Optimization of a Search for Ultra-High Energy Neutrinos in Four Years of Data of ARA Station 2

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Brian A. Clark, M.Sc.

Graduate Program in Physics

The Ohio State University

2019

Dissertation Committee:

Professor Amy Connolly, Advisor

Professor John Beacom

Professor James Beatty

Professor Ezekiel Johnston-Halperin

ABSTRACT

Ultra-high energy neutrinos are key messengers to the distant, high energy universe, as other particles typically used in high energy astronomy—such as cosmic rays and gamma rays—are either deflected or destroyed en-route to earth from distant cosmic accelerators. Detection of these rare neutrinos requires enormous detectors, and the focus of this thesis is on analysis and hardware work completed for a next-generation pathfinder experiment, the Askaryan Radio Array (ARA). First, we present a study of radio emission observed during an X-class solar flare in the ARA prototype instrument. This radiation is the first in ARA to be reconstructed to an extraterrestrial source on an event by event basis, and is a key demonstration of ARA's ability to locate astrophysical objects in the coordinate system of the celestial sphere. Second, we present the design, characterization, and performance of a new signal conditioning module designed for ARA. The module has a small spatial and power footprint, and features a bank of eight digital, tunable step attenuators which increase the dynamic range of the instrument. Finally, we present the optimization of an analysis of four years of data of station A2, as well as lay the groundwork for the extension of that study to A3. We present the design and performance of a new event-level thermal noise rejection algorithm, and for the first time in an ARA analysis, filter events of contaminating man-made noise. Finally, we present a procedure for optimizing cuts to achieve a best expected limit taking into account separate periods of data taking livetime.

Looking forward, as ARA further accumulates livetime, and our analysis efficiency improves, we anticipate being able to set world-leading limits above approximately 10 EeV by 2022. With next generation neutrino telescopes coming online—e.g., RNO, IceCube-Gen2, PUEO, etc.—along with other observatories like aLIGO, LSST, and HAWC, we anticipate an exciting next decade for the field of multimessenger astronomy.

To all the people who never stopped believing in me.

ACKNOWLEDGMENTS

The list of people to thank is surely too long to fit on a few pages, but I will make the effort to acknowledge at least some of the people who have supported me in this endeavor. It takes a village—and the one supporting me has been nothing short of extraordinary.

First and foremost, I would like to thank my advisor Amy Connolly, who has challenged me over the course of the last five years to grow as a scholar, scientist, mentor, and person. Thank you for always supporting my ambitions, for expecting more of me than I ever thought myself capable, and for teaching me so much about doing good physics—always optimize your cuts, don't use jargon, and if it's not on Dropbox, it didn't happen!

To members of the Connolly and Beatty groups, both past and present, thank you for being incredible mentors, peers, and mentees. To research scientist Patrick Allison, thank you for teaching me how to solder, and how to think about the design and operation of hardware. Thanks also for being my "surrogate advisor" in the last weeks leading up to my unblinding request. To postdocs Carl Pfendner, Mike Sutherland, and Jordan Hanson, thank you for always tolerating my many questions. And to Carl, thank you for mentoring me through my first project on `AraSimQC`. To my peers in the Connolly-Beatty group who have already graduated—Brian Dailey, Sam Stafford, Jacob Gordon, and Oindree Banerjee—it was a joy to grow alongside you. Special thanks are due to Oindree Banerjee, for the many runs to Panera and Pita Pit, and always constant willingness to laugh, cry, or scream, sometimes at the same time.

To my more recent colleagues of Jorge Torres, Keith McBride, Andres Medina, and Lauren Ennesser—thank you for being a constant source of energy, which has buoyed me in these last two months. Thanks especially to Keith for always being willing to get coffee and for listening to me rant. And thanks to Jorge for helping me finish all of the necessary limit plots for the diffuse analysis. I am equally grateful to all of the many undergraduate students I have had the opportunity to mentor here at OSU. I learned an enormous amount from you. Special thanks are due to Suren Gourapura, Lucas Smith, and Jason Torok for being the first group of undergrads I ever supervised, and for doing such an awesome job with building ARA stations. To the latest three "ASE" students—Alex Patton, Scott Janse, and Eliot Ferstl—I am appreciative of all your laughter and enthusiasm for flipping through ten-thousand reconstruction plots to identify surface noise.

To my classmates here in the physics building—thank you. I would not have survived coursework without you. I cannot express enough thanks to Humberto Gilmer and Tony Lefeld, who were my steadfast companions through innumerable hours of homework and panic over core course exams. Paulo Montero, it was great to have you join the fold in QFT. I am especially grateful to my former office mate Khalida Hendricks, who was so supportive during my final weeks, and let me come play with her beagles to relax. Khal—congrats on your awesome new job. To my classmate Dylan Roderick and former classmate Kasandara Sullivan, thank you for letting me squat at your kitchen tables and for cooking dinner for me during my final weeks. I am also grateful to Michael Newburger, Kathryn Nicolich, and Chris Cappiello. To my colleagues down on the mezzanine, thank you for always answering my questions about neutrino and dark matter theory, particularly Mauricio Bustamante and Katie Auchettl. Special thanks to Shirley Li for being the first person to sit me down and seriously encourage me to to pursue a postdoc.

I am thankful to my collaborators on the Askaryan Radio Array, and for the opportunity I received working on ARA to go to the South Pole. In particular, I have appreciated the camaraderie of Rishabh Khandelwal, Uzair Latif, and Ming-Yuan Lu. I am thankful to Dave Besson for being a supportive "god-parent" on the solar paper and for writing letters in support of my postdoctoral applications. I am thankful to Abby Vierreg, Eric Oberla, and Simon Archambault for making my trip to Antarctica so much fun. To all of the people who work in Antarctica as part of USAP—thank you for making my research possible.

I am appreciative of my committee members Jim Beatty, John Beacom, and Zeke Johnston-Halperin for their time. Zeke, thank you for agreeing to serve on this committee. I am also grateful to former member Linn Van Woerkom, who has been with me since candidacy. Jim, thanks for always stopping by in the morning with a hearty *Buongiorno!* John, I will always be appreciative of your insightful advice.

To my past mentors and teachers, thank you for supporting me and nurturing my love of physics. In particular, thanks to Erin Mulanax, my high school physics teacher in the Pattonville School District (Go Pirates!), who inspired my love of physics for the first time. I am appreciative of my undergraduate advisors at Washington University in St. Louis: Henric Krawczynski, who introduced me to astrophysics, and James Schilling, who took a chance on a random freshman and sponsored my first summer research project.

I am thankful to The Ohio State University for providing an extraordinary place to be a student. A special thank you is due to program coordinator Kristina Dunlap for all her work in supporting me and my fellow students in the graduate program. I am thankful to all the organizations who have supported my research and education financially, particularly the Department of Physics for first-year fellowship support, and to the Physics Department and Astronomy Department for support as a Teaching Assistant. I am forever indebted to the

National Science Foundation for their support through the Graduate Research Fellowship Program (GRFP) Award DGE-1343012 which allowed me to have flexible and ambitious research goals. Lastly, I am grateful to the organizations which provided infrastructure support to my research— namely the Ohio Supercomputer Center (OSC) and the CCAPP Antarctic Research Test Facility (CART). I am particularly grateful to Bryan Dunlap and Andrew Canale of the Department of Physics Computing office, who were always available at short-notice to help me fix my Linux box or recover from a hard-drive issue on my mac.

Last, but certainly not least, I want to thank my family. To Mom and Dad, thank you for always believing in me, and encouraging me to chase my dreams, even when that involved doing things as crazy as spending a month at the South Pole. The advice to take it one day at a time was indispensable in staying motivated in these last months. To Marie and Brandon, you are the best siblings I could have asked for—thanks for always reminding me of the bigger picture. To my 94 year old grandmother, Marie, you inspire me everyday with your tenacity. I am eternally grateful that you “keep-a-going”, and are going to come to Ohio to see me walk in Commencement.

VITA

August 1992	Born—Creve Ceour, MO USA
May 2014	B.S., Washington University in St. Louis St. Louis, MO USA
June 2016	M.Sc., The Ohio State University Columus, OH USA

Publications

P. Allison *et. al.* for the ARA Collaboration, including B. A. Clark. *Design and Performance of an Interferometric Trigger Array for Radio Detection of High-Energy Neutrinos.* Nuclear Instruments and Methods A Vol 930 Pg 112-125 (2019).

P. Allison *et. al.* for the ARA Collaboration, including B. A. Clark. *Measurement of the real dielectric permittivity ϵ_r of glacial ice.* Astroparticle Physics Vol 108 Pg 63-73 (2017).

F. Kislat , **B. Clark**, M. Bielicke, H. Krawczynski. *Analyzing the Data from X-ray Polarimeters with Stokes Paramters.* Astroparticle Physics Vol 68 Pg 45-51 (2015).

Fields of Study

Major Field: Physics

Studies in:

Ultra-High Energy Neutrino Astrophysics	Prof. Amy Connolly
X-Ray Astronomy	Prof. Henric Krawczynski

Table of Contents

Appendices

List of Figures

List of Tables

Table **Page**

Chapter 1

Introduction

Ultra-high energy neutrinos (> 10 PeV) are unique messengers to the distant, high energy universe. As chargeless and weakly interacting particles, neutrinos offer observational advantages over the other standard model messengers. Hadrons (cosmic rays) above approximately $10^{19.5}$ eV are destroyed by interactions with the Cosmic Microwave Background (CMB) and are also deflected by galactic and intergalactic magnetic fields, meaning they do not point back to their sources and cannot be used for astronomy. Photons (gamma rays) pair annihilate with the CMB and Extragalactic Background Light (EBL) above a few TeV, and can also be stopped by space dust. So neutrinos stand as an appealing messenger; however, detection of these valuable neutrinos is challenging due to low incident fluxes and low cross sections, and experiments designed to detect them therefore require enormous instrumented volumes.

In this chapter, I will summarize the physics targeted by modern high-energy neutrino observatories. I will then discuss the experimental signature of high-energy neutrinos, which are the optical and radio-Cherenkov effects. Finally, I will survey the current experimental landscape, including the IceCube, Askaryan Radio Array (ARA), Antarctic Impulsive Transient Antenna (ANITA), and Antarctic Ross Ice-Shelf Antenna Neutrino Array (ARIANNA) detectors.

1.1 Target Physics

The goals of ultra-high energy neutrino detectors are broad. By measuring *cosmogenic* neutrinos, which are byproducts of the aforementioned interactions between cosmic rays and the CMB, we will learn about the nature of the cosmic-ray flux at high energies and from high redshifts ($z > 1$). By measuring the *astrophysical* flux of neutrinos, we will study the nature of astrophysical sources that are capable of accelerating particles to very high energies.

1.1.1 Astrophysical Neutrinos

As can be seen in Fig. 1.1, the existence of a flux of extragalactic astrophysical neutrinos has been confirmed by the IceCube instrument [1] up to 6 PeV, with a spectral index ranging from $E^{-2.2-2.5}$, depending on whether the measurement is made with through-going muon tracks [2] which generates a harder spectral index, or a joint fit to both the through-going tracks and the cascade-like events [3] which results in a softer spectral index. However, no neutrino sources have been definitively identified, nor have any neutrinos been observed above 10 PeV. An enticing hint was delivered in 2018 by the observation of a high-energy neutrino in IceCube which was spatially coincident with the blazar TXS0506+056, that was found by HESS, *Fermi*, and others to be flaring in gamma-rays [4, 5].

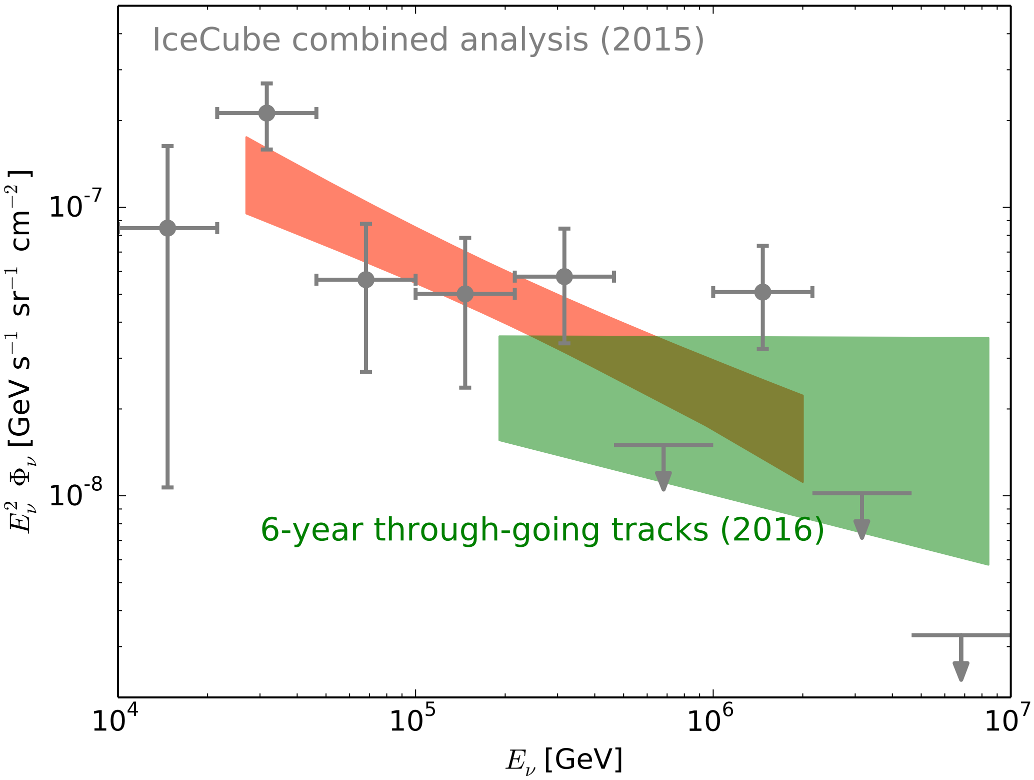


Figure 1.1: The measured astrophysical neutrino spectrum from IceCube, with the all-channel measurement in gray/pink [2] and the measurement from through-going muons in green [3]. Figure from Aartsen *et. al.* 2018 [6].

This flux of high-energy neutrinos at and above the PeV scale, termed the *astrophys-*

ical neutrino flux, is expected from the high energy accelerators in the universe, such as Active Galactic Nuclei (AGN), Gamma Ray Bursts (GRBs), and others. The observation of neutrinos from a source provides unambiguous proof that hadronic particles are being accelerated; this is relevant for identifying whether the most energetic transients in the universe, like gamma ray bursts (GRBs) [7], are hadronic or leptonic in origin. Further, there is a strong multimessenger connection between neutrinos and gamma-rays. This is because a high-energy accelerator will produce both charged pions and neutral pions; charged pions decay to neutrinos through the reaction in Eq. 1.1 (and its charge conjugate)

$$\begin{aligned} \pi^+ &\to \mu^+ + \nu_\mu \\ &\to \mu^+ + e^+ + \nu_e + \bar{\nu}_\mu \end{aligned} \tag{1.1}$$

meanwhile, neutral pions decay through the reaction in Eq. 1.2

$$\pi^0 \to \gamma + \gamma \tag{1.2}$$

As such, joint measurements of both gamma rays and neutrinos should constrain neutrino source models, as has been demonstrated by Murase *et. al.* and others [8, 9].

Measuring the flux of high-energy neutrinos above 10 PeV will test whether the population of accelerators currently being observed by IceCube continue to higher energies, or if nature switches to a new, more powerful class of astrophysical objects. Identifications of the first neutrino sources—particularly in a multimessenger and multi-wavelength fashion in the spirit of TXS0506+056 or the binary-neutron star merger observed by LIGO [10]—will provide unprecedented information about the accelerators themselves.

1.1.2 Cosmogenic Neutrinos

A flux of ultra-high energy neutrinos above 100 PeV, termed the *cosmogenic neutrino flux*, is expected from the interaction of ultra-high energy cosmic rays (UHECR) with the the CMB. The destruction of UHECR through interactions with the CMB was first suggested by Greisen, Zatsepin, and Kuzmin [11, 12]. In the simple case of a cosmic-ray proton interacting with a CMB photon, for example $p+\gamma \to n+\pi^+$, the CMB photon would have energy $E_\gamma \approx 3 \times 10^{-4}$ eV and the cosmic ray would require $E_p \approx 10^{19.5}$ eV. A suppression of the UHECR flux is observed at this energy [13] and can be seen as a flux suppression in the cosmic ray spectrum in Fig. 1.2, lending support to the existence of this so-called "GZK effect". Cosmic ray primaries heavier than protons (such as helium, iron, etc.) are also attenuated by the GZK effect, but as composite particles, are also susceptible to photodisintegration, e.g. the giant dipole resonance [14]. Note that because the mean-free path of a UHECR in the CMB is approximately 50 Mpc, cosmic rays do not arrive at earth from sources with a redshift higher than $z \sim 1$, creating a effective "GZK horizon" for cosmic ray observations.

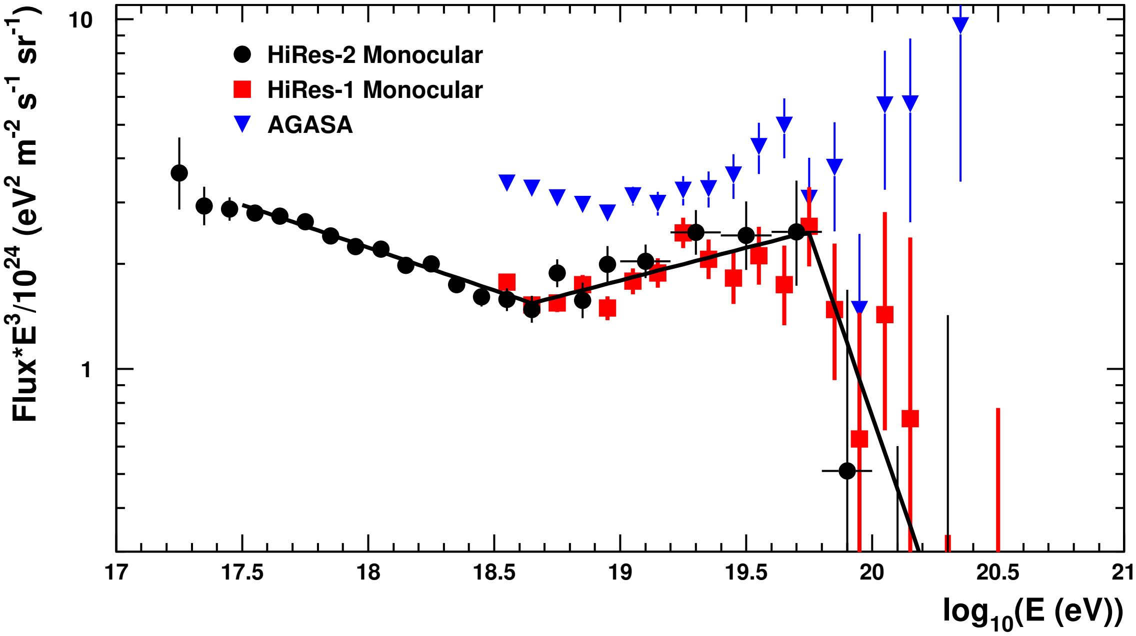


Figure 1.2: The observation of the GZK flux suppression by the HiRes Collaboration. The GZK flux suppression is interpreted as the cause of the sudden change in spectral index at approximately $10^{19.5}$ eV. Figure from Abbasi *et. al.* 2007 [13].

Fortunately, these pions that result from the GZK interaction will later decay into a flux of neutrinos: the pion decays into a μ^+ and ν_μ, and the anti-muon further through $\mu^+ \rightarrow e^+ + \nu_e + \bar{\nu}_\mu$. This flux of neutrinos from the GZK interaction comprises the Berezinsky-Zatsepin (BZ) flux [15]. The nature of the accelerators which are injecting the source cosmic rays—specifically their elemental composition, maximum acceleration energies, spectral features, and red-shift distributions—will imprint on the BZ flux of neutrinos [16–18]. As such, measuring cosmogenic neutrinos will provide direct constraints on the cosmic ray accelerators, and will be a unique window on the nature and distribution of UHECR accelerators at high redshift, where the cosmic rays and gamma-rays cannot be observed directly.

These cosmogenic neutrinos are expected to arrive at Earth with an incident flux of $\sim 10\ \nu/\text{km}^2/\text{yr/sr}$, but the low cross-section of $\sigma \sim 10^{-32}$ cm^2 gives an interaction rate in ice ($\rho = 0.9$ g/cm^3) of $< 3 \times 10^{-3}\ \nu\text{km}^3/\text{yr/sr}$, and so detectors with effective volumes of order 100 km^3 are required to make regular detections.

1.2 Experimental Signature

High-energy neutrinos are never detected directly; rather, they are detected by observation of daughter particles produced by neutrino-matter interactions. Fig. 1.3 shows the Feynman diagrams for the two main deep-inelastic scattering (DIS) channels. On the left is the "charged-current" (CC) $W^{\pm}$-vector-boson mediated interaction, and on the right is the "neutral-current" (NC) Z^0-vector-boson mediated interactions. In the case of a CC-interaction, $\nu_\ell + N \to \ell + X$, a neutrino ν_ℓ exchanges a $W^{\pm}$ with a nucleon N, producing a daughter lepton ℓ and an electromagnetic shower X. In the case of a NC-interaction, $\nu + N \to \nu + N^*$, a neutrino ν exchanges a Z^0 with a nucleon N, exciting the nucleon which spallates an induces a hadronic shower. Additionally, in the case of a CC-interaction, the daughter lepton can further decay and produce sub-showers.

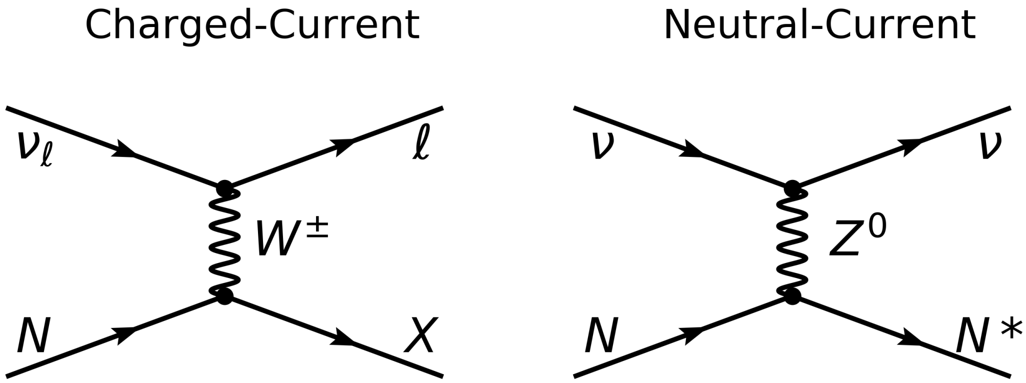


Figure 1.3: Feynman diagrams of deep-inelastic neutrino-nucleon scattering, with the charged-current interaction on the left and the neutral-current interaction on the right.

These electromagnetic or hadronic showers are moving relativistically ($\beta = v/c \approx 1$ where v is the speed of the shower and c is the speed of light). As such, if these relativistic neutrino-induced showers occur in a dense dielectric medium with an index of refraction greater than unity ($n > 1$), the showers will be moving faster than the phase velocity of light in the medium, and will emit Cherenkov radiation [19]. This optical-Cherenkov radiation is dominated by blue and near-UV light, and is emitted preferentially in a cone with an opening angle relative to the shower axis of θ_C, where $\cos(\theta_C) = 1/n\beta$ depends on the index-of-refraction n and relativistic factor β. The opening angle θ_C is termed the Cherenkov angle. If the dielectric medium is transparent to electromagnetic radiation—such as ice, salt, water, etc.—then this optical-Cherenkov radiation can be observed by a light sensor such as a photo-multiplier tube. This detection of neutrinos through the optical-Cherenkov emission was first suggested in the 1960's by Markov [20], and is used in a number of modern neutrino detectors, including Super-Kamiokande [21] and IceCube [22].

Additionally, these showers also generate radio-Cherenkov emission, interchangeably referred to as Askaryan radiation. This radiation occurs because these showers obtain a net negative charge asymmetry of $\sim 10-20\%$ due to the destruction of positrons in the shower and electrons in the medium, as well as the creation of electrons through Compton scattering and the photoelectric effect. This relativistically moving charge distribution radiates, and for wavelengths approximately the size of the longitudinal width of the shower, the radiation adds coherently. The transverse size of the shower will be $\mathcal{O}(10)$ cm in size, and the radiation will be a broadband (10-1200 MHz), fast ($\mathcal{O}(100)$ ps) radio pulse [23, 24]. This was first suggested by Askaryan, for whom the radiation is named. The existence of Askaryan radiation in both sand and ice has been confirmed by firing bunches of high-energy electrons into fixed targets at the Stanford Linear Accelerator (SLAC) [25, 26].

The Askaryan signal is strongest on the Cherenkov cone and the radiation is polarized perpendicular to the cone along the line-of-sight to the shower. The Askaryan pulse broadens in time, loses its high frequency content, and decreases in amplitude as the observer moves further "off" the Cherenkov cone. Quantitatively, the parameterization by Alvarez-Muniz *et. al.* predicts that between 10-1200 MHz, the electric field ($E(f)$) at a given frequency (f) is approximately inversely proportional to the distance to the interaction (R), dependent in a Gaussian fashion on the angle from the Cherenkov cone ($\Omega = \theta - \theta_C$), and linearly proportional to the neutrino induced shower energy (yE_ν), where y is the inelasticity and governs what fraction of the neutrino energy E_ν is deposited in the shower as shown in Eq. 1.3. This can also be seen graphically in Fig. 1.4.

$$E(f) \propto \frac{yE_\nu}{R} \times \frac{f}{1150\text{MHz}} \times \exp\left[-\frac{1}{2}\left(\frac{f}{1\text{ GHz}} \times \frac{\Omega}{2.2^\circ}\right)^2\right] \tag{1.3}$$

The Askaryan emission is promising because unlike optical light, which has scattering lengths of ~ 20 m and extinction lengths of ~ 100m [28], Antarctic and Arctic ice is measured to have attenuation lengths on the order of a several hundred meters to a kilometer [29–32]. As such, an instrument designed to measure the radio-Cherenkov effect can be deployed roughly an order of magnitude more sparsely than an instrument designed to detect optical-Cherenkov light, allowing for larger volumes to be instrumented.

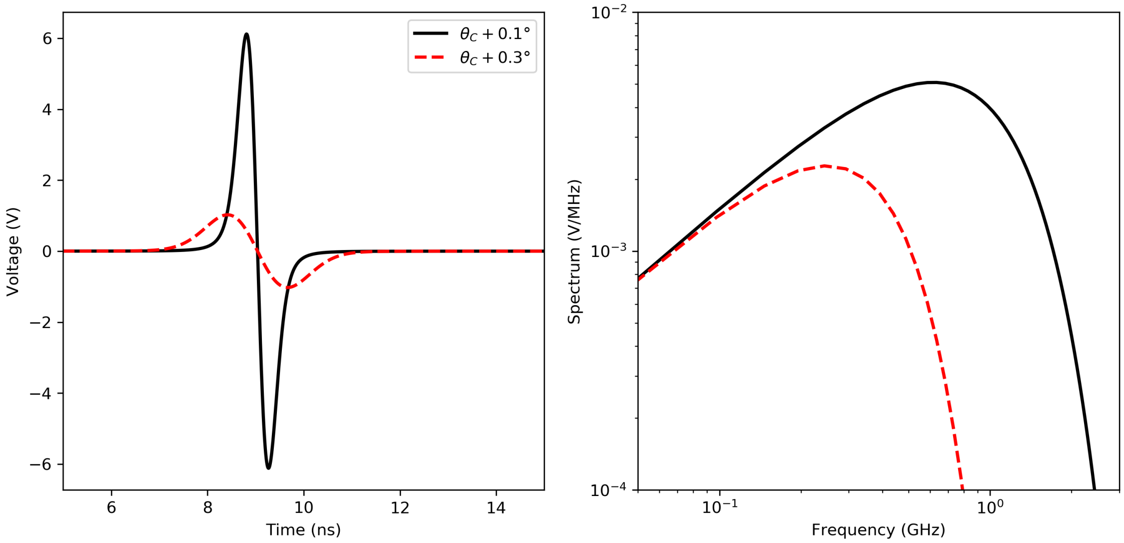


Figure 1.4: A model by Alvarez-Muniz *et. al.* [27] for the time-domain waveform (left) and frequency-domain spectra (right) of Askaryan radiation produced by a 10^{20} eV neutrino inducing an electromagnetic shower viewed from 1 m from the interaction location. I show the predicted results for two different choices of viewing angle.

1.3 The Current Experimental Landscape

To provide an overview of the current experimental landscape, in Fig. 1.5 I show the current leading limits on the diffuse flux of ultra-high energy neutrinos, along with several cosmogenic neutrino models. Along the y-axis, I plot the flux of the number of particles, and on the x-axis, I show the logarithm of the neutrino energy in GeV. As can be seen by the number of experimental efforts, the last two decades have seen an enormous interest in the detection of neutrinos across a broad energy range from a few TeV up to and above a ZeV. At energies below $10^{19.5}$ eV, experiments searching for the optical signatures from neutrino-nucleon showers have dominated the field. Above $10^{19.5}$ eV, the most stringent limits arise from experiments looking for the radio signatures from neutrino-nucleon showers.

1.3.1 Optical Cherenkov Experiments: Neutrinos up to 10 EeV

As mentioned above, looking for the optical-Cherenkov radiation from neutrino-nucleon showers, and their daughter particles, has been an incredibly successful method of measuring neutrinos up to 10 PeV, and for setting limits up to $10^{19.5}$ eV. The experiments work in a relatively straightforward fashion and rely on proven, decades old technology: arrays of photomultiplier tubes are deployed in an optically transparent medium. When a neutrino-induced shower, or one of its daughter leptons, is emitting Cherenkov radiation and traverses

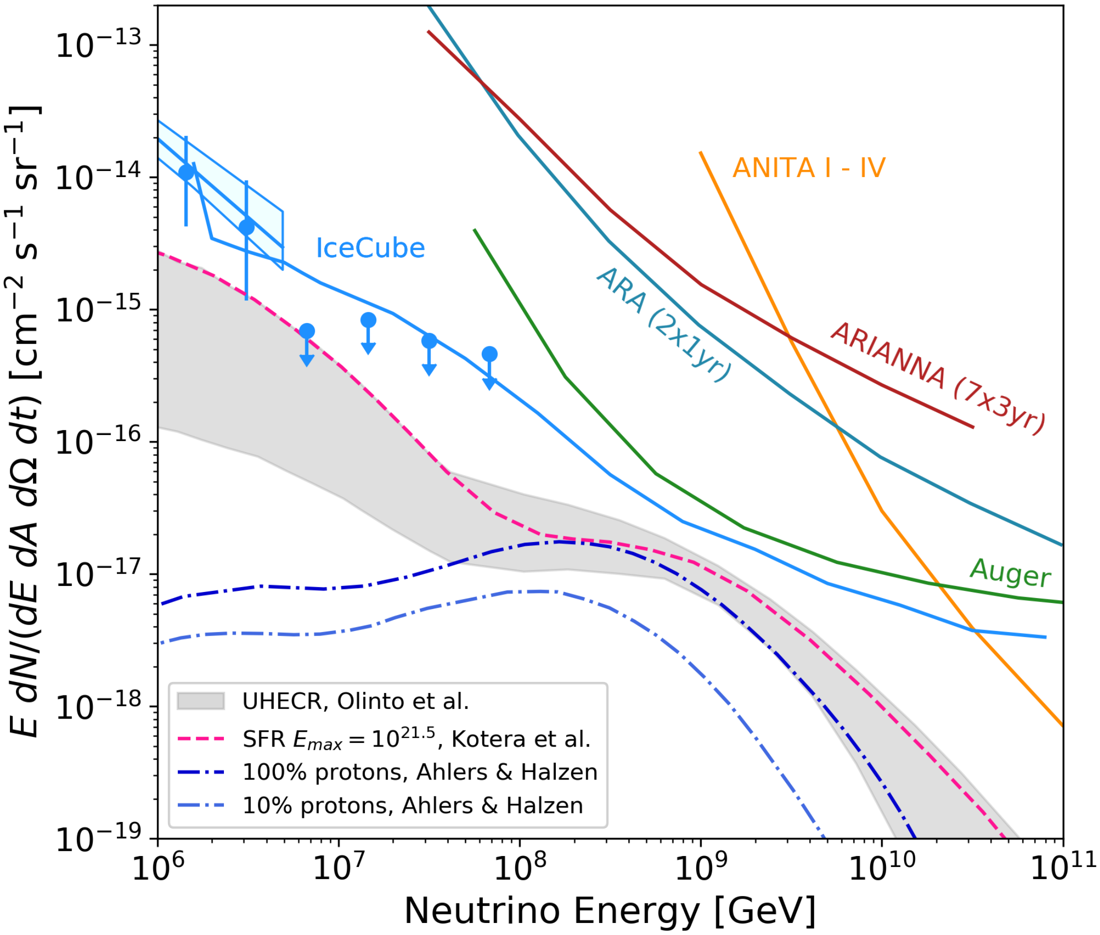
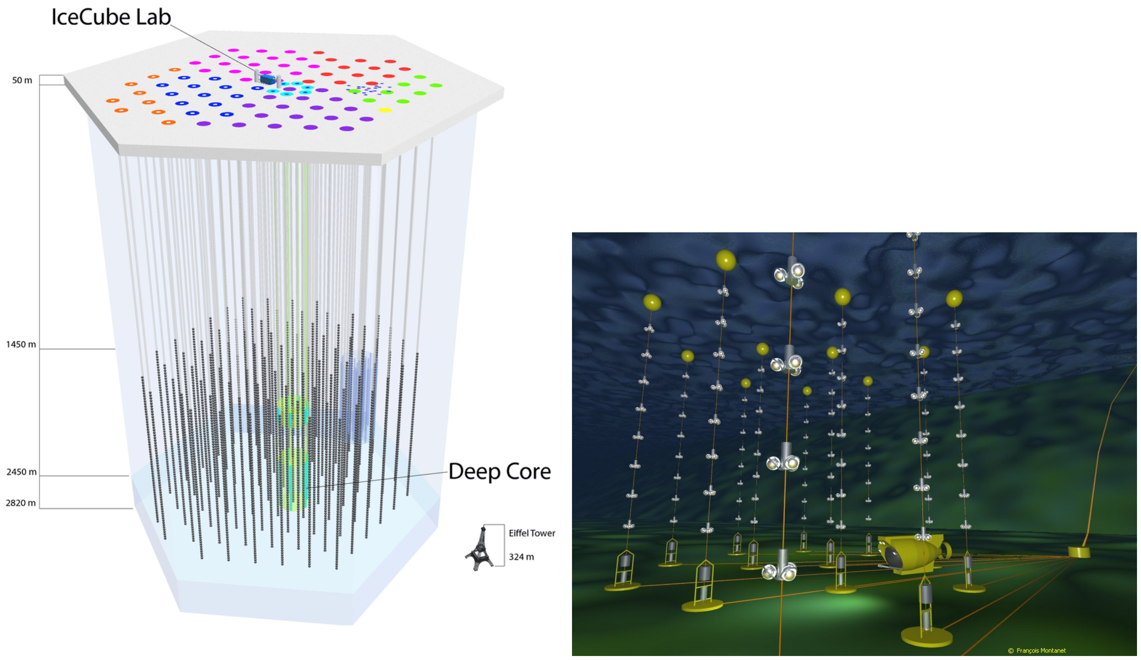


Figure 1.5: The current leading limits from existing ultra-high energy neutrino detectors along with theoretical predictions for the cosmogenic flux of neutrinos. The data is the measured astrophysical neutrino flux from IceCube [2], and the limits are from IceCube [33], Auger [34], ANITA [35], ARA [36], and ARIANNA [37]. The dashed lines and shaded bands are theoretical models from Olinto *et. al.* [16], Kotera *et. al.* [17], and Ahlers and Halzen [18].

near the photomultiplier tube, a light signal is recorded. The quantity of light, and its pattern as recorded in the array, are then analyzed to reconstruct and constrain the incoming neutrino's direction, energy, and flavor.

Chief among current experiments leveraging this technique are the IceCube [22] and ANTARES instruments [38], as depicted in Fig. 1.6. For IceCube, this is a cubic-kilometer of ice at the South Pole in Antarctica. IceCube has deployed 86 strings, each with 60 photomultiplier tubes enclosed in Digital Optical Modules (DOMs), and was completed in 2011. For ANTARES, the medium is sea water in the Mediterranean Sea off the coast of

Toulon, France. ANTARES has deployed twelve strings of photomultiplier tubes, each with 75 optical modules, and was completed in 2008. The IceCube and ANTARES experiments are complimentary in their field of view; deployed at the South Pole, IceCube has the best sensitivity to high-energy neutrinos coming from the Northern sky, where the Earth serves as a shield to atmospheric muons and neutrinos. In contrast, by being deployed in the Mediterranean Sea, ANTARES preferentially observes the Southern sky.



(a) Schematic of the IceCube experiment.

(b) Schematic of the ANTARES experiment.

Figure 1.6: Schematics of the IceCube and ANTARES instruments from [39] and [40] respectively.

1.3.2 Radio Cherenkov Experiments: Neutrinos above 10 EeV

Detection of neutrinos through the radio-Cherenkov effect focus on a higher energy regime than the aforementioned optical experiments, and currently set world leading limits above $10^{19.5}$ eV. Experiments searching for radio emission can be subdivided into two categories: instruments that seek to observe the radio-transparent medium *remotely*, and those which try to observe the medium *in-situ*. Because the amplitude of a radio wave scales inversely with the distance to the emitter, being *in-situ* allows for the observation of lower-amplitude signals. The trade-off however is that such arrays can observe less medium for a given instrumentation budget, and have correspondingly smaller exposures. We again emphasize that in the design of these radio experiments, there is a constant trade-off between being close to the interaction and having a large instrumented volume.

Of the remote experiments, the Antarctic Impulsive Transient Antenna (ANITA) experiment plays a leading role. ANITA seeks to observe Askaryan radiation by flying an array of dual-polarized antennas above the Antarctic ice-sheet on a NASA Long-Duration Balloon (LDB) dozens of kilometers above the earth surface [41]. The ANITA instrument, along with its various detection channels, are shown schematically in Fig. 1.7. The neutrino detection channel for which ANITA was originally intended manifests as a neutrino interacting in the Antarctic ice, and the Askaryan radiation from the resulting particle cascade refracting out of the ice surface and arriving at the payload. ANITA has flown four campaigns and currently sets the most stringent limits on the diffuse flux of neutrinos between $10^{21} - 10^{23}$ eV [35] using this channel. As can also be seen in the diagram, ANITA has several other detection channels; one exciting, but yet unrealized channel, is the observation of a upgoing tau neutrino [42].

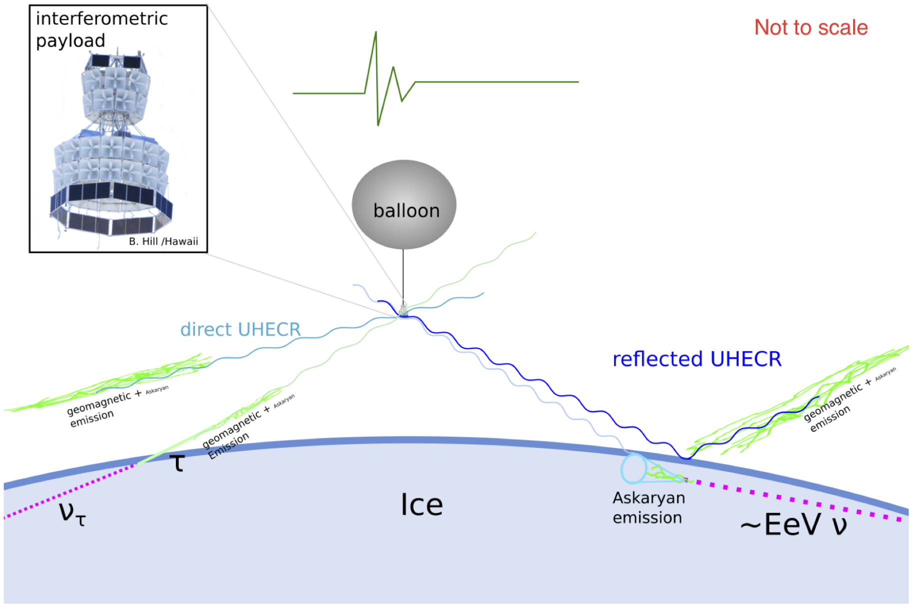


Figure 1.7: A schematic of the ANITA detection concept, highlighting all of the various ways in which ANITA could observe a physics signal–namely, Askaryan emission from neutrinos, direct and reflected geomagnetic emission from cosmic-ray air showers, and potentially radio emission from earth-penetrating tau neutrinos. Figure by Cosmin Deaconu [43].

A diagram showing the lunar detection concept is in Fig. 1.8, which is utilized by the NuMoon experiment to search for ultra-high energy neutrinos [44]. The NuMoon project currently holds the best limit above 10^{23} eV. The extremely high energy threshold arises because of the large distance between the radio telescope and the neutrino interactions on the lunar surface. Access to the large lunar regolith allows for a competitive limit however.

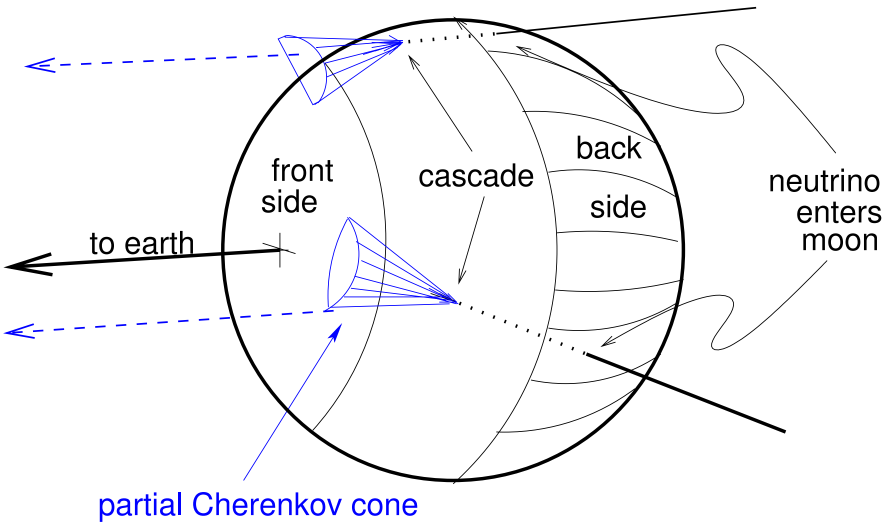


Figure 1.8: A schematic of the interaction of a neutrino with the lunar surface. The radio waves produced could then be observed at earth. Figure from [45].

By deploying the radio antennas inside the ice itself, the energy threshold of the experiment can be lowered nearly three orders of magnitude relative to the ANITA experiment; the trade-off is a reduction in the effective volume of the instrument by essentially the same order of magnitude . Classes of detectors which instrument the ice volume itself are *in-situ* experiments. The first such project to propose burying the antennas in the ice was the Radio Ice Cherenkov Experiment (RICE)[46], which was deployed alongside the IceCube pathfinder AMANDA. Since that time, two successor experiments have been established—the Askaryan Radio Array (ARA) [30] and the Antarctic Ross Ice-Shelf ANtenna Neutrino Array (ARIANNA) [47].

ARA, which is located at the South Pole a few kilometers from the IceCube experiment, is the primary focus of this thesis. ARA consists of a roughly cubical lattice of 16 antennas deployed down boreholes in the Antarctic ice sheet. Each sub-set of antennas, or station, operates autonomously. ARA has deployed a prototype station, the “Testbed”, along with one station at 100m depth (A1), and four stations at 200m depth (A2-5). Each individual

ARA station is denoted A1, A2, etc., while the entire five station array is denoted “ARA5”.

A schematic of an ARA station along with the neutrino detection principle is provided in Fig. 1.9. The neutrino detection channel for ARA consists of a neutrino interacting in the Antarctic ice, and the Askaryan radiation traveling to the array to be observed and recorded. ARA, being located at the South Pole and having access to the South Pole power grid, also has a 100% uptime.

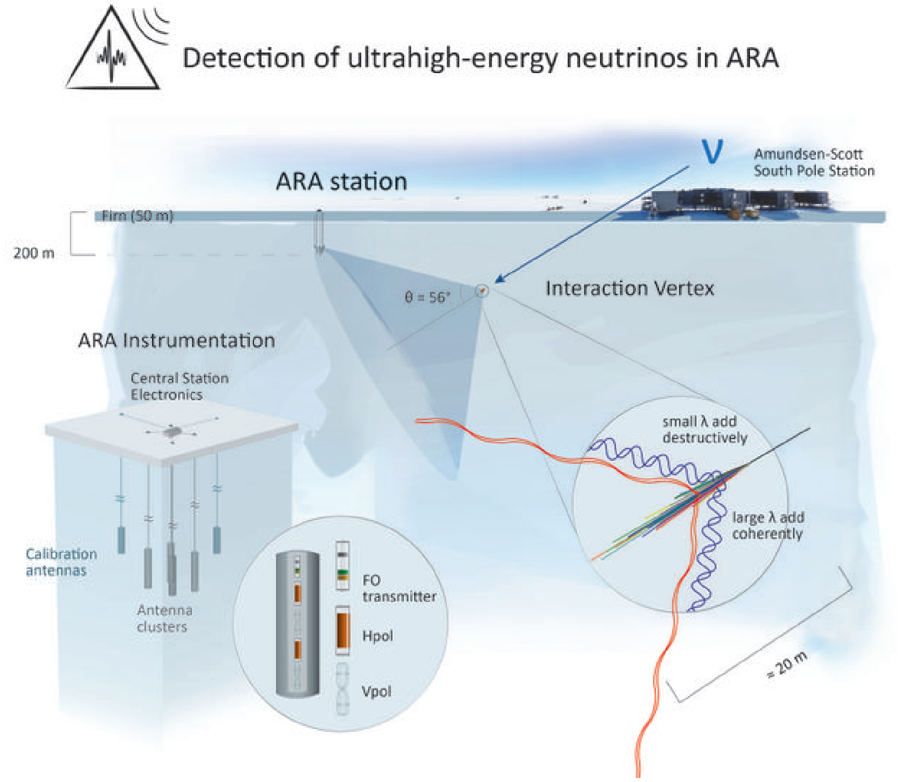


Figure 1.9: A schematic of the ARA detection concept, along with a inset showing the station design. Figure from [48].

ARIANNA, which is located at Moore’s Bay in Antarctica, takes a complimentary approach to ARA, deploying relatively high-gain log-periodic dipole antennas near the ice surface. A schematic of an ARIANNA station along with the neutrino detection principle is provided in Fig. 1.10. ARIANNA has deployed approximately seven stations so far. ARIANNA benefits from the relatively radio-quiet environment of Moore’s Bay, as well as simpler technical challenges during deployment on account of not needing to drill 200 m deep holes. The remote location restricts the experiment to roughly 60% uptime throughout the year, as ARIANNA can only operate when the sun is up. ARIANNA can observe neutrinos through two primary channels: detecting Askaryan radiation which travels directly to the station from the neutrino interaction, and detecting radiation that bounces off the ice-water interface beneath the ice sheet. The experiment also observes cosmic rays approximately once per day [49].

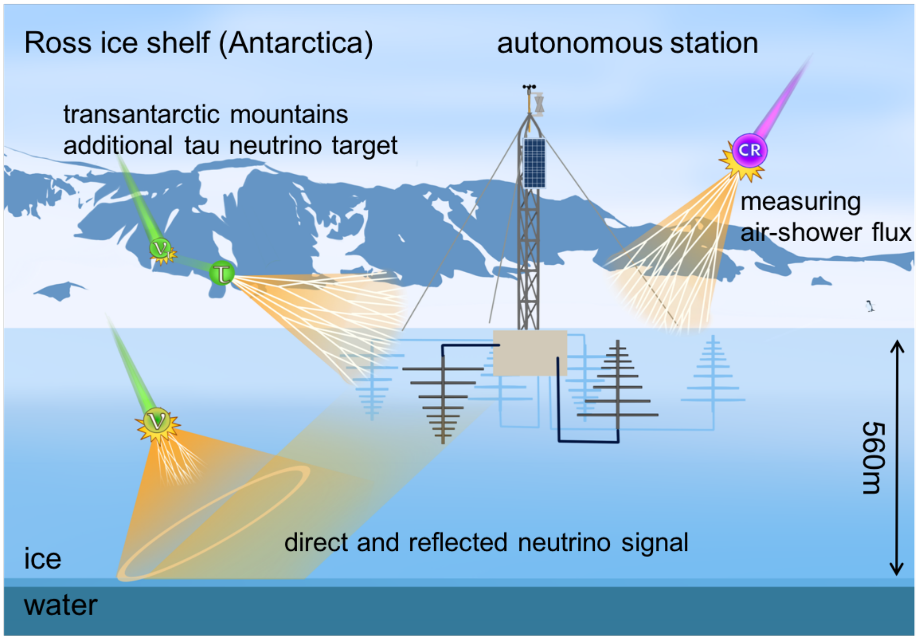


Figure 1.10: A schematic of the ARIANNA detection concept, along with a cartoon of the station design. Figure from [50].

1.4 My Contributions to ARA as a Graduate Student

As a graduate student working on the Askaryan Radio Array (ARA), I have had the opportunity to be involved with all elements of the ARA project, ranging from hardware, to simulation, to analysis. In this section, I will provide a brief summary of my contributions.

1.4.1 Analysis

I have contributed to two major analysis projects. The first was the observation of reconstructable radio emission from a solar flare in the ARA prototype “Testbed” station. This was the first ARA analysis to reconstruct natural radio emission to an extraterrestrial source on an event by event basis. This paper is the topic of Chapter 2 of this thesis document. I have additionally worked on an analysis of four years of data from ARA station 2 and have laid much of the groundwork for an extension of the analysis to ARA station 3. This is the topic of Chapters 4 and 5 of this thesis document.

1.4.2 Hardware

I led the construction of two new ARA stations, A4 and A5, which have since been deployed to Antarctica and are currently taking data. I also led the refurbishment of stations A1 and A3, which were redeployed to Antarctica and are also taking data again. In the process of building new stations, and refurbishing the old, we implemented a new, more modular power and control system. In A4 and A5, we also implemented a new signal conditioning module, called the ARA Front End, or ARAFE. The ARAFE has variable attenuators that allow for gain matching between channels as well as allowing us to prevent saturation of the digitizer system when intentionally injecting bright calibration signals. The design and characterization of the ARAFE is the topic of Chapter 3 of this thesis

1.4.3 Simulation

Lastly, I have worked on two simulation projects. The first is called `AraSimQC`, which stands for AraSim quality and control. The QC software runs the AraSim simulation package and generates a standard set of output plots. The goal is to see how changes committed to the code from the development team affect these physics outcomes. The second major simulation project is work on a next-generation Monte Carlo simulation framework known as `NuRadioMC`, and a paper has been prepared describing it [51]. I am responsible for implementing a C++ based ray tracing algorithm which led to a factor of 20 speed up in the signal propagation methods. In addition, I added new trigger modes to the AraSim software which have allowed us to directly compare AraSim results to other software packages. This is valuable in our work in optimizing the detector design for next generation radio arrays.

Chapter 2

OBSERVATION OF RECONSTRUCTABLE RADIO EMISSION COINCIDENT WITH AN X-CLASS SOLAR FLARE IN THE ASKARYAN RADIO ARRAY PROTOTYPE STATION

The contents of this chapter are largely taken from a publication I have submitted to the Journal of Astroparticle Physics on behalf of the ARA collaboration. We are in the process of revising and resubmitting the paper after receiving reviewer requests for minor revisions. The full citation, along with the arXiv number, is available below.

> P. Allison *et. al.* for the ARA Collaboration. *Observation of Reconstructable Radio Emission Coincident with an X-Class Solar Flare in the Askaryan Radio Array Prototype Station.* arXiv:1807.03335 (2018).

2.1 Abstract

The Askaryan Radio Array (ARA) reports an observation of radio emission coincident with the "Valentine's Day" solar flare on Feb. 15th, 2011 in the prototype "Testbed" station. We find ~ 2000 events that passed our neutrino search criteria during the 70 minute period of the flare, all of which reconstruct to the location of the sun. A signal analysis of the events reveals them to be consistent with that of bright thermal noise correlated across antennas. This is the first natural source of radio emission reported by ARA that is tightly reconstructable on an event-by-event basis. The observation is also the first for ARA to point radio from individual events to an extraterrestrial source on the sky. We comment on how the solar flares, coupled with improved systematic uncertainties in reconstruction algorithms, could aid in a mapping of any above-ice radio emission, such as that from

cosmic-ray air showers, to astronomical locations on the sky.

2.2 Introduction

The Askaryan Radio Array (ARA) is a radio array designed to detect ultra-high energy (UHE) neutrinos via their radio-Cherenkov emission in ice [23, 24]. The array consists of 150-850 MHz antennas buried in the ice at 200 m depth. The signature of a UHE neutrino would be an impulsive signal that originates from a neutrino-induced cascade of charged particles in the otherwise quiet ice.

The ARA Collaboration plans to deploy a 100-km^2 array at 200 m depth in the ice near the South Pole. So far, five deep stations have been deployed–one in 2012, two in 2013, and another two in 2018. In 2011, prior to the deep deployments, an initial prototype "Testbed" station was deployed at ~ 30 m depth to assess the level of anthropogenic and natural backgrounds at radio frequencies, clarity of ice near the South Pole in our frequency band, and to test potential designs.

We report on an observation of radio emission from the sun that is coincident in time with the Feb. 15th, 2011 "Valentine's Day" solar flare. The radio emission arrives with relative time delays among our antennas that indicate a direction to its origin that tracks the sun's movement–with a 2° systematic offset in azimuth, and 10° systematic offset in elevation–over the 70 minute duration of the flare. The radiation was generated at the sun and propagated to earth, and is not the neutrino-associated Askaryan radiation the array was designed to detect.

The events are the first reported by ARA to reconstruct radio on an event-by-event basis to an extraterrestrial source on the sky. As systematic uncertainties in reconstruction algorithms – such as errors in antenna positions and modeling of the depth-dependent index of refraction – are reduced, such events from a moving high-statistics, above ice emitter can be used to improve ARA's mapping of above-ice RF sources to locations on the sky. This will be especially applicable to studies of the direction of cosmic rays which produce geomagnetic radio emission through extensive air showers.

The solar flare events reported here are different from solar-related emission previously reported by ARA from two days earlier on Feb. 13th [52] in that they have tight reconstructions and track the sun for an extended period of time; a more extensive discussion of the Feb. 13th flare can be found in Appendix A.4.

These events also aid ARA's understanding of the radio sky. Sources of radio emission, both natural and human-made, can interfere with both triggering and data analysis. We would also like to be open to observations of astrophysical radio sources given ARA's unusually large field of view with a broad bandwidth. To this end, we are able to build spectrograms that demonstrate the observation of type-II and type-IV solar radio emission that agrees with other radio observatories.

The paper is organized as follows. In Sec. 2, we describe the ARA Testbed instrument. In Sec. 3, we outline the main features of the radio emission observed from the flare, in terms of spectra and reconstructions. In Sec. 4, we present a signal analysis of the events demonstrating evidence for the hypothesis that they are correlated thermal noise from a point source. In Sec. 5, we discuss the potential of this and future solar flare detections in ARA, namely their potential role in understanding ARA's response to above-ice RF sources like cosmic-ray air showers. In Sec. 6, we discuss a search for other solar flares with the Testbed and with deep ARA stations. In Sec. 7, we summarize our results.

2.3 The ARA Testbed Instrument

The Askaryan Radio Array (ARA) "Testbed" instrument is an array of 14 "high"-frequency (150-850 MHz) broadband antennas buried up to 30 m deep, and two "low"-frequency (30-300 MHz) antennas placed at the surface. All are located at the South Pole with the goal of detecting ultra-high energy neutrinos via their radio-Cherenkov emission in ice. The Testbed was a prototype instrument deployed in January 2011 and operated until January 2013. The station's main purpose was to assess the radio noise environment, measure ice properites, and explore design choices later utilized in a mature station; a full instrument description of the Testbed is available in [52]. Of the 14 higher-band antennas, the analysis presented in this paper primarily uses data from eight which were buried at a depth of ∼30 m in "boreholes": four vertically polarized (VPol) bicone antennas and four horizontally polarized (HPol) bowtie-slotted cylinder antennas. These antennas were selected for the analysis because, (a) for each polarization, the antennas have the same basic dipole-like design (bicone for VPol antennas, bowtie-slotted cylinder for HPol antennas), and (b) they have the same sampling rate of 2 GHz.

The station consists of two primary elements; "downhole" components which receive and transmit the signal, and "surface" components which provide data acquisition and system control. Downhole components include the antennas, signal conditioning, and transmission to the surface. Surface components include additional signal conditioning, and triggering and digitization electronics.

The RF signal chain provides amplification and filtering in preparation for digitization and triggering. In the ice, the signal undergoes $\sim$ 40 dB of low-noise amplification, a 450 MHz notch filter to remove South Pole communications signals, and is transmitted to surface electronics via radio-over-fiber (RFoF). At the surface, a transceiver returns it to electronic RF, and the signal undergoes an additional 40 dB of second stage amplification, a 150-850 MHz band-pass filter to remove out-of-band amplifier noise, and a splitting to trigger and digitizing paths.

In ARA, the trigger is designed to identify transient impulsive events and the digitizer records these signals with high fidelity. The trigger path routes to a tunnel diode, which functions as a ~ 5 ns power integrator. Any signal exceeding 5-6 times the mean thermal noise power in 3 out of 8 borehole antennas within a 110 ns time window triggers the digitization of all 16 antennas in the array. Signal to be digitized is sampled at 2 (1) GHz for borehole (shallow) antennas with a 12-bit custom digitizer. The digitization window is 250 ns wide, centered to within 10 ns on the trigger time. Impulsive events, which produce large integrated power over short time windows, naturally satisfy this trigger. The trigger will also fire if the blackbody thermal noise background of the ice randomly fluctuates high in a sufficient number of channels, or if some radio loud source illuminates the array.

2.4 Characteristics of Observed Radio Emission

The ARA team first noticed the events of interest in March 2014 during the final stages of a data analysis aimed at searching for UHE neutrinos in the ARA prototype Testbed station [53]. In that analysis, we searched for events that did not have strong continuous-wave (CW) contamination, and that originated in the ice below the Testbed station. Their location of origin was determined through a peak in the cross-correlation of waveforms from different antennas across the station, accounting for delays in any hypothesized direction of origin.

The events of interest were discovered during final tests performed in the neutrino search, and were determined to originate from the Feb. 15th flare due to time coincidence and directional reconstruction that tracked the sun's location. By removing the original search requirement that the event's origin be in the ice below the station, we found 244 events were recorded in the time period from 02:04-02:58 Feb. 15th 2011 UTC in only 10% of the dataset used until "unblinding" the final stage of analysis. When the full 100% ublinded data set is examined, 2323 events between 02:01-02:58 UTC pass cuts. This time period was found to be coincident with the February 15th, 2011 X-2.2 solar flare, which the RHESSI Flare Survey webpage [54] reports to have begun at 01:43:44 UTC the same day in X-ray measurements. These events on Feb. 15th follow closely events from an M-class flare on Feb. 13th, which were reported on in the prototype instrument paper and are discussed further in Appendix A.4.

2.4.1 Trigger Rates and Spectra

Figure 2.1 shows that during the hour over which these 244 events were observed in 10% of the data, our trigger rates rose from a few Hz during a normal period to as much as 25 Hz. This maximum trigger rate is set by the deadtime during readout whenever a trigger is fired.

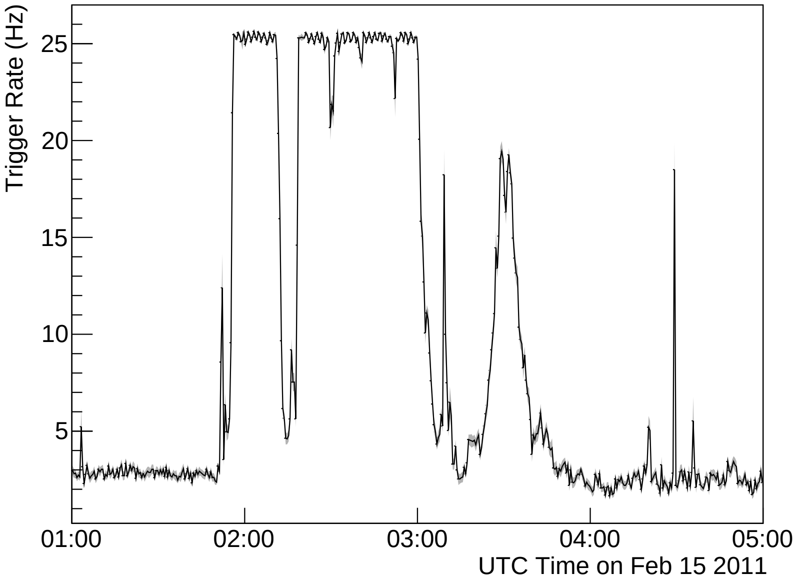


Figure 2.1: The Testbed global trigger rate during the solar flare. Before and after the flare, the instrument triggers near the pre-determined “thermal-noise riding” rate [52]. The black line is the mean trigger rate per thirty second time bin, and the grey band is the standard error on that mean (that is, the square root of the number of entries per bin).

In Fig. 2.2, we show average spectra from one antenna (VPol) for the 2323 events observed in 100% of the data, compared with an average spectra over 2300 events from the forced trigger sample on Feb 11, dominated by thermal noise. One can see that over the period of the flare, the excess emission is predominantly in the 200-400 MHz portion of the band.

We create a spectrogram for the 01:30-03:00 UTC period and compare with those made by solar radio astronomers e.g. [55, 56]. In Fig. 2.3, we plot a background-subtracted spectrogram of a borehole VPol channel over the entire hour of the flare. The background spectrogram is constructed from an identical time window on February 11th, 2011, when the sun would have been in approximately the same position, but was not active.

The radio emission begins to saturate our trigger at approximately 01:57 UTC, roughly coinciding with the peak of the soft x-ray flux. Such coincidence is a feature typical of type-II flare radiation [55]. Two reference measurements of the peak soft x-ray flux are given by the RHESSI instrument (3-10 keV), which reports peaking at 01:55 UTC [54] and by the Fermi GBM (>10 keV coinciding with GOES observations), which reports peaking at 01:59 UTC [57].

The spectrograms, and their associated features between 150 and 500 MHz, agree with

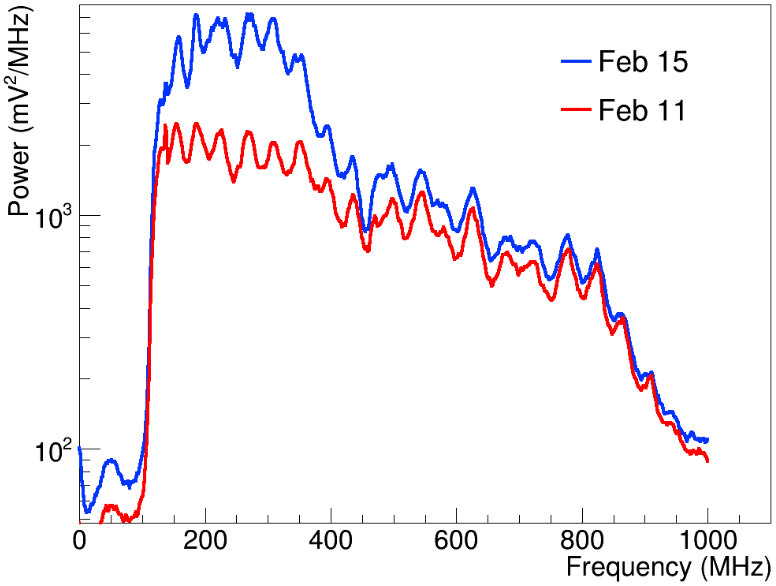


Figure 2.2: Average spectra of Channel 2 (VPol) for the 2323 events that pass analysis cuts designed for the neutrino search on Feb 15$^{\text{th}}$ (blue) and 2300 events between 2 AM and 3:00 AM that are thermal noise forced triggers (red) from Feb 11$^{\text{th}}$. This shows the power excess observed between 200-400 MHz during this particular time of the flare. The small scale oscillations, present in both the flare and thermal noise sample, are attributed to instrumental characteristics. The lines are the mean of the distribution of spectral amplitudes in a given frequency bin; the distribution is an exponential, as is expected for the power spectrum of thermal noise plotted in the units of mV^2/MHz.

other radio observatories that classify this flare emission as type-II followed by type-IV. The Culgoora telescope, with sensitivity from 18-1800 MHz, records spectrograms comparable to our own and lists this flare in their type-II catalog [58].

2.4.2 Directional Reconstruction

The most striking feature of these events is how well they "reconstruct" uniquely to the sun, tracking the motion of the sun in azimuth during that hour. In other words, considering all hypothesized directions across the sky, we found the highest cross-correlation values in a direction within 2° in azimuth of the sun, without distinctly different directions also giving competitive cross-correlations for the same event. The azimuthal angle of the reconstruction peak has a 2° systematic offset from the true value of the sun's azimuth. The events also track the solar position in elevation, but with a significant systematic offset, appearing $\sim 10°$ higher in the sky than the true solar elevation. In this section, first we will review the method that we use to calculate cross-correlation values associated with positions on

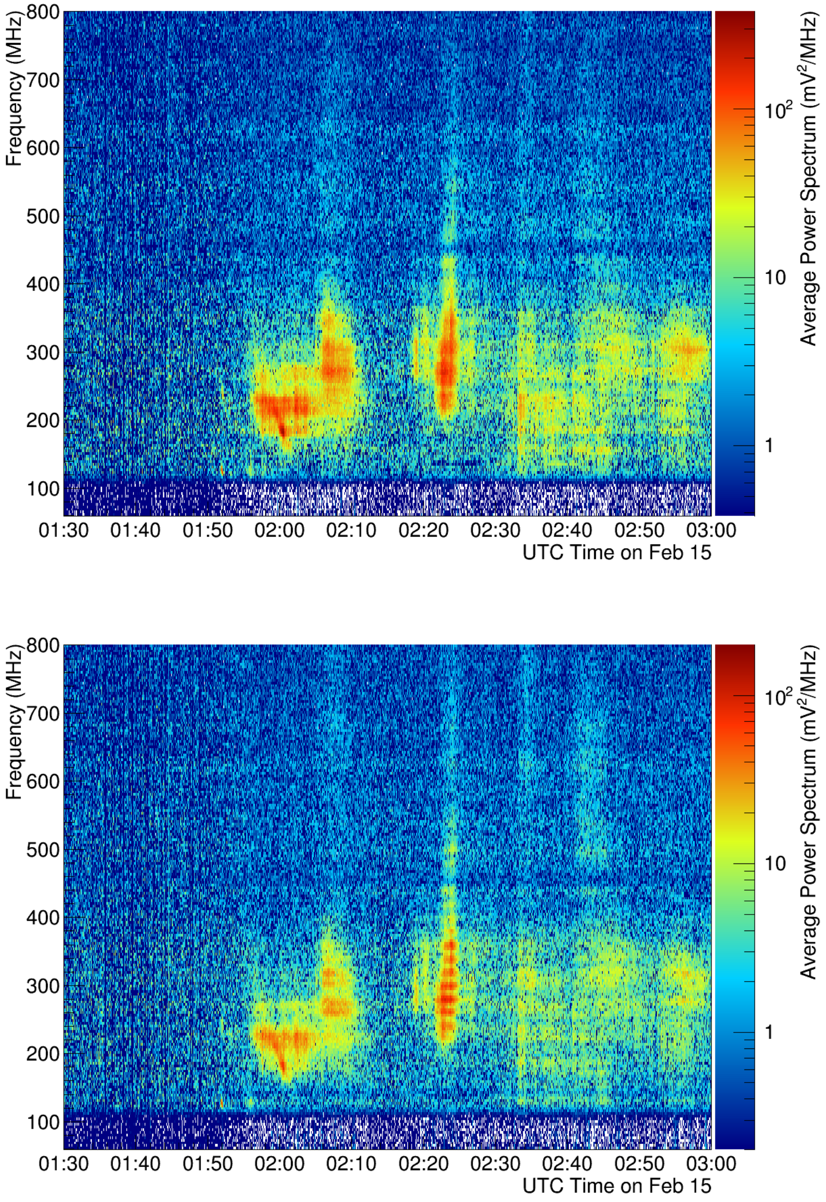


Figure 2.3: (Top) A background-subtracted spectrogram of Channel 2 of the Testbed, for the hour from the beginning of the flare. The background sample is taken from an identical time period on Feb. 11th (four days prior) when the sun was not flaring. (Bottom) A background-subtracted spectrogram produced from the coherently-summed waveform from two channels (Channels 3 and 4) given the delay relative to one another that gives the strongest cross-correlation value. Note that the coherently-summed spectrogram demonstrates that many features are shared between antennas. We only use two channels, as opposed to four channels combined with a directional hypothesis, to avoid a reduction in the signal strength due to any timing misaligments for a given directional hypothesis.

the sky before showing cross-correlation maps for a typical flare event. We also describe our calculations of coherently summed waveforms that we use to investigate the nature of the correlated noise component of these events.

Reconstruction in ARA Analyses

In order to determine the direction of the source of radio signals, we use an interferometric technique similar to the one used in a search for a flux of diffuse neutrinos using data from the ARA Testbed station that takes into account the index of refraction of the ice surrounding the antennas [53]. We first map the cross-correlation function from pairs of waveforms from two different antennas to expected arrival delays from different putative source directions. For each direction on the sky, the mapped cross-correlations from many pairs of antennas are added together and where the delays between different pairs of antennas signals have the strongest agreement, there is a peak in the correlation map.

For a given pair of antenna waveforms, the cross-correlation between the voltage waveform on the i-th antenna ($f_i(t)$) and the voltage waveform on the j-th antenna (($g_j(t)$) as a function of time lag τ can be found from Eq. 2.1:

$$C_{i,j}(\tau) = \sum_{t=-\infty}^{\infty} f_i(t) g_j(t+\tau) \tag{2.1}$$

The time lag τ depends on the position of the source relative to the array, characterized as elevation angle θ, azimuthal angle ϕ, and the distance R to the source; the origin of this coordinate system is defined as the average of the positions of the antennas contributing to the map. As in [53], we only consider two hypothesized distances, 30 m and 3000 m. In the original diffuse analysis, these were chosen because 30 m is roughly the distance of the local calibration pulser, and 3000 m is a estimate for a far-field emitter like a neutrino interaction. The total cross-correlation strength for a given point on the sky (θ,ϕ) is given by summing over all like polarization pairs of antennas as in Eq. 2.2

$$C(\theta, \phi; R) = \sum_{i=1}^{n_{ant}-1} \sum_{j=i+1}^{n_{ant}} C_{i,j}[\tau(\theta, \phi; R)] \tag{2.2}$$

Plotting this total cross-correlation strength for all points on the sky yields a correlation map, as is shown in Fig. 2.4. The azimuth and elevation correspond to the position on the sky of a putative emitter relative to the station center.

The arrival delays τ are found by calculating a path from a hypothesized source location to an antenna through an ice model that accounts for the changing index of the Antarctic firn. We consider a constant $n = 1$ index of refraction in the air. The ray-tracing method

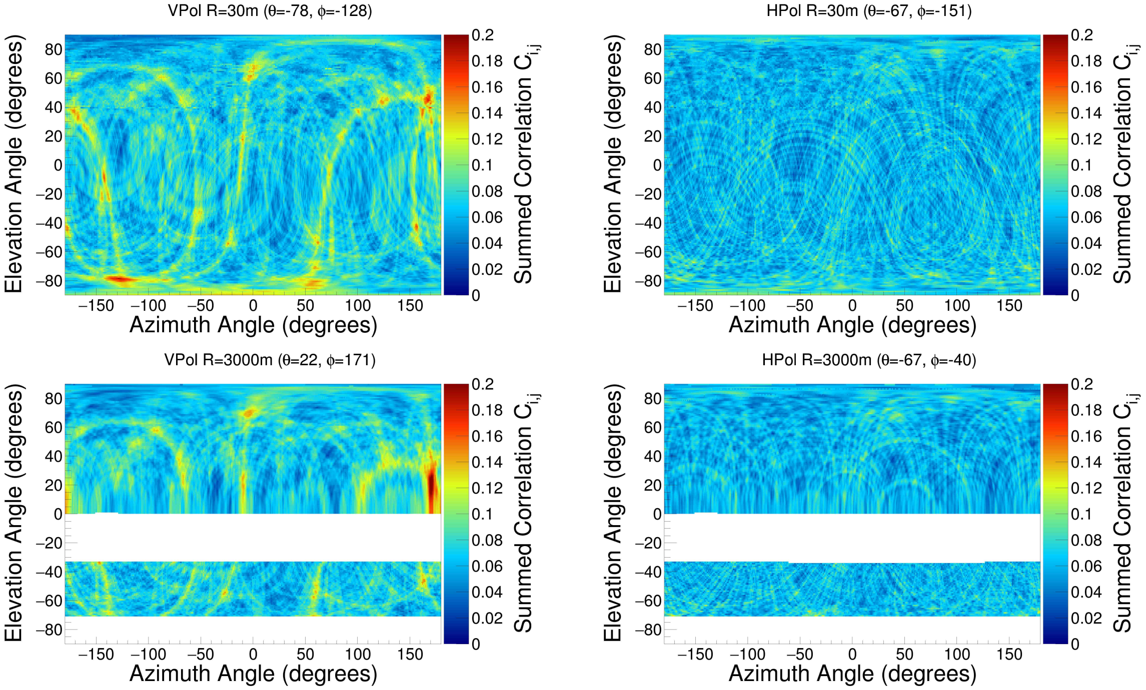


Figure 2.4: Map of cross-correlation values associated with directional hypotheses across the sky for the same event from the flare period as the one shown above in Figs. A.1 and A.3. The upper two figures show reconstruction maps for the hypothesis of 30 m distance, while the lower two figures show reconstruction maps for the 3000 m distance hypothesis. The maps are created in the Testbed local coordinate system. The white gaps in the 3000 m maps correspond to areas of ice where there is either no ray tracing solution (upper white band) or where there is bedrock (lower white band).

models the changing index of refraction as follows:

$$n(z) = 1.78 - 1.35e^{0.0132z} \tag{2.3}$$

This index of refraction model is taken from a fit to data taken by the RICE experiment [59], and is the same as the one used in [53]. To smooth uncertainties in this ice model and other systematics, we calculate the Hilbert envelope of the cross-correlation function before summing over pairs, as is done in previous analyses [53, 60].

Reconstructed Directions Track the Sun

Figs. 2.4 shows maps of cross-correlation values obtained by considering directions across the sky for the same event as shown in Figs. A.1 in the Appendix. Comparison plots from a quiescent period can be found in Appendix A.1. Note that we find the strongest cross-correlation values when we consider the $R = 30$ m hypothesis rather than the $R = 3000$ m hypothesis. Although, the peak in the cross-correlation map has an elevation that is closer to the true elevation of the sun in the $R = 3000$ m case.

We attribute the offsets in the azimuth and elevation of the reconstruction peak, as well as the higher correlation peak value at the near radius, to coupled systematics uncertainties associated with our reconstruction algorithms. This includes the orientation of the local surface normal $\hat{n}$, the parameterization of the depth-dependent index of refraction $n(z)$, variations of the index-of-refraction at the surface due to blowing snow and human activity, unknown cable delays at the level of 1-2 ns, and surveying uncertainties at the level of 10-20 cm. The relative importance of these various systematics is under ongoing study.

The top panel of Fig. 2.5 show the reconstructed directions for all events passing cuts as a function of time for the duration of the flare, taken from the peak of the maps for the $R = 3000$ m hypothesis. The events during this time period reconstructed to a direction within a few degrees in azimuth and approximately 10° in elevation angle of the sun's position in the sky, appearing higher than the known location of the sun (13.3°). We note that the second increase in trigger rates in Fig. 2.1 around 03:30 UTC does include a set of ~ 200 events that also reconstruct to the sun, but do not pass any analysis cuts. We do not study them in depth, and instead focus on the emission from 01:50-03:00 UTC where many events pass analysis cuts and we have direct overlap with results from other solar observatories.

2.4.3 Coherently Summed Waveforms

In addition to building a correlation map, we can construct coherently summed waveforms (CSWs) from antennas of like polarization by first delaying the waveforms with respect to each other by a τ that is specific to an antenna pair, and then summing. For a set of τ's corresponding to the direction of the source, the CSW should have uncorrelated noise suppressed relative to any correlated component.

If we wish to observe this CSW under a source direction hypothesis, we derive the lags τ from the ray-tracing method described above. We can also find the CSW in a way that does not depend on the index of refraction model. To do this for a pair of antennas i and j, we consider all potential lags τ and choose the one that provides the highest $C_{i,j}$.

2.5 Correlated Thermal Noise from the Sun

We conclude that the source of the reconstruction of the events is the presence of correlated thermal noise between antennas. This is consistent with the sun acting as a bright thermal point source that illuminated all antennas. The observation that a thermal source generates fields that are correlated between spatially separated detectors is expected and understood from classical theories of statistical optics; for example, see *Goodman* [61] Chapters 5 and

6.

Thermal noise would be characterized by spectral amplitudes, in a given frequency bin, whose distribution over many events would follow Rayleigh distributions [61]. To verify this thermal nature, we looked at the spectral amplitudes for different frequency bins and attempted to fit each against a Rayleigh distribution. For a sliding three-minute time bin across the duration of the flare, we fit the distribution of spectral amplitudes x in a given frequency bin to a Rayleigh:

$$H(x) = \frac{Ax}{\sigma^2} \exp - \frac{x^2}{2\sigma^2}. \tag{2.4}$$

as was done to characterize the noise environment in other analyses [36, 53]. The normalization of the fit A and the characteristic width parameter σ are allowed to float.

The top panel of Fig. 2.6 shows two spectral amplitude distributions for 210 and 250 MHz with their best fit Rayleigh superimposed. For comparison, the bottom panel of Fig. 2.6 shows a set of Rayleigh distributions from February 11th, 2011 when the sun was not active. The quality of the fit is evaluated by computing the reduced χ^2 test statistic for each spectral and time bin. The distribution of the reduced χ^2 test statistic is Gaussian and centered about 1 for all time bins except for the five minutes after 2:20 AM. This is shown in the bottom panel of Fig. 2.5 for a two-channel CSW (described below). Note that this short period coincides with a gap in the passing events that track with the Sun's position. Note also that the σ parameter, which represents the power in a spectral bin, is higher on Feb 15 than on Feb 11, consistent with the brightened thermal emission over background observed in Sec. 2.4. Fits for a larger variety of frequencies on February 15th are available in Appendix A.6.

In the top panel of Fig. 2.7, we plot the evolution of the fit parameter σ as a function of time. We find that for frequencies between 200-400 MHz, σ increases at times coincident with the brightening of the flare observed in spectrograms. Based on the quality of the single-channel fits, we find the excess power during the flare in the 200-400 MHz band to be consistent with that of thermal emission. We do not mean that the observed emission has a relationship between the intensity of spectral modes that follows a blackbody spectrum.

We also perform the Rayleigh fits for the CSWs described in Sec. 2.4.2, and best fit values of the σ parameter are shown as a function of time on the bottom of Fig. 2.7 for comparison to those derived from the single-channel fits. These particular CSWs are formed out of only two antennas of like polarization and using the time lag that gives the highest cross-correlation value. This is likely to give us a time lag that is more appropriate for picking out the correlated component than a time lag that is derived from a directional hypothesis, given that we observe timing uncertainties associated with our directional hypotheses. By fitting Rayleigh distributions to the spectral amplitudes derived from the CSWs, we are investigating the nature of the emission that is correlated between the antennas, as the

CSW suppresses the uncorrelated component relative to the correlated component. The radiation that is correlated between the antennas is also consistent with thermal, and looks similar to the distribution of total excess emission seen in a single channel in the left panel.

In the progress of determining the thermal nature of the events, we also investigated the possibility that the reconstructability of the events was due to the presence of (1) low signal-to-noise ratio transients, (2) continuous-wave (CW) contamination, (3) time-dependent features like chirps that might be expected in solar radio emission. We found evidence for only the thermal emission described in this section, though we review the findings of these other investigations in Appendix A.3.

2.6 Implications

Although the detection of solar flares by ARA is not the experiment's purpose, such events provide the opportunity to enhance the information extracted from above-ice RF sources, especially the astronomical coordinates of cosmic-ray air showers. Being sourced from the sun, the solar flare events enable validation of software used to project an event's RF reconstruction direction onto celestial coordinates.

The power of these flare events is, at present, limited by systematic uncertainties associated with ARA reconstruction algorithms. Each antenna has both an associated surveying uncertainty and also an associated cable delay uncertainty. Additionally, there is an overall uncertainty in the details of the ice medium transporting the radio signal - specifically, the refractive index profile as well as the surface slope at the optical entry point into the ice. Once those uncertainties can be narrowed to the level of the statistical error on the ARA channel-to-channel timing cross-correlation ($\sim$100 ps), solar flare events can then be used to help point reconstructed ultra-high energy cosmic rays. Specifically, the solar flare data is useful because it contains many events ($>$ 1500 events/min) which have clear reconstructions on an event-by-event basis. The sun is a true far-field plane wave emitter at a known location, and moves slowly enough to be treated as stationary while statistics accumulate, and still sweeps out a series of azimuthal angles.

Finally, if events like these are observed again, the flare can serve as a calibration source for the entire array. As future ARA stations are deployed, they will be positioned further from pulsers deployed near South Pole infrastructure (like the pulser atop the IceCube Counting Laboratory, or "ICL rooftop pulser"). As this distance increases, the effectiveness of the ICL Pulser will diminish as the signal will be considerably weaker at the far stations. With a solar flare as a calibration source, the strength of the signal remains consistent across all stations. The utility of solar flare events to calibrate radio-frequency instruments has been demonstrated by experiments seeking to observe the radio emission from cosmic-ray air showers–for example, see work done by the LOPES Collaboration in Horneffer *et.*

al.[62].

2.7 Other Solar Flares in ARA Data

2.7.1 Search for Other Solar Flares with the Testbed

In addition to the Feb. 15th events found through the interferometric analysis described here, and the Feb. 13th events discussed in Appendix A.4, we also located bursts of radio emission coincident with three other solar flares in 2012. Events were found to correlate with solar flares from March 5 2012, March 7 2012, and Nov 21 2012.

We checked the reason these other three flare do not pass the interferometric analysis. For this purpose, we again removed the cut which required the events come from within the ice. For the March 5 flare, no single cut is responsible for the rejection of all events. For March 7 and November 21, all events are rejected by a cut designed to remove events contaminated by continuous-wave (CW) emission; more than 86% of recorded events are also simultaneously rejected by the "Reconstruction Quality Cut" which imposes requirements on the height and size of the peak in the interferometric maps. We do not explore these events further here.

2.7.2 Search for Solar Flares with Deep ARA Stations

A search was performed to attempt to detect similar solar flare emission in the deep stations around the times of the X-class solar flares listed in Tab. A.1 in Appendix A.4. In examining data from one of the deep stations (A2) taken during the 2013 season (data already analyzed for purposed of a diffuse search [36]), the cut criteria were altered to attempt to identify clusters of events that reconstruct to the Sun's position. Because of the differences between the Testbed station and the deep station, the same set of cuts were not used but rather only one cut was implemented: the event must have a reconstructed correlation value greater than 0.18 with a assumed reconstruction radius of 3000 m. To test the validity of this method's ability to reconstruct events above the ice, it was applied to ICL rooftop pulser events (described in Sec. A.1.3), which reconstructed regularly to the correct position above the ice. After applying this cut on the data taken during the 2013 solar flare periods, we find no events passing these cuts. However, this search was not explicitly designed to detect solar events, and it is also the case that ARA has yet to analyze data from 2014 forward, which was a more active part of the solar cycle than 2013 [63], and so some deep station cases may still be in archival data.

2.8 Conclusions and Discussion

The ARA Testbed has recorded events during a solar flare on February 15th, 2011 that contain a broadband signal and uniquely and tightly reconstruct within a few degrees of the Sun. The excess radio emission is found to contain a component that is coherent across the antennas, consistent with correlated thermal emission expected from the sun being a bright thermal point source.

This is the first observation by ARA, and by an ultra-high energy Antarctic radio neutrino telescope, of emission that can be identified as being produced at an extraterrestrial source on an event-by-event basis. The events offer a high statistics data set that, coupled with improved systematic uncertainties, can validate the software ARA uses to map above-ice RF sources, such as cosmic-ray air showers, to celestial coordinates.

From the perspective of solar physics, we offer a comparison of spectrograms derived from the total observed emission and compare to the spectrograms obtained with the correlated component amplified. This may provide a view of the mechanisms through which the correlated and uncorrelated components of the emission are produced. On a basic level, due to our 2 GHz sampling of 250 ns waveforms, ARA may be able to contribute to the knowledge of solar flare phenomena on smaller time scales than can be probed elsewhere, and thus smaller length scales.

ARA's location in the southern hemisphere and radio quiet environment may add data to the worldwide network of solar burst spectrometers. At present, most spectrometers are located in the Northern Hemisphere (e.g. Greenbanks in Virginia, Palehua in Hawaii, Ondrejov in the Czech Republic). Some reside in the south–for example, Culgoora and Learmouth–but are contaminated by steady-state man-made spectral features, for example the noise lines in Culgoora spectrographs near 180 and 570 MHz. The radio community has already expressed a need for a so-called "Frequency Agile Solar Radiotelescope" [56], to which ARA bears remarkable resemblance in bandwidth and digitization electronics. This could also be true for other radio neutrino-telescopes deployed to Antarctica, such as ARIANNA [64].

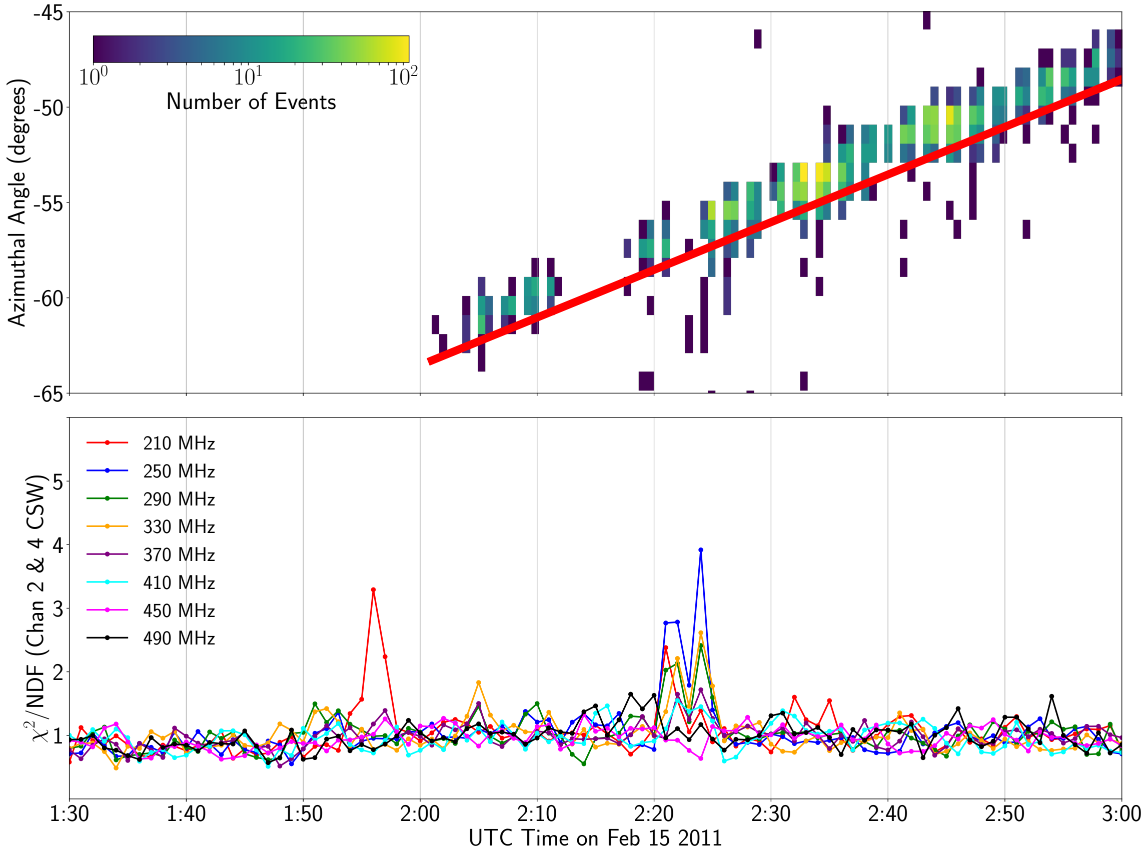


Figure 2.5: (Top panel) Reconstructed direction in azimuth of events occurring during the flare, and the expected location of the sun drawn as a red line. The azimuthal angle here follows the continent global coordinate system described in the Appendix. (Bottom panel) The χ^2/NDF values for fits of the spectral amplitudes to Rayleigh distributions of the CSW formed from channel 2 and 4 (see section 2.5) given the delay relative to one another that gives the strongest cross-correlation value. We only use two channels to avoid a reduction in the signal strength due to any timing misalignments for a given directional hypothesis. Note the increased deviation from the Rayleigh fit in the few minutes after 2:20 UTC which matches a period when the events no longer pass the analysis cuts.

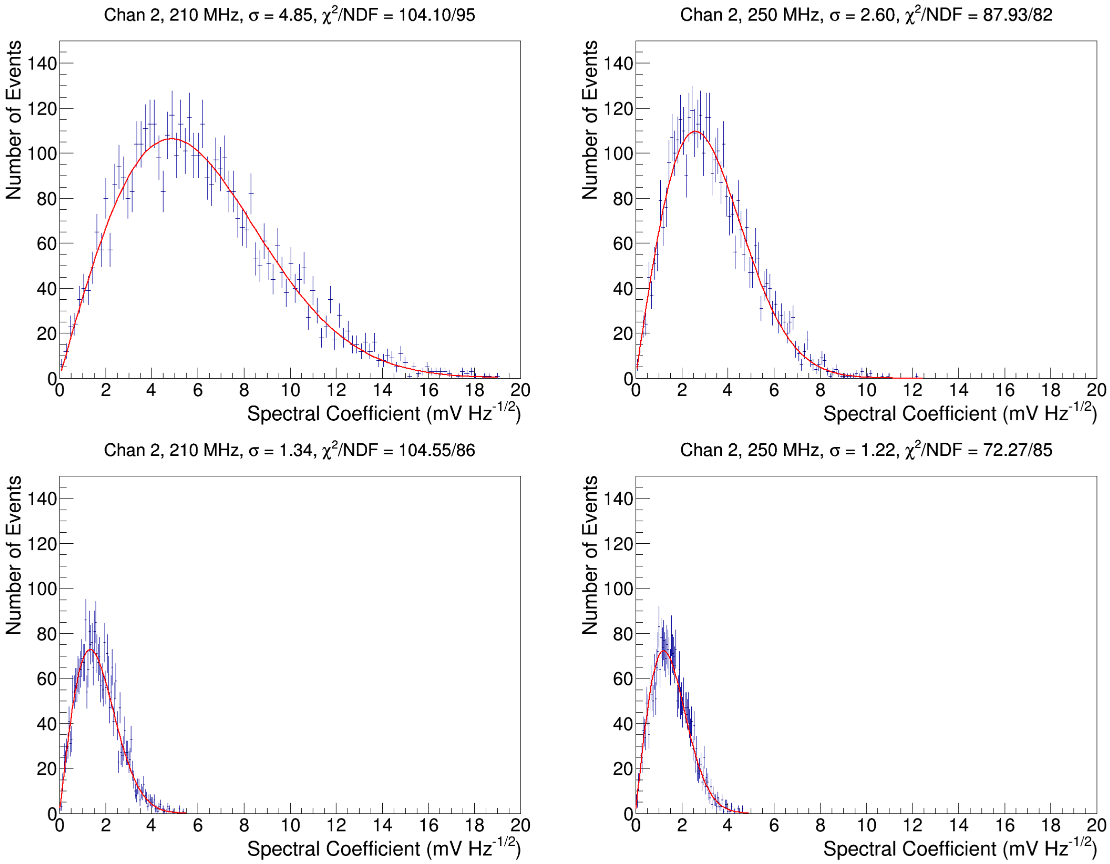


Figure 2.6: (Top): Distributions of amplitude spectra at two frequencies where the emission was strong, 210 MHz and 250 MHz, during the period of flare. These events satisfied the RF trigger between 1:59 and 2:01 UTC. (Bottom): The same as the top figure, but during a time when the sun was not active. These events are from the forced triggered data sample recorded on February 11th, 2011 between 01:40 and 03:00.

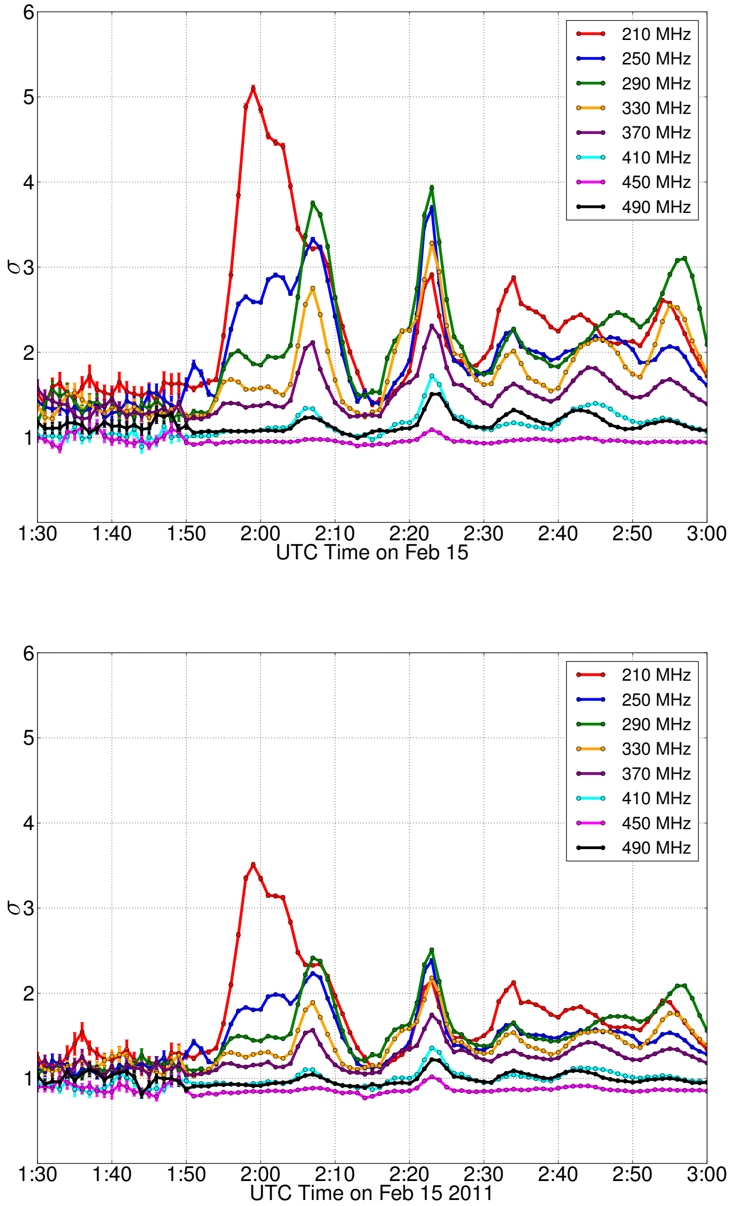


Figure 2.7: (Top) The time evolution of the Rayleigh fit parameter for channel two as a function of time. (Bottom) The same for the CSW from channels 2 and 4. In both, the error bars are the one-standard deviation errors on the fit parameter. The larger error bars before 1:50 are due to lower statistics before the trigger saturates after ∼1:50AM.

Chapter 3

The design, characterization, and performance of a new radio-frequency front-end module for the Askaryan Radio Array

The contents of this chapter are largely taken from a publication I am preparing on behalf of the ARA collaboration which will be submitted to Journal of Nuclear Instruments and Methods A.

3.1 Abstract

The Askaryan Radio Array (ARA) is an *in-situ* ultra-high energy ($> 10^{16}$ eV) neutrino detector under phased construction at South Pole. The latest two of five deployed stations were installed in Austral Summer 2018, both containing new ARA Front End (ARAFE) systems. The ARAFE is a redesign of the detector's second-stage signal conditioning. The modules increase the robustness of the systems' circuit protection, and feature onboard microcontrollers to dynamically tune digital attenuators in the radio-frequency signal chain. This dynamic adjustability allows for gain matching between channels of the instrument to 0.25 dB precision. The modules are spatially compact (< 25 in^3) and low power (<0.56 W/channel), and result in a $\sim$ \$10 k (20%) reduction in the cost of the station electronics. We demonstrate the performance of the ARAFEs in Antarctica by using them to avoid saturation of the A4 by high amplitude calibraiton pulsers.

3.2 Introduction

The Askaryan Radio Array (ARA) is an ultra-high energy (UHE) neutrino telescope ($>$ 10 PeV) under construction at the South Pole [30]. ARA seeks to observe neutrinos through the radio-Cherenkov (Askaryan) emission that is predicted to be produced by neutrino-induced particle cascades in the Antarctic ice [23, 24]. The signature of the Askaryan emission is a broadband ($150-1200$ MHz), fast (~ 500 ps) radio-frequency (RF) impulse.

ARA searches for this signal by burying a roughly 30 m cubical lattice of antennas 200 m deep inside boreholes in the South Pole ice sheet. A prototype station ("Testbed") was deployed at 30 m depth in 2011 [30], a first "deep" station (A1) at 100 m depth in 2012, two more (A2, A3) at 200 m depth in 2013 [36], and yet another two (A4, A5) at 200 m depth in 2018 [65, 66].

A signal observed by an antenna down a borehole must be transmitted to the surface and amplified and filtered before digitization. In ARA, these tasks proceed in two stages: in "first-stage" antenna-mounted systems, and in a "second-stage" surface system.

In ARA stations A4 and A5, a new design was used for the second-stage amplification and filtering. These new modules, named the ARA Front Ends (ARAFEs) are more compact, more cost effective, and more flexible in capabilities than in previous station generations. A key part of the flexibility is a bank of eight tunable digital step attenuators that enable ARA to gain-match the amplification on different channels to 0.25 dB precision to more-effectively leverage the dynamic range of the digitizer and triggering systems.

In this paper, we review the design and impact of the ARAFE. In Sec. 3.3 we review the ARA signal chain, and motivate the development of the ARAFEs. In Sec. 3.4 we provide an overview of the ARAFE modules themselves. In Sec. 3.5 and Sec. 3.6 respectively, we cover the design of the two main components of the modules, the Power and Control (PC) board and the Radio Frequency (RF) board. In Sec. 3.7 we characterize the RF properties of the modules and in Sec. 3.8 we demonstrate the performance of the ARAFE in the Antarctica in station A4.

3.3 The ARA Signal Chain

The ARA signal chain is designed to amplify the signals that we exepct to observe from neutrino-induced relativistic particle cascades. Askaryan models predict a microvolt signal [27, 67] that once folded with our antenna response must be boosted to the millivolt level for digitization. The signal chain must also notch filter the 450 MHz South Pole radio communications channel and filter out-of-band amplifier noise. The full signal chain, from being recorded at the antenna through digitization and storage, is depicted in Fig. 3.1.

Signal processing begins "downhole". The ARA antennas are sensitive to frequencies between 150-850 MHz [30, 36]; the physical constraints of placing the antennas down 6"

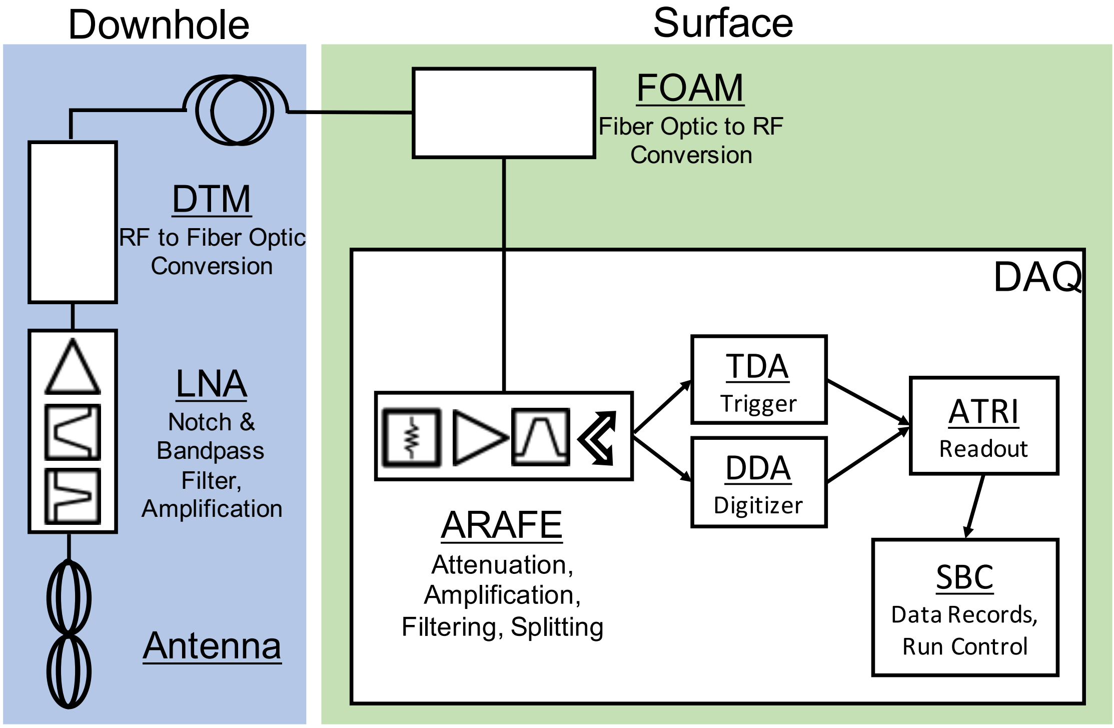


Figure 3.1: A schematic of the full ARA signal and data path for a single RF channel.

boreholes defines this range, and sets the bandwidth of the system. After exciting an antenna, a signal is first deeply notched ($>$ 80 dB) at 450 MHz to remove South Pole radio communications, is band-pass filtered, and amplified by $\sim$34 dB. The notching and filtering is purchased from LARK engineering. Downhole amplification and filtering occurs inside of an antenna-mounted Low-Noise Amplifier (LNA).

The RF outputs of the LNA then enter the Downhole Transmission Module (DTM) and are converted to a fiber-optic signal to avoid attenuation during the 200 m transmission to the surface. At the surface, a transceiver stored in the Fiber Optic Amplification Module (FOAM) returns the optical signal to electronic RF and applies another $\sim$16 dB of gain. The signal is then ready for "second-stage" conditioning.

The second stage conditioning must provide an additional $\sim$ 23 dB of gain, filter out-of-band amplifier noise, and route the signal to the Data Acquisition Electronics (DAQ). The DAQ is located on the surface of the ice, and is responsible for triggering and recording signals, packaging the data, and transmitting data back to permanent storage in the IceCube Laboratory.

In previous stations (A1-A3), the second-stage conditioning was spatially divided; the amplifiers were stored outside of the DAQ box inside the FOAMs, while the filtering and splitting occurred inside the DAQ box. The filtering was conducted with the same LARK filters as are used inside the LNAs. The division of conditioning between the FOAM and Data Acquisition (DAQ) lead to numerous mechanical connections and poor serviceability during deployment in Antarctica. Further, use of the LARK filters was unnecessarily costly, as the LARK filters provide better out-of band rejection than is necessary at the second-stage, while contributing almost \$600 to the per-channel cost of the station.

The ARAFEs were designed to address the challenges listed above. All of the second-stage conditioning was relocated to the ARAFEs, which are housed inside the DAQ box. This means the second-stage conditioning has a much smaller spatial footprint in A4-A5 than in A1-A3 and have fewer mechanical connections. The ARAFEs also consolidated the amplifiers and filters onto the same printed-circuit board, and utilize a much lower cost ($\sim$ \$10/channel) board-mounted band-pass filter. Finally, we were able to outfit each channel with a tunable attenuator that allows for gain matching between channels to 0.25 dB.

3.4 Design Overview

3.4.1 Operational Design

The ARAFE modules have three critical functions. First, they must provide power to the fiber-optic transceiver modules in the boreholes. Second, they must perform second-stage amplification of the RF signal and filter out-of-band amplifier noise. Third, they must split the signal into a digitizer and trigger path for further analysis by the ARA DAQ. Each ARAFE module is equipped with four channels of conditioning, and so one module is necessary per ARA detector string.

The first function of an ARAFE module, supplying power to the fiber-optic transceivers, is provided via a bias-tee. Through the bias-tee, the coaxial cable which transports RF signal from the transceiver to the ARAFE can also support transmission of the 12 V direct-current (DC) voltage line that powers the transceiver. To generate the 12 V power line, the ARAFEs contain a low-ripple ($\sim 100\mu$V) DC-to-DC voltage regulator to down-convert the 15 V powering the modules to the 12 V required by the transceiver.

The second function of an ARAFE module, amplifying and filtering of RF signals, is provided via a series of two modest-gain (11 dB and 17 dB) amplifiers, followed by a high-pass (145 MHz) and low-pass (830 MHz) filter. Each channel also has two dedicated, digital, step attenuators that allow for tuning of the gain of each channel in 0.25 dB increments.

The RF amplifiers and step attenuators are powered with a 5V DC voltage supply line. To generate this second voltage supply line, the ARAFEs contain a second low-ripple DC-to-DC regulator. The design is the same as the 12V case, with slightly different values for the passive components (resistors and capacitors) that configure the voltage.

The third function of an ARAFE module, the splitting of the signal between the digitizer and trigger paths, is carried out by a 10 dB coupler, after the signal has been band-pass filtered. A coupler allows most of the power to proceed along a “digitizer path”, but a copy is “coupled-off”. This copy is reduced in power by 10 dB, and is therefore a small fraction of the total power in the signal. This -10 dB copy is routed to a “trigger path”. The digitizer path leads to the Digitizing Daughter Boards (DDAs) that are responsible for recording signals. The triggering path leads to a tunnel diode that acts as a square-law detector, before being fed into the Triggering Daughter Boards (TDAs), which govern the triggering procedure.

Control of the ARAFE modules is extremely granular. Each ARAFE contains a microcontroller that can communicate with a user via a serial interface. The microcontrollers can enable/disable and monitor the voltage of the DC-to-DC converters powering the RF amplifiers and the fiber-optic transceivers. The microcontroller also controls electronic fuses on the bias-tees, and can shut-down individual transceivers. Finally, the microcontroller programs the digital step attenuators in the signal chain. The attenuator settings and programming clock are broadcast to all attenuators, with the microcontroller sending an enable pulse to a user-specified attenuator.

3.4.2 Physical Design

To isolate the sensitive RF signal chain from the noisy DAQ electronics, the modules are built as two-stacked boards. One board, named the Power and Control (PC) board, houses the power generation and control tools, and is described in Sec. 3.5. A second board, named the Radio Frequency (RF) board, contains the RF signal chain, and is described in Sec. 3.6.

The PC board consists of circuit protection electronics (Sec. 3.5.2), the microcontroller (Sec. 3.5.3), the 12 V and 5 V DC-to-DC converters that power the RF amplifiers and fiber-optic transceivers (Sec. 3.5.4), and the electronic switches which regulate power to the fiber-optic transceivers (Sec. 3.5.5).

The RF board contains the four channels of signal conditioning. A channel consists of a bias-tee (Sec. 3.6.1), digital step-attenuators (Sec. 3.6.2), amplifiers, filters, and coupler (Sec. 3.6.3).

The PC and RF boards stack together and are mounted inside a custom machined aluminum enclosure, modeled after Compac 51235-138-1 [68], but with custom holes for all RF cabling to pass through. The enclosures are about 4”×4”×1.5” in size, making them very compact at 24 in^3 volume. The enclosures underwent MIL-DTL-5541F, type-II, class 3 [69] chromate conversion (or iriditing) to generate a conductive surface and provide conductive coupling to the rest of the DAQ enclosure.

3.5 Power and Control (PC) Board

3.5.1 Communications over Power

Power and communications to each PC board is provided by a single power-ground wire pair, minimizing the number of individual conductors needed to operate the module. This is done by adapting the "broadband over power lines" (BPL) technique that has been successful in the medium voltage housing electric grid [70]. An 8 MHz carrier is superimposed on the 15 V line powering the PC board, and the carrier is modulated by data bits. The presence of the carrier is a logical "1" and its absence a logical "0", otherwise known as on-off keyed (OOK) transmission. The PC board performs edge-detection on the signal (with a series of Schottkey diodes and a comparator) to recover the data bits, which are in turn interpreted by the serial receive (RX) port of an on-board microcontroller.

We adapt the implementation of *Truchsess* [71], which was engineered for low-voltage applications. We specifically use a half-duplex asynchronous serial connection, implemented with a standard 9600/8-N-1 protocol (9600 baud rate, 8 data bits, no parity, 1 stop bit).

Our circuit is the same as in *Truchsess* so we do not reproduce it here. We do make a few important modifications for robustness and performance. The changes make the initial bit of a transmission more closely resemble the following bits and allows for smoother turn-on. The discrepancy between the first and later bits is evident in Fig. 3 of *Truchsess*, where initial rise time for the first bit is extremely slow, exceeding 1 bit time. The following are the changes, all referenced to Fig. 1 in *Truchsess*. First, D1 and D4 were changed to BAS70 diodes for reduced reverse leakage, while D2 and D3 were changed to BAT54S for better conduction over temperature ranges when the diode is forward conducting. Second, a 10 kΩ termination resistor to ground was added at the node labeled 1, which is otherwise left floating. Third and finally, a pull-up/pull-down pair to 3.3V/GND was added after R1 to bias the point to 3.3V/2V, which is the DC average of the carrier output.

3.5.2 Circuit Protection

The PC board protects all electronics—including both the first-stage fiber-optic transceiver and second-stage RF amplifiers—from overvoltage (OV), undervoltage (UV), and reverse bias (RB) events through a comprehensive power controller, the LTC4365 from Linear Techonology [72]. If the supply voltage exceeds or recedes above or below specified "lockout" points, the LTC4365 will disconnect the supply from the load. The protection ensures smooth turn-on of the sensitive RF electronics, shields the RF amps from transient events in the field, and also makes the modules more robust in the laboratory test environment (e.g., accidentally being connected to power supplies with power and ground cables reversed).

The LTC4365 essentially functions as an autonomous electronic switch between the 15 V power supply and the rest of the PC board. It can shield downstream systems very quickly, as it can disconnect the supply from the load in less than 25 μs. We set the system to have a UV lockout at $11.28^{+0.21}_{-0.22}$ V and a OV lockout at $18.25^{+0.36}_{-0.35}$ V. The UV and OV levels are set by a network of resistors external to the LTC4365; uncertainties in the lockout points result from variations in the precise value of those resistors.

The LT4365 functions as a switch by toggling the gate voltage of a back-to-back N-channel MOSFET network (we use Vishay Si4946BEY [73]). The gates of the two MOSFET's are mutual, as are the sources. The drain of one is connected to the input supply, and the drain of the other connected to the load. When in normal operation mode, the LT4365 drives the gate of the network high relative to the load, forcing the network into a conducting state. When experiencing a OV/UV event, the LT4365 idles the gate voltage below the input voltage, shutting off the MOSFET connected to the power supply. In an RB event, the LT4365 connects the gate directly to the supply, disabling the MOSFET connected to the load.

3.5.3 Microcontroller

We utilize an ultra-low power MSP430G2153 microcontroller by Texas Instruments [74], which maintains the small power consumption of the modules. When the microcontroller is in active mode, it consumes 2.2 mA at 3.3V, or 7 mW. Most of the time, the device idles in "standby" mode, where it consumes more than a factor of 200 less power. The device can be woken from its standby mode into active mode in less than 1μs.

This specific microcontroller family suits the needs of our application well. Specifically, the microcontroller contains an on-chip comparator necessary for recovering data bits in our communications-over-power scheme (see Sec. 3.5.1). The microcontroller is additionally equipped with at least 14 general purpose input/output (GPIO) pins—enough to manage the four bias-tee switches, two on-board DC converters, and the RF board's eight attenuators. At least one GPIO pin possesses an oscillator to program the attenuators, and six GPIO pins have analog-to-digital converters (ADCs) to provide voltage monitoring of the two low-ripple DC converters and the four bias-tee switches.

The microcontroller is powered by a dedicated on-board DC regulator. The regulator—an LP2951CD-3.3R2G by On Seminconductor [75]—converts between the 15V line powering the PC board and the 3.3 V needed by the microcontroller.

3.5.4 Low-Ripple Voltage Regulation

To down-convert the 15 V powering the ARAFE to the 12 V and 5 V needed to power the fiber-optic transceivers and RF amplifiers, we utilize a low-ripple, high-efficiency switching voltage regulator from Linear Technologies, the LT8611 [76].

By heavily filtering the output of the LT8611 with a circuit of capacitors and inductors, we achieve a 100 μV (0.75 mV) peak-to-peak switching transient on the input (output) of the second amplifier on the RF board. This is nearly a factor 50 smaller than the $\sim$ 40 mV thermal noise floor at the same place in signal chain.

We achieve an efficiency of 93% for the 12 V converter and 82% for the 5 V converter. The efficiency of a voltage regulator (η) is the fraction of power delivered to the converter at high-voltage (V_{in}) which is converted to power at a lower voltages (V_{out}). The results are about 10% lower than manufacturer specifications, which we attribute to the large quantity of filtering at the output of the regulator. These efficiencies are however higher, by 13% and 49% for the 12 V and 5 V case respectively, than their linear regulator counterparts, which have an efficiency of $\eta = V_{out}/V_{in}$.

Both regulators can be enabled, disabled, and have their current monitored by the ARAFE's microcontroller. Physically, both regulators are covered by Laird BMI-S-202-S [77] aluminum shields to prevent any electromagnetic interference (EMI) from affecting the neighboring RF circuitry.

The switching regulators must be low-ripple because the two second-stage RF amplifiers are connected one after another. As they are powered through a bias-tee on their outputs, any transient generated by the regulator will be amplified by the second amplifier (17 dB) in the RF chain. Use of a linear regulator gave an unacceptably low efficiency in cases which require large differences between V_{out} and V_{in}, such as our 15 V →5 V case.

3.5.5 Bias-Tee Power Switches

The 12 V voltage lines which supply power to the fiber-optic transceiver units can be disabled individually by TPS25921A eFuses by Texas Instruments [78]. The microcontroller operates the eFuse as a switch by controlling the enable pin of the TPS25921A.

The TPS25921A, as a fuse, is equipped with over-current protection. We configure the TPS25921A to shut-down, disconnecting the fiber-optic transceivers from the 12 V DC regulators, if the current draw exceeds 535 ± 5 mA. This is a factor four above the projected consumption of $\sim$ 145 mA.

The presence of these fuse-switches contributes to the granular control of the signal-conditioning subsystem. The switches enable us to disable single fiber-optic amplifiers in the debugging and installation process, or if one amplifier is contributing excessive noise to the system. The fuses prevent a short in a single channel from affecting other normally operating channels.

3.6 Radio Frequency (RF) Board

A single RF board contains four channels of analog signal conditioning. The schematic for a single channel is given in Fig. 3.2. The RF board is manufactured on 31 mil (31/1000

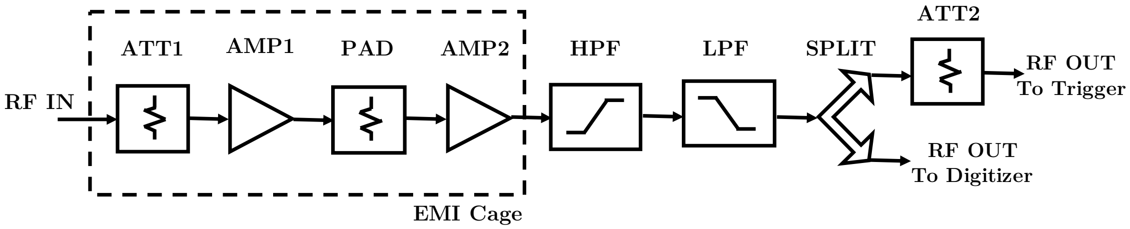


Figure 3.2: A simplified schematic of a single channel of an RF board.

of an inch) thick FR-4 (flame-redardant grade four) printed circuit board material, but is designed for compatibility with even lower loss fabrication materials.

RF signal is carried into the RF board by high-isolation IsoRate cable assemblies by Samtec Technologies [79] (model IJ5H-04-0400-S-2-01SP1). The cable is RG316, terminated on one end by 30 μm gold coated SMA-A and on the other by Samtec's proprietary "EdgeRate" latching connector. Both ends are terminated to 50 Ω for quality signal transmission. The SMA end of the hybrid cables preserve backwards compatibility with the previous generation of ARA Data Acquisition enclosures, as well as input to the ditigization electronics. The EdgeRate end simplifies the mating process to the RF boards, reducing the connections from four torque-SMAs in A1-3 to one latched connector A4-5.

After amplifying, filtering, and splitting the signal, the ARAFEs must route signal to the digitizer and triggering systems in the DAQ box. Signal is sent to the digitizer system by the same model of IsoRate cables that carried signal into the ARAFE modules. Signal is sent to the triggering system by RG316 coaxial cable connected to four board-mounted SMA-As. There was insufficient room on the board to support a third EdgeRate connector for the trigger signal, which would have eliminated SMAs from the modules altogether.

In the following sections, we will describe the stages an RF signal undergoes during conditioning. We provide a summary table of the technical specifications of all relevant parts in Tab. 3.1.

3.6.1 Bias-Tee

Power is provided to the fiber-optic transceivers in the FOAMs through the RG-3616 coaxial cable which carries RF signal into the ARAFE. This occurs through a bias-tee, with the inductor 742792022 by Würth Electronik [80] and a 1 μF ceramic capacitor.

At DC frequencies, the inductor serves as a low-impedance coupling between the 12V regulator and the fiber-optic transceivers, while the capacitor serves as a DC block. At RF frequencies, the inductor is high-impedance (220-330 Ω) relative to the 50 Ω capacitor so only RF frequencies pass into the amplifiers.

Part Name	Technical Specifications
RFSA3713 (ATT1/2)	5 MHz-6 GHz operating range; 1.4 dB insertion loss at 2 GHz; 0.9 mW power consumed (180μA at 5V)
AG201-63 (AMP1)	At 400 MHz: Gain of 11 dB, P1dB of 6.5dB, Noise Figure of 4.4 dB; 100 mW power consumed
ADL5544 (AMP2)	At 400 MHz: Gain of 17.4 dB, P1dB of 17.6dB, Noise Figure of 2.9 dB; 275 mW power consumed
HFCV-145+ (LPF)	1.5 dB insertion loss in passband, $>$ 30 dB rejection in stop band, 3 dB point of 132 MHz
LFCN-630+ (HPF)	1.2 dB insertion loss in passband, $>$ 20 dB rejection in stop band, 3 dB point of 830 MHz

Table 3.1: The technical specifications of parts on the RF boards.

3.6.2 Attenuators

After entering the RF board and passing through the bias-tee, the signal is attenuated by a Qorvo RFSA3713 digital step attenuator [81] before being fed into a series of two amplifiers. The RFSA3713 allows for a maximum of 31.75 dB in attenuation, in 0.25 dB increments.

The presence of the programmable step attenuator on every channel allows for ARA to compensate for gain-mismatches in downhole equipment digitally. Because the ARA system only triggers when three out of eight antennas receive an impulsive signal, channels with significantly lower gain than others suppress the instrument sensitivity. The new-step attenuators allow us to gain balance the channels without changing physical hardware in Antarctica. Further, the attenuators are useful during the calibration procedure, especially when the calibration sources saturate the ARA digitizers. This is visible in such calibrations as the IceCube Deep pulser runs described in *Allison et. al.* [82].

The first digital step attenuator (ATT1 in Fig. 3.2) is placed before the series of amplifiers so that the attenuator can be used to moderate the input power level to the amplifiers. In this way, the step attenuators allow us to place the input power to the amplifiers in the linear operating regime, and prevent the amplifiers from going into compression.

The second digital step attenuator (ATT2 in Fig. 3.2) is placed after the amplification, filtering, and digitizer-trigger split. This attenuator only affects the trigger path, and allows ARA to moderate the power input to the tunnel diodes in the triggering system. This is done to ensure the diodes are operating as square-law detectors. This limited range where the tunnel diode is operating optimally as a square-law detector can clearly be seen in a lab measurement of a diode in Fig. 3.3. Between 2-45 mV the diode output voltage (as a function of input voltage) is well described by a power-law, but deviates elsewhere.

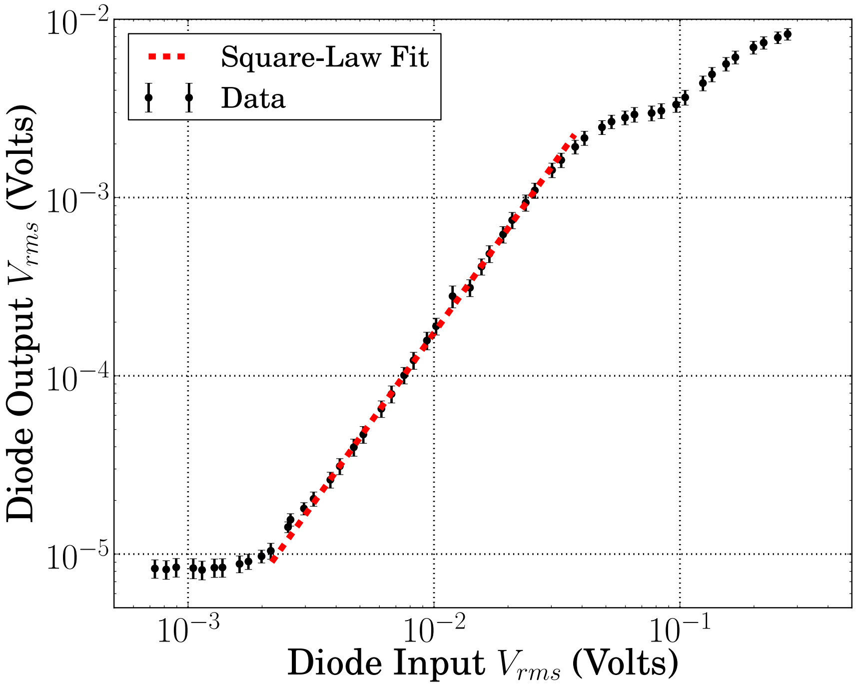


Figure 3.3: The output voltage of the tunnel diode as a function of input voltage. The red dashed line shows the region (approximately 2-45 mV) where the tunnel diode exhibits the behavior of a square-law detector, that is, where the output is proportional to the input squared. The best-fit slope in log-log space is 1.95, close to the expected value of 2.

3.6.3 Amplifiers, Filters, Coupler

After being attenuated, a series of two amplifiers provide the second-stage amplification. The two amplifiers, identified as AMP1 and AMP2 in Fig. 3.2, are an AG201-63 from Qorvo [83] and an ADL5544 by Analog Devices [84] respectively. The two amplifiers provide a nominal combined gain of 28 dB. A 2 dB attenuator (PAT0510S from SSM Thin Film Technologies [85]) is placed in between the two amplifiers to prevent them from oscillating (this is PAD in Fig. 3.2). With a combined power draw of 375 mW/channel, the amplifiers account for the majority of the module's power consumption.

The first attenuator (ATT1) and both amplifiers (AMP1 and AMP2) are covered by the Laird BMI-S-209-C [77] shield to protect the amplifiers from external EMI.

Following amplification, the signal is filtered of out-of-band noise. The filters, labeled HPF and LPF in Fig. 3.2 are a high-pass filter at 145 MHz and a low-pass filter at 630 MHz, respectively. We use the board-mounted HFCV-145+ and LFCN-630+ from Mini-Circuits [86, 87]. These two filters set the band-pass filter for the modules.

Finally, after filtering, the signal is split by a coupler into a trigger and digitization path (this is SPLIT in Fig. 3.2). We use DBTC-10-13-LX+ from Mini-Circuits [88]. The coupler functionally operates as an inductive coil from 5-1000 MHz. The induced "couple" copy is reduced in power by 10 dB. The original signal (or "mainline") is only reduced by 1.4 dB, which accounts for the insertion loss of the coupler and the power lost to the induced signal copy.

The majority of the signal, or "mainline", is routed out through the IsoRate cable assemblies to the ARA digitizer system. The -10 dB copy, or "couple", is sent through one more RFSA3713 step attenuator before exiting through a board-mounted SMA to the ARA triggering system.

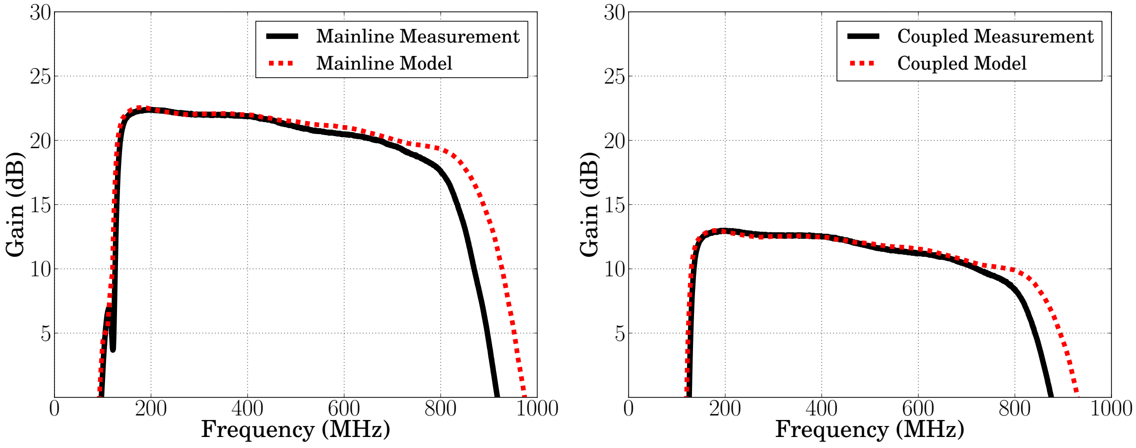
3.7 Characterization

The gain of the ARAFE was measured in the laboratory with a vector network analyzer, and is plotted in Fig. 3.4; we show the mainline (digitizer) path on the top left in Fig. 3.4a and the coupled (trigger) path on the top right in Fig. 3.4b. Along the digitizer path, the modules achieve a maximal gain of 22 dB with a 3 dB point of 730 MHz. As expected, the trigger path has approximately 10 dB less gain than the digitizer path. These measurements represent the full end-to-end realized gain of the modules, including any cable losses.

The phase response of the modules is shown for the mainline (digitizer) path on the bottom left in Fig. 3.4c and the coupled (trigger) path on the bottom right in Fig. 3.4d. The phase response is linear in the range of interest (150-850 MHz), making the modules non-dispersive. This is important because ARA triggers on the rising edge of a $\sim$ 5 ns power envelope of a signal, making it sensitive to fast transients. Triggering on a fast transient is ideal for ARA as the Askaryan emission is expected to be an approximately 500 ps wide pulse. As such, any dispersion introduced by the electronics reduces the amount of power localized into a small time window, and reduces sensitivity.

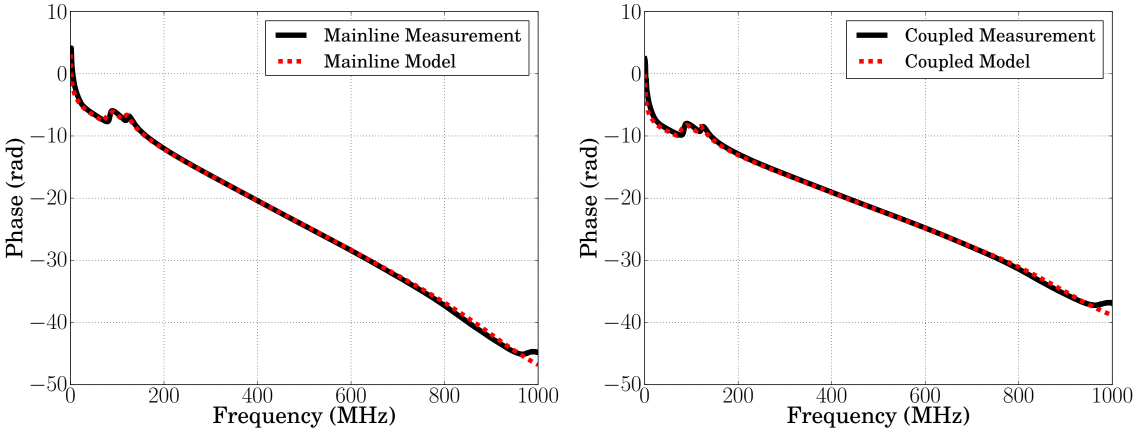
We validate the measurements with a Qucs ("Quite Universal Circuit Simulator") model of the RF board, implemented in QucsStudio [89]. Full two-port scattering parameter files were utilized for all board mounted parts, and we also simulated the RG316 in the IsoRate cables. The gain measurements show excellent agreement with simulation, to within 0.5 dB, below 750 MHz. The phase measurements show agreement across the band.

The slower-than-observed roll-off at high frequencies in the gain of the simulation relative to the measurements is likely explained by inaccurate modeling of the cutoff in the low-pass filter and unaccounted for losses in the RG316 and micro-strip traces on the RF boards. All three more strongly affect high-frequencies than low-frequencies. Some of this high-frequency loss could be recovered by shortening the IsoRate cables or printing the RF boards on material with lower dielectric losses than FR-4. We did not focus on improving the high-frequency behavior for A4 and A5 because the gain of the ARA antennas and LNAs is already falling above 700 MHz and the system temperature is rising.



(a) “Mainline” Path Gain

(b) “Coupled” Path Gain



(c) “Mainline” Path Phase

(d) “Coupled” Path Phase

Figure 3.4: The measured (black lines) and Qucs models (red dashed lines) of the gains (top row) and phases (bottom row) for the ARAFE modules. On the left (a,c) is the mainline path and on the right (b,d) is the coupled path.

We demonstrate the dynamic attenuation capabilities of the modules in Fig. 3.5 by scanning the attenuation setting of the mainline attenuator (ATT1) over five different attenuations settings in 3 dB increments. The shape of the gain curve remains the same as a function of attenuation setting, ensuring the attenuators are altering only the overall amplitude of a signal, and not changing the relative spectral content.

Based on the specifications of the parts used, the gain of the modules decreases with increasing temperature, by about 0.2 dB/10° C. We expect the modules to primarily operate between −20° C and 30° C, a temperature range over which the gain is therefore stable to ∼1 dB. This small variation with temperature is ideal in the Antarctic environment where seasonal variations in the ambient temperature on the order of 60° C induce temperature

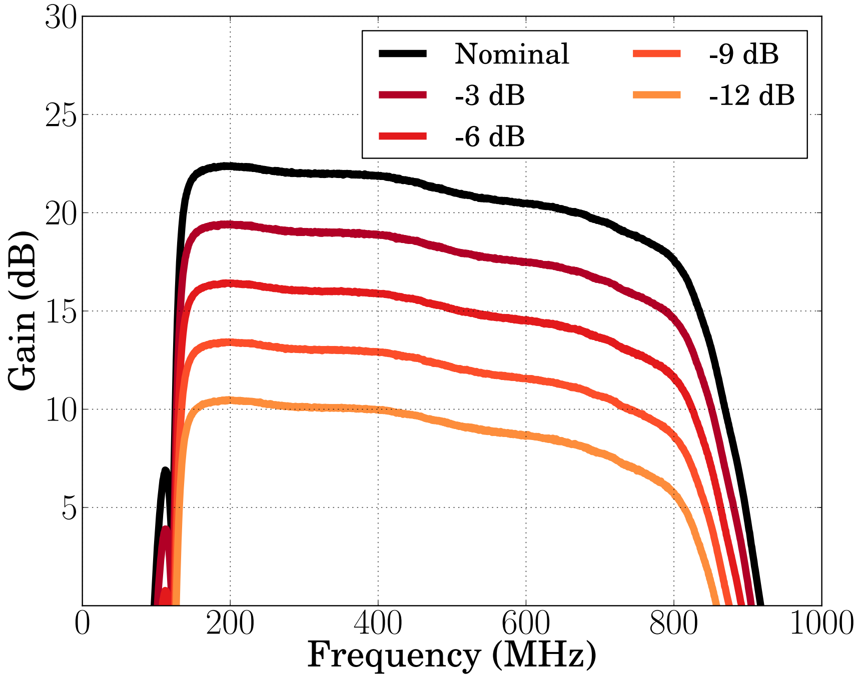


Figure 3.5: A mainline channel of the ARAFE being scanned through five attenuation settings in 3 dB increments.

changes inside the insulated DAQ enclosures on the order of 20°-30° C.

3.8 Antarctic Performance

In this section, we give a demonstrations of the ARAFE's performance in the Antarctic environment. One useful calibration source for ARA is a bright, fast pulser deployed on one of the last few IceCube strings. One pulser was deployed at 2500m depth, at the bottom of IceCube, and two at 1500m depth, near the top of IceCube. When the 1500 m pulsers are operated, the amplitude of the pulses typically saturated the ARA digitizers, which makes studying the spectral content of the pulses or comparing pulses from one station to another impossible (aside from time of flight measurements). However, with the addition of the ARAFE to stations A4 and A5, we are able to raise the attenuation of the signal chain before the pulse arrives at the digitizers. We demonstrate this capability in Fig. 3.6, where on the left it can be seen that the deep pulser saturates the ARA digitizer, and on the right, with a higher attenuation value, the pulser does not. The ability to compare the non-saturated signals between stations A4 and A5 is enabling the longest-ever horizontal baseline measurement of the ice attenuation length. A publication addressing this measurement is forthcoming.

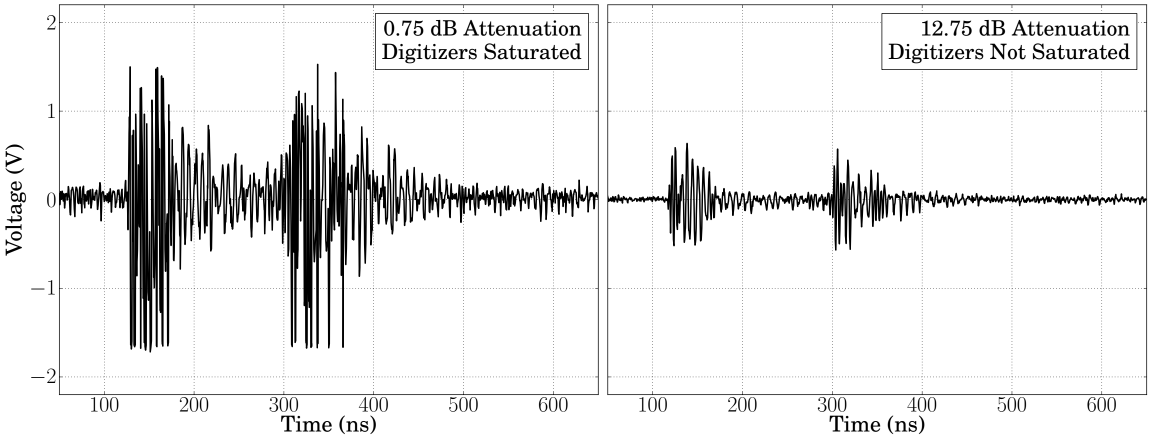


Figure 3.6: Waveforms of the IceCube deep pulser recorded in channel 1 of A4 with lower attenuation (left) and higher attenuation (right), demonstrating that by utilizing a higher ARAFE attenuation, we prevent the saturation of the ARA digitizers.

3.9 Conclusion

In this paper we have outlined the design and characterization of a new radio-frequency front-end module for the Askaryan Radio Array. We have also given a demonstration of its performance in the Antarctic environment, which shows its potential to enable new physics measurements by preventing saturation of the ARA digitizer system.

Chapter 4

Diffuse Neutrino Search: Analysis Methods

4.1 Introduction

In this chapter, I present the data analysis methods used in the neutrino search in A2 data from the years 2013-2016. This includes a description of all cuts used in the analysis; in some sections, I lay the groundwork to repeat this process in station A3. This analysis excitingly leverages move livetime than the previous 10-month analysis, and is the largest period of ARA livetime analyzed to date.

4.2 The ARA2 Instrument

The ARA2 instrument is here defined to be the two-detector array of A2 and A3. A1, which was deployed before A2 and A3, has so far been excluded from the analysis due to challenges calibrating the instrument, and because it is at 100 m depth, and is therefore expected to be less sensitive to neutrinos. A map of ARA2, in the context of several other major South Pole landmarks, is presented in Fig. 4.1. In Fig. 4.2, we additionally show the current five-station ARA instrument ARA5. A2 and A3 were both deployed in the 2012-2013 Pole Season, with drilling in November and December of 2012, instrumentation deployment in December and January of 2013, and commissioning completed by mid-February 2013.

4.2.1 Hardware

An ARA station is a roughly cubical lattice of sixteen antennas buried up to 200 m deep, as well as associated trigger and readout electronics. A diagram of an ARA station is presented in Fig. 4.3. The antennas are deployed at the corners of the cube, with the cube side length being approximately 30 m. Four antennas are deployed along each edge of the cube, comprising a “string”. At the top of a string sits a pair of antennas—one sensitive to horizontally-polarized (HPol) light, and one sensitive to vertically-polarized (VPol) light.

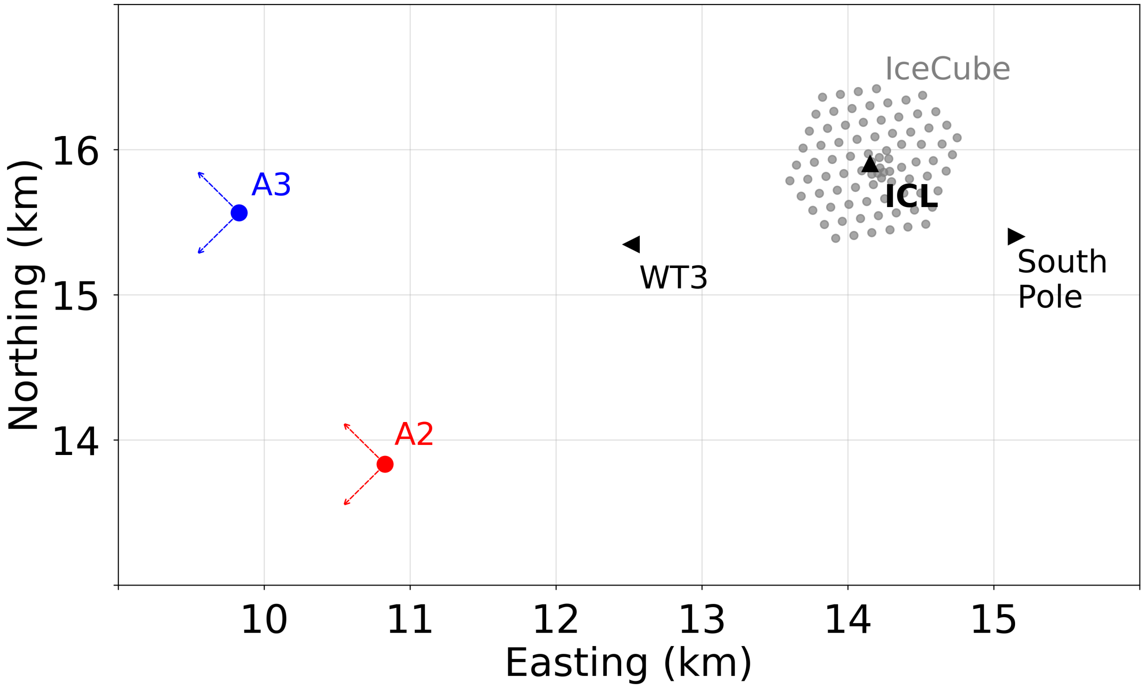


Figure 4.1: A map of the ARA2 instrument at the South Pole, in surveyor coordinates. Marked in red and blue are the A2 and A3 instruments, respectively. In gray is the IceCube instrument. In black are three important landmarks for the ARA project: Wind Turbine 3 (WT3), the IceCube Counting Laboratory (ICL), and the South Pole.

Another pair of HPol and VPol antennas sits at the bottom of the string. In addition to these four measurement strings, each station is equipped with two local-calibration pulsers, which are deployed approximately 40 m away from the station center. On each "cal-pulser" string, there two additional antennas, one VPol and one HPol, each connected to a fast, broadband pulser. By controlling these local cal-pulsers, and illuminating the measurement strings, ARA is able to calibrate the array and verify livetime. The four measurement strings and the two calibration strings comprise a full station.

The locations of the measurement and calibration strings in A2 and A3 are shown in Fig. 4.4a and Fig. 4.4b respectively. Because the instruments were deployed in the same pole season and by the same deployment team, the geometries in the $\hat{x}-\hat{y}$ plane are very similar. The only differences are minor variations, on the order of a half a meter, in the exact $\hat{x}-\hat{y}$ locations of the strings and the depths of the antennas. A more extensive discussion of the coordinate systems of Antarctica and the stations is presented in App. B.1; the appendix also includes the locations of the major landmarks in the local coordinate systems of the stations.

At the hardware level, the stations consist of two main components—a downhole com-

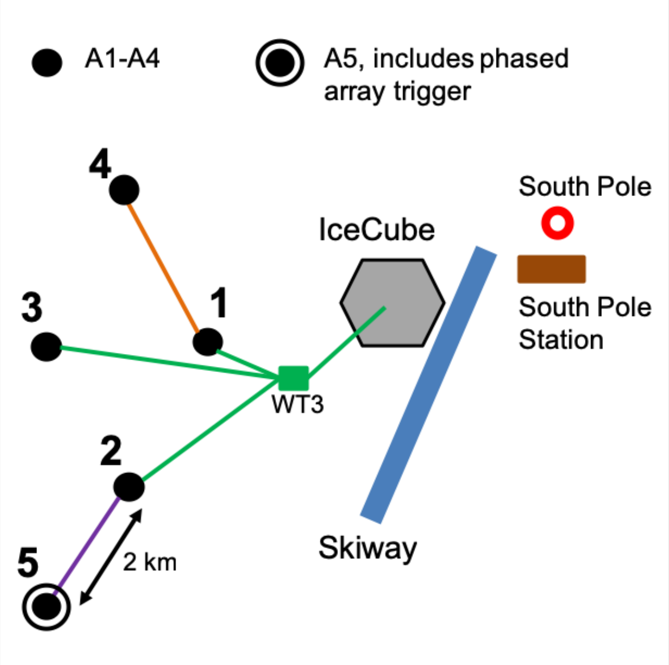


Figure 4.2: A schematic map of the ARA5 instrument at the South Pole.

ponent and a surface component. Signal processing begins in the downhole components. A signal excites an antenna, and an antenna-mounted Low Noise Amplifier (LNA) module performs the "first-stage" signal conditioning. In the first stage, the signal is deeply notched (> 80 dB) at 450 MHz to remove South Pole radio communications, is band-pass filtered from 150-850 MHz, and is amplified by approximately 34 dB. The bandwidth of the system is set by a series of constraints. The Askaryan emission itself is only predicted to be coherent from a few dozen MHz up to a few gigahertz. However, the spectrum of the noise is not flat over this entire extent. Above $\sim$ 130 MHz, the noise in the system is dominated by the 230 K blackbody thermal radiation from the Antarctic ice. Below $\sim$ 130 MHz [90] however, radio noise from the galaxy begins to exceed the blackbody emission of the ice. The lower edge of the ARA bandwidth was chosen to eliminate this dominant galactic radio noise. The upper edge of the band is set by the physical constraints of the borehole. The ARA team found it difficult to design a broadband antenna of only six inches in diameter with an upper frequency range substantially higher than 850 MHz. Further, the very high frequency modes above 1 GHz matter only when the Askaryan emission is observed very near the Cherenkov angle ($\Omega < 0.2°$), which is a small fraction of available phase space.

After an excites a voltage in an antenna, it must be conditioned, transmitted to the surface, and recorded. This happens in the following way. After being amplified in the LNA, RF signal then enters the Downhole Transition Module (DTM) where it is converted

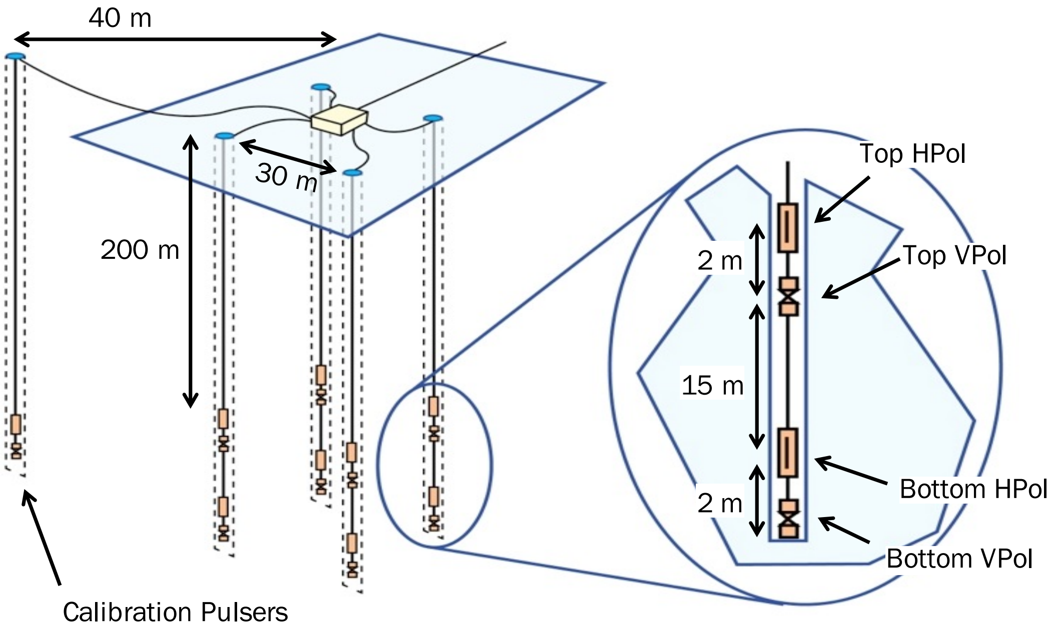
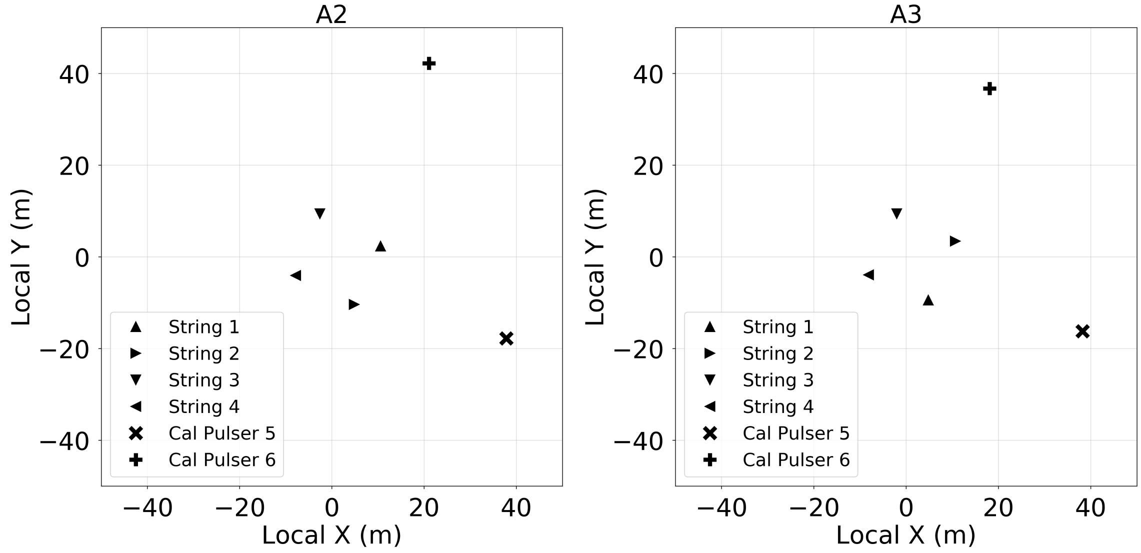


Figure 4.3: A diagram of an ARA station with important dimensions labeled. The inset on the right highlights the VPol and HPol antennas.

to a fiber-optic signal for low-loss transmission to the surface, which is nearly 200 m away. Once the signal has reached the surface, it enters the Fiber Optic Amplification Module (FOAM) and begins its "second-stage" of signal conditioning. In the FOAM, the fiber-optic signal is converted back into electronic RF, and is further amplified by approximately 40 dB. At this point, the signal will enter the Data Acquisition (DAQ) box at the surface. Inside the DAQ box, the signal is once more band-pass filtered. This second stage of filtering is necessary to remove out-of-band noise added by the amplifiers in the FOAM. After this second stage filtering, the signal is split. Half of the power from the signal is routed to the Digitizing Daughter Boards (DDAs), which are responsible for digitizing the signals. The other half of the power is routed through tunnel diodes to the Triggering Daughter Boards (TDAs), which are responsible for triggering. The tunnel diodes act as passive ~ 5 ns power integrators, which improves the signal-to-noise ratio of the event.

4.2.2 Triggering and Readout

The station is designed to trigger when 3/8 antennas of the same polarization exceed five to six times the average thermal noise level within 170 ns. The trigger window is set to this value because 170 ns is the largest amount of time a plane wave requires to casually traverse the array. The trigger threshold of each individual channel is actively adjusted by a servo



(a) A2 station map in local coordinates.

(b) A3 station map in local coordinates.

Figure 4.4: Maps of the A2 and A3 stations in the station-local coordinates. The strings are marked with triangles, while the calibration pulsers are marked with crosses. Note that these local X and Y coordinate systems are not the same as as the continent coordinate system, which have cardinal directions Easting and Northing; the local coordinate systems are rotated approximately 126 degrees counter-clockwise from the continent coordinate system.

to maintain a global station trigger rate of 5 Hz. In addition to these 5 Hz of so called "RF" triggers, we additionally take 1 Hz of calibration pulser data, and 1 Hz of "software" triggers.

The TDAs and DDAs are both controlled by the central ARA Triggering and Readout Interface (ATRI) board. The ATRI board receives the individual-channel triggering information from the TDA, determines if a global trigger has occurred, and then requests the digitizers on the DDAs be read out.

The digitizers used in ARA are Ice Ray Sampler 2 (IRS2) chips. The IRS2 is a switched capacitor array (SCA) digitizer, which essentially operates as a large bank of capacitors, where the stored charge is proportional to the voltage waveform. The bank is organized into 512 blocks, each of which contains eight channels of 64 samples. Each block represents 20 ns of time. The blocks are constantly being overwritten by new data and therefore act as a cyclic buffer. The chips nominally sample the RF signal at a rate of 3.2 GHz providing a a Nyquist frequency of 1.6 GHz. In practice, this is only achieved in the first two channels. The second two channels are effectively sampled at half this rate due to known issues within the chip itself. The first two channels are both reserved for the VPol antennas, while the second two are used on the HPol antennas. The poorer performing channels of the IRS2 are utilized on the HPol antennas because the HPol antennas in ARA are known to have lower bandwidths and lower gains than the VPol. The remaining four channels are used in the following way. The fourth and fifth channels can be connected to other antennas of interest, which in A2 and A3, are connected to surface antennas. The sixth channel is terminated, and the seventh channel is connected to the 100 MHz system clock.

When the ATRI board requests readout, sampling is stopped, and the requested blocks are read out. After being read out, the event is packaged by the ATRI board, and forwarded to a Single Board Computer (SBC) over a Peripheral Component Interconnect express (PCIe) interface. The event is stored in Random Access Memory (RAM) on the SBC for a short amount of time, typically a few tens of seconds, before being transmitted over a fiber-optic network to permanent storage in the IceCube Counting Laboratory (ICL) building.

An ARA "event" is the combination of the 32 digitized waveforms, along with affiliated monitoring and housekeeping data. We record the time the event was read out by the SBC (what is known as the "`unixTime`"), the number of clock cycles since the last Pulse Per Second (PPS) pulse (what is known as the "`timeStamp`"), as well as the trigger rates and thresholds at the time of event readout.

4.3 Analysis Process and Tools

4.3.1 Data Blinding

To help reduce and control for analyst bias, we separate the full ARA data set into two pieces before analysis begins. The first is a 10% "burn" sample which is used to understand backgrounds, tune cuts, and find detector efficiencies. The second is a 90% "full" sample, which will be used to search for a discovery signal, or in the absence of an observation, to set an upper limit on the size of the signal. The bias we are trying to avoid is two-fold: we want to avoid an analyst designing cuts to eliminate potentially interesting signal events. We also want to avoid an analyst designing cuts which will save types of events which appear "good". Based on existing limits on the flux of cosmogenic neutrinos, we expect that the entire data set will contain at most a handful of neutrinos, and so in examining only the 10% data set, we can isolate a sample which is background only.

The splitting of the data sets into "burn" and "analysis" samples is handled by the ARA collaboration. For every set of 10 events, a random event is selected. This is done by throwing a random number between 0 and 1. If the random number is less than 0.1, the first event is selected. If the random number is less than 0.2 but greater than 0.1, the second event is selected, and so forth. This method avoids systematically selecting events with any periodicity (e.g., the tenth event is always corrupted), and also allows for calibration pulsers and software trigger to enter the dataset in a proportion equal to that with which they are

Year	Station 2	# of Station 2 Files	Station 3	# of Station 3 Files
2013	1449-2792	953	473-1907	1001
2014	2820-4763	1327	1930-3785	1272
2015	4765-6638	1506	3786-6150	1143
2016	6645-8246	1142	6159-7808	1251

Table 4.1: A table of the run numbers for each year of the analysis time period.

Config	L1 Trigger Mask	Readout Window (ns)	Trigger Window (ns)	Trigger Delays Active	Livetime (days)
1	none	400	110	yes	179.07
2	none	400	110	no	142.55
3	D4BH	400	110	yes	94.54
4	D4BH	520	170	yes	439.02
5	D4BH	520	170	no	287.28

Table 4.2: A table of the configurations used in A2.

found in the full data set.

4.3.2 Data Taking Configurations

The optimization discussed in this thesis address all of the data taken in stations A2. Because we will also extend this analysis to A3, in some sections, we will discuss the A3 data as well. For A2, the analysis period encompasses run numbers 1449 to 8246, and for A3, it is run numbers 473-7808. Which runs belong to which year of data taking is broken down in Tab. 4.1. The total accumulated livetime, before any cuts addressing bad runs or known human activity, is 1142.46 days in A2 and 1078.52 days in A3.

Over the time period of data taking, the stations were run in several different configurations, representing several different combinations of critical operating parameters. In particular, the operations team varied the size of the triggering window, the length of the waveform being read out, whether or not channel specific trigger delays were applied, and if some channels were prevented from taking part in the trigger. In some cases, the operating parameters were changed because the analysis team had learned more information about the stations; for example, in A2 it was not realized until configuration 3 that the last HPol channel was not working. In some cases, the operations team did not realize a setting, like the trigger delays, were not active.

Config	L1 Trigger Mask	Readout Window (ns)	Trigger Window (ns)	Trigger Delays Active	Good String 4 (see text)
1	none	400	110	yes	yes
2	none	400	110	no	yes
3	none	520	170	yes	no
4	none	520	170	no	no
5	none	400	1100	yes	no

Table 4.3: A table of the configurations used in A3.

Run Numbers	Configuration
0 to 4	1
11 to 60	4
120 to 2274	2
2275 to 3463	1
3465 to 4027	3
4029 to 6481	4
6500 to 8097	5
8100 to 8246	4

Table 4.4: A table of mapping between run numbers and configurations for A2.

We split up the data taking period for each station into five different configurations, with the parameters outlined in Tab. 4.2 and Tab. 4.3. Which configuration was used when does not map linearly in time—for example, in A2, configuration 2 was actually the earliest used configuration in the data taking, while configuration 4 was used through the end of 2015, and was resumed for a few days at the end of 2016. A table connecting specific run numbers to their affiliated configurations are presented in Tab. 4.4 and Tab. 4.5.

It is worth noting that for some periods of livetime, not all channels in all stations were operational. In particular, for the entirety of the data taking period, A2 Ch15 was found to have very high noise levels, thought to be due to a broken antenna-LNA connection. For 75% of the livetime (2014-2016), all of string four of A3 was non-operational (Ch3, 7, 11, 15). These "bad-channels" are removed from the analysis for the time periods for which they were not producing physics quality data.

Run Numbers	Configuration
0 to 3	1
470 to 1448	2
1449 to 1901	1
2275 to 3463	1
1902 to 3061	5
3063 to 6004	3
6005 to 7653	4
7658 to 7808	3

Table 4.5: A table of mapping between run numbers and configurations for A3.

4.3.3 Analysis Framework

The code base of this analysis is constructed on the `AraRoot` analysis framework that is designed and maintained by the ARA collaboration, with myself as a current lead developer. The framework is currently stored on GitHub. The analysis framework provides a common access class to the ARA data, which allows us to extract information about every event—including waveforms, triggering information, etc. `AraRoot` also provides a way to access information about each individual station, including the location of antennas within a station, and the location of each station in various coordinate systems. Access to geometry information is achieved through the `AraGeomTool` class, and is used in the following way:

```
AraGeomTool *geom = AraGeomTool::Instance();
geom->getStationInfo(station)->getAntennaInfo(i)->antLocation[ii];
```

The `AraGeomTool` loads geometry information stored in three SQLite databases: `AntennaInfo.sqlite` contains the information about the measurement antennas, `CalPulserInfo.sqlite` contains information about the calibration pulsers, and `AraArrayCoords.sqlite` contains the location of the stations in a global coordinate system. Information can be drawn from the databases externally using standard SQL calls.

The rest of the analysis is built on top of this core framework. Tools to run the analysis at a high level, such as getting graphs in large batches, making plots, collecting summary information about events, etc., are stored in a header-only C++ library at my personal GitHub, and were originally started by Carl Pfendner. All of the header libraries have the naming convention `tools_ToolName.h`, with the title of the tool being self explanatory. For example, `tools_WaveformFns.h` contains utilities for manipulating and plotting waveforms, while `tools_Cuts.h` holds functions to perform cuts using high-level variables such as run times, reconstruction directions, etc. Most elements of this `tools_` analysis

Coordinate	Original Value	New Value
CP6 V X	17.882	21.082
CP6 V Y	35.616	42.2156
CP6 V Z	-173.289	-175.089
Ch6 X	-190.642	-191.242
Ch6 delay	199.517	206.317

Table 4.6: A table of the updated geometry for A2.

library are original to this analysis. The reconstruction algorithms are implemented in a secondary class, the `RayTraceCorrelator`, which was originally designed by Eugene Hong and Carl Pfendner. This tool is stored in a GitHub repository https://github.com/osu-particle-astrophysics/RayTraceCorrelator. To run this analysis, an analyst will need the `a23_2019_analysis` branch of the code, which includes updates to the interferometry described in Sec. 4.7, and some of the CW filtering tools described in Sec. 4.8.2. Full details of how to process the 10% "burn-sample" of the data are provided in App. B.4.

4.3.4 Recovering the Full Geometry Calibration of A2 and A3

In order for the analysis to proceed, we had to recover the full geometry calibration of both A2 and A3. The geometry for the stations had previously been calibrated during an analysis of the first year of data by Thomas Meures [91]. Due to time constraints, the corrections to the geometry were coded in-line to his analysis code, and were not formally included in `AraRoot`. We took these informal changes and added them correctly to `AraRoot`. The details of this process can be found in App. C.

In addition to the changes described in App. C, we make one final change to the geometry of A2. Ming-Yuan Lu found that we had trouble consistently reconstructing the location of calibration pulser CP6 in A2. He therefore recalibrated the location of the CP6 pulser, and adjusted the X coordinate and cable delay of Ch6 in A2. In total, we updated the (X,Y,Z) location of CP6V and updated the depth (Z) and delay of Ch6. Tab. 4.6 summarizes the changes. Note that because we cannot move CP6V without also moving CP6H (the two cal pulsers are in the same borehole), the (X,Y) coordinates of CP6H were also updated, but not the depth. We anticipate details concerning this process to be found in Ming-Yuan's thesis, which will be forthcoming this fall; in the meantime, documentation is available internal to the ARA collaboration at DocDB-1817.

4.4 Data Conditioning

Before ARA data can be used for analysis, the raw analog-to-digital converter counts recorded by the Data Acquisition System (DAQ) must be converted into voltages, and the data must further be conditioned. This conditioning is performed by the data analysis access framework `AraRoot`. We perform a series of three data conditioning steps:

1. Removing the First Block of a Waveform
2. Setting the Mean of the Waveform to Zero
3. Inverting Channels 0, 8, and 11 of A3

For the first time, we implemented these standard data conditioning steps in an `AraRoot` class titled the `AraEventConditioner` which automatically performs all conditioning steps. The user can access a list of conditioning steps performed to their data by calling it from the data pointer: `vector<string> list = ev->fConditioningList`.

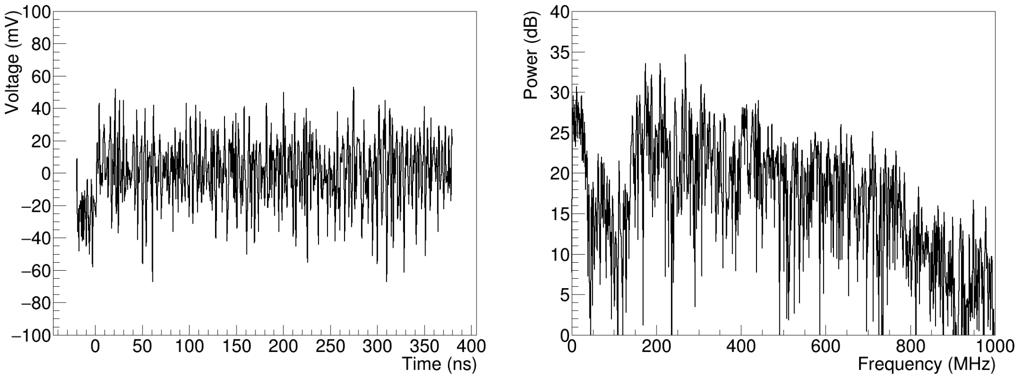
4.4.1 First Block Removal and Correcting the Waveform Mean

During the readout of the IRS2 digitizers, for some configurations, the first block (64 samples) of data recorded has a systematic offset of -20 mV to -30 mV. This can be observed in the time domain on the left side of Fig. 4.5a, and as artificial spectral power at low frequencies on the right side of the same figure. This offset at zero time also leads to an artificial enhancement of the correlation between waveforms at zero lag.

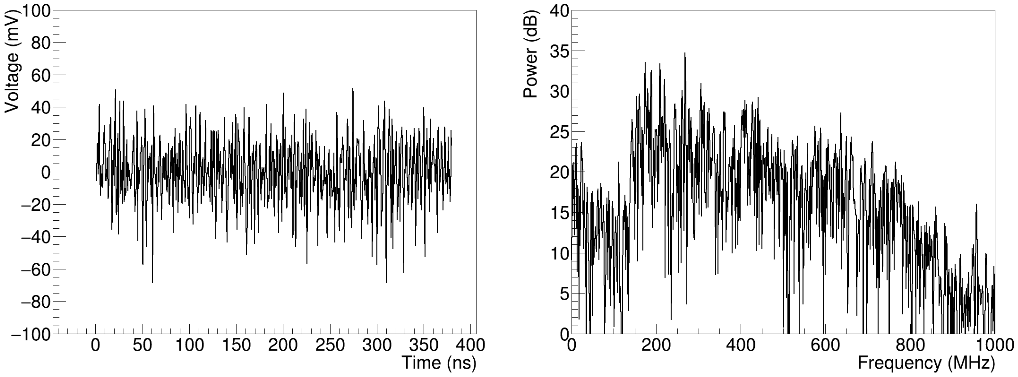
To fix this first block readout error, we remove the first block, and set the mean of the remaining blocks to zero. This second step is necessary because during the final step of the calibration process implemented in `AraRoot`, the mean of the *entire* waveform is set to zero. Because the first block is systematically offset by ~ -30 mV, the remaining blocks have an average positive offset of $\sim +2$ mV. This will lead to an artifical excess of power at low frequencies. After the first block trimming, and after setting the mean of the remaining blocks to zero, there is a reduction in the artificial low-frequency power, as can be seen in Fig. 4.5b. In the conditioner class, this is accomplished by the two functions `AraQualCuts::trimFirstBlock` and then `AraQualCuts::makeMeanZero`.C

4.4.2 A3 Waveform Inversion

Thomas Meures found during his studies [91] of A3 that channels 0, 8, and 11 exhibit inverted polarity relative to those expected. The source of this inversion is not known. To correct this inversion, we multiply channel 0, 8, and 11 waveform by -1. This is accomplished in the conditioner class by the function `AraQualCuts::invertA3Chans`.



(a) Before Correction



(b) After Correction

Figure 4.5: The waveform and power spectrum of an event with a first block readout error in A2. This is channel 5 of event 102 from A2 run 2818.

4.5 Data Quality Cuts

Before beginning the analysis, we apply a series of basic data quality cuts. These quality cuts are divided into two classes: for each run, and cuts applied for each event.

4.5.1 Run Level Cuts

Cuts that are applied to entire cuts are designed to remove periods of livetime where the station was either not operating in a normal mode or when the collaboration was conducting calibration activities. An example of a non-standard operating mode is when the trigger coincidence window of the station is set to an artificially small or large value to investigate the behavior of the triggering system. An example of a calibration activity is the carrying of a pulser system directly above the station or the operation of the IceCube deep pulser.

We choose to remove these livetime periods pre-unblinding because were a neutrino-like signal to be found during these periods, we would reject the event after unblinding because

A2 Excluded Run(s)	Reason
3, 11, 59, 60, 71	First runs after extended downtime
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
8100+	Consistently Misreconstructing CP5

Table 4.7: The list of excluded runs in A2.

we would assume that the event originated from something related to operations (e.g. sparks from the engine of a snowmobile or from a hand-held radio), potential extraneous behavior of the pulsing systems, or non-standard operating parameters in the DAQ system.

A table of the runs which we exclude in the analysis, along with an explanation for their exclusion, is provided in Tab. 4.7 for A2 and in Tab. 4.8 for A3. One large period of excluded runs is all runs with a run number greater than 8100. This was done because during this time period, we cannot robustly reconstruct the calibration pulser. Exclusion of this livetime removes 31 days of livetime from the analysis. In future analyses, further development in reconstruction algorithms will make the livetime analyzable. For example, my colleague Ming-Yuan Lu is able to analyze these last 31 days because his analysis relies on an iterative reconstruction process, where

To perform run exclusion in practice, we first build a `std::vector` containing a list of the bad runs. This list is constructed by calling `tools_Cuts::BuildBadRunList`. Afterwards, bad run identification is performed with the function `tools_Cuts::isBadRun`.

4.5.2 Bad Livetime Periods

In addition to the known periods of calibration activity discussed in Sec. 4.5.1, which result in the exclusion of entire runs from our analysis, we also exclude specific shorter periods of livetime which have identified anthropogenic activity. Identification of these time periods is currently a labor intensive process. We perform reconstructions (see Sec. 4.7), and examine

A3 Excluded Run(s)	Reason
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 IceCube Deep Pulsing
3977, 3978	2015 Surface Pulsing
6041	2015 ICL Rooftop Pulsing
3977	"Noisy Surface" Run

Table 4.8: The list of excluded runs in A3.

the distribution of events on the sky as a function of time. In several cases, structure can clearly be seen. An example of such structure is demonstrated in Fig. 4.6. What can be seen is the reconstructed azimuthal direction of events in A2 run 4775 over a 30 minute period. It can clearly be seen that a source rapidly crosses nearly the entire field of view of the instrument, starting at about zero degrees at 15:40 and ending at -100 degrees around 15:45 UTC. No true astrophysical source—whether it be neutrinos, or some flaring radio source like the sun—could sweep across the sky so quickly. So, we are able to identify this as not being astrophysically interesting, and most likely is a source of anthropogenic noise. In any case, we should exclude this livetime from the analysis, as we would reject an event in this period after unblinding for the same reasons we would reject an event taking during a calibration run period.

To locate all periods of bad livetime, three of my undergraduate colleagues, Alex Patton, Eliot Ferstl, and Scott Janse, examined such plots of reconstruction directions for all runs in the A2 and A3 data set (nearly 10,000 runs in total) and manually flagged such structures for livetime removal. A comprehensive table of removed livetime is provided in App. B.2 in Tab. 4.7 and Tab. 4.8. To check if an event is inside a period of excluded livetime, we use the analysis library function `tools_Cuts::isBadLivetime`.

Effective Livetime for A2

The total accumulated livetime for A2 is given in the left column of Tab. 4.9, while the total good livetime that is used for the diffuse analysis is given in the middle column. What fraction of the total livetime is "good" is provided in the rightmost column. We are able to analyze ~98% of our total recorded livetime, which is a substantial improvement over the 62% that was able to be analyzed in the prototype analysis.

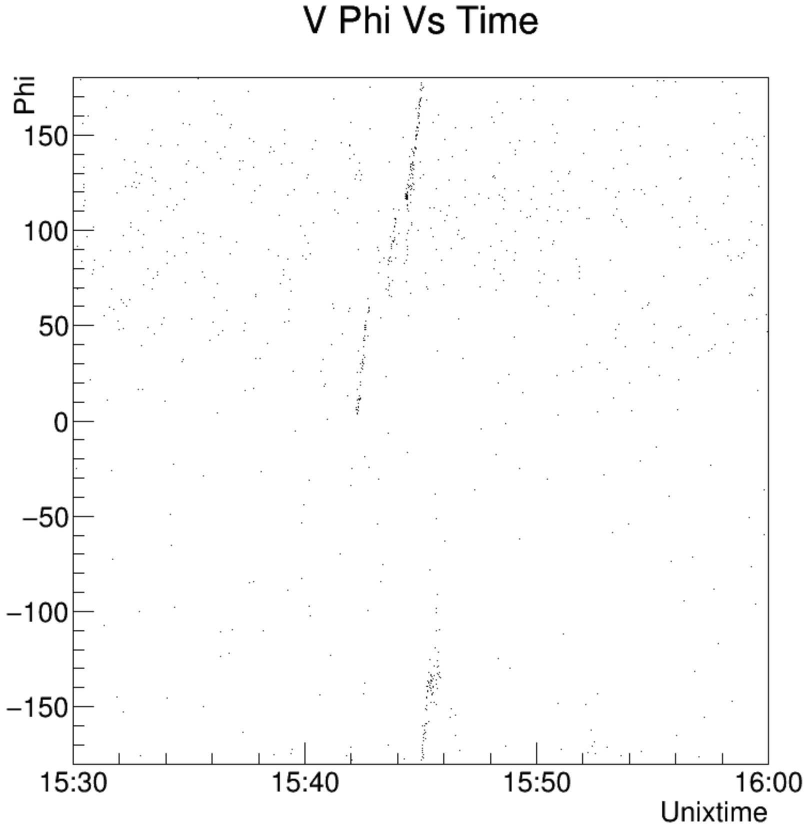


Figure 4.6: A plot of the reconstruction direction of events in run 4775 over the course of 30 minutes, clearly showing anthropogenic activity.

4.5.3 Event Level Quality Cuts

Cuts applied to individual events are designed to remove two classes of events. The first are events that enter the data set because the DAQ system itself triggered them. The second class of events are corrupted due to digitizer or system readout errors in the DAQ.

The first class of events, those that enter the data set because the DAQ itself triggered them, consist of the local calibration pulsers and software triggers. Local calibration pulsers are removed through the `AraRoot` function `AraRoot::AraEvent::isCalpulserEvent`, which was designed by the broader ARA collaboration. The function cuts events based on a timing criterion. The firing of the cal-pulser is triggered by the PPS (pulse-per-second) clock on the station. It is observed in the data that the calibration pulser event will be read out within the first 10,254 100 MHz clock cycles after the PPS. So the function removes any event which occurs within 10,254 clock cycles. Software triggers are removed by checking the triggering information about the event. If the event is a software trigger, the third

Config	Total Livetime (days)	Good Livetime (days)	Percentage Good
1	185.08	179.62	97
2	143.58	143.57	99
3	100.07	94.45	94
4	413.01	409.86	99
5	265.73	263.76	99

Table 4.9: The amount of livetime that was recorded, and was good for analysis, in A2.

triggering bit (bit 2) of the event metadata will be set "high" (to a value of 1). This is implemented in `AraRoot` through the function `AraRoot::AraEvent::isSoftwareTrigger`.

The second class of events, those that enter the data set because of DAQ errors, are removed by the following cuts. Most of these are new to this analysis, and represent previously unknown classes of errors in the ARA data, and were found by myself and my colleague Ming-Yuan Lu:

1. Short Waveform Cut
2. Timing Error Cut
3. Block Gap Cut
4. First Five Events Cut
5. Offset Block Events Cut
6. Bad Spare Channel Cut

For the first time, we implemented these standard data quality cuts in an `AraRoot` class titled `AraQualCuts`, which automatically performs the relevant checks and returns a "good" or "not good" for analysis flag to the user. The quality cuts tool can be accessed in the following way:

```
AraQualCuts *q = AraQualCuts::Instance();
q->isEventGood(UsefulAtriStationEvent *ev)
```

We point out that this particular class of readout error could in principle be caught at the hardware level in real time by the acquisition software, and the station could be programmed to recover by, for example, resetting its data taking buffers.

Short Waveform Cut

We cut any event with a waveform that has fewer than 64 samples. This is because a correctly read out IRS2 digitizer block should contain 64 samples, and all correctly read out events should have more than one block. So if a waveform contains fewer than 64 samples, it should not be analyzed. In the quality cuts class, this is implemented by the function `AraQualCuts::hasTooFewBlocks`.

Additionally, to reject any untagged software triggers or other waveform readout errors, we reject any event with any waveform that is shorter than 500 samples; this cut was chosen by previous analysts Carl Pfendner. This cut is partially redundant with the cut on events with fewer than 64 samples in a waveform described above, but is observed to identify some additional readout errors.This cut, relative to the distribution of all events, can be seen in Fig. 4.7. A cut at 500 samples separates the software triggers in VPol (just left of the red line in the top eight panels) from the beginning of the HPol RF triggered events (just right of the red line in the bottom eight panels). Why the distribution of events has so much variance is not known. Some variance can be explained by the analysis team doing test runs with longer readout windows than others; for example, when operating the IceCube deep pulser, it can be desirable to have a longer readout window because of the presence of a direct and reflected/refracted pulse.

Timing Error Cut

We cut any event with a waveform that has misordered timing samples. That is, if at any time, the $i+1$ sample has a smaller time than the i-th sample: $t_{i+1} < t_i$. Such a waveform would be acausal, and should not be analyzed. This error was discovered by attempting to interpolate such events, which will fail if the sample timing is not monotonically increasing. The cause of this error is not known, but seems to be an error in the sample timing calibration process, as it is not observed in the raw data as best we can identify. In the quality cuts class, this is implemented by the function `AraQualCuts::hasTimingError`.

Block Gap Cut

We cut any event with a waveform that does not have contiguous block readout. For a correctly readout ARA event, the blocks are digitized contiguously. If the hardware requests five blocks to be read-out, and read-out began on block 25, then the five read-out blocks should be [25, 26, 27, 28, 29]. Any other event, e.g. [25, 26, 27, 40, 41], will be acausal across the block boundary, and cannot be analyzed. Therefore, we cut any event for which `final_block != first_block + num_readout_blocks`. The cut also accounts for the "wraparound" case where an event spans the boundary between block 512 (the last block in the capacitor array) and block 0 (the first block in the capacitor array). In the

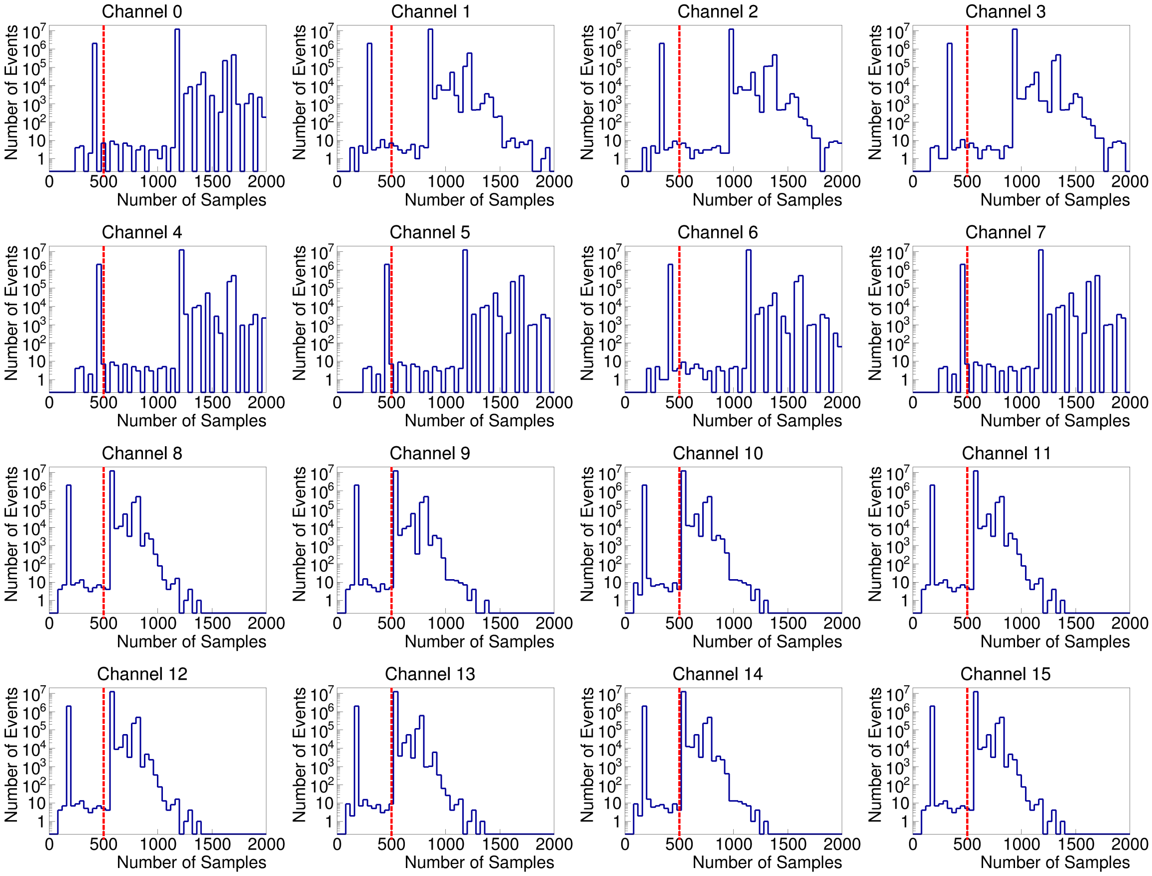


Figure 4.7: Distributions of the number of samples per waveform for all events in the A2 2013 data sample.

quality cuts class, this is implemented by the function `AraQualCuts::hasBlockGap`.

First Five Events Cut

We cut any event in A2 or A3 which occurs in the first five events of a run. We generally observe the first five event corruption for A2 after November 25, 2015. In A3, this error is present beginning in 2014. An example of a corrupted event is shown in Fig. 4.8. In the quality cuts class, this is implemented by the function `AraQualCuts::hasFirstEventCorruption`. The precise cause of this error is not known, though hardware experts have suggested that it is due to incorrect resetting of the data taking buffers between the end of one run and the beginning of a new run.

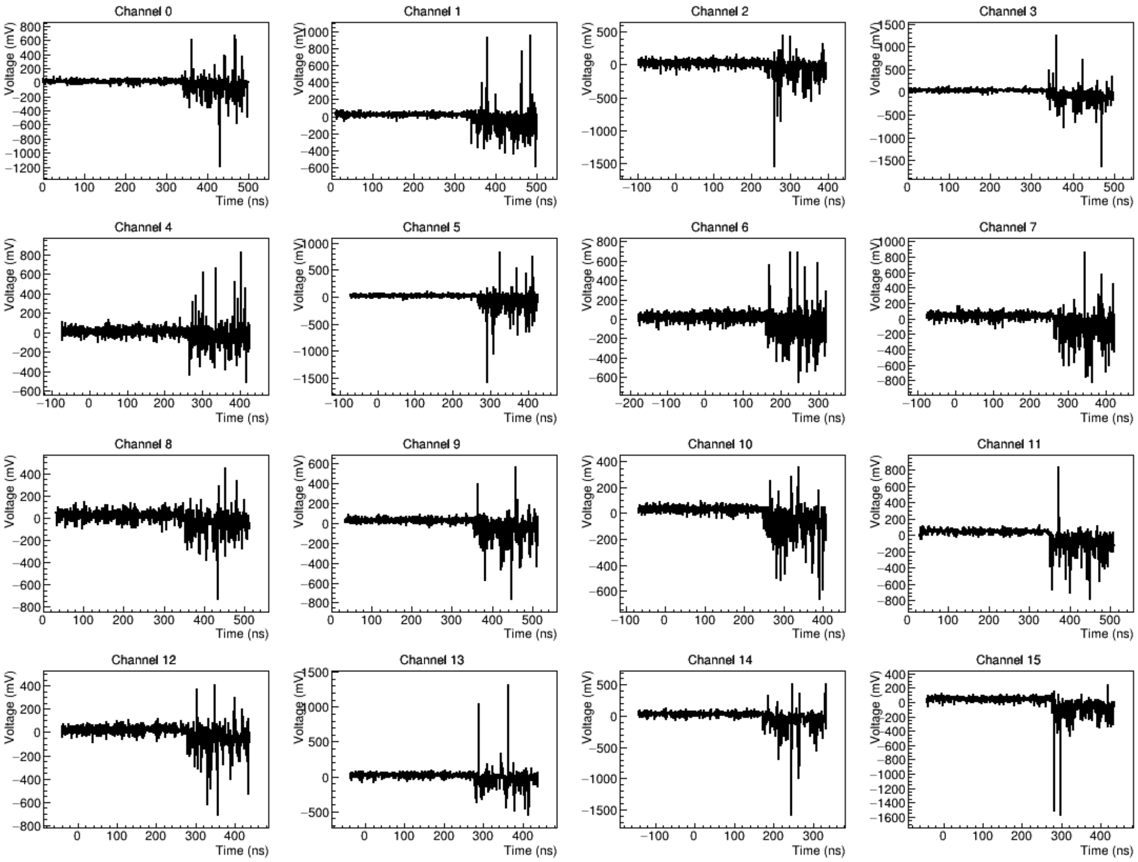


Figure 4.8: The sixteen RF channel waveforms exhibited often by the first four events read out on A2 after November 25 2015. This is A2 run 6570 event 3.

Offset Block Cut

We cut any event with a digitizer error that presents as a single digitization block in the middle of the waveform having a non-zero mean, that is, is "offset". Such an event is presented in Fig. 4.9. The criterion for identifying these events was designed by Ming-Yuan Lu, and is implemented in AraRoot through the function `AraQualCuts::hasOffsetBlocks`.

We perform this cut by first computing a rolling 64-sample mean (the length of one block) of each graph. If this rolling mean exceeds 20 mV in VPol or 12 mV in HPol, we flag that time in the waveform as having a potential offset block. For the event to qualify as an offset block event, we then require that in two strings, both VPol channels and at least one HPol channel all have such an excursion within 10 ns. If this criterion is met, the event is flagged as containing an offset block. The values for this cut were chosen empirically by my colleague Ming-Yuan Lu to remove events which stood out in the analysis; a more complete discussion is expected when his thesis is completed later this year.

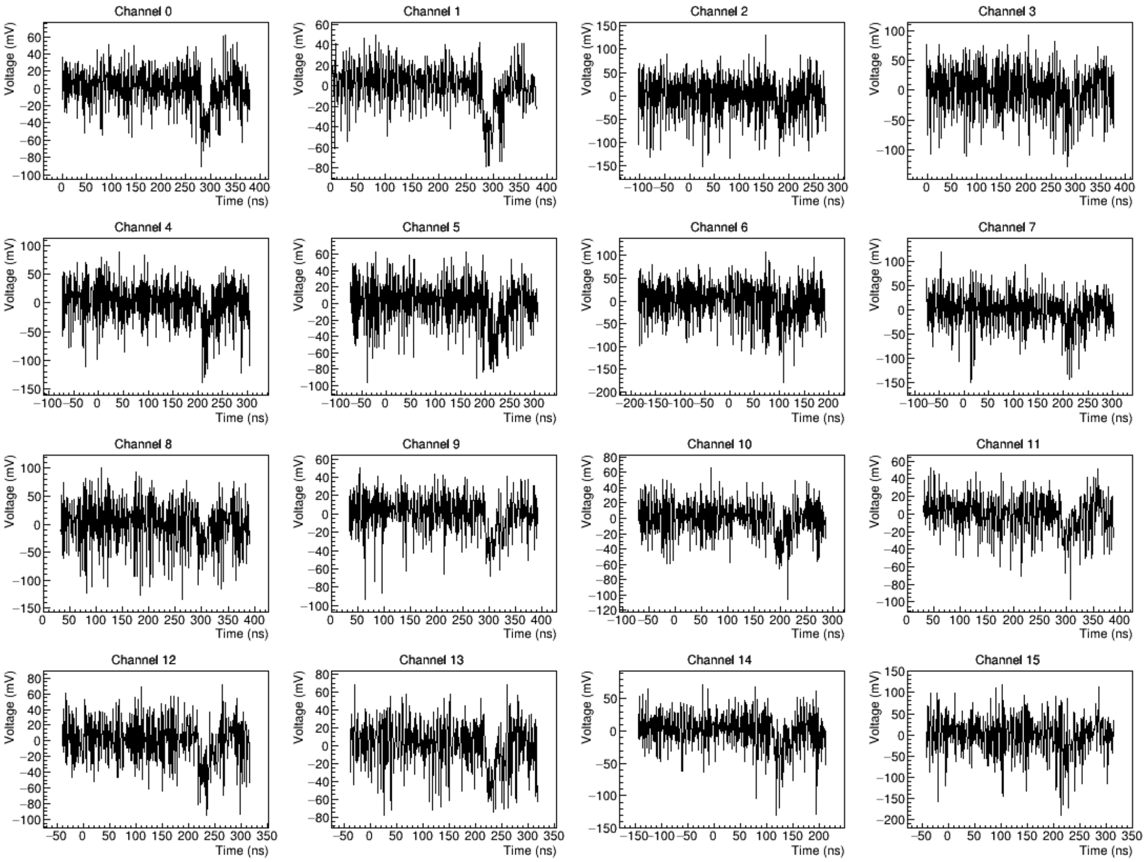


Figure 4.9: The sixteen RF channel waveforms of an event which has a digitizer error which manifests itself as a block mid-waveform that is systematically shifted relative to the remaining blocks. This is A2 run 2823 event 95294.

Bad Spare Channel Cut

We cut any events that exhibit a large root-mean-square (RMS) voltage in the terminated channels of the ARA digitizer boards. The use of the spare channels to identify digitizer glitches is new to this analysis. As discussed previously in Sec. 4.2.2, there is one spare channel per DDA, the sixth channel, which is terminated with 50Ω. As such, the digitizer for this channel should be recording zero, with some small noise floor which includes cross-talk from the neighboring clock channel. When a digitizer glitch occurs, it will generate a large RMS in this channel, which can be identified. Identifying glitches in this fashion in the spare channel is easier than in the other digitizer channels because the noise floor will be much lower than in the first five channels, which are connected to antennas. The first five channels will have the noise of the amplifier and filter chain, making the RMS of the channel a value that is typically between 30 and 50 mV. In the spare digitizer channel the average RMS is approximately 4 mV. Removal of events with a bad spare channel is done with the function `tools_Cuts::hasSpareChannelIssue`. We check the spare channels in the first three DDAs, and cut an event if in two of three spare channels the RMS exceeds 20 mV. We also cut an event if in any spare channel the RMS exceeds 60 mV. We only check the first three because we expect to not have a functioning fourth DDA for 75% of the livetime in A3.

We find that these last five glitch identification techniques remove less than 0.07% of the events from the data set, representing a negligible reduction in the efficiency of the analysis.

4.6 Thermal Noise Wavefront RMS Filter

The ARA station records data at a rate of 5 Hz, accumulating over 157 million events per year of livetime. In this data set of $\mathcal{O}(10^8)$ events, we expect to observe at most a handful of neutrino candidates. As such, nearly all events ($> 99.9999\%$) are thermal noise fluctuations. To reduce this data set, we apply a first-level thermal noise event filter called the *wavefront-RMS* filter. The goal of this filter is to reduce the number of events to be analyzed with more computationally intensive interferometric reconstruction algorithms as described in Sec. 4.7.

4.6.1 Design and Performance

The wavefront-RMS parameter is new to this analysis and was originally designed by Carl Pfendner. It takes advantage of the regular geometry of an ARA station. The roughly cubical shape is composed of 12 edges, where the edges are the baselines between antennas at the corner of the cube. For example, there are four baselines roughly parallel to the local $\hat{x}$ unit vector, which we might designate as A-type baselines and label as A_i, where $i = 0, 1, 2, 3$. A schematic of two different kinds of baselines, with the A-type labeled, can be seen in Fig. 4.10. We could call the four parallel to the $\hat{y}$ unit vector B-type, etc. Each baseline pair is designated a face. For a station with eight fully operational channels per polarization, there are a total of twelve faces: the four side faces of the cube, the top and bottom, and six slanted faces. In A2, where Ch15 is non-operational, the number of faces is reduced to six. In A3 for 2014-2016, where an entire string is removed, the number of faces is three.

For each antenna in a baseline A_i, we calculate a sliding five-nanosecond power envelope of that antenna's waveform. We find the two highest peaks of the envelope for both antennas, and compute the time delay between the first peaks in the graphs, between the second peaks in the graph, and their cross combination. In order to qualify as a peak, we require that the size of the peak be an SNR of 2.0 or greater. Because of ray-bending in the firn, simulation shows that ARA will occasionally observe a direct and refracted/reflected signal from the same physics interaction; this is why we compute all four timing delays, as opposed to just

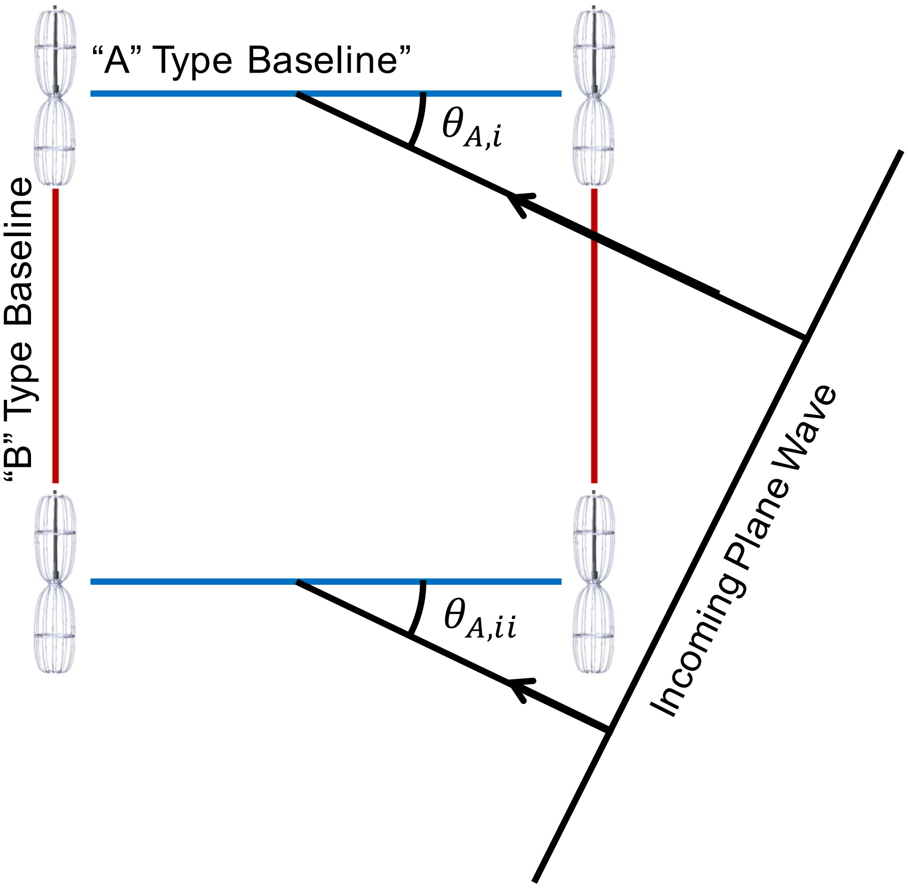


Figure 4.10: A drawing of two baseline types in the wavefront-RMS filter.

the delay between the two higher peaks. For a given baseline A_i, there are therefore four contributed delays $\Delta t_{A,i,j}$, where j designates a specific computed delay, running from 1-4, for the i-th A-type baseline.

Each $\Delta t_{A,i,j}$ can be converted into an incident angle $\cos(\theta_{A,i,j})$ through equation 4.1, where c is the speed of light in vacuum, $n = 1.78$ is the index of refraction of deep South Pole ice, and $\Delta d_{A,i}$ is the distance between the antennas

$$\cos(\theta_{A,i,j}) = \frac{c\Delta t_{A,i,j}}{n\Delta d_{A,i}} \tag{4.1}$$

Utilizing the cosine of the incident angle, as opposed to the time delay itself, has the benefit of being independent of baseline length.

Each baseline pair, or *face*, for example A_i and $A_{i'}$, will have four mean $\overline{\cos(\theta_{A_i,A_{i'}})}$

(where the bar denotes a mean) as defined in Eq. 4.2

$$\overline{\cos(\theta_{A_i,A_{i'}})} = \frac{\cos(\theta_{A_i}) + \cos(\theta_{A_{i'}})}{2} \tag{4.2}$$

and also four RMS values about the four means, ξ=RMS$\big(\overline{\cos(\theta_{A_i,A_{i'}})}\big)$ defined in Eq. 4.3.

$$\xi = \mathrm{RMS}\big(\overline{\cos(\theta_{A_i,A_{i'}})}\big) = \sqrt{\frac{\Big(\cos(\theta_{A_i}) - \overline{\cos(\theta_{A_i,A_{i'}})}\Big)^2 + \Big(\cos(\theta_{A_{i'}}) - \overline{\cos(\theta_{A_i,A_{i'}})}\Big)^2}{2}} \tag{4.3}$$

This *wavefront-RMS* (ξ) should be small for an Askaryan neutrino pulse, as we expect $\cos(\theta_{A_i}) \approx \cos(\theta_{A_{i'}})$. The opposite is true for a thermal noise fluctuation which should on average generate random timing delays, resulting in comparatively larger values. For a given event and polarization, we cut on the smallest wavefront-RMS of any face.

In Fig. 4.11, we show the distribution of wavefront-RMS values for VPol (left) and HPol (right) in A2 configuration 1. In blue, we show the distribution of RF-triggered events, and in red, the tagged calibration pulsers, the bulk of which are clearly visible at a lower value of ξ. The peak of events at $\xi = 3$ are events which did not meet the SNR threshold requirement. In A2, we require $\log_{10}(\mathrm{RMS}(\cos(\theta))) < -1.3$ in VPol and < -1.4 in HPol.

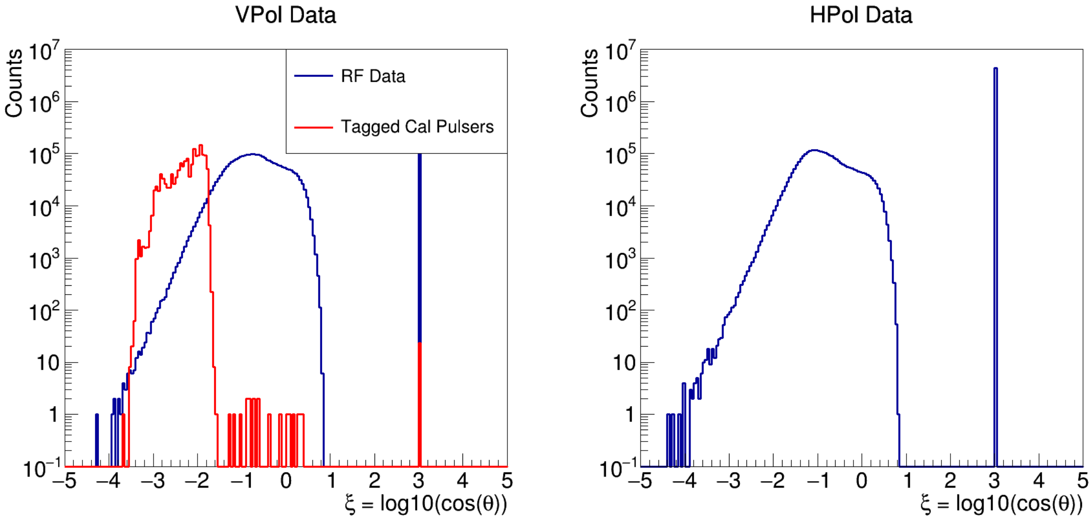


Figure 4.11: The distribution of wavefront-RMS parameter in A2 configuration 1 data in VPol (left) and HPol (right), with RF-triggered background events in blue, and tagged calibration pulsers in red.

Distributions in this variable for a flux of simulated neutrinos is shown in Fig. 4.12.

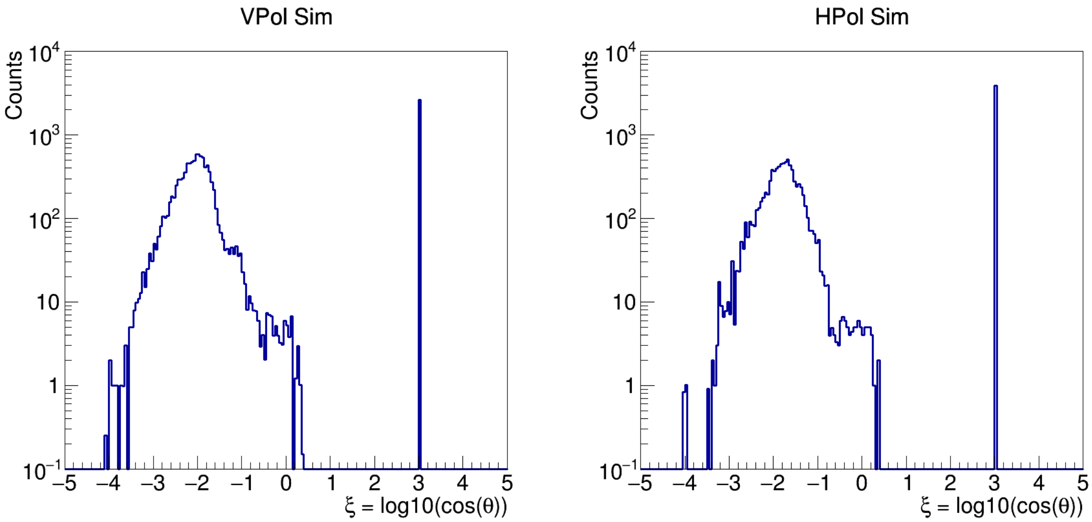


Figure 4.12: The distribution of wavefront-RMS parameter for A2 configuration 1 simulation in VPol (left) and HPol (right).

These neutrinos were generated from a distribution expected from a benchmark cosmogenic neutrino flux by *Kotera et. al.* [17], with energies ranging from 10^{17} to $10^{20.5}$ eV. For simulation, it is particularly important to know the *efficiency* of the filter. The efficiency is defined as the fraction of neutrino events which pass the filter. In particular, we are interested in the efficiency of the filter as a function of the signal-to-noise ratio. We are concerned about the efficiency-vs-SNR because we anticipate that we should be able to efficiently recover events with a high-SNR, while we expect events with a low-SNR to be difficult to both reconstruct and confidently identify as being different from thermal-noise fluctuations. The SNR of the event is defined by the third-highest V_{peak}/RMS, which is the same as what was employed in previous Testbed analyses. For all eight same-polarization antennas, we compute the peak in their waveform V_{peak} and their voltage noise RMS. We then select the antenna in the set with the third highest V_{peak}/RMS. As the station requires a 3/8 antenna coincidence for a trigger, the third strongest V_{peak}/RMS sets the threshold for whether the event triggered or not, and is thus expected to be a sensitive variable.

The efficiency of the wavefront-RMS filter as a function of SNR for signal events is presented in Fig. 4.13. We find that on average the filter has an efficiency of 75% in VPol and 60% in HPol, with a combined total efficiency of 90%. The efficiency of the filter broken out by configuration and polarization is presented in Tab. 4.10.

We point out that the threshold values governing this filter—the SNR inclusion threshold of 2.0, and the cut values of -1.3 and -1.4—are not fine tuned, and a future analyst could include them in an optimization. We chose these values with the goal of reducing the size of the data set to be analyzed while rejecting less than 5% of neutrino from the aforementioned

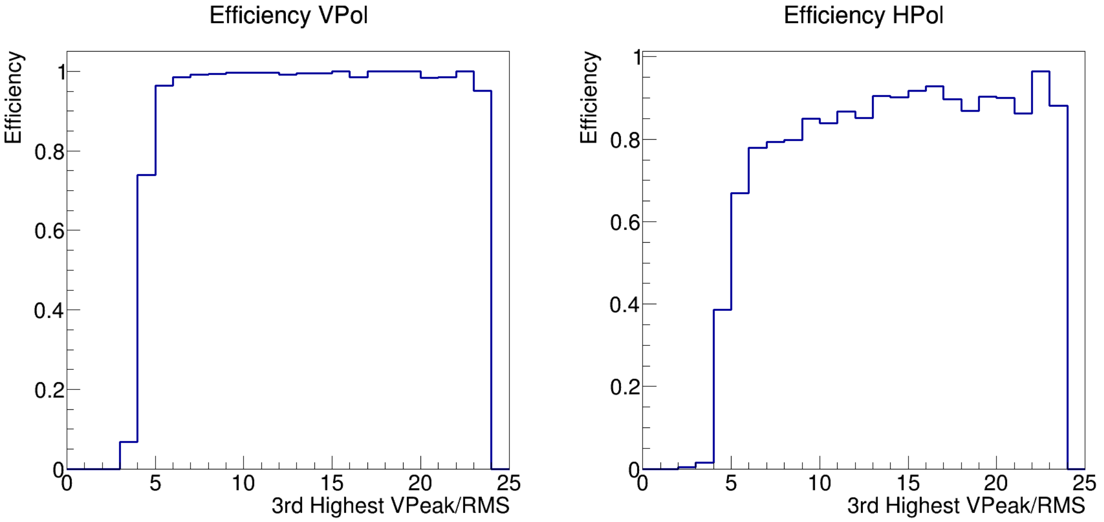


Figure 4.13: The efficiency of the wavefront-RMS parameter as a function of SNR, where the SNR is defines as the third-highest VPeak/RMS, for simulated neutrinos in A2 configuration 1 in VPol (left) and HPol (right).

Config	VPol Efficiency	HPol Efficiency	Efficiency to Pass VPol or HPol
1	74.7	58.0	89.9
2	69.8	48.1	85.2
3	75.6	58.1	91.1
4	75.0	58.7	90.4
5	76.4	59.4	91.7

Table 4.10: A table of the efficiency of the wavefront-RMS filter for the five different data taking configurations in A2.

Kotera spectrum of events with this cut alone. We find that with these loosely tuned values, we accept between 5-10% of events in the dataset per polarization, while achieving the requisite fraction of events cut as last cut.

4.6.2 Code Implementation

In code, the process begins by extracting copies of the sixteen waveforms of an event through the function `tools_PlottingFns::makeGraphsFromRF` and interpolating them with the function `tools_PlottingFns::makeInterpolatedGraphs`. To construct the rolling-mean graphs, we use the function `tools_WaveformFns::makeIntegratedBinPowerGraphs`. To convert these values to a signal-to-noise ratio (SNR), we divide by the run-average RMS of the RF triggers in the run. We find the peaks in these integrated SNR graphs, along with their hit times, using the function `tools_WaveformFns::getAbsMaximum_N`.

Now that we have the waveforms set-up to be analyzed by the filter, we must construct a list of all faces that are to be considered by the filter. This is accomplished with the function `tools_WaveformFns::setupFaces`. We pass this list of faces, the list of heights of the SNR peaks, and the list of hit times to the function `tools_WaveformFns::getRms_Faces_Thresh_N`, which will return the wavefront-RMS parameter for each face under consideration.

4.7 Interferometric Event Reconstruction and Geometric Cuts

4.7.1 Interferometric Reconstruction

For events that pass the wavefront-RMS filter in either polarization, we perform an interferometric-based reconstruction [92] to determine the direction of the source of measured incoming radio waves. This interferometric reconstruction technique has been used in other ARA analyses [36, 53, 93, 94] and in the ANITA experiment. The interferometric technique relies on the fundamental relationship connecting (1) the location of an emitting source in space and (2) the time delay between when that signal will be observed in two physically displaced measurement antennas.

For a given pair of antenna waveforms, the cross correlation between the voltage waveform on the i-th antenna (V_i) and the voltage waveform on the j-th antenna (V_j) as a function of lag τ can be found in Eq. 4.4:

$$C_{i,j}(\tau) = \frac{SNR_i \times SNR_j \times \sum_{t=-\infty}^{\infty} V_i(t) V_j(t+\tau)}{N_{overlap} \times RMS_i \times RMS_j} \tag{4.4}$$

where the SNR_i and SNR_j are the signal to noise ratio of the waveforms contributing to the correlation function. The SNR in this case is defined as the highest V_{peak} over the run-average RMS. The lag τ defines the the time delay of one antenna waveform relative to the other. The lag τ depends on the position of the source emitter relative to the array center, characterized as an elevation angle (θ), an azimuthal angle (ϕ), and a distance to the source (R). The array center is defined as the average position of all sixteen measurement antennas in the station. The total cross-correlation strength for a given point on the sky is given by summing over all like polarization pairs of antennas as in Eq. 4.5:

$$C_{sky}(\theta,\phi;R) = \frac{1}{\sum_{i=1}^{n_{ant}-1}\sum_{j=i+1}^{n_{ant}} SNR_i \times SNR_j} \sum_{i=1}^{n_{ant}-1} \sum_{j=i+1}^{n_{ant}} C_{i,j}[\tau(\theta,\phi;R)] \tag{4.5}$$

The factors of SNR in the denominator serve to normalize the SNR weighting in $C_{i,j}$. The cross-correlation function for an individual pair of antennas, $C_{i,j}$, is expected to be maximal when the lag is equal to the true difference in the arrival times of a signal at the two different antennas. The sky map is therefore expected to have a peak at the putative source direction. We take the peak summed correlation value in this interferometric map to be the location of the RF source. An example interferometric map is presented in Fig. 4.14; the map has a peak near the expected calibration pulser location of $(\theta, \phi) = (5.9°, 65.2°)$.

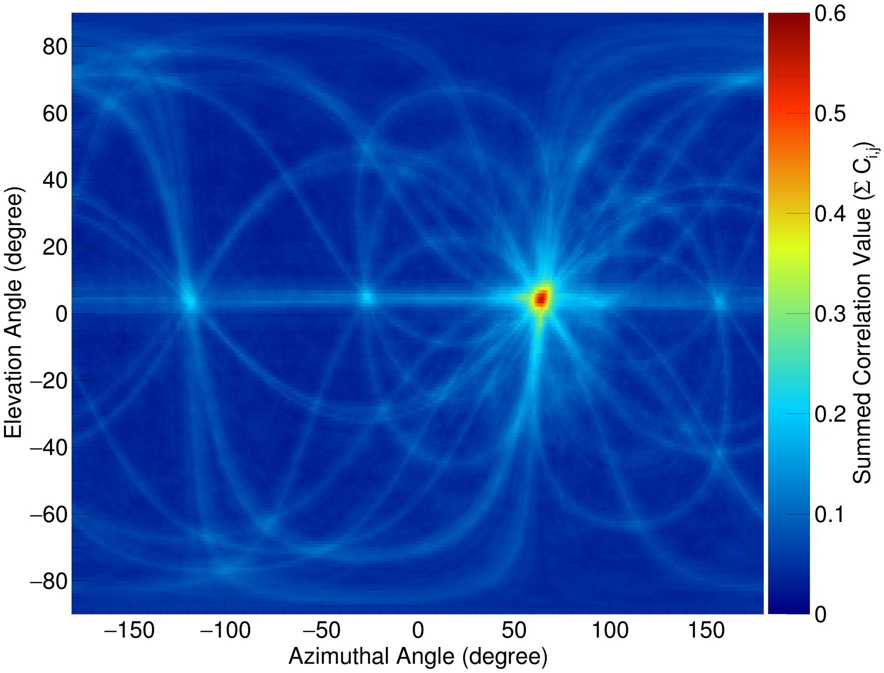


Figure 4.14: An example of an interferometric skymap for a calibration pulser in A2.

The time lags τ for a given point on the sky θ, ϕ are computed by calculating the path a light ray would take from a hypothesized source location to an antenna. The calculation accounts for the changing index of refraction of the Antarctic firn, which causes the light rays to follow curve trajectories, not straight one.

The ray-tracing method models the changing index of refraction as follows, where $n(z)$ is the depth-dependent index-of-refraction, and z is the depth from the ice surface:

$$n(z) = 1.78 - 1.35e^{0.0132z} \tag{4.6}$$

We consider the index to be unity above the surface. This index of refraction model was developed by the RICE experiment based on data collected in Antarctica [95]. To smooth uncertainties in this ice model and other systematics (such as differences in the phase responses of the various contributing antennas), we calculate the Hilbert envelope of the cross-correlation function before summing over pairs, as is done in previous analyses.

We perform the reconstruction for two radius hypotheses: 41 m and 300 m. The 41 m map is used exclusively to reject untagged calibration pulser events, as described in Sec.4.7.3. We select 41 m because that is the actual approximate geometric distance between the station-center and the calibration pulsers. The 300 m map is used as a far-field proxy. We select 300 m because for A2 and A3, this is the largest radius for which all θ,ϕ combinations on the sky have a ray-tracing solution. For larger radii, some sky points become classically inaccessible due to ray-bending in the firn. We therefore selected a distance radius for which all points on the sky may contribute.

4.7.2 Improvements to the Single Pair Correlation Function $C_{i,j}$

This single-pair correlation function $C_{i,j}$ has two changes which are improvements from the correlation function used in previous analyses. Firstly, the function is weighted by the product of the signal-to-noise ratio (SNR) of the contributing waveforms. In this way, channels with higher SNR contribute more strongly than others. Second, the correlation function has a factor of $N_{overlap}$ in the denominator. This is the number of samples overlapping between V_i and V_j for a given lag τ. This is important because the correlation function without this factor of $N_{overlap}$ is an inherently biased estimator of the similarity between the two graphs. This arises because the numerator will grow with the number of overlapping samples, but the denominator would otherwise be a constant. This leads to an envelope in the amplitude of the correlation function. The factor of $N_{overlap}$ serves to reduce this bias. The inclusion of the factor of $N_{overlap}$ is a typical option in other digital signal processing tools, such as MATLAB [96].

The need for this revised correlation function, particularly the factor of $N_{overlap}$, was noticed when examining the distribution of reconstruction locations on the sky using the default correlation normalization in the `RayTraceCorrelator`—that is, it has no correction for $N_{overlap}$. In Fig. 4.15, I show the distribution of reconstructed directions on the sky, on the left for VPol and on the right for HPol. It is immediately clear that the distribution of events is *not* isotropic, as would be expected for random thermal noise fluctuations. In fact, it is clear that (1) the events preferentially reconstruct to a stripe in $\theta \sim -42°$, and (2), one side of the sky appears "hotter", centering near $\phi \sim 100°$.

We have reasonable hypotheses for both sources of anisotropy in the reconstruction direction. Evidence suggests the anisotropy in azimuth is related to an asymmetry in the deployed hardware: when A2 and A3 were both deployed, a mistake led to the fiber-optic cable on one string being made 100 m too long. In A2, that is string 3, and in A3, that

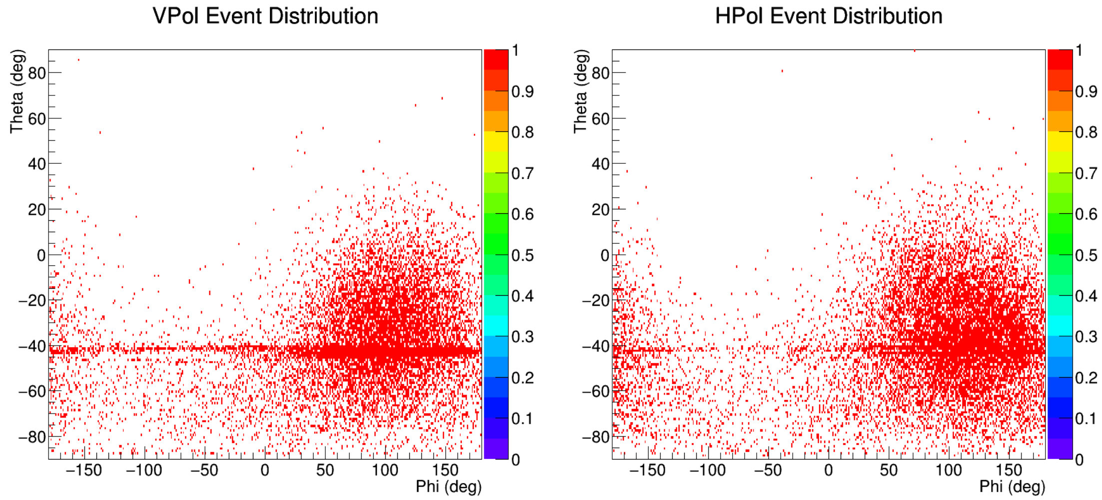


Figure 4.15: The distribution of reconstruction directions on the sky utilizing the default waveform correlation function for run 1449. VPol is on the left, and HPol is on the right.

is string 2. This means that, in A2 for example, all of the waveforms for string 3 must always be translated ~ -100ns in time, such that they begin earlier than the other three strings. In cases where a large signal dominates in the waveform, e.g., a calibration pulser, the correlation function is still able to identify a peak at the appropriate lag. However, when the thermal noise dominates, as is the case in the vast majority of events, the extra trace length at the beginning of string 3 means that all of the waveforms on string 3 appear to *systematically* lag all the other strings, and this pulls the direction of the reconstruction toward the azimuth of the longer string. In A2, this means we expect the hot-spot to be observed at the ϕ of the longer string, which is 111°, precisely the location of the observed excess. In A3, which is not shown here, the excess is observed near a ϕ of 30°, which is near the expected location of 21° for A3 string 2.

The anisotropy in zenith—the stripe at $\theta \sim -42^\circ$—we suspect is also related to the hardware. The top and bottom antennas are separated by $\sim$ 19m of cable, in which light travels 0.255m/ns, amounting to $\sim$ 75 ns of delay between the two. In the example of A2 D1TV and D1BV, the known geometric distance between the two antennas is 19.26 m. If we assume that a hypothetical signal has a time delay between the antennas of 75 ns, and solve for the angle of arrival ($\theta = \sin^{-1}(Dc\Delta t/n)$),would expect what signal to arrive from there at -41°. As such, we conclude that there must be unaccounted for cross-talk between the top and bottom antenna pairs on each string. Indeed, when we compute the average correlation function between the top and bottom antennas, we observe an artificial enhancement at 75 ns that we cannot otherwise account for; this can be seen clearly in Fig. 4.16 which was prepared by my colleague Ming-Yuan Lu. We hypothesize that like with the azimuthal bias, in cases where the signal strength is high, it will dominate the correlation function. But when just thermal noise is present, this leads to a systematic bias in the correlation function, which pulls the reconstructions in the direction of $\theta \sim -42°$.

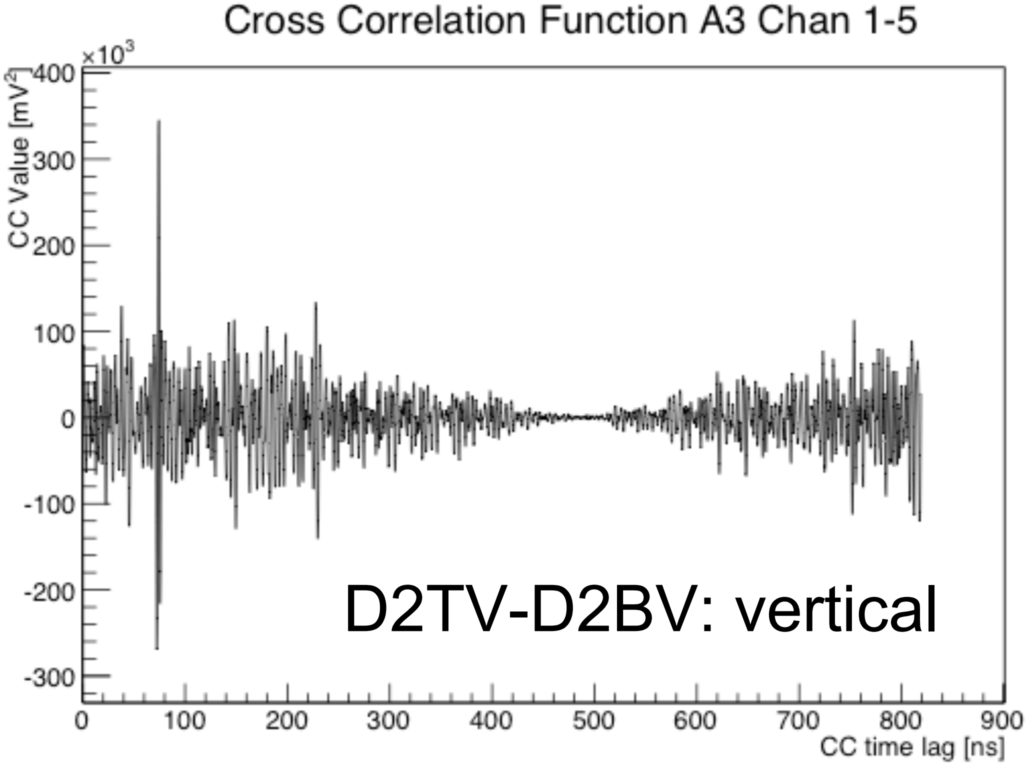


Figure 4.16: The average correlation function between D2TV and D2BV in A3, showing a clear bias for 75 ns. This plot was made by Ming-Yuan Lu.

Our new, proposed correlation function, which addresses the factor of $N_{overlap}$, reduces the anisotropy in both azimuth and zenith. This can be seen in Fig. 4.17. Qualitatively, the bias in azimuth has essentially been eliminated, and quantitatively, the bias in zenith has been reduced by roughly a factor of two. With the old normalization, we found that 38.7% of events in VPol, and 20.9% of events in HPol, reconstructed inside the zenith bias stripe between $-45°$ and $-39°$; with the new correlation function, we have reduced this to 20.8% and and 10.2% for VPol and HPol respectively. We can note also that a much higher fraction of events reconstruct above the surface, which would be expected if the events were actually distributed randomly.

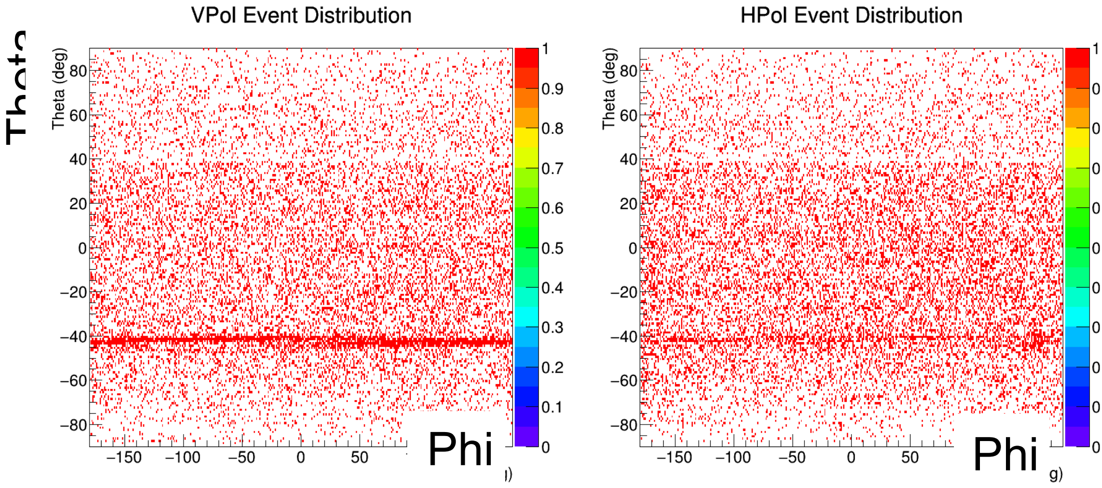


Figure 4.17: The distribution of reconstruction directions on the sky utilizing the new and improved correlation function for run 1449. VPol is on the left, and HPol is on the right.

4.7.3 Geometric Cuts

After reconstructing an event, we implement two geometric cuts that reject an event if it reconstructs to a particular region of the sky. First, in the 300 m map, we reject an event if it reconstructs to or above the surface in either polarization: $\theta_{300m} >= 37°$. We make this cut because we expect neutrino events to interact in the ice, and therefore the radio waves from the neutrinos should reconstruct within the ice. The cut angle of 37° is arrived at through geometric arguments. The intersection of the 300 m source sphere, centered on the station, intercepts the surface at an elevation angle of 37°.

Secondly, we reject an event if in the 41 m map, in either polarization, it reconstructs near the expected location of the calibration pulser systems. There are two calibration pulsers: one is deployed on string five, and is designated CP5, and one is deployed on string six, and is designated CP6. The expected geographic locations of the calibration pulsers are given in Tab. 4.11. We point out that these reconstruction directions are in the local coordinate system of the station, as described in App. B.1.1.

The cut regions for θ and ϕ are presented in Tab. 4.12. By comparing with the expected locations in Tab. 4.11, we can see that the cut regions are essentially defined as 10° wide boxes in azimuth and 15° wide boxes in zenith, centered on the expected locations of the calibration pulser region. The precise boundaries of these regions are not yet precisely tuned, and will be updated before unblinding so that the number of events expected outside the cut box will be an order of magnitude lower than the other expected backgrounds. The exception is the region identified as the “CP5 Mirror”, which is exclusive to A2 in configuration 5. Both Ming-Yuan Lu and myself identify this spot on the sky as a place to which tagged

Pulser	Geometric ϕ	Geometric θ
A2 CP5 H	-25.02	-19.53
A2 CP5 V	-25.02	-23.56
A2 CP6 H	65.19	9.76
A2 CP6 V	65.19	5.85
A3 CP5 H	-23.55	-11.70
A3 CP5 V	-23.55	-16.10
A3 CP6 H	65.53	-9.84
A3 CP6 V	65.54	-14.28

Table 4.11: The expected locations of the calibration pulser system in A2 and A3.

Pulser	ϕ Cut Region	θ Cut Region
A2 CP5	-30 to -20	-29 to -14
A2 CP5 Mirror	-28 to -18	0 to 15
A2 CP6	60 to 70	30 to 41
A3 CP5	-28 to -18	-20 to -5
A3 CP6	60 to 70	-20 to -5

Table 4.12: The calibration pulser rejection regions for A2.

calibration pulser events consistently mis-reconstruct due to errors in the trigger settings of the station during this time.

An example of a distribution of tagged calibration pulser events is presented in Fig. 4.18 for A2 configuration 2. We can see that for this configuration and pulser, we are able to reconstruct the tagged calibration pulser events to within a tight distribution a few degrees wide, and that is centered to within a degree or so of the expected true location of the calibration pulser. A broader selection of the distribution of cal-pulser reconstruction directions for A2, including the CP5 "Mirror", is available in App. B.3.

4.7.4 Code Implementation

The reconstruction algorithms are stored in the `RayTraceCorrelator` class originally created by Eugene Hong and Carl Pfendner. To generate a reconstruction sky-map, we use the function `RayTraceCorrelator::getInterferometricMap_RT_select_NewNormalization_SNRweighted`. The use of the function with the `_select` designation allows for the explicit exclusion of channels which are known to be corrupted, e.g., Ch15 in A2. This function will return a ROOT `TH2D` two-dimensional histogram object. We then identify the peak correlation value and the location of the peak using the function

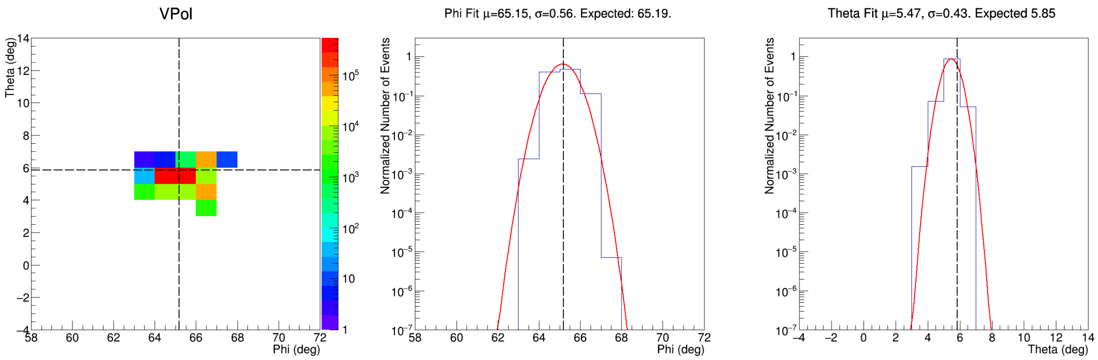


Figure 4.18: The spatial distribution of all tagged calibration pulser events reconstructing in the direction of CP6 in A2 config 2. The left panel shows the distribution of events in θ,ϕ space. The middle panel shows the normalized projection of this distribution along the ϕ axis. The right panel shows the same for the θ axis. The dashed lines mark the expected geometric locations of the calibration pulsers.

`tools_RecoFns::getCorrMapPeak_wStats`. Identification of the calibration pulser is accomplished with the function `tools_Cuts::identifyCalPulser`.

4.8 Continuous Wave (CW) Identification and Removal

We would like to address the issue of continuous-wave interference in the ARA data. The CW emission we are concerned with is anthropogenic (human made) in origin, and is characterized by a strong spectral peak at one or a few frequencies within a relatively narrow band. The most common type of CW encountered in ARA is generated by the ~403 MHz radiosonde attached to NOAA weather balloons that are launched once or twice daily from the South Pole. The presence of CW is generally easily identified visually as a strong spectral peak in the Fourier transform of an event. An example of a 400 MHz balloon event can be seen in Fig. 4.19, where the green circles surround the CW peak. To guide the eye, we also plot in red the average FFT of all events in the run.

In this section, we address the two ways that we identify CW contamination in the dataset. We then explain the method we use to remove the CW contamination from events, a process we achieve by using a filtering technique developed by Brian Dailey for the ANITA experiment at Ohio State.

4.8.1 CW Identification

We identify CW frequencies for filtering using two techniques. The first identifies narrow peaks in the spectrum over an established baseline. The second identifies stability in the phasor at a given frequency between neighboring events.

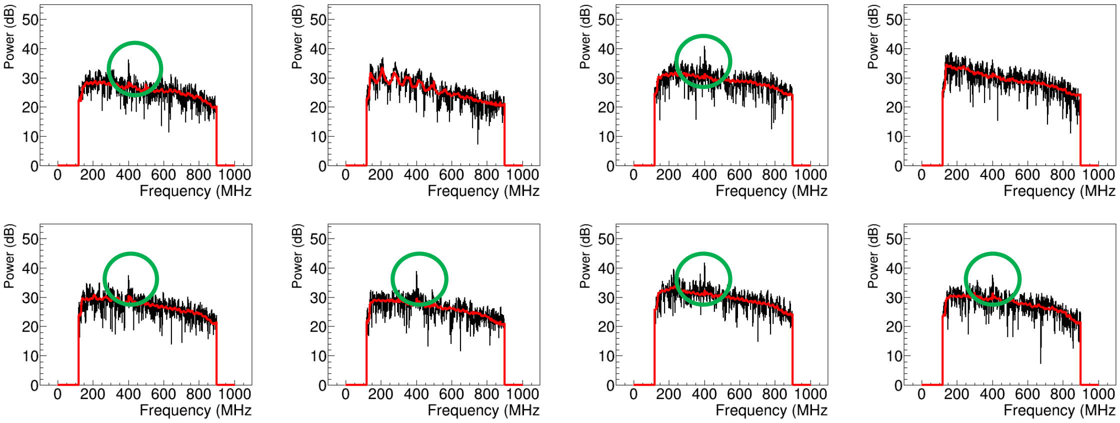


Figure 4.19: An example of a 400 MHz CW contaminated event. These are the FFTs of Ch0-7 (the VPol channels) in event 20695 in run 1548. The green circles highlight the clear CW peak, while the red lines are the average FFT of all events in the run.

Peaks Over Baselines

The first technique utilized for identifying CW emission identifies narrow spectral peaks over established baselines. This technique has been previously relied upon in analyses of ARA prototype data [53]. To build the baselines, we take the average FFT of all radio-frequency events that trigger the station in a given run, excluding software triggers and calibration pulsers. We exclude software triggers because they have a different waveform length than RF triggers, and therefore require different quantities of zero-padding in the FFT and necessarily have less overall spectral power. We exclude calibration pulsers, as they have clear non-thermal spectral content. We only form the baselines between the frequency of 150 MHz and 850 MHz, which is approximately the bandwidth of the ARA system. Fig. 4.20 demonstrates how we form the baseline by averaging together 20 events in run 1501 (black lines), and the average they form (red line). The low-level oscillations in the baseline are taken to be an instrumental feature; they are thought to be an artifact of the ARA data calibration process. ARA data is calibrated in 20 ns data segments, and therefore has breaks across boundaries every 25 MHz or so, which is roughly the observed frequency of the oscillation.

We flag a frequency for filtering if the spectral peak in that frequency bin is both *high and coincident* and *narrow*. A frequency is tagged as *high and coincident* if in a channel, the spectral power exceeds 6 dB over the baseline, and if three other channels of the same polarization also have a peak 6dB above the baseline within 5 MHz. The frequency is tagged as *narrow* if in a single channel, less than 50% of the bins in the neighboring 40 MHz have peaks that are 5.5 dB over the baseline. These values are modeled after those selected in

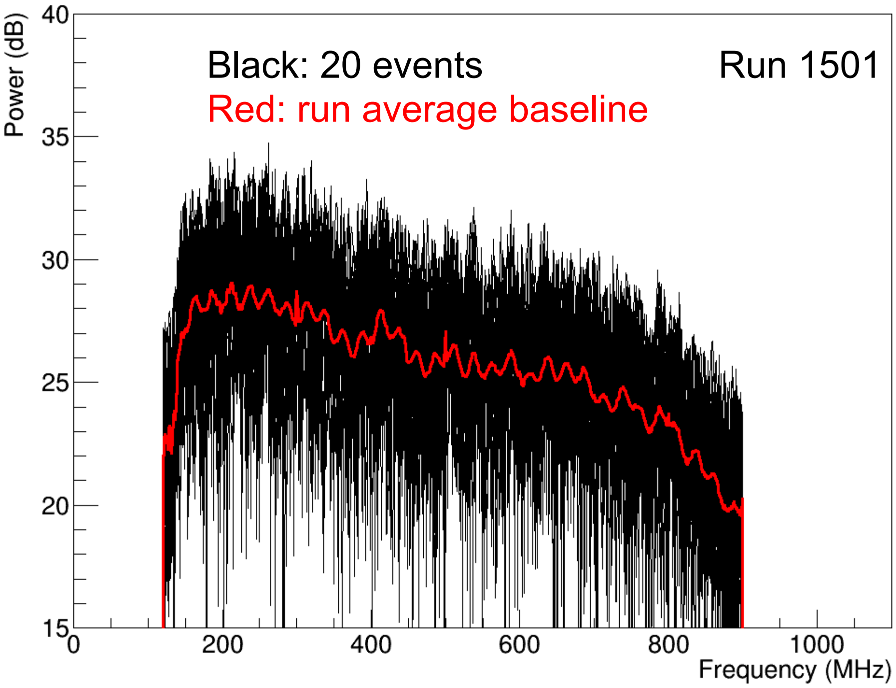


Figure 4.20: An example of forming the spectral baseline for run 1501 by averaging FFTs of events in the run. In black are 20 RF event FFTs, and in red is the baseline they form.

the ARA prototype analyses (the prototype analysis had used power excess values half a dB higher than what are used here).

In code, these frequencies are identified by using the function `tools_CW::CWCut_TB`, which takes as arguments: the FFT of the event in question, the baseline averages for the run the event is drawn from, the height of the CW peak, the number of channels for coincidence, and a list of channels not to check for CW contamination (such as Ch15 in A2 or string four in A3).

Phase Variance

The second method we use to identify CW frequency, termed the *phase variance* technique, is inspired by work from the LOFAR collaboration[97] and was originally designed by Carl Pfendner. The logic of the phase variance filter is as follows: if a continuous-wave source has entered the field-of-view of the instrument, and is illuminating a pair of antennas, it should arrive at an antenna with roughly the same relative FFT phase (φ) in the CW frequency bin across neighboring events. So, the difference between the phases of two antennas, $\varphi_2 - \varphi_1$, should be a constant between neighboring events if the CW is present. As such, if we add up the phase difference $\varphi_2 - \varphi_1$ across neighboring events, as in Eq. 4.7

$$V = 1 - \frac{1}{N}\left|\sum_{k=1}^{N}\exp(i(\varphi_2 - \varphi_1))\right| \tag{4.7}$$

we expect to recover a constant value. In this equation, N is the number of neighboring events, φ_1 is the phase in one channel, and φ_2 is the phase in a second channel. This behavior for CW is different than what is expected for thermal noise, whose phasor sum should approximate a random walk. Fig. 4.21 illustrates the concept of the phase variance identification method diagramatically. For every event we consider, we compute both a "forward-looking" and "backward-looking" phase-variance. In the "forward" case, the event in question is the first of fifteen events which are used to construct V. In the "backwards" case, the event is the last of fifteen used to to construct V.

Fig. 4.22 shows a plot of the phase-variance for two cases. On the right, is the phase-variance during a "quiet" uncontaminated period. On the left is a period which is contaminated with 400 MHz balloon noise, which can be seen as a clear deviation in the phase variance at the 400 MHz frequency bin. As can be seen on the right, the phase variance technique additionally identifies other frequencies, even in the period when we do not observe obvious peaks in the amplitude spectrum. In fact, we observe that in *most* events, the phase-variance technique identifies three frequencies for filtering: 125 MHz, 300 MHz, and 500 MHz. Experts on the construction and operation of the DAQ have suggested that these frequencies could be subtle contamination that is generated by the electronics in the DAQ itself. For example, there is a 100 MHz clock which maintains synchronization between all digitizer and readout boards. This clock could leak into the digitizer channels, and we could be observing harmonics of this clock at 300 MHz and 500 MHz. In any case, we do not mark an event as requiring filtering if the only frequencies identified as being CW are 125 MHz, 300 MHz, or 500 MHz because they do not stand out as spectral peaks and cause misreconstructions in the same way as the 400 MHz balloon events.

To turn the phase variance into a value that can be cut on, we want to know how atypical the phase variance in a given frequency bin is compared to the other frequency bins. To accomplish this, we compute the median of the distribution of phase variances for a single event, subtract the phase variance in every frequency bin from the median, and divide by the width of the phase variance distribution. This essentially tells us how many standard deviations away from the median of the distribution of phase variances a given frequency bin is. We flag a frequency bin for filtering if the 15 event average has a deviation more than 1.5 standard deviations from the median. This value is not fine tuned, and was observed to identify CW events that stood out in the analysis.

In code, the phase as a function of frequency of a waveform is calculated using the function `tools_CW::getFFTPhase`, and the difference between the phases of two waveforms

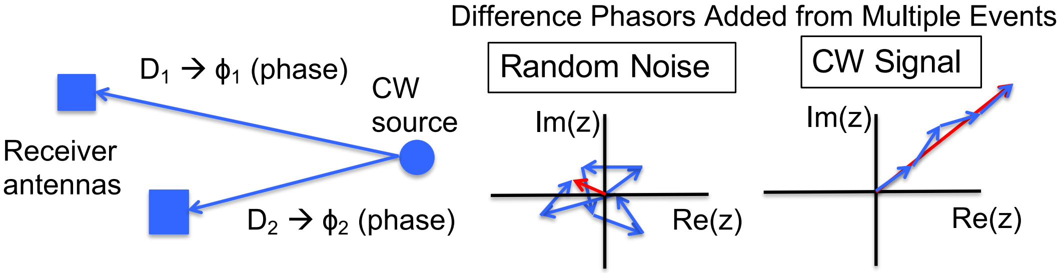


Figure 4.21: A schematic showing the concept of the phase variance CW frequency identification technique. Figure originally designed by Carl Pfendner.

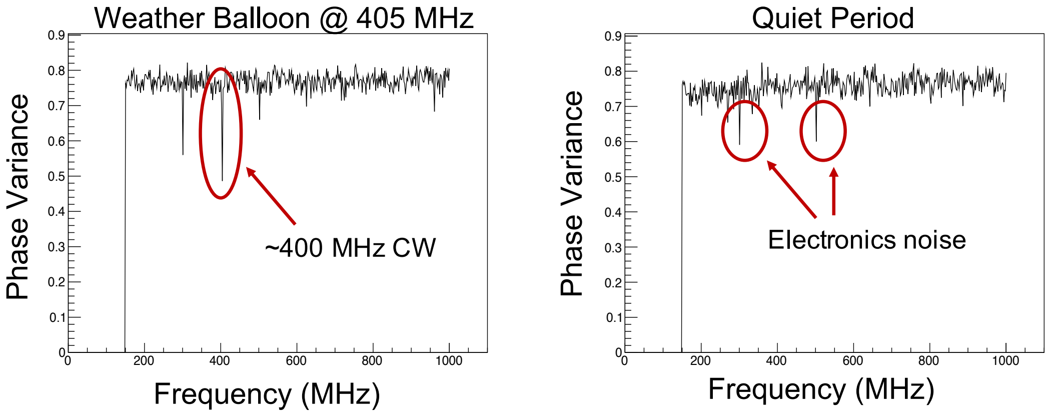


Figure 4.22: A plot of the phase variance for a 400 MHz CW event (left) and for thermal noise (right) in A2.

is calculated with `tools_CW::getPhaseDifference`. The phase differences are then used as inputs to the function `tools_Cw::getPhaseVariance`, which computes the phase variance itself. This function returns the average multiple of sigma away from the median each frequency bin is for the fifteen-event average; to compute the median and sigma of the distribution, we use the function `tools_CW::getMedian`.

4.8.2 CW Filtering: The ANITA Geometric Filter

Once frequencies have been identified by either the peaks-over-baseline or the phase-variance technique, they are filtered from the waveform by use of the ANITA *geometric filter*. This filter was originally designed by Brian Dailey here at Ohio State [98], and has been used in searches for a diffuse flux of neutrinos in the ANITA-II and ANITA-III experiment [99]. The geometric filter works on frequencies that are placed inside *notches*, or frequency ranges that are identified to have CW contamination. Before being notched, individual frequencies are first combined into notches. If two or more frequencies are within 10 MHz of one another, they are combined into a single notch. If a frequency bin is identified for filtering and is "standalone"—with no other frequencies nearby—it is given a 6 MHz wide notch. The function `RayTraceCorrelator::pickFreqsAndBands` performs this identification process.

As a reminder, when taking the Fourier transform of a waveform, we calculate the amplitude and phase spectrum for that waveform. The Geometric Filter works by notching frequencies in amplitude spectrum, and adjusting the phases within the notch. The majority of the source code for the Geometric filter was found in the ANITA `AnalysisFramework` GitHub repository, which was originally coded by a combination of Brian Dailey and Sam Stafford. I imported this filter into the ARA analysis framework.

The filter takes place in two stages. In the first stage, the contaminated spectra have their amplitudes notched. This is accomplished by simply interpolating across the notch using an Akima interpolation scheme. This is done with the function `RayTraceCorrelator::interpolatedFilter`.

In the second stage, the Geometric Filter adjusts the phase of the FFT within the notch. Full details of this process are available in Brian Dailey's thesis [98]. In brief: the algorithm attempts to identify the magnitude and phase of the contaminating CW phasor, and then solves for the phase had the contaminating phasor been absent. It then sets the phase to this reconstructed value. This is accomplished with the function `RayTraceCorrelator::GeometricFilter`.

An example amplitude and phase spectrum for a CW contaminated event which has had the Geometric Filter applied can be seen in Fig. 4.23. The notched amplitude spectrum can be seen in the lower left panel, and the "repaired" phase can be seen in the lower right panel. To protect against cases where the filter removes excessive quantities of the original power in the waveform, we reject an event if the third-largest amount of power removed in a graph exceeds 6% of the original power. We chose the cut of 6% by looking at the distribution of fractions of power cut by the filter, and removing only a few percent of the total events notched by the filter.

After an event has been filtered of its contaminating frequencies, it is reconsidered by the wavefront-RMS filter. Additionally, we reconstruct the event again using the filtered waveforms, and again re-check the geometric cuts described in Sec. 4.7.3. The reconstruction of the filtered events is accomplished with the function `RayTraceCorrelator::getInterferometricMap_RT_FiltMany_select_NewNormalization_SNRweighted`. We state that the events are re-checked only because in the analysis as is it currently coded, events are reconstructed in a "first-round" before being filtered and reconstructed in a "second-round" post-filtering. The pre-filtering reconstructions are only used as an initial

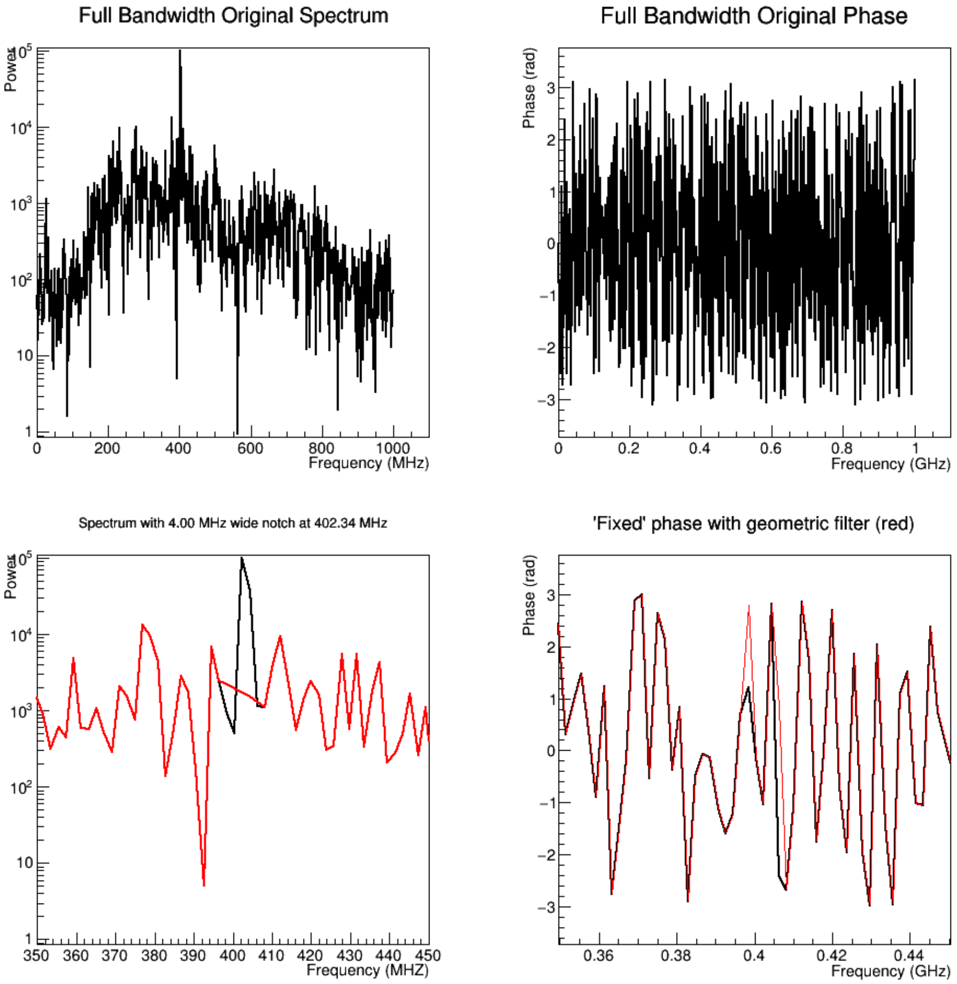


Figure 4.23: An example of a CW contaminated event which has been put through the Geometric Filter, in particular, run 1548 event 20695.

rejection of the calibration pulser region on the sky. The post-filtering reconstructions are used to again check if the event is inside the cal pulser geometric cut box and to check the surface cuts. A improvement for the analysis would be to only check the calibration pulser cut region after the events have been filtered.

4.9 Conclusion

In this chapter, I presented the analysis methods for station A2. In the following chapter, I will describe the optimization of the final cut value, and present the expected limit of the analysis.

Chapter 5

Diffuse Neutrino Search: Optimization

5.1 Introduction

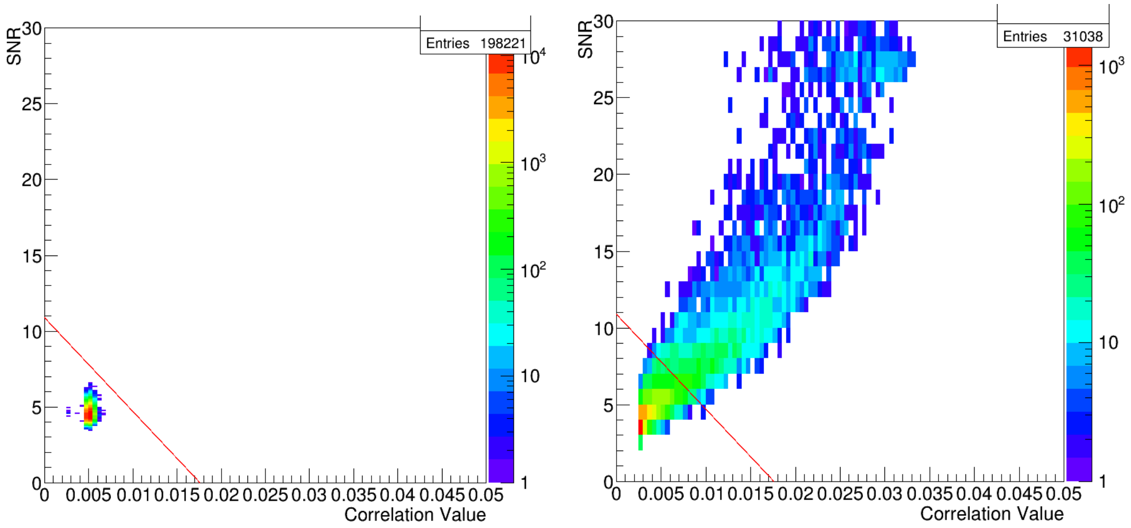
In this chapter, I present the optimization of an analysis of data from A2 from the years 2013-2016. In the previous chapter, I described the analysis methods. In this chapter, I will present the optimization of the final cut values, which will be the slope and intercept of a slanted line in the cross-correlation vs SNR plane. I will also present the expected number of background events, the efficiency of the analysis on simulated neutrinos, and an expected limit.

5.2 Final Cut and Background Estimate

5.2.1 The Rotated Cross-Correlation Cut (RCut)

After applying the wavefront-RMS filter, removing CW contamination from contaminated events, and rejecting events that reconstruct to the direction of the calibration pulser and reconstruct above the surface, we must set a final cut. We set this final cut in a two-dimensional parameter space. The first parameter is the third-highest V_{peak}/RMS, which was introduced previously in Sec. 4.6.1. The second parameter is the peak value of the correlation map, C_{sky}, that we computed in Sec. 4.7. The final cut is intended to separate the background events from signal events.

The final cut is drawn as a slanted-line in this SNR vs correlation plane, which we term the "RCut". An example of such a line in the 2D plane is shown in Fig. 5.1. The z-axis marks the number of events in the plane, in this case for all of the data at this stage in the analysis in A2 configuration 1. Notice that in the left plot, the background events are isolated in the lower left corner. However, in the right plot, many neutrinos extent out and to the right of the line, to higher SNR and correlation values.



(a) A2 config 1 data.

(b) A2 config 1 simulated Kotera neutrinos

Figure 5.1: The two-dimensional final cut parameter space for A2 in configuration 1 in the VPol channel;. Data is on the left, while simulation is on the right.

To set this final cut, we must select both a slope for the line m, and a y-intercept d. An event will therefore pass this cut if it has an SNR which meets the criteria in Eq. 5.1:

$$SNR \geq C_{sky} \times m + d \tag{5.1}$$

Because of existing constraints on the flux of high-energy neutrinos, we do not anticipate making an observation with this dataset. As a result, we want to optimize our final cut to set the strongest possible upper limit on the flux of neutrinos. The background estimate and choice of the final cut line are intimately connected in this analysis. This is because the number of background events admitted by the analysis directly impacts the strength of the limit we are able to set. Given that we expect to have at most a few signal events in the entire dataset, we intuitively expect our number of backgrounds to be well below unity. The question is, by how much? We are able to answer this question by performing an optimization which will explicitly select the cuts which generate the strongest expected limit by considering both the expected number of background events and the expected number of signal events simultaneously.

Choosing the final cuts is more complicated for this analysis than in previous analyses because we must in fact find *ten* cut parameters, not two, as we can in principle have every livetime configuration i have its own slope m_i and y-intercept d_i. More importantly, we cannot optimize the configuration bins separately as five different experiments and combine the results at the end. This will lead to a weaker overall limit than if we optimize all cuts simultaneously and set one limit leveraging the sensitivity of the five configuration bins together.

The reason why optimizing each configuration separately is non-optimal becomes more obvious if we perform a thought experiment. Consider what would happen if the livetime of each configuration were to become very short—for example, one day per configuration, with more than 1000 configurations total. Were we to optimize each of these experiment separately, each would find an expected number of background events per configuration that would be a fraction of an event. But the total number of backgrounds expected would then be a fraction of an event times 10^3, which is far more than the single signal event which might be in the data set.

Later, we will describe our prescription for locating these optimized cut values, taking into account the relative background and signal contribution from each configuration. But first, to understand the optimization procedure and its fundamental components, let us consider the procedure for optimizing the cuts for the best limit within a single configuration bin i only.

5.2.2 Optimizing the RCut within a Single Configuration Bin

Background Estimation in a Single Configuration Bin

First, let us consider the question of how to estimate the expected number of background events b. To estimate the number of background events, we generate a model for the background which follows the distribution observed in the data. If we make a plot of the differential number of events rejected by the RCut as a function of y-intercept, we find that the distribution is consistent with an exponential. This can be seen in Fig. 5.2. In this particular case, I have assumed a slope for the RCut line of $m = -620$.

The fit to the tail of the differential distribution is an exponential of the form $e^{\beta_1 \times SNR + \beta_2}$, where on a log-linear plot, β_1 sets the slope of the line and β_2 sets the intercept. The fit is performed from half-way down the tail (that is, midway between the maximum filled bin and the last filled bin), to two beyond above the last filled bin. In this example for A2 configuration 1, which is shown in Fig. 5.2, we find $\beta_1 = -5.96$ and $\beta_2 = 58.29$. The figure is not the final fit result for A2 config 1, as we adjusted the precise fitting routine as described in this text after this figure was made, and so we do not provide the uncertainties on the fit parameters here.

We then must evaluate the goodness of fit. We do so by computing the binned log-likelihood $-2\ln(L)$:

$$-2\ln(L) = 2\sum_{i=1}^{N}\left[\mu_i - n_i + n_i \ln\frac{n_i}{\mu_i}\right] \tag{5.2}$$

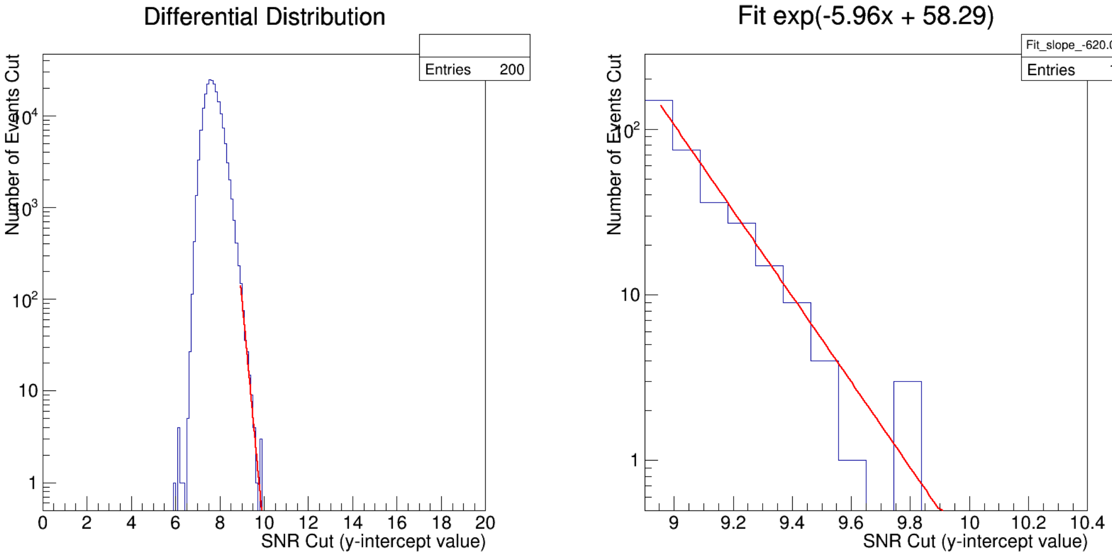


Figure 5.2: The "differential distribution" of number of events cut in A2 configuration 1. The right panel shows the tail of the distribution on the left.

where N is the total number of bins in the likelihood estimation, μ_i is the expected number of backgrounds predicted by the exponential distribution in an SNR bin i, and n_i is the observed number of events in that SNR bin. For this fit, we find a log-likelihood of -14.69. To place this log-likelihood value in context, we compute an ensemble of 10,000 pseudo-experiments, each of which follows the exponential distribution provided by our model, and each of which is filled with number of events drawn from a Poisson distribution predicted by our background model. For each one of these toy pseudo-experiments, we again compute the log-likelihood.

The distribution of the log-likelihoods for these 10,000 pseudo-experiments can be seen in Fig. 5.3. To evaluate the goodness of fit, we find the p-value p, which is the fraction of pseudo-experiments with smaller likelihoods than the one that we observed. We require $p > 0.05$ to be considered reasonable. This is because a p-value less than 0.05 is usually considered the threshold for a statistically significant non-typical result. For our example, we find a p-value of 0.40, which is reasonable, and suggests that our observation is typical.

Finally, having established that our model for the background is reasonable, we can estimate the number of events that will pass our cuts for a selected value of y-intercept d. This number of events will be the background b, and can be found by integrating our model for the background from our cut value of d to infinity:

$$b = \int_{d}^{\infty} e^{\beta_1 \times SNR + \beta_2} = -\frac{1}{\beta_1} e^{\beta_1 d + \beta_2} \tag{5.3}$$

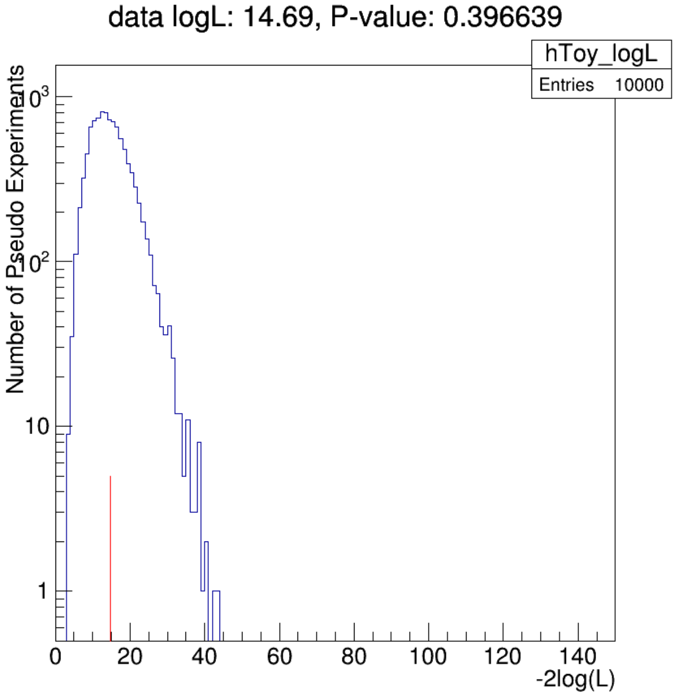


Figure 5.3: The distribution of log-likelihoods for the 10,000 pseudo-experiments modeled on A2 configuration 1, where the red line is the p-value we observed in the data.

We additionally multiply our background estimate from the 10% burn sample by a factor of 10 to account for the additional number of events we will expect in moving to the 100% sample, which is ten times larger. We multiply by 10, and not 9, because in ARA our “full” data sample contains the 10% burn sample, and so we must account for the 10% of livetime already analyzed and the 90% sample has so far been blinded.

Cut Optimization in a Single Configuration Bin

We are making a measurement, and would like to know the 90% confident level (CL) upper limit on the number of signal events that a model might produce and still be consistent with our observations of backgrounds b in the absence of signal. The probability for detecting n events total given s signal events and a known number of background events b is given by the Poisson distribution:

$$P(n|s) = \frac{(s+b)^n}{n!}e^{-(s+b)} \tag{5.4}$$

We can find the upper limit on the number of signal events S_{up} by using Bayes' Theorem with a flat prior:

$$1-\alpha = 0.9 = \int_0^{S_{up}} p(s|n)ds = \frac{\int_0^{S_{up}} P(n|s)ds}{\int_0^{\infty} P(n|s)ds} \tag{5.5}$$

where $1-\alpha$ is the confidence level, and $p(s|n)$ is probability of having mean signal events s when we have detected n. This is different from $p(n|s)$, which is the probability of detecting n events given s mean signal events. Because we expect no signal events, we set $s = 0$ and $n = b$. In this analysis, $n = b$ always, but in the later section, we will describe how n is allowed to vary randomly taking into account uncertainties in b. By solving Eq. 5.5, we will be able to locate S_{up}. It is worth noting that in this limit setting scenario where no signal events are expected, S_{up} depends *only* on the expected number of background events b. In my analysis framework, solving for S_{up} is done with the function `tools_Stats::GetS_up`.

The value of S_{up} must be compared to the number of neutrinos passing all cuts for the same choice of slope m and intercept d; this number of neutrinos passing all cuts is the signal, S. These are the neutrinos up and to the right of the red line in Fig. 5.1. What we do in practice is, for a given m, is step through our choices for d, and for each d, we compute the corresponding background expectation b, which generates an S_{up}. For this d, we can also compute the number of signal events passing all cuts. This allows us to define an S/S_{up} for each choice of d, and we will select the cut that provides the maximum value of S/S_{up}. A plot of S/S_{up} for A2 configuration 1 can be seen in Fig. 5.4, which has a peak at 10.9, which would be the optimized choice of cut value for the example slope of -620.

5.2.3 Optimizing the RCut Across Configurations

Up to now, I have shown how to optimize each configuration individually. The question remains: how to set a limit utilizing the livetime in all configurations simultaneously. We utilize a straightforward extension of the single configuration case. We will search for a global S/S_{up}. This time however, we will use the *summed* background contributions, $b_{tot} = \sum b_i$ from each configuration to determine $S_{up,tot}$, and will utilize the *summed* signal passing all cuts in each configuration $S_{tot} = \sum S_i$.We vary both the slope m and intercepts d_i to maximize $S_{tot}/S_{up,tot}$.

Note that because we simulated the same number of signal neutrinos in the simulation data sets for each configuration, we must weight the number of neutrinos contributing from a given configuration by the relative livetime contribution of that configuration. For example, configuration 1 only contributes approximately 18% of the livetime in the search, and so only 18% of the signal events passing in the configuration 1 simulation will contribute to the total summed signal S_{tot}. We accomplish this global optimization with a exhaustive grid scan. Which is to say, we check all combinations of m, and all combinations of d_i, one-by-one in our search for a global optimum.

There is one last question, which is how to account for the uncertainties in the expo-

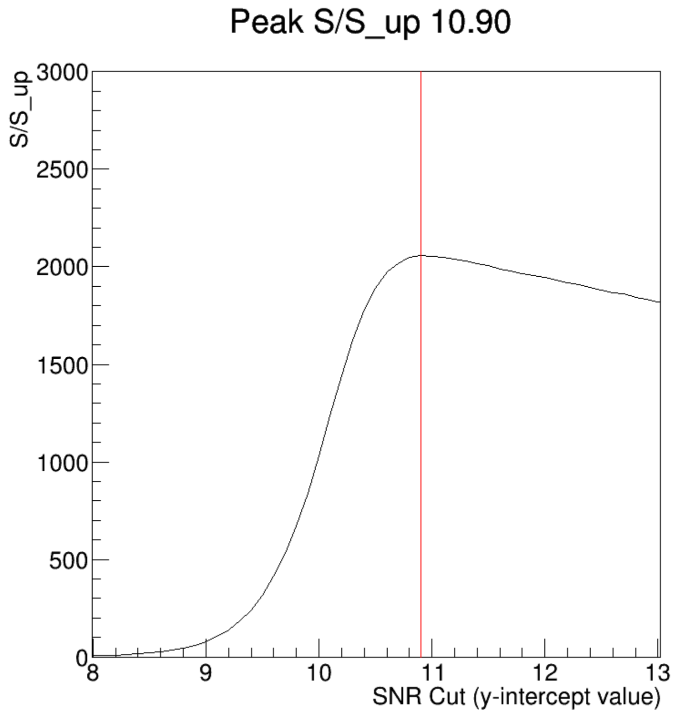


Figure 5.4: The calculation of S/S_{up} for A2 configuration 1.

nential fit parameters β_1 and β_2, which we ignored in the previous section. These lead to systematic uncertainties in our background estimate. In the previous section, in estimating the background, and computing the resultant S_{up}, we assumed the central value of the fit parameters for estimating the true value of the expected number of background events. This of course fails to take into account the fact that the fit parameters have uncertainties, which we designate σ_{β_1} and σ_{β_2}. The fit parameters cannot be determined completely independently of one-another—that is, they are correlated.

Fortunately, there is an established way of taking into account these uncertainties and the correlation between the fit parameters: it is accomplished with pseudo-experiments, and is described in the Monte Carlo section of the Particle Data Group handbook [100]. We will start by assuming that the values we retrieved for the fit are each the mean of a Gaussian, and that the fit uncertainties represent the standard deviation of these Gaussians.

What we would like to know is the following: if we repeated the experiment again in the future, what does the *distribution* of the observed number of background events look like given our model for the background. And given this distribution of observed number of backgrounds, what does the distribution of expected S_{up} look like. We will optimize the mean of this distribution of expected S_{up}, which will be pulled to a higher value than if we had assumed the central value, , making the expected limit weaker.

The pseudo-experiments to build the distribution of the number of observed background events and the distribution of S_{up} are thrown as follows. In a single configuration i, we have the fit parameters $\beta_{1,i}$ and $\beta_{2,i}$, with associated uncertainties $\sigma_{\beta_1,i}$ and $\sigma_{\beta_2,i}$. When we performed the exponential fitting in ROOT, we also retrieved the correlation coefficient ρ_i between the fit parameters, which is equal to the covariance divided by the product of the uncertainties: $\rho_i = \text{Cov}(\beta_{1,i}, \beta_{2,i})/\sigma_{\beta_1,i}\sigma_{\beta_2,i}$. For the pseudo-experiment, we select a new value for the slope and intercept of the exponential fit: these new values we will designate with primes, so that the slope of the exponential for this pseudo-experiment is $\beta'_{1,i}$ and the new value for the intercept is $\beta'_{2,i}$. The equation for $\beta'_{1,i}$ is provided by Eq. 5.6, where η_1 is a random number drawn from a Gaussian distribution with a mean of zero and a width of unity.

$$\beta'_{1,i} = \beta_{1,i} + \sigma_{\beta_1,i}\eta_1 \tag{5.6}$$

What this says is that we will select a new slope from a distribution of slopes that is centered about the nominal value, but that varies in a Gaussian fashion with an average deviation of $\sigma_{\beta_1,i}$. To select a new value for the intercept, we follow Eq. 5.7

$$\beta'_{2,i} = \beta_{2,i} + \rho_i\sigma_{\beta_2,i}\eta_1 + \sigma_{\beta_2,i}\eta_2\sqrt{1-\rho_i^2} \tag{5.7}$$

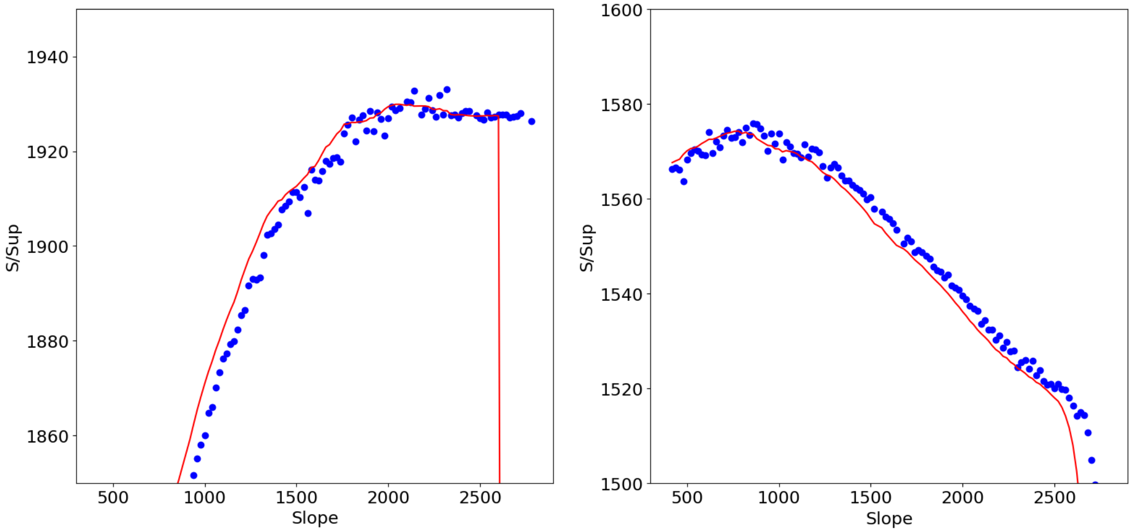
where ρ_i is the correlation coefficient described above, and η_2 is a second Gaussian random number with a mean of zero and a width of unity. Like before, this says that we expect the new value for the intercept to selected from a distribution centered on the nominal central value. But in this case, the new amplitude is also influenced by the variation in the new slope and the correlation between the parameters.

Using these new values for the slope and intercept, we can compute a new value for the expected number of background events for this configuration b'_i. We draw a new expected number of background events for each configuration i, sum them into a new total background estimate b'_{tot}, and compute a new $S'_{up,tot}$ for this pseudo-experiment. By performing many thousands of pseudo-experiments, we can build up a distribution of $S'_{up,tot}$, and we will optimize using the mean of this distribution of upper limits.

5.2.4 Optimized Cut Values and Background Estimate

We performed the grid scan, including the use of pseudo-experiments to systematically smear the background distributions according to systematic uncertainties, and found optimal slope values of -2060 for VPol and -780 for HPol. This can be seen in Fig. 5.5.

For these set of slopes, the y-intercepts for VPol and HPol are presented in Tab. 5.1. With these final cut values, we are able to determine the efficiency-vs-SNR for both VPol and HPol. As an example, these are presented in Fig. 5.6 for A2 configuration 1. We also find the efficiency as a function of neutrino energy allowing a neutrino event to pass in



(a) VPol slope optimization for A2.

(b) HPol slope optimization for A2.

Figure 5.5: The plot of S/S_{up} as a function of slope for VPol and HPol.

Config	VPol Y-Intercept	HPol Y-Intercept
1	20.5	13.6
2	22.4	14.7
3	20.4	13.1
4	21.4	14.7
5	21.5	14.5

Table 5.1: The optimized y-intercepts for the RCut for VPol and HPol for all configurations.

either VPol or HPol, which is presented in Fig. 5.7. The analysis achieves a total efficiency for neutrinos ranging from approximately 30% near 10^{17} eV, to almost 65% near 10^{20} eV. The final efficiency as a function of energy for all configurations can be seen in Fig. 5.8. The total efficiency of the analysis for the Kotera max flux in A2 configuration 1, averaged across energy bins, is approximately 49%. The efficiency for each configuration bin is presented in Tab. 5.2.

Additionally, we can place the efficiency of this analysis in context by comparing the efficiency to what has been achieved previously. This can be seen in Fig. 5.9, where our result (black) is compared with the previously achieved final efficiency of the prototype analysis in blue [53]. We make this comparison specifically because the prototype analysis also set their final cut in the correlation vs SNR plane. We have achieved a roughly factor of two to four increase in efficiency over the prototype analysis.

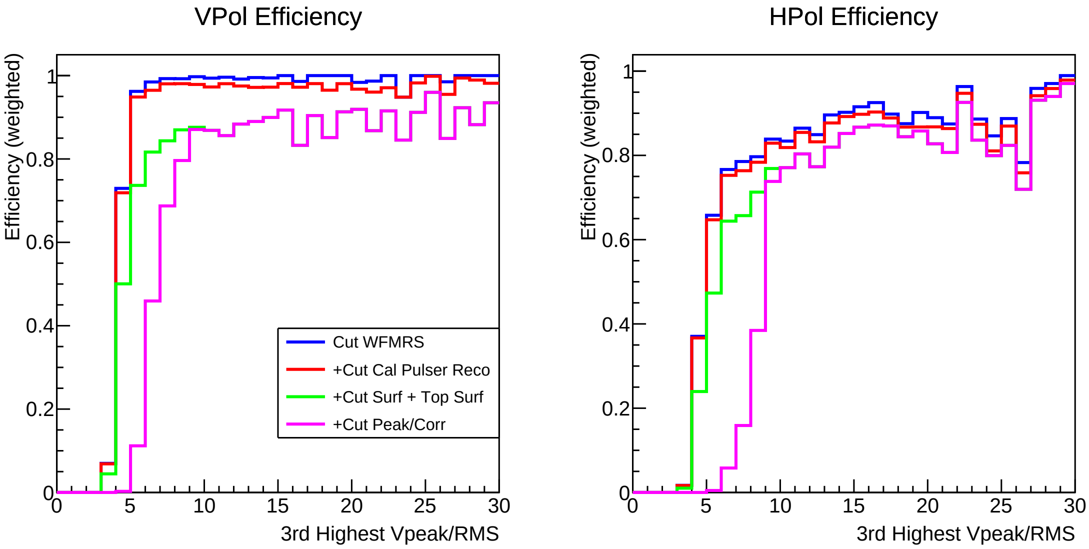


Figure 5.6: The efficiency as a function of SNR for all cuts in the analysis for VPol and HPol for A2 configuration 1.

Config	VPol Efficiency	HPol Efficiency	Efficiency to Pass VPol or HPol
1	40.2	33.5	49.0
2	32.4	19.7	36.8
3	41.0	34.5	50.8
4	38.2	31.5	47.0
5	38.8	32.3	47.7

Table 5.2: The efficiency of the VPol and HPol analysis separately, and combined.

With our selected final cuts values, we are also able to generate our “cut-table”, which reports the number of events that are cut by each cut. The cut table for A2 configuration 1 is presented in Tab. 5.3. We report the fraction of events cut “as first cut”, “in sequence”, and “as last cut”. The “as last cut” number is a measure of the orthogonality of the cuts, and indicates how many events are rejected by that cut alone.

For these cut values, we can also estimate the expected number of background events. The distribution of the expected number of background events for A2 is presented in Fig. 5.10, with VPol on the left and HPol on the right. The expected number of events is $0.0104\times^{+0.0029}_{-0.0036}$ in VPol and $0.0157\times^{+0.0026}_{-0.0029}$ in HPol. The upper and lower error bars are the 1σ points of the background distributions.

	Data						**Simulation**					
Num Total	10655428						12053					
Good Events	7642534						12053					
	V Data			*H Data*			*V Sim*			*H Sim*		
Fraction Cut By	As First	In Seq	As Last	As First	In Seq	As Last	As First	In Seq	As Last	As First	In Seq	As Last
WFRMS	9.507E-01	9.507E-01	1.018E-03	9.449E-01	9.449E-01	2.198E-05	2.557E-01	2.557E-01	2.439E-03	4.270E-01	4.270E-01	4.263E-02
Box	1.353E-02	7.202E-04	1.099E-05	1.353E-02	7.783E-04	5.365E-06	1.566E-02	1.212E-02	7.516E-03	1.566E-02	1.005E-02-02	6.264E-03
Surf & Top Surf	3.103E-01	2.171E-02	1.265E-04	3.098E-01	2.355E-02	2.074E-04	1.808E-01	1.182E-01	3.962E-02	1.493E-01	6.461E-02	1.436E-02
RCut	9.983E-01	2.684E-02	2.684E-02	9.996E-01	3.080E-02	3.080E-02	5.466E-01	2.115E-01	2.115E-01	5.981E-01	1.636E-01	1.636E-01

Table 5.3: The as last cut table for A2 configuration 1. At the top of the table is the number of events surviving the basic quality cuts. In the bottom half of the table is the fraction of events cut by each cut.

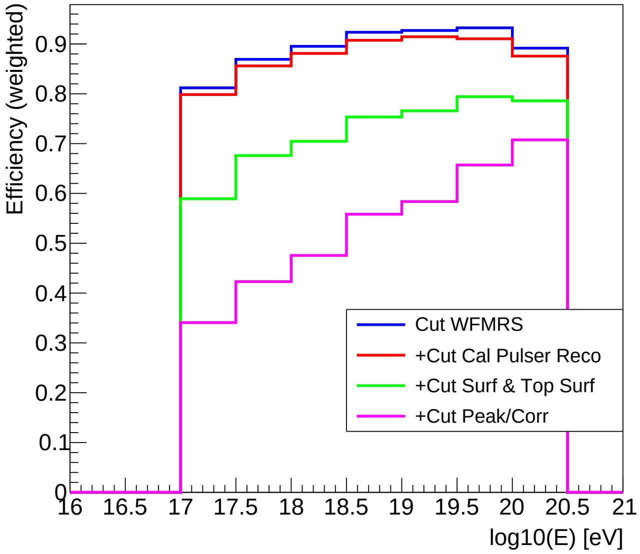


Figure 5.7: The efficiency as a function of neutrino energy for all cuts in the analysis for A2 configuration 1. Neutrinos are allowed to pass in either the VPol or HPol channel.

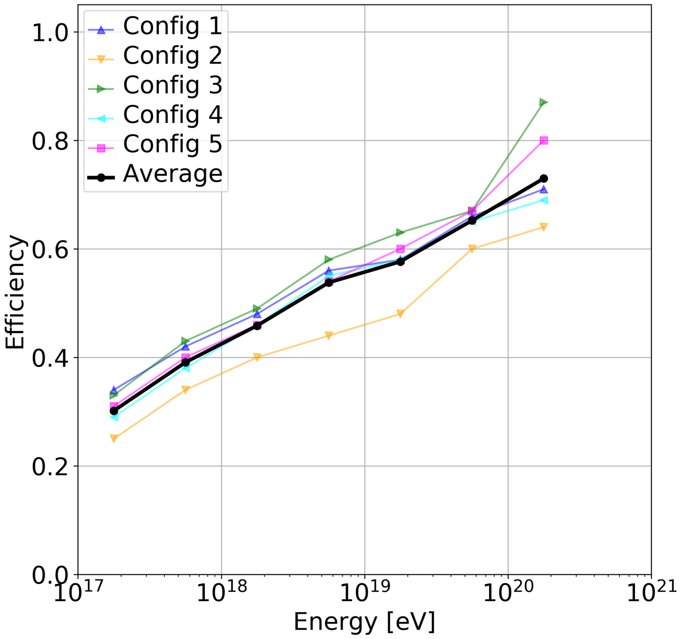


Figure 5.8: The efficiency of this analysis as a function of energy for all configurations.

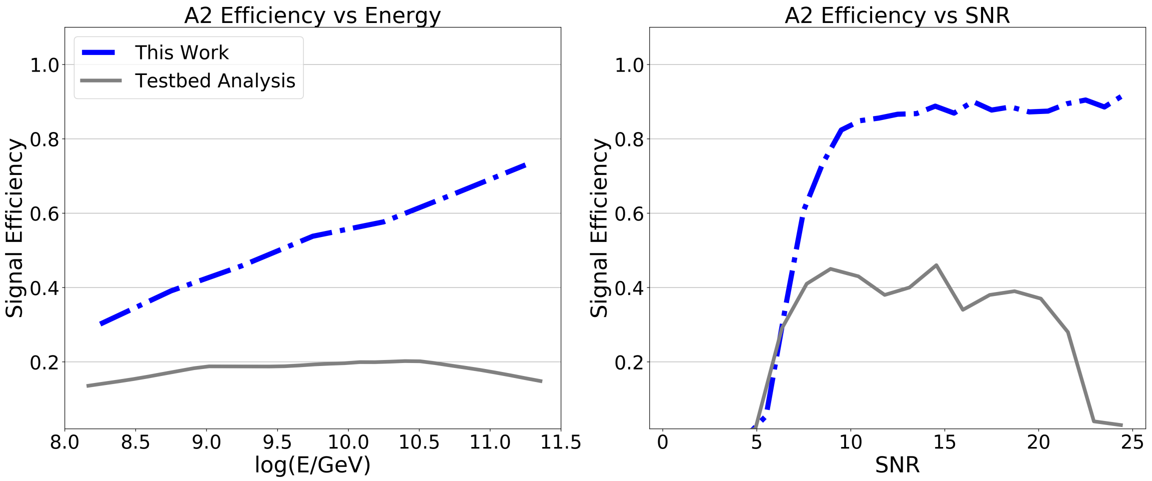
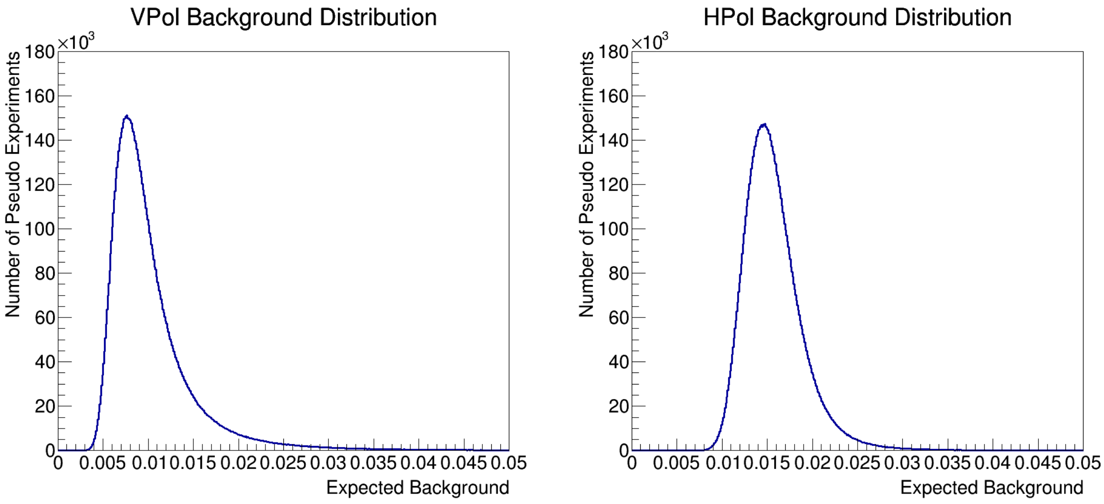


Figure 5.9: A comparison of the efficiency of this analysis to the previous prototype analysis. On the left is the efficiency as a function of energy, and on the right, as a function of SNR.



(a) VPol.

(b) HPol.

Figure 5.10: The distribution of pseudo-experiments for the number of background events in A2 for VPol (left) and HPol (right).

5.3 Expected Limit

To set our expected limit, we begin by computing the effective volume of the ARA instrument. This is done with the ARA simulation package `AraSim` [53]. We throw a known number of neutrinos N_{thrown} in an interaction volume V_{thrown} isotropically, and record the number of events which pass the trigger N_{det}. The effective volume times steradians, $[V\Omega]_{eff}$, can then be computed from Eq. 5.8. We plot the resulting $[V\Omega]_{eff}$ for all configurations in Fig. 5.11.

$$[V\Omega]_{eff} = \frac{N_{det}}{N_{thrown}} \times V_{thrown} \times 4\pi \tag{5.8}$$

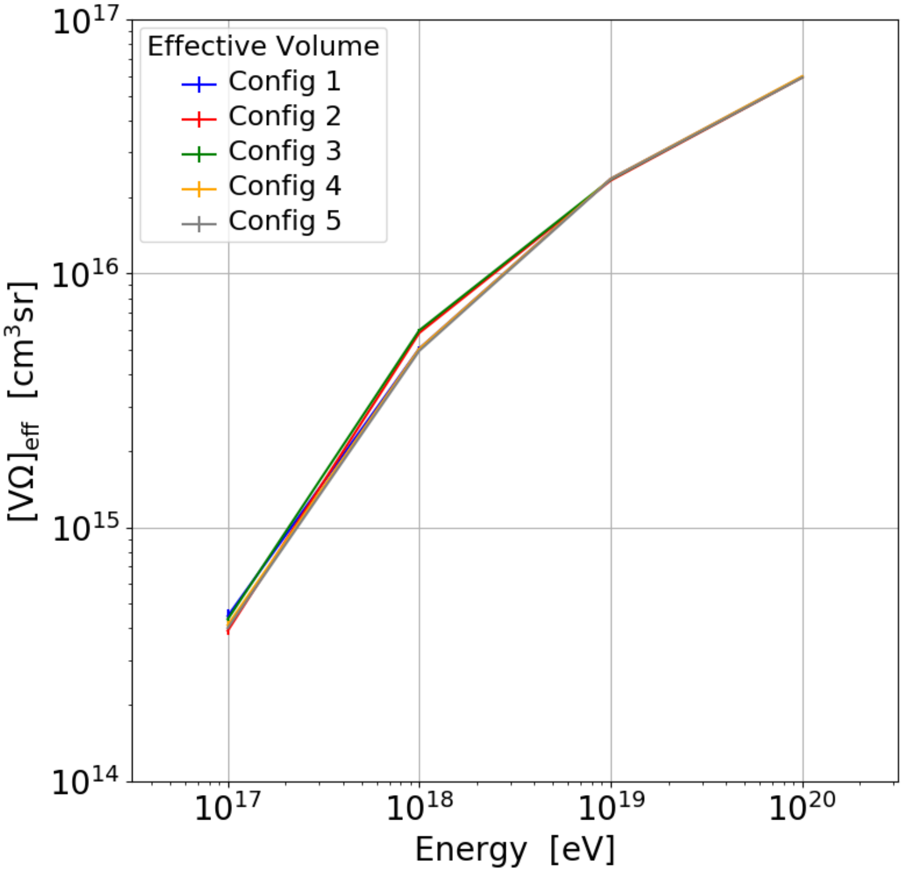


Figure 5.11: The trigger level effective volumes of all A2 configurations.

To convert between effective volumes and effective areas, we utilize the thin-target approximation:

$$[A\Omega]_{eff} = \frac{[V\Omega]_{eff}}{\mathcal{L}_{int}} \tag{5.9}$$

where $[A\Omega]_{eff}$ is the effective area times steradians, $[V\Omega]_{eff}$ is the effective volume times steradians, and $\mathcal{L}_{int}$ is the interaction length of a neutrino in the earth. We may find the interaction length:

$$\mathcal{L}_{int} = \frac{m_N}{\rho\ \sigma_{\nu N}} \tag{5.10}$$

where $m_N = 1.67 \times 10^{-24}$g is the mass of a nucleon, $\rho = 3.8$ g/cm^2 is the average density of the earth, and $\sigma_{\nu N}$ is the neutrino-nucleon cross section in the units of cm^2. An approximation for $\sigma_{\nu N}$ can be found in Connolly *et. al.* [101].

Given our knowledge of the effective volume of the instrument, we can compute the number of events we expect to observe in the j-th energy bin, N_j according to Eq. 5.11.

$$N_j = \epsilon_j \left(\frac{dN}{d\log_{10}E\ dA\ dt\ d\Omega} \right)_j d\log_{10}E\ dA_j\ d\Omega_j\ dt \tag{5.11}$$

where $\frac{dN}{d\log_{10}E\ dA\ dt\ d\Omega}$ is the energy-dependent flux $F(E)$ in a given logarithmic energy bin, dt is the livetime, $dA\ d\Omega$ is the aperture, $d\log_{10}E$ is the logarithmic energy bin width, and ϵ_j is the analysis level efficiency. We can use the properties of logarithms to rearrange the $F(E)$ term:

$$d\log_{10}E = \frac{d\ell nE}{\ell n10} = \frac{1}{E}\frac{1}{\ell n10}dE \tag{5.12}$$

such that Eq. 5.11 maybe rearranged as

$$\begin{aligned} N_j &= \epsilon_j E_j \left(\frac{dN}{dE\ dA\ dt\ d\Omega} \right)_j \ell n10\ d\log_{10}E\ dA_j\ d\Omega_j\ dt \\ &= \epsilon_j E_j\ F(E)_j\ \ell n10\ d\log_{10}E\ dA_j\ d\Omega_j\ dt \end{aligned} \tag{5.13}$$

This new quantity $EF(E)$ has units of cm^{-2} sr^{-1} s^{-1}. Generally we express the aperture as a single effective quantity $[A\Omega]_{eff,j}$, and take dt as the livetime of the experiment T because the detector observes all energy bins with equal livetime. We choose to set our limit in decade wide bins in the logarithm of the energy so that $d\log_{10}E = 1$. For notational convenience, let us define the product of the efficiency, livetime, and effective area in a single energy bin j as the *exposure* Λ_j as in Eq. 5.14:

$$\Lambda_j = \epsilon_j \times [A\Omega]_{eff,j} \times T_j \tag{5.14}$$

Because we have five livetime configurations i for A2, each with their own efficiencies ($\epsilon_{j,i}$), apertures ($[A\Omega]_{eff,j,i}$) and livetimes ($T_{j,i}$), we must sum over the five configurations to find the total Λ_j we will use in setting the expected limit as in Eq. 5.15:

$$\Lambda_j = \sum_{\text{configs}\, i=1}^{5} \epsilon_{j,i} \times [A\Omega]_{eff,j,i} \times T_{j,i} \tag{5.15}$$

In the expected absence of a detection ($N_j = 0$), we compute an expected *limit* on the *maximum* value of $EF(E)$. At the 90% confidence-level (CL), the largest number of events in a given bin which is compatible with a non-observation is 2.44 [102]. So, starting from Eq. 5.13, we set $N_j = 2.44$, substitute for the exposure, set $d\log_{10}E = 1$, and solve for $EF(E)$:

$$EF(E)_j = \frac{2.44}{\ell n10\ \Lambda_j} \tag{5.16}$$

The expected limit curve as a function of energy is presented in Fig. 5.12, where the solid black line labeled "ARA (1x4yr)" is this analysis. We also show in a purple band the projected trigger level sensitivity of the full five-station station array (ARA5) by 2022, taking into account the livetime that is already archived but not yet analyzed, in addition to that which will be accumulated by 2022. The upper edge of the band is the 90% confidence-level upper limit assuming a non-observation; the lower edge of the band is the single event sensitivity. As can be seen, ARA5 will have accumulated sufficient data to set world leading limits above 10^{19} eV within the next four years, and will have a real chance at observing the first ultra-high energy neutrino candidate. Importantly, this measurement with ARA5 will constrain several cosmogenic neutrino flux models, and will probe directly if the astrophysical neutrino flux measured by IceCube continues unbroken to higher energies. Such pathfinding will be essential to setting the scale for the next generation of neutrino observatories, e.g. RNO, IceCube-Gen2, GRAND, etc.

5.4 Conclusion

In this chapter, I presented the optimization of an analysis of four years of A2 data, including a description of all cuts, the expected number of background events, the energy dependent efficiency of the analysis, and the expected limit assuming a non-detection. I also laid the groundwork for this process to be repeated quickly in A3. This analysis introduced a new, efficient thermal noise rejection filter and for the first time filtered ARA events of CW contamination. The analysis is the largest amount of ARA livetime analyzed to date, demonstrating ARA's ability to manage its growing bank of livetime.

In the coming weeks, our plan is to finish unblinding the A2 data set. We will do this in several stages, opening sidebands of the experiment first. We will first reverse the in-ice cut to study events which come from above the surface. We will look for potential events of interest, including solar flares and periods of livetime which might contain anthropogenic activity which was not previously identified. We will then open the box for the in-ice events, but that fail the RCut, which is the last cut in the analysis. In this way, we can test whether the distribution of background events in the 100% sample follows that observed in the 10% sample. Finally, we will open the box on events coming from inside the ice, but passing the

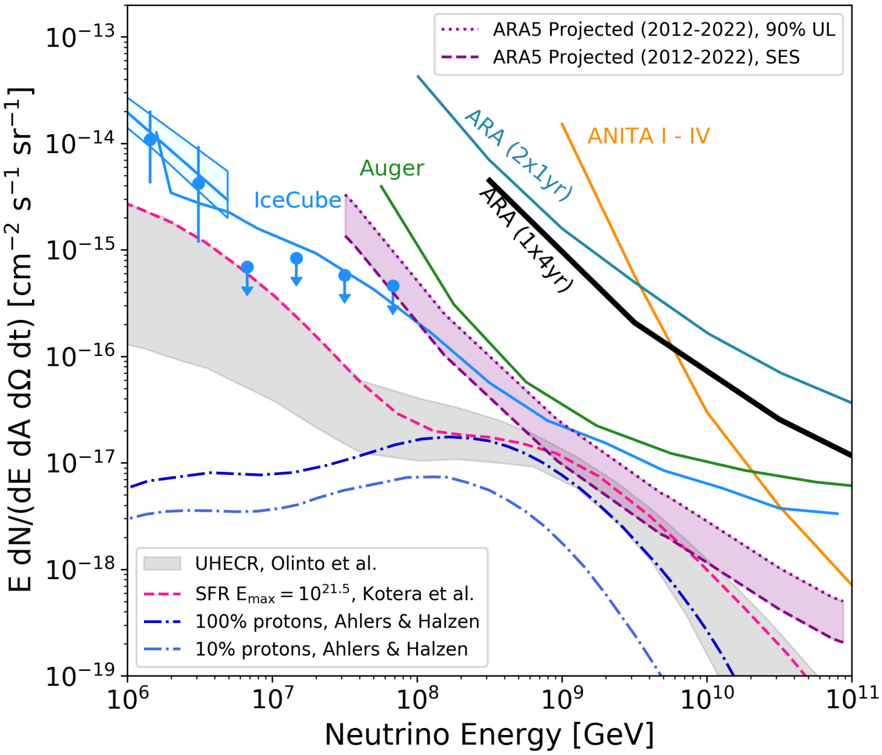


Figure 5.12: The expected limit for the A2 four year analysis.

RCut, which is expected to be the neutrino signal region.

Chapter 6

Diffuse Neutrino Search: Lessons Learned and Potential Improvements

6.1 Lessons Learned

There were many lessons learned in the process of optimizing this analysis. I will list some of them here, hopefully for the benefit of future students and collaborators.

First, the data set is very large—larger than the field has ever analyzed before in a single analysis. The 10% data sample for A2 has nearly 58 million events; the 100% sample therefore has over half of a billion events, and takes up almost 40 TB of space on computing clusters. For these reasons, the design and tuning of cuts takes more time than expected, because often when you want to design a new cut, or test a new threshold for a cut, you must loop over these many millions of events. Additionally, because the data is split up over approximately 5000 runs per station for the four year analysis, 10,000 run files total, there is a scripting challenge to organize the many files, and record when certain jobs fail so they may be restarted. This challenge will only continue to grow as ARA accumulates more livetime.

Second, the analysis complexity could likely be reduced with improvements to the ARA operations procedures. For example, roughly 1/5 of the excluded runs for this analysis were discovered during the analysis process, and confirmed by manually checking old email archives, old DocDB posts, etc. Ideally, these bad run periods would have been carefully recorded and logged in a central database and could have been removed more quickly. Additionally, the number of different configurations could have been reduced if some sub-optimal operating modes had been caught more quickly. For example, configuration five has no trigger delays enabled due to an operational oversight. But this oversight effectively increased the complexity of the analysis by 20% by adding a fifth configuration.

Third, the analyses could benefit from more sharing and standardization of basic data quality algorithms. For example, my fellow analyst Ming-Yuan Lu and I did not realize until several months into the analysis that we had done some redundant work in removing glitches and bad runs from our dataset. However, this could have been handled by a mutual piece of code, and would have saved both of us valuable time. Indeed, was the inspiration for the `AraQualCuts` class. We should continue to do this in the future so every analyst does not have to begin by designing their own basic quality cuts.

6.2 Potential Improvements

There are a few improvements to the analysis that could be made that occur to me immediately, and I am confident future readers could arrive at others. The cut values for some cuts are not finely tuned. This was stated, for example, about the wavefront-RMS filter. One could consider exploring the threshold-cut space more thoroughly to determine if there is another combination that achieves comparable neutrino efficiency but with a lower thermal-noise passing rate. The cut could also be made part of the optimization procedure.

Further, concerning the wavefront-RMS event filter: as originally designed, it was specifically intended to leverage the regular geometry of the ARA station. The reduction in the number of faces contributing to the filter—from twelve in a fully operational station to six in the case of one lost antenna, to three in the case of two lost antennas—suggests that there might be a modification to the filter to algorithm to make it less reliant on the use of a complete station.

There are a few places where the code could be more computationally efficient. For example, the phase-variance technique requires the computation of the phase vs frequency for the events around an event being considered for CW contamination, which can be time-expensive to calculate. A solution, for example, might be to store the FFTs of the events ahead of time, but this would essentially require making and storing a copy of the raw data, which would require large storage space. A second computation improvement would be to make the interferometry more efficient; doing interferometry on an event takes approximately 1 second, which is computationally prohibitive for large data samples. This could potentially be improved by leveraging parallelization techniques, e.g. open-MP, etc.

Finally, the identification of surface-related anthropogenic activity was done before the improvement to the correlation function. With the improved correlation function, it may be possible to additionally identify periods of livetime which are contaminated with anthropogenic activity.

Chapter 7

CONCLUSION

In this thesis, I have described an analysis of radio emission from an X-class solar flare in chapter 2, presented the design, characterization, and performance of a new radio-frequency conditioning module for ARA in chapter 3, and lastly presented the analysis of four years of data from station A2 in chapters 4 and 5

The solar flare paper of chapter 2 was the first analysis by ARA to reconstruct radio emission to an extraterrestrial source on an event-by-event basis. This proved a valuable testing ground to confirm ARA's reconstruction projections onto celestial coordinates, and should aid ARA in constraining the direction of other above-ice sources we would hope to detect in the future—for example, cosmic rays.

The ARAFE's discussed in chapter 3 presented the next-generation of signal-conditioning modules for ARA, which will hopefully serve as a launching point for the amplification and filtering modules for future experiments. In particular, their compact design and small power footprint make them ideal for small deployment payloads with tight power budgets. The inclusion of the tunable attenuators increases the dynamic range of the instrument, as we demonstrated by using them to prevent saturation of the digitizer system during a calibration test run.

Finally, in chapters 4 and 5, we presented the optimization of an analysis of four years of data of station A2, including an explanation of all cuts, the tuning of the final cuts, the energy dependent efficiency of the analysis, the expected number of background events, and a expected limit. We also laid the groundwork for the analysis of station A3, which is ongoing. The analysis uses a new thermal noise rejection algorithm, and for the first time filters ARA data of CW contamination instead of outright rejecting contaminated events. The analysis leverages more of the livetime of A2 that has been analyzed previously, and demonstrates ARA's ability to study our increasingly large data sets.

Bibliography

[1] M. G. Aartsen et al. Observation of High-Energy Astrophysical Neutrinos in Three Years of IceCube Data. *Phys. Rev. Lett.*, 113:101101, 2014.

[2] M. G. Aartsen et al. Observation and Characterization of a Cosmic Muon Neutrino Flux from the Northern Hemisphere using six years of IceCube data. *Astrophys. J.*, 833(1):3, 2016.

[3] M. G. Aartsen et al. A combined maximum-likelihood analysis of the high-energy astrophysical neutrino flux measured with IceCube. *Astrophys. J.*, 809(1):98, 2015.

[4] M. G. Aartsen et al. Neutrino emission from the direction of the blazar TXS 0506+056 prior to the IceCube-170922A alert. *Science*, 361(6398):147–151, 2018.

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

[6] M. G. Aartsen et al. Astrophysical neutrinos and cosmic rays observed by IceCube. *Adv. Space Res.*, 62:2902–2930, 2018.

[7] Eli Waxman and John N. Bahcall. High-energy neutrinos from cosmological gamma-ray burst fireballs. *Phys. Rev. Lett.*, 78:2292–2295, 1997.

[8] Kohta Murase, Markus Ahlers, and Brian C. Lacki. Testing the Hadronuclear Origin of PeV Neutrinos Observed with IceCube. *Phys. Rev.*, D88(12):121301, 2013.

[9] Markus Ahlers and Francis Halzen. Opening a New Window onto the Universe with IceCube. *Prog. Part. Nucl. Phys.*, 102:73–88, 2018.

[10] B. P. Abbott et al. GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral. *Phys. Rev. Lett.*, 119(16):161101, 2017.

[11] Kenneth Greisen. End to the cosmic ray spectrum? *Phys. Rev. Lett.*, 16:748–750, 1966.

[12] G. T. Zatsepin and V. A. Kuz'min. Upper Limit of the Spectrum of Cosmic Rays. *Soviet Journal of Experimental and Theoretical Physics Letters*, 4:78, August 1966.

[13] R. U. Abbasi et al. First observation of the Greisen-Zatsepin-Kuzmin suppression. *Phys. Rev. Lett.*, 100:101101, 2008.

[14] B. L. Berman and S. C. Fultz. Measurements of the giant dipole resonance with monoenergetic photons. *Rev. Mod. Phys.*, 47:713–761, 1975.

[15] V. S. Berezinsky and G. T. Zatsepin. Cosmic rays at ultrahigh-energies (neutrino?). *Phys. Lett.*, 28B:423–424, 1969.

[16] A. V. Olinto, K. Kotera, and D. Allard. Ultrahigh Energy Cosmic Rays and Neutrinos. *Nucl. Phys. Proc. Suppl.*, 217:231–236, 2011.

[17] K. Kotera, D. Allard, and A. V. Olinto. Cosmogenic Neutrinos: parameter space and detectabilty from PeV to ZeV. *JCAP*, 1010:013, 2010.

[18] Markus Ahlers and Francis Halzen. Minimal Cosmogenic Neutrinos. *Phys. Rev.*, D86:083010, 2012.

[19] John David Jackson. *Classical Electrodynamics*. Wiley, New York, NY, 3rd ed. edition, 1999.

[20] M. A. Markov. On high energy neutrino physics. In *Proceedings, 10th International Conference on High-Energy Physics (ICHEP 60): Rochester, NY, USA, 25 Aug - 1 Sep 1960*, pages 578–581, 1960.

[21] Y. Fukuda et al. The Super-Kamiokande detector. *Nucl. Instrum. Meth.*, A501:418–462, 2003.

[22] M. G. Aartsen et al. The IceCube Neutrino Observatory: Instrumentation and Online Systems. *JINST*, 12(03):P03012, 2017.

[23] G. A. Askar'yan. Excess negative charge of an electron-photon shower and its coherent radio emission. *Sov. Phys. JETP*, 14(2):441–443, 1962. [Zh. Eksp. Teor. Fiz.41,616(1961)].

[24] G. A. Askar'yan. Coherent Radio Emission from Cosmic Showers in Air and in Dense Media. *Soviet Journal of Experimental and Theoretical Physics*, 21:658, Sep 1965.

[25] David Saltzberg, Peter Gorham, Dieter Walz, Clive Field, Richard Iverson, Allen Odian, George Re, Paul Schoessow, and Dawn Williams. Observation of the Askaryan effect: Coherent microwave Cherenkov emission from charge asymmetry in high-energy particle cascades. *Phys. Rev. Lett.*, 86:2802–2805, 2001.

[26] P. W. Gorham et al. Observations of the Askaryan effect in ice. *Phys. Rev. Lett.*, 99:171101, 2007.

[27] J. Alvarez-Muniz, R. A. Vazquez, and E. Zas. Calculation methods for radio pulses from high-energy showers. *Phys. Rev.*, D62:063001, 2000.

[28] M. Ackermann et al. Optical properties of deep glacial ice at the South Pole. *J. Geophys. Res.*, 111(D13):D13203, 2006.

[29] S. Barwick, D. Besson, P. Gorham, and D. Saltzberg. South Polar in situ radio-frequency ice attenuation. *J. Glaciol.*, 51:231–238, 2005.

[30] P. Allison et al. Design and initial performance of the Askaryan Radio Array prototype EeV neutrino detector at the South Pole. *Astroparticle Physics*, 35:457–477, February 2012.

[31] Taylor Barrella, Steven Barwick, and David Saltzberg. Ross ice shelf (antarctica) in situ radio-frequency attenuation. *Journal of Glaciology*, 57(201):61–66, 2011.

[32] J. Avva, J. M. Kovac, C. Miki, D. Saltzberg, and A. G. Vieregg. An in situ measurement of the radio-frequency attenuation in ice at Summit Station, Greenland. *J. Glaciol.*, 61:1005–1011, 2015.

[33] M. G. Aartsen et al. Differential limit on the extremely-high-energy cosmic neutrino flux in the presence of astrophysical background from nine years of IceCube data. *Phys. Rev.*, D98(6):062003, 2018.

[34] Alexander Aab et al. Improved limit to the diffuse flux of ultrahigh energy neutrinos from the Pierre Auger Observatory. *Phys. Rev.*, D91(9):092008, 2015.

[35] P. W. Gorham et al. Constraints on the ultra-high energy cosmic neutrino flux from the fourth flight of ANITA. 2019.

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

[37] Chris Persichilli. *Performance and Simulation of the ARIANNA Pilot Array, with Implications for Future Ultra-high Energy Neutrino Astronomy.* PhD thesis, University of California - Irvine, 2018.

[38] M. Ageron et al. ANTARES: the first undersea neutrino telescope. *Nucl. Instrum. Meth.*, A656:11–38, 2011.

[39] Icecube Collaboration. An absence of neutrinos associated with cosmic-ray acceleration in γ-ray bursts. *Nature*, 484(7394):351–354, Apr 2012.

[40] ANTARES Collaboration. Virtual reality views and animations. http://antares.in2p3.fr/Gallery/3D/index.html, 2011. [Online; accessed 15-June-2019].

[41] P. W. Gorham et al. The Antarctic Impulsive Transient Antenna Ultra-high Energy Neutrino Detector Design, Performance, and Sensitivity for 2006-2007 Balloon Flight. *Astropart. Phys.*, 32:10–41, 2009.

[42] A. Romero-Wolf et al. Comprehensive analysis of anomalous ANITA events disfavors a diffuse tau-neutrino flux origin. *Phys. Rev.*, D99(6):063011, 2019.

[43] Cosmin Deaconu. Anita. https://www.hep.ucl.ac.uk/uhen/anita/, 2018. [Online; accessed 15-June-2019].

[44] O. Scholten, S. Buitink, J. Bacelar, R. Braun, A. G. de Bruyn, H. Falcke, K. Singh, B. Stappers, R. G. Strom, and R. Al Yahyaoui. First results of the NuMoon experiment. *Nucl. Instrum. Meth.*, A604:S102–S105, 2009.

[45] Peter W. Gorham, C. L. Hebert, K. M. Liewer, C. J. Naudet, D. Saltzberg, and D. Williams. Experimental limit on the cosmic diffuse ultrahigh-energy neutrino flux. *Phys. Rev. Lett.*, 93:041101, 2004.

[46] I. Kravchenko et al. Updated results from the rice experiment and future prospects for ultra-high energy neutrino detection at the south pole. *Phys. Rev. D*, 85:062004, Mar 2012.

[47] S. W. Barwick et al. A First Search for Cosmogenic Neutrinos with the ARIANNA Hexagonal Radio Array. *Astropart. Phys.*, 70:12–26, 2015.

[48] ARA Collaboration. Askaryan radio array. https://ara.wipac.wisc.edu/home, 2016. [Online; accessed 15-June-2019].

[49] S. W. Barwick et al. Radio detection of air showers with the ARIANNA experiment on the Ross Ice Shelf. *Astropart. Phys.*, 90:50–68, 2017.

[50] A. Anker et al. Targeting cosmogenic neutrinos with the ARIANNA experiment. 2019.

[51] Christian Glaser et al. NuRadioMC: Simulating the radio emission of neutrinos from interaction to detector. 2019.

[52] P. Allison et al. Design and Initial Performance of the Askaryan Radio Array Prototype EeV Neutrino Detector at the South Pole. *Astroparticle Physics*, 35:457–477, 2012.

[53] P. Allison et al. First Constraints on the Ultra-High Energy Neutrino Flux from a Prototype Station of the Askaryan Radio Array. *Astropart. Phys.*, 70:62–80, 2015.

[54] Rhessi flare list. `https://hesperia.gsfc.nasa.gov/rhessi3/data-access/rhessi-data/flare-list/index.html`. Accessed: 2017-05-13.

[55] Stephen M. White. Solar Radio Bursts and Space Weather. *Asian Journal of Physics*, 16:189–207, 2007.

[56] Zhiwei Liu, Dale E. Gary, Gelu M. Nita, Stephen M. White, and Gordon J. Hurford. A subsystem test bed for the frequency agile solar radiotelescope. *Publications of the Astronomical Society of the Pacific*, 119(853):303, 2007.

[57] Fermi solar flare x-ray and gamma-ray observations. `https://hesperia.gsfc.nasa.gov/fermi_solar/`. Accessed: 2017-05-13.

[58] Sws - culgoora observatory - type ii events. `http://www.sws.bom.gov.au/Solar/2/5/1`. Accessed: 2017-05-12.

[59] I. Kravchenko, D. Besson, and J. Meyers. In situ index-of-refraction measurements of the South Polar firn with the RICE detector. *Journal of Glaciology*, 50:522–532, 2004.

[60] M. Lu, C. Pfendner, and A. Shultz. Ultra-high energy neutrino search with the askaryan radio array. *Proceedings of Science (ICRC2017)*, 2017.

[61] J. Goodman. *Statistical Optics*. 1985.

[62] A. Horneffer, W. D. Apel, J. C. Arteaga, T. Asch, J. Auffenberg, F. Badea, L. Bahren, K. Bekk, M. Bertaina, P. L. Biermann, J. Blumer, H. Bozdog, I. M. Brancus, M. Bruggemann, P. Buchholz, S. Buitink, E. Cantoni, A. Chiavassa, F. Cossavella, K. Daumiller, V. de Souza, F. Di Pierro, P. Doll, R. Engel, H. Falcke, M. Finger, D. Fuhrmann, H. Gemmeke, P. L. Ghia, R. Glasstetter, C. Grupen, A. Haungs, D. Heck, J. R. Horandel, T. Huege, P. G. Isar, K. . Kampert, D. Kang, D. Kickelbick, Y. Kolotaev, O. Kromer, J. Kuijpers, S. Lafebre, 2. Luczak, H. J. Mathes, H. J. Mayer, J. Milke, B. Mitrica, C. Morello, G. Navarra, S. Nehls, A. Nigl, J. Oehlschlager, S. Over, M. Petcu, T. Pierog, J. Rautenberg, H. Rebel, M. Roth, A. Saftoiu, H. Schieler, A. Schmidt, F. Schroder, O. Sima, K. Singh, M. Stumpert, G. Toma, G. C. Trinchero, H. Ulrich, W. Walkowiak, A. Weindl, J. Wochele, M. Wommer, J. Zabierowski, and J. A. Zensus. Detecting radio pulses from air showers. In *2008 IEEE Nuclear Science Symposium Conference Record*, pages 3339–3346, Oct 2008.

[63] Noaa space weather prediction center, solar cycle progression. `https://www.swpc.noaa.gov/products/solar-cycle-progression`. Accessed: 2018-05-12.

[64] S. W. Barwick et al. Design and Performance of the ARIANNA HRA-3 Neutrino Detector Systems. *IEEE Trans. Nucl. Sci.*, 62(5):2202–2215, 2015.

[65] Rishabh Khandelwal, Ming-Yuan Lu, and Albrecht Karle. Askaryan radio array neutrino detector: status and design considerations for the future. APS April Meeting, 2018.

[66] R. Bard et al. Design and Performance of an Interferometric Trigger Array for Radio Detection of High-Energy Neutrinos. 2018.

[67] Jordan C. Hanson and Amy L. Connolly. Complex Analysis of Askaryan Radiation: A Fully Analytic Treatment including the LPM effect and Cascade Form Factor. *Astropart. Phys.*, 91:75–89, 2017.

[68] Compac Development Corporation. Catalog. Accessed: 2018-05-29.

[69] Defense Logistics Agency. Chemical conversion coatings on aluminum and aluminum alloys, 2006. Accessed: 2018-05-29.

[70] IEEE. 1901-2010 - ieee standard for broadband over power line networks: Medium access control and physical layer specifications. 2010.

[71] Julia Truchsess. Simple circuit communicates over low-voltage power lines, 2013. Accessed: 2018-05-09.

[72] Linear Technology. Ltc4365 - overvoltage, undervoltage, and reverse suply protection controller, 2016.

[73] Vishay Intertechnology. Dual n-channel 60-v (d-s) 175c mosfet, 2009.

[74] Texas Instruments. Msp430g2x53, msp430g2x13 mixed signal microcontroller. Accessed: 2018-05-23.

[75] On Semiconductor. Lp2950, lp2951, ncv2951: 100 ma, low power, low dropout voltage regulator, 2018. Accessed: 2018-05-24.

[76] Linear Technology. Ltc4365 - overvoltage, undervoltage, and reverse suply protection controller, 2016. Accessed: 2018-06-06.

[77] Laird. Bmi-s-202 two piece shield, 2018. Accessed: 2018-06-06.

[78] Texas Instruments. Tps25921a 18v, 1.6a, 90 mohm efuse with adjustable accurate current limit, 2015. Accessed: 2018-06-07.

[79] Samtec Technologies. Isorate rf systems, 2016.

[80] Wurth Electronik. We-cbf smd emi suppression ferrite bead. Accessed 2018-06-11.

[81] Qorvo Semiconductor. Rfsa3713, 2016.

[82] P. Allison et al. Measurement of the real dielectric permittivity epsilon_r of glacial ice. *ArXiv e-prints*, December 2017.

[83] Qorvo Semiconductor. Ag201-63g, 2009. Accessed 2018-06-08.

[84] Analog Devices. Adl5544. Accessed 2018-06-08.

[85] Susumu International. Precision chip attenuators pat series. Accessed 2018-06-08.

[86] Mini-Circuits. Hfcv-145+ high pass filter. Accessed 2018-06-11.

[87] Mini-Circuits. Lfcn-630+ low pass filter. Accessed 2018-06-11.

[88] Mini-Circuits. Dbtc-10-13lx+ directional coupler. Accessed 2018-06-11.

[89] QucsStudio. About qucsstudio. Accessed 2018-06-12.

[90] John Daniel Kraus. *Radio Astronomy.* Cygnus-Quasar Books, New York, NY, 1986.

[91] Thomas Meures. *Development of a Sub-glacial Radio Telescope for the Detection of GZK Neutrinos.* PhD thesis, Vrije Universiteit Brussel, Springer International Publishing, 2015.

[92] A. Romero-Wolf et al. An interferometric analysis method for radio impulses from ultra-high energy particle showers. *Astropart. Phys.*, 60:72–85, 2015.

[93] P. Allison et al. Constraints on the Ultra-High Energy Neutrino Flux from Gamma-Ray Bursts from a Prototype Station of the Askaryan Radio Array. *Astropart. Phys.*, 88:7–16, 2017.

[94] P. Allison et al. Observation of Reconstructable Radio Emission Coincident with an X-Class Solar Flare in the Askaryan Radio Array Prototype Station. 2018.

[95] I. Kravchenko, D. Besson, and J. Meyers. In situ index-of-refraction measurements of the South Polar firn with the RICE detector. *Journal of Glaciology*, 50:522–532, 2004.

[96] Math Works. xcorr. `https://www.mathworks.com/help/matlab/ref/xcorr.html`, 2019. [Online; accessed 20-June-2019].

[97] P. Schellart et al. Detecting cosmic rays with the LOFAR radio telescope. *Astron. Astrophys.*, 560:A98, 2013.

[98] Brian Dailey. *Analysis of the second flight of the ANtarctic Impulsive Transient Antenna with a focus on filtering techniques.* PhD thesis, The Ohio State University, 2017.

[99] P. W. Gorham et al. Observation of an Unusual Upward-going Cosmic-ray-like Event in the Third Flight of ANITA. *Phys. Rev. Lett.*, 121(16):161102, 2018.

[100] M. Tanabashi and others. Review of particle physics. *Phys. Rev. D*, 98:030001, Aug 2018.

[101] Amy Connolly, Robert S. Thorne, and David Waters. Calculation of High Energy Neutrino-Nucleon Cross Sections and Uncertainties Using the MSTW Parton Distribution Functions and Implications for Future Experiments. *Phys. Rev.*, D83:113009, 2011.

[102] Gary J. Feldman and Robert D. Cousins. A Unified approach to the classical statistical analysis of small signals. *Phys. Rev.*, D57:3873–3889, 1998.

[103] Plataforma solar de almeria sun position algorithm. `http://www.psa.es/sdg/sunpos.htm`. Accessed: 2017-01-15.

[104] Manuel Blanco-Muriel, Diego C. Alarcón-Padilla, Teodoro López-Moratalla, and MartÍn Lara-Coira. Computing the solar vector. *Solar Energy*, 70(5):431 – 441, 2001.

[105] Mukul R. Kundu. *Solar Radio Astronomy.* 1965.

[106] M. J. Reiner, M. L. Kaiser, J. Fainberg, J.-L. Bougeret, and R. G. Stone. Remote Radio Tracking of Interplanetary CMEs. In A. Wilson, editor, *Correlated Phenomena at the Sun, in the Heliosphere and in Geospace*, volume 415 of *ESA Special Publication*, page 183, December 1997.

[107] Stereo / waves science investigation. `https://swaves.gsfc.nasa.gov/swaves_science.html`. Accessed: 2017-05-12.

[108] Markus Aschwanden. *Physics of the Solar Corona.* 2005.

[109] J P Wild. Observations of the spectrum of high-intensity solar radiation at metre wavelengths. ii. outbursts. *Australian Journal of Scientific Research A*, 3:399, 1950.

[110] J.L. Meléndez, H.S. Sawant, F.C.R. Fernandes, and A.O. Benz. Statistical analysis of high-frequency decimetric type iii bursts. *Solar Physics*, 187:77–88, 1999.

[111] A. R. Jacobson, S. O. Knox, R. Franz, and D. C. Enemark. Forte observations of lightning radio-frequency signatures: Capabilities and basic results. *Radio Science*, 34(2):337–354, March 1999.

[112] K.-S.. Cho et al. A high-frequency type II solar radio burst associated with the 2011 February 13 Coronal Mass Ejection. *Astropart. Phys.*, 765:148–156, 2013.

Appendix A

APPENDICES TO THE SOLAR FLARE PAPER

A.1 Supporting Figures

A.1.1 Waveforms and Spectra

Fig. A.1 shows waveforms from an event at 02:04 GMT, near the beginning of the flare period. For comparison, in Fig. A.2 we show a typical event from February 11, 2011, four days before the flare of interest when the sun was not active. Figs. A.3 and A.4 show corresponding spectral amplitudes, from the Fourier transforms of the waveforms in the previous two figures.

A.1.2 Rayleigh Fits

An example of fits for four borehole antennas is given in Fig. A.5 and the fit to the CSW of those four boreholes is given in Fig. A.7. This demonstrates again that the single channel spectra, and the CSW, are both thermal in nature.

We also show in Fig. A.6 spectral amplitude distribution and their associated Rayleigh fits for a broader selection of frequencies.

A.1.3 Polarization

As seen in Fig. 2.4, we find higher peak cross-correlation values from waveforms from VPol antennas than from the HPol antennas. We decided to investigate whether this VPol dominance had originated in the sun, or rather was due to higher gains in VPol antennas than HPol antennas. Although some solar flares do emit radio that is circularly polarized–thought to be associated with how plasmas generating the radio emission orient themselves with respect to dynamic magnetic field lines–we would expect it to be a coincidence if the emission would happen to prefer the vertical direction at the site of the ARA Testbed.

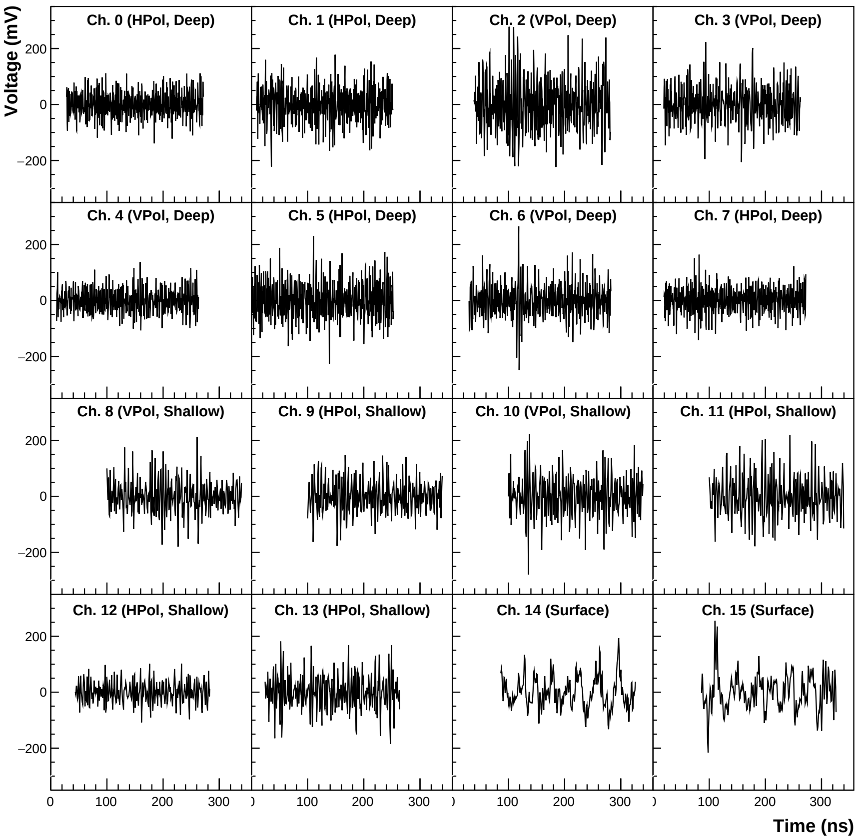


Figure A.1: Example waveform for a typical event during the flare period, recorded at 02:04, Feb 15th 2011 UTC.

In order to measure the relative response of the VPol and HPol antennas in the Testbed *in situ*, we used recorded waveforms from a pulser mounted to the top of the IceCube Counting Laboratory (ICL) at the South Pole, nicknamed the "ICL Pulser." This is a Seavey broadband ($\sim 200-1200$ MHz), dual-polarized quad-ridged horn antenna, the same as is used for the ANITA payload [41], that transmits pulses in both polarizations. The ICL Pulser sits 2.1 km from the Testbed at a 13 m height, so that the pulse destined for a Testbed antenna would approach the ice at approximately $\sim 0.36°$ above the surface. This Seavey antenna is ideally suited for this measurement because its response across the band is very similar in both polarizations [41].

For this study, we used ICL Pulser data taken on January 26th, 2013, during two different forty-minute long runs where the pulser was either transmitting in VPol or HPol mode. We use forty events in each run where we found the highest cross-correlation value for a direction within 10° in azimuth of the expected ICL Pulser position and a zenith angle above the surface. We take the Fourier Transform of each waveform and, for each antenna, average

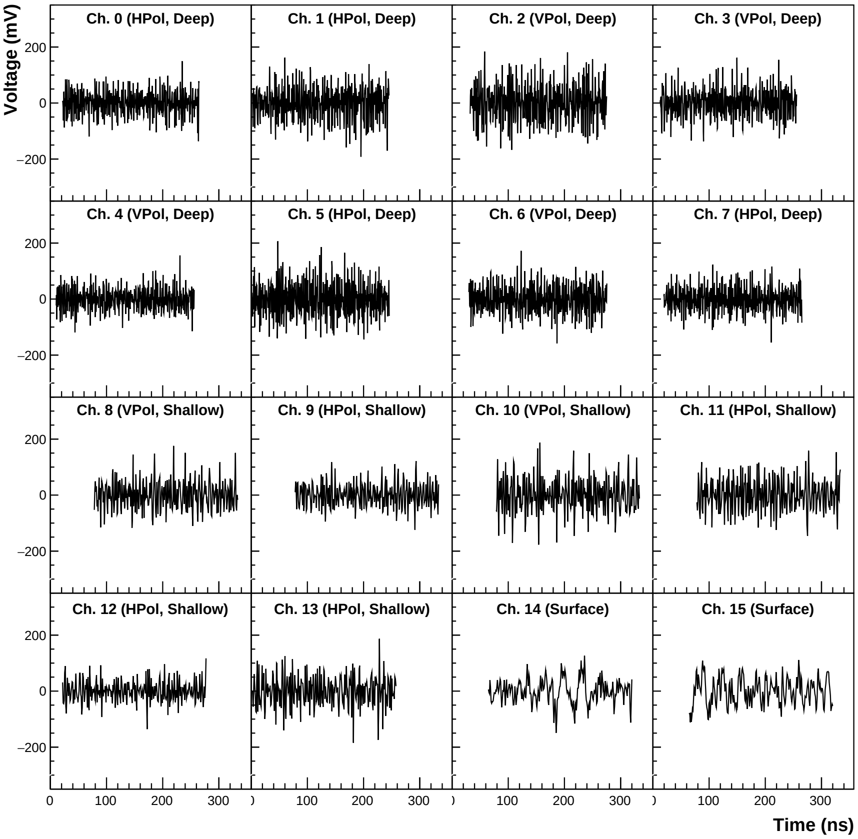


Figure A.2: Example waveforms for a typical non-flare event, recorded at 02:04, Feb 11th 2011 UTC.

the spectral amplitudes for the forty events in a run.

We use the following procedure to measure the frequency-dependent ratio of HPol to VPol spectral amplitudes. We considered pairs of antennas that were deployed in the same borehole and within 3-6 m of each other in depth, and for each pair, took the ratio of the average spectral amplitudes measured in the HPol during HPol transmission mode to the average spectral amplitudes measured in the VPol during VPol transmission mode, after accounting for the different Fresnel coefficients governing the fraction of field-amplitude transmitted at the air-ice boundary for each polarization , t_H and t_V. For the 89.64° angle with respect to normal at air-to-ice interface, the transmission coefficients are $t_V = 0.0185$ for VPol signals and $t_H = 0.0138$ for HPol signals using an index of refraction at the surface of 1.35. The four ratios are then averaged together.

To compute the Fresnel coefficients for the antenna calibration, we assume that the ICL roof-top pulser is 13.16m above the ice at a distance of 2099.06m away (accounting for differences in ice elevation at the location of Testbed and the ICL). This leads to a angle

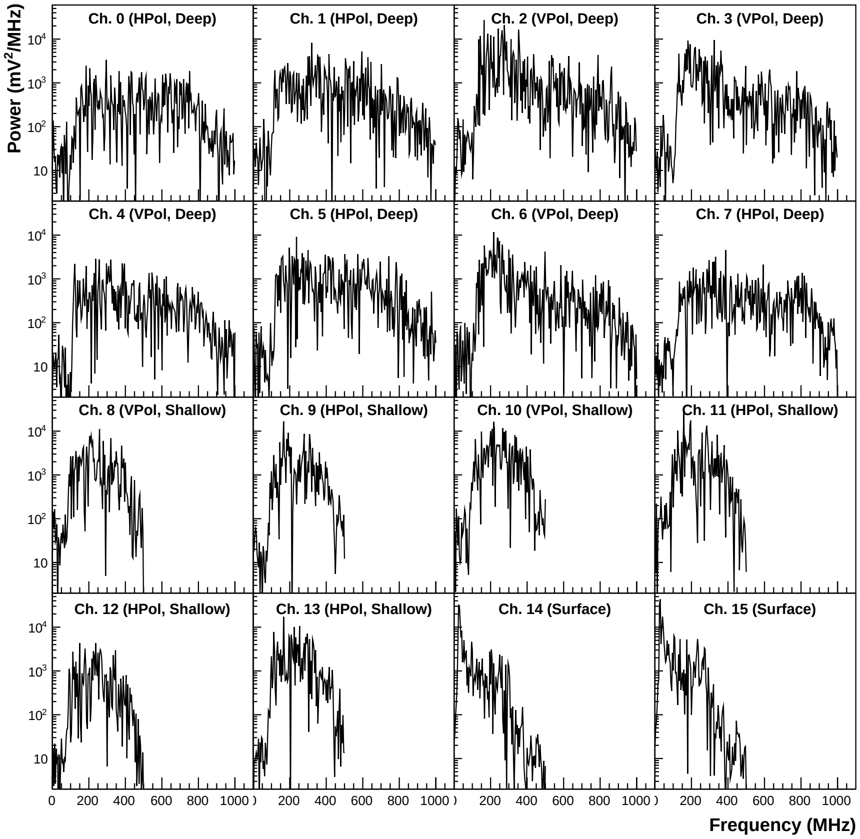


Figure A.3: Example amplitude spectra for the same event during the flare as in Fig. A.1, recorded at 02:04, Feb 15th UTC.

from the surface of $\theta_{pulser} = \tan^{-1}(\frac{13.16\text{m}}{2099.06\text{m}}) = 0.36°$ or $89.64°$ from surface normal.

The transmission coefficient in *voltage* for the polarization parallel to the normal (perpendicular or the surface of the ice, our VPol) is given by t_V:

$$t_V = \frac{2n_{\text{air}}\cos\theta_i}{n_{\text{air}}\cos\theta_t + n_{\text{firn}}\cos\theta_i} = \frac{2\sin\theta_t\cos\theta_i}{\sin(\theta_i+\theta_t)\cos(\theta_i-\theta_t)} \tag{A.1}$$

The transmission coefficient in *voltage* for the polarization perpendicular to the normal (parallel to the surface of the ice, our HPol) is given by t_H:

$$t_H = \frac{2n_{\text{air}}\cos\theta_i}{n_{\text{air}}\cos\theta_i + n_{\text{firn}}\cos\theta_t} = \frac{2\sin\theta_t\cos\theta_i}{\sin(\theta_i+\theta_t)} \tag{A.2}$$

where $n_{\text{air}} = 1$ is the index of refraction in air, $n_{\text{firm}} = 1.35$ is the assumed index of refraction in the shallow ice, θ_t is the transmission angle with respect to normal, and θ_i is the incidence angle with respect to normal, and the relationship between θ_t and θ_i is found by Snell's law.

For the solar flare sample–namely all 2323 events passing cuts– we first build coherent

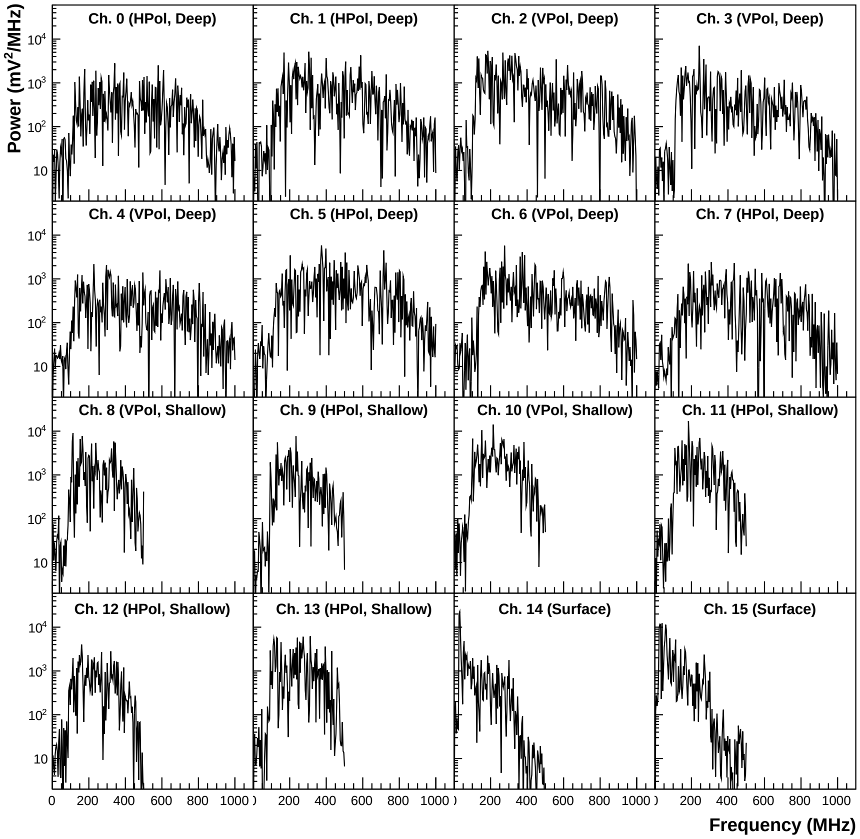


Figure A.4: Amplitude spectra for the same typical event in Fig. A.2 from a time where the sun was quiescent, at 02:04, Feb 11th UTC.

sums of every event in both polarizations. The coherent sum waveforms are then Fourier transformed and averaged, accounting for Fresnel coefficients for the incidence angle of the sun (which is 13.3° above the surface). The ratio of the spectral averages is then corrected by the known relative antenna responses found through the ICL calibration procedure above. Fig. A.9 summarizes the results of this investigation of the polarization of the flare emission. The plots show (left) that for these shallow incidence angles, the HPol antenna response is intrinsically weaker than the VPol antenna response across the band from ~220 MHz to 500 MHz by about a factor of three, and that (right) once this is accounted for, the emission from the flare is consistent with having equal contributions in HPol and VPol. The anomalies at ~150 MHz are likely due to a difference in the gains of the two types of antennas in this region, and can be seen as a calibration artifact in the left panel.

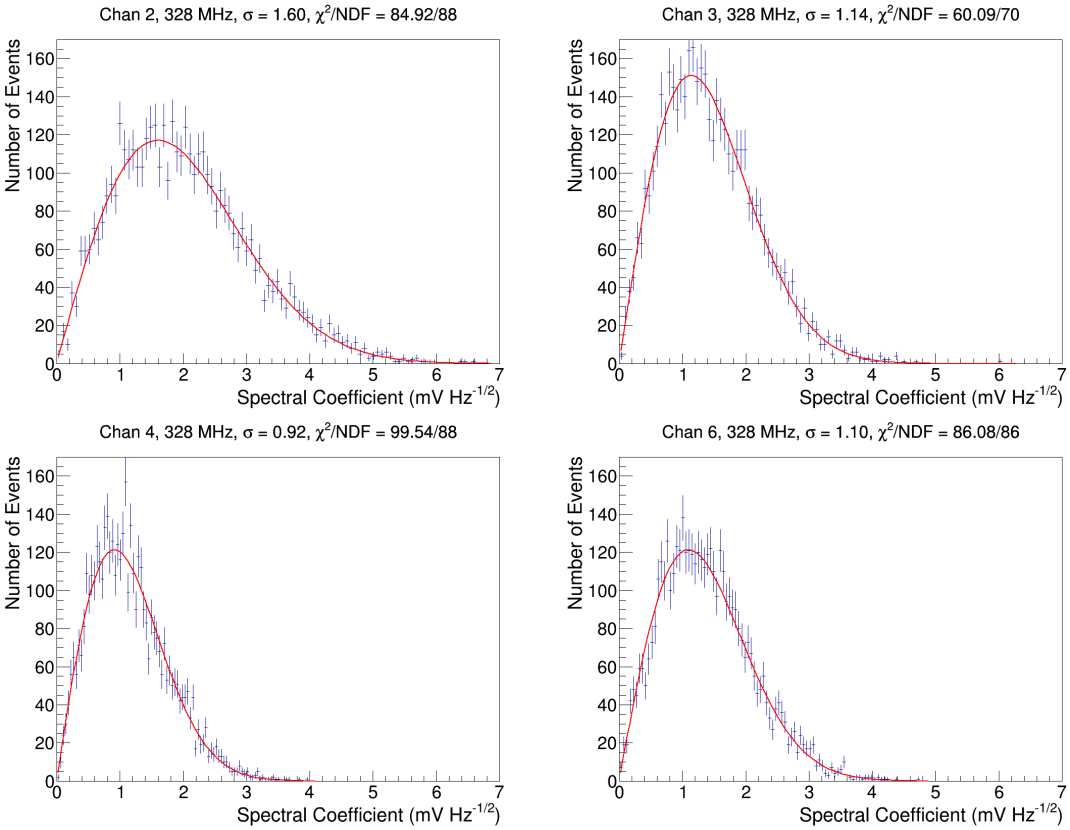


Figure A.5: Distribution of spectral amplitudes at 328 MHz for the four borehole VPol antennas in the Testbed.

A.2 Coordinate Systems

A.2.1 Solar Coordinates

The location of the sun is computed with a publicly available C library [103] developed by the Platforma Solar de Almería (PSA). For a given unixtime, latitude, and geographic longitude, the program returns the position of the sun (both azimuth and zenith angle) to within 0.5 arcminutes. The PSA algorithm is chosen for its simplicity of implementation while preserving sub-degree solar position accuracy for the years 1999-2015 [104].

The PSA method returns the sun location in a coordinate system where zero in azimuth is defined by the meridian of the observer, in this case, the longitude of the prototype station, which is 74.22° W from Prime Meridian (rounded to the nearest hundredth). We first rotate this reported value to align with the geographic grid of the South Pole, where zero azimuth is defined by the Prime Meridian. This “continent” coordinate system natively follows a *clock-wise* convention where N(orth) = 0° = Prime Meridian, E(ast) = 90°, S(outh) = 180°, and W(est)= 270°. We further transform coordinate systems to one which proceeds *counter-clockwise* in the physics tradition, so E = 0°, N = 90° = Prime Meridian, W = 180°, and S = 270°. We finally rescale the azimuth range from $[0, 360]$ to $[-180, 180]$ to match the output of the analysis interferometric map methods described in section A.2.2. The result is that the solar azimuth is reported as a counter-clockwise angle in the range of $[-180, 180]$ in a coordinate system where E = 0°, N = 90° = Prime Meridian, W = 180°, and S = −90°.

A.2.2 Interferometric Map Coordinates

The interferometric map measures azimuthal angles in a counter-clockwise coordinate system with zero azimuth corresponding to the direction of Antarctic continent ice-flow. At the location of the prototype station, this corresponds to 36.77° W from Prime Meridian (rounded to the nearest hundredth). We rotate the map by this constant offset to align the map with the same "continent" coordinate system as used for solar positions in section A.2.1, so that E = 0°, N = 90° = Prime Meridian, W = 180°, and S = −90°. The interferometry routines are already designed to report answers in the range of $[-180, 180]$ as in [53], so unlike the PSA routine, a range adjustment is not required.

A.2.3 A note On Convention Selection

We note that the azimuthal (ϕ) coordinate system utilized here (E = 0°, N = 90° = Prime Meridian, W = 180°, and S = −90°) is 90° out of phase with the ISO 6709 definition of longitude. It does however agree with the methodology used by South Pole surveyors in determining coordinates for both ARA and IceCube.

A.3 Other Hypotheses For Unique Reconstruction to the Sun

We concluded, as described in the main body of the paper, that the events were consistent with bright thermal emission. In this Appendix, we detail other studies we undertook to explicitly rule out the hypotheses of CW emission, low signal-to-noise transients, broad-band but non thermal emission produced by solar flares.

A.3.1 Continuous Wave

We looked for evidence of the emission being CW-like by looking for a strong peak or peaks at unique frequencies in the spectra. Any impulse whose duration would be shorter than our waveforms of ∼250 ns in length when produced at the sun would be dispersed over several μs in the ionosphere, so we looked for a chirp-like behavior in the waveforms.

Continuous wave signal at frequencies higher than $\sim 1/200ns = 5$ MHz will repeat over timescales less than $\sim 200\ ns$. The ARA Testbed high-pass filters remove frequencies below 150 MHz, so the unique reconstruction to the sun that we are seeing cannot be due to signals being at sufficiently low frequencies that their period is not contained in our waveforms. Fig. A.10 shows a reconstruction map from continuous wave interference at $\sim$403 MHz (the radiosonde on South Pole weather balloons), where one can see peaks in the cross-correlation function corresponding to directions all over the map.

A.3.2 Transients

Any signal that only begins, or ceases, to be observed at some time between the beginning and end of the waveforms would also give a unique reconstruction direction. For example, impulsive signals reconstruct cleanly, as demonstrated from a measured calibration pulser in Fig. A.10. However, any impulsive signal that might be produced by the flare would be dispersed as it exits the sun as well as in the earth's ionosphere.

To check whether the cause of the unique reconstruction of these events to the sun was some emission with start or end times observed in our waveforms, we considered the channel with the largest peak voltage compared to the RMS noise voltage, and divided up the waveforms in the event into three sections: an 80 ns window surrounding the time of the peak, and the periods before and after that window (See Fig. A.11). The windows in the waveforms from each antenna are delayed to be consistent with the Sun's direction so that the windows in each waveform correspond with each other. As seen in Fig. A.12, we find that the event reconstructs uniquely to the sun using any of the three time periods on their own. The same test applied to a calibration pulser event shows a strong reconstruction using only the "impulsive" part of the waveform and no reconstructability to the calibration pulser's direction using the segments before and after the impulsive portion. This suggests that the signal correlated with the solar flares is not impulsive in nature but rather that the signal correlating with the Sun's direction is spread across the waveform.

A.3.3 Broadband, non thermal

Time-dependent spectral features expected from solar flares

Some decimeter-wavelength (decimetric) radio emission associated with solar flares is due to excitations of the coronal plasma at the plasma frequency f_p [105]. This plasma frequency depends on the square root of the electron density n_e, so that $f_p \propto \sqrt{n_e}$. The coronal electron density is in turn proportional to the inverse-square of the distance from the sun's center R so that $n_e \propto 1/R^2$ [106, 107]. Therefore, as the electrons in the ejecta plasma are carried outward from the sun, the coronal density falls as $1/R^2$, leading to a chirp from higher to lower frequencies. This chirp is characterized by its so-called "frequency drift rate" df/dt, where $df/dt > 0$ indicates the electrons are being ejected from the sun, and $df/dt < 0$, or "reverse drift", indicates they are being dragged back in.

The drift rate is a measure of how quickly the density of the solar atmosphere changes from the perspective of an ejecting electron and divides flares with plasma emission into two categories [108], type-II and type-III [1]. The faster the ejecta, the more quickly the change in the plasma frequency, and the greater the magnitude of the frequency drift rate. Type-II flares are characterized by df/dt on the order of 10^{-1} MHz/s, and are usually attributed to magneto-hydrodynamic shocks (shock waves in the solar plasma excited by the evolving magnetic fields) moving through the solar corona [109]. Type-III flares drift significantly faster, upwards of 10^3 MHz/s [110], and are caused by electrons traveling outward along open field lines at speeds upward of $0.5c$.

Looking for features in spectrograms

The time scales of the chirps from both types of flares are orders of magnitude too long to be observable on our individual waveforms of length $\sim$ 250 ns, and cannot explain the correlation that we are seeing between waveforms on an event by event basis. In our bandwidth (approximately 125-850 MHz), the fastest drift rates are those associated with type-III flares. Melendez *et. al.* [110] suggest that for our center frequency of 350 MHz, $df/dt \sim 700$ MHz/s, which would generate roughly 170 Hz of drift over the waveform. Such a chirp should only be visible on a timescale of $\sim 1/170\text{Hz} \sim 6$ ms, which is over 10,000 times longer than our waveform.

Impulsive events with time-scales of nanoseconds would be dispersed in the ionosphere due to the electron plasma frequency resonance in the 10-100 MHz range –as has been observed by other impulsive radio experiments like FORTE [111]. The ionospheric plasma frequency is even greater during active solar periods thus producing an even greater dispersive effect with group delays potentially reaching 1μs at 300 MHz. Therefore, if we were seeing the effect of impulsive signals that were of $\sim$ns or 10's of ns duration when produced at the sun, one may expect to find a chirp on the scale of 10's to 100's of nanoseconds. Because of the 25-ms deadtime between triggered events, μs-scale and ms-scale spectrograms cannot be obtained. However, we can still check for features in spectrograms from a single event over a 250 ns time scale, and using $\sim$ 50 events we can produce spectrograms over timescales of order a second.

Using $\sim$ 50 events, we can produce second-scale spectrograms, an example of which can be seen in Fig. A.13. No clear pattern emerges from these spectrograms.

By examining subsections of each event, we produce event spectrograms,which demonstrate no clear chirp-like behavior on the time-scale of the event length ($\sim$250 ns). We divide each waveform into many segments and calculate the spectrum for each segment.

---

[1] The numbering on the flare types is arbitrary and historical. At the time of naming, type-II flares were "slow-drift" bursts, and type-III were "fast-drift bursts", referring to the time scale of the frequency drift. Type-I flares were "noise-storms", type-IV were "broadband-continuum emission," and type-V were "continuum emission at meter wavelengths."

Date	Time	Class
2011/02/15	01:43:44	X2.2
2011/03/09	23:10:28	X1.5
2011/09/22	10:53:16, 11:06:40, 11:36:04	X1.4
2011/09/24	09:19:28, 9:30:00	X1.9
2012/03/05	02:41:48, 03:12:20, 03:16:28, 03:56:04	X1.1
2012/10/23	03:14:36	X1.8
2013/10/25	07:52:44	X1.7
2013/10/25	14:56:16	X2.1
2013/10/28	01:45:24, 01:51:44	X1.0

Table A.1: Table summarizing X-class flares during the livetime of the ARA Testbed and three deep stations and the periods when the Sun is above the horizon at the South Pole. Flares with multiple times indicate an active period where X-ray emission rose and fell repeatedly within a short time frame but the X-ray emission is integrated over those subflares. This list of flares was extracted from the RHESSI catalog [54].

Although the frequency resolution is made coarser by this process than when analyzing the full waveform, it allows the examination of the power spectrum as a function on time during the period covered by the waveform.

We confirm the absence of single-waveform chirp behavior by plotting spectrograms for single waveforms as in Fig. A.14.

A.4 Other solar flares during the ARA livetime

In order for ARA to observe radio emission from a solar flare during the livetime of the instrument, the sun should be up in Antarctica (from mid-September to mid-March only, and 24 hours per day during that period), the flare must face the earth. Table A.1 summarizes other X-class flares that satisfied the above criteria for the ARA Testbed and three deep stations from 2011 to 2013 using the RHESSI flare catalog.

A.4.1 The Feb. 13th solar flare reported by ARA

The ARA team observed emission from an M-class flare in the ARA Testbed that occurred on Feb. 13th, 2011, just two days before the flare that was the subject of this paper, and reported it in a paper about the performance of the Testbed instrument in 2011 [52]. This Feb. 13th flare was sought by ARA analyzers after the Green Bank Solar Radio Burst Spectrometer detected a series of strong type-II solar radio bursts in the 10 MHz-1 GHz band [112]. ARA reported a spectrogram from the Feb. 13th flare with many features similar to ones in the spectrogram report by Green Bank, demonstrating that the ARA Testbed system was sensitive to low-level radio emission, and also that the noise environment of the

South Pole ARA site is sufficiently quiet to allow for such a detection.

The Feb. 13th and Feb. 15th flares had many differing characteristics that account for the different circumstances under which ARA analyzers encountered the two events. The emission from the Feb. 13th flare peaked at frequencies below about 150 MHz, and the ARA Testbed search for diffuse neutrinos rejected events with more than 10% of the power accounted for below 150 MHz. This cut was made to reduce anthropogenic background noise. In addition, the Feb. 13th flare did not produce any events that survived the "Reconstruction Quality Cut" imposed in the neutrino search, which required that the region in the sky that gives a high cross-correlation be narrow and unique. Fig. A.15 shows a typical event during the time of the Feb. 13th flare. This map shows a broader range of directions that give cross-correlation values similar to the peak value, compared to the narrow peak typical of an event from the Feb. 15th flare. Additionally, the results shown in the previously published work on the Feb. 13th flare [52] only used the surface antennas which operate in a lower frequency band (30-300 MHz) than the borehole antennas used in this work (150-850 MHz).

The solar flare events from Feb. 13th failed to pass the "Reconstruction Quality Cut," which requires that the reconstruction map of a "well-reconstructed" event be both well-defined and unique. The key parameter of the cut is the area of the 85% contour around the absolute peak of the map. The conditions of the cut are two-fold. First, this area must be less than 50 square degrees in magnitude. This indicates that the reconstruction is "well-defined," and does not allow for a broad range of possible reconstruction locations within the peak contour. Second, the area around this peak must be no less than two-thirds of the total area on the map with a value of 85% of the peak or greater. This condition forces the event to be "unique," in that no other location on the sky reconstructs as well. A fuller description of the requirement can be found in [53].

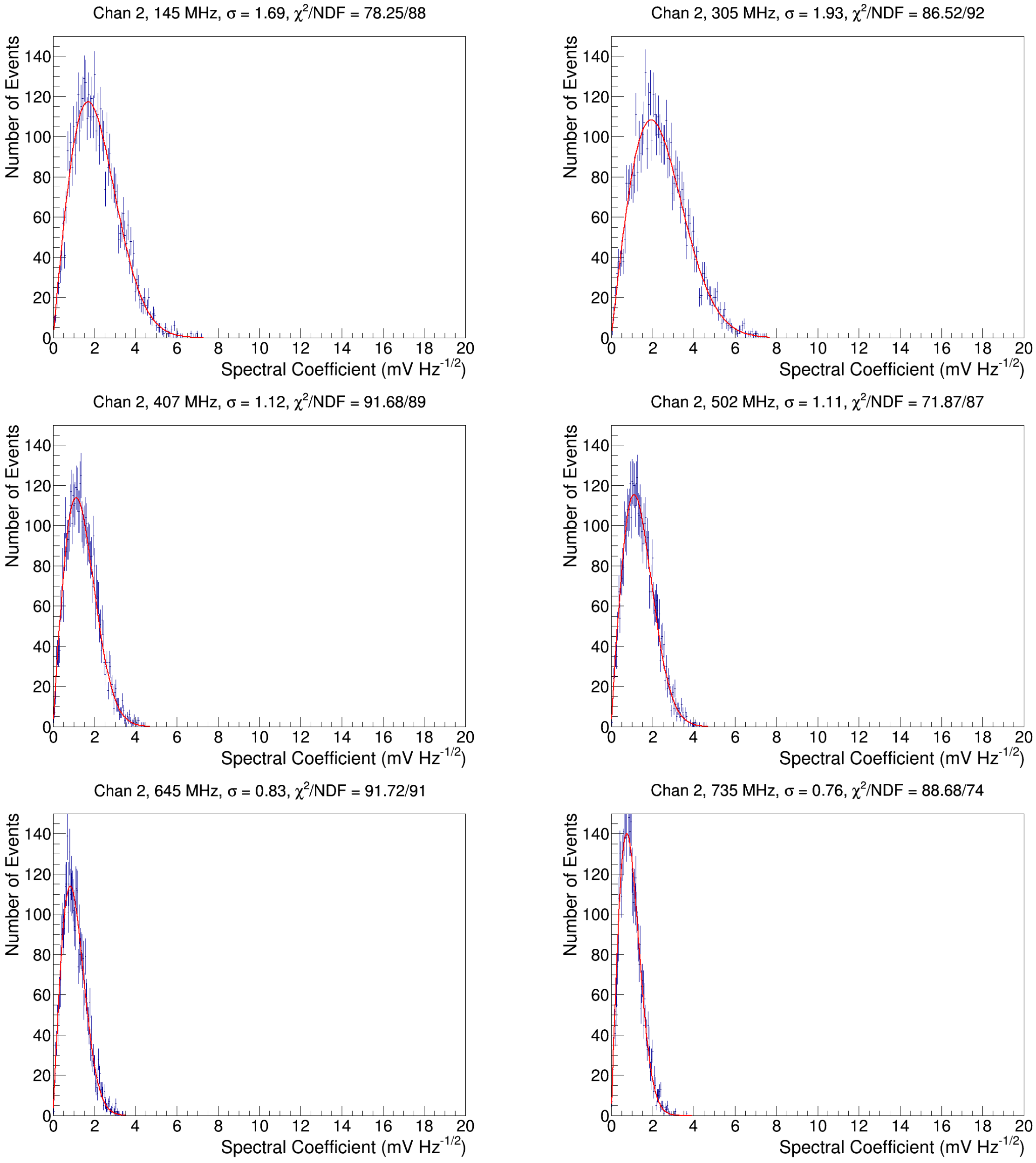


Figure A.6: A selection of spectral amplitudes for several frequencies across our band with their best fit Rayleigh's superimposed.

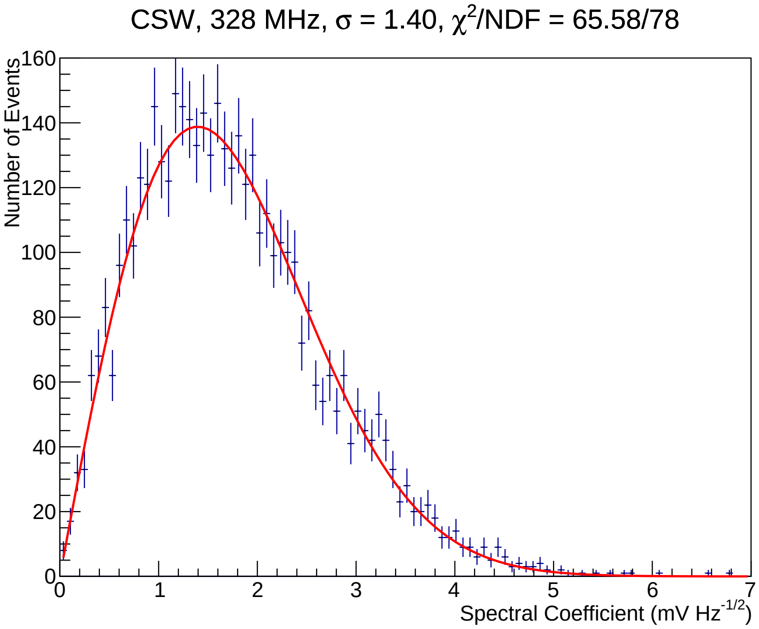
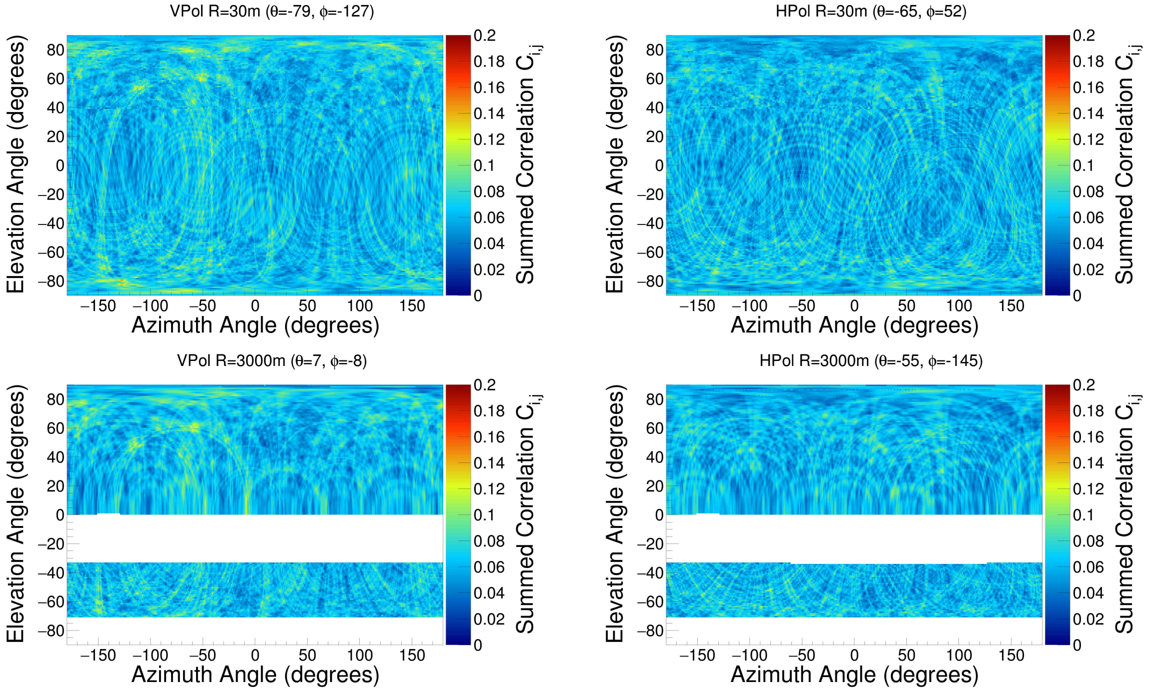


Figure A.7: Distribution of spectral amplitudes at 328 MHz for the coherently summed waveform derived from the four borehole VPol antennas in that are shown in Fig. A.5. The CSW is made with directional hypothesis time lags corresponding to the correlation peak on the map for a given event.

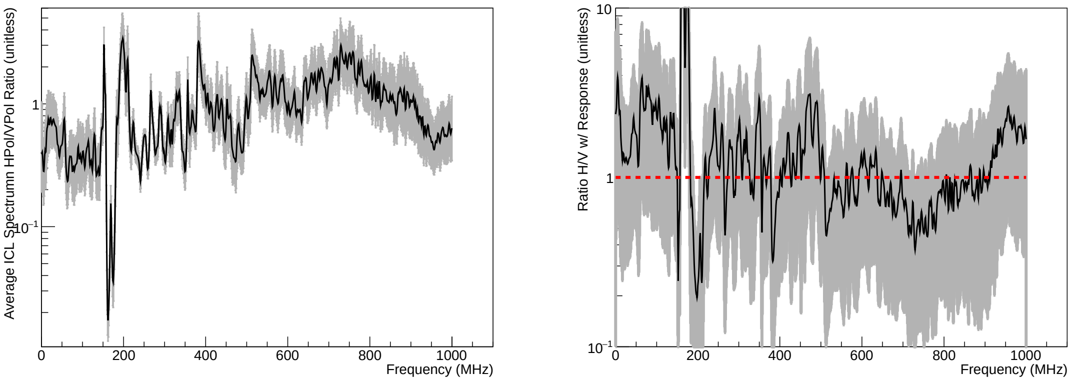

VPol R=30m (θ=-79, ϕ=-127)

HPol R=30m (θ=-65, ϕ=52)

VPol R=3000m (θ=7, ϕ=-8)

HPol R=3000m (θ=-55, ϕ=-145)

Figure A.8: Map of cross-correlation values associated with directional hypotheses across the sky for the same event from a time when the sun was quiescent as the one shown above in Figs. A.2 and A.4. The upper two figures show reconstruction maps for the hypothesis of 30 m distance, while the lower two figures show reconstruction maps for the 3000 m distance hypothesis. The maps are created in the Testbed local coordinate system.



(a) H/V ratio of the Antenna Response

(b) H/V ratio of the solar flare spectra

Figure A.9: (a) The antenna response of HPol antennas in ARA Testbed compared to that of VPol antennas, showing that the HPol antennas have lower gains by about a factor of three. (b) The observed emission from the solar flare in VPol and HPol after accounting for the different gains from the two types antennas, showing that the observed emission during the flare was consistent with being observed equally in horizontal and vertical polarizations. Note the artifact around 150 MHz is related to the calibration and is visible in the left figure, and not due to the solar flare emission.

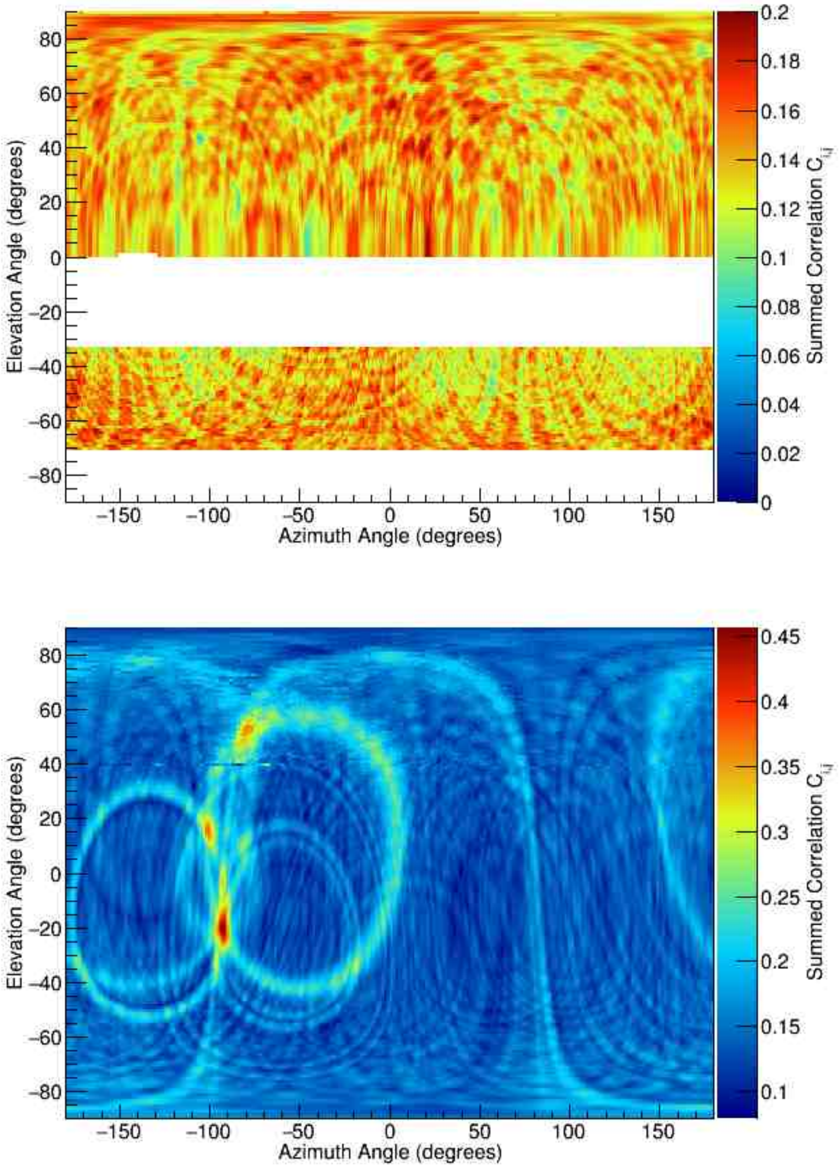


Figure A.10: (Top) Reconstruction map from known continuous wave interference at ∼403 MHz (the radiosonde on South Pole weather balloons), showing peaks in the cross-correlation function for time delays corresponding to many different directions. The behavior differs from the reconstruction maps seen for the flare as shown in Fig. 2.4. The map was made assuming a source distance of 3000m, and the angles are all in the local Testbed coordinate system. (Bottom) Reconstruction map from an impulsive waveform from an in-ice calibration pulser. The map was made assuming a source distance of 30m, and the angles are all in the local Testbed coordinate system.

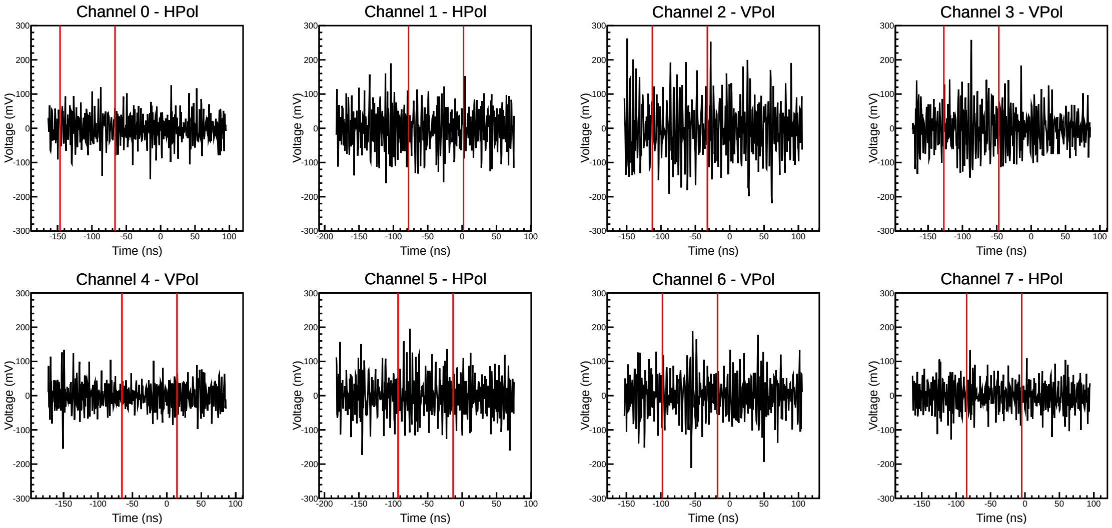
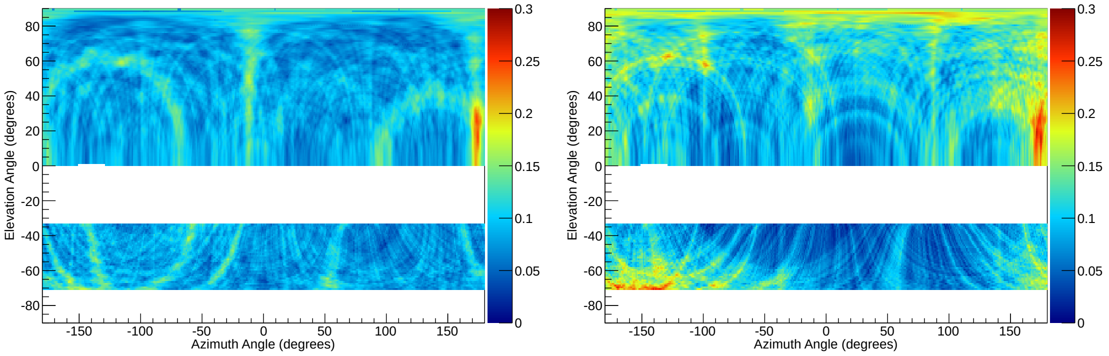
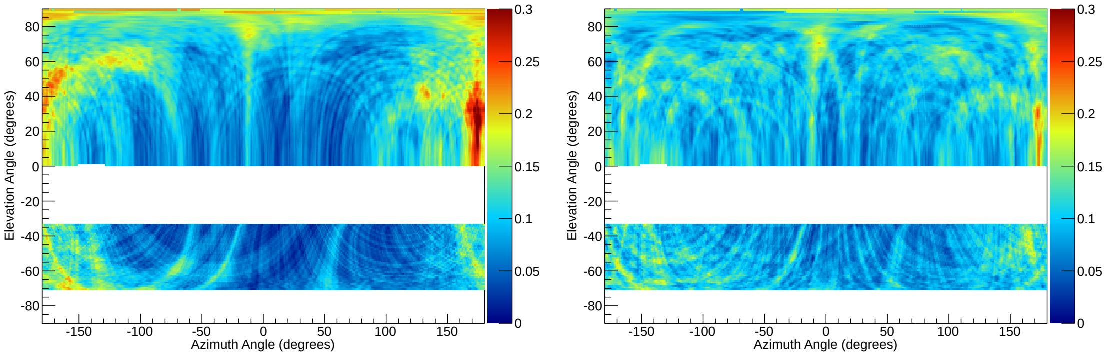


Figure A.11: The waveforms from an event that reconstructs to the Sun's direction. To test whether the reconstruction is caused by an impulsive signal, the impulsive portion is defined by identifying the largest V_{peak}/RMS values across all channels contributing to the map (i.e. VPol borehole antennas) and selecting an 80-ns window around that time. For other channels, the window is defined by shifting the window in the peak waveform by the delay consistent with the peak reconstruction direction. The "impulsive" period is shown as the region between the red vertical lines.



(a) Full waveform map

(b) "Before" map



(c) "During" map

(d) "After" map

Figure A.12: The reconstruction maps for the event shown in Fig. A.11 using (a) the full waveform, (b) only the "before" section of the waveform, (c) only the "during section, and (d) only the "after" section.

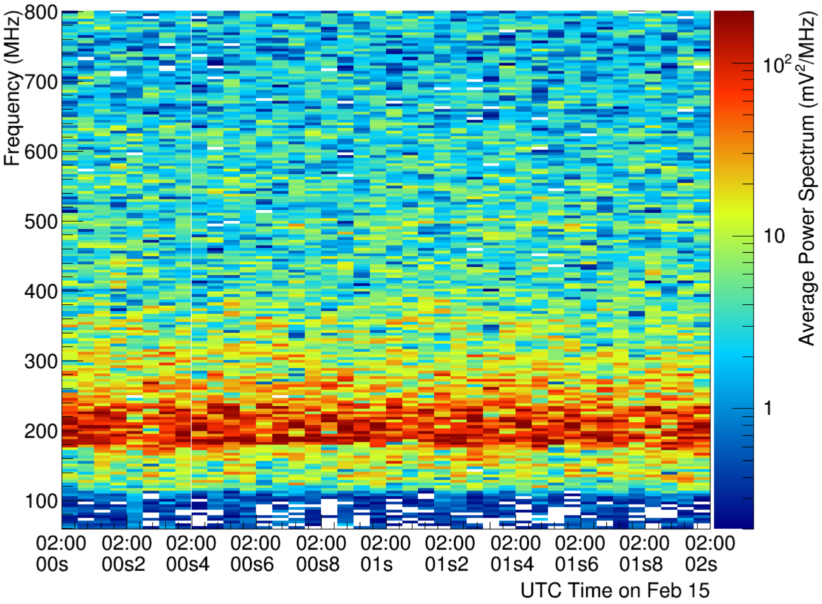
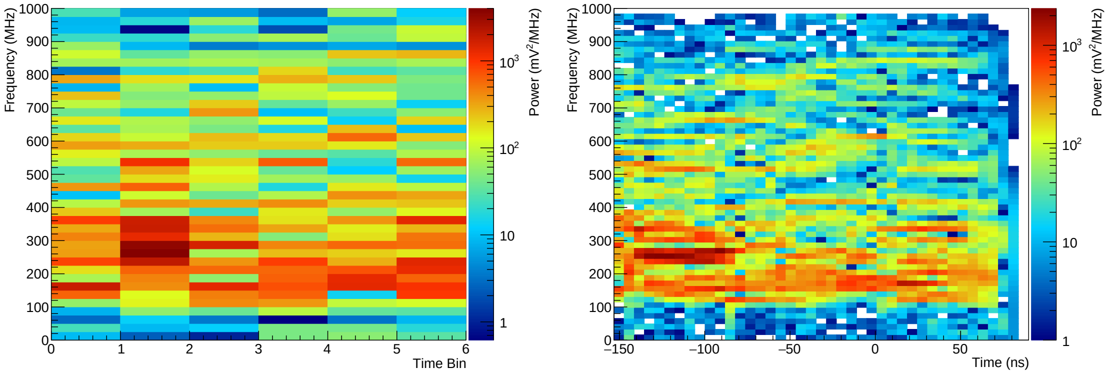


Figure A.13: A spectrogram produced over a two second-long scale using 48 sequential events during the flare period. The events were recorded as RF triggers between 2:00:00 and 2:00:02 on Feb. 15th, 2011.



(a) Spectrogram with independent time bins

(b) Spectrogram from running sub-samples of the waveform

Figure A.14: Spectrogram created from a single waveform, measured in Channel 2, from the same event as shown in Figs. A.1, A.3 and 2.4Figure (a) shows the spectrogram from 6 independent sub-samples of the waveform. Figure (b) shows the spectrogram taken from a running sub-window of the waveform. Although each sample is correlated with its neighbors in time, this approach yields finer frequency resolution due to the greater number of samples allowed.

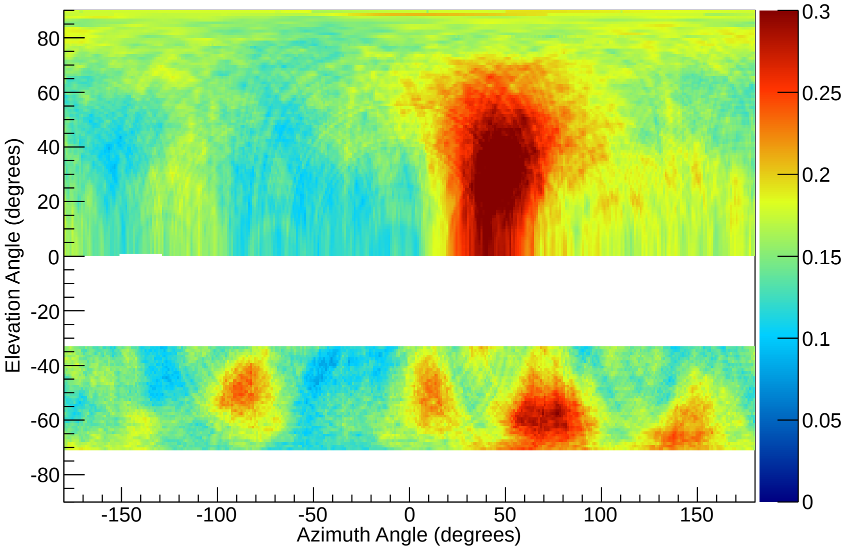


Figure A.15: Reconstruction map for a typical event during the time of the Feb. 13th flare. This map shows a broader range of directions that give cross-correlation values similar to the peak value, compared to the narrow peak typical of an event from the Feb. 15th flare.

Appendix B

Supporting Information for the A2 and A3 Diffuse Analysis

B.1 Coordinate Systems

The coordinate system utilized by the South Pole surveyors has an origin at the location of the former South Pole Station, known as the Dome. The origin of this system is set to a value of (50000' Easting, 50000' Northing) (in feet) to avoid negative Eastings and Northings in most situations. This is (15.240 km, 15.250 km). The $\hat{x}$-axis of the SP surveyor coordinate system is parallel with the local "Easting" vector, and the $\hat{y}$-axis of the SP surveyor coordinate system is parallel with the local "Northing" vector. The Northing vector, and therefore the local $\hat{y}$-axis, is defined to lie along the Prime Meridian. In this way, the definition of geographic North is the Prime Meridian, and and coincides with an azimuth of 90°, so that E = 0°, N = 90° = Prime Meridian, W = 180°, and S = −90°.

B.1.1 A2 and A3 Local Coordinates

To transform into station local coordinates for stations A2 and A3, a translation and rotation is needed. First, the origin of the coordinate system is moved to the location of the DAQ box, which are presented in South Pole surveyor coordinates in Tab. B.1. Then, the orientation of the coordinate system is rotated in azimuth to align with the direction of continental ice-flow, which is 36.773° W. In this way, the station-local coordinate systems has the $\hat{x}$-axis rotated 126.773° counter-clockwise from the surveyor coordinate system. Note that because the ice is flowing, the location of fundamentally fixed landmarks, like the *true* geographic South Pole, move by about ten meters every year. The values reported here for the location of the South Pole were the measured values on January 1, 2011. Objects that sit on the ice flow and are co-moving with the coordinate system origin, such as the ICL and WT3, do not have time-dependent positions.

The locations of the major landmarks at the South Pole are presented in Tab. B.2, in both global surveyor coordinates and station-local coordinates. I also provide the azimuth

Station	Global X	Global Y
A2	10.829	13.833
A3	9.829	15.565

Table B.1: Table of the coordinates of A2 and A3 in the coordinate system of the South Pole surveyor, in kilometers, rounded to the nearest meter.

Object	Global X	Global Y	A2 X	A2 Y	A2 ϕ	A3 X	A3 Y	A3 ϕ
SP	15.156	15.401	-1.342	-4.407	$-106.9°$	-3.329	-4.169	$-128.6°$
WT3	12.516	15.346	-0.330	-3.902	$-94.8°$	-2.317	-3.665	$-122.3°$
ICL	14.152	15.903	0.203	-2.258	$-84.9°$	-1.783	-2.021	$-131.4°$

Table B.2: Table of the coordinates of major geographic landmarks in both global and station-local coordinates, in units of kilometers, rounded to the nearest meter. The table also presents the azimuth (ϕ) of the landmarks in the station local coordinates, rounded to the nearest tenth of a degree.

of the landmarks in station local coordinates, which is where the objects would be found in reconstruction skymaps.

B.2 Excluded Livetime

The following are the periods of livetime that are excluded from the analysis because of anthropogenic activity as identified by my undergraduate colleagues. They identified approximately 27 hours of bad livetime in A2 and 54 hours of bad livetime in A3. The full list of excluded livetimes for A2 is given in Tab. B.3 and for A3 in Tab. B.4.

Table B.3: The bad livetime periods identified for A2.

A2 Bad Livetime Periods		
A2 Run	Unixtime Start	Unixtime Stop
2868	1389381600	1389384000
4775	1420317600	1420318200
6507	1449189600	1449190200
2575	1382712900	1382713500
2589	1382972700	1382973300
2642	1383884400	1383886200
2652	1384060200	1384061100
2677	1384487400	1384489800
2678	1384489980	1384491060
2698	1384856520	1384856640
2868	1389381600	1389383700
2868	1389398700	1389400200
2884	1389665100	1389666300
3099	1393288800	1393289400
2235	1376731800	1376733000
3605	1400276700	1400277300
4184	1409986500	1409988000
4408	1413898200	1413899100
4418	1414083900	1414086000
4434	1414350300	1414351200
4452	1414674300	1414674780
4471	1414986600	1414987200
4483	1415223000	1415223900
4493	1415380500	1415381400
4503	1415558100	1415559000
4513	1415742300	1415743800

Continued on next page...

A2 Run	Unixtime Start	Unixtime Stop
4541	1416207000	1416212100
4814	1420978200	1420978800
4579	1416905100	1416910500
4621	1417677000	1417678200
4631	1417836000	1417837500
4763	1420097100	1420098300
4774	1420293300	1420294200
4775	1420317600	1420318200
4814	1420978200	1420978800
4817	1421024400	1421025300
4872	1421713200	1421718600
4873	1421718000	1421725800
4873	1421733300	1421733900
4876	1421783400	1421794200
5801	1435623000	1435623600
5915	1437601200	1437602700
6086	1440581700	1440582480
6396	1445934900	1445935500
6397	1445960400	1445961000
6408	1446165600	1446166200
6433	1446607800	1446608640
6445	1446784200	1446784800
8143	1477511700	1477512600
8241	1479231900	1479232500
6513	1449280500	1449281100
6531	1449610200	1449612000
6584	1450536000	1450537200
6669	1452008100	1452009000
6679	1452197700	1452198600
6680	1452213600	1452214200
6684	1452282000	1452282600
6685	1452298200	1452298800
6690	1452385500	1452386400
6696	1452494100	1452495000
6709	1452715200	1452716100
6724	1452972300	1452973440

Continued on next page...

A2 Run	Unixtime Start	Unixtime Stop
6747	1453408500	1453409400
6776	1453930200	1453931400
6916	1456200900	1456201800
6927	1456392600	1456393800
6962	1456997400	1456999200
7474	1466501400	1466503200

Table B.4: The bad livetime periods identified for A3.

A3 Bad Livetime Periods		
A3 Run	Unixtime Start	Unixtime Stop
1584	1381008600	1381010400
1670	1382476200	1382477400
1682	1382687400	1382688600
1684	1382712600	1382713800
1698	1382972700	1382973300
1739	1383688800	1383691500
1761	1384060200	1384060800
1770	1384208700	1384209900
1786	1384486200	1384492800
1980	1389399600	1389400800
2001	1389744000	1389747600
2025	1390176600	1390182000
2079	1391027700	1391028900
2235	1393652400	1393660800
2328	1394846400	1394856000
2363	1395437400	1395438600
2526	1397856300	1397857800
3533	1390176600	1390182000
3216	1409954100	1409956200
3217	1409986800	1409988600
3332	1412026200	1412028000
3347	1412284920	1412287020

Continued on next page...

A3 Run	Unixtime Start	Unixtime Stop
3362	1412544120	1412546400
3377	1412803620	1412805780
3439	1413897900	1413899100
3440	1413914400	1413922200
3449	1414083600	1414086300
3478	1414674000	1414675500
3520	1415380500	1415381400
3524	1415460600	1415461500
3540	1415742000	1415744100
3568	1416207300	1416209700
3579	1416457800	1416459000
3605	1416909600	1416910680
3608	1416951000	1416952500
3647	1417676400	1417679400
3651	1417742400	1417743600
3656	1417836600	1417839300
3800	1420317000	1420318200
3810	1420493700	1420494600
3811	1420513200	1420515000
3816	1420598700	1420600500
3830	1420857900	1420859700
3840	1421019000	1421020200
3863	1421101800	1421103600
3910	1421723400	1421723940
3912	1421750700	1421751720
3977	1421868600	1421881200
3978	1421881200	1421884680
3987	1422048900	1422049800
3995	1422307200	1422308100
4132	1423660800	1423661700
4200	1424819880	1424820720
4412	1428529500	1428531000
4445	1429094400	1429095600
4473	1429615800	1429617600
4474	1429616700	1429627500
4482	1429733400	1429734600

Continued on next page...

A3 Run	Unixtime Start	Unixtime Stop
4557	1431034500	1431036900
4693	1433365500	1433367900
4829	1435755600	1435756500
4832	1435791000	1435791600
4867	1436393700	1436395500
7658	1476740100	1476741300
7704	1477511400	1477518300
7709	1477604700	1477605900
7729	1477950300	1477951500
7802	1479231600	1479235800
6009	1448959200	1448960100
6046	1449610500	1449611400
6077	1450119900	1450120500
6098	1450536360	1450536720
6188	1452116100	1452116700
6193	1452196800	1452198600
6194	1452213600	1452214200
6198	1452282300	1452282900
6199	1452298500	1452299100
6203	1452385800	1452386400
6206	1452457800	1452458700
6208	1452494100	1452494700
6212	1452544980	1452545580
6213	1452561120	1452561480
6219	1452637020	1452637260
6223	1452715320	1452715680
6239	1452972660	1452973020
6259	1453325400	1453326300
6295	1453930500	1453931100
6328	1454535000	1454536200
6349	1454911200	1454911800
6350	1454911200	1454912100
6397	1455746400	1455747300
6433	1456374300	1456374900
6501	1457559300	1457560500
6618	1460843100	1460844600

Continued on next page...

A3 Run	Unixtime Start	Unixtime Stop
7052	1467927840	1467929640
7458	1473371280	1473372180
7562	1475186100	1475187000
7584	1475530500	1475531700
7625	1476221400	1476222600

B.3 Other Calibration Pulser Reconstruction Distributions

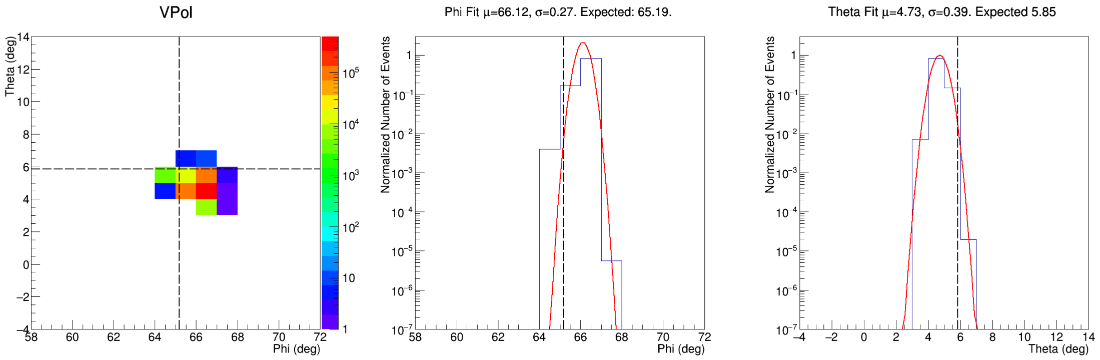


Figure B.1: The spatial distribution of all tagged calibration pulser events reconstructing in the direction of CP6 in A2 config 1, along with their projections along the θ and ϕ axes.

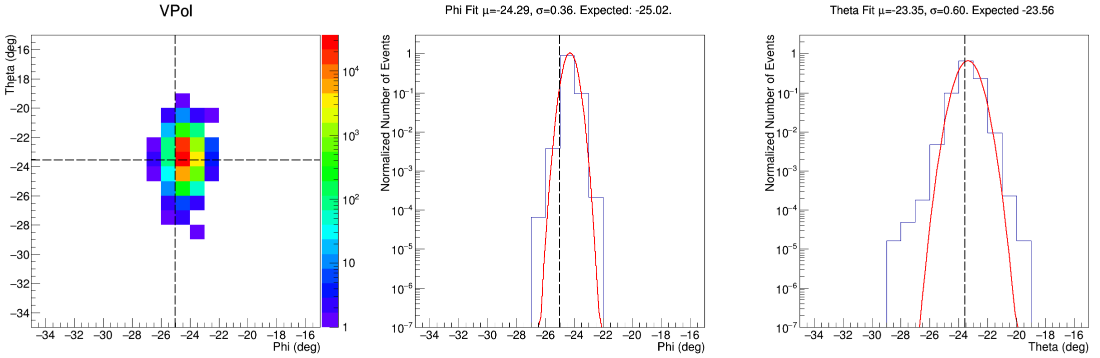


Figure B.2: The spatial distribution of all tagged calibration pulser events reconstructing in the direction of CP6 in A2 config 2, along with their projections along the θ and ϕ axes.

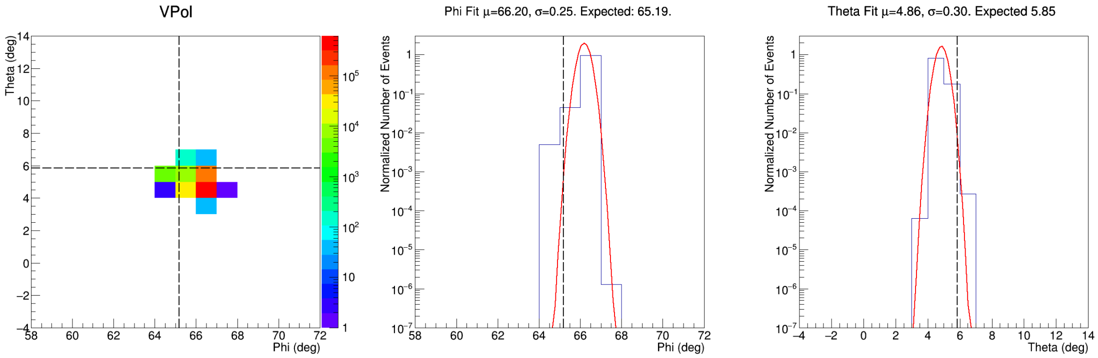


Figure B.3: The spatial distribution of all tagged calibration pulser events reconstructing in the direction of CP6 in A2 config 3, along with their projections along the θ and ϕ axes.

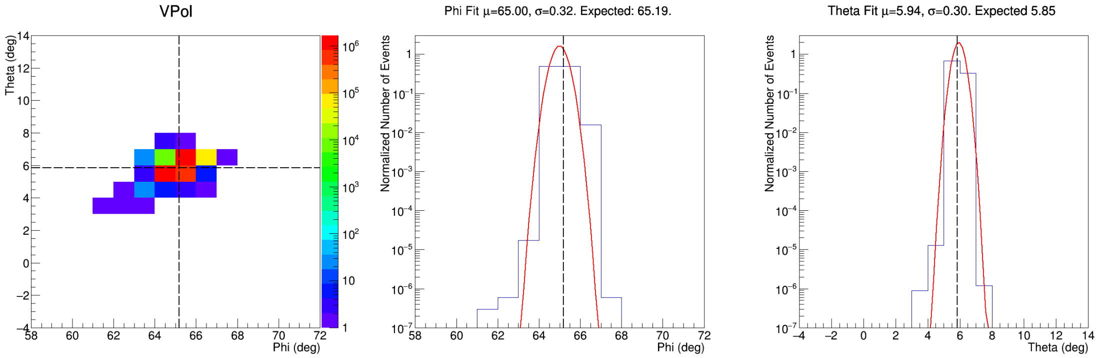


Figure B.4: The spatial distribution of all tagged calibration pulser events reconstructing in the direction of CP6 in A2 config 4, along with their projections along the θ and ϕ axes.

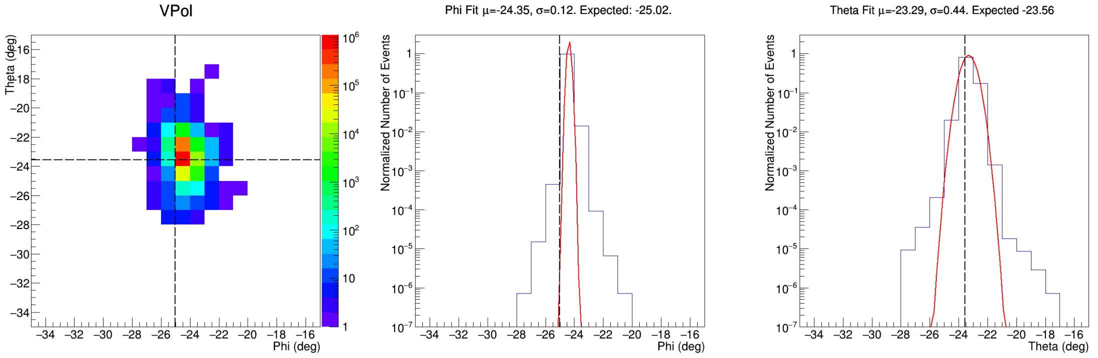


Figure B.5: The spatial distribution of all tagged calibration pulser events reconstructing in the direction of CP5 in A2 config 5, along with their projections along the θ and ϕ axes.

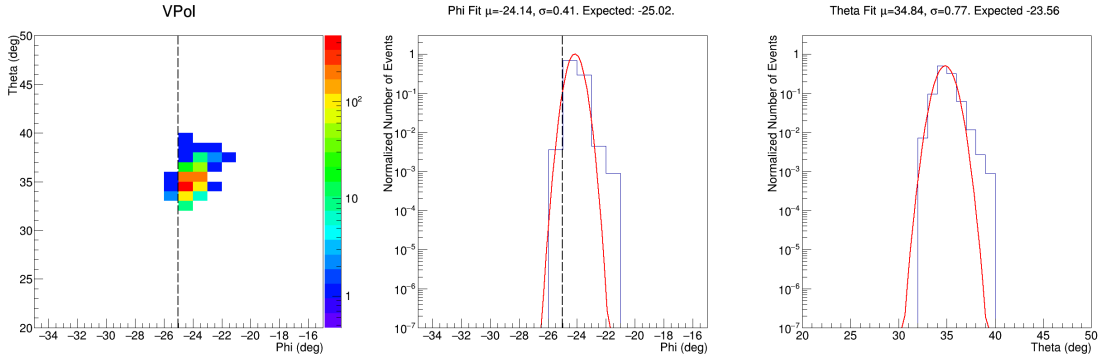


Figure B.6: The spatial distribution of all tagged calibration pulser events reconstructing in the direction of CP5 “Mirror” in A2 config 5, along with their projections along the θ and ϕ axes.

B.4 How to Run the 10% Analysis

B.4.1 Preparation Phase

There is an initial “data prep” phase before the meat of the analysis can begin. In this phase, we make a list of all data runs we want to analyze and identify their associated pedestal files. Then, we generate run summaries, run spectral baselines, and identify CW in events.

Making Run Lists and Pairing Pedestal Files

First, we must make a list of the data files we wish to analyze. This is done with the scripts in step0-make_run_list, and is particularly controlled by the bash file makeA23FileList.sh.

After making a list of data files, we must also generate a list of pedestal files, and pair each data file with its associated pedestal file. Because the pedestal files can vary at the percent level across the months of data taking we are examining, when possible, we utilize a pedestal file from nearby in time to when a run was recorded. The list of pedestal files is generated by the script makeA23PedFileList.sh in the step1-make_ped_pairs folder.

The pairing of a data file with its associated pedestal file is accomplished with the script makeA23DataPedPairs.sh. For this script to function, it must be in the presence of the tools.sh file, which contains auxiliary functions to identify the nearest pedestal function based on the run number and available list of pedestal files.

This process of making a run list, a pedestal list, and pairing runs and pedestals, generally only needs to be completed once at the data warehouse at the University of Wisconsin-Madison. It can then be copied over to other computing environments.

Run Summaries and Baseline Generation

In the run summaries, we essentially store a run average of the RMS of various types of events (software, RF, etc.). We also compute a run average FFT of all the RF triggered events in the run (but no software triggers or cal pulsers or events with < 500 samples). This will be used in our CW identification. The C++ program v2_analysis_run_summary, is the program which generates the output. It is steered by the step2-make_run_summary scripts.

Identify CW

This code decides if an event is CW contaminated. It constructs two metrics for contamination. First, it compares the FFT to the baseline (computed above) and decides if there is an excess of low frequency power as described in Sec. 4.8.1. It also applies Carl’s phase variance technique as described in Sec.4.8.1. The C++ program v2_analysis_CWID, is the program which does the identification process. It is steered by the step2c-check_cw scripts.

B.4.2 Analysis Phase

This is the meat of the analysis, and requires the most time computationally. In this phase, we: thermal filter the events with the wavefront-RMS filter as described in Sec.4.6, perform reconstruction on the events as described in Sec. 4.7, and merge the outputs of the filtering and reconstruction stages into single analysis files. We run the filter and two radii reconstructions as separate jobs, and join them at the end, in order to enable as much parallelization as possible in the processing.

Wavefront RMS Filter

In this stage, we run our thermal noise filter on ever event. It saves if an event passes or fails this filter criterion for a variety of combinations of the cut threshold values. The C++ program v2_analysis_filter is the program which executes the filtering. It is steered by the step3-filter scripts.

Reconstruction

In this stage, we perform interferometric reconstruction on every event. This one stage occupies most of the run time required to analyze the 10% data set. We save the location of the reconstruction peak (theta, phi), the amplitude of the peak, and a few other measurements that currently go unused (like the RMS of the map). Currently we perform reconstructions with two radial hypotheses: 41m and 300m. And so the output files will be tagged with these values. The C++ program v2_analysis_reco is the program which executes the filtering. It is steered by the step4-reco scripts.

Merging Filter and Reconstruction Output

To reduce the number of files that need to be read in at later stages, the outputs of the filter (one file) and reconstruction (two files, 41m and 300m) stages are merged into a single ”joined” file. The C++ program v2_join_filter_41_300.cxx is the program which merges the outputs. It is steered by the step5-join_reco scripts.

B.4.3 Post-Processing Phase

In this phase we do some final post processing on the outputs of the analysis phase. We: save out the final values needed for cuts (SNR and correlation map peaks) and can make plots.

Saving Interesting Values

In this stage, we loop over all events, and save out the SNR and correlation map peaks to a final file which we can loop over quickly to make final cut and efficiency plots. An important note at this stage is that if a event passed the thermal filter and reconstructs inside the ice, but is tagged as being contaminated with CW, then that event is filtered of its CW contamination using the Geometric Filter adapted from ANITA. This particular phase was placed here to allow for easy adjustment of the filtering technique etc. and to avoid re-doing the interferometry over again for all events. This way we only need to fix events which pass the thermal filter. After CW filtering, they are re-analyzed by the thermal noise filter. The C++ program v2_save_vals.cxx is the program which saves these values. It is steered by the step6-save_vals_for_cuts scripts.

Making Plots

Finally, we can make plots in the final cut parameter space of SNR vs Correlation Value. This runs over the output of the ”saving interesting values” phase. The C++ program v2_final_plots.cxx is the program which does this final plotting.

Appendix C

Recovering the Full Geometry Calibration of A2 and A3

This appendix will go over how we will recover the exact geometric and time delay calibrations utilized by Thomas Meures in his 10 month study of ARA stations 2 and 3. This more formally compiles results originally presented on an Analysis Focus call on Wed, Jan 9 2019 (see DocDB 1790).

C.1 Introduction to AraRoot

To be useful for analysis, the `AraRoot` analysis software needed to be updated. `AraRoot` was kept on the SVN repository at the University of Wisconsin-Madison at `http://svnmsn.icecube.wisc.edu/ara/Software/AraRoot`. The version recommended for analyses when these updates were issued was the `trunk` version, which was on revision 2908. Let's call this `AraRoot-2908`. Please note that since these updates were made, `AraRoot` has been migrated to a GitHub repository.

C.1.1 Where Geometry Information is Stored in AraRoot

`AraRoot` stores geometry information about the station in three member classes of the `AraEvent` class. `AraAntennaInfo` handles the measurement antenna information, `AraStationInfo` handles station information, and `AraCalAntennaInfo` handles the cal-pulser antenna information.

C.1.2 AraStationInfo

The `AraEvent/AraStationInfo` class handles information about the location of the ARA station broadly, namely the location of the ARA station in global coordinates, the Easting and Northing, etc. This class draws its information from the `AraArrayCoords.sqlite` file found in `AraRoot/AraEvent/calib` file.

AraAntennaInfo

The `AraEvent/AraAntennainfo` class handles information about the locations of the measurement antennas within an ARA station. It can report on things like the antenna polarization, and most importantly for this documentation, the coordinates (`X,Y,Z`) of the antennas as well as the cable delay for a given antenna. This class draws its information from the `AntennaInfo.sqlite` file found in `AraRoot/AraEvent/calib`.

The `AntennaInfo.sqlite` is created by a user from scripts in `AraRoot`. First, a user enters the antenna coordinates in the `AraRoot/dbScripts/makeAntInfoATRI.sql` file. In this file, each station is given a SQL table. And each antenna in a station is added to the respective table for that station, including information about the polarization, location, and cable delays. The creation of these tables and addition of antennas to the tables are really just SQL commands.

After the `makeAntInfoATRI.sql` file has been populated with information, the file is compiled into the SQLite table by running the command `sqlite3 AntennaInfo.sqlite < makeAntInfoATRI.sql`.

AraCalAntennaInfo

The `AraEvent/AraCalAntennainfo` class handles information about the locations of the cal-pulser antennas within an ARA station. Like `AraAntennainfo`, it can report on the coordinates (`X,Y,Z`) of the antennas. This class draws its information from the `CalPulserInfo.sqlite` file found in `AraRoot/AraEvent/calib`. Also like `AraAntennainfo`, it is created by a user from a source file, specifically `makeCalPulserInfo.sql` in the `AraRoot/dbScripts/` directory.

C.2 Updating The SQLite Information

It has been pointed out by Dave Seckel during in studies of the ARA IceCube deep pulser that over time we have failed to keep the `AntennaInfo.sqlite` file fully updated. He produced a summary of these inaccuracies at DocDB 1508. `CalPulserInfo.sqlite` is likewise out of date. What we want to do in this document is go from the results currently recorded in `AraRoot-2908`, and copied into Tab. C.1, Tab. C.2, Tab. C.3, Tab. C.4 in App. C.4.1 and recover the correct answer. This is challenging because the corrections to the geometry stored in `AraRoot-2908` are scattered throughout a script used by Thomas Meures and distributed to other analysts informally. The two main sources for the corrections come from two files, `mainAnalysis.cxx` and `calibrationTools.cxx`. They have the same measurement antenna corrections, while only the latter corrects the cal-pulser antennas.

The corrections Thomas makes in his script can be broken down into two categories: (1) corrections stored in auxiliary files and (2) corrections made “inline” in the analysis script. The corrections stored in files are stored in two files called `geometryResultsARA2E.txt` and `geometryResultsARA3E.txt` files in Thomas' analysis directory: `/home/meures/analysis/analysisCode/` and are attached to this note in code Listings C.5 and C.6 in App. C.5.

Lines 202-221 of `mainAnalysis.cxx` tell us how to understand the contents of these files (this code is reproduced in Listing C.1 below). It tells us that the files contain four x-coordinate corrections, four y-coordinate corrections, four z-coordinate corrections, 4 delay corrections, and then corrections to the two cal pulser locations.

Listing C.1: The lines of mainAnalysis.cxx (which are identical in to those in calibrationTools.cxx) describing the contents of geometryResultsARA files.

```
//here the station correction file is read. the rows are the different strings
//the 4 coloms are the three coordinate-corrections and then the delay-correction
//The structure of the file is:
//  4 X-corr
//  4 Y-corr
//  4 Z-corr
//  4 delay-corr
//  5 calpulser corrections:
//    NON-reference(So far this is always D5):
//      Z-corr
//      X-corr
//      Y-corr
//    REFERENCE-pulser:
//      Z-corr
//      distance-corr ( to be applied as: (1 + distance-corr)*x-coord. or y-coord.  )
//  1 light speed correction
//  4 slack corrections.
//  The last five are not valid!
//
```

The antenna position corrections and delay corrections are loaded into into two dimensional array titled `posDelayArray`. This array has four entries for each string; for each string, it stores the following values in this order: the x-correction, y-correction, z-correction, and the delay correction. So `posDelayArray[0][0]` stores the x-coordinate correction for string 0, where `posDelayArray[2][3]` stores the delay correction for string 2, and so on. Five of the entries are the corrections to the cal-pulser antenna locations. Cal-pulser D5 has all three (X,Y,Z) values corrected, while cal-pulser D6 only has its depth and absolute distance ($R = \sqrt{X^2 + Y^2}$) corrected. In `calibrationTools.cxx`, these are loaded into the array `pulserCorr` in that order.

C.2.1 Antenna Position Corrections

Measurement Antennas

The corrections to the measurement antenna positions are executed in lines 257-282 of `mainAnalysis.cxx` and are given below in Listing C.2. Note that every antenna is corrected by `posDelayArray`, and two antennas are additionally corrected in an ad-hoc fashion. A2 channel 0 is adjusted by 1.68m (line 270) and A3 channel 10 is adjusted by 2.01m (line 2.01).

The corrections to the measurement antenna positions are executed in lines 255-282 of `mainAnalysis.cxx` and are given below in Listing C.2. Note that every antenna is corrected by `posDelayArray`, and two antennas are additionally corrected in an ad-hoc fashion. A2 channel 0 is adjusted by 1.68m (line 270) and A3 channel 10 is adjusted by 2.01m (line 2.01).

I

Listing C.2: The lines of mainAnalysis.cxx (which are identical in effect to those in calibrationTools.cxx) correcting the position of the measurement antennas.

```
//Here station coordinates are read. the geometry correction is included as well as
//two specific reading errors, which were not corrected in araroot.
//also is the station center set to 180m under the ice.

    AraGeomTool *geom = AraGeomTool::Instance();
    std::vector<double> antl;
    std::vector<std::vector<double> > ant_loc;
    Double_t *antloc=0;

    for(int a=0;a<16;a++){
        antloc = geom->getStationInfo(rawAtriEvPtr->stationId)
                    ->getAntennaInfo(a)
                    ->getLocationXYZ();
        if(stationId==2 && a==0) antloc[2] = antloc[2] + 1.68;
        if(stationId==3 && a==10) antloc[2] = antloc[2] + 2.01;

        antl.push_back(antloc[0] + posDelayArray[a%4][0]);
        antl.push_back(antloc[1] + posDelayArray[a%4][1]);
        if((a/4)%2==1)antl.push_back(antloc[2]+180.0
                                + posDelayArray[a%4][2]
                                + slackArray[a%4]);
        else antl.push_back(antloc[2]+180.0 + posDelayArray[a%4][2]);

        cerr << "antDepth[1]["<<a<<"] = " << antl[2] << endl;

        ant_loc.push_back(antl);
        antl.clear();
    }
```

We will point out two additional things that are no longer necessary:

1. The correction by 180m.

2. The contents of `slackArray` is always zero, so we will not carry that correction forward.

We summarize the *total* corrections, including those from the files and those made in-line, for all measurement antennas for A2 in Tab. C.5 and for A3 in Tab. C.6, both in App. C.6.1. It is worth noting that in these correction tables, one string always has *no* geometry correction; that is because in Thomas' procedure, the position of one string was always taken as fixed in order to execute the fit. Said a different way, Thomas' chose the location of a single string to be (0,0,0). In A2, that is string 0 (channels 0, 4, 8, 12) and in A3, that is string 2 (channels 2, 6, 10, 14).

To demonstrate a correction, consider in A2 channel 2. We can read from Tab. C.1 that the nominal location is `[-2.68200, 8.68450, -170.38620]`. We can lookup the position correction in Tab. C.5 as `[0.10072, 0.69365, -1.20298]`. So, we find the corrected positions to be `[-2.58128, 9.378175, -171.58918]`. We repeat the exercise for every antenna, and provide the final result in Tab. C.7 and Tab. C.8 in App. C.6.1.

Cal-Pulser Antennas

The corrections to the cal-pulser antenna positions are executed in lines 136-165 of `calibrationTools.cxx` and are given below in Listing C.3. Note that the (X,Y,Z) of D5 are corrected outright. Meanwhile, the correction to D6 is only in the Z-coordinate, and the radial distance ($R = \sqrt{X^2 + Y^2}$) to the pulser. This was done, as we understand it, because there were not sufficient free parameters to do otherwise.

Listing C.3: The lines of calibrationTools.cxx correcting the position of the cal-pulser antennas.

```
cout<<"Number of of calpulsers for station "<<rawEvPtr->stationId<<": "
    <<geom->getStationInfo(rawEvPtr->stationId)->getNumCalAnts()<<endl;

for(int c=0; c<geom->getStationInfo(rawEvPtr->stationId)->getNumCalAnts(); c++){
    antloc = geom->getStationInfo(rawEvPtr->stationId)
                ->getCalAntennaInfo(c)->getLocationXYZ();
    string locName(&geom->getStationInfo(rawEvPtr->stationId)
                        ->getCalAntennaInfo(c)->locationName[0]);
    cout<<"locName: "<<locName<<endl;
    if( locName[locName.length()-1]  == '5' ){
        antl.push_back( antloc[0] + pulserCorr[1]  );
        antl.push_back( antloc[1] + pulserCorr[2]  );
        antl.push_back( antloc[2] + stationCenterDepth + pulserCorr[0]  );
    } else if ( locName[locName.length()-1] == '6'){
        cout<<"D6 xOrg: "<<antloc[0]
            <<" yOrg: "<<antloc[1] <<" zOrg: "<<antloc[2]+stationCenterDepth
            <<" pulserCorr(xy): "<<pulserCorr[4]
            <<" pulserCorrZ: "<<pulserCorr[3]<<endl;
        antl.push_back( (1.+pulserCorr[4]) * antloc[0]  );
        antl.push_back( (1.+pulserCorr[4]) * antloc[1]  );
```

```
            antl.push_back( antloc[2] + stationCenterDepth + pulserCorr[3] );
        } else {
            cerr<<"Pulser name undefined\n";
            //return -1;
        }
        printf("Corrected pulser %d X: %f Y: %f Z: %f\n"
                               , c, antl[0], antl[1], antl[2]);
        pul_loc.push_back(antl);
        antl.clear();
    }
```

We summarize the corrections to the positions of the cal-pulser antennas for A2 in Tab. C.9 and for A3 in Tab. C.10 of App. C.6.2. The final results are in Tab. C.11 and Tab. C.12 in App. C.6.2.

C.2.2 Antenna Delay Corrections

The corrections to the measurement antenna delays are executed in lines 304-339 of `mainAnalysis.cxx` and are shown in Listing C.4 below.

Listing C.4: The lines of mainAnalysis.cxx correcting the delay of the measurement antennas.

```
{//Loop through all 16 channels and write the data to TGraphs

    //Set up the delays properly(some are not taken care of in the simulation):
    addDelay = 0.0;
    delay= geom->getStationInfo(calEvPtr->stationId)->getCableDelay(a);
    if(a/4==0){addDelay+=(4.0 + posDelayArray[a%4][3]);}
    if(a/4==1){addDelay+=(12.0 + posDelayArray[a%4][3]);}
    if(a/4==2){addDelay+=(0.0 + posDelayArray[a%4][3]);}
    if(a/4==3){addDelay+=(8.0 + posDelayArray[a%4][3]);}

    //This gets the graph for the specified RF channel:
    gr_s1[a] = calEvPtr->getGraphFromRFChan(a);
    //We need to transcribe the Graphs to another array gr_s2, to adjust a few things:
    gr_s2[a] = new TGraph();

    if(calEvPtr->stationId==3 && (a==0||a==4||a==8)){invertGraph(gr_s1[a]);}
    average[a]=0;
    //*** The following is to avoid reading corrupted waveforms. ***//
    //*** I encountered only a few of them, so maybe this is not ***//
    //*** really neccessary anymore. ******************************//
    if(gr_s1[a]->GetN()<100 ){
        cerr<< "event: " << eve
        << " wow wow wow "
        << a << ", points: "
        << gr_s1[a]->GetN() << endl;cutWaveAlert=1;continue;}
    //Cut of the first 20 ns in new TGraph, because they are often corrupted.
    int pc = 0;
    for(int p=0;p<gr_s1[a]->GetN();p++){
        gr_s1[a]->GetPoint(p, times, volts);
        if(times>20.0 && times - delay - addDelay<9999.0){
```

```
            gr_s2[a]->SetPoint(pc, times - delay - addDelay, volts);
            average[a]+=volts;
            pc++;
        }
    }
}
```

There are several delays to keep track of in this case:

1. `delay`: this variable (line 309) is the delay originally stored in the SQLite table.
2. `addDelay`: this variable is composed of
   (a) Delays loaded from the geometryResults files and stored in from `posDelayArray`
   (b) Delays added to compensate for the cables running through the antennas themselves; these are the 0ns, 4ns, 8ns, and 12 ns in lines 310-313. Recall antennas are deployed down the hole, top to bottom, in the order of TH, TV, BH, BV. This mean the signals in BV must travel through patch cables in the BH, TV, and TH antennas, or 12ns of patch cable. The BH antenna must only travel through the TV and TH antennas, and requires 8 ns of delay. The TV only travels through TH, or 4 ns. And TH travels through no other antenna, and requires no delay. Note that antennas are loaded into AraRoot in the order TV, BV, TH, BH. So we expect the delays to show up in AraRoot in the order of 4ns, 12ns, 0ns, 8ns. And that is true in the code.

The value recorded on the x-axis of an adjusted waveform is found at line 331: `NewTime = OldTime - delay - addDelay`, where `NewTime` is what I am calling the fully calibrated time, `OldTime` is the time drawn from the original waveform TGraph, and `delay` and `addDelay` are described above.

In `AraRoot-2908`, the `delay` part of the timing correction is applied during the waveform calibration procedure. So we do not need to subtract `delay` off again. Rather, we need to update the SQLite file to include the delays in `posDelayArray` and the cable throughput delays. We compile the list of corrections to `AraRoot-2908` in Tab. C.13 and Tab. C.14 in App. C.7.1.

To demonstrate a correction, consider in A2 channel 2. We can read from Tab. C.1 that the nominal delay is 115.44ns. We can lookup the delay correction in Tab. C.5 as 10.1066. So, we find the corrected total delay to be 125.5466. We repeat the exercise for every antenna, and provide the final result in Tab. C.15 and Tab. C.16 in App. C.7.2.

C.3 Conclusion

We have shown in this document how to compute the current best estimates for the antenna positions (both measurement and cal-pulser) and delays for A2 and A3 using what is known in `AraRoot-2908` and in analysis scripts provided by Thomas. These are the same results as were obtained by Dave Seckel in June 2017 for the measurement antenna positions.

We have formally made these changes in `makeAntInfoATRI.sql` and pushed them to the SQLite file `AntennaInfo.sqlite` in revision 2909 of AraRoot. Changes to `CalPulserInfo.sqlite` are in revision 2913. All future analysts should proceed exclusively with these updated files.

C.4 2908 Antenna Positions and Delays

C.4.1 Antenna Positions and Delays

Antenna	X (m)	Y (m)	Z (m)	Delay (ns)
0	10.5874	2.3432	-171.927	15.44
1	4.8235	-10.384	-170.489	15.44
2	-2.682	8.6845	-170.386	115.44
3	-7.7232	-4.4671	-170.672	15.44
4	10.5873	2.3428	-189.502	81.41
5	4.8234	-10.3844	-189.542	81.41
6	-2.6821	8.6841	-189.439	181.41
7	-7.7234	-4.4674	-189.56	81.41
8	10.5874	2.3128	-167.492	7.24
9	4.8235	-10.384	-167.57	7.24
10	-2.682	8.6846	-167.265	107.24
11	-7.7232	-4.467	-167.715	7.24
12	10.5873	2.3429	-186.546	73.2
13	4.8234	-10.3844	-186.257	73.2
14	-2.6821	8.6841	-186.319	173.2
15	-7.7232	-4.4674	-186.275	73.2

Table C.1: The `AraRoot-2908` positions and delays for A2 measurement antennas.

Antenna	X (m)	Y (m)	Z (m)	Delay (ns)
0	4.4113	-9.3933	-173.394	15.44
1	10.6919	3.5089	-173.973	115.44
2	-2.0103	9.4124	-174.105	15.44
3	-8.099	-3.7098	-173.548	15.44
4	4.4111	-9.3936	-192.447	81.41
5	10.6916	3.5086	-192.698	181.41
6	-2.0102	9.412	-192.666	81.41
7	-8.0992	-3.7101	-192.601	81.41
8	4.4114	-9.3932	-170.273	7.24
9	10.6919	3.509	-171.017	107.24
10	-2.01	9.4124	-172.627	7.24
11	-8.0989	-3.7097	-170.591	7.24
12	4.4111	-9.3935	-189.195	73.2
13	10.6917	3.5087	-189.742	173.2
14	-2.0102	9.4121	-189.709	73.2
15	-8.0992	-3.71	-189.48	73.2

Table C.2: The `AraRoot-2908` positions and delays for A3 measurement antennas.

C.4.2 Pulser Antenna Positions

Cal-Pulser	X (m)	Y (m)	Z (m)
BH5 Horizontal	37.8746	-18.0508	-189.641
BH5 Vertical	37.8746	-18.0508	-192.926
BH6 Horizontal	17.7477	35.3359	-164.654
BH6 Vertical	17.7477	35.3359	-167.939

Table C.3: The `AraRoot-2908` positions for cal-pulsers in A2.

Cal-Pulser	X (m)	Y (m)	Z (m)
BH5 Horizontal	38.0057	-16.3989	-194.898
BH5 Vertical	38.0057	-16.3989	-198.183
BH6 Horizontal	17.9354	36.3856	-184.726
BH6 Vertical	17.9354	36.3856	-188.011

Table C.4: The `AraRoot-2908` positions for cal-pulsers in A3.

C.5 Auxiliary Files by Thomas

```
0.0  0.028169719032520312  0.1007192958885017
   −0.11790695789547029  0.0    −0.014120738328222979
   0.69365469216027864  0.40918517119628395  0.0
   0.14192507885017425  −1.2029843129384439  −4.7053124970964015
    0.0    0.42056236716608586  6.1066035001161492
   3.135874423577234    −4.6202589895470378
   −0.016119269379674783  0.28036008549941882
   −5.3502286411149811  0.0079167000000000005  0  0.0  0.0  0.0
   0.0
```

Listing C.5: The geometryResultsARA2E.txt file used by Thomas' mainAnalysis.cxx program

```
0.41211138297872346  −0.049042692978723372  0.0
   −0.047544037234042 56   −0.024736354361702125
   −0.10044319574468084  0.0  −0.22320156297872346
   0.78712646808510667  −2.2353778723404254  0.0
   1.3336945106382978   −0.2863143718085106  6.8085269361702112
   0.0  2.1214127187234042   4.9054484042553197
   0.22793788690425526  0.18414589138297871
   −3.9500453510638298  0.0089554300000000003  0  0.0  0.0  0.0
   0.0
```

Listing C.6: The geometryResultsARA3E.txt file used by Thomas' mainAnalysis.cxx program

C.6 Antenna Position Tables

C.6.1 Measurement Antennas

Measurement Antenna Position Corrections to AraRoot-2908

Antenna	X-Correction (m)	Y-Correction (m)	Z-Correction (m)
0	0	0	1.68
1	0.0281697	-0.0141207	0.141925
2	0.100719	0.693655	-1.20298
3	-0.117907	0.409185	-4.70531
4	0	0	0
5	0.0281697	-0.0141207	0.141925
6	0.100719	0.693655	-1.20298
7	-0.117907	0.409185	-4.70531
8	0	0	0
9	0.0281697	-0.0141207	0.141925
10	0.100719	0.693655	-1.20298
11	-0.117907	0.409185	-4.70531
12	0	0	0
13	0.0281697	-0.0141207	0.141925
14	0.100719	0.693655	-1.20298
15	-0.117907	0.409185	-4.70531

Table C.5: The total corrections for the positions of the A2 measurement antennas relative to those found in AraRoot-2908 (includes geometryResultsARA2E.txt corrections and inline corrections.)

Antenna	X-Correction (m)	Y-Correction (m)	Z-Correction (m)
0	0.412111	-0.0247364	0.787126
1	-0.0490427	-0.100443	-2.23538
2	0	0	0
3	-0.047544	-0.223202	1.33369
4	0.412111	-0.0247364	0.787126
5	-0.0490427	-0.100443	-2.23538
6	0	0	0
7	-0.047544	-0.223202	1.33369
8	0.412111	-0.0247364	0.787126
9	-0.0490427	-0.100443	-2.23538
10	0	0	2.01
11	-0.047544	-0.223202	1.33369
12	0.412111	-0.0247364	0.787126
13	-0.0490427	-0.100443	-2.23538
14	0	0	0
15	-0.047544	-0.223202	1.33369

Table C.6: The total corrections for the positions of the A3 measurement antennas AraRoot-2908 (includes geometryResultsARA3E.txt corrections and inline corrections.)

Measurement Antenna Final Corrected Positions

Antenna	X (m)	Y (m)	Z (m)
0	10.5874	2.3432	-170.247
1	4.85167	-10.3981	-170.347
2	-2.58128	9.37815	-171.589
3	-7.84111	-4.05791	-175.377
4	10.5873	2.3428	-189.502
5	4.85157	-10.3985	-189.4
6	-2.58138	9.37775	-190.642
7	-7.84131	-4.05821	-194.266
8	10.5874	2.3128	-167.492
9	4.85167	-10.3981	-167.428
10	-2.58128	9.37825	-168.468
11	-7.84111	-4.05781	-172.42
12	10.5873	2.3429	-186.546
13	4.85157	-10.3985	-186.115
14	-2.58138	9.37775	-187.522
15	-7.84111	-4.05821	-190.981

Table C.7: The final, correct positions for the A2 measurement antennas.

Antenna	X (m)	Y (m)	Z (m)
0	4.82341	-9.41804	-172.607
1	10.6429	3.40846	-176.209
2	-2.0103	9.4124	-174.105
3	-8.14654	-3.933	-172.214
4	4.82321	-9.41834	-191.66
5	10.6426	3.40816	-194.933
6	-2.0102	9.412	-192.666
7	-8.14674	-3.9333	-191.267
8	4.82351	-9.41794	-169.486
9	10.6429	3.40856	-173.252
10	-2.01	9.4124	-170.617
11	-8.14644	-3.9329	-169.257
12	4.82321	-9.41824	-188.408
13	10.6427	3.40826	-191.977
14	-2.0102	9.4121	-189.709
15	-8.14674	-3.9332	-188.146

Table C.8: The final, correct positions for the A3 measurement antennas.

C.6.2 Cal-Pulser Antennas

Cal-Pulser Antenna Position Corrections to AraRoot-2908

Cal-Pulser	X-Correction (m)	Y-Correction (m)	Z-Correction (m)
BH5 Horizontal	-0.0161193	0.28036	-4.62026
BH5 Vertical	-0.0161193	0.28036	-4.62026
BH6 Horizontal	0.140503	0.279744	-5.35023
BH6 Vertical	0.140503	0.279744	-5.35023

Table C.9: The total corrections for the positions of the A2 cal-pulser antennas relative to those found in AraRoot-2908.

Cal-Pulser	X-Correction (m)	Y-Correction (m)	Z-Correction (m)
BH5 Horizontal	0.227938	0.184146	4.90545
BH5 Vertical	0.227938	0.184146	4.90545
BH6 Horizontal	0.160619	0.325849	-3.95005
BH6 Vertical	0.160619	0.325849	-3.95005

Table C.10: The total corrections for the positions of the A3 cal-pulser antennas relative to those found in AraRoot-2908.

Cal-Pulser Antenna Final Corrected Positions

Antenna	X (m)	Y (m)	Z (m)
BH5 Horizontal	37.8585	-17.7704	-194.261
BH5 Vertical	37.8585	-17.7704	-197.546
BH6 Horizontal	17.8882	35.6156	-170.004
BH6 Vertical	17.8882	35.6156	-173.289

Table C.11: The final, correct positions for the A2 cal-pulser antennas.

Antenna	X (m)	Y (m)	Z (m)
BH5 Horizontal	38.2336	-16.2148	-189.992
BH5 Vertical	38.2336	-16.2148	-193.277
BH6 Horizontal	18.096	36.7114	-188.676
BH6 Vertical	18.096	36.7114	-191.961

Table C.12: The final, correct positions for the A3 cal-pulser antennas.

C.7 Antenna Delay Tables

C.7.1 Measurement Antenna Delay Corrections to AraRoot-2908

Antenna	Delay Correction (ns)
0	4
1	4.42056
2	10.1066
3	7.13587
4	12
5	12.4206
6	18.1066
7	15.1359
8	0
9	0.420562
10	6.1066
11	3.13587
12	8
13	8.42056
14	14.1066
15	11.1359

Table C.13: The total corrections for the delays of the A2 measurement antennas relative to those found in AraRoot-2908 (includes geometryResultsARA2E.txt corrections and inline corrections.)

Antenna	Delay Correction (ns)
0	0.412111
1	-0.0490427
2	0
3	-0.047544
4	0.412111
5	-0.0490427
6	0
7	-0.047544
8	0.412111
9	-0.0490427
10	0
11	-0.047544
12	0.412111
13	-0.0490427
14	0
15	-0.047544

Table C.14: The total corrections for the delays of the A3 measurement antennas relative to those found in AraRoot-2908 (includes geometryResultsARA3E.txt corrections and inline corrections.)

C.7.2 Measurement Antenna Final Corrected Delays

Antenna	Delay (ns)
0	19.44
1	19.8606
2	125.547
3	22.5759
4	93.41
5	93.8306
6	199.517
7	96.5459
8	7.24
9	7.66056
10	113.347
11	10.3759
12	81.2
13	81.6206
14	187.307
15	84.3359

Table C.15: The final, correct delays for the A2 measurement antennas.

Antenna	Delay (ns)
0	19.1537
1	126.249
2	19.44
3	21.5614
4	93.1237
5	200.219
6	93.41
7	95.5314
8	6.95369
9	114.049
10	7.24
11	9.36141
12	80.9137
13	188.009
14	81.2
15	83.3214

Table C.16: The final, correct delays for the A3 measurement antennas.