THE LATERAL DISTRIBUTION OF RADIO SIGNALS FROM COSMIC RAY AIR SHOWERS

Zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

von der Fakultät für Physik des
Karlsruher Institut für Technologie (KIT)

genehmigte

Dissertation

von

BENJAMIN LOTHAR FUCHS

aus Esslingen am Neckar

Tag der mündlichen Prüfung: 26. Oktober 2012
Referent: Prof. Dr. Johannes Blümer, Karlsruher Institut für Technologie
Korreferent: Prof. Dr. Günter Quast, Karlsruher Institut für Technologie
Betreuer: Dr. Tim Huege, Karlsruher Institut für Technologie

13. Februar 2013

In Memoriam

Josef Landsbeck

CONTENTS

ACRONYMS

4nec2 antenna characteristic simulation

ADC analog-to-digital converter

AERA Auger Engineering Radio Array

AGN active galactic nuclei

AMIGA Auger Myons and Infill to the Ground Array

BLS Balloon Launching Station

CDAS central data acquisition system

CERN European Organization for Nuclear Research

CMB cosmic microwave background

CODALEMA COsmic ray Detection Array with Logarithmic Electro Magnetic Antennas

COREAS REAS in CORSIKA

CORSIKA cosmic ray simulation for KASCADE

CRS central radio station

CS coordinate system

EVA Electric fields, using a Variable index of refraction in Air shower simulations

FD fluorescence detector

FFT Fast Fourrier Transformation

GZK Greisen Zatsepin Kuz'min effect

HEAT High Elevation Auger Telescope

KASCADE KArlsruhe Shower Core and Array DEtector

LDF lateral distribution function

LHC Large-Hadron-Collider (CERN)

LIDAR LIght Detection And Ranging

LOFAR LOw Frequency ARray

LOPES LOFAR prototype station

LPDA logarithmic periodic dipole antenna

MAD median absolute deviation

MGMR Macroscopic model of Geo-Magnetic Radio emission

MSU Moscow State University

RD radio detectors

REAS3 Radio simulation of Extended Air Showers

RMS root mean square

ROOT CERN analysis package

SAD space angle difference

SD surface detectors

SNR signal-to-noise ratio

UHECR ultra-high energy cosmic rays

VEM vertical equivalent muon

xml extendable markup language

ABSTRACT

This work deals with the reconstruction of the radio signals emitted by air showers and of the characteristics of the initiating cosmic rays. Air showers are created by cosmic rays hitting the Earth atmosphere. The air shower emits radio signals, which are detectable with radio antennas at ground level.

AERA, the Auger Engineering Radio Array measures these radio signals at the Pierre Auger Observatory in Argentina. One of many goals of AERA is the reconstruction of the characteristics of cosmic rays from radio measurements. This requires the development of new analysis and reconstruction methods, like the ones discussed in this work.

First a reconstruction method for the complete $\vec{E}$-field vector from partial measurements is developed. AERA is therefore currently the only radio air shower experiment with access to a reconstruction of the three-dimensional $\vec{E}$-field vector. Subsequently a detailed discussion on the quality and treatment of the antenna characteristic is given, as it is an important part of the $\vec{E}$-field reconstruction.

The now available $\vec{E}$-field provides then the foundation for a comparison of measured $\vec{E}$-fields with state of the art radio simulations. In average, a factor of two difference between COREAS simulations and AERA measurements is observed. EVA simulations exhibit a factor of three difference.

Based on the $\vec{E}$-field, also a reconstruction of the cosmic ray energy using the LDF is presented. In a first step a new parametrisation of the lateral distribution of radio signals is discussed. This includes a comparison of a new parametrisation based on the Gaussian distribution function with the common exponential parametrisation. Subsequently, the parameters of the Gaussian parametrisation are used to derive an energy estimator for radio measurements in general and in special for AERA. On basis of radio simulations a theoretical energy resolution of 13 % below 10^{18} eV can be reached. The energy resolution estimated for AERA measurements is larger with 16 %. In addition, a feature in the energy estimator at around 10^{18} eV is observed. Simulations indicate a relation between the feature and the cosmic ray primary particle type. Two physical explanations are discussed, one connected to the altitude of the observing experiment the other to the coherence of the emitted radio signal.

Finally, a radio core reconstruction method using the new Gaussian parametrisation is presented for radio simulations. A relative core shift of about 20 m to the east between the radio and the particle core is observed. This result is compatible with published results.

ZUSAMMENFASSUNG

Diese Arbeit beschäftigt sich mit der Rekonstruktion der Radiosignale von Luftschauern und darauf aufbauend mit den Eigenschaften der auslösenden kosmischen Strahlung. Luftschauer entstehen wenn kosmische Strahlung auf die Atmosphäre der Erde trifft und emittieren ein Radiosignal.

Das Auger Engineering Radio Array (AERA) misst diese Radiosignale am Pierre Auger Observatorium in Argentinien. Eines der Ziele von AERA ist die Untersuchung der Radiosignale und die Rekonstruktion der Eigenschaften der auslösenden kosmischen Strahlung. Im Zuge dessen ist die Entwicklung neuer Analyse- und Rekonstruktionsmethoden unerlässlich, wozu diese Arbeit ihren Beitrag leistet.

Zunächst wird eine neue Rekonstruktionsmethode vorgestellt, welche es AERA als erstem Radioluftschauer-Experiment ermöglicht, den vollständigen elektrischen $\vec{E}$-Feldvektor des Radiosignals zu rekonstruieren. Dabei wird auch die technische Behandlung und Eigenschaften der Antennencharakteristik diskutiert.

Anschließend wird ein Vergleich des gemessenen $\vec{E}$-Feldes mit modernsten Radiosimulationen gezogen. Es wird dabei ein durchschnittlicher Unterschied von einem Faktor zwei zwischen COREAS Simulationen und gemessenen Radiosignalen festgestellt. EVA Simulationen weisen im Durchschnitt einen Unterschied um Faktor drei auf.

Auf Basis des $\vec{E}$-Feldes und der Lateralverteilung der Radiosignale wird anschließend eine Methode zur Energiebestimmung der kosmischen Strahlung für Radiomessungen vorgestellt. Dazu wird die Lateralverteilung der Radiosignale mittels einer Gauss-Funktion parametrisiert und der bisher üblichen exponentiellen Parametrisierung gegenübergestellt. Anschließend wird ein Energieschätzer auf Basis der Gauss-Parametrisierung abgeleitet, welcher die Bestimmung der Energie der kosmischen Strahlung allgemein und insbesondere für AERA Messungen ermöglicht. Die theoretische Energieauflösung mittels Simulationen wird zu 13 % unterhalb von 10^{18} eV bestimmt, die Energieauflösung für AERA Messungen auf 16 %. Außerdem wird eine Änderung im Energieschätzer nahe 10^{18} eV näher untersucht, welche von der Art des Primärteilchens abhängt. Es werden zwei physikalische Ursachen diskutiert, welche auf der Höhenlage des Experiments und der Kohärenz des Radiosignals basieren.

Zum Schluss wird auf Basis der neuen Parametrisierung eine Rekonstruktionsmethode für das Zentrum der Radioemission am Boden diskutiert. Die Ergebnisse deuten auf eine Verschiebung des Radioschauerzentrum gegenüber dem Teilchenschauerzentrum um 20 m nach osten hin. Dies Ergebnis deckt sich mit anderen publizierten Ergebnissen.

1 INTRODUCTION

At the time of writing, it is exactly 100 years since Victor Franz Hess made his Nobel prize winning discovery of cosmic rays. His discovery raised many questions about the sources of cosmic rays or their composition and even today, physicists all over the world are occupied with answering them. In comparison, the first attempts to detect a radio signal from cosmic rays are only half as old.

Due to the limitations of the available radio technology back in the 1970s, the radio detection of cosmic rays air showers was nearly forgotten for about 30 years. Only ten years ago has technological progress renewed these efforts and new radio experiments like LOFAR prototype station (LOPES) or COsmic ray Detection Array with Logarithmic Electro Magnetic Antennas (CODALEMA) have accepted the challenge to investigate the radio signals once again. Therefore, the investigation of radio emission from cosmic ray air showers is today, even after a long break, a very active community of researchers throughout the world. In this global effort, Auger Engineering Radio Array (AERA), as part of the Pierre Auger Observatory contributes as one of the largest and most advanced radio detectors currently deployed.

The radio detection technique was and is still especially interesting as a promising alternative to fluorescence detectors, since radio measurements are not limited to the night. Therefore radio detectors easily provide large statistics. Furthermore, the radio emission is in principle sensitivity to the energy and composition of the cosmic rays. All together radio detection combines the benefits of the established particle and fluorescence / Čerenkov detection techniques.

The Pierre Auger Observatory consists of two large scale detectors, i.e. surface detectors (SD) and fluorescence detector (FD), that both measure cosmic rays in coincidence. The SD consists of water Čerenkov particle detectors and the FD comprises of 27 fluorescence telescopes overlooking the SD array. The AERA extension is dedicated from technical point of view to the development of new large scale radio detection techniques. From the physics point of view, on the investigation of radio emission processes, the emitted radio signals and finally will contribute to the understanding of air shower physics. For these physics challenges, it is mandatory to improve and develop existing and new analysis techniques.

In three main parts, this work contributes to the technical aspects by advancing and improving existing analysis methods; to our understanding of the radio signals from cosmic ray air showers by analysis of measured and simulated radio signals and finally to the reconstruction of cosmic ray air showers on basis of radio measurements by presenting an improved energy estimator on basis of the lateral distribution of cosmic rays.

The first part of this work is a contribution to the reconstruction of the radio signal itself. As an electro-magnetic wave, the radio pulse from cosmic ray air showers is completely described by the three-dimensional $\vec{E}$-field timeseries. With the methods introduced here, it is now possible to reconstruct this important quantity even from measurements that cover only two components of the radio signal. This is also the first time that an experiment has this method at its disposal, as former experiments relied on partial reconstructions of the $\vec{E}$-field. This fundamental quantity is nowadays widely used within AERA for all kinds of analysis beginning with polarization studies and investigations on basis of the magnitude of the electric field.

As a fundamental quantity of many different kinds of analyses, the $\vec{E}$-field required special attention towards identifying sources of systematic uncertainties during the reconstruction. Therefore the antenna response pattern, as they represent a central part of the detector description for this kind of reconstruction, are carefully investigated for their influence on the reconstruction. This part lead to several major improvements of the available antenna response patterns and also improved the methods how these are applied during reconstruction. As a fundamental part of the detector description, this also touches the aspect of detector simulations for the analysis of radio simulations. Therefore, both reconstructions of measurements as well as simulations will profit from this work.

With the complete $\vec{E}$-field vector and the improved detector description available, it is natural to compare the results of measurements with the ones of radio simulations. To this extend common radio simulations codes, e.g. REAS or MGMR, are compared with the first measurements of a prototype setup for AERA and also with first AERA measurements.

The final chapter is dedicated to an energy estimator for cosmic rays on basis of AERA measurements. In a first step, rooted on most recent results from LOPES and CODALEMA, a new parametrisation of the lateral distribution function (LDF) of cosmic rays is proposed on basis of radio simulations. This new parametrisation is a derivative of a Gaussian distribution function. In the next step it is compared to the common parametrisation of the LDF on basis of an exponentially decreasing function and proves to be superior in both describing single radio events and

radio simulations in general. Finally the Gaussian LDF is used to acquire an energy estimator for cosmic rays, which proves to be successful for measurements in comparison to SD reconstructions.

2

THEORY OF COSMIC RAY AIR SHOWERS AND THEIR RADIO EMISSION

2.1 THEORY OF COSMIC RAY AIR SHOWERS

2.1.1 *Cosmic rays: sources, propagation and interactions*

In 1912 Victor Francis Hess discovered during his famous balloon flights that the ionising radiation in the Earth's atmosphere increases in intensity with the altitude. The increase was interpreted as a hint towards an extra-terrestrial origin of the ionizing radiation that is nowadays known as cosmic rays. This discovery brought up numerous questions, of which some remain unanswered even today.

2.1.1.1 *Sources*

One of these questions concerns the sources of cosmic rays. Especially the highest energetic cosmic ray particles around or above 10^{20} eV require sources with similar acceleration energies. Based on the input from cosmic ray theory on acceleration mechanisms, it is possible to classify known astronomical objects by their acceleration potential. The most prominent acceleration mechanism, widely acknowledged throughout the cosmic ray community, is based on acceleration of charged particles in magnetic fields. The charged particles are forced by the Lorentz force on circular trajectories. A sufficiently large and strong magnetic field is thereby able to confine charged particles to the region of space where they are accelerated. The favoured electro-magnetic process for accelerating cosmic rays within an astronomical object is the first order Fermi acceleration as reviewed to some detail in the work of Drury [2012]. The maximum energy of this kind of electro-magnetic accelerator is limited by the particle energy required to break free from the magnetic field and leave the acceleration region.

The first order Fermi acceleration assumes an acceleration at shock fronts (c.f. Fermi [1949]). It requires a dense plasma shock front traversing through a comparable thin surrounding medium. At the boundary between the plasma shock front and the surrounding medium, a dense shock front manifests. This shock front is moving slower from the point of view of the particles behind the shock front, since the shock front is slowed down by the surrounding medium. Thus the particles behind the shock front see a closing shock front.

The situation is similar for the particles of the surrounding medium. These are slower than the shock front and therefore see also an approaching shock front. If charged particles hit the shock front they are reflected by the magnetic fields with an average energy gain through a stochastic process.

Since the reflected particles move within an magnetic field (which is at least partially generated by the shock front itself), they are deflected and have a chance to travel on a curved trajectory back into the shock front. If this happens, the particles can repeatedly reflected and gain additional energy with each crossing of the shock front. The process is limited by the energy required by one particle to break free from the confining magnetic field. Therefore the size and strength of the magnetic field provided by the accelerator corresponds to the maximum energy reached by the emitted particles.

Recently, theoretical calculations and simulations (c.f. Drury [2012]) indicated, that the magnetic field configuration of shock fronts can produce unbound magnetic field lines. Particles travelling alongside these field lines are naturally extracted from the source at energies far below the maximum acceleration energy. This effectively reduces the efficiency of the cosmic accelerator. The same simulations also showed, that the velocity of the shock wave has some influence on the efficiency of the acceleration. Theoretical calculations discussed in the same publication [Drury, 2012] also revealed that ultra relativistic shock velocities actually reduce the efficiency significantly. This makes non-relativistic shocks the most efficient accelerators for cosmic rays at highest energies currently known.

Another limiting factor on the maximum particle energy can be the limited live time of the cosmic ray accelerator. This is the case if the source object ceases to exists before the particles can gain the theoretical maximum energy accessible (c.f. Drury [2012]). The maximum energy of this kind of incomplete acceleration is then given by the lifespan of the source and no longer by its magnetic field size and strength. This eliminates several source candidates with respect to high energy particles. In summary the potential maximum acceleration energy of an astronomical source object depends primarily on its energy potential, magnetic field strength and size and secondary on its life time.

This fundamental relations can be investigated by Hillas-Diagrams, introduced in 1984 by Hillas [1984]. A recent example is given in figure 2.1.1. A number of known cosmic objects show magnetic fields large and strong enough providing enough energies to accelerate particles to the highest observed energies. It is however true, that only a few sources, e.g. active galactic nuclei, provide enough energy and magnetic field strength to accelerate particles up to the observed energy at several 10^{20} eV.

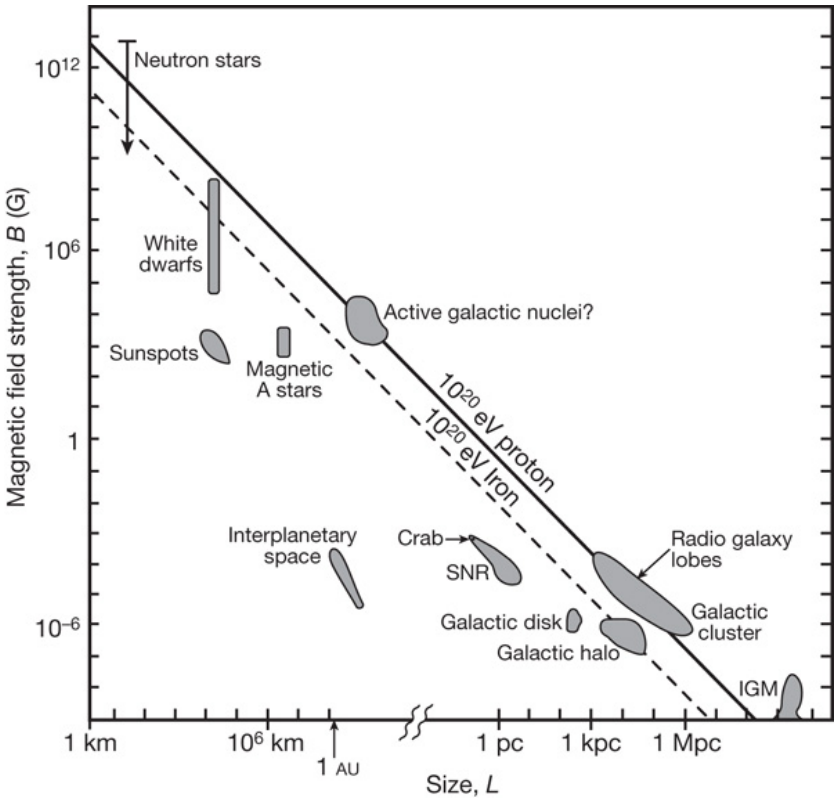


Figure 2.1.1: Hillas diagram as published by Bauleo and Martino [2009]. The magnetic field strength of cosmological objects vs. their size is displayed.

2.1.1.2 *Propagation*

At energies between 3×10^{18} eV and 10^{20} eV magnetic fields of galaxies are insufficient to provide an effective confinement and charged particles can break free to travel between galaxies.

Below 3×10^{18} eV the particles are limited to propagate inside the galaxy of their source object. Accordingly, all sources for particles below this energy are expected to be situated inside our galaxy. Namely supernova remnants[1] are potential accelerators for these particles and also one of the first and most discussed candidates.

At high and ultra-high energies cosmic rays can escape the magnetic fields of galaxies. Thanks to their high energies, these particles experience only a small deflection in the magnetic fields they encounter. This makes them especially interesting for astronomy in astro-particle physics (cf. Waxman et al. [1997], Takami and Sato [2008]), since their source position correlates with their arrival direction on earth. At lower energies, magnetic fields result in adiabatic deflections. This changes only the direction but not the energy or type of the cosmic ray particle. There-

1 large expanding gas shells from a supernova explosions, traversing the galactic medium

fore, the flight direction of the cosmic ray particle is no longer pointing backwards to the source after some time in a magnetic environment like our galaxy. The result is a flat arrival direction distribution. The magnetic fields effectively prevent cosmic ray astronomy at these energies.

Alongside to magnetic interactions also interactions with other particles can occur and change the type of the cosmic ray particle. These particle interactions do not interfere with the flight direction. Depending on the type (proton, iron, etc.) different interactions are possible with the cosmic microwave background (CMB), infrared, optical and ultra-violet photons the particles encounter during their journey.

For nuclei heavier than proton, i.e. with $A > 1$, three interaction processes with the background photons are important. These are the Giant Dipolar Resonance, the Quasi Deuteron Process and the Baryonic Resonance. All three processes require cosmic ray particle energies of the order of several 10^{16} eV. As a side effect, these processes do not only change the composition of the cosmic rays but also the energy, since the interaction products have a lower energy than the original cosmic ray particle before interaction.

Similar processes exist for proton cosmic rays interaction with the background. The dominant processes are the photo-hadronic interactions:

$$\begin{aligned} p\,\gamma &\rightarrow N + n\cdot\pi \\ p\,\gamma &\rightarrow p\,e^{+}\,e^{-} \end{aligned}$$

During these interactions, numerous pions together with a new nucleon or an electron-positron pair are produced (Bethe-Heitler Process). Both processes require cosmic ray particle energies of the order of 10^{20} eV for interactions with CMB photons. A much discussed additional effect is the Greisen Zatsepin Kuz'min effect (GZK) as proposed by Greisen, Zatsepin and Kuz'min in [1966]. It describes the interactions with the CMB and predicts for sufficiently long travel distances a cut off in the cosmic ray energy spectrum at around 10^{20} eV. Particles below this energy do not suffer from the effect, since the interaction with the CMB photons is suppressed as it is below the energy of the Giant Dipolar Resonance.

The cross section of the CMB interaction can be used to estimate the interaction length for protons. This interaction length immediately translates into a shell (horizon) around the observer (Earth) on cosmic scales. It is around 100 Mpc, which in general enables observations of higher energetic particles at Earth from sources within a shell with this radius. Outside the 100 Mpc horizon, the interaction probability increases and the particles

loose their energy through the occurring GZK interactions before they reach Earth. Accordingly, all sources of observed ultra-high energy cosmic rays (UHECR) are expected to be within the horizon of the GZK-effect. In recent years, observations of the Pierre Auger Observatory were correlated with available maps of active galactic nuclei (AGN) surveys (cf. the Pierre Auger Collaboration [2007], P. Abreu et al. - Pierre Auger Collaboration [2010, 2011a]), since AGN are a promising source for UHECR as displayed in figure 2.1.1. The comparison however showed only a slight correlation between the known AGN positions and the reconstructed incident directions of the UHECR on Earth. These first results are however not yet conclusive. Therefore the anisotropy of UHECR are under further investigation by several experiments (cf. the Pierre Auger Collaboration [2007], P. Abreu et al. - Pierre Auger Collaboration [2010, 2011a], M. Santander et al. - IceCube Collaboration [2012], T. Abu-Zayyad et al. - Telescope Array Collaboration [2012]).

Another explanation for the observed cut off in the spectrum could be in the nature of the sources itself. Since every source can only provide a limited amount of acceleration energy, another possible explanation for the cut of is a random coincidence of this natural limit with the GZK predicted cut off. This does in no case imply, that the GZK effect does not exist. In this case only its effect would be greatly suppressed, since the required energies are barely reached.

A differentiation between the two effects is however difficult. The composition of the UHECRs, if it can be associated with a distinct source type, could provide access to the question just like the observation of photons produced during GZK interactions.

Up to now, only the charged particles, which make up the majority of the cosmic rays, were discussed. Those can also be accompanied by uncharged cosmic rays like neutrinos or photons. A possible source of both are the above described interactions during the journey but also other sources like decay products of dark matter decays or annihilations are possible. Photons are in principle not affected by the interstellar magnetic fields and therefore suitable for cosmic ray astronomy.

Neutrinos on the other hand, are even more interesting for cosmic ray astronomy, since they travel unaffected by magnetic fields at near light speed for large distances without interactions with the cosmic material on the way. In case of super-nova explosions, measurements were already conducted and it was shown that the arrival time of the emitted light coincides with the arrival time of the neutrinos emitted from the signal-to-noise ratio (SNR) event (cf. Hirata et al. [1988]). This on one hand indicates no interactions of the neutrinos with the source material after emission and on the other hand supports theoretical mod-

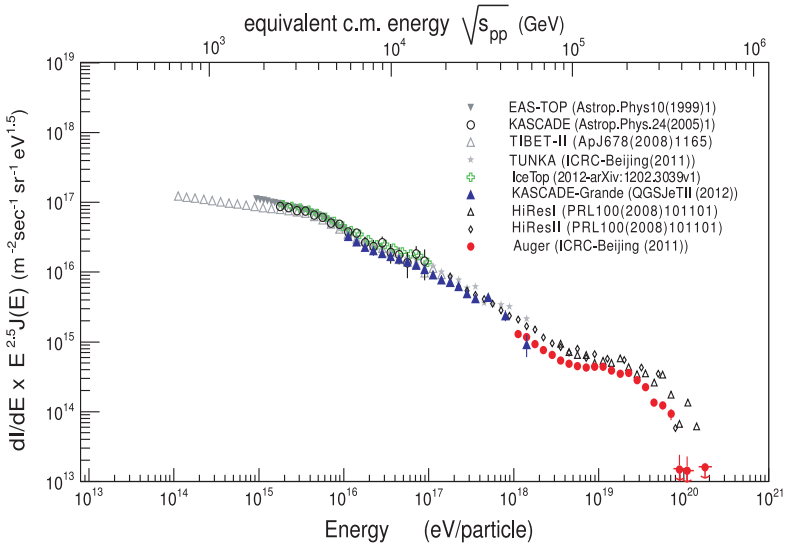


Figure 2.1.2: Combined cosmic ray energy spectrum above 10^{14} eV as measured by several experiments.

els for supernovae. Accordingly, neutrino induced air showers could be used to pin-point the source position in the sky.

Alongside of the highest energetic cosmic rays, also low energetic particles hit the Earth atmosphere. At energies up to 10^3 eV, the primary source is within our solar system: the Sun. Through its continuous ejection of solar winds, the Sun causes high particle fluxes, which also interact with the Earth atmosphere.

2.1.2 *The cosmic ray energy spectrum*

All effects discussed so far, either related to the source or to the journey, leave their marks in the energy spectrum of cosmic rays as measured on Earth and are described by Kotera and Olinto [2011] in their comprehensive review on UHECR.

In figure 2.1.2 the energy spectrum for cosmic rays is shown. The spectrum follows a power law with a spectral index of $s \simeq 2.7$ between 10^{10} eV and 10^{15} eV. Above 10^{15} eV the spectrum becomes slightly steeper with a spectral index of $s \simeq 3$. This feature is called the "knee" of the spectrum. This knee is generally accounted to a composition effect. Around this energy, first the proton flux decreases, followed by the heavier nuclei up to iron. The decrease of the all particle flux was extensively investigated by the KArlsruhe Shower Core and Array DEtector (KASCADE) experiment as described by W.D. Apel et al. - KASCADE-Grande Collaboration [2011]. The fundamental cause for this particle type dependent flux decrease can be described by two different theoretical models.

On the one hand, the galactic sources are simply reaching their maximum acceleration energy. The dependency on the particle type originates in this model from the higher acceleration of heavier nuclei, since these have a higher charge.

On the other hand, the knee can also be described by the modified leaky box model. In this model, the galactic magnetic field is no longer strong enough to confine the cosmic ray particles. Therefore an increasing amount of particles escape from our galaxy. Again, the energy threshold for escaping the galaxy depends on the charge of the particle, i.e. on the particle type, since the Lorentz force caused by the galactic magnetic field depends on the charge of the particles.

In principle, also a combination of both theoretical models can be the reason for the steepening of the spectrum. Only if the leaky box model is the dominating effect, then the observed flux of cosmic ray particles above the energy threshold has to be accounted to cosmic rays of extra galactic origin. However, the energy spectrum steepens at energies between 10^{15} eV to 3×10^{18} eV.

Around 3×10^{18} eV, at the so called ankle feature, the energy spectrum changes again its index. It then follows a power law with $s \simeq 2.6$. Up to today, there are differing theoretical models for this effect. All the models share, that they depend both on the injection spectra, i.e. the energy distribution of the particles before acceleration by the astronomical source, and on the composition of the cosmic rays at the highest energies. Therefore the measurement of the energy spectrum at the highest energies and of their composition is a still ongoing effort throughout the cosmic ray community. Accordingly a variety of experiments like the Pierre Auger Observatory [P. Abreu et al. - Pierre Auger Collaboration, 2011b] or the Telescope Array [Y. Tsunesada et al. - Telescope Array Collaboration, 2011] measure the composition at energies above 10^{18} eV.

Above several 10^{20} eV, the energy spectrum exhibits the above discussed cut off. This feature can be accounted to two different effects or a combination of both. First the GZK effect could be responsible, reducing effectively the energy of higher cosmic rays to a level below the threshold. And second, the energy of the sources could be insufficient to reach higher energies. Only future measurements with high statistic by large scale experiments like the Pierre Auger Observatory or the Telescope Array can provide the necessary insights on the sources and the spectrum to solve this mystery.

In summary, the cosmic ray energy spectrum is an important characteristic of cosmic rays, as it contains information about the sources, the composition and interactions with the interstellar or intergalactic medium.

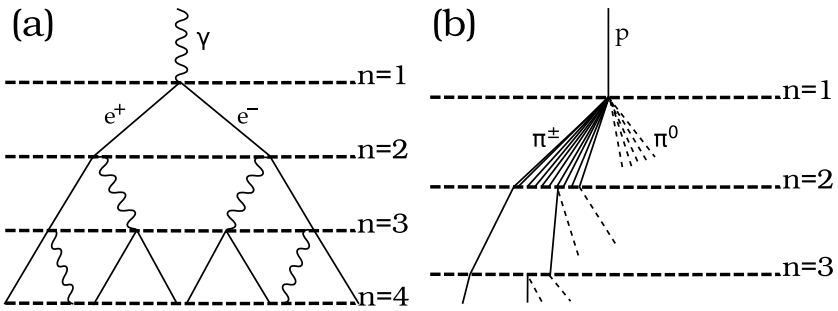


Figure 2.1.3: schematic model of air shower cascade interactions with (a) the electro-magnetic cascade (Heitler model) and (b) the hadronic cascade (Heitler-Matthews model).

2.1.3 *Air shower particle cascades*

Cosmic ray particles arriving at the Earth, interact with the molecules in the atmosphere. The primary particles have energy E_p, which is transferred through interactions with the atmosphere molecules and develop secondary particle cascades. One description of these air shower cascades is based on two similar models as a comprehensive introduction is given by Kampert and Unger [2012]. One model describes hadronic interactions (Heitler-Matthews model [Matthews, 2005]) and one describes the electro-magnetic interactions (Heitler model Heitler [1949]). A schematic overview over both models is given in Figure 2.1.3 on page 22.

The Heitler-Matthews model as described by Matthews [2005], assumes a primary inelastic interaction of the cosmic ray particle with the atmosphere resulting in secondary particles. The first interaction occurs after a distance that equals in average the mean free path of the cosmic ray particle in air. This distance therefore depends on the particle energy and on the cross section, i.e. the type of the particle[2]. The produced particles are primarily hadrons with enough energy to likewise produce more particles in inelastic follow up interactions. The result is a hadronic particle cascade in the atmosphere, i.e. the hadronic component of the air shower. It develops through several generations of interactions and starts to fade out, as soon as the mean free part of the produced hadrons exceeds their decay length. At this point, the hadrons start to decouple from the air shower and decay.

In the air shower cascade, the produced neutral Kaons decay into π^0 and contribute to the electro-magnetic shower as it is discussed below. The charged Kaons on the other hand, decay

2 The type of the particle is given by its mass, e.g. proton and iron are different particle types.

into their charged pion counterparts, which themselves decay further into charged muons via

$$\pi^{\pm} \rightarrow \mu^{\pm}\nu_{\mu}$$

The resulting muons[3] have a significantly larger lifetime than pions and therefore do not decay in great numbers during their travel through the atmosphere. They form the comparably small muonic component of the air shower.

The original hadronic component on the other hand, starts to die out as decay or radiation losses begin to dominate over the interactions. As a result, the air shower fades out in average after $n \geqslant 5$ interactions for primary particle energies above 10^{15} eV.

Over the whole air shower development, one third of the energy of each interaction is transferred from the hadronic component to the electro-magnetic component. This energy transfer is the result of the decay of the short lived neutral pions into photons. The neutral pion production ceases as soon as the mean free path of the particles starts to be larger than their decay length. Therefore the energy of the electro-magnetic component is

$$E_{em} = E_{primary}\left(1 - \left(\frac{2}{3}\right)^{n}\right)$$

For more than 5 interactions, the majority of the primary energy is transferred from the hadronic into the electro-magnetic component. Therefore this energy is also known as the calorimetric energy of the air shower, since the atmosphere behaves like a large calorimeter, which makes measurements of the electro-magnetic energy E_{em} sensitive to the air shower development.

One characteristic of the air shower cascade is the slant depth X_{max}, i.e. how deep in the atmosphere the maximum number of particles in the air shower is reached. This shower maximum is given by the development of the electro-magnetic cascade, which is usually initialised during the primary particle interaction as the interaction with the highest inelasticity (Heitler [1949]). In this interaction, a large number of neutral pions is generated (Figure 2.1.3 on page 22), which decay dominantly by two channels

$$\pi^0 \rightarrow \gamma\gamma \quad (99\%)$$
$$\pi^0 \rightarrow \gamma e^+ e^- \quad (1\%)$$

3 In general also charged muons contribute via decay into electrons to the electro-magnetic shower. The process is however negligible due to the large lifetime of the relativistic muon.

These decays form the electro-magnetic component of the air shower, which is predominantly made up of electrons and photons. The high energetic electrons and positrons emit Bremsstrahlung[4], which has enough energy for a process called pair production. Pair production produces in the vicinity of a nuclei out of a high energetic photon electron-positron pairs. The result is the electro-magnetic cascade developing alongside the hadronic cascade of the air shower. The electro-magnetic cascade fades out as the ionization energy of the particles exceeds their Bremsstrahlung losses. At this point, the electrons and positrons are absorbed by the atmosphere. The maximum of the electro-magnetic cascade and therefore of the air shower is approximately reached at:

$$\langle X^{em}_{max} \rangle \approx X_0 \cdot \ln {}^{E}/_{\varepsilon_c^{em}}$$

One secondary emission of the electro-magnetic component of the air shower is within the frequency range of the radio regime. This radio emission of cosmic ray air showers is discussed in more detail in the following section, as it is the source for the detection technique used throughout this work.

The air shower development shows statistical fluctuations as it is dominated by the mean free path of interactions and decays of its particles, which both are stochastic processes. Therefore it is not possible to predict the exact air shower development for a given primary particle even at fixed energy. In addition, today's accelerator experiments, like the Large-Hadron-Collider (CERN) (LHC) at European Organization for Nuclear Research (CERN), are unable to reach energies common to UHECRs. Therefore the cross-sections of the first high energetic interactions are extrapolations of known cross-sections at lower energies and accordingly only known with some degree of uncertainty. All together the developing air shower cascade can therefore vary widely. A high energy proton air shower, can for example either develop far up in the atmosphere, if the primary interaction occurred early, or deep in the atmosphere if it occurred later. This phenomena is called shower-to-shower fluctuations.

Even as the shower maximum X_{max} varies due to the shower-fluctuations, it still provides important information on the primary particle type. Due to the larger cross-section of heavier nuclei, e.g. iron, compared to protons, heavier nucleons interact in average earlier in the atmosphere than proton showers of the same energy. This however does not allow a differentiation between different particle types, since protons can also interact early in the atmosphere due to the stochastic nature of the in-

4 For muons produced in the air shower, the same is true, but due to their larger mass compared to electrons, the Bremsstrahlung from muons is suppressed compared to the one from electrons.

teraction. Therefore it is not possible to identify for a single air shower the primary particle type on the basis of an X_{max} measurement. However, it is possible to derive statistical information on the primary particle type for a set of air showers with similar primary particle energy by calculating the variation of the X_{max} values. Since proton air showers interact in average deeper in the atmosphere than iron air showers, it is possible to differentiate between iron pure or proton pure air showers for a sufficiently large statistic on the basis of the scattering of the X_{max} values. For this kind of analysis it is obviously important to know the energy of the air showers, preferably from an independent measurement. One of these techniques applied at the Pierre Auger Observatory in Argentina, is the measurement of the fluorescence light emitted by the air shower.

The fluorescence light is emitted, as the charged particles of the air shower cascade excite nitrogen molecules of the air into fluorescence. The emitted light is strongest within the wavelength of 300 nm to 400 nm.

2.1.3.1 *radio emission from cosmic ray air showers*

The radio emission from air showers depends dominantly on the spatial distribution and movement of the electrons and positrons in the air shower (Ludwig et al. [2012] and Scholten et al. [2012]). Contributions of heavier charged particles like muons are significantly suppressed due to their larger mass and can be neglected.

The emission can be described by several emission mechanisms. The radio emission of an air shower is however complex and therefore always a combination of all emission mechanisms.

One class of emission mechanisms is immediately caused by the geomagnetic field. The charged particles of the air shower are deflected by the Earth magnetic field since they experience the Lorentz force:

$$\vec{F} = q \cdot \left(\vec{E} + \vec{v} \times \vec{B} \right) \tag{2.1.1}$$

This separates electrons and positrons in the plane orthogonal to the air shower axis from each other, since electrons are deflected into the opposite direction compared to positrons. The result is a transverse current orthogonal to the shower axis and a net charge excess over the air shower evolution (Kahn and Lerche [1966]).

In addition, the separation of electrons and positrons causes a dipole momenta orthogonal to the air shower axis (Scholten et al. [2008]). The strength of this dipole varies over time and also emits a radio signal, therefore this emission is commonly called varying dipole contribution.

Another immediate contribution of the geomagnetic field is the emission of synchrotron radiation due to the change in the flight direction and thereby acceleration of the charge particles (Huege and Falcke [2003]). The geo-synchrotron emission is however only a minor contribution to the total radio signal.

Alongside the emissions from geomagnetic deflections, a net charge excess occurs as described by Askaryan [1962]. This charge excess is caused by the knock out of electrons from atmospheric molecules by the electrons and positrons of the air shower and, to a minor degree, by annihilation of positrons with electrons already present in the atmosphere. Both effects increase the number of electrons relative to positrons and result in a net negative charge excess varying over time. This time variation then contributes to the radio emission. This is the strongest emission of the non geomagnetic emissions[5].

Additional contributions result from relativistic effects, namely the Čerenkov effect. As a result, a radio emission with the typical Čerenkov cone is emitted. Another effect of the refractive index of the atmosphere is influencing the radio signal propagation in air as described by Ludwig and Huege [2011], Werner et al. [2012]. Depending on the density of the atmosphere, i.e. the hight above the ground, also the refractive index varies. Therefore radio signals from early stages of the air shower can arrive at an observer at the same time as radio signals from later stages of the air shower. This results in a compression or decompression of the radio signals at the ground, depending on the air shower geometry.

Another side effect of the relativistic movement of the air shower through the atmosphere is the beamed emission of the radio signal. This and the small pancake-like shower front with a thickness in the order of centimetres results in constructive interference of all emitted electro-magnetic waves with wavelengths in the order of the thickness of the shower front. Accordingly, the radio emission covers, depending on the geometry, the frequency range from 0 GHz to 10 GHz. Without this beamed constructive interference in forward direction no measurable radio signal would be emitted.

2.2 THE PIERRE AUGER OBSERVATORY AND THE AERA EXTENSION

The Pierre Auger Observatory [J. Abraham et al. - Pierre Auger Collaboration, 2010, I. Allekotte et al. - Pierre Auger Collaboration, 2008] is located in Argentina close to the city of Malargue. It consists of two detector systems, operated together to measure

5 In general, all described emissions from the geomagnetic field are commonly called the geomagnetic emission.

extensive air showers caused by cosmic rays. The bulk of the observatory is made of a 3000 km² Čerenkov detector array known as the SD. The Čerenkov detectors are placed in a hexagonal grid with a spacing of 1.5 km. Each SD unit consists of a water tank with three top mounted photo-multipliers. These SD tanks measure the Čerenkov light emitted by charged particles (in this case dominated by muons) with velocities near the speed of light as they pass through water. The SD detectors measure in principle day and night over a full year without pause and are only limited by maintenance down times. Thus SD achieves a duty cycle of close to 90 %.

The array is overseen by four FD telescope stations, Coihueco, Los Leones, Los Morados and Loma Amarilla. Each telescope station consists of 6 telescopes with a field of view close to 180°, which are only in clear dark nights able to measure the fluorescence light emitted by the air shower. Since the sensitive photo-multipliers necessary to detect the faint fluorescence light can be operated only during pitch black nights without a moon or bright stars in the field of view, the operational time of the fluorescence detectors is limited. Thus FD achieves duty cycles of only 13 %.

Both the FD and SD detectors send their data to the central campus where the central data acquisition is situated. There, the telescope operations are overseen and managed together with extensive atmospheric monitoring by LIght Detection And Ranging (LIDAR) and Laser measurements of the atmospheric conditions during FD measurements.

The Pierre Auger Observatory furthermore accommodates several extensions to both the SD and FD installations as well as new detector concepts.

Four of these extensions are located close to the Coihueco FD station. There the High Elevation Auger Telescope (HEAT) telescopes were installed, which are three additional fluorescence telescopes mounted on inclinabel mounts, such that their field of view is covering the sky above the existing Coihueco telescopes. The HEAT extension was installed to increase the sensitivity of FD towards low energetic air showers at energies of several 10^{17}eV. Those air showers develop higher in the atmosphere and have a weaker fluorescence light emission than higher energetic air showers. They thus emit their fluorescence light at a greater height, which is inaccessible to the other FD detectors.

For the same reason, the detection of low energetic air showers, another extension was installed within the existing SD detector field in front of Coihueco. The Infill installations are additional standard SD tanks deployed in a narrower grid with a spacing of 750 m. Since low energetic air showers have a smaller foot print of charged particles on the ground, the existing SD array

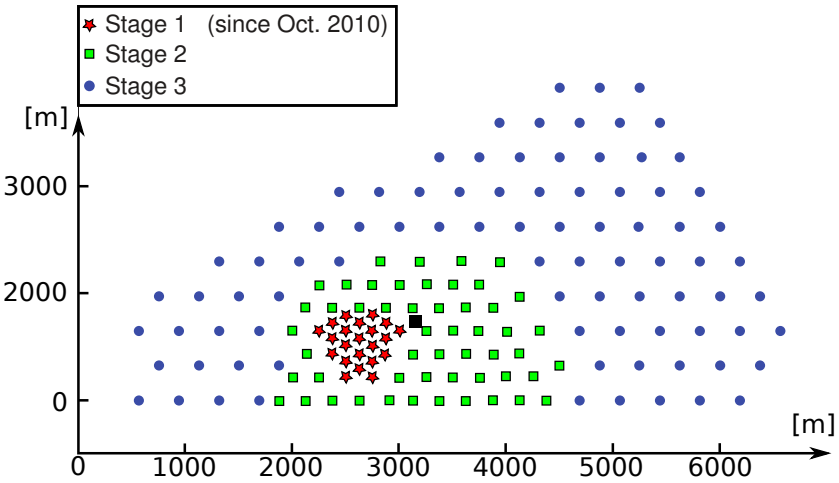


Figure 2.2.1: Map of the RD station layout to be deployed for AERA at the Pierre Auger Observatory.

was unable to detect these showers with enough stations for a reliable reconstruction of the air shower. Alongside the Infill extension to the SD array the Auger Myons and Infill to the Ground Array (AMIGA) [Kruppke-Hansen, 2011] extension is currently deployed. Buried into the ground, scintillation counters are used to measure high-energy muons. Both extensions aim to lower the energy threshold of SD from $\gtrsim 10^{18}$eV to $\gtrsim 10^{17}$ eV.

In the same area, AERA is going to be installed. The detector is able to measure the radio emission of air showers for energies above $\gtrsim 10^{17}$ eV, and complements the measurements of the Infill, HEAT and AMIGA detectors in the region. AERA has been designed based on the experiences made with the prototype set-ups previously deployed at the Pierre Auger Observatory and the LOPES [T. Huege et al. - LOPES Collaboration, 2010] and CODALEMA [Ravel et al., 2010] experiments. After completion, it will cover an area of approximately20 km² with 160 radio antennas (figure Figure 2.2.1 on page 28) . The antennas are deployed in a hexagonal grid, like the other ground detectors and will be installed in three stages. Stage one is operated since October 2010 and consists of 24 radio detectors (RD) stations with a spacing of 100 m.

An AERA detector station [S. Fliescher et al. - Pierre Auger Collaboration, 2010a] of the currently operated stage 1 as shown in figure 2.2.2 , consists of the radio antenna (logarithmic periodic dipole antenna (LPDA)) including the electronics, a solar panel for power supply, a GPS clock for timing information and a fence. The antenna type is only sensitive to the NS and EW polarization of the radio signal. The vertical component of the radio signal is available through reconstruction. The radio stations are connected via an optical fibre to the central radio station (CRS), a container housing the data acquisition hard-

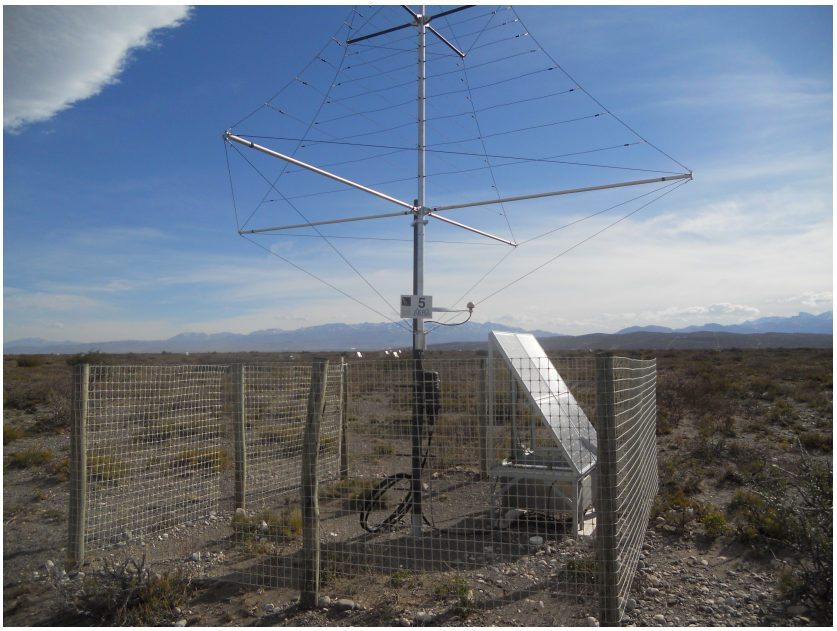


Figure 2.2.2: AERA RD station as deployed during October 2010. The station consists of the antenna and related electronics, a solar panel and a fence.

ware. The optical fibre facilitates the development of communication systems during stage 1. For stages 2 and 3, a wireless communication solution is foreseen. Each RD station is measuring autonomously, transmitting events to the CRS for high-level triggering and data acquisition. AERA is furthermore able to process an external trigger, for example from the SD or the Infill, alongside the self-triggered operation mode on basis of a three station coincidence. One difference between self-triggered and externally-triggered events is the amount of acquired data. Self-triggered events consist only of data gathered by triggered stations, while externally triggered events contain the gathered data of all stations, including stations, which have not seen a radio signal from an air shower. The RD stations measure at a bandwidth of 30 MHz to 80 MHz with a 200 MHz sampling rate. The current data rate is 5 MHz to 20 MHz depending on the applied trigger conditions and ambient radio noise level.

In preparation for AERA, several prototype stations were deployed. One of these is the so called BLS setup, which is named after the location close by the Balloon Launching Station. The Balloon Launching Station (BLS) was erected for weather and atmospheric monitoring purposes, and provided the infrastructure needed for several radio antennas. Therefore a total of nine radio stations were deployed in 2006 in an equilateral triangle with a baseline of 100 m as described in A. M. van den Berg et al. - Pierre Auger Collaboration [2007]. Alongside the radio antennas a particle detector was installed to provide an external

trigger for the radio antennas. The deployed radio antennas are a precursor of the LPDA antennas used at AERA. The setup has a bandwidth of 25 MHz to 75 MHz. The BLS setup is also deployed inside the SD field, which allows the combined analysis of SD coincident BLS measurements.

To exploit the full potential of AERA, hybrid measurements together with the SD and later with the FD detectors has to be done. Today, SD and RD gathered data can already be combined. Since both detectors trigger independently, these SD-RD hybrid events are associated after data acquisition. This association is currently based on matching the trigger timestamps and in a second step a matching of the reconstructed incident directions of the two detectors within their respective uncertainties. For this task and for further analysis and reconstruction of the radio events, a high capacity analysis framework is important.

2.2.1 *The Offline framework*

The Offline framework was originally developed for the analysis of SD and FD data as described by S. Argiro et al. - Pierre Auger Collaboration [2007], J. Allen et al. - Pierre Auger Collaboration [2008] and most recently by Gonzalez [2012]. With the design and development of AERA, additional functionality was incorporated into the framework, especially addressing the needs of the analysis of RD data and radio simulations as discussed by T. Huege et al. - Pierre Auger Collaboration [2011]. From the beginning, Offline was designed to be flexible enough to provide also a basis for data analysis of future extensions to the Pierre Auger Observatory like it is AERA. The structure of Offline is therefore modular and by intend flexible and adaptable. The radio analysis extension sticks to this design and provides a flexible frame for the analysis of both data and radio simulations even from differing set-ups.

A typical analysis of data with Offline consists of a sequence of analysis modules, which are provided by the framework and configured through extendable markup language (xml) configuration files. These modules can easily be exchanged providing the necessary flexibility to analyse different types of data like SD, FD or in this case RD.

This also allows for an easy comparison of analysis methods or algorithms implemented in modules, since they can be easily interchanged. The result of the reconstruction can afterwards visualised by an event browser. In Figure 2.2.3 on page 31, a typical reconstructed radio trace is given as an example for the representation given in the event browser.

To improve the usability of the framework, several standard and example application like Offline configurations are available.

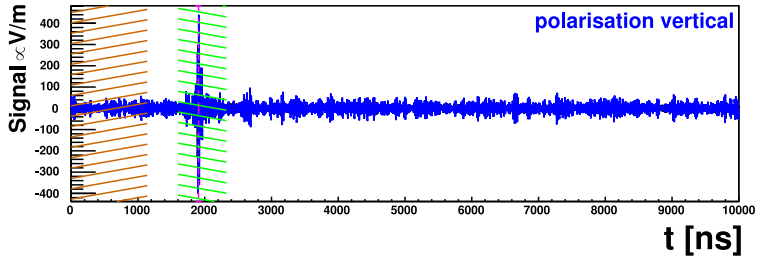


Figure 2.2.3: Example radio trace for one radio station for a typical radio reconstruction as reconstructed by Offline and displayed in the event browser. The signal and noise windows are marked in green and red, as well as the position of the reconstructed peak (magenta line).

These provide the user with all he needs to adapt an analysis to its needs. All modules can be provided with parameters for the implemented algorithms to influence their behaviour. Therefore the modules can be adapted to suffice the needs of different setup, data formats or analyses. As an example, the signal search window is given, which makes it possible to provide information on when and within which uncertainties the radio signal is expected. This information can be available from external triggers, e.g. by SD measurements. One central module (algorithm) fundamental to every radio reconstruction is the RdSignalReconstructor, which is discussed below.

By design the detector related information is separated from the measured data and currently available through xml files within the framework. For the radio extension, the detector description is extended to cover also time dependent detector information. This is especially important, since AERA is in a secondary role used to develop the necessary hardware for present and future radio detectors and thus experiences frequent changes in its hardware setup. Accordingly it is crucial to provide within the framework a flexible method for describing the detector at any given time.

Offline and the radio extension [T. Huege et al. - Pierre Auger Collaboration, 2011] provide through their modules an easy to use interface to the underlying functionality of the framework. Thus the end-user only needs basic knowledge in C++ to perform an analysis.

As it is not uncommon for a complex software like it is the Offline framework, mistakes during implementations of analysis algorithms can occur. These can result in a large variety of problems, which are not always easy to spot. At short notice for this work, such a mistake was discovered in the source code of the radio analysis algorithms in early August 2012. The problem was introduced during the first implementations of radio reconstruction algorithms to Offline in 2007 and specifically touches

the analysis of radio simulations. A possible influence on the reconstruction of radio measurements is not yet established but still under investigations. Therefore it can be necessary to update several analyses of the past couple of years. In this work, only the BLS analysis discussed in 5 is affected, as the available time was insufficient for a complete update of this analysis.

Misreconstructions can in principle occur for radio simulations with an azimuth incident direction between 270° and 360°. For this azimuth range, the $\vec{E}$-field vector seems to be subtly changed as a consequence of an interface problem.

2.2.1.1 *The radio signal reconstruction algorithm of* $\overline{\text{Off}}\underline{\text{line}}$

The heart of the signal reconstruction algorithm is the non-trivial signal definition. In practice, this includes all information available on the expected radio signal for a given radio timeseries, including shape, arrival time, expected signal strength, etc. In the most general case implemented in $\overline{\text{Off}}\underline{\text{line}}$, it is assumed that the signal is defined by its strength relative to the noise level, i.e. the SNR. The noise level necessary for this calculation is currently acquired by calculating the standard deviation for a given part of the timeseries. This part, the noise window, has to be chosen such, that no signal is expected in it. For the time being, this information has to be provided setup specific by the user. It is furthermore possible to switch from a simple SNR to the squared SNR. This is especially beneficial, for timeseries of single $\vec{E}$-field components, since the oscillations are between positive and negative vales (c.f. Figure 2.2.3 on page 31). Finally it is possible to limit the algorithm on a part of the timeseries, the signal window.

If the algorithm is applied, it searches the trace for any radio peak fulfilling the requirement of the SNR and saves the pulse position in the timeseries, i.e. the signal time for later reference. If no signal was identified on basis of the SNR cut, this information is also stored for later reference.

Since the introduction of the three-dimensional $\vec{E}$-field in $\overline{\text{Off}}\underline{\text{line}}$, it is also possible to specify if one single component or the magnitude of the $\vec{E}$-field timeseries should be used. This can be important for polarization studies, that are only interested in one component of the $\vec{E}$-field.

For this work, all described analyses are using the magnitude of the $\vec{E}$-field and the signal is identified by an squared SNR if not stated otherwise.

Part I

ON THE THREE-DIMENSIONAL ELECTRIC FIELD OF COSMIC RAY AIR SHOWER RADIO SIGNALS

3

RECONSTRUCTION OF THE ELECTRIC FIELD VECTOR

The radio emission from cosmic ray air showers was if not exclusively but certainly most intensely investigated in recent years. The early LOPES experiment, the CODALEMA experiment, the LOw Frequency ARray (LOFAR) experiment and even the prototype stations for the AERA experiment are witnesses of this effort. Together with older experiments, they share one communality when it comes to the $\vec{E}$-field: All radio experiments measure only an incomplete projection of the $\vec{E}$-field.

Usually two projections, the $\vec{E}$-field component of the east-west and the north-south direction, are measured. This incomplete information on a three-dimensional quantity like the $\vec{E}$-field is limiting any subsequent analysis. As a result, the LOPES experiment ventured forth and deployed in its last stage radio stations, which are able to measure all three projections of the $\vec{E}$-field, i.e. east-west, north-south and vertical. Without such a direct and complete measurement of the $\vec{E}$-field vector, the only option is a reconstruction of the $\vec{E}$-field using additional information to compensate for the missing projection measurement.

This chapter describes the first method to do such a reconstruction for radio signals emitted by air showers on basis of the incident direction and by exploiting the transverse nature of the radio signal. This new reconstruction method makes AERA the first radio detector capable of reconstructing the complete $\vec{E}$-field from an incomplete two-dimensional projected $\vec{E}$-field measurement.

The radio emission of cosmic ray air showers is a transverse electro-magnetic wave evolving with time and described by a timeseries of three dimensional real $\vec{E}$-field vectors. This timeseries $\vec{E}(t)$ is equivalent to the frequency spectrum $\vec{E}(\nu)$, which is accessible through Fourier transformations. In subsequent discussions, for the sake of transparency but w.l.o.g., a single $\vec{E}$-field vector of the frequency domain is considered. This complex vector contains complete amplitude and phase information of the radio signal.

Measurements of the $\vec{E}$-field in the MHz range are done with radio antennas like the LPDA antenna at AERA. The $\vec{E}$-field of the radio signal induces a current into the radio antenna that

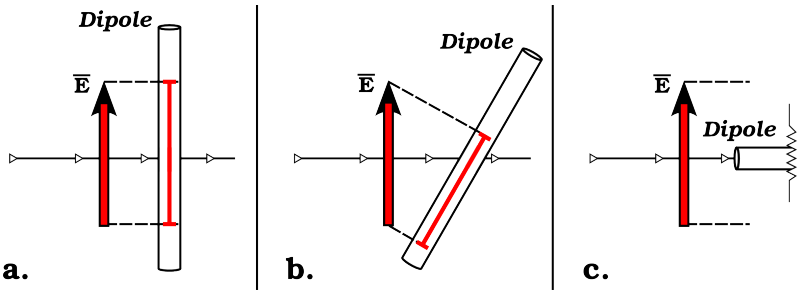


Figure 3.0.2: Projections of different linear polarised $\vec{E}$-field vectors onto a canonical dipole. a) full $\vec{E}$-field, b) projected $\vec{E}$-field and c) no $\vec{E}$-field measured

results into a measurable voltage difference at the terminal of the antenna.

It is of special interest to reconstruct from this voltage the physical three dimensional $\vec{E}$-field vector, since it contains the complete information on all components of the radio signal. Based on the $\vec{E}$-field vector, it is e.g. possible to directly compare measurements to simulations (c.f. chapter 5), to do investigations on the radio LDF (c.f. chapter 6) and to do reconstructions of air shower quantities like the primary particle energy of the cosmic ray (c.f. chapter 6).

Figure 3.0.1: The three crossed dipoles of a LOPES-3D antenna station

A reconstruction of the $\vec{E}$-field vector naturally starts with the measurement process itself. This includes two effects, projections of the $\vec{E}$-field on the antenna and any distorting influence of the antenna like dampening or dispersion effects. The projections are best described on the example of a dipole antenna, which is only sensitive to $\vec{E}$-fields parallel to its main axis. Therefore a measurement with a single dipole is accompanied by a loss of information. Figure 3.0.2 illustrates this for $\vec{E}$-field vectors linearly polarised in the plane of the dipole.

Since the measurement with a single dipole is insufficient to reconstruct the original three dimensional vector, a common practice evolved to combine several antennas at the same position but aligned in different directions. That way they are sensitive to

different projections of the $\vec{E}$-field vector. This combination of antennas, including all associated electronics is called a station. For the discussed example dipole antenna, a combination of three orthogonal aligned dipoles as shown in figure 3.0.1 form a station sensitive to the full $\vec{E}$-field vector[1]. This approach is only applied in the most recent deployment of the LOPES experiment [D. Huber et al. - LOPES Collaboration, 2011], which makes it so far the only experiment measuring the complete $\vec{E}$-field. Alongside the projections, also the influence of the antenna's characteristic has to be taken into account. The antenna characteristic provides information on the sensitivity of the antenna, which in general depends on the frequency of the received signal. Since the radio signal covers a broad range in the frequency domain (broadband signal), the antenna shows a frequency dependent influence on the amplitude and phase of the radio signal.

Both the effects from the projections and the dispersive effects of the antenna characteristic are described by a complex matrix of effective antenna heights[2]: the antenna response pattern.

This matrix description is similar to a quantum mechanical description of measurement processes. There a measurement operator, which also can be written as a matrix, is applied to an eigenstate to describe a measurement. In a similar way, the response pattern describes the measurement process with a radio antenna. The dimension of this matrix is given by the number of projections plus the incident direction and the frequency range. In case of AERA, this adds up to 5 dimensions, three for the frequency and incident direction and two for each one of the crossed antennas. The process is described by the equation below, which results in the voltage at the terminal of the antenna:

$$\vec{U} = \bar{\bar{R}}(\theta, \phi, \nu) \cdot \vec{E}(\theta, \phi, \nu) \tag{3.0.1}$$

with:

θ	zenith angle of the incident direction
ϕ	azimuth angle of the incident direction
ν	frequency
$\vec{E}$	$\vec{E}$-field vector of the incoming wave
$\bar{\bar{R}}$	two dimensional response matrix
$\vec{U}$	vector of measured voltages at the terminal of the antenna for different projections of the $\vec{E}$-field

1 The before mentioned LPDA antenna of AERA is actually a combination of two orthogonally oriented LPDA antennas, one facing to the east-west and one facing to the north-south.

2 the effective antenna height gives both amplitude and phase corrections for the electric field.

Algorithm 3.1 Offline module sequence for reconstruction of AERA measured events

```
<moduleControl>
    <loop numTimes="unbounded">
        <module> EventFileReaderOG                  </module>
        <module> RdEventInitializer                 </module>
        <module> RdChannelADCToVoltageConverter     </module>
        <module> RdChannelPedestalRemover           </module>
        <module> RdChannelResponseIncorporator      </module>
        <module> RdChannelUpsampler                 </module>
        <loop numTimes="unbounded">
            <module> RdAntennaChannelToStationConverter </module>
            <module> RdStationSignalReconstructor       </module>
            <module> RdDirectionConvergenceChecker      </module>
            <module> RdPlaneFit                         </module>
        </loop>
        <module> RdEventPostSelector                </module>
        <module> RecDataWriterNG                    </module>
    </loop>
</moduleControl>
```

A reconstruction of the $\vec{E}$-field requires a solution of the inverse this matrix equation:

$$\bar{\bar{R}}^{-1}(\theta,\phi,\nu)\cdot\vec{U}(\theta,\phi,\nu) \quad = \quad \vec{E}(\theta,\phi,\nu) \tag{3.0.2}$$

Both equations are defined on the frequency spectrum of the radio signal and not on the measured timeseries, since in the frequency domain amplitude and phase information are easily described by complex numbers. Only then, the application of the response pattern is a simple matrix multiplication with the complex $\vec{E}$-field as given in equation (3.0.1) and equation (3.0.2).

3.1 Reconstruction of the $\vec{E}$-field with the Offline-framework

In case of a partial $\vec{E}$-field measurement like at AERA the situation is complicated, since only two orthogonal projections of the $\vec{E}$-field vector are measured instead of three. Therefore the equation system cannot be solved without further information.

It is however possible to reconstruct the physical $\vec{E}$-field vector, if the incident direction of the transverse $\vec{E}$-field is known. Following the reconstruction of the $\vec{E}$-field for this case is discussed on the example of a standard reconstruction sequence of Offline for AERA measured radio events:

In the first step, the raw measured analog-to-digital converter (ADC) timeseries[3] of all participating stations for the radio event are read into Offline (RdEventFileReaderOG), followed by defin-

3 Instead of a single $\vec{E}$-field vector or voltage, a timeseries of $\vec{E}$-field vectors is usually measured, since the radio signal is a broadband signal with a flat frequency spectrum. Obviously the signal is affected by the bandwidth of the setup, which spans in case of AERA the region of 30MHz - 80MHz.

ition of a local coordinate system on the ground (RdEventInitializer), conversion of the raw ADC counts into voltages at the terminal of the antenna (RdChannelADCToVoltageConverter), corrections for characteristics of the electronics (RdChannelPedestalRemover and RdChannelResponseIncorporator) concluded by up-sampling of the coltage timeseries (RdChannelUpsampler). At this point, the voltage timeseries is independent of electronic characteristics like cable delays or dispersion effects and ready for the reconstruction of the physical $\vec{E}$-field timeseries.

Since at this stage of the reconstruction the incident direction of the air shower is still unknown, I developed an iterative reconstruction method to reconstruct both the $\vec{E}$-field and the incident direction simultaneously.

At the beginning of the iterative loop in line nine of module sequence 3.1, an initial incident direction is assumed[4] and via a Fast Fourrier Transformation (FFT) the frequency spectrum of the radio signal is calculated. The matrix equation system

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

is now solved in the frequency domain (line 10, RdAntennaChannelToStationConverter) for the assumed incident direction and reduces to:

$$\vec{U}\,(\nu) \quad = \quad \bar{\bar{R}}(\nu) \cdot \vec{E}\,(\nu)$$

This three dimensional equation can be further reduced to a two dimensional problem by transformation into the coordinate system of the transverse wave. This coordinate system is defined by two coordinates:

- Φ orthogonal on the z-axis and parallel to the ground ($\parallel$ to the azimuth plane)
- Θ orthogonal on the z- and Φ-axis, pointing away from the zenith

The coordinate system is also illustrated in figure 3.1.1.

4 The initial direction is set to 45° azimuth and 5° zenith, since the two antennas of one AERA station face to the E-W (0°) and to the N-S (90°). Investigations showed, that the reconstruction is independent of the initial direction, if the timeseries is not dominated by ambient radio noise.

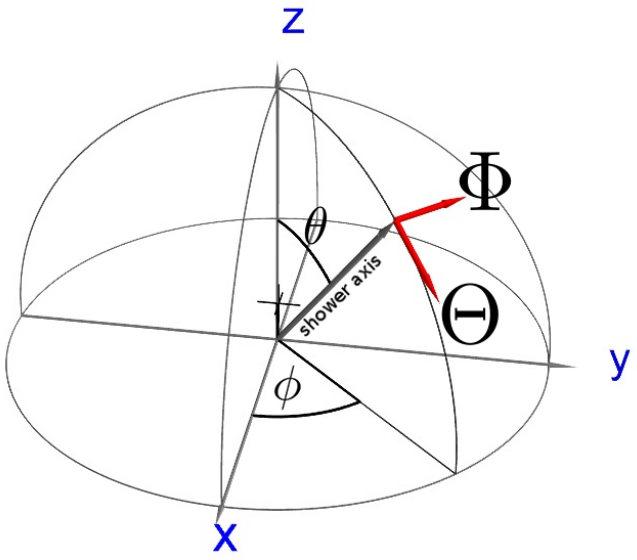


Figure 3.1.1: Response pattern related coordinate systems. The shower coordinate system is defined by the Θ and Φ coordinates (red axes). Θ faces away from the zenith, Φ is orthogonal on Θ and the shower axis and parallel to the ground. Together they form a Cartesian coordinate system. The Antenna on the other hand is in the centre of the local x, y, z coordinate system. In case of the Offline convention, the x-axis faces to the east, y to the north and z towards the zenith.

Since the $\vec{E}$-field vector is orthogonal on the incident direction, the equation system is reduced to a two-dimensional problem in this representation[5]:

$$\begin{aligned}
\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} &= \begin{pmatrix} R_{1\Theta} & R_{1\Phi} \\ R_{2\Theta} & R_{2\Phi} \end{pmatrix} \cdot \begin{pmatrix} E_\Theta \\ E_\Phi \end{pmatrix} \\
\Rightarrow U_1 &= R_{1\Theta} \cdot E_\Theta + R_{1\Phi} E_\Phi \text{ and} \\
U_2 &= R_{2\Theta} \cdot E_\Theta + R_{2\Phi} \cdot E_\Phi \\
\Rightarrow E_\Theta &= \frac{U_1 - R_{1\Phi} \cdot E_2}{R_{1\Theta}} \text{ and} \qquad (3.1.1) \\
E_\Phi &= \frac{U_2 \cdot R_{1\Theta} - U_1 \cdot R_{2\Phi}}{R_{1\Theta} \cdot R_{2\Phi} - R_{2\Theta} \cdot R_{1\Phi}} \qquad (3.1.2)
\end{aligned}$$

In a first step the E_Φ-component is calculated. The result of equation (3.1.2) is then inserted into equation (3.1.1) to calculate E_Θ. A trivial coordinate transformation from the two-dimensional coordinate system (CS) given by the coordinates (Θ, Φ) to the global three-dimensional coordinate system with coordin-

5 The response pattern has also to be defined in the reduced coordinate system, which is common practice, since the requirements for storing the response pattern also reduces from three to two dimensions.

ates (east-west,north-south,vertical) yields the three-dimensional $\vec{E}$-field.

After an inverse FFT, the resulting $\vec{E}$-field vector timeseries can now be used to identify the radio pulse in the timeseries[6] (RdSignalReconstructor). The radio pulse is in this case defined as the highest magnitude of the $\vec{E}$-field vector in a window of the timeseries. The pulse time is given by the time at which the highest magnitude occurred[7].

Based on the pulse arrival times reconstructed in this manner for each station of the radio event, the time differences between the stations is used to reconstruct the incident direction by a simple plane fit. In the RdPlaneFit module, the pulse arrival times of the stations are fitted under variation of the incident direction to the plane wave model.

The resulting incident direction is then used in the next iteration to solve again the equation system (3.0.1) based on the new incident direction.

Both the plane fit and the $\vec{E}$-field reconstruction mutually affect each other, since on the one hand each new direction results in slightly different antenna responses that on the other hand change the pulse timing, which in the next step changes again the reconstructed incident direction.

It has to be stressed, that the plane fit assumes a plane wave for the radio signal from the cosmic ray air showers. This is well motivated on small scales and small distances between the stations, since the uncertainties on the reconstructed arrival times makes the detector insensitive to possible curvature effects of the wave front. In general, however, a spherical wave front is better motivated by electro-dynamics and recent investigations by F. G. Schröder et al. - LOPES Collaboration [2011] even indicate that the radio signals of cosmic ray air showers show a conical wave front. The Offline framework is in principle providing the facilities to test other models, like the conical wave front, in the future.

This iterative reconstruction can be continued indefinitely, if no convergence criterion is provided to stop the iterative reconstruction as soon as a consistent reconstruction of the incident direction and the $\vec{E}$-field is reached.

The applied criterion in this case, is a limit on the change in the reconstructed incident directions in between two iterations. Per default it is required that the angular difference between

6 To avoid confusion, the also common term "signal" is not used in this work for this context, since the radio pulse in the timeseries is only a transformation of the radio signal of the cosmic ray air shower.

7 The signal window position is given by the time delay between triggering and readout of the measured ADC timeseries in the field. Its width is given by the uncertainties.

two successive incident direction reconstructions is below 10^{-5} degrees.

Furthermore it is necessary to avoid uncontrolled behaviour of this in principle endless iterative loop. The RdDirectionConvergenceChecker module therefore terminates the reconstruction latest after 10 iterations unsuccessfully.

During the analysis of BLS and AERA data, the algorithm converged in general within 3 to 4 iterations. Also no dependence on the initial conditions for the reconstructed incident direction was observed. However, in absence of a significant radio pulse in the timeseries, which is the case if the timeseries is dominated by ambient radio noise, the iterative reconstruction is not conclusive and therefore terminated by the RdDirectionConvergenceChecker module.

The reconstruction sequence is concluded by the RdEventPostSelector module, which ensures that only events with a successful reconstruction are written to disk by the RecDataWriter module.

With the presented method, AERA is the first air shower radio experiment having the ability to reconstruct the full three-dimensional $\vec{E}$-field from measured voltages including the full information content of the antenna response. This is a major improvement not only for the experiment but also of the reconstruction algorithms compared with former approaches for which only partial information of the antenna response was considered or for which the measured projections of the $\vec{E}$-field strongly depended on the antenna orientation.

For example, studies of components of the radio signal (like the east-west or the north-south portion) required carefully oriented antennas to measure directly the respective projection (e.g. the LOPES 30 setup). This is no longer required by the iterative reconstruction method, which depends only on a general knowledge of the orientation of the antenna, which can be arbitrarily chosen. As a benefit, the reconstructed $\vec{E}$-field and portions of it independent of the underlying antenna orientation.

Former applied approaches to treat the antenna response, like the one applied in the LOPES experiment, did not include the complex phase information. This resulted in increased systematic uncertainties for the reconstruction, which now can be avoided, since the antenna response is applied in the frequency domain where both amplitude and phase effects are covered. Only recently this improved treatment of the response pattern was also incorporated into the LOPES data analysis, as the LOPES experiment was reconfigured to LOPES 3D [D. Huber et al. -

LOPES Collaboration, 2011, W.D. Apel et al. - LOPES Collaboration, 2012a].

The three-dimensional $\vec{E}$-field also improves subsequent analysis, since it holds a lot more information than single projections. For instance, studies on the emission mechanism of radio signals from cosmic ray air showers depend on the east-west and north-south portions of the measured signal; quantities that directly profit from the improved reconstruction. Based on the magnitude of the $\vec{E}$-field, it is now also possible to investigate correlations between the total power of the radio signal and the energy of the primary cosmic ray particle as for example in chapter 6.

However, the precision of the three-dimensional $\vec{E}$-field reconstruction is severely depending on how well the antenna characteristic is known, i.e. the antenna response patterns (c.f. equation (3.0.2)). Since it was expected, that the methods applied to prepare the antenna response patterns for $\overline{\text{Off}}\underline{\text{line}}$ could be insufficient to guarantee a high data quality, it was mandatory to have a closer look into this matter.

4

RESPONSE PATTERNS OF RADIO ANTENNAS AND THEIR INFLUENCE ON RECONSTRUCTIONS

The reconstruction of the complete three-dimensional $\vec{E}$-field vector is an important step, as this quantity forms the basis for many different subsequent analyses like the lateral distribution of radio signals, polarization studies or reconstructions of the energy or X_{max}. Therefore special attention has to be paid to sources for systematic uncertainties like the quality of the antenna response patterns. Therefore it is mandatory to do a thorough check of these patterns and the applied methods to improve the whole reconstruction.

4.1 GENERAL CONSIDERATIONS ON RESPONSE PATTERNS

In general, response patterns are available from two different sources.

1. direct measurements in a controlled environment

2. simulated by software like the antenna characteristic simulation (4nec2)

Both methods provide response patterns for discrete incident directions and frequencies. To use these patterns during reconstruction requires an interpolation. Late in this chapter this issue will be addressed by introducing a comparison of two interpolation methods.

Response patterns of antennas are usually governed by the geometrical shape of the antenna and the environment (e.g. the reflectivity of the ground). The technical aspects and material related characteristics like the resistance of the antenna are relative to the above mentioned of secondary importance.

4.1.1 *Structure of response patterns*

Mathematically, response patterns are described by a matrix of effective antenna heights (also known as antenna factors). Effective antenna heights are complex numbers for a fixed incident direction and frequency, giving the fraction of the incident elec-

tric field strength with respect to the voltage at the terminal of the antenna (in $\frac{1}{m}$):

$$\begin{aligned} U &= \vec{H}(\varphi, \theta, \nu) \cdot \vec{E}(\varphi, \theta, \nu) \\ &= H_\Theta \cdot E_\Theta + H_\Phi \cdot E_\Phi \end{aligned} \tag{4.1.1}$$

with:

$\vec{E}$	$\vec{E}$-field vector of the incident radio signal
$\vec{H}$	response pattern, here a vector of effective antenna heights
U	measured voltage at the terminal of the antenna

The response pattern $\vec{H}$ as well as the electric field strength of the radio signal $\vec{E}(\theta, \varphi, \nu)$ are always defined in the two dimensional coordinate system with the components (Θ, Φ) (see Chapter 3). The coordinate system is illustrated in Figure 3.1.1 on page 40 (red coordinate system). A common convention is to give the voltage U with respect to a resistance of 50Ω at the terminal of the antenna.

From (4.1.1) it is evident, that the response pattern separates into two components in analogy with the underlying coordinate system (Θ, Φ): H_Θ, H_Φ.

Therefore, a response pattern for one antenna actually consists of two three-dimensional matrices of complex numbers, one for each component (Θ, Φ) (with the dimensions ν, θ, ϕ). Each matrix element represents one effective antenna height factor.

For the subsequent discussion of the response patterns available in $\overline{\text{Off}}\underline{\text{line}}$, several plots are presented, which follow a common scheme. Since it is difficult to visualise the three-dimensional response pattern directly, it was chosen to display two-dimensional slices for sake of transparency:

1. azimuth sweeps at fixed frequency and zenith angle
2. zenith sweeps at fixed frequency and azimuth angle
3. frequency sweeps at fixed azimuth and zenith angle

Each sweep[1] is composed of six plots. One for the amplitude of the effective antenna height, one for the phase of the effective antenna height and one for each component Θ, Φ and the frequency ν. For instance the sweeps for the Φ-component of the Alu-LPDA antenna are shown in figures 4.1.1a, 4.1.1b and 4.1.1c.

The fixed parameters of the sweeps are chosen to be well inside the parameter range of the response patterns in the $\overline{\text{Off}}\underline{\text{line}}$ framework:

1 A sweep defines one parameter, e.g. the frequency, is varied within its limits while the other parameters, e.g. zenith and azimuth, are fixed to one given value. This corresponds to a one-dimensional slice of the response matrix.

frequency (ν)	50 MHz
azimuth (ϕ)	30°
zenith (θ)	60°

Sweeps were done for the following antenna response patterns, of which only a selection can be presented here:

- Alu-LPDA (AluLPDA_ground2[2])
- Blackspider at the BLS (BlackSpider_atBLS_ground2)
- Butterfly (Butterfly_freespace and Butterfly_ground2)
- SALLA (Salla_freespace)
- Small Blackspider at the BLS (SmallBlackSpider_atBLS_ground2)
- Small Blackspider (SmallBlackSpider_ground2)

With the SmallBlackSpider_atBLS_ground2 pattern being the most important, since it describes the antenna currently used at AERA. These patterns are first simulated by the 4nec2 software. Afterwards they are manually prepared for use in Offline and included into the framework as XML files. The simulation and preparation of the antenna patterns for Offline was done by Stefan Fliescher [Fliescher, 2009 - 2011].

4.1.2 *Physical features in response patterns*

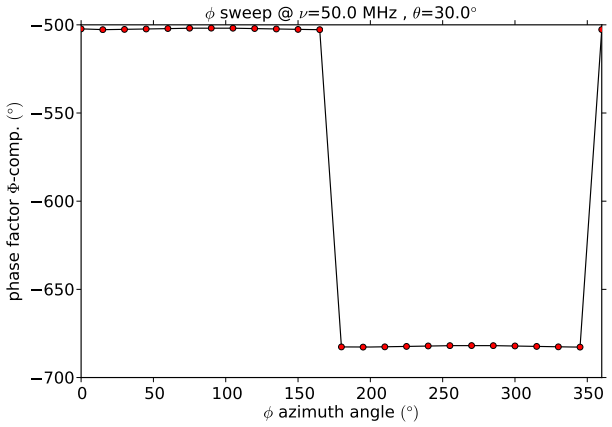
In Figure 4.1.1 on page 48 several characteristic sweeps of the AluLPDA_ground2 pattern are presented. Several prominent features can be observed, which have a physical source.

In 4.1.1a an azimuth sweep is shown of the phase of the effective antenna height. Several distinct phase jumps of 180° can be observed for two directions, 180° and 360°. These are the insensitive directions of the corresponding antenna. For these directions, the amplitude of the effective antenna height approaches zero (Figure 4.1.1b on page 48) and therefore the measured signal vanishes as expected for this direction and a phase jump occurs.

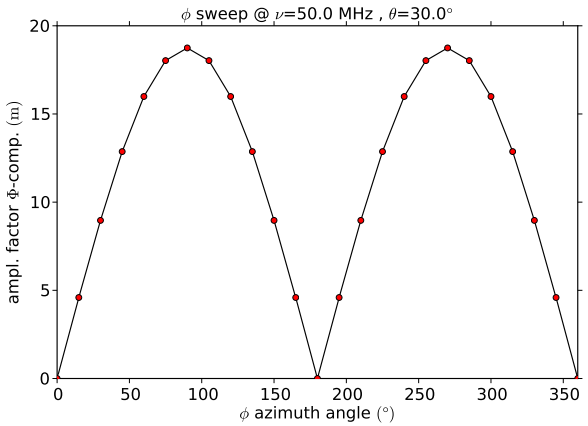
In principle, similar phase jumps can occur in the zenith sweep, but since the antennas are oriented in the azimuth plane, no insensitive direction exists towards the zenith and therefore no phase jumps are present.

The Figure 4.1.1c on page 48 shows a frequency sweep of the amplitude of the effective antenna height, which feature several dips caused by the structure of the antenna. In figure 4.1.2 the

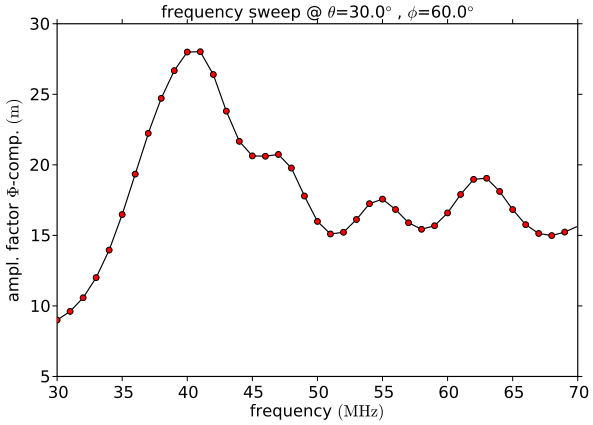
2 The Offline internal name of the response pattern is given in brackets.



(a) Phase correction factor of H_Φ against the azimuth



(b) Amplitude correction factor of H_Φ against the azimuth



(c) Amplitude correction factor of H_Φ over the frequency

Figure 4.1.1: ALU LPDA antenna response pattern, frequency sweep after interpolation. The ROOT interpolation is indicated by a black line, the sample points of the simulated pattern by red dots.

structure of an LPDA antenna [Fliescher et al.] is sketched. The LPDA has a bandwidth of 26.7 MHz to 85.0 MHz with a radiation efficiency of 93.8 MHz. Each dipole has a resonance for one signal wavelength and its multiples. If the frequency of the signal differs significantly from the innate frequency of the receiving dipole, the sensitivity decreases. Therefore several dipoles of increasing length combined together are sensitive to a broader range of frequencies. Nevertheless, the sensitivity of the resulting antenna still decreases in between the dipole frequencies, which results in dips of the frequency sweep.

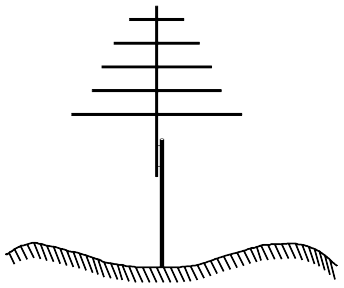


Figure 4.1.2: LPDA schematics

4.1.3 *General remarks on interpolation*

So far only discrete response patterns were discussed, but a successful reconstruction depends on a response function continuous in azimuth, zenith and frequency. Therefore, a linear interpolation of the discrete pattern was provided from the beginning in Offline.

The interpolation depends also on the representation of the complex antenna heights. For complex numbers two forms are available:

rectangular	Euler
$a + \imath \cdot b$	$A \cdot e^{\imath \cdot \phi}$

Simulated response patterns as provided by 4nec2 are usually given in rectangular form, in which an independent interpolation of the amplitude and phases is impossible. A trivial coordinate transformation to the Euler representation resolves this dependency. In the Euler representation, phase ϕ and amplitude A are orthogonal parameters and can be interpolated independently.

However, one obstacle remains. Since the phases are limited to the interval $(0°, 360°]$ in the rectangular form, artificial phase jumps occur after the transformation to Euler representation at the limits of the interval.

Since interpolations assume a continuity, discontinuities can have a negative influence on the interpolation and has to be avoided by unwrapping the phases in Euler form. In Euler form phases are well defined on the interval $(-\infty, +\infty)$. This can be used to prevent the phase jump in the phase series:

385°, 359°, 0°, 1°

which transforms into the continuous phase series

385°, 359°, 360°, 361°

in Euler form. In rectangular form both series cannot be distinguished as they result in the same value. However for interpolations the second form is beneficial and therefore preferred. The continuous phases are called unwrapped phases, since the phase space is unlimited. In Offline, the unwrapped continuous response patterns are provided from the simulations and stored as xml files. The unwrapping from rectangular to Euler form has to be prepared by hand and in case of Offline was done by Stefan Fliescher [Fliescher, 2009 - 2011]. The manual unwrapping process can mistakenly introduce artificial phase jumps as it occurred for the first generation of response patterns in Offline and therefore has to be thoroughly tested to avoid any errors during this step.

4.2 INVESTIGATIONS ON THE QUALITY OF THE RESPONSE PATTERNS

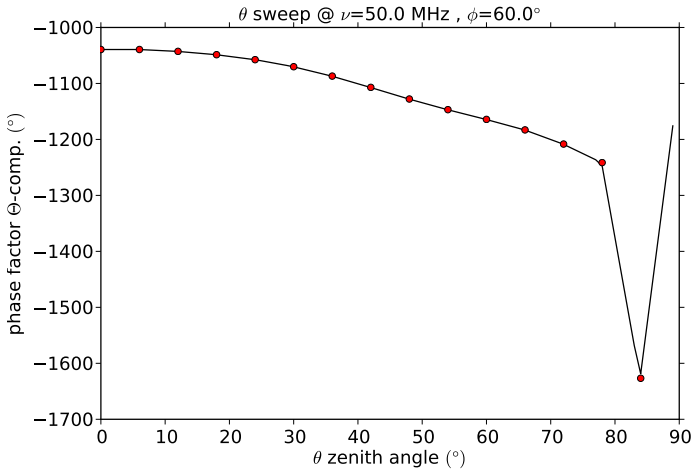
During the investigations of the response patterns in Offline, several defects surfaced, which are exclusively related to the phase of the effective antenna height. One example is presented in figure 4.2.1, where a dip occurs close to the horizon. An examination of the discrete response pattern shows, that one sample point is not aligned (lower figure) with its neighbours. This sample point is 360° phase shifted towards the preceding and following points. In this case, a physical phase jump due to insensitive directions can be excluded, since the sample points before and after are well aligned. More importantly, the Θ component is defined in the zenith plane, which is orthogonal to the insensitive directions of the LPDA antenna in the azimuthal plane (Φ). Therefore, the observed phase jumps are artificial.

This type of discontinuity can be traced back to the unwrapping of the phases provided by the rectangular response patterns simulated with 4nec2.

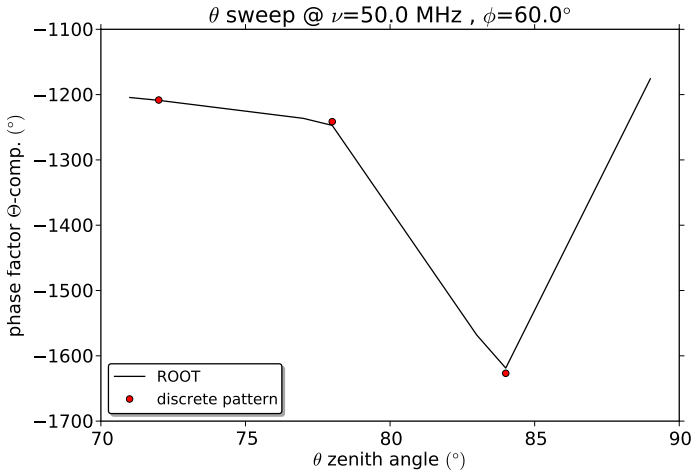
To prevent the non-physical features, the pattern points have to be realigned by applying a phase shift of multiples of 360°.

My investigation revealed similar discontinuities in various response pattern sweeps, both for the phases of the Θ and Φ component, which were as a result resolved by an updated set of response patterns[3] that were rechecked to ensure their quality.

3 The updated response patterns were provided by Stefan Fliescher from RWTH Aachen



(a) Phase of H_Θ after interpolation



(b) Detail for zenith: $70° < \theta < 90°$

Figure 4.2.1: Black-Spider antenna at BLS response pattern, zenith sweep after interpolation. The ROOT interpolation is indicated by a black line, the sample points of the simulated pattern by red dots.

It has to be noted, that no pattern defects were found relating to the amplitude factors.

4.2.1 *Interpolation of response patterns*

Commonly, response simulations like 4nec2 are used to determine the patterns H_Θ and H_Φ of the two components of the coordinate system. These simulations have the draw-back that they are time intensive, so an interpolation of the discrete patterns is applied to provide a continuous response function during analysis.

As response patterns are three-dimensional matrices, the interpolation has to be done in three dimensions. As discussed, amplitude and phase of the complex effective antenna height are orthogonal in Euler representation and can be interpolated separately.

In $\overline{\text{Off}}\underline{\text{line}}$, the interpolation of the response patterns consists of two steps:

- Read in of two response patterns from file, one matrix for each the θ- and φ-component (simulated by 4nec2)
- Three-dimensional interpolation of the amplitude and phase for both the θ- and φ-matrix

4.2.1.1 *Interpolation methods and technical requirements*

The standard $\overline{\text{Off}}\underline{\text{line}}$ interpolation method is a linear interpolation provided by the CERN analysis package (ROOT). ROOT is an analysis package for physic data analysis and visualisation. Among others it provides plotting, fitting, statistical functionality and the here used interpolation routines.

An alternative spline based interpolation method is compared to this standard interpolation. The spline-based interpolation is provided by the software library *Einspline [Esler, 2012]*. Einspline provides efficient routines for spline interpolations of equidistant input sample points in one, two and three dimensions. The library has low computation time demands thanks to highly optimised routines. More details on spline interpolation algorithms can be found in Press et al. [2007] and on spline theory in Deboor [1978]. Spline interpolations also provide a differential interpolation without kinks and discontinuities, which is not the case for linear interpolations. Even as differentiability is not mandatory for response patterns, they are nevertheless expected to be continuous and differentiable and therefore better described by splines than linear interpolations.

Both interpolation algorithms have to take the discrete response patterns simulated by 4nec2 in Euler representation as input. The

interpolation has to provide a continuous response function over the full sky and frequency band.

4.2.1.2 *Technical requirements for the ROOT interpolation*

There are no special technical requirements on the discrete input response patterns for ROOT. But the ROOT interpolation has one limitation as it does not support a full three-dimensional interpolation. This requires multiple two-dimensional interpolations in the ν-ϕ-, ν-θ- and ϕ-θ-planes, which are afterwards combined to a three-dimensional interpolation.

4.2.1.3 *Technical requirements for the Einspline interpolation*

The *Einspline* based interpolation has additional requirements as equidistantly spaced input points are mandatory. The reason is an underlying optimisation of the implemented algorithms. Without equidistant sample points in ν, ϕ and θ, the computation speed of the interpolation decreases significantly.

4.2.2 *Observed artefacts of the interpolation algorithms*

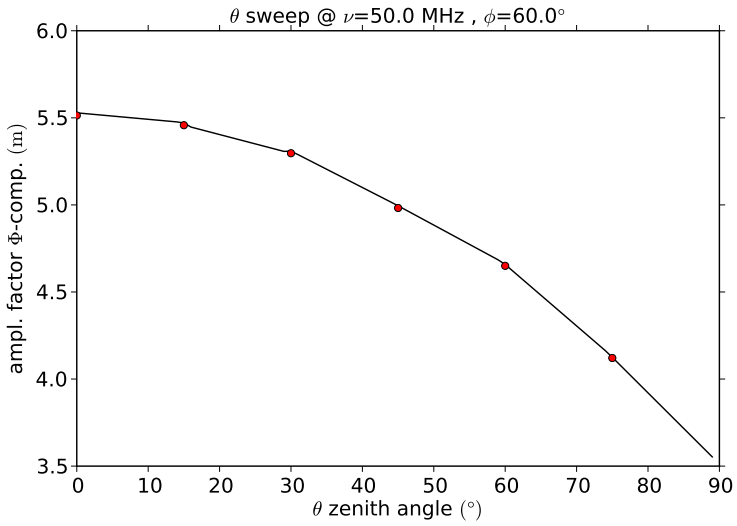
The investigations revealed defects, unrelated to the discrete response pattern, which are directly related to the interpolation method used. These defects cannot be avoided, only reduced in some cases.

4.2.2.1 *ROOT, linear interpolation*

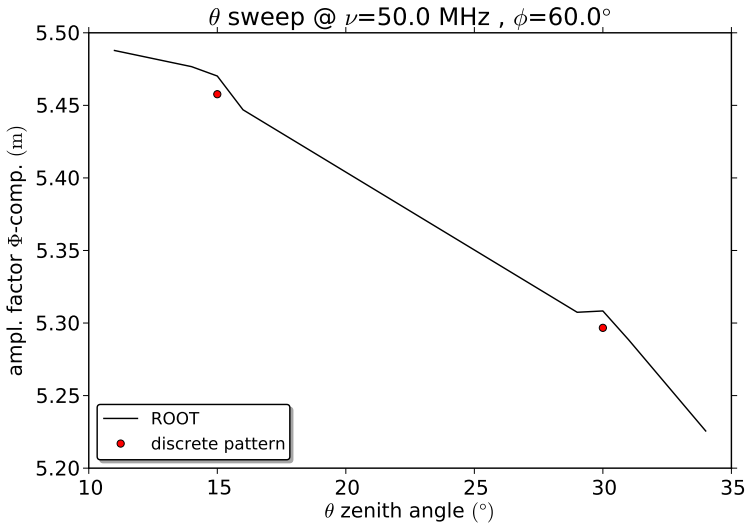
During examination of ROOT interpolated response patterns several artificial features became obvious. One general limitation is the linear nature of the ROOT interpolation, which gives cause to kinks at sample points. These kinks have no physical motivation, since even around insensitive directions the response is still expected to be continuous. This is especially true for interpolations of amplitude antenna factors. At around 84° in figure 4.2.1, immediately after the last sample point, the interpolation shows a kink and afterwards non-physical high phase values. In comparison, the region below 84° is continuous as expected.

Another artefact is present at the last sample point. There the interpolated values do not include the sampling point at all. In technical terms, the interpolation is not an interpolation but an approximation. This is solely related to the interpolation and can result in non-physical results.

Another example is given in Figure 4.2.2 on page 54. After the interpolation, a significant difference between the discrete response pattern (red dots) and the interpolation provided by ROOT is present. Especially small kinks approximately at the



(a) Amplitude of H_Φ



(b) Detail for zenith angle $10° < \theta < 35°$

Figure 4.2.2: SALLA antenna response pattern, zenith sweep after interpolation
black line: ROOT interpolation. The ROOT interpolation is indicated by a black line, the sample points of the simulated pattern by red dots.

position of the expected discrete pattern points are noticeable (see zoom in figure below). Furthermore the difference between the interpolation and the input values decreases with decreasing amplitude factors (see again figure 4.2.2), which excludes an absolute and constant effect. Also, no misaligned sampling points can be identified as the source for the failed interpolation. It has therefore to be attributed solely to the interpolation routine. These kind of failed amplitude interpolations are observed for most of the patterns interpolated with ROOT, which makes the artefact independent of the specific antenna or response pattern.

The observed differences towards the sample points introduce a systematic uncertainty on the reconstructed $\vec{E}$-field (especially for the amplitude factors), which is in general avoidable.

4.2.2.2 *Einspline, spline interpolation*

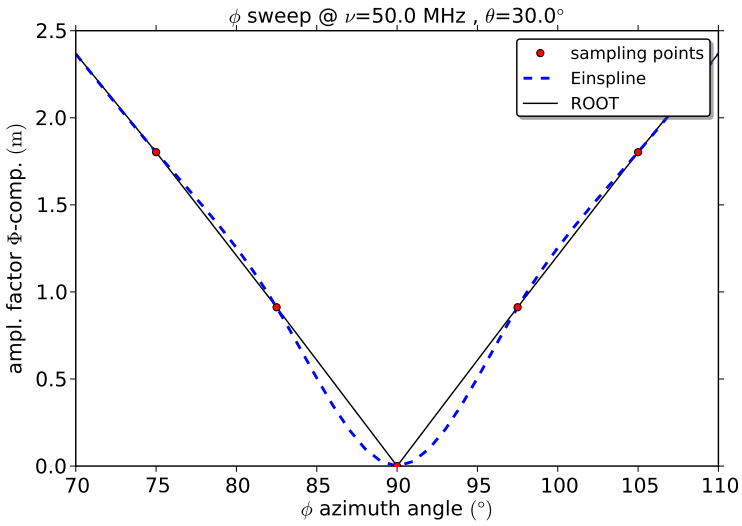
For the Einspline based interpolation, no significant differences between discrete input response patterns and interpolations are observed. Instead other prominent artefacts close to the before discussed physical phase-jumps at insensitive directions occur, which are related to the very nature of splines.

In figure 4.2.3 a detail of the phase factor for the Φ-component is shown. Both the ROOT and the *Einspline* interpolations are included. Two oscillations of the Einspline interpolation are obvious around the phase jump. These oscillations occur, since splines assume an underlying continuous connection and are thereby not suited for interpolations of discontinuities.

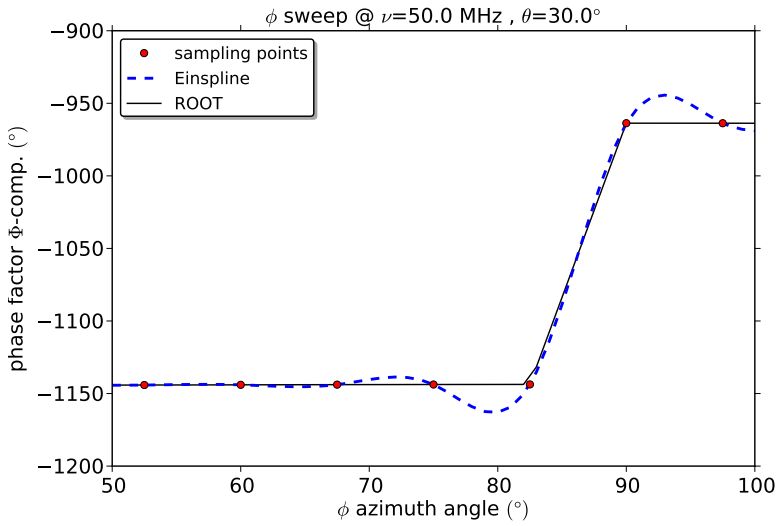
An analogue thought experiment enlightens the situation. A spline can be compared to a "spring metal bar" bend around fixed "poles". The "poles" represent the discrete pattern points in this picture. As long as the "poles" show some continuity, without discontinuities on small scales, the "bar" will be able to form a continuous curved line encompassing the poles without breaking. If a discontinuity is introduced, like in figure 4.2.3, the "bar" has not enough freedom to compensate the "stress" with a mild curvature. The "bar" has then two options. Either it breaks, which translates into a kink in the spline interpolation, or it bends around the poles accompanied by strong oscillations to compensate for the "stress" induced by the discontinuity.

In case of splines the underlying spline-equations and their boundary conditions resemble the limited ability of the "bar" to bend without breaking. These equations also enforce differentiability, which excludes a kink in the interpolation. The only remaining freedom for splines is therefore an oscillation around the discrete pattern points in the vicinity of the discontinuity.

The oscillations of spline interpolations are observed to spread over three to four discrete pattern points. Afterwards spline and



(a) Amplitude of H_Φ



(b) Phase of H_Φ

Figure 4.2.3: Small Black-Spider antenna response pattern, azimuth sweep after interpolation. The displayed are the Einspline interpolation (blue dashed line), the ROOT interpolation (black line) and the sampling points (red dots)

linear interpolation differ insignificantly. Also, no immediate dependence on the sample point spacing was observed.

Since the oscillations are non-physical, it is important to ensure that they show no significant influence There are two possible approach to address this effect.

First, the number of discrete pattern points around the discontinuity can be decreased. This reduces the amplitude of the oscillations. The difference between the Einspline interpolation and the linear ROOT interpolation decreases as a result but on the expense of a larger affected region of the pattern.

The second approach is the opposing strategy to the first one. By introducing additional discrete pattern points and exploiting that the oscillations depend strongly on the sample points and not on the distance to the discontinuity, the oscillations can be confined to the direct vicinity of the discontinuity. Since the amplitude factors are small for physical discontinuities at insensitive directions, the effect of the oscillations is naturally suppressed in the immediate vicinity. Therefore, the second approach is here preferred over the first one.

It has to be noted, that this is only true for the common use of response patterns during the reconstruction of measured radio signals. If the response pattern is applied during a detector simulation, as it is the case during the processing of radio simulations within Offline, the inverted response pattern is applied. Therefore small amplitude factors translate into large ones and the influence of the oscillations increases likewise. Nevertheless the second approach is favoured, as in principle the problem is also present for the first approach.

Another effect occurs for interpolations of amplitude factors. At insensitive directions, the amplitude factors approach zero. This results for a linear interpolation in a discontinuous kink. The spline interpolation on the other hand provides a continuous approach in the vicinity of the kink. The result is a smooth approach of the insensitive direction as presented in figure 4.2.3. An inevitable difference between the linear and the spline based interpolations occurs for events with arrival directions close to the insensitive direction.

4.2.3 *Requirements for response pattern interpolations based on Einspline*

The Einspline interpolation requires on the technical side an equidistant discrete input pattern. To address the potentially problematic oscillations, a small distance between the discrete pattern points is advised. The binning is especially important for the azimuth plane, since there physical discontinuities due to insensitive directions are common.

For an Einspline interpolation the following spacing requirements are suggested:

$$\begin{aligned} \text{frequency} &: 1\,\text{MHz} \\ \text{azimuth} &: 7.5° \\ \text{zenith} &: 3° \end{aligned}$$

In general, there is no indication, that further reduction of the spacing will result in any ill effects. However, there is an increase on the initialisation time for both interpolation methods expected and of course on the amount of required disk space. Since Einspline requires equidistant discrete input patterns, a reduction of the spacing to counter the described oscillations, ultimately results in an increase of the total amount of pattern points to be stored and read.

4.3 COMPARISON OF *Einspline* AND ROOT INTERPOLATION IN LIGHT OF DATA ANALYSIS

4.3.1 *Data sets and parameters*

To compare the influence of *Einspline* to the results of the standard radio reconstruction provided by $\overline{\text{Off}}\underline{\text{line}}$, a set of BLS measured events is used as they are described in Chapter 5. This limits the analysis to the response pattern of the SmallBlackSpider_atBLS_ground2 model.

The events for the comparison were reconstructed by $\overline{\text{Off}}\underline{\text{line}}$ using ROOT and *Einspline* interpolation. The parameters for this analysis with $\overline{\text{Off}}\underline{\text{line}}$ are the same as described by Fuchs et al. [2011] and in Chapter 5. To suppress noise during the reconstruction an SNR cut of $S^2/N^2 > 36$ was applied[4]. This resulted in 28 events with a total of 84 stations and is sufficient for a direct comparison of reconstructed quantities:

- signal time
- incident direction reconstruction
- magnitude of the $\vec{E}$-field vector

For the exact reconstruction algorithms please refer to the work of S. Fliescher et al. - Pierre Auger Collaboration [2010b]. The interpolations of the discrete response patterns is done with ROOT and with *Einspline* (both interpolate the updated patterns, which showed so far no pattern defects). The Einspline interpolation is also done within $\overline{\text{Off}}\underline{\text{line}}$ by an adaptation of the corresponding

4 This cut requires, that the signal power is by a factor of 36 larger than the noise power.

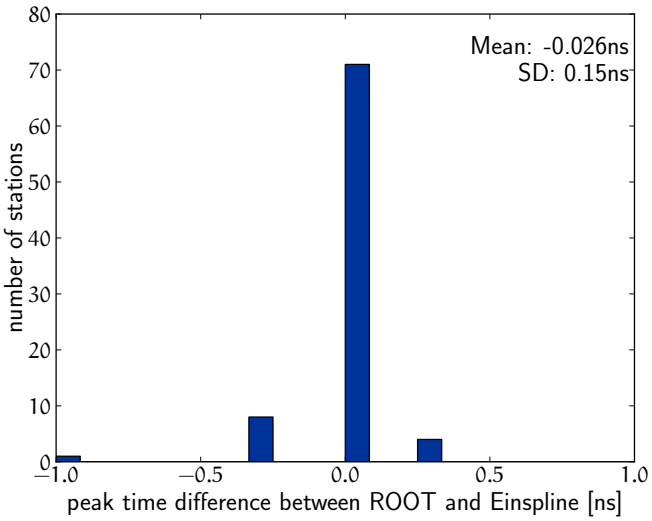


Figure 4.3.1: Small Black Spider at BLS antenna, difference in the reconstructed signal time between the ROOT and the Einspline reconstruction based on the updated pattern.

framework part (AntennaType class) that I introduced to provide an interface to the Einspline interpolation.

The result of the reconstruction are two data sets:

1. updated unwrapped pattern corrected for discontinuities and obeying the mentioned quality criteria (second generation). Reconstructed with ROOT

2. updated unwrapped pattern corrected for discontinuities and obeying the mentioned quality criteria (second generation). Reconstructed with Einspline

The reconstruction of the simulated data set results for both reconstructions in the same 28 events. The interpolation method therefore shows no influence on the number of reconstructed events. This is in general expected.

4.3.2 *Signal time comparison*

The signal time is given as the position of the signal magnitude in the $\vec{E}$-field timeseries with nanosecond precision.

Figure 4.3.1 on page 59 presents a histogram of the difference between the ROOT interpolation and the *Einspline* interpolation both based on the second generation response pattern. The distribution of the differences in the signal time for 84 stations has a mean of -0.026 ns and a standard deviation is 0.15 ns. No major influence on the reconstructed arrival times was found.

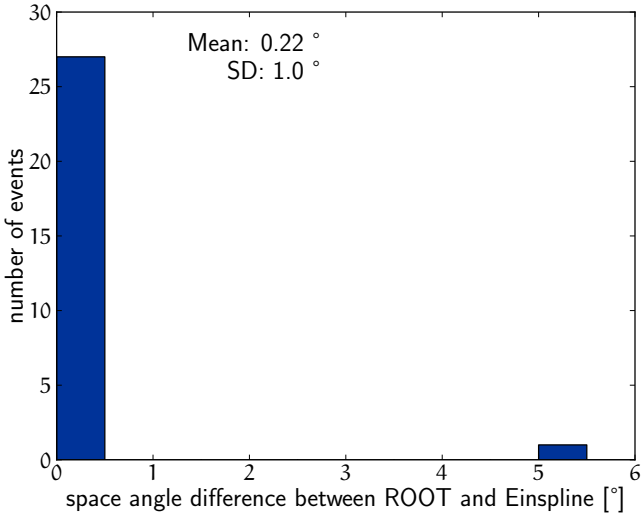


Figure 4.3.2: Small Black Spider at BLS antenna, difference in the reconstructed incident direction between the Einspline reconstruction and the ROOT reconstruction, both based on the updated pattern.

4.3.3 *Incident direction comparison*

The incident direction is calculated from the reconstructed azimuth and zenith angle. For the comparison, the space angle difference (space angle difference (SAD)) between different reconstructed incident directions is calculated. Since the incident direction is a communal quantity of all stations participating in an event, 28 space angle differences are calculated.

Figure 4.3.2 on page 60 presents a histogram of the difference between a ROOT interpolation and an *Einspline* interpolation with the updated pattern. The difference of the reconstructed space angle difference is given for all 28 events. Here, the influence of the response pattern can reach 5° with a mean of 0.22° and a standard deviation of 1°. In general the majority of the events is unaffected by the interpolation algorithm. The variation is in the same order of the typical direction reconstruction uncertainty for radio experiments of ~1° (e.g. as given for the LOPES experiment in [F. G. Schröder et al. - LOPES Collaboration, 2011]).

Event 3913273 is an interesting exception as it shows a difference of around 5° between the ROOT and *Einspline* reconstruction. The ROOT reconstructed incident direction of $\phi = 136.84°$ and $\theta = 28.8°$ does not match to the SD reconstructed incident direction of $\phi = 147.48°$ and $\theta = 29.98°$ as well as the *Einspline* reconstructed incident direction $\phi = 147.84°$ and $\theta = 29.56°$. The *Einspline* interpolation agrees significantly better with the SD reconstruction for this event.

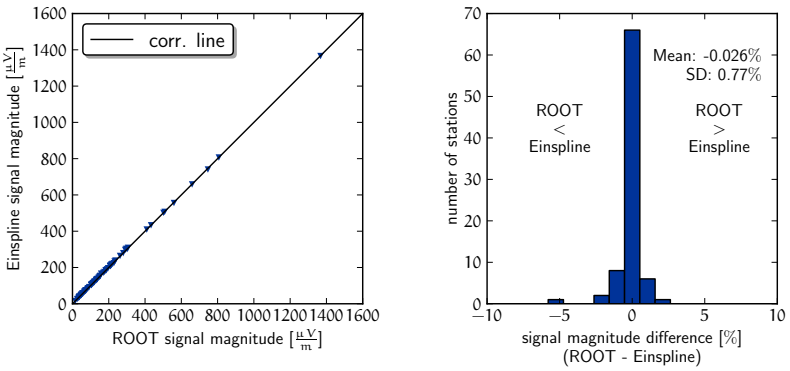


Figure 4.3.3: Small Black Spider at BLS antenna, difference in the reconstructed signal strength between the Einspline reconstruction and the ROOT reconstruction based on the updated pattern.

4.3.4 *Comparison of the magnitude of the $\vec{E}$-field*

For the comparison, the signal strength for single stations for different reconstructions is calculated and compared. This resulted in 84 reconstructed $\vec{E}$-fields for all stations.

Figure 4.3.3 on page 61 presents a correlation plot and a histogram of the difference between the Einspline interpolation and the ROOT interpolation. On the left hand side, the reconstructed magnitude of the $\vec{E}$-field based on the Einspline interpolation is plotted versus the ROOT based interpolation. The diagonal line indicates a total match of both magnitudes. On the right hand side, a histogram of the relative difference between the two reconstructions is plotted. The mean of -0.026% ($-0.015\ \frac{\mu V}{m}$) and standard deviation of 0.77% ($1.8\ \frac{\mu V}{m}$) shows no significant difference between ROOT and *Einspline*.

Summarising, in the first generation of response patterns defects are observed related to the manual unwrapping of the phase of the effective antenna height. These defects are now resolved by correcting these patterns. In addition, several ambiguities like kinks or miss-interpolations revealed for the applied interpolation method on basis of ROOT. To reduce this influence, a substitution on basis of the *Einspline* interpolation library is implemented. The subsequent comparisons of the *Einspline* and the ROOT based interpolation however shows no significant impact on the reconstruction for a representative data set measured by the BLS. The investigated quantities are: the magnitude of the $\vec{E}$-field, the incident direction and the signal arrival time. Therefore a significant influence on the reconstruction of measured data cannot be established.

Since the incident directions of measured data sets are usually biased towards zenith angles smaller than 60° as a result of the detector sensitivity. All differences between interpolations on the other hand are observed for zenith angles above 60°. Therefore the only one event of the data set experiencing a significant difference between the ROOT and *Einspline* interpolation has to be especially emphasised. The difference is significant for the reconstructed incident direction, as event 3913273 shows a deviation of 5° in space angle between the ROOT and the *Einspline* reconstructed incident direction. This difference cannot be linked to a feature, i.e. insensitive direction of the response pattern. Event 3913273 is therefore most likely sensitive to influences from the interpolation method. The event can also be compared to the SD reconstruction of the air shower that emitted the radio signal. This SD reconstruction is consistent with the *Einspline* based reconstruction and inconsistent with the ROOT based reconstruction. This indicates, that the *Einspline* based reconstruction can potentially improve the reconstruction for certain incident directions.

Unfortunately, several unsolved technical incompatibilities between the *Einspline* library and the Offline framework prevent a regular employment of Einspline for the time being. Therefore, all subsequently discussed analyses are still based on the ROOT interpolation. This does however not concern the response patterns, which are nevertheless upgraded and available in Offline. Consequently, the improved response patterns are used for the subsequent analyses.

5

COMPARISON OF RADIO MEASUREMENTS WITH RADIO SIMULATIONS

On basis of the three-dimensional $\vec{E}$-field, a comparison of simulations and measurements is now possible. This is important to asses how well radio simulations actually describe real measurements. Comparisons also further the understanding of the measurement process itself and on the relative contributions of different emission mechanisms and other physics effects.

In the case of Offline, it is not only possible to compare the three-dimensional $\vec{E}$-field, but to include furthermore a full detector simulation. Therefore the presented comparisons includes all the effects and influences from the detector itself and from reconstruction algorithms. The first time, this kind of study was done on a BLS data set of 494 measured events in preparation for the upcoming analyses of AERA data. At this time AERA was still in deployment and no data available. Only recently, it was possible to repeat the analysis also for gathered AERA data and with improved radio simulations from up-to-date radio simulation codes. Therefore the comparison also provides viable information on how the radio simulations evolved.

5.1 COMPARISON OF BLS DATA WITH RADIO SIMULATIONS

As a preparation for data analysis of the AERA setup the combined BLS data set of the years 2007/2008 was analysed for this study. The total number of measured BLS events coincident with SD events is 494. 313 of these were measured in 2007 and 181 were measured in 2008. The data was collected with three stations in the field as described by **?**.

For each BLS event an SD event is triggered within the same time and therefore an SD reconstruction of the air shower is available. However, this does not imply that all radio measured events can actually be reconstructed, since on the one hand a reconstruction requires a signal in all three radio stations and on the other hand the ambient radio noise level during the measurement can interfere. Therefore the number of reconstructed radio events differs significantly from the number of measured radio events.

The SD reconstructed air shower quantities (e.g. core position) of the associated events, are used as input parameters of the radio simulations. This bases the radio simulations on the independent SD reconstruction. Accordingly, there is no bias due to

the input parameters of the radio simulations for comparisons between radio measurements and simulations. The radio simulations are prepared on a strict event per event basis for each of the BLS measurements. The simulations are done with two independent radio simulation codes Radio simulation of Extended Air Showers (REAS3) [Ludwig and Huege, 2010] and Macroscopic model of Geo-Magnetic Radio emission (MGMR) [Scholten et al., 2008].

In addition to the event per event simulations, a second set was prepared. These varied simulations were designed to cover the uncertainty of the SD reconstructed quantities. This enabled an investigation of the impact of the SD uncertainties on the reconstructed radio simulations. To achieve this goal, the input parameters of the radio simulations were varied within their SD reconstruction uncertainties. The result is a set of simulations for each measured BLS event with varied initial parameters. For a large enough statistical sample of varied simulations per measured event, an estimation of the uncertainty on the quantities reconstructed from the radio simulations can be derived[1].

Altogether this results in three data sets of simulations and measurements for the amplitude comparison:

1. measured BLS data set of the years 2007/2008
2. simulated radio simulations, one for each measured event
3. varied simulated events for each measured event for the purpose of uncertainty estimation

The amplitude comparison is based on traces with a length of 10 µs. All the traces contain radio signals from air showers, since they were measured in coincidence with SD. The radio signal of an event is determined by a signal search algorithm as described in 2.2.1.1. The signal search algorithm yields the largest amplitude of the $\vec{E}$-field vector within a detector specific signal window. The signal window used during this analysis, is the standard for the BLS setup as provided by Offline and covers the region between 1600 ns and 2325 ns.

The noise level of a measured event is determined by the root mean square (RMS) of the $\vec{E}$-*field* vector amplitudes within a selected part of the trace. This part of the trace is called the noise window, which is chosen accordingly to the BLS standard provided by Offline. The window covers the region between 0 ns and 1140 ns. The so calculated noise level is especially important for the analysis, since it provides an approach to estimate the uncertainty of the measured radio signal as discussed later.

1 The varied simulations were kindly provided by E. D. Fraenkle from Groningen University, Netherlands. The method is successfully applied throughout the collaboration for both radio simulations of the BLS and AERA setups.

At this point a recent development within the analysis routines has to be mentioned, which especially influences the to be presented BLS analysis. At short notice, early in August 2012, a mistake in the source code of the radio analysis algorithms was discovered. The problem was introduced during the first implementations of radio reconstruction algorithms to Offline in early 2007 and especially concerns the analysis of radio simulations. An influence on the reconstruction of radio measurements is not yet established but cannot be ruled out at the time.

Misreconstructions can occur for radio simulations with an azimuth incident direction between 270° and 360°. For this azimuth range, the $\vec{E}$-field vector is subtly changed as a consequence of an interface problem. Therefore the following BLS analysis has to be updated in the near future, as this is not possible for this work due to short notice.

Fortunately it was possible, to at least apply an updated Offline version for the AERA analysis presented in Section 5.2 and following discussions. The drawn conclusions of the BLS analysis is therefore limited but still valid, since the AERA analysis in general supports the BLS results.

5.1.1 *The measured data set and applied selection cuts*

During reconstruction, the iterative direction reconstruction algorithm as discussed in chapter 3 is applied to provide both the incident direction and the vectorial $\vec{E}$-field.

After the reconstruction, 60 events remain before any selection or SNR cuts. This dramatic reduction in the number of events is a result of most events showing reconstructed signals only in one or two stations. For a successful direction reconstruction all 3 stations need to show significant radio signals. As a result the majority of events cannot be reconstructed by the direction reconstruction algorithm. As this principle limitation to a minimum of three stations cannot be avoided, an improvement of the direction reconstruction is not expected for soon.

In addition the events split into two groups, due to two changes in the hardware setup. The first group was measured with a bandwidth of $40 - 70\,\mathrm{MHz}$ (referenced as DA), while the second group was measured with a bandwidth of $50 - 70\,\mathrm{MHz}$ (referenced as DB). Both groups are separately processed considering their appropriate filter setup. The $40 - 70\,\mathrm{MHz}$ events form a data set of 52 events, while the $50 - 70\,\mathrm{MHz}$ events include only eight due to the short time period this configuration was in use.

detector bandwidth	40-70 MHz	50-70 MHz
selection	D_A	D_B
reconstructed events before cuts	52	8
events remaining after SNR cut	36	8
events remaining after SNR and SAD cut	32	8
events remaining for analysis	**32**	**8**

Table 5.1.1: BLS data sets for the two bandwidths 40 MHz to 70 MHz and 50 MHz to 70 MHz and their evolution with applied cuts. Of the remaining 40 events, 3 have been identified as thunderstorm events.

To prevent a reconstruction of ambient transient noise, it is furthermore necessary to exclude events dominated by noise from the analysis and to ensure that only events with a clean radio signals are used during analysis. To identify and select such events, an SNR cut, like $(\text{Signal}/\text{Noise})^2$, has proven to be an appropriate method. An SNR cut rejects events with an SNR below the cut. This reduces the probability of a background fluctuation and the probability to accidentally reconstruct a transient noise signal instead of a cosmic ray air shower radio signal. Such a misreconstruction corresponds to a type I error known in the field of statistics[2].

Aside the SNR cut, also a cut on the angle between the direction reconstructed by radio and SD can be applied. If the SAD between the SD and RD direction exceeds the cut threshold, the event is rejected. This reduces the amount of accidental reconstructed noise, since the position of the noise source unlikely coincides with the air shower. Therefore the SAD cut ensures, that both the RD and SD measured signals are emitted by the same source/air shower. A very small cut threshold on the SAD, however, is ill advised, since the resolution of the RD direction reconstruction is unknown. A too strict cut would lead to an increase in the type II error probability, since well reconstructed radio signals from air showers would be rejected by mistake.

For this analysis the SAD and SNR cuts are derived from a two-dimensional histogram as shown in figure 5.1.1. This method has the benefit, that it also covers cross-correlations between the SAD and SNR cut. The histogram shows the chosen cuts (black dot) on basis of the 40-70 MHz data set, since the 50-70 MHz data set provides a too small statistical basis for this kind of analysis.

2 Type I error: accepting by mistake a noise event as a signal.
Type II error: rejecting by mistake a signal as a noise event.

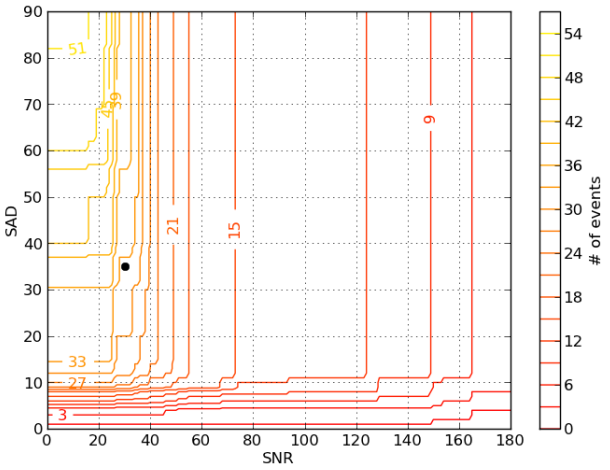


Figure 5.1.1: Influence of the SAD cut (SD-RD incident direction) and the SNR cut on the number of reconstructed events. The chosen cuts for the analysis are marked with a black dot. (selection DA, only 40-70 MHz events included)

In a plateau region, as found near the values marked with the black dot, the number of selected events is approximately independent from small variations of the SNR and SAD cut thresholds. The existence of such a plateau region indicates two populations of events. The first population has high SAD and low SNR (upper left region of the histogram), which is expected for the noise events we want to reject.

The second population consists of events with high SNR and low SAD (lower right region of the histogram). These are the event characteristics ideally expected for real cosmic ray events in absence of noise and reconstruction uncertainties.

To separate these two populations, the cuts are chosen such, that small variation of the cuts has no big influence on the number of surviving events. This is the case for the plateau region, the final cuts are chosen from. Based on these cuts, both populations can be reliably separated.

To minimise the probability of a type II and a type I error, the cuts are furthermore chosen to show an as high SAD and simultaneously as low SNR as possible within the limits of the plateau (black dot):

$$\begin{aligned} \frac{\text{Signal}^2}{\text{Noise}^2} &> 30 \\ \text{SAD}\,(\text{SD}-\text{RD}) &< 36^\circ \end{aligned}$$

The SNR is calculated on the magnitude of the $\vec{E}$-field. After both cuts are applied to the measured data set, 32+8 events remain for the analysis.

This data set still includes events measured during thunderstorms. It is well known that thunderstorms can interfere with the radio emission of air showers as described by W.D. Apel et al. - LOPES Collaboration [2011], Melissas et al. [2010] and Buitink et al. [2007]. Since, however, the exact influence of the thunderstorms on the radio signal is still under investigation, a correct treatment of the effect is currently not possible. For this reason three events measured during thunderstorms are excluded from the analysis. The thunderstorm events are identified by independent measurements of the atmospheric electric field with a so called $\vec{E}$-field mill.

Finally the uncertainty on the measured and reconstructed amplitude for the BLS events is estimated based on the noise level of the trace. The method is well motivated, since both the noise and the signal are radio waves, which can interfere constructively or destructively. Due to this interference of noise and signal waves, the noise can be treated as an estimate for the uncertainty on the signal strength. Systematic deviations are nevertheless expected for small SNR values, which was investigated among others by Schröder et al. [2010]). This again supports an as high as possible SNR cut for the analysis, since then systematic deviations are not significant for the analysis.

5.1.2 *The single and varied radio simulation data sets*

For each of the 32+8 measured and reconstructed radio events (DA & DB including thunderstorms), the radio simulations are available both for REAS3 (based on the URQMD+QGSJETII models) and MGMR.

detector bandwidth	40-70 MHz		50-70 MHz	
selection	MA	RA	MB	RB
MGMR / REAS3 simulated events	**32**	**32**	**8**	**8**

Table 5.1.2: simulated data sets, based on the BLS data selections Da and Db. Mx indicates MGMR simulations, Rx REAS3 simulations.

The input parameters for the radio simulations were taken from the central data acquisition system (CDAS) reconstruction of the SD data of the air shower [X. Bertou et al. - Pierre Auger Collaboration, 2010]. The SD data reconstruction is the standard Radio-Herald reconstruction as used and described by **?**. No additional quality cuts aside of the described ones have been applied.

The available REAS3 and MGMR simulations for this study do not cover second order effects like the influence of the refractive index of the atmosphere. These were only recently introduced into both simulation codes and resulted in changes of the simulated $\vec{E}$-field strength. This introduces additional uncertainties on the here used simulations, that are not yet estimated accurately. Therefore a difference of a factor 2 in amplitudes between the simulations and the measurements can be expected, as it was also observed by LOPES [Ludwig, 2012].

The reconstruction of the simulated data with Offline differs from the one applied to measured data. In a first step, the detector response is applied to the simulations to perform a complete detector simulation. The resulting two traces per station on the channel level include all known characteristics of the detector. This corresponds to the state after reading into Offline the raw measured data. From there on the simulated events are reconstructed as described for measured data but without applying the SNR and SAD cuts.

Since the SD quantities are used as initial parameters of the radio simulations, the uncertainties on the SD (CDAS) reconstructed quantities have some influence on the simulations output and reconstructed quantities. To cover this influence, a second set of varied radio simulations, as mentioned above, was prepared. The propagation of the SD uncertainty through the radio simulation requires a variation of the initial parameters of the simulations within its SD uncertainty. The primary shower parameters reconstructed by CDAS are:

- T_0: the arrival time
- U, V: the projection of the (normalised) shower axis on the x, y-Plane
- X_{Core}, Y_{Core}: the position of the shower core
- S_{1000}: the signal in vertical equivalent muon (VEM) at 1000 meters from the core
- X_{max}: the maximum of the shower (vertical)
- R: the curvature radius of the shower front

Only the estimated primary particle energy and the event geometry (core position and arrival direction) are input parameters to the radio simulations. Additional quantities listed here are included for the sake of completeness, but have no primary impact on the analysis.

Together these parameters can be expressed as a vector $\vec{X}$

$$\vec{X} = (T_0, U, V, X_{Core}, Y_{Core}, S_{1000}, X_{Core}, R)^T$$

detector bandwidth	40-70 MHz		50-70 MHz	
selection	MVA	RVA	MVB	RVB
number of simulated showers per event	100	25	100	25
MGMR / REAS3 based uncertainty estimation events	**23**	**23**	**7**	**7**

Table 5.1.3: varied simulated data sets, based on selections Da and Db. RVx indicates REAS3 varied simulations, MVx MGMR varied simulations. The number of selected events reduces to 23+7, since only a subset of varied simulations were prepared. The selections exclude the three thunderstom events.

accompanied by a 8×8 covariance matrix. Based on the covariance matrix and the assumption of normal distributed uncertainties, the dependency between the uncertainties of SD quantities can be treated correctly to estimate the resulting uncertainties on the radio simulation input parameters[3].

This method to cover correlations between SD quantities and uncertainties and their influence on the radio simulations was used for some time within the AERA group and described in more detail by Fuchs et al. [2011]. Furthermore the varied simulations cover also the effect of shower-to-shower fluctuations (see Chapter 2) and the thereof resulting uncertainties.

For each radio event a set of varied simulations was prepared in this manner and each of them reconstructed by Offline.

The result is a set of reconstructed quantities for each radio event (e.g. the magnitude of the $\vec{E}$-field, incident direction, etc.). The uncertainty on one reconstructed quantity can be estimated by calculating the RMS of each set of values for a reconstructed quantity. In physics it is common to favour the RMS estimator for the task but since in this case the data set is limited in size, the median absolute deviation (MAD) estimator is preferred. The number of available varied radio simulations per event is limited due to the large computation time of the radio simulation codes. Especially for simulations with cosmic ray simulation for KASCADE (CORSIKA)+REAS3 this is a major obstacle. This hinders the preparation of large sets of varied radio simulations for a measured data set in feasible times.

Aside of the uncertainties resulting from initial parameters of the radio simulations also the noise is a governing influence on the simulations. Unfortunately the Offline framework currently lacks a reasonable radio noise simulation. Accordingly both the single radio simulations and the varied radio simulations do not

3 The method was developed by Daniël Fraenkle from Groningen University, Netherlands.

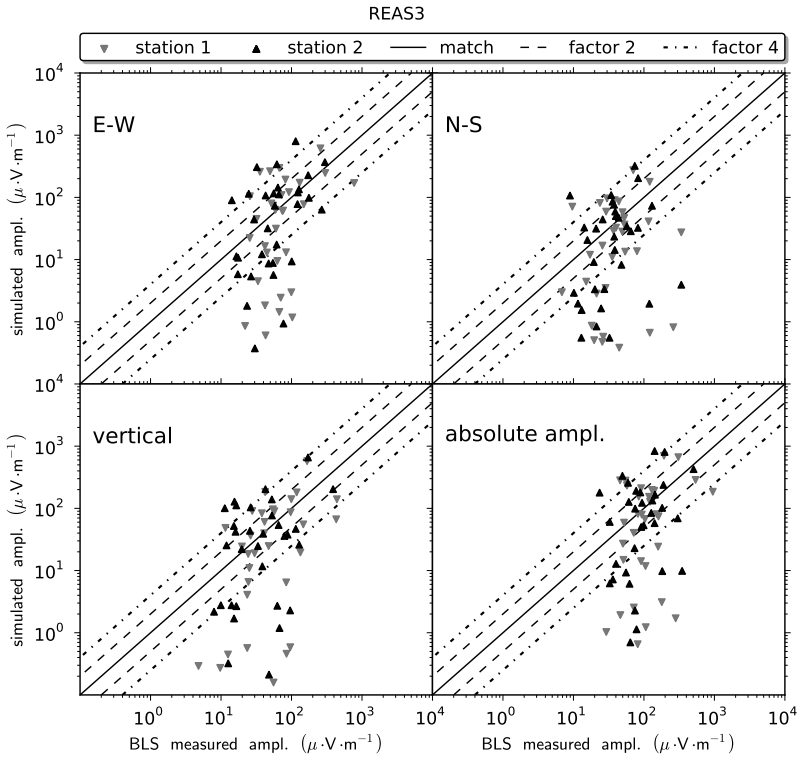


Figure 5.1.2: REAS3 simulated and measured amplitude for the 3 polarization components and the magnitude of the $\vec{E}$-field [selection Ra]. The data were taken with a bandwidth of 40-70 MHz between 2007 and 2008. For events on the solid line, measured and simulated amplitude match. The dashed and dotted lines indicate a factor of 2 and 4 difference in the amplitudes. The east-west, north-south and vertical amplitude are reconstructed with the iterative reconstruction algorithm for the $\vec{E}$-field discussed in Chapter 3.

incorporate a noise simulation during the analysis. However, the noise influence is only a governing factor for an analysis of small signals close to the noise level. Thus the influence of the noise level is small on the analysis in mind, since it is focused on measured events with a high SNR.

5.1.3 *Comparison of measured and simulated amplitudes*

As described above, the full 3D vectorial $\vec{E}$-field was reconstructed both from simulations (including a full detector simulation) and measured data. The vectorial $\vec{E}$-field consists of three components, the EW-, NS- and vertical one. Together with the magnitude of the $\vec{E}$-field, they form a set of 4 reconstructed $\vec{E}$-field quantities. In Figure 5.1.2 on page 71, an individual comparison

category	difference $\frac{\text{measurement}}{\text{simulation}}$
I	2 stations between a factor 3 and 4
II	2 stations greater than a factor 4
III	only 1 station greater than a factor 4

Table 5.1.4: categories for events based on their deviation in measured and simulated amplitude

between the measured and the REAS3 simulated $\vec{E}$-field is shown for each of the vectorial components and the magnitude of the $\vec{E}$-field. The focus of this analysis is primarily on the magnitude of the $\vec{E}$-field and not in the individual components of the signal, which are not considered in the following.

Figure 5.1.3 on page 75 details the ratio of measured to simulated magnitude of the $\vec{E}$-field for each event individually (selection RA). Similar plots for the other selections are going to be discussed below, but beforehand it is necessary to address some general features common to all selections.

The central line of the plots indicates a match between the simulated and measured amplitude. The lines outward denote on both sides a factor of two and a factor of four difference.

Obviously, not every event matches in every component of the $\vec{E}$-field the simulations well. To discuss the events with a significant deviation, it is prudent to define three categories with different degrees of differences.

Category I contains all events of a selection with a difference between a factor of three and four. Events with a difference above a factor of four in both stations are assigned to category II. Finally there are events remaining in between category I and II, which show a difference of above a factor of four only for one station. These events form the third category (see also Table 5.1.4 on page 72).

As later described, these categories are not arbitrarily chosen, but rather genuinely motivated by the estimated uncertainties.

Figure 5.1.3 on page 75 shows the events of selection RA. The selection contains 14 reconstructed events out of the total 32, which fall into category I, showing a difference within a factor of three to four for both stations. A total of 10 events fall into the second category with a larger than factor four difference in both stations. Finally eight events remain with a factor of four difference only in one station (category III).

For the selection MA of MGMR simulations the situation is similar (see Figure 5.1.4 on page 76). 15 of the total 32 events belong category I. Eleven events show a difference above a factor of four for both stations (category II) and six events show the same difference but only in one station (category III).

It is remarkable, that out of the events of the second and third category (difference of a factor of four or more for one or two stations) the majority exhibits a small simulated amplitude. This pattern is consistent for both the REAS3 (18 events) and MGMR (17 events) simulations. The measured amplitude of those events is significantly larger than the simulated one (e.g. the 12 events on the right hand side of Figure 5.1.3 on page 75 and Figure 5.1.4 on page 76).

This signature of a large difference between simulation and measurement, together with a strong underestimation of the $\vec{E}$-field strength by the simulations, is not unknown for air showers measured during thunderstorms.

In theory the strong atmospheric $\vec{E}$-fields and lightnings present in the atmosphere above the detector during thunderstorms can influence the radio signal from air showers. An amplification of the radio signals from air showers can be expected, as discussed by Melissas et al. [2010]. However the fundamental processes resulting in an amplification are not yet full understood, which makes it impossible to describe the exact strength of the amplification for the simulations. As a result, simulations for such thunderstorm events tend to underestimate the measured amplitude, if they ignore the influence.

Since atmospheric $\vec{E}$-field measurements are available for the BLS data set, thunderstorm
events can be easily identified and excluded from the analysis. In Figures 5.1.3 and 5.1.4 all thunderstorm events are marked for the four selections RA, RB, MA and MB.

As expected, all identified thunderstorm events belong to the second category by showing at least a factor of four difference in both stations. Again both the REAS3 and MGMR simulations are consistent.

After exclusion of these thunderstorm events, there still remain nine unexplained events of the second category, which require deeper investigations. It has to be stressed that for those events the atmospheric $\vec{E}$-field measurements give no evidence for the presence of thunderstorms during the measurement (see: appendix). This is in particular true for the events 4701978 and 4193458, which exhibit very strong radio pulses, very similar to the expectation from a thunderstorm period.

Further investigation requires to take the uncertainties of the measurement and simulations into account, to explore the significance of the observed differences. This includes the uncertainties of the SD reconstruction, which forms the basis for all the radio simulations as discussed in 5.1.2. The estimated uncertainties of the radio simulations arising from the SD reconstruction and the estimated uncertainty of the measurement based on the

noise level of the respective event are combined by application of Gaussian error propagation:

$$\begin{aligned}\sigma_{\text{ratio}} &= \sqrt{\left(\frac{\sigma_{\text{Signal}}}{A_{\text{Signal}}}\right)^2 + \left(\frac{\sigma_{\text{Sim}}}{A_{\text{Sim}}}\right)^2} \cdot \frac{A_{\text{Signal}}}{A_{\text{Sim}}} \\ &= \sqrt{\left(\frac{1}{\text{SNR}}\right)^2 + \left(\frac{\text{MAD}}{A_{\text{Sim}}}\right)^2} \cdot \frac{A_{\text{Signal}}}{A_{\text{Sim}}}\end{aligned}$$

with: $A_{\text{Signal/Sim}}$ = amplitude of the simulated/measured signal.

In Figure 5.1.5 the combined uncertainty is displayed for each event. Please note that thunderstorm events are now excluded. Since varied simulations are not available for all of the 29+8 events of the selections RA and RB as well as MA and MB, the comparison data set reduces to 23+7 events.

From Figures 5.1.5 and 5.1.6 it is obvious, that the events of the introduced categories are almost all within their uncertainties compatible in their magnitudes of the $\vec{E}$-field[4]. Out of the 23+7 events, a total of 13+6 events show a difference smaller than a factor of four. 10+1 of the 23+7 events belong to category I or II and thereby show a deviation of at least a factor of four in one or two stations. Only a small subset of 6+0 events shows a small simulated amplitude, which is incompatible with the measured data even after consideration of the uncertainties arising from the SD reconstruction. The source of the deviations of these 6 events are subsequently of primary interest.

It is possible to interpret the result as an unknown influence on the measured data causing similar differences in the magnitude of the $\vec{E}$-field like the influence from thunderstorms. However, before this can be taken for granted other possible sources have to be ruled out. Especially the estimated uncertainties from the SD reconstruction can be another prime source for the observed differences.

As mentioned, the SD reconstruction provides the foundation of the radio simulations, providing the beneficial independence of the radio simulations from the radio measurements. The SD reconstruction was originally optimised for the standard geometry of the SD array and has proven so far very successful. Since, however, an additional tank was deployed within the SD-field, to improve the coincident SD-RD measurements, the geometry of the SD array has changed.

4 The actual amplitude in the point is the one resulting from the non-varied simulation based on the SD measurement.

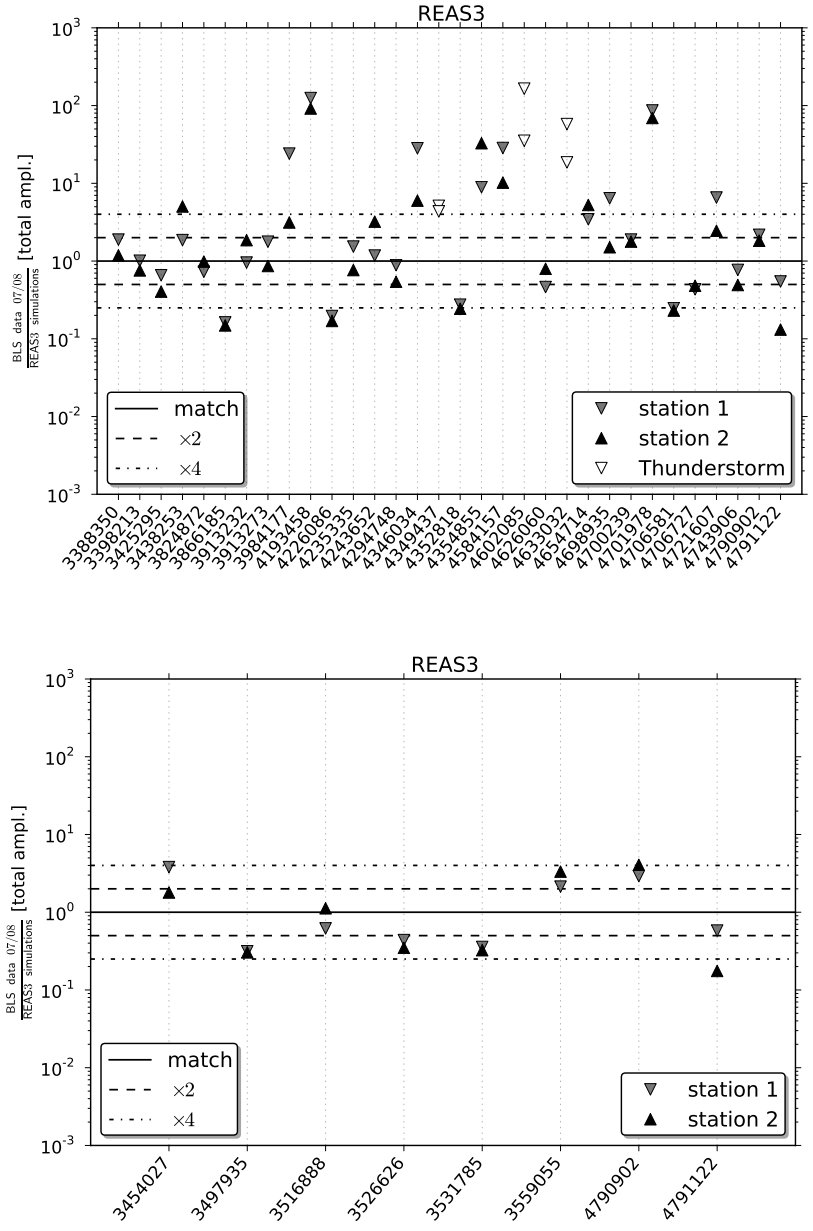


Figure 5.1.3: REAS3 selection RA, 32 events (upper plot) and selection RB, 8 events (lower plot)
magnitude of simulated versus the measured amplitude (thunderstorm events are marked in white if present). For events on the solid line, measured and simulated amplitude match. The dashed and dotted lines indicate a factor of 2 and 4 difference.

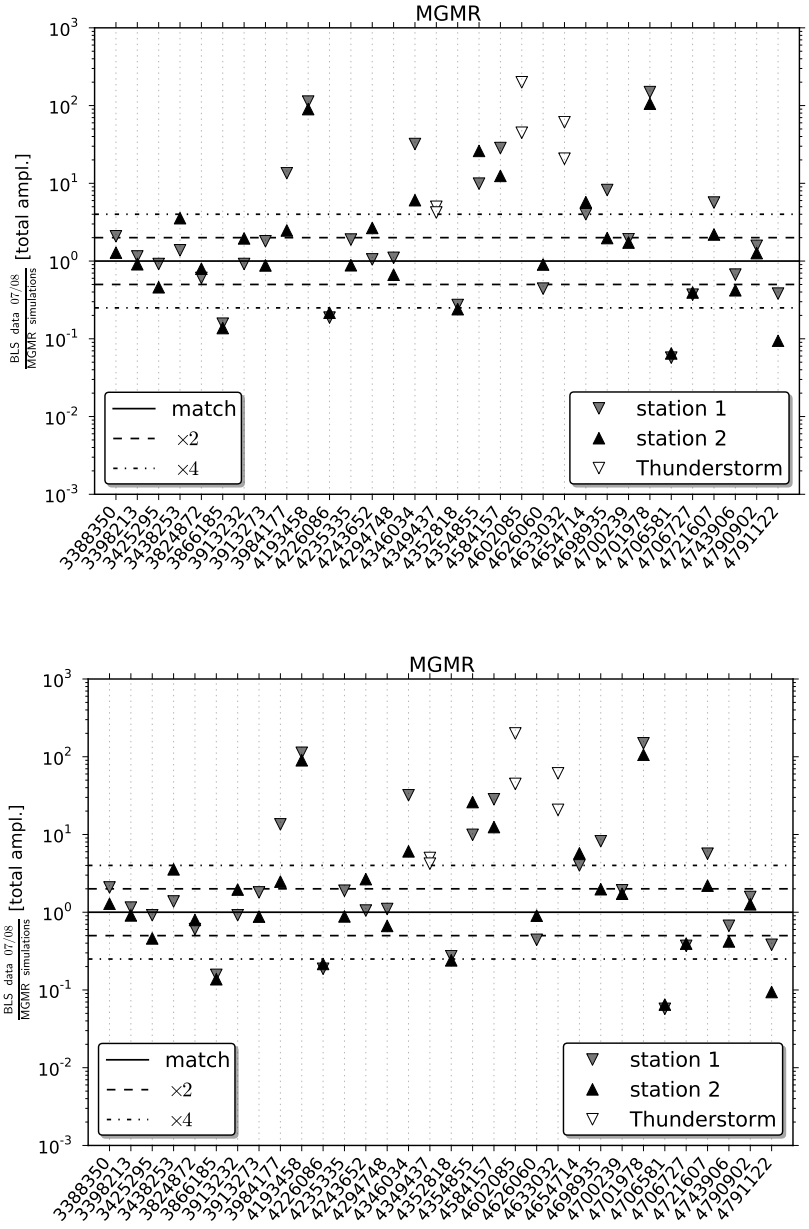


Figure 5.1.4: MGMR selection MA, 32 events (upper plot) and selection MB, 8 events (lower plot)
magnitude of reconstructed versus measured amplitude (thunderstorm events are marked in white if present). For events on the solid line, measured and simulated amplitude match. The dashed and dotted lines indicate a factor of 2 and 4 difference.

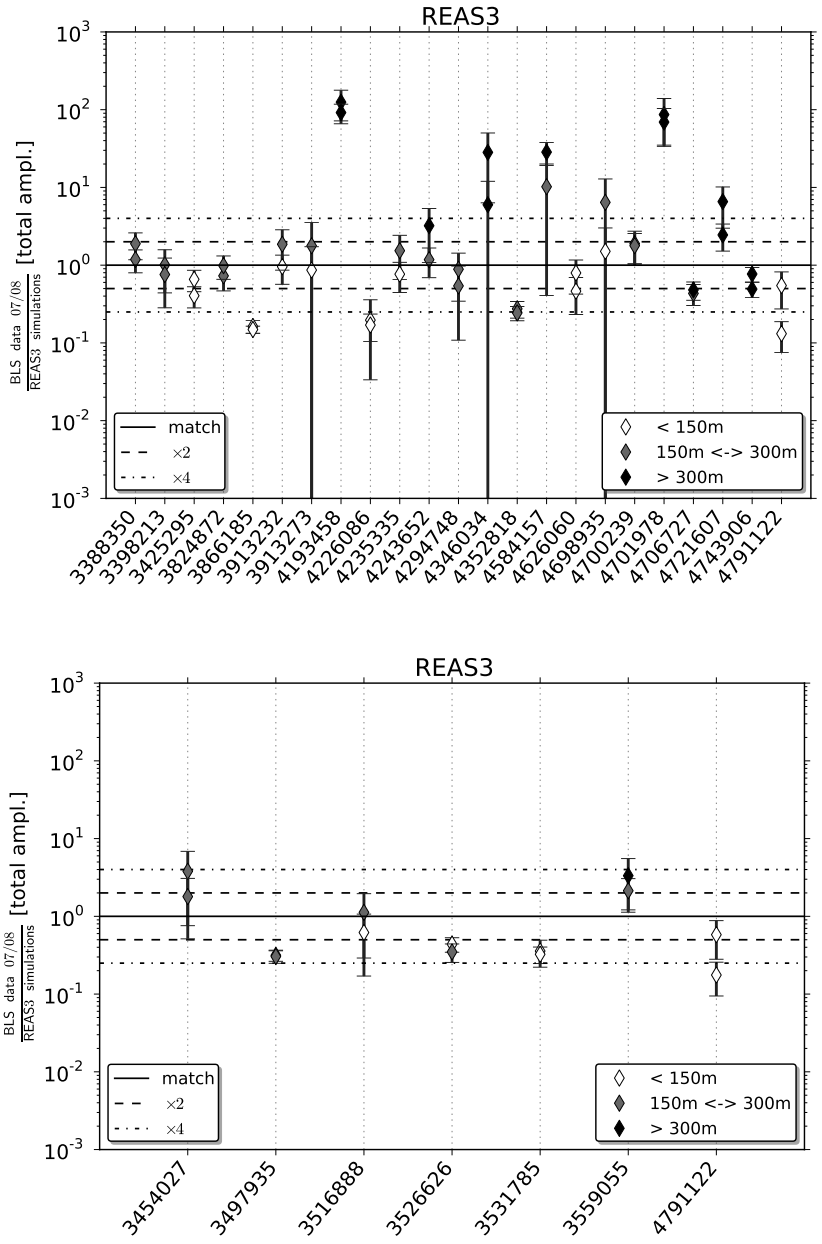


Figure 5.1.5: REAS3 selection RVA, 23 events (upper plot) and selection RVB, 7 events (lower plot)
magnitude of reconstructed versus measured amplitude excluding thunderstorm events. For events on the solid line, measured and simulated amplitude match. The dashed and dotted lines indicate a factor of 2 and 4 difference in the amplitudes. The events are separated in three bins (<150m, 150m-300m, >300m) by their distance to the shower axis.

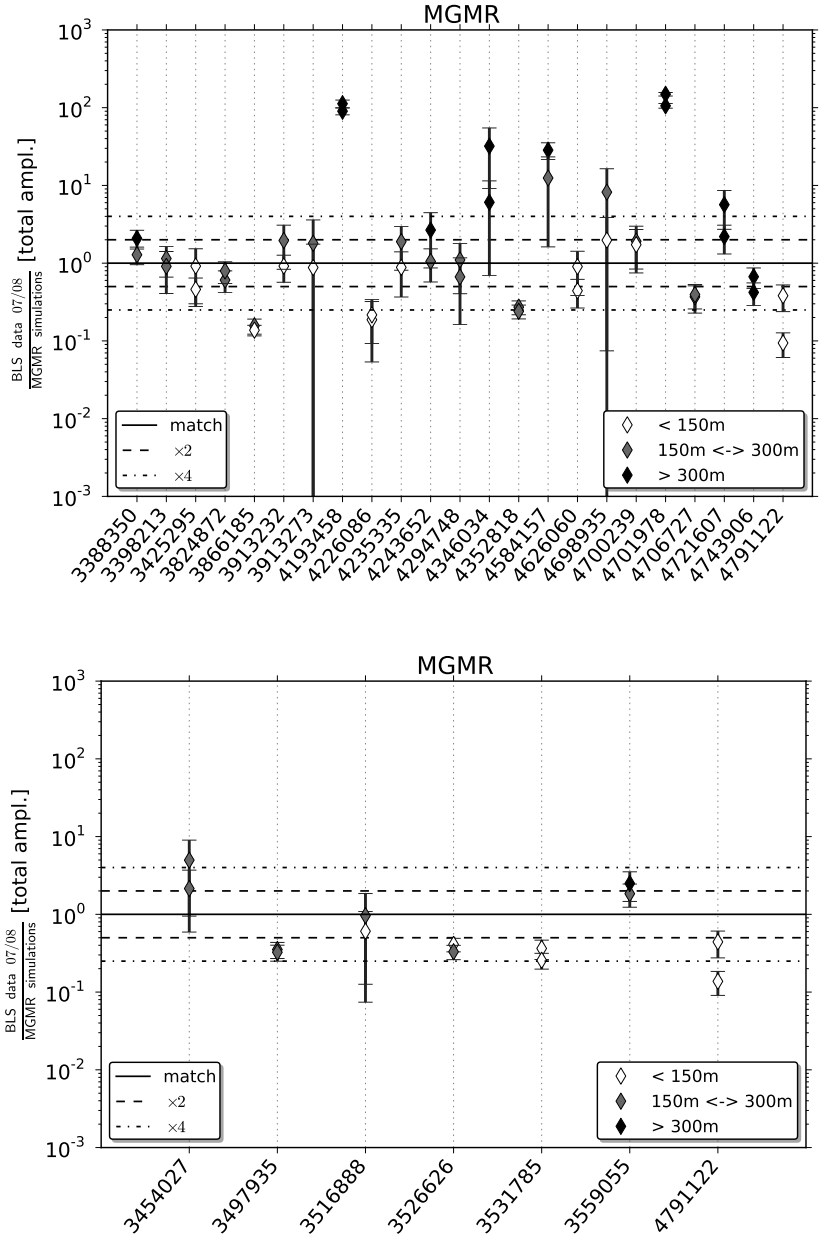


Figure 5.1.6: MGMR selection MVA, 23 events (upper plot) and selection MVB, 7 events
magnitude of reconstructed versus measured amplitude excluding thunderstorm events. For events on the solid line, measured and simulated amplitude match. The dashed and dotted lines indicate a factor of 2 and 4 difference in the amplitudes. The events are separated in three bins (<150m, 150m-300m, >300m) by their distance to the shower axis.

The here applied SD reconstruction does not cover the implications introduced by the additional tank on the reconstructed quantities and their estimated uncertainties[5]. This touches especially the shower core position reconstructed with CDAS for SD, since it is one major initial parameter of the radio simulations. A misreconstruction of this core position by SD could have a significant impact on the reconstructed $\vec{E}$-field strength from simulations. While the lateral distribution of the radio signals from air showers is still under investigation, results from smaller-scaled experiments provide indications for an exponential lateral distribution function with a scale factor in the order of 100-200 meters (compare the work of W.D. Apel et al. - LOPES Collaboration [2010] and P. Lautridou et al. - CODALEMA Collaboration [2009]). This LDF shows fast decreasing radio signals with growing distance to the air shower, a behaviour that can also be expected for BLS measurements. Assuming, that the SD reconstruction has mistakenly reconstructed the air shower core to a position further away from the antennas than it really was situated, would result in similar observations as the described ones. Radio simulations, based on the mistakenly reconstructed to far away core position, would predict a significantly smaller signal than actually measured by the antennas. Since the signature depends on the distance between the source and the antennas, it is necessary to sort the events into three distance bins[6]:

0m-150m, 150m-300m and $>$ 300m

In figures 5.1.5 and 5.1.6 the events are marked accordingly to their distance for each selection. A correlation between the distance towards the shower axis and the difference of the measured to the simulated amplitude is indeed present. The 6+0 events with too small simulated amplitudes as discussed before, show for five of the six events in at least one station an axis distance greater than 300m. This distance is well above the scale factor observed from small scale experiments.

In either case, without being a hard evidence, this result could be explained by a misreconstructed core position. This obviously affects the estimated SD uncertainties for the core position as they would be also underestimated. Also, it has to be stressed that the SD reconstruction in use was never thoroughly tested at the low energies typical for these radio measurements (well below 10^{18} eV).

Under the impression of the analysis results and implications, it thus seems to be more adequate to base the reconstruction of the BLS and later the AERA coincident SD events on the infill reconstruction. This was already described by the Pierre Auger

5 At the time of the analysis, an SD reconstruction covering these effects was not available.

6 The distance depends implicitly on the core position, since it fixes the position of the shower axis on the ground.

Collaboration 2011 as being superior to the regular SD reconstruction with CDAS.

The observations made during this analysis are also complementary to the findings of the analysis of Neuser et al. [2011], which demonstrated that the inclusion of one additional tank (like the Olaia tank at the BLS), can lead to significant differences in the regular SD reconstruction. This further strengthens the need to exchange the regular for the infill reconstruction in case of BLS measurements. The same holds up for AERA measurements, since it is situated within the SD-infill.

Alongside events with larger measured than simulated amplitude, a second group of events with an inverted behaviour also have to be mentioned. Several simulated events, close to the shower core, tend towards overestimating the measured radio signal. Recent studies of the LOPES experiment indicate for such events a decrease in the radio signal close to the shower core. This is today attributed to the influence of the refractive index on the propagation of the radio signal in the atmosphere. Refractive index effects were therefore only recently introduced to both simulation codes. First updated radio simulations are already available for the AERA setup but not yet for the here discussed BLS setup. The next section is therefore addressing AERA, as first radio measurements of the AERA setup are now also available for a comparison.

5.2 Comparison of AERA measurements with simulations

A comparison like the one described above for BLS data, is in case of AERA limited by two factors. First the available data set is not yet as large as the BLS data set. Second, the preparation of radio simulations, especially of varied simulation data sets for estimating the influence of SD uncertainties, is time consuming. Therefore the available data set for this analysis contains only 17 measured events. For those events varied radio simulations of the most recent radio simulation codes are available. The analysis is however not limited by the discussed misreconstruction due to a mistake within the analysis routines, as the data set is processed with an updated Offline version.

The data set does not include events measured during thunderstorms, as those are identified by data from the weather station installed at the CRS.

The radio simulation data sets are simulated with the REAS in CORSIKA (COREAS) (improved version of $REAS_3$) and the Electric fields, using a Variable index of refraction in Air shower simulations (EVA) (improved version of MGMR) simulation code. Both simulations are based on coincident SD measurements of

Algorithm 5.1 Offline module sequence applied to measured AERA coincident events.

```
<moduleControl>
    <loop numTimes="unbounded">
        <!-- SD reconstruction of the hybrid data -->
        <module>  RdChannelADCToVoltageConverter </module>
        <module> RdChannelPedestalRemover        </module>
        <module> RdChannelMedianFilter           </module>
        <module> RdChannelResponseIncorporator   </module>
        <module> RdEventInitializerHybrid        </module>
        <module> RdChannelUpsampler              </module>
        <module> RdChannelTimeSeriesWindower     </module>
        <module> RdPreWaveFitter </module>
        <loop numTimes="unbounded">
            <module> RdAntennaChannelToStationConverter </module>
            <module> RdStationBandpassFilter </module>
            <module> RdStationSignalReconstructor </module>
            <module> RdDirectionConvergenceChecker </module>
            <module> RdWaveFit </module>
        </loop>
        <module> RdEventPostSelector </module>
        <module> RecDataWriterNG </module>
    </loop>
</moduleControl>
```

the emitting air shower. The preparation of the varied radio simulations follows the lines of the BLS analysis. They include both the influence of SD uncertainties and of shower-to-shower fluctuations on the radio simulations.

As explained above (c.f. chapter Chapter 5), the noise level is estimated by calculating the RMS of the noise window. The so estimated noise level provides also an estimate on the uncertainty of the AERA measured $\vec{E}$-field.

5.2.1 *Reconstruction of the measured and simulated data sets*

The 17 measured events are all recorded by the first stage of AERA (24 antennas). For each radio event the coincident SD reconstruction of the air shower is available. Therefore it is possible to do a hybrid reconstruction of the radio data set by including the SD reconstructed incident direction. As a side effect, it is now also possible to reconstruct events with only one or two radio stations. Therefore a SAD cut is also not necessary, since the incident direction is not reconstructed from radio.

Also an SNR cut is not applied, as the radio events are coincident with the SD measurement. The applied module sequence for the reconstruction is given in 5.1.

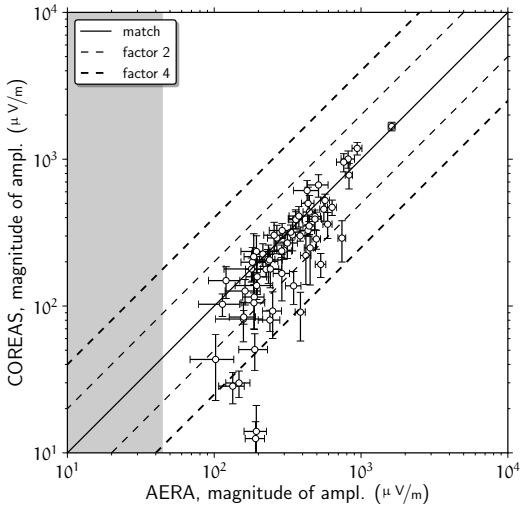
The two simulation codes used to prepare the radio simulations are COREAS and EVA. Both simulation codes are improvements of the REAS3 and MGMR simulation codes, used for the BLS comparison. They both include for the first time the influence of the refractive index of the atmosphere on the radio signal.

Algorithm 5.2 Offline module sequence applied to the EVA and COREAS simulations

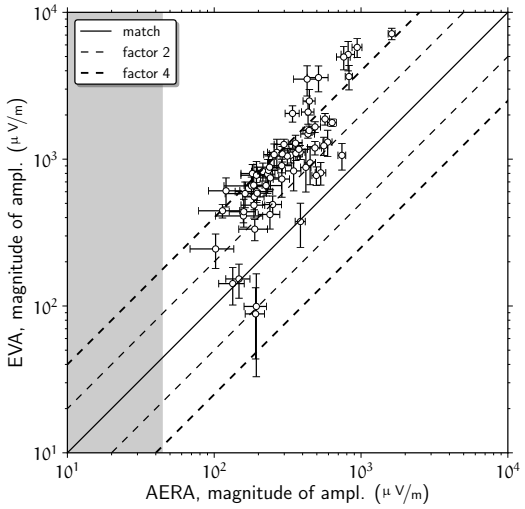
```
<moduleControl>
      <loop numTimes="unbounded">
        <module> EventFileReaderOG </module>
        <!-- Begin of detector simulation -->
        <module> RdStationAssociator </module>
        <module> RdAntennaStationToChannelConverter </module>
        <module> RdChannelResponseIncorporator </module>
        <module> RdChannelResampler </module>
        <module> RdChannelTimeSeriesClipper </module>
        <module> RdChannelVoltageToADCConverter </module>
        <!-- Begin of reconstruction -->
        <module> RdEventInitializer </module>
        <module> RdChannelADCToVoltageConverter </module>
        <module> RdChannelPedestalRemover </module>
        <module> RdChannelResponseIncorporator </module>
        <module> RdChannelUpsampler </module>
        <module> RdPreWaveFitter </module>
        <loop numTimes="unbounded">
          <module> RdAntennaChannelToStationConverter </module>
          <module> RdStationSignalReconstructor </module>
          <module> RdDirectionConvergenceChecker </module>
          <module> RdWaveFit </module>
        </loop>
        <module> RdEventPostSelector </module>
        <module> RecDataWriterNG </module>
      </loop>
</moduleControl>
```

The radio simulations are done for all 24 radio stations of the AERA setup, not only the radio stations participating in the measured radio event. Therefore it is possible to apply a standard radio reconstruction without relying on the incident direction of the SD reconstruction. As for the BLS comparison, no further SNR or SAD cuts are applied to the radio simulations and also the simulation of the ambient noise level is unavailable.

The module sequence applied to the simulation reconstruction is given in 5.2.

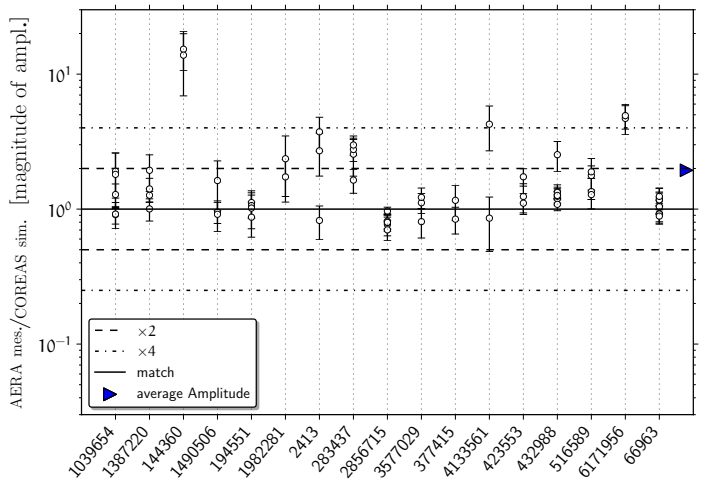


(a) COREAS simulations

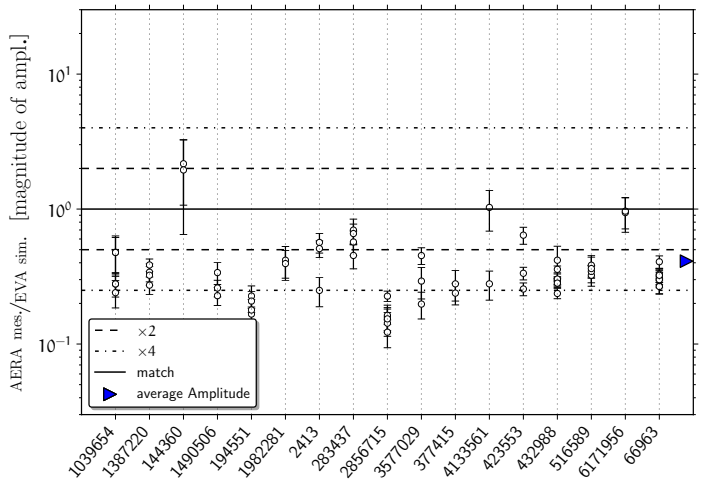


(b) EVA simulations

Figure 5.2.1: Amplitude measured by 64 AERA radio stations (17 events) vs. the predicted amplitude by COREAS / EVA simulations. The grey-shaded area indicates the average noise level of the measurements. COREAS simulations (upper figure) tend to underestimate small measured amplitudes. The EVA simulations (lower figure) systematically overestimate the measured amplitudes.



(a) COREAS simulations



(b) EVA simulations

Figure 5.2.2: Ratio between AERA measured and simulated COREAS / EVA amplitudes for 17 events (several stations per event). The error bars visualise the uncertainties on the measurement from the ambient noise level, the uncertainties on the radio simulations from shower-to-shower fluctuations and the SD reconstruction. The COREAS simulations (upper figure) agree in average (blue triangle) within a factor of 2 with the measurement. The EVA simulations (lower figure) agree in average (blue triangle) within a factor of 3 with the measurement, but systematically overestimate the measured amplitudes.

5.2.2 *Amplitude comparison*

Figure 5.2.1a on page 83 and 5.2.1b show the COREAS / EVA simulated magnitude of the amplitude vs. the AERA measured magnitude of the amplitude.

For the COREAS simulations, the results of the BLS study are in general confirmed, as they agree in average within a factor of two with the measurement. This is an improvement, which can be accounted to the improved COREAS simulation code, as the difference reduces from a factor of four to a factor of two.

The EVA simulations on the other hand, are systematically overestimating the AERA measured amplitude. They are therefore not as good in agreement with the original BLS comparison, which showed no systematics. The average difference in the magnitude of the amplitude is still a factor of three, and therefore comparable with the differences observed in the BLS comparison.

For small amplitudes, both simulation data sets tend to underestimate the measurement. Due to the small applied SNR cut, small measured amplitudes are dominated by the ambient noise level. In figure Figure 5.2.1a on page 83 and 5.2.1b is therefore the average noise level of all events indicated by the grey shaded area[7]. The influence of the noise level on the measured amplitude is not included in any of the two radio simulation data sets, as an ambient noise simulation is currently not available within Offline.

The observed underestimation of the radio signal by simulations is therefore expected for events with a small measured amplitude, since the radio noise can constructively interfere with the radio signal. Accordingly, events with small amplitudes are dominated by the ambient noise level, which can influence the reconstructed amplitudes. The result is sometimes an increased measured amplitude in comparison to the predictions of the radio simulations due to constructive interference.

In Figure 5.2.2a on page 84 and 5.2.2b the same data set is displayed separated by events. The ratio of the measured vs. the simulated magnitude of the amplitude is shown on the y-axis. The event ID is displayed on the x-axis. For each event, the participating radio stations are displayed as one data point each. The number of radio stations varies between events, since the number of radio stations depends on aspects like the primary particle energy and the shower geometry.

The average amplitude ratio over all events is indicated for both simulations as a triangle on the right hand side of the figure.

7 The noise level is estimated by the RMS of the timeseries noise window of each event. The mean of the noise levels of all events is calculated and displayed as the grey area.

For COREAS simulations the average amplitude ratio is 1.94. The COREAS simulations therefore agree within a factor of two with the measurement and slightly tend towards underestimating the measured amplitude.

The average amplitude ratio of the EVA simulations is 0.41. The EVA simulations therefore agree within a factor of 2.5 with the measurement but unlike the COREAS simulations, the EVA simulations systematically overestimate the measured amplitude.

5.3 SUMMARY OF THE BLS AND AERA STUDY

In the presented analysis of AERA and BLS data, the introduced improvements discussed in previous chapters, like the three-dimensional reconstruction of the $\vec{E}$-field vector and the updated response patterns, are applied to their full extend. In addition, updated station positions with increased precision for the BLS setup are used and a full detector simulation is included for radio simulations.

The preparatory BLS study for AERA data analysis compared the radio measurements performed at the BLS with per-event simulations carried out with the simulation codes REAS$_3$ and MGMR. Many events agree between the simulated and measured magnitude of the electric field within the uncertainties on the radio simulations arising from the SD reconstruction. A significant fraction of events, however, exhibits deviations from the simulated electric field amplitudes larger than a factor of four. For most of these events, the field strength predicted by simulations is lower than the actual measured field strength. This signature is familiar for events measured during thunderstorms, which is the case for three events with a vastly too low predicted field strength.

About a third of the events still shows deviations which are larger than a factor of four in at least one station after exclusion of thunderstorm influenced events. A closer investigation on the dependency between the axis distance and low predicted $\vec{E}$-field strengths reveals a pattern. All remaining events with a larger than factor of four difference between simulations and measurements exhibit an axis distances of > 300m.

A possible explanation, is a core position adopted further away from the antennas than it is actually the case. A possible explanation for this could be a misreconstruction of the SD core position in combination with a possible underestimation of the reconstruction uncertainties. A deeper analysis of the SD reconstruction quality in the parameter space of typical AERA measurements — which is outside the well-tested region of paramet-

ers — would therefore be beneficial but is not yet available. At the moment it cannot be ruled out, that also other effects, leading to the adoption of a mismatched core position for the radio simulations beside a SD misreconstruction, could be the cause of the observed differences between simulations and measurements. Therefore the reliability of the underlying SD reconstruction should be revisited in the future as it is clear that at energies well below 10^{18} eV, the standard SD reconstruction cannot simply be taken at face value. Instead, the infill reconstruction should be used as it is done in the presented AERA study.

For this study, radio events measured by AERA were compared to radio simulations. The AERA measurements are also in coincidence with SD measurements and reconstructed with the Infill reconstruction. The radio simulations used, are improved versions of the REAS$_3$ now COREAS, and MGMR now EVA simulation codes, which include refractive index effects of the atmosphere. It is therefore also possible to asses the progress in the field of radio simulations alongside a comparison with measured data.

The comparison is in good agreement with the results of the BLS study. With Thunderstorm events excluded, the average difference between measured and simulated $\vec{E}$-field strength is a factor of two for COREAS and a factor of two and a half for EVA simulations. Notably is a systematic effect visible in EVA simulations, that is not observed in MGMR and recent COREAS simulations. This results in a systematic overestimation of the $\vec{E}$-field strength by EVA simulations.

In both simulations (if the systematic effect is neglected), one outlier is observed, that exhibits a larger measured amplitude than the simulation prediction. At the moment, it is not possible to relate this outlier to uncertainties of the SD reconstruction. However, the event is dominated by the ambient noise level during measurement. This noise level is not yet included into the radio simulations and therefore could be a possible explanation for the observed underestimation, as thunderstorm events are excluded from the analysis. The BLS study does not suffer from the effect as a strong SNR cut is applied to exclude events dominated by noise. This approach was however forgone for the AERA study as the amount of available data is still small. Future studies, on basis of a larger data set, should again exclude noise dominated events.

In direct comparison with the BLS study and the there applied older versions of the radio simulation codes, COREAS shows a reduction from a factor of four difference (REAS$_3$) to a factor of two difference in average. This result agrees with recent publications of the LOPES collaboration (cf. Ludwig, 2012) MGMR simulations also improved with the introduction of EVA but to a less extend, as they now exhibit a systematic effect and therefore

show an average difference of a factor of two and a half. However, both simulation codes have improved, since the BLS study with COREAS being at the moment slightly superior to EVA.

Part II

SENSITIVITY OF AERA MEASUREMENTS ON AIR SHOWER CHARACTERISTICS

6

ENERGY ESTIMATION FOR RADIO MEASUREMENTS BASED ON THE LDF

Right from the beginning of radio measurements of cosmic ray air showers in the 1960's, the long-term goal was the development of another detection technique alongside particle and Čerenkov detectors. This requires the successful reconstruction of air shower quantities, like the primary particle energy or X_{max}. One promising approach, investigated with the LOPES experiment, is the lateral distribution of the radio signals on ground, which should be depending in its shape on both the energy and X_{max} (cf. W.D. Apel et al. - LOPES Collaboration, 2012b).

First attempts for such analyses in the scope of AERA were recently undertaken by Glaser [2012] in a very similar way to the one performed with the LOPES experiment (cf. Palmieri, 2012). This chapter presents an alternative approach for an analytical description of the LDF, which in addition improves the description of single radio events over former methods. The new LDF description takes the most recent results from the LOPES and LOFAR experiments on the shape of the LDF into account. As a first step, the derived LDF is compared with the usually applied method as it is for example described by Palmieri [2012]. Afterwards the derived LDF is applied to determine an energy estimator for AERA and to investigate the dependency on X_{max}. In the last part, a reconstruction of the air shower core is done on basis of the derived LDF description. The results of this reconstruction are also compared to recent results of the CODALEMA experiment (cf. the CODALEMA collaboration, 2011).

6.1 INTRODUCTION TO LDF OF RADIO SIGNALS

The lateral distribution, i.e. how the radio signal strength evolves with increasing distance to the shower axis, can be described by an analytic function. Unfortunately, this analytic form is not available from theoretical predictions, since the relative contributions of different emission mechanisms are still under investigation. In addition, the shape of the LDF also depends on the sensitive frequency range of the experiment. Therefore a common approach is the derivation of an appropriate parametrisation based on measurements or simulations.

Already the very first radio experiments back in the 1960's followed this approach and fitted parametrisations to the LDF. Based on measurements, the Haverah Park experiment estim-

ated an exponentially decreasing LDF, a result incompatible with measurements of the Moscow State University (MSU)-array, which determined a power law.

Only recently, the LOPES experiment was able to shed new light on this contradiction, as the LOPES results support the results of the Haverah Park experiment (cf. W.D. Apel et al. - LOPES Collaboration, 2010).

Therefore an exponential parametrisation of the lateral distribution is nowadays widely accepted as the appropriate LDF function to describe radio measurements in the typical measurement range of 100 m to 600 m.

This result was challenged by other results of the LOPES experiment (cf. W.D. Apel et al. - LOPES Collaboration, 2010) as well as the LOFAR experiment (cf. A. Nelles et al. - Pierre Auger Collaboration, 2012), which both observed a flattening of the LDF towards the shower axis for several radio events.

A flattening towards the shower axis is not described by a monotone exponential decrease. It is therefore necessary to modify the parametrisation. My approach is based on the Gaussian distribution function, which naturally shows a flattening towards its maximum. The ultimate goal is a description of events with a pure exponential and likewise of events with a flattening LDF towards the shower core. Both parametrisations are compared with each other below, followed by the derivation of an energy estimator based on the Gaussian parametrisation. The subsequent discussions are all based on several data sets, which separate into radio simulations and first AERA measurements.

6.1.1 *Data sets for the analysis*

Radio simulations[1] provide an ideal basis for the development and test of new analysis methods as they provide information on the air shower, which are sometimes only accessible by sophisticated reconstructions of data, taken with other detection methods, i.e. SD. One example are input parameters of radio simulations, which can be used for correlation studies as discussed later in this chapter. The Offline framework provides the option to apply a full detector simulation with a subsequent reconstruction of radio observables like the $\vec{E}$-field for radio simulations. This is a widely used method within AERA, and also applied for this study. The simulation data sets prepared for this study are all

1 The radio simulations are based on an air shower simulation provided by the CORSIKA code, which applies the URQMD and QGSJETII model to describe the particle interactions in the air shower.

specific to the geometry of the AERA setup[2] and rooted on the following input parameters:

- the shower core position on the ground for the particle shower
- the incident direction
- the primary particle energy (further referred to as p.p. energy)
- the primary particle type (here either a proton or an iron nucleus)

The input parameters are fixed. They are either chosen beforehand and therefore precisely known or reconstructed air shower parameters from other detection methods, i.e. SD or FD can be used. The second approach is usually applied, when measured air showers are simulated, as the radio simulation is in this case still independent from the radio measurement. If the input parameters are reconstructed quantities of SD, the uncertainties on the radio simulations have to include the SD uncertainties and should be part of the simulation analysis. Due to the non-linear nature of radio simulations, an indirect approach has to be applied to estimate the radio simulation uncertainties. Within the SD uncertainties, e.g. for the energy, 25 initial parameter sets are randomly chosen and the same number of varied radio simulations is prepared. The final simulation dataset then contains $25 + 1$ radio simulations per event; one radio simulation for the SD reconstructed initial parameters and 25 for the randomly chosen initial parameter sets. A simple calculation of the standard deviation of the simulation results yields an estimate of the uncertainty of the SD measurement on the radio simulations[3]. The same approach includes shower-to-shower fluctuations as discussed in the theoretical discourse in the beginning of this work.

The only aspect currently not included in the radio simulations is the influence of the ambient radio noise present during measurement. At the moment, the ambient radio noise cannot be simulated for AERA within $\overline{\text{Off}}\underline{\text{line}}$. The influence of the ambient noise level on the here presented analysis should therefore be assessed as soon as a noise simulation is available.

However, even as physical radio noise is absent from current radio simulations, numerical noise from the radio simulations and incoherent contributions to the radio signal still exist. The

2 Since the BLS data sets of the previous chapters consist only of three stations, a complex LDF fit requiring at least 4 stations is impossible. The dataset is thereby not suited for this kind of study.

3 This is of course only true, if the simulation uncertainties are Gaussian distributed as it is generally assumed.

results are non vanishing radio signals even for radio stations far away from the shower axis (>500 m). Even as the observed noise levels are significantly smaller than physical noise levels, they behave similar for this analysis and have to be considered during analysis.

6.1.1.1 *The E100p data set*

It contains 100 radio simulations of which 98 can be reconstructed by Offline. They are flat distributed in the primary particle energy between 10^{17}eV $- 10^{19}$eV and have proton as the primary particle[4]. The input parameters of the data set are:

core position	inside the AERA field
incident direction	orthogonal to the geomagnetic field
p.p. energy	10^{17}eV $- 10^{19}$eV
p.p. type	proton

The incident direction is chosen orthogonal on the geomagnetic field to remove the influence of varying contributions from different radio emissions with the incident direction. This data set therefore represents an ideal case, since the radio emission is expected to be strongest for this configuration.

6.1.1.2 *The E100I data set*

It contains 100 radio simulations of which 99 can be reconstructed by Offline. They are flat distributed in the primary particle energy between 10^{17}eV $- 10^{19}$eV and have iron as the primary particle. The input parameters of the data set are:

core position	inside the AERA field
incident direction	orthogonal to the geomagnetic field
p.p. energy	10^{17}eV $- 10^{19}$eV
p.p. type	iron

Again, the incident direction is chosen orthogonal on the geomagnetic field, which therefore represents an ideal case.

6.1.1.3 *The AvarCp data set*

This simulation data set contains 15 simulations based on AERA-SD coincident measurements. Each event is simulated 25+1 times

4 p.p. = primary particle

with input parameters taken from the SD reconstruction of the air shower. The data sets can be averaged to estimate the uncertainties from shower-to-shower fluctuations and the input parameters as discussed in chapter 5. It is explicitly stated, if the data set is used without averaging.

core position	from SD
incident direction	from SD
p.p. energy	from SD
p.p. type	proton
variations	25+1

6.1.1.4 *The AvarCI data set*

This simulation data set is similar to the AvarCp data set, with the exception of an iron nucleus as primary particle.

core position	from SD
incident direction	from SD
p.p. energy	from SD
p.p. type	iron
variations	25+1

As the simulation data sets AvarCp/I cover only the energy range between 10^{17} eV to 2×10^{18} eV, the E100p/I data set is combined to derive an energy estimator up to 10^{19} eV. The only difference between both data sets is the incident direction. The E100p/I data set has a fixed incident direction, while the AvarCp/I data sets have varying incident directions. Keeping this in mind, it is nevertheless possible to combine both data sets into one.

6.1.1.5 *Acoinc data set*

Measured coincident AERA-SD events, recorded during November 2011 to July 2012 are used. This data set contains 129 radio events, of which 40 are excluded as they exhibit not enough radio stations for an LDF fit. After the LDF fit and all applied cuts 21 events remain for analysis.

The LDF fit is applied together with a radio core variation for the measured radio events. The applied cuts and the excluded events are discussed in more detail in the following section.

Furthermore the average noise level of the of the data set is calculated and used as uncertainty estimates for radio simulations.

Algorithm 6.1 Offline module sequence applied to measured AERA coincident events for the LDF study.

```
<moduleControl>
    <loop numTimes="unbounded">
        <!-- SD reconstruction of the hybrid data -->
        <module>  RdChannelADCToVoltageConverter </module>
        <module> RdChannelPedestalRemover        </module>
        <module> RdChannelMedianFilter           </module>
        <module> RdChannelResponseIncorporator   </module>
        <module> RdEventInitializerHybrid        </module>
        <module> RdChannelUpsampler              </module>
        <module> RdChannelTimeSeriesWindower     </module>
        <module> RdPreWaveFitter </module>
        <loop numTimes="unbounded">
            <module> RdAntennaChannelToStationConverter </module>
            <module> RdStationBandpassFilter </module>
            <module> RdStationSignalReconstructor </module>
            <module> RdDirectionConvergenceChecker </module>
            <module> RdWaveFit </module>
        </loop>
        <module> RdEventPostSelector </module>
        <module> RecDataWriterNG </module>
    </loop>
</moduleControl>
```

6.1.2 *Analysis of the data sets with Offline*

The applied module sequence for measured AERA radio events in coincidence with SD is displayed in 6.1. For hybrid measurements, the SD events are reconstructed with the standard observer reconstruction before the radio reconstruction. The standard observer reconstruction is commonly used within the Pierre Auger Collaboration for SD data analysis.

The reconstruction of radio events applies a SNR cut of

$$\left(\text{Signal}/\text{Noise}\right)^2 \quad = \quad 6.5$$

and excludes events with less than three radio stations as an incident direction cannot be reconstructed for these events.

Radio simulations are reconstructed with Offline by the module sequence given in 6.2. The first part is the detector simulation, while the second part contains a standard reconstruction sequence as it is commonly applied to measured radio data.

After the Offline reconstruction, several additional quality cuts are applied during the LDF fitting procedure:

- At least 1 radio station with SNR > 6.5 is required within 300 m from the shower axis. This cut excludes events, for which the LDF is an extrapolation on basis of a few radio stations at a relatively large distance from the air shower.

- For simulations it is furthermore required that the angle between the radio reconstructed incident direction and the simulation input incident direction is smaller than 5°. This

Algorithm 6.2 Offline module sequence applied to simulated AERA events for the LDF study (data sets AvarCp/I, E100p/I).

```
<moduleControl>
      <loop numTimes="unbounded">
        <module> EventFileReaderOG </module>
        <!-- Begin of detector simulation -->
        <module> RdStationAssociator </module>
        <module> RdAntennaStationToChannelConverter </module>
        <module> RdChannelResponseIncorporator </module>
        <module> RdChannelResampler </module>
        <module> RdChannelTimeSeriesClipper </module>
        <module> RdChannelVoltageToADCConverter </module>
        <!-- Begin of reconstruction -->
        <module> RdEventInitializer </module>
        <module> RdChannelADCToVoltageConverter </module>
        <module> RdChannelPedestalRemover </module>
        <module> RdChannelResponseIncorporator </module>
        <module> RdChannelUpsampler </module>
        <module> RdPreWaveFitter </module>
        <loop numTimes="unbounded">
          <module> RdAntennaChannelToStationConverter </module>
          <module> RdStationSignalReconstructor </module>
          <module> RdDirectionConvergenceChecker </module>
          <module> RdWaveFit </module>
        </loop>
        <module> RdEventPostSelector </module>
        <module> RecDataWriterNG </module>
      </loop>
</moduleControl>
```

cut excludes radio events with a misreconstructed incident direction.

- The LDF fit requires at least four radio stations with SNR > 6.5.
- The log-likelihood value of the LDF fit (quality), has to be smaller than 1000. This cut helps to exclude radio events, which are noise dominated. It is applied for the analysis of measured radio events.

6.2 PARAMETRISATIONS OF THE LDF

The standard method to parametrise the LDF of radio signals is a fit of a chosen analytical function (parametrisation) to measured radio events. Depending on the analysis in mind, different components of the radio signal can be used[5]. In this work I focused on the estimation of the energy of the primary particle. This energy is expected to be proportional to the amount of charged particles in the air shower and therefore to the radio signal $|\vec{E}|$.

In a general definition, the LDF is a two-dimensional function of the position on the ground $\vec{x}$. Under realistic conditions, the deployed stations are spread out to cover an as large area as pos-

5 For example, also one single component of the $\vec{E}$-field vector, e.g. EW, can be used to investigate the LDF of that polarisation.

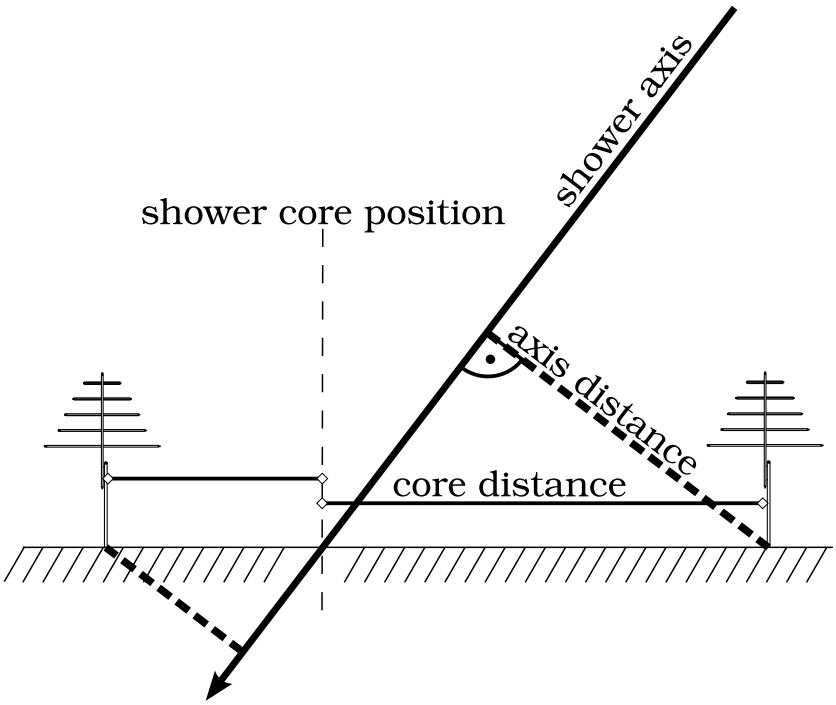


Figure 6.2.1: Geometry of the axis distance and radio stations with respect to the shower axis. Note that the distance of a radio station to the shower axis is defined orthogonal onto the shower axis.

sible. This increases the sensitive area and also reduces the costs of the antenna array. On the downside the average number of stations close to the shower axis decreases. Since the radio signal is expected to be strongest near the shower axis, a lack of radio stations close to the shower axis usually provides insufficient information for a fit of a two dimensional LDF. A common method to solve this problem, is a projection onto one dimension. By projection, the amount of fit parameters of the LDF reduces as the amount of radio stations close to the shower axis increases. This enables a fit of a one-dimensional LDF. One possible projection for this reduction to one dimension, is the orthogonal distance from the radio station to the shower axis (the shower axis is in this representation at 0) as illustrated in figure 6.2.1 . The projection implicitly assumes that the radio emission shows a radial symmetry around the shower axis.

However, as the emitted radio signal strength depends on the Lorentz force, i.e. the angle α between the geomagnetic field and the incident direction, an azimuthal asymmetry is expected. Therefore it is necessary to normalise the $\vec{E}$-field, i.e. the field strengths are divided by $\sin(\alpha)$.

In this work a log-likelihood fit of the squared absolute field strength $|\vec{E}|^2$ is chosen to estimate the parameters of the one-dimensional parametrisations. $|\vec{E}|^2$ is chosen over $\vec{E}$, as this quant-

ity is proportional to the power of the radio signal. Therefore the power of the radio noise can be treated as a constant offset with regard to LDF parametrisations. In case of $\vec{E}$, this is not possible, since de-/constructive interference between the radio signal and the radio noise is possible. The de-/constructive interference of the radio signal with radio noise can result in a de-/increase of the resulting signal. Therefore the radio noise cannot be treated as a constant noise level for $\vec{E}$.

The radio simulations do not provide estimates for the uncertainties of the simulated amplitudes. Furthermore, the uncertainties cannot be estimated on basis of the noise like for measured data, as no radio noise simulation is available. Therefore the uncertainties on the simulated $|\vec{E}|^2$ are estimated by taking the average noise level of the Acoinc measured events, which is 539.53 —V^2/m^2. Even as this approach cannot substitute a full noise simulation, it still provides an estimate for the influence expected from the ambient noise level at AERA on the radio simulations.

The standard parametrisation for the LDF is an exponentially decreasing function (cf. figure 6.2.2):

$$f(x) \quad = \quad p_{e0} \cdot e^{p_{e2} \cdot x} + p_{e3}$$

with the following parameters:

- p_{e0}, which corresponds to the signal at the shower axis
- p_{e2}, giving the slope and therefore how fast the signal decreases with distance
- p_{e3}, a parameter mutually connected to the ambient noise level of the measurement.

The ambient noise level, omnipresent during measurements, results in non-zero radio signals even for radio stations that did not see the radio event. The reconstructed radio signals of those stations usually do not pass applied filters, e.g. signal to noise ratio cuts. The piece of information that no radio signal was measured at a given distance however contains still information on the form of the LDF[6]. A complete integration of this evidence however requires dedicated statistical investigations with large data sets and a good understanding of the detector performance. This is short after the deployment of AERA not possible, but it is still possible to include stations far away from the shower axis by the p_{e3} parameter of the LDF. This parameter then corresponds to the average ambient noise level of the measurement.

In this case, the exponential parametrisation has three free parameters. The number of free parameters constrains the number of required stations (# of free parameters +1) for the fit.

6 This is also known as the concept of silent stations, e.g. in SD analysis.

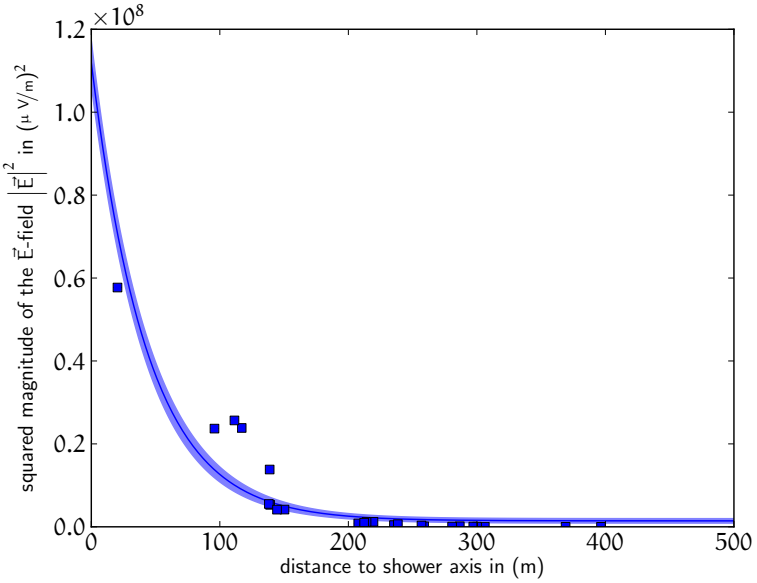


Figure 6.2.2: Example event (80114) of the E100p simulation data set. LDF fitted with the exponential parametrisation. $\left|\vec{E}\right|^2$ is proportional to the power of the radio signal. The errors on $\left|\vec{E}\right|^2$ are the average ambient noise level of measured events (Acoinc).

The exponential parametrisation is not suited to describe a flattening towards the shower axis as displayed in figure 6.2.2.

Therefore a new approach based on a Gaussian distribution (cf. figure 6.2.3) is subsequently investigated:

$$f(x) \quad = \quad p_{g0} \cdot \exp\left(\left(p_{g1}+x\right)^2/2\cdot\left(p_{g2}\right)^2\right) + p_{g3}$$

This parametrisation has four instead of three parameters:

- The p_{g0} parameter is again corresponding to the maximum signal
- p_{g1} corresponds to the position of the maximum signal relative to the shower axis.
- The p_{g2} parameter gives the width of the Gaussian distribution function and therefore has a similar interpretation like the p_{e2} parameter of the exponential parametrisation.
- The final parameter, p_{g3} models once again the ambient noise floor.

The Gaussian parametrisation in general has four free parameters. Under the assumption that the highest radio signal occurs at the shower axis, i.e. the source of the radio emission, the Gaussian distribution is fixed around 0. This reduces the free para-

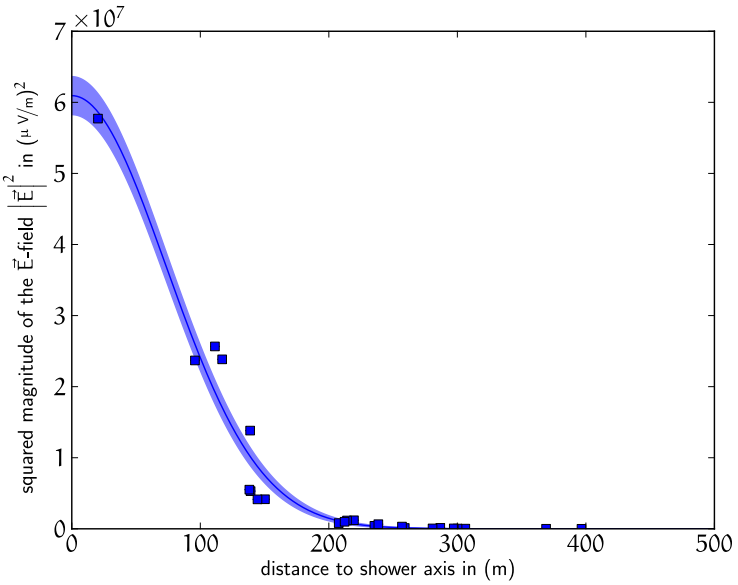


Figure 6.2.3: Example event (80114) of the E100p simulation data set. LDF fitted with the Gaussian parametrisation. $\left|\vec{E}\right|^2$ is proportional to the power of the radio signal. The errors on $\left|\vec{E}\right|^2$ are the average ambient noise level of measured events (Acoinc).

meters to three and increases the number of events to which the method can be applied.

In figure 6.2.2 an example lateral distribution of simulated radio signals is displayed. The radio signals usually cover a large range of signal strengths, with at least several orders of magnitude. This has a negative influence on the stability of the fit, since the parameter space for p_{g0} or p_{e0} also covers several orders of magnitude. Regression algorithms, based on χ^2 or log-likelihood estimators, are in those situations sensitive to the initial parameters of the fit. This can be countered by reducing the parameter space[7].

In case of the exponential function, a possible solution to the problem is the estimation of the initial parameter values by a prefit on a mapped parameter space. This prefit can be evaluated in ln-scale, i.e. the mapping is the ln function. In this representation the exponential function is given by a line. This reduces the allowed parameter space, which makes the fit converge more reliably. The estimated parameters of the prefit are subsequently used as initial parameters of the main fit. This avoids to my knowledge a bias on the initial parameters of the main fit, since the initial parameters of the prefit are unbiased.

7 In general, this introduces a bias on the fit result.

For the Gaussian distribution function, this approach is not suitable, since the Gaussian distribution does not reduce to a simple line in ln-scale. Instead, it is possible to chose the initial parameters of the fit directly from the data. The p_{g0} parameter directly corresponds to the maximum signal of the LDF. Therefore a good initial value for this parameter is the maximum signal in the data set. For the width p_{g2} a fixed starting value proves to be sufficient. In general, it is observed that radio signals at distances larger than 500 m from the shower axis are in the order of the noise level. Therefore the initial value of the width can be set to 250 m. The p_{g3} parameter corresponds to the power of the radio noise, which is measured by radio stations far away from the shower axis and can be estimated by averaging over the radio signals measured at distances larger than 500 m.

In figure 6.2.2 a simulated radio event with 24 stations is shown. It is fitted with an exponential parametrisation and for comparison with an Gaussian parametrisation (figure 6.2.3). The event shows a distinct flattening of the radio signals towards the shower axis. As expected, the Gaussian parametrisation describes the distribution of the radio signals better than the exponential parametrisation. A comparison of the quality (log-likelihood values) for both fits supports this[8]. The Gaussian fits has a quality value of about 318 million, which is one fifth of the quality value of the exponential fit (1.5 billion). However, not every radio event exhibits a flattening towards the shower axis and therefore the exponential parametrisation can yield comparable results as the Gaussian distribution for single events. Based on the data set E100, the quality of the exponential and Gaussian parametrisations are compared. Again the Gaussian parametrisation proves to be slightly better. Out of the 98 events approximately 1/3 (29 events), show a flattening towards the shower core.

For all 98 reconstructed events of this data set, a successful fit of the Gaussian parametrisation is possible. An exponential fit is only for 95 events successful. The average quality of the E100p simulation data set for the Gaussian LDF is 322 million, while it is significantly worse for the exponential LDF: 748 million.

The uncertainty of the fit parameters are estimated by bootstrapping (cf. section §A.2 in the appendix). The resulting uncertainties are included in figure 6.2.3 as error bands.

For a Gaussian parametrisation, it is now possible to investigate correlations between LDF parameters and the primary particle energy or X_{max}. For this task, radio simulations provide a good basis, as both quantities are initial parameters and therefore precisely known.

8 Smaller log-likelihood values denote a better fit of the model to the data and therefore a better quality.

6.3 ESTIMATION OF THE PRIMARY PARTICLE ENERGY BASED ON A GAUSSIAN PARAMETRISATION OF THE LDF

Any information on the primary particle energy available from the LDF should be accessible through the parameters of the LDF parametrisation. This is investigated in figure 6.3.1 , which shows the LDF parameter value against the primary particle energy of the simulation input.

A dependency of the p_{g0} parameter is observed and expected as this parameter corresponds to the maximum signal in the LDF, which correlates with the emitted radio signal strength. In addition the p_{g2} parameter, i.e. the width of the LDF, shows some weak dependency below 10^{18} eV but without a clear correlation to the primary particle energy.

To improve the correlation and reduce the scatter, a combination of the two parameters p_{g0} and p_{g2} is investigated by calculation of the integral over the LDF.

In figure 6.3.2 the energy dependency on the integral over the LDF is displayed. The uncertainty estimated for the LDF parameters by bootstrapping, are propagated by integrating the error bands (cf. A.1). In addition, the uncertainties resulting from the initial parameters of the data sets AvarCP and AvarCI and from shower-to-shower fluctuations are included. Gaussian error propagation provides the final uncertainty estimates for all subsequent figures (cf. section §A.1).

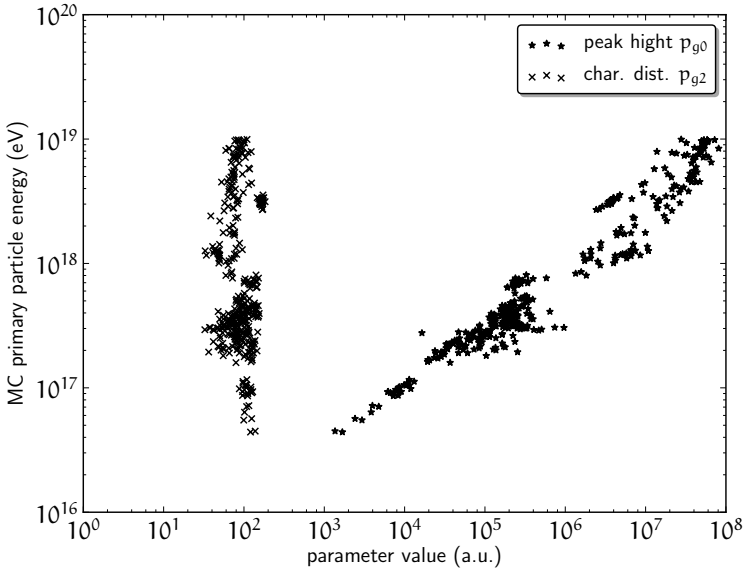
The integral over the around 0 centred LDF is given by the interval $[0, p_{g2}]$ and the following equation:

$$I_{LDF} = \int_0^{p_{g2}} dx \left[p_{g0} \cdot \exp\left((p_{g1}+x)^2 / 2 \cdot (p_{g2})^2 \right) + p_{g3} \right]$$

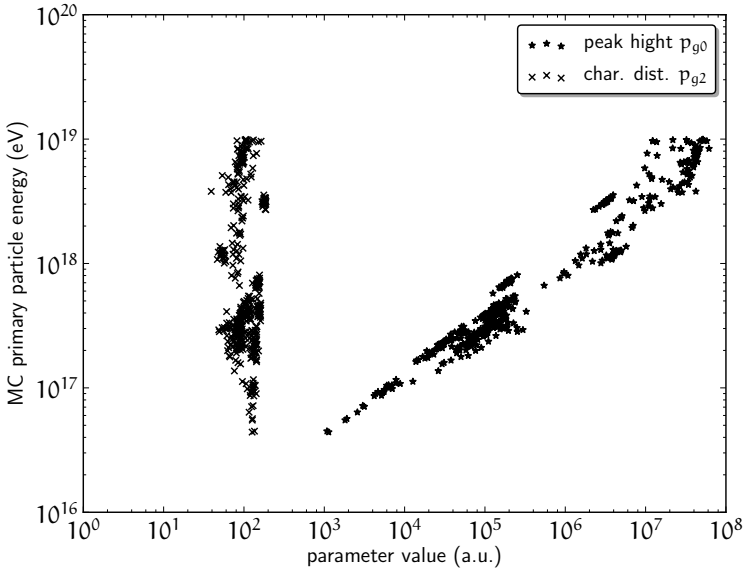
The limit of the integration is depending on the event as p_{g2} characterises the LDF. Henceforth p_{g2} is called the characteristic distance of the LDF.

The integral up to the characteristic distance improves the correlation and reduces the scatter as figure 6.3.2 proves. It is possible to describe the correlation by a linear function. This description is only limited by a feature around 10^{18} eV. At this energy, the proton and iron simulations start to exhibit a growing scatter and a kink occurs in the spectrum. Noteworthy, the scatter seems to increase at higher energies for iron simulations (3×10^{18} eV) compared to proton simulations (10^{18} eV).

The feature at around 10^{18} eV is also supported by figure 6.3.2, where a linear fit to the simulations is displayed over the full energy range. A single linear function is not suited to describe the full energy range for both data sets, as the simulations are systematically below the linear function for low energies. Therefore two separate correlation functions are fitted to the low and

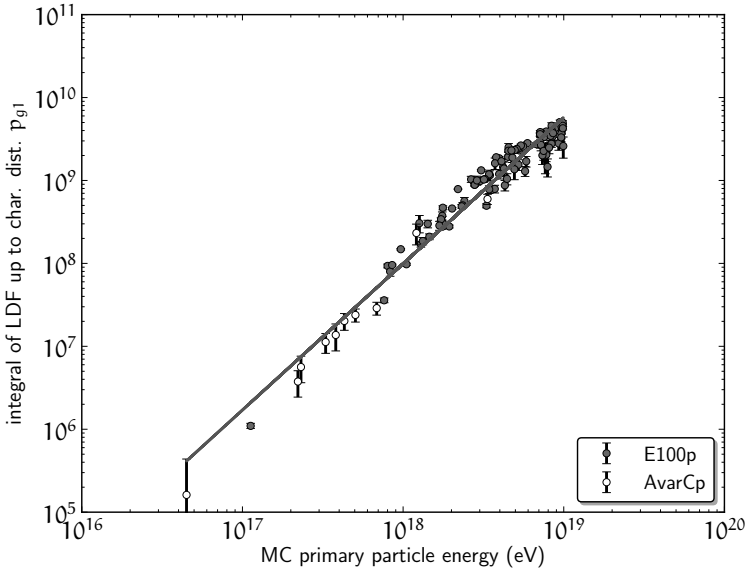


(a) data sets AvarCp (no averaging) and E100p

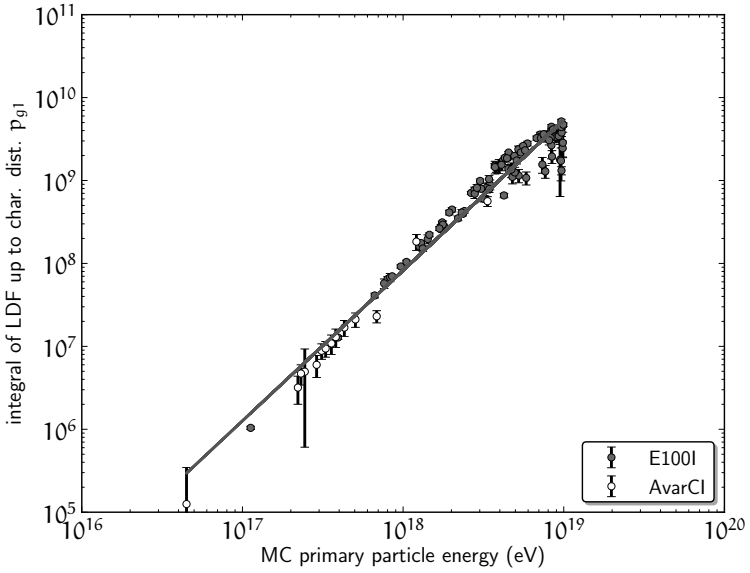


(b) data sets AvarI (no averaging) and E100I

Figure 6.3.1: Gaussian LDF parameters peak height, p_{g0} (stars) and characteristic distance, p_{g2} (crosses) vs. the input primary particle energy of the simulation for proton and iron simulations. Only the peak height shows a dependency on the primary particle energy. The p_{g3} parameter, i.e. the floor, does not show any correlation and is excluded from the plot for sake of transparency.

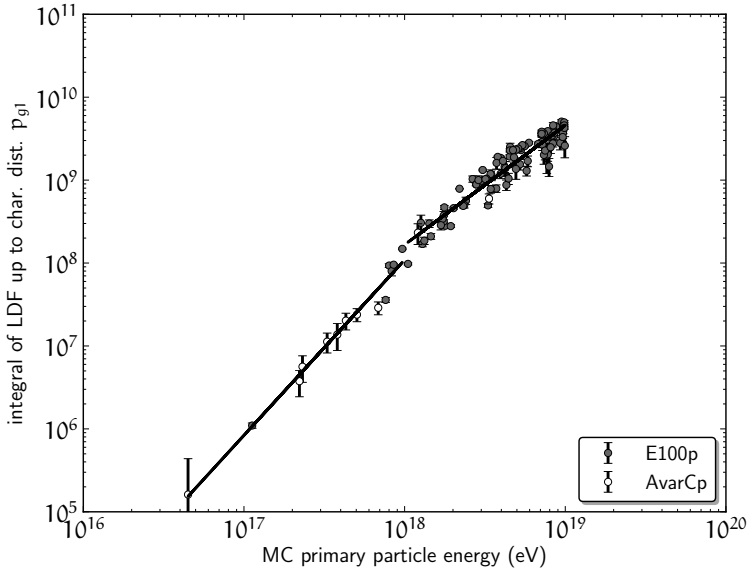


(a) data sets AvarCp and E100p

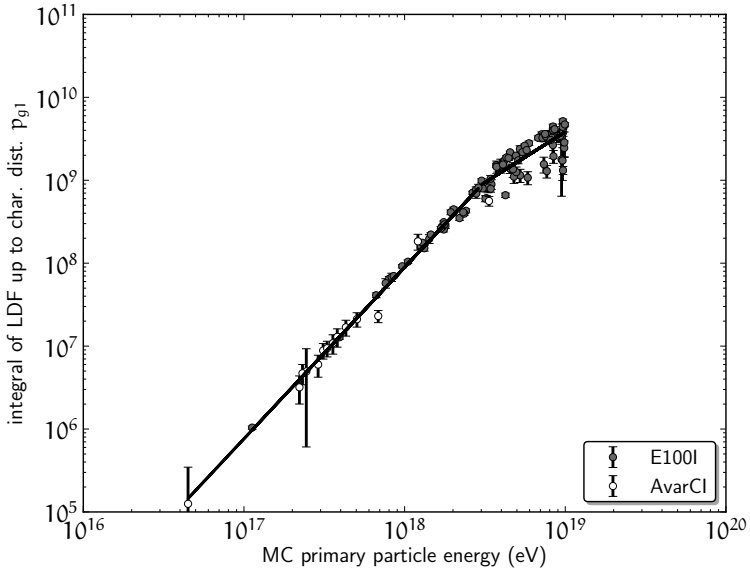


(b) data sets AvarCI and E100I

Figure 6.3.2: Primary particle energy vs. the integral over the LDF for proton and iron simulations. The errorbars include the LDF fit uncertainties and for the AvarCp/I data sets the uncertainties on the initial parameters of the simulations. A linear correlation function is assumed and fitted to the data.



(a) data sets AvarCp and E100p. The kink is assumed at 10^{18} eV.



(b) data sets AvarCI and E100I. The kink is assumed at 3×10^{18} eV.

Figure 6.3.3: Primary particle energy vs. the integral over the LDF for proton and iron simulations. The uncertainties include the LDF fit uncertainties and for AvarCp/I data stets the uncertainties on the initial parameters of the simulations. Two linear correlation functions are fitted separately to the low and high energy regime.

high energy range in figure 6.3.3. The low energy range is described by a linear correlation function with slope $a = 2.11$ for proton and $a = 2.07$ for iron simulations. The high energy range above 10^{18} eV exhibits a slope of 1.44 for proton and 1.22 for iron

simulations. A slope of about $a = 2$ as observed for the low energy regime, is expected as the primary particle energy is proportional to the $\vec{E}$-field (cf. Palmieri, 2012). In the chosen $|\vec{E}|^2$ representation of the LDF, a slope of $a = 2$ is therefore naturally expected[9]. The observed increase in scatter and decrease in the slope above some 10^{18} eV is discussed later.

Both the low and high energy correlation functions have the same analytical form[10]:

$$\log I_{LDF} = a \cdot \log E_{pp} + c$$

with:

I_{LDF}	the integral of the LDF
a	the slope of the fitted correlation function
c	the offset of the fitted correlation function
E_{pp}	the primary particle energy

The equation for the correlation line can be inverted to calculate the energy for a given integral value I_{LDF} over the LDF:

$$\begin{aligned} \frac{\log I_{LDF}}{a} - \frac{c}{a} &= \log \frac{E_{pp}}{eV} \\ \Rightarrow 10^{\log I_{LDF} - c/a} &= E_{pp} \end{aligned} \tag{6.3.1}$$

This provides a calibration function, which can be used to reconstruct the primary particle energy from the integral over the LDF. The parameters a and c are given by the linear correlation fits for the high and low energy regime:

	proton	iron
c (high)	−17.7	−13.6
a (high)	1.44	1.22
c (low)	−30.0	−29.3
a (low)	2.11	2.07

Based on the two calibration functions, it is also possible to estimate the uncertainty of the energy reconstruction for the simulation data sets. Therefore the standard deviation of the absolute differences between the estimated energy on basis of equation (6.3.1) and the input energy of the radio simulation is calculated. The uncertainty of the energy estimation is then derived to 13 % for the low energy regime and 21 % for the high

9 Note that all figures are given in log scale as also the correlation function is fitted in log scale. Accordingly a quadratic correlation of fundamental quantities is expressed by a linear function with slope 2.

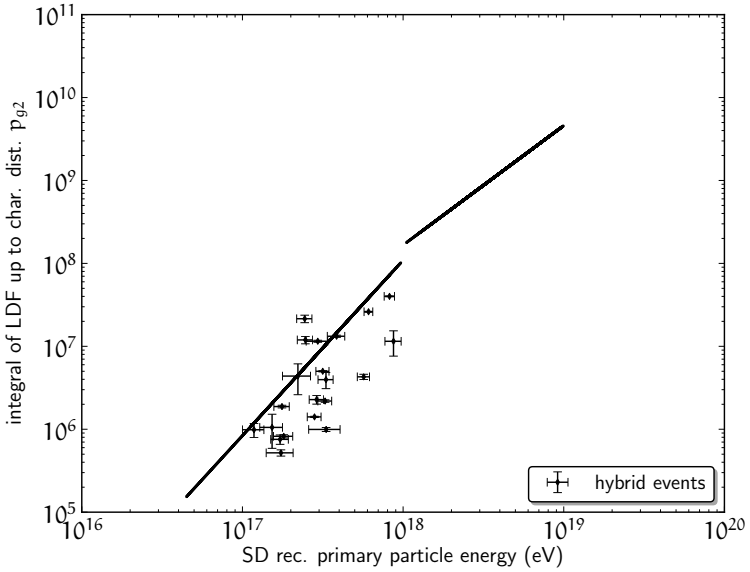
10 log indicates the logarithm to the base 10 in this case, as quadratic function corresponds to a line with slope 2 in this representation.

energy regime in case of proton simulations. For iron simulations similar uncertainties are estimated with 8 % for the low energy range. In the high energy range, the large scatter increases likewise the uncertainty of the energy estimation to 34 %. The low energy results are compatible with the results of the analysis presented by Glaser [2012] that covers only energies below 10^{18} eV.

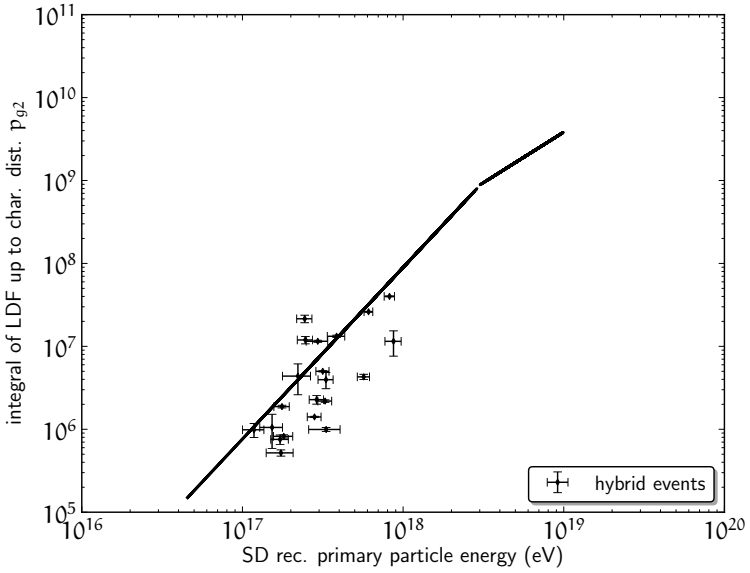
On basis of the available measured data sets, the energy estimation is tested. In figure 6.3.4 the measured events of the data set Acoinc are displayed alongside the two linear fits for the low and high energy regime.

Not all measured radio events are described by the correlation line within their uncertainties. About half (10) of the 21 events show a large difference and tend to underestimate the energy reconstructed by SD. A possible explanation is the noise level, as Offline does not include a noise simulation for radio simulations at the moment. As the ambient noise and the radio signal from cosmic ray air showers are both electro-magnetic waves, constructive and destructive interference can occur between them. This can unpredictably increase or decrease the measured radio signal. This effect is limited to the strength of the radio noise. Therefore the influence of the ambient noise level reduces, as the radio signal strength increases with the energy of the primary particle. The noise level has even more importance, as the applied SNR cut on the measured radio events is low if for example compared to the BLS study in chapter 5. With the applied SNR cut of $S^2/N^2 = 6.5$ the signal is in average only a factor of ~ 2.5 stronger than the noise. A high SNR cut is in principle not necessary for identification of the radio signal, as the available SD measurement of the air shower can be used to separate accidental radio noise signals from cosmic ray signals by comparison of the incident direction and the arrival time of the radio signal with the SD reconstruction. On the downside, the influence of the radio noise on the measured radio signal is increased and therefore contributes to the observed difference between the energy estimator and the SD reconstructed energy.

The scatter of the energy estimator is further increased by the observed difference between radio simulations and radio measurements as described in chapter 5, which is in average a factor of two. The CoREAS radio simulations furthermore tend towards underestimating the measured radio signal. Accordingly at least a factor four difference between the energy estimator and the SD reconstructed energy is expected, as the squared $\vec{E}$-field is used for the LDF. To correct for this effect, the SD reconstructed energy is divided by a factor of four and compared with the estimated energy. This results in an energy resolution of 16 %, which is



(a) correlation line, as derived for proton primary particles



(b) correlation line, as derived for iron primary particles

Figure 6.3.4: AERA measured radio events of the data set Acoinc. The energy of the primary particle is estimated by integration over the LDF and compared to the SD reconstructed energy. The displayed uncertainties cover the SD uncertainties on the energy reconstruction and the uncertainties of the LDF fit and integration.

slightly worse than the estimated energy resolution derived only from radio simulations.

Altogether, the influence of the radio noise on the measured radio events has to be investigated in more detail together with radio simulations, as soon as an ambient noise simulation for radio simulations is available.

6.3.1 *The feature around* 10^{18} eV

The kink around 10^{18} eV can be related to numerous sources, such as technical, statistical or physical.

First of all, a technical source for the kink is disfavoured, as all data sets are processed with the same algorithm, independent of the energy range. But a bias of the data sets used for deriving the correlation functions cannot be excluded. As the low energy range is dominantly populated by events from the AvarCp/I data sets and the high energy range by events from the E100p/I data sets, it is well possible that the data sets themselves cause the kink. Since the data sets only differ in their incident direction distribution, the kink would in this case be related to the incident direction of the air shower. This is however disfavoured, as all events of the E100p/I data sets below 10^{18} eV, are in good agreement with the AvarCp/I events. Furthermore, one naturally expects an increased scatter for a influence of the incident direction, as the incident direction and cosmic ray energy are uncorrelated. Therefore the incident direction of the radio events is it is disfavoured as the source of the kink. Nevertheless other unknown differences between the data sets cannot be ruled out as the source of the kink.

Another observation, however, makes a bias of the data sets seem unlikely. For iron simulations, a shift of the kink towards higher energies can be observed. Due to the shift of the kink, it is also possible to fit two correlation lines for only the E100I data set. The result of the fits is displayed in 6.3.5. Again, the slope of the low energy fit a $= 2.06$ is in good agreement with the combined AvarCI, E100I and the proton simulation results. The high energy fit gives a slope of a $= 1.19$, which is within the expectation from the other data sets. The shift of the kink furthermore indicates a sensitivity of the LDF on the primary particle type and could provide future access to the composition of cosmic rays.

Aside technical and data set related issues, also underlying physics can be responsible for the kink. As a decrease in the slope is observed, it is expected for the physical effect capable of describing the kink to result in a reduced strength of the emitted radio signal at high energies. On the face of it, this seems to be a contradiction, as the amount of charged particles in the air

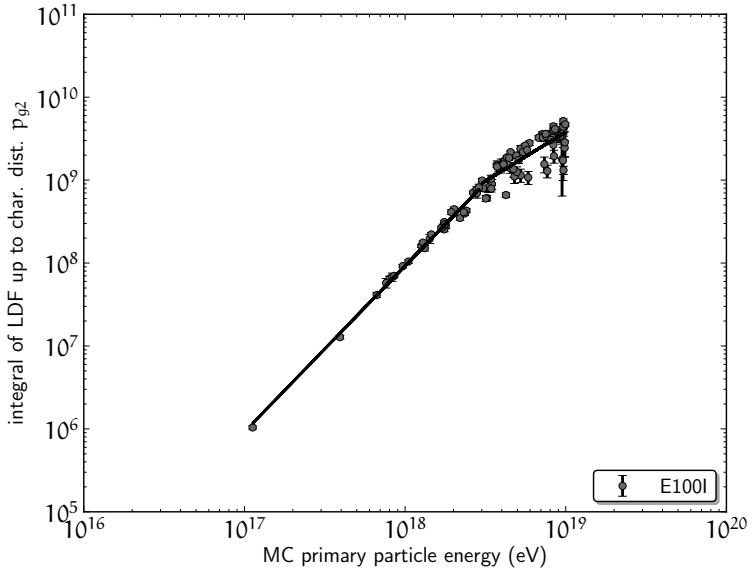


Figure 6.3.5: Data set E100I on basis of iron simulations. The energy estimator on basis of the LDF is displayed vs. the input primary particle energy of the simulation. Two correlation functions are fitted separately below and above the kink at 3×10^{18} eV. The kink is given by the increasing scatter.

shower rises with the primary particle energy and therefore also the strength of the radio emission. Accordingly the assumption is made that the air shower completely evolves and all charged particles contribute to the radio emission.

This assumption does not necessarily hold up at high energies and for detectors at high altitudes as it is the case for the Pierre Auger Observatory. It is well known that in average the shower maximum X_{max} is reached deeper in the atmosphere as the energy increases. For detectors at high altitudes, as it is the case for the Pierre Auger Observatory, it henceforth occurs at high energies that the shower maximum is reached close to the ground. As the ground absorbs the charged particles of the dying air shower cascade, also the strength of the radio signal emitted by the air shower can be expected. Experiments at different altitudes could provide additional insight, as with increasing atmosphere, i.e. decreasing altitude, the kink is then expected to shift to higher energies.

The already in simulations observed shift of the kink towards higher energies for iron simulations does agree well with this interpretation. Heavy nuclei like iron, tend to interact higher in the atmosphere than lighter nuclei due to their higher cross-section. Therefore the shower maximum of iron air showers is expected to reach the ground at higher energies. This results in a shift of the kink towards higher energies for heavier nuclei as

it is observed. The kink could even provide insights in the interplay of different radio emission mechanisms, as they are also depending on the shower evolution and should be influenced by a truncated shower development.

Another physical interpretation of the kink, is a increasing decoherence of the radio signal at these energies. A linear correlation (slope $a = 1$) between the $\vec{E}$-field and the primary particle energy is indeed expected for a de-coherent emission, while a quadratic dependency (slope $a = 2$) is expected for coherent emission. The observed slope for proton simulations of $a = 1.44$ above the kink can thereby also be an indication for an increased de-coherence of the radio signal. In this case, the kink should be independent of the observer altitude, and therefore can be investigated with experiments on differing altitudes.

In any case, final conclusions cannot be drawn at the moment, as the available data sets do not yet provide the necessary statistic for further investigations. Especially the currently available measured data sets have insufficient statistic in the energy range between 10^{18} eV to 10^{19} eV due to the reduced flux of high energy cosmic rays. This will change in the future and it will be possible to either confirm or reject the existence of the kink on basis of measured data. At the moment also a discrimination between a physical source for the kink and a bias of the data sets cannot be done, even as the observations disfavour a bias of the data. Cross-calibrations of AERA with the FD of the Pierre Auger Observatory are one possible approach to investigate this further in the future.

6.3.2 X_{max} *dependency of the LDF*

In principle, alongside the energy dependency of the LDF, also a dependency on X_{max} is expected from earlier studies of the LOPES experiment (cf. Palmieri [2012]). This is of special interest as it provides a reconstruction method for X_{max} from radio measurements alone. Therefore correlations of the LDF parameters with X_{max} are investigated and displayed in figure 6.3.6. In contrast to the primary particle energy, no significant correlation between any of the parameters and X_{max} can be observed. Only a weak dependency is present for the peak height, p_{g0}. This is insufficient for a reconstruction of the X_{max} value on basis of single LDF parameters, as the scatter is too large. This can be further investigated in the future on basis of other LDF features, like integrals or derivatives.

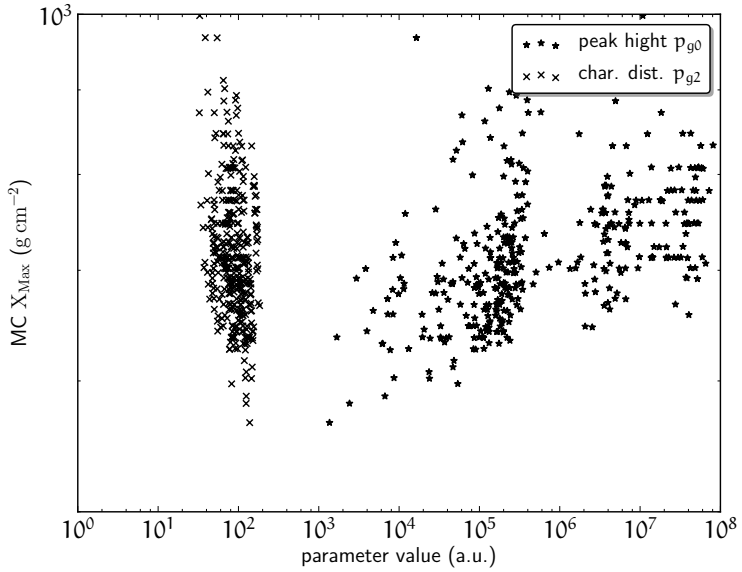


Figure 6.3.6: Shower maximum X_{max} vs. the peak height and characteristic distance of the LDF. No strong correlation between X_{max} and both parameters can be observed.

6.4 INVESTIGATION AND RECONSTRUCTION OF THE RADIO SHOWER CORE

Aside the possibility to derive energy estimators, the LDF provides also an approach to reconstruct the shower core of the radio emission on the ground. This radio core position does not necessarily have to be the same like the particle shower core. This was demonstrated only recently by the CODALEMA experiment (cf. the CODALEMA collaboration [2011]), which presented a radio core reconstruction based on measurements with the assumption of an exponential LDF. The observed core shift towards the particle shower core is observed to be depending on the geomagnetic angle (c.f. the CODALEMA collaboration [2011] and figure 6.4.2). The method to estimate the core position from the radio LDF can also be applied to AERA measurements and simulations on basis of a Gaussian LDF parametrisation.

In the one-dimensional projection on the shower axis, the position of the shower axis is always at 0 m. This is implicitly connected to the core position, as the distance to the shower axis depends on the core position at the ground. Under the assumption that the chosen parametrisation describes the lateral distribution of radio signals, the radio core can be reconstructed by adapting the data to the model.

For this reconstruction, the shower core is varied on a 200 m $\times$ 200 m rectangular grid with a step size of 5 m around the initial core position. The initial core position can be given by

the input core position of the radio simulations, which coincides with the particle core position. Subsequently, the LDF parametrisation is fitted for each core position on the grid. Based on a predefined criteria, which is in this case the fit with the best quality, a core position is selected.

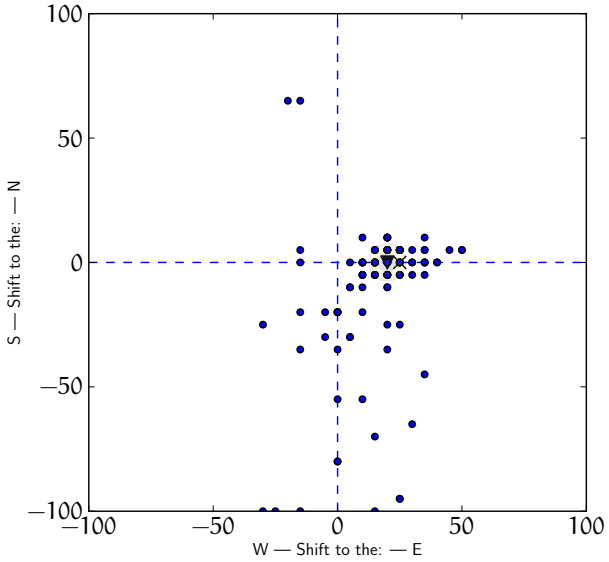
At this point, it has to be stressed that this approach is not model independent. The grid scan is a fit of the data to the model (here the parametrisation), since the "best" fit is chosen. This is not the case for the before discussed LDF fits, where the model parameters are fitted to the data. This introduces a bias to the resulting core position, as it can change depending on the LDF parametrisation in use. This has to be kept in mind, when treating with the resulting core position.

For the above presented simulation data sets E100p and E100I the radio core position is estimated by core variation. In figure 6.4.1 the difference in metres between the reconstructed radio core position and the particle core position of the radio simulation is shown. Two distinct features are present:

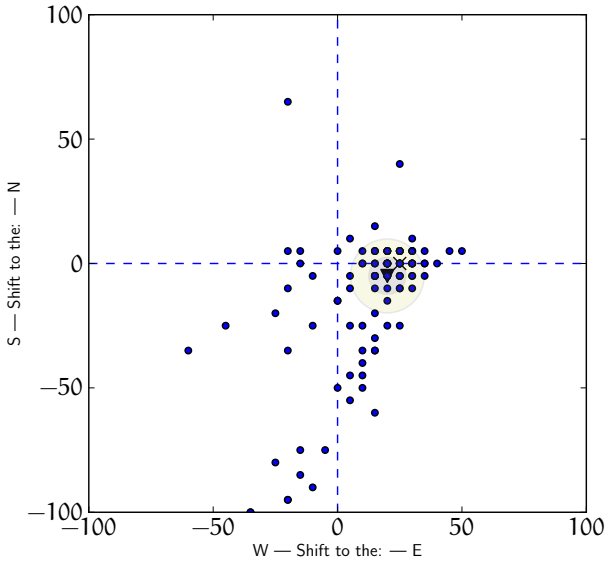
First the median core position of all simulations is shifted towards the east with respect to the initial core position[11]. The uncertainties for this core shift are estimated by the MAD estimator. The median shift itself is 20.0 ± 11 m to the east and 0 ± 19 m to the north. Iron shower show a similar behaviour, with a median core shift of 20.0 ± 14 m to the east and 5 ± 20 m to the south. Both simulation data sets are therefore compatible with the results presented by the CODALEMA collaboration [2011] (~25 m to the east). The simulations accordingly support the CODALEMA observations as displayed in 6.4.2. A possible explanation for the core shift can be given by the superposition of the two main radio emission mechanisms, charge excess radiation and geomagnetic radiation. The two emissions show a distinct difference in the polarisation of the radio signal at the ground. While the stronger geomagnetic radiation results in a eat-west polarised $\vec{E}$-field, the charge excess radiation shows a radial to the shower axis inside facing polarised $\vec{E}$-field at the ground. If both emissions are superimposed constructive and destructive interference of the $\vec{E}$-fields occurs. This results in an effective shift of the total $\vec{E}$-field towards the east compared to the core of the particle shower as schematically shown in figure 6.4.3. This core shift is in a sense virtual and only an expression for a rotational asymmetry around the shower axis in the one-dimensional projection.

Therefore it is either necessary, to describe the lateral distribution of radio signals with a two-dimensional LDF parametrisation or to derive a model, which introduces corrections for the

11 Instead of the mean core position of the simulations, the median core position is chosen, since the distance distribution exhibits a heavy tail (c.f. Appendix 3)

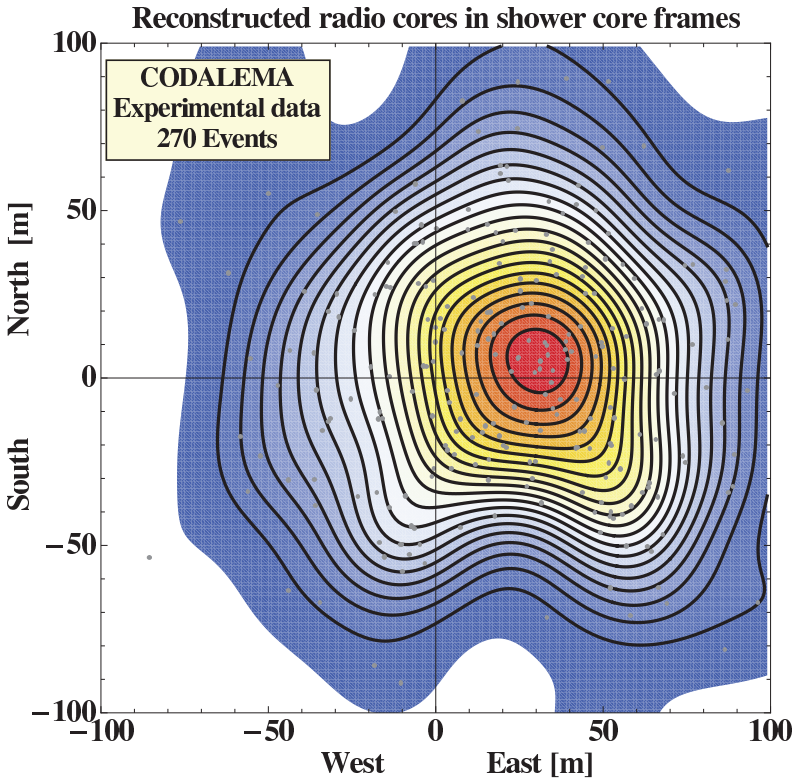


(a) proton simulation data sets E100p and AvarCp

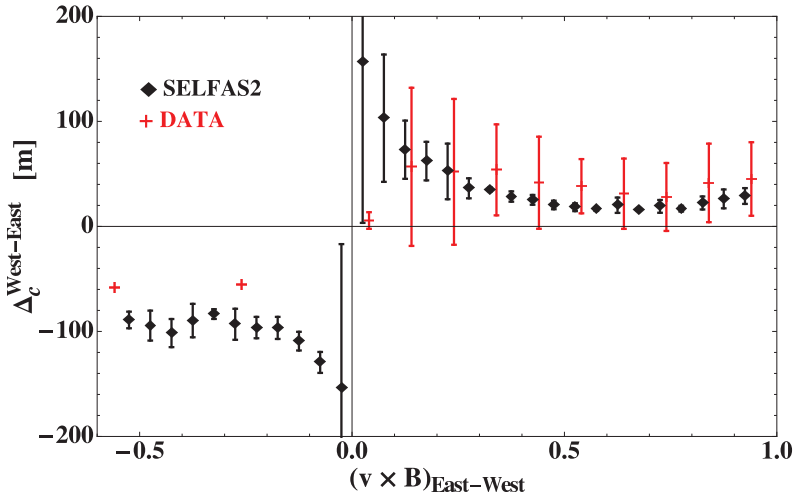


(b) iron simulation data sets E100I and AvarCI

Figure 6.4.1: Difference between the simulation input core (particle shower core) at 0 and the radio reconstructed core by a grid scan (data sets AvarCp, AvarCI, E100p and E100I). The black cross denotes the expected east-west core shift reported by the CODALEMA experiment. The median of the east-west and north-south core shifts for this data set are given by the black triangle.



(a) core shift as reconstructed by CODALEMA for 270 measured events.



(b) Predicted core shift by the CODALEMA experiment for SELFAS2 simulations and measurements.

Figure 6.4.2: Core shift as reported by the CODALEMA collaboration and published in the CODALEMA collaboration, 2011. The core shift is about 25 m to the east with a small shift to the north. The north shift is a result of the geomagnetic emission, which is stronger for air showers arriving orthogonal to the geomagnetic field. Therefore air showers from the north emit a stronger radio signal than radio signals from the south on the northern hemisphere. For the Pierre Auger Observatory on the southern hemisphere, the situation is inverted. Accordingly a core shift to the east and south is expected for the southern hemisphere.

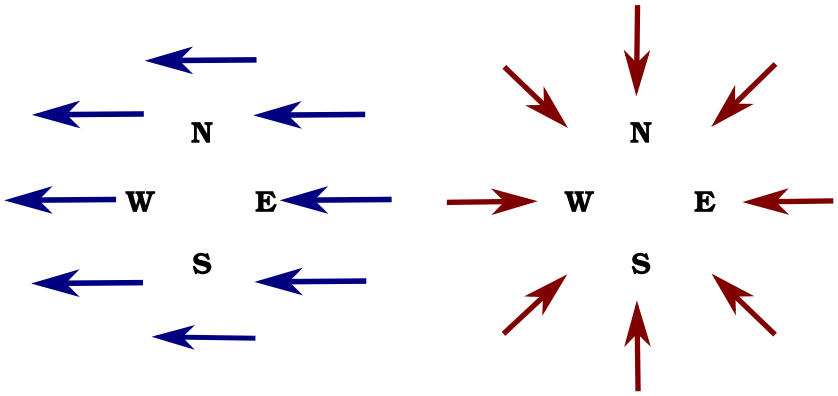


Figure 6.4.3: $\vec{E}$-field as observed at the ground for the two main emission mechanisms geomagnetic emission (l.h.s.) and charge excess (r.h.s.). The superposition of both results in an effective core shift to the east, since the partial radio emissions experience constructive interference in the east and destructive interference in the west.

asymmetry. Since an increase of the dimensions automatically increases the amount of free parameters, e.g. for the Gaussian parametrisation from four to six, a comparable amount of radio stations with reconstructed radio signals is required for the fit. This approach is at the moment not feasible at AERA, since only a small amount of measured radio events fulfil this requirement.

6.5 SUMMARY

The introduced Gaussian LDF has shown to be superior to the common exponential LDF both overall and in describing single events. On basis of the Gaussian LDF I was able to derive an energy reconstruction for radio measurements with AERA as demonstrated for simulations and measurements. The simulation results for the low energy regime are also in good agreement with similar results of the LOPES experiment, as they predict a linear correlation between the energy and the $\vec{E}$-field strength. The correlation function derived from simulations was tested on hybrid AERA-SD measurements. About half of the investigated hybrid events agreed well with the derived correlation within their uncertainties. The other half exhibits a larger difference between the estimated energy and the SD reconstructed energy than expected and tend towards larger energies than estimated. This observation is partially related to the observed factor two difference between simulated and measured $\vec{E}$-fields (cf. chapter 5). In addition, the ambient noise level is not yet included into the radio simulations, which also contributes to the difference between the simulation based correlation function and measured data. Fi-

nally, a low SNR cut is applied, which results in an increased influence of the radio noise on the measured radio signal.

Therefore it is necessary to investigate the radio noise and its influence on the radio signal in more detail in the future. In this context, the introduction of a radio noise simulation for radio simulations is of importance.

The presented energy estimator furthermore exhibits a feature at 10^{18} eV, which is currently not accessible with measurements due to low statistics. This kink like feature is best described by two separate linear correlations, one for the low and one for the high energy regime. At the moment, the source of the kink is unclear, as both a bias from the applied data sets or a physical explanation are possible. The comparison of proton and iron simulations supports however a physical explanation. One physical explanation here discussed links the feature to the air shower development. A kink can be expected for truncated air shower developments, if the distance between the shower maximum and the ground is indeed the source of the feature. Experiments at low altitudes, e.g. sea level, should in this case observe a shift of the feature towards higher energies as the shower maximum has a greater distance to the ground even at higher energies.

Another physical explanation can be an increased de-coherence of the radio signal at higher energies, which also results in a weaker correlation.

However, a final conclusion on this topic is only possible on basis of large statistic measured and simulated data sets in the future.

In addition to an energy reconstruction, a reconstruction of the radio core is now available for the first time in the context of AERA. The results indicate a shift towards the east for the used radio simulations, which is in good agreement with a similar analysis recently presented by the CODALEMA experiment.

7

SUMMARY AND OUTLOOK

As a first topic of this work a new method for the reconstruction of the complete $\vec{E}$-field of the radio signal from a partial measurement is introduced. This makes AERA the first radio experiment for cosmic ray air shower, capable of a reconstruction of the complete three-dimensional $\vec{E}$-field. The $\vec{E}$-field provides the complete information on the radio signal, emitted by the air shower and is therefore the basis for high-level analyses, like LDF or polarisation studies.

Since several years, radio simulations play an important role in understanding the radio emission from cosmic rays and development of new analysis methods. Therefore a comparison between simulations and measurements is important, especially for the new reconstruction of the $\vec{E}$-field vector. I present two such comparisons in this work, which are also a document for the development of radio simulation codes in the past two years.

The first comparison is based on measurements of the BLS, a prototype station for AERA. The results indicated for the two well recognised radio simulation codes MGMR and REAS$_3$ an agreement between the simulated and measured electric field up to a factor of four. In a more recent comparison on basis of first AERA measurements I could also show, that the difference reduced to about a factor of two for the most recent radio simulation codes EVA and COREAS.

Based on the $\vec{E}$-field vector and the gathered knowledge on radio simulations, I derived a new parametrisation of the lateral distribution function. This became necessary, as recent measurements of both the LOPES and LOFAR experiment observed radio events are not described by an exponential LDF as the radio signal strength levels near the shower axis (also known as flattening events). This observation is supported by recent COREAS simulations. Therefore a parametrisation of the LDF on basis of a Gaussian distribution function is derived, which is capable of describing most of the measured and simulated events including flattening radio events.

With this new description of the LDF, it is possible to investigate a primary energy estimator for radio measured cosmic ray induced air showers. This is an important first step towards a reliable energy reconstruction of the primary particle energy of the cosmic ray air shower. I demonstrated on basis of COREAS radio simulations, that the integral over the LDF up to a characteristic distance provides a reliable energy estimator. I estimated

the uncertainty on this energy estimator on basis of proton radio simulations to be 13 % below 10^{18} eV.

The energy estimator also exhibits a kink around 10^{18} eV. It is shown, that the energy depends quadratic on the proton primary particle energy of the air shower below the feature. Above 10^{18} eV the correlation between the energy estimator and the proton primary particle energy has a slope of 1.44 with an energy resolution of 21 %. An investigation of iron radio simulations, furthermore indicates a shift of the feature towards higher energies. This disfavours a bias of the used data sets.

Therefore, two possible physical effects are discussed. The first effect is related to the shower evolution and depends on the altitude of the experiment. In this case, the feature would not be observed by low altitude experiments. The second effect relates to the coherence of the radio signal. From theory it is known, that a slope of 2 indicates a coherent radio emission, while a slope of 1 denotes a de-coherent emission. Therefore the feature can indicate a de-coherence of the radio signal at high primary cosmic ray particle energies. Based on a larger data set measured with AERA and including more high energetic events, it will be possible to investigate the kink in more detail in the future and to determine its origin.

However, the results below 10^{18} eV agree with recent results of the LOPES experiment and an independent study of AERA measurements.

Alongside the derivation of an energy estimator, a model dependent method to reconstruct the core position of the radio emission is presented. The reconstruction heavily depends on the applied LDF parametrisation and provides the core position on basis of a grid scan. The result of a 20.0 ± 11 m core shift to the east with respect to the particle shower on basis of proton COREAS simulations is in good agreement with a similar analysis of the CODALEMA collaboration. Furthermore, this result confirms today's knowledge on radio emission mechanisms, as it indicates a dominant role of geomagnetic emission.

In addition, the investigations demonstrate, that a one-dimensional LDF is insufficient in describing azimuthal asymmetries in the lateral distribution of the radio signals. This suggest the derivation of a two-dimensional LDF in the future.

Summarizing, I improved with the Gaussian parametrisation of the LDF the description of flattening events, introduced a new method for radio measurements to estimate the energy of cosmic rays and included a reconstruction method for the radio core position for AERA.

Part III

APPENDIX

A

ERROR PROPAGATION AND UNCERTAINTY ESTIMATION

A.1 UNCERTAINTY ESTIMATION FOR LDF PARAMETERS

After fitting the LDF the uncertainties on the resulting parameter values have to be estimated. This is done by bootstrapping. Therefore the residuals of the data points towards the fitted function are calculated:

$$r_i = |f(x_i) - y_i|$$

Out of the set of n residuals, n new ones are randomly chosen (flat distributed) R_i. N.B. that single residuals can occur more than once in R while others do not occur. On basis of this new test data set R, the fit is repeated and the resulting parameter values saved. This is repeated 500 times. Out of the set of 500 times fitted parameters, the uncertainties on the original parameter value can be calculated by the standard deviation. The result is then included as 86 % error bands into the LDF fits.

A.2 UNCERTAINTY ESTIMATION ON INTEGRATED QUANTITIES

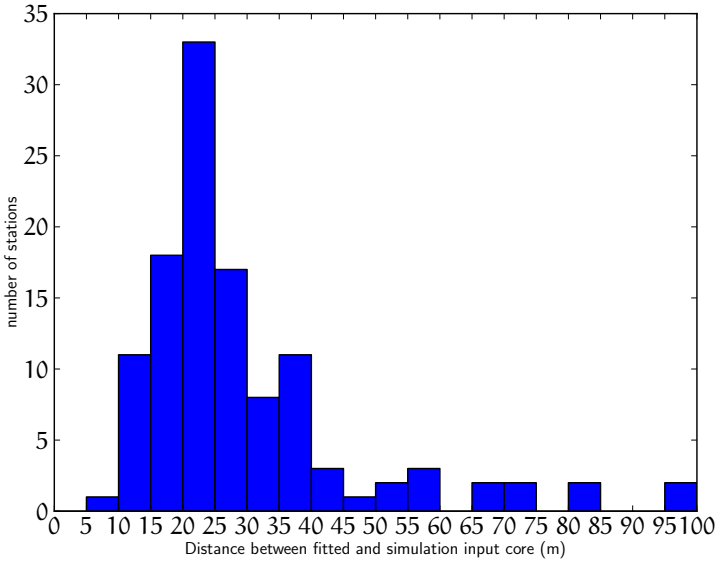
The uncertainty on integrated quantities, e.g. the energy estimator, are based on the uncertainty of the underlying function. As discussed in section §A.1, the uncertainties on the integration function are known. Each, the upper and the lower uncertainty estimates $\left(\int \sigma_{low}, \int \sigma_{high}\right)$ are integrated and the uncertainty calculated by:

$$\Sigma = \left| \int f(x) - \int \sigma_i \right|$$

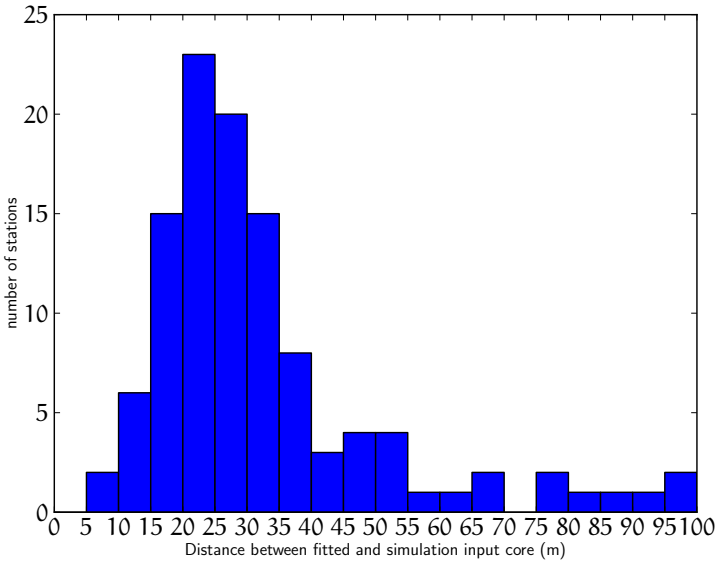
The numerical uncertainties of the integrations are neglected in this approach.

A.3 MEDIAN CORE SHIFT CALCULATION

In figure A.3.1 the direction independent relative core shift of the radio reconstructed core towards the particle shower core is given. The large tail of events with a core shift larger than 50 m influences the mean core position. Therefore the median

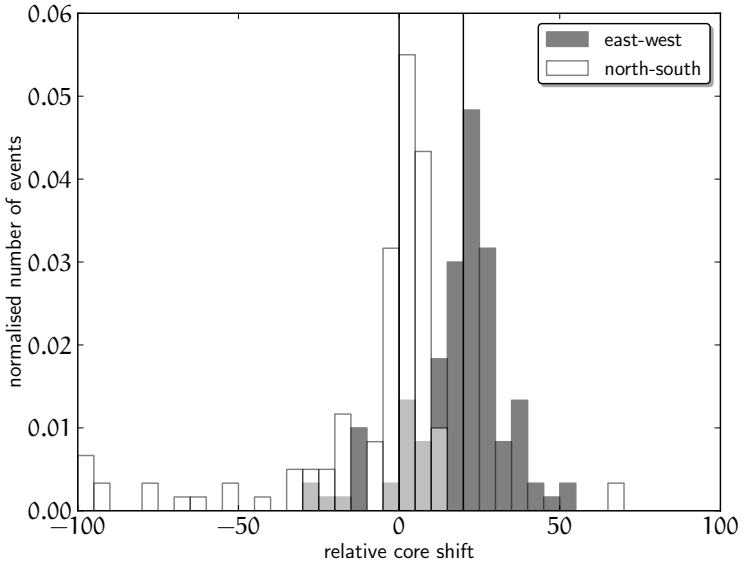


(a) proton simulation data sets E100p and AvarCp

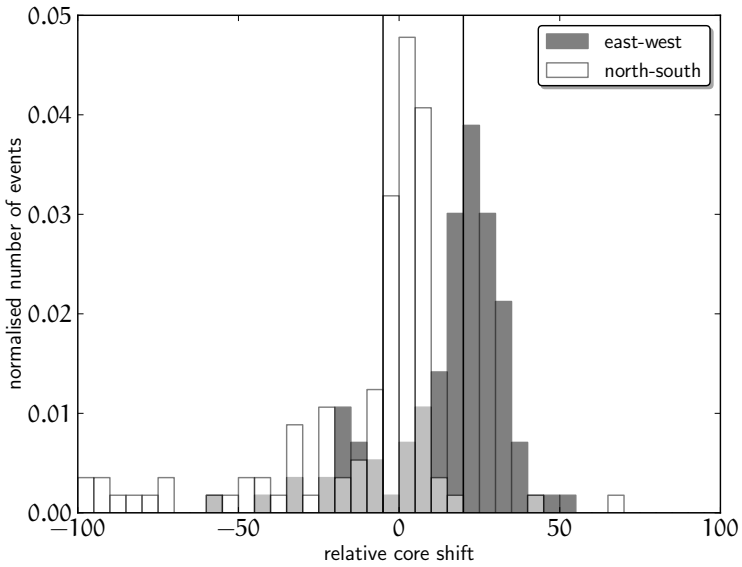


(b) iton simulation data sets E100I and AvarCI

Figure A.3.1: Histogram of the relative distance (m) between the particle and the radio reconstructed shower core. Two distinct features are present. First, most of the events experience an direction independent core shift of about 25 m. Second, several events exhibit large core shifts above 50 m.



(a) proton simulation data sets E100p and AvarCp



(b) iron simulation data sets E100I and AvarCI

Figure A.3.2: Histogram of the relative distance (m) between the particle and the radio reconstructed shower core for the east-west (black) and north-south (white) direction. The corresponding median is given by the vertical lines.

core position is calculated. In figure A.3.2 the distribution of the east-west and north-south core shift in m is given. The median describes well the centre of the distributions. Therefore the core shift of the sample is calculated by median instead of the mean.

"Computer. What is the nature of the universe?"–**Beverly Crusher**
"The universe is a spheroid region 705 meters in diameter."–**Computer**

— Star Trek: The Next Generation, *(Remember Me)*

ACKNOWLEDGMENTS

Es liegt in der Natur einer Doktorarbeit, dass meist viel mehr Personen Anteil an ihrem Gelingen haben als auf den ersten Blick ersichtlich ist. Mein Dank gilt daher auch einer ganzen Reihe Menschen, deren Namen ich an dieser Stelle leider nicht alle nennen kann.

Besonders hervorheben möchte ich deshalb nur meinen Doktorvater, Professor Dr. Johannes Blümer, welcher mir diese Arbeit ermöglichte und Prof. Dr. Günter Quast, der bereitwillig das Koreferat für diese Arbeit übernommen hat. Fernerhin gilt mein Dank meinem Betreuer Herrn Dr. Tim Huege und zu gleichen Teilen Herrn Dr. Andreas Haungs, welche beide viel zum Gelingen dieser Arbeit beigetragen haben. Ich möchte auch nicht meine Kollegen aus der Radio-Gruppe und dem IK vergessen, denen ich für ihre vielfältige Unterstützung danke. Die anregenden Diskussionen bei einer Tasse Kaffee und die vielen erheiternden und freundlichen Stunden, die wir zusammen verbrachten, haben mich sehr unterstützt.

Privat danke ich meinen vielen Freunden, denen ich sicher das eine oder andere Mal nicht so gut gelaunt begegnete und die mir große Unterstützung und den wichtigen Ausgleich von der Arbeit bescherten.

Zuletzt möchte ich mich besonders herzlich bei meiner Familie bedanken, die nicht nur während meiner Doktorarbeit sondern auch während meines Studiums und all die Jahre zuvor immer für mich da waren und mich auf meinem Lebensweg nach Kräften unterstützten. Insbesondere meiner Mutter Gudrun und meiner Großmutter Helga danke ich für alles.

EIDESSTATTLICHE ERKLÄRUNG

Hiermit versichere ich an Eides statt, daß diese Arbeit selbständig und ohne unerlaubte fremde Hilfe angefertigt und keine anderen, als die von mir angegebenen Schriften und Hilfsmittel benutzt wurden.

Karlsruhe, 13. Februar 2013

Benjamin Lothar Fuchs

BIBLIOGRAPHY

A. M. van den Berg et al. - Pierre Auger Collaboration. First detection of radio signals from cosmic rays at the Pierre Auger Observatory., 2007. (Cited on page 29.)

A. Nelles et al. - Pierre Auger Collaboration. t.b.a. In *Proceedings of the ARENA conference, Erlangen*, 2012. (Cited on page 92.)

G. A. Askaryan. Excess negative charge of an electron-photon shower and its coherent radio emission. *Soviet Physics JETP*, 14:441, 1962. (Cited on page 26.)

Pablo M. Bauleo and Julio Rodriguez Martino. The dawn of the particle astronomy era in ultra-high-energy cosmic rays. *Nature*, 458(7240):847–851, Apr 2009. ISSN 0028-0836. doi: 10.1038/nature07948. URL http://dx.doi.org/10.1038/nature07948. (Cited on page 17.)

S. Buitink et al. Amplified radio emission from cosmic ray air showers in thunderstorms. *Astron.Astrophys.*, 467:385–394, 2007. doi: 10.1051/0004-6361:20066006. (Cited on page 68.)

D. Huber et al. - LOPES Collaboration. LOPES 3D reconfiguration and first measurements. In *International Cosmic Ray Conference*, volume 3 of *International Cosmic Ray Conference*, page 72, 2011. (Cited on pages 37 and 42.)

C. Deboor. *A Practical Guide to Splines*. Springer-Verlag Berlin and Heidelberg GmbH & Co. K, 1978. ISBN 3540903569. URL http://www.worldcat.org/isbn/3540903569. (Cited on page 52.)

Luke O'C. Drury. Origin of Cosmic Rays. 2012. (Cited on pages 15 and 16.)

Kenneth P. Esler. The einspline library webpage. webpage, 07 2012. URL http://einspline.sourceforge.net/. (Cited on page 52.)

F. G. Schröder et al. - LOPES Collaboration. Investigation of the Radio Wavefront of Air Showers with LOPES and REAS3. In *International Cosmic Ray Conference*, volume 3 of *International Cosmic Ray Conference*, page 64, 2011. (Cited on pages 41 and 60.)

E. Fermi. On the origin of the cosmic radiation. *Phys. Rev.*, 75:1169–1174, Apr 1949. doi: 10.1103/PhysRev.75.1169. URL

http://link.aps.org/doi/10.1103/PhysRev.75.1169. (Cited on page 15.)

S. Fliescher. private communication, 2009 - 2011. (Cited on pages 47 and 50.)

S. Fliescher, M. Erdmann, T. Hebbeker, S. Maurice, and K. Weidenhaupt. The antennas and preamplifiers with integrated filters for the auger engineering radio array (phase 1). Internal GapNote 2011-139. (Cited on page 49.)

Benjamin Fuchs, Daniël Fraenkel, Tim Huege, and Maximilien Melissas. A comparison of the reconstructed E-field from BLS radio data with simulations based on Offline. *Internal GapNote 2011-135*, 2011. (Cited on pages 58 and 70.)

C. Glaser. Energy Estimation of Cosmic Rays Measured with the Auger Engineering Radio Array. internal publication of the Pierre Auger Observatory (in preparation), 2012. (Cited on pages 91 and 108.)

Javier G. Gonzalez. The Offline Software of the Pierre Auger Observatory: Lessons Learned. 2012. (Cited on page 30.)

Kenneth Greisen. End to the cosmic ray spectrum? *Phys.Rev.Lett.*, 16:748–750, 1966. doi: 10.1103/PhysRevLett.16.748. (Cited on page 18.)

W. Heitler. Theory of meson production. *Rev. Mod. Phys.*, 21:113–121, Jan 1949. doi: 10.1103/RevModPhys.21.113. URL http://link.aps.org/doi/10.1103/RevModPhys.21.113. (Cited on pages 22 and 23.)

A M Hillas. The origin of ultra-high-energy cosmic rays. *Annual Review of Astronomy and Astrophysics*, 22 (1):425–444, 1984. doi: 10.1146/annurev.aa.22.090184.002233. URL http://www.annualreviews.org/doi/abs/10.1146/annurev.aa.22.090184.002233. (Cited on page 16.)

K. S. Hirata, T. Kajita, M. Koshiba, M. Nakahata, Y. Oyama, N. Sato, A. Suzuki, M. Takita, Y. Totsuka, T. Kifune, T. Suda, K. Takahashi, T. Tanimori, K. Miyano, M. Yamada, E. W. Beier, L. R. Feldscher, W. Frati, S. B. Kim, A. K. Mann, F. M. Newcomer, R. Van Berg, W. Zhang, and B. G. Cortez. Observation in the kamiokande-ii detector of the neutrino burst from supernova sn1987a. *Phys. Rev. D*, 38:448–458, Jul 1988. doi: 10.1103/PhysRevD.38.448. URL http://link.aps.org/doi/10.1103/PhysRevD.38.448. (Cited on page 19.)

Tim Huege and Heino Falcke. Radio emission from cosmic ray air showers: Coherent geosynchrotron radiation. *As-*

tron.Astrophys., 412:19–34, 2003. doi: 10.1051/0004-6361:20031422. (Cited on page 26.)

I. Allekotte et al. - Pierre Auger Collaboration. The surface detector system of the pierre auger observatory. *NIMA*, 586(3):409 – 420, 2008. ISSN 0168-9002. doi: 10.1016/j.nima.2007.12.016. (Cited on page 26.)

J. Abraham et al. - Pierre Auger Collaboration. The Fluorescence Detector of the Pierre Auger Observatory. *NIMA*, 620(2-3):227 – 251, 2010. ISSN 0168-9002. doi: 10.1016/j.nima.2010.04.023. (Cited on page 26.)

J. Allen et al. - Pierre Auger Collaboration. The Pierre Auger Observatory offline software. *J.Phys.Conf.Ser.*, 119:032002, 2008. doi: 10.1088/1742-6596/119/3/032002. (Cited on page 30.)

F. D. Kahn and I. Lerche. Radiation from Cosmic Ray Air Showers. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 289(1417):206–213, 1966. doi: 10.1098/rspa.1966.0007. URL http://rspa.royalsocietypublishing.org/content/289/1417/206.abstract. (Cited on page 25.)

Karl-Heinz Kampert and Michael Unger. Measurements of the Cosmic Ray Composition with Air Shower Experiments. *Astropart.Phys.*, 35:660–678, 2012. doi: 10.1016/j.astropartphys.2012.02.004. (Cited on page 22.)

Kumiko Kotera and Angela V. Olinto. The Astrophysics of Ultrahigh Energy Cosmic Rays. 2011. (Cited on page 20.)

Daniel Kruppke-Hansen. Extending the Pierre Auger Observatory to explore the ankle of the cosmic ray spectrum. *Advances in Space Research* , 2011. ISSN 0273-1177. doi: 10.1016/j.asr.2011.04.028. (Cited on page 28.)

M. Ludwig. Arena proceeding 2012. ARENA Conference 2012, Erlangen Germany, 2012. (Cited on pages 69 and 87.)

M. Ludwig and T. Huege. REAS3: Monte Carlo simulations of radio emissions from cosmic ray air showers using an 'end-point' formalism. *Astropart. Phys.*, 2010. doi: 10.1016/j.astropartphys.2010.10.012. (Cited on page 64.)

M. Ludwig and T. Huege. Analysis of air shower radio signals with reas3. *Proceedings of the 32nd ICRC, Bejing, China, 2011*, 2:19, 2011. (Cited on page 26.)

Marianne Ludwig, Tim Huege, Olaf Scholten, and Krijn D. de Vries. A detailed comparison of REAS3 and MGMR simulations for radio emission from EAS. 2012. (Cited on page 25.)

M. Santander et al. - IceCube Collaboration. Anisotropy of TeV and PeV cosmic rays with IceCube and IceTop. 2012. (Cited on page 19.)

J. Matthews. A Heitler model of extensive air showers. *Astroparticle Physics*, 22:387–397, January 2005. doi: 10.1016/j.astropartphys.2004.09.003. (Cited on page 22.)

M. Melissas, T. Huege, and A. Haungs. Measurements of high amplitude radio events during thunderstorms. GAP-2010-065, 2010. Auger internal publication, GAPNote. (Cited on pages 68 and 73.)

J. Neuser, K.-H. Kampert, and J. Rautenberg. Influence of an additional infill SD-station on the directional and energy reconstruction, 2011. (Cited on page 80.)

P. Abreu et al. - Pierre Auger Collaboration. Update on the correlation of the highest energy cosmic rays with nearby extragalactic matter. *Astropart.Phys.*, 34:314–326, 2010. doi: 10.1016/j.astropartphys.2010.08.010. (Cited on page 19.)

P. Abreu et al. - Pierre Auger Collaboration. Anisotropy and chemical composition of ultra-high energy cosmic rays using arrival directions measured by the Pierre Auger Observatory. *JCAP*, 1106:022, 2011a. doi: 10.1088/1475-7516/2011/06/022. (Cited on page 19.)

P. Abreu et al. - Pierre Auger Collaboration. The Pierre Auger Observatory II: Studies of Cosmic Ray Composition and Hadronic Interaction models. 2011b. (Cited on page 21.)

P. Lautridou et al. - CODALEMA Collaboration. Results of the radio-detection experiment CODALEMA. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 604:S13 – S19, 2009. doi: 10.1016/j.nima.2009.03.164. (Cited on page 79.)

N. Palmieri. *Determination of energy and mass of cosmic rays using air shower radio emission*. PhD thesis, Karlsruher Institut für Technologie, 2012. (Cited on pages 91, 107, and 112.)

William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press, New York, NY, USA, 3 edition, 2007. ISBN 0521880688, 9780521880688. (Cited on page 52.)

O. Ravel et al. The CODALEMA experiment. In *Nucl. Instr. and Meth. A, Proceedings of the ARENA 2010 conference, Nantes, France*, 2010. doi: 10.1016/j.nima.2010.12.057. (Cited on page 28.)

S. Argiro et al. - Pierre Auger Collaboration. The Offline Software Framework of the Pierre Auger Observatory. *Nucl. Instrum. Meth.*, A580:1485–1496, 2007. (Cited on page 30.)

S. Fliescher et al. - Pierre Auger Collaboration. Radio detection of cosmic ray induced air showers at the Pierre Auger Observatory. In *NIMA Proceedings of the ARENA 2010 conference, Nantes, France*, 2010a. doi: 10.1016/j.nima.2010.11.045. (Cited on page 28.)

S. Fliescher et al. - Pierre Auger Collaboration. The radio extension of Auger Offline. *Auger internal GAPNote GAP-2010-056*, 2010b. (Cited on page 58.)

O. Scholten, K. Werner, and F. Rusydi. A macroscopic description of coherent geo-magnetic radiation from cosmic-ray air showers. *Astropart. Phys.*, 29:393–411, 2008. (Cited on pages 25 and 64.)

Olaf Scholten, Krijn D. de Vries, and Klaus Werner. What the radio signal tells about the cosmic-ray air shower. 2012. (Cited on page 25.)

F.G. Schröder, others, and the LOPES Collaboration. On noise treatment in radio measurements of cosmic ray air showers. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 2010. doi: 10.1016/j.nima.2010.11.009. (Cited on page 68.)

T. Abu-Zayyad et al. - Telescope Array Collaboration. Search for Anisotropy of Ultra-High Energy Cosmic Rays with the Telescope Array Experiment. 2012. (Cited on page 19.)

T. Huege et al. - LOPES Collaboration. The LOPES experiment - recent results, status and perspectives. In *Nucl. Instr. and Meth. A, Proceedings of the ARENA 2010 conference, Nantes, France*, 2010. doi: 10.1016/j.nima.2010.11.081. (Cited on page 28.)

T. Huege et al. - Pierre Auger Collaboration. Advanced functionality for radio analysis in the Offline software framework of the Pierre Auger Observatory. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 635(1):92 – 102, 2011. ISSN 0168-9002. doi: DOI:10.1016/j.nima.2011.01.049. (Cited on pages 30 and 31.)

Hajime Takami and Katsuhiko Sato. Implications to Sources of Ultra-high-energy Cosmic Rays from their Arrival Distribution. 2008. (Cited on page 17.)

the CODALEMA collaboration. Charge excess signature in the codalema data. interpretation with selfas2. *Proceedings of the 32nd ICRC, Bejing, China, 2011*, 2011. (Cited on pages 91, 113, 114, and 116.)

the Pierre Auger Collaboration. Correlation of the highest-energy cosmic rays with nearby extragalactic objects. *Science*, 318(5852):938–943, 2007. doi: 10.1126/science.1151124. URL http://www.sciencemag.org/content/318/5852/938.abstract. (Cited on page 19.)

the Pierre Auger Collaboration. The Pierre Auger Observatory I: The Cosmic Ray Energy Spectrum and Related Measurements. *Proceedings of the 32nd ICRC, Bejing, China, 2011*, 2011. (Cited on page 79.)

Eli Waxman, Karl B. Fisher, and Tsvi Piran. The Signature of a Correlation between Cosmic-Ray Sources above 1019 eV and Large-Scale Structure. *The Astrophysical Journal*, 483(1):1, 1997. URL http://stacks.iop.org/0004-637X/483/i=1/a=1. (Cited on page 17.)

W.D. Apel et al. - KASCADE-Grande Collaboration. KASCADE-Grande - Contributions to the 32nd International Cosmic Ray Conference, Beijing. 2011. (Cited on page 20.)

W.D. Apel et al. - LOPES Collaboration. Lateral distribution of the radio signal in extensive air showers measured with lopes. *Astroparticle Physics*, 32(6):294 – 303, 2010. doi: 10.1016/j.astropartphys.2009.09.007. (Cited on pages 79 and 92.)

W.D. Apel et al. - LOPES Collaboration. Thunderstorm observations by air-shower radio antenna arrays. *Advances in Space Research*, 48(7):1295 – 1303, 2011. ISSN 0273-1177. doi: 10.1016/j.asr.2011.06.003. URL http://www.sciencedirect.com/science/article/pii/S0273117711004297. (Cited on page 68.)

W.D. Apel et al. - LOPES Collaboration. Lopes-3d, an antenna array for full signal detection of air-shower radio emission. *NIMA*, (0):–, 2012a. ISSN 0168-9002. doi: 10.1016/j.nima.2012.08.082. URL http://www.sciencedirect.com/science/article/pii/S0168900212009850?v=s5. (Cited on page 43.)

W.D. Apel et al. - LOPES Collaboration. Experimental evidence for the sensitivity of the air-shower radio signal to the longitudinal shower development. *Phys.Rev.*, D85:071101, 2012b. doi: 10.1103/PhysRevD.85.071101. (Cited on page 91.)

Klaus Werner, Krijn D. De Vries, and Olaf Scholten. Radio Emission from Cosmic Ray Air Showers based on Shower Simula-

tions: On the Importance of a Realistic Atmospheric Index of Refraction. 2012. (Cited on page 26.)

X. Bertou et al. - Pierre Auger Collaboration. CDAS & Herald Software. webpage, 2010. URL https://auger.colostate.edu/private/herald/. link to webpage of Herald software. (Cited on page 68.)

Y. Tsunesada et al. - Telescope Array Collaboration. Highlights from Telescope Array. 2011. (Cited on page 21.)

G. T. Zatsepin and V. A. Kuz'min. Upper Limit of the Spectrum of Cosmic Rays. *Soviet Journal of Experimental and Theoretical Physics Letters*, 4:78, August 1966. (Cited on page 18.)