Optimization of Radio Reconstruction for Inclined Air Showers with AERA at the Pierre Auger Observatory

Master Thesis

zur Erlangung des akademischen Grades eines
Master of Science (M.Sc.) im Studiengang Physik

Fakultät Mathematik und Naturwissenschaften
Fachgruppe Physik
der Bergischen Universität Wuppertal

vorgelegt von
Jelena Petereit
März 2022

BERGISCHE
UNIVERSITÄT
WUPPERTAL

Erstgutachter:
Prof. Dr. Karl-Heinz Kampert
Bergische Universität Wuppertal

Zweitgutachter:
Dr. Tim Huege
Karlsruher Institut für Technologie

Abstract

This thesis describes the adaptations and improvements that have been made on the AERA analysis with the Offline reconstruction framework. Using CoREAS simulations for measured event geometries, noise extracted from data can be added to a simulated pure signal. Various parameters for identifying noise dominated stations for the rejection in the geometry reconstruction, such as the time difference between the pure signal and the signal with noise, are examined and modified in order to improve the event reconstruction. The false rejection rate of the Cluster Finder noise rejection module was reduced from $\sim 7.9\%$ to $\sim 0.3\%$ and the correct rejection rate was increased from $\sim 53\%$ to $\sim 58\%$.

Contents

Chapter 1

Introduction

Cosmic rays were discovered by Victor Hess in his famous balloon experiments in 1912. He observed that, contrary to the belief at the time, radiation increases with higher altitude. Furthermore, after measuring radiation during a solar eclipse, he concluded that the source of the radiation had to originate from further out in space than the Sun [1]. Since their discovery in 1912, various characteristics of cosmic rays, such as their energies and mass composition, have been explored. However, many aspects including their origin and propagation processes are largely unknown. Especially for high-energy cosmic rays, questions surrounding these aspects are still unanswered. Contrary to the cosmic rays measured by Hess, high-energy cosmic rays can best be measured indirectly using ground-based detectors. They observe extensive air showers which are induced when high-energy cosmic rays enter the Earth's atmosphere. In order to obtain the necessary information, a large detection area as well as highly-developed detection methods are required.

The Pierre Auger Observatory is the world's largest cosmic ray observatory. It is designed to measure extensive air showers that are induced by high-energy cosmic rays in the Earth's atmosphere. Using a hybrid detection of various methods, it observes properties such as the arrival direction and energy spectrum, mainly for energies above $10^{17}\,\mathrm{eV}$ [2]. Its Auger Engineering Radio Array (AERA) consists of more than 150 antenna stations that cover an area of about $17\,\mathrm{km}^2$ and is used to detect radio signals emitted by extensive air showers [3]. These measurements are used to reconstruct properties of the primary cosmic rays that induce the air showers. The main scientific goals of AERA are the calibration of the radio emission from air showers, demonstrating the energy, angular and mass resolutions of the radio technique and the measurement of the cosmic ray composition in the range of both Galactic and extragalactic cosmic rays [2]. In contrast to some other detection methods, the observation of air showers using radio detection can be done during both day and night. Radio signals offer a high precision on discerning the direction and energy of air showers. As they are sensitive to the electromagnetic component of an air shower, radio signals also provide information on the depth of the shower maximum and thus the mass of the incoming cosmic ray.

This thesis describes the improvements that have been made on the AERA analysis with the $\overline{\text{Off}}\underline{\text{line}}$ reconstruction framework. Using CoREAS simulations for measured event geometries, noise extracted from data can be added to a simulated pure signal. Various parameters for identifying noise dominated stations for the rejection in the geometry reconstruction, such as the time difference between the pure signal and the signal with noise, are examined and modified in order to improve the event reconstruction.

Chapter 2

Cosmic Rays

This chapter covers the theoretical background of cosmic rays, including the characteristics of the cosmic ray energy spectrum and the evolution of extensive air showers. On this basis, the emission of radio signals, which is most relevant for this thesis, is discussed.

2.1 Cosmic Rays

Cosmic Rays (CR) are subatomic high-energy particles of extraterrestrial origin, which propagate through space. Nuclei make up the majority of their composition with 98%, out of which 87% are protons, 12% are α-particles and 1% nuclei of heavy elements. The remaining 2% consist of electrons [4].

CR can have different energies, which mainly depend on the mass and charge of the particles. Solar CR originate from solar flares and coronal mass ejections of our Sun and reach energies of approximately 10^7–10^{10} eV. The highest energetic CR that reach Earth form in the region outside our solar system within our Galaxy. These so-

called Galactic CR typically reach energies of up to 10^{10}–10^{15} eV. Finally, there are the so-called extragalactic CR, which reach energies above 10^{15} eV. Their sources are unidentified, but popular candidates are gamma-ray bursts, galaxy clusters, active galactic nuclei or starburst galaxies.

Cosmic Ray Energy Spectrum

Energy flux describes the rate of energy that is transferred through a surface. The energy flux of CR is highest for lowest energies of $\sim 10^9$ eV and decreases towards the highest energies $\simeq 10^{19}$ eV. The highest energetic solar CR make up the peak of the energy spectrum. Below energies of 10^{10} eV, the cosmic ray flux is suppressed by heliospheric magnetic fields that are induced by solar winds [5]. For higher energies, the differential flux of cosmic rays Φ can be described by

$$\Phi(E) = \frac{\mathrm{d}^4 N}{\mathrm{d}E\,\mathrm{d}\Omega\,\mathrm{d}A\,\mathrm{d}t} \propto E^{-\gamma} \tag{2.1}$$

where E is the energy, N is the number of particles, Ω is the solid angle, A is the area, and t is the time. The differential flux of cosmic rays follows a power law with a varying spectral index γ [6].

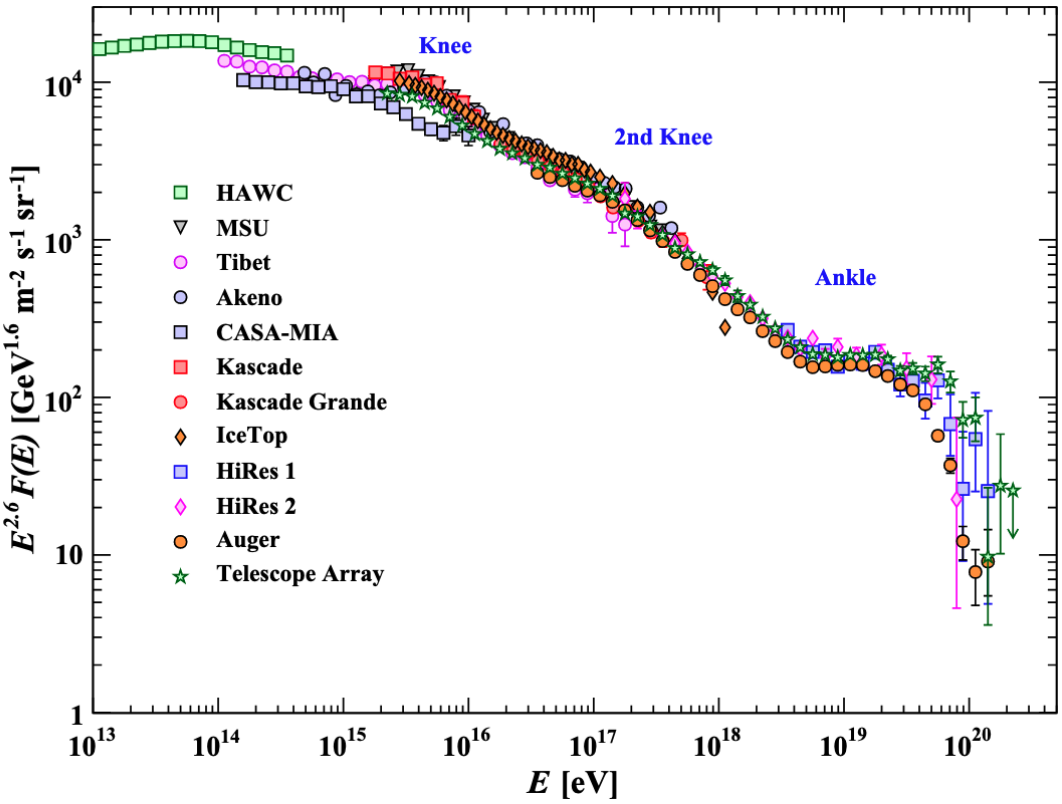

Figure 1: The observed energy spectrum for primary cosmic rays, measured by various air shower experiments. The flux is scaled with $E^{2.6}$ to highlight the characteristic features of the spectrum, which are labeled as *knee*, *second knee* and *ankle*. Figure taken from [7].

The observed energy spectrum of CR is shown in Fig. 1. The flux is scaled with $E^{2.6}$ to highlight the characteristic features of the spectrum above 10^{13} eV. The depicted spectrum begins with Galactic CR until it reaches the so-called *knee* at about $5 \cdot 10^{15}$ eV. There, Galactic accelerators are believed to have reached their maximum acceleration energy for protons and the spectral index slowly increases from $\gamma = 2.7$ to $\gamma = 3.1$. Below the knee, the majority of CR consists of heavier particles that are believed to be accelerated in supernova remnants within our Galaxy. At the so-called *second knee* at about 10^{17} eV, the spectral index increases further

to $\gamma = 3.2$. The next break in the spectrum at $\sim 10^{18}$ eV marks the so-called *ankle*. It indicates the transition from Galactic to extragalactic CR. There, the spectral index decreases to $\gamma = 2.6$. Finally, at energies above 10^{19} eV, the spectrum reaches the so-called *cutoff* and decreases rapidly [8].

The cutoff most likely is due to either the maximum cosmic ray acceleration energy of the CR sources or due to the Greisen–Zatsepin–Kuzmin (GZK) effect, or a combination of both [9]. According to the GZK efffect, CR lose energy as they interact with photons in the cosmic microwave background while propagating through space. This leads to a theoretical upper limit of $\sim 5{\cdot}10^{19}$ eV for CR energies for protons. For heavy nuclei, it manifests in photodisintegration [10].

2.2 Extensive Air Showers

An extensive air shower (EAS) is a cascade of ionized particles and electromagnetic radiation, which is induced when a primary CR enters the atmosphere. EAS can be divided into electromagnetic, muonic and hadronic components, according to their secondary particles. Electromagnetic air showers can be described by the Heitler model [11]. They are almost exclusively induced by pair production ($\gamma \rightarrow e^+ + e^-$) and Bremsstrahlung ($e \rightarrow e + \gamma$). Their interaction lengths remain constant for low energies and the interactions remain electromagnetic throughout the cascade [12]. For hadronic air showers, all types of interactions which include hadrons are possible. Their interaction lengths and energies vary and they can induce both electromagnetic and muonic air showers [11]. Figure 2

shows schematic views of both an electromagnetic and a hadronic air shower.

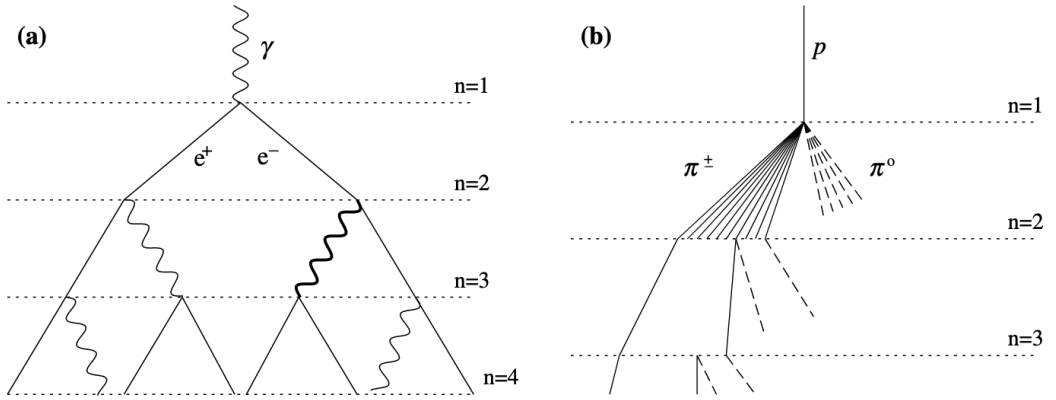

Figure 2: Schematic views of **(a)** an electromagnetic air shower and **(b)** a hadronic air shower. In the electromagnetic shower, squiggly lines represent photons and straight lines represent electrons or positrons. In the hadron shower, dashed lines represent neutral pions and solid lines represent charged pions. Both diagrams are taken from [11].

The evolution of an EAS and the division into its three components is shown in Fig. 3. The shower is initiated when a primary CR particle enters the Earth's atmosphere and interacts with its particles. These interactions produce pions, out of which one third are neutral and induce the electromagnetic component. They quickly decay into pairs of photons, which then interact with the nuclei in the atmosphere. The photons interacting with the nuclei in the atmosphere then create electron-positron pairs, which finally induce photons by Bremsstrahlung and lead to a cascade of particles. The remaining two-thirds of the initially produced pions induce the hadronic component of the air shower, which repeatedly produces neutral pions and thus induces more electromagnetic showers [13].

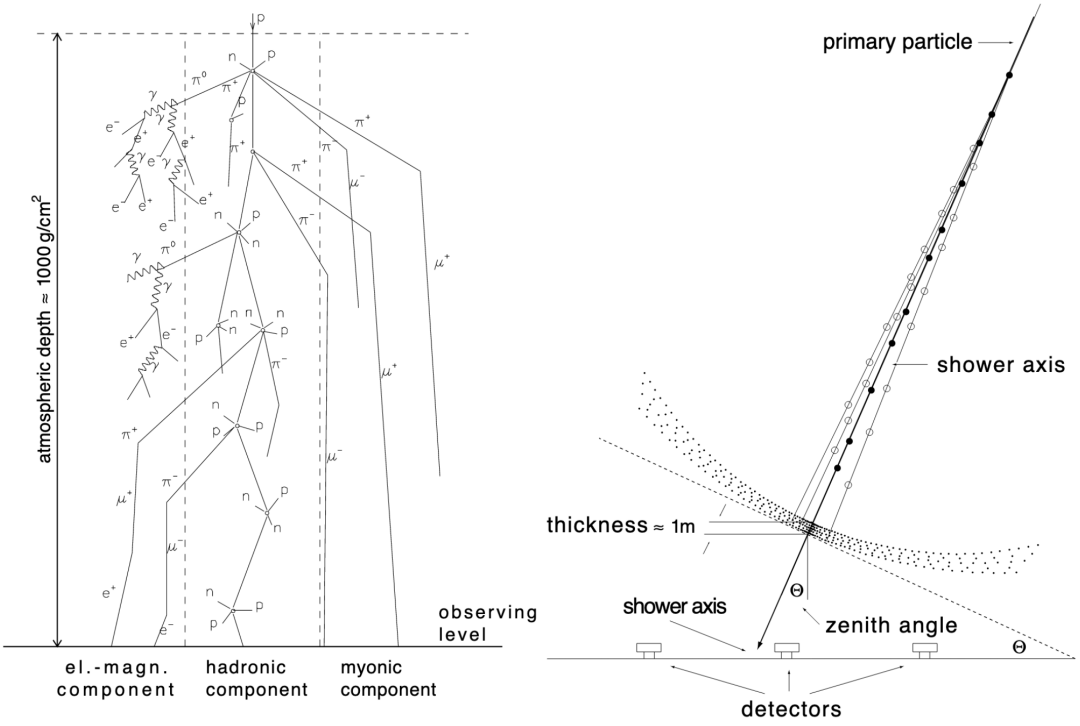

Figure 3: The production of an extensive air shower with its three components (left) and the shower axis of an extensive air shower (right), from [14].

2.2.1 Mass Composition

The longitudinal profile of an EAS is measured via the atmospheric slant depth X. The maximum air shower development $X_{\max}$ is related to the mass of the primary particle of the air shower. Using the Heitler Model, it is given by

$$X_{\max} = \lambda \log_2 (N_{\max}) = \lambda \log_2 \left(\frac{E_0}{E_c} \right) \tag{2.2}$$

where λ is the interaction length, $N_{\max}$ is the maximum number of particles in the cascade and E_0/E_c is the ratio of the primary energy and the critical energy [15].

Figure 4 shows measurements of first and second moment of $X_{\max}$ as a function of the primary energy. The predicted values for protons and iron are included in the figure, as CR are considered to be dominated by nuclei of elements up to iron.

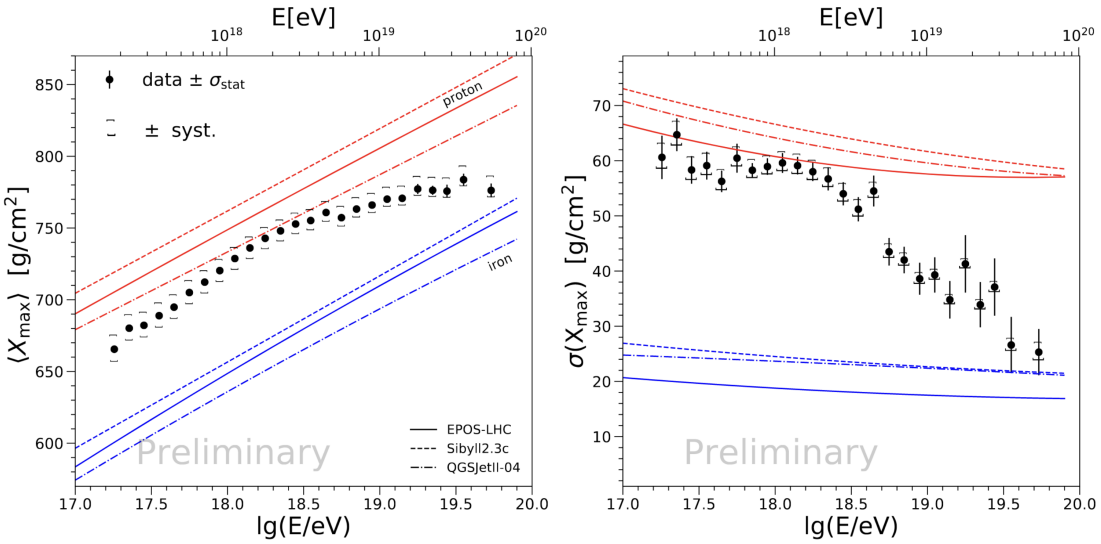

Figure 4: Measurements of $\langle X_{\max} \rangle$ (left) and $\sigma(X_{\max})$ (right) at the Pierre Auger Observatory. The values for proton and iron primaries are predicted using different hadronic interaction models. Figure taken from [16].

2.2.2 Radio Signals from Extensive Air Showers

Radio signals are emitted from the electromagnetic component of an air shower and can be explained by the geomagnetic [17] and the Askaryan effect [18]. Schematic diagrams of both effects are shown in Fig. 5.

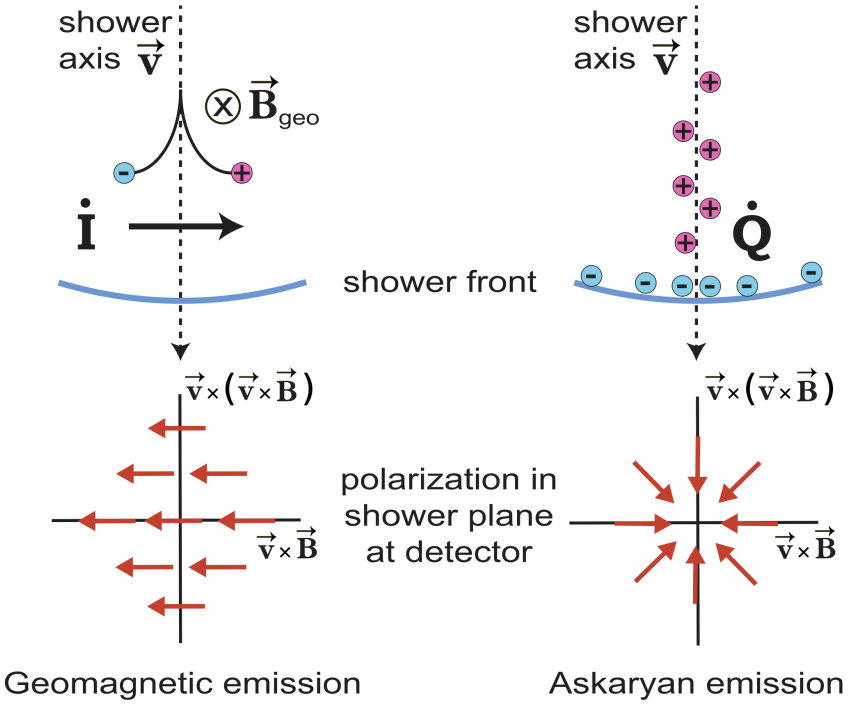

Figure 5: A diagram showing the charge distributions and the polarizations for the two main radio emission mechanisms, with the geomagnetic emission (left) and the Askaryan emission (right). Figure taken from [19].

The *geomagnetic effect* describes the emission that originates from the deflection of charged particles in the air shower. The Earth's magnetic field induces a Lorentz force, which accelerates electrons and positrons in the air shower in opposite directions. The deflection of the charged particles yields a time-dependent transverse drift current throughout the length of the air shower. The resulting radio emission is linearly polarized in the direction of the Lorentz force

$$|F_L| = |q| \cdot |v| \cdot |B| \cdot \sin\alpha \tag{2.3}$$

where q is the charge of the air shower's particles, v is the particles' velocity, B is the magnitude of the Earth's magnetic field and α describes the angle between v and B. Since the amplitude of

the emitted radio waves is proportional to B and α, the detection threshold and the efficiency of the reconstruction depends on the direction of the air shower [20].

The *Askaryan effect* describes radio emission due to the excess of negative charges in air showers. This negative charge excess is caused by the production of electrons while the air shower propagates and the annihilation of positrons as they decay into photons upon interaction with atmospheric electrons. Thus, the further the air shower propagates, the more the charge excess increases until it reaches its maximum and decreases again [20]. This time-dependent charge excess leads to the emission of radio signals that are radially linearly polarized towards the shower axis.

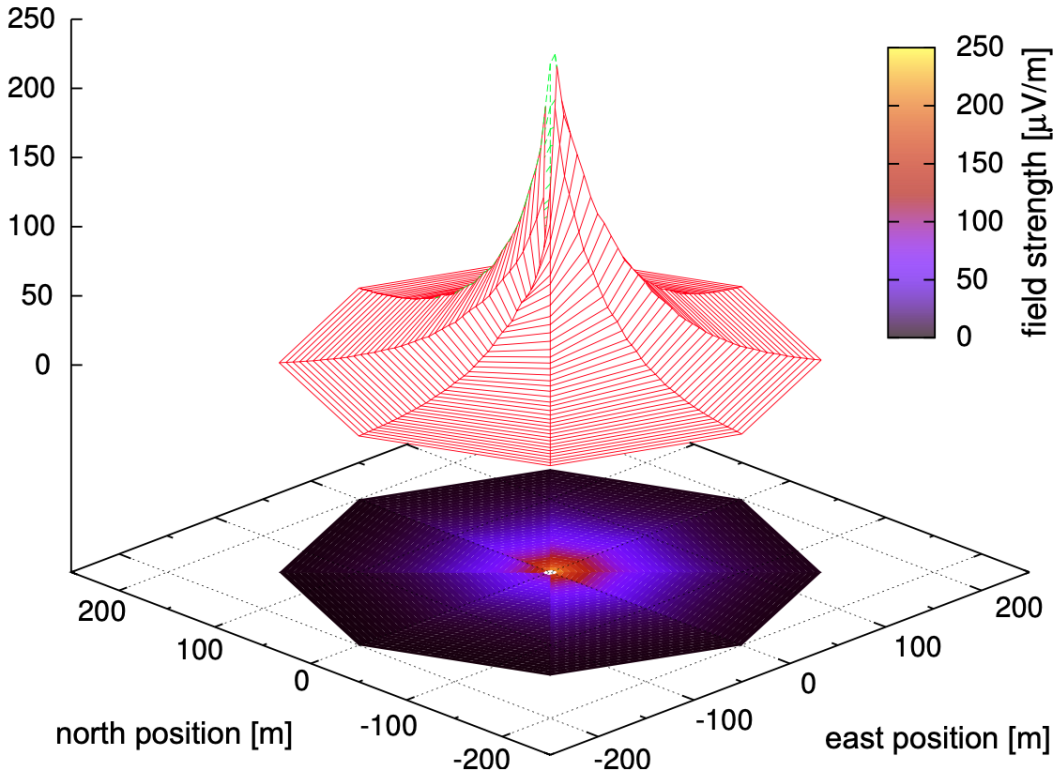

Figure 6: Radio footprint of the $40 - 80\,\mathrm{MHz}$ total field strength of a vertical $10^{17}\,\mathrm{eV}$ air shower induced by a proton primary. Figure from [21].

While the geomagnetic emission is typically the dominating component in air shower emission, the radio emission of EAS can be described as a superposition of both effects [3]. The radio emission expands in the shape of a cone, as can be seen in Fig. 6.

2.2.3 Inclined Air Showers

Besides the division into their shower components, EAS can also be distinguished according to their zenith angle (cf. Fig. 3). Air showers with zenith angles $\theta < 60°$ are called *vertical air showers*, while air showers with zenith angles $\theta > 60°$ are called *inclined air showers*[1].

First of all, the zenith angle has a considerable effect on the position of the shower maximum. For more inclined air showers, the shower will propagate further inside the Earth's atmosphere until it reaches its maximum than it does for vertical air showers. This affects the size of the footprint resulting from the air shower's radio emission, which is much larger for inclined air showers than it is for vertical air showers. Simulated footprints of the radio emission of an EAS for different zenith angles are shown in Fig. 7.

[1] Inclined air showers were called horizontal air showers (HAS) in the past. The abbreviation HAS is still used.

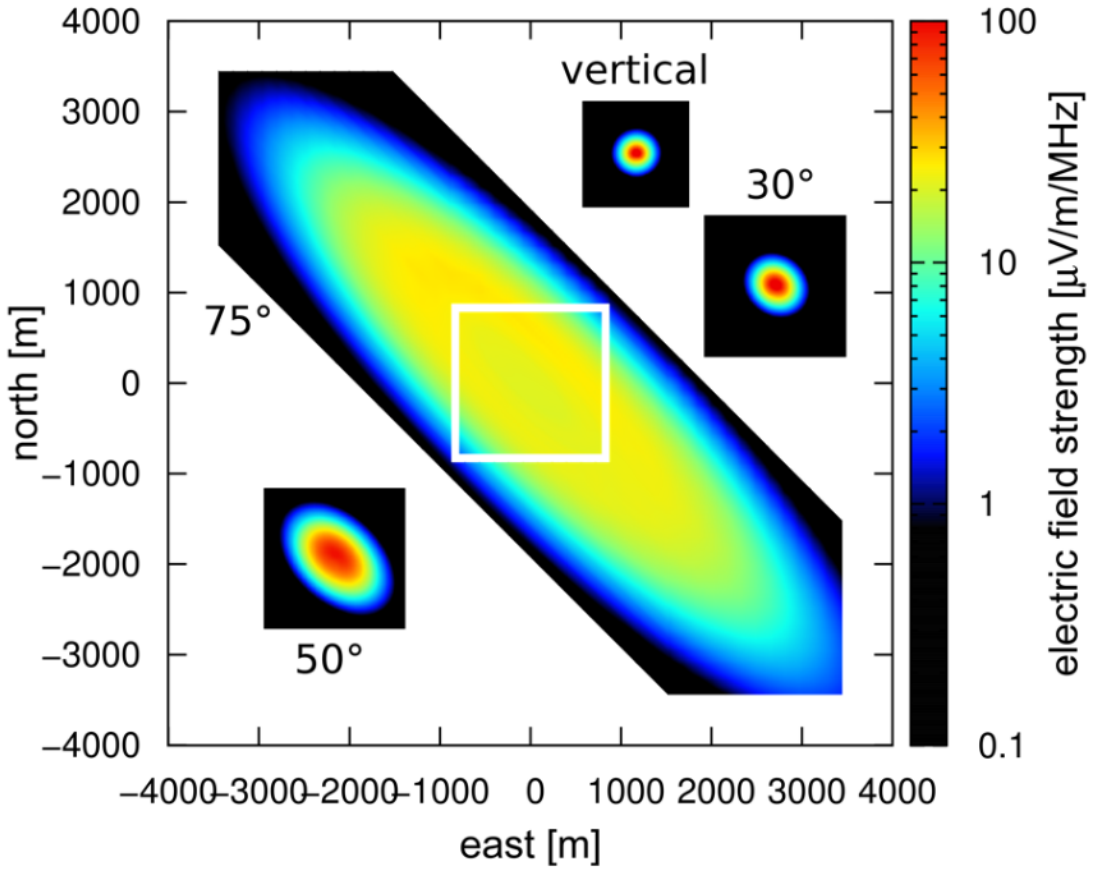

Figure 7: Radio footprints of extensive air showers simulated with various zenith angles in the $30 - 80\,\mathrm{MHz}$ band for an air shower with an energy of $5 \cdot 10^{18}\,\mathrm{eV}$. The white rectangle indicates the size of the 50° inset. Figure from [21].

The differences in footprint size of radio emission for inclined and vertical showers justify different detection setups. For inclined air showers, the radio footprint is larger and the average electric field amplitude is lower. For vertical air showers, the radio footprint is smaller and the average electric field amplitude is higher [22]. Thus, a sparse antenna grid is more sensitive to inclined air showers, while a more dense antenna grid is more sensitive to vertical air showers.

2.3 Simulating Air Showers

In this thesis, air showers simulated by the CORSIKA (**CO**smic **R**ay **SI**mulation for **KA**SCADE) [23] and its radio extension CoREAS (**CO**RSIKA-based **R**adio **E**mission from **A**ir **S**howers) [24] are used.

CORSIKA offers detailed simulations of extensive air showers that are induced by high energy cosmic rays. It allows to select primary particles, such as protons, light nuclei and photons, which are then tracked through the atmosphere while they interact with other particles and finally decay. With CoREAS, the radio emission of the electrons and positrons is calculated using the endpoint formalism [24].

A simulated air shower has to account for all particles of the shower along with their decays and interactions with other particles in the Earth's atmosphere. Consequently, both hadronic and electromagnetic interactions have to be modeled at sufficiently high energies. They have to be included along with an elaborate representation of the atmosphere that the particles are tracked through.

Chapter 3

The Pierre Auger Observatory

The Pierre Auger Observatory is located on the planes of Pampa Amarilla 1400m above sea level next to the town of Malargüe in western Argentina. Covering an area of $3000\,\mathrm{km}^2$, it is the world's largest cosmic ray observatory. It is designed to detect extensive air showers with energies of $10^{17}\,\mathrm{eV}$ up to $10^{20}\,\mathrm{eV}$ by using a combination of various detection methods [2].

The observatory's two baseline components are the Surface Detector (SD) and the Fluorescence Detector (FD). The SD is an array that consists of more than 1600 water-Cherenkov detectors [25], while the FD is made up of 24 telescopes located at four different sites overseeing detection area [26], as shown in Fig. 8. In the western part of the detection area, additional components are installed to further detect CR. The Auger Muon and Infill Ground Array (AMIGA) combines measurements of water Cherenkov detectors and underground scintillators in order to measure the muon content of air showers. The High Elevation Auger Telescope (HEAT) consists of three additional fluorescence telescopes with an elevated field of view and works independently of other FD sites. Both AMIGA

and HEAT enable measurements of low-energy CR [2]. Finally, the most important detection site for this thesis is the Auger Engineering Radio Array (AERA). It consists of more than 150 antenna stations that cover an area of about 17 km^2 and is used to detect radio signals in the range of 30-80 MHz [3]. A more detailed description of AERA is given in Sec. 3.3.

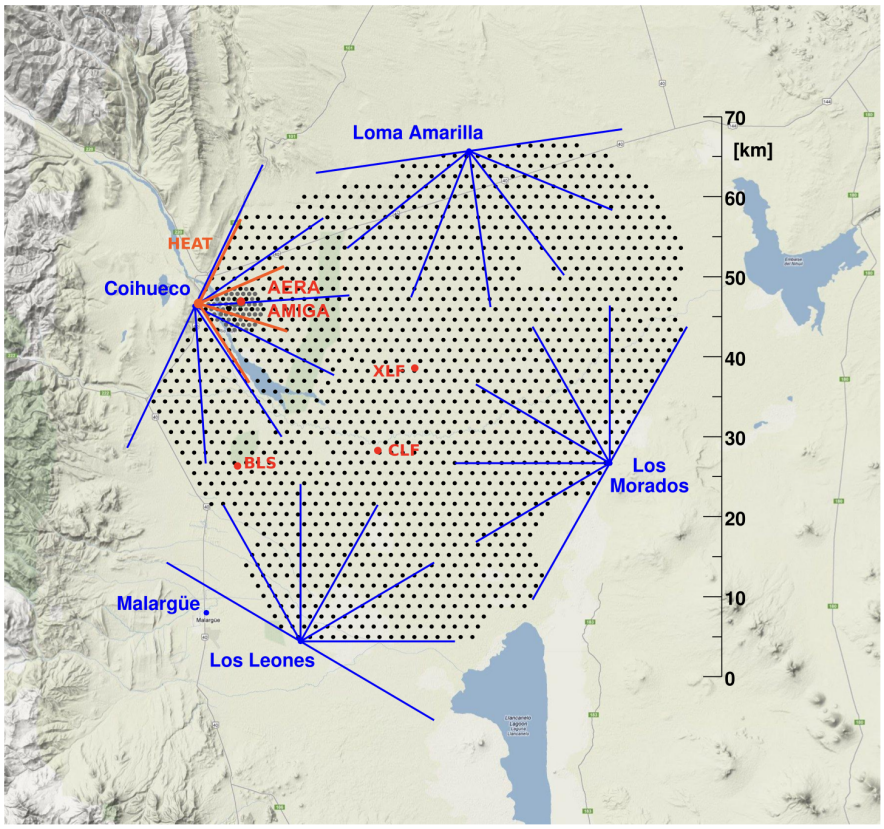

Figure 8: A map of the Pierre Auger Observatory. SD stations are represented by black dots. The blue lines indicate the fields of view of the 24 FD telescopes and the red lines indicate the fields of view of HEAT. Figure from [27].

The key feature of the Pierre Auger Observatory is its hybrid design. The various detection methods can be combined to simultaneously measure different aspects of CR. Typically, the two baseline components SD and FD are used as an example for this. The SD array observes the air shower at ground level, measuring the elec-

tromagnetic and muonic components of the shower, while the FD monitors the longitudinal development of the shower cascade in the atmosphere [2].

3.1 Surface Detector

The SD is an array that includes more than 1600 water-Cherenkov detectors. They are distributed on a hexagonal grid at distances varying between $433 - 1500\,\mathrm{m}$[1]. Each SD station consists of a water tank of 3.6 m in diameter and 1.2 m in height, which contains a sealed liner with a reflective inner surface and $12\,\mathrm{m}^3$ of ultra-pure water. Each SD station operates independently. A solar panel powers the photomultiplier tubes (PMTs) as well as further electronics such as a processor, a GPS receiver, a radio transceiver and a power controller. A picture of an SD station is shown in Fig. 9.

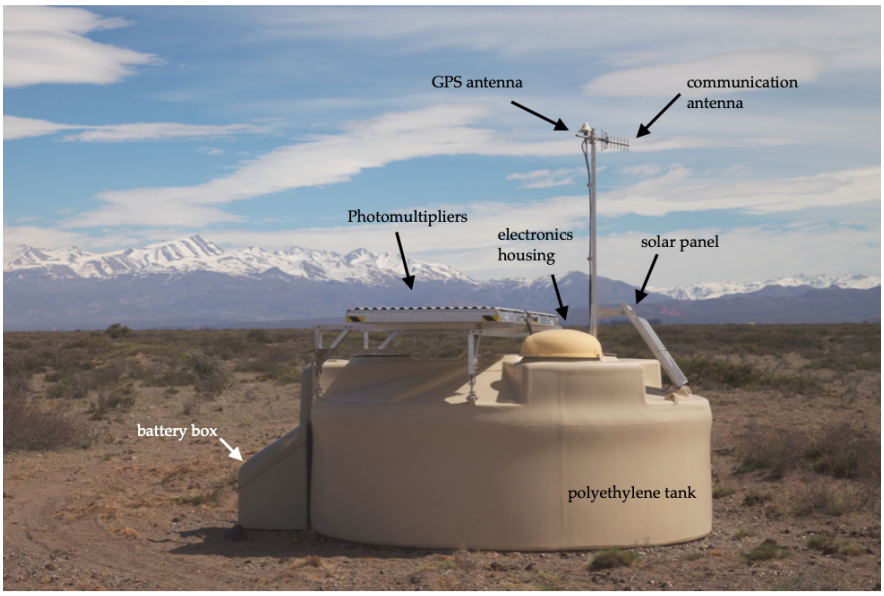

Figure 9: Labeled picture of a water-Cherenkov detector and its components. Photo adapted from [28].

[1]The main spacing is 1500 m, while the infill spacing is 433 m or 750 m.

SD stations observe air showers at ground level, measuring the signal footprint on the ground at the time of arrival. Specifically, Cherenkov light is induced by an air shower's electromagnetic and muonic components in the tank. The Cherenkov light is then reflected off the reflective inner surface of the tank until it reaches one of the three 9-inch diameter (PMTs) that are installed on top of the tank. If the measured signal passes the station's local trigger, the local triggers are combined to an air shower trigger that initiates the acquisition of the measured signal traces of all associated stations [25].

3.2 Fluorescence Detector

The FD consists of 24 telescopes that are situated at four different sites (Los Leones, Los Morados, Loma Amarilla and Coihueco) of the detection area. Each FD station houses six independent telescopes in a clean climate controlled building, which can be seen in Fig. 10. The telescopes each have a field of view of $30^\circ \times 30^\circ$ in azimuth and elevation, with a minimum elevation of 1.5° above the horizon. As can be seen on the map in Fig. 8, the telescopes are directed towards the center of the detection area, which results in a combined azimuthal field of view of 180° [26].

The FD is only utilized during dark and moonless nights with clear weather conditions, which reduces its operation time to $\sim 15\%$ [2]. Due to the aforementioned hybrid detection, the FD is always controlled by the SD triggers. This enables a synchronized measurement of the same air showers by both SD and FD. The FD monitors the longitudinal development of the shower cas-

cade in the atmosphere by detecting fluorescent light that is emitted by the air shower. The CR particles and the atmosphere produce photons through various interactions. In particular, UV light is emitted when atmospheric nitrogen is excited by the particles of the air shower.

Figure 10: A picture of the FD building with closed shutters, from [2].

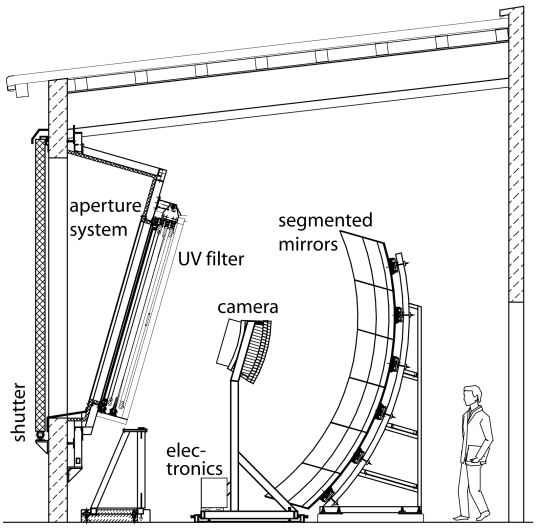

Figure 11: An illustration of a fluorescence telescope with its main components, from [26].

Figure 11 shows an illustration of an FD telescope with its main components. The aperture system admits fluorescent light in the UV range into the main room of the telescope, where it is reflected off of a 13 m^2 segmented spherical mirror onto a camera with PMTs [26].

3.3 Auger Engineering Radio Array

The Auger Engineering Radio Array (AERA) is an enhancement of the Pierre Auger Observatory. It consists of more than 150 antenna stations which are used to detect radio signals emitted by extensive air showers [3]. These measurements are used to reconstruct properties of the primary CR that induce the air showers. The main scientific goals of AERA are the calibration of the radio emission from air showers, demonstrating the energy, angular and mass resolutions of the radio technique and the measurement of the CR composition in the range of both Galactic and extragalactic CR [2].

In contrast to other detection methods such as FD, the observation of air showers using radio detection can be done during both day and night. Radio signals offer a high precision on discerning the direction and energy of air showers. As they are sensitive to the electromagnetic component of an air shower, radio signals also provide information on the depth of the shower maximum and thus the mass of the incoming CR.

AERA Phases and Antenna Types

The deployment of AERA was performed in three phases. Figure 12 shows the distribution of the AERA stations and their phases. In 2010, the 24 stations of AERA phase I were installed, using logarithmic dipole antennas (LPDA). A picture of an LPDA antenna is shown in Fig. 13.

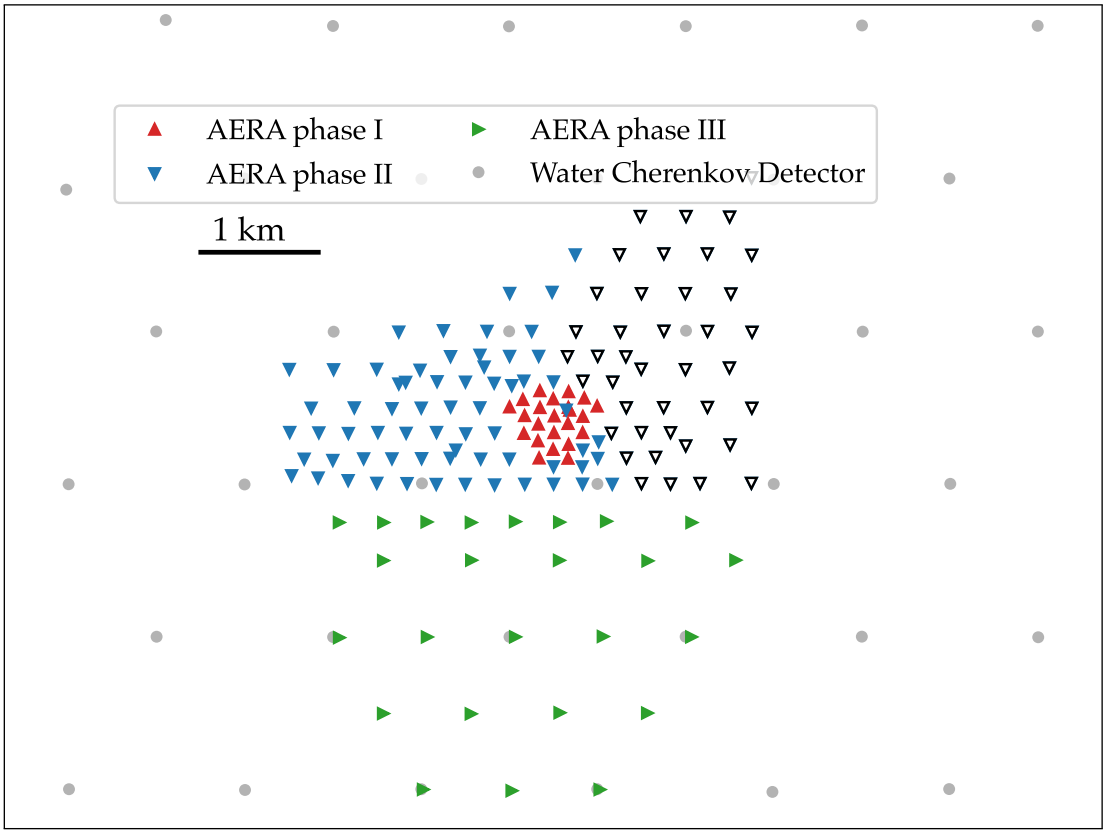

Figure 12: A map of the stations of the three AERA phases. AERA phase I stations are marked by an upward triangle, AERA phase II stations by a downward triangle and AERA phase III stations by a triangle pointing to the right. German stations are colored in while Dutch stations are uncolored. Water Cherenkov detectors are marked by gray circles.

In May 2013, AERA phase II was installed with 100 stations using butterfly antennas. A picture of a butterfly antenna is shown in Fig. 14. This new type of antenna was developed for better narrow pulse detection as compared to the LPDA antennas [29]. In March 2015, the stations for AERA phase III were deployed, increasing the total number of radio detection stations to 153 [2]. As can be seen in the map shown in Fig. 12, the spacing between the stations differs for each phase. While the stations in phases I and II are less than $\sim 400\,\mathrm{m}$ apart, the distances between the stations in phase III are

up to $\sim 750\,\mathrm{m}$. As mentioned in Sec. 2.2.3, a sparse antenna grid is more sensitive to inclined air showers, while a more dense antenna grid is more sensitive to vertical air showers. Thus, AERA phases I-II are more convenient for detecting vertical air showers, whereas AERA phase III is best suited for detecting inclined air showers.

Figure 13: Picture of an AERA LPDA antenna, from [30].

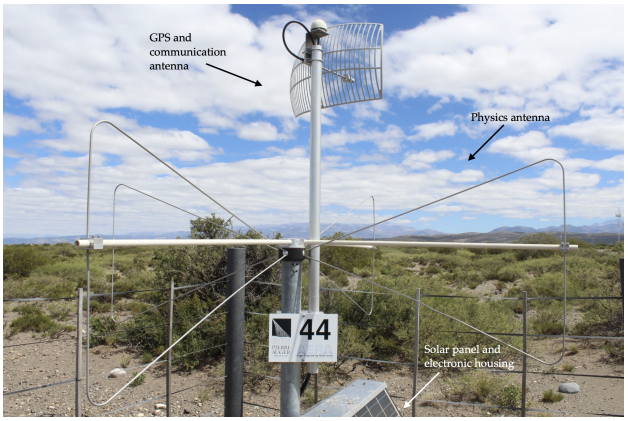

Figure 14: Labeled picture of an AERA butterfly antenna, from [31]

Each AERA station is equipped with a dual polarization antenna, which detects the electric field in the north/south and east/west directions, and is sensitive to radio signals between $30 - 80\,\mathrm{MHz}$. Similar to the SD stations, each station is powered by a solar panel and is equipped with further electronics and communication antennas. The read-out of the radio data is initiated by a self-trigger or an external trigger. However, most radio stations receive external triggers from SD or FD [32].

Chapter 4

Simulations and Radio Reconstruction with $\overline{\text{Offline}}$

This chapter covers the simulations used in this thesis as well as the basics on the Auger Reconstruction Framework $\overline{\text{Offline}}$.

4.1 Simulation

To evaluate the performance of the noise rejection for AERA III, four different sets of inclined simulations are used.

- *set A* [33]:
  contains reconstructed real events, which passed a high quality event selection.
- *set B* [34]:
  contains reconstructed real events, which passed the same high quality event selection. In this set, the energy is artificially increased.

- *set C* [35]:
  contains generic events where shower core, direction and energy are randomly selected. In order to simulate air showers that are relevant to AERA, the number of antennas near the shower axis is crucial. In this case, a minimum of five antennas are within three Cherenkov radii from the shower axis.

- *set D* [36]:
  contains generic events where shower core, direction and energy are randomly selected. In order to simulate air showers that are relevant to AERA, the number of antennas near the shower axis is crucial. In this case, a minimum of ten antennas are within one Cherenkov radius from the shower axis.

Combined, these four sets of simulations cover an energy range of $10^{18.3}\,\mathrm{eV} < E < 10^{19.8}\,\mathrm{eV}$ and a zenith angle range of $58^\circ < \theta < 84^\circ$, as can be seen in Fig. 15. The individual distributions are shown in Appendix B.

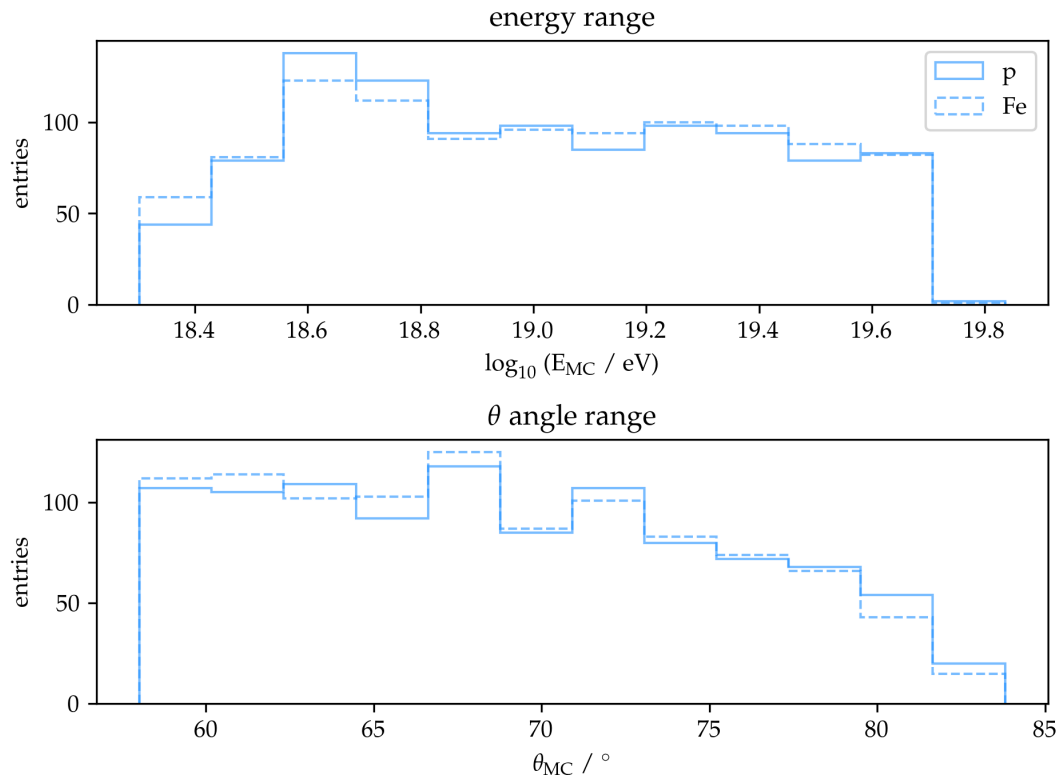

Figure 15: The energy range (top) and zenith angle range (bottom) of the simulation used in this thesis for proton (solid) and iron (dashed).

For the reconstruction, randomly selected measured noise can be added to pure simulated signals with Offline. Since simulating noise accurately is nearly impossible for radio signals, measured noise is being used instead. An example of a trace from one simulated event is shown in Fig. 16. The upper plot shows the full traces of the simulated signal peak and the plot at the bottom of the figure shows the same simulated signal peak with added measured noise. For inclined air showers, the eastern and northern polarizations can be measured, but the vertical component can only be reconstructed.

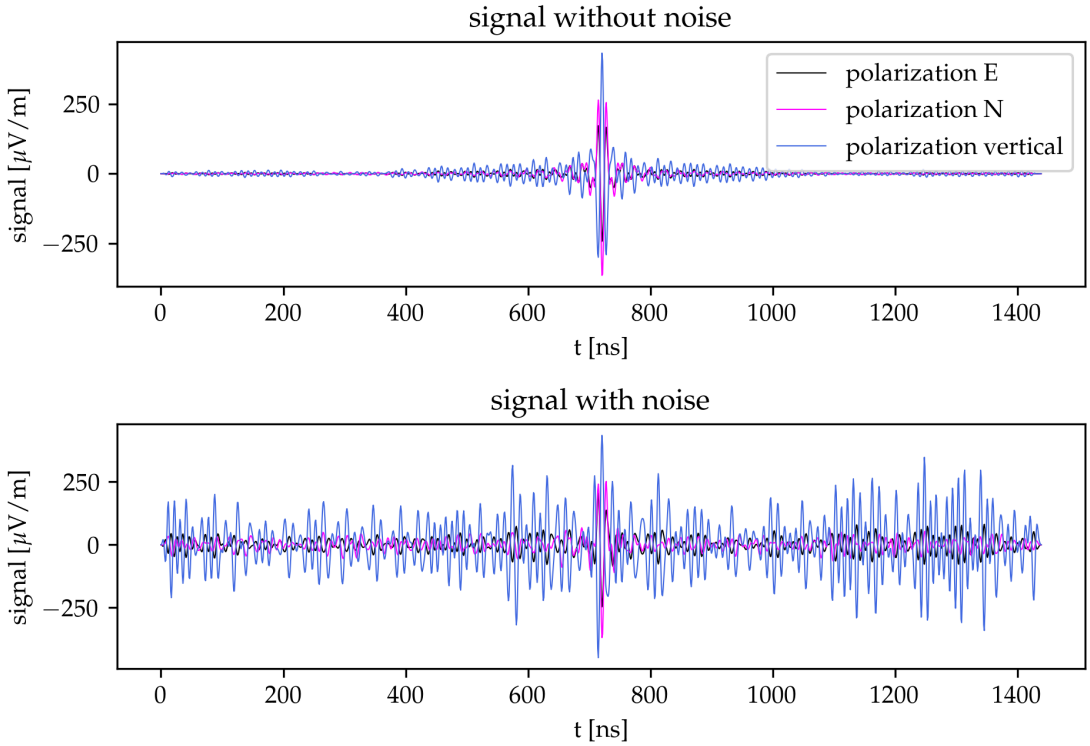

Figure 16: A reconstructed signal trace without noise (top) and with noise (bottom). The polarization in eastern direction is shown in blue, the polarization in northern direction is shown in pink and the vertical polarization is shown in black. Note that the vertical polarization is reconstructed, because it cannot be measured.

4.2 The Reconstruction Framework Offline

The analysis software Offline is a `C++` framework that provides various methods to analyze measured data and simulated air showers [37], [38]. Originally, it was designed to reconstruct air showers using information from SD and FD.

As shown in Fig. 17, Offline can be divided into three fundamental parts. The *Detector Description* provides information on the different components of the detector. Time-varying information

is stored in `MySQL` databases, while consistent information such as detector positions is stored in `XML` files. The *Data Processing* contains the various modules that perform the different parts of the event reconstruction. The modules are run consecutively, reading from and storing information in the *Event Data*. The Event Data contains different types of event data, including both measurements and reconstructed information.

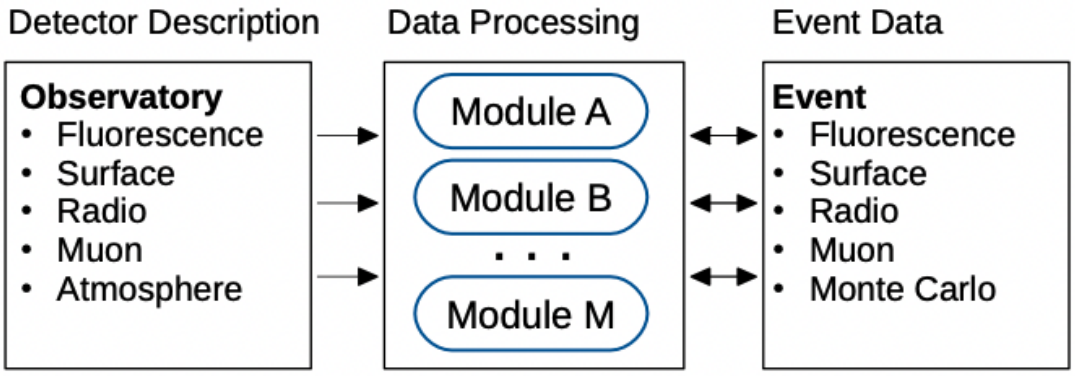

Figure 17: The three fundamental parts of the Offline framework, from [39].

This thesis uses the `RdHASSimulationObserver`, which is designed to reconstruct simulations of inclined air showers. It contains detailed detector simulations and allows the addition of measured noise [40]. A reconstruction is defined by the `ModuleSequence.xml` file, which describes the sequence of applied Data Processing modules, and the `bootstrap.xml` file, which can modify pre-defined parameters in order to produce different kinds of reconstructions.

Reconstructed events are written to an `ADST.root` (Advanced Data Summary Tree) file, which is based on the `C++` framework `ROOT` [41]. The `ADST` file contains information on the detector and the event data as well as the applied Offline settings. The produced `ADST` file can be opened using Offline's `Event Browser`, which vi-

sualizes the reconstructed events and raw event information of the different detectors.

4.3 Noise Identification in Offline

One main challenge that this thesis aims to address is the high frequency in which radio frequency interference (RFI) pulses are observed. In order to avoid misinterpreting an RFI pulse as a signal pulse, a *signal search window* is defined. It is typically centered around the expected signal arrival times based on the reconstructed SD geometry. Since the range of the *signal search window* is large compared to the typical length of a radio pulse of a few nano seconds, RFI pulses are still misinterpreted as signal pulses. In order to distinguish the signal pulse from the RFI pulses, Offline includes several methods for rejecting noise stations.

A station is either classified as signal-dominated when it is not rejected or classified as noise-dominated when it is rejected by Offline. Particularly, these rejections are performed by the four noise rejection modules

1. `RdStationPulseShapeRejector`
2. `RdStationPolarizationRejector`
3. `RdClusterFinder`
4. `RdTopDownStationSelector`

which are typically run in this order. This section will provide an overview of the four modules and discuss issues caused by the `RdClusterFinder`, which will be analyzed further in this thesis.

Pulse Shape Rejection

In order to distinguish a noise pulse from a signal pulse, the `RdStationPulseShapeRejector` matches the shape and width of a detected pulse to that of an expected radio pulse.

Polarization Rejection

The `RdStationPolarizationRejector` uses the superposition of geomagnetic and charge excess emission as well as SD geometry to compare the measured polarization to the prediction of superposition. Since it has been shown that its false rejection rate increases for large zenith angles [42], this method is not used for inclined air showers. The deactivation is done using azimuthal bins of 30°. For some 30° bins the rejection is performed up until 50° in zenith, for others it goes up to 70° in zenith. Therefore, few polarization rejections may still occur despite the module's deactivation.

Cluster Finder Rejection

The `RdClusterFinder` rejects lonely stations or isolated clusters that are far away from the shower axis. For this, the Cluster Finder module uses two criteria. First, there is the so-called *Lonely Station Rejection*, which rejects stations that have either no neighboring signal stations within a radius of 400 m or only one neighboring

signal station within a radius of 800 m. The second criterion is the so-called *Distance Rejection*. It calculates the distance to the shower axis for each station and compares the results. If neighboring stations have a difference in shower axis distance of more than 500 m, they are rejected.

The distances that are applied in the *Lonely Station Rejection* have not been adjusted since the addition of AERA phase III. As can be seen in Fig. 12, the distances between stations in phase III are up to three times larger than the ones in phases I and II. As a result, the phase III stations on a 750 m grid will always be rejected. Since the *Distance Rejection* is independent of the distances between stations, but instead relies on the SD geometry, it should not be affected by the different setups of AERA phases.

Top Down Rejection

The `RdTopDownStationSelector` detects stations with wrong signal arrival times. It begins with inspecting the signal arrival time of the three stations that are closest to the SD core. If they each match the expectation for a signal pulse, the procedure moves on to examine the next station. In an iterative procedure, an additional station is included in the wavefront fit. The new station is included if it does not decrease the χ^2 probability of the fit below 5 %. Otherwise, it is classified as noise-dominated station and the procedure continues with the next station.

In a previous analysis, AERA I-II reconstructions using all four rejection modules yield a true rejection rate of 92 % for both ver-

tical and inclined air showers [42]. However, this analysis neither includes the challenges introduced by the different AERA III setup for the Cluster Finder's *Lonely Rejection*, nor does it include the deactivation of the polarization rejection. Moreover, a succeeding analysis [39] replaced the plane wavefront fit of the Top Down module with a spherical one for inclined air showers. This thesis will revisit the analysis of the rejection modules with a special emphasis on the *Lonely Rejection* for AERA III.

Chapter 5

Noise Rejection Methods

This chapter covers the modifications that were applied to the methods for noise rejection that are already included in the Auger Reconstruction Framework $\overline{\text{Off}}\underline{\text{line}}$. In order to evaluate the performance of the existing methods for noise rejection, CoREAS simulations for measured event geometries are used. Noise extracted from data is added to a simulated pure signal, making the simulation more correlative to measured data and thus the evaluation more reliable. Particularly, changes have been made on the *Cluster Finder* module of the $\overline{\text{Off}}\underline{\text{line}}$ framework in order to improve noise rejection.

5.1 Cluster Finder

As mentioned in Sec. 4.2, the Cluster Finder module uses two criteria to perform noise rejections. The so-called *Lonely Station Rejection* rejects stations that have no neighboring signal stations within a radius of 400 m or only one neighboring station within a radius of 800 m. The second criterion is the so-called *Distance Rejection*,

which rejects stations that are further away than 500 m from the reconstructed SD axis.

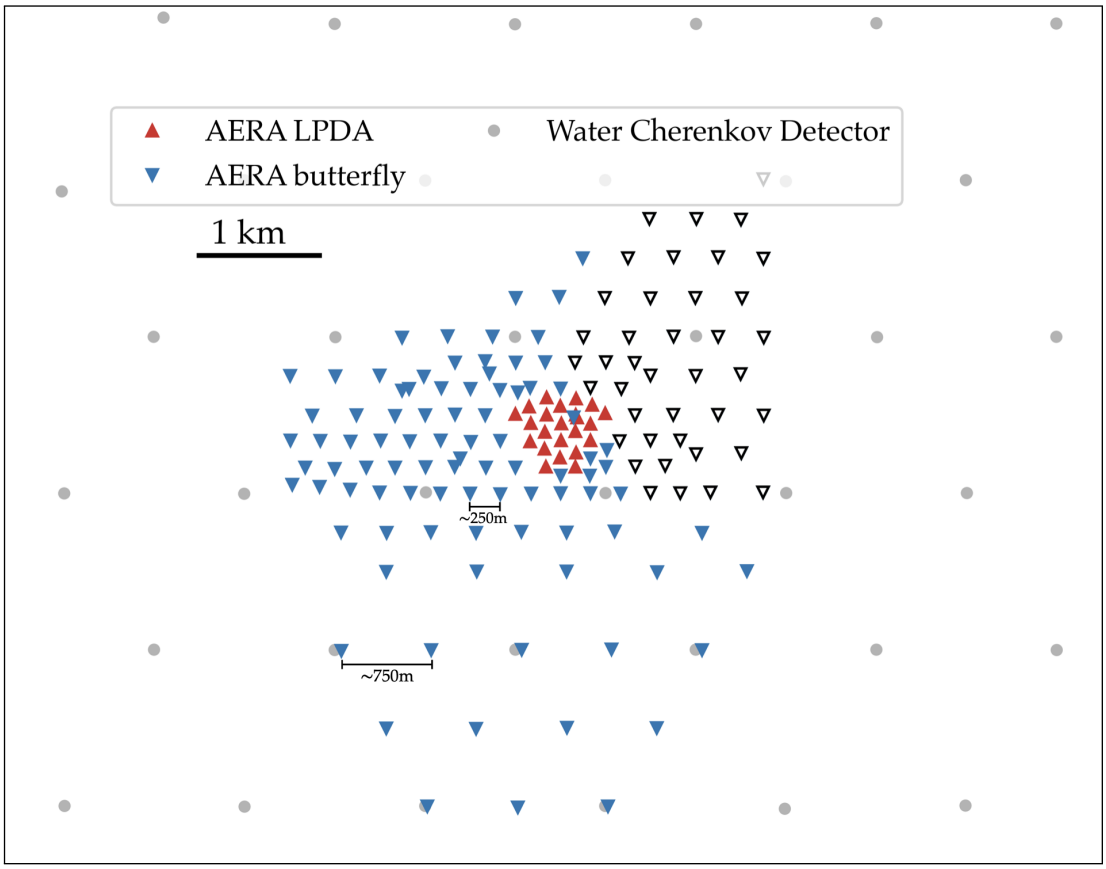

Figure 18: A map of the AERA stations. AERA LPDA stations are marked by an upward triangle and AERA butterfly stations are marked by a downward triangle. German stations are colored in while Dutch stations are uncolored. Water Cherenkov detectors are marked by gray circles. The distances between two AERA phase II stations and between AERA phase III stations are indicated.

While the method itself was developed well, the distances that are applied in the *Lonely Station Rejection* have not been adjusted since the addition of the later AERA phases. As can be seen in Fig. 18, the distances between stations in phase III are up to three times larger than the ones in phases I and II. As a result, phase III

stations are more likely to be falsely rejected whenever they measure a signal, simply because they are further than 400 m apart.

As a first approach to improving the noise rejection performed by the Cluster Finder, examining the numbers of rejections according to the *Distance Rejection* as well as the *Lonely Station Rejection* can offer more insights. Since the *Distance Rejection* uses the distance of a station from the SD shower axis, the differences in distances between the stations themselves should have no effect on this.

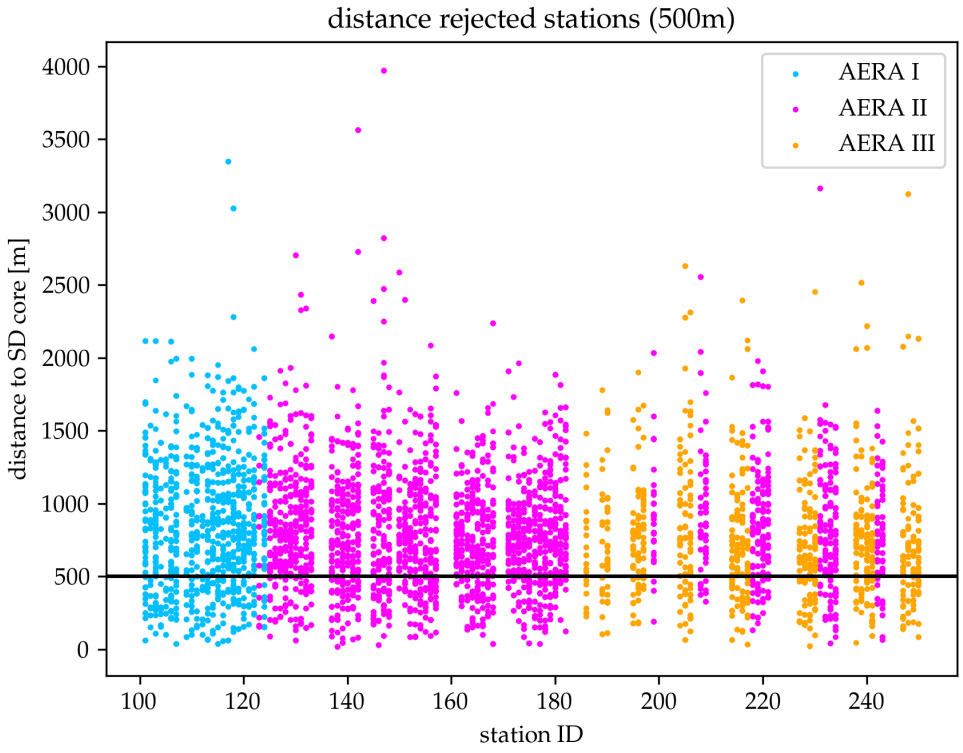

Figure 19: The distance to the SD core per station ID using the standard Cluster Finder for 3500 events. A black line indicates the rejection distance of 500 m. AERA phase I stations are marked in blue, AERA phase II stations in pink and AERA phase III stations in orange.

Figure 19 shows the distance to the SD core for each station ID using the standard Cluster Finder for 3500 events. The individual points indicate the numbers of stations that were rejected due to their distance to the SD shower axis. As expected, the numbers of rejections are on the same level for all three AERA phases. Thus, these results show that the differences in distances of the three AERA phases have no effect on the *Distance Rejection*.

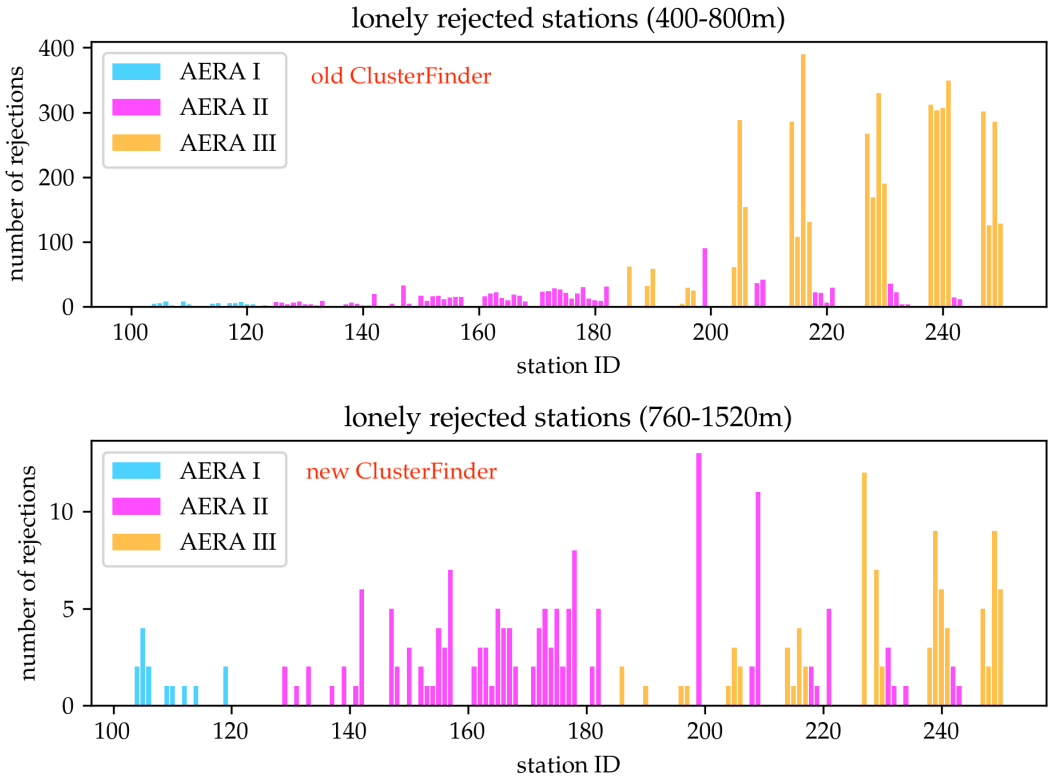

Figure 20: Numbers of lonely rejected stations per station ID using the standard Cluster Finder (top) and a modified Cluster Finder (bottom) for 3500 events. AERA phase I stations are marked in blue, AERA phase II stations in pink and AERA phase III stations in orange. Note the different scalings of the y-axes.

As a next step, the *Lonely Station Rejection* needs to be examined. In its standard settings, it rejects stations that have either no neighboring signal stations within a radius of 400 m or only one

neighboring signal station within a radius of 800 m. Figure 20 shows the numbers of lonely stations for each station ID using the standard Cluster Finder (top) and a modified Cluster Finder (bottom) for 3500 events. As expected, the numbers of *Lonely Station Rejection* using the standard 400-800 m are much higher for phase III stations than for phase I and II stations.

The functionality of the existing Cluster Finder is extended with respect to the distances between stations in each AERA phase. The corresponding phase is detected by station ID. For phases I and II, the corresponding stations with IDs in the range of 1-157 are included. For phase III, the corresponding stations with IDs in the range of 86-117 are included. All distances that are used for the rejections can be configured using the existing `xml` parameters

- `LonelyIfNoneInDistance`
- `LonelyIfOneInDistance`

and the newly implemented `xml` parameters

- `LonelyIfNoneInDistancePhase3`
- `LonelyIfOneInDistancePhase3`

Examples of configuration files with these parameters are listed in Appendix C.

Typically, stations in phase I and II are less than 250 m apart, whereas stations in phase III are more than 750 m apart. Thus, the standard settings of 400-800 m are maintained for phase I and II stations, while for phase III stations the distances for *Lonely Station*

Rejection are increased to 760-1520 m according to the positions of the phase III stations. These modifications decrease the absolute number of phase III rejections to the same level as the phase I and II rejections.

Although reducing the total number of rejections is a good first indicator of reducing the number of falsely rejected signals, the quality of the rejections still needs to be evaluated. When analyzing measured data, it is possible to count the numbers of rejections, but it is not possible to clearly identify signal and noise. Therefore, it is not possible to reliably verify the quality of the performed noise rejections. Instead, these analyses can be performed using a simulation where signal and noise are clearly defined in advance and can be compared to the rejections.

5.2 Identification of Noise Pulses

In this section, the effect of the noise background that is added to simulated signal traces is explored as well as the definition of signal and noise stations. Subsequently, the parameter Δt is introduced as a means for identifying noise pulses. Finally, in the last part of this section, the quality of the rejections is examined by calculating the respective probabilities as well as their true and false positive rates. The PYTHON code that was developed for this analysis can be found on `GitLab`[1].

[1] https://git.uni-wuppertal.de/buw-auger/radioanalysis

5.2.1 Comparing Signal and Noise

In this section, a possible method for comparing signal and noise is examined. It becomes clear that finding the precise difference between both sets is difficult due to the combination of the three polarizations *East*, *North* and *vertical* for each set. In order to simplify the comparison of the pure signal and the signal with noise, a Hilbert transform can be used to calculate the respective analytic signals.

In signal processing, an analytic signal is a complex function of time, where the imaginary part is given by the Hilbert transform of the real-valued signal. It is complex differentiable and thus contains no negative frequencies [43]. The analytic signal $x_a(t)$ can be described as

$$x_a(t) = x(t) + i \cdot h\left(x(t)\right) \tag{5.1}$$

where $x(t)$ is the real-valued signal and $h\left(x(t)\right)$ is the Hilbert transform of $x(t)$. In this thesis, the analytic signals for each polarization are calculated using PYTHON's `scipy.signal.hilbert` function.

Finally, the envelope $H(t)$ for each set of traces is calculated as

$$H(t) = \sqrt{h\left(x_E(t)\right)^2 + h\left(x_N(t)\right)^2 + h\left(x_V(t)\right)^2} \tag{5.2}$$

where $h\left(x_j(t)\right),\ j \in (E, N, V)$ is the Hilbert transform for each polarization. Figure 21 shows the Hilbert envelopes for the pure simulated signal peak in the upper plot and the simulated signal peak with added measured noise in the lower plot.

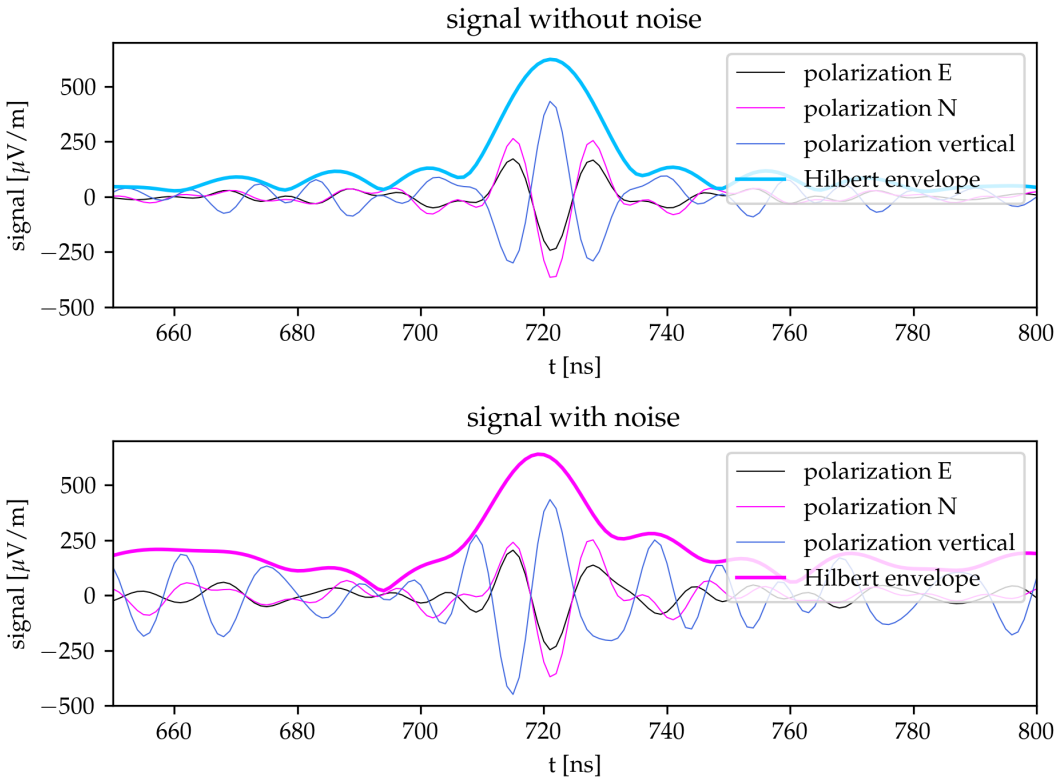

Figure 21: The zoomed-in traces of the signal peak of a simulated event for one station without noise (top) and with noise (bottom). The polarization in eastern direction is shown in blue, the polarization in northern direction is shown in pink and the vertical polarization is shown in black. The Hilbert envelope of the signal without noise is shown in blue and the Hilbert envelope of the signal with noise is shown in pink.

5.2.2 Time Differences

In preparation for creating a method for noise rejection, a technique to distinguish a noise pulse from a signal pulse has to be defined first. So far, the difference in signal timing has been considered for this in the existing methods [42]. Using the Hilbert envelopes calculated in the previous section, it makes sense to determine the

offset between the respective signal peaks in order to enhance and extend the methods for noise rejection.

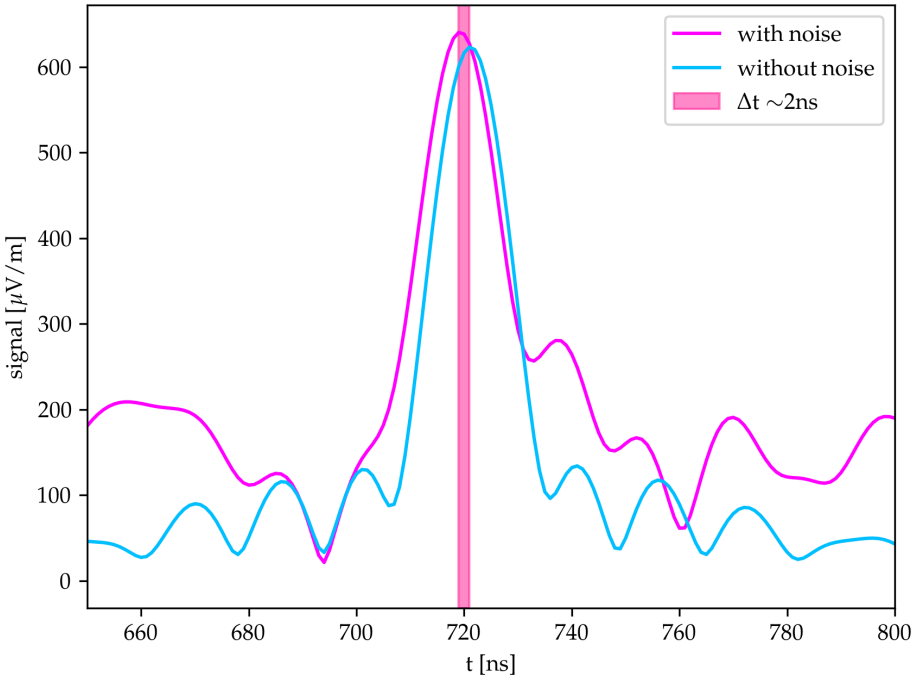

Figure 22: A visualization of the time difference between a signal with noise and a pure signal without noise. The Hilbert envelope of the signal without noise is shown in blue and the Hilbert envelope of the signal with noise is shown in pink. The time difference between the two is marked by a pink band and amounts to approximately 2 ns in this example.

As can be seen in Fig. 22, there is a notable offset between the time of the pure signal peak without noise and the time of the signal peak with noise. This time difference will be applied as distinguishing parameter and it will be referred to as Δt in the following. It can be obtained from Offline as

$$\Delta t \texttt{ = eSignalTime - eTraceStartTime - eSimPulseSignalTime} \tag{5.3}$$

where the `eSignalTime` describes the time at which the signal was found relative to the event start time, the `eTraceStartTime` describes the start time of the trace with respect to the event time and the `eSimPulseSignalTime` describes the time at which the simulated pulse was found relative to the trace start time. The value of Δt varies for each station of every event and thus has to be examined further in order to determine sensible thresholds.

5.2.3 Signal Stations

Generally, signal-dominated stations are classified by having a pulse, which means that the signal intensity is above a certain threshold, and by not being rejected by any of the noise rejection modules in Offline. For the purpose of this thesis, signal-dominated stations will be referred to as *signal stations*. Figure 23 shows the Gaussian fits to the histograms of the Δt distributions for LPDA and butterfly signal stations. Signal stations are not grouped by events and thus considered independently.

Ideally, the time difference between the signal and noise of one station would be zero. As can be seen in Fig. 23, that is not exclusively the case. The Δt distributions peak around zero, which shows correctly identified signals, but large Δt are likely to indicate noise pulses that are falsely identified as signals. The Gaussian fit parameters are listed in Tab. 1. The values confirm that the distributions for both antenna types peak around zero.

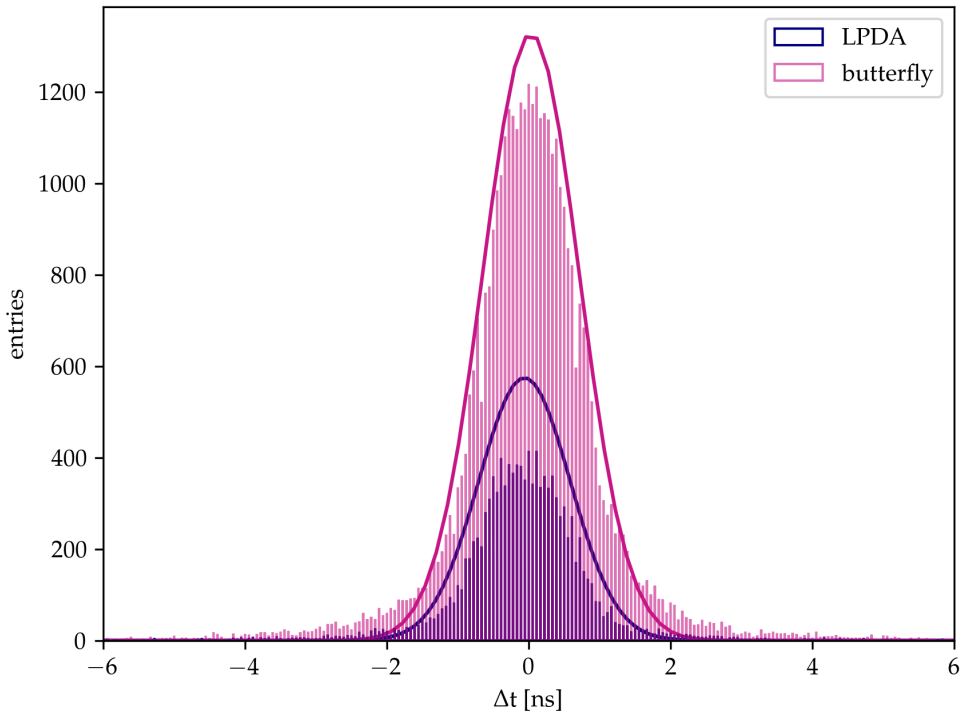

Figure 23: The Δt distribution together with a Gaussian fit for LPDA (blue) and butterfly (pink) signal stations.

	μ [ns]	σ [ns]
LPDA	-0.006 ± 0.003	0.065 ± 0.003
butterfly	0.003 ± 0.002	0.068 ± 0.002

Table 1: The Gaussian fit results for the Δt distribution for LPDA and butterfly signal stations.

Signal Classification Intervals

Finally, after considering all aspects that are included in the classification of Δt, a choice on its best possible value has to be made. The idea is to use Δt on the simulations to evaluate and improve the

different rejection reasons. Since there is no Δt for measured data, it can only be used to evaluate how many stations are correctly rejected in simulations. The better the method works for simulated data, the better it will also perform for measured data.

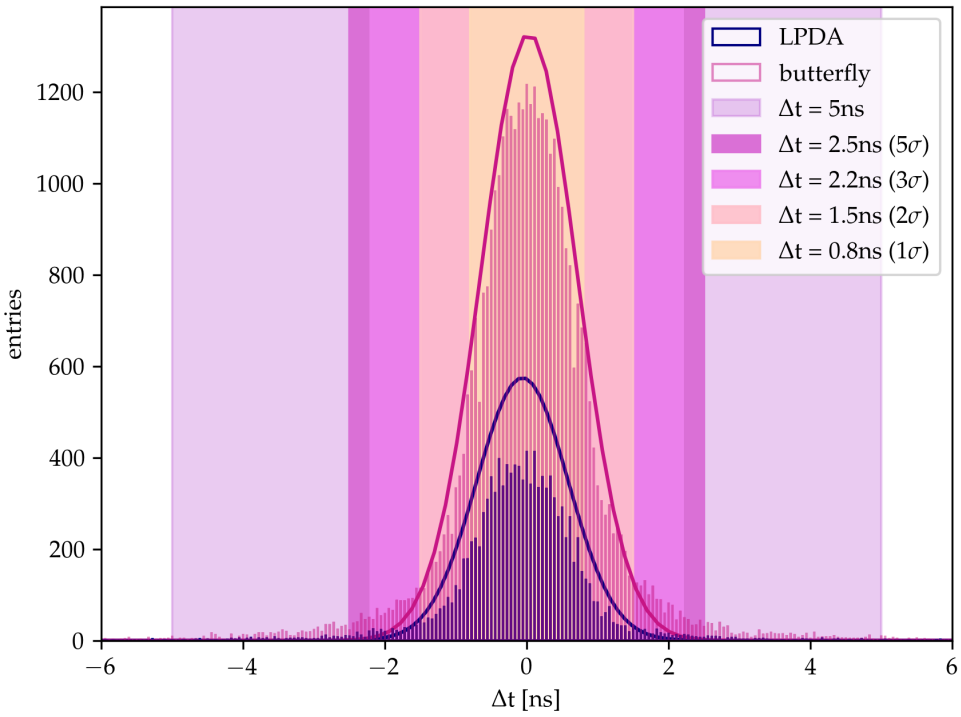

Figure 24: The Δt distribution for LPDA (blue) and butterfly (pink) signal stations. The intervals for $\Delta t \in [-5, 5]$ ns, $\Delta t_{5\sigma} \in [-2.5, 2.5]$ ns, $\Delta t_{3\sigma} \in [-2.2, 2.2]$ ns, $\Delta t_{2\sigma} \in [-1.5, 1.5]$ ns and $\Delta t_{2\sigma} \in [-0.8, 0.8]$ ns are marked in different colors.

Figure 24 shows different Δt intervals on the Δt distribution for signal stations from before (cf. Fig. 23). It visualizes how much of the peak is included for each value of Δt. In the existing methods, $\Delta t \in [-5, 5]$ ns is used [42]. Since the intervals of both $\Delta t_{3\sigma} \in [-2.2, 2.2]$ ns and $\Delta t_{1\sigma} \in [-1.5, 1.5]$ ns include the entire signal peak and less of the background in comparison to $\Delta t = 5$ ns, it can be

assumed that the existing methods still include many stations that are falsely classified as signal stations.

Finally, the choice of the best possible value for Δt can be made using Bayes' theorem [44]

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)} \tag{5.4}$$

With $P(A)$ the probability of signal stations total, $P(B)$ the probability of stations within a given interval of Δt and $P(B|A)$ the probability of signal stations within the same interval of Δt, eq. 5.4 yields the probability P_{signal} that a station within this interval of Δt is a true signal station. The resulting probabilities for the three intervals shown in Fig. 24 are listed in Tab. 2.

Δt_{signal} [ns]	P_{signal}
$\in [-0.8, 0.8]$	$\sim 99.95\%$
$\in [-1.5, 1.5]$	$\sim 99.90\%$
$\in [-2.2, 2.2]$	$\sim 99.86\%$
$\in [-2.5, 2.5]$	$\sim 99.84\%$
$\in [-5.0, 5.0]$	$\sim 99.71\%$

Table 2: The Bayesian probabilities that a station is a signal station for three different values of Δt.

The 1σ interval of $\Delta t_{1\sigma} \in [-0.8, 0.8]$ ns yields the highest probability that a station within the interval is correctly classified as signal station. However, it contains very low statistics and thus little reliability compared to larger intervals. With a probability of $\sim 99.90\%$ for a station within the interval being a correctly classi-

fied signal station, the 2σ interval of $\Delta t_{2\sigma} \in [-1.5, 1.5]$ ns offers the most sensible choice.

5.2.4 Noise Background

Noise-dominated stations are classified by having a pulse, which means that the signal intensity is above a certain threshold, and by being rejected by the noise rejection modules in Offline. For the purpose of this thesis, noise-dominated stations will be referred to as *noise stations*. For analytical purposes, noise traces were obtained by substituting simulated traces with randomly selected measured noise when reconstructing the peak time. These noise traces are shown in Fig. 25. Analogously to the distribution for signal stations, noise stations are not grouped by events and thus considered independently.

Since the noise traces that are added to the simulations are selected randomly, they are different for each reconstruction and are thus not reproducible. However, they all share a similar shape and time range. Especially the asymmetry of the peak with its shift towards the negative time values is an important aspect to take note of.

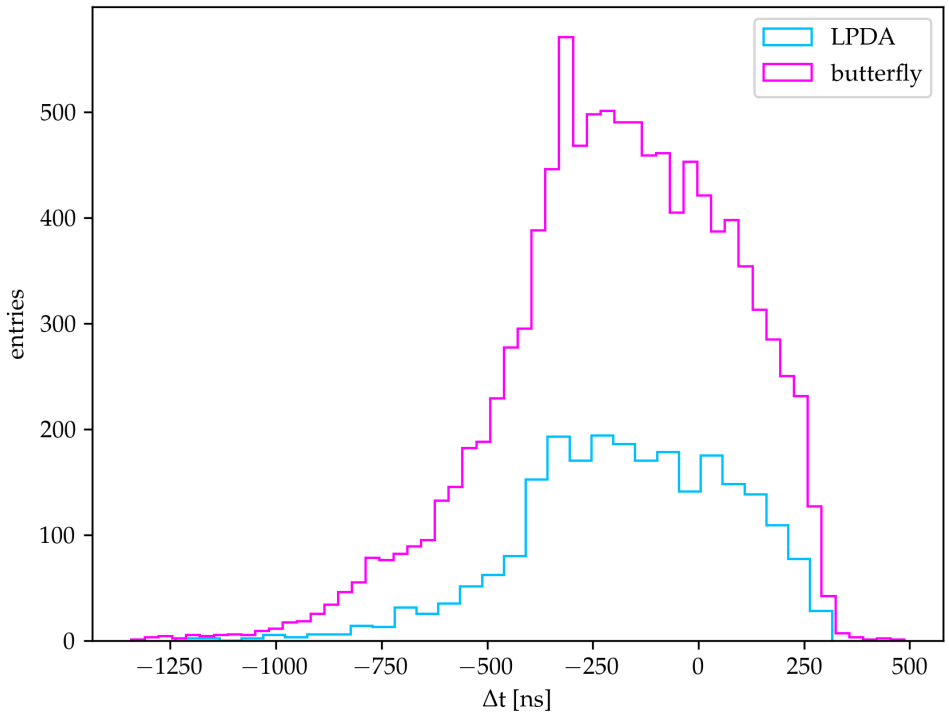

Figure 25: The Δt distribution of measured noise traces, which were obtained by substituting simulated traces with randomly selected measured noise when reconstructing the peak time.

Noise Classification Intervals

Currently, the existing methods do not apply separate classification intervals for signal and noise [42]. Further analysis shows that applying different intervals for signal and noise can significantly reduce the number of stations that are falsely classified.

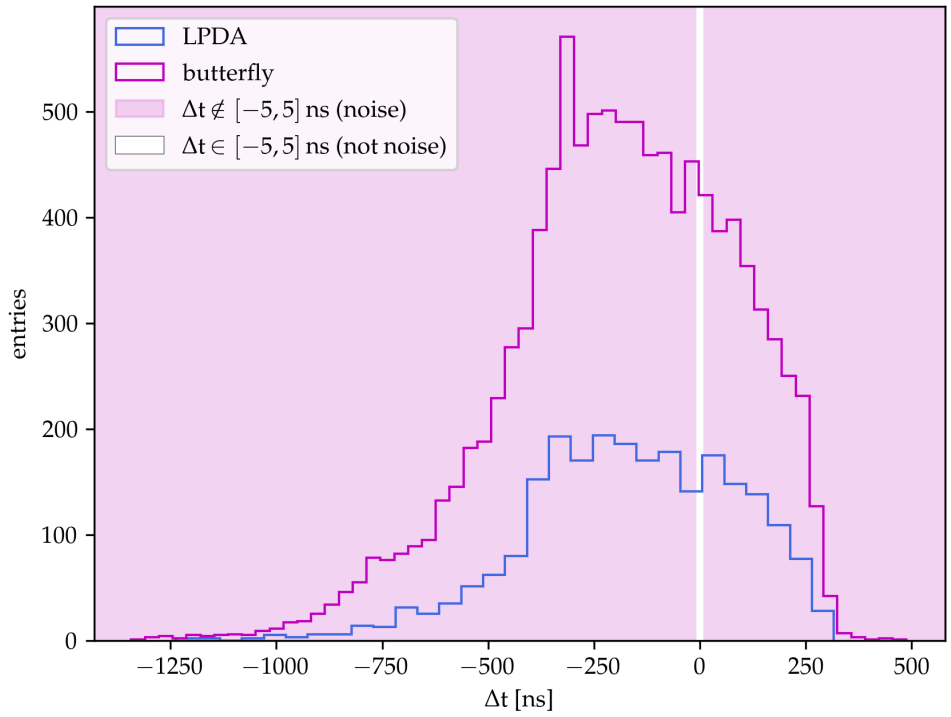

Figure 26: The Δt distribution for LPDA (blue) and butterfly (pink) noise stations. The noise area for $\Delta t \notin [-5, 5]$ ns is marked in purple.

Figure 26 shows the Δt distribution for noise traces from before (cf. Fig. 25) with a noise area of $\Delta t \notin [-5, 5]$ ns. It visualizes how much of the distribution is included for each value of Δt.

The choice of the best possible value for Δt can again be made using Bayes' theorem (eq. 5.4). Here, $P(A)$ is the probability of noise stations total, $P(B)$ is the probability of all stations within the noise area and $P(B|A)$ is the probability of noise stations within the noise area, which yields a probability P_{noise} that a station within the given noise area is a genuine noise station.

Δt_{noise} [ns]	P_{noise}
$\notin [-0.8, 0.8]$	$\sim 45.64\%$
$\notin [-1.5, 1.5]$	$\sim 76.72\%$
$\notin [-2.2, 2.2]$	$\sim 78.83\%$
$\notin [-2.5, 2.5]$	$\sim 81.38\%$
$\notin [-5.0, 5.0]$	$\sim 91.61\%$

Table 3: The Bayesian probabilities that a station is a noise station for three different values of Δt.

5.3 Bayesian Probabilities for Signal and Noise

In order to summarize the results from Secs. 5.2.3-5.2.4, Fig. 27 shows a combined Δt distribution for signal and noise, including intervals for the respective Δt thresholds.

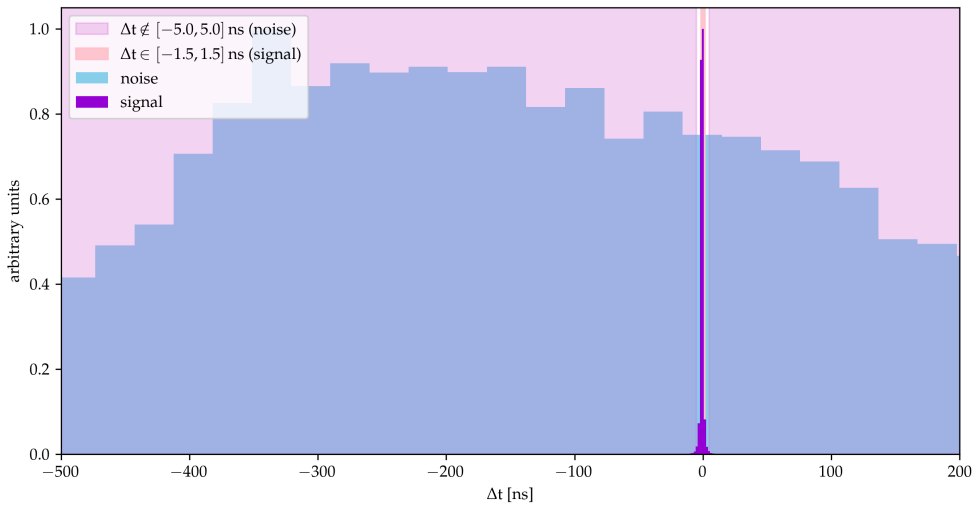

Figure 27: The Δt distributions for LPDA signal (purple) and noise (blue). The signal interval for $\Delta t_{2\sigma} \in [-1.5, 1.5]$ ns and the noise area for $\Delta t \notin [-5, 5]$ ns are indicated in different colors. Note that both distributions are normalized and that the y-axis shows arbitrary units.

The purpose of this figure is to simply visualize the proportions between the signal and noise distributions and highlight the different Δt thresholds. Therefore, each distribution is normalized and the y-axis shows arbitrary units. A second visualization, where the same distribution is zoomed-in on the signal peak, is shown in Appendix B.

Finally, Fig. 28 shows the Bayesian probabilities for signal and noise-dominated stations for different values of Δt.

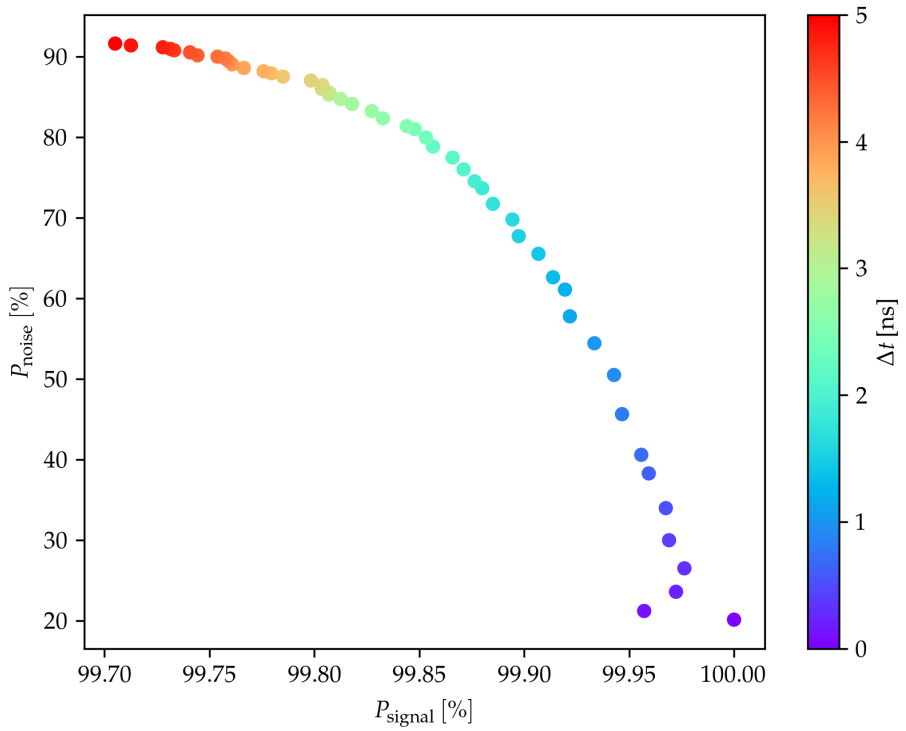

Figure 28: A visualization of the Bayesian probabilities for signal and noise-dominated stations for different values of Δt. The Bayesian probability P_{noise} for finding a noise station vs. the Bayesian probability P_{signal} for finding a signal station is shown. The values of Δt are marked by color in a range of $[0.0, 5.0]$ ns.

It can be seen that the smaller the interval of Δt, the higher the probability of finding a signal station and the larger the interval of Δt, the higher the probability of finding a noise station. At very small values of Δt, fluctuations are due to the low statistics.

5.4 Rejection Rates

Finally, the performance of the adapted methods for noise rejection can be evaluated. The performances can be described by the *correct rejection rate* and *false rejection rate* for each rejection module.

The correct rejection rate can be calculated as *true positive rate*, TPR:

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{5.5}$$

where TP denotes the *true positive* and FN the *false negative*. In this case, the true positive is the number of correctly rejected stations, i.e. the number of noise stations which are correctly identified as noise stations, and the false negative is the number of falsely not rejected stations, i.e. the number of noise stations that are falsely identified as signal stations.

The false rejection rate can be calculated as *false positive rate*, FPR:

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}} \tag{5.6}$$

where FP denotes the *false positive* and TN the *true negative*. Here, the false positive is the number of correctly not rejected stations, i.e. the number of signal stations that were correctly identified as signal stations, and the false positive is the number of falsely rejected

stations, i.e. the number of signal stations falsely identified as noise stations.

The four noise rejection modules included in $\overline{\text{Offline}}$ are run in the order Pulse Shape, Polarization, Cluster Finder and Top Down. This has the effect that the modules do not reject stations that have already been rejected by previous modules. The order has been determined as the best possible rejection sequence in previous analyses [42].

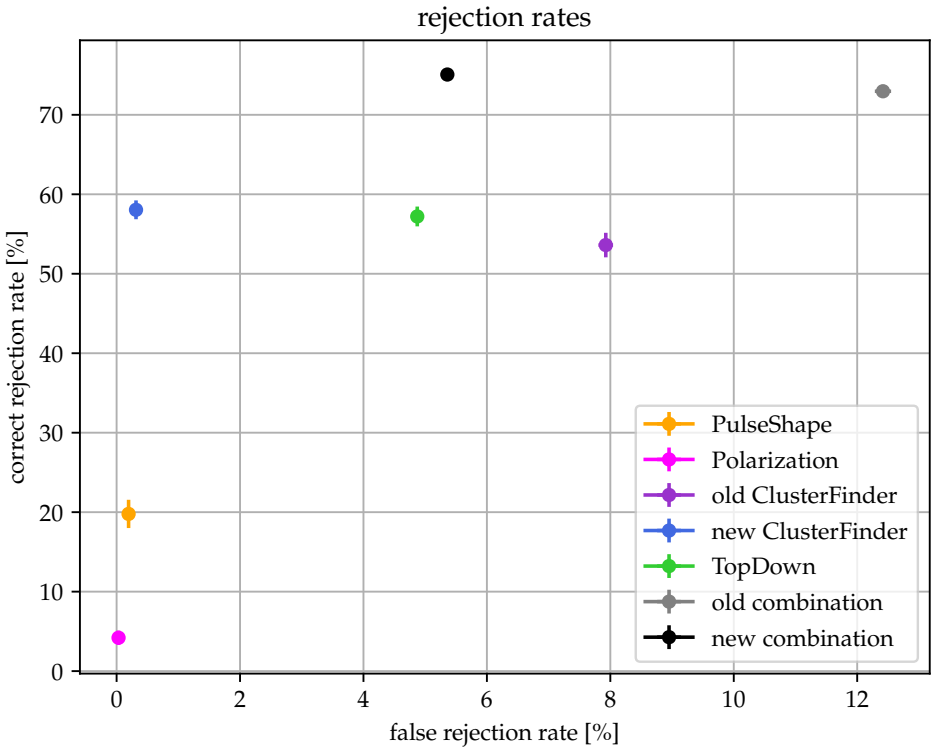

Figure 29: A visualization of the rejection rates of the different noise rejection modules. The false rejection rate in percent is shown on the x-axis, the true rejection rate in percent is shown on the y-axis. The rejection rates of the noise rejection modules are marked in different colors. A combination of the "old" Cluster Finder and the other three modules is marked in black and a combination of the "new" Cluster Finder and the other three modules is marked in gray.

Figure 29 shows the false rejection rates and the correct rejection rates for each noise rejection module. The values are listed in Tab. 4 in Appendix A. The values for the other modules besides the Cluster Finder simply serve as references for the overall performance of the noise rejection method.

The performance of the "old" and "new" versions of the Cluster Finder can be compared via their rejection rates. The false rejection rate of the Cluster Finder noise rejection module was reduced from $\sim 7.9\%$ to $\sim 0.3\%$ and the correct rejection rate was increased from $\sim 53\%$ to $\sim 58\%$.

5.4.1 Event Example

When using the old version of the Cluster Finder, events which are dominated by AERA III stations are lost due to the large distances between these stations. Figure 30 shows an example of an event that is reconstructed using the modified Cluster Finder. It contains multiple AERA III stations, which would have been rejected by the old Cluster Finder. The traces measured by one of the AERA III stations involved in the same event is shown in Fig. 31. The shape of the signal pulse indicates that this particular station should be classified as signal station.

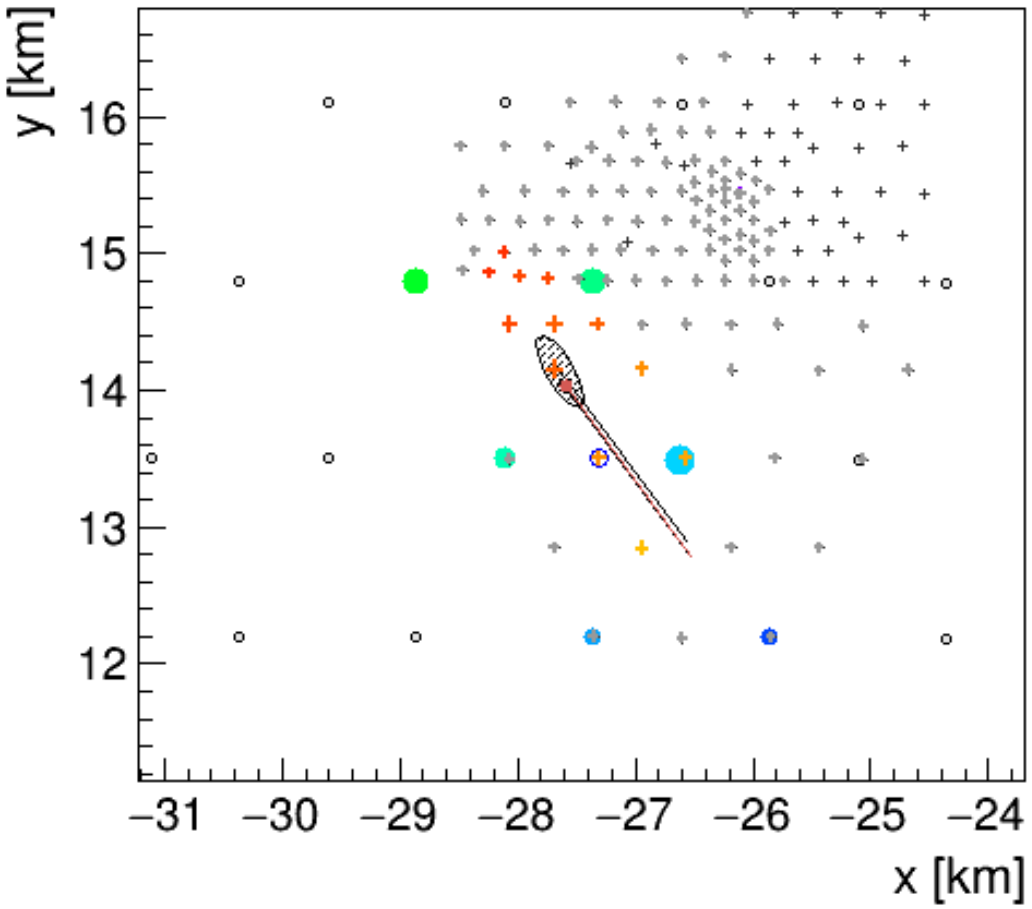

Figure 30: An example of an event which contains AERA III stations, reconstructed using the modified Cluster Finder. Gray circles show SD stations and plus-symbols show AERA stations. Triggered AERA stations are marked in red and orange according to the intensity of the signal. For SD stations, the size of the circle indicates the intensity of the measured signal, while the color range from dark blue to light green indicates the time of measurement. The shower axis is shown as a gray line and the shower footprint is represented by a gray ellipsis. AERA station number 128 is encircled in dark blue.

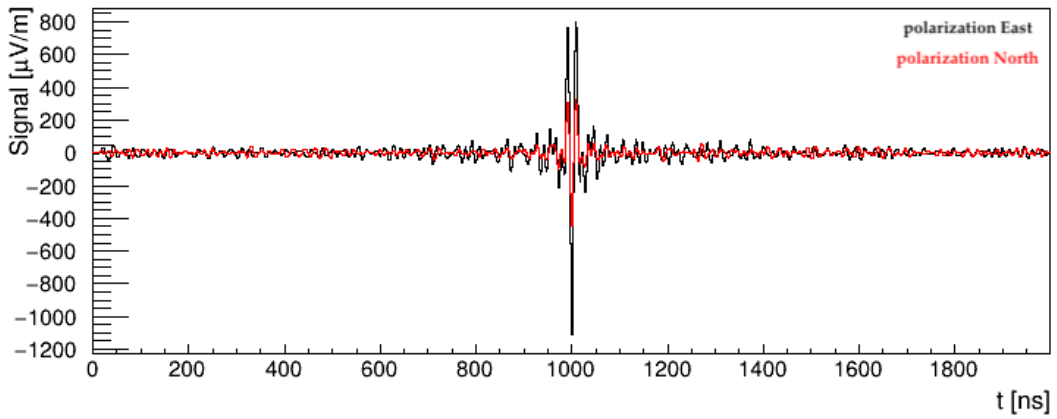

Figure 31: An example of a signal trace measured by AERA III station number 128. The eastern polarization is shown in black and the northern polarization is shown in red.

Chapter 6

Conclusion

The Auger Engineering Radio Array (AERA) is an enhancement of the Pierre Auger Observatory, which is designed to detect radio signals emitted by extensive air showers. The main scientific goals of AERA are the calibration of the radio emission from air showers, demonstrating the energy, angular and mass resolutions of the radio technique and the measurement of the cosmic ray composition in the range of both Galactic and extragalactic cosmic rays. Due to their differences in antenna spacing, AERA phases I-II are more convenient for detecting vertical air showers, whereas AERA phase III is best suited for detecting inclined air showers. Since the existing methods are only optimized for AERA phases I-II, the accuracy of radio reconstruction for inclined air showers has room for improvement.

In this thesis, enhancements have been made to the AERA analysis using the $\overline{\text{Off}}\underline{\text{line}}$ reconstruction framework with regard to radio reconstruction for inclined air showers. Various parameters for identifying noise-dominated stations have been examined and modified in order to improve the performance of the noise rejection methods

in Offline. The parameter Δt, which compares the offset between a simulated pure signal and the same signal with added noise, was used to verify the rejections done by Offline. The magnitude of the parameter indicates the probability that a noise pulse is falsely identified as signal. In particular, the Cluster Finder module, which rejects lonely stations or isolated clusters that are far away from the shower axis, was modified during the work carried out for this thesis. Furthermore, the performance of the modified Cluster Finder can be described in terms of a false rejection rate and a correct rejection rate. The adaptations yield a false rejection rate of $(0.31 \pm 0.03)\%$ and a correct rejection rate of $(58.05 \pm 1.18)\%$ for the Cluster Finder noise rejection module. A comparison to the other noise rejection modules in Offline shows that this modified version of the Cluster Finder offers the best performance values for inclined air showers. Especially the Top Down module can still be improved. However, due to time constraints, the adaptation of a second module could not be included in the context of this thesis.

The improved analytical performance of the Cluster Finder module will lead to a higher precision and accuracy in future research regarding the radio emissions of air showers using Offline. Furthermore, the developed methods can easily be conferred onto similar analyses regarding both the effects of antenna spacing on measurements, as well as the detection of signal among noise pulses.

Appendix A

Tables

module	false rejection rate [%]	correct rejection rate [%]
Pulse Shape	0.19 ± 0.03	19.78 ± 1.61
Polarization	0.03 ± 0.00	4.20 ± 0.91
"old" Cluster Finder	7.93 ± 0.12	53.61 ± 1.55
"new" Cluster Finder	0.31 ± 0.03	58.05 ± 1.18
Top Down	4.87 ± 0.09	57.21 ± 1.23

Table 4: The rejection rates of the different noise rejection modules.

Appendix B

Figures

Simulations

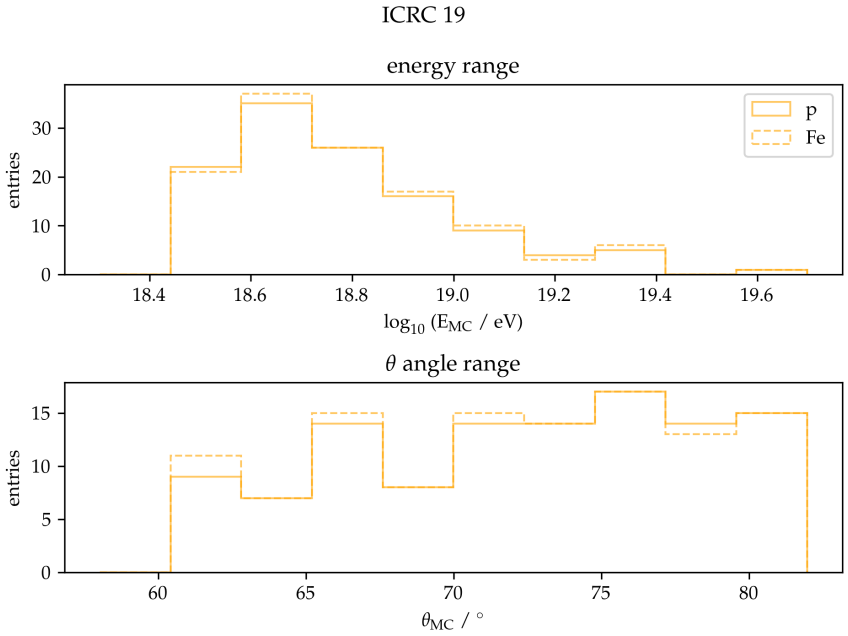

Figure 32: The energy range (top) and zenith angle range (bottom) of simulation A from [33] used in this thesis for proton (solid) and iron (dashed).

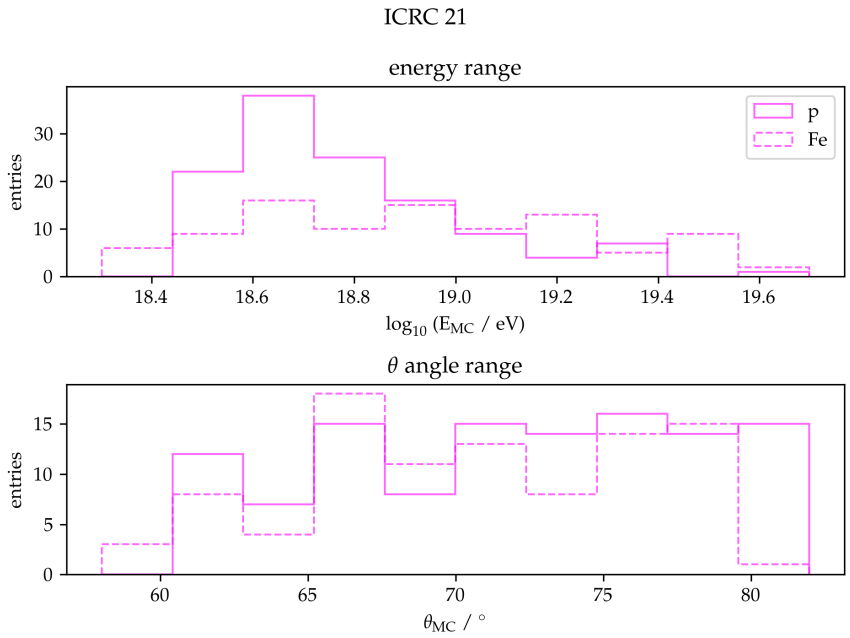

Figure 33: The energy range (top) and zenith angle range (bottom) of simulation B from [34] used in this thesis for proton (solid) and iron (dashed).

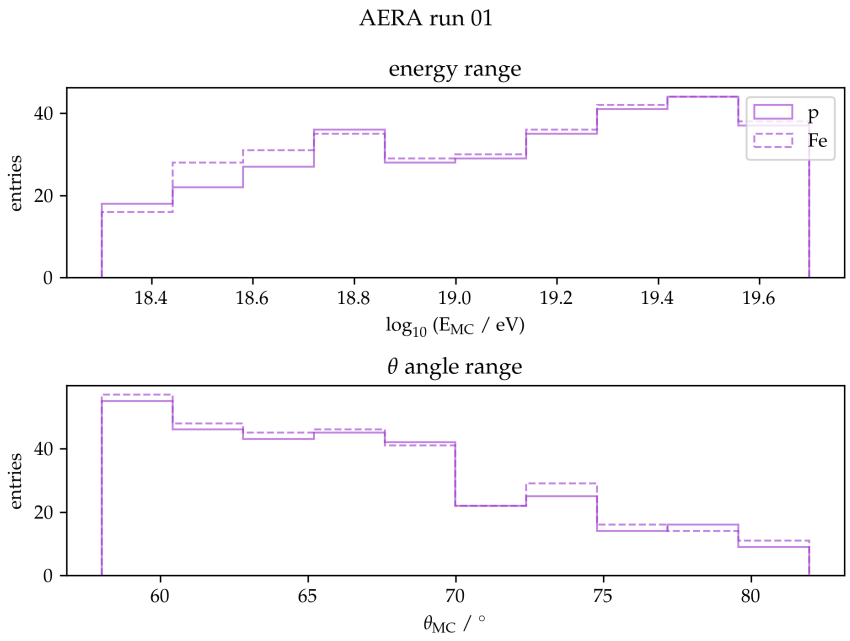

Figure 34: The energy range (top) and zenith angle range (bottom) of simulation C from [35] used in this thesis for proton (solid) and iron (dashed).

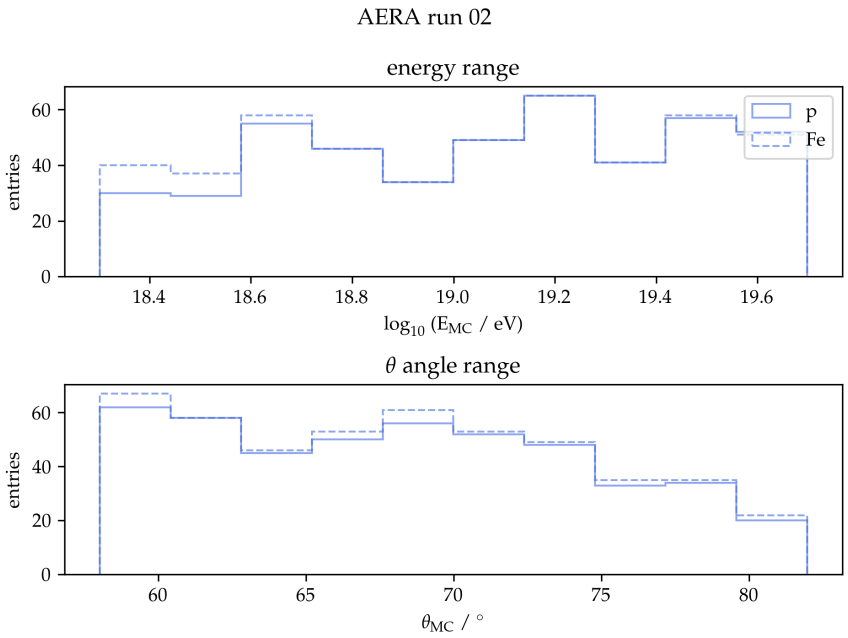

Figure 35: The energy range (top) and zenith angle range (bottom) of simulation D from [36] used in this thesis for proton (solid) and iron (dashed).

Classification Intervals

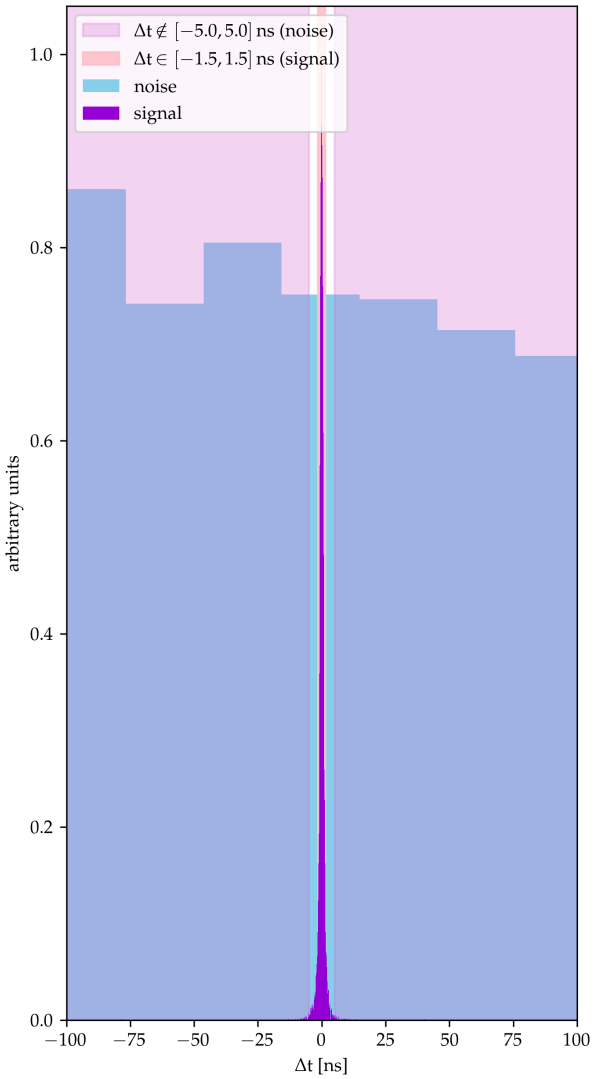

Figure 36: The Δt distributions for LPDA signal (purple) and noise (blue), zoomed in on the signal peak. The signal interval for $\Delta t_{2\sigma} \in [-1.5, 1.5]$ ns and the noise area for $\Delta t \notin [-5, 5]$ ns are indicated in different colors. Note that both distributions are normalized and that the y-axis shows arbitrary units.

Appendix C

Offline Configuration

The Offline used in this thesis was based on the main version from the framework repository on `GitLab`[1] retrieved on 2022-02-10[2]. Excerpts from the configured Offline files are listed here.

RdClusterFinder.xml

This version of the `RdClusterFinder.xml` file includes the developed modifications regarding the different AERA phases.

```
1 <?xml version="1.0" encoding="iso-8859-1"?>
2 <!-- Configuration of Module RdClusterFinder -->
3 <!-- In the RdClusterFinder tag we define the path and name of
      the schema file
4 (RdClusterFinder.xsd) used to validate this xml file. When make
      is invoked, /offline/install/share/auger-offline/config gets
      replaced by the path to the directory in the install area
      where
5 configuration files are copied. -->
6
```

[1] https://gitlab.com/auger-observatory/offline/offline

[2] Commit SHA: 50f033959bd2ac694ffec0b1c64d86a325e42005

```
7 <RdClusterFinder xmlns:xsi="http://www.w3.org/2001/XMLSchema-
    instance"
8                  xsi:noNamespaceSchemaLocation='/offline/
    install/share/auger-offline/config/RdClusterFinder.xsd'>
9
10  <!-- verbosity 0=none, 1=final result, 2=intermediate,
11  3=most (obscure), 4=minuit included -->
12  <InfoLevel> 1 </InfoLevel>
13
14  <!-- maximum discontinuity in distribution of the euclidean
     distance between hit stations and Sd axis -->
15  <MaximumDiscontinuity unit="m"> 500.0 </MaximumDiscontinuity>
16
17  <!-- select whether lonely stations should be rejected or not
     (0: don't reject, 1: reject) -->
18  <RejectLonelyStations> 0 </RejectLonelyStations>
19  <!-- if no other signal station is in this distance the current
      station will be flagged as lonely -->
20
21  <!-- Lonely for AERA Phases 1 and 2 -->
22  <LonelyIfNoneInDistance unit="meter"> 400 </
     LonelyIfNoneInDistance>
23  <!-- if not more than one station is in this distance the
     current station will be flagged as lonely -->
24  <LonelyIfOneInDistance unit="meter"> 800 </
     LonelyIfOneInDistance>
25
26  <!-- Lonely for AERA Phase 3 -->
27  <LonelyIfNoneInDistancePhase3 unit="meter"> 760 </
     LonelyIfNoneInDistancePhase3>
28  <!-- if not more than one station is in this distance the
     current station will be flagged as lonely -->
29  <LonelyIfOneInDistancePhase3 unit="meter"> 1520 </
     LonelyIfOneInDistancePhase3>
30
31  <!-- If set to 0, rejected stations will be treated as silent
     stations. If set to 1 manually rejected stations
32  and stations without noise data (for simulations) are treated
     as signal stations. If set to 2 all rejected
33  stations will be treated as signal stations -->
34  <IgnoreRejectedStations> 2 </IgnoreRejectedStations>
35 </RdClusterFinder>
```

RdClusterFinder.xsd

This version of the `RdClusterFinder.xsd` file includes the parameters for the developed modifications regarding the different AERA phases.

```
1 <xs:schema xmlns:xs='http://www.w3.org/2001/XMLSchema'
2            xmlns:auger="http://www.auger.org/schema/types">
3 <!-- Following instruction includes data types defined in
     AugerSchemaTypes.xsd.  These
4 include types useful for Auger applications, like doubleWithUnit,
      listOfDoublesWithUnits, etc.
5 NOTE the definition of the auger namespace in the previous line.
       -->
6
7 <xs:import namespace="http://www.auger.org/schema/types"
8            schemaLocation="/offline/install/share/auger-offline/
     config/AugerSchemaTypes.xsd"/>
9
10 <xs:element name="RdClusterFinder">
11   <xs:complexType>   <!-- specifies that RdClusterFinder is an
      element comprising sub-elements and possibly attributes -->
12     <xs:all>         <!-- indicates that all the following sub-
     elements must be present in RdClusterFinder -->
13
14        <!-- types that have xs:something come pre-defined in the
     standard xml schema .  xs just
15             defines the namespace where they live (a bit like the
      std namespace).  -->
16        <xs:element name="InfoLevel" type="auger:InfoLevelType" />
17        <xs:element name="MaximumDiscontinuity" type="
     auger:doubleWithUnit"/>
18        <xs:element name="RejectLonelyStations" type="
     xs:nonNegativeInteger" />
19        <xs:element name="LonelyIfNoneInDistance" type="
     auger:doubleWithUnit"/>
20        <xs:element name="LonelyIfOneInDistance" type="
     auger:doubleWithUnit"/>
21        <xs:element name="LonelyIfNoneInDistancePhase3" type="
     auger:doubleWithUnit"/>
```

```
22        <xs:element name="LonelyIfOneInDistancePhase3" type="
    auger:doubleWithUnit"/>
23       <xs:element name="IgnoreRejectedStations" type="
    xs:nonNegativeInteger"/>
24      </xs:all>
25   </xs:complexType>
26 </xs:element>
27 </xs:schema>
```

bootstrap.xml - new

This version of the `bootstrap.xml` file applies the developed modifications regarding the different AERA phases, using *LonelyRejection* distances of 400 and 800 m for phases I and II and distances of 760 and 1520 m for phase III.

```
1     <configLink id="RdClusterFinder">
2       <RdClusterFinder>
3         <RejectLonelyStations> 1 </RejectLonelyStations>
4         <LonelyIfNoneInDistance unit="meter"> 400 </
    LonelyIfNoneInDistance>
5         <LonelyIfOneInDistance unit="meter"> 800 </
    LonelyIfOneInDistance>
6   <LonelyIfNoneInDistancePhase3 unit="meter"> 760 </
    LonelyIfNoneInDistancePhase3>
7         <LonelyIfOneInDistancePhase3 unit="meter"> 1520 </
    LonelyIfOneInDistancePhase3>
8         <IgnoreRejectedStations> 2 </IgnoreRejectedStations>
9       </RdClusterFinder>
10    </configLink>
```

bootstrap.xml - standard

This excerpt uses the same version of the `bootstrap.xml` file as above. Here, the *LonelyRejection* distances for phase III are overwritten with 400 and 800 m. This can be used to compare performance the modified Cluster Finder to previous versions.

```
1     <configLink id="RdClusterFinder">
2       <RdClusterFinder>
3         <RejectLonelyStations> 1 </RejectLonelyStations>
4         <LonelyIfNoneInDistance unit="meter"> 400 </
      LonelyIfNoneInDistance>
5         <LonelyIfOneInDistance unit="meter"> 800 </
      LonelyIfOneInDistance>
6    <LonelyIfNoneInDistancePhase3 unit="meter"> 400 </
      LonelyIfNoneInDistancePhase3>
7         <LonelyIfOneInDistancePhase3 unit="meter"> 800 </
      LonelyIfOneInDistancePhase3>
8         <IgnoreRejectedStations> 2 </IgnoreRejectedStations>
9       </RdClusterFinder>
10    </configLink>
```

List of Figures

List of Tables

Acronyms

AERA **A**uger **E**ngineering **R**adio **A**rray.

AMIGA **A**uger **M**uon and **I**nfill **G**round **A**rray.

CoREAS **CO**RSIKA-based **R**adio **E**mission from **A**ir **S**howers.

CORSIKA **CO**smic **R**ay **SI**mulation for **KA**SCADE.

CR **C**osmic **R**ay.

EAS **E**xtensive **A**ir **S**hower.

FD **F**luorescence **D**etector.

GZK **G**reisen–**Z**atsepin–**K**uzmin.

HAS **H**orizontal **A**ir **S**hower.

HEAT **H**igh **E**levation **A**uger **T**elescope.

LPDA **L**ogarithmic **P**eriodic **D**ipole **A**ntenna.

PMT **P**hoto**M**ultiplier **T**ube.

RFI **R**adio **F**requency **I**nterference.

SD **S**urface **D**etector.

Bibliography

[1] V.F. Hess. "Über Beobachtungen der durchdringenden Strahlung bei sieben Freiballonfahrten". In: *Phys. Z.* 13 (1912), pp. 1084–1091.

[2] The Pierre Auger Collaboration. "The Pierre Auger Cosmic Ray Observatory". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 798 (2015), pp. 172–213. DOI: https://doi.org/10.1016/j.nima.2015.06.058.

[3] F. G. Schröder. "Radio detection of high-energy cosmic rays with the Auger Engineering Radio Array". In: *Nuclear Instruments and Methods in Physics Research* 824 (July 2016), pp. 648–651. DOI: 10.1016/j.nima.2015.08.047.

[4] H.V. Klapdor-Kleingrothaus and K.Zuber. *Teilchenastrophysik*. B.G. Teubner Stuttgart, 1997.

[5] S.E.S Ferreira and M.S Potgieter. "Galactic cosmic rays in the heliosphere". In: *Advances in Space Research* 34.1 (2004), pp. 115–125. DOI: https://doi.org/10.1016/j.asr.2003.02.057.

[6] "Features of the Energy Spectrum of Cosmic Rays above 2.5×10^{18} eV Using the Pierre Auger Observatory". In: *Phys. Rev. Lett.* 125 (12 Sept. 2020), p. 121106. DOI: 10.1103/PhysRevLett.125.121106.

[7] M. Tanabashi et al. "Review of Particle Physics". In: *Phys. Rev. D* 98 (3 Aug. 2018), p. 030001. DOI: 10.1103/PhysRevD.98.030001.

[8] A. Zilles. "Introduction to Cosmic Rays and Extensive Air Showers". In: *Emission of Radio Waves in Particle Showers: Validation of Microscopic Simulations with the SLAC T-510 Experiment and their Potential in the*

Future Square Kilometre Array. Springer International Publishing, 2017, pp. 1–13. ISBN: 978-3-319-63411-1. DOI: 10.1007/978-3-319-63411-1_1.

[9] K.-H. Kampert and P. Tinyakov. "Cosmic rays from the ankle to the cutoff". In: *Comptes Rendus Physique* 15.4 (2014), pp. 318–328. DOI: https://doi.org/10.1016/j.crhy.2014.04.006.

[10] K. Greisen. "End to the Cosmic-Ray Spectrum?" In: *Phys. Rev. Lett.* 16 (17 Apr. 1966), pp. 748–750. DOI: 10.1103/PhysRevLett.16.748.

[11] J. Matthews. "A Heitler model of extensive air showers". In: *Astroparticle Physics* 22.5 (2005), pp. 387–397. DOI: https://doi.org/10.1016/j.astropartphys.2004.09.003.

[12] M. V. S. Rao and B. V. Sreekantan. *Extensive Air Showers*. WORLD SCIENTIFIC, 1998. DOI: 10.1142/3307.

[13] Z. Szadkowski. *Cosmic Rays*. Rijeka: IntechOpen, Aug. 2018. DOI: 10.5772/intechopen.72533.

[14] O.C. Alkhofer. *Introduction to Cosmic Radiation*. Verlag Carl Theimig, 1975.

[15] T. Stanev. *High Energy Cosmic Rays*. Springer, 2004.

[16] A. Yushkov. "Mass Composition of Cosmic Rays with Energies above $10^{17.2}$ eV from the Hybrid Data of the Pierre Auger Observatory". In: *PoS* ICRC2019 (2019), p. 482. DOI: 10.22323/1.358.0482.

[17] D. Ardouin et al. "Geomagnetic origin of the radio emission from cosmic ray induced air showers observed by CODALEMA". In: *Astroparticle Physics* 31.3 (2009), pp. 192–200. DOI: https://doi.org/10.1016/j.astropartphys.2009.01.001.

[18] G. A. Askaryan. "Excess negative charge of an electron-photon shower and the coherent radio emission from it". In: *Zhur. Eksptl'. i Teoret. Fiz.* Vol. 41 (Aug. 1961). URL: https://www.osti.gov/biblio/4833087.

[19] F. G. Schröder. "Radio detection of cosmic-ray air showers and high-energy neutrinos". In: *Progress in Particle and Nuclear Physics* 93 (2017), pp. 1–68. DOI: https://doi.org/10.1016/j.ppnp.2016.12.002.

[20] C. Grupen. *Einstieg in die Astroteilchenphysik*. 2nd ed. Springer, Berlin, 2018.

[21] Tim Huege. "Radio detection of cosmic ray air showers in the digital era". In: *Physics Reports* 620 (Mar. 2016), pp. 1–52. DOI: 10.1016/j.physrep.2016.02.001.

[22] Thierry Gousset, Olivier Ravel, and Christelle Roy. "Are vertical cosmic rays the most suitable to radio detection?" In: *Astroparticle Physics* 22.1 (Oct. 2004), pp. 103–107. DOI: 10.1016/j.astropartphys.2004.05.004.

[23] D. Heck et al. "CORSIKA: A Monte Carlo Code to Simulate Extensive Air Showers". In: *FZKA Report 6019* (1998). URL: https://digbib.ubka.uni-karlsruhe.de/volltexte/fzk/6019/6019.pdf.

[24] T. Huege, M. Ludwig, and C. W. James. "Simulating radio emission from air showers with CoREAS". In: *AIP Conference Proceedings* 1535.1 (2013), pp. 128–132. DOI: 10.1063/1.4807534.

[25] I. Allekotte et al. "The surface detector system of the Pierre Auger Observatory". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 586.3 (2008), pp. 409–420. DOI: https://doi.org/10.1016/j.nima.2007.12.016.

[26] J. Abraham et al. "The fluorescence detector of the Pierre Auger Observatory". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 620.2 (2010), pp. 227–251. DOI: https://doi.org/10.1016/j.nima.2010.04.023.

[27] D. Veberic. https://web.ikp.kit.edu/darko/auger/auger-array/.

[28] E. M. Holt. "Combined Detection of Muons and Radio Emission of Cosmic-Ray Air Showers". PhD thesis. Karlsruher Institut für Technologie (KIT), 2018. 182 pp. DOI: 10.5445/IR/1000083318.

[29] P Abreu et al. "Antennas for the detection of radio emission pulses from cosmic-ray induced air showers at the Pierre Auger Observatory". In: *Journal of Instrumentation* 7.10 (Oct. 2012), P10011–P10011. DOI: 10.1088/1748-0221/7/10/p10011.

[30] A. F. Nelles. "Radio emission of air showers - The perspective of LOFAR and AERA". PhD thesis. Radboud University Nijmegen, 2014.

[31] J. Pawlowsky. private communication.

[32] J.L. Kelley. "Data acquisition, triggering, and filtering at the Auger Engineering Radio Array". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 725 (Oct. 2013), pp. 133–136. DOI: 10.1016/j.nima.2012.11.153.

[33] Available on iRODS. URL: /pauger/Papers/ICRC2019_AERA_HAS/CoREAS.

[34] Available on iRODS. URL: /pauger/Radio/Simulations/library/HAS-AERA-ICRC-21-energy-scaled.

[35] Available on iRODS. URL: /pauger/Radio/Simulations/library/AERA_HAS_generic_2022/run01.

[36] Available on iRODS. URL: /pauger/Radio/Simulations/library/AERA_HAS_generic_2022/run02.

[37] S. Argirò et al. "The offline software framework of the Pierre Auger Observatory". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 580.3 (Oct. 2007), pp. 1485–1496. DOI: 10.1016/j.nima.2007.07.010.

[38] P. Abreu et al. "Advanced functionality for radio analysis in the Offline software framework of the Pierre Auger Observatory". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 635.1 (2011), pp. 92–102. DOI: https://doi.org/10.1016/j.nima.2011.01.049.

[39] M. Gottowik. "Radio Hybrid Reconstruction and Analysis of Inclined Air Showers with AERA of the Pierre Auger Observatory - Measuring the Hadronic Shower Development and Cosmic Ray Mass Composition". PhD thesis. Bergische Universität Wuppertal, 2021.

[40] O. Kambeitz. "Radio Detection of Horizontal Extensive Air Showers". PhD thesis. Karlsruher Institut für Technologie (KIT), 2016. DOI: 10.5445/IR/1000055758.

[41] I. Antcheva and Bothers. "ROOT —A C++ framework for petabyte data storage, statistical analysis and visualization". In: *Computer Physics Communications* 180.12 (2009), pp. 2499–2512. DOI: https://doi.org/10.1016/j.cpc.2009.08.005.

[42] C. Welling. *Identification of radio signals from cosmic ray induced air showers with the Auger engineering radio array.* MA thesis. 2017.

[43] J.O. Smith. *Mathematics of the Discrete Fourier Transform (DFT) with Audio Applications*. W3K Publishing, 2007. ISBN: 978-0-9745607-4-8.

[44] Bronstein et al. *Taschenbuch der Mathematik*. 9th ed. Europa Lehrmittel, 2013. ISBN: 978-3-8085-5673-3.

Declaration of Authorship

I hereby declare that the thesis submitted is my own unaided work. All direct or indirect sources used are acknowledged as references. This thesis was not previously presented to another examination board and has not been published.

Eidesstattliche Erklärung

Hiermit versichere ich, die vorliegende Arbeit selbstständig und unter ausschließlicher Verwendung der angegebenen Literatur und Hilfsmittel erstellt zu haben. Die Arbeit wurde bisher in gleicher oder ähnlicher Form keiner anderen Prüfungsbehörde vorgelegt und auch nicht veröffentlicht.

Wuppertal, den _____ _____

(Jelena Petereit)