Gamma Hadron Separation using Pairwise Compactness Method with HAWC

Zigfried Hampel-Arias[*][a] **and Stefan Westerhoff**[a] **for the HAWC Collaboration**[b]

[a]*Department of Physics and WIPAC, University of Wisconsin-Madison, Madison, WI, USA*
[b]*For a complete author list, see www.hawc-observatory.org/collaboration/icrc2015.php.*
  Email: zhampel@icecube.wisc.edu

The High-Altitude Water Cherenkov (HAWC) Observatory is a ground based air-shower array deployed on the slopes of Volcan Sierra Negra in the state of Puebla, Mexico. While HAWC is optimized for the detection of gamma-ray induced air-showers, the background flux of hadronic cosmic-rays is four orders of magnitude greater, making background rejection paramount for gamma-ray observations. On average, gamma-ray and cosmic-ray showers are characterized by different topologies at ground level. We will present a method to identify the primary particle type in an air-shower that uses the spatial relationship of triggered PMTs (or "hits") in the detector. For a given event hit-pattern on the HAWC array, we calculate the mean separation distance of the hits for a subset of hit pairs weighted by their charges. By comparing the mean charge and mean separating distance for the selected hits, we infer the identity of the event's primary. We will report on the efficiency for identifying gamma-rays and the performance of the technique with simulation.

---

[*]Speaker.

---

1. Introduction

Gamma-ray observations at GeV-TeV energies are obtained by ground-based experiments in two categories. Imaging air-Cherenkov telescopes (IACTs), including VERITAS, MAGIC, and H.E.S.S., have a low duty cycle (~10%) but excellent sensitivity over a narrow field-of-view (<5°). Air-shower arrays, including Milagro, ARGO-YBJ, and HAWC, have comparatively lower instantaneous sensitivity but have a nearly continuous duty cycle and wide field-of-view (~2*sr*) of the overhead sky. This advantage gives air-shower arrays the ability to make unbiased surveys of the GeV-TeV gamma-ray sky, and will complement IACT observations at higher energies (up to 100 TeV), with greater sensitivity for extended sources.

The challenge for air-shower arrays to observe gamma-ray sources is the large background of hadronic cosmic rays. Thus, the identification and rejection of this background is paramount. Fortunately, the nature of the primary particle determines the subsequent shower development, and provides air-shower arrays such as HAWC a manner in which to discriminate event types.

Gamma-ray induced showers are nearly purely electromagnetic, which develop primarily via pair-production and bremsstrahlung. Such showers are characterized by compact, smooth lateral shower profiles. Hadronic showers are dominated by large fluctuations caused by nuclear interactions, and subsequent generation and decay of pions. These sub-showers and further production of muons provide more transverse momentum as compared to gamma-ray showers. This results in hadronic showers generally having larger signals farther from the shower's principal axis, or core (fig. 1a, 1b). We exploit this difference to suppress the hadronic background.

2. The HAWC Observatory

The High-Altitude Water Cherenkov Observatory, or HAWC, is an air-shower array optimized for the detection of gamma-ray showers. HAWC is located at 4100 m above sea level on the slopes of Sierra Negra, Mexico, and comprises a $22,000$ m^2 densely packed array of 300 water Cherenkov detectors (WCDs). Each WCD is a light-tight steel tank 4.5 m high and 7.3 m in diameter containing 200 kL of purified water and four upward-facing photomultiplier tubes (PMTs). In its now completed configuration, HAWC is most sensitive to high-energy gamma-rays and cosmic-rays in the energy range 100 GeV – 100 TeV.

When an incident primary particle interacts in the Earth's atmosphere, a subsequent air-shower develops, containing many lower-energy charged particles and gamma-rays. As the secondary particles pass through the WCDs, they emit Cherenkov light, which is then detected by the PMTs. By measuring the timing, charge, and spatial information of the triggered PMTs in an air-shower, the arrival direction, energy, and type of the primary particle can be identified.

As shown here, topological event information can be used to discriminate the large background of hadronic cosmic-ray showers (10 kHz event rate) from the electromagnetic air-showers produced by gamma-rays. Suppressing this background, we can conduct a variety of studies including producing sky maps of extended and point-like gamma-ray sources [1, 2], performing dark matter searches [3], and measuring the $e^- + e^+$ flux [4]. While this study illustrates one such gamma/hadron separation technique, another method in development can be found in [5]. Further

details on the operation of the HAWC detector and the event reconstruction are given in [6]. We use the 250 tank configuration of the detector (HAWC-250) for this study.
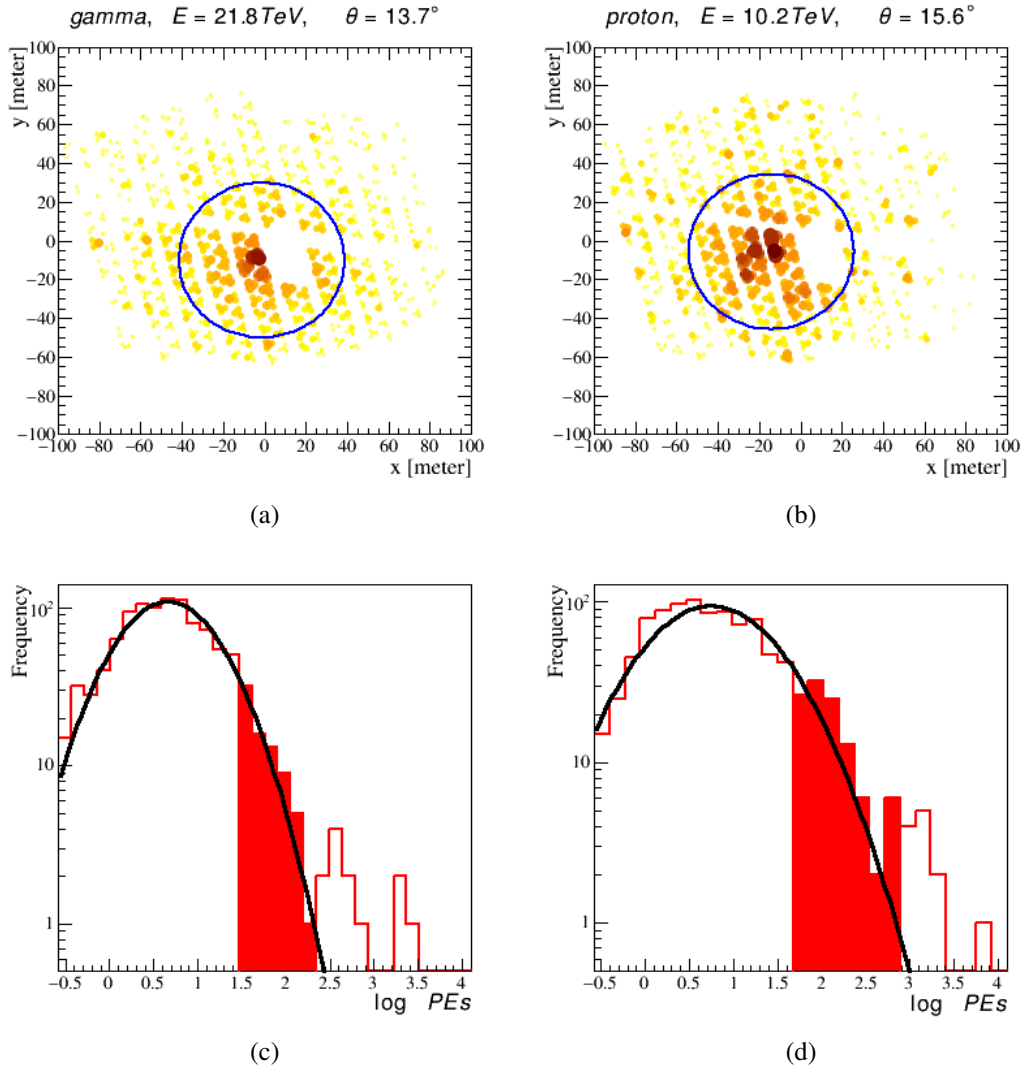

Figure 1: Comparison of gamma (a) and proton (b) events of similar energy in HAWC. The saturation and size of the dots is proportional to the signal strength at that location. Note the gamma event's well-defined core region and smooth lateral distribution beyond. Despite having a core region, the proton event is relatively dispersed and scattered. The blue circle corresponds to the 40 m radius used in the compactness method described in section 3.1. The signal distributions (in units of photoelectrons, or PEs) for each shower are show in (c) and (d). The pair compactness hit selection, S (eq. 3.2), corresponds to the shaded region. The largest signals corresponding to the core region are located beyond S.

3. Gamma/Hadron Separation Methods

3.1 Compactness

One of the simplest gamma/hadron separation techniques used in HAWC analysis is *compactness* [6, 7]. The method involves identifying the largest PMT signal, or hit, beyond 40 m from the fit shower core (*CxPE*40), and dividing the total number of triggered PMTs (*nHit*) by this value:

$$compactness = \frac{\text{nHit}}{\text{CxPE40}}, \tag{3.1}$$

where *CxPE*40 is measured in photoelectrons (PEs).

This measure has the intended effect of comparing detected air-shower events with similar sizes (*nHit*) and qualifying their gamma- or hadronic-like nature by exploiting fluctuations well outside the core region. As illustrated in fig. 1a, 1b, the choice of 40 m ensures that the core region, and its high concentration of large hits, is excluded from *CxPE*40. Hadronic showers, which tend to exhibit larger *CxPE*40 values than gamma showers, will be characterized by lower values of compactness than gamma showers. Compactness can only be calculated when a fit to the core position has been successful.

3.2 Pairwise Compactness

The *pair compactness* method is an extension of compactness in that it involves measuring the spatial connectivity of a selection of hits. Where compactness identifies the relationship of the largest hit beyond 40 m from the core, the pair compactness algorithm identifies the mean charge and pairwise spatial separation for a subset of hits.

The hit selection is made by observing that the distribution of the logarithm of the charges, $\{q\}$, in detected air-showers follows a nearly Gaussian distribution (fig. 1c, 1d). Using the Freedman-Diaconis Rule for optimal binning [8], we fit a Gaussian to $\{\log q\}$ and identify the mean and standard deviation ($\bar{q}, \sigma_q$). If the fit fails, we simply extract $\bar{q}$ and σ_q from the binned distribution. Those hits within the range 1.5σ and 3σ from $\bar{q}$ comprise the n elements of the pair compactness hit selection, S:

$$S = \{i | \bar{q} + 1.5\sigma_q < q_i < \bar{q} + 3\sigma_q\}, \tag{3.2}$$

$$n = \{i | q_i \in S\}. \tag{3.3}$$

The mean charge of this sub-range is calculated and identified as q_{pc},

$$q_{pc} = \frac{1}{n} \sum q_i. \tag{3.4}$$

The spatial relationship is quantified as the weighted mean of the separation of hit pairs in the hit selection, r_{pc} (eq. 3.5), where the weight assigned to each pair is the larger charge of the pair (eq. 3.6).

$$r_{pc} = \frac{1}{\sum\limits_{i \neq j} w_{ij}} \sum_{i \neq j} w_{ij} |r_i - r_j|, \tag{3.5}$$

4

where

$$w_{ij} = \max q_i q_j. \tag{3.6}$$

Similar to compactness, the pairwise compactness method aims to compare showers of similar size, but does so with the mean charge q_{pc} of S in lieu of $nHit$. Thus, the physical size of a shower represented by the mean separation distance, r_{pc}, is compared to showers of similar q_{pc}. In general, we expect hadronic showers to have greater r_{pc} for a given q_{pc}, compared to gamma showers. Based on the distribution of q_{pc} and r_{pc} for gamma showers (fig. 2), we make a fit of the form

$$y = a + \frac{b}{x-c} + d \cdot x, \tag{3.7}$$

then cut away all events that fall in the range $\log r_{pc} > y$, as shown in fig. 2. In this manner we identify gamma-like events with the pair compactness technique.

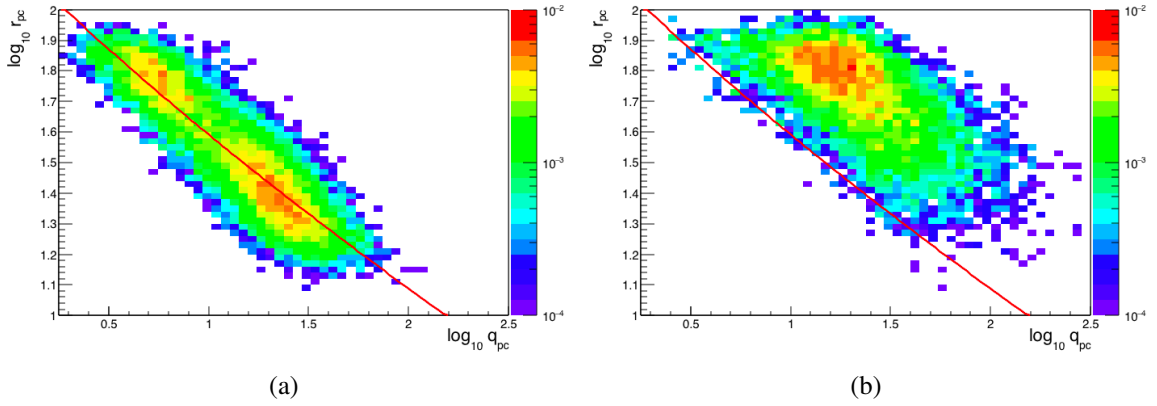(a)   (b)

Figure 2: Normalized pair compactness distributions for gammas (a) and protons (b) for analysis bin 4, as defined in table 1. The curve which maximizes the Q-factor (section 4.1) is shown in red.

4. Comparison of Methods

4.1 Simulation of HAWC

To compare the performance of the separation methods, we simulate the HAWC detector response to gamma and hadronic air-showers. For an incident primary particle, the air-shower produced when it interacts in the atmosphere is simulated using the CORSIKA [9] package, with FLUKA [10, 11], and QGSJET II [12] packages as the particle physics interaction models. The HAWC detector response to the secondary particles is simulated using GEANT 4 configured to the HAWC detector configuration, tanks, and photomultipliers. For this study, simulation of gamma, proton, ^4He, ^{12}C, ^{16}O, ^{20}Ne, ^{24}Mg, ^{28}Si, and ^{56}Fe primaries was used to optimize and subsequently compare the separation capabilities of compactness and pair compactness.

We quantify the performance of the gamma/hadron rejection methods by calculating its Q-factor as defined here:

$$Q = \varepsilon_\gamma / \sqrt{\varepsilon_{CR}}, \tag{4.1}$$

where ε_γ and $\varepsilon_{\mathrm{CR}}$ are the efficiency to identify simulated gamma-rays and cosmic rays, respectively. In this study, we adjust the cut values of the methods scanning for the maximal Q-factor. We also require $\varepsilon_\gamma > 0.5$ to retain as many signal events as possible after cuts.

4.2 Analysis Bins

The current HAWC analysis defines a set of *nHit* ranges in which cuts to compactness are optimized for the detection of the simulated transit of a Crab-like source. Included in the optimization of Q-factors for compactness are further cuts to goodness-of-fit results from the core and angular reconstruction. These additional cuts are not included in the optimization of Q for pair compactness which is done in the *nHit* ranges defined in table 1. For each bin, we calculate the Q-factors for compactness and pair compactness, allowing a comparison of the two methods.

Bin	*nHit* Range
0	38 - 46
1	47 - 70
2	71 - 111
3	112 - 171
4	172 - 253
5	254 - 353
6	354 - 466
7	467 - 580
8	581 - 679
9	680 - 865

Table 1: Ranges of *nHit* defining analysis bins.

4.3 Comparison of Methods

Figure 3 shows the Q-factor comparison of compactness and pair compactness for each analysis bin. Between bins 0 and 4, pair compactness has higher separation capability, with nearly matching Q-factors in bin 5. This may be due to the difficulty in event reconstruction, especially in the lowest *nHit* bins. Both low-energy primaries landing on the array and high-energy primaries landing off the array produce lower-*nHit* events which are difficult to fit to the steeply falling lateral distribution functions used in core fitting.

Beyond bin 6, compactness demonstrates significantly higher Q-factors. The decreased performance in pair compactness is caused by the saturation effect of the core region into S. Though the pair compactness algorithm attempts to exclude the core in the hit selection as shown in fig. 1, for high *nHit* showers, the core region can be large enough in the number of hits to leak into S, dominating the weighting factors in the calculation of r_{pc}. The result is the convergence of the gamma-ray and cosmic-ray pair compactness r_{pc} distributions, i.e., the two types of showers begin to look similar at high *nHit*. Of course, the uncertainties in the event reconstruction drop as *nHit* increases, contributing to the improved gamma/hadron separation from compactness in the higher analysis bins.
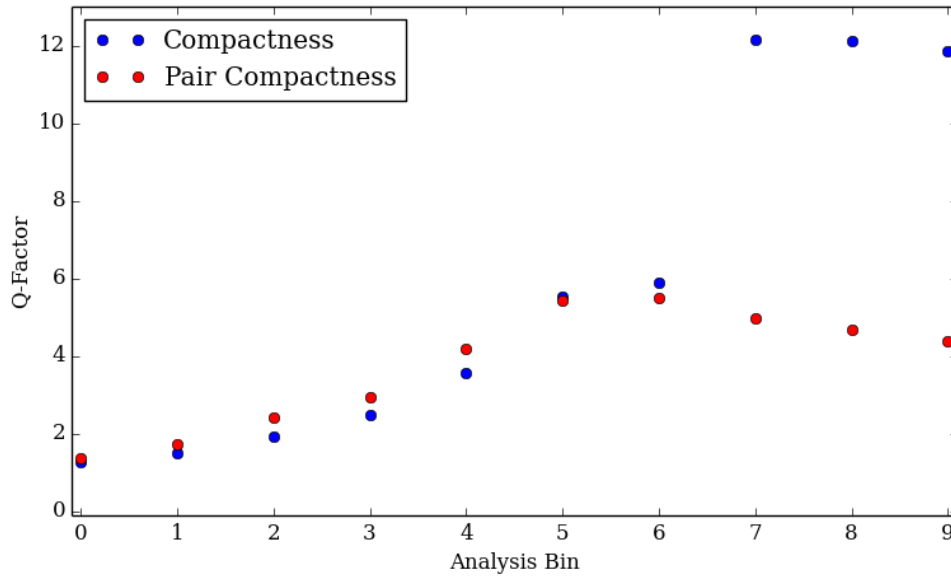
Figure 3: Q-factor comparison of compactness and pair compactness for HAWC-250 analysis bins. In the higher analysis bins, the core region takes over the pair compactness calculation, reducing its efficacy.

5. Conclusion

HAWC is a high-uptime air-shower array that is ideal for the study of both point-like and extended gamma-ray sources. By employing algorithms that exploit topological differences between signal and background events, we can increase the sensitivity of HAWC to these sources, as well as extend HAWC science capabilities to a variety of other studies. The pair compactness technique is a promising gamma/hadron separation method, especially for lower *nHit* events, where there is an abundance of data and difficulties in fitting the shower parameters such as the core location required by compactness. Further improvements to the performance of pair compactness may be possible by identifying an improved hit selection, *S*, and by employing more sophisticated weighting schemes.

Acknowledgments

References

[1] **HAWC** Collaboration, H. Ayala, *Fermi Bubbles with HAWC*, in *Proc. 34th ICRC*, (The Hague, The Netherlands), August, 2015.

[2] **HAWC** Collaboration, M. Hui, *HAWC Observation of Supernova Remnants and Pulsar Wind Nebulae*, in *Proc. 34th ICRC*, (The Hague, The Netherlands), August, 2015.

[3] **HAWC** Collaboration, P. Harding, *Dark Matter Annihilation and Decay Searches with the High Altitude Water Cherenkov (HAWC) Observatory*, in *Proc. 34th ICRC*, (The Hague, The Netherlands), August, 2015.

[4] **HAWC** Collaboration, S. Benzvi, *Measuring the e+ e- Flux Above 1 TeV with HAWC*, in *Proc. 34th ICRC*, (The Hague, The Netherlands), August, 2015.

[5] **HAWC** Collaboration, T. Capistran, *New method for Gamma/Hadron separation in HAWC using neural networks*, in *Proc. 34th ICRC*, (The Hague, The Netherlands), August, 2015.

[6] **HAWC** Collaboration, A. Smith, *HAWC: Design, Operation, Reconstruction and Analysis*, in *Proc. 34th ICRC*, (The Hague, The Netherlands), August, 2015.

[7] A. Abeysekara et al., *Sensitivity of the High Altitude Water Cherenkov Detector to Sources of Multi-TeV Gamma Rays*, *Astropart.Phys.* **50-52** (2013) 26–32, [arXiv:1306.5800].

[8] P. Freedman, D; Diaconis, *On the histogram as a density estimator: L2 theory*, *Probability Theory and Related Fields* **57 (4)** (1981) 453–476.

[9] D. Heck, G. Schatz, T. Thouw, J. Knapp, and J. Capdevielle, *CORSIKA: A Monte Carlo code to simulate extensive air showers*, .

[10] G. Battistoni et al., *The FLUKA code: Description and benchmarking*, in *Proceedings of the Hadronic Shower Simulation Workshop 2006*, vol. 896 of *AIP Conference Proceedings*, p. 31, 2007.

[11] A. Ferrari, P. Sala, A. Fasso, and J. Ranft, *FLUKA: a multi-particle transport code*, . CERN-2005-10 (2005), INFN/TC_05/11, SLAC-R-773.

[12] **GEANT4** Collaboration, S. Agostinelli et al., *GEANT4: A Simulation Toolkit*, *NIM* **A506** (2003) 250–303.