

The Pennsylvania State University
The Graduate School
Eberly College of Science

**MEASUREMENT OF THE COSMIC RAY FLUX ABOVE 100 PeV
AT THE PIERRE AUGER OBSERVATORY**

A Dissertation in
Physics
by
Alan Coleman

© 2018 Alan Coleman

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

December 2018

The dissertation of Alan Coleman was reviewed and approved* by the following:

Miguel A. Mostafá
Professor of Physics and Astronomy & Astrophysics
Dissertation Advisor, Chair of Committee

Stephane Coutu
Professor of Physics and Astronomy & Astrophysics

Doug Cowen
Professor of Physics and Astronomy & Astrophysics

Abraham D. Falcone
Research Professor of Astronomy & Astrophysics
Outside Committee Member

Piera Luisa Ghia
Director of Research, Institut de Physique Nucléaire Orsay
Special Member

Nitin Samarth
Professor of Physics
Head of Physics

*Signatures are on file in the Graduate School.

Abstract

Cosmic rays are particles and nuclei that arrive at Earth and act as messengers, informing us of the nature of celestial objects and events throughout the universe. Some of the highest energy events, with over a Joule of energy, are more energetic than what can be made on Earth in modern particle accelerators. In the ultra-high energy range $\gtrsim 100$ PeV, the arrival rate is low enough (~ 1 per km^2 per year, or less) that there are still many outstanding questions concerning their origin and what types of particles they are.

Further, their rarity presents an experimental challenge as direct detection of cosmic rays above this energy would require decades to centuries of data collection. Instead, the particles cascades created by ultra-high energy cosmic rays interacting with air molecules high in the atmosphere, called *air showers*, can be detected using observatories that cover $\sim 10\text{-}1000 \text{ km}^2$. The Pierre Auger Observatory includes a number of methods to detect air showers and covers over 3000 km^2 in the Mendoza Province of Argentina. The Observatory includes a hybrid design by which air showers can be detected by fluorescence detectors while they are developing in the air as well as by an array of surface detectors on the ground.

This thesis includes an extensive update to the reconstruction methods used to estimate the trajectory and energy of a cosmic ray using a hexagonal array of water Cherenkov detectors with 750 m spacing. The update was motivated by the inclusion of a new set of particle triggers that were installed in the local stations that make up the array. These triggers were designed to be insensitive to muons which make up the primary background for individual stations. Thus, they increase the sensitivity of the array to lower energy parts of air showers and lower energy showers in general.

A major component of this work was the characterization of the array's abilities to detect and sample air showers with the new trigger algorithms. On the level of individual stations, the triggering efficiency and distribution of signals was studied. On the array-level, the efficiency with which the 750 m array detects showers was also calculated for two sets of cosmic ray masses using a dedicated set of air shower simulations based on hadronic interaction models.

The second component of this thesis was an improvement of Auger’s model of air shower development. Due to a lack of understanding of hadronic cross sections in the ultra-high energy regime, empirical models are used to characterize the temporal and spatial distribution of particles within the cascade. The distribution of signal as a function of distance from an air shower’s central axis was updated, benefiting from the $\lesssim 30\%$ more sampling of the shower front by new-triggered stations. This model is particularly important as it is used to find the expected signal at a fixed reference distance from the axis to estimate the shower’s size, a quantity highly correlated to energy. These size estimations were then corrected for a number of systematic biases to produce a more precise energy estimator. Finally, the energy estimators were cross-calibrated with the nearly calorimetric energy measurements made by the fluorescence detectors. This allowed for the surface detector to directly estimate energies accurate to within $\Delta E/E = 14\text{--}15\%$.

Motivated by the parameter space where the array can detect showers with full efficiency, two semi-joint data sets were chosen which included energies and zenith angles ($E > 10^{17}$ eV, $\theta < 40^\circ$) and ($E > 10^{17.3}$ eV, $\theta < 55^\circ$). With these data sets the cosmic ray flux was measured down to energies half an order of magnitude lower than what was previously possible by the surface array. As a result of the decreased energy threshold, an additional spectral feature, sometimes called the *second knee* or *iron knee*, was observed where spectral index of the otherwise approximately power law spectrum changes value. This feature, at $\simeq 100$ PeV as well as a second observed feature at $\simeq 5$ EeV, the *ankle*, are important as they are signatures of change in origin and/or types of arriving cosmic rays. The measurement of the flux agrees well with that of other experiments within the systematic errors.

Table of Contents

List of Figures	ix
List of Tables	xii
Acknowledgments	xiii
Chapter 1	
Introduction to Ultra-High Energy Cosmic Ray Physics	1
1.1 A Brief History of Cosmic Ray Physics	2
1.2 Sources of Cosmic Rays	5
1.2.1 Top Down Models	5
1.2.2 Bottom Up Models	6
1.3 Cosmic Ray Flux Spectrum	7
1.3.1 The Knee(s)	10
1.3.2 The Ankle and Suppression	11
Chapter 2	
Extensive Air Showers	15
2.1 Electromagnetic Heitler Model	15
2.1.1 Development of Electromagnetic-Initiated Showers	16
2.2 Hadronic Shower	17
2.2.1 Hadronic Maximum	18
2.2.2 Nuclear Mass Scaling	19
2.3 Air Shower Age and Universality	20
2.3.1 Air Shower Age	20
2.3.2 Particle Distributions in Air Showers	21
2.4 Methods For Studying Air Showers	22
2.4.1 Air Shower Simulations	23
2.4.2 Air Shower Detection Methods	24

Chapter 3	
The Pierre Auger Observatory	27
3.1 The Fluorescence Detectors	28
3.1.1 FD Telescopes	29
3.1.2 FD Event Reconstruction	30
3.2 The Surface Detectors	32
3.2.1 SD Calibration	33
3.2.2 SD Event Reconstruction	35
3.3 Selected Physics Results	37
3.3.1 Energy Spectrum	37
3.3.2 Cosmic Ray Mass Measurements	38
3.3.3 Large-Scale Arrival Directions	41
Chapter 4	
The Auger SD Triggers	43
4.1 Station-Level Triggers	43
4.1.1 The Old Triggers	44
4.1.2 The New Triggers	45
4.2 Array-Level Triggers	47
4.2.1 T3 Triggers	47
4.2.2 Physics Triggers	49
4.3 The New Triggers in the 750 m Array	50
4.3.1 Impact of the New Triggers on Individual Stations	51
4.3.2 Station Triggering Efficiency	53
4.3.3 Summary of the New Triggers	56
Chapter 5	
Event Reconstruction Using New Triggers	58
5.1 Station Selection with New Triggers	58
5.1.1 Selection Methodology	59
5.1.2 Update to the Selection Window	60
5.1.3 Additional Selections	63
5.2 The Air Shower Reconstruction	64
5.2.1 Geometric Reconstruction	66
5.2.2 Lateral Distribution Reconstruction	69
5.2.3 Update of the LDF Shape	72
5.2.3.1 The LDF Tuning Data Set	72
5.2.3.2 Parameterization of the LDF Slope	73
5.2.4 The Global Reconstruction	79

Chapter 6	
Air Shower Energy Assignment	80
6.1 The LDF Bias at 450 m	81
6.2 Correction for the Atmospheric Conditions	84
6.3 Constant Intensity Cut	87
6.3.1 The CIC Tuning Methodology	90
6.3.2 Results of the CIC Tuning	93
6.4 Energy Calibration Using FD Events	95
6.4.1 Selection of Golden Hybrid Events	95
6.4.2 The Calibration Fit	97
6.4.3 SD Energy Uncertainties	101
6.5 Conclusion	105
Chapter 7	
The Efficiency and Exposure of the 750 m Array	107
7.1 Triggering Efficiency of the 750 m Array	108
7.1.1 Efficiency Study Using T2 Triggers	109
7.1.2 Efficiency Study Using Hadronic Models	111
7.1.2.1 Conversion to the Auger Energy Scale	112
7.1.2.2 Efficiency Calculation in the FD Energy Scale	115
7.1.3 Parameterization of the T4 Efficiency	118
7.2 Array Exposure	121
7.3 Conclusion	123
Chapter 8	
Measurement of the Energy Spectrum Above 100 PeV	125
8.1 The Raw Spectrum	126
8.1.1 Zenith Dependence	127
8.2 The Unfolded Energy Spectrum	130
8.2.1 The Forward Folding Procedure	131
8.2.2 Impact of the Spectral Model on the Unfolding	133
8.3 Comparison with Previous Auger Measurements	136
8.4 Identification of the Spectral Features	142
8.5 Systematic Uncertainty of the Spectrum	143
8.6 Comparison with other Experiments	145
8.7 Conclusion	148
Chapter 9	
Summary	150
9.1 Inclusion of the New Triggers	150

9.2 Energy Spectrum Using New Triggers	152
9.3 Outlook	153
Appendix A	
Additional Analyses	155
A.1 Surface Detector Station Calibration	155
A.2 Top-Down Station Selection	159
A.3 The Time Variance Model	162
A.3.1 Update to the Time Variance Model	164
A.3.2 The Calculation of Geometrical Errors	168
A.4 The T4 Efficiency for Old Triggers	170
A.5 Energy Dependence of the CIC	172
Appendix B	
Additional Plots and Information	178
B.1 Weather Correction	178
B.2 Constant Intensity Cut	180
B.3 FD Energy Resolution	181
B.4 Monte Carlo Energy Scale Offset	183
B.5 Unfolding Correction Factors	184
B.6 Flux Measurements Using the 750 m Array	185
Appendix C	
List of Acronyms	186
Bibliography	188

List of Figures

1.1	Flux spectra of the HiRes and AGASA experiments	4
1.2	Hillas plot of potential cosmic ray sources	8
1.3	The cosmic ray flux spectrum above 1 GeV	9
1.4	Mass spectrum measured by KASCADE-Grande	11
1.5	Flux correction factor for the Dip Model	12
1.6	Example of the mixed composition above the ankle	14
2.1	Schematic of the Heitler model for EM and hadronic showers	17
2.2	Distribution of electrons in hadronic air shower	22
2.3	Agreement between models	24
3.1	Schematic of the Auger Observatory layout	28
3.2	Photo of an FD site	29
3.3	Schematic of the FD telescopes and the shower detector plane	30
3.4	Example of a hybrid event	31
3.5	Photo of an SD station	33
3.6	SD station calibration histograms	34
3.7	Example of an SD Event	36
3.8	The Auger energy spectrum	38
3.9	The Auger X_{\max} measurements	39
3.10	Combined fit of the energy spectrum and X_{\max} distributions	40
3.11	Arrival directions of Auger events above 8×10^{18} eV	41
4.1	Schematic of trigger levels for SD	44
4.2	Example traces of the four T2 station triggers	46
4.3	Examples of T3 configurations	48
4.4	Change in station multiplicity when including TOTd and MoPS	52
4.5	The distribution of measured station signals	53
4.6	Triggering efficiency of the T1/T2 algorithms	54
4.7	Overall T1/T2 triggering efficiency versus zenith angle	57
5.1	Schematic of the station selection process	60

5.2	The signal timing residuals compared to a plane front	61
5.3	The timing residual CDFs representing the selection windows	62
5.4	Example of lonely stations	64
5.5	Example of the digitized trace created by nearby lighting	65
5.6	Schematic of the Auger shower model	66
5.7	Difference in altitude of neighboring stations in the 750 m array	67
5.8	Example of the geometric fit of an SD event	68
5.9	Schematic of a shower front incident on an surface array	70
5.10	Example of the LDF fit of an SD event	72
5.11	Comparison of the measured lateral distribution and the LDF	76
5.12	Residuals of the lateral distribution fit as a function of zenith angle and shower size	77
5.13	Residuals of the lateral distribution fit for various shower sizes and trigger types	78
6.1	Bias of the LDF estimation at 450 m from the shower axis	82
6.2	Dependence of the LDF bias parameters on zenith angle	84
6.3	Local atmospheric pressure and density measurements	85
6.4	The number of observed showers in $\sin^2 \theta$ bins	89
6.5	The smearing function used in the CIC fit	92
6.6	The number of CIC-corrected showers as a function of $\sin^2 \theta$	94
6.7	Quality of the CIC fit for various shower sizes	95
6.8	Example of an event seen by four FD telescopes	96
6.9	Distribution of events used in the energy calibration	98
6.10	Results of the FD calibration	100
6.11	Systematic uncertainty from the calibration procedure	101
6.12	Systematic uncertainty of the SD's energy scale	102
6.13	The energy resolution of the SD	104
6.14	The total correction of the shower size	106
7.1	T4 trigger efficiency calculated via T2 efficiency	110
7.2	Results of the CIC prescription on Monte Carlo events	113
7.3	Calibration of the Monte Carlo events	114
7.4	Array efficiency calculated via CORSIKA air shower simulations	116
7.5	Comparison of the QGSJet and EPOS efficiency results	117
7.6	Array efficiency calculated via CORSIKA air shower simulations	120
7.7	Diagram of an exposure hexagon	122
7.8	Exposure of the 750 m array over four years	123
8.1	Raw number of counts as a function of energy	126

8.2	The raw flux spectrum	127
8.3	Raw number of counts as a function of energy for different zenith angle bins	128
8.4	T4 trigger efficiency calculated via T2 efficiency	131
8.5	The folded and unfolded spectra using new triggers	135
8.6	Comparison of the flux to the ICRC 2017 measurements	138
8.7	Comparison of the flux from ICRC 2017 using Code-1	139
8.8	Explanation of the offset in the flux from ICRC 2017	140
8.9	The ratio of the SD and FD energies with time	141
8.10	Empirical fit of the 750 m array	142
8.11	The systematic uncertainty on the flux	144
8.12	Comparison of the spectrum with other experiments up to the suppression	146
8.13	Comparison of the spectrum with other experiments near the second knee	147
A.1	Example of a station-calibration histogram	156
A.2	Example of an abnormal station calibration histogram	157
A.3	Comparison of the old and new station calibration methods	159
A.4	The individual station residuals used for the top-down selection	161
A.5	The combined station residuals used for the top-down selection	163
A.6	The agreement of the time variance model with the data	166
A.7	The variance of the signal start times versus signal and axial radius	167
A.8	Chi-square probability distributions for the geometric fit	169
A.9	Comparison of the T4 efficiency for with and without new triggers	171
A.10	Energy dependence of the CIC correction	174
A.11	Overall shift in the event energy when using an energy dependent CIC	176
A.12	Effects of the CIC’s energy dependence on the spectrum	177
B.1	Effects of atmospheric conditions on the event rate	179
B.2	The weather correction-model parameters	179
B.3	The intensity of air showers	180
B.4	Energy resolution of the FD	182
B.5	The offset between the MC and FD energy scales	183
B.6	The correction factors for the unfolding using three spectral models	184

List of Tables

4.1	The T1, T2, T3, and T4 trigger rates in the 750 m array	51
5.1	Updated values of the selection window by trigger type	63
5.2	Updated values for the LDF shape parameters	75
5.3	Typical uncertainties of the shower reconstruction	79
6.1	Results of the CIC using the 750 m array	93
6.2	Results of the new trigger energy calibration	101
7.1	Values for the CIC and energy calibration for Monte Carlo data . .	115
7.2	Parameterization of the T4 trigger efficiency for new triggers . . .	119
8.1	Comparison of the spectral models used in the unfolding	136
8.2	Fit of the spectral features for the new trigger data sets	143
8.3	Comparison of values describing the second knee	148
A.1	The limits for the top-down selection for each trigger type	162
A.2	Best fit values for the time variance model	168
A.3	Values for the CIC and energy calibration for Monte Carlo old-trigger data	170
A.4	Parameterization of the T4 trigger efficiency	172
B.1	The values of the spectral points for the new trigger data set	185

Acknowledgments

This thesis is a testament to the amount of opportunities and support I have been given throughout my academic career as much as to any of my own doing. This has never been lost on me and is something for which I am truly grateful.

Foremost, I owe a special thanks to Miguel Mostafá and Piera Luisa Ghia for being the best advisors I ever could have hoped for. You have both been simply fantastic mentors in your advice and insight, your careful editing of my papers, and your willingness to take the time to discuss ideas when I barged into your offices. Thank you for your direction to help me to navigate my way though this thesis and at times keep me on course when I tried to chase interesting side topics into the weeds. A particular thanks is due to Piera for doing most of this from Paris via over 1000 emails and many dozens of video calls as well as for her uncanny ability to totally demolish my manuscripts while still somehow making me feel like I did something right.

Secondly, I would like to thank the people of the Pierre Auger Collaboration. My experience working with the many interesting and insightful researchers across the world has been fantastic. They have impacted me significantly as a scientist but also have given me many cultural experiences that have greatly shaped my world view. In particular, I would like to thank everyone at Institut de Physique Nucléaire for inviting me to work with them in Orsay and for sharing their expertise on the SD and French cheese, especially Isabelle, Olivier, and Quentin. An additional thanks is due to the members of KIT for the aid, cross-checks, and for hosting me, particularly to Daniela, Darko, David, and Markus. Finally, a big thanks is owed to Valerio for lending me his expertise on the FD, for calibrating my data more times than I can count, and for showing me the beautiful Italian countryside.

Of course, none of this would be possible without the great people at Penn State, in physics and otherwise. Firstly, a thanks to the PSU Auger group, Foteini, George, Stephane, and Sylvia for being wonderful and supportive people to work with since my very first Auger analysis. Secondly, a thanks to my office mates: Jamyang for being my cultural guide to Tibetan traditions and for always having a good puzzle to work on, and to Colin and Jimmy for all of the laughs and for

letting me pester you with questions.

A big thanks to all of my friends for the game nights, the hikes throughout Happy Valley, the jam sessions, the culture dinners, and the many adventures. In particular I want to thank Dasha, Mark, and Teng who kept me on an even keel every day. This goes double for Sam who has been one of my biggest sources of confidence and support.

Last but not least, thanks to my family for always being behind me and giving me ample encouragement. Without the advice of my grandpa, I would not have gotten into physics and none of this would have happened. The support of my parents, brother, and sister has always been evident no matter how far away I was.

Chapter 1

Introduction to Ultra-High Energy Cosmic Ray Physics

The study of cosmic rays is the study of nature at its most extreme. These particles span the gap between seemingly disparate disciplines that make up modern astrophysics, relativity and quantum mechanics. The first two govern the large, describing phenomena like the death of stars, the birth of galaxies, and the bending of space-time. The latter describes the extremely small such as how two particles can annihilate into pure energy and be gone forever. Cosmic rays are elementary particles as well as nuclei that are created in some of the most violent events in the universe such as accretion onto a black hole and large stellar explosions. After traveling in the near-vacuum of space for millions of light years these particles might arrive at Earth. In this way, cosmic rays live in both extremes, exhibiting properties of both large-scale relativity and particle-sized quantum mechanics.

Ultra-high energy cosmic ray ([UHECR](#)) physics includes the study of how particles with a macroscopic amount of energy came to exist. If they are the byproduct of even more energetic phenomena, then understanding their origin can help explain stellar and galactic structure and evolution. If [UHECR](#)s are instead a relic of the very early universe, then they may give some insight into exotic physics and new particles that would otherwise be hard to study on Earth. Either way, with center-of-mass energies an order of magnitude or more than what can be made by humans, they present a way to test particle interactions in a regime that is otherwise not directly accessible in accelerator collisions today.

1.1 A Brief History of Cosmic Ray Physics

The field of cosmic ray physics began unexpectedly in the early 20th century. At that time, the theories of radiation had been well established for a few decades and there were a number of experiments studying the cause of ionization in air using electroscopes. Leading theories postulated that this ionization was a result of radiation emitted by Earth's surface. In an effort to test this hypothesis, Theodor Wulf created a portable electroscope which was taken to the top of the Eiffel Tower (~ 300 m). If the radiation was indeed emitted from the ground, the ionization of the air should decrease with altitude. While Wulf's results were inconclusive [1], it led to a follow up measurement by Victor Hess. Using Wulf's portable electroscope design, Hess performed ionization measurements at much more extreme altitudes. Using a balloon, Hess and his team reached an altitude of 5400 m. Surprisingly, he found that the ionization first decreased then significantly *increased* with altitude [2]. This lead to the realization that the radiation was of cosmic origin for which Hess won the 1936 Nobel Prize.

Following Hess' discovery, many new experiments were performed to further explore the downward flux of particles. Many experiments [3–5] quickly identified that the particles were positively charged by noting a higher flux from the west than the east caused by their deflection in Earth's magnetic field. In 1937, Pierre Auger made an important observation [6] that cosmic rays, observed using an array of Geiger counters, showed temporal correlations over several square meters. He hypothesized that an energetic primary cosmic ray would interact with air molecules high in the atmosphere which would then create an extensive air shower (EAS) of secondary particles which were causing the observed ionization over extended areas on the ground.

This insight was important for the field of cosmic ray physics as it led to the development of dedicated experiments to measure the temporal and spatial distributions of EASs. John Linsley, in the 1960's, led a team that created such an experiment at Volcano Ranch in New Mexico. This was done by constructing an array of scintillators on the ground which were able to measure the lateral distribution and arrival times of EAS particles to determine the direction and energy of the primary cosmic rays. The most surprising result from the Volcano Ranch array was a measurement of an air shower with an energy above 10^{20} eV

($\simeq 16$ J) [7]. Today we know that air showers with such an energy are extremely rare, but this observation breathed life into the future experiments which would continue the investigation into [UHECRs](#), with energies above 10^{18} eV.

Following the Volcano Ranch result, two important observations were made which had large impacts for the field of [UHECRs](#). In 1964 two radio astronomers discovered the presence of very isotropically distributed photons throughout space [8] called the Cosmic Microwave Background ([CMB](#)). The implication of this for [UHECRs](#) was independently realized in 1966 by Kenneth Greisen [9] and Vadim Kuzmin & Georgiy Zatsepin [10]. Termed the Greisen Zatsepin Kuz'min ([GZK](#)) Effect, this is a description of photo-pion production, the interaction between [CMB](#) photons and cosmic ray nuclei with energies above $\sim 5 \times 10^{19}$ eV. The largest photo-pion channel, via a Δ resonance, is detailed below for cosmic ray protons:

$$p + \gamma_{\text{CMB}} \rightarrow \Delta^+ \rightarrow \begin{cases} p + \pi^0 \\ n + \pi^+ \end{cases} .$$

The result of this interaction is a degradation of the proton energy. This implies that for cosmic rays with above-[GZK](#) energies, the cosmic rays would never make it to Earth or would at least be bounded on the scale of the [GZK](#) interaction length. If there were many more cosmic rays with energy around 10^{20} eV as seen at Volcano Ranch, the signature of the [GZK](#) effect would be apparent by looking at the [UHECR](#) flux.

In the 1990-2000's two prominent experiments tried to test the [GZK](#) hypothesis. The High Resolution Fly's Eye ([HiRes](#)) experiment, which ran for around a decade, included a battery of fluorescence telescopes which are able to directly image the nitrogen emission lines given off when a developing EAS excites air molecules. HiRes, which was able to measure cosmic rays above 10^{17} eV, reported the observation of the GZK cutoff at 6×10^{19} eV [11]. However, this result was in conflict with that of the Akeno Giant Air Shower Array ([AGASA](#)). This experiment instead used scintillation detectors placed in a 100 km^2 array to measure the secondary particles in [EASs](#) on the ground. [AGASA](#) did not observe a decrease in flux in the [GZK](#)-relevant energy range [12]. This discrepancy can be seen in figure 1.1 which shows the measured flux of the two experiments. This plot shows the differential number of observed cosmic rays per area per time per steradian per energy. Due to

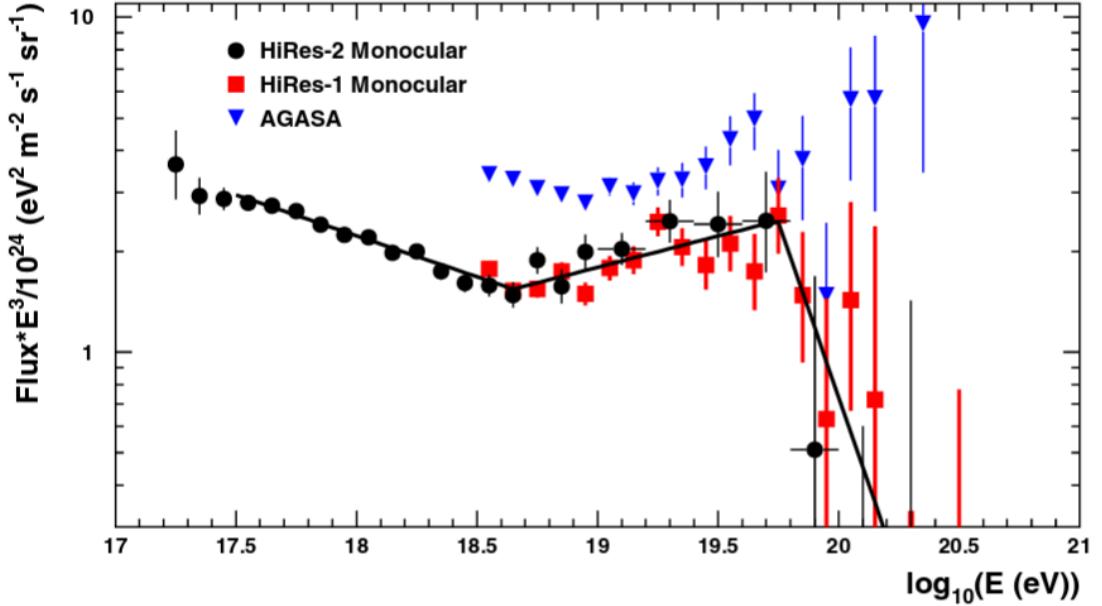


Figure 1.1: The results from the [HiRes](#) and [AGASA](#) experiments shown here have a disagreement on the presence of the GZK suppression which is expected to cause a dip in the flux around $\sim 5 \times 10^{19}$ eV. Due to the power-law decrease in cosmic ray rate with energy, the flux spectra are scaled by E^3 to exemplify the changes in slope. Image taken from [11].

the steeply declining flux ($\frac{dN}{dE} \propto E^{-\gamma}$), the vertical axis has been scaled by E^3 for visual clarity. While the red and black points of [HiRes](#) show the suppression around $\sim 5 \times 10^{19}$ eV, the blue points of [AGASA](#) do not. In hindsight, this can be resolved by noting that the spectra have a generally similar shape. Both spectra include an inflection point where the flux values change from a negative to positive slope. However, the inflection points are misaligned by $0.5 \log(E/\text{eV})$ which indicates that the experiments have an offset in their energy scales.

To help shed some light on this disagreement in flux, the Pierre Auger Observatory was designed to make high statistics measurements in the ultra-high energy range ($> 10^{18}$ eV). Auger combines an array of fluorescence detectors with an array of surface detectors which cover over 3000 km^2 making it the world's largest detector (details of the observatory are covered in chapter 3). After having run for over a decade, Auger has observed a [GZK](#)-like suppression in the flux, though further studies have shown that the suppression might be due to other physical effects. Indeed, much work at Auger, including this thesis, is focused on identifying

the various causes for the changes in slope throughout the cosmic ray spectrum (see section 3.3 and chapter 8).

1.2 Sources of Cosmic Rays

In the energy range above 1 PeV, the sources of cosmic rays are still largely unknown. The models for where UHECRs come from can be split up into two groups, top down and bottom up models. The top down models for cosmic rays postulate that the particles observed are the decay products of yet more energetic particles. The bottom up models assume that cosmic rays are accelerated in some medium, gaining energy, until they escape and propagate to Earth.

As the field of cosmic rays has progressed, more detailed observations have been made for which any source model must adhere. Any theory must explain the flux spectrum of cosmic rays arriving at Earth, their arrival direction, and the types of particles being observed. Given the current cosmic ray measurements, it is extremely unlikely that one single source or process can provide a full characterization at all energies. Instead, a full description of $> 1 \text{ PeV}$ cosmic rays must embrace a combination of multiple processes and/or multiple sources. Some of the leading theories in the top down and bottom up classes are given below.

1.2.1 Top Down Models

The top down models suggest that UHECRs are the result of relics from the early universe. They propose that some system, such as a particle or a topological defect, with sufficient energy could have been created long ago during one of the inflationary periods which then generated a cosmic ray which traveled to Earth.

Various types of exotic or at least beyond the standard model particles have been theorized to be the source of UHECRs. There are explanations as to how a particle with a mass on the order of a few Joules could be created in the early universe [13, 14]. Such extremely massive particles would interact or decay and directly produce $\sim 1 \text{ J}$ cosmic rays. One of the main constraints on such theories is the explanation as to how these particles, which were created in the early universe, would still be abundant today. The interaction time scales which produce UHECRs must be on the order of the age of the universe to still be producing particles.

To match the observed particle flux requires small coupling constants or a high-dimension Lagrangian describing the particle interactions [15]. There is still a window for top down models to be the progenitors of **UHECRs** by invoking a very heavy dark matter particle for which the basic assumption of long lifetime is inherently fulfilled.

Another top down particle candidate is a magnetic monopole which again would have been created in the early universe [16]. While cosmic rays with electric charge are deflected in the interstellar magnetic fields, magnetic monopoles would be directly accelerated. Since they are stable, their existence today would not be a concern like in other top down theories. However, there have been many experiments which have searched for signatures of relativistic monopoles as the source of cosmic rays and so far none have been found [17–20].

An even more exotic subset of top-down models propose topological defects, such as cosmic strings, which could themselves be the sources of **UHECRs**. However, studies of the **CMB** disfavor the existence of topological defects [21–24].

In general, top down models are very unlikely candidates for **UHECR** sources. These theories were far more popular in the wake of **AGASA**'s lack of an observation of a **GZK**-like flux suppression. They require very precise tunings to produce any significant flux of cosmic rays today. Further, most of them fail to produce a spectral shape which is consistent with what is observed on Earth and produce too many ultra-high energy (**UHE**) photons. Indeed Auger has placed the strongest constraints [25, 26] on top down models which have ruled out many **UHECR**-production scenarios.

1.2.2 Bottom Up Models

Bottom up source models propose that **UHECRs** are the result of ordinary particles or nuclei, such as protons, being incrementally accelerated up to the highest energies. Usually the acceleration is a result of the interactions between charged particles and inhomogeneous electromagnetic fields.

The foundation of bottom up models is known as Fermi Acceleration [27]. It is based on high velocity shockwaves (such as a in a supernova, gamma ray burst, etc.) imparting energy to particles caught in the path. This is a multi-step process where each scattering off the non-uniform magnetic fields would progressively increase

the energy on average. Fermi Acceleration models are attractive as they naturally produce power-law spectra ($dN/dE \propto E^{-\gamma}$) which is similar to what is observed by cosmic ray experiments.

One can then ask about the types of environments in which such an acceleration model could be possible. The crucial component to bottom up acceleration models is that the particles must be accelerated in stages and must therefore be contained within the magnetic region for some period of time. A first-order estimate of maximum energy for sources consistent with Fermi Acceleration is given in the equation

$$E_{\max} \sim ZBR. \quad (1.1)$$

From this coarse estimation, astrophysical sources can be ruled out simply based on the accelerated particle's charge Z and the acceleration medium's radial size, R , and magnetic field strength, B . The Hillas plot, shown in figure 1.2, places a number of possible candidate sources in the B - R plane to show the possible **UHECR** source candidates including active galactic nuclei (**AGN**), gamma-ray bursts (**GRBs**), supernova remnants (**SNRs**), and intergalactic medium (**IGM**) shocks. In this plot, the diagonal lines show the values required to produce a 10^{20} eV proton (blue) or iron (red). From this plot it is clear that the types of known candidate sources are few and just barely meeting the requirements at the highest energies. Note that the Hillas plot is an optimistic view as it does not take into account any energy loss mechanisms (synchrotron, curvature, etc.) which would make it more difficult to accelerate a **UHECR**.

1.3 Cosmic Ray Flux Spectrum

The flux spectrum (dN/dE vs E) of cosmic rays above 1 PeV has been observed by many air shower experiments. It is extremely important because it highlights both the open questions in cosmic ray physics as well as the challenges with research in this field. The spectrum can be seen in figure 1.3 which shows the arrival rate of cosmic rays on Earth.

One of the main experimental hurdles highlighted in this plot which must be overcome to study **UHECRs** is the extremely low event rate. Cosmic rays with an energy around 1 PeV arrive at Earth at the rate of one per m^2 per year. At 10^{20} eV,

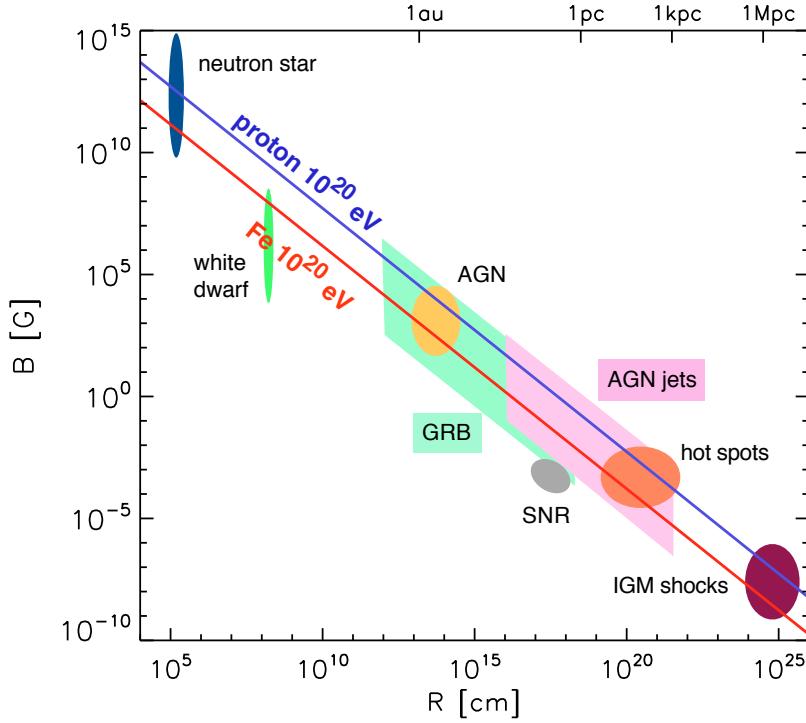


Figure 1.2: Fermi Acceleration is based on many incremental increases in energy as cosmic ray particles are scattered off turbulent magnetic fields. Given the accelerating region's size and magnetic field strength, the Hillas Plot above gives an estimate of the maximum energy which can be reliably imparted. Many potential source candidates are given along with the lines corresponding to the lower limits for 10^{20} eV protons (blue) and iron (red). Image taken from [28].

corresponding to the most energetic event observed at Volcano Ranch, particles fluxes are near 1 per km^2 per century.

This decrease of the observed cosmic ray flux is approximately described by a power law $\frac{dN}{dE} \propto E^{-\gamma}$ where γ is known as the *spectral index*. Probably most interesting are the inflection points where the spectral index has a change in value which likely reflect the transitions between sources and/or acceleration mechanisms. Due to the shape of this spectrum (particularly when scaled by some power of E) resembling a leg, the two most rapid changes in spectral index are referred to as the *knee* at $\sim 5 \times 10^{15}$ eV and the *ankle* at $\sim 5 \times 10^{18}$ eV which can be clearly seen in figure 1.3. It has been the convention to describe the spectrum as a set of broken

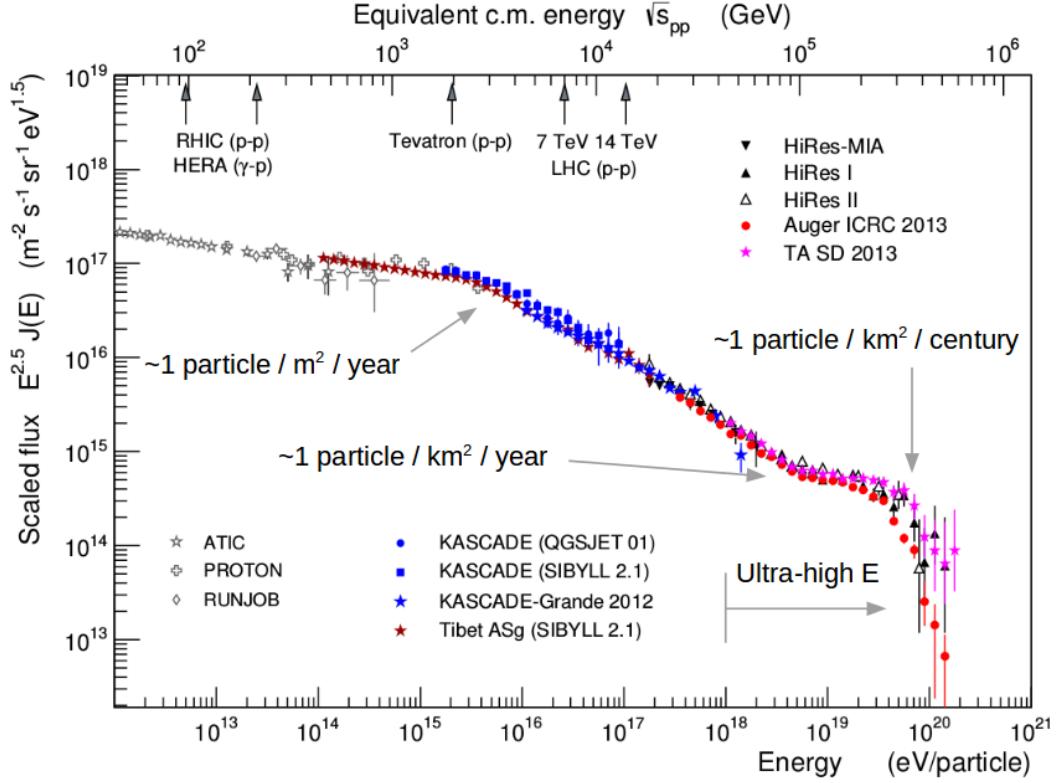


Figure 1.3: The all-particle flux spectra of many experiments are overlayed above. The flux is scaled by $E^{2.5}$ to further exemplify the changes in slope. The knee can clearly be seen at $\sim 3 \times 10^{15}$ eV as well as the ankle at $\sim 5 \times 10^{18}$ eV. The flux suppression has also been seen by three experiments in the range of energies near the GZK predicted limit. Relevant particle rates are given for these three features. Image adapted from [29].

power laws

$$\frac{dN}{dE} \propto \begin{cases} E^{-2.7}, & E < 10^{15.7} \text{ eV} \\ E^{-3.3}, & 10^{15.7} < E < 10^{18.7} \text{ eV} \\ E^{-2.7}, & 10^{18.7} < E < 10^{19.7} \text{ eV}. \end{cases} \quad (1.2)$$

In fact, the spectrum has more features which have only become apparent as the precision of cosmic ray measurements has increased. Indeed exploring the small features in the spectrum at lower energies than the ankle is at the heart of this thesis and will be covered in subsequent chapters. The rest of this section, due to their prominence in the spectrum, will discuss the knee, the ankle, and the suppression along with a brief summary of the models explaining their existence.

1.3.1 The Knee(s)

The energy range surrounding the steepening of the spectrum near 3×10^{15} has been studied by many ground-based observatories such as KASCADE-Grande [30], IceTop [31], EAS-TOP [32], and Tunka [33]. It is generally accepted that this energy marks the beginning of the end for the bulk of galactic cosmic rays.

Given the basic magnetic containment requirement from equation 1.1 it is likely that the end of the galactic component of cosmic rays will be charge dependent. Supposing a bottom up method of acceleration, the maximum attainable energy will be proportional to Z . However, even without proposing an acceleration method, this equation can still be applied to any cosmic rays created in the Milky Way. Indeed, given the galactic magnetic field strength, cosmic rays should only be confined in our galaxy up to some maximal energy, $E_{\max} \simeq 3Z \times 10^{15}$ eV [34]. Beyond this energy the cosmic rays would leak out and not be detected on Earth, directly creating a knee-like structure.

Such a charge dependent model is further motivated by the mass-dependent studies performed by the KASCADE experiment [35] and its successor KASCADE-Grande. These experiments were able to split the arriving cosmic rays roughly into “light” and “heavy” components. They showed that the knee-energy corresponds to a sharp decrease in the observed flux of light nuclei [36]. Following up on this, KASCADE-Grande was able to measure a second knee, sometimes referred to as the *iron knee*, at 8×10^{16} eV [37]. At this energy it was noted that there is a distinct softening of the spectrum for heavy elements. This can be seen in figure 1.4. At the iron knee the flux of cosmic rays is mostly comprised of heavy (blue stars) and medium (green squares) masses. The heavy component undergoes a change in spectral index shifting from a positive to negative slope (since the y-axis is scaled by $E^{2.5}$). These two measurements indicate that the two distinct knees in the all-particle spectrum are due to the changes in the spectra of individual elements. Further, the ordering of these knees for the different mass groups provides validation for the charge-dependent interpretation of the spectrum in the knee-to-ankle energy range.

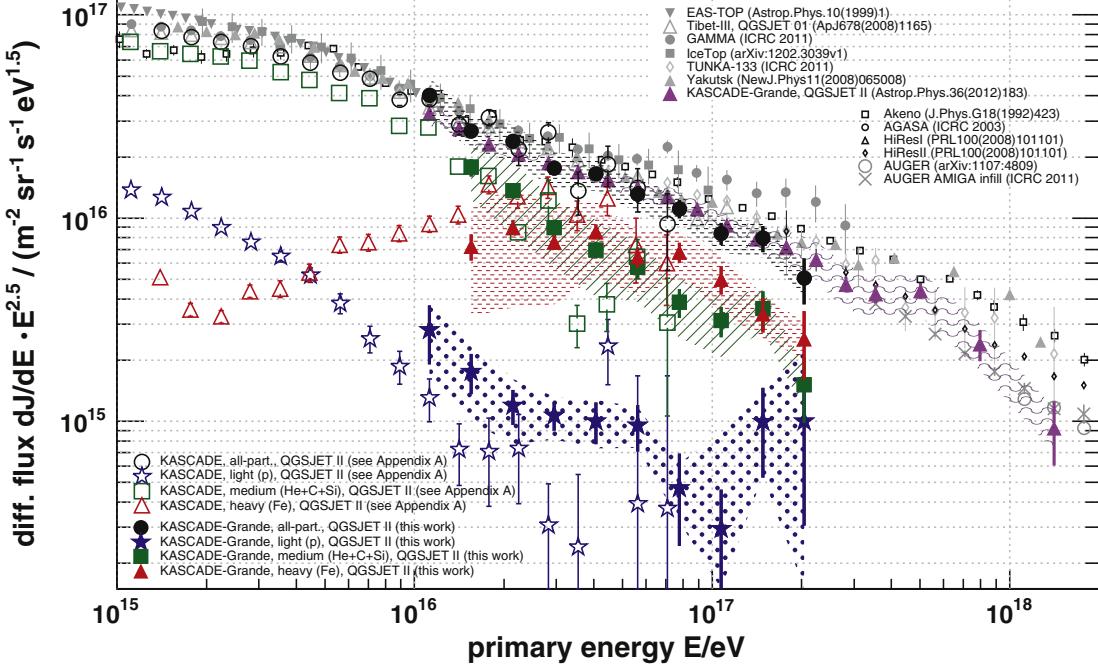


Figure 1.4: KASCADE-Grande was able to split their measurements of EAS into light (blue stars), medium (green squares), and heavy (red triangles) components. Using this, they were able to note two distinct locations where the individual components change spectral index, the knee and the iron knee. The empty and filled markers correspond to two analysis methods by which the data was split into the mass groups (see [38]).

1.3.2 The Ankle and Suppression

The current generation of UHECR detectors are in total agreement about the existence of both the ankle feature near $\sim 5 \times 10^{18}$ eV and the sharp suppression of flux near $\sim 4 \times 10^{19}$ eV (though the exact energies differ by $\sim 10\%$). Beyond the ankle, cosmic rays have been observed to be of extra-galactic origin [39] (see section 3.3.3) and are thus traveling large distances between their source(s) and Earth. So these two features marking the end of the cosmic ray spectrum are best explained as propagational effects of cosmic rays traveling through the CMB.

The most important propagational effect for the ankle region is the pair-production mechanism whereby a cosmic ray interacting with the CMB will create an electron-positron pair. Given a cosmic ray with mass number A , this interaction

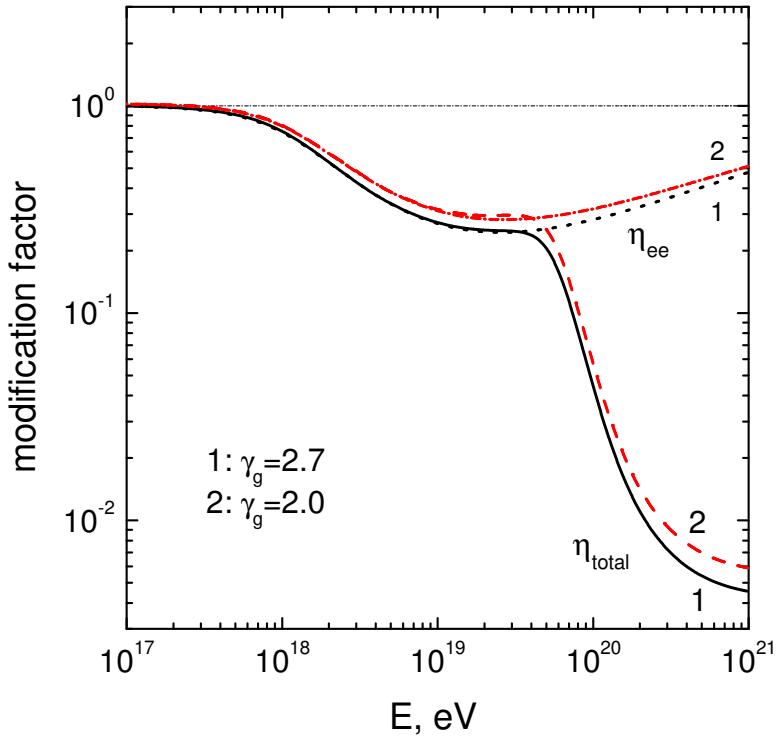


Figure 1.5: The Dip Model proposes that the ankle and knee are the results of UHECRs interacting with the CMB. The curves show the correction factors which modulate the spectra due to pair production (η_{ee}) as well as the GZK effect (η_{total}). The black and red lines show the results for two spectral indexes, γ_g . Image taken from [41].

is described by

$$A + \gamma_{\text{CMB}} \rightarrow A + e^+ + e^- . \quad (1.3)$$

The Dip Model [40] is based on the pair production interaction under the assumption that all UHECRs are protons. The pure-proton composition was a fair assumption to make due to the lack of precision mass measurements at ultra-high energies. Pair production is the major source of energy loss for UHECRs above 2×10^{18} eV until the GZK losses take over around 5×10^{19} eV [41]. The energy lost in creating (potentially multiple) pairs of e^\pm results in a dip in the spectrum and a “pile up” of cosmic rays around the ankle energy. Eventually the GZK interaction begins to dominate which gives rise to the suppression. The Dip Model modification factor, $\eta(E) = J_{\text{obs}}(E)/J_o(E)$, which gives the ratio of observed and source-injected fluxes is shown in figure 1.5. The red and black curves correspond to two injection spectral indexes which have been affected by pair production (labeled η_{ee}) as well

as GZK effects (labeled η_{total}). This model describes well the flux as observed by Telescope Array (TA) for an injected spectral index of 2.6 – 2.7. However, its underlying assumption of a pure proton composition is at odds with the recent mass measurement by Auger (see section 3.3.2) which has shown an increasing average mass number beginning near the ankle.

The model which best describes the Auger results for energies $\gtrsim 10^{18.5}$ is based on a mixed composition of UHECRs. In this model, the maximum attainable acceleration energy of the sources is rigidity dependent. This directly gives rise to the observed steady increase in average mass above $\sim 10^{18.5}$ eV. Various studies have been conducted to simultaneously fit the spectrum and the mass composition above the ankle to measure this source evolution and are discussed in section 3.3.2. The mixed composition model then explains the suppression as largely due to the energy budget at the source rather than *purely* a propagational effect.

Finally, this model also requires an additional galactic component of light particles (i.e. protons and helium) to account for the flux below the ankle. Note that this light particle component is necessarily distinct from the galactic components which create the knee features. An example of a mixed composition fit using Auger spectrum and mass measurements is shown in figure 1.6. The various elemental groups are shown by solid lines which peak in order of their charge. The extra invoked light galactic flux is shown by the dotted line which bridges the gap between the iron knee and the ankle. Due to this thesis work, Auger is now able to make measurements of the cosmic ray flux above the iron knee to help validate the presence of this additional galactic component.

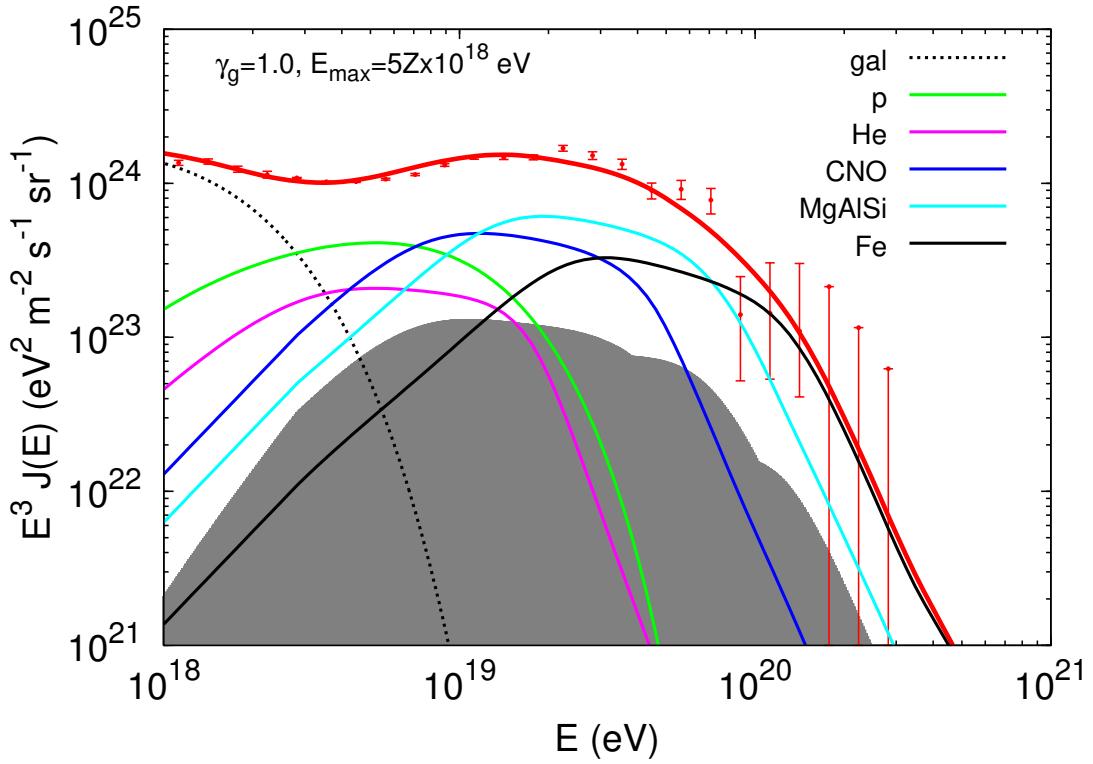


Figure 1.6: This plot shows one realization for a potential mixed composition scenario which fits the Auger spectrum and X_{\max} measurements. Given the increasing average mass of cosmic rays above the ankle, it is likely that different masses all have individual spectra which peak at higher energies for higher Z. It requires an extra galactic component of light particles (dotted line), but is otherwise dominated by mass groups (solid lines). Image taken from [42].

Chapter 2

Extensive Air Showers

Due to the large flux of cosmic rays below 1 PeV, primary particles are directly observed high in Earth’s atmosphere using detectors either mounted on balloons or flown in orbit. However, above this energy, the flux of cosmic rays is ~ 10 events per m^2 per year and quickly decreases. At these rates, it is no longer feasible to launch large enough instruments for direct detection. Instead, cosmic ray physics involves measuring [EASs](#) which are initiated when a cosmic ray interacts with air.

The lack of a direct detection provides an experimental challenge as the properties of the initial cosmic ray (energy, mass, charge, etc.) must be inferred from the resulting particle cascade. A good understanding of how an [EAS](#) develops in the atmosphere is required to be able to extract information about the primary particle.

This chapter summarizes important aspects of air shower development. The first section describes the development of electromagnetic cascades. This is done by using simplifying approximations to estimate the particle production. An extension of this model to hadronic-initiated showers is given in section [2.2](#). A universal model of the way air showers develop is presented in section [2.3](#). Finally, this chapter concludes with a brief description of the modern methods of studying [EASs](#) in section [2.4](#).

2.1 Electromagnetic Heitler Model

The development of an electromagnetic shower is based on the cyclic conversion between photons and leptons. The two dominant processes which govern an

electromagnetic cascade are pair production and Bremsstrahlung. Pair production occurs when a high energy photon creates an electron and a positron, $\gamma \rightarrow e^+ + e^-$. The Bremsstrahlung interaction involves a high energy electron (or positron) scattering off a Coulomb field to produce a photon, $e^\pm \rightarrow e^\pm + \gamma$. It is important to note that both of these processes create one extra high energy particle. This results in the exponential growth of the shower until the particle energies decrease to the point where the photons can no longer pair-produce.

The Heitler model [43, 44] only considers these two interactions to describe the growth and eventual attenuation of an electromagnetic shower and neglects all other energy loss mechanisms. Since both processes are described by the same field equations, they also occur at the same rate and can be treated symmetrically. Specifically, both of these interactions would occur with a length scale, λ_{EM} , which is given in units of traversed matter, termed *grammage*.

The Heitler model begins by noting that the number of high energy electromagnetic particles¹, N , doubles every λ_{EM} ,

$$N(X) = 2^{X/\lambda_{\text{EM}}}. \quad (2.1)$$

Here, X is the traversed grammage and $\lambda_{\text{EM}} = 37 \text{ g/cm}^2$. This is shown schematically in the left panel of figure 2.1.

2.1.1 Development of Electromagnetic-Initiated Showers

The cascade's growth can be sustained as long as there is energy left to continue particle production. As the number of shower particles increases, the average energy per particle decreases as $\langle E(X) \rangle = E_0/N(X)$ where E_0 is the energy of the primary. Eventually, the average particle will have insufficient energy to continue further particle production and the shower will begin to attenuate in size. This *critical energy*, $E_c^{(\text{EM})}$, is typically taken to be the point at which the dominant energy loss mechanism switches from pair production to ionization², about 86 MeV in air.

Rearranging equation 2.1 gives that the maximum number of particles, $N_{\text{max}} =$

¹For the rest of this thesis *electromagnetic particles* refers only to photons, electrons, and positrons.

²Aside from marking the critical energy, ionization is otherwise ignored in the Heitler method.

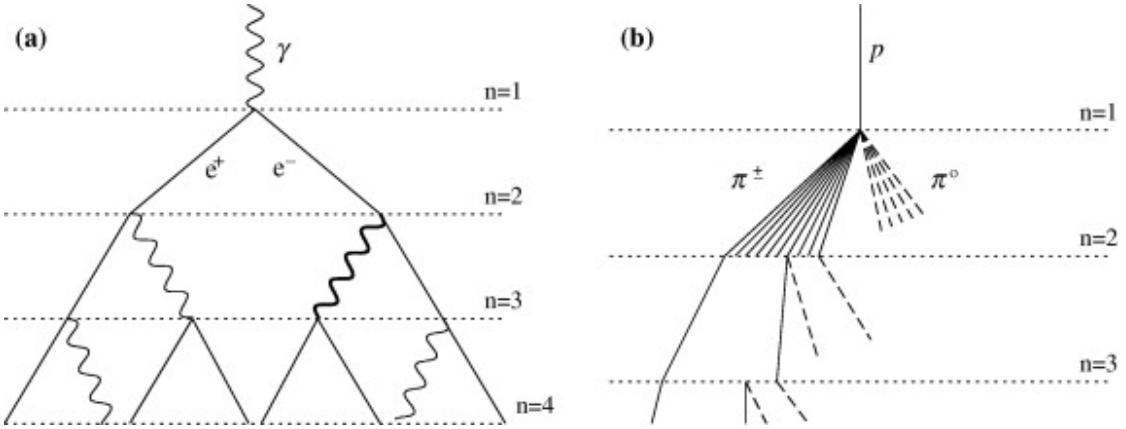


Figure 2.1: The Heitler model is a simplified description of an air shower cascade. Figure (a) shows the Heitler model for an electromagnetic shower which doubles every interaction length, numerated by n . Figure (b) shows a hadronic cascade where charged pions continue the hadronic shower but the neutral pions result in their own electromagnetic shower. Image taken from [45].

$E_0/E_c^{(\text{EM})}$ occurs at a grammage of

$$X_{\max} = \lambda_{\text{EM}} \ln \left(\frac{E_0}{E_c^{(\text{EM})}} \right). \quad (2.2)$$

For a purely electromagnetic shower initiated by a 10^{18} eV primary, these equations give $\sim 10^{10}$ particles at an X_{\max} of 860 g/cm^2 .

Due to the simplifying assumptions, the Heitler Model underestimates the ratio of photons to leptons. The non-symmetric production of e^\pm and γ is partly due to Bremsstrahlung interactions occasionally producing more than one photon. Further, model overestimates the shower X_{\max} partly due to the lack of considering other energy loss mechanisms. However, this model does a good job describing the observed linear relationship of N_{\max} and the logarithmic relationship of X_{\max} with primary energy.

2.2 Hadronic Shower

A hadronic air shower is more complicated to describe than a pure electromagnetic cascade as it involves many multi-particle interactions as well as various decay processes. Further, the scaling of the hadronic interactions with energy is

unknown in the ultra-high energy regime. However, the hadronic shower can still be approximated using an analogous Heitler model via the description of pions.

The interaction of a hadronic primary with an air molecule will produce many high energy pions in an approximately 1:1:1 ratio of charged and neutral flavors, π^+ , π^- , and π^0 , see right panel of figure 2.1. The neutral pions decay almost immediately into two photons ($\tau \sim 10^{-17}$ s) which produce their own electromagnetic sub-showers. The charged pions will either interact, producing more pions in an equal ratio which continues the hadronic cascade, or decay into muons and neutrinos, but on a much shorter timescale ($\tau \sim 10^{-8}$ s) than the π^0 s. The three dominant pion decays considered in this model are listed below.

$$\pi^0 \rightarrow 2\gamma \quad (2.3)$$

$$\pi^+ \rightarrow \mu^+ + \nu_\mu \quad (2.4)$$

$$\pi^- \rightarrow \mu^- + \bar{\nu}_\mu \quad (2.5)$$

The Heitler model has been applied to hadronic cascades [45] by following the population of charged pions over the course of many interaction lengths, $\lambda_{\text{Had}} \simeq 120 \text{ g/cm}^2$ [46]. After n such generations, the energy remaining in the hadronic cascade will be $E^{(\text{Had})} = E_0(2/3)^n$ where the remaining 1/3 of the energy at each interaction is transferred into electromagnetic sub-showers via π_0 decay. This exponential decrease in energy results in the electromagnetic component of an EAS containing 95% of the air shower energy in about $7\lambda_{\text{Had}}$.

2.2.1 Hadronic Maximum

Again, it is important to look at the point where the energy per π^\pm reaches the hadronic critical energy, $E_c^{(\text{Had})}$, at which point the hadronic cascade begins to decline in size. The $E_c^{(\text{Had})}$ refers to the limit where the distance between hadronic interactions becomes comparable to the decay length. Note that the decay length is dependent on the air density and pion energy but occurs around $E_c^{(\text{Had})} \simeq 10\text{--}30 \text{ GeV}$ (decreasing with energy). At this point the hadronic component begins to be converted into muons which generally propagate without interacting to the ground³ which ends the hadronic cascade.

³As an example, a muon with an energy of 2.4 GeV will have a decay length of 15 km [47] which is comparable to the height above Earth's surface at which they would be produced.

For a proton primary that initially created N_{tot} total particles of which N_{ch} are charged pions, the model predicts that the hadronic shower maximum will occur after

$$n = \frac{\ln(E_0/E_c^{(\text{Had})})}{\ln\left(\frac{3}{2}N_{\text{ch}}\right)} \quad (2.6)$$

generations. At this point there will be $N_{\pi^\pm} = (N_{\text{ch}})^n = N_\mu$ charged pions which will each decay into one muon. In terms of grammage, this shower maximum occurs at

$$X_{\max}^{\text{Prot}} \simeq \lambda_{\text{Had}} + \lambda_{\text{EM}} \ln\left(\frac{E_0}{2N_{\text{tot}} E_c^{(\text{EM})}}\right). \quad (2.7)$$

Note that the logarithmic scaling of the shower maximum with energy is exactly the same as the electromagnetic case.

2.2.2 Nuclear Mass Scaling

The description above of the hadronic shower maximum for a proton can be extended to higher mass primaries. For nuclei with mass A , the superposition model is used which simply treats the first interaction as A simultaneous proton showers all with an equal share of the nucleus' energy.

The shower observables, such as the number of particles and the depth of shower maximum, can be found by replacing $E_0 \rightarrow E_0/A$ and summing over the A concurrent showers:

$$X_{\max}^A = X_{\max}^{\text{Prot}} - \lambda_{\text{Had}} \ln(A) \quad (2.8)$$

$$N_\mu^A = N_\mu^{\text{Prot}} A^{1 - \frac{\ln N_{\text{ch}}}{\ln N_{\text{tot}}}}. \quad (2.9)$$

These equations show the scaling as a function of nuclear mass compared to that of a proton shower. Equation 2.8 describes the reduction in grammage of the shower maximum with increasing nuclear mass. Another interesting feature is the non-linear scaling of the number of muons, N_μ , with A (equation 2.9). Even with the reduced number of interaction lengths, there will still be more created muons for more massive hadronic primaries.

Both the reduced X_{\max} and the increased muon number with increasing A are thus important air shower observables for distinguishing the mass distribution of UHECRs. Equations 2.8 and 2.9 estimate that iron primaries will produce 1.8 times

the number of muons at an X_{\max} which occurs $4\lambda_{\text{Had}}$ earlier in the atmosphere. Thus any detector which is sensitive to these observables will be sensitive to the mass distribution of the primary cosmic rays⁴.

2.3 Air Shower Age and Universality

The previous sections covered the Heitler model which estimates various observables of an air shower. This model showed that all electromagnetic or hadronic air showers exhibit similar scaling with energy and primary mass. In fact, similarities between air showers are even more profound and extend beyond the shower depth and particle number. More air shower properties such as the distributions of particles which make up the various cascades have been shown to be dependent on a reduced number of parameters which do not include primary energy and mass to first order. This concept is known as air shower *universality*.

2.3.1 Air Shower Age

Universality, like the Heitler model, is based on the large total number of particles ($\sim 10^9 - 10^{10}$) such that the relative statistical fluctuations are small. Thus any predictive model will work best around the shower maximum. So it is natural to introduce the *shower stage*, t , which is centered around the shower maximum

$$t(X) = \frac{X - X_{\max}}{\lambda_{\text{EM}}}. \quad (2.10)$$

The shower stage gives the number of electromagnetic radiation lengths from X_{\max} .

Near X_{\max} of an electromagnetic cascade the number of electrons and photons can be treated continuously as two coupled differential equations as a function of energy and shower stage. It is useful to seek a solution to these electromagnetic equations in the form for a simple differential equation $\frac{dN(t)}{dt} = \lambda(s)N(t)$ where $\lambda(s)$ gives rate of change in the total particle number and is controlled by a single parameter, s . Lambda would have positive values when the shower is growing, the maximum would occur at $\lambda(s) = 0$, and negative values would describe the rate of attenuation.

⁴These models are only valid on average, and thus will not give significant mass estimates for an individual air shower. However, they can be applied to a collection of EAS measurements.

The power of universality is seen in the expression for lambda below [48,49]

$$\frac{dN(t)}{dt} = \lambda(s)N(t) = \frac{1}{2} \left[\frac{3X}{X + 2X_{\max}} - 1 - 3 \log \left(\frac{3X}{X + 2X_{\max}} \right) \right] N(t). \quad (2.11)$$

Since the rate of growth and decay of the shower is controlled via $\lambda(s)$, it is easy to extract the parameter s , known as *shower age*, which controls the rate of particle increase

$$s = \frac{3X}{X + 2X_{\max}}. \quad (2.12)$$

The form of the shower age equation is important as it implies that the profile of a shower is independent of any initial conditions of the electromagnetic cascade and only depends on the grammage from X_{\max} (where $s = 1$ and $\lambda(1) = 0$). Thus, while a more energetic shower might have more particles in total, the overall shape of any two showers will be the same on average and they will just differ by a size normalization.

2.3.2 Particle Distributions in Air Showers

The notion of universality extends beyond the overall shape to include the distribution of particles within air showers. Early work on universality included the description of the lateral distribution of electrons from the shower axis. This was solved analytically [50,51] for purely electromagnetic showers which resulted in the Nishimura-Kamata-Greisen (NKG) distribution. This shows how the density of particles at distance r from the shower axis changes with shower age

$$\rho(r, s, N_e) = \frac{N_e}{r_M^2} \frac{\Gamma(4.5 - s)}{2\pi \Gamma(s) \Gamma(4.5 - 2s)} \left(\frac{r}{r_M} \right)^{s-2} \left(1 + \frac{r}{r_M} \right)^{s-4.5}. \quad (2.13)$$

Here r_M is the Moliere radius and N_e is an overall electron normalization. Such a universal description is important in UHECR physics as lateral distributions can be directly sampled by surface detectors such as at Auger.

Due to a lack of Joule-scale hadronic physics, such an analytical treatment cannot be applied to the hadronic case. Rather, hadronic Monte Carlo (MC) simulations have been performed to study the details of air shower particles. In [52,53], it was shown that the electrons in a proton-induced shower have an energy and angular distribution which can be described using only shower age and that these

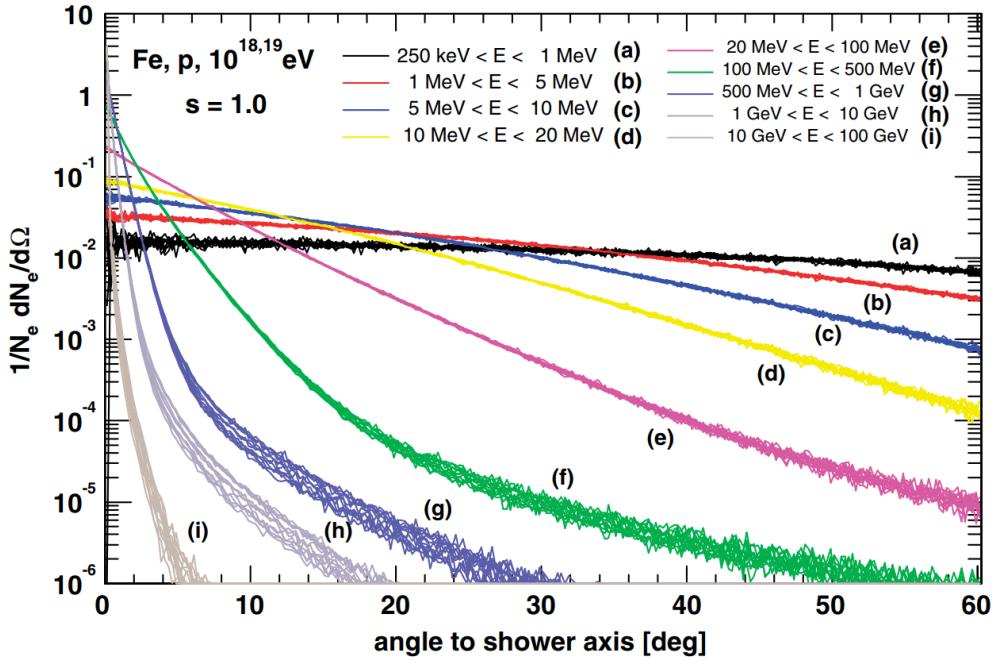


Figure 2.2: The figure above shows the angular distributions of electrons with the shower axis at X_{\max} ($s = 1$). The simulations were done for many iron and proton primaries with energy of 10^{18} eV and 10^{19} eV. The various colors correspond to different energy bins for the electrons. Image taken from [53].

distributions do not strongly depend on the energy or mass of the particle which initiated the shower. An example of universality is given in figure 2.2 which includes many showers with different energies and primary masses. This plot shows that the electron angular distributions with respect to the shower axis, are identical for various energy bins. Were universality not to hold, there would be distinct differences between the angular distributions of the various primary groups. Further simulation-based studies [52–55] have also shown that the lateral distribution and energy distribution of particles in hadronic showers can also be universally described in terms of shower age.

2.4 Methods For Studying Air Showers

The modern study of EASs is conducted both out in the field and in the lab. Computer models of air showers are created using powerful simulations packages which have to take into account all of the standard model. Experimental techniques

employ various detection schemes which use cutting edge technology to precisely measure the secondary particles. Often these methods are used in tandem where simulations help provide context for the observations of air shower experiments.

2.4.1 Air Shower Simulations

The rise of more powerful computer processors has provided the opportunity to simulate the hadronic and/or electromagnetic cascades of air showers from different primaries. The various simulation packages all attempt to encapsulate our current knowledge of the standard model. Some of the common simulation packages for studying air shower physics are Dual Parton Model with JETs ([DPMJET](#)) [56], [EPOS](#) [57], Quark Gluon String Model with JETs ([QGSJET](#)) [58,59], and Sibyll [60]. These models differ in their methods to determine interaction cross sections but generally are based on parton exchanges between interacting hadrons.

All of these models are constantly being updated to conform with modern measurements of hadronic interactions. Likewise, the lack of data in the ultra-high energy range provides the greatest uncertainty in the simulated results. The highest energy data currently available is from the 13 TeV center of mass energy results from the Large Hadron Collider ([LHC](#)). Note that since 13 TeV in the center of mass translates to roughly 10^{17} eV in the lab frame, all simulations performed in the ultra-high energy range will be extrapolations. Figure 2.3 gives an idea of the agreement between models for three common air shower interactions, proton-proton, proton-air, and pion-air. The left panel demonstrates how all the models, which have been tuned to collider data, have strong agreement. Conversely, the right panel shows the cross sections of protons and pions with air for which there are no available measurements. The models do not agree nearly as well and differences are on the order of tens of a percent.

Even with the uncertainties associated with the hadronic models, simulations are still useful for providing a reference for air shower observations. In the previous sections, it was explained how two shower parameters, N_μ and X_{\max} , are dependent on the mass of the primary particle (equations 2.9 and 2.8, respectively). Any experiment which can make measurements of either of these parameters will thus sample the cosmic ray mass distribution. However, these measurements are relative and simulations are often used to provide an absolute scaling, i.e. what the average

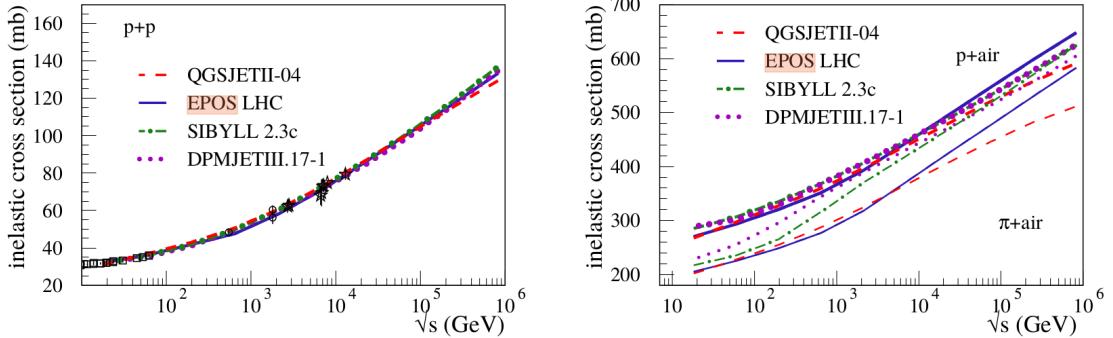


Figure 2.3: Four hadronic interaction models were compared to show their agreement for various inelastic cross sections. The left panel shows the cross sections for the four models which have been tuned to collider data (black points [61] and black stars [62]). The right panel shows the cross sections for air with protons and pions for which the agreement between models is not nearly as good. Image taken from [63].

value of N_μ or X_{\max} is for a mass of $A = 1$. Such a simulation-based study using the N_μ measurements of KASKADE-Grande (see figure 1.4) was used to identify the iron knee at $\simeq 10^{17}$ eV. Similarly, X_{\max} measurements from the Pierre Auger Observatory have been used in conjunction with simulations to study the mass above $10^{17.2}$ eV (see section 3.3.2).

EAS simulations are also used to test the capabilities of various detectors or to explore the expected signature of secondary particles in the detectors. One of the key ingredients in measuring the energy spectrum is knowing the efficiency of the detector to measure an air shower. Since this amounts to knowing what types of showers a given detector is *not* able to measure, this is not an easy problem so simulations provide a natural solution. Such a simulation-based efficiency calculation is done using two hadronic models in section 7.1.

2.4.2 Air Shower Detection Methods

There are a diverse set of techniques to measure the secondary particles produced in the air shower cascades. These can be roughly split into two groups. The first group are telescopes which are able to observe the air shower via the photons which are *not* produced as part of the electromagnetic cascade. The second group are the surface based particle detectors which measure the footprint of the cascades on the ground.

The two styles of air shower telescopes differ in the types of photons that they observe. Cherenkov telescopes view the Cherenkov radiation produced by the relativistic charged particles of an [EAS](#). This light is beamed forward in a cone and can be used to determine the trajectory and energy of the air shower. The cone will have an opening angle given by $\cos \theta = \frac{v}{c n}$ where n is the index of refraction and v is the particle velocity. At large energy, this angle can be quite small. Thus they require the shower axis is pointed roughly at the telescope and are not typically used in the [UHE](#) range where the fluxes are low.

Fluorescence telescopes instead view the light emitted by nitrogen molecules which have been excited by secondary particles. This light is given off isotropically and can be directly observed by a telescope. Since the fluorescence light is relatively weak compared to Cherenkov radiation, these telescopes work best at the highest cosmic ray energies. Both fluorescence and Cherenkov detectors have a drawback of only working in dark conditions which gives them a relatively low duty cycle.

The ground-based detectors come in various forms but all attempt to measure the footprint of the secondary particles on the ground. Early detectors, such as in Pierre Auger's 1937 experiment, used Geiger counters, but this has given way to other detector types. Scintillators are a common option as they act as particle counters for charged particles. Secondary particles travel through the detector medium and produce light which can be measured to estimate the particle densities on the ground. Further, these types of detectors can be partially shielded (or buried) to act as muon counters as the electromagnetic component is much more easily attenuated.

Another type of ground-based detector is a water Cherenkov detector. This works by viewing the Cherenkov light given off as the relativistic secondaries enter the water volume. Again these detectors are used to measure particle density on the ground. More discussion of water Cherenkov detectors will be given in the next chapter.

Ground based detectors have the advantage of working at all times of the day and having a large field of view. Their down side is that you must have many of them to fully map out the footprint of the shower on the ground. This means that they provide only a single snapshot of the shower development. Further, they do not provide a true energy measurement and must depend on other methods, such as simulations, to estimate the cosmic ray energy.

One last type of detector observes the radio- and micro-waves emitted by air showers via the Askaryan effect [64]. This occurs due to the asymmetric distribution of charged particles from deflections in the geomagnetic field along with the excess of negative particles in the shower front. Coherent light is emitted and is observed by radio antennas on the ground. These measurements view the entire development of the shower, not unlike fluorescence detectors.

Chapter 3

The Pierre Auger Observatory

The Pierre Auger Observatory is an air shower detector located in the Mendoza Province of Argentina. The observatory covers over 3000 km^2 of the Pampa Amarilla, making it the world's largest cosmic ray detector. The large area is one of its key features as it is necessary to make precise measurements of the extremely small flux of cosmic rays at the highest energies. The flat pampa, at an altitude of 1400 m, equivalent to 880 g/cm^2 of vertical depth, is an ideal location to observe cosmic rays with an energy greater than 10^{17} eV .

The observatory's design includes a hybrid detection scheme which incorporates both a fluorescence detector (**FD**) and a surface detector (**SD**). The **FD** sites are situated on hills at the edges of the observatory and overlook the **SD** array (see figure 3.1). This design allows for the observation of air showers both during development (on dark, clear nights) and once they reach the ground, providing redundant and complementary measurements of the shower geometry and energy.

This chapter includes a description of the two detectors including their design and the method by which they identify the relevant properties of an air shower. The **FD** telescopes and the basic reconstruction principles are presented in section 3.1. The **SD** design and reconstruction are likewise presented in section 3.2. Greater attention is given to the **SD** as it is more pertinent to this thesis. This chapter concludes with some selected Auger measurements which are important to the cosmic ray community and relevant to this thesis work.

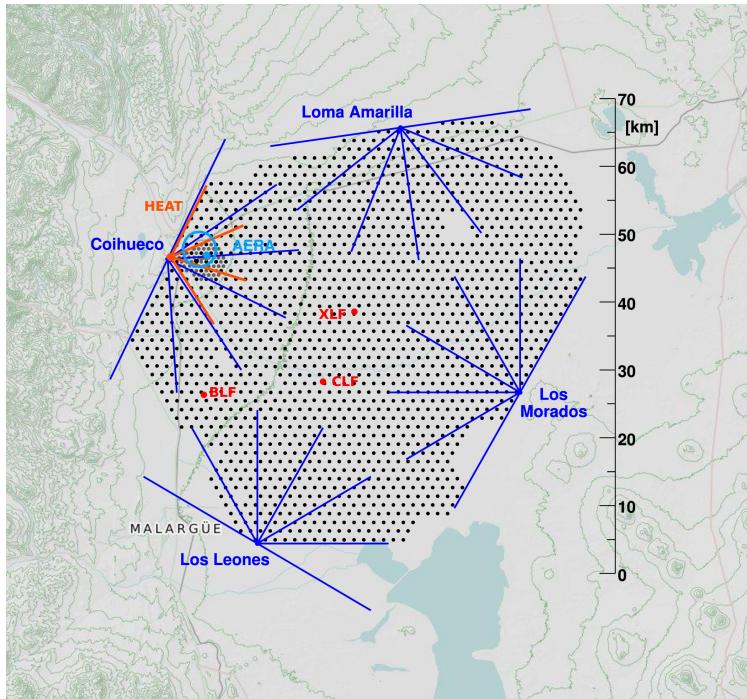


Figure 3.1: The Pierre Auger Observatory consists of two detector schemes. The four **FD** sites and their respective fields of view are shown in blue on the edges of the array. The **SD** stations, arranged in a hexagonal lattice are represented by the black circles. The low energy extensions for the **FD** (known as High Elevation Auger Telescope (**HEAT**), orange lines) and **SD** are located in the top left of the array. Also shown are the Central Laser Facility (**CLF**), the eXtreme Laser Facility (**XLF**), the Balloon Launch Facility (**BLF**), and the array of radio detectors call the Auger Engineering Ratio Array (**AERA**).

3.1 The Fluorescence Detectors

The **FD** observes an air shower's *longitudinal development*, the amount of energy deposited as a function of traversed matter. This is done by directly viewing the isotropically emitted ultra-violet light from excited nitrogen molecules in the vicinity of a developing shower. The amount of fluorescence light given off is proportional to the energy deposited by the electromagnetic component. Thus the **FD** is able to directly probe the energy content of an air shower.

The **FD** includes four sites located on hills which overlook the **SD**, Los Leones, Loma Amarilla, Coihueco, and Los Morados. An additional low-energy extension to the **FD**, known as High Elevation Auger Telescope (**HEAT**), is located at the



Figure 3.2: An **FD** site (Los Morados) in operating mode is shown above. The white doors, which normally protect the **FD** eyes during the daytime, are open, exposing the UV filters and cameras inside the building. The tower behind the **FD** building is a communications tower for both the **FD** and **SD** arrays.

Coihueco site. **HEAT** includes telescopes which can be tilted towards the zenith to view lower energy showers. Such showers would not be otherwise observed by the **FD** as they become attenuated much earlier in the atmosphere.

3.1.1 FD Telescopes

Each **FD** site includes six telescopes each of which covers a $30^\circ \times 30^\circ$ field of view giving a combined 180° coverage in azimuth. Each telescope is housed inside a climate controlled bay which also includes all the **FD** electronics and serves as a local communications hub for the nearby **SD** stations.

During operation the bay doors are opened to expose a 1.1 m UV filter which is mounted inside the doors (see figure 3.2). The filters remove the background light to help isolate the nitrogen fluorescence which boosts the signal to noise ratio. The light is then focused using a segmented spherical mirror onto the camera. Each camera includes 440 pixels each of which is an individual photomultiplier tube (**PMT**). The **PMT** voltages are digitized into 100 ns time bins. A diagram of an **FD** telescope is shown in the left panel of figure 3.3.

Various measurements of the local atmosphere are taken throughout each **FD** data taking period. Aerosol contamination is measured using two banks of lasers,

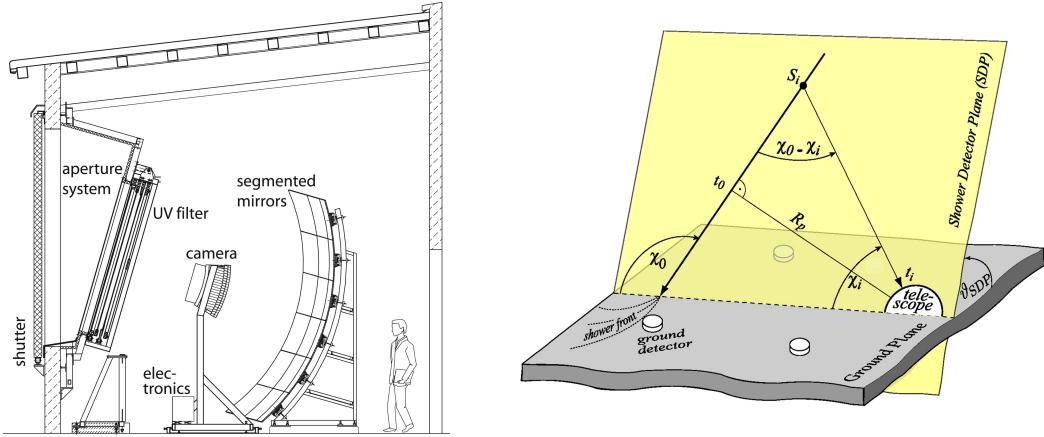


Figure 3.3: Left: The **FD** telescope design is shown schematically including the UV filter, the reflective mirrors, and the camera. Right: The timing information from triggered **FD** pixels is sufficient to fit the shower detector plane (**SDP**). This is the plane that includes the telescope and the triggered pixels. The shower axis must lie in this plane which can be uniquely determined using additional timing information from the **SD** stations.

the Central Laser Facility (**CLF**) and the eXtreme Laser Facility (**XLF**). Both the **CLF** and **XLF** are located roughly in the center of the array and include a laser which produces 355 nm light. The energy of each laser shot corresponds to a 10^{20} eV air shower and can be viewed from each **FD** site. Additional electronics at the **FD** sites measure the ground temperature and pressure which are used in the **FD** and **SD** reconstructions. Each **FD** site also includes a lidar which can be aimed in various directions to measure the cloud coverage above the array.

3.1.2 FD Event Reconstruction

An air shower in the **FD** field of view is seen by many camera pixels and can also be seen by multiple telescopes depending on the trajectory. For relatively high energy showers, it is also possible for simultaneous observation by multiple **FD** sites which provides a stereo (or better) view of the shower. Upon detection of an air shower, the triggered pixels' timing and signal are used to reconstruct the event.

The **FD** reconstruction begins by identifying the shower detector plane (**SDP**). This is the plane which includes the telescope and the shower axis, shown schematically in the right panel of figure 3.3. The **SDP** can be easily determined by fitting the triggered pixels to a line (see figure 3.4, left) and projecting this outward from

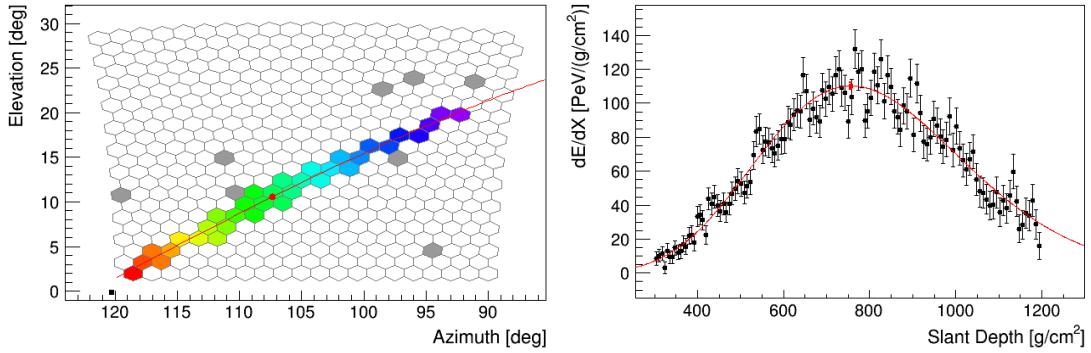


Figure 3.4: The figures above show the results of a hybrid reconstruction of a 7.6×10^{19} eV shower with a zenith angle of 54° . Left: The air shower is seen over many pixels as it develops. The colors indicate the timing in which the pixels were triggered (earlier is purple). The red line indicates the fitted **SDP** with a red dot signifying the location of the shower maximum. Right: The measurement of the energy deposited as a function grammage (black circles) has been fit to a Gaisser-Hillas function (red line). The shower maximum is indicated by the red circle at the peak.

the telescope. The orientation of the shower axis within the **SDP** can be narrowed down by using the timing information from the pixels. However, some trajectories can leave a large degeneracy of possible orientations within this plane. A unique shower trajectory can be found by also including the timing information from triggered **SD** stations. This is known as a hybrid reconstruction and is one of the benefits of the observatory’s multi-detector design.

Upon determining the shower geometry (the orientation of the shower axis), the **FD** is able to measure the air shower’s longitudinal profile which describes the amount of energy deposited in the atmosphere as a function of traversed matter, or *slant depth*. This quantity is determined via the signal intensity in the **FD** pixels and correcting for attenuation between the shower axis and the camera. Such measurements provide an almost¹ calorimetric measurement of the primary’s energy. An example of a longitudinal profile from an Auger event is shown in figure 3.4, right. The signal intensity measurements (black points) of the differential energy deposit (dE/dX) are fit to a Gaisser-Hillas function [67] (red line). From the fit, the location of X_{\max} can be directly extracted from the fitted peak and an integration

¹The measurements are not exact due to the muons, neutrinos, etc. which carry some fraction of the energy to the ground. However, this energy fraction has been studied at Auger via simulations [65] and real data [66] and is corrected for after the reconstruction.

over X yields the calorimetric air shower energy.

Both the X_{\max} and energy measurements make the **FD** an invaluable component to the observatory. While the former can help identify the mass distribution of cosmic rays, the latter provides the observatory's energy scale. Unfortunately, the **FD** is only able to operate on clear nights with a low moon fraction which yields an average duty cycle of 15%.

3.2 The Surface Detectors

The **SD** stations measure the energy density of secondary particles on the ground. Whereas each **FD** telescope can independently measure an air shower, each **SD** station only provides a single measurement of the energy density and time of arrival of the secondary particles. Thus each event requires multiple stations to map out the shower footprint. By combining the measurements of three or more stations, a reconstruction can be performed which produces the air shower trajectory along with an energy estimator.

The **SD** is comprised of around 1600 water Cherenkov detectors (**WCDs**) spread out in a hexagonal array with 1500 m spacing. The cylindrical containers are 1.2 m tall and hold 12,000 L of ultra-pure water. The **WCDs** observe the Cherenkov radiation given off by highly boosted secondary particles that enter the water volume. The inside of each water tank is lined with a reflective material which helps collect this light into the three 9 in diameter **PMTs** mounted to the top. The trio of **PMTs** provides redundant measurements of the energy deposited by the air shower particles to increase the sampling precision. Note that due to the light-tight design of the stations, the **SD** has a 24-hour operating time giving it a $> 99\%$ duty cycle.

The **PMT** pulse is read out on both the dynode and the anode. This provides a larger dynamic range (x32) which is important when sampling the large energy densities close to the shower axis which would otherwise saturate the **PMTs**. The outputs of each **PMT** are digitized using a flash analog to digital converter (**FADC**) in 25 ns time bins. These traces are analyzed and sent wirelessly to the nearest **FD** station on the edges of the array. All onboard electronics, including a GPS receiver and processor, are powered by a solar panel and battery which make each station completely self sufficient. An image of an **SD** station is shown in figure 3.5.



Figure 3.5: An [SD](#) station is shown here with the Los Leones [FD](#) site in the background. The solar panel mounted to the top of the water tank can be seen along with the GPS receiver and transmission antenna attached to the pole.

Nested within the [SD](#) is a low energy extension which is comprised of identically designed [WCDs](#) but with half the lattice spacing, 750 m. This 750 m array can be seen in the top left of the figure 3.1 and is overlooked by [HEAT](#). While the 1500 m array is fully efficient² for air showers with zenith angles of less than 60° and energies greater than $10^{18.5}$ eV, the 750 m array is efficient for showers with angles less than 55° and $10^{17.5}$ eV. However, this thesis provides an update for the 750 m array which lowers this energy threshold down to 10^{17} eV, near the iron knee.

3.2.1 SD Calibration

The primary background for the [SD](#) is the constant flux of primarily muons (as well as electrons and photons) which are created in low (not ultra-high) energy showers. These background particles produce signals in the [WCDs](#) which set a barrier to the types of low-energy signals that can be measured. However, muons also provide the method by which the SD stations are calibrated.

The traces produced by the digitized [PMT](#) signals are in units of analog to

²Full efficiency is the lower energy limit above which an air shower that falls on the array is guaranteed to be detected.

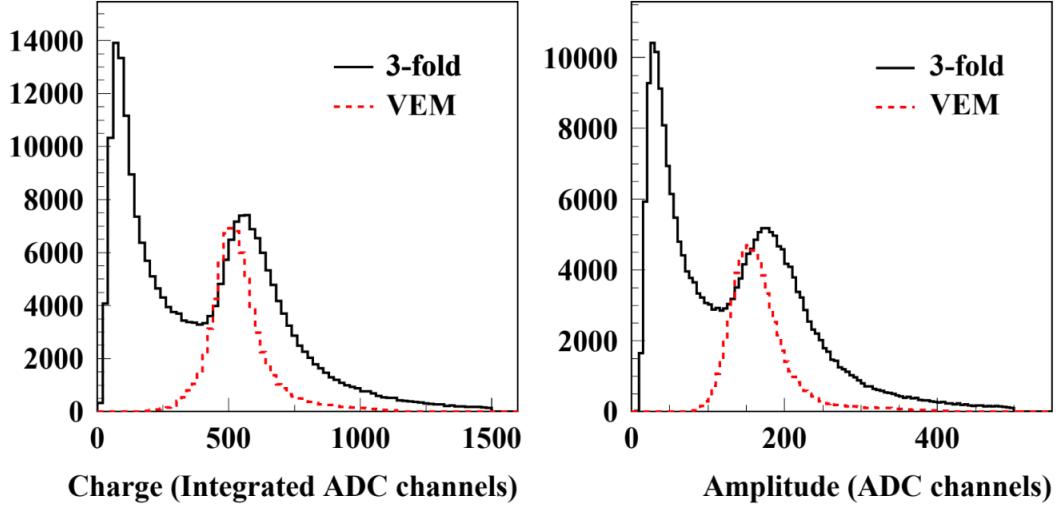


Figure 3.6: The left and right plots show histograms of the charge and amplitude of background signals measured in an [SD](#) station. The black curves correspond to any signals measured in all three [PMTs](#) which were five [ADC](#) channels above the baseline. The red curves were created using an external muon telescope which distinguished only vertical muons. A slight offset is seen between the two peaks due to the black curve being the sum of the muon peak with the falling exponential. Image taken from [69].

digital converter ([ADC](#)) channels. To account for gain differences between [PMTs](#), the traces are converted into a standardized unit by using the background muons. While the water volume is a calorimeter for electrons and photons, it is a scintillator for muons. The energy lost in a water volume is proportional to the muon track length and largely independent of the muon energy up to ~ 100 GeV [68]. Thus the chosen standardized unit is a vertical equivalent muon ([VEM](#)). This is the [PMT](#) signal created by a muon with a vertical trajectory.

Each [SD](#) is constantly monitoring the area (Q) and amplitude (I) of the digitized signals³. At a background-particle event rate of about 25,000 Hz, each station quickly builds up a large set of measurements to determine the conversion between [ADC](#) channels and [VEM](#). An example histogram of the integrated background signals and amplitudes is shown in figure 3.6. These figures show the measured distributions (solid black lines) of Q (left) and I (right) and the respective measurements due to only vertically going muons (dashed red lines) as determined by an external muon

³Note that the digitized signals are a measure of the current within the [PMT](#). Thus the signal area is sometimes referred to as the *charge*, Q , whereas the signal amplitude is a *current*, I .

telescope. The black curve shows the steeply falling distribution of low energy particles and corner-clipping muons along with a second red peak due to vertical muons. There is a slight offset between the true muon peak and the one given by the black curve due to it being the sum of two distributions. An onboard computer performs an initial calculation to identify the muon peaks and correct for this offset. This provides the first estimate of the charge and amplitude of a vertical muon, Q_{VEM} and I_{VEM} , respectively.

A further, more accurate determination of the station calibration is performed offline during reconstruction. When an air shower is detected (see section 4.2), 60 s of calibration measurements (i.e. the black histograms in figure 3.6) are sent to the main Auger campus to provide up to date calibration information for each event. The offline calibration includes a more robust method to find the Q_{VEM} peak which includes fitting a parabola with negative curvature. The offline calibration is more reliable when the muon peak is relatively small. For more information of the offline treatment of the calibration data, see section A.1.

3.2.2 SD Event Reconstruction

Using the calibrated traces from the SD stations, the air shower arrival direction can be determined using the timing of the signals. The secondary particles arrive on the ground in a curved shower front. The shower direction can be determined by finding the zenith and azimuth angles for which a shower front would arrive at each station at the moment of its triggering. The accuracy of the shower reconstruction improves with the number of stations included in this determination and is generally around a few degrees. An SD event is shown in figure 3.7, left. The black line indicates the direction of the air shower and ends at the shower *core* (where the shower axis intersects the ground) in the center of the event. The size of the circles is related to the signal deposited in the stations and the color indicates the start time of the signals. The dark-blue stations were the first to be hit by the shower front.

The shower geometry is determined almost entirely by the start times of the signals in the stations. The secondary particles are assumed to arrive in a shower front that is cylindrically symmetric around the shower axis. The orientation of the axis can be determined by comparing the signal start times to the arrival time

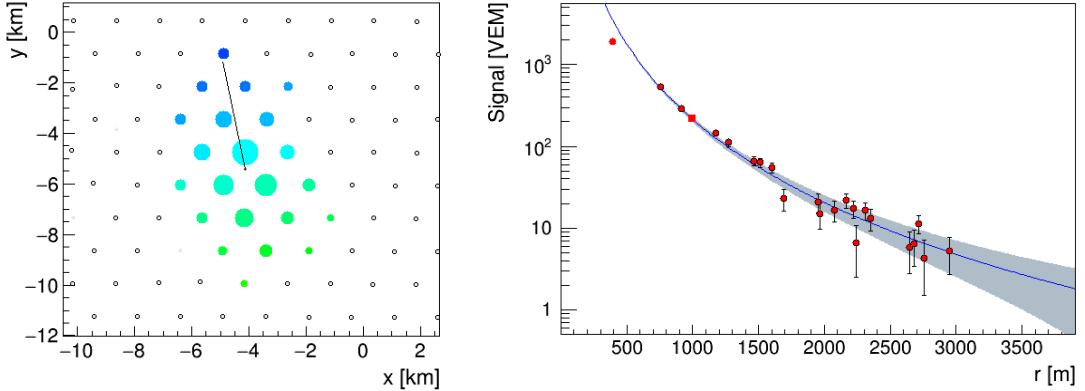


Figure 3.7: An Auger SD event with energy 7.5×10^{19} eV and zenith angle 54° is shown above. Left: The triggered stations in the event are indicated by colored circles. Earlier stations are shown in dark blue and later stations are shown in green. The size of the circle corresponds to the deposited signal in the station. The shower axis and core location are indicated by the black arrow. Right: The measured signal (red dots) as a function of distance to the shower axis is fit to a lateral distribution function (blue line).

of the shower front at a particular station. See section 5.2.1 for more details.

The SD makes sparse measurements of the deposited signal. Thus a calorimetric measurement of the energy is not possible. Instead the SD reconstruction produces an energy estimator based on the size of the shower at ground level. The shower size is determined by the lateral distribution of the signals. First the amplitudes of the station signals are plotted as a function of distance from the identified shower axis. Then this distribution is fit using an empirically chosen lateral distribution function (LDF) (see section 5.2.2 for more information). From this fit, the shower size (in VEM) is given by the expected signal at a reference distance from the axis. For the 1500 m array, the reference distance is 1000 m, $S(1000)$, and for the 750 m array shower size is measured at 450 m, $S(450)$. These distances have been chosen via Monte Carlo simulations such that the functional form of the LDF has the smallest systematic error on the size estimator. An example of an LDF fit is shown in the right plot of figure 3.7. The measured signals (red circles) have been fit to an exponentially falling LDF (blue line). The shower size estimator is shown with a red square and has a value of $S(1000) = 213$ VEM.

The $S(1000)$ or $S(450)$ values are then corrected for the shower age. Air shower universality describes how any two showers will have similar properties at the same

shower age. However, due to the daily and yearly fluctuations in the atmospheric conditions as well as the changing slant depth with zenith angle, showers are not sampled at a consistent age. Thus an atmospheric correction [70] is applied based on the local weather measurements taken at the FD sites. Finally, a zenith-dependent correction is applied to account for more highly inclined showers being sampled at a later age than vertical showers. This second correction is described in section 6.3.

A scaling is finally applied which converts the corrected shower size into an energy. The relationship between the size in VEM and the air shower energy is found via a calibration with the FD. This process uses events which can be reconstructed independently by the two arrays and is described in section 6.4.

It is important to note that, regarding their construction, the 1500 m and 750 m arrays are identical aside from the stations' separation. Likewise, the methodology in the event reconstruction is identical for the two arrays. However, due to their differing size, they generally have different systematics and thus Auger uses two independent reconstructions. This thesis will focus almost exclusively on an update to the 750 m array reconstruction.

3.3 Selected Physics Results

With over 73,000 km² sr yr of combined exposure accumulated over 13 years, the Auger collaboration has published some of the most important experimental results in the ultra-high energy regime. This section includes a selected subset of those results which are important to the cosmic ray community and relevant to this thesis. This section includes only previously published work. Updates which are part of this work will be presented in the subsequent chapters.

3.3.1 Energy Spectrum

The energy spectrum is independently measured at Auger by the FD, the 1500 m array, and the 750 m array. Figure 3.8 shows the combined spectrum [26] of these various data sets. Note that the flux, $J(E)$, has been scaled by E^3 to emphasize the changes in spectral index. The spectrum is shown above $10^{17.5}$ eV where (prior to this thesis) the 750 m array was fully efficient.

The combined data set allows for a very precise measurement of the spectral

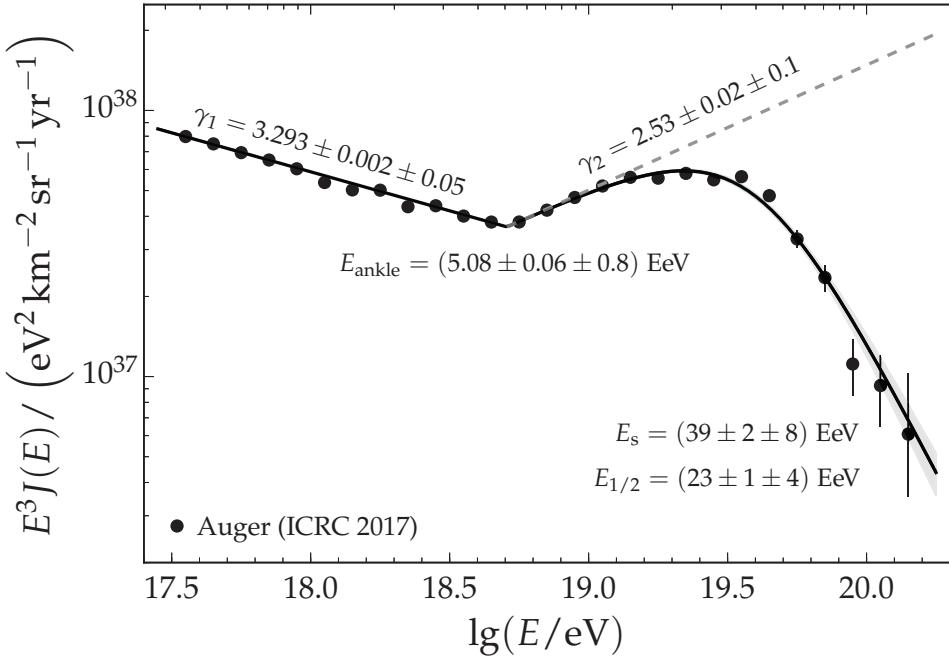


Figure 3.8: The energy spectrum measured by Auger is shown above. The data has been fit to a broken power-law with a smooth suppression. The various spectral indexes and energies corresponding to the inflection points of the spectrum are given along with their statistical and systematic errors. Image taken from [26].

indexes on either side of the ankle (γ_1, γ_2) along with the energy at which this change occurs (E_{ankle}). The spectrum has been fit to a broken power law with a smooth suppression

$$J(E) = \begin{cases} J_0 \left(\frac{E}{E_{\text{ankle}}} \right)^{\gamma_1}, & E \leq E_{\text{ankle}} \\ J_0 \left(\frac{E}{E_{\text{ankle}}} \right)^{\gamma_2} \left[1 + \left(\frac{E_{\text{ankle}}}{E_s} \right)^{\Delta\gamma} \right] \left[1 + \left(\frac{E}{E_s} \right)^{\Delta\gamma} \right]^{-1}, & E > E_{\text{ankle}} \end{cases}. \quad (3.1)$$

The ankle, at 5.08 EeV, is clearly visible along with the suppression, E_s , at the highest energies. This thesis includes an update in chapter 8 to the Auger spectrum which now extends down to 100 PeV.

3.3.2 Cosmic Ray Mass Measurements

The masses of the primary particles are not directly observed in the ultra-high energy regime. However, as described in chapter 2, the primary mass does affect the depth at which air showers reach their maximum size. Auger is able to sample

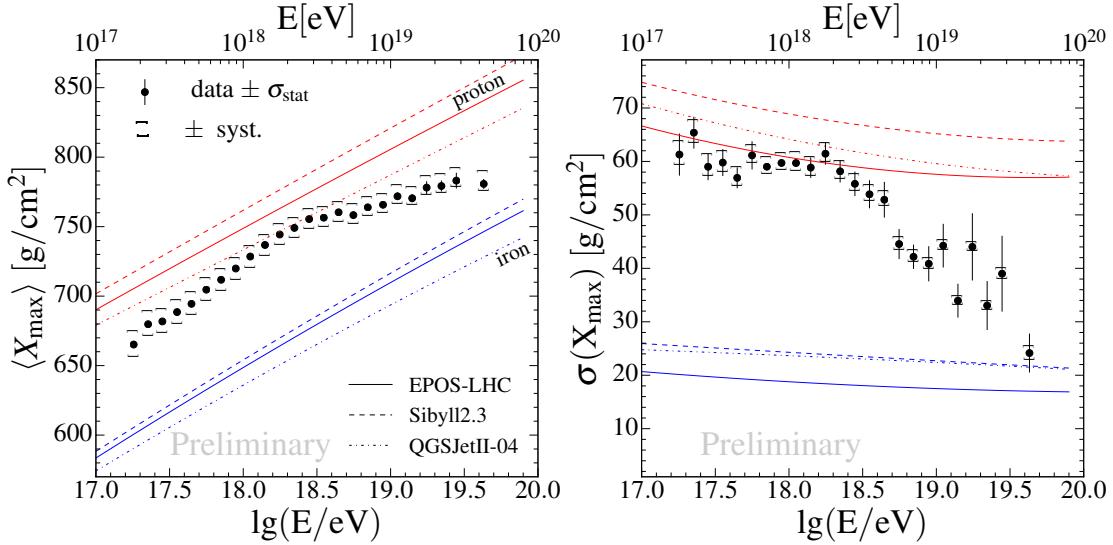


Figure 3.9: The measurements of X_{\max} from the FD give Auger a sensitivity to the mass distribution of cosmic rays. The average grammage along with the standard deviation as a function of energy are shown in the left and right plots, respectively. The Auger data (black dots) are compared to the results from simulations of proton (red) and iron (blue) primaries using three hadronic models. Image taken from [71].

this distribution using the lateral profiles from the FD. With a resolution better than 25 g/cm^2 above 10^{18} eV , Auger can make precise measurements to help infer the primary mass distribution.

Figure 3.9 shows the first two moments of the observed X_{\max} distributions along with the results from three air shower simulation packages. The lines corresponding to the average X_{\max} for simulated proton and iron primaries are indicated in red and blue respectively. The mean X_{\max} distribution on the left shows a composition which grows increasingly light up to $10^{18.4} \text{ eV}$ after which the average mass grows increasingly heavy. The standard deviation of X_{\max} is shown in the right plot. The data shows a trend of an increasingly narrow X_{\max} distribution above $10^{18.4} \text{ eV}$.

Using these X_{\max} measurements, it is only possible to determine the average tendency of the UHECR mass distribution which leads to degeneracy in the composition. This can be broken by also including the Auger energy spectrum measurements along with a model of the UHECR sources. As described in chapter 1, a leading model to explain the UHECR sources involves the maximum attainable energy of acceleration to be rigidity dependent. Specifically, higher Z elements would dominate the spectrum at higher energies. The X_{\max} measurements combined

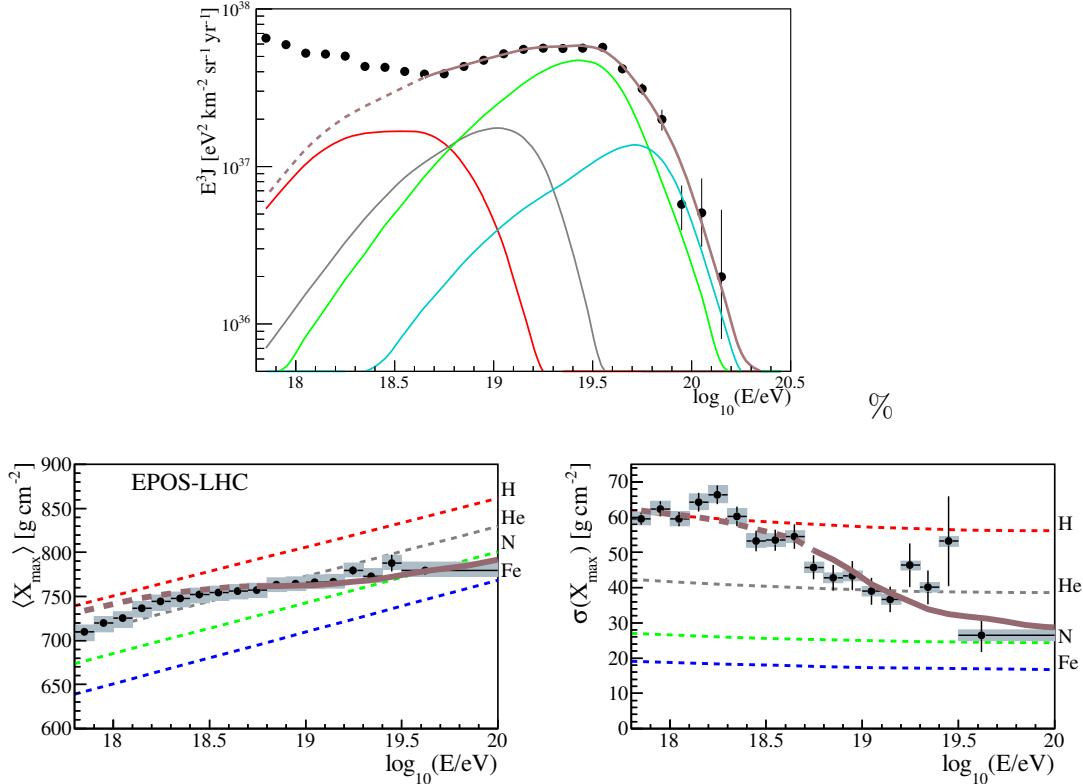


Figure 3.10: By using the distribution of X_{\max} along with the energy spectrum measured at Auger, it is possible to make a combined fit of the data to determine fractional composition by mass. The figures above show the results of a rigidity dependent source model which has been fit to the Auger measurements. The various colors show different components (see text). The net result of the fit is shown in brown. Note that the fit was only performed above $10^{17.8}$ eV (solid brown lines). Image taken from [72].

with the flux measurements can then be fit simultaneously to break the composition degeneracy. This was done in [72] and is shown in figure 3.10. In this study, the spectrum was split up into being the sum of various mass groups: ${}^1\text{H}$ (red), ${}^4\text{He}$ (grey), ${}^{14}\text{N}$ (green), and ${}^{56}\text{Fe}$ (blue). The cosmic rays were assumed to be accelerated with an initial power-law distribution and subsequently attenuated to account for propagational effects (GZK, etc.). The remaining flux can then be summed to produce the Auger energy spectrum and X_{\max} distributions. The amplitude of the four mass groups, the source spectral index, the rigidity cut-offs, and the overall normalization were fit to match the Auger data (black dots) above

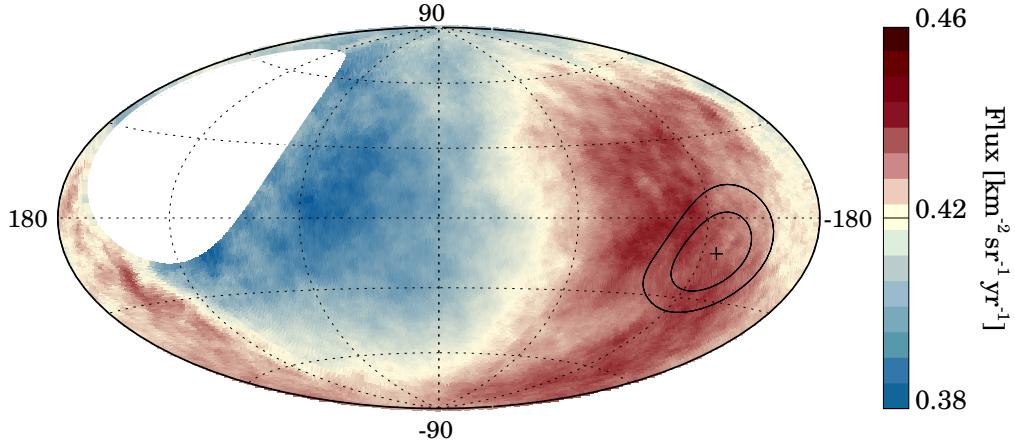


Figure 3.11: The large-scale distribution of cosmic rays was studied using the Auger event set. The fluxes of cosmic rays, smoothed by a 45° top-hat function, above 8×10^{18} eV are shown above in galactic coordinates. The 5.2σ significance dipole is shown by a black cross. The black lines enclose the 68% and 95% confidence intervals. Image taken from [39].

the ankle, resulting in the brown lines in the three plots.

3.3.3 Large-Scale Arrival Directions

Cosmic rays, due to their charge, are deflected in the intergalactic and galactic magnetic fields. Modern models of the galactic magnetic field indicate that deflections can be a few tens of degrees [73], depending on the trajectory. This makes point source searches difficult. However, with $\sim 10^\circ$ deflections, analyses of the large-scale structure of the arrival distributions remain tenable. These studies aim to identify deviations from an isotropic arrival distribution. Figure 3.11 is the result of a such an analysis which studied the dipole structure in the observed cosmic ray arrival directions.

This analysis identified the line along which the arrival directions are most polarized. The figure shows the location on the sky of $E > 8 \times 10^{18}$ eV cosmic rays which have been smoothed with a 45° top-hat function. The direction of the dipole relating to maximal flux is indicated by the black cross and encircled by the 68% and 95% confidence intervals. This figure is shown in galactic coordinates with the galactic center at the origin. At a significance of 5.2σ , this is the first result in the

ultra-high energy regime for which anisotropy has been observed⁴. Further, the dipole direction at 100° in right ascension gives compelling evidence that above 8×10^{18} eV, cosmic rays have an extra-galactic origin.

⁴TA has also noted a “hotspot” in the northern sky (not within view of the Auger detector) but with a significance of 3.4σ [74].

Chapter 4

The Auger SD Triggers

To search for the relatively rare signature of [EAS](#)s within the heavy flux of background particles, the [SD](#) uses a set of quality triggers which range in scope from the individual station-level up to the array-level. At the individual station-level, triggers are used to identify when the amount of light observed inside a [WCD](#)'s water volume is unlikely to be due to background particles. These triggers are described in section 4.1. Section 4.2 details how the array-level triggers search for configurations of triggered stations that are consistent with an air shower footprint. While these trigger levels have been used for the duration of Auger, an additional set of station-level trigger algorithms were implemented in mid-2013. Section 4.3 is an updated analysis of the various trigger levels including a calculation of the trigger rates, the typical amplitudes of signals, and the triggering efficiencies.

4.1 Station-Level Triggers

The goal of the station-level triggers, termed T1 and T2, is to identify patterns in the digitized signal traces which are likely to be due to secondary particles from an [EAS](#) using pre-set algorithms. This is not a trivial task as the same types of particles which make up the signal also account for most of the background. These background particles, with a rate of $200 \text{ m}^{-2}\text{s}^{-1}$, are the secondary particles of $\sim 100 \text{ GeV}$ showers and are mostly muons but also electromagnetic particles [75].

The T1 and T2 triggers [76] include four algorithms, two *old triggers* which have been in use since 2004 and two *new triggers* which were installed in mid-2013. A schematic diagram summarizing the T1 and T2 trigger types is shown in figure 4.1.

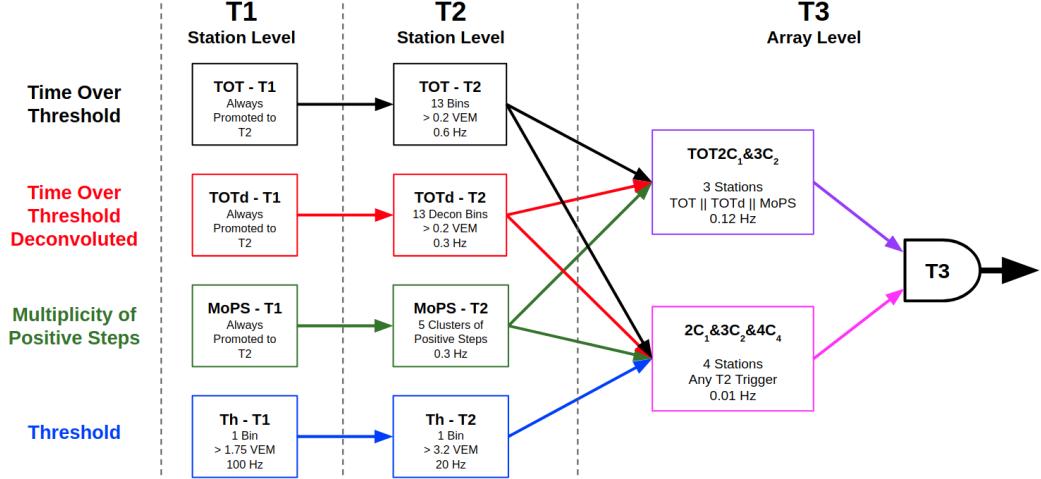


Figure 4.1: This diagram shows the requirements for the various T1, T2, and T3 level triggers. TOT, TOTd, and MoPS are always promoted directly to T2 and are the only three triggers from which a TOT2C₁&3C₂ T3 trigger can be made.

The distinction between T1 and T2 is important as the lowest array-level trigger can only be created from T2-triggered stations. The T1 triggers simply provide extra information which can be used during the reconstruction.

4.1.1 The Old Triggers

The two old trigger algorithms are called threshold (**Th**) and time-over-threshold (**TOT**). Both of these algorithms look for temporal correlations of elevated signal amplitude between the three **PMTs** in a station.

The **Th** trigger requires that all three **PMTs** in a station observe a pulse height in the same time bin which is some multiple of the typical background signals, I_{VEM} (see section 3.2.1). The **Th** trigger comes in two levels, Th-T1 and Th-T2. Th-T1 triggers require correlated pulse heights of $1.75 I_{VEM}$ yielding a trigger rate of 100 Hz (per station). Likewise, the Th-T2 trigger has a stricter amplitude requirement of $3.2 I_{VEM}$ and thus a lower rate of 20 Hz. This threshold was set to reduce the ~ 3 kHz background particle rate to a level that can be handled by the limited bandwidth between the individual detectors and the Central Data Acquisition System (**CDAS**).

Signals that pass *only* the **Th** trigger are sharp, short duration spikes (see the top-left trace in figure 4.2) which is the typical signal structure created by a muon

traversing the water volume. This makes the **Th** triggers particularly good at measuring older showers for which the electromagnetic component has been highly attenuated and the signals have a higher muon fraction.

The **TOT** trigger requires a moderately high signal that is long in duration. To fulfill a **TOT** trigger, at least two **PMTs** must identify 13 or more signal bins with amplitudes over $0.2 I_{VEM}$ within a sliding 120 bin ($3 \mu s$) window. An example of a **TOT** trigger can be seen in the top-right of figure 4.2. Because of how unlikely it is that background particles can mimic such a signal pattern, **TOT** triggers are automatically promoted to the higher status of T2.

4.1.2 The New Triggers

In mid-2013, two new types of T2 triggers were installed in the **SD** which were designed to measure small amplitude signals. Note that the requirements for the **Th** and **TOT** algorithms directly give the minimum integrated signal that can be measured. These are 1.75, 3.2, and 2.6 VEM for the Th-T1, Th-T2, and **TOT** triggers, respectively¹. Attempting to measure smaller signals than this requires the design of an algorithm that can contend with the much higher background in the signal range of the typical background muon, ($\simeq 1$ VEM by construction). Thus the two new triggers, time-over-threshold deconvoluted (**TOTd**) and multiplicity of positive steps (**MoPS**), were designed specifically to not be sensitive to muon-like signals.

The **TOTd** algorithm uses the average decay time of light inside the reflective water volume to identify signals which are too small in amplitude to pass the **TOT** criteria. The algorithm is based on the idea that signals with a significant electromagnetic component will have a characteristic shape like the **TOT** trace. Such traces have a sharp increase followed by an approximately exponential decrease with time constant, τ , which is given by how fast the Cherenkov light is absorbed. For an ideal signal with a truly exponential tail, a deconvolution would result in a flat trace

$$D_i = \frac{S_i - S_{i-1} e^{-\Delta t/\tau}}{1 - e^{-\Delta t/\tau}}. \quad (4.1)$$

¹In fact signals are typically much larger than this as the traces are not shaped like a single binned pulse or a 13 bin by $0.2 I_{VEM}$ box. Further, in some rare cases signals have been observed below these limits when the online **VEM** calibrations significantly differed from the offline calibrations.

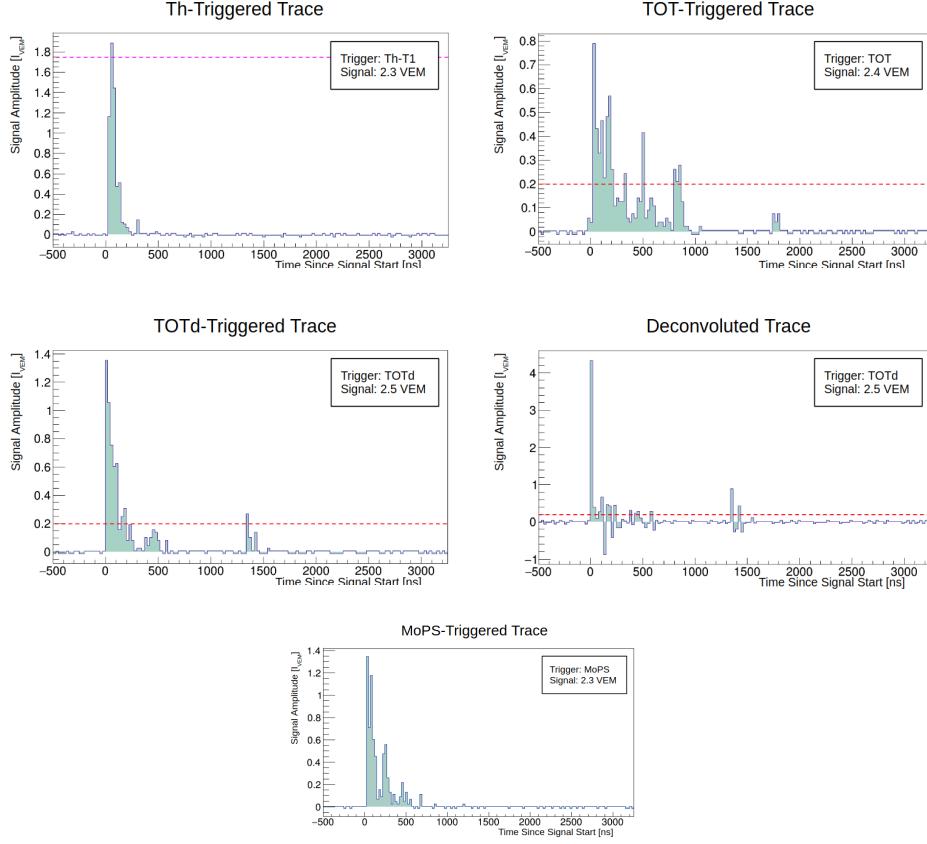


Figure 4.2: The traces above are examples of ~ 2.4 VEM signals for the four T2 trigger types used in the SD. The top row contains the old triggers, Th-T1 (left) and TOT (right). The middle row shows a TOTd trigger on the left and its deconvolution (see equation 4.1) on the right. The bottom trace is an example of a MoPS trigger. See text for a description of the various trigger algorithms. Where applicable, a red line has been drawn to show the TOT and TOTd (0.2 I_{VEM}) threshold and a pink line for the Th-T1 (1.75 I_{VEM}).

Here D_i is the deconvoluted signal in bin i , Δt is the width of one bin (25 ns), τ is the average decay time of a signal (67 ns), and S_i is the measured signal.

The tail of a real signal is approximately an exponential decay, but with added Poissonian fluctuations (approximately $\propto \sqrt{S_i}$). Thus, after the deconvolution is performed, the set $\{D_i\}$ will be a series of pulses where each bin has a random sign but will decrease like $\sqrt{\exp(-\Delta t/\tau)}$, i.e. $\tau \rightarrow 2\tau$. Once the deconvoluted trace, $\{D_i\}$, has been created, the TOT algorithm is applied to it, requiring that at least two PMTs find 13 or more bins with an amplitude greater than 0.2 I_{VEM} within a

window of 120 bins. This results in a trigger rate of 0.3 Hz.

The deconvolution has the added benefit whereby changes in signal that are far from the timescale, τ , will be suppressed making it unlikely that the short spikes from muons will trigger the [TOTd](#) algorithm. An example of a [TOTd](#)-triggered trace along with its deconvolution are shown in the middle row of figure 4.2.

[MoPS](#) is a unique trigger type in that it is completely divorced from the [VEM](#) calibration. The algorithm keeps track of the number of times consecutive bins increase in signal and the amplitude of that increase, j . For such clusters of bins, j must be above a minimum threshold, chosen to be above the average noise, and below that of the average consecutive-bin increase due to a vertical muon, $3 < j \leq 31$. Each of these clusters which pass the upper and lower cuts are then counted towards the multiplicity, m . A [MoPS](#) trigger is accepted when $m > 4$ within a 120 bin window. This also results in a trigger rate of 0.3 Hz.

Both [TOTd](#) and [MoPS](#) are directly promoted to T2 triggers as they accept very little background. Together these two triggers are able to lower the sensitivity of individual stations without sacrificing confidence that the triggers are indeed the result of an [EAS](#)'s secondaries.

4.2 Array-Level Triggers

The calculation of T2 triggers is performed in real time by the stations. With each trigger type having a rate of 0.3–20 Hz, the 1600 [SD](#) stations are constantly sending information to the [CDAS](#) located at the Auger campus. Using this information, the [CDAS](#) then tries to find clusters of T2-triggered stations. This is the lowest array-level trigger, T3.

4.2.1 T3 Triggers

This search for temporal and spatial groupings is done using two types of T3 algorithms. Both of them look for T2-triggered stations that are in a compact configuration. The compactness is calculated by the number of *crowns* over which the T2 stations are spread. A crown is a way of counting hexagons outward from a central station. The first crown, C_1 , around a station consists of the hexagon outlined by that station's six nearest neighbors. The next crown, C_2 , would

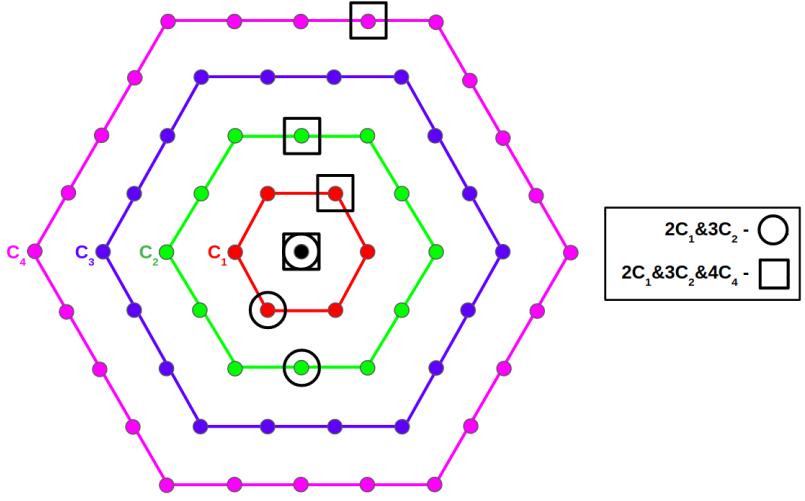


Figure 4.3: The first four crowns, C_1 , C_2 , etc. around the central black station are indicated by the red, green, etc. hexagons. The T3 trigger is always centered on a T2 station (black dot) and requires other nearby T2 stations to be within a certain number of crowns (see text). An example of the stations which would have to trigger for a $TOT2C_1\&3C_2$ event are designated by the black circles. Likewise, an example $2C_1\&3C_2\&4C_4$ event is designated by the black squares.

then be the hexagon concentrically outward from the first crown. This is shown schematically in figure 4.3.

The first T3 trigger is the $TOT2C_1\&3C_2$ which requires that there be at least three stations which pass the **TOT** algorithm. Since the inclusion of the new T2 triggers, this definition has been expanded to include any combination of three **TOT**, **TOTd**, or **MoPS** stations. These three stations must be compact enough that two of them are contained within one crown ($2C_1$) and all three are contained within two crowns ($3C_2$). This is a fairly pure trigger for which 95% of $TOT2C_1\&3C_2$ events will be promoted to the next trigger level. An example of this trigger is schematically shown by the open circles in figure 4.3.

The second type of T3 trigger is the $2C_1\&3C_2\&4C_4$. This requires that there be four stations which registered any type of T2 trigger. As the name would suggest, the compactness requirement is the same as $TOT2C_1\&3C_2$ but with the additional requirement that all four stations must be contained within four crowns. Since the $2C_1\&3C_2\&4C_4$ permits Th-T2 stations, it tends to be triggered by more inclined events where the shower is more muonic at the ground. However, it also accepts

mostly background events and only 5% of $2C_1 \& 3C_2 \& 4C_4$ triggers are promoted to the next trigger level. An example of a $2C_1 \& 3C_2 \& 4C_4$ trigger is shown by the squares in figure 4.3.

T2-triggered stations that pass either the $TOT2C_1 \& 3C_2$ or $2C_1 \& 3C_2 \& 4C_4$ algorithms are then subjected to a temporal requirement. The three (or four) stations that make up the trigger must also have triggered within $(6 + 5 C_n) \mu s$ of the central station (i.e. the station shown in black in figure 4.3).

The successful identification of a T3 trigger is the first time the shower information is permanently recorded. When a T3 trigger is identified, all of the trace information, calibration data, etc. from T1 and/or T2 triggered-stations are requested by the [CDAS](#) and stored on the local servers. Thus, the T3 identification is the final step in the trigger hierarchy where the analysis is completely done in real time.

4.2.2 Physics Triggers

The next level trigger, T4, is used to identify T3-triggered events that are the result of an air shower. The T4 trigger comes in two types, T4-3TOT and T4-4C1, one for each of the T3 triggers. The T4-3TOT algorithm has the following requirements:

- The event includes three stations which have passed any combination of [TOT](#), [TOTd](#), or [MoPS](#) triggers.
- These three stations are in a triangle where two of the legs are one lattice spacing long (i.e. 750 m or 1500 m).
- The difference in start time of the signals, Δt , is less than the distance between the stations, Δr , divided by the speed of light, $\Delta t < \Delta r/c$.

The T4-4C1 algorithm has the requirements:

- The event includes four stations with a T2 trigger.
- These four stations are all contained within one crown centered on one of the stations.
- The difference in start time of the signals, Δt , is less than the distance between the stations, Δr , divided by the speed of light, $\Delta t < \Delta r/c$.

Both of these triggers do not allow for configurations of aligned tanks since the shower axis could not be uniquely determined. The ability of an air shower to pass either of the two T4 criteria effectively sets the energy and zenith angle requirements for full efficiency of the array.

The highest level trigger, T5, demands that the station with the largest signal in the event have 6 active neighbors at the time of the event. Given that the measured signals will be largest near the shower axis, the T5 trigger effectively requires that the shower cores are within one station of the array edge and do not land near any “holes” in the hexagonal lattice. Such a requirement firstly provides a way to determine the effective area of the array (see section 7.2). Secondly, it avoids the poor reconstructions that result when the shower lands outside of the array. In these cases, the shower core location is not well constrained and the shower size estimation can be biased.

In some studies, the T5 trigger is relaxed to help maximize the statistics. In these cases the number of active neighbors is reduced to 5 or 4 stations. For this reason, the designation XT5 is sometimes used to specify the X number of required neighbors. Studies in this thesis have all been performed only on events which passed the 6T5 requirement unless otherwise specified.

4.3 The New Triggers in the 750 m Array

The new triggers increase the SD’s sensitivity to low energy parts of showers by decreasing the size of signals that can be distinguished from background. This also increases Auger’s sensitivity to lower energy showers in general. This increase can be directly seen via the large boost in the event rate after the implementation of [TOTd](#) and [MoPS](#). Table 4.1 gives the trigger rates for the lowest four Auger trigger levels. The T2 rates of [TOTd](#) and [MoPS](#) are relatively low, especially compared to Th-T1 and Th-T2. However, even with this modest increase in the overall station triggering rate, the array-trigger rates improve dramatically. For the TOT2C_{1&3}C₂ and T4-3TOT triggers, the event rate more than doubles when the new triggers are included (from 1300 to 3200 per day). Conversely, the rate of TOT2C_{1&3}C₂ and T4-4C1 types did not seem to be affected by the new triggers at all. This is partly due to these algorithms being sensitive to muon-rich cascades but also because the TOT2C_{1&3}C₂ trigger mostly picks up background. This indicates that

Trigger Rates in the 750 m Array					
	T1/T2 Triggers				
	Th-T1	Th-T2	TOT	TOTd	MoPS
Single Station [Hz]	100	20	0.6	0.3	0.3
T3 Triggers					
	TOT2C ₁ &3C ₂		2C ₁ &3C ₂ &4C ₄		
Old Triggers [min ⁻¹]	1.0		0.4		
New Triggers [min ⁻¹]	2.4		0.5		
T4 Triggers					
	T4-3TOT		T4-4C1		
Old Triggers [min ⁻¹]	0.9		0.02		
New Triggers [min ⁻¹]	2.3		0.02		

Table 4.1: The table above gives the rates of the various Auger trigger levels. The array-level triggers are given before and after the inclusion of the new T2 triggers, TOTd and MoPS. See text for the definition of the trigger types.

the new triggers do not increase the array efficiency at high zenith angle. A more quantitative analysis of the increased efficiency to detect air showers in the 750 m array will be given in section 7.1.

4.3.1 Impact of the New Triggers on Individual Stations

Nearly half of the events that are currently being recorded by the 750 m array would also exist if the new triggers were turned off. It is interesting to see how the new triggers impact these events. Figure 4.4 shows a histogram of the number of stations belonging to an event when considering only old triggers versus including new triggers in the 750 m array. Note that in this plot, the events have already been cleaned to remove coincidentally triggered stations (see section 5.1). The black circles show the average value of the new-trigger multiplicity in each bin. There is an increase of around 1 station at low old-trigger multiplicities, though this increases to 2 or more once the events begin to grow in size. Since most events previously had a station multiplicity of three to five, an additional station is a large relative increase which results in a better constrained shower reconstruction.

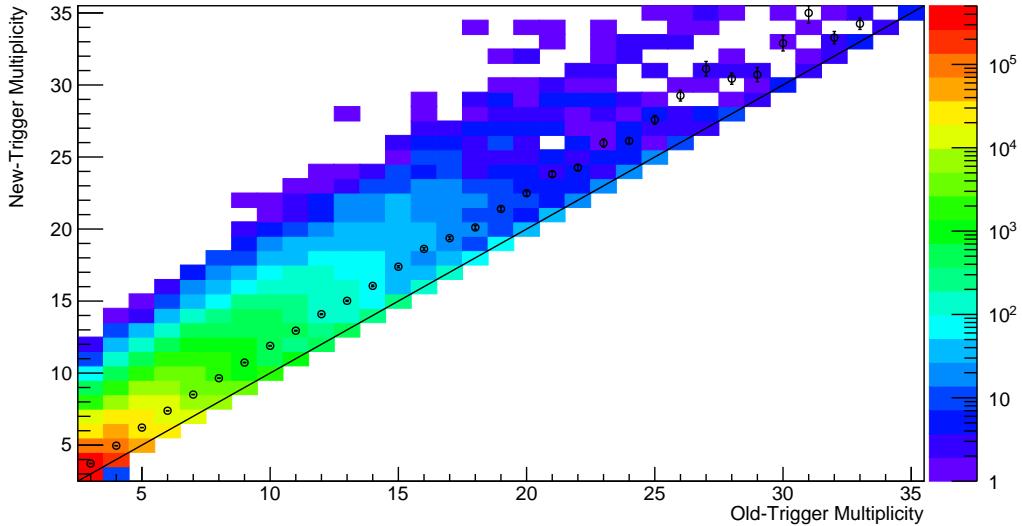


Figure 4.4: This histogram shows increase in multiplicity of events when including the TOTd and MoPS triggers. The solid black line is the 1-to-1 line representing events where the multiplicity did not change. The black circles show the average multiplicity when using the new triggers.

[TOTd](#) and [MoPS](#) were designed to measure smaller signals than was possible with the [Th](#) and [TOT](#) algorithms. Due to the exponentially decreasing signal amplitude as a function of distance from the shower axis, the additional stations added to an event are found on the outskirts of the shower footprint where the signals are small. The distributions of the integrated signal for the new and old triggers are shown in figure 4.5. The histograms shown in this plot give the number of stations which passed any of the old triggers (solid black) or any of the new triggers (solid red) for a given signal size. The exclusive distributions are also shown for stations which passed one of the new triggers and none of the old triggers (dashed red) and vice versa (dashed black). Thus, the histogram representing the additional signals gained by the inclusion of the [TOTd](#) and [MoPS](#) algorithms is given by the dashed red distribution. These typically have a signal from 0.3 to 4 VEM. However, from the solid red distribution it is clear that the new triggers are found up to the highest amplitudes. For signal amplitudes greater than 5 VEM, there are more stations which pass the old triggers than the new triggers. This is due in part to stations triggered mostly by muons which tend to pass only the Th algorithm (see figure 4.7).

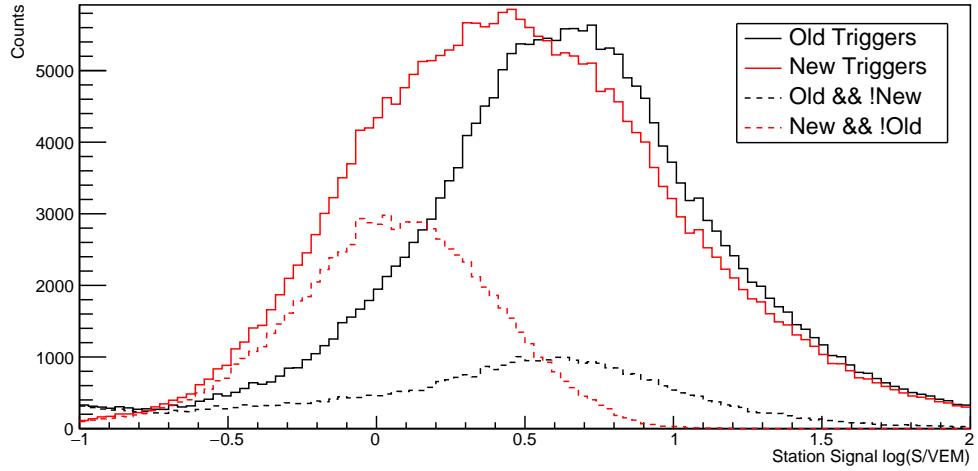


Figure 4.5: The histograms above show the distribution of measured station signals. The solid lines show the number of instances where a station passed at least one of the old (black) or new (red) triggers. The dashed lines show the exclusive distributions where the station passed only one of the trigger groups and not the other.

4.3.2 Station Triggering Efficiency

The distributions shown in figure 4.5 are a convolution of two competing effects. The first is the decrease in signal amplitude as a function of distance from the shower axis. This results in events having a few stations with large signals and many stations with small signals. The second effect is the decreasing T2 triggering efficiency at small signal sizes. It is clear from the distributions of old and new triggered stations that this efficiency has become a bit better with the inclusion of the new triggers.

To characterize the improved efficiency, a study was performed using 3 years of 750 m array events with zenith angles less than 55° , approximately 3.2 million instances. The 750 m array includes eight doublet stations which are auxiliary WCDs that have been installed 10 m from one of the standard stations to be used in diagnostic studies. In these cases, the doublet is treated as a slave and does not contribute to T3's and is not used in the standard reconstruction. Given the 10 m separation, the doublets measure approximately the same part of the shower as their paired station. The triggering probability can then be calculated by noting the T1/T2 status of the doublet stations.

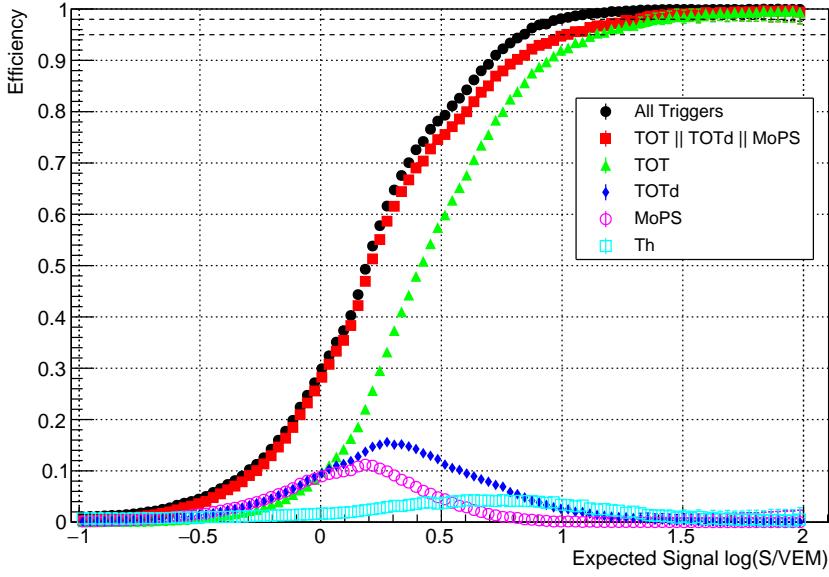


Figure 4.6: The T1/T2 trigger efficiency was calculated for the various trigger types using the doublet stations. The efficiencies shown above are differential such that the **TOT** (green triangles) distribution includes stations which passed at least the **TOT** algorithm, the **TOTd** (blue diamonds) includes stations which were (**TOTd** and not-**TOT**) etc. The overall (black circles) is also shown along with the triggers contributing to the T3 TOT2C₁&3C₂ trigger (red squares).

The events were reconstructed and the expected signal (given by the fit LDF) was calculated for any of the doublet stations for which its counterpart station was triggered. The doublet's triggering status was then recorded from which the efficiency, ε , could be calculated

$$\varepsilon = \frac{N_{\text{Trig}}(S_{\text{exp}})}{N_{\text{Trig}}(S_{\text{exp}}) + N_{\text{NoTrig}}(S_{\text{exp}})} = \frac{N_{\text{Trig}}(S_{\text{exp}})}{N_{\text{Total}}(S_{\text{exp}})}. \quad (4.2)$$

Here N_{Trig} is the number of triggered stations, N_{NoTrig} is the number of non-triggered stations, and N_{Total} is the total number of stations for an expected signal, S_{exp} .

Figure 4.6 shows the triggering probability for the various T1/T2 algorithms as a function of expected signal. The trigger algorithms are listed non-inclusively such that the **TOT** (green triangles) distribution includes stations which passed at least the TOT algorithm, the **TOTd** (blue diamonds) includes stations which were not **TOT** but were at least **TOTd**, etc. Thus the increased efficiency for T3

TOT2C₁&3C₂ triggers can be directly seen by the difference between the TOT and **TOT||TOTd||MoPS** (red squares) distributions. The increased efficiency when including the new triggers is around 10-20% for expected signals between 1 to 3 VEM. The new triggers also lower the 95% efficiency threshold from 10.5 VEM to 6.8 VEM. Note that Th distribution includes both Th-T1 and Th-T2 triggers and largely resembles that distribution of signals which passed the old triggers but not the new triggers in figure 4.5 (dashed black lines).

The overall triggering efficiency is shown by the black circles. This distribution represents the efficiency for all of the showers in the data set. However, there are slight differences in the total efficiency as a function of zenith angle. For more inclined showers, the electromagnetic secondaries will be more strongly attenuated by the time the **EAS** reaches the ground. Thus, there will be a higher muon fraction which are less likely to pass the **TOTd** or **MoPS** algorithms and the overall efficiency will decrease.

An accurate description of the triggering efficiency is important as it is used in the event reconstruction (see section 5.2.2). Motivated by this, the T1/T2 efficiency was parameterized as a function of expected signal and zenith angle. An empirical functional form was chosen to describe the data which is the sum of a Gaussian and a hyperbolic tangent

$$\varepsilon(S_{\text{exp}}; \theta) = N(\theta) \exp\left(-\frac{(\log(S_{\text{exp}}) - \mu(\theta))^2}{\sigma^2}\right) + \frac{1}{1 + \exp\left(-\frac{\log(S_{\text{exp}}) - X_0(\theta)}{\Delta(\theta)}\right)}. \quad (4.3)$$

Here $\log(S_{\text{exp}})$ is the log of the expected signal in the station and N , σ , μ , X_0 , and Δ are free parameters to be fit. Note that since the hyperbolic tangent function is normally bounded between -1 and 1, the second term has already been scaled such that it fits the more natural bounds of 0 and 1.

The fit was performed using a maximum likelihood method. Station triggering is a Bernoulli process and thus the Wilson [77] estimator and confidence interval, $\hat{\varepsilon} \pm \Delta_{\hat{\varepsilon}}$, were used. This estimator and interval are given below

$$\hat{\varepsilon} \pm \Delta_{\hat{\varepsilon}} = \frac{N_{\text{Trig}} + \frac{z^2}{2}}{N_{\text{Total}} + z^2} \pm \frac{z}{N_{\text{Total}} + z^2} \sqrt{\frac{N_{\text{Trig}} N_{\text{NoTrig}}}{N_{\text{Total}}} + \frac{z^4}{4}}. \quad (4.4)$$

Here z is the chosen $1 - \alpha/2$ confidence quantile of a standard normal distribution². This interval is commonly used when there are few samples or when the probabilities are close to 0 or 1. Using this mean and interval the likelihood analysis was performed using a chi-square

$$\chi^2 = \sum_i \left(\frac{\hat{\varepsilon}_i - \varepsilon(S_{\text{exp}}; \theta)}{\Delta_{\hat{\varepsilon},i}} \right)^2. \quad (4.5)$$

The functional forms for $N(\theta)$, $\mu(\theta)$, etc. which produced the best fit per degree of freedom are given below

$$\begin{aligned} \mu(\theta) &= (0.174 \pm 0.039)z + (0.261 \pm 0.002) \\ \sigma(\theta) &= 0.097 \pm 0.007 \\ N(\theta) &= (-0.252 \pm 0.049)z + (0.091 \pm 0.084) \\ X_0(\theta) &= (-0.246 \pm 0.048)z^2 + (0.170 \pm 0.004)z + (0.173 \pm 0.006) \\ \Delta(\theta) &= 0.225 \pm 0.006 \end{aligned} \quad (4.6)$$

where $z = \sin^2 \theta$.

The result of the fit are shown in figure 4.7. The agreement between the model and the data is seen to be quite good and deviations are typically within 2-4%. A only large zenith dependence can be seen at $\simeq 2$ VEM which is a result of the new triggers (see their respective distributions in figure 4.6). At higher zenith angles, the **TOTd** and **MoPS** algorithms are less likely to be satisfied by the increasingly large muonic fraction of the shower front. As will be described in section 5.2.2, the station triggering efficiency is used in the shower reconstruction. For such applications, deviations between the model and data of a few percent are acceptable.

4.3.3 Summary of the New Triggers

This chapter provided an update to the capabilities of the 750 m array when including the new triggers. Due to their relatively weak sensitivity to background muons, the new triggers extend the range of the stations to lower signal sizes. **TOTd** and **MoPS** triggers increase the **SD**'s ability to detect 0.3 to 5 VEM signals

²For example, a confidence interval of $\alpha = 68\%$ corresponds to $z = 1$ and for an $\alpha = 95\%$ interval, $z = 1.96$

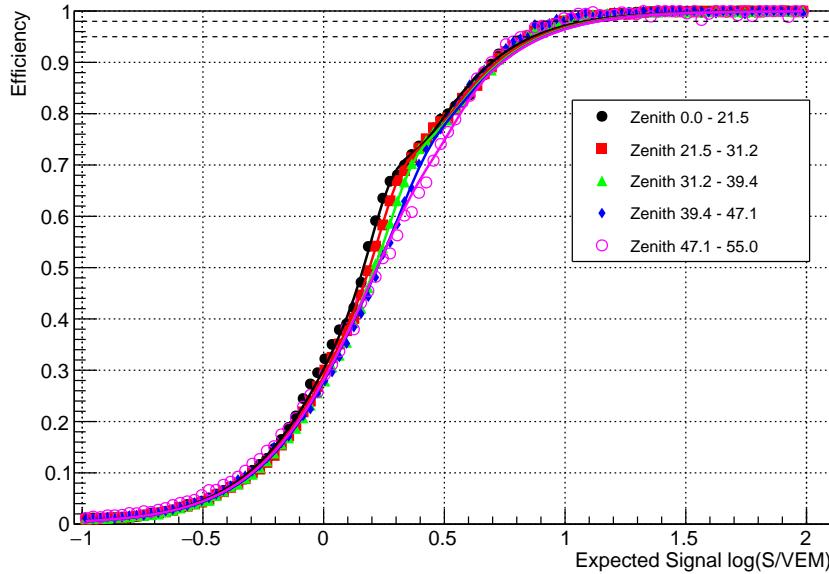


Figure 4.7: The overall T1/T2 triggering efficiency is a function of zenith angle. The plot above shows the efficiency as a function of the expected signal for a number of zenith bins equal in $\sin^2(\theta)$. The solid lines show the result of the fitted model given by equations 4.4 and 4.6.

and improve the station triggering efficiency from 10.5 to 6.8 VEM.

The first benefit of this increased sensitivity to small signals is the additionally triggered stations in the events. On average, observed events gained an additional station though for larger showers, an additional two or three stations was quite common.

Secondly, the **TOTd** and **MoPS** triggers improved the sensitivity of the 750 m array to lower energy showers. After these triggers were implemented, there was a factor of 2.5 increase in the T4 event rate. The extension of the energy region where the array is fully efficient is a primary focus of this thesis as it opens up a new window for Auger to investigate cosmic rays with unprecedented statistics.

Chapter 5

Event Reconstruction Using New Triggers

The [TOTd](#) and [MoPS](#) triggers are not equally sensitive to all the types of secondary particles that make up a hadronic air shower. Because of this, it is imperative that the shower reconstruction chain be reevaluated to ensure the best possible estimation of the [EAS](#) properties. The following chapter details the method by which air showers are reconstructed and various updates to the algorithm for the 750 m array.

This chapter begins with a description of how the events are cleaned to remove coincidentally triggered stations (section [5.1](#)). The selection criteria have been updated for use with the new triggers and are presented in section [5.1.2](#). An overview of the reconstruction algorithm is presented in section [5.2](#). This describes the way that the shower direction (section [5.2.1](#)) and lateral distribution of particles (section [5.2.2](#)) are fit using the T1- and T2-triggered stations. As part of the reconstruction update, the shape of the [LDF](#) was retuned using [TOTd](#) and [MoPS](#) data and is presented in section [5.2.3](#).

5.1 Station Selection with New Triggers

Before any reconstruction is performed on a T4 event, the event is cleaned to remove stations that were triggered by background particles. A station selection is applied to all of the triggered stations in an event to find a configuration which is likely to be due to an air shower. This process is done by comparing the start

times of the signals to the expected start time of a plane shower front moving at the speed of light.

5.1.1 Selection Methodology

The first step in the selection is to get a good estimate of the shower arrival direction. This is done by identifying the *seed* which is comprised of the three stations which pass the following requirements.

1. The stations are in a triangle where at least two of the sides are one lattice spacing long, i.e. 750 m (or 1500 m, depending on the array).
2. All stations are any combination of **TOT**, **TOTd**, **MoPS**, and **Th-T2**.
3. The three stations have the largest summed signal (in units of **VEM**) of any combination which pass requirements 1 and 2.

Note that these conditions are less strict than the T4 requirements. Thus every T4 event is guaranteed to have a seed. The seed becomes the foundation of the event with the motivation that these three stations make up the smallest triangle that encloses the shower's impact point with the ground. Further, the use of the largest signal stations decreases the signal start time uncertainty (see section A.3) which makes the estimation of the shower direction more accurate.

Using the difference in start time between any two seed stations, Δt_{ij} , and their separation, $\Delta \vec{x}_{ij}$, a plane shower front can be fit by analytically¹ solving a system of equations

$$\begin{bmatrix} \hat{n}_x \\ \hat{n}_y \end{bmatrix} = \begin{bmatrix} \Delta x_{12} & \Delta y_{12} \\ \Delta x_{23} & \Delta y_{23} \end{bmatrix}^{-1} \begin{bmatrix} c \Delta t_{12} \\ c \Delta t_{23} \end{bmatrix}. \quad (5.1)$$

where, \hat{n} , is the shower-direction unit vector.

For each station in the event, the difference between the signal start time, t_i , and the expected start time from the fit plane, t_{exp} , are compared, $dt = t_i - t_{\text{exp}}$. Any stations which fall outside of the window, $A < dt < B$, are rejected. This is shown schematically in figure 5.1. The timing window was previously tuned using the old triggers, $A = -500$ ns and $B = 1400$ ns, though an update with the new triggers will be shown in the next section.

¹The approximation is taken that all stations have the same altitude which is true up to a few meters, see figure 5.7.

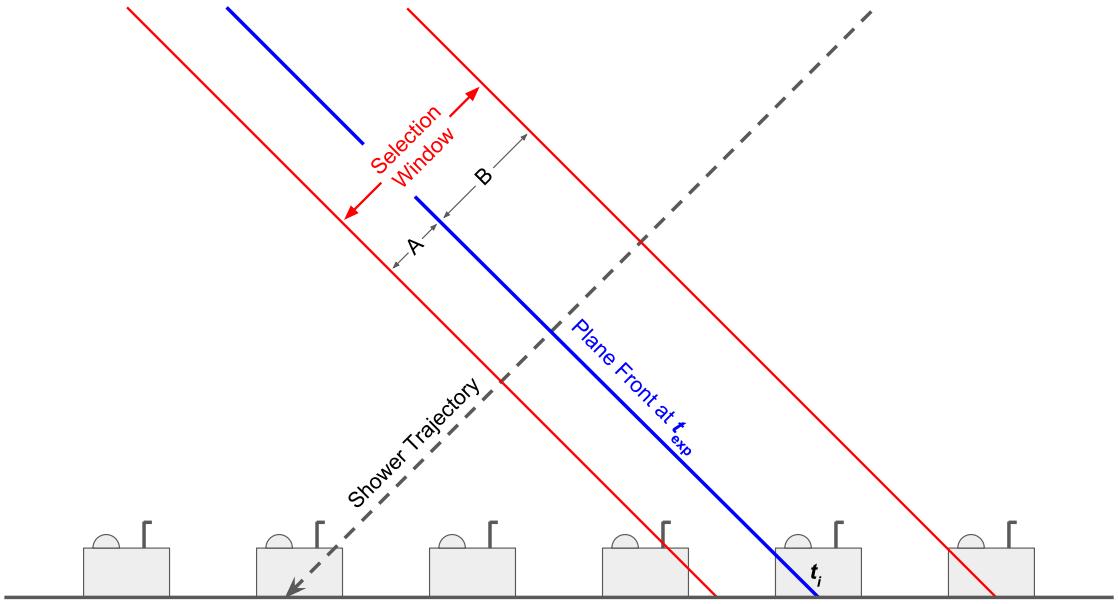


Figure 5.1: The selection is based on a comparison of the expected arrival time, t_{exp} , of a plane shower front and the signal's start-time in each station, t_i . Stations are rejected if the signal begins outside of fixed selection window, indicated by the red lines.

This process of selecting stations by adding them to the seed is known as the bottom-up selection. There is also a second selection method which instead assumes all stations are part of the event and iteratively removes stations which have large timing residuals. This selection is known as the top-down selection and is discussed in section A.2.

5.1.2 Update to the Selection Window

There are a few reasons to suspect that the selection limits (A and B) may not necessarily be appropriate for the new triggers. The preferential sensitivity to the electromagnetic component means that the start times measured by TOTd or MoPS-triggered stations will be delayed with respect to the muon-rich leading edge of the shower front. Further, in large events, stations that pass the new triggers tend to be on the edge of the events. Since the selection does not take into account the shower front's curvature, stations further from the shower axis will inherently appear delayed compared to a plane front. To create a selection that is

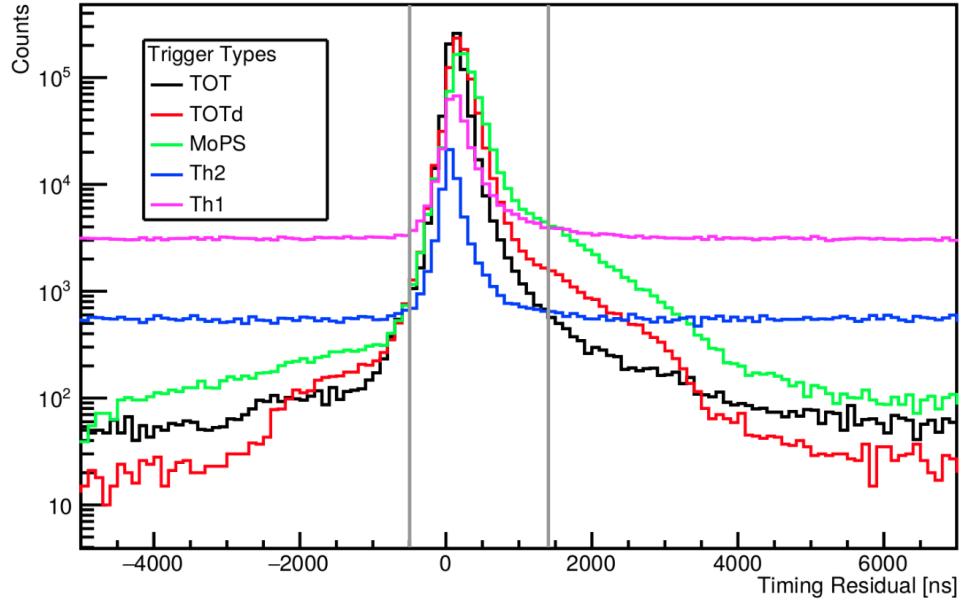


Figure 5.2: The histogram above displays the timing residuals in the stations by trigger type based on the expected times from a plane shower front. The triggers have been organized in a hierarchy (see text). The vertical lines show the current residual limits. Negative values correspond to start times which are earlier than predicted and positive values are late.

consistent for the different types of T1 and T2s, the timing windows were updated independently for each trigger.

The windows were calculated using 1.5 years of new trigger data from the 750 m array. For each event, the shower direction was estimated using the seed and dt was calculated for all of the triggered stations. Since the seed stations inherently have a residual of zero, they were then removed from the data set². This resulted in roughly 14.5 million timing residuals.

Since a single trace can satisfy multiple triggers at once (particularly for large signal amplitudes), the various trigger types were analyzed according to the trigger hierarchy used in section 4.3.2 (**TOT** → **TOTd** → **MoPS** → **Th-T2** → **Th-T1**). A histogram of the timing residuals according to this hierarchy is shown in figure 5.2. For each of the trigger types, there is a clear peak which is caused by the correlated air shower particles. This distribution is both asymmetric and slightly off-center from zero. This is a result of the entries to these histograms being non-seed stations

²This means that only events with a multiplicity of ≥ 4 stations were used in the analysis.

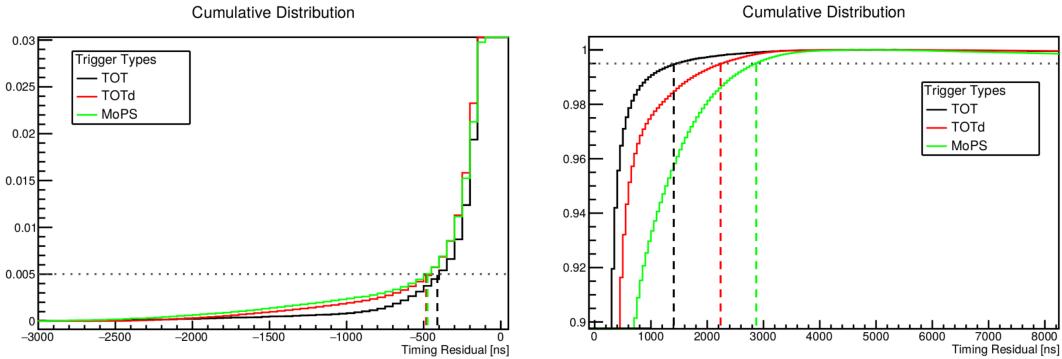


Figure 5.3: The background-subtracted histograms of the timing residuals were used to create the CDFs above. The two plots are the high and low ends from which the selection window values were chosen to select the middle 99% of the distribution. The vertical dashed lines represent the bounds for each trigger type, limit A shown in the left plot and limit B shown in the right plot.

(and thus relatively far from the shower axis) combined with the shower front curvature. The delay of secondary particles in a curved front with respect to the plane fit used in this selection process increases with axial distance and is on the order of ~ 100 ns (see figure 5.8 for an example). Additionally, the shower front becomes thicker far from the axis, increasing the variance of the arrival times (see section A.3). The new triggers also show an additional component at 1000 - 4000 ns which may be due to heavily delayed electromagnetic particles. These extra tails for the TOTd and MoPS triggers (red and green lines) do not fit well within the timing window used for the old triggers (grey vertical line).

The threshold triggers both have very large backgrounds as indicated by the flat, constant offset. This is unsurprising as they are more suited to identify short, muon-like traces. The purity of the triggers can also be directly seen as the ratio of counts above and below this background line.

The determination of the timing window limits was done by first subtracting off the constant background which was calculated on the interval of 8000 - 9000 ns. The background subtracted distributions were then integrated to create a cumulative distribution function (CDF) for each of the trigger types. Finally, the selection window was then defined to be the limits that select the middle 99% of the data, i.e. $CDF(A) = 0.005$ and $CDF(B) = 0.995$. The two plots in figure 5.3 show the upper and lower tails of the CDFs for the TOT, TOTd, and MoPS trigger types along with their respective limits (vertical dashed lines). The values for all five trigger

Trigger Type	<i>A</i> [ns]	<i>B</i> [ns]
TOT	−397	1454
TOTd	−468	2285
MoPS	−477	2883
Th2	−485	1379
Th1	−651	2360
Previous	−500	1400

Table 5.1: The updated selection windows for each trigger type are listed above along with the previous values.

types are shown in table 5.1. While the lower limits generally did not change much, the previous upper limit of $B = 1400$ ns was only selecting 96%-98% of the new triggers. The recalculated upper limits for both of TOTd and MoPS are much larger (~ 100 's of ns) which was expected given the residual plot in figure 5.2. The values of A and B did not have a significant change for TOT and Th-T2 as these are both old triggers.

5.1.3 Additional Selections

After the time window selection is applied to an event, two final cuts are used to remove any further coincidentally triggered stations. Given that the stations are spread quite far apart, there should not be large gaps in the lateral distribution of particles on the order of hundreds of meters. Thus, the first cut is based on the assumption that triggered stations should be roughly contiguous. This cut removes the so-called *lonely* stations from an event which is defined as having less than one triggered station within 900 m (1800 m) and less than two triggered stations within 2500 m (5000 m) for the 750 m (1500 m) arrays. This essentially only allows for *pairs* of non-contiguous stations. An example of the types of configurations that would pass this cut is shown in figure 5.4.

The second cut looks for stations which have been triggered by lightning. The occasional lightning discharges near the array create strongly oscillating traces which are able to pass the T1/T2 triggers. An example of such a PMT trace with the characteristic 2.5 - 3 MHz oscillation is shown in figure 5.5. Thus, lightning candidates are identified by applying a Fourier Transform to the trace. Due to

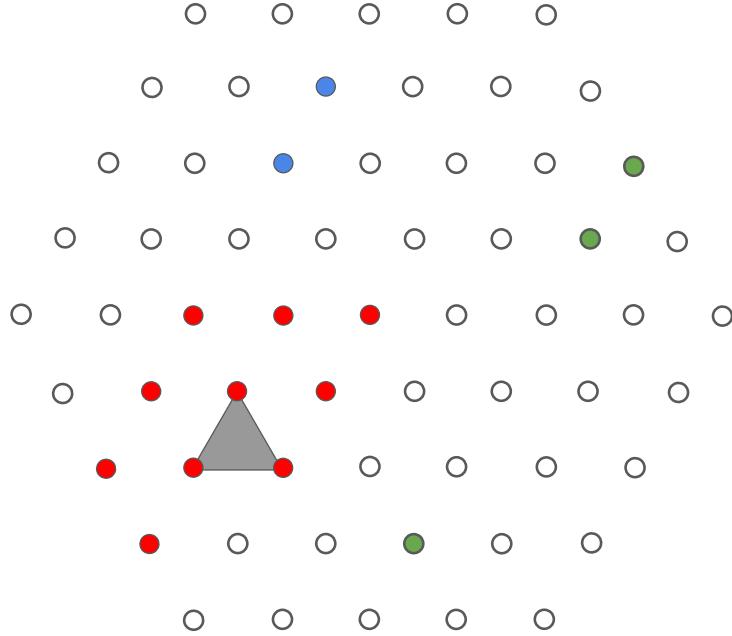


Figure 5.4: The diagram above is an example of an event with isolated stations. The seed (grey triangle) along with the contiguous stations are shown in red. The isolated group in blue would pass the lonely station criteria while all the green stations would not and would be removed from the event.

the size of the array, not every station in an event will be affected and thus it is sufficient to simply remove the individual problematic stations as necessary.

Following the removal of lightning-triggered stations, one final check is performed to ensure that the event still passes the T4 requirements (see section 4.2.2). If the requirements are still met, the remaining stations are then used to reconstruct the air shower.

5.2 The Air Shower Reconstruction

The [EAS](#) reconstruction algorithm is based on a simplified model of an air shower which macroscopically describes the location and density of the secondary particles. The model approximates the secondaries as arriving in a thin, curved shower front moving at the speed of light towards the ground. This is shown schematically in figure 5.6. The centroid of the front defines the shower axis which moves in the $-\hat{n}$

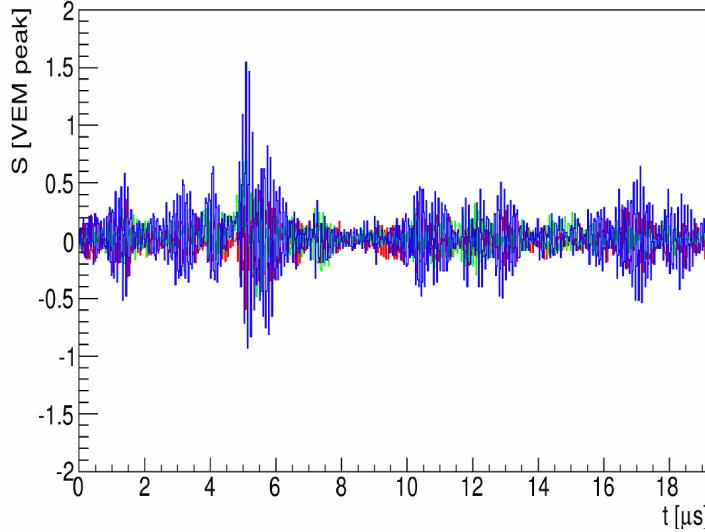


Figure 5.5: The traces from three PMTs are shown that were recorded as a result of nearby lighting. The large positive *and* negative fluctuations are distinct from the types of signals seen in standard cosmic ray traces. Image from [78].

direction and intersects the ground at the shower core, \vec{x}_c , at time, T_0

$$\vec{x}_{\text{centroid}} = \vec{x}_c - c(t - T_0)\hat{n}. \quad (5.2)$$

The curvature of the front is approximated as being spherically shaped with radius, R . The overall size of the shower, which describes the universal scaling with energy (see chapter 2), is estimated by the expected signal that would be measured by a station at a reference distance of 450 m, $S(450)$.

This model is split into two parts, one for the geometric construction which is found using the timing information, and one for the lateral distribution which depends mostly on the signal sizes. These two parts are fit to the data by maximizing a likelihood function which gives the probability of the shower model parameters \hat{n} , T_0 , R , \vec{x}_c , and $S(450)$ given the signal sizes and start times of the stations. This process is done using a numerical extrema finder, MINUIT [79].

The following sections describe the way that a shower's geometry (section 5.2.1) and lateral distribution (section 5.2.2) are fit to the model including their respective likelihood functions.

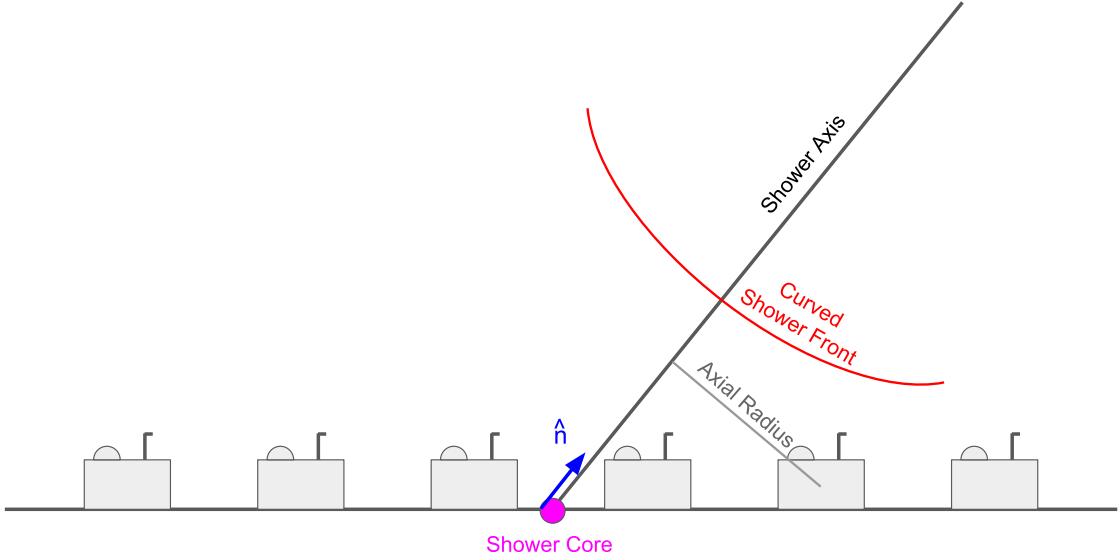


Figure 5.6: The shower model used in the reconstruction describes the secondary particles as being distributed along a spherically-curved shower front. The motion of the center of the front defines the shower axis which intersects the ground at the shower core. The particles in the shower create signals in the stations which decrease exponentially with increasing axial radius.

5.2.1 Geometric Reconstruction

The geometric reconstruction uses the timing information from the stations to identify the most likely \hat{n} , T_0 , and R . Because each triggered station in a T4 event contributes one degree of freedom to the geometric fit, the four parameters can be under-constrained for small, three-station events. For events with fewer than 5 triggered stations, the shower front's curvature is ignored and the expected times are calculated from a plane front, similar to what was done for the station selection. Since these are the only free parameters during the geometric fit, a good estimation of \vec{x}_c and $S(450)$ is needed to calculate the likelihood and to provide good initial guesses for the numerical fitting routine.

The shower core location and time are first estimated using a weighted barycenter of the selected stations located at \vec{x}_i with start time t_i

$$\vec{x}_{c,\text{est}} = \frac{\sum \sqrt{S_i} \vec{x}_i}{\sum \sqrt{S_i}}, \quad T_{0,\text{est}} = \frac{\sum \sqrt{S_i} t_i}{\sum \sqrt{S_i}}. \quad (5.3)$$

The choice of weight being the square root of the station signal, S_i , has been found

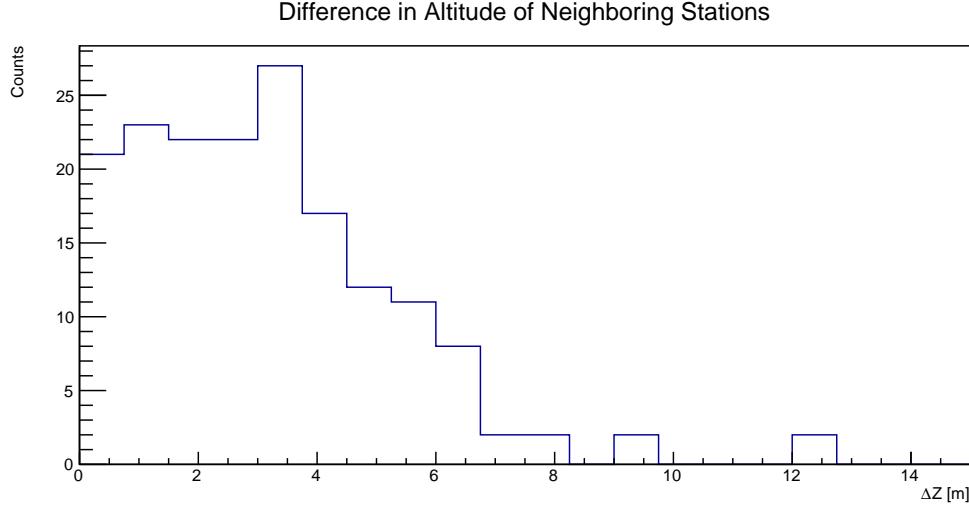


Figure 5.7: The 750 m array is relatively flat. The histogram above shows the absolute value of the difference in altitude between nearest-neighbor stations. These small changes in altitude mean that a line between any two stations is within one degree of being level.

to result in an estimation which had the best agreement with the reconstructed values. Auger’s location was well chosen such that any two stations are essentially at the same altitude. Figure 5.7 shows the change in altitude between neighboring stations which are typically on the order of a few meters. Such small deviations allow the core’s altitude to be fixed, $z_c \equiv z_{c,\text{est}}$, which reduces the model’s parameter space without sacrificing accuracy.

The shower direction is estimated by fitting a plane to the seed stations. This process is identical to what was done during the station selection, equation 5.1.

According to the shower model, the station signals should begin just as the shower front passes through the water volume. In the fit, the station locations³ are assumed to be exactly known and the observed start-times, t_i , are approximated as Gaussian fluctuations around the front-impact time, t_{exp} . This assumption produces the geometric likelihood function given below

$$-\ln \mathcal{L}_{\text{Geom}} = \sum_i^{\text{Stations}} \frac{[t_i - t_{\text{exp}}(\vec{x}_i)]^2}{2\sigma_{t,i}^2}. \quad (5.4)$$

³For clarity, all positions will be given with respect to the shower core $\vec{x} \equiv \vec{x} - \vec{x}_c$ and all axial distances will be written as $r \equiv |\hat{n} \times \vec{x}|$ for the remainder of the chapter.

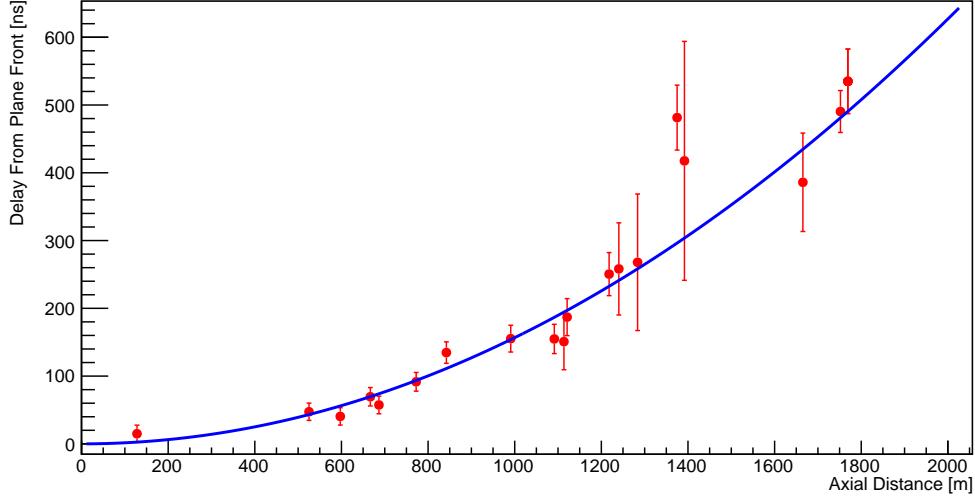


Figure 5.8: The geometric fit is done by fitting a curved shower front using the signal timing information. The signal start times (red points) are shown with respect to the fitted plane front to highlight the curvature. The expected time is represented by the blue line which has a radius of curvature of $R = 10.7$ km. This shower had an energy of 3 EeV and a zenith angle of 37° .

Here the sum is over all of the triggered stations that passed the selection. The start-time variance, σ_t^2 , has been parameterized using the doublet stations and is dependent on the signal size and the shower's zenith angle, see section A.3.

In general, the expected start times in the stations are given by

$$t_{\text{exp}}(\vec{x}_i) = T_0 - \frac{\hat{n} \cdot \vec{x}_i}{c} + \frac{r_i^2}{2cR}. \quad (5.5)$$

The first two terms describe the impact time of a plane shower front. The last term is a further correction for the delay due to the spherical curvature. For a spherical front, this additional delay is $\frac{1}{c}(R - \sqrt{R^2 - r_i^2})$. However, to increase the linearity of the model, the first order expansion in $(r/R)^2$ is used⁴ to produce equation 5.5. As previously described, for showers with fewer than 5 triggered stations, the shower curvature is ignored ($R \rightarrow \infty$).

An example of the results from the geometric fit is shown in figure 5.8. The delay is shown with respect to a plane shower front such that the curvature of

⁴The typical shower front radius of curvature for an event observed by the 750 m array is approximately 6 to 12 km and $r/R \sim 1/5$.

the shower can be seen. The radius of curvature for this shower is 10 km which is typical given its energy, 3 EeV.

5.2.2 Lateral Distribution Reconstruction

Using the geometric results from the previous section, the second step in the reconstruction is the fit of the lateral distribution of measured signals. The decreasing signal, $S(r)$, as a function of axial distance is generically described by

$$S(r) = S(450) f_{\text{LDF}}(r). \quad (5.6)$$

Here, f_{LDF} is normalized such that $f_{\text{LDF}}(450 \text{ m}) = 1$. Within Auger, there are two commonly used functional forms of $f_{\text{LDF}}(r)$, an [NKG](#)-inspired (see equation 2.13) function and a log-log parabola

$$f_{\text{LDF}}^{\text{NKG}}(r) = \left(\frac{r}{r_{\text{ref}}}\right)^{\beta} \left(\frac{r + r_0}{r_{\text{ref}} + r_0}\right)^{\beta+\gamma} \quad (5.7)$$

$$f_{\text{LDF}}^{\text{LLP}}(r) = \begin{cases} \exp[\beta \rho + \gamma \rho^2], & r > r_{\text{cut}} \\ \exp[\beta \rho + \gamma (2\rho - \rho_{\text{cut}}) \rho_{\text{cut}}], & r \leq r_{\text{cut}} \end{cases}. \quad (5.8)$$

The [NKG](#)-like function includes a normalization length, $r_0 = 700 \text{ m}$, and the reference distance, $r_{\text{ref}} = 750 \text{ m}$. The log-log parabola is a function of $\rho = \ln\left(\frac{r}{r_{\text{ref}}}\right)$ and is split into two parts. Below $r_{\text{cut}} = 150 \text{ m}$, a log-linear extension is used, where $\rho_{\text{cut}} = \ln\left(\frac{r_{\text{cut}}}{r_{\text{ref}}}\right)$, which better describes the shape of the signals at small axial distances. For both functions, the slope is controlled by β and γ . Note that in this work, only $f_{\text{LDF}}^{\text{LLP}}(r)$ will be used⁵.

During the reconstruction, the two slope parameters are not fit as there are often not enough signal measurements to constrain the [LDF](#) shape well⁶. Instead, β and γ are kept fixed to a parameterization based on Auger events with high multiplicities. This has been done for the new triggers using the $f_{\text{LDF}}^{\text{LLP}}(r)$ shape and is presented in section 5.2.3.

⁵It has been shown [80] that the exact functional form describing the [LDF](#) shape does not have a strong impact on the shower size estimation given that a suitable reference distance (450 m in this case) is used.

⁶If a shower only triggers three stations which are all the same distance from the shower axis, the [LDF](#) slope is under-constrained since the three stations effectively provide only one measurement of the lateral distribution of signal.

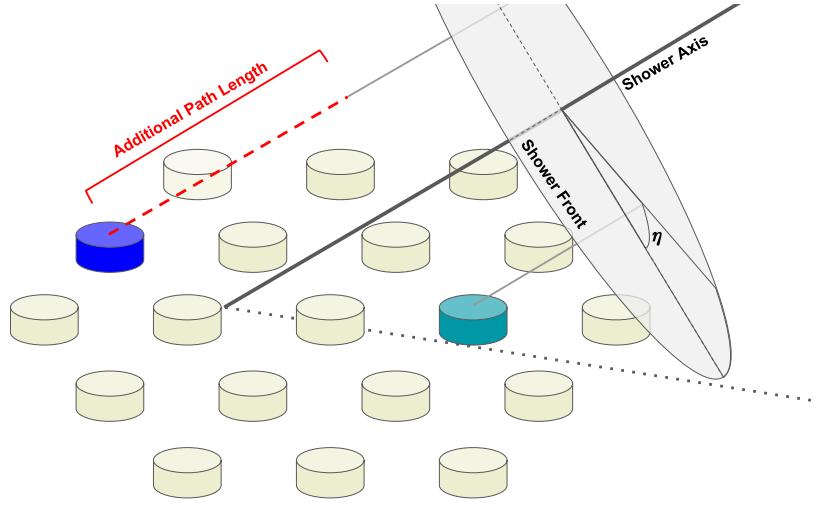


Figure 5.9: The schematic above highlights the difference in path length traveled for two stations (green and blue) which are at the same distance from the shower axis (solid black line). The shower front (grey circle) will first arrive at the green station and must travel an additional distance (red dashed line) before reaching the blue station.

Both of the Auger LDFs are cylindrically symmetric about the shower axis. In fact, this is not generally the case for showers that are observed by surface arrays. This is a result of the inconsistent shower age at which the secondary particles are measured for non-vertical trajectories, shown schematically in figure 5.9. As the shower front approaches the ground, stations that are under the shower axis ($\eta \simeq 0^\circ$) will trigger before stations above the shower axis ($\eta \simeq 180^\circ$). So in the figure, the green station will have triggered before the blue station. The shower will continue to attenuate as it traverses the additional length (represented by the red dashed line) and will result in a smaller measured signal even if the two stations are at the same radial distance. This additional path length can be several electromagnetic shower lengths (see section 2.1) for very inclined trajectories.

During the reconstruction, the cylindrical symmetry is restored by scaling the measured signals. The scaling amplitude, $\alpha(r_i, \theta)$, has been previously parameterized using simulations

$$S_i \rightarrow \frac{S_i}{1 + \alpha(r_i, \theta) \cos(\eta_i)}. \quad (5.9)$$

Here η_i is the cylindrical azimuthal angle about the shower axis of the i^{th} station.

The reconstruction aims to maximize the likelihood of measuring the observed

set of azimuthally-corrected signals, $\{S\}$, given the shower parameters \vec{x}_c and $S(450)$

$$\mathcal{L}_{\text{LDF}}(\{S\}|\vec{x}_c, S(450)) = \prod_k P(S_k|S(r_k)). \quad (5.10)$$

When a T3 event is identified, only trace information from stations which triggered is recorded. For all other stations, no signal is available. However, the fact that a station did not trigger can still inform the fit of the shower model. The binary distinction between triggered and untriggered stations naturally splits the likelihood function into two respective terms

$$\mathcal{L}_{\text{LDF}}(\{S\}|\vec{x}_c, S(450)) = \prod_i^{\text{Trig}} P(S_i|S(r_i)) \prod_j^{\text{Not-Trig}} P(S_j = 0|S(r_j)). \quad (5.11)$$

The first factor corresponds to the triggered stations for which the signal fluctuations are roughly Gaussian about the expected value, $S(r_i)$, with a variance, σ_i^2 . This variance has been found using Auger doublet data [81]

$$\sigma_i = \sqrt{S(r_i)} \times \begin{cases} 0.10, & \text{Saturated PMT and } S_i < 2000 \text{ VEM} \\ 1.06, & \text{Saturated PMT and } S_i \geq 2000 \text{ VEM} \\ 0.34 + 0.46 \sec \theta, & \text{else} \end{cases}. \quad (5.12)$$

Note that the variance is treated differently for saturated stations. When a large amount of signal (≥ 1000 VEM) is deposited in the stations, the voltage output of the [PMTs](#) can overflow the range of the digitizing electronics and the trace becomes truncated. In such cases, the true signal size and errors can no longer be determined in the usual manner.

The second term in equation 5.11 describes the probability for a station to not have triggered given the expected signal, $S(r_i)$. The complement of this quantity was calculated in section 4.3.2, $\varepsilon(S(r_i), \theta)$. Thus the probability to *not* trigger is given by $1 - \varepsilon(S(r_i), \theta)$ and the log-likelihood function is

$$-\ln \mathcal{L}_{\text{LDF}} = \sum_i^{\text{Trig}} \frac{[S_i - S(r_i)]^2}{2\sigma^2} - \sum_i^{\text{Not-Trig}} \ln [1 - \varepsilon(S(r_i), \theta)]. \quad (5.13)$$

The best fit parameters using the shower model are found by minimizing this log-likelihood function via MINUIT.

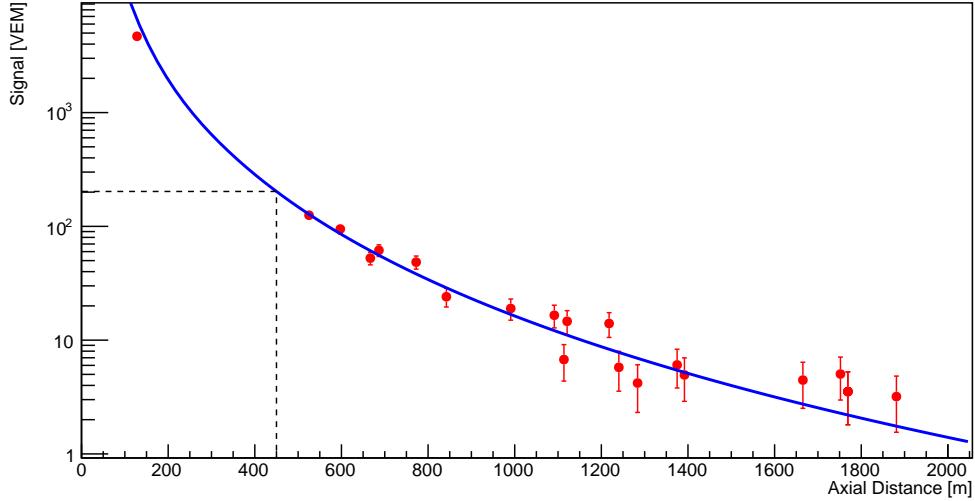


Figure 5.10: The lateral distribution is fit using a log-log parabola. An example of the reconstructed LDF is shown above for a 3 EeV shower with a zenith angle of 37° . The measured signal values are shown in red and the LDF is shown as a blue line. The dashed black line shows the reference signal, $S(450) = 202$ VEM.

An example of a fitted LDF is shown in figure 5.10. This 3 EeV shower (corresponding to figure 5.8) was found to have a reference signal of 202 VEM (shown by the black dashed line). The signals in the triggered stations are shown in red and the expected signal, $S(r)$, is shown in blue. The station closest to the axis was saturated (this is common for signals above ~ 1000 VEM).

5.2.3 Update of the LDF Shape

The LDF functions used in Auger (equations 5.7 and 5.8) were chosen empirically due to a lack of knowledge about the hadronic interactions in the ultra-high energy range. Their slope is controlled by two parameters β and γ which are left fixed to a parameterization during the shower reconstruction. This section includes an update to the parameterization using new triggers.

5.2.3.1 The LDF Tuning Data Set

For this study, an ideal set of events from the 750 m array were used for which the slope could be determined with minimal bias. These events were identified using

the following cuts.

- Number of stations ≥ 4 .
- Statistical uncertainty of the core location was $< 75\text{ m}$.
- The reduced $\chi^2 < 7$ for the LDF fit.
- None of the stations signals were saturated.
- Closest station to the reconstructed axis was $> 150\text{ m}$.
- $S(450) > 50\text{ VEM}$.

The first three cuts provide some assurance that there were enough data points to constrain the LDF and that the fit was good. The fourth and fifth requirements avoid tuning the LDF shape using signals that are larger than the stations are able to accurately measure. The final cut is used to minimize the impact of upward fluctuations in the station signals.

The WCDs are not particle counters, rather, they measure the number of Cherenkov photons that are produced by the secondary particles in the water. This process is subject to Poissonian fluctuations which smear the measurement around an average value. For signal sizes near the station-triggering threshold, upward fluctuations in the number of Cherenkov photons (or photo electrons in the PMT) may pass the T1/T2 trigger whereas downward fluctuations will not. This results in an excess of larger-than-average signals being measured near the triggering threshold. Thus, the restriction to large-sized events⁷ reduces the contribution of this triggering artifact.

Using 3.5 years of new trigger data from the 750 m array, these cuts resulted in about 7500 events and 85,000 stations.

5.2.3.2 Parameterization of the LDF Slope

To determine a good choice of variables by which to parameterize β and γ , it is important to consider the way showers are observed by the WCDs. The WCDs are neither particle counters nor are they calorimeters for muons. This results in

⁷For scale, 50 VEM is roughly equal to $10^{17.7}\text{ eV}$, relatively high energy for the 750 m array.

the measured lateral distribution of signal having a bias depending on the muonic fraction in the cascade.

As described in chapter 2, the ratio of electromagnetic to muonic secondaries changes during the shower development and is energy and mass dependent. To account for these effects, the slope parameters are allowed to vary with shower size, $S(450)$, zenith angle, θ , and the combination of the two. The parameterizations used are given below

$$\begin{aligned}\beta &= (b_0 + b_1x + b_2x^2)(1 + b_3y) \\ \gamma &= g_0 + g_1y \\ x &= \sec \theta - 1.27 \\ y &= \ln S(450) - 5\end{aligned}. \quad (5.14)$$

The x and y variables describe the zenith dependence and log-size and are normalized to the center of the viewing range (35°) and the average shower size observed above 50 VEM, respectively.

Finding the optimal slope parameters thus consists of finding the values of $\{b_i, g_i\}$ which best describe the data. This was done via a likelihood analysis. For different values of $\{b_i, g_i\}$, all of the events which passed the cuts were reconstructed using the LDF with fixed slope as defined by equation 5.14. The quality of the LDF fits was assessed using a chi-square

$$\chi^2 = \frac{1}{N_{\text{stn}}} \sum_{\text{Events}} \sum_{S_i > 5 \text{ VEM}}^{\text{Stations}} \frac{(S(r_i) - S_i)^2}{\sigma_i^2}. \quad (5.15)$$

This quantity gives the square of the residuals from all stations in all events divided by the total number of stations, N_{stn} . The best $\{b_i, g_i\}$ were taken to be the values which produced the smallest residuals over the entire data set. These are listed in table 5.2.

Figure 5.11 shows a comparison of the measured lateral distribution of signals and the expected values given by the LDF for zenith angles $15^\circ < \theta < 45^\circ$ and shower sizes $50 < S(450) < 60$ VEM as an example. The top plot gives the signals in VEM while the bottom plot shows the residuals. Both of these plots show good agreement up to about 1200 m at which point the expected signal consistently underestimates the measured value. An explanation of this can be seen in the top plot where the measured values reach a signal floor of 2 VEM and become constant

Parameter	Value	Error
b_0	2.95	0.02
b_1	−1.04	0.47
b_2	0.673	0.052
b_3	0.018	0.027
g_0	0.260	0.047
g_1	−0.022	0.010

Table 5.2: The values above are the best fit values for the parameterization of the LDF shape parameters, β and γ . The set $\{b_i, g_i\}$ describe the way the shape of the LDF changes with reference signal and zenith angle, see equation 5.14.

for all distances. This value roughly corresponds to the signal at which the T1/T2 triggers are 50% efficient (see figure 4.7).

The expected changes in the lateral distribution of particles for showers of differing ages motivated the choice of a zenith and size dependency on the slope parameters. Additional checks were performed to ascertain how well these dependencies were accounted for. The residuals of the LDF fit are shown as a function of these two parameters in figure 5.12. Both show the fractional residuals of all stations within 1200 m of the shower axis from showers with reference signals greater than 50 VEM and zenith angles less than 55° . The top plot is the residual distribution as a function of shower size. The residuals for large shower sizes are consistent with zero though there is an increasing trend for showers below 100 VEM. This is unsurprising and should be the case as smaller showers will contain more stations with upward fluctuations within 1200 m. The bottom plot shows the distribution as a function of zenith angle in $\sin^2 \theta$ bins. Here the residuals deviate from zero only by a few percent.

Most importantly, the LDF does not show any bias at 450 m where the shower size is estimated. However, for smaller sized showers, the constant 2 VEM floor due to upward fluctuations will occur closer to the shower axis. The approach of the 2 VEM floor with decreasing shower size is shown in figure 5.13. This plot again shows the residuals as a function of axial distance but for five different integral cuts in shower size. The decreasing distance at which the residuals diverge from zero with decreasing shower size is clear. The smallest shower size bin ($10 \text{ VEM} \simeq 10^{17.1} \text{ eV}$) begins to show the effects of the upward fluctuations around 700 m from the axis. Further, there is a slight but steady trend of the residuals to deviate

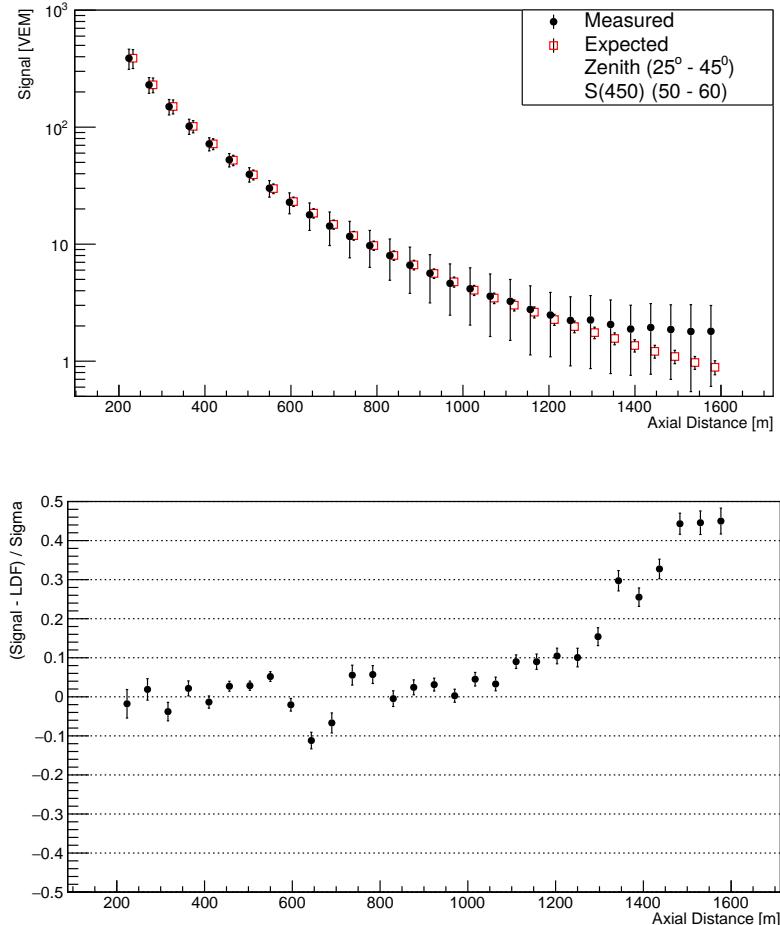


Figure 5.11: The plots above show how well the parameterized LDF shape fits to the measured data with zenith angles $15^\circ < \theta < 45^\circ$ and shower sizes $50 < S(450) < 60$ VEM as an example. Top: The measured signal (black dots) and expected signal (red squares) from the LDF as a function of axial distance. The error bars are the standard deviation of each bin. Bottom: The residuals of the expected and measured signals divided by the signal uncertainty. The error bars indicate the uncertainty of the mean. Both plots show good agreement up to 1200 m where trigger biases begin to dominate, see text.

upward from zero at 450 m as the shower size cut is decreased. This results in an inability to estimate the $S(450)$ parameter for arbitrarily small shower sizes in an unbiased way which will later be addressed in section 6.1.

One final analysis was performed to check the compatibility of the updated LDF shape for the reconstruction with old triggers. Again all of the events in the

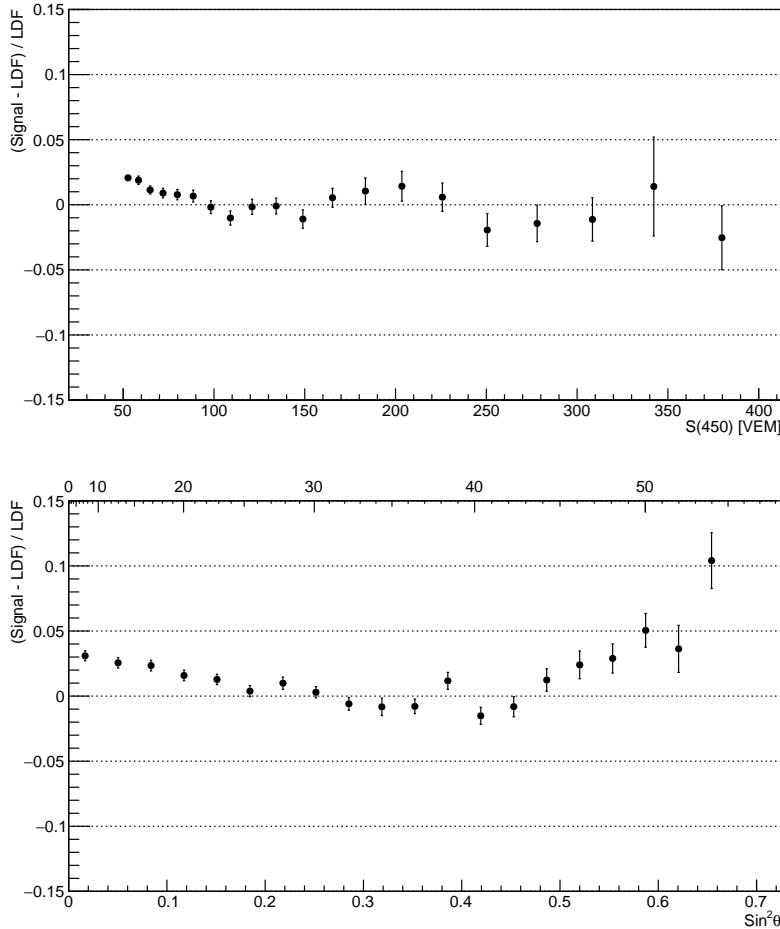


Figure 5.12: The fractional residuals are shown for shower sizes > 50 VEM and zenith angles $< 55^\circ$. All stations within 1200 m of the shower axis are included in these plots. The residuals are shown as a function of shower size (top) and zenith angle (bottom).

study were reconstructed but without the new triggers being used (that is, stations that passed only the new triggers were removed in the station selection process). The residuals are shown in the bottom plot of figure 5.13. New trigger residuals are shown by the black circles and old trigger residuals are shown by the red triangles. For distances close to the shower axis, there is total agreement between the two trigger sets. However, the tail, caused by the dominance of upward fluctuations, begins a bit earlier for the old triggers, around 1000 m from the shower axis. This is a direct result of the increased signal floor for the old triggers being at 4-5 VEM instead of 2 VEM. Ultimately, this plot indicates that the updated LDF shape could

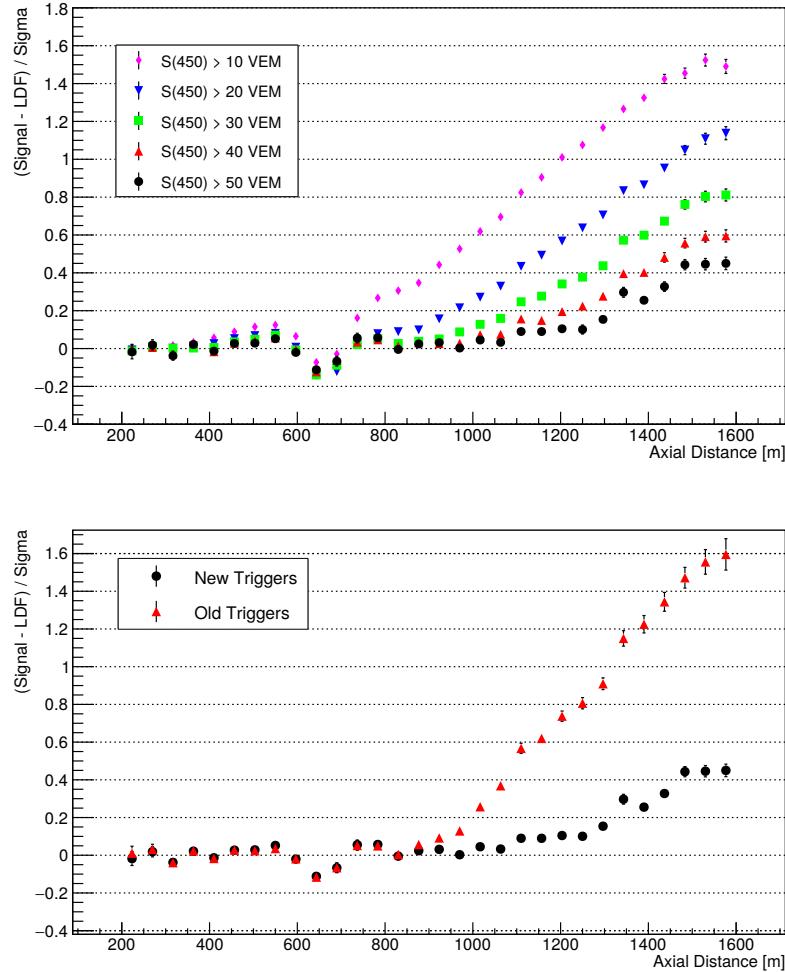


Figure 5.13: The fractional residuals are shown for zenith angles $< 55^\circ$. All stations within 1200 m of the shower axis are included in these plots. The top plot includes the residuals for various cuts of $S(450)$ to emphasize the decreasing distance at which the upward fluctuations dominate for small showers. The bottom plot shows a comparison of the residuals for the new (black circles) and old (red triangles) triggers.

also be used for the old triggers which was originally tuned using a much smaller data set.

Model Parameter	Typical Uncertainty	
	Statistical	Systematic
Core Location	40 m - 70 m	40 m
Shower Direction	1° - 2°	1° - 5°
Shower Size	5% - 7%	10% - 25%

Table 5.3: The typical resolutions for the shower model parameters are given above.

5.2.4 The Global Reconstruction

The choice of splitting the shower reconstruction into two parts was motivated by the weak coupling between the geometric and lateral distribution fits. The former is largely dependent on the timing while the latter is mostly dependent on the signal sizes. However, these two fits cannot be disentangled as any calculation of an axial distance requires both \vec{x}_c and \hat{n} .

To ensure the best reconstruction one final, global fit is performed during which all the shower model parameters are allowed to vary⁸. The global reconstruction is done via a combination of $\mathcal{L}_{\text{Geom}}$ (equation 5.4) and \mathcal{L}_{LDF} (equation 5.13)

$$-\ln \mathcal{L}_{\text{Global}} = -\ln \mathcal{L}_{\text{Geom}} - \ln \mathcal{L}_{\text{LDF}}. \quad (5.16)$$

The shower parameters which minimize this combined likelihood function are taken to be the final estimators of the shower.

The typical uncertainties of the reconstruction are listed in table 5.3. Events with a larger station multiplicity will normally result in a better resolution. The uncertainty in the shower size given here is simply from the statistical errors on $S(450)$. This value is corrected to remove a number of biases (see chapter 6) and thus these values do not represent the complete size uncertainty. For a more detailed discussion on the shower size and energy resolution, see section 6.4.3.

The final step in the reconstruction is the conversion between the observed shower size and the shower energy. This is a multi-part process involving removing a number of systematic biases and a calibration with the FD. These steps are described in the following chapter.

⁸For events where the geometry was previously fit to a plane, this is still the case for the global fit.

Chapter 6

Air Shower Energy Assignment

The [SD](#) stations do not allow for a direct measurement of an air shower's energy since they only sample the shower at one point in the development. Instead, the reconstruction produces an energy estimator based on a shower's size, $S(450)$, for the 750 m array. While this quantity is highly correlated with air shower energy, there are a number of higher order effects that cause $S(450)$ to be biased. Some of these effects, such as the statistical fluctuations in the cascade development (so called shower-to-shower fluctuations), cannot be accounted for on an event-by-event basis. Others are more pathological and are generally related to the inconsistent age at which showers are observed by a ground-based array.

In this chapter, a number of these higher order effects will be described along with the methods used to unbias the size estimator. First, an ad hoc correction is introduced which is required to remove an observed bias in the [LDF](#) fit at 450 m for small shower sizes (section 6.1). The second correction accounts for the changing amount of grammage between the [SD](#) and the top of the atmosphere due to the constantly evolving atmospheric conditions. A method to rectify this effect based on local weather measurements is described in section 6.2. The final correction is geometric in nature and accounts for the increasing slant depth of the array with increasing zenith angle leading to a more attenuated shower on the ground. This method relates the observed shower size to the one that would have been measured if the shower instead arrived at a chosen reference zenith angle (section 6.3). Finally, in section 6.4, the cross calibration between the corrected shower size estimators and the nearly calorimetric [FD](#) energies is presented.

6.1 The LDF Bias at 450 m

In section 5.2.3, the shape of the LDF was updated using a relatively high energy ($S(450) > 50 \text{ VEM} \simeq 10^{17.8} \text{ eV}$) subset of the Auger events seen by the 750 m array. The tuned shape, $\text{LDF}(r)$, was shown to agree with data at the reference distance of 450 m to within a few percent (see figure 5.13, top). However, the agreement of the LDF and the measured signal at 450 m is not expected to be as good for all energies. Particularly for small shower sizes, observed events will begin to be more strongly influenced by station and array-level triggering biases.

For example, the influence of upward fluctuations, which was previously discussed in section 5.2.3, causes a systematic bias in the measured station signal. It was shown how the upward fluctuations begin to dominate near the T1/T2 triggering threshold and cause a type of signal floor of around 2 VEM when using the new triggers (see figure 5.11). This results in biased measurements of the lateral distribution beyond r_{up} for which $2 \text{ VEM} \simeq S(450) \times \text{LDF}(r_{\text{up}})$. For relatively low energy showers, the small events will include a larger fraction of stations at or beyond r_{up} which will impact the estimation of the shower size. This was previously highlighted by the global residuals of the LDF (figure 5.13, top). A slight under-estimation of the signals at 450 m was observed, the magnitude of which increases with decreasing shower size.

A second potential source of bias in the LDF fit is related to the shape of the lateral distribution for various secondary particle types. The observed lateral distribution of signal by the WCDs, $S(r)$, is neither the lateral distribution of energy nor particle density. Instead, the signal in a station is a convolution of the properties of the secondaries entering the water (their trajectory, energy, etc.) and the response of the SD hardware/electronics. This response function is non-uniform for all particle types (i.e. muonic, hadronic, and electromagnetic) and the observed signal can be thought of as a sum of individual distribution functions, $\mathcal{LDF}_i(r)$, and relative signal amplitudes, \hat{S}_i ,

$$S(r) = S(450) \text{LDF}(r) = S(450) \times \sum_i^{\text{Types}} \hat{S}_i \mathcal{LDF}_i(r). \quad (6.1)$$

In this form, it is more clear that what $\text{LDF}(r)$ is describing is quite complex. The way its shape evolves with zenith angle and shower size must reflect both the

change of the individual LDFs ($\mathcal{LDF}_{\text{Muonic}}$, etc.) and their relative amplitudes. The shape parameters describing the evolution of $\text{LDF}(r)$ were tuned (see section 5.2.3) using only high-energy events and are extrapolated to lower energies.

At high energies, showers will trigger many SD stations and will include many of the trigger types. However, at low energies, the shower will be smaller and a higher fraction of the stations will pass only the TOTd and MoPS triggers. Since the triggers were designed to not be sensitive to muons, an observed event will not have sampled the signal distribution given in equation 6.1, but will instead have sampled a sum over only the electromagnetic and hadronic particle types. This can cause problems since lower energy showers interact higher in the atmosphere on average and arrive at the ground with a *large* muonic fraction.

A study was conducted to examine the net bias in the shower size estimation. Four years (2014 - 2017) of 750 m array events were reconstructed using the new triggers. The relative LDF fit residuals were calculated for stations with an axial radius of 450 ± 20 m. The results are shown in figure 6.1 for ten bins of equal size in $\sin^2(\theta)$ as a function of $\log(S(450)/\text{VEM})$. For large shower sizes, $S(450) > 10 \text{ VEM}$,

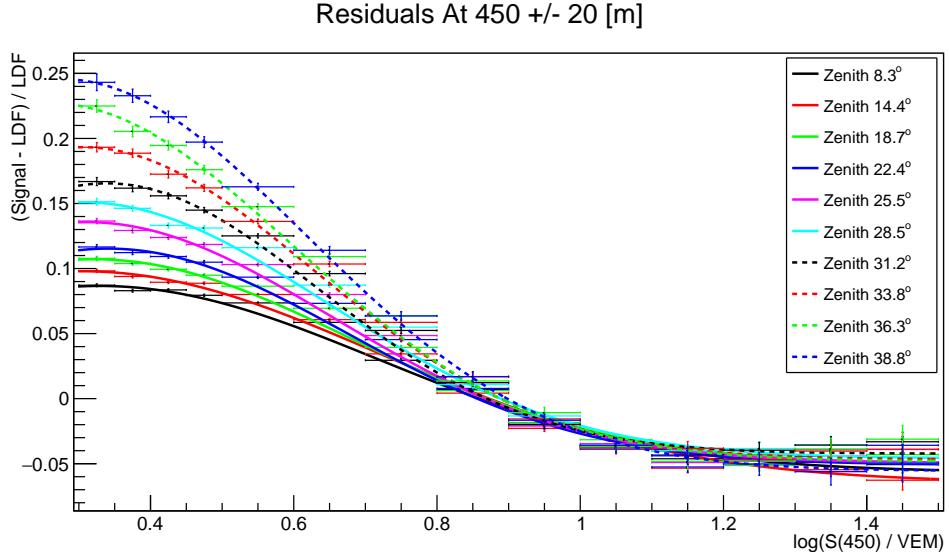


Figure 6.1: The bias in the shower size assignment was assessed by comparing the observed and expected signal sizes for stations near 450 m from the shower axis. The relative differences are shown for 10 zenith angle bins as a function of shower size (as given by the LDF).

the signal is overestimated by the LDF by about 4%. Note that a flat bias is not

a problem since the absolute shower size will ultimately be absorbed in the FD calibration. However, it is important that there are no zenith dependencies which are seen for small shower sizes, $S(450) < 10$ VEM. Here the LDF is underestimating the signals measured in the stations. The bias is as large as 25% and is most pronounced at high zenith angles.

The bias in the $S(450)$ values was fit empirically to an unnormalized Gaussian

$$\mathcal{B}(S(450), \theta) = 1 + A(\theta) \exp\left(-\frac{(\log S(450) - \mu(\theta))^2}{\Delta(\theta)}\right) + C(\theta). \quad (6.2)$$

Here the model parameters are the amplitude, A , the mean, μ , the rate, Δ , and the offset, C were allowed to be functions of shower inclination to capture the clear zenith dependence of \mathcal{B} . For each zenith angle bin, the model (specifically $\mathcal{B} - 1$) was fit to the data and is shown by the solid and dashed lines in figure 6.1. The four model parameters were then fit using polynomials. Their values are shown in figure 6.2 (black points) as a function of $\sin^2 \theta$ as well as the polynomial fits (red lines). The zenith dependence of each of the parameters is given below, statistical errors are given in parentheses

$$\begin{aligned} A(\theta) &= 0.1468(53) + 0.004(65) \sin^2 \theta + 0.94(17) \sin^4 \theta \\ \mu(\theta) &= 0.3175(86) \\ \Delta(\theta) &= 0.286(31) - 0.26(13) \sin^2 \theta \\ C(\theta) &= -0.0626(48) + 0.131(51) \sin^2 \theta - 0.26(12) \sin^4 \theta. \end{aligned} \quad (6.3)$$

The LDF bias can then be removed by using the model

$$S(450) \rightarrow \mathcal{B}(S(450), \theta) \times S(450). \quad (6.4)$$

Note that this analysis was only performed up to 40° . Below 10 VEM, where the bias begins to be relevant, the 750 m array is no longer fully efficient for more inclined showers (see section 7.1). The reduced triggering efficiency results in too few events with stations near 450 m from the shower axis to robustly fit the bias. Above 40° an extrapolation will be sufficient since the bias is flat above 10 VEM and the physics analyses in this work will only include events in the parameter space where the array is fully efficient.

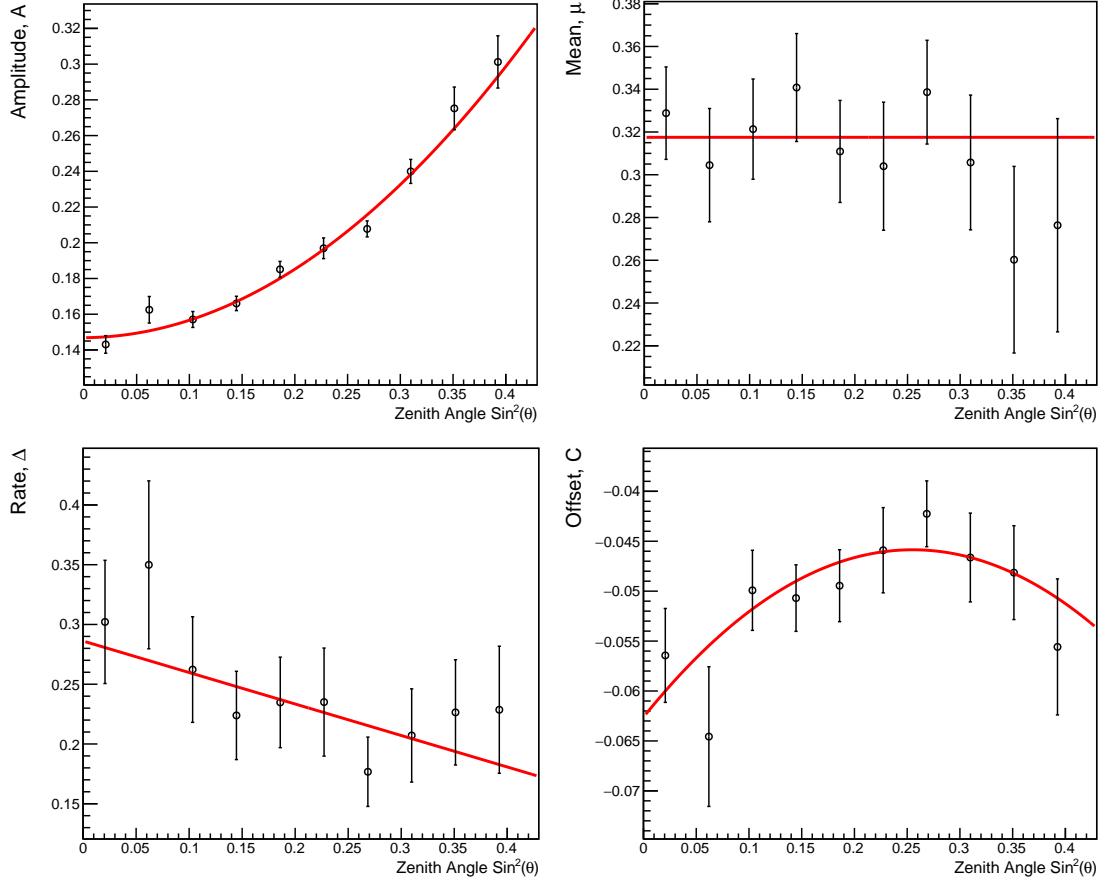


Figure 6.2: The plots above show the zenith dependencies of the four LDF-bias parameters (see equation 6.2). The black points are the best fit values for ten zenith angle bins of equal size in $\sin^2 \theta$. The red lines show the fit polynomials to describe these dependencies (see equation 6.3).

6.2 Correction for the Atmospheric Conditions

The detection of air showers relies on using the dynamically changing atmosphere as a radiator and absorber. The daily and yearly cycles of the local weather produce a systematic bias in the observed shower sizes by the SD. The local air pressure changes the atmospheric overburden while fluctuations in temperature change the Molière radius. Since the SD only samples showers on the ground, the fluctuating overburden results in changes in the average age at which showers are observed. Likewise, changes in the Molière radius affect the lateral spread of particles on the ground. While shower sizes can be estimated via the signal at a particular axial

distance, a better metric would be the signal at a particular number of Molière radii. The impact of the dynamically changing weather on the shower size estimation must be corrected for when attempting to produce an accurate estimation of an air shower's energy.

At Auger, the temperature, T , and pressure, P , are monitored every five minutes by a variety of weather stations located at the [FD](#) sites and the [CLF](#). These measurements are used to find the air density, ρ , using the dry-air relation, $\rho = (M/R)(P/T)$ where $M/R = 0.3484 \frac{\text{kg}}{\text{m}^3 \text{ K}}$. The daily and yearly cycles of these values are shown in figure 6.3. The air density is seen to have daily cycles in

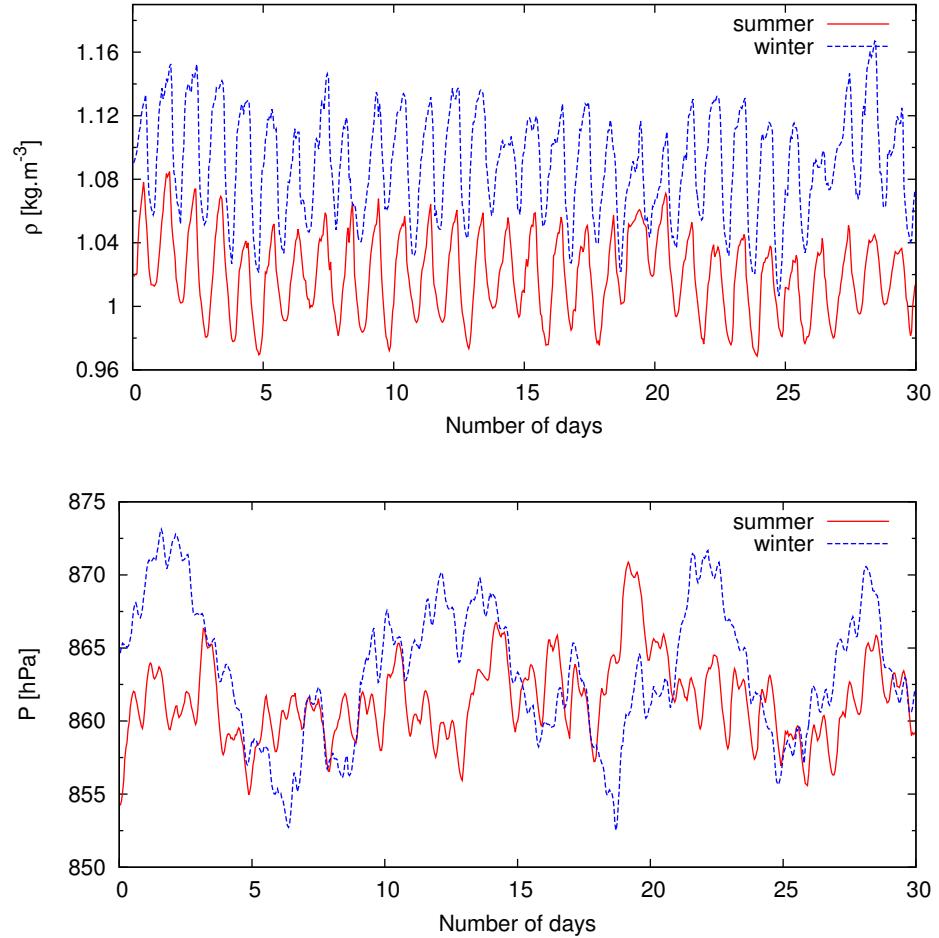


Figure 6.3: The plots above show the local air density (top) and pressure (bottom) at Auger. The hourly measurements for one summer (solid red) and winter (dashed blue) month are shown to emphasize the contribution of the daily and yearly variance. Figure from [82].

amplitude of 3% though there is a stronger yearly amplitude of around 6%. The air pressure shows no consistent yearly fluctuations and is instead dominated by daily fluctuations.

The effects can be directly seen by looking at the number of events above some shower-size threshold S_{cut} . For showers whose sizes, if observed during the average atmospheric conditions, are near this border, $S_{\text{cut}} - \varepsilon < S(450) < S_{\text{cut}} + \varepsilon$, then slight changes in the weather may cause them to cross from one side to the other. For instance, during periods of relatively low air pressure, showers will arrive at the **SD** with a less attenuated electromagnetic component and would thus be observed as having a larger-than-average size.

On top of simply adding an additional source of systematic uncertainty to the shower size over the course of a day/year, this can significantly affect the results of any analysis for which these effects are not averaged out. One example of such a study is the cross calibration between the **SD** shower sizes, and the **FD** energy measurements. The fluorescence telescopes only take data at night during which the air density is typically higher than average. This means that showers will likewise be observed with lower than average sizes. Since the **FD** already accounts for changes in the atmospheric conditions, this will produce a bias in the energy calibration (though ultimately the resultant energy bias is rather small, $\simeq 0.5\%$).

Large scale anisotropy studies are also impacted by the cyclical nature of the weather conditions. For such an analysis, the daily atmospheric patterns will produce a bias in the number of events above a given S_{cut} (or equivalently, an energy cut) over a sidereal day. The net migration of events across S_{cut} can produce an observed dipole on the sky. Given that in the ultra-high energy range, the amplitude of anisotropy is $\lesssim 5\%$ [39], this effect cannot be ignored.

To correct for this shift in energy, the rate of air showers detected above an S_{cut} were studied as a function of the atmospheric conditions. The arrival rate, R , of cosmic rays should follow

$$\frac{dR}{d\theta} = 2\pi \sin \theta \cos \theta \int_{S_{\text{cut}}}^{\infty} dS \epsilon(S, \theta) \frac{d\Phi}{dE} \frac{dE}{dS}. \quad (6.5)$$

Here $\frac{d\Phi}{dE}$ is the differential flux of cosmic rays with energy which is approximately described by a power law ($\propto E^{-\gamma}$), $\frac{dE}{dS}$ describes the way the cosmic ray energy

changes with signal ($E \simeq AS^B$, see section 6.4), and ε is the triggering efficiency of the array (see section 7.1). Since the *observed* sizes will be scaled by the weather conditions, \mathcal{W} , a function of the local air density and pressure, this can be rewritten to first order as

$$\frac{dR}{d\sin^2\theta} \propto \mathcal{W} \int_{S_{\text{cut}}}^{\infty} dS \epsilon(S, \theta) S^{-B\gamma+B-1}. \quad (6.6)$$

A model for \mathcal{W} was developed [70] to describe scaling of $S(450)$ with the atmospheric conditions. The model includes three parameters, the pressure at the time of the event, the average density within twelve hours of the event, ρ_d , and the density two hours previous to the event, $\tilde{\rho}$ (all calculated at ground level). This last parameter, $\tilde{\rho}$, is used to account for the delay in the atmospheric conditions between the **SD** and two radiation lengths (~ 1 km) above the ground. This model relates the observed signal, $S(450)$, to the expected signal, S_0 , that would be measured in the yearly-averaged local atmosphere with $\rho_0 = 1.06 \text{ kg m}^{-3}$ and $P_0 = 862 \text{ hPa}$

$$S(450) = S_0 \mathcal{W}(P, \rho_d, \tilde{\rho}) = S_0 [1 + \alpha_P(P - P_0) + \alpha_\rho(\rho_d - \rho_0) + \beta_\rho(\tilde{\rho} - \rho_0)]. \quad (6.7)$$

The best fit of the model parameters, α_P , α_ρ , and β_ρ was found by determining the values for which the detection rate given by equation 6.6 was constant as a function of time (see [70]). Further, they were also parameterized as a function of zenith angle to account for the changing slant depth (see B.1).

For each event, the weather observations made by the centrally located **CLF** (see figure 3.1) are used to produce measurements of the pressure and density. These measurements are typically recorded every five minutes¹. Using equation 6.7, each observed shower size, $S(450)$, is replaced with the unbiased value, S_0 .

6.3 Constant Intensity Cut

An unavoidable consequence of observing cosmic rays with a ground-based detector is the variable slant depth of the array depending on the zenith angle. The increased grammage for showers originating closer to the horizon results in an older shower being sampled by the **SD**. Observationally, two identical showers with differing

¹In cases where there are large gaps in the data (> 3 hr), atmospheric measurements from the nearest **FD** site are used after being corrected for the difference in altitude.

trajectories would thus produce a larger value of $S(450)$ for the more vertical of the two². A method, known as a Constant Intensity Cut (**CIC**) [67, 83] has been developed to correct for this effect without requiring a detailed model of the hadronic interactions and the atmospheric attenuation.

The **CIC** method is based on the simple assumption that locally there is no preferred arrival direction of cosmic rays to Earth. Thus the number of observed showers of a given energy should not have any local-zenith angle dependence. This is a reasonable assumption given that the only observed deviation from isotropy in the **UHE** regime is that of a dipole feature on the sky with a $\simeq 5\%$ anisotropic fraction [39]. This feature is spread out over the entire **SD** field of view (**FOV**) and is thus roughly sampled by all local zenith angles equally. Further, this dipole was only observed for cosmic rays with energies above 8 EeV. The 750 m array will only observe ~ 10 such events per year which will be completely washed out by the zenith angle distribution of the ($\sim 100,000$) lower energy events. Additionally, at energies less than 1 EeV, the deflections of cosmic rays in the Galactic and intergalactic magnetic fields are on the order of the **SD**'s **FOV**, and significantly smear out any local source. So since a fairly strong and localized source has not been seen in the 0.1 - 4 EeV range, thus the isotropic assumption is well motivated.

For a flat detector, such as the **SD**, the number of arriving cosmic rays is distributed as

$$\frac{dN}{d\theta} = \cos \theta \sin \theta. \quad (6.8)$$

The cosine term accounts for the decreasing projected area of the array with increasing zenith angle while the sine term describes the differential spherical segment. This can be made linear by using a change of variables³ to produce

$$\frac{dN}{d(\sin^2 \theta)} = \text{Const.} \quad (6.9)$$

This equation implies that in the absence of attenuation (and any other observational effects), the arrival direction of cosmic rays should be constant in bins of equal size in $\sin^2 \theta$.

The impact of the attenuation can be seen in figure 6.4 which gives the number

²This assumes that the array is deeper in the atmosphere than X_{\max} for all the observed showers, like at Auger. Otherwise, the shower would be growing and the effect would be reversed.

³ $\cos \theta \sin \theta d\theta = -\cos \theta d(\cos \theta) = \frac{1}{2}d(\sin^2 \theta)$

of observed showers as a function of weather-corrected shower size and $\sin^2 \theta$. This

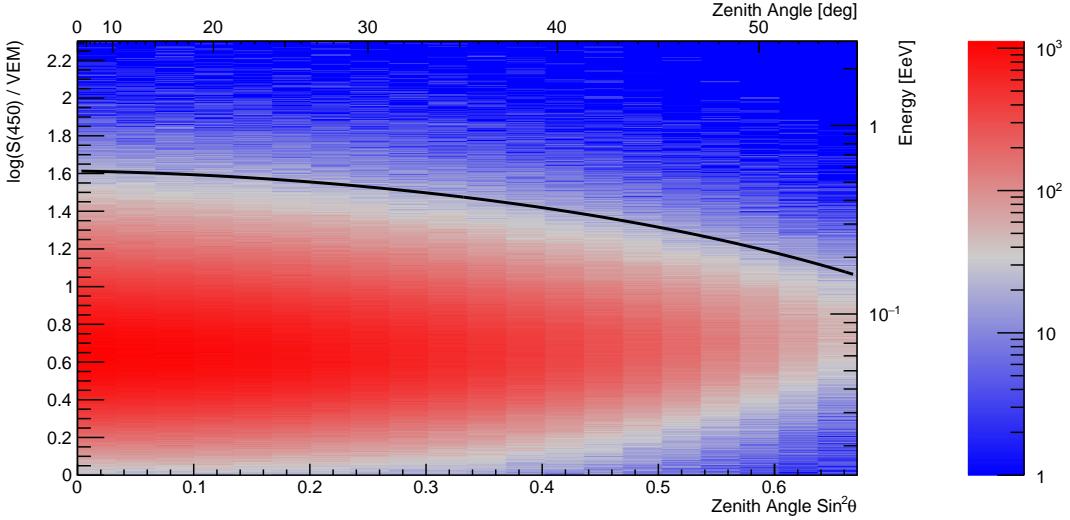


Figure 6.4: The log of the number of observed showers by the 750 m array with new triggers turned on is shown above in equal $\sin^2 \theta$ bins. The effects of attenuation are evident above 10 VEM from the inconsistent number of events at a given shower size. The bottom part of the histogram shows the decrease in efficiency where the SD no longer triggers on every shower. The black line represents an example CIC curve above which there is a consistent number of events in every angular bin.

histogram shows a number of important observational effects. In the horizontal direction, there is a deficit of observed air showers at high angles. Since the slant depth of the array is larger for more inclined trajectories, such showers will arrive at the ground with a more attenuated electromagnetic component. The reduced number of particles will therefore deposit less signal at 450 m from the axis. So the net effect is for showers to migrate downwards as the attenuation becomes stronger with increasing slant depth.

In the vertical direction the number of observed events first increases with decreasing shower size as a result of the power-law flux of cosmic rays. The distribution then peaks around $\simeq 8$ VEM and begins to decrease. This final decrease is caused by the reduced efficiency of the array (the probability of showers to pass the T4 triggers, see section 7.1). Again the effect is seen most prominently at high zenith angles which is also a product of attenuation effects.

The CIC corrects the downward migration of events by scaling the observed shower sizes in a way that the zenith-angle distribution of events follows the expected

distribution from equation 6.9.

6.3.1 The CIC Tuning Methodology

The observed size of a shower at the ground, S_0 , is equal to the maximum size, S_{\max} , (which would occur at X_{\max}) scaled by the absolute atmospheric attenuation, $F_{\text{atten}}(\theta)$,

$$S_0(\theta) = S_{\max} F_{\text{atten}}(\theta). \quad (6.10)$$

Since the attenuation function is unknown, the **CIC** method relies instead on the *relative* attenuation, $f_{\text{CIC}}(\theta)$,

$$\frac{S_0(\theta)}{S_0(\theta_{\text{ref}})} = \frac{F_{\text{atten}}(\theta)}{F_{\text{atten}}(\theta_{\text{ref}})} = f_{\text{CIC}}(\theta). \quad (6.11)$$

Here $S_0(\theta_{\text{ref}})$ is the shower size that would be expected to be observed if the shower had instead arrived at a reference zenith angle⁴ of θ_{ref} . Equation 6.11 implies that $S_0(\theta_{\text{ref}})$ is a quantity which is proportional to the observed shower size and is independent of zenith angle (i.e. it is a measure of the shower size with the attenuation effects removed),

$$S_0(\theta_{\text{ref}}) = \frac{S_0(\theta)}{f_{\text{CIC}}(\theta)}. \quad (6.12)$$

For simplicity and to emphasize this zenith independence, the reference size will be referred to as simply $S_0(\theta_{\text{ref}}) \rightarrow S_{\text{CIC}}$.

Since the scaling of the observed shower size by $f_{\text{CIC}}(\theta)^{-1}$ removes the attenuation effects (the reduction in observed size at higher zenith angle), the distribution of arrival directions for a given value of S_{CIC} should match that of cosmic rays at the top of the atmosphere (i.e. that of equation 6.9). The **CIC** method thus finds an appropriate description of $f_{\text{CIC}}(\theta)$ such that this assertion is true.

Of course, this would only be valid for energies/shower sizes for which there are no other zenith angle dependent observational effects. As previously described, at low energies, a zenith dependence is also caused by the triggering efficiency which would make this invalid. Thus the tuning can only be applied in the region where

⁴The reference angles were chosen to be the median expected zenith angle under the isotropic assumption, $\frac{1}{2}(\sin^2 \theta_{\max} + \sin^2 0^\circ) = \sin^2 \theta_{\text{ref}}$, where θ_{\max} is the maximum viewing angle of the detector.

the array is fully efficient (100% for $\theta < \theta_{\max}$, the maximum viewing angle). This is done by instead only considering the number of events with $S > S_{\text{cut}}$. A proper choice of this cut needs to ensure that the array is fully efficient at least down to observed shower sizes of $S_{\text{cut}} f_{\text{CIC}}(\theta)$. However, a choice that is too large will decrease the statistical precision of the analysis. Based on these motivations, a value of $S_{\text{cut}} = 30$ VEM was used⁵. The attenuation function $30 \times f_{\text{CIC}}(\theta)$ is shown by the black line as an example in figure 6.4.

Equation 6.11 implies that $f_{\text{CIC}}(\theta_{\text{ref}}) = 1$. Polynomials of increasing order were tried, making sure that the data was not being over-fit. The best results were found using a third order function

$$f_{\text{CIC}}(\theta) = 1 + a x + b x^2 + c x^3, \quad x = \sin^2 \theta_{\text{ref}} - \sin^2 \theta. \quad (6.13)$$

This cubic function includes three model parameters, a , b , and c .

The model was then fit by noting that the isotropic assumption results in a CDF which increases linearly with $\sin^2 \theta$ (equation 6.9)

$$\text{CDF}(\theta) = \frac{\sin^2 \theta}{\sin^2 \theta_{\max}}. \quad (6.14)$$

Likewise, the CDF of the N observed events with shower sizes $S_0 > S_{\text{cut}} f_{\text{CIC}}(\theta)$, $\text{cdf}(\theta)$, can be written as

$$\text{cdf}(\theta) = \frac{1}{\text{Norm.}} \sum_i^N \Theta(S_{0,i} - S_{\text{cut}} f_{\text{CIC}}(\theta_i)) \Theta(\theta - \theta_i). \quad (6.15)$$

Here Θ is the Heaviside function. Note that the three CIC parameters are contained within $f_{\text{CIC}}(\theta)$ and will thus be fit implicitly.

The attenuation function that best describes the data was then found by checking the compatibility of $\text{cdf}(\theta)$ with $\text{CDF}(\theta)$ using the Anderson-Darling test [84]

$$\mathcal{A} = N \int_0^{\sin^2 \theta_{\max}} \frac{[\sin^2 \theta - \text{cdf}(\theta)]^2}{\sin^2 \theta (\sin^2 \theta_{\max} - \sin^2 \theta)} d(\sin^2 \theta). \quad (6.16)$$

The value, \mathcal{A} , is a test statistic to determine whether a data set has been drawn

⁵The choice of this cut value does not have a large impact on the result of the CIC tuning, see section A.5

from a given parent distribution (in this case, the linear CDF). The best fit values would be the ones for which \mathcal{A} is minimal.

Because the optimal CIC parameters were determined via a numerical minimization routine, care was taken to reduce the impact of the discrete nature⁶ of the observed shower sizes. Instead of simply treating the events as Heaviside functions, the events were smeared out using a density function, n , with length scale, σ ,

$$n(x, \sigma) = \begin{cases} 0, & x < -\sigma \\ \frac{1}{2} \left(1 + \frac{x}{\sigma}\right)^2, & -\sigma \leq x < 0 \\ 1 - \frac{1}{2} \left(1 - \frac{x}{\sigma}\right)^2, & 0 \leq x < \sigma \\ 1, & x \geq \sigma. \end{cases} \quad (6.17)$$

This function is shown in figure 6.5. Using the smearing function, the CDF of the

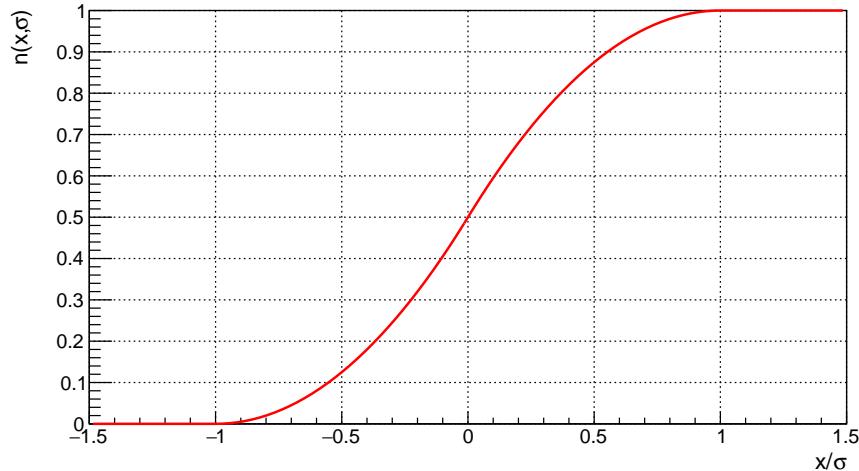


Figure 6.5: To aid the numerical minimization during the fit of the CIC parameters, each event was smeared according to its statistical size uncertainty, σ . See equation 6.17.

observed events can be written as

$$\text{cdf}(\theta) = \frac{1}{\text{Norm.}} \sum_i^N n_i (S_{0,i} - S_{\text{cut}} f_{\text{CIC}}(\theta_i), \sigma_{S_{0,i}}) \Theta(\theta - \theta_i) \quad (6.18)$$

⁶In this fit, the numerical routine MINUIT was used. This routine requires a good estimation of the parameter-space gradient, which is also calculated numerically using small steps. This can cause problems if the function to be minimized has a discontinuous derivative.

Zenith Angles		Shower Size		CIC Parameters		
θ_{\max}	θ_{ref}	S_{CIC}	$S_{\text{cut}} [\text{VEM}]$	a	b	c
40°	27°	S_{27}	30	1.14 ± 0.05	-1.27 ± 0.18	-1.22 ± 1.79
55°	35°	S_{35}	30	1.71 ± 0.02	-1.00 ± 0.08	-2.15 ± 0.32

Table 6.1: The best fit values of the CIC parameters (equation 6.13) for the 750 m array using new triggers.

where $\sigma_{S_{0,i}}$ is the statistical uncertainty of the shower size (the shower size uncertainty will be discussed in section 6.4.3).

6.3.2 Results of the CIC Tuning

The tuning was performed using four years of 750 m array events (2014 - 2017) with the new triggers included (2.9 million observed events with $\theta < 55^\circ$). This process was repeated for two zenith angle ranges where the respective CIC parameters ($\theta_{\max}, \theta_{\text{ref}}, S_{\text{CIC}}$) are $(40^\circ, 27^\circ, S_{27})$ and $(55^\circ, 35^\circ, S_{35})$. The numbers of events with $S_{\text{CIC}} > 30$ VEM were 23,000 and 37,000, respectively. These data sets were chosen as they correspond to zenith angle ranges above which the array is fully efficient at $10^{17.0}$ eV and $10^{17.3}$ eV, respectively (see section 7.1).

The results of the fit for the 55° data set are shown by the black line in figure 6.4 (for another visualization of the CIC data for both data sets, see figure B.3). The CIC parameters are given in table 6.1 along with their bootstrapped [85] errors.

Using these CIC parameters, each shower was assigned a corrected shower size according to equation 6.11,

$$S_{\text{CIC}} = \frac{S_0}{f_{\text{CIC}}(\theta)}. \quad (6.19)$$

The results of the CIC tuning can be seen qualitatively in figure 6.6 which shows the logarithm of the number of events in bins of equal size in $\log(S_{35}/\text{VEM})$ and $\sin^2 \theta$. The attenuation effects have been removed and the histogram now shows horizontal bands of constant color (counts) in the top half of the plot, as expected. Further, the triggering efficiency of the array can also be more clearly seen for smaller signal sizes and can be estimated based on the value of S_{35} where the counts begin to decrease with decreasing values of S_{35} ($\simeq 10^{1.1}$ VEM at 55°).

A more quantitative assessment of the CIC fit is given in figure 6.7 which shows

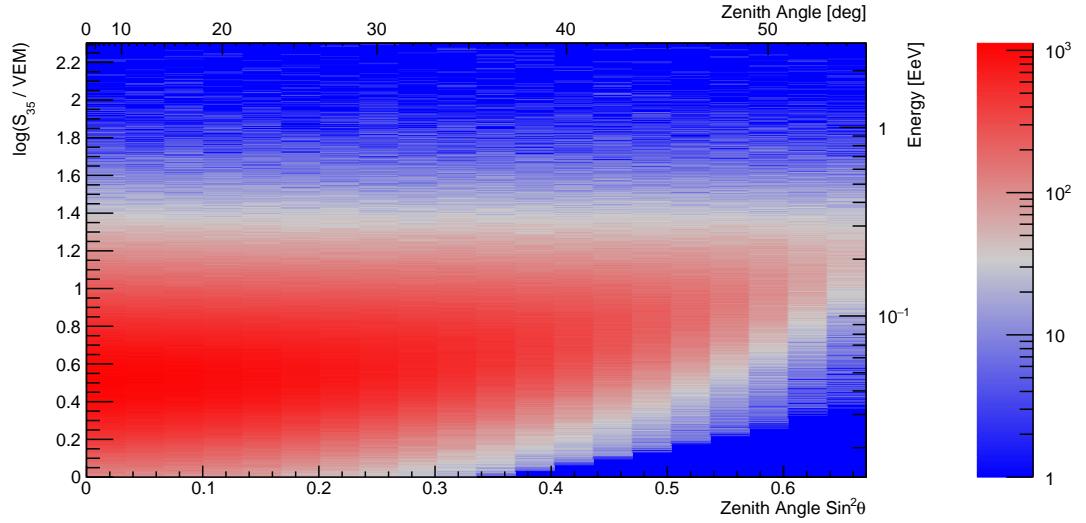


Figure 6.6: The histogram shows the log of the number of events as a function of S_{35} and $\sin^2 \theta$. The quality of the correction can be seen by the flatness of the horizontal bands of color. At small shower size $\lesssim 10$ VEM, the triggering effects become apparent.

the number of events with an S_{35} that is larger than a given S_{cut} . The flatness of the data is estimated by the square of the residuals away from the average value. For all shower sizes, the data agrees well with being flat indicating that there is not a large impact on the [CIC](#) tuning based on the value of S_{cut} . This is important because the choice of 30 VEM, while well motivated, was, in some sense, arbitrary.

The flatness also has more important implications on the cosmic ray and shower physics. In this section, the motivation for the way that the correction factor was applied did not consider an energy dependence of $f(\text{CIC}(\theta))$. In fact, it is expected that the correction factor *should* change with energy to some extent. For example, a shift to heavier primaries would create a more muonic shower front which is less attenuated by the atmosphere. If the average primary mass is changing over the observed energy range, the shape of $f_{\text{CIC}}(\theta)$ would not be constant for different values of S_{cut} . The energy dependence of the [CIC](#) procedure has been studied and is small for the 750 m array (see section [A.5](#)).

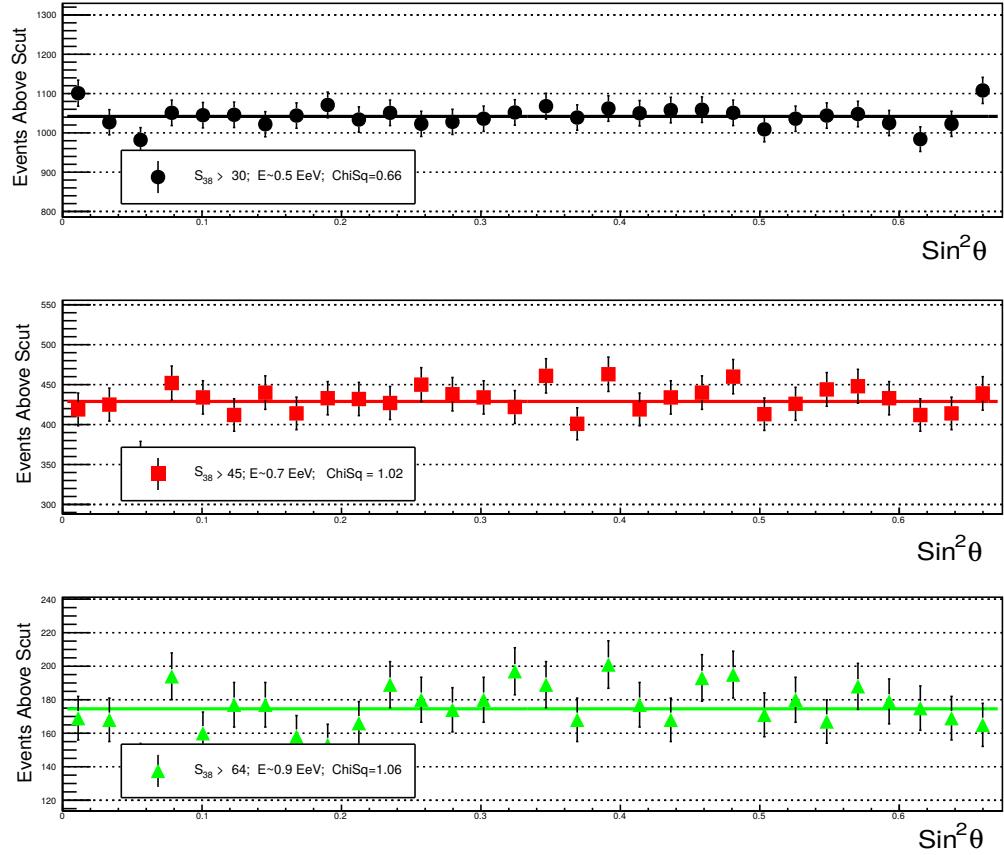


Figure 6.7: The plots above show the number of events with $S_{35} > S_{\text{cut}}$ for various values of S_{cut} . For each cut, the approximate corresponding energy and the reduced χ^2 , estimating the agreement with flatness, are given. The **CIC** tuning was performed using a cut value of 30 VEM (shown in the top plot).

6.4 Energy Calibration Using FD Events

The final step in assigning an energy to **SD** events is the cross-calibration between the surface and fluorescence detectors. This is done using a special set of showers, called *golden hybrid* events, which can be independently reconstructed by the two detectors. These can then be used to find a relationship between the energy measurements by the **FD** and the corrected signal sizes from the **SD**.

6.4.1 Selection of Golden Hybrid Events

The **FD** observes the evolution of **EAS**s in the atmosphere provided that they produce enough fluorescence light to be seen by the telescopes. At low energies,

this can be restrictive for golden hybrid showers which need to be very close to the telescopes and must also land in the 750 m array (and pass the 6T5 condition).

Beginning with the set of all such events, a number of cuts were applied to ensure a high quality **FD** reconstruction and to avoid acceptance biases. Recall that the longitudinal profile seen by the telescopes is fit to a Gaisser-Hillas function (see figure 3.4 for an example). The first set of cuts requires that the statistical error on the fit for X_{\max} is less than 40 g/cm^2 (typical uncertainties are 20 g/cm^2) and that the calorimetric energy uncertainty is less than 12%. Further, each shower is required to be viewed through a grammage of at least 200 g/cm^2 to define well the curvature of the profile (smaller showers are typically viewed only near X_{\max} where the longitudinal profile is relatively flat).

The second set of cuts requires that the shower is well contained within the **FOV**. Since a single shower can be observed by multiple telescopes, no events are used which have gaps in the longitudinal profile larger than 20 g/cm^2 . Such an event is shown in figure 6.8 which was seen by four telescopes. Each white

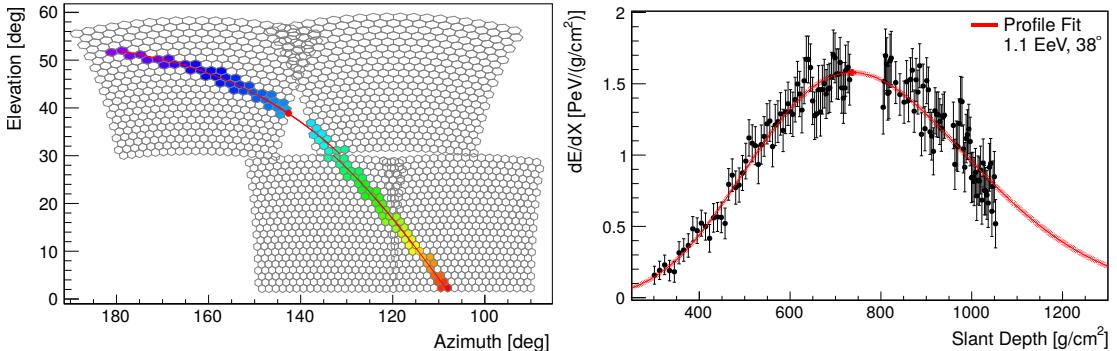


Figure 6.8: Left: The figure above shows a shower that was seen by four **FD** telescopes. The top two **FOV**s correspond to **HEAT** while the two in the bottom ones are the **FOV**s of Coihueco. The colors indicate the timing in which the pixels triggered (earlier is purple). The red line is the fitted **SDP** with a red circle at the shower maximum. Right: The combined longitudinal profile is shown for the same event. The gap in the observation can be seen where the shower crossed the hole between the **HEAT FOV**s.

hexagon represents the **FOV** of one **PMT** from one telescope. This event was seen by two telescopes from **HEAT** (top left and top right) and two from Coihueco (bottom center and right). There is a gap in the combined **FOV** during which the shower was not observed and thus would be rejected. Further, X_{\max} must also be

visible in the **FOV**. This ensures that enough data points are seen to fit the curved longitudinal profile and also acts as a fiducial cut to avoid primary-mass dependent biases. The profile is also shown in figure 6.8 for the same event. In this case, the shower maximum was just barely contained within the **FOV** (red circle).

The final cut requires that there is less than 25% coverage of the sky by clouds. This is determined by the dedicated LIDAR measurements made at the **FD** sites. The reflection of the nitrogen fluorescence off of clouds can produce significant amounts of indirect light to be seen by the telescopes which hampers the determination of event geometry and timing.

Due to missing aerosol and cloud data⁷, the calibration could only be performed over two years of **FD** events, 2014-2015. After applying the cuts listed above, 1031 golden hybrid events with zenith angles less than 55° and 998 events with zenith angles less than 40° remained. The distributions of the **FD** energies and zenith angles are shown independently in the top plots of figure 6.9 and their correlations are shown in the bottom plot. Each red dot represents an individual event and the average tendencies are indicated by the blue line. Since the arrival direction of cosmic rays should be evenly distributed in $\sin^2 \theta$, a detector with perfect efficiency would have a zenith distribution with a mean of $\sin^2 35^\circ = 0.33$ (shown by the grey dashed line). The average values shown in the bottom plot highlight that the exposure for golden hybrid events is greatest for more vertical orientations, particularly at low energies. Also note that most of the calibration events are clustered in the range $10^{17.4} - 10^{17.7}$ eV and decrease rather rapidly at higher and lower energies. This is a result of the exponentially falling cosmic ray spectrum and the restrictive golden hybrid acceptance for low-energy, dim showers, respectively.

6.4.2 The Calibration Fit

The **FD** makes nearly calorimetric measurements of an **EAS**'s energy, E_{cal} . However, some of a shower's energy is carried away by particles that reach the ground (muons, neutrinos, etc.) and is not observed by the telescopes. The amount of this so called *invisible energy*, E_{invis} , has been evaluated using simulations and very inclined ($> 80^\circ$) **SD** events, which are nearly composed purely of muons [65, 66]. The total

⁷These types of atmospheric data must first be cleaned which has not yet been performed through 2017.

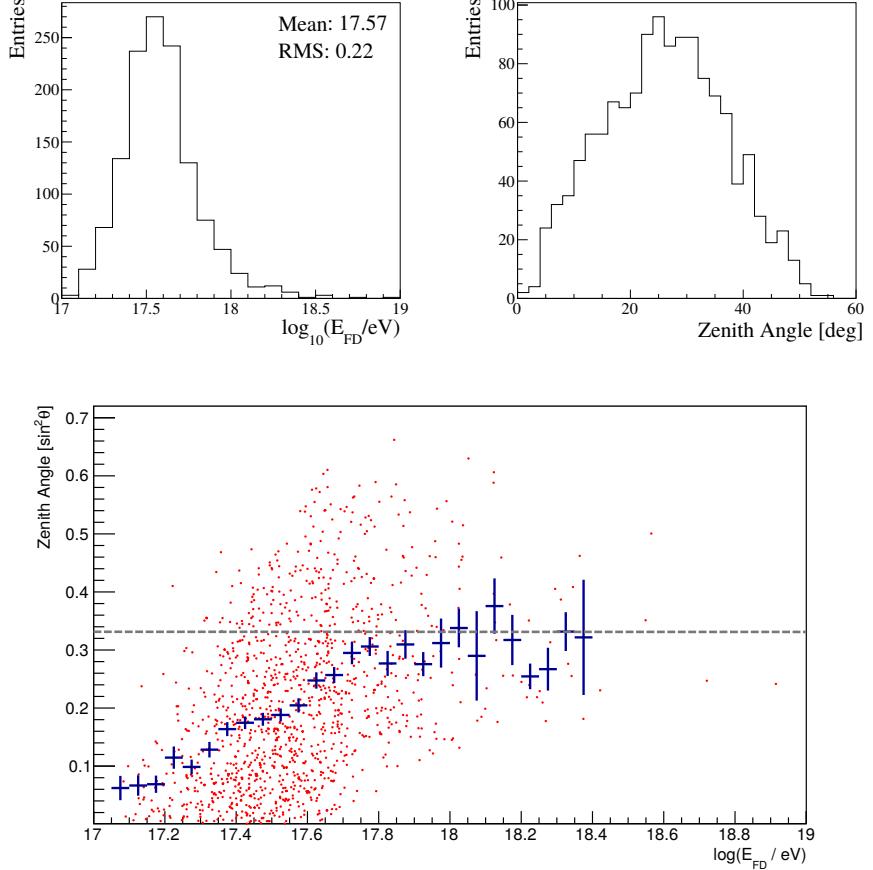


Figure 6.9: The distributions of golden hybrid events that passed a set of quality cuts are shown. The top left plot is a histogram of the number of events in various energy bins. The top right plot instead shows the distribution of events as a function of zenith angle. The bottom plot shows the correlations between the two (red points). The central tendency is also indicated by the blue point which indicates the strong relationship between energy and zenith angle. The dashed grey line shows the median expected arrival direction for the SD (35°).

air shower energy is then the sum of these two values for a given event,

$$E_{FD} = E_{\text{cal}} + E_{\text{invis}}. \quad (6.20)$$

It is this total value, E_{FD} , that is used to find a relationship between air shower energy and S_{CIC} for which a power law has been used,

$$E_{FD} = A (S_{\text{CIC}})^B. \quad (6.21)$$

Here A and B are the calibration parameters to be fit.

The fit was performed using the likelihood function given below,

$$\ln \mathcal{L} = \sum_k \ln \left[\sum_i \frac{e^{-\frac{1}{2} \frac{(E_{FD,k} - E_{FD,i})^2}{\sigma_{E_{FD},i}^2}}}{\sigma_{E_{FD},i}} \frac{e^{-\frac{1}{2} \frac{(S_{SD,k} - S(E_{FD,i}))^2}{\sigma_{S_{SD},i}^2}}}{\sigma_{S_{SD},i}} \right]. \quad (6.22)$$

This function takes into account the systematics of both the **FD** and **SD**. The **FD**'s energy resolution, $\sigma_{E_{FD}}$, is estimated on an event-by-event basis. However, the size resolution of the **SD**, $\sigma_{S_{SD}}$, is comprised of two terms, the detector sampling fluctuations, $\sigma_{S_{det}}$, and the shower-to-shower fluctuations, $\sigma_{S_{sh}}$, for which the latter cannot be determined for an individual event (see section 6.4.3 for further discussion). The index, k , runs over the golden hybrid events which passed the cuts defined in the previous section (i.e. the 1031 or 998 events, depending on the zenith and energy cuts).

However, i runs over all of the **FD** events that pass the cuts defined in the previous section and fall within the 750 m array, but are not necessarily seen by the **SD**. To describe the expected distribution of observed shower sizes in golden hybrid events, one must also know the distributions of observed shower energies and zenith angles for the **SD**, $g(E, \theta)$, and the **FD**, $h(E, \theta)$. This is the convolution of the cosmic ray spectrum and the aperture of the respective detectors. For the **FD**, this is hard to calculate, so the sum over i is a way of approximating this distribution, $h(E, \theta) \simeq \frac{1}{N} \sum_i^N \delta(E - E_{FD,i}) \delta(\theta - \theta_i)$. Such an approximation is not needed for the **SD**, since the calibration is only performed on energies for which the array is fully efficient ($g(E, \theta) = 1$).

The likelihood is then the product of two Gaussians, one for the **FD** events with energy, E_{FD} , and one for the **SD** with measured signal, S_{SD} , and expected signal, $S(E_{FD})$, given by the inverted calibration function (equation 6.21).

This fit was performed on the two event sets previously defined ($\theta_{\max} < 40^\circ$, $E_{FD} > 10^{17.0}$ eV) and ($\theta_{\max} < 55^\circ$, $E_{FD} > 10^{17.3}$ eV). The results of the calibrations are given in table 6.2 and shown in figure 6.10. The left plots show the distribution of events in the calibration data (black points) as well as the fitted calibration curve (red line, equation 6.21). The quality of the fit can be seen in the right plots which give the average ratio of the **SD** and **FD** energies for each event after applying the

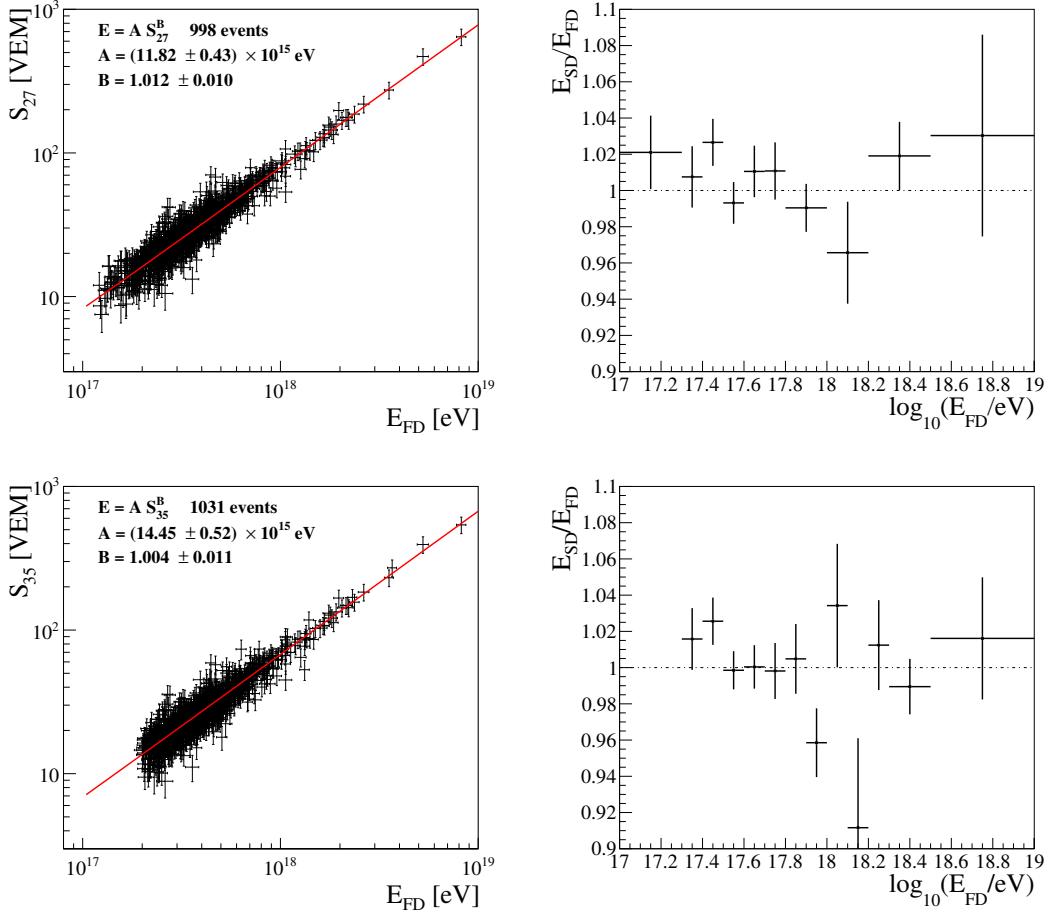


Figure 6.10: The plots above show the results of the **FD** calibration with the 40° (top) and 55° (bottom) data sets. The energy and signal measurements (black crosses) along with the calibration fit (red line) to a power law (equation 6.21) are shown on the left. The right plots provide a look at the binned residuals after converting the signal measurements into energy.

calibration. The distributions of $\frac{E_{\text{SD}}}{E_{\text{FD}}}$ for individual energy bins can be found in figure 6.13.

The calibration normalization, A , gives the energy of an event arriving at the reference angle and having $S_{\text{CIC}} = 1$. Because the reference angle is different for the two data sets, it is expected that A should differ for the two calibrations. The reduced attenuation at 27° results in $S_{27} > S_{35}$ and thus a smaller value of A for the $\theta_{\max} = 40^\circ$ set. The exponent, B , is compatible with unity in both cases which implies almost a direct correlation between the measured shower size and energy.

Data Set		Calibration Parameters	
θ_{\max}	θ_{ref}	A [PeV]	B
40°	27°	11.82 ± 0.43	1.012 ± 0.010
55°	35°	14.45 ± 0.52	1.004 ± 0.011

Table 6.2: The calibration constants for the 40° and 55° data sets are given above.

6.4.3 SD Energy Uncertainties

The systematic uncertainty on the **SD**'s energy scale is a combination of the uncertainty from the **FD** and an additional uncertainty from the calibration procedure, $\Delta E_{\text{sys,SD}}^2 = \Delta E_{\text{sys,FD}}^2 + \Delta E_{\text{sys,Calib}}^2$. The contribution from the calibration is comprised of two parts. The first comes directly from the uncertainty on the calibration fit (i.e. from the error on A and B). The second part accounts for the stability of the calibration procedure. This was studied by calibrating over different time periods and results in a 5% relative systematic uncertainty [66]. The combination of these two is shown in figure 6.11. The calibration stability is the dominating

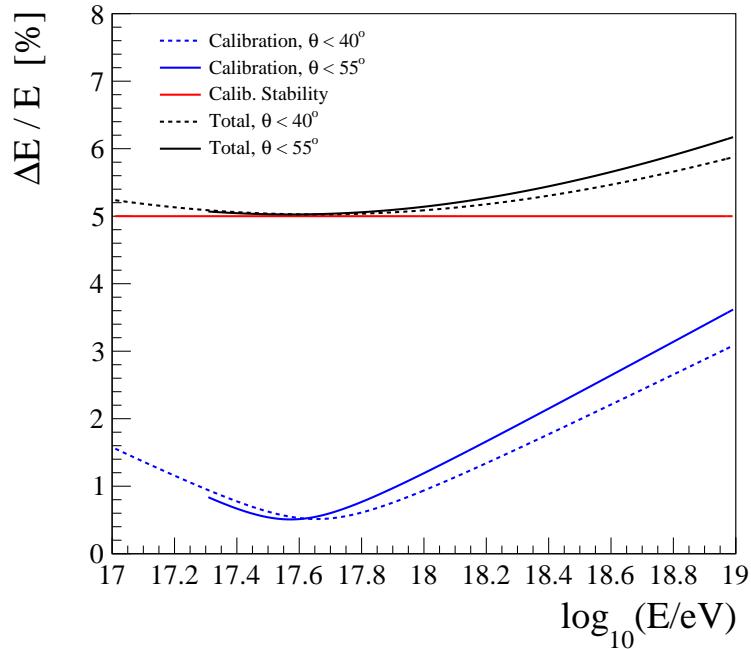


Figure 6.11: The contributions to the systematic uncertainty of the energy scale from the calibration procedure are shown above. This includes the uncertainty in the fit (blue lines) and the temporal stability of the calibration (red line).

term on top of which the uncertainty from the fit adds less than a percent to the total systematics.

The $< 6\%$ relative uncertainty coming from the calibration is a small addition to the much larger systematics of the **FD**. The leading factor of $\Delta E_{\text{sys,FD}}$ is from the **FD**'s own calibration for which the relative uncertainty is nearly 10%. The remaining systematics are due to uncertainties in the atmospheric conditions, the fluorescence yield, the invisible energy, and the fit of the longitudinal profile (for more, see [66]). Figure 6.12 shows the contributions to the total relative uncertainty. For both detectors, the total uncertainty (black lines) is stable to within one percent

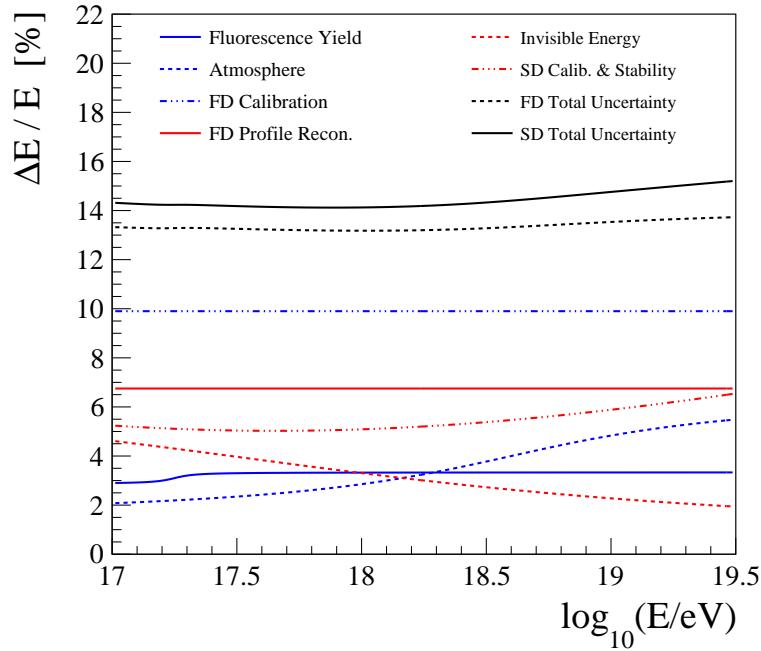


Figure 6.12: The contributions to the systematic uncertainty of the energy scale are shown above. This includes uncertainties from the fluorescence yield, the atmospheric properties, the **FD**'s energy calibration, the reconstruction of the longitudinal profile, the invisible energy, and the **SD**'s energy calibration (for 55°).

for all relevant energies, about 13.5% and 14.5% for the **FD** and **SD**, respectively.

The systematic uncertainty is comparable to that of the energy resolution. For a given event, there are two contributions to the energy resolution of a surface detector, $\sigma_{\text{SD}}^2 = \sigma_{\text{det}}^2 + \sigma_{\text{sh}}^2$. The first comes from the sampling fluctuations of the

detector,

$$\sigma_{\text{det}}^2 = \sigma_S^2 + \sigma_\theta^2 + \sigma_{\text{CIC}}(\theta)^2 + \sigma_{\text{LDF-bias}}(S(450), \theta)^2 + \sigma_{\text{Weather}}(\theta)^2 \quad (6.23)$$

that is, σ_{det} , is the squared-sum of the uncertainties from the reconstructed size and zenith angle (σ_S, σ_θ) and the correction terms that make up the S_{CIC} assignment ($\sigma_{\text{CIC}}, \sigma_{\text{LDF-bias}}, \sigma_{\text{Weather}}$). On an event-by-event basis, σ_{det} can be determined by propagating the reconstruction uncertainties on S and θ into the size correction terms.

The second contribution to the energy resolution comes from the shower-to-shower fluctuations, σ_{sh} . These are a result of the stochastic processes that govern **EAS** development. Even for a surface detector with arbitrarily small σ_{det} , the sparse sampling of the lateral distribution introduces a sensitivity to local fluctuations in the shower front. Thus a surface array would not measure the same particles for two showers with the exact same initial conditions. There is no way to estimate σ_{sh} on an event-by-event basis since it is completely separate from the reconstruction.

Within Auger, the shower-to-shower fluctuations have historically been estimated using a dedicated set of simulations [86] and with σ_{det} to calculate the total uncertainty, σ_{SD} . However, reliance on simulations for this uncertainty contribution can be avoided entirely by directly calculating σ_{SD} using the golden hybrid events. For the golden events, each of the detectors will make an independent measurement of the true shower energy smeared by their respective energy resolution. These observed energies are assumed to be Gaussian fluctuations around the true energy (as was done in the calibration, equation 6.22).

For two Gaussian-distributed variables, X and Y , the distribution of their ratio, $z = X/Y$, will depend on the means and standard deviations of X and Y , $\text{PDF}(z; \mu_X, \sigma_X, \mu_Y, \sigma_Y)$ (see equation B.1). The ratio of the energy measurements by the **FD** and **SD**, $z = E_{\text{SD}}/E_{\text{FD}}$, for golden hybrid events will likewise create a distribution dependent on $\mu_{E_{\text{FD}}} \equiv 1$, $\sigma_{E_{\text{FD}}}$, $\mu_{E_{\text{SD}}}$, and $\sigma_{E_{\text{SD}}}^2 = \sigma_{E_{\text{det}}}^2 + \sigma_{E_{\text{sh}}}^2$.

The ratios, z_i , were calculated for the events in the calibration data set and fit to $\text{PDF}(z; \mu_{E_{\text{SD}}}, \sigma_{E_{\text{SD}}}, 1, \sigma_{E_{\text{FD}}})$ in eight energy bins. As an example, the distributions of z_i and the respective fits are shown for two energy bins in the bottom plots of figure 6.13. During the fit, $\mu_{E_{\text{SD}}}$, $\sigma_{E_{\text{SD}}}$, and an overall normalization were left as

free parameters while, σ_{FD} , was fixed to its average value⁸ for the given energy bin (see figure B.4). The resulting values for $\sigma_{E_{SD}}$ are shown in figure 6.13. The red

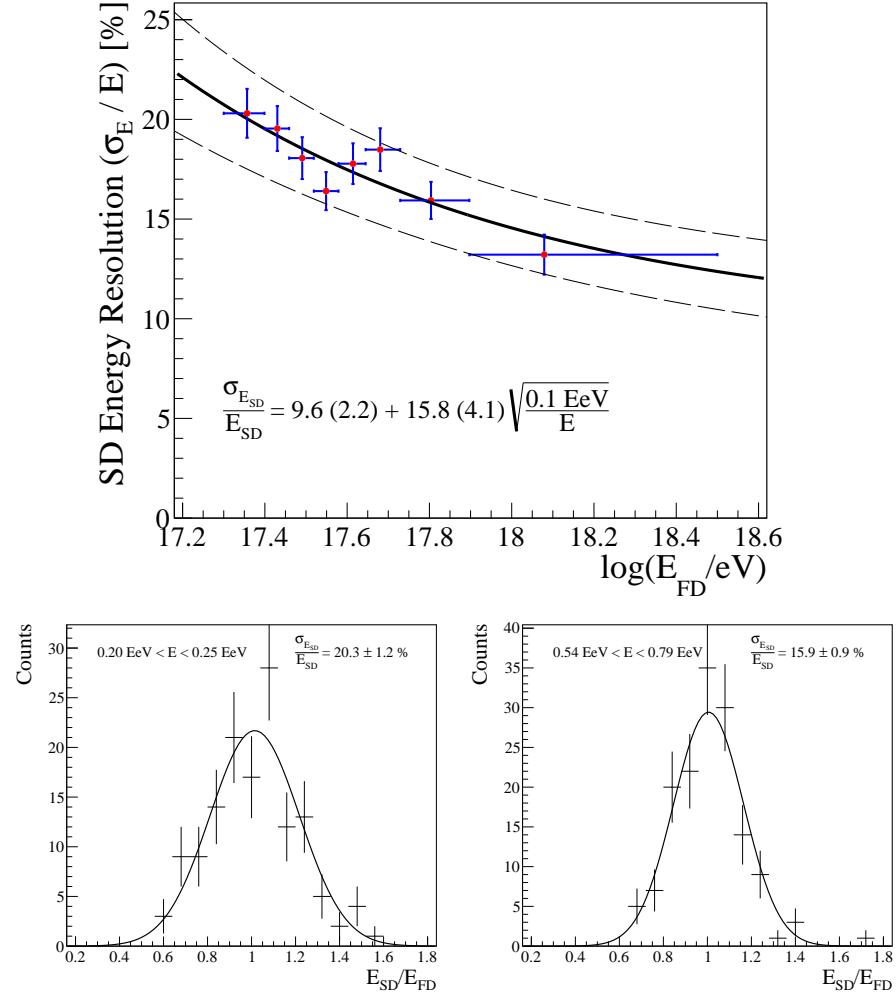


Figure 6.13: The energy resolution of the SD was determined using the golden hybrid events. Top: The values for $\sigma_{E_{SD}}$ were calculated in energy bins (blue crosses) and fit using an empirical function (solid black line). The dashed lines represent the uncertainty of the fit. Bottom: Examples of the energy-ratio distributions are shown for two of the energy bins.

points are the average energy within each bin and the horizontal error bars indicate the bin's width. The vertical error bars are the uncertainty on $\sigma_{E_{SD}}$ from the fit.

⁸Since the FD directly observes the shower development, it is not subject to shower-to-shower fluctuations and the total statistical energy uncertainty can be estimated for each event.

An empirical shape was chosen to describe the energy resolution as a function of E_{FD} ,

$$\frac{\sigma_{E_{\text{SD}}}}{E} = 9.6(2.2) + 15.8(4.1)\sqrt{\frac{0.1 \text{ EeV}}{E}}. \quad (6.24)$$

The best fit is shown by the solid black lines with the one sigma uncertainty bands represented by the dashed line which was calculated using the errors given in parentheses above.

With this result, it is possible to work backwards to determine the contribution of the shower-to-shower fluctuations. Since the **SD**'s detector resolution, σ_{det} , can be estimated for each event, σ_{sh} would be given by,

$$\sigma_{\text{sh}}(E)^2 = \sigma_{\text{SD}}(E)^2 - \sigma_{\text{det}}(E)^2. \quad (6.25)$$

An attempt was made to calculate the shower-to-shower contribution using this method, however, with the current number of golden hybrid events, a meaningful determination of $\sigma_{\text{sh}}(E)$ was not possible.

6.5 Conclusion

The analysis chain presented in this chapter allows for an estimation of an air shower's energy using events observed by the **SD**. The three corrections detailed in the beginning of the chapter reduce the systematic uncertainty of the shower size (and ultimately the energy) estimation. The smallest of these is the correction due to the dynamically changing weather above the array. This has a maximal change of $\sim 5\%$ on the size estimation but on average has no net scaling of $S(450)$ since it relates the observed shower sizes to what would be measured on average over a year. The correction with the strongest impact is the **CIC** which affects showers of all sizes. At high angles, this correction can scale the shower sizes by a factor of 2 - 2.5.

The net impact of the three corrections described in this chapter is shown in figure 6.14 as a function of zenith angle and uncorrected showers size. At high energies, the correction is essentially due to the **CIC** alone while at low energies, it is a combination of the **CIC** and **LDF** bias correction, $\mathcal{B}(S(450), \theta)/f_{\text{CIC}}(\theta)$. Note that roughly half of the events (white \rightarrow blue) are scaled down in energy as a result

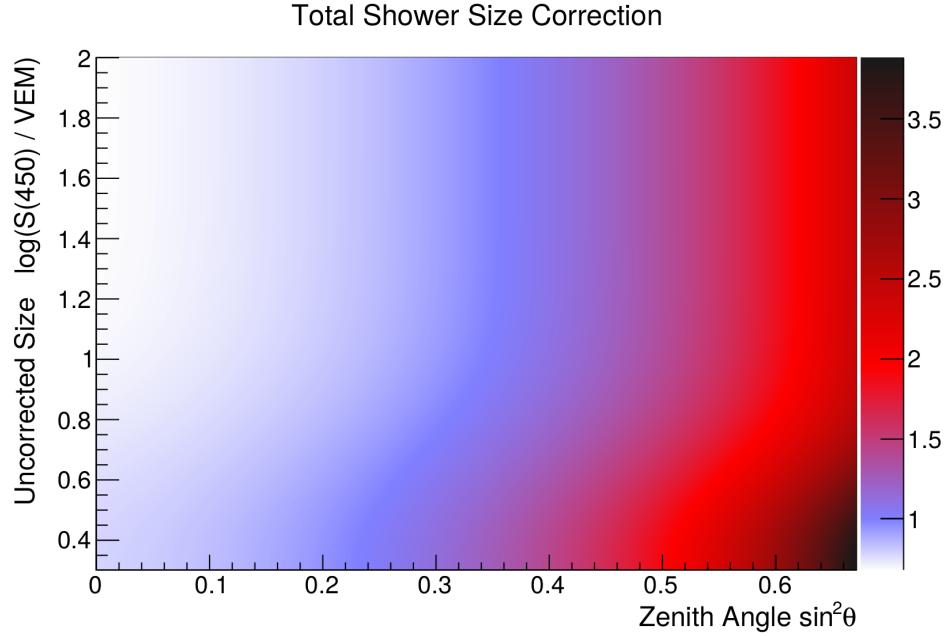


Figure 6.14: The plot above shows the total correction of the shower size ($S_{35}/S(450)$) depending on the zenith angle and uncorrected shower size. This is a combination of the [LDF](#) bias, weather, and [CIC](#) corrections for the $\theta_{\max} = 55^\circ$ data set.

of the choice of reference angle, $\sin^2 35^\circ \simeq 0.33$, for the [CIC](#).

One of the most important steps in the energy assignment is the calibration of the corrected shower sizes with the energy measurements from the [FD](#). This highlights another benefit of the hybrid detector scheme at Auger. By using only the ~ 1300 golden hybrid events to cross-calibrate, it is possible to make energy measurements that are self-consistent between detector systems and can be viewed with the 100% duty cycle of the [SD](#).

The accuracy of the energy assignment is set by the typical detector sampling resolution and the stochastic fluctuations of [EASs](#) during development. For the relevant energy range observed by the 750 m array, the energy resolution is 10 - 23% (largest at lower energies). The systematic uncertainty of the energy scale is comparable to this, 14 - 15%. This is dominated primarily by the calibration of the [FD](#), 10%. The cross-calibration with the [SD](#) results in a total of 6% to the [SD](#)'s energy scale uncertainty. In chapter 8, these uncertainties will be used to unbias the observed spectrum from detector effects and to determine the systematic uncertainty on the spectrum measurement.

Chapter 7

The Efficiency and Exposure of the 750 m Array

The cosmic ray flux, J , is the differential number of cosmic rays, dN , that arrive per energy (dE), time (dt), area (dA), and solid angle ($d\Omega$)

$$J(E) = \frac{dN}{dE dt dA d\Omega} = \frac{dN}{dE d\mathcal{E}}. \quad (7.1)$$

The final three factors can be absorbed into a single quantity describing the differential exposure, $d\mathcal{E} = dt dA d\Omega$.

In practice, a measurement of this differential spectrum is done by counting the number of events, N , in integral bins (centered at energy, E_i , with width, ΔE) and dividing by the exposure of the array during the data-taking period, \mathcal{E}

$$J(E_i) = \frac{N_i}{\Delta E_i \mathcal{E}}. \quad (7.2)$$

For a meaningful calculation of the flux, it is therefore required to accurately determine the array exposure.

For the symmetrically spaced stations which comprise Auger's **SD**, the calculation can ultimately be reduced to a simple geometry problem dependent on the acceptance angle, surface area, and live-time of the array. However, such a calculation is only valid for zenith angles, energies, etc. where the array is fully efficient and a shower is guaranteed to pass the T4 conditions¹.

¹Ultimately, such a guarantee can never be made in full confidence and this work will adopt a

While it is still possible to work below the full efficiency limit, any uncertainty in the triggering probability of the array will be propagated to the exposure and ultimately will increase the systematic uncertainty of a flux measurement. Since showers initiated by different primaries will not produce the same composition of secondaries, the triggering efficiency below full efficiency is primary-mass dependent. Given the uncertainties of the cosmic ray mass distribution, is it prudent to work in the conservative limit for which the spectrum will be largely unchanged when interpreted using different hadronic models².

With this motivation, the work presented in this chapter will cover two topics. The first is an estimation of the T4 triggering probability of the 750 m array using the TOTd and MoPS triggers. This includes two independent studies, the first (section 7.1.1) uses the LDF and T2-triggering efficiency parameterizations that were previously presented. The second is based on simulations of EASs using modern hadronic models, see section 7.1.2. The second topic of this chapter is the calculation of the geometric exposure which is presented in section 7.2.

7.1 Triggering Efficiency of the 750 m Array

Determining the efficiency of a cosmic ray detector is an inherently difficult problem as it is a measure of how many cosmic rays are *not* being detected. However, for the SD arrays it is one of the most important parameters to understand since it can strongly impact (and has been implicitly used in) the CIC, FD calibration, and any other studies that are sensitive to triggering effects.

This section contains two simulation studies to find a model describing the efficiency, ϵ , of the 750 m array using new triggers. In both, the zenith and energy parameter space is explored by generating EASs and applying the T4 algorithm.

The first study includes a toy-model MC analysis which uses the previously parameterized station-trigger probability along with the updated LDF shape. While this study is simplistic in nature, it provides a calculation of ϵ that depends only on models derived from real air showers.

The second study involves using modern hadronic models to fully simulate air

more realistic limit defined by the 98% triggering probability threshold.

²Only with very severe energy cuts can this be completely avoided. However, the primary-mass dependent effects, with reasonable cuts produce a $\leq 1.5\%$ change in the flux (see section 8.5).

shower cascades. These hadronic models are known to disagree in the UHE range, both with each other (see figure 2.3) and with observed data [87, 88]. However, using the CIC curve and the FD energy calibration from the previous chapter, these effects can be mitigated to produce more reliable results.

The methodology of the two studies are shown in sections 7.1.1 and 7.1.2, respectively. Section 7.1.3 includes a parameterization of $\epsilon(E, \theta)$ for both of the models and a brief comparison. Results with the new triggers turned off can be found in appendix A.4.

7.1.1 Efficiency Study Using T2 Triggers

As a first estimate of the array's T4 efficiency, a study was conducted using a toy model. First, an array with 750 m spacing was generated. The array was six crowns large (see figure 4.3), slightly larger than Auger's 750 m array. Next, the energies and trajectories of showers were randomly chosen. The zenith angles were chosen randomly on a range of 0° - 55° with a distribution, $dN/d\theta \propto \sin \theta \cos \theta$, to match the assumed UHECR flux (see section 6.3). The shower energies were likewise taken from a distribution³, $dN/dE \propto E^{-1}$, on a range from 10^{16} - 10^{18} eV. Finally, the core was randomly placed on the array with the constraint that the closest station to the core was not a border station (i.e. it would pass the 6T5 criteria).

The shower energies were converted into shower sizes using the FD calibration (equation 6.21) and the CIC (equation 6.13) tuned at 55° . The expected signal size, $S(r, \theta)$, for each station was calculated using via $S(r, \theta) = S(450) \text{LDF}(r, \theta)$. Then the stations were randomly triggered according to the T1/T2 probability calculated in section 4.3.2.

Finally, the T4 criteria (section 4.2.2) were applied to the simulated event. In this study, an approximation was used whereby the stations were not assigned a particular trigger type (i.e. TOT, TOTd, etc.). As a result, there is no distinction between the T4-3TOT and T4-4C1 algorithms. Further, no background particles were included in this simulation and thus the only pertinent T4 constraints were that the three stations are triggered in a triangular shape and that two of them were separated by 750 m.

³This choice of energy distribution was used such that there were the same number of events in bins of equal size in $\log(E)$.

A total of 5,000,000 showers were thrown on the array and the T4 triggering probability was calculated in bins of zenith angle and shower energy. This probability is given by the ratio of triggered (N_{T4}) and total (N_{Total}) events

$$\epsilon(E, \theta) = \frac{N_{\text{T4}}(E, \theta)}{N_{\text{Total}}(E, \theta)}. \quad (7.3)$$

The results are shown in figure 7.1 where the efficiency is plotted as a function of energy for ten zenith angle bins of equal size in $\sin^2 \theta$. The bottom plot shows lines

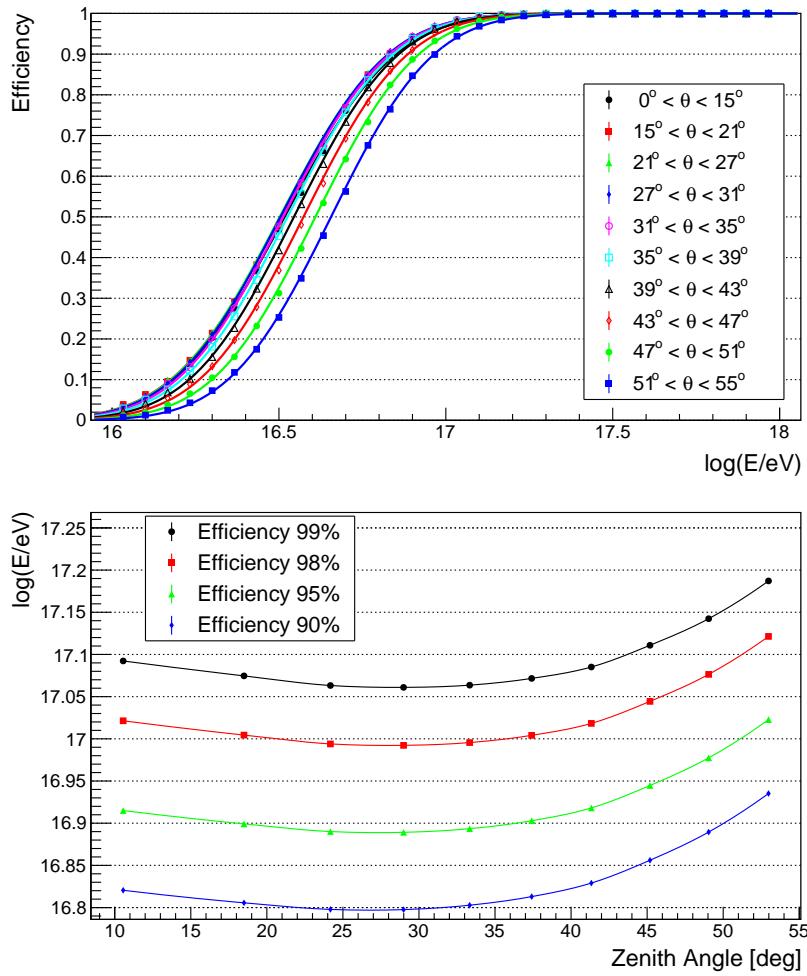


Figure 7.1: The plots above show the T4 efficiency of the 750 m array as calculated by a toy model. The top plot gives the efficiency as a function of energy in ten differential zenith angle bins of equal size in $\sin^2 \theta$. The bottom plot shows lines of constant efficiency for a number of values.

of constant efficiency in E - θ space for a number of efficiency values. The array is shown to be 98% efficient for energies above $10^{17.1}$ eV and zenith angles less than 50° .

The zenith dependence of the efficiency shows a trend that is consistent with what is expected from the [CIC](#) studies. Figure 6.6 showed how there are fewer events observed at higher zenith angles as a result of the triggering efficiency. This also makes physical sense since inclined showers will arrive at the array with a more attenuated electromagnetic component than vertical ones. The slight reduction in efficiency for the most vertical showers may be a product of reduced size of the shower footprint on the ground which scales to first order as $\propto \sec \theta$. For vertical showers, the footprints may be smaller than the distance between the stations.

7.1.2 Efficiency Study Using Hadronic Models

The array efficiency was cross-checked using air showers simulated using different hadronic models. The [EAS](#) development was propagated using the CORSIKA [89] framework. This software package simulates primary cosmic rays with user-chosen initial conditions (energy, zenith angle, mass, etc.) and calculates all of the relevant standard model interactions to produce a list of the secondaries and their properties on the ground⁴. To speed up computation time, the showers were *thinned*, meaning that during the simulation particles with an energy less than some fraction of E_{MC} , the energy of the primary, were combined using weights. This thinning was set to $E_{\text{MC}} \times 10^{-6}$.

The particles generated by CORSIKA were then injected into an array of [SD](#) stations with 750 m spacing. The Cherenkov-light production of each of the injected secondary particles was simulated using GEANT4 [90]. Digitized traces were then generated according to the observed response of the station electronics ([PMTs](#), [FADC](#), etc.). Finally, the hierarchy of trigger algorithms (T1→T4) was applied to the event.

For this study, 15,000 showers were simulated using CORSIKA for proton and iron primaries using QGSJetII-04 [59] and EPOS-LHC [57] as hadronic models. Further, each shower was injected onto the array ten times, each with a unique, random, core location, resulting in 600,000 total events. For T4 events, a recon-

⁴The altitude of the ground, local magnetic field, atmospheric profile, etc. can also be supplied by the user to produce a more realistic simulation.

struction using the new triggers was performed. This process resulted in six data sets, four for each of the model-primary combinations and two for each model, containing a 50/50 mix of proton and iron.

7.1.2.1 Conversion to the Auger Energy Scale

There are known discrepancies between the measurements of **EASs** and hadronic simulations. Such deviations between simulations and observed **EASs** have specifically been observed via an underproduction of muons in simulated cascades [87, 88]. Since a shower with more muons will create more signal on average in a **WCD**⁵, the energy scale of simulations is offset from that of data since a real and simulated shower with the same energy will not produce the same observed shower size on average. In light of this, it is important to be careful when choosing quantities with which to describe the efficiency. Specifically, it is vital to address the difference in the **MC/FD** energy scales. If the average signal on the ground is not the same for a simulated shower with energy $E_{\text{MC}} = 1 \text{ EeV}$ and a real shower with measured energy $E_{\text{FD}} = 1 \text{ EeV}$, then $\epsilon(E_{\text{MC}})$ is not a useful quantity to calculate.

A better quantity to parameterize the efficiency would be the shower size $S(450)$. This is a particularly good choice because it directly encapsulates what ϵ is trying to describe, the amount of signal needed to pass a T4 trigger. Because the conversion between $S(450)$ and E_{FD} has already been determined, the problem of differing energy scales can be resolved by calculating an equivalent **FD** energy for each event

$$E_{\text{MC}} \rightarrow S_{35,\text{MC}} \rightarrow S(450) \rightarrow S_{35,\text{Data}} \rightarrow E_{\text{FD}}. \quad (7.4)$$

Note that the subscripts will be used throughout this chapter to distinguish between the MC and FD/Data scales.

The conversion from E_{MC} to $S(450)$ was done using the same logic presented in chapter 6, but in reverse order. Given the large number of simulations, the **CIC** calibration was performed directly on the **MC** data following the steps in section 6.3. The parameters from this calibration are given in table 7.1 for the six data sets. The results of the 50/50 mix using QGSJet are shown as an example in figure 7.2. Due to the flat cut in maximum simulated energy, the impact of the atmospheric

⁵A muon will create more light in a station's water volume than an electromagnetic particle with the same energy and trajectory.

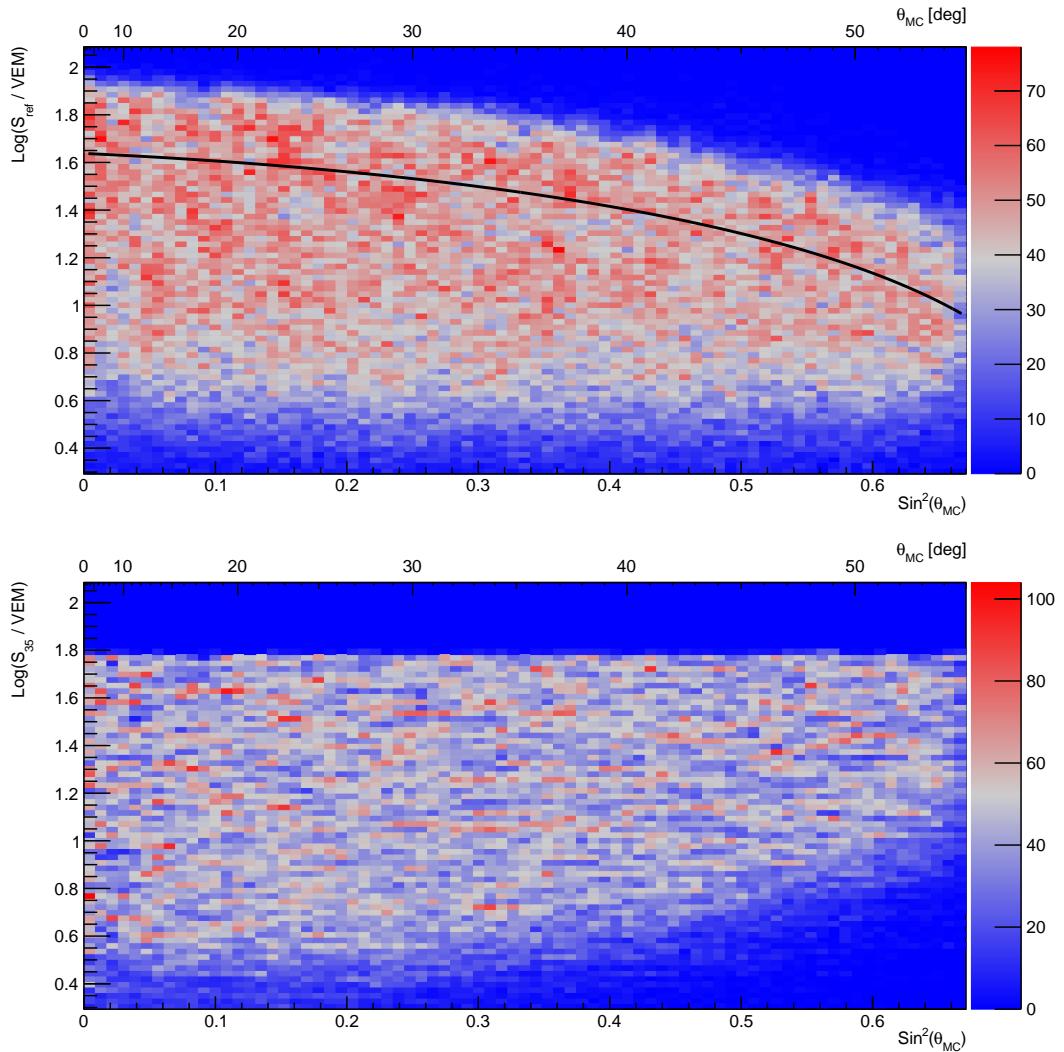


Figure 7.2: The **CIC** (section 6.3) method was applied to a 50/50 mixture of simulated proton and iron primaries using QGSJet as a hadronic model. Top: The distribution of measured shower sizes is shown as a function of zenith angle. The **CIC** was applied to the data resulting in the black curve, tuned to 30 VEM. Bottom: The distribution of attenuation-corrected shower sizes is shown as a function of zenith angle.

attenuation can clearly be seen by the sharp top-edge which decreases with zenith angle. The fitted **CIC** attenuation curve is shown by the black line.

The bottom plot shows the attenuation-corrected shower sizes, $S_{35,\text{MC}}$, as a function of zenith angle. The top of the distribution is flat as expected from the **CIC** method. As previously seen in the real data (figure 6.6), the bottom edge

of the distribution is highly curved and represents the border where the trigger efficiency of the array begins to decline. Unlike in the toy model study, a strong zenith angle dependence can be seen.

To complete the conversion to $S(450)$, an energy calibration was performed on each data set. For this, the likelihood was slightly changed to be a simple chi-squared minimization

$$\chi^2 = \sum_i^{E_{\text{MC}} > 0.3 \text{ EeV}} \left(\frac{S_{450}(E_{\text{MC}}, \theta_{\text{MC}}) - S(450)_i}{\sigma_{S_{450},i}^2} \right)^2 \quad (7.5)$$

where

$$S_{450}(E_{\text{MC}}, \theta_{\text{MC}}) = \text{CIC}(\theta_{\text{MC}}) \times A \left(\frac{E_{\text{MC}}}{1 \text{ EeV}} \right)^B. \quad (7.6)$$

Note that the fit was only performed on events with energy greater than 0.3 EeV to avoid trigger biases. As an example, the fit for a 50/50 mixture of QGSJet iron and

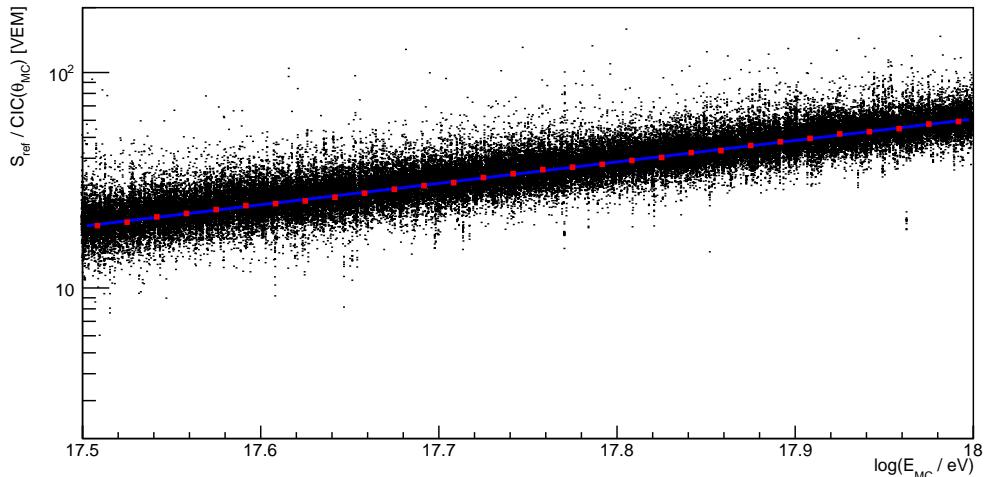


Figure 7.3: The calibration was performed on the MC data to derive a conversion between Monte Carlo energy and reference signal size. The plot above shows the observed showers sizes for a 50/50 mixture of iron and proton primaries using the QGSJet hadronic model. The individual events are shown as a function of $S_{35,\text{MC}}$ and E_{MC} (black points) and are also shown in energy bins for visual clarity (red squares). The fitted calibration curve (equation 7.6) is shown in blue.

proton primaries is shown in figure 7.3. The black points represent each calibration event. Because each CORSIKA simulation was thrown on the array ten times, this

Data Set	Constant Intensity Cut			Energy Calibration	
	a	b	c	A [VEM]	B
QGSJet P	1.74 ± 0.05	-1.69 ± 0.13	-1.78 ± 0.63	59.8 ± 1.0	0.98 ± 0.02
QGSJet Fe	1.82 ± 0.05	-0.36 ± 0.15	0.26 ± 0.70	61.7 ± 1.0	1.00 ± 0.02
QGSJet 50/50	1.79 ± 0.03	-1.03 ± 0.10	-0.84 ± 0.47	60.9 ± 0.9	1.00 ± 0.02
EPOS P	1.71 ± 0.05	-2.42 ± 0.12	-3.98 ± 0.67	67.3 ± 1.1	0.99 ± 0.01
EPOS Fe	1.90 ± 0.05	-0.02 ± 0.16	-0.54 ± 0.75	66.7 ± 1.0	1.00 ± 0.02
EPOS 50/50	1.79 ± 0.04	-1.26 ± 0.10	-2.11 ± 0.51	67.2 ± 0.9	1.00 ± 0.02

Table 7.1: The CIC and energy calibration were performed on the various combinations of simulated data. For each hadronic model the analysis chain was repeated for proton and iron primaries and a 50/50 mix of the two. The calibration constants and their statistical errors are given.

causes groups of points to be vertically aligned. The best fit line for the data is shown by the blue line. For visual clarity, the data was also binned in energy bins shown by the red squares. The best fit parameters for all of the simulated data sets are given in table 7.1.

By combining the MC energy calibration and CIC with those calculated in the previous chapter, the conversion to E_{FD} can be calculated as

$$E_{\text{FD}}(E_{\text{MC}}, \theta) = A_{\text{FD}} \left[A_{\text{MC}} \left(\frac{E_{\text{MC}}}{1 \text{ EeV}} \right)^{B_{\text{MC}}} \times \left(\frac{\text{CIC}_{\text{MC}}(\theta)}{\text{CIC}_{\text{Data}}(\theta)} \right)^{B_{\text{FD}}} \right]. \quad (7.7)$$

The subscripts have been used here to distinguish the calibration parameters (A, B) for the MC and FD calibrations. The ratio of the MC and FD energies are shown in figure B.5.

7.1.2.2 Efficiency Calculation in the FD Energy Scale

Using equation 7.7, the Monte Carlo energy of each event was converted into an equivalent energy in the FD energy scale. Then $\epsilon(E_{\text{FD}}, \theta)$ was calculated in the same manner as was done for the toy model analysis (equation 7.3). As an example, the triggering probability is shown for a 50/50 mix of proton and iron primaries simulated with QGSJet in figure 7.4. The two plots are the differential (top) and integral (bottom) efficiency in ten zenith angle bins of equal size in $\sin^2 \theta$ and thirty bins of equal size in $\log(E)$.

The results look qualitatively similar to the triggering probability calculated

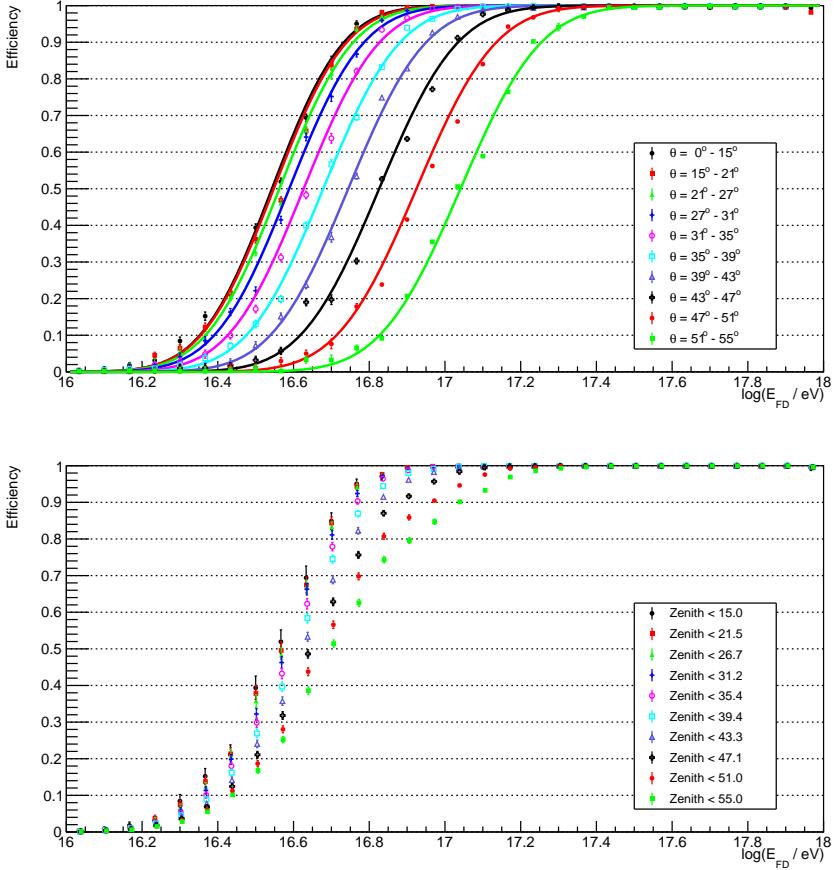


Figure 7.4: The array efficiency was calculated by simulating air showers using hadronic models. Both figures above are the results of a 50/50 mix of proton and iron primaries. The top plot shows the T4 triggering probability as a function of theta for ten differential zenith angle bins of equal size in $\sin^2 \theta$. The data was parameterized using an empirical function (see section 7.1.3) which is shown for each zenith angle bin by the solid lines. The bottom plot is the same triggering probability distribution but for integral zenith angle bins.

using the toy model (figure 7.1, top). The efficiency is best for more vertical showers and is worst at high angles. However, the differences between the most vertical and most horizontal bins are much larger for the results shown here, a factor of $\simeq 3$ in energy versus a factor of $\simeq 1.5$.

Unlike in the toy model study, the simulations using hadronic models allows for a study of the array efficiency as a function of cosmic ray mass. As described in section 2.2, EASs initiated by hadronic primaries produce a non-linearly increasing

number of muons as a function of the atomic mass. Given the reduced sensitivity of the **TOTd** and **MoPS** triggers to muons, this could produce a difference in the triggering efficiency for the different primaries. Following the same arguments, the efficiency can also disagree between the two models as they are known to produce differing numbers of muons on the ground [87, 88]. Some evidence of this can be seen in the calibration constant A in table 7.1. This parameter gives the expected value of $S(450)$ for a shower with energy 1 EeV. EPOS, which is known to produce more muons than QGSJet, has values of A that are 10% larger.

A comparison of the efficiencies for the various model-primary combinations is shown in figure 7.5. The top two plots show the agreement between the two

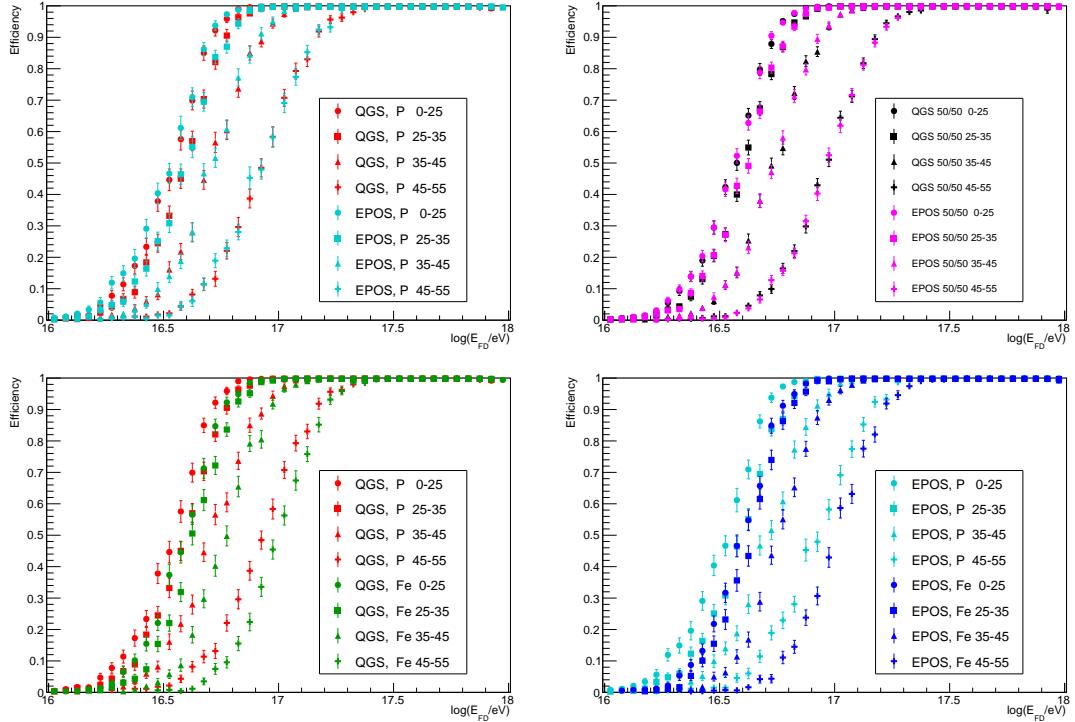


Figure 7.5: The plots above show comparisons of the differential efficiency for the various data sets. The top plots compare the two hadronic models, the left for proton primaries and the right for a 50/50 mixture of proton and iron. The bottom plots compare the efficiency of proton and iron for QGSJet (left) and EPOS (right)

hadronic models for proton primaries (left) and a 50/50 mix of proton and iron (right). In these comparisons, the efficiency results of QGSJet and EPOS are essentially indistinguishable. This is fortunate as it removes the need to choose one model over the other. The bottom two plots show the comparison between

proton and iron primaries for QGSJet (left) and EPOS (right). Here the agreement is not as good and the array seems to be more efficient in detecting proton showers. As described before, this is expected due to the reduced muonic content of [EAS](#)s produced by low-mass, hadronic primaries. Since the two primaries do not produce the same efficiency pattern and the true mass composition of [UHECR](#)s is unknown, the full efficiency limit will be determined by the most conservative result.

7.1.3 Parameterization of the T4 Efficiency

From the results above the energy and zenith angle limits for 98% efficiency can already be roughly determined. However, it will become useful to have an analytical model of the triggering probability for the creation of the energy spectrum. The differential distributions for all of the efficiency studies in this chapter fit to normalized error functions

$$\epsilon(E_{\text{FD}}, \theta) = \frac{1}{2} \left(1 + \text{Erf} \left(\frac{\log E_{\text{FD}} - \mu(\theta)}{\sigma(\theta)} \right) \right). \quad (7.8)$$

The variable $\mu(\theta)$ represents the energy at which the efficiency reaches 50% and $\sigma(\theta)$ controls the rate at which the efficiency asymptotes to unity, both of which are functions of theta

$$\begin{aligned} \mu(\theta) &= \mu_0 + \mu_1 \sin^2(\theta) + \mu_2 \sin^4(\theta) + \mu_3 \sin^6(\theta) \\ \sigma(\theta) &= \sigma_0 + \sigma_1 \sin^2(\theta). \end{aligned} \quad (7.9)$$

The fit was performed using the same chi-squared analysis as was done for the T1/T2 efficiency study presented in section 4.3.2. The data was split into 30 bins of equal size in $\log(E/eV)$ and then subdivided into 10 zenith bins of equal size in $\sin^2 \theta$. The best fit values were then found using the Wilson estimator and confidence interval, $\hat{\epsilon} \pm \Delta_{\epsilon}$, (see equation 4.4)

$$\chi^2 = \sum_i \left(\frac{\hat{\epsilon}_i - \epsilon(E_{\text{FD}}, \theta)}{\Delta_{\epsilon,i}} \right)^2. \quad (7.10)$$

The resultant values of $\{\mu_i, \sigma_j\}$ are given in table 7.2 for the six hadronic-model data sets and the toy model study. The curves described by these values have been overlayed on the differential efficiency plots in the top panels of figures 7.1 and 7.4.

Data Set	μ_0	μ_1	μ_2	μ_3
QGSJet P	16.521 ± 0.010	0.00 ± 0.12	0.82 ± 0.42	0.60 ± 0.41
QGSJet Fe	16.578 ± 0.009	-0.11 ± 0.12	1.11 ± 0.40	0.47 ± 0.40
QGSJet 50/50	16.540 ± 0.007	-0.03 ± 0.09	0.84 ± 0.31	0.68 ± 0.30
EPOS P	16.482 ± 0.011	0.25 ± 0.14	0.24 ± 0.48	1.01 ± 0.47
EPOS Fe	16.593 ± 0.009	-0.17 ± 0.12	0.90 ± 0.42	0.96 ± 0.41
EPOS 50/50	16.539 ± 0.007	0.01 ± 0.09	0.61 ± 0.31	0.97 ± 0.31
Toy Model	16.520 ± 0.001	-0.19 ± 0.02	0.40 ± 0.05	0.36 ± 0.05

Data Set	σ_0	σ_1
QGSJet P	0.205 ± 0.006	0.074 ± 0.015
QGSJet Fe	0.212 ± 0.006	-0.003 ± 0.014
QGSJet 50/50	0.209 ± 0.004	0.042 ± 0.011
EPOS P	0.218 ± 0.006	0.036 ± 0.015
EPOS Fe	0.209 ± 0.006	0.007 ± 0.014
EPOS 50/50	0.212 ± 0.004	0.028 ± 0.011
Toy Model	0.361 ± 0.001	-0.040 ± 0.002

Table 7.2: The differential T4 efficiency distributions were fit as a function of energy and zenith angle, equations 7.8 and 7.9. The best fit values are given in the table above for the six hadronic-model simulations as well as the toy model.

Likewise, a comparison between the fits for the data sets in this chapter can be seen in figure 7.6 for two different angles. Both of the plots show the efficiency for showers with zenith angles less than θ_{\max} . As a reference, the analysis using hadronic models was repeated for the old triggers (see appendix A.4). A clear improvement can be seen, especially for vertical showers, when the new triggers are used.

Using these functions, it is possible to then find a good description of the threshold which defines full efficiency. In this work the full efficiency threshold will define the limiting energy and zenith angle for which any EAS that lands within the fiducial area (see the definition of 6T5 in section 4.2.2) will pass the T4 trigger 98% of the time. In principle, this is a two dimensional function, $g(E, \theta)$, which is defined via the parameterized efficiency functions as, $g(E, \theta) = \epsilon^{-1}(0.98)$. An example of such a curve was previously shown in the bottom plot of figure 7.1.

However, to simplify the analyses, flat cuts in zenith angle and energy will be used, at the cost of exposure. A maximum zenith angle, θ_{\max} , and minimum energy

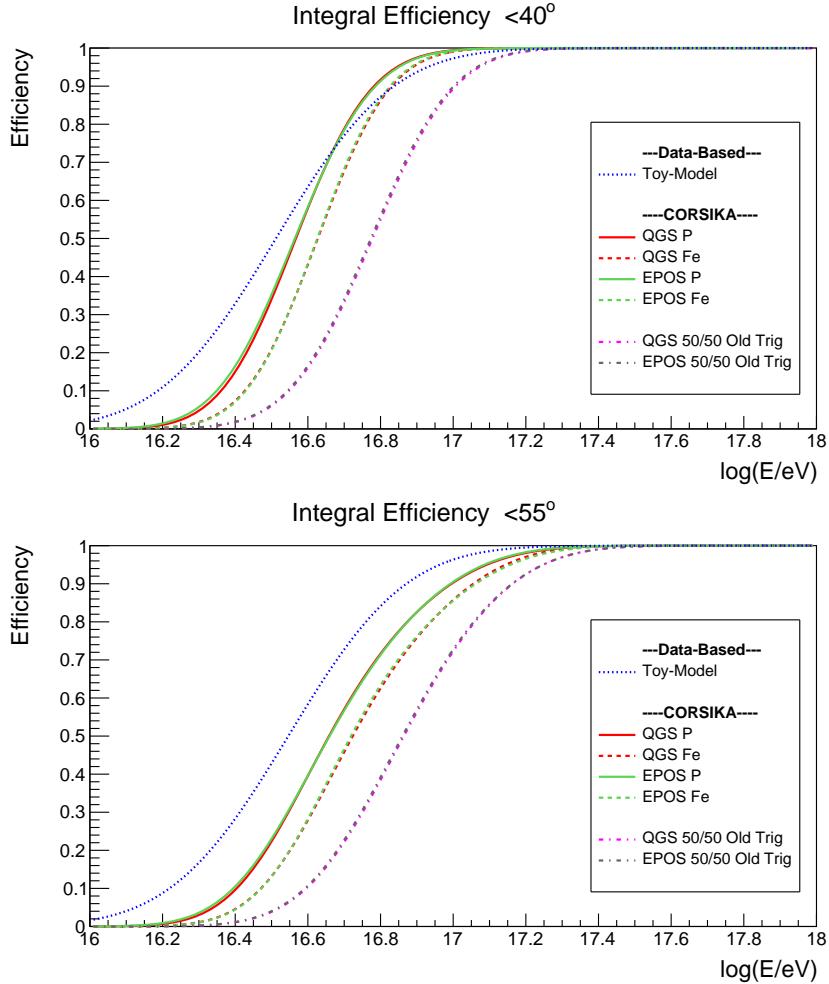


Figure 7.6: The plots above give the integral T4 trigger efficiency of the 750 m array using the new triggers for zenith angles less than 40° (top) and 55° (bottom). The distributions shown are given by the parameterization in table 7.2 for the toy model (blue dotted) and the hadronic model simulations (solid and dashed). The results for the old triggers (see appendix A.4) are also given as a baseline comparison (dot-dashed).

threshold, E_{\min} , can be selected by hand with the constraint that the integral efficiency is greater than 98% for $\theta < \theta_{\max}$ and $E > E_{\min}$. This makes the choice somewhat arbitrary as many choices of θ_{\max} and E_{\min} will fulfill this requirement. However, a particular choice can be well motived via a cut which maximizes the scientific value of the desired analysis. For an analysis of the cosmic ray energy spectrum, such a cut would be the one with the minimum energy threshold.

Two sets of threshold cuts, defining two semi-joint data sets, were selected

for analyses in this thesis. The first cut was chosen to coincide with the zenith angle range of the 750 m array that has been previously used in Auger publications, $\theta_{\max} = 55^\circ$ and $E_{\min} = 10^{17.3}$ eV. Note that this energy cut is an improvement from the previous limit when using only the old triggers, $10^{17.5}$ eV. The second set of cuts was chosen to significantly increase the energy range of the 750 m array, making use of the increased efficiency from the new triggers, $\theta_{\max} = 40^\circ$ and $E_{\min} = 10^{17}$ eV. The efficiencies for both of these zenith angle ranges were shown in figure 7.6.

7.2 Array Exposure

The exposure, \mathcal{E} , of the SD array is given by the time integration of the array's aperture. The aperture, \mathcal{A} , of the SD is proportional to the triggering efficiency so it is dependent on the shower energy and zenith angle

$$\frac{d\mathcal{E}}{dt} = \mathcal{A}(E, \theta) = A_0 \Omega \epsilon(E, \theta). \quad (7.11)$$

Here A_0 is the effective area of the array and Ω is the solid angle over which cosmic rays are observed. In this form it is easy to see the justification to work only in the parameter space where the array is fully efficient since the aperture is constant, $\mathcal{A}(E, \theta) = A_0 \Omega$. Additionally, this removes the need to rely on a well parameterized description of the triggering efficiency for which any uncertainty will be propagated to the exposure. The results from the previous sections should be treated as estimations so it is prudent to avoid having to rely on the models of $\epsilon(E, \theta)$ if possible.

The SD's effective area is well defined as a result of its repeating hexagonal layout combined with the 6T5 trigger (see section 4.2.2). Each station that passes the 6T5 requirement of having six immediate, data-taking neighbors, contributes a *cell* of area, A_{cell} . Such an exposure cell is shown schematically in figure 7.7. For a station spacing of λ , the cell area is given by $A_{\text{cell}} = \frac{\lambda^2 \sqrt{3}}{2} = 0.487 \text{ km}^2$ for the $\lambda = 750 \text{ m}$ array⁶.

By integrating this over the field of view, the aperture of a cell can then be

⁶The 1500 m array would thus have $A_{\text{cell}} = 1.949 \text{ km}^2$.

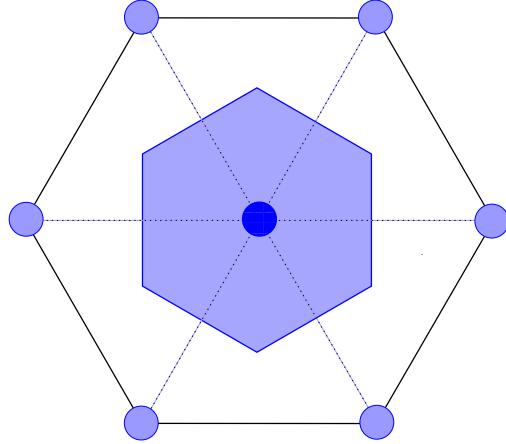


Figure 7.7: The diagram above represents the effective area of one hexagon as defined by the 6T5 triggers. The central dark blue circle represents the station with the largest signal in an event with the nearest neighbors being shown as the lighter-blue circles. The active area is then the blue hexagon.

calculated

$$\mathcal{A}_{\text{cell}} = A_{\text{cell}} \int d\Omega \cos \theta = A_{\text{cell}} \int d\phi \int d\theta \cos \theta \sin \theta = \pi A_{\text{cell}} \sin^2 \theta_{\max}. \quad (7.12)$$

This yields $0.632 \text{ km}^2 \text{ sr}$ for $\theta_{\max} = 40^\circ$ and $1.027 \text{ km}^2 \text{ sr}$ for $\theta_{\max} = 55^\circ$. Finally, the aperture of the whole array can be calculated by simply summing over all of the 43 cells that make up the 750 m array. The maximum apertures for the array are thus $21.19 \text{ km}^2 \text{ sr}$ and $45.16 \text{ km}^2 \text{ sr}$, respectively.

Since stations occasionally will have some problem that causes them to become inactive, the total exposure is calculated as an integration over the live-time of the 43 potentially-active cells, j , during the data period being used

$$\mathcal{E} = \int \frac{d\mathcal{E}(t)}{dt} dt = \int \left(\sum_j^{\text{Cells}} \mathcal{A}_{\text{cell},j}(t) \right) dt. \quad (7.13)$$

In practice this integration is done by determining the working status of the stations via their reporting of T2 triggers to the [CDAS](#). A verification of this method has been performed by computing the event rate to the calculated exposure for different time intervals. Combined with the added uncertainties from the communication system, the overall uncertainty of the exposure is 4% [76]. Thus

equation 7.13 becomes a summation over time for each of the candidate active cells

$$\mathcal{E} = \sum_j^{\text{Cells}} \sum_i \mathcal{A}_{\text{cell},j}(t_i). \quad (7.14)$$

This summation as a function of time is given in figure 7.8. The plot gives the

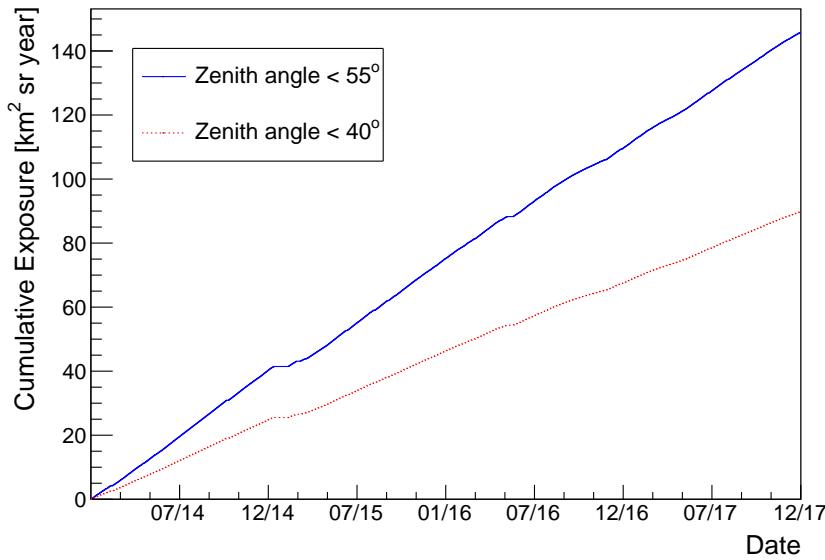


Figure 7.8: The integral exposure is shown for the 750 m array during 2014 - 2017 for 55° (solid blue) and 40° , (dashed red).

integral exposure of the array for the two data sets defined in the previous section. Given that both sets include the same effective area, the two zenith angle ranges differ by a scaling equal to the ratio of their respective solid angles, $\simeq 0.6$.

7.3 Conclusion

The analyses presented in this chapter comprise the work required to calculate the exposure for the 750 m array. The discussion in section 7.2 detailed the way \mathcal{E} can be simply calculated based on the number of active cells and the opening angle. This calculation is only valid where the array is fully efficient and thus most of this chapter was instead devoted to determining the T4 efficiency.

Below full efficiency, the calculation of the exposure can be done. However, it ultimately requires a detailed understanding of the efficiency for which any

error in its calculation would be propagated to the measurement of the flux. The efficiency studies in this chapter should be treated as estimations since they are all extrapolations. The toy model is based on an extrapolation of the LDF shape and FD energy calibration while the hadronic models themselves are based on extrapolations of collider data to higher energies. With this in mind, the decision to work above 98% efficiency is really a trade-off to minimize errors at the cost of viewing angle and energy range.

To keep the analysis rather straightforward two data ranges were selected which respect the full efficiency requirement. The first is efficient for showers with zenith angles less than 40° and energies above $10^{17.0}$ eV and the second is valid for zenith angles less than 55° and energies above $10^{17.3}$ eV. Both of these data sets will be used for physics analyses in the rest of this thesis.

Chapter 8

Measurement of the Energy Spectrum Above 100 PeV

The analyses in the previous sections included all of the detailed work that is required to incorporate the new station triggers, **TOTd** and **MoPS**, into the 750 m array. With a design that included a reduced sensitivity to the background muons, they allow for the observation of lower energy regions of the shower front that are normally obscured by the flux of background muons. This results in more stations being triggered and thus a better sampling of the **LDF**, but more importantly, the new triggers allow the **SD** to observe lower energy showers than was previously possible at Auger with full efficiency.

Using only the old triggers, the **SD** arrays were able to make measurements of the cosmic ray flux above $10^{17.5}$ eV. This energy range included the ankle observed at $\simeq 10^{18.7}$ eV and the suppression at $\simeq 10^{19.5}$ eV. With the full incorporation of the new triggers into the air shower reconstruction algorithm, this energy range has been extended by half an order of magnitude down to $10^{17.0}$ eV.

In this chapter, the measurement of the cosmic ray spectrum using the 750 m array will be presented. This will be done for both of the energy sets defined in the previous section (zenith angles $\theta < 40^\circ$ and $\theta < 55^\circ$). First the raw spectrum will be shown in section 8.1 along with a check for zenith angle dependencies. Next, a method will be presented to remove the detector effects to produce an unbiased flux measurement (section 8.2). Finally, a comparison between the new trigger spectrum and those previously measured by the 750 m array will be presented (section 8.3) as well as a comparison to the results of other experiments (section 8.6).

8.1 The Raw Spectrum

Using the exposure, \mathcal{E} , calculated in the previous chapter, the flux of cosmic rays as a function of energy, $J(E)$, is given by

$$J(E) = \frac{N}{\Delta E \mathcal{E}}. \quad (8.1)$$

Here, N , is the number of observed events in an energy bin with width, ΔE . The bin size was chosen such that the half-width is approximately equal to the total energy uncertainty, $\Delta E = 0.1$ in $\log(E/\text{eV})$.

To give a sense of the raw event counts in the two data sets, the number of events detected by the 750 m array is shown in figure 8.1 as a function of energy for years 2014 - 2017. The number of observed events first increases due to the

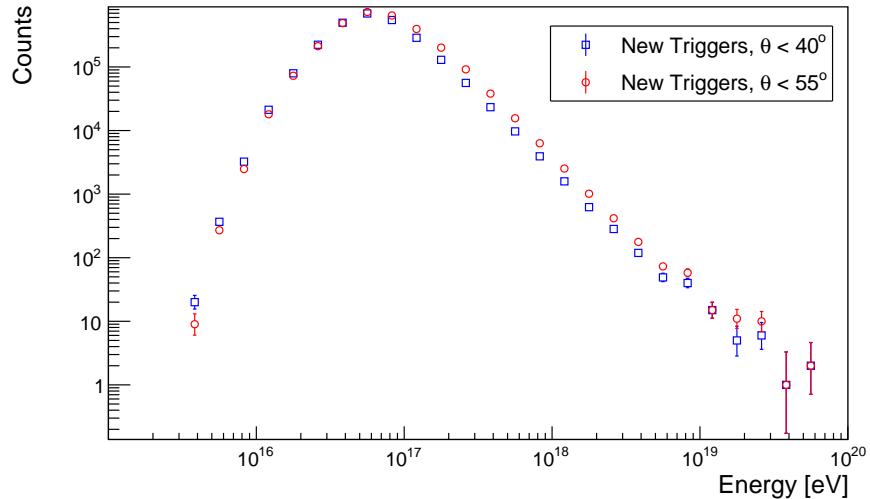
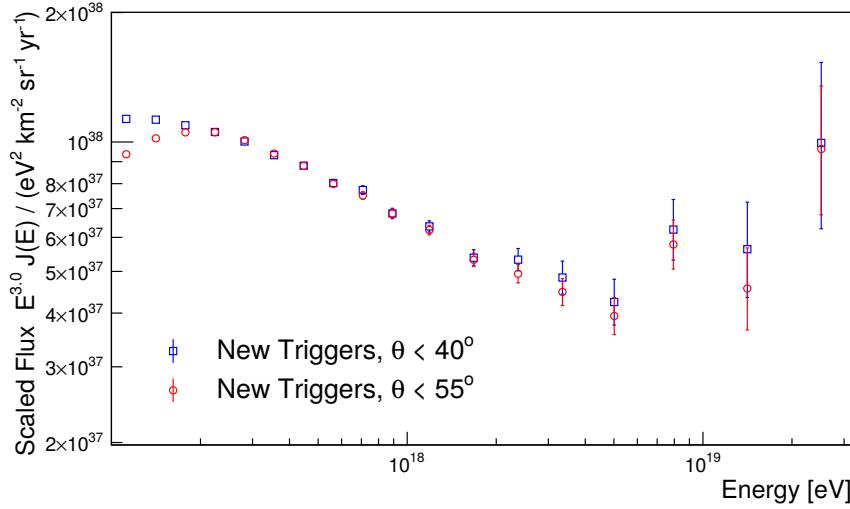


Figure 8.1: The raw number of counts detected by the 750 m array is shown in the figure above. The two data sets are shown with zenith angle cuts of $\theta < 40^\circ$ (blue squares) and $\theta < 55^\circ$ (red circles).

saturation of the trigger efficiency. Likewise, when the array becomes fully efficient, $\gtrsim 10^{17} - 10^{17.3}$ eV, the decreasing number of counts is the result of the steep, power law spectrum. The low number of counts above 3×10^{19} eV inhibits the possibility of observing the flux suppression, like can be done with the 1500 m array. However, the 750 m array does have enough exposure to observe the ankle at $\simeq 5 \times 10^{18}$ eV.

To create the spectrum, the number of events was divided by the bin width

and the constant exposure, $145.8 \pm 5.8 \text{ sr yr km}^2$ for the 55° data set and $89.8 \pm 3.6 \text{ sr yr km}^2$ for the 40° data set. The result is shown in figure 8.2 which has been scaled by a factor of E^3 to emphasize the spectral features. The two data sets are



in this thesis and have been corrected for via the LDF-tuning (section 5.2.3), the size-bias correction (section 6.1), and the CIC tuning (section 6.3). A comparison of the flux as a function of local zenith angle provides a good test of how well these corrections do to produce an unbiased energy assignment.

The top plot of figure 8.3 shows the spectrum in four zenith angle bins. Each

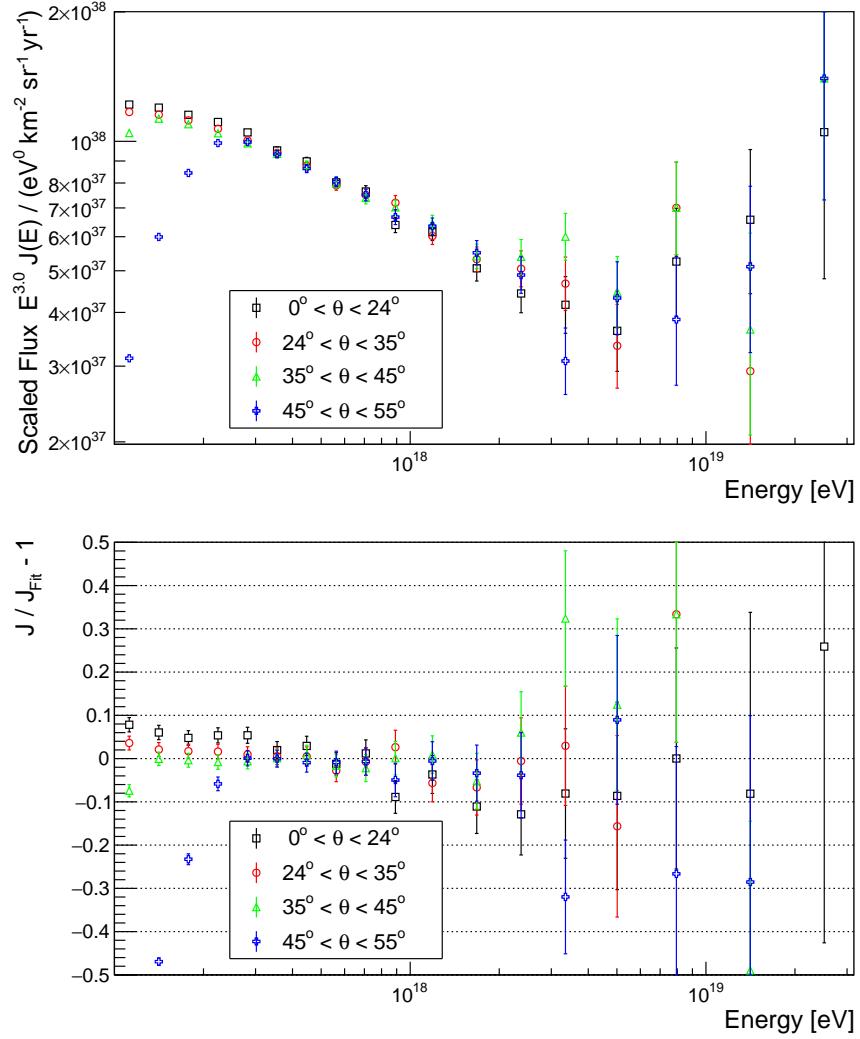


Figure 8.3: Top: The raw spectrum is shown for four zenith angle bins of equal exposure. Bottom: The residuals of the measured fluxes for the zenith angle bins are shown with respect to a fit function (equation 8.2).

bin is of equal size in $\sin^2 \theta$ such that the exposures are the same. The effects of the triggering efficiency can clearly be seen by the departure of the most inclined bin (blue crosses) from the other distributions below $\simeq 2 \times 10^{17}$ eV. As shown in

section 7.1, zenith angles of 45° - 55° are not fully efficient for energies less than 1.5 - 2.5×10^{17} eV.

The agreement of the zenith angle bins is shown a bit more clearly in the bottom plot of figure 8.3. This plot gives the residuals of the flux with respect to an empirically chosen function that was fit to the 40° data set (blue squares in figure 8.2),

$$J(E) = J_0 \times \begin{cases} E^{-\gamma_0} \left(1 + \left(\frac{E}{E_{\text{Break}}}\right)^k\right)^{\frac{\gamma_0 - \gamma_1}{k}}, & E < E_{\text{Ankle}} \\ \text{Norm} \times \left(\frac{E}{E_{\text{Ankle}}}\right)^{-\gamma_2} \left(1 + \left(\frac{E}{E_{\text{Supp}}}\right)^{\Delta\gamma}\right), & E \geq E_{\text{Ankle}}. \end{cases} \quad (8.2)$$

This function includes three energies, E_{Break} , E_{Ankle} , and E_{Supp} which mark where the spectral index changes, $\gamma_0 \rightarrow \gamma_1 \rightarrow \gamma_2 \rightarrow \gg \gamma_2$, respectively. The equation describes a sharp break at E_{Ankle} and is normalized such that the spectrum is continuous. The other breaks at E_{Break} and E_{Supp} are smooth and the rate of change of the spectral index is controlled by k and $\Delta\gamma$, respectively. This equation cannot be fully constrained in the $\gtrsim 10^{19}$ eV range by the 750 m array so the suppression energy and spectral index were fixed to the values measured by the 1500 m array [26]. However, for the time being, the fitted parameters are not important since the raw spectrum still includes many detector effects and this is just being used as a reference.

The residuals as a function of zenith angle are completely compatible above 3×10^{17} eV. For these energies all of the zenith angle bins are fully efficient and there are no signs of a systematic bias. Below this energy, the three most vertical bins show some systematic differences for which the spread is constrained within ± 3 - 4% . To understand what this means with respect to the energy assignment, note that the spectrum is measured in integral bins. Thus for a pure power-law spectrum (which is approximately the case for any individual bin) the relative scaling of the flux is given by

$$\frac{dJ}{J} = \frac{dN}{N} = \frac{d(\int E^{-\gamma} dE)}{\int E^{-\gamma} dE} = (\gamma - 1) \frac{dE}{E}. \quad (8.3)$$

Assuming that near 3×10^{17} eV the spectral index is $3 \lesssim \gamma \lesssim 3.3$, then the $\simeq 3.5\%$ separation in the flux in the various zenith bins can be accounted for by a $\simeq 1.5\%$ shift in the energy assignment. The uncertainty on the CIC correction is as large

as 2% at 0° and 55° . For example, this separation may be a result of the [CIC](#)'s energy dependence which was tuned at $10^{17.6}$ eV (see section [A.5](#)). Whatever the cause, the vertical ordering of the spectra with zenith angle indicates that there may still be some lingering bias at the lowest energies.

8.2 The Unfolded Energy Spectrum

The observation of an [EAS](#) by the [SD](#) includes statistical uncertainty from the stochastic development of the shower, the production of light in the [PMTs](#), and the estimation of the size during the reconstruction (including the assignment of $S(450)$, the [CIC](#), etc.). These fluctuations cause events to migrate between energy bins and obscures the true spectrum. While some events will migrate to a lower-energy bin, the steep, power-law shape of the spectrum causes the net migration to be asymmetric (favoring higher energies). The combination of this and triggering effects produces an observed spectral shape that differs from that of the true spectrum in a non-trivial way.

An example of this effect is shown in figure [8.4](#) for a toy model detector with similar characteristics to that of the 750 m array. The top plot shows an example spectrum (solid red line) which includes a smooth break at 10^{17} eV, a sharp break at the ankle and a smooth rolloff above $10^{19.5}$ eV. Additionally two spectra are shown which represent what would be observed by a detector with a finite energy resolution. The solid blue line is the observed spectrum if the detection efficiency is ideal while the dashed blue line includes efficiency effects. The corresponding energy resolution and detection efficiency are shown in the bottom plot in dotted blue and solid red, respectively. The observed spectra show a clear indication of a steepening which is most pronounced at low energies. Both black lines in the bottom plot give the ratio of the true and observed spectra if the detection efficiency is included (solid black) or ignored (dashed black).

The raw spectra shown in the previous section have these detector effects folded into them. Without correcting for this, it is not possible to accurately determine the spectral features and indexes. The next sections will describe a method by which the finite energy resolution and triggering efficiency effects can be unfolded from the spectrum measurement.

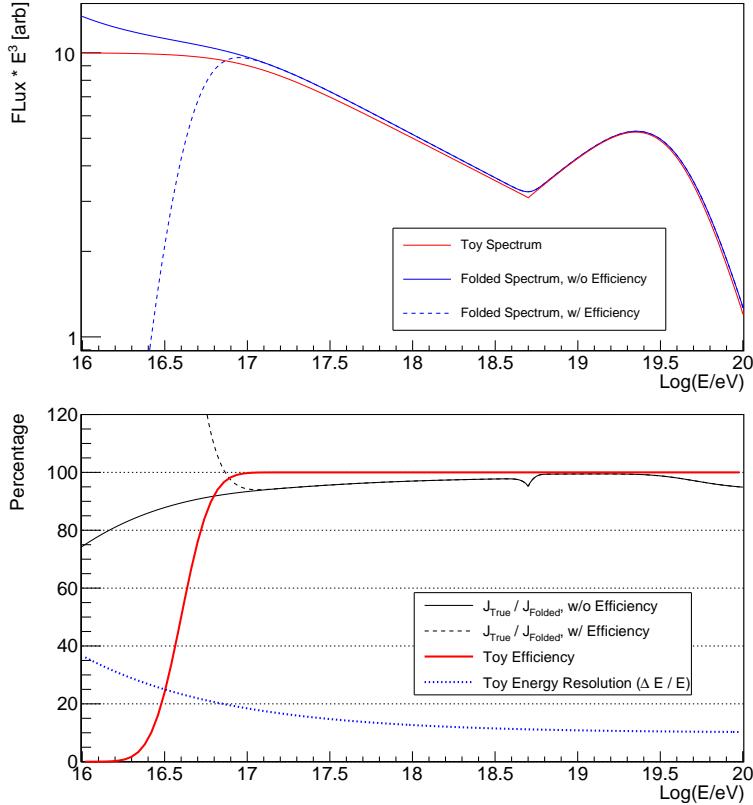


Figure 8.4: Top: An example toy spectrum is shown in red along with the spectra that would be observed by a detector with finite energy resolution. The solid blue line is the spectrum for a detector with an ideal triggering efficiency while the dashed blue line shows the observed spectrum for a detector with an efficiency comparable to that of the 750 m array. Bottom: The energy resolution and detection efficiency of the toy detector are shown by the dotted blue and solid red lines, respectively. The black lines indicate the net result of the event migration, the ratio of the true and observed spectra.

8.2.1 The Forward Folding Procedure

The folded energy spectrum for a surface array, $J_{\text{Folded}}(E)$, is a convolution of the true spectrum, $J_{\text{True}}(E)$, the detector's energy resolution, R , and the triggering efficiency, ϵ ,

$$J_{\text{Folded}}(E) = \int dE' \int d\mathbf{p} J_{\text{True}}(E') \mathcal{F}[E, R(E'), \mathbf{p}] \epsilon(E', \mathbf{p}). \quad (8.4)$$

Here, \mathcal{F} , is a response function which describes the probability that events with true energy, E' , will be assigned an energy, E . In general, this function and the trigger efficiency can both be dependent on a number of other parameters, \mathbf{p} , such as zenith angle, time, etc. However, for Auger, this parameter space is well described by the zenith angle and energy, $R(E)$ and $\epsilon(E, \theta)$ (see section 7.1.3). The energy resolution has been characterized using the golden hybrid events (see section 6.4.3) for which the statistical fluctuations of SD energies are consistent with a Gaussian distribution. So the smearing function can be replaced by, $\mathcal{F}[E, R(E'), \mathbf{p}] \rightarrow G[E, E', R(E')]$, where G is a Gaussian with mean, E' , and standard deviation, $R(E')$. So for the SD array, the folded spectrum is thus given by

$$J_{\text{Folded}}(E) = \int_0^\infty dE' J_{\text{True}}(E') G[E, E', \sigma(E')] \epsilon(E') \frac{\int_0^{\theta_{\max}} d\theta \epsilon(E', \theta) \sin \theta \cos \theta}{\int_0^{\theta_{\max}} d\theta \sin \theta \cos \theta}. \quad (8.5)$$

Since the only non-energy parameter remaining from equation 8.4 is the zenith angle, the integration over $d\mathbf{p}$ is replaced by $d\theta$, given in the bottom equation. The zenith dependence of the array's efficiency has been convolved with the assumed isotropic distribution of arrival directions (see section 6.3).

To extract the true spectrum, equation 8.5 must be inverted. However, the discreteness of the flux measurements make this impossible since only changes in $J_{\text{True}}(E)$ that are on the length scale of the spectrum binning can be resolved. Instead, a forward-folding procedure will be used. This method folds the detector effects into a supplied model for the true spectrum (the choice and impact of a particular model will be discussed in section 8.2.2) using equation 8.5 then compares the folded spectrum to the observed spectrum. The number of expected (observed) counts, N_{exp} , in a bin with center E_i can be calculated by solving for N in equation 8.1 and integrating over the bin width, ΔE ,

$$N_{\text{exp}}(E_i, \Delta E_i) = \mathcal{E} \int_{E_i - \frac{1}{2}\Delta E_i}^{E_i + \frac{1}{2}\Delta E_i} dE' J_{\text{Folded}}(E'). \quad (8.6)$$

Note that the inclusion of the triggering efficiency in the determination of the folded spectrum means that N_{exp} is valid for any energy, not just where the array is fully efficient.

For each spectrum bin, N_{exp} was calculated using the forward folding procedure, equations 8.5 and 8.6, then compared to the observed number of counts using Poissonian statistics. This produces the log-likelihood function

$$\ln \mathcal{L} = \sum_i^{\text{Energy}} N_{\text{exp},i} - N_{\text{obs},i} \ln N_{\text{exp},i}. \quad (8.7)$$

Using the best fit values, the spectrum measurements were unfolded by scaling them with a correction factor, $C(E)$, which is defined as the ratio of the true and folded spectra

$$C(E) = \frac{J_{\text{True}}(E')}{J_{\text{Folded}}(E')}. \quad (8.8)$$

An example of a correction factor can be seen in black in the bottom plot of figure 8.4 for the toy model.

8.2.2 Impact of the Spectral Model on the Unfolding

The forward-folding procedure requires a model of the spectral shape into which the detector effects can be folded. As shown in the toy model spectrum at the beginning of this section, the true and folded spectra look similar. For instance, the finite energy resolution smooths out the sharpness of the ankle, but does not completely obscure it². So a pragmatic choice of a spectral model would be one that resembles the raw spectrum using the fewest number of parameters. As a test, the spectrum unfolding procedure was performed using three models for $J_{\text{True}}(E)$.

From figure 8.2 it is clear that the raw spectrum includes a negative spectral index that is changing near 100 PeV. Further, the spectrum has negative curvature below 1 EeV. As shown in the toy model (figure 8.4), the unfolding will make the curvature *more* negative, so a single power-law will not be able to describe the spectrum for $10^{17} \text{ eV} < E < E_{\text{Ankle}}$. Thus the three models will have to include a term to account for the change in spectral index at $E \lesssim 10^{17.5}$.

Model-1, $J_{\text{M-1}}(E)$, is a simple set of *three* broken power laws, with spectral

²The magnitude of how obscured sharp features are is anti-correlated with the energy resolution.

indexes $(\gamma_0, \gamma_1, \gamma_2)$ and breaks E_{Break} and E_{Ankle} ³

$$J_{\text{M-1}}(E) = J_0 \times \begin{cases} \left(\frac{E}{E_{\text{Break}}}\right)^{-\gamma_0}, & E < E_{\text{Break}} \\ \left(\frac{E}{E_{\text{Break}}}\right)^{-\gamma_1}, & E_{\text{Break}} \leq E < E_{\text{Ankle}} \\ \left(\frac{E_{\text{Ankle}}}{E_{\text{Break}}}\right)^{-\gamma_1} \left(\frac{E}{E_{\text{Ankle}}}\right)^{-\gamma_2}, & E_{\text{Ankle}} \leq E. \end{cases} \quad (8.9)$$

The data does not seem to follow a sharp break near 1 EeV, however, this model will be kept as a minimum-parameter benchmark and to study the sensitivity of the unfolding procedure. Model-2 is similar in shape, but includes a smooth roll-off at E_{Break} instead of the sharp one of Model-1

$$J_{\text{M-2}}(E) = J_0 \times \begin{cases} E^{-\gamma_0} \left(1 + \left(\frac{E}{E_{\text{Break}}}\right)^k\right)^{\frac{\gamma_0 - \gamma_1}{k}}, & E < E_{\text{Ankle}} \\ E_{\text{Ankle}}^{-\gamma_0} \left(1 + \left(\frac{E_{\text{Ankle}}}{E_{\text{Break}}}\right)^k\right)^{\frac{\gamma_0 - \gamma_1}{k}} \left(\frac{E}{E_{\text{Ankle}}}\right)^{-\gamma_2}, & E \geq E_{\text{Ankle}}. \end{cases} \quad (8.10)$$

Model-3 also includes a roll-off at E_{Break} but employs a different functional form

$$J_{\text{M-3}}(E) = J_0 \times \begin{cases} \left(\frac{E}{E_{\text{Break}}}\right)^{\frac{\gamma_0 + \frac{\gamma_1 - \gamma_0}{\Delta}}{1 + \left(\frac{E}{E_{\text{Break}}}\right)^{\Delta}}}, & E < E_{\text{Ankle}} \\ \left(\frac{E_{\text{Ankle}}}{E_{\text{Break}}}\right)^{\frac{\gamma_0 + \frac{\gamma_1 - \gamma_0}{\Delta}}{1 + \left(\frac{E_{\text{Ankle}}}{E_{\text{Break}}}\right)^{\Delta}}} \left(\frac{E}{E_{\text{Ankle}}}\right)^{-\gamma_2}, & E \geq E_{\text{Ankle}}. \end{cases} \quad (8.11)$$

The results of the unfolding using these three models are shown in the top plot of figure 8.5. The error bars include the statistical errors on the number of counts and the unfolding factor (see figure B.6) added in quadrature and scaled by E^3 . The raw spectrum (grey triangles), as with the toy model example, shows a larger flux than that of the unfolded spectra. There are no large deviations between the three results which indicates that the unfolding procedure is rather insensitive to the choice of J_{True} .

To highlight this, the spectral points, unfolded using Model-2 (blue circles), were fit using the $J_{\text{M-2}}(E)$ shape (shown by the blue dashed line in figure 8.5, top). The bottom plot of figure 8.5 shows the ratio of the three unfolded spectra with respect to this fit. The largest deviation between any two is only a few percent and

³Since the suppression is not well observed by the data sets used in this thesis, the models will not include terms for this feature.

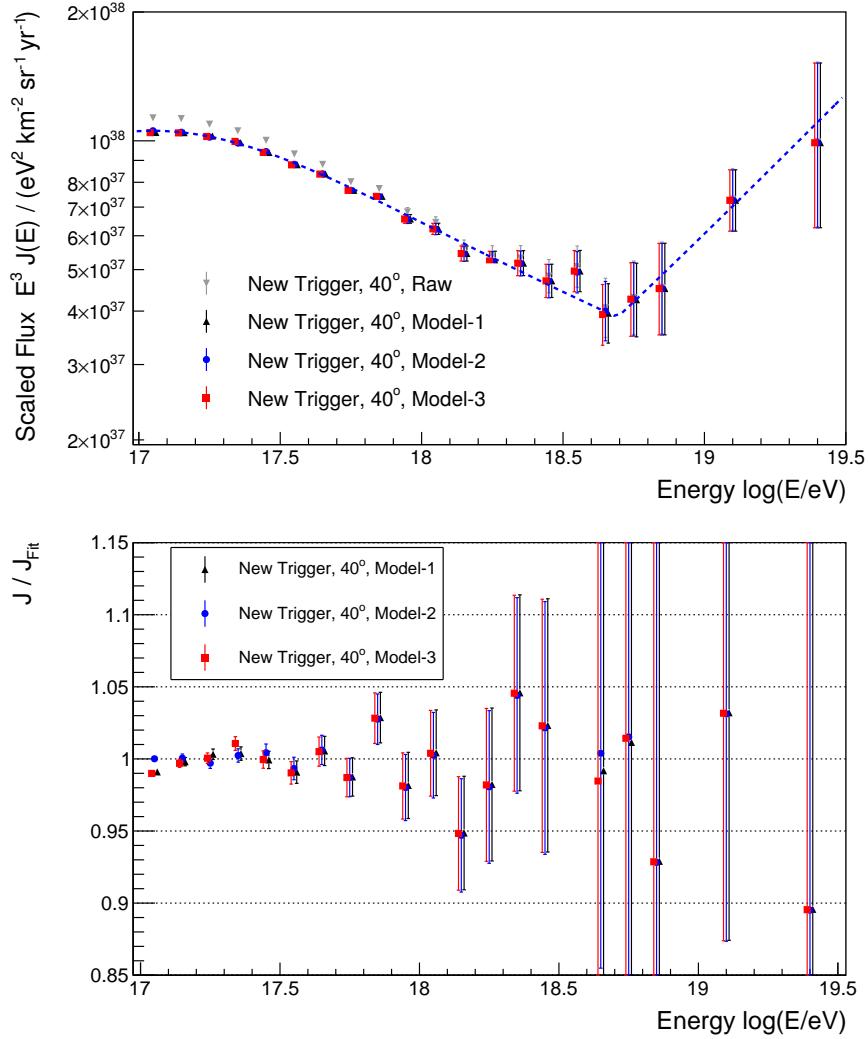


Figure 8.5: Top: The raw spectrum (grey triangles) is shown for showers with zenith angles less than 40° . The results of the unfolding using three models (equations 8.9, 8.10, and 8.11) are also shown, slightly offset for visual clarity. The blue dashed line is the result of fitting the Model-2 function to its respective data set (blue circles). Bottom: The ratio of the unfolded spectra with respect the fit line in the top plot is shown.

is largest near the two breaks.

Even though the various models produce spectra which are compatible to within a few percent, they do not necessarily describe the unfolded result equally well. To show this, an analysis was performed which fit each of the three sets of unfolded spectral points, individually, using each of the model equations. The reduced- χ^2

values for each of the nine fits are given in table 8.1. The results highlight two

Unfolding Model	Spectral Fitting Model		
	$J_{M-1}(E)$ (Eq. 8.9)	$J_{M-2}(E)$ (Eq. 8.10)	$J_{M-3}(E)$ (Eq. 8.11)
Model-1	30.17	1.03	1.17
Model-2	29.51	0.97	0.94
Model-3	27.97	0.98	0.96

Table 8.1: The raw spectrum has been unfolded using three models for the flux (equations 8.9, 8.10, and 8.11). The three unfolded spectra (rows) were then fit using the flux models (columns). The reduced χ^2 value for each fit is given above.

important points. First, regardless of which model was used in the unfolding process, when the resultant spectral points were fit, they were inconsistent with the simple three-power-law shape of Model-1 ($\chi^2 \simeq 30$). Such a trend is not seen for the other two models which show equivalently good agreement. Second, the fits do not seem to be significantly worse when Model-1 was used in the unfolding (see the top row). This gives an indication of just how insensitive the unfolding process is to the choice of a particular model. In light of these two points, for the rest of this thesis, all spectra will be unfolded using Model-2.

8.3 Comparison with Previous Auger Measurements

The full integration of the TOTd and MoPS triggers into Auger’s reconstruction algorithm required an update to almost every step of the analysis. This included the way that stations are rejected as background (section 5.1.2), the model of the lateral distribution of signal (section 5.2.3), the way that the untriggered stations enter into the fit (section 4.3.2), the entire energy assignment (chapter 6), and the T4 efficiency and energy resolution models used in the unfolding. While each of these changes are individually rather small, they can compound to produce a spectrum that does not necessarily agree with previous measurements.

Auger’s most recently published 750 m spectrum was presented at the 2017 International Cosmic Ray Conference (ICRC) [26]. This is thus the best candidate for a comparison to the new-trigger spectra. However, there are a number of details in how the ICRC spectrum was created that should be discussed. Firstly,

it was created using the old triggers and only includes energies above $10^{17.5}$ eV and angles $\theta < 55^\circ$. Data was used from the years 2008-2016 resulting in an exposure of $288 \text{ km}^2 \text{ sr yr}$, a factor of 2 larger than is possible with the new triggers. The unfolding was also performed using a slightly different equation for $J_{\text{True}}(E)$, two broken power-laws⁴

$$J(E) = J_0 \times \begin{cases} \left(\frac{E}{E_{\text{Ankle}}}\right)^{-\gamma_1}, & E < E_{\text{Ankle}} \\ \left(\frac{E}{E_{\text{Ankle}}}\right)^{-\gamma_2}, & E \geq E_{\text{Ankle}}. \end{cases} \quad (8.12)$$

Finally, the **ICRC** results were created using a separate reconstruction code. Within the collaboration, two separate reconstruction algorithms have been maintained (for clarity, the reconstruction framework updated in this thesis will be referred to as Code-1 whereas the other algorithm will be called Code-2) which have the same general philosophy such as the same rejection of background stations, station calibration, etc., but differ in small ways. For instance, Code-2 fits the lateral distribution of signal with an **NKG**-like function (equation 5.7) instead of a log-log parabola. These small differences are useful for performing cross-checks between code sets as well as estimating the systematic error associated with the various empirically chosen aspects of the reconstruction (i.e. the **LDF** shape, treatment of saturated **PMTs**, etc.). As a result, this section will include a comparison to the published **ICRC** spectrum (which was created using Code-2), but it will prove more illuminating to compare the new trigger spectra to the one produced using Code-1.

The (Code-1) unfolded spectra from this work and the **ICRC** (Code-2) results for the 750 m and 1500 m arrays are shown in figure 8.6. The top plot is the unfolded spectra which have been scaled by E^3 . The black line is a fit to the 1500 m spectrum (black crosses) and has been extrapolated down to lower energies

$$J(E) = J_0 \times \begin{cases} \left(\frac{E}{E_{\text{Ankle}}}\right)^{-\gamma_1}, & E < E_{\text{Ankle}} \\ \text{Norm} \times \left(\frac{E}{E_{\text{Ankle}}}\right)^{-\gamma_2} \left(1 + \left(\frac{E}{E_{\text{Supp}}}\right)^{\Delta\gamma}\right), & E \geq E_{\text{Ankle}}. \end{cases} \quad (8.13)$$

Using the fit-line as a reference, it is clear that there is an offset between the new trigger and old trigger results. The magnitude of this difference can be seen more

⁴Because the spectrum is only measured above $10^{17.5}$ eV with the old triggers, this simplified shape is sufficient to use in the unfolding.

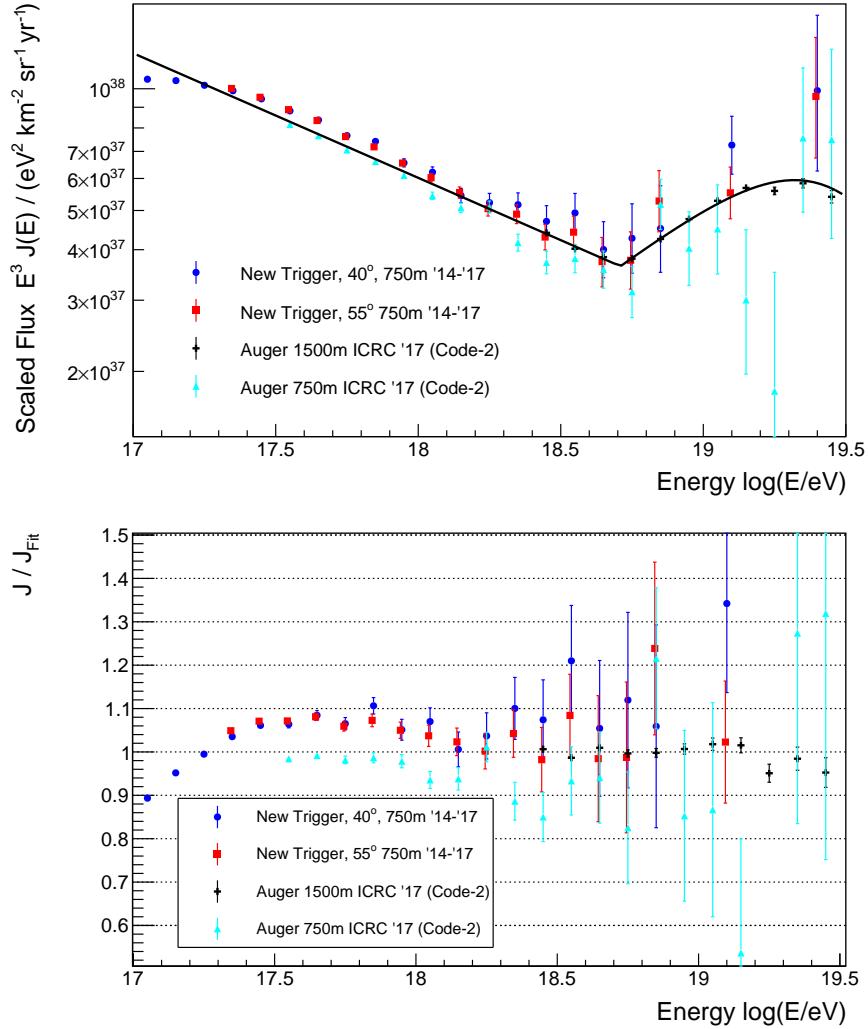


Figure 8.6: Top: The unfolded spectra from this work as well as the most recent results from the [ICRC](#) 2017 using Code-2 (see text) are shown. The black line is a fit to the 1500 m spectrum, extrapolated below $10^{18.4}$ eV. Bottom: The ratio of the spectrum to the spectral fit in the top plot is shown.

easily in the bottom plot which shows the ratio of the individual spectra and the fit line.

From this plot, the benefit of the new triggers is clear. Not only does this measurement simply include a wider range of energies but the data sets (particularly the 40° set) allow the array to see the change in spectral index near the second knee for the first time. This has not been done before by Auger and makes this work a significant improvement to the Observatory.

However, the difference in the flux of the new and old trigger spectra for the 750 m array is not negligible, $\simeq 10\%$. Due to the shape of the spectrum this means that there is a net energy offset of $\simeq 4\%$ which is quite large given that the energies were tuned using the same set of golden hybrid events. The rest of this section will be dedicated to showing that the offset is mainly due to improvements to the **SD** reconstruction and the energy scale.

To understand this better, it is easier to instead compare the new trigger spectra to the old trigger spectrum created using Code-1. The respective ratios are compared in figure 8.7. The grey triangles are the unfolded spectral points for the

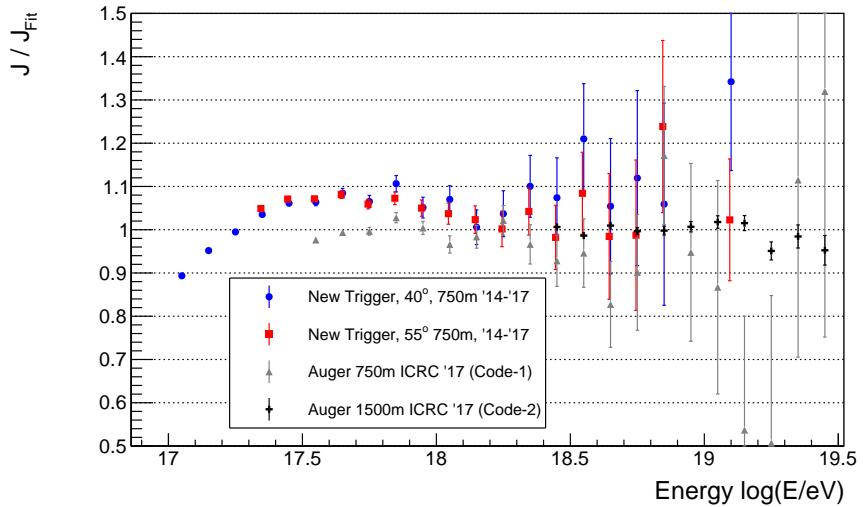


Figure 8.7: The unfolded spectra for the two new trigger data sets (blue circle and red squares) are compared to the old trigger spectrum using the **ICRC** 2017 data sets (grey triangles and black crosses). The ratio of the measured spectral points are shown with respect to a fit line (see figure 8.6, top).

ICRC 2017 data set using Code-1. Even with the same reconstruction framework, these points do not agree with the measurements using the new triggers. As it turns out, this discrepancy is not related to the use of new triggers versus old triggers but instead is a result of the updates to the air shower reconstruction and the data period being used.

One of the major changes to the reconstruction since the **ICRC** publication was the update to the **LDF** fit, specifically the term in the likelihood function describing the untriggered stations (see sections 4.3.2 and 5.2). To test the impact of this change, the events in the **ICRC** data set were reconstructed again using the updated

untriggered station term. The energy assignment (**CIC** and **FD** calibration) was retuned and the unfolding procedure was repeated. The residuals of the spectrum are shown by the green circles in figure 8.8. The agreement is already much better,

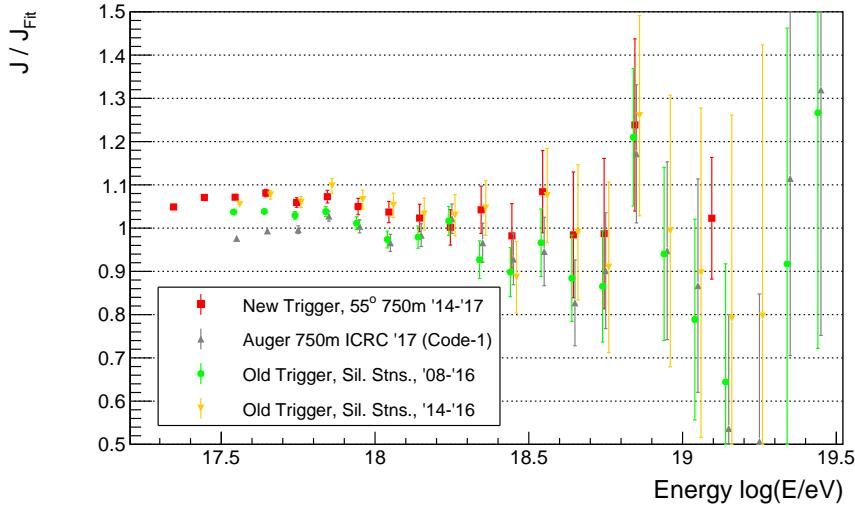


Figure 8.8: To understand the offset in the flux measured using the new triggers, the **ICRC** data set was reconstructed using a consistent treatment of the untriggered stations (green circles) and time period (yellow triangles). The ratio of the fluxes is shown with respect to the empirical fit shown in figure 8.6, top.

cutting the flux separation in half at the lowest energies. The reason for this is subtle and has to do with the fact that untriggered stations are only considered out to a fixed axial radius (2500 m for the 750 m array). One of the effects of including the untriggered stations is that the **LDF** gets pulled down, producing smaller values of $S(450)$. Since untriggered stations are only considered out of a fixed radius, the smallest showers, which trigger only $\simeq 3$ stations, and the most inclined showers, for which the intersection of an oblique cylinder (with 2500 m radius) with the ground will cover the largest area, there will be the most silent stations used in the fit. Thus when the treatment of the untriggered stations is changed, it will affect showers with both size and zenith angle dependencies. Because the **FD** does not have the same aperture as that of the **SD**, the golden hybrid showers used in the energy calibration include an oversampling of specific regions of zenith angle and shower size space (see figure 6.9). So the angular and size dependence of the energy estimator produced a shift.

The remaining offset can almost entirely be explained by examining the ratio

of the energies (E_{SD}/E_{FD}) for the golden hybrid events used in the reconstruction. These values are shown in figure 8.9 for the years used in the [ICRC](#) energy calibration, 2010 - 2016. Each of the grey circles represents the energy ratio for one event and

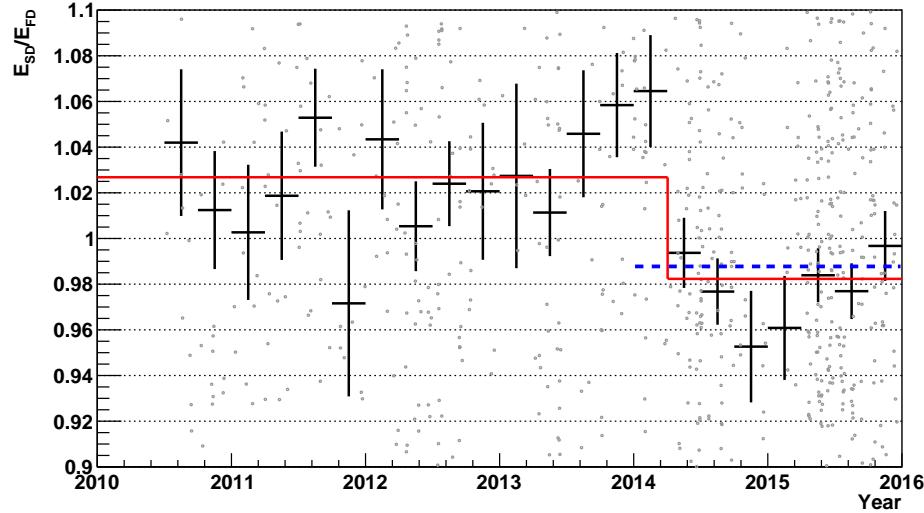


Figure 8.9: The ratio of the [SD](#) and [FD](#) energies used in the calibration of the [ICRC](#) data set is shown above for each event (grey circles) and in 3-month bins (black crosses). A fit has been performed for the periods before and after the shift in the energy scale (solid red line) as well as for the years used in the new trigger calibration (dashed blue line).

the black crosses show the average trend in 3-month bins. By construction, the average ratio over the entire six years should be 1, however, there is a clear trend in the data, a discontinuity in mid-2014. The total shift in the energy scale is shown by the red line, $+2.7\% \rightarrow -1.8\%$. However, since the new trigger event begins at the start of 2014, the effect of the energy shift is mitigated somewhat and is shown by the blue dashed line, -1.2% . Thus, to produce the most consistent check against the previous Auger results, the [ICRC](#) data set (with the updated treatment of untriggered stations) was recalibrated over a consistent time period, 2014 - 2017. The resulting spectrum is shown in yellow triangles in figure 8.8. The combination of these two changes to the [ICRC](#) data set produces a spectrum that is completely consistent with those of the new triggers.

8.4 Identification of the Spectral Features

The flux measurements shown in this thesis include enough exposure to clearly see the ankle feature in the cosmic ray spectrum. With the inclusion of the new triggers, Auger is now also sensitive to the change in spectral index near 100 PeV, sometimes referred to as the *second knee*. Both of these features mark energies where the arriving cosmic rays are transitioning in some respect (i.e. mass distribution, source, etc.). Thus understanding where these features are is important to constrain models describing UHECRs.

Both of the data sets were fit using Model-2 (equation 8.10). The fit is shown for the $\theta < 40^\circ$ data set is shown in figure 8.10. A quantitative estimate of the spectral

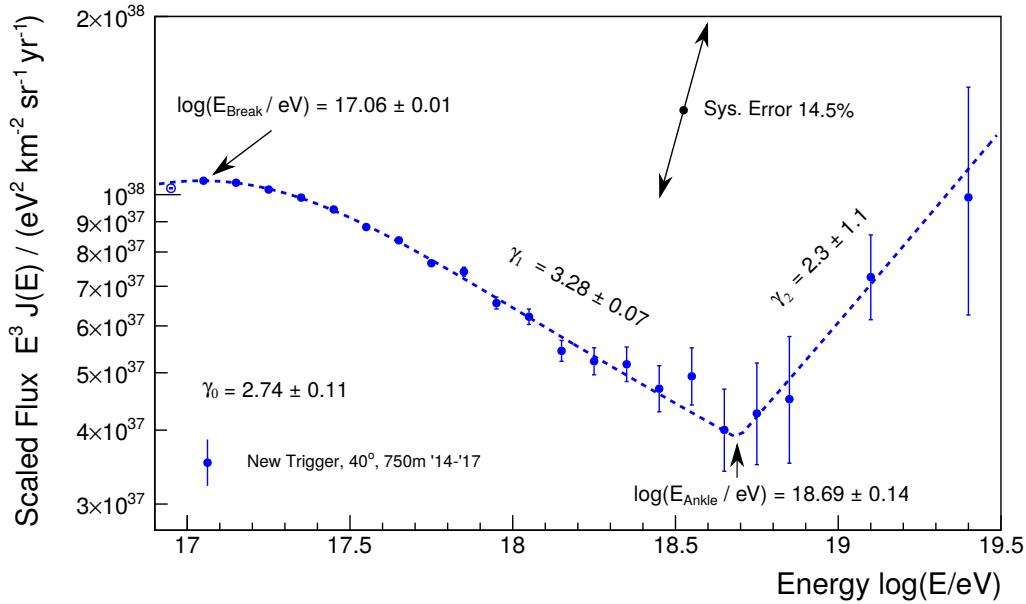


Figure 8.10: The 750 m array spectrum was fit to Model-2 (equation 8.10). The $\theta < 40^\circ$ data set (see section B.6 for flux values) and the fit function are shown above. Statistical errors are shown, scaled by E^3 .

features was performed by fitting the unfolded spectrum to an empirical function, Model-2 (equation 8.10). A fit to the 40° data set and the best fit parameters are shown in figure 8.10. The unfolded spectral points (solid blue circles) and their respective statistical errors, scaled by E^3 , are shown. One unfolded spectral point below full efficiency (empty blue circle) is also shown but was not included in the fit. The agreement is seen to be good and produces a reduced chi-square of 0.97

New Trigger Spectral Fit		
Parameter	Data Set	
	$(\theta < 40^\circ, E > 10^{17.0} \text{ eV})$	$(\theta < 55^\circ, E > 10^{17.3} \text{ eV})$
γ_0	2.74 ± 0.11	2.52 ± 0.47
$\log(E_{\text{Break}}/\text{eV})$	17.06 ± 0.01	17.03 ± 0.02
γ_1	3.28 ± 0.07	3.27 ± 0.06
$\log(E_{\text{Ankle}}/\text{eV})$	18.69 ± 0.14	18.68 ± 0.12
γ_2	2.3 ± 1.1	2.5 ± 0.9
k	0.44 ± 0.21	0.38 ± 0.14

Table 8.2: The values above give the best fit values for the two data sets using Model-2 (equation 8.10).

(see section 8.2.2).

The best fit parameters are given for both the 40° and 55° data sets in table 8.2. Both data sets agree with one another within the statistical uncertainties. These values also agree with the most recent published Auger results from the 750 m array using the old triggers, $\gamma_1 = 3.29$, $\log(E_{\text{Ankle}}/\text{eV}) = 18.71$, and $\gamma_2 = 2.53$ [26].

The energy range covered by the 750 m array does not extend low enough to cover the entire break in the spectrum at $\simeq 10^{17}$ eV. Thus the fit values describing E_{Break} and γ_0 should not be treated as true measurements but of features of the model. A measurement would require a self consistent observation of the flux which extends down to an energy for which the spectral index becomes roughly constant again. However, the energy spectrum produced in this work is vital to constraining the shape and location of the energy where the flux returns to a power law shape leading up to the ankle.

These results will be compared to those of other experiments in section 8.6.

8.5 Systematic Uncertainty of the Spectrum

To make a meaningful comparison of the unfolded flux spectrum to that of other experiments, it is crucial to first understand the systematic errors. By far, the dominant source of uncertainty for the spectrum comes from the energy scale. For the 750 m array, the energy uncertainty is between 14 - 15% (see section 6.4.3). The steep spectrum further amplifies this such that the relative uncertainty on the

flux is closer to $\Delta J/J \simeq 30\%$ depending on the value of the spectral index (see equation 8.3). This is shown by the red dot-dashed line in figure 8.11.

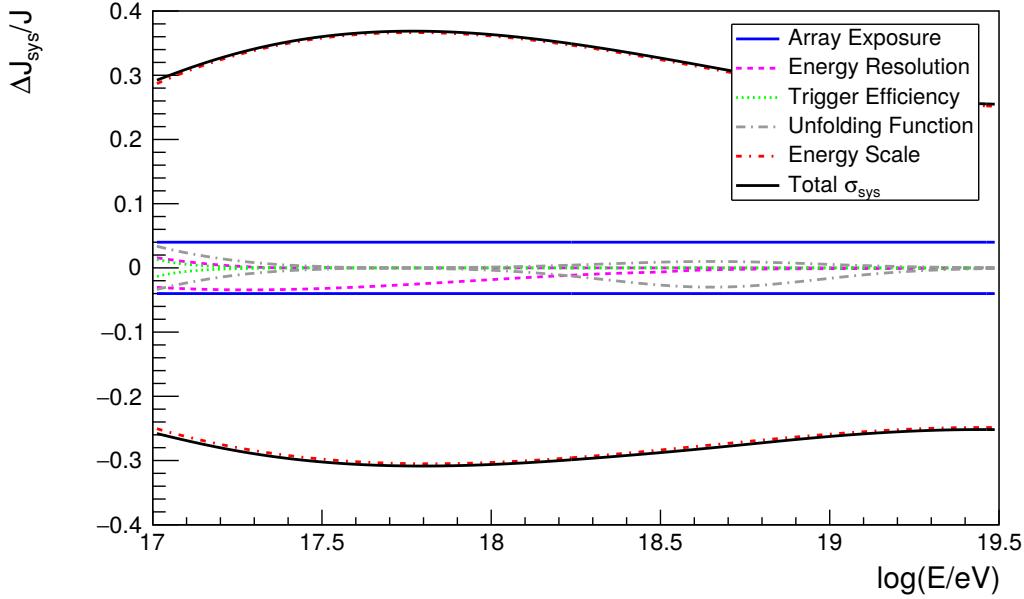


Figure 8.11: The contributions to the systematic uncertainty on the unfolded flux spectrum are shown above. These were calculated by changing the flux systematics by $\pm\sigma$ or conservative estimates were calculated using edge cases.

The next largest uncertainty comes from the exposure of the array. As described in section 7.2, the exposure is calculated by summing over all of the active stations during the data taking period and multiplying by the aperture. The uncertainty on the exposure is the result of the stability of this summing procedure and is accurate to within 4% (blue solid line in figure 8.11).

The rest of the contributions to the systematic uncertainty comes from the unfolding procedure. Already the impact of choosing a particular unfolding function has been analyzed in section 8.2.2. Model-1, three broken power-laws, was shown to not be compatible with the observed spectral shape for which the impact on the spectrum was a few percent or less. This was used as an estimation of the uncertainty from the choice of a particular spectral model (grey dot-dashed lines). The unfolding procedure also relies on the estimate of the triggering efficiency. Studied heavily in section 7.1, the descriptions of the T4 trigger probability using simulations of proton and iron were used to find the impact on the spectrum. The final uncertainty contribution is from the energy resolution used in the unfolding

process. This was studied in section 6.4.3 using hybrid events. The unfolding procedure was repeated with the resolution set to $\pm\sigma$ of the fit. All of these three parameterizations used in the unfolding process amount to a 4% error and impact the spectrum most at low energies where the correction factor is the largest.

The total energy uncertainty was determined by adding all of these uncorrelated contributions in quadrature. The result is shown by the solid black line in figure 8.11 for which the energy scale uncertainty can clearly be seen to have the largest impact.

8.6 Comparison with other Experiments

The Pierre Auger Observatory is the world’s largest cosmic ray detector. The data collected over 14 years has been used to make the most precise measurement of the cosmic ray spectrum in the decade of energy leading up to the ankle and including the flux suppression. With this work, the 750 m is now able to also observe one additional spectral feature at 100 PeV. This is important for the cosmic ray community because while the flux spectrum has been measured by many experiments using various detection schemes, there is still a disagreement on the location of the spectral features.

For energies above $10^{16.5}$ eV, there are many experiments that have made measurements of the cosmic ray spectrum. To give a sense of the spread of the flux measurements, the energy spectra according to AGASA [91], AKENO [92], Auger [26], Fly’s Eye [93], GAMMA [94], HiRes [11], IceTop-73 [95], KASCADE-Grande [96], TA [97–99], Tibet-III [92], TUNKA-133 [100], and Yakutsk [101] are shown along with the results from this work for the 40° data set (solid blue circles) in figure 8.12. It is clear that there is a fairly large spread across the various experiments. For scale, the average systematic uncertainty in the energy for the new trigger reconstruction ($\sim 14.5\%$) is shown. Within these error bars all of the measurements near the break at 100 PeV agree with the results from this work and can be considered as compatible at least in the overall flux. However, near the ankle this is not so as some significant divergence is seen up to $\sim 30\%$ for the AGASA, AKENO, and Fly’s Eye measurements in particular. The difference is likely due to the energy scale which was set by a simulation code that is now three decades old [102].

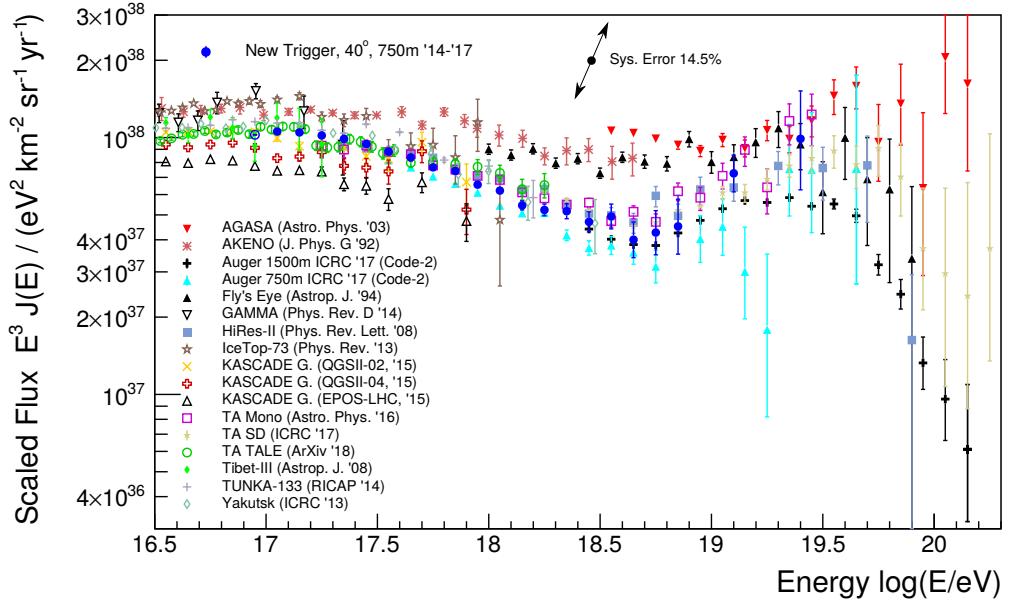


Figure 8.12: The observed spectra by many other experiments (see text for references) are shown along with the 40° spectrum from this work (solid blue circles).

The results from this work are most important to the Observatory and the cosmic ray community in its extension of the **SD** into the break region near 100 PeV. Of the measurements shown in figure 8.12, six of the results quote an observation of a break between the knee and ankle. A more close up view of this energy range is shown in figure 8.13. This plot highlights the usefulness of the extending Auger’s ability to view an extra spectral feature. When trying to understand the shape of the spectrum, the offsets between the flux measured by various experiments can be understood better by comparing the relative location of the various spectral features. That is, since the location of the features in the *true* spectrum are fixed, they can be used as standard candles to understand the relative systematic offsets between experiments.

In this respect, figure 8.13 highlights a second feature of the **SD**’s flux measurement. For many experiments, it is quite common to rely entirely on simulations to set the energy scale. As an example, the spectrum measured by KASCADE-Grande is shown for an interpretation of the data [96] using three hadronic models, **QGSJETII-02** (gold Xs), **QGSJETII-04**⁵ (red crosses), and **EPOS-LHC** (black tri-

⁵The most important difference between **QGSJETII-02** and **QGSJETII-04** is that they have been tuned using pre- and post-**LHC** data, respectively.

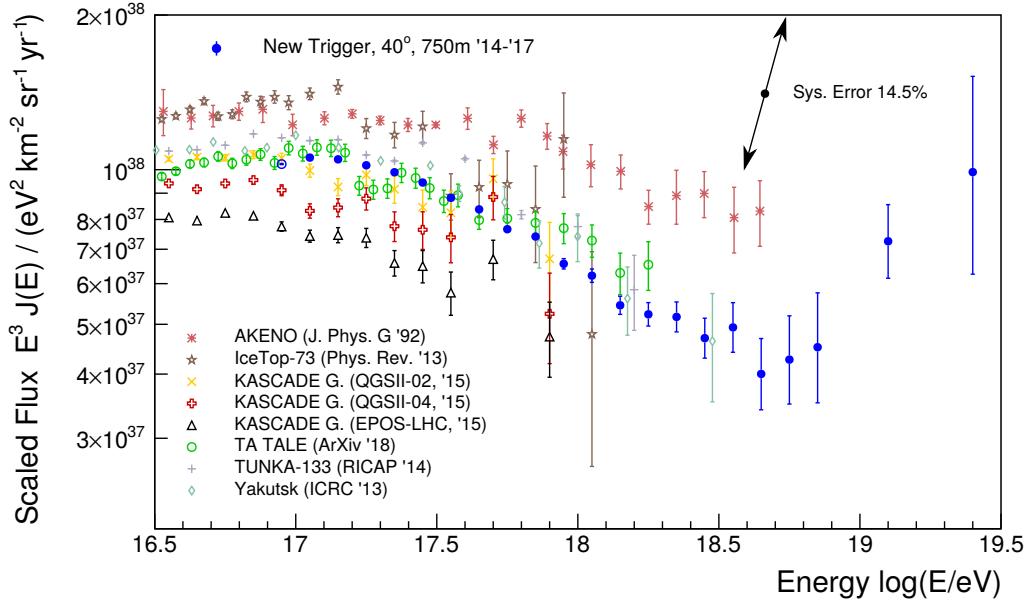


Figure 8.13: The measurements of the spectra by six experiments that observe a second knee are shown along with the 40° spectrum from this work.

angles). The separation between models is considerable, about 20 - 30% in flux. This is not the case for the spectrum measurement in this thesis. The algorithms used in the reconstruction have all been derived using Auger data. Only in the unfolding were simulations used, but this has been shown to contribute < 2% to the flux uncertainty and has been crosschecked using data.

With these concepts in mind, a comparison can be performed between the observed spectral features near the second knee with the six experiments shown in figure 8.13. The values are shown in table 8.3 (note that break values without error bars are estimations given by the respective collaboration). Before providing an analysis of how these values stack up, it should be noted that since the second knee is only observed on one side by Auger, there may be a slight bias in the various values. Additionally, most of the spectra were fit using simple power laws (i.e. something like equation 8.9) instead of the smooth break shown in figure 8.10. However, even with these in mind, it is still illustrative to compare the results from this work to other measurements.

The values given by other experiments have a relatively large spread for the location of the break, $10^{16.86} \rightarrow 10^{17.8}$ eV. However, most of these measurements are fairly close to 100 PeV. Excluding the estimations and the value from this work,

Measurement	$\log(E_{\text{Break}}/\text{eV})$	γ_1
New Trigger $\theta < 40^\circ$	17.06 ± 0.01	3.28 ± 0.07
AKENO [92]	17.8	3.16 ± 0.08
IceTop-73 [95]	17.1 ± 0.1	3.37 ± 0.05
KASCADE-Grande (QGSII-02) [96]	16.88 ± 0.16	3.15 ± 0.05
KASCADE-Grande (QGSII-04) [96]	16.91 ± 0.24	3.15 ± 0.10
KASCADE-Grande (EPOS-LHC) [96]	16.86 ± 0.10	3.17 ± 0.05
TA TALE [99]	17.04 ± 0.04	3.19 ± 0.02
TUNKA-133 [100]	17.5	3.35 ± 0.11
Yakutsk [101]	17	3.12 ± 0.03

Table 8.3: The values above compare the location of the break and γ_1 (see equation 8.10) as measured by six experiments. Values without error bars correspond to the estimated break points according to the respective collaboration.

the weighted mean is $\log(E/\text{eV}) = 17.02 \pm 0.03$. Similarly, the weighted mean of the spectral index is $\gamma_1 = 3.19 \pm 0.02$. Given these values, the spectral fit from this work is in fairly good agreement with other measurements. The location of the break is well within one sigma of the global average. The spectral index, γ_1 , from this work is a bit larger than the average, however, this might again be a result of the smooth versus sharp break used in the fit. That is, in the fit from section 8.4 (using Model-2), γ_2 has the interpretation of being the index asymptotically far from the 100 PeV break where as for a sharp break, γ_1 is the average slope.

Overall, the spectrum presented in this work agrees well with the measurements of these other experiments. The systematic uncertainty in the energy, shown by the black arrows in figure 8.13, indicates that any of the spectra can be made to coincide with the 750 m array results by a one sigma shift or less in energy. Further, the identified spectral features are fairly consistent with those of other experiments that observe the second knee.

8.7 Conclusion

This chapter is the culmination of the many individual studies needed to incorporate the **TOFd** and **MoPS** triggers into the 750 m array and to understand how their inclusion affected the capabilities of the detector. With their inclusion, the Auger Observatory is now able to measure the energy spectrum above 10^{17} eV with a nearly 100% duty cycle. Using four years of data, the $145 \text{ km}^2 \text{ sr yr}$ of exposure

allowed for the observation of a steepening at 10^{17} eV, the so-called second knee, as well as the ankle.

Using the forward folding method, the detector effects were removed from the spectrum measurement to produce an unbiased result. The robustness of this method was analyzed for which the total uncertainty is less than 5% and did not show much dependence on the assumptions of spectral model.

Since the new triggers required an update to nearly every part of the reconstruction algorithm, a crosscheck was performed to verify compatibility with previous measurements. In a comparison to the most recent results, there was a significant offset in the flux measurements compared to that of the most recently published Auger results for the 750 m ($\simeq 10\%$). However, the difference was shown to be a result of updates to the **LDF** fit procedure and the temporal stability of the **FD** energy scale. That is, when treated in a consistent fashion, the new trigger data set agrees perfectly well with that of the old triggers.

Finally, a comparison was shown between the new trigger spectrum and many of the other experiments that observe the second knee. Despite the general inter-experiment disagreement, they agree with the Auger measurement within the $\simeq 14.5\%$ uncertainty in energy ($\simeq 30\text{-}35\%$ uncertainty in flux).

Chapter 9

Summary

In the **UHE** regime, cosmic rays are not directly observed but are instead studied via the air showers that are created via their interaction with air molecules high in the atmosphere. The Pierre Auger Observatory is the world’s largest cosmic ray detector which employs a hybrid method of observing **EASs** including a **FD** and a **SD**. The **FD** is a crucial component of Auger as it allows for the direct observation of the longitudinal profile by which the **FD** can make nearly calorimetric energy measurements. Combined with the ground-based observations of the secondary particles by the **SD**, Auger has a $\simeq 100\%$ duty cycle which is necessary to study cosmic rays with ultra-high energies.

This thesis was focused on the low-energy extension of the **SD**, a nested array of **WCD** stations with 750 m spacing, half that of the larger array. The work presented above included two main components. The first was the inclusion of two new station-level triggers into the reconstruction algorithm of the 750 m array. The second component was a measurement of the cosmic ray spectrum. With the improved reconstruction and detector capabilities, this measurement extended into a lower energy range than what was previously possible at the Observatory.

9.1 Inclusion of the New Triggers

The work in this thesis was motivated by a set of new station-level triggers, **TOTd** and **MoPS**, which were designed to be less sensitive to the low energy muons which make up the primary background for **WCDs**. They are thus able to increase the sensitivity of the stations to regions of the shower front where the particle densities

and/or energies are low.

On an individual station basis, these triggers were shown to increase the detection capability of a [WCD](#) to signals with sizes 1/3 to 3 times that which would be made by a vertically going muon. Given that they are mostly beneficial at small sizes, they are typically found far from the central shower axis and on average increase the station multiplicity in an event by $\simeq 10\text{-}30\%$.

Further, [TOTd](#) and [MoPS](#) also increase the 750 m array's sensitivity to lower energy showers in general. To study this sensitivity, the probability of the array to (T4) trigger on a shower was estimated using air shower simulations based on modern hadronic interaction models. These values were also crosschecked using a simple air shower model based on Auger data. The largest increase in triggering probability was shown to be for more vertical showers. This is consistent with the new triggers' design since increasingly inclined showers have a more attenuated electromagnetic component and are thus muon rich. The 750 m array was shown to trigger with $> 98\%$ efficiency for showers with energies and zenith angles in the ranges ($E > 10^{17}$ eV and $\theta < 40^\circ$) or ($E > 10^{17.3}$ eV and $\theta < 55^\circ$). This is a large improvement over the previous triggering algorithms for which this efficiency limit was met for showers with ($E > 10^{17.5}$ eV and $\theta < 55^\circ$). Due to the steep power law shape of the energy spectrum, the inclusion of the new triggers produced a doubling of the overall event rate, $0.9 \rightarrow 2.3 \text{ min}^{-1}$.

The additional sampling by [TOTd](#)- and/or [MoPS](#)-triggered stations is crucial in the reconstruction of an air shower event. At Auger, the [SD](#) reconstruction is based on an air shower model which macroscopically describes the distribution of secondary particles within the shower front. The model can almost completely be decoupled into two parts. The trajectory of the shower can be determined almost exclusively by the start times of the signals in the stations. Likewise an air shower's energy is estimated via the distribution of signal as a function of axial radius. Since a surface array only observes the shower in a particular time slice, it cannot directly make energy measurements. However, a shower's size, a value highly correlated with its energy, can be estimated by the expected signal a station would measure at a fixed axial distance ($S(450)$ for the 750 m array). The [LDF](#), an empirical model of lateral distribution, was updated using the new triggers which included energy and zenith angle dependences to account for the changing muonic fraction of showers at the ground. This step is crucial in making an unbiased measurement

of the size estimator, $S(450)$.

Finally, a number of systematic biases were removed from the size estimation. The total grammage a shower will pass through before arriving at the array is not constant in time or zenith angle. The first correction removed the effects of the cyclic fluctuations in atmospheric overburden and Molière radius as a result of the dynamically changing local weather at the array. Second an ad-hoc correction was applied to account for the $\sim 10\%$ offset in the observed shower sizes at 450 m from the shower axis. This correction mainly affected low energy, highly inclined showers. Finally, a constant intensity correction was applied to offset the increasing slant depth of the array as a function of zenith angle. The unbiased energy estimators, S_{27} and S_{35} , were calculated for the two data sets used in this thesis, ($E > 10^{17}$ eV, $\theta < 40^\circ$) and ($E > 10^{17.3}$ eV, $\theta < 55^\circ$), respectively.

Whereas other experiments would then have to rely on simulations to find a conversion between the unbiased energy estimator and air shower energy, S_{27} and S_{35} were cross calibrated with the energy measurements of the [FD](#). The hybrid design of the array allows for the independent estimation of air shower properties by both detector systems. Using $\simeq 1000$ showers that were well reconstructed by both the [FD](#) and [SD](#), a conversion was found for both data sets. The overall systematic uncertainty in the [SD](#) energy scale is 14 - 15% which is dominated by the 13.5% uncertainty from the [FD](#).

9.2 Energy Spectrum Using New Triggers

One of the most precise measurements that can be made by the [SD](#) is of the cosmic ray flux. Such a measurement was made using the four years (2014 - 2017) of data from the 750 m array. This was done for both of the $\theta < 40^\circ$ and $\theta < 55^\circ$ data sets.

Due to the regularity of the [SD](#)'s layout and the choice of only using data for which the array was fully efficient, the exposure was calculated geometrically. The number of stations with six working neighbors was added up to calculate the aperture of the array. This summation was performed over the four year data taking period to determine the total exposure, $89.8 \pm 3.6 \text{ sr yr km}^2$ for the $\theta < 40^\circ$ data set and $145.8 \pm 5.8 \text{ sr yr km}^2$ for the $\theta < 55^\circ$ data set.

For such an exposure, $\sim 500,000$ events were observed with energies $E > 10^{17}$ eV, enough to see the ankle at $\simeq 18^{18.7}$ eV but not the suppression at $\simeq 10^{19.5}$ eV.

However, this did allow for an observation of a change in spectral index near 100 PeV, a feature of the energy spectrum historically too low in energy to be seen by Auger.

The raw measurement of the cosmic ray flux includes detector effects as a result of the finite energy resolution. Though the migration of an individual event is roughly symmetric, the steepness of the spectrum causes the net migration to be asymmetric. Using a forward folding procedure, these detector effects were removed. As part of this forward folding method, a model for the true spectrum must be supplied. Three models were tested which describe a power law spectrum with a break at near 100 PeV and one at the ankle. Ultimately, these produced compatible unfolded results to within 5% in flux though the best models were the two for which the break near 100 PeV was described by a smooth transition.

The unfolded spectrum was compared to that of the most recently published 750 m results which used only the old trigger algorithms. There was a difference in the flux on the order of 10% but this offset was shown to be a result of the difference in reconstruction methods and time period over which the data was calibrated. Additionally, the spectral features ultimately agreed with that of previous Auger measurements, with the ankle being located at $\log(E_{\text{Ankle}}) = 18.69 \pm 0.13$ and with a spectral index of $\gamma_1 = 3.28 \pm 0.07$ leading up to the break. The additional break, seen primarily by the $\theta < 40^\circ$ data set, is seen near 100 PeV.

The 14 - 15% systematic uncertainty in the energy scale likewise produces a ≈ 30 - 35% systematic uncertainty in the flux. Many of the spectrum measurements made by other experiments are compatible with the result of this thesis. However, the measurement using the 750 m is somewhat unique as it has a very minimal reliance on simulations since the energy scale was determined by a cross calibration with the [FD](#). Further, the ability of Auger to also observe the spectrum down to 100 PeV using the [SD](#) is important as the Observatory is now able to see three spectral features using a consistent detector scheme.

9.3 Outlook

The second knee is of particular interest because it has been identified as a transition point for the mass composition of particles arriving at Earth [37]. A knee-like feature for the heavy components near $10^{16.8}$ eV and an ankle-like feature for the

light components create an overall change in the spectral index of the all-particle flux near 100 PeV. These measurements are important because they help constrain the theoretical models explaining how and where cosmic rays at [UHE](#) are created.

The work in this thesis, to this end, includes two contributions. The first is the most precise measurement of the energy spectrum near the second knee. The second is the extension of the reconstruction framework itself. Moving forward, the reconstruction chain developed in this work can be used to study the arrival directions and average masses down to lower energies than were previously possible with the [SD](#).

Large scale anisotropy studies using the 750 m array (potentially combined with the results of the 1500 m array) can give some indication of where the transition from a Galactic to an extragalactic origin occurs. The increased statistics afforded by the nearly 100% duty cycle of the [SD](#) may even be able to show a quadrupole moment expected by cosmic rays coming from the galactic disk. Additionally, there have already been a number of methods currently being pursued at Auger to find subtle observables within the [PMT](#) signals to make X_{\max} measurements using the surface array. Such studies can be complementary to those of the [FD](#) to both provide reassuring crosschecks of the mass distribution of cosmic rays and to increase the statistics by up to two orders of magnitude.

In these respects, the extension of the 750 m reconstruction down to the second knee makes possible the prospect of making some of the most precise measurements of the cosmic rays in an energy range where there are theoretical motivations to expect a transition from galactic to extragalactic sources. The measurement of the energy spectrum shown in this work is an important first step towards a much better understanding of [UHECRs](#).

Appendix A

Additional Analyses

A.1 Surface Detector Station Calibration

As described in section 3.2.1,

The amount of signal produced by a muon in the water volume of a WCD is roughly proportional to its track length [68]. This consistent emission of Cherenkov photons is used as a standard candle to calibrate the digitized signal output by a station’s onboard electronics. At Auger, the unit of signal in the stations is the VEM, the amount of light that is produced by a vertically going muon.

SD stations constantly monitor the integrated area, Q , and amplitude, I^1 , of the signals created by background particles. When a T3 trigger is identified by the CDAS, each of the T1 and/or T2 triggered stations sends its relevant traces for the event, but also sends the charge and current histograms for each of the PMTs. An example of a charge histogram is shown in figure A.1. This histogram shows the typical, double-peaked shape. The first peak (located at $\simeq 920$) is the result of low energy electromagnetic particles and corner-clipping muons while the second is due to the muons that traverse significant portions of the water volume (located at $\simeq 1070$). The charge corresponding to the location of the second peak, $Q_{\text{VEM}}^{\text{peak}}$, is the quantity of interest as its center is proportional to the charge deposited by a vertically going muon, Q_{VEM} ,

$$Q_{\text{VEM}}^{\text{peak}} = 1.045 \times Q_{\text{VEM}}. \quad (\text{A.1})$$

¹The quantities Q and I are sometimes called the charge and current since this is what is being measured in the PMTs.

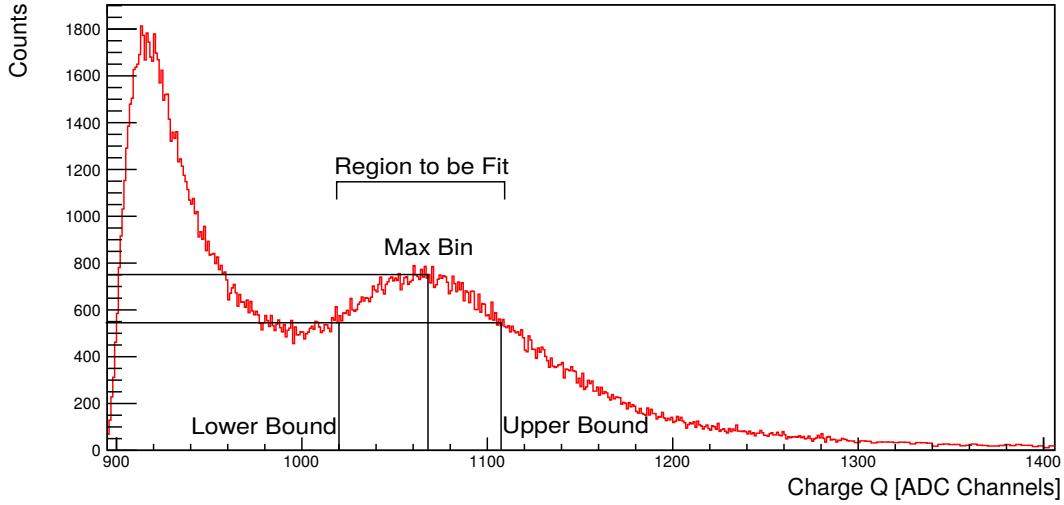


Figure A.1: An example of a charge histogram is shown above. The bin heights represent the number of observed background events with an integrated signal corresponding to the given ADC channel. The muon peak, centered at Max Bin is the product of muons that traversed significant portions of the WCD's water volume. The peak centered at $\simeq 920$ is caused by background electromagnetic particles and corner-clipping muons.

Thus, by fitting the peak, Q_{VEM} can be used to convert the digitized traces into units of **VEM**.

Previous to this work, the method to find the peak included the following steps:

1. Starting from the last bin (bin $\simeq 1400$ in figure A.1), iterate backwards until a bin of at least 300 counts is found².
2. Continuing, the tallest iterated bin (Max Bin) is kept track of until three consecutive bins are found which all have height $h_i < (\text{Max Bin})/1.3$. This is the Lower Bound.
3. The iteration reverses direction until three more consecutive bins are found with the same height criteria. This is the Upper Bound.

In figure A.1, the Max Bin is located at $Q \simeq 1070$ and the two bounds identified in steps 2 and 3 are marked as Lower Bound and Upper Bound. A rudimentary calculation is performed online by each station's onboard electronics which gives

²Because of the 25 kHz rate of background particles entering the water volume, the total number of counts in the charge histograms is fairly stable.

an estimate of the bin corresponding to the peak, $Q_{\text{VEM}}^{\text{est}}$. If this estimated value is not contained within the Upper and Lower Bound, the algorithm quits and the estimated onboard value is kept, $Q_{\text{VEM}}^{\text{peak}} = Q_{\text{VEM}}^{\text{est}}$. Otherwise, all of the bins in this range are fit to a parabola for which the maximum is used as an estimate of $Q_{\text{VEM}}^{\text{peak}}$. This method works well 98% of the time.

The typical pathology that leads to a failed fit occurs when both of the peaks are of comparable size. In such a situation, the dip between the peaks is too shallow to satisfy the $h_i < (\text{Max Bin})/1.3$ requirement. An example of such a charge histogram is shown in figure A.2. The Lower and Upper Bounds will instead

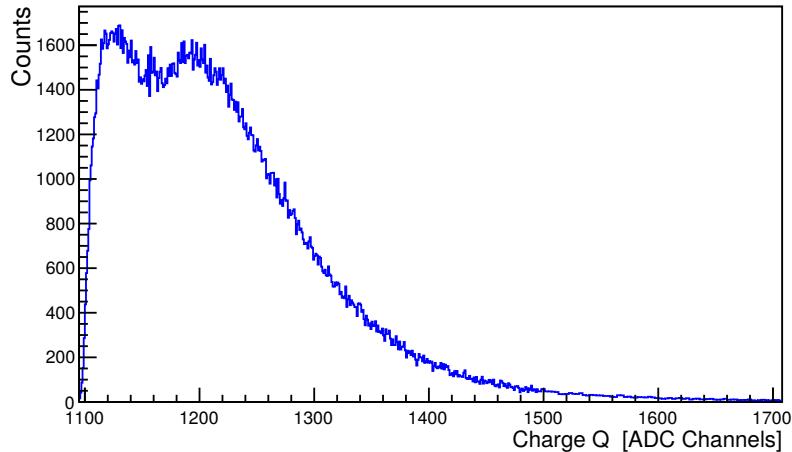


Figure A.2: The charge histogram shown above is an example of a pathology where the previous calibration method failed to find the correct peak.

surround both peaks which will, together, be fit to a parabola, producing a bad estimate of $Q_{\text{VEM}}^{\text{peak}}$.

An update to the peak-finding routine was devised which does not fail in these edge cases. This method includes two fits, one for the muon peak and one for the dip. Before any fit is performed, the charge histogram is first smoothed by performing an average:

$$y_i = \frac{h_{i-2} + 2h_{i-1} + 3h_i + 2h_{i+1} + h_{i+2}}{9}. \quad (\text{A.2})$$

The resulting distribution is then fit using a two-step routine. The first step involves finding a suitable range of i over which to fit the values of y_i . The second step involves an iterative fit to ensure that the estimate of $Q_{\text{VEM}}^{\text{peak}}$ is robust.

The initial guess of the range is given by $(C \pm R) \times Q_{\text{VEM}}^{\text{est}}$ where the default values are $C = 1$ and $R = 0.25$. That is, the initial fitting range includes the bins centered at $Q_{\text{VEM}}^{\text{est}} \pm 25\%$ (note that this is by bin number, not charge value). This range is then fit to a parabola for which the fit's curvature must be negative and the Max Bin must be located in bins $20 \leq i \leq 500$ (again by bin number, not the value of Q). If either of these requirements is not met, the values of R and C are individually changed, $0.2 < R < 0.33$ and $0.8 < C < 1.2$, to see if a good fit can be found on a different range of bins.

If a valid fit is found, a small feedback loop is set up by which the maximum value found using the fit parabola is then used to redefine the fit range again $(C \pm R) \times Q_{\text{VEM}}^{\text{peak}}$. This loop is repeated until the parabola maxima converge (if it does not converge, the fit is considered to have failed).

Once a peak is found, the same process is applied to the dip with two modifications. First the initial guess of the dip location is given by $0.65 Q_{\text{VEM}}^{\text{peak}}$. Second the curvature of the fit parabola must be positive.

Because an accurate calibration of the stations is essential to a high quality event reconstruction, the new and old methods were compared to ensure that this update did not include any biases. A comparison of the identified values of muon peak is shown in figure A.3. This shows the ratio of the muon peak using the old and new methods for 10 years of calibration data. The agreement is seen to be good up to 0.1% though there is a distribution which is non-Gaussian and not centered at zero. This is the result of events where the old method was fitting both peaks in the charge histogram but not explicitly failing.

The updated station-calibration method has two benefits. First, it fails an order of magnitude less (0.3% of the time) and provides a better calibration in the edge cases where the original did not explicitly fail. Second, it provides a useful statistic by which the aging effects of the detectors can be studied (for example, the decrease in water-liner reflectivity). As the stations age, the average area-over-peak of signals measured in individual stations degrades. Since the charge histogram essentially measures the scale of signals, station aging also leads to a decrease in the separation of the two peaks. Using the new calibration method, station aging can be characterized by studying the relative location of the muon peak and the dip.

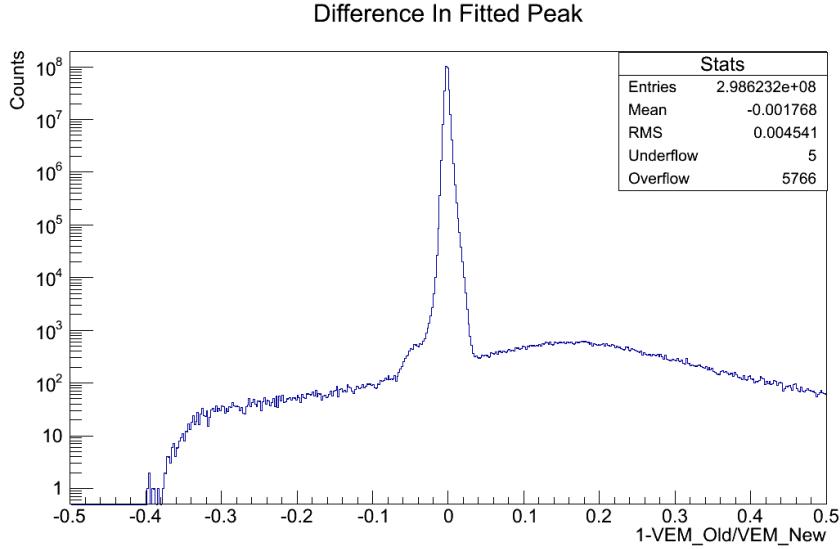


Figure A.3: The histogram above shows a comparison of the location of the muons peaks for the new and old calibration methods. The ~ 100 count background is a result of events where the old method was fitting both peaks in the charge histogram but not explicitly failing.

A.2 Top-Down Station Selection

In section 5.1.2, a method was described by which coincidentally triggered stations are removed from an event. The algorithm, called the bottom-up method, proceeds by finding the three neighboring stations with largest summed signal, fitting a plane shower front using their signal start times, and adding other stations to the event that are temporally consistent with the plane's arrival time. For consistency, the timing window, which defines compatibility with the plane, was updated in section 5.1.2 for each of the T1/T2 trigger types.

A second selection method has been developed within Auger which identifies coincidentally triggered stations but works in the opposite direction, called top-down. The algorithm begins by first removing the lonely stations (see section 5.1.3). All the remaining stations are considered to be part of the event and are fit to a curved shower front. Events with only three stations will immediately pass or fail the top-down selection depending on whether they fulfill the T4 requirements. For higher multiplicity events, the difference between the start time of the signal, t_i ,

and the expected time from the shower front, t_{exp} , is calculated for each station,

$$dT_i = t_i - t_{\text{exp}}, \quad (\text{A.3})$$

as well as a metric for the combined residuals,

$$\mathcal{K} = \sqrt{\frac{\sum_i^M dT_i^2}{M - 3}}, \quad (\text{A.4})$$

where M is the total number of currently selected stations in the event. For the station to pass the top-down selection, every station must pass,

$$|dT_i| < C(M - 2) \cos \theta, \quad (\text{A.5})$$

and the event must pass,

$$\mathcal{K} < D(M - 2) \cos \theta, \quad (\text{A.6})$$

where C and D are the selection tolerance values and θ is the shower's estimated zenith angle calculated during the fit. For events that pass both cuts, the selection is considered completed. However, if either of these criteria is not met, the station with the smallest signal is removed from the event and the criteria are checked again. If they are *still* not met, the removed station is replaced and the station with the next smallest signal is removed instead. This replacing and removing process will continue until both selection requirements are satisfied. If each station has been tested and still the criteria are not met, stations are removed in pairs, then trios, etc.

As with the update to the bottom-up selection, the top-down method was augmented to include trigger-dependent selection tolerances. Shown previously in figure 5.2, the distributions of dT for the various trigger types is quite different. For example, the Th1 and Th2 triggers caused by UHE showers occur within the thinnest dT window but also include the largest background. However, the TOTd and MoPS triggers have a very large range of dT but have a very small background. So the update to the top-down selection will also include finding new tolerance values, C_i and D_i , which are trigger-type specific. For this, equation A.5 needs to

only be updated by replacing $C \rightarrow C_i$,

$$|dT_i| < C_i (M - 2) \cos \theta \quad (\text{A.7})$$

but the criterion in equation A.6 has to be slightly augmented,

$$\mathcal{K} < \sum_i^{N_i} D_i N_i \cos \theta. \quad (\text{A.8})$$

Here N_i is the number of currently selected stations in the event with trigger type i meaning that D_i represents the maximum expected contribution \mathcal{K} for trigger i .

To update the top-down limits, 2.5 years of 750 m events for zenith angles less than 60° (3.7 million events) were used. For each event, the shower direction was estimated using the three seed stations (see section 5.1.2). Since the selection criteria are zenith angle and multiplicity dependent, the scaled timing residuals, c_i ,

$$c_i = \frac{|dt|}{(M - 2) \cos \theta}, \quad (\text{A.9})$$

were calculated for each station. The distributions of c_i for each of the trigger types are shown in the left panel of figure A.4. The distributions are calculated

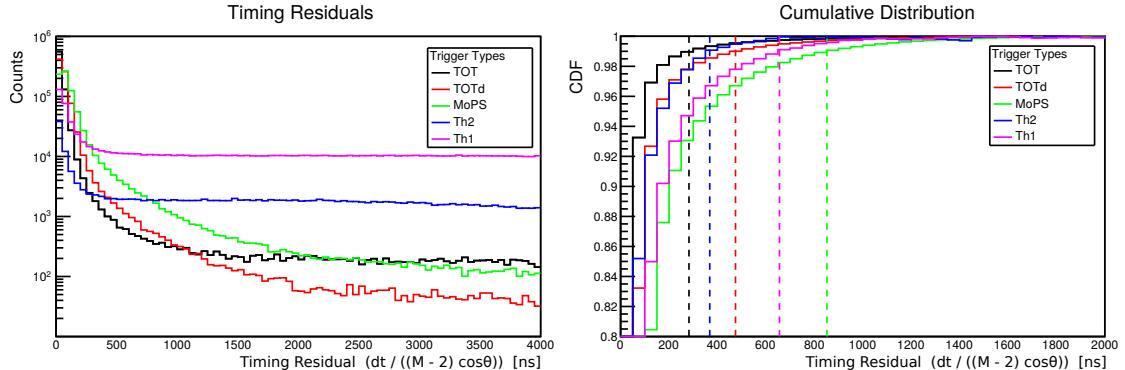


Figure A.4: Left: The scaled timing residuals used in the top-down selection are shown for the five trigger types. Right: The respective CDFs of the background-subtracted residuals are shown along with their respective 99% limits (dashed lines).

differentially according to the hierarchy TOT → TOTd → MoPS → Th2 → Th1 to be consistent with the bottom-up selection update³. The disparity in the amount

³Events in the MoPS distribution represent stations which passed the MoPS algorithm but

Trigger Type	C_i [ns]	D_i [ns]
TOT	282	604
TOTd	474	374
MoPS	852	458
Th2	383	137
Th1	670	225

Table A.1: The updated selection limits for the top-down method are given above for the individual selection criterion, C_i , (equation A.7) and for the combined residual criterion, D_i , (equation A.8).

of background is quite clear for the various trigger types with the threshold triggers being the largest.

As with the update to the bottom-up method, the selection limits were found by subtracting off the background and making a CDF. The values for C_i were set such that $\text{CDF}(C_i) = 0.99$. In figure A.4, these distributions are shown in the right panel and the limits are given by the vertical dashed lines (and can also be found in table A.1).

A similar study was conducted to calculate D_i . This requirement involves a sum over potentially different trigger types. So the quantity of interest is then the share of the total residual by each trigger type, d_i ,

$$d_i = \mathcal{K} \cos \theta \frac{N_i}{M}. \quad (\text{A.10})$$

The background-subtracted distribution of d_i for each trigger type can be found in the left panel of figure A.5. The respective CDFs for the highest three trigger types are shown in the right panel. Again the limits for D_i were chosen such that $\text{CDF}(D_i) = 0.99$. These values are given in table A.1.

A.3 The Time Variance Model

The geometric reconstruction (see section 5.2.1) involves fitting a model of the shower front to find the arrival direction, shower core impact time, and shower curvature. This fit is performed using the start times of the signals in the selected stations. Since the likelihood function being used is a χ^2 (equation 5.4), to get

not the TOT and TOTd triggers.

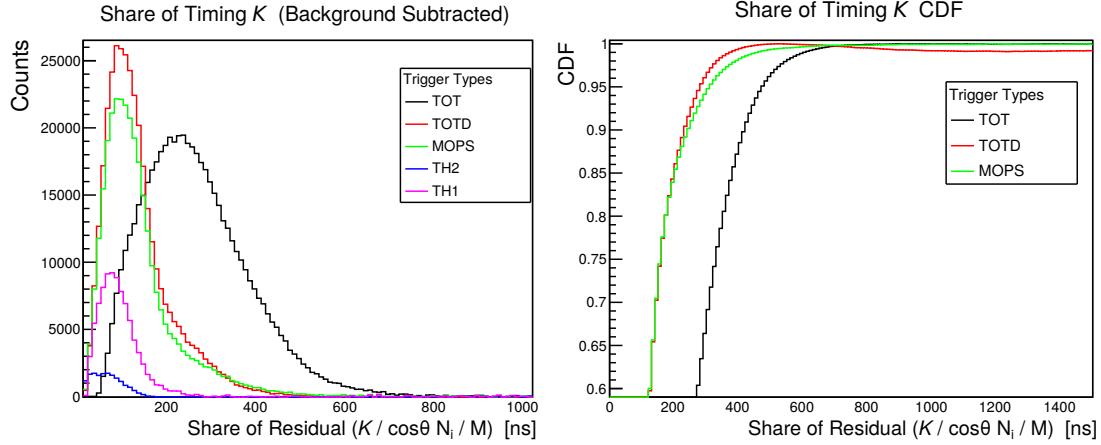


Figure A.5: Left: The contribution to the event’s residuals used in the top-down selection are shown for the five trigger types. The constant background has already been subtracted. Right: The respective CDFs of the background-subtracted residuals are shown for the top three triggers in the hierarchy.

meaningful errors⁴ it is then necessary to have an accurate model of the variance in start times, σ_t^2 . With a proper assignment of σ_t^2 , the angular uncertainty can be estimated on a shower-by-shower basis.

The variance model used in Auger is based on the idea that the secondary particles enter a station’s water volume over some time interval, \mathcal{T} . The distribution of their arrival times is assumed to be an exponentially decreasing, $dN/dt \propto \exp(-t/\tau)$ (i.e. most particle are located at the leading edge of the shower front). For n arriving particles, the variance of the time when the first particle enters a particular station is given by:

$$V = \left(\frac{\mathcal{T}}{n}\right) \left(\frac{n-1}{n+1}\right). \quad (\text{A.11})$$

Because of the way that the Auger WCDs observe secondaries, the true number of particles is unknown. However, with the assumption that most of the signals are created by muons, n can be approximated by:

$$n(S_i, \theta) = S_i [\text{VEM}] \frac{2 h_{\text{stn}} \sin \theta + \pi R_{\text{stn}} \cos \theta}{\pi R_{\text{stn}}} \quad (\text{A.12})$$

⁴The errors on parameters are calculated by finding the parameter values for which the chi-square value increased by 1 with respect to the best fit values.

where h_{stn} and R_{stn} are the height (1.2 m) and radius (1.8 m) of a station and θ is the shower's zenith angle. Because the duration over which the signals are arriving is also unknown, this interval is estimated by finding the time at which the signal reaches half of its full value, T_{50} , and doubling it, $\mathcal{T} \simeq 2T_{50}$,

$$2 \times \int_0^{T_{50}} \frac{dS}{dt} dt = \int_0^{\infty} \frac{dS}{dt} dt. \quad (\text{A.13})$$

Finally, a constant offset is added to account for various effects from the electronics and data-taking rate. Thus, the time variance model is given by [103]

$$\sigma_t^2 = A(\theta)^2 \left(\frac{2T_{50}}{n(S_i, \theta)} \right) \left(\frac{n(S_i, \theta) - 1}{n(S_i, \theta) + 1} \right) + B(\theta)^2. \quad (\text{A.14})$$

The amplitudes, $A(\theta)$ and $B(\theta)$, are allowed to vary with angle,

$$\begin{aligned} A^2(\theta) &= 0.951 - 1.685 \cos \theta + 0.956 \cos^2 \theta \\ B^2(\theta) &= 89.5 + 94.8 \cos \theta - 7.4 \cos^2 \theta \end{aligned} \quad (\text{A.15})$$

This model was chosen and tuned before the new triggers were designed. As a result, the approximation that the signals are mostly created by muons is no longer valid for the **TOTd** and **MoPS** triggers. Thus, the use of the track length to calculate $n(S_i, \theta)$ is not a good estimate. Further, the function resulting from the assumption of an exponentially decreasing arrival time has serious problems for small signals. Given that signals which pass only the new triggers regularly result in values of $n(S_i, \theta) < 1$, equation A.14 can produce negative variances⁵.

A.3.1 Update to the Time Variance Model

To study the variance, the doublet⁶ stations were used. Showers were reconstructed as per normal and the timing residuals, dT_1 , (observed time minus expected time) were calculated for stations that had a doublet partner (eight such stations exist inside the 750 m array). Then the shower was reconstructed again after removing a

⁵In the code, this is avoided by using a lower limit of $n \geq 2$.

⁶These are stations that are placed $\simeq 10$ m from a normal station for studying various systematics of the **WCDS**.

standard station and including its doublet partner instead. This time the timing residual was calculated for the doublet, dT_2 . Replacing the station instead of just calculating both dT at once removes the systematic dependencies of the residuals as a result of the chosen shower front model.

The difference in time residuals, ΔT , can be defined,

$$\Delta T = dT_1 - dT_2, \quad (\text{A.16})$$

for which its variance, $\text{Var}[\Delta T]$, will be the sum of the variances⁷ of dT_i

$$\text{Var}[\Delta T] = \text{Var}[dT_1] + \text{Var}[dT_2]. \quad (\text{A.17})$$

Thus if the time variance model is describing the data well, then

$$\Omega = \Delta T / \sqrt{\text{Var}[\Delta T]} \quad (\text{A.18})$$

will follow a standard normal distribution (due to the central limit theorem) with a mean of 0 and a variance of 1.

This provides a natural way to test how well the data is being described by the current model. Over 400,000 6T5 events from the 750 m array were reconstructed. For events with at least one (triggered) doublet station, ΔT was calculated for all the doublet pairs. The data was then split up into new and old trigger sets (if a doublet or its pair had a new trigger ΔT was added to the new trigger set). The variance of Ω as a function of average signal and average T_{50} (the average of a doublet and its pair) are shown in the top left and top right panels of figure A.6, respectively. The new trigger data set is shown in red and the old trigger data set is shown in black. Deviations from 1 indicate where the current time variance model is not describing the true variance well. As a function of signal, the new and old triggers both show a similar trend. For large signal sizes, the model predicts a larger variance than what is measured while the opposite is true for small signal sizes. As a function of T_{50} , there is a large difference between the new and old triggers and the trends are not at all similar. Overall, the old triggers are being described fairly well but the new triggers are not.

⁷Note that σ_t^2 , referred to in the beginning of this section, will instead be referred to as $\text{Var}[dT]$ from hereon, for clarity.

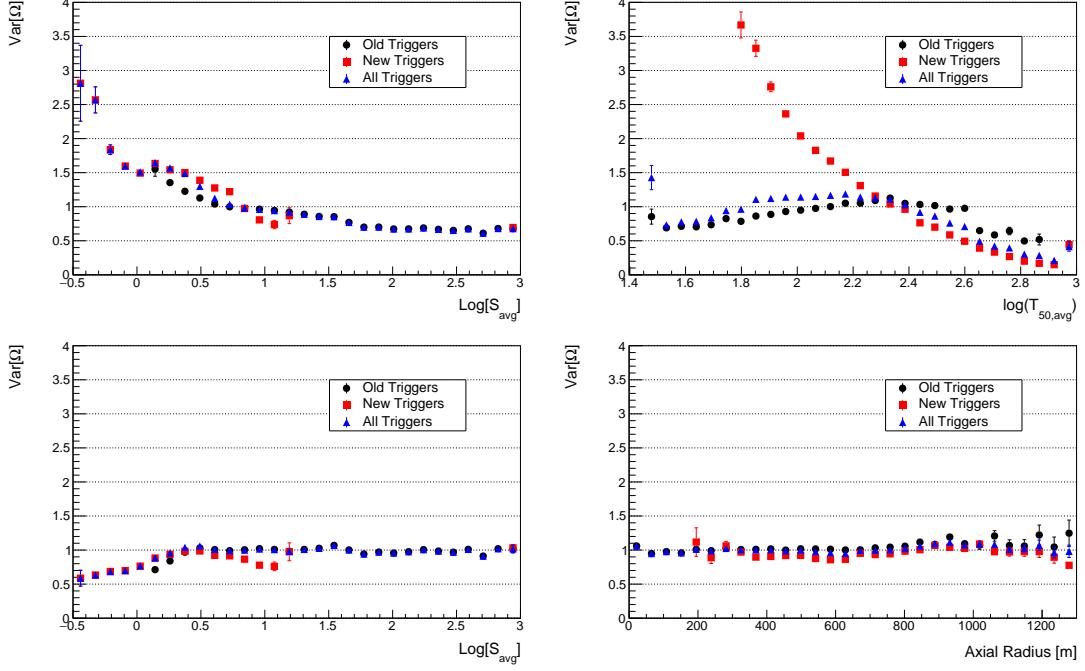


Figure A.6: The figures above show the variance of Ω (equation A.18). This is shown for the old triggers (black circles), the new triggers (red squares), and the combination of the two (blue triangles). The old variance model was parameterized according to signal (top left) and T_{50} (top right). The updated model uses signal (bottom left) and axial radius (bottom right). Error bars show the uncertainty of the variance.

In place of the current time variance model, an empirical one was developed with the following motivations in mind. Since it describes a variance, it should always produce positive values for all reasonable inputs. Further, it should depend on measurable quantities for which the signal variance is already strongly correlated. The form of the empirical model is given below

$$x = A \log(r) + B \log(S)$$

$$\log(\text{Var}[dT]) = \frac{C}{1 + \exp(-\frac{x-D}{E})} + F. \quad (\text{A.19})$$

The model includes six parameters, $\{A, B, C, D, E, F\}$, and depends on the station's axial radius, r , and signal size, S . This model was chosen because the variance of the start times was most strongly correlated with $\log(S)$ and generally followed a sigmoid shape (see figure A.7).

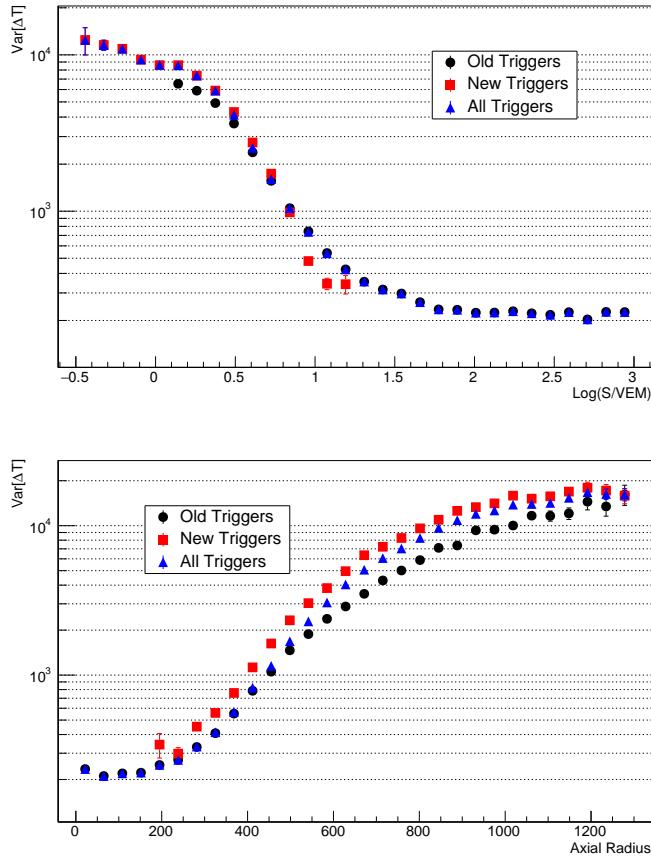


Figure A.7: The figures above show the variance of ΔT (equation A.17). This is shown for the old triggers (black circles), the new triggers (red squares), and the combination of the two (blue triangles) as a function of the signal (top) and axial radius (bottom) Error bars show the uncertainty of the variance.

The best model parameters were found by minimizing the log-likelihood function

$$-\ln \mathcal{L} = \sum_i \frac{\Delta T_i^2}{\text{Var}[dT_{1,i}] + \text{Var}[dT_{2,i}]} + \ln [2\pi(\text{Var}[dT_{1,i}] + \text{Var}[dT_{2,i}])]. \quad (\text{A.20})$$

The respective values are given in table A.2. To see how well the data is described by the updated model, $\text{Var}[\Omega]$ was plotted as a function of the station signal and axial radius in the bottom panels of figure A.6. The variance of Ω is now near 1 for all axial radii, though a small departure is seen for very small signals.

Paramter	Value	Error
A	-0.88	0.09
B	0.31	0.24
C	-2.27	0.08
D	-2.29	0.04
E	1.94	0.46
F	4.31	0.09

Table A.2: The values above are the best fit parameters for the time variance model (equation A.19) which describes the uncertainty on the start times of signals in stations.

A.3.2 The Calculation of Geometrical Errors

During the geometric reconstruction, the variance model is used as a weight in the chi-square fit (equation 5.4). So, the true test of the model is to see if the distribution of chi-sq probabilities is flat. The 6T5 events from 2014 were reconstructed using the updated and previous variance models. The χ^2 probabilities, which depend on the number of degrees of freedom, were calculated for each event. Recall that for station multiplicities of 3 or 4, the shower curvature is fixed so the numbers of free parameters are 0 and 1 respectively. So three station events were not used in this analysis. For station multiplicities $M > 4$, the curvature is fit and the number of degrees of freedom is $M - 4$.

The chi-square probabilities are shown in figure A.8 using the old model (top) and the new model (bottom). The distributions are shown for inclusive station multiplicity cuts. In the ideal case, the histograms would be flat indicating that the start times are sampled well out of a Gaussian distribution and that $\sqrt{\text{Var}[dT]}$ represents 1 standard deviation. In both cases, there is an excess of low-probability events. Some information about why this occurs can be gleaned by noting that the relative height of the low-probability peak with respect to the rest of the distribution is larger for high multiplicity events. In such events, most stations would have a low signal size and would be at $\gtrsim 450$ m. Note that the previous model underestimated the variance of small signals while the updated model overestimates the variance. Since both models produce the low-probability peaks, this may be evidence that the model of the shower front curvature may be the problem. During the geometric fit, the likelihood function (equation 5.4) assumes that the stations' start times are

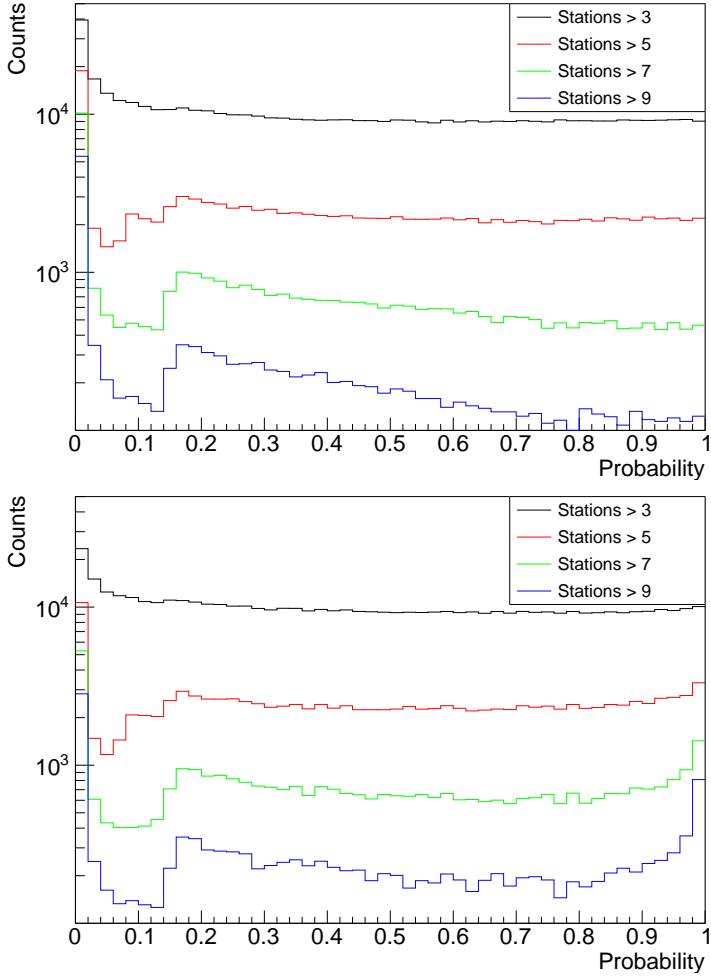


Figure A.8: The distribution of chi-square probabilities are presented above for the old (equation A.14) and new (equation A.19) time variance models, shown in the top and bottom panels, respectively. These distributions are given for various integral cuts in station multiplicity.

Gaussian fluctuations about a mean value given by the arrival time of the front. However, if the functional form of the shower front model (a spherical shape) does not describe that of the true shape, then this assumption will be invalid since the timing residuals will no longer be symmetric about the expected value.

There are also a number of other differences between the probability distributions for the two models. The old model includes a negative slope while the new model has an additional high-probability peak. Both of these features are again most evident at high multiplicities.

While the updated model produces a better description of the time variance

Data Set	Constant Intensity Cut			Energy Calibration		
	a	b	c	A [VEM]	B	
QGSJet p	1.72 ± 0.05	-1.69 ± 0.13	-1.79 ± 0.63	58.5 ± 0.9	0.99 ± 0.01	
QGSJet Fe	1.80 ± 0.05	-0.42 ± 0.15	0.02 ± 0.69	59.7 ± 1.0	0.99 ± 0.01	
QGSJet 50/50	1.77 ± 0.03	-1.09 ± 0.10	-1.01 ± 0.47	59.3 ± 0.9	0.99 ± 0.01	
EPOS p	1.66 ± 0.05	-2.46 ± 0.13	-3.67 ± 0.65	65.5 ± 1.0	0.99 ± 0.01	
EPOS Fe	1.88 ± 0.05	-0.11 ± 0.15	-0.82 ± 0.71	65.3 ± 0.9	1.00 ± 0.01	
EPOS 50/50	1.76 ± 0.04	-1.33 ± 0.10	-2.27 ± 0.50	65.6 ± 1.0	1.00 ± 0.01	
		a	b	c	A [PeV]	
Old Trig Data		1.68 ± 0.02	-1.16 ± 0.05	-2.19 ± 0.25	15.1 ± 0.6	0.99 ± 0.01

Table A.3: The CIC and energy calibration were performed on the various combinations of simulated old-trigger data. For each hadronic model the analysis chain was repeated for proton and iron primaries and a 50/50 mix of the two. The bottom line includes the respective parameters for the actual Auger old trigger data set. The calibration constants and their statistical errors are given.

for individual stations (see bottom panels of figure A.6), ultimately neither model produces the results that would be expected from data sampled from a normal distribution. This may be evidence that the way errors are assigned to the start times is being approached from the wrong angle, or it may be that the shower front model does not describe the data well at large axial radii. Further work is needed to discern the origin of the peaks in the chi-square probability distributions to understand how better to improve the variance model.

A.4 The T4 Efficiency for Old Triggers

Section 7.1 included two studies to calculate the T4 trigger efficiency of the 750 m array including the **TOTd** and **MoPS** triggers. One of the studies involved simulating air showers using two hadronic models. The analysis was repeated using the same 600,000 simulated events but with the rejection of stations that *only* passed the new triggers.

The first steps were to perform the energy calibration and **CIC** on the Monte Carlo data to find a conversion from the **MC** energy scale to that of the **FD**. The results of this process are shown in table A.3 for each of the model-primary combinations. To complete the conversion, the relevant parameters were also calculated for the old triggers on real Auger data for the years 2014–2017. Unlike

what was done for the new trigger data set in section 6.4, the energy calibration was only performed above $10^{17.5}$ eV. The respective parameters for the CIC and FD are given in the last row of the table.

A comparison of the efficiencies for the new and old triggers is shown in figure A.9 in differential energy bins. The top four plots compare model-primary

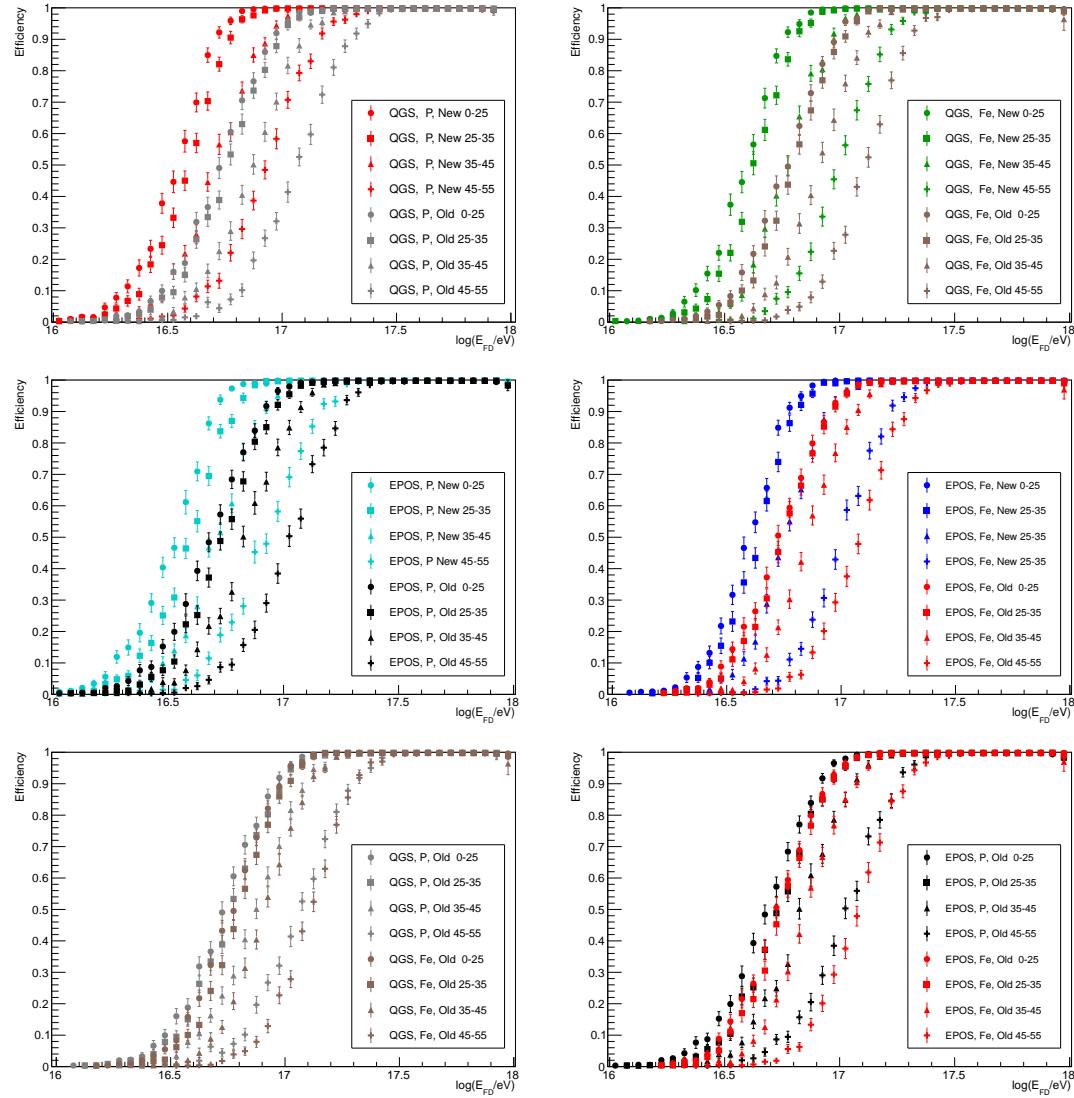


Figure A.9: The plots above compare the T4 efficiency as calculated by hadronic simulations of air showers. The top four plots compare the results from old and new triggers for different model-primary combinations. The bottom two plots show the efficiency agreement between primaries for a single model using only the old triggers.

Data Set	μ_0	μ_1	μ_2	μ_3
QGSJet p	16.736 ± 0.010	-0.38 ± 0.14	1.67 ± 0.48	-0.27 ± 0.46
QGSJet Fe	16.768 ± 0.010	-0.35 ± 0.14	1.64 ± 0.47	-0.20 ± 0.45
QGSJet 50/50	16.540 ± 0.007	-0.03 ± 0.09	0.84 ± 0.31	0.68 ± 0.30
EPOS p	16.684 ± 0.010	0.22 ± 0.14	-0.01 ± 0.47	1.10 ± 0.46
EPOS Fe	16.771 ± 0.009	-0.27 ± 0.12	0.83 ± 0.40	0.94 ± 0.39
EPOS 50/50	16.727 ± 0.007	-0.04 ± 0.09	0.43 ± 0.32	1.03 ± 0.30

Data Set	σ_0	σ_1
QGSJet p	0.262 ± 0.006	-0.023 ± 0.015
QGSJet Fe	0.245 ± 0.005	-0.052 ± 0.014
QGSJet 50/50	0.209 ± 0.004	0.042 ± 0.011
EPOS p	0.262 ± 0.006	-0.035 ± 0.015
EPOS Fe	0.248 ± 0.006	-0.058 ± 0.015
EPOS 50/50	0.252 ± 0.004	-0.046 ± 0.010

Table A.4: The differential T4 efficiency distributions using only the old triggers were fit as a function of energy and zenith angle, equations 7.8 and 7.9. The best fit values are given in the table above for the six hadronic-model simulations.

combinations with and without the new triggers. The difference in the efficiency is clear particularly for high zenith angles. A comparison between the primaries for the two models is shown in the bottom two plots, QGSJet on the left and EPOS on the right. The agreement between primaries is better here than was seen for the new triggers in figure 7.5 (which showed the array being significantly more efficient for proton showers).

Using these parameters, the same efficiency model (equations 7.8 and 7.9) was fit to the data as described in section 7.1.3. The results are shown in table A.4. A direct comparison between the efficiency parameterizations with and without including the new triggers has already been shown in figure 7.6 for two integral zenith angle ranges.

A.5 Energy Dependence of the CIC

The **CIC** method (see section 6.3) is a way of correcting for the additional attenuation that an **EAS** undergoes at high zenith angles. This is done by scaling the shower

sizes such that the distribution of local arrival directions is consistent with isotropy. The correction factor used in this thesis, $f_{\text{CIC}}(\theta)$ (equation 6.11), did not include any energy dependence. However, there are some reasons to expect that $f_{\text{CIC}}(\theta)$ *should* be energy dependent. With increasing energy, showers reach their maximum deeper in the atmosphere⁸ and thus have a smaller grammage over which to be attenuated before reaching the ground. Further, if the primary-mass composition evolves with energy, then the muonic fraction of the secondaries will also change. Since muons have a longer interaction length than electromagnetic particles, showers initiated by heavier primaries will not experience as strong of an attenuation over the same slant depth.

While these (and other) effects could introduce energy dependence to the relative attenuation factor, $f_{\text{CIC}}(\theta)$, it is hard to gauge the strength of their respective impacts. Instead, the net impact of all energy dependent effects can be studied by simply retuning the **CIC** at various shower-size amplitudes. The **CIC** method involves finding a description of $f_{\text{CIC}}(\theta)$ such that the number of events with $S_0/f_{\text{CIC}}(\theta) > S_{\text{cut}}$ is distributed according to what would be expected from isotropy (equation 6.9). Changing the value of S_{cut} , is analogous to tuning the **CIC** using only events above some equivalent energy cut, E_{cut} , to first order.

The tuning process, described in section 6.3, was repeated for the 55° data set using various values of S_{cut} . The values were chosen such that: the smallest cut would not be affected by triggering effects, the largest value would include a reasonable amount of statistics, and the spacings between cuts were larger than the typical uncertainty in shower size. The **CIC** curves are shown in the top panel of figure A.10 for eight values of $20 \leq S_{\text{cut}} \leq 100$. The difference between them is only apparent for vertical showers for which the largest difference in the correction factors is 15% ($1.25 \rightarrow 1.40$). The correction factors tuned at the largest values of S_{cut} correspond to the smallest corrections. However, since the tunings are performed on integral data sets, the results in this study are correlated.

The energy dependence of the **CIC** was then fit using a single analytical function. For the eight values of S_{cut} , the best fit parameters, a , b , and c , were individually fit to a parabola (i.e. $b(S) = b_0 + b_1 S + b_2 S^2$, etc.). Then b and c were fixed to their parameterized values ($b(S)$ and $c(S)$, respectively) and the **CIC** was tuned once

⁸For a given primary, the shower maximum occurs half an electromagnetic radiation length further into the atmosphere per decade in energy (see figure 3.9).

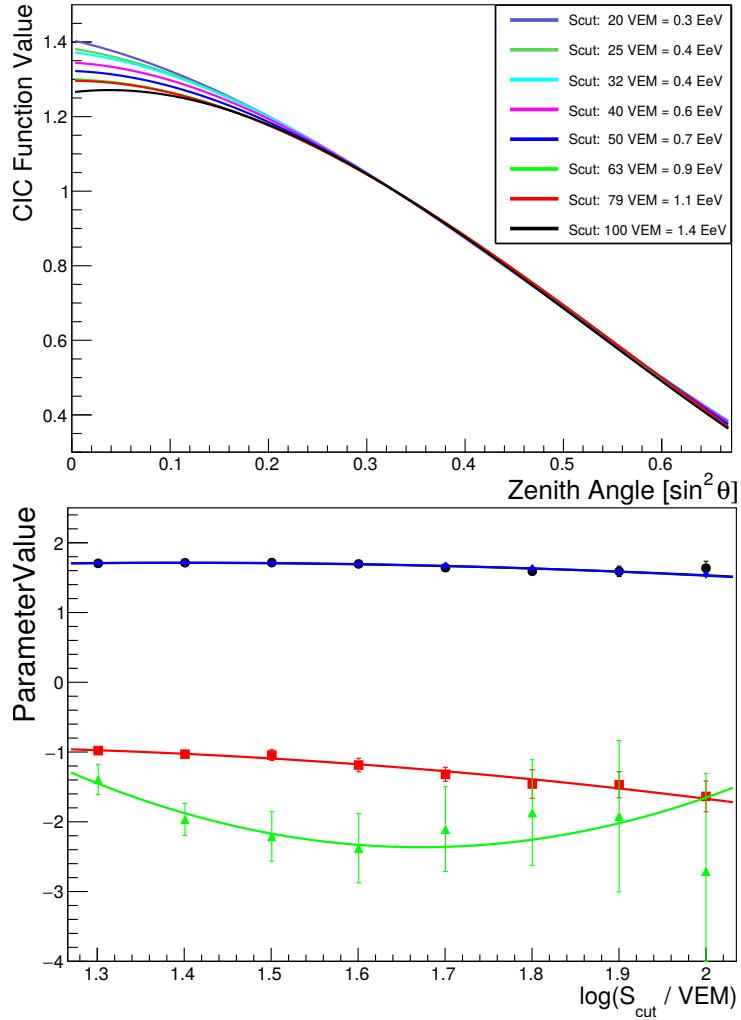


Figure A.10: Top: The CIC curves, $f_{\text{CIC}}(\theta)$, are shown for eight different values of S_{cut} . Bottom: The best fit parameters are shown (a in blue diamonds and black circles, see text, b in red squares, and c in green triangles) for the various cut thresholds.

again for the eight values of S_{cut} . Finally, $a(S)$ was fit. This second tuning iteration was done to take into account the correlations of a , b , and c . The parameter values and the fits are shown in the bottom panel in figure A.10. The values for a are shown by the black circles and blue diamonds for the first and second fits, respectively. Likewise, the values of parameters b and c are shown in red squares

and green triangles. The energy dependent **CIC** correction is then

$$\begin{aligned}
f_{\text{CIC}}(\theta, S) &= 1 + a(S)x + b(S)x^2 + c(S)x^3 \\
x &= \sin^2 35^\circ - \sin^2 \theta \\
a(S) &= 1.41 + 0.564S - 0.246S^2 \\
b(S) &= -1.83 + 1.73S - 0.826S^2 \\
c(S) &= 15.3 - 21.0S + 6.21S^2.
\end{aligned} \tag{A.21}$$

From the top plot of figure A.10, it is clear that the energy dependence only has strong implications for more vertical showers. What is not clear is how this will affect the energy assignment (recall that the calibration events do not have the same angular distribution as the **SD** events since they must also trigger the **FD**) nor how this will affect the spectrum. Using the **CIC** from equation A.21, the events were calibrated using the same set of events described in section 6.4. The calibration constants A and B did not change outside of the error bars. However, since the energy of an event is given by

$$E_{\text{SD}} = A \left(\frac{S}{f_{\text{CIC}}(\theta, S)} \right)^B, \tag{A.22}$$

the energies for (mostly vertical) events will change since $f_{\text{CIC}}(\theta) \neq f_{\text{CIC}}(\theta, S)$ in general. A comparison of the energy assignments is shown in figure A.11 for various energy bins. This plot shows the ratios of the energy calculated using the energy dependent **CIC** (equation A.21) and the energy independent one (equation 6.13). As expected, the largest change ($\sim 5\%$) is for the most vertical events (black). The smallest change is for the zenith bin $35^\circ < \theta < 45^\circ$. This is a result of the **CIC** being normalized at 35° and of the energy dependence of $f_{\text{CIC}}(\theta)$ being smaller for $\theta > 35^\circ$ than for $\theta < 35^\circ$ (see figure A.10).

To understand how this affects the spectral measurement, the cosmic ray flux was calculated with and without this correction. The energy spectrum was fit to an empirical model, Model-2 (see equation 8.10), then was broken up into four zenith bins of equal exposure. The results are shown in figure A.12. The top and bottom figures show the ratio of the fluxes with respect to the spectrum fit for the fixed and energy dependent **CICs**, respectively. As previously seen (figure 8.3), there is a statistically significant 5-8% spread in the flux for the various zenith bins

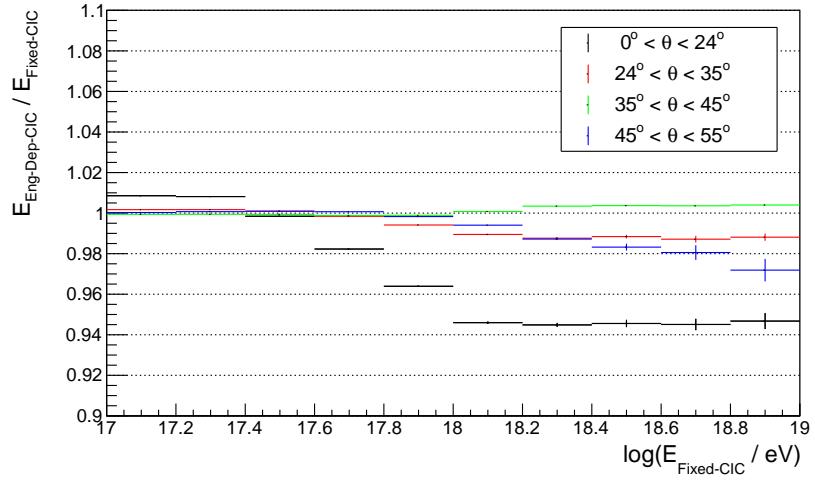


Figure A.11: The ratio of the energies assigned to each event when switching from a single **CIC** calibration to an energy dependent one are shown above. The events were split into four zenith angle bins that are of equal size in $\sin^2 \theta$.

when the fixed **CIC** is used. This spread is, however, absent above 400 PeV which is the energy at which the **CIC** was tuned ($\simeq 30$ VEM). In the bottom panel, the residuals indicate that this zenith dependence is removed in all but the first bin.

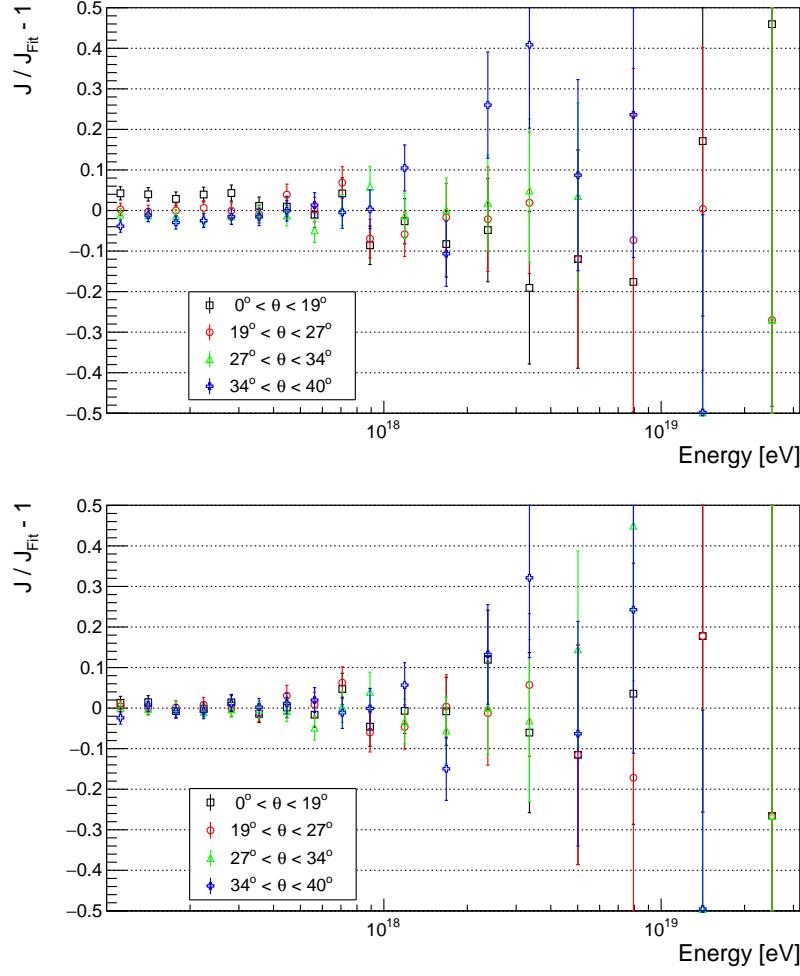


Figure A.12: The figures above show the residuals of the flux measurements with respect to an empirical fit of the combined spectrum in various zenith bins of equal exposure. The top plot is the spectrum without any energy dependent **CIC** correction while in the bottom the energy dependence is included.

Appendix B

Additional Plots and Information

B.1 Weather Correction

Cosmic rays in the ultra-high energy range are observed via their air showers. By using the Earth’s atmosphere as a radiator and absorber, the cascade of particles can be viewed while it is developing and once it reaches the ground. Because the atmosphere is dynamically changing, the [SD](#)’s slant depth and the average transverse momentum of the particles in the air shower are not time-independent. The impact that these weather effects have on air shower reconstructions was studied in [70].

A model to correct for these effects was described in section 6.2. Figure B.1 shows the agreement of the model with the data. The red points show the number of events measured as a function of hour (top) and day of the year (bottom). The black points indicate the expected values from the model which show good agreement. The model parameters (equation 6.7) were parameterized as function of zenith angle to account for the change in slant depth of the detector. The values calculated for the 750 m array are shown in figure B.2.

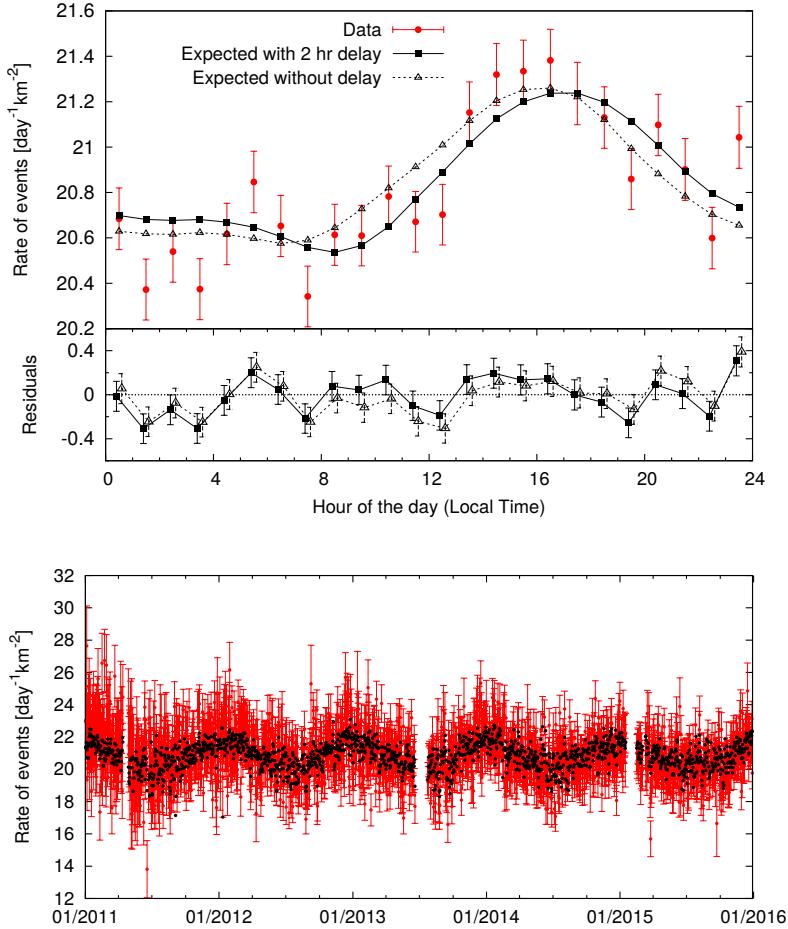


Figure B.1: The expected and measured rate of particles over the course of a day (top) and year (bottom).

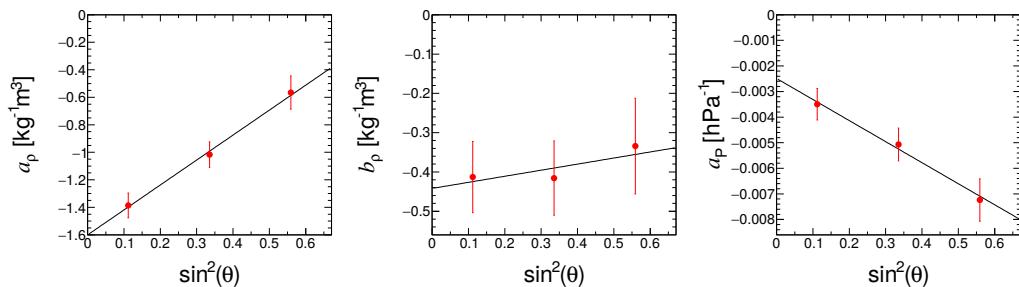


Figure B.2: The parameters used to describe the change in signal size with temperature and pressure (see equation 6.7) are shown above for the 750 m array.

B.2 Constant Intensity Cut

The **CIC** method described in section 6.3 finds a scaling of the observed shower sizes to correct for the attenuation in the atmosphere. This scaling is found by demanding that the number of events being measured has no local zenith angle dependence. That is, the distribution of observed events with shower sizes greater than $S_{\text{cut}} f_{\text{CIC}}(\theta)$ must have a linear CDF as a function of $\sin^2 \theta$. Historically, the **CIC** was performed by first choosing a particular “intensity” (though in practice, one would choose a number of events, N). Then the data would be separated into zenith angle bins with equal size in $\sin^2 \theta$. Finally, for each bin, the signal size above which there are N events would be identified. The shower size correction could then be found by fitting these signal size values to $\hat{S} f_{\text{CIC}}(\theta)$, where \hat{S} is an amplitude which corresponds to the signal size above which there are N events at θ_{ref} . Such a plot is shown for the 40° and 55° data sets in figure B.3. The

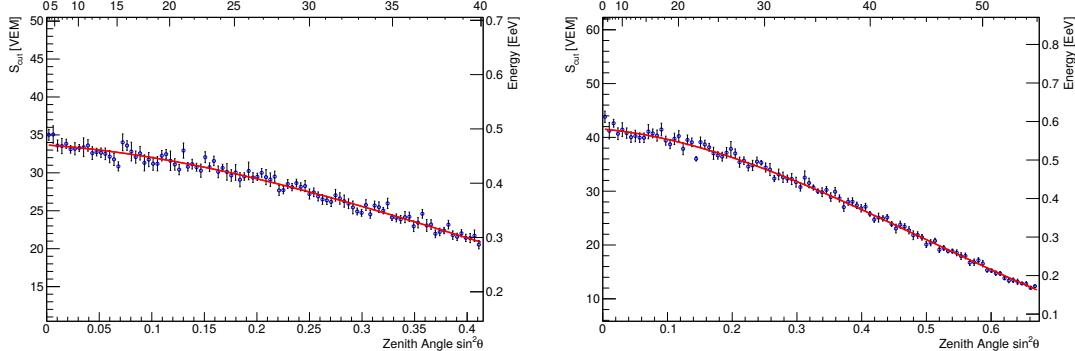


Figure B.3: The data above shows the signal size above which the number of events in each zenith bin is constant for the 40° (left) and 55° (right) data sets. The red lines are results of the **CIC** tuning using the unbinned method described in section 6.3.

blue circles show the signal values above which the number of events was constant ($N = 327$ and $N = 317$ events per zenith bin, respectively). The red lines are the results of the unbinned **CIC** tuning as described in the main text of the thesis i.e. it is not a direct fit to the blue points.

B.3 FD Energy Resolution

In section 6.4.3, the energy resolution of the **SD** was described. It can be broken up into two uncorrelated parts. The first part is related to the resolution of the shower size by the detector, σ_{det} , as a result of the sampling uncertainty. The second is due to shower-to-shower fluctuations, σ_{sh} . On an event-by-event basis, only σ_{det} can be estimated and thus the total energy resolution, $\sigma_{\text{SD}}^2 = \sigma_{\text{det}}^2 + \sigma_{\text{sh}}^2$, cannot.

Using the golden hybrid events, σ_{SD}^2 can be measured by calculating the ratio of the energy measurements by the two detectors. Assuming that the errors on the energy are Gaussian with width σ_{SD} and σ_{FD} , then the ratio of the energies, $z = x/y = E_{\text{SD}}/E_{\text{FD}}$, should follow the distribution given below [104]:

$$\text{PDF}(z; \mu_x, \sigma_x, \mu_y, \sigma_y) = \frac{b(z) d(z)}{a^3(z)} \frac{1}{\sqrt{2\pi}\sigma_x\sigma_y} f(z) + \frac{1}{a^2(z)\pi\sigma_x\sigma_y} e^{-c/2} \quad (\text{B.1})$$

$$\begin{aligned} a(z) &= \sqrt{\frac{1}{\sigma_x^2} z^2 + \frac{1}{\sigma_y^2}} & b(z) &= \frac{\mu_x}{\sigma_x^2} z + \frac{\mu_y}{\sigma_y^2} \\ c &= \frac{\mu_x^2}{\sigma_x^2} + \frac{\mu_y^2}{\sigma_y^2} & d(z) &= \exp\left(\frac{b^2(z) - c a^2(z)}{2a^2(z)}\right) \\ f(z) &= \Phi\left(\frac{b(z)}{a(z)}\right) - \Phi\left(-\frac{b(z)}{a(z)}\right) & \Phi(x) &= \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du. \end{aligned}$$

Unlike for the **SD**, the **FD** is able to estimate the complete energy uncertainty on an event-by-event basis since it observes the longitudinal profile. The **FD** energy uncertainties are shown in figure B.4 for the events used in the energy calibration data set. Thus, for individual energy bins, the energy resolution of the **SD** can be calculated by fitting the energy ratios, z_i , to equation B.1. Since σ_{FD} is already known and $\mu_{\text{FD}} \equiv 1$, the **SD**'s total energy resolution can be estimated from the fitted parameters. The results of this process have been shown in section 6.4.3.

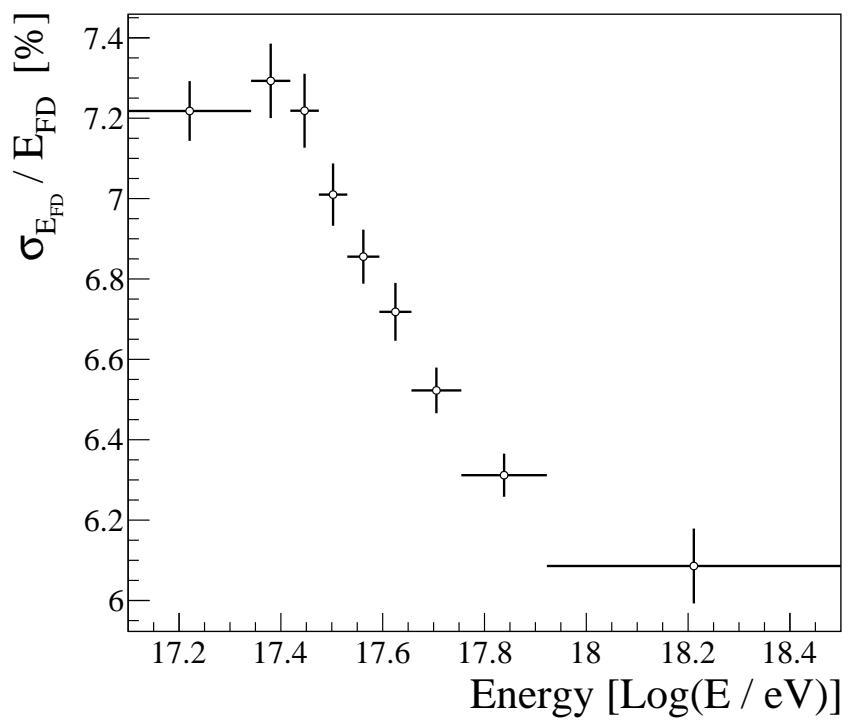


Figure B.4: The figure above shows the estimated energy resolution of the **FD** events used in the energy calibration.

B.4 Monte Carlo Energy Scale Offset

In section 7.1.2, the T4 triggering efficiency was calculated using simulated air showers. The simulations used two hadronic models, QGSJetII-04 [59] and EPOS-LHC [57]. In dedicated studies, these models have been found to under-produce muons compared to that of observed air showers [87, 88]. Since muons deposit more energy in WCDs than electromagnetic or hadronic particles, a real and simulated shower with the same energy will not produce the same amount of signal on the ground. This is just one effect that creates an offset in the MC and FD energy scales.

However, a conversion between the two scales can be found by using the CIC corrections and energy calibrations of the MC and real data sets (see equation 7.7). The relative energy shift is shown in figure B.5 as a function of zenith angle. The zenith dependence of the energy scale is a product of the difference in CIC

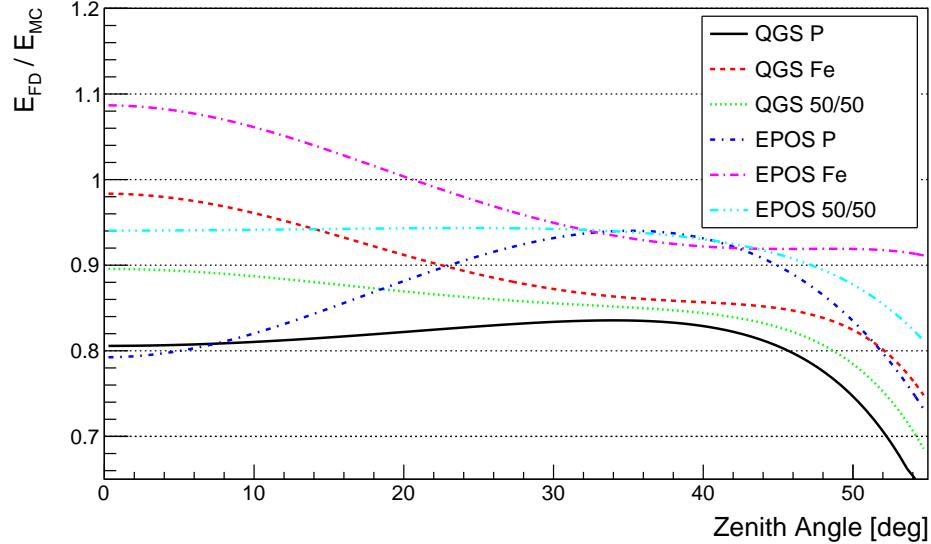


Figure B.5: The curves above show the ratio of the MC and FD energy scales as a function of zenith angle (see equation 7.7).

calibrations. Thus the flattest curves in figure B.5 correspond to data sets for which $f_{\text{CIC, MC}}(\theta) \simeq f_{\text{CIC, FD}}(\theta)$.

B.5 Unfolding Correction Factors

The forward folding procedure includes a scaling of the spectral points given by equation 8.8. As a requisite for the forward folding, a model for the true spectrum must be supplied. In section 8.2.2, unfolded spectra were compared using three models. The correction factors are shown in figure B.6. Generally they all have a

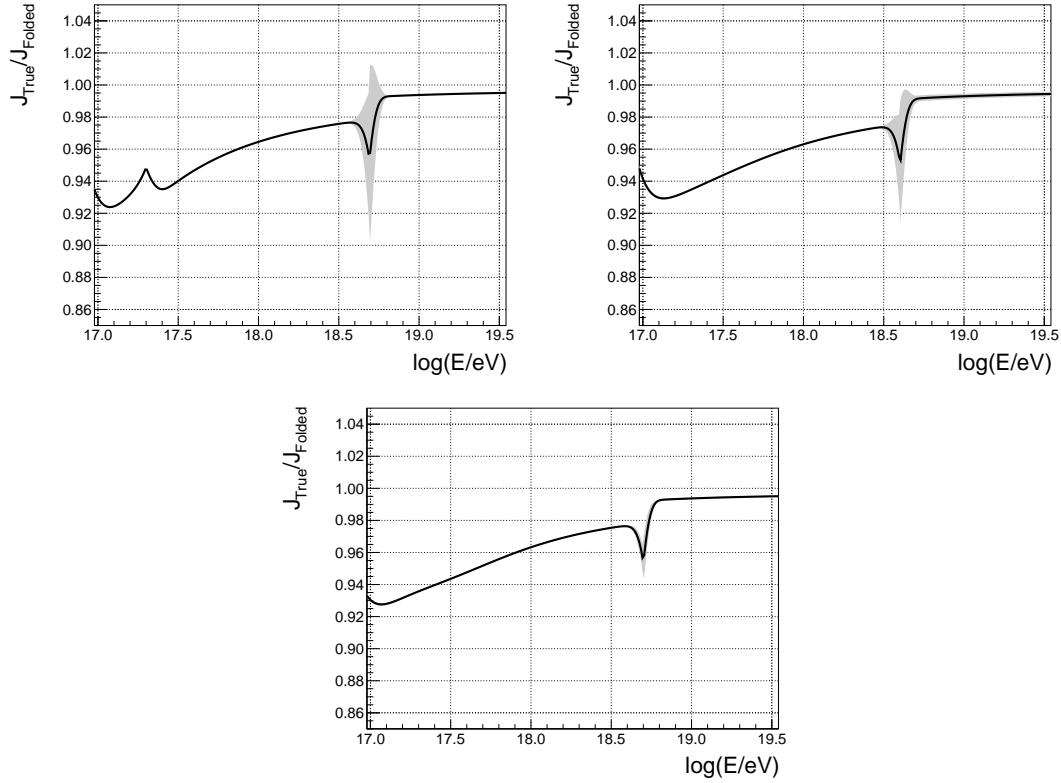


Figure B.6: The correction factors from the unfolding are shown above. They were calculated using Model-1 (top left, equation 8.9), Model-2 (top right, equation 8.10), Model-3 (bottom, equation 8.11). Statistical errors are represented by the grey band.

similar overall shape with a sharp break at the ankle ($\simeq 10^{18.70}$ eV). However, Model-1 (top left) includes another kink since it also has a sharp break near the second knee (unlike the other two models which use a smooth break in the spectrum).

B.6 Flux Measurements Using the 750 m Array

The measurement of the flux was shown in chapter 8. The corresponding values for the flux using the data set for zenith angles below 40° and energies above 10^{17} eV are given in table B.1.

$\log\left(\frac{E}{\text{eV}}\right)$	Scaled Flux $J(E) E^3$ [eV 2 km $^{-2}$ sr $^{-1}$ yr $^{-1}$]		
	$\frac{dN}{dE d\epsilon} E^3 \times 10^{-37}$	$\sigma_{\text{stat.}}^- \times 10^{-37}$	$\sigma_{\text{stat.}}^+ \times 10^{-37}$
17.05	10.56	0.03	0.03
17.15	10.47	0.03	0.03
17.25	10.20	0.04	0.04
17.35	9.88	0.05	0.05
17.45	9.44	0.06	0.06
17.55	8.82	0.07	0.07
17.65	8.37	0.08	0.09
17.75	7.66	0.10	0.10
17.85	7.41	0.13	0.13
17.95	6.56	0.15	0.15
18.05	6.22	0.18	0.19
18.15	5.44	0.22	0.23
18.25	5.23	0.27	0.28
18.35	5.16	0.34	0.36
18.45	4.70	0.40	0.44
18.55	4.93	0.52	0.58
18.65	4.00	0.59	0.69
18.75	4.27	0.77	0.93
18.85	4.51	1.00	1.24
19.10	7.26	1.11	1.29
19.40	9.89	3.63	5.30

Table B.1: The values above correspond to the measurement of the flux scaled by the energy ($J E^3$) as shown in figure 8.10.

Appendix C

List of Acronyms

ADC	analog to digital converter	33
AGASA	Akeno Giant Air Shower Array	3
AGN	active galactic nuclei.....	7
CDAS	Central Data Acquisition System	44
CDF	cumulative distribution function	62
CIC	Constant Intensity Cut	88
CLF	Central Laser Facility	30
CMB	Cosmic Microwave Background	3
DPMJET	Dual Parton Model with JETs.....	23
EAS	extensive air shower	2
EPOS	Energy conserving quantum mechanical multiple scattering approach based on partons, off-shell remnants, and splitting of parton ladders	
FADC	flash analog to digital converter	32
FD	fluorescence detector	27
FOV	field of view	88
GRB	gamma-ray burst	7
GZK	Greisen Zatsepin Kuz'min	3
HEAT	High Elevation Auger Telescope	28
HiRes	High Resolution Fly's Eye	3

ICRC	International Cosmic Ray Conference.....	136
IGM	intergalactic medium	7
QGSJET	Quark Gluon String Model with JETs.....	23
LDF	lateral distribution function	36
LHC	Large Hadron Collider.....	23
MC	Monte Carlo.....	21
MoPS	multiplicity of positive steps.....	45
NKG	Nishimura-Kamata-Greisen	21
PMT	photomultiplier tube	29
SD	surface detector	27
SDP	shower detector plane	30
SNR	supernova remnant	7
TA	Telescope Array	13
Th	threshold.....	44
TOT	time-over-threshold	44
TOTd	time-over-threshold deconvoluted	45
UHE	ultra-high energy	6
UHECR	ultra-high energy cosmic ray	
VEM	vertical equivalent muon.....	34
WCD	water Cherenkov detector	32
XLF	eXtreme Laser Facility	30

Bibliography

- [1] WULF, T. (1910) “Beobachtungen über Strahlung hoher Durchdringungsfähigkeit auf dem Eiffelturm,” *Physikalische Zeitschrift*, **11**, pp. 811–813.
- [2] HESS, V. F. (1912) “Über Beobachtungen der durchdringenden Strahlung bei sieben Freiballonfahrten,” *Physikalische Zeitschrift*, **13**, p. 1084.
- [3] JOHNSON, T. H. (1933) “The azimuthal asymmetry of the cosmic radiation,” *Physical Review*, **43**(10), p. 834.
- [4] ALVAREZ, L. and A. H. COMPTON (1933) “A positively charged component of cosmic rays,” *Physical Review*, **43**(10), p. 835.
- [5] ROSSI, B. (1934) “Directional measurements on the cosmic rays near the geomagnetic equator,” *Physical Review*, **45**(3), p. 212.
- [6] AUGER, P. ET AL. (1939) “Extensive cosmic-ray showers,” *Reviews of Modern Physics*, **11**(3-4), p. 288.
- [7] LINSLEY, J. (1963) “Evidence for a primary cosmic-ray particle with energy 10^{20} eV,” *Physical Review Letters*, **10**(4), p. 146.
- [8] PENZIAS, A. A. and R. W. WILSON (1965) “A Measurement of Excess Antenna Temperature at 4080 Mc/s.” *The Astrophysical Journal*, **142**, pp. 419–421.
- [9] GREISEN, K. (1966) “End to the cosmic-ray spectrum?” *Physical Review Letters*, **16**(17), p. 748.
- [10] ZATSEPIN, G. T. and V. A. KUZ’MIN (1966) “Upper limit of the spectrum of cosmic rays,” *ZhETF Pisma Redaktsiiu*, **4**, p. 114.
- [11] ABBASI, R. U. ET AL. (2008) “First Observation of the Greisen-Zatsepin-Kuzmin Suppression,” *Physical Review Letters*, **100**, p. 101.
- [12] TAKEDA, M. ET AL. (1998) “Extension of the cosmic-ray energy spectrum beyond the predicted Greisen-Zatsepin-Kuz’mín cutoff,” *Physical Review Letters*, **81**(6), p. 1163.

- [13] CHUNG, D. J. H. ET AL. (1999) “Production of massive particles during reheating,” *Physical Review D*, **60**, p. 063504.
- [14] BOYANOVSKY, D. ET AL. (1995) “Linear versus nonlinear relaxation: Consequences for reheating and thermalization,” *Physical Review D*, **52**, pp. 6805–6827.
- [15] DE VEGA, H. and N. SANCHEZ (2003) “Extreme energy cosmic rays: bottom-up vs. top-down scenarii,” in *The Early Universe and the Cosmic Microwave Background: Theory and Observations*, Springer, pp. 433–449.
- [16] PORTER, N. (1960) “The Dirac monopole as a constituent of primary cosmic radiation,” *Il Nuovo Cimento*, **16**(5), pp. 958–959.
- [17] AYNUTDINOV, V. ET AL. (2008) “Search for relativistic magnetic monopoles with the Baikal Neutrino Telescope,” *Astroparticle Physics*, **29**(6), pp. 366–372.
- [18] DETRIXHE, M. ET AL. (2011) “Ultrarelativistic magnetic monopole search with the ANITA-II balloon-borne radio interferometer,” *Physical Review D*, **83**(2), p. 023513.
- [19] ABBASI, R. ET AL. (2013) “Search for relativistic magnetic monopoles with IceCube,” *Physical Review D*, **87**(2), p. 022001.
- [20] AAB, A. ET AL. (2016) “Search for ultrarelativistic magnetic monopoles with the Pierre Auger Observatory,” *Physical Review D*, **94**(8), p. 082002.
- [21] NETTERFIELD, C. ET AL. (2002) “A measurement by BOOMERANG of multiple peaks in the angular power spectrum of the cosmic microwave background,” *The Astrophysical Journal*, **571**(2), p. 604.
- [22] LEE, A. ET AL. (2001) “A High spatial resolution analysis of the MAXIMA-1 cosmic microwave background anisotropy data,” *The Astrophysical Journal Letters*, **561**(1), p. L1.
- [23] HALVERSON, N. ET AL. (2002) “Degree angular scale interferometer first results: a measurement of the cosmic microwave background angular power spectrum,” *The Astrophysical Journal*, **568**(1), p. 38.
- [24] ADE, P. ET AL. (2014) “Planck 2013 results. XXV. Searches for cosmic strings and other topological defects,” *Astronomy & Astrophysics*, **571**, p. A25.

- [25] SCHERINI, V., THE PIERRE AUGER COLLABORATION, ET AL. (2011) “Search for ultra-high energy photons with the Pierre Auger Observatory,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **630**(1), pp. 226–229.
- [26] F. FENU FOR THE PIERRE AUGER COLLABORATION (2017) “The cosmic ray energy spectrum measured using the Pierre Auger Observatory,” in *The 35th International Cosmic Ray Conference*, vol. 301.
- [27] FERMI, E. (1949) “On the origin of the cosmic radiation,” *Physical Review*, **75**(8), p. 1169.
- [28] KOTERA, K. and A. V. OLINTO (2011) “The astrophysics of ultrahigh-energy cosmic rays,” *Annual Review of Astronomy and Astrophysics*, **49**, pp. 119–153.
- [29] ENGEL, R. ET AL. (2011) “Extensive air showers and hadronic interactions at high energy,” *Annual Review of Nuclear and Particle Science*, **61**, pp. 467–489.
- [30] NAVARRA, G. ET AL. (2004) “KASCADE-Grande: a large acceptance, high-resolution cosmic-ray detector up to 10^{18} eV,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **518**(1), pp. 207–209.
- [31] ABBASI, R. ET AL. (2013) “IceTop: The surface component of IceCube,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **700**, pp. 188–220.
- [32] AGLIETTA, M. ET AL. (1999) “The EAS size spectrum and the cosmic ray energy spectrum in the region 10^{15} – 10^{16} eV,” *Astroparticle Physics*, **10**(1), pp. 1–9.
- [33] BEREZHNEV, S. ET AL. (2012) “The Tunka-133 EAS Cherenkov light array: status of 2011,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **692**, pp. 98–105.
- [34] PTUSKIN, V. ET AL. (1993) “Diffusion and drift of very high energy cosmic rays in galactic magnetic fields,” *Astronomy & Astrophysics*, **268**, pp. 726–735.
- [35] KLAGES, H. ET AL. (1997) “The KASCADE experiment,” *Nuclear Physics B-Proceedings Supplements*, **52**(3), pp. 92–102.
- [36] ANTONI, T. ET AL. (2005) “KASCADE measurements of energy spectra for elemental groups of cosmic rays: Results and open problems,” *Astroparticle Physics*, **24**(1), pp. 1–25.

- [37] APEL, W. ET AL. (2011) “Kneelike structure in the spectrum of the heavy component of cosmic rays observed with KASCADE-Grande,” *Physical Review Letters*, **107**(17), p. 171104.
- [38] APEL, W. ET AL. (2013) “KASCADE-Grande measurements of energy spectra for elemental groups of cosmic rays,” *Astroparticle Physics*, **47**, pp. 54–66.
- [39] AAB, A. ET AL. (2017) “Observation of a large-scale anisotropy in the arrival directions of cosmic rays above 8×10^{18} eV,” *Science*, **357**(6357), pp. 1266–1270.
- [40] HILL, C. T. and D. N. SCHRAMM (1985) “Ultrahigh-energy cosmic-ray spectrum,” *Physical Review D*, **31**(3), p. 564.
- [41] ALOISIO, R. ET AL. (2007) “A dip in the UHECR spectrum and the transition from galactic to extragalactic cosmic rays,” *Astroparticle Physics*, **27**(1), pp. 76–91.
- [42] ALOISIO, R. ET AL. (2014) “Ultra high energy cosmic rays: implications of Auger data for source spectra and chemical composition,” *Journal of Cosmology and Astroparticle Physics*, **2014**(10), p. 020.
- [43] CARLSON, J. and J. OPPENHEIMER (1937) “On multiplicative showers,” *Physical Review*, **51**(4), p. 220.
- [44] HEITLER, W. (1936) *Quantum theory of radiation*, Oxford: Clarendon Press.
- [45] MATTHEWS, J. (2005) “A Heitler model of extensive air showers,” *Astroparticle Physics*, **22**(5), pp. 387–397.
- [46] GAISSER, T. K., R. ENGEL, and E. RESCONI (2016) *Cosmic rays and particle physics*, Cambridge University Press.
- [47] PATRIGNANI, C., P. D. GROUP, ET AL. (2016) “Review of particle physics,” *Chinese Physics C*, **40**(10), p. 100001.
- [48] ROSSI, B. and K. GREISEN (1941) “Cosmic-ray theory,” *Reviews of Modern Physics*, **13**(4), p. 240.
- [49] LIPARI, P. (2009) “Concepts of age and universality in cosmic ray showers,” *Physical Review D*, **79**(6), p. 063001.
- [50] KAMATA, K. and J. NISHIMURA (1958) “The lateral and the angular structure functions of electron showers,” *Progress of Theoretical Physics Supplement*, **6**, pp. 93–155.

- [51] GREISEN, K. (1960) “Cosmic ray showers,” *Annual Review of Nuclear Science*, **10**(1), pp. 63–108.
- [52] GILLER, M. ET AL. (2004) “Energy spectra of electrons in the extensive air showers of ultra-high energy,” *Journal of Physics G: Nuclear and Particle Physics*, **30**(2), p. 97.
- [53] NERLING, F. ET AL. (2006) “Universality of electron distributions in high-energy air showers - Description of Cherenkov light production,” *Astroparticle Physics*, **24**(6), pp. 421–437.
- [54] GILLER, M. ET AL. (2005) “Similarity of extensive air showers with respect to the shower age,” *Journal of Physics G: Nuclear and Particle Physics*, **31**(8), p. 947.
- [55] GILLER, M. ET AL. (2015) “An extended universality of electron distributions in cosmic ray showers of high energies and its application,” *Astroparticle Physics*, **60**, pp. 92–104.
- [56] ROESLER, S., R. ENGEL, and J. RANFT (2001) “The Monte Carlo Event Generator DPMJET-III,” in *Advanced Monte Carlo for Radiation Physics, Particle Transport Simulation and Applications*, Springer, pp. 1033–1038.
- [57] WERNER, K., F.-M. LIU, and T. PIEROG (2006) “Parton ladder splitting and the rapidity dependence of transverse momentum spectra in deuteron-gold collisions at the BNL Relativistic Heavy Ion Collider,” *Physical Review C*, **74**(4), p. 044902.
- [58] OSTAPCHENKO, S. (2006) “Nonlinear screening effects in high energy hadronic interactions,” *Physical Review D*, **74**(1), p. 014026.
- [59] OSTAPCHENKO, S. (2011) “Monte Carlo treatment of hadronic interactions in enhanced Pomeron scheme: QGSJET-II model,” *Physical Review D*, **83**(1), p. 014018.
- [60] RIEHN, F. ET AL. (2015) “A new version of the event generator Sibyll,” *arXiv preprint arXiv:1510.00568*.
- [61] CASO, C. ET AL. (1998) “Review of particle physics,” *European Physical Journal C*, **3**(1-4), pp. 1–+.
- [62] AABOUD, M. ET AL. (2016) “Measurement of the Inelastic Proton-Proton Cross Section at $s=13$ TeV with the ATLAS Detector at the LHC,” *Physical Review Letters*, **117**(18), p. 182002.

- [63] PIEROG, T. and I. KIT (2017) “Air Shower Simulation with a New Generation of post-LHC Hadronic Interaction Models in CORSIKA,” in *The 35th International Cosmic Ray Conference*.
- [64] SALTZBERG, D. ET AL. (2001) “Observation of the Askaryan effect: Coherent microwave Cherenkov emission from charge asymmetry in high-energy particle cascades,” *Physical Review Letters*, **86**(13), p. 2802.
- [65] BARBOSA, H. M. ET AL. (2004) “Determination of the calorimetric energy in extensive air showers,” *Astroparticle Physics*, **22**(2), pp. 159–166.
- [66] V. VERZI FOR THE PIERRE AUGER COLLABORATION (2013) “The energy scale of the Pierre Auger Observatory,” *The 33rd International Cosmic Ray Conference*.
- [67] GAISSER, T. K. and A. M. HILLAS (1977) “Reliability of the method of constant intensity cuts for reconstructing the average development of vertical showers,” in *International Cosmic Ray Conference*, vol. 8, pp. 353–357.
- [68] KLIMUSHIN, S. ET AL. (2001) “Precise parametrizations of muon energy losses in water,” *arXiv preprint hep-ph/0106010*.
- [69] BERTOU, XAVIER FOR THE PIERRE AUGER COLLABORATION (2006) “Calibration of the surface array of the Pierre Auger Observatory,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **568**(2), pp. 839–846.
- [70] AAB, A. ET AL. (2017) “Impact of atmospheric effects on the energy reconstruction of air showers observed by the surface detectors of the Pierre Auger Observatory,” *Journal of Instrumentation*, **12**(02), p. P02006.
- [71] BELLIDO, J. (2017) “Depth of maximum of air-shower profiles at the Pierre Auger Observatory: Measurements above $10^{17.2}$ eV and Composition Implications,” .
- [72] AAB, A. ET AL. (2017) “Combined fit of spectrum and composition data as measured by the Pierre Auger Observatory,” *Journal of Cosmology and Astroparticle Physics*, **2017**(04), p. 038.
- [73] FARRAR, G. R. (2014) “The Galactic magnetic field and ultrahigh-energy cosmic ray deflections,” *Comptes Rendus Physique*, **15**(4), pp. 339–348.
- [74] ABBASI, R. ET AL. (2014) “Indications of intermediate-scale anisotropy of cosmic rays with energy greater than 57 EeV in the northern sky measured with the surface detector of the telescope array experiment,” *The Astrophysical Journal Letters*, **790**(2), p. L21.

- [75] AAB, A. ET AL. (2015) “The Pierre Auger cosmic ray Observatory,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **798**, pp. 172–213.
- [76] ABRAHAM, J. ET AL. (2010) “Trigger and aperture of the surface detector array of the Pierre Auger Observatory,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **613**(1), pp. 29–39.
- [77] WILSON, E. B. (1927) “Probable inference, the law of succession, and statistical inference,” *Journal of the American Statistical Association*, **22**(158), pp. 209–212.
- [78] R. COLALILLO FOR THE PIERRE AUGER COLLABORATION (2017) “Peculiar lightning-related events observed by the surface detector of the Pierre Auger Observatory,” in *The 35th International Cosmic Ray Conference*.
- [79] JAMES, F. (1994) *MINUIT function minimization and error analysis: reference manual version 94.1*, Tech. rep.
- [80] NEWTON, D., J. KNAPP, and A. WATSON (2007) “The optimum distance at which to determine the size of a giant air shower,” *Astroparticle Physics*, **26**(6), pp. 414–419.
- [81] AVE, M. ET AL. (2007) “The accuracy of signal measurement with the water Cherenkov detectors of the Pierre Auger Observatory,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **578**(1), pp. 180 – 184.
- [82] A. COLEMAN FOR THE PIERRE AUGER COLLABORATION (2017) “The influence of weather effects on the reconstruction of extensive air showers at the Pierre Auger Observatory,” in *The 35th International Cosmic Ray Conference*.
- [83] HERCZEGH, J. ET AL. (1961) “Observations of extensive air showers near the maximum of their longitudinal development,” *Physical Review Letters*, **6**(1), p. 22.
- [84] ANDERSON, T. W. and D. A. DARLING (1952) “Asymptotic theory of certain goodness of fit criteria based on stochastic processes,” *The annals of mathematical statistics*, pp. 193–212.
- [85] EFRON, B. and R. J. TIBSHIRANI (1994) *An introduction to the bootstrap*, CRC press.

- [86] M. AVE FOR THE PIERRE AUGER COLLABORATION (2007) “Reconstruction accuracy of the surface detector array of the Pierre Auger Observatory,” *The 30th International Cosmic Ray Conference*.
- [87] PIEROG, T. and K. WERNER (2008) “Muon production in extended air shower simulations,” *Physical Review Letters*, **101**(17), p. 171101.
- [88] AAB, A. ET AL. (2015) “Muons in air showers at the Pierre Auger Observatory: Mean number in highly inclined events,” *Physical Review D*, **91**, p. 032003.
- [89] HECK, D. ET AL. (1998) *CORSIKA: A Monte Carlo code to simulate extensive air showers, Tech. rep.*
- [90] AGOSTINELLI, S. ET AL. (2003) “GEANT4-a simulation toolkit,” *Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **506**(3), pp. 250–303.
- [91] AMENOMORI, M. ET AL. (2008) “The all-particle spectrum of primary cosmic rays in the wide energy range from 10^{14} to 10^{17} eV observed with the Tibet-III air-shower array,” *The Astrophysical Journal*, **678**(2), p. 1165.
- [92] NAGANO, M. ET AL. (1992) “Energy spectrum of primary cosmic rays above $10^{17.0}$ eV determined from extensive air shower experiments at Akeno,” *Journal of Physics G: Nuclear and Particle Physics*, **18**(2), p. 423.
- [93] BIRD, D. ET AL. (1994) “The cosmic-ray energy spectrum observed by the Fly’s Eye,” *The Astrophysical Journal*, **424**, pp. 491–502.
- [94] TER-ANTONYAN, S. (2014) “Sharp knee phenomenon of primary cosmic ray energy spectrum,” *Physical Review D*, **89**(12), p. 123003.
- [95] AARTSEN, M. G. ET AL. (2013) “Measurement of the cosmic ray energy spectrum with IceTop-73,” *Physical Review D*, **88**(4), p. 042004.
- [96] BERTAINA, M. ET AL. (2015) “KASCADE-Grande energy spectrum of cosmic rays interpreted with post-LHC hadronic interaction models,” *PoS*, p. 359.
- [97] ABBASI, R. ET AL. (2016) “The energy spectrum of cosmic rays above $10^{17.2}$ eV measured by the fluorescence detectors of the Telescope Array experiment in seven years,” *Astroparticle Physics*, **80**, pp. 131–140.
- [98] TSUNESADA, Y. ET AL. (2017) “Energy Spectrum of Ultra-High-Energy Cosmic Rays Measured by The Telescope Array,” in *The 35th International Cosmic Ray Conference*.

- [99] ABBASI, R. ET AL. (2018) “The Cosmic-Ray Energy Spectrum between 2 PeV and 2 EeV Observed with the TALE detector in monocular mode,” *arXiv preprint arXiv:1803.01288*.
- [100] PROSIN, V. ET AL. (2016) “Results from Tunka-133 (5 years observation) and from the Tunka-HiSCORE prototype,” in *EPJ Web of Conferences*, vol. 121, EDP Sciences, p. 03004.
- [101] KNURENKO, S. ET AL. (2013) “Ultra-high energy cosmic rays: 40 years retrospective of continuous observations at the Yakutsk array: Part 1. Cosmic ray spectrum in the energy range 10^{15} - 10^{18} eV and its interpretation,” .
- [102] DAI, H. ET AL. (1988) “On the energy estimation of ultra-high-energy cosmic rays observed with the surface detector array,” *Journal of Physics G: Nuclear Physics*, **14**(6), p. 793.
- [103] BONIFAZI, C. ET AL. (2008) “A model for the time uncertainty measurements in the Auger surface detector array,” *Astroparticle Physics*, **28**(6), pp. 523–528.
- [104] HINKLEY, D. V. (1969) “On the ratio of two correlated normal random variables,” *Biometrika*, **56**(3), pp. 635–639.

Vita

Alan Coleman

EDUCATION

Ph.D. Pennsylvania State University, <i>University Park, PA</i>	Fall 2018
B.S. University of Michigan, <i>Ann Arbor, MI</i>	Winter 2012

FELLOWSHIPS AND HONORS

Duncan, Downsborough, and Miller Graduate Fellowship	Dec 2017
Chateaubriand Fellowship	Jan 2017
Helmholtz Alliance for Astroparticle Physics iPROGRESS Grant	June 2016
Physics Department Graduate Teaching Award	Apr 2015

SELECTED PUBLICATIONS

- “Combined fit of spectrum and composition data as measured by the Pierre Auger Observatory”, The Pierre Auger Collaboration, *Journal of Cosmology and Astroparticle Physics* 04 (2018) 038
- “Inferences on Mass Composition and Tests of Hadronic Interactions from 0.3 to 100 EeV using the water-Cherenkov Detectors of the Pierre Auger Observatory”, The Pierre Auger Collaboration, *Physical Review D* 96 (2017) 122003
- “Observation of a Large-scale Anisotropy in the Arrival Directions of Cosmic Rays above 8×10^{18} eV”, The Pierre Auger Collaboration, *Science* 357 (2017) 1266-1270
- “The influence of weather effects on the reconstruction of extensive air showers at the Pierre Auger Observatory”, Alan Coleman for the Pierre Auger Collaboration, in *The 35th International Cosmic Ray Conference*