Improving H.E.S.S Monoscopic Gamma-Hadron Separation at Low Energies and Application to the Vela Pulsar

Master's Thesis

submitted to

Indian Institute of Science Education and Research Tirupati

in partial fulfillment of the requirements for the

BS-MS Dual Degree Programme

by

Prajath B R

Roll Number: 20191043

Supervisor: Prof. Dr. James Anthony Hinton

Max-Planck-Institut für Kernphysik, Heidelberg, Germany

April, 2024

©Prajath B R and MPIK (2024)
All rights reserved

Acknowledgements

I would like to express my deepest gratitude to my supervisor Prof. Dr. Jim Hinton, and co-supervisors Dr. Lars Mohrmann and Dr. Laura Olivera Nieto for their unwavering guidance and support throughout my thesis at Heidelberg. Under their mentorship, I have grown both professionally and personally, and I am truly thankful to have had the opportunity to work alongside such esteemed mentors. I would like to thank my co-supervisors for their patience in explaining the same concepts a hundred times and correcting me over my mistakes. I especially thank Dr. Lars Mohrmann for processing the required datasets that were used in the thesis, and Dr Laura Olivera Nieto for tirelessly proofreading my thesis during her leisure time at conferences. I would also like to thank Dr. Richard White for agreeing to be the expert member and being part of the evaluation committee. I would like to extend special appreciation to Dr. Simon Steinmaßl for his assistance with the finer details, even amidst his busy schedule.

I would like to thank everyone at IISER Tirupati and MPIK for making this collaboration possible and allowing me to pursue a thesis project at MPIK. I also thank MPIK and the organizing bodies for providing the necessary funding throughout my time in Germany. I thank all the IISER Tirupati faculty whose lectures have taught me all the necessary concepts required to excel in a research career. I also thank the thesis committee for their guidance throughout the process.

I would like to thank the H.E.S.S collaboration for the significant contributions to the H.E.S.S framework that allowed me to work on my thesis. In particular, I would like to acknowledge the invaluable assistance of Dr. Rodrigo Guedes Lang and Tim Unbehaun. I also thank Ms. Anita Deka Baruah for allowing me to use the reconstruction weights for the thesis and engaging in all the discussions that were very helpful in developing a core understanding of several concepts.

I would like to thank all the members of the Hinton division for providing an interactive and productive workplace, with a special thanks to Georg Schwefer for helping with theoretical details and also for sharing insights into navigating life in Germany.

I thank my dear friends and family for all the support I needed throughout my time in Germany. I thank my parents for supporting all my decisions and my sister Manasa B R for her cooking recipes that I did not do a good job replicating. I also thank Ainesh Sanyal and Rakshith Mallesh for the much-needed emotional support during my stay outside India. I would like to thank Ankush Kaushik for the random cooking sessions at his place.

The thesis would not have been possible without the support of friends and family during my undergraduate years at IISER Tirupati. I would like to express my gratitude to all my friends with whose support I was able to grow personally. I thank Rakshith Mallesh, Akasha T and Vasanth Kashyap for all the fun times from watching movies to swimming and skateboarding. I would also thank Ainesh Sanyal for the emotional support and random table tennis/coffee breaks, Anita Deka Baruah and Aryan Patel for their support and friendship, and Sunanda Biswas for the late-night movies. I thank Alen James for all the late-night gaming sessions and snacking.

Finally, I would like to express gratitude towards my parents Ramesh B S and Umabai for their support throughout my life and backing up all my decisions while also being the world's greatest mentors.

Abstract

Gamma-rays of astrophysical origin having energy $E_\gamma \gtrsim 10$ GeV produce a cascade of secondary particles called an extensive air-shower upon their interaction with the particles in the atmosphere. Detecting these gamma rays is crucial to understanding the nature of high-energy particles and non-thermal emissions in the universe. The High Energy Spectroscopic System (H.E.S.S) uses an array of Cherenkov telescopes to detect the emission from the air-shower particles. H.E.S.S has been successful in detecting TeV gamma rays from various sources like supernovae, AGNs and pulsars. This thesis involves the use of monoscopic observations of the Vela pulsar from the central CT5 telescope of H.E.S.S to detect pulsations in the sub-100 GeV energy range.

The analysis process requires separating the gamma-ray induced showers from the background-dominated cosmic-ray induced showers. This process of separation is carried out in H.E.S.S using a Boosted Decision Tree classifier (BDT) by training on point-source Monte-Carlo simulations of gamma-ray showers. This separation is comparatively difficult for low energies using single telescope observations due to high similarities in the observed properties of the shower. Having additional features that are different for the two classes would be beneficial for this process. This thesis presents the improvement in the performance of the classifier by adding new input parameters to the BDT training process. The parameter among them that is the most helpful in the discrimination is discussed. The results from training the classifier on two different ranges of shower intensities are also presented.

The selection cuts that define the decision boundary between what is considered signal or background are optimized to get the best significance of signal over background. Further, a consistency check using diffuse Monte-Carlo simulations of gamma-ray showers is performed to investigate the effect the trained model could have on showers recorded at different quadrants of the camera.

An improved background separation at low energies would benefit in better distinguishing the pulsar signal from the background. This thesis also presents the application of the improved performance of the classifier to the Vela pulsar to get a spectral energy distribution in the 30 to 100 GeV energy range. Finally, the results are compared with the previous detections of Vela with H.E.S.S.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Non-thermal radiation

The radiation emitted by a black body can be associated with a well-defined spectral energy distribution according to Planck's radiation law. This radiation is dependent on the temperature of the black body and is called thermal or black body radiation. While a lot of the radiation coming from space can be associated with thermal processes, there exists radiation in certain wavelengths that cannot be linked to a radiating black body and are said to have *non-thermal* origins. Their spectral energy distribution is very different and most commonly resembles a power law. Examples of such radiation are high-energy γ-rays, fast moving positively charged nuclei, radio and X-rays. Fig 1.1 shows the galactic plane in different wavelengths and it can be seen that a large part of the center is very luminous in wavelengths other than optical and infrared. The discovery of sources responsible for the production of non-thermal radiation and the underlying processes is an active area of research.

Earth is constantly being bombarded by charged particles traveling close to the speed of light. Having energies orders of magnitude greater than TeV, they are capable of penetrating and ionizing the upper atmosphere. First confirmed by Victor Hess using balloon experiments that their origin has little to do with the Sun as their spatial distrubution is isotropic. The reason being that charged particles can deviate from their original direction through multiple deflections from interstellar and intergalactic magnetic fields. Apart from the most energetic of these particles, their direction serves no purpose in determining the source of emission. The energy spectrum of these charged particles, termed *cosmic rays*, can be approximated by a power law with an index of -2.7 [1]. There are no known thermal mechanisms that can produce particles with such energies and spectral distribution.

Astrophysical γ-rays on the other hand follow a rectilinear path, undeflected by

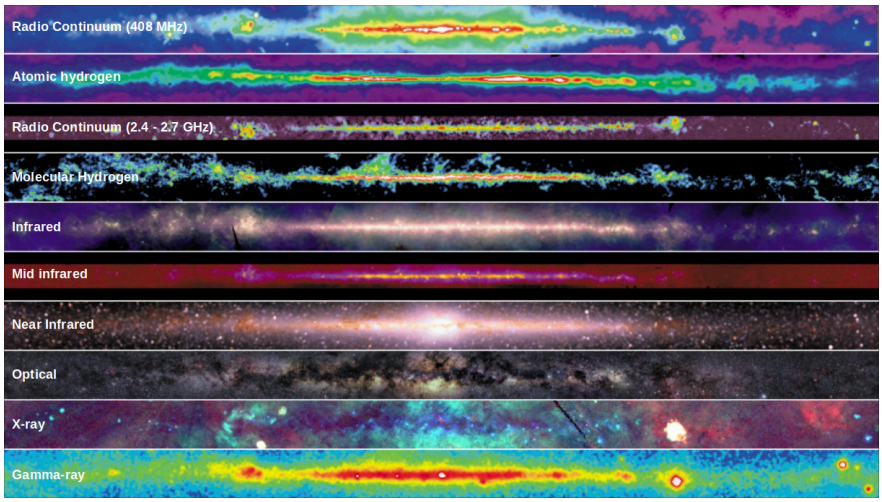


Fig 1.1: A multi-wavelength view of the galactic plane (credit: NASA)

magnetic fields and point to the direction of their source. For a black body's radiation to peak at gamma-rays they would need to have a temperature of over 10^9 K. Such temperatures can be achieved in some cases like at the core of a star during a supernova, but γ-rays have been detected from sources that are not known to have such temperature, for example, a neutron star [2]. These γ-rays are most often the products of radiation losses form high energy charged particles. Leptonic process involve electrons and positrons losing their energy to synchrotron and Inverse Crompton (IC) radiation. Synchrotron radiation is responsible for the observed radio and X-ray spectra, whereas the IC scattering of thermal photons like the Cosmic Microwave Background (CMB) produces γ-rays in the high energy (GeV) and very high energy (TeV) range [3]. γ-rays from Hadronic processes involve the decay of neutral pions $\pi^0 \rightarrow \gamma + \gamma$ which result from the collision of the charged particle with ambient nuclei [4].

The requirement of the presence of energetic charged particles for the production of γ-rays make them an important tool to discover the source of cosmic-ray acceleration and their distribution. The sources capable of having such acceleration power are called cosmic accelerators. Some of them are Active Galactic Nuclei (AGNs), pulsars and supernova remnants. This thesis is mostly concerned with non-thermal γ-ray emission from pulsars.

1.2 Pulsars

Pulsars are neutron stars that are rapidly rotating with periods ranging from milliseconds to a few seconds. Resulting from core-collapse supernovae, their fast period

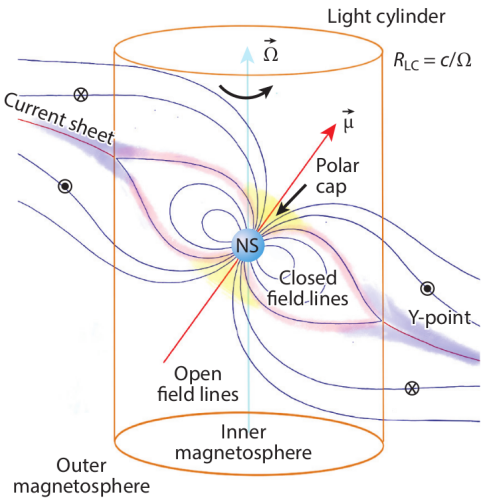


Fig 1.2: The magnetosphere of a pulsar and its terminologies [7].

is a consequence of angular momentum conservation. They are very dense and comprise of degenerate neutrons. This dense nature results in a very high magnetic field strength and calculations show that they are of the order of 10^{12} Gauss [5]. Due to their rotation, their detection can be characterised by pulsed signals periodic in time and this pulsation can sometimes be seen in multiple wavelengths. The integrated pulse profile of the Vela pulsar is shown in Fig 1.3 as an example. It is important to know the morphology of the pulsar magnetosphere in order to understand the emission mechanisms.

1.2.1 Magnetosphere

Pulsars have bipolar magnetic fields whose magnetic axis is usually at an angle from the rotation axis (refer to Fig 1.2). For a pulsar with angular velocity Ω, the light cylinder defines an imaginary surface whose radius is defined as the distance from the rotation axis where the rotational velocity equals the speed of light, $R_{LC} = c/\Omega$. The magnetic field lines which are within this cylinder are closed whereas those outside the cylinder are open and extend to infinity. Theoretical models usually assume the pulsar as a perfectly conducting sphere where the electric field inside the surface is zero, i.e. $\mathbf{E} = -(\mathbf{\Omega} \times \mathbf{r}) \times \mathbf{B}/c$ [6]. Continuity of the electric field requires that there is a non-zero component of the electric field parallel to the magnetic field, outside the surface ($\mathbf{E} \cdot \mathbf{B} \neq 0$). This electric field is strong enough to pull the surface charges away from the surface and these charges further screen the electric field, giving a force-free state in the plasma-filled magnetosphere as well.

Particle acceleration

The acceleration of charges needs the presence of strong electric fields, meaning that the force-free nature of charges has to be broken. This would imply that there need to be pockets of low density where $\mathbf{E} + (\mathbf{\Omega} \times \mathbf{r}) \times \mathbf{B}/c \neq 0$ and the residual $\mathbf{E}$ field can accelerate particles. Two main regions have been identified where these *acceleration gaps* can exist. One is between the closed and open field lines near the light cylinder, called the outer gap and the other is the polar gap located at the magnetic poles. This gives rise to two locations of particle acceleration and subsequent non-thermal emission.

Polar gap: The edges of the polar gap region are given by the last open magnetic field line tangential to the light cylinder. The residual magnetic field accelerates the primary particles which suffer radiation losses through curvature radiation and inverse Compton effect [8]. Curvature radiation is a type of synchrotron radiation where the charged particles radiate under the influence of *curved* magnetic fields. The inverse Compton scattering of the surface X-rays is responsible for the observed γ-ray flux according to this model. The presence of strong magnetic fields facilitates the production of electron-positron pairs through a single-photon conversion process. These pairs move along the magnetic field lines further producing more γ-rays, and this cascade creates a secondary pair plasma. The secondary plasma streams out along the magnetic field lines with large γ factors $\sim 10 - 10^3$. These secondary plasma particles are proposed to produce the measured radio flux, at a distance from the surface. The attenuation of γ-rays to $e^{\pm}$ pair-production in the $\mathbf{B}$ field produces a characteristic sharp exponential cut-off in the observed spectrum.

Outer gap: The emission region in the outer gap model is situated near the light cylinder along the null field lines ($\mathbf{\Omega} \cdot \mathbf{B} = 0$) which separate the charges of opposite polarity [9]. The gap results from the inefficiency of the surface charges to replace the charges flowing out along the open field lines. This creates the acceleration gap that accelerates the charged particles. Being far from the surface, the density of thermal photons is low at the outer gap, making curvature and synchrotron the main emission mechanism for GeV photon energies. However, TeV photons are likely to be the result of IC scattering of infrared photons. Unlike the polar gap scenario, the emission is not attenuated by $\gamma - B$ pair creation. Limited mostly by the curvature radiation, the exponential cut-off starts at higher energies and is expected to be less steep.

There have been a lot more modifications to the theoretical models both analytical and also with the help of detailed simulations. A recent review of the emission models can be referred to [7]. The energy radiated in the form of non-thermal radiation is transferred from the pulsar's rotational kinetic energy. This loss of energy is responsi-

ble for the slowing down of pulsars over time. Through precise timing measurements, this can be measured in the form of a period derivative. Although small, this becomes important for time-averaging the pulsed signal from years of observational data. For a brief description of pulsar and pulsar astronomy, refer to [10]

1.3 γ-ray detection techniques

The observation of γ-rays can be categorised into direct and indirect detection. The atmosphere is responsible for absorbing most of the γ-rays and there is little to no flux on the ground. Therefore direct detection involves the use of space-based telescopes like the Fermi-LAT, INTEGRAL, etc, which have a relatively small collection area of around a few square metres. This approach is limited to photon energies of a few 100 GeV after which the detection becomes flux-limited due to most sources having a power law like flux. Their detection requires a much larger collection area and increased observation times. Increasing the collection area of space-based telescopes is not a viable option due to the requirement of launching them on space crafts and hence other methods are required.

High energy γ-rays ($E_\gamma >\sim 10$ GeV), however, on interacting with particles in the atmosphere create a cascade of secondary particles known as *extensive air-showers*. Detection of these secondary particles and reconstruction of initial γ-ray population is the basis of very high-energy γ-ray astronomy. There are a few different ways in which this is carried out. Imaging Atmospheric Cherenkov Telescopes (IACTs) like H.E.S.S and MAGIC make use of the Cherenkov light emitted by the secondary particles in the atmosphere, arrays like HAWC (High Altitude Water Cherenkov) detect the Cherenkov light emitted by secondary particles in water tanks, and there are some radio telescopes that detect the radio emission from the shower. These methods offer a large effective collection area (over 10^5 m^2) which is required for the low flux in this energy range. This thesis is focused on the γ-ray observations of the Vela pulsar from the CT5 telescope in the High Energy Spectroscopic System (H.E.S.S) array. The details of the detection technique employed in H.E.S.S is discussed in Chapter 2.

1.4 Vela Pulsar

The Vela pulsar (PSR 0833-45) is one of the fastest spinning pulsar known with a period of 89 ms. The γ-ray pulsations from the pulsar were first detected by the SAS-2 telescope at energies above 35 MeV [12]. Later it was also detected by the Fermi-LAT [13]. The phase-averaged pulse from the Fermi-LAT is shown in Fig 1.3.

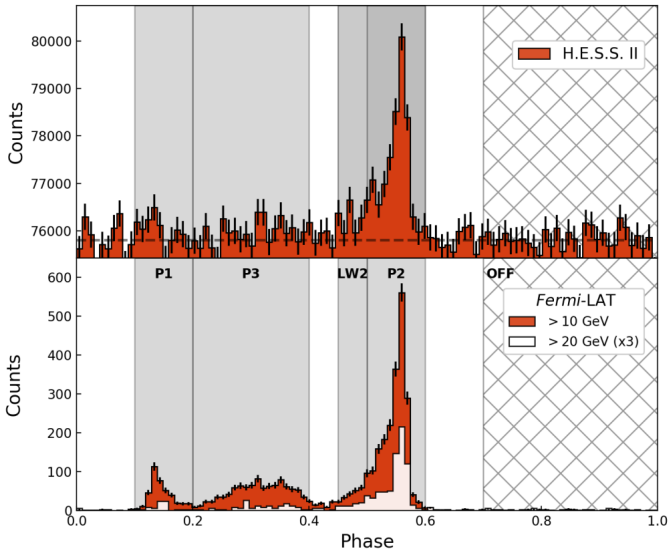


Fig 1.3: Pulse profile from the Fermi-LAT (lower) and the H.E.S.S (upper) observations. In the bottom part, the double-peaked emission is clearly visible [11].

The pulse profile from the Fermi-LAT consists of two peaks P1 and P2 with P2 being the most significant among the two. There is also a bridged emission P3 that connects the two peaks. P1 and P3 cannot be seen in the H.E.S.S pulse profile.

Spectral analysis showed that the best-favoured model was a power law with an exponential cut-off

$$dN/dE = \phi_0(E/E_0)^{-\Gamma} \exp -(E/E_c)^b \tag{1.1}$$

with $b < 1$. The test statistic was favored against a super-exponential cut-off ($b > 1$) implying that the outer gap scenario is likely responsible for the high-energy emission in Vela. In addition, the Fermi analysis also showed that the emission height in Vela is comparable to R_{LC}. Later, H.E.S.S was also successful in the detection of the pulsar above 20 GeV using its CT5 camera with evidence of curvature [11], and a recent H.E.S.S study using a stereoscopic approach reported the pulsations to extend beyond TeV [14]. The model which best explained the observed TeV flux was associated with synchrotron/IC rather than curvature/IC, and the emission height was restricted to 1-2R_{LC}. It was also shown that the high-energy (GeV) flux detected by the CT5-only approach and the very high-energy flux (TeV) from the stereo analysis could be best reproduced if two different polulations of electrons(and positrons) were assumed.

This thesis intends to improve upon the background separation at low energies and the application of the improved analysis on the Vela pulsar.

Chapter 2

Ground-based detection

This section presents an overview of the detection techniques that are employed by IACTs, particularly in H.E.S.S. To construct a map of the sky, two main quantities are required: the direction of the photon and its energy. The basic principle employed by IACTs is to use the information gathered from the light emitted by the air-shower particles to reconstruct the direction and energy of the γ-ray that initiated the shower.

2.1 Air showers

γ-rays with $E_\gamma \gtrsim 10$ GeV, upon interacting with the nuclei of molecules in the atmosphere, initiate the shower by the creation of $e^\pm$ pair. The pairs have relativistic speeds and further produce γ-rays through the process of Brehmmstrahlung. The two processes, pair production and Brehmmstrahlung continue to create a cascade of γ-rays and $e^\pm$, with the energy being divided among the particles/photons with each step. The cascade continues until the energy of γ-rays has reached the critical energy required for pair creation, and the rest of the energy of $e^\pm$ is lost through ionization processes. A simulated air-shower from a γ-ray is shown in Fig 2.1. The shower contains a core that consists of most of the particles and a fraction of particles having larger transverse velocities are away from the core.

Most particles in the air-shower are moving at relativistic speeds that exceed the speed of light in the medium. This produces Cherenkov radiation which is radiated at an angle of $cos\theta = 1/\beta n$ from the direction of the moving particle, for a medium with refractive index n. The collective emission is radiated at an angle of $\theta \sim 1^0$ from the direction of the shower core. The Cherenkov radiation is in the visible range and mostly near the blue part of the spectrum. The flash of light lasts a few nano-seconds and IACTs rely on fast photomultipliers to detect them. The amount of Cherenkov light produced depends on the absolute number of particles in the shower, which,

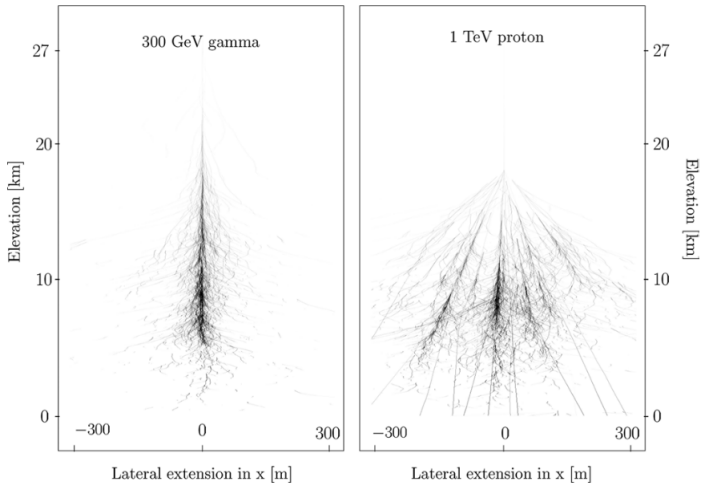


Fig 2.1: Simulated shower images of a γ-ray and a proton. [15]

is dependent on the initial energy of the shower-inducing particle. Therefore the Cherenkov light yield is proportional to the γ-ray energy.

Complications arise due to the fact that air-showers can also be initiated by high-energy cosmic-rays. It becomes important to differentiate whether the Cherenkov light detected by the telescope came from a γ-ray or a cosmic-ray shower. Fortunately, there are some key differences between the showers which can be used as a starting point to separate the two kinds.

2.1.1 Hadronic vs electromagnetic showers

Electromagnetic showers (induced by a γ-ray) contain mainly electrons, positrons and γ-rays. The shower profile is fairly homogeneous and the transverse spread is small. Whereas, in the case of hadronic showers (induced by cosmic-rays), the shower consists of the following components:

- nuclear fragments resulting from the collision with molecules in the atmosphere
- decay components like π, K mesons
- γ-rays from π^0 decay, which can initiate their own electromagnetic cascade
- muons from the decay of charged mesons

The presence of these sub-components gives the hadronic showers an irregular profile, as can be seen from Fig 2.1. This also results in them being wider than their electromagnetic counterparts.

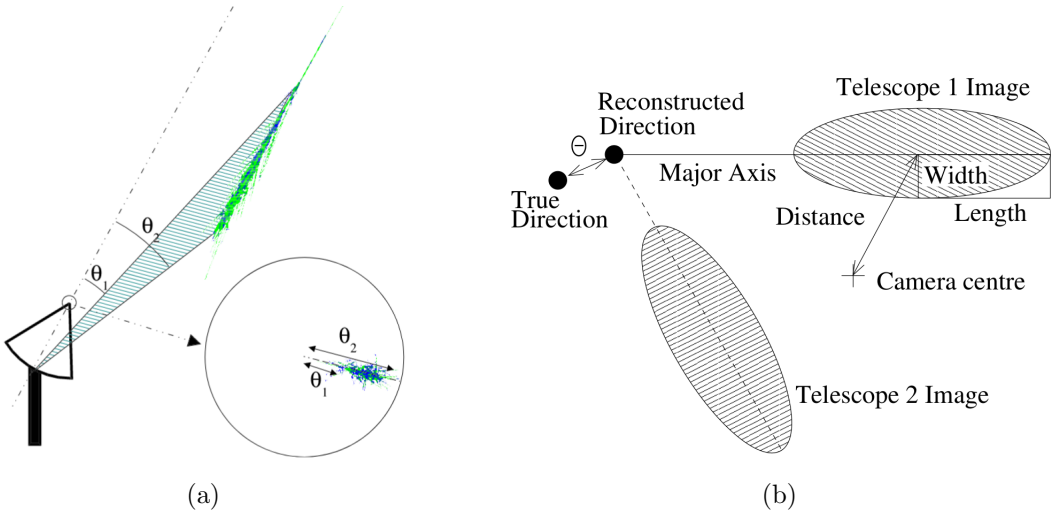


Fig 2.2: (a) The angular profile of the air-shower being captured by a Cherenkov telescope. [16] **(b)** The parametrization of the shower image and stereoscopic reconstruction using images from two telescopes [17]

2.2 H.E.S.S

Located in Namibia, the H.E.S.S array consists of four 12 m diameter telescopes named CT1-4 and one larger 28 m diameter CT5 telescope. CT5 is placed at the center of a square with CT1-4 at each corner. The CT5 telescope can collect more light and hence is able to capture more low-energy showers compared to CT1-4. The array is used in different configurations, depending on the science case, to perform an analysis. The **mono** mode makes use of only the events from CT5 and offers an increased sensitivity to low-energy events. The **stereo** mode makes use of the CT1-4 telescopes, having the requirement that at least two of the four telescopes be triggered at a time. This method is sensitive to higher energies and gives improved direction and energy reconstruction. A third **hybrid** method makes use of all five telescopes.

IACTs reflect the incoming Cherenkov light with the help of mirrors onto the focal plane. A detector centered on the focal plane collects this light, whose pixels are composed of photomultiplier tubes. When the Cherenkov light from the air-shower is in the field of view of the telescope, the detector captures the angular profile of the shower, as shown in Fig 2.2. The Cherenkov light pool covers a large area on the ground, which offers this method a large effective area of collection (over 10^5 m^2). An individual detection of an air-shower is called an event. The higher the energy of the particle that induced the shower, the brighter the shower image produced by the photomultipliers, given the same observation conditions.

Since the Cherenkov light is in the blue wavelength, the telescope cannot be

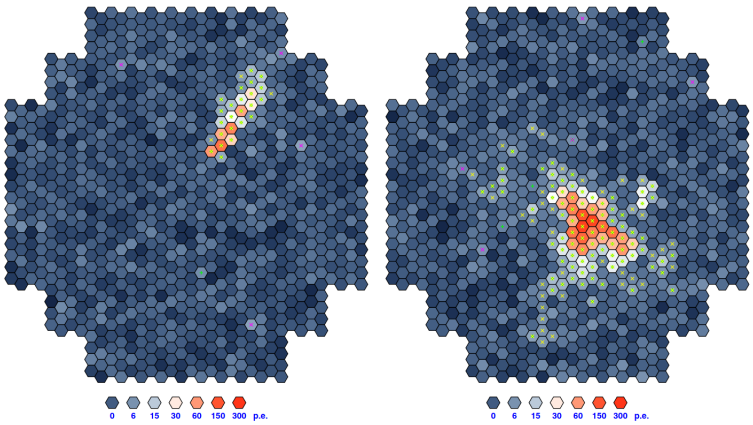


Fig 2.3: An extreme example of the images of two showers **left** from a γ-ray, **right** from a proton. [15]

operated during the day due to contamination from sunlight or also during nights with bright moon light. During the night, however, starlight and other stray light constitute the unwanted background referred to as the Night Sky Background (NSB). Thus the captured image consists of noise pixels and needs to be cleaned to retain only the shower images. In H.E.S.S, two cleaning algorithms are used:

- **Tailcut:** This method uses a nearest-neighbor criteria, where those pixels are kept that have a minimum amplitude say of 9 p.e. (photoelectrons) and a certain number of neighboring pixels that have at least say 16 p.e. amplitude. The minimum number of neighbors and amplitude threshold can be varied.
- **Time-cleaning:** This method makes use of a clustering algorithm that takes into account both the spatial and temporal clustering of pixels. The shower pixels are going to be correlated in both space and time, whereas the noise pixels will be de-correlated. This method is more efficient in removing the noise, while retaining the shower pixels.

The cleaned images will now contain only the required pixels needed for further data reduction. The events are parametrised by using the moments of the shower image and are required for reconstruction. The meaning of these parameters are further explained in section 3.2.

2.2.1 Reconstruction

The major axis of the shower image gives the projected direction of shower propagation. Somewhere along this axis is the point of impact of the γ-ray that initiated the

shower. The statistical nature of air-showers makes it impossible to get the energy from the shower directly. No two showers would look identical even if they were produced from a γ-ray with the same energy. Depending on the telescope configurations, the reconstruction of energy and direction is carried out differently. If two or more telescopes detected the same shower (refer to Fig 2.2b), as in the case of stereo configuration, they each see a different projection of the same shower and the intersection of the major axes gives the point of origin of the shower (stereoscopic reconstruction). This is the simplest approach. For the energy reconstruction, in H.E.S.S, a pixel-wise template fitting method has been in use recently for events involving two or more telescopes [18].

For the case of mono, stereoscopic reconstruction is not possible. A neural network is instead used to predict the source displacement along the major axis of the image. The training of the neural network is performed using simulated γ-ray events. Detailed Monte Carlo simulations of air-showers, telescope optics and electronics are carried out for a range of input energies, directions and impact points on the ground, to produce thousands of shower images. The model trained on the parameters derived from the images is then used for predicting the direction of events from a real observation. A separate neural network is trained using the same parameters to predict the energy of the γ-ray.

2.2.2 Gamma-hadron separation

The telescopes capture many more background events coming from a cosmic-ray than γ-ray events. The observations are mostly background-dominated and can have over ten times more background events than γ-ray events. For performing an analysis, the background events need to be separated from the required events. The differences in the shower profiles between γ-ray and hadron showers are also reflected in the images captured by the telescopes. Fig 2.3 shows an example of the two shower images, one from a 1 TeV γ-ray and the other from a 2.6 TeV proton. As the γ-ray showers are more narrow and regular, their shower images are more elliptical. In comparison, the hadronic images are wider and irregular owing to the presence of multiple subcomponents. This difference can be used to differentiate between the said categories. Fig 2.3 is an extreme example of the two showers where the difference can be clearly seen, but this is often not the case and expecially when dealing with energies below a few hundred GeV the showers are very similar.

A parameter selection based approach was initially used for event selection [19]. This approach uses the fact that the parameter distribution of the two sets of shower differs, as can be seen in Fig 2.4. A cut on Monte Carlo-based distribution value

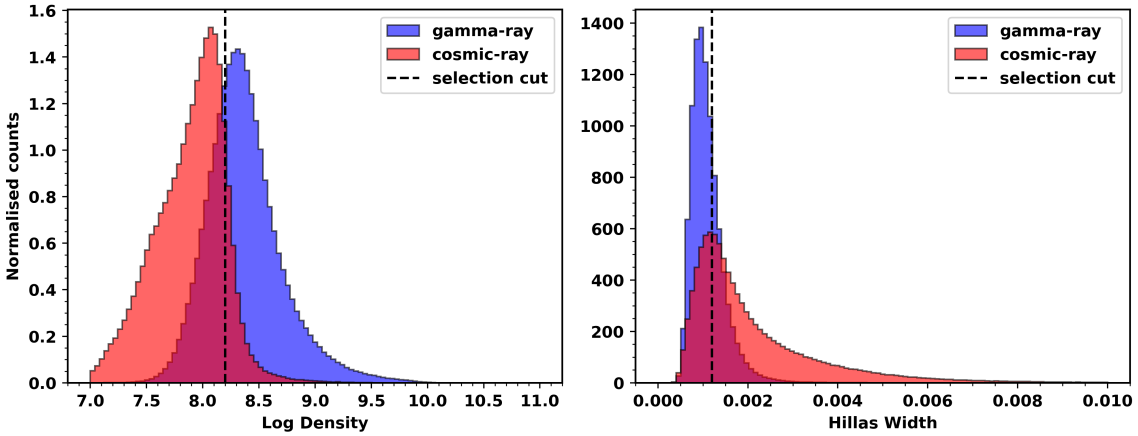


Fig 2.4: Shown are the distribution of two shower image parameters for both γ-ray and cosmic-ray showers. Examples of selection cuts are depicted with the dashed black line.

would be decided that would represent the most γ-like events. An event that passes these cuts would be classified as a γ-ray event. This method is not very efficient as it does not take into account the cross-correlations between the different parameters and results in more background events passing the cuts.

More recently, newer methods have made use of machine learning tools like neural networks and decision trees. In H.E.S.S, the gamma-hadron separation is done using a Boosted Decision Tree (BDT) classifier [20]. This is a multivariate method that takes the parameter distribution as input and outputs a probability value ζ that is related to how likely the input parameters are from a γ-ray shower. More information on BDT is presented in the next chapter. This method has been proven to be much better than a simple cut-based selection as it also incorporates the correlation between the various parameters [20]. A value of ζ is selected based on the science case, whether the spectrum is *hard* or *soft*. High-energy γ-ray events have well spearated parameter distributions and therefore get a higher value of ζ. A higher cut on ζ would be ideal for a hard spectrum and vice versa.

The BDT is trained on the set of parameters derived from Monte Carlo simulations for γ-ray shower. The simulations are carried out for a *point* source with an assumed spectral index, which is at a specific offset in the field of view of the telescope. For the background, the parameters are taken from real observations of cosmic-ray showers, called OFF-runs. These observations are carried out in regions without any known γ-ray sources, so the recorded showers mostly contain cosmic-ray showers. From the point of view of what the telescope sees, the γ-ray showers originate from a point in the field of view (when talking about point sources) which is at some offset from the pointing position, whereas the background is diffuse and originates from every

direction in the field of view. The idea is to differentiate between the signal events from the diffuse background and therefore the background events are not selected based on offset from the pointing position. There are also MC simulations that simulate diffuse signal events and more information about them is presented in section 3.8.

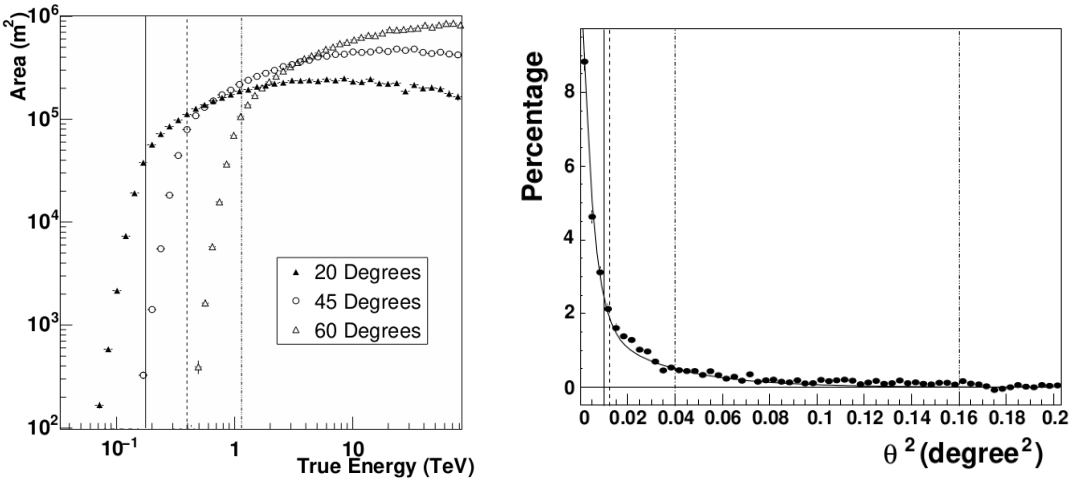
To ensure good separation between signal and background, the training is carried out separately in zenith angle bands. A cut on a given value of ζ is associated with a signal efficiency (fraction of signal events remaining after the cut) and background fraction (fraction of background events remaining after the cut). The ideal case would be to get a high signal efficiency and a very low background fraction after a cut. Since the performance of the classifier is not the same across zenith bands, the classifier output ζ is scaled to a value that corresponds roughly to the signal efficiency, called ζ_{BDT}. A cut on ζ_{BDT} of say 0.5 would correspond to roughly 50% signal efficiency across all the zenith angle bands, making the process uniform.

2.2.3 IRFs and lookups

Using the above techniques, the events that are most γ-like can be determined, along with their reconstructed direction and energy. However, the required flux cannot be calculated from this alone. Relating the count data to the flux requires the knowledge of the response of the instrument to any signal, termed the Instrument Response Functions (IRFs). The IRFs are produced using simulated events that pass two selection cuts: the ζ_{BDT}, which represents the most γ-like events and an angular size cut around the simulated source position, called the θ^2 cut. The θ^2 represents the squared angular difference between the true and reconstructed directions. The following are the important IRFs:

- **Energy dispersion:** The energy dispersion relates the reconstructed energy to the most probable simulated true energy. This is required to present the flux in terms of the true energy of the γ-ray. To avoid having large differences between true and reconstructed energy, usually, the analysis is carried out in the energy range where the relative error with respect to the true energy is less than 10 percent.

- **Effective area:** The effective area IRF (refer to Fig 2.5a) is the energy-wise collection area of the instrument, calculated from the simulated γ-ray events that pass the selection cuts. The effective area, along with the livetime (time of observation), are divided from the counts to represent the results in physical units of flux. The rising edge of this IRF, along with the safe range of energy dispersion determines the overall lower energy threshold.



(a) (b)

Fig 2.5: (a) The effective area of the H.E.S.S array as a function of simulated energy. **(b)** A Monte-Carlo derived PSF, normalised to counts binned in θ^2. [17]

- **PSF:** The Point Spread Function (refer to Fig 2.5b) refers to the overall angular spread of a point source mostly due to instrumental and environmental uncertainties. This is approximated by the 68 percent containment radius of the events that pass the ζ_{BDT} cut. The smaller the spread, the better the angular resolution. The differential performance of the classifier/neural network on different event energy makes the containment radius and hence the PSF an energy-dependent quantity.

The selection cuts for which the IRFs are produced are normally chosen based on an optimization process, that optimises for the best significance of the signal with respect to the background, after weighting the simulations for an assumed source spectrum. Since the instrument response varies within the field of view, a set of IRFs for each zenith, azimuth and offset are normally produced that form the lookups that will be referred to when performing an analysis.

For a given bin in true energy ΔE_{true} and an assumed spectral model $\phi(E_{true})$, to get the predicted counts N_{pred} at a given position in the field of view, the above mentioned IRFs at that specfifc position are utilized in the following manner,

$$N_{pred} = N_{bg} + \int_{\Delta E_{true}} AEFF(E_{reco}) \cdot PSF(E_{reco}) \cdot EDISP(E_{true}) \cdot \phi(E_{true}) dE_{true} \tag{2.1}$$

where N_{bg} is the predicted background counts from the background model. The background estimation for a pulsar analysis is briefly described in section 4.1.

Chapter 3

Improving gamma-hadron separation

This section presents the improvement in gamma-hadron separation by adding new parameters to the separation training, starting with information on the classifier being used for the training.

3.1 Boosted Decision Trees (BDT)

A binary decision tree is a machine learning algorithm that repeatedly splits the training instances into two sets depending on a condition, which ultimately minimizes the loss function. A simple schematic of a decision tree is shown in Fig 3.1a. Each split is called a *node*(shown in blue) and each instance ends up on a *leaf node*(green) where no further splitting takes place. Width greater than 0.001 is one example of a decision at one node. The instances are split based on whether the decision is true or false. For each node, all combinations of such decisions are chosen from the parameter space and the one which best minimizes the loss function is assigned. Every instance on a leaf node gets assigned the same score which ultimately decides the class(signal or background) of the leaf node. After a tree is trained, in order to predict the class of a new instance, its parameters are passed through the trained tree and is assigned the class of the leaf node it ends up in.

In the random forest approach, an ensemble of decision trees are trained in parallel and the prediction from each tree is averaged to get the final probability score. In this case, the package used is called **xgboost**[1] which also uses tree boosting. This method used gradient descent starting from a single tree, to determine at each iteration the

[1]xgboost package: `https://github.com/dmlc/xgboost.git`

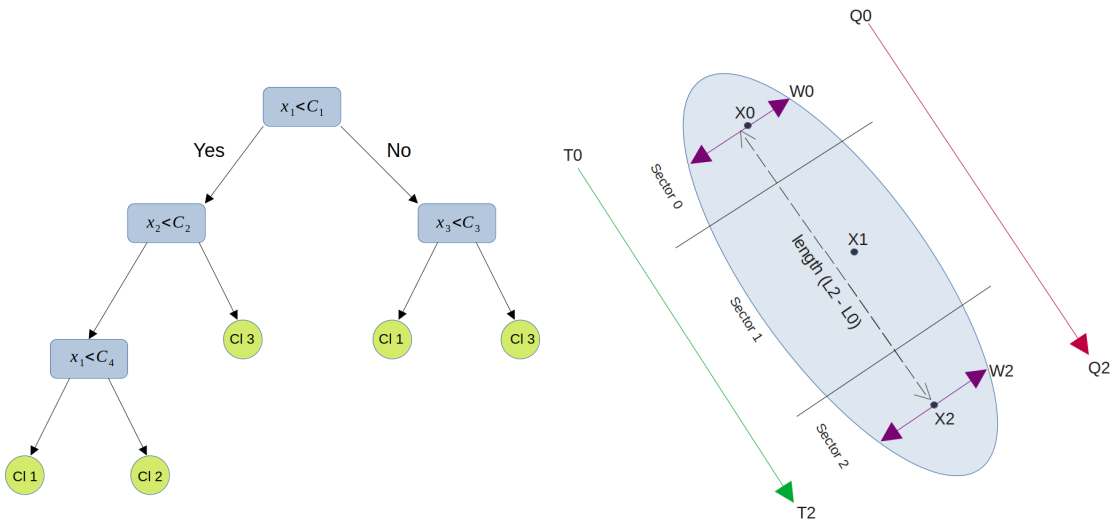


Fig 3.1: (a) Schematic of a simple decision tree. Each node shown in blue represents a condition to be satisfied. The leaf nodes are shown in green. Each represents a single class. **(b)** The definition of the new parameters by dividing the shower image into three sectors.

tree whose output when added gives the best reduction in the loss [21]. The evaluation metric normally employed in binary classification is the binary log loss function given by

$$logloss = \frac{1}{N}\sum_{i=1}^{N} -(y_i * log(p_i) + (1 - y_i) * log(1 - p_i)) \tag{3.1}$$

where N is the number of instances in the training set, y_i is the true label and p_i is the prediction of the classifier.

The xgboost package uses several hyperparameters that set the learning rate, step size and control overfitting. The hyperparameters were chosen using a parameter search in which each hyperparameter was varied separately and those were selected that gave the best evaluation score. Most of the hyperparameters produce little variation in performance for a wide range of values. The distribution of the classifier output is normally compared with both the training and test set to check for signs of overfitting, in which case the test set would perform poorly compared to the training set. All the performance plots(like ROC) referred to in the following sections were plotted using the test set unless otherwise stated.

Further, the BDT training is performed on Monte-Carlo simulation of point sources with an assumed spectral index of -2. The training is carried out in zenith angle bands and the simulated source position at 0.5 degrees offset from the pointing position. The

Parameter name	Definition
Hillas Length l	Length of the major axis of the image
Hillas Width w	Length of the minor axis of the image
abs(Hillas Skewness)	Absolute value of skewness along the major axis
Hillas Kurtosis	Kurtosis along the major axis
Log Density	log (total intensity / (l * w))
Length over log size	l / log(total intensity)

Table 3.1: The definition of the parameters currently used in the training, a.k.a old parameters.

following discussions are specifically based on the zenith angle band of 20 deg, 180 deg azimuth and 0.5 deg offset, but the same can be extended to other zenith angle bands as well (see section 3.7).

3.2 New parameters

The parameters currently in use for gamma-hadron separation in mono are the Hillas parameters, log density (*intensity/length*width*) and length over log size. Refer to Table 3.1 and Fig 2.2b for the definition of these variables. These parameters are defined using the moments of the cleaned shower image. The Hillas length and width relate to the dimensions of the image. The skewness measures the symmetry of the image along the major axis and kurtosis measures the tailedness along the same axis. The BDT will learn the correlations between these parameters for signal and background and hence classify the events.

Although these variables result in a good separation for high-energy events, in the case of low-energy events these are not enough. There are a few problems that need to be overcome when it comes to mono. Since low energy events are only seen by one telescope (CT5), constraining the image parameters is less precise compared to stereo where they are averaged over all the telescopes that see the shower. Low energy events especially are smaller and have low image intensities. Due to this reason, the images look very similar between signal and background and most of the parameter distributions overlap heavily. Fig 3.2 shows the parameter distribution for two different energy ranges. The distributions in the 60-100 p.e. range overlap heavily compared to the 200-300 p.e. range. Having distributions that overlap to a large extent would result in a bad performance of a classifier that relies on decisions based on the distributions.

Parameter name	**Definition**
Time gradient	(T2 - T0)/(X2 - X0)
Charge asymmetry	(Q2 - Q0)/(Q0 + Q1 + Q2)
Length	Length between X0 and X2
Width1	Width of sector 0
Width2	Width of sector 2
Radius	Radius of curvature of X0, X1 and X2

Table 3.2: The definition of the new parameters.

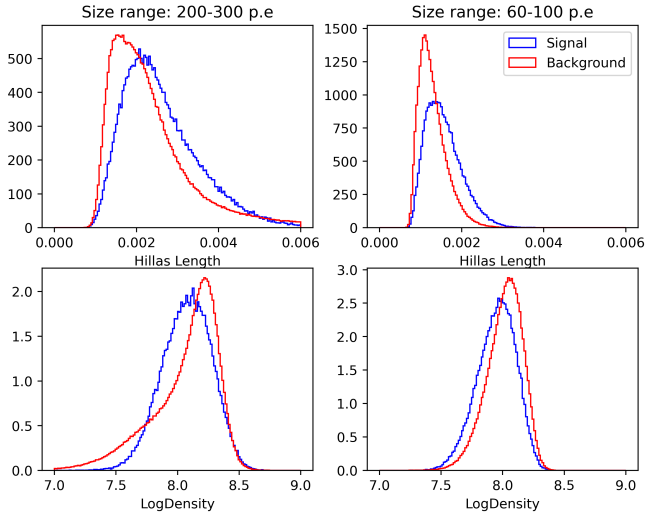


Fig 3.2: Plots showing the parameter distributions for two intensity/size ranges: 200-300p.e. and 60-100p.e.

Another issue is that mono events are contaminated by Cherenkov emission from muons that fell far from the telescope. Muons are produced from the decay of mesons and are a characteristic sign of hadronic showers. The ringed emission from a muon, when far from the telescope resembles an arc that would be hard to distinguish from a *gamma*-ray shower image at low energies. If there were multiple telescopes involved, the muon ring might not be visible in all the telescopes and the event could be rejected, but that is not the case in mono. Therefore, separating a muon ring from a γ-ray shower would further add to the background rejection.

Additional parameters can be added to the separation training that could potentially increase the performance of the classifier. A few new parameters were developed that utilize additional information from the shower including the time of arrival and charge induced in the pixels. These parameters were defined by splitting the shower

image into three sectors based on the Hillas length, as shown in Fig 3.1b. Then the center of gravity of each of the sectors is calculated (X0, X1 and X2). Based on this, the additional parameters are:

- **Time gradient:** Each pixel has a time associated with it when the pixel gets activated from the Cherenkov light. The time gradient is the difference in intensity weighted average time of arrival between the tail and head sector
- **Charge Asymmetry:** Each pixel also has a charge associated with the photomultiplier tube. The charge asymmetry is the difference in average charge between the tail and head sectors, weighted by the total charge in all the sectors respectively.
- **Length:** The distance between the center of gravity of the head and tail sector.
- **Width1:** The width of the tail sector.
- **Width2:** The width of the head sector.
- **Radius:** The radius of curvature of the center of gravities of the three sectors.

An important thing to note here is that since the shower images can have arbitrary directions with respect to the camera/detector coordinates, the distribution of width 1 and width 2 would be meaningless as they could correspond to either the head or tail, without explicitly knowing the direction of the shower propagation. As can be inferred from Fig 3.1b, this is due to the fact the depending on the orientation of the shower image in the camera, sector 0 and sector 2 can correspond to either the head or tail of the shower. Also meaningless are the signs of the time gradient and charge asymmetry due to the same reason. However, the skewness carries the information about the shower propogation as the shower is always skewed in the same manner with respect to the position of the source. Therefore the sign of the skewness can be used to reshuffle the distribution of widths 1 and 2 such that their distributions now correspond to the same part of the shower, i.e. the tail and the head respectively. Similarly, the signs of the time gradient and the charge asymmetry can be changed based on the sign of the skewness.

After the corrections, ideally, the background events should have an uneven distribution of charge and time of arrival compared to signal events. The radius would help separate out events contaminated by muon rings, as they would have a curvature that would otherwise not be present in a γ-ray shower image. These additional parameters are used to improve the performance on two configurations/analysis cuts called *safe*

and *loose*, explained in the next section. The definitions of the new parameters are also summarised in Table 3.2.

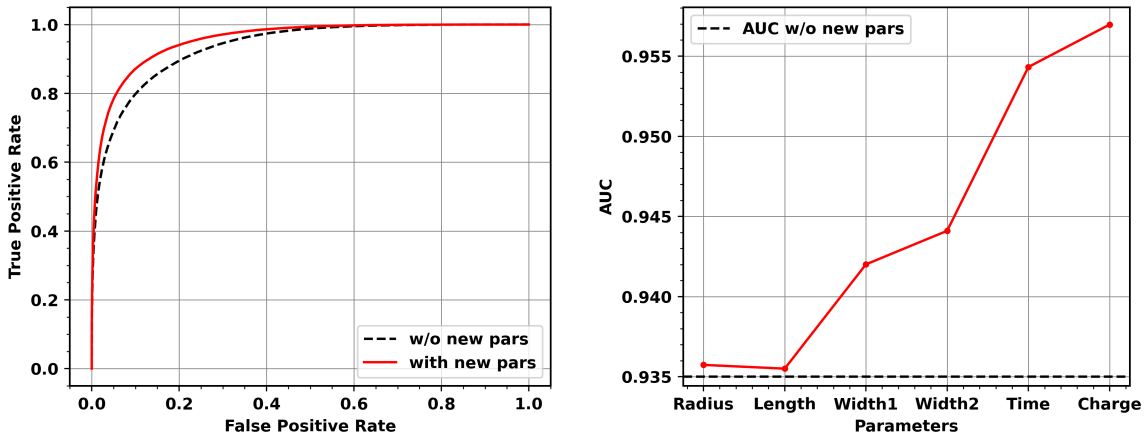
3.3 Safe configuration

A configuration or config in short is a set of analysis cuts that are tuned for a specific type of analysis. For example, the safe configuration contains images above 250 p.e. total image amplitude and a minimum of 10 pixels. Having a 250 p.e. cut would throw away a lot of low-energy events below 100 GeV. So, this configuration is ideal for sources having a spectrum like the crab nebula, with a spectral index of around 2.5. Similarly, the loose config consists of different analysis cuts and would be suited for a different analysis.

The BDT classifier was trained on events from the safe config using two sets of parameters: one with the new parameters and the other without. The performance of the BDT classifier can be estimated using the Reciever Operating Characteristics (ROC) curve that measures the true vs false positive rate (evaluated on the test set) for different classifier output thresholds. The true positive rate is the proportion of positive instances that are classified correctly, which in this case are signal events that are classified as signal. On the other hand, the false positive rate refers to the proportion of negative instances being classified as positive, that is background events being classified as signal. The requirement is an increase in the true positive rate and a reduction in the number of false positives. In the ROC curve, this implies that the curve should be more towards the left-hand corner, and equivalently the maximum area under the curve (AUC).

Fig 3.3a shows the ROC curve of the two classifier models. It is seen that the AUC increases from 0.93 to 0.96 upon adding the new parameters mentioned in the previous section. This indicates that the classifier performs better with the addition of the new variables. Fig 3.3b shows the increase in the AUC as the new parameters are added one at a time in succession. This helps to see which of the new variables contribute most to the increase in AUC. Of all the variables, the time gradient contributes the most to the increased performance. The radius and length do not provide a significant performance boost. The classifier output with the new variables is shown in Fig 3.8a.

Although the performance plots portray well the comparison of the two classifier models, an absolute comparison of the two can be done by estimating the signal efficiency as a function of the classifier output. As seen from Fig 3.5a, there is a general increase in the signal efficiency across the classifier output, for the model trained with the new parameters. Not only does the signal efficiency (ϵ_s) increase,



(a) (b)

Fig 3.3: **(a)** Comparison of the performance of the classifier with and without the new parameters, on the safe config. **(b)** The increase in the area under the curve as parameters are added in succession.

the background fraction (ϵ_b) is also reduced. Both these points are summarised in Fig 3.5b, which shows $\epsilon_s/\sqrt{\epsilon_b}$ also known as Q-value as a function of signal efficiency. For the same value of signal efficiency, the Q-value shows an improvement with the new parameters. The distribution peaks towards low signal efficiency with some statistical fluctuations near zero, implying that cutting on low signal efficiency (hard cuts) could be beneficial in terms of getting the best signal over background significance. Also, having a cut that allows for a higher signal efficiency would still benefit from the reduced background fraction.

An ideal cut would be to have the least amount of background as possible, meaning that the cut needs to be more towards the right of Fig 3.5a. But since the signal efficiency also decreases, reduction in background fraction would have diminishing results toward the extreme right. In order to choose the final cuts for the safe configuration, the MC simulation events are re-weighted to simulate a crab nebula-like spectrum with a spectral index of -2.5 and a normalization of 10 percent. The weighted spectrum is optimized by considering bins in both ζ_{BDT} and θ^2. To get the number of background events within the angular size θ from the simulated source position, a ring is considered around the pointing position and the number of background events remaining after the ζ_{BDT} within the ring is counted. This count is scaled with respect to the ratio of angular size between the ring and the ON region (the region around the simulated source), and the observation time. The significance is calculated using the Li&Ma method which is against the assumption that all the counted events within the ON region are background [22].

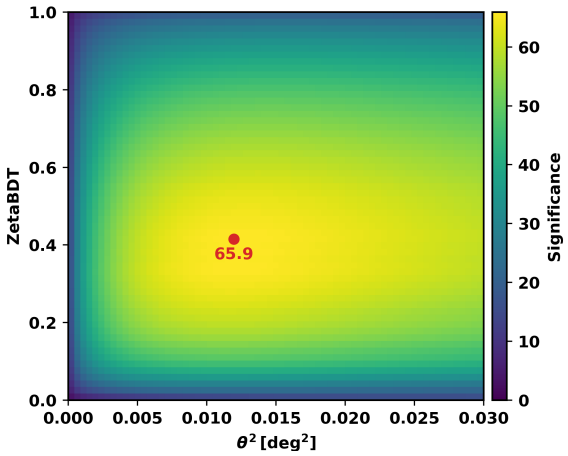


Fig 3.4: The 2D surface of significance as a function of ζ_{BDT} and θ^2. The red dot represents the global maximum of the surface.

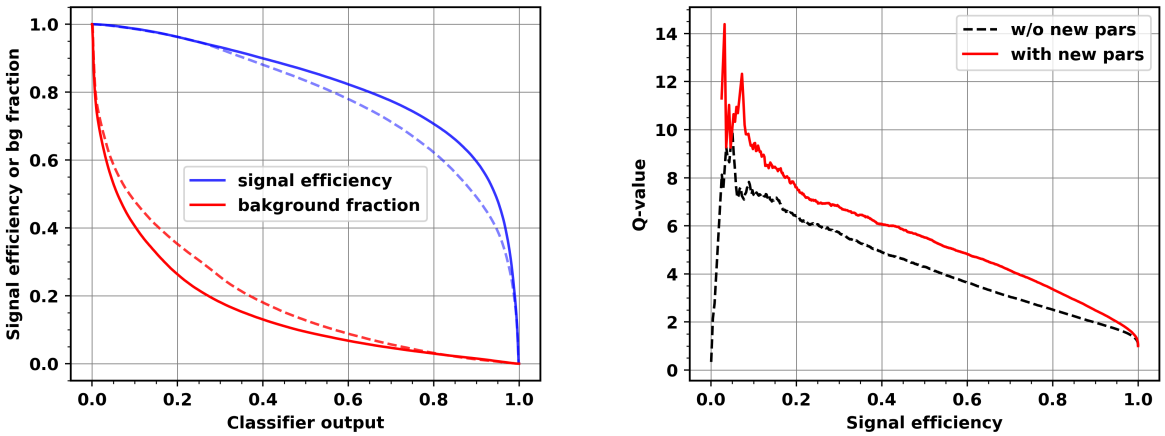
The optimization resulted in a ζ_{BDT} cut of 0.4 and a θ^2 cut of 0.012 deg^2. The significance surface is shown in Fig 3.4 as a function of ζ_{BDT} and θ^2. The peak appears to be rather broad at the optimized value, and therefore choosing slightly different cut values would not reduce the significance by a large amount. Therefore the final cut value for ζ_{BDT} is chosen to be 0.5. The cuts are summarised in table 3.3. The optimized cuts give an overall signal efficiency of around 18 percent.

3.4 Loose configuration

The loose config, as the name suggests has relaxed selection cuts, with a lower image intensity cut of 60 p.e. and a minimum of 6 pixels in the image. This would allow for more low energy signal events but that would also be accompanied by a significant increase in background events. Such a config would be more suitable for sources that have a steep spectrum, like that of pulsars.

The classifier output of the model trained on the loose events without the new parameters is compared to that trained on the safe cuts, in Fig 3.6. The loose model is comparitively worse with a lot of events that are grouped near the middle of the classifier output. This reduction in performance is expected due to the the fact that the events that are extra in the loose config are those events that have an amplitude of 60-250 p.e. including some with a pixel count of less than 10. For such events, the similarity between the signal and background is more, making the classifier perform worse for those events.

To confirm the above-mentioned point, the same loose BDT model can be used to predict the output for two image size ranges, one 60-250 p.e. and the other 250



(a) (b)

Fig 3.5: **(a)** The signal efficiency and the background fraction as a function of the classifier output for the safe config. The solid lines represent the model with the new parameters while the dashed lines represent without them. **(b)** The Q-value as a function of the signal efficiency for the safe config, with the black line representing the model without the new parameters.

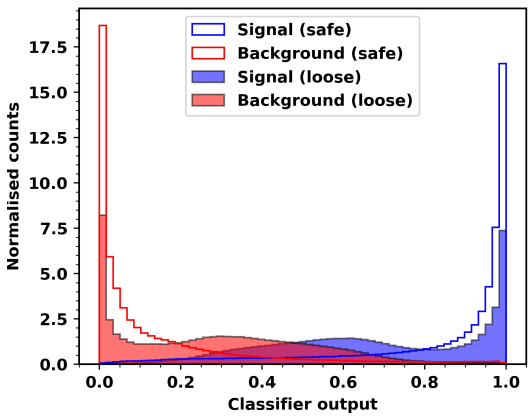
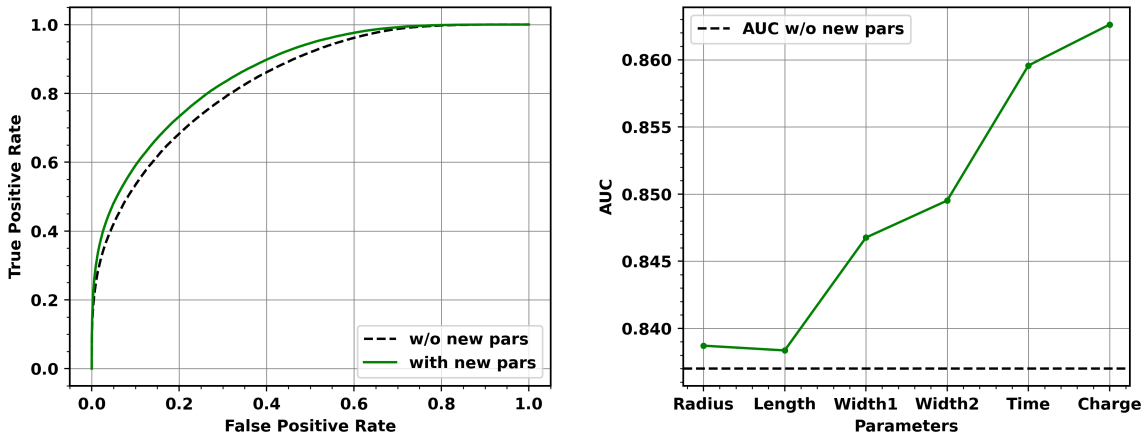


Fig 3.6: The classifier output of BDT trained on the safe (hollow histogram) vs the loose config (filled histogram) with the new parameters.



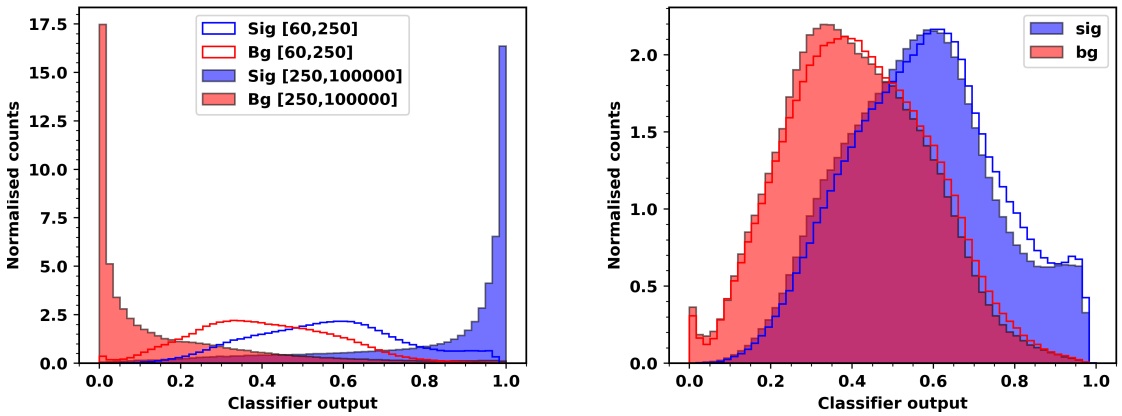
(a) (b)

Fig 3.7: (a) Comparison of the performance of the classifier with and without the new parameters, on the loose config. **(b)** The increase in the area under the curve as parameters are added in succession.

p.e. above. This comparison shown in Fig 3.8a confirms the fact that the events that are grouped near the center are mostly events from the former size range. A question that could be asked here is whether a classifier trained completely on events in the size range of 60-250p.e would perform any better? This point is further discussed in detail in section 3.6. However, the ROC curves comparing the models (see Fig 3.7a)that were trained with the two sets of parameters do show an improvement with the new parameters, although not by a lot, with an increase from 0.84 to 0.86. Looking at which parameter provides the largest increase in AUC (Fig 3.7b), the results are similar to the safe config with the time gradient having the highest effect on the AUC. The radius and length do not have a major effect on the performance.

3.4.1 Low intensity events

The low intensity events that are classified poorly in Fig 3.8 form a second peak in some of the parameter distributions. The distribution of width1 is shown in Fig 3.9a. There is clearly a sharp peak near the value of 0.00058. This feature is also seen in width2 and Hillas Width. It is present in both the signal and background distributions, and curiously near the same value. As the events are selected for a higher cutoff value of amplitude and pixel counts, the peak starts to reduce(Fig 3.9b) and is not present for events with a pixel cut of 15 p.e. Selecting for the exact inverse of the cuts in Fig 3.9a and by fixing the pixel count to less than 11 pixels, the majority of the events that belong at the peak can be seen, and can be confirmed that most



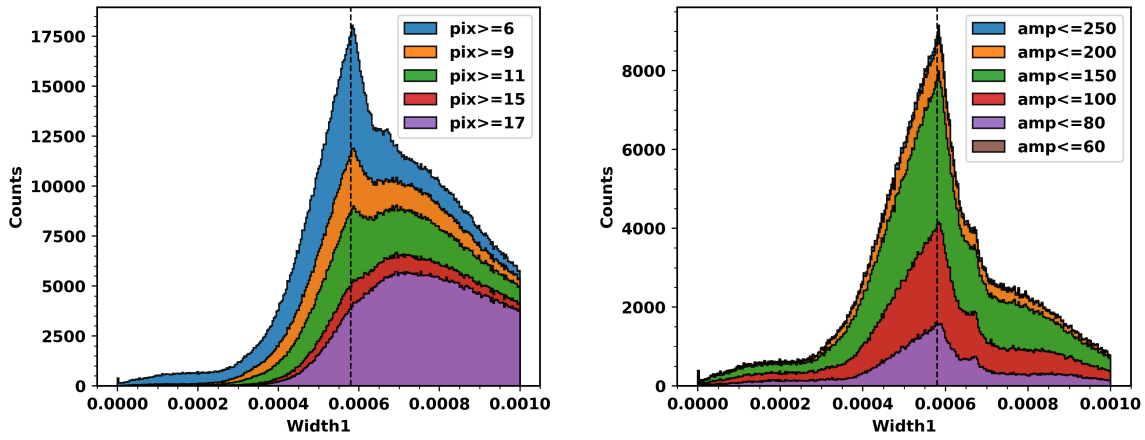
(a) (b)

Fig 3.8: (a) The comparison of the loose model trained with the new parameters on two different size ranges, one 60-250p.e. and the other 250-100000p.e. **(b)** The classifier output for a model trained with the new variables exclusively on events in the size range 60-250p.e. The solid histogram represents the model trained on the whole loose config while the hollow histogram represents model trained in the 60-250p.e. range.

of them are events with an amplitude less than 200 p.e.

The peak is not just a feature of the specific dataset used, but is also present in datasets with different zenith angles and offsets and are roughly located at the same value. Since the parameters are calculated based on the weighted averages (length, width, amplitude etc) over each pixel with respect to the center of gravity, it could be a possibility that having only a few pixels one could end up with values that are very close to each other making the distribution peak at a particular value. The division of the shower into sectors itself would be ambiguous when dealing with a few pixels. However, it is not clear what the physical value of the peak, i.e. 0.00058, actually represents. The sector-wise pixel makeup of events in a small range around the peak is shown in Fig 3.10. It becomes clear that the events in the peak can be classified into two distinct distributions: the events from Fig 3.9b constitute the blue distribution which indeed have a low pixel count in each sector and the rest that have a higher sector-wise pixel count, shown in red.

The above-mentioned points suggest that the peak is likely a geometric effect resulting from the finite size of the pixel. This is not an issue with high-energy events as they have larger images and more number of pixels, but it becomes important for low-energy events where the dimensions of the images are comparable to the size of the pixels themselves.



(a) (b)

Fig 3.9: **(a)** The distribution of Width1 for different pixel cuts, showing a second peak near the value 0.00058. The peak reduces as number of pixels is increased. **(b)** The width1 distribution for events with less than 11 pixels shown for different upper limits of image amplitude.

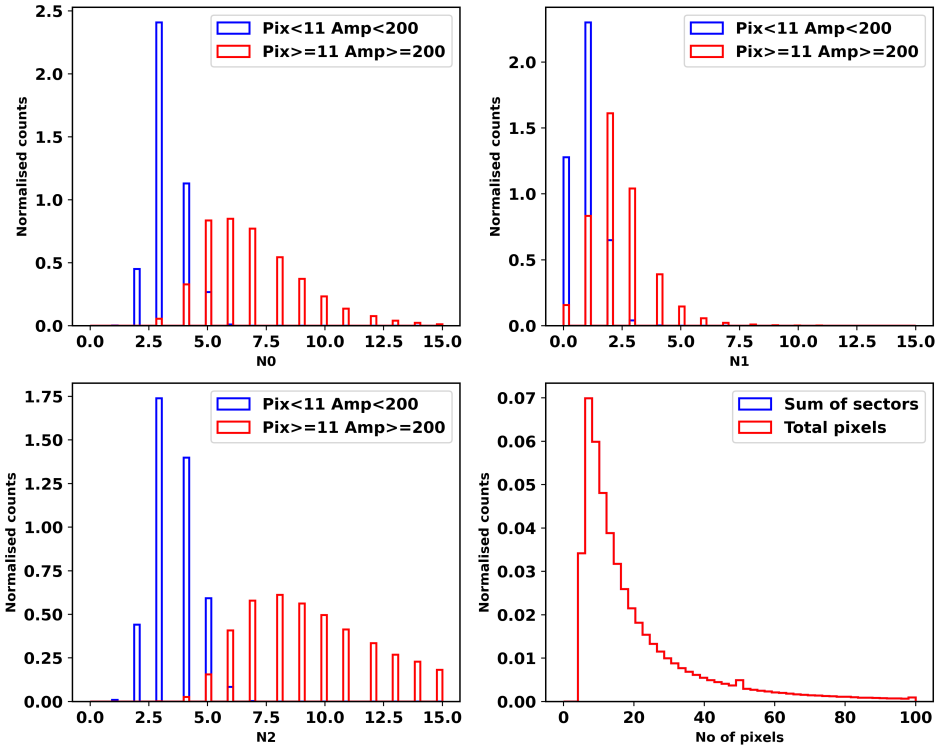


Fig 3.10: Sector-wise decomposition of pixel makeup (N0, N1, N2) of events having width1 between 0.00056 to 0.0006. The pixel makeup is decomposed into two distributions in each sector. The blue histograms show the pixel composition of the events from Fig 3.9b, while the red ones represent the rest of the events in the mentioned width1 range.

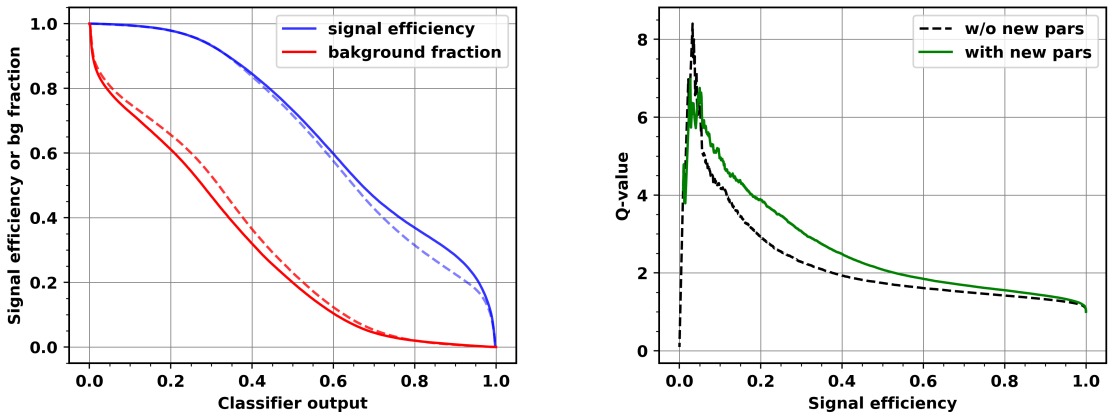
The question now is whether to use these events for the gamma-hadron separation and ultimately for the analysis. Considering gamma-hadron separation, including these events would not be an issue as the peak is present at the same value for signal and background, and is consistent with zenith angles and offsets. Although this would result in a poor classification of these events (as in Fig 3.8), for standard or hard cuts that select for events with say over 0.8 of classifier output, these events will anyhow be excluded.

However, for loose cuts, these events might present an issue. Since they are spread out over a wide range of classifier output, any inconsistency between real data and MC simulations would result in a varying γ-ray efficiency. In other words, since more events have the possibility of being misclassified as signal/background due to the inconsistency, this could result in a larger error. In the case of pulsars though, having more background is affordable due to the pulsed nature of the signal. As will be explained in section 4.1, by knowing the rotation period of the pulsar, it is expected when the pulsed signal belonging to the pulsar would arrive. Based on this, those events that are outside the pulse duration can be considered as background and be excluded from the analysis.

3.4.2 Signal efficiency

The signal efficiency is worse compared to the safe config, as is expected due to the addition of the 60-250 p.e. events that are poorly classified. The background fraction is also higher for the same reason. The new parameters offer increased signal efficiency and reduced background fraction. The trend is similar to the safe config with an increase in Q-value after the addition of the new parameters and the peak of the Q-value distribution is towards the lower side of signal efficiency. However, the Q-value at the peak seems to be higher without the new parameters in the training. The improvement in signal efficiency is more between 0.6 to 1.0 of the classifier output while the reduction in background fraction is more between 0 to 0.6.

The optimization carried out for a steeper spectrum of index -3 shows two peaks in the significance plot in Fig 3.12a: one with a ζ_{BDT} cut value of around 0.3, θ^2 cut of around 0.015 (call it peak 1) and another with a more loose cut of 0.78 for ζ_{BDT} and a θ^2 cut of around 0.05 (call it peak 2). The red dot represents peak 2 which is the global maximum. The two peaks could be explained by observing that the classifier output in Fig 3.6 has two peaks, one around 0.5 and the other at the extreme right. Judging from the signal efficiency curve in Fig 3.11a, peak 2 represents a hard cut on the classifier output of around 0.9, which would retain mostly high-energy events above 250 p.e. Whereas, peak 2 would represent a classifier output cut of about 0.45,



(a) (b)

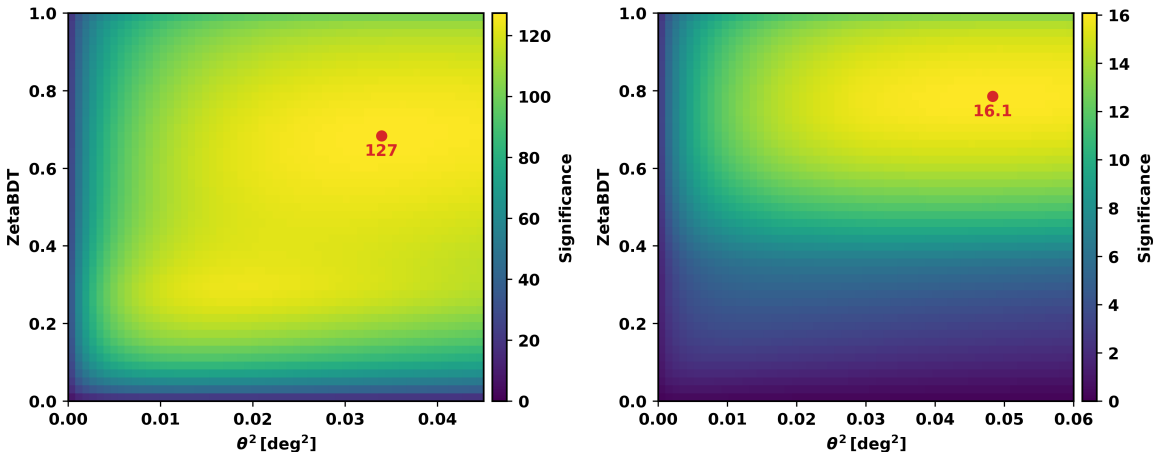
Fig 3.11: **(a)** The signal efficiency and the background fraction as a function of the classifier output for the loose config. The solid lines represent the model with the new parameters while the dashed lines represent without them. **(b)** The Q-value as a function of the signal efficiency. The dashed black line represents the model without the new parameters.

config	**amplitude**	**pixel**	ζ_{BDT}	θ^2
safe	250	10	0.5	0.012
ultraloose	60	6	0.78	0.05

Table 3.3: The finalized selection cuts for safe and loose

that would also retain some of the low energy events within the size range 60-250 p.e. This cut would also inevitably let a lot of the background through, but as discussed in the previous section, would not be a large problem for pulsars when also selecting in the time domain.

For performing a pulsar analysis using this config, it would benefit to optimize for a steeper spectrum that most pulsars exhibit. Fig 3.12b shows the significance plot for a Vela-like spectrum with an index of -4. There is only one peak in this distribution which is near peak 2 of Fig 3.12a. For such a steep spectrum the high energy flux is low and a loose cut that this optimization suggests would be required to keep the low energy events from being thrown away. From this optimization, the cuts were decided to be 0.78 for the ζ_{BDT} and a θ^2 cut of 0.05. The final cuts chosen for the IRF production is summarised in Table 3.3.



(a) (b)

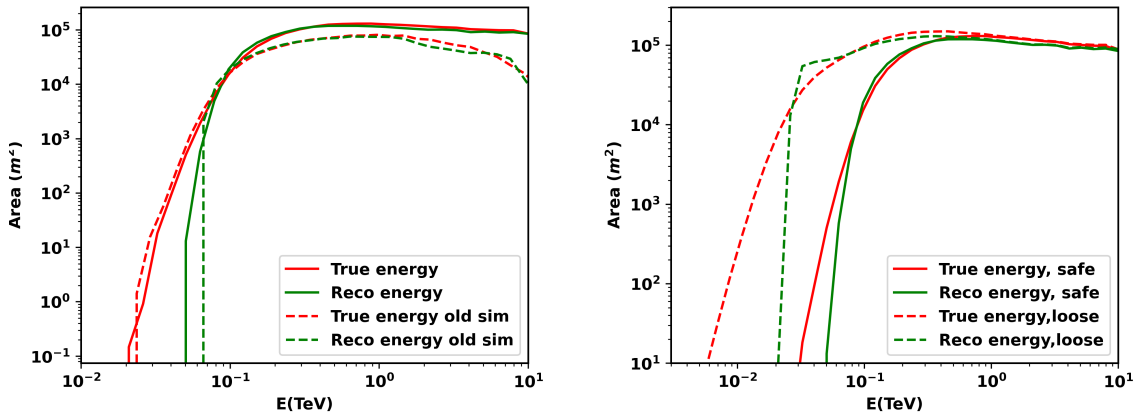
Fig 3.12: The significance plot of the loose config for **(a)** a crab nebula-like spectrum with index -3 and normalization 0.1 at 1 TeV, **(b)** the Vela spectrum with index -4.

3.5 IRFs

The effective area for the safe config with the ζ_{BDT} cut in table 3.3 and without any θ^2 cut is compared to the ones produced previously for older simulations (see Fig 3.13a. The effective area is presented both as a function of true and reconstructed energy. The energy threshold (approx 10 percent of the peak) is slightly higher with the new model, with the difference that the effective area after 100 GeV is higher for the current model. This is directly related to a higher signal efficiency after ∼100 GeV compared to the previous training. The effective area as a function of true energy has a lower threshold than the reconstructed one due to a higher bias in energy reconstruction which reconstructs events at a higher energy than the true energy. Above 100 GeV, the effective area with respect to true and reconstructed energy are consistent with one another.

The effective area for the ultraloose cut on the other hand has a lower threshold (less than 100 GeV) than the safe config. This is because the loose cuts retain more of the low-energy events and have a higher signal efficiency for events in the 60-250 p.e. range. The lower threshold would be beneficial to detect the flux below 100 GeV from pulsars or AGNs. Above 1 TeV the area is more comparable between the two models as both the safe and loose cuts would retain a similar fraction of events at this energy.

The overall angular resolution depends on both the reconstruction and gamma-hadron separation. For the same reconstructed events, if the gamma-hadron cut selects more signal events, then the containment radius would be larger and therefore



(a) (b)

Fig 3.13: **(a)** The effective area after ζ_{BDT} cut as a function of true and reconstructed energy for the safe configuration compared with the old simulations. **(b)** the effective area after ζ_{BDT} cut compared between the safe and loose configuration.

the angular resolution would also be larger. Since the loose cuts (peak 1) select more low energy events that also have poor reconstruction compared to high energy events, the resolution is worse. Whereas in the case of safe cuts which have a relatively small number of low-energy events, the resolution is comparitively better. For energy above 1 TeV, the plots converge as the classifier performance is more or less the same in both the cases and the cuts would select for a similar fraction of events.

The ultraloose cuts at peak 2 allow for even more of the low-energy events to pass through the selection cuts and combined with comparatively worse reconstruction, the containment radius is the highest of all the three cuts below 1 TeV.

3.6 Size Ranges

The model that was trained on events of all sizes in the loose config performs poorly for events in the 60-250 p.e. range, as evident from Fig 3.8a. It could be a possibility that since the classifier has to learn the features for all size ranges its classifing power is not evenly distributed among events of all size ranges. It would be interesting to see whether the classifier would perform better by training on events in a specific size range, and in this case that would be 60-250 p.e. Since the new parameters do help improve the performance, this training is performed including those parameters. The classifier output for the training carried out on these specific events is shown in Fig 3.8b, compared to the prediction for the events in the same size range by a model

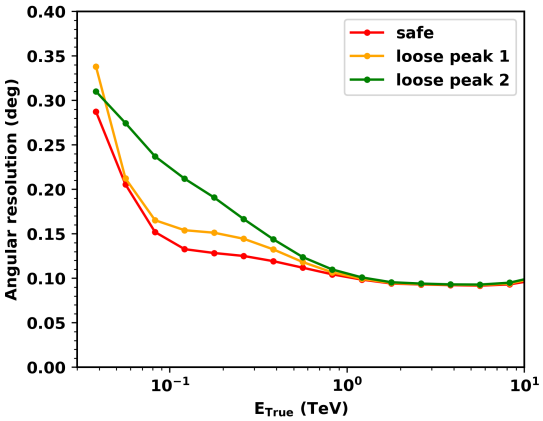


Fig 3.14: The angular resolution as a function of true energy for the safe cuts and the two loose cuts. They represent the 68% containment radius of events that pass the ζ_{BDT} cut.

trained on the events of all sizes in the loose config. The distributions are very similar to each other, indicating that the classifier does not perform significantly better when trained in this size range. The hollow signal distribution does shift slightly towards the right indicating a slight performance improvement, although the peak is still near 0.7. The rest of the events in the size range of 250 p.e and above have already been discussed in section 3.3.

3.7 Performance with distance from zenith

The trend in the performance of the classifier after including the new parameters in training across zenith angles is similar to what was discussed earlier. That is, the new parameters do help improve the performance (see Fig 3.15a) across all zenith angle bands. Further, there seems to be an increase in AUC as the zenith angle increases. This can be associated with the fact that as the zenith angle of observation increases, the Cherenkov light needs to travel a longer distance before reaching the telescope optics. This implies that light from low-energy showers is absorbed more in comparison to high-energy showers, and for the same source spectrum, the energy distribution shifts along the energy axis as the zenith angle increases. Plotting the peak of the energy distribution for different zenith angles shows this shift (Fig 3.15b). Therefore, there are relatively more high-energy events as the zenith angle increases. As seen earlier in Fig 3.2, since the parameter distributions between the signal and background are more separated for high-energy events, the performance increases for higher zenith angles.

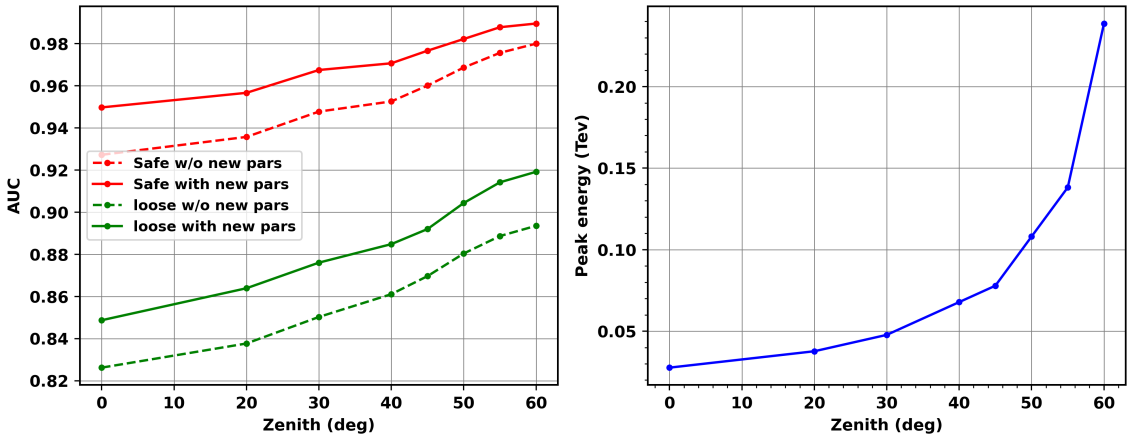


Fig 3.15: Left The performance of the classifier as a function of angle from the zenith. **Right** The shift in the peak value of the true energy distribution, as zenith angle increases.

3.8 Diffuse gammas

The datasets (MC simulations) used for the training are based on point source simulations at 0.5 deg offset and contain shower images that originate from one part of the sky. In H.E.S.S, the source is normally simulated on the x-axis of the camera coordinate system. Due to this reason, the number of shower images would be more on the right-hand side of the y-axis. On the contrary, the background events are isotropic and the events are more or less the same in each quadrant of the camera. In real observations though, the source might not be in the same position at every instance and would certainly not be on the x-axis for the majority of the time. It is important to make sure during training that the BDT does not learn where the events originate from. In other words, if the trained model were to be tested on events originating from an entirely different position in the camera, the performance should not differ. If it does differ, then it would mean the ζ_{BDT} cut would produce varying signal efficiency in different parts or quadrants of the camera.

Diffuse MC simulations can be used to perform this consistency check. These have events that originate from every direction with respect to the camera. Fig 3.16 shows the performance of the BDT trained with point source simulations, on the events from the diffuse simulations divided with respect to each camera quadrant. It can be seen that the performance is not the same in each of the quadrants. Moreover, it seems that the performance of the 1st and 2nd quadrants are comparable to that of the 4th and 3rd quadrants respectively. The performance decreases (the peak is less sharp) for events from the left side of the y-axis. This would imply that the classifier is learning that the events coming from Q1 and Q4, which is where the MC

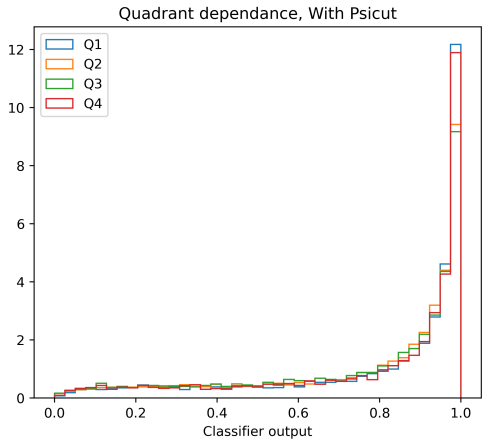
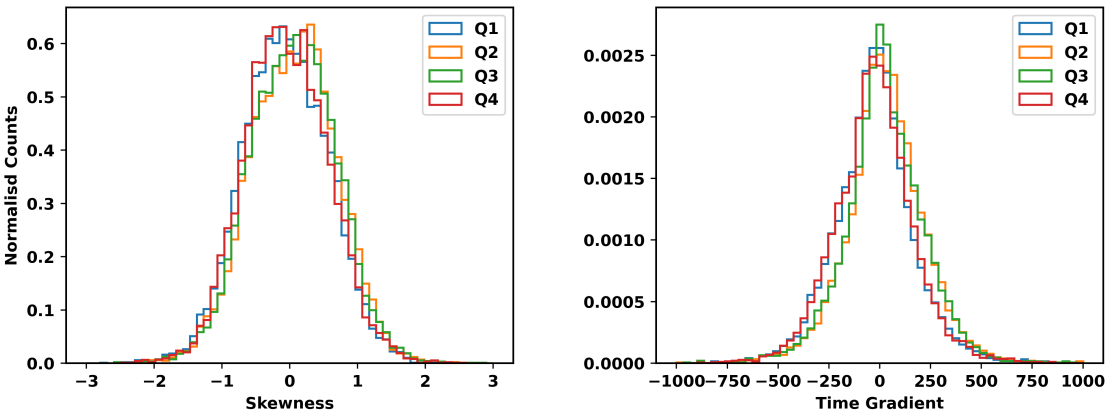


Fig 3.16: The classifier trained on point source simulations being evaluated on showers from different camera quadrants from the diffuse simulations. The performance is quadrant dependent. The psicut here refers to events selected between an angular size of 0.4 to 0.6 deg

simulations are based, are more likely to be signal and get a higher probability score whereas relatively more events from Q2 and Q3 are classified with a lower probability.

One way the quadrant dependence could arise would be due to the inclusion of some parameter in the training that could directly relate to the location of the source. For example, the center of gravities of the images are not used in training as the classifier could directly learn the position of the source. But, from the definition of the variables, none of them have a dependence on the location in any way so this cannot be the issue. Since the BDT classifier uses the parameters for the training, if any parameter happens to be quadrant-dependent, then this could result in a quadrant-dependent performance as the parameter distribution of the simulated source would mostly be localized to two quadrants. Upon using the diffuse simulations to check for this, the parameters that seem to have a quadrant dependence are the time gradient, charge asymmetry, Width1, Width2 and skewness.

Apart from skewness, the rest of the parameters are not inherently quadrant-dependent, but when the skewness is used to flip the sign of the distribution of say the time gradient, the resulting distribution does vary with the quadrant. The distribution of skewness, and the time gradient after flipping using skewness is shown in Fig 3.17. The skewness peaks at two different values depending on the quadrants. For both the plots in Fig 3.17, two distinct distributions seem to be separated, with Q1 and Q2 having the same distribution as Q4 and Q3 respectively, which is similar to the trend seen in the classifier output. Since the BDT is trained on events whose showers are more in Q1 and Q4, the classifier would predict with a lower probability



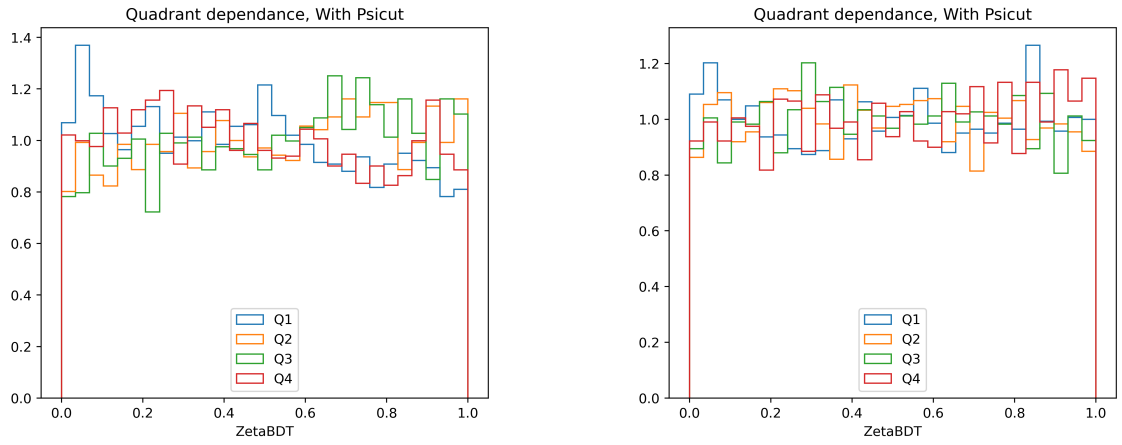
(a) (b)

Fig 3.17: (a) The skewness distribution based on camera quadrants for events within an angular cut of 0.4 to 0.6 deg. It peaks at two different values depending on the quadrant. **(b)** The time gradient distribution after flipping the sign using the skewness also showing a similar quadrant dependence.

for showers from Q2 and Q4 due to them having a slightly different distribution.

The effect the quadrant dependence can have on the ζ_{BDT} distribution is shown in Fig 3.18a, for a model trained with the Hillas Parameters and only the time gradient and with skewness used to change its sign. The signal efficiency varies between the two halves of the camera for a ζ_{BDT} value of around 0.8. Comparatively, Fig 3.18b shows the ζ_{BDT} distribution for a model with the same parameters but the only difference being that the absolute value of time gradient is used instead of using the skewness to flip the sign. This distribution does not exhibit a quadrant dependence like before. Although the skewness distribution varies with quadrant, the absolute value of skewness is quite consistent across the four quadrants due to which after the addition of the absolute value of time gradient to the old set of parameters, the ζ_{BDT} distribution is not expected to be quadrant dependent.

The systematic uncertainty in the flux measurements is estimated to be around 20 percent [17]. This uncertainty arises due to contributions from various sources of error like the MC atmospheric simulations, MC shower interaction models, the effect of missing pixels, etc. Although the variation of signal efficiency across the quadrants is not ideal, it is within the uncertainty limits. Therefore, for the purpose of this analysis, the distributions are flipped using the sign of the skewness.



(a) (b)

Fig 3.18: The ζ_{BDT} distribution of classifier trained with **(a)** Hillas parameters and time gradient with signs flipped using skewness **(b)** Hillas parameters and the absolute value of time gradient.

Chapter 4

Application to the Vela pulsar

The following section involves the application of the improved training of the loose config on the Vela pulsar. The datasets used contain observations/runs from two periods: dataset 1 with runs from 2019 - 2020 and dataset 2 with runs from 2022 (refer Table 4.1). The datasets have been processed using the BDT model from the new training with a ζ_{BDT} cut of 0.78 (ultraloose cut). The datasets contain the selected events with each event having the information of reconstructed energy, direction, time of arrival and several other parameters. They also contain the IRFs that were calculated for the ζ_{BDT} cut . The **gammapy** [1]package is used for the analysis of the datasets.

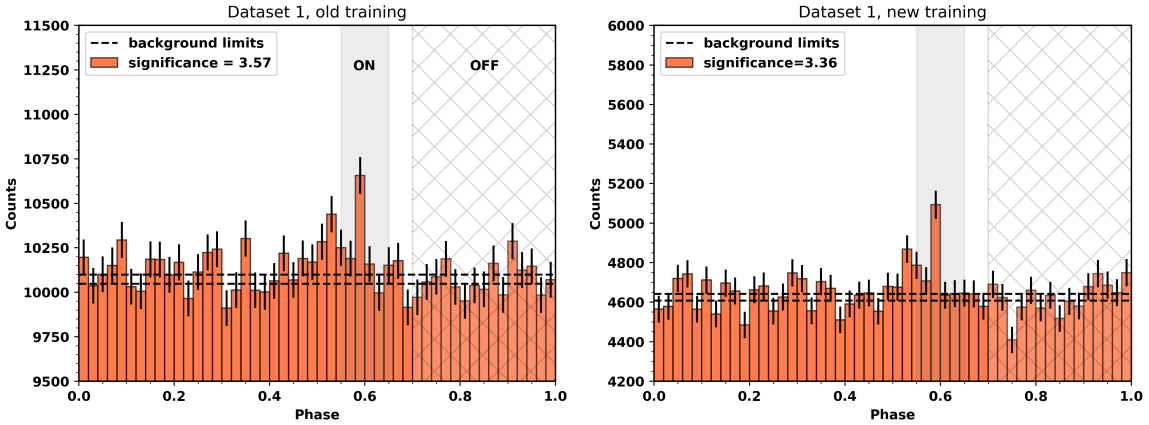
Dataset Name	**Year**	**Trigger threshold**
Dataset 1	2019-2020	nominal
Dataset 2	2022	nominal
Dataset 3	2021	reduced

Table 4.1: Time of observation of datasets used and their trigger threshold

4.1 Phasogram

The signal arrives at short intervals from the pulsar separated by a duration equal to the rotation period of the pulsar. The events recorded outside this interval are considered entirely background events. By knowing the period and period derivative of the pulsar, each event can be assigned a quantity called *phase* having a value between 0 to 1. The phase interval that contains the pulsed emission is called the ON

[1]gammapy package: `https://github.com/gammapy/gammapy`



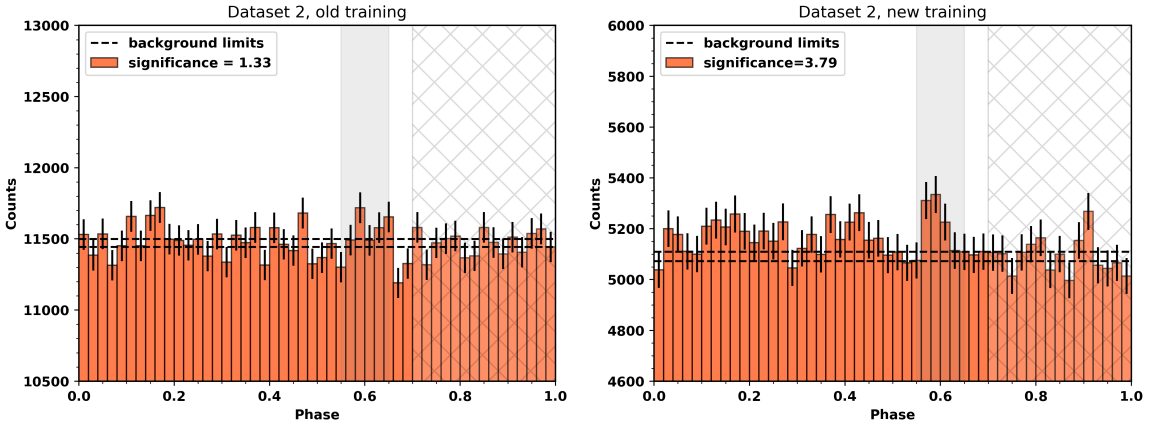
(a) (b)

Fig 4.1: (a) The phasogram of Vela from dataset 1 with the old training. Overlayed on top are the ON and OFF regions, with the horizontal lines representing the upper and lower limits of the background counts estimated from the OFF region . **(b)** Phasogram of Vela from dataset 1 with the new training.

region and the interval that contains only the background counts is called the OFF region (refer Fig 4.1a). The phase can be thought of in the same way any rotating body could be characterized by a phase angle. The signal from the pulsar would arrive in phase whereas the background is arbitrary. To get better statistics on the signal, the events with the same phase are stacked on one another in a process called phase folding. With enough statistics, any pulsed signal would be visible as a peak with respect to the background.

The phase folding is carried out using a package called **PINT** [2] which uses an ephemeris file containing the known information of the period and period derivatives of the pulsar. The histogram of the phases called the phasogram for dataset 1 for events selected within a radius of 0.2 deg from Vela's position is shown in Fig 4.1b. Having a large radius would ensure that most of the signal events are selected. The sharp peak from the pulsar is visible at a phase value slightly less than 0.6. For comparison, the phasogram obtained using the training previously used in H.E.S.S on the same dataset is shown in Fig 4.1a. The peak is at the same phase for both datasets but there are some differences. The average counts from the new training is almost half as much as the old training. In addition, the fluctuations are more with the old training. In the case of dataset 2, the peak from the new training (see Fig 4.2) is also near the same phase as before but can be seen more clearly compared to the old training where the peak is not very evident.

[2]PINT package : `https://github.com/nanograv/PINT`



(a) (b)

Fig 4.2: The phasogram of Vela from dataset 2 with **(a)** the old training **(b)** the new training.

The ON region also contains background counts that need to be estimated to know how significant the signal is compared to the background. The number of background events N_{off} can be estimated from the OFF region where it is known beforehand that it contains no signal from the pulsar. Since the background events occur at random, the number of background events in a given phase interval is supposed to be the same wherever this interval is considered in the phasogram. If α is the ratio of the ON region to the OFF region, the number of excess counts within the ON region can be calculated using the below equation,

$$N_{excess} = N_{on} - \alpha N_{off} \tag{4.1}$$

The significance obtained for the ON phase range (0.55-0.65) and OFF phase region (0.7-1.0) for dataset 1 having calculated N_{excess} and N_{off} is 3.4 σ above the background with the new training which is slightly lower in comparison to the old training in which case the significance is 3.6 σ. The older training seems to be performing marginally better in this case. The difference is most likely due to statistical fluctuations and the choice of the ON region. For dataset 2 the significance is 1.3 σ with the old training but is 3.8 σ with the new one. In this case, the new training seems to be beneficial.

One thing to note is that for around half the number of observations in dataset 2 the zenith angle of observation is over 40 deg. Low energy events are more likely to get absorbed in the atmosphere before reaching the telescope for large zenith angles. For the ones that do reach, the performance of the gamma-hadron separation

would influence whether or not those events get selected. From the significances and phasograms in Fig 4.2 it looks like the new training performs better at larger zenith angles of observation.

4.2 Spectrum

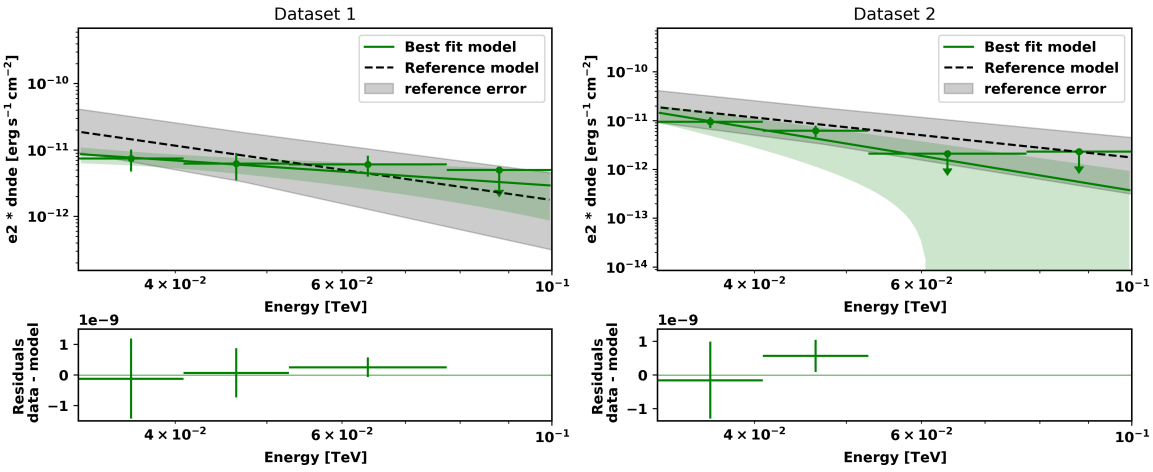
The counts binned in reconstructed energy need to be now related to the flux to obtain a spectrum as a function of energy. To perform this, a spectral model is assumed which in this case is a power law of the form

$$dN/dE = \phi_0(E/E_0)^{-\Gamma} \tag{4.2}$$

with the free parameters being the spectral index Γ and the flux normaliation ϕ_0, for a reference energy E_0 which is fixed. The IRFs are then used to convert the flux units to counts as a function of reconstructed energy using the effective area, livetime (observation time) and energy dispersion (Eq 2.1). The counts from the model are fit to the dataset by varying Γ and ϕ_0 and maximizing the likelihood ratio with respect to the null hypothesis assumption. This fitting provides a global spectral model.

The actual flux points might vary individually with respect to the global model. To account for this the same model is then re-fitted by fixing the Γ obtained from the previous fit and then fitting only for the normalization ϕ_0 for resampled energy bins. The resulting normalizations per energy bin are the flux points.

The spectrum with the global model and the flux points for dataset 1 and 2 is presented in Fig 4.3 with its error band. Overlayed on top is the error margin including systematic uncertainties and the global model from the previous H.E.S.S Vela detection [11]. From both the plots it can be seen that although the fitted spectral model is not the same as the previous model, they are well within the previous uncertainty limits. The last two flux points in Fig 4.3b suffer from a lack of statistics resulting in only upper limits. Although the difference between the models could be resulting from the new training, it is currently not straightforward to identify the exact reason due to the fact that there are multiple systematic uncertainties involved like having a newer simulation, a different IRF and energy correction scheme, etc. It is however a good indication that both the models are within the uncertainty margin of the previous detection, including most of the flux points. The spectrum obtained by stacking the two datasets is presented in Fig 4.4a with better statistics, and is well within the uncertainty margin except for the first flux point. This could be due to the effective area being very steep near the threshold resulting in larger fluctuations.



(a) (b)

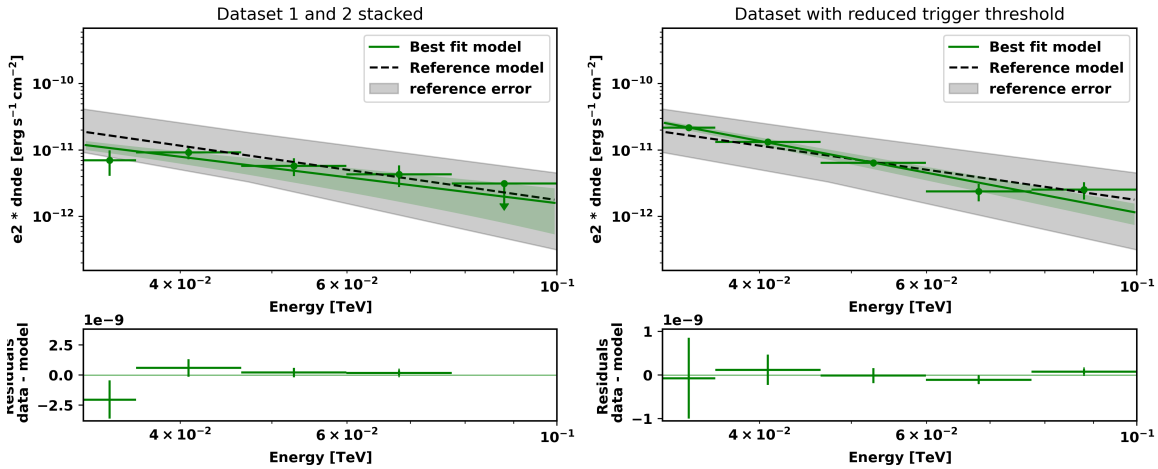
Fig 4.3: (a) The spectrum of Vela obtained from dataset 1. The solid green line is the global best-fit model with its error bands, the dotted black line is the H.E.S.S best-fit model from [11] with its error bands in grey. The green dots are the flux points. **(b)** The spectrum of Vela obtained from dataset 2.

4.3 Reduced threshold dataset

This section presents the application of the new training on a Vela dataset that contains observations recorded at a reduced trigger threshold (Dataset 3). The lower trigger threshold enables more low-energy events to trigger the telescope and get recorded. The same training and ultraloose ζ_{BDT} cuts were applied in the production of the dataset, although the training was performed for nominal threshold MC simulations.

The phasogram for this dataset is compared between the new and the old training (also trained with nominal threshold) in Fig 4.5. For both the trainings, the peak is quite significant compared to the previous two datasets. This difference can majorly be attributed to the reduced threshold as the duration of observation is comparable between this dataset (37.5 hours) and the total combined duration of dataset 1 and 2 ($\sim$ 30 hours). The peak consists of a rising edge and a sharp cutoff after the peak, characteristic of the Vela pulse profile seen in Fig 1.3.

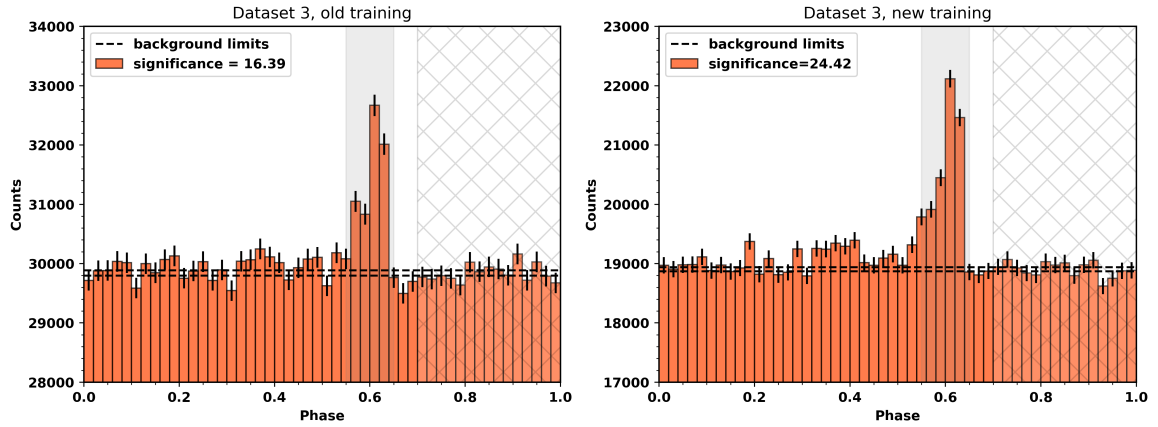
There is also a considerable difference between Fig 4.5a and Fig 4.5b . One, the significance of the pulse in the ON phase range is comparatively greater with the new training, with 16.4 σ with the old training and 24 σ with the new one. The peak is more sharp with the rising edge of the peak having more structure. Second, the other two features of the pulse profile mentioned in section 1.4 can now be faintly seen in



(a) (b)

Fig 4.4: **(a)** The spectrum obtained by stacking datasets 1 and 2 **(b)** The spectrum obtained from the dataset with a reduced trigger threshold.

the phasogram above the background. These two features are the P1 peak (refer Fig 4.5b) and the bridge emission between the main peak (P2) and P1. These features have been previously observed in the γ-ray domain by space-based telescopes like the Fermi-LAT, but this is the first time hints of this emission can be seen with an IACT.



(a) (b)

Fig 4.5: The phasogram of Vela from the reduced trigger threshold dataset processed with the old training and compared with the new training. Hints of peak P1 and bridge emission seen for the first time with an IACT.

The spectrum from this dataset with the new training is presented in Fig 4.4b. The best-fit model and the flux points are well constrained within the previous uncertainty limits, showing an agreement with the previous model. Any difference between the

current and previous model could arise from the fact that dataset 3 was processed with the new training performed on simulations with a nominal trigger threshold. This could add to the error from reconstruction and gamma-hadron separation and it is unclear how this would affect the systematic uncertainities. However, from the spectrum, it can be said that the effect is not severe.

Chapter 5

Conclusion

The boosted decision tree classifier was trained on point source MC simulations of γ-ray showers and OFF runs for cosmic ray showers using two sets of parameters, one including the new variables and the other without. The training performed on events from two configurations, safe and loose showed an increase in performance with the addition of new parameters. The increase in area under the curve was more for safe compared to loose. In both cases, the most important parameter among the new ones turned out to be the time gradient. The increase in performance upon the inclusion of the new variables was consistent across the zenith angle bands for both configurations.

For the loose config, the events in the 60-250p.e. range were grouped around a second peak in the distribution of all three widths, and for both signal and background. Evidently, the classifier performed poorly on these events. Further, a classifier trained specifically in this size range also did not improve the performance by a large margin. The low pixel and amplitude events in this size range are most likely the reason for the poor performance due to an ambiguity in properly defining some of the parameters on them. The IRFs produced from this config for the ultraloose cut reached a threshold lower than 100 GeV with a large containment radius in comparison to the safe cuts.

The application of the models on diffuse MC simulations revealed that the performance was quadrant dependent, with the classifier performing the best on the quadrants where the point source MC simulations were based on. The ζ_{BDT} distribution showed varying signal efficiency with respect to the quadrants. This quadrant dependence arose due to the skewness parameter having a different distribution on the left and right sides of the y-axis. When used to determine the sign of the time gradient, the skewness induced a quadrant dependency similar to the one seen in the classifier output. The ζ_{BDT} variation was within the 20 percent systematic uncer-

tainty estimated in H.E.S.S. Using the absolute value of time gradient instead showed that the quadrant dependence on the ζ_{BDT} could be reduced. Although parameters with the skewness flip werr used for the purpose of this analysis, using the absolute value or a quantity other than skewness that does not have a quadrant dependence to flip the signs might be beneficial to reduce the systematic errors.

The phasogram of the Vela pulsar using datasets processed with the ultraloose ζ_{BDT} cut and taken at nominal trigger threshold showed that the new training performs either similar or better in comparison to the old training. The significance of the pulse from Dataset 1 using the new training was comparable to the training previously done in H.E.S.S. In addition, the fluctuations in the number of counts across different bins seemed to have reduced. A different choice of the ON phase region might produce different results. With Dataset 2 on the other hand the peak was over twice more significant with the new training. Dataset 2 having many runs at a higher zenith, the result from the new training might suggest that it can handle large zenith angles better.

The reduced trigger threshold dataset produced the most significant pulse in comparison to the other datasets. Moreover, the pulse was more significant with the new training. The new training was also beneficial in capturing more structure of the main peak. Two other features that were visible in the Fermi-LAT phasogram, i.e. the P1 peak and the bridge emission were now faintly visible above the background, observed for the first time with an IACT.

The fitted spectrum in the energy range of 30 to 100 GeV for Datasets 1 and 2 were in agreement with the previous H.E.S.S detection, within the uncertainty limits from [11]. It is currently not clear the reason for the variation in the fitted models due to some processes in the analysis having changed since the previous detection. The spectral model from the reduced threshold dataset was the best-constrained within the said uncertainty limits compared to all the datasets. Judging from this spectrum, using the training performed on nominal threshold simulations does not seem to have a significant effect on the performance, but this needs to be further investigated.

Including the low intensity events for the analysis does produce a spectrum within the uncertainity limits. By using a background estimation technique that relies on the timing information, low-energy signal events can be separated from low energy background events despite having a lot more background events after the ultraloose cut. But the effect of any inconsistency in the gamma-hadron separation, due to the inclusion of these events, on any other background estimation technique needs to be explored by using a different background estimation technique or performing an analysis with the ultraloose cuts on a different source. Lastly, it might also be

beneficial to further reduce the image amplitude cut (<60 p.e.) and the number of pixels (<6) followed by the optimization of training to detect more low-energy events from pulsars.

Bibliography

[1] A. Apanasenko and et al, "*Composition and energy spectra of cosmic-ray primaries in the energy range 1013–1015 eV/particle observed by Japanese–Russian joint balloon experiment*", Astroparticle Physics **16**, 13 (2001).

[2] S. Ansoldi and et al, "*Teraelectronvolt pulsed emission from the crab pulsar detected by magic*", A&A **585**, A133, A133 (2016).

[3] J. Hinton and W. Hofmann, "*Teraelectronvolt Astronomy*", Annual Review of Astronomy and Astrophysics **47**, 523–565 (2009).

[4] T. K. Gaisser, R. J. Protheroe, and T. Stanev, "*Gamma-Ray Production in Supernova Remnants*", The Astrophysical Journal **492**, 219 (1998).

[5] J. Truemper, W. Pietsch, C. Reppin, W. Voges, R. Staubert, and E. Kendziorra, "*Evidence for strong cyclotron line emission in the hard x-ray spectrum of hercules x-1.*", Astrophysical Journal **219**, L105 (1978).

[6] P. Goldreich and W. H. Julian, "*Pulsar electrodynamics*", Astrophysical Journal **157**, 869 (1969).

[7] A. Philippov and M. Kramer, "*Pulsar Magnetospheres and Their Radiation*", Annual Review of Astronomy and Astrophysics **60**, 495 (2022).

[8] J. K. Daugherty and A. K. Harding, "*Gamma-ray pulsars: emission from extended polar cap cascades*", Astrophysical Journal **458**, 278 (1996).

[9] K. S. Cheng, C. Ho, and M. Ruderman, "*Energetic radiation from rapidly spinning pulsars. i. outer magnetosphere gaps*", Astrophysical Journal **300**, 500 (1986).

[10] D. R. Lorimer and M. Kramer, *Handbook of Pulsar Astronomy*, Vol. 4 (2004).

[11] H. E. S. S. Collaboration and et al, "*First ground-based measurement of sub-20 gev to 100 gev γ-rays from the vela pulsar with h.e.s.s. ii*", A&A **620**, A66, A66 (2018).

[12] D. J. Thompson, C. E. Fichtel, D. A. Kniffen, and H. B. Ogelman, "*Sas-2 high-energy gamma-ray observations of the vela pulsar.*", The Astrophysical Journal **200**, L79 (1975).

[13] A. A. Abdo, M. Ackermann, and et al, "*THE VELA PULSAR: RESULTS FROM THE FIRST YEAR OF FERMILAT OBSERVATIONS*", The Astrophysical Journal **713**, 154–165 (2010).

[14] H. et al, "*Discovery of a radiation component from the vela pulsar reaching 20 teraelectronvolts*", *Nat Astron* **7**, 1341–1350 (2023).

[15] H. J. Völk and K. Bernlöhr, "*Imaging very high energy gamma-ray telescopes*", Experimental Astronomy **25**, 173–191 (2009).

[16] M. de Naurois and D. Mazin, "*Ground-based detectors in very-high-energy gamma-ray astronomy*", Comptes Rendus Physique **16**, 610 (2015).

[17] F. Aharonian and et al, "*Observations of the crab nebula with hess*", A&A **457**, 899 (2006).

[18] R. D. Parsons and J. A. Hinton, "*A monte carlo template based analysis for air-cherenkov arrays*", Astroparticle Physics **56**, 26 (2014).

[19] A. M. Hillas, "*Cerenkov light images of eas produced by primary gamma rays and by nuclei*", in 19th international cosmic ray conference (icrc19), volume 3, Vol. 3, International Cosmic Ray Conference (Aug. 1985), p. 445.

[20] S. Ohm, C. van Eldik, and K. Egberts, "γ*/hadron separation in very-high-energy* γ*-ray astronomy using a multivariate analysis method*", Astroparticle Physics **31**, 383 (2009).

[21] T. Chen and C. Guestrin, "*Xgboost: a scalable tree boosting system*", arXiv e-prints, arXiv:1603.02754, arXiv:1603.02754 (2016).

[22] T. P. Li and Y. Q. Ma, "*Analysis methods for results in gamma-ray astronomy.*", Astrophysical Journal **272**, 317 (1983).