A SURVEY OF THE HIGHEST-ENERGY ASTROPHYSICAL SOURCES WITH THE HAWC OBSERVATORY

A Dissertation in
Physics
by
Kelly Malone

© 2018 Kelly Malone

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

December 2018
The dissertation of Kelly Malone was reviewed and approved* by the following:

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

Stephane Coutu  
Professor of Physics and of Astronomy and Astrophysics

Doug Cowen  
Professor of Physics and of Astronomy and Astrophysics

Peter Mészáros  
Eberly Chair Professor of Astronomy and Astrophysics, Professor of Physics

Richard Robinett  
Professor of Physics  
Associate Head for Undergraduate and Graduate Studies

*Signatures are on file in the Graduate School.
Abstract

Since Galactic cosmic rays have been observed up to at least PeV energies, Galactic sources capable of accelerating particles to these energies should exist. These sources are called “PeVatrons”. Characteristic gamma-ray signatures of a PeVatron include a hadronic, hard spectrum that extends past 100 TeV without any signs of a cutoff or attenuation. Until recently, no gamma-ray experiment had the sensitivity to make detections at 100 TeV.

This dissertation outlines the first extremely high energy catalogs of the High Altitude Water Cherenkov (HAWC) Observatory. Two such catalogs are created: one of sources emitting above 56 TeV, and one of sources emitting above 100 TeV. HAWC consists of an array of 300 Water Cherenkov Detectors (WCDs) located at 4100 meters above sea level, near the peak of the extinct volcano Sierra Negra in the state of Puebla, Mexico. Each WCD is equipped with four photomultiplier tubes (PMTs) in a tank of water and can detect the Cherenkov light emitted when the charged particles from the extensive air shower created when a gamma-ray hits the Earth’s atmosphere make it to ground level. Due to the high duty cycle (nearly 24/7) and large field-of-view (∼2 sr) of the experiment, it is well-poised to perform all-sky surveys.

These catalogs are created using a new energy estimation technique that uses the PMT charge 40 meters from the air shower axis. This technique has been validated on the Crab Nebula and then applied to the Galactic plane. There are six sources in the Galactic plane that emit above 56 TeV. Three of them continue to emit past 100 TeV. These are the highest energy gamma-ray sources ever detected. While the gamma ray emission in some of these sources is likely purely leptonic in origin, for some of them a hadronic component cannot be ruled out and they may be considered PeVatron candidates.
# Table of Contents

List of Figures vii

List of Tables xvii

Acknowledgments xix

Chapter 1

**Introduction** 1

1.1 Production mechanisms of gamma rays 1

1.1.1 Leptonic mechanisms 1

1.1.2 Hadronic mechanisms 3

1.2 Sources of Galactic gamma rays 4

1.3 The highest-energy gamma rays 6

1.3.1 Applications to multi-messenger astronomy: searches for PeVatrons 7

1.3.2 Existing measurements 10

Chapter 2

**Very high-energy gamma-ray detection techniques** 12

2.1 Air shower physics 12

2.2 Cherenkov telescopes 16

2.3 Extensive air shower arrays 18

Chapter 3

**HAWC** 21

3.1 The HAWC detector 21

3.1.1 Electronics, data acquisition, and calibration 23

3.2 Air shower reconstruction 26

3.2.1 Core and angle fits 26

3.2.2 Event selection 30
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2.2.1 Fraction hit bins</td>
<td>30</td>
</tr>
<tr>
<td>3.2.2.2 Gamma/hadron separation</td>
<td>32</td>
</tr>
<tr>
<td>3.2.2.3 Angular resolution</td>
<td>35</td>
</tr>
<tr>
<td>3.2.2.4 Data quality cuts and dataset</td>
<td>35</td>
</tr>
<tr>
<td>3.3 Map-making, background estimation, and likelihood analysis</td>
<td>36</td>
</tr>
<tr>
<td>3.4 HAWC simulation</td>
<td>38</td>
</tr>
<tr>
<td>3.5 Sensitivity</td>
<td>38</td>
</tr>
<tr>
<td>3.6 Science goals</td>
<td>41</td>
</tr>
<tr>
<td>Chapter 4</td>
<td></td>
</tr>
<tr>
<td>Energy estimation method</td>
<td>43</td>
</tr>
<tr>
<td>4.1 Determination of the optimal radius</td>
<td>44</td>
</tr>
<tr>
<td>4.1.1 Cross-check with minimization of the function</td>
<td>47</td>
</tr>
<tr>
<td>4.1.2 Saturated tanks</td>
<td>49</td>
</tr>
<tr>
<td>4.2 Energy assignment method</td>
<td>50</td>
</tr>
<tr>
<td>4.3 Performance on simulation</td>
<td>53</td>
</tr>
<tr>
<td>4.3.1 Mixing matrix</td>
<td>53</td>
</tr>
<tr>
<td>4.3.2 Bias and resolution</td>
<td>54</td>
</tr>
<tr>
<td>4.3.3 Bin purity</td>
<td>56</td>
</tr>
<tr>
<td>4.4 Systematics in the energy assignment</td>
<td>58</td>
</tr>
<tr>
<td>4.4.1 Effect of using one optimal radius for all events</td>
<td>58</td>
</tr>
<tr>
<td>4.4.2 Laser effect</td>
<td>59</td>
</tr>
<tr>
<td>4.5 Improvements over $f_{\text{hit}}$</td>
<td>65</td>
</tr>
<tr>
<td>Chapter 5</td>
<td></td>
</tr>
<tr>
<td>Spectral fits with the energy estimator</td>
<td>68</td>
</tr>
<tr>
<td>5.1 A 2D binning scheme for events</td>
<td>68</td>
</tr>
<tr>
<td>5.2 Gamma/hadron separation and additional data quality cuts</td>
<td>70</td>
</tr>
<tr>
<td>5.3 Background estimation for low-statistics bins</td>
<td>71</td>
</tr>
<tr>
<td>5.4 Likelihood fit</td>
<td>74</td>
</tr>
<tr>
<td>5.4.1 Bins used in the fit</td>
<td>75</td>
</tr>
<tr>
<td>5.4.2 Flux points technique</td>
<td>79</td>
</tr>
<tr>
<td>5.4.2.1 Location of points on the x-axis</td>
<td>80</td>
</tr>
<tr>
<td>5.4.2.2 A note about extended sources</td>
<td>80</td>
</tr>
<tr>
<td>5.5 Spectral fit systematics</td>
<td>81</td>
</tr>
<tr>
<td>5.6 Application to the Crab Nebula</td>
<td>84</td>
</tr>
<tr>
<td>5.6.1 Systematics in the Crab spectral fit</td>
<td>88</td>
</tr>
<tr>
<td>5.6.2 Purity of the highest energy bins</td>
<td>90</td>
</tr>
<tr>
<td>5.6.3 Scientific significance</td>
<td>92</td>
</tr>
</tbody>
</table>
# List of Figures

1.1 All sources in the TeVCat as of August 2018. The horizontal line of sources is the Galactic plane. Image made using the source viewer at http://tevcat.uchicago.edu. .......................................................... 4

1.2 The cosmic ray spectrum. All cosmic rays up to at least the knee are expected to be of Galactic origin. Source: Particle Data Group [1]. 8

2.1 A simple schematic showing the basics of the Heitler model for an electromagnetic air shower. .......................................................... 13

2.2 An electromagnetic air shower initiated by a 1 TeV gamma ray (left) and a hadronic air shower initiated by a 10 TeV proton (right). Note the presence of non-electromagnetic components in the hadronic shower (the green and blue lines) (source: created by F. Schmidt and J. Knapp and retrieved from [2]). .......................................................... 14

2.3 A diagram of a hadronic shower initiated by a proton. Note the presence of neutrinos, muons, and pions, absent in the model of an air shower initiated by a photon (source: hawc-observatory.org). .. 16

2.4 The geometry involved in Cherenkov radiation. \( \theta \) is the Cherenkov angle. Created by Arpad Horvath and used under a Creative Commons BY-SA 2.5 license. ................................. 17

2.5 Aerial view of the Milagro detector. The red dots denote the location of outrigger tanks that were later added to help with reconstruction. (source: Milagro Collaboration, https://web.archive.org/web/20180402221713/umdgrb.umd.edu/cosmic/milagro.html). ........... 19

3.1 An image of the completed HAWC-300 array, taken just before the inauguration in early 2015 (source: Jordan Goodman). ............... 21

3.2 A 10-inch Hamamatsu R7081 PMT (left) and an 8-inch Hamamatsu R5912 PMT (right). .......................................................... 22
3.3 A schematic of an air shower particle hitting one of HAWC’s tanks and emitting Cherenkov light. Each tank has 4 PMTs; the one in the center is high-quantum efficiency. A person is shown for scale.  

3.4 A schematic of the HAWC array. Each of the blue circles is a tank; the gap in the center is due to a structure known as the counting house that holds all the electronics. The green circles denote the location of the VAMOS engineering test array, which operated in 2011 and 2012.  

3.5 A two edge event. The curve is the pulse from the PMT. Note that it does not cross the second, higher threshold. Image used with permission of Zigfried Hampel-Arias.  

3.6 A four edge event. Unlike the two-edge case, this PMT pulse crosses both thresholds. Image used with permission of Zigfried Hampel-Arias.  

3.7 The HAWC event display showing the distribution of charge for a gamma-like event at the location of the Crab Nebula, which is used as a standard candle in gamma-ray astronomy. 686 PMTs were hit during this event. This information is used in the core fit. The star denotes the location of the fitted core; the circle has a radius of 40 meters and is important for gamma/hadron separation, along with the PMT outlined in red, which is the PMT with the highest charge outside that circle (discussed in section 3.2.2).  

3.8 The same event as in Figure 3.7, but this time showing the timing information. This is information used to reconstruct the geometry of the shower.  

3.9 The distribution of true energies within each $f_{hit}$ bin, assuming a transiting source with a power law index of -2.63 and a declination of 20 degrees N. Note that $f_{hit}$ acts as a rudimentary energy estimator. An increase in bin is roughly correlated with an increase of energy, but there are large overlaps in some bins. Additionally, Bin 9 functions as an overflow bin that contains a large portion of HAWC’s energy range. This figure is made using HAWC Monte Carlo simulations.  

3.10 The inverse-compactness distribution for events near the Crab Nebula where more than 75% of the array is hit (the inverse is plotted because the gamma rays are easier to see on the plot). Gamma rays can be seen as the excess on the left-hand side of the plot.
3.11 The PINCness distribution for events near the Crab Nebula where more than 75% of the array is hit. Gamma rays from the Crab Nebula are immediately evident by noting the excess above the surrounding region at low values (the bump between PINCness values of 1.0 and 2.0). 

3.12 The HAWC point source skymap, created with 1017 days of data. A power-law index of 2.7 is assumed. This map includes data from all $f_{hit}$ bins. In Chapter 6 this map will be broken down into individual energy bins.

3.13 The differential sensitivity of HAWC for a point source with an $E^{-2.63}$ spectrum. The dark red line is the fit to the flux in each $f_{hit}$ bin required to detect the source at $5\sigma$ 50% of the time (the nine light red lines). Lines for 50 hour observations from various IACTs are shown for comparison. See the text of section 3.5 for more information.

4.1 A sample event showing how the optimal radius is computed. Each black point is $log_{10}(q_{eff})$, where $q_{eff}$ is the effective charge. The solid red line is the best fit to Equation 4.1, while the light red band denotes the ensemble of fits obtained by varying $s$ by $\pm 10\%$. Note that the band is smallest around $r=40m$ (the small red circle).

4.2 Distribution of the optimal radius for different event sizes. Plot (a) starts at 6.7% of the array hit as this is the threshold for the first analysis bin. The percentage of the array hit is defined by $nHitSP20/nChAvail$. Each histogram contains 5000 simulated events. No gamma/hadron separation cuts were applied, however a cut to remove events with reconstructed cores off the array was applied along with the standard data quality cuts and a requirement that the difference between the reconstructed and true zenith angle of the event is $< 0.5^\circ$.

4.3 The optimal radius as a function of zenith angle. Only events where more than 75% of the array is hit are used. The x-error bar shows the size of the zenith bin while the y-error bars denote one standard deviation.

4.4 Cross-check between minimizing the function, as is done in [3] and the method presented here. The mean is 5.168e-5 and the RMS is 9.194e-3. The histogram includes 5000 events of all event sizes.
4.5 An example showing the correlation between effective charge and \( E_{\text{true}} \) for a small range of zenith angles \( (14.77^o < \theta < 15.42^o) \). The black line is the line of the best fit. Note the change in slope around \( \log_{10}(s_{r_{40}}) = \sim 0.7 \). This change in slope is known as \( s_{c_{\text{crit}}} \). Colors in the histogram denote the relative probabilities assuming an isotropic source at a declination of 20°. This fit is repeated for every zenith bin so that the best fit parameters can be parameterized by zenith angle.

4.6 Examples for the parameterization by zenith angle for different fit parameters. All fit functions are chosen empirically. Figure 4.6a shows the parameterization of \( b_1 \), while Figure 4.6b shows the parameterization of \( m_2 \). Horizontal error bars are the RMS of the zenith distribution in each bin. Note that the line of best fit for \( m_2 \) has a sharp change in slope at \( \theta = 30^o \). It is not clear \( a \text{ priori} \) what causes this.

4.7 The mixing matrix, assuming isotropic gamma rays with an \( E^{-2} \) spectrum. The dotted line is the 1:1 line (i.e. events that lie on this line have energies equal to their true energy).

4.8 The same mixing matrix as shown in Figure 4.7, but with the Monte Carlo weighted to resemble a transiting source with an \( E^{-2} \) spectrum (a) and a transiting source with an \( E^{-3} \) spectrum (b). Both sources transit at a declination of 20 degrees.

4.9 The bias and resolution as a function of the true (MC) energy.

4.10 The bias and resolution as a function of the reconstructed energy.

4.11 The bin purity for two different simulated sources transiting at a declination of 20 degrees. The bin purity is worse for the \( E^{-3} \) source because the spectrum is softer meaning that there are far more lower-energy events with upward fluctuations than there are truly high-energy events.

4.12 The bin purity for an \( E^{-2} \) source transiting at a declination of 20 degrees, compared to an \( E^{-2} \) source transiting at a declination of -15 degrees.

4.13 The difference in the standard estimated energy, with the signal measured at 40 meters, and the estimated energy with the algorithm optimized on the signal 30 meters from the shower axis. The mean and RMS of the histogram are -0.04 and 0.09, respectively.

4.14 The difference in the standard estimated energy, with the signal measured at 40 meters, and the estimated energy with the algorithm optimized on the signal 65 meters from the shower axis. The mean and RMS of the histogram are -0.06 and 0.19, respectively.
4.15 A sample distribution showing the arrival times of PEs in a PMT. 0 is defined as the time the first PE reaches the PMT. This plot includes all PMTs with a distance from the shower axis between 0 and 40 meters, and all showers having $3 < \log_{10}(nPE) < 3.5$. $nPE$ is the number of PEs.

4.16 Schematic of the setup for the simulated air shower runs.

4.17 The average ToT-log nPE curves for all possible combinations of ToT and PMT size: high Tot/8 inch PMT (top left), high Tot/10 inch PMT (top right), low ToT/8 inch PMT (bottom left), and low ToT/10 inch PMT (bottom right). The red data points are the average points for PMTs receiving light from the delayed fibers during the modified calibration run. The blue data points correspond to the average curves from a standard calibration curve. The yellow lines are the fits to curves.

4.18 The ratio of the low ToT-log nPE curve (left) and the ratio of the high Tot-nPE curves (right). The ratio is calculated as the curve from the modified calibration run divided by the curve from a standard calibration run. The grey line corresponds to 8-inch PMTs while the black line corresponds to 10-inch PMTs.

4.19 The difference between the standard energy estimate and the energy estimate obtained when the signal is reduced by 5%.

4.20 The difference between the standard energy estimate and the energy estimate obtained when the signal is increased by 5%.

4.21 The distribution of true MC energies in each reconstructed energy bin. This plot should be compared to the identical plot for the $f_{hit}$ analysis (Figure 3.9). The plot is made assuming a transiting source at a declination of 20 degrees N and with a power law index of -2.63. Plot made by John Pretz.

4.22 The differential sensitivity of HAWC for the energy analysis. The lines are the sensitivities in each energy bin. A point source with an $E^{-2.63}$ is assumed. Plot made by John Pretz.
5.1 The 68% containment of events as a function of $f_{hit}$ bin and energy, as measured on the Crab Nebula. The number refers to the $f_{hit}$ bin (defined in Table 3.1) and the letter refers to the energy bin (defined in Table 5.1). To ensure a good fit, only bins where the Crab Nebula is $>3\sigma$ in data are shown. As $f_{hit}$ increases across an energy bin, the angular resolution noticeably improves because the shower size increases. The upturn in the angular resolution above 56 TeV is believed to be because the shower curvature model used during event reconstruction has not yet been updated for the highest energies.

5.2 The local coordinates ($\theta$ and $\phi$) for all events passing gamma/hadron cuts with an energy between 100 TeV and 177 TeV. There are 87,586 such events. A random $\theta/\phi$ pair is pulled from this distribution and used to calculate the background, as described in the text. Note that the two local coordinates are correlated, making the 2D histogram a necessity.

5.3 The background for all events with energies between 100 TeV and 177 TeV, using the direct integration method. Note that the background is not very smooth. The map is in equatorial coordinates.

5.4 The background for all events with energies between 100 TeV and 177 TeV, using the new method developed for low statistics bins. The scale is the same as Figure 5.3. Note that the background is now smooth. The map is in equatorial coordinates.

5.5 Histogram of the right ascension of all events with energies between 100 and 177 TeV and declinations between 15 and 25 degrees. Note that due to the low statistics, the data is prone to fluctuations, and the direct integration background follows those fluctuations. The randomized background is much smoother, but still follows general trends in the data.

5.6 The zenith angle distribution for events between 100 TeV and 177 TeV, for different time periods. The dataset used in this dissertation has been divided into three epochs with roughly equal livetime. The histograms have been scaled so that the integrals of all three histograms are the same. The zenith distribution is stable over time.

5.7 A depiction of how the bins are chosen. This is the simulated estimated energy distribution for all events in $f_{hit}$ bin 6, assuming a source transiting at a declination of -15$^o$ with a power law index of 2.0 and no exponential cutoff. The gray shaded band denotes the middle 99% of the events. This corresponds to 2D bins 6f, 6g, 6h, and 6i (energies between 5.62 TeV and 56.2 TeV).
5.8 The important bins at a declination of 20 degrees (roughly overhead for HAWC). The x-axis is the estimated energy and the y-axis is the $f_{hit}$ bin. Bins shaded in black are important regardless of what transiting source hypothesis is assumed. Bins shaded in yellow are only important in certain scenarios. For example, gamma rays in bin 9l ($> 177$ TeV) are only expected for a very hard source with no cutoff. The union of the black and yellow bins is used in the likelihood spectral fit.

5.9 The 68% containment of events on the Crab Nebula, compared to the value from simulation. Only bins where the Crab Nebula is detected $> 3\sigma$ in data are shown. Note the systematic discrepancy.

5.10 The Crab Nebula as a function of energy. To reduce the number of images, half-decade bins are used instead of quarter decade bins.

5.11 The Crab excess per transit, along with the residual, which is defined as (measured-expected)/expected. The expected counts are computed from Monte Carlo weighted to resemble the overall fitted spectrum.

5.12 The spectral energy distribution of the Crab Nebula obtained using the ground parameter energy estimate and the flux points method outlined above. The dotted black line is the forward-folded best fit. For comparison, HAWC’s previously published Crab spectrum, a forward-folded fit using the $f_{hit}$ based analysis, is also shown. The red band denotes statistical uncertainties on the previous HAWC fit, while the dark grey band denotes statistical uncertainties on the fit documented in this Chapter.

5.13 The size of the uncertainty in $E^2dN/dE$ as a function of energy for each source of systematic uncertainty.

5.14 The amount each flux point shifts on the x-axis for each source of systematic uncertainty. This tells us about the overall energy scale of the experiment.

5.15 Identical to Figure 5.12, but with a systematic uncertainty band added. This band contains all fits obtained varying assumptions about the modeling of the detector to investigate the effects discussed in Section 5.5, added in quadrature. For comparison, the previously published $f_{hit}$ based Crab spectrum [4] with its systematic uncertainties is also shown. Improved modeling has greatly decreased the size of the systematic uncertainty band.
6.1 The HAWC sky above 56 TeV in reconstructed energy. This plot assumes a source spectrum with a power law index of 2.0 and a source morphology of a disk with a radius of 0.5 degrees. Sources in the Galactic plane are clearly visible. The map is in equatorial coordinates.

6.2 The Crab Nebula above 56 TeV in reconstructed energy. The map assumes a point source morphology. The open circle denotes the location of the Crab in the 2HWC all-energy catalog paper, while the black circle outlined in white is the local maximum of the > 56 TeV test statistic.

6.3 The Crab Nebula above 100 TeV in reconstructed energy. The map assumes a point source morphology. The open circle denotes the location of the Crab in the 2HWC all-energy catalog paper. The Crab is not detected at 5 sigma above 100 TeV. The maximum significance is 3.98$\sigma$ at (83.45, 21.94) degrees.

6.4 The Galactic plane above 56 TeV in reconstructed energy. The map assumes a point source morphology. The open circles denote the locations of sources in the 2HWC all-energy catalog paper, while the black circles outlined in white are the local maxima of the > 56 TeV test statistic. The middle portion of the plane (44 < $l$ < 68 is omitted as there are no high energy sources present there.

6.5 The Galactic plane above 100 TeV in reconstructed energy. The map assumes a point source morphology. The open circles denote the locations of sources in the 2HWC all-energy catalog paper, while the black circles outlined in white are the local maxima of the > 100 TeV test statistic. The middle portion of the plane (44 < $l$ < 68 is omitted as there are no high energy sources present there.

6.6 The Galactic plane above 177 TeV in reconstructed energy. The map assumes a point source morphology. The open circles denote the locations of sources in the 2HWC all-energy catalog paper. No high energy sources are significantly detected.

6.7 The spectrum of 2HWC J1825-134. The grey and light pink bands are the statistical and systematic uncertainties, respectively.

6.8 The spectrum of 2HWC J1839-057. The grey and light pink bands are the statistical and systematic uncertainties, respectively.

6.9 The spectrum of 2HWC J1844-032. The grey and light pink bands are the statistical and systematic uncertainties, respectively.

6.10 The spectrum of 2HWC J1908+063. The grey and light pink bands are the statistical and systematic uncertainties, respectively.
6.11 The spectrum of 2HWC J2019+367. The grey and light pink bands are the statistical and systematic uncertainties, respectively. 108
6.12 The spectrum of 2HWC J2031+415. The grey and light pink bands are the statistical and systematic uncertainties, respectively. 108
6.13 Results for 2HWC J1825-134. These plots show the emission at all energies (1 TeV < $E_{\text{reco}}$ < 316 TeV). 109
6.14 Results for 2HWC J1825-134. These plots concentrate on the high energy emission (56 TeV < $E_{\text{reco}}$ < 100 TeV). 110
6.15 Results for J1839-057. These plots show the emission at all energies (1 TeV < $E_{\text{reco}}$ < 316 TeV). 111
6.16 Results for J1839-057. These plots concentrate on the high energy emission (56 TeV < $E_{\text{reco}}$ < 100 TeV). 112
6.17 Results for 2HWC J1844-032. These plots show the emission at all energies (1 TeV < $E_{\text{reco}}$ < 316 TeV). 113
6.18 Results for 2HWC J1844-032. These plots concentrate on the high energy emission (56 TeV < $E_{\text{reco}}$ < 100 TeV). 114
6.19 Results for 2HWC J1908+063. These plots show the emission at all energies (1 TeV < $E_{\text{reco}}$ < 316 TeV). 115
6.20 Results for 2HWC J1908+063. These plots concentrate on the high energy emission (56 TeV < $E_{\text{reco}}$ < 100 TeV). 116
6.21 Results for 2HWC J2019+367. These plots show the emission at all energies (1 TeV < $E_{\text{reco}}$ < 316 TeV). 117
6.22 Results for 2HWC J2019+367. These plots concentrate on the high energy emission (56 TeV < $E_{\text{reco}}$ < 100 TeV). 118
6.23 Results for 2HWC J2031+415. These plots show the emission at all energies (1 TeV < $E_{\text{reco}}$ < 316 TeV). 119
6.24 Results for 2HWC J2031+415. These plots concentrate on the high energy emission (56 TeV < $E_{\text{reco}}$ < 100 TeV). 120
6.25 The extension of 2HWC J1825-134 as a function of energy. Error bars are statistical only. To ensure enough statistics, only energy bins where the source is > 3$\sigma$ are plotted. 121
6.26 The same as Figure 6.25, but for J1839-057. 121
6.27 The same as Figure 6.25, but for 2HWC J1844-032. 122
6.28 The same as Figure 6.25, but for 2HWC J1908+063. 122
6.29 The same as Figure 6.25, but for 2HWC J2019+367. 122
6.30 The same as Figure 6.25, but for 2HWC J2031+415. 123
6.31 Diffusion fits to each of the highest energy sources. Errors are statistical only. 129
7.1 Drone footage of the HAWC array. The outriggers can be seen as white dots outside the main array. Picture taken by Alberto Carramiñana. 135

7.2 The projected sensitivity incorporating the outriggers and the energy analysis. Plot courtesy of John Pretz. 136
### List of Tables

3.1 The definitions for each $f_{hit}$ bin, where $f_{hit}$ is the number of PMTs hit that are within 20 ns of the shower plane during the air shower event divided by the total number of available PMTs after calibration. 30

4.1 $R_{opt}$ for different shower sizes. The optimal radius slowly increases as the showers get bigger. 46

5.1 The binning scheme used for energy fits. 69

5.2 Bins to use by declination. 78

5.3 The best fit parameters for the Crab Nebula. The fit spans from 1 TeV to 316 TeV in reconstructed energy. 84

6.1 Result of the catalog source search above 56 TeV in reconstructed energy. The names of the sources from the 2HWC catalog and the 2HWC locations are given along with the locations of the peaks of the high energy emission. “Distance” is the distance between the 2HWC catalog location and the high energy catalog location. “Search” denotes which search the source was found in (point source, 0.5 degree extended, or 1.0 degree extended). Some strong sources may be found in multiple searches. “TS” is the test statistic of the high energy emission, assuming a power law with a spectral index of 2.0. 97

6.2 The same as Figure 6.1, but for the catalog source search above 100 TeV in reconstructed energy. 98

6.3 The best fits to Equation 6.2 for each source that emits with $E_{reco} > 56$ TeV. Errors are statistical only. Extension is the Gaussian sigma from the fit. The units of $\phi_o$ are cm$^{-2}$ TeV$^{-1}$ s$^{-1}$. “PL diff” shows how much the power law with a cutoff is preferred over a pure power-law with no exponential cutoff. 104
6.4 The systematic uncertainties on the index and cutoff for each high energy source. ......................... 104
6.5 The bin purity for $E_{\text{reco}} > 56$ TeV events. In general, anywhere from 1/2 to 2/3 of the events with a reconstructed energy above 56 TeV have true energies above that value according to the Monte Carlo simulation. The remainder of the events are mis-reconstructed lower energy events, which are much more numerous due to the properties of astrophysical spectra. Uncertainties quoted here are systematic. 123
6.6 The bin purity for $E_{\text{reco}} > 100$ TeV events. Uncertainties quoted here are systematic. 123
6.7 Each 2HWC source emitting above 56 TeV in reconstructed energy. The nearest non-HAWC TeVCat source and the distance to it is listed. “Source type” is the source identification from the TeVCat catalog: “PWN” is “pulsar wind nebula”; “UNID” is “unidentified”. 124
6.8 Diffusion constants for the three sources emitting above 100 TeV in reconstructed energy. The diffusion constant is valid for 300 TeV electrons creating 100 TeV gamma rays only. The uncertainty on the diffusion constant is statistical only and comes from the statistical uncertainty on the fitted diffusion radius value, propagated through to the final result. 132
6.9 Diffusion constants for the three sources that emit above 56 TeV in reconstructed energy but not above 100 TeV. The diffusion constant is valid for 200 TeV electrons creating 50 TeV gamma rays only. The uncertainty on the diffusion constant is statistical only and comes from the statistical uncertainty on the fitted diffusion radius value, propagated through to the final result. 132
Acknowledgments

This dissertation never would have been completed without the help of several people. First of all, I would like to thank my advisor, Miguel Mostafà. Thanks for guiding me through how to become a scientist and for creating an environment where it was possible for me to ask all of my dumb questions.

I would also like to thank John Pretz, who finally convinced me to learn Python after months of badgering.

The group at Los Alamos National Lab also deserves a lot of credit. Thanks for hosting me for a six month stay! In particular, Brenda Dingus, Pat Harding, and Andrea Albert provided helpful insights and encouragement. The idea to fit several sources to a diffusion model (Chapter 6) arose during a visit to Los Alamos.

My thesis committee (Stephane Coutu, Doug Cowen, and Peter Mészáros) asked many helpful questions. I would also like to thank Stephane for going through the entire thesis with a fine-tooth comb and finding every single typo and problem with subject/verb agreement.

I would like to thank the friends I met during graduate school. At PSU, Susan and Jayson Kempinger as well as Cody Messick and Kasey Cannon provided many happy memories.

Thank you to the HAWC graduate students, in particular Mehr Un Nisa, Chad Brisbois, Hugo Ayala, Sam Marinelli, Kristi Engel, and Chang Rho, for providing a constant stream of gifs on Slack, for answering my dumb coding questions, and providing company when traveling. Airport layovers, collaboration meetings, and conferences would have been a lot more boring without you weirdos around!

And of course, thanks to my family as well, even if you still have no clue what a gamma ray is. Special thanks to Allison, who decided to willingly put up with someone who would occasionally ignore everything except for astrophysics.

This material is based upon work supported by the National Science Foundation under Grants PHY-1506145 and PHY-1806854. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. This
material is also based upon work supported by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists, Office of Science Graduate Student Research (SCGSR) program. The SCGSR program is administered by the Oak Ridge Institute for Science and Education for the DOE under contract number DE-SC0014664
Dedication

This dissertation is dedicated to my father, Philip J. Malone.
Chapter 1  |  Introduction

1.1 Production mechanisms of gamma rays

The detection of the Crab Nebula in TeV gamma rays by the Whipple Observatory in 1989 [5] ushered in the dawn of a new era in particle astrophysics. Since that discovery, numerous experiments have surveyed our Galaxy looking for other sources of gamma rays (see Chapter 2 for an overview of the techniques used by various experiments).

These gamma rays, which sit at the highest-frequency end of the electromagnetic spectrum and have energies above 1 MeV or so, can be produced in a variety of ways. There are two main classes of gamma-ray production: leptonic, where the gamma rays come from electrons, and hadronic, where the gamma rays come from particle decay. In the discussion below, the emphasis will be placed on mechanisms that create TeV gamma rays, although lower energies will also be briefly discussed.

1.1.1 Leptonic mechanisms

Synchrotron emission occurs when electrons are accelerated relativistically around magnetic field lines, traveling in a curved orbit. Electromagnetic radiation is emitted during this process, as would be expected from classical electrodynamics. The radiated power is described by the relativistic Larmor formula [6]:

$$ P = -\frac{dE}{dt} = \frac{2e^2a^2}{3c^3} $$

(1.1)
where $a$ is the centripetal acceleration of the particle and $E$ is the energy, $c$ is the speed of light, and $e$ is the charge of the electron. By making the proper substitutions into the above equation for acceleration in a circular orbit ($a = \Gamma^2 \frac{v^2}{R}$) and using the definition of the Larmor radius, the above equation can be rearranged into a form where it can be seen that the energy loss rate of the electron depends on the square of the energy of the electron as well as the square of the magnetic field. From [6], we see that in the Thompson (low-energy) regime:

$$-\frac{dE}{dt} = \frac{4}{3} \sigma_T c U_B \Gamma^2 \tag{1.2}$$

where $\sigma_T$ is the Thompson cross section, $U_B$ is the magnetic field energy density ($\frac{B^2}{8\pi}$), and $\Gamma = \frac{E}{m_e c^2}$.

Synchrotron radiation in astrophysical sources follows power-law energy spectra; the gamma ray spectral index is equal to $(\alpha_e - 1)/2$, where $\alpha_e$ is the spectral index of the parent electron population [6]. If the emission region is dense, this synchrotron emission may be partially absorbed inside the source in a process called synchrotron self-absorption; this modifies the observed spectrum [7]. Synchrotron radiation peaks well below gamma rays on the electromagnetic spectrum, mainly creating infrared and x-ray photons. However, the synchrotron component is connected to the TeV-emitting inverse Compton component, described below.

The same electron population that produces the synchrotron emission can scatter off of ambient photons in the vicinity of the source. The photon receives some of the electron’s energy during this process, boosting its energy. This process is known as “inverse Compton scattering”. Since this ambient photon may be the synchrotron photon discussed above, these two processes together are sometimes referred to as “synchrotron self-Compton”. This leads to a spectral energy distribution that has two peaks: one at lower energies ($\sim$1 keV) for the synchrotron component, and another in the GeV-TeV range for the inverse Compton component. The exact location of the peak is determined by a wide variety of factors, such as the magnetic field.

Like synchrotron radiation, inverse Compton scattering in the Thompson regime results in a gamma ray spectral energy distribution that has an index equal to $(\alpha_e - 1)/2$, where $\alpha_e$ is once again the spectral index of the parent electron population [6].
The relationship between the synchrotron and inverse Compton components is

\[
\frac{P_{\text{synch}}}{P_{\text{IC}}} = \frac{U_B}{U_{\text{ph}}}. \tag{1.3}
\]

As described in [7], \(P_{\text{synch}}\) is the power in the synchrotron radiation, \(P_{\text{IC}}\) is the power in the inverse Compton component, \(U_B\) is once again the energy density contained in the magnetic field, and \(U_{\text{ph}}\) is the energy density of the photons. Note that this relation does not necessarily hold in the Klein-Nishina regime (discussed in Section 1.3) or if either energy density is very large or very small [8].

A third type of leptonic emission, bremsstrahlung, primarily occurs at lower energies. German for “braking radiation”, it occurs when a charged particle is deflected by a second charged particle, resulting in a loss of kinetic energy. A photon must be emitted to preserve conservation of energy. Bremsstrahlung will not be discussed further in this dissertation.

### 1.1.2 Hadronic mechanisms

The primary decay mode of neutral pions is

\[
\pi^0 \rightarrow 2\gamma. \tag{1.4}
\]

Astrophysically, neutral pions are created when cosmic rays interact with the interstellar medium, either with another cosmic ray or with an ambient photon. Therefore, gamma rays can act as a tracer of the cosmic ray distribution in our galaxy. Note that charged pions, which decay to neutrinos, are also created in similar processes. Neutrinos can also act as a tracer of cosmic rays, although they are significantly harder to detect. This is discussed in much more detail in Section 1.3.1.

Regardless of the actual energy distribution of the neutral pions or their parent protons, each photon created has an energy of 67.5 MeV in the rest frame of the neutral pion, which is half of the pion rest energy \(m_\pi c^2\) [9]. However, the gamma ray spectrum translated into the laboratory frame can extend orders of magnitude above this, and TeV gamma rays from neutral pion decay are possible.
Following the derivation in [6], we see that the energy in the laboratory frame is

\[ E_\gamma = \Gamma E_{\gamma}^* + \beta \Gamma p_{\gamma}^* \cdot \cos(\theta)^* \]  

where the * denotes quantities in the center of mass frame, \( \Gamma = \frac{E_\pi}{m_\pi} \) and \( \beta = \frac{|p_\pi^*|}{E_\pi} \).

By differentiating with respect to \( \cos(\theta)^* \), one can show that the energy spectrum, \( E_\gamma^2 \frac{dN}{dE_\gamma} \), is

\[ E_\gamma^2 \frac{dN}{dE_\gamma} = \frac{1}{2} \frac{1}{\beta \Gamma p_{\gamma}^*} \]  

This causes what is known as the “pion decay bump”, with the emission rising steeply in the MeV range and then tracing the distribution of the parent protons above that [6].

1.2 Sources of Galactic gamma rays

As of this writing, TeVCat [10], which tracks very high energy (E > 50 GeV) gamma ray sources, has 152 entries (Figure 1.1). A large portion of these sources lie along the Galactic plane, identified in various surveys of subsets of the region [11–13]. The rest of this dissertation will focus solely on Galactic gamma rays.

A wide variety of source classes can be found emitting gamma rays in our...
Galaxy. This includes supernova remnants, pulsar wind nebula/TeV halos, binaries, molecular clouds, and the Galactic center. These sources can vary largely in size, from being seen as point-like to covering a few degrees of the sky.

Supernova remnants (SNRs) are the remains of stellar explosions. SNRs are characterized by a rapidly expanding shock wave with particles accelerated via diffusive shock acceleration (sometimes called first-order Fermi acceleration) [14]. This acceleration occurs when a particle moves between two moving clouds of stellar material that have a shock front in between them. As the particle bounces back and forth between the two clouds, its energy increases with each successive encounter. Fermi acceleration results in a power-law energy spectrum of particles.

The gamma rays in a SNR are hadronic in origin. SNRs are thought to be the main source of cosmic rays in our galaxy; this is easily achievable if SNRs release about 10% of their energy in the form of cosmic rays [15].

Pulsars are rapidly rotating neutron stars that are created after a massive star undergoes a supernova explosion. From our vantage point on Earth, their emission appears pulsed, since it is only visible when the beam of emission is pointing toward us. Pulsars can have a period ranging from seconds all the way down to milliseconds. Pulsars are magnetized; this induces a relativistic electron-positron pulsar wind known as a “pulsar wind nebula” [15]. Electrons are accelerated during interactions with the interstellar medium (ISM). In this case, the gamma-ray emission is leptonic (synchrotron radiation dominates at low energies while inverse Compton scattering is responsible for the higher energy TeV emission).

Some in the gamma-ray community have begun to object to the phrase “pulsar wind nebula” (PWN) to describe TeV emission around a pulsar. Citing the fact that this term was originally borrowed from X-ray astronomers but the TeV emission is significantly more extended than the X-ray emission, they prefer the term “TeV halo” [16].

Binaries consist of a compact Galactic object orbiting a companion star. Although relatively common in X-rays, only a handful of gamma-ray emitting binaries have been observed. One such example is LS 5039 [17].

Molecular clouds consist of dense patches of material that serve as a target for particles to interact with and create gamma rays. For example, it has been suggested that cosmic ray sources could be identified when particles that escape a SNR shell travel to a nearby molecular cloud and interact there, illuminating it in
gamma rays [18].

Additionally, there are large-scale structures that also emit gamma rays. Diffuse emission permeates our Galaxy and can be hadronic or leptonic in origin (for example, neutral pions created in interactions with the interstellar medium can decay to gamma rays, or electrons can create gamma rays when they inverse Compton scatter off of the interstellar radiation field. Thus, diffuse emission is an important tool for setting constraints on Galactic cosmic ray acceleration and propagation.

The Fermi Bubbles, discovered using data from the Fermi-LAT telescope [19], are the largest-scale structures in the Galaxy, extending approximately 50 degrees above and below the Galactic plane. The origin is currently unknown.

A fairly large percentage of gamma-ray emitters are unidentified in nature. Conclusively determining the type of source an object is often relies on a multi-wavelength analysis. Occasionally, data are not available at other wavelengths, or the area has been observed and no source has been identified. Additionally, the telescope energy and angular resolution are not always sensitive enough to disentangle different possible gamma-ray emission mechanisms. Identifying the production mechanisms of these gamma-ray emitters is a major goal of the field.

1.3 The highest-energy gamma rays

The study of the highest-energy gamma rays, defined here as those above 50 TeV, is inherently interesting. Gamma rays observed at these energies must be relatively close to us (< 100 Mpc), as high-energy gamma rays cannot propagate over large distances due to attenuations by interactions with both the cosmic microwave and infrared backgrounds. This restricts high-energy gamma rays to our Galaxy, the local group, and nearby clusters, and excludes the possibility that any high-energy gamma rays could come from active galactic nuclei or have a cosmological origin [20].

At high energies, the leptonic inverse Compton component is suppressed due to the Klein-Nishina effect. The Klein-Nishina cross-section is smaller than the Thompson cross-section, so scattering is less efficient. As the energy increases, the electron spectrum becomes harder, which also changes the gamma-ray spectrum. The gamma-ray spectral index, which is \((\alpha_e - 1)/2\) in the Thompson regime, changes to \(\alpha_e\) in the ultra-relativistic Klein-Nishina regime [6]. The transition begins around
\( E_\gamma \approx 10 \text{ TeV} \) [15]. If the energy density is small enough compared to the magnetic field, there is expected to be a sharp break in the inverse Compton spectrum [8]. If the gamma-ray energy is high enough, the inverse Compton component is expected to be completely suppressed, leaving a dominantly hadronic component.

An unambiguously hadronic high-energy detection of a gamma-ray source has implications for the sources of Galactic cosmic rays (see Section 1.3.1 for an overview of this possibility.)

High-energy gamma-ray sources are the highest-energy sources of any type of particle ever detected. While higher-energy particles have been observed, they both seem to be predominantly extragalactic and their actual sources have not yet been identified. For example, the Pierre Auger Observatory has observed cosmic rays with energies \( > 10^{20} \text{ eV} \) [21], and the IceCube Neutrino Observatory has seen neutrinos with PeV energies [22].

1.3.1 Applications to multi-messenger astronomy: searches for PeVatrons

Despite Herculean efforts by thousands of scientists over the last century, the sources of cosmic rays are still unknown. These particles, which mainly consist of high-energy protons but can also contain atomic nuclei, are charged. They bend in magnetic fields on their way to the Earth, and by the time they are detected no longer point directly back to their sources. This is not to say that their distribution is completely uniform over the sky. At low energies, a small-scale anisotropy is observed [23], while at higher energies, a large-scale anisotropy can be seen [24]. The sizes of these anisotropies are \( 10^{-4} \) and 6%, respectively.

Studying the sources and acceleration mechanisms of cosmic rays is inherently interesting due to the sheer amount of energy cosmic rays transport across galaxies, making them very closely connected to the evolution of stars and galaxies.

One can immediately see how it would be advantageous to use gamma rays to search for the sources of cosmic rays. Recall from Section 1.1.2 that the main hadronic mechanism of gamma ray production occurs when neutral pions, which have been created in cosmic ray interactions, decay to gamma rays. Since these gamma rays are neutral, they point back to their source and therefore can in principle be used to identify cosmic ray sources. This assumes that it is possible
Figure 1.2: The cosmic ray spectrum. All cosmic rays up to at least the knee are expected to be of Galactic origin. Source: Particle Data Group [1].

to unambiguously identify the gamma-ray emission as hadronic, a challenge given that there are many competing leptonic components. This is where the > 50 TeV measurements are critical, as the leptonic emission is suppressed.

The cosmic ray spectrum (Figure 1.2) decreases over several orders of magnitude, with only a few well-known bumpy features where the slope of the spectrum changes. At lower energies, the cosmic rays are believed to be of Galactic origin. The “knee” occurs at $\sim 3$ PeV and is believed to mark a transition region where the production and acceleration of Galactic cosmic rays become less efficient. There is a second feature at $\sim 3$ EeV called the “ankle”. The cosmic rays above this point are believed to be dominantly extragalactic [7].

Since cosmic rays of Galactic origin are expected to exist up to at least the PeV range, some Galactic sources that accelerate to these energies should exist. Searches for these sources, called “PeVatrons” are currently an active area of study in gamma-ray astrophysics.
A PeV cosmic ray creates a gamma ray that is approximately an order of magnitude less energetic. The gamma-ray signature of a PeVatron would be a hard power-law spectrum ($\sim E^{-2}$) that extends to approximately 100 TeV without any sort of attenuation, spectral break, or exponential cutoff [25]. The choice of spectral index is motivated by the spectrum at the source from Fermi acceleration.

Note that pair production off the interstellar radiation field and the cosmic microwave background may lead to a steepening of the high-energy tail of the spectra. Studies of interstellar radiation field models near the Galactic center show that the observed spectral cutoff and the intrinsic cutoff could be different by a factor of two (i.e. the observed spectral cutoff underestimates the actual cutoff at the source) [26].

Exactly what source classes can accelerate Galactic cosmic rays to PeV energies is still up for debate. SNRs are an attractive candidate, as the energetics work out perfectly: they could easily provide the whole Galactic cosmic ray contribution provided that one SNR occurs approximately every 100 years and approximately 10% of the energy goes into gamma-ray production, both reasonable assumptions [6].

However, the SNR hypothesis is not without some major problems. In order to get the particles to high enough energy, some sort of magnetic field amplification at the SNR shock is required [27]. Additionally, SNRs are only expected to act as PeVatrons for a short time period (less than 1000 years [18]), meaning that the probability of observing one during a time period it is acting as a PeVatron is statistically low.

Gabici and Aharonian [18] note that “delayed” gamma-ray emission from a PeVatron could be used to indirectly detect these sources. If PeV particles escape the shell of an SNR and travel through the interstellar medium to a nearby molecular cloud, it could act as a target for gamma-ray production. The resulting gamma rays would be lower in flux than the original SNR, but still within the realm of what presently operating experiments can detect. This emission would also last for at least an order of magnitude more time, making it easier to detect. In order for this to be conceivable, the gas cloud must be within < 100 pc of the SNR.

Assuming that SNRs are PeVatrons, currently operating experiments should be able to probe some fraction ($\sim 25\%$) of the Galactic population. However, due to low photon statistics, they may not be able to reconstruct the parent proton spectrum up to 1 PeV, required for definitively identifying them as such [28].
Since no SNR has been identified as a PeVatron, some alternative theories have been developed. One such idea is that galactic star clusters may operate as PeVatrons. The acceleration may take place near the stars or in the superbubbles themselves [29]. Superbubbles are cavities inflated in the ISM around stellar clusters as the result of stellar explosions and winds. Examples of clusters that are PeVatron candidates include the Westerlund 1 and Westerlund 2 complexes as well as the Cygnus OB2 area.

To date, only one PeVatron candidate has been observed: in 2016, the HESS Collaboration announced that the Galactic center region is acting as a PeVatron, providing a possible alternative to SNR remnants [30]. There is ample molecular gas in this region that could be illuminated by a PeVatron, but the black hole at the Galactic center is not currently active enough to provide a large contribution to the Galactic cosmic rays. While it may have been more active in the past, more present-day PeVatrons are expected to exist.

Tangentially, some fraction of the IceCube astrophysical neutrinos may also be associated with PeVatrons. Most of these neutrinos are expected to be extragalactic in origin, but any Galactic portion could originate from PeVatrons as cosmic-ray interactions also create charged pions that decay to neutrinos. A non-observation of certain Galactic plane sources with IceCube would rule them out as cosmic-ray accelerators [31].

### 1.3.2 Existing measurements

Up to this point, measurements above 50 TeV have been sparse and extremely incomplete due to the sharply decreasing sensitivity of most gamma-ray experiments in this energy regime.

The High Energy Gamma Ray Astronomy (HEGRA) experiment observed the Crab Nebula up to 80 TeV in 2004 [32]. However the last data point in the spectrum, which was centered at 74 TeV, only had a significance of 2.7σ.

The High Energy Stereoscopic System (HESS) array has recently determined the spectrum of one PWN, HESS J1825-137, up to 70 TeV [33].

Above ~ 80 TeV, many experiments have searched for high-energy gamma rays over the course of the last two decades with only null results. See, for example, upper limits on emission from the Crab Nebula > 100 TeV from both CASA-MIA [34]
and the Tibet Air Shower Array [35],

The completion of the High Altitude Water Cherenkov Observatory (HAWC) has allowed a much more systematic survey of > 50 TeV gamma rays. HAWC is the most sensitive currently-operating gamma-ray observatory in the world above 10 TeV. The HAWC detector is discussed in Chapter 3, and the results of the first > 50 TeV and > 100 TeV sky surveys are discussed in Chapter 6.
Chapter 2 | Very high-energy gamma-ray detection techniques

There are many experiments, both space-based and ground-based, designed to observe gamma rays. Space-based observatories have many advantages: for example, they have a wide field-of-view and a duty cycle of nearly 100%. This makes them well-poised to perform surveys of the entire sky or to detect transients such as gamma-ray bursts. However, they are typically designed to work in the MeV-GeV energy range. Since they are satellites, the space-based observatories have small effective areas which prevent them from being sensitive to the relatively rare TeV gamma rays. This chapter will focus on ground-based experiments, which have larger collection areas and are optimized to detect gamma rays in the GeV and TeV regimes.

2.1 Air shower physics

Gamma rays cannot be directly detected from the Earth’s surface; the atmosphere is opaque to them. Instead, they interact with molecules in the atmosphere to create what is known as an extensive air shower (EAS).

On the surface, a gamma-ray initiated shower is fairly simple. In a model originally formulated by Heitler and documented extensively in [36], a primary gamma ray with energy $E_\gamma$ interacts with the atmosphere to create an electron-positron pair. These particles will travel some distance and then in turn go through the bremsstrahlung process, emitting another gamma ray (see Figure 2.1). The distance traveled in each step is approximately one “radiation length” (the distance
Figure 2.1: A simple schematic showing the basics of the Heitler model for an electromagnetic air shower.

an electron can travel before losing all but $1/e$ of its energy) [37]. One radiation length in air is roughly $37 \text{ g cm}^{-2}$ [6].

Due to conservation of energy, the resulting gamma ray will have less energy than the primary gamma ray that initiated the air shower. However, it is still capable of producing another electron-positron pair. Thus, the cycle continues, with each step creating exponentially more particles, until the ionization energy losses dominate over radiative losses (bremsstrahlung) and the shower begins to die out. This point is known as the shower maximum; the energy at which this occurs is the critical energy, $E_c$ ($\sim$80-85 MeV in air). Up until the shower reaches the maximum, it will have $2^n$ particles, where $n$ is the number of splittings that the particles have gone through.

Heitler’s model is not perfect, mainly because bremsstrahlung emission can result in multiple photons being created, but otherwise describes actual air showers remarkably well - for example, it correctly predicts that the shower maximum will be related to the primary energy logarithmically [36]. At the shower maximum, there are $2^n \approx E_o/E_c$ particles present [6].

For a 1 TeV gamma ray, the shower maximum will be $\sim$10 km above sea level [37]. See the left panel of Figure 2.2 for an illustration of a gamma-ray initiated air shower. Because the shower tends to die out well before reaching the ground, experiments that detect the particles from the air shower directly tend to be located at high altitude since they are closer to this shower maximum.

The age of the shower, $s$, parameterizes the longitudinal shower development
and is given by

\[ s = \frac{3}{1 + 2y/t} \]  \hspace{1cm} (2.1)

where \( y = \log\left(\frac{E_\gamma}{E_c}\right) \) and \( t \) is a dimensionless variable: the pathlength divided by the radiation length. Therefore, \( s = 1 \) at shower maximum [37].

As the shower develops, the particles will have some lateral spread. This is partially due to the opening angles of the pair production and bremmstrahlung, but largely due to multiple Coulomb scattering [7]. In order to determine the lateral distribution of electrons in an air shower, a system of cascade equations must be solved that incorporates the energy, the location, the zenith angle, and the time. In the 1950s, Greisen, Kamata, and Nishimura obtained approximate solutions to this formula, which is now known as the NKG function. One form of it can be derived from [7]:

\[ f(x) \propto x^{s-2} (1 + x)^{s-4.5}. \]  \hspace{1cm} (2.2)

Here \( x \) is equal to \( r/r_m \), where \( r \) is the distance from the shower core and \( r_m \) is the Molière radius, a constant that depends on the atmospheric density at the location.
of the array.

This leads to a second equation for the particle density [38]:

$$\rho_N(r, t) = \frac{N_e(t)}{r_m^2} \frac{\Gamma(4.5 - s)}{2\pi \Gamma(s) \Gamma(4.5 - 2s)} f(x)$$

(2.3)

where $N_e$ is the number of particles in the shower after $t$ radiation lengths have passed. We will return to the NKG function in Chapter 4, where it will be used as the starting point to assign energies to the primary gamma rays.

The location where the gamma ray first interacts in the atmosphere is unknown to observers and fluctuates from shower to shower. These fluctuations contribute to the uncertainty in estimating the energy of the primary gamma ray. There are also fluctuations in the development of the shower itself that also complicate energy estimation.

Cosmic rays (charged particles) hitting the Earth’s atmosphere also initiate extensive air showers. These cosmic ray showers outnumber the ones initiated by gamma rays by orders of magnitude and are the background for any ground-based gamma-ray experiment. These hadronic showers have an electromagnetic (EM) component just like photon-induced ones do, but in this case the EM particles come mostly from the different decays of neutral pions and eta particles. Cosmic ray showers also contain charged pions, which in turn create muons and neutrinos (Figure 2.3). Muons make up approximately 10% of a hadronic shower [7]. These muons have high transverse momenta and will therefore cause the shower to have a larger lateral footprint when it reaches the ground (see the right panel of Figure 2.2).

Matthews has extended the Heitler model described above to hadronic showers in [36]. Assuming that the initial interaction creates some number of charged pions ($N_{ch}$) charged pions and 0.5$N_e h$ neutral pions (which feed the electromagnetic part of the shower by decaying to photons) he finds a relationship between the primary energy of the cosmic ray and energy in the hadronic and EM parts of the shower. The hadronic portion will have $(\frac{2}{3})^n$ of the initial energy, where $n$ is the number of interactions, while the electromagnetic portion will get the rest. This means that the electromagnetic portion gains most of the available energy fairly quickly.

Currently, two main types of experiments have been developed to detect gamma-ray initiated air showers: imaging atmospheric Cherenkov telescopes (IACTs) and
Figure 2.3: A diagram of a hadronic shower initiated by a proton. Note the presence of neutrinos, muons, and pions, absent in the model of an air shower initiated by a photon (source: hawc-observatory.org).

Extensive Air Shower (EAS) arrays. They are documented in Sections 2.2 and 2.3 respectively.

2.2 Cherenkov telescopes

As the particles in the air shower traverse the Earth’s atmosphere at relativistic speeds, they give off a type of radiation known as Cherenkov radiation. This electromagnetic radiation occurs when a charged particle is traveling through a medium with a velocity that is greater than the phase velocity of light in that same medium (i.e. the refractive index is >1). Cherenkov light has a characteristic blue glow and occurs because the charged particle disrupts the electromagnetic field, polarizing the medium. The resulting disturbance is radiated away as a coherent shockwave, which is sometimes described as analogous to the “sonic boom” created when an object travels faster than the speed of sound.

The emission is conical in nature and the Cherenkov angle is defined as:

\[
\cos(\theta) = \frac{1}{\beta n},
\]  

(2.4)
where $\beta$ is $v/c$ and $n$ is the index of refraction. Here, $v$ is the velocity of the particle and $c$ is the speed of light. The index of refraction relies slightly on the altitude. In air, $\cos(\theta)$ is $\sim 0.8$ degrees at 10 km [6]. If the medium is water, the angle is approximately 40 times larger. (This will become relevant in the discussion about water Cherenkov detectors in Section 2.3). Figure 2.4 shows the geometry of Cherenkov radiation.

The current generation of Cherenkov telescopes uses mirrors to focus the Cherenkov light from the air shower onto a camera, creating an image. Because of this, they are often known by their acronym “IACTs”.

Currently operating IACTs include VERITAS [39], MAGIC [40], and HESS [41]. They use a stereoscopic technique, where the air shower is viewed simultaneously from more than one telescope. This is important for background rejection. All of these experiments have multiple telescopes with 1000-2000 pixels in each camera, giving an average field of view of $\sim 5$ degrees [42]. Another IACT, FACT [43], uses Geiger-mode avalanche photodiodes as the photosensors. As of this writing, the Cherenkov Telescope Array (CTA) is under development [44] and will offer this method a boost in sensitivity.

With the ability to significantly detect the Crab Nebula, the brightest gamma-ray source in the sky, in a matter of minutes, the IACTs are the most sensitive detectors in the world for observing $>50$ GeV gamma rays. This makes them
well-suited to do in-depth studies of the morphology of sources. However, since they operate in a manner similar to traditional optical telescopes with respect to pointing, they cannot easily survey the entire sky or monitor it for gamma-ray transients. They also have a fairly low duty cycle (\(~10\%) as the night sky background limits their data taking capabilities to dark, moonless nights with clear weather.

Because of the way IACTs estimate the background, they cannot observe very extended objects. “Extended” here means that the object is comparable in size or larger than the field of view. This can occur either because the source is quite large or if the source is located very close to Earth.

### 2.3 Extensive air shower arrays

Unlike the IACTs, extensive air shower arrays detect the secondary particles from the air shower directly. This is typically done using either an array that contains scintillation material or with water Cherenkov detectors. This method alleviates some of the weaknesses that are inherent with the IACT method, but introduces some new weaknesses. For example, air shower arrays can in principle operate around-the-clock regardless of weather or moonlight conditions, but they have worse angular and energy resolution than IACTs. The two techniques can be thought of as complementary to each other.

The Tibet AS\(\gamma\) collaboration [45] operates the most notable currently-operating gamma-ray EAS detector of the scintillator type. Called Tibet-III, it has been operating at the Yangbajing site in Tibet since 1999.

EAS experiments relying on the water Cherenkov method consist of photomultiplier tubes (PMTs) placed in water. When the charged particles from the air shower hit this water, they give off Cherenkov light that is recorded by the PMTs. The Cherenkov angle, calculated using equation 2.4 analogously to the case of Cherenkov emission in air, is \(~41\) degrees.

This water Cherenkov technique was pioneered by the Haverah Park experiment, which studied cosmic rays. It was later applied to very high energy gamma rays by the Milagro Collaboration, who used PMTs in a pool of water (Figure 2.5) located in the Jemez Mountains in New Mexico. Milagro was the first water Cherenkov experiment to see and even discover TeV gamma-ray sources. Most notably, Milagro was the first experiment to see TeV emission from Geminga [46]
and from the Cygnus region [47].

PMTs are a type of vacuum tube capable of multiplying very weak signals. When light hits the PMT photocathode, electrons are ejected due to the photoelectric effect with some probability (the quantum efficiency). The electrons are accelerated towards a dynode (electrode) by the electric field and are multiplied via secondary emission before being accelerated towards a second dynode held at a higher voltage. This dynode chain continues, multiplying the original signal by many orders of magnitude. The last step in the chain is an anode, which results in an easily detectable current.

Due to the Milagro pool design, a muon skimming across the surface of the pool could trigger every single PMT, causing confusion between the hadronic background and legitimate gamma-ray showers. For this reason, next-generation experiments use tanks of water instead of a single pool. The PMTs from Milagro were later reused in the High Altitude Water Cherenkov Observatory (discussed in Chapter 3), which employs an array design and has 15 times the sensitivity that Milagro did. This sensitivity increase comes from a combination of its higher altitude placing it
closer to the shower maximum, the PMT optical isolation mentioned above, and an increase in the effective area [48].
Chapter 3  
HAWC

3.1 The HAWC detector

The High Altitude Water Cherenkov Gamma-Ray Observatory (Figure 3.1) is an array of 300 Water Cherenkov detectors located at an altitude of 4100 meters in the state of Puebla, Mexico, on the saddle point of the volcano Sierra Negra. This location (19.0° N) makes the observatory sensitive to sources with a declination range of -26° to 64°.

Each WCD consists of 4 upward-facing PMTs: in the center, there is a 10-inch high quantum-efficiency Hamamatsu R7081 PMT, surrounded by three 8-inch

Figure 3.1: An image of the completed HAWC-300 array, taken just before the inauguration in early 2015 (source: Jordan Goodman).
Figure 3.2: A 10-inch Hamamatsu R7081 PMT (left) and an 8-inch Hamamatsu R5912 PMT (right).

Figure 3.3: A schematic of an air shower particle hitting one of HAWC’s tanks and emitting Cherenkov light. Each tank has 4 PMTs; the one in the center is high-quantum efficiency. A person is shown for scale.
Figure 3.4: A schematic of the HAWC array. Each of the blue circles is a tank; the gap in the center is due to a structure known as the counting house that holds all the electronics. The green circles denote the location of the VAMOS engineering test array, which operated in 2011 and 2012.

Hamamatsu R7081 PMTs that were reused from Milagro (Figure 3.2). They are arranged in an equilateral triangle whose center is the 10-inch PMT; each 8-inch PMT is \( \sim 1.4 \) meters from the central PMT. The addition of the high quantum-efficiency PMT increases sensitivity to sub-TeV air showers. See Figure 3.3 for a schematic of a tank, and Figure 3.4 for a schematic of the entire array. Each tank is 7.3 meters in diameter and holds 200,000 liters of purified water in a bladder, for a water height of 4.5 m. The outer enclosure is made of stainless steel and has a roof to protect the instrument from rain and snow.

### 3.1.1 Electronics, data acquisition, and calibration

A structure known as the “counting house” is located in the middle of the array. An RG59 cable is run from each PMT to this structure. This cable both provides the high voltage to the PMT (chosen for each individual PMT based on the gain) and carries the signal back to the data acquisition system (DAQ) in the counting
Figure 3.5: A two edge event. The curve is the pulse from the PMT. Note that it does not cross the second, higher threshold. Image used with permission of Zigfried Hampel-Arias.

Like the PMTs, the front end electronics were reused from Milagro. The PMT signals travel through the cable and lightning protection spark gaps and enter the custom-built front end boards. If the signal crosses a certain discriminator threshold, it is given a time stamp and digitized by the CAEN VX1190A multi-hit time to digital converters (TDCs). The TDCs measure a quantity called “Time over Threshold” (ToT), which is defined as the difference between the leading and trailing edges of a signal. Two thresholds are in use, 0.25 photoelectrons and 5 photoelectrons [49]. Low charge events will have 2 edges, and high charge events will have four, from crossing both thresholds (Figures 3.5 and 3.6). This makes the
trigger entirely software-based, with an air shower trigger rate of $\sim 25$ kHz. All the TDCs are kept synchronized by a GPS Timing and Control (GTC) system. The resulting information is then read out to single board computers (SBCs) that are connected to an Ethernet network.

Signals are fed to the two data acquisition systems and the online reconstruction mentioned in the preceding section is then run. The online analysis is done in real time and relies on ZeroMQ, a software used for data transfer, and the Analysis and Event Reconstruction Integrated Environment (AERIE), which is HAWC’s dedicated analysis software. AERIE is built on a C++ framework. There are two DAQs: the main one used for all analyses, and a scalar DAQ that simply counts the signal rates in a specified time window and is mainly used for gamma-ray burst analyses. The DAQ and online processing system are further documented in [50]. Raw data are saved to disk and a more in-depth analysis is run offline. All of the analyses in this dissertation were run off-site.

Figure 3.6: A four edge event. Unlike the two-edge case, this PMT pulse crosses both thresholds. Image used with permission of Zigfried Hampel-Arias.
All 1200 PMTs are calibrated using short light pulses from a 532 nm wavelength laser; neutral density filters can vary the light levels over 7 orders of magnitude. This range in intensity is necessary to calibrate the PMTs from one photoelectron to several thousand.

As described in [51], “calibration” is broken down into charge calibration and timing calibration. Charge calibration aims to translate the ToT described above to a number of photoelectrons. Timing calibration determines the relative timing between PMT channels and accounts for effects such as the PMT slewing time, the fact that light will take a different path to each PMT, and time residuals (the difference between the PMT time and what would be expected from fitting the air shower). We will return to this discussion of calibration in Section 4.4.2, which describes some systematic errors in the energy estimation algorithms.

### 3.2 Air shower reconstruction

After data are removed from the site, the full reconstruction and analysis chain is run. This begins with a core and angle fit followed by a step where the gamma-like events are selected. These are then made into skymaps and a likelihood method is used to study sources.

#### 3.2.1 Core and angle fits

An air shower hitting the HAWC array fulfills the single multiplicity trigger condition if a certain number of PMTs (28 as of this writing) are hit within 150 nanoseconds of each other. PMT hits in the range (-500 ns, +1000 ns) from the trigger time are saved and used in the reconstruction. The pattern of these PMT hits (Figure 3.7) is used to reconstruct the air shower core, and the timing of the hits (Figure 3.8) is used to reconstruct the air shower zenith angle. Some PMT hits are removed if they occur directly after a high-charge hit in the same PMT, since PMT afterpulsing is a concern.

Recall from the discussion of air shower physics in Chapter 2 that the shower will have the greatest number of particles along the axis that the primary particle would have traversed if it had not interacted in the atmosphere (see Figure 2.2). The point where this axis intersects the ground is called the core and denotes the
area where the greatest amount of charge will be measured by the PMTs. The lateral spread could be fit to the NKG/$r$ function:

$$\rho(r) = A \left( \frac{r}{r_m} \right)^{s-3} \left( 1 + \frac{r}{r_m} \right)^{s-4.5} \quad (3.1)$$

where $r_m$ is once again the Molière radius and $A$ is the overall normalization. NKG/$r$ is used instead of the traditional NKG function (Equation 2.2) because it has been empirically found to measure energy density instead of the particle density that the NKG function measures. However, the algorithm to find the core using NKG/$r$ takes a large amount of CPU resources due to the presence of gamma functions embedded in the normalization (see Equation 2.2) and a pole at $r = 0$ in the formula. Instead, the current core fitter is known as the “Super Fast Core Fitter” (SFCF) and, as the name suggests, has been chosen for its speed with which it can fit the core. The SFCF functional form is a combination of an NKG and a Gaussian function. A simple Gaussian function cannot be used because physical showers have tails that are much longer than a Gaussian function has.

The SFCF assumes that the signal in the $i$th PMT will be

$$S_i = A \left( \frac{1}{2\pi\sigma^2} e^{-|\vec{x}_i - \vec{x}|^2/2\sigma^2} + \frac{N}{(0.5 + |\vec{x}_i - \vec{x}|/r_m)^2} \right) \quad (3.2)$$

Here, $\vec{x}$ is the location of the core, $\vec{x}_i$ is the location of the $i$-th PMT, $r_m$ is the Molière radius, $\sigma$ is the traditional sigma associated with a Gaussian fit, $A$ is the overall normalization, and $N$ is a normalization of the tail component.

The angle fitter uses the relative timing of the PMT hits. One might naively expect for the air shower to arrive as a perfect plane, with the timing being dependent on the zenith angle of the air shower. This is true to first approximation, but further study shows that the shower front actually has a conical shape. This is because air shower particles further from the core are subject to multiple scattering and thus travel a slightly further distance, delaying their arrival. Additionally, there is an artificial effect caused by the decreased density of the shower particles far from the core, leading to a lesser probability of a PMT detecting them. These two effects are large enough to necessitate the introduction of a correction in the reconstruction, known as the “curvature/sampling correction” [4]. This correction is on the nanosecond time scale and is a function of both the distance of the PMT
Figure 3.7: The HAWC event display showing the distribution of charge for a gamma-like event at the location of the Crab Nebula, which is used as a standard candle in gamma-ray astronomy. 686 PMTs were hit during this event. This information is used in the core fit. The star denotes the location of the fitted core; the circle has a radius of 40 meters and is important for gamma/hadron separation, along with the PMT outlined in red, which is the PMT with the highest charge outside that circle (discussed in section 3.2.2).

to the shower core and of the charge measured in the PMT. Once this correction has been made, the hits can then be approximated as a plane and a $\chi^2$ fit is performed.

Determination of the core and zenith angle is a multi-step, iterative process. Before the SFCF fit is performed, a simple center-of-mass fit is carried out. The result serves as the guess for the SFCF fit, which is the second step in the core/angle fit chain. This is immediately followed by the directional fit. The entire process is then repeated, using only hits that fall within 50 nanoseconds of the shower plane. This has been known to increase the significance of the Crab Nebula. The leading hypothesis is that a multi-step core/angle fit does a better job dealing with the random noise that is inherently present in the PMTs, some of which is due to
accidental air shower muons [4].

The gamma-ray energy estimates that will be discussed in Chapter 4 rely on the signal a certain distance from the shower axis, and treat the core fit result as correct with no uncertainty. Therefore, an accurate core fit is very important for precise estimates of the energy. The SFCF returns fairly small errors for showers that have cores landing on the array; even for small events, where less than a quarter of the array triggered during an air shower, the median error is \( \sim 4 \) meters [4]. This error becomes much larger for cores that land off the array, often reaching an order of magnitude higher. For this reason, only events with cores that are reconstructed on the array are used in the energy-based analysis.
Table 3.1: The definitions for each $f_{hit}$ bin, where $f_{hit}$ is the number of PMTs hit that are within 20 ns of the shower plane divided by the total number of available PMTs after calibration.

<table>
<thead>
<tr>
<th>Bin number</th>
<th>$f_{hit}$ percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6.7-10.5</td>
</tr>
<tr>
<td>2</td>
<td>10.5-16.2</td>
</tr>
<tr>
<td>3</td>
<td>16.2-24.7</td>
</tr>
<tr>
<td>4</td>
<td>24.7-35.6</td>
</tr>
<tr>
<td>5</td>
<td>35.6-48.5</td>
</tr>
<tr>
<td>6</td>
<td>48.5-61.8</td>
</tr>
<tr>
<td>7</td>
<td>61.8-74.0</td>
</tr>
<tr>
<td>8</td>
<td>74.0-84.0</td>
</tr>
<tr>
<td>9</td>
<td>84.0-100.0</td>
</tr>
</tbody>
</table>

3.2.2 Event selection

3.2.2.1 Fraction hit bins

Not every event that is reconstructed is used in the standard gamma-ray analysis. Events are divided into bins based on shower size. Shower size is defined by $\frac{n_{hit}SP20}{nChAvail}$. $n_{hit}SP20$ is the number of PMTs hit within 20 nanoseconds of the shower front; $nChAvail$ is the number of channels available at the time of the event. These bins are known as fraction-hit ($f_{hit}$ bins)$^1$. Since HAWC was built in a modular fashion, with tanks being turned on as they were constructed, $f_{hit}$ bins were chosen instead of simply using the number of PMTs hit in an event because they allow for comparisons in size between different construction epochs. Even now, with construction completed, it is still a useful quantity as portions of the detector may occasionally be turned off for maintenance. Data taking proceeds with a partial array during these time periods. Table 3.1 shows the $f_{hit}$ bin definitions used in this dissertation.

$f_{hit}$ bins are necessary because the angular resolution of HAWC is a function of event size, ranging from about 1 degree in the lowest bin to better than 0.2 degree in the highest. Background rejection capabilities also vary strongly between

$^1$Note that this is a slightly different binning scheme than what was used in previously published HAWC papers, such as [4] and [11]. There, $f_{hit}$ bins are defined slightly differently: $n_{hit}/nChAvail$. The new definition, which only selects within 20 nanoseconds of the shower plane, prevents noise hits from contributing to the fraction of PMTs hit.
Figure 3.9: The distribution of true energies within each $f_{hit}$ bin, assuming a transiting source with a power law index of -2.63 and a declination of 20 degrees N. Note that $f_{hit}$ acts as a rudimentary energy estimator. An increase in bin is roughly correlated with an increase of energy, but there are large overlaps in some bins. Additionally, Bin 9 functions as an overflow bin that contains a large portion of HAWC's energy range. This figure is made using HAWC Monte Carlo simulations.

$f_{hit}$ bins: events with more PMTs hit are much rarer, but they have much better angular resolution and gamma/hadron separation power is much better.

The threshold for Bin 1 is 6.7%. Note that in principle “Bin 0” does exist, consisting of everything just above the trigger threshold (28 PMTs hit), but as of this writing it is not used in analyses as there are some problems with its sensitivity that are poorly understood. Rectifying this is an active area of research.

The 9 $f_{hit}$ bins can also be used as a proxy for energy (in fact, in all HAWC papers up to this point, including [4] and [11] it is the only energy proxy) by using simulation to determine the mean gamma-ray energy of each bin. However, it is only a mediocre energy estimator. While the size of the event is correlated with energy, $f_{hit}$ fails to take into account other variables that also have an effect on the energy, such as the angle at which the primary gamma ray enters the atmosphere. Additionally, there is a large overlap in energy between bins (see Figure 3.9).

There is also a saturation problem related to energy estimation and the highest
$f_{hit}$ bin (Bin 9). Here, nearly every PMT in the array is hit during an air shower event. The mean energy of this bin is a few tens of TeV but HAWC’s sensitivity extends up to 100 TeV and beyond. All of those highest-energy events are included in this bin, but there is a complete lack of dynamic range at the highest energies if $f_{hit}$ is used as the sole energy estimator, as it can only determine the mean energy of the bin. Note that in Figure 3.9, the energy distribution for Bin 9 is much broader than that of the others. Chapter 4 describes a separate event-by-event energy estimator that can be used to solve this problem.

### 3.2.2.2 Gamma/hadron separation

The hadronic background is orders of magnitude larger than the desired gamma ray signal - the air shower trigger rate is $\sim 25$ kHz, but only approximately 400 gamma rays arrive per day from HAWC’s brightest source, the Crab Nebula. This necessitates the development of gamma/hadron separation algorithms to remove as much of the background as possible.

Recall from Chapter 2 that gamma-ray induced air showers look physically different from those induced by hadrons. Most notably, a gamma-ray shower is smoother and somewhat more compact due to the absence of particles such as muons that have high transverse momenta and hit the ground far from the shower core.

Currently, there are two gamma/hadron separation variables in use in HAWC: compactness and PINCness. The two variables vary in their gamma/hadron separation abilities, so using both of them in tandem has been shown to improve performance on the Crab Nebula. The specific values of compactness and PINCness to cut on are optimized in each analysis bin.

Compactness ($C$) is defined by

$$C = \frac{n_{hit}SP20}{CxPE40}$$

(3.3)

where again $n_{hit}SP20$ is the number of PMTs hit within 20 nanoseconds of the reconstructed shower front, and $CxPE40$ is the highest charge more than 40 meters from the shower core. Events with a high $CxPE40$ value are more likely to be background, meaning that events with a higher compactness value are more gamma-like (see Figure 3.10). The value of the compactness cut in the standard $f_{hit}$-based
Figure 3.10: The inverse-compactness distribution for events near the Crab Nebula where more than 75% of the array is hit (the inverse is plotted because the gamma rays are easier to see on the plot). Gamma rays can be seen as the excess on the left-hand side of the plot.

analysis ranges from 3.0 to 18, depending on the $f_{hit}$ bin.

See Figure 3.7 for a visualization of how compactness works. In this figure, the PMT outlined in red directly to the left of the circle (which has a radius of 40 meters) is the highest charge PMT and is known as $CxPE40$. It has a charge of 49 PEs and an $n_{hit}SP20$ value of 675, giving it a compactness of 13.77. This is far above the compactness cut of 3.0 that is used in the analysis bin this event belongs to.

While compactness separates gamma rays from hadrons by using the single largest charge recorded far from the core, the Parameter for Identifying Nuclear Cosmic rays ($PINCness$) uses all of the hits to compute how “smooth” the lateral distribution of the charge is.

This variable is at its core a chi-square fit. First, for each PMT, an expected logarithmic charge is computed by averaging up all of the logarithmic charges in the 5-meter wide ring (centered at the shower core) that contains this PMT.
Figure 3.11: The PINCness distribution for events near the Crab Nebula where more than 75% of the array is hit. Gamma rays from the Crab Nebula are immediately evident by noting the excess above the surrounding region at low values (the bump between PINCness values of 1.0 and 2.0).

\[
\langle \log_{10}(q_i) \rangle. \text{ Then, PINCness } (P) \text{ is computed:}
\]

\[
P = \frac{1}{N} \sum_{i=0}^{N} \frac{(\log_{10}(q_i) - \langle \log_{10}(q_i) \rangle)^2}{\sigma_{\log_{10}(q_i)}^2}.
\]  

(3.4)

Here, \(q_i\) is the effective charge in the i-th PMT, while \(N\) is the total number of hits. \(\sigma\) was determined by selecting a sample of high \(f_{\text{hit}}\), extremely gamma-like events and fitting the distribution of measured charges for a variety of different predicted charge levels. This algorithm can be thought of as measuring the charge resolution directly from the data.

The smoother the lateral distribution of the event is, the more gamma-like it is and the smaller the value of PINCness is (see Figure 3.11). The standard \(f_{\text{hit}}\)-based analysis selects events with a PINCness value less than 1.6-3.0, depending on the analysis bin.

The effectiveness of the gamma/hadron separation varies depending on the analysis bin. In the lowest \(f_{\text{hit}}\) bin, approximately 80% of the gamma rays pass the
gamma/hadron separation cuts, but more than 10% of the hadrons also do. Since there are orders of magnitude more hadronic showers, the lower analysis bins are therefore mostly background even after cuts. To contrast this, in the highest $f_{hit}$ bin the background rejection is much stronger: as before, 80% of the gamma rays pass the gamma/hadron cuts, while only $\sim$0.1% of the hadrons do.

3.2.2.3 Angular resolution

Using Monte Carlo simulations, the difference between the true arrival direction of the incoming gamma ray and the reconstructed direction can be investigated. The fit to this is known as the point spread function. In HAWC, the point spread function is parameterized by a combination of two 2D Gaussian fits.

In general, angular resolution improves as the $f_{hit}$ bin increases. This is because a larger portion of the array is hit and there is more information available to reconstruct the event.

3.2.2.4 Data quality cuts and dataset

The detector was inaugurated in the spring of 2015, but data taking began long before that, with the tanks being added as they were completed. Science operations formally began on August 1, 2013 with 100 tanks (HAWC-100). This dissertation uses only uses data from the time after the 1000th PMT was installed, with the earliest data being taken in June 2015.

Not all data are reconstructed equally well. Some data quality cuts are made before binning the events in right ascension (RA) and declination to create maps of the sky. Events must successfully pass the core and directional fits. Additionally, only events where at least 800 PMTs\(^2\) were switched on and $nChAvail/nChTot > 0.9$ are used. $nChAvail$ is the number of PMTs used after edge finding and calibration is done, while $nChTot$ is the number of PMTs that were switched on during the subrun\(^3\). These two cuts require that most of the array is operational and removes air showers where large numbers of PMTs are thrown out during calibration, which could impact the $f_{hit}$ bin to which the event is assigned.

---

\(^2\)Even though data with $>1000$ installed PMTs is used here, portions of the array are somewhat frequently switched off for maintenance.

\(^3\)Events are divided into $\sim$24 hour chunks of data known as runs, which are further subdivided into subruns that hold a few million air shower events each.
Additionally, the energy-based analysis requires an even stricter event selection. This will be described later, in Chapter 5.

After the cuts are made, an additional event selection is performed to remove subruns where the detector was determined to be unstable. This is determined by looking at the variance of the zenith angle distribution. This is necessary because the background estimation method used (discussed in Section 3.3) requires 2-hour periods of data where the data are recorded at a fairly uniform rate. The HAWC data frequently has small amounts of dead time, and the DAQ often crashes due to weather conditions such as lightning strikes. Removing periods of time where the detector is unstable ensures that the background estimation is accurate. The amount of data removed during this step is small (∼ 2% of the data).

Once the event selection is completed, skymaps are generated using standard HAWC analysis software.

### 3.3 Map-making, background estimation, and likelihood analysis

In order to analyze the data, skymaps must be created. In the standard HAWC analysis, maps are created for each $f_{hit}$ bin, but they will be created using a 2D $f_{hit}$/energy binning scheme in the future, as described in Chapter 5.

Maps are created using a Hierarchical Equal Area isoLatitude Pixelization (HEALPIX) map [52]. The pixelation scheme has many features that make it useful for astrophysics, including pixels that are equal area and pixel centers that lie in discrete rings of latitude, with the number of rings being dependent on the resolution needed. Maps shown in this dissertation all have a spacing between the pixel centers of 0.06 degrees, which is less than the angular resolution in any bin and is therefore more than sufficient for analysis.

After the raw data maps, which are merely a histogram of events numbers across the sky, are created, the background estimate is performed. A method known as “direct integration”, originally pioneered by the Milagro Collaboration, is used [53]. Assuming that the detector is stable over a 2 hour time period and also assuming that the cosmic ray background is isotropically distributed (both of which have
been validated), the background is calculated according to

\[ N = \int \int E(\text{ha}, \delta) R(t) \epsilon(\text{ha}, \text{RA}, t) dt d\Omega \]  

(3.5)

where \( E \) is the efficiency in local coordinates (hour angle and declination) and is found from the data map, \( R(t) \) is the event rate as a function of time, and \( \epsilon \) either 0 or 1 depending on whether the local bin contributes to the bin of right ascension and declination at a particular time. In practice, the background is found by summing the efficiency map over the observed events.

HAWC has better background rejection than Milagro, so an additional smoothing of the background by 0.5 degree is necessary. This compensates for the relatively low statistics in the highest \( f_{\text{hit}} \) bin, and prevents having pixels where the background estimate is zero. Pixels with zero background cause problems in a likelihood analysis.

Note that when a bin has extremely low statistics, direct integration cannot be used as it requires a minimum number of events. This can happen in a variety of cases, such as in rare high-energy bins or during transients when only a short time period of data is being analyzed. A variety of background estimation methods have been created to account for this breakdown. In this dissertation, low energy bins will use a background randomization method instead of direct integration. This will be discussed in Section 5.3.

The significance of sources is computed using a likelihood analysis given a specific source geometry and spectrum. This is done for both the whole sky and for specific sources. HAWC’s analysis framework is called LiFF (Likelihood Fitting Framework) [54]. This framework takes into account a given source model, the data and background maps, and the expected detector response and calculates a binned Poisson log-likelihood value. A likelihood ratio test is used to compute the test statistic:

\[ TS = 2 \ln \frac{\mathcal{L}(\text{Alternative hypothesis}); N_{\text{obs}}}{\mathcal{L}(\text{Null hypothesis}); N_{\text{obs}}}. \]  

(3.6)

Here \( \mathcal{L} \) is the standard definition of the likelihood. Taking Wilks’ Theorem [55] into account, the significance is simply \( \sqrt{TS} \) for a nested model since the test statistic is distributed as the chi-square distribution with the number of degrees of freedom equal to the number of free parameters between the hypotheses.

All-sky maps, such as the ones in [11], are created assuming a point source with
a power-law spectrum (index 2.7), meaning that the flux is the only free parameter. 2.7 is a typical spectral index of sources detected by HAWC. Searches for extended sources are also run, assuming a power-law spectrum of 2.0. A HAWC point source skymap, made with the first 1017 days of HAWC data, can be seen in Figure 3.12.

LiFF is also used to determine the spectra of sources through a likelihood maximization (i.e. the parameters of the source model that maximize the test statistic are found).

### 3.4 HAWC Simulation

In order to determine how the detector responds to air showers, simulations are necessary. Air showers (both gamma-ray and hadronic) are simulated using the CORSIKA software package [56]. CORSIKA is used to propagate the particles in the air shower from the point of the first interaction all the way to the detector level. Then, a detector simulation based on GEANT 4 [57], a toolkit developed by CERN and capable of simulating the movement of particles through matter, is used to propagate those same particles through the HAWC water tanks and tracks the Cherenkov light that is created.

The simulated events are then reconstructed using the same analysis software (AERIE) that is used to reconstruct data. During this process, noise is added, and the PMT efficiency is matched to what is observed in data by scaling the vertical muon information. Vertical muons provide a constant light source. The events are thrown with a nonphysical distribution in order to probe rarer parts of the parameter space so the Monte Carlo must be weighted after the fact. For example, one may weight a gamma-ray sample to model a source with a certain spectral index transiting overhead, or the events may be isotropically weighted.

### 3.5 Sensitivity

HAWC is optimized to detect gamma rays from 100 GeV to 100 TeV, but as will be seen in Chapter 4, the energy range can be extended higher by measuring the gamma-ray energies on an event-by-event basis.

Figure 3.13 shows a plot of differential sensitivity to a point source with a power-law spectrum of $E^{-2.63}$ using the standard $f_{hit}$-based analysis. The numerous
Figure 3.12: The HAWC point source skymap, created with 1017 days of data. A power-law index of 2.7 is assumed. This map includes data from all $f_{\text{cut}}$ bins. In Chapter 6 this map will be broken down into individual energy bins.
Figure 3.13: The differential sensitivity of HAWC for a point source with an $E^{-2.63}$ spectrum. The dark red line is the fit to the flux in each $f_{hit}$ bin required to detect the source at $5\sigma$ 50% of the time (the nine light red lines). Lines for 50 hour observations from various IACTs are shown for comparison. See the text of section 3.5 for more information.

Light red lines denote the flux required in each $f_{hit}$ bin for the source to be detected at the $5\sigma$ level 50% of the time. This is then fit to obtain the curve labeled “HAWC 507-day”.

HAWC is sensitive to sources with spectra at least as faint as 50 mCrab (integral power-law spectra of $\sim 5 \times 10^{-13} \text{ cm}^{-2} \text{ sec}^{-1}$ above 2 TeV with a year of data taking). This number was published in the original HAWC sensitivity paper [58], which was written before construction was completed and had a large uncertainty related to lack of knowledge of the background above 10 TeV. Since HAWC’s background estimate is derived from the data, this uncertainty has been eliminated.
now that years of data are available.

3.6 Science goals

As HAWC was designed as a survey instrument with a wide variety of applications, the scientific goals of the experiment are broad. Most of these goals relate to the biggest questions in the field of particle astrophysics: how are cosmic rays and other particles accelerated, and how do they propagate through the universe?

The search for high-energy (> 50 TeV) gamma-ray sources that will be described in the later chapters of this dissertation contributes to this goal by looking for spectral features that may be indicative of a cosmic ray acceleration site.

However, even sources that do not emit at these energies can contribute to our knowledge of particle acceleration models. For example, leptonic and hadronic sources are expected to have different spectral features in the TeV range, and HAWC will be able to distinguish between these two source classes. Currently, most of these sources only have lower energy observations available, where it is very difficult (or impossible) to distinguish between leptonic and hadronic emission.

In addition to spectral and morphological studies, HAWC also functions as a discovery instrument. Of the 39 sources identified in the 2HWC (second HAWC) catalog [11], 16 are more than a degree away from known TeV sources. While some of these sources can be classified as pulsar wind nebulae, blazars, etc. due to observations at other wavelengths, most of them remain unidentified in nature.

HAWC is particularly adept at discovering and probing extended TeV sources. As discussed in Chapter 2, IACTs often have trouble studying extended objects that are larger than or comparable in size to their field of view. HAWC, meanwhile, has confirmed the Milagro Collaboration’s detection of Geminga and discovered a second source close to it, coincident with the location of the middle-aged pulsar PSR B0656+14 [59]. Both of these are a few degrees across in size. These close, middle-aged pulsars have been postulated to be the source of the positron excess measured at Earth, but the HAWC measurements have created tension with that hypothesis. Studies of other extremely large-scale structures have also been carried out, such as the Fermi Bubbles [60].

Studying the diffuse emission (emission that cannot be resolved into a specific source) can also help probe cosmic rays by looking at the gamma ray signatures
that form when the cosmic rays interact with gas and dust. Diffuse emission can be of the Galactic variety [61], which can set constraints on cosmic ray acceleration and propagation within our Galaxy (as mentioned previously in Section 1.2), or it can be isotropic [62], which has contributions from cosmic ray interactions with the intergalactic medium, as well as a possible dark matter contribution. Note that isotropic diffuse emission is expected to accompany the astrophysical neutrinos seen by IceCube, provided that the neutrinos are within the gamma-ray horizon (gamma rays too far away from us will never reach the Earth because they are attenuated by the extragalactic background light).

HAWC is a powerful transient detector as well. Although as of this writing, no gamma-ray bursts have been detected, numerous flares of the AGN Markarian 421 have been observed. This has led to Astronomer’s Telegrams notifying the community [63,64] and resulted in contemporaneous observations with other instruments, namely FACT. While the first few science goals discussed above mainly concern Galactic science, studying both AGN and GRBs helps us learn more about extragalactic cosmic rays.

The ability to detect transients has immediate consequences for multimessenger astrophysics. Due to its high duty cycle, HAWC can either notify other observatories of gamma-ray transients, or can follow up on reports from other experiments. HAWC has performed follow-up observations of high-energy neutrinos from IceCube [65] and has searched for electromagnetic counterparts of gravitational waves detected by LIGO [66]. HAWC and IceCube are especially good partners because they observe roughly the same part of the northern sky, and their energy ranges are compatible.

Although primarily designed to study gamma rays, HAWC is also capable of using its immense background to probe the cosmic ray anisotropy [23]. This can directly tell us about the distribution of cosmic ray sources. Cosmic rays can also be studied more indirectly; for example, limits on the proton/antiproton flux have been made by studying the Moon shadow [67].

Lastly, HAWC can investigate fundamental physics phenomena, such as dark matter searches [68] or investigations of Lorentz invariance violation [69].
Chapter 4  
Energy estimation method

As discussed in the preceding chapter (Section 3.2.2.1), HAWC employed $f_{hit}$ as a rudimentary energy proxy in its first published papers. This method has two main drawbacks. First, it is only weakly correlated with energy since it does not take into account additional variables that have an effect on the energy, most notably the zenith angle and how well-contained the shower is within the array. Second, there is a complete loss of dynamic range above 10 TeV. At these energies, almost every PMT in the array is hit during an air shower. With the $f_{hit}$ method, it is only possible to extract the mean energy of an analysis bin. This is an especially egregious loss given that HAWC is the only currently operating experiment with sensitivity up to 100 TeV.

Clearly, the development of an event-by-event energy estimator is warranted. One technique for doing this was developed in the early 1970s by Hillas [70] and is frequently used in large cosmic ray experiments [3, 71]. This method relies on measuring the charge on the ground some distance from the shower axis. Several modifications must be made to account for the differing array design between the average cosmic-ray and gamma-ray experiment (cosmic-ray experiments are typically orders of magnitude larger, with much greater separation between individual tanks). This is the first time this technique has been applied to purely electromagnetic gamma-ray showers.

Note that the HAWC Collaboration has developed two completely independent energy estimation algorithms. The other method, a neural network [72], will not be discussed here.
4.1 Determination of the optimal radius

In order to translate the charge on the ground into energy, first one must decide at what distance from the air shower axis the charge should be measured. Measuring too close to the shower axis will result in a sub-optimal measurement due to the presence of large fluctuations from shower-to-shower. Conversely, measuring too far away introduces the possibility that threshold effects from the electronics may adversely affect the measurement.

Additionally, it is not immediately clear to what function the lateral distribution should be fit. The NKG function (Equation 2.2) is a purely empirical function, and it is not immediately clear that it will work better than other potential functions for energy estimation.

To determine the optimal radius ($r_{opt}$) at which to measure the charge, the
effective charge\(^1\) as a function of distance to the shower axis (the lateral distribution function) is fit to an NKG-like function (Figure 4.1). The fit is done in log-space because several important charged-related distributions, such as the charge error, are log-normal. The fit function is of the form

\[
\log_{10}(NKG) = \log(A) + s[\log_{10}(r/124.21) + \log_{10}(1 + r/124.21)] - 3\log_{10}(r/124.21) - 4.5\log_{10}(1 + r/124.21)
\] (4.1)

where \(A\) is the overall amplitude of the fit, \(s\) is a parameter related to the shower age, \(r\) is the distance from the PMT to the shower axis in meters, and 124.21 m is the value of the Molière radius at HAWC. Note that this is not the traditional form of the NKG function described in Equation 2.2. NKG/\(r\) was chosen because it describes energy density, while the NKG function only describes particle density. This is due to the fact that the higher energy particles in the air shower get scattered through smaller angles and therefore are closer to the shower axis [73]. Therefore, \(s\) is not the shower age. A conversion factor must be used to get the true age, and \(s \neq 1\) at the shower maximum.

The charge error on each PMT hit in the lateral distribution is the same as the \(\sigma\) used in the PINCness calculation (Equation 3.4), determined by selecting a sample of high \(f_{hit}\) extremely gamma-like events and fitting the distribution of measured charges for different charge levels.

In order to find the distance where the uncertainty in the correct lateral distribution function is minimized, \(s\) is varied by \(\pm 10\%\) from the best fit value obtained in Equation 4.1. Since there are only two free parameters, this forces a change in \(A\). All possible fits from varying \(s\) are plotted (the band in Figure 4.1). The optimal radius is the point where this band is the smallest; the uncertainty in knowing what the exact form of the lateral distribution function is smallest here.

In determining the optimal radius, only a few quality cuts were applied to the simulation: in addition to the standard quality cuts outlined in Section 3.2.2.4, the event must be reconstructed with its shower core on the array and the difference between the reconstructed and true zenith angles of the event must be < 0.5\(^\circ\). A zenith angle cut of 45\(^\circ\) is applied, as studies have shown that the energy estimation method is less effective above this angle. No gamma/hadron separation cuts were

\(^1\)Recall that the central PMT in each tank is a larger, high quantum-efficiency PMT. A lateral distribution consisting of measurements from two different types of PMT necessitates the use of an effective charge that normalizes the differences between the two.
Table 4.1: $r_{opt}$ for different shower sizes. The optimal radius slowly increases as the showers get bigger.

<table>
<thead>
<tr>
<th>Percent of array hit</th>
<th>Mean $r_{opt}$ (m)</th>
<th>RMS (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.7-25</td>
<td>30.34</td>
<td>4.91</td>
</tr>
<tr>
<td>25-50</td>
<td>34.87</td>
<td>5.07</td>
</tr>
<tr>
<td>50-75</td>
<td>40.09</td>
<td>5.66</td>
</tr>
<tr>
<td>75-100</td>
<td>45.07</td>
<td>6.30</td>
</tr>
</tbody>
</table>

Figure 4.2: Distribution of the optimal radius for different event sizes. Plot (a) starts at 6.7% of the array hit as this is the threshold for the first analysis bin. The percentage of the array hit is defined by $nHitSP20/nChAvail$. Each histogram contains 5000 simulated events. No gamma/hadron separation cuts were applied, however a cut to remove events with reconstructed cores off the array was applied along with the standard data quality cuts and a requirement that the difference between the reconstructed and true zenith angle of the event is $< 0.5^\circ$. 

46
applied, as these can in theory be optimized for each individual analysis. These same cuts will be used when the energy estimation algorithm is determined (see Section 4.2).

Figure 4.2 shows the distribution of $r_{opt}$ for different $f_{hit}$ levels ($f_{hit}$ is again defined as $nHitSP20/nChAvail$). This histogram, along with all others in this chapter, is weighted to resemble events coming from an isotropic distribution of events with a spectrum of $E^{-2}$. This spectrum is harder than most astrophysical sources, but was chosen because it ensures good statistics at the highest energies and gives an energy assignment that is robust no matter what the actual spectrum of the source is.

Note that the optimal radius slowly increases as air shower events get larger, going from a mean value of 30.24 m when less than a quarter of the array is hit to 45.07 m when more than three quarters is hit. Table 4.1 summarizes the optimal radii for different event sizes.

For simplicity, an optimal radius of 40 meters is chosen for the analysis, used irrespective of $f_{hit}$ bin. This provides a good balance between the preferred optimal radius for different event sizes. The signal at this distance will be referred to as $s_{r_{opt}}$ throughout the rest of this dissertation. Later in this chapter (in Section 4.4), it will be shown that the systematic uncertainty from using one optimal radius for all events is negligible.

The optimal radius only shows a slight trend with zenith angle (see Figure 4.3). This mirrors findings from [3], which found that the optimal radius depends largely on array geometry and very little on other characteristics such as zenith angle of the event.

4.1.1 Cross-check with minimization of the function

Newton et al. give an alternative method for calculating the optimal radius in [3]; this can be used to cross-check that the fitting technique described above gives the expected results.

The variation in the slope is the smallest when the signal is minimized with respect to the slope parameter ($\frac{\partial(NKG)}{\partial s}$). This requires the assumption that the amplitude in the NKG function (Eq. 4.1) has a dependence on the age (i.e., $A = A(s)$). The math for the minimization is done in [3], but we redo it here to
account for the fact that we are working in log space, not in natural log space as Newton et al. assume. Differentiating Equation 4.1 gives:

\[
\frac{\partial \log_{10}(NKG)}{\partial s} = \frac{\partial \log_{10}(A)}{\partial s} + \log_{10}(r/124.21) + \log_{10}(1 + r/124.21).
\] (4.2)

Setting this equal to 0 and solving for \( r \) gives:

\[
\frac{r_{\exp}}{124.21} = \frac{1}{2} (-1 \pm \sqrt{1 + 4 \times 10^{-\partial \log_{10}(A)/\partial s}}).
\] (4.3)

In practice, \( \partial \log_{10}(A)/\partial s \) is calculated by plotting several values of \( \log_{10}(A) \) vs. \( s \) and fitting them to find the slope.

The difference between this “expected optimal radius” \( r_{\exp} \) and the \( r_{\text{opt}} \) calculated above can be seen in Figure 4.4. The two methods give essentially the same result.
Figure 4.4: Cross-check between minimizing the function, as is done in [3] and the method presented here. The mean is 5.168e-5 and the RMS is 9.194e-3. The histogram includes 5000 events of all event sizes.

4.1.2 Saturated tanks

Cosmic ray scientists have spent much time studying the effect on the optimal radius of tanks where the signal is saturated. Newton et al. found that the presence of saturated tanks at the Pierre Auger Observatory can change the optimal radius by as much as 500 meters [3]. Due to the differences between cosmic-ray and gamma-ray experiments, this is not as concerning in HAWC. If PMT hits with effective charges near their maximum calibrated charge are removed from the fit, the average optimal radius only increases by $\sim 1$ meter. As will be seen at the end of this chapter, using a radius that is off from the “true” one by tens of meters introduces no noticeable error. Additionally, the maximum calibrated charge in each PMT is different and removing these hits would increase the computation time greatly. Therefore, saturated signals are not removed in the HAWC implementation of the optimal radius calculation.
4.2 Energy assignment method

During the standard HAWC event reconstruction, each event is fit to the NKG/r function (Equation 4.1). The amplitude and modified shower age of this fit are saved in the reconstructed event data so that the signal at $r_{40}$ can be calculated:

$$
\log_{10}(s_{r_{40}}) = \log_{10}(A) - 0.371024s + 0.930696.
$$

(4.4)

The numbers come from replacing $r$ in Equation 4.1 with 40 meters.

The next step is to associate the energy with the signal at $r_{40}$. The basic technique is to use simulation to create a 2D histogram with $\log_{10}(s_{r_{40}})$ on one axis and $\log_{10}(E_{\text{true}})$ on the other, and then find the best fit line between charge and energy:

$$
\log_{10}(E_{\text{est}}) = f(\log_{10}(s_{r_{40}})).
$$

(4.5)

However, recall that the relationship between charge and energy depends on the zenith angle of the primary gamma ray, due to the differing amount of atmosphere showers at different inclinations must travel through before reaching the detector. Therefore, the energy function becomes a function of both charge and zenith angle:

$$
\log_{10}(E_{\text{est}}) = f(\log_{10}(s_{r_{40}}), \theta).
$$

(4.6)

The 2D charge–energy histogram is therefore created in many narrow bands of zenith (chosen so that each bin has an equal number of simulated events) and the best fit values are parameterized by zenith angle. An example of the 2D histogram in a narrow zenith band can be seen in Figure 4.5.

We have no a priori assumptions as to what function the charge vs. energy histogram should be fit to. In keeping with tradition of other air shower-related equations, the fit function is found empirically. The histogram in Figure 4.5 has a representative shape that is seen in all zenith bands. There are two linear segments. The fit is the line of best fit through the bin that is the most populated in each vertically projected slice of the histogram. This was found to give a better energy estimate than using the mean or median of each slice, for reasons that are not entirely clear.

The fit is constrained so that the best fit line is continuous at the point where
Figure 4.5: An example showing the correlation between effective charge and $E_{\text{true}}$ for a small range of zenith angles \((14.77^o < \theta < 15.42^o)\). The black line is the line of the best fit. Note the change in slope around \(\log_{10}(s_{40}) = \sim 0.7\). This change in slope is known as $s_{\text{crit}}$. Colors in the histogram denote the relative probabilities assuming an isotropic source at a declination of 20°. This fit is repeated for every zenith bin so that the best fit parameters can be parameterized by zenith angle.

the slope changes. This point where the change in slope occurs is called the “critical signal” \((\log_{10}(s_{\text{crit}}))\). This constraint gives four free parameters: the slope of the first segment \((m_1)\), the y-intercept of the first segment\((b_1)\), the slope of the second segment\((m_2)\), and the critical signal \((s_{\text{crit}})\). The y-intercept of the second segment is constrained to make the two lines continuous and is equal to \(b_1 + \log_{10}(s_{\text{crit}})(m_1 - m_2)\). The change in slope means that events with $s_{40}$ below $s_{\text{crit}}$ have one functional form for the energy formula, while those with $s_{40}$ above
Figure 4.6: Examples for the parameterization by zenith angle for different fit parameters. All fit functions are chosen empirically. Figure 4.6a shows the parameterization of $b_1$, while Figure 4.6b shows the parameterization of $m_2$. Horizontal error bars are the RMS of the zenith distribution in each bin. Note that the line of best fit for $m_2$ has a sharp change in slope at $\theta = 30^\circ$. It is not clear \textit{a priori} what causes this.

$s_{\text{crit}}$ get another.

The fit parameters are then plotted as a function of zenith angle and parameterized (Figure 4.6). The fit functions are once again chosen empirically and are different for each parameter. For example, $b_1$ is fit very well to a second degree polynomial, while $m_1$ exhibits no trend with zenith angle and is fit to a straight, horizontal line.

The slope of $m_2$ exhibits a sharp change around $\theta = 30^\circ$. It is therefore necessary to give events with a $\theta < 30^\circ$ a different energy assignment than those events with $\theta > 30^\circ$. Coupled with the differing energy assignments depending on whether the signal is above or below the critical signal, there are three possible formulas for the energy, depending on the zenith angle and signal.

The critical signal is also parameterized by zenith angle:

$$
log_{10}(s_{\text{crit}}) = 0.649 - 0.0014\theta.
$$

(4.7)
Finally, the formula for the energy assignments is:

\[
\log_{10}(E) = \begin{cases} 
2.055\theta * \log_{10}(s_{40}) + 2.335 + 0.00184\theta + 0.000448\theta^2 & s_{40} \leq s_{\text{crit}} \\
(0.986 - 0.00303\theta) * \log_{10}(s_{40}) + 3.029 + 0.00231\theta + 0.000444\theta^2 & s_{40} > s_{\text{crit}}; \theta \leq 30^\circ \\
(1.345 - 0.015\theta) * \log_{10}(s_{40}) + 2.796 + 0.0106\theta + 0.000427\theta^2 & s_{40} > s_{\text{crit}}; \theta > 30^\circ 
\end{cases}
\] (4.8)

4.3 Performance on simulation

The performance of the ground parameter has been extensively studied with Monte Carlo simulations. In addition to the standard quality cuts outlined in Section 3.2.2.4, all plots in this subsection were made with a requirement that the core be reconstructed on the array, a zenith angle cut at 45°, and a requirement that the difference between the reconstructed and true zenith angles of the event be < 0.75°. There is also a requirement that > 6.7% of the array is hit, corresponding to the beginning of \( f_{\text{hit}} \) Bin 1.

Only purely gamma ray Monte Carlo simulations have been used (i.e. none of the cosmic ray background is included).

4.3.1 Mixing matrix

The mixing matrix (a histogram of true MC energy vs. estimated energy) can be seen in Figure 4.7.

The mixing matrix will naturally change if a different spectrum is assumed. Figure 4.7 shows what the mixing matrix looks like assuming an isotropic \( E^{-2} \) spectrum. Recall that this is the same spectrum that was used to assign the energies. To compare, Figures 4.8a and 4.8b show what the mixing looks like if the Monte Carlo is weighted to resemble a transiting source with an \( E^{-2} \) or \( E^{-3} \) spectrum, respectively. Both sources transit at a declination of 20 degrees, which is roughly

\footnote{Note that this is a slightly relaxed requirement from the 0.5° cut used in the preceding two sections. This is done because it is not uncommon to have an angular resolution worse than half a degree for the smallest showers.}
overhead for HAWC. Note that lower energy events are more numerous with an $E^{-3}$ spectrum, pushing the mean energy to the lower left-hand corner of the plot.

### 4.3.2 Bias and resolution

Following the example of the MAGIC collaboration [74], bias and resolution are computed by taking the distribution $(\log_{10}(E_{\text{est}}) - \log_{10}(E_{\text{true}}))/\log_{10}(E_{\text{true}})$ in quarter-decade bins in true energy and fitting to a Gaussian. The mean of the Gaussian fit is the bias (reported as a percentage), while the resolution is the standard deviation. This calculation is done in log10 space rather than linear space.
Figure 4.8: The same mixing matrix as shown in Figure 4.7, but with the Monte Carlo weighted to resemble a transiting source with an $E^{-2}$ spectrum (a) and a transiting source with an $E^{-3}$ spectrum (b). Both sources transit at a declination of 20 degrees.
Figure 4.9: The bias and resolution as a function of the true (MC) energy.

used in [74] because these distributions are only symmetric in the HAWC data if calculated in log10 space.

Figure 4.9 shows the bias and resolution as a function of true energy. Note that below 1 TeV, there is a very large, positive bias. This occurs because the threshold cut (6.7% of the array must be hit) removes most of the sub-TeV showers from the sample, leaving only showers with overfluctuations. These showers hit more of the array than the average sub-TeV shower, and they are reconstructed with systematically high energies.

Figure 4.10 shows the bias and resolution as a function of the reconstructed energy.

4.3.3 Bin purity

Bin purity is measure of the contamination of a bin by poorly reconstructed events. It is defined as the fraction of events in a half-decade reconstructed energy bin whose true (MC) energy is also within that bin.

The bin purity for two different simulated sources transiting at a declination of 20 degrees can be seen in Figure 4.11. Note that the bin purity is worse than would
be naively expected from just looking at the bias and resolution, both of which are only a few percent. This can be attributed to the properties of exponential spectra: there are far more lower-energy events and even a small percentage of them being reconstructed with high energies can cause them to fluctuate into another bin. These mis-reconstructed lower-energy events are much more numerous than the truly high-energy events. Note that the bin purity is worse for the $E^{-3}$ spectrum than the much harder $E^{-2}$ spectrum.

Bin purity is also a function of declination. A source transiting directly overhead will be seen at all zenith angles, while a source transiting near the edge of HAWC’s field-of-view will only be seen at high zenith angles. This translates to a higher energy threshold and different biases in the reconstructed energy. Figure 4.12 compares two identical simulated sources with a spectrum of 2.0, but with one transiting at a declination of 20 degrees (roughly overhead) and one transiting at a declination of -15 degrees (corresponding to a spot in the Galactic plane, closer to the edge of the field of view). Note that for most energies, bin purity is better at 20 degrees.

Note that the addition of a cutoff to a power-law spectrum can decrease bin purity. We will return to this point later after fitting real astrophysical sources in
Figure 4.11: The bin purity for two different simulated sources transiting at a declination of 20 degrees. The bin purity is worse for the $E^{-3}$ source because the spectrum is softer meaning that there are far more lower-energy events with upward fluctuations than there are truly high-energy events.

Chapters 5 and 6.

4.4 Systematics in the energy assignment

There are two potential sources of systematic error stemming from the method. The first is the effect of using 40 meters as the optimal radius for every event. The second effect is known as the “broad pulse” effect and comes from mismatch in the arrival time distributions at the PMT between data and Monte Carlo (this effect, called the “late light” in [4], was the largest source of systematic error in the first analysis of the Crab Nebula by HAWC).

4.4.1 Effect of using one optimal radius for all events

Recall from Section 4.1, and from Figure 4.2 in particular, that the distribution of $r_{opt}$ is somewhat broad: $r_{opt}$ is $30.3 \pm 4.6$ for events where $< 25\%$ of the array is hit, and grows to $44.4 \pm 5.9$ for events where more than $75\%$ of the array is hit.
Figure 4.12: The bin purity for an $E^{-2}$ source transiting at a declination of 20 degrees, compared to an $E^{-2}$ source transiting at a declination of -15 degrees.

It would be computationally prohibitive to develop an algorithm for the energy estimate that uses the signal at the true $r_{opt}$ for each event, so 40 meters was chosen as the distance at which the signal is measured for all events. This may introduce a small systematic error from not using the true optimal radius for each event.

To investigate the effect of this, two new energy estimates were determined. One was based on the signal at 30 meters ($s_{30}$) from the shower axis, while the second was based on the signal at 65 meters ($s_{65}$) from the shower axis. These two values are approximately the smallest and largest optimal radii seen in simulation, respectively.

Figures 4.13 and 4.14 show the difference in the standard energy estimate, based on $s_{40}$, and the two new energy estimates. Both histograms have a mean at approximately zero, showing that the choice to measure the signal at 40 meters is not systematically introducing any biases into the measurement.

4.4.2 Laser effect

For reasons that are unknown, the raw PE distribution in HAWC exhibits a fairly large data/MC discrepancy above 50 PEs. The size of this effect suggests that
Figure 4.13: The difference in the standard estimated energy, with the signal measured at 40 meters, and the estimated energy with the algorithm optimized on the signal 30 meters from the shower axis. The mean and RMS of the histogram are -0.04 and 0.09, respectively.

Figure 4.14: The difference in the standard estimated energy, with the signal measured at 40 meters, and the estimated energy with the algorithm optimized on the signal 65 meters from the shower axis. The mean and RMS of the histogram are -0.06 and 0.19, respectively.
simulation may be underestimating the charge at this level by as much as a factor of two. Additionally, one of the biggest sources of systematic error in [4] is how light arriving later in the air shower is treated. Simulation says that much of the light from an air shower should arrive within 10 nanoseconds, but data suggests a broader distribution.

One hypothesis as to the source of this discrepancy stems from the laser used in the calibration of the experiment. Since a typical laser pulse lasts for much less time than a typical air shower, there is a concern that the slewing effect is undercalibrated due to the laser. Widening the length of time the laser pulses last may shed light on this discrepancy. It is also possible to investigate any potential effects on the energy assignment, which is intertwined with the measured charge.

The distribution of the arrival time of PEs at a PMT during an air shower was determined from simulation (Figure 4.15). To reduce the effect of random noise hits, only PMTs with more than 20 hits in an event were used. This arrival time distribution was determined for several different air shower PE levels and for several different PMT distances to the air shower axis. Regardless of the total number of PEs and the distance of the PMT to the air shower axis, the bulk of the light arrives within 5-10 nanoseconds after the first PE arrives at the PMT.

An apparatus to delay the laser pulse by this amount was created at Los Alamos National Laboratory in late 2016. Light leaving a laser was sent to an optical splitter, and some of the light was delayed by adding more fiber for it to travel through. A second splitter was used to combine these fibers back together; the light was then sent to the PMTs. See Figure 4.16 for a schematic of the setup. An oscilloscope was used to verify that the bulk of the light was in fact being delayed by a few nanoseconds.

These fibers were later taken to the site in Mexico, connected to the HAWC calibration system, and incorporated into full calibration runs. Initial processing of the data showed that the ToT-nPE calibration curves (Figure 4.17) did not vary greatly between this modified calibration run and a standard calibration run. (Recall the discussion of the calibration system in Section 3.1.1 for a definition of ToT). The ratio of the high ToT-nPE curves for this study and a standard calibration run varied between 0.95 and 1.05 (Figure 4.18). This suggests that the data/MC discrepancy is not caused by the laser.

There do seem to be variations in what the curves look like between different
Figure 4.15: A sample distribution showing the arrival times of PEs in a PMT. 0 is defined as the time the first PE reaches the PMT. This plot includes all PMTs with a distance from the shower axis between 0 and 40 meters, and all showers having $3 < \log_{10}(nPE) < 3.5$. $nPE$ is the number of PEs.

Figure 4.16: Schematic of the setup for the simulated air shower runs.
Figure 4.17: The average ToT-log nPE curves for all possible combinations of ToT and PMT size: high Tot/8 inch PMT (top left), high Tot/10 inch PMT (top right), low ToT/8 inch PMT (bottom left), and low ToT/10 inch PMT (bottom right). The red data points are the average points for PMTs receiving light from the delayed fibers during the modified calibration run. The blue data points correspond to the average curves from a standard calibration curve. The yellow lines are the fits to curves.

calibration runs taken with this modified setup. Since there was only one set of fibers, data could only be collected for a portion of the array at a time. Data was collected over the course of a few days. This suggests that small variations in calibration may be expected on a short timescale. Typically, calibration runs are only taken after PMTs are taken into or out of service, so it is unclear if this effect is expected. It may be a good avenue for future study.

This study does suggest that the laser used for calibration may introduce an uncertainty of ±5% into the measured charge. Propagating this through to the energy by replacing the \( \log_{10}(s_{40}) \) in Equation 4.8 with \( 0.95\log_{10}(s_{40}) \) and \( 1.05\log_{10}(s_{40}) \) shows a small effect - roughly ± 0.04 in \( \log_{10}(E) \) (Figures 4.19 and 4.20).

Note that the source of the data/MC PE discrepancy still has not been deter-
Figure 4.18: The ratio of the low ToT-log nPE curve (left) and the ratio of the high Tot-nPE curves (right). The ratio is calculated as the curve from the modified calibration run divided by the curve from a standard calibration run. The grey line corresponds to 8-inch PMTs while the black line corresponds to 10-inch PMTs.

Figure 4.19: The difference between the standard energy estimate and the energy estimate obtained when the signal is reduced by 5%.
Figure 4.20: The difference between the standard energy estimate and the energy estimate obtained when the signal is increased by 5%.

mined. While the signal at 40 meters is typically less than 50 PEs, this discrepancy is not expected to affect the energy estimate much. However, it does affect the overall flux levels determined during a spectral fit. This has been incorporated as a systematic error during the spectral fit process (to be discussed in Chapter 5) and is investigated by changing the PE levels in the Monte Carlo.

4.5 Improvements over $f_{\text{hit}}$

The creation of an event-by-event energy estimator allows for numerous improvements to the standard analysis method used by HAWC. Recall that in the first HAWC papers, events were merely divided into 9 $f_{\text{hit}}$ bins based on the fraction of PMTs hit during the air shower. After the spectral fit was done using the likelihood fitter, simulation would be used to determine the mean true MC energy in each analysis bin according to the fitted spectrum. The distribution of true energies in each bin is very broad (see Figure 3.9), as this method does not take into account information about the shower that has an impact on the energy, such as the zenith angle of the air shower.

The biggest improvement is at the highest energies. The highest $f_{\text{hit}}$ bin
functions as an overflow bin in the \( f_{hit} \) analysis. The threshold for this bin is 84% of the array hit, which includes every shower with an energy above \( \sim 20 \) TeV. This leads to only a single data point above 20 TeV being included in the spectral fit, even though this parameter space covers a large part of HAWC’s energy sensitivity.

With the ground parameter, it becomes possible to instead bin events in estimated energy (Figure 4.21). Using a binning scheme of a quarter decade width in log-reconstructed energy space creates 5 bins with true-MC energy peaks above 20 TeV. The energy resolution also becomes much better at all energies; the energy overlap between the bins becomes a much smaller problem.

The ground parameter also gives a more accurate sensitivity estimate. Recall the differential sensitivity plot shown in Figure 3.13. Figure 4.22 is the analogous plot for the ground parameter analysis. The last \( f_{hit} \) bin has been subdivided into many energy bins, clearly showing the recovery of dynamic range at the highest energies.
Figure 4.22: The differential sensitivity of HAWC for the energy analysis. The lines are the sensitivities in each energy bin. A point source with an $E^{-2.63}$ is assumed. Plot made by John Pretz.
Chapter 5  Spectral fits with the energy estimator

This chapter describes how the energy estimation method described in the preceding chapter can be applied to spectral fits that use HAWC data. Since the ground parameter solves the problem of the lack of dynamic range of $f_{hit}$, it can be used to more accurately determine high energy cutoffs.

5.1 A 2D binning scheme for events

One may naively expect that with the advent of an energy estimator, the $f_{hit}$ variable can be abandoned completely. This is not the case. The angular resolution depends strongly on the size of the shower, so $f_{hit}$ plus the energy estimate parameterizes the point spread function better than solely binning in energy (see Figure 5.1). The loss of information about the PSF when binning solely in energy increases the statistical errors on the fit parameters.

Additionally, some parts of the parameter space have a worse signal to background ratio than others. If the events are binned in a 2D binning scheme, the likelihood fit method can recognize those bins and give them less weight in the fit. If only a 1D-energy binning scheme is used, this information is also lost and the test statistic of the fit will be smaller. Studies using HAWC data show that the difference in test statistic between a 2D binning scheme and a 1D binning scheme can be up to a factor of three. Obviously this poses problems for claiming statistically significant detections of weaker sources.

For these reasons, a 2D binning scheme for the events was chosen. The 9 $f_{hit}$...
Table 5.1: The binning scheme used for energy fits.

<table>
<thead>
<tr>
<th>Bin number</th>
<th>Bin low edge (TeV)</th>
<th>Bin high edge (TeV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>0.316</td>
<td>0.562</td>
</tr>
<tr>
<td>b</td>
<td>0.562</td>
<td>1.00</td>
</tr>
<tr>
<td>c</td>
<td>1.00</td>
<td>1.78</td>
</tr>
<tr>
<td>d</td>
<td>1.78</td>
<td>3.16</td>
</tr>
<tr>
<td>e</td>
<td>3.16</td>
<td>5.62</td>
</tr>
<tr>
<td>f</td>
<td>5.62</td>
<td>10.0</td>
</tr>
<tr>
<td>g</td>
<td>10.0</td>
<td>17.8</td>
</tr>
<tr>
<td>h</td>
<td>17.8</td>
<td>31.6</td>
</tr>
<tr>
<td>i</td>
<td>31.6</td>
<td>56.2</td>
</tr>
<tr>
<td>j</td>
<td>56.2</td>
<td>100</td>
</tr>
<tr>
<td>k</td>
<td>100</td>
<td>177</td>
</tr>
<tr>
<td>l</td>
<td>177</td>
<td>316</td>
</tr>
</tbody>
</table>

Figure 5.1: The 68% containment of events as a function of $f_{hit}$ bin and energy, as measured on the Crab Nebula. The number refers to the $f_{hit}$ bin (defined in Table 3.1) and the letter refers to the energy bin (defined in Table 5.1). To ensure a good fit, only bins where the Crab Nebula is $> 3\sigma$ in data are shown. As $f_{hit}$ increases across an energy bin, the angular resolution noticeably improves because the shower size increases. The upturn in the angular resolution above 56 TeV is believed to be because the shower curvature model used during event reconstruction has not yet been updated for the highest energies.
Bins (described in Table 3.1) are subdivided into 12 energy bins (defined in Table 5.1), for a total of 108 bins. Each energy bin spans a quarter decade in log10 space. Combining the information from the two tables, one can see how a 2D binning scheme falls into place. For example, bin 8j contains all of the events with between 74% and 84% of the array hit that have energies between 56 and 100 TeV.

Of course, not all 108 bins are physical. In general, for a grid with $f_{hit}$ on one axis and energy on the other, the physical bins are the bins that lie along the diagonal connecting the low $f_{hit}$/low energy events and the high $f_{hit}$/high energy events. For example, bin 9a, a high $f_{hit}$/low energy bin, is completely unphysical. There are no GeV events in bin 9. However, bin 9k (consisting of events with > 84% of the array hit and energies between 100 and 177 TeV) is physical. Additionally, some bins in the transition region between “high statistics” and “completely unphysical” may have low enough statistics that they are either not modeled well in the Monte Carlo or it is impossible to compute their background correctly.

The exact bins which are expected to contribute to a source’s significance are dependent on the declination of the source. Sources transiting directly overhead of HAWC’s field of view will have a lower energy threshold than those that transit at high zenith angles. This will be discussed further in Section 5.4.1.

In this thesis, energy bins a and b, which consist of events with energies less than 1 TeV, are excluded from spectral fits. These two bins are highly biased by the PMT trigger cut. To fall in $f_{hit}$ bin 1, at least 6.7% of the PMTs must be triggered during an event. This requirement removes the vast majority of events with true energies below 1 TeV, leaving only those with upward fluctuations in the number of PMTs hit. Improvements to the energy estimation algorithm may allow for the recovery of these bins in the future.

## 5.2 Gamma/hadron separation and additional data quality cuts

The same data quality cuts used in the $f_{hit}$ analysis (described in Section 3.2.2.4) are also used in the energy-binned analysis. Two additional data quality cuts are also added.

Events whose reconstructed cores lie off the array are discarded. This cut
removes the greatest number of events from the analysis. The energy estimation algorithm works best when the core is known to within a few meters. Events whose cores lay off the detector are often reconstructed very poorly. While this leads to any given source being observed with less significance, it is mainly poor angular resolution events that are removed.

Additionally, all events with a zenith angle above 45 degrees are removed, as the energy estimator becomes more biased above this energy.

PINCness and compactness, which were described in Section 3.2.2.2, are once again used as the gamma/hadron separation variables. The cut values were optimized in each 2D bin on a simulated Crab-like source \textit{a priori}, before looking at any data. The 2D binning scheme was used for cut optimization because, like angular resolution, there is some dependence on both $f_{hit}$ and energy for the optimal values. There is an explicit requirement that the cuts in each bin have at least a 50% gamma-ray efficiency. The efficiency in a bin may be anywhere from the 50% minimum to nearly 100%.

5.3 Background estimation for low-statistics bins

Recall from Section 3.3 that HAWC uses a method called “direct integration” for background estimation. In this method, the efficiency at a particular point in the sky is convolved with the event rate over a $\sim$2 hour time period. This method is very good at estimating the background when the data rate is high - for example, it removes the effect of the cosmic ray anisotropy in the lower $f_{hit}$ bins.

However, direct integration begins to break down when the event rate in a given bin is low. The data in low-statistics bins is prone to fluctuation, and the direct integration background follows those fluctuations. The result is a background estimate that is not smooth. If the background rate is low enough, there may even be pixels in the HEALPIX skymap where the background estimate is completely unphysical (zero events). This causes problems for HAWC’s likelihood software.

This problem with the background is mainly seen in the highest energy 2D $f_{hit}$/energy bins. $f_{hit}$ bin 9, which is the highest $f_{hit}$ bin, only comprises a small percentage of HAWC data. Subdividing it into 6 or so quarter decade energy bins causes the event rate per bin to fall even further.

To solve this problem, a new background estimation method known as the
Figure 5.2: The local coordinates ($\theta$ and $\phi$) for all events passing gamma/hadron cuts with an energy between 100 TeV and 177 TeV. There are 87,586 such events. A random $\theta/\phi$ pair is pulled from this distribution and used to calculate the background, as described in the text. Note that the two local coordinates are correlated, making the 2D histogram a necessity.

"randomized background method" has been developed. This method is similar to direct integration, but instead of using 2 hours as the standard integration time, the entire dataset is used. This method only works in the low-statistics regime, where the event rate is low and there are not large anisotropies in the background due to the cosmic ray anisotropy.

Computing the background with this new technique is a two-step process. First, a 2D histogram of the local coordinates (theta and phi) in the bin of interest is built (see Figure 5.2 for an example). Then, the background is calculated by randomly pulling a theta/phi pair from this histogram and using that with the time of the candidate gamma-ray event to compute a right ascension and declination on the sky. It is obvious from Figure 5.2 that there is some correlation between the two...
Figure 5.3: The background for all events with energies between 100 TeV and 177 TeV, using the direct integration method. Note that the background is not very smooth. The map is in equatorial coordinates.

local coordinates, making it a necessity to pull the random $\theta/\phi$ pair from this distribution instead of pulling these numbers separately from their individual 1D distributions. This process is repeated many times ($\sim 10,000$) for each event, and then the background map is normalized to the actual number of events in that bin. As is done for the direct integration method, the entire background map is then smoothed by half a degree.

Figure 5.3 shows the background in $f_{hit}$ bin 9/energy bin k (which is all events with energies between 100 TeV and 177 TeV), computed using the direct integration method. This map is not smooth and exhibits many fluctuations. Figure 5.4 shows the background for the exact same events, but computed using the randomized background method. Figure 5.5 shows a histogram of the right ascension of all events within a small declination band along with the number of background events computed using both methods. By eye, one can see that the randomized background method is much smoother but still follows general trends in the data.

This background estimation method has the potential to be systematically biased if the theta and/or phi distributions are not stable in time. These distributions were checked for high energy events, where this algorithm is intended to be used.
Figure 5.4: The background for all events with energies between 100 TeV and 177 TeV, using the new method developed for low statistics bins. The scale is the same as Figure 5.3. Note that the background is now smooth. The map is in equatorial coordinates.

Even though there have been changes in calibration and reconstruction methods, etc. over the livetime of the HAWC detector, the zenith distribution at the highest energies remains remarkably stable. See Figure 5.6 for an illustration.

In the remainder of this dissertation, this new background estimation technique is used when the event rate in a 2D bin is less than 500 events per day. Detailed studies showed that this point is approximately where the direct integration background and the randomized background have the same smoothness.

### 5.4 Likelihood fit

Once the maps have been binned in the 2D $f_{hit}$/energy binning scheme and the background has been re-computed following the new technique as described above, one can determine the spectrum of any source using the likelihood technique described in Section 3.3.
Figure 5.5: Histogram of the right ascension of all events with energies between 100
and 177 TeV and declinations between 15 and 25 degrees. Note that due to the low
statistics, the data is prone to fluctuations, and the direct integration background
follows those fluctuations. The randomized background is much smoother, but still
follows general trends in the data.

5.4.1 Bins used in the fit

First, one must determine which of the 108 2D bins should be used in the analysis.
In an ideal situation, one would find out which bins are physical and use those
bins in every fit. This is not feasible in reality, for a few reasons. Due to HAWC’s
exposure, some of the bins only have data in them of the sky. For example, as one
moves away from declination 20 degrees N (which is directly overhead for HAWC),
the energy threshold slowly increases. This means that the list of bins that are
physical is not consistent across the entire sky. One can immediately visualize an
extreme example where a bin is so unpopulated at a specific declination that the
background estimate in that region is zero. This causes problems for HAWC’s
likelihood software. Additionally, if a bin only has a handful of events, it may be
poorly modeled in the Monte Carlo simulation and adversely affects the spectral fit.

The second concern is one of computational power - a likelihood fit with more
than 100 bins in it takes valuable resources and time, so even if it was possible to
use every bin for all fits, one would not necessarily want to.
Figure 5.6: The zenith angle distribution for events between 100 TeV and 177 TeV, for different time periods. The dataset used in this dissertation has been divided into three epochs with roughly equal livetime. The histograms have been scaled so that the integrals of all three histograms are the same. The zenith distribution is stable over time.

Instead, Monte Carlo is used to \textit{a priori} determine a declination-dependent set of bins that should be used to fit the spectrum of sources. For each $f_{\text{hit}}$ bin, a histogram of the base-10 logarithm of estimated energy is created assuming a transiting source with some spectrum. The central 99\% of events in this histogram is then selected, which has the effect of only cutting out bins that the Monte Carlo says are unphysical at that declination. This histogram is created for several different transiting source scenarios:

- Transiting source with a power law index of 2.0 and no cutoff
- Transiting source with a power law index of 2.0 and an exponential cutoff at 5 TeV
- Transiting source with a power law index of 3.0 and no cutoff
- Transiting source with a power law index of 3.0 and an exponential cutoff at 5 TeV

Figure 5.7 shows a physical depiction of this process.
Figure 5.7: A depiction of how the bins are chosen. This is the simulated estimated energy distribution for all events in $f_{hit}$ bin 6, assuming a source transiting at a declination of -15° with a power law index of 2.0 and no exponential cutoff. The gray shaded band denotes the middle 99% of the events. This corresponds to 2D bins 6f, 6g, 6h, and 6i (energies between 5.62 TeV and 56.2 TeV).

Most HAWC sources have a power law index between these two values, so the union of bins that are important for these four source hypotheses should give us the important bins for any hypothetical source. In general, different source models at the same declination are different by only a few bins, at most. The source declination has a far greater effect on bin selection, since the zenith angle and therefore the energy distribution is declination-dependent. As an example, the bins that are important for declination +20° can be seen in Figure 5.8. Table 5.2 gives the list of bins that are important at every declination.

The choice to keep the bins that contain 99% of the events will be discussed in the systematics section (Section 5.6.1). As mentioned above, bins “a” and “b” are removed in this dissertation.
Table 5.2: Bins to use by declination

<table>
<thead>
<tr>
<th>Dec(°)</th>
<th>Bins to use</th>
</tr>
</thead>
<tbody>
<tr>
<td>-35</td>
<td>not accessible due to zenith angle cut of 45°</td>
</tr>
<tr>
<td>-30</td>
<td>not accessible due to zenith angle cut of 45°</td>
</tr>
<tr>
<td>-25</td>
<td>1d 1e 1f 1g 2e 2f 2g 2h 3e 3f 3g 3h 4f 4g 5h 5i 6g 6h 6i 7g 7h 7i 8i 8j 9j 9k 9l</td>
</tr>
<tr>
<td>-20</td>
<td>1d 1e 1f 1g 2d 2e 2f 2g 3e 3f 3g 4e 4f 4g 4h 5f 5g 5h 6g 6h 6i 7h 7i 7j 8h 8i 8j 9i 9j 9k 9l</td>
</tr>
<tr>
<td>-15</td>
<td>1c 1d 1e 1f 2d 2e 2f 3d 3e 3f 3g 4e 4f 4g 4h 5f 5g 5h 6f 6g 6h 6i 7g 7h 7i 8h 8i 9h 9i 9j 9k 9l</td>
</tr>
<tr>
<td>-10</td>
<td>1c 1d 1e 1f 2c 2d 2e 2f 3d 3e 3f 3g 4d 4e 4f 4g 4h 5e 5f 5g 5h 6f 6g 6h 6i 7g 7h 7i 8h 8i 9h 9i 9j 9k 9l</td>
</tr>
<tr>
<td>-5</td>
<td>1b 1c 1d 1e 1f 2c 2d 2e 2f 3e 3d 3e 3f 4d 4e 4f 4g 5d 5e 5f 5g 5h 6f 6g 6h 6i 7g 7h 7i 8g 8h 8i 8j 9h 9i 9j 9k 9l</td>
</tr>
<tr>
<td>0</td>
<td>1b 1c 1d 1e 1f 2c 2d 2e 2f 3c 3d 3e 3f 4d 4e 4f 4g 5d 5e 5f 5g 5h 6e 6f 6g 6h 6i 7f 7g 7h 7i 8g 8h 8i 8j 9g 9h 9i 9j 9k 9l</td>
</tr>
<tr>
<td>5</td>
<td>1b 1c 1d 1e 1f 2b 2c 2d 2e 2f 3c 3d 3e 3f 4c 4d 4e 4f 4g 5d 5e 5f 5g 5h 6e 6f 6g 6h 6i 7f 7g 7h 7i 8g 8h 8i 8j 9g 9h 9i 9j 9k 9l</td>
</tr>
<tr>
<td>10</td>
<td>1b 1c 1d 1e 2b 2c 2d 2e 2f 3b 3c 3d 3e 3f 4c 4d 4e 4f 4g 5d 5e 5f 5g 5h 6e 6f 6g 6h 6i 7f 7g 7h 7i 8f 8g 8h 8i 8j 9g 9h 9i 9j 9k 9l</td>
</tr>
<tr>
<td>15</td>
<td>1a 1b 1c 1d 1e 2b 2c 2d 2e 2f 3b 3c 3d 3e 3f 4c 4d 4e 4f 4g 4d 5e 5f 5g 5h 6e 6f 6g 6h 6i 7f 7g 7h 7i 8f 8g 8h 8i 8j 9g 9h 9i 9j 9k 9l</td>
</tr>
<tr>
<td>20</td>
<td>1a 1b 1c 1d 1e 2a 2b 2c 2d 2e 2f 3b 3c 3d 3e 3f 4c 4d 4e 4f 4g 4d 5e 5f 5g 5h 6e 6f 6g 6h 6i 7f 7g 7h 7i 8g 8h 8i 8j 9g 9h 9i 9j 9k 9l</td>
</tr>
<tr>
<td>25</td>
<td>1a 1b 1c 1d 1e 2b 2c 2d 2e 2f 3b 3c 3d 3e 3f 4c 4d 4e 4f 4g 5d 5e 5f 5g 5h 6e 6f 6g 6h 6i 7f 7g 7h 7i 8g 8h 8i 8j 9g 9h 9i 9j 9k 9l</td>
</tr>
<tr>
<td>30</td>
<td>1a 1b 1c 1d 1e 2b 2c 2d 2e 2f 3b 3c 3d 3e 3f 4c 4d 4e 4f 4g 5d 5e 5f 5g 5h 6e 6f 6g 6h 6i 7f 7g 7h 7i 8f 8g 8h 8i 8j 9g 9h 9i 9j 9k 9l</td>
</tr>
<tr>
<td>35</td>
<td>1b 1c 1d 1e 1f 2b 2c 2d 2e 2f 3c 3d 3e 3f 4c 4d 4e 4f 4g 5d 5e 5f 5g 5h 6e 6f 6g 6h 6i 7f 7g 7h 7i 8g 8h 8i 8j 9g 9h 9i 9j 9k 9l</td>
</tr>
<tr>
<td>40</td>
<td>1b 1c 1d 1e 1f 2b 2c 2d 2e 2f 3c 3d 3e 3f 4d 4e 4f 4g 5d 5e 5f 5g 5h 6e 6f 6g 6h 6i 7f 7g 7h 7i 8g 8h 8i 8j 9h 9i 9j 9k 9l</td>
</tr>
<tr>
<td>45</td>
<td>1b 1c 1d 1e 1f 2c 2d 2e 2f 3c 3d 3e 3f 4d 4e 4f 4g 5e 5f 5g 5h 6f 6g 6h 6i 7g 7h 7i 8g 8h 8i 8j 9h 9i 9j 9k 9l</td>
</tr>
<tr>
<td>50</td>
<td>1c 1d 1e 1f 2c 2d 2e 2f 3d 3e 3f 3g 4d 4e 4f 4g 4h 5e 5f 5g 5h 6f 6g 6h 6i 7g 7h 7i 8h 8i 8j 9h 9i 9j 9k 9l</td>
</tr>
<tr>
<td>55</td>
<td>1c 1d 1e 1f 2d 2e 2f 2g 3d 3e 3f 3g 4e 4f 4g 4h 5f 5g 5h 6f 6g 6h 6i 7g 7h 7i 8h 8i 8j 9h 9i 9j 9k 9l</td>
</tr>
<tr>
<td>60</td>
<td>1d 1e 1f 1g 2d 2e 2f 2g 3e 3f 3g 4f 4g 4h 5f 5g 5h 5i 6g 6h 6i 7h 7i 7j 8i 8j 9i 9j 9k 9l</td>
</tr>
<tr>
<td>65</td>
<td>not accessible due to zenith angle cut of 45°</td>
</tr>
<tr>
<td>70</td>
<td>not accessible due to zenith angle cut of 45°</td>
</tr>
<tr>
<td>75</td>
<td>not accessible due to zenith angle cut of 45°</td>
</tr>
</tbody>
</table>
5.4.2 Flux points technique

Using the bins determined above, the forward-folded spectrum is determined over the entire energy range of interest. This is described in Section 3.3 and is identical to what was done in earlier papers with the $f_{\text{hit}}$ analysis, such as [4].

Once the best fit source model (and for extended sources, source geometry) is found via the forward-folded likelihood fit, flux points in each energy bin are determined. This is done by fixing all of the fitted parameters except for the normalization to their best-fit value from the overall forward-folded fit. For example, if the best fit model to a source is found to be a power law with an exponential cutoff:

$$\frac{dN}{dE} = \phi_o (E/E_o)^{-\alpha} \exp(-E/E_{\text{cut}})$$  \hspace{1cm} (5.1)

where $\phi_o$ is the flux normalization, $E_o$ is the pivot energy, $\alpha$ is the power-law index and $E_{\text{cut}}$ is the exponential cutoff. $\alpha$ and $E_{\text{cut}}$ would be fixed to their overall best-fit values (the pivot energy is chosen \textit{a priori} and is not allowed to float in any fit). The likelihood fit would then be performed for each group of $f_{\text{hit}}$ bins that contribute to a given energy bin. For example, to get the flux point in energy bin e (3.16-5.62 TeV) for a source at a declination of 20 degrees, $f_{\text{hit}}$ bins 1e, 2e, 3e, 4e, 5e, and 6e would be fitted together. In this fit, the only free parameter is $\phi_o$, and in this manner the flux normalization for each quarter decade energy bin
is determined. It is then very easy to plot the flux point for the spectral energy
distribution by plugging all of the fitted parameters into the formula for \( E^2 dN/dE \).

This method is not a true deconvolution of the spectrum, but it does show
which energy bins do not exhibit good agreement with the overall forward-folded
fit. There is precedent for this method. The Fermi-LAT Collaboration has used a
similar method in the past [75].

5.4.2.1 Location of points on the x-axis

It is not correct to place each flux point in the center of the energy bin because
this can give misleading impressions about trends in the data when the bin is wide.
Given the relatively large energy resolution of water Cherenkov experiments, the
true energy distribution for a given reconstructed energy bin is quite wide, with
tails that extend well beyond the edges of the reconstructed energy bin.

Lafferty and Wyatt’s prescription for where to place data points in wide bins
is used [76]. The data point should be placed where the value of the predicted
function is equal to its mean value over the wide bin. Equation 6 from that paper
gives the formula:

\[
f(x_{lw}) = \frac{1}{\Delta x} \int_{x_1}^{x_2} f(x) dx
\] (5.2)

where \( f(x) \) is the predicted distribution, taken here to be the fitted spectral energy
distribution: \( E^2 dN/dE \). \( x_{lw} \) is the location the data point should be placed at (“lw”
stands for “large width”). \( x_1 \) and \( x_2 \) are the bin edges, and \( \Delta x \) is \( x_2 - x_1 \). Here
we take the bin edges to be the edges of the true Monte Carlo energy distribution
in the reconstructed energy bin, weighted for a transiting source with the overall
fitted spectrum. This formula can then be solved either analytically or numerically
for \( x_{lw} \), depending on the form of \( E^2 dN/dE \).

5.4.2.2 A note about extended sources

Some extended sources may exhibit energy dependent morphology. Before going
through the procedure to get the flux points, one may want to study the observed
extent as a function of energy to determine whether this parameter should be fixed
when determining the flux points or should be allowed to float in each energy bin.
Either are valid options.
5.5 Spectral fit systematics

There are many systematic effects that introduce uncertainties into the overall spectral fit, predominantly related to the modeling of our detector. The largest contributors to the systematics are discussed below. These systematic uncertainties and their effect on the Crab Nebula spectrum will be discussed later in this chapter, in section 5.6.1. Many of these systematic uncertainties were discussed in HAWC’s $f_{\text{hit}}$-based Crab paper [4] but a few are specific to the energy-based analysis.

In general, the effect of a systematic uncertainty is investigated by re-fitting the spectrum of the source with the modeling of the detector slightly changed. For example, charge uncertainty is estimated to be $\sim \pm 15\%$. The spectrum would then be refit under two different assumptions: one where all the charges are increased by 15\% and one where the charges are decreased by the same amount.

**Charge uncertainty:** No two PMTs are exactly the same. It is expected that there is relative differences in PMT detection efficiency from PMT to PMT. Additionally, a single PMT’s measurement may vary for a fixed amount of input light. These two effects are combined into one systematic called “charge uncertainty”. Charge uncertainty is investigated by studying tanks that have downward-going vertical muons in them. Muons are used because they have a known geometry and it is easy to compute at what time light from a muon should arrive at each PMT in a tank. Taking several muon events and plotting the number of PEs a PMT sees gives something known as a “muon peak”. Looking at the width of the muon peak gives a rough estimate of the charge uncertainty. This number is estimated to be approximately 15\%.

**PMT quantum efficiency:** The probability that a photon hitting a PMT surface will be converted into a photoelectron is known as the “efficiency”. It is impossible to accurately know the absolute PMT efficiency because HAWC’s calibration system cannot precisely enough determine the optical path of the calibration light to the PMT. PMT efficiency is estimated by plotting the muon peak (in PEs) against the PE threshold required for the PMT to have a rate of 100 Hz. A linear fit is then performed. This fit gives the estimated PE threshold as a function of the muon peak. If PMT efficiency was known exactly, this would be perfectly linear with no spread. Instead, there is some spread of $\pm 15\%$. When this process is repeated with different rates (500 Hz and 10 Hz), the spread is slightly
larger, so we take ±20% as the spread in PMT quantum efficiencies.

**Time dependence:** It is essentially an impossible task to accurately simulate the layout of the detector at all times. Occasionally water leaks out of the tanks, and PMTs are turned off for maintenance. Very rarely, it becomes necessary to turn off the high voltage to portions of the array for this maintenance. The calibration for a PMT may change after it is repaired, but calibration runs are only performed on a monthly time scale. In order to determine if the spectral fit is affected by the constantly changing detector layout, the dataset is divided into two halves roughly equal in livetime and the spectral fit is performed in both.

**Angular resolution data/MC mismatch:** The 68\% containment radius shows a systematic mismatch between data from the Crab Nebula and the Monte Carlo simulation (see Figure 5.9). The mean of this discrepancy is \(\sim 0.05\) degrees (determined using the Crab Nebula).

The cause of this is not immediately clear, but it is thought to be at least partially because the curvature model used during event reconstruction has not yet been updated to have an energy dependence; it currently only depends on the PMT charge and the distance to the shower core. This systematic uncertainty has been studied by artificially increasing changing the point spread function in the simulation to match what is seen in data.

**Late light effect:** One of the largest sources of uncertainty in the spectral fits is due to a systematic mis-modeling of how light that arrives late in the air shower is treated. Simulation implies that the arrival time distribution of the photoelectrons should largely occur within 10 nanoseconds. However, the raw photoelectron distributions in data show a discrepancy above 50 PEs. It was thought that this is because the calibration laser has a very narrow pulse and air showers are much broader. Recall that this was investigated in Section 4.4.2 by creating a broader laser pulse, but it was found that this does not seem to account for the entire discrepancy. A correction has been added to the simulation to make the PE distribution match what is seen in the data. This correction has been shown to not matter much to the gamma ray observables, such as estimated energy, but it does affect the distributions of the gamma/hadron separation variables. To investigate the effect of this correction on the spectral fit, a new simulation with the correction completely turned off has been created.

**PMT threshold:** The PMT threshold is fixed at 0.55 PEs in the simulation.
Figure 5.9: The 68% containment of events on the Crab Nebula, compared to the value from simulation. Only bins where the Crab Nebula is detected $> 3\sigma$ in data are shown. Note the systematic discrepancy.

so that the single PE peak matches what is seen in data. However, the cosmic ray rate implies that the true threshold may be much lower, possibly as low as 0.3 PEs. A special simulation set with the lowered threshold has been created.

**Bins used in the fit:** Recall the bin selection criteria discussed previously in this chapter (Section 5.4.1). The energy bins that keep 99% of each $f_{hit}$ bin are used in the spectral fit. Varying the number of bins introduces a small, sub-dominant systematic uncertainty. The spectral fit is repeated using all of the bins that have any data in them, as well as using the energy bins that keep only 90% of each $f_{hit}$ bin.

**PMT acceptance model:** Many of the systematic uncertainties discussed above are deeply tied to the PMT acceptance model. This is especially true for the PMT threshold. The PMT acceptance model, which is a probability distribution function of the recorded charge as a function of the position of the PE incident on the face of the PMT, is over ten years old and was originally developed for the Milagro detector. Updating this model is an avenue of active study in the HAWC Collaboration as of this writing. When it is finished, it is expected that the size of many of these systematic uncertainties will decrease dramatically. This will be
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Statistical Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi_0$</td>
<td>$3.10 \times 10^{-13}$ (TeV cm$^2$ s)$^{-1}$</td>
<td>$4.961 \times 10^{-15}$ (TeV cm$^2$ s)$^{-1}$</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>2.732</td>
<td>0.018</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.121</td>
<td>0.011</td>
</tr>
</tbody>
</table>

Table 5.3: The best fit parameters for the Crab Nebula. The fit spans from 1 TeV to 316 TeV in reconstructed energy.

incorporated into future papers but will not be discussed further here.

### 5.6 Application to the Crab Nebula

The flux points procedure described above has been tested on the Crab Nebula, which is the brightest source HAWC sees and is frequently used as a standard candle in TeV gamma-ray astronomy. This measurement uses 837.2 days of data, spanning from June 2015 (when the HAWC's 100th PMT was installed) to December 2017.

Figure 5.10 shows the Crab Nebula in half-decade energy bins, assuming a Crab-like spectrum ($E^{-2.63}$). The Crab is detected in every energy range shown in the figures; therefore we expect to be able to determine flux points across HAWC’s entire energy range.

From simulation, we expect 60.68 gamma rays above 1 TeV per Crab transit. In data, $59.9 \pm 2.1$ gamma rays are observed, matching expectations. Figure 5.11 shows the expected and observed number of events from the direction of the Crab per transit, along with the residual.

The overall spectrum is determined via a likelihood fit, following the analysis chain described in Section 3.3. A log parabola shape is preferred over a simple power law by 13.14$\sigma$:

$$\frac{dN}{dE} = \phi_0 (E/E_0)^{-\alpha - \beta \ln(E/E_0)}.$$  \hspace{1cm} (5.3)

The pivot energy, $E_0$, is the energy where the correlations between the other fit parameters are minimized. It is fixed at 7 TeV here. The values where the likelihood test statistic is maximized can be seen in Table 5.3.

Figure 5.12 shows the flux points determined using the ground-parameter energy method. In the last quarter-decade energy bin, bin $l$ ($177 < E_{\text{reco}} < 316$ TeV), the Crab is not detected. The maximum likelihood test statistic in this bin is only 0.33
(a) $1 \text{ TeV} < E_{\text{reco}} < 3.16 \text{ TeV}$. Maximum significance is $84.91\sigma$.

(b) $3.16 \text{ TeV} < E_{\text{reco}} < 10 \text{ TeV}$. Maximum significance is $82.10\sigma$.

(c) $10 \text{ TeV} < E_{\text{reco}} < 31.6 \text{ TeV}$. Maximum significance is $82.45\sigma$.

(d) $31.6 \text{ TeV} < E_{\text{reco}} < 100 \text{ TeV}$. Maximum significance is $21.39\sigma$.

(e) $100 \text{ TeV} < E_{\text{reco}} < 316 \text{ TeV}$. Maximum significance is $4.12\sigma$.

Figure 5.10: The Crab Nebula as a function of energy. To reduce the number of images, half-decade bins are used instead of quarter decade bins.
Figure 5.11: The Crab excess per transit, along with the residual, which is defined as \((\text{measured-expected})/\text{expected}\). The expected counts are computed from Monte Carlo weighted to resemble the overall fitted spectrum.

\((\sim 0.57\sigma)\). Instead of placing a flux point in this bin, a 95\% upper confidence limit is set using the method of Feldman and Cousins [77].

For comparison, Figure 5.12 also shows the forward-folded band originally published in [4]. This band was obtained using the \(f_{\text{hit}}\) based analysis. The two results agree well. Note that the energy-based spectrum extends much higher in
Figure 5.12: The spectral energy distribution of the Crab Nebula obtained using the ground parameter energy estimate and the flux points method outlined above. The dotted black line is the forward-folded best fit. For comparison, HAWC’s previously published Crab spectrum, a forward-folded fit using the \( f_{\text{hit}} \)-based analysis, is also shown. The red band denotes statistical uncertainties on the previous HAWC fit, while the dark grey band denotes statistical uncertainties on the fit documented in this Chapter.

energy. The last flux point is well past 100 TeV. This is a big improvement over the \( f_{\text{hit}} \)-based analysis, which was only valid until 37 TeV due to the saturation of the array in the last \( f_{\text{hit}} \) bin.

Errors are statistical only. Systematic uncertainties will be discussed in the next section. The y-error bar on each point is simply the error from the likelihood fit on the fitted flux normalization in each energy bin, propagated through to the spectral energy distribution. Since the Crab Nebula is best fit to a log-parabola spectrum, for each energy flux point we have:
\[ E^2 \frac{dN}{dE} = E^2 (N \pm \delta_N) \left( \frac{E}{E_0} \right)^{-\alpha - \beta \ln \left( \frac{E}{E_0} \right)} \]. \quad (5.4)

Recall that \( \alpha \) and \( \beta \) were held constant to their value from the overall global fit when determining the flux points, so there is no error associated with either parameter. Propagating the error on the flux normalization through, the error on each flux point is simply

\[ E^2 \delta_N \left( \frac{E}{E_0} \right)^{-\alpha - \beta \ln \left( \frac{E}{E_0} \right)} \]. \quad (5.5)

The x-error bars are determined by varying the overall fit parameters by their statistical errors and calculating the maximum shift in the calculated x-axis location.

### 5.6.1 Systematics in the Crab spectral fit

The effect of all of the systematic uncertainties discussed previously (Section 5.5) have been evaluated on the Crab Nebula. In general, these systematic uncertainties are a function of energy, with some only contributing at lower energies while others are dominant at the highest energies.

Figure 5.13 shows the percentage by which each systematic uncertainty changes the flux points of the spectral energy distribution as a function of energy. Note that at the highest energies, the time dependent systematic is the largest. This is because the > 100 TeV events are predominantly found in the second half of the dataset. It is unclear at this time whether this is a real physical effect or if it is merely the result of a changing detector layout. Borione et al. [34] noted that the 100 TeV flux from the Crab may be expected to vary on timescales of \( \sim \) 1 year, since the MeV synchrotron component has been shown to vary on this timescale and this same population is involved in the higher-energy inverse Compton scattering. In this study, each half of the dataset has a livetime of slightly more than one year. More data should allow us to determine any possible time-dependence of the Crab Nebula spectrum.

Figure 5.14 shows the amount each flux point shifts on the x-axis for each source of systematic uncertainty. This can be used to determine the uncertainty on the absolute energy scale of HAWC.

Figure 5.15 shows the spectral energy distribution with a shaded systematic
error band around it. The systematic error contains all the flux points obtained for each systematics study, added in quadrature. The previously published $f_{hit}$-based Crab spectrum [4] is also shown with its systematic uncertainty band. Improved modeling of the detector has greatly decreased the size of the systematic uncertainty band with the ground parameter analysis.

The effect of the systematic uncertainties on the index ($\alpha$) and on $\beta$ from the log parabola fit has also been computed. Adding all of the systematic uncertainties in quadrature, the final values are $\alpha = 2.732 \pm 0.018_{stat} - 0.147_{sys} + 0.053_{sys}$ and $\beta = 0.121 \pm 0.011_{stat} - 0.075_{sys} + 0.062_{sys}$.
Figure 5.14: The amount each flux point shifts on the x-axis for each source of systematic uncertainty. This tells us about the overall energy scale of the experiment.

### 5.6.2 Purity of the highest energy bins

This measurement is the highest energy measurement of the Crab Nebula to date. Previously, the highest energy measurement came from the HEGRA experiment [32]. There, the last bin was centered at 75 TeV (covering a range between 56 and 100 TeV) but only had a significance of 2.7σ. Here, the Crab Nebula is 10.2σ between 56 and 100 TeV in reconstructed energy and 4.2σ between 100 and 177 TeV.

Of course, significance in a particular energy bin is essentially meaningless if the energy assignments in that bin are biased and/or affected by feed-down. Recall the concept of “bin purity” from Section 4.3.3. Since astrophysical sources tend to emit following power-law spectra, there are far more lower-energy events than higher-energy ones. If even a small fraction of the low energy events are
Figure 5.15: Identical to Figure 5.12, but with a systematic uncertainty band added. This band contains all fits obtained varying assumptions about the modeling of the detector to investigate the effects discussed in Section 5.5, added in quadrature. For comparison, the previously published $f_{hit}$ based Crab spectrum [4] with its systematic uncertainties is also shown. Improved modeling has greatly decreased the size of the systematic uncertainty band.

As this dissertation is mainly concerned with the highest-energy sources, the bin purity of the highest energy bins for the Crab Nebula has been investigated using Monte Carlo simulations. For events with a reconstructed energy above 56 TeV, $39^{+17}_{-25}\%$ of the events have true energies below 56 TeV. $61^{+25}_{-17}\%$ of the events have true energies above 56 TeV. The uncertainties quoted are systematic and are computed by taking the best spectral fit for each systematic uncertainty, recomputing the bin purity assuming that fit, and adding the differences between that and the nominal best fit in quadrature. Given the relatively high significance of the Crab above 56 TeV ($\sim 11\sigma$), we can conclude that the Crab is seen significantly
above 56 TeV in true energy despite the 60% bin purity.

Above 100 TeV in reconstructed energy, the numbers are a bit worse. $49^{+18}_{-34}$% of the events have true energies below 100 TeV, while $51^{+32}_{-17}$% of events have true energies above this amount. Furthermore, the Crab Nebula is not yet 5$\sigma$ in this bin. Claims of a detection above 100 TeV should be treated with caution until more data is collected.

### 5.6.3 Scientific significance

While the Crab Nebula spectrum was intended to validate the energy estimation method and ensure that it agrees with the $f_{hit}$-based analysis at lower energies, there are scientific implications of detecting the Crab above 56 TeV.

While this measurement is the highest-energy observation of the Crab Nebula to date, it is not a completely unexpected result. Both de Jager and Harding [78] and Atoyan and Aharonian [79] predicted detectable high-energy fluxes from the Crab Nebula, for $> 50$ TeV and $> 30$ TeV respectively, as early at the 1990s. They used magnetohydrodynamic flow models. Atoyan and Aharonian incorporated the effect of the microwave background radiation field into their work, which had been neglected in the past. They found that inverse Compton scattering on the CMB leads to a significant increase at the highest energies, with the flux at 100 TeV being very close to upper limits set by gamma ray experiments in the 1990s. Horns and Aharonian later predicted gamma rays at least up to 80 TeV [80]. They incorporated a recently (at the time) discovered mm-wavelength non-thermal emission region into their predictions. They note that in order to interpret emission above 50 TeV as inverse Compton in origin, the PeV electrons must exist. A predominantly, if not entirely, leptonic origin of emission is expected from the Crab Nebula’s status as a pulsar wind nebula and from observations at lower energies. A comparison of the TeV gamma-ray flux and the lower-energy synchrotron component can help us learn more about the magnetic field of the Crab Nebula [35].

Recall from Section 1.3 that Klein-Nishina effects change the observed inverse Compton gamma ray spectrum starting at about 10 TeV. Deviations from what is expected from Klein-Nishina effects may indicate a hadronic component of the emission. To date, no hadronic component of the Crab Nebula emission has been found, although some have pointed out the possibility that this component may
exist [79,81]. If a hadronic component were to be observed, this would identify the Crab Nebula as a source of Galactic cosmic rays. This question is not addressed here and is a potential avenue for future study.

Now that the energy-based analysis has been validated on the Crab Nebula, the doors have been opened for HAWC to study the highest energies. At the highest energies (above 56 TeV or so) leptonic and hadronic gamma-ray sources have different signatures. Hadronic sources could be responsible for a fraction of the IceCube astrophysical neutrinos [22].

If a hadronic source that extends high enough in energy were found, it could be identified as a PeVatron. The next chapter of this dissertation will document HAWC’s highest-energy catalogs and discuss whether any of these sources can be identified as PeVatron candidates.

High energy observations can also be used to study Lorentz invariance, as discussed by Martinez-Huerta [69]. This will not be discussed further.
Chapter 6

Highest-energy source catalog

The energy estimator described in the preceding two chapters can be used to create a catalog of HAWC’s highest energy sources. Then, their spectra can be examined to see if any of them could be PeVatron candidates.

Section 6.1 of this Chapter describes the method used to create the first catalogs of sources that emit above 56 TeV and 100 TeV in reconstructed energy. These are the highest energy gamma-ray sources ever detected. Section 6.2 gives the spectral fits of each of these sources. Section 6.3 discusses the bin purity of the high energy bins to determine whether these are significant high energy detections. Lastly, Section 6.5 discusses whether these sources can be identified as hadronic or leptonic.

6.1 Catalog search method

Two high-energy source catalogs are created: one of sources emitting above 56 TeV in reconstructed energy and one of sources emitting above 100 TeV in reconstructed energy. The method used to construct these catalogs is identical to the method used to construct HAWC’s all-energy catalog of sources, hereafter called the “2HWC catalog” [11].

The same dataset used to compute the spectrum of the Crab Nebula in the previous chapter is used here, consisting of 837.2 days of livetime between June 2015 and December 2017. The events are again binned in the 2D binning scheme of quarter-decade bins of estimated energy and \( f_{hit} \) (Tables 3.1 and 5.1).

All-sky significance maps at the highest energies are created by moving the location of a theoretical source across the sky and using the likelihood ratio test
described in Section 3.3 to compute a test statistic. The input maps are all the 2D estimated energy/$f_{hit}$ maps that contribute to the given energy range at any declination (i.e. for the $>56$ TeV map bins 7j, 8j, 9j, 9k, and 9l are used, while for the $>100$ TeV map bins 9k and 9l are used).

When making these significance maps, a source with a power-law spectrum with an index 2.0 is assumed:

$$\frac{dN}{dE} = \phi_0 (E/7\text{TeV})^{-\alpha}.$$  \hfill (6.1)

This is a slight difference from the 2HWC catalog, where an index of 2.7 was assumed. This change was made because higher-energy sources are expected to have harder spectra. Recall that a PeVatron candidate should have a power-law index of $\sim2.0$.

These $>56$ TeV and $>100$ TeV significance maps are created for several different assumed source morphologies:

- Point source
- Disk with a radius of 0.5 degree
- Disk with a radius of 1.0 degree.

It is necessary to use different extensions because a very extended source may not be found in the point source search, and vice versa.

Figure 6.1 shows an example high-energy significance map. The Galactic plane can be seen by eye.

The 2HWC catalog construction method is then applied to the significance maps. This method relies on the distribution of test statistics across the sky. For each map, all areas of the sky where the test statistic (TS) is $>25$ are flagged. These are then separated into primary and secondary sources. Primary sources must be separated from their next-closest source by a valley of $\Delta \sqrt{TS} > 2$. The secondary source classification was needed in the 2HWC all-energy catalog since multiple local maxima where often found near each other in the crowded Galactic plane. In this high-energy catalog, no secondary sources are identified.

Table 6.1 shows the sources that are found in the $>56$ TeV catalog source search and Table 6.2 shows the sources found in the $>100$ TeV catalog source search.
Figure 6.1: The HAWC sky above 56 TeV in reconstructed energy. This plot assumes a source spectrum with a power law index of 2.0 and a source morphology of a disk with a radius of 0.5 degrees. Sources in the Galactic plane are clearly visible. The map is in equatorial coordinates.
<table>
<thead>
<tr>
<th>2HWC name</th>
<th>2HWC location (RA, Dec)°</th>
<th>High energy location (RA, Dec)°</th>
<th>Distance</th>
<th>Search</th>
<th>TS</th>
</tr>
</thead>
<tbody>
<tr>
<td>2HWC J0534+220</td>
<td>(83.63,22.02)</td>
<td>(83.54,22.02)</td>
<td>0.09</td>
<td>PS</td>
<td>118</td>
</tr>
<tr>
<td>(Crab)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>(83.54, 21.94)</td>
<td></td>
<td>0.12</td>
<td>0.5</td>
<td>71</td>
</tr>
<tr>
<td>-</td>
<td>(84.07, 22.27)</td>
<td></td>
<td>0.51</td>
<td>1.0</td>
<td>34.2</td>
</tr>
<tr>
<td>2HWC J1825-134</td>
<td>(276.46,-13.40)</td>
<td>(276.50,-13.06)</td>
<td>0.34</td>
<td>PS</td>
<td>120</td>
</tr>
<tr>
<td>-</td>
<td>(276.42, -13.29)</td>
<td></td>
<td>0.12</td>
<td>0.5</td>
<td>146</td>
</tr>
<tr>
<td>-</td>
<td>(276.24, -13.29)</td>
<td></td>
<td>0.25</td>
<td>1.0</td>
<td>141</td>
</tr>
<tr>
<td>2HWC J1837-065</td>
<td>(279.36, -6.58)</td>
<td>(279.84, -5.79)</td>
<td>0.92</td>
<td>PS</td>
<td>35.9</td>
</tr>
<tr>
<td>-</td>
<td>(279.80, -5.60)</td>
<td></td>
<td>1.07</td>
<td>0.5</td>
<td>43.6</td>
</tr>
<tr>
<td>-</td>
<td>(279.62, -5.75)</td>
<td></td>
<td>0.87</td>
<td>1.0</td>
<td>41.3</td>
</tr>
<tr>
<td>2HWC J1844-032</td>
<td>(281.07, -3.25)</td>
<td>(281.03, -3.43)</td>
<td>0.18</td>
<td>0.5</td>
<td>26.9</td>
</tr>
<tr>
<td>-</td>
<td>(280.99, -3.55)</td>
<td></td>
<td>0.31</td>
<td>1.0</td>
<td>31.3</td>
</tr>
<tr>
<td>2HWC J1908+063</td>
<td>(287.05, 6.39)</td>
<td>(286.96, 6.32)</td>
<td>0.11</td>
<td>PS</td>
<td>41.5</td>
</tr>
<tr>
<td>-</td>
<td>(286.92, 6.35)</td>
<td></td>
<td>0.14</td>
<td>0.5</td>
<td>69.3</td>
</tr>
<tr>
<td>-</td>
<td>(286.83, 6.13)</td>
<td></td>
<td>0.34</td>
<td>1.0</td>
<td>69.3</td>
</tr>
<tr>
<td>2HWC J2019+367</td>
<td>(304.94, 36.80)</td>
<td>(304.98, 36.75)</td>
<td>0.06</td>
<td>PS</td>
<td>77.4</td>
</tr>
<tr>
<td>-</td>
<td>(304.81, 36.75)</td>
<td></td>
<td>0.14</td>
<td>0.5</td>
<td>78</td>
</tr>
<tr>
<td>-</td>
<td>(304.89, 37.31)</td>
<td></td>
<td>0.51</td>
<td>1.0</td>
<td>40</td>
</tr>
<tr>
<td>2HWC J2031+415</td>
<td>(307.93, 41.51)</td>
<td>(307.70, 41.26)</td>
<td>0.34</td>
<td>PS</td>
<td>32.2</td>
</tr>
<tr>
<td>-</td>
<td>(307.79, 41.26)</td>
<td></td>
<td>0.29</td>
<td>0.5</td>
<td>36.4</td>
</tr>
<tr>
<td>-</td>
<td>(307.13, 41.21)</td>
<td></td>
<td>0.85</td>
<td>1.0</td>
<td>25.1</td>
</tr>
</tbody>
</table>

Table 6.1: Result of the catalog source search above 56 TeV in reconstructed energy. The names of the sources from the 2HWC catalog and the 2HWC locations are given along with the locations of the peaks of the high energy emission. “Distance” is the distance between the 2HWC catalog location and the high energy catalog location. “Search” denotes which search the source was found in (point source, 0.5 degree extended, or 1.0 degree extended). Some strong sources may be found in multiple searches. “TS” is the test statistic of the high energy emission, assuming a power law with a spectral index of 2.0.
Table 6.2: The same as Figure 6.1, but for the catalog source search above 100 TeV in reconstructed energy.

<table>
<thead>
<tr>
<th>2HWC name</th>
<th>2HWC location (RA, Dec)°</th>
<th>High energy location (RA, Dec)°</th>
<th>Distance</th>
<th>Search</th>
<th>TS</th>
</tr>
</thead>
<tbody>
<tr>
<td>2HWC J1825-134</td>
<td>(276.46, -13.40)</td>
<td>(276.42, -12.98)</td>
<td>0.42</td>
<td>PS</td>
<td>37.1</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>(276.37, -13.40)</td>
<td>0.09</td>
<td>0.5</td>
<td>39.3</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>(275.98, -13.36)</td>
<td>0.47</td>
<td>1.0</td>
<td>41.4</td>
</tr>
<tr>
<td>2HWC J1908+063</td>
<td>(287.05, 6.39)</td>
<td>(286.92, 6.35)</td>
<td>0.14</td>
<td>PS</td>
<td>27.2</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>(286.74, 6.28)</td>
<td>0.33</td>
<td>0.5</td>
<td>41.9</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>(287.14, 6.17)</td>
<td>0.23</td>
<td>1.0</td>
<td>39.9</td>
</tr>
</tbody>
</table>

Figure 6.2: The Crab Nebula above 56 TeV in reconstructed energy. The map assumes a point source morphology. The open circle denotes the location of the Crab in the 2HWC all-energy catalog paper, while the black circle outlined in white is the local maximum of the > 56 TeV test statistic.
Figure 6.3: The Crab Nebula above 100 TeV in reconstructed energy. The map assumes a point source morphology. The open circle denotes the location of the Crab in the 2HWC all-energy catalog paper. The Crab is not detected at 5 sigma above 100 TeV. The maximum significance is $3.98\sigma$ at $(83.45, 21.94)$ degrees.

Figures 6.2 through 6.5 show the highest energy emission. Figure 6.2 shows the Crab Nebula for $E_{\text{reco}} > 56$ TeV, while Figure 6.3 shows the analogous figure for $E_{\text{reco}} > 100$ TeV. Figure 6.4 shows the Galactic plane for $E_{\text{reco}} > 56$ and Figure 6.5 shows the Galactic plane for $E_{\text{reco}} > 100$ TeV. For each of these figures, the open circles are the location of 2HWC catalog sources while the black circles outlined in white are the local maxima in the high energy maps. Maps are point source maps. All high energy sources are labeled, regardless of whether they were found in the point source search or one of the extended searches. If a high-energy source was found in multiple searches, the location is taken from the smallest possible extension the source is found in. This is because several nearby point sources could theoretically be merged into one extended source when the map is smoothed.

Note that the Crab Nebula is not detected significantly for $E_{\text{reco}} > 100$ TeV. The maximum significance near the Crab Nebula is $3.98\sigma$ at $(83.45, 21.94)$ degrees.
Figure 6.4: The Galactic plane above 56 TeV in reconstructed energy. The map assumes a point source morphology. The open circles denote the locations of sources in the 2HWC all-energy catalog paper, while the black circles outlined in white are the local maxima of the > 56 TeV test statistic. The middle portion of the plane (44 < l < 68 is omitted as there are no high energy sources present there.
Figure 6.5: The Galactic plane above 100 TeV in reconstructed energy. The map assumes a point source morphology. The open circles denote the locations of sources in the 2HWC all-energy catalog paper, while the black circles outlined in white are the local maxima of the $>100$ TeV test statistic. The middle portion of the plane ($44 < l < 68$) is omitted as there are no high energy sources present there.
Also note that 2HWC J2019+367 is just below the threshold for detection in the 100 TeV map. This will be discussed in the next section.

All of the highest energy sources are coincident with sources from the 2HWC catalog (defined as less than 0.5 degrees), with one exception. The high energy emission near 2HWC J1837-065 is nearly a degree away from the 2HWC source location. Looking at the 2HWC catalog, one sees significant all-energy emission at the location of the highest energy emission, but it was not tagged as its own source. It is likely that 2HWC J1837-065 is actually two different sources with different energy cutoffs and this was missed in the catalog paper because the all-energy test statistic difference between the two sources ($\Delta \sqrt{T S}$) is not $> 2$. This is a sign that more sophisticated methods may need to be developed for use in the Galactic plane. For the remainder of this dissertation, the high-energy component of this source will be known as J1839-057 and the location of the peak of the high-energy emission will be used as the location when performing spectral fits.

Since all high energy sources are coincident with known HAWC emission, no studies of the false positive rate are required.

This catalog source search was also run on the $>$ 177 TeV map but no sources were significantly detected (Figure 6.6).

### 6.2 Spectra of the highest-energy sources

This section gives simple spectra for the sources listed in the tables above (with the exception of the Crab Nebula, which was already discussed in Chapter 5). Given that all of these sources are located in the Galactic plane, contamination from the Galactic diffuse emission is a potential concern. This emission is not accounted for when determining the spectra. More detailed modeling should be done in the future.

When performing the spectral fits, the source location is fixed to the location given in the 2HWC catalog. The one exception is J1839-057, where the location is fixed to the peak of the high energy emission. This choice was made because the nearest 2HWC source is nearly 1 degree away.

All of the sources in the Galactic plane are best fit to a power-law with a cutoff:

$$\frac{dN}{dE} = \phi_0 \left( \frac{E}{10 \text{TeV}} \right)^{-\alpha} \exp\left(-\frac{E}{E_{\text{cut}}}\right)$$  

(6.2)
Figure 6.6: The Galactic plane above 177 TeV in reconstructed energy. The map assumes a point source morphology. The open circles denote the locations of sources in the 2HWC all-energy catalog paper. No high energy sources are significantly detected.
Table 6.3: The best fits to Equation 6.2 for each source that emits with $E_{\text{reco}} > 56$ TeV. Errors are statistical only. Extension is the Gaussian sigma from the fit. The units of $\phi_o$ are cm$^{-2}$ TeV$^{-1}$ s$^{-1}$. “PL diff” shows how much the power law with a cutoff is preferred over a pure power-law with no exponential cutoff.

<table>
<thead>
<tr>
<th>Name</th>
<th>Fit $\sigma$</th>
<th>Ext. ($\circ$)</th>
<th>$\phi_o \times 10^{-13}$</th>
<th>$\alpha$</th>
<th>$E_{\text{cut}}$ (TeV)</th>
<th>PL diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>2HWC J1825-134</td>
<td>36.0</td>
<td>0.55±0.03</td>
<td>3.11$^{+0.21}_{-0.20}$</td>
<td>2.01±0.06</td>
<td>98$^{+27}_{-21}$</td>
<td>5.1</td>
</tr>
<tr>
<td>J1839-057</td>
<td>36.8</td>
<td>1.01±0.05</td>
<td>3.81$^{+0.34}_{-0.31}$</td>
<td>2.23±0.05</td>
<td>50$^{+11}_{-9}$</td>
<td>7.2</td>
</tr>
<tr>
<td>2HWC J1844-032</td>
<td>25.9</td>
<td>1.29±0.07</td>
<td>2.81$^{+0.29}_{-0.26}$</td>
<td>2.19±0.06</td>
<td>73$^{+22}_{-17}$</td>
<td>5.3</td>
</tr>
<tr>
<td>2HWC J1908+063</td>
<td>32.7</td>
<td>0.64±0.03</td>
<td>1.27±0.11</td>
<td>2.17±0.05</td>
<td>96$^{+32}_{-26}$</td>
<td>4.4</td>
</tr>
<tr>
<td>2HWC J2019+367</td>
<td>28.6</td>
<td>0.29±0.02</td>
<td>0.76±0.08</td>
<td>1.63±0.11</td>
<td>41$^{+10}_{-8}$</td>
<td>7.2</td>
</tr>
<tr>
<td>2HWC J2031+415</td>
<td>18.6</td>
<td>0.42±0.05</td>
<td>0.70$^{+0.13}_{-0.11}$</td>
<td>2.10±0.12</td>
<td>67$^{+40}_{-26}$</td>
<td>4.9</td>
</tr>
</tbody>
</table>

Table 6.4: The systematic uncertainties on the index and cutoff for each high energy source.

<table>
<thead>
<tr>
<th>Name</th>
<th>Ext. ($\circ$)</th>
<th>$\alpha$</th>
<th>$E_{\text{cut}}$ (TeV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2HWC J1825-134</td>
<td>0.55±0.02</td>
<td>2.01$^{+0.30}_{-0.40}$</td>
<td>98$^{+159}_{-57}$</td>
</tr>
<tr>
<td>J1839-057</td>
<td>1.01±0.03</td>
<td>2.23$^{+0.34}_{-0.31}$</td>
<td>50$^{+47}_{-22}$</td>
</tr>
<tr>
<td>2HWC J1844-032</td>
<td>1.29±0.08</td>
<td>2.19$^{+0.18}_{-0.31}$</td>
<td>73$^{+81}_{-35}$</td>
</tr>
<tr>
<td>2HWC J1908+063</td>
<td>0.64±0.02</td>
<td>2.17$^{+0.15}_{-0.24}$</td>
<td>96$^{+66}_{-46}$</td>
</tr>
<tr>
<td>2HWC J2019+367</td>
<td>0.29±0.05</td>
<td>1.63$^{+0.25}_{-0.44}$</td>
<td>41$^{+28}_{-19}$</td>
</tr>
<tr>
<td>2HWC J2031+415</td>
<td>0.42±0.04</td>
<td>2.10$^{+0.26}_{-0.36}$</td>
<td>67$^{+40}_{-33}$</td>
</tr>
</tbody>
</table>

and are seen as extended sources. No detailed modeling of the source morphology is performed. A Gaussian source morphology is assumed for all sources. It is assumed that the sources remain constant in size as a function of energy. Table 6.3 gives the best-fit parameters for each spectral fit, while Table 6.4 gives the statistical errors associated with each fit parameter. To calculate statistical errors, the fit is redone for each of the sources of systematic uncertainty discussed in Section 5.5 and then the results are added in quadrature. Figures 6.7 through 6.12 show the spectral fits for each source, while Figures 6.13 through 6.24 show the model and residual maps for both the overall emission and the $E_{\text{reco}} > 56$ TeV emission. Only flux points that have a significance above 3$\sigma$ are plotted.

When 2HWC J2019+367 is fit to this spectral shape and function, the emission in the 100 < $E_{\text{reco}}$ < 177 TeV bin has a test statistic of 31.6. Recall that in the blind catalog search described above, this source was just below the detection threshold for a > 100 TeV source. Therefore, we conclude that there are three
sources emitting above 100 TeV in reconstructed energy.

### 6.2.1 Extensions at the highest energies

Since all of the sources are best fit to an extended source morphology, it is interesting to see if this extension continues up to the highest energies. Studying the extension can give clues to emission mechanisms. Electrons cool relatively quickly so an extremely extended source may imply hadronic emission as the highest-energy particles are not able to diffuse great distances from the center of a source without cooling. We will return to this discussion in Section 6.5, where we will fit each of these sources to an alternative electron diffusion model and investigate whether particles could diffuse that far within the age of the system.

Figures 6.25 through 6.30 give the extension as a function of energy for all six sources in the Galactic plane. To ensure that there are enough statistics to fit the size, only energy bins where the test statistic is greater than 3\( \sigma \) are plotted. Even with this requirement, the statistical error bars are somewhat large. However, it can be clearly seen that all six sources remain extended at high energies. 2HWC
Figure 6.8: The spectrum of 2HWC J1839-057. The grey and light pink bands are the statistical and systematic uncertainties, respectively.

J1908+063 is the largest above 56 TeV, with the extension from the Gaussian morphological fit remaining essentially unchanged from the lower energies.

To contrast 2HWC J1908+063, many of the sources are dramatically smaller in extent at the highest energies. For example, 2HWC J2031+415 has a Gaussian sigma of 1.7 degrees at $\sim$3 TeV but shrinks to 0.2 degrees above 56 TeV.

6.3 Bin purity

The purity of the highest energy events has been investigated in a manner analogous to the discussion of the Crab Nebula bin purity in Section 5.6.2. The results can be found in Table 6.5 for the events above $E_{\text{reco}} > 56$ TeV and Table 6.6 for events above $E_{\text{reco}} > 100$ TeV. Only sources with significant emission above 100 TeV are included in the second table. A conservative statement can be given that at least half of the observed events are truly high energy, with 60-70% bin purity not unusual. Given that most of these sources are well above $5\sigma$ in the highest energy bins, it is fair to conclude that these truly are high energy detections, as reducing the number of counts by 60% would not decrease the significance by an appreciable
Figure 6.9: The spectrum of 2HWC J1844-032. The grey and light pink bands are the statistical and systematic uncertainties, respectively.

Figure 6.10: The spectrum of 2HWC J1908+063. The grey and light pink bands are the statistical and systematic uncertainties, respectively.
Figure 6.11: The spectrum of 2HWC J2019+367. The grey and light pink bands are the statistical and systematic uncertainties, respectively.

Figure 6.12: The spectrum of 2HWC J2031+415. The grey and light pink bands are the statistical and systematic uncertainties, respectively.
(a) The radial profile of 2HWC J1825-134. This contains all gamma-ray events with reconstructed energies between 1 TeV and 316 TeV.

(b) The significance map for 2HWC J1825-134, assuming the best-fit spectral parameters. Valid for reconstructed energies between 1 TeV and 316 TeV. TeVCat sources are labeled.

(c) The model map for 2HWC J1825-134, for energies between 1 TeV and 316 TeV.

(d) Residual map for 2HWC J1825-134, for energies between 1 TeV and 316 TeV. Note that significant diffuse emission a few degrees from the source is still present.

Figure 6.13: Results for 2HWC J1825-134. These plots show the emission at all energies (1 TeV < $E_{reco}$ < 316 TeV).
(a) The significance map for 2HWC J1825-134, assuming the best-fit spectral parameters. Valid for reconstructed energies between 56 TeV and 316 TeV. TeVCat sources are labeled.

(b) The model map for 2HWC J1825-134, (c) Residual map for 2HWC J1825-134, for energies between 56 TeV and 316 TeV. energies between 56 TeV and 316 TeV.

Figure 6.14: Results for 2HWC J1825-134. These plots concentrate on the high energy emission ($56 \text{ TeV} < E_{\text{reco}} < 100 \text{ TeV}$).
(a) The radial profile of J1839-057. This contains all gamma-ray events with reconstructed energies between 1 TeV and 316 TeV.

(b) The significance map for J1839-057, assuming the best-fit spectral parameters. Valid for reconstructed energies between 1 TeV and 316 TeV. TeVCat sources are labeled.

(c) The model map for J1839-057, for energies between 1 TeV and 316 TeV.

(d) Residual map for J1839-057, for energies between 1 TeV and 316 TeV. Note that 2HWC J1837-065 is not removed.

Figure 6.15: Results for J1839-057. These plots show the emission at all energies ($1 \text{ TeV} < E_{\text{reco}} < 316 \text{ TeV}$).
(a) The significance map for J1839-057, assuming the best-fit spectral parameters. Valid for reconstructed energies between 56 TeV and 316 TeV. TeVCat sources are labeled.

(b) The model map for J1839-057, for energies between 56 TeV and 316 TeV.

(c) Residual map for J1839-057, for energies between 56 TeV and 316 TeV.

Figure 6.16: Results for J1839-057. These plots concentrate on the high energy emission ($56 \text{ TeV} < E_{\text{reco}} < 100 \text{ TeV}$).
(a) The radial profile of 2HWC J1844-032. This contains all gamma-ray events with reconstructed energies between 1 TeV and 316 TeV.

(b) The significance map for 2HWC J1844-032, assuming the best-fit spectral parameters. Valid for reconstructed energies between 1 TeV and 316 TeV. TeVCat sources are labeled. Significant spillover from the much brighter J1839-057/2HWC J1837-065 complex can be seen.

(c) The model map for 2HWC J1844-032. (d) Residual map for 2HWC J1844-032, for energies between 1 TeV and 316 TeV.

Figure 6.17: Results for 2HWC J1844-032. These plots show the emission at all energies (1 TeV < $E_{reco}$ < 316 TeV).
(a) The significance map for 2HWC J1844-032, assuming the best-fit spectral parameters. Valid for reconstructed energies between 56 TeV and 316 TeV. TeVCat sources are labeled.

(b) The model map for 2HWC J1844-032, for energies between 56 TeV and 316 TeV.

(c) Residual map for 2HWC J1844-032, for energies between 56 TeV and 316 TeV.

Figure 6.18: Results for 2HWC J1844-032. These plots concentrate on the high energy emission ($56 \text{ TeV} < E_{\text{reco}} < 100 \text{ TeV}$).
(a) The radial profile of 2HWC J1908+063. This contains all gamma-ray events with reconstructed energies between 1 TeV and 316 TeV.

(b) The significance map for 2HWC J1908+063, assuming the best-fit spectral parameters. Valid for reconstructed energies between 1 TeV and 316 TeV. TeVCat sources are labeled.

(c) The model map for 2HWC J1908+063. (d) Residual map for 2HWC J1908+063, for energies between 1 TeV and 316 TeV.

Figure 6.19: Results for 2HWC J1908+063. These plots show the emission at all energies (1 TeV < \( E_{\text{reco}} < 316 \) TeV).
(a) The significance map for 2HWC J1908+063, assuming the best-fit spectral parameters. Valid for reconstructed energies between 56 TeV and 316 TeV. TeVCat sources are labeled.

(b) The model map for 2HWC J1908+063, for energies between 56 TeV and 316 TeV. for energies between 56 TeV and 316 TeV.

(c) Residual map for 2HWC J1908+063, for energies between 56 TeV and 316 TeV.

Figure 6.20: Results for 2HWC J1908+063. These plots concentrate on the high energy emission ($56 \text{ TeV} < E_{\text{reco}} < 100 \text{ TeV}$).
(a) The radial profile of 2HWC J2019+367. This contains all gamma-ray events with reconstructed energies between 1 TeV and 316 TeV.

(b) The significance map for 2HWC J2019+367, assuming the best-fit spectral parameters. Valid for reconstructed energies between 1 TeV and 316 TeV. TeVCat sources are labeled.

(c) The model map for 2HWC J2019+367.

(d) Residual map for 2HWC J2019+367, for energies between 1 TeV and 316 TeV.

Figure 6.21: Results for 2HWC J2019+367. These plots show the emission at all energies (1 TeV < $E_{reco}$ < 316 TeV).
(a) The significance map for 2HWC J2019+367, assuming the best-fit spectral parameters. Valid for reconstructed energies between 56 TeV and 316 TeV. TeVCat sources are labeled.

(b) The model map for 2HWC J2019+367. (c) Residual map for 2HWC J2019+367, for energies between 56 TeV and 316 TeV. for energies between 56 TeV and 316 TeV.

Figure 6.22: Results for 2HWC J2019+367. These plots concentrate on the high energy emission (56 TeV < $E_{\text{reco}}$ < 100 TeV).
(a) The radial profile of 2HWC J2031+415. This contains all gamma-ray events with reconstructed energies between 1 TeV and 316 TeV.

(b) The significance map for 2HWC J2031+415, assuming the best-fit spectral parameters. Valid for reconstructed energies between 1 TeV and 316 TeV. TeVCat sources are labeled.

(c) The model map for 2HWC J2031+415. (d) Residual map for 2HWC J2031+415, for energies between 1 TeV and 316 TeV.

Figure 6.23: Results for 2HWC J2031+415. These plots show the emission at all energies (1 TeV < $E_{reco}$ < 316 TeV).
(a) The significance map for 2HWC J2031+415, assuming the best-fit spectral parameters. Valid for reconstructed energies between 56 TeV and 316 TeV. TeVCat sources are labeled.

(b) The model map for 2HWC J2031+415. (c) Residual map for 2HWC J2031+415, for energies between 56 TeV and 316 TeV. for energies between 56 TeV and 316 TeV.

Figure 6.24: Results for 2HWC J2031+415. These plots concentrate on the high energy emission (56 TeV < $E_{\text{reco}}$ < 100 TeV).
6.4 Possible source associations

Table 6.7 gives each high energy source along with the nearest non-HAWC source in the TeVCat catalog [10]. The distance between the two sources is also stated, along with the source identification from TeVCat, if available. When computing the distance between the two sources, the best-fit location of the high energy emission is used (given in Table 6.1). The high-energy emission location is taken from the
Figure 6.27: The same as Figure 6.25, but for 2HWC J1844-032.

Figure 6.28: The same as Figure 6.25, but for 2HWC J1908+063.

Figure 6.29: The same as Figure 6.25, but for 2HWC J2019+367.
Figure 6.30: The same as Figure 6.25, but for 2HWC J2031+415.

Table 6.5: The bin purity for $E_{\text{reco}} > 56$ TeV events. In general, anywhere from 1/2 to 2/3 of the events with a reconstructed energy above 56 TeV have true energies above that value according to the Monte Carlo simulation. The remainder of the events are mis-reconstructed lower energy events, which are much more numerous due to the properties of astrophysical spectra. Uncertainties quoted here are systematic.

<table>
<thead>
<tr>
<th>2HWC name</th>
<th>$E_{\text{true}} &lt; 56$ TeV (%)</th>
<th>$56 &lt; E_{\text{true}} &lt; 316$ TeV (%)</th>
<th>$E_{\text{true}} &gt; 316$ TeV (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2HWC J1825-134</td>
<td>$34^{+18}_{-15}$</td>
<td>$65^{+14}_{-18}$</td>
<td>$0.15^{+1.0}_{-0.2}$</td>
</tr>
<tr>
<td>J1839-057</td>
<td>$49^{+13}_{-22}$</td>
<td>$51^{+27}_{-15}$</td>
<td>$0.00^{+0.14}_{-0.00}$</td>
</tr>
<tr>
<td>2HWC J1844-032</td>
<td>$41^{+15}_{-20}$</td>
<td>$59^{+20}_{-15}$</td>
<td>$0.01^{+0.14}_{-0.02}$</td>
</tr>
<tr>
<td>2HWC J1908+063</td>
<td>$36^{+21}_{-21}$</td>
<td>$64^{+20}_{-21}$</td>
<td>$0.04^{+0.23}_{-0.05}$</td>
</tr>
<tr>
<td>2HWC J2019+367</td>
<td>$44^{+19}_{-39}$</td>
<td>$56^{+30}_{-19}$</td>
<td>$0.00^{+0.04}_{-0.00}$</td>
</tr>
<tr>
<td>2HWC J2031+415</td>
<td>$40^{+24}_{-22}$</td>
<td>$60^{+22}_{-23}$</td>
<td>$0.01^{+0.22}_{-0.02}$</td>
</tr>
</tbody>
</table>

Table 6.6: The bin purity for $E_{\text{reco}} > 100$ TeV events. Uncertainties quoted here are systematic.

<table>
<thead>
<tr>
<th>2HWC name</th>
<th>$E_{\text{true}} &lt; 100$ TeV (%)</th>
<th>$100 &lt; E_{\text{true}} &lt; 316$ TeV (%)</th>
<th>$E_{\text{true}} &gt; 316$ TeV (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2HWC J1825-134</td>
<td>$31^{+16}_{-17}$</td>
<td>$69^{+11}_{-16}$</td>
<td>$0.6^{+3.3}_{-0.7}$</td>
</tr>
<tr>
<td>2HWC J1908+063</td>
<td>$47^{+29}_{-31}$</td>
<td>$52^{+31}_{-28}$</td>
<td>$0.2^{+0.9}_{-0.2}$</td>
</tr>
<tr>
<td>2HWC J2019+367</td>
<td>$60^{+23}_{-39}$</td>
<td>$40^{+39}_{-23}$</td>
<td>$0.01^{+0.16}_{-0.01}$</td>
</tr>
<tr>
<td>2HWC name</td>
<td>Nearest TeVCat source</td>
<td>Distance (°)</td>
<td>Source type</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------------------</td>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>2HWC J0534+220</td>
<td>Crab</td>
<td>0.08</td>
<td>PWN</td>
</tr>
<tr>
<td>2HWC J1825-134</td>
<td>HESS J1826-130</td>
<td>0.02</td>
<td>UNID</td>
</tr>
<tr>
<td>J1839-057</td>
<td>HESS J1841-055</td>
<td>0.45</td>
<td>UNID</td>
</tr>
<tr>
<td>2HWC J1844-032</td>
<td>HESS J1843-033</td>
<td>0.31</td>
<td>UNID</td>
</tr>
<tr>
<td>2HWC J1908+063</td>
<td>MGRO 1908+06</td>
<td>0.05</td>
<td>UNID</td>
</tr>
<tr>
<td>2HWC J2019+367</td>
<td>VER J2019+368</td>
<td>0.11</td>
<td>UNID</td>
</tr>
<tr>
<td>2HWC J2031+415</td>
<td>PSR J2032+4127</td>
<td>0.33</td>
<td>Binary</td>
</tr>
</tbody>
</table>

Table 6.7: Each 2HWC source emitting above 56 TeV in reconstructed energy. The nearest non-HAWC TeVCat source and the distance to it is listed. “Source type” is the source identification from the TeVCat catalog: “PWN” is “pulsar wind nebula”; “UNID” is “unidentified”.

The smallest extension map in which the source is significantly detected (i.e. if a source is found in both the 0.5 degree and 1.0 degree extended source searches, the location is taken from the 0.5 degree search). This limits source confusion in the crowded Galactic plane.

In the following section, each individual source in the Galactic plane will be discussed and possible source associations will be identified. As there is no ambiguity that 2HWC J0534+220 is the Crab Nebula, it is not discussed further.

All three of the sources that emit above 100 TeV are located fairly close (0.36 degrees or closer) to pulsars from the ATNF radio pulsar catalog\(^1\) [82] and it has been suggested that they may be TeV halos [16]. This would imply a leptonic origin of the emission. However, all three of these sources are fairly young systems; the characteristic ages range from 17.2-21.4 kiloyears. The authors of [16] state that this may imply significant contamination from the associated supernova remnant.

Of the sources emitting above 56 TeV but not above 100 TeV, 2HWC J2031+415 is considered a TeV halo. This source is much older (181 kiloyears) than the three 100 TeV sources. However, it has a flux at 7 TeV about 12 times brighter than the TeV halo model presented in [16] predicts. 2HWC J1837-065 also has a potential TeV halo identification, although like the > 100 TeV sources it is also a younger system that may have contamination from a supernova remnant. This leaves 2HWC J1844-032 as the only high energy source without a potential identification as a teV halo.

\(^1\)An updated version of the ATNF pulsar catalog can be found at http://www.atnf.csiro.au/research/pulsar/psrcat/
The all-energy emission of 2HWC J1825-134 is located almost directly in between two HESS sources: HESS J1825-137 [83] (identified as a TeV pulsar wind nebula) and HESS J1826-130 (which is unidentified). The high-energy emission appears to be much closer to HESS J1826-130 (0.02 degrees vs. 0.79 degrees). However, both sources are extended and the HESS spectrum of HESS J1826-130 is contaminated by emission from the much brighter PWN J1825-137 [84]. The HESS Collaboration has published reports of gamma rays up to 70 TeV from the pulsar wind nebulae [33]. Therefore, it is not clear at this time if the HAWC emission is originating from the unidentified source HESS J1826-130 or whether it comes from the PWN J1825-137, with there possibly being some offset in the highest energy emission.

The pulsar associated with HESS J1825-137, PSR J1826-1334, has no directly observed SNR, although an Hα rim is present which may be evidence of the progenitor SNR [85]. There are two additional supernova remnants near HESS J1826-130: G018.6-00.2 and G018.1-00.1 (both shell-type supernova remnants) as well as a second pulsar, PSR J1826-1256 [84].

J1839-057: J1839-057 is the high energy source located the furthest from a known TeV source. HESS J1841-055 is ∼0.45 degrees away. However, the HAWC emission is quite extended, and the location of HESS J1841-055 falls within the 3σ HAWC contour. It is possible that this source is HESS J1841-055 and that the highest energy emission is offset from the nominal source value.

HESS J1841-055 is unidentified. There are two SNRs in the region, Kes 73 and G26.6-0.1. There are also three pulsars and an x-ray binary. The main problems in determining which one is responsible for the emission are explained in [86]. For example, the pulsars would need to have an extremely high efficiency to be able to create all of the emission. Kes 73 is at the edge of the emission. G26.6-0.1 alone cannot contribute all of the emission. It is likely that some combination of all of these sources contributes to the TeV emission.

2HWC J1844-032: 2HWC J1844-032 lies in a line of HESS sources. The high-energy location is closest to HESS J1843-033, while the overall 2HWC emission is closer to HESS J1844-030.

HESS J1843-033 [87] is not very well-studied in the TeV band. One possible counterpart is AX J1843.8-0352, a supernova remnant candidate observed in the
X-ray band by *Chandra* [88]. There are no radio counterparts to the source. The most recent HESS analysis of the source states that this source likely consists of several overlapping sources [13].

SNRs Kes 75 and G28.8+1.5 are also located in the vicinity of this source, although offset [87]. It has been suggested that the outer shell of G28.8+1.5 may be interacting with molecular clouds, contributing to this emission [11].

HESS J1844-030 is seen by HESS as a faint point source [13] coincident with a mysterious radio source, PMN J1844-0306 that may be an SNR candidate.

**2HWC J1908+063:** 2HWC J1908+063 is almost certainly associated with MGRO J1908+06, originally discovered by the Milagro experiment [46]. This source is officially unidentified in TeVCat, but some classify it as a PWN since the Fermi-LAT discovery of a radio-quiet gamma-ray pulsar, PSR J1907+0602, nearby [89]. However, VERITAS observations of the source show that it does not look like other pulsar wind nebulae of similar age: the spectrum does not soften with distance from the pulsar, which would be expected from electron cooling [90]. Therefore, the nature of this source is still largely unknown.

The supernova remnant (G40.5-0.5) has been identified [91]. The VERITAS observations of the area noted that the TeV emission extends toward the supernova remnant, but is much larger than the radio SNR. The SNR cannot solely account for the TeV emission unless there is a large amount of dense molecular matter nearby [90].

**2HWC J2019+367:** 2HWC J2019+367 is located in the Cygnus region. This region is very bright in diffuse gamma-rays and is frequently included when discussing possible Galactic cosmic-ray sources. Positionally, the high energy emission is closest to VER J2019+368, which is an unidentified source located in the region of MGRO J2019+37. MGRO J2019+37 was Milagro’s most significant source detection [92]. It is classified as a TeV pulsar wind nebula.

MGRO J2019+37 was later resolved into two unidentified sources by VERITAS: VER 2019+368 and VER 2016+371. VER 2019+368 accounts for the bulk of the emission and is coincident with PSR J2021+3651, as well a star-forming region (SH 2-104). At the time, it was thought that this source is likely the PWN seen by Milagro and powered by the pulsar [93]. VER J2019+368 has recently been further resolved into two source candidates [67]. There have also been additional pulsars observed in the region. The exact source association is still unclear.
No SNR has been observed near 2HWC J2019+367.

**2HWC J2031+415**: 2HWC J2031+415 is also located in the Cygnus region. This region is very crowded, with several reported TeV sources that are very close together and that may be associated with each other. The nearest TeVCat source to the high-energy emission is the binary PSR J2032+4127/MT91 213 system [94].

The all-energy 2HWC emission is closest to TeV J2031+4130, which is a steady source. This is coincident with the Cygnus Cocoon. This GeV emission, originally detected by the Fermi-LAT, is a superbubble of freshly accelerated cosmic rays [95]. ARGO-YBJ has reported the observation of the TeV counterpart of the Cygnus Cocoon [96]. The region has also been called a PWN and has been detected by both air shower experiments (notably Milagro [92]) and IACTs [12]. In general, the flux reported by IACTs is systematically different from the flux reported by air shower experiments, which signals that there may be many different components of the emission.

It has recently been suggested that this galactic cluster region (Cyg OB2) could continuously inject cosmic rays that could reach energies of 1 PeV [29]. Cyg OB2 along with the Westerlund clusters and the Galactic center could explain the flux of Galactic CRs without having to invoke supernova remnants and the problems they bring to PeVatron theory (see Section 1.3.1).

### 6.5 Possible emission mechanisms

In this section, possible emission mechanisms for these sources are discussed. Recall from Section 1.3 that starting around 10 TeV, the leptonic inverse Compton component is suppressed due to the Klein-Nishina effect. However, it is not completely negligible. The high energy detection of the Crab Nebula, an established leptonic source of gamma rays, in the previous chapter is proof.

Subsection 6.5.1 provides alternative spectral fits to an electron diffusion model. This model assumes diffusion from the center of the source. The resulting diffusion constant is then analyzed to determine if particles could diffuse from the center of each of these sources within the electron cooling time. This is a concern since electrons cool relatively quickly and all of these sources remain extended at the highest energies. Unrealistic values for the diffusion constant could imply hadronic emission mechanisms.
If these sources are TeV halos, the diffusion constant near the source is expected to be suppressed by a factor of 100-1000 as compared to nominal interstellar medium values [97]. This is because the cosmic ray gradient creates Alfvén waves that then resonantly scatter the same cosmic ray population, creating cosmic-ray self-confinement.

6.5.1 Diffusion fits

The diffusion model used is identical to the one used in the HAWC Collaboration’s analysis of the PWN Geminga [59] and models the gamma ray spectrum that would result in positrons and electrons diffusing away from a source:

\[
\frac{d^2 N}{dE d\Omega} = \phi_o \left( \frac{E}{20\text{ TeV}} \right)^{-\alpha} \exp\left( -\frac{E}{E_{\text{cut}}} \right) \frac{1.22}{\pi^{3/2} \theta_d(E) (\theta + 0.06 \theta_d(E))} \exp\left( -\theta^2 / \theta_d^2(E) \right).
\]

This formula assumes that the electrons and positrons are continuously being injected from a point source. \( E \) is the energy of the gamma rays, while \( \phi_o, \alpha \), and \( E_{\text{cut}} \) are the usual normalization, index, and cutoff of the power-law with a cutoff fit. \( \theta_d \) is the diffusion angle. This parameter is a function of the energy, the magnetic field, and the CMB energy density. Here, the magnetic field is fixed at 3 microGauss.

Figure 6.31 shows the spectral fits to this model. In general, the fitted parameters for the power-law with a cutoff (\( \alpha \) and \( E_{\text{cut}} \)) do not change much from the Gaussian morphology fit results (Table 6.3). The one exception is the fit to 2HWC J2031+415, which has an exponential cutoff that changes dramatically. For the Gaussian morphology, the exponential cutoff for this source is \( 67^{+40}_{-26} \) TeV. Here, that cutoff changes to \( 108^{+90}_{-50} \) TeV. The cutoffs are still compatible within statistical errors, and may have to do with the change in choice of pivot energy. This source has a slightly higher test statistic for the diffusion model.

Similar to the method used to obtain the spectral flux points, the overall index and cutoff are held constant and the fit is done in each energy bin, fitting only the normalization and \( \theta_d \). Strictly speaking, letting the diffusion angle float should not be necessary for an ideal source. If one assumes a Kolgomorov turbulence model [98], the diffusion constant can be defined as
Figure 6.31: Diffusion fits to each of the highest energy sources. Errors are statistical only.
\[ D(E_e) = D_0 (E_e / 10 \text{GeV})^\delta \]  
\hspace{1cm} (6.4)

where \( \delta \) is fixed to 1/3. Therefore, \( \theta_d \) is expected to change in size with energy and the amount of this variation is known. However, the plots of source extension as a function of energy in Section 6.2.1 show that not all sources vary with energy by the same amount. Most notably, 2HWC J1908+063 does not appear to change in size with energy at all and therefore the choice is made to float \( \theta_d \). In the following section, \( \theta_{d,56} \) refers to the fitted diffusion radius in the 56-100 TeV bin and \( \theta_{d,100} \) refers to the fitted diffusion radius in the 100-177 TeV bin.

The definition of the diffusion coefficient (sometimes called the diffusion constant) is also defined as

\[ r_d = 2 \sqrt{D(E_e) t} \]  
\hspace{1cm} (6.5)

where \( t \) is the time. If the pulsar is a very young system, this will be the age of the pulsar. However, electrons cool somewhat rapidly and in most cases this is the cooling time of the electrons, which is dependent on the electron energy. Since the goal of this section is to determine if the highest energy events could have diffused before the electrons cooled, only the diffusion coefficients at 56 TeV and 100 TeV will be calculated.

From [9], the relation between the electron energy and gamma ray energy in inverse Compton scattering processes is:

\[ < E_e > \approx 17 < E_\gamma > 0.54 + 0.046 \log_{10} ( < E_\gamma > / \text{TeV}) \]  
\hspace{1cm} (6.6)

where \( < E_e > \) is the mean electron energy and \( < E_\gamma > \) is the mean gamma ray energy. 50 TeV gamma rays imply \( \sim 200 \) TeV electrons, and 100 TeV gamma rays imply \( \sim 300 \) TeV electrons. The cooling times for 200 TeV and 300 TeV electrons in a 3 microGauss magnetic field are 5776 years and 4136 years, respectively [99].

\( r_d \) in Equation 6.5 is the diffusion radius. If the distance to the source is known, it is simple trigonometry to convert from the diffusion angle, \( \theta_d \) to \( r_d \):

\[ \theta_d = \frac{180}{\pi} \frac{r_d}{d} \]  
\hspace{1cm} (6.7)

where \( d \) is the distance to the source. Once \( r_d \) and the cooling time are known, it
is trivial to calculate the diffusion coefficient.

Tables 6.8 and 6.9 give the calculated diffusion constants for sources emitting above 100 TeV and 56 TeV, respectively. Diffusion radii are calculated using the distance to the nearest pulsar in the ATNF pulsar catalog\(^{2}\) that has a distance estimate. In the case of 2HWC J1825-134, the diffusion constant is calculated twice: once using the distance to PSR J1826-1334, which is the pulsar closest to the source, and once using the distance to PSR J1826-1256, which is in the vicinity of the unidentified source HESS J1826-130. Recall from Section 6.4 that it is unclear at this time which of two nearby HESS sources is responsible for the highest-energy HAWC emission.

For comparison, the interstellar medium value for the diffusion constant is approximately $3 \times 10^{28} \text{cm}^2\text{s}^{-1} E_{\text{GeV}}^{1/3}$ [97]. This translates to $\sim 1.75 \times 10^{30} \text{cm}^2\text{s}^{-1}$ for 200 TeV electrons and $\sim 2 \times 10^{30} \text{cm}^2\text{s}^{-1}$ for 300 TeV electrons.

It is interesting to investigate whether any of these sources exhibit the suppression in their diffusion constant, as [97] predicts a suppression on the order of a factor of 100-1000 for TeV halos.

We start with 2HWC J2031+415, since it was one of the six sources identified in [16] as having the highest probability of being a TeV halo. There is indeed a suppression in the diffusion constant compared to the ISM here. This suppression is a factor of $69^{+753}_{-44}$. 2HWC J2019+367 is almost certainly a TeV halo as well and therefore leptonic. The suppression for that source is a factor of $482^{+834}_{-243}$. Furthermore, the fitted cutoff for 2HWC J2019+367 is the lowest out of all the high energy sources and is completely consistent with what would be expected from Klein-Nishina effects.

The results for 2HWC J1825-134 change dramatically depending on which pulsar is used. If PSR J1826-1334, which is closest to the all-energy emission is used, there is only a small amount of suppression (factor of $28^{+33}_{-15}$). However, if PSR J1826-1256 is used instead, the suppression is much larger and consistent with a TeV halo hypothesis (factor of $153^{+179}_{-83}$). More modeling is needed to understand the emission mechanisms in this region.

On the contrary, the other three sources do not show any suppression in their diffusion coefficient at all. J1839-057 and 2HWC J1844-032 both have diffusion constants that are roughly the ISM value. The diffusion constant 2HWC J1908+063

\(^{2}\)http://www.atnf.csiro.au/research/pulsar/psrcat/
<table>
<thead>
<tr>
<th>2HWC name</th>
<th>Pulsar Name</th>
<th>Pulsar distance (kpc)</th>
<th>$\theta_{d,100}$ (°)</th>
<th>$r_{d,100}$ (pc)</th>
<th>Diff. constant (cm$^2$ s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2HWC J1825-134</td>
<td>PSR J1826-1334</td>
<td>3.61</td>
<td>0.99$^{+0.47}_{-0.32}$</td>
<td>63$^{+30}_{-20}$</td>
<td>7.1$^{+8.3}_{-3.9}$ $\times$ 10$^{28}$</td>
</tr>
<tr>
<td></td>
<td>PSR J1826-1256</td>
<td>1.55</td>
<td>-</td>
<td>27$^{+13}_{-9}$</td>
<td>1.3$^{+1.5}_{-0.7}$ $\times$ 10$^{28}$</td>
</tr>
<tr>
<td>2HWC J1908+063</td>
<td>PSR J1907+0602</td>
<td>2.58</td>
<td>1.98$^{+0.89}_{-0.60}$</td>
<td>89$^{+40}_{-31}$</td>
<td>1.5$^{+1.6}_{-0.8}$ $\times$ 10$^{29}$</td>
</tr>
<tr>
<td>2HWC J2019+367</td>
<td>PSR J2021+3651</td>
<td>1.8</td>
<td>0.48$^{+0.20}_{-0.19}$</td>
<td>15$^{+6}_{-6}$</td>
<td>4.2$^{+4.4}_{-2.6}$ $\times$ 10$^{27}$</td>
</tr>
</tbody>
</table>

Table 6.8: Diffusion constants for the three sources emitting above 100 TeV in reconstructed energy. The diffusion constant is valid for 300 TeV electrons creating 100 TeV gamma rays only. The uncertainty on the diffusion constant is statistical only and comes from the statistical uncertainty on the fitted diffusion radius value, propagated through to the final result.

<table>
<thead>
<tr>
<th>2HWC name</th>
<th>Pulsar Name</th>
<th>Pulsar distance (kpc)</th>
<th>$r_{d,56}$ (°)</th>
<th>$r_{d,56}$ (pc)</th>
<th>Diff. constant (cm$^2$ s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>J1839-057</td>
<td>PSR J1838-0549</td>
<td>4.06</td>
<td>4.07$^{+2.44}_{-1.32}$</td>
<td>352$^{+100}_{-157}$</td>
<td>1.6$^{+1.2}_{-1.1}$ $\times$ 10$^{30}$</td>
</tr>
<tr>
<td>2HWC J1844-032</td>
<td>PSR J1844-0310</td>
<td>5.97</td>
<td>3.30 $\pm$ 1.4</td>
<td>344$^{+146}_{-146}$</td>
<td>1.6$^{+1.6}_{-1.0}$ $\times$ 10$^{30}$</td>
</tr>
<tr>
<td>2HWC J2031+415</td>
<td>PSR J2032+4127</td>
<td>1.33</td>
<td>1.90$^{+1.24}_{-1.35}$</td>
<td>44$^{+29}_{-31}$</td>
<td>2.6$^{+4.4}_{-2.3}$ $\times$ 10$^{28}$</td>
</tr>
</tbody>
</table>

Table 6.9: Diffusion constants for the three sources that emit above 56 TeV in reconstructed energy but not above 100 TeV. The diffusion constant is valid for 200 TeV electrons creating 50 TeV gamma rays only. The uncertainty on the diffusion constant is statistical only and comes from the statistical uncertainty on the fitted diffusion radius value, propagated through to the final result.
is only suppressed by a factor of $14_{-7}^{+18}$. The emission from these sources may be purely hadronic, or perhaps a mix of leptonic and hadronic.

### 6.5.2 Possible hadronic origins of the emission

Alternatively, we can consider hadronic origins of the emission. Here, we will discuss 2HWC J1908+063. Recall from the previous section that this source has a diffusion constant that is an order of magnitude higher than the other two sources that emit with $E_{\text{reco}} > 100$ TeV.

It is advantageous here to turn to neutrinos. Several of the sources discussed in this Chapter have been identified as potential neutrino emitters for IceCube. For example, Halzen et al. identified 2HWC J1908+06, 2HWC J2019+367 and 2HWC J2031+41 as potential neutrino emitters [31].³ The authors note that a non-observation of J1908+06 in 15 years of IceCube data will rule out the source as an accelerator of cosmic rays. Unfortunately, IceCube’s sensitivity rapidly drops below $\sim 5$ degrees in declination, so it is not possible to use IceCube to look for possible neutrino counterparts of 2HWC J1825-134.

[100] describes a search by the IceCube collaboration for point-like sources in the Northern Hemisphere using muon-neutrino events. While no sources were significantly detected, 1908+06 has the highest post trial p-value out of 34 neutrino candidates that were selected a priori to the search being performed. The p-value is 10.2%.

More modeling is needed to know if any of these sources are hadronic. The most likely candidate at this time appears to be 2HWC J1908+063. This prediction stems from its large diffusion constant and from IceCube observations. More data from neutrino experiments, taken in tandem with the HAWC results presented here, will give a clearer picture of this source.

³In the paper, they are referred to by their Milagro names of MGRO J1908+06, MGRO J2019+37, and MGRO J2031+41 respectively.
Chapter 7  |  Conclusions

7.1 Summary

A new energy estimation technique, known as the *ground parameter*, was presented in Chapter 4. This method, which relies on the PMT signal 40 meters from the core as well as the zenith angle of the shower, assigns energies on an event-by-event basis. This restores dynamic range to events with $E > 20$ TeV; this was missing with HAWC’s previously-published $f_{hit}$ analysis.

This method has been validated on the Crab Nebula. Figure 5.15 shows the flux points obtained using the ground parameter compared to the previously-published $f_{hit}$-analysis spectrum [4]. At the lower energies, the two spectra agree within the systematic uncertainties. The extension past 100 TeV is the highest energy measurement of the Crab Nebula to date. The detection of a flux above 56 TeV confirms theoretical predictions [78–80].

This energy estimation method opens the door for HAWC to study the highest-energy gamma-ray sources, including searches for PeVatrons. Chapter 6 gives the results of HAWC’s first highest-energy catalog. There are seven sources that exhibit significant emission above 56 TeV in reconstructed energy. Six of these sources are in the Galactic plane; the last one is the Crab Nebula. Three sources are significantly detected above 100 TeV in reconstructed energy.

These sources are fit to an electron diffusion model to determine whether a leptonic origin of the gamma ray emission is possible. For the $> 100$ TeV sources, two of them exhibit the suppression in their diffusion constant predicted by [97]. The third, MGRO J1908+06, would have to be diffusing very close to the ISM.
value in order for electrons to diffuse from the center of the source. This source may have a hadronic component to its emission.

### 7.2 The outrigger array and future improvements

The HAWC Collaboration has recently completed an upgrade to its detector. Known as the “outrigger array”, this consists of a sparse array of 350 smaller water Cherenkov detectors that are located outside the main array. Each one of these tanks is 1.55 meters in diameter and 1.65 meters in height, and contains a single 8-inch PMT [101].

This outrigger array will help with the determination of the shower core for large events by breaking the degeneracy between deeply penetrating, high-energy showers with cores far from the array and shallower, lower-energy showers with cores closer to the array. The footprint of these two types of events looks roughly the same in the main array. The instrumented area is 4x bigger than the main HAWC array, and is expected to boost sensitivity above 10 TeV by a factor of two or more (Figure 7.2).
Recall that the energy estimation method detailed above does not use events whose shower cores are reconstructed off the main array because those events have poor angular resolution. Extending the method to incorporate information from the outrigger PMTs is an avenue for future study.

With the increase in sensitivity, it is possible that HAWC will discover more sources emitting at the highest energies soon.
Bibliography


URL http://indico.nucleares.unam.mx/event/4/session/40/contribution/378/material/paper/0.pdf

URL http://arxiv.org/abs/1702.02992

URL http://arxiv.org/abs/1805.05989


URL http://adsabs.harvard.edu/abs/1954ApJ...119....1F


URL http://arxiv.org/abs/1703.09704


URL http://arxiv.org/abs/1005.5480


URL http://arxiv.org/abs/1709.03619


URL http://www.astronomerstelegram.org/?read=9936

URL http://www.astronomerstelegram.org/?read=11194


URL https://pos.sissa.it/301/868/pdf

URL http://adsabs.harvard.edu/abs/1971ICRC....3.1001H


URL http://dx.doi.org/10.1093/mnras/278.2.525


URL http://arxiv.org/abs/0712.1173

URL http://arxiv.org/abs/0710.3528


URL http://arxiv.org/abs/1404.7185

URL http://www.mrao.cam.ac.uk/surveys/snrs/


Vita
Kelly Malone

Selected Papers (in prep)

- Observation of > 100 TeV sources in the Galactic plane with the HAWC Observatory
- The Crab Nebula spectrum extending up to 100 TeV as seen by the HAWC Gamma-Ray Observatory

Education

Pennsylvania State University
Ph.D. in particle astrophysics
2018

University of Massachusetts Amherst
Bachelor of Science in Physics with honors, May 2013
Bachelor of Science in Astronomy, May 2013

Selected Awards and Fellowships

Peter Eklund Memorial Lectureship (Penn State), Honorable Mention, 2017

PA Space Grant Graduate Fellowship (NASA), 2016-2017

Office of Science Graduate Student Research Award (US Department of Energy), 2016

David C. Duncan Graduate Fellowship in Physics (Penn State), 2015-2017

Rising Research Award (University of Massachusetts), 2013