Detection of the Crab Nebula with the first TAIGA IACT

Dissertation

zur Erlangung des Doktorgrades
an der Fakultät für Mathematik, Informatik und
Naturwissenschaften
Fachbereich Physik
der Universität Hamburg

vorgelegt von

Michael Blank

Hamburg
2023

Gutachter der Dissertation:	Prof. Dr. Dieter Horns
	Prof. Dr. Yvonne Becherini
Zusammensetzung der Prüfungskommission:	Prof. Dr. Dieter Horns
	Prof. Dr. Yvonne Becherini
	Prof. Dr. Michael Potthoff
	Prof. Dr. Marcus Brüggen
	Dr. Martin Tluczykont
Vorsitzender der Prüfungskommission:	Prof. Dr. Michael Potthoff
Datum der Disputation:	18.08.2023
Vorsitzender des Fach-Promotionsausschusses PHYSIK:	Prof. Dr. Günter H. W. Sigl
Leiter des Fachbereichs PHYSIK:	Prof. Dr. Wolfgang J. Parak
Dekan der Fakultät MIN:	Prof. Dr.-Ing. Norbert Ritter

Abstract

The recent observation of gamma rays close to PeV energies has opened up a new, yet unexplored window to the sky. To identify the sources capable of producing such highly energetic particles and to understand the underlying acceleration mechanism, continued observations at energies above 100 TeV are a promising field of study. Instruments designed to observe at these energies require high sensitivity. The TAIGA experiment aims to implement a hybrid detection technique of cosmic and gamma rays at TeV to PeV energies. It combines the Cherenkov light sampling array HiSCORE with Imaging Air Cherenkov Telescopes (IACTs) and seeks to introduce a new reconstruction technique that complements the angular and shower core resolution of HiSCORE with the gamma-hadron separation power of the imaging telescopes.

In this work, a simulation and analysis chain of the first TAIGA IACT starting from raw data and concluding with a high-level spectral reconstruction of Crab Nebula data of only the first IACT, taken between October 2019 and December 2020 is presented.

A good agreement of higher-level image parameters with simulation is demonstrated. Using the image parameters, origin reconstruction, energy estimation, and background suppression were implemented with Random Forests, a commonly used machine learning technique. Overall, an angular resolution of $<0.2^\circ$ between 10 TeV and 100 TeV degrading towards higher and lower energies was achieved. The energy resolution is $<25\,\%$ and the bias is $<5\,\%$ between 10 TeV and 100 TeV. For the background suppression with the Random Forest, a quality factor of 4 for a Crab-like spectrum was achieved. Above 30 TeV, a quality factor of 6 was reached.

Together with the instrument response functions from the Random Forest analysis, calibrated, cleaned, reconstructed, and background suppressed events were used in the higher-level analysis with Gammapy. 122 h of good-weather-selected Crab Nebula data taken in wobble mode resulted in 204.6 excess events at a significance of 9.7σ for an analysis threshold of approximately 6 TeV. A forward-folding power law fit resulted in a flux of $\phi_0 = (4.3 \pm 0.5) \times 10^{-10}\,\mathrm{m}^{-2}\,\mathrm{s}^{-1}\,\mathrm{TeV}^{-1}$ at a reference energy of 11 TeV with a photon index of $\Gamma = 2.97 \pm 0.15$.

A rough estimate of the systematic uncertainties was attempted and resulted in an uncertainty of the energy calibration of $\sim 20\,\%$.

Kurzfassung

Die Beobachtung von Photonen nahe an PeV-Energien öffnete ein neues, bis jetzt wenig erforschtes Feld. Weitergehende Beobachtungen oberhalb von 100 TeV stellen ein vielversprechendes Forschungsfeld dar, um jene Quellen zu finden, die fähig sind, solch hochenergetische Teilchen zu erzeugen, und um die zugrundeliegenden Beschleunigungsprozesse zu verstehen. Instrumente, die für Beobachtungen in diesem Energiebereich entworfen werden, erfordern eine hohe Sensitivität. Das Ziel des TAIGA Experiments ist es, eine hybride Detektionsmethode für Gamma- und kosmische Teilchen im Bereich von TeV bis PeV zu realisieren. Das Experiment kombiniert HiSCORE, eine Anordnung von optischen Stationen zur Aufzeichnung von Cherenkov-Licht, mit abbildenden Luft-Cherenkov-Teleskopen (IACT). Ziel ist es, neue Rekonstruktionstechniken einzuführen, die die Winkel- und Kernortauflösung von HiSCORE mit dem Gammateilchen-Hintergrund Separationsvermögen der abbildenden Teleskope kombinieren.

In dieser Arbeit wird eine Simulations- und Analyseprozedur des ersten abbildenden Teleskops von TAIGA vorgestellt. Dieser Prozess beginnt bei der Verarbeitung der Rohdaten und endet mit der spektralen Rekonstruktion der Daten des Krebsnebels, die ausschließlich vom ersten abbildenden Teleskop stammen und zwischen Oktober 2019 und Dezember 2020 aufgenommen wurden.

Es wird eine gute Übereinstimmung abstrakterer Bildparameter mit der Simulation gezeigt. Unter Verwendung dieser Bildparameter werden Richtungsrekonstruktion, Energieabschätzung und Hintergrundunterdrückung mithilfe von Random Forests umgesetzt, einer weit verbreiteten Technik des maschinellen Lernens. Dabei wird eine Winkelauflösung von $<0.2°$ zwischen 10 TeV und 100 TeV erreicht. Die Energieauflösung und -genauigkeit bleiben $<25\,\%$ bzw. $<5\,\%$ zwischen 10 TeV und 100 TeV. Bei der Hintergrundunterdrückung mit dem Random Forest wird über alle Energien ein Qualitätsfaktor von 4 erreicht, der oberhalb von 30 TeV auf 6 ansteigt.

Zusammen mit den Instrumenten-Antwortfunktionen der Random Forest Analyse wurden kalibrierte, gesäuberte, rekonstruierte und hintergrundunterdrückte Ereignisse für eine Auswertung mit Gammapy genutzt.

122 h der Beobachtung des Krebsnebels im Wobbel-Modus unter guten Wetterbedingungen ergaben 204.6 Überschussereignisse mit einer Signifikanz von 9.7σ bei einer Energieschwelle von etwa 6 TeV. Die Anpassung eines Potenzgesetzes mittels Vorwärts-Faltung ergab einen Fluss von $\phi_0 = (4.3 \pm 0.5) \times 10^{-10}\,\mathrm{m}^{-2}\,\mathrm{s}^{-1}\,\mathrm{TeV}^{-1}$ bei einer Referenzenergie von 11 TeV und einen Exponenten von $\Gamma = 2.97 \pm 0.15$.

Eine grobe Abschätzung der systematischen Fehler ergab eine Ungenauigkeit der Energiekalibration von $\sim 20\,\%$.

Contents

Abstract **iii**

Kurzfassung **v**

Contents **vi**

1 Introduction **1**
- 1.1 Motivation 4
- 1.2 The Crab Nebula 5
- 1.3 Extensive Air Showers 6
  - 1.3.1 Cherenkov Light 9
  - 1.3.2 Imaging Air Cherenkov Telescopes 10
- 1.4 The Hybrid Approach 13

2 TAIGA **17**
- 2.1 Tunka Overview 17
  - 2.1.1 Coordinate Systems 19
- 2.2 HiSCORE 20
  - 2.2.1 Raw data processing 21
  - 2.2.2 Calibration 22
- 2.3 The First Imaging Air Cherenkov Telescope 23
  - 2.3.1 Camera 24
  - 2.3.2 Photomultiplier Tubes 26
  - 2.3.3 Camera Electronics 26
  - 2.3.4 Trigger and Readout Logic 29
  - 2.3.5 Calibration 31
  - 2.3.6 Pointing accuracy 32
  - 2.3.7 Optical Point Spread Function 34
  - 2.3.8 Raw data processing 36
  - 2.3.9 Pedestal estimation 37
  - 2.3.10 Cleaning 41
  - 2.3.11 Image Parameters 42
  - 2.3.12 Trigger Threshold 44
- 2.4 The Second Imaging Air Cherenkov Telescope 45
- 2.5 Hybrid Event Finding 50

3 Simulation **51**

3.1 Monte Carlo Dataset . . . 51
  3.1.1 Simulated Array . . . 53
  3.1.2 Atmosphere . . . 54
  3.1.3 Sample Weighting . . . 56
3.2 Simulation of the First Imaging Air Cherenkov Telescope . . . 58
  3.2.1 Mirrors . . . 59
  3.2.2 Camera and PMTs . . . 62
    3.2.2.1 Winston Cones . . . 63
    3.2.2.2 Quantum Efficiency and Single Photoelectron Response . 63
  3.2.3 Signal . . . 64
  3.2.4 Pedestal and Night Sky Background . . . 66
  3.2.5 Trigger and Readout . . . 67
  3.2.6 Comparison with Data . . . 70
  3.2.7 Expected Background Rate . . . 73
3.3 Simulation of HiSCORE . . . 75

4 Reconstruction **77**
4.1 IACT Random Forest Analysis . . . 77
  4.1.1 Random Forest . . . 78
  4.1.2 Training and Test Sets . . . 81
  4.1.3 Origin Reconstruction . . . 82
  4.1.4 Energy Reconstruction . . . 87
  4.1.5 Gamma-Background Separation . . . 90
  4.1.6 Effective Area . . . 94
  4.1.7 Energy Threshold . . . 96
4.2 HiSCORE . . . 99
  4.2.1 Core Reconstruction . . . 99
    4.2.1.1 Center of Gravity . . . 99
    4.2.1.2 Amplitude Distribution Function Fit . . . 100
  4.2.2 Origin Reconstruction . . . 100
    4.2.2.1 Planar Light Front Fit . . . 101
    4.2.2.2 Curved Light Front Fit . . . 101
    4.2.2.3 Arrival Time Model Fit . . . 101

5 Single IACT Results **103**
5.1 Available Data . . . 103
5.2 Good-Weather-Selection . . . 104
5.3 Crab Analysis . . . 107
  5.3.1 Instrument Response Functions . . . 107
  5.3.2 Source and Background Regions . . . 108
  5.3.3 Significance Determination . . . 111
  5.3.4 Spectral Fit . . . 112
  5.3.5 Flux Point Determination . . . 114
5.4 Summary of Crab Spectrum . . . 115
5.5 Systematic Uncertainties . . . 117
  5.5.1 Consistency in Subsets . . . 117
  5.5.2 Optical Transmission . . . 118

5.5.3 Other Uncertainties 120

6 First Hybrid Steps **123**
6.1 Stereo and Hybrid Core Reconstruction 123
6.2 IACT as additional Station 127

7 Summary and Outlook **131**

A Software and Data Availability **133**

B Processing Flow **135**

List of Figures **137**

List of Tables **147**

Abbreviations **149**

References **151**

Acknowledgements **164**

Chapter 1

Introduction

Humans have been observing the sky for thousands of years and for most of that time, the observations were constrained to purely photons at optical wavelengths visible by the human eye. Then, around 1912 another unknown source of radiation was detected with the balloon experiments of Victor Hess[1], the cosmic rays. In the last century, with advancements in technology and understanding, most of the electromagnetic spectrum outside of the optical range has been covered. Radio waves from the Milky Way were first observed in 1933 with a directional radio antenna[2]. At the other end of the spectrum, in 1962, the first emission from an object outside the solar system at X-ray energies was detected with an Aerobee rocket[3]. With the invention of spacecraft, the higher energies were further explored and the first detection of cosmic gamma rays took place in the late 1970s. The Vela spacecrafts[4], a satellite system with the purpose of monitoring the compliance with the 1963 Treaty Banning Nuclear Weapon Tests in the Atmosphere, in Outer Space and Under Water, detected transient events with increased gamma ray flux, Gamma Ray Bursts (GRB). Because the flux of most sources quickly drops with increasing energy, the limited effective area a satellite can offer practically determines the upper detection threshold. The brightest known source with mostly steady emission is the Crab Nebula. It serves as a standard candle in X-ray and gamma ray astronomy[5], and its TeV emission has been observed by numerous instruments[e.g. 6–14]. A brief description of the Crab Nebula is given in Section 1.2. It has an approximate flux of $\sim 3 \times 10^{-13}\,\mathrm{s^{-1}\,cm^{-2}}$ at energies above 10 TeV, corresponding to 0.2 photons per year at an effective area of $2\,\mathrm{m^2}$.

Reaching effective areas several orders of magnitude larger (typically $>1 \times 10^4\,\mathrm{m^2}$), ground based observatories are sensitive to the small fluxes at the highest energies. These large areas are the result of indirect detection techniques, where these instruments detect extensive air showers (EAS) caused by energetic gamma and cosmic rays when entering the atmosphere. EAS are explained in more detail in Section 1.3.

The beginning of TeV astronomy was the detection of the Crab Nebula in 1989 by Whipple[6], an Imaging Air Cherenkov Telescope (IACT). It measured the flashes of Cherenkov light produced in EAS, a method elaborated in more detail in Section 1.3.2.

Ever since, the field continuously expanded and photons of increasingly higher energies are being detected, so that nowadays hundreds of sources at TeV energies are identified.[1] Recently, LHAASO observed several photons above 400 TeV[15], with one photon reaching as high as 1.4 PeV. Interestingly, most of the photons could not be firmly localized and identified with a known source. One exception is the Crab Nebula, where LHAASO measured the spectrum to ∼1 PeV[13].

To find the sources capable of producing such highly energetic particles and to understand the underlying acceleration mechanisms, continued observations at the highest energies above 100 TeV are necessary and promise unprecedented insight.

Observations at these energies require very good sensitivities and large effective areas. In order to observe 1 photon above 100 TeV in 100 h from the Crab Nebula with an approximate flux of $\sim 3 \times 10^{-15}\,\mathrm{s}^{-1}\,\mathrm{cm}^{-2}$, an effective area of approximately $1 \times 10^{5}\,\mathrm{m}^{2}$ is required. The approach being pursued by LHAASO KM2A is to build a large array of water Cherenkov detectors and scintillators for muon detection spread over an area of $\sim 1.3\,\mathrm{km}^{2}$.[16] Another existing approach is to use the largest IACTs (e.g., CTA Large Size Telescope (LST) or HESS CT 5 as indicated by [17]) at zenith angles of >70°. This was already demonstrated for observations of the Crab Nebula with the MAGIC telescopes[12], with the finding that the effective area increases from the typical $\sim 0.1\,\mathrm{km}^{2}$ to $\sim 1\,\mathrm{km}^{2}$ at 100 TeV at the cost of worsening energy and angular resolution by a factor of about 1.5 to 2 with the expected potential to be reduced by a dedicated analysis.

A straight forward way of increasing the effective area is to scale IACT arrays (HESS, MAGIC, VERITAS, etc.) to larger areas, as planned for CTA. On the southern site, CTA plans to distribute 37 Small-Sized Telescopes (SST) over an area of $\sim 3\,\mathrm{km}^{2}$ (Alpha Configuration)[18, 19]. The close spacing is required so that for every shower at least two telescopes are triggered, allowing the stereoscopic reconstruction of the origin and other parameters of the primary particle causing the EAS.

However, to reach effective areas $>10\,\mathrm{km}^{2}$, this approach requires a large number of telescopes, so another idea is to combine images created by individual IACTs with a Cherenkov light sampling array. Instead of using traditional stereoscopy, the image of individual IACTs is complemented with the time and amplitude information recorded at the stations, potentially allowing a separation power comparable to IACT arrays.

---

[1] `http://tevcat.uchicago.edu/`

To investigate this hybrid approach, the TAIGA project was started. The existing pilot array already features more than 100 stations and 2 IACTs. This work is one step in a series of efforts to prove that the hybrid method is a viable alternative for future detectors observing at energies >100 TeV. The main part of this work was to implement the analysis chain of the first TAIGA IACT in a standalone mode, from raw data to the differential energy spectrum of the Crab Nebula. Chapter 2 introduces the current TAIGA array and Chapter 3 describes the simulation of the IACTs and demonstrates a good match of higher level parameters between data and simulation. For the first TAIGA IACT, a Random Forest analysis is presented in Chapter 4 and used to reconstruct the differential spectrum of the Crab Nebula in Chapter 5. A first look at hybrid events is taken in Chapter 6 by comparing the shower core reconstruction between the simulation and data and by using the telescope as an additional station in the array reconstruction.

1.1 Motivation

As mentioned above, the acceleration mechanisms of gamma sources are not yet fully understood and it is likely that many sources are still undetected. In addition to that, there are several unresolved questions as identified by Tluczykont et al. [20], which may be answered with observations of EAS above several TeV. Two particular cases are:

Cosmic rays up to energies of ∼3 PeV are thought to be of galactic origin as charged particles below this energy could be confined by the galactic magnetic field. The sources capable of accelerating cosmic rays to those energies, called Pevatrons, are not yet identified. Supernova remnants (SNR) are thought to be suitable candidates possibly accelerating charged particles via Fermi acceleration[21]. Associating gamma rays at multi-TeV energies either via direct emission or via emission from molecular clouds when accelerated particles hit the cloud could help to identify the Pevatrons. [22, 23]

Furthermore, in this energy range, the gamma rays are attenuated by pair production in photon fields at microwave wavelengths. In this band, the well-known cosmic microwave background (CMB) and the galactic interstellar radiation fields (ISRF) are the main contributions to the opacity. Depending on the line of sight through the galaxy and on the energy of the observed gammas, either the absorption from the ISRF or the absorption from the CMB dominates. If the intrinsic spectrum of a galactic source is known, the ISRF density or the distance to the source could be estimated. [20, 24, 25]

Other possibilities mentioned in Tluczykont et al. [20] are:

- Alteration of the expected opacity by photon-axion oscillations or modified pair production thresholds by Lorentz invariance violation.
- Gammas could be produced in intergalactic pair cascades initiated by ultra-high-energy cosmic rays from the local supercluster when interacting with the CMB. This emission would match the large scale of the supercluster and could be observed by measuring anisotropy.
- Measurement of the proton-air and consequently the proton-proton cross section by reconstruction of the depth of the shower maximum.

For a more complete list of unanswered questions, see e.g. Tluczykont et al. [20] or Funk [23].

Additionally, multi-TeV observations could also help to shine more light on the brightest steady gamma source, the Crab Nebula, also analyzed in this work. Long term observations at the highest energies of the Crab Nebula could be used to constrain a possible long term variation of the spectrum ≫1 TeV as indicated by Dirson and Horns [26] based on the comparison of spectra taken over several years and from different instruments. Additionally, the radial extent of the magnetic field could be constrained by the shape of the spectrum at the highest energies as pointed out in Dirson and Horns [26].

1.2 The Crab Nebula

The analysis presented in this work is tested on data from the Crab Nebula, the brightest steady gamma source. It is a remnant of a supernova explosion visible from earth in the year 1054 AD and is used as a standard candle in gamma astronomy. This brief description of the Crab Nebula is based on Bühler and Blandford [27] and a recent model fit presented in Dirson and Horns [26].

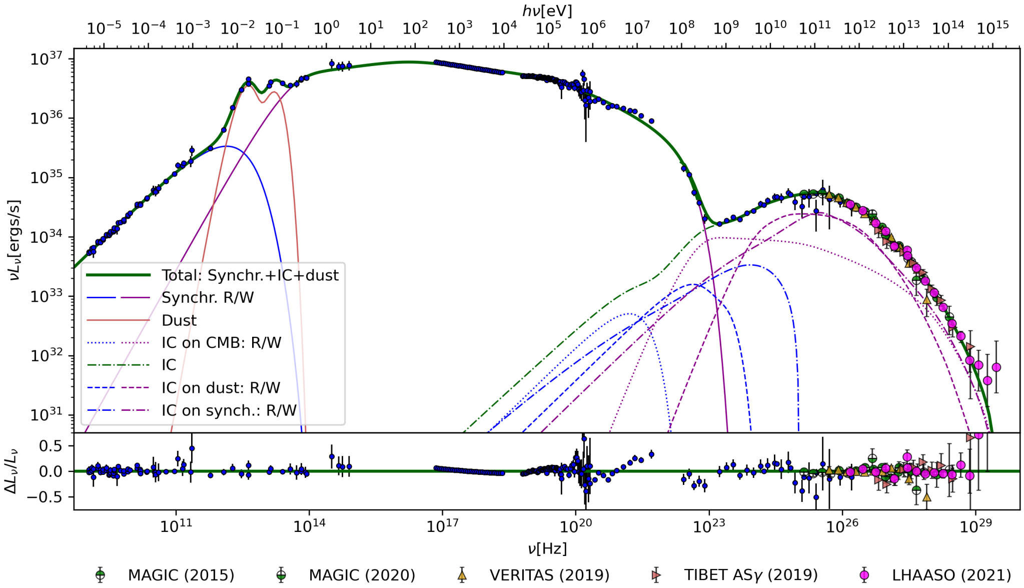


Figure 1.1: Spectral energy distribution of the Crab Nebula from radio to gamma ray energies. It was taken from Dirson and Horns [26] and shows measurements as well as a model fit. Additionally, the synchrotron and inverse Compton (IC) scattering components resulting from two assumed electron populations, the radio (R) and the wind (W) electrons, are shown. See [26] for details.

In typical models, it has a magnetized and spinning neutron star at its center. In the magnetic field of the pulsar, electrons and positrons are created and accelerated in an outward direction up to relativistic speeds, forming the pulsar wind. This wind terminates in a standing shock when it hits the surrounding remnant. Downstream of the shock, the wind slows down. [27]

Relativistic electrons in the shock can explain the nonthermal radio emission with synchrotron radiation. The nonthermal gamma rays are produced when the relativistic electrons undergo inverse Compton (IC) scattering with the photons present in the nebula. These photons can be photons from synchrotron radiation, photons of the cosmic microwave background (CMB), or possibly infrared photons from dust emission. [26]

Figure 1.1 shows the spectral energy distribution from radio to gamma ray energies. It exhibits two broad peaks, the synchrotron and the IC peak starting at $\sim$1 GeV. Because of its brightness at the highest energies, the Crab Nebula is ideally suited to validate and test an analysis chain. The result of a Random Forest analysis of Crab Nebula data taken with the first TAIGA IACT is shown in Chapter 5. In this analysis, an effective area of $\sim 2 \times 10^5\,\mathrm{m}^2$ at $\sim$100 TeV was reached. Single IACTs can reach such large areas by detecting the Cherenkov light front produced in extensive air showers (EAS) as described in Section 1.3.

1.3 Extensive Air Showers

When detecting gamma rays via extensive air showers (EAS), the largest background are EAS produced by charged primary particles. There is a constant flux of cosmic rays, mainly protons ($\sim$90 %) and heavier nuclei impinging on the atmosphere. Figure 1.2 shows the isotropic all particle flux which approximately follows a power law $\propto E^{-\alpha}$ with $\alpha \approx 2.7$ up to the so-called knee at $\sim 3\,\mathrm{PeV}$, from where it follows an exponent of $\alpha \approx 3.1$ before changing to $\alpha \approx 2.6$ at $\sim 3\,\mathrm{EeV}$, the ankle. [28]

The sources of the cosmic rays are still not completely identified even though the knee is believed to be connected to the end of the spectrum of galactic accelerators and the cosmic rays above the ankle are believed to be of extragalactic origin.[28]

When a cosmic or gamma ray with sufficient energy reaches the atmosphere, extensive air showers (EAS) develop. This brief description is based on de Naurois and Mazin [30] and Hofmann and Hinton [31]. For a gamma primary, the process starting the electromagnetic shower is the electron-positron pair production in the vicinity of a field of an atmospheric nucleus or electron. The created electrons and positrons will mainly undergo bremsstrahlung[32] and emit photons when deflected by the nuclei. These secondary photons will again pair produce and this cascade goes on until Compton scattering becomes more dominant for photons and ionization losses dominate over bremsstrahlung for electrons and positrons. At the critical energy of $E_c = 83\,\mathrm{MeV}$, the losses from bremsstrahlung are equal to the ionization losses. In air, the radiation length for bremsstrahlung is $X_0 = 36.7\,\mathrm{g\,cm^{-2}}$, which is $\frac{7}{9}$ of the mean free path for pair production. The value is given in units of slant depth X, which is a measure of the traversed matter with density ρ along the shower track $X = \int \rho(l)dl$. The atmosphere

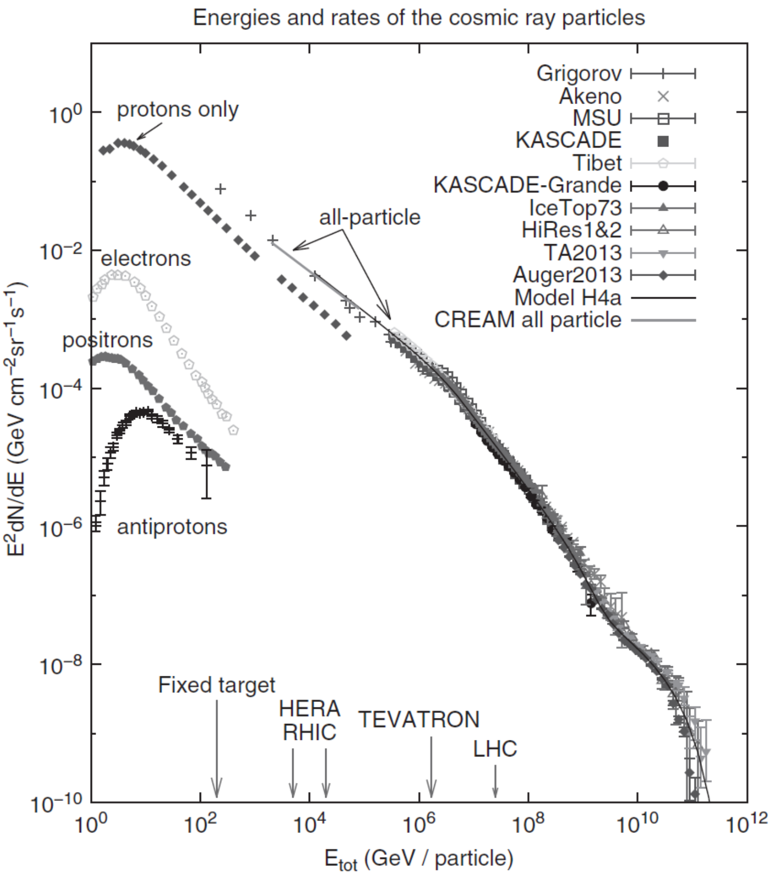


Figure 1.2: Cosmic ray flux. Taken from [28].

at 0° zenith distance has a total of $X \approx 1000\,\mathrm{g\,cm^{-2}}$, corresponding to approximately 27 radiation lengths. Due to the strong dependence of ρ on altitude, one radiation length is $\sim$300 m at sea level and $\sim$1 km at an altitude of 10 km. Figure 1.3a shows an example of a simulated shower for a 100 TeV gamma. A simple model is the Heitler model. It considers only pair production and bremsstrahlung and assumes that both have the same radiation length, neglecting $\frac{7}{9}$. In this model, every particle with energy E_0 produces two new particles, with each of the new particles having an energy of $E_0/2$. It gives a good approximation of the position with the maximum number of particles in the shower, the shower maximum X_{max} as

$$X_{max} = X_0 \ln \frac{E_0}{E_c},$$

which is $\sim 430\,\mathrm{g\,cm^{-2}}$ for 10 TeV and $\sim 600\,\mathrm{g\,cm^{-2}}$ for 1 PeV.

For other primary particles like protons and heavier nuclei, hadronic interactions involving strong force such as elastic scattering, inelastic scattering and nuclear disintegration occur. In these processes, additional secondary particles, mainly pions, are produced.

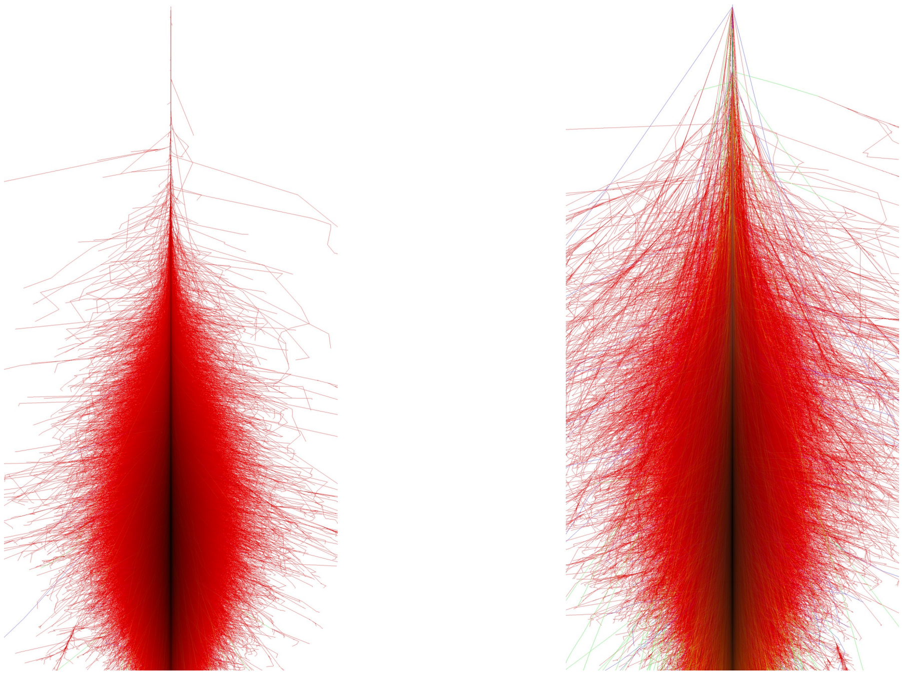

(a) 100 TeV Gamma (b) 100 TeV Proton

Figure 1.3: Examples of simulated extensive air showers. Electrons, positrons and gammas are shown in red. Muons are shown in green. The vertical axis shows the altitude from 0 km to 30.1 km, while the horizontal axis spans ±5 km. Taken from [29].

Neutral pions π^0 almost immediately convert to two photons

$$\pi^0 \to \gamma + \gamma,$$

which may in turn start secondary electromagnetic showers, while positive and negative pions $\pi^\pm$ survive long enough to interact again before decaying. π^+ decay into muons μ^+ and neutrinos ν, while negative pions π^- decay into μ^- and antineutrinos $\overline{\nu}$:

$$\pi^+ \to \mu^+ + \nu, \text{ and}$$

$$\pi^- \to \mu^- + \overline{\nu}.$$

Because of these interactions, hadronic showers show larger fluctuations, and with the larger transverse momentum transfer, they appear wider as the showers caused by primary gammas with comparable energy. An example shower initiated by a 100 TeV proton is shown in Figure 1.3b. Hadronic showers have $\sim \frac{1}{3}$ of the number of charged particles of electromagnetic showers with the same energy.

The shower development is a stochastic process and every shower is individual, limiting the theoretically achievable reconstruction accuracy. A major contribution to the fluctuation is the depth of the first interaction from where the cascade starts. For gamma showers, the depth of the first interaction fluctuates by a root mean square of $\sim 1X_0 \approx 36.7\,\mathrm{g\,cm^{-2}}$ and for protons it is one interaction length of $\sim$90 g cm^{-2}.

1.3.1 Cherenkov Light

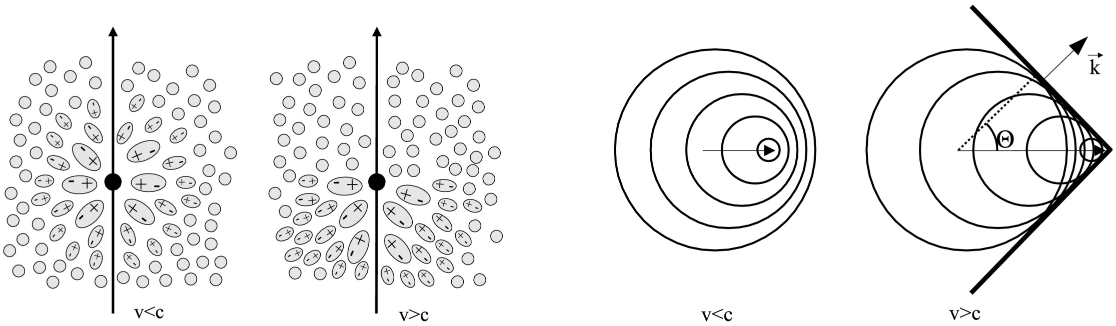


Figure 1.4: On the left, the polarisation of a medium is shown for the cases where the speed V of the charged particle is larger and smaller than the speed of light c in the medium. On the right, the constructive interference producing the Cherenkov wavefront with an angle Θ to the trajectory of the particle is shown. Taken from [30].

During their path through the dielectric medium, the charged particles of the air shower polarize the molecules along their tracks. If the speed v of the charged particles is greater than the speed of light $c = \frac{c_0}{n}$ in the air with the refractive index n and the speed of light in vacuum c_0, the depolarization causes constructively interfering Cherenkov wavefronts to be produced as illustrated in Figure 1.4. At a given point along the trajectory, the opening angle Θ of the Cherenkov light cone is described by $\cos\Theta = \frac{c_0}{vn}$. This opening angle varies from $\sim$1.5° at sea level to 0.8° at an altitude of 10 km. The resulting spectrum of Chrenkov light follows $\propto \frac{1}{\lambda^2}$, where λ is the wavelength of the light. Considering the atmospheric opacity and the spectrum of the light of the night sky background (NSB), the useful range is $\sim$300 nm to $\sim$600 nm.

Figure 1.5 shows how the overlay of the main emission directions of particles close to the shower axis for a shower at vertical incidence causes a characteristic distribution on the ground. This distribution has a maximum intensity at a circular annulus with a radius of $\sim$150 m around the core position. On the upper right, the photon intensity on the ground per radial distance to the shower core is shown. On the lower right, the distribution of the time delay for light emitted at different altitudes is shown. Right at the core location, the light emitted last is arriving first and the light emitted earlier in the shower arrives later, because the shower front travels faster than the speed of light. At larger radial distances, the effect reverses, as the light emitted last has to travel larger distances, compensating the speed difference between the speed of the shower front and the speed of the light in air.

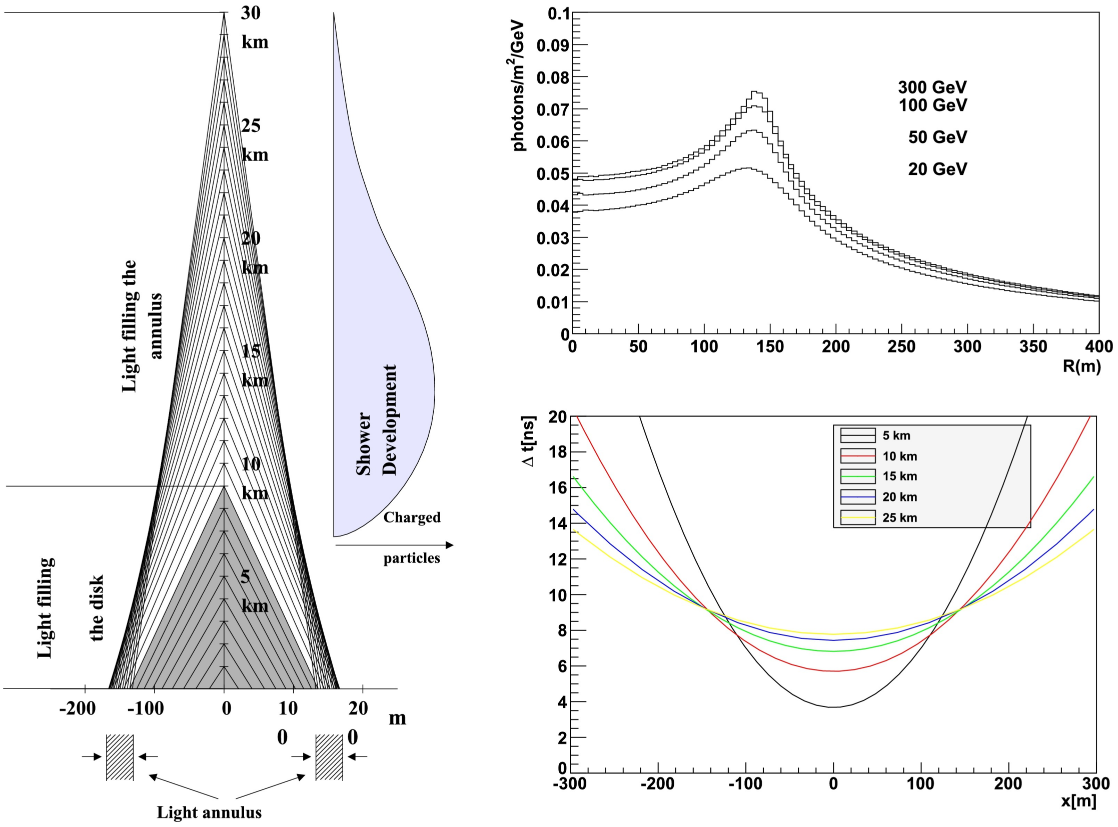


Figure 1.5: On the left, the overlap of the Cherenkov light on the ground is shown (the positive labels of the horizontal axis miss a 0). Here, only the change of the Cherenkov opening angle with altitude of particles close to the shower axis is considered. On the upper right, the density of the Cherenkov light on the ground is plotted against the lateral distance, while on the lower right, the arrival time of the Cherenkov light on the ground for different emission heights is shown as a function of the lateral distance. Taken from [30].

There are two methods to observe the Cherenkov light on the ground. Detection of the arrival time and intensity of the Cherenkov light front on several positions on the ground, e.g. via arrays of optical stations, and IACTs, taking images of the Cherenkov light.

The arrays typically record the lateral amplitude distributions and the arrival times of the showers shown on the right in Figure 1.5. From that information, e.g., the core position and arrival direction of the primary particle is reconstructed.

1.3.2 Imaging Air Cherenkov Telescopes

IACTs record the Cherenkov light in several pixels, taking an image of the shower as it develops in the atmosphere. Figure 1.6 shows an example image. Typically, the image is analysed with a moment analysis proposed in Hillas [33]. It describes the amplitude distribution as an ellipse and defines, among other parameters, the second moments along the semi-minor and semi-major axes as image width and length.

When interested in gamma rays, the images of all EAS caused by other primary particles are considered background. Typically, the gamma fraction of all recorded events is $\sim 1 \times 10^{-4}$ and an effective method of background suppression is required to improve the sensitivity. The parameters of the image can be used to define effective cuts, reducing the number of background events in an analysis.

A convenient simplification to describe the significance of an instrument recording N_{sig} signal events from a point source inside a circular region with radius $\boldsymbol{\theta}$ with a large enough isotropic background of N_{BG} events is

$$S = \frac{N_{sig}}{\sqrt{N_{BG}}} = \frac{t\ \epsilon_{sig}\ \epsilon_{\boldsymbol{\theta}} R_{sig}}{\sqrt{t \epsilon_{BG} \Omega R_{BG}}}, \tag{1.1}$$

where R_{sig} is the expected signal rate and R_{BG} is the background rate per solid angle. Both R refer to the rates before background suppression and ϵ is the fraction of events surviving the background suppression. $\Omega = 2\pi(1-\cos\boldsymbol{\theta}) \approx \pi\boldsymbol{\theta}^2$ is the solid angle defined by the radius $\boldsymbol{\theta}$ of the circular region and $\epsilon_{\boldsymbol{\theta}}$ is the fraction of signal events reconstructed to that region as expected from the point spread function of the instrument. S can then be written as

$$S = \sqrt{t} \frac{R_{sig}}{\sqrt{R_{BG}}} \frac{\epsilon_{sig}}{\epsilon_{BG}} \frac{\epsilon_{\boldsymbol{\theta}}}{\boldsymbol{\theta}} = Q \frac{\epsilon_{\boldsymbol{\theta}}}{\boldsymbol{\theta}} \sqrt{t} \cdot \text{const.}, \tag{1.2}$$

where $Q = \frac{\epsilon_{sig}}{\sqrt{\epsilon_{BG}}}$ is the quality factor indicating the separation power of the analysis. R_{sig} can be approximated as $R_{sig} = \int \phi_{sig} A_{sig} dE$, where ϕ_{sig} is the expected source flux and A is the effective area (before background suppression and $\boldsymbol{\theta}$-cut). To get the sensitivity, one chooses a significance level and an observation time and solves for ϕ_{sig}. Even though this is a simplified approach and normally the formula derived in Li and Ma [34] is used to estimate the sensitivity, the scaling of the significance / sensitivity with Q is still similar.

If an image is recorded by more than one telescope, stereoscopic analysis is possible. Using the geometric information available from several images as shown in Figure 1.6, the origin and the core position can be reconstructed. The origin is reconstructed by intersecting the lines defined by the major axes of the ellipses. To reconstruct the core position, the planes defined by the telescope location and by the shower axes are intersected on the ground level. It is evident, that the angle between the axes, also called stereo angle, is important for the reconstruction quality. This purely geometric reconstruction improves the angular resolution up to ~0.06° in current IACT arrays such as MAGIC[35]. The background suppression was also improved by scaling the image parameters with the expectation value determined from the mean value expected at the reconstructed core position. Two examples are the Scaled Width (SCW) and the Scaled Length (SCL). Typically, the mean of the SCW or SCL of all triggering telescopes is used, further improving the suppression power.

However, in order to use the stereoscopic reconstruction, at least two telescopes need to trigger and because the geometric reconstruction generally degrades when the stereo angle is small, the usable core locations relative to the telescopes are limited. This requires a close spacing of the IACTs, drastically increasing the cost to scale it to $>10\,\mathrm{km}^2$.

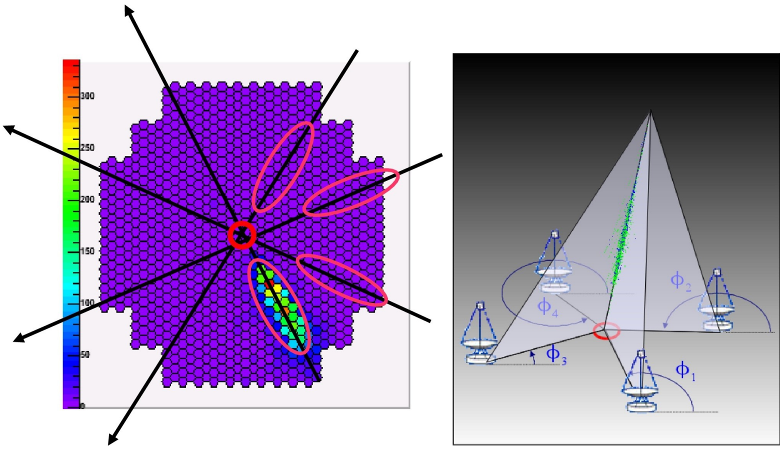


Figure 1.6: On the left, the geometric intersection of main axes of the superimposed images taken by four telescopes to reconstruct the origin of the primary particle. On the right, the planes through the shower axis and through the telescope positions show how the impact point is reconstructed. Taken from [30].

1.4 The Hybrid Approach

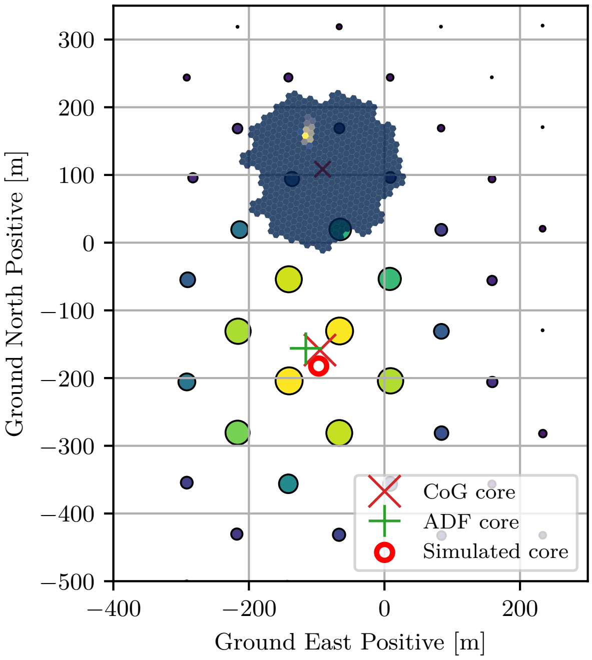


Figure 1.7: A simulated hybrid event of a 34 TeV gamma. The image recorded by the IACT is overlain at the telescope position and the color scale and size of the circles indicate the maximum amplitude recorded at the optical station. The circle marks the core position of the shower and the green and red crosses mark the core positions reconstructed from the center of gravity (CoG) and amplitude distribution function (ADF) fit methods of the array reconstruction.

Scaling a Cherenkov light timing array to larger areas is cheaper and easier than scaling conventional IACT arrays. However, the arrays typically have poor background rejection power. To overcome this, the hybrid approach combines the information of the images of single IACTs with the amplitude and timing information recorded by the array of optical stations. Figure 1.7 shows a simulated example event recorded with an array of optical stations with the image taken by an IACT overlain. This combination of instruments allows new analysis techniques (potentially newly developed ones) and can provide separation power comparable to stereoscopic observations. One possibility already explored in Kunnas [36] is the Hybrid Scaled Width $HSCW$, where the expected width at the core location reconstructed from the array is used to scale the width. Because the core position is determined from the array, it is independent of the core position relative to the telescope. As shown in Kunnas [36], the quality factor Q expected from $HSCW$ depends on the core distance and on the energy.

Figure 1.8 shows Q against the core distance for different energies. For energies between 20 TeV to 38 TeV, $Q \gtrsim 5$ is reached up to core distances of 250 m. For higher energies,

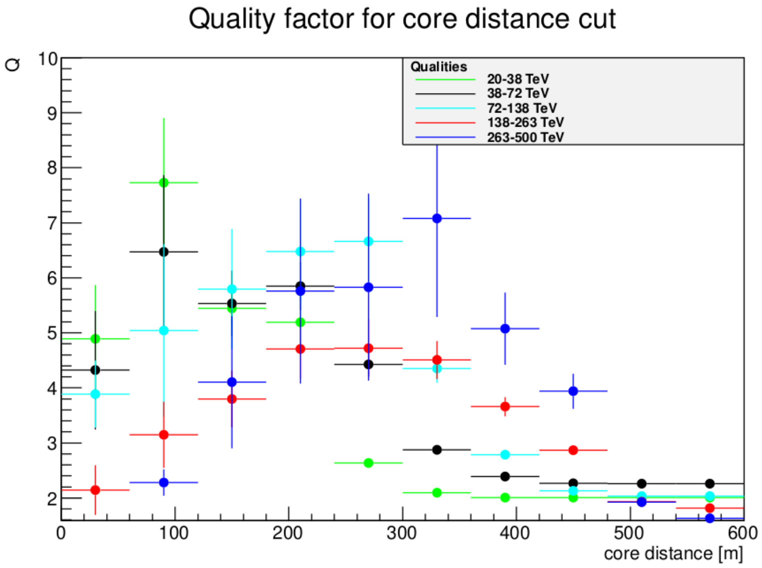


Figure 1.8: Achieved quality factor for several energies against the distance of the core to the telescope at a zenith distance of 0°. Taken from [36].

Q decreases close to the telescope but stays high even for core distances of $\sim$400 m, indicating the possibility of large distances between any two telescopes.

The same procedure as for $HSCW$ can also be applied to other image parameters like length, further increasing the separation potential. Using this approach, the telescope spacing can be >500 m, greatly reducing the number of required IACTs as compared to arrays consisting solely of IACTs.

To investigate this approach, the TAIGA project (Chapter 2) was started. It aims to combine the HiSCORE array with IACTs. After establishing the hybrid method with a pilot array of 1 km^2, the array is expected to be scaled to 5 km^2. Its expected integral sensitivity for 300 h is shown in Figure 1.9. However, the future is unclear, especially the German participation in future stages.

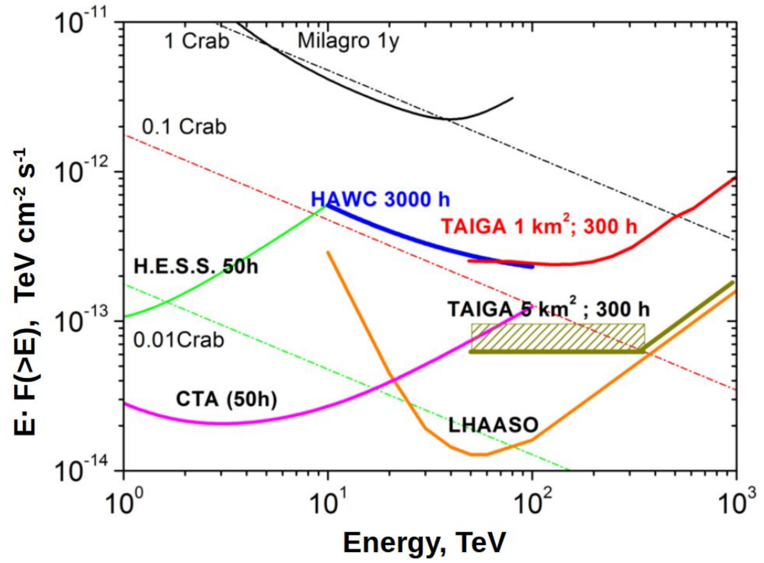


Figure 1.9: Integral flux sensitivity expected for TAIGA with 300 h point source observation. Taken from [37].

Chapter 2

TAIGA

The Tunka Advanced Instrument for Gamma- and cosmic-ray Astronomy (TAIGA) is an experiment for the observation of gamma and cosmic rays via the detection of extensive air showers (EAS). It is observing gamma rays above a few TeV and cosmic rays from ~100 TeV to about 1 EeV and consists of three main parts, the shower front sampling array Hundred *i* Square-km Cosmic ORigin Explorer (HiSCORE)[38–40], muon detectors (Tunka-Muon and TAIGA Muon) [41, 42] and IACTs [43]. Using the combination of the IACT image information with the HiSCORE reconstruction in the hybrid approach (see Section 1.4), a cost effective coverage of large areas is possible.

While already taking data, a pilot array currently covering ~1 km^2 is under continuous construction. It hosts several IACTS spaced at distances of more than 250 m. The first IACT is described in detail in Section 2.3 and Section 2.4 provides details of the second IACT and gives an overview of the planned IACTs. A brief description of HiSCORE is given in Section 2.2.

If not noted otherwise, the information presented in this Chapter is taken from the description of TAIGA written by Almeida and Tluczykont [44].

2.1 Tunka Overview

Figure 2.1 shows an image of a HiSCORE station with IACT 1 in the background. The snow visible in the image is an indication of the harsh observation conditions at the TAIGA site. In winter, clear days are common and the temperatures can drop down to −50 °C.[45]

TAIGA is located in the Tunka valley on the site of the Tunka-133 experiment[46] 50 km southwest of lake Baikal in Siberia (51°48′35″N, 103°4′2″E). The layout of the

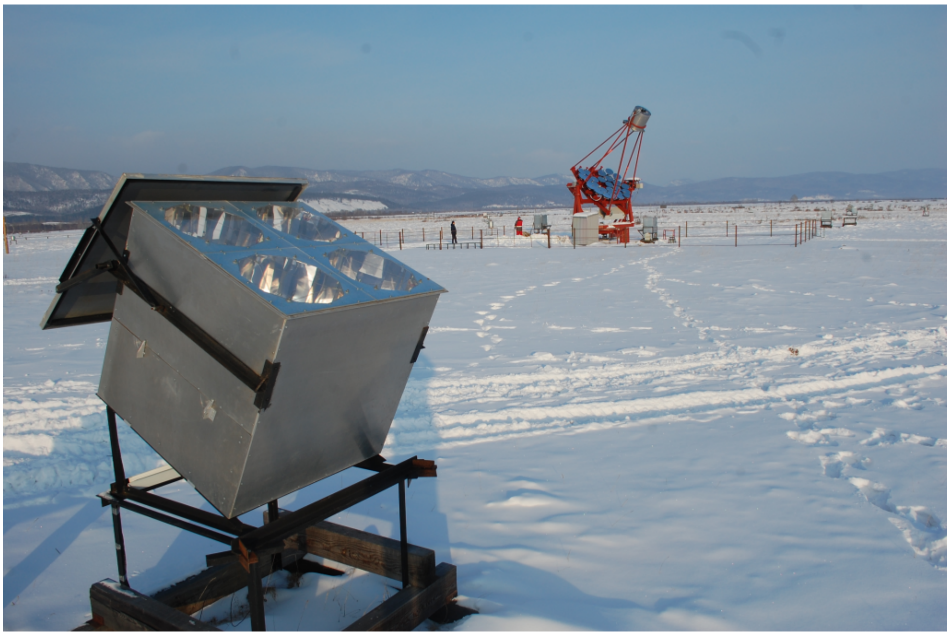


Figure 2.1: A HiSCORE station with its lid open in the foreground and IACT 1 in the background. The bucket-like instrument between the station and the IACT is a Tunka-133 detector. Photo courtesy TAIGA collaboration and available at `https://taiga-experiment.info/gallery`.

observation site with the IACTs and optical stations is shown in Figure 2.2. The current array consists of 121 optical stations, 3 IACTs and 3 TAIGA muon detectors.

The array spans an area of approximately 1.12 km^2 and is divided into 4 clusters. (Cluster 1: station ids below 100, cluster 2: station ids 1xx, cluster 3: 2xx, cluster 4: 3xx) Throughout the seasons analysed in this work, clusters 1 and 2 were always active. The first day of operation with cluster 3 was 2020-01-14 and the first day of observation with cluster 4 active was 2021-12-14.

All stations and IACTs have their own local trigger (no array trigger). The acquired and timestamped data is transferred to a central file server.

In order to achieve an angular resolution of $\sim 0.1°$ a timing accuracy of better than 1 ns is required. This is achieved with a synchronisation procedure similar to that used in the Tunka-133 array. It is based on a square 100 MHz reference signal created by the so-called MEGA-Host and is distributed to all clients via optical fibre. For some stations and the IACTs, the White Rabbit system[48] is additionally used for cross-checking. The general layout of the TAIGA Data Acquisition (DAQ) system[49] is shown in Figure 2.3. The DAQs of the IACTs and of each station are connected to cluster centers which are in turn connected to the central DAQ where the storage server is located.

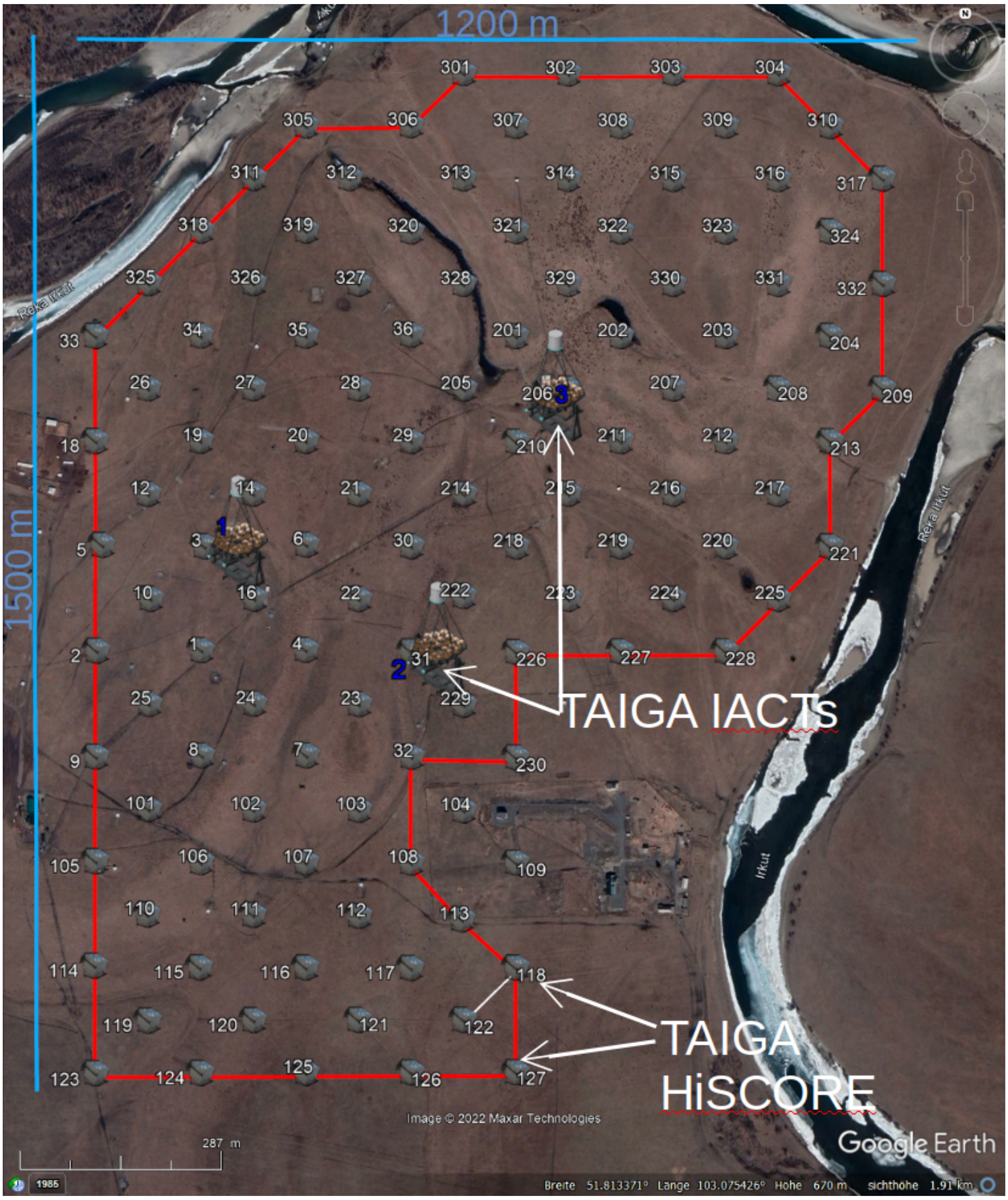

1200 m
1500 m
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
33
34
35
36
201
202
203
204
26
27
28
205
206
3
207
208
209
18
19
20
29
210
211
212
213
12
14
21
214
215
216
217
5
1
3
6
30
218
219
220
221
10
16
22
222
223
224
225
2
1
4
2
31
226
227
228
25
24
23
229
TAIGA IACTs
9
8
7
32
230
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
TAIGA HiSCORE
123
124
125
126
127
Reka Irkut
Irkut
Image © 2022 Maxar Technologies
287 m
Google Earth
1985
Breite 51.813371° Länge 103.075426° Höhe 670 m sichthöhe 1.91 km

Figure 2.2: Layout of the current TAIGA detector. The HiSCORE stations are labeled with their station ID in white color. The IACTs are labelled in blue. The blue lines mark the scales of the experiment while the red line indicates the perimeter of the active array components (Clusters 1, 2, 3 and 4, 1.12 km^2). Taken from [47].

2.1.1 Coordinate Systems

The array coordinate system has its origin at the first Tunka-133 station located at (51.813 194 44°N, 103.071 194 44°E) at an altitude of 675 m above sea level. It is a local euclidean system (no curvature) where X points towards north, Y towards the east and Z is pointing up. Over the several hundred meters spanned by the detectors, the

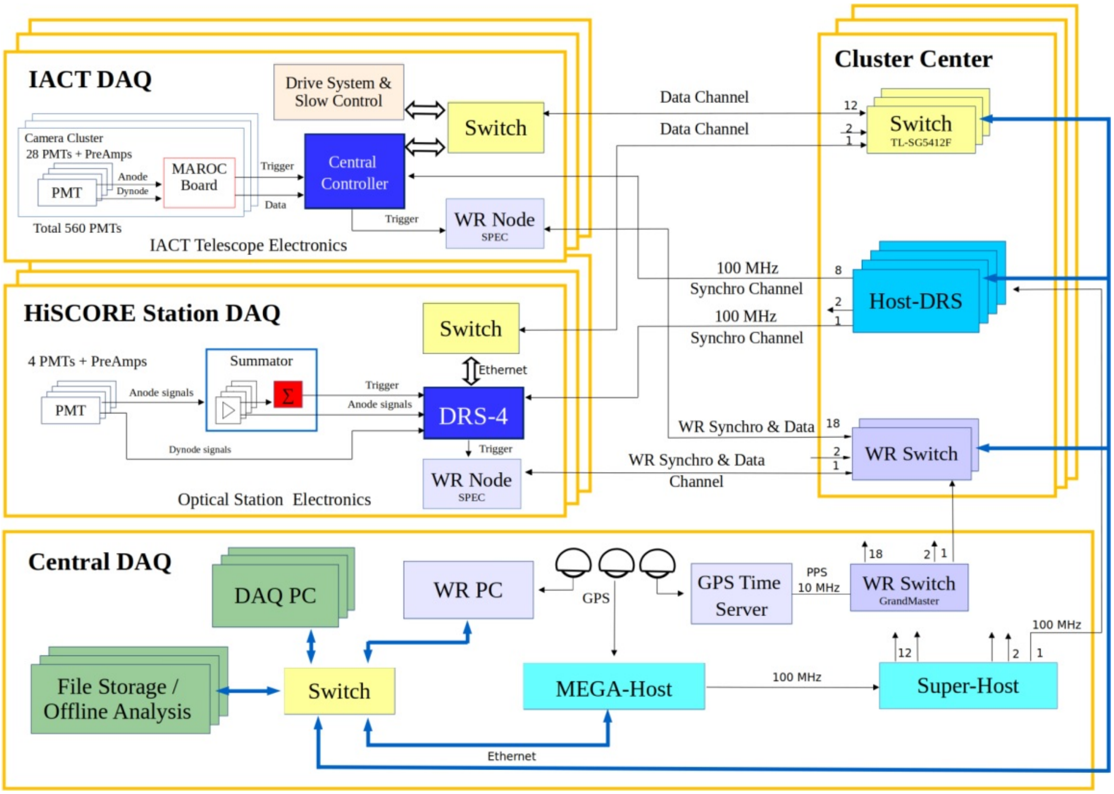

IACT DAQ
Drive System & Slow Control
Switch
Camera Cluster
28 PMTs + PreAmps
PMT
Anode
Dynode
MAROC Board
Trigger
Data
Central Controller
Trigger
WR Node SPEC
Total 560 PMTs
IACT Telescope Electronics
HiSCORE Station DAQ
Switch
4 PMTs + PreAmps
Summator
PMT
Anode signals
Dynode signals
Ethernet
Trigger
Anode signals
DRS-4
Trigger
WR Node SPEC
Optical Station Electronics
Cluster Center
Data Channel
Data Channel
Switch TL-SG5412F
100 MHz Synchro Channel
100 MHz Synchro Channel
Host-DRS
WR Synchro & Data
WR Synchro & Data Channel
WR Switch
Central DAQ
DAQ PC
WR PC
GPS
GPS Time Server
PPS 10 MHz
WR Switch GrandMaster
File Storage / Offline Analysis
Switch
MEGA-Host
100 MHz
Super-Host
100 MHz
Ethernet

Figure 2.3: Layout of the TAIGA Data Acquisition (DAQ) system. Taken from Budnev et al. [49]. All HiSCORE stations and IACTs are connected to so called Cluster Centers va optical fibers for the data transmission and for the 100 MHz time synchronisation channel. Some stations and the IACTs are additionally equipped with an optical fiber for the White Rabbit (WR) time synchronisation system. The several Cluster Centers are connected to the Central DAQ via Ethernet for the data transmission and via two optical fibers for the time synchronisation systems.

difference to a geographical (curved) system is already a few cm but practically still negligible. For the IACTs, the pointing is given in (altitude / zenith distance, azimuth) coordinates with (0° / 90°, 0°) when pointing towards the north. Turning towards the east is counted as a positive azimuth change and south lies at 180°. In the camera coordinate system, X points left (west) and Y points up (sky) when looking at the camera while standing on the ground in front of the mirrors with the telescope pointed at (0° / 90°, 0°). The zero point is in the center of the camera.

2.2 HiSCORE

HiSCORE is a Cherenkov light sampling detector array with currently 121 stations in operation. The stations are arranged on a regular grid (see Figure 2.2) with about 106 m between each of the stations spanning a total area of roughly 1 km^2.

As already elaborated in Section 2.1, the array is divided into 4 independent clusters (see also Figure 2.3).

Each station is tilted by 25° to the south and equipped with four PMTs and their respective Winston cones protected by a Plexiglas cover. The Winston cones reduce the collected NSB light by confining the radial acceptance to $\sim$0.6 sr and increase the light collection area to $0.5\,\mathrm{m}^2$. Two different PMTs are in use: Electron Tubes 9352KB and Hamamatsu R5912.

Figure 2.1 shows one station, and Figure 2.4 shows a sketch of the contained components.

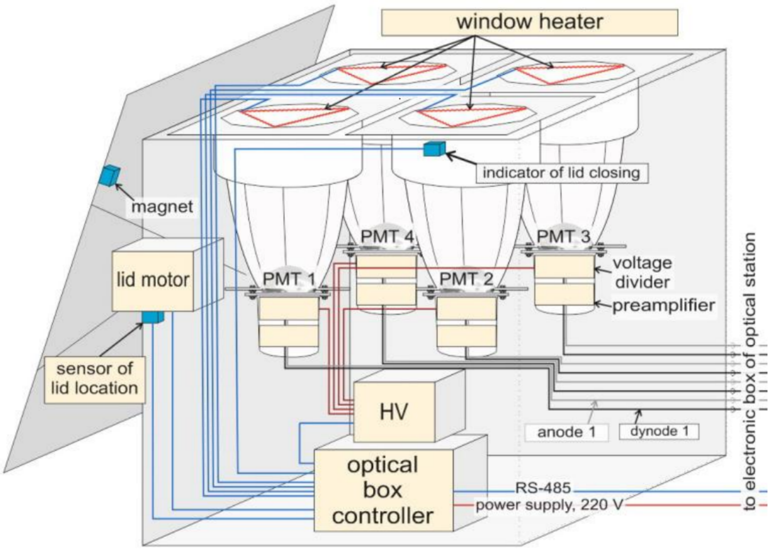

window heater
magnet
indicator of lid closing
PMT 4
PMT 3
PMT 1
PMT 2
voltage divider
preamplifier
lid motor
sensor of lid location
HV
anode 1
dynode 1
optical box controller
RS-485
power supply, 220 V
to electronic box of optical station

Figure 2.4: Sketch of the electronics of a HiSCORE station. Taken from Budnev et al. [43].

The trigger is independent for each station and is formed when the sum of the anode signals of all four PMTs reaches a preset threshold of $\sim$200 p.e.. For each event, the waveforms of the dynode and anode signal of each PMT, as well as the signal of the 100 MHz clock are read out. Each of the 9 channels of the DRS 4 provides 1024 slices sampled at 2 GHz. Figure 2.5 shows a recorded time slice.

2.2.1 Raw data processing

To increase the signal-to-noise ratio, the traces of all four PMTs are aligned and summed. One example signal is shown in Figure 2.5 together with its 100 MHz synchronisation signal in the lower panel. From the sum signal, several parameters are reconstructed: The maximum amplitude A, the integral charge Q, the full width at half maximum (FWHM) and the time delta between the time when the rising edge reaches 50 % of the maximum amplitude to the next rising phase of the clock signal. The parameters are provided in text files, the *tim*-files (see [37, 50]), storing the reconstructed and matched events (window of 2 µs parameters with one line per station. These text-based *tim*-files are converted to the widely used binary HDF5[1] format to save space and improve the read performance.

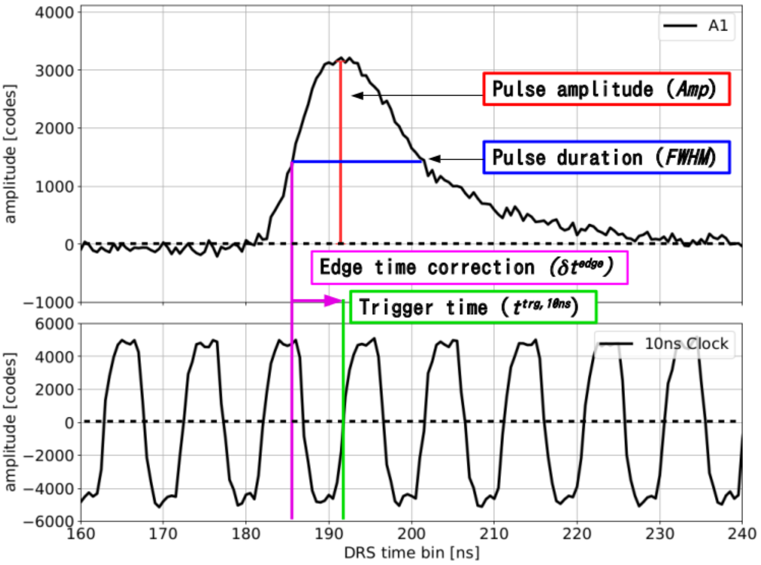


Figure 2.5: Basic parameters describing the waveform of the sum of the signals recorded by a HiSCORE station. The lower panel shows the digitized signal of the 100 MHz synchronisation. Taken from Porelli [37].

2.2.2 Calibration

The amplitude and time calibration is done with LED runs from a flasher located several hundred meters away from the stations. As the LED is located outside the field of view, a white reflective screen is mounted on top of the stations with a 45° angle. This time-consuming process is only done for a fraction of the stations. The calibration constants of the remaining stations are then determined indirectly via analysis of the expected time and amplitude based on events with a large number of stations triggering. For more details, see Porelli [37] and Epimakhov [50].

The accuracy of the timing calibration was also confirmed with the tracking of the passage of the CATS-LIDAR aboard the ISS.[37]

---

[1] `https://www.hdfgroup.org/solutions/hdf5/`

2.3 The First Imaging Air Cherenkov Telescope

The design of the TAIGA IACTs is derived from the design of the HEGRA telescopes[51]. Figure 2.6 shows an image of IACTs 1 and 2.

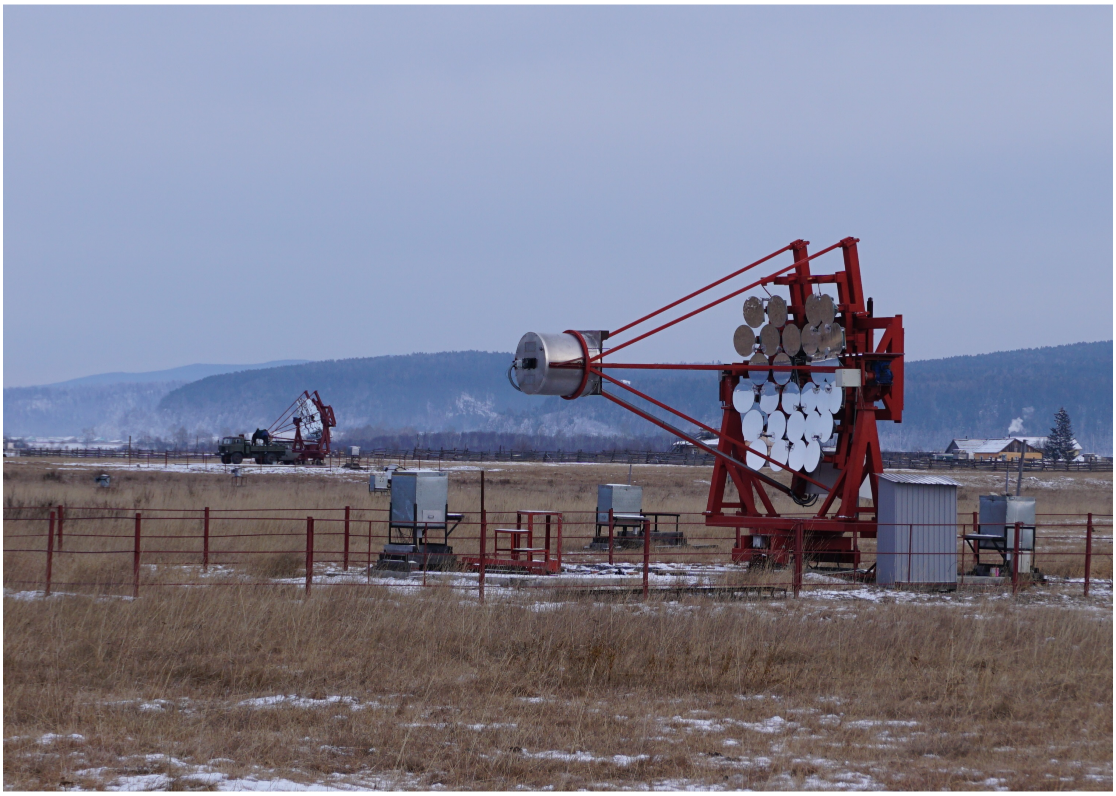


Figure 2.6: The first TAIGA imaging air cherenkov telescope with the second telescope in the background. Photo courtesy TAIGA collaboration.

Each of the telescopes has 34 mounting points for individual mirror segments spread over the diameter of 4.3 m. The segments have a diameter of 60 cm and a focal length of $f = (4.75 \pm 0.01)\,\mathrm{m}$ [36] and are arranged in a Davies-Cotton design.

For the first telescope, 29 round segments are used forming a reflective area of $8.2\,\mathrm{m}^2$. They were produced by Galactica in Armenia.On 2020-12-09, five hexagonal mirrors built by our colleagues[52] were added at the missing spots, increasing the reflective surface area to $\sim 9.7\,\mathrm{m}^2$.

At the central spot, no mirror is attached, and two LED flashers with diffusers and one CCD camera are installed for calibration purposes.

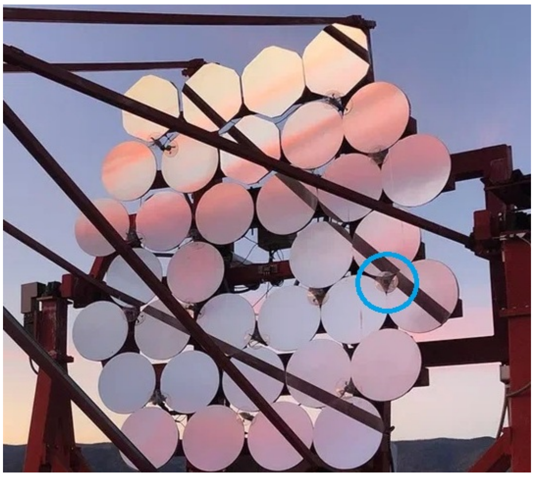
To prevent ice formation during cold observation conditions, hot air is blown onto small reflection shields located a few centimeters in front of the mirrors. In Figure 2.7a), one of the reflection shields is marked with a circle.

The telescope camera is mounted on 6 support arms and positioned at $(475 \pm 1)\,\mathrm{cm}$ distance from the mirrors such that the entrance of the Winston cones lies directly in the focal plane, adjusting the focus point of the telescope to infinity.

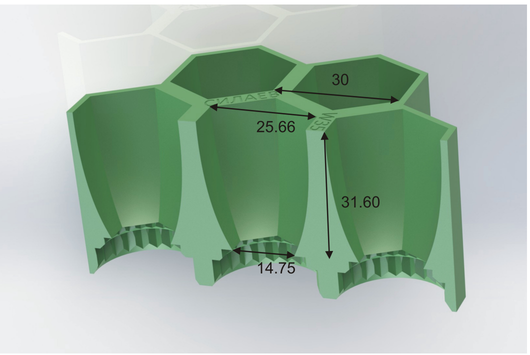
With the pixel width $w_{pix} = 3\,\mathrm{cm}$ and the nominal focal length of $f = 475.2\,\mathrm{cm}$ the conversion factor s from a distance in the focal plane to the angular distance seen by the telescope in the center of the camera is given by

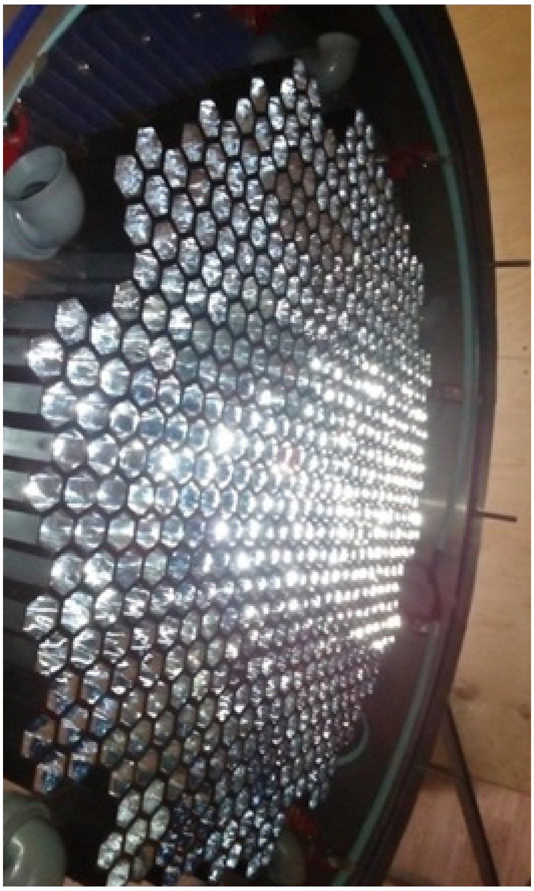
$$s = 2\frac{\arctan\left(0.5\frac{w_{pix}}{f}\right)}{w_{pix}} \approx 0.121\,^{\circ}\,\mathrm{cm}^{-1}. \tag{2.1}$$

2.3.1 Camera



(a) Mirrors of IACT 1 with the five additional hexagonal mirrors





(b) Winston Cone Sketch

(c) Front view of the Camera

Figure 2.7: Images courtesy TAIGA collaboration. (a) Image taken from www.taiga-experiment.info and cropped, (b) Taken from TAIGA internal presentation of Andrey Grinyuk, (c) Taken from TAIGA internal presentation of Martin Tluczykont. The blue circle in (a) highlights one of the small reflection shields used to blow hot air onto the mirrors to prevent ice formation.

The camera sits in a well insulated cylindrical housing with an outer diameter of 110 cm and a length of 107 cm. From the front, the camera is protected by a plexiglass window with a thickness of 1.5 cm and an average transmission probability of ∼0.82 [53] (see Figure 2.9b for wavelength-dependent transmittance). Between the plexiglass window and the entrance of the Winston cones, there is a gap of 5 cm that allows the circulation of hot air. Figure 2.7c provides a front view of the camera with the Winston cones and the grey hot air outlets of the camera heating system. Additionally, eight LEDs are installed on the camera to allow the determination of the camera alignment with respect to the CCD camera.

A sketch of the Winston cones is shown in Figure 2.7b. The pixel spacing of the IACT is $w_{pix} = 3$ cm and the cone opening is 2.566 cm. Due to the wall thickness of 4.33 mm, the sensitive area is ∼80 % of the sensitive area of a cone with a wall thickness of 1.5 mm. It is planned to install Winston cones with thinner walls in the future.

Figure 2.8 shows the transmission probability against the angle θ of the incoming photons with respect to the optical axis of the Winston cone with PMT assembly. The measurement was carried out by Nima Lubsandorzhiev by illuminating the cone with a LED (+ diaphragm (30 cm from the LED) from 7 m distance to the Winston cone with a PMT attached. The average anode current for different orientation angles θ of the cones is recorded. As the visible cross section of the cone entrance reduces with increasing θ, the currents are scaled by $\frac{1}{\cos\theta}$ to obtain the relative transmission probability. The efficiency of the Winston cone e_{cone} was estimated to be $\frac{I_{cone}}{I_{naked}} \approx 2.6$, where I_{cone} is the current with the cone attached and I_{naked} is the current without the cone.

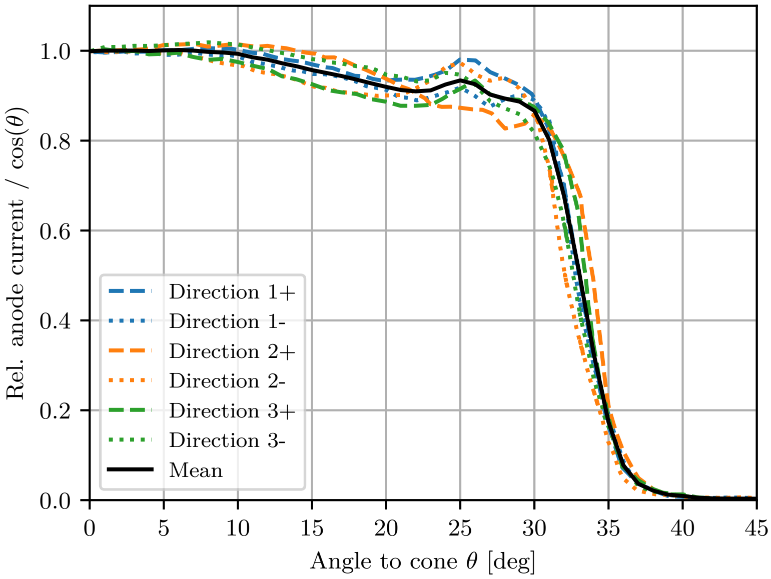
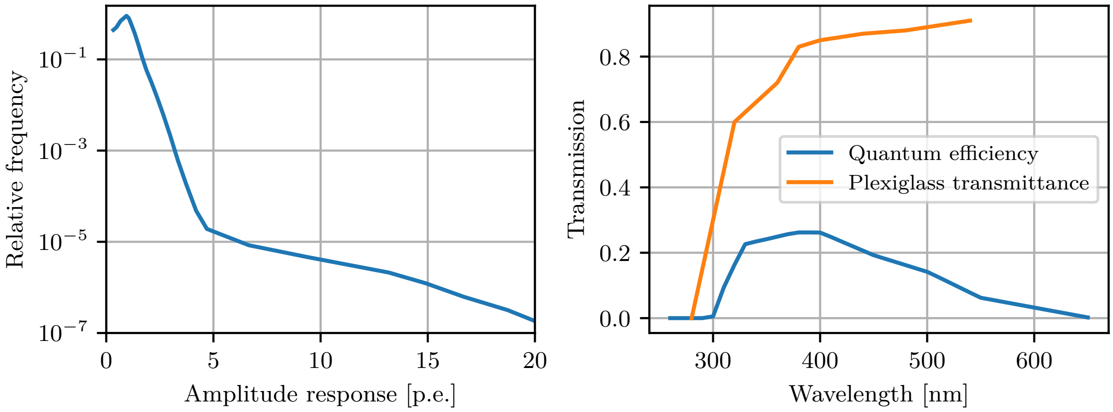


Figure 2.8: Measurement of the TAIGA funnel acceptance by Nima Lubsandorzhiev. The measurement was conducted for three different funnel rotations. + and - denote the rotation in the opposing rotational directions.

2.3.2 Photomultiplier Tubes

The camera uses 560 Philips XP1911 PMTs on IACT 1 with 15 mm active cathode diameter. Figure 2.9b shows the wavelength dependent quantum efficiency peaking at 0.26. Another important characterisation is the response of the PMT to a single photoelectron. It is given as a probability distribution and should describe the average behaviour of the PMTs in the experiment, including the effects of the wavelength, angle and position dependence of the photon collection efficiency. No specific experiment using the same position and Winston cones was conducted. However, a measurement of the PMT was carried out in [54]. To obtain the probability distribution shown in Figure 2.9a, the measurement of the trigger rate over different trigger thresholds was differentiated and transformed into photon equivalents p.e. with the conversion factor of 90 mV p.e.^{-1} given in the text. Starting at about 1.5 p.e. it appears to follow a steep exponential distribution that fades into a flatter exponential distribution caused by the afterpulsing of the PMT (see [54]).



(a) XP1911 Single photoelectron response

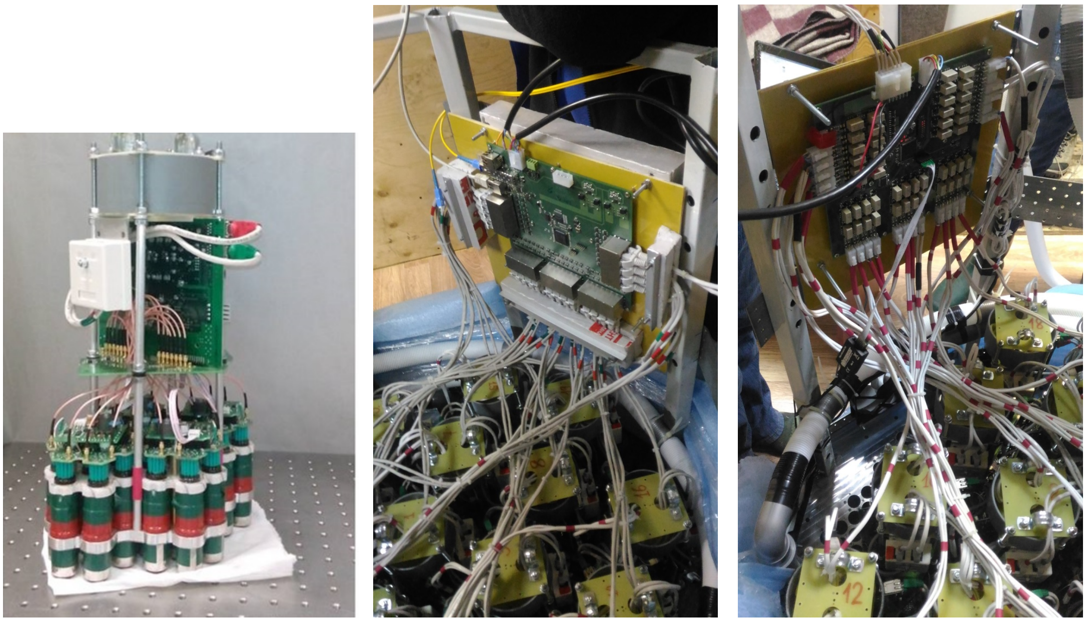
(b) XP1911 Quantum efficiency

Figure 2.9: (a) Single photoelectron response as derived from a measurement by Mirzoyan et al. [54]. (b) Quantum efficiency of the XP1911 PMTs as measured bei Razmik Mirzoyan.

2.3.3 Camera Electronics

The PMTs are operated at a gain of $\sim 1 \times 10^5$ and a corresponding current of $\sim 1\,\mu\text{A}$ with a negative high voltage (HV) of ~ 1.1 kV connected to the cathode. To prevent discharge to the tube surrounding the PMT, the tube is connected to half of the amplitude of the HV.

Groups of seven PMTs (one central PMT and its 6 next neighbours) are directly attached to the cross-boards hosting the HV source and the voltage dividers for the dynodes. As each board only has one adjustable HV level, PMTs with a similar product of gain and



(a) Cluster assembly

(b) Central Controller (CC)

(c) 12 V and 24 V Power Management Board

Figure 2.10: (a) One full cluster assembly with the MAROC readout board on top and four times the PMT assembly with Divider Board and High Voltage Supply directly attached (7 PMTs each). (b) Central Controller installed above the clusters in the camera. (c) Power Management Board installed in the Camera. The visible pipes are for the camera heating system. Images courtesy TAIGA collaboration.

quantum efficiency are selected. Four groups are then attached to one readout board. Figure 2.10a shows the assembly of one of the clusters. The full camera consists of up to 24 clusters. The readout board (MAROC board) is based on the MAROC 3A chip for amplitude readout and trigger forming of up to 64 channels. An FPGA (Cyclone I on the first boards, Cyclone III on later revisions) is used to control the MAROC chip and communicate with the Central Controller (CC) of the camera.

On the MAROC board, the anode signal of each PMT is fed to a current measurement circuit and to two channels of the MAROC chip (see Figure 2.11), one for low gain and one for high gain.

Then, the anode signal is processed by the MAROC 3A readout chip. An outline of the components is shown in Figure 2.12.

The first stage on the MAROC ASIC itself is the pre-amplification via current mirrors by factors between 0 and 4 in steps of 0.0156. The high and low gain channels are set to values of ~150 and 15, which correspond to amplifications of ~2.3 and 0.234.

After the preamplification, the signal is fed to the Slow Shaper for amplitude readout and, for the high gain, also to the half bipolar Fast Shaper FSB2 (where half means that

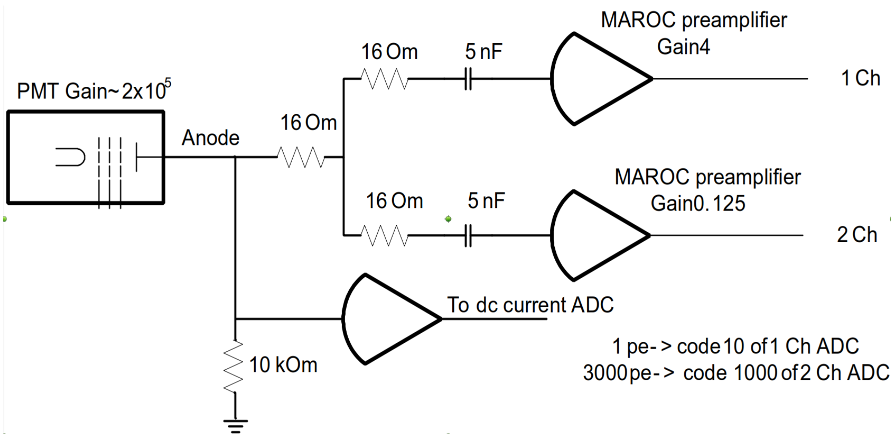

MAROC preamplifier
Gain4
16 Om
5 nF
1 Ch
PMT Gain~2x10^5
Anode
16 Om
MAROC preamplifier
Gain0.125
16 Om
5 nF
2 Ch
To dc current ADC
1 pe-> code 10 of 1 Ch ADC
3000pe-> code 1000 of 2 Ch ADC
10 kOm

Figure 2.11: Connection of the XP1911 PMTs to the MAROC chip. Taken from Kompaniets [55].

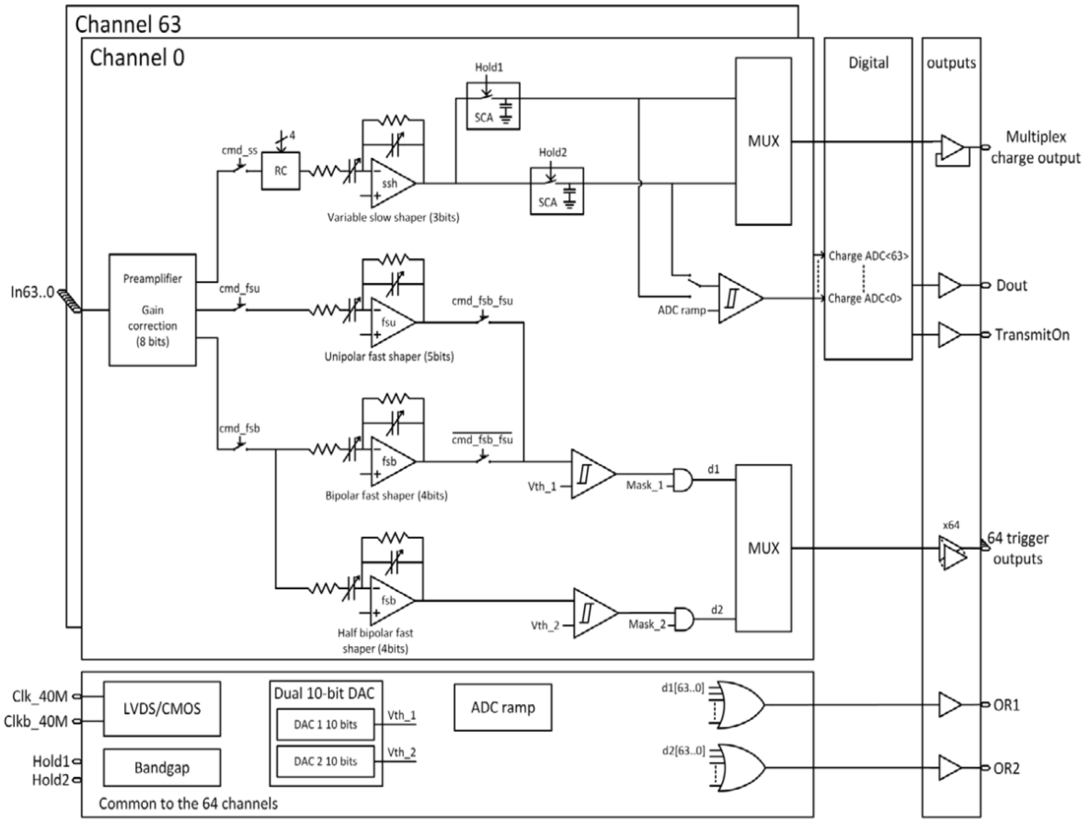

Channel 63
Channel 0
Hold1
SCA
4
cmd_ss
RC
ssh
Variable slow shaper (3bits)
Hold2
SCA
MUX
Digital
outputs
Multiplex
charge output
In63..0
Preamplifier
Gain correction (8 bits)
cmd_fsu
fsu
Unipolar fast shaper (5bits)
cmd_fsb_fsu
ADC ramp
Charge ADC<63>
Charge ADC<0>
Dout
TransmitOn
cmd_fsb
fsb
Bipolar fast shaper (4bits)
cmd_fsb_fsu
Vth_1
Mask_1
d1
MUX
x64
64 trigger
outputs
fsb
Half bipolar fast shaper (4bits)
Vth_2
Mask_2
d2
Clk_40M
Clkb_40M
LVDS/CMOS
Dual 10-bit DAC
DAC 1 10 bits
DAC 2 10 bits
Vth_1
Vth_2
ADC ramp
d1[63..0]
OR1
Hold1
Hold2
Bandgap
d2[63..0]
OR2
Common to the 64 channels

Figure 2.12: Overview of the MAROC3 ASIC. Taken from Fig. 2 of *MAROC 3A Datasheet* [56].

the signal amplitude is halfed) which is used for triggering.

For IACT 1, the FSB2 is used with a capacitance C of $50\,\mathrm{fF} + 20\,\mathrm{fF} = 70\,\mathrm{fF}$ and a resistance R of $100\,\mathrm{k}\Omega$. With these settings, FSB2 has a time constant of $\tau = RC \approx 7\,\mathrm{ns}$. A typical response to a short signal is shown in Figure 2.13. The Fast Shaper signals of all channels are fed to discriminators which have a common threshold set to $\sim$10 p.e.

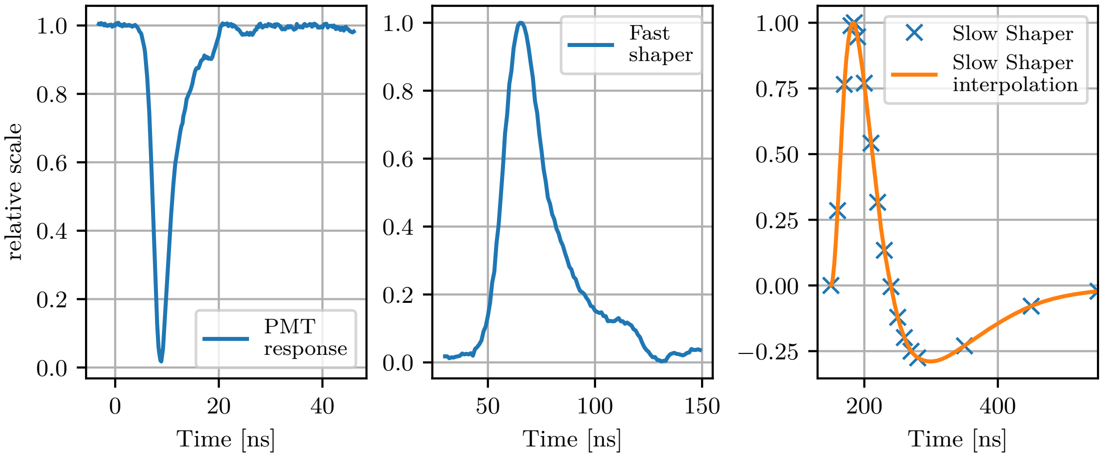


Figure 2.13: Measured responses of the electronics to short 'single p.e.' like signals. The PMT response and Fast Shaper signals were measured by N. Lubsandhorziev and The Slow Shaper waveform was measured by E. Kompaniets.

(∼12 p.e. after mirror addition) and are connected to output pins (one for each channel) routed to the FPGA.

The RC buffer before the Slow Shaper is not used and the signal is fed directly to the Slow Shaper which is set to a feedback capacitance of 0.3 pF with a feedback resistance of 100 kΩ in IACT 1. For the input RC pair, the capacitance is 0.9 pF and the resistance is 33 kΩ, resulting in a time constant of 30 ns. A typical response to a short signal is shown in Figure 2.13.

For the Analog to Digital Converter (ADC), the MAROC 3A internal Wilkinson ADC is used in 12-bit mode, giving an analog count range from 0 to 4095 codes.

2.3.4 Trigger and Readout Logic

An overview of the timing is given in Figure 2.14. When one of the triggers on the MAROC chip fires, the FPGA waits for the hold delay t_h before emitting the local hold signal to the Hold1 pin of the MAROC chip. This signal causes the amplitudes of the Slow Shapers to be stored in analogue memories. The local hold signal will stay active for ∼100 ns. The second task on the MAROC board is to check the coincidence of 2 triggers within a trigger window of 15 ns. When a local coincidence of any 2 pixels (no next neighbour condition) is met, the MAROC board will notify the Central Controller of a trigger and transmit the binary trigger information per channel.

On the Central Board, the trigger time is assigned at the next phase of the 100 MHz reference signal (giving an accuracy of ≈10 ns) and the so-called global hold signal is sent to all MAROC boards. On each MAROC board, it is also forwarded to the Hold1

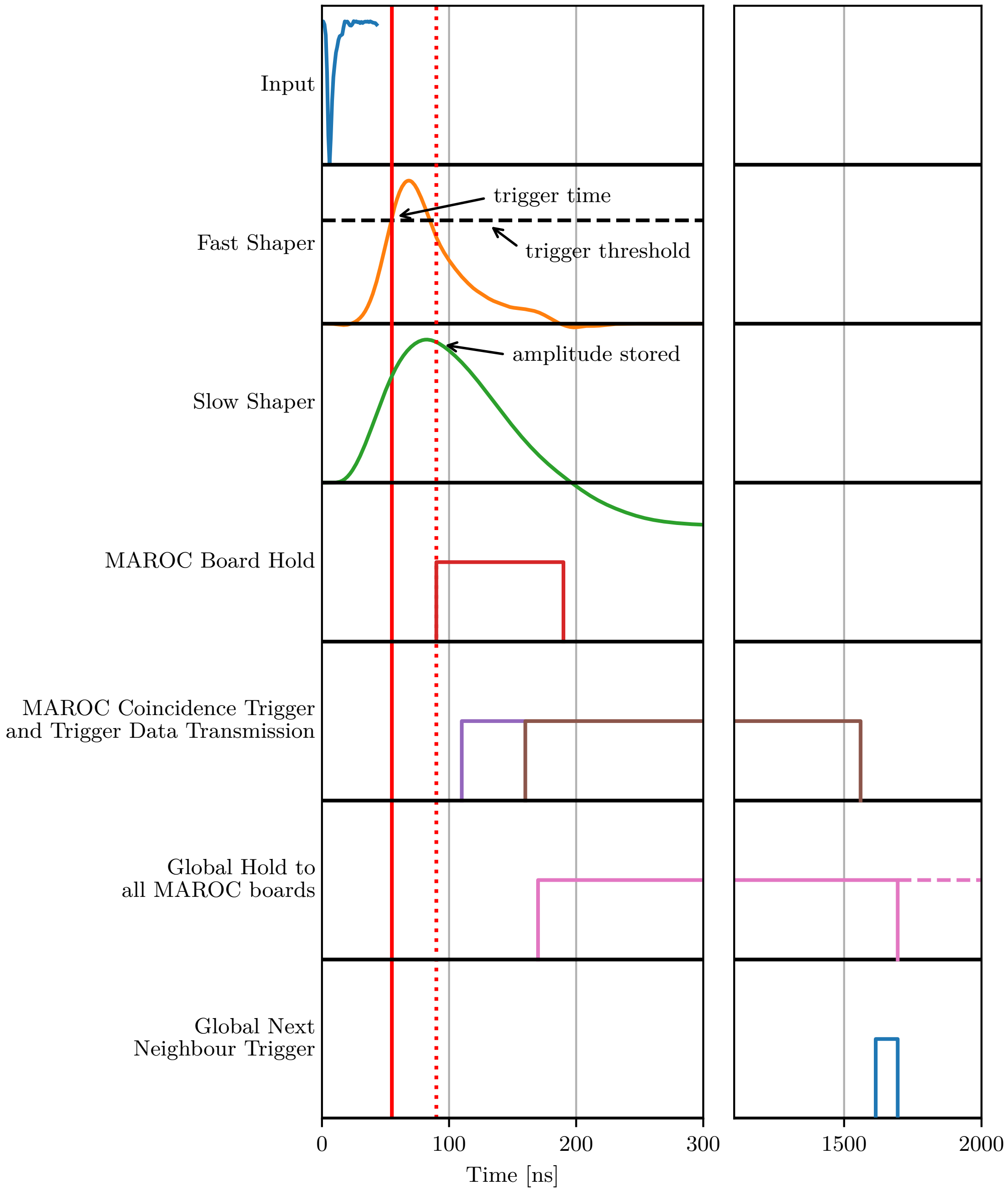


Figure 2.14: Trigger and readout scheme of the first TAIGA IACT camera. When the Fast Shaper signal reaches the trigger threshold, the discriminator fires (red vertical line). After the hold delay time t_h, the MAROC board issues a local hold signal, storing the amplitudes of the Slow Shaper signals of all channels on this MAROC board in a Track and Hold circuit (red dotted vertical line). When two PMTs trigger within 15 ns on this MAROC board, the trigger information is sent to the Central Controller, where a global hold signal is sent to all MAROC boards of the camera. Then, the topology trigger condition is checked (next neighbours) and if not met, the hold signal is reset. On success, the global hold signal is kept and the readout command is sent. Then, all MAROC boards in local hold will start the ADC and transmit the amplitude information. This scheme is an adaption of figures from Kompaniets [55] and *MAROC 3A Datasheet* [56].

pin of the MAROC chip, just like the local hold signal. The boards which are in local hold will then be kept in hold for the duration of the global hold.

The trigger information of all participating MAROC boards is then checked for the topological trigger condition of two next neighbours per MAROC board. When the trigger condition is not met, the global hold is reset, and the system is ready for the next trigger. If the trigger condition is met, the global hold is kept and the data readout of the triggered MAROC boards is started. The digitisation and transmission of the signal takes ≈180 µs.

The hold delay t_h is set such that the amplitude ∼10 ns after the time of the maximum of the Slow Shaper response signal to a short signal at the triggering channel is saved.[2] This setting ensures that very bright signals (Early trigger, well before maximum) and signals arriving later on other PMTs on the same MAROC board still give amplitude readings close to the maximum of the Slow Shaper waveform.

2.3.5 Calibration

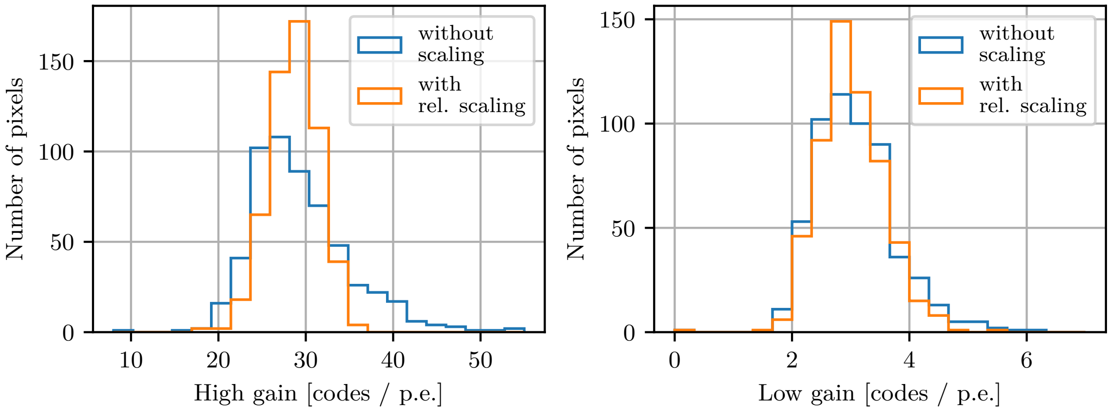
In order to achieve a roughly homogeneous response throughout the camera, the HV of each group of 7 PMTs is adjusted such that all groups have equal average current draw. For the high-gain channels, further readout homogenisation is achieved by the adjustment of the MAROC preamplifier setting for each PMT (see also Section 2.3.3). For the high gain, the amplification is set to ∼2.3 and the possible steps of 0.0156 give a sufficient dynamical range. However, this is not done for the low-gain channel with an amplification of 0.234 where one step in the preamplification setting corresponds to an amplification change of ∼6.7 %, not sufficient for a more precise adjustment. For calibration, the LED with diffuser mounted on the mirror frame a few cm offset from the center is operated in flashing mode. The calibration constants k are then determined with the F-factor method [57, 58] using the equation

$$k = \frac{\sigma^2_{LED} - \sigma^2_{ped}}{F^2(\mu_{LED} - \mu_{ped})}, \tag{2.2}$$

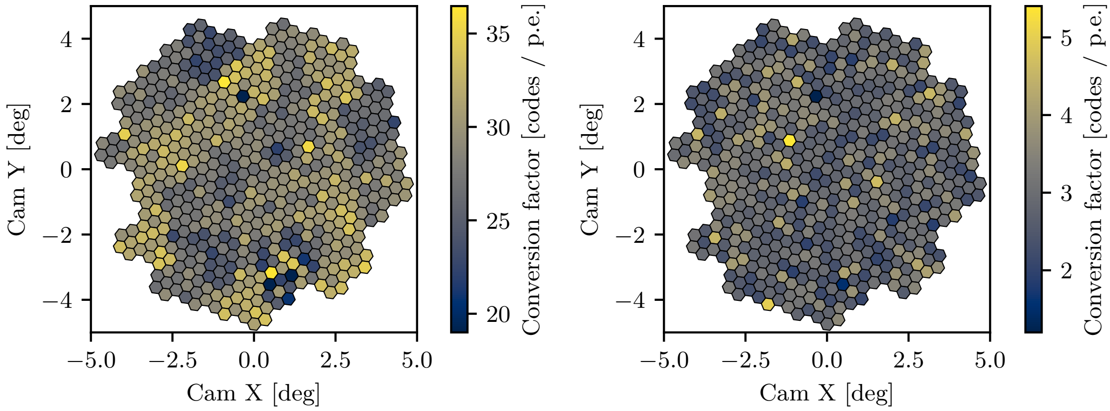
where F is the excess noise factor and μ and σ^2 are the mean value and variance of normal fits to the raw code distribution of the LED or pedestal signal. In this work, the calibration factors obtained by Lubsandorzhiev [59] with $F^2 = 1.3$ are used. Additionally, relative calibration factors between the pixels are determined from deviations to a two dimensional fit of the expected illumination pattern.

---

[2] Internal communication



(a) High Gain Histogram (b) Low Gain Histogram



(c) High Gain Cam (d) Low Gain Cam

Figure 2.15: Histograms and per pixel plots of the calibration constants of IACT 1. The values were obtained by the same author with the same method as described in Lubsandorzhiev et al. [60]. The small reduction of the spread with relative scaling for the low gain is a result of flatfielding with the preamplification. Because of the dynamic range of the preamplifier, only the high gain is adjusted (see text).

Figure 2.15 shows histograms of the conversion factors k mapping digitized values from codes to photon equivalents p.e.. It is apparent that the calibration factors without the relative calibration show a distinct shoulder to larger values. For the high gain, the spread is minimized with the relative calibration. For the low gain, the effect of the relative calibration is minimal because no flatfielding with the preamplification can be applied.

2.3.6 Pointing accuracy

As described by Zhurov et al. [61, 62], the telescope azimuth and altitude axes are controlled with Phytron hybrid stepper motors controlled with a PhyMOTION control unit and the position is monitored via 17-bit shaft encoders. However, it was found, that due to a hardware problem, relying only on the encoder feedback alone was not precise and reliable enough. Therefore, on-the-fly pointing updates calculated from CCD-camera images were implemented. [61]

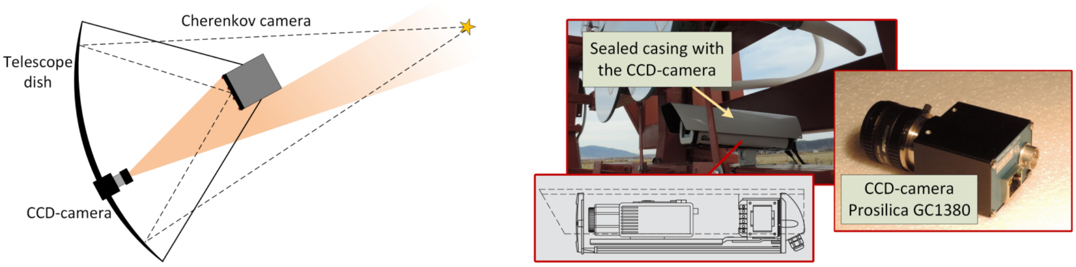

Cherenkov camera
Telescope dish
CCD-camera
Sealed casing with the CCD-camera
CCD-camera Prosilica GC1380

Figure 2.16: On the left, a sketch visualises the field of view of the CCD camera. On the right, images of the camera prior to installation and when mounted on the telescope are shown. Taken from Zhurov et al. [61].

The CCD camera is located slightly behind the mirror plane 1 m from the center at a distance of 483 cm from the center of the cameras Winston cone entrance. It has a field of view of 31.4° × 23.6° at a resolution of of 1360 px × 1024 px resulting in a resolution of 0.023° per pixel. With automated detection of the stars and of the camera positioning LEDs (8 LEDs around the camera perimeter), the position and orientation of the camera with respect to the observed sky can be reconstructed. CCD images are taken with an integration time of 5 s every 20 s[49] during observation.

Figure 2.16 shows the Prosilica GC1380 CCD camera and its position on the telescope. An example image taken during observations is displayed in Figure 2.17.

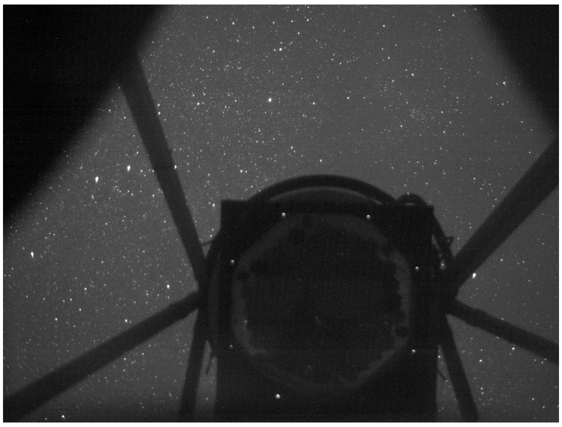


Figure 2.17: An image taken by the CCD camera of the pointing system. The image was taken during normal observation on 2019-11-07. It shows the camera and its support masts as well as a part of the sky also within the field of view of the telescope. The brighter spots around the inner hexagonal perimeter of the camera are the positioning LEDs.

Processing takes place online to provide on-the-fly feedback to the telescope control software. A crosscheck of the CCD image position determinations against the positions determined from current scans as shown in Figure 2.18 resulted in an accuracy estimate of <0.02°. For the current scans, the telescope is steered such that a star moves across a pixel in two perpendicular directions, forming a cross. [62]

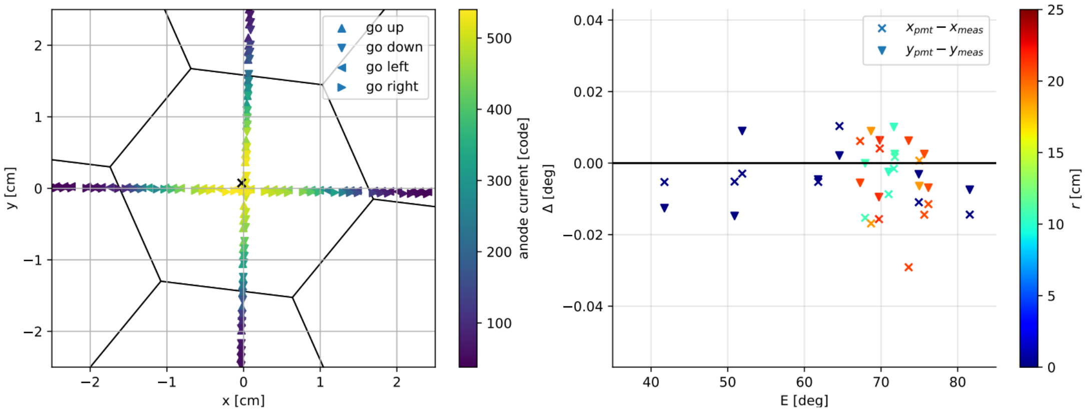


Figure 2.18: On the left, a current scan through the central pixel of the camera is shown. The color scale of the marker codes the measured anode current at that position. ∼80 codes correspond to ∼1 µA. On the right, the distance between the center position estimated with the CCD camera to the center position estimated with the PMT currents is shown as a function of the telescope altitude. The image was taken from Zhurov et al. [62].

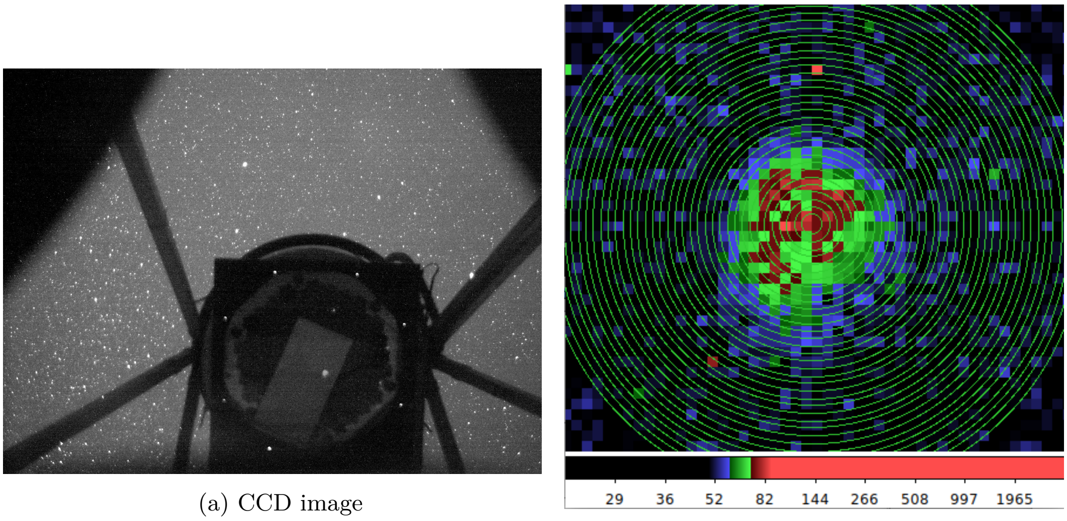
2.3.7 Optical Point Spread Function

Another important parameter of the IACTs performance is the quality of the telescope optics usually described by the optical Point Spread Function (PSF). It is a measure of the size of the spot of a point-like light source on the Winston cone entrance plane. One way to estimate the PSF is to measure the spot sizes with the CCD-camera used for the pointing corrections. Using the image available on the TAIGA data server and shown in Figure 2.19, the PSF was estimated (i.e. the following calculation is my work).

Figure 2.19 shows an image taken while the telescope was tracking a bright star.

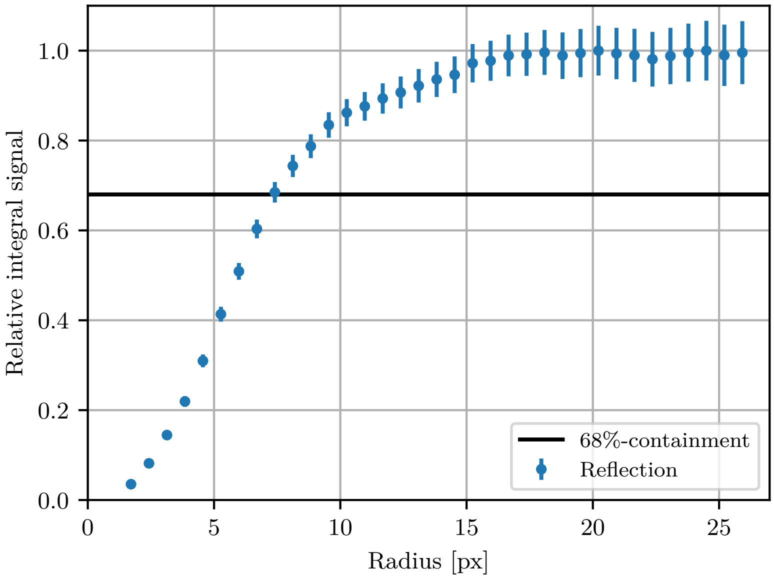
To make the reflection of the star visible, a screen was positioned in front of the Winston cones and behind the plexiglass as shown in Figure 2.19a. The reflection was analysed with the Annulus region tool of the SAOImageDS9 [63] application. Figure 2.19b shows the placement of the annuli in the reflection. Figure 2.19c shows the relative integral signal N_I above the baseline B of all pixels n_i

$$N_I(r) = \sum_{i}^{n_r} N_i - B \tag{2.3}$$



(a) CCD image

(b) Placement of the annuli on the reflection



(c) Relative integral signal of the reflection

Figure 2.19: (a) CCD image taken when the telescope was pointing to a bright star (Mirach). On the camera, the reflection of the star on a screen installed in front of the light guides is visible. The image was taken on 2019-09-28 at an approximate elevation of 38° (b) Screenshot of the annuli selected in SAOImageDS9 [63]. (c) Relative integral signal of the reflection assuming a background pixel intensity of 48.6.

contained within the annuli of radius r. N_i is the signal of one pixel. The uncertainty σ_{N_I} is estimated via

$$\sigma_{N_I}(r) = \sqrt{\sum_{i}^{n_r} N_i}. \tag{2.4}$$

A baseline value of $B = 48.6$ is assumed for each pixel and it is chosen such that the integral above a radius of 18 px becomes flat (no signal above baseline). For Figure 2.19c the normalisation value is taken from the points above $r = 15$ px. The value of r where 0.68 is crossed corresponds to the 68%-containment radius r_{68}. For the uncertainty estimation, slight changes in the baseline of 0.1, slight movements of the center position and the statistical errors were taken into account, resulting in an estimate of

$$r_{68} = (7.4 \pm 0.5)\,\mathrm{px}.$$

As the analysis is based on the raw counts given by the CCD camera, two main assumptions are made: (a) a similar gain and baseline for every CCD-pixel and (b) a sufficiently small PSF of the CCD camera. Due to the flatness of N_I above $r \approx 18\,\mathrm{px}$, (a) appears to be an appropriate assumption when considering the average behaviour of several pixels. As the intensity distribution of the reflection stretches over several pixels, a constraint of at least 10 pixels per annulus could be met. To estimate the PSF of the CCD-camera itself, regions around stars with a brightness below the saturation level of the CCD-camera were checked, giving an estimate of the CCD PSF of $\leq 1.5\,\mathrm{px}$. Because of the CCD PSF and because of the slight image distortion expected from the second passage through the plexiglass, r_{68} can be considered as an upper limit.

One CCD-camera pixel corresponds to about

$$2d_{ccd}\tan\left(\frac{\omega_{ccd}}{2}\right) \approx 0.194\,\mathrm{cm\,px^{-1}} \tag{2.5}$$

in the center of the entrance plane of the Winston cone, where $d_{ccd} = 483\,\mathrm{cm}$ is the distance of the CCD camera to the cones and $\omega_{ccd} = 0.023°$ is the FoV of one CCD pixel. This value can be further converted to angular extensions as seen by the telescope camera (see Equation 2.1), giving

$$r_{68} \leq (7.4 \pm 0.5)\,\mathrm{px} \approx (1.43 \pm 0.10)\,\mathrm{cm} \approx (0.17 \pm 0.01)°, \tag{2.6}$$

what is compatible with the range of values estimated in Kunnas [36]. The only other comparison value available is the assumed blurring of 10 mm in the focal plane used in Postnikov et al. [64]. It is however unclear to what quantity it corresponds to. If it corresponds to the standard deviation of a symmetric 2D normal distribution, it is ~0.18° and close to the estimated value. If it is the 68%-containment radius, it is ~0.12° and closer to but slightly smaller than the values resulting from the simulation described in Section 3.2.1 (see also Figure 3.6), where the PSF according to the simulation is investigated in more detail.

2.3.8 Raw data processing

The data is stored in binary files for each run portion per MAROC board. Each run has 22 subdirectories, one for the binary files of each MAROC board. Depending on the length of the run, a run may have more than 100 portions, each with a length of ∼2 min. To form the image of an event, the files of the 22 clusters need to be checked and merged.

Additionally, data of some events may be split across portions. Data at the end of one portion for one cluster may, for another cluster, be stored in the beginning of the binary file for the next portion.

Potentially useful metadata is stored in several text files in the run directory. The status log of the observation program contains information about disabled channels and trigger settings (for the first seasons, this was moved to two separate files in the 21/22 season) and of the currently synchronised time of the GPS subsystem and of the system time of the machine controlling the IACT. The current measurements conducted between portions are stored in a separate file. From the pointing subsystem (Section 2.3.6), the tracking information such as telescope altitude, azimuth and celestial pointing is stored in yet other files. Furthermore, for some runs, there are further files containing metadata of the observation program such as wobble changes.

The C/C++ program chesnok[3] was written to read the individual binary files, merge the events and write them to HDF5 file format, which is a widely adopted format with readers available in many languages and packages. After making sense of the information found in the various text files, parsing algorithms were added to store the metadata in the HDF5 file. Whenever the available files or the format of the files changed, the code was adapted to allow the analysis of the new data.

In addition to the merging process, the program also corrects the timestamps of the events for a potential time difference of the MEGA host and the 'Computer time'.

Based on the HDF5-file, the rest of the data analysis is carried out in python using mainly numpy[65], scipy[66] and pandas[67, 68]. For coordinate transformations, astropy[69, 70] was used and HDF5 handling was done with h5py[71]. The code for the transformation from camera coordinates to right ascension / declination or altitude / azimuth was based on the implementation in pyfact[4].

2.3.9 Pedestal estimation

Without any additional signal, each of the channels will give readings around a baseline value, called pedestal. This nonzero pedestal is usually required for the electronics to work in the desired operation range.

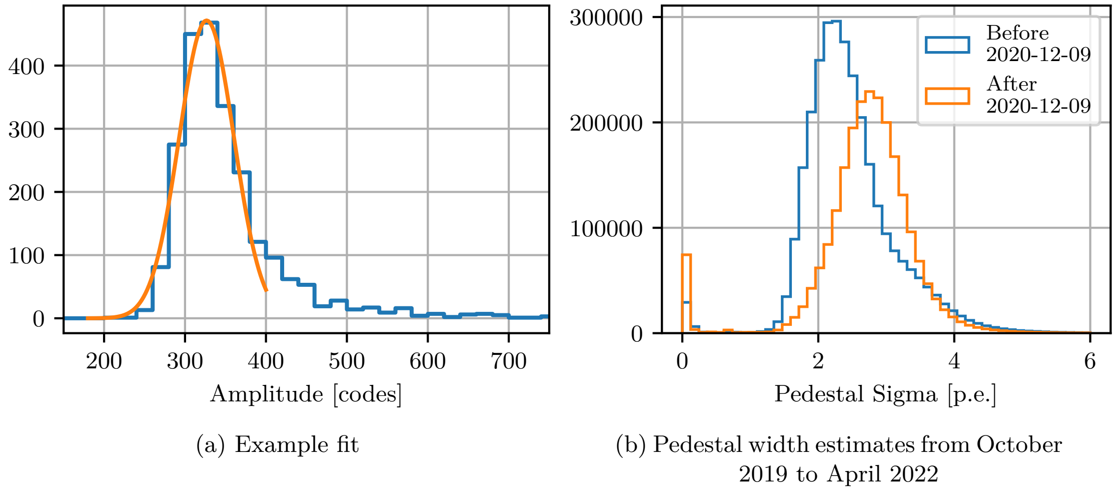
---

[3] `https://gitlab.rrz.uni-hamburg.de/TAIGA/io_taiga`

[4] `https://github.com/fact-project/pyfact/tree/master/fact/coordinates`

If the camera lid is open and NSB light hits the pixels, the spread around the pedestal value is dominated by the NSB level. However, as the signal of air showers and PMT afterpulsing also contributes, there is a tail to large amplitudes.

To correctly interpret the signal amplitudes in units of ADC bits, termed codes, this pedestal value needs to be estimated.



(a) Example fit

(b) Pedestal width estimates from October 2019 to April 2022

Figure 2.20: (a) Histogram of the raw data amplitudes of one pixel of one chunk (usually 100000 consecutive events) of one run. The orange curve visualizes the result of the fit of a normal distribution to the bins around the pedestal maximum. The fit results in a pedestal estimate of 317.5 with a width of 2.15 p.e. (b) Histogram of all pedestal width estimates of 3 seasons. The blue histogram is without and the orange histogram is with the five additional hexagonal mirrors installed on 2020-12-09.

As no regular pedestal measurements are conducted throughout the observation, the pedestals need to be estimated from the raw data available. This is done based on the raw amplitude distributions of a set of events, one chunk. Typically, 100k events are used for one chunk to ensure enough entries for clusters which have a lower trigger rate. In a two-stage process, a histogram with values y_i for bins i with a binning of 20 codes between 20 codes and 2000 codes is analysed. First, the first peak of the histogram is determined using the `scipy` function `find_peaks` (settings: $\mathit{prominence} = 0.05$, $\mathit{distance} = 5$) on the relative histogram values (y_i/y_{max}) and then the pedestal value is estimated as the mean μ of a fit of a normal distribution

$$N(x) = Ae^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{2.7}$$

to 7 bins before and 4 bins after the maximum bin determined. A is a normalisation constant and σ is the standard deviation used as the pedestal width estimate. An example fit is shown in Figure 2.20a. Additionally, pixels with a σ below 1 and above 5 are excluded from the analysis. This removes pixels with non-standard behaviour (e.g. star in FoV, electronic defect, etc.).

Because the width of the amplitude distribution depends on the NSB level, the standard deviation σ of the fit is also a good indicator of the NSB level. Figure 2.21 shows the median μ and σ of all pixels for each of the chunks over several months.

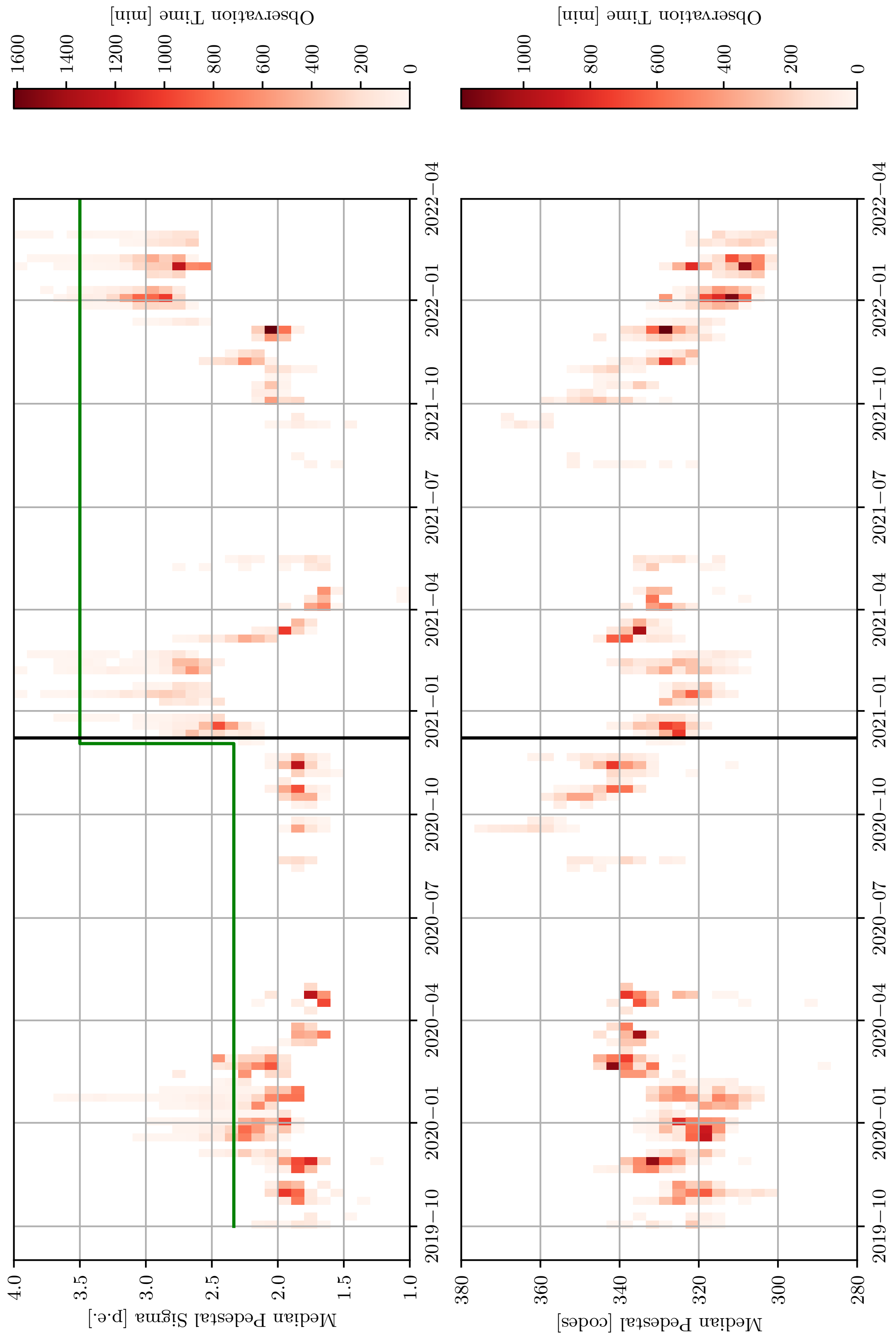
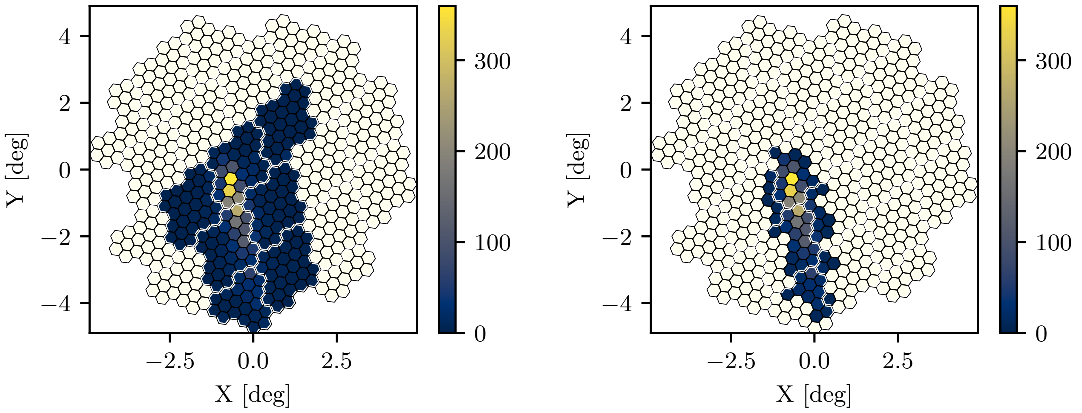


Figure 2.21: Evolution of the median value of the pedestal fit standard deviations and pedestal estimates over the course of the seasons in weekly binning. The vertical black line indicates the date of the installation of the hexagonal mirrors on the five remaining spots.

2.3.10 Cleaning

For most of the events, the light from the air shower is confined to a few pixels. All other pixels show an unwanted signal dominated by fluctuations caused by NSB light. To keep only the light of the showers and to disregard the noise, a cleaning procedure is applied. All pixels above a threshold l_1 will be kept. Additionally, all pixels directly neighbouring pixels above l_1 will be checked, and if they read amplitudes above l_2, they are also kept.

Figure 2.22 shows an example of an image before and after cleaning. The cleaning thresholds used in this analysis are $l_1 = 14\,\text{p.e.}$ and $l_2 = 7\,\text{p.e.}$ from October 2019 to December 2020 (split 1) and $l_1 = 21\,\text{p.e.}$ and $l_2 = 10.5\,\text{p.e.}$ starting from 2020-12-08 (split 2). For a pedestal with $\sigma = 2\,\text{p.e.}$, $l_1 = 14\,\text{p.e.}$ corresponds to $\approx 7\sigma$, giving a very low probability for one pixel to be kept due to NSB fluctuation. The value is chosen such that it is still reasonable at higher NSB level, so that it is reasonable to use one level for each of the splits. For a more detailed estimation of the amplitude fluctuations caused by the NSB light, see Sec. 3.2.4.



(a) Calibrated camera image

(b) Cleaned image

Figure 2.22: (a) Calibrated image of an event recorded on 2019-11-23. The double white lines indicate the 22 clusters. The ivory colored pixels are from clusters which were not read out. (b) Cleaned version of the image in (a). The cleaning levels used are $l_1 = 14\,\text{p.e.}$ and $l_2 = 7\,\text{p.e.}$.

2.3.11 Image Parameters

On the cleaned images, a moment analysis is performed as first proposed by Hillas [33]. Figure 2.23 shows a sketch explaining the basic parameters. The main parameters for a camera image of n pixels with intensities I_i and coordinates x_i, y_i are:

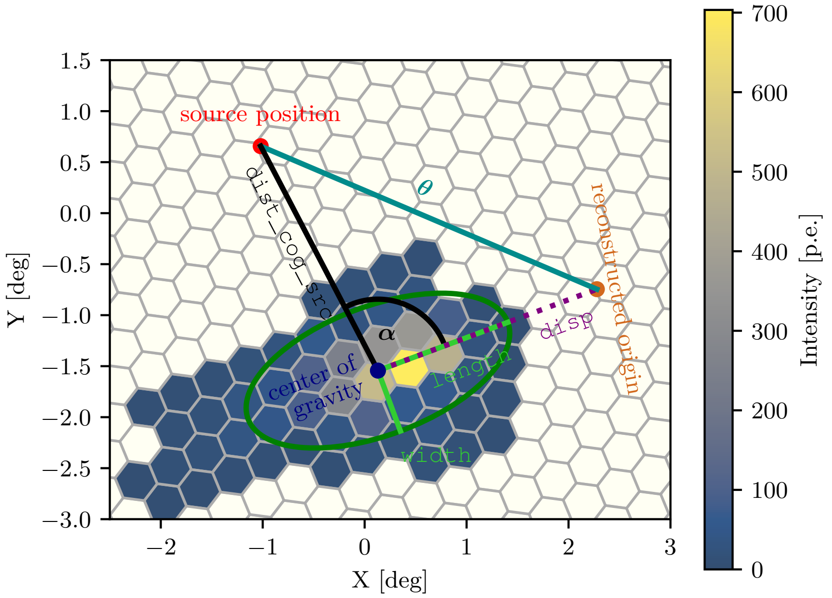


Figure 2.23: Basic image parameters.

n_pixels	Number of pixels surviving the cleaning
size	Total sum of amplitudes $\sum_i^n I_i. \qquad (2.8)$
cog_x	First moment (Center of Gravity (CoG)) along the x-axis of the camera $\texttt{cog\_x} = m_{10} = \frac{\sum_i^n I_i x_i}{\texttt{size}}. \qquad (2.9)$
cog_y	First moment along the y-axis of the camera $\texttt{cog\_y} = m_{01} = \frac{\sum_i^n I_i y_i}{\texttt{size}}. \qquad (2.10)$
length	Square root of the second central moment along the semi-major axis.

The coordinate system x', y' along the semi-major and semi-minor axes of the ellipse is defined by the perpendicular eigenvectors of the covariance matrix Cov defined as

$$Cov = \begin{pmatrix} \frac{\sum_i^n I_i x_i^2}{\texttt{size}} - {m_{10}}^2 & \mu_{11} \\ \mu_{11} & \frac{\sum_i^n I_i y_i^2}{\texttt{size}} - {m_{01}}^2 \end{pmatrix} \tag{2.11}$$

with $\mu_{11} = \frac{\sum_i^n I_i x_i y_i}{\texttt{size}} - m_{10} m_{01}$. In the (x', y')-basis with its zero-point in the CoG position, `length` is defined as

$$\texttt{length} = \sqrt{\frac{\sum_i^n I_i {x'}_i^2}{\texttt{size}}}. \tag{2.12}$$

`width` Square root of the second central moment along the semi-minor axis. In the same notation as used for `length`, the `width` is defined as

$$\texttt{width} = \sqrt{\frac{\sum_i^n I_i {y'}_i^2}{\texttt{size}}}. \tag{2.13}$$

`disp` Distance (displacement) between the CoG and the reconstructed source position.[72]

`dist_cog_src` Distance between the CoG and the source position in the camera.

`dist_cog_center` Distance between the CoG and the center of the camera.

θ Distance between reconstructed source position and nominal source position.

α Angle between the major axis and the vector CoG to source position in the camera.

`M3l` Third central moment along the major axis. In the same notation as used for `length`, it is defined as

$$\texttt{M3l} = \frac{\sum_i^n I_i {x'}_i^3}{\texttt{size}}. \tag{2.14}$$

`second_hottest` Second highest pixel intensity of the event. It is used instead of the highest pixel intensity because it should be less prone to fluctuations, especially for low intensity events. Furthermore, when using a next-neighbour trigger and a lower cleaning threshold well below the trigger thresholds, its distribution can be used to infer the trigger threshold.

`conc2`	Measure of the concentration of the image. It is defined as the fraction of the signal in the two brightest pixels with the indices j and k

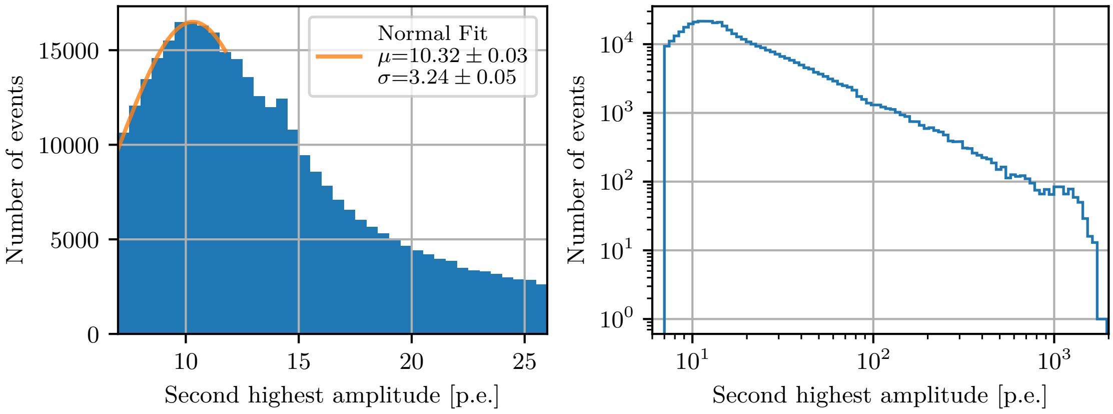
$$\texttt{conc2} = \frac{I_j + I_k}{\texttt{size}}. \tag{2.15}$$

`leakage`	Measure of the truncation of the image. It is defined as the fraction of the signal in the border pixels to the total signal.[73]

$$\texttt{leakage} = \frac{\sum_{\text{border}} I_b}{\texttt{size}}. \tag{2.16}$$

2.3.12 Trigger Threshold

When the trigger condition is set to equal to or more than two pixels, the distribution of `second_hottest` will have its maximum around the trigger threshold of the camera. Figure 2.24 shows the distribution for one run (cleaning $l_1 = 14$ p.e. and $l_2 = 7$ p.e.).



(a) Small values (b) Full range

Figure 2.24: Distribution of the second highest amplitude per event for one run of the first camera. The peak and spread around the peak in (a) is an indication of the trigger threshold of the camera when a ≥ 2-pixel trigger condition is used. It is a combined effect of the thresholds of all pixels. In (b) one can see that the amplitude distribution follows a power law for amplitudes above ~20 p.e.. The slight dip at ~90 p.e. is caused by the switch from high to low gain.

The rising part including the peak should be caused by the combination of the trigger levels and spread of the trigger levels of all pixels participating in the trigger. The normal fit indicates that the trigger threshold (set to ~300 codes of the FSB) corresponds to (10.3±3.2) p.e.. This value was staying roughly constant throughout the first observation period. A slight increase of the threshold to ∼320 codes became necessary after the addition of the hexagonal mirrors.

Above ∼20 p.e., `second_hottest` follows a power law distribution, as expected. The small dip at ∼90 p.e. seen in all runs indicates the switch from high to low gain (switch at a high gain of 3500 codes) and indicates a slight mismatch. At values of ∼1000 p.e. a pileup is visible. This is caused by the maximum range of the ADC for each pixel (4096 codes).

2.4 The Second Imaging Air Cherenkov Telescope

On the TAIGA pilot site, the positions of five telescopes are already planned (see also Figure 3.1). The second IACT is already fully equipped and operational since Feb. 2020 and the third IACT recently received its camera but is not yet in regular operation. The fourth and fifth telescopes are in the planning phase.

Here, IACT 2 is described in more detail. All further telescopes will follow the revised design of IACT 2. The second IACT is similar to the first IACT described in Section 2.3. The main differences are:

- Full camera readout for every event (all clusters).
- Hexagonal mirror tiles on all 34 mounts with a total area of ∼10.6 m^2.
- 595 instead of 560 pixels.
- Larger Slow Shaper capacity (increased integration time).
- No magnetic shielding on individual PMTs.

The main change are the increased pixel size and the accomodation of a full camera readout (as opposed to the clusterwise readout of IACT 1). Unfortunately, the mu-metal housings of the individual PMTs were removed and the sensitivities of the PMTs change by up to 50 % depending on the alignment w.r.t. the earths magnetic field[5], limiting the use in data analysis. The addition of a magnetic shield around the whole camera housing is planned to resolve the situation. Despite this dependence of the gain on the alignment towards the earths magnetic field, IACT 2 is used for the stereo core reconstruction technique in Section 6.1. For this purpose, only the center of gravity position and the direction of the major axis are required, making the magnetic field dependent amplitude variations less important.

---

[5] Presentation of Nima Lubsandorzhiev, TAIGA collaboration meeting 2021-02-18

The full camera readout was realized with an additional Fast Hold Board. On the MAROC boards, the local trigger is now checked for coincidence with a "fast coincidence check" before a hold signal is issued to all MAROC boards in the camera via the Fast Hold Board. Figure 2.14 shows the adapted trigger and readout scheme.

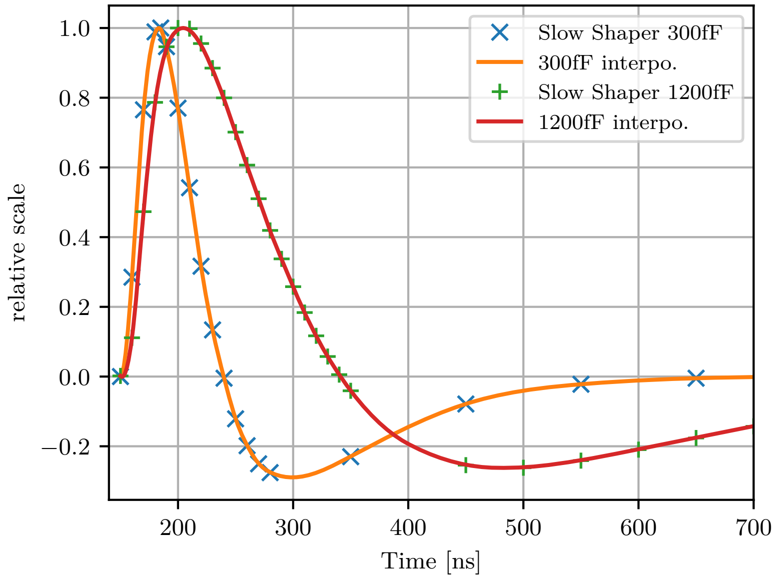


Figure 2.25: Slow Shaper response to short 'single p.e.' like signals as measured by E. Kompaniets via hold delay t_h variation, where t_h is the difference between the readout time and the trigger time. IACT 1 uses a feedback capacitance of 0.3 pF. On IACT 2, a full camera readout is implemented with the Fast Hold Board. The board distributes the hold signal created by any MAROC board to all other MAROC boards. To compensate for the additional transit time, the feedback capacitance is set to 1.2 pF. The waveforms were measured by E. Kompaniets.

As the hold signal distribution takes several nanoseconds, it would arrive well after the Slow Shaper pulse peak when using the settings of the first IACT. Therefore, IACT 2 uses a feedback capacitance of 1.2 pF and a capacitance of 3.6 pF on the input RC-pair, increasing the time constant from 30 ns to 120 ns. A typical response to a short signal is shown in Figure 2.25.

Figure 2.27 shows the calibration constants for the second camera determined via the F-factor method (see Section 2.3.5).

However, as indicated by the distribution of `second_hottest` shown in Figure 2.28, when using these factors for the high as well as the low gain, there is a mismatch visible as the dip at values of ∼100 p.e. in the case of IACT 2. The low gain seems to give too large estimates. As the calibration is done for both channels with the same LED intensity, the high gain shows a stronger signal and is expected to be more reliable. Therefore, the low gain calibration constants are determined from the linear relation apparent when plotting the low gain in raw codes against the high gain in codes.

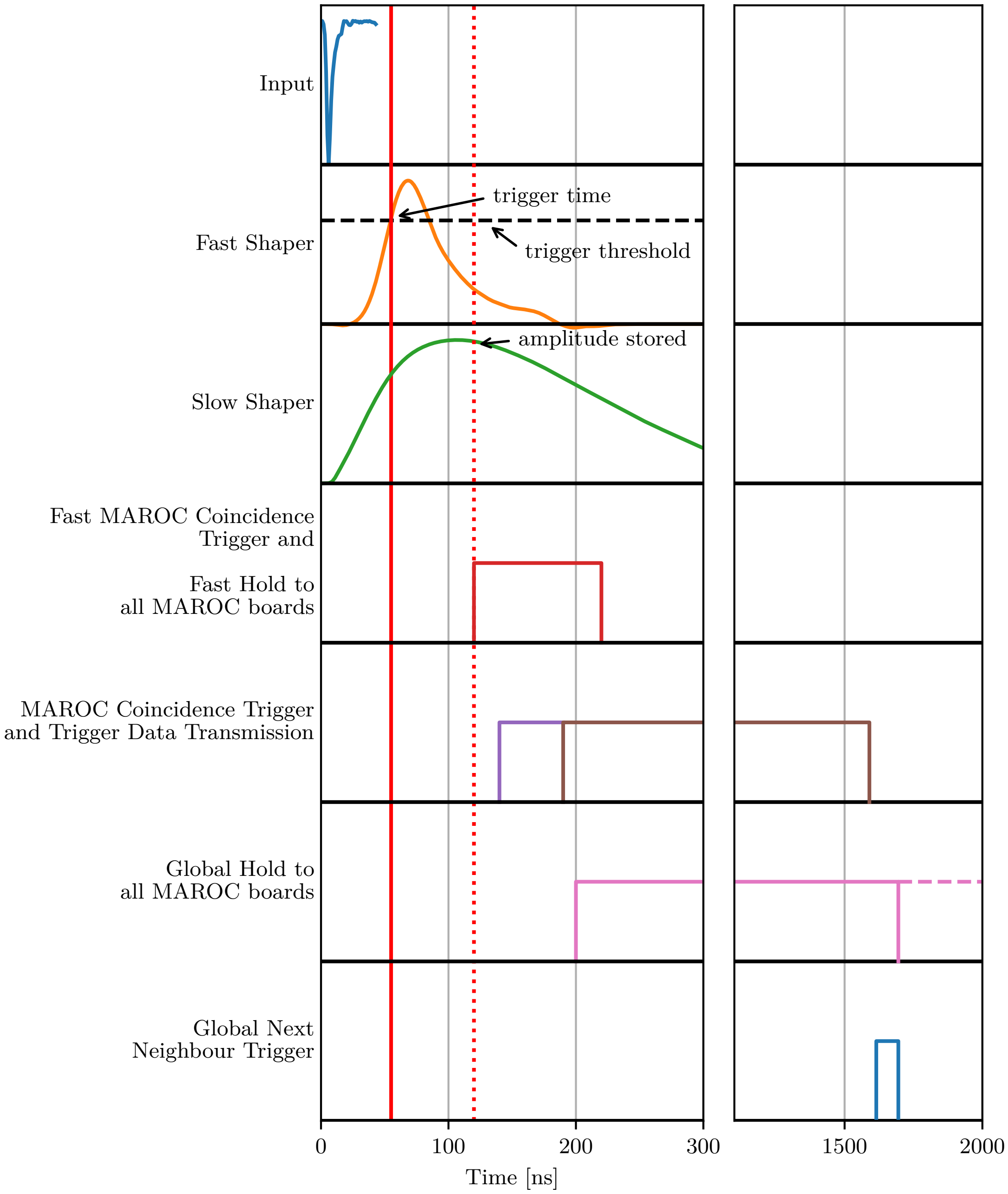
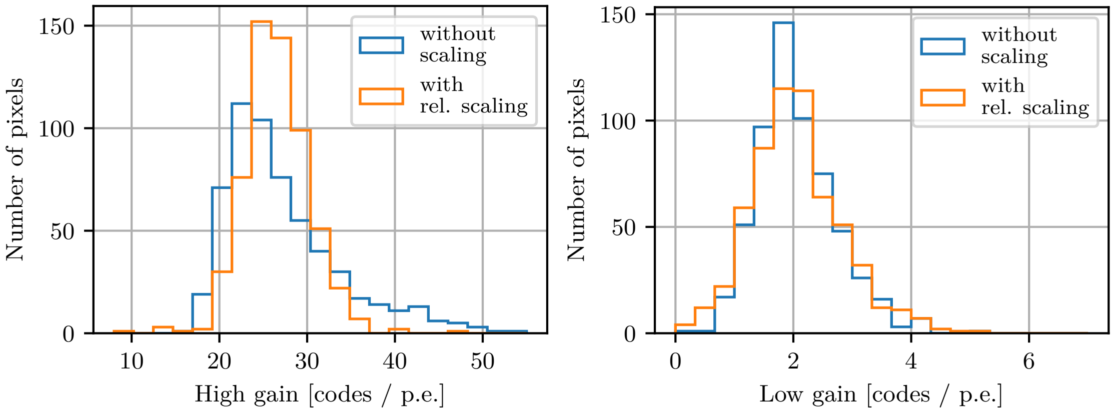

Input
Fast Shaper
trigger time
trigger threshold
amplitude stored
Slow Shaper
Fast MAROC Coincidence
Trigger and
Fast Hold to
all MAROC boards
MAROC Coincidence Trigger
and Trigger Data Transmission
Global Hold to
all MAROC boards
Global Next
Neighbour Trigger
0
100
200
300
1500
2000
Time [ns]

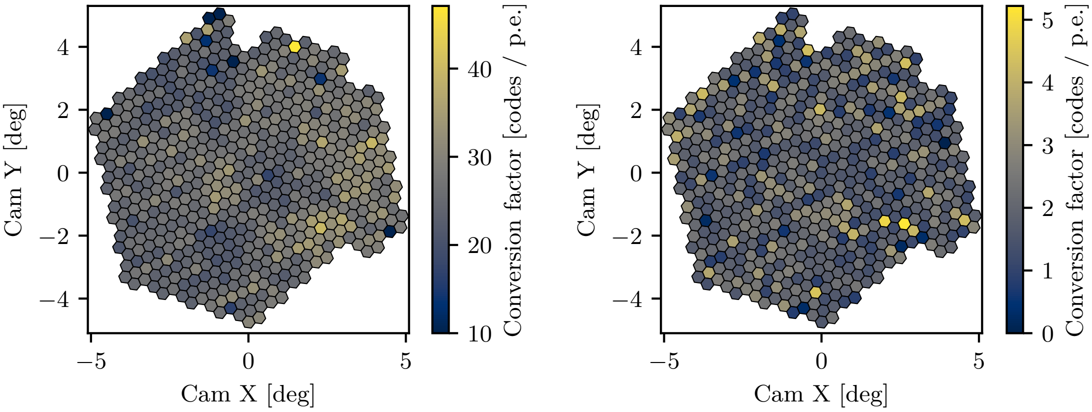
Figure 2.26: Trigger and readout scheme of the second TAIGA IACT camera. When the trigger threshold on the Fast Shaper is reached, the discriminator fires (red vertical line). If at least two pixels are above threshold (fast coincidence check), the MAROC board issues a hold signal after the hold time t_h to all MAROC boards, storing the amplitudes of the Slow Shapers in a Track and Hold circuit (red dotted vertical line). When two PMTs trigger within 15 ns, the trigger information is sent to the Central Controller, where a global hold signal is sent to all MAROC boards of the camera. Then, the topology trigger condition is checked (next neighbours) and if not met, the hold signal is reset. On success, the global hold signal is kept and the readout command is sent. Then, all MAROC boards will start the ADC and transmit the amplitude information. This scheme is an adaption of figures from Kompaniets [55] and *MAROC 3A Datasheet* [56].

The determination is done by a linear fit to the range between a high gain of 1200 and 2500 codes. As shown by the orange histogram in Figure 2.28, the `second_hottest` distribution indeed matches well when using the factors determined by the proportion between low and high gain.

In the run used for the distribution in Figure 2.28, cleaning levels of $l_1 = 21$ p.e. and $l_2 = 10.5$ p.e. are used. This distribution is not suitable for the trigger threshold estimation, as a three or more pixel coincidence trigger was used instead of the two or more pixels setting used for Figure 2.24.

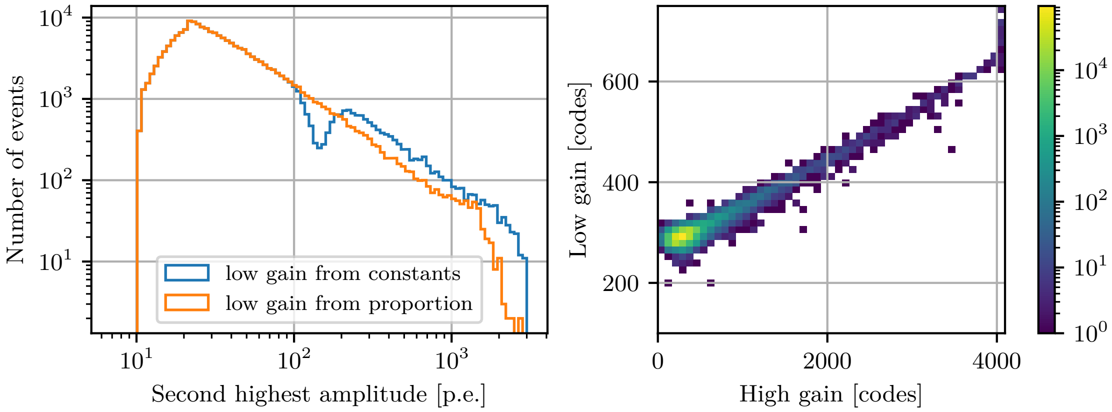


(a) High Gain Histogram (b) Low Gain Histogram



(c) High Gain Cam (d) Low Gain Cam

Figure 2.27: Histograms and per pixel plots of the calibration constants of IACT 2. The values were obtained by the same author with the same method as described in Lubsandorzhiev et al. [60].



(a) Second brightest pixel distribution

(b) Low Gain against High Gain

Figure 2.28: (a) Distribution of the second highest amplitudes for one run of the second camera. In blue, the provided calibration constants (see text) are used for the high and the low gain. The clear dip starting at ∼100 p.e. is caused by a mismatch of the high and low gain calibration factors. To mitigate this, the low gain calibration is estimated from the high gain calibration factors via the linear relation evident when plotting (b) the low against the high gain amplitudes in raw ADC codes.

2.5 Hybrid Event Finding

Events with both, IACT and HiSCORE triggering are the aim of the hybrid analysis with the goal of implementing a method with high performance and the possibility to instrument very large areas at the same time (see Section 1.4). As each station and the IACTs have independent trigger systems and as the array has a much larger FoV and energy threshold, most of the events in both event lists are not coincident events. Therefore, the coincident events need to be found. As the absolute accuracy of the event times is smaller than the dead times of 182 µs for the IACTs and 378 µs for the HiSCORE stations, the following scheme can be used to find the coincident events: For each event in one event list, calculate the time differences Δt_- and Δt_+ to the closest previous and next events of the other event list. Then, the median time of all events with $\Delta t < 10\,\mu\text{s}$ is determined. All events within a window of $\pm 1.5\,\mu\text{s}$ around this median value are considered coincident events. An example histogram of Δ_t for coincident events of both IACTs on 2020-01-27 is shown in Figure 2.29. The resulting distribution depends on the arrival times of the light fronts. As the IACTs were pointing to azimuth angles between 100° and 240° and zenith distances between 30° and 15°, a wide distribution is expected. For events with an azimuth angle along the connection line of the telescopes and with a zenith distance of 30°, the expected time delay between the two IACTs for a planar light front is ∼600 ns. The coincident events are used in Chapter 6.

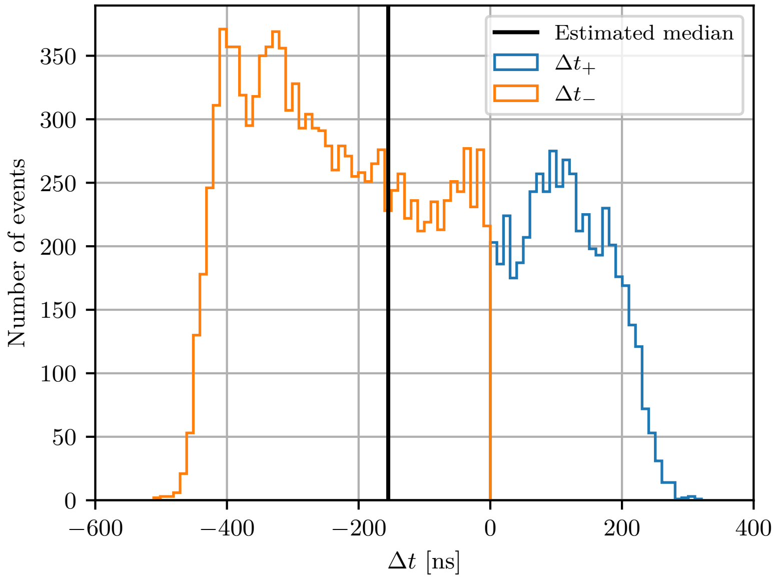


Figure 2.29: Distributions of the relative arrival times $\Delta t = T_2 - T_1$ between events recorded by IACT 1 at T_1 and by IACT 2 at T_2. The events survived the respective image cleaning and were recorded on 2020-01-27.

Chapter 3

Simulation

For ground-based gamma and cosmic-ray detectors, the performance and all response functions need to be simulated, as no calibration source with controlled emission exists yet. For this task, CORSIKA[74] is used. It utilizes Monte Carlo (MC) methods to track the primary and EAS particles through the atmosphere. Section 3.1 describes the CORSIKA configuration used and introduces the simulated datasets. Details of the simulation of IACT 1 are presented in Section 3.2 and a brief summary of the HiSCORE simulation is given in Section 3.3.

The code used for the configuration of CORSIKA and for the simulation of the IACTs and HiSCORE can be found on GitLab[1].

3.1 Monte Carlo Dataset

For the simulation of EAS, CORSIKA in v7.64 [74] with the IACT/ATMO package v1.50 [75] is used. For low- and high-energy hadronic interactions, UrQMD[76, 77] and QGSJET[78] are used, respectively. The viewcone option is enabled and is used to simulate diffuse and isotropic emission in a cone with an outer opening radius of Ω around the position specified by the zenith distance θ from the south ($\phi = 0^\circ$ in CORSIKA). The volume detector option is enabled and the longitudinal shower development is sampled in vertical steps of $20\,\mathrm{g/cm^2}$ to allow estimates of the height of the shower maximum X_{max}. The Cherenkov photons are tracked in bunches with up to 10 photons and stored in binary EventIO format.

---

[1] `https://gitlab.rrz.uni-hamburg.de/TAIGA/Simulation`

particle	E_l [TeV]	E_h [TeV]	θ [°]	Ω [°]	X_s [m]	Y_s [m]	N_{runs}	c	N_{evt} [10^4]
γ	0.5	5	16	10	1096	1159	500	3	1500
γ	5	50	16	10	”	”	500	3	150
γ	50	500	16	10	”	”	500	3	15
γ	500	5000	16	10	1296	1359	1000	3	3
γ	0.5	5	26	10	1096	1159	500	3	1500
γ	5	50	26	10	”	”	500	3	150
γ	50	500	26	10	”	”	500	3	15
γ	500	5000	26	10	1296	1359	1000	3	3
γ	0.5	5	36	10	1096	1159	1000	1	1000
γ	5	50	36	10	”	”	1000	1	100
γ	50	500	36	10	”	”	1000	1	10
γ	500	5000	36	10	1296	1359	1000	3	3
γ	0.5	5	46	10	1096	1159	500	3	1500
γ	5	50	46	10	”	”	500	3	150
γ	50	500	46	10	”	”	500	3	15
γ	500	5000	46	10	1296	1359	1000	3	3
H	0.6	6.8	25	35	1096	692	81992	1	9038.7
H	6.8	68.2	25	35	”	”	57000	1	900
H	68.2	682	25	35	”	”	1000	1	10
He	0.7	7.4	25	35	”	”	73997	1	3917
He	7.4	74.3	25	35	”	”	13000	1	340
He	74.3	743	25	35	”	”	13998	1	5.6
N	0.8	8.3	25	35	”	”	1000	1	1000
N	8.3	83	25	35	”	”	1000	1	100
Fe	0.9	9.6	25	35	”	”	999	1	999
Fe	9.6	96	25	35	”	”	1000	1	100

Table 3.1: The simulation sets. Each set was simulated with a power law following $\propto E^{-1}$ between a lower energy E_l and a higher energy E_h. The origin was spread evenly in a circle with the diameter Ω around the nominal zenith distance θ. On the ground, the core locations are scattered between $\pm X_s$ in the X direction and $\pm Y_s$ in the Y direction. N_{runs} is the total number of simulated runs in the set and N_{evt} is the total number of events. If a shower was reused c times by randomly scattering it over the allowed core area, it was also counted c times.

A magnetic field of 18.5685 µT in the x direction (North) and 57.3985 µT in the z direction (downwards) is used. The component of ~1.1 µT towards West is neglected. The energy is simulated with an index of -1 ($\propto E^{-1}$), giving a flat histogram with logarithmic binning. The observation level is set to 675 m.

Table 3.1 lists the different datasets produced. For the background set, protons, helium, nitrogen and iron are used with a viewcone of $\Omega = 35°$ around the HiSCORE pointing with a zenith distance of $\theta = 25°$. Because of computing power and disk space constraints, higher energies were only simulated for proton and helium. The sets of gamma particles were simulated with $\Omega = 10°$ for θ of 16° 26° 36° and 46°. With the camera's FoV of about 9.6°, this gives a useful range for telescope zenith distances of 11° to 51°.

3.1.1 Simulated Array

The simulated array consists of 143 HiSCORE stations and 8 IACTs. It is shown and compared to the most recent existing array in Figure 3.1. Some of the recently installed stations are located up to 2.5 m away from the anticipated positions when the simulation configuration was created. Additionally, three stations on the east end and five stations on the north are installed outside of the boundary of the simulated array (cluster 4). IACT number 3 and 4 are a bit offset from the real position and for the position where IACT 5 is going to be installed, no IACT was simulated. This will be addressed with an updated simulation configuration in the future. For this work, the array matches well enough, as the differences of the simulated positions for the stations close to IACT 1 and 2 are small and cluster 4 was not yet in operation. For the 8 IACTs, the radius r_s of the sphere which Cherenkov photons need to pass to be stored in the EventIO file is set to $r_s = 3.3\,\mathrm{m}$. For the HiSCORE stations, it is set to $r_s = 1\,\mathrm{m}$.

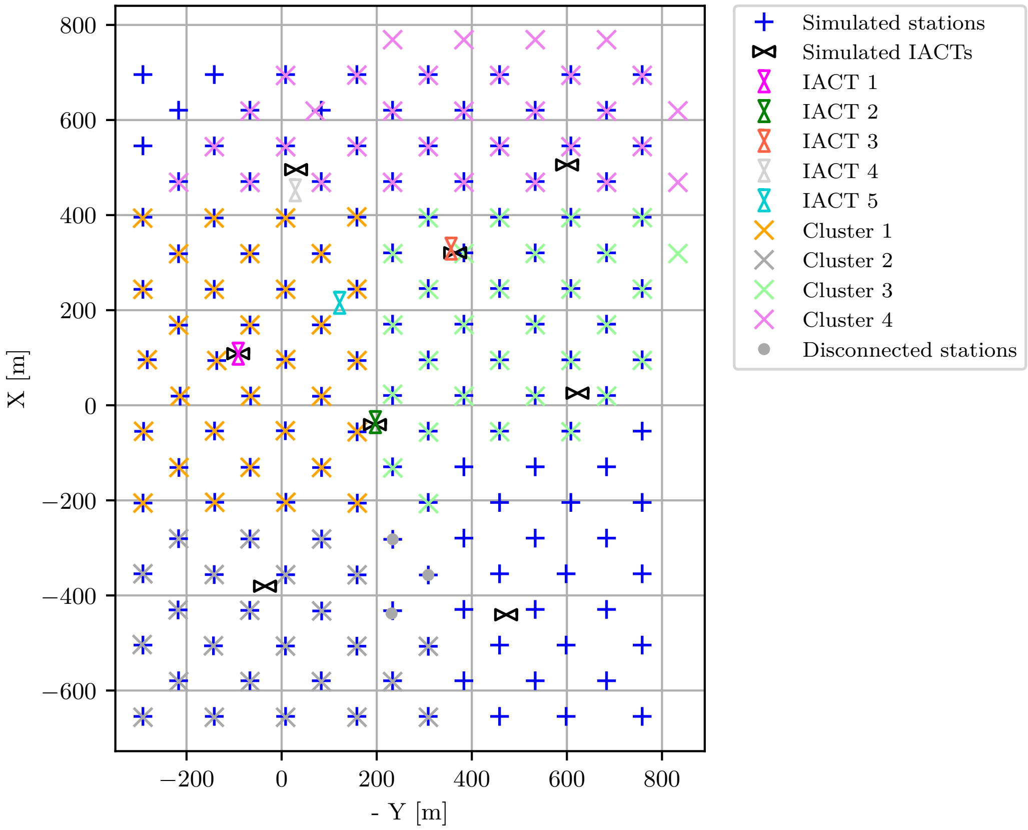


Figure 3.1: Simulated and existing array configuration.

3.1.2 Atmosphere

The development of the air shower, the generation of Cherenkov light, and the consequent absorption depend on the profile of the atmosphere above the detector. In CORSIKA, the default option of the US standard atmosphere is used.

According to Bernlöhr [79], the expected variation of the Cherenkov light density in a radius of ∼120 m around the core position is on the order of 10 % to 20 % for mid-latitudes.

As no direct measurements of the atmosphere above the Tunka valley are available, the gdastool (see [80]), a python script coming with CORSIKA was used to obtain the atmosphere profiles of select days. This tool queries the Global Data Assimilation System (GDAS)[2] and retrieves the model profiles at the closest grid point (51.8°N, 103.1°E) for altitudes from sea level to ∼25 km. It is not known how accurate it is at the Tunka site, but it was found that the accuracy at the Pierre Auger site is adequate.[81] As the tool returns the refractive index n for radio waves, it was adapted to calculate the refractive index at 350 nm for this work by using the equations derived by Ciddor [82] and implemented in the python package `ref_index`[3].

Figure 3.2 shows a comparison of the air density ρ and of $n-1$ compared to the standard atmosphere of the USA (atmprof6 in the IACT package) for profiles of three days and for the mid-latitude winter and subarctic winter profiles provided with IACT/ATMO (atmprof3 and atmprof5).

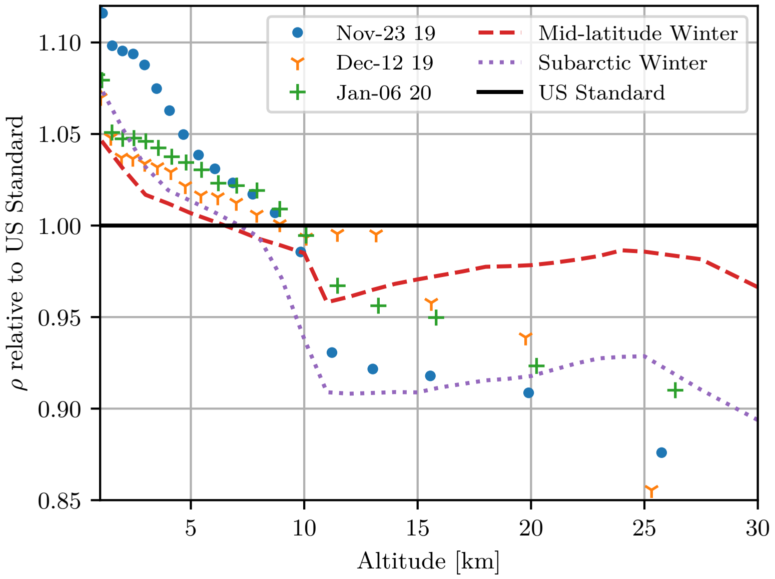
Bernlöhr [79] showed, that the influence caused only by the refractive index is typically smaller than the differences arising from absorption and density. From Figure 3.2, it is apparent that the density gradient of the standard atmosphere used in the simulation is smaller. The crossing point, where all curves match best is around ∼10 km altitude. This is close to the typical height of the shower maximum X_{max}. As the density above that altitude appears to be lower for the three days, the mass overburden at that height is likely smaller, systematically shifting X_{max} closer to the ground.

For altitudes below ∼15 km, the mid-latitude winter profile seems to fit best to the three exemplary days obtained from GDAS, even though a difference of a few percent, increasing to low altitudes is apparent. As the shower maxima are typically below this altitude, the mid-latitude winter profile is chosen for atmospheric absorption (see Figure 3.3).

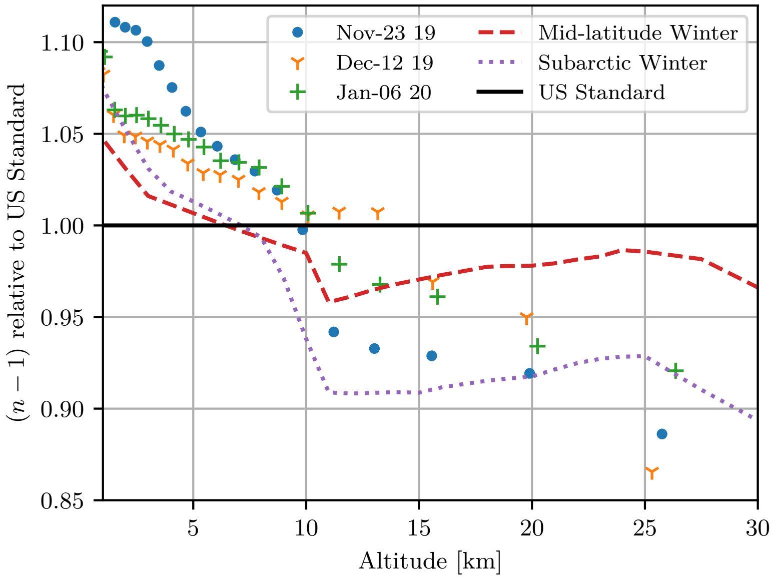
---

[2] `https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-data-assimilation-system-gdas`

[3] On PyPI, `https://pypi.org/project/ref_index/#description`



(a) Density



(b) Refractive index

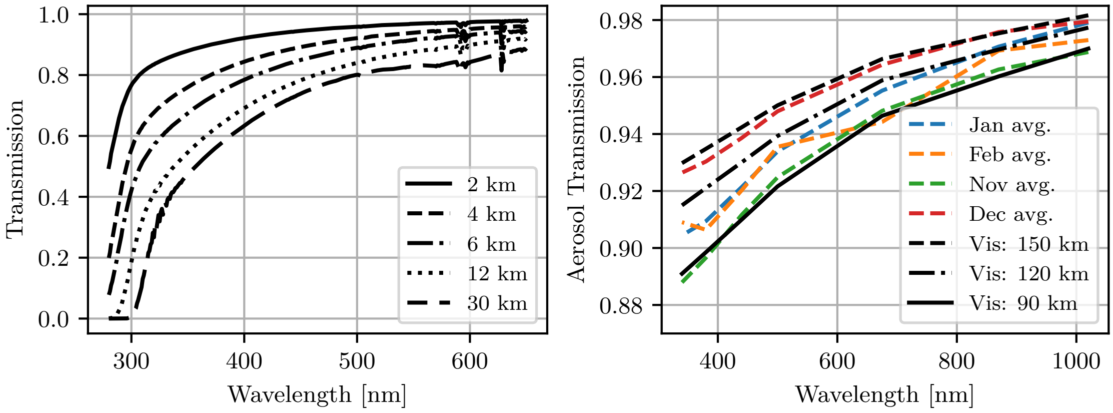
Figure 3.2: Comparison of three exemplary profiles obtained from the Global Data Assimilation System with the US standard, mid-latitude winter and subarctic winter atmospheric profiles provided with the IACT/Atmo package. (a) Density relative to the US standard atmosphere as a function of the altitude above sea level. (b) $n-1$ relative to the US standard atmosphere, where n is the optical refractive index, as a function of the altitude above sea level.

The absorption is provided as a table of optical depth $\tau(h, \lambda)$ values for vertical heights h and photon wavelengths λ. τ is connected to the transmission t obtained at a zenith distance θ through

$$t = e^{\frac{-\tau(h,\lambda)}{\cos\theta}}. \tag{3.1}$$

Another important factor influencing the atmospheric absorption are the aerosols in the atmosphere. The total optical depth from free space to the ground $\tau_{tot} = \tau_{rem} + \tau_a$ is a combination of the contribution by the Aerosol Optical Depth (AOD) τ_a and by other constituents τ_{rem}.

Monthly averages of estimates of the Aerosol Optical Depth (AOD) τ_a for specific wavelengths are available from solar observations at the Tunka site as presented in Dementeva et al. [45] (in this article, the Tunka site is referred to as Tory site[4]). MODTRAN3 v1.5 [83] is used to create the atmospheric absorption table. The mid-latitude winter profile is used with the rural haze option set to a visibility of 150 km, reflecting the average τ_a of the Tunka site for December. Figure 3.3b shows the atmospheric transmission corresponding to the average τ_a for November, December, January and February and the simulated total aerosol contributions for the visibility settings of 90 km, 120 km and 150 km. For the simulation, the transmission is estimated by subtracting the τ obtained without the aerosol model from the value with the aerosol model activated. Even though the aerosol component from ground to free space seems to be sufficiently described, some uncertainty remains as no information about the height profile of the aerosol distribution is available.



(a) Total transmission

(b) Aerosol transmission

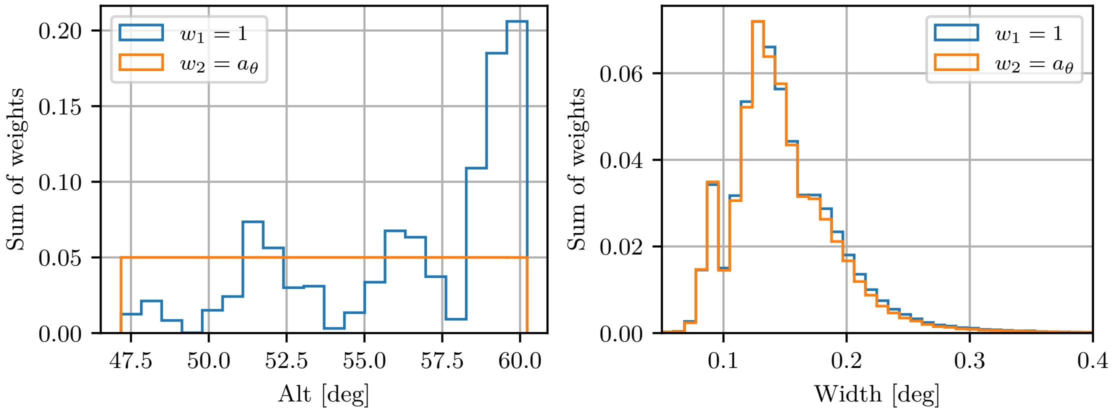
Figure 3.3: (a) Atmospheric transmission to 675 m at a zenith angle of $\theta = 30°$. (b) Average optical transmission from free space to the Tunka site ($\theta = 0°$) as found in [45] and compared to transmission simulated with MODTRAN3 [83] for a midlatitude winter atmosphere and rural haze set to different visibilities.

3.1.3 Sample Weighting

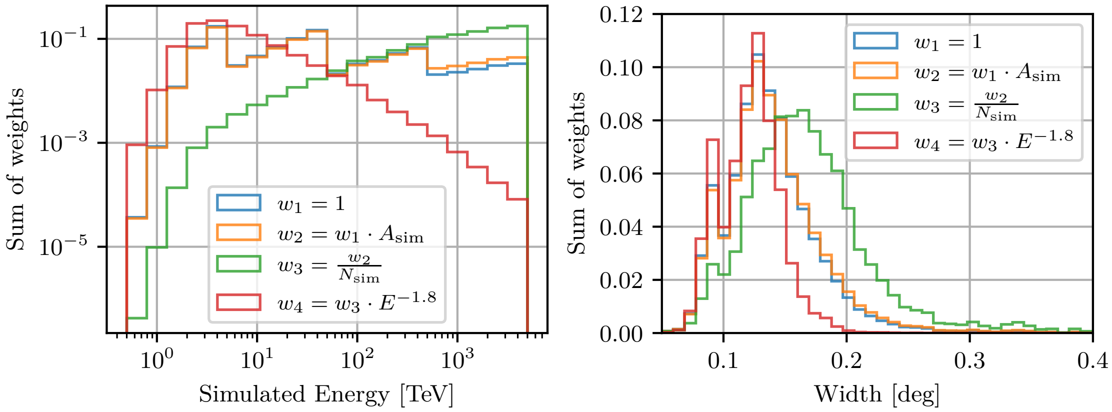
When using the simulated sets to obtain, for example, histograms of some resulting parameters, care must be taken as to how the simulated events are distributed. The most obvious example is the number of events per simulated energy range. Because the number of events in each simulation set of one decade in energy is chosen such that a similar computing time is required, the number of simulated events goes down by a factor of 10 with every decade in energy and the resulting distribution in energy has jumps (see Figure 3.4c).

---

[4] This station is part of the Aerosol Robotic Network AERONET as Irkutsk station (`https://aeronet.gsfc.nasa.gov/`)



(a) Data (b) Data



(c) Gamma (d) Gamma

Figure 3.4: Illustration of the sample weighting. For data, the altitude distribution is shown in (a) and the `width` is shown in (b). Here, the weights a_θ produce a flat distribution in altitude. For gamma simulation data, the weights take into account: The number of events simulated for each energy range, the simulated area and the spectral shape. The histogram over the true energy is shown in (c) and the `width` is shown in d).

The easiest way to use the simulation sets with realistic distributions would be to select a subset of the events with the desired composition / distribution. As this results in a loss of statistics, especially at higher energies, weights w are assigned to each event, so that the desired distributions are achieved. Histograms are then created by the sum W of the weights w_i of the events i falling into the bin (instead of counting the number of events in the bin)

$$W = \sum_i w_i, \tag{3.2}$$

and when an estimate of an error σ_W is required, it is calculated as

$$\sigma_W = \sqrt{\sum_i w_i^2}. \tag{3.3}$$

The weights are defined as

$$w_i = \frac{f(E_i) A_{\mathrm{sim},i}}{N_{\mathrm{sim},i} a_{\theta,i}}, \tag{3.4}$$

where f is a flux term depending on the particle energy E, resulting in a specific primary energy spectrum (e.g. a power law with photon index Γ: $\propto E^{-\Gamma}$). Then, for every range of simulated energies, the total number of events in that range N_{sim} is used to obtain a continuous spectrum without jumps at the edges of the energy ranges. Additionally, for each set, the scattering area $A_{\mathrm{sim}} = 2Y_s \cdot 2X_s$ used in CORSIKA is used to compensate the different event densities on the ground.

The zenith distance distributions of the gamma sample is different to the distribution found in data. For this reason, a scaling factor a_θ is introduced. It is calculated from a histogram in telescope pointing zenith distance with small enough bins j, such that the resulting zenith distance distribution is mostly flat ($a_\theta = \frac{\sum_j N_j}{N_j}$). When not required, it is set to $a_\theta = 1$.

For helium, iron and nitrogen, $f(E)$ is determined according to the corresponding flux predicted by the polygonato model [84]. For proton and helium, recent results by DAMPE are used [85, 86]. Helium is used to represent the atomic numbers $A = 2 - 5$, nitrogen for $A = 6 - 21$, and iron for $A = 22 - 70$.

For gamma, f is typically set close to $E^{-1.8}$ to give a Crab-like differential power law spectrum with the resulting exponent of -2.8. Because of the higher energy and the expected steepening, a slightly softer value is selected as compared to -2.6 used in other TeV experiments.

Typically, the weights are normalized such that the total sum of the weights is one. Figure 3.4c shows the effect of the weights for real data and for a gamma simulation sample.

3.2 Simulation of the First Imaging Air Cherenkov Telescope

For the IACT simulation, an adapted version of `sim_telarray`[75] is used. The readout and conversion to the HDF format (also used for real data) is then done with `read_taiga`, a C/C++ programme based on an example programme provided with `sim_telarray`. `sim_telarray` was developed for the use with HEGRA and HESS and is also used for CTA. It provides extensive configuration possibilities and required only little changes to the code.

The photon bunches created in CORSIKA are traced through the atmosphere, reflected on the mirrors (Section 3.2.1) and then traced to the camera. There, it is checked if a conversion from photon to photoelectron occurs (Section 3.2.2). The electronic response of the Fast Shaper (discriminator in `sim_telarray`) and of the Slow Shapers (Flash Analog To Digital Converter (FADC) in `sim_telarray`) are simulated accordingly (Section 3.2.2.2).

The following list describes the simulation steps and the approximate probability of a photon to survive the step:

- Atmospheric absorption ~75 % at 30 deg and (Section 3.1.2)
- Shadowing by camera and support structure ~85 % (Section 3.2.1)
- Mirror reflection and alignment ~83 % (Section 3.2.1)
- Plexiglass protection (Camera filter) ~84 %(Section 3.2.1)
- Winston cones ~80 % ((Section 3.2.2.1)
- Dead space between Winston cones ~73 % ((Section 3.2.2.1)
- Quantum efficiency ~20 % (Section 3.2.2.2)
- Single photoelectron response (Section 3.2.2.2).

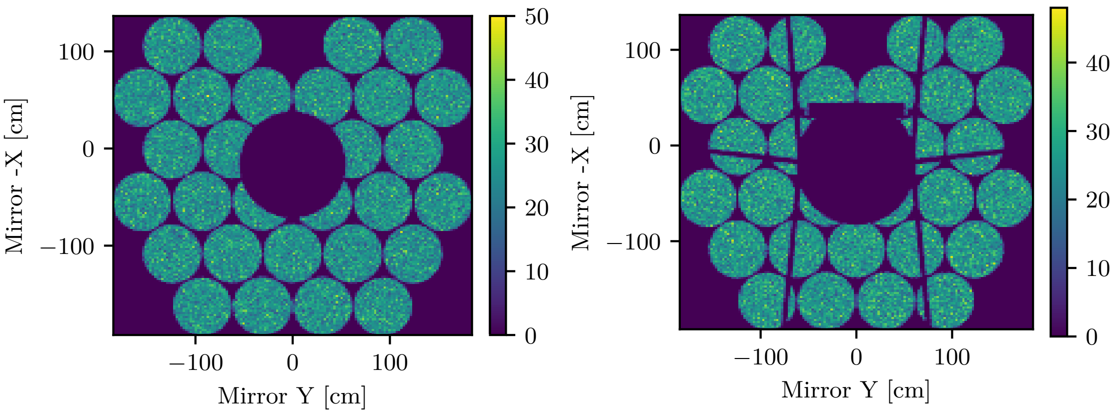
In total, only about ~7 % of the light reaching the telescope is detected.

3.2.1 Mirrors

After atmospheric absorption (Section 3.1.2) and on the path towards the mirror, each photon bunch might hit the camera housing or the support structure of the camera. The camera housing is modelled with a diameter of 110 cm and is included in every run. The shadowing by the support structure is computationally expensive and is therefore approximated with the constant transmission probability of $t = 0.90$. This factor is determined with a `sim_telarray` compile time option[5]. Figure 3.5 shows the intersection points of photons with the mirror plane for the cases with and without shadowing of the support structure. The rays are emitted from a point source at infinity at a zenith distance of $\theta = 28°$ and the telescope is pointing to $\theta = 30°$.

---

[5] $-DRAYTRACING\_INTERSECT\_RODS$



(a) Only Camera Shadowing

(b) Including Support Structure

Figure 3.5: Photon hit points in the mirror plane. Hypothetical point source at infinity (zenith distance $\theta = 28°$ and telescope pointing to $\theta = 30°$.

The mirrors are arranged in a Davies-Cotton configuration as described in Section 2.3 and modelled with a constant reflectivity of 0.9 [64]. The nominal focal length is set to $f = 4.75\,\mathrm{m}$. However, in this simulation, only the central mirror has a distance f from the focal point. The outer mirrors are then arranged on a radius of 4.32 m. Each mirror is modeled to have a random reflection angle offset following a normal distribution around zero with a width of $\sigma_{refl} = 0.0066°$.

Additionally, several parameters are sampled once for every run of `sim_telarray`. Each facet is allowed to have a normally distributed distance to the focal point with $\sigma = 2\,\mathrm{cm}$ around 4.75 m. The variations of the focal length for each tile are set to 0. Furthermore, the misalignment of the mirrors is modeled with a zenith distance dependent normal distribution. It is split into a horizontal misalignment with $\sigma_{horizontal} = \sqrt{(0.0035°)^2 + (0.023°\,(\cos\theta - \cos 88°))^2}$ and a vertical misalignment with $\sigma_{vertical} = \sqrt{(0.0034°)^2 + (0.01°\,(\cos\theta - \cos 88°))^2}$. This is used to describe the worsening of the PSF when deviating from the estimated mirror alignment position at $\theta = 88°$. The last parameter influencing the PSF is the distance of the entrance plane of the Winston cones to the nominal plane, the focus distance d. This value is set to $d = 2.8\,\mathrm{cm}$. Figure 3.6 shows the resulting PSF for stars at infinity for different off axis angles. The PSF is quantified with the 68 %-containment radius r_{68}. For larger offset values, the PSF gets worse and asymmetric, as also shown in Kunnas [36]. The difference of the PSF in positive and negative zenith distance direction is caused by the missing outermost mirrors at the top. Without the contribution of the mirrors the farthest away, the PSF is much better confined.

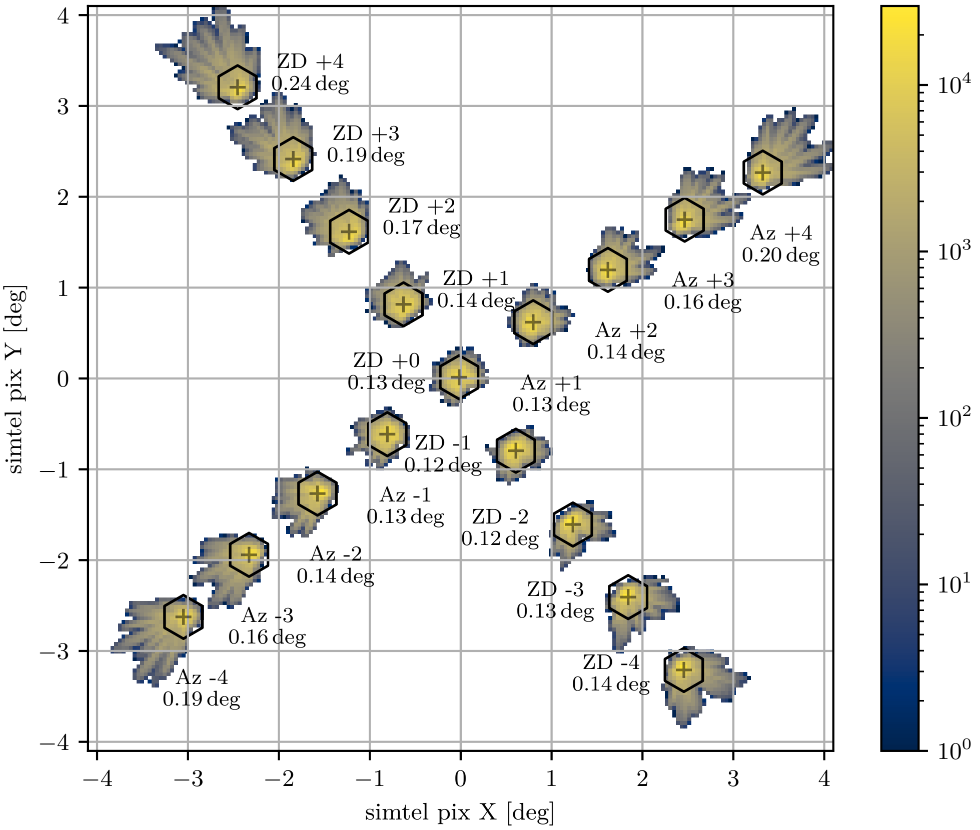


Figure 3.6: Containment radii r_{68} for various telescope pointings relative to a star at a zenith distance of 30°.

A value of $r_{68} = \sim 0.13°$ corresponds to ~1.1 cm which corresponds to a standard deviation of ~0.73 cm when assuming a 2D normal distribution. Comparing this value to the ones in Kunnas [36], this corresponds to a misalignment value of $\sigma = 0.045°$.

The configuration with $d = 2.8\,\text{cm}$ was chosen because it was close to the default setting and it describes the resulting distributions well (see Section 3.2.6). In the experiment $d = (0 \pm 1)\,\text{cm}$ and a corresponding focus at infinity is used. (See Section 2.3). According to the thin lens equation ($\frac{1}{G} + \frac{1}{B} = \frac{1}{f}$, where G is the distance from the object and B the distance from the image), a focus distance $d = 2.8\,\text{cm}$ corresponds to an approximate focus point at a distance of $B = \left(\frac{1}{f} - \frac{1}{f+d}\right)^{-1} \approx 0.8\,\text{km}$. Figure 3.7 shows r_{68} for point sources against the distance to the telescope. The change between 12 km and 4 km is about 10 %. From there, the PSF drops significantly towards closer distances. However, using $d = 0$ no close match of the image parameter distributions was achieved. It is conceivable, that the actual distance between Winston cones and mirrors is larger, because the cones were mounted in the nominal focus point of the construction but the actual distance of the mirrors to the Winston cone plane could be larger. The final position of the mirrors depends on the actual thickness of the mirrors and on the adjustment of the adjustable mirror mounts.

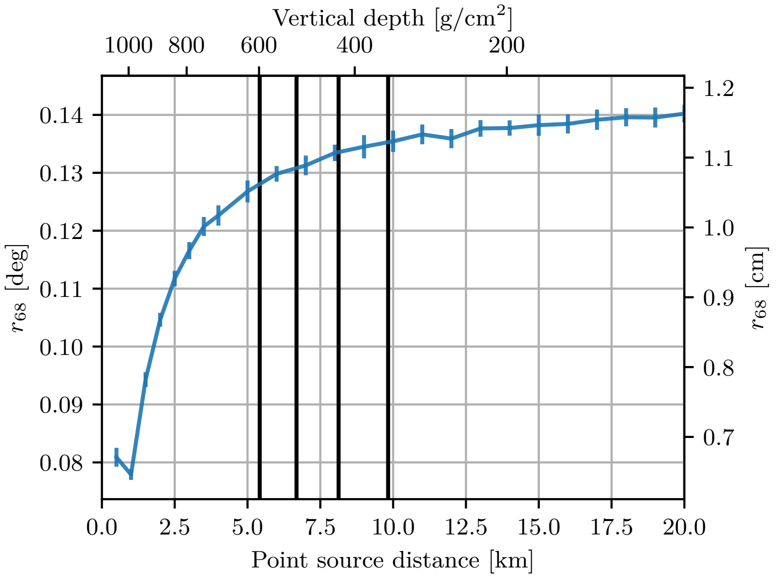


Figure 3.7: Containment radius versus the distance of a point source for a camera entrance positioned $d = 2.8\,\text{cm}$ behind the focal plane. The black lines indicate the expected height of the shower maximum for gammas with an energy of 1000 TeV, 100 TeV, 10 TeV and 1 TeV from left to right.

In the future, the behaviour of the PSF across zenith distances and for different off-axis angles should be either measured in more detail or, if possible, inferred from available data. In principle, this could be achieved by an analysis of the smearing of showers with constrained characteristics, such as muon rings[e.g. 87]. Possibly, other showers at selected core positions and with a well reconstructed shower maximum could be used.

Another factor is the ray offset caused by the refraction of the light when entering and exiting the plexiglass protection window. For every mirror, the angle under which the window is hit is systematically different. For the outer mirrors and the corresponding angle to the plexiglass of ∼25°, the expected shift at the exit of the plexiglass is approximately 0.4 cm when assuming an index of refraction of $n = 1.5$ and a thickness of 1.5 cm. This value is small enough as compared to the expected spot size to be neglected. Furthermore, it can be compensated to some extent during the mirror alignment procedure. The plexiglass is modelled with the wavelength-dependent transmission probability shown in Figure 2.9 (see Section 2.3.2).

3.2.2 Camera and PMTs

After reflection off the mirror and passing through the plexiglass, the light hits the entrance opening of the camera. The camera is modelled according to the description in Section 2.3.1, with each of the 560 PMTs having a hexagonal Winston cone with a flat-to-flat width of 2.566 cm at a pixel spacing of 3 cm. The camera is rotated by −37.5°, corresponding to 52.5° rotation in the `sim_telarray` coordinate system.

3.2.2.1 Winston Cones

When a photon bunch hits the sensitive area of the Winston cone, the average angular acceptance as shown in Figure 2.8 is used with a scaling factor t_s, disregarding the position dependence of the overall and angular acceptances.

$t_s = 0.81$ is the approximate transmission probability for photons uniformly distributed over the full Winston cone opening. It is estimated from the Winston cone efficiency of $e_{cone} = \frac{I_{cone}}{I_{naked}} \approx 2.6$ (Measured by Nima Lubsandorzhiev, see also Section 2.3.1), where I_{cone} is the current with the cone attached and I_{naked} is the current without the cone. I_{cone} must be proportional to $t_s A_{entrance}$ where $A_{entrance} = 2\sqrt{3}d^2$ is the area corresponding to the flat-to-flat width d of the entrance opening. I_{naked} should be directly proportional to the active photocathode area A_{cat}, which is assumed to be a circle with a diameter of 15 mm. Therefore, the estimate for t_s is

$$t_s \approx e_{cone} \frac{A_{cat}}{A_{entrance}} \approx 0.806. \tag{3.5}$$

3.2.2.2 Quantum Efficiency and Single Photoelectron Response

When passing the Winston cone, the light hits the PMT, where its probability of conversion to a photoelectron is described with wavelength-dependent quantum efficiency as shown in Figure 2.9. If a photoelectron is created, the actual amplitude in the electronic response is derived according to an average probability distribution, the single photoelectron (SPE) response distribution. It is derived from the measurement shown in Figure 2.9. It includes several effects, such as the incomplete photoelectron collection. There are two cases: The prompt response and the background response including afterpulsing. For the prompt response to a photoelectron created by a signal photon, the afterpulsing is not relevant, as it typically occurs only several hundred ns later. However, for the background distribution, it is required as it can lead to responses with an amplitude corresponding to >10 p.e.. The probability for a given amplitude response in photo equivalents (p.e.) is approximated with the following function:

$$p(A) = \begin{cases} 0.5 & \text{for } A < 0.775\,\text{p.e.}, \\ 0.85 e^{-\left(\frac{A-(1.125\,\text{p.e.})}{2(0.35\,\text{p.e.})}\right)^2}, & \text{for } 0.775\,\text{p.e.} \leq A < 1.625\,\text{p.e.} \\ 46 e^{-\frac{A-(0.125\,\text{p.e.})}{0.3\,\text{p.e.}}} + p_{afterpulsing}, & \text{for } A \geq 1.625\,\text{p.e.} \end{cases} \tag{3.6}$$

With $p_{afterpulsing} = 0$ for the prompt and $p_{afterpulsing} = 3 \times 10^{-5} \cdot e^{-\frac{A-(0.125\,\mathrm{p.e.})}{5.5\,\mathrm{p.e.}}}$ for the background case. It is scaled so that on average 1 p.e. is returned. This normalisation is consistent with the conversion factors of the experiment determined with the F-factor method (see Section 2.3.5).

3.2.3 Signal

When a photoelectron is created and its amplitude A in photo equivalents p.e. is determined according to the SPE distribution, the electronic signals are updated. This is done by adding the signal shapes (see Section 2.3.3 and Figure 2.13) with an amplitude of $A \cdot G$ to the sampled signals at the photoelectron arrival time t, where G is the gain of the corresponding channel. For each pixel, there are three channels: The high gain and low gain (Slow Shaper) for the readout and the discriminator (Fast Shaper) for the trigger. The high and low gain channels are simulated in steps of 1 ns in a window of 128 ns. For the discriminator, steps of 0.5 ns are used. The number of bins required was determined by checking if the hold times occur within the simulated window.

The starting point t of where to add the pulse shapes with the amplitude of $Ab \cdot G$ to the corresponding signal of the channels is determined by the arrival time of the photoelectron with an additional random jitter of $\sigma = 1$ ns.

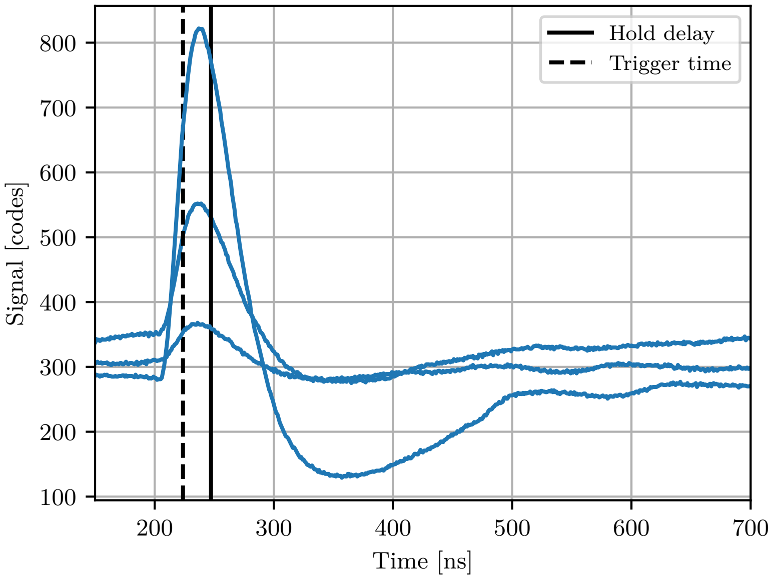
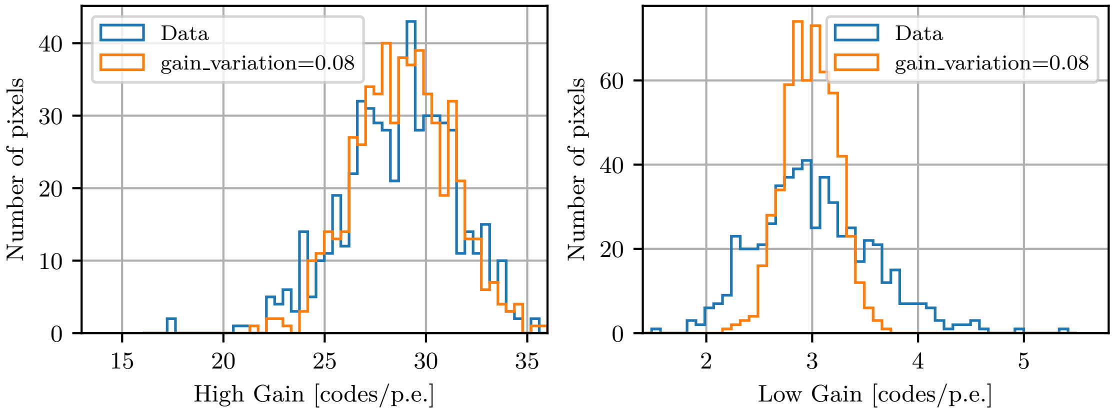


Figure 3.8: Simulated low gain signal of three pixels of a telescope positioned at a distance of 100 m from the impact point of a 10 TeV gamma shower with a zenith angle of 25°. The trigger time is indicated by the dashed line and the solid line marks the time at which the amplitudes are buffered for a readout.

An example of three simulated low-gain signal slices is shown in Figure 3.8. The three pixels have more than >20 p.e. and the telescope was positioned 100 m from the impact point of a 10 TeV gamma shower with a zenith angle of 25°. A window of 1024 ns instead of 128 ns was used to show the whole signal.

The gain values used in the simulation are specified by normal distributions. For the high gain, $\mu = 29\,\text{codes}\,\text{p.e.}^{-1}$ and for the low gain $\mu = 3\,\text{codes}\,\text{p.e.}^{-1}$ and the channel to channel variation is specified by a relative fluctuation of $\sigma = 0.08$. Figure 3.9 shows histograms of the low and high gain values in the experiment and of one realisation in the simulation. For the high gain, the match seems appropriate, while for the low gain the actual fluctuation is not reflected properly. Variations in the relative gains of the high- and low-gain channels are caused by the dynamic range of the MAROC amplifier settings (see Section 2.3.5). In more recent versions of `sim_telarray`, the *HG_LG_VAR* parameter could be used to adjust this slightly better. In most cases the effect is expected to be negligible. It can however show in the resulting distributions when single pixel amplitudes start to reach saturation.

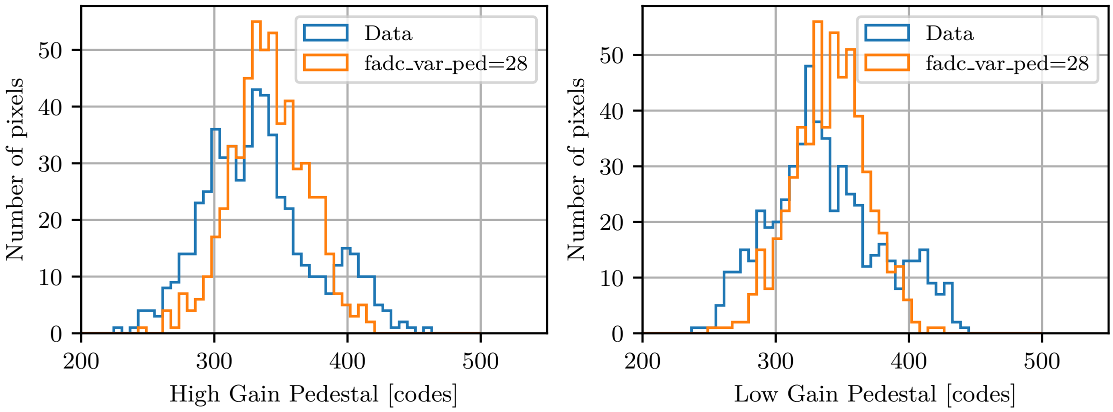
For the pedestals, values of $\mu = 340$ codes with $\sigma = 28$ codes are used. A comparison of one realisation in the simulation to the values seen in the experiment is shown in Figure 3.10. The spread corresponds to the spread around the first peak at ∼330 codes in the data and the mean is shifted slightly upward to also incorporate the pedestal values of the second peak at ∼400 codes. A shift of ∼10 codes corresponds to ∼0.34 p.e. for the high gain and ∼3.3 p.e. for the low gain.



(a) High gain

(b) Low gain

Figure 3.9: Comparison of the gain values in the simulation to the ones seen in data. The gains are sampled from a normal distribution around $\mu = 29\,\text{codes}\,\text{p.e.}^{-1}$ for the high gain and $\mu = 3\,\text{codes}\,\text{p.e.}^{-1}$ for the low gain. The spread is specified by a relative fluctuation of $\sigma = 0.08$. The difference between data and simulation for the low gain is a result of the dynamic range of the MAROC preamplifier for high and low gain (see text).



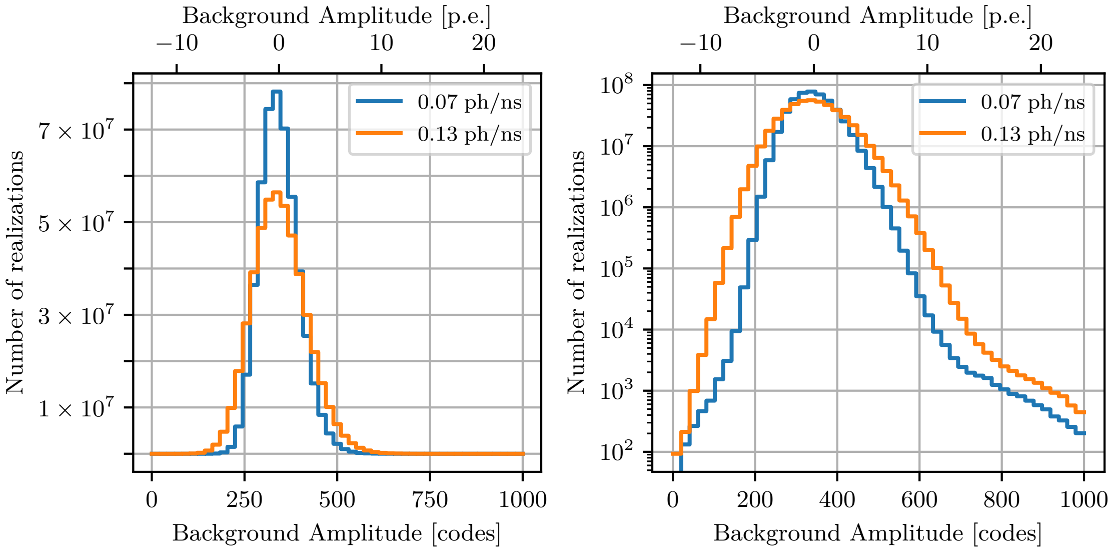
(a) High gain

(b) Low gain

Figure 3.10: Comparison of the pedestal values in the simulation to the measured ones. The simulated pedestals are sampled from a normal distribution around $\mu = 340$ codes with $\sigma = 28$ codes.

3.2.4 Pedestal and Night Sky Background

Each of the signals fluctuates around its pedestal value. The main contribution to this fluctuation is caused by the NSB. In the simulation, the NSB is specified as a rate r_{NSB} of photoelectrons ph per time in units of ph/ns.



(a) Linear scale

(b) Logarithmic scale

Figure 3.11: Resulting background distribution in one time slice for two background photoelectron rates. A pedestal value of 343 codes and a gain of 27.61 codes/p.e. was used. The tail above 10 p.e. to large values is an effect of the afterpulsing.

For each bin of the time slices, the number of photoelectrons is drawn from a Poisson distribution according to the rate r_{NSB}. For each of the resulting photoelectrons, the amplitude is then drawn from the SPE distribution, including afterpulsing, and added to the signal as described in Section 3.2.3. The resulting background distribution is shown in Figure 3.11 for rates of 0.07 ph/ns and 0.13 ph/ns.

To compare this with the background estimated in the data as described in Section 2.3.9, the same fit procedure as used for the real data analysis is applied to several values of r_{NSB}. Figure 3.12 shows the correlation of the width obtained by fitting as a function of r_{NSB}. This correlation can be used to determine the r_{NSB} setting for a given pedestal width estimate of the data.

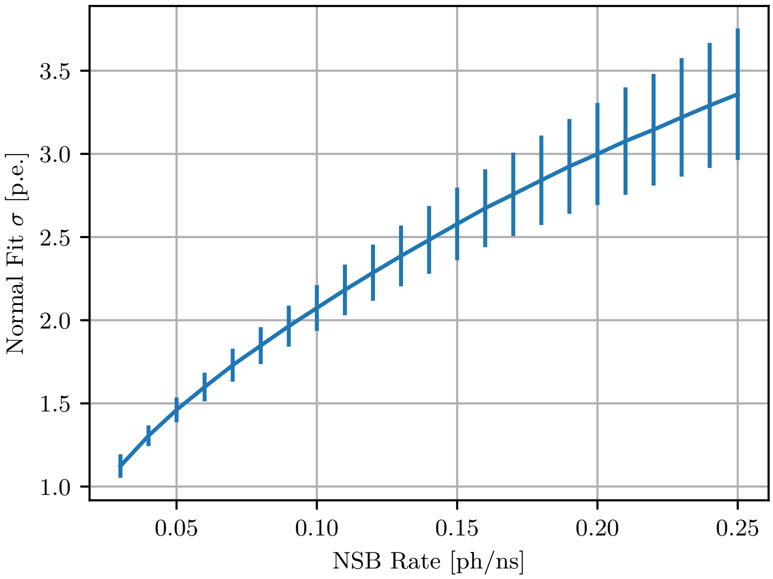


Figure 3.12: Normal distribution width values obtained via a fit of a normal distribution around the peak of the expected background distributions for different photoelectron background rates. For each rate, the fit is repeated 10000 for histograms with 50000 entries each. The resulting median of the obtained σ values and the standard deviation are plotted.

In the simulation, the signal shapes used for the background distribution are cut at the 5 %-level to save computing time. Because the algorithm to determine the signal within the 5 %-level in `sim_telarray` cuts away the downswing part of our signal shapes (see Figure 3.8), the code was adjusted to keep it. Keeping the downswing part for the background estimate has a pronounced effect on negative fluctuations below the pedestal level.

3.2.5 Trigger and Readout

For individual pixels, the MAROC FSB and the trigger discriminator are represented by the shapes of the discriminator signal and the discriminator threshold settings in `sim_telarray`. However, the trigger logic and the readout process explained in Section 2.3.4 cannot be implemented in `sim_telarray`, as partial readout of individual clusters at individual times is not supported. A cluster (MAROC board) produces a local trigger when any pixel is above the threshold. When any two pixels are above threshold within 15 ns, the Central Controller is informed. A telescope event is only recorded when the Central Controller checked the next neighbour condition. On a successful check of a trigger pair, all clusters which were in a local trigger and sent the local hold signal are read out. (The local hold signal is the signal which causes the momentary amplitudes to be stored in a buffer for a digitization.)

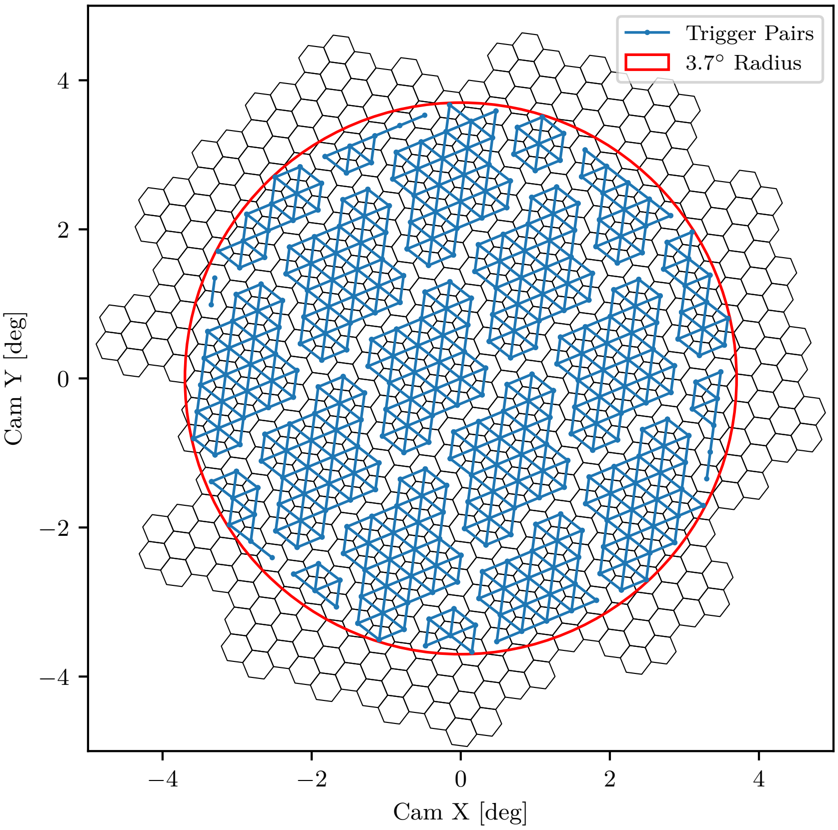
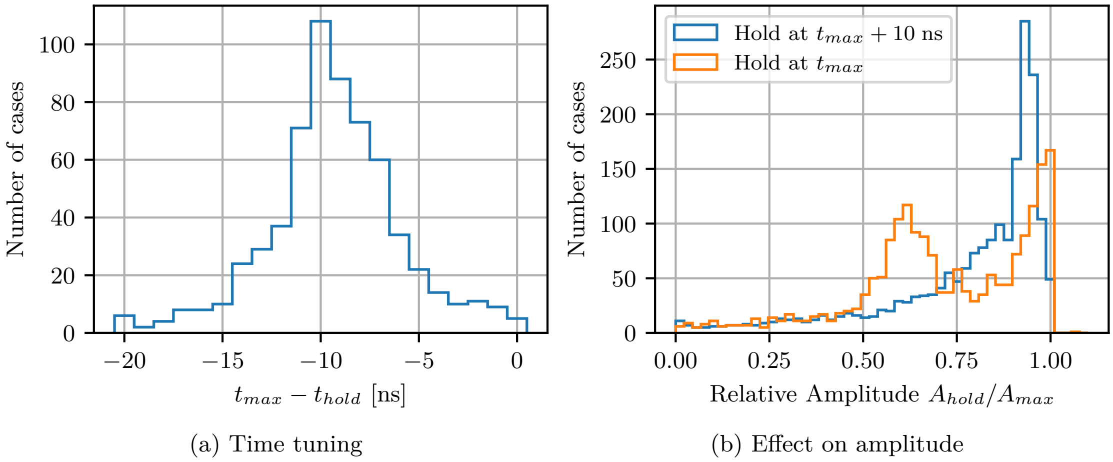


Figure 3.13: The 798 trigger pairs of next neighbours within a radius of 3.7°.

This two stage trigger mechanism is implemented with the following workaround: In `sim_telarray`, trigger groups, called sectors are defined for (a) the next neighbour conditions shown in Figure 3.13 and (b) for the conditions, that any pixel of a cluster fires. All events fulfilling any trigger of (a) or (b) are saved to the EventIO output file in sampling mode. Then, when `read_taiga` is run to convert the data to the HDF5 format, the triggering sectors are checked. Only when at least one of the next-neighbour conditions (a) is fulfilled, the event is actually considered as a triggering event. To simulate the cluster-wise readout via the hold delay, the first trigger that occurs in a cluster is used to determine the trigger time t_{trig} of the cluster. Usually, this is one of the sectors in (b). Then, the sampled pulse is read out at $t_{trig} + t_{hold}$, where t_{hold} is the hold delay.

The hold delay t_{hold} is tuned in a similar way as in the experiment. It is set such that it typically occurs 10 ns after the pulse shape maximum. Figure 3.14a shows the time difference between the time of the maximum time slice and the hold time $t_{max} - t_{hold}$ for $t_{hold} = 23.5$ ns for 100 10 TeV showers with a zenith distance of $\theta = 25°$ and an impact point at a distance of 100 m from the telescope. The telescope is pointing to $\theta = 26.2°$.



(a) Time tuning

(b) Effect on amplitude

Figure 3.14: (a) Hold delay tuned such that the readout occurs ∼10 ns after the peak. b) Fraction of the amplitude read out at hold time to the amplitude at the maximum of the signal. Two cases are considered: Readout at the time of the maximum and readout 10 ns after the maximum. Only pixels with a signal of more than 5 p.e. are considered.

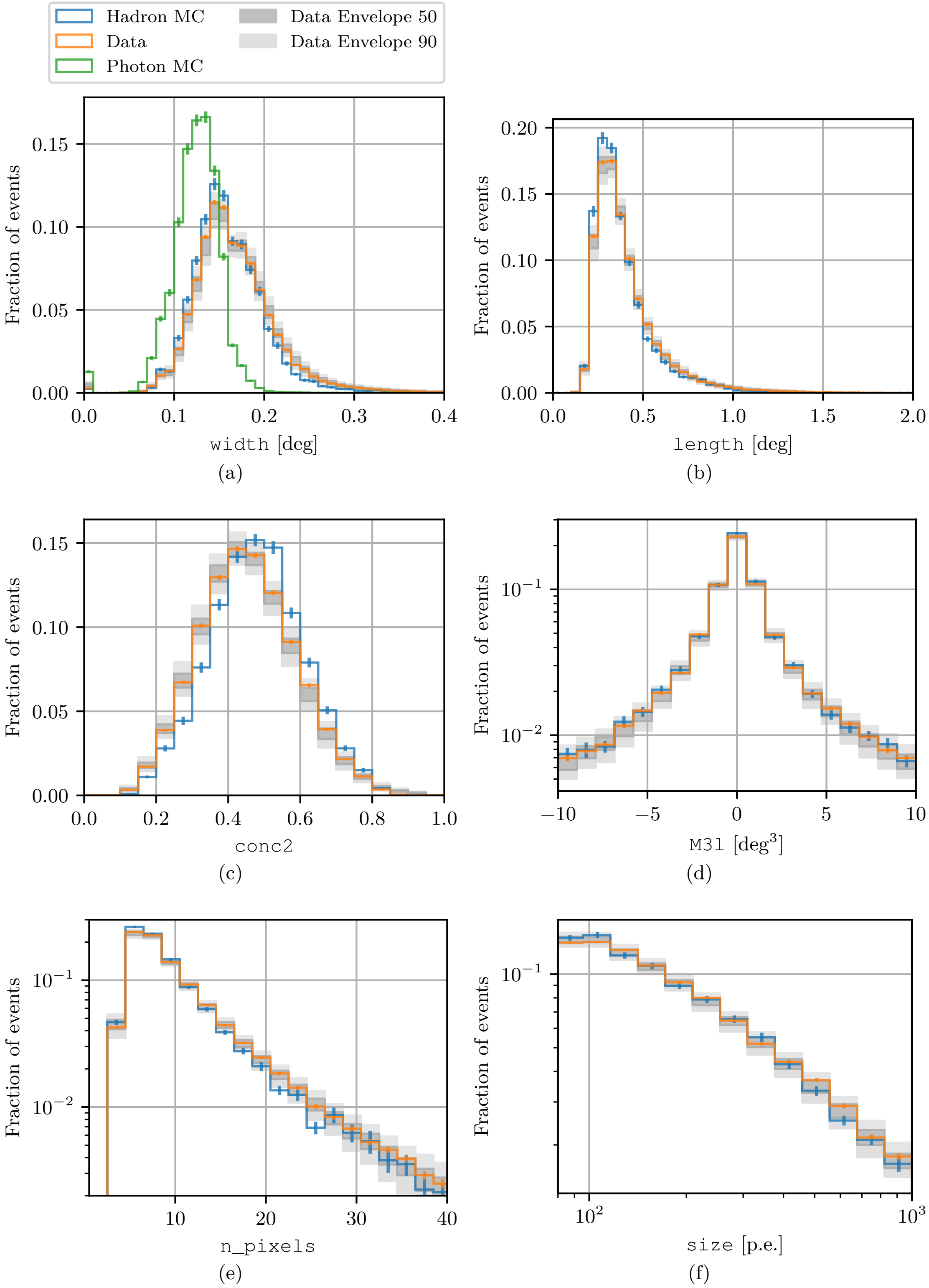
As explained in Section 2.3.4, t_{hold} is set to occur behind the maximum to prevent a readout before the peak for late signals. This should be more noticeable at larger core distances. To demonstrate that this can, for some showers, improve the average readout, the relative amplitude A_{hold}/A_{max} between the amplitude at the pulse maximum to the amplitude at the hold time is compared for $t_{hold} = 23.5$ ns and $t_{hold} = 13.5$ ns. Figure 3.14b shows A_{hold}/A_{max} for pixels with a signal of more than 5 p.e. for 50 TeV showers with an impact parameter of 200 m and $\theta = 25°$ and the telescope pointing to $\theta = 26.2$. When the readout occurs on average at t_{max}, there is a strong peak at 1 and a second peak at 0.6. For the case of 10 ns after the maximum, the peak at 0.6 vanishes and a wider and more pronounced peak at 0.9 occurs. In this way, some channels read a smaller fraction, but overall, the readout is more homogeneous.

The trigger threshold is set to ∼10 p.e. and the coincidence window of the discriminator is set to 15 ns.

3.2.6 Comparison with Data

To check whether the simulation describes the data well enough, the distributions of the image parameters used for the analysis are compared to real data. Figure 3.15 and Figure 3.16 show distributions of several image parameters for a zenith distance of $\theta = 31°$ after a basic set of cuts: `size` $> 80\,$p.e., `leakage` < 0.15 and `n_pixels` > 3. The 50 % and 90 % containment envelopes of the data are estimated from a combined probability density determined from the histogram values and uncertainties of numerous data intervals, each containing 8000 events.

Overall, all distributions show a suitable agreement. For `conc2` (Figure 3.15c), the deviation appears to be the most pronounced. The concentration in the data appears to be systematically lower. Possible reasons are a too small NSB rate and an underestimation of the PSF. For `width` (Figure 3.15a) and `length` (Figure 3.15b), the simulation seems to be slightly shifted to smaller values. This also indicates a higher concentration in the simulation and might hint at a mismatching PSF or a too small NSB rate. One reason for the underestimation at larger values also visible for the `n_pixels` (Figure 3.15e) and `size` (Figure 3.15f) distributions could be the missing high energy simulations of heavier particles which result in wide and long images. High energy events were only simulated for helium and proton. In the distribution of the `dist_cog_center` (Figure 3.16a), the number of events drops sharply after passing $\sim$3.7°. Pixels outside of that radius are disabled from the trigger. The simulation seems to accept slightly more events towards the border of the telescope. This is likely caused by the non-triggering cluster in the lower center, also visible as a void in the histogram of the CoG distribution for data (Figure 3.16f). The other obvious difference between the CoG distributions of simulation (Figure 3.16e) and data (Figure 3.16f) are the three voids caused by malfunctioning pixels. In the simulation, they are excluded for the higher level analysis but are enabled at the trigger level. Otherwise, the distributions seem to agree reasonably well. The `gammaness` (Figure 3.16b) parameter used in (g) is explained in Section 4.1.5. The distribution shows a good match except for an underestimation close to 0. Here, some noise triggers could be the reason.



(a)

(b)

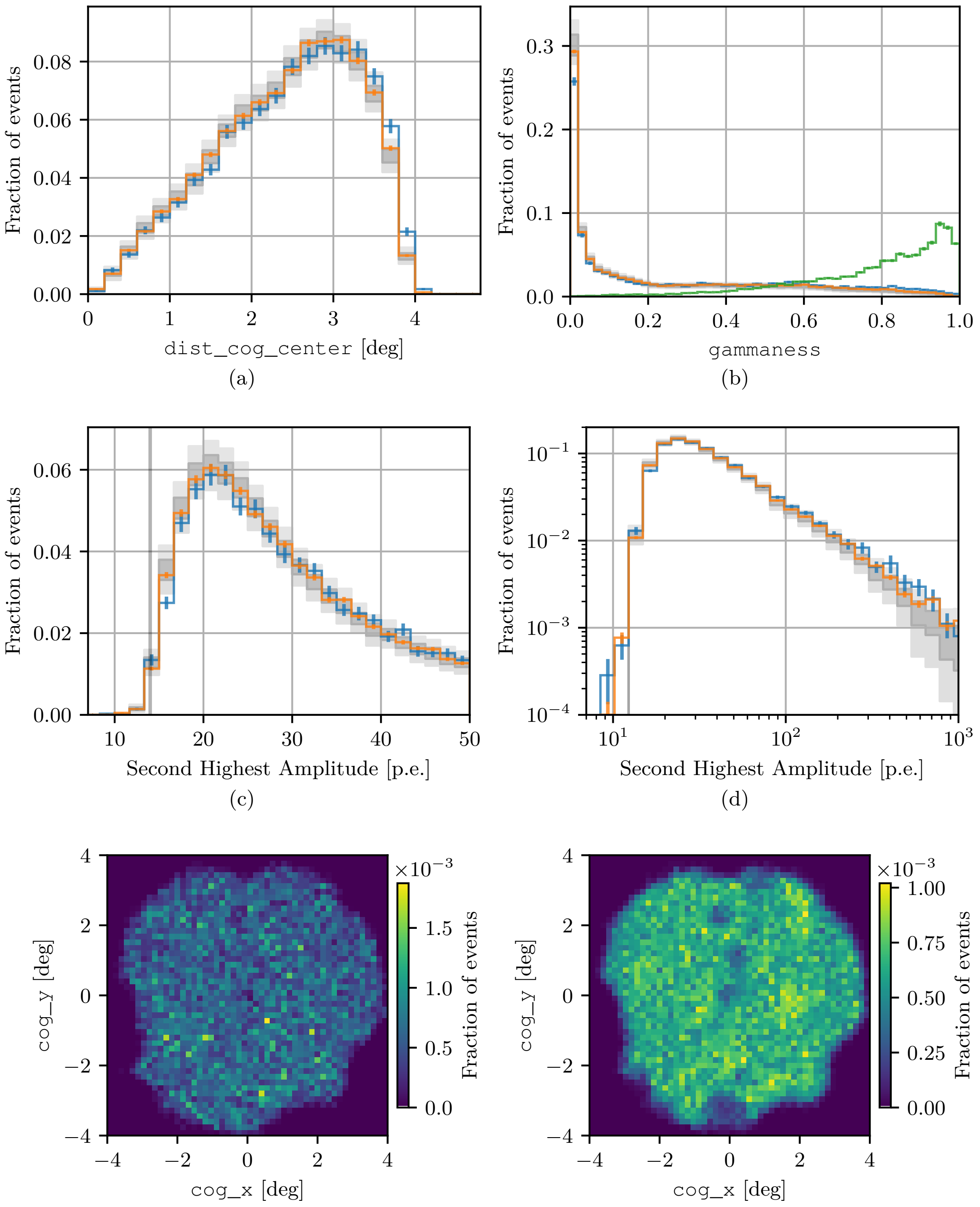
(c)

(d)

(e)

(f)

Figure 3.15: Comparison of simulation with data (1/2).



(a) (b)

(c) (d)

(e) Simulation (f) Data

Figure 3.16: Comparison of simulation with data (2/2).

3.2.7 Expected Background Rate

To estimate the expected rate in the experiment, the effective area is calculated and the expected rate is determined according to the polygonato model and the DAMPE fits (see Section 3.1.3). The effective area $A_{eff,i}$ of an energy bin i is calculated as

$$A_{eff,i} = \frac{h_i}{H_i} A_{sim} \Omega_{sim}, \tag{3.7}$$

where h_i is the number of events that survived the cleaning and the `size` $> 80\,$p.e. cut and H_i is the number of events simulated in the bin. When the events of multiple telescopes are plotted together to increase the satistics, H_i is multiplied by the number of telescopes. A_{eff} and Ω_{sim} are the simulated area and field of view and are related to the simulated viewcone radius Ω and the scattering rectangles X_s and Y_s via

$$\Omega_{sim} = 2\pi(1 - \cos\Omega) \approx 1.14\,\mathrm{sr}, \mathrm{and} \tag{3.8}$$

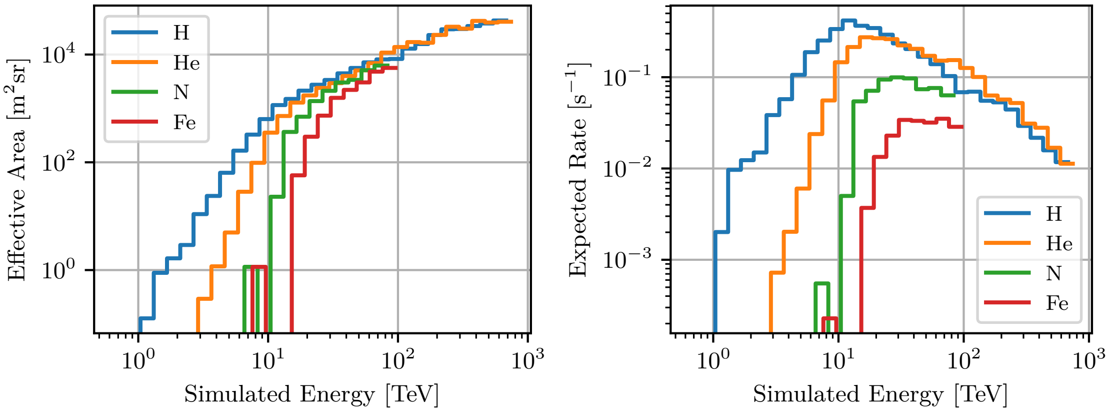
$$A_{sim} = 2X_s \cdot 2Y_s \approx 3.03\,\mathrm{km}^2. \tag{3.9}$$

The expected rate r_i for each type of particle is then calculated for the lower and upper bin boundaries E_l^i and E_h^i as

$$r_i = A_{eff}^i \int_{E_l^i}^{E_h^i} f(E)dE, \tag{3.10}$$

where $f(E)$ is the differential flux according to the polygonato models and DAMPE fits as described in Section 3.1.3.

Figure 3.17 shows the effective areas and rates for proton, helium, nitrogen and iron for a cut of `size` $> 80\,$p.e. and $l_1 = 14\,$p.e. and $l_2 = 7\,$p.e. cleaning. For each of the particles, the binning was chosen such that the bin boundaries at each decade fall to the boundaries of the energy ranges in the simulation sets. The total expected rate for all particles is $r_{80} \approx 7.1\,\mathrm{s}^{-1}$. The rate scaled to the zenith is $s_{80} = \frac{r_{80}}{\cos\theta} \approx 8.3\,\mathrm{s}^{-1}$. This falls in the range used for good-weather selection of 7.3 Hz to 12.2 Hz described in Section 5.2.



(a) Effective area (b) Expected rate

Figure 3.17: Effective area (a) and the expected rate (b) after a cut of `size` > 80 p.e.. The sum of all the rates in the plot is $\sim 7.1\,\mathrm{s}^{-1}$.

3.3 Simulation of HiSCORE

Starting from the same EventIO files as the IACT simulation, the HiSCORE stations are simulated with `sim_score` which is based on a skeleton program provided with `sim_telarray` as described in Tluczykont et al. [38], Hampf [88], Epimakhov [50], and Porelli [37]. The output is stored as HDF5 in the same format as that used for the real data (see Section 2.2.1). One major change of the code was the removal of all root-based functionality to drop the cumbersome dependency and to speed it up.

Each photon bunch is ray traced to the Winston cone entrances tilted by 25° to the south. Here, the code was modified to handle station heights different from 0 correctly in the arrival time calculation.

After passing the cones, wavelength-dependent plexiglass absorption, angle-dependent Winston cone acceptance and wavelength-dependent quantum efficiency for an Electron Tubes PMT 9354KB (plus fixed photoelectron collection efficiency of 0.9) are checked. For each of the four PMTs, a signal channel with a window of 4000 ns with steps of 1 ns is simulated. On arrival of a photoelectron, the electronic response[50] is added to the signal without the application of a SPE response.

When all photon bunches are processed, a random slice of precalculated NSB noise is added. The NSB noise corresponds to a rate of 18 photoelectrons/ns and includes the effects of the pulse shape and of a SPE parametrisation including afterpulsing.

The station triggers if the sum of all four channels is above a threshold for 10 ns. The threshold is set such that a typical signal with 180 p.e. causes a trigger.

When a station triggers, a pulse analysis (see Section 2.2.1) is performed and the determined amplitude, edge time, FWHM and integrated charge are written to the HDF5 output file.

Chapter 4

Reconstruction

After conversion to HDF5-files, all further analysis is performed in python. For the IACT data, the reconstruction is done with Random Forests. After starting with a brief introduction to Random Forests, the regression of `disp`-Parameter for origin reconstruction is outlined in Section 4.1.3, the regression of the primary energy in 4.1.4 and the gamma/background separation with a Random Forest classifier is described in Section 4.1.5. The analysis presented here is an enhanced version of the one used in [47] with more data and increased MC statistics.

The HiSCORE reconstruction is done with the fit methods developed in Hampf [88], Hampf et al. [39], and Porelli [37]. For the shower impact position reconstruction, the center of gravity and the fit to the amplitude distribution function (Section 4.2.1.2) are used. For origin reconstruction, three fit methods to the arrival time distributions are employed: a fit with a planar light front (Section 4.2.2.1), with a curved wavefront (Section 4.2.2.2) and a fit with a model for the arrival time (Section 4.2.2.3).

4.1 IACT Random Forest Analysis

The basic task of IACT reconstruction is to infer the primary particle type, from where it hit the atmosphere and what its energy was. In this analysis, Random Forests are used to accomplish the three tasks. They are trained on the higher level image parameters using MC and data and were first used with IACTs in [89]. Now, they are widely used by many major collaborations, as they provide a robust method giving a good estimate in many situations.

As opposed to more recent deep learning techniques trained on the telescope images, this approach on the more abstract image parameters is less computationally intensive and does not require the telescope images to match the real data as precisely.

4.1.1 Random Forest

In this analysis, the `scikit-learn` [90, 91] implementation of the Random Forest algorithm is used. The Random Forest concept was developed by Breiman [92] in 2001.

A Random Forest consists of several Decision Trees. Decision Trees are a Machine Learning approach to regression or classification tasks. Given a set of input parameters, they are grown by creating branches by splits on a value of one of the parameters. Each following branch can again be split into further branches. The final nodes are called leafs, and the number of split levels is called depth. How the image parameters and values for each split are determined depends on the implementation and configuration of Decision Trees. In the `scikit-learn` learn implementation, the split value v and parameter par splitting the n samples at a node are determined by the decrease of the impurity function G depending on the value of a loss function H at the left ($par \leq v$) and right ($par > v$) child nodes:

$$G = \frac{\sum_i^{n_{left}} w_i}{\sum_i^n w_i} H(left) + \frac{\sum_i^{n_{right}} w_i}{\sum_i^n w_i} H(right) \tag{4.1}$$

Here, w_i is the weight of the i-th sample of all target samples x in the node. n_{left} and n_{right} are the corresponding number of samples in the left and right subsets of the samples x.

In this analysis, two loss functions are used. In the case of regression, it is the weighted mean squared error MSE:

$$MSE(x,p) = \frac{\sum_i^n w_i \left(x_i - p_i\right)^2}{\sum_i^n w_i}, \tag{4.2}$$

where x_i is the i-th value of the n target samples x and p_i is the i-th value of the corresponding predictions p. For a given node, the predicted value is determined by the mean of the samples $\overline{x}$ at the node:

$$\overline{x} = \frac{\sum_i^n w_i x_i}{\sum_i^n w_i} \tag{4.3}$$

$$H_{MSE}(x) = \frac{\sum_i^n w_i \left(x_i - \overline{x}\right)^2}{\sum_i^n w_i} = \overline{x^2} - \overline{x}^2. \tag{4.4}$$

Therefore, the regression loss function is the variance at the node. If a good separation is found, the impurity G at the node will decrease.

For the classification, it is the Gini impurity[1]. The fraction of gammas at node f_{Gamma} is

$$f_{Gamma} = \frac{\sum_i^{n_{Gamma}} w_i}{\sum_i^n w_i}, \tag{4.5}$$

and the fraction of background events is f_{BG}

$$f_{BG} = \frac{\sum_i^{n_{BG}} w_i}{\sum_i^n w_i}. \tag{4.6}$$

The Gini loss function is then defined as

$$H_{Gini} = f_{Gamma}(1 - f_{Gamma}) + f_{BG}(1 - f_{BG}). \tag{4.7}$$

This function will have a maximum at $0.5^2 + 0.5^2 = 0.5$ when the sum of weights in the gamma fraction is the same as the sum of weights in the BG fraction. For any split having unequal shares, the value will decrease and if a pure split with e.g. only gammas is achieved, it would have a value of 0. An example of the first four levels of one Decision Tree for the gamma/BG-classifier is shown in Figure 4.1.

In a Random Forest, numerous Decision Trees are grown on different random subsets of the data. The individual results of the ensemble of Decision Trees are then averaged to provide a better estimate of the true value in the case of a regression. In a classification task, the averaging over the Decision Trees leads to a probability-like value representing the fraction of the trees voting for a class. In the case of the gamma/BG separation, this probability-like value is called `gammaness` if it is chosen such that gammas tend to have values closer to 1 and background particles closer to 0.

Generally, Random Forests show good generalization and give a small variance.[92][2]

More information about the Random Forests can be found in the documentation of the implementation[3].

To evaluate how well the Random Forests perform overall, so-called metrics are used. `scikit-learn` comes with a variety of different metrics.

---

[1] `https://scikit-learn.org/stable/modules/tree.html#tree-mathematical-formulation`

[2] `http://scikit-learn.org/stable/modules/ensemble.html#forest`

[3] `http://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html`

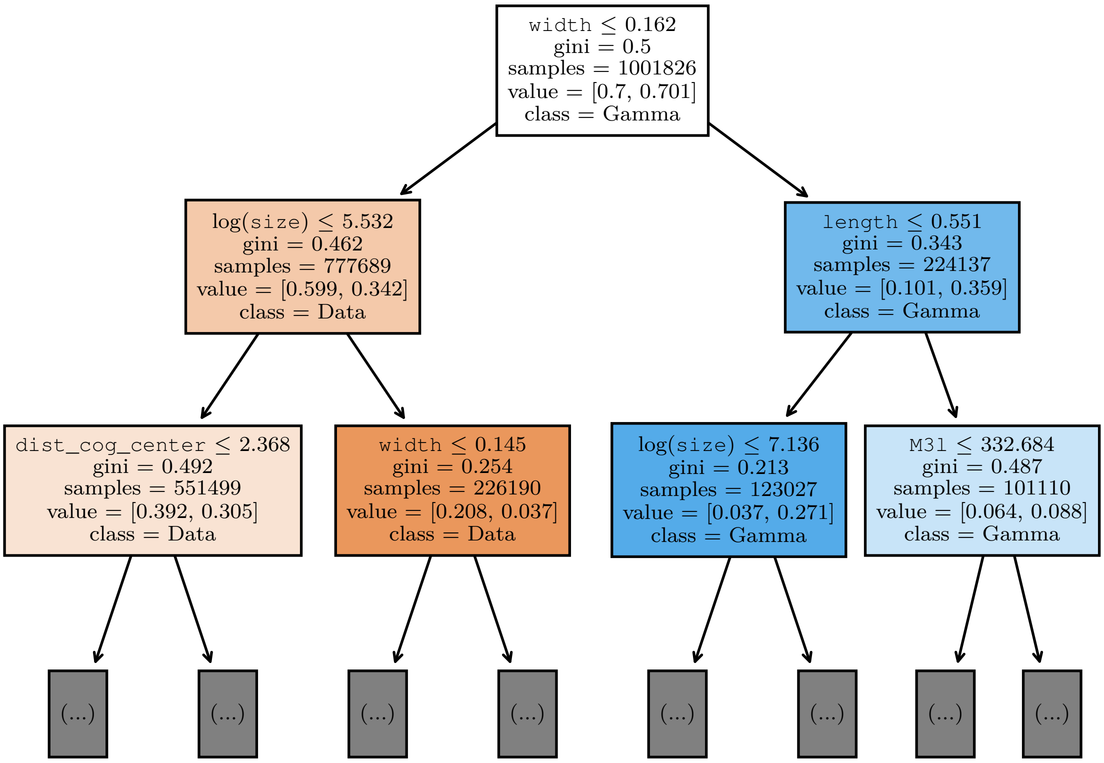

width ≤ 0.162
gini = 0.5
samples = 1001826
value = [0.7, 0.701]
class = Gamma
log(size) ≤ 5.532
gini = 0.462
samples = 777689
value = [0.599, 0.342]
class = Data
length ≤ 0.551
gini = 0.343
samples = 224137
value = [0.101, 0.359]
class = Gamma
dist_cog_center ≤ 2.368
gini = 0.492
samples = 551499
value = [0.392, 0.305]
class = Data
width ≤ 0.145
gini = 0.254
samples = 226190
value = [0.208, 0.037]
class = Data
log(size) ≤ 7.136
gini = 0.213
samples = 123027
value = [0.037, 0.271]
class = Gamma
M31 ≤ 332.684
gini = 0.487
samples = 101110
value = [0.064, 0.088]
class = Gamma
(...)

Figure 4.1: The first two levels of one of the Decision Trees of the Random Forest gamma/background-classifier. Each box represents one node. The first line is the chosen image parameter and cut, the second line shows the *Gini* value of the node. Samples is the number of samples in the node. For value, the first of both values is the sum of the weight of the background samples and the second value is the sum of the weight of the gamma samples in the node. The last line refers to the dominating class in this node (The class the classifier would vote for if it were to stop here). The color of the boxes corresponds to the class. The higher the *Gini* value, the lighter the coloring.

For the classification task, the *accuracy* metric is the standard metric. It is defined as

$$accuracy = \frac{\sum_{i}^{n_c} w_i}{\sum_i^n w_i}, \tag{4.8}$$

with the number of all samples n and the number of correctly predicted samples n_c. Here, correctly predicted means that all values with `gammaness` > 0.5 are classified as a gamma and else as BG. However, in the case of the gamma/BG separation, it is not very important to classify the maximum number of events in both classes correctly. Instead, it is more important to minimize the wrongly identified background events while at the same time keeping most of the gammas. A better metric is the Q-factor (Equation 4.11) at a given fraction of gammas surviving, as explained in 4.1.5. The `gammaness` cut value is then chosen based on Q and the fraction of gammas surviving.

For the two regression tasks of origin estimation via `disp`-regression and for determining the energy, two metrics are used. The first metric is the r^2 score, also called coefficient of determination. It is the standard metric in the `scikit-learn` implementation and is defined as

$$r^2 = \left(1 - \frac{\sum_i^n w_i \left(p_i - x_i\right)^2}{\sum_i^n w_i \left(x_i - \overline{x}\right)^2}\right). \tag{4.9}$$

If the regressor gave the mean value $\overline{x}$ (Equation 4.3) for each event, it would be 0. For better predictions, the nominator becomes smaller, and r^2 approaches 1. For regressions such as energy regression, where the residuals for predictions of larger values are generally larger, r^2 is strongly influenced by those larger values. Even though this effect is somewhat compensated by the low probability of the higher energies, when comparing two Random Forests for the same set, a larger r^2 value might not always reflect the overall better regressor. To overcome this, the second metric, the mean absolute percentage error *MAPE* is used:

$$MAPE(x, p) = \frac{1}{\sum_i^n w_i} \sum_i^n w_i \frac{|x_i - p_i|}{x_i}. \tag{4.10}$$

By only using relative values, this metric is well suited for the power law distribution of the energy of the events.

For the three tasks, the following parameters will be used: `size`, `n_pixels`, `width`, `length`, `M3l`, `dist_cog_center`, `conc2` and $wl = \frac{\texttt{width}}{\texttt{length}}$. The estimates of the metrics will be checked via 10-fold cross-validation. This means, that the metric is calculated for 10 subsamples randomly split from the test set.

4.1.2 Training and Test Sets

The gamma MC samples with a viewcone of $\Omega = 10°$ around θ_0 presented in Table 3.1 of Section 3.1 are used in two ways: As a diffuse set and as a point source set.

For the diffuse set, the zenith distance θ of the telescopes is randomly selected between $\theta_0 \pm 5°$ for each of the runs of each of the samples. This ensures that the full FoV of the camera is covered for all pointings.

In the point source case, the telescopes change to a pointing at the wobble offset for every event. To achieve this, the *RANDOM_VIEWING_RING* option of `sim_telarray` was modified. For every event, the telescope viewing direction is set to a random position in a uniform ring around the origin with inner and outer radii of 1.15° and 1.25°. This introduces an additional uncertainty, as the ZD dependent random mirror alignment is only computed for the nominal pointing position at the beginning of each run. At a viewcone of 10°, the resulting PSF difference is ~0.02° between zenith distances of 46° and 56°, if the assumption of PSF scaling in Section 3.2.1 holds true.

To increase the statistics, the triggering events of up to four telescopes set to the IACT 1 configuration are combined.

For the gamma/background classifier, real data is used as the background (BG).

70 % of the events are used for the training set and 30 % for the test set. The gamma samples used for training are weighed to $\propto E^{-2}$. For testing, a Crab-like slope of $\propto E^{-2.8}$ is used. As the sets do not have a uniform distribution in ZD, the sets are weighted so that the weighted ZD distribution is flat (see Section 3.1.3).

In total, the point source set has 1.38×10^6 triggering events in the training set and 0.59×10^6 events in the test set. For the diffuse set, 209×10^3 and 90×10^3 events are available. For the background set, 0.32×10^6 and 0.14×10^6 are used. The total number of simulation events is 61.17×10^6 (times the number of telescopes used).

4.1.3 Origin Reconstruction

For the origin reconstruction, the `disp` parameter[72] is used.

As the shower develops around the trajectory the primary particle would have had, the image of the Cherenkov photons in the camera is elongated along this direction. With the description of the shower as an ellipse and using this assumption, that the particle must have originated from somewhere along the elongated axis, the origin lies somewhere on the projection of the major axis. To find where on the axis, the distance `disp` of the center of gravity to the origin needs to be estimated. As `disp` is an absolute value, the ambiguity in which direction the origin lies also needs to be resolved. Here, the sign of the third moment along the major axis, the `M3l` parameter is used. As photons from further away than the shower maximum will have smaller angular distances to the CoG position, a skewness of the image towards the origin can be expected.

The Random Forest is set to grow 50 trees with a maximum depth of 15 levels. At each split, a minimum of 16 samples are required and the minimum number of samples in a leaf is eight.[4]

To prevent a bias towards a fixed source position in the camera, the diffuse set is used for training. On the test set split into eight random subsets, the regressor achieves scores of $r^2 = 0.813 \pm 0.005$ and $MAPE = 0.235 \pm 0.006$, indicating a good regression.

---

[4] $n\_trees = 50$, $max\_depth = 16$, $min\_samples\_split = 16$, $min\_samples\_leaf = 8$

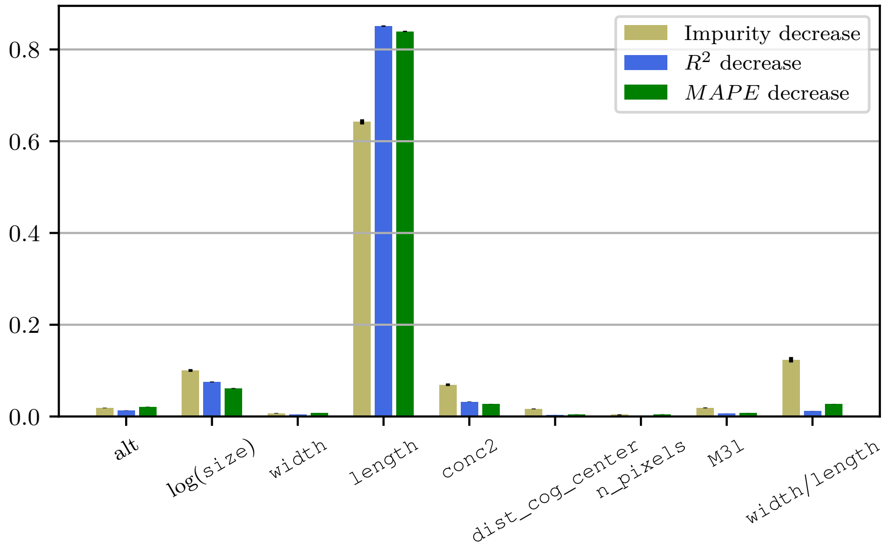


Figure 4.2: Feature importances for the `disp` regression Random Forest.

Feature	`disp` [%]	Energy [%]	G/H [%]
alt	1.9 ± 0.1	11.42 ± 0.04	2.5 ± 0.2
`size`	10.0 ± 0.3	6.33 ± 0.08	17.4 ± 3.5
`width`	0.7 ± 0.1	0.23 ± 0.02	26.7 ± 8.2
`length`	64.2 ± 0.6	55.78 ± 0.10	6.8 ± 3.1
`conc2`	6.9 ± 0.3	5.52 ± 0.10	9.2 ± 4.9
`dist_cog_center`	1.7 ± 0.1	3.31 ± 0.03	18.5 ± 3.0
`n_pixels`	0.4 ± 0.1	16.58 ± 0.07	4.2 ± 2.1
`M3l`	1.9 ± 0.1	0.54 ± 0.01	2.4 ± 1.1
wl	12.4 ± 0.6	0.29 ± 0.02	12.3 ± 5.6

Table 4.1: Importances of the features in the Random Forests for the different tasks. The importances are based on the mean decrease in impurity.

The feature importances as reported by the `disp`-regressor based on the mean impurity decrease can be found in Table 4.1. Figure 4.2 shows the relative feature importances determined via the mean impurity decrease of the nodes and for the permutation importance for the two metrics. The permutation importances are determined via the following procedure: For a sample, the baseline score is determined. Then, the values of one of the features is randomly shuffled (while keeping the others the same). The observed decrease in the score is the permutation importance of the shuffled feature.

The most important parameter is by far the `length`. It would be expected to be larger for showers with larger `disp`. The low importance of *wl* in comparison is surprising, as it is usually used to find a parametrisation of `disp` on basic image parameters.

Figure 4.3 shows a histogram of the `disp` estimate $\texttt{disp}_{est}$ against the true value $\texttt{disp}_{sim}$ and Figure 4.4 shows the relative accuracy as a function of the simulated

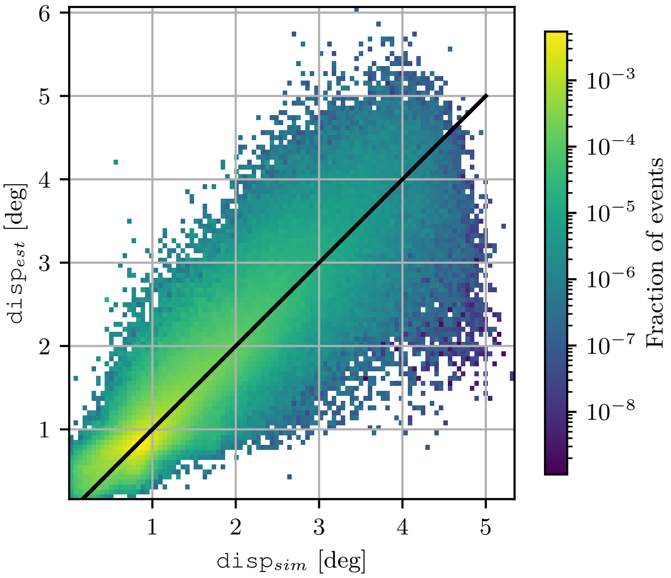


Figure 4.3: Estimate of the Random Forest regressor for the `disp` parameter against the true `disp` of the simulated events. The line with a slope of one visualizes the result a perfect estimator would give. The score of the regression is $r^2 = 0.813 \pm 0.005$ and $MAPE = 0.235 \pm 0.006$. The data shows the test sample after quality cuts.

energy E_{sim}. The relative accuracy is defined as $\frac{\texttt{disp}_{est}}{\texttt{disp}_{sim}} - 1$, where $\texttt{disp}_{est}$ is the estimated and $\texttt{disp}_{sim}$ is the true value. On this distribution, the median is called bias and the half of the interquantile distance between 16 % and 84 % is called resolution (half of the interval containing 68 % of the data and therefore corresponding to the containment in one σ for a normal distribution). The reconstruction works well between `disp` values of 1° and 4° with a bias below 5 % and a resolution below 20 %. The events below 1° have strongly increasing bias and variance, so they are reconstructed to too large values. For these events, the showers appear almost circular and exhibit stronger fluctuations, making them harder to reconstruct. At `disp` values above 4°, the bias deviates strongly while the variance reduces slightly. Here, the most likely explanation is, that this deviation at the border is caused by the averaging process of the Random Forest. For each Decision Tree, the maximum obtainable value is given by the mean value of the leaf with the highest values. Additionally, only some fraction of the trees will reconstruct to the highest leaf and a number of trees will give smaller estimates. When averaging the averaged values, this causes a systematic underestimation which increases closer to the edge values of the simulated parameter space.

Figure 4.5 shows the angular resolution (defined as the 68 %-percentile) as a function of the simulated energy E_{sim}. The blue (orange) curves show only the events with (without) correct `disp`-reconstruction direction (sign of `M3l` gives real direction). In the realistic case with a cut on `gammaness` > 0.9, the direction reconstruction achieves a resolution of about 0.2° for energies between 10 TeV and 50 TeV. For larger energies, the resolution

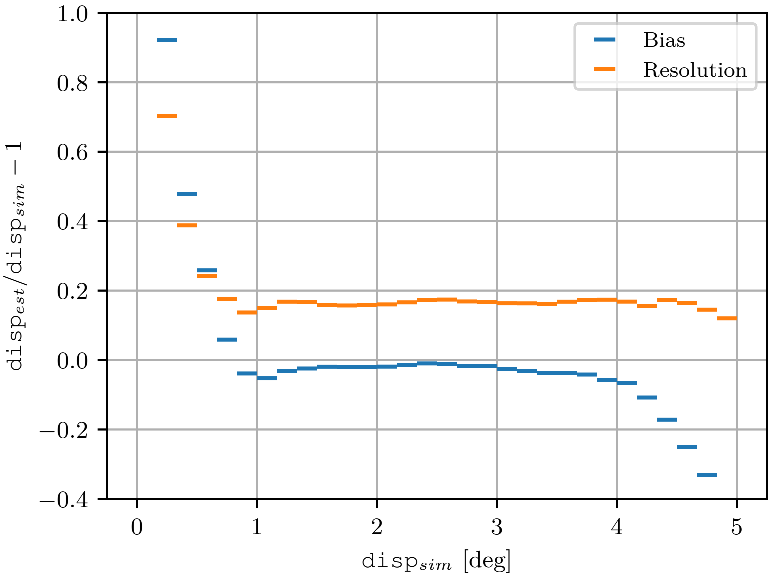


Figure 4.4: Median (bias) and half of the 16 %- to 84 %-quantiles (resolution) of $\frac{\texttt{disp}_{est}}{\texttt{disp}_{sim}} - 1$ of the Random Forest `disp` parameter reconstruction as a function of the simulated energy E_{sim}. The score of the regression is $r^2 = 0.813 \pm 0.005$ and $MAPE = 0.235 \pm 0.006$. The data shows the results for the test sample after quality cuts.

degrades to about 0.45°, even though a better resolution would be expected because the showers are brighter and thus potentially easier to reconstruct. However, at larger energies, showers with large impact parameters get more common and the fraction of events with large values of `disp` increases. As the `disp` resolution is roughly constant at 20 % between 1° and 4°, the absolute values of the uncertainty increases. Additionally, at higher energies, the fraction of events reconstructed in the wrong direction increases. This is likely caused by a worse accuracy of the third moment for events which are cut off at the edges and for events with saturating pixels. A resolution of ∼0.2° is acceptable but likely has some room for improvement. At energies of ∼100 TeV, where the resolution starts to get worse, HiSCORE should already trigger and provide a similar angular resolution improving with increasing energy. Additionally, also at lower energies, for events where only a few and possibly even only one or two stations trigger, the angular resolution can be improved as investigated in Wertz [93]. It should also be possible to resolve the ambiguity of the direction reconstruction without the `M3l` parameter.

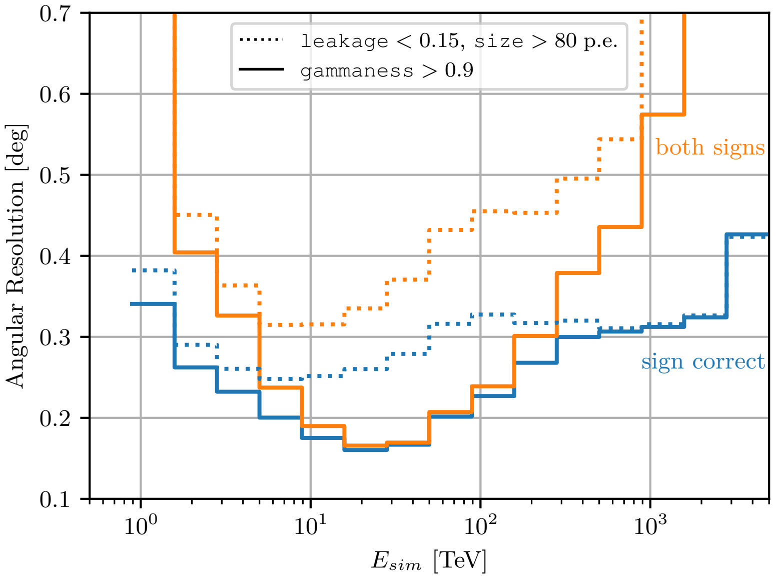


Figure 4.5: Angular resolution (68 % containment) for the origin reconstruction with the Random Forest regressor for the `disp` parameter as a function of the simulated energy. For the dotted lines, `size` > 80 and `leakage` < 0.15 are applied. For the solid lines, `gammaness` > 0.9 is additionally applied. The blue curves represent the events where `disp` points towards the source position, while the orange curves represent all events.

4.1.4 Energy Reconstruction

For the energy regressor, the Random Forest is set to the same parameters as the `disp` regressor. 50 trees are grown with a maximum depth of 15 levels. At each split, a minimum of 16 samples are required and the minimum number of samples in a leaf is eight.[5]

The point source set is used and the the regression results in $r^2 = 0.829 \pm 0.001$ and $MAPE = 0.25 \pm 0.06$ on the eight random splits of the test set, indicating a good regression. The feature importances as reported by the regressor are listed in Table 4.1. Figure 4.6 additionally shows the importances obtained via permutation (Section 4.1.3). Interestingly, the most important parameter when looking at the decrease in impurity of the nodes is the `length`, while the `size` seems very unimportant. One would expect `size` to be relatively important, as the number of photons in the showers increases with increasing energy. The permutation importances show this expected behaviour and give `size` as the most important parameter at 28 %. For the permutation importances, `length` drops to 25 %, followed by altitude and `dist_cog_center` with 21 % and 17 %. The `dist_cog_center` and `length` importances are expected, as these parameters correlate with the changing impact and X_{max} positions. As opposed to the `disp` regression, the altitude seems to have a significant impact on the energy estimate. `width` has no influence on the energy reconstruction.

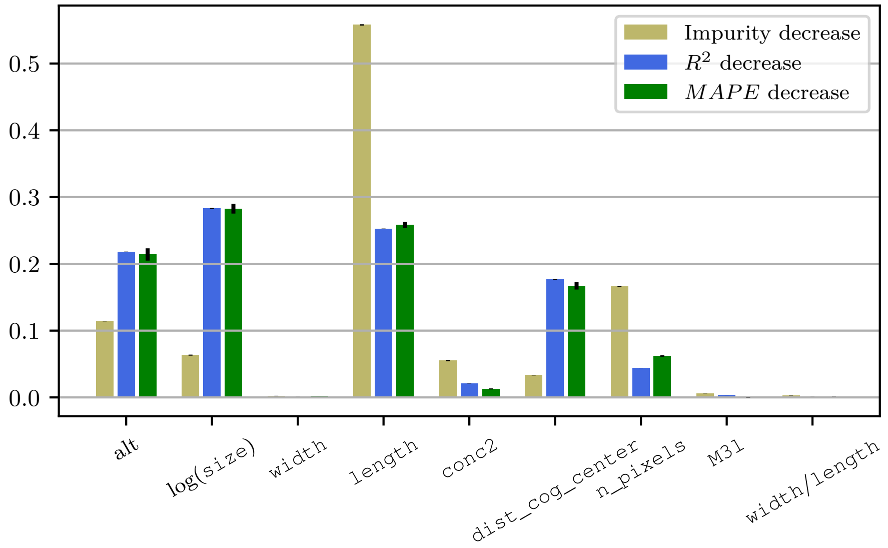


Figure 4.6: Feature importances for the energy regression Random Forest.

The histogram in Figure 4.7 of true energy E_{sim} and the energy estimated by the RF E_{est} indicates a linear performance in the energy range of roughly 15 TeV to 1 PeV. A better

---

[5] $n\_trees = 50$, $max\_depth = 16$, $min\_samples\_split = 16$, $min\_samples\_leaf = 8$

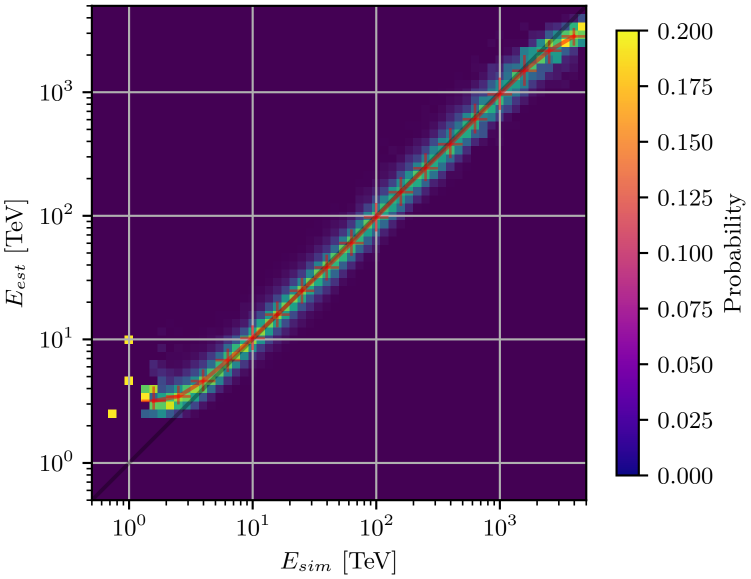


Figure 4.7: Energy estimate of the Random Forest regressor as a function of the true energy. The line with a slope of one visualizes the result a perfect estimator would give. The scores of the regression are $r^2 = 0.829 \pm 0.001$ and $MAPE = 0.25 \pm 0.06$. The red markers show the median and 16 %- to 84 %-quantiles.

way of representing the performance is by plotting the relative differences $\frac{E_{est}-E_{sim}}{E_{sim}}$ against the true energy. Figure 4.8 indicates, that, overall, the energy estimator shows roughly linear behaviour between 7 TeV to 1000 TeV. At small true energies, the energies are overestimated. This is, because of all low-energy events, only the ones with large `size` survive and the energies get overestimated. At larger energies, the larger images might not be fully contained inside of the camera and central pixels might saturate, which can explain a part of the systematic deviation for these energies. However, as already explained in Section 4.1.3, closer to the higher end of the simulated energy range, systematic overestimation arises from the averaging of the regressor. To quantify the differences at different energies, indicators of the distributions of ΔE are the mean, the median, the variance and the half of the 16 %- and 84%-quantile differences. Fig. 4.9 shows these quantities as a function of the simulated energy.

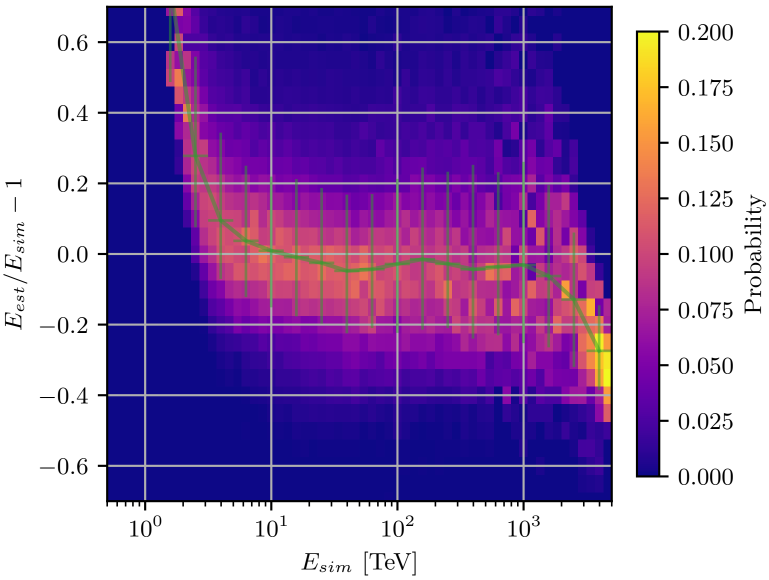


Figure 4.8: Relative accuracy of the energy estimate of the energy Random Forest regressor as a function of the true energy. The line at 0 visualizes the result a perfect estimator would give. The scores of the regression are $r^2 = 0.829 \pm 0.001$ and $MAPE = 0.25 \pm 0.06$. The green line visualises the median and 16 %- to 84 %-quantiles.

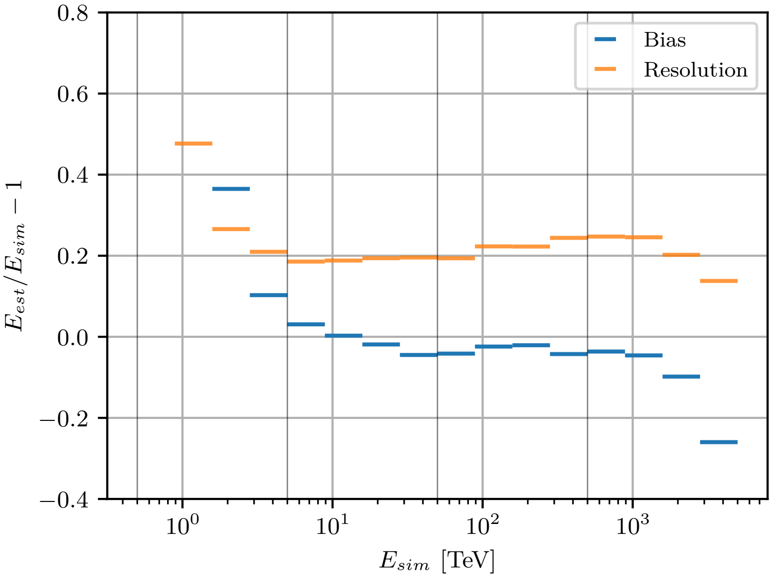


Figure 4.9: Median (bias) and half of the 16 %- to 84 %-quantiles (resolution) of $\frac{E_{est}}{E_{sim}} - 1$, where E_{est} and E_{sim} are the reconstructed and simulated energy, respectively. The scores of the regression are $r^2 = 0.829 \pm 0.001$ and $MAPE = 0.25 \pm 0.06$.

4.1.5 Gamma-Background Separation

To determine if the particle was a gamma or any other particle considered background in this analysis, a Random Forest classifier is trained. The background sample is real data taken during the 2019-20 season. The runs are chosen such that they cover the simulation zenith distances and that enough statistics at higher estimated energies remain. For the observation data, weights are assigned in such a way, that the weighted events follow a flat altitude distribution (see Section 3.1.3).

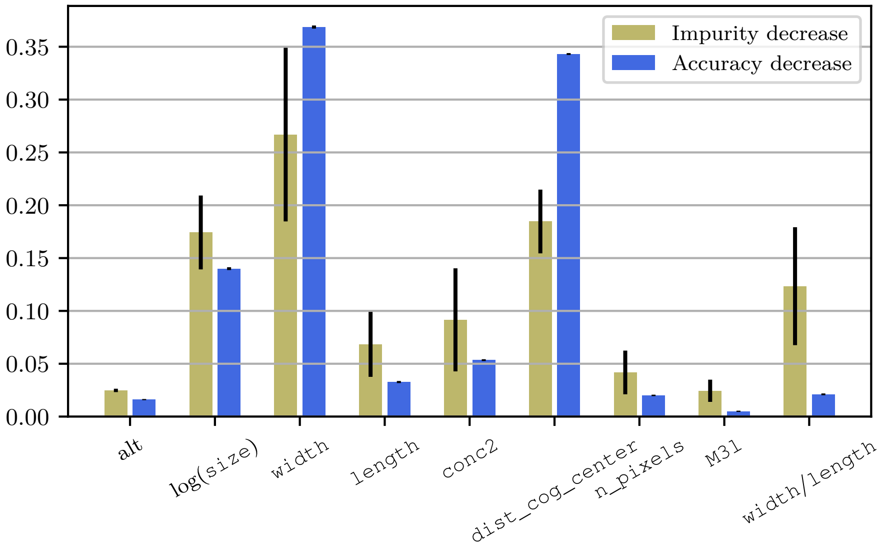


Figure 4.10: Feature importances for the gamma/background Random Forest classifier.

With the same settings as the regressors, the Random Forest classifier grows 50 trees with a maximum depth of 15 levels. At each split, a minimum of 16 samples are required and the minimum number of samples in a leaf is eight.[6] The accuracy on eight random subsets of the test set is $accuracy = 0.857 \pm 0.002$.

Figure 4.10 shows the feature importances of the classifier. As expected, `width` and `dist_cog_center` seem to be the parameters with the highest discrimination power.

To quantify the performance of the gamma/BG discrimination in the data analysis, the accuracy score defined for `gammaness` > 0.5 (majority vote of the trees in the forest) has limited use, as one is only interested in a high fraction of gammas with a low fraction of BG. A more useful metric is the quality factor Q (see also Equation 1.2)

$$Q = \frac{\epsilon_\gamma}{\sqrt{\epsilon_{BG}}}, \tag{4.11}$$

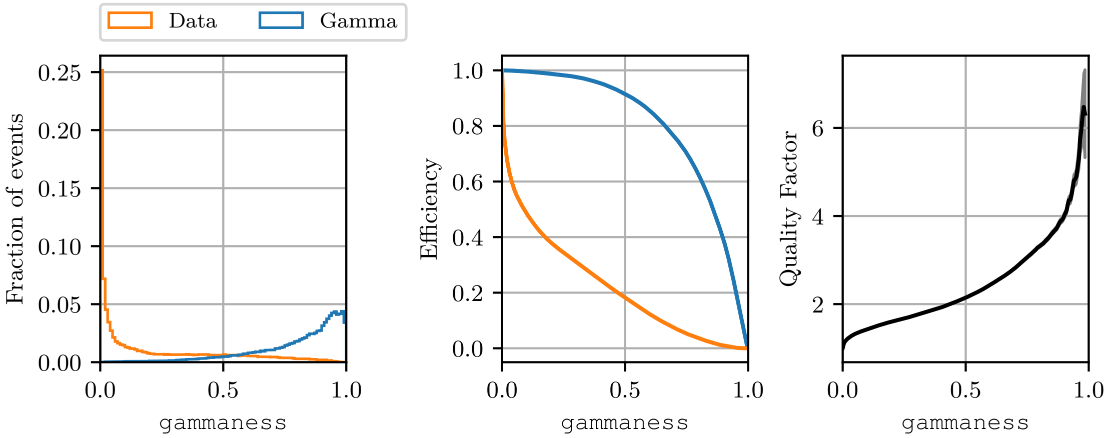
---

[6] $n\_trees = 50$, $max\_depth = 16$, $min\_samples\_split = 16$, $min\_samples\_leaf = 8$

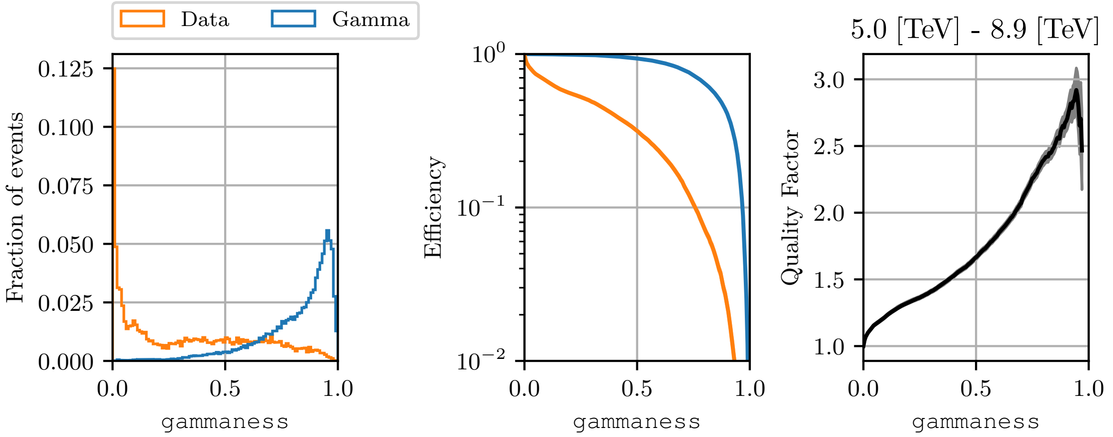
where ϵ_γ (ϵ_{BG}) is the gamma (BG) efficiency, the fraction of gammas (BG events) surviving to the total number of gammas (BG events). Q at a given ϵ_γ determines the separation power. In practice, the optimal values also depend on the expected fluxes of gammas and background, on the angular resolution and other parameters.

Figure 4.11a shows the `gammaness` histogram for the full set of data and gamma. It also shows ϵ and Q as a function of the `gammaness` cut value. Figure 4.11b and Figure 4.11c show examples at low and medium energies. Q increases with increasing energy and, equivalently, the ϵ_γ curve (blue) separates further from the ϵ_{BG} curve. The background distribution clearly accumulates at a `gammaness` close to 0, while the distribution of the gammas peaks close to 1. For a Crab-like source, the overall quality factor is $Q = 4.0$ at a cut of `gammaness` > 0.9. The corresponding background efficiency is about 0.95 % and the remaining gamma efficiency is 39 %.

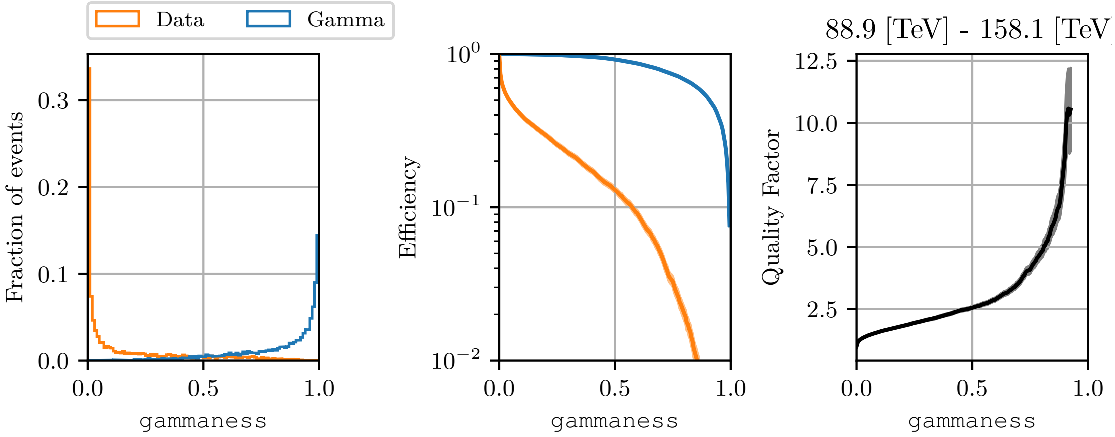
Figure 4.12 shows Q, ϵ_γ and ϵ_{BG} as a function of the estimated energy E_{est} for three `gammaness` cut values. With increasing energy, Q increases from about $Q = 1.5$ to $Q = 6$ at an estimated energy of ∼30 TeV. ϵ_γ reaches a minimum at ∼30 TeV from where it starts to increase towards higher energies. For ϵ_{BG}, the efficiency drops sharply until ∼30 TeV, from where it seems to rise slightly with increasing energy before dropping further. For higher estimated energies, Q was not estimated because of too few background events.



(a) Full test sample

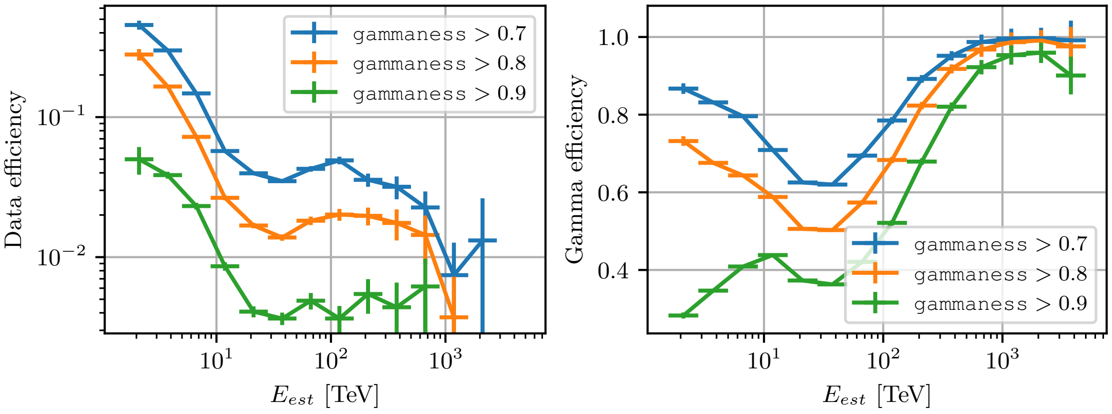


(b) Example at lower energies (5.0 TeV to 8.9 TeV)



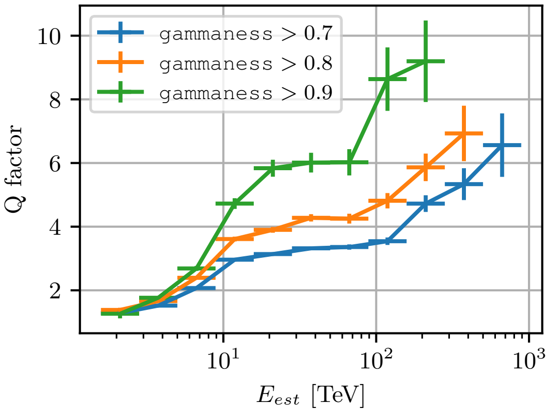
(c) Example at higher energies (88.9 TeV to 158.1 TeV)

Figure 4.11: Distribution, efficiency and quality factor against `gammaness` for the full test set and two example energy bins. The histograms on the left show the distribution of the simulated gamma and of the data (considered background). In the middle, the efficiencies for different `gammaness` cut values are shown and on the right, the corresponding Q-factors are shown. At about 0.9 the gamma efficiency is 39 % and the background efficiency is about 0.95 % giving a Q-factor of about 4.0. The Q-factors are only calculated if more than 10 background events remain after the cut.



(a) Background efficiency

(b) Simulated gamma efficiency



(c) Quality factor

Figure 4.12: (a) Background efficiency, (b) gamma efficiency and (c) quality factor for three `gammaness` cut values against the energy estimated with the Random Forest regressor. The quality factor is only calculated for entries with more than 10 surviving background events. The errors give the uncertainties estimated from the event statistics.

4.1.6 Effective Area

With origin reconstruction and gamma/BG-separation available, the expected point source effective area of the analysis can be calculated. The areas are calculated according to Equation 3.7, but with $\Omega_{sim} \rightarrow 1$ for the point source emission of the gamma rays

$$A_{eff,i} = \frac{h_i}{H_i} A_{sim}, \tag{4.12}$$

where h_i are the weighted counts in the true energy E_{sim} bin i surviving the reconstruction and H_i are the weighted sums of the simulated events corresponding to the conditions of h.

Figure 4.13 and Figure 4.14 show the effective area against the simulated energy for zenith angles of $40° < \theta < 50°$ and $10° < \theta < 20°$. Before `gammaness` and Θ cuts, the maximum is reached for energies of 200 TeV. After `gammaness` > 0.9 especially medium energy events are cut away, while for the largest energies, the gamma efficiency remains high. As apparent from the purple curves, the origin reconstruction seems to be systematically worse at higher energies, decreasing the effective area considerably. For a full hybrid analysis, no such degradation is expected. In fact, at ~100 TeV, HiSCORE has already a comparable angular resolution, improving even further with higher energies. Additionally, as also stated in Section 4.1.3, already at lower energies, for events where only a few and possibly even only one or two stations trigger, the angular resolution can be improved as investigated in Wertz [93]. It should also be possible to resolve the ambiguity of the direction reconstruction without the `M31` parameter.

Towards larger zenith distances, the effective area increases at higher energies and shrinks at low energies. This is due to the increasing airmass and due to the geometric effect, that the same distance to the telescope in the tilted shower plane corresponds to a larger distance on the ground. Additionally, because of the increased airmass, the showers reach X_{max} further away from the telescope and light from showers with larger impact parameters still fit into the camera. At lower energies, the light of the showers is not sufficiently bright to penetrate the increased airmass and a lower fraction of events is recorded. This also has a strong influence on the energy threshold of the telescope.

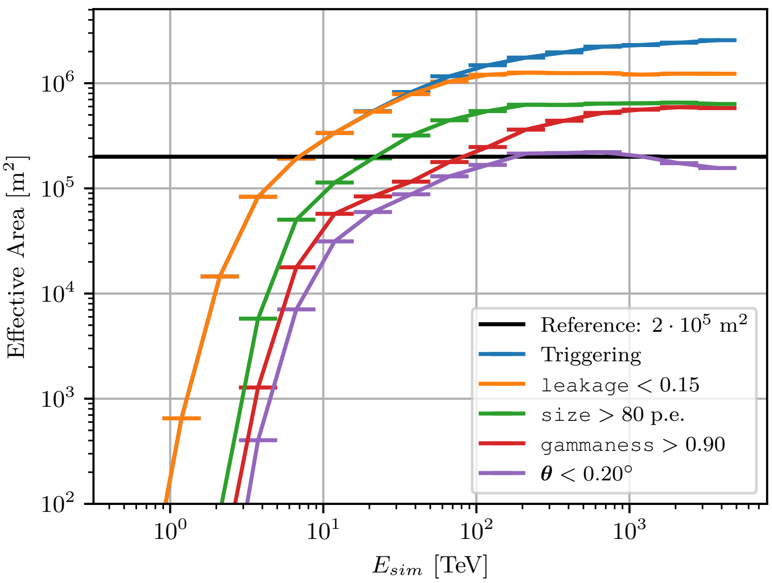


Figure 4.13: Effective Area of the Random Forest analysis with different cuts applied for zenith distances from $\theta > 40°$ to $\theta < 50°$. The blue curve shows the triggering area after cleaning. The other curves show the area after different additional analysis cuts.

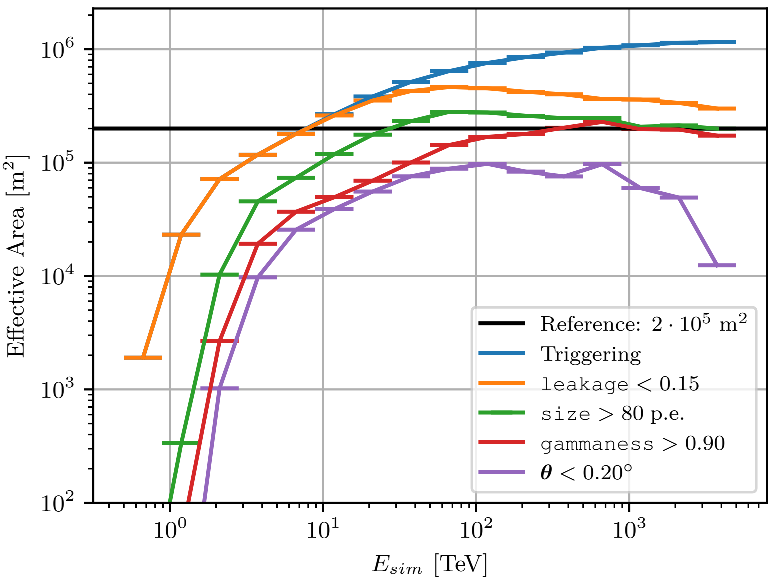


Figure 4.14: Effective Area of the Random Forest analysis with different cuts applied for zenith distances from $\theta > 10°$ to $\theta < 20°$. The blue curve shows the triggering area after cleaning. The other curves show the area after different additional analysis cuts.

4.1.7 Energy Threshold

The energy threshold of an IACT is defined as the true energy at which the highest event rate is expected. It is determined as the maximum of the histogram in E_{sim} of the events surviving the cuts. Figure 4.15 and Figure 4.16 show the histograms for zenith angles of $40° < \theta < 50°$ and $10° < \theta < 20°$. With an example cut of `gammaness` > 0.9 and $\boldsymbol{\theta} < 0.2°$, the respective thresholds for a primary spectrum $\propto E^{-2.8}$ are approximately 8 TeV to 10 TeV and ∼3 TeV. The trend of increasing threshold with stricter cuts is related to the fact that lower energy events with less light are harder to reconstruct and rarely reach higher `gammaness` values as shown in Figure 4.11.

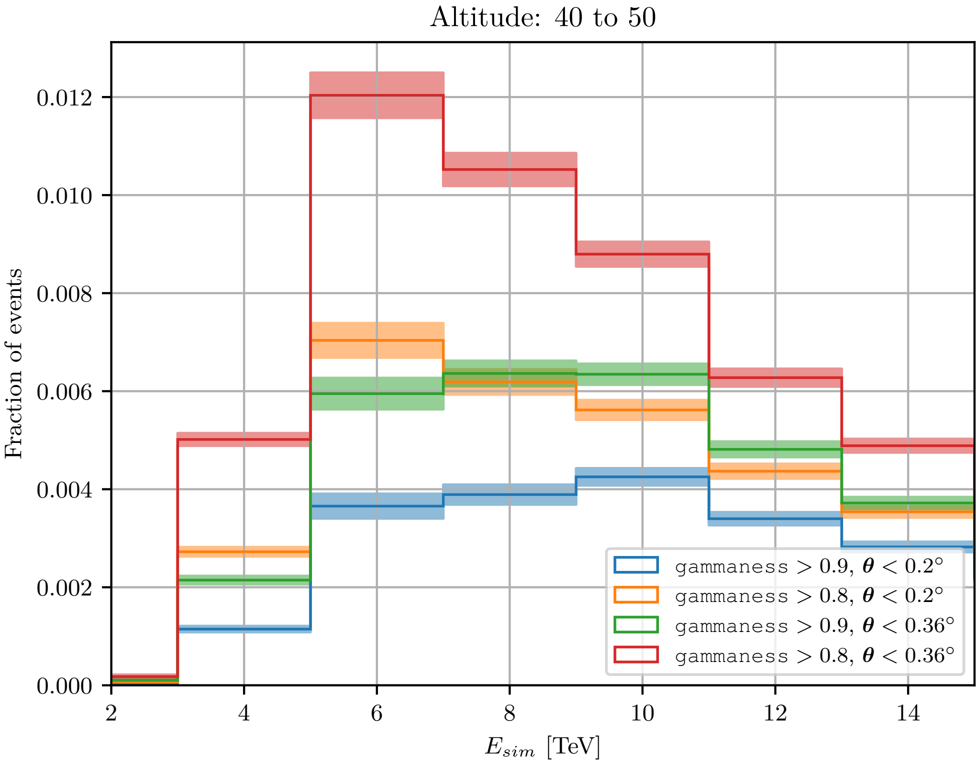


Figure 4.15: Histogram of the events surviving `gammaness` and $\boldsymbol{\theta}$ cut at zenith angles from $\theta > 40°$ to $\theta < 50°$ weighted to Crab-like power law spectrum $\propto E^{-2.8}$. The maximum of the histograms indicate the energy threshold.

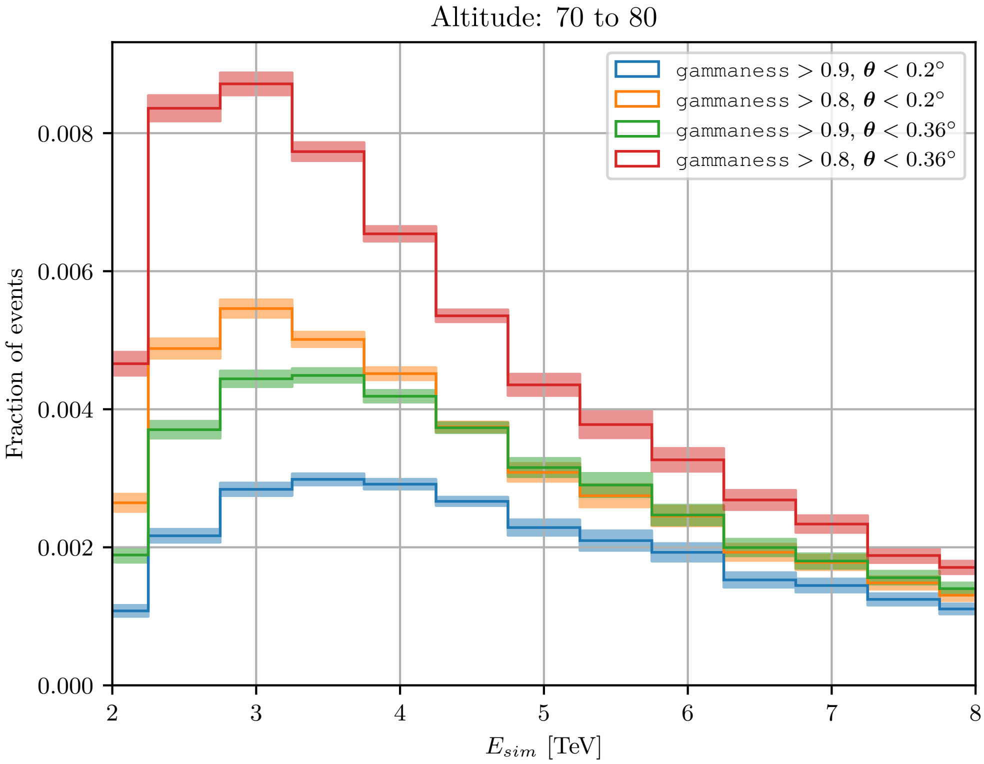


Figure 4.16: Histogram of the events surviving `gammaness` and $\boldsymbol{\theta}$ cut at zenith angles from $\theta > 10°$ to $\theta < 20°$ weighted to Crab-like power law spectrum $\propto E^{-2.8}$. The maximum of the histograms indicate the energy threshold.

4.2 HiSCORE

From the parameters introduced in Section 2.2.1 describing the waveform at each station, the impact position of the shower on the ground called core position (x_C, y_C) and the origin of the shower in zenith distance θ and azimuth Φ are reconstructed. The fits are done in python based on `reco_score` written by Hampf et al. [39] and Hampf [88] by fitting the arrival times (edge times) or amplitudes to models of the expected times or amplitudes at each station. The software was further improved by Porelli [37]. For this work, the required part of the implemented software was copied and an interface to the data stored in the binary HDF5 format was written. Furthermore, the code[7] was sped up by the more extensive use of numpy arrays and by optimizing the fit functions with numba[94], such that the reconstruction is 15 times faster without the use of parallelisation.

4.2.1 Core Reconstruction

The core of the shower is reconstructed via the calculation of the center of gravity, via a fit to the amplitude distribution function or via a fit to the lateral light density function.

4.2.1.1 Center of Gravity

A straight forward method to determine the impact position of the shower on the ground is the center of gravity of the signal recorded by the stations.

$$\vec{\mathrm{CoG}} = \frac{\sum_i A_i \vec{X}_i}{\sum_i A_i} \tag{4.13}$$

---

[7] `https://gitlab.rrz.uni-hamburg.de/TAIGA/andrea_software/`

4.2.1.2 Amplitude Distribution Function Fit

Another possibility is to fit the amplitude measured by the stations to the amplitude distribution function (ADF) empirically derived for the Tunka-133 experiment.[37, 95, 96]

It describes the amplitude at a station A with three independent parameters: The distance from the impact point R, the steepness b_A and the amplitude $A(200\,\mathrm{m})$ at $R = 200\,\mathrm{m}$. It is defined as

$$A(R) = A(200\,\mathrm{m}) f(R, b_A) \tag{4.14}$$

where f is split into four distinct regions

$$f(R, b_A) = \begin{cases} \left(\frac{R_{kn}}{200\,\mathrm{m}}\right)^{-b_2} exp\left(\frac{R_{kn}-R}{R_0}\left(1+\frac{3}{R+2}\right)\right), & \text{for } R \leq R_{kn} \\ \left(\frac{R}{200\,\mathrm{m}}\right)^{-b_2}, & \text{for } R_{kn} < R \leq 200\,\mathrm{m} \\ \left(\frac{\frac{R}{400\,\mathrm{m}}+a}{1+a}\right)^{-b_A}\left(\frac{0.5+a}{1+a}\right), & \text{for } 200\,\mathrm{m} < R \leq 400\,\mathrm{m} \\ \left(\frac{\frac{R}{400\,\mathrm{m}}+1}{2}\right)^{-b_A}\left(\frac{0.5+a}{1+a}\right), & \text{for } R > 400\,\mathrm{m} \end{cases} \tag{4.15}$$

using

$$\begin{aligned} d &= \begin{cases} \log_{10} 4, & \text{for } b_A > 9 \\ \log_{10} b_A - 5, & \text{for } 9 > b_A > 5.4 \\ \log_{10} 0.4, & \text{for } b_A \leq 5.4 \end{cases} \\ R_0 &= 10^{2.44-2d} \cdot 1\,\mathrm{m} \\ R_{kn} &= 145\,\mathrm{m} - d \cdot 115\,\mathrm{m} \\ a &= 0.89 - 0.29d \\ b_2 &= \begin{cases} 2.4 + 2(d-0.15), & \text{for } d \geq 0.15 \\ 2.4, & \text{for } d < 0.15. \end{cases} \end{aligned} \tag{4.16}$$

To fit this function, the amplitude of each station A_i is compared to $A(R_i)$ where R_i is the distance of that station (x_i, y_i) to the core location (x_C, y_C) in the shower plane. This gives four parameters to be optimized (x_C, y_C, b_A, $A(200\,\mathrm{m})$), requiring at least 5 triggering stations for the event.

4.2.2 Origin Reconstruction

The origin is at first coarsely reconstructed via fitting the arrival times to the expectation by a planar wavefront and then refined via fits to a curved wavefront and to the arrival time model.

4.2.2.1 Planar Light Front Fit

The first approximation is a plane wavefront

$$t_i^{PW} = \frac{1}{c}\left(x_i \sin\theta\cos\phi + y_i \sin\theta\sin\phi - z_i\cos\theta\cos\phi\right) + t_0 \tag{4.17}$$

with the speed of light c and a constant offset t_0. It has three free parameters: θ, ϕ, t_0.

4.2.2.2 Curved Light Front Fit

This method is also used in the Tunka-133 array and adds a curvature parameter to the arrival time parametrisation:

$$t_i = t_i^{PW}(\theta, \phi, t_0) + \frac{(R_i + 250\,\mathrm{m})^2}{c(4\times10^4\,\mathrm{m})}. \tag{4.18}$$

With the same free parameters as the plane wave fit, it also requires 4 stations in the event.

4.2.2.3 Arrival Time Model Fit

Another possibility is the description based on Stamatescu et al. [97], derived in Hampf [88] and extended to arrays which are not flat (z-component of the station position different for each station) in Porelli [37]. It assumes that the emission happens at a point of height Z and that the refractive index is proportional to the density following the barometric formula. The arrival times are parametrized with

$$t_i = \frac{1}{c}\sqrt{k}\left(1 + \frac{\eta_0 h_0(\exp\left(-z_i/h_0\right) - \exp -Z/h_0}{Z - y_i}\right) + t_0, \tag{4.19}$$

where

$$k = r_i^2 + \frac{Z^2}{\cos^2(\theta)} - 2Z(z_i - r_i\tan(\theta)\cos(\phi_i - \phi). \tag{4.20}$$

With the free parameters θ, ϕ, Z and t_0 it requires 5 stations in the event.

Chapter 5

Single IACT Results

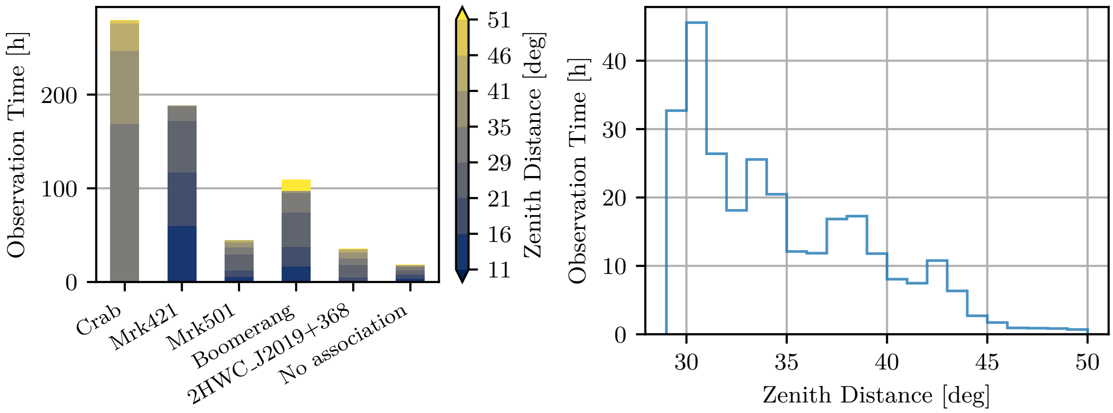
For the first IACT, the data of the 2019-20 and beginning of 2020-21 seasons was used to create a spectrum for the Crab Nebula up to 100 TeV. A more detailed description of the data used is given in Section 5.1. The selection of good weather is explained in Section 5.2. The spectrum is presented in Section 5.4 and an overview of the uncertainties is given in Section 5.5.

5.1 Available Data

For this analysis, data from the 2019/2020 and 2020/2021 seasons taken before the date of the addition of 5 hexagonal mirrors is used. This corresponds to the data taken from October 2019 to 2020-12-08. During that time, IACT 1 was taking data regularly. Figure 5.1a shows the available observation time per source. It is the time of data taken under all weather conditions and includes dead time. Of the total observation time of 657 h, 279 h of Crab Nebula data taken with IACT 1 are available, 188 h data on Mrk421 and 190 h on the remaining sources.

The Crab culminates at $\theta{\sim}30°$ and was observed in wobble mode[98], where the telescope pointed at two positions $\pm 1.2°$ in right ascension away from the position of the Crab (83.633° right ascension, 22.014° declination). Figure 5.1b shows the available observation time distributed over the observation zenith distance. About 70 % of the available data was taken below a zenith distance of $\theta = 40°$.

Because the observation also includes observation conditions that were not simulated (clouds, atmosphere changes, etc.), much of the time is not usable for analysis and only periods of good weather need to be selected.



(a) Observation time per source

(b) Crab observation time per zenith distance.

Figure 5.1: (b) The total available observation time per source including bad weather and dead time. (a) The observation time for the Crab Nebula as a function of the zenith distance

5.2 Good-Weather-Selection

Good weather is selected by cutting on the IACT event rate after cleaning and `size` cut of 80 p.e.. Applying a size cut and cleaning before calculating the rate cuts away small events that might be coincidence triggers due to NSB fluctuations or bright objects in the sky. Furthermore, the `size` > 80 p.e. cut is also used in the analysis to ensure a homogeneous CoG distribution in the camera and matching distributions between data and MC. The scaled rate is defined as

$$s_{80} = \frac{r_{80}}{\cos\theta}, \tag{5.1}$$

where θ is the zenith distance and r_{80} is the rate before scaling. With the scaling, one cut can be used for a range of different zenith distances. As the air mass increases with increasing distance, only the Cherenkov light of showers with higher energy or late first interaction reach the telescope and the rate is expected to decrease. For a plane-parallel atmosphere, the air mass scales with $\frac{1}{\cos\theta}$. Although the effect on the trigger rate is not expected to be linear, the scaling is good enough for the observed zenith distances of up to 60°.

Figure 5.2 shows a histogram of the mean rate per portion versus altitude. The red lines at 7.3 Hz and 12 Hz define the good weather selection for the portions. The bulk of the portions follows a linear trend and falls into the corridor defined by the lines. The bright spot at an altitude of 60° is caused by the culmination of the Crab Nebula. When visualizing the data as a histogram in s_{80}, two peaks are apparent. The blue curve in Figure 5.3 shows the portions recorded from October 2019 to 2020-12-08 (date of the installation of the additional hexagonal mirrors) with the exception of January, February and March. A fit of a normal distribution peaks at $\mu = (10.23 \pm 0.08)$ Hz and has a standard deviation of $\sigma = (1.0 \pm 0.1)$ Hz. The orange line shows the histogram of the mean scaled rate per portion for the three months of January 20, February 20 and March 20. For these portions, a fit of a normal distribution gives $\mu = (8.74 \pm 0.09)$ Hz and $\sigma = (1.3 \pm 0.1)$ Hz. The lower rate is caused by a trigger condition that requires three neighbouring pixels instead of two. The shoulder at ~10 Hz indicates that some of the portions in the three months were collected with the setting of two next neighbouring pixels. In order to also keep some data during the period with the three-pixel trigger, the cuts for the good weather portions are set to

$$7.3\,\mathrm{Hz} < s_{80} < 12.2\,\mathrm{Hz}. \tag{5.2}$$

In this range, it is assumed that the atmospheric conditions were sufficiently similar and that the camera electronics showed stable behaviour. At lower rates, bad weather is assumed and the higher rates are excluded as they could be caused by artificial light sources or very bright NSB conditions.

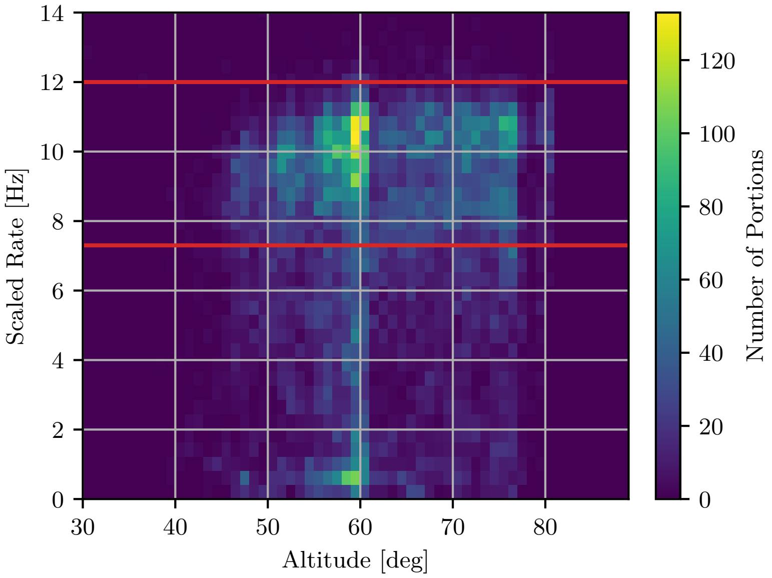


Figure 5.2: Histogram of the mean scaled rate of the portions over the mean altitudes. The data was taken between October 19 and December 20. One portion corresponds to ~120 s of observation time. The good-weather cut is defined as all portions contained between the red lines at 7.3 Hz and 12.2 Hz.

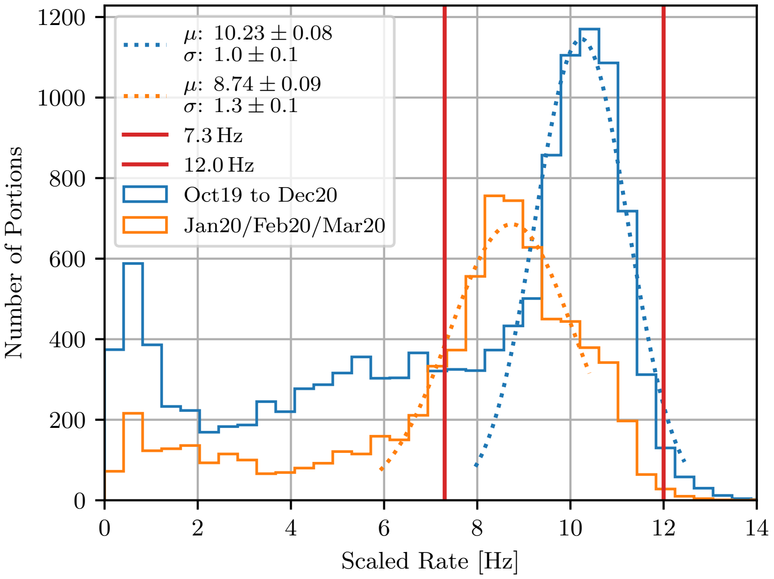


Figure 5.3: Histogram of the scaled rate per portion for two periods: Portions taken during January 2020, February 2020 and March 2020 are plotted in orange and portions taken during the remaining month between October 2019 and 2020-12-08 are shown in blue. The dashed lines correspond to fits of Normal distributions. The good-weather cut is defined as all portions contained between the red lines at 7.3 Hz and 12.2 Hz.

5.3 Crab Analysis

With the calibrated and quality-selected data, the high-level analysis is performed with Gammapy[99, 100]. In order to use Gammapy, the data is selected into continuous intervals with constant celestial pointing. These intervals are called observations and have an average duration of ~15 min. The corresponding high-level information is stored in the FITS file format (Gamma Astro Data Formats (GADF)-FITS format)[1]. For each of the observations, the Instrument Response Functions (IRFs) are calculated individually and converted to the GADF-FITS format with *pyirf*[101].

5.3.1 Instrument Response Functions

The gammas for the calculation of the IRF are selected according to the altitude of the observation. If the observation spans more than 5° in altitude, all gammas of the test and training set with matching altitude are selected. Otherwise, the covered altitude range is extended symmetrically until at least 5° are covered. Using at least 5° in altitude covered and both sets, the test and training set, ensures sufficient statistics for the IRF. The IRFs calculated with the test set were compared to the IRFs calculated with the training set, and the differences were minor, justifying the use of the full data for the calculation of the IRFs. In addition to the `leakage` < 0.15 and `size` > 80 p.e. cuts, the `gammaness` > 0.9 cut as used for the analysis is applied. The following IRFs are calculated:

Energy Migration The energy migration matrix $M_{i,j}$ is specified as a two dimensional histogram in bins j in $rel = \frac{E_{est}}{E_{sim}}$ and bins i in true energy E_{sim}. This is similar to the plot shown in Figure 4.8, but not subtracted by 1. The binning is 22 bins between 0.1 and 10 for rel and 24 logarithmic bins between 0.5 TeV and 5000 TeV. It is normalised along the rel dimension as

$$\sum_j M_{i,j} = 1. \tag{5.3}$$

PSF The point spread function $PSF_{i,j}$ is specified in the radial bins j of the angular separation between the reconstructed and true position $\boldsymbol{\theta}$ and in the true energy bins i. 60 bins between 0° and 4° are used for $\boldsymbol{\theta}$ and the bins in E_{sim} are the same as the ones for the energy migration matrix. The normalisation condition is

---

[1] `https://github.com/open-gamma-ray-astro/gamma-astro-data-formats`

$$\sum_j PSF_{i,j} \cdot \omega_j = 1, \tag{5.4}$$

where ω_j is the solid angle covered by bin j:

$$\omega_j = 2\pi \left(\cos \boldsymbol{\theta}_{min,j} - \cos \boldsymbol{\theta}_{max,j}\right), \tag{5.5}$$

with the edges of the bin $\boldsymbol{\theta}_{min,j}$ and $\boldsymbol{\theta}_{max,j}$.

Effective Area For the effective area $A_{eff,i}$, the same bins i in true energy E_{sim} as for the energy migration and PSF are used. See Section 4.1.6.

5.3.2 Source and Background Regions

All events reconstructed to a position inside a circle with a radius of $\boldsymbol{\theta}_s$ around the position of the source are considered signal events. In this analysis, no optimisation was performed and the value was set to $\boldsymbol{\theta}_s = 0.2°$, the 68 %-containment value at 10 TeV to 50 TeV.

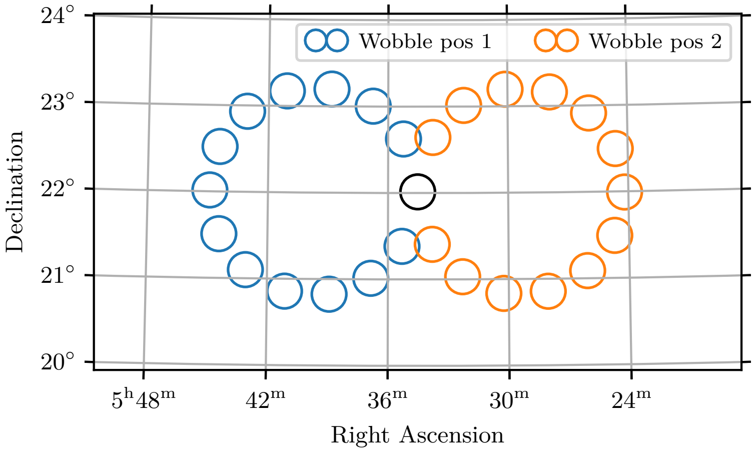


Figure 5.4: Placement of the analysis regions. The center is the on source region with a radius of $\boldsymbol{\theta}_s = 0.2°$. The circles on the two arcs to the right and to the left are the off regions positioned at the same distance to the two wobble pointing positions of the telescope.

In order to estimate the expected background in the source region, 13 background regions of the same size as the source region are aligned on an arc with the same distance to the center as the source position (called reflected regions in Gammapy). On this arc, the average acceptance is expected to be close to the acceptance of the on source region.

Figure 5.4 shows the placement of the regions in galactic coordinates. There are two visible arcs, one around each of the two wobble positions.

To visualize a possible excess around the source region, a histogram in bins of the squared angular distance $\boldsymbol{\theta}^2$ between the reconstructed event positions to the center of the regions is common. Each of the bins covers the same angular area. For the background positions, a mostly flat distribution is expected, while the source should show as a peak at values close to 0. Figure 5.5 confirms an increase in counts N_{on} for the on region. The background counts N_{off} are scaled by the ratio of the acceptances $\alpha = \frac{1}{13}$ between on and off and show a slight increase of ~10 counts towards small values of $\boldsymbol{\theta}^2$. At values above $\sim 0.06\,{}^{\circ 2}$, the on and background distributions overlap, indicating that the average acceptance of the off regions matches the acceptance of the on region. The plotted range up to a radius of approximately $\boldsymbol{\theta} = 0.6°$ is located well within the FoV of the camera where its acceptance should be mostly flat. Therefore, the observed decreasing trend of the off counts is slightly larger than expected. One possible cause is that the background suppression trained with gammas coming from a source with an offset of 1.2° from the camera center favours the events which end up being reconstructed to origin positions close to that circle at 1.2°, explaining the decrease of events with larger $\boldsymbol{\theta}$. Another reason could be the systematic reconstruction bias with the `disp` method to these positions. This is however less likely, as the `disp` Random Forest is trained with a diffuse set to avoid this bias and the reconstructed positions before background suppression are distributed rather homogeneously.

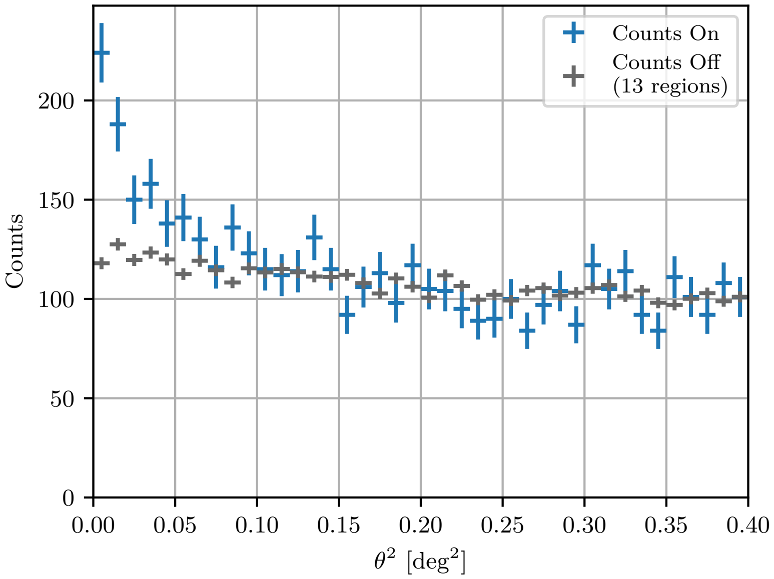


Figure 5.5: Counts around the position of the Crab Nebula N_{on} and around the off positions αN_{off} as a function of the angular distance to the region centers $\boldsymbol{\theta}$. The 13 off regions give an acceptance ratio of $\alpha = \frac{1}{13}$. The errors are estimated assuming a Poisson distribution of the counts.

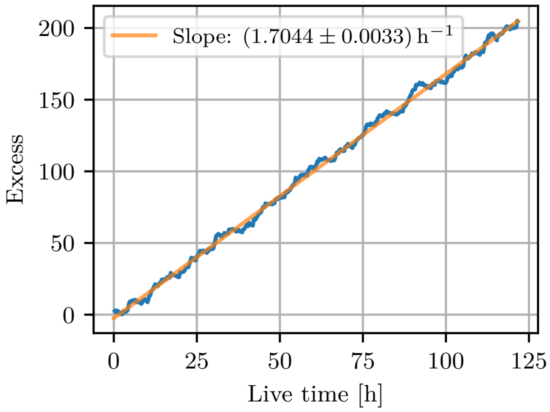
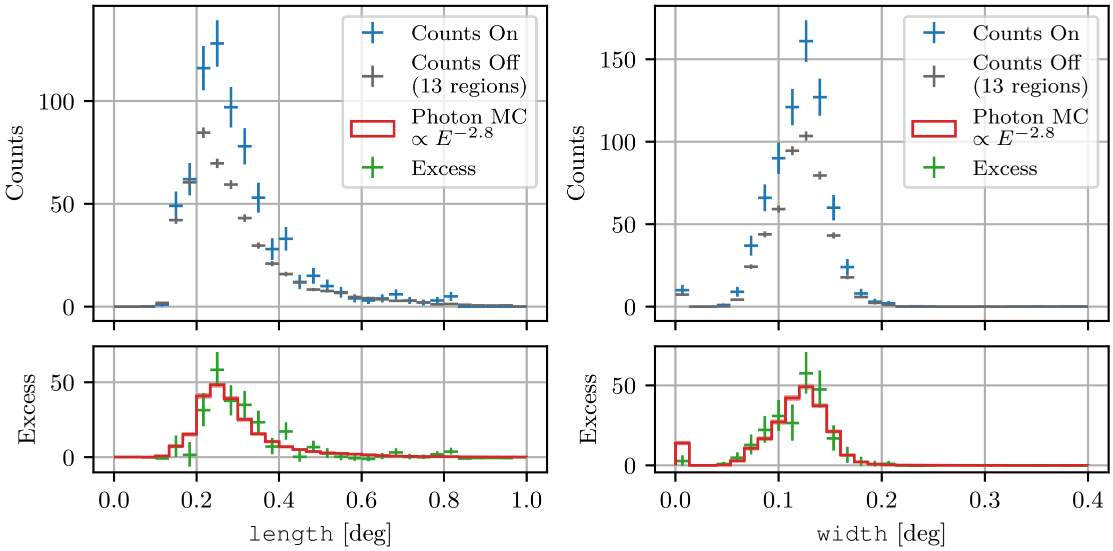


Figure 5.6: Excess $N_e = N_{on} - \alpha N_{off}$ as a function of the live time with $\alpha = \frac{1}{13}$. The orange line is the result of a linear fit to the data, giving a slope of $(1.7044 \pm 0.0033)\,\mathrm{h}^{-1}$.



(a)

(b)

Figure 5.7: Counts of the on position N_{on} and scaled counts of the off positions αN_{off} with $\alpha = \frac{1}{13}$ as functions of (a) the `length` and (b) the `width`. The lower panel shows the excess events $N_e = N_{on} - \alpha N_{off}$ with gamma simulation superimposed. The simulation is taken at a zenith distances between 30° and 40° and weighted to a primary spectrum of $\propto E^{-2.8}$. It is scaled such that the sum of the weights matches the sum of the excess events of the histogram.

To calculate the counts in the region, only the events in the estimated energy bins are counted where the effective area in true energy is larger than 5% of the maximum effective area (Gammapy SafeMaskMaker).

This results in a total of $N_{on} = 553$ events for the on-source and $N_{off} = 4524$ for the 13 background regions. Using the ratio of the acceptances in the on and off regions $\alpha = \frac{1}{13}$, this gives

$$N_e = N_{on} - \alpha N_{off} = 204.6 \tag{5.6}$$

excess events N_e. Figure 5.6 shows the accumulation of excess events during the live time of the telescope. The accumulation follows a linear trend with ∼1.7 excess events per hour.

Those excess events should also appear in the distribution of `width` and `length`. Figure 5.7 shows the histograms for the on and off positions. The lower panels show a good match of the excess with the gamma simulation expectation. The weights of the gamma sample are scaled to sum to the total number of excess events. The errors of the excess counts in the lower panels are estimated by finding values of μ_s, where $W(N_{on}, N_{off}, \alpha, \mu_s) - W(N_{on}, N_{off}, \alpha, N_{on})$ (Equation 5.8) changes by 1.

5.3.3 Significance Determination

To determine, how significant the detection is, a likelihood ratio test is used. Assuming a Poisson distribution of the on and background counts, a test statistic TS is defined based on a likelihood ratio of the likelihood with the on source signal and the likelihood with only the background signal as first described in Li and Ma [34] and based on the theorem in Wilks [102]. A description of the implementation in Gammapy can also be found in Naurois [103] and in the documentation[2].

Writing the expected background signal as μ_b and the expected counts from the source as μ_s, the expected signal from the on region is given as $\mu_s + \alpha\mu_b$ and the likelihood to observe N_{on} and N_{off} is

$$L(N_{on}, N_{off}, \alpha, \mu_s, \mu_b) = \frac{(\mu_s + \alpha\mu_b)^{N_{on}}}{N_{on}!} e^{-\mu_s - \alpha\mu_b} \frac{(\mu_b)^{N_{off}}}{N_{off}!} e^{-\mu_b}. \tag{5.7}$$

Taking $-2 \ln L$ and neglecting the constant terms only depending on N_{on} and N_{off} yields for the general case

$$W(N_{on}, N_{off}, \alpha, \mu_s, \mu_b) = 2\left(\mu_s + (1+\alpha)\mu_b - N_{on} \ln\left(\mu_s + \alpha\mu_b\right) - N_{off} \ln\left(\mu_b\right)\right). \tag{5.8}$$

In Gammapy, if there is no model available for μ_b, an analytical estimate is calculated using the profile likelihood by solving $\frac{d \ln L}{d\mu_b} = 0$. The resulting estimate is

$$\mu_b = \frac{C + \sqrt{C^2 + 4(\alpha+1)\alpha N_{on} N_{off}}}{2\alpha(\alpha+1)} \tag{5.9}$$

with

$$C = \alpha(N_{on} + N_{off}) - (\alpha+1)\mu_s. \tag{5.10}$$

---

[2] `https://docs.gammapy.org/1.0.1/user-guide/stats/wstat_derivation.html`

Then, the likelihood ratio for the significance determination is calculated by assuming only background $\mu_s = 0$ as the null hypothesis and expecting $\mu_s = N_{on}$ as the alternative hypothesis:

$$TS = W(N_{on}, N_{off}, \alpha, 0) - W(N_{on}, N_{off}, \alpha, N_{on}). \tag{5.11}$$

Using the theorem in Wilks [102] and following the argument in Li and Ma [34], TS is approximately $\chi^2(1)$ distributed if the null hypothesis (only BG) is true and if N_{on} and N_{off} do not have too few entries. Then, the significance of the alternative hypothesis (signal from source) is directly given as

$$S = \sqrt{TS} = 9.7. \tag{5.12}$$

This method to estimate the significance is equivalent to Equation 17 in Li and Ma [34]. Figure 5.8 shows the value of $\sqrt{TS}$ as a function of the observation live time. It follows a $\propto \sqrt{t}$ behaviour.

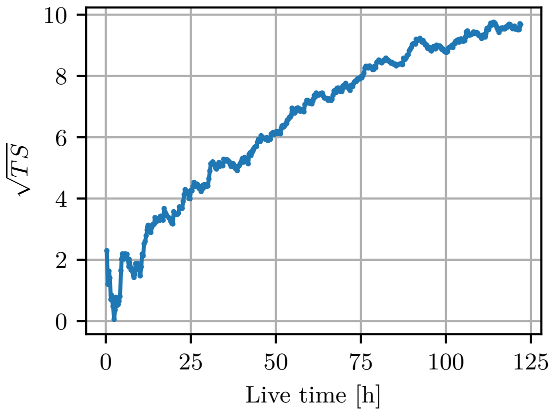


Figure 5.8: Significance $S = \sqrt{TS}$ as a function of the live time.

5.3.4 Spectral Fit

When the signal from the source is established, a spectral fit can be carried out. In this work, two models are used. The power law

$$\phi(E) = \phi_0 \cdot \left(\frac{E}{E_0}\right)^{-\Gamma}, \tag{5.13}$$

with the flux normalisation ϕ_0, the reference energy E_0 and the spectral index Γ. And the log parabola

$$\phi(E) = \phi_0 \left(\frac{E}{E_0}\right)^{-\alpha-\beta\log\left(\frac{E}{E_0}\right)}, \tag{5.14}$$

with the two parameters α and β.

For the fitting, a forward folding technique is employed. For each bin in $E_{est,k}$, the expected number of signal counts $\mu_{s,k}$ is determined with a convolution of the expected flux ϕ with the IRF in bins l of true energy $E_{sim,l}$ with the edges of the bin $E_{min,l}$ and $E_{max,l}$. The distinction between i, j (Section 5.3.1) and l, k of Equation 5.15 is important, as the binning used for the calculation of the IRF is not equivalent to the binning used in the reconstruction. $\mu_{s,k}$ is calculated as

$$\mu_{s,k} = \sum_l \overline{D}_{l,k}\epsilon_l \int\limits_{E_{min,l}}^{E_{max,l}} \phi(E)dE, \tag{5.15}$$

where $\overline{D}_{l,k}$ is the average energy migration matrix of all observations and ϵ_l is the total exposure. These average values are determined in a step called stacking, where the observations m are combined according to their respective exposure. The total exposure ϵ_l is the sum of the exposures ϵ_l^m of each observation

$$\epsilon_l = \sum_m \epsilon_l^m = \sum_m t_m A_{eff,l}^m \sum_{j=0}^{\boldsymbol{\theta}_s} PSF_{l,j}^m \omega_j, \tag{5.16}$$

with the livetime of the observation t_m and the containment value of the $PSF_{i,j}^m$ (Equation 5.4) at the radius of the on source region $\boldsymbol{\theta}_s$ (ω_j is the solid angle of the bin of the PSF, see Equation 5.5). The interpolation from the energy in bins i to bins l and the linear interpolation at the solid angle cut $\boldsymbol{\theta}_s$ is done on the cumulative version of the PSF (along the angular axis $\boldsymbol{\theta}$ in bins j). The effective area $A_{eff,l}^m$ at the simulated energy bin l of the observation m is determined from the effective area in the binning $A_{eff,i}^m$ with linear interpolation in the logarithm of the energy. The average energy migration matrix is calculated as the exposure weighted average

$$\overline{D}_{l,k} = \frac{\sum_m D_{l,k}^m \epsilon_l^m}{\epsilon_l} \tag{5.17}$$

of the migration matrices $D_{l,k}^m$ of the observations m, which are determined from the energy migration $M_{i,j}^m$. For clarity, the superscripts m, indicating that the values change for every observation, are left out in the following paragraph. First, the values $M_{i,j}$ are rebinned to $M_{l,j}$ with linear interpolation in logarithmic energy. Then, the bins in $E_{est,k}$ are constructed from the relative energy bins $\left(\frac{E_{est}}{E_{sim}}\right)_j$ by taking the linear interpolation of the cumulative sum $I_{l,r} = \sum_{j=0}^{r} M_{l,j}$ at the positions of the energy edges $L_{l,k} = I\left(\frac{E_{min,k}}{E_{mid,l}}\right)$ and $U_{l,k} = \left(\frac{E_{max,k}}{E_{mid,l}}\right)$ and taking the difference between these values $D_{l,k} = U_{l,k} - L_{l,k}$.

The exposure weighting ensures that the energy migration corresponds to the sum of the counts of the observations:

$$N_{on,k} = \sum_m N^m_{on,k} \tag{5.18}$$

and

$$N_{off,k} = \sum_m N^m_{off,k}. \tag{5.19}$$

For each of the models $\phi(E)$, the parameters are varied and the total sum W_{tot} of the negative log-likelihoods W (Equation 5.8) for each of the bins with the expected number of counts according to Equation 5.15 is minimized

$$W_{tot}(\phi) = \sum_k W(N_{on,k}, N_{off,k}, \alpha, \mu_{s,k}(\phi)). \tag{5.20}$$

For the power law, the result is $\Gamma = 2.97 \pm 0.15$ and $\phi_0 = (4.3 \pm 0.5) \times 10^{-10}\,\mathrm{m^{-2}\,s^{-1}\,TeV^{-1}}$ at a reference energy of $E_0 = 11\,\mathrm{TeV}$.

A log-parabola fit was also tested, but was not preferred. The resulting test statistics difference of $\Delta TS = 4.58$ between both models corresponds to a p-value of 0.032 when assuming a χ^2-distribution with one degree of freedom.

The expectation of the power law fit and the excess counts are plotted against the estimated energy in Figure 5.9, indicating a consistent fit.

5.3.5 Flux Point Determination

Flux points are estimated with the Gammapy FluxPointsEstimator[3] as described in [104] and are specified in bins of estimated energy E_{est}. For every flux point, one or more bins $E_{est,k}$ are selected. The minimum and maximum energy of the range of the bins are E_1 and E_2. To estimate the flux in the energy range between E_1 and E_2, the best fit model parameters are fixed to the value at the logarithmic mid point E_{ref} of the energy range. Then, a relative normalisation is introduced and the likelihood fit is conducted with only the relative normalisation as a free parameter. Similarly as in the case of the significance calculation, this allows the definition of a test statistic difference TS between the best fit model and the model that has the relative normalisation as a free parameter. TS is again $\chi^2(1)$ distributed and $\sqrt{TS}$ corresponds to the significance S of the best fit model over the model with another norm. The obtained differences

---

[3] `https://docs.gammapy.org/0.19/api/gammapy.estimators.FluxPointsEstimator.html`

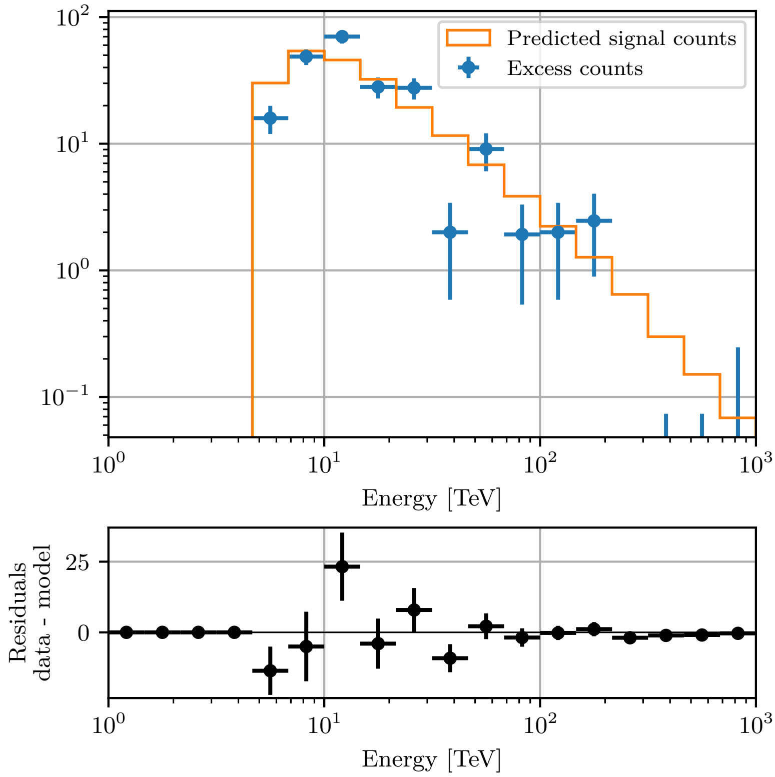


Figure 5.9: Raw excess events per simulated energy of the Crab Nebula observation. The predicted counts of the best fit power law are shown in orange. The lower plot visualizes the residuals between predicted and excess counts.

in the log-likelihood for different values of the normalisation can be used to find the normalisation values at which the $S = 1\sigma$ positive and negative deviations lie. If the significance of a flux point is $S < 2$, the upper limits are assumed at 2σ positive deviation from the best fit value of the negative log-likelihood W.

The resulting flux points for the observations are listed in Table 5.1 and plotted in Figure 5.10. The lowest energy bin ranges from 4.64 TeV to 10 TeV and has a significance of $\sqrt{TS} = 4.75$, while the highest flux point is between 46.4 TeV and 100 TeV and still has 11 excess events at a significance of $\sqrt{TS} = 2.23$. At higher energies, two upper limits are calculated.

5.4 Summary of Crab Spectrum

Of the signal from the source established with a significance of $S = 9.7$ at $N_e = 204.6$ excess events, the best fit power law model $(4.30 \pm 0.50) \times 10^{-10}\,\mathrm{m}^{-2}\,\mathrm{s}^{-1}\,\mathrm{TeV}^{-1} \cdot (E/11\,\mathrm{TeV})^{-2.97\pm0.15}$ and the flux points are shown in Figure 5.11. For comparison, a model fit to several VHE datasets as derived in Dirson and Horns [26] is also plotted. The

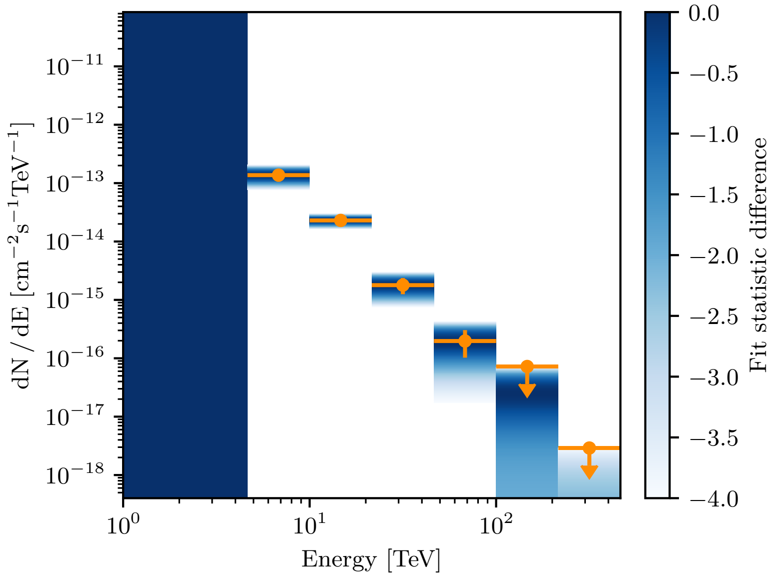


Figure 5.10: Flux points of the Crab Nebula as determined with a forward folding fit implemented in Gammapys FluxPointsEstimator. The color scale indicates the change of the test statistic TS of a model with only the norm as a free parameter. The model is constructed by multiplying a normalisation parameter to the value of the best fit power law in the energy range of the bin.

E	E$_1$	E$_2$	flux	error +	error −	$\sqrt{TS}$	μ_s	N_{on}
TeV			10^{-15} TeV^{-1} cm^{-2} s^{-1}					
6.813	4.64	10.00	136.2	32.37	31.0	4.75	64.37	213
14.678	10.00	21.54	22.91	3.57	3.41	7.74	98.00	215
31.623	21.54	46.42	1.850	0.58	0.54	3.76	30.61	82
68.129	46.42	100.00	0.198	0.11	0.095	2.23	11.00	30
146.780	100.00	215.44	-	< 0.02	-	1.40	4.378	12
316.228	215.44	464.16	-	< 0.003	-	-1.42	-2.231	1

Table 5.1: Flux points as determined with a forward folding fit implemented in Gammapys FluxPointsEstimator for a source flux model with $4.30\,\mathrm{m^{-2}\,s^{-1}\,TeV^{-1}} \cdot (E/11\,\mathrm{TeV})^{-2.97}$. The obtained $\sqrt{TS}$ gives the significance of that bin. μ_s is the number of predicted signal events (excess) for that bin and N_{on} is the number of counts in the on region.

VHE datasets were scaled to match the Fermi flux and the model with the radial magnetic field dependence $\propto r^{-\alpha}$ downstream of the shock with $\alpha = 0.51$ is shown. Among the datasets, the observation by LHAASO[13] between 2019-12-27 and 2021-02-28 was partly contemporaneous.

The obtained spectrum agrees well with previous measurements. For the lowest flux point, the largest deviation to the model is seen. This is also the flux point where the largest systematic uncertainty is expected as it is at the energy threshold of the analysis.

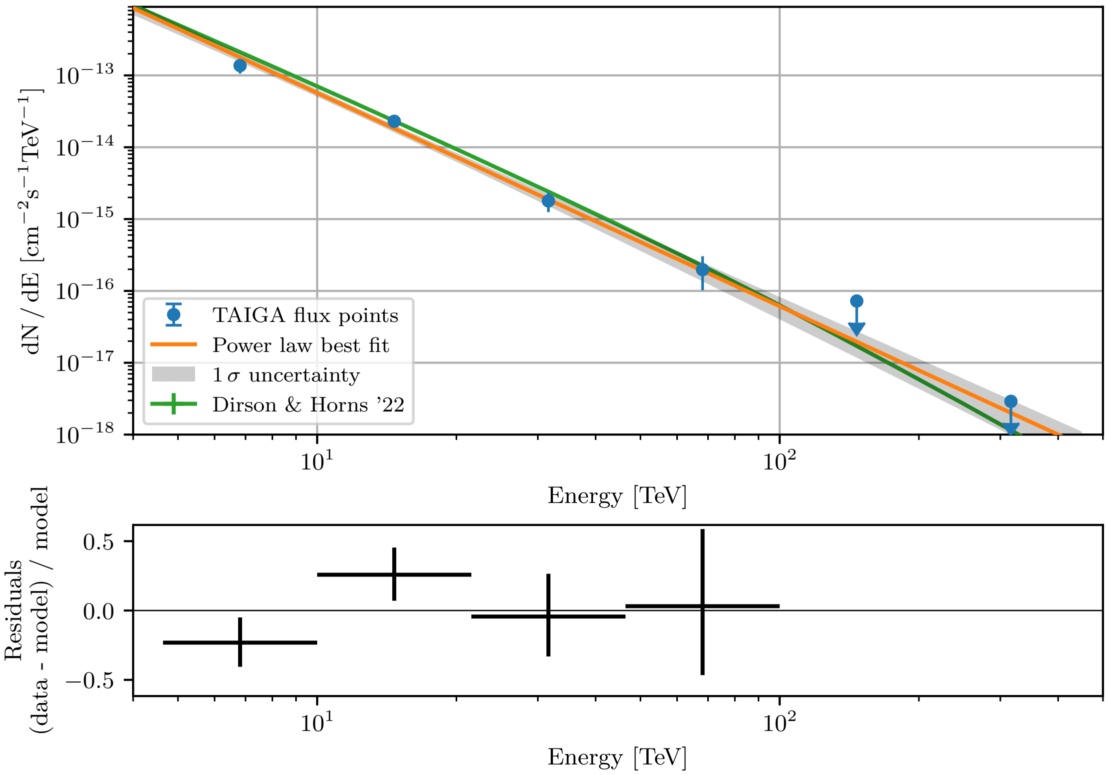


Figure 5.11: Spectrum of the Crab Nebula obtained with the Random Forest analysis of the first TAIGA IACT. The total live time after good-weather selection is about 122 h. The orange line is the best fit power law. The $1-\sigma$ statistical error is shown in grey. The flux points resulting from a forward folding fit assuming the best fit power law are shown in blue. For comparison, a model fit to data of several experiments in [26] is shown in green. The lower plot shows the difference of the flux points to the power law fit (orange).

5.5 Systematic Uncertainties

The assessment of the systematic uncertainty of the telescope is an extensive subject, and many points cannot be fully addressed in this work. However, a brief estimate of the systematic uncertainties is attempted. Section 5.5.1 explores the consistency of the results when analysing various subsets of the data. An approximate uncertainty resulting from a change in the total transmission of the telescope is estimated in Section 5.5.2 and other potential sources of uncertainties which were not quantified are given in Section 5.5.3.

5.5.1 Consistency in Subsets

To check if a systematic effect is present in different subsets of the data, the model fit result and the count statistics are assessed for two splits: (1) First and second half of the observation period and (2) for each of the wobble positions. The first half of the

Split	Γ	ϕ_0 10^{-15} $\mathrm{TeV^{-1}\,cm^{-2}\,s^{-1}}$	cov_{11}	N_{on}	N_e	$\sqrt{TS}$	t h
Full set	2.97 ± 0.15	4.30 ± 0.50	0.061	552	205.62	9.67	121.6
First half	2.82 ± 0.17	4.57 ± 0.72	0.246	260	102.92	7.17	59.5
Second half	3.28 ± 0.28	3.98 ± 0.70	-0.274	292	101.96	6.54	62.1
Wobble 1	2.86 ± 0.24	3.61 ± 0.70	0.174	255	85.00	5.81	58.8
Wobble 2	3.06 ± 0.20	4.90 ± 0.71	-0.035	297	119.62	7.83	62.8

Table 5.2: Results for various splits of the Crab data. Normalization ϕ_0 and photon index Γ of a power law fit $\phi_0 E^{-\Gamma}$ are listed together with the off-diagonal element of the covariance matrice c_{11}. Additionally, the overall number of counts in the on region N_{on}, the number of excess events N_e, the resulting significance $\sqrt{TS}$ and the corresponding live time t are given. The splits considered are the first half of the data (October 19 to January 20), the second half of the data set, and the data taken in each of the two wobble positions (±1.2° in right ascension).

observations span 59.5 h from October 19 to 2020-01-09 and the second half of the data comprises 62.1 h between 2020-01-09 and 2020-12-08. Of the wobble position splits, the first position (+1.2° in right ascension) has a total live time of 58.8 h and the second wobble position (−1.2° in right ascension) has 62.8 h. The results are listed in Table 5.2.

For all splits, the best-fit values for Γ and ϕ_0 are compatible within their respective estimates of the 1σ statistical error. The first and second wobble positions give $85.0^{+16.7}_{-16.1}$ and $119.6^{+17.9}_{-17.3}$ excess events N_e, respectively. With a difference of 34.6 counts, the values seem far apart; however, when considering the live time difference of 4 h, they are compatible within their respective errors. The errors are calculated by finding the value of μ_s, where $W(N_{on}, N_{off}, \alpha, \mu_s) - W(N_{on}, N_{off}, \alpha, N_{on})$ (Equation 5.8) changes by 1.

No inconsistency between the subsets was found.

5.5.2 Optical Transmission

The main component of the systematic uncertainties in the IACT analysis is the accurate description of the telescope and the atmosphere in the simulation. The first obvious step is the atmosphere itself, where factors like changing weather conditions are not fully addressed in the simulation and only minimized by using the good-weather cut. Furthermore, the varying profiles of the atmosphere lead to systematic changes in the development of the shower and especially in the optical transmission of the atmosphere. In Section 3.1.2 some properties of the profiles and of the optical transmission are discussed. Leaving out the systematic shift in X_{max} and only looking at the optical transmission, the uncertainty associated with the atmosphere used in this work is estimated to be approximately 10 % to 15 %.

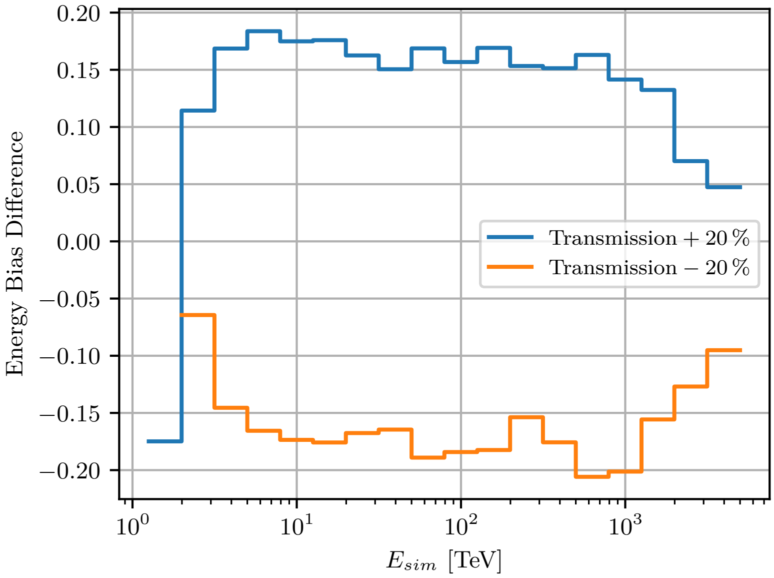


Figure 5.12: Change of the energy reconstruction bias (median) at ±20 % change in optical transmission against the simulated energy of the Random Forest Regressor for gammas at zenith distances from 31° to 41°.

The next components with their respective estimated error (educated guesses partly based on collaboration internal communication, no measurements were made) are the mirror reflectivity of ∼8 %, the plexiglass cover transmission ∼7 %, the Winston cone transmission of approximately 8 %, and the quantum efficiency (+collection efficiency) of the PMTs of ∼15 %.

The total squared sum of the estimated uncertainties of the optical transmission is 22.4 %. To study the influence of a change of ±20 % on the energy estimation of the Random Forest regressor, `sim_telarray` was run for both cases. Figure 5.12 shows the change in energy bias over the simulated energy. As is evident, for a change of 20 %, the change in bias is between 15 % and 18 % for energies of 10 TeV to 800 TeV. For −20 % transmission, the bias changes between −15 % and −20 % for energies of 10 TeV to 800 TeV.

Energies above 1 PeV are not suitable for analysis due to the systematic effect of the energy regression explained in Section 4.1.4. For energies below 10 TeV, the systematic effects are harder to assess, as this is in the range of the energy threshold where effects of the telescope electronics are of increased importance.

Assuming a power law $\propto E^{-3.0}$, a systematic energy bias of ±17 % corresponds to a change in flux at a given energy of 75 % and −37 %.

This value falls into the range also seen in other experiment. For example, MAGIC estimated an energy scale uncertainty of 15 % to 17 %[105] in a more detailed investigation.

5.5.3 Other Uncertainties

Further uncertainties might arise from the telescope electronics and from the choice of the simulation parameters. Some potential sources which were not investigated in detail are listed here. It is not a complete list of all potential sources.

Only one NSB level for the entire dataset: The simulation is processed with only a single NSB level for the whole period. However, higher levels of NSB increase the noise in the image and have a systematic effect on the distributions. Since it would cause the images to get wider and longer and therefore appear more like background at the training NSB level, it would probably lead to a higher fraction of rejected gammas.

PSF and focus of the camera: As described in Section 3.2.1, the PSF and camera focus settings were adjusted to give a good fit of the overall distributions, as no measurement of the PSF in the full camera plane is available. If the PSF in the simulation were completely off, the comparison of the distributions from simulation to data would not match at all. Some systematic shift of the distributions of the events could, however, be possible. It could be possible that if the PSF is overestimated in the simulation, the background contamination would increase, and if it is underestimated, the gamma fraction would increase. The camera focus position is set to ∼800 m in the simulation and it is possible that its closer to infinity in the real experiment. If that holds, a systematic effect on the image for showers at different distances from the telescope might be expected. This would potentially cause a systematic shift of the distributions for higher energies and close core distances (more light from closer to the telescope).

Cluster structure: Missing cluster in event: It can happen that a cluster has a pixel above threshold just before an actual event occurs. It is then in dead time when the event arrives and only the neighbouring cluster triggers and is read out. Even though the local dead time is small, it is not accurately known how the interplay between the central controller and the MAROC boards operates. Wrong amplitude in a cluster: If a cluster sends an early hold signal due to NSB, it is still counted in the event, but the amplitude is off. Although this also occurs in the simulation, it could introduce a systematic shift of the distributions and of the IRF if it is not correctly simulated. Events affected by both cases might get reconstructed to the wrong position and with a wrong energy. In the best case, these events do not survive the analysis and can be neglected.

Non-linearity of the PMT readout: The MAROC board has a slight non-linearity (on the order of 4 %) which is not simulated.

PMT gains / calibration: The calibration was only done at the beginning of the observation period. During the time of the observation, they may have changed considerably. Furthermore, the variation of the gains seen in the experiment could not be met in the simulation for both gains, as they are assumed to be directly linked together. On the MAROC board however, the low gain channel takes the anode signal with no amplification, and the high gain has an adjustable amplifier, allowing a more homogeneous gain setting for the high gain channel. (See Figure 3.9 in Section 2.3.2) Therefore, images with a high amplitude in many pixels might show offset image parameter distributions. This could be especially true if many of the pixels are above low gain saturation, as the spread of the maximum amplitude per pixel is larger in the experiment than in the simulation.

MAROC internal trigger discriminators: In `sim_telarray`, the trigger is simulated as a true comparator response without delays. However, in the real experiment, the MAROC board internal discriminator will not trigger if a signal is infinitely short above the set threshold. The trigger characteristics of the MAROC board should be measured more precisely. This should not have a large impact as the `size` > 80 p.e. cut should shift the effective threshold to higher values.

Inacurate SPE response curve: The SPE response curve was measured with the PMT on a different base and without a Winston cone attached (see Section 2.3.2). However, the value of the HV, the setting of the voltage divider, and the impact angles on the photocathode have an influence on the low-amplitude events (below 0 and the SPE peak) and on the afterpulsing probability. Low-amplitude signals as well as afterpulses can have an effect on the effective trigger probability as pointed out in Mirzoyan et al. [54]. Additionally, a large fraction of low-amplitude signals can increase the variability of the amplitudes seen in a pixel and therefore have an effect on the distributions of the image parameters.

Chapter 6

First Hybrid Steps

With the analysis of the first IACT established, hybrid events can be explored. In Section 6.1, the core positions reconstructed from stereo events are compared to the core positions reconstructed with HiSCORE and Section 6.2 explores the possibility of treating the IACT as another HiSCORE station for the core reconstruction.

6.1 Stereo and Hybrid Core Reconstruction

To compare the reconstructed core position of HiSCORE (CoG and ADF) with the core position of the stereo reconstruction, events that trigger IACTs 1 and 2 and HiSCORE are selected.

For the comparison, $\sim$140 min of simultaneous data taken on 2020-02-16 at IACT pointing zenith distances from 30° to 41° and azimuth angles from 170° to 230° are used. During the night, HiSCORE clusters 1, 2 and 3 were in operation. For IACT 1, the typical 14-7 cleaning is applied, while for IACT 2, a 18-9 cleaning is used. The higher cleaning levels for IACT 2 are necessary because of the longer integration time and because of the larger mirror area (see also Section 2.4). All events with four or more HiSCORE stations are reconstructed with the ADF fit (Section 4.2.1.2).

In total, 1707 real data events and 28827 simulated gammas remain.

The stereo core is reconstructed by first calculating the corresponding altitude and azimuth values for the CoG of the image. Then, the altitude and azimuth for a position of one unit length along the reconstructed major axis are determined. For each of the telescopes, the two pointing position vectors are used to span a plane through the respective telescope position. Finally, both planes are intersected at the observation

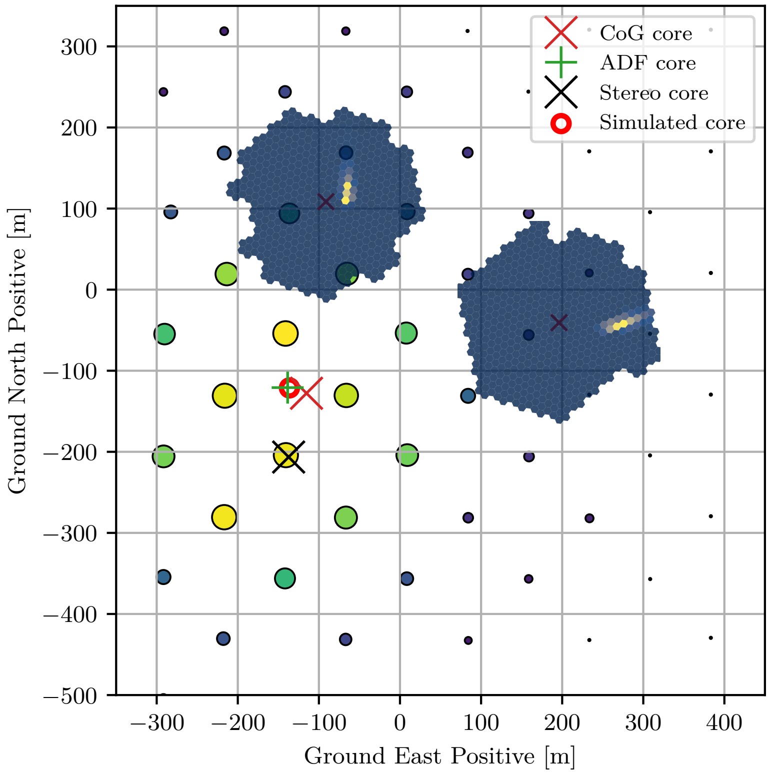
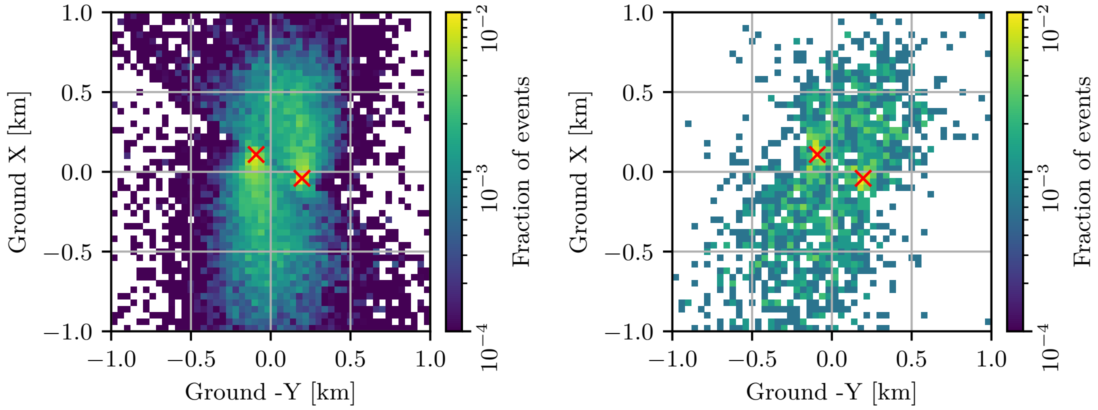


Figure 6.1: Example of a simulated 300 TeV gamma event observed by both telescopes and HiSCORE. The color of the circles and their size show the maximum amplitude measured at the station. On the telescope positions, the amplitude distributions of the cleaned images are superimposed.

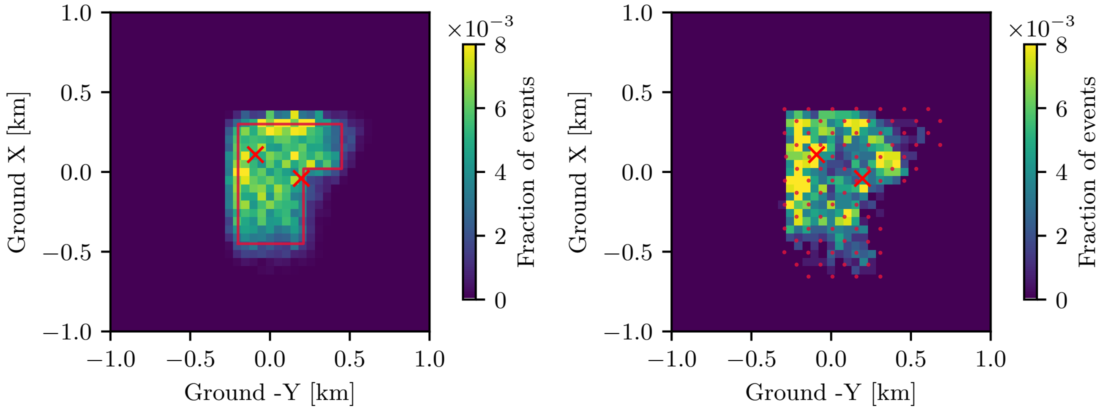
height. Figure 6.1 shows an example event of a simulated 300 TeV gamma. The core positions of the CoG, ADF and stereo reconstruction are marked. Because the core is located between the telescopes but offset, the angle between the intersecting planes is close to the optimal angle of ∼90° and the event can be well reconstructed with the method described.

Figure 6.2 shows the ground distribution of the core positions obtained with the stereo method and with the ADF fit of HiSCORE for the hybrid events for real data and for gamma simulation. Comparing the stereo core reconstructed events from real data to observation, the bulk of the distribution is centered in two lobes left and right of the telescope connection line. For the simulation, the lobe seems to be more aligned along the north-south connection line than seen in the data. This is caused by the different azimuth angles. While the real data events have a mean azimuth angle of 210° (coming from the lower left), the simulation events are centered around 180° (directly from the south). Both distributions also show an accumulation of events at the telescope positions. These are likely events that should be reconstructed to other positions close to the line connecting the telescopes. Two effects cause this: Events closer to one telescope produce circular images, so that the direction of the major axis is badly constrained. Furthermore, events in the middle between both telescopes have almost parallel intersection planes, making the reconstruction position less well defined. Both effects lead to a misreconstructed core position. When looking at the ADF distributions, the determination of the core position is independent of the core position relative to the telescope, and all events are spread more evenly over the array. Additionally, the particles reconstructed in the real experiment are mostly cosmic ray background data, while the simulation set shows only gamma events. Although they should perform similarly in the array analysis, an additional effect cannot be excluded.



(a) Stereo core simulation

(b) Stereo core real data

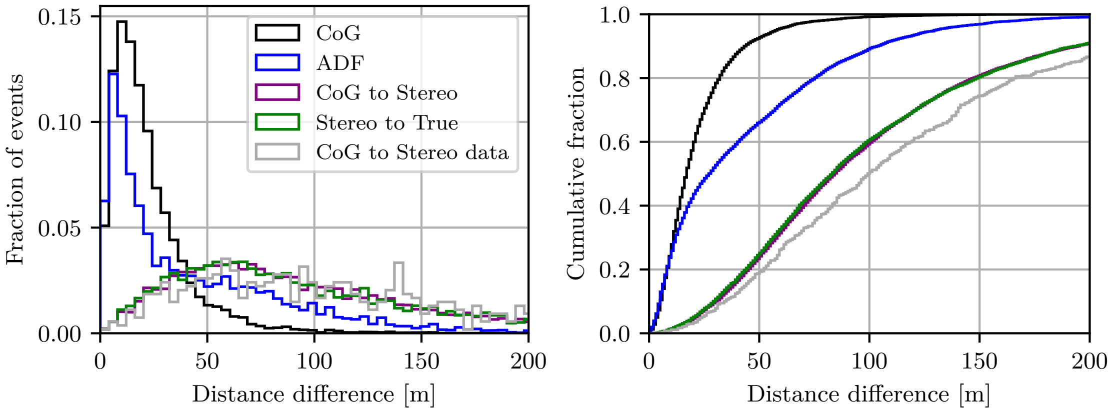


(c) CoG core simulation

(d) CoG core real data

Figure 6.2: Reconstructed core positions of coincident events triggering both telescopes and at least 4 HiSCORE stations for simulation and real data. (a, b) Stereo reconstructed core positions. The red crosses mark the positions of IACT 1 and 2. (c, d) The same events reconstructed via the center of gravity (CoG) of the maximum amplitudes of the HiSCORE stations. The red dots in (d) mark the positions of the HiSCORE stations of the three active clusters and the red perimeter in (c) shows the cut used for selecting the core positions inside of the array for a further analysis.

To compare the overall reconstruction performance, the core positions used for the evaluation are confined to core positions reconstructed to positions well inside the array. The selected inner core area is shown as the red perimeter in Figure 6.2c. Figure 6.3



(a) (b)

Figure 6.3: Overall core resolutions obtained for several methods. For center of gravity (CoG), amplitude distribution function (ADF), and for the stereo reconstruction method, the resulting distances to the simulated core are shown. Furthermore, the distances between the core positions reconstructed with CoG to the position reconstructed with the stereo method are compared for data (grey) and simulation (purple).

Simulation				Data
CoG	ADF	stereo	CoG to stereo	
24 m	54 m	117 m	119 m	137 m

Table 6.1: 68 %-containment values for center of gravity (CoG), amplitude distribution function (ADF), and for the stereo reconstruction method. Futhermore, the 68 %-containment values for the distance of the core reconstructed via CoG to the core reconstructed with the stereo method is shown for simulation and for data.

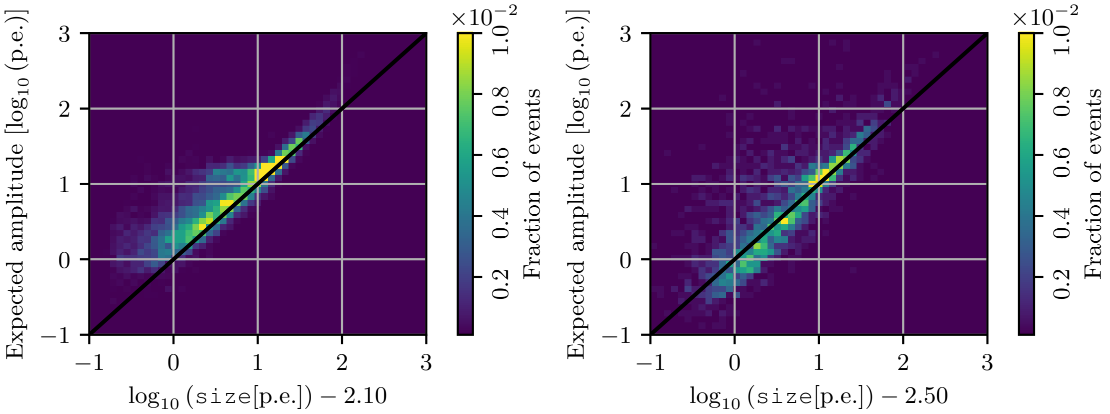
shows the distribution of the distances between the simulated and reconstructed core positions. It also shows the distance between the position reconstructed with the HiSCORE CoG to the core position reconstructed with the stereo method. As can be seen, the resolution of the stereo method is considerably worse, and the dominating part of the difference between the CoG core and the stereo core is the resolution of the stereo core reconstruction. For all events, the CoG method (black line) gives a resolution of 24 m, the ADF (blue) a resolution of 54 m and the stereo (green) a resolution of 117 m.

For the gamma simulation, the distance (purple line) of the CoG reconstructed core to the stereo reconstructed core has a 68 %-containment value at 119 m. In data, however, the 68 %-containment of the distance (grey line) between CoG and stereo reconstructed core is 137 m. The containment values are summarized in Table 6.1.

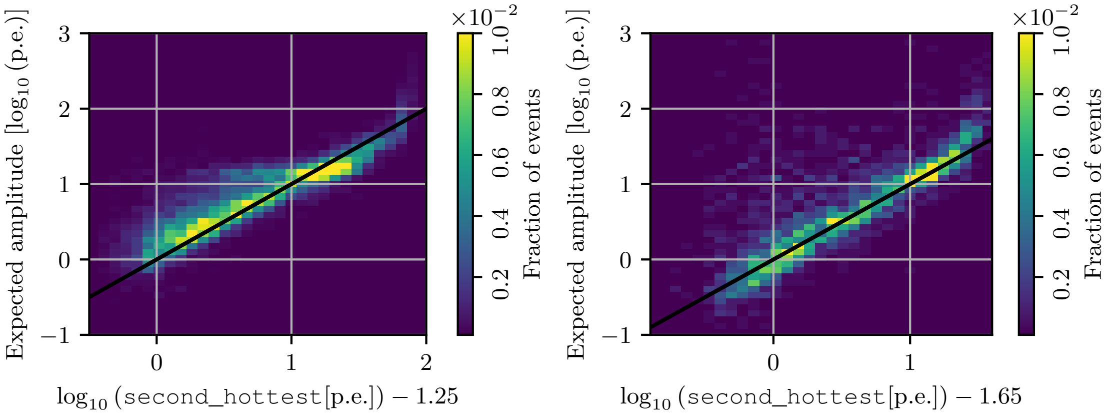
The CoG reconstruction shows the best overall resolution. This is the case because low energy and low number of station events dominate. For these events, CoG works best, as will also be shown in Section 6.2. At higher energies, ADF works better, and a threshold value of when to switch from CoG reconstruction to ADF fit should be chosen well in a hybrid analysis.

6.2 IACT as additional Station

One of the obvious possibilities of a hybrid analysis is the use of the telescope as a more sensitive HiSCORE station. Figure 6.4 shows the amplitude of the ADF function (Section 4.2.1.2) at the telescope position for the best fit obtained with HiSCORE against the shifted logarithms of the IACT parameters `size` and `second_hottest`. The offset by which the distributions are shifted is a result of the differing sensitivities of stations and IACT. It is influenced by numerous factors such as light sampling area and optical efficiency. Both `size` and `second_hottest` exhibit roughly linear behaviour. `second_hottest` seems to show a slightly better correlation with a simple scaling. $\log_{10}(\texttt{second\_hottest}) - 1.65$ gives a good approximation of the expected amplitude for real data. At large values of $\log_{10}(\texttt{second\_hottest}) - 1.65$, saturation occurs and the amplitude is underestimated. For the simulation, the roughly matching offset is -1.25. The difference of 0.4 arises because for the simulation of HiSCORE the integral



(a) Simulation (b) Real data



(c) Simulation (d) Real data

Figure 6.4: (a, b) Amplitude expected at the position of the telescope from the best fitting Amplitude Distribution Function (ADF) against the shifted $\log_{10}(\texttt{size})$ for gamma simulation and a real data run. (c, d) shows the same expected amplitude against the `second_hottest`.

pulse shape corresponds to one p.e. and the amplitude is not calibrated separately, while for the conversion in data, an additional calibration is determined.

Using $\log_{10}$ (`second_hottest`)-1.25 for the gamma simulation set gives a core position resolution as shown in Figure 6.5. As can be seen, the core resolution at energies $<$ 80 TeV is approximately 30 m with the CoG reconstruction method. For higher energies, the CoG core resolution worsens, while the ADF fit resolution improves to ∼5 m at ∼1 PeV. Using the telescope as a station, the improvement of the ADF fit at low energies and small station multiplicities is noticeable. At the highest energies, the effect gets smaller. For the CoG method, the additional telescope information worsens the 68 % containment value already at low multiplicities. This is potentially caused by a mismatch between the IACT and station calibration. Even a slight offset might cause the sensitive IACT to influence the CoG reconstruction.

In Hampf [88], the expected core resolution at ∼1 PeV is also ∼5 m for the lateral light density function (LDF) fit which performs similarly to the ADF fit. However, the CoG resolution decreases with increasing energy, while it is observed to increase in this work. This difference can be explained by the small array size considered here, leading to more truncated events at higher energies and consequently worsening CoG core estimates towards the edge of the array, where a large fraction of events is located.

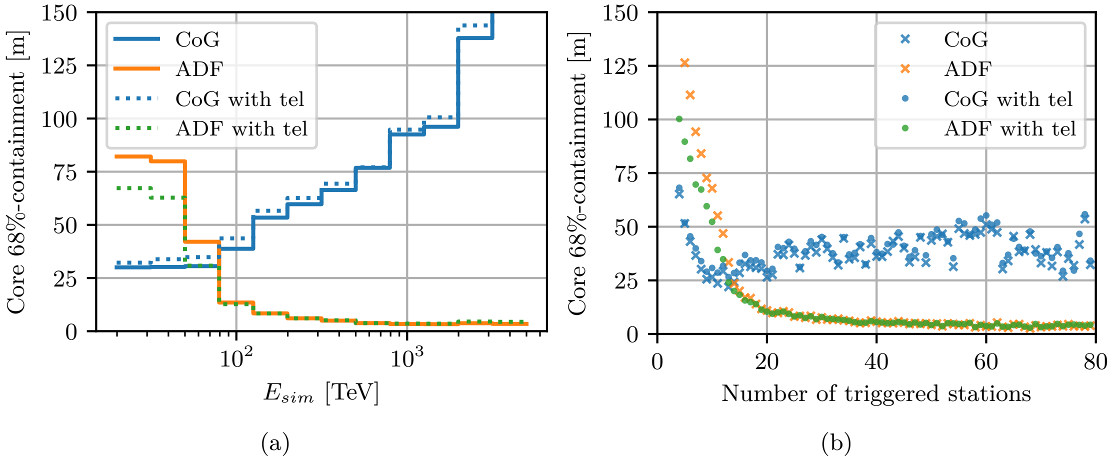


Figure 6.5: Core resolution (68 % containment) of a gamma simulation sample with mean zenith distance of 33° as a function of (a) the energy, and (b) the number of triggered stations. The Center of Gravity (CoG) and fit of the Amplitude Distribution Function (ADF) methods for the core reconstruction are used. For the reconstruction, the maximum amplitude of each station is used and in the case of 'with tel', the telescope is used as an additional station by scaling the `second_hottest` parameter of image.

This result suggests that using the telescope as an additional station improves the resolution when using the ADF fit. This is even more evident at low multiplicities below 10, where the resolution improves even further than if one more additional station were available for the fit. At 5 available stations, the resolution is $\sim$127 m, while with 4 stations and using the telescope, it is $\sim$100 m.

For the CoG method, the telescope should not be used. The thresholds of when to switch from CoG to ADF or LDF estimation should be investigated in detail for a hybrid analysis.

Chapter 7

Summary and Outlook

In this work, an analysis chain of the first TAIGA IACT starting from raw data and concluding with a high level spectral reconstruction was presented and validated on 122 h of Crab Nebula data.

Starting from the raw data, the pedestal values were estimated from the amplitude distributions with normal fits. After the calibration, a two-level next-neighbour cleaning was used and higher level image parameters were calculated.

The air shower simulations were conducted with CORSIKA and first processed with an adapted version of `sim_telarray` in sampling mode. In a second program, the two stage trigger of the MAROC boards and the readout after the hold delay was implemented. A good agreement of the higher level image parameters between data and simulation was demonstrated.

Using the image parameters, origin reconstruction, energy estimation and background suppression were implemented with Random Forests, a commonly used machine learning technique. The training of the Random Forests was done with gamma simulation data and, in case of the background suppression, also with real data as background. The IRFs derived from the test set of the simulation give an angular resolution of $<0.2^\circ$ between 10 TeV and 100 TeV, degrading towards higher and lower energies. The achieved energy resolution and bias are <25 % and well below 5 % between 10 TeV and 100 TeV. For the background suppression, a quality factor over all energies of $Q = 4$ for only the `gammaness` > 0.9 cut was found for a Crab like spectrum. Between 20 TeV and 100 TeV, $Q \approx 6$ was reached.

Applying the higher level Gammapy analysis on the Crab Nebula data resulted in an excess of $N_e = 204.6$ at a significance of $S = 9.7$ for an analysis threshold of ∼6 TeV. A power law fit resulted in a flux of $\phi_0 = (4.3 \pm 0.5) \times 10^{-10}\,\mathrm{m}^{-2}\,\mathrm{s}^{-1}\,\mathrm{TeV}^{-1}$ at a reference energy of 11 TeV with a photon index of $\Gamma = 2.97 \pm 0.15$. This result is compatible with previous and partly contemporaneous measurements.

The distribution of `width` and `length` of the excess events was found to match the simulation.

Furthermore, the core position reconstructed from stereo events was compared to the core position reconstructed with HiSCORE and considered compatible given the differences between data and the simulation set.

Using the telescope as an additional station was found to improve the core resolution at lower energies with the ADF fits.

In this first analysis of the data, the $\boldsymbol{\theta}$ cut was chosen at the level of the 68 % containment at ~100 TeV and the cleaning levels were fixed to one set for the full observation period. To refine the analysis, optimized cleaning levels also depending on the varying NSB levels and potentially energy-dependent `gammaness` and $\boldsymbol{\theta}$ cuts optimized for a given signal level could improve the sensitivity, especially at lower energies.

The angular resolution achieved with the `disp` method starts to worsen at ~100 TeV. At this energy, the angular resolution of HiSCORE is expected to become comparable or better, further improving the separation capability. Furthermore, if the hybrid scaled width and other scaled parameters are employed, the achievable separation is expected to improve significantly.

This result brings the TAIGA project one step closer to its goal of a full hybrid reconstruction up to several hundred TeV.

Now, with HiSCORE and the first IACT reasonably well understood, the hybrid method is on the verge of being established.

Appendix A

Software and Data Availability

As this work concerns the TAIGA collaboration, the software and data access is under restriction.

The software and scripts used or written for this thesis can be found on the universities GitLab server:

- Configs and programs used for simulation.[1]
- Taigapy and other IACT reconstruction software.[2]
- HiSCORE reconstruction software.[3]
- Cosmic ray fluxes for background weights.[4]

Additionally, a permanent copy of the software/scripts and a selection of the higher level data is stored on the Forschungsdaten Repositorium of the Universität Hamburg under [106].[5]

---

[1] `https://gitlab.rrz.uni-hamburg.de/TAIGA/Simulation`

[2] `https://gitlab.rrz.uni-hamburg.de/TAIGA/io_taiga`

[3] `https://gitlab.rrz.uni-hamburg.de/TAIGA/andrea_software`

[4] `https://gitlab.rrz.uni-hamburg.de/TAIGA/sensitivity`

[5] `https://doi.org/10.25592/uhhfdm.12554`

Appendix B

Processing Flow

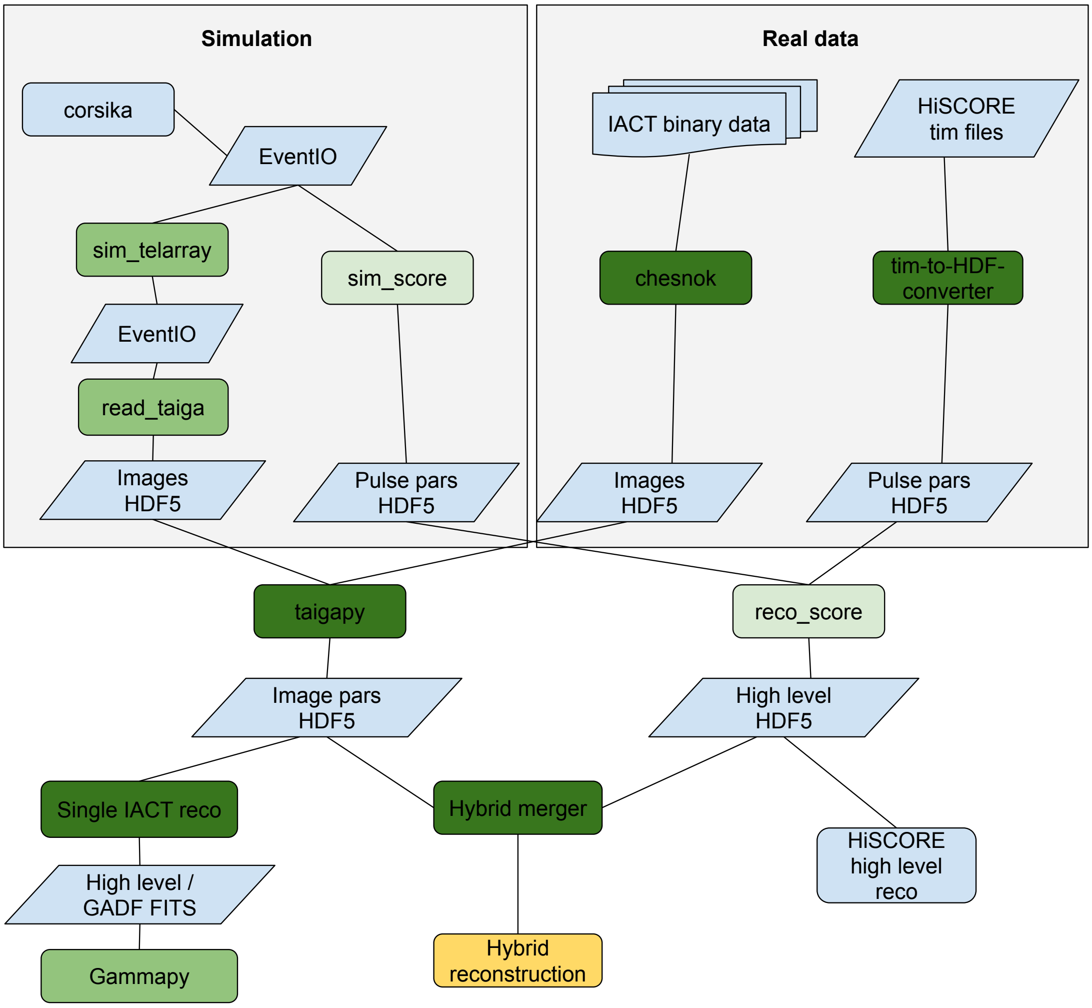

Simulation
Real data
corsika
EventIO
sim_telarray
sim_score
EventIO
read_taiga
Images HDF5
Pulse pars HDF5
IACT binary data
HiSCORE tim files
chesnok
tim-to-HDF-converter
Images HDF5
Pulse pars HDF5
taigapy
reco_score
Image pars HDF5
High level HDF5
Single IACT reco
Hybrid merger
HiSCORE high level reco
High level / GADF FITS
Gammapy
Hybrid reconstruction

Figure B.1: Processing flow of simulated and real data. The rounded boxes represent processing steps or programs and the skewed squares represent stored data. Boxes where either the program or the configuration was adapted / newly programmed are shown in green. A lighter color means less modification. The orange coloring of hybrid reconstruction means that the effort here was just barely started.

List of Figures

1.1 Spectral energy distribution of the Crab Nebula from radio to gamma ray energies. It was taken from Dirson and Horns [26] and shows measurements as well as a model fit. Additionally, the synchrotron and inverse Compton (IC) scattering components resulting from two assumed electron populations, the radio (R) and the wind (W) electrons, are shown. See [26] for details. 5
1.2 Cosmic ray flux. Taken from [28]. 7
1.3 Examples of simulated extensive air showers. Electrons, positrons and gammas are shown in red. Muons are shown in green. The vertical axis shows the altitude from 0 km to 30.1 km, while the horizontal axis spans ±5 km. Taken from [29]. 8
1.4 On the left, the polarisation of a medium is shown for the cases where the speed V of the charged particle is larger and smaller than the speed of light c in the medium. On the right, the constructive interference producing the Cherenkov wavefront with an angle Θ to the trajectory of the particle is shown. Taken from [30]. 9
1.5 On the left, the overlap of the Cherenkov light on the ground is shown (the positive labels of the horizontal axis miss a 0). Here, only the change of the Cherenkov opening angle with altitude of particles close to the shower axis is considered. On the upper right, the density of the Cherenkov light on the ground is plotted against the lateral distance, while on the lower right, the arrival time of the Cherenkov light on the ground for different emission heights is shown as a function of the lateral distance. Taken from [30]. 10
1.6 On the left, the geometric intersection of main axes of the superimposed images taken by four telescopes to reconstruct the origin of the primary particle. On the right, the planes through the shower axis and through the telescope positions show how the impact point is reconstructed. Taken from [30]. 12
1.7 A simulated hybrid event of a 34 TeV gamma. The image recorded by the IACT is overlain at the telescope position and the color scale and size of the circles indicate the maximum amplitude recorded at the optical station. The circle marks the core position of the shower and the green and red crosses mark the core positions reconstructed from the center of gravity (CoG) and amplitude distribution function (ADF) fit methods of the array reconstruction. 13
1.8 Achieved quality factor for several energies against the distance of the core to the telescope at a zenith distance of 0°. Taken from [36]. 14

1.9 Integral flux sensitivity expected for TAIGA with 300 h point source observation. Taken from [37]. 15

2.1 A HiSCORE station with its lid open in the foreground and IACT 1 in the background. The bucket-like instrument between the station and the IACT is a Tunka-133 detector. Photo courtesy TAIGA collaboration and available at `https://taiga-experiment.info/gallery`. 18

2.2 Layout of the current TAIGA detector. The HiSCORE stations are labeled with their station ID in white color. The IACTs are labelled in blue. The blue lines mark the scales of the experiment while the red line indicates the perimeter of the active array components (Clusters 1, 2, 3 and 4, 1.12 km^2). Taken from [47]. 19

2.3 Layout of the TAIGA Data Acquisition (DAQ) system. Taken from Budnev et al. [49]. All HiSCORE stations and IACTs are connected to so called Cluster Centers va optical fibers for the data transmission and for the 100 MHz time synchronisation channel. Some stations and the IACTs are additionally equipped with an optical fiber for the White Rabbit (WR) time synchronisation system. The several Cluster Centers are connected to the Central DAQ via Ethernet for the data transmission and via two optical fibers for the time synchronisation systems. 20

2.4 Sketch of the electronics of a HiSCORE station. Taken from Budnev et al. [43]. 21

2.5 Basic parameters describing the waveform of the sum of the signals recorded by a HiSCORE station. The lower panel shows the digitized signal of the 100 MHz synchronisation. Taken from Porelli [37]. 22

2.6 The first TAIGA imaging air cherenkov telescope with the second telescope in the background. Photo courtesy TAIGA collaboration. 23

2.7 Images courtesy TAIGA collaboration. (a) Image taken from www.taiga-experiment.info and cropped, (b) Taken from TAIGA internal presentation of Andrey Grinyuk, (c) Taken from TAIGA internal presentation of Martin Tluczykont. The blue circle in (a) highlights one of the small reflection shields used to blow hot air onto the mirrors to prevent ice formation. 24

2.8 Measurement of the TAIGA funnel acceptance by Nima Lubsandorzhiev. The measurement was conducted for three different funnel rotations. + and - denote the rotation in the opposing rotational directions. 25

2.9 (a) Single photoelectron response as derived from a measurement by Mirzoyan et al. [54]. (b) Quantum efficiency of the XP1911 PMTs as measured bei Razmik Mirzoyan. 26

2.10 (a) One full cluster assembly with the MAROC readout board on top and four times the PMT assembly with Divider Board and High Voltage Supply directly attached (7 PMTs each). (b) Central Controller installed above the clusters in the camera. (c) Power Management Board installed in the Camera. The visible pipes are for the camera heating system. Images courtesy TAIGA collaboration. 27

2.11 Connection of the XP1911 PMTs to the MAROC chip. Taken from Kompaniets [55]. 28

2.12 Overview of the MAROC3 ASIC. Taken from Fig. 2 of *MAROC 3A Datasheet* [56]. 28

2.13 Measured responses of the electronics to short 'single p.e.' like signals. The PMT response and Fast Shaper signals were measured by N. Lubsandhorziev and The Slow Shaper waveform was measured by E. Kompaniets. 29

2.14 Trigger and readout scheme of the first TAIGA IACT camera. When the Fast Shaper signal reaches the trigger threshold, the discriminator fires (red vertical line). After the hold delay time t_h, the MAROC board issues a local hold signal, storing the amplitudes of the Slow Shaper signals of all channels on this MAROC board in a Track and Hold circuit (red dotted vertical line). When two PMTs trigger within 15 ns on this MAROC board, the trigger information is sent to the Central Controller, where a global hold signal is sent to all MAROC boards of the camera. Then, the topology trigger condition is checked (next neighbours) and if not met, the hold signal is reset. On success, the global hold signal is kept and the readout command is sent. Then, all MAROC boards in local hold will start the ADC and transmit the amplitude information. This scheme is an adaption of figures from Kompaniets [55] and *MAROC 3A Datasheet* [56]. 30

2.15 Histograms and per pixel plots of the calibration constants of IACT 1. The values were obtained by the same author with the same method as described in Lubsandorzhiev et al. [60]. The small reduction of the spread with relative scaling for the low gain is a result of flatfielding with the preamplification. Because of the dynamic range of the preamplifier, only the high gain is adjusted (see text). 32

2.16 On the left, a sketch visualises the field of view of the CCD camera. On the right, images of the camera prior to installation and when mounted on the telescope are shown. Taken from Zhurov et al. [61]. 33

2.17 An image taken by the CCD camera of the pointing system. The image was taken during normal observation on 2019-11-07. It shows the camera and its support masts as well as a part of the sky also within the field of view of the telescope. The brighter spots around the inner hexagonal perimeter of the camera are the positioning LEDs. 33

2.18 On the left, a current scan through the central pixel of the camera is shown. The color scale of the marker codes the measured anode current at that position. ~80 codes correspond to ~1 µA. On the right, the distance between the center position estimated with the CCD camera to the center position estimated with the PMT currents is shown as a function of the telescope altitude. The image was taken from Zhurov et al. [62]. 34

2.19 (a) CCD image taken when the telescope was pointing to a bright star (Mirach). On the camera, the reflection of the star on a screen installed in front of the light guides is visible. The image was taken on 2019-09-28 at an approximate elevation of 38° (b) Screenshot of the annuli selected in SAOImageDS9 [63]. (c) Relative integral signal of the reflection assuming a background pixel intensity of 48.6. 35

2.20 (a) Histogram of the raw data amplitudes of one pixel of one chunk (usually 100000 consecutive events) of one run. The orange curve visualizes the result of the fit of a normal distribution to the bins around the pedestal maximum. The fit results in a pedestal estimate of 317.5 with a width of 2.15 p.e. (b) Histogram of all pedestal width estimates of 3 seasons. The blue histogram is without and the orange histogram is with the five additional hexagonal mirrors installed on 2020-12-09. 38
2.21 Evolution of the median value of the pedestal fit standard deviations and pedestal estimates over the course of the seasons in weekly binning. The vertical black line indicates the date of the installation of the hexagonal mirrors on the five remaining spots. 40
2.22 (a) Calibrated image of an event recorded on 2019-11-23. The double white lines indicate the 22 clusters. The ivory colored pixels are from clusters which were not read out. (b) Cleaned version of the image in (a). The cleaning levels used are $l_1 = 14$ p.e. and $l_2 = 7$ p.e.. 41
2.23 Basic image parameters. 42
2.24 Distribution of the second highest amplitude per event for one run of the first camera. The peak and spread around the peak in (a) is an indication of the trigger threshold of the camera when a $\geq$ 2-pixel trigger condition is used. It is a combined effect of the thresholds of all pixels. In (b) one can see that the amplitude distribution follows a power law for amplitudes above ∼20 p.e.. The slight dip at ∼90 p.e. is caused by the switch from high to low gain. 44
2.25 Slow Shaper response to short 'single p.e.' like signals as measured by E. Kompaniets via hold delay t_h variation, where t_h is the difference between the readout time and the trigger time. IACT 1 uses a feedback capacitance of 0.3 pF. On IACT 2, a full camera readout is implemented with the Fast Hold Board. The board distributes the hold signal created by any MAROC board to all other MAROC boards. To compensate for the additional transit time, the feedback capacitance is set to 1.2 pF. The waveforms were measured by E. Kompaniets. 46
2.26 Trigger and readout scheme of the second TAIGA IACT camera. When the trigger threshold on the Fast Shaper is reached, the discriminator fires (red vertical line). If at least two pixels are above threshold (fast coincidence check), the MAROC board issues a hold signal after the hold time t_h to all MAROC boards, storing the amplitudes of the Slow Shapers in a Track and Hold circuit (red dotted vertical line). When two PMTs trigger within 15 ns, the trigger information is sent to the Central Controller, where a global hold signal is sent to all MAROC boards of the camera. Then, the topology trigger condition is checked (next neighbours) and if not met, the hold signal is reset. On success, the global hold signal is kept and the readout command is sent. Then, all MAROC boards will start the ADC and transmit the amplitude information. This scheme is an adaption of figures from Kompaniets [55] and *MAROC 3A Datasheet* [56]. 47
2.27 Histograms and per pixel plots of the calibration constants of IACT 2. The values were obtained by the same author with the same method as described in Lubsandorzhiev et al. [60]. 48

2.28 (a) Distribution of the second highest amplitudes for one run of the second camera. In blue, the provided calibration constants (see text) are used for the high and the low gain. The clear dip starting at ∼100 p.e. is caused by a mismatch of the high and low gain calibration factors. To mitigate this, the low gain calibration is estimated from the high gain calibration factors via the linear relation evident when plotting (b) the low against the high gain amplitudes in raw ADC codes. 49

2.29 Distributions of the relative arrival times $\Delta t = T_2 - T_1$ between events recorded by IACT 1 at T_1 and by IACT 2 at T_2. The events survived the respective image cleaning and were recorded on 2020-01-27. 50

3.1 Simulated and existing array configuration. 53

3.2 Comparison of three exemplary profiles obtained from the Global Data Assimilation System with the US standard, mid-latitude winter and sub-arctic winter atmospheric profiles provided with the IACT/Atmo package. (a) Density relative to the US standard atmosphere as a function of the altitude above sea level. (b) $n-1$ relative to the US standard atmosphere, where n is the optical refractive index, as a function of the altitude above sea level. 55

3.3 (a) Atmospheric transmission to 675 m at a zenith angle of $\theta = 30°$. (b) Average optical transmission from free space to the Tunka site ($\theta = 0°$) as found in [45] and compared to transmission simulated with MODTRAN3 [83] for a midlatitude winter atmosphere and rural haze set to different visibilities. 56

3.4 Illustration of the sample weighting. For data, the altitude distribution is shown in (a) and the `width` is shown in (b). Here, the weights a_θ produce a flat distribution in altitude. For gamma simulation data, the weights take into account: The number of events simulated for each energy range, the simulated area and the spectral shape. The histogram over the true energy is shown in (c) and the `width` is shown in d). 57

3.5 Photon hit points in the mirror plane. Hypothetical point source at infinity (zenith distance $\theta = 28°$ and telescope pointing to $\theta = 30°$. 60

3.6 Containment radii r_{68} for various telescope pointings relative to a star at a zenith distance of 30°. 61

3.7 Containment radius versus the distance of a point source for a camera entrance positioned $d = 2.8\,\mathrm{cm}$ behind the focal plane. The black lines indicate the expected height of the shower maximum for gammas with an energy of 1000 TeV, 100 TeV, 10 TeV and 1 TeV from left to right. 62

3.8 Simulated low gain signal of three pixels of a telescope positioned at a distance of 100 m from the impact point of a 10 TeV gamma shower with a zenith angle of 25°. The trigger time is indicated by the dashed line and the solid line marks the time at which the amplitudes are buffered for a readout. 64

3.9 Comparison of the gain values in the simulation to the ones seen in data. The gains are sampled from a normal distribution around $\mu = 29\,\mathrm{codes\,p.e.}^{-1}$ for the high gain and $\mu = 3\,\mathrm{codes\,p.e.}^{-1}$ for the low gain. The spread is specified by a relative fluctuation of $\sigma = 0.08$. The difference between data and simulation for the low gain is a result of the dynamic range of the MAROC preamplifier for high and low gain (see text). 65

3.10 Comparison of the pedestal values in the simulation to the measured ones. The simulated pedestals are sampled from a normal distribution around $\mu = 340$ codes with $\sigma = 28$ codes. 66

3.11 Resulting background distribution in one time slice for two background photoelectron rates. A pedestal value of 343 codes and a gain of 27.61 codes/p e was used. The tail above 10 p.e. to large values is an effect of the afterpulsing. 66

3.12 Normal distribution width values obtained via a fit of a normal distribution around the peak of the expected background distributions for different photoelectron background rates. For each rate, the fit is repeated 10000 for histograms with 50000 entries each. The resulting median of the obtained σ values and the standard deviation are plotted. 67

3.13 The 798 trigger pairs of next neighbours within a radius of 3.7°. 68

3.14 (a) Hold delay tuned such that the readout occurs ~10 ns after the peak. b) Fraction of the amplitude read out at hold time to the amplitude at the maximum of the signal. Two cases are considered: Readout at the time of the maximum and readout 10 ns after the maximum. Only pixels with a signal of more than 5 p.e. are considered. 69

3.15 Comparison of simulation with data (1/2). 71

3.16 Comparison of simulation with data (2/2). 72

3.17 Effective area (a) and the expected rate (b) after a cut of `size` $>$ 80 p.e.. The sum of all the rates in the plot is $\sim 7.1\,\mathrm{s}^{-1}$. 74

4.1 The first two levels of one of the Decision Trees of the Random Forest gamma/background-classifier. Each box represents one node. The first line is the chosen image parameter and cut, the second line shows the *Gini* value of the node. Samples is the number of samples in the node. For value, the first of both values is the sum of the weight of the background samples and the second value is the sum of the weight of the gamma samples in the node. The last line refers to the dominating class in this node (The class the classifier would vote for if it were to stop here). The color of the boxes corresponds to the class. The higher the *Gini* value, the lighter the coloring. 80

4.2 Feature importances for the `disp` regression Random Forest. 83

4.3 Estimate of the Random Forest regressor for the `disp` parameter against the true `disp` of the simulated events. The line with a slope of one visualizes the result a perfect estimator would give. The score of the regression is $r^2 = 0.813 \pm 0.005$ and $MAPE = 0.235 \pm 0.006$. The data shows the test sample after quality cuts. 84

4.4 Median (bias) and half of the 16 %- to 84 %-quantiles (resolution) of $\frac{\mathtt{disp}_{est}}{\mathtt{disp}_{sim}} - 1$ of the Random Forest `disp` parameter reconstruction as a function of the simulated energy E_{sim}. The score of the regression is $r^2 = 0.813 \pm 0.005$ and $MAPE = 0.235 \pm 0.006$. The data shows the results for the test sample after quality cuts. 85

4.5 Angular resolution (68 % containment) for the origin reconstruction with the Random Forest regressor for the `disp` parameter as a function of the simulated energy. For the dotted lines, `size` > 80 and `leakage` < 0.15 are applied. For the solid lines, `gammaness` > 0.9 is additionally applied. The blue curves represent the events where `disp` points towards the source position, while the orange curves represent all events. 86
4.6 Feature importances for the energy regression Random Forest. 87
4.7 Energy estimate of the Random Forest regressor as a function of the true energy. The line with a slope of one visualizes the result a perfect estimator would give. The scores of the regression are $r^2 = 0.829 \pm 0.001$ and $MAPE = 0.25 \pm 0.06$. The red markers show the median and 16 %- to 84 %-quantiles. 88
4.8 Relative accuracy of the energy estimate of the energy Random Forest regressor as a function of the true energy. The line at 0 visualizes the result a perfect estimator would give. The scores of the regression are $r^2 = 0.829 \pm 0.001$ and $MAPE = 0.25 \pm 0.06$. The green line visualises the median and 16 %- to 84 %-quantiles. 89
4.9 Median (bias) and half of the 16 %- to 84 %-quantiles (resolution) of $\frac{E_{est}}{E_{sim}} - 1$, where E_{est} and E_{sim} are the reconstructed and simulated energy, respectively. The scores of the regression are $r^2 = 0.829 \pm 0.001$ and $MAPE = 0.25 \pm 0.06$. 89
4.10 Feature importances for the gamma/background Random Forest classifier. 90
4.11 Distribution, efficiency and quality factor against `gammaness` for the full test set and two example energy bins. The histograms on the left show the distribution of the simulated gamma and of the data (considered background). In the middle, the efficiencies for different `gammaness` cut values are shown and on the right, the corresponding Q-factors are shown. At about 0.9 the gamma efficiency is 39 % and the background efficiency is about 0.95 % giving a Q-factor of about 4.0. The Q-factors are only calculated if more than 10 background events remain after the cut. 92
4.12 (a) Background efficiency, (b) gamma efficiency and (c) quality factor for three `gammaness` cut values against the energy estimated with the Random Forest regressor. The quality factor is only calculated for entries with more than 10 surviving background events. The errors give the uncertainties estimated from the event statistics. 93
4.13 Effective Area of the Random Forest analysis with different cuts applied for zenith distances from $\theta > 40°$ to $\theta < 50°$. The blue curve shows the triggering area after cleaning. The other curves show the area after different additional analysis cuts. 95
4.14 Effective Area of the Random Forest analysis with different cuts applied for zenith distances from $\theta > 10°$ to $\theta < 20°$. The blue curve shows the triggering area after cleaning. The other curves show the area after different additional analysis cuts. 95
4.15 Histogram of the events surviving `gammaness` and $\boldsymbol{\theta}$ cut at zenith angles from $\theta > 40°$ to $\theta < 50°$ weighted to Crab-like power law spectrum $\propto E^{-2.8}$. The maximum of the histograms indicate the energy threshold. . 96
4.16 Histogram of the events surviving `gammaness` and $\boldsymbol{\theta}$ cut at zenith angles from $\theta > 10°$ to $\theta < 20°$ weighted to Crab-like power law spectrum $\propto E^{-2.8}$. The maximum of the histograms indicate the energy threshold. . 97

5.1 (b) The total available observation time per source including bad weather and dead time. (a) The observation time for the Crab Nebula as a function of the zenith distance . . . 104
5.2 Histogram of the mean scaled rate of the portions over the mean altitudes. The data was taken between October 19 and December 20. One portion corresponds to ∼120 s of observation time. The good-weather cut is defined as all portions contained between the red lines at 7.3 Hz and 12.2 Hz. . . . 105
5.3 Histogram of the scaled rate per portion for two periods: Portions taken during January 2020, February 2020 and March 2020 are plotted in orange and portions taken during the remaining month between October 2019 and 2020-12-08 are shown in blue. The dashed lines correspond to fits of Normal distributions. The good-weather cut is defined as all portions contained between the red lines at 7.3 Hz and 12.2 Hz. . . . 106
5.4 Placement of the analysis regions. The center is the on source region with a radius of $\boldsymbol{\theta}_s = 0.2°$. The circles on the two arcs to the right and to the left are the off regions positioned at the same distance to the two wobble pointing positions of the telescope. . . . 108
5.5 Counts around the position of the Crab Nebula N_{on} and around the off positions αN_{off} as a function of the angular distance to the region centers $\boldsymbol{\theta}$. The 13 off regions give an acceptance ratio of $\alpha = \frac{1}{13}$. The errors are estimated assuming a Poisson distribution of the counts. . . . 109
5.6 Excess $N_e = N_{on} - \alpha N_{off}$ as a function of the live time with $\alpha = \frac{1}{13}$. The orange line is the result of a linear fit to the data, giving a slope of $(1.7044 \pm 0.0033)\,\mathrm{h}^{-1}$. . . . 110
5.7 Counts of the on position N_{on} and scaled counts of the off positions αN_{off} with $\alpha = \frac{1}{13}$ as functions of (a) the `length` and (b) the `width`. The lower panel shows the excess events $N_e = N_{on} - \alpha N_{off}$ with gamma simulation superimposed. The simulation is taken at a zenith distances between 30° and 40° and weighted to a primary spectrum of $\propto E^{-2.8}$. It is scaled such that the sum of the weights matches the sum of the excess events of the histogram. . . . 110
5.8 Significance $S = \sqrt{TS}$ as a function of the live time. . . . 112
5.9 Raw excess events per simulated energy of the Crab Nebula observation. The predicted counts of the best fit power law are shown in orange. The lower plot visualizes the residuals between predicted and excess counts. . 115
5.10 Flux points of the Crab Nebula as determined with a forward folding fit implemented in Gammapys FluxPointsEstimator. The color scale indicates the change of the test statistic TS of a model with only the norm as a free parameter. The model is constructed by multiplying a normalisation parameter to the value of the best fit power law in the energy range of the bin. . . . 116

5.11 Spectrum of the Crab Nebula obtained with the Random Forest analysis of the first TAIGA IACT. The total live time after good-weather selection is about 122 h. The orange line is the best fit power law. The $1 - \sigma$ statistical error is shown in grey. The flux points resulting from a forward folding fit assuming the best fit power law are shown in blue. For comparison, a model fit to data of several experiments in [26] is shown in green. The lower plot shows the difference of the flux points to the power law fit (orange). 117
5.12 Change of the energy reconstruction bias (median) at ±20 % change in optical transmission against the simulated energy of the Random Forest Regressor for gammas at zenith distances from 31° to 41°. 119

6.1 Example of a simulated 300 TeV gamma event observed by both telescopes and HiSCORE. The color of the circles and their size show the maximum amplitude measured at the station. On the telescope positions, the amplitude distributions of the cleaned images are superimposed. 124
6.2 Reconstructed core positions of coincident events triggering both telescopes and at least 4 HiSCORE stations for simulation and real data. (a, b) Stereo reconstructed core positions. The red crosses mark the positions of IACT 1 and 2. (c, d) The same events reconstructed via the center of gravity (CoG) of the maximum amplitudes of the HiSCORE stations. The red dots in (d) mark the positions of the HiSCORE stations of the three active clusters and the red perimeter in (c) shows the cut used for selecting the core positions inside of the array for a further analysis. . . . 125
6.3 Overall core resolutions obtained for several methods. For center of gravity (CoG), amplitude distribution function (ADF), and for the stereo reconstruction method, the resulting distances to the simulated core are shown. Furthermore, the distances between the core positions reconstructed with CoG to the position reconstructed with the stereo method are compared for data (grey) and simulation (purple). 126
6.4 (a, b) Amplitude expected at the position of the telescope from the best fitting Amplitude Distribution Function (ADF) against the shifted $\log_{10}$(`size`) for gamma simulation and a real data run. (c, d) shows the same expected amplitude against the `second_hottest`. 127
6.5 Core resolution (68 % containment) of a gamma simulation sample with mean zenith distance of 33° as a function of (a) the energy, and (b) the number of triggered stations. The Center of Gravity (CoG) and fit of the Amplitude Distribution Function (ADF) methods for the core reconstruction are used. For the reconstruction, the maximum amplitude of each station is used and in the case of 'with tel', the telescope is used as an additional station by scaling the `second_hottest` parameter of image. 128

B.1 Processing flow of simulated and real data. The rounded boxes represent processing steps or programs and the skewed squares represent stored data. Boxes where either the program or the configuration was adapted / newly programmed are shown in green. A lighter color means less modification. The orange coloring of hybrid reconstruction means that the effort here was just barely started. 135

List of Tables

3.1 The simulation sets. Each set was simulated with a power law following $\propto E^{-1}$ between a lower energy E_l and a higher energy E_h. The origin was spread evenly in a circle with the diameter Ω around the nominal zenith distance θ. On the ground, the core locations are scattered between $\pm X_s$ in the X direction and $\pm Y_s$ in the Y direction. N_{runs} is the total number of simulated runs in the set and N_{evt} is the total number of events. If a shower was reused c times by randomly scattering it over the allowed core area, it was also counted c times. 52

4.1 Importances of the features in the Random Forests for the different tasks. The importances are based on the mean decrease in impurity. 83

5.1 Flux points as determined with a forward folding fit implemented in Gammapys FluxPointsEstimator for a source flux model with $4.30\,\mathrm{m}^{-2}\,\mathrm{s}^{-1}\,\mathrm{TeV}^{-1}\cdot(E/11\,\mathrm{TeV})^{-2.97}$. The obtained $\sqrt{TS}$ gives the significance of that bin. μ_s is the number of predicted signal events (excess) for that bin and N_{on} is the number of counts in the on region. 116

5.2 Results for various splits of the Crab data. Normalization ϕ_0 and photon index Γ of a power law fit $\phi_0 E^{-\Gamma}$ are listed together with the off-diagonal element of the covariance matrice c_{11}. Additionally, the overall number of counts in the on region N_{on}, the number of excess events N_e, the resulting significance $\sqrt{TS}$ and the corresponding live time t are given. The splits considered are the first half of the data (October 19 to January 20), the second half of the data set, and the data taken in each of the two wobble positions (±1.2° in right ascension). 118

6.1 68 %-containment values for center of gravity (CoG), amplitude distribution function (ADF), and for the stereo reconstruction method. Futhermore, the 68 %-containment values for the distance of the core reconstructed via CoG to the core reconstructed with the stereo method is shown for simulation and for data. 126

Abbreviations

TAIGA	Tunka Advanced Instrument for Gamma- and cosmic-ray Astronomy
HiSCORE	Hundred *i* Square-km Cosmic ORigin Explorer
ADC	Analog to Digital Converter
ADF	Amplitude Distribution Function
AOD	Aerosol Optical Depth
BG	Background
CCD	Charge Coupled Device
CMB	Cosmic Microwave Background
CoG	Center of Gravity
DAQ	Data Aquisition
EAS	Extensive Air Shower
FITS	Flexible Image Transport System
FOV	Field of View
FWHM	Full Width Half Maximum
GADF	Gamma Astro Data Formats
GDAS	Global Data Assimilation System
GRB	Gamma Ray Burst
GPS	Global Positioning System
HDF	Hierarchical Data Format
HSCW	Hybrid Scaled Width
HV	High Voltage
IACT	Imaging Air Cherenkov Telescope
IC	Inverse Compton
IRF	Instrument Response Function
ISRF	Interstellar Radiation Field
LED	Light Emitting Diode
LDF	Lateral Light Density Function
MAPE	Mean Absolute Percentage Error
MC	Monte Carlo
MSE	Mean Squared Error
NSB	Night Sky Background
PMT	Photomultiplier Tube

PSF	Point Spread Function
QE	Quantum Efficiency
RF	Random Forest
SCL	Scaled Length
SCW	Scaled Width
SNR	Supernova remnants
SPE	Single Photoelectron
TS	Test Statistics
UHE	Ultra High Energy
VHE	Very High Energy
WR	White Rabbit
ZD	Zenith Distance

References

[1] V. F. Hess. “Über Beobachtungen der durchdringenden Strahlung bei sieben Freiballonfahrten”. In: *Phys. Z.* 13 (1912), pp. 1084–1091.

[2] K. G. Jansky. “Radio Waves from Outside the Solar System”. In: Nature 132.3323 (July 1933), p. 66. DOI: `10.1038/132066a0`.

[3] R. Giacconi, H. Gursky, F. R. Paolini, and B. B. Rossi. “Evidence for x Rays From Sources Outside the Solar System”. In: *Phys. Rev. Lett.* 9 (11 Dec. 1962), pp. 439–443. DOI: `10.1103/PhysRevLett.9.439`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.9.439`.

[4] R. W. Klebesadel, I. B. Strong, and R. A. Olson. “Observations of Gamma-Ray Bursts of Cosmic Origin”. In: ApJ 182 (June 1973), p. L85. DOI: `10.1086/181225`.

[5] M. Meyer, D. Horns, and H.-S. Zechlin. “The Crab Nebula as a standard candle in very high-energy astrophysics”. In: *A&A* 523 (2010), A2. DOI: `10.1051/0004-6361/201014108`. URL: `https://doi.org/10.1051/0004-6361/201014108`.

[6] T. C. Weekes, M. F. Cawley, D. J. Fegan, et al. “Observation of TeV Gamma Rays from the Crab Nebula Using the Atmospheric Cerenkov Imaging Technique”. In: ApJ 342 (July 1989), p. 379. DOI: `10.1086/167599`.

[7] F. Aharonian, A. Akhperjanian, M. Beilicke, et al. “The Crab Nebula and Pulsar between 500 GeV and 80 TeV: Observations with the HEGRA Stereoscopic Air Cerenkov Telescopes”. In: ApJ 614.2 (Oct. 2004), pp. 897–913. DOI: `10.1086/423931`. arXiv: `astro-ph/0407118 [astro-ph]`.

[8] F. Aharonian, A. G. Akhperjanian, A. R. Bazer-Bachi, et al. “Observations of the Crab nebula with HESS”. In: *A&A* 457.3 (2006), pp. 899–915. DOI: `10.1051/0004-6361:20065351`. URL: `https://doi.org/10.1051/0004-6361:20065351`.

[9] K. Meagher. *Six years of VERITAS observations of the Crab Nebula.* 2015. arXiv: `1508.06442 [astro-ph.HE]`.

[10] A. U. Abeysekara, A. Albert, R. Alfaro, et al. “Measurement of the Crab Nebula Spectrum Past 100 TeV with HAWC”. In: *The Astrophysical Journal* 881.2 (Aug. 2019), p. 134. DOI: `10.3847/1538-4357/ab2f7d`. URL: `https://dx.doi.org/10.3847/1538-4357/ab2f7d`.

[11] M. Amenomori, Y. W. Bao, X. J. Bi, et al. “First Detection of Photons with Energy beyond 100 TeV from an Astrophysical Source”. In: *Phys. Rev. Lett.* 123 (5 July 2019), p. 051101. DOI: `10.1103/PhysRevLett.123.051101`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.123.051101`.

[12] MAGIC Collaboration, V. A. Acciari, S. Ansoldi, et al. “MAGIC very large zenith angle observations of the Crab Nebula up to 100 TeV”. In: A&A 635, A158 (May 2020), A158. DOI: `10.1051/0004-6361/201936899`. arXiv: `2001.09566 [astro-ph.HE]`.

[13] T. L. Collaboration*†, Z. Cao, F. Aharonian, et al. “Peta-electron volt gamma-ray emission from the Crab Nebula”. In: *Science* 373.6553 (2021), pp. 425–430. DOI: `10.1126/science.abg5137`. eprint: `https://www.science.org/doi/pdf/10.1126/science.abg5137`. URL: `https://www.science.org/doi/abs/10.1126/science.abg5137`.

[14] S. Lombardi, O. Catalano, S. Scuderi, et al. “First detection of the Crab Nebula at TeV energies with a Cherenkov telescope in a dual-mirror Schwarzschild-Couder configuration: the ASTRI-Horn telescope”. In: *A&A* 634 (2020), A22. DOI: `10.1051/0004-6361/201936791`. URL: `https://doi.org/10.1051/0004-6361/201936791`.

[15] Z. Cao, F. A. Aharonian, Q. An, et al. “Ultrahigh-energy photons up to 1.4 petaelectronvolts from 12 γ-ray Galactic sources”. In: Nature 594.7861 (June 2021), pp. 33–36. DOI: `10.1038/s41586-021-03498-z`.

[16] X.-H. Ma, Y.-J. Bi, Z. Cao, et al. “Chapter 1 LHAASO Instruments and Detector technology *”. In: *Chinese Physics C* 46.3 (Mar. 2022), p. 030001. DOI: `10.1088/1674-1137/ac3fa6`. URL: `https://dx.doi.org/10.1088/1674-1137/ac3fa6`.

[17] A. Baktash and D. Horns. *EASpy: Fast simulation of fluorescence and Cherenkov light from extended air showers at large zenith angles.* 2023. arXiv: `2305.08710 [astro-ph.IM]`.

[18] C. T. A. Observatory and C. T. A. Consortium. *CTAO Instrument Response Functions - prod5 version v0.1.* Version v0.1. Zenodo, Sept. 2021. DOI: `10.5281/zenodo.5499840`. URL: `https://doi.org/10.5281/zenodo.5499840`.

[19] K. Bernlöhr, A. Barnacka, Y. Becherini, et al. “Monte Carlo design studies for the Cherenkov Telescope Array”. In: *Astroparticle Physics* 43 (2013). Seeing the High-Energy Universe with the Cherenkov Telescope Array - The Science Explored with the CTA, pp. 171–188. ISSN: 0927-6505. DOI: `https://doi.org/10.1016/j.astropartphys.2012.10.002`. URL: `https://www.sciencedirect.com/science/article/pii/S0927650512001867`.

[20] M. Tluczykont, T. Kneiske, D. Hampf, and D. Horns. *Gamma-ray and Cosmic Ray Astrophysics from 10 TeV to 1 EeV with the large-area (>10 km^2) air-shower Detector SCORE.* 2009. arXiv: `0909.0445 [astro-ph.IM]`.

[21] E. Fermi. “On the Origin of the Cosmic Radiation”. In: *Phys. Rev.* 75 (8 Apr. 1949), pp. 1169–1174. DOI: `10.1103/PhysRev.75.1169`. URL: `https://link.aps.org/doi/10.1103/PhysRev.75.1169`.

[22] S. Gabici and F. A. Aharonian. “Searching for Galactic Cosmic-Ray Pevatrons with Multi-TeV Gamma Rays and Neutrinos”. In: *The Astrophysical Journal* 665.2 (Aug. 2007), p. L131. DOI: `10.1086/521047`. URL: `https://dx.doi.org/10.1086/521047`.

[23] S. Funk. “Ground- and Space-Based Gamma-Ray Astronomy”. In: *Annual Review of Nuclear and Particle Science* 65.1 (2015), pp. 245–277. DOI: `10.1146/annurev-nucl-102014-022036`. eprint: `https://doi.org/10.1146/annurev-nucl-102014-022036`. URL: `https://doi.org/10.1146/annurev-nucl-102014-022036`.

[24] I. V. Moskalenko, T. A. Porter, and A. W. Strong. “Attenuation of Very High Energy Gamma Rays by the Milky Way Interstellar Radiation Field”. In: *The Astrophysical Journal* 640.2 (Mar. 2006), p. L155. DOI: `10.1086/503524`. URL: `https://dx.doi.org/10.1086/503524`.

[25] S. Vernetto and P. Lipari. “Absorption of very high energy gamma rays in the Milky Way”. In: *Phys. Rev. D* 94 (6 Sept. 2016), p. 063009. DOI: `10.1103/PhysRevD.94.063009`. URL: `https://link.aps.org/doi/10.1103/PhysRevD.94.063009`.

[26] L. Dirson and D. Horns. “Phenomenological modelling of the Crab Nebula´s broadband energy spectrum and its apparent extension”. In: *A&A* 671 (2023), A67. DOI: `10.1051/0004-6361/202243578`. URL: `https://doi.org/10.1051/0004-6361/202243578`.

[27] R. Bühler and R. Blandford. “The surprising Crab pulsar and its nebula: a review”. In: *Reports on Progress in Physics* 77.6 (June 2014), p. 066901. DOI: `10.1088/0034-4885/77/6/066901`. URL: `https://dx.doi.org/10.1088/0034-4885/77/6/066901`.

[28] T. K. Gaisser, R. Engel, and E. Resconi. “Cosmic ray data”. In: *Cosmic Rays and Particle Physics*. 2nd ed. Cambridge University Press, 2016, pp. 12–29. DOI: `10.1017/CBO9781139192194.004`.

[29] F. Schmidt and J. Knapp. *CORSIKA Shower Images*. 2005. URL: `https://www-zeuthen.desy.de/~jknapp/fs/showerimages.html`.

[30] M. de Naurois and D. Mazin. “Ground-based detectors in very-high-energy gamma-ray astronomy”. In: *Comptes Rendus Physique* 16.6 (2015). Gamma-ray astronomy / Astronomie des rayons gamma, pp. 610–627. ISSN: 1631-0705. DOI: `https://doi.org/10.1016/j.crhy.2015.08.011`. URL: `https://www.sciencedirect.com/science/article/pii/S1631070515001462`.

[31] W. Hofmann and J. Hinton. “Detectors for high-energy messengers from the Universe”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 907 (2018). Advances in Instrumentation and Experimental Methods (Special Issue in Honour of Kai Siegbahn), pp. 31–45. ISSN: 0168-9002. DOI: `https://doi.org/10.`

1016/j.nima.2018.03.020. URL: https://www.sciencedirect.com/science/article/pii/S0168900218303553.

[32] G. R. Blumenthal and R. J. Gould. "Bremsstrahlung, Synchrotron Radiation, and Compton Scattering of High-Energy Electrons Traversing Dilute Gases". In: *Rev. Mod. Phys.* 42 (2 Apr. 1970), pp. 237–270. DOI: 10.1103/RevModPhys.42.237. URL: https://link.aps.org/doi/10.1103/RevModPhys.42.237.

[33] A. M. Hillas. "Cerenkov light images of EAS produced by primary gamma". In: *ICRC* 3 (1985), pp. 445–448.

[34] T.-P. Li and Y.-Q. Ma. "Analysis methods for results in gamma-ray astronomy." In: *Astophys. J.* 272 (Sept. 1983), pp. 317–324. DOI: 10.1086/161295.

[35] J. Aleksić, S. Ansoldi, L. A. Antonelli, et al. "The major upgrade of the MAGIC telescopes, Part II: A performance study using observations of the Crab Nebula". In: *Astroparticle Physics* 72 (Jan. 2016), pp. 76–94. DOI: 10.1016/j.astropartphys.2015.02.005. arXiv: 1409.5594 [astro-ph.IM].

[36] M. Kunnas. "Studies of the performance of an IACT system for the TAIGA array". PhD thesis. Universität Hamburg, 2017. URL: https://ediss.sub.uni-hamburg.de/handle/ediss/7582.

[37] A. Porelli. "TAIGA-HiSCORE: a new wide-angle air Cherenkov detector for multi-TeV gamma-astronomy and cosmic ray physics". PhD thesis. Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät, 2020. DOI: http://dx.doi.org/10.18452/21610.

[38] M. Tluczykont, D. Hampf, D. Horns, et al. "The ground-based large-area wide-angle γ-ray and cosmic-ray experiment HiSCORE". In: *Advances in Space Research* 48.12 (2011), pp. 1935–1941. DOI: https://doi.org/10.1016/j.asr.2011.08.004.

[39] D. Hampf, M. Tluczykont, and D. Horns. "Event reconstruction techniques for the wide-angle air Cherenkov detector HiSCORE". In: *NIMA* 712 (2013), p. 137.

[40] M. Tluczykont, D. Hampf, D. Horns, et al. "The HiSCORE concept for gamma-ray and cosmic-ray astrophysics beyond 10 TeV". In: *Astr.Part.Phys.* 56 (2014), p. 42.

[41] I. Astapov, P. Bezyazeekov, A. Borodin, et al. “Scintillation detectors for the TAIGA experiment”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 936 (2019). Frontier Detectors for Frontier Physics: 14th Pisa Meeting on Advanced Detectors, pp. 254–256. DOI: `https://doi.org/10.1016/j.nima.2018.10.081`.

[42] on behalf of TAIGA collaboration, I. Astapov, P. Bezyazeekov, et al. “Optimisation studies of the TAIGA-Muon scintillation detector array”. In: *Journal of Instrumentation* 17.06 (June 2022), P06022. DOI: `10.1088/1748-0221/17/06/P06022`. URL: `https://dx.doi.org/10.1088/1748-0221/17/06/P06022`.

[43] N. Budnev, I. Astapov, P. Bezyazeekov, et al. “The TAIGA experiment: from cosmic ray to gamma-ray astronomy in the Tunka valley”. In: *Journal of Physics: Conference Series* 718.5 (May 2016), p. 052006. DOI: `10.1088/1742-6596/718/5/052006`. URL: `https://dx.doi.org/10.1088/1742-6596/718/5/052006`.

[44] U. B. de Almeida and M. Tluczykont. “Future developments in ground-based gamma-ray astronomy”. To be published.

[45] A. Dementeva, G. Zhamsueva, A. Zayakhanov, and V. Tcydypov. “Interannual and Seasonal Variation of Optical and Microphysical Properties of Aerosol in the Baikal Region”. In: *Atmosphere* 13.2 (2022). DOI: `10.3390/atmos13020211`. URL: `https://www.mdpi.com/2073-4433/13/2/211`.

[46] S. F. Berezhnev, D. Besson, N. M. Budnev, et al. “The Tunka-133 EAS Cherenkov light array: Status of 2011”. In: *NIMA* 692 (2012), p. 98.

[47] M. Blank, M. Tluczykont, A. Porelli, et al. “Detection of the Crab Nebula using a random forest analysis of the first TAIGA IACT data”. In: *Monthly Notices of the Royal Astronomical Society* (Jan. 2023). stad276. ISSN: 0035-8711. DOI: `10.1093/mnras/stad276`. eprint: `https://academic.oup.com/mnras/advance-article-pdf/doi/10.1093/mnras/stad276/48943735/stad276.pdf`. URL: `https://doi.org/10.1093/mnras/stad276`.

[48] M. Brückner and R. Wischnewski. “A White Rabbit Setup for Sub-nsec Synchronization, Timestamping and Time Calibration in Large Scale Astroparticle Physics Experiments”. In: *International Cosmic Ray Conference*. Vol. 33. International Cosmic Ray Conference. Jan. 2013, p. 3151.

[49] N. Budnev, I. Astapov, P. Bezyazeekov, et al. “TAIGA—an advanced hybrid detector complex for astroparticle physics and high energy gamma-ray astronomy in the Tunka valley”. In: *Journal of Instrumentation* 15.09 (Sept. 2020), p. C09031. DOI: `10.1088/1748-0221/15/09/C09031`. URL: `https://dx.doi.org/10.1088/1748-0221/15/09/C09031`.

[50] S. Epimakhov. “Exploring cosmic ray origins with ground-based EAS arrays Tunka and HiSCORE”. PhD thesis. Universität Hamburg, 2015.

[51] R. Mirzoyan, R. Kankanian, F. Krennrich, et al. “”The first telescope of the HEGRA air Cherenkov imaging telescope array””. In: *Nucl. Instr. Meth. Phys. Res. A* 351 (1994), pp. 513–526.

[52] A. Borodin, V. Grebenyuk, A. Grinyuk, et al. “Imaging Atmospheric Cherenkov Telescope for the TAIGA Observatory—JINR Participation”. In: *Physics of Atomic Nuclei* 83.2 (Aug. 2020), pp. 268–271. DOI: `10.1134/S1063778820020076`.

[53] A. Grinyuk, E. Postnikov, and L. Sveshnikova. “Monte Carlo Simulation of the TAIGA Hybrid Gamma-Ray Experiment.” In: *Phys. Atom. Nuclei* 83 (2020). doi: https://doi.org/10.1134/S106377882002012X, pp. 262–267.

[54] R. Mirzoyan, E. Lorenz, D. Petry, and C. Prosch. “On the influence of afterpulsing in PMTs on the trigger threshold of multichannel light detectors in self-trigger mode”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 387.1 (1997). New Developments in Photodetection, pp. 74–78. DOI: `https://doi.org/10.1016/S0168-9002(96)00964-3`.

[55] K. Kompaniets. “Development of front-end electronics of TAIGA-IACT”. Poster presented at the International Symposium on Cosmic Rays and Astrophysics (ISCRA). 2019. URL: `https://iscra2019.mephi.ru/content/public/files/posters/Poster_4_07_KompanietsK.pdf`.

[56] *MAROC 3A Datasheet*. 1.1. 1.1. Weeroc. Nov. 2017.

[57] R. Mirzoyan. *Conversion Factor Calibration for MAGIC Based on The Use of Measured F-Factor of PMTs.* MAGIC internal note. 2000.

[58] R. Mirzoyan and E. Lorenz. “On the Calibration Accuracy of Light Sensors in Atmospheric Cherenkov Fluorescence and Neutrino Experiments”. In: *International Cosmic Ray Conference (ICRC1997).* International Cosmic Ray Conference. 1997, pp. 265–268.

[59] N. Lubsandorzhiev. “The TAIGA-IACT camera: construction, calibration, performance”. In: *36th International Cosmic Ray Conference (ICRC2019).* Vol. 36. International Cosmic Ray Conference. 2019, 730, p. 730.

[60] N. Lubsandorzhiev, I. Astapov, P. Bezyazeekov, et al. “Camera of the first TAIGA-IACT: construction and calibration”. In: *35th International Cosmic Ray Conference (ICRC2017).* Vol. 301. International Cosmic Ray Conference. 2017, p. 757.

[61] D. Zhurov, O. A. Gress, I. Astapov, et al. “Software design for the TAIGA-IACT telescope pointing and control system”. In: *Proceedings of 35th International Cosmic Ray Conference — PoS(ICRC2017).* Vol. 301. 2017, p. 785. DOI: `10.22323/1.301.0785`.

[62] D. Zhurov, O. A. Gress, D. S. Lukyantsev, et al. “TAIGA-IACT pointing control and monitoring software status”. In: *Proceedings of 37th International Cosmic Ray Conference — PoS(ICRC2021).* Vol. 395. 2021, p. 690. DOI: `10.22323/1.395.0690`.

[63] W. A. Joye and E. Mandel. “New Features of SAOImage DS9”. In: *Astronomical Data Analysis Software and Systems XII.* Ed. by H. E. Payne, R. I. Jedrzejewski, and R. N. Hook. Vol. 295. Astronomical Society of the Pacific Conference Series. Jan. 2003, p. 489.

[64] E. Postnikov, I. Astapov, P. Bezyazeekov, et al. “Monte Carlo Simulation of the TAIGA”. In: *Experiment. Bull. Russ. Acad. Sci. Phys.* 83 (2019), pp. 955–958. DOI: `https://doi.org/10.3103/S1062873819080331`.

[65] C. R. Harris, K. J. Millman, S. J. van der Walt, et al. “Array programming with NumPy”. In: *Nature* 585.7825 (Sept. 2020), pp. 357–362. DOI: `10.1038/s41586-020-2649-2`. URL: `https://doi.org/10.1038/s41586-020-2649-2`.

[66] P. Virtanen, R. Gommers, T. E. Oliphant, et al. "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". In: *Nature Methods* 17 (2020), pp. 261–272. DOI: `10.1038/s41592-019-0686-2`.

[67] W. McKinney. "Data Structures for Statistical Computing in Python". In: *Proceedings of the 9th Python in Science Conference.* Ed. by S. van der Walt and J. Millman. 2010, pp. 56–61. DOI: `10.25080/Majora-92bf1922-00a`.

[68] J. Reback, W. McKinney, jbrockmendel, et al. *pandas-dev/pandas: Pandas 1.2.1.* Version v1.2.1. Jan. 2021. DOI: `10.5281/zenodo.4452601`. URL: `https://doi.org/10.5281/zenodo.4452601`.

[69] Astropy Collaboration, T. P. Robitaille, E. J. Tollerud, et al. "Astropy: A community Python package for astronomy". In: *A&A* 558, A33 (Oct. 2013), A33. DOI: `10.1051/0004-6361/201322068`. arXiv: `1307.6212 [astro-ph.IM]`.

[70] Astropy Collaboration, A. M. Price-Whelan, B. M. Sipőcz, et al. "The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package". In: *The Astronomical Journal* 156.3, 123 (Sept. 2018), p. 123. DOI: `10.3847/1538-3881/aabc4f`. arXiv: `1801.02634 [astro-ph.IM]`.

[71] A. Collette, T. Kluyver, T. A. Caswell, et al. *h5py/h5py: 3.2.1.* Version 3.2.1. Mar. 2021. DOI: `10.5281/zenodo.4584676`. URL: `https://doi.org/10.5281/zenodo.4584676`.

[72] R. Lessard et al. "A new analysis method for reconstructing the arrival direction of TeV gamma rays using a single imaging atmospheric Cherenkov telescope". In: *Astropart. Phys.* 15.1 (2001), pp. 1–18. DOI: `https://doi.org/10.1016/S0927-6505(00)00133-X`.

[73] T. Schweizer. "Analysis of the Large Gamma Ray Flares of Mkn 421 as Observed with HEGRA CT1 on the Island La Palma in 2001". PhD thesis. Universitat Autònoma de Barcelona, 2002. URL: `http://hdl.handle.net/10803/3343`.

[74] D. Heck, J. Knapp, J. N. Capdevielle, G. Schatz, and T. Thouw. *CORSIKA: a Monte Carlo code to simulate extensive air showers.* Feb. 1998.

[75] K. Bernlöhr. "Simulation of imaging atmospheric Cherenkov telescopes with CORSIKA and sim_telarray". In: *Astroparticle Physics* 30 (Oct. 2008), pp. 149–158.

[76] S. Bass, M. Belkacem, M. Bleicher, et al. "Microscopic models for ultrarelativistic heavy ion collisions". In: *Progress in Particle and Nuclear Physics* 41 (1998), pp. 255–369. DOI: `https://doi.org/10.1016/S0146-6410(98)00058-1`.

[77] M. Bleicher, E. Zabrodin, C. Spieles, et al. "Relativistic hadron-hadron collisions in the ultra-relativistic quantum molecular dynamics model". In: *Journal of Physics G: Nuclear and Particle Physics* 25.9 (Sept. 1999), p. 1859. DOI: `10.1088/0954-3899/25/9/308`. URL: `https://dx.doi.org/10.1088/0954-3899/25/9/308`.

[78] N. Kalmykov, S. Ostapchenko, and A. Pavlov. "Quark-gluon-string model and EAS simulation problems at ultra-high energies". In: *Nuclear Physics B - Proceedings Supplements* 52.3 (1997), pp. 17–28. DOI: `https://doi.org/10.1016/S0920-5632(96)00846-8`.

[79] K. Bernlöhr. "Impact of atmospheric parameters on the atmospheric Cherenkov technique". In: *Astroparticle Physics* 12.4 (2000), pp. 255–268. DOI: `https://doi.org/10.1016/S0927-6505(99)00093-6`.

[80] P. Mitra, A. Bonardi, A. Corstanje, et al. "Reconstructing air shower parameters with LOFAR using event specific GDAS atmosphere". In: *Astroparticle Physics* 123 (2020), p. 102470. DOI: `https://doi.org/10.1016/j.astropartphys.2020.102470`.

[81] P. Abreu, M. Aglietta, M. Ahlers, et al. "Description of atmospheric conditions at the Pierre Auger Observatory using the Global Data Assimilation System (GDAS)". In: *Astroparticle Physics* 35.9 (2012), pp. 591–607. DOI: `https://doi.org/10.1016/j.astropartphys.2011.12.002`.

[82] P. E. Ciddor. "Refractive index of air: new equations for the visible and near infrared". In: *Appl. Opt.* 35.9 (Mar. 1996), pp. 1566–1573. DOI: `10.1364/AO.35.001566`. URL: `https://opg.optica.org/ao/abstract.cfm?URI=ao-35-9-1566`.

[83] F. Kneizys, D. Robertson, L. Abreu, et al. *The MODTRAN 2/3 Report and LOWTRAN 7 MODEL*. Tech. rep. Contract F19628-91-C-0132. PL/GPOS, 29 Randolph Road, Hanscom AFB, MA 01731-3010: Phillips Laboratory, Geophysics Directorate, Jan. 1996.

[84] J. R. Hörandel. “On the knee in the energy spectrum of cosmic rays”. In: *Astroparticle Physics* 19.2 (May 2003), pp. 193–220. DOI: `10.1016/S0927-6505(02)00198-6`. arXiv: `astro-ph/0210453 [astro-ph]`.

[85] Q. An, R. Asfandiyarov, P. Azzarello, et al. “Measurement of the cosmic ray proton spectrum from 40 GeV to 100 TeV with the DAMPE satellite”. In: *Science Advances* 5.9 (Sept. 2019), eaax3793. DOI: `10.1126/sciadv.aax3793`. arXiv: `1909.12860 [astro-ph.HE]`.

[86] F. Alemanno, Q. An, P. Azzarello, et al. “Measurement of the Cosmic Ray Helium Energy Spectrum from 70 GeV to 80 TeV with the DAMPE Space Mission”. In: *Phys. Rev. Lett.* 126 (20 May 2021), p. 201102. DOI: `10.1103/PhysRevLett.126.201102`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.126.201102`.

[87] M. Gaug, S. Fegan, A. M. W. Mitchell, et al. “Using Muon Rings for the Calibration of the Cherenkov Telescope Array: A Systematic Review of the Method and Its Potential Accuracy”. In: *The Astrophysical Journal Supplement Series* 243.1 (July 2019), p. 11. DOI: `10.3847/1538-4365/ab2123`. URL: `https://dx.doi.org/10.3847/1538-4365/ab2123`.

[88] D. Hampf. “Study for the wide-angle air Cherenkov detector HiSCORE and time gradient event reconstruction for the H.E.S.S. experiment”. PhD thesis. Universität Hamburg, Fakultät für Mathematik, Informatik und Naturwissenschaften, Fachbereich Physik, 2012.

[89] J. Albert, E. Aliu, H. Anderhub, P. Antoranz, et al. “Implementation of the Random Forest method for the Imaging Atmospheric Cherenkov Telescope MAGIC”. In: *Nuclear Instruments and Methods in Physics Research A* 588.3 (Apr. 2008), pp. 424–432. DOI: `10.1016/j.nima.2007.11.068`. arXiv: `0709.3719 [astro-ph]`.

[90] F. Pedregosa, G. Varoquaux, A. Gramfort, et al. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[91] O. Grisel, A. Mueller, Lars, et al. *scikit-learn/scikit-learn: scikit-learn 0.24.1*. Version 0.24.1. Jan. 2021. DOI: `10.5281/zenodo.4450597`. URL: `https://doi.org/10.5281/zenodo.4450597`.

[92] L. Breiman. “Random Forests”. In: *Mach. Learn.* 45.1 (Oct. 2001), pp. 5–32. DOI: `10.1023/A:1010933404324`. URL: `https://doi.org/10.1023/A:1010933404324`.

[93] M. Wertz. “Entwicklung einer hybriden Luftschauerrekonstruktion für TAIGA”. MA thesis. Universität Hamburg, 2019.

[94] S. K. Lam, stuartarchibald, A. Pitrou, et al. *numba/numba: Version 0.55.2*. Version 0.55.2. May 2022. DOI: `10.5281/zenodo.6584302`. URL: `https://doi.org/10.5281/zenodo.6584302`.

[95] V. Prosin, S. Berezhnev, N. Budnev, et al. “Tunka-133: Results of 3 year operation”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 756 (2014), pp. 94–101. DOI: `https://doi.org/10.1016/j.nima.2013.09.018`.

[96] N. Budnev, A. Chiavassa, O. Gress, et al. “The primary cosmic-ray energy spectrum measured with the Tunka-133 array”. In: *Astroparticle Physics* 117 (2020), p. 102406. DOI: `https://doi.org/10.1016/j.astropartphys.2019.102406`.

[97] V. Stamatescu, G. Rowell, J. Denman, et al. “Timing analysis techniques at large core distances for multi-TeV gamma ray astronomy”. In: *Astroparticle Physics* 34.12 (2011), pp. 886–896. DOI: `https://doi.org/10.1016/j.astropartphys.2011.03.008`.

[98] V. P. Fomin, A. A. Stepanian, R. C. Lamb, et al. “New methods of atmospheric Cherenkov imaging for gamma-ray astronomy. I. The false source method”. In: *Astroparticle Physics* 2.2 (May 1994), pp. 137–150. DOI: `10.1016/0927-6505(94)90036-1`.

[99] C. Deil, R. Zanin, J. Lefaucheur, et al. “Gammapy - A prototype for the CTA science tools”. In: *35th International Cosmic Ray Conference (ICRC2017)*. Vol. 301. International Cosmic Ray Conference. Jan. 2017, 766, p. 766. arXiv: `1709.01751 [astro-ph.IM]`.

[100] A. Donath, C. Deil, R. Terrier, et al. *gammapy/gammapy: v.0.19*. Version v0.19. Nov. 2021. DOI: `10.5281/zenodo.5721467`. URL: `https://doi.org/10.5281/zenodo.5721467`.

[101] M. Nöthe, M. Peresano, J. Sitarek, et al. *cta-observatory/pyirf: v0.6.0 –2022-01-10*. Version v0.6.0. Jan. 2022. DOI: `10.5281/zenodo.5833284`. URL: `https://doi.org/10.5281/zenodo.5833284`.

[102] S. S. Wilks. "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses". In: *The Annals of Mathematical Statistics* 9.1 (1938), pp. 60–62. DOI: `10.1214/aoms/1177732360`. URL: `https://doi.org/10.1214/aoms/1177732360`.

[103] M. de Naurois. "Very High Energy astronomy from H.E.S.S. to CTA. Opening of a new astronomical window on the non-thermal Universe". PhD thesis. Ecole Polytechnique, 2012.

[104] H.E.S.S. Collaboration, H. Abdalla, A. Abramowski, et al. "The H.E.S.S. Galactic plane survey". In: *A&A* 612 (2018), A1. DOI: `10.1051/0004-6361/201732098`. URL: `https://doi.org/10.1051/0004-6361/201732098`.

[105] J. Aleksić, E. Alvarez, L. Antonelli, et al. "Performance of the MAGIC stereo system obtained with Crab Nebula data". In: *Astroparticle Physics* 35.7 (2012), pp. 435–448. ISSN: 0927-6505. DOI: `https://doi.org/10.1016/j.astropartphys.2011.11.007`. URL: `https://www.sciencedirect.com/science/article/pii/S0927650511002064`.

[106] M. Blank. *Resources for thesis: Detection of the Crab Nebula with the first TAIGA IACT*. June 2023. DOI: `10.25592/uhhfdm.12554`. URL: `https://doi.org/10.25592/uhhfdm.12554`.

Acknowledgements

First of all, I would like to express my sincere gratitude to Prof. Dr. Dieter Horns for giving me the opportunity to work in his group and for providing an environment with the freedom to pursue own ideas. His expertise and insightful feedback greatly helped to shape the path of this endeavor. I am truly thankful for his guidance and support.

I would like to express my heartfelt gratitude to Dr. Martin Tluczykont for his assistance and mentorship as my project lead. His knowledge and expertise have formed the outcomes of this work, and I am deeply thankful for his insights and fruitful coffee time discussions. His support and encouragement have been vital throughout the whole time.

I am deeply grateful to Prof. Dr. Yvonne Becherini for evaluating this thesis.

Furthermore, I would like to thank Ali Baktash for insightful discussions in the office. I would also like to thank my officemates Delia, Josep and Dr. Ludmilla Dirson and also Dr. Le Hoang Nguyen and the other members of the group for the positive atmosphere and for all the academic and non-academic chatter. Thank you for providing such a nice time and for the invaluable BBQ sessions.

I am indebted to Le Hoang Nguyen, Charlotte Hartmann, Nina and especially Martin Tluczykont and Ali Baktash for proofreading.

The members of the gravitational wave group should also not go unmentioned: Thank you for the enjoyable lunch break debates.

I am grateful to my parents for their unwavering love and support.

Furthermore, I am grateful to Nina for her support, patience and understanding, especially in the moments of self-doubt.

Finally, I thank all the members of the TAIGA collaboration for making this project possible, for their helpfulness and for valuable discussions. This work was supported by the Deutsche Forschungsgemeinschaft (DFG, TL 51/6-1). I acknowledge support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy — EXC 2121 "Quantum Universe" — 390833306.

Eidesstattliche Versicherung

Declaration on oath

Hiermit erkläre ich an Eides statt, die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Hilfsmittel und Quellen benutzt zu haben.

Hamburg den
18. Juni 2023
Unterschrift

Michael Blank