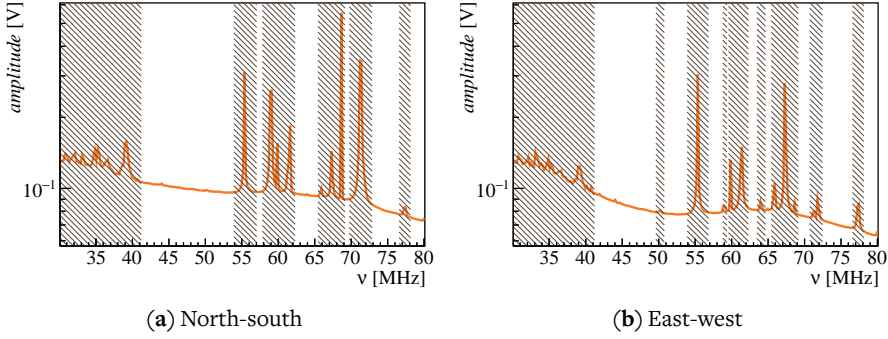


Figure 6.3: Measured noise spectra in the north-south and east-west antenna channels averaged over all stations in the scintillator-triggered setup and all times in the period from 20 September 2013 to 22 October 2013 (the calibration data set). The hatched areas indicate rejected frequency bands.

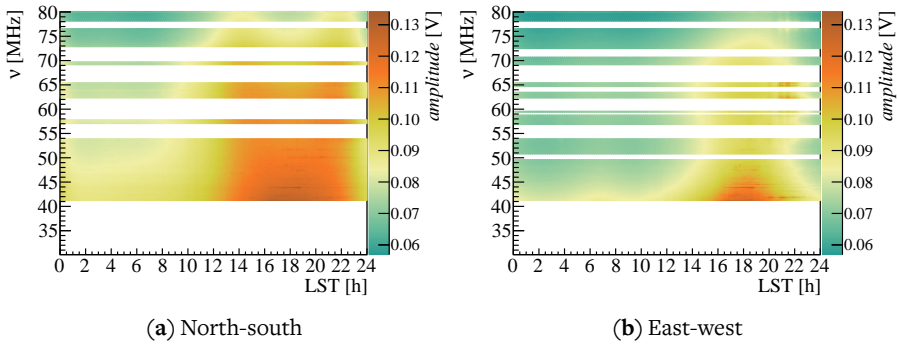


Figure 6.4: Dynamic average frequency spectrum as a function of LST of both antenna channels, for the period from 20 September 2013 to 22 October 2013. Broad-band and narrowband RFI were removed. The periodic galactic imprint is clearly visible.

Because of the dense forest of emitters in the region between 30.0 MHz and 41.0 MHz, we will reject this region in its entirety. Whereas [61] had identical rejected frequency bands in both channels, we have made the bands channel-specific in order to maximize the number of usable frequency bins. When the frequencies with narrowband noise are removed from the spectra, the periodic signal of the galaxy as a function of LST becomes

much more pronounced, as is visible in the cleaned dynamic spectra of figure 6.4.

6.1.4 Interpolation

This method of RFI rejection leaves gaps in the frequency-dependent calibration constants, which renders the data in these parts of the spectrum unusable for further analysis. To solve this, we choose to interpolate these gaps. Large features in the frequency-dependent gain patterns which deviate more than 1 % from the mean usually extend over many frequency bins. Therefore, the patterns are assumed to be more or less smooth, and we can attempt to fill up the gaps left by the narrowband emitters with a linear interpolation. In figure 6.5, the frequency-dependent gain patterns for four channels are plotted, including the linear interpolation of the gaps.

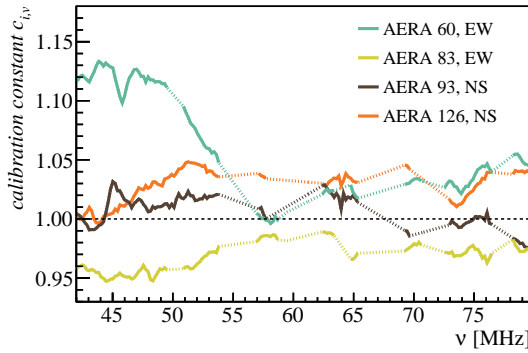


Figure 6.5: Calibration constants as a function of frequency, with in dotted lines the interpolated gaps. Four channels have been selected from the data set from 20 September 2013 to 22 October 2013.

Because of the constant presence of most of the narrowband noise disturbances in the full data set, it is impossible to validate this method on existing gaps left by narrowband transmitters. We can, however, remove chunks of an arbitrary size from the relatively wide noise-free band from 41.0 MHz to 54.5 MHz in the north-south polarization arm (see figure 6.3(a)) and interpolate these gaps. The clean frequency band has a width of 65 bins, and we will use a sliding window of a certain number of bins over this entire frequency range in which we will interpolate. We can then compare the interpolated constants $c_{i,v,\text{inter}}$ of these artificial gaps with the original calibration constants $c_{i,v,\text{real}}$. In the left panel of figure 6.6, the distribution of the difference $c_{i,v,\text{real}} - c_{i,v,\text{inter}}$ is plotted for a gap size of 10 bins (≈ 2.0 MHz), sliding over the clean band in all active stations in the data set of the scintillator-triggered setup of 20 September 2013 to 22 October 2013. The standard deviation of this distribution is 2.55 % with respect to the average, which is comparable with the precision of the calibration method itself, as will be shown in

the next section. The standard deviations of the distributions of the differences for interpolated gaps of up to 30 bins are plotted in the right panel of figure 6.6. As long as the gap does not become too big, the linear interpolation has an accuracy that is still well below the typical initial distribution of offsets between stations before calibration, as we will also see in the next section.

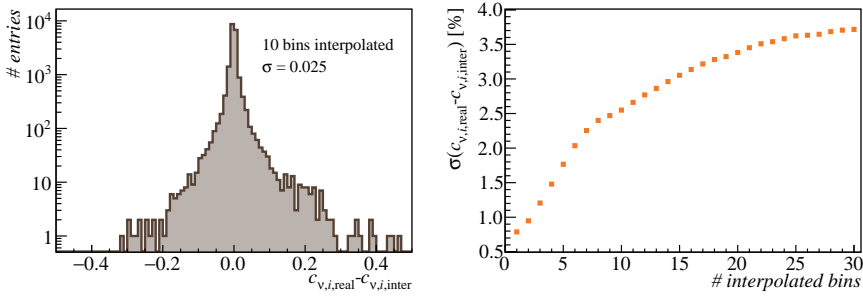


Figure 6.6: Difference between interpolated calibration constants and the real calibration constants. *Left:* distribution of absolute differences between the real and interpolated calibration constants for a gap size of 10 bins; *right:* standard deviation of the distribution as a function of gap size.

6.1.5 Application to AERA-II data

In order to see if the calibration constants calculated within one period of AERA-II data can be applied to all other periods, we can look at the time-evolution of the frequency-dependent calibration constants. This has been done for all data taken with the scintillator-triggered setup between May 2013 until September 2014, with calibration constants calculated for every 10-day period. The calibration constants of two typical stations are displayed in figure 6.7.

The main frequency-dependent features in both channels appear to remain constant over the whole period of data taking. Variations in the intensity of the features can be observed, but these are most likely due to changes in the underlying average as the collection of active stations constantly changes over time. Therefore, we conclude that it is possible to apply the calibration constants calculated within one period of time to the entire data set with good results. A remarkable feature of station AERA 44 is the increased amplitude around 43 MHz in the full plotted period. This enhancement is most likely because of defective hardware, and similar features can be observed in some other stations. We will continue to calibrate stations with significant enhancements in their average spectra, but ignore them when calculating the average spectral amplitudes $\bar{A}(\nu, t_k)$ to prevent biases. For the scintillator-triggered setup the stations that are excluded from the average are: AERA 44, AERA 85, AERA 100, AERA 101, AERA 102, AERA 113 and AERA 124.

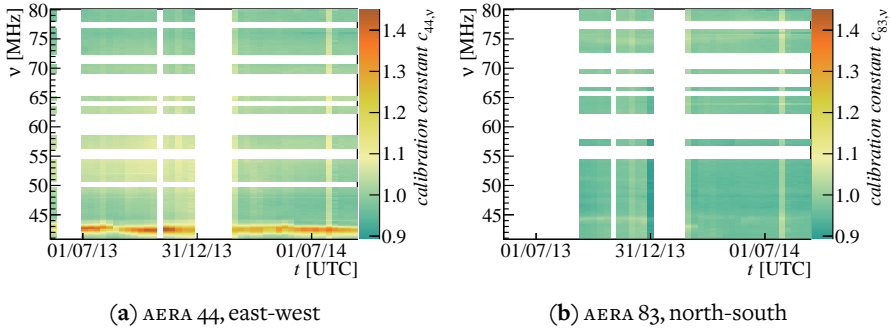


Figure 6.7: Frequency-dependent calibration constants as a function of time in bins of 10 days. The empty periods in the time domain are periods in which the station was not active or flagged as bad (according to the criteria from section 5.2).

We now need to specify a period that we can use as a *calibration data set*. A period with a large number of active stations will help preventing biases in the average spectrum as function of LST, and it will also make the calculated calibration constants applicable to most of the stations. According to section 5.2, the period from 20 September 2013 to 22 October 2013 (run 525/526) has the highest number of active stations of the data set of the scintillator-triggered setup, so we will use it to calculate the calibration constants and apply them to the rest of the data. If we first correct the data for the response of the analog chain as was measured in the lab, we get the distributions for $c_{i,v}$ that are plotted in figure 6.8. The conjunct distribution of the north-south and east-west channels has a standard deviation σ of 5.66 %.

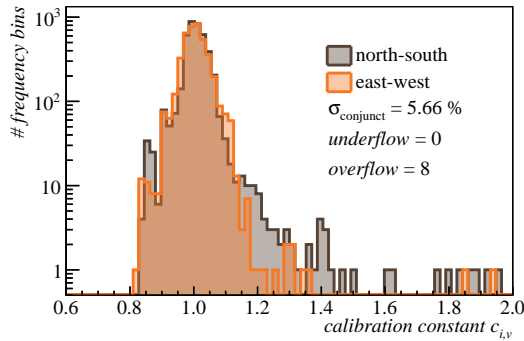


Figure 6.8: Calibration constants for all stations in both antenna channels of the scintillator-triggered setup from 20 September 2013 to 22 October 2013, the calibration data set.

We validate the method and the calibration constants by applying them to a validation data set in another period, namely from 2 June 2014 to 6 July 2014 (run 550). The average noise spectra of this period, together with the flagged frequency bands, are plotted in figure 6.9. The noise circumstances in this validation data set are slightly different from the calibration data set, which enables us to gain some insight in the performance of the interpolated calibration constants. In total there are 15 interpolated frequency bins (out of a total of 256 bins) in the north-south antenna channel and 24 interpolated frequency bins in the east-west antenna channel that can be probed in the validation data set.

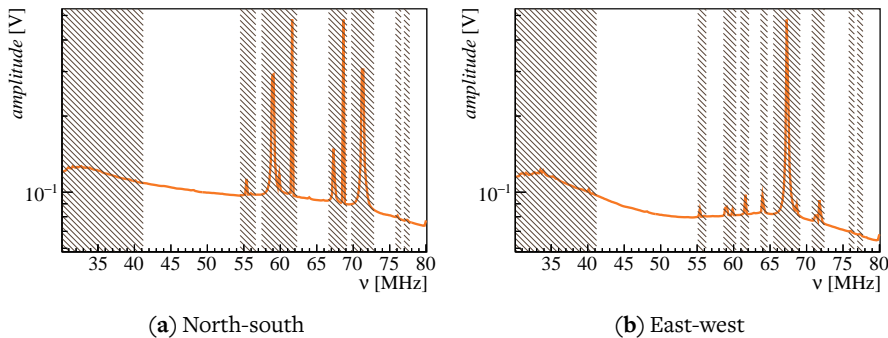


Figure 6.9: Measured noise spectra in the north-south and east-aest antenna channels averaged over all stations of the scintillator-triggered setup and all times in the period from 2 June 2014 to 6 July 2014 (the validation data set). The hatched areas indicate rejected frequency bands.

First, we have calculated the calibration constants for the validation data set, so we know what the initial distribution of the offsets between the stations was. This distribution is displayed in figure 6.10(a), and has a standard deviation of 5.21 %. After this, we have applied the calibration constants of the calibration data set to the validation data set, and again calculated the calibration constants. This was done with and without interpolation of the gaps left by the narrowband transmitters, and the resulting distribution is visible in figure 6.10(b).

After applying the calibration, the standard deviation of the distribution has reduced to 2.37 %. When the interpolation is applied, the total number of bins naturally increases, and the standard deviation of 2.34 % is very much comparable with the uninterpolated calibration. From this, we conclude that the galactic calibration reduces the spread of the offsets between stations by more than a factor of 2. The interpolation of gaps in the frequency-dependent calibration constants does not seem to negatively influence the calibration performance, and comes with the big advantage that the full frequency spectrum is available for further analysis.

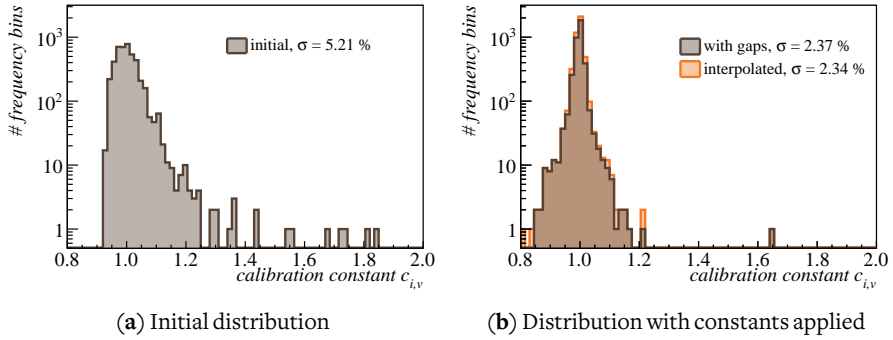


Figure 6.10: Improvement of the offsets between stations in the validation data set when the calibration constants are applied.

We can apply the same analysis to calibrate the stations of the externally triggered setup. For this we have selected a calibration data set running from 1 July 2014 to 1 August 2014. Just like for the scintillator-triggered setup, several stations with enhanced spectral regions were ignored when calculating the average spectral amplitudes $\hat{A}(\nu, t_k)$ to prevent biases, namely AERA 42, AERA 45, AERA 57, AERA 61, AERA 64, AERA 66 and AERA 108. The resulting calibration constants $c_{i,v}$ are plotted in figure 6.11. The distribution is very similar to that from the scintillator-triggered setup in figure 6.8. There are a few more frequency bins with high offsets, which are probably due to contamination from narrowband sources, which are less distant to the stations from the externally triggered setup and therefore have a stronger influence on the variation between stations. The frequency bins with a calibration constant below ~ 0.8 are caused by a single station (AERA 41), which has a relatively low gain in a broad range of the frequency spectrum.

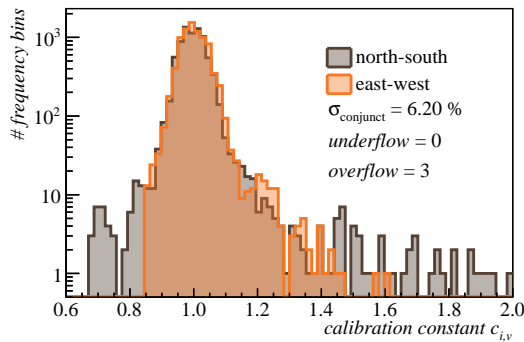


Figure 6.11: Calibration constants for all stations in both antenna channels of the externally triggered setup from 1 July 2014 to 1 August 2014.

6.1.6 Integration into the software framework

Two modules have been created to integrate the relative galactic calibration in the *Offline* framework: the `RdChannelGalacticConstantsGenerator`, which can be used to calculate the calibration constants from any given data set, and the `RdChannelGalacticBackgroundCalibrator`, which is made to apply calibration constants to data within an analysis pipeline. Configurable options include a list of stations that need to be excluded when calculating the average $\bar{A}(\nu, t_k)$, and which average spectral density threshold to use for broadband noise filtering. When applying the calibration constants, a linear interpolation to fill gaps in the calibration constants left by narrowband filtering can be performed.

6.2 Absolute calibration of AERA-II stations

6.2.1 Motivation

The vector effective length (VEL) as a function of frequency, electric field field polarization and arrival direction that forms the antenna pattern is calculated with computer simulations of the antenna characteristics (see section 4.3.4). To verify these simulated antenna patterns, dedicated measurements in the field can be performed with the help of radio emitters attached to balloons, cranes or octocopters (see [104] and [98] for details concerning calibration measurement campaigns of AERA antennas). From [98], we know that the simulated VEL of the Butterfly antenna differs significantly from the VEL inferred from the calibration measurements, as can be seen in figure 6.12.

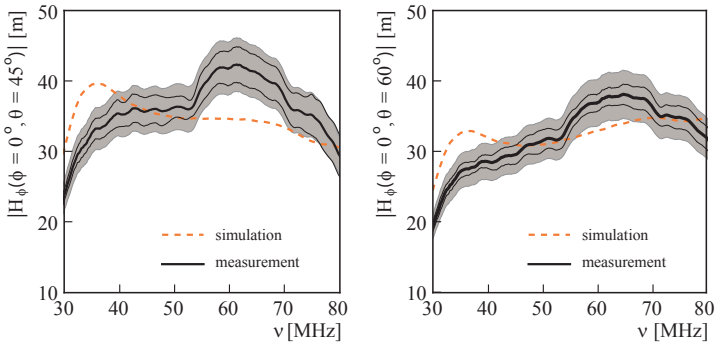


Figure 6.12: Calibration measurement of the VEL using an octocopter and the simulated VEL at two different zenith angles for the east-west antenna arm (modified from [98]). The filled areas indicate systematic uncertainties.

Within the bandwidth of AERA, significant enhancements of the VEL (H) are visible around 60 MHz in both arrival directions that were measured. This also has obvious ramifications for the spectra taken from air shower induced radio pulses. In figure 6.13,

two measured amplitude spectra of the reconstructed electric field at different observer angles are plotted, in which the simulated antenna pattern has been applied. The differences with the simulated spectra of figure 3.9 are that instead of decreasing exponential spectra, we now have spectra that peak around 60 MHz. This can be explained by the same deviations of the simulated antenna pattern that were also visible in the calibration measurements. If we want to perform a physics analysis based on the spectrum of air shower induced pulses, we need to compensate for these deviations. Unfortunately, there has been no calibration measurement of the antenna pattern covering the complete parameter space at the time of writing of this thesis.

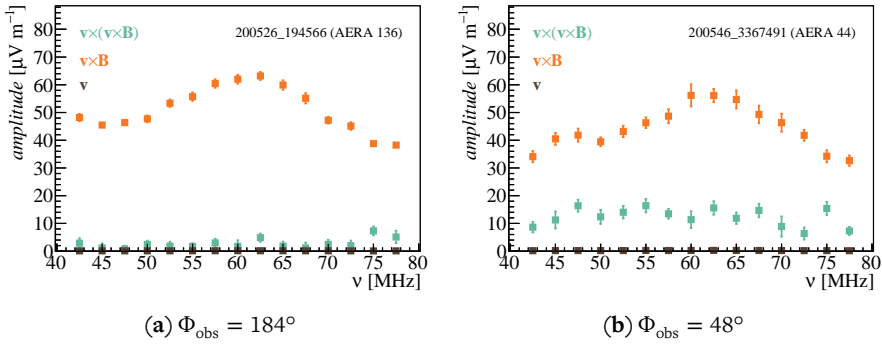


Figure 6.13: Measured amplitude spectra of the electric field, which was calculated using the simulated antenna pattern.

There is, however, an indirect method to validate and calibrate the antenna pattern. This is possible by using the radio background of the sky, as was already demonstrated in [104]. We can propagate a model of the full radio sky through the simulated antenna pattern, and compare this with the average spectra recorded by the stations. Like with the relative calibration using the background noise, we can only use it to probe the gain offsets of the antenna model and not the phase information.

6.2.2 Method

As a model for the radio background we use the output of the LFMmap program [114], which interpolates measurements of the full radio sky at several frequencies to create maps of the radio sky in antenna temperature T_a at any particular frequency. An example of the radio sky at 60 MHz produced by the model is visible in figure 6.14.

The antenna temperature can be converted to sky brightness $B_\nu(T_a)$ – or radiated power per unit frequency – by using the Rayleigh-Jeans approximation of a black body radiator:

$$B_\nu(T_a) = \frac{dP}{d\nu} = \frac{2k_B}{c^2} \nu^2 T_a, \quad (6.3)$$

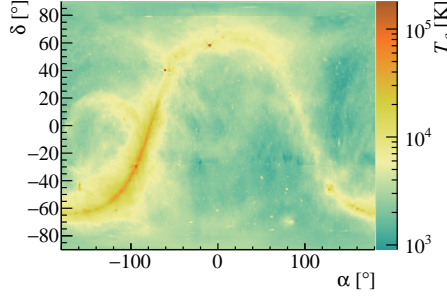


Figure 6.14: The radio sky at 60 MHz in antenna temperature T_a , produced by LFMmap. The equatorial coordinate system is used, with right ascension α and declination δ .

in which k_B is the Boltzmann constant and c the speed of light. The sky brightness is in units of $\text{W m}^{-2} \text{sr}^{-1} \text{Hz}^{-1}$. If we want to propagate this radiated power through the simulated antenna pattern, we have to calculate the total magnitude of the VEL of each antenna arm i (see also section 4.3.4). Because the radiation from the radio background sky is mostly unpolarized, we use:

$$||H_i|| = \sqrt{|H_{i,\theta}|^2 + |H_{i,\phi}|^2}. \quad (6.4)$$

This can be done for all frequencies. An example of the magnitude of the VEL of the simulated antenna patterns of both antenna arms is visible in figure 6.15, which is calculated at a frequency of 60 MHz.

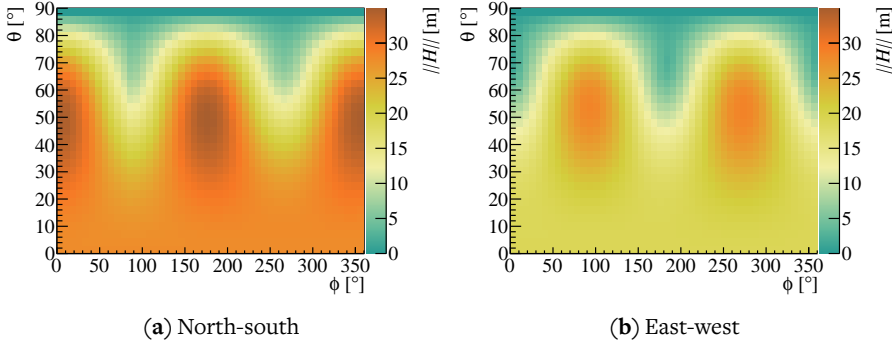


Figure 6.15: Magnitude of the VEL $||H||$ at 60 MHz in the horizontal coordinate system for both antenna arms. The asymmetry in the magnitude between the two polarization arms is clearly visible, and is probably caused by the modeling of the electronics box.

We can map the antenna pattern, which is defined in the horizontal coordinate system

(see e.g. figure 4.14) for observers at the Pierre Auger Observatory, to the equatorial coordinate system in which the radio sky is modeled. This mapping is a function of local sidereal time at the observer position. Once the antenna pattern is mapped to the equatorial coordinate system we calculate the spectral intensity $I_{\Omega,v}$:

$$I_{\Omega,v}(\alpha, \delta) = B_v(\alpha, \delta) ||H_{i,v}(\alpha, \delta)||^2, \quad (6.5)$$

of which the unit is $\text{W sr}^{-1} \text{Hz}^{-1}$. Figure 6.16 shows the spectral intensity at 60 MHz and $\text{LST} = 17 \text{ h}$ for both antenna arms. These can be interpreted as the intensity with which the antenna arms observe features in the radio sky.

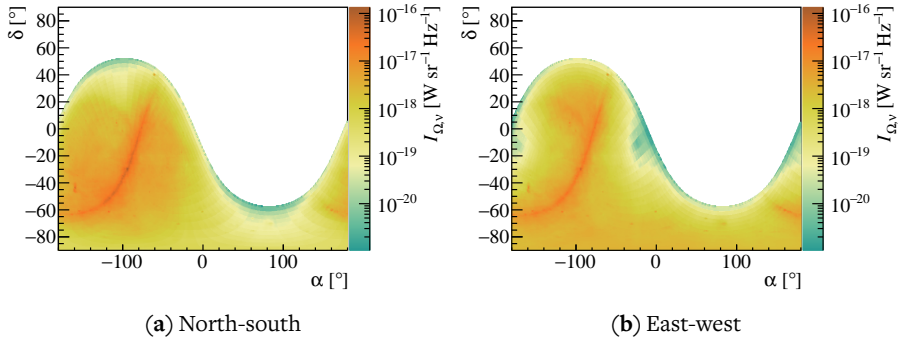


Figure 6.16: Spectral intensity $I_{\Omega,v}$ at 60 MHz for both antenna arms at $\text{LST} = 17 \text{ h}$.

From here, we can integrate over the sky to get the received power P_v per unit frequency (see [104] for a derivation of the prefactor):

$$P_v = \frac{1}{2} \frac{Z_0}{Z_L} \int_{\Omega} I_{\Omega,v}(\alpha, \delta) d\Omega, \quad (6.6)$$

where Z_0 is the impedance of free space of $120\pi \Omega$ and Z_L is the load impedance of 50Ω . Because we use a map with discrete binning it is convenient to rewrite this as a sum over the right ascension α and the declination δ :

$$P_v = \frac{1}{2} \frac{Z_0}{Z_L} \sum_{\alpha} \sum_{\delta} I_{\Omega,v}(\alpha, \delta) \cos \delta \Delta\delta \Delta\alpha. \quad (6.7)$$

From electric power we can easily convert to the voltage U_v which would be measured per unit frequency:

$$U_v = \sqrt{P_v Z_L}. \quad (6.8)$$

6.2.3 Results

We have created temperature maps of the radio sky with a step size of 1 MHz, and calculated for each frequency the amplitude spectrum as it would be measured through the

simulated antenna patterns for steps in LST of one hour. An example of an amplitude spectrum at $\text{LST} = 17 \text{ h}$ is plotted in figure 6.17(a). These spectra can be compared with actual measurements of the radio sky, and for this we use the average amplitude spectra as a function of LST created for the relative amplitude calibration. We calibrate each experimental setup of AERA separately, so after the calibration procedure the experimental setups will also be cross-calibrated via the radio sky, a step we have omitted so far. For the scintillator-triggered setup we use the validation data set defined in section 6.1.5, which ran from 2 June 2014 to 6 July 2014, and for the externally triggered data set we use data from 1 July 2014 to 1 August 2014. Both sets have been corrected for the effects of the analog chain, and the gaps left by the removal of the narrowband noise are linearly interpolated so the spectra are more or less smooth. This data is also binned in LST in bins of one hour. As a comparison, the measured amplitude spectrum at $\text{LST} = 17 \text{ h}$ from the scintillator-triggered setup is shown in figure 6.17(b).

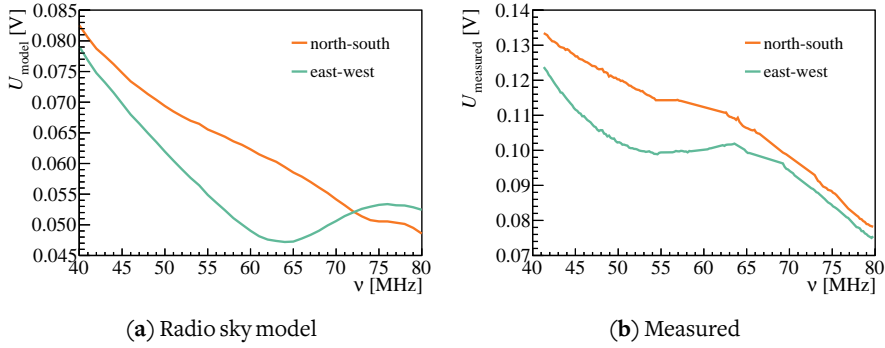


Figure 6.17: Comparison between the propagated amplitude spectrum of the radio sky model and the measured spectrum for both antenna arms at $\text{LST} = 17 \text{ h}$.

Apart from the difference in the shape of the spectrum there is also a significant gain offset between the two amplitude spectra. The origin of this gain offset is not known, but since the offset is present in the data from both experimental setups it could be caused by an underestimation of the amplification in the LNAs. To compensate for this gain offset we assume there is a constant scale parameter C_0 between the average amplitude produced by the model, $\langle U_{\text{model}} \rangle$, and the contribution of the radio sky to the average amplitude received by the antennas $\langle U_{\text{received}} \rangle$:

$$\langle U_{\text{model}} \rangle = \langle U_{\text{received}} \rangle \cdot C_0. \quad (6.9)$$

We also assume that the measured amplitude U_{measured} is a superposition of the amplitude received from the radio sky U_{received} and the electronic noise of the detector system U_{noise} :

$$U_{\text{measured}}^2 = U_{\text{received}}^2 + U_{\text{noise}}^2. \quad (6.10)$$

From equations (6.9) and (6.10) it follows that we can relate the constant scale parameter and average noise amplitude to the measured and modeled radio sky amplitudes as:

$$\langle U_{\text{measured}} \rangle^2 = \frac{\langle U_{\text{model}} \rangle^2}{C_0^2} + \langle U_{\text{noise}} \rangle^2. \quad (6.11)$$

This means that the scale parameter C_0 and the average noise amplitude U_{noise} follow from a linear fit of U_{measured}^2 versus U_{noise}^2 of both antenna arms for all frequency and LST bins, as is plotted in figure 6.18.

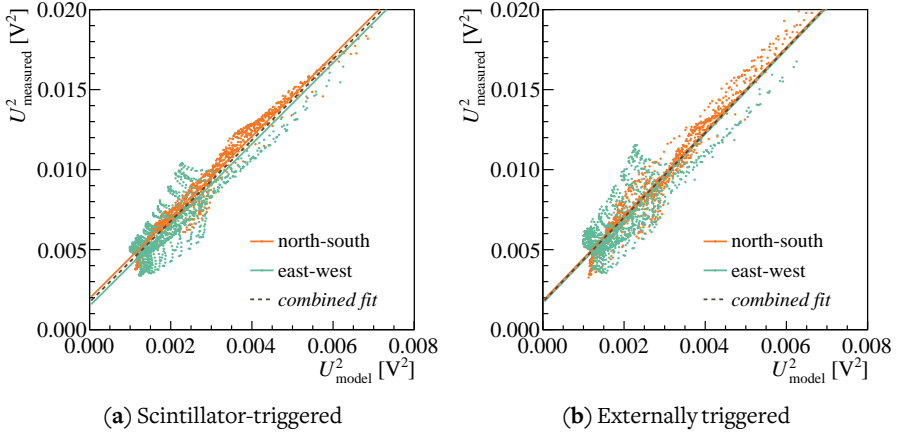


Figure 6.18: Linear fit to all data points of both experimental setups to derive the constant scale parameter C_0 to describe the gain offset between the model and the measurement, and the average noise amplitude $\langle U_{\text{noise}} \rangle$. The fanning pattern of the data points around the fit is caused by the frequency-dependent offsets between the model and the measurement.

Because we want to calibrate both measured antenna channels to the radio sky model, and we assume that the observed gain offset affects both channels equally, the same average scale parameter has to apply to both channels. The noise contribution to both measured channels, however, might be different. Therefore, we will first perform a combined fit to determine C_0 , given by the slope, and then perform two independent fits of the channels with the slope fixed to determine the average noise amplitudes. We do this for both experimental setups separately because the gain offsets as well as the electronics noise could have their origins in the different hardware that is used. The resulting values are listed in table 6.1. The uncertainty of the noise amplitude in the east-west antenna channel, $\langle U_{\text{noise,EW}} \rangle$, is much higher than that of the noise amplitude in the north-south channel, $\langle U_{\text{noise,NS}} \rangle$. This is because the deviations of the simulated antenna model for the east-west arm are much larger than for the north-south arm, as we shall see later on.

Table 6.1: Fit results to determine the constant scaling parameter and noise amplitudes.

	scintillator-triggered	externally triggered
C_0	1.591 ± 0.005	1.625 ± 0.006
$\langle U_{\text{noise, NS}} \rangle$ [mV]	44.2 ± 0.2	42.5 ± 0.3
$\langle U_{\text{noise, EW}} \rangle$ [mV]	38 ± 19	41 ± 20

We can apply the scale parameter and the noise amplitudes to the measured amplitudes using equations (6.9) and (6.10) to recover U_{received} , the contribution of the radio sky to the measurements in each frequency bin. In figure 6.19, the average amplitude of both the model and the received radio sky over the whole frequency range as a function of LST is plotted for both antenna arms of the scintillator-triggered setup. The model and the measurement follow each other remarkably well, which means that even though there might be differences in the sensitivity as a function of frequency, the directional sensitivity is described well by the simulated antenna pattern.

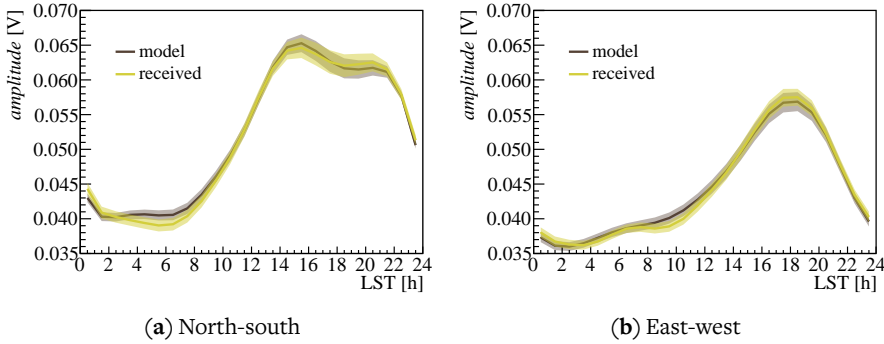


Figure 6.19: Spectral amplitude averaged over all frequencies of both the radio sky model propagated through the antenna pattern and the radio sky as received by the scintillator-triggered setup as a function of LST after a correction for the constant gain offset and the electronic noise. The shaded areas indicate the standard errors of the average values.

It is also possible to look at the deviations in the simulated antenna pattern as a function of frequency. For this, we have divided the U_{received} spectra by the U_{model} spectra in each LST bin, and then averaged each 1 MHz spectrum bin over all LST bins. The resulting average deviations in both experimental setups are plotted in figure 6.20. If the simulated antenna pattern would have perfectly described the actual antenna response, the ratio would be unity over the whole frequency range. The response pattern of the north-south antenna arm is reasonably well described by the antenna simulations, at

least until ~ 72 MHz. However, the response pattern of the east-west antenna arm has a large enhancement in the 55 MHz – 70 MHz range, which might have the same origin as the enhancements that are visible in the calibration measurements (figure 6.12) and the air shower spectra (figure 6.13). The global features are similar in both experimental setups, and the variation between the deviations might be the result of differences between the measured and actual frequency response of the analog chains of the two experimental setups.

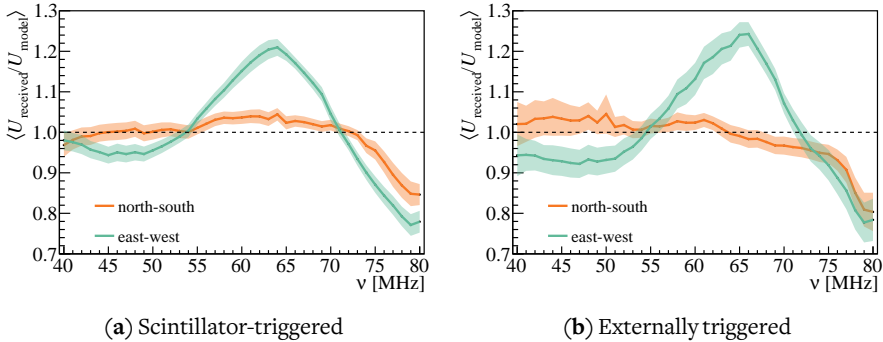


Figure 6.20: Average deviations of the received radio sky in both experimental setups with respect to the modeled radio sky with the incorporation of the simulated antenna response pattern. The measured radio sky has been corrected by the constant gain offset C_0 and by an estimate of the electronic noise to recover the received signal. The shaded areas indicate one standard deviation.

The average deviations can be interpreted as calibration constants that need to be inversely applied to the data before the electric field is reconstructed with the simulated antenna response pattern. Because they were calculated with measured spectra that have been corrected for the analog chain between the digitizer and the output terminals of the LNA, the constants are applied to all stations with Butterfly antennas. The calibration constants are integrated in the *Offline* framework by simply multiplying them with the relative calibration constants calculated in the previous section. This way, the absolute and relative calibration can both be applied by the `RdChannelGalactic-BackgroundCalibrator`.

6.3 Suppression of narrowband RFI in the time domain

As previously mentioned in section 5.1.1, several sources of narrowband RFI influence the data recorded at AERA. Removing the contribution of these noise sources from the data is an important task, especially in physics analyses where the frequency spectrum is concerned. Several different techniques have been developed in the past to remove

the influences of these sources from the data. Most of them utilize the frequency domain. One of the simplest methods is to set all frequency bins which are influenced by narrowband transmitters to zero. This will naturally remove all information of this frequency bin, including physically relevant information. The *median filter* is a similar method, but this sets the amplitude of the affected frequency bin to the average of the surrounding frequency bins. Although this recovers to a certain extent the amplitude of the frequency bin, the phase of the frequency bin will remain unchanged and as much affected by the narrowband transmitter as before. In addition, both of these methods do not address the problem of *spectral leakage*, which occurs when the frequency of the narrowband transmitter is not exactly in the center of a frequency bin in a discrete Fourier transform, as is illustrated in figure 6.21. This makes it very hard to precisely remove the influence of narrowband noise using the frequency domain, as many bins surrounding the transmitter frequency might be affected.

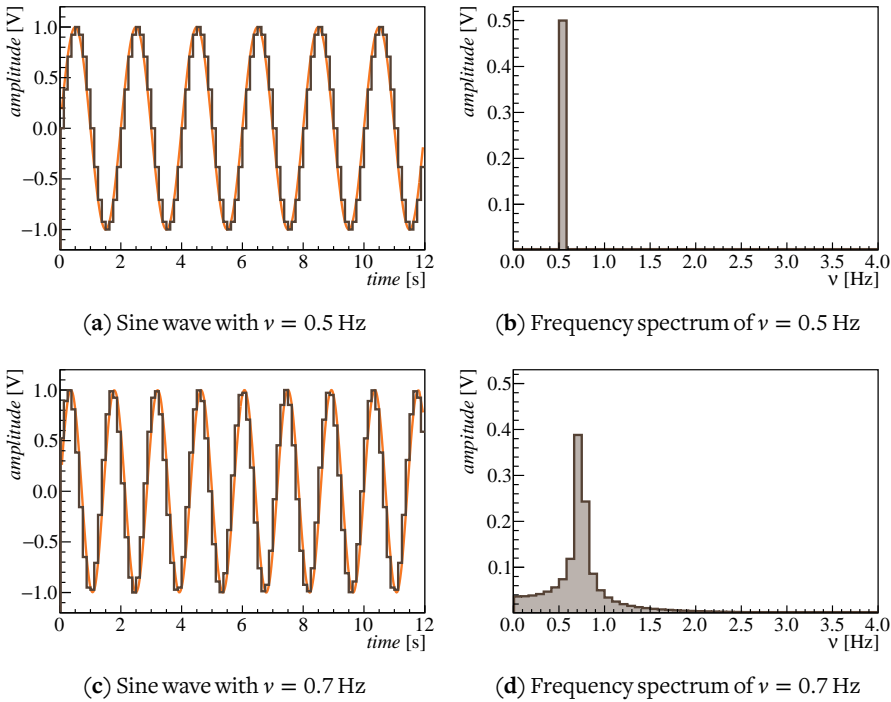


Figure 6.21: Example of spectral leakage at a sampling frequency of 8 Hz. The sine wave at the top has a frequency exactly in the center of a frequency bin, while the bottom sine wave has a frequency which is off-center to the frequency bin. The bottom frequency shows excessive leakage.

As was first demonstrated in [61], the problem of the phase information and the spectral leakage can be tackled by suppressing RFI in the time domain instead.

6.3.1 Method

The contribution of RFI with a constant frequency ν to the time trace has the form of pure sine waves:

$$f(t) = A \cdot \sin(2\pi\nu t + \phi), \quad (6.12)$$

with A as amplitude and ϕ the phase of the wave. We can fit these sine waves to the time traces, and then subtract them. In order to do so we have to know approximate frequencies of the narrowband sources, as well as an estimation of their amplitude.

6.3.2 Dynamic RFI identification and parameter estimation

Although the beacon frequencies are exactly known, there is also other narrowband RFI which we will try to suppress using this method. To find start values for the fit we dynamically identify narrowband noise using the frequency spectrum. In the frequency spectrum, the amplitudes of ordinary instrumental noise approximately follow a Rayleigh distribution, which has the following probability density function:

$$f(x) = \frac{x}{\sigma^2} \exp\left(\frac{-x^2}{2\sigma^2}\right), \quad (6.13)$$

where σ is the scale parameter, equal to the mode of the distribution. Any frequency bins which deviate significantly from this distribution are probably affected by narrowband noise, as is illustrated in figure 6.22, where the amplitudes in two frequency bands are plotted. One of the bands is clean of known RFI, while in the other two beacon frequencies are located. The bins located at the beacon frequencies clearly deviate from the Rayleigh distribution. Because the average amplitude of the frequency spectrum is not constant over the whole frequency range (see for instance the slope of the spectra in figure 6.3), the Rayleigh distribution is only a good approximation of the true shape within small spectral bands of a few MHz.

By estimating the shape of the distribution we can determine an amplitude threshold above which bins are likely to be influenced by narrowband transmission. For the Rayleigh distribution, the scale parameter σ is estimated using its maximum likelihood:

$$\sigma \approx \sqrt{\frac{1}{2N} \sum_{i=1}^N x_i^2}, \quad (6.14)$$

in which N is the total number of samples x_i . The estimation corresponds almost exactly to the peak of the distribution in the clean frequency band of figure 6.22(a). In the contaminated band of figure 6.22(b) it is however slightly overestimated. Using the scale parameter we calculate the cumulative distribution function $F(x)$ to set a threshold at a certain fraction:

$$F(x) = 1 - \exp\left(\frac{-x^2}{2\sigma^2}\right). \quad (6.15)$$

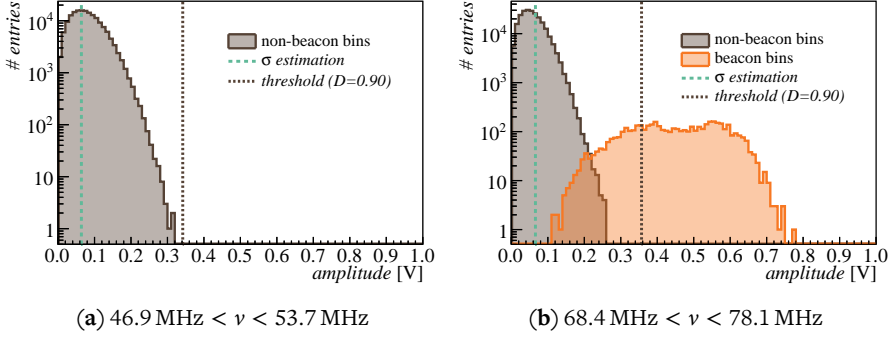


Figure 6.22: Rayleigh distribution of the amplitudes in the frequency spectrum in two spectral bands, with the associated scale parameter estimations and noise thresholds. The distributions contains the north-south channel of all active stations from 100 periodically triggered events from 17 October 2014 (run 556).

To make the probability of a bin crossing the threshold similar for any number of sampled frequency bins, we determine this fraction as a function of the total number of bins N :

$$F(x) < D^{\frac{1}{N}}, \quad (6.16)$$

where we have determined $D = 0.90$ as a good discriminator. The resulting thresholds are indicated in figure 6.22. In figure 6.22(a) all the noise is contained, while in figure 6.22(b) a significant amount of bins affected by the beacon cross the threshold.

We select all frequency bins which have amplitudes that cross the threshold as possible narrowband source frequencies. This is done using a sliding window of 100 bins in the frequency spectrum to negate the effect of the changing noise level over the full frequency range. If several bins in a row are selected, we only use the bin with the highest amplitude. This is because they are probably all caused by the same transmitter, which is either wider than a single frequency or has significant spectral leakage. As start value for the frequency ν of the fit we use the center of the selected frequency bin, and as range we take the size of one complete frequency bin around it. The amplitude A of the sine wave is estimated from the amplitude in the frequency spectrum M . Depending on the amount of spectral leakage the amplitude of the sine wave is:

$$\frac{2M}{N} \leq A \leq \frac{\pi M}{N}, \quad (6.17)$$

where N is the total number of frequency bins. As start value for the amplitude we use the minimum amplitude, and as upper limit the amplitude resulting from maximum spectral leakage. We take the modal noise level σ/N as the lower limit for the amplitude.

After performing the fit, only sine waves with a fitted amplitude higher than the threshold are subtracted from the time trace. The phase ϕ of the sine wave is also estimated from the Fourier transform, but will be unbounded in the fit.

6.3.3 Results

A comparison between average spectra with and without subtraction of the fitted sine waves is displayed in figure 6.23. Because the beacon is located almost exactly due west from the array, the four beacon frequencies are mostly visible in the north-south channel. They are clearly reduced to below the standard deviation of the noise. Other narrowband emitters are also strongly reduced in both channels, but not all to the same level as the beacon. This is most likely because they are not as narrow in frequency as the beacon, so only the main component is removed, such as the carrier frequency in the case of an FM transmitter.

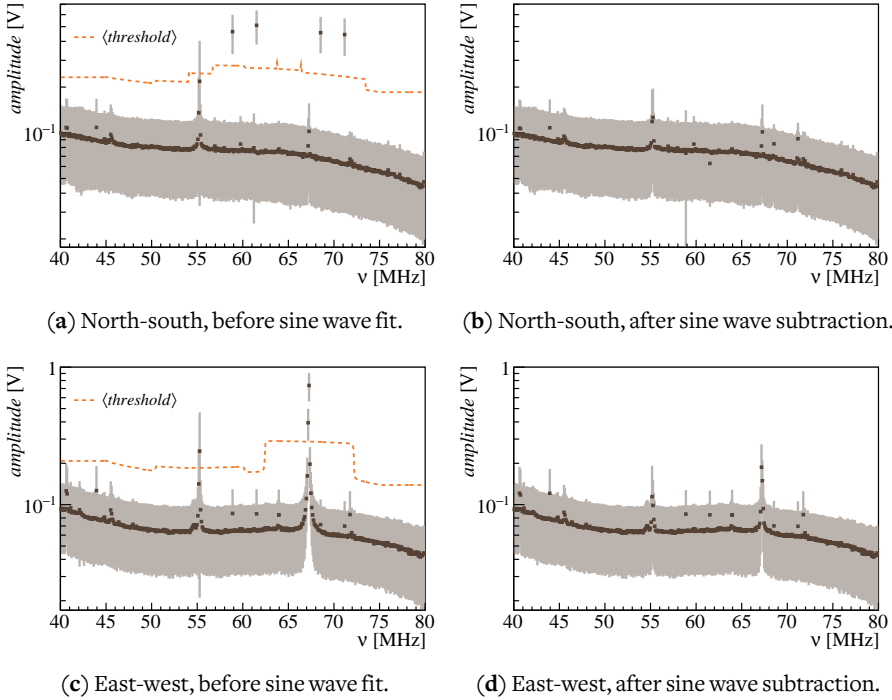


Figure 6.23: Average spectra with and without subtraction of fitted sine waves. The shaded bands indicate one standard deviation of the average values, and the average noise threshold is shown as a dashed line. All active stations from the scintillator-triggered setup from 100 periodically triggered events from 17 October 2014 (run 556) were used.

6.3.4 Integration into the software framework

A module to apply the RFI-suppression in the time domain has been created for the *Offline* framework: the `RdChannelSineWaveSuppressor`. The module dynamically identifies narrowband RFI in the frequency domain and fits sine waves in the time domain using the `Minuit` minimization program [115]. Configurable options of the module include the size of the sliding window in which the mode of the Rayleigh distribution is estimated and the value for D to calculate the noise threshold.

Results: Measuring shower development with AERA

IN THIS CHAPTER we will analyze the spectral index of radio pulses and estimate from them the depth of shower maximum X_{max} . First, we will determine the spectral index from the pulse spectrum and estimate its uncertainties. We will then reconstruct X_{max} using the parameterization of the spectral index derived in chapter 3, and compare it to measurements of X_{max} made by the fluorescence detector. After this cross-check, we will calculate the average depth of shower maximum as a function of cosmic ray energy, from which we can infer cosmic ray composition.

7.1 The spectral index in AERA data

7.1.1 Extraction of the pulse spectrum

Before we are able to study the radio pulse spectrum, we must reconstruct the electric field of the events that have survived all of the selection steps discussed in chapter 5. While we perform the reconstruction, we apply the calibration and cleaning procedures as described in chapter 6. An overview of all the steps we take to reconstruct the full three-dimensional electric field from the raw ADC traces of the selected events is shown in figure 7.1. This figure also serves as a road map for the analysis described in this chapter, going from the electric field all the way to the estimation of the depth of shower maximum X_{max} .

The next step is to rotate the electric field from the Cartesian coordinate system to the coordinate system in the shower reference frame along the \mathbf{v} , $\mathbf{v} \times \mathbf{B}$ and $\mathbf{v} \times (\mathbf{v} \times \mathbf{B})$ directions, similar to our simulation study of chapter 3. For this, we need the direction of the geomagnetic field \mathbf{B} at the AERA site at the time of the event. Several models exist to calculate the geomagnetic field as a function of position around the globe and time. We use the IGRF-12 model [116] to calculate the field at the AERA site. The direction of the calculated field as a function of time is plotted in figure 7.2. The uncertainty of

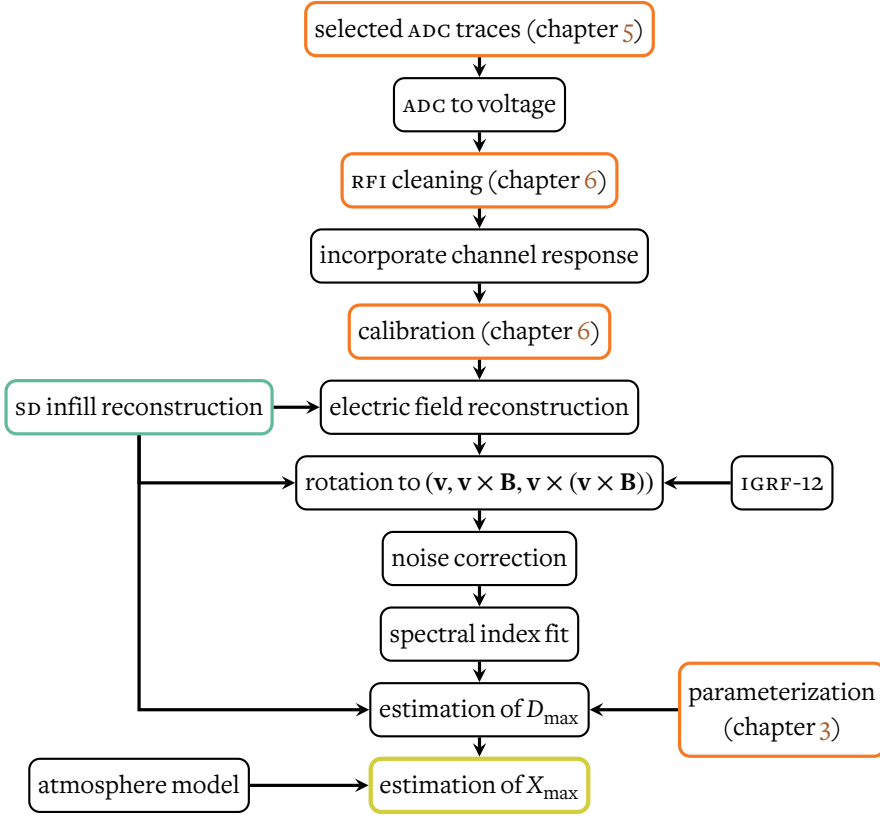


Figure 7.1: Reconstruction chain from selected radio traces to the estimation of depth of shower maximum X_{\max} . The analysis steps from the electric field reconstruction onward are described in this chapter. The shower geometry from SD is used as input to reconstruct the three-dimensional electric field, rotate the field in the shower reference frame and estimate the position of shower maximum when applying the parameterization of the spectral index.

the orientation of the magnetic field as quoted from the model guidelines is negligible compared to the uncertainties in the shower geometry, as we shall see in the next section.

After the coordinate transformation we determine the pulse time in the $\mathbf{v} \times \mathbf{B}$ component of the electric field, and extract the signal pulse in a window of 400 ns around the peak value of the Hilbert envelope, as was discussed in section 5.4.3. We then calculate the amplitude spectrum of all three components of the electric field in this window by performing a discrete Fourier transform. From the same rotated time traces, we extract as many noise traces as we can with the same 400 ns length outside the signal window.

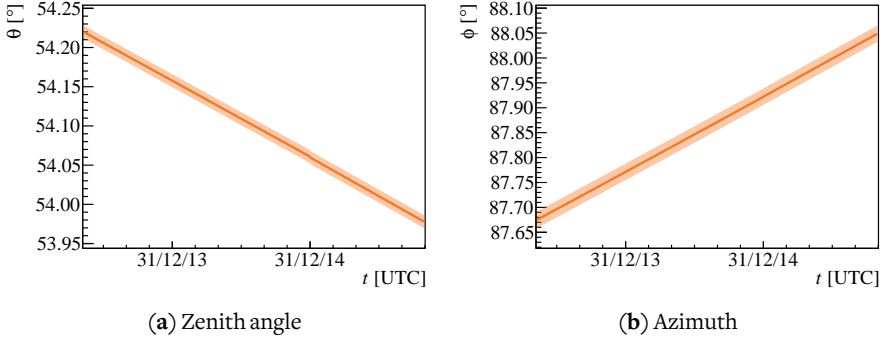


Figure 7.2: Geomagnetic field orientation as a function of time at the AERA site as approximated by the IGRF-12 model. Uncertainties are 0.008° for the zenith angle (inclination) and 0.016° for the azimuth (declination).

We use these to calculate average noise spectra in all three components. The resulting spectra will have frequency bins with a width of 2.5 MHz.

A limited noise correction can be performed on the individual frequency bins of the spectrum using the average noise spectrum, as was demonstrated in [61]. Because of the random phase of the noise contribution to each frequency bin of the signal spectrum, it is difficult to estimate the true signal amplitude at a certain frequency. Following the prescription in [61], an estimation of the signal amplitude can be made by assuming a vectorial sum of the signal and noise contributions and integrating over the expected phase difference between the true signal and the noise background ϕ . Using the amplitudes of the signal spectrum S_m and the noise spectrum S_n , we can calculate a corrected signal S_{cor} :

$$|S_{\text{cor}}(\nu)| = \frac{1}{2\pi} \int_0^{2\pi} \sqrt{S_m(\nu)^2 - (S_n(\nu) \sin \phi)^2} d\phi, \quad (7.1)$$

where ν is the central frequency in the spectrum bin. The amplitudes of the frequency bins in the average noise spectrum will furthermore be used as uncertainties of the signal spectrum, which will serve to weigh the data points when we perform the fit of the spectral index. An example of a calibrated signal spectrum with uncertainties and associated average noise spectrum can be seen in figure 7.3.

7.1.2 The spectral index and its uncertainties

To determine the spectral index b of the pulse we fit equation (3.2) to the corrected amplitude spectra in the range from 41 MHz to 79 MHz. This is the clean range in which we have been able to calibrate our stations, and which is reasonably free from narrowband noise sources.

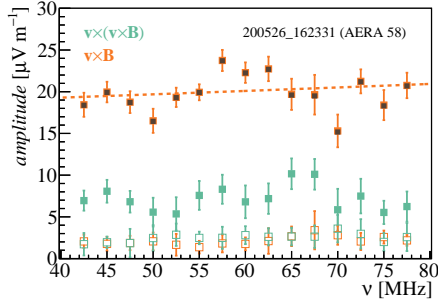


Figure 7.3: Example of calibrated signal spectra (solid markers) with average noise spectra (open markers) of the two components of the reconstructed electric field. The uncertainties of the signal spectrum amplitudes are obtained from the average noise spectrum, and the uncertainties of the noise spectra are obtained from the rms of the average noise amplitudes. The corrected signal amplitudes S_{cor} are drawn in brown, and a fit of equation (3.2) is shown as a dashed line.

Noise background

The level of the noise background will have an influence on the frequency spectrum of the measured electric field. This results directly from the fact that the measured spectrum is a superposition of noise and signal. As an example, it is easy to imagine a high powered white noise spectrum, which will result in a flattening of the measured spectrum when it is added to a signal, and thus bringing the spectral index closer to zero. Additionally, noise is further amplified when we calculate the three-dimensional electric field from the two measured polarizations. This is an intrinsic property of the field recovery method, as was demonstrated in section 4.3.4. To estimate the effect of the noise background we first quantify the amount of noise by means of the signal-to-noise ratio, where we will use the integrated definition of equation (4.5), and use the average noise spectrum to calculate the noise level. This definition is more independent of the pulse shape than the SNR definitions that rely on the peak amplitude of the pulse. The distribution of the SNR of all selected events is displayed in figure 7.4.

In section 4.3.4, we used a toy Monte Carlo model to show that the uncertainty of the spectral index is reasonably constant as a function of arrival direction, as long as the zenith angle θ is smaller than 60° . We use the same model here to determine the influence of the noise level on the spectral index, and to estimate the effects on the uncertainty. We again compare the real spectral index b_{real} before the addition of Gaussian white noise and the application of the antenna model with the spectral index measured afterwards, b_{measured} . This time, we express the difference as a function of the SNR, determined after the electric field reconstruction. The average value and standard de-

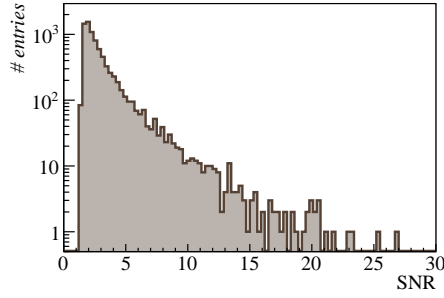


Figure 7.4: Distribution of the SNR as defined by equation (4.5) of all selected events, disregarding weather conditions.

viation of the difference $b_{\text{real}} - b_{\text{measured}}$ calculated using 5 million pulses generated with random arrival directions of $\theta < 60^\circ$ are displayed in figure 7.5. The calculated average differences as a function of SNR will be used to correct the fitted spectral indices of the data. The obtained standard deviations will be used to assign an uncertainty to the spectral index based on the SNR.

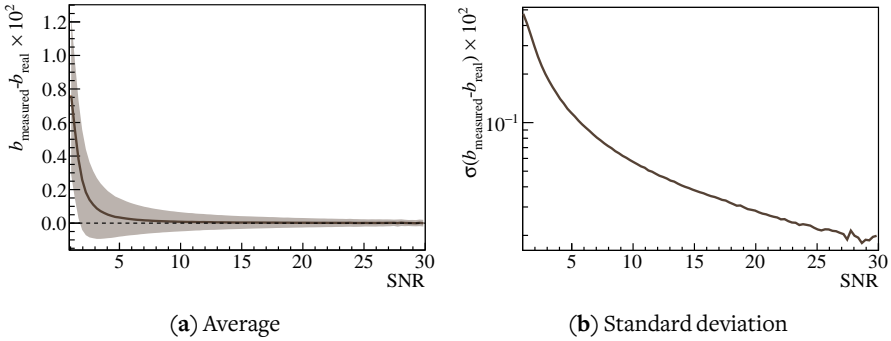


Figure 7.5: Average difference between the real spectral index b_{real} and the measured spectral index b_{measured} after the addition of Gaussian white noise and application of the antenna model as a function of SNR. The shaded band indicates one standard deviation of the average in (a). A close-up of the standard deviation is provided in (b).

SD reconstruction uncertainty

The uncertainty of the reconstructed shower arrival direction and core impact location have a direct influence on the knowledge of the station distance to the shower axis d . As we know from chapter 3, this is one of the main dependencies of the parameterized spectral index. With the covariance matrix $\text{cov}(i, j)$ of the core location reconstruction

provided by the sd lateral distribution fit we can propagate the uncertainty of the core impact location to the uncertainty of the distance to the shower axis $\sigma_{d,\text{core}}$:

$$\sigma_{d,\text{core}}^2 = \frac{1}{d^2} [d_1^2 \cdot \text{cov}(d_1, d_1) + d_2^2 \cdot \text{cov}(d_2, d_2) + 2d_1d_2 \cdot \text{cov}(d_1, d_2)], \quad (7.2)$$

where d_1 and d_2 are two vector components perpendicular to the direction of movement of the shower \mathbf{v} that define the vector from the station position to the shower axis with length d , in such a way that $d^2 = d_1^2 + d_2^2$. The distribution of the uncertainties of the distance to the shower axis of all selected events is plotted in figure 7.6.

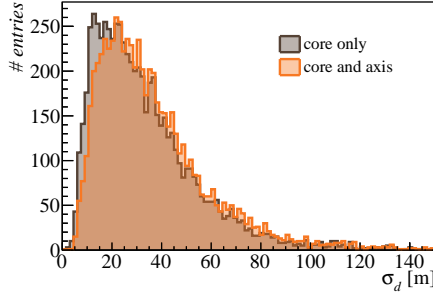


Figure 7.6: The uncertainty of the distance from the station to the shower axis for all selected events, calculated once by using only the core uncertainty with equation (7.2), and once as the quadratic addition of the core and axis uncertainties via equation (7.4).

The reconstructed arrival direction of the shower also has an uncertainty, but this is not explicitly included in the standard sd reconstruction because it is usually small compared to the core uncertainty. Using Gaussian error propagation, we can estimate the influence of the uncertainty of the shower axis σ_θ on the uncertainty of the distance from a radio station to the shower axis:

$$\sigma_{d,\text{axis}} \approx \frac{|d_v| \sigma_\theta}{\cos^2 \theta}, \quad (7.3)$$

where θ is the zenith angle of the shower and d_v the distance along the shower axis from the core impact point to the point where a plane shower front would intersect the radio station position, as is illustrated in figure 7.7. The upper limit on the uncertainty of the shower axis direction is 2° [117]. Given that the typical distance to the shower axis d in our data set is 150 m, and the most common zenith angle θ is 45° , this would result in an uncertainty of less than about 10 m of the distance to the shower axis, which is indeed small compared to the uncertainties propagated from the core location. Although the uncertainty of the reconstructed sd axis and core impact location are not independent, we can make a conservative estimation of the total uncertainty of the distance to the

shower axis σ_d by adding them quadratically as:

$$\sigma_d = \sqrt{\sigma_{d,\text{core}}^2 + \sigma_{d,\text{axis}}^2}. \quad (7.4)$$

The resulting distribution of the total uncertainty of the station distance to the shower axis is also plotted in figure 7.6, where we have used the upper limit of $\sigma_\theta = 2^\circ$.

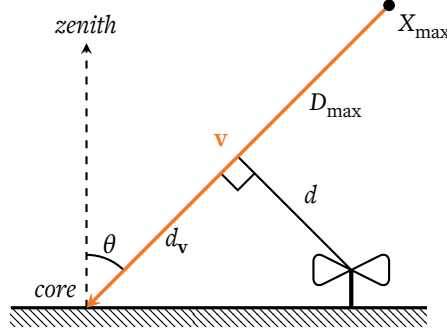


Figure 7.7: Reconstructed shower geometry and relative position of the radio station.

The distance to the shower axis d is the shortest distance from the station position to the shower axis \mathbf{v} , running perpendicular to it. The distance from this intersection to X_{max} is D_{max} , and the distance from the intersection to the shower core is d_v .

The uncertainty of the shower arrival direction (i.e. the direction the \mathbf{v} vector is pointing) also has an influence on the way we perform the coordinate transformation to the shower reference frame. An erroneous axis direction might result in the mixing of other components into the $\mathbf{v} \times \mathbf{B}$ component, in which we are interested. The potential influence of this mixing on the spectral index is determined by the relative strength of the emission in the $\mathbf{v} \times \mathbf{B}$ and the $\mathbf{v} \times (\mathbf{v} \times \mathbf{B})$ components of the field and their spectral indices. If we follow equations (3.8), (5.4) and (5.5), we find that the relative strength of the emission is a function of geomagnetic angle α and the observer angle Φ_{obs} with uncertainty σ_Φ .

The uncertainty of the observer angle, given by $\Phi_{\text{obs}} = \arctan(d_2/d_1)$, is propagated from the uncertainty of the two vector components perpendicular to the shower axis, as was done in equation (7.2):

$$\sigma_\Phi^2 = \left(1 + \left(\frac{d_2}{d_1} \right)^2 \right)^{-2} \left[\frac{d_2^2}{d_1^4} \text{cov}(d_1, d_1) + \frac{1}{d_1^2} \text{cov}(d_2, d_2) - \frac{d_2}{d_1^3} \text{cov}(d_1, d_2) \right], \quad (7.5)$$

where the covariance matrix is again provided by the SD lateral distribution fit. We neglect the influence of the uncertainty of the shower axis here, because even in our most

conservative estimation it only has a small effect. The distribution of the uncertainty of the observer angle in our selected data set is displayed in figure 7.8.

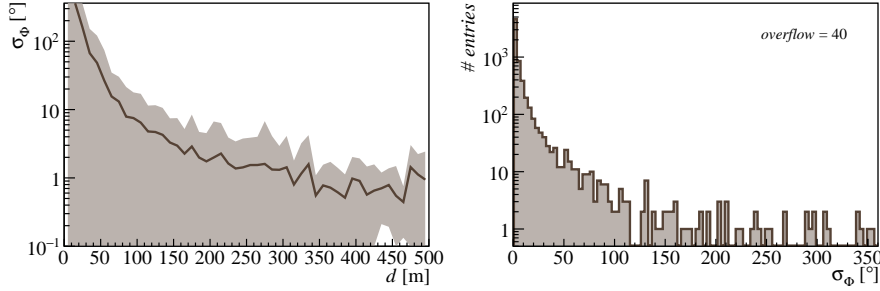


Figure 7.8: Uncertainty of the observer angle in the selected data set, as calculated with equation (7.5). *Left:* average uncertainty as a function of distance to the shower axis d (shaded bands indicate one standard deviation), *right:* total distribution.

We can now go back to our toy Monte Carlo, and modify it in such a way that it can simulate contamination of the $\mathbf{v} \times \mathbf{B}$ component by the $\mathbf{v} \times (\mathbf{v} \times \mathbf{B})$ component, and estimate the uncertainty of the spectral index as a function of α , Φ_{obs} and σ_Φ . This is done by randomly drawing pulses from a set of 200 events simulated at $\Phi_{\text{obs}} = 90^\circ$ and $100 \text{ m} \leq d \leq 400 \text{ m}$ in the non-discrete simulation set described in section 3.3.2. Because of this selection of the observer angle, each pulse contains a geomagnetic and charge-excess component that are perpendicular and can be separated. This ensures that we mix pulses with realistic values for the spectral indices in the two electric field contributions. We then assign a random arrival direction with zenith angle $\theta < 60^\circ$ and a random observer angle Φ_{obs} . We also select a random uncertainty of the observer angle σ_Φ between 0° and 10° , because this is the likely value in our region of interest of $100 \text{ m} \leq d \leq 400 \text{ m}$, as can be seen in figure 7.8.

The next step is to compose the $\mathbf{v} \times \mathbf{B}$ and $\mathbf{v} \times (\mathbf{v} \times \mathbf{B})$ components and their relative power with the selected geomagnetic and charge-excess pulses as they would be measured by an observer, based on the arrival direction, the associated geomagnetic angle and observer angle using equations (3.8), (5.4) and (5.5). The pulse is then propagated through the simulated antenna model to get the two-component voltage, after which a random variation is added to the shower axis direction using a two-dimensional Gaussian with $\sigma = 2^\circ$. Likewise, a random variation is added to the observer angle using a Gaussian with a width of σ_Φ . After this, we apply the inverse of the antenna model to recover the three-dimensional field again, and we calculate the spectral index of the field in the $\mathbf{v} \times \mathbf{B}$ component of the field, b_{measured} , which we can compare with the spectral index of the original pulse, b_{real} .

In similar fashion to what we did for the noise level, we now estimate the uncertainty of the spectral index as a function of the three parameters by using the standard deviation of the average difference between b_{measured} and b_{real} . Two projections of the final distribution after generating 100 million pulses with the Monte Carlo are displayed in figure 7.9. The deviations are the largest at observer angles where the two emission contributions are perpendicular. This is because at these angles the relative power of the pulse in the $\mathbf{v} \times (\mathbf{v} \times \mathbf{B})$ component is the largest, and will therefore influence the $\mathbf{v} \times \mathbf{B}$ component the most when there is mixing. Using the same logic, we deduce that for small geomagnetic angles the relative power in the $\mathbf{v} \times (\mathbf{v} \times \mathbf{B})$ is the largest, and will therefore also produce a large uncertainty. This can also be observed in the plots of figure 7.9. The uncertainty of the observer angle hardly has any influence on the spectral index for the range we have selected. We will therefore assign uncertainties to the spectral indices resulting from uncertainties in the SD reconstruction only as a function of α and Φ_{obs} , and derive these from the Monte Carlo.

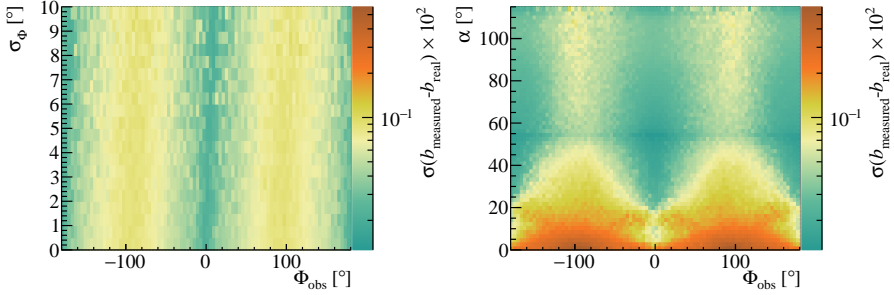


Figure 7.9: Projections of the standard deviations of the average difference between real and measured spectral indices produced by our toy Monte Carlo as a result of mixing between the two electric field components. *Left:* observer angle Φ_{obs} versus its uncertainty σ_ϕ ; *right:* observer angle versus geomagnetic angle α .

The contributions to the uncertainty from the noise background and the SD reconstruction are assumed to be independent, so the total uncertainty of the spectral index b is calculated by the quadratic addition of the two.

7.2 Dependence of the spectral index on showerdevelopment

In this section, we use the RD-FD data set described in section 5.4.5 to test the parameterization of the spectral index as a function of shower development – which we derived in chapter 3. First, we examine the general dependencies of the measured spectral index. In particular, we investigate its relation with the distance to the shower axis d and the geometrical distance to shower maximum D_{max} . We then use the parameterization

of the dependencies of the spectral index to reconstruct the depth of shower maximum on an event-by-event basis and compare these with FD measurements. Finally, we use the measurements of the depth of shower maximum in a subset of the selected data set from chapter 5 to calculate the average depth of shower maximum as a function of energy.

7.2.1 Measurements of the spectral index

The spectral indices of all events in the RD-FD data set are plotted in figure 7.10. The data is divided into six bins of distance to shower axis d of 50 m each. The geometrical distance from the station to shower maximum D_{max} and its uncertainty are derived from the atmospheric depth of shower maximum as reconstructed from the FD measurements using average monthly atmospheric density profiles. In each plot, the parameterization of the spectral index of a purely geomagnetic signal at the center of the distance bin as a function of D_{max} is indicated by a solid line. The parameterizations of the spectral index at the edges of the distance bin for extreme values of the observer angle ($\Phi_{\text{obs}} = 180^\circ$ at the lower edge and $\Phi_{\text{obs}} = 0^\circ$ at the upper edge) are marked by dashed lines to indicate the full range of the parameterized spectral index. Considerable spread can be observed in the measured indices, but the majority of the data points are in agreement with the region defined by the parameterization within one standard deviation.

Average profile histograms of the four distance bins in which we have a large number of events are visible in figure 7.11. The average spectral index is calculated in bins of 1 km of D_{max} . The average measured spectral index is compatible with the parameterization, indicated again by solid and dashed lines, especially when there is a significant number of measured values in a bin. Because of the constructive interference of geomagnetic and charge-excess emission at $\Phi_{\text{obs}} = 0^\circ$, signals measured at this observer angle are more likely to exceed the SNR-cut. Therefore, the measurements in this plot are biased towards this observer angle, which results in a lower average spectral index.

7.2.2 Event-by-event reconstruction of shower maximum

Unfortunately, the number of events in the RD-FD data set is not sufficient to perform a full data-driven analysis along the lines of the simulation study of chapter 3 to derive the dependencies of the spectral index. Instead, we rely on the parameterization derived from simulations for the extraction of information on the depth of shower maximum from the radio signal.

In order to find the best fitting value of the geometrical distance from the shower core to shower maximum $D_{\text{max, core}}$, we minimize the following function for events consisting

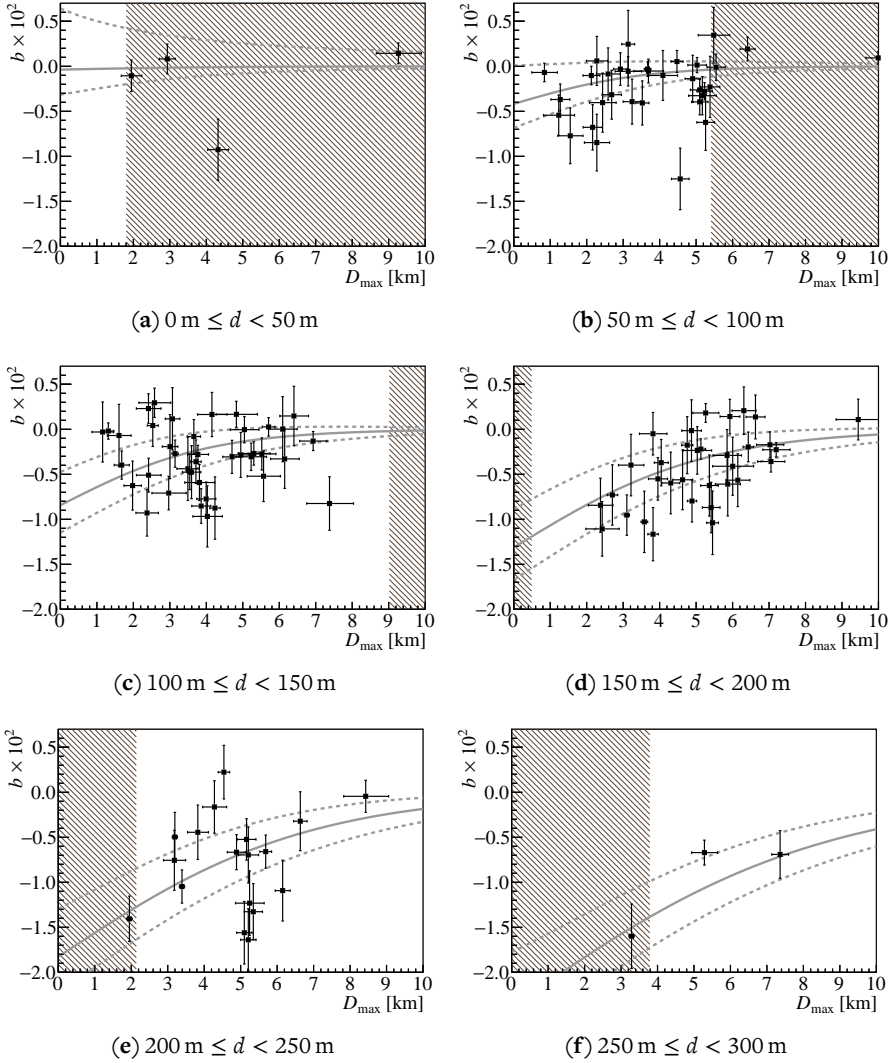


Figure 7.10: Dependence of the measured spectral index b on the distance to shower maximum D_{max} as determined by the FD for different bins of the distance to the shower axis d . The lines indicate the predicted spectral index based on the parameterization at the bin center (solid) and the bin edges (dashed). Regions in which the parameterization is not valid as calculated for the center of the bin are indicated with hatches.

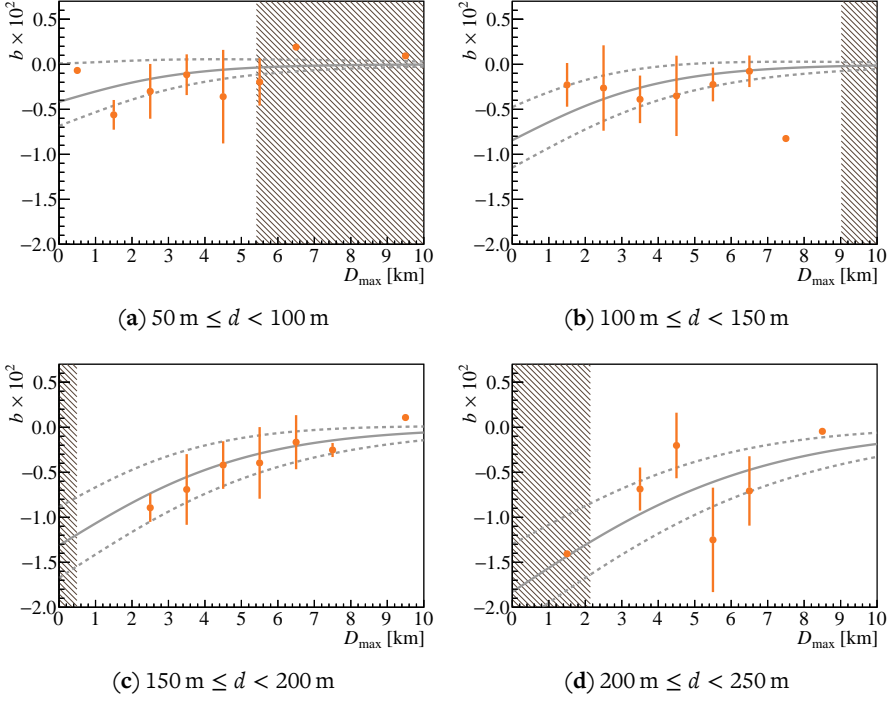


Figure 7.11: Average spectral index in bins of 1 km of D_{\max} for several bins of distance to shower axis d . The error bars indicate one standard deviation of the average in each bin. The data points without error bars contain only one measurement and the uncertainty could not be estimated. The lines and hatched regions have the same meaning as in figure 7.10.

of a total of n stations with a measurement of the spectral index:

$$\chi^2 = \sum_{i=0}^n \left(\frac{b_{\text{model},i}(D_{\max,i}(D_{\max,\text{core}})) - b_{\text{measured},i}}{\sigma_{b,i}} \right)^2, \quad (7.6)$$

where $b_{\text{measured},i}$ and $\sigma_{b,i}$ are the measured spectral index at station i and its assigned uncertainty. From equation (3.6) and the parameterized dependencies in chapter 3 we calculate the parameterized spectral index $b_{\text{model},i}$ as a function of the distance from the station to shower maximum $D_{\max,i}$ and find the best fitting value. The distances of each station to shower maximum within an event are coupled via the shower geometry and the only free parameter of the fit: the geometrical distance measured from the shower core to shower maximum $D_{\max,\text{core}}$. We derive the distance to the shower axis and observer angle for each station from the SD inflill reconstruction of the shower geometry.

To calculate a first estimate of the distance to shower maximum we only use the zenith angle of each shower and the average X_{\max} value of 700 g cm^{-2} , which we convert to a geometrical distance using the monthly average atmospheric density profiles. As fitting range we use the full valid range of the parameterization, which is between 0 km and 14 km.

Stations for which the lower uncertainty limit $b_{\text{measured},i} - \sigma_{b,i}$ 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 Φ_{obs} will be rejected from the fit. An iterative fit procedure makes sure stations are excluded when their fit result of D_{\max} places them outside the region of validity of the parameterization as defined by table 3.1. The fit is redone until all remaining stations are within the valid region or until there are no more stations left, when the fit is considered to have failed. Fits will also be rejected when they do not converge and the parameters are set to the fit range boundaries, which may happen when for instance the measured spectrum has a strong positive index. Afterwards, the fitted value of D_{\max} is converted to an atmospheric depth using the monthly density profiles to get the depth of shower maximum X_{\max} . Two examples of events with just a single station that has survived the selection criteria within the valid range of the parameterization are plotted in figure 7.12. The plots contain the measured spectral index with its uncertainty, and the value predicted by the parameterization when the distance to shower maximum from the FD measurement, $D_{\max, \text{FD}}$, is used.

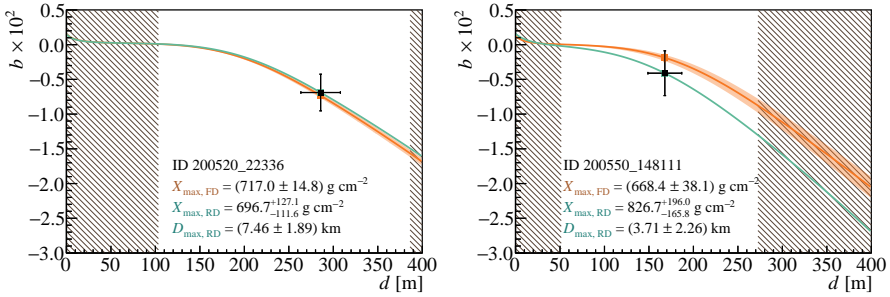


Figure 7.12: Two examples of events containing one station that survived the selection procedure. The measured spectral index is plotted in black, and is compared with the parameterized spectral index as calculated from the value of X_{\max} measured by FD (with the colored bands indicating the uncertainty) and the best fit of the parameterization to the measured index (RD). These were calculated using the observer angle as measured from SD over a wide range of distance to the shower axis d to show their dependence. The regions of d in which the parameterization is invalid for the fitted value of $D_{\max, \text{RD}}$, measured from the shower core impact location, are marked with hatches.

These plots demonstrate that a reconstruction of 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. This uncertainty is decreased if we perform the fit for events with more than just one station. Two examples of events where equation (7.6) is minimized for $n \geq 2$ stations are plotted in figure 7.13.

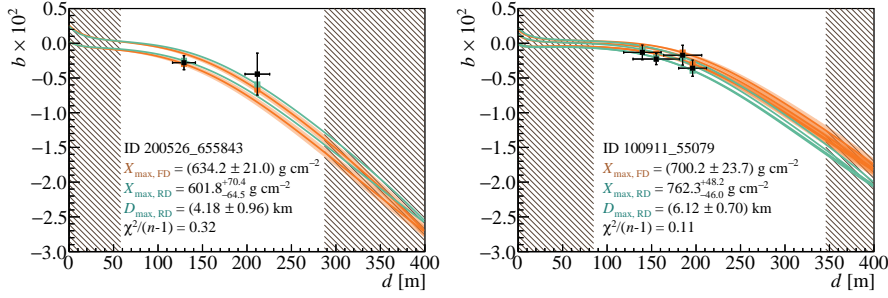


Figure 7.13: Two examples of events containing multiple stations that survived the selection procedure. For each station, the spectral index as predicted from the parameterization using the FD measurement of X_{\max} and the best fit of the combined RD event are drawn.

7.2.3 Comparison with fluorescence measurements

Of the 146 selected signal stations in the RD-FD data set, 94 stations can successfully be used to reconstruct X_{\max} with the parameterization of the spectral index. The failed reconstructions not only include single station events which fall outside the valid region of the parameterization or for which the fit does not converge, but also events with multiple stations in which stations are removed during the iterative fitting procedure. A total of 18 events with more than one signal station have a successful reconstruction of shower maximum using the parameterization of the spectral index. A comparison between the distance to shower maximum $D_{\max, \text{RD}}$ as it is reconstructed from the fit of the parameterization of the spectral index and the value derived from the FD measurement, $D_{\max, \text{FD}}$, is plotted in figure 7.14. Both geometrical distances were calculated relative to the shower core impact location.

The *Pearson correlation coefficient* (r) is calculated for the full data set, as well as for the subset of events with more than one signal station that could be used for the fit. As expected, the uncertainties of the fitted values of $D_{\max, \text{RD}}$ are smaller for the multi-station events. The correlation coefficient is also considerably higher for the multi-station subset. Because of the strong correlation between the zenith angle and the geometrical distance to shower maximum, it is better to compare the reconstructed values of depth of shower maximum, as is done in figure 7.15.

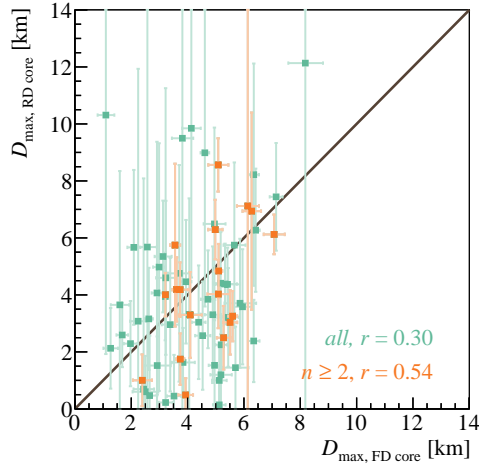


Figure 7.14: Geometrical distance of the shower core to shower maximum (D_{\max}) as measured by FD and from a fit of the parameterization of the spectral index (RD). A distinction is made between all events and events for which a combined fit of multiple signal stations ($n \geq 2$) is performed.

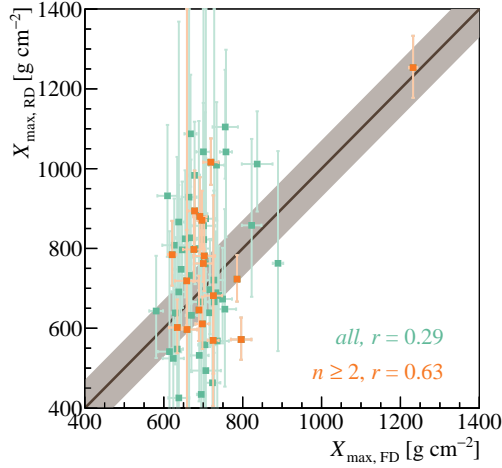


Figure 7.15: Comparison between X_{\max} measured by the FD and the best fit of the parameterization of the spectral index (RD). The shaded band marks the theoretical uncertainty of the parameterization of 71 g cm^{-2} for single station events. The plot contains the same events as figure 7.14, with again the distinction between events with single stations and events for which a combined fit could be performed.

The distributions of the difference between the reconstruction from the FD measurement and the fit of the parameterization of the spectral index to determine the depth of shower maximum are plotted in figure 7.16. From the rms of these distributions we can estimate the precision of our method compared to FD. For all events, the vast majority containing just one radio signal, the position of X_{\max} can be determined with a precision of 2.6 km relative to the FD measurement, which corresponds for this ensemble of events to an atmospheric depth of 168 g cm^{-2} . This is reduced to a resolution of X_{\max} of 138 g cm^{-2} if only events with more than one signal station are included. This resolution is considerably lower than the theoretical limit of 71 g cm^{-2} for single stations determined in chapter 3, but this was determined without the addition of noise and with a fully determined three-dimensional electric field.

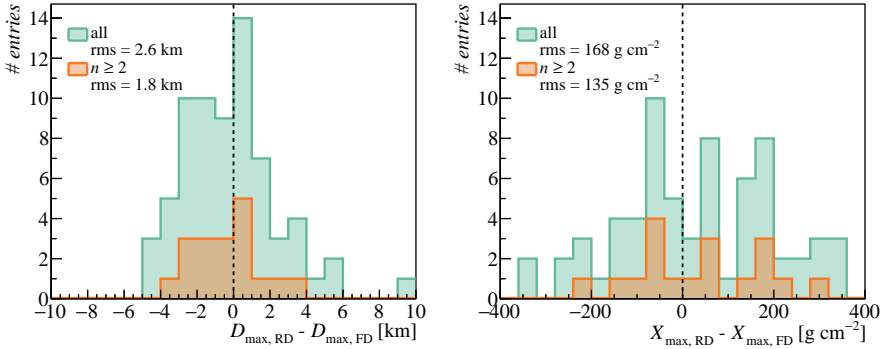


Figure 7.16: Distributions of the difference between the FD measurement and the fit of the parameterization of the spectral index (RD), for the geometrical distance to shower maximum D_{\max} (left) and atmospheric depth of shower maximum X_{\max} (right).

To test the reconstruction method and the estimation of the uncertainties we create a *pull distribution* of our results. The pull ΔX_{\max} of each reconstructed event is calculated using:

$$\Delta X_{\max} = \frac{X_{\max, \text{RD}} - X_{\max, \text{FD}}}{\sqrt{\sigma_{X_{\max, \text{RD}}}^2 + \sigma_{X_{\max, \text{FD}}}^2}}, \quad (7.7)$$

where $\sigma_{X_{\max, \text{RD}}}$ and $\sigma_{X_{\max, \text{FD}}}$ are the uncertainties of the fit using the parameterization of the spectral index and the uncertainty of the FD measurement of shower maximum respectively. The distribution of the pull of all events is plotted in figure 7.17, including a Gaussian fit of the distribution of all events including those with only one signal station. According to the mean of the Gaussian fit, there is a slight bias towards overestimation of the depth of shower maximum with the fit of the parameterization of the spectral index. The standard deviation of the Gaussian fit is close to unity, from which we can conclude that the uncertainties have been estimated reasonably well.

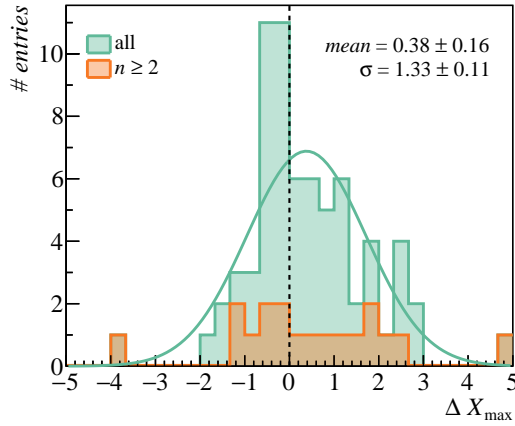


Figure 7.17: Pull distribution of the reconstruction of shower maximum from a fit of the parameterization in comparison with the value measured by FD. A Gaussian fit of the distribution of all events is drawn.

7.2.4 Average depth of shower maximum in AERA data

In the previous section we cross-checked the results from the event-by-event reconstruction of shower maximum using the parameterization of the spectral index with the FD measurements of X_{\max} . Now we will move on to calculate the depth of shower maximum for our full selected data set, which was described in chapter 5. This will enable us to perform a measurement of the average cosmic ray composition as a function of energy in the energy range of AERA, just as was described for the baseline Auger data in section 2.2.3. The advantage of using the full selected data set as compared to the RD-FD data set is that we have a much higher number of events. This is especially relevant for events with many stations, which provide a reconstruction of shower maximum with high precision. Several examples of these large events are displayed in figure 7.18. When compared to the events in figures 7.12 and 7.13, the uncertainty of X_{\max} is considerably lower.

Selection of high quality data

To investigate the average depth of shower maximum we need a high-quality and bias-free subset of the data. In figure 7.19, the average uncertainty of the geometrical distance to shower maximum (D_{\max}) as well as the atmospheric depth of shower maximum (X_{\max}) is plotted as a function of the final number of stations used in the fit of the parameterization of the spectral index, n . All selected events were used to create this plot, excluding events recorded during thunderstorm conditions and events for which the fit of the parameterization is considered a failure based on the criteria specified in

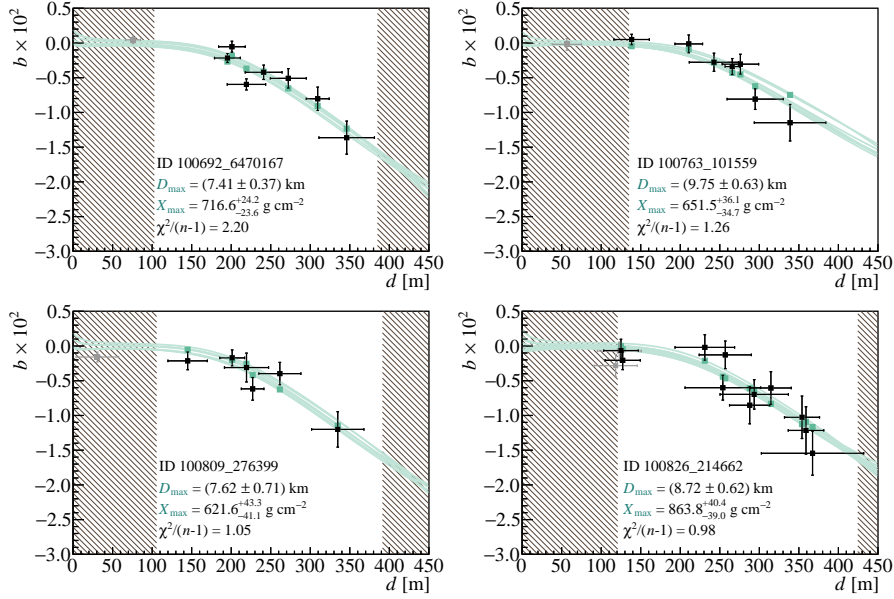


Figure 7.18: Several examples of events with more than five stations but no FD measurement, with a fit of the parameterization of the spectral index. The hatches and lines have the same meaning as in figure 7.12.

section 7.2.2. In the case of X_{\max} , the uncertainty is asymmetric because of the variable density in the atmosphere, so the average between the lower and upper uncertainty bounds is taken. This plot supports the notion that the uncertainty decreases as the number of stations per event in the reconstruction increases. At $n \geq 4$ the uncertainty on the position of shower maximum stabilizes and is less than 1 km. Therefore, we only use events with more than three stations in our analysis.

The selection criterion on the station number n reduces the data set to 200 events. We can further improve the quality of the data set by taking the fit quality into account. This is quantified by the minimum value of χ^2 from equation (7.6) divided by the number of degrees of freedom, which in our case is $n - 1$. The distribution of this value for the 200 remaining events is plotted in figure 7.20, which also includes a fit of a pure χ^2 probability density function. To prevent the inclusion of events which have a low fit quality, we only use events for which $\chi^2/(n - 1) < 10.0$.

A bias in the average depth of shower maximum is introduced by the zenith angle of the shower. An inclined shower will have traversed more atmosphere, and the average atmospheric depth of shower maximum will therefore be biased towards higher values. At the same time, vertical showers will have traversed less atmosphere, and the average

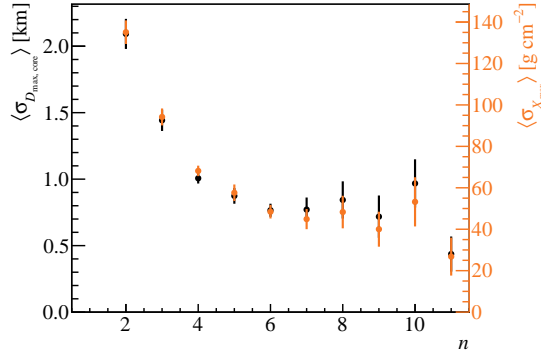


Figure 7.19: Average uncertainty of the geometrical distance to shower maximum D_{\max} and its atmospheric depth X_{\max} as a function of the final number of stations n used in the fit of the parameterization of the spectral index.

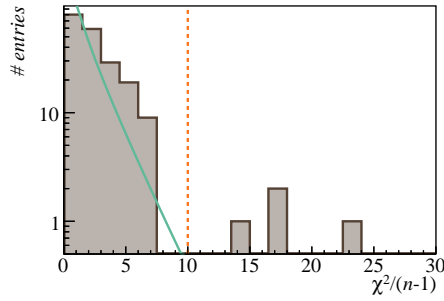


Figure 7.20: Distribution of χ^2 of the fit of the parameterization of the spectral index divided by the number of degrees of freedom, $n - 1$. A fit of a pure χ^2 probability density function is drawn by a solid line. Only data for which $\chi^2/(n - 1)$ is lower than 10.0 will be selected.

will therefore have a bias towards lower depths of shower maximum. This is illustrated in figure 7.21, where a profile histogram of the average depth of shower maximum as a function of zenith angle θ is plotted. An increase in the average depth of shower maximum as a function of zenith angle can be observed in the full range of zenith angles. To limit the bias, we select only events with zenith angles between 28° and 50° .

To study the depth of shower maximum as a function of energy we cluster the events in 10 logarithmic energy bins between 10^{17} eV and 10^{19} eV, for which we use the cosmic ray energy reconstruction of the SD. In each of the bins, the average depth of shower maximum is calculated. One thing we need to take into account is the efficiency of AERA in combination with our data selection method as a function of energy, as this might

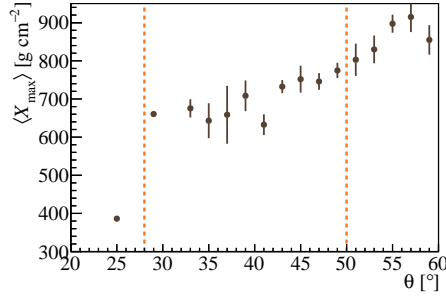


Figure 7.21: Average depth of shower maximum X_{\max} as a function of zenith angle θ . The bias is limited by selecting only events with zenith angles between 28° and 50° . The error bars represent the standard error on the average value.

also result in biases. When we consider the energy distribution – also in 10 logarithmic energy bins – of our selected events in figure 7.22, we notice that it peaks at $\sim 10^{18}$ eV. This means that below this energy, our sampling of the incoming cosmic rays is not fully efficient, and we measure mostly upward fluctuations of the real event distribution. In other words, we measure only the showers with a signal that barely exceeds the ambient noise level, while a large part of the other showers at these energies produce pulses which have too low power to make it through our selection criteria.

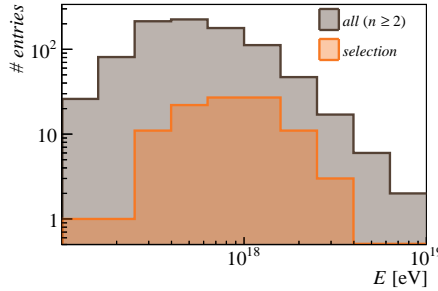


Figure 7.22: Distribution of the energy of the data set we have selected so far. As a comparison, the distribution of the energy of the initial set of events with $n \geq 2$ is also plotted.

A possible bias in the measurement of the average depth of shower maximum is difficult to investigate in the absence of simulations that incorporate a realistic noise environment. We can, however, qualitatively study the effect by using the scale parameter A from the spectral index fit of equation (3.2), which is related to the strength of the signal pulse. The average value of A as a function of D_{\max} in two rings of antennas at different distance d from our simulation study is plotted in figure 7.23. The charge-excess compo-

ment of the radio emission becomes much stronger when shower maximum is close to the ground, especially for observers near the shower axis. The strong pulses measured from showers with low D_{\max} will dominate the event sets in bins of low energies, which will result in a bias towards higher average X_{\max} values in these bins. We will therefore not include the energy range below $10^{17.8}$ eV in our data set.

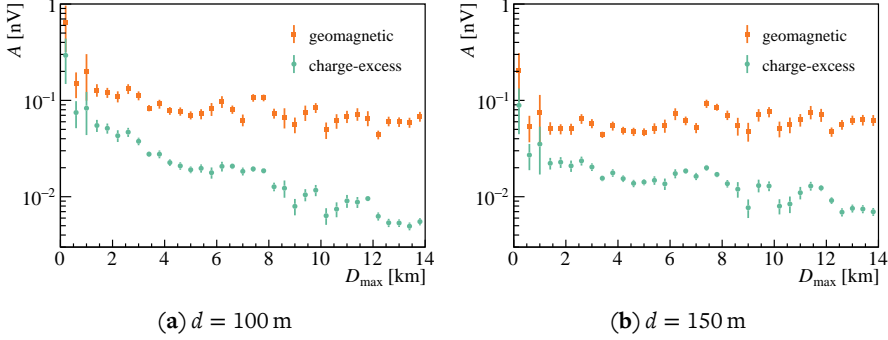


Figure 7.23: The average scale parameter A as a function of D_{\max} from all showers of the simulation study of chapter 3. Only antenna positions were used where the geomagnetic and charge-excess components could be completely separated.

This last step reduces the data set to a total of 68 events distributed over four energy bins. An overview of all data selection steps is visible in table 7.1.

Table 7.1: Selected number of events after the application of each selection cut, for the measurement of the average depth of shower maximum.

$n \geq 4$	200
$\chi^2/(n-1) < 10.0$	196
$28^\circ < \theta < 50^\circ$	104
$E > 10^{17.8}$ eV	68

Of the 68 selected events, 10 events were recorded in periods when the weather station was down. We include these events because the chance that a significant number of them is affected by thunderstorms is small due to our selection cut on the polarization offset β . The distribution of the depth of shower maximum of the selected events is plotted in figure 7.24. The distributions of depth of shower maximum X_{\max} in each of the four energy bins are plotted in figure 7.25.

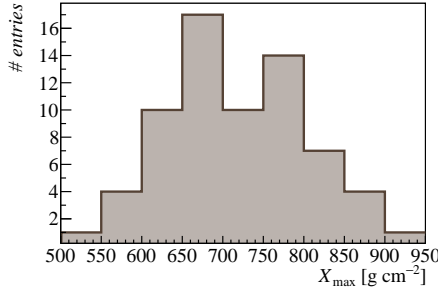


Figure 7.24: Distribution of the depth of shower maximum of the final data set.

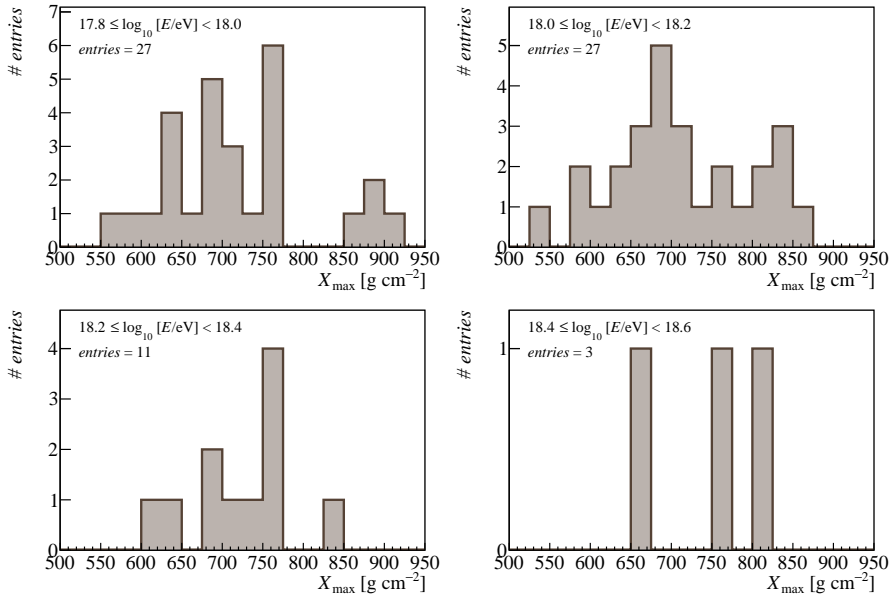


Figure 7.25: Distributions of the reconstructed values of X_{\max} in the four energy bins that still contain events after all selection steps.

Estimation of systematic uncertainties

Several sources of systematic uncertainty of the average value of X_{\max} can be identified. A possibly important source of systematic uncertainty is the frequency dependent gain calibration of the antenna. We can estimate the effect on the spectral index by creating a toy Monte Carlo model that randomly fluctuates the spectrum according to a Gaussian distribution described by the average uncertainties from the calibration (plotted in figure 6.20(b)) as a function of frequency. The uncertainty is in the order of 5 % in the full frequency range. We use the uncertainties from the calibration of the externally

triggered setup because the vast majority of the events in our high quality data set are from this experimental setup due to our cut on the minimum number of stations. For each of the four energy bins we propagate the average value of X_{max} through the atmosphere using the average values of d and θ calculated in the energy bin to get an average value of the distance to shower maximum. We then use the same average shower geometry to calculate an average value of the spectral index b in the energy bin. Using equation (3.2) we convert this spectral index into a spectrum with arbitrary scale parameter, which we modify with the Monte Carlo according to the uncertainties of the calibration and fit the spectral index again. A distribution of the difference δb between the original spectral index and the index of the modified spectrum after generating one million spectra is plotted in figure 7.26.

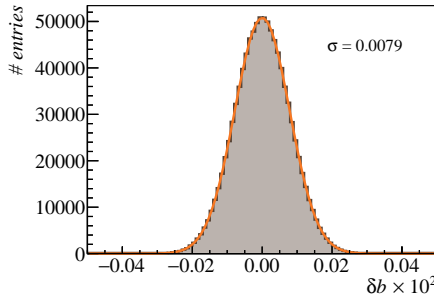


Figure 7.26: Distribution of the difference δb between the original spectral index and the spectral index of the modified spectrum after one million runs of the Monte Carlo, calculated for the average spectral index of the energy bin between $10^{18.4}$ eV and $10^{18.6}$ eV. A Gaussian fit is drawn as a solid line.

We use the standard deviation of a Gaussian fit to this distribution as the systematic uncertainty of the spectral index resulting from the gain calibration. The uncertainties of the average spectral index were very similar for all energy bins. Using the average geometry of the energy bin, we then use the parameterization again and propagate back through the atmosphere to recover the uncertainty limits of $\langle X_{\text{max}} \rangle$. The systematic uncertainties resulting from the gain calibration are listed in table 7.2 for all four energy bins.

A further potential source of systematic uncertainty is the spectral noise correction. We estimated the correction for the noise background in section 7.1.2 with a Monte Carlo which assumed a flat noise spectrum. This estimation is only correct if the noise in the measured data can also be described by a flat spectrum. The distribution of the spectral index of a fit to the spectrum of a 400 ns noise trace from all selected events from chapter 5 is displayed in figure 7.27. The mean value of a Gaussian fit to this distribution is -0.05×10^{-2} . From this we deduce that, even though the galactic noise imprint has

a distinct spectral shape, this is drowned in random noise in individual measurements, which is on average well described by a flat spectrum. The method of narrowband RFI suppression – which was described in section 6.3 – was applied to the noise traces to which the spectral index was fitted. This means that this method to suppress RFI does not result in a significant bias in the spectral index either. We conclude therefore that the several methods of noise correction do not contribute substantially to the systematic uncertainty of the measurement of the average depth of shower maximum, and the fluctuations in the noise spectra are small with respect to the statistical uncertainties, and we neglect this as a possible source of systematic uncertainty.

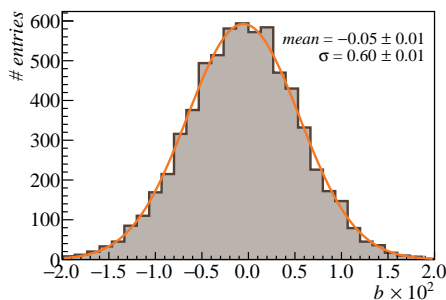


Figure 7.27: Distribution of the spectral index calculated in a 400 ns noise trace for all selected events. A fit of a Gaussian function is drawn as a solid line.

Finally, the parameterization of the spectral index itself is a source of systematic uncertainty. Only the limited number of events in the RD-FD set could be used in section 7.2.3 to test the parameterization. It is therefore difficult to determine the uncertainty of the parameterization from the comparison with data. We can instead use the bias in X_{\max} resulting from the application of the parameterization to the simulations, as was determined in section 3.5. Depending on the observer angle and distance to the shower axis, the bias in the depth of shower maximum is between 10.0 g cm^{-2} and 20.0 g cm^{-2} for reconstructions using information from a single station, as can be seen in figure 3.17. As a conservative estimate, we will use 20.0 g cm^{-2} as the systematic uncertainty resulting from the parameterization of the spectral index.

The two contributions to the systematic uncertainty are independent and are therefore added quadratically to obtain the total systematic uncertainty. An overview of all systematic uncertainties as a function of energy bin is listed in table 7.2.

Average depth of shower maximum

The average value of X_{\max} as a function of energy is plotted in figure 7.28, where it is compared with measurements of shower maximum made by the fluorescence detector (see also figure 2.8). In addition, it is compared with the average X_{\max} as a function of

Table 7.2: Overview of systematic uncertainties of the average depth of shower maximum as a function of bin energy in the selected high quality data set.

$\log_{10}[E/\text{eV}]$	17.9	18.1	18.3	18.5
gain calibration [g cm^{-2}]	4.5	4.2	3.7	3.4
parameterization [g cm^{-2}]	20.0	20.0	20.0	20.0
total [g cm^{-2}]	20.5	20.4	20.3	20.3

energy produced by three different interaction models for showers with proton and iron primaries. The statistical uncertainty of the average value can be estimated by calculating the standard error of the average, and is represented in the plot by the error bars.

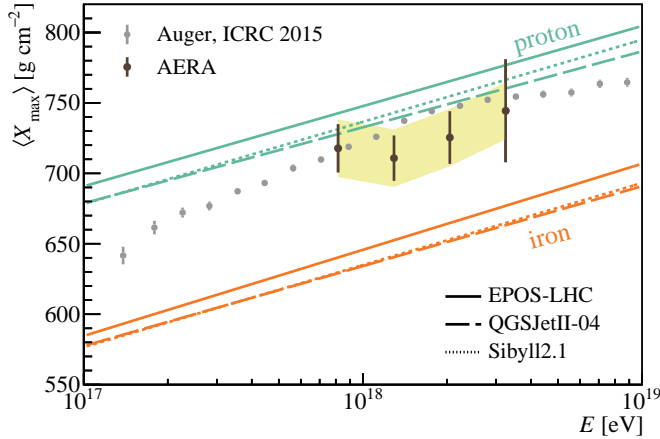


Figure 7.28: Average depth of shower maximum as function of cosmic ray energy as measured by AERA using the spectral index compared with the fluorescence measurements from the Pierre Auger observatory [33]. The error bars of the AERA data points represent the standard error of the average. The colored band indicates the systematic uncertainty. Predictions of the depth of shower maximum of iron and proton primaries from three different interaction models are plotted as lines.

The average depth of shower maximum as a function of energy as measured with AERA using the parameterization of the spectral index of the radio pulse is compatible with the measurements of the fluorescence detector of the Pierre Auger Observatory. Both measurements are consistent with a predominantly light cosmic ray composition in the energy range from $10^{17.8}$ eV to $10^{18.6}$ eV.

Discussion and outlook

IN THE PREVIOUS CHAPTER, we used AERA data to demonstrate that the spectral index of the radio pulse is sensitive to air shower development. We have used a fit of the parameterization of the spectral index derived in chapter 3 to reconstruct the depth of shower maximum and compared the results with FD measurements. This confirmed the possibility to perform a measurement of shower development by using data from a single radio station, provided that the shower geometry is known.

8.1 Discussion of results

Compared to FD measurements, however, the precision of these measurements was shown to be low, with a standard deviation of 168 g cm^{-2} . This precision was improved by performing a fit of the parameterization using events with multiple signal stations, as was discussed in section 7.2.3. Due to the low number of events however, only a very limited sample of multi-station events could be compared to FD measurements. By using mostly events with two and three stations, we were able to improve the precision to 135 g cm^{-2} . As the average uncertainty of the reconstructed position of shower maximum plotted in figure 7.19 shows, we can expect the precision to reach about 60 g cm^{-2} for events with four stations or more in the current experimental setting.

From the pull distribution of figure 7.17 we can deduce that the presented experimental method is sound, and that the uncertainties are properly estimated. In addition to increasing the number of measured signals per event, we should therefore also strive to decrease the measurement uncertainties to increase the precision. One important contribution to the measurement uncertainties is the uncertainty of the SD shower geometry. In particular the uncertainty of the core impact location, which is related to the station distance to the shower axis, one of the dependencies in the parameterization of the spectral index. The median uncertainty in the distribution of figure 7.6 is 25 m, a distance over which the spectral index can change significantly, as was demonstrated

in the simulation study and is shown in figure 3.11. Because the radio detectors are distributed on a denser grid than the SD infill array, using their timing and lateral signal strength distribution when reconstructing the shower geometry might improve the precision of the shower core impact location measurement.

More important even is the intrinsic uncertainty introduced by reconstruction of the three-dimensional electric field from a measurement in two dimensions in a noisy environment. This effect was demonstrated by a toy Monte Carlo in sections 4.3.4 and 7.1.2. It will be difficult to overcome this noise contribution without adding a third antenna arm to measure all three components of the electric field, which means redesigning the detector.

By using a subset of 68 events with four or more signal stations to decrease the influence of the measurement uncertainties, we were able to calculate the average depth of shower maximum as function of the cosmic ray energy. The average depth of shower maximum is in agreement with measurements from the FD, and is compatible with a predominantly light composition.

Unfortunately, we were not able to perform a fully data driven analysis and parameterization of the dependencies of the spectral index of the radio pulse. This would have produced a method independent of models, but could also have been used to verify and possibly confirm the results acquired from the full Monte Carlo simulations described in chapter 3. We were, however, severely limited by the small size of the RD-FD coincident data set. The causes of this small data volume are manifold. On the one hand, there is the scintillator-triggered setup which has a spacing of 375 m, which is too sparse to sample the interesting region of the air shower with high efficiency. This was aggravated because a large fraction of the stations from this experimental setup was down at any time within our period of data collection, as was shown in section 5.2.2. On the other hand, the more densely spaced part of the array, mostly consisting of the externally triggered setup, is located too close to the fluorescence detector to capture shower maximum within the field of view of the telescopes, as was discussed in section 5.4.5. This is especially true if the HEAT telescopes are tilted downwards, which was for example the case in November 2014. This led to the majority of radio events coincident with a fluorescence measurement to be rejected outright based on the FD selection criteria.

8.2 Outlook and recommendations

The parameterization of the spectral index of the radio pulse unveils interesting new possibilities to study air shower properties through their radio emission. Until now, the frequency content of the radio signal has hardly been utilized to reconstruct air shower parameters. Because of its sensitivity to shower geometry, the spectral index can also be used to reconstruct the shower core and axis in addition to the geometrical distance

to shower maximum, which was explored in this thesis. Using the spectral index in this context would provide an independent measurement of shower geometry, which is complementary to existing radio techniques using the pulse timing and the lateral signal power distribution (see also section 2.3.3).

As was demonstrated in the study of the lateral distribution of the radio signal power (see e.g. [64]), the important features in the radio signal of the air shower are located within the first few hundred meters from the shower axis. This is no less true for the spectral index, the features of which flatten off after about 350 m, as could be seen in figure 3.11. Future radio observatories employing these methods should therefore have antenna spacings of not much more than 200 m to adequately sample this region. To decrease the uncertainties of radio measurements in general, efforts should be directed towards designing and modeling antennas to accurately measure the electric field in all three dimensions.

BIBLIOGRAPHY

- [1] V. F. Hess, *Über Beobachtungen der durchdringenden Strahlung bei sieben Freiballonfahrten*, *Z. Phys.* **13** (1912) 1084 (cited on page 5).
- [2] P. Auger et al., *Extensive cosmic-ray showers*, *Rev. Mod. Phys.* **11** no. 3–4 (1939) 288 (cited on page 5).
- [3] O. Adriani et al. (PAMELA Collaboration), *Cosmic-ray electron flux measured by the PAMELA experiment between 1 and 625 GeV*, *Phys. Rev. Lett.* **106** no. 20 (2011) 201101 (cited on page 6).
- [4] T. Sanuki et al. (BESS Collaboration), *Precise measurement of cosmic-ray proton and helium spectra with the BESS spectrometer*, *Astrophys. J.* **545** no. 2 (2000) 1135 (cited on page 6).
- [5] R. Decker et al., *Voyager 1 in the foreshock, termination shock, and heliosheath*, *Science* **309** no. 5743 (2005) 2020–2024 (cited on page 6).
- [6] S. Swordy, *The energy spectra and anisotropies of cosmic rays*, in *The Astrophysics of Galactic Cosmic Rays*. Springer, 2001 (cited on page 6).
- [7] M. A. Lawrence et al., *The cosmic ray energy spectrum above 4×10^{17} eV as measured by the Haverah Park array*, *J. Phys. G Nucl. Partic.* **17** no. 5 (1991) 733 (cited on page 6).
- [8] N. Chiba et al. (AGASA Collaboration), *Akeno giant air shower array (AGASA) covering 100 km² area*, *Nucl. Instrum. Meth. A* **311** no. 1 (1992) 338–349 (cited on page 6).
- [9] V. P. Egorova et al., *The spectrum features of UHECRs below and surrounding GZK*, *Nucl. Phys. B-Proc. Sup.* **136** (2004) 3–11 (cited on page 6).
- [10] R. U. Abbasi et al. (High Resolution Fly’s Eye Collaboration), *Measurement of the flux of ultrahigh energy cosmic rays from monocular observations by the High Resolution Fly’s Eye experiment*, *Phys. Rev. Lett.* **92** no. 15 (2004) 151101 (cited on page 6).
- [11] M. Nagano and A. A. Watson, *Observations and implications of the ultrahigh-energy cosmic rays*, *Rev. Mod. Phys.* **72** no. 3 (2000) 689 (cited on page 7).
- [12] J. R. Hörandel, *On the knee in the energy spectrum of cosmic rays*, *Astropart. Phys.* **19** no. 2 (2003) 193–220 (cited on page 7).
- [13] I. Valiño for the Pierre Auger Collaboration, *The flux of ultra-high energy cosmic rays after ten years of operation of the Pierre Auger Observatory*, in *Proc. of the 34th ICRC*. The Hague, Netherlands, 2015 (cited on page 7).
- [14] A. M. Hillas, *Cosmic rays: Recent progress and some current questions*, *arXiv preprint: 0712.2832* (2006) (cited on page 8).
- [15] T. K. Gaisser et al., *Cosmic ray energy spectrum from measurements of air showers*, *Front. Phys.* **8** no. 6 (2013) 748–758 (cited on page 8).
- [16] M. Unger et al., *Origin of the ankle in the ultra-high energy cosmic ray spectrum and of the extragalactic protons below it*, *arXiv preprint: 1505.02153* (2015) (cited on page 8).

- [17] V. Berezhinsky, *Extragalactic cosmic rays and their signatures*, *Astropart. Phys.* **53** (2014) 120–129 (cited on page 8).
- [18] A. di Matteo for the Pierre Auger Collaboration, *Combined fit of spectrum and composition data as measured by the Pierre Auger Observatory*, in *Proc. of the 34th ICRC*. The Hague, Netherlands, 2015 (cited on page 8).
- [19] E. Fermi, *On the origin of the cosmic radiation*, *Phys. Rev.* **75** no. 8 (1949) 1169 (cited on page 8).
- [20] A. Hillas, *The origin of ultra-high-energy cosmic rays*, *Annu. Rev. Astron. Astr.* **22** (1984) 425–444 (cited on pages 8 and 9).
- [21] K. Greisen, *End to the cosmic-ray spectrum?*, *Phys. Rev. Lett.* **16** no. 17 (1966) 748 (cited on page 9).
- [22] G. T. Zatsepin and V. A. Kuz'min, *Upper limit of the spectrum of cosmic rays*, *Zh. Eksp. Teor. Fiz.* **4** (1966) 114 (cited on page 9).
- [23] D. Harari et al., *On the ultrahigh energy cosmic ray horizon*, *J. Cosmol. Astropart. P.* **2006** no. 11 (2006) 012 (cited on pages 9 and 10).
- [24] P. P. Kronberg, *Extragalactic magnetic fields*, *Rep. Prog. Phys.* **57** no. 4 (1994) 325 (cited on page 10).
- [25] M. Risse, *Properties of extensive air showers*, *arXiv preprint: astro-ph/0402300* (2004) (cited on pages 12 and 14).
- [26] D. Heck et al., *CORSIKA: A Monte Carlo code to simulate extensive air showers*, FZKA-6019, Forschungszentrum Karlsruhe, 1998 (cited on pages 12 and 21).
- [27] S. Sciutto, *AIRES: A system for air shower simulations (version 2.2.0)*, *arXiv preprint: astro-ph/9911331* (1999) (cited on page 12).
- [28] R. Engel et al., *Extensive air showers and hadronic interactions at high energy*, *Annu. Rev. Nucl. Part. S.* **61** (2011) 467–489 (cited on page 12).
- [29] W. Heitler, *The quantum theory of radiation*, vol. 86. Courier Corporation, 1954 (cited on page 13).
- [30] 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 15th ICRC*. Plovdiv, Bulgaria, 1977 (cited on page 14).
- [31] R. U. Abbasi et al. (Telescope Array Collaboration), *Study of ultra-high energy cosmic ray composition using Telescope Array's middle drum detector and surface array in hybrid mode*, *Astroparticle Physics* **64** (2015) 49–62 (cited on page 14).
- [32] M. Unger for the Pierre Auger Collaboration and the Telescope Array Collaboration, *Report of the Working Group on the Composition of Ultra-High Energy Cosmic Rays*, in *Proc. of the 34th ICRC*. The Hague, Netherlands, 2015 (cited on page 14).
- [33] A. Porcelli for the Pierre Auger Collaboration, *Measurements of X_{\max} above 10^{17} eV with the fluorescence detector of the Pierre Auger Observatory*, in *Proc. of the 34th ICRC*. The Hague, Netherlands, 2015 (cited on pages 15 and 129).

- [34] T. Pierog and K. Werner, *Muon production in extended air shower simulations*, *Phys. Rev. Lett.* **101** no. 17 (2008) 171101 (cited on page 15).
- [35] A. Aab et al. (Pierre Auger Collaboration), *Muons in air showers at the Pierre Auger Observatory: Measurement of atmospheric production depth*, *Phys. Rev. D* **90** no. 1 (2014) 012012 (cited on page 15).
- [36] J. Jelley et al., *Radio pulses from extensive cosmic-ray air showers*, *Nature* **205** (1965) 327–328 (cited on page 15).
- [37] F. D. Kahn and I. Lerche, *Radiation from cosmic ray air showers*, *P. Roy. Soc. Lond. A Mat.* **289** no. 1417 (1966) 206–213 (cited on page 15).
- [38] G. Askar’yan, *Excess negative charge of an electron-photon shower and its coherent radio emission*, *Zh. Eksp. Teor. Fiz.* **41** (1961) 616–618 (cited on page 16).
- [39] H. Allan et al., *Radio pulses from extensive air showers*, *Nature* **227** no. 5263 (1970) 1116–1118 (cited on page 16).
- [40] D. Ardouin et al. (CODALEMA Collaboration), *Radio-detection signature of high-energy cosmic rays by the CODALEMA experiment*, *Nucl. Instrum. Meth. A* **555** no. 1 (2005) 148–163 (cited on page 16).
- [41] H. Falcke et al. (LOPES Collaboration), *Detection and imaging of atmospheric radio flashes from cosmic ray air showers*, *Nature* **435** no. 7040 (2005) 313–316 (cited on page 16).
- [42] A. Aab et al. (Pierre Auger Collaboration), *Probing the radio emission from air showers with polarization measurements*, *Phys. Rev. D* **89** no. 5 (2014) 052002 (cited on pages 16 and 75).
- [43] O. Scholten et al., *A macroscopic description of coherent geo-magnetic radiation from cosmic-ray air showers*, *Astropart. Phys.* **29** no. 2 (2008) 94–103 (cited on page 17).
- [44] K. Werner et al., *A realistic treatment of geomagnetic Cherenkov radiation from cosmic ray air showers*, *Astropart. Phys.* **37** (2012) 5–16 (cited on page 17).
- [45] T. Huege et al., *Simulating radio emission from air showers with COREAS*, in *Proc. of the ARENA2012 Conf.* Erlangen, Germany, 2013 (cited on page 17).
- [46] J. Alvarez-Muñiz et al., *Monte Carlo simulations of radio pulses in atmospheric showers using ZHAires*, *Astropart. Phys.* **35** (2012) 325–341 (cited on pages 17 and 28).
- [47] J. D. Jackson, *Classical electrodynamics*. Wiley New York, 1962 (cited on page 17).
- [48] C. W. James et al., *General description of electromagnetic radiation processes based on instantaneous charge acceleration in “endpoints”*, *Phys. Rev. E* **84** no. 5 (2011) 056602 (cited on page 17).
- [49] E. Zas et al., *Electromagnetic pulses from high-energy showers: Implications for neutrino detection*, *Phys. Rev. D* **45** no. 1 (1992) 362 (cited on page 17).
- [50] T. Huege, *Theory and simulations of air shower radio emission*, *arXiv preprint: arXiv:1301.2135* (2013) (cited on page 17).
- [51] C. W. James, *Electromagnetic radiation in the Tamm problem*, in *Proc. of the ARENA2012 Conf.* Erlangen, Germany, 2013 (cited on page 18).
- [52] V. Bugaev et al., *Comparison of ZHAires and COREAS radio emission simulations in the Ultra-High Frequency band*, in *Proc. of the 33rd ICRC*. Rio de Janeiro, Brazil, 2013 (cited on page 18).

- [53] K. D. de Vries et al., *Coherent Cherenkov radiation from cosmic-ray-induced air showers*, *Phys. Rev. Lett.* **107** no. 6 (2011) 061101 (cited on pages 18 and 30).
- [54] A. Corstanje et al. (LOFAR Collaboration), *The shape of the radio wavefront of extensive air showers as measured with LOFAR*, *Astropart. Phys.* **61** (2015) 22–31 (cited on page 18).
- [55] A. Nelles et al., *A parameterization for the radio emission of air showers as predicted by COREAS simulations and applied to LOFAR measurements*, *Astropart. Phys.* **60** (2015) 13–24 (cited on page 18).
- [56] C. Glaser for the Pierre Auger Collaboration, *The energy content of extensive air showers in the radio frequency range of 30–80 MHz*, in *Proc. of the 34th ICRC*. The Hague, Netherlands, 2015 (cited on pages 18 and 24).
- [57] F. G. Schröder, *Investigation of the radio wavefront of air showers with LOPES and REAS3*, in *Proc. of the 32nd ICRC*. Beijing, China, 2011 (cited on page 18).
- [58] W. Apel et al. (LOPES Collaboration), *The wavefront of the radio signal emitted by cosmic ray air showers*, *J. Cosmol. Astropart. P.* **2014** no. 09 (2014) 025 (cited on page 18).
- [59] S. Buitink et al., *Shower X_{\max} determination based on LOFAR radio measurements*, in *Proc. of the 33rd ICRC*. Rio de Janeiro, Brazil, 2013 (cited on page 18).
- [60] J. Schulz, *Cosmic radiation, reconstruction of cosmic-ray properties from radio emission of extensive air showers*. PhD thesis, Radboud University Nijmegen, 2016 (cited on page 18).
- [61] S. Grebe, *Finger on the pulse of cosmic rays*. PhD thesis, Radboud University Nijmegen, 2013 (cited on pages 18, 19, 22, 59, 83, 85, 86, 100, and 107).
- [62] NOAA and US Air Force, *US standard atmosphere*, 76–1562, NOAA-S/T, 1976 (cited on page 21).
- [63] T. Huege et al., *The convergence of EAS radio emission models and a detailed comparison of REAS3 and MGMR simulations*, *Nucl. Instrum. Meth. A* **662** (2012) S179–S186 (cited on page 22).
- [64] A. Nelles, *Radio emission of air showers: The perspective of LOFAR and AERA*. PhD thesis, Radboud University Nijmegen, 2014 (cited on pages 24 and 133).
- [65] A. Fasso et al., *New developments in FLUKA modelling hadronic and EM interactions*, in *Proc. of the SARE3 Conf.* Tsukuba, Japan, 1997 (cited on page 24).
- [66] N. N. Kalmykov and S. S. Ostapchenko, *The nucleus-nucleus interaction, nuclear fragmentation, and fluctuations of extensive air showers*, *Phys. Atom. Nucl.* **56** (1993) 346–353 (cited on page 24).
- [67] F. Suarez for the Pierre Auger Collaboration, *The AMIGA muon detectors of the Pierre Auger Observatory: overview and status*, in *Proc. of the 33rd ICRC*. Rio de Janeiro, Brazil, 2013 (cited on pages 25 and 42).
- [68] H. Schoorlemmer, *Tuning in on cosmic rays*. PhD thesis, Radboud University Nijmegen, 2012 (cited on pages 32, 65, 68, and 75).
- [69] J. Abraham et al. (Pierre Auger Collaboration), *Properties and performance of the prototype instrument for the Pierre Auger Observatory*, *Nucl. Instrum. Meth. A* **523** no. 1 (2004) 50–95 (cited on page 39).

- [70] I. Allekotte et al. (Pierre Auger Collaboration), *Site survey for the Pierre Auger Observatory*, *J. Phys. G Nucl. Partic.* **28** no. 6 (2002) 1499–1510 (cited on page 39).
- [71] T. Suomijarvi for the Pierre Auger Collaboration, *Performance of the Pierre Auger Observatory surface detector*, in *Proc. of the 30th ICRC*. Merida, Mexico, 2007 (cited on page 39).
- [72] C. Bonifazi for the Pierre Auger Collaboration, *The monitoring system of the Pierre Auger Observatory: on-line and long-term data quality controls*, in *Proc. of the 33rd ICRC*. Rio de Janeiro, Brazil, 2013 (cited on page 39).
- [73] M. Healy for the Pierre Auger Collaboration, *Composition-sensitive parameters measured with the surface detector of the Pierre Auger Observatory*, in *Proc. of the 30th ICRC*. Merida, Mexico, 2007 (cited on page 39).
- [74] G. Aar, *TbD*. PhD thesis, Radboud University Nijmegen, 2016 (cited on page 39).
- [75] J. Abraham et al. (Pierre Auger Collaboration), *The fluorescence detector of the Pierre Auger Observatory*, *Nucl. Instrum. Meth. A* **620** no. 2 (2010) 227–251 (cited on pages 39 and 43).
- [76] 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** no. 12 (2014) 122005 (cited on pages 40, 79, 144, and 154).
- [77] I. Allekotte et al. (Pierre Auger Collaboration), *The surface detector system of the Pierre Auger Observatory*, *Nucl. Instrum. Meth. A* **586** (2008) 409–420 (cited on page 40).
- [78] J. Abraham et al. (Pierre Auger Collaboration), *Trigger and aperture of the surface detector array of the Pierre Auger Observatory*, *Nucl. Instrum. Meth. A* **613** no. 1 (2010) 29–39 (cited on page 41).
- [79] M. Roth for the Pierre Auger Collaboration, *The lateral distribution function of shower signals in the surface detector of the Pierre Auger Observatory*, in *Proc. of the 28th ICRC*. Tsukuba, Japan, 2003 (cited on page 41).
- [80] D. Newton et al., *The optimum distance at which to determine the size of a giant air shower*, *Astropart. Phys.* **26** no. 6 (2007) 414–419 (cited on page 42).
- [81] A. Schulz for the Pierre Auger Collaboration, *The measurement of the energy spectrum of cosmic rays above 3×10^{17} eV with the Pierre Auger Observatory*, in *Proc. of the 33rd ICRC*. Rio de Janeiro, Brazil, 2013 (cited on page 42).
- [82] F. Sanchez for the Pierre Auger Collaboration, *The AMIGA detector of the Pierre Auger Observatory: an overview*, in *Proc. of the 32nd ICRC*. Beijing, China, 2011 (cited on page 42).
- [83] I. Maris for the Pierre Auger Collaboration, *The AMIGA infill detector of the Pierre Auger Observatory: performance and first data*, in *Proc. of the 32nd ICRC*. Beijing, China, 2011 (cited on page 42).
- [84] M. Kleifges for the Pierre Auger Collaboration, *Extension of the Pierre Auger Observatory using High-ElevAtion fluorescence Telescopes (HEAT)*, in *Proc. of the 31st ICRC*. Łódź, Poland, 2009 (cited on page 43).
- [85] T. Mathes for the Pierre Auger Collaboration, *The HEAT telescopes of the Pierre Auger Observatory status and first data*, in *Proc. of the 32nd ICRC*. Beijing, China, 2011 (cited on page 43).

- [86] P. Sommers, *Capabilities of a giant hybrid air shower detector*, *Astropart. Phys.* **3** no. 4 (1995) 349–360 (cited on page 44).
- [87] F. Nerling et al., *Universality of electron distributions in high-energy air showers – Description of Cherenkov light production*, *Astropart. Phys.* **24** no. 6 (2006) 421–437 (cited on page 44).
- [88] 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** no. 9 (2012) 591–607 (cited on page 44).
- [89] J. Abraham et al. (Pierre Auger Collaboration), *A study of the effect of molecular and aerosol conditions in the atmosphere on air fluorescence measurements at the Pierre Auger Observatory*, *Astropart. Phys.* **33** no. 2 (2010) 108–129 (cited on page 44).
- [90] Pierre Auger Collaboration, *Techniques for measuring aerosol attenuation using the Central Laser Facility at the Pierre Auger Observatory*, *J. Instrum.* **8** no. 04 (2013) P04009 (cited on page 44).
- [91] S. Y. BenZvi et al., *The Lidar system of the Pierre Auger Observatory*, *Nucl. Instrum. Meth. A* **574** no. 1 (2007) 171–184 (cited on page 44).
- [92] J. Chirinos for the Pierre Auger Collaboration, *Cloud monitoring at the Pierre Auger Observatory*, in *Proc. of the 33rd ICRC*. Rio de Janeiro, Brazil, 2013 (cited on page 44).
- [93] A. Schmidt et al., *Third level trigger for the fluorescence telescopes of the Pierre Auger Observatory*, *Nucl. Instrum. Meth. A* **601** no. 3 (2009) 347–353 (cited on page 44).
- [94] S. Acounis et al., *Results of a self-triggered prototype system for radio-detection of extensive air showers at the Pierre Auger Observatory*, *J. Instrum.* **7** no. 11 (2012) P11023 (cited on page 45).
- [95] J. Coppens for the Pierre Auger Collaboration, *Observation of radio signals from air showers at the Pierre Auger Observatory*, *Nucl. Instrum. Meth. A* **604** no. 1 (2009) S41–S43 (cited on page 45).
- [96] 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** no. 10 (2012) P10011 (cited on pages 45 and 47).
- [97] D. Charrier et al. (CODALEMA Collaboration), *Antenna development for astroparticle and radioastronomy experiments*, *Nucl. Instrum. Meth. A* **662** (2012) S142–S145 (cited on page 45).
- [98] K. Weidenhaupt, *Antenna calibration and energy measurement of ultra-high energy cosmic rays with the Auger Engineering Radio Array*. PhD thesis, RWTH Aachen, 2014 (cited on pages 45 and 92).
- [99] C. Rühle, *Entwicklung eines schnellen eingebetteten Systems zur Radiodetektion kosmischer Strahlung*. PhD thesis, Karlsruher Instituts für Technologie, 2014 (cited on page 48).
- [100] A. Schmidt, *Realization of a Self-triggered Detector for the Radio Emission of Cosmic Rays*. PhD thesis, Karlsruher Instituts für Technologie, 2011 (cited on page 49).
- [101] J. Kelley for the Pierre Auger Collaboration, *Data acquisition, triggering, and filtering at the Auger Engineering Radio Array*, *Nucl. Instrum. Meth. A* **725** (2013) 133–136 (cited on page 49).

- [102] S. Argiro et al., *The Offline software framework of the Pierre Auger Observatory*, *Nucl. Instrum. Meth. A* **580** no. 3 (2007) 1485–1496 (cited on page 52).
- [103] 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** no. 1 (2011) 92–102 (cited on page 52).
- [104] S. Fliescher, *Measurement of radio emission from cosmic ray induced air showers at the Pierre Auger Observatory*. PhD thesis, RWTH Aachen, 2011 (cited on pages 52, 59, 92, 93, and 95).
- [105] F. G. Schröder, *Instruments and methods for the radio detection of high energy cosmic rays*. PhD thesis, Karlsruher Institut für Technologie, 2011 (cited on page 58).
- [106] R. Gupta, *Television Engineering and Video Systems*. Tata McGraw-Hill, 2007 (cited on page 58).
- [107] N. Mandolesi et al., *Radio pulses from extensive air showers during thunderstorms—the atmospheric electric field as a possible cause*, *J. Atmos. Terr. Phys.* **36** no. 8 (1974) 1431–1435 (cited on page 65).
- [108] S. Buitink et al. (LOPES Collaboration), *Amplified radio emission from cosmic ray air showers in thunderstorms*, *Astron. Astrophys.* **467** no. 2 (2007) 385–394 (cited on pages 65 and 78).
- [109] J. Rautenberg for the Pierre Auger Collaboration, *Lightning detection at the Pierre Auger Observatory*, in *Proc. of the 34th ICRC*. The Hague, Netherlands, 2015 (cited on page 65).
- [110] T. Huege et al., *Dependence of geosynchrotron radio emission on the energy and depth of maximum of cosmic ray showers*, *Astropart. Phys.* **30** (2008) 96–104 (cited on page 66).
- [111] Pierre Auger Observatory, *CDAS Public Web page*. <http://www.auger.org.ar/CDAS-Public/>. Accessed: 1 September 2015 (cited on page 67).
- [112] P. Schellart et al. (LOFAR Collaboration), *Polarized radio emission from extensive air showers measured with LOFAR*, *J. Cosmol. Astropart. P.* **2014** no. 10 (2014) 014 (cited on page 75).
- [113] 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** no. 3 (2008) 433–441 (cited on page 79).
- [114] E. Polisensky, *Lfmap: A low frequency sky map generating program*, 111, Long Wavelength Array Memo Series, 2007 (cited on page 93).
- [115] F. James, *MINUIT, CERN Program Library Long Writeup D506*, 1994 (cited on page 104).
- [116] International Association of Geomagnetism and Aeronomy, *International Geomagnetic Reference Field 12*. <http://www.ngdc.noaa.gov/IAGA/vmod/igrf.html>. Accessed: 1 September 2015 (cited on page 105).
- [117] C. Bonifazi for the Pierre Auger Collaboration, *The angular resolution of the Pierre Auger Observatory*, *Nucl. Phys. B-Proc. Sup.* **190** (2009) 20–25 (cited on page 110).

SUMMARY

WHILE READING THIS SUMMARY, you are exposed to a torrent of microscopic particles. This never-ending stream of electrons, neutrinos, muons and many other particle types is the debris from collisions between cosmic particles and molecules in the earth's atmosphere. The cosmic particles are charged atomic nuclei, ranging from light hydrogen nuclei (protons) to heavy iron nuclei. They originate from faraway astronomical sources, and are what we – for historical reasons – call *cosmic rays*.

The cosmic rays that arrive on Earth have an enormous range in energy. There are eleven orders of magnitude in energy between cosmic rays with the lowest and the highest energies we can measure. As the energy of the particles goes up, we observe them less and less frequent. Every square centimeter on Earth is struck more than once per second by a cosmic ray from the low end of the energy spectrum, while cosmic rays of the highest energy – about 10^{20} eV – are so rare we have to wait longer than a century to be likely to measure one within one square kilometer.

The existence of these rare highly energetic cosmic particles is shrouded in mystery. It is not precisely known how these particles are accelerated to these huge energies. Because the charged nuclei are deflected by magnetic fields of unknown orientation and strength during their journey between the source and Earth, it is difficult to reconstruct their path and origin, as is illustrated in figure S1.1. Therefore, the sources and acceleration mechanisms of cosmic rays of the highest energies are as of yet not experimentally determined. The key to answer these fundamental questions lies in the measurement of the *mass composition* of cosmic rays.

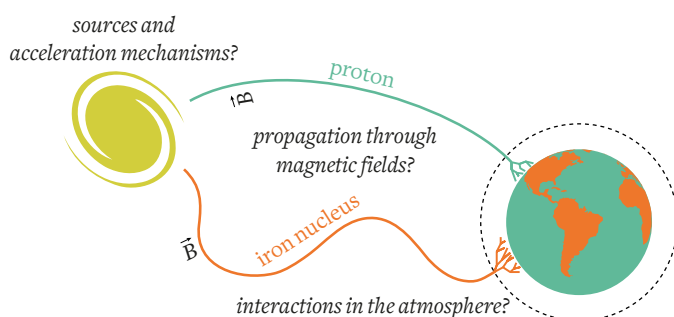


Figure S1.1: The journey of cosmic rays to Earth is shaped by several fundamental and unanswered questions. Magnetic fields (\vec{B}) have a stronger influence on **iron nuclei** than on **protons**.

The key: mass composition

Cosmic rays with low mass and charge – like protons – are deflected by a small amount by the (inter)galactic magnetic fields. Nuclei with a larger mass also have an increased charge, and are much more influenced by magnetic fields. It is therefore thought that if we only consider very highly energetic light cosmic rays, it is more likely that their arrival direction corresponds with the position of their sources, as is shown in figure S1.1. Consequently, distinguishing between light and heavy cosmic rays may lead to better understanding of their origins. Measuring the rare high-energy cosmic rays and determining their mass composition is however no simple undertaking.

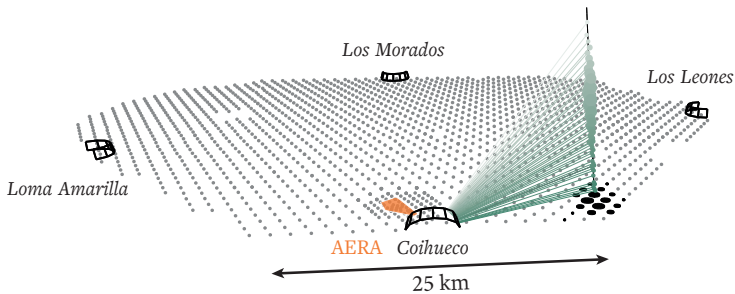


Figure S1.2: Overview of the Pierre Auger Observatory, with particle detectors indicated by gray dots. The air shower is observed by one of the four fluorescence telescopes, and is measured by the black particle detectors. The orange region marks the location of AERA (based on [76]).

Fortunately, nature lends us a hand. Whenever cosmic particles collide with atmospheric molecules they produce a cascade of secondary particles: the *air shower*. In the case of high-energy cosmic rays, this shower of reaction products can cover a surface area of tens of square kilometers at ground level. It is therefore possible to create a vast detector surface sensitive to high-energy cosmic rays with a relatively small number of sparsely placed particle detectors. This principle is used at the Pierre Auger Observatory in Argentina, where more than 1600 particle detectors are deployed in a region with an area of about 3000 km^2 – slightly larger than the country of Luxembourg. An overview of the observatory is drawn in figure S1.2. It is possible to calculate the energy and arrival direction of the incoming cosmic ray using the measured number of particles and their arrival times in the particle detectors.

To determine the mass composition of cosmic rays, fluorescence telescopes overlook the observatory from four locations. On dark and cloudless nights, these telescopes register the light emitted through fluorescence by atmospheric nitrogen after it is excited by charged particles in the air shower. From these measurements, the development of the air shower as a function of atmospheric depth can be determined. The development

of the air shower is related to the mass composition of the cosmic ray. This is because of a difference in interaction cross-section: an air shower initiated by a heavy nucleus will develop higher in the atmosphere than an air shower initiated by a light nucleus. One drawback of the fluorescence method is that the measurement conditions are only favorable about 13 % of the time. Therefore, it can only determine the mass composition of a small number of the rare high-energy cosmic rays.

Radio pulses from the air shower

In addition to direct detection of particles in the air shower and the fluorescence light, cosmic rays can be detected through their *radio emission*. This emission is mainly caused by the Lorentz force – which is exerted on the charged particles in the air shower by the geomagnetic field – that pushes electrons and positrons in opposite directions, as is visible in figure s1.3. Just like moving electrons in a wire this induces a current. This current results in radio emission, which is measurable by a simple antenna as a pulse in the megahertz domain – the frequency band of FM radio. In 2013, 95 relatively simple radio antennas were deployed as the second phase of the *Auger Engineering Radio Array* (AERA). They are located within a small region of the Pierre Auger Observatory, indicated in figure s1.2. With this setup, it is possible to measure radio pulses from the air shower in coincidence with observations from the particle detectors and the fluorescence telescopes.

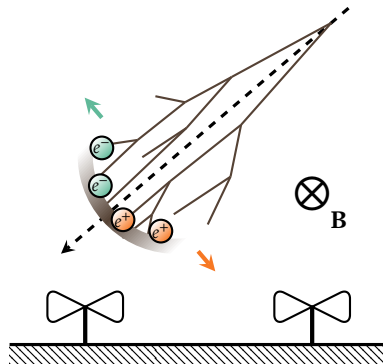


Figure s1.3: Influenced by the geomagnetic field (**B**), the electrons and positrons move in opposite directions in the air shower, which travels towards the earth's surface along the dashed arrow. This acceleration of charges results in radio emission.

It is possible to determine the arrival direction of the cosmic ray by considering the arrival time of the radio signal at different antennas. In addition, the signal strength of the radio pulse can be used to calculate the cosmic ray energy. The last piece of the puzzle is to measure cosmic ray mass composition using the radio signal, and this has been the

subject of my PhD research. The big advantage of a method that uses radio signals to measure the mass composition is that it is not dependent on dark nights – in contrast to the fluorescence technique. With radio signal detection, it is possible to develop a method that operates almost 100 % of the time. I have studied if the pulse length of the radio signal can be used to measure the development depth of the air shower – and thereby the mass composition of the cosmic ray.

Determining mass composition from the pulse length

The length and shape of the radio pulse as it is measured at ground level depends on the location of the radio emission region in the air shower with respect to the antenna position. This is illustrated by the *path length difference*, which is displayed in figure s1.4. The path length difference for an air shower that emits radiation close to the ground is larger than the difference for an air shower that emits radiation high in the atmosphere. The height of the emission region is related to the development depth of the air shower. Therefore, the length of the radio pulse is a parameter sensitive to mass composition. A cosmic ray with a light mass composition – such as a proton – will typically generate an air shower close to ground level, and therefore produce a long radio pulse. A heavier cosmic ray – such as an iron nucleus – will develop an air shower higher up in the atmosphere, and the measured radio pulse will be shorter.

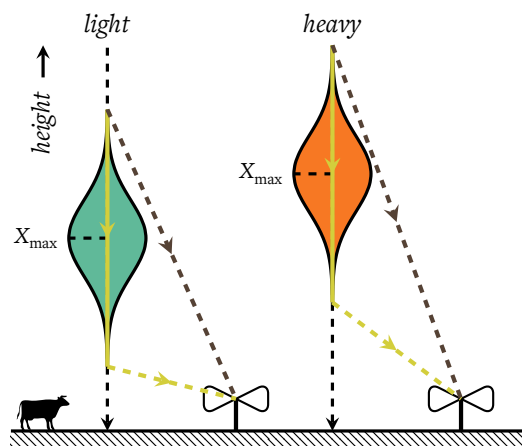


Figure s1.4: The length of the radio pulse at ground level as a path length difference: the difference between the time it takes for a radio signal to travel from the start of the emission region to the antenna (brown), and the time it takes for the signal to reach the antenna from the end of the emission region plus the total time in which the air shower emits radiation (yellow). The time difference for a **light** cosmic ray is typically larger than for a **heavy** cosmic ray.

As a measure for the pulse length I used the *spectral index* of the frequency spectrum of the pulse. The spectral index describes the decay of the pulse spectrum as a function of frequency, and can be used as a measure of the pulse length. A long pulse contains relatively more power in lower frequencies than a short pulse. This results in a steeper spectrum and a lower (more negative) spectral index. Conversely, a short pulse contains relatively more power in higher frequencies than a long pulse, which results in a flatter spectrum and a higher spectral index, as is illustrated in figure s1.5.

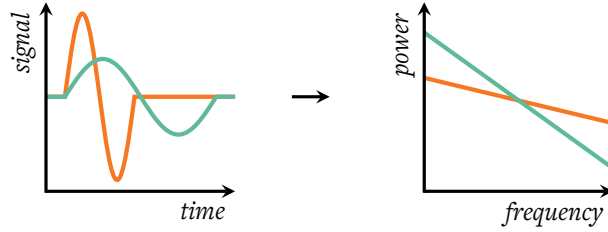


Figure s1.5: A short pulse results in a less steep spectrum than a long pulse, and therefore has a higher spectral index. The color coding corresponds to the air showers in figure s1.4.

The path length difference – and therefore also the spectral index – not only depends on the depth of air shower development. It also strongly depends on the arrival direction of the cosmic ray and the distance of the antenna to the air shower. Additionally, the travel time of the radio signal depends on the local speed of light in the atmosphere, which is not constant but decreases while moving closer to ground level. All in all, this results in a complex combination of different dependencies, which makes it difficult to relate the spectral index directly to the air shower development depth.

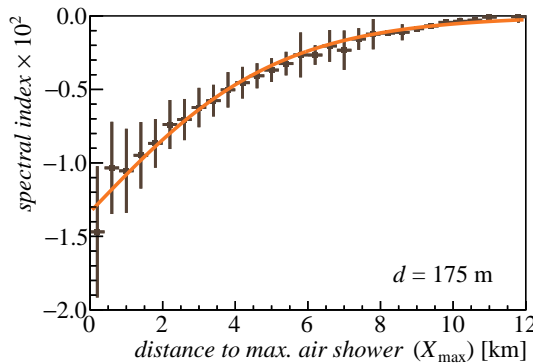


Figure s1.6: Parameterization of the spectral index from the simulation study for an antenna located at a distance of 175 m from the air shower

In order to study these dependencies, I performed a study of computer simulations of air showers and their radio emission. I examined many different arrival directions, antenna positions and development depths of the air shower, which I compared to the spectral index. Using this simulation study, it was possible to develop a parameterization that describes the spectral index as a function of development depth of the air shower, if the rest of the air shower geometry is completely known. The parameterization of the spectral index for one specific antenna distance is plotted in figure s1.6, in which the development depth of the air shower is expressed in terms of X_{max} : the atmospheric depth at which the maximum number of particles in the air shower is reached before it starts to die out, measured from the top of the atmosphere (see also figure s1.4).

Mass composition measurements with AERA

Following the simulation study, I have applied the parameterization of the spectral index to radio pulses measured by the AERA antenna stations. First, I have made a selection of radio pulses that are genuinely originating from air showers. The radio antennas are extremely sensitive to ambient noise, and some noise sources produce radio pulses that are difficult to distinguish from pulses from air showers. By accurately comparing the time structure and polarization of the pulses with information about the air shower from the Auger particle detectors I have made a pure selection of radio pulses from air showers. Some radio antenna stations are equipped with small particle detectors (scintillators), which made a useful contribution to this selection process.

Before the selected radio pulses could be used in the analysis, I have calibrated the frequency-dependent gain of the radio antennas to the background radiation from the Galaxy. Because of this calibration I have been able to measure the spectrum of the radio pulse with a much higher accuracy than in previous studies. I have also refined and applied a method to suppress narrowband interference – such as television channels – from our data by describing them with sine waves in the time domain.

A direct comparison of our method with the measurements of the Auger fluorescence telescopes shows that a low precision is achieved when only one measured radio pulse from an air shower is used to determine the depth of air shower development. The precision increases when a combined analysis is performed on multiple radio pulses from the same air shower measured by different radio antennas, as is plotted in figure s1.7.

By applying the parameterization of the spectral index to the complete selected AERA data set, I was able to determine the average depth of air shower development as a function of cosmic ray energy. The results of these measurements are visible in figure s1.8. The average depth of shower development that I measured by using the radio pulses is in agreement with measurements from the Auger fluorescence telescopes. Predictions of the average development depth from different models for two different mass com-

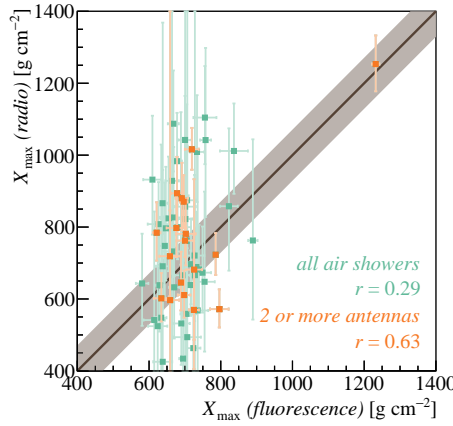


Figure s1.7: Direct comparison between measurements of the depth of air shower development – expressed in X_{\max} – from the fluorescence telescopes and the pulse shape of the radio signal. A 1:1 correlation would place all points on the diagonal line. The precision is low, but is increased when multiple radio measurements from the same air shower are used, as is indicated by the higher *Pearson correlation coefficient* (r).

positions of the cosmic rays are displayed in the plot as lines. From this, I conclude that cosmic rays have a predominantly light mass composition in the energy range in which AERA is sensitive.

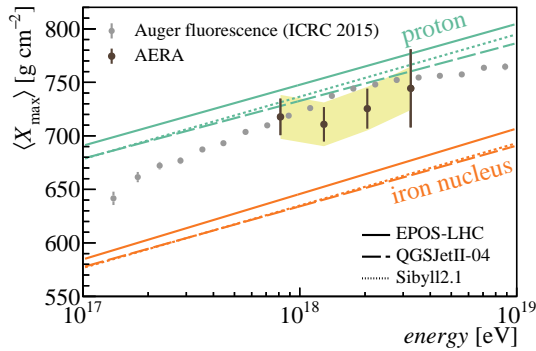


Figure s1.8: The average depth of air shower development – expressed in X_{\max} – as a function of cosmic ray energy. The results from our analysis with radio data from AERA are in agreement with the measurements from the Auger fluorescence telescopes. The diagonal lines are the predicted average depths of shower development based on several models with two different mass compositions.

Where do we go from here?

It follows from my PhD research that the spectral index of the radio pulse can be used to measure the mass composition of cosmic rays. This opens up new possibilities to use the radio signal to perform cosmic ray physics. A combined analysis of this technique with other methods to extract cosmic ray parameters from the radio signal will lead to an even greater applicability of the radio detection of cosmic rays. This is especially relevant for future large scale applications. Together with the development of new methods to measure the mass composition with the particle detectors of the Pierre Auger Observatory, this will enable us to better separate between light and heavy cosmic rays in the future. We will then be a little bit closer to answering the fundamental questions about the nature and origin of cosmic rays.

LIST OF PUBLICATIONS

As a member of the Pierre Auger Collaboration, Stefan Jansen has co-authored 33 publications in refereed journals as of 1 February 2016. Publications presenting results of AERA or that are otherwise related to the radio emission of extensive air showers are listed below.

S. Acounis et al.

Results of a self-triggered prototype system for radio-detection of extensive air showers at the Pierre Auger Observatory

J. Instrum. **7** no. 11 (2012) P11023.

A. Aab et al.

Probing the radio emission from air showers with polarization measurements

Phys. Rev. D **89** no. 5 (2014), 052002.

A. Aab et al.

Nanosecond-level time synchronization of autonomous radio detector stations using a reference beacon and commercial airplanes

J. Instrum. **11** no. 01 (2016) P01018.

A. Aab et al.

Energy Estimation of Cosmic Rays with the Engineering Radio Array of the Pierre Auger Observatory

Submitted to *Phys. Rev. D.*, [arXiv preprint: 1508.04267](#) (2015).

A. Aab et al.

First Measurement of the Energy in the Radio Signal of Extensive Air Showers

Submitted to *Phys. Rev. Lett.*

Internal Auger publications

S. Jansen & C. Timmermans

Timing and synchronization of AERA

GAP2012-067

S. Grebe, S. Jansen, C. Timmermans

Suppression of self-introduced narrowband RFI in the time domain

GAP2013-012

LIST OF PUBLICATIONS

C. Timmermans et al.

Description of the scintillator triggered AERA-II stations

GAP2013-074

J. Schulz et al.

The Scintillators of the Auger Engineering Radio Array

GAP2014-122

S. Jansen, S. Grebe, C. Timmermans

Towards standardized selection criteria for AERA data

GAP2015-007

S. Jansen, C. Timmermans, S. de Jong

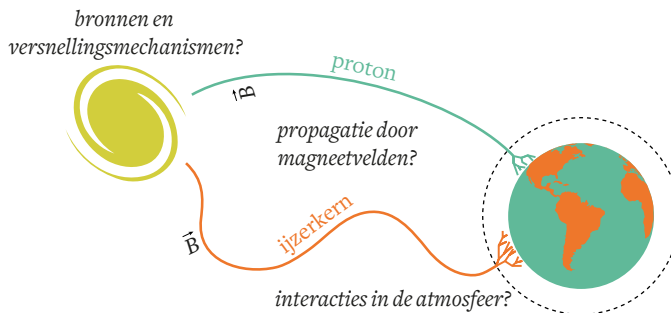
Single radio station sensitivity to shower development: a Monte Carlo study

GAP2015-054

SAMENVATTING

TIJDENS HET LEZEN van de samenvatting van mijn proefschrift word je blootgesteld aan een stortvloed van microscopische deeltjes. Dit is een constante stroom van muonen, neutrino's, elektronen en vele andere soorten elementaire deeltjes, die de restanten vormen van botsingen tussen kosmische deeltjes en moleculen in de atmosfeer van de aarde. De kosmische deeltjes zijn geladen atoomkernen, variërend van lichte waterstofkernen (protonen) tot zware ijzerkernen. Ze zijn afkomstig van ver weg gelegen astronomische bronnen, en zijn wat we – om historische redenen – *kosmische stralen* noemen.

De energie die de kosmische stralen hebben als ze de aarde bereiken varieert enorm. Er zitten maar liefst elf ordes van grootte – dat is een factor van honderd miljard – tussen de kosmische stralen met de laagste en de hoogste energie die we kunnen meten. Naarmate de energie van de deeltjes omhoog gaat, nemen we ze met steeds minder grote regelmaat op aarde waar. Elke vierkante centimeter op aarde wordt vaker dan eens per seconde geraakt door een kosmische straal met een lage energie, terwijl de kosmische stralen met de hoogst meetbare energie – ongeveer 10^{20} eV – zo zeldzaam zijn dat je doorgaans meer dan een eeuw moet wachten om er één per vierkante kilometer te kunnen meten.



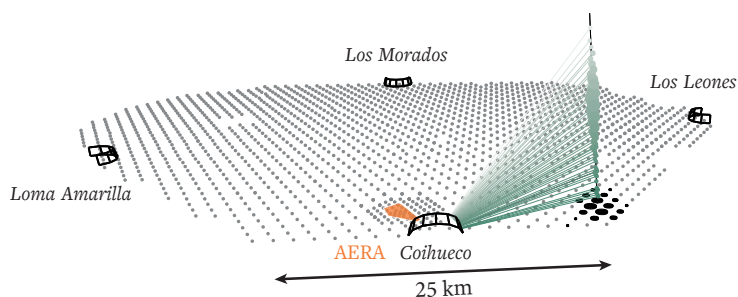
Figuur s2.1: De reis van kosmische stralen naar de aarde voert langs enkele onbeantwoorde fundamentele vraagstukken. Magneetvelden (\vec{B}) hebben een grotere invloed op **ijzerkernen** dan op **protonen**.

Het bestaan van deze zeldzame hoogenenergetische deeltjes is omgeven door raadselen. We weten ten eerste niet precies hoe deze deeltjes versneld kunnen worden tot een dermate hoge energie. Daarnaast worden de geladen deeltjes afgebogen door magneetvelden met onbekende sterkte en oriëntatie in hun reis tussen de bron en de aarde, waardoor het erg lastig is de herkomst van de deeltjes te achterhalen. Daardoor zijn de

bronnen en versnellingsmechanismen van de kosmische stralen met de hoogste energie vooralsnog niet experimenteel vastgesteld. De sleutel om deze fundamentele vraagstukken op te lossen ligt in het meten van de *massacompositie* van de kosmische stralen.

De sleutel: massacompositie

Lichte kosmische deeltjes met weinig lading – zoals protonen – worden in kleine mate afgebogen door magnetevelden tussen en binnen sterrenstelsels. Hoe zwaarder de atoomkern, hoe groter de lading, en des te sterker is de invloed van de magnetevelden op deze deeltjes. Door in zoektochten naar bronnen van kosmische straling alleen de aankomstrichting van zeer hoog energetische lichte deeltjes mee te nemen is het waarschijnlijker dat de aankomstrichting overeenkomt met de ligging van de bron, zoals te zien is in figuur s2.1. Het meten van de zeldzame hoogenergetische kosmische straling en het bepalen van hun massacompositie is echter geen eenvoudige opgave.



Figuur s2.2: Overzichtstekening van het Pierre Auger Observatorium, met de deeltjesdetectoren als grijze stippen. De deeltjeslawine wordt gezien door één van de vier fluorescentietelescopieën, en wordt gemeten door de zwarte deeltjesdetectoren. Het oranje gebied markeert AERA (gebaseerd op [76]).

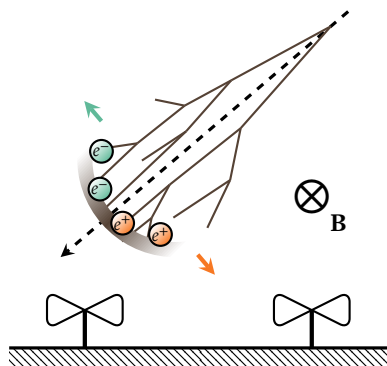
Gelukkig helpt de natuur ons hier een handje. Wanneer de kosmische deeltjes botsen met moleculen in de atmosfeer produceren ze een cascade van secundaire deeltjes die we de *deeltjeslawine* noemen. In het geval van hoogenergetische kosmische straling kan deze lawine van reactieproducten op grondniveau een oppervlakte bestrijken van enkele tientallen vierkante kilometers. Hierdoor is het mogelijk om met relatief weinig deeltjesdetectoren een gigantisch detectoroppervlak gevoelig voor hoogenergetische kosmische straling te creëren. Van dit principe wordt gebruikt gemaakt bij het Pierre Auger Observatorium in Argentinië, waar meer dan 1600 deeltjesdetectoren zijn geplaatst in een gebied met een oppervlakte van ongeveer 3000 km^2 – ongeveer zo groot als de provincie Zuid-Holland. Een overzichtstekening van het observatorium is te zien in figuur s2.2. Met behulp van het gemeten aantal deeltjes en hun aankomsttijd in de deeltjesdetectoren is het mogelijk om de energie en aankomstrichting van de oorspronkelijke kosmische straal te berekenen.

Om de massacompositie van de kosmische straal te kunnen bepalen staan er op vier locaties rondom het observatorium fluorescentietelescopieën opgesteld die gericht zijn op de atmosfeer boven de deeltjesdetectoren. Deze kunnen op heldere donkere nachten het fluorescentielicht meten dat door atmosferisch stikstof wordt uitgezonden nadat het aangeslagen is door de geladen deeltjes in de lawine. Hierdoor kan de ontwikkeling van de deeltjeslawine als functie van atmosferische diepte bepaald worden. Deze ontwikkeling is gerelateerd aan de massacompositie van de kosmische straal: vanwege de grotere botsingsdoorsnede zal de deeltjeslawine geproduceerd door bijvoorbeeld een ijzerkern gemiddeld hoger in de atmosfeer tot ontwikkeling komen dan die van een proton. Het nadeel van de fluorescentietechniek is dat er maar ongeveer 13 % van de tijd geschikte meetomstandigheden zijn, waardoor slechts van een klein aantal van de toch al zeldzame hoogenergetische kosmische stralen de massacompositie bepaald kan worden.

Radiopulsen uit de deeltjeslawine

Behalve via de deeltjes in de lawine of het fluorescentielicht zijn kosmische stralen ook te detecteren met behulp van *radiostraling*. Deze straling wordt hoofdzakelijk veroorzaakt doordat de Lorentzkracht – die door het aardmagnetisch veld op de geladen deeltjes in de deeltjeslawine wordt uitgeoefend – de elektronen en positronen een tegengestelde richting op drukt, zoals is afgebeeld in figuur [s2.3](#). Net als bewegende elektronen in een draad veroorzaakt dit een stroom, die op zijn beurt weer resulteert in radiostraling. Deze straling laat zich met een simpele radioantenne meten als een puls in het megahertz domein – vergelijkbaar met FM-radio. Sinds 2013 vormen 95 relatief eenvoudige radioantennes de tweede fase van de *Auger Engineering Radio Array* (AERA), geplaatst op een klein oppervlak binnen het Pierre Auger Observatorium, zoals is aangegeven in figuur [s2.2](#). Met deze opstelling is het mogelijk de radiopulsen uit de deeltjeslawine te meten in combinatie met waarnemingen van de deeltjesdetectoren en de fluorescentietelescopieën.

Met behulp van de aankomsttijd van het radiosignaal bij verschillende antennes is het mogelijk om de aankomstrichting van de kosmische stralen te bepalen. Daarnaast weten we sinds kort dat we met behulp van de signaalsterkte van de puls ook de energie van de kosmische straal kunnen uitrekenen. Het laatste puzzelstukje is het meten van de massacompositie met radiostraling, en dit vormde het onderwerp van mijn promotieonderzoek. Het grote voordeel van compositiebepaling met radiostraling is dat deze niet afhankelijk is van de duisternis – in tegenstelling tot de fluorescentietechniek. We kunnen zo in principe een methode ontwikkelen die bijna 100 % van de tijd werkt. Ik heb onderzocht of de pulslengte van het radiosignaal bruikbaar is als maat voor de ontwikkelingsdiepte van de deeltjeslawine in de atmosfeer – en dus om de massacompositie van de kosmische straal te bepalen.



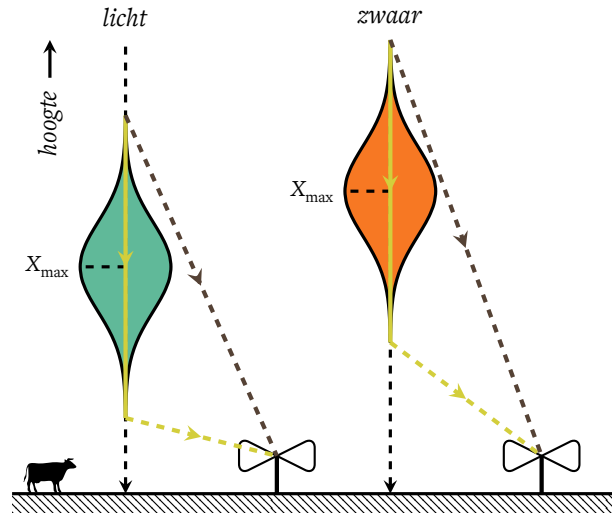
Figuur s2.3: Onder invloed van het aardmagneetveld (**B**) bewegen de elektronen en positronen in tegengestelde richting in de deeltjeslawine, die langs de gestreepte pijl naar het aardoppervlak raast. Dit veroorzaakt emissie van radiostraling.

Massacompositiebepaling met de pulslengte

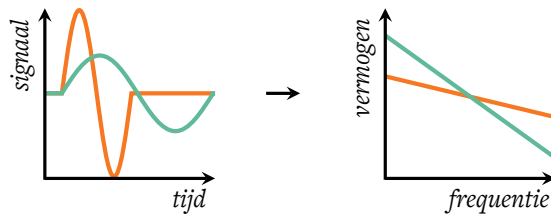
De lengte en vorm van de radiopuls zoals we deze meten op de grond hangt af van de locatie van het radio-emissiegebied in de deeltjeslawine ten opzichte van de antennepositie. Dit kunnen we illustreren met het *padlengteverschil*, zoals is weergegeven in figuur s2.4. Het padlengteverschil bij een deeltjeslawine die laag bij de grond radiostraling uitzendt is groter dan het padlengteverschil bij een deeltjeslawine die hoog in de atmosfeer radiostraling uitzendt. Omdat de hoogte van het emissiegebied gerelateerd is aan de ontwikkelingsdiepte van de lawine, is de lengte van de radiopuls hierdoor een compositiegevoelige parameter. Kort samengevat zal een kosmische straal met een lichte massacompositie – zoals een proton – laag in de atmosfeer een deeltjeslawine genereren dus een lange radiopuls opleveren. Een zwaardere kosmische straal – zoals een ijzernucleus – zal daarentegen eerder in de atmosfeer een lawine ontwikkelen waardoor een korte radiopuls gemeten zal worden.

In de praktijk is het padlengteverschil niet alleen afhankelijk van de ontwikkelingsdiepte van de deeltjeslawine, maar hangt het ook af van de aankomstrichting van de kosmische straal en de afstand van de antenne tot de deeltjeslawine. Daar komt nog bij dat de reistijd van het radiosignaal afhankelijk is van de lokale lichtsnelheid in de atmosfeer, die niet constant is maar afneemt wanneer men zich dichterbij de grond bevindt. Alles bij elkaar vormt dit een complex samenspel van verschillende afhankelijkheden die het ingewikkeld maakt de pulslengte direct te relateren aan de ontwikkelingsdiepte van de deeltjeslawine.

Om deze relaties te onderzoeken heb ik een studie uitgevoerd op computersimulaties van deeltjeslawines en hun radiostraling. Hierbij heb ik veel verschillende aan-



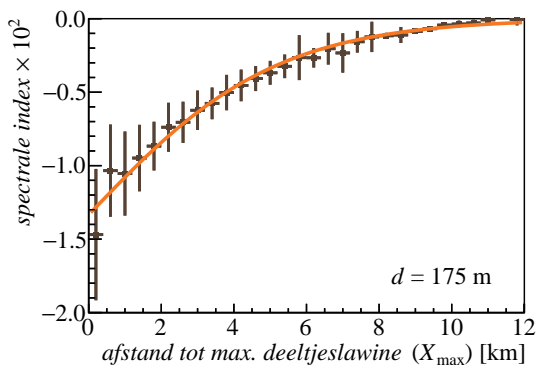
Figuur s2.4: De lengte van de radiopuls op grondniveau als padlengteverschil: het verschil tussen de tijd die een radiosignaal reist van het begin van het emissiegebied naar de antenne (bruin), en de tijd die een radiosignaal nodig heeft om van het eind van het emissiegebied naar de antenne te reizen plus de totale tijd waarin de deeltjeslawine radiostraling uitzendt (geel). Het lengte verschil voor een **lichte** kosmische straal is typisch groter dan voor een **zware**.



Figuur s2.5: Een **korte** puls produceert een minder steil spectrum dan een **lange** puls, en heeft daardoor een hogere spectrale index. De kleurcode correspondeert met de deeltjeslawines van figuur s2.4.

komstrichtingen, antenneposities en ontwikkelingsdieptes van de deeltjeslawine bekeken. Deze heb ik vervolgens vergeleken met de pulslengte, of om precies te zijn de *spectrale index* van het frequentiespectrum behorende bij de radiopuls. De spectrale index beschrijft het verval van het pulsspectrum als functie van frequentie, en is te gebruiken als maat voor de pulslengte. In een lange puls zit naar verhouding meer vermogen in de lage frequenties dan in een korte puls. Dit resulteert in een steiler spectrum en een

lagere (sterker negatieve) spectrale index. Andersom zit in een korte puls naar verhouding meer vermogen in de hoge frequenties, wat resulteert in een vlakker spectrum en een hogere spectrale index, zoals is weergegeven in figuur s2.5. Met behulp van de simulatiestudie heb ik een parametrisering ontwikkeld die de spectrale index beschrijft als functie van de ontwikkelingsdiepte van de deeltjeslawine, mits de rest van de geometrie van de deeltjeslawine van de kosmische straal volledig bekend is. In figuur s2.6 is de parametrisering van de spectrale index weergegeven voor één bepaalde antennepositie. De ontwikkelingsdiepte drukken we uit in X_{\max} : de atmosferische diepte waar het maximale aantal deeltjes in de deeltjeslawine wordt bereikt voordat hij uitsterft, gemeten vanaf de bovenkant van de atmosfeer (zie ook figuur s2.4).



Figuur s2.6: Parametrisering van de spectrale index uit de simulatiestudie voor het geval de antenne 175 m van de deeltjeslawine af staat.

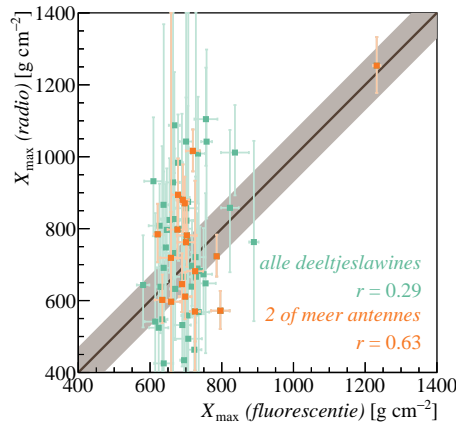
Massacompositie zoals gemeten door AERA

De parametrisering van de spectrale index heb ik vervolgens toegepast op radiopulsen gemeten met de AERA antennestations. Eerst heb ik echter een selectie gemaakt van gemeten radiopulsen die daadwerkelijk afkomstig zijn van deeltjeslawines. De radioantennes zijn erg gevoelig voor omgevingsruis, en sommige ruisbronnen produceren radiopulsen die moeilijk te onderscheiden zijn van pulsen uit de deeltjeslawine. Door nauwkeurig de tijdstructuur en polarisatie van de pulsen te vergelijken met informatie uit de Auger deeltjesdetectoren heb ik een zo zuiver mogelijke selectie van pulsen gemaakt. Hierbij heb ik ook handig gebruik gemaakt van kleine deeltjesdetectoren (scintillatoren) waarmee een aantal van de radioantennestations uitgerust zijn.

Voordat de geselecteerde radiopulsen gebruikt konden worden in de analyse heb ik de frequentiegevoeligheid van de radioantennes gekalibreerd op de achtergrondruis van de Melkweg. Hierdoor kon ik het spectrum en de spectrale index van de puls nauwkeuriger bepalen dan bij eerdere studies. Ook heb ik een methode om smalbandige omgevingsruis – zoals televisiezenders – te onderdrukken door ze te beschrijven met

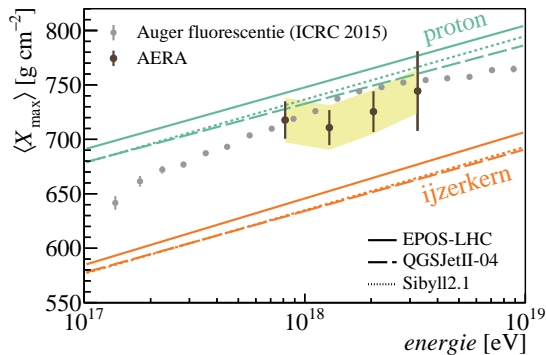
sinusgolven in het tijdsdomein verder uitgewerkt en toegepast op de gemeten radiosignalen.

Uit een directe vergelijking met metingen van de Auger fluorescentietelescopen blijkt dat de nauwkeurigheid van de radiomethode niet erg groot is wanneer er maar één radiopuls uit de deeltjeslawine gebruikt wordt om de ontwikkelingsdiepte te bepalen. De nauwkeurigheid neemt echter toe als een gecombineerde analyse uitgevoerd wordt op meerdere radiopulsen afkomstig van dezelfde deeltjeslawine, gemeten door verschillende antennes. Deze vergelijking is weergegeven in figuur s2.7.



Figuur s2.7: Directe vergelijking tussen metingen van de ontwikkelingsdiepte van deeltjeslawine – uitgedrukt in X_{\max} – van de fluorescentietelescopen en met de vorm van de radiopuls. Een 1:1 correlatie zou alle punten op de diagonale lijn plaatsen. De nauwkeurigheid is laag, maar neemt toe als meerdere radiometingen uit dezelfde deeltjeslawine worden gebruikt, zoals is aangegeven door de hogere *Pearson correlatiecoëfficiënt* (r).

Door de parametrisering van de spectrale index toe te passen op de complete geselecteerde AERA dataset heb ik de gemiddelde ontwikkelingsdiepte van de deeltjeslawine kunnen meten als functie van energie van de kosmische straal. De resultaten van deze meting zijn weergegeven in figuur s2.8. Mijn metingen van de gemiddelde ontwikkelingsdiepte met behulp van het radiosignaal zijn in overeenstemming met metingen van de Auger fluorescentietelescopen. De gemiddelde ontwikkelingsdiepte die verschillende modellen voorspellen voor twee verschillende massacomposities van de kosmische straal zijn in de plot aangegeven met lijnen. Aan de hand hiervan is te zien dat we voornamelijk kosmische stralen met een lichte massacompositie meten in het energiegebied waarin AERA gevoelig is.



Figuur s2.8: De gemiddelde ontwikkelingsdiepte van de deeltjeslawine – uitgedrukt in X_{max} – als functie van de energie van de kosmische straal. De resultaten gebaseerd op de analyse met radiodata van AERA zijn in overeenstemming met de fluorescentiemetingen. De diagonale lijnen zijn de verwachte gemiddelde ontwikkelingsdieptes gebaseerd op enkele modellen met verschillende massacomposities.

Hoe nu verder?

Uit mijn promotieonderzoek blijkt dat de spectrale index van de radiopuls bruikbaar is voor massacompositiemetingen van kosmische stralen. Dit betekent dat we behalve de signaaltijden -sterkte ook het frequentiespectrum van de radiopuls kunnen aanwenden om de eigenschappen van kosmische straling te bepalen. Een gecombineerde analyse van al deze informatie zorgt voor een grotere bruikbaarheid van de radiodetectietechniek, wat interessant is met het oog op toekomstige toepassingen op grotere schaal. Samen met de ontwikkeling van methodes om ook met de deeltjesdetectoren van het Pierre Auger Observatorium massacompositiemetingen te verrichten biedt dit de mogelijkheid om in de toekomst veel beter onderscheid te maken tussen lichte en zware kosmische stralen. Hierdoor komt het beantwoorden van de fundamentele vraagstukken omtrent de oorsprong van hoogenergetische kosmische straling weer een stukje dichterbij.

CURRICULUM VITAE

Stefan Jansen was born on 18 August 1985 in Ludvika, Sweden. After being moved to the Netherlands, he first attended primary school at Basisschool 't Startblok in Elst, and then high school at the Over Betuwe College in Elst and Bommel. He started his studies in physics and astronomy at the then Katholieke Universiteit Nijmegen in 2003.

He received his bachelor's degree in 2009, and in August 2011 he obtained his master's degree in physics and astronomy in Nijmegen. For his research project he spent one year at the Florida Institute of Technology in Melbourne, Florida. Here, he modeled superoutbursts and accretion disks of SU UMa binary stars under the supervision of Prof. Matt Wood. During his stay in the United States he got the opportunity to visit the Pierre Auger Observatory for a shift controlling the fluorescence telescopes for two weeks. This experience aroused his curiosity in the field of astroparticle physics and cosmic rays.

In November 2011, he therefore started his PhD via Nikhef Amsterdam at the Institute of Mathematics, Astrophysics and Particle physics (IMAPP) under the supervision of Dr. Charles Timmermans and Prof. Sijbrand de Jong. There, he worked on research involving composition measurements with the AERA radio detector at the Pierre Auger Observatory, the results of which were presented in this thesis.

ACKNOWLEDGEMENTS

WRITING THIS THESIS would not have been possible without the support of my colleagues, friends and family. Most of all this is true for the delightful coworkers in the pleasant working environment that is our high energy physics department and the Auger group in Nijmegen. An environment which encourages creative thinking¹, but also allowed me to relax and have a laugh in between endless typing, clicking and compiling.

First I would like to thank my supervisors, Charles and Sijbrand, for giving me the chance to develop myself in this environment. Without their guidance, provocative coaching and scientific discussions I would never have been able to produce the results presented in my thesis. Back when I was still a young and prospective PhD student, Harm, Stefan and Anna taught me all the ins and outs of our experiment and the radio analysis, but also all the perks of working in an international group like AERA. During my years as a PhD student, I collaborated most closely with Guus and Johannes. We have had many fruitful conversations about our research, but of course also lots of nice travels to remote places such as Malargüe and Beekbergen – where the lost wooden leg of Seyss-Inquart reputedly still haunts the castle. Johannes also provided me with the bulk of the shower simulations used in my analysis, something for which I also owe Anna, Christian Glaser and Washington Carvalho Jr. gratitude. Later, our group was joined by Giuseppe, Alex and Fabrizia. It is good to know that our work of maintaining the inhabitants of our office jungle will be continued by such a great group of people. Outside of the department, exchanges with Clancy James and Raphael Krause were very helpful in solving some of the major problems I faced when working on my analysis.

Coffee gatherings, (cheese) lunches and borrels at the department provided ample opportunity for small talk and intriguing arguments about such diverse topics as plastic islands, Sicilian ice cream sandwiches, the merits of gearless bikes, biking trips, climbing times on bikes and bikes in general. I want to thank Antonia, Cristina, Frank, Geert-Jan, Irene, Jan-Willem, Jeroen, Jeroen, Jins, Katharina, Luca, Magda, Marcel, Marie, Melvin, Nicolo, Rene, Robert, Ronald, Sascha, Thei, Thijs, Vincent and Wim for their input and company. Most importantly, the great working atmosphere at the department would not have been possible without our secretariat keeping everything in running condition: Gemma, Marjo and Annelies.

Buiten onze afdeling en ver weg van mijn werkcomputer was er nog genoeg tijd om te wielrennen, muziek te maken en bier te drinken. Met vrienden, oud-studiegenoten en

¹A.F. Jarosz et al. *Conscious. Cogn.* 21 no. 01 (2012) 487–493.

mede-promovendi die ik tegen het lijf bleef lopen in de brede gangen van onze faculteit. Ik wil Bernard, Corneel, Frank, Jonas, Jord, Lennert, Karel, Leendertjan, Martijn, Merel, Sander, Sjoerd, Thomas, Tom en Youri bedanken, want dankzij hen waren mijn vier jaren als promovendus meer dan draaglijk. Ook de vrijwel dagelijkse chatsessies met Mario zorgden voor het nodige vermaak, en ik wil hem daarnaast bedanken voor alle grafische tips die ik van hem kreeg voor allerlei projecten en vooral ook voor het werk dat hij in mijn kافت gestopt heeft.

Ten slotte wil ik mijn familie en schoonfamilie bedanken voor hun gezelligheid en geduld. In het bijzonder Papa, Mama, Ingrid en Oma, voor alle ondersteuning, maar natuurlijk ook voor de verre en minder verre reizen die we door de jaren heen samen gemaakt hebben. Zonder titel geen proefschrift, en omdat een kratje bier in dit geval toch een beetje een sigaar uit eigen doos zou zijn wil ik deze gelegenheid ook graag aangrijpen om Carlijn te bedanken voor het bedenken hiervan. Niet alleen voor de titel trouwens, maar ook voor alle ondersteuning, liefde en het doorstaan van jaren gezeur over mijn proefschrift en de vraag of het ooit nog wat ging worden.

RADIO FOR THE MASSES

*Cosmic ray mass composition measurements
in the radio frequency domain*



The key to unravel the mysterious origins of high energy cosmic rays is to measure their mass composition. Stefan Jansen presents in this thesis a new method to determine this mass composition. For this, he uses radio detection of the air shower, which is initiated by the cosmic ray in the atmosphere.



The shape of the measured radio pulse can be analyzed in the frequency domain, and is related to the mass composition. He applied the new method to data from the AERA experiment at the Pierre Auger Observatory in Argentina.

Using the shape of the radio pulse he determined that cosmic rays consist primarily of light particles in the energy range of the experiment. This is in agreement with measurements from other available instruments at the observatory. The method described in this thesis opens up new possibilities for the application of radio detection of cosmic rays.

ISBN 978-94-028-0073-9



9 789402 800739 >