Gamma/hadron discrimination at high energies through the azimuthal fluctuations of air shower particle distributions at the ground

R. Conceição$^{a,b}$ L. Gibilisco$^{a,b}$ M. Pimenta$^{a,b}$ B. Tomé$^{a,b}$

$^a$Laboratório de Instrumentação e Física Experimental de Partículas (LIP),
Av. Prof. Gama Pinto, 2, P-1649-003 Lisbon, Portugal

$^b$Departamento de Física, Instituto Superior Técnico (IST),
Universidade de Lisboa, Av. Rovisco Pais 1, 1049-001, Lisbon, Portugal

Abstract. Wide field-of-view gamma-ray observatories must fight the overwhelming cosmic ray background to identify very-high-energy astrophysical gamma-ray events. This work introduces a novel gamma/hadron discriminating variable, $LC_m$, which quantifies the azimuthal non-uniformity of the particle distributions at the ground. This non-uniformity, due to the presence of hadronic sub-showers, is higher in proton-induced showers than in gamma showers. The discrimination power of this new variable is then discussed, as a function of the air shower array fill factor, in the energy range 10 TeV to 1 PeV, and compared to the classical gamma/hadron discriminator based on the measurement of the number of muons at the ground. The results obtained are extremely encouraging, paving the way for the use of the proposed quantity in present and future large ground-array gamma-ray observatories.
1 Introduction

The recent detection of gamma rays with energies up to the PeV [1, 2] has opened a new window to the extreme energy Universe. Due to the low fluxes and the high background charged cosmic ray, such detection is only possible at ground-based gamma-ray observatories with large surface areas (of the order of higher than $1\text{ km}^2$) and able to discriminate, with very high efficiency, hadron showers from gamma showers.

At high energies (above tens of TeV), high background rejection factors may be reached by studying the distribution of the particles at the ground as a function of the distance to the shower core, identifying the existence of energetic sub-showers, or measuring the steepness, compactness or bumpiness of the Lateral Distribution Function (LDF) [3–5], or, in the case of Cherenkov telescopes, the differences on the shower longitudinal development [6, 7]. On the other hand, the measurement of the number of muons arriving at the ground [8–10] is, whenever possible, one of the best discriminators. The measurement of muons can be done by absorbing the electromagnetic component of the shower by shielding the detectors using earth (e.g. [11–13]), water (e.g. [14], [15]), concrete or some other inert material. Alternatively, muons might be detected by studying, in detectors with several light sensors, the differences in time and/or intensity of the collected signals (e.g. [16–18]). In the end, a global rejection factor of the order or higher than $10^4$ should be achieved.

In this article, we explore the azimuthal non-uniformity of the particle distributions at the ground, introducing a new variable, $C_k$. This quantity is computed in successive circular rings centred at the shower core with a radius $r_k$ in the shower transverse plane. From the $C_k$ distribution as a function of $r_k$ is then defined a new gamma/hadron discriminating variable, $LCm$, being simply the value of the log($C_k$) distribution at a given $r_k$. The paper is organised as follows: the simulation sets used to evaluate all the findings presented in this paper are described in section 2; the variable $C_k$ is introduced in section 3, while the discriminator quantity $LCm$ is presented and discussed in section 4. There, the discrimination power of $LCm$ is tested for different array fill factors ($FF$) in the energy range from 10 TeV to 1 PeV; in section 5 the correlation of this new variable with the logarithm of the number of muons at the ground is analysed; finally, in section 6 the use of this new variable in the present and future large ground array gamma-ray observatories is discussed.
2 Simulation sets

CORSIKA (version 7.5600) [19] was used to simulate gamma-ray and proton-induced vertical showers assuming an observatory altitude of 5200 m a.s.l. The shower energy ranged from 10 TeV up to 2 PeV, being generated with an $E^{-1}$ energy spectrum. FLUKA [20, 21] and QGSJET-II.04 [22] were used as hadronic interaction models for low and high energy interactions, respectively.

A ground detector array was emulated by a 2D-histogram with cells with an area of $\sim 12 \text{ m}^2$ covering all the available ground surface with a fill factor equal to one (FF=1). Fill factors smaller than one were obtained by masking the 2D-histogram with regular patterns.

Each cell represents a station. The signal in each station was estimated as the sum of the expected signals due to the particles hitting the station, using dedicated parameterizations as a function of the particle energy for protons, muons and electrons/gammas. These curves were obtained by injecting vertical particles sampled uniformly on top of a small water Cherenkov detector station with four PMTs placed at the bottom [23]. The parameterizations were built for the mean signal in the station and the signal distribution standard deviation. Through the use of these two numbers, it is then possible to emulate the WCD signal response fluctuations due to the stochastic processes of particle interactions and light collection. Additionally, for muons, it was also included the fluctuation in their tracklength due to geometry variations. Such was achieved using the distribution of the muon taken from proton-induced shower simulations ran over a Geant4 simulation, which provided the geometry of the WCD array and stations.

To make the comparisons fair and mimic realistic experimental conditions, the energy reconstruction is emulated conservatively by taking the total electromagnetic signal at the ground for each event. An energy bin with size $\log(E/\text{GeV}) = 0.2$ is taken for gammas, and cuts on the electromagnetic signal are derived. These cuts were defined taking $\mu \pm \sigma$, where $\mu$ is the mean of the distribution of the sum of the electromagnetic signal collected in the stations and $\sigma$ is its standard deviation.

These cuts are applied to the proton simulations over an extended energy range. In this way, we are comparing shower events with the same total signal at the ground.

3 The $C_k$ variable

The new variable $C_k$ is defined for each radial ring $k$ as:

$$C_k = \frac{2}{n_k(n_k - 1)} \left( \frac{1}{\langle S_k \rangle} \sum_{i=1}^{n_k-1} \sum_{j=i+1}^{n_k} (S_{ik} - S_{jk})^2 \right),$$

where $n_k$ is the number of stations in ring $k$, $\langle S_k \rangle$ is the mean signal in the stations of the ring $k$, and $S_{ik}$ and $S_{jk}$ are the collected signals in the stations $i$ and $j$ of the ring $k$, respectively. The term $\frac{2}{n_k(n_k - 1)}$ is the inverse of the number of two-combinations for $n_k$ stations, $\binom{n_k}{2}$. In this work, each circular ring $k$ is centred around the shower core position and has a width of 10 m.

The variable $C_k$ is just the mean sum of the square of the differences between the collected signals in any pair of two stations in the ring $k$ normalised to the mean signal observed in one station of the ring $k$. This normalisation should minimise the possible correlations between the dependence in $r_k$ of $C_k$ and the lateral distribution function. On a side note, the information
Figure 1. Mean value of $\log(C_k)$ as a function of $r_k$ for gamma showers with a primary energy $\in [100;160]$ TeV (blue, lower points) and for proton showers (red, upper points) with similar energies at the ground considering an array $FF = 1$. The errors bars are the RMS of the $\log(C_k)$ distributions in each $r_k$ bin. The full lines represents the best fit using the parametrisation expressed in equation 3.3.

in the LDF could be combined with the azimuthal asymmetry information, $C_k$, to enhance the discrimination capability. This is out of the scope of this paper and will be addressed in a future work.

For a completely uniform azimuthal distribution, $C_k$ is, by construction, equal to zero, while it should be greater than zero otherwise. Due to the presence of hadronic sub-showers, the pattern at the ground of a proton shower is more complex than the pattern of a gamma shower, whose development is basically driven by electromagnetic interactions. Thus, $C_k$ is expected to be higher in the case of proton-induced showers compared to gamma showers of equivalent energies at the ground.

In fact, this behaviour is clearly seen for $r_k$ greater than $40$ m in figure 1, where the distributions of the mean values of the $\log(C_k)$ variable are represented as a function of $r_k$ for gamma (blue points) and proton (red points) showers with $\sim 100$ TeV considering a $FF = 1$. The energy of the gamma-induced showers are in the energy interval of $100$ TeV to $160$ TeV, while proton-induced showers have similar energies at the ground.

The $C_k$ distributions as a function of $r_k$ depend on the primary energy and the array fill factor, but, if $C_k$ would be somehow correlated to the average number of detected muons at ground (see section 5), they should scale with a $K$ factor defined as:

$$K = E^\beta \times FF,$$

where $E$ is the primary gamma-ray energy (in TeV), $\beta$ is the index of the power dependence of the mean number of muons at the ground, and $FF$ the fill factor defined in the interval $[0;1]$. The parameter $\beta$ was fixed to 0.925, which is a typical mean value used in hadronic shower simulations [24].

Hence, for identical $K$ factors but different energies and fill factors, the mean values of the $\log(C_k)$ distributions is expected to be essentially identical. This crude rule of thumb is
Figure 2. Difference of the average distributions of $\log(C_k)$ as a function of $r_k$ of two simulation sets with different energies $E$ and fill factors $FF$ (see legend for details) but with similar $K \sim 10.6$ factors, computed using equation 3.2. This plot was produced using vertical proton-induced showers. The displayed error bars represent the propagated statistical uncertainty.

Successfully verified, for $r_k > 100$ m, at a level of a few percent in figure 2. In this figure are shown the difference of the average distributions of $\log(C_k)$ as a function of $r_k$ such that the compared showers have similar $K$ factors, namely: $E \approx 13$ TeV$^1$, $FF = 1 \rightarrow K = 10.67$; $E \approx 130$ TeV, $FF = 0.12 \rightarrow K = 10.77$; and $E \approx 1.3$ PeV, $FF = 0.014 \rightarrow K = 10.58$.

Finally, it should be also noted that the global behaviour of the $\log(C_k)$ distributions in the region $r_k > 40$ m can be described by the following parametrisation (full lines in figure 1):

$$\log(C_k) = a + \frac{b}{\log\left(\frac{r_k}{40\text{ m}}\right) + 1}$$

(3.3)

where $a$ and $b$ are free real parameters. In the limit $r_k \rightarrow 40$ m, $\log(C_{k,40}) = a + b$ and in the limit $r_k \rightarrow \infty$, $\log(C_{k,\infty}) \rightarrow a$.

This simple parametrisation complies with the two asymptotic limits and indeed captures well the evolution of $C_k$ as a function of $r_k$. The parameters $a$ and $b$ are determined, for each event, by fitting the corresponding $\log(C_k)$ distributions to the above parametrisation, using as errors the RMS of the $\log(C_k)$ distributions in each $r_k$ bin, which are also displayed in figure 1.

The overall quality of the fit is good ($\chi^2/n.d.f. \sim 1$) with a very small tail of events ($< 1\%$) with $\chi^2/n.d.f. > 2$. The analysis of these tail events is out of the scope of the present article, so these events were simply discarded. However, this type of analysis may tag events where extreme fluctuations in the shower development had occurred, such as, for instance, the so-called double-bang events or, more speculatively, exotic physics in the air shower development.

$^1$the mean value of the chosen bin assuming an $E^{-1}$ spectrum as a balance between computational time and statistics at the highest energies.
4 The gamma/hadron discrimination variable: $LCm$

A simple new $\gamma$/h discriminator is defined as the value of the log$(C_k)$ at a given value of $r_k = r_m$, computed using the parametrisation introduced in equation 3.3:

$$LCm \equiv \log \left( C_k \right)_{r_k=r_m},$$

(4.1)

The value of $r_m$ could be optimised for each energy bin as the value corresponding to the largest separation between the gamma and proton distributions.

The optimal $r_m$ value should grow slowly in a monotonous way as a function of the shower energy $E$. Nevertheless, to allow a more straightforward comparison between the values of $LCm$ at the different energies in the interval between 10 TeV and 1 PeV, and considering that the sparse array may have a radius of many hundreds of meters, $r_m$ was fixed to be:

$$r_m = 360 \text{m}.$$  

(4.2)

The $LCm$ distributions as well as their cumulative distributions are shown in figure 3 for gamma showers with energies $\sim 100$ TeV (blue points) and for proton showers (red points) with similar energies at ground, considering an array with $FF = 0.12$. Within the currently limited statistics\(^2\) no background events remain at a gamma efficiency close to 100%.

The rule of thumb introduced in the previous section (equation 3.2) should also apply, by construction, to the $LCm$ distributions, as it can be verified in figure 4, which is similar to figure 3, but now for the 1 PeV gamma energy bin and a $FF = 0.014$. In fact, $LCm$ is a function of $K$ which may be parameterised as (figure 5):

$$\langle LCm_i \rangle (K) \sim A_i + \frac{B_i}{\sqrt{K}}.$$  

(4.3)

\(^2\text{O}(10^4)\) events for proton showers.
where $A_i$ and $B_i$ are constants defined for a given primary particle type ($i = \gamma, \text{proton}$). Asymptotically when $K \to \infty$, $LCm_i(K) \to A_i$. The values of $A_i$ were found to be 2.049, 2.867 for $\gamma$ and proton, respectively, while the corresponding $B_i$ are equal to $-0.510, -0.874$.

Furthermore, the width of the $LCm(r_m)$ distributions for a given energy and fill factor was found to be reasonably well parametrized as function of $K$ (see figure 6) using the following equation:

$$\sigma_{LCm_i}(K) \sim \Delta_i + \frac{C_i}{\sqrt{K}},$$

where $\Delta_i$ and $C_i$ are constants defined for each primary particle type. Assuming Gaussian distributions, the values of $\Delta_i$ were found to be 0.036, 0.055 for $\gamma$ and proton, respectively,
Figure 6. Width of the $LCm$ distributions as a function of $K$ for gamma showers (blue, lower points) and for proton showers (red, upper points). The empty circles, squares and triangles correspond to samples with different energies, as defined in the inserted legend, and the full circles to the mean value for the several sets in each K bin. The line is the best fit using the parametrisation defined by equation 4.4.

while the corresponding $C_i$ were 0.024, 0.075.

These empirical parametrisations are thus well verified for all the simulated sets. However, to better establish the range of their validity and interpret their parameters in terms of the physics of the shower development, detailed studies implying larger simulated sets, different generators and realistic detector models will be needed.

5 Correlation of $LCm$ with the total number of detected muons

The number of detected muons is, as mentioned in section 1, often used as a $\gamma/h$ discriminator. Hence, some degree of correlation between $LCm$ and $N_\mu$ is expected. Muons are produced in the hadronic sub-showers, which are also the origin of the azimuthal non-uniformity.

In figure 7 we show, for the 100 TeV energy bin, the correlation of the $LCm$ distribution with the number of muons that hit the array stations located at a distance greater than 40 m from the shower core ($N_\mu$). The blue points are the gamma showers and the red points are the proton-induced showers. A uniform array with a fill factor of 12% and a radius of 1000 m was considered. The shower cores were placed at the centre of the array and the detection efficiency of the muons in the stations was considered to be 1. There is a clear almost linear correlation between $LCm$ and $\log(N_\mu)$ for $N_\mu \gtrsim 15$. It may be noted that one gamma event has a high positive fluctuation on the detected number of muons with a similar positive fluctuation of the $LCm$ values. Therefore, within the present limited statistics, both $N_\mu$ and $LCm$ lead to equivalent background rejection factors for a high-efficiency gamma event tagging at an energy of 100 TeV.

The same distributions are reported in figure 8 without considering now the contributions to the signal of the muons that hit the stations. The patterns shown in figures 7 and 8 are basically identical. This non-trivial attribute suggests that the $LCm$ variable is indeed
Figure 7. $LCm$ vs $\log(N_\mu)$ distributions for gamma showers with a primary energy of about 100 TeV (bluish histogram, to the left) and for proton showers (reddish histogram, to the right) with similar energies at ground, considering an array with $FF = 0.12$ and a radius of 1000 m and placing the shower cores at the centre of the array.

Figure 8. Same as figure 7, but not considering the contributions of muons and hadrons to the total signal of the stations.

sensitive to the sub-clusters structure of the showers. It is important to note that while these distributions are shown for an energy bin, the same correlation is present when analysing showers with fixed energy. This confirms that the correlation arises from intrinsic shower features rather than primary energy.
6 Discussion and conclusions

The recent identification in the Northern sky of more than 12 sources of gamma rays with energies up to the PeV opens a new era. These extreme energy events will be measured by present and future large wide-field gamma-ray observatories covering large surfaces and able to attain high background rejection factors. The measurement of the number of muons arriving at the ground is recognised, so far, as the best approach to reach such levels of rejection ($>10^4$), but it leads to very costly detection solutions.

In this article, we show that the quantification of the azimuthal non-uniformity in the pattern of the shower at the ground as a function of the distance to the shower core may be an alternative way to access the intrinsic differences in the development of electromagnetic and hadronic showers without implementing any costly strategy to absorb the electromagnetic component of the shower.

A new quantity is introduced which is simple and easy to compute. This variable presents a discrimination power similar to the one attained by the detected number of muons at the ground. It is worth noticing that while very effective in rejecting hadronic-induced showers, the ultimate discrimination power should be reached by combining this new quantity with other available discriminators accessible in the experiment, particularly with the LDF based variables.

It should be noted that, while the simulations presented in this work were generated for vertical events, the discriminator $LCm$ has a similar discrimination capability for inclined showers, provided that $r_k$ is computed in the plane perpendicular to the shower axis. Such was verified using a small set of simulations with zenith angles up to $35^\circ$.

The study of the precise requirements for the design of future large ground array gamma-ray observatories, like the future Southern Wide-field Gamma-ray Observatory (SWGO) [25], are out of the scope of this article. Such study would imply much larger sets of simulations both at the generator level and at the detector level to realistically include the performance of the chosen detector stations. However, we envisage that the station unit should be sensitive both to the muonic and to the electromagnetic component of the shower and it should have a low energy threshold. In this sense, water-Cherenkov detectors may be a good possibility. Moreover, the sparse array should cover in a way as uniform as possible surfaces of the order of 1 km$^2$ or higher, with variable fill factor as a function of the distance to the centre of the array, ensuring $K$ factors of about 10 for each targeted energy in the relevant regions. For PeV energies, this implies fill factor values that may be lower than 1% in the outer regions of the sparse array.

Finally, just a word to remark that this new proposed estimator may also contribute significantly to the study of charged cosmic rays and, namely, in the determination of the composition and in the test and validation of hadronic models in an extended energy range.

Acknowledgments

We would like to thank Jaime Alvarez-Muñiz, Pedro Brogueira, Roger Clay, Antonio Bueno, Alessandro De Angelis, Giorgio Matthiae, Jakub Vicha, Alan Watson and Enrique Zas for carefully reading the manuscript and providing useful comments. This work has been financed by national funds through FCT - Fundação para a Ciência e a Tecnologia, I.P., under project PTDC/FIS-PAR/4300/2020. R. C. is grateful for the financial support by OE - Portugal, FCT, I. P., under DL57/2016/ cP1330/cT0002.
**Dedication**  We dedicate this article to the memory of our friend, Prof. Ronald C. Shellard (Ron), who passed away recently, December 2021, and with whom some of us have worked closely during many years, developing new ideas and strategies, namely in the field of high energy charged and neutral cosmic rays.

**References**

[1] M. Amenomori et al., *First Detection of sub-PeV Diffuse Gamma Rays from the Galactic Disk: Evidence for Ubiquitous Galactic Cosmic Rays beyond PeV Energies*, Phys. Rev. Lett. **126** (2021) 141101.

[2] Cao, Zhen and et. al, *Ultrahigh-energy photons up to 1.4 petaelectronvolts from 12 γ ray Galactic sources*, Nature **594** (2021) 33.

[3] K. Greisen, *Cosmic ray showers*, Annual Review of Nuclear Science **10** (1960) 63.

[4] Abeysekara and et. al, *Observation of the Crab Nebula with the HAWC Gamma-Ray Observatory*, The Astrophysical Journal **843** (2017) 39.

[5] P. Assis et al., *Design and expected performance of a novel hybrid detector for very-high-energy gamma-ray astrophysics*, Astropart. Phys. **99** (2018) 34 [1607.03051].

[6] A.M. Hillas, *Cerenkov light images of eac produced by primary gamma*, in 19th Intern. Cosmic Ray Conf-Vol. 3, no. OG-9.5-3, 1985.

[7] T.C. Weekes et al., *Observation of TeV gamma rays from the Crab nebula using the atmospheric Cerenkov imaging technique*, Astrophys. J. **342** (1989) 379.

[8] P.R. Blake and W.F. Nash, *Muons in extensive air showers. i. the lateral distribution of muons*, Journal of Physics G: Nuclear and Particle Physics **21** (1995) 129.

[9] N. Hayashida and et. al a, *Muons (≥ 1 GeV) in large extensive air showers of energies between 1016.5 ev and 1019.5 ev observed at akeno*, Journal of Physics G: Nuclear and Particle Physics **21** (1995) 1101.

[10] W. Apel and et. al, *The kascade-grande experiment*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **620** (2010) 202.

[11] A. Borione and et. al, *A large air shower array to search for astrophysical sources emitting γ-rays with energies ≥ 10^{14} eV*, Nuclear Instruments and Methods in Physics Research A **346** (1994) 329.

[12] A. Aab and et. al, *Calibration of the underground muon detector of the Pierre Auger Observatory*, Journal of Instrumentation **16** (2021) P04003.

[13] LHAASO collaboration, *Design and performances of prototype muon detectors of LHAASO-KM2A*, Nucl. Instrum. Meth. A **789** (2015) 143.

[14] P. Abreu et al., *MARTA: a high-energy cosmic-ray detector concept for high-accuracy muon measurement*, Eur. Phys. J. C **78** (2018) 333 [1712.07685].

[15] A. Letessier-Selvon, P. Billoir, M. Blanco, I.C. Mariş and M. Settimo, *Layered water Cherenkov detector for the study of ultra high energy cosmic rays*, Nucl. Instrum. Meth. A **767** (2014) 41 [1405.5699].

[16] A. Zuñiga-Reyes, A. Hernández, A. Miranda-Aguilar, A. Sandoval, J. Martínez-Castro, R. Alfaro et al., *Detection of vertical muons with the HAWC water Cherenkov detectors and its application to gamma/hadron discrimination*, arXiv preprint arXiv:1708.09500 (2017).

[17] B.S. González and et. al, *Using convolutional neural networks for muon detection in WCD tank*, Journal of Physics: Conference Series **1603** (2020) 012024.
[18] P. Assis et al., The Mercedes water Cherenkov detector, 2203.08782.

[19] D. Heck, J. Knapp, J. Capdevielle, G. Schatz and T. Thouw, A monte carlo code to simulate extensive air showers, Report FZKA 6019 (1998).

[20] A. Ferrari et al., FLUKA: A multi-particle transport code, CERN-2005-010, SLAC-R-773, INFN-TC-05-11 (2005).

[21] T. Böhlen et al., The FLUKA Code: Developments and Challenges for High Energy and Medical Applications, Nuclear Data Sheets 120 (2014) 211.

[22] S. Ostapchenko, Monte Carlo treatment of hadronic interactions in enhanced Pomeron scheme: I. QGSJET-II model, Phys. Rev. D 83 (2011) 014018 [1010.1869].

[23] R. Conceição, B.S. González, A. Guillén, M. Pimenta and B. Tomé, Muon identification in a compact single-layered water Cherenkov detector and gamma/hadron discrimination using machine learning techniques, Eur. Phys. J. C 81 (2021) 542 [2101.10109].

[24] T. Bergmann et al., One-dimensional hybrid approach to extensive air shower simulation, Astropart. Phys. 26 (2007) 420 [astro-ph/0606564].

[25] P. Huentemeyer et al., The Southern Wide-Field Gamma-Ray Observatory (SWGO): A Next-Generation Ground-Based Survey Instrument, BAAS 51 (2019) 109.