Towards measurement of UHECR with the ARA experiment

Uzair Abdul Latif

B.S Physics, LUMS, 2014

Submitted to the graduate degree program in Department of Physics and Astronomy and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Committee members

Dave Besson, Chairperson

Ian Lewis

Chris Rogan

Kyoungchul Kong

Benjamin Sikes, External Reviewer

Date defended: July 10, 2020

The Dissertation Committee for Uzair Abdul Latif certifies
that this is the approved version of the following dissertation :

Towards measurement of UHECR with the ARA experiment

Dave Besson, Chairperson

Date approved: July 22, 2020

Abstract

The Askaryan Radio Array (ARA) experiment at the South Pole aims to detect GZK (Greisen–Zatsepin–Kuzmin) neutrinos (of energies from some PeV to higher energies) from extra-galactic sources. ARA looks for Askaryan emission in radio-wavelengths (150 to 800 MHz) from showers induced by neutrino-ice molecule collisions in Antarctica. This thesis was aimed at trying to identify the UHECR (Ultra High Energy Cosmic Ray) signal in the ARA stations (or detectors), with each station consisting of radio antennas inside 200 m deep holes in the ice sheet. The CoREAS (CORSIKA-based Radio Emission from Air Showers) simulation was used to simulate radio emissions from cosmic ray air showers and predict voltage signals produced by UHECR in the ARA detectors. UHECR event templates were made for the ARA detectors. These templates were used in a template matching analysis method in order to identify possible UHECR shower candidates in the ARA-2 and ARA-3 "burn" sample data for years 2013 to 2017 (the 'burn', or 'unblinded' sample data set consists of 10% of the full data set and is made by randomly selecting 1 in every 10 events in each data run with the goal of cut-tuning prior to analysis of the complete data sample).

Acknowledgements

This thesis is by no means the product of only my hard work. It took a village – people who provided me with thoughtful feedback throughout, others who provided all kinds support and guidance. Here, I would like to acknowledge some of those individuals. None of this would have been possible without their support.

I am immensely grateful to Dave Besson for giving me a chance to be his graduate student and for being an incredibly supportive supervisor throughout my PhD. I am highly indebted to him for all I have learnt throughout my PhD and for all the research opportunities I got under his tutelage. Dave always helped me push my boundaries and challenge myself which has helped me grow as a scientist. I look forward to collaborating with him and learning from him in all of my future research endeavours.

I am also grateful to my ARA collaborators for teaching me everything that I know about astroparticle physics, radio cosmic ray and neutrino astronomy and ice layers. All the feedback that I got from my collaborators not only improved my research, it made me a better researcher.

I also want to recognise my ARA drill team colleagues James Roth, Robert Young, Thomas Meures and Terry Benson for all that they taught me about drilling, ice sheets and hardware while I was working at the South Pole.

I would like to thank my lab mates Steven Prohira, Alexander Novikov and Brendon Madison and my roommate Majed Sofiani for their collegial support. I especially appreciate the discussions I had with my friend, colleague and former roommate Haider Al-Hazmi who was always there to listen and help me come out of creativity blocks while dealing with complex problems.

I appreciate all the dinners and hangouts with my friends Alexander Novikov and Galina Bulgakova. You two added much soul to my life in Lawrence and I will certainly miss my time with you.

I would like to acknowledge my Pakistani friends in Lawrence which include Amirali Shalwani, Hamza Rehman and Daniyal Ahmed. Spending Saturday nights with you guys was always a pleasure, always made me feel closer to home and provided me with some much needed refuge from the pressure that came with graduate school. I certainly have gained considerable perspective and knowledge from our late night discussions on all the various religious, political and societal topics.

My friend Ammar Khalid certainly played a huge role in helping me stay sane and make sense of all that was going on around me, while at the same time always pushing me to be more critical of myself and become a better person.

These last few months of thesis writing, with the COVID-19 pandemic raging across the world, have been very stressful. So I would like to acknowledge my fiancé, Bushra Siddiqui, who was always a beacon of support and hope for me. My conversations with her played an immense role in keeping my anxiety levels in check and helping me get to the finish line in these highly uncertain times.

Most importantly, I would like to thank my parents Sabahat Latif and Farzana Latif for always supporting me in every way possible from day one, and always having my back in the darkest of times. I would also like to thank my siblings and their partners Haris Latif, Shehwar Latif, Saad Latif, Maleeha Zubairi, Ali Azeem and Summaya Latif for being the best siblings one could have ever asked for and for always covering for me and helping me get back up every time I fell down. All of you helped me maintain my self-belief and helped me in countless ways which I cannot start to cover here as the list would be infinitely long. This PhD is dedicated to my loving family and it belongs to my parents and siblings as much as it belongs to me.

Contents

List of Figures

List of Tables

Preface

The work done for this thesis represents the first UHECR analysis for an in-ice radio experiment at the South Pole. The in-ice and in-air raytracing codes that were written in the process of this thesis have had a significant impact on this UHECR analysis and have also been widely used in other applications. The raytracing codes have been utilized in:

1. The upcoming version of CoREAS: Raytracing has been incorporated into CoREAS and has allowed it to calculate radio emissions from UHECR shower particles even after the showers enter the ice. The simulation would be able to do precise calculations of in-air and in-ice Askaryan, geomagnetic, and transition radiations that might be observed at the in-ice radio antennas. This would be the first time any simulation would be able to account for all of these radio emissions from UHECR showers for in-ice antennas.

2. The optimization of the South Pole ice refractive index model: In the optimization process, data recorded in the ARA stations during the 2018-2019 SPUNK (South Pole UNiversity of Kansas) radio pulser drop in the SPICE (South Pole ICE) core borehole was used to extract the observed time difference between the Direct and Reflected/Refracted (DnR) rays ($dt(D,R)$). The $dt(D,R)$ times were obtained from data for each antenna for all ARA stations and were then compared to $dt(D,R)$ times calculated using raytracing. The simulated times were then adjusted (by adjusting the refractive index model parameters) until they matched the data times. This was the first time $dt(D,R)$ times were used to infer the South Pole ice refractive index model, and the whole process yielded a promising South Pole ice refractive index model for the ARA stations.

3. The optimization of in-ice Askaryan radio detector geometries: The DnR rays were used in simulations to calculate the effective volumes of the detectors. The DnR ray paths were used

to do vertex reconstruction of neutrino showers and calculate neutrino energy. This allowed simulation users to find optimal positions (i.e., that maximized the overall effective volume of the detector) for antennas inside the ice.

4. In-ice radar simulations for UHECR and neutrino showers: The in-ice radar setup involves a transmitting antenna which emits radiowaves that get reflected by an in-ice shower towards a receiving antenna. In the simulation, all ray paths between the transmitting antenna, shower particles and receiving antenna were calculated through raytracing and used to obtain the final electric field observed on the receiving antenna.

5. ARA calibration activities: The cross-polarization response of the ARA Hpol calibration pulsers caused the emission of Hpol and Vpol pulses. The relative time delays between the received Hpol and Vpol pulses in the ARA receiving antennas were compared to the expected time delays calculated from raytracing. This exercise led to proposed adjustments in the time delays of ARA channels, making the source vertex reconstructions, for ARA stations, more accurate.

The ARA UHECR analysis that was carried out for this thesis is undoubtedly the first iteration and has helped set the stage for all future UHECR analysis for in-ice radio experiments at the South Pole. The currently deployed ARA stations are uniquely positioned to observe UHECR showers that hit the South Pole ice surface and emit in-air and in-ice Askaryan, geomagnetic and transition radiation. The next generation CoREAS which will be used in the future iterations of this UHECR analysis, will account for all of these emissions and result in a more detailed and accurate analysis.

Chapter 1

Cosmic Rays

Cosmic rays are defined as charged particles that come from outer space. They interact with air particles in our atmosphere to produce EM (Electro-Magnetic) and Hadronic showers, which then propagate and evolve through the atmosphere as they approach the surface. At times, other known particles like neutrinos or photons are also called cosmic rays when they are of relatively high energy and interact with the atmosphere to produce showers (Nelles, 2014).

1.1 Significance of CRs

Cosmic Rays (CR) were discovered by Victor Hess in 1912, when he took a free balloon flight and observed that the radiation detected by his electrometers increased at higher altitudes. He then took a similar flight during a solar eclipse and made similar observations, which made him realize that the sun was not the sole source of this radiation. Since most CR particles are charged, they get deflected by the multitude of magnetic fields (i.e., galactic, planetary and solar) that they encounter along their journeys from their source to their final destination. Therefore, discovering a CR particle source from its trajectory presents an enormous challenge and is one of the biggest unresolved mysteries in Science. Some possible sources of UHECRs and UHE neutrinos include super-massive black-holes, neutron stars, mergers involving neutron stars and black-holes and supernova explosions.

Cosmic Rays are penetrating our atmosphere all the time. Everything on Earth is exposed to them. Some particles like muons (which are weakly interacting, negatively charged and 200 times more massive than electrons) can easily pass through physical barriers. CR particles inter-

act with the molecules in our bodies and everything around us. It has been shown that there is a possible correlation between CRs and the number of observed mutations in bacterial DNA (Gonzalez, 2014). Certainly, intense exposure to cosmic rays will cause damage to human cells and mutations in DNA strands. The Earth's atmosphere and magnetic fields play a massive role in protecting humans from CR radiation. Charged particles that get trapped in Earth's magnetic fields lead to auroras on Earth's polar region when they interact with atmospheric particles. Astronauts are particularly exposed to a high influx of CR radiation, and studies indicate that CR radiation on long-term missions (i.e., planned missions to Mars) would pose a multitude of health risks and could cause cancer, damage to central nervous system, cataracts, circulatory diseases and acute radiation syndromes (Cucinotta & Cacao, 2017). Therefore, it is critical to understand the effects of CR radiation on human bodies, not just for space exploration, but also for humans living on Earth.

The evolution of humans and animals has also been affected by CR radiation. Researchers have also linked supernovae explosions, that happened around 2.6 million years ago, to humans learning to walk on two feet. The supernovae are believed to have occurred within the "immediate" vicinity of Earth within a distance range of around 50 parsecs (163 light-years). They believe that the supernovae caused an avalanche of cosmic rays on Earth, which heavily ionized the atmosphere and led to an increase in the number of lightning storms. These lead to large scale fires around the world and burned down forests transforming them into savannahs. Humans then learned to walk on two feet as they had to walk larger distances between trees, and also standing up helped them spot predators from far away. The researchers also linked these storms to higher amounts of radiation dosage in animals and humans, leading to a possible general decline in their populations (Melott & Thomas, 2019; Melott, 2016).

Electronics are also known to be adversely affected by CRs. High-energy neutrons coming from the sky can cause computer errors known as Single-Event Upsets (SEUs) when they collide with a nucleus on a computer board (Cooper, 2012). The collision produces a burst of electric charge that either causes a change in a single digit in the computer memory or causes a logic

circuit to produce an erroneous result that may hang up (or crash) a computer program. This can be extremely harmful to electronics on board airplanes and satellites and can endanger human lives. Similarly, these neutrons have also been known to crash supercomputers (Scoles, 2018). Therefore, it is critical to understand the effects of CRs on electronics and mitigate them so that our electronics can run more efficiently and reliably.

Studying CRs has been particularly useful from multiple perspectives:

1. The measurement of the lifetime of cosmic ray muons led to another confirmation of Einstein's theory of relativity. The measured lifetime of muons ($\sim 2.2\ \mu$s) came out to be less than the time taken by the muons to reach Earth's surface ($\sim 50\ \mu$s). This paradox was only correctly explained by the theory of time dilation, which was postulated by Einstein.

2. Tomography is the process of using penetrative radiation to image an object internally by looking at cross-sectional slices. Muon and neutrino tomographies have proved to be useful in several different ways:

   (a) Muon tomography has been used in a wide variety of applications which include:

      i. The ScanPyramids mission aimed at studying the underlying structures and possible hidden chambers inside of the Great Pyramids of Giza (Morishima et al., 2017).

      ii. Non-proliferation of nuclear weapons and curtailing the movement of contraband by using detectors at ports that use tomography to scan for such materials in cargo (Katz, 2014).

      iii. The reclamation of the Fukushima Daiichi Nuclear complex using tomography to understand the extent of the structural damage to the nuclear cores during the March 2011 tsunami. This survey will help in making plans for the eventual reactor dismantlement.

   (b) Neutrino tomography has been used to study the internal structures of Earth. A recent study using IceCube data successfully determined the mass of Earth, the mass of

Earth's core and the moment of Inertia of Earth (Donini et al., 2019). It was also able to establish that the core is denser than the mantle.

Multi-messenger astronomy seems to be the future of astro-particle physics, where multiple experiments looking at the sky through various methods corroborate each other's observations, and CRs and neutrinos will play a significant part in these observations. Recently a multi-messenger observation of a gamma-ray blazar (called TXS 0506+056) was made through a neutrino observed by IceCube, through gamma rays observed by Fermi-LAT and several ground-based Imaging Atmospheric Cherenkov Telescopes (IACTs) including High-Energy Stereoscopic System (HESS), Very Energetic Radiation Imaging Telescope Array System (VERITAS) and Major Atmospheric Gamma Imaging Cherenkov (MAGIC) (Aartsen et al., 2018). This observation was the first of its kind and groundbreaking. Multi-messenger observations are incredibly insightful because they shed more light on the physics of these high energy events and increase the confidence in the measurements being made by all these science instruments.

In the last two decades, astro-particle physics has made significant advances in narrowing down the sample of possible sources that produce UHECRs and UHE neutrinos. Sources having supermassive black-holes like Centaurus-A and TXS 0506+056 blazar are promising candidates. However, even more data is needed to be able to make conclusions and to understand the underlying physics at these high energies. In this regard, multi-messenger astronomy has undoubtedly given a considerable boost to astro-particle physics, and with more technologically advanced detectors coming in the future, the next few decades are going to be very exciting and enriching for this field.

1.2 Energy Spectrum and GZK suppression

Figure 1.1 shows that the flux of these cosmic rays starts to rapidly drop to zero at around $10^{19.5}$ eV. This effect is known as the GZK suppression, and has been measured to high accuracy by both the Telescope Array (TA) and the Pierre Auger Observatory (PAO) experiments (Ogio, 2019; Castellina, 2020). At energies above this threshold energy, the cross-section for the interaction of cosmic

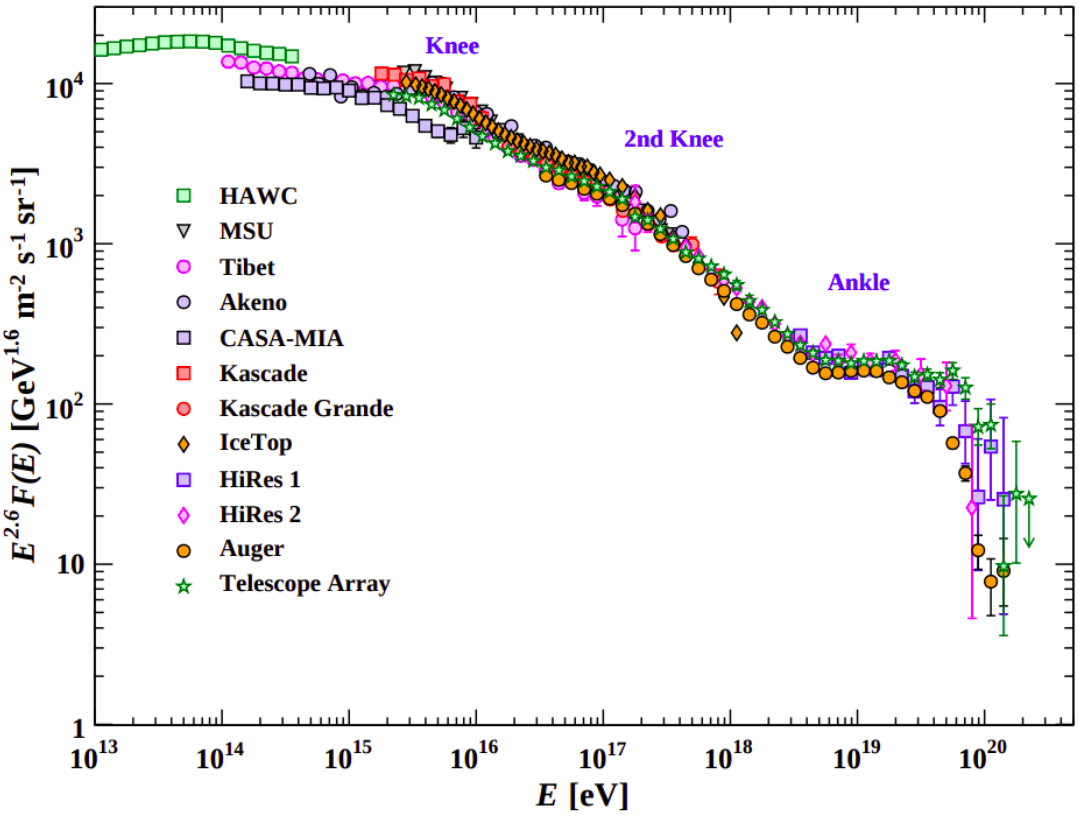


Figure 1.1: Measurements of the cosmic ray flux for all types of particles from various experiments in units of energy per nucleon. The differential energy spectrum has been multiplied by $E^{2.6}$ to display the features of the steep spectrum that are otherwise difficult to discern. The steepening that occurs between 10^{15} and 10^{16} eV is known as the "knee" of the spectrum. The feature around $10^{18.5}$ eV is called the "ankle" of the spectrum. Figure adapted from (Tanabashi et al., 2018).

rays with CMB photons increases dramatically, and their measured flux on Earth is correspondingly attenuated. GZK suppression is named after Greisen, Zatsepin and Kuzmin, who predicted it in 1966 (Greisen, 1966; Zatsepin & Kuzmin, 1966). If we take the case of protonic cosmic rays, we can write:

$$p^+ + \gamma_{CMB} \rightarrow \Delta^+ \tag{1.1}$$

$$\Delta^+ \rightarrow p^+ + \pi^0 \rightarrow p^+ + 2\gamma \tag{1.2}$$

$$\Delta^+ \rightarrow n + \pi^+ \rightarrow n + \mu^+ + \nu_\mu \tag{1.3}$$

$$\mu^+ \rightarrow e^+ + \bar{\nu}_\mu + \nu_e \tag{1.4}$$

Similar equations can be written for anti-particles. Another way to understand Equation 1.1 is that, if we boost into the rest frame of the proton, the Δ^+ resonance threshold is crossed at a photon energy of around 340 MeV (Hanson, 2013). As a result of this, protons effectively lose 95% of their energy in each interaction (Zatsepin & Kuzmin, 1966). The distance scale of such an energy loss is around 10 Mpc (Greisen, 1966).

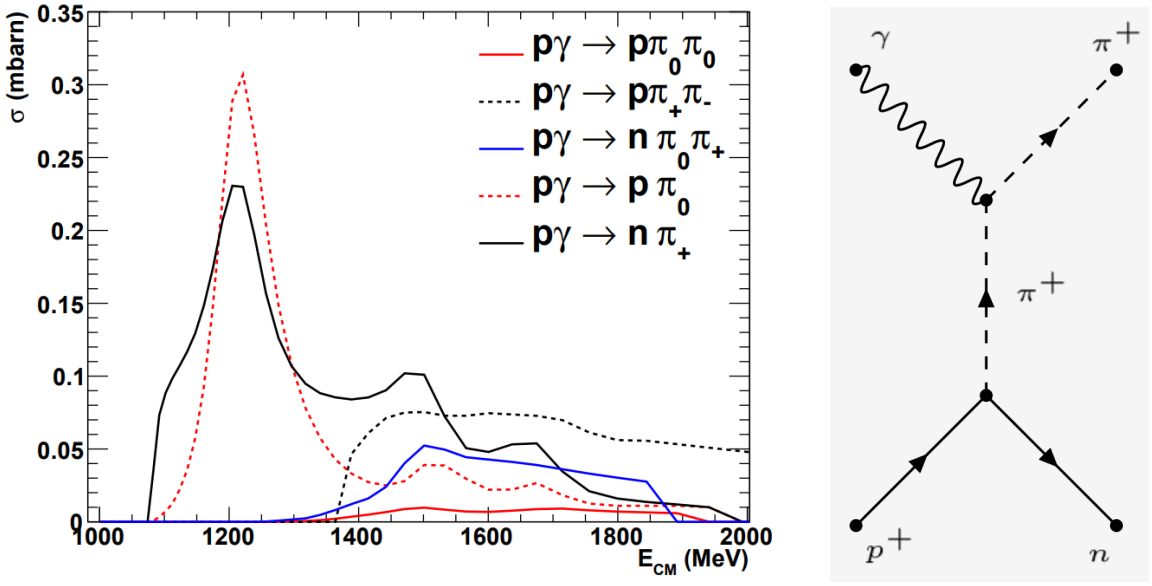


Figure 1.2: Pion production through Δ^+ resonance. (**Left**) Cross section for photoproduction of pions vs. E_{CM} (center of mass energy of the primary proton and CMB photon), dominated by Δ^+ resonance at 1232 MeV. (**Right**) The tree level t-channel Feynman diagram for direct pion production via a virtual meson exchange. The Figure on the left has been adapted from (Andersen & Klein, 2011).

This resonant Δ^+ production produces twice as many neutral pions than charged pions, as can be inferred from the cross-sections in Figure 1.2 on the left. Direct pion production via a virtual meson exchange also occurs, although this contributes only about 20% to the total cross-section, in part because this exclusively happens for charged pions (Ahlers et al., 2010). Hence in the $p\gamma$ interactions, equal numbers of π^+ and π^0 are produced. The Feynman diagram for the direct pion exchange is shown on the right in Figure 1.2 (Yu et al., 2011), including an EM interaction vertex

and a strong interaction vertex. It can be thought of as EM scattering by virtual mesons, which are quantum-mechanical representations of the strong force within a baryon. In the interaction, the virtual mesons gain enough momentum to come on-shell. The EM vertex requires that the particle that couples to the photon is charged; hence direct interactions with neutral pions are forbidden. Since the cross-section for the direct scattering process scales with $m_{virtual}^{-2}$ (where $m_{virtual}$ is the mass of the virtual meson), therefore, contributions from heavier mesons are suppressed (Mucke et al., 2000).

Photons produced by π^0 in this process typically have GeV-TeV gamma-ray energies. The process $\pi^0 \rightarrow 2\gamma$ has a branching ratio of about 99%. Each of the three neutrinos produced in the processes given by Equations 1.3 and 1.4 obtain about 5% of the initial cosmic ray energy.

If the parent cosmic ray is heavier (up to Z=iron), then the high-energy neutrino flux is significantly reduced; in order to take part in the GZK process, the iron nuclei must first photo-disintegrate to produce protons. In this case, the protonic cosmic rays will have $E_{p^+} \approx E_{Fe^{+56}}/56$, which implies that to take part in the GZK process, the energy of the iron nuclei should be at least 56 times the energy of a proton. This essentially shifts the cosmic ray spectrum, which in turn lowers the UHE neutrino flux from heavier cosmic rays by about a factor of 100 (Hanson, 2013).

The dip at the 'ankle' of Figure 1.1 is caused by electron-positron pair production by these proton CRs in their interactions with CMB photons:

$$p^+ + \gamma_{CMB} \rightarrow p^+ + e^+ + e^- \tag{1.5}$$

This is also known as Bethe-Heitler pair production. This process dominates at the 'ankle' but its cross-section greatly decreases at the GZK energy limit. The pair production dip and GZK cutoff are generic features of proton CRs.

1.3 Composition

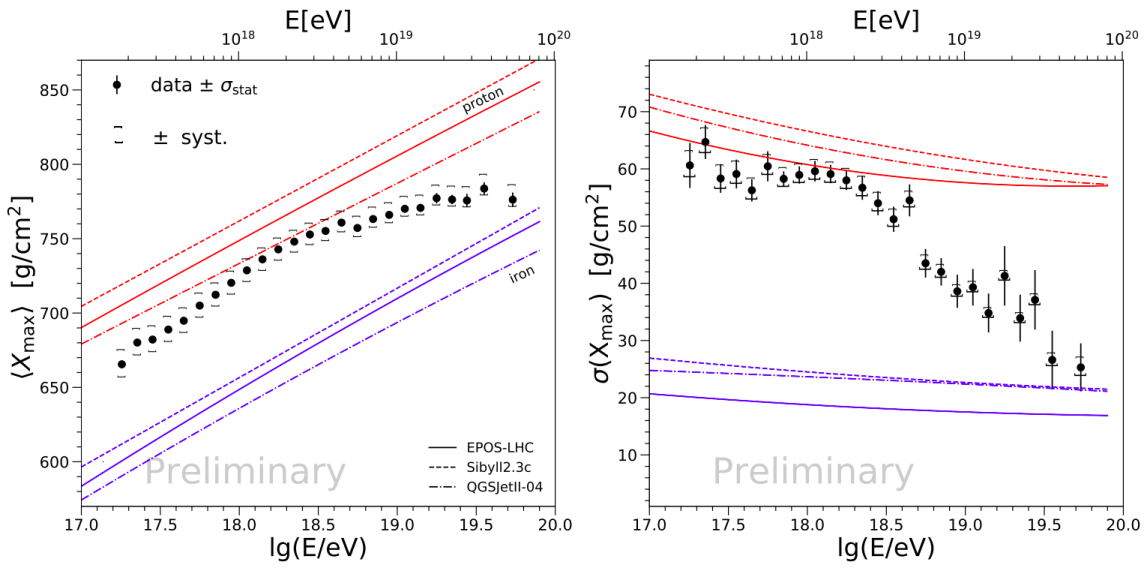
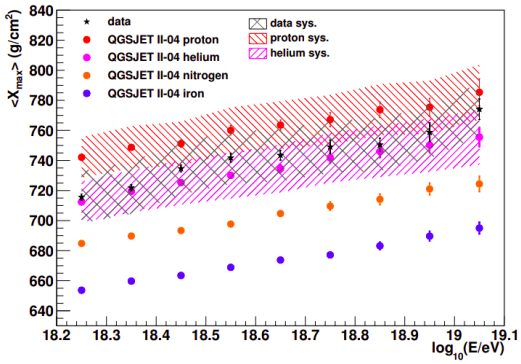


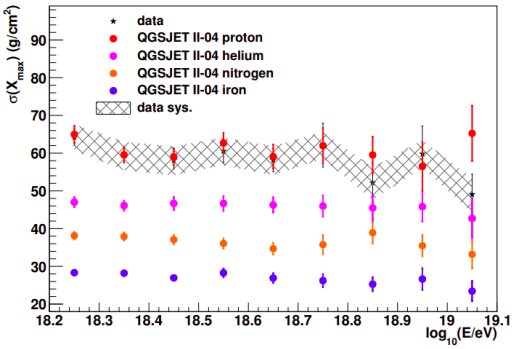
Figure 1.3: Most recent results from Pierre Auger Observatory (PAO) reported at the ICRC 2019 conference. The red lines are simulated predictions for proton cosmic rays and the blue lines are for iron. There are three dashed lines for each of the three simulation models: EPOS-LHC, Sibyll2.3 and QGSJETII-04, all tuned to match LHC data. **Left**: Average measured X_{max} for the given cosmic ray energies. **Right**: Measured and simulated values of the standard deviations $\sigma(X_{max})$ for each of the measured X_{max} values. Figure adapted from (Castellina, 2020)

.

Figures 1.3 and 1.4 show recent results from the Pierre Auger Observatory (PAO) and the Telescope Array (TA) for CR shower maximum (X_{max}) data. The PAO $\langle X_{max} \rangle$ data appears to favor a mixture of lighter nuclei (protons and possibly helium) at lower energies and heavier at higher energies. The TA $\langle X_{max} \rangle$ data appears to favor a light mixture of helium and protons for all energies. The TA $\sigma(X_{max})$ data appears to favor a pure proton CR spectrum at lower energies, whereas, at higher energies, the trend favors helium nuclei; the PAO $\sigma(X_{max})$ distribution shows a similar trend compared to PAO $\langle X_{max} \rangle$ distributions. Both collaborations show general agreement in their UHECR composition data at lower energies, favoring lighter mass compositions in UHECRs but show a disagreement at higher energies (Ogio, 2019; Castellina, 2020; Aloisio, 2018).



(a) TA 10 year hybrid $\langle X_{\max} \rangle$



(b) TA 10 year hybrid $\sigma(X_{\max})$

Figure 1.4: Most recent results from Telescope Array experiment, submitted to the ICRC 2019 Conference. The red circles are for protons, purple for helium, orange for nitrogen and blue for iron. All of the simulation data was made using QGSJETII-04. Dark black stars show data from hybrid events taken from TA BR (Black Rock)/LR (Long Ridge) FD (Fluorescence Detectors), in coincidence with that experiment's SD (Surface Detectors) array. For each energy bin, the mean and width of the distribution are overlaid for data as well as for a variety of Monte Carlo simulated compositions. The number above each bin gives the total number of events in that bin. **Left**: Average measured X_{max} for the indicated cosmic ray energies. **Right**: Measured and simulated values of the distribution standard deviations $\sigma(X_{max})$ as a function of energy. Figure adapted from (Hanlon, 2020).

1.4 Origins and Observed Anisotropy

To create these UHECRs requires powerful sources of magnetic shocks and magnetic fields. The acceleration of these particles is achieved via Fermi acceleration, and the particles must remain confined in the source and accelerated several times before they can reach such high energies. These requirements can be summarized by the following formula (Nelles, 2014):

$$E \approx 10^{18}\ \text{eV}.\frac{Z.\beta}{2}.\frac{R}{\text{kpc}}.\frac{B}{\mu\text{G}} \tag{1.6}$$

where E is the energy that can be achieved by the particle, R is the size of the source, B is the strength of the magnetic field, Z is the charge of the particle and β is the relativistic shock velocity. The larger the sources and the stronger their magnetic fields, the greater the likelihood of the particle reaching ultra-high energies.

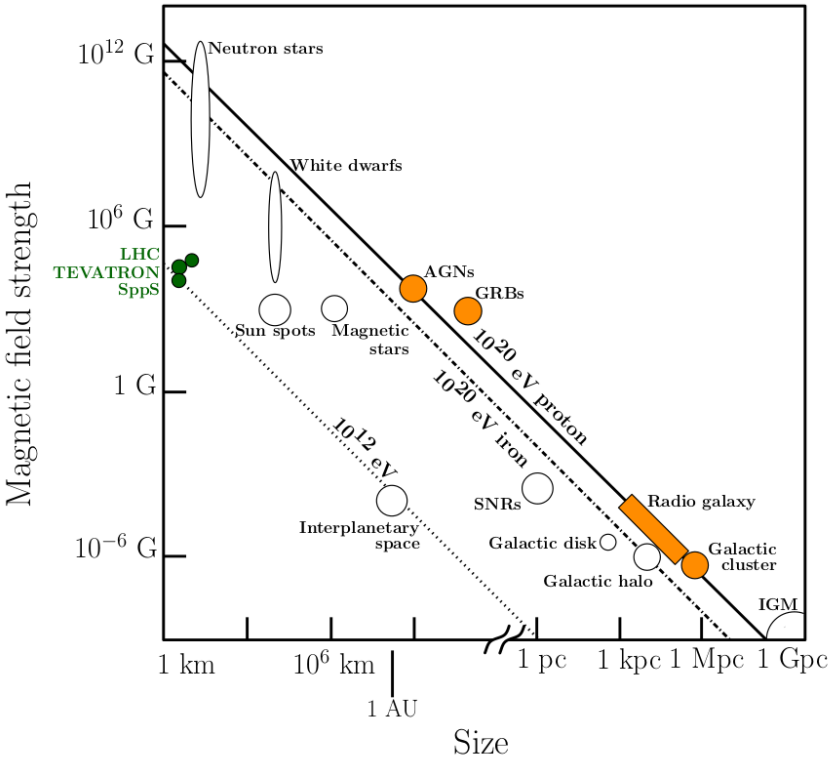


Figure 1.5: Hillas plot (Hillas, 1984). Size and magnetic field strengths of the various sources (denoted by squares and circles) that can accelerate charged particles. Sources whose shapes intersect the shown energy lines can accelerate charged particles at those energies, given the corresponding size and magnetic field values. Figure adapted from (Nelles, 2014).

One example of first-order Fermi acceleration is the case when a charged particle collides with an oncoming shock wave from a supernova and then gets reflected back and forth from the front and back sides of the shock wave that accelerates the particle. Multiple fractional energy gains of the particle can accelerate the particle to at least 100 TeV (Hanson, 2013). This is sometimes referred to as Diffusive Shock Acceleration (DSA) and occurs in SuperNova Remnants (SNR).

Cosmic rays of energy from GeV to PeV seem to be linked to regions of denser gas in our galaxy. These regions correspond to star-forming regions $4-5$ kpc from the center. The distribution of these regions is roughly consistent with the distribution of supernovae, especially of types II, Ib and Ic. The flux of these cosmic rays seems to extend up to the 'knee' energy region around 4×10^{15} eV (Hillas, 2005). Although no direct measurements have been made so far, the

presence of cosmic rays is expected from the gamma rays produced from their inelastic collisions ($p^+ + p^+ \rightarrow \pi^0 + X$ followed by $\pi^0 \rightarrow 2\gamma$) with the interstellar gas. The Fermi-LAT experiment has shown some strong evidence for these galactic cosmic rays using this technique (Ackermann et al., 2013).

In the energy region higher above PeV, the UHECR sources are currently unknown; AGNs or blazars or neutron stars are some of the possible candidates amongst some others, as shown by the Hillas plot in Figure 1.5.

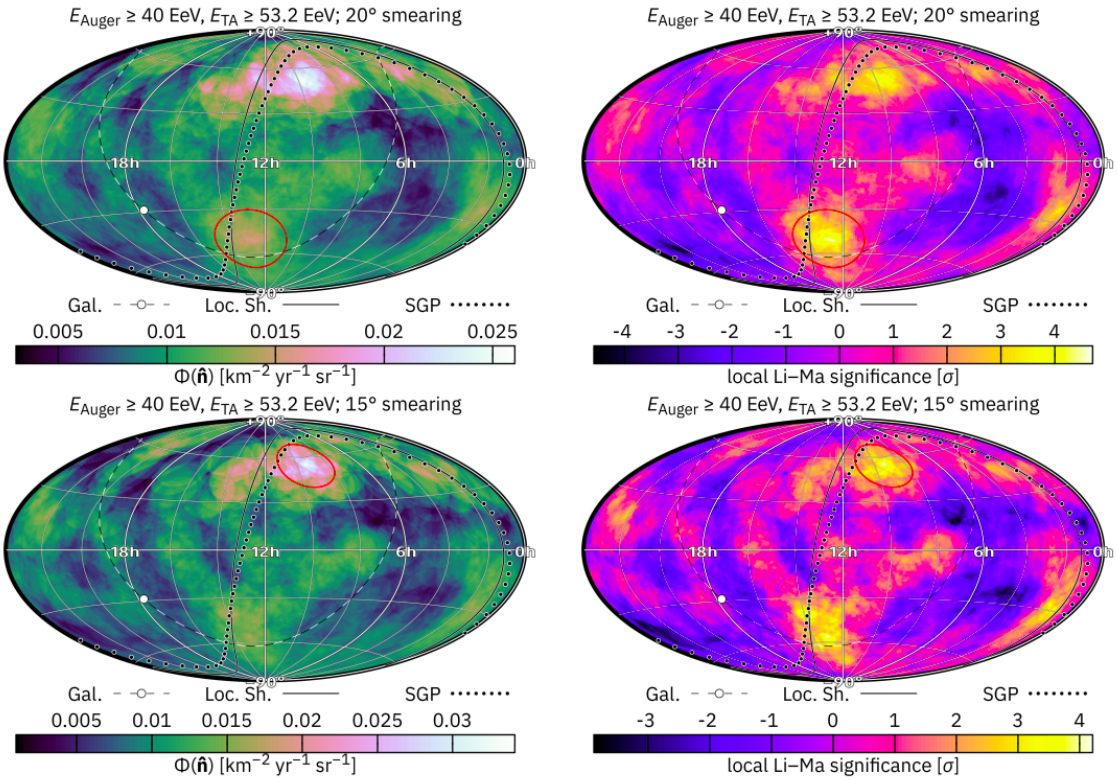


Figure 1.6: Sky maps showing observed cosmic ray flux (**left**) and corresponding local statistical significances against the isotropic null hypothesis (**right**). Sky Maps with (**top**) 20° smearing angle (due to scrambling of galactic and intergalactic magnetic fields) and (**bottom**) with 15° smearing angle. The Galactic, Super Galactic and the Local Sheet planes have been indicated by different dashed lines. The plots indicate several different excesses above $E_{\mathrm{Auger}} > 40$ EeV and $E_{\mathrm{TA}} > 53.2$ EeV. These energy thresholds were chosen in order to match the computed integral fluxes in the declination band $-12° \leq \delta \leq +42°$ which is visible to both detectors. Figure adapted from (di Matteo et al., 2020).

A combined blind search was performed by PAO and TA for medium-scale anisotropies at the

highest energies and is shown in Figure 1.6 (di Matteo et al., 2020). The top and bottom plots have smearing angles of 20° and 15° , respectively. The smearing angle is a variable that acts as an effective search radius and takes into account the unknown deflections of the UHECRs in intervening magnetic fields that they encounter as they travel from their source towards Earth. The study was carried out for a range of smearing angles from 5° to 35° in steps of 5° but the two most significant excesses in UHECR flux were found for smearing angle values of 15° and 20° which is why results for only those two angles were reported in the study.

The sky maps shown in Figure 1.6 also show various types of galactic planes for reference, which are namely (di Matteo et al., 2020):

1. The Galactic Plane (GP) of the Milky Way galaxy.

2. The Super Galactic Plane (SGP) is widely used and was defined in 1953 by Gérard de Vaucouleurs. It is based on the distribution of matter on the scale of several tens of Mpc and is tilted by 84° w.r.t. GP.

3. The Local Sheet (LS) includes nearly all galaxies ≤ 6 Mpc and mostly consists of twelve large galaxies in a ring ("Council of Giants") and two more in the middle (Milky Way and M31) plus their satellites. It is tilted by 89° w.r.t. GP and 8° w.r.t. SGP.

The combined PAO and Auger study reported that for the 20° smearing angle an excess 4.7σ pre-trial (2.2σ post-trial) was found and for the 15° smearing angle an excess 4.2σ pre-trial (1.5σ post-trial) was found (di Matteo et al., 2020). It can be seen in Figure 1.6 that the three most visible excesses are very close to the SGP. The excess regions marked by the red circles in Figure 1.6 (top) include NGC 4945, Centaurus-A and Circinus; M83 is also present and is within 3° of the SGP edge; M81 and M82 are within 1° of the SGP edge in Figure 1.6 (bottom).

Centaurus-A is one example of an AGN that is about 4 Mpc away from our galaxy. This analysis by PAO and TA has shown strong hints of an excess of high energy cosmic ray events that seem to be coming from the direction of Centaurus-A. Star Burst Galaxies (SBGs), which have

shown an excess of UHECR flux in this study, include NGC 4945 and M83, and both of them are present along the SGP.

AGN models used for this study were based on catalogs of highly luminous AGNs detected by Fermi-LAT and *Swift*-BAT and for SBGs based on HCN (Hydrogen Cyanide) emission SBG survey catalogs. It was assumed that for AGNs the UHECR intensity is proportional to integral gamma-ray flux between 50 GeV and 2 TeV and that for the SBGs the 1.4 GHz radio flux serves as a UHECR proxy. It was noted that for SBGs detected in gamma rays, the radio flux scales linearly with the gamma-ray luminosity (di Matteo et al., 2020, & references therein). The search for sources is complicated by the fact that, since cosmic rays are charged, they are easily deflected by galactic, inter-galactic and intra-galactic magnetic fields. This is especially true for cosmic rays passing through strong magnetic fields. The greater the distance the particle travels and the higher the particle charge, the greater will be the particle deflection. The galactic and especially the intergalactic magnetic fields are still not well-known. Recent efforts in trying to model them give improved results, particularly for UHECRs, compared to previous models (Unger & Farrar, 2017). After testing out new alternative parameterizations to the field model, using more recent synchrotron emission maps from the WMAP (Wilkinson Microwave Anisotropy Probe) and Planck telescopes, and using newer models for thermal and cosmic ray electron densities in our galaxy, the authors have shown they can backtrack particles of high rigidity (e.g., protons at 6×10^{19} eV) with reasonable accuracy.

Chapter 2

Extensive Air Showers (EAS)

2.1 Development of Air Showers

The initiation of a shower occurs when a primary cosmic ray has an inelastic collision with an air nucleus (mostly from either oxygen or a nitrogen atom). The energy of the primary particle is transferred to secondary particles which include baryons (p, n..), mesons (π, K..) and leptons, and dominated by pions, with twice as many charged pions than neutral pions.

We can estimate the decay lengths (L_D) of the particles in the lab frame by using the formula:

$$L_D = \gamma\beta c\tau = \gamma\sqrt{1-\frac{1}{\gamma^2}}c\tau \tag{2.1}$$

here the relativistic gamma factor is given by $\gamma = E_{particle}/m_{rest}c^2$ and τ is the proper lifetime of the particle. In this case, in order to calculate the decay length for the shower particles $E_{particle} =$ 1 GeV will be used.

The neutral pions form the EM part of the shower. They almost immediately decay into photons after a decay length of $L_D \approx 1.87 \times 10^{-7}$ m, using $c\tau = 25.5$ nm and $\gamma = 7.41$(Glaser, 2017). These photons then produce electron-positron pairs through pair production, which then produces more photons. The number of interactions by ionization, Compton scattering and Bremsstrahlung keeps increasing until the average energies fall below the critical energy (E_c). Below E_c average collisional (i.e. ionization) energy losses begin to exceed radiative (i.e. Bremsstrahlung) losses. At 1 GeV energy scales, most of the radio emissions arise from the EM part of the shower. Charged pions and kaons constitute the hadronic component of the shower. Initially charged pions and

kaons have longer decay lengths, $L_D \approx 55.34$ m and $L_D \approx 6.52$ m (using $c\tau = 7.8$ m and $\gamma = 7.16$ for pions and $c\tau = 3.7$ m and $\gamma = 2.03$ for kaons), relative to neutral pions. This gives them a higher probability, relative to neutral pions, of interacting before decaying. The hadronic part of the shower keeps growing until typical energies are small enough so that interaction lengths are smaller than the decay lengths. Charged pions almost exclusively decay to muons and neutrinos:

$$\pi^+ \rightarrow \mu^+ + \nu_\mu \tag{2.2}$$

$$\pi^- \rightarrow \mu^- + \bar{\nu}_\mu \tag{2.3}$$

and muons then decay to:

$$\mu^- \rightarrow e^- + \nu_\mu + \bar{\nu}_e \tag{2.4}$$

$$\mu^+ \rightarrow e^+ + \bar{\nu}_\mu + \nu_e \tag{2.5}$$

However, for muons, we have $c\tau = 658.64$ m, and unless they interact en route, they almost always survive to reach the ground.

After the critical energy point has passed, and the shower has reached X_{max}, the particle number starts to decrease. Some of these particles decay, and some of them are captured by air nuclei, with the remainder traveling towards the ground. The atmospheric depth of X_{max} depends on the initial cosmic ray energy and also on the type of particle. Iron nuclei, being heavier than proton nuclei, interact earlier in the atmosphere, and their showers correspondingly reach X_{max} values sooner than a proton's shower. Higher energies always lead to greater X_{max}.

The left plot in Figure 2.1 shows a schematic of particle shower development in the atmosphere and highlights some of the important particle production processes that have been discussed in this section. It also shows the emission of Cherenkov radiation and fluorescence radiation that are produced by the shower particles. The right plot of Figure 2.1 shows the number of hadrons, muons, electrons and photons produced as a function of atmospheric depth as the shower progresses through the atmosphere. This plot was generated using the CORSIKA (COsmic Ray

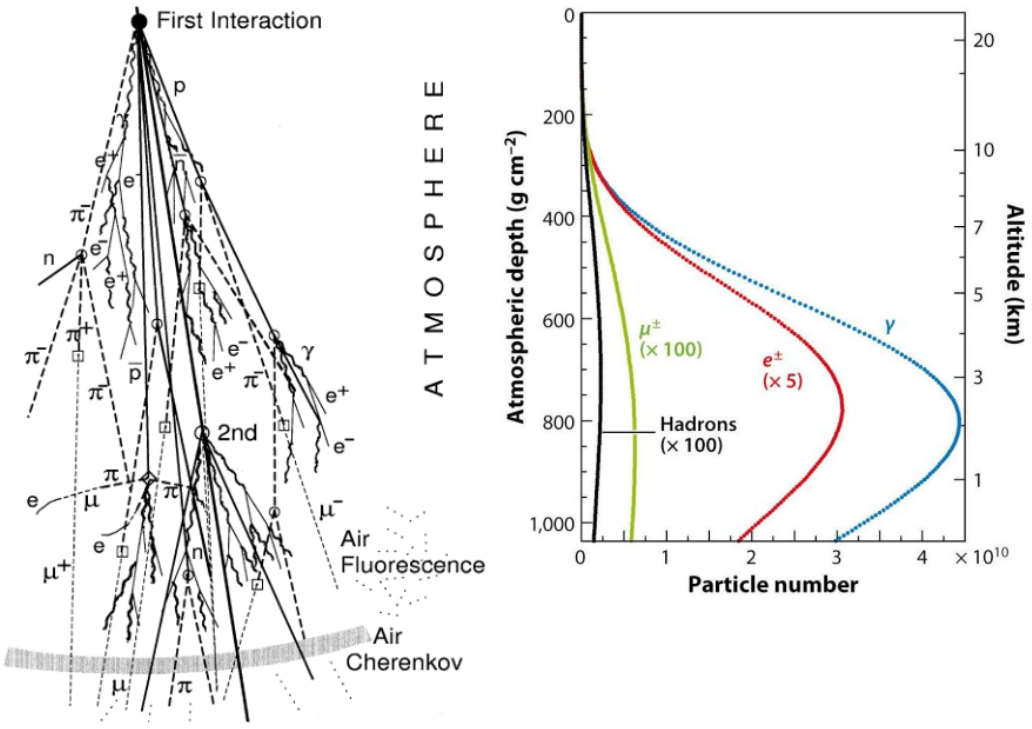


Figure 2.1: (**Left**) Schematic view of a cosmic ray extensive air shower. (**Right**) Evolution of the most dominant particles in the shower. Muon, electron and hadron populations have been multiplied by the numbers shown to highlight their trends. Figure adapted from (Glaser, 2017). The plot on the right was taken from (Glaser, 2017; Schröder, 2017).

Simulations for KAscade) simulation with 10^{19} eV proton primaries, averaged over one hundred showers. The X_{max} for showers at this energy seems to come around 800 g/cm^2 as after that particle populations start to decline.

2.2 Radio Emissions from Air Showers

There are essentially two main mechanisms that cause radio emissions from a cosmic ray shower. The first is geomagnetic emission. The Earth's magnetic field interacts with particles on the shower front via the Lorentz force, which is proportional to $q\vec{v} \times \vec{B}$. Positively-charged particles are pushed to one side of the shower front and negatively charged ones are pushed to the other side. This causes the formation of a transverse current along the shower front. As they spiral, the particles emit synchrotron RF (Radio Frequency) radiation. The polarization of this emission is along the

direction of the Lorentz force. The magnitude of the emitted radiowaves is directly proportional to the strength of the magnetic field $\vec{B}$, the speed of the particles (which is proportional to the initial primary cosmic ray energy) and the sine of the angle α between the field $\vec{B}$ and the velocity $\vec{v}$. About 90% of the emission results from this mechanism.

The second emission mechanism is known as the "Askaryan effect" and this was postulated by Gurgen Askaryan in 1962. A charge anisotropy develops as the shower propagates through the atmosphere, due to Compton-scattered atomic electrons and positrons depleted due to electron-positron annihilation. Both of these effects, combined, produce a $\sim 20\%$ net negative charge. This excess charge is moving faster than the speed of light in the dielectric (air in this case), so it emits coherent Cherenkov radiation at radio wavelengths (based on the size of the shower front). The Cherenkov radiation is emitted at the Cherenkov angle (θ_{Ch}) which is given by:

$$\cos(\theta_{Ch}) = \frac{v}{c.n(h)} \tag{2.6}$$

here the refractive index is a function of height (h) and also dictates the emission angle. The Cherenkov angle is generally of order 1° in this case.

The charge anisotropy at the shower front also causes "relativistic time compression" at the Cherenkov angle. This effect arises, owing to the fact that the atmosphere has an altitude-dependent refractive index. Since $n(h) > 1$, the particle shower front propagates faster than the speed of light and its emission. This essentially 'compresses' the emission in time, equivalent to having the emission add coherently, particularly at very high frequencies, and makes the signal more broadband. This effect causes the formation of an emission ring at the Cherenkov angle around the shower axis, which is especially visible in more energetic cosmic ray showers.

At the Cherenkov angle, any observer would observe all the radio emissions to arrive almost simultaneously, as the time compression effect would be the strongest at that angle. Therefore radio signals at the Cherenkov angle would be highly coherent at higher frequencies. At observer positions off the Cherenkov angle, the time compression would not be as strong. The shower radio

emissions would arrive over a broader range of time, which would lead to weaker coherence at lower frequencies leading to a fall-off in signal strength off the Cherenkov cone. Due to coherence effects, the radio emission off the Cherenkov angle is strongest below 100 MHz (Glaser, 2017). There is generally little power at higher frequencies except for the enhancement observed at the Cherenkov angle due to relativistic time compression.

Figure 2.2 shows a schematic describing the geomagnetic and Askaryan mechanisms. Both of these mechanisms produce nanosecond-duration, broadband radio signals.

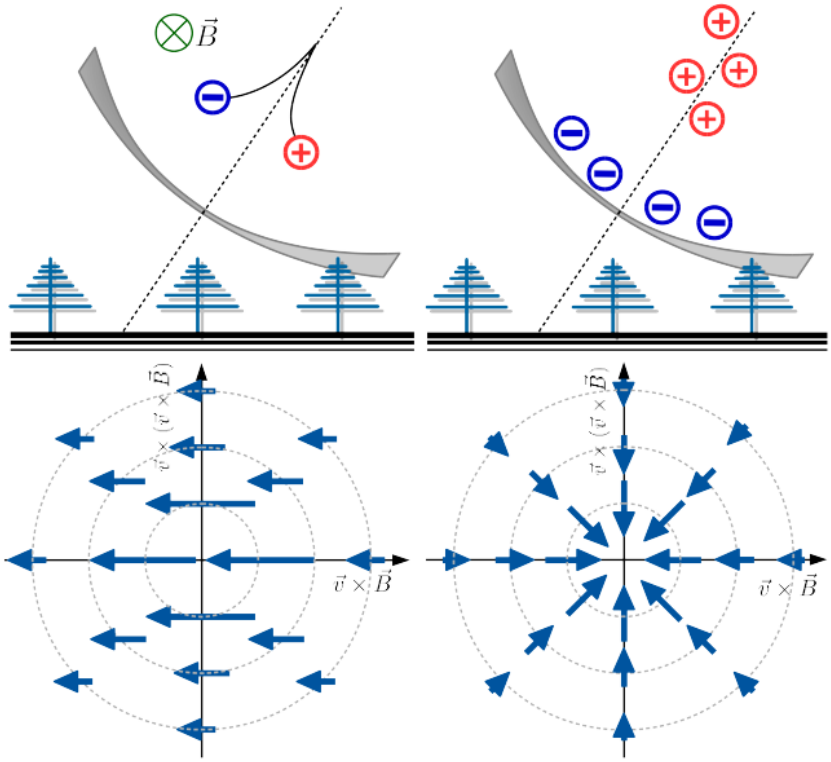


Figure 2.2: Schematic of the two main mechanisms that cause radio emissions from the cosmic ray shower. (**top left**) Geomagnetic emission, for which charges separate in opposite directions; (**top right**) Askaryan emission, for which a charge excess causes the emission; (**bottom left**) Received polarizations of the E-fields on the ground for geomagnetic emission where $\vec{E}$ is aligned with $\vec{v}\times\vec{B}$; (**bottom right**) Received polarizations of the E-fields on the ground for Askaryan emission where $\vec{E}$ is directed towards the center of the shower. Figure adapted from (Glaser, 2017).

The two mechanisms also introduce an asymmetry in the footprint of the shower. It can be observed from Figure 2.2 that the two polarizations, from the two emission mechanisms, add in

the direction of $\vec{v} \times \vec{B}$ but cancel opposite to that direction resulting in signal suppression on one side and signal enhancement on the other side of the radio footprint causing the asymmetry. This asymmetry is also apparent in the radio footprints obtained from simulations of UHECR showers by CoREAS as shown in Figure 2.3.

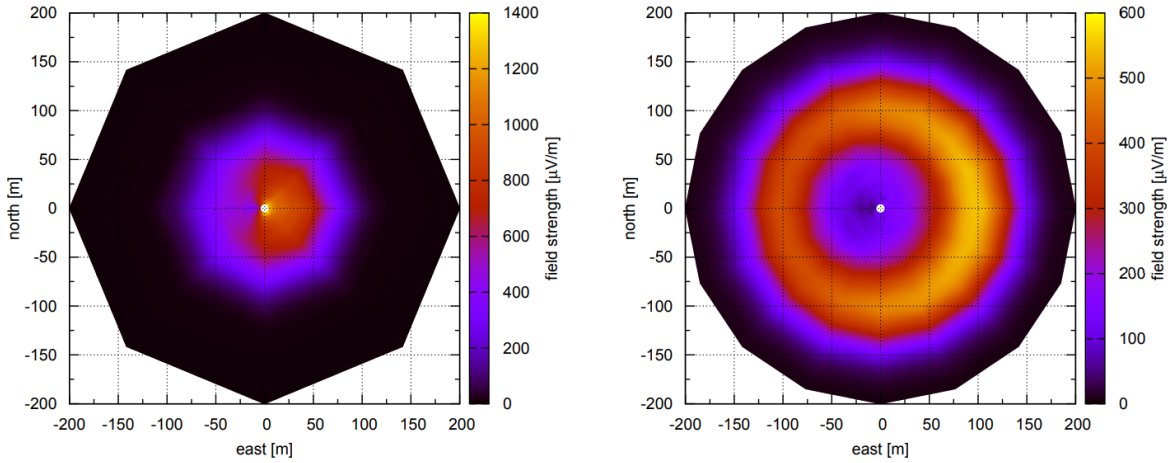


Figure 2.3: Two-dimensional profiles of the magnitude of radio emissions (in the 300 to 1200 MHz bandwidth) at the LOPES site. Both of the showers shown here are simulated in CoREAS for a 10^{17} eV primary with 0° zenith angle. Iron is shown on the **right** and protons are shown on the **left**. The asymmetric nature of the emission is clearly visible in both cases. The expected Cherenkov ring is also visible for the iron shower but not for the proton one – since the X_{max} of the iron shower is much shallower as compared to the protonic shower, the Cherenkov emission is visible within the lateral range shown on the larger scale. No ring is expected at lower frequencies and the 2-D profile of both showers will be similar to that shown on the left. Figure adapted from (Huege et al., 2013).

Given the abundance of anthropogenic noise which permeates all radio frequencies (particularly at lower frequencies) and the irreducible Galactic noise floor (which follows an exponential spectrum starting from a couple of MHz and extending through the entire MHz range), it becomes a bit harder to detect cosmic rays of lower energies. The practical detection threshold is around 10^{16} eV (Huege, 2016).

Thus, every observer around the shower observes an electric field of varying amplitude and polarization. A very basic way to quantify the general E-field pulse amplitude trends observed on the ground is to use *Allan's Formula* to describe the Lateral Distribution Function (LDF) (Allan et al.,

1970; Nelles et al., 2015). This formula is based on empirical observations and some intuition:

$$E_f = C.\left(\frac{E_p}{10^{17}\text{eV}}\right)\sin(\alpha)\cos(\theta)e^{\frac{-R}{R_0(f,\theta)}}\frac{\mu\text{V}}{\text{m MHz}} \tag{2.7}$$

Here E_f is the pulse amplitude per unit bandwidth, θ is the zenith angle and α is the angle between the magnetic field $\vec{B}$ and the arrival direction of the cosmic ray shower as given by its velocity vector $\vec{v}$. For values of 10^{17} eV$< E_p < 10^{18}$ eV and radial distances from the shower axis $R <$ 300 m, Allan gives the scaling parameter $C = 20$, $R_0 = 100 \pm 10$ m for $f = 54$ MHz and zenith angles $\theta < 35°$.

Equation 2.7 indicates that the magnitude of the pulse on the ground is directly proportional to the energy of the primary cosmic ray and the sine of the angle (α). Allan's expression also predicts that the lateral spread of the shower increases; as the inclination increases, the peak amplitude of the E-field pulses on the ground decreases. This effect, however, is predicted to oppose the increased geomagnetic emission as a shower traverses an increased amount of atmosphere. It is expected that away from the shower axis, the pulse amplitude falls off exponentially.

It should be noted that although this formula describes the general properties of the E-field pulse amplitude caused by a shower, it is limited to particular energy, azimuth angle, elevation angle and frequency ranges. Various experiments like LOFAR (Low-Frequency Array), LOPES (LOFAR PrototypE Station), CODALEMA and Tunka-Rex have used different variants of this formula to make their LDFs to estimate the energy of the primary particle (Nelles et al., 2015; Kostunin et al., 2016).

Chapter 3

CoREAS simulation

To understand the complexities of the radio emissions from CR showers better, people have developed several simulations over the past decade or so. These include CoREAS and ZHAireS (the ZHS formalism built into the AIRES code), which treat the shower at the microscopic level, treating emissions from separate particles individually. Macroscopic models like MGMR and EVA treat the shower as a whole by parameterizing the transverse currents and overall charges in the shower as it evolves (Huege, 2016). The latter approach is faster but tends to ignore particle-level correlations, which can lead to an overestimation of the emission (Huege, 2016).

CoREAS is a Monte Carlo code for the simulation of radio emissions from EAS. It implements the endpoint formalism for the calculation of EM radiation directly in CORSIKA (Huege et al., 2013). The simulation makes no assumptions on the emission mechanism for the radio signals and takes into account the complete complexity of the electron and positron distributions as simulated by CORSIKA (Heck et al., 1998).

The endpoint formalism describes particle motion via a series of discrete, instantaneous acceleration events, or 'endpoints', with each event being treated as a source of emission (James et al., 2011). In the final step, the radiation from all particles is superimposed, resulting in the macroscopically observed radio pulse. The electric fields for the endpoint formalism can be directly calculated from the Lienard-Wiechert potentials (Glaser, 2017):

$$\phi(\vec{x},t) = \left[\frac{e}{(1-n\vec{\beta}.\hat{r})R}\right]_{ret} , \vec{A}(\vec{x},t) = \left[\frac{e\vec{\beta}}{(1-n\vec{\beta}.\hat{r})R}\right]_{ret} \tag{3.1}$$

here R is the distance from the point of emission to observer, $\hat{r}$ is the unit vector to the direction of

the observer, $\vec{\beta}$ is the velocity of the particle divided by the speed of light and n is the refractive index of the medium. The subscript "*ret*" denotes that the equation needs to be evaluated at the retarded time $t' = t - nR/c$. Using these potentials we can calculate the E-field (James et al., 2011; Glaser, 2017):

$$\vec{E}(\vec{x},t) = q\left[\frac{\hat{r} - n\vec{\beta}}{\gamma^2(1-n\vec{\beta}.\hat{r})^3\ R^2}\right]_{ret} + \frac{q}{c}\left[\frac{\hat{r}\times[(\hat{r}-n\vec{\beta})\times\dot{\vec{\beta}}]}{(1-n\vec{\beta}.\hat{r})^3\ R}\right]_{ret} \tag{3.2}$$

here $\dot{\vec{\beta}}$ is the time derivative of β and γ is the relativistic gamma factor. The first term is the near-field term and reduces to Coulomb's Law for $\beta = 0$. Since, it has $1/R^2$ dependence, it is therefore neglected in this formalism. The particle motion is subdivided into straight track segments and the emission of the E-field is calculated knowing the start point and the end point of each of those tracks. The radiation at a specific observer location is then given by summing up (or superimposing) all the contributions from all the "endpoints". A downside of this formalism is that it breaks down near the Cherenkov cone angle as the denominator goes to zero because the Cherenkov condition (i.e., $n\vec{\beta}.\hat{r} = 1$) gets fulfilled.

In that case, we can use the ZHS formalism (Glaser, 2017). In the ZHS formalism, we start with the same subdivided tracks but then use the Fraunhofer approximation to calculate the E-field emissions. Here the track lengths need to be small w.r.t the wavelength and distance from the observer. This works well in dense media like ice or sand but in air, the shower size is comparable to the distance to the observer, and uncertainties arise in the calculations. Therefore in air showers, the track lengths need to be made small enough so that this approximation condition is always fulfilled; making computation very slow (Glaser, 2017).

CORSIKA allows one to choose from a multitude of high and low energy hadronic models. In this thesis, QGSJETII.04 will be used as the high energy model and UrQMD 1.3 will be the low energy one. However, as the radio emissions mostly arise from the EM part of the shower, the choice of hadronic models should not affect the radio emission results (Nelles, 2014).

CoREAS has been used and tested by several experiments extensively. Experimental measure-

ments agree with simulations with an uncertainty of around 20% (Schröder, 2017). Although the CoREAS simulation has been primarily developed to predict signals in the MHz region, it can also make predictions in the GHz region. However, predictions of CoREAS at lower-than-MHz frequencies should be taken with caution as at these frequencies, although the positive ions left behind in the air shower may play an important role, they are neglected in most of the codes (Huege, 2016). The positive ions left behind have relatively long lifetimes, around $O(\mu s)$, and, together with the shower front, form an electric dipole that can be as long as a kilometer along the shower axis. The moving electric dipole thus contributes to radio emissions at a wavelength of a kilometer or so well below a MHz.

3.1 Particle Thinning

To speed up the CoREAS simulation, one can apply thinning. When thinning is active, all particles below an adjustable fraction of the primary energy which emerge from an interaction are "thinned". The algorithm is based on the thinning fraction, which is equivalent to a threshold in energy and is given by:

$$\varepsilon_{thin} = \frac{E_{thin}}{E_p} \tag{3.3}$$

If particles fall below this threshold, only one of the particles is selected, and an appropriate weight is given to it, while the other particles are neglected. This reduces the number of particles for which the emission needs to be calculated and makes the simulation drastically faster. The weights are calculated as follows:

$$w_i = \frac{1}{p_i} \tag{3.4}$$

Here $p_i = E_i/E_{thin}$ and E_i is the energy of the ith secondary particle. The particle which passes the thinning process has its weight increased by w_i. For CoREAS simulations in this thesis, if used, the thinning fraction is set at $\varepsilon_{thin} = 1 \times 10^{-6}$.

3.1.1 Thinning noise

Thinning can lead to artificial coherence and increased radio signal power at high frequencies, caused by the increased localization of the particles when only one of them is selected for propagation. Due to this, the simulated signal power does not reach zero at large distances to the shower axis on the ground, which requires an ex-post-facto correction, and for high energy particles, such artifacts can reach amplitudes above the noise floor (Huege, 2016; Nelles, 2014). Therefore care needs to be taken whenever thinning is used in the simulation.

3.2 Atmospheric Refractive Index Profile

In the most recent version of CoREAS (v1.3), the simulation also takes into account the atmosphere's refractive index profile. This is done using the GDAS-tool ("Global Data Assimilation System"). This database contains atmospheric parameters as used by the National Center for Environmental Prediction's (NCEP) Global Forecast System (GFS) in weather forecasting. It features atmospheric parameters at 24 altitudes in the atmosphere at a 1×1 degree grid on Earth, updated every three hours.

The atmosphere mainly consists of N_2, O_2 and Ar with corresponding respective volume fractions of 78.1%, 21.0% and 0.9%. In the simulation, the atmosphere is basically divided into five layers. In the lower four layers, the density follows an exponential dependence on altitude, which leads to a relationship between mass overburden $T(h)$ (g/cm^2) and altitude as:

$$T(h) = a_i + b_i.e^{-h/c_i} \quad i = 1, ..., 4; \tag{3.5}$$

in the fifth (topmost) layer the mass overburden decreases linearly with altitude:

$$T(h) = a_5 - b_5 h/c_5. \tag{3.6}$$

The boundary of the atmosphere in this model is defined as the altitude where the mass overburden

vanishes, corresponding to h=112.8 km for the US standard atmosphere (Heck et al., 1998). The *a*, *b* and *c* parameters for each layer are provided by the GDAS-tool.

The GDAS atmospheric database provides, for a given location and time, a corresponding density and humidity profile. The density profile is fitted to generate the 5-layer atmosphere that is fed into CORSIKA. At the same time, a consistent, tabulated refractivity profile is fed to CoREAS, which includes humidity effects in the refractive index profile (Huege et al., 2013).

Chapter 4

The ARA detector

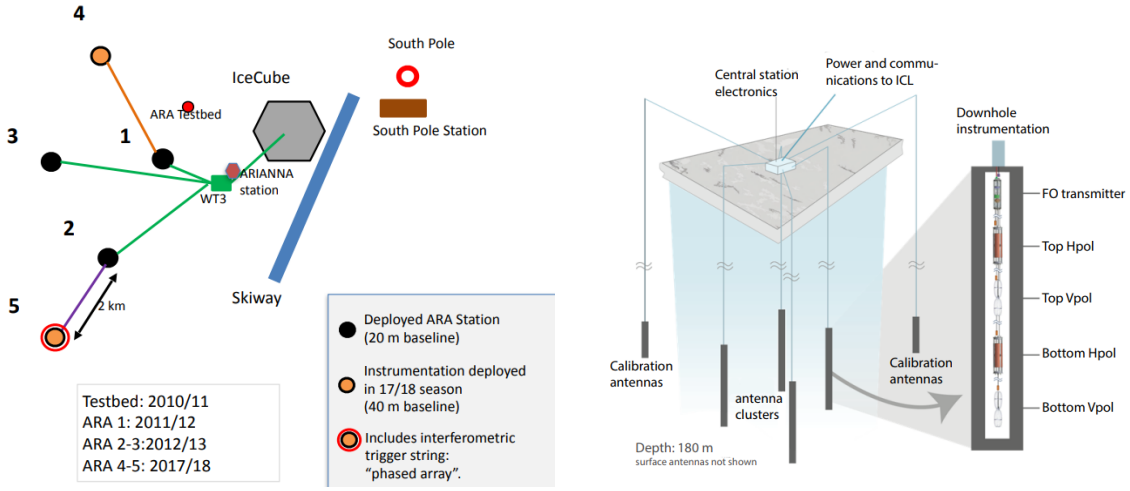


Figure 4.1: (**Left**) The current layout of the ARA stations that have already been deployed at the South Pole. The lines indicate the cabling of the stations, which are powered from the IceCube Lab (ICL). The ARA Testbed, the recently deployed ARIANNA (Antarctic Ross Ice Shelf Antenna Neutrino Array) station and the Amundsen-Scott South Pole Station are also shown for reference. (**Right**) A schematic drawing of a typical ARA station; ARA-5 also has an extra string in the middle for the Phased Array. The figure on the right was adapted from (Allison et al., 2016) and the figure on the left was adapted from A. Karle's presentation at ARENA 2018.

The Askaryan Radio Array is a planned hexagonal array of 37 stations (or detectors) in the South Polar ice sheet. Currently, five stations have been deployed: ARA-1 deployed in the 2011-2012 austral summer season, ARA-2 and ARA-3 deployed in the 2012-2013 austral summer season, and then ARA-4 and ARA-5 deployed in the 2017-2018 austral summer season. The array is located next to the IceCube experiment; all the power supplies for all the stations are housed in the IceCube Lab (ICL). The distance between neighboring stations is 2 km. The array proposed to

cover an overall area of the order 100 km^2. Such a large array is required by the low expected flux of GZK neutrinos (Allison et al., 2016). The reason for having the experiment at the South Pole is the low radio noise (there are very few anthropogenic noise sources) and also the long km-scale attenuation length of radiowaves in ice.

ARA-2, ARA-3, ARA-4 and ARA-5 stations have antennas at depths ranging from $170 - 200$ m. For ARA-1, a drill malfunction led to antennas being deployed around 40% of the planned 200 m depth. Each station has 16 antennas in 4 strings (or holes), with 4 in each string. The strings are configured in a square with side 20 m. There are also two more strings that contain calibration pulsers that emit a RF pulse every second. The calibration pulser strings are 40 m away from the center of the square and in-between two of the four square strings horizontally.

In each of the strings, the antennas are ordered, top-to-bottom as: Hpol (horizontally polarized) top-most, VPol 2 m below, 20 m a second HPol antenna and 22 m the final VPol antenna. The station layout is shown in Figure 4.1. The horizontally polarized antennas are quad-slot cylinder antennas and the vertically polarized antennas are the birdcage bicone antennas (Allison et al., 2012). The stations are sensitive in the bandwidth range from 150 to 850 MHz.

4.1 The ARA Signal Chain

Figure 4.2 shows the signal chain for each of the ARA channels. The received radio signal from the antennas first passes through a notch filter, which removes 450 MHz frequency components, corresponding to the band used for radio communications at the South Pole. The signal is then passed to a Low-Noise Amplifier (LNA) with a gain of about 20 dB. The signal is then transported over RFoF (RF over Fiber) to a receiver where it is amplified first by the 20 dB Optical Zonu receiver, and then amplified by a second-stage 40 dB RF amplifier. The signal is then bandpass filtered with a High Pass Filter of 150 MHz and a Low Pass Filter of 850 MHz, after which it is split by a 50:50 power splitter.

The split signal is then passed to the Trigger Daughter board for ARA (TDA) and the Digitizing Daughter board for ARA (DDA). On the TDA, the signal is read into the digital electronics and

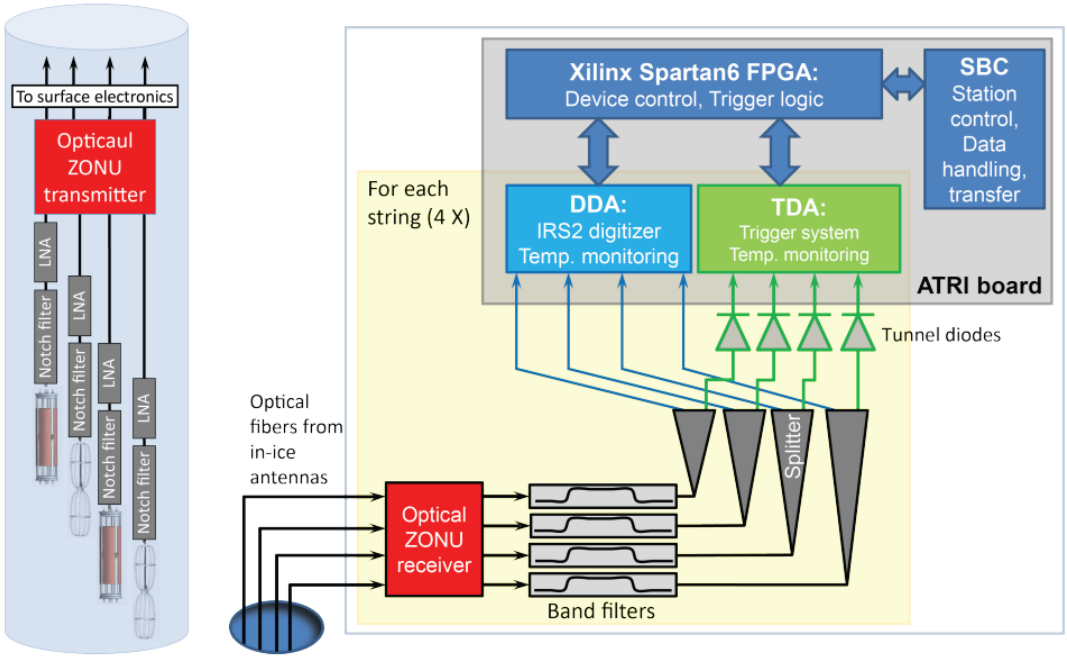


Figure 4.2: A schematic of the ARA signal chain for each channel (or antenna). The components in the blue cylinder on the left are the components unique to each string. The yellow box on the right shows the components which are on the surface at the DAQ box and are common to all strings. Figure adapted from (Allison et al., 2016).

processed by the logic implemented in an FPGA (Field Programmable Gate Array), which makes a trigger decision. In the DDA, the data is sampled by the IRS2 ASIC, a digitization chip capable of sampling up to a rate of 4 GS/s. In ARA, the sampling is done at 3.2 GS/s. The power consumption per channel is around 20 mW (Allison et al., 2016).

Both the TDA and DDA boards are mounted on the ARA Triggering and Readout Interface (ATRI), which provides all logic for the DAQ system in a single Spartan-6 FPGA. This FPGA is programmed by a Single Board Computer (SBC), which handles the data transfer to storage on disks in the ICL (Allison et al., 2016).

The trigger condition for the stations requires that at least 3 antennas of one polarization must have a pulse that crosses the trigger threshold (set to be around 6σ above noise) within 170 ns. Trigger thresholds are automatically adjusted to maintain a global event trigger rate (data written to tape) of 5 Hz for each station. The calibration pulsers add another 1 Hz from the calibration

events, with an additional 0.5 Hz added by forced software triggers.

In total the ARA system response gives about 80 dB gain. Figure 4.3 shows the total gain of the system in dB, the unwrapped phase in radians and the group delay in ns versus frequency in MHz.

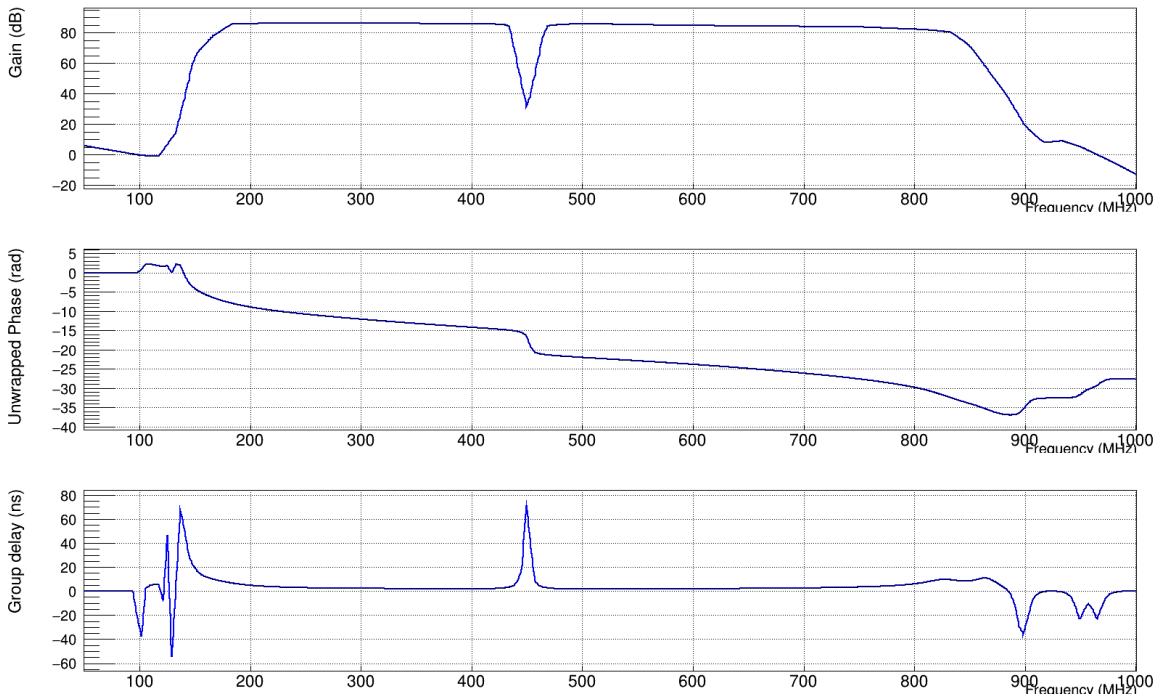


Figure 4.3: Total ARA system response. (**Top**) Gain (dB), (**Middle**) Unwrapped Phase (rad) and (**Bottom**) Group Delay (ns) vs Frequency in MHz.

Chapter 5

RayTracing in South Pole Ice

Once we have propagated the simulated E-fields from air shower particles to the ice surface, we have to understand the propagation mechanism that takes them from the ice surface to the ARA antennas inside the ice sheet. For this, we need to perform raytracing, i.e., tracing the path of the radiowaves as they travel through the ice sheet. In so far as ARA is a detector of neutrino-induced showers in the ice sheet, raytracing forms a crucial part of the neutrino vertex reconstruction required to perform energy reconstruction of the neutrino. Rays are refracted in the ice sheet owing to the depth-dependent density, and therefore the depth-dependent index of refraction. The density profile of the South Pole ice is expected to follow an exponential form which corresponds to a refractive index profile:

$$n(z) = A + Be^{Cz} \tag{5.1}$$

here if z is negative, then C has to be positive, and if C is positive, then z has to be negative, with z defined as depth inside the ice. For the model currently used by ARA these parameters have the values $A = 1.78$, $B = -0.43$ and $C = 0.0132$ m^{-1} (Kelley et al., 2018).

This causes the bending of rays and creates 'shadow zones' which are areas from which we should technically not see any radio signals. This is important as it helps quantify the effective area or volume over which the ARA stations can detect neutrino-induced showers.

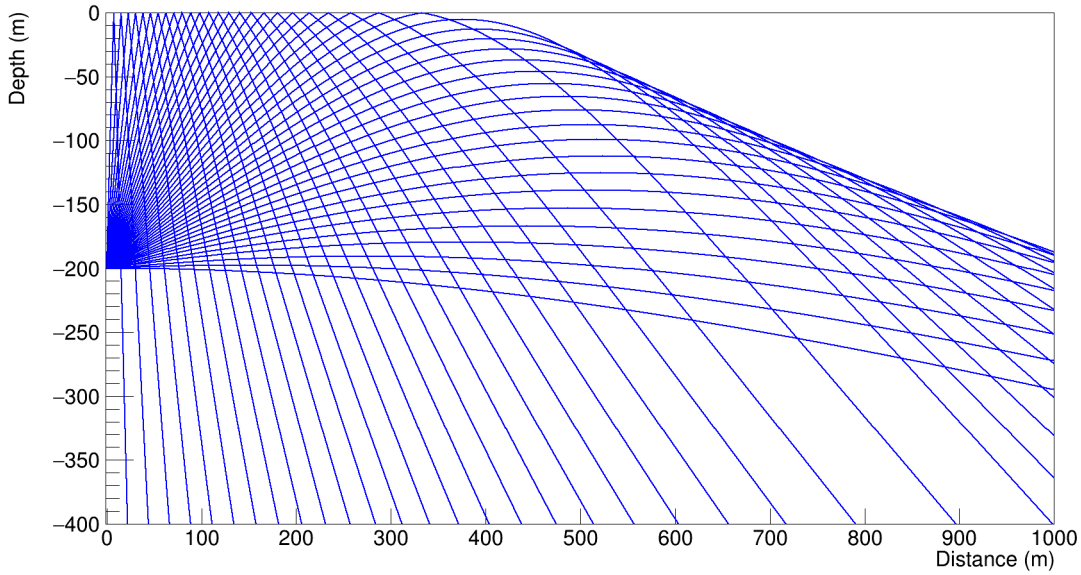


Figure 5.1: Raytracing simulation for a source at $z = -200$ m depth, emitting radiowaves in launch angle steps of 2°. Rays are reflected first from the ice surface at $z = 0$ then begin to refract downward as the source migrates laterally from the receiver. The area where the rays do not penetrate is called the 'shadow zone'.

As is visible in Figure 5.1, there are in general two possible (and mutually exclusive) indirect ray trajectories from each source – those that reflect off the surface and those which refract downwards because of the refractive index profile. The refracted ray that takes the shortest possible time to reach the target is called the direct ray.

5.1 Numerical Raytracing using the Runge-Kutta Method

There are two ways ray trajectories can be calculated: numerically and analytically. The set of following equations can be solved using the fourth order Runge Kutta Method to trace out the ray path:

$$\frac{dx}{ds} = \sin(\beta)\ , \quad \frac{dz}{ds} = \cos(\beta)\ , \quad \frac{dt}{ds} = \frac{n(z)}{c} \tag{5.2}$$

$$\frac{d\beta}{ds} = -\frac{\sin(\beta)}{n(z)} \frac{dn(z)}{dz}\Big|_z\ , \quad \frac{dA_0}{ds} = \frac{A_0}{L(z,f)} \tag{5.3}$$

here β is the ray angle w.r.t the vertical, from below, t is the arrival time of the ray, A_0 is the initial signal amplitude and $L(z, f)$ is the attenuation length, which is dependent on frequency and the depth of the ray. A schematic of these conventions is shown in the right panel of Figure 5.2.

To solve these equations, we need initial conditions. The coordinates of the source (x_0, z_0) and also the coordinates of the target (x_1, z_1) are given, but we do not know, *a priori*, the initial launch angle θ_0 required to reach the receiver. The unknown launch angle can be found in several different ways. We can either use trial-and-error, and keep launching rays until we hit the target, or we can try to derive an analytical expression that gives us the launch angle for the reflected and the direct rays, as discussed in the next section.

The advantage of the brute force trial-and-error method is that it will work with any refractive index model. However, to predict the ray times accurately requires very small step sizes, of order 1 mm. If the rays are being traced over lengths on the order of kilometers (the usual case), then the computational time is large.

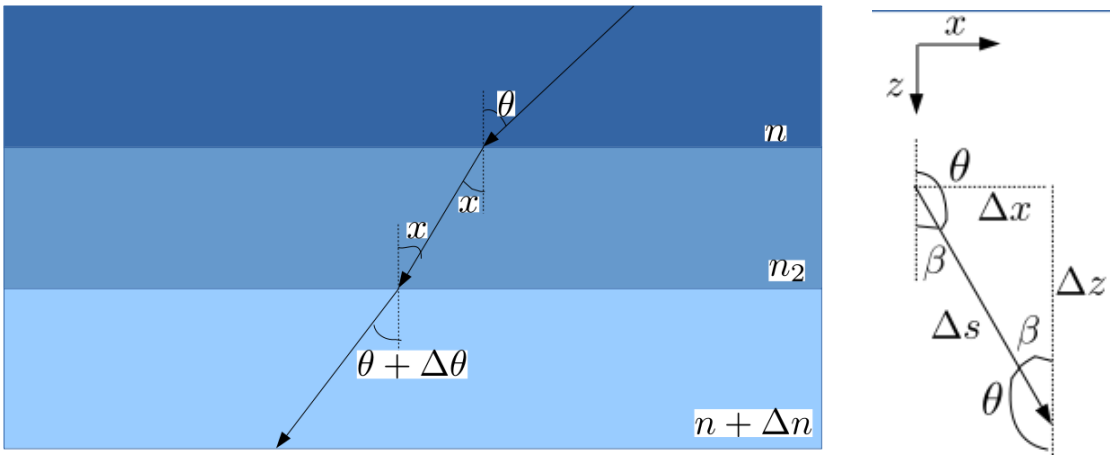


Figure 5.2: (**Left**) A schematic of the ice sheet, approximated as having multiple layers. (**Right**) A schematic of the ray path conventions being used in the equations.

5.2 Analytical Solution for Raytracing

To start the derivation we start from the ice layer schematic shown on the left of Figure 5.2. Here the ice sheet is being approximated as a combination of multiple thin layers with different values of refractive index. Using Snell's law, we can write:

$$n\sin(\theta) = n_2\sin(x), \;\; n_2\sin(x) = (n+\Delta n)\sin(\theta+\Delta\theta). \tag{5.4}$$

Substituting one of the above equations into the other gives us:

$$\frac{\sin(\theta+\Delta\theta)}{\sin(\theta)} = \frac{n}{n+\Delta n}. \tag{5.5}$$

Using double angle identities and some algebraic manipulation we can write:

$$\frac{\sin(\theta)\cos(\Delta\theta)}{\sin(\theta)} + \frac{\cos(\theta)\sin(\Delta\theta)}{\sin(\theta)} = \left(1+\frac{\Delta n}{n}\right)^{-1}; \tag{5.6}$$

taking the limits $\delta\theta \ll 1$ and $\Delta n \ll 1$ we can rewrite the above equation as:

$$1+\cot(\theta)\Delta\theta = 1-\frac{\Delta n}{n} \;\; \Rightarrow \;\; \frac{\Delta n}{n} = -\cot(\theta)\Delta\theta, \tag{5.7}$$

reducing to:

$$\boxed{\frac{d\theta}{dn} = -\frac{\tan(\theta)}{n(z)}} \tag{5.8}$$

From here, using Equation 5.8 and the chain rule, we obtain:

$$\frac{d\theta}{dz} = \frac{d\theta}{dn}.\frac{dn}{dz} \;\; \Rightarrow \;\; -\frac{\tan(\theta)}{n(z)}\, n'(z), \text{ here we have } n'(z) = \frac{dn}{dz} \tag{5.9}$$

If we integrate both sides of the above equation we obtain:

$$L = n(z)\sin(\theta), \tag{5.10}$$

where θ is defined in the schematic on the right of Figure 5.2 and L is the integration constant. In this case, L corresponds to the initial condition of the ray and its value is obtained from the source depth (z_0) and the launch angle (θ_0):

$$\boxed{L = n(z_0)\sin(\theta_0)} \tag{5.11}$$

Using Equation 5.10 and the value of L we can now write:

$$\theta = \arcsin\left(\frac{L}{n(z)}\right), \tag{5.12}$$

Now using the above result, we can write for any ray path that:

$$\frac{dx}{dz} = \tan(\theta) = \tan\left(\arcsin\left(\frac{L}{A+Be^{Cz}}\right)\right), \tag{5.13}$$

Here the functional form of $n(z)$ has been explicitly stated. If we integrate this differential equation, we finally obtain the analytic form of the ray path:

$$\boxed{x(L,z) = \frac{L}{C}\frac{1}{\sqrt{A^2-L^2}}\left(Cz - \log\left(A(A+Be^{Cz}) - L^2 + \sqrt{A^2-L^2}\sqrt{(A+Be^{Cz})^2-L^2}\right)\right)} \tag{5.14}$$

The analytic solution that we found in Equation 5.14 has several important properties:

1. When $A = L$ or $A = n(z_0)\sin(\theta_0)$, the solution becomes undefined. The condition $A < L$ therefore must always be fulfilled, putting a limit on our launch angle (θ_0) for a given source depth, and limiting launch angles to less than 90°. Since ray paths are reversible, the target and the source can be swapped. Thus this technique allows us to trace rays even if the source is shallower than the target.

2. This solution has two roots: one for $C > 0$ and $z < 0$ and the other for $C < 0$ and $z > 0$.

3. The solution also becomes undefined when $n(z) = L$ or $A + Be^{Cz} = L$. This gives us a limit on the maximum depth (peak point, or turning point) that a ray launched at a certain angle θ

will attain. This point is called z_{max}, which can be written as:

$$\boxed{z_{max} = \frac{1}{C}\log\left(\frac{L-A}{B}\right)} \tag{5.15}$$

5.2.1 Finding the Launch Angle

Given (x_0, z_0) and (x_1, z_1), we still do not have the value of θ_0 which is needed so that the value of L can be calculated. Then the ray can be traced out between the source and the target.

To find θ_0, the analytical solution needs to be minimized w.r.t the launch angle. We start by writing the solution in Equation (5.14) as $x = f(L,z)$. First we start the ray from zero distance, such that x_0 corresponds to $x = 0$. To fulfill this condition, it is required:

1. For Direct rays: $f_1'(L,z) = f_1(L,z) - f_1(L,z_0)$.
2. For Reflected and Refracted rays: $f_2'(L,z) = f_2(L,z) - f_2(L,z_0) - 2x_{max}$.

Here x_{max} is the value of x at the turning point (x_{max}, z_{max}). Then, if the target coordinates are to lie on this ray path, they must fulfill the condition:

1. For Direct rays: $f_1'(L,z_1) - x_1 = 0$.
2. For Reflected rays: $f_2'(L,z_1) - x_1 = 0$, for which $x_{max} = f_2(L,0) - f_2(L,z_0)$.
3. For Refracted rays: $f_2'(L,z_1) - x_1 = 0$, for which $x_{max} = f_2(L,z_{max}) - f_2(L,z_0)$.

Now the value of L or the launch angle θ_0 for which these conditions are fulfilled can be calculated. This minimiser procedure can be done in a variety of ways. In ROOT, a function `TF1::GetX()` can be used which uses Brent's method and returns the value of the launch angle for the zeroes of the above function (Brun & Rademakers, 1996). GSL (GNU Scientific Libraries) minimiser libraries can also be used to perform this routine.

5.2.2 Finding the Travel Time of the Rays

For calculating the travel time, one can use Fermat's least time principle, which leads to the integral:

$$t = \int dz \sqrt{1 + \left(\frac{dx}{dz}\right)^2} \frac{n(z)}{c} \tag{5.16}$$

Here we can substitute Equation (5.13) for dx/dz and integrate to get:

$$\begin{aligned} t(L,z) = & \left(\frac{1}{c\,C\sqrt{n(z)^2 - L^2}}\right) \\ & \left[n(z)^2 - L^2 + \left[Cz - \log\left(A\,n(z) - L^2 + \sqrt{A^2 - L^2}\sqrt{n(z)^2 - L^2}\right)\right] \frac{A^2\sqrt{n(z)^2 - L^2}}{\sqrt{A^2 - L^2}}\right. \\ & \left. + A\sqrt{n(z)^2 - L^2} \log\left[n(z) + \sqrt{n(z)^2 - L^2}\right]\right] \end{aligned} \tag{5.17}$$

Equation (5.17) gives us the time at a particular point; to find the time taken by the ray between two points we have:

1. For Direct rays: $\Delta t = t(L, z_0) - t(L, z_1)$
2. For Reflected rays: $\Delta t = t(L, z_0) - t(L, 0) + t(L, z_1) - t(L, 0)$
3. For Refracted rays: $\Delta t = t(L, z_0) - t(L, z_{max}) + t(L, z_1) - t(L, z_{max})$

5.2.3 A, B and C Parameters for Air and Ice Refractive Index Models

For in-ice raytracing we assume a single layer exponential refractive index model with parameter values $A = 1.78$, $B = -0.43$ and $C = 0.0132$ m^{-1}. This is the default ARA $n(z)$ model.

For in-air raytracing we have to take into account a five layer refractive index model for the atmosphere, based on the five layer CORSIKA atmospheric model for mass overburden. The GDAS-tool, as discussed in Section 3.2, makes a file that is used by CORSIKA and CoREAS as

it contains the tabulated air refractive index profile, the boundaries of the five atmospheric layers and the a, b and c parameter values for the mass overburden layers of the atmosphere. Although there certainly was a refractive index profile present in the GDAS file, which would take into account humidity, it was not used for analytic raytracing. Adding humidity effects would cause the refractive index profile to lose its exponential nature. It should be noted that currently, CoREAS uses the tabulated refractive index profile for air (obtained from the GDAS tool) to do numerical raytracing in air; there is no raytracing in ice.

The A, B and C values for each of the exponential refractive index layers for air are given in Table 5.1.

Table 5.1: A, B and C values for the five exponential refractive index layers of the South Pole Atmosphere.

Layer	Altitude Range (m)	A	B	C (m^{-1})
1	0 to 3217.48	1	0.000328911	0.000123309
2	3217.48 to 8363.54	1	0.000348817	0.000141571
3	8363.54 to 23141.80	1	0.000361006	0.000145679
4	23141.80 to 100000	1	0.000368118	0.000146522
5	$>$ 100000	1	0.000368117	0.000146522

In order to obtain the B and C values, reported in Table 5.1, for the exponential atmospheric refractive index layers, the following process using the GDAS file was followed:

1. The air refractive index value at sea level (i.e., 0 m altitude) was obtained from the GDAS file.

2. The value of parameter C at sea level was obtained by using the c value of the mass overburden profile from the GDAS file to calculate C using the relationship $C = 1/c$.

3. Using the fact that at sea level, the values of A, C and the refractive index of air were known, the value of B was calculated.

4. The exponential profile of all the layers was then reconstructed from bottom to top using the same process at each layer boundary. This was possible since we already knew the atmospheric layer altitude boundaries and the value of the C parameter for each layer from the GDAS file.

A values for air refractive index layers were for always taken be to 1.

5.2.4 Limitations of Current Raytracing Model

The current analytic raytracing method assumes a pure exponential profile for the refractive index models of air and ice and, in some cases, gives results that contradict observations. The analytic raytracing leads to a 'shadow' region not only in ice but also in air, where sources at extremely inclined angles would not be visible to an in-ice antenna beyond a certain horizontal distance. For example, this code predicts that an in-air radio source 1 m above a 3000 m altitude air/ice boundary broadcasting to an in-ice antenna 200 m below the surface would be shadowed beyond a horizontal distance (between the antenna hole and the source) of 205.064 m on the ice surface. This is contradictory by data, as ARA stations certainly observe anthropogenic RF sources from South Pole station, the IceCube Laboratory and Wind Turbine 3. All of these sources are several kilometers away from the ARA station and emit RF pulses towards the station at extremely steep angles in air. This prediction of a 'horizontal cut-off' (or shadow region) in air is purely the result of refraction of radiowaves in air as it disappears if we assume air to have a constant refractive index value of 1. It is important to note that several things are ignored or not taken into account in the proposed analytic raytracing method that may explain this discrepancy:

1. It is assumed that Earth has a flat surface and curvature effects are ignored.

2. The raytracer does not take into account local variations in elevation (also known as *Sastrugi*) at the ice surface around the South Pole. These non-uniformities in the ice surface can certainly cause the rays to emerge from (or to enter into) the ice surface from unexpected locations.

3. The raytracer also ignores near-surface Fresnel effects. This is another case where a Huygens wavelet approach will be more exact (as in a Finite-Difference Time-Domain (FDTD) simulation of EM ray propagation) around the air-ice boundary.

All of the above-listed possibilities explain why the predictions of the current analytic raytracer disagree with observations.

It should also be pointed out that there are also 'non-conventional' propagation modes for radiowaves, which allow the reception of radiowaves from sources that would be present in 'shadow' zones. Classical methods, such as this analytic raytracer, dictate that those sources should not be observed. However, these 'non-conventional' propagation modes have been observed by the ARIANNA collaboration at Moore's Bay (Lahmann, 2019). These propagation modes could also explain the observation of RF activity at the South Pole Station in the ARA stations.

Chapter 6

Simulating EAS for ARA using CoREAS and Producing UHECR Templates

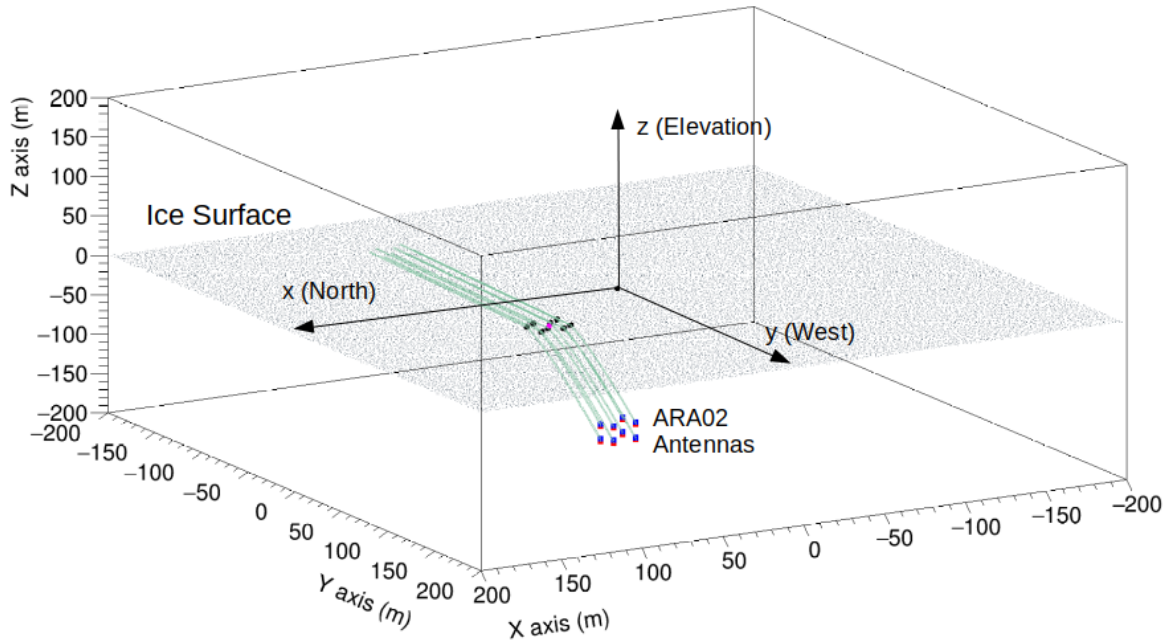


Figure 6.1: Schematic of the approach used to simulate CoREAS showers for the ARA stations. In this coordinate system, the x-axis points towards Magnetic North. Green lines represent radio emissions from the shower entering the ice. This particular shower was simulated for the direction $\theta_s = 45°$and $\phi_s = 35°$. The pink circle is the shower core, the 16 black circles the radiation entry point into the ice, such that it hits the ARA antennas; red (Vpol) and blue (Hpol) squares are the ARA antennas. The origin of this coordinate system is located at the SE corner of the ARA station DAQ box which is roughly in the center of the station.

Although the ARA detector is primarily designed to detect GZK neutrinos, anything that produces radio waves presents a potential background to the experiment. Radio emissions from cosmic ray showers are particularly interesting as their signals are very similar to the predicted Askaryan

emissions from neutrinos. The neutrinos are expected to be down-coming and have trajectories similar to UHECR. The difference between them lies in the fact that the in-ice Cherenkov cone for neutrinos is upcoming after it's in-ice interaction. Upcoming neutrinos cannot be detected since, at such high energies, they are "absorbed" by Earth on their way to the radio detector.

CoREAS, given an antenna (or observer) position on the ground, calculates the emissions at that antenna point for a given shower. The user specifies the hadronic interaction models, shower direction, energy, shower core position on the ground, the primary cosmic ray particle type and the antenna positions. Using these pre-defined input parameters, CoREAS returns the E-field at that observer position in a Cartesian coordinate system (x, y, z) as a function of time:

$$\vec{E}(t) = E_x(t)\ \hat{x} + E_y(t)\ \hat{y} + E_z(t)\ \hat{z} \tag{6.1}$$

To simulate showers in CoREAS for ARA and make UHECR templates the general scheme is:

1. Select a shower's zenith and azimuthal (θ_s, ϕ_s) angles and its energy.

2. Find the points (observer) on the ice where a shower of that (θ_s, ϕ_s) needs to intersect such that radiation travels and refracts through the ice to 'hit' a given ARA antenna.

3. Convert the E-fields from the Cartesian coordinate system to spherical coordinates.

4. Propagate the E-field down through the ice using analytic raytracing. Raytracing returns the values of propagation time and distance and the angle of reception at the antenna inside the ice.

5. Convolve the E-fields with the antenna height effective (h_{eff}) (which is the complex antenna response function) to find the voltage induced on the antennas.

6. Apply the ARA system response to the voltage waveforms to obtain the final waveforms as they should be observed in the data.

7. Adjust the timing of the simulated waveforms to correctly mimic the trigger latch and also to take into account cable delays.[1]

All of these steps will now be discussed in detail.

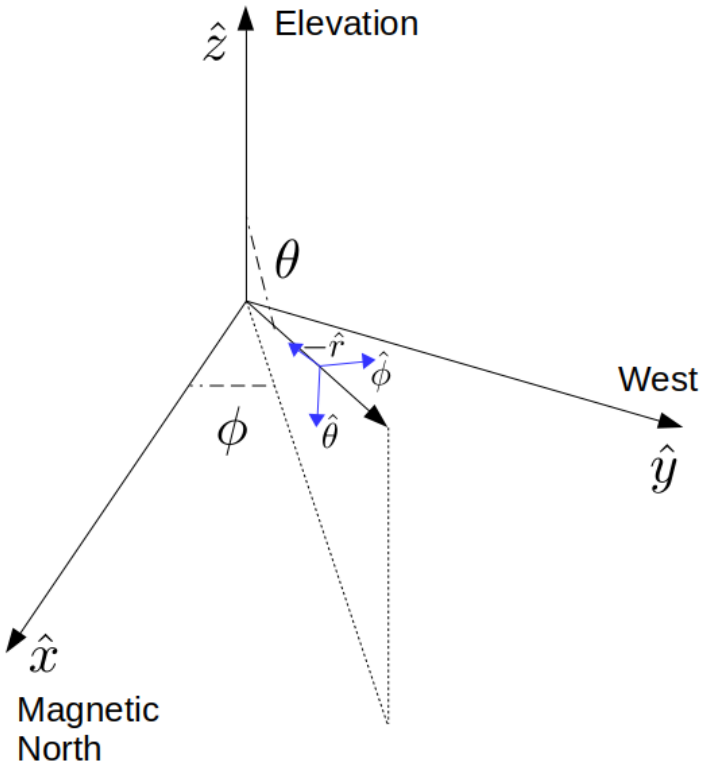


Figure 6.2: CoREAS coordinate system.

6.1 Coordinates of the ARA Stations

CORSIKA and CoREAS follow a coordinate system where the x-axis points towards Magnetic North, the y-axis points west and the z-axis points upwards. θ is the zenith angle of the incoming primary cosmic ray. It is defined as the angle between the particle momentum vector (directed towards negative $\hat{r}$) and the negative z-axis, and ϕ is the angle between the positive x-axis and the horizontal component of the particle momentum vector, with positive defined as counter-clockwise.

[1] N.B. This is not being done currently as CoREAS UHECR templates are now used for "shape" matching to extract viable UHECR candidates from data. Interferometry will then be performed on data events to reconstruct their directions. Therefore timing the template pulses is not necessary in this context.

All of these conventions are shown in Figure 6.2.

The ARA stations have their own coordinate system (the ARA Station Centric (ASC) coordinate system). At the ARA station sites at the South Pole, the ice sheet flows along a line near 36.773° W longitude by about 10 m/year. In the ASC system, the x-axis points in the direction of this ice flow. The origin of this coordinate system is the SE corner of the DAQ boxes at the surface, roughly in the middle of the four strings. To obtain the ARA station coordinates for CoREAS, we have to perform two rotations, as shown in Figure 6.3.

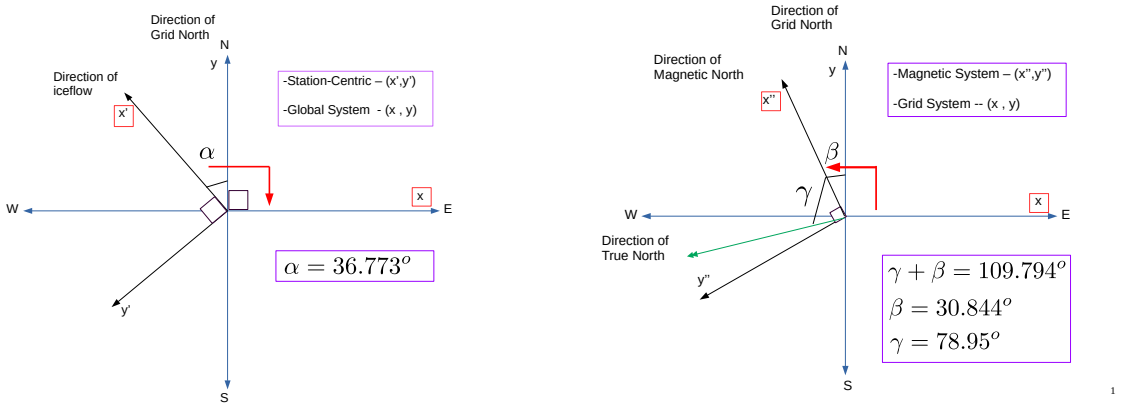


Figure 6.3: The two rotations to get from ARA Station Coordinate System to get to Magnetic North Coordinate system. (**Left**) The rotation from ASC to the Grid coordinate system (**Right**) The rotation from the Grid to the Magnetic North coordinate system.

1. Rotation from ASC to Grid coordinates: If we take ASC coordinates to be (x',y') and the Grid (or Global) coordinate system to be (x,y), we rotate clockwise with an angle $\alpha + 90°$ where $\alpha = 36.773°$, as defined by the ice flow direction.

2. Rotation from Grid coordinate system to Magnetic North coordinate system: If we take Magnetic North coordinates to be (x'',y'') and the Grid (or Global) coordinate system to be (x,y), we rotate counterclockwise with angle of $\beta + 90°$, where β is the angle between grid north and magnetic north and is equal to 30.884°. This angle was calculated using a combination of magnetic field calculators, which included the online version from NOAA's website

(National Oceanic and Atmospheric Administration) and the one available with GeographicLib's C++ classes (NOAA, 2018; Karney, 2019). The magnetic field calculators were made to use the IGRF-12 (International Geomagnetic Reference Field) model for the Earth's magnetic field. The angle of declination was calculated to be $\gamma = 78.95°$ E, which gives the angle between grid north and true north to be $\gamma + \beta = 109.794°$. The arguments given to the calculator were the GPS coordinates of the ARA-2 station's SE DAQ box corner (89.9588 S, 109.794 W and 9293.2712 ft MSL). The time argument used for the calculation was given (somewhat arbitrarily) as 03/20/2019.

6.2 Calculating Radiation Transect Points on Ice

Assuming that (most of the) emission is collinear with the primary cosmic ray momentum, we find the point on the ice such that the signal refracts to hit the ARA antennas in the ice.

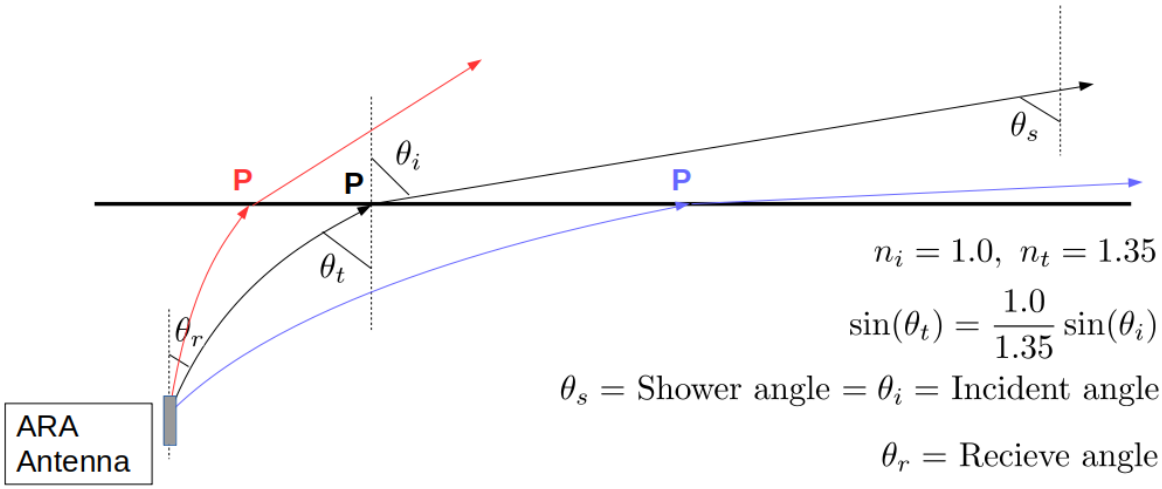


Figure 6.4: Dependence of the air/ice transect point P on the incoming shower angle. The red and blue ray paths show how the variation of the shower elevation can change the position of this point w.r.t the receiver. The refractive index of ice at the surface at the South Pole is taken to be 1.35.

These positions depend on the incoming angle of the shower (or the primary cosmic ray), as shown in Figure 6.4. Depending on the elevation direction of the incoming shower (θ_s), this point P (on the ice surface) can be further or nearer to the receiving antenna. Once the incoming shower

zenith angle is fixed, raytracing codes can be used to find this point P for that specific antenna. These 16 P coordinates are then passed to COREAS to yield the E-fields at these points.

To specify P, we proceed as follows. We know the start coordinates (x_0, z_0) for the ARA antenna and the angle θ_t at which the ray enters (or leaves) the ice for fixed shower elevation θ_s. For $x_0 = 0$, we find the distance x_1 (the point P) on the ice, by minimizing:

$$f(z_0, z_1, \theta_t, \theta_r) = \tan\left(\arcsin\left(\frac{L(\theta_r, z_0)}{A + Be^{Cz_1}}\right)\right) - \tan(\theta_t) = 0, \tag{6.2}$$

and finding the value of θ_r which satisfies this equation.

Once we obtain the launch angle θ_r, we can easily find x_1 (or the point P) at the ice surface for each of the ARA antennas. Then since we know the horizontal distance of each of these points from their respective antennas and the azimuth of the incoming shower, these points can be situated correctly on the ice surface after doing the required translation and rotation.

6.3 Propagating the E-fields from the surface to the ARA antennas

Once we have the E-fields in the Cartesian coordinate system from CoREAS at each of our hit points on the ice surface, we convert the E-fields from Cartesian coordinate $(\hat{x}, \hat{y}, \hat{z})$ to spherical coordinates $(\hat{r}, \hat{\theta}, \hat{\phi})$. This conversion uses the standard matrix transformation:

$$\begin{bmatrix} \hat{x} \\ \hat{y} \\ \hat{z} \end{bmatrix} = \begin{bmatrix} \sin(\theta)\cos(\phi) & \cos(\theta)\cos(\phi) & -\sin(\phi) \\ \sin(\theta)\sin(\phi) & \cos(\theta)\sin(\phi) & \cos(\phi) \\ \cos(\theta) & -\sin(\theta) & 0 \end{bmatrix} \begin{bmatrix} \hat{r} \\ \hat{\theta} \\ \hat{\phi} \end{bmatrix} \tag{6.3}$$

The transformed unit vectors are then substituted into Equation (6.1) to obtain:

$$\vec{E}(t) = E_r(t)\,\hat{r} + E_\theta(t)\,\hat{\theta} + E_\phi(t)\,\hat{\phi} \tag{6.4}$$

where:

$$\begin{aligned} E_r(t) &= E_x(t)\sin(\theta)\cos(\phi) + E_y(t)\sin(\theta)\sin(\phi) + E_z(t)\cos(\theta) \\ E_\theta(t) &= E_x(t)\cos(\theta)\cos(\phi) + E_y(t)\cos(\theta)\sin(\phi) - E_z(t)\sin(\theta) \\ E_\phi(t) &= -E_x(t)\sin(\phi) + E_y(t)\cos(\phi) \end{aligned} \tag{6.5}$$

Here θ and ϕ are fixed by the incoming shower direction (θ_s, ϕ_s). E_r will be ignored as it contributes no signal; only the E_θ and E_ϕ components are propagated in ice.

Before propagation in ice, we multiply the E-fields by the transmittance factor (determined from the standard Fresnel Equations). The fraction of the incident power that is reflected from the air-ice interface is given by the reflectance R, and the fraction that is transmitted is given by the transmittance T $(= 1 - R)$.

In cases where two media have permittivities such that $\mu_1 \approx \mu_2 \approx \mu_0$ (approximately true for our air-ice interface), we can write the reflectance for s-polarization (ϕ-polarization) as:

$$R_s = \left| \frac{n_1\cos(\theta_i) - n_2\sqrt{1 - \left(\frac{n_1}{n_2}\sin(\theta_i)\right)^2}}{n_1\cos(\theta_i) + n_2\sqrt{1 - \left(\frac{n_1}{n_2}\sin(\theta_i)\right)^2}} \right|^2 \quad \Rightarrow \quad T_s = 1 - R_s \tag{6.6}$$

so the modified ϕ component is $E^*_\phi = E_\phi\sqrt{T_s}$. The p-polarized (θ polarization) is given by:

$$R_p = \left| \frac{n_1\sqrt{1 - \left(\frac{n_1}{n_2}\sin(\theta_i)\right)^2} - n_2\cos(\theta_i)}{n_1\sqrt{1 - \left(\frac{n_1}{n_2}\sin(\theta_i)\right)^2} + n_2\cos(\theta_i)} \right|^2 \quad \Rightarrow \quad T_p = 1 - R_p \tag{6.7}$$

so the modified θ component is $E^*_\theta = E_\theta\sqrt{T_p}$. Here in both of these cases I have taken $n_1 = 1$, $n_2 = 1.35$, with θ_i to be the angle of incidence on ice.

6.4 Using the Complex Effective Height to Obtain the Signal Voltage

Once we have propagated the E-field (V/m) to the antenna by raytracing, we need to take the dot product of the E-field vector with the complex height effective (h_{eff} in m) vector to give us the voltage (V) recorded at the antenna. The effective height is a measure of the voltage induced on the open-circuit terminals of the antenna when an E-field wave impinges upon it, and is a function of frequency and direction (θ, ϕ). The h_{eff} vector can be written as:

$$\vec{\tilde{h}}_{eff}(f,\theta,\phi) = \tilde{h}_{eff,\theta}(f,\theta,\phi)\ \hat{\theta} + \tilde{h}_{eff,\phi}(f,\theta,\phi)\ \hat{\phi} \tag{6.8}$$

Before we take the dot product of the E-fields with h_{eff}, we Fourier Transform into frequency space, such that the following complex dot product can be calculated to give the complex voltage in frequency space:

$$\tilde{V}_{OC}(f) = \vec{\tilde{E}}(f,\theta,\phi).\vec{\tilde{h}}^{*}_{eff}(f,\theta,\phi) \tag{6.9}$$

It should be noted that there should be a fractional distance factor that accounts for the distance traveled by the EM fields in ice. Assuming that all the radio emission comes from the shower maximum (X_{max}), the fractional distance factor is defined as the total optical path length from X_{max} to the point on the surface divided by the total optical path of from X_{max} to the in-ice receiving antenna.[2]

We assume conjugate impedance matching which basically means that for simple antennas such as dipoles the induced voltage into a matched receiver load is given by (Allison et al., 2012)

$$\tilde{V}_{induced}(f) = \frac{\tilde{V}_{OC}(f)}{2} \tag{6.10}$$

[2]This fractional distance factor was not included in the initial calculations done for the template matching analysis described herein, but is now included by default.

The functional form of the effective height for the ARA antennas is given by (Allison et al., 2012):

$$|\vec{\tilde{h}}_{eff}| = 2\sqrt{\frac{\Re(Z_L)A_{eff}}{Z_0}}; \tag{6.11}$$

here $Z_0 = 120\pi\ \Omega$ (the impedance of free space), $\Re(Z_L) = 50\ \Omega$ (the load impedance), $A_{eff} = (G(\theta,\phi)c^2)/(4\pi f^2)$ is the effective area of the antenna and $G(\theta,\phi)$ is the gain of the antenna. We can then rewrite Equation (6.11) as:

$$|\vec{\tilde{h}}_{eff}(f,\theta,\phi)| = 2\sqrt{\frac{\Re(Z_L)}{(Z_0)}\left(\frac{c}{f}\right)^2\frac{G(\theta,\phi)}{4\pi}} \tag{6.12}$$

In ice we must rescale $c \Rightarrow c/n$, $Z_0 \Rightarrow Z_0/n$ and $Z_L \Rightarrow Z_L/n$ and therefore the formula becomes:

$$\boxed{|\vec{\tilde{h}}_{eff}(f,\theta,\phi,z)| = 2\sqrt{\frac{\Re(Z_L)}{(Z_0)}\left(\frac{c}{f.n(z)}\right)^2\frac{G(\theta,\phi)}{4\pi}}} \tag{6.13}$$

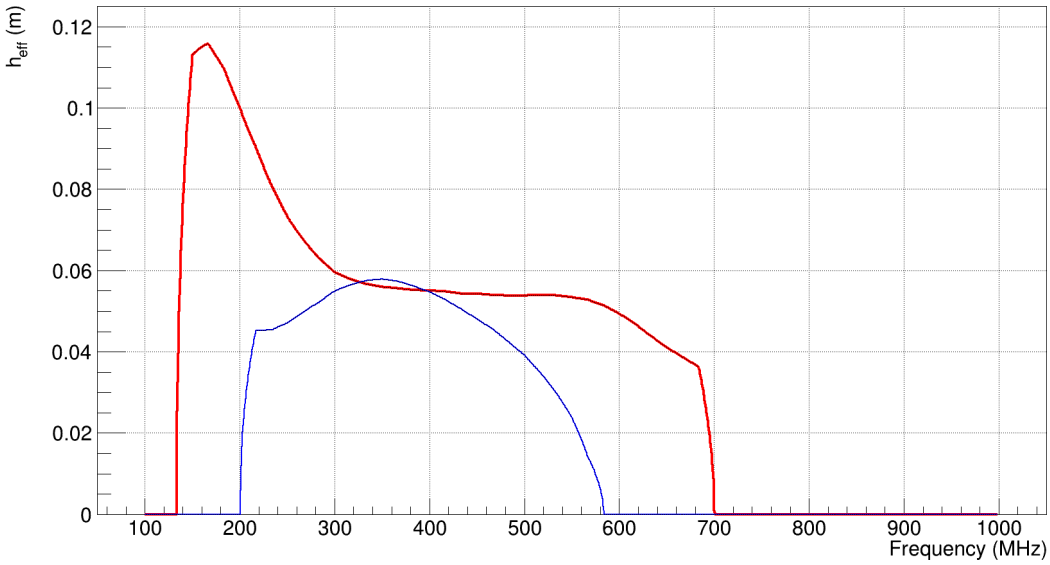


Figure 6.5: Effective height calculated for the ARA antennas at $n = 1.78$, $\theta = 45°$ and $\phi = 45°$. Blue line is for the quad slot cylinder antenna and the red line for the birdcage bicone antenna.

A more careful derivation of this formula is also given in Appendix A of Jordan Hanson's thesis (Hanson, 2013). Equation 6.13 is used to calculate the effective height for the quad slot cylinder antennas (the Hpol antennas) of ARA, using simulated gain $G(\theta,\phi)$. The values of the gain $G(\theta,\phi)$ were calculated using the XFDTD simulation by our colleagues in Chiba University. We assume (Allison et al., 2012):

$$\vec{\tilde{h}}_{eff} = |\vec{\tilde{h}}_{eff}|\ \hat{\phi}; \tag{6.14}$$

based on the azimuthal symmetry of the antenna, and for the Vpol birdcage bicone antenna it can be written as:

$$\vec{\tilde{h}}_{eff} = |\vec{\tilde{h}}_{eff}|\ \hat{z}; \tag{6.15}$$

which is consistent with expectation for the vertically polarized antennas.

6.5 Convolving with the ARA System Response

We now include the ARA system response, shown in Figure 4.3. We use the phase to multiply the complex gain by the complex voltage in frequency space:

$$\tilde{V}_{final}(f) = \tilde{V}_{induced}(f).\tilde{G}_{ARA}(f).\frac{1}{\sqrt{2}}; \tag{6.16}$$

here the $1/\sqrt{2}$ factor arises from the power splitter in the ARA data acquisition signal chain. The $\tilde{G}_{ARA}(f)$ takes into account all the amplification and filtration that happens in the ARA signal chain.

6.6 Final Template Waveforms

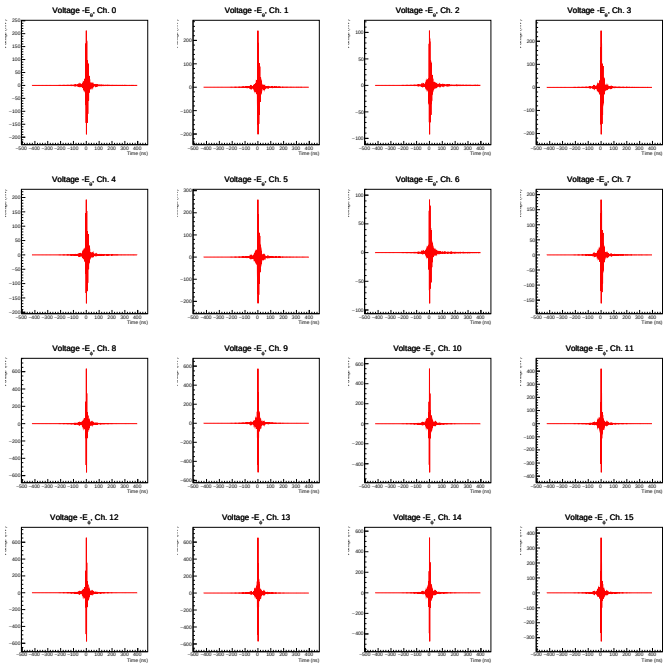


Figure 6.6: Simulated voltages (mV) vs. time (ns) in all the 16 channels of an ARA station. The top row shows the four Top Vpol antennas in string 1 (first column), string 2 (second column), string 3 (third column) and string 4 (fourth column) respectively, the second row shows the Bottom Vpol channels, the third row shows the Top Horizontal channels and the fourth row shows the Bottom Hpol channels. The channels are numbered from 0 to 15 such that channels 0-3 (from left to right) are in the first row, channel 4-7 are in the second row, 8-11 are in the third row and 12-15 are in the fourth and last row.

We now show results for a shower simulated with $\theta = 30°$, $\phi = 90°$ and 1×10^{18} eV. The hadronic models used were QGSJETII.04 for high energy showers and UrQMD 1.3 for low energy showers. This simulation was done with thinning ON, with a thinning fraction of 10^{-6}. The Hpol voltages are relatively stronger than the Vpol ones in Figure 6.6, since the geomagnetic radio emission is primarily Hpol. This is because of the fact that at the South Pole the magnetic field is vertical so the geomagnetic radio emission has a stronger signal in Hpol.

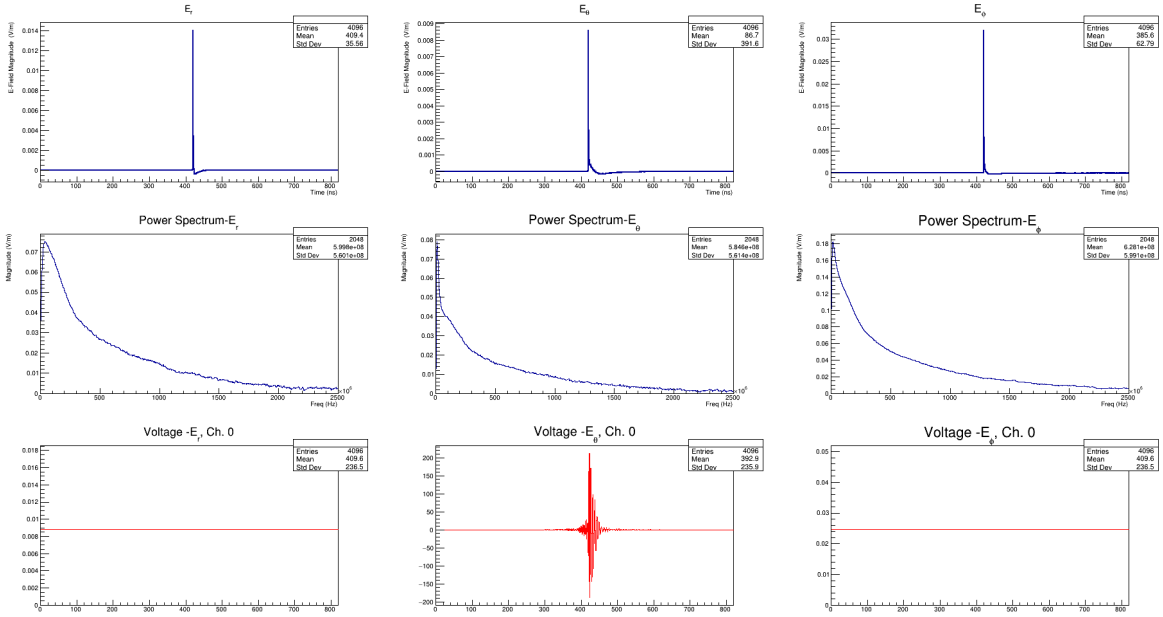


Figure 6.7: Top row shows E_r, E_θ and E_ϕ, for Vpol Channel 0, where the latter two components have been multiplied by their transmittance. The second row shows their DFTs and the third row shows the final voltage waveforms we obtain in Channel 0 after simulating the full signal chain.

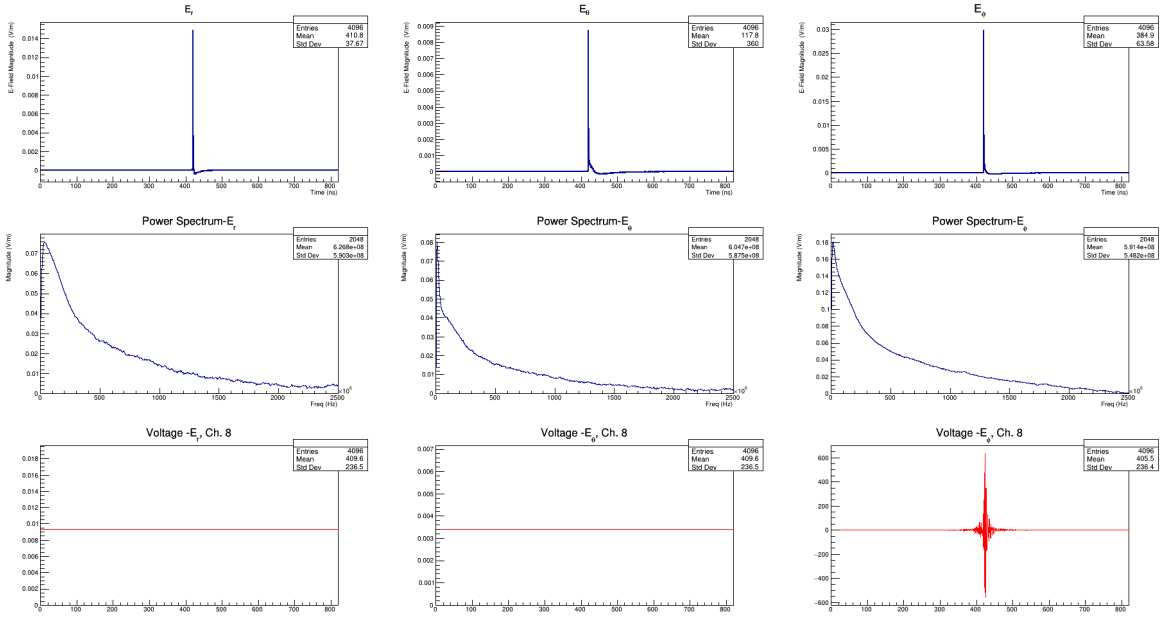


Figure 6.8: Top row shows E_r, E_θ and E_ϕ, for Hpol Channel 8, where the latter two components have been multiplied by their transmittance. The second row shows their DFTs and the third row shows the final voltage waveforms we obtain in Channel 8 after simulating the whole signal chain.

Chapter 7

Finding UHECRs in ARA-2 10% Unblinded Data Sample

ARA data sets are broken up into two types of data sets: the "burn" sample (10% unblinded data sample) and the full sample (100% blinded data sample). The "burn" sample data set is made by randomly selecting 1 in every 10 events in each data run and is used in the phase prior to unblinding, in which an analyst refines cuts, understands backgrounds and finds detector efficiencies for the "signal", which in this case is a UHECR shower signal. Once the analysis cuts are refined on the "burn" sample, and the backgrounds have been characterized, the full sample data set can then be analyzed to search for the presence of UHECR signal. Breaking up the data set, and therefore the analysis, into two steps helps the analyst avoid introducing bias in the cuts by enabling the designing of cuts that will not eliminate events of interest but instead will prevent them from being rejected as "background" events.

The goal of this analysis was to identify a candidate UHECR sample in the ARA-2 and ARA-3 2013 to 2016 "burn" sample data. The templates made by CoREAS, as described in Chapter 6, would be used in a template matching analysis method to extract UHECR candidate events from ARA data. The following summarizes the analysis structure:

1. Introducing and setting up the UHECR template matching method.

2. Verifying the feasibility of the template matching method on previously identified UHECR candidate events, which were found in an ARA-2+ARA-3 neutrino search.

    (a) Since, in this case, the UHECR candidate events had already been identified in ARA-2, and the event geometries (i.e., reconstructed zenith and azimuth angles) already

known through reconstruction, "geometry-specific" UHECR templates were generated for each of the UHECR candidate events for the template matching process.

3. Setting up UHECR templates in order to carry out the UHECR candidate event search in ARA data.

   (a) In this case, "geometry-specific" UHECR templates could not be generated as no event geometries, for possible UHECR candidate events were known *a priori* the UHECR search. In order to conduct the UHECR search in the ARA data, UHECR templates covering a variety of directions were generated to get an average template match score for each ARA event in the data set.

4. Carrying out the UHECR event search for ARA-2 in the "burn" sample data for 2013 to 2016.

   (a) At this stage cuts to identify and isolate viable UHECR candidate events were defined. These included:

      i. Level-0 cuts: These were made to cut out any non-impulsive events in the "burn" sample data just based on the template match scores. Low template match scores with UHECR templates meant the event was, most likely, not from an impulsive radio source.

      ii. Background identification: These cuts, applied to the "burn" sample events that passed Level-0 cuts, were specifically made to identify and characterize anthropogenic impulsive radio sources present around the South Pole and the ARA stations throughout the year.

      iii. Level-1 cuts: These cuts were made to identify viable UHECR candidate events present in the "burn" sample data events that passed the Level-0 criteria. The Level-1 cuts were based on the geometry-specific template match score and how isolated the UHECR candidate event was from other events that passed the Level-0 cuts.

iv. Calibration pulser (or calpulser) events and forced trigger events (or Software trigger events) were also rejected in the ARA data sample used for this analysis. Calpulser events are triggers caused by the radio pulses that are emitted every second by calibration pulsers. Software triggers are forced triggers where the ARA station is forced, by the on-board station computer, to trigger. These events are pre-tagged in the ARA data sample.

5. Following the ARA-2 analysis, a similar analysis was conducted for the 2013-2016 ARA-3 burn sample data.

6. Finally, the expected UHECR detected rate for four-years of ARA burn-sample data is calculated for comparison.

In this chapter, after first discussing the template matching method, the ARA-2 UHECR search analysis will be discussed. The ARA-3 UHECR search analysis will be discussed in the following chapter.

7.1 Template Matching Proof-of-principle

Before full-scale application of template matching, we wish to demonstrate the veracity of the procedure. As a part of his neutrino search in ARA-2 data for the years of 2013 to 2016, Ohio State University (OSU) graduate student Brian Clark also found several UHECR candidate events in the 100% data set for those four years (Clark, 2019). These events fulfilled very rudimentary requirements: they were required to be down-coming, have more or equal power in Hpol as compared to Vpol and were required to be isolated enough in time from other impulsive events. Therefore these UHECR candidate events (Table 7.1) provided a good starting point to test out the template matching method. It should be noted that ARA data runs from the full (100%) sample were used for this feasibility study with UHECR candidate events found in the OSU neutrino analysis.

Table 7.1: The UHECR candidate events that were found by Ohio State University graduate student Brian Clark as a part of his neutrino search analysis for ARA-2. (**right**) Table of events that passed Vpol&Hpol cut of the OSU analysis; (**left**) Table of events passed the Hpol cut only of the OSU analysis.

Sr. No.	Run No.	Event No.
1	1455	22814
2	1571	120910
3	2779	8977
4	2871	41557
5	2937	111277
6	3202	106330
7	3206	126972
8	3325	81004
9	3392	84494
10	3529	34113
11	5009	28578
12	6674	35879
13	6861	57376
14	7170	147

Sr. No.	Run No.	Event No.
1	2165	121109
2	2869	29331
3	3663	180627
4	5505	39072
5	6610	36035
6	6675	78233

The direction (i.e., zenith and azimuth angles) of each UHECR candidate event was reconstructed using interferometry. Interferometry uses the hit time of the pulses in each channel (or antenna) relative to each other in order to triangulate the most likely source position with respect to the receiver stations. The interferometer used for this study took into account the bending of radiowaves due to refraction and returned the angles at which the station received the rays.

It was assumed that the radio emissions from a UHECR shower are collinear with the shower propagation direction and that the events given in Table 7.1 were due to a UHECR shower. Using the aforementioned assumptions, the angle of incidence of the UHECR shower on the ice surface was determined using the zenith angle of incidence at the station to trace the ray back to the ice surface. Events which gave incidence angles larger than the critical angle in ice at the ice-air boundary were rejected, as the radio pulses in that case likely originated from an in-ice source. UHECR shower cores hitting the ice could also be possible in-ice sources; however, it would be challenging to correctly identify them using in-air UHECR templates. Doing so would require that CoREAS propagate showers deep into the ice (that work is currently in progress). All the events listed in Table 7.1 pointed to sources above the surface.

After the UHECR candidate event geometry was obtained, a "geometry-specific" template was generated for the event using CoREAS at 1×10^{18} eV energy. The simulated event template was then matched with its corresponding data event and all the other RF events in the corresponding run.

RF events are events in ARA data that have been triggered by some radio activity. They mostly consist of thermal noise events and events having an anthropogenic origin but can also consist of UHECR and neutrino events.

Whenever a UHECR template was matched with an ARA event, this was achieved by cross-correlating each channel waveform of the UHECR template with its corresponding ARA event channel waveform. The maximum correlation score obtained from matching each ARA event channel waveform with its corresponding channel waveform in the UHECR template was then assigned to be the correlation score of that channel. The average of all eight Vpol channel correlation scores was then taken to get an average correlation score, which was defined as the Vpol correlation score for that ARA event for the given UHECR template. The same process was repeated with the eight Hpol channels to get a Hpol correlation score. This process was repeated for all the UHECR candidate events in Table 7.1 to get correlation scores for each of them for each polarization. It should be noted that in all the calculations for ARA-2 analysis, channel 15, an Hpol channel, was excluded as that channel was mostly corrupted with noise.

7.1.1 Comparing Event Performance through *CorSNR*

As can be seen from Figure 7.1, the correlation scores for the UHECR candidate events were distinctively higher than the surrounding RF events in the corresponding run. This Figure nearly has a delta function-like feature; most of the UHECR candidate events listed in Table 7.1, showed this feature.

To compare the template matching scores of UHECR candidate events relative to other UHECR

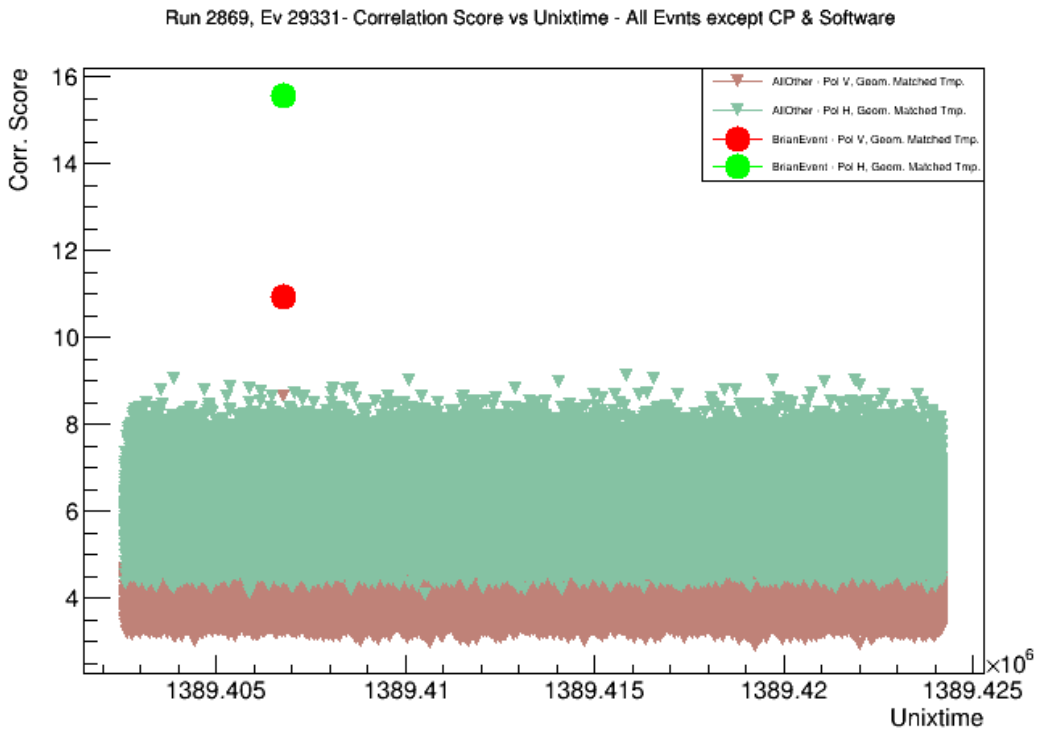


Figure 7.1: Correlation scores of Vpol and Hpol vs. unixtime of the events in run 2869 of ARA-2. These were obtained as a result of template matching for event 29331 where the "geometry specific" template for that event was matched with all the remaining RF events in the run. The red color is for Vpol scores and green is for Hpol scores. Circles signify score for event 29331 (UHECR candidate event) and triangles show the scores of the remaining RF events in the run.

candidate events, a new normalized quantity was defined:

$$CorSNR = \frac{CorScore_{Event} - AvgCorScore_{RFEvents}}{RMSCorScore_{RFEvents}} \tag{7.1}$$

here $CorScore_{Event}$ is the correlation score of the event in question, $AvgCorScore_{RFEvents}$ is the average correlation score of all the RF events in the given run and $RMSCorScore_{RFEvents}$ is the RMS of the correlation scores of all the RF events in the given run. Each event would then have a *CorSNR* value corresponding to each of its correlation scores from each polarization.

CorSNR values for all UHECR candidate events, listed in Table 7.1, were calculated using the definition given in Equation 7.1. The results are summarised by Figures 7.2 and 7.3.

Figures 7.2 and 7.3 show the *CorSNR* distributions of the UHECR candidate events (red dis-

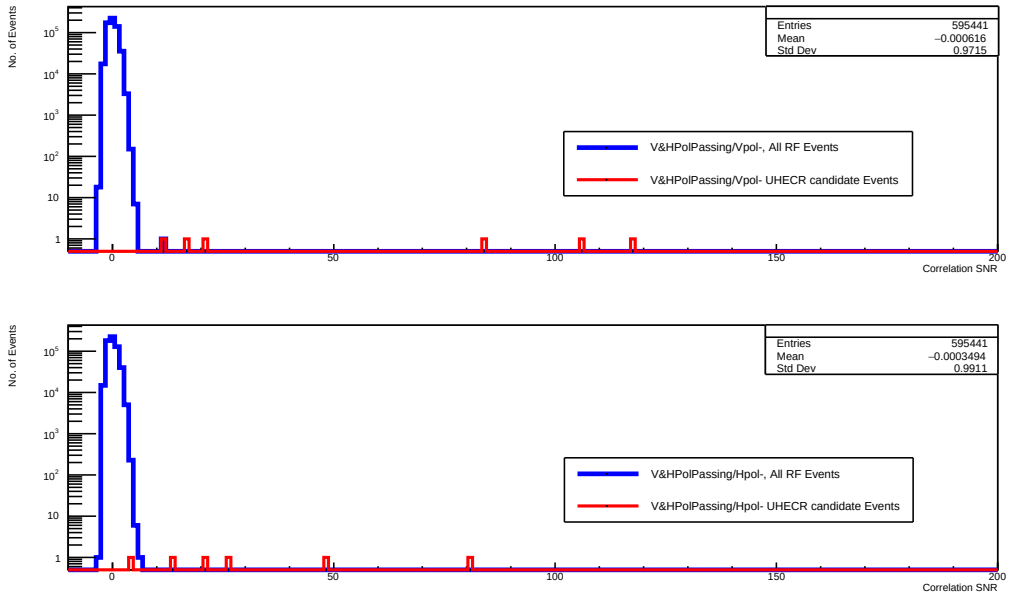


Figure 7.2: Distributions of *CorSNR* values of UHECR candidate events which passed V&Hpol cuts of the OSU neutrino analysis. The (**top**) Vpol and (**bottom**) Hpol *CorSNR* distributions are shown separately. RF event *CorSNR* distributions, consisting of RF events taken from all the runs corresponding to UHECR candidate events are represented by blue histograms and UHECR candidate event *CorSNR* distributions are represented by red histograms.

tribution) and RF events (blue distribution) that passed the V&Hpol and the Hpol-only cuts of the OSU neutrino analysis respectively. Figure 7.2 shows that *CorSNR* values for UHECR candidate events are all much higher than the RF event *CorSNR* values. UHECR candidate events that passed the V&Hpol cut of the OSU analysis have relatively higher Vpol *CorSNR* values, above the RF event *CorSNR* level, than the Hpol *CorSNR* values. Figure 7.3 shows that *CorSNR* values for UHECR candidate events are all relatively close to the RF event *CorSNR* values. Since Figure 7.3 consists of UHECR candidate events that passed the Hpol-only cut of the OSU analysis, the Hpol *CorSNR* values for UHECR candidate events are relatively higher than the RF event *CorSNR* level, as compared to the Vpol *CorSNR* values, consistent with the UHECR hypothesis.

It can also be observed from the *CorSNR* distributions in Figures 7.2 and 7.3 that there are two RF events that rise distinctly above the blue distribution of RF event *CorSNR* values. Figure 7.2 has one RF event, event 29309 from run 2869, which has a slightly higher *CorSNR* value in Vpol

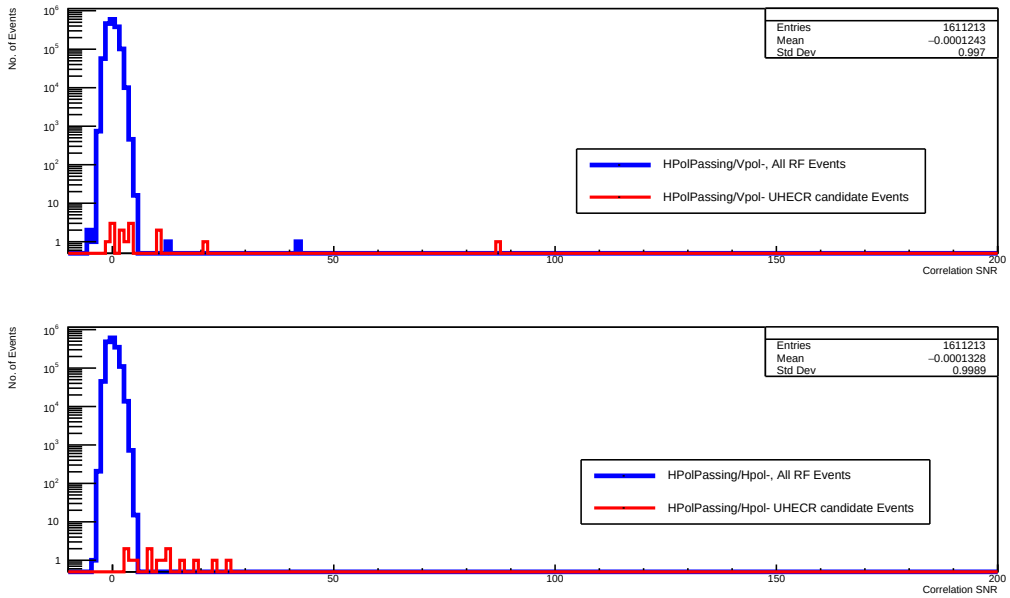


Figure 7.3: Distributions of *CorSNR* values of UHECR candidate events which passed the Hpol cut only of the OSU neutrino analysis. The (**top**) Vpol and (**bottom**) Hpol *CorSNR* distributions are shown separately. RF event *CorSNR* distributions, consisting of RF events taken from all the runs corresponding to UHECR candidate events, are represented by blue histograms and and UHECR candidate event *CorSNR* distributions are represented by red histograms.

as compared to other RF event Vpol *CorSNR* values. Figure 7.3 has two RF events that have Vpol *CorSNR* values that are slightly higher above the RF event *CorSNR* level namely: event 43391 from run 6861 and event 28579 from run 5009. These three events are not present in the list of UHECR candidate events in Table 7.1 but are each present in the runs corresponding to three of the UHECR candidate events. Therefore these events are "adjacent" to the UHECR candidate events; waveforms for these adjacent events can be found in Chapter A of the Appendix.

The template matching exercise clearly demonstrated that impulsive events, like the UHECR candidate events, were easily picked out by the template matching method as they generally gave relatively high *CorSNR* values in at least one of the polarization.

7.2 Fake Rate of RF Events

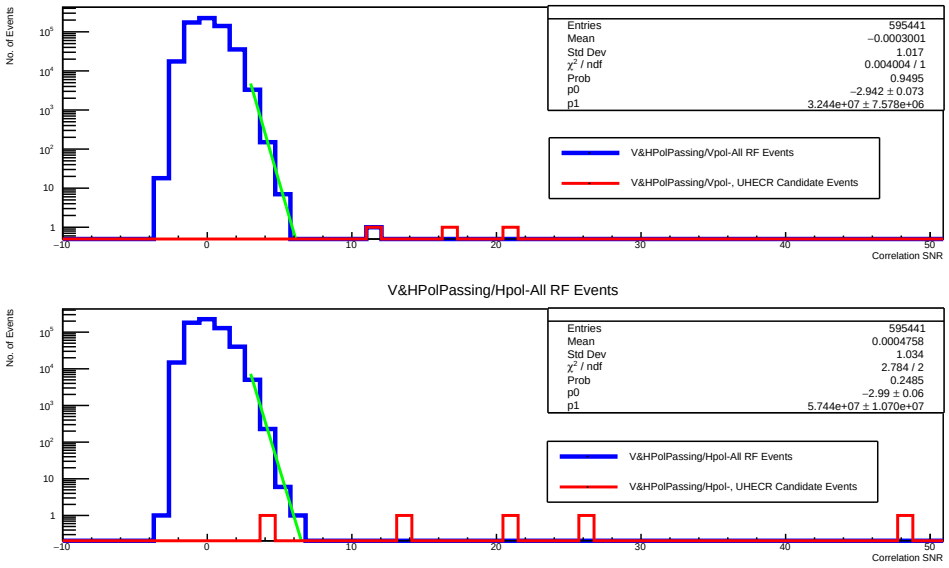
The fake rate of RF events, which are mostly thermal, is a measure of the probability of a RF event mimicking an impulsive event like a UHECR event. If this rate is high, then a large number of thermal noise events can be misidentified as UHECR candidate events through the template matching process, which can be problematic. The misidentification would happen because, in this case, RF events would have a *CorSNR* value of around 8, bringing them into the *CorSNR* region where UHECR candidate events, listed in Table 7.1, generally reside.

In order to calculate the fake rates the first step was to fit the tails of the RF event distributions (blue distributions), on the side where *CorSNR* values were greater than zero, in Figures 7.2 and 7.3 with a simple exponential function. The exponential function was defined as:

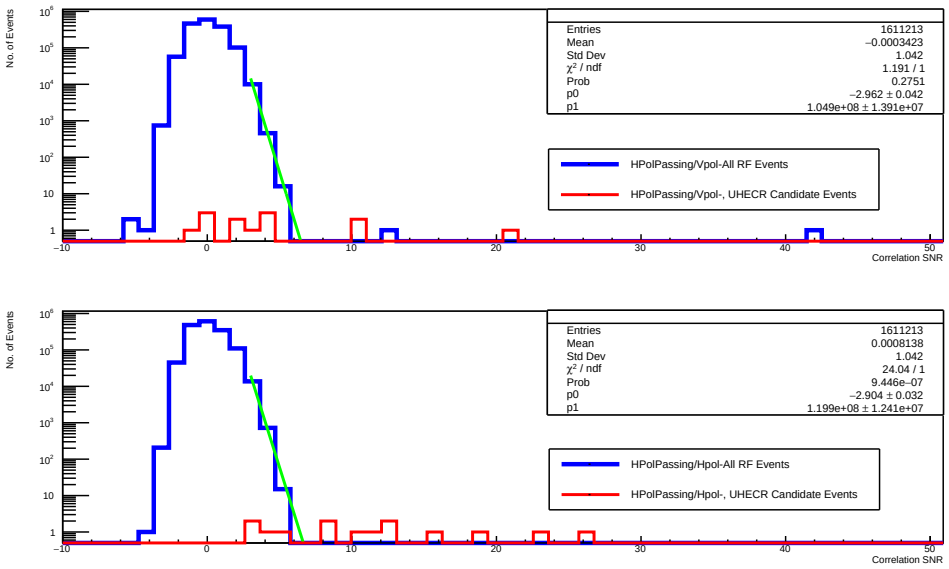
$$f(x) = \text{Par}[1]\ \exp(-x\ \text{Par}[0]) \tag{7.2}$$

where x is *CorSNR*, $f(x)$ is the number of events and the values of Par[0] and Par[1] were determined by fitting. The fit range was set from a *CorSNR* value of 2.975 to 8 as that was where the tails with high *CorSNR* value, of the blue distributions, were found. Fits to the overall shape of of the RF event *CorSNR* distributions have been discussed in Chapter B of the Appendix.

Using the fit values of Par[0] and Par[1], the fit function given in Equation 7.2 was then integrated from a *CorSNR* value of 8 to 1000 to get the sum of all the possible events in that range. 1000 was used as a proxy for infinity because of the limitations of the numerical integrator being used. The result of the integration was then divided by the total number of events to obtain the fake rate. The fake rates for UHECR candidate events, together with the fit results, are listed in Table 7.2. The fits to the RF event *CorSNR* distributions are shown in Figures 7.4a and 7.4b.



(a) Events from runs containing UHECR candidate events that passed the V&Hpol cut of the OSU analysis.



(b) Events from runs containing UHECR candidate events that passed the Hpol cut of the OSU analysis.

Figure 7.4: Fits to the tails of distributions of RF event *CorSNR* values taken from runs containing UHECR candidate events that passed V&Hpol (**Figure (a)**) and Hpol-only (**Figure (b)**) cuts of the OSU neutrino analysis. The top and bottom plot of Figures (a) and (b) contain Vpol and Hpol *CorSNR* distributions, respectively. RF event *CorSNR* distributions, consisting of RF events taken from all the runs corresponding to UHECR candidate events, are represented by blue histograms and UHECR candidate event *CorSNR* distributions are represented by red histograms. Figures (a) and (b) are zoomed-in versions of Figures 7.2 and 7.3.

Table 7.2: Exponential Tail Fit Results & Fake Rates. The fit range for the exponential tail covered $2.975 < CorSNR < 8$.

No.	UHECR Event Type-Pol	Par[0]	Par[1] $(\times 10^7)$	χ^2/NDF	Fake Rate $(\times 10^{-9})$
1.	V&HPol Events-Vpol	2.942 ± 0.073	(3.244 ± 0.758)	$0.004/1$	1.113
2.	V&HPol Events-Hpol	2.990 ± 0.059	(5.744 ± 1.070)	$2.784/2$	1.317
3.	HPol Events-Vpol	2.962 ± 0.042	(10.487 ± 1.391)	$1.191/1$	1.126
4.	HPol Events-Hpol	2.904 ± 0.032	(11.993 ± 1.241)	$24.04/1$	2.084

7.2.1 The Dependence of Fake Rates on Fit Range

Table 7.3: Exponential Tail Fit Results & Fake Rates. The fit range for the exponential tail covered $3.975 < CorSNR < 9$.

No.	UHECR Event Type-Pol	Par[0]	Par[1] $(\times 10^7)$	χ^2/NDF	Fake Rate $(\times 10^{-9})$
1.	V&HPol Events-Vpol	2.919 ± 0.376	(2.940 ± 4.677)	$(8.25 \times 10^{-9})/0$	1.223
2.	V&HPol Events-Hpol	3.421 ± 0.397	(36.386 ± 60.764)	$0.697/1$	0.232
3.	HPol Events-Vpol	3.192 ± 0.246	(28.082 ± 29.143)	$(4.88 \times 10^{-9})/0$	0.442
4.	HPol Events-Hpol	3.691 ± 0.254	(355.830 ± 379.662)	$(0.67 \times 10^{-9})/0$	0.090

Investigation of the dependence of fake rates on the fit range was carried out by adjusting the fit range and re-fitting the tails to calculate the new fake rates. First, the whole fit range was shifted by $+1$ *CorSNR* units, which made the fit range start from a *CorSNR* value of 3.975 to end at 9. The fit results for this range are listed in Table 7.3.

It can be inferred from the fit results reported in Table 7.3 that shifting the whole fit range by $+1$ *CorSNR* units certainly deteriorated the quality of the fits. The errors on the fit parameters increased by orders of magnitude and the χ^2/NDF values also became nonphysical. The whole fit range was then shifted by -1 *CorSNR* units which made the fit range start from a *CorSNR* value

of 1.975 to end at 7. The fit results for this range are listed in Table 7.4.

Table 7.4: Exponential Tail Fit Results & Fake Rates. The fit range for the exponential tail covered $1.975 < CorSNR < 7$.

No.	UHECR Event Type Pol	Par[0]	Par[1] $(\times 10^7)$	χ^2/NDF	Fake Rate $(\times 10^{-9})$
1.	V&HPol Events-Vpol	2.394 ± 0.014	(0.513 ± 0.015)	119.5/2	17.328
2.	V&HPol Events-Hpol	2.251 ± 0.011	(0.434 ± 0.010)	502.6/3	48.839
3.	HPol Events-Vpol	2.367 ± 0.008	(1.393 ± 0.024)	482/2	21.746
4.	HPol Events-Hpol	2.239 ± 0.006	(1.160 ± 0.017)	1267/2	53.622

The fit results from Table 7.4 clearly show that shifting the fit range by -1 *CorSNR* units reduces the fit errors on the parameters and increases the χ^2/NDF values by at least two orders of magnitude. The reduced errors on the fit parameters indicate that the fitter gave better convergence for the fit parameters, and the high χ^2/NDF values indicate that the tails certainly do not follow an exponential profile. It should also be noted that the fake rates have increased by order of magnitude compared to the fake rate values given by the original fit range covering a *CorSNR* range from 2.975 to 8.

The fit results of Table 7.4 imply that the fake rates provided in Table 7.2 have probably been underestimated. This is because even though the exponential function, given in Equation 7.2, provided a reasonable fit for the tails, we have to be cautious in the interpretation of the χ^2/NDF values since the χ^2 is mostly determined by the largest-statistics bins, which generally results in an overestimate of the slope magnitude. Therefore these fake rate values can be taken as a lower limit on the actual fake rate values. The upper limit can be taken as the inverse of the total number of events, which gives a value of around 10^{-6} for the fake rates.

7.3 UHECR Template Analysis for 10% Unblinded Data Sample.

In order to start the template-matching analysis for the "burn" sample, an 'average' UHECR template is needed *a priori*, as interferometry and vertex reconstruction have not yet been performed,

and there is also no way of knowing *a priori* which possible UHECR candidates might be present in the data set. Therefore no "geometry correct" templates can be used in the pre-analysis stage.

Eight UHECR templates were therefore generated at an energy of 1×10^{18} eV that covered a variety of zenith and azimuth angles. The templates have been listed in Table 7.5 and their waveforms are presented in Chapter C of the Appendix.

Table 7.5: Summary of azimuth and elevation (CORSIKA coordinate system) of CoREAS templates used for first-pass (pre-vertex reconstruction) UHECR candidate event selection for 10% unblinded data sample.

Template No.	Zenith (°)	Azimuth (°)
Template 0	30	0
Template 1	30	90
Template 2	30	180
Template 3	30	270
Template 4	60	0
Template 5	60	90
Template 6	60	180
Template 7	60	270

These eight templates were then used to identify possible UHECR candidates in the "burn" sample data in a new template matching method called the "eight template average method". In this method, these eight UHECR templates were individually matched with a given data event, and *CorSNR* values for each polarization for each template were calculated. The final Vpol *CorSNR* value for the given data event was set to be the average of the eight Vpol *CorSNR* values obtained from each of the eight matched templates. The same template matching method was also applied to calculate a Hpol *CorSNR* for the data event. The final *CorSNR* values for each polarization were then used to determine if the data event in question was a possible UHECR candidate or not.

Table 7.6: *CorSNR* values for all of UHECR candidates that were listed in Table 7.1. Columns 4 and 5 show the eight template averaged *CorSNR* values for Vpol and Hpol respectively. Columns 6 and 7 show the "geometry correct" template *CorSNR* values for Vpol and Hpol respectively. Columns 8 and 9 show the reconstructed zenith and azimuth values of the possible UHECR shower that would have possibly produced that ARA event.

	Run	Event	8Tmp.Avg. Vpol *CorSNR*	8Tmp.Avg. Hpol *CorSNR*	Geom.Corr. Vpol *CorSNR*	Geom.Corr. Hpol *CorSNR*	UHECR Shower Reco. Zen.(°)	UHECR Shower Reco. Azi.(°)
1	2165	121109	132.841	88.506	117.730	47.954	86.513	222.976
2	2869	29331	11.365	11.001	16.839	13.245	63.667	307.976
3	3663	180627	110.110	43.000	106.511	26.180	52.184	131.929
4	5505	39072	83.969	88.552	84.069	80.845	45.053	322.976
5	6610	36035	6.436	0.817	11.920	3.960	63.667	320.976
6	6675	78233	12.375	3.022	20.619	21.091	52.184	334.976
7	1455	22184	1.087	4.360	1.704	8.711	60.465	314.976
8	1571	120910	2.752	4.283	3.687	22.626	60.465	23.929
9	2779	8977	1.791	3.136	3.377	15.225	60.465	201.976
10	2871	41557	9.368	3.250	10.872	8.030	60.465	169.929
11	2937	111277	1.486	3.792	1.795	19.312	86.513	209.976
12	3202	106330	-0.370	1.241	-0.205	3.239	40.681	175.929
13	3206	126972	9.588	17.193	10.097	12.054	51.062	14.929
14	3325	81004	3.335	3.662	3.910	11.538	60.465	123.929
15	3392	84494	0.643	7.647	0.481	26.341	76.678	172.929
16	3529	34113	2.203	7.251	3.891	12.852	63.667	217.976
17	5009	28578	89.864	4.674	87.433	3.399	63.667	298.976
18	6674	35879	-0.239	1.500	-0.921	3.894	49.712	298.976
19	6861	57376	-0.092	0.762	-0.284	5.678	63.667	91.929
20	7170	147	12.022	1.987	21.010	10.305	67.240	165.929

The eight template average method was then applied to the known UHECR candidate events, listed in Table 7.1. This was done to verify that this method was adequate, compared to matching single "geometry correct" templates, in identifying an initial UHECR candidate event sample. The

results of using the eight template average method for the given UHECR candidate events are shown in Table 7.6.

The *CorSNR* values of UHECR candidate events, summarized in Table 7.6, show that using the eight template average method certainly lowers the *CorSNR* values slightly, as compared to the "geometry correct" template matching method, but still returns reasonably high *CorSNR* values. UHECR candidate events which well-matched with their UHECR templates, using the "geometry correct" template matching method, had *CorSNR* values of 8 or above. Therefore to analyze the "burn" sample data using eight template average method, in our search for viable UHECR candidate events, it was required that events have an eight template average *CorSNR* value greater than or equal to 6 so that the probability of missing a possible UHECR candidate was minimized.

It can be concluded that using the eight template average method to analyze the "burn" sample data, for identifying UHECR events, was certainly reasonable. A possible UHECR candidate would still give distinctly high *CorSNR* values, as compared to the surrounding RF events in the same run, using the eight template average method.

7.4 Runs Excluded from ARA-2 10% Unblinded Data Sample Analysis

Runs that were excluded initially from the ARA-2 analysis are given in Table 7.7. A run could be excluded due to several reasons:

1. The run contained some station calibration activity with a surface or an in-ice calibration pulser and hence was full of impulsive events which were of anthropogenic origin. Calibration pulser amplitude sweep runs certainly lie in this category. In these runs, the calpulser amplitude was adjusted by increasing the attenuation of the calpulsers, and calpulser triggers were recorded at the ARA station at each attenuation setting. The calibration pulser amplitude sweep runs helped identify the voltage SNR of an impulsive signal at which ARA stations would have a 50% trigger (or detection) efficiency.

2. The ARA station was operating in a non-standard configuration. An example of this would

be when the ARA station's trigger window size was being adjusted to optimize the window size so that the station could have maximal trigger efficiency for impulsive RF events. The trigger window size is an important feature of the ARA station data acquisition system as the trigger condition of the station requires that at least three antennas of either polarization get triggered (i.e., get a RF signal which crosses the trigger threshold) within the trigger window time.

Table 7.7: List of runs that were excluded from ARA-2 analysis as they either contained some calibration activity with one of the calibration pulsers or the station had some non-standard configuration setting (i.e. varying trigger window size) during those runs as it was being tested.

ARA-2 Excluded Run(s)	Reason for Exclusion
2884-2918, 2938-2939	2014 Surface Pulsing
3120, 3242	2014 ICL Rooftop Pulsing
3139-3162, 3164-3187 3289-3312	2014 Calibration Pulser Amplitude Sweep
3464-3504	2014 L2 Scaler Masking Investigation
3578-3598	2014 Trigger Window Scan
4785, 4787, 4795-4800	2015 IceCube Deep Pulsing
4820-4825, 4850-4854 4879-4936, 5210-5277	2015 Calibration Pulser Noise Mode Tests
4872, 4873, 4876	2015 Surface Pulsing
6513	2015 Calibration Pulser Depth Lift
6527	2015 ICL Rooftop Pulsing
7625-7686	2016 Calibration Pulser Amplitude Sweep

7.5 UHECR Candidate Search Results: ARA-2 10% Unblinded Data Sample

Figures 7.5 and 7.6 show the results of template matching with the eight template average method described in Section 7.3. Based on the conclusions from section 7.3, a Level-0 cut was defined where only events having an eight template average *CorSNR* value of 6 or greater in either polarization were allowed to pass.

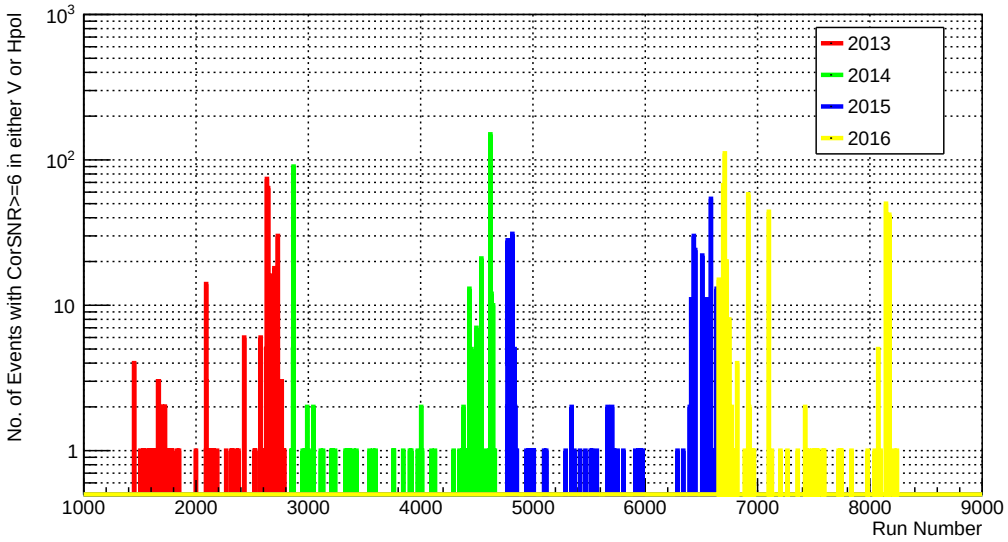


Figure 7.5: Distribution of number of events passing Level-0 cut as a function of run number, over all four years, in the ARA-2 10% unblinded data sample (i.e. "burn" sample data).

Figure 7.5 shows the distribution of events that passed the Level-0 cut as a function of run number, which is being used here as a proxy for time. It can be seen that the density of these events increases at the start or end of each year as that period coincides with the austral summer season, and human activity at the South Pole increases. Throughout the rest of the year, the events are scattered and sparse as expected.

Figure 7.6 shows the run number distribution of *CorSNR* values for all the events found in the ARA-2 "burn" sample data set that passed the Level-0 cut for all four years. There were in total 1542 such events. It should be noted that there are events with *CorSNR* values lower than 6 in

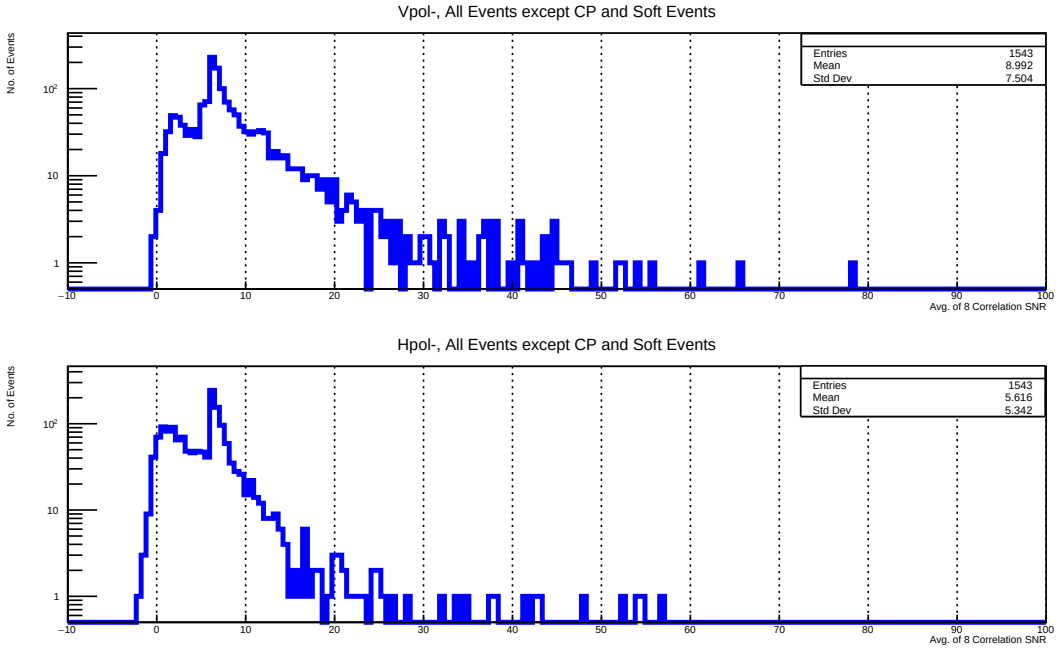


Figure 7.6: Distribution of *CorSNR* values for all of the ARA-2 10% unblinded data sample (i.e. "burn" sample data) events that passed the Level-0 cut found in all four years. The (**top**) Vpol and (**bottom**) Hpol *CorSNR* distributions are shown separately.

either polarization. This indicates that there are ARA events that can well-match with the UHECR templates, i.e., *CorSNR* value greater than or equal to 6, in one polarization but then give low *CorSNR* values, or a poor match, in the other polarization.

7.5.1 Reconstructing Event Directions and Characterising Noise Sources

Using interferometry, the possible source direction for each Level-0 event was reconstructed. The results of reconstruction for Vpol and Hpol channels for all events that passed the Level-0 cut are shown in Figure 7.7.

Some known sources and their azimuth directions are with respect to ARA-2 (in ARA station coordinates) are:

1. South Pole Station: 253.06°
2. IceCube Lab: 265.16°

3. WindTurbine 3: 275.15°

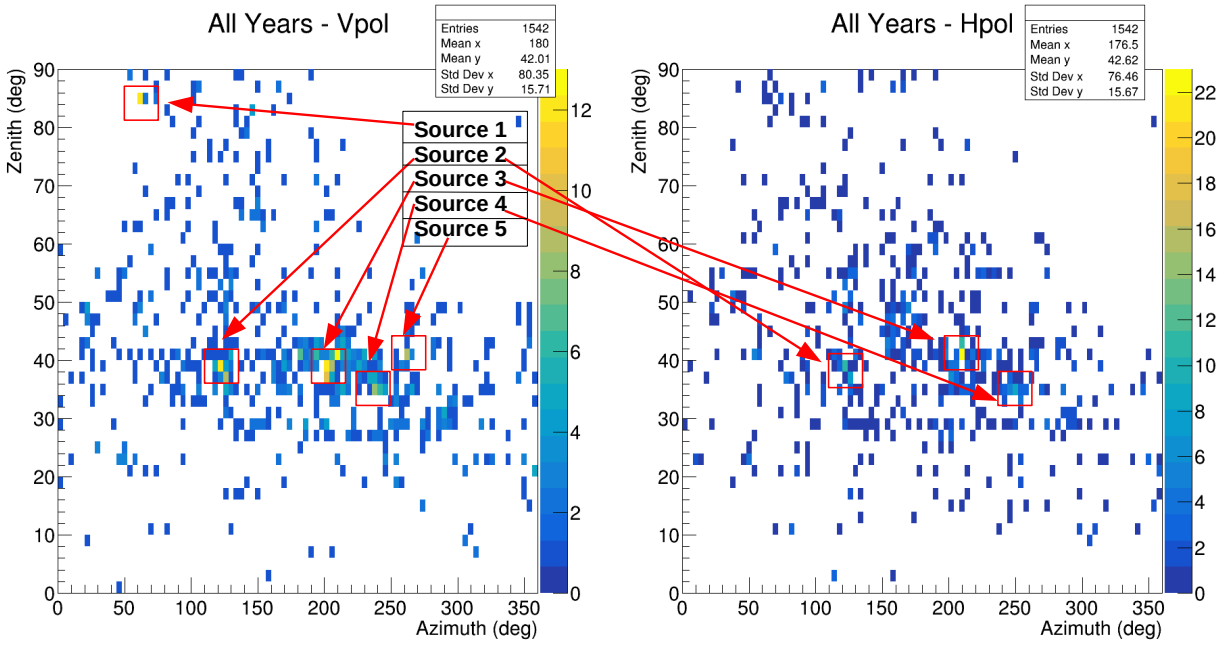


Figure 7.7: Reconstructed zenith and azimuth directions for all the 1542 events remaining after the Level-0 cut was applied on the ARA-2 10% unblinded data sample (i.e. "burn" sample data). Reconstructed directions are shown for (**left**) Vpol channels and (**right**) Hpol channels. Five different noise 'hot spots' or sources can be identified in these two plots.

Background events are defined as events that are not thermal noise and would most probably arise from an on-surface anthropogenic noise source. Events which passed the Level-0 cut were classified as background events if:

1. their reconstructed zenith angles were less than 90° in either Hpol or Vpol channels (i.e., the Vpol or the Hpol signal was coming from above the ARA station).

2. there were other events with a similar topology within ± 10 s of their recorded trigger times.

There were 645 events in total that were thus classified as background events. Runs having more than 10 of these events have been listed in Table 7.8.

Table 7.8: Summary of runs which had more than 10 events that passed the Level-0 cut in the ARA-2 10% unblinded data sample (i.e. "burn" sample data) analysis and were classified as background events. The mean reconstructed zenith and azimuth values are reported below, along with the standard error in the mean values for each polarization.

	Run No.	No. of Events	Reco. Vpol Zenith (°) (Mean ± Std. Err.)	Reco. Vpol Azimuth (°) (Mean ± Std. Err.)	Reco. Hpol Zenith (°) (Mean ± Std. Err.)	Reco. Hpol Azimuth (°) (Mean ± Std. Err.)
1	2090	13	58.527 ± 9.108	196.373 ± 15.506	85.583 ± 12.502	192.121 ± 19.916
2	2631	60	43.682 ± 1.815	128.190 ± 2.890	57.493 ± 4.527	126.863 ± 3.427
3	2642	55	39.090 ± 0.292	206.509 ± 0.491	44.428 ± 2.676	205.368 ± 3.293
4	2677	13	33.262 ± 2.084	144.231 ± 27.933	55.119 ± 9.048	188.441 ± 19.966
5	2698	11	70.543 ± 13.626	251.849 ± 22.892	71.348 ± 11.205	211.721 ± 28.390
6	2868	33	38.465 ± 1.770	163.958 ± 16.952	40.827 ± 3.447	152.623 ± 16.770
7	4541	19	37.499 ± 1.635	170.156 ± 12.579	49.1634 ± 6.473	164.723 ± 14.697
8	4621	77	62.963 ± 3.022	181.205 ± 4.381	52.651 ± 1.686	181.413 ± 3.255
9	4631	11	32.310 ± 3.185	308.220 ± 12.614	80.812 ± 14.336	230.840 ± 27.139
10	4775	27	49.580 ± 1.587	96.203 ± 12.491	52.908 ± 4.214	96.994 ± 12.692
11	4817	26	34.813 ± 1.671	255.772 ± 2.076	76.4398 ± 8.682	168.137 ± 16.020
12	6433	24	44.369 ± 1.428	188.001 ± 4.167	63.789 ± 7.754	174.096 ± 10.742
13	6445	18	23.360 ± 2.806	220.838 ± 28.947	22.322 ± 3.151	245.763 ± 25.910
14	6584	17	55.372 ± 6.933	223.959 ± 15.808	64.1069 ± 8.444	217.555 ± 16.093
15	6916	40	75.778 ± 9.237	148.008 ± 9.997	52.0393 ± 7.043	177.449 ± 9.416
16	8143	17	44.298 ± 6.997	76.052 ± 18.017	57.2707 ± 6.632	57.725 ± 9.113
17	8173	34	36.116 ± 3.926	261.867 ± 7.690	52.1321 ± 7.102	231.066 ± 12.543

Based on the reconstruction results from Table 7.8 and Figure 7.7, we summarize each of the noise sources, or hot spots, as follows:

- Source 1 is consistent with an ARA-2 calpulser. The radio pulses are evident only in Vpol channels. These events are found in run 4645, but they did not make it to the list in Table 7.8 as the events are not within ± 10 s of each other. During normal operation, calpulsers emit one radio pulse per second; however, these events were separated in time by more than a

second. This was probably because only some of them made it to the "burn" sample from the full data set via standard 'pre-selection' procedure. These calpulser events were not rejected because they were untagged, so the standard calpulser event rejection cut failed to flag them.

- Source 2 is an unknown source and is visible in both Vpol and Hpol channels. It is found around 123° in azimuth in Vpol. The events reconstructing to this source are primarily found in run 2631, indicating a transient/episodic background.

- Source 3 is an unknown source and is visible in both Vpol and Hpol channels. It is found around 205° in azimuth in Vpol. The events reconstructing to this source are primarily found in run 2642, indicating a transient/episodic background.

- Source 4 is an unknown source and is visible in both Vpol and Hpol channels. It is found around 238° in azimuth in Vpol. There seems to be a moving source in run 4621.

- Source 5 is around 263° in azimuth and is most probably from ICL. This source is only visible only in Vpol channels. The events reconstructing to this source are primarily found in run 8173, indicating a transient/episodic background.

7.5.2 Cuts to Identify UHECR Candidates

For an event to be identified as a possible UHECR candidate, another set of cuts was defined, which were called the Level-1 cuts. These required that the ARA UHECR candidate event:

1. had to be coming from above the station (i.e., it should have a reconstructed zenith angle of less than 90° in both Hpol and Vpol channels)

2. should have a *CorSNR* greater than or equal to 6 in both Hpol and Vpol

3. should have its *CorSNR* Hpol greater than its *CorSNR* Vpol

4. should have its voltage SNR Hpol (defined below) greater than its voltage SNR Vpol, consistent with the expectation that UHECR events have more HPol than VPol signal power.

Here voltage SNR was simply defined as the peak-to-peak amplitude divided by the noise RMS of the waveform. The channel with the third-highest voltage SNR was chosen as the voltage SNR for that polarization for that ARA event, using the same definition used for studies of the ARA station trigger efficiency.

A UHECR template Monte Carlo was made to verify that UHECR events would pass cuts 3 and 4. It was concluded from the MC that around 90% of the simulated UHECR events do seem to pass cuts 3 and 4, which provided some validation to use these cuts to identify possible UHECR candidate events in ARA data. A more detailed description of the Monte Carlo can be found in Chapter F.

The Level-1 cuts were applied on all 1542 events that passed the Level-0 cut; only two events survived. These were:

1. Run 2868, Event 9482. The time to last Level-0 event was 121 s

2. Run 3206, Event 126972. The time to last Level-0 event was 400868 s.

Run 2868 had 33 events that were classified as background events as shown in Table 7.8. The interferometric reconstructions of the events in run 2868 return azimuth angles, which change in an incremental fashion across the events and similar zenith angles. This indicates that a dynamic source is probably responsible for the impulsive RF events in run 2868. We, therefore, conclude that the event from this run was most probably from an anthropogenic noise source. However, the event from run 3206 passed all Level-1 cuts and was isolated in time from other Level-0 events.

Event 126972 is also one of the UHECR candidate events, listed in Table 7.1, and its eight template average *CorSNR* values for Vpol and Hpol are 9.587 and 16.762, respectively. The eight template average *CorSNR* values reported here are slightly different from the eight template average *CorSNR* values given in Table 7.6. The eight template average *CorSNR* values in Table 7.6 are for full sample (100%) ARA data runs and the values quoted here were calculated from the "burn" sample (10%) data runs. Since the "burn" sample and the full sample have different numbers of RF events, the RMS of the RF event correlation scores changes slightly, which leads

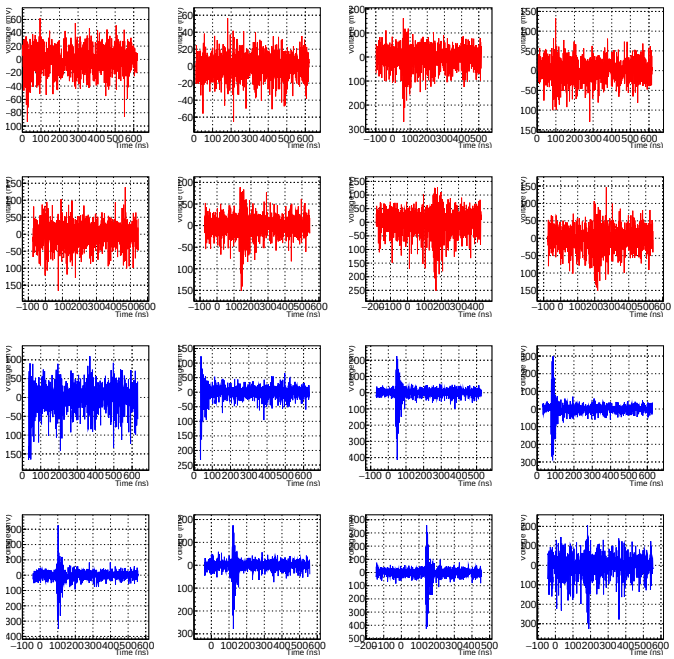


Figure 7.8: Waveform for event 126972 from run 3206. The individual channel axes are Voltage (mV) vs. Time (ns). The top row shows the four Top Vpol antennas in string 1, string 2, string 3 and string 4 respectively, the second row shows the Bottom Vpol channels, the third row shows the Top Horizontal channels and the fourth row shows the Bottom Hpol channels. The channels are numbered from 0 to 15 such that channels 0-3 (from left to right) are in the first row, channel 4-7 are in the second row, 8-11 are in the third row and 12-15 are in the fourth and last row.

to slightly different *CorSNR* values.

Figure 7.8 shows the waveforms for event 126972 from run 3206. Using interferometry the reconstructed zenith and azimuth directions for Vpol antennas were 37.198° and 9° respectively and for Hpol they were 38.744° and 5.192° respectively. The reconstructed angles have been reported in the ARA station coordinate system.

It should also be noted that the "geometry correct" *CorSNR* value is less than the eight template average *CorSNR* value in Hpol for event 12972. This could be due to several reasons:

1. In this study, each simulated UHECR shower template represents a very special case of a shower from that direction where the shower core lands directly between the 16 transect points on the ice surface. The shower could have actually hit a slightly different point, which would cause the core hit point to not be in-between the 16 transect points on the ice. This

would lead to the 16 points sampling a different part of the shower radio footprint than what was actually simulated. This may point to underlying fundamental problems in the way these UHECR templates were generated using CoREAS. It will be shown, for example, in Chapter D of the Appendix, that displacing the shower core by 10 m to 20 m from its original position can result in 5% to 20% change in the *CorSNR* values.

2. It could also be due to the reconstructed zenith and azimuth angles being slightly off from their actual values and, as a result giving a UHECR shower template for a slightly different direction. This may point to accuracy issues with the reconstructed angles being returned by the interferometer. The angular resolution of the interferometer used in this thesis is around 1.8°, limited somewhat by the pixelation in the interferometry. The interferometer has been known to give reconstructed zenith angle values which are systematically off by a degree and can give a large spread in the reconstructed azimuth values if less than eight channels are being used for reconstruction for a given polarization. The accuracy of the interferometer has been discussed in detail in Chapter E of the Appendix.

7.5.3 UHECR Purity Check on Event 126972 from Run 3206

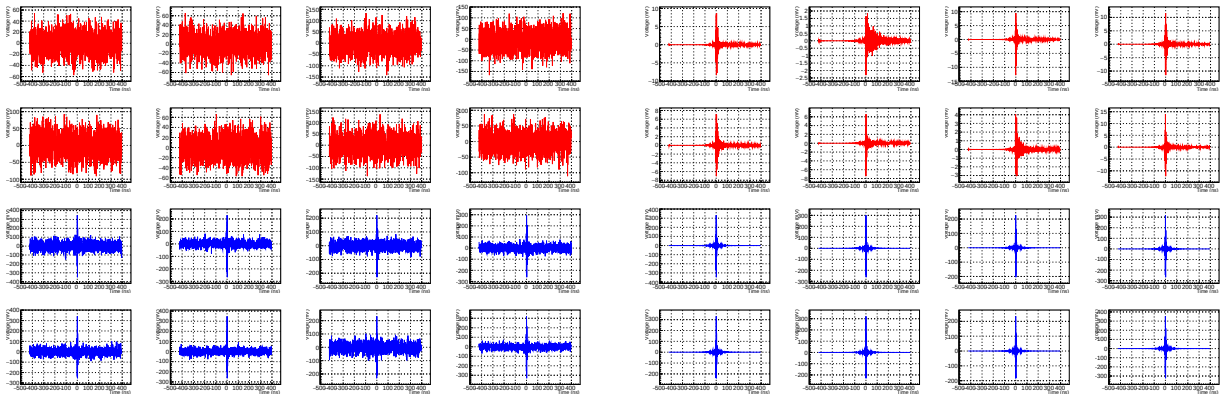


Figure 7.9: Waveforms for the template of event 126972 from run 3206. The individual channel axes are Voltage (mV) vs Time (ns). (**left**) Channel waveforms with Gaussian noise added; (**right**) Channel waveforms with pure signal.

It was known, for event 126972, from the high *CorSNR* values in Table 7.6, that the "geometry correct" template well-matched the event. In order to quantify how closely the relative channel-by-channel signal voltage amplitudes in data matched with their simulation counterparts, the ratio of simulated template channel amplitude to the data channel amplitude was calculated for all channels. To conduct the aforementioned exercise for event 126972, two UHECR templates were made; one with noise and one without noise. Gaussian noise was used to simulate noise, as described in Chapter F of the Appendix. The waveforms for the UHECR template with and without noise can be seen in the right and left plots of Figure 7.9.

Table 7.9: Channel signal voltage amplitude ratio values for UHECR template to data for all the channels for event 126972 from run 3206. There are two columns; one for the UHECR template with Gaussian noise added to the channels and one for without noise.

Ch. no.	Ratio for Temp. without Noise	Ratio for Temp. with Noise
0.	0.143	0.785
1.	0.039	1.092
2.	0.061	0.595
3.	0.102	0.940
4.	0.065	0.744
5.	0.061	0.542
6.	0.019	0.710
7.	0.104	0.594
8.	2.408	2.875
9.	1.529	1.350
10.	0.666	0.634
11.	0.960	0.931
12.	0.922	0.825
13.	1.305	1.316
14.	0.531	0.594
15.	1.400	1.328

Analyzing the channel voltage amplitude ratio values given in Table 7.9, it is noted that when the template has noise, the amplitude ratio values are generally closer to 1 as compared to when there is no noise in the template. This can be explained by the fact that Vpol signals in the waveform of event 126972, shown in Figure 7.8, have relatively low signal amplitudes, with some channels (chs. 2 and 4, e.g.) exhibiting no evident signal in the waveforms. In the UHECR tem-

plates, for event 126972, the Vpol signals have much lower signal amplitudes than the RF noise levels in the channels, and the signals are entirely lost when noise is added to the template channels. Comparing the event waveform in Figure 7.8 with the simulated template waveforms in Figure 7.9 it can be said the Vpol signal is weaker than expected from the UHECR template waveform. This indicates that the template channels' signal amplitude and perhaps the UHECR energy at which the UHECR template was simulated should be higher. Since signal amplitude in the ARA channels scales linearly with UHECR energy, the UHECR template would have given channel amplitude ratios of around 1 had it been simulated at slightly higher energy. It should also be noted that if the radio pulses, in data events, have much lower amplitudes as compared to the RF noise levels in the channel waveforms, then this can certainly adversely affect the quality of our template matches and give us lower *CorSNR* values.

Although event 126972 in run 3206 is a promising UHECR candidate, there is certainly a need to do a deeper investigation in order to find the right UHECR shower template that would perfectly match this ARA data event. The things that could have improved the template match in this context would have been:

1. A higher UHECR energy for the UHECR template.

2. The 'correct' placement of the shower core on the ice surface for the UHECR template. A possible way to do that is suggested in Chapter D of the Appendix.

Chapter 8

Finding UHECRs in ARA-3 10% Unblinded Data Sample

It is important to note that in ARA-3, the whole of string 4 (i.e., channels 3, 7, 11 and 15) was excluded in the analysis of ARA-3 data from 2014 to 2016. String 4 was functioning in a perfectly normal way until the end of 2013 but developed some problems after that and was corrupted by noise for the rest of the years.

8.1 Runs Excluded from ARA-3 10% Unblinded Data Sample Analysis

A list of runs that were initially excluded in the ARA-3 analysis is given in Table 8.1. A run could be excluded due to several reasons, as discussed in Section 7.4. In addition to those reasons, some ARA-3 runs were also excluded if they had an excessively high percentage of software triggers.

Table 8.1: List of runs that were excluded from ARA-3 analysis as they contained some calibration activity with one of the calibration pulsers during those runs. ARA-3 also had some runs that had an excess of software triggers and at times almost purely consisted of them which were also excluded.

ARA-3 Excluded Run(s)	Reason for Exclusion
2235, 2328	2014 ICL Rooftop Pulsing
2251-2274, 2376-2399	2014 Calibration Pulser Amplitude Sweep
3811, 3810, 3820-3822	2015 IceCube Deep Pulsing
3844-3860, 3881-3891, 3916-3918, 3920-3975, 4009-4073	2015 Calpulser Noise Tests
3977, 3978	2015 Surface Pulsing
6041	2015 ICL Rooftop Pulsing
3977	"Noisy Surface" Run
1796-1802, 1804-1807 1809-1812, 1814, 1126, 1129-1140, 1143, 1228, 1231, 1322, 1428, 3421-3424, 3426-3429 3788, 3861, 3892, 3919, 4978, 5014, 5024 7756-7758, 7760-7763 7765-7768, 7710-7712 7125, 7312, 7561, 7570 2000, 2037-2038 2042-2043, 2466-2469 2471-2473	Runs containing an excess of Software triggers. Most of these runs contained 90% or more Software triggers and the rest of them contained around 30% Software triggers

8.2 UHECR Candidate Search Results of ARA-3 10% Unblinded Data Sample Analysis

Figures 8.1 and 8.2 show the results of template matching with the eight template average method described in Section 7.3 for the ARA-3 "burn" sample data of 2013 to 2016.

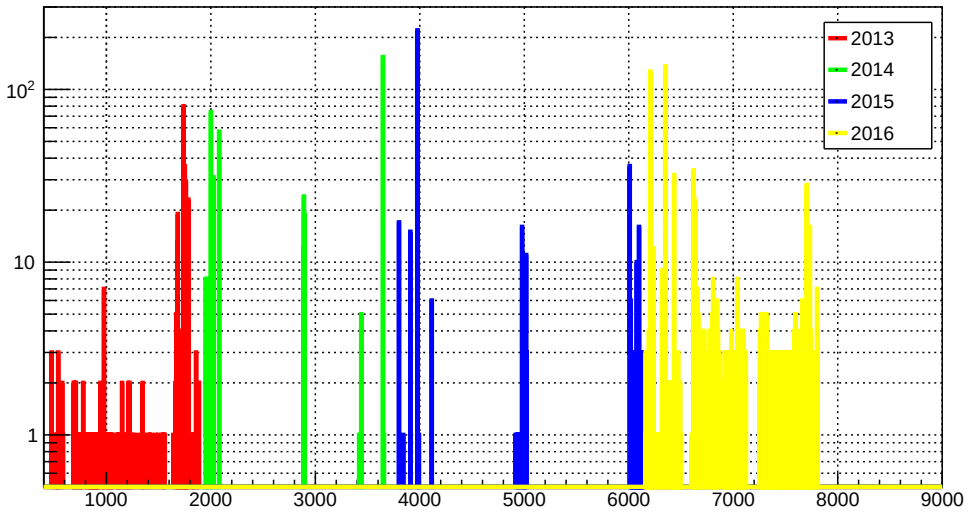


Figure 8.1: Run number distribution for four years of ARA-3 10% unblinded data sample (i.e "burn" sample data) that passed the Level-0 cut.

Figure 8.1 shows the distribution of events which passed the Level-0 cut vs the run number, which is being used here as a proxy for time. The four peaks in event numbers, in Figure 8.2, coincide with a period of time that consists of the austral summer season where human activity at the South Pole increases. 2014 and 2015 are sparsely populated as compared to the other years. Around 50% of the events are from 2016 which seems to be an extremely noisy year.

Figure 8.2 shows the distribution of *CorSNR* values for all the events that passed the Level-0 cut for all four years. There were in total 2764 such events. Similarly to ARA-2 Level-0 event *CorSNR* values, shown in Figure 7.6, it can also be seen for ARA-3 that there are events which have *CorSNR* values much lower than 6, around zero, in either polarization. That is expected as our Level-0 cut only required that the UHECR template should well-match the data event for at

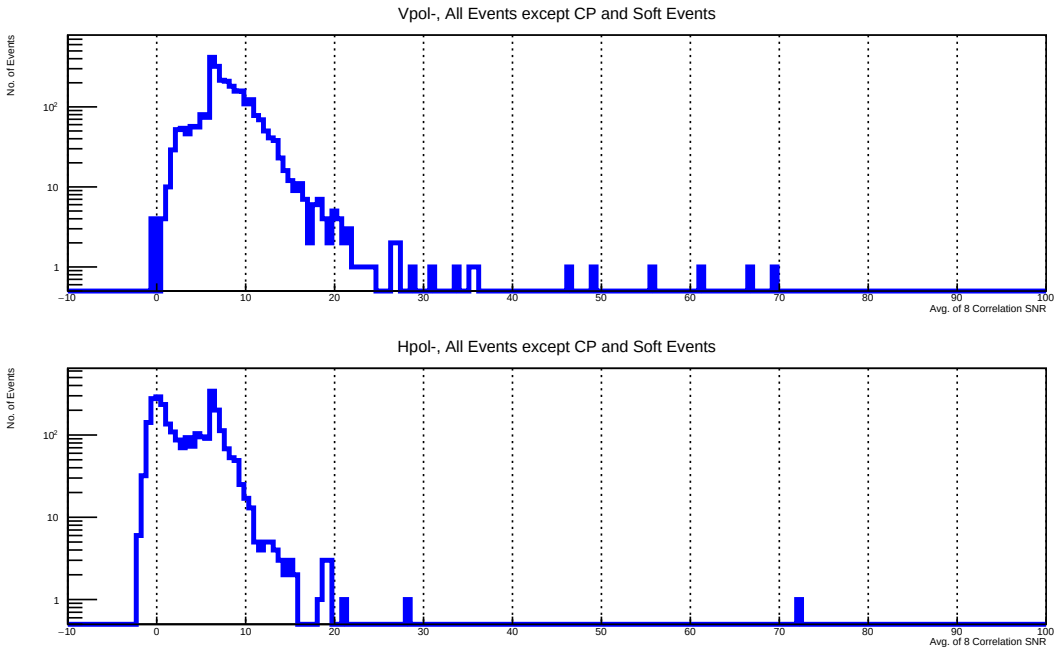


Figure 8.2: Distribution of *CorSNR* values for all of the ARA-3 10% unblinded data sample (i.e., "burn" sample data) events that passed the Level-0 cut found in all the four years. The (**top**) Vpol and (**bottom**) Hpol *CorSNR* distributions are shown separately.

least one polarization (i.e., *CorSNR* value of greater than or equal to 6 in that polarization).

8.2.1 Reconstructing Event Directions and Characterising Noise Sources

Using interferometry, the possible source directions for all the Level-0 ARA-3 events were reconstructed. The results of the reconstruction for the Vpol and Hpol channels for all events that passed the Level-0 cut are shown in Figure 8.3.

The known sources previously enumerated for ARA-2, and their azimuth directions with respect to ARA-3 (in ARA station coordinates) are:

1. South Pole Station: 231.39°
2. IceCube Lab: 237.7°
3. WindTurbine 3: 228.57°

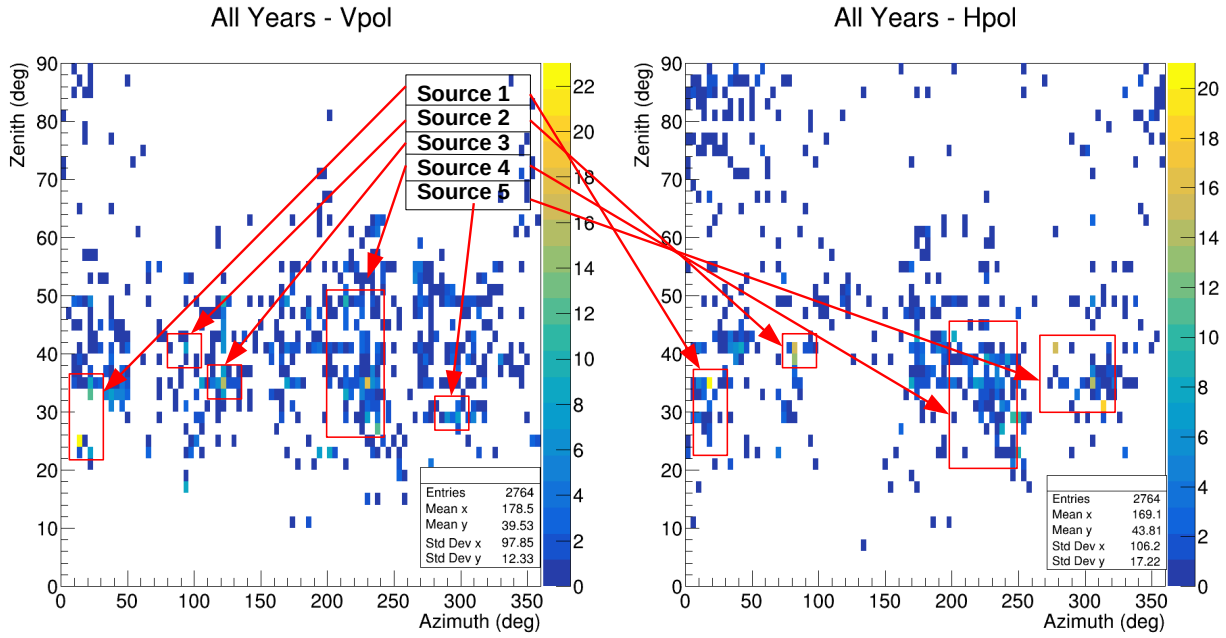


Figure 8.3: Reconstructed zenith and azimuth directions for all the 2764 events were left after the Level-0 cut. Reconstructed directions are shown for (**left**) Vpol channels and (**right**) Hpol channels. Five different noise hot spots, or sources, can also be identified by a visual inspection of these zenith and azimuth angle distributions.

ARA-3 station had 653 events in total got classified as background events. Runs having more than 10 of these events have been listed in Table 8.2.

Table 8.2: Summary of runs which had more than 10 events that passed the Level-0 cut for ARA-3 and were classified as background events. The mean reconstructed zenith and azimuth values have been reported together with the standard error in the mean values for each polarization.

	Run No.	No. of Events	Reco. Vpol Zenith (°) (Mean ± Std. Err.)	Reco. Vpol Azimuth (°) (Mean ± Std. Err.)	Reco. Hpol Zenith (°) (Mean ± Std. Err.)	Reco. Hpol Azimuth (°) (Mean ± Std. Err.)
1	1684	11	44.998 ± 1.785	13.202 ± 1.800	136.978 ± 4.497	97.816 ± 36.657
2	1739	73	39.392 ± 0.680	122.247 ± 1.161	125.373 ± 3.230	125.204 ± 12.153
3	1751	33	42.358 ± 1.279	214.874 ± 0.226	49.825 ± 6.157	207.508 ± 9.695
4	1761	26	40.200 ± 5.501	231.829 ± 3.780	130.579 ± 5.133	133.596 ± 25.291
5	1786	22	45.979 ± 2.448	184.151 ± 8.716	81.333 ± 8.409	140.652 ± 12.849
6	2001	72	43.775 ± 0.884	139.130 ± 14.483	57.097 ± 4.769	148.028 ± 14.785
7	2025	22	49.235 ± 1.523	264.582 ± 4.296	62.234 ± 7.715	261.61 ± 4.537
8	2079	29	43.108 ± 3.555	236.225 ± 17.178	39.281 ± 3.054	241.246 ± 18.198
9	3647	32	46.001 ± 3.123	212.624 ± 11.816	39.548 ± 1.183	215.675 ± 4.694
10	3910	14	41.485 ± 1.902	165.797 ± 6.883	38.325 ± 2.623	172.075 ± 7.031
11	3978	63	30.291 ± 1.885	249.130 ± 10.056	34.721 ± 1.3159	254.878 ± 10.063
12	6009	16	50.740 ± 5.561	259.050 ± 15.348	124.319 ± 6.203	222.882 ± 30.737
13	6349	63	42.582 ± 3.763	45.705 ± 3.045	64.360 ± 5.748	64.161 ± 9.490
14	6350	16	51.243 ± 3.880	275.249 ± 27.925	74.363 ± 10.679	256.412 ± 29.950
15	6433	27	45.020 ± 3.766	127.798 ± 5.693	99.858 ± 7.151	105.25 ± 21.161

We summarize the A3 reconstruction results from Table 8.2 and Figure 8.3 as follows:

- Source 1 is an unknown source and is visible in both Vpol and Hpol channels. It is found around 13° in azimuth in Vpol. The events reconstructing to this source are primarily found in run 1684.

- Source 2 is an unknown source and is visible in both Vpol and Hpol channels. It is found around 95° in azimuth in Vpol. The events reconstructing to this source are found in run 6433.

- Source 3 is an unknown source and is visible in just Vpol channels. It is found around 122° in

azimuth in Vpol. The events reconstructing to this source are found in run 1739. Sources 2 and 3 are so close to each other it could be that they could be the same source.

- Source 4 includes a combination of events from South Pole Station, IceCube Lab and Wind-Turbine 3 and is visible in both Vpol and Hpol channels. These events are spread across multiple runs.

- Source 5 is an unknown source and is visible in both Vpol and Hpol channels. It is found around 95° in azimuth in Vpol. The events reconstructing to this source are present in runs 3978 and 6350.

8.2.2 Level-1 Cuts to Identify UHECR Candidates in ARA-3

For an event to be identified as a possible UHECR candidate the same Level-1 cuts that were applied on ARA-2 were also applied on ARA-3 Level-0 events. The Level-1 cuts were applied on all 2764 events that passed the Level-0 cut, and only six events survived in the ARA-3 "burn" sample data analysis. These were:

1. Run 3647, Event 14023 passed Level 1 cut. The time to last Level-0 event was 23 s.

2. Run 3647, Event 14123 passed Level 1 cut. The time to last Level-0 event was 63 s.

3. Run 3978, Event 6183 passed Level 1 cut. The time to last Level-0 event was 29 s.

4. Run 3978, Event 11943 passed Level 1 cut. The time to last Level-0 event was 70 s.

5. Run 6350, Event 1478 passed Level 1 cut. The time to last Level-0 event was 18 s.

6. Run 6350, Event 1513 passed Level 1 cut. The time to last Level-0 event was 51 s.

Run 3647 had 32 events that were classified as background events as shown in Table 8.2. So the events from this run were most probably from an anthropogenic noise source. The same can be said about events from runs 3978 and 6350, which had 63 and 16 background events, respectively.

Therefore it can be concluded that no viable UHECR candidates were found in the ARA-3 “burn” sample data.

Chapter 9

Expected UHECR flux for an ARA Station

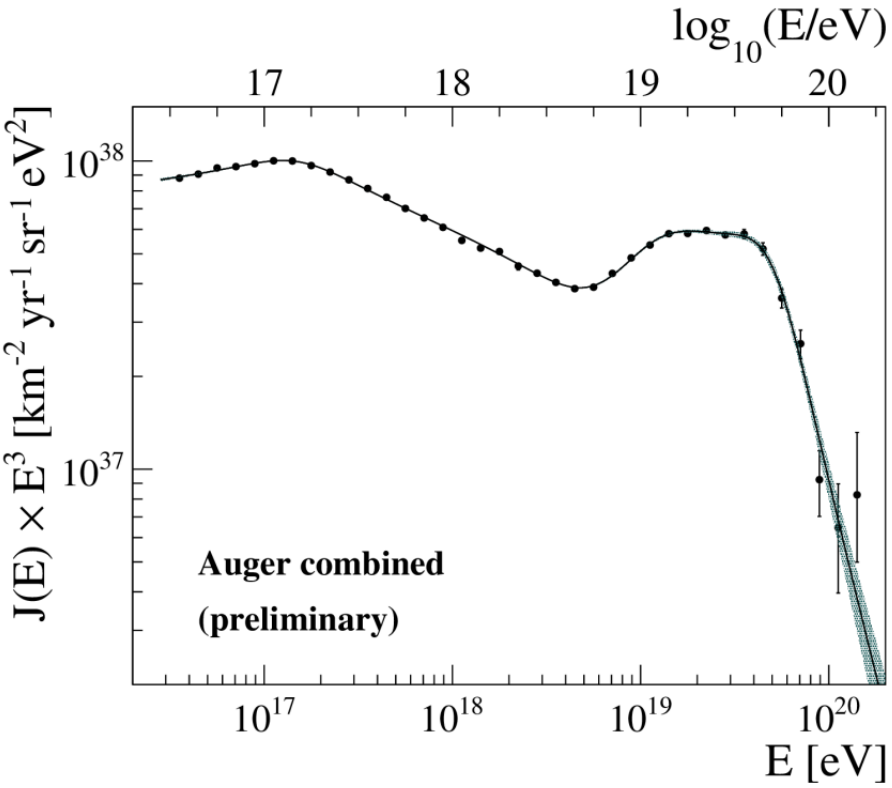


Figure 9.1: PAO UHECR spectrum data, fitted by the function given in Equation 9.1. The spectrum was built by combining data from all the different detectors available at the PAO site. Figure was adapted from (Aab et al., 2019).

In the UHECR searches carried out for ARA-2 and ARA-3 "burn" sample data for the years 2013 to 2016, in Chapters 7 and 8, only one viable UHECR candidate was found in ARA-2; there were none found for ARA-3. These UHECR search results need to be compared to the expected UHECR event rate for four years worth of "burn" sample data in both the stations. To calculate the expected measured UHECR flux for an ARA station, PAO's UHECR flux spectrum had to be

folded in with the ARA UHECR detection efficiency, $\varepsilon(E)$, which is a function of the UHECR energy (Aab et al., 2019). The UHECR flux was then integrated over UHECR energy, effective UHECR detection area for an ARA station on the ice surface and ARA station livetime. The obtained UHECR flux per second was then used to calculate the expected number of UHECR events for ARA-2 and ARA-3 in the "burn" sample data of 2013 to 2016.

PAO's best-fit curve to the UHECR normalised flux spectrum data, as a function of UHECR energy, is shown in Figure 9.1. The reported best-fit function has been defined as (Aab et al., 2019):

$$J(E) = J_0 E^{-\gamma_0} \frac{1+(E/E_{01})^{\gamma_0}}{1+(E/E_{01})^{\gamma_1}} \frac{1+(E/E_{12})^{\gamma_1}}{1+(E/E_{12})^{\gamma_2}} \frac{1+(E/E_{23})^{\gamma_2}}{1+(E/E_{23})^{\gamma_3}} \frac{1+(E/E_{34})^{\gamma_3}}{1+(E/E_{34})^{\gamma_4}} \tag{9.1}$$

The values for the fit parameters of the fit function given in Equation 9.1 are $E_{01} = (0.15 \pm 0.02) \times 10^{18}$ eV, $E_{12} = (6.2 \pm 0.9) \times 10^{18}$ eV, $E_{23} = (12 \pm 2) \times 10^{18}$ eV, $E_{34} = (50 \pm 7) \times 10^{18}$ eV, $\gamma_0 = 2.92 \pm 0.05$, $\gamma_1 = 3.27 \pm 0.05$, $\gamma_2 = 2.2 \pm 0.2$, $\gamma_3 = 3.2 \pm 0.1$ and $\gamma_4 = 5.4 \pm 0.6$, where the errors include the statistical and systematic uncertainties (Aab et al., 2019).

In these expressions, J_0 is the overall normalisation factor. An online plot digitizer was used to extract the value of J_0 from the right plot in Figure 9.1. Using a curve digitizer, a point on the function was extracted: at $E = 9.05738 \times 10^{16}$ eV the function was found to have a value of $E^3 J(E) = 9.7872 \times 10^{37}$ eV^{2}km^{-2}sr^{-1}yr^{-1}. The value of J_0 was then obtained by substituting the E and $E^3 J(E)$ values for the selected point back into the fit function given by Equation 9.1, and found to be 4.17655×10^{36}.

To obtain the expected UHECR flux per second, J_s, for an ARA station, $J(E)$ had to be convolved with the ARA UHECR detection efficiency, $\varepsilon(E)$, and integrated over energy, solid angle and area. Since $\varepsilon(E)$ had not yet been measured or calculated, the Level-1 cut efficiency for UHECR event identification, shown in Figure F.3, was used as a proxy for the ARA UHECR detection efficiency.

Integration over energy was performed from a UHECR energy of 1×10^{16} eV, the energy at which radio signals from UHECR showers would start to overcome the Galactic noise floor,

to 1×10^{20} eV, the energy by which the expected UHECR shower flux has been significantly suppressed due to the GZK effect.

The azimuth angle was integrated from 0° to 360° covering all the azimuth directions. For zenith integration, the critical angle of the ice-air boundary interface had to be calculated. The critical angle puts a limit on the in-ice incidence angle at which an above-surface source can be observed, and can be calculated using Snell's Law to be $\theta_C = \arcsin(1.00/1.35) = 47.8°$. An in-ice incidence angle of 47.7° returns an in-air incidence angle of 86.6° using Snell's Law. This shows how sharply the in-air angle changes for a small change in the in-ice angle. Therefore the zenith angle integration was performed from 0° to 90°.

To calculate the expected ARA UHECR observed flux, the effective UHECR detection area of an ARA station had to be calculated. The horizontal distance at which a UHECR shower with an incoming zenith angle of $\theta = 90°$ would hit the ice surface such that its EM rays hit at least one of the antennas at -200 m depth would be around 158 m. This distance was calculated using analytic raytracing. Therefore, accounting for distance from the center of the station and all the ARA antennas, one can take a circle of 200 m radius centered around the station origin as a conservative estimate.

The triple integral of $J(E)$, given in Equation 9.1, over energy, solid angle and area has been summarised in Equation 9.2:

$$J_s = \frac{2\pi \times \pi(0.200\ \text{km})^2 \times \int_0^{90} \int_{10^{16}\ \text{eV}}^{10^{20}\ \text{eV}} J(E)\varepsilon(E) dE \sin(\theta) d\theta}{365 \times 24 \times 60 \times 60} \tag{9.2}$$

here J_s is the UHECR flux per second for an ARA station, 2π comes from integration over the azimuth angle, $\varepsilon(E)$ is the ARA UHECR detection efficiency and the factors of $365 \times 24 \times 60 \times 60$ are present in order to normalize the flux to units of inverse seconds.

The UHECR flux came out to be $J_s = 7.18 \times 10^{-6}\ \text{s}^{-1}$. The total livetimes for all four years for the ARA stations were taken from the ARA diffuse analysis paper (Allison et al., 2019a). The livetimes for ARA-2 and ARA-3 were 98582400 s and 88300800 s, respectively. Taking $1/10$th of

the total livetime, to account for the "burn" sample, and multiplying that by J_s gives an expected yield of 70.74 UHECR events for ARA-2 and 63.36 UHECR events for ARA-3 for the four years in the "burn" sample data for the two stations.

This is certainly higher than what has been found. We believe one primary reason is that the detection efficiencies have been overestimated as only special UHECR shower geometries are being considered, which have shower cores placed in between the 16 transect points while generating templates. These UHECR geometries lead to an overestimate of our predicted radio signal strength, and hence the ARA detection efficiency because radio emissions from UHECR showers, in the ARA bandwidth, largely happen around the central core region and the Cherenkov ring region. Therefore placing the shower core in between and close to the 16 transect points on the ice biases our UHECR templates by ensuring that we always see a strong radio signal in the ARA antennas.

In the UHECR analysis, the radio footprint on the ice surface is just being sampled with 16 points, with just a single ray from each of those points, which also biases our efficiencies. In reality, EM rays from the entire radio footprint, especially from the core and Cherenkov ring regions, on the ice surface, would contribute to the E-field signal at the ARA antennas. A better way to generate the UHECR templates would be to sample the entire footprint, propagate E-fields from each of those points to each of the 16 in-ice antennas and manually sum the fields at each of the antennas. These issues and possible ways to improve upon them are discussed in Section 10.1.

The ARA antenna gain models and ARA system response model also need to be taken with caution. These models heavily influence the shape and amplitude of the received radio pulses in the ARA antennas and, thus, profoundly affect the trigger efficiencies. A study was done where two ARA-2 calpulser events were used as templates and were matched with UHECR candidate events, listed in Table 7.1. The performance of the calpulser templates was then compared with the "geometry specific" UHECR templates by comparing the *CorSNR* values obtained from the matching of both of these templates with the UHECR candidate events. It was concluded that the calpulser templates performed equally as well as the "geometry specific" UHECR templates in

identifying the UHECR candidate events. This is because of the fact that impulses from UHECR events and calpulser events have a shorter time-scale than the ARA system response such that the system response dominates. Therefore since UHECR templates and calpulser templates have similar impulse time scales, they perform equally well in identifying viable UHECR events in ARA data. This study has been described in detail in Chapter G of the Appendix. This exercise emphasizes the importance of having an accurate ARA system response model for generating UHECR templates. It is evident from the results of the study how vital the ARA system response is in affecting the pulse shape, and therefore the ARA UHECR trigger efficiency.

To get an expected flux of one and zero events in ARA-2 and ARA-3 "burn" sample data, the Level-1 efficiency curve, shown in Figure F.3, needs to be shifted to the right by about $+0.80$ and $+1.80$ energy exponent units respectively. Although using the Level-1 cut efficiency for UHECR event identification as a proxy for ARA UHECR detection efficiency is inaccurate, this exercise gives an estimate of the ARA UHECR detection efficiency as a function of energy. Also, it provides insight into the influence of the ARA UHECR detection efficiency on the ARA expected detected UHECR flux.

Chapter 10

ARA UHECR Search: Potential Improvements And Conclusion

10.1 Potential Improvements

The UHECR analysis that has been done in this thesis is certainly a first iteration and had some limitations. One of the biggest issues with the current UHECR template generation method was that it only used 16 transect points on the ice surface to sample the entire radio footprint. This would certainly lead to a bias in the UHECR template sample as we were only looking for radio emissions from a specific part of the UHECR shower radio footprint. In reality, the whole radio footprint would contribute to the received E-fields at the in ice antennas.

It was also assumed that most of the EM emission is collinear with the shower propagation direction. However, for this assumption to hold, the shower maximum (X_{max}) has to be significantly far away from the ice surface, at least around a kilometer in distance. For protonic UHECR showers at 1×18 eV, this assumption starts to break down as the distance between X_{max} and the ice surface significantly decreases (approaching O(100 m) or less), especially for showers having near-vertical zenith angles.

Possible areas of improvement for the current analysis framework and all future iterations of this analysis are:

1. Only the ARA-2 and ARA-3 "burn" sample data for 2013 to 2016 was analyzed. The full sample data for ARA-2 and ARA-3 for 2013 to 2016 still needs to be analyzed and will certainly contain more UHECR candidates.

2. This UHECR analysis needs to be extended to include more ARA stations (ARA-1, ARA-4

and ARA-5) and data beyond 2016.

3. The interferometer's angular resolution needs to be improved by increasing the number of pixels in the interferometric sky map.

4. The ARA antenna gain and system response models need to be thoroughly vetted and their accuracy verified. It was clearly shown in Chapter G of the Appendix, how heavily the ARA system response model influenced the radio pulse shapes in the ARA UHECR event templates and, therefore, the overall expected UHECR flux for the ARA detectors.

5. While calculating the final received E-fields at the ARA antennas in this UHECR analysis, the distance traveled by the E-fields between the space points on the ice surface and the ARA antennas was not taken into account. For any distance traveled by the E-fields, there should certainly be some loss in signal strength. To account for this loss, a new unit-less quantity called the fractional distance was introduced. Fractional distance (d_{FD}) was defined as the total optical path length from the CR shower maximum (X_{max}) to a surface transect point, divided by the total optical path length from X_{max} to an in-ice ARA antenna. d_{FD} can be written as:

$$(|\vec{r}_s - \vec{r}_{SM}|_{optical})/(|\vec{r}_s - \vec{r}_{SM}|_{optical} + |\vec{r}_{Rx} - \vec{r}_s|_{optical}) \tag{10.1}$$

   here $\vec{r}_s$ is the position vector for a given space point on the ice surface, $\vec{r}_{Rx}$ is the position vector for an in-ice receiver (i.e. ARA antenna), $\vec{r}_{SM}$ is the position vector for X_{max} and the $optical$ subscript represents the fact that we are only taking into account the optical path lengths.

6. The UHECR template generation method using CoREAS can be developed further:

   (a) The ideal scenario would be to have CoREAS raytrace all the radio emissions into the ice towards the ARA antennas and sum up all the E-field combinations directly at the in-ice ARA antennas. This is something that is currently being worked upon.

(b) The procedure described in this thesis is based on using a single ray to derive the estimated electric field strength at a given ARA antenna. In reality, the signal at a given in-ice ARA antenna is collected from all illuminated source points on the ice surface. As a short-term alternative, until the aforementioned approach is implemented, we propose a new procedure which numerically calculates the net received E-field from a UHECR shower for a given in-ice antenna (or receiver). The new procedure can be summarized in the following steps:

i. Instead of having 16 space points on the ice surface for which CoREAS returns an E-field vector, there would be 160 space points on the ice surface covering the whole radio shower footprint. The 160 points would be in a star shape configuration, with eight radial arms having equally-spaced space points. For a perfectly vertical UHECR shower, the star would have a circular shape and a radius of 250 m. For an inclined UHECR shower, the star would be stretched in the required direction (giving it an elliptical shape) to cover as much of the radio footprint as possible.

ii. The electric field at each space point i at location $\vec{r}_{s,i}$ (having a magnitude $E_{s,i}$) would now be propagated into the ice to the receiver at $\vec{r}_{Rx}$. Assuming that the original source was at the UHECR shower maximum, with a coordinate $\vec{r}_{SM}$, the electric field at the location of the receiver, for a given space point, would then have a magnitude proportional to:

$$E_{s,i} \times \mathbf{P} \times d_{FD} \tag{10.2}$$

here $\mathbf{P}$ is a polarization factor given by the dot product of the propagating radiation electric field with the effective height of the receiver antenna at that azimuthal and elevation viewing angle and d_{FD} is the fractional distance given by $(|\vec{r}_{s,i} - \vec{r}_{SM}|_{optical})/(|\vec{r}_{s,i} - \vec{r}_{SM}|_{optical} + |\vec{r}_{Rx} - \vec{r}_{s,i}|_{optical})$. The signal for E-field given by E_s

would arrive at the receiver at a position-dependent time t_{Rx} and with a frequency-dependent phase ϕ_i.

iii. The total CoREAS-modeled radio-frequency power, in the ARA bandwidth, on the surface (P_0(surface)$\sim \Sigma_i(E_{s,i}^2)$) would then be calculated; here the sum would loop over all the 160 space points which would be in a star configuration on the ice surface. This value would be retained as a normalization factor for the final result, which would be the calculated E-fields at the in-ice antennas.

iv. The contributions for all the simulated surface space points would then be summed in a given retarded time bin. The E-field signal with the shortest path lengths to a given receiver would be the fastest to arrive. For surface space points having the longest path lengths, the E-field signal would arrive the latest. In performing the sum over all the space positions, the relative phases of all E-field signals arriving at a given time would be properly summed, such that there would be some mitigation of the final E-field signal through destructive interference. Each contribution from each surface point i would also require an additional weighting factor of $E_{s,i}^2/P_0$. This is to ensure that the final result (i.e., calculated E-field at the in-ice antenna) is insensitive to the number of surface points chosen for the sum and the highest signal-strength surface E-field vectors properly receive the highest weighting.

7. If a future UHECR analysis uses the same UHECR template generation method as used in this thesis, considerable attention and care need to be given to:

   (a) Taking into account the actual X_{max} position instead of assuming it be significantly far away from the ice surface. CoREAS also returns the atmospheric depth of X_{max} in its output files, which can be used to reconstruct the X_{max} coordinates. The 16 transect points on the ice would then be recalculated and rays would be then be traced directly from X_{max} to the ARA antennas. This would completely remove the need to make the assumption that the shower emission is collinear with shower propagation direction. It

should be noted that this method requires the recalculation of the E-fields at the new 16 transect points. This can be achieved through two-dimensional interpolation, where the E-field emissions across the whole shower footprint on ice are sampled with multiple space (or observer) points. Then interpolation is used to reconstruct the E-fields at the new transect points.

(b) The placement of UHECR shower core on the ice surface for the CoREAS simulation. It was shown in Chapter D that if the shower core position was adjusted by a small amount on the ice surface, it affected the UHECR templates and, therefore, their match scores with the respective ARA event by a significant amount. A method to find the 'correct' position of the shower core for a given UHECR template for an ARA event was also suggested in Chapter D.

The suggestions given above to improve this UHECR template generation method further would help curtail limitations of the current template generation method and bring the UHECR templates for ARA antennas closer to reality.

10.2 Conclusion

Chapters 1 to 6 provided background information on the ARA detector, UHECR showers and how they might be observed in an ARA station. This served as the foundation for setting up the CoREAS simulation to simulate the radio emissions from UHECR showers and generate UHECR event templates for the ARA detectors.

Chapter 7 discussed in detail the "burn" sample analysis of ARA-2 2013 to 2017. All the cuts were outlined and the noise sources were characterized. This led to the identification of a single UHECR candidate event in ARA-2 "burn" sample. A purity check was done on that event and it appears to be a promising candidate. It can be inferred that if one viable UHECR candidate was found in the "burn" sample data for those four years, then ten should be found in the full ARA-2 data set, for those four years, if the same analysis was repeated. As noted in Chapter 8, the same

UHECR candidate event identification process was repeated for ARA-3 "burn" sample data for 2013 to 2016 but no viable UHECR candidates were found in the analysis.

Chapter 9 centered around the discussion of the possible UHECR flux that an ARA station might observe. To achieve that, the Level-1 cut efficiency was used as a proxy for ARA UHECR detection efficiency. This was certainly an approximation and was employed to get an idea of the ARA UHECR detection efficiency as a function of UHECR energy, and therefore the expected rate. The Level-1 cut efficiencies were likely overestimated because they were only made using special cases of the most optimistic UHECR shower geometries that were simulated through CoREAS for the MC. In these special UHECR shower geometries, the shower core was made to hit a particular position on the ice surface between the 16 transect points, that were found for each of the ARA antennas. The process of generating templates could also be improved by sampling a wider area of the shower footprint by using more than 16 transect points. It was determined that, with the current template generation method, to get the observed expected flux of one and zero events in ARA-2 and ARA-3 in their "burn" sample data, the Level-1 cut efficiency curve had to shift higher by around $+0.80$ and $+1.80$ energy exponent units respectively.

Overall, the template matching process proved to be effective in identifying viable UHECR candidates in the ARA data. As UHECR event impulses have a shorter time-scale than the ARA system response, the system response dominates. Due to this, any UHECR template would perform reasonably well in identifying viable UHECR events in data. Although, to get a deeper understanding of the event geometry and energy, a more detailed and exhaustive simulation setup will be required.

References

Aab, A. et al. (2019). The Pierre Auger Observatory: Contributions to the 36th International Cosmic Ray Conference (ICRC 2019).

Aartsen, M. et al. (2018). Multimessenger observations of a flaring blazar coincident with high-energy neutrino IceCube-170922A. *Science*, 361(6398), eaat1378.

Ackermann, M. et al. (2013). Detection of the Characteristic Pion-Decay Signature in Supernova Remnants. *Science*, 339, 807.

Ahlers, M., Anchordoqui, L. A., Gonzalez-Garcia, M. C., Halzen, F., & Sarkar, S. (2010). GZK Neutrinos after the Fermi-LAT Diffuse Photon Flux Measurement. *Astropart. Phys.*, 34, 106–115.

Allan, H. R., Clay, R. W., & Jones, J. K. (1970). Radio Pulses from Extensive Air Showers. *Nature*, 227, 1116–1118.

Allison, P. et al. (2012). Design and Initial Performance of the Askaryan Radio Array Prototype EeV Neutrino Detector at the South Pole. *Astropart. Phys.*, 35, 457–477.

Allison, P. et al. (2016). Performance of two Askaryan Radio Array stations and first results in the search for ultrahigh energy neutrinos. *Phys. Rev.*, D93(8), 082003.

Allison, P. et al. (2019a). Constraints on the Diffuse Flux of Ultra-High Energy Neutrinos from Four Years of Askaryan Radio Array Data in Two Stations.

Allison, P. et al. (2019b). Long-baseline horizontal radio-frequency transmission through polar ice.

Aloisio, R. (2018). The Physics of UHECRs: Spectra, Composition and the Transition Galactic-Extragalactic. *JPS Conf. Proc.*, 19, 011008.

Andersen, K. K. & Klein, S. R. (2011). High energy cosmic-ray interactions with particles from the Sun. *Phys. Rev.*, D83, 103519.

Barwick, S. et al. (2018). Observation of classically 'forbidden' electromagnetic wave propagation and implications for neutrino detection. *JCAP*, 07, 055.

Brun, R. & Rademakers, F. (1996). Root - an object oriented data analysis framework. In *AIHENP'96 Workshop, Lausane*, volume 389 (pp. 81–86).

Castellina, A. (2020). Highlights from the Pierre Auger Observatory (ICRC2019). *PoS*, ICRC2019, 004.

Clark, B. (2019). *Optimization of a Search for Ultra-High Energy Neutrinos in Four Years of Data of ARA Station 2*. PhD thesis, Ohio State University.

Cooper, N. (2012). The Invisible Neutron Threat — National Security Science. [Online; accessed 16-July-2020].

Cucinotta, F. & Cacao, E. (2017). Non-Targeted Effects Models Predict Significantly Higher Mars Mission Cancer Risk than Targeted Effects Models. *Scientific Reports*, 7, 1832.

di Matteo, A. et al. (2020). Full-sky searches for anisotropies in UHECR arrival directions with the Pierre Auger Observatory and the Telescope Array. *PoS*, ICRC2019, 439.

Donini, A., Palomares-Ruiz, S., & Salvado, J. (2019). Neutrino tomography of Earth. *Nature Phys.*, 15(1), 37–40.

Glaser, J. C. (2017). *Absolute energy calibration of the Pierre Auger observatory using radio emission of extensive air showers*. Dissertation, RWTH Aachen University, Aachen. Veröffentlicht auf dem Publikationsserver der RWTH Aachen University; Dissertation, RWTH Aachen University, 2017.

Gonzalez, A. (2014). Mutagenesis and Background Neutron Radiation. *arXiv e-prints*, (pp. arXiv:1406.6641).

Gorski, K., Hivon, E., Banday, A., Wandelt, B., Hansen, F., Reinecke, M., & Bartelman, M. (2005). HEALPix - A Framework for high resolution discretization, and fast analysis of data distributed on the sphere. *Astrophys. J.*, 622, 759–771.

Greisen, K. (1966). End to the cosmic-ray spectrum? *Phys. Rev. Lett.*, 16, 748–750.

Hanlon, W. (2020). Telescope Array 10 Year Composition. *PoS*, ICRC2019, 280.

Hanson, J. (2013). *The Performance and Initial Results of the ARIANNA Prototype*. PhD thesis, University of California, Irvine.

Heck, D., Knapp, J., Capdevielle, J. N., Schatz, G., & Thouw, T. (1998). *CORSIKA: a Monte Carlo code to simulate extensive air showers.*

Hillas, A. M. (1984). The Origin of Ultrahigh-Energy Cosmic Rays. *Ann. Rev. Astron. Astrophys.*, 22, 425–444.

Hillas, A. M. (2005). Can diffusive shock acceleration in supernova remnants account for high-energy galactic cosmic rays? *Journal of Physics G: Nuclear and Particle Physics*, 31(5), R95.

Huege, T. (2016). Radio detection of cosmic ray air showers in the digital era. *Phys. Rept.*, 620, 1–52.

Huege, T., Ludwig, M., & James, C. W. (2013). Simulating radio emission from air showers with CoREAS. *AIP Conf. Proc.*, 1535(1), 128.

James, C. W., Falcke, H., Huege, T., & Ludwig, M. (2011). General description of electromagnetic radiation processes based on instantaneous charge acceleration in “endpoints”. *Phys. Rev. E*, 84, 056602.

Karney, C. F. F. (2019). Geographiclib, version 1.49 (2017-10-05). `https://geographiclib.sourceforge.io/1.49`. Online: accessed 21-March-2019.

Katz, J. (2014). Scanning of Vehicles for Nuclear Materials. *AIP Conf. Proc.*, 1596, 225.

Kelley, J., Lu, M.-Y., Seckel, D., Pan, Y., Besson, D. Z., & Collaboration, F. T. A. (2018). Observation of two deep, distant (1.4, 4)km impulsive RF transmitters by the Askaryan Radio Array (ARA). *PoS*, ICRC2017, 1030.

Koci & Kuivinen (1983). A 237-meter ice core from South Pole Station. *Antarctic Journal of the U.S*, 18, 113–114.

Kostunin, D., Bezyazeekov, P. A., Hiller, R., Schröder, F. G., Lenok, V., & Levinson, E. (2016). Reconstruction of air-shower parameters for large-scale radio detectors using the lateral distribution. *Astropart. Phys.*, 74, 79–86.

Kravchenko, I., Besson, D., & Meyers, J. (2004). In situ index-of-refraction measurements of the South Polar firn with the RICE detector. *Journal of Glaciology*, 50(171), 522–532.

Lahmann, R. (2019). Investigations of ice and emitter properties from radio signals recorded with ARIANNA. In *36th International Cosmic Ray Conference (ICRC2019)*, volume 36 of *International Cosmic Ray Conference* (pp. 939).

Melott, A. (2016). Stellar astrophysics: Supernovae in the neighbourhood. *Nature*, 532, 40–41.

Melott, A. L. & Thomas, B. C. (2019). From cosmic explosions to terrestrial fires? *J. Geology*, 127, 475–481.

Morishima, K. et al. (2017). Discovery of a big void in Khufu's Pyramid by observation of cosmic-ray muons. *Nature*, 552(7685), 386–390.

Mucke, A., Engel, R., Rachen, J. P., Protheroe, R. J., & Stanev, T. (2000). SOPHIA: Monte Carlo simulations of photohadronic processes in astrophysics. *Comput. Phys. Commun.*, 124, 290–314.

Nelles, A. (2014). *Radio emission of air showers: The perspective of LOFAR and AERA*. PhD thesis, Radboud University Nijmegen.

Nelles, A., Buitink, S., Falcke, H., Hörandel, J., Huege, T., & Schellart, P. (2015). A parameterization for the radio emission of air showers as predicted by CoREAS simulations and applied to LOFAR measurements. *Astropart. Phys.*, 60, 13–24.

NOAA (2018). Magnetic field calculator using igrf (1590-2019). `https://www.ngdc.noaa.gov/geomag-web/`. Online: accessed 21-October-2018.

Ogio, S. (2019). Highlights from the Telescope Array Experiment. *PoS*, ICRC2019, 031.

Schröder, F. G. (2017). Radio detection of Cosmic-Ray Air Showers and High-Energy Neutrinos. *Prog. Part. Nucl. Phys.*, 93, 1–68.

Scoles, S. (2018). Cosmic Ray Showers Crash Supercomputers. Here's What to Do About It — WIRED. [Online; accessed 16-July-2020].

Tanabashi, M. et al. (2018). Review of Particle Physics. *Phys. Rev.*, D98(3), 030001.

Unger, M. & Farrar, G. R. (2017). Uncertainties in the Magnetic Field of the Milky Way.

Yu, B. G., Choi, T. K., & Kim, W. (2011). Regge phenomenology of pion photoproduction off the nucleon at forward angles. *Phys. Rev.*, C83, 025208.

Zatsepin, G. T. & Kuzmin, V. A. (1966). Upper limit of the spectrum of cosmic rays. *JETP Lett.*, 4, 78–80. [Pisma Zh. Eksp. Teor. Fiz.4,114(1966)].

Appendix A

Waveforms for Events Adjacent to ARA-2 UHECR Candidate Events

These are the waveforms for events that were found to be "adjacent" (i.e., in the same data run) to ARA-2 UHECR Candidate Events found in OSU neutrino analysis. These adjacent events had distinctively high Vpol *CorSNR* values (i.e., *CorSNR* values of above 6), which implied that they well-matched the "geometry specific" template of the given UHECR candidate event and were also impulsive in nature. The UHECR candidate and their adjacent events were referred to in Section 7.1. In the figures, the top row shows the four Top Vpol antennas in string 1 , string 2 , string 3 and string 4 respectively, the second row shows the Bottom Vpol channels, the third row shows the Top Horizontal channels and the fourth row shows the Bottom Hpol channels. The channels are numbered from 0 to 15 such that channels 0-3 (from left to right) are in the first row, channel 4-7 are in the second row, 8-11 are in the third row and 12-15 are in the fourth and last row.

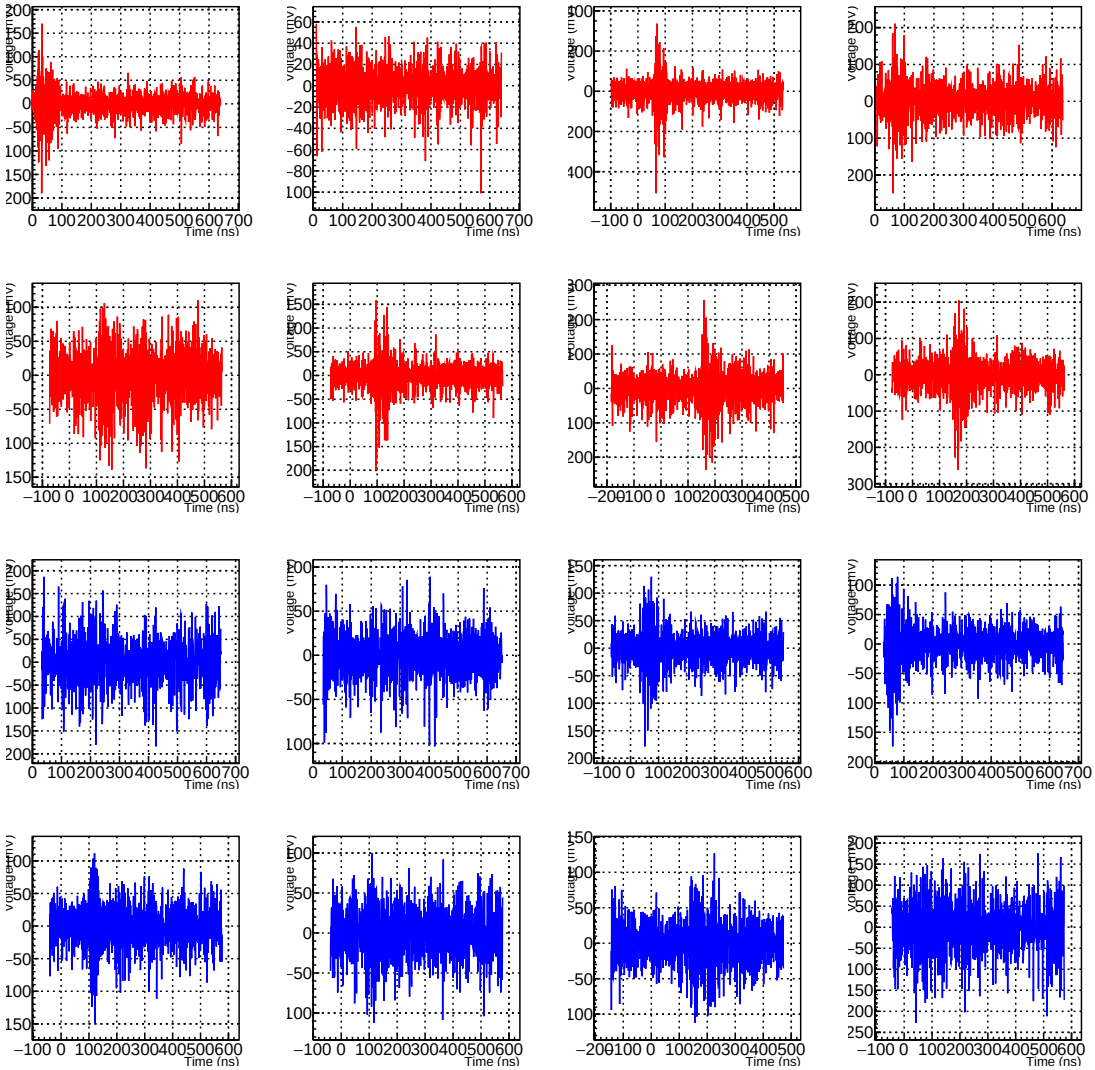


Figure A.1: Waveform for event 29309 from run 2869. The individual channel axes are Voltage (mV) vs Time (ns).

.

Figure A.1 shows the waveform for event 29309 from run 2869 which is adjacent to UHECR candidate event 29331. The pulses in the channels have some double pulses features, which can be noticed if the event waveform is given a closer visual inspection. Using interferometry the reconstructed zenith and azimuth directions for Vpol antennas were 32.6005° and 308.864° respectively and for Hpol they were 28.0562° and 305.526° respectively. These angles have been reported in the ARA station coordinate system.

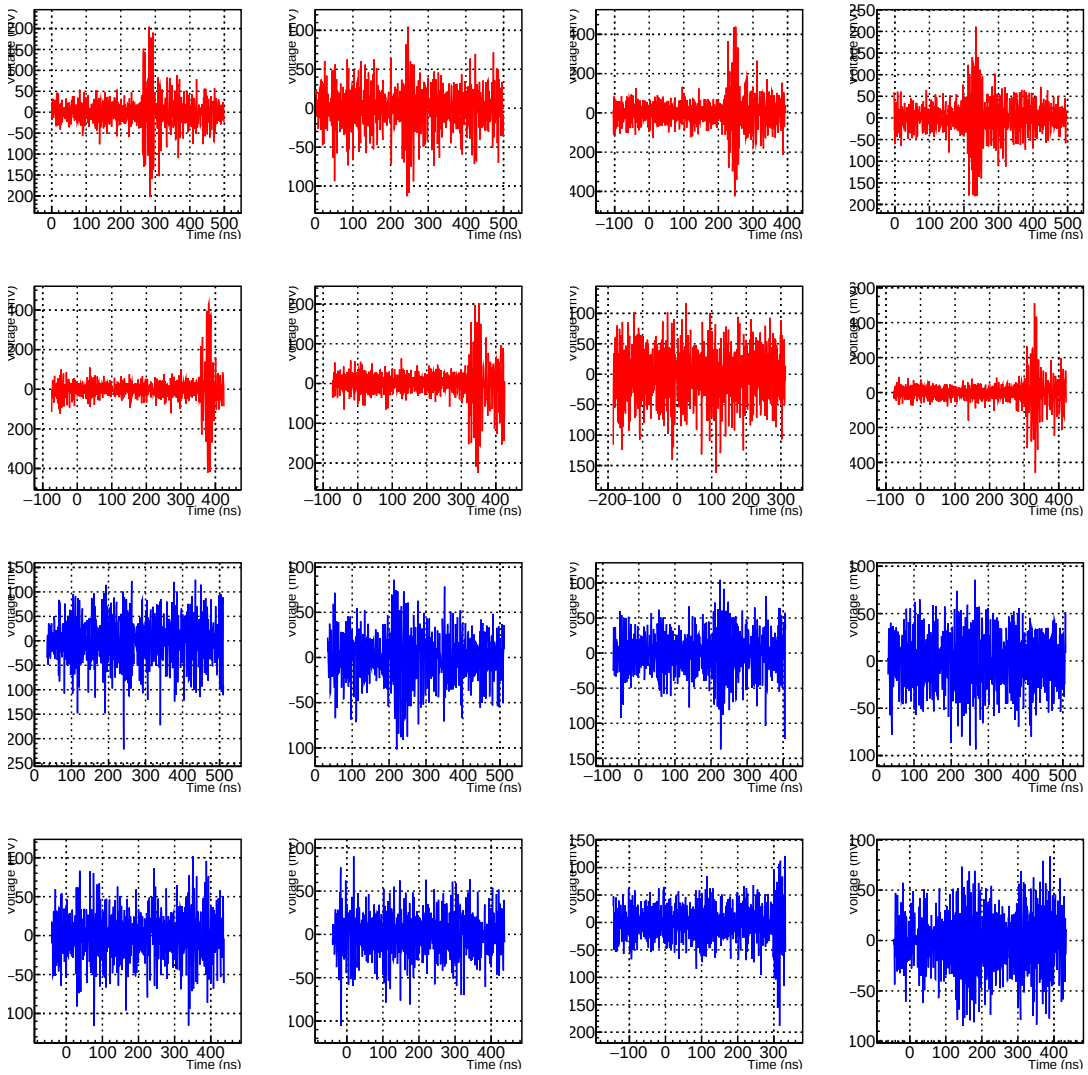


Figure A.2: Waveform for event 28579 from run 5009. The individual channel axes are Voltage (mV) vs Time (ns).

.

Figure A.2 shows the waveform for event 28579 from run 5009 which is adjacent to UHECR candidate event 28578. This event coincidentally comes right after the UHECR candidate event 28578, which makes it likely that both of these events could be coming from an anthropogenic source. Using interferometry the reconstructed zenith and azimuth directions for Vpol antennas were 40.2976° and 188.333° respectively and for Hpol they were 37.1981° and 203.4° respectively. These angles have been reported in the ARA station coordinate system.

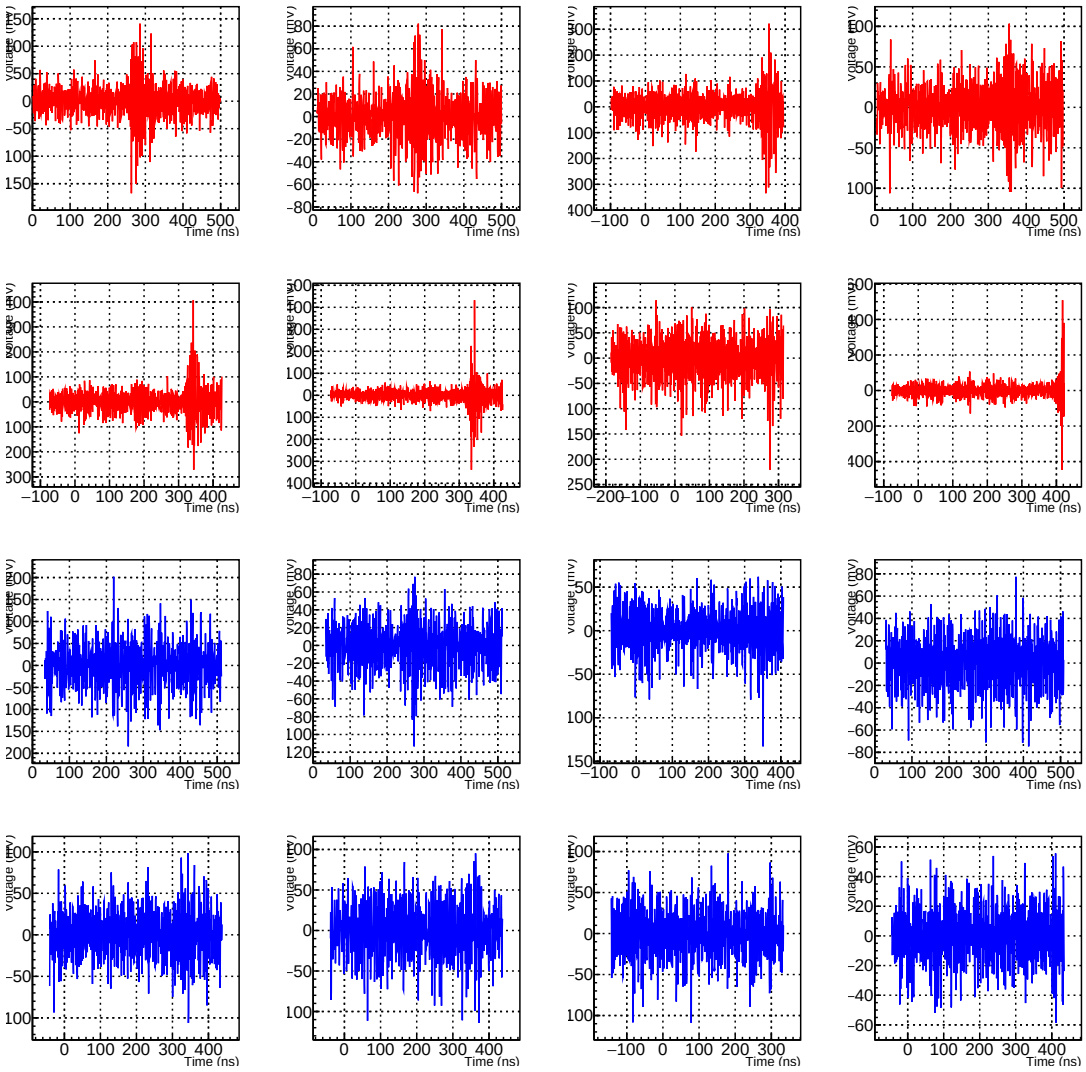


Figure A.3: Waveform for event 43391 from run 6861. The individual channel axes are Voltage (mV) vs Time (ns).

.

Figure A.3 shows the waveform for event 43391 from run 6861 which is adjacent to UHECR candidate event 57376. This event has recorded more power in its Vpol channels as compared to Hpol channels. Using interferometry the reconstructed zenith and azimuth directions for Vpol antennas were 58.6118° and 337.5° respectively and for Hpol they were 148.92° and 45° respectively. These angles have been reported in the ARA station coordinate system.

Appendix B

Shape of the RF event *CorSNR* distributions

The high *CorSNR* tails of the RF event *CorSNR* distributions were fitted with exponential functions in Section 7.2. To determine the overall shape, the RF event *CorSNR* distributions were fitted with Gaussian and Poisson distribution functions.

The top and bottom plots of Figure B.1a show the results of Gaussian fits to RF event *CorSNR* distributions (blue distributions). The events in these Figures came from runs corresponding to UHECR candidate events that passed the V&Hpol cuts of the OSU analysis. The Gaussian function in ROOT that was used for the fits was given by:

$$f(x) = \text{Par}[0] \times \texttt{TMath::Gaus}(x, \text{Par}[1], \text{Par}[2]) + \text{Par}[3] \tag{B.1}$$

where x was the *CorSNR* value, $f(x)$ was the number of events, `TMath::Gaus` is the pre-defined Gaussian function in ROOT and the values of Par[0], Par[1], Par[2] and Par[3] were determined by fitting. The fit range was set from a *CorSNR* value of -10 to 10 as that covered the whole range of the RF event distributions. The χ^2/NDF values obtained from this fit function for all the RF event *CorSNR* distributions have been listed down in Table B.1.

Table B.1: χ^2/NDF values for Gaussian distribution fits to RF event *CorSNR* distributions. The fit range covered $-10 < CorSNR < 10$.

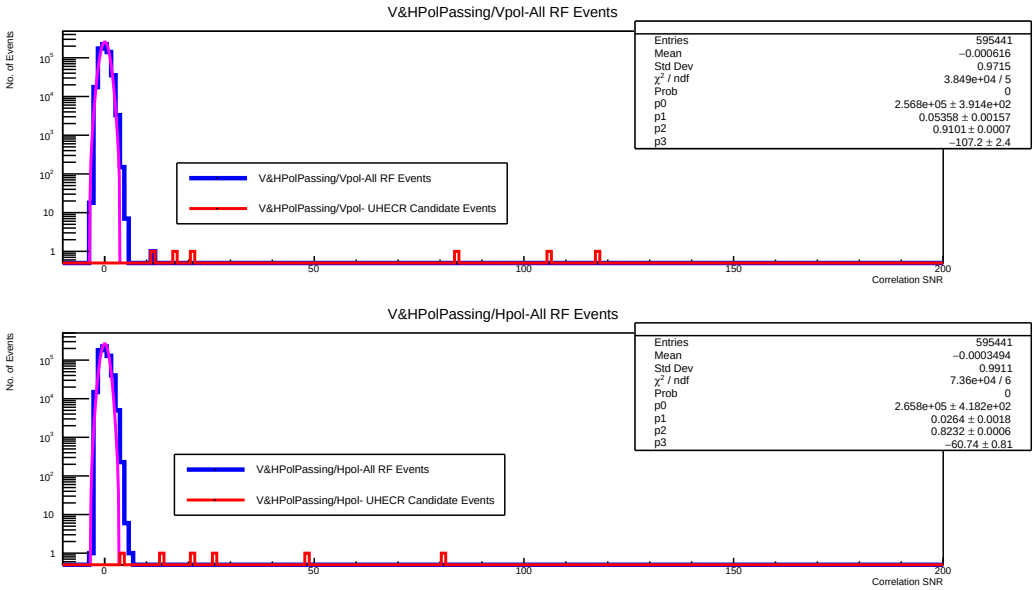
No.	UHECR Event Type-Pol	χ^2/NDF
1.	V&HPol Events-Vpol	$(3.85 \times 10^4)/5$
2.	V&HPol Events-Hpol	$(7.36 \times 10^4)/6$
3.	HPol Events-Vpol	$(5.34 \times 10^4)/7$
4.	HPol Events-Hpol	$(9.79 \times 10^4)/6$

The Gaussian fit function gave χ^2/NDF values of around $O(10000)$, and the high χ^2/NDF values indicate that the Gaussian distribution certainly does not represent the shape of the RF event *CorSNR* distributions. It can be seen from the fits in Figures B.1a that the fitted Gaussian distribution shows a big deviation from *CorSNR* data on the high-end tails which results in the high χ^2 values.

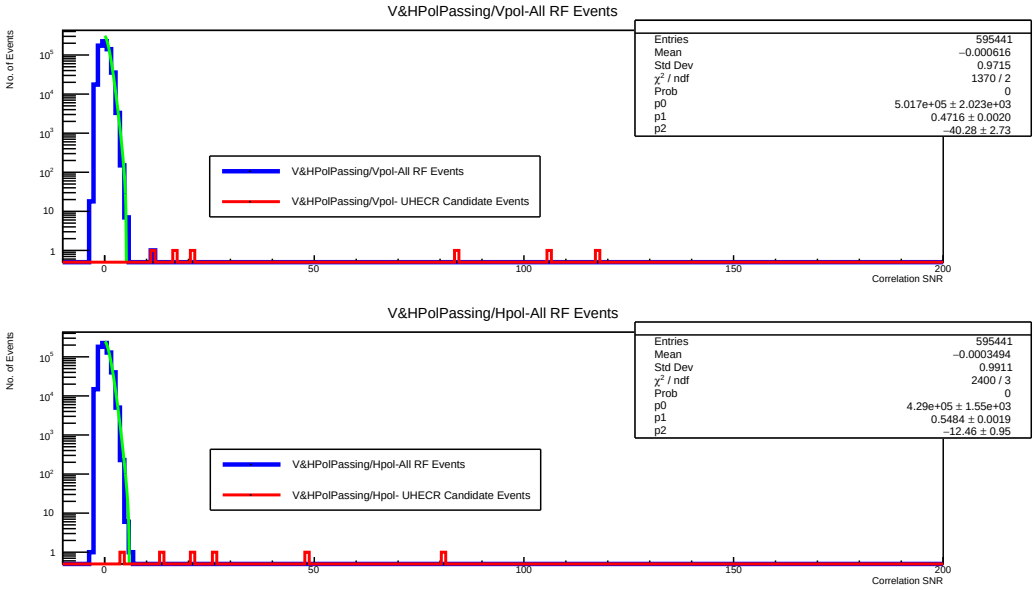
The top and bottom plots of Figure B.1b show the results of Poisson fits to RF event *CorSNR* distributions (blue distributions). The Poisson function in ROOT that was used for the fits was given by:

$$f(x) = \text{Par}[0] \times \texttt{TMath::Poisson}(x, \text{Par}[1]) + \text{Par}[2] \tag{B.2}$$

where x was the *CorSNR* value, $f(x)$ was the number of events, `TMath::Poisson` is the pre-defined Poisson function in ROOT and the values of Par[0], Par[1] and Par[2] were determined by fitting. The fit range was set from a *CorSNR* value of 0 to 10 because Poisson function was only valid for positive values of x. The χ^2/NDF values obtained from this fit function for all the RF event *CorSNR* distributions have been listed down in Table B.2.



(a) Gaussian distribution fits to the RF event distributions.



(b) Poisson distribution fits to the RF event distributions.

Figure B.1: Fits to RF event *CorSNR* distributions consisting of events from runs containing UHECR candidate events that passed V&Hpol cuts of the OSU neutrino analysis. The RF event distributions were fitted with Gaussian (**Figure (a)**) and Poisson (**Figure (b)**) distributions. The top and bottom plot of Figures (a) and (b) contain Vpol and Hpol *CorSNR* distributions, respectively. RF event *CorSNR* distributions, consisting of RF events taken from all the runs corresponding to UHECR candidate events, are represented by blue histograms and UHECR candidate event *CorSNR* distributions are represented by red histograms.

Table B.2: χ^2/NDF values for Poisson distribution fits to RF event *CorSNR* distributions. The fit range covered $0 < CorSNR < 10$.

No.	UHECR Event Type-Pol	χ^2/NDF
1.	V&HPol Events-Vpol	$1370/2$
2.	V&HPol Events-Hpol	$2400/3$
3.	HPol Events-Vpol	$4215/2$
4.	HPol Events-Hpol	$4564/2$

The Poisson fit function gave χ^2/NDF values of around $O(1000)$, which indicates that Poisson distribution does a reasonable job of fitting the high-end tails of the RF event *CorSNR* distributions.

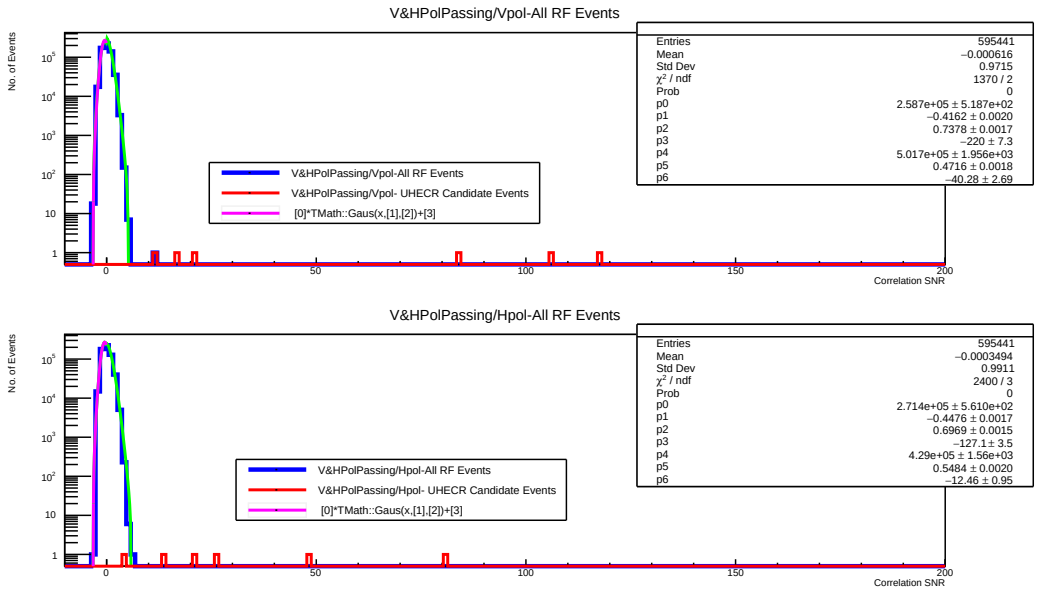


Figure B.2: Fits using a hybrid function (consisting of a mixture of Gaussian and Poisson distributions) to RF event *CorSNR* distributions consisting of events from runs containing UHECR candidate events that passed V&Hpol cuts of the OSU neutrino analysis. The top and bottom plot of Figures (a) and (b) contain Vpol and Hpol *CorSNR* distributions, respectively. RF event *CorSNR* distributions, consisting of RF events taken from all the runs corresponding to UHECR candidate events, are represented by blue histograms and UHECR candidate event *CorSNR* distributions are represented by red histograms.

Therefore based on the χ^2/NDF values reported in Tables B.1 and B.2 the shape of the RF

event *CorSNR* distributions in Figures B.1a and B.1b can be described by a mixture of Poisson and Gaussian distributions. A hybrid function was made which was mixture of the two distributions where for $CorSNR \leq 0$ the Gaussian function (Equation B.1) was used and for $CorSNR > 0$ the Poisson function (Equation B.2) was used. Figure B.2 shows the results of hybrid function fits to RF event *CorSNR* distributions (blue distributions). The χ^2/NDF values obtained from this fit function for all the RF event *CorSNR* distributions have been listed down in Table B.3.

Table B.3: χ^2/NDF values for the hybrid Gaussian+Poisson function fits to RF event *CorSNR* distributions. The fit range covered $-10 < CorSNR < 10$.

No.	UHECR Event Type-Pol	χ^2/NDF
1.	V&HPol Events-Vpol	$1370/2$
2.	V&HPol Events-Hpol	$2400/3$
3.	HPol Events-Vpol	$4361/4$
4.	HPol Events-Hpol	$4889/3$

The χ^2/NDF values listed in Table B.3 show that the hybrid fit brought a slight improvement in the case of the RF distributions obtained from runs that contained the UHECR candidate events that passed the Hpol cut of the OSU analysis. For the V&Hpol case, there was no improvement in the χ^2/NDF values compared to the values in Table B.2, which shows that the Gaussian part of the hybrid function did not contribute much to the χ^2/NDF values. Since the hybrid fit function gave the lowest χ^2/NDF values and covered the whole range of the RF event *CorSNR* distributions, it can be considered a viable fit function these distributions.

Appendix C

CoREAS Template Waveforms for 10% Unblinded Data Sample Analysis

This chapter contains the waveforms for the eight UHECR templates, which were used in the 10% unblinded data sample (i.e., "burn" sample data) analysis for ARA-2 and ARA-3 and were referred to in Section 7.3. These templates were made using CoREAS with cosmic ray energy set to 1×10^{18} eV. In the figures, the top row shows the four Top Vpol antennas in string 1 , string 2 , string 3 and string 4 respectively, the second row shows the Bottom Vpol channels, the third row shows the Top Horizontal channels and the fourth row shows the Bottom Hpol channels. The channels are numbered from 0 to 15 such that channels 0-3 (from left to right) are in the first row, channel 4-7 are in the second row, 8-11 are in the third row and 12-15 are in the fourth and last row.

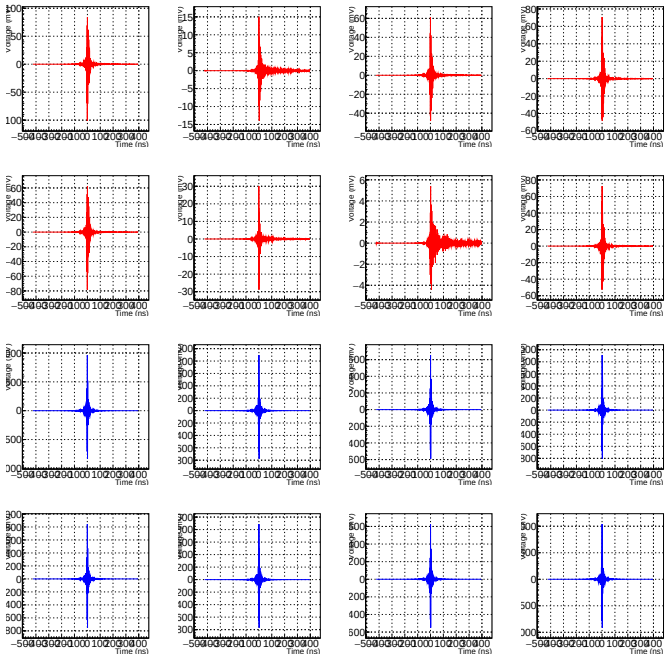


Figure C.1: Waveform for Template 0 which has a zenith angle of 30° and azimuth of 0°. The individual channel axes are Voltage (mV) vs Time (ns).

.

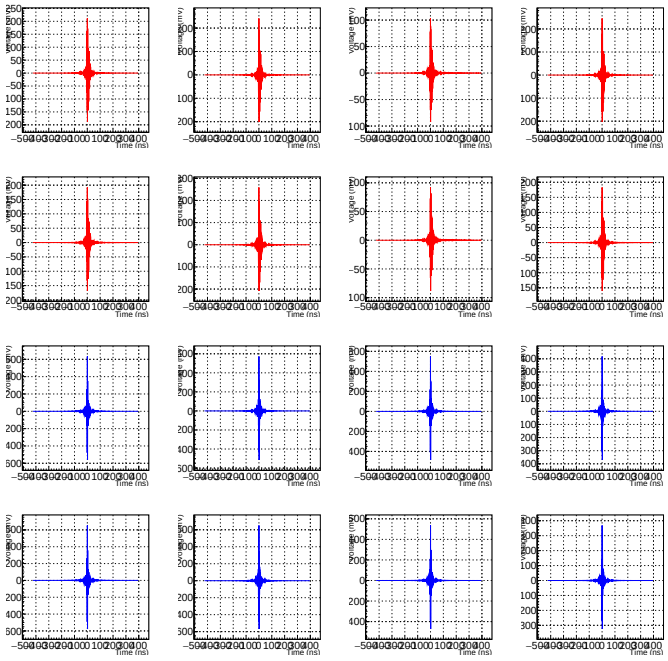


Figure C.2: Waveform for Template 1 which has a zenith angle of 30° and azimuth of 90°. The individual channel axes are Voltage (mV) vs Time (ns).

.

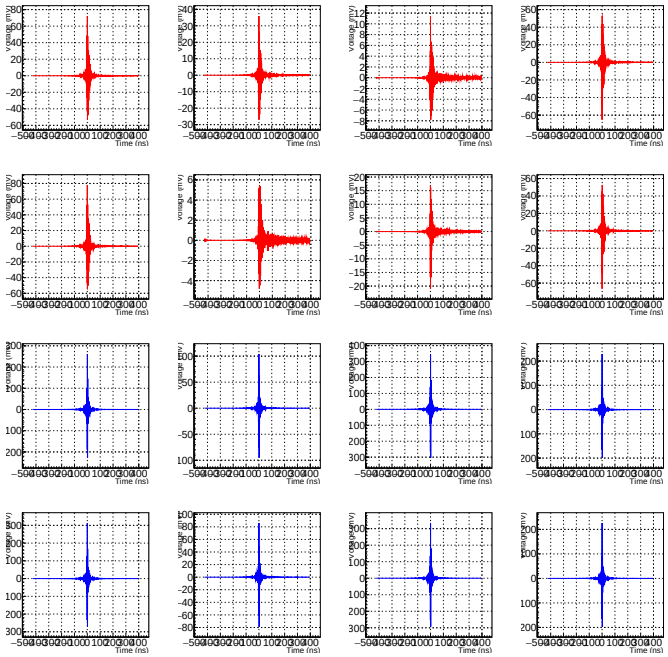


Figure C.3: Waveform for Template 2 which has a zenith angle of 30° and azimuth of 180°. The individual channel axes are Voltage (mV) vs Time (ns).

.

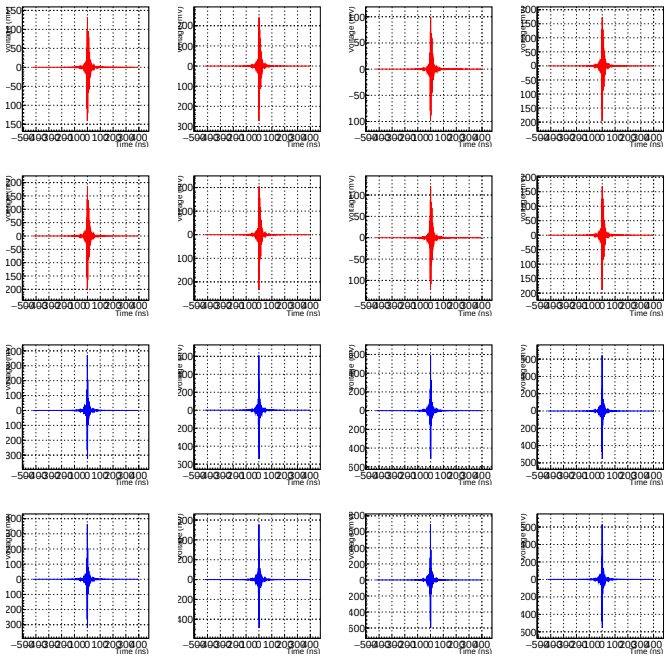


Figure C.4: Waveform for Template 3 which has a zenith angle of 30° and azimuth of 270°. The individual channel axes are Voltage (mV) vs Time (ns).

.

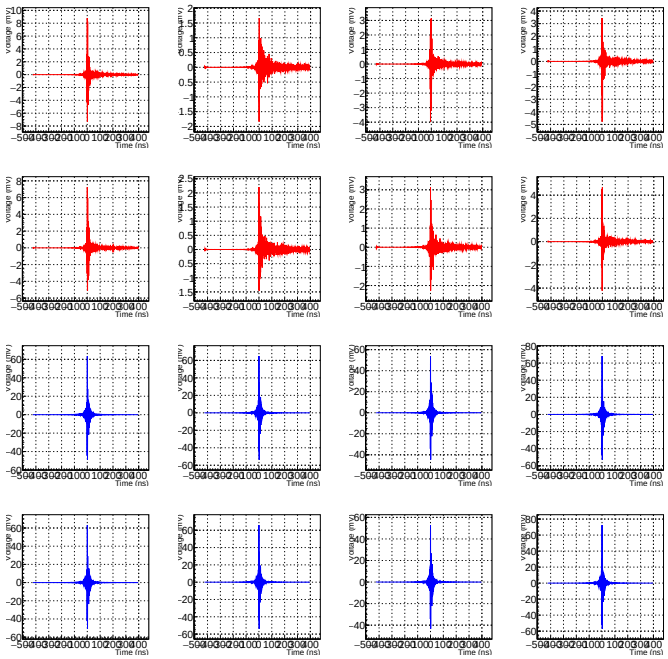


Figure C.5: Waveform for Template 4 which has a zenith angle of 60° and azimuth of 0°. The individual channel axes are Voltage (mV) vs Time (ns).

.

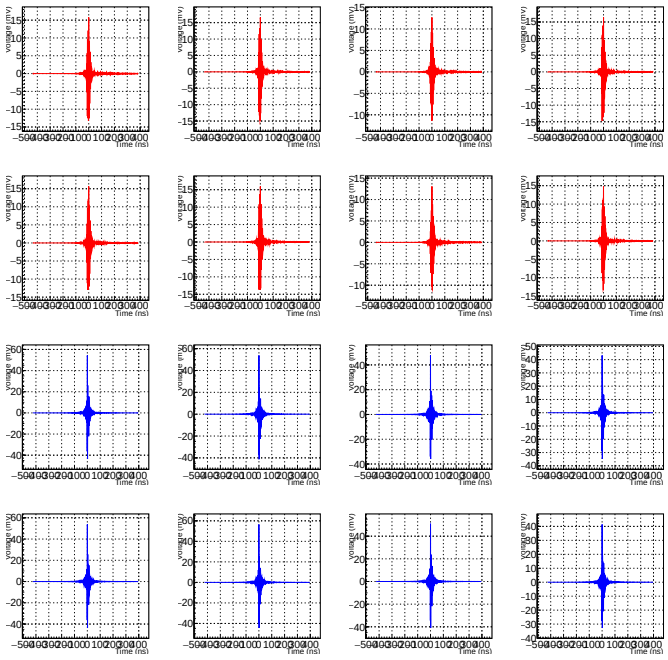


Figure C.6: Waveform for Template 5 which has a zenith angle of 60° and azimuth of 90°. The individual channel axes are Voltage (mV) vs Time (ns).

.

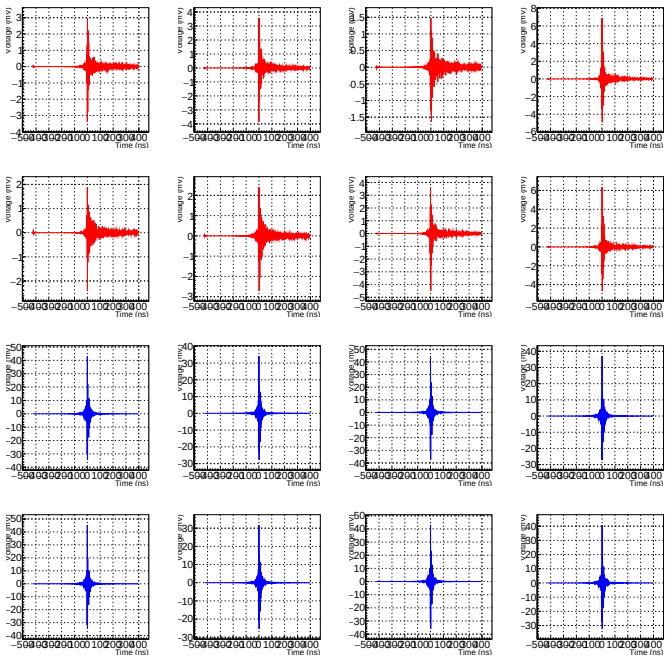


Figure C.7: Waveform for Template 6 which has a zenith angle of 60° and azimuth of 180°. The individual channel axes are Voltage (mV) vs Time (ns).

.

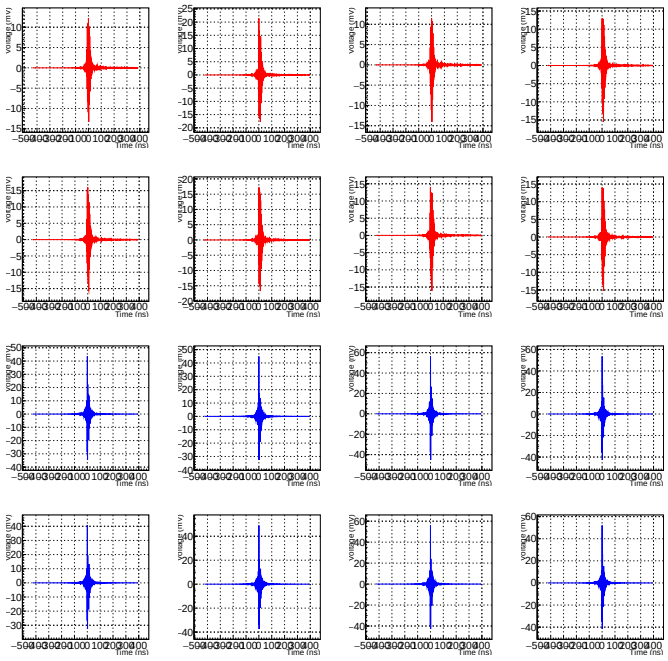


Figure C.8: Waveform for Template 7 which has a zenith angle of 60° and azimuth of 270°. The individual channel axes are Voltage (mV) vs Time (ns).

.

Appendix D

CorSNR Dependence on UHECR Shower Core Position

A UHECR "geometry specific" template was generated using CoREAS for a given ARA UHECR candidate event once the event geometry (i.e., reconstructed zenith and azimuth angles) had been obtained using interferometry. The zenith and azimuth angles obtained from the event reconstruction were used to calculate the shower propagation direction. The shower propagation direction was then used to calculate the 16 transect points on the ice surface for each of the 16 ARA antennas. In the absence of more shower details, it was assumed that the emitted EM rays were collinear with the shower propagation direction and the transect point on the ice surface for each ARA antenna was selected such that the emitted EM rays from the shower would enter the ice such that they would hit the ARA antenna.

One of the most important things that had to be determined before starting the CoREAS simulation was the placement of the UHECR shower core on the ice. For all the UHECR templates generated in this thesis, the average x and y coordinates of all the 16 transect points on the ice surface were taken to be the x and y coordinates of the shower core. However, UHECR templates having their shower cores placed in the average 16 transect point position represented a very special case of a UHECR shower for a given direction.

ARA-2 UHECR candidate event 126972 from run 3206 was therefore used to conduct a study to quantify the change caused by placing the shower core in a different position (as compared to the original position) in the "geometry specific" UHECR template. A change in the UHECR template would lead to a change in the overall template match score given by the *CorSNR* value.

The shower core position was adjusted four times by adjusting the displacement of the core position with respect to the ARA station origin by $+10$ m, -10 m, $+20$ m and -20 m. For each

shower core position, a corresponding UHECR template was generated for event 126972. For each new template, the zenith and azimuth angles and the positions of the 16 transect points on the ice were always kept at the original values that were calculated for the "geometry specific" template for event 126972 and the UHECR energy was fixed at 1×10^{18} eV. The four new templates, each having a different shower core position, were then matched with event 126972 resulting in Vpol and Hpol *CorSNR* values for each template and the *CorSNR* values have been reported in Table D.1.

Table D.1: *CorSNR* values as a function UHECR template shower core position. Four new UHECR templates were matched with ARA-2 UHECR candidate event 126972 from run 3206 and the *CorSNR* values for each polarization were calculated. The four templates used in this study were made by using the zenith and azimuth angles and the positions of the 16 transect points that were calculated for the original "geometry specific" template for event 126792. The only difference in the setup of the four templates was the displacement of the shower core with the respect to the ARA station origin which was adjusted by the amounts listed in the table. The UHECR energy of all four templates was kept constant at 1×10^{18} eV. The first entry for template 0 in the table is for the original "geometry specific" template for event 126792.

Template No.	Shower Core Displacement (m)	*CorSNR* Vpol	*CorSNR* Hpol
0.	0	10.150	12.036
1.	+10	10.560	12.798
2.	+20	12.023	14.844
3.	−10	10.758	13.579
4.	−20	11.654	14.049

The *CorSNR* results from Table D.1 indicate that a change in the UHECR template shower core position on the ice surface can result in at least 5% to 20% change in the original *CorSNR* values depending on the magnitude and direction of displacement of the shower core from its original position (which was the average of the 16 transect point positions on the ice surface).

Translating the shower core position essentially means translating the full radio footprint on the ice, for a given UHECR template. One way of locating the 'correct' shower core position for a template for a given ARA event, in the context of the current UHECR template generation method, would be :

- For a UHECR template, which was made for a given direction and energy, sample the full radio footprint on the ice surface, with as many observer (or antenna) positions as possible.
- Interpolation would then be used to reconstruct the complete radio footprint on the ice surface. The interpolation would use the CoREAS-returned E-fields at the specified observer positions to reconstruct the E-fields across the footprint.
- The radio footprint would then be translated across the ice surface; with every translation, new E-fields at the 16 transect points, and therefore at the ARA antennas, would be recalculated using interpolation. This would give a 'new' UHECR template for each footprint translation for the ARA event.
- All of the 'new' templates would be matched with the ARA event.
- The translated position of the shower radio footprint that returns a UHECR template that gives the maximal match score with the ARA event would then be chosen as the 'correct' position of the radio footprint, and therefore the shower core, for that UHECR template.

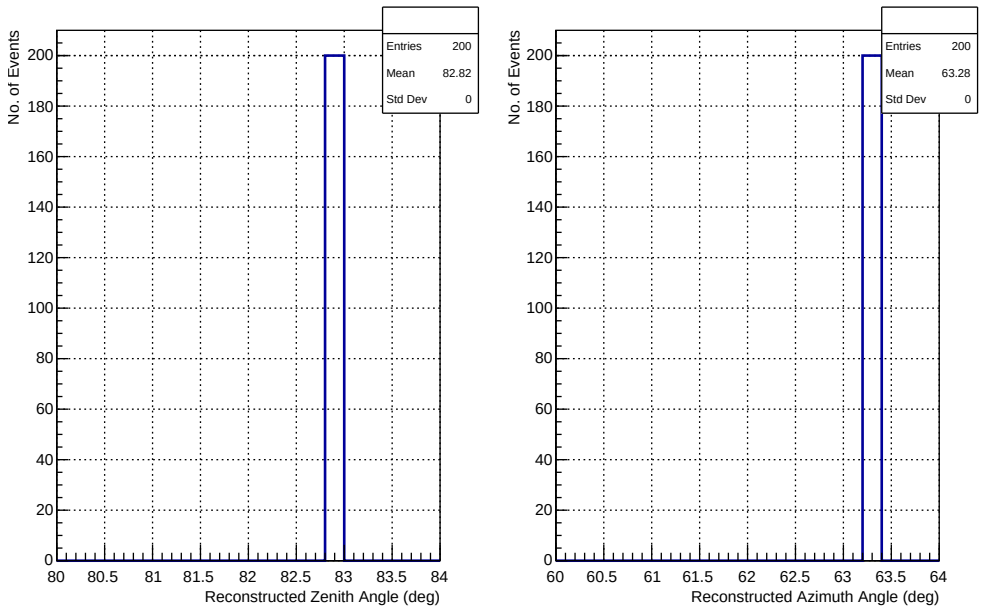
This exercise just emphasizes the importance of placing the shower core in the 'correct' position as it can significantly affect the template match scores given by the *CorSNR* values. Hence to fully understand and recreate an ARA UHECR candidate event and have a maximal template match score, it is crucial to find the 'correct' (or the actual) UHECR shower core position on the ice surface.

Appendix E

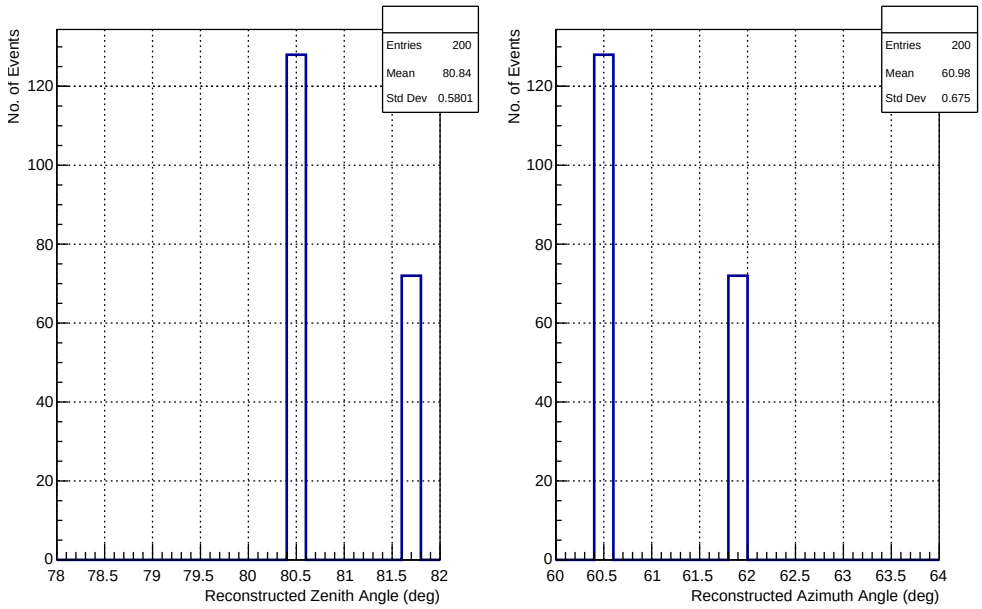
Accuracy and Precision of the Interferometer

Interferometric reconstruction was used in order to reconstruct the zenith and azimuth angles of all the data events for ARA-2 and ARA-3 stations. The reconstructed angles assisted in the identification of possible noise sources and also of UHECR candidate events. In order to carry out interferometric reconstruction, radiowave propagation times associated with all directions at a given radius from the station were computed for each ARA channel using raytracing. The channel waveforms were then shifted in time by adding radiowave propagation times to them, and the waveforms were then cross-correlated with each other. The correlation scores were then summed to get a 'coherence' score for each given direction. The direction with the highest coherence value was chosen to be the reconstructed direction of the source.

To estimate the uncertainties of the interferometer's angular reconstructions, 400 ARA-2 calpulser events were reconstructed. 200 events came from run 9129 in which the ARA-2 string 6 Vpol calpulser was emitting pulses towards the station and 200 events came from run 9185 in which the ARA-2 string 6 Hpol calpulser was emitting pulses towards the station. In both cases, the calpulsers operated at 0 dB attenuation and triggered ARA-2 once every second. All eight Vpol channels were used to reconstruct string 6 Vpol calpulser events and seven Hpol channels (excluding channel 15) were used to reconstruct string 6 Hpol events. The results of the reconstruction of all 400 events are shown in Figure E.1.



(a) Reconstructed zenith and azimuth values for ARA-2 string 6 Vpol calpulser



(b) Reconstructed zenith and azimuth values for ARA-2 string 6 Hpol calpulser

Figure E.1: Distributions of reconstructed zenith and azimuth values of ARA-2 string 6 Vpol and Hpol calpulsers. The string 6 Vpol calpulser distributions consist of 200 events taken from run 9129 and string 6 Hpol calpulser distributions consist of 200 events taken from run 9185. In both runs the respective calpulsers were operating at 0 dB attenuation.

The 'true' (or calculated) zenith and azimuth values of string 6 Vpol calpulser are 84.059° and 63.46° respectively; for string 6 Hpol calpulser, the corresponding values are 79.537° and 63.46° respectively. Figure E.1a shows that all 200 string 6 Vpol calpulser events reconstructed to the same zenith and azimuth values which were 82.82° and 63.28°. It should also be noted that there was a difference of $+1.24°$ and $+0.18°$ between the 'true' and the reconstructed zenith and azimuth values respectively. Figure E.1b shows the reconstruction results for zenith and azimuth values for string 6 Hpol pulser. The precision slightly deteriorates in the case of string 6 Hpol pulser as the reconstruction points to two possible directions for both the zenith and azimuth values. The average reconstructed zenith and azimuth values for the Hpol calpulser were 80.84° and 60.98° with standard deviations of 0.58° and 0.68° respectively. In the case of the Hpol calpulser there was a difference of $-1.303°$ and $+2.48°$ between the 'true' and the average reconstructed zenith and azimuth values respectively.

To summarize, the reconstructed zenith values for both Vpol and Hpol channels were off by around a degree, and the reconstructed azimuth and zenith values for Hpol channels were less precise as compared to VPol channels. Therefore based on these results, conservative estimates on the errors of reconstructed zenith and azimuth values from the interferometer can be taken to be:

1. Vpol zenith: $\pm 1.24°$
2. Hpol zenith: $\pm 1.42°$
3. Vpol azimuth: $\pm 0.18°$
4. Hpol azimuth: $\pm 2.57°$

Here the values of Hpol zenith and azimuth errors were obtained by adding the standard deviations and the differences between the 'true' and reconstructed values in quadrature. The fact that Hpol reconstruction uses one less channel than the Vpol reconstruction could be causing the Hpol reconstructions to have larger uncertainties.

The interferometer uses the HEALPix (Hierarchical Equal Area isoLatitude Pixelization) software to do reconstruction and make interferometric sky maps (Gorski et al., 2005). These sky

maps consist of pixels and a larger number of pixels implies a better angular resolution for the interferometric reconstruction. However, that comes at the cost of increased computation time. The HEALPix manual reports that the number of pixels on either side (i.e., zenith side or azimuth side) of the map is given by 2^{ResInd} where *ResInd* is the resolution index and was set to a value of 5 for this thesis. The total number of pixels in a map is given by $12 \times (2^{ResInd})^2$ and for a *ResInd* value of 5 the total number of pixels on the sky map comes out to be 12288. According to the HEALPix manual, this means that the mean pixel spacing (or angular resolution) for the sky maps used in this thesis comes out to be 1.8323°. The reconstructed zenith and azimuth values for ARA-2 calpulsers are certainly within the reported angular resolution of the 'true' zenith and azimuth values except for Hpol azimuth values, which seem to be slightly more spread out than this angular range.

Appendix F

Monte Carlo for UHECR candidate events

A Monte Carlo (MC) was designed to study the efficiency of the Level-1 cuts used to search for UHECR candidate events in the ARA-2 and ARA-3 "burn" sample data for 2013 to 2016. UHECR template events were generated using CoREAS for a variety of energies and directions, and then noise was added to the UHECR radio signals in the templates. The eight template average method was then applied to the noise-added UHECR template events, and the efficiency of Level-1 cuts in identifying UHECR candidate events was calculated. The efficiency calculations were done only with the ARA-2 station geometry as ARA-2 and ARA-3 have very similar station structures (i.e., placement of strings and depth of antennas), so their efficiencies should be very similar.

For this MC study, 1008 UHECR templates were initially generated with UHECR energies set to 1×10^{18} eV using CoREAS. The templates covered a zenith angle range from 5° to 70° in 5° steps and an azimuth range of 0° to 355° in 5° steps. In order to conduct this MC study the following steps were taken:

1. Since the UHECR radio signal amplitude in the templates scales linearly with UHECR energy, to generate UHECR templates for different UHECR energies the radio signals in the original 1008 template events were just scaled up or down in amplitude depending on the energy. UHECR templates were generated for energies $\log(\text{Energy}) = 16 \rightarrow \log(\text{Energy}) = 21$ with an energy exponent step size of 0.25. In total, 20160 UHECR templates were generated.

2. Gaussian noise was then added to the template channel waveforms in order to make UHECR template events. The average noise RMS values for each channel were obtained from data (listed in Table F.1). The channel RMS was used as the sigma for the Gaussian function and

a random number generator was used to extract voltage values from the Gaussian Probability Distribution Function (PDF).

3. A sample of 1000 pure noise events were also generated to mimic thermal noise triggers. Thermal noise events makeup most of the ARA RF events after the calpulser and the software events have been rejected. These simulated 1000 noise events were then used to calculate the RMS of RF event correlation scores. The RMS of the correlation scores of RF events (which in this case are all noise events) needs to be calculated as it enters in the denominator of the *CorSNR* definition given in Equation 7.1.

4. All of the UHECR template events were then run through the same eight template average analysis previously described (as defined in Section 7.3). The *CorSNR* values for each UHECR template event were then calculated for each polarization using the correlation scores obtained from template matching.

Table F.1: ARA-2 channel noise RMS values

Ch. no.	RMS (mV)	RMS error (mV)
0.	18.911	0.008
1.	20.716	0.050
2.	39.744	0.015
3.	44.049	0.021
4.	32.523	0.012
5.	20.579	0.008
6.	42.241	0.022
7.	29.596	0.010
8.	42.204	0.207
9.	22.726	0.046
10.	26.314	0.073
11.	26.284	0.130
12.	29.134	0.072
13.	23.683	0.176
14.	30.655	0.079
15.	21.861	0.026

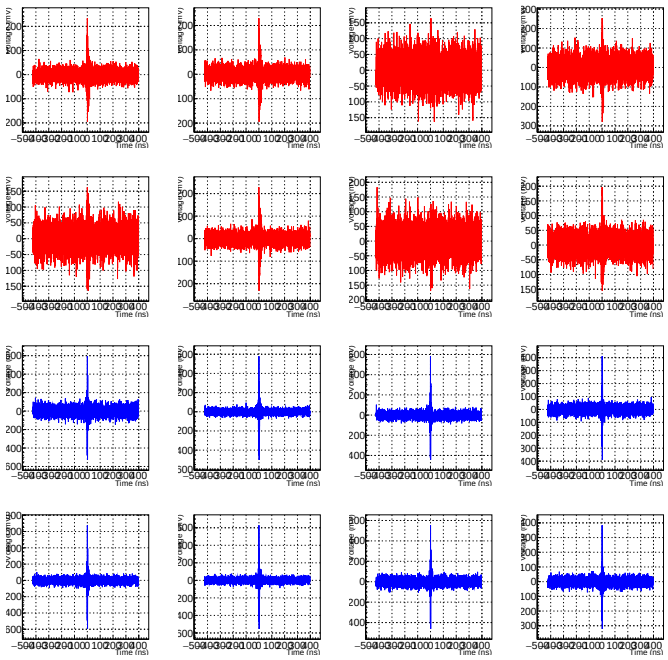


Figure F.1: Waveform for a CoREAS template with Gaussian noise added to the signals. The template was made at an energy of 1×10^{18} eV, zenith angle of 30° and an azimuth angle of 90°. The individual channel axes are Voltage (mV) vs. Time (ns). The top row shows the four Top Vpol antennas in string 1 , string 2 , string 3 and string 4 respectively, the second row shows the Bottom Vpol channels, the third row shows the Top Horizontal channels and the fourth row shows the Bottom Hpol channels. The channels are numbered from 0 to 15 such that channels 0-3 (from left to right) are in the first row, channel 4-7 are in the second row, 8-11 are in the third row and 12-15 are in the fourth and last row.

.

Figure F.1 shows an example of a UHECR template event after Gaussian noise has been embedded in its waveforms.

It should be noted that the voltage SNR for a channel waveform was defined as the maximum peak-to-peak channel amplitude divided by the noise RMS of the waveform. The third highest channel SNR was then defined as the voltage SNR for that polarization of the UHECR template event.

To identify a UHECR candidate from the "burn" sample data, Level-1 cuts were applied (defined in Section 7.5.2). Cut 3 of the Level-1 cuts was defined as *CorSNR* Hpol>*CorSNR* Vpol, and events that passed this cut had a better template match score in Hpol rather than Vpol. Cut 4 of

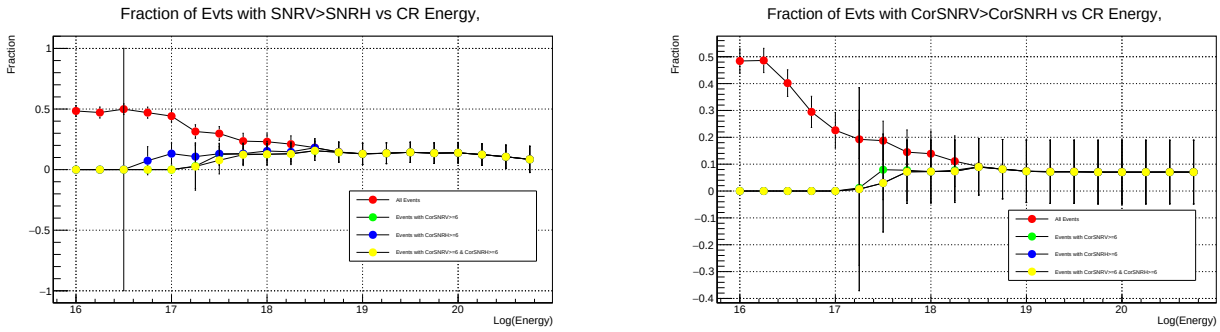


Figure F.2: Fraction of events at each energy bin (**left**) having voltage SNR Vpol>voltage SNR Hpol and (**right**) having *CorSNR* Vpol>*CorSNR* Hpol. There are four curves, each for a different type of UHECR template event: **1)** All events (red curve), **2)** events with *CorSNR* Vpol$>=6$ (green curve), **3)** events with *CorSNR* Hpol$>=6$ (blue curve) and **4)** events with *CorSNR* Vpol$>=6$ and *CorSNR* Hpol$>=6$ (yellow curve).

the Level-1 cuts was defined as voltage SNR Hpol> voltage SNR Vpol and it implied that events passing this cut had to have a higher voltage SNR in Hpol rather than Vpol. Cut 4 comes from the fact that the magnetic field at the South Pole is almost vertical; therefore, because of the Lorentz force (i.e., $v \times B$), the strongest emission happens in the horizontal direction. Cut 4 is connected to cut 3 by the fact that events having a higher voltage SNR in Hpol would also naturally give a better match (i.e., higher *CorSNR* value) in Hpol because noise would not be able to distort the shape of the higher SNR pulses.

In order to verify cut 3 of the Level-1 cuts, the fraction of UHECR template events not passing this cut was calculated for each energy bin. The results are shown in the right plot of Figure F.2. The relevant curve in this context is in yellow and is for events that fulfill the condition *CorSNR* Vpol$>=6$ and *CorSNR* Hpol$>=6$ (which is also cut 2 of our Level-1 cuts). It is evident that about 10% of the UHECR template events for high UHECR energy values do not pass cut 3 (i.e., have a better template match score in Vpol as compared to Hpol) and this points to the fact that for some UHECR event geometries the received Vpol power in the ARA antennas might be slightly stronger than Hpol power. At low energies, all the UHECR template events which pass cut 2 also pass cut 3.

Cut 4 also has similar trends, as compared to cut 3, for the fraction of events that do not pass cut

4, as can be seen from the yellow curve in the left plot of Figure F.2. The reasons for the similarity in trends for cuts 3 and 4 are connected as UHECR template events which do not pass cut 3 (i.e., have a better template match score in Vpol as compared to Hpol) also do not pass cut 4 (i.e., have received more radio power in Vpol ARA antennas than Hpol). It can also be deduced from the yellow curves in Figure F.2 that the energy bin of $10^{17.5}$ eV is approximately the point at which the UHECR template Vpol pulses, and therefore Hpol pulses too, have become strong enough to be visible above noise.

It can be concluded that having cuts 3 and 4 as a part of our Level-1 cuts is reasonable since we will always have 90% or more UHECR events that will pass those cuts. The Monte Carlo was then used to calculate the efficiency of Level-1 cuts in identifying possible UHECR events, as a function of energy, and the results are shown in Figure F.3.

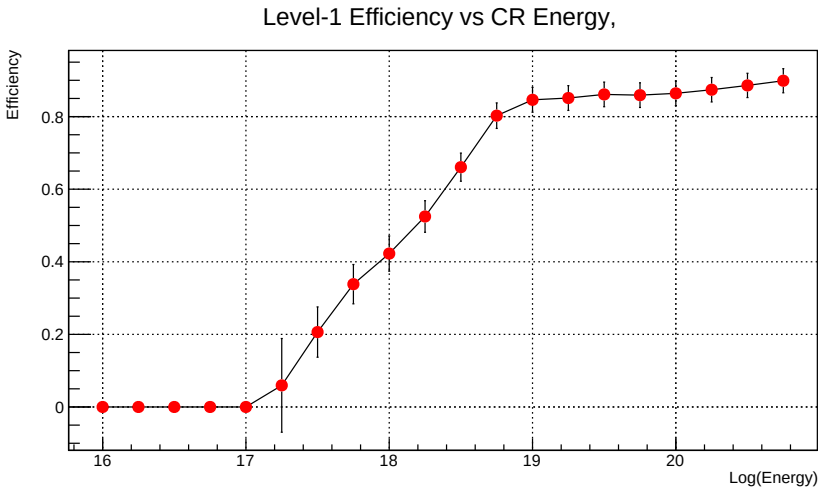


Figure F.3: Efficiency of Level-1 cuts as a function of UHECR energy.

Figure F.3 does not show the UHECR trigger efficiency of the ARA station but the efficiency of identifying a UHECR candidate using the template matching method given the ARA station already triggered.

Appendix G

Calibration Pulser Templates for UHECR Search in ARA data

The ARA calibration pulsers (or calpulsers) were primarily designed to be impulsive and periodically verify that the ARA stations can trigger on impulsive radio signals. All ARA stations have two calpulser strings, in addition to the four strings consisting of the receiving antennas, except ARA-4 and ARA-5, which only have one calpulser string. Each calpulser string has a Vpol and a Hpol calpulser. At any point in time, only one of the station calpulsers is active and emits one radio pulse per second.

In this thesis, the focus was primarily on using UHECR templates to identify UHECR candidate events in ARA data using the template matching method. However, we wanted to verify that identification of UHECR events could be done with templates made from any impulsive event (i.e., calpulser events). Therefore, in this study, templates made out of calpulser events were matched with known UHECR candidate events in ARA-2, listed in Table 7.1, that were found in the OSU neutrino analysis. This was done to compare how efficiently the calpulser templates could identify the UHECR candidate events (as compared to the UHECR templates) by comparing the *CorSNR* values obtained from the template matching process.

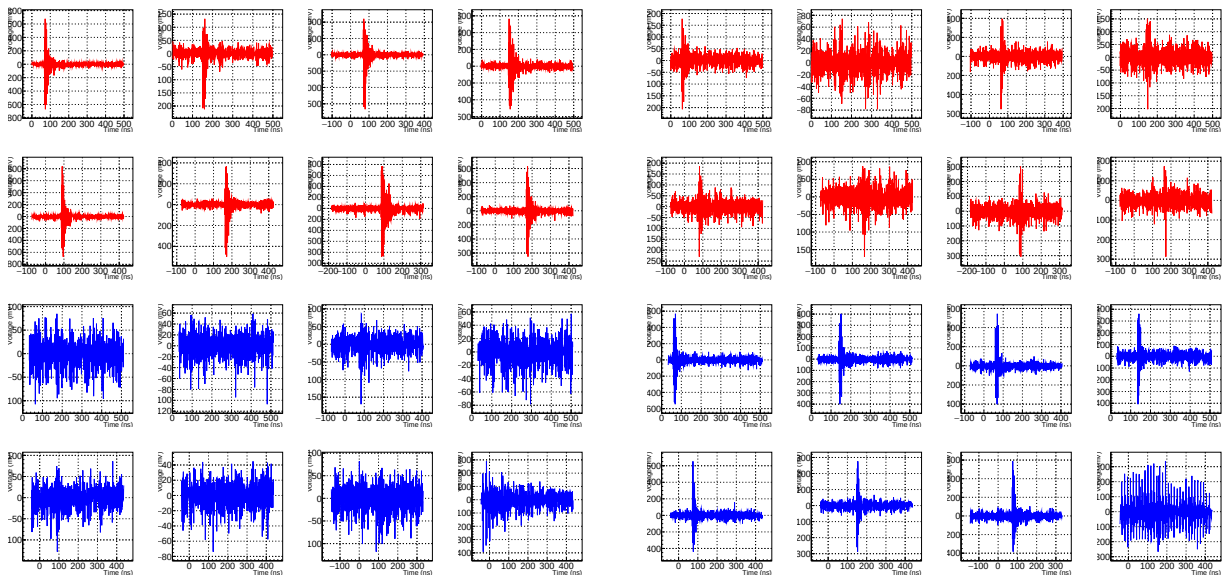
Two ARA-2 calpulser events were chosen to make two calpulser templates. The two events were:

1. The first calpulser event of run 9129, which had calpulser triggers from string 6 Vpol calpulser. Figure G.1a shows the waveform for this event. As expected for the Vpol calpulser, we only have radio pulses in the eight Vpol ARA antennas (red-colored channel waveforms).

2. The first calpulser event of run 9185, which had calpulser triggers from string 6 Hpol

calpulser. Figure G.1b shows the waveform for this event. It is apparent that, not only do we have radio pulses in the eight Hpol antennas (blue-colored channel waveforms), we also have significantly strong radio pulses in the eight Vpol antennas (red-colored channel waveforms). This is unexpected for a Hpol calpulser. This could be caused by, either the Hpol calpulser antenna emitting pulses in both Vpol and Hpol, or the receiving Vpol antennas having a strong acceptance for Hpol pulses. In this case, it is probably a combination of both of those factors.

In both of these events, the calpulsers were operating at 0 dB attenuation, so the calpulsers were emitting radio pulses at maximum available power. This ensured the calpulser pulses had high voltage SNR values and were much stronger than the RF noise levels in the respective channels.



(a) Calpulser event in run 9129 from string 6 Vpol calpulser

(b) Calpulser event in run 9185 from string 6 Hpol calpulser

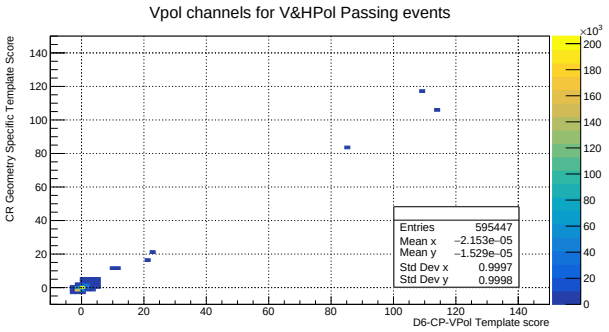
Figure G.1: Event waveforms for calpulser triggers from ARA-2 string 6 Vpol calpulser (**left**) and Hpol calpulser (**right**). The calpulsers were operating at 0 dB attenuation in both of these cases.

The two calpulser templates were matched with all the ARA-2 UHECR candidate events (that were found in the OSU analysis) and the RF events present in the runs corresponding to the UHECR candidate events. After the matching, *CorSNR* values were calculated for each polar-

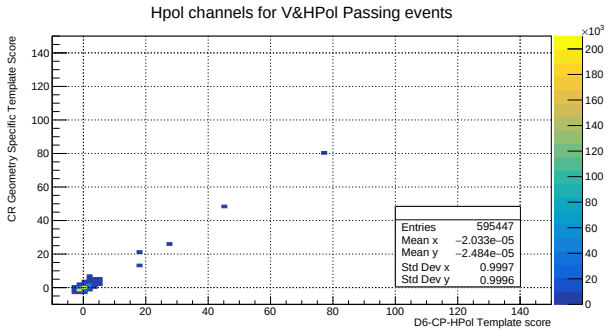
ization for each event. The Vpol *CorSNR* values of the Vpol calpulser template (i.e. event from run 9129) were compared with the Vpol *CorSNR* values of the "geometry specific" UHECR template. The Hpol *CorSNR* values of the Hpol calpulser template (i.e., event from run 9185) were compared with Hpol *CorSNR* values of "geometry specific" UHECR template.

Figures G.2a, G.2b, G.2c and G.2d are two dimensional histograms which have *CorSNR* values obtained from the matching of calpulser templates on the x-axis and *CorSNR* values obtained from the matching of "geometry specific" UHECR templates on the y-axis. The z-axis represents the number of events in each *CorSNR* bin. Figures G.2a and G.2b, show the *CorSNR* values for UHECR candidate events that passed the V&HPol cut of the OSU analysis and their corresponding RF events. Figures G.2c and G.2d, show the *CorSNR* values for UHECR candidate events that passed the HPol-only cut of the OSU analysis and their corresponding RF events. Figures G.2a and G.2c, consist of Vpol *CorSNR* values. Figures G.2b and G.2d consist of Hpol *CorSNR* values for all events.

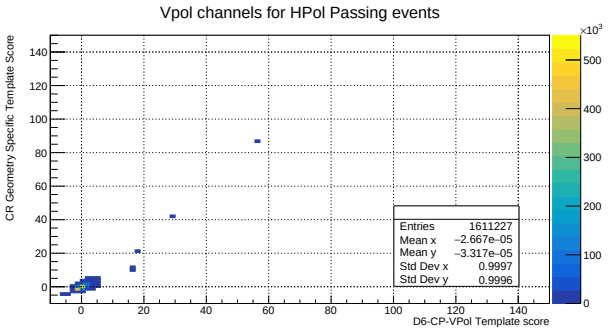
All four *CorSNR* distribution histograms shown in Figure G.2 demonstrate that most of the events have *CorSNR* values around zero for both the calpulser and the UHECR templates. This cluster of events mostly consists of RF events with only a handful of UHECR candidate events. The handful of events that deviate outside the big cluster of events located around the origin are the UHECR candidate events, which have relatively high *CorSNR* values for both the calpulser template and the "geometry specific" UHECR templates. The most important thing that should be noted about all four figures is that events with high *CorSNR* values follow a linear trend. This implies that for a given UHECR candidate event, calpulser templates provide an equally high *CorSNR* value as compared to "geometry specific" UHECR templates.



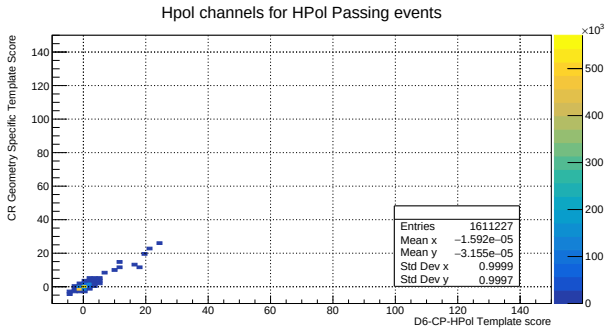
(a) Vpol *CorSNR*s for UHECR candidate events that passed the V&Hpol cut and the RF events found in the corresponding runs.



(b) Hpol *CorSNR*s for UHECR candidate events that passed the V&Hpol cut and the RF events found in the corresponding runs.



(c) Vpol *CorSNR*s for UHECR candidate events that passed the Hpol-only cut and the RF events found in the corresponding runs.



(d) Hpol *CorSNR*s for UHECR candidate events that passed the Hpol-only cut and the RF events found in the corresponding runs.

Figure G.2: All UHECR candidate events (found in the ARA-2 data in the OSU neutrino analysis) and the RF events found in the runs corresponding to the UHECR candidate events were matched with calpulser templates and "geometry specific" UHECR templates, and *CorSNR* values were calculated for each template matching case for each polarization for each event. These Figures show the combined *CorSNR* distributions of UHECR candidate events and the corresponding RF events. *CorSNR* values obtained from the matching of calpulser templates are on the x-axis and *CorSNR* values obtained from the matching of "geometry specific" UHECR templates on the y-axis. The z-axis represents the number of events in each *CorSNR* bin. (**top row**) *CorSNR* distributions for UHECR candidate events that passed the V&Hpol cut of the OSU analysis and their corresponding RF events; (**bottom row**) *CorSNR* distributions for UHECR candidate events that passed the Hpol-only cut of the OSU analysis and their corresponding RF events; (**left column**) Vpol *CorSNR* values for all UHECR candidate events and RF events; (**right column**) Hpol *CorSNR* values for all UHECR candidate events and RF events. It should be noted that the Vpol *CorSNR* values of the Vpol calpulser template (i.e. event from run 9129) were compared with the Vpol *CorSNS* values of the "geometry specific" UHECR template. The Hpol *CorSNR* values of the Hpol calpulser template (i.e. event from run 9185) were compared with Hpol *CorSNR* values of "geometry specific" UHECR template.

This result might seem unexpected at first, as one would assume that the "geometry specific" UHECR templates should always give a better match to the respective UHECR candidate event as compared to any other template. However, the calpulser templates seem to perform well in matching with the UHECR candidate events because both calpulser and UHECR templates have impulses of similar time durations and a shorter time-scale than the ARA system response. Hence, the ARA system response dominates, and this causes the radio pulses to have a similar shape in the channel waveforms for both calpulser and UHECR templates.

Appendix H

Optimization of South Pole Ice $n(z)$ Model Using Travel Time Differences Between Direct and Reflected/Refracted rays ($dt(D,R)$)

The simulation of propagation radio waves (or raytracing) in South Pole ice needs to be done with extreme accuracy as it is central to generating UHECR templates for ARA stations. The accuracy of raytracing is subject to having an accurate South Pole ice $n(z)$ model. South Pole ice $n(z)$ models have been derived in various ways(Kravchenko et al., 2004):

1. Using relative radio pulse hit times at multiple receiving antennas from a pulsing transmitting antenna. This method leads to the measurement of the average $n(z)$ profile between the transmitter and the receiver.

2. Using ice density data to calculate the ice refractive index as a function of depth. This method uses the Schytt model that relates ice density to refractive index and is given by $n(z) = 1 + 0.845\rho(z)$. Here $\rho(z)$ is given in units of Mg/m^3.

In this chapter, the measurement of South Pole ice $n(z)$ model will be attempted by using the travel time differences between Direct (D) and Reflected/Refracted (R) rays ($dt(D,R)$) values from ARA data. This $dt(D,R)$ data was taken during the 2018-2019 summer season at the South Pole when the SPUNK (South Pole UNiversity of Kansas) pulser was lowered (or "dropped") into the SPICE (South Pole ICE) core borehole and was made to emit radio pulses at one pulse per second. This caused emission of a direct and refracted or reflected pulse towards each of the ARA receivers (or channels), enabling the collection of the D and R radio pulses, as a function of SPUNK pulser

depth, in each of the ARA channels. As seen in Section 5.2.2, the expressions for the travel times of direct, reflected and refracted rays are essentially a function of the refractive index model, $n(z)$, and the L parameter which is the initial condition of the ray and relates to the initial launch angle. Therefore $dt(D,R)$ can be written as:

$$dt(D,R) = f(A,B,C,z_{Tx},z_{Rx},Distance) \tag{H.1}$$

here *Distance* is the horizontal distance between the receiver (Rx) and transmitter (Tx), z_{Tx} is the depth of Tx and z_{Rx} is the depth of Rx. Here it can be clearly seen that the time difference $dt(D,R)$ is a function of the refractive index model parameters A, B and C (ABC). It should be noted that we have assumed $n(z)$ to have an exponential profile. L is not explicitly written in full here as L is also a function of the ABC parameters, z_{Tx}, z_{Rx} and the *Distance*. For the given values of these three parameters, two L values are calculated; one for the Direct ray and one for the Reflected/Refracted ray. The procedure to find the value of L, using these parameters, has been explained in Section 5.2.1.

Therefore if the values of z_{Tx}, z_{Rx} and *Distance* are known, and the ABC values are also known, then the value of $dt(D,R)$ can be found analytically. The reverse is also true: if the value of $dt(D,R)$ is known from data, the values of the parameters ABC can be adjusted until the analytical value of $dt(D,R)$ matches the empirical one. This is basically a minimization procedure that can be simply written as:

$$dt(D,R)_{Data} - dt(D,R)_{Analytical} = 0 \tag{H.2}$$

Using this procedure, the ABC parameter values can be inferred from the data-derived values of $dt(D,R)$.

Figure H.1 shows the dependence of $dt(D,R)$ as function of the SPUNK pulser depth in the ARA-2 station channels. The $dt(D,R)$ data, shown in Figure H.1, was taken during the 2018-2019 summer season when the SPUNK pulser was "dropped" into the SPICE core borehole. It can be seen from Figure H.1 that using the default ARA South Pole ice refractive index model

(or the AraSim model), given by $n(z) = 1.78 - 0.43e^{0.0132z}$, for raytracing the D and R rays between SPUNK pulser and ARA antennas, offers a relatively high degree of agreement between the $dt(D,R)$ time predictions and data values but there is definitely room for improvement (Allison et al., 2019b).

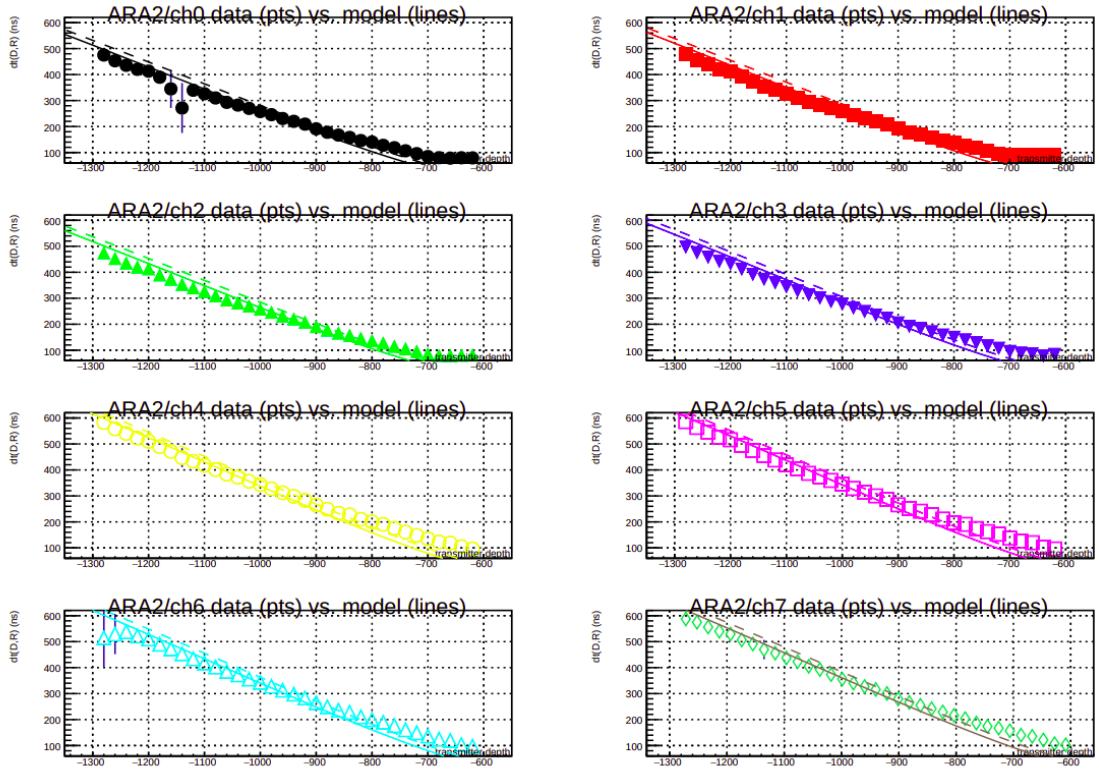


Figure H.1: $dt(D,R)$ data (in ns) recorded as a function of SPUNK depth (in m) in ARA-2 channels 0 to 7 during one of SPUNK pulser runs conducted in the 2018-2019 season. The data $dt(D,R)$ is shown as the points. The solid line is the $dt(D,R)$ prediction using the default ARA ice refractive index model ($n(z) = 1.78 - 0.43e^{0.0132z}$ where z is negative depth in meters). The dashed line is for a slightly tweaked model given by $n(z) = 1.78 - 0.43e^{0.0125z}$. Figure adapted from (Allison et al., 2019b)

.

One way of doing the minimisation procedure given in Equation H.2 is by using $dt(D,R)$ vs. Tx depth priors for the station channels as shown in the plots in Figure H.1. This can be done by fixing the values of:

1. A to 1.78 since that is the asymptotic value of the ice refractive index at the South Pole and is well-measured.

2. The *Rx* depth since we know the depths of the ARA receiving antennas.

3. The *Distance* since we know the distance from each of the ARA antenna holes to the SPICE core borehole.

Therefore Equation H.1 can then be rewritten as:

$$dt(D,R) = f(B,C,z_{Tx}) \tag{H.3}$$

Equation H.3 can be now used to fit the data shown in Figure H.1, with B and C left as free parameters extracted from the fit and $dt(D,R)$ as a function of the Tx (in this case SPUNK) depth.

H.1 Fitting the 2018-2019 ARA $dt(D,R)$ data from the SPUNK Pulser Drop

H.1.1 Setting up the Fit Function for the ARA Stations

One important thing to note is that ARA-1 is a relatively shallow station as compared to the other stations. All the ARA-1 antennas have depths of around -70 m, whereas all the other stations have antennas with depths ranging from -180 to -200 m. Due to this, the ARA-1 station receives double pulse combinations of Direct and Reflected pulses and Direct and Refracted pulses as it is so close to the ice surface. By contrast, all the other deeper stations receive double pulse combinations of Direct and Refracted pulses from the SPUNK pulser as it moves down the SPICE core borehole.

For all the stations except ARA-1, the $dt(D,R)$ data trends, as a function of SPUNK depth, were similar to what is shown on the right plot of Figure H.2. The $dt(D,R)$ trends are linear in nature for all antennas of all stations except for ARA-1. For ARA-1, it seems like two linear lines with slightly different gradients are joined together at a certain SPUNK pulser depth. This depth is called the transition depth ($z_{Transition}$) for the SPUNK pulser and was found to be around 880 m for the ARA-1 channels. Above the transition depth $dt(D,R)$ times, for ARA-1, come from a combination of Direct and Refracted pulses and below that they come from a combination of Direct and Reflected pulses.

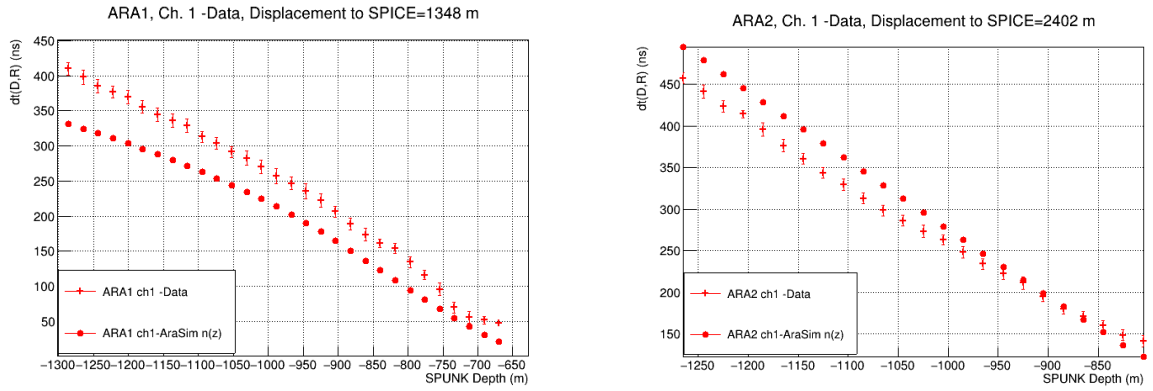


Figure H.2: $dt(D,R)$ data (in ns) vs SPUNK depth (in m) for ARA antennas. The $dt(D,R)$ values were extracted from ARA data taken during the SPUNK pulser drop at the SPICE core borehole in the 2018-2019 summer season at the South Pole. (**left**) ARA-1 channel 1; (**right**) ARA-2 channel 1. The crosses represent the $dt(D,R)$ values taken from ARA data and the solid circles represent the $dt(D,R)$ values calculated analytically using the default ARA $n(z)$ model with raytracing.

For ARA-1, the $dt(D,R)$ function was therefore split such that:

$$dt(D,R) = \begin{cases} dt(D,R)_{Refracted} & z_{SPUNK} \geq -880 \text{ m} \\ dt(D,R)_{Reflected} & z_{SPUNK} < -880 \text{ m} \end{cases} \tag{H.4}$$

For ARA-2 and the rest of the stations the $dt(D,R)$ function was just:

$$dt(D,R) = dt(D,R)_{Refracted} \tag{H.5}$$

H.1.1.1 $n(z)$ Ice Models and RICE $n(z)$ Discontinuity

Before starting the fitting process, it is essential to try to understand how many layers need to be considered for the South Pole ice $n(z)$ model. There can be multiple $n(z)$ layers at different depths and each layer can have different ABC parameters for the respective $n(z)$ models. To get a sense of the number of $n(z)$ layers required in this fitting exercise, we will study some data taken by RICE (Radio Ice Cherenkov Experiment).

It was observed during a pulser drop test with RICE antennas in 2003 that there seemed to be a

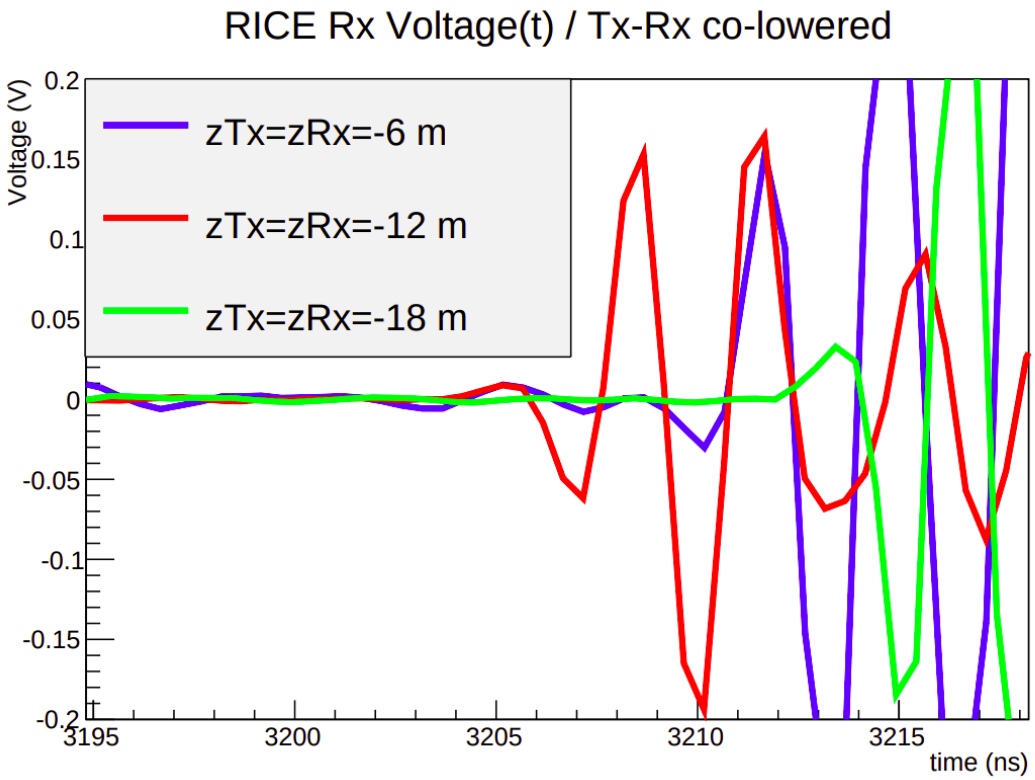


Figure H.3: Signal arrival times of a radio transmitting antenna to a radio receiving antenna which were 30 m apart and were being co-lowered in two different holes in the ice. This pulser drop test was done as a part of ice studies carried out by the RICE (Radio Ice Cherenkov Experiment) experiment (Barwick et al., 2018).

discontinuity in the $n(z)$ profile of South Pole ice (Barwick et al., 2018). The results of the pulser drop can be seen in Figure H.3. A radio pulser antenna and a receiving antenna were co-lowered into the ice and the pulser was made to transmit pulses as it was lowered into the ice. The Figure shows the results for the depths of -6 m, -12 m and -18 m. It should be noted that the radio pulse came later in time for -18 m depth as compared to -12 m depth relative to -6 m depth. This provides an obvious indication that the refractive index does not necessarily increase with depth, and there is probably a discontinuity in the refractive index profile around -12 m depth.

This type of discontinuity tends to show up around the critical density depths where $\rho_c = 550$ kg.m^{-3} and $\rho_c = 830$ kg.m^{-3}. The critical density points in ice represent the depths below which the rate of compactification of ice increases as the snow grains get more densely packed

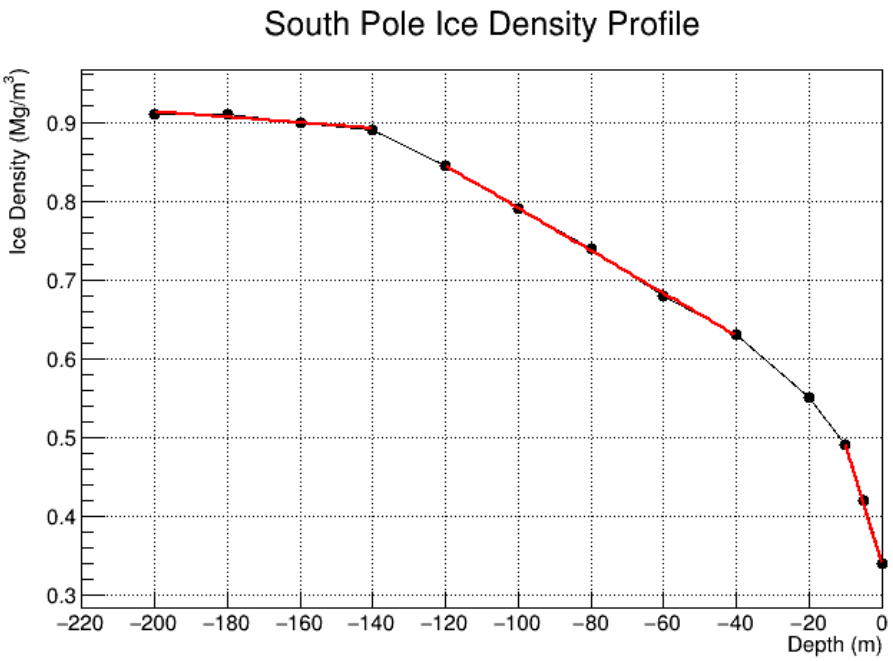


Figure H.4: Density of ice as a function of depth at the South Pole station. The two critical density points at $\rho_c = 550$ kg.m^{-3} and $\rho_c = 830$ kg.m^{-3} can be clearly seen as the gradient of the trend line seems to change twice; once around -10 m and once around -120 m (Koci & Kuivinen, 1983).

with depth. The Greenland ice layer around the Summit Station has its first critical density point, of $\rho_c = 550$ kg.m^{-3}, around a depth of -14.9 m. To calculate the critical density depth for the South Pole, ice density data was taken from the Koci and Kuivinen density data set and then fitted (Koci & Kuivinen, 1983). For the South Pole the first critical density point came out to be around -14.58 m in depth, and that is very close to the depth at which RICE also observed the $n(z)$ discontinuity.

Therefore a bi-layer $n(z)$ model was used for fitting of ARA-1 $dt(D,R)$ data where layer 1 was from 0 to -14.58 m depth and then layer 2 was below -14.58 m depth. This meant that there were two different B and C values for each layer. To ensure there was no sharp discontinuity in the bi-layer model a sigmoid function was used to connect the two layers. The final bi-layer $n(z)$ model was:

$$n(z) = n_1(z) \times (1 - \sigma(z,\mu)) + n_2(z) \times (\sigma(z,\mu)) \tag{H.6}$$

where, for negative z values, the sigmoid function was defined as:

$$\sigma(z,\mu) = \frac{1.0}{(1+\mathrm{e}^{-\frac{-14.58-z}{\mu}})} \tag{H.7}$$

here $n_1(z) = 1.78 + B_1 e^{C_1 z}$ and $n_2(z) = 1.78 + B_2 e^{C_2 z}$ were the refractive index models of the two layers, μ was the slope factor for the sigmoid function and was fixed to a value of 8 for this case.

A single layer $n(z)$ model was used to fit the $dt(D,R)$ data from rest of the ARA stations as the refracted rays in those cases never came to such a shallow depth and never crossed into layer 1 above -14.5 m depth.

H.2 *B* and *C* Fit Results

The fits to the $dt(D,R)$ data trends, as a function of SPUNK depth, obtained from ARA data resulted in values of B and C parameters for the $n(z)$ models for each ARA station. The fitting was done for only the Vpol ARA channels as $dt(D,R)$ data was only available for those channels. All eight Vpol channels were used for ARA-1 and ARA-2; Channels 0, 2, 4, 5 and 6 were used for ARA-3; Channels 0, 1, 4 and 5 were used for ARA-4; Channels 4 and 5 were used for ARA-5. Some channels were excluded from ARA-3, ARA-4 and ARA-5 because of the poor quality of data in those channels. The $dt(D,R)$ data, as a function SPUNK depth, from ARA-1, ARA-2, ARA-3, ARA-4 and ARA-5 was fitted with the $dt(D,R)$ fit functions as described in the Section H.1.1. The bi-layer $n(z)$ function (Equation H.6) and split $dt(D,R)$ function (Equation H.4) were used to fit the $dt(D,R)$ data trends of ARA-1. The single layer $n(z)$ and $dt(D,R)$ (Equation H.5) functions were used to fit the $dt(D,R)$ data trends of ARA-2, ARA-3, ARA-4 and ARA-5.

If the obtained values for B and C fit parameters turn out to be the same for ARA-2, ARA-3, ARA-4 and ARA-5 then it would imply that South Pole ice has the same structure between the SPICE core borehole and these ARA stations and that the same $n(z)$ model can be used for raytracing between these ARA stations and the SPICE core borehole. However, if the values for B and C fit parameters turn out to be different for different stations, then it would imply that there is

an azimuth dependence in the South Pole ice structure and to trace rays between ARA-2, ARA-3, ARA-4 or ARA-5 station and the SPICE core borehole we would need to use the $n(z)$ model for the respective ARA station.

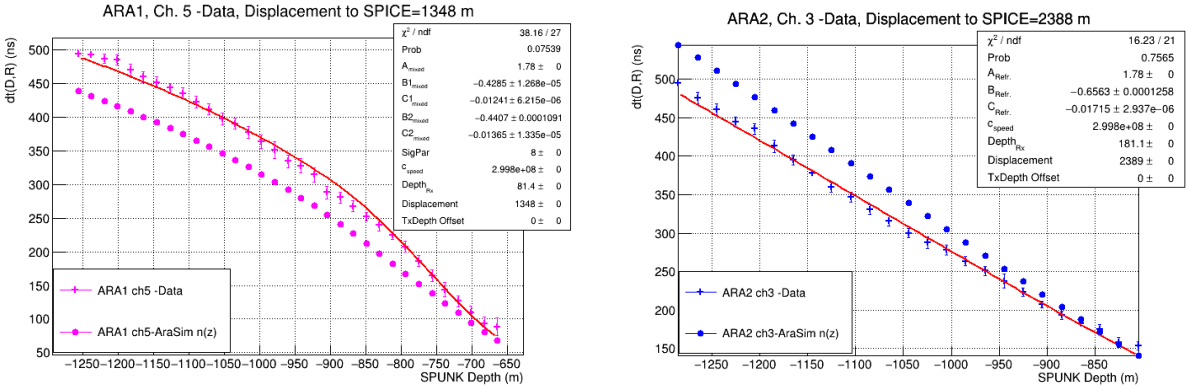


Figure H.5: $dt(D,R)$ (in ns) data vs SPUNK depth (in m) for ARA antennas fitted with the $dt(D,R)$ functions described in Section H.1.1. (**left**) ARA-1 channel 5 and (**right**) ARA-2 channel 3. The crosses represent the $dt(D,R)$ values taken from data and the solid circles represent the $dt(D,R)$ values calculated analytically using the default ARA South Pole ice $n(z)$ model for raytracing. The red line shows the fit to the data using the new $dt(D,R)$ functions.

Figure H.5 shows fit results for channel 5 from the ARA-1 station and for channel 3 from ARA-2 station. The goodness of fit is indicated by χ^2/NDF values of around 1 for both the fits of ARA-1 channel 5 and ARA-2 channel 3. The fit results for the rest of the channels of ARA-1 and ARA-2 similarly were also very promising and can be seen at the end of this chapter in Section H.4.

The average B and C fit values with their standard errors that were obtained from fitting for ARA-1 channels were $B_1 = -0.43337 \pm 0.00244$, $C_1 = -0.01277 \pm 0.00011$ m^{-1}, $B_2. = -0.438887 \pm 0.00430$ and $C_2. = -0.01343 \pm 0.00008$ m^{-1} and for ARA-2 channels were $B = -0.63915 \pm 0.01440$ and $C = -0.01703 \pm 0.00024$ m^{-1}.

The B and C fit values for ARA-1 and ARA-2 were considerably different from each other. One reason for this could be that ARA-1 data has not been properly calibrated for time and voltage. Currently, ARA-1 uses ARA-3 calibration files. Miscalibration in timing can adversely affect the time intervals between points in a waveform and therefore can cause systematic errors in the

$dt(D,R)$ values extracted from the waveforms in ARA data. As for the ARA-3, ARA-4 and ARA-5 stations, the fits were of poor quality. The $dt(D,R)$ functions here seemed to be extremely sensitive to the *Distance* parameter; if the *Distance* parameter was allowed to vary, then the fit would converge on huge (unreasonable) corrections to the already-known *Distance* values. Conversely, if the *Distance* parameter was fixed, the fit would return non-physical values of B and C parameters. The distance measurements between SPICE core borehole and ARA station channels have an uncertainty of around ± 0.22 m[1]. It should also be noted that ARA-4, like ARA-1, also has uncalibrated data and uses calibration files of ARA-3. To counter the problem of not having $dt(D,R)$ functions for ARA-3, ARA-4 and ARA-5 stations, the ARA-2 single exponent $n(z)$ model was applied on all of them. The performance of the ARA-2 single exponent model for all three problematic stations will be discussed in Section H.3.

H.3 Comparing the Performance of Different $n(z)$ Models

After analysing the fit results of B and C parameters, in Section H.2, obtained from fitting the $dt(D,R)$ ARA data that was taken during the SPUNK pulser drop, two new refractive index models for South Pole ice were discovered in addition to the AraSim model:

1. AraSim $n(z)$ (ARA default): $1.78 - 0.43e^{0.0132z}$
2. ARA-2 Deep Ice $n(z)$: $1.78 - 0.63915e^{0.01703z}$
3. ARA-1 Double Exponent $n(z)$: $(1.78 - 0.43337e^{0.01277z}) \times (1 - \sigma(z,8))$
   $+ (1.78 - 0.438887e^{0.01343z}) \times \sigma(z,8)$

The three $n(z)$ models were then used to calculate the $dt(D,R)$ values using raytracing, as a function of SPUNK depth, for all Vpol channels of each ARA station. χ^2 values were calculated in order to quantify how closely each $n(z)$ model was able to recreate the ARA $dt(D,R)$ data values, as shown by Equation H.8, for each Vpol channel of each ARA station for each $n(z)$ model. The

[1]The Eastings and Northings of the ARA stations and SPICE borehole have been measured to within ± 5 ft by the ASC (Antartic Support Contractor) surveyors.

$dt(D,R)$ trends for each channel, as a function of SPUNK depth, were broken up into two regions: -400 m $> z > z_{lin}$ (Region A) and $z_{lin} > z >$ 1500 m (Region B). Here z_{lin} was defined as the SPUNK depth at which $dt(D,R)$ time values, extracted from ARA data taken during the SPUNK pulser drop, would start to plateau off horizontally and become constant. This was a result of how well the algorithm, designed to extract $dt(D,R)$ values from the ARA data taken during the SPUNK pulser drop, worked. It could only extract times to a minimum a threshold time value and was limited by its ability to resolve double pulses in a channel waveform. The value z_{lin} depth for each station was determined by doing visual inspection of $dt(D,R)$ data trends, as a function of SPUNK depth, obtained from the SPUNK pulser drop ARA data. Some of the $dt(D,R)$ data trends for ARA-2 have been shown in Figure H.1. Therefore the χ^2 values for each channel of each station for a given $n(z)$ model were given by:

$$\begin{aligned} \chi^2_A &= \sum_i \left(\frac{dt(D,R)_{Data,i} - dt(D,R)_{Model,i}}{\delta dt(D,R)_{Data,i}} \right)^2 \quad \text{for} \quad -400 \text{ m} > z > z_{lin} \\ \chi^2_B &= \sum_i \left(\frac{dt(D,R)_{Data,i} - dt(D,R)_{Model,i}}{\delta dt(D,R)_{Data,i}} \right)^2 \quad \text{for} \quad z_{lin} > z > -1500 \text{ m} \end{aligned} \tag{H.8}$$

here χ^2_A are the χ^2 values for Region A, χ^2_B are the χ^2 values for Region B, $dt(D,R)_{Data}$ values were extracted from ARA data, taken during the SPUNK pulser drop, for each channel, $dt(D,R)_{Model}$ values were calculated by raytracing using the South Pole ice $n(z)$ model in question for each channel, $\delta dt(D,R)_{Data}$ are the errors on the $dt(D,R)_{Data}$ values and z_{lin} was the SPUNK depth at which the $dt(D,R)_{Data}$ times values became almost constant (i.e., the double pulses became almost indistinguishable in data) and therefore $dt(D,R)$ times stopped varying significantly, as a function of SPUNK depth, in ARA data. To calculate the χ^2 values for Region A and Region B the sum was performed over all the $dt(D,R)$ points available in that SPUNK pulser depth range in the respective region. $\delta dt(D,R)_{Data}$ values were derived from the algorithm that extracted the $dt(D,R)$ values from ARA data. $dt(D,R)_{Data}$ values having no errors were automatically assigned a $\delta dt(D,R)_{Data}$ value of ± 1 ns in order to prevent the χ^2 values from blowing up. For a given $n(z)$ model the χ^2 value for a station was obtained by summing the χ^2 values of all eight Vpol channels for that

station.
Table H.1: χ^2 values for Region A and Region B for all the stations for all the three South Pole ice $n(z)$ models.

Station	$n(z)$ model	χ^2_A	χ^2_B	z_{lin} (m)
ARA-1	AraSim	18550.1	15718.1	-725
	ARA-2 Deep Ice	1535.67	208664	
	ARA-1 Dbl. Exp.	18554.8	7563.01	
ARA-2	AraSim	5939.91	13651.2	-700
	ARA-2 Deep Ice	2811.72	1791.9	
	ARA-1 Dbl. Exp.	5073.61	14470	
ARA-3	AraSim	28921.6	35555.4	-850
	ARA-2 Deep Ice	15867.6	18617.5	
	ARA-1 Dbl. Exp.	26093.5	31091.5	
ARA-4	AraSim	1511.6	1669.11	-900
	ARA-2 Deep Ice	1663.24	704.48	
	ARA-1 Dbl. Exp.	1360.51	1401.34	
ARA-5	AraSim	279.309	13762.3	-950
	ARA-2 Deep Ice	490.632	6839.91	
	ARA-1 Dbl. Exp.	297.34	12151.5	

Table H.1 gives the χ^2 values for Region A and Region B for all the stations for all the three South Pole $n(z)$ ice models. The $dt(D,R)$ time extraction algorithm worked perfectly in Region B, which covered around 90% of the ARA $dt(D,R)$ data, taken during the SPUNK pulser drop. Therefore, the χ^2 values for Region B were the relevant values in order to determine the level of agreement between the $dt(D,R)$ predictions from each of the $n(z)$ models and ARA $dt(D,R)$ data. It can be inferred from the χ^2_B values, in Table H.1, regarding the accuracy of each of the three South Pole ice $n(z)$ models that:

1. ARA-1 Double Exponent model and the ARA-2 Deep Ice model were the most accurate $n(z)$ models, specifically for their respective stations. This is indicated by the relatively low χ^2_B

values as compared to the performance of the other two $n(z)$ models for ARA-1 and ARA-2.

2. ARA-2 Deep Ice model was the preferred $n(z)$ model for ARA-3, ARA-4 and ARA-5 stations and seems to offer the highest degree of agreement between the $dt(D,R)$ data values and the $n(z)$ model predicted $dt(D,R)$ values.

3. The AraSim model was the third-best for most of the ARA stations as compared to the other two $n(z)$ models.

After studying the χ^2_B results, it seems that the ARA-2 deep ice model performs the best for all the stations except ARA-1, in predicting $dt(D,R)$ times between SPICE core borehole and ARA stations. The ARA-1 Double Exponent model does perform slightly better than the AraSim model for all the stations and is generally second best to ARA-2 Deep Ice model for all the stations except ARA-1. However, if a single model were to be used for all raytracing required in neutrino shower simulations and UHECR shower simulations for all the ARA stations, the ARA-2 Deep Ice model would be the obvious choice. Since ARA-1 is not yet correctly calibrated, it seems unreasonable to have a model that works well for an uncalibrated station but not for the calibrated stations; therefore, the ARA-1 model is not the chosen model.

H.4 $n(z)$ Fit Results for All Channels of ARA-1 and ARA-2

Fit results for B and C parameters of South Pole ice $n(z)$ model for $dt(D,R)$ data, as a function of SPUNK depth, for all channels of ARA-1 and ARA-2 have been reported in this section. The data was taken with ARA stations during the SPUNK pulser drop in the 2018-2019 austral summer season. The fit results for all eight Vpol channels for ARA-1 are shown in Figure H.6 and for ARA-2 they are shown in Figure H.7. The fit methodology for doing these fits was discussed in Section H.1.1.1.

The precision of the fit results, for all the channels from both the stations, is indicated by the fact that all the χ^2/NDF are around 1 for the $dt(D,R)$ curves and on visual inspection, the fits are also smooth. However, there are two ARA-1 channels, 1 and 2, which show small kinks on the fit lines around the $z_{Transition}$ depth. This is because $z_{Transition}$ is not precisely at -880 m for these two channels and is slightly off from that value. The fits for all channels were sensitive to the $z_{Transition}$ value, and adjusting the $z_{Transition}$ value would make the fits worse for other channels; -880 m was chosen as a compromise value for the fits for all the channels of ARA-1.

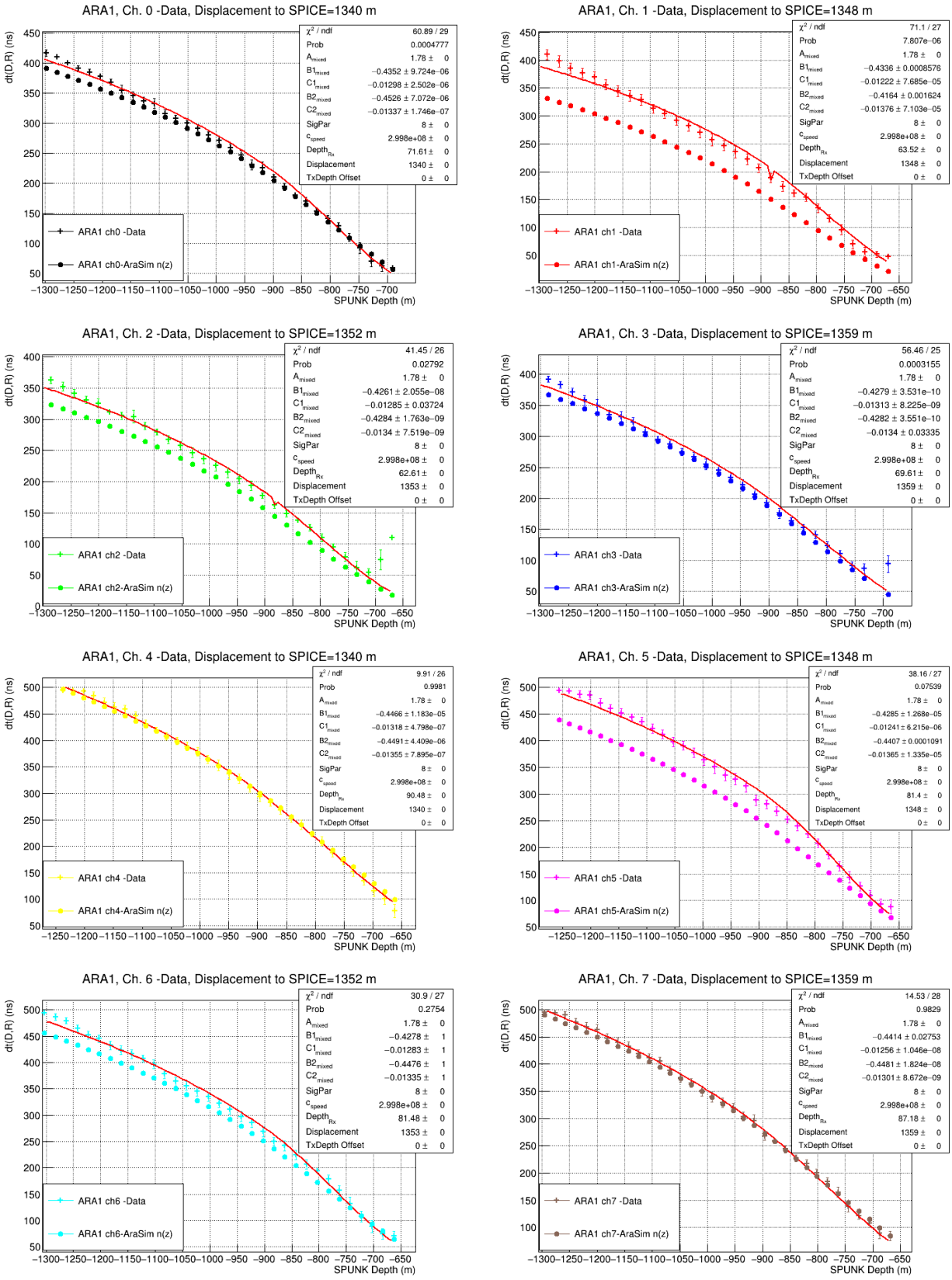


Figure H.6: Fit results for the $dt(D,R)$ vs SPUNK depth (m) for all eight of ARA-1's Vpol channels.

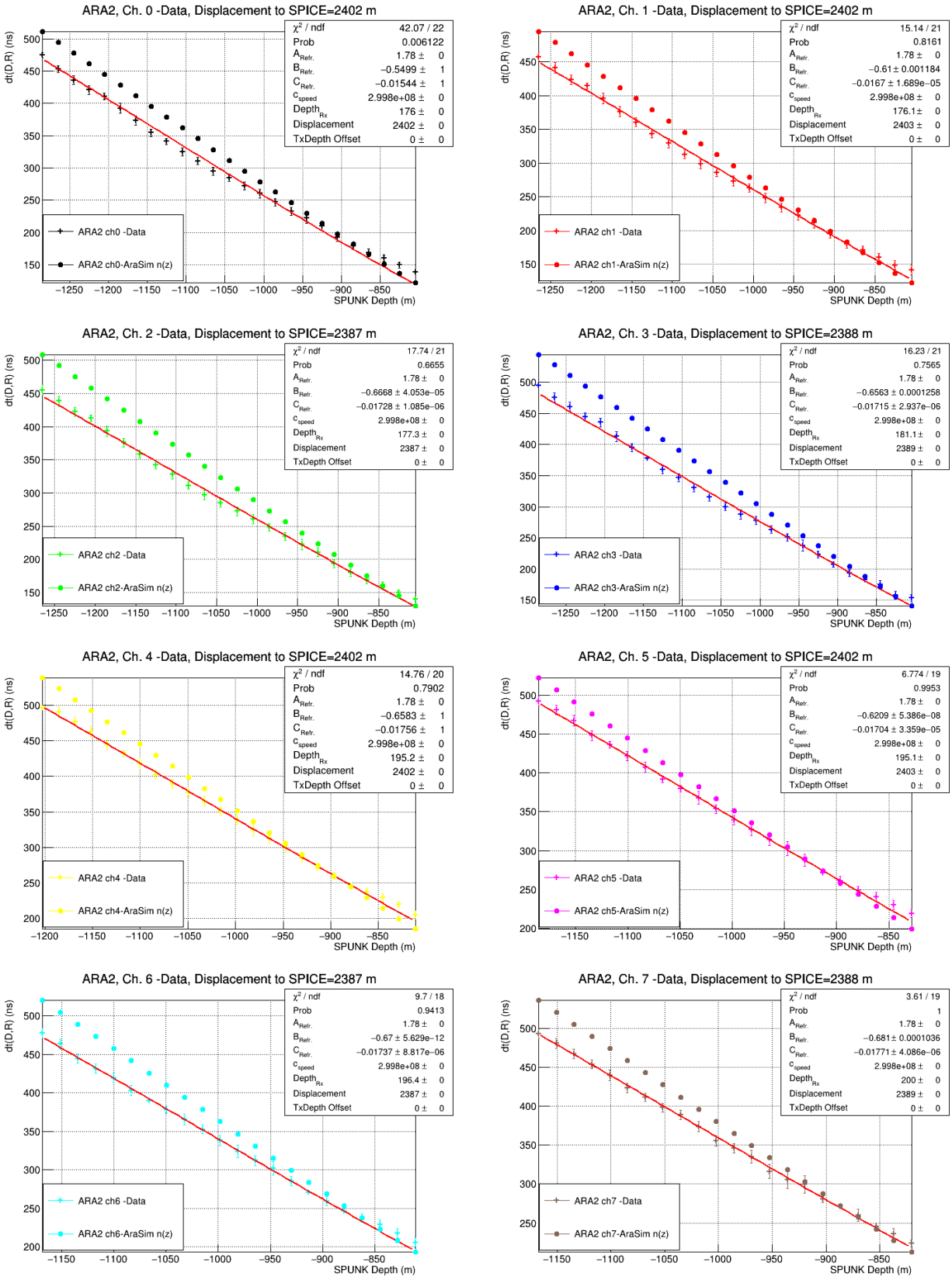


Figure H.7: Fit results for the $dt(D,R)$ vs SPUNK depth (m) for all eight of ARA-2's Vpol channels.