Universität Siegen / Naturwissenschaftlich-Technische Fakultät

MASTERARBEIT

Hybrid Search for Photons with the Low-Energy Extensions of the Pierre Auger Observatory

Zur Erlangung des akademischen Grades **Master of Science**
(M. Sc.) im Studiengang Physik

von

Tim Lukas Fehler

eingereicht an der

Universität Siegen
Naturwissenschaftlich-Technische Fakultät
Department Physik

Siegen, den 10. Januar 2024

Betreuer & Erster Gutachter	Prof. Dr. Markus Risse
Zweiter Gutachter	Prof. Dr. Markus Cristinziani

Contents

Introduction

Every second, an enormous number of highly energized particles originating from outer space strike the Earth's atmosphere. Fortunately for life on Earth, the flux decreases rapidly towards higher energies and the geomagnetic field and upper atmospheric layers provide a rather effective protection from this constant radiation. In the case of solar particles, this protection is the reason for the stunning phenomenon of the aurora borealis/australis (see Fig. 0.1). However, the origin of the higher energy radiation evidently lies outside the solar system. The so called cosmic-ray particles are composed of a variety of subatomic particles, including protons, electrons, and even atomic nuclei and can possess a wide range of energies.

Fig. 0.1 Auroral beads seen from the International Space Station on Sept. 17, 2011. [1].

With recent experimental and technological advancements, there has been a growing interest in studying the ultra-high-energy (UHE[1]) regime of the cosmic-ray spectrum, i.e. above approximately 10^{17} eV. While the basic production processes for low energies are relatively well-known, the sources of the elusive UHE cosmic rays still remain a mystery. Part of the problem lies in the fact that cosmic rays are composed of electrically charged particles, which therefore experience substantial deflection in interstellar and extragalactic magnetic fields, losing most directional information. Of particular interest, hence, are neutral particles, i.e. without electric charge, emitted in association with the primary cosmic radiation near their acceleration sites or during propagation. The (aim for) observation of neutral particles, among them photons, forms an important building block in multi-messenger astronomy and may potentially lead to a deeper understanding of many astrophysical phenomena in the future.

Air-shower arrays, among them the Pierre Auger Observatory, which were primarily designed to efficiently observe cosmic rays at the highest energies, have proved to be capable of reliably detecting UHE photons. The diverse detector systems of the Observatory have been used to impose stringent upper limits on the diffuse, integral flux of UHE photons across several orders of magnitude in energy.

This thesis was written as part of the work on an update to the currently leading photon search for cosmic rays in the 10^{17} to 10^{18} eV range [2], which is based on the low-energy extensions of the Pierre Auger Observatory.

[1] As an aid to the reader, a list of acronyms and abbreviations used throughout this thesis is provided in Section B of the Appendix.

Physical and Technical Background

The following chapter will lay the necessary groundwork to commence with the study of this thesis. It starts with a concise review of the early historical development of astroparticle physics and its connection to experimental particle physics (Sec. 1), after which the current state of knowledge about cosmic rays is presented (Sec. 2). The mechanisms leading to extensive air showers will be detailed (Sec. 3), before moving on to their detection. The design of the Pierre Auger Observatory is discussed in detail (Sec. 4). Finally, this part concludes with a brief overview of the topic of ultra-high-energy (UHE) photons and their identification as primary particles in air showers using the Observatory's detectors (Sec. 5).

1. The Early Days of Cosmic-Ray Physics

1.1. Discovery of Cosmic Radiation

The history of cosmic rays began with the observation of the remnant conductivity of air more than 200 years before this thesis was written, in the late 18th century. Charles-Augustin de Coulomb described the naturally occurring gradual discharge of an electroscope in an electrically insulated enclosure [3], which could not be ascribed to ambient humidity. Coulomb attributed this phenomenon to dust particles in the surrounding air, but his explanation was considered inadequate, since the discharge process was not impeded by the insulation of the electroscope against leakage currents to the outside.

The search for a better explanation continued for another century, when the phenomenon of radioactivity and its associated ionizing radiation was discovered by Henri Becquerel and Marie Skłodowska-Curie in 1896 [4], while studying phosphorescent materials. Terrestrial radioactivity, i.e. the radiation from naturally occurring radioactive ores in the Earth's crust, seemed to be a plausible culprit for the measured ionization values of the air. Radioactive elements emit charged particles, which ionizes the gas therein, thus causing the discharge of electroscopes. Since it was assumed that the vast majority of the ionization is caused by natural radioactivity originating from the upper layer of the Earth's crust, a theoretical decrease with height above ground was calculated [5]. The newly constructed Wulf two-string electrometer allowed its inventor and eponym Theodor Wulf much more accurate measurements and was quickly adopted by the scientific community as the new standard tool [6]. Using his apparatus, Wulf performed systematic studies regarding the origin of the γ-radiation in the atmosphere, including taking measurements on top of the Eiffel Tower in Paris [7, 8]. He measured a minor but not significant decrease in the ionization values several hundred meters above sea level, contrary to the expectations for a pure soil based origin. The meteorologist Domenico Pacini, also made extensive measurements of radioactivity over the ground and over the oceans and concluded from his data and earlier investigations that an appreciable part of the penetrating radiation has an origin independent of the direct action of radioactive substances in the soil and was the first to

propose an extraterrestrial origin for a part of the ionizing radiation. He estimated the excess ionization to be 2 ion pairs per cm^3 per s, which is about the amount of ionization due to cosmic rays at sea level as we know it today [9, 10].

The initial measurements were conducted ground-based, but in the following years together with technological advancements the era of balloon measurements began. The first use of balloons was made in 1904 under the direction of Franz Linke in Göttingen [11]. In addition, Wulf and Albert Gockel worked together in 1907 to measure the ionization rate on various peaks in the Swiss Alps [12]. It was Gockel who later used balloons to measure the ionization rate at altitudes up to 4500 m [13]. He was able to observe a decrease in the values, but also had to deal with major problems with his instruments, which were not adapted to the atmospheric conditions at such high altitudes.

Victor Franz Hess succeeded in improving the accuracy of Wulf's electrometers by an order of magnitude by calibrating them with radium sources [15]. He also used pressure-tight electrometers to overcome the issues encountered by Gockel and others during high altitude balloon flights. With this measurement setup, significant improvements were now expected. In 1912 the ionization rate was recorded by Hess during seven balloon flights to an altitude of 5 km, where, after a brief decline immediately above ground, the values increased significantly [16]. At the apex of the flight Hess noted ionization levels up to a factor of four compared to the ground in both of his independent detectors. The experiment was set up in a way to minimize the impact of radioactive contamination and excluded the sun as a source candidate due to lack of day-night modulations. Hess proposed a new kind of radiation from outside the atmosphere, not connected to solar activity, as the primary source of the phenomenon. Hardly acknowledged at first, his findings were later independently confirmed by Werner Kohlhörster, who flew balloons to altitudes exceeding 9 km, measuring even higher ionization levels (see Fig. 1.1) [14]. International recognition finally came in 1936 with the Nobel Prize in Physics, awarded to Hess "for the discovery of cosmic radiation" [17].

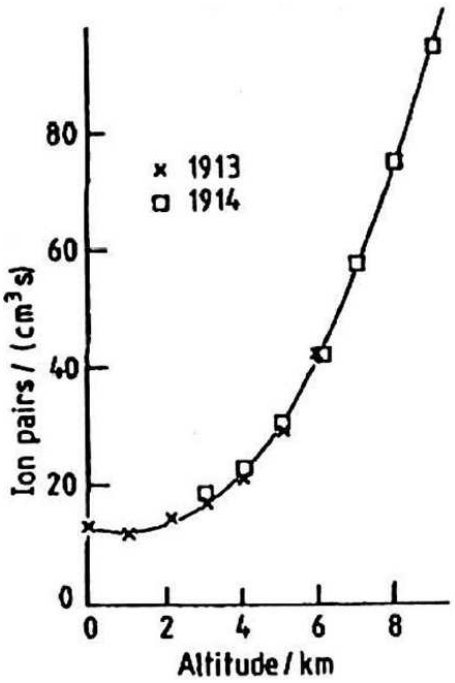

Fig. 1.1 Kohlhörster's measurement of ionization vs altitude during two balloon flights in 1913 and 1914 following and extending the discoveries of Hess [14].

The early explorers refereed to this radiation by the German terms *Höhenstrahlung* (high-altitude radiation) or *Ultrastrahlung* (ultra radiation). Then, in 1926, while attempting to prove this radiation was γ-rays from the nucleosynthesis of the common C and O elements, Robert Andrews Millikan coined the now more common term *cosmic rays* [18–20].

1.2. Extensive Air Showers

Another important step in developing understanding of the true nature of cosmic rays was the discovery of extensive air showers. By the 1930s, research was well advanced and some basic properties, such as the corpuscular and electrically charged nature of cosmic rays, had been determined. In 1935, Erich Regener and Georg Pfotzer studied the intensity of vertically incident cosmic rays with triple coincidence using three Geiger-Müller counting tubes [21]. They were able to reproduce the increasing radiation intensity with altitude already known from the early balloon experiments. However, when converted to vertical incidence, they also found a radiation maximum at about 15 km altitude. This "Pfotzer maximum" was consistent with what one would expect if the cosmic rays measured here at the Earth's surface and in the lower atmospheric layers were actually just a byproduct of collisions of primary cosmic rays with atomic nuclei in the upper atmosphere, similar to the particle avalanches caused by high energy radiation in lead plates that were already known from experiments by Bruno Benedetto Rossi [22]. In his writings, Rossi used the Italian term *sciami* for this avalanche phenomenon, which was later transferred to English as *shower* by Patrick Blackett [23]. The phenomenon of particle cascades received a theoretical foundation with the work of Homi Jehangir Bhabha and Walter Heitler on electromagnetic showers [24], based on the ideas of the by then developing theory of Quantum Electrodynamics (QED).

However, Pierre Victor Auger is credited with the actual experimental discovery of extensive air showers triggered by cosmic rays. In 1938, together with his collaborators, he installed spatially separated Geiger-Müller particle counters with improved time resolution at several high-altitude mountain sites, including the Jungfraujoch (3500 m a.s.l.) in Switzerland [25, 26]. He measured the coincidence rate at various distances separating the detectors and was able to confirm coincidences up to 300 m that clearly exceeded the expectations from chance coincidence. From his measurements, Auger estimated that the total number of particles in such showers can easily reach 10^6 and thus the energy spectrum of the primary cosmic-ray particles must extend at least up to 10^{15} eV [27]. This represented a great sensation at the time, because even today, almost a century later, the most powerful particle accelerator built by humans, the Large Hadron Collider (LHC) situated below Geneva at CERN, does not reach energies of this kind. It was also Auger who in his works established the term *extensive air showers* for the particle cascades of this order of magnitude.

1.3. Driving Force of Early Particle Physics

Looking back, the early development of cosmic-ray physics can be seen as the cradle of modern high-energy particle physics. With the newly discovered high-energy cosmic rays, a natural particle accelerator and collider was available for researchers to probe even deeper into the fundamental constituents of matter than had previously been possible. However, this also required novel methods of particle detection and consequential identification. A key milestone for the field of high-energy particle physics was the invention of the cloud chamber by Charles Thomsom Ress Wilson in 1912, which enabled the trajectories of ionizing radiation to be observed by eye alone [28]. Through the application of external electric

and magnetic fields, it became possible to investigate the properties of the particles and form a classification. Since the conditions for operation were not continuously maintained in the original "pulsed" chambers, the researchers quickly realized that the poor yield of particle tracks by random expansion could be significantly improved by triggering the chamber with an array of Geiger-Müller counters [9]. This made it possible to repeatedly and reliably observe cosmic-ray tracks, many of them showers. Later implementations of the concept of the cloud chamber increasingly circumvented the problem through advanced designs.

The Wilson chamber allowed the first observation of the positron, the electron's counterpart in the antimatter domain, by Carl David Anderson in 1932 (see Fig. 1.2) [29, 30], which earned him the Nobel Prize in Physics in the same year as Victor Franz Hess [17]. This initial discovery followed a rich abundance of newly established particles in the subsequent years. It was also Anderson, together with Seth Henry Neddermeyer, who, in 1937, discovered the muon μ in cloud chamber pictures of cosmic radiation [31]. This was later followed by the neutral and charged kaons $K^{0,\pm}$ in 1947 [32] and the Λ^0 [33], Ξ [34] and Σ [35] hyperons in the early 1950s by different researchers.

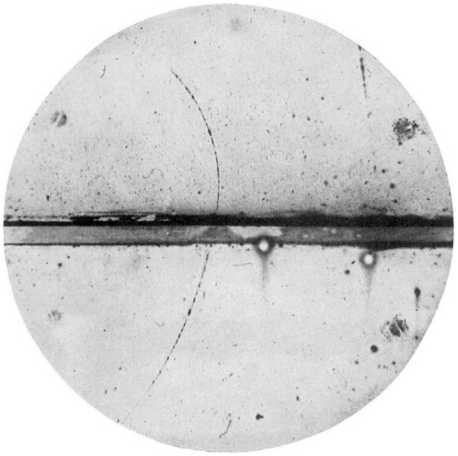

Fig. 1.2 Anderson's famous photograph of the positron trajectory in a Wilson chamber [29].

Eventually, in the middle of the 20th century, the field divided into two groups: The analysis of the variety of newly discovered particles moved to the man-made particle accelerators, while the research on the origin and astrophysics of the primary cosmic radiation continued in the newly established field named *astroparticle physics*. Many of the experiments followed the technological advances and relocated satellite-based into space, with one of the exceptions being the study of extensive air showers, which is primarily driven by the use of large-scale ground-based detector arrays, such as the Pierre Auger Observatory discussed in Section 4.

2. Cosmic Rays

The measurements of cosmic rays have come a long way from the early days of discovery and speculation. Today we know that primary cosmic rays are high-energy and stable particles that arrive at Earth after originating from unknown astrophysical sources. They consist of protons, fully ionized atomic nuclei and electrons, whereby the latter are easily decelerated by their motion in local magnetic fields and thus only factor in when considering close sources. Unstable particles and isotopes can also generally be neglected due to the extended propagation time of cosmic rays. Interactions of the main radiation near (or close) to the acceleration sites can produce secondary charged or neutral particles, among them photons and neutrinos, which are stable and may arrive at Earth in association with

the primary component. All the particles produced during the propagation of the primary component and upon impact in the atmosphere are generally denoted as secondary cosmic rays. However, in some cases, the terminology is inconsistent, and the particles that initiate the first interaction of an extensive air shower in the atmosphere are referred to as primaries, since their origin is not relevant for the air shower considered.

The next subsections aim to present the generally accepted current state of knowledge on cosmic rays, but at the same time to show which contemporary challenges and fundamental issues still remain to be solved and what role the search for UHE photons takes in this context.

2.1. Energy Spectrum

The energy spectrum of primary cosmic rays is well known today and has been measured in numerous experiments over many orders of magnitude. It extends from energies below 1 GeV to beyond 100 EeV ($= 10^{20}$ eV). Above some 10 GeV, where effects of the geomagnetic field and solar activity can be neglected, the all-particle differential flux quite accurately follows a smooth power law of the form

$$\frac{\mathrm{d}\phi}{\mathrm{d}E} \propto E^{\gamma}, \tag{2.1}$$

with an approximate spectral index of $\gamma \simeq -3$. Due to the steep spectral index γ, the flux covers a range of over 30 orders of magnitude. It extends from several thousand particles per square meter per second around 1 GeV to less than one particle per square kilometer per year above 1 EeV (see Fig. 2.1).

Experiments that involve balloons or satellites are typically utilized to measure the lower end of the spectrum ($\lesssim 1$ PeV). Direct measurements become unfeasible above this energy region, as the flux decreases significantly, due to the steep spectral index, making it increasingly difficult to gather sufficient statistics. Therefore, above this energy, indirect observation through extended air showers is relied upon almost exclusively. On the ground, considerably larger areas can be utilized as a base for detection. Detector components are commonly dispersed throughout arrays covering square kilometers, measuring individual air showers with sufficient precision. Using this data, the properties of the primary particle responsible for initiating the air shower are subsequently inferred. There are methods available for reconstructing the primary particle's type, energy, and arrival direction solely based on the signals detected from the secondary particles within the shower.

2.1.1. Structures in the Spectrum

Upon closer examination of the energy spectrum, it becomes evident that despite the flux closely following a simple power law, there are additional structures present. At various distinct positions along the spectrum, the spectral index γ changes its value noticeably. By multiplying the flux with E^3, these variations are accentuated and become apparent (see Fig. 2.2). There are at least four such features that have been generally agreed upon.

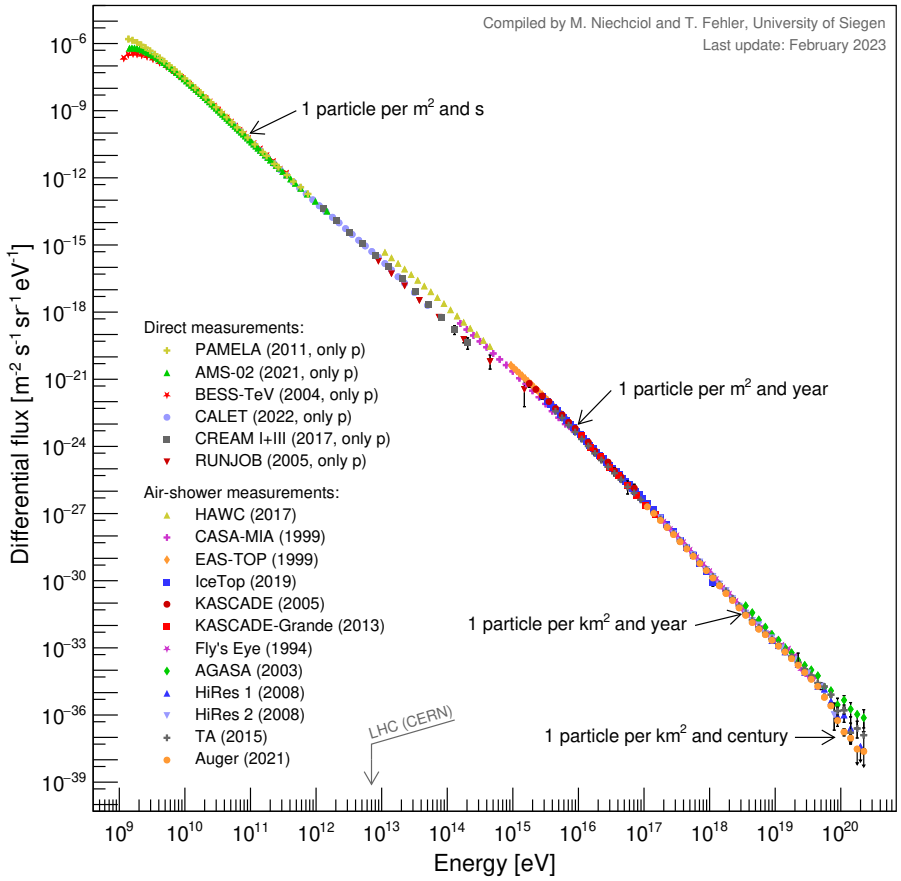

Fig. 2.1 Energy spectrum of cosmic rays as measured by various experiments. For the direct measurements only the flux of protons is considered.

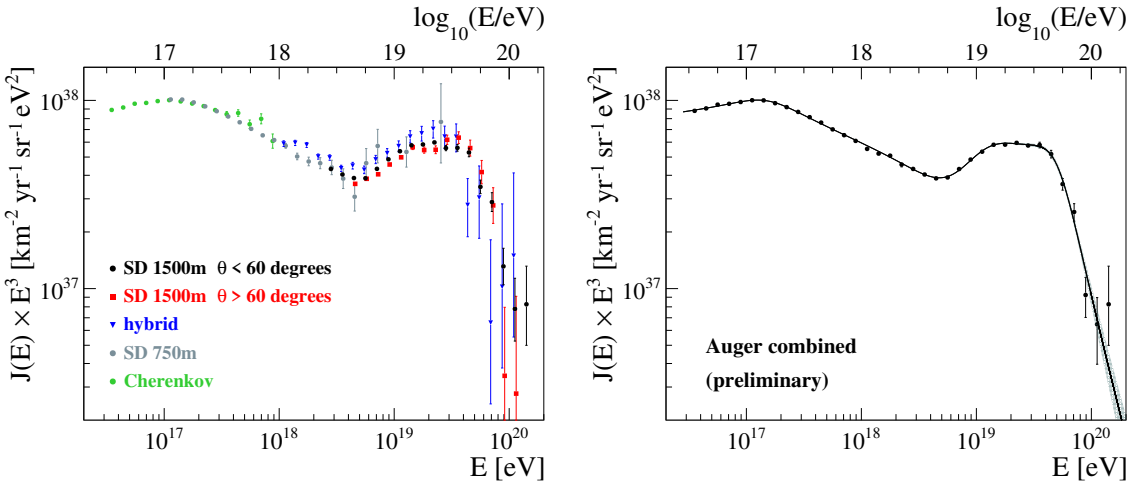

Fig. 2.2 Energy spectra measured with the various detectors of the Pierre Auger Observatory above $10^{16.5}$ eV (left) and corresponding spectrum obtained by combining the different measurements (right). In both plots, the spectrum has been multiplied by E^3 to enhance the structures along it [36].

Their notation refers to the course of the spectrum and borrows from the shape of the human leg. The spectral index can be considered relatively constant between these points, although there is evidence for further minor features in the spectrum [37, 38].

Note, that different nuclear species that are present in the cosmic-ray flux demonstrate spectral characteristics at varying energies. However, especially in the high-energy range, it becomes difficult to decompose individual components due to the indirect nature of the observation, therefore only the all-particle spectrum will be discussed here for the time being.

Knee

At about $4 \times 10^{15}\,\mathrm{eV}$, the spectral index γ transitions from -2.7 to -3.1. This feature is commonly referred to as the *knee*. Several theories exist to explain the existence of this feature, among them some who link the knee to a change in the acceleration mechanism at this point in the flux. There may be a charge-dependent upper limit for cosmic ray emission from galactic sources and accelerators and a subsequent shift to an extragalactic origin near this energy. The current state of knowledge on possible sources and acceleration mechanisms is presented in more detail in Section 2.3. Scenarios are also conceivable in which the feature is caused by the influence of single dominant sources in the local galactic vicinity or the presence of previously unknown effects during propagation [39, 40].

Second Knee

The flux steepens again at approximately $4 \times 10^{17}\,\mathrm{eV}$, resulting in a spectral index of $\gamma = -3.3$ above this *second knee*. The name reflects the widely accepted assumption that this feature is closely related to the former one. Assuming that the cosmic-ray flux is composed as a superposition of the fluxes of the individual elementary components, and that these are subject to a charge-dependent upper limit of acceleration or propagation ability, then the second knee could be associated with the heaviest nuclei, while the first knee could be attributed to the lightest elements, starting with the proton [41]. However, the verification of this theory from the measured data is insufficient and there are still some unresolved discrepancies. Moreover, there also exist alternative explanations for both of the knee features [41].

Ankle

At an energy of $5 \times 10^{18}\,\mathrm{eV}$, the spectrum hardens again to a spectral index of $\gamma = -2.6$. This point is denoted as the *ankle* of the spectrum. Again, the exact cause is not known with complete certainty. The simplest and traditional explanation is related to the presumed transition from galactic to extragalactic sources at the two knee features. Assuming a flat source spectrum of extragalactic cosmic rays, the ankle may form the point of transition. From there, this previously subdominant harder component takes over and governs the overall shape [42].

In addition, the concave feature of the ankle could be explained by considering an associate particle physics process that may occur during the propagation of cosmic rays. In the dip

model [43], the interaction of the proton component of the flux with photons from the Cosmic Microwave Background (CMB) leads to the observed attenuation at the energy level of the ankle. The CMB, a remnant of the primordial universe, can be regarded as the thermal afterglow of the big bang, filling the entire observable universe with an extremely homogeneous microwave radiation with a black-body spectrum corresponding to a temperature of 2.725 K. The process of Bethe-Heitler pair production

$$p + \gamma_{\mathrm{CMB}} \rightarrow p + e^{-} + e^{+} \tag{2.2}$$

is suspected to cause the flattening of the spectrum in the energy range of the ankle. However, this also presupposes a proton-dominated composition at and beyond the ankle, for which sufficient evidence is lacking [44].

Instep

A subtle steepening in the all-particle spectrum around 10^{19} eV has been recently detected by the Pierre Auger Observatory [45]. Possible explanations for this observation include a distinctive spectrum from a local source emitting protons, which significantly contributes to the overall intensity. However, local anisotropies are notably absent in this energy range. Moreover, spectral differences would be anticipated to manifest in certain areas of the sky, yet such measurements are currently lacking. Consequently, any contribution of this nature is ruled out by the limits imposed on extragalactic magnetic fields, as they tend to obscure conspicuous features in the arrival direction of cosmic rays. Scenarios involving multiple such proton-emitting sources, while simultaneously mitigating any directional effects, would necessitate substantial fine-tuning and are hence considered less likely. Additionally, compositional fit results indicate contributions from various nuclear components to the total intensity at this energy level.

Cutoff

Above an energy of approximately 4×10^{19} eV, measurements [46, 47] confirms a significant suppression of the flux. There is a potential exponential cutoff at 2×10^{20} eV. However, at the highest energies, the size of the measured cosmic-ray data set becomes so small that the situation remains unclear. Such a strong suppression was postulated already in 1966 by Kenneth Greisen [48] and independently by Georgiy Zatsepin and Vadim Kuz'min [49] as the logical consequence of the experimental discovery of the CMB. The theorized GZK effect, named after the conceptual fathers, concludes a possible interaction between CMB photons and UHE protons from the cosmic ray flux. Above a threshold of approximately 5×10^{19} eV, some of the energy of individual protons is carried away through pion production, with the $\Delta(1232)^{+}$ resonance acting as an intermediate state:

$$p + \gamma_{\mathrm{CMB}} \rightarrow \Delta(1232)^{+} \rightarrow \begin{cases} p + \pi^{0} \\ n + \pi^{+} \end{cases} \tag{2.3}$$

This would lead to an energy-cutoff of the cosmic-ray proton spectrum from distant sources, since the universe becomes progressively opaque for cosmic-ray protons with increasing

energy, imposing an effective GZK horizon for ultra-high-energy cosmic rays (UHECR). According to this model, the mean free path length for a 10^{20} eV proton would be limited to approximately 100 Mpc [50]. A similar effect occurs for heavy nuclei, but at a slightly higher energy threshold. Although the GZK effect has not been experimentally verified, it would produce a considerable amount of secondary neutral particles, including photons and neutrinos, by decaying high-energy pions. Therefore, searching for UHE photons is also a way to test the nature of cosmic ray flux suppression at the highest energies.

Another plausible explanation is the postulation of an upper limit in the acceleration capabilities of the UHECR sources. Such a maximum would likely be charge dependent and thus the highest energies in the range of 2×10^{20} eV could only be reached by iron nuclei [51]. This upper limit suppresses the flux without triggering the onset of the GZK process. Consequently, there should be a lower flux of secondary UHE neutrinos and photons.

A detailed survey of the uppermost end of the spectrum, in particular regarding the elemental composition, together with a targeted search for UHE neutrinos and photons, could serve as a discriminator between various cosmic-ray models and provide clarity in the future whether the flux suppression is the result of the GZK effect, an upper limit on the sources' accelerating capabilities, a combination of both or an entirely different phenomenon.

2.2. Composition

It is known that cosmic rays consist of charged particles and not of γ-rays, as some had expected, at least since Arthur Holly Compton's research and measurement expeditions at different geomagnetic latitudes in the 1930s [52]. Their sign of charge, i.e. the fact that cosmic-ray particles are predominantly positively charged, was established with the observation of the *east-west effect* [53]. Considering the direction of incidence from lower-energy cosmic-ray particles, one finds a higher intensity from the west compared to the east. This relates to the fact that some possible trajectories from easterly directions do not reach out into space, due to the curvature imposed by the influence of the geomagnetic field, which reduces the overall intensity from these directions.

However, the state of knowledge has improved considerably with the extensive measurements over recent decades. Primary cosmic rays mainly encompass protons ($\approx$ 85%), helium nuclei ($\approx$ 12%), and a minute fraction of heavier elements ($\approx$ 3%). Electrons, with a flux about two orders of magnitude below the all-particle spectrum, are also detected at the lower end of the spectrum, primarily originating from solar wind contributions. There even is a tiny fraction of antiparticles, mainly positrons, measured in the cosmic ray flux. Their abundance is consistent with the expectations from presumed interactions with interstellar gas and pair production occuring during the propagation phase of primary cosmic rays [54].

Given the potential of direct measurements, the composition at lower energies, up to about 10^{14} eV, is relatively well known. The elemental composition for nuclei in the cosmic-ray flux is comparable with the abundance of elements in the solar system, which indicates that

the dominant production process for the elements in cosmic rays is stellar nucleosynthesis. Certain deviations from this fundamental pattern can be accounted for by fragmentation and spallation processes during the propagation phase [54].

In the pursuit of energies surpassing 10^{14} eV, the survey of cosmic rays via direct detection is becoming increasingly difficult. A transition towards indirect observation of extensive air showers in the UHE realm is imperative, and, consequently, the accuracy of primary-particle identification degrades due to the unavailability of direct access and the need to discriminate solely on the basis of the characteristics exhibited by the air showers (see Sec. 3). Furthermore, due to the rapidly diminishing particle flux and the associated progressively smaller statistics, making reliable predictions about the elemental composition, especially at the highest energies beyond 10^{20} eV, becomes challenging. As a result, the exact composition of cosmic rays at energies above 10^{14} eV is to large extent unknown territory and the subject of ongoing research. However, there is certainly circumstantial evidence pointing to a heavier mixed mass composition at the higher energies $> 10^{17}$ eV (see e.g. [44, 55, 56]).

Neutral particles, including photons, neutrinos, and neutrons, are typically not categorized as primary cosmic rays. Nevertheless, their presence is intricately linked to the acceleration and propagation of UHE cosmic rays near their point of origin. Specifically, within this framework, high-energy photons and neutrinos arise as decay products of pions (and muons), which originate either directly from the sources or emerge during the journey of charged cosmic rays through interactions with the surrounding background fields. The observation of neutral particles can provide additional insights into the investigation of charged cosmic rays and contribute to the identification of their sources. In instances where direct observation is not feasible, these primary particles are discerned through the characteristic signatures in the individual extensive air showers they initiate within the atmosphere. In the context of photons, the specific attributes associated with these events are discussed in greater detail in Sec. 5.2.

2.3. Origin and Acceleration of Cosmic Ray Particles

Since cosmic rays are subject to chaotic deflections in galactic and intergalactic magnetic fields, the observed events do not point directly back to their sources. The identification of the cosmic-ray sources is hence experimentally challenging and has so far proved inconclusive. From a theoretical point of view, the quest to uncover the origin of cosmic rays revolves around two core inquiries: Firstly, what inherent mechanisms accelerate these charged particles to the energies we currently observe? Secondly, within which astrophysical settings or objects can the pre-established theoretical prerequisites for such extreme acceleration be realized? The spectral shape indicates that the acceleration of cosmic rays is not of thermal origin. An active acceleration mechanism is required to maintain a power-law dependence over several orders of magnitude. For the energy transfer process in general, two types of models are viable:

Top-Down Models

In top-down models, energetic cosmic rays are produced in the decay of hypothetical massive particles, topological defects, domain walls, or cosmic strings, which could be relics of the Big Bang, circumventing the problem of acceleration. These particles are, however, subject to certain restrictions, which none of the known elementary and composite particles of the Standard Model (SM) can fulfill. First of all, such particles would have to possess a rest mass $\gg 10^{20}\,\mathrm{eV}/c^2$ and, in addition, have to be distributed approximately homogeneously in the universe with a suitable density in order to explain the uniformly distributed particle flux, without, however, having been discovered so far. Therefore, in the context of dark matter searches one usually refers to them as Weakly Interacting Massive Particles (WIMPs). If such particles are remnants of the Big Bang, their lifetime must be of the order of the age of the universe, or they must be continuously reproduced. Some models link these hypothetical particles to a subset of dark matter, in which case stronger anisotropies in the direction of arrival would be expected than are observed today [57].

A common feature of top-down models is their prediction of a substantial flux of UHE photons due to the decay of supermassive particles. However, existing upper bounds on the integral photon flux in the UHE regime severely constrain the viability of most top-down models. In fact, these bounds are stringent enough to reject certain models altogether (see e.g. [58, 59]). As a result, top-down models tend to be less preferred than bottom-up models today.

Bottom-Up Models

In the more conventional bottom-up models, the energy is transferred to the cosmic-ray particle by an acceleration process in a unique astrophysical environment. This can manifest as either a single acceleration event or through a series of stochastic processes that on average yield an energy gain for the particle. Assuming a simple one-shot acceleration with a conventional electric field, potential drops of up to $10^{20}\,\mathrm{V}$ are only possible in the most extreme astrophysical environments, e.g. pulsars, where, however, unfavorable conditions prevail at the same time, since the cross-section for energy loss processes is comparatively high. For lower energies (up to the GeV range), on the other hand, this process may well be applicable even at normal stars.

Within the realm of UHECRs, the progressive acceleration in a recurring stochastic process finds greater applicability. The inception of this concept can be traced back to 1949, when Enrico Fermi introduced the idea in his work on the origin of cosmic radiation [60]. The process Fermi discussed relied on the reoccurring deflection of the individual particle with a magnetized cloud of interstellar matter. The average energy gain of a single interaction is then proportional to u^2, where u is the velocity of the magnetized cloud relative to a stationary observer. Due to this quadratic dependence, this variant is often called Fermi mechanism of 2nd order. Even though this process appears capable of accelerating cosmic rays up to ultra-high energies and results in a power-law spectrum of particle energies, its big drawback is the inefficiency due to the usually non-relativistic cloud velocities ($u \ll 1$) and the quadratic dependence, resulting in very small energy gains per individual collision and hence long acceleration times. To be viable for particle acceleration, considering the

intermediate energy loss due to interactions with interstellar or intergalactic gas between two collisions, this mechanism requires an effective minimum injection energy.

A more efficient way of energy transfer was theorized in 1978, with the stochastic acceleration on relativistic shock fronts [61]. Such shock fronts are commonly observed in supernova remnants (SNRs) as well as various other astrophysical objects. High-energy particles may get scattered back and forth between the upstream and downstream regions of the shock front by magnetic irregularities, obtaining a statistical amount of energy at each pass. This energy gain is directly proportional to the shock front velocity, denoted as u, leading to the nomenclature of this mechanism as the *first-order Fermi mechanism*. Similarly to the former one, this mechanism naturally yields a power-law distribution for the energy spectrum of the accelerated particles. The exact spectral index is solely dependent upon the shock compression ratio [61]. The inherent linearity in u leads to an improved efficiency in the energy transfer. Consequently, this mechanism is favored in describing cosmic-ray acceleration within the realm of high-energy particles.

2.3.1. Source Candidates

At the low end of the energy spectrum, the sun is evidently the clearest cosmic-ray source in the vicinity of the Earth. Under realistic assumptions, merging sunspots and solar flares exhibit suitable conditions to obtain particle energies in the GeV range [54], as mentioned in the previous section.

For cosmic rays of higher energies, the picture is not as clear. Since they show no correlation with the sun, they evidently originate from outside the solar system. The identification of individual sources, however, is hindered by the substantial deflection imposed by uncharted (extra)galactic magnetic fields. Primary particles of energies up to the ankle are commonly attributed to galactic sources, with SNRs generally thought to be viable source candidates [62], although direct evidence is still lacking. However, theorized to lay in the extragalactic domain, the origin of cosmic-ray particles beyond the ankle is still largely unresolved and remains a subject of ongoing research.

A valuable indication for the search may be the observation of anisotropies in the arrival directions of incoming cosmic-ray particles of a certain energy level and elemental composition. Even though the arrival distributions of cosmic rays have been shown to be remarkably isotropic, several experiments have found small but significant excesses of events at various angular scales. For energies in the TeV to PeV range, a dipolar structure was observed. While for the lower energies its right ascension phase is oriented parallel to the local magnetic field [63], at about 200 TeV the direction flips to alignment with the galactic center direction (see e.g. [64]). This is consistent with the expectation in the case of isotropic diffusion with a smooth distribution of sources on the galaxy, since the diffusive dipole is expected to simply align with the highest concentration of sources in the sky, i.e. towards the galactic center. Another potential origin of a weak dipolar anisotropy lies in the *Compton-Getting effect*. This effect arises from the relative motion of an observer in relation to the frame in which cosmic ray distributions appear isotropic [65]. In this

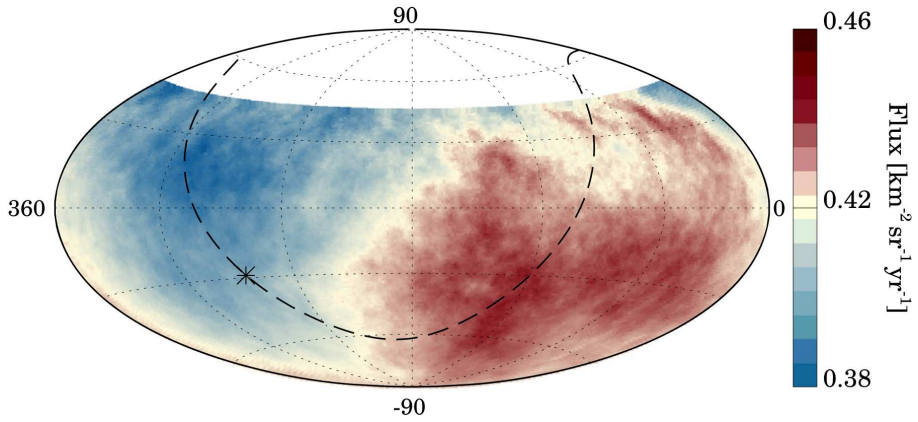

Fig. 2.3 Map in equatorial coordinates of the cosmic-ray flux for $E \geq 8\,\mathrm{EeV}$, smoothed in windows of 45°. The galactic plane is represented with a dashed line, while the galactic center is indicated by an asterisk [67].

energy range, it could encompass factors like the intrinsic motion of galactic arms, while, as a broader illustration, even the Earth's orbital motion around the sun can introduce subtle yet discernible contributions to a dipolar anisotropy.

At energies in excess of $\sim 0.5\,Z$ EeV, the Larmor radius of individual particles becomes larger than the galactic disk, and hence galactic sources, if present at this energy, should contribute to significant anisotropies in the direction of the galactic center and the galactic plane. Since no such anisotropies were observed, the galactic contribution is strongly constrained in the $1-5\,\mathrm{EeV}$ range [66]. Instead, results from the Pierre Auger Observatory indicate a significant dipole amplitude emerging for energies $> 4\,\mathrm{EeV}$, pointing $\sim 125^\circ$ away from the galactic-center direction, indicating an extragalactic origin of the cosmic-ray flux above the ankle [67, 68]. Fig. 2.3 shows a smoothed map of the flux for $E > 8\,\mathrm{EeV}$ in equatorial coordinates, in which the dipole pattern is apparent.

The substantial anisotropy observed cannot be solely attributed to the Compton-Getting effect [69]. Instead, a more plausible explanation lies in the dipolar modulation stemming from the diffuse propagation of cosmic rays within the turbulent extragalactic magnetic fields. Generally, lower densities are expected to yield greater anisotropy, as the overall flux becomes distributed across fewer discrete sources in such scenarios. Consequently, the dipole's orientation would predominantly align with the direction of the nearest or most luminous sources. As the energy threshold increases, the maximum redshift from which ultra-high-energy cosmic rays can reach Earth diminishes due to interactions with cosmic background fields, constraining their origins to nearby structures or objects. Given the inhomogeneous distribution of galaxies in our local neighborhood, if cosmic ray sources mirror the matter distribution in the Universe, this alone would result in a substantial dipolar anisotropy at sufficiently high energies, further distorted by local magnetic fields. Reasonable estimates naturally place the predicted dipole amplitude at $5-20\%$ at 10 EeV, aligning with the observed levels [70]

At smaller angular scales, there is a keen interest in studying anisotropies to establish a

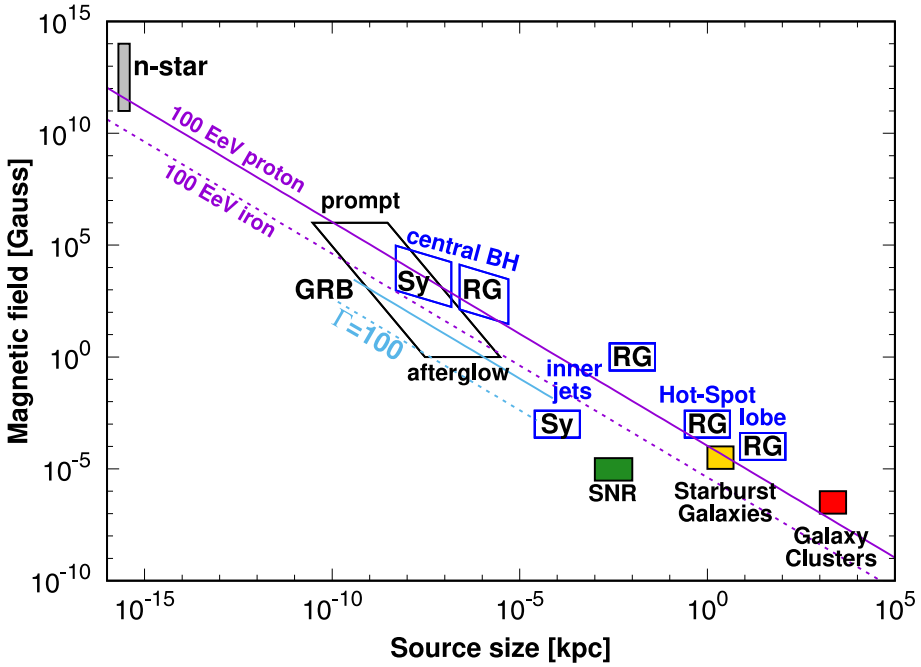

Fig. 2.4 Hillas plot illustrating the positions of potential sources of ultra-high-energy cosmic rays (UHECR) on a graph that compares their characteristic magnetic field against their size [72]. The sources include neutron stars (n-stars), gamma-ray bursts (GRB), central black holes (BH) within Seyfert galaxies (Sy) and radio galaxies (RG), accompanied by their respective jets, lobes, and hot-spots, as well as supernova remnants (SNR), starburst galaxies, and galaxy clusters. The diagonal lines on the plot represent the energy threshold above which a source is deemed capable of accelerating particles (here protons and iron nuclei) to energies exceeding 10^{20} eV.

direct connection between the observed cosmic-ray flux and individual sources. Extensive searches have been undertaken to detect intrinsic anisotropies and to explore correlations with the positions of known astrophysical objects in the sky (see e.g. [71]), but as of yet, no significant results have emerged in either of these domains. This absence of small-scale anisotropies, even at the highest energies, suggests that galactic and/or extragalactic magnetic fields likely exert a substantial influence on the trajectories of UHECRs, effectively smoothing out small-scale anisotropies. This outcome aligns with expectations in scenarios featuring a mixed composition, where the cosmic-ray composition becomes progressively heavier as energy increases. Such a scenario finds support in the composition measurements conducted by the Pierre Auger Collaboration [55, 56].

Regarding possible source objects respectively source regions for UHECR, there is a simple geometrical criterion which can be employed to obtain a rough classification of the suitability of an astrophysical object for acceleration by two parameters. Using the elongation L of the object and its average magnetic field B, if the Lamour radius of a particle is required to be contained inside the object volume during the acceleration process, the corresponding maximum attainable energy $E_{\max}$ can be expressed by the formula

$$E_{\max} \simeq \Gamma e Z B L, \tag{2.4}$$

where eZ is the charge of the nucleus. In addition, the relation allows for a Lorentz boost factor Γ in the case of acceleration in a relativistic jet. A graphical depiction of this Hillas criterion for various astrophysical candidates for the acceleration of UHECR is shown in Fig. 2.4. It is named after Antony Michael Hillas, who in 1984 was the first to compile such a classification [73]. The diagonal lines on the plot depict the exclusion thresholds below which potential source candidates would not be anticipated to achieve the acceleration of protons (or alternatively, iron nuclei) to energy levels surpassing 10^{20} eV.

Active galactic nuclei (AGN), gamma-ray bursts (GRB), quasars, or supernova remnants (SNR) are generally thought to be excellent source candidates for high-energy cosmic rays, but direct evidence is still lacking.

3. Extensive Air Showers

In the ultra-high-energy regime, the flux of cosmic rays is so small and the interaction probability with air molecules so high that the direct detection method ceases to be viable. For this reason, most of the detection of high-energy cosmic rays relies on ground-based indirect detection of air-shower events in the atmosphere and the secondary particles produced therein. The air-shower process represents one of the most complex phenomena observed in ultra-high-energy physics. Detailed numerical simulations are a challenge due to the immense number of particles involved.

When particles are energetic enough to surpass the geomagnetic field and penetrate into the Earth's atmosphere, they may interact with its constituent air molecules primarily composed of nitrogen (N_2), oxygen (O_2), or argon (Ar). This interaction initiates a chain reaction, leading to the proliferation of a particle cascade, analogous to the shower development in a calorimeter of variable density with a vertical thickness of more than 11 interaction lengths and 26 radiation lengths [23]. Consequentially, the Earth's atmosphere is practically opaque for ultra-high-energy radiation. While, strictly speaking, *particle cascade* refers to the physical phenomenon of exponential particle multiplication due to incident high-energy radiation in a dense material, and *air shower* refers to its practical realization in the context of cosmic ray research, the terminology will be used interchangeably, henceforth, to describe the latter.

Due to the highly variable air density, it is impractical to describe the evolution of such cascades solely in terms of their altitude above the Earth's surface. In line with man-made calorimeters, one usually relies on the concept of *(vertical) atmospheric depth* as a measure of the material traversed and, consequently, as a proxy for the longitudinal axis of the shower. When the density ρ of air as a function of altitude h is known, the atmospheric depth X can be computed using the following integral:

$$X(h) = \int_h^\infty \rho(h')\,\mathrm{d}h' \tag{3.1}$$

Integrated from outer space to sea level, the atmosphere has a total column density of about $1030\,\mathrm{g\,cm^{-2}}$ [74]. However, it's essential to consider that cosmic rays arrive from

various directions, exhibiting nearly isotropic characteristics. Thus, it cannot be assumed that these showers strike the atmosphere perpendicular to its surface. Neglecting the curvature of Earth's atmosphere as a preliminary approximation (*flat Earth approximation*), the actual *slant depth* X_{inc} for an inclined shower with a zenith angle θ is derived from straightforward geometric considerations:

$$X_{\mathrm{inc}}(h, \theta) \simeq \frac{X(h)}{\cos(\theta)} \tag{3.2}$$

This section seeks to establish a theoretical framework for comprehending the development and behavior of extensive air showers within the Earth's atmosphere. The ultimate goal is to gain insight into how the observed properties of such showers can be leveraged to deduce the nature of its primary particle. First, the de facto standard case of the nucleon-initiated air shower will be discussed here, before the peculiarities of other primary particles will be considered.

3.1. Nature of Particle Cascades

Since the Earth's atmosphere does not exhibit a clear upper bound, but rather shows an exponential decrease of air pressure and consequentially density as described by the barometric formula, the height of the first interaction of the air shower can vary a lot and is best described by the slant depth X. The depth of the first interaction X_1 is distributed as $\exp(-X_1/\lambda_{\mathrm{int}})$, where λ_{int} denotes the hadronic interaction length, and is one of the main sources for shower fluctuations [74]. Typical altitudes for the first interaction for the case of air showers initiated by hadrons are between 15 km and 20 km [75].

The energetic secondary particles resulting from both initial and subsequent collisions behave alike as they propagate deeper and deeper into the atmosphere, giving rise to particle number multiplication in an exponential fashion. The collision processes are typically hadron dominated, contributing to the formation of a hadron cascade that progresses longitudinally along the direction of incidence. Within a particle multiplication process (hadronic scattering, pair creation, etc.), the energy of the initial particle will be distributed among the secondary products. Simultaneously with the multiplication of the quantity of particles, the mean kinetic energy per particle involved decreases. Furthermore, additional energy loss and absorption processes also arise in the entire duration of the shower. If the individual particle energy falls below a certain critical threshold, the creation of new particles slows down significantly and as a result of decay processes the total number drops again. The longitudinal profile of the particle number $N(t)$ in the air shower can be parameterized in a rough simplification by

$$N(t) \sim t^{\alpha} e^{-\beta t}, \tag{3.3}$$

where $t = x/X_0$ is the shower depth in units of the radiation length and α and β are free fit parameters [54]. This typical development of an air shower is also shown schematically in Figure 3.1. A realistic depiction of the extent of particle trajectories inside an air shower was created with computer simulations and is shown in Figure 3.2. Even here and in spite

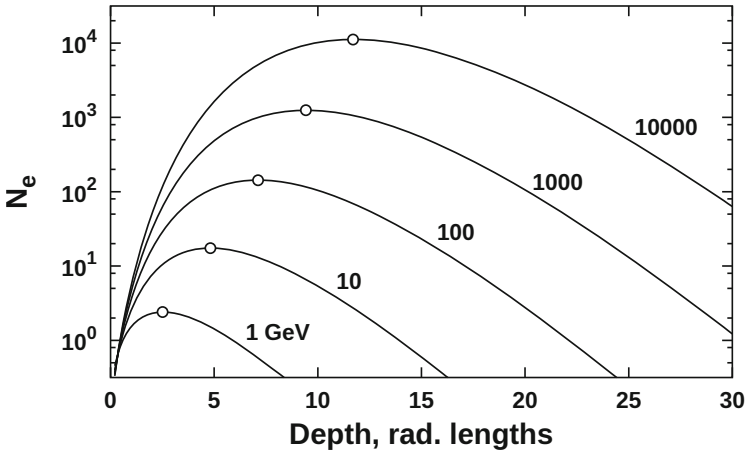

Fig. 3.1 Shower profiles for electrons in a purely electromagnetic air shower. The primary energy is noted next to the respective profiles, while the depth of the shower maximum is indicated with circles [74].

of the immense number of individual trajectories, the behaviors of particle multiplication, maximum development at a specific height and slow fade out are distinctly visible.

Thus, for each shower that completely passes through this evolution, a distinct point on the longitudinal axis can be identified at which the cumulative number of particles is at a maximum. Termed as $X_{\max}$ when measured in units of atmospheric depth, this parameter assumes critical importance in the systematic study of air showers. It is noteworthy that the depth of maximum development exhibits a strong dependency on energy (cf. Fig. 3.2), moderate sensitivity to the mass of the primary particle, and other factors, making it a central element in the identification of the primary particle. Section 3.6 details the possibilities for direct ground-based measurement of this parameter.

The total particle yield in an air shower is referred to as the *shower size*, denoted by N. Various methods have been developed to obtain this parameter through ground-based measurements as well. Typically, only charged particles are considered for summation, given that the neutral constituents, mainly gamma rays and neutrinos, tend to escape detection, either partially or entirely. The shower size is primarily a function of the energy of the primary particle E_0 (cf. Fig. 3.2), angle of incidence θ and the height of the first interaction, designated as either h_1 or X_1 [75]. However, the shower size exhibits a limited sensitivity to the mass of the primary particle A. As a result, it proves to be a viable tool for estimating the primary particle's energy.

For the purposes of geometrical reconstruction, the *shower axis* plays an important role. It is defined as the linear continuation of the momentum vector of the primary particle, and it provides the longitudinal axis of the shower. Intersected in the lateral plane, it is usually forming the center of the dispersion and the point of highest particle density in the lateral direction as the *shower core*.

It is imperative to emphasize that, at any specific moment during the progression of the air shower's development, the spatial distribution of the particles involved is essentially confined to a nearly flat, disk-like configuration oriented perpendicular to the shower axis. This *shower front* has a thickness of only a few meters and travels earthward at a velocity

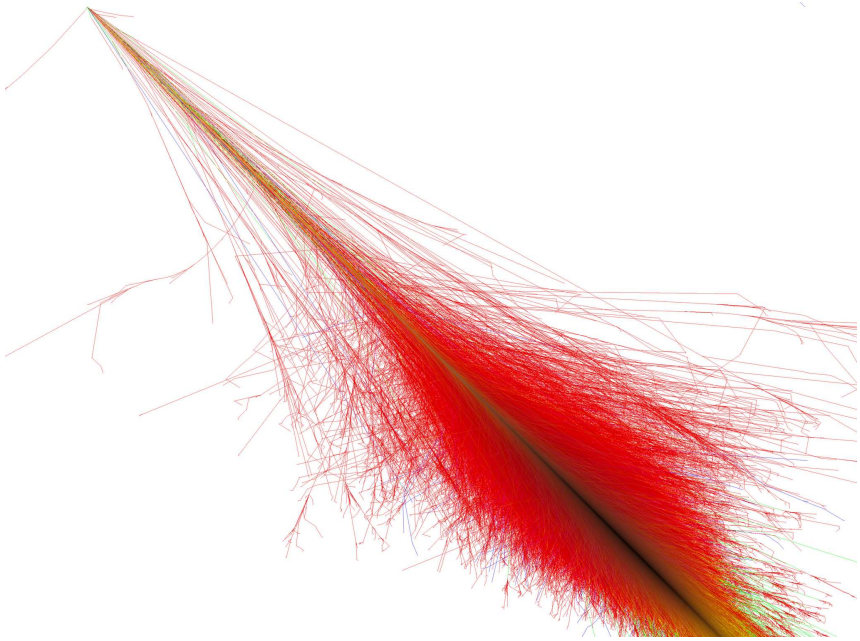

Fig. 3.2 Visualization depicting the individual particle trajectories within a simulated air shower triggered by a proton with an energy of 10^{15} eV at an angle of inclination of 45°. The image encompasses spatial dimensions of about 70 km in both horizontal and vertical directions. It represents a time-integrated trajectory of the event rather than a snapshot of a particular moment in time [76].

close to the speed of light. Therefore, the temporal resolution capabilities of a potential detector system play a crucial role in the accuracy of the reconstruction of the geometry of such a shower.

3.2. Components of an Air Shower

An extensive air shower has an electromagnetic, a muonic, a hadronic and a neutrino component. Here, the ordinary case of a nucleus-induced shower shall be discussed first and all distinctions that arise for air showers of photons will be addressed in the next part.

Fig. 3.3 shows the different components of an ordinary, i.e. nucleus-induced, air shower. The hadronic cascade is the backbone of the shower, providing a steady supply of high-energy pions, which in the case of neutral pions initiate independent electromagnetic cascades, or in the case of charged pions form a kind of parallel hadronic cascade themselves. Finally, muons and neutrinos are not driving elements of the air shower, but rather final states that can be used as markers for the evolution of the shower.

The relative composition of particle types in an air shower depends on a number of factors, above all on the stage of shower development, often referred to as *shower age*, denoted by s. Based on computer simulations, it can be roughly estimated that an average (hadronic-

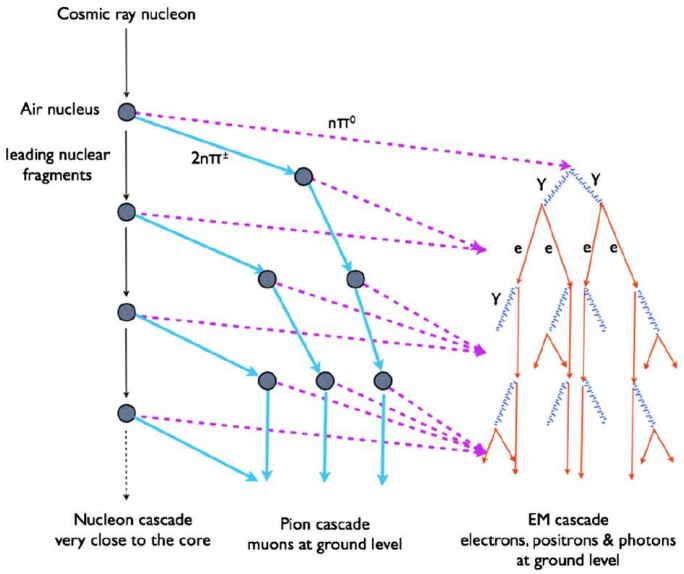

Fig. 3.3 Schematic depiction of the evolution of an air shower and its components. At each step roughly 1/3 of the energy is transferred from the hadronic cascade to the electromagnetic one. [23]

induced) air shower at sea level consists of around 80% photons, 18% electrons/positrons, 1.7% muons, and approximately 0.3% hadrons [75] (cf. Fig. 3.4). Neutrinos are even more abundant, but usually escape detection. However, the actual measured fractions can differ greatly from this due to the individual efficiencies of the various detection systems.

Aspects of the individual components are briefly described in the following subsections. Since the electromagnetic component is the cleanest and self-contained, while it shapes the overall shower evolution due to its dominant nature, it will be discussed first, after which the basic concept will be abstracted to hadronic showers.

3.2.1. Electromagnetic Component

Electromagnetic cascades possess a simpler straightforward configuration owing to the well understood interactions and the limited number of different elementary particles. They can, therefore, serve as a template for all particle cascades. The Heitler model [24] provides a simplified explanation of their development and chief characteristics. A simpler variation of this model is initially presented [74] that incorporates cascades with a single type of equivalent particles and only one consistently behaving interaction. Furthermore, the evolution of the cascade is one-dimensional. Every time a particle passes the interaction length λ, it splits into two new particles, each carrying half the original energy. The cross-section of the interaction, and therefore λ, remains energy independent and constant. Thus, beginning with a single primary particle of energy E_0, there is a doubling of the number of particles and a corresponding halving of the individual particle energy after each stage of the shower evolution. At a certain atmospheric depth $X = n\lambda$, the cascade consists of $N = 2n$ particles, each carrying energy $E = E_0/2^n$. The process of particle multiplication

is repeated until the individual particle energies fall below a critical value E_c, under which the theorized interaction promptly ceases to produce secondary particles, and the maximum attainable shower size is reached. The applicability of the Heitler model ends here and any intermediate energy loss or degeneration processes that would eventually lead to a fade-out of the cascade after this maximum are beyond its scope.

Despite the model's crude assumptions and severe simplifications, it makes two predictions that are qualitatively in line with findings from controlled experiments and Monte Carlo (MC) simulations: The shower size, i.e. the number of particles in the cascade, is proportional to the primary energy, while the depth of the maximum $X_{\max}$ is proportional to the logarithm of the primary energy.

The more extensive Heitler model describes realistic electromagnetic showers, that consist of self-perpetuating circuits of multiple interactions from QED, which are formed by photons γ and electrons/positrons $e^{\pm}$. In the atmosphere, high-energy photons from the cascade, in combination with a nucleus of the surrounding air, will produce electron-positron pairs:

$$\gamma + N_{\text{air}} \rightarrow e^{-} + e^{+} + N_{\text{air}} \tag{3.4}$$

In turn, electrons/positrons will emit photons in bremsstrahlung processes in the Coulomb field of the nuclei in air:

$$e^{\pm} + N_{\text{air}} \rightarrow e^{\pm} + \gamma + N_{\text{air}} \tag{3.5}$$

On first account, the two processes listed above each lead to a doubling of the shower particles, with the original energy of the shower particle being shared among them. The nucleus of an air molecule is required for its Coulomb field, but usually does not gain a considerable share of the primary energy in the scattering and can therefore be neglected in the shower development. Now, if the assumptions are made that the two interactions are the only ones that the particles undergo and that their mean free path lengths are identical and energy independent, then a cascade can be constructed according to the above model and exactly the same results regarding the shower size and depth of maximum development can be obtained for the case of the electromagnetic cascade.

Electromagnetic showers in the atmosphere can be induced by any high-energy incident particle that interacts electromagnetically, e.g. ultra-high-energy photons initiating pair-production processes. In the context of hadronic showers, they are usually started by the decay of energetic π^0 produced as a by-product of the hadronic cascade.

Due to their relatively short interaction length – for electrons the characteristic radiation length is $X_{\text{rad}} = 37\,\text{g}\,\text{cm}^{-2}$ – an electromagnetic cascade is rapidly absorbed in the atmosphere. However, it should be noted that in the case of a hadronic shower the high-energy neutral pions that are produced as a by-product constantly spawn new electromagnetic sub-cascades, which leads to the aforementioned dominance of this component in a nucleus-induced air shower.

3.2.2. Hadrons

The basic principles of an electromagnetic cascade can be extended to the case of the hadronic component of an air shower [77], which can be considered the back-bone of the whole shower evolution.

Before proceeding any further, it is essential to emphasize that hadronic interactions in the energy domain present in extensive air showers are not well understood. This is due to the absence of controlled measurements within this range, which results in essential quantities such as cross-sections having to be derived from extrapolation over several orders of magnitude from the results of collider experiments. Multiple interaction models are available for this purpose. Nonetheless, this introduces considerable systematic uncertainties into the theoretical modeling of air showers, particularly with regard to hadronic primaries.

An air shower and with it the hadronic cascade is initiated by a hadronic interaction of a nucleon or heavier nucleus. A considerable fraction of the initial nucleon's energy will then be transferred to secondary mesons. Apart from second-order collective effects the same happens in nucleus-nucleus interactions, where more than one nucleon of the primary can participate in the interaction. Further information about the theoretical approach with nucleus primaries is given in Section 3.3, while initially the case of plain hadronic shower will be discussed here first.

The simplified model in [77] assumes that about $1/2$ of the primary energy is transferred to the secondaries, which are compound of both charged and neutral pions. The rest of the energy is retained by the secondary nucleon, which traverses further until it interacts again, in an analogous manner. Analog to the electromagnetic cascade, the interaction length λ_{int} is assumed to be energy-independent and fixed. The process continues until dense target material, e.g. the ground, is reached or the energy of the nucleon falls below the interaction threshold.

Each hadronic interaction produces a multitude of high-energy secondary pions, which are on average subdivided into $1/3$ π^0 and $2/3$ $\pi^{\pm}$. Due to their very short mean lifetime ($(8.4 \pm 0.6) \times 10^{-17}$ s [75]) the neutral pions are assumed to decay practically instantaneous at the interaction point via

$$\pi^0 \rightarrow \gamma + \gamma, \tag{3.6}$$

and produce high-energy photons that initiate subsequent electromagnetic cascades. These overlapping sub-showers provide the main observable features of the air shower, due to their numerical dominance.

The charged pions possess a much longer lifetime and are therefore able to traverse considerable distances in the direction of incidence before decaying. Thus, there exists the possibility for further interactions with nuclei of the air. Interacting charged pions produce new generations of $1/3$ π^0 and $2/3$ $\pi^{\pm}$, which contribute to the second particle generation, until they finally decay. The competition between the decay and the re-interaction channel depends on the variable density of the surrounding material, i.e. the constantly changing atmospheric conditions and primarily the altitude. While at high altitudes the decay with-

out intermediate interactions is more probable, at lower altitudes, towards the ground, the likelihood of numerous interactions is becoming higher [74].

Hadronic showers behave much in the same way as electromagnetic ones. The cascade consists of two interrelated process each one not dissimilar to the cascade from the Heitler model, while producing overlapping electromagnetic cascades. Below $10^{14}\,\mathrm{eV}$, the interaction length of the participating hadrons, albeit energy-dependent, can be approximated as $\sim 100\,\mathrm{g\,cm^{-2}}$ among its species, which is longer than the typical radiation length in air [74]. This leads to a larger fraction of the primary particle energy to be carried further along the shower axis than would be the case for a pure electromagnetic cascade. However, this is somewhat counteracted by the higher multiplicity of secondaries and the resulting faster energy dissipation. The determination of the depth of the shower maximum is far more complex in the case of a hadronic shower, due to the superposition of the electromagnetic sub-showers and the effective energy transfer between the components, which depends on numerous dynamical quantities including cross-section, inelasticity of the interaction, secondary particle multiplicity, and re-interaction rate for charged pions.

3.2.3. Muons and Neutrinos

The development of the hadronic component can be observed by its muonic output. Muons are produced in copious quantities as decay products of unstable charged secondaries, mainly pions and kaons, during the evolution of the hadronic cascade, with the following decay processes being the most relevant contributors [75]:

$$\begin{aligned} \pi^{\pm} &\rightarrow \mu^{\pm} + \overset{(-)}{\nu}_{\mu}, \\ K^{\pm} &\rightarrow \mu^{\pm} + \overset{(-)}{\nu}_{\mu}, \\ K^{\pm} &\rightarrow \pi^{0} + \mu^{\pm} + \overset{(-)}{\nu}_{\mu}, \\ \tau^{\pm} &\rightarrow \mu^{\pm} + \overset{(-)}{\nu}_{\mu} + \overset{(-)}{\nu}_{\tau}. \end{aligned} \tag{3.7}$$

Overall, the number of muons N_{μ} in an air shower scales almost linearly with its energy [77].

All longer lived unstable mesons that contribute with their decays to the overall muon count, e.g. charged pions, are suspect to a competition between decay and re-interaction, that depends primarily on their current altitude. Very inclined showers spend a considerable time in the higher regions of the atmosphere, where the decay channel for energetic charged pions is more prominent. This shifts the atmospheric depth at which the bulk of the muonic component is produced to lower values compared to vertical incident showers and leads to an overall zenith angle dependence of the muon number in an air shower.

An additional source for muons in an air shower is the photoproduction process. Instead of the more common electron-positron pair, a high-energy photon will sometimes create a doublet of the second lepton family. This allows pure electromagnetic showers to develop a muon component as well. Even though this channel is suppressed for kinematic reasons, the

sheer number of photons in an extensive air shower will lead to a significant contribution to the total muon sum.

After passing the apex the number of muons is hardly reduced toward sea level, since the probability for energy-loss processes, like bremsstrahlung and ionization, is relatively low compared to electrons [54]. Even though muons themselves are unstable particles and will decay in nearly all cases via

$$\mu^- \to e^- + \nu_\mu + \bar{\nu}_e \tag{3.8}$$

and its charge-conjugate channel, due to relativistic time dilation occurring for the energetic particles, the decay length for a 1 GeV muon is about 6 km (cf. Sec. C.1) and in most cases sufficient for the unstable particles to be directly detected at sea level.

Neutrino production in an air shower is closely related to the muonic one. The bulk of decay processes that yield neutrinos are two-body decays with an associated muon-neutrino ν_μ (or muon-antineutrino $\bar{\nu}_\mu$). Consequently there exist a parallel neutrino component that evolves as the shower propagates in the atmosphere and develops in age. However, *atmospheric neutrinos* are often disregarded in the context of air showers for their low interaction probability, since they do not advance the evolution of the shower and are very difficult to detect directly. Whereas the steady production of neutrinos in the atmosphere associated with air showers can certainly be detected in dedicated neutrino detectors and forms an often times undesired low-energy background for the analysis of cosmic neutrino radiation.

3.3. Superposition Model for Nuclear Primaries

An air shower that was initiated by a nucleus ${}^A_Z X$ with primary energy E_0 can be assumed to develop like a superposition of A independent plain nucleon-induced showers, each with an energy E_0/A. This approach is commonly known as the *superposition approximation* for nucleus primaries. If the primary energy is scaled by $1/A$ in the extended shower model by Matthews [77], the depth of the shower maximum $X_{\max}$ can be obtained in relation to a plain proton-induced shower of the same primary energy:

$$X_{\max} = X^p_{\max} - X_0 \ln A, \tag{3.9}$$

where X_0 is the radiation length relevant for the contained electromagnetic sub-showers. This implies that air showers with the same primary energy develop their maximum earlier in the atmosphere, when the mass of the primary nucleus is greater. This shift is overall somewhat exaggerated by the numerous simplifications of the model, but nonetheless affirms the usability of the observable $X_{\max}$ for composition studies. The shower size, however, is not changed by the transformation, but the earlier development may lead to showers induced by heavier nuclei getting strongly absorbed before reaching the observation level.

If the same procedure is applied to the total number of muons N_μ in an air shower, the result

$$N^A_\mu = A^{1-\beta} N^p_\mu \tag{3.10}$$

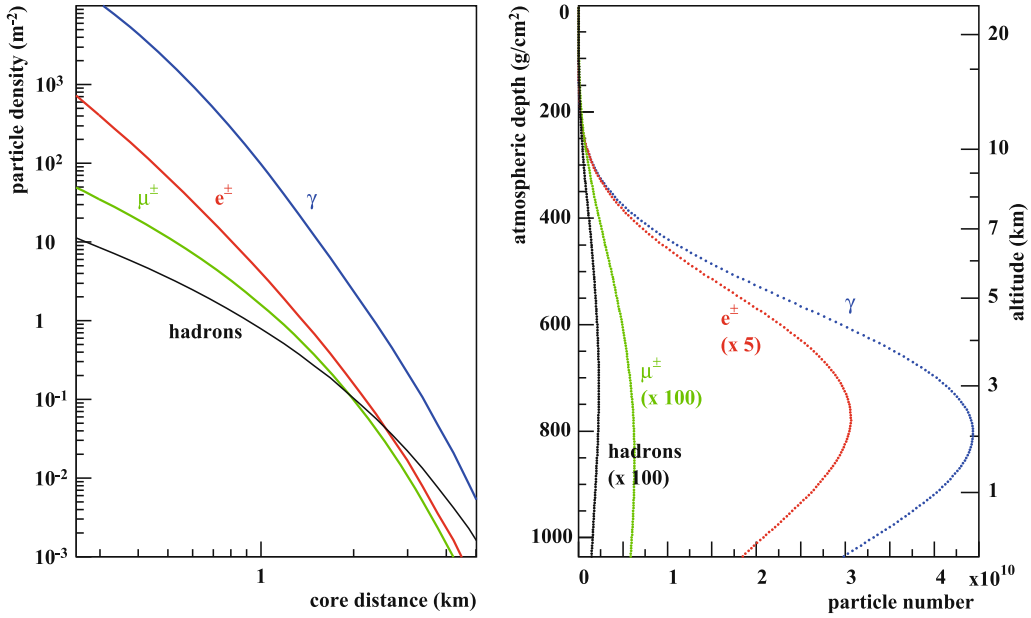

Fig. 3.4 Simulated lateral (left) and longitudinal (right) shower profiles for vertical proton-induced air showers of 10^{19} eV. The lateral distribution has been simulated for an atmospheric depth of 870 g cm^{-2}, which corresponds to the altitude of the detector array of the Pierre Auger Observatory for vertical incidence [78].

suggests that for constant primary energy the number of muons scales with primary mass A, with $\beta = 0.85$ being a fundamental parameter of muon multiplication in Matthew's model. Hence, the determination of the muon content in an extensive air shower, may give additional valuable insight for distinguishing between showers induced by different primary particles.

3.4. Longitudinal Development of an Air Shower

The concept of *air-shower universality* states that the longitudinal development of the electromagnetic component of nucleus-induced showers can be completely described in terms of two key parameters, the primary energy E_0 and the shower age s. This implies that all information about the primary particle can in principle be deduced from the measurement of three distinct parameters, like $X_{\max}$, E_0 and the number of muons in the shower N_μ. In addition, information about the first-interaction cross-section can be retrieved by fitting the $X_{\max}$ distribution in fixed energy bins [23].

Fig. 3.4 shows the longitudinal development of a simulated air shower. The overall shape of the shower profile is dominated by the electromagnetic component. The parameterization according to Thomas Gaisser and Anthony Hillas can describe the longitudinal development of a single electromagnetic cascade as well as the average development of an air shower [79]. It constitutes the extension to Eq. (3.3). The number of particles $N(X)$ as a function of

the atmospheric depth is expressed as

$$N_e(X) = N_{\max} \left(\frac{X - X_1}{X_{\max} - X_1} \right)^{\frac{X_{\max} - X_1}{\lambda}} \exp\left(\frac{X_{\max} - X}{\lambda} \right), \tag{3.11}$$

where $N_{\max}$ is the maximal number of particles, attained at atmospheric depth $X_{\max}$. X_1 is the depth of the first interaction and λ the mean free path length. This form of the longitudinal profile is commonly referred to as the *Gaisser-Hillas function* and is widely employed for the reconstruction of air shower profiles from longitudinal particle number measurements. There exist also other functions used for this purpose, among them a modified Gaussian [75]. Even if parts of the profile are missing, a fit can reveal the location of the maximum and the total shower size.

3.5. Lateral Distribution of Particles

Due to transverse momentum of the secondaries emerging from the collisions in the air shower and as a consequence of scattering processes, the cascade spreads out laterally as well. The lateral distribution of different particle species in an air shower is depicted in Fig. 3.4. On the basis of the shower axis, the widening occurs uniformly in all directions (except for strongly inclined showers). The figure clearly shows that in particular the hadronic component is more widely spread, while nearer to the shower core the electromagnetic particles are dominant. At the highest energies (10^{19} eV), the measurable spread can be on the order of several km.

For the case of electromagnetic cascades Koichi Kamata, Jun Nishimura [80] and independently Kenneth Greisen [81] provided an analytic approximate solution to the transport equations in the late 1950s and proposed a lateral distribution function (LDF) for particles contained in the shower:

$$\frac{\mathrm{d}N_e}{\mathrm{d}r^2} \propto r_M^{-2} \left(\frac{r}{r_M} \right)^{s-2} \left(1 + \frac{r}{r_M} \right)^{s-4.5} \tag{3.12}$$

Here s is the shower age parameter

$$s = \frac{3X}{X + 2X_{\max}}, \tag{3.13}$$

that generalizes the stage of shower development, with $s = 1$ at shower maximum, and $r_{\mathrm{M}} \simeq 90\,\mathrm{m} \times (10^{-3}\,\mathrm{g\,cm^{-3}}/\rho_{\mathrm{air}})$ is the Molière radius, which depends on the density of air ρ_{air}, and accounts for the typical spread of electrons in a material of specific density [72]. This form of LDF is sometimes referred to by the name of *NKG function*.

However, the actual electromagnetic component of a nucleus-induced shower behaves slightly differently, due to the constant re-spawning of sub-showers, and is more spread. In general, shower components, like the muonic, that develop higher in the atmosphere exhibit a flatter lateral profile [54].

The area density of particles is relevant for modeling the response of large spread ground-based detector systems, since it can be used to deduce several air shower quantities, among

them the primary energy. Various parametrizations of the LDF have been developed for the different experimental setups, that directly include the expected signal strengths at varying distances from the shower core [23].

3.6. Principles of Observation

The minimum energy for a primary particle to be reasonably well measured at sea level via the particles produced in the air shower is about $10^{14}\,\mathrm{eV}$ [54]. This value thus marks the lower limit above which the systematic study of extended air showers can be usefully employed for the survey of cosmic rays. Further limits are only set by the design of the detector systems that are used, and here too a variety of solutions have been developed for the different energy ranges.

Air shower detectors, typically situated at ground level, may benefit from various detection altitude ranges due to the scaling of the shower depth with primary energy, since fluctuations during the development of the shower are smallest near the particle maximum, X_{max}. That said, it is important to ensure that the premature absorption of the shower by the ground does not cause a loss of critical information.

The classical technique for measuring extensive air showers is the sampling of the shower particles with scintillators or water-Cherenkov counters that are installed on ground level. Energetic particles passing through the detection volume induce photon emission by fluorescence or the Cherenkov effect, which is converted into digital pulses with precise temporal resolution. Typically, each of these detectors covers an area on the order of $1\,\mathrm{m}^2$. For the observation of extended air showers individual detector stations are networked in regular arrays, that cover surface areas in direct proportion to the cosmic ray flux in the energy range of interest. For the studies on the high end of the energy spectrum ($E > 10^{19}\,\mathrm{eV}$) surface areas of a few thousand km^2 are required to collect sufficient statistics in reasonable time frames [23]. The spacing of air shower arrays is also related to the primary energy and the magnitude of the subsequent lateral shower spread expected. When a cosmic ray event falls within the boundaries of the array, the subsample of detectors close enough to the shower core will observe the radiation reaching the ground. As long as three individual stations record the event in coincidence, the shower properties can be reconstructed, with increasing precision for growing station number, based on the temporal onset of the shower front and the spatial distribution of signal densities. Such detector arrays can operate day and night and provide inherent redundancy. However, their main drawback is that they sample particle densities at a fixed atmospheric depth, and thus only provide a single snapshot of shower evolution at development stages that vary from shower to shower, because of the underlying depth fluctuations.

Another popular approach, that somewhat circumvents the above issue, utilizes the atmosphere as a calorimeter for the observation of air shower events: During the evolution of the shower, charged particles that propagate through the atmosphere will lose some of their energy by excitation of nitrogen atoms which in turn emit fluorescence light isotropically in the ultraviolet (UV) range (see e.g. [82]). If the atmospheric conditions allow it, this faint

glow along the shower axis can be observed by UV telescopes and used to reconstruct the longitudinal development of the particle number, which together with geometrical information provides direct access to the value of the key observable $X_{\max}$. For primary particles with energies exceeding 10^{17} eV, the fluorescence light is sufficiently intense to be recorded by telescopes at sea level in the presence of the diffuse background of starlight [54]. The number of emitted fluorescence photons is proportional to the energy deposited in the atmosphere, due to electromagnetic energy losses by charged particles. After accounting for attenuation effects and correcting for missing energy carried away by neutrinos and muons [83], the emission can even be used to estimate the primary energy, when integrated along the shower axis.

Similarly to the measurement of fluorescence light in the atmosphere, efforts have been made to exploit the emission of Cherenkov light that is released as a consequence of the passage of energetic charged particles in air. However, due to its highly directional nature, detector arrays have to be spread over large surface areas on the ground to gather sufficient statistics. Today, ground-based Cherenkov arrays and sophisticated imaging systems are the backbone of dedicated high-energy gamma-ray astronomy [84].

The muon component of an air shower can be explored by employing underground scintillation counters that use the soil above them as shielding from conventional electromagnetic radiation. For the detection of GeV muons, the shielding should be of the order of $500\,\mathrm{g\,cm^{-2}}$, which amounts to about 2 m of rock [74, 85].

In addition, equipment and procedures have been developed to measure the radio emission from extensive air showers (see e.g. [86–88]). The charged particles in the shower interact with the geomagnetic field, which causes electromagnetic emission in the MHz range. Current efforts still focus on engineering work, i.e. to demonstrate the capabilities to measure air shower properties like arrival direction, energy, primary mass and prove the applicability of the radio technique to large scale observatories, but show promising results for further prevalence [89].

4. Pierre Auger Observatory

The Pierre Auger Observatory is presently the most extensive and capable facility for the detection and measurement of extensive air showers induced by ultrahigh-energy cosmic rays. It is located in the Argentinian province of Mendoza, northeast of the small town of Malargüe, from where it is monitored and technically supported by local staff day and night. The recorded data of the Observatory is preprocessed and distributed at regular intervals to the collaborating institutes around the globe. Over 400 scientists from 18 different countries are part of the Pierre Auger Collaboration, that manages the Observatory and analyses the data streaming from it.

The Observatory has been under construction until 2008, but due to its modular structure, it had been used to detect cosmic rays since 2004, practically without interruption. Much of the recent knowledge about the nature of high-energy cosmic rays is based on data

measured at the Pierre Auger Observatory. Its main scientific objectives include the study of the flux suppression at the highest energies and its connection to the theorized GZK effect, composition and anisotropy measurements for extragalactic cosmic rays and the search for UHE neutral particles.

The detection is performed in essence by two independent detector systems, the Surface Detector (SD) and Fluorescence Detector (FD), covering an area on the ground of $3000\,\mathrm{km}^2$. Their layout is depicted on the map in Fig. 4.1; Fig. 4.2 shows the detector components deployed in the field. The combination of the two well-established systems into a hybrid detector is a key feature of the Observatory and provides data of unprecedented precision for air-shower event reconstruction. The two techniques present complementary ways of observing air showers, which provides important cross-checks and measurement redundancy. Since the FD can only operate in suitable viewing conditions, amounting to a total uptime of $\sim 15\%$ [85], its data can be used to improve the reconstruction accuracy of sole SD measurements. With its uptime of $\sim 100\%$, the SD contributes to a large extent to the annual accumulation of about $6 \times 10^3\,\mathrm{km}^2\,\mathrm{sr\,yr}$ of exposure at $3 \times 10^{18}\,\mathrm{eV}$, above which the full detector is expected to obtain full efficiency for extensive air showers [92]. In addition to these two main detector systems, extensions have been available for several years to augment the detection and interpretation of air-shower events.

The detection area is located on the high plain of the Pampa Amarilla (engl.: *yellow prairie*), with a mean altitude of $\sim 1400\,\mathrm{m}$ a.s.l., corresponding to a vertical air column of $\sim 870\,\mathrm{g\,cm}^{-2}$ [85]. The atmospheric conditions above the Observatory are of particular interest for the shower reconstruction and are thus continuously monitored by dedicated facilities in addition to installations at the four FD sites.

In the following, some necessary details about the various relevant detector systems will be given.

4.1. Surface Detector

The surface detector (SD) of the Pierre Auger Observatory is used to sample the lateral particle density of an extensive air shower at ground level. Its regular configuration comprises about 1600 cylindrical water-Cherenkov detectors (WCD), with a diameter of 3.6 m, which are set up in a triangular grid, each 1500 m apart on the ground area of the Observatory. Fig. 4.3a shows such a WCD unit, with its main components annotated. The detector stations operate completely autonomously and continuously, powered by individual solar modules on top of the tanks, with batteries for the nighttime. Communication with the central data acquisition system (CDAS) is managed with wireless technology and timing information is provided by a GPS antenna.

Each SD station is filled with $12\,\mathrm{m}^3$ ultra-pure water and the inner perimeter of the tank is lined with a diffuse reflective Tyvek foil. When relativistic charged particles associated with cosmic-ray induced air showers pass through the water, they may produce Cherenkov light, which is registered by three 230 mm diameter photomultiplier tubes (PMTs) installed

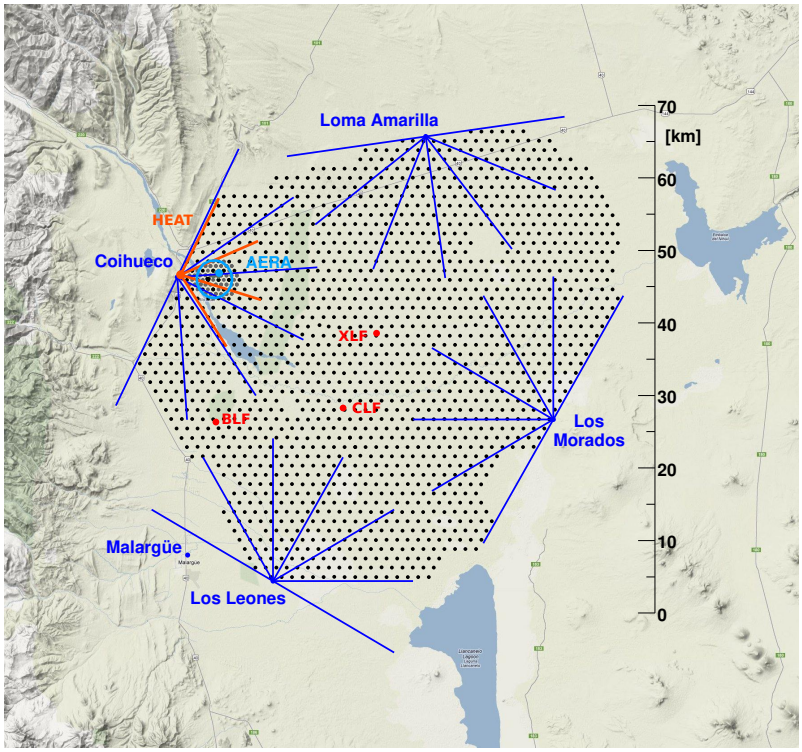

Fig. 4.1 Map of the different detector components of the Pierre Auger Observatory. Each black marker corresponds to an autonomous SD station. The four FD sites (Los Leones, Coihueco, Loma Amarilla and Los Morados) are shown, each with the 30° field-of-view of its six telescopes. In addition, the additional ground sites for atmospheric monitoring and calibration are also depicted (CLF, XLF, BLF). Located on the north-western border of the array are the low-energy extensions (see Sec. 4.3) of the Observatory, together with the prototype radio detection array AERA [90].

Fig. 4.2 A single SD station deployed in the field with the Los Leones FD site on the hill top in the background [91].

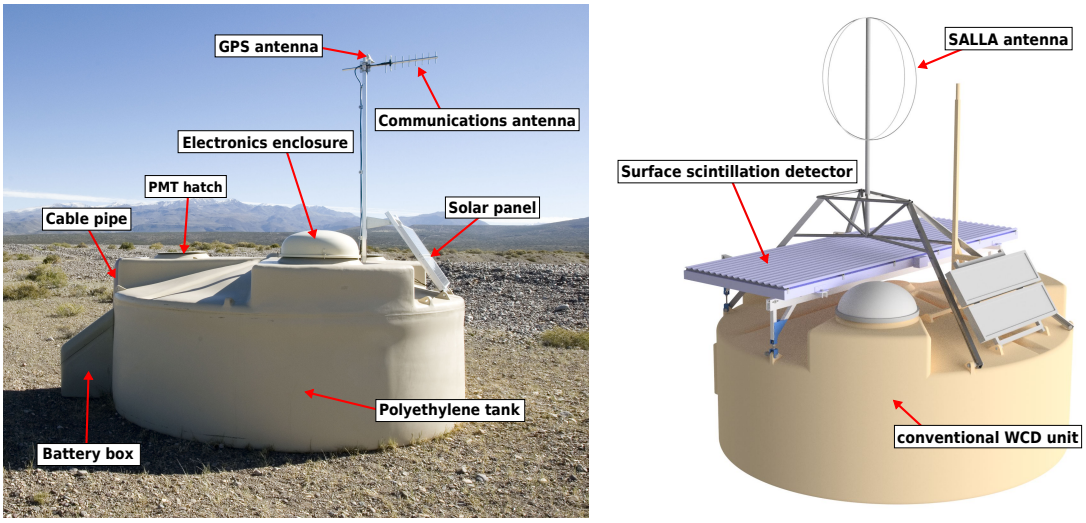

Fig. 4.3 SD station as deployed in the field, showing its main components. (a) Plain water-Cherenkov detector (WCD) as deployed during initial installation, with its main components annotated. (b) Components added during the Auger Prime Upgrade (see Sec. 4.1.1). Pictures are taken from [91] and modified.

symmetrically on the tank surface and looking downward through optical coupling into the water volume. In general, an SD station responds to both the electromagnetic and the muonic component of an air shower. The tank height of 1.2 m makes it also sensitive to high energy photons converting to electron-positron pairs within the water volume. The PMT signals are sampled by flash analog-to-digital converters (FADCs) with a time resolution of 25 ns, which allows for a good discrimination between electrons and muons entering the detection volume from the top. The recorded sequence of signals is referred to as the *signal trace*, with the intensity usually given in units of vertical equivalent muons (VEM) [85].

Initially two types of trigger conditions were imposed for an event to be registered: The local trigger (T2) operates on the level of an individual station. For the trigger condition to be satisfied either the signal must exceed a certain threshold in at least on time bin (threshold trigger), or the signal exceeds a much lower threshold in at least 13 bins within a 3 µs time window (Time-over-threshold (ToT) trigger). A superordinate level receives the local T2 trigger and builds a global, array-wide trigger (T3) to register events. It requires a relatively compact configuration of at least three local detector stations compatible in time, each satisfying the ToT trigger, or four triggered stations within slightly larger than light travel time with any type of T2. Additional trigger cases were added over the years. With the complete SD array, the global T3 trigger rate is about two events per minute, with one third of the showers having a primary energy above 3×10^{17} eV [85].

Individual SD stations are calibrated continuously using the passage of muons from low-energy showers. The calibration parameters are returned to the CDAS with each event and stored along with the event data. For details regarding this procedure, refer to [93].

4.1.1. AugerPrime Upgrade

With the original commissioning of the Pierre Auger Observatory, the suppression of the cosmic-ray flux toward the highest energies was clearly established. Its origin, however, is not yet understood. As a major step forward, an extension of the Observatory has been planned, which will facilitate addressing questions about the composition even with the low level of statistics present in the UHE regime.

The *AugerPrime* upgrade includes the installation of additional detector components across the array and a major overhaul of the SD station electronics. The primary goal is to extend the composition sensitivity of the Observatory up to the flux suppression region, which includes the newly introduced measurements of composition-specific observables on a shower-by-shower basis. The deployment of this upgrade commenced in late 2016 with an initial engineering array and has now progressed to the final stages of upgrading the full-scale SD array [94].

As part of this enhancement, each SD station will be equipped with an additional scintillation detector (SSD) positioned above and triggered by the larger WCD unit beneath it. This setup serves as a complementary method to aid in particle identification and measurement of the muon density on the ground. The SSD is read out by a dedicated PMT within its encasing. Coupled to the deployment is the installation of a low-cost radio antenna on top of each SD station, laying the foundation for a future large-scale radio detector [95]. The rendering in Fig. 4.3b depicts the additional equipment to be installed on top of a SD station [96].

With the instrumentation additions, there also will be an update to the SD electronics, enabling a faster sampling rate, an expanded dynamic range for PMT measurements in the WCD, and the incorporation of a SSD signal readout. Importantly, these ongoing upgrades are not anticipated to disrupt the continuous data collection. Apart from the installation of new solar panels, the existing power and communication systems will remain largely unchanged, with minor software adjustments to accommodate the added detection channels [96].

Apart from the changes to the SD station setup, the upgrade encompasses other projects, such as a dedicated Underground Muon Detector (UMD) setup as an addition to the existing AMIGA project (see Sec. 4.3.1). There will also be an extension of FD capabilities to allow for a further increase in the currently limited duty cycle [96].

As a consequence, dedicated photon searches are expected to profit in separation power from the complementary detection channels and the direct access to composition-sensitive observables, such as the muon content, for each shower. If this were to bring the potency of SD stand-alone analyses to the level of present-time hybrid studies, the previous state of upper bounds could simply be improved by nearly an order of magnitude through the expanded statistics (cf. Sec. 5.3).

4.2. Fluorescence Detector

The excitation of nitrogen molecules in the air by the charged particles within an air shower and the subsequent emission of fluorescence light has already been mentioned in Sec. 3.6 as a potential gateway for the direct measurement of the longitudinal profile of an extensive air shower. The fluorescence detector (FD) of the Pierre Auger Observatory is designed to detect this light emission within the airspace above the ground array using telescopes stationed at four FD sites. The prerequisite for this being that there are no other light sources in the field-of-view of the telescopes at the time of the shower and that the weak glow is not absorbed by other effects. Thus, the fluorescence detector can only be operated on clear and moonless nights, resulting in a duty cycle of approx. 15%, presently.

At each of the four FD sites of the Observatory (Los Leones, Coihueco, Loma Amarilla and Los Morados) a clean, climate-controlled building accommodates six individual and independent UV telescopes. The photography in Fig. 4.4a depicts the semicircular front of such a building. Each telescope possesses a field of view spanning $30^\circ \times 30^\circ$ in azimuth and elevation, collectively covering a 180° range in azimuth with minimal overlap. These buildings are strategically positioned to oversee the entire SD array from four different vantage points on small hills at the Observatory's periphery. The aim is to observe air showers impinging on the SD array above a certain threshold energy with at least one FD site. Although stereo observations (i.e., the triggering of two FD sites) were not a primary design criterion due to the already exceptional hybrid geometry reconstruction, they are still quite common at the highest energy levels.

Fig. 4.4b shows a schematic view of a single FD telescope. A shutter system safeguards the sophisticated optics from external elements and excessive daylight that could harm the highly sensitive electronics. When environmental conditions permit, the shutters are opened during the night for observations. Nitrogen light enters through the large UV-passing filter window and a Schmidt optics corrector rings in front of the telescope. Without this filter element, the fluorescence signal would be lost in the noise of visible photons. It also works as sealant against dust infiltration. The fluorescence light is focused by a $13\,\mathrm{m}^2$ mirror onto a camera with 440 pixels consisting of PMT light sensors. The main optical mirror is segmented to reduce both cost and weight of the optical system. The pixels of an FD telescope are arranged in a 22×20 grid matrix, with each pixel covering a hexagonal area. Light pulses in the PMTs are digitized every 100 ns. Similar to the SD design, there exists a hierarchy of trigger levels that involve trigger information from the ground array. Additionally, when an event is registered in the FD, it triggers the SD externally, and its data is recorded to facilitate hybrid reconstruction. Showers detected by both systems are termed *hybrid events* and provide more precise information regarding the event geometry and primary energy [97].

Calibration of the telescopes is achieved on two distinct levels. Firstly, for the absolute calibration, a portable, drum-shaped light source with known intensity provides a uniform pulsed flux of photons. With it, the FD telescopes are calibrated about once per year. Secondly, the relative calibration is automatically conducted before and after each observation run. This serves the purpose of monitoring the stability of the overall detector response over time and tracking any variations in the performance of individual components. To

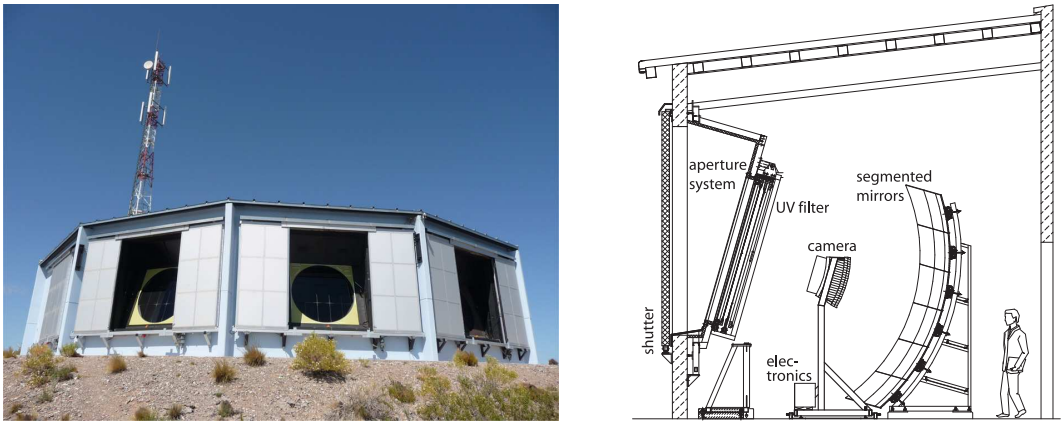

Fig. 4.4 The fluorescence detector (FD) of the Pierre Auger Observatory [85]. (a) FD building at Los Leones with its shutters opened during the day for maintenance. (b) Schematic cross-section of a single fluorescence telescope and its main components.

perform this relative calibration, controllable light sources are permanently integrated into the detector setup. These sources direct their light through optical fibers to various components along the optical path of the telescope. Additionally, the geometrical reconstruction of the FD system can be assessed using the Observatory's dedicated laser facilities (CLF, XLF).

4.3. Low-Energy Extensions

Once the baseline construction showed the effectiveness of hybrid measurements, major extensions were added to extend the capabilities of the Observatory toward lower energies. For the further course of this thesis, the two projects AMIGA and HEAT are essential; their operational methods will therefore be briefly explained here.

4.3.1. AMIGA – Auger Muons and Infilled Ground Array

The Auger Muon and Infilled Ground Array (AMIGA) facilities encompass a joint arrangement of additional WCDs and buried scintillation detectors, organized in a denser triangular sub-array configuration with an inter-station spacing of 750 m. The installation covers an area of 27.5 km^2 and is situated within the northwestern boundaries of the regular baseline array (1500 m spacing), positioned 6 km to the east of the Coihueco FD site.

Fig. 4.5 illustrates the spatial distribution of the associated SD stations in the field and the boundary defined by this infill structure. Some of these detector stations are shared with the regular array. At each station multiple scintillation modules are buried at depths of 1.3 m and 2.3 m in the ground, forming a dedicated detection system for the muonic component of air showers. It is worth noting that the inclusion of the muon detector in this context is primarily for comprehensiveness, as its relevance to the subsequent sections

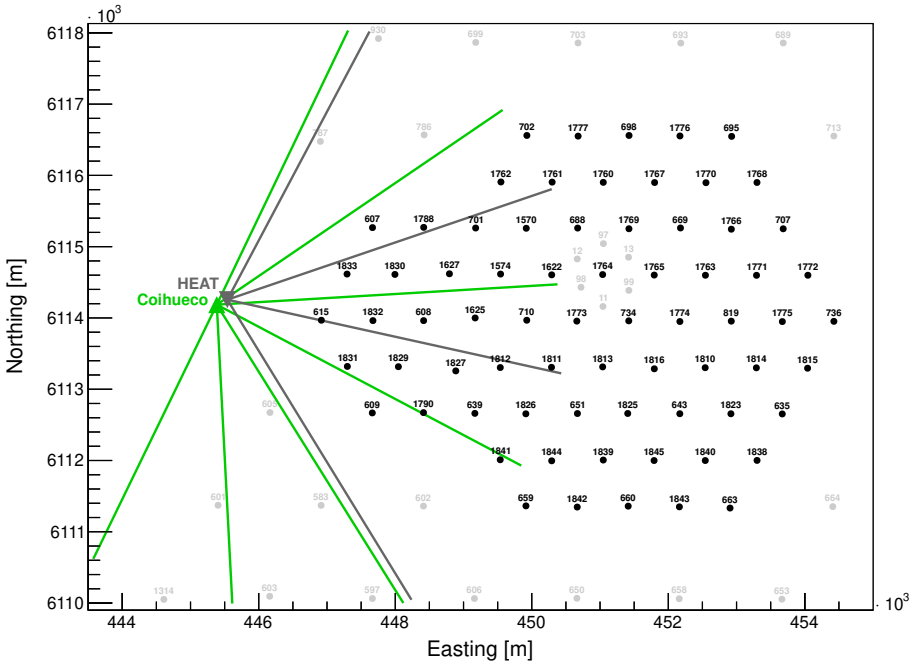

Fig. 4.5 Setup of the 750 m spaced infill array. Each black dot represents the position of a singular autonomous SD station on the ground. The gray dots are SD stations of different baseline arrays, which are usually not included in 10^{17} to 10^{18} eV analyses. The airspace over the infill array is monitored by the combined FD system of the Coihueco and HEAT site (referred to as HeCo), indicated by field-of-view in azimuth. [98]

of this study is negligible [85].

The 750 m infill array was completed in September of 2011. It is fully efficient for air showers above 3×10^{17} eV and with zenith angle $\leq 55^{\circ}$ [99]. The proven methods and tools used for the analysis of the 1500 m SD array have been adopted to reconstruct the lower-energy events. Although the infill area is much smaller than the regular array (covering less than 1%), the strong increase of the cosmic-ray flux towards lower energies allows a significant number of events to be observed and thus already enables the study of the cosmic-ray spectrum in the region between second knee and ankle.

4.3.2. HEAT – High Elevation Auger Telescopes

Located about 180 m away from the Coihueco FD site, three additional fluorescence telescopes were installed in 2009 which cover an elevated field-of-view. The High Elevation Auger Telescopes (HEAT) are very similar in design to the regular fluorescence telescopes, but are each housed in an enclosure that can be tilted by a hydraulic system in its entirety by 29° upward. A photographic depiction of the telescopes is given in Fig. 4.6. Together they form the fifth FD site of the Observatory, that, independent of the others, overlooks the 750 m infill array of AMIGA (cf. Fig. 4.5). This allows them to work complementary with the Coihueco site as a combined system, which is sometimes referred to as HeCo.

Fig. 4.6 The three HEAT telescopes in the upward tilted configuration [85].

The HEAT telescopes are specifically designed to cover the elevation range from 30° to 58°, a region beyond the field-of-view of other telescopes. This design facilitates the observation of cosmic-ray showers developing at higher altitudes, thereby encompassing a lower energy range, which enables the unbiased determination of both the cosmic-ray spectrum and the distribution of X_{max} up to the ankle and below the second knee. When combined with the SD infill array, this extends the energy range for high-quality hybrid air shower measurements down to 1×10^{17} eV. Furthermore, the system allows for the detection of close-inclined shower events up to the highest energies, without introducing any mass-dependent bias or requiring specific cuts [85].

4.4. Hybrid Event Reconstruction

Air shower events can be measured and reconstructed with the SD alone, but this leads to significant uncertainties in primary energy and direction of arrival, especially at the lower energies. However, if these measurements are supplemented by data from the FD, i.e., it then constitutes a hybrid event, the atmospheric depth of X_{max}, for example, can be determined much more accurately. Accordingly, there are also two distinct reconstruction procedures, depending on which detection data is provided. For sake of brevity, only the hybrid event reconstruction will be illustrated here. Based on the corresponding description in [85], it can be separated into five distinct, sequential stages:

1 Pulse Reconstruction at the FD

At the beginning of the reconstruction procedure for every FD pixel, the baseline and background is subtracted from the digitized signal. Subsequently, every triggered pixel is searched for the start and stop of a shower signal. The pulse time of the signal is then given by the centroid time, i.e. the signal weighted time.

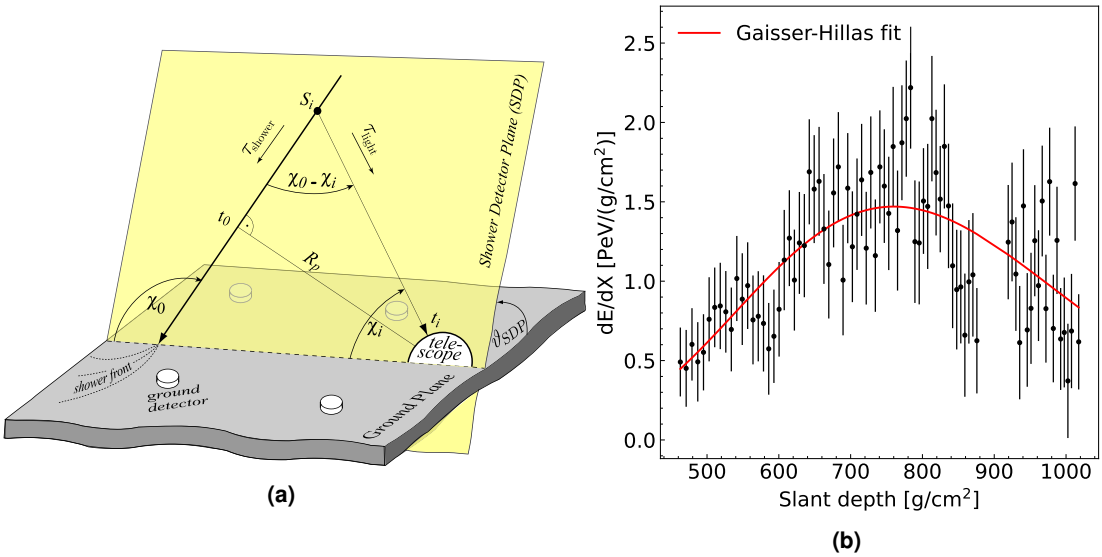

Fig. 4.7 Different steps of the hybrid reconstruction process. (a) Geometry in the shower-detector plane [100]. (b) Fit to the longitudinal energy deposit of an exemplary air shower recorded at Los Morados, which results in $X_{\max} = (760 \pm 11)\,\mathrm{g\,cm^{-2}}$ and $E_{\mathrm{hybrid}} = (9.7 \pm 0.7) \times 10^{17}\,\mathrm{eV}$. The event with ID 52094323300 is part of the Auger Open Data release [101, 102].

2 Fit of Shower-Detector Plane

The shower-detector plane (SDP) is the plane containing both the shower axis, and with it the shower core on the ground as recorded by the SD, and the triggered fluorescence telescopes. Its geometry is depicted in Fig. 4.7a. The SDP is reconstructed using the pointing directions of the triggered pixels.

3 Hybrid Time Fit

When the SDP has been determined, the shower axis is only free in a single parameter, χ_0, since it has to intersect the ground core position fixed by the SD measurement. The angular movement of the shower is then calculated using the timing information of the pulse pixels and of the triggered ground stations.

4 Light Collection

The total light at the aperture as a function of time is attained by adding the signals of the camera pixels at each time bin. Only the pixels with a pointing direction within a certain angular distance to the vector from the telescope to the shower core at the corresponding time are included.

5 Profile Reconstruction

With the known geometry of the air shower, the light flux at the aperture as a function of time can be converted to energy deposited as a function of slant depth ($\frac{\mathrm{d}E}{\mathrm{d}X}(X)$). For this all possible light-attenuation effects in the atmosphere as well as undesired external

influences, such as direct or scatter Cherenkov light or artificial light sources have to be taken into account. The energy deposited is calculated from the recorded fluorescence intensity based on the predetermined fluorescence yield, where the atmospheric conditions during the observation are once again an elementary factor. With the measurements of the energy deposit along the shower axis provided, the longitudinal development can be fitted with a Gaisser-Hillas function, analog to Eq. (3.11), and subsequently $X_{\max}$ determined (see Fig. 4.7b). A calorimetric energy is obtained by integrating the Gaisser-Hillas function over the total shower path. The primary energy is then estimated by correcting for invisible energy carried away by neutrinos and high-energy muons.

5. Ultra-High-Energy Photons

Astronomical observations started with humans observing and measuring visible light. Although dedicated particle and radiation detectors have assumed the role of observers today, the electromagnetic channel remains a crucial element for modern-day astronomy, with the photon being the main messenger particle for the exploration of the Universe. Energy-wise, the spectrum of measured gamma radiation was dramatically expanded over the last decades and now extends well into the regime of cosmic rays. The current upper limit of observed photon energies is on the order of $10^{15}\,\mathrm{eV}$ [103].

Any occurrence of ultra-high-energy (UHE) photons ($E \gtrsim 10^{17}\,\mathrm{eV}$) could be closely linked to other messenger particles, like charged cosmic rays and neutrinos, since various models predict UHE photon fluxes as by-products in a variety of acceleration and propagation scenarios for cosmic rays. Thus, observing such photons, in particular, would be key to completing the experimental multi-messenger approach aimed at understanding the most extreme processes in our universe capable of producing cosmic rays up to the energies observed today, while taking advantage of the fact that photons are not deflected by (inter)galactic magnetic fields.

Although the neutral photons are not commonly considered to be part of cosmic rays, they can similarly induce extensive air showers in the atmosphere when sufficiently energetic. This is exploited to systematically search for their occurrence with exposures surpassing any direct observational approach. The key for this is the efficient discrimination of such air showers from the nuclei-induced background through their characteristic signature.

This section aims to outline the peculiarities of the general search for UHE photons. After some details on their production and propagation (Sec. 5.1), in Sec. 5.2 the identification of photons as air-shower primaries will be detailed to the extent necessary to comprehend the design of the original analysis in [2]. The current status of photon searches at the Pierre Auger Observatory will also be reviewed (Sec. 5.3).

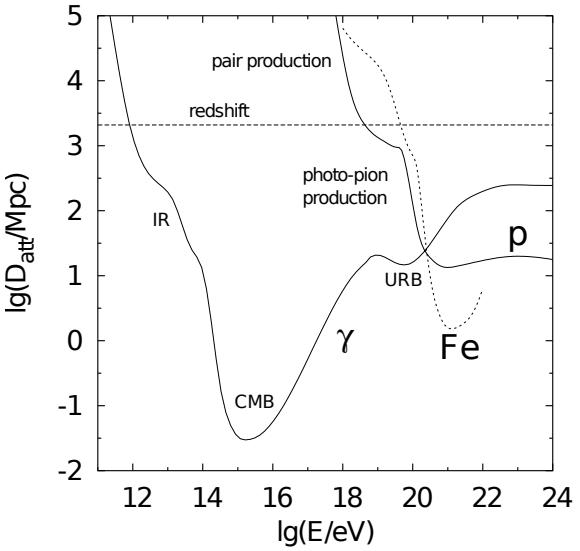

Fig. 5.1 Energy loss length for photons for interactions with infared (IR), cosmic microwave background (CMB) and universal radio background (URB). Furthermore the constant adiabatic energy loss, due to the expansion of the Universe is indicated (redshift). For comparison, the corresponding curves for energetic protons and iron nuclei are shown too [104].

5.1. Origin and Propagation of UHE Photons

Typical production and energy gain processes, like synchrotron radiation, inverse Compton scattering or bremsstrahlung, are not considered potent enough to yield photons of energies in the UHE range [54]. As viable source mechanisms only the annihilation or decay of UHE particles remain, which in themselves may be produced by some primary process.

Similar to the origin of cosmic rays (see Sec. 2.3) the production of UHE photons can be subdivided into two categories: In non-acceleration models, the primary process is given by the decay or annihilation of some super-massive primordial relics beyond the SM. However, the more conventional bottom-up approach is the production of energetic pions during the acceleration or propagation of cosmic rays [104], which inherently connects these photons to other forms of UHE radiation, among them neutrinos produced by the decay of charged pions, and in turn motivates multi-messenger observations:

$$\text{primary process} \rightarrow \pi^0(+\pi^\pm) + \ldots \rightarrow 2 \times \gamma_{\text{UHE}}(+\nu_{\text{UHE}}) + \ldots \tag{5.1}$$

Other possible sources for photons in this scenario include the delayed production by UHE neutrino scattering [105].

The most relevant primary process connecting UHE photons to cosmic rays is the theorized GZK effect, which entails the resonant photo-pion production of UHE nucleons with the CMB and has been discussed in Sec. 2.1.1 as a possible explanation for the observed cutoff in the cosmic ray spectrum. The energy of each these GZK photons would be reduced by a factor of ~ 10 compared to the initial cosmic ray nucleus [104].

Depending on the source scenario, UHE photons have to traverse significant distances in the extragalactic medium before they may get observed on Earth, which influences their spectrum. During the propagation phase, UHE photons can undergo electron-positron pair production, when interacting with one of the several present background fields. This initiates an electromagnetic cascade, which multiplies the particle number, while at the

same time lowering the individual momenta:

$$\gamma_{\mathrm{UHE}} + \gamma_{\mathrm{background}} \rightarrow e^{\pm} \rightarrow \cdots \rightarrow \gamma_{\mathrm{GeV-TeV}} + \ldots \tag{5.2}$$

Even though the alternating processes have diminishing cross-sections compared to their air-shower equivalents, on galactic and extragalactic propagation scales this effect still leads to significant energy degradation (see Fig. 5.1). The magnitude of this effect is not well known, primarily due to large uncertainties on the universal radio background (URB) [106], which enables the pair production at $E_\gamma \sim 10^{20}$ eV. For lower energies the influences of CMB and infrared background become relevant. Typical energy loss lengths for UHE photons range between 7–15 Mpc at 10^{19} eV and 5–30 Mpc at 10^{20} eV. The electromagnetic cascade continues down to GeV–TeV energies upon which the Universe becomes increasingly transparent for photons and the redshift, due to the expansion of the universe, becomes the dominant energy loss mechanism [104].

5.2. Experimental Measurement of Photon-Induced Air Showers

5.2.1. Methods of Discrimination

UHE photons do not penetrate the atmosphere unhindered. Analogous to cosmic rays, they also trigger extensive air showers, which, however, differ in essential aspects from the simple nucleus-induced case discussed in Sec. 3.2. The two main differences are schematically depicted in Fig. 5.2.

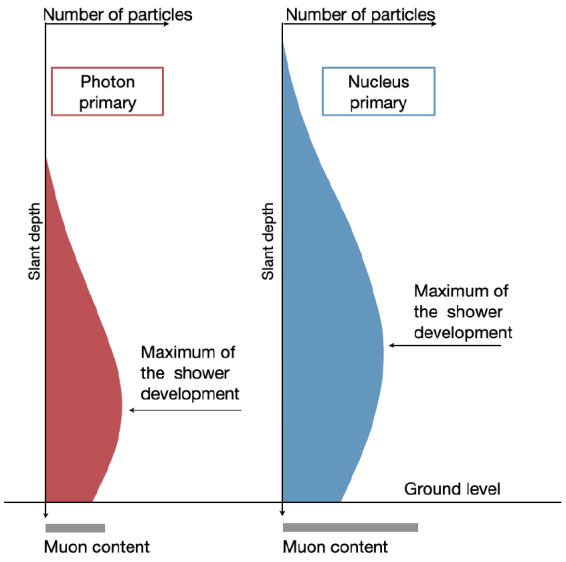

Fig. 5.2 Main features of photon- and nucleus-induced air showers [107].

When compared to nucleus primaries of similar energy, photon-induced showers develop later in the atmosphere, which leads on average to higher values of X_{max}. Earlier it was mentioned that this observable can be used to distinguish between different nuclear primaries, and here again the criterion provides the best discriminating power.

In addition, photon showers are almost purely electromagnetic, with alternating cascade steps of pair production and bremsstrahlung. Although suppressed by the factor $(m_e/m_\mu)^2$, the former process occasionally leads to the creation of muon-antimuon pairs. An additional source for them are photonuclear interactions that produce secondary hadrons and subsequent muons, with a cross-section more than two orders of magnitude smaller than that for pair production [104]. Both effects allow the development of a muon content, denoted with the total number N_μ, in a photon-induced air shower, which is not expected to be as numerous as in a comparable nucleus-induced shower. The lower

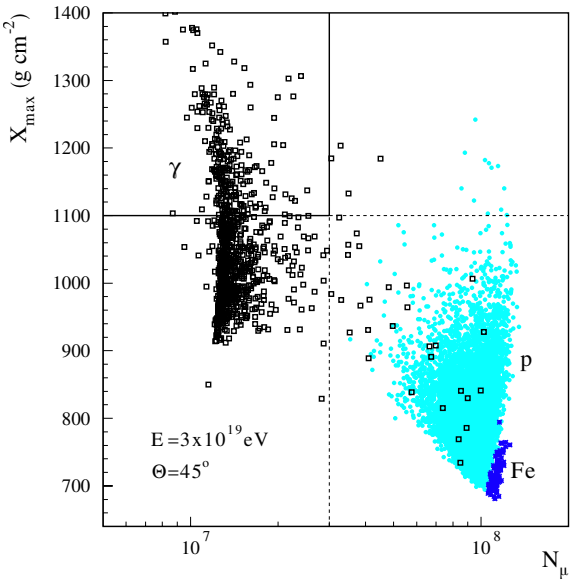

Fig. 5.3 Scatter plot of the main observables $X_{\max}$ vs. N_μ used for discriminating photon ind air showers from the nuclear background. Shown are MC simulations of photon (1000 events), proton (10 000 events) and iron showers (100 events), which include the pre-shower effect but do not consider any detector efficiencies. Further details about the theoretical model can be found in the source reference [104].

number of hadronic interactions allows a considerably more precise theoretical modeling of the shower, since the main limitation, the lack of knowledge about hadronic interactions at high energies, is mitigated. The observables of simulated photon showers at specific energies therefore carry much less systematic uncertainty.

Together, these two aspects form the key quantities used to distinguish photon showers from the nuclear background. Fig. 5.3 shows the separation between simulated air showers of different primaries in the two-dimensional observable space. While $X_{\max}$ can be measured directly with fluorescence telescopes, N_μ has influences on several observables, among them the temporal structure of the shower front signal in surface detectors. The characteristics of photon primaries in both quantities lead to a lower lateral signal spread on the ground, which observables like S_b [108, 109] or L_{LDF} [110, 111] employ for discrimination. As apparent from the iron population in the figure and simple considerations following the superposition model, heavier nuclei of similar energy behave even more extremely in this observable space. It is thus often sufficient to design searches for photon-proton separation, since proton-induced air showers can be taken as conservative worst-estimates for the nuclear background.

Due to different elongation rates $\mathrm{d}X_{\max}/\mathrm{d}\log E_0$ and the energy dependence of shower-to-shower fluctuations, the characteristics become more significant with increasing primary energy and subsequently allow for a more efficient separation between primaries.

5.2.2. High-Energy Effects

At the highest energies, the evolution of a photon-induced air shower is affected by two major incipient effects, that are not present for nucleus primaries. Both influence the average depth of the shower maximum, $\langle X_{\max}\rangle$, competing with each other: *Pre-showering* reduces $\langle X_{\max}\rangle$ and its event-by-event fluctuations, while the *LPM effect* increases both. In the following, the two phenomena will be outlined briefly.

Pre-Showering

In the geomagnetic field, UHE photons can convert into an $e^{\pm}$ pair before even entering the atmosphere, with the initial energy approximately halved for the two secondaries [112]. The leptons and numerous additional high-energy photons emerging from subsequent bremsstrahlung processes progress further toward Earth, enter the atmosphere and initiate respective electromagnetic air showers. The individual sub-showers are too superposed in spatial and temporal dimensions for current generation detectors to segregate them and are thus registered as only one event [113]. The pre-shower effect can therefore be considered as a jump-start to the regular air shower evolution, which shifts $X_{\max}$ to lower values.

Pre-shower probabilities depend strongly on the energy of the initial photon and the concrete trajectory through the magnetosphere. Typical altitudes for the initial pair production from a 10^{20} eV photon are ~ 1000 km, while the regular air shower development starts at ~ 100 km [104].

LPM effect

The LPM effect was postulated by Lev Landau and Isaak Pomeranchuk [114, 115] as an extension to the Bethe-Heitler formulation [116] of the cross-sections for pair production and bremsstrahlung at high energies in matter. The mathematical formulation by Arkady Migdal followed [117]. It states that in a sufficiently dense medium, the cross section for pair production is increasingly reduced due to destructive interference for high energies. The corrected cross-section σ_{LPM} depends on the energy of the photon E_γ and the density of the medium ρ:

$$\sigma_{\mathrm{LPM}} \propto (\rho E_\gamma)^{-\frac{1}{2}}, \tag{5.3}$$

In addition, the bremsstrahlung process is suppressed in a similar manner [118]. Therefore, due to the elongation of the radiation lengths for photon and electron, the development of electromagnetic air shower is delayed, which leads to higher values for $X_{\max}$ on average. Due to the negative correlation $\sigma_{\mathrm{LPM}}(X_2) < \sigma_{\mathrm{LPM}}(X_1)$ for depths $X_2 > X_1$, fluctuations of this quantity are also increased.

5.2.3. Detector Response

Differences in the shower development can affect the detector acceptance to and the energy reconstruction of UHE photons, when compared to nuclear primaries. To account for this effect, specific analysis approaches appropriate for the considered experimental apparatus have to be developed.

An example for this are near vertical photon showers that, due to their delayed evolution compared to nucleus primaries of similar energy, may reach the ground before being fully developed. The bias this introduces can be corrected for with detailed MC simulations. Another problematic aspect is the systematic deviation for the energy reconstruction of photon showers when the fluorescence method is used. The energy scales for photon and nucleus-induced showers nearly match, but the missing energy correction (see Sec. 3.6)

overestimates the primary energy for photon showers by about 10%. This leads to problems with the boundaries in energy of dedicated searches. For the reconstruction based on ground array measurement, even higher deviations from the actual primary energy may occur. However, when the plain photon flux rather than the fraction is calculated, only the photon energy scale is required, which means that the resulting values are not subject to this systematic uncertainty [104].

5.3. Current Status of Photon Searches in the UHE Regime

Above several 10^{15} eV [103] no astrophysical photons have been unambiguously identified so far. Nevertheless, even their non-observation can provide valuable insight for multi-messenger studies. Subsequent upper limits have been used to severely constrain source and propagation models for cosmic rays that predict specific levels of photon fluxes.

Limits on the integral photon flux can be formulated with the Feldman-Cousins approach [119]. The exact form depends on the design of the analysis. In the exemplary case of the photon search in [2], the upper limit on the integrated photon flux above E_0 at 95% confidence level (C.L.) is given by the formula

$$\Phi_{\mathrm{UL}}^{0.95}(E_\gamma > E_0) = \frac{N^{0.95}(E_\gamma > E_0)}{\epsilon_{\mathrm{cand}} \ (1 - f_{\mathrm{burn}}) \ \mathcal{E}_\gamma(E_\gamma > E_0 | E_\gamma^{-\Gamma})}, \tag{5.4}$$

where $N^{0.95}$ is the Feldman-Cousins upper limit at 95% C.L. on the number of photon candidate events assuming zero background, ϵ_{cand} is the efficiency of the photon candidate cut, f_{burnt} is the fraction of the data used in a burn sample, and $\mathcal{E}_\gamma$ is the integrated efficiency-weighted exposure for photons, which in turn is defined by the multidimensional integral

$$\mathcal{E}_\gamma(E_\gamma > E_0 | E_\gamma^{-\Gamma}) = \int_{E_0}^{\infty} \mathrm{d}E_\gamma \int \mathrm{d}S \int \mathrm{d}t \int \mathrm{d}\Omega \, \frac{E_\gamma^{-\Gamma}}{c_E} \, \epsilon_\gamma(E_\gamma, t, \theta, \varphi, x, y). \tag{5.5}$$

Here $\epsilon_\gamma(\ldots)$ is the overall efficiency for photons, depending on the photon energy E_γ, the time t, the arrival direction (zenith angle θ and azimuth angle φ), and the impact point (coordinates x and y) of the air shower on ground. The integration is performed over area S, time t, solid angle Ω, and photon energy E_γ. The normalization factor c_E includes just the integration of the spectrum $E^{-\Gamma}$ over all values $E_\gamma > E_0$ [2].

Various air shower detection facilities have measured the integrated photon flux in the UHE regime over the years, including the Pierre Auger Observatory. The four currently most stringent analyses of the Pierre Auger Collaboration are presented in Tab. 5.1; the limits established on the photon flux may be found in the respective publications. By using the different detector components, a wide energy range can be covered. The different energy lower limits of the analyses are also listed in the table. Updating the analysis in [2], that employs the low-energy extensions of the Observatory for its search, is the overarching focus of this thesis.

Detector	E_0 [eV]	Reference
SD 433 m + UMD	$(5, 8, 12, 20) \times 10^{16}$	[120]
HeCo + SD 750 m	$(2, 3, 5, 10) \times 10^{17}$	[2]
FD + SD 1500 m	$(1, 2, 3, 5, 10) \times 10^{18}$	[121]
SD 1500 m	$(1, 2, 4) \times 10^{19}$	[111]

Tab. 5.1 Current diffuse searches for UHE photons using various facilities of the Pierre Auger Observatory.

The limits on the integral photon flux above a given threshold it imposes are displayed in Fig. 5.4, in a compilation with the results of the other analyses, and compared with expected photon-flux bands from several still viable astrophysical and cosmogenic origin scenarios for various lower energy bounds between 10^{17} eV and 10^{20} eV. Furthermore, the calculated state-of-the-art upper limits of some other air shower detection facilities are added for context and comparison.

In addition to efforts to search in a diffuse manner, there have also been analyses in the past that used measurement data from the Pierre Auger Observatory to specifically search for point sources of UHE photons. The analysis in [127] performed a blind search for excesses in arrival direction of photon-like events in the energy range between $10^{17.3}$ eV and $10^{18.5}$ eV, while its successor [128] searched in the same energy range for cluster in correlation with the known position of various astrophysical targets in the sky. However, no such significant excesses were found in either case.

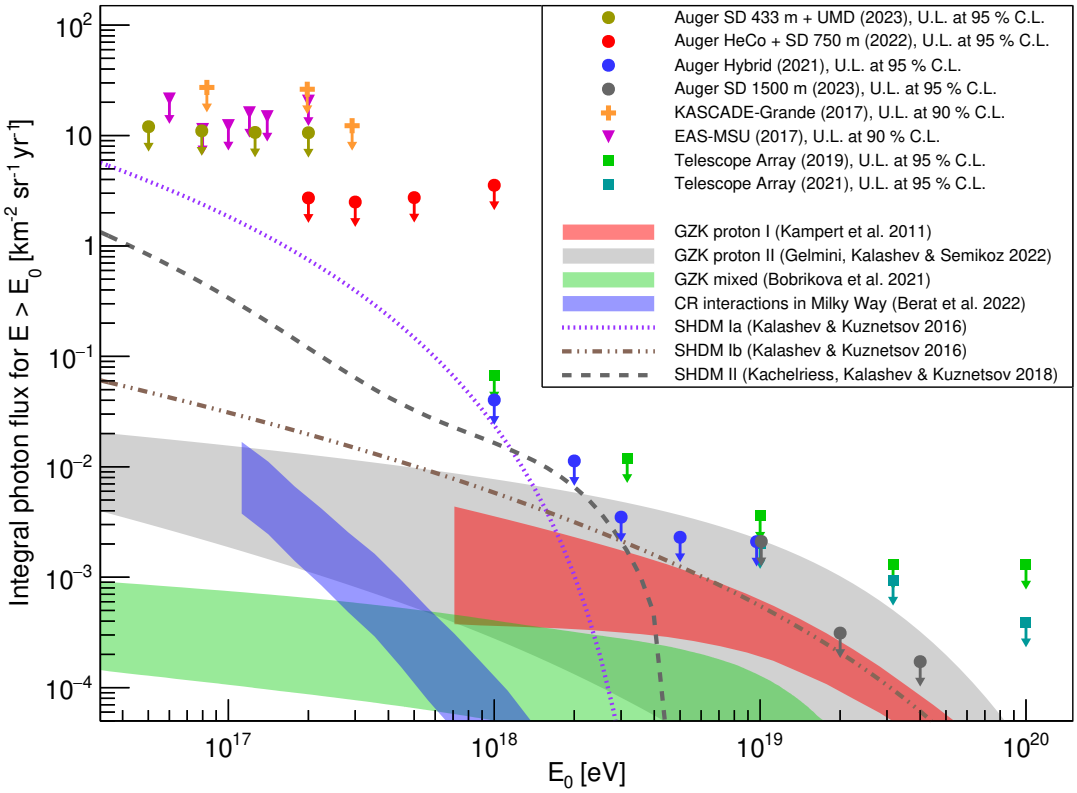

Fig. 5.4 The current status of limits on the integrated flux of UHE photons from measurements at the Pierre Auger Observatory [2, 111, 120, 121], Telescope Array [122, 123], KASCADE-Grande [124] and EAS-MSU [125]. Note that the last two, different from the rest, are given at 90% confidence level. Expected levels from several astrophysical, among them cosmogenic GZK models, are represented by the colored bands [126].

Updating the hybrid search for photons above 10^{17} eV

This part of the thesis will now draw from the physical and technical foundations laid previously to first introduce the research work (Sec. 6) and the data sample that is to be updated (Sec. 7), before taking a closer look in Sec. 8 at particular aspects of the analysis, which will be studied individually, before finally issuing recommendations for the further progress of the update. The chapter concludes with accompanying results of the extension of the data period through the initial production of a new extensive simulation set (Sec. 9) and an outlook on the forthcoming update steps (Sec. 10).

6. Design of Original Analysis

The original analysis was published in July 2022 [2]. It extends the photon search of its predecessor [129], published in April 2017, that was designed for the energy range above 10^{18} eV, toward lower energies, while employing an analog hybrid design with the low-energy extensions of the Pierre Auger Observatory, namely HEAT and AMIGA, that have been detailed previously. The original publication of the study was already accompanied by the idea of tailoring the analysis further to the specific detector setup and the lower energies, while extending the data period to its maximum attainable range, the full Phase-1 dataset, with an ensuing publication toward a later date. The rough design of the original analysis is outlined below to the extent necessary to formulate the subsequent update steps. For more details, see [2].

The analysis is based on hybrid data collected by the Coihueco and HEAT telescopes, henceforth referred to as "HeCo", in addition to the 750 m SD array of AMIGA, which in turn is now designated as the infill array, between 2010-06-01 and 2015-12-31. In addition, realistic detector and reconstruction simulations were created to optimize the sensitivity of the analysis for photons. The simulation set includes approximately $^2/_3$ photon-induced air showers and $^1/_3$ background events, for which proton showers were chosen as a conservative estimate for nuclei-induced events (cf. Sec. 5.2.1). The high quality of the recorded events was ensured by a variety of selection criteria, which were applied to both the data set and the two simulation samples. Details about the creation and selection process for the different samples are provided in the subsequent section.

In order to obtain the desired sensitivity for a small photon signal, a multivariate analysis (MVA) combining various event-specific observables into a single classifier was employed, as is common in high-energy physics. Specifically, a boosted decision tree (BDT), as implemented in the TMVA toolkit, version 4.20, [130], was used. While it is possible to use practically raw data for event classification with its common alternative, an artificial neural network, preprocessing is required for a BDT, since more tangible physical parameters are used as input variables. In contrast to convolutional neural networks, the influence of the

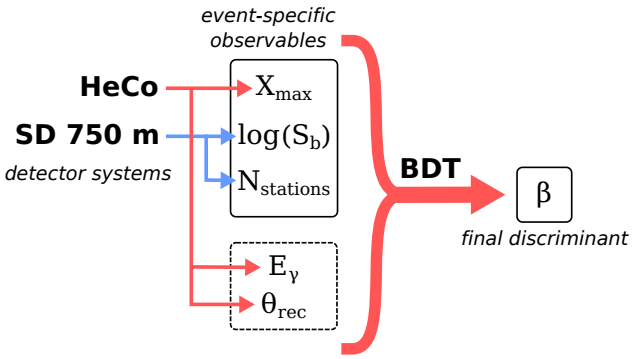

Fig. 6.1 Schematic design of the event-by-event classification process of the original analysis. Both HeCo and infill array measurements are used to define discriminatory quantities ($X_{\mathrm{max}}, \log(S_b), N_{\mathrm{stations}}$), which are used by the BDT to calculate a final scalar discriminant β. Due to dependencies of the observables, E_γ and θ_{rec} are also included to avoid data binning.

various input parameters is much more transparent. In addition, the time required to train the MVA is usually shorter with BDTs. Nonetheless, similar discriminatory power can be achieved in both methods [131, 132].

For the MVA, the two major structural distinctions between photon- and hadron-induced air showers were exploited: Photon shower develop later in the atmosphere and contain overall a lower number of muons. Fig. 6.1 illustrates the classification process of the original analysis using a BDT. The employed input variables are physical observables which are sensitive to one or both of the above differences. Firstly, there is X_{max}, the atmospheric depth of the shower maximum, which is directly observed by the HeCo telescopes. Furthermore, the observable S_b is used for discrimination. Originally proposed in [108] as a composition-sensitive parameter, its potential for photon searches was realized in a subsequent publication [109]. It is defined as follows:

$$S_b = \sum_i S_i \times \left(\frac{R_i}{1000\,\mathrm{m}} \right)^b , \tag{6.1}$$

where the sum runs over all the triggered SD stations. S_i is their corresponding signal (in arbitrary units) and R_i their distance to the reconstructed shower axis. b is a free parameter, chosen as $b = 4$ in accordance with [129]. By construction, S_b is sensitive to the lateral particle distribution, which depends on both the depth of air shower development and the number of contained muons. Finally, the total count of triggered SD stations, referred to as N_{stations}, is included as a rather simple observables, since it has been shown in [129] to have the potential to significantly improve the overall performance of the analysis. Additional dependencies of the discriminatory observables are taken care of by including the photon energy estimate E_γ as well as the reconstructed zenith angle θ_{rec}, to mitigate the need for event binning in the training and deployment stages of the MVA.

The BDT is trained with a randomly chosen subset of the simulation sample containing $^2/_3$ of the provided events, $^1/_3$ is devoted to testing and validation. For both training and testing, the events are weighted on an individual basis according to a power-law spectrum $E^{-\Gamma}$ with $\Gamma = 2$ as in previous photon searches. The separation capability of the BDT can be quantified with the specification of the background rejection at a specific signal efficiency. For 50% signal efficiency, this is the ratio of background events beyond a potential cut at the median of the photon distribution. With the above described setup, the BDT reaches a background rejection at 50% signal efficiency of $(99.91 \pm 0.03)\,\%$ above 2×10^{17} eV.

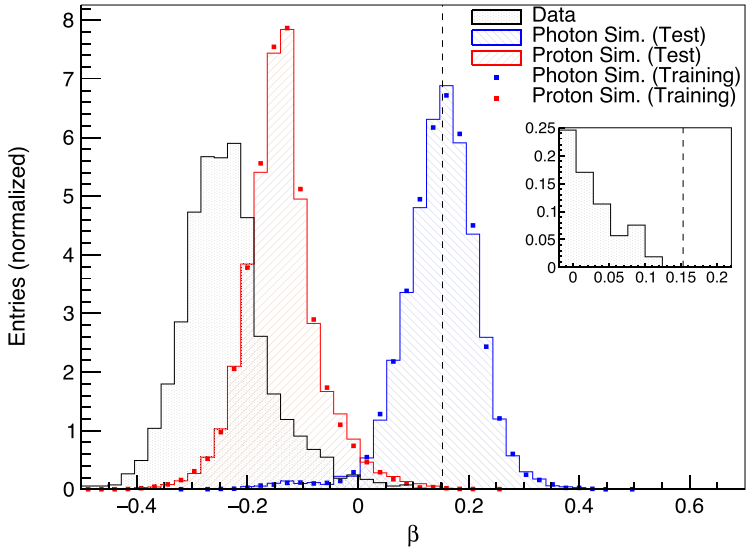

Fig. 6.2 Normalized distributions of the final discriminant β for events with $E_\gamma > 2 \times 10^{17}$ eV. The photon sample is shown in blue, the proton sample in red and the data sample in black. For both simulation samples, training and test sub-sets are shown. The dashed line denotes the median of the photon distribution, which marks the photon candidate cut of the analysis. The inlay shows a close-up of the data distribution near the photon candidate cut. No data events pass the candidate cut [2].

Based on the calculated differential exposure determined in part from the simulation data, the lowest energy limit of the analysis was chosen as 2×10^{17} eV. An analogous procedure was carried out for the lower limits 3×10^{17} eV, 5×10^{17} eV and 1×10^{18} eV. Similar to earlier works, the median of the β distribution of the photon simulations above 2×10^{17} eV was chosen as the candidate cut in all energy bins. Due to this fixed value the efficiency for photon events deviates from the 50% for the higher energy bounds.

Classification of the 2204 data events that remain after the selection cuts and the subtraction of the burnt sample leads to no identification of photon candidates. The distribution of β for the data samples is depicted in Fig. 6.2, together with the corresponding simulation distributions. For a pure-proton background, 1.98 ± 0.66 events would have been expected from the background alone. The non-observation has been translated with Eq. (5.4) to subsequent upper limits on the integral photon flux, which have already been presented earlier in Fig. 5.4. For the exact values and details about their calculation, refer to [2]. Their relative differences are largely influenced by the energy-dependent acceptance of the low-energy detector setup and the declining photon spectrum.

7. Data and Software

The original burn-sample and the simulations have been employed for the fundamental studies in the subsequent section. For this reason some details on their scope and compo-

sition, as well as the creation process for the simulations will be given in this section. It is important to emphasize, that for the studies in this thesis, only the selections of data and simulated event preceded by the application of the quality-ensuring selection cuts of [2] have been accessed. The section concludes with details on additional software usage that has been essential for this thesis and the studies therein.

7.1. Original Data Sample

The original data sample contains 557 944 HeCo events in the aforementioned time period. After all selection levels, 2308 events above 2×10^{17} eV remain, which are further subdivided into 104 events that constitute the sample employed for testing and optimizing the design of the study and the remaining 2204 events. Due to possible a-posteriori adjustments of the method, the corresponding events of the former are excluded from the final analysis, constituting the burn-sample. Hence, the above results are only based on the second subsample. In order to achieve a sufficiently random, yet repeatable and extensible collection of events for the burn-sample, the selection was realized with the simple prescription $T_{\mathrm{GPS}} \mod 20 = 0$, where T_{GPS} denotes the time the event was recorded in units of GPS seconds [2]. Accordingly, the targeted share of events in the burn sample is 5%.

7.2. Original Simulation Sample

For the simulation samples, previously prepared air shower simulations of the *Napoli+Praha* library [133] were used. They have been performed with CORSIKA [134], version 7.6400, employing EPOS-LHC [135, 136] as the high-energy hadronic interaction model. A smaller additional sample of showers has been produced in Siegen with the same CORSIKA settings. Although they do not have a significant impact on the shower development in the target range $< 10^{18}$ eV, both pre-showering and the LPM effect have been taken into account. The showers are approximately uniformly distributed in $\log(E)$ and $\sin^2(\theta)$.

The detection and reconstruction process is simulated with the $\overline{\mathrm{Off}}\underline{\mathrm{line}}$ Software Framework, revision 30689 of the v3r3 branch from the Auger SVN [137]. Here, every simulated shower is resampled five times with a newly associated core position and time to form five unique events. Whereby the event time is randomly drawn between 2010-06-01 and 2015-12-31, with the drawn time only accepted if the status logs indicate concurrent uptime fractions for all HeCo telescopes above 95%, to ensure that the detector was actually taking data in this time period. Moreover, the position of the shower core on the ground is randomly determined within the boundaries of the infill array plus a 1.5 km margin around this area. Both procedures have been implemented to ensure the creation of a realistic MC sample that resembles the corresponding data closely. Accordingly, the status of the detector components is read from the logs and is realistically reproduced in the simulation environment as it were during the randomly determined event time.

After the aforementioned selection cuts, 55 158 photon-induced and 33 279 proton-induced simulated air showers remain for studying. Due to the variable detector acceptance and the

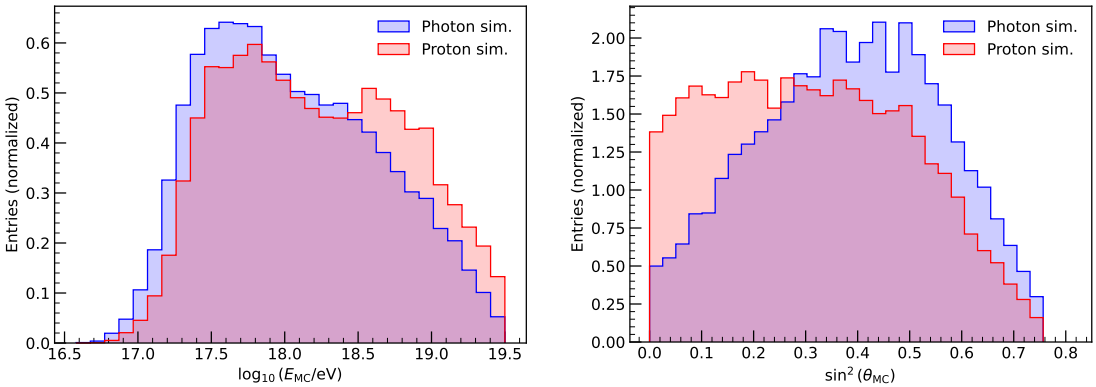

Fig. 7.1 Distributions of $\log(E/\mathrm{eV})$ (left) and $\sin^2(\theta)$ (right) for the original simulation samples after the quality ensuring selection cuts of the original analysis. The photon distribution is shown in blue, while proton-induced events are indicated by red color. The two histograms show the normalized densities of the events for better comparability between the different sized samples.

implemented quality cuts, the distributions of the events in energy and arrival direction, as illustrated in Fig. 7.1, are slightly modified in the process.

7.3. Further Software Usage

Besides the creation of simulations, the data analysis stage of the original study was also performed with Offline, using the same version as specified above, and the underlying ROOT framework [138], implemented in C++. As part of this thesis, the original analysis code was transferred to Offline v3r99p3, ICRC '23 production release 1, compiled with g++ 9.3.0. It has been rerun and checked for consistency, while relying solely on the internally published documentation as a means of checking it for completeness and comprehensibility at the same time. No deviations from the published results of [2] were identified in the process.

The studies conducted in the context of this thesis employ the same updated version of the Offline framework, with the included ADST package being used to access the data of recorded or simulated air showers. Subsequent data analysis and plotting was either performed within the ROOT framework and C++, if performance was critical, or within Python 3.10.12. For the latter, ROOT storage files were directly accessed with *uproot* 5.0.11 [139] and plotting was performed with the *matplotlib* 3.7.2 library [140].

8. Fundamental Studies and Proposed Updates

The following subsections each address a single aspect of the original analysis design. Whereby, in each case, the available data from the original analysis, i.e. burn sample and photon/proton simulations, detailed above, was used as a directly available data basis

in order to ensure that the expenditure of time was kept within reasonable bounds for a Master's thesis. The results presented here should therefore be regarded as suggestions for a potential future study design, accordingly verified when the updated data samples are attained and also tested for compatibility with one another.

8.1. Optimizing the S_b exponent

The air-shower observable S_b has been used as an SD-only composition sensitive parameter in the original analysis. In the context of photon searches by the Pierre Auger Collaboration, it has become popular in recent years, with several studies employing it. Even beyond the initial field of application, S_b has great influence. For instance, the observable M_b, which was designed for the muon detector UMD in [141], is directly based on the definition of S_b respectively $\log(S_b)$. Another example would be the adoption of S_b for the usage in the future Southern Wide-field Gammay-ray Observatory (SWGO), that is currently ongoing and complicated due to the non-constant density of the detector array [142].

The observable S_b was originally proposed in [108] for the purpose of shower-by-shower discrimination between various nuclear species in the UHE regime of the cosmic-ray flux. In Eq. (6.1), the definition of S_b and the associated quantities has been given within the preceding section. Per construction, the exponent b of the normalized distance term in the sum is left free but constant. The original publication contained an analytic study in combination with a numerical approach, that placed the optimal value for proton-iron separation in the UHE region near $b = 3$. Furthermore, the dependence on the array configuration, zenith angle and hadronic interaction model were reported as negligible.

Following the original publication, a subsequent study [109] expanded upon this concept exploring the observable as a means of separating photons in the cosmic-ray flux from the exceedingly dominant nuclear background on the basis of the lateral signal topology. Here again, a well-defined value for b exists that maximizes the discrimination potential. At $E \simeq 10^{19}$ eV the optimal value is again reported as $b \simeq 3$.

In both cases, the merit factor η was employed for the optimization procedure as the key metric of separation power and consequently maximized. It is defined as follows:

$$\eta = \frac{\left|\mathrm{E}[q_\gamma] - \mathrm{E}[q_\mathrm{p}]\right|}{\sqrt{\mathrm{Var}[q_\gamma] + \mathrm{Var}[q_\mathrm{p}]}}, \tag{8.1}$$

where $\mathrm{E}[q_i]$ and $\mathrm{Var}[q_i]$ are the mean, respectively variance of the photon (γ) and proton (p) distribution of the observable S_b. When varying the value of b, the merit factor of the observable distributions shows a clear peak. The underlying causes for the formation of this maximum in separation power are explained in more detail in [109].

The merit factor is particularly well suited for distributions closely resembling a Gaussian shape as higher distribution momenta are neglected. For observables that exhibit highly asymmetrical distributions, it is therefore more appropriate to utilize a different parameter. Such a metric that is suited for quantifying the separation performance in an arbitrary

two-category classification problem is the background rejection $\rho(\varepsilon)$ at a specific signal efficiency ε. It is essentially counting the events of the (expected) background sample, here proton-induced, that are located in the observable space within a specific percentile of the signal distribution, here photon showers. As has been mentioned in the preceding sections, in the context of photon searches, the median of the photon distribution has often been utilized as the photon candidate cut. Consequently, and when the S_b observable is used, the definition of the background rejection would look like

$$\rho(\varepsilon = 0.5) \equiv \frac{N\left(S_b^{(p)} > S_{b,\mathrm{median}}^{(\gamma)}\right)}{N_{\mathrm{total}}}. \tag{8.2}$$

Previous usage of the S_b observable by the Pierre Auger Collaboration has been inconsistent. After its inception, it was first adopted by the two directional photon searches [127, 128], in 2014 and 2017, that were mentioned earlier. Both operated in the photon energy range between $10^{17.3}$ eV and $10^{18.5}$ eV. A BDT was used as a classifier, with S_b among the input variables, whereby $b = 3$ was chosen, citing the optimization in [109]. A sort of paradigm change took place with its next published usage in 2017. The study in [129] again employed a BDT above 10^{18} eV to discriminate photon events from the nuclear background. Here, $\log(S_b)$ was used as an input, with $b = 4$ optimized for separation power above 10^{18} eV. Further details on this choice were omitted from the publication.

The study of [2], utilizing the low-energy extensions, was designed for consistency with [129], and consequently inherited the usage of $\log(S_b)$ with $b = 4$. However, there is no direct justification for the choice of this value. Additionally, the findings of the original optimization [109] show a slight dependence on energy for the optimal value of b when using the plain S_b, though this trend was only examined within the range $18.5 < \log_{10}(E/\mathrm{eV}) < 19.6$. Furthermore, for $\log(S_b)$ no such optimization has been published, and it is unclear if the previously reported independence from array configuration, i.e. array spacing, also holds for the logarithmic variant, and therefore whether previous findings can be transferred to the infill array at all. Moreover, there is no inherent reason for sticking to integer values of b.

The above reasons provide motivation for a renewed study regarding the usage of S_b in the energy range of the low-energy extensions of the Pierre Auger Observatory. The aim is to assess the usage of $\log(S_b)$ over plain S_b, and provide recommendations for the optimal value of b. Due to the envisaged energy range of the study to be updated and considerations arising from the efficiency-weighted exposure of the detector setup, as well as the range covered by the optimization in [109], the following study was constrained between 10^{17} eV and 10^{18} eV. However, it is quite conceivable that a follow-up may extend the procedure at least up to $10^{18.5}$ eV.

8.1.1. Numerical Study

The study was conducted before updated detector simulations for both photon- and proton-induced air showers were available. Therefore, the simulation samples of the original anal-

ysis, as detailed previously, were used as a basis. With the constraint of the primary energies between 10^{17} eV and 10^{18} eV and the preceding quality ensuring selection cuts, 26 665 photon- and 14 880 proton-induced hybrid-events remained for analysis, approximately logarithmic-uniformly distributed in energy. Regarding the zenith angle, the events are uniformly distributed in $\sin^2(\theta)$ with the limit $\theta < 60^\circ$. For consistency reasons with [109], the merit factor η was chosen as the relevant measure of separation power.

The simulations were divided into logarithmic energy intervals of width $\log_{10}(E/\text{eV}) = 0.2$. The number of simulated events in each energy bin is given in Tab. 8.1. Within each of these, the b value was varied in increments of 0.1 over the range of 1.0 to 6.0, and the resulting S_b distributions were evaluated using the merit factor η. Furthermore, the uncertainty of the merit factor was estimated for each b value through the use of a bootstrapping method, with 10 000 iterations performed to strike a balance between computational efficiency and accuracy. The exact same procedure was conducted for the logarithmic variant $\log(S_b)$ with the only difference being a widening of the scanned b range toward an upper limit of 10.0 for reasons detailed later on. Since the logarithm is not a linear transformation, the position of the optimal b value can be expected to change slightly, requiring this renewed analysis. Here, the results of the study for plain S_b will be presented first, after which the findings for $\log(S_b)$ will be addressed.

As an illustrative example, the visual representation of the impact of different components described by Eq. (8.1) on peak formation can be observed in Fig. 8.1. The two curves depicting the mean values exhibit a rather strict decrease as the parameter b grows. This behavior arises due to the individual distances between the shower core and the detector stations being smaller on average than the reference distance of 1000 m. Consequently, a higher exponent in the second term of Eq. (8.1) leads to a suppression of the overall sum, resulting in progressively smaller values that tend to approach zero asymptotically. This effect is particularly prominent when using the 750 m array configuration and dealing with lower primary energies. However, the actual reason behind the improvement or deterioration of the discriminating capability is not manifested in the separation between the mean values of the distribution, as both exhibit a similar trend. Instead, it is reflected in the broadening of both distributions towards the boundaries of the examined range of b values. This broadening ultimately yields small values of η for both small and large b values during the scan across the range of the parameter, as illustrated by the example

$\log_{10}(E/\text{eV})$	photon events	proton events
17.0–17.2	1585	777
17.2–17.4	4370	2385
17.4–17.6	6844	3732
17.6–17.8	7218	4082
17.8–18.0	6648	3904
Σ	26 665	14 880

Tab. 8.1 Number of simulated events in the energy bins.

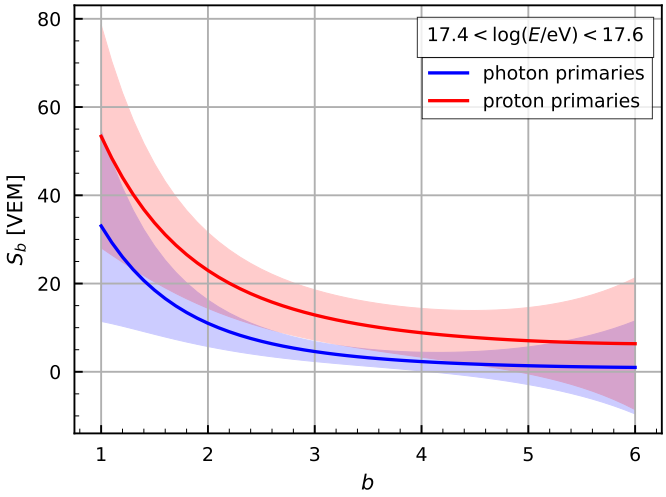

Fig. 8.1 Mean S_b distributions in the energy bin $17.4 < \log_{10}(E/\text{eV}) < 17.6$. The standard deviation is indicated by the shaded regions. Note that the observable S_b is strictly positive defined and the widening for high values for b is therefore of asymmetric nature. The corresponding plots for the other energy bins can be found in the appendix in Fig. D.1.

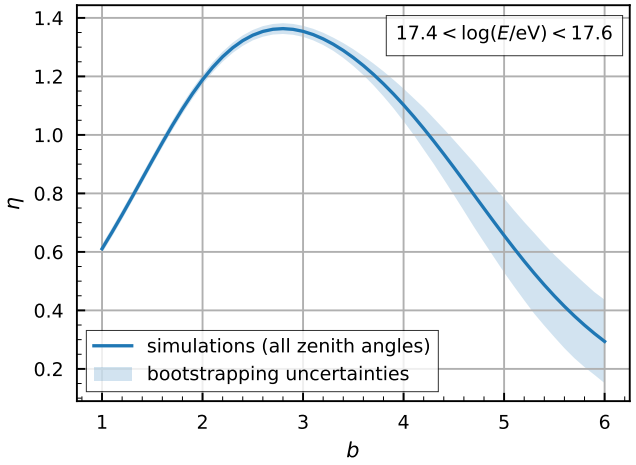

Fig. 8.2 Merit factor η and its uncertainties for S_b as a function of b in the energy bin $17.4 < \log_{10}(E/\text{eV}) < 17.6$. The uncertainties have been determined through a bootstrapping method. Corresponding plots for the other energy bins can be found in the appendix in Fig. D.2.

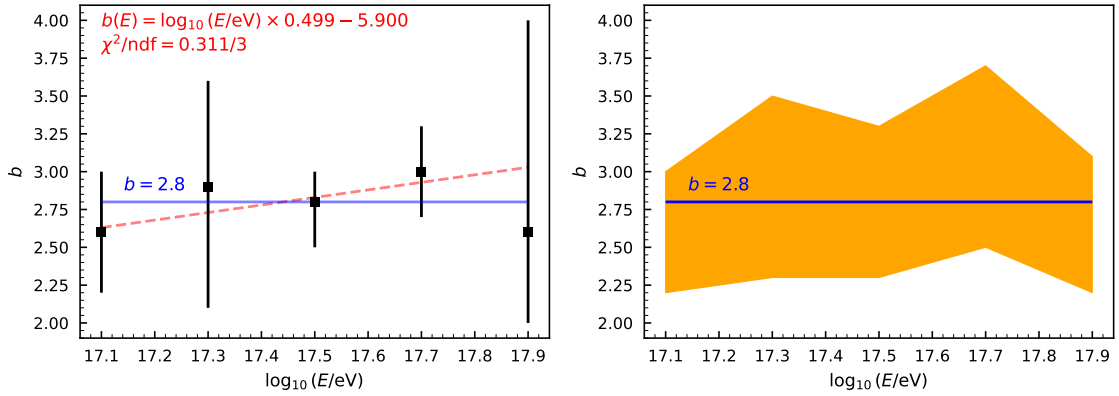

Fig. 8.3 Left: Optimal value b_{opt} as a function of primary energy, when using plain S_b, together with a linear fit (red) and the constant values as proposed in the text (blue). The uncertainties are estimated from the width of the bootstrapping band at the peak in Fig. 8.2 and corresponding scans. Right: Band that represents a 5% variation in η, together with the constant value.

depicted in Fig. 8.2 for a specific energy bin. The widening of the distributions primarily originates from the variances associated with individual detector-to-shower-axis distances and signals. At low b values, only the detector stations in close proximity to the shower core significantly contribute to the overall sum, while at high b values, the opposite effect occurs, with only the stations furthest from the core playing a significant role. Typically, the averaging process over multiple signals at varying distances masks individual fluctuations. However, when single stations become the predominant contributors to the sum represented by the S_b value, this masking effect is no longer applicable and conversely a broadening of the distributions takes place.

The analysis of data from the five energy intervals compiled in Fig. 8.3 reveals that, when using the plain S_b observable, the preferred values for the exponent are in good agreement with previous findings above $10^{18.5}$ eV [109], suggesting that the optimal values for b in the energy range studied lie close to $b = 3$. The application of a linear model (Fig. 8.3, left) shows a slight incline with increasing energy, similar to the one published in the higher energy range. As there is no theoretical justification for restricting the values of b, 2.8 may be used in the stand-alone case, since this is near the best for a constant value and yields a merit factor that is well within 5% of the maximum value observed in any of the energy intervals analyzed (see Fig. 8.3). It has been previously presented that the value optimized in photon and proton simulations also provides the best separation when a significant contribution of heavier nuclei are considered [109], which may be the case in this energy range.

Recent photon searches of the Pierre Auger Collaboration appear to have progressively switched to using the logarithmic variant as an input for multivariate analyses, instead of the plain observable S_b. While the direct effects on the discriminatory potential, especially for complex analysis, are unclear, an immediate advantage can be identified. Due to its strong energy dependence, the numerical range of the observable typically spans several orders of magnitude, when analyses are not restricted to a very narrow energy range. The logarithm, however, compresses these values into a manageable numerical range that is better suited for human handling and data visualization.

Due to the strictly monotonic nature of the logarithm, the distributions of $\log(S_b)$ are expected to exhibit a peak in separation power at a distinct value of b as well. When the aforementioned analysis procedure is applied to the logarithmic variant, its non-linearity, however, will result in a different behavior of the observable when varying b. The effect of the logarithmic mapping on the position and width of the observable distributions is depicted in Fig. 8.4. As expected, the shape of the merit factor scan in Fig. 8.5 differs from the plain S_b case. The peak shifts to slightly higher b values and the region above the maximum stays at a relatively stable η value for much longer. Overall, the profile displays similarities with a saturation effect. To ensure that the 5% variation interval could still be measured, the range of b values scanned was extended up to 10.0.

Fig. 8.6 presents the optimal values of the S_b exponent along with the corresponding peak width measurement. It is evident from the linear fit that the location of the maximum is now strongly dependent on the energy. Nevertheless, due to the wide 5% band in the

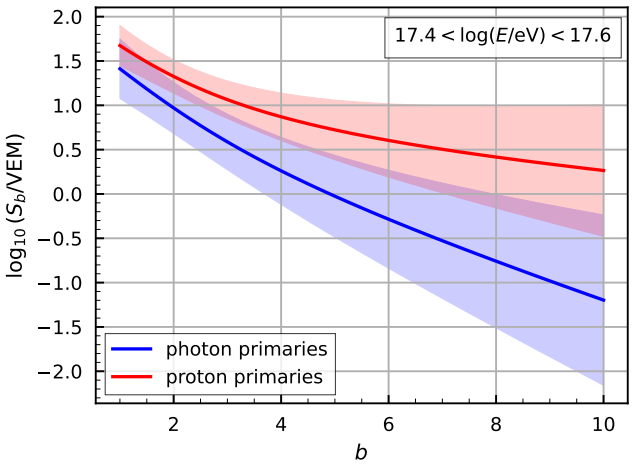

Fig. 8.4 Mean $\log(S_b)$ distributions in the energy bin $17.4 < \log_{10}(E/\mathrm{eV}) < 17.6$. The standard deviation is indicated by the shaded regions. The corresponding plots for the other energy bins can be found in the appendix in Fig. D.3.

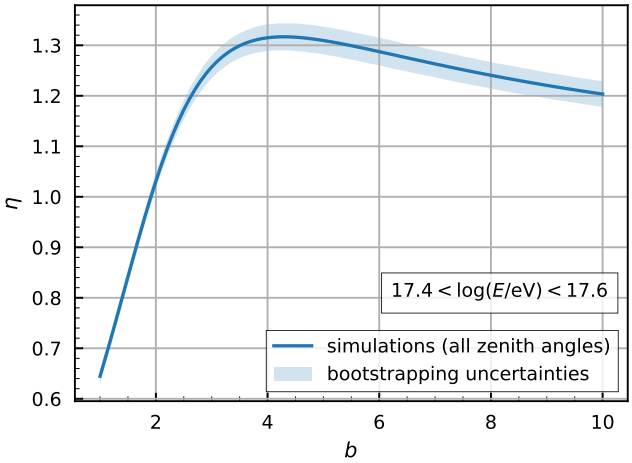

Fig. 8.5 Merit factor η and its uncertainties for $\log(S_b)$ as a function of b in the energy bin $17.4 < \log_{10}(E/\mathrm{eV}) < 17.6$. The uncertainties have been determined through a bootstrapping method. Corresponding plots for the other energy bins can be found in the appendix in Fig. D.4.

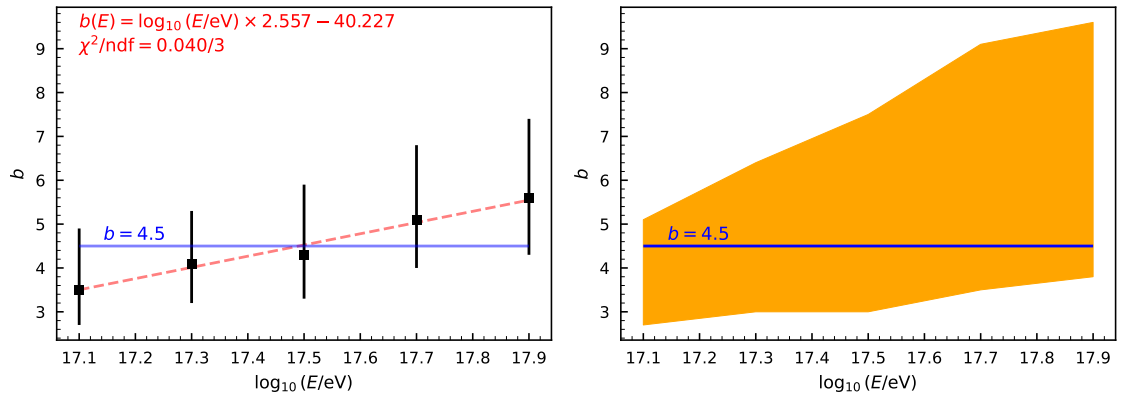

Fig. 8.6 Left: Optimal value b_{opt} as a function of primary energy, when using plain $\log(S_b)$, together with a linear fit (red) and the constant values as proposed in the text (blue). The uncertainties are estimated from the width of the bootstrapping band at the peak in Fig. 8.2 and corresponding scans. Right: Band that represents a 5% variation in η, together with the constant value.

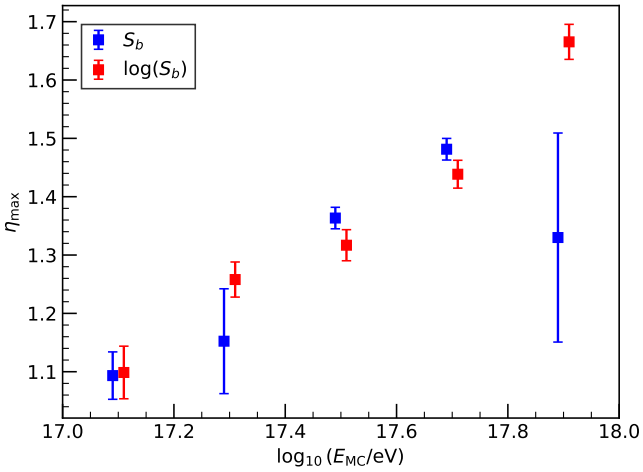

Fig. 8.7 Maximum attainable value of the merit factor when using the optimal value b_{opt} for each energy. Similar discriminatory power can be achieved with S_b and $\log(S_b)$. The energy values are slightly shifted near the center of the bins for better visibility.

direction of higher b values, an overshooting of the optimal value is not as critical. A fixed value can still be found with a merit factor that deviates only slightly from the optimal value throughout the entire energy range. In the following section both $b = 4.0$ and 4.5 will be examined in an application-oriented manner, as they exhibit promising results in this study. The former value is the one currently in use in various multivariate analyses, while the latter provides a better fit in this energy range and for this array configuration as manifest in the results presented here (Fig. 8.6).

The findings are again in accordance with the current usage in the original analysis [2], while providing a documented rationale for $b = 4$. However, the optimal value of the exponent for $\log(S_b)$ is significantly less stable with varying energy. A linear extrapolation of the findings beyond 10^{18} eV would not be compatible with an optimal value of $b = 4$, that has been used there [129]. Assuming that a similar optimization has been performed for the baseline array, this shows that there has to be a certain degree of dependency on array spacing, in contrast to the findings of [109] which, among other results, found no such dependency for plain S_b. Moreover, this pronounced energy dependence may motivate a parametrization of b and subsequently $\log(S_b)$.

What remains is a direct comparison of the discriminatory power of these two observables. To visualize their relative stand-alone performance, Fig. 8.7 illustrates the maximum attainable merit factor η_{max} for S_b and $\log(S_b)$ in the five energy bins. Both variants exhibit a similar increase with raising energy, indicating the relative decrease of shower-to-shower fluctuations and the subsequent improvement in the classification. In addition, it is not clear which of the two is superior in terms of separation power, as both show a similar performance in the energy range considered here and any differences between the two respective levels lie within the range of the statistical uncertainties determined with the bootstrapping method.

8.1.2. Application in Multivariate Analysis

The conventional usage of S_b sees the observable as part of an input to a multivariate analysis (MVA), particularly combinations with $X_{\max}$ resulted in superior separation power between cosmic-ray protons and photons. The analysis in [2] used a BDT to combine five input variables into a single discriminant, upon which a photon candidate cut is applied in the end. The input contained the energy and zenith angle estimate to account for the dependencies of S_b (and $X_{\max}$), which mediated the need for a binning of training and testing events.

The original BDT training and testing procedure has been redone for various observable configurations. Apart from the input variables, no other parameters where changed. Each observable configuration is tested with the four b values 2.8, 3.0, 4.0 and 4.5, which are traditionally used or showed promising performance in the two stand-alone cases of the previous section. The simulations samples are subdivided randomly in training and test samples with a ratio of 2:1. After the training had been conducted, the performance of the BDT was tested in six energy bins between $10^{16.5}$ and $10^{19.5}$ eV. Again the merit factor was chosen as the measure of separation power between photon- and proton-initiated air-shower events. The results are displayed in Fig. 8.8. The original observable combination from [2], i.e., $X_{\max}$, $\log(S_b)$, N_{stations}, θ and E_{photon} together with $b = 4$, is also included.

From the results of the original analysis, it is known that the uncertainties of the tested merit factor are of order $\sigma_\eta \approx 0.2$ for each energy bin, when testing and training sample are rebuilt and therefore dominant over any statistical fluctuations in the testing process. Bootstrapping estimations are not applied here, because of the rapid increase in running time, due to the time needed for the training of the BDT at each step.

Several conclusions can be made from the results above: To start, combinations containing $\log(S_b)$ in general perform better than those with plain S_b, even though the maximum merit factor in the stand-alone case is of similar value. A possible reason for this is the non-linearity of the logarithm that accounts very well for the strong energy dependence and sort of stretches the values of lower magnitude. As a result, the BDT has a higher resolution for binary decisions in this region. Secondly, as expected from the results of the stand-alone study, setups with higher values of b generally perform better in the higher energy bins, while setups with lower values of b are better situated in the lower energy ranges. This is especially apparent for the use of $\log(S_b)$ instead of S_b and lower numbers of input observables, while, for more complete combinations, the effect decreases within the systematical uncertainties associated with the division of the simulated events in training and test sample.

The background rejection at 50% signal efficiency was also used as a separate metric of separation power to evaluate the performance of the trained BDT. The same effects and results in the relative performance of the observable combinations were observed as when the merit factor was used, which further reinforces the findings. The corresponding plots can be found in Fig. D.5 of the appendix.

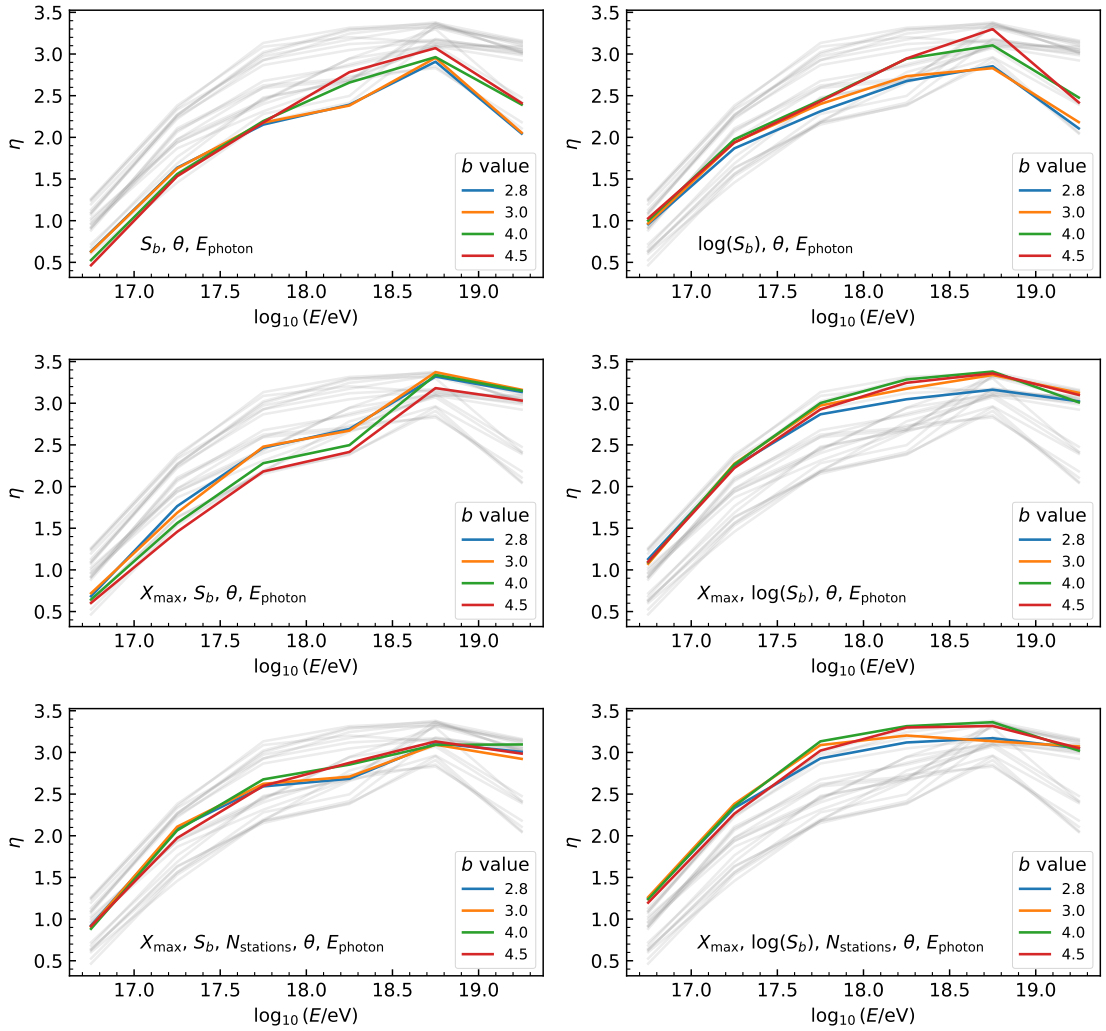

Fig. 8.8 Merit factor of BDT discriminant in various observable configurations measured in energy bins of width $\log_{10}(E/\text{eV}) = 0.5$. The left column depicts combinations containing S_b, while in the right column instead of the former $\log(S_b)$ is used. Results of the other combinations are indicted by gray lines.

8.1.3. Conclusions

In this study, a numerical approach was employed to analyze the observable S_b both in its standard form and logarithmic variant, to determine their relative performance, and provide recommendations for the optimal value of the S_b exponent, within the energy range 10^{17} eV to 10^{18} eV, which is covered by the low-energy extensions of the Pierre Auger Observatory. For this, a numerical approach was chosen, which was based on the similar optimization published in [109]. Notably, essential findings of it and preceding publications were reproduced, and extended within this lower energy range.

When using plain S_b, the optimal value for photon-proton separation was again located near $b = 3$ and found relatively stable with increasing energy, the best constant fit for the energy range 10^{17} eV to 10^{18} eV being approximately $b = 2.8$. Conversely, when using the logarithmic version $\log(S_b)$, the ideal value shifts slightly higher to around $b = 4$, but shows a considerable energy dependence now. With the same precision as above, $b = 4.5$ yields the best constant fit in the energy range considered here. Nonetheless, deviations from the optimal value are less severe compared to the linear case, as the range in which the merit factor deviates by no more than 5% from its maximum value is considerably wide. The maximum attainable separation performance is of similar level for both versions of the observable.

Promising values were tested in the application of a multivariate use case. The logarithmic variant was found to perform better for this purpose. Based on the findings of this study, for a potential update of the study in [2], it seems reasonable to keep the observable configuration already used ($X_{\max}, \log(S_b), N_{\text{stations}}, \theta, E_\gamma$), with the S_b exponent chosen as $b = 4$. Even though the performance for other b values reaches a comparable level within the uncertainties, but for consistency and comparability with the studies above 10^{18} eV, a change does not seem to be sufficiently justified for now. The potential for some follow-up consideration that may alter this recommendation is detailed in the subsequent section.

For accessibility of these results, an earlier version of here detailed optimization study has been published to the Collaboration as an internal note [143] in June of 2023.

8.1.4. Additional Considerations for Future Studies

Some aspects of the study would profit from a more thorough treatment, while other questions have been left completely unanswered here, which both may provide motivation for a follow-up investigation. On the basis of this, the following prospects should shape a picture of the considerations made, that did not find their way into the study.

The optimization procedure of [108] and [109] found no significant dependence on the hadronic interaction model and zenith angle of the simulated air showers. Without the time- and computing-intensive production of new simulation samples, the first aspect cannot be investigated, at least for this array configuration. Furthermore, the dependence on zenith angle was not addressed, due to the low number of events in the individual energy bins, which makes further subdivision in zenith angle ranges unappealing, when the level

of statistics should be retained. Besides, the logarithmic mapping is not expected to alter these fundamental aspects of S_b considerably. When, however, the resource are available, it may be interesting to look into both aspects, considering that for $\log(S_b)$ no such studies have been performed yet.

Due to the realistic nature of the MC simulations, smearing and efficiency effects are included naturally. For the sake of completeness, however, these aspects could be considered separately, providing the additional insight whether S_b or $\log(S_b)$ is more stable regarding these sources of uncertainty.

In [108], it was suggested that the usage of a modified merit-factor definition would be more applicable for asymmetric distributions, as present for S_b, considering its strong energy dependence and the considerable wide energy intervals of the study presented here. Instead of the means, the medians of the distributions are chosen and the width of the distributions is defined on the basis of the quantiles at 16% and 84%:

$$\mathrm{Var}[S_b] \equiv \left(\frac{q_{84} - q_{16}}{2} \right)^2 \tag{8.3}$$

While the optimization in [108] avoided its usage, due to its semi-analytic approach, this alternative definition could be considered here. Originally, however, the asymmetrical nature of the observable distribution was to be taken into account by using the background rejection in addition to the merit factor in order to evaluate the separating power in the individual energy bins. Due to the discrete counting nature of this metric, the wide separation of the observables and the low number of events in each energy bin, however, lead to no background contamination above a certain level, upon which no further optimization was possible. Hence, the merit factor was employed as the sole metric in the presented study.

The last aspect listed here transcends the pure character of an optimization and should rather be considered in the context of a possible application in an updated photon search. The study presented here tried to provide a recommendation for an optimal value of the S_b exponent to be used in an MVA. For this, the discriminatory potential was maximized in each energy bin and the recommended value was obtained with a constant fit, approximately averaging the individual results, while leaving behind the previous restriction to integer values. Due to the inherent energy dependence, slightly lower values perform better near 10^{17} eV, while a larger value would be most appropriate near 10^{18} eV. This, however, does not take the expected number of photons in this energy range into account. For this, values of b could be weighted according to the expected photon spectrum $E^{-\Gamma}$, with $\Gamma = 2$, which would decrease the optimal value slightly. Furthermore, when also considering the detector acceptance and efficiency, folding the linear model with the integrated efficiency weighted exposure, as calculated for the original analysis (or a potential update), would provide an S_b observable optimized near the energy region of highest sensitive for such a study.

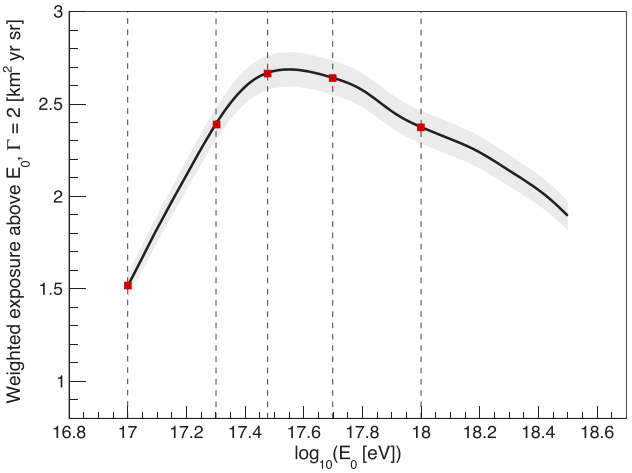

Fig. 8.9 Integrated efficiency-weighted hybrid exposure for photons taken from [2] and modified. An additional dashed line indicates the new lowest energy threshold to be introduced, while the red markers were added to emphasize the different levels of integrated exposure, that determine the subsequent integrated upper limits for the photon search. Due to the decreased detector acceptance for lower energies, the exposure for the 1×10^{17} eV threshold is far from the levels of the current energy limits.

8.2. Additional Initial Cut Regarding $X_{\mathbf{max}}$ Reconstruction

This section describes the studies on a novel initial quality cut. While the precise definition or specific value of this threshold remains open to further in-depth exploration, the outlined procedure can be acknowledged as an initial framework. Key considerations, such as efficiency and inherent bias, crucial for the introduction of any new quality cut, have been taken into account.

As a basis, the events contained in the original photon- and proton-simulation samples, together with the burn sample of the original analysis were utilized. In this subsection, for the sake of consistency with [2], the photon energy E_γ was employed for any energy assignments. Given that the overarching goal of the analysis is to determine upper limits on the total photon flux, using this energy scale ensures the appropriate evaluation of events.

8.2.1. Motivation

A key objective in updating the low-energy photon search was to extend the analysis towards a lower energy threshold of 1×10^{17} eV. This lower limit in the previous analysis was based on the results of the integrated efficiency-weighted exposure calculated for the original study, as illustrated in Fig. 8.9. The figure highlights a notable performance difference between this additional energy threshold and the four previously established limits. Nonetheless, there is no compelling reason to overlook this energy range, especially considering that other analyses based on low-energy extension measurements have also incorporated this limit. In addition, one of the overarching goals should be to narrow the energy gap to the recent photon measurements in the PeV range [103] leveraging the relatively favorable exposure compared to other experiments and potentially identifying photon candidates in this range at some point.

However, individual examination of the events in the burn sample reveals the energy range 1×10^{17} eV to 2×10^{17} eV, which would potentially be added, to encompass a

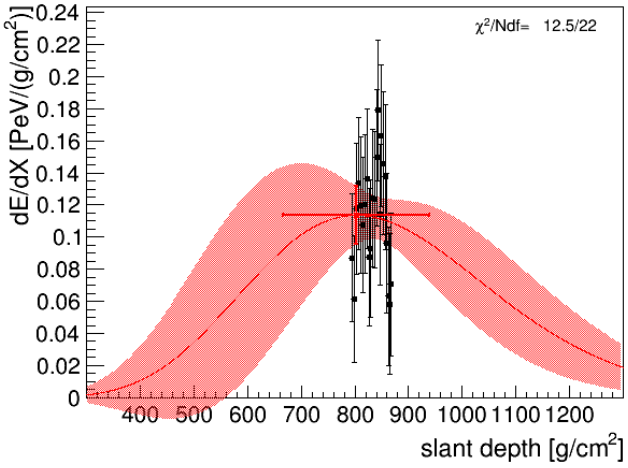

Fig. 8.10 Example of the longitudinal fit for such a low-quality event still present in the burn sample. Measurement data of the HeCo telescopes is shown in black. The red marker corresponds to the estimate for X_{max} and its uncertainties. The red band displays the uncertainties on the model obtained in the fitting procedure.

notable quantity of events with suboptimal longitudinal reconstruction quality, based on the fluorescence-light measurements of the HeCo telescopes. An example for the energy-deposit fit of the Gaisser-Hillas function to such an event is illustrated in Fig. 8.10. The depicted event is still present in the burn sample of the original analysis, however located slightly below the potential energy cutoff at 1×10^{17} eV.

A unifying feature within this category of events lies in the limited measurement of longitudinal air shower development by the telescopes, due to the event geometry and detector acceptances. Consequently, this leads to a suboptimal fit and significant uncertainties in determining the position of the shower maximum X_{max}. Given the pivotal role of this value in the classification of the primary particle, this study aims to devise an effective filter for such events, which have been subjectively deemed questionable in their reconstruction. The challenge is to achieve this without unduly diminishing statistics in the low-energy range, which is already affected by reduced detector acceptances, or compromising the overall performance to an unacceptable extent.

Another factor to take into account pertains to the training phase of the BDT. This is because identical selection criteria are applied to both simulation samples utilized as input during the training process. Consequently, it can be inferred that events meeting these criteria are likely present in the artificially generated data as well, albeit potentially at a lower frequency due to its controlled nature. Expanding the scope to lower energies would introduce more of these events into the BDT training stage, potentially having a detrimental impact on the separation performance across all energy levels. This is because only a single model is trained to encompass the entire energy range.

8.2.2. Current Cuts

The current selection of quality ensuring cuts applied to the measurement and simulation samples is specified to some extent in [2]. Further details can be found in the corresponding GAP notes [98, 144]. The cuts contain a subset that addresses the quality of the longitudinal energy-deposit fit. The following list provides an excerpt from this subset,

encompassing individual cuts that, according to the current perception, already address the aforementioned issue to some degree:

`photonEnergyError 0.2` The relative error of the photon energy is constrained below 20%. This is essential for the analysis to ensure correct binning and subsequent training and deployment of the BDT model. Furthermore, it also enforces a certain quality of fit for the Gaisser-Hillas function indirectly.

`XmaxErrorLessThenXmax` This cut limits the relative error of the X_{max} below 1. It is designed similar to the candidates that will be discussed in the following. Due to the usage of the X_{max} value as one of the key discriminants input in the BDT, the accuracy of the measured values is of particular importance.

`profileChi2Sigma 5. -1.42` Here, a certain quality of fit is directly enforced using the metric of χ^2. Deviations of the normal-transformed Gaisser-Hillas χ^2 are only allowed within the specified sigma-multitude threshold, on the assumption of five fit parameters.

`maxDepthHole 30` The cut requires that any gaps in the recorded longitudinal evolution, e.g. due to air showers crossing the field-of-view of several telescopes that are not exactly aligned, or due to individual detector efficiencies, amount to less than 30% of the total observed tracklength.

The above definitions can serve as a guideline for devising a prospective cut. It should be noted that an adjustment of the listed levels of the above cuts may also present an opportunity to efficiently circumvent the problematic event class. Nevertheless, this study exclusively explores the feasibility of introducing a novel definition. Care was therefore taken to ensure that the new cut synergistically supplements the existing ones, while avoiding significant overlap with their operational domain.

8.2.3. Definition and Specificity

For the definition of a new cut, the aspect of specificity was a key concept. It states that the influence of a new initial cut compromises the existing performance of the original analysis to the smallest possible extent and its state is preserved as possible. The operating range should therefore be limited as far as possible to the newly introduced energy range between 1×10^{17} eV and 2×10^{17} eV. It would be desirable if this property already inherently emerges from the fundamental definition and such a restriction to the low-energy range does not have to be artificially introduced by a second layer.

For a potential cut, different variants of tracklength, ΔX_{max} and combinations of them were analyzed in this study. A simple cut to the value of the observed tracklength was quickly discarded, as its effect was almost uniform across all energy ranges. The next quantity that was considered was the uncertainty of the X_{max} estimate, where both the relative and the absolute value were analyzed. To convert the threshold value of a single variable

into a corresponding value for another variable, one can utilize a graphical representation of the distribution of both variables in a scatter plot. The strong intrinsic correlation between absolute and relative error results in the formation of a distinct curve, serving as a mapping function between the two quantities. Subsequently, comparable levels of event rejection criteria were established for both absolute and relative ΔX_{max}. Upon comparing their performance, it was observed that the relative variant significantly compromised the discriminatory potential of the BDT. Consequently, the relative variant was excluded from further deliberations.

According to statistical considerations (relative) errors should follow distinct distributions, which could be employed to look for anomalies in data. This, however, was not pursued further in this context, as the burn sample has a small sample size, which would elicit considerable uncertainties in the modeling.

Overall, a basic cut based on the X_{max} uncertainty was deemed not satisfactory regarding the criterion of specificity.

A final construction combined these two aforementioned approaches into a single metric, based on their ratio. The initial cut proposed here follows the definition

$$\frac{\Delta X_{\mathrm{max}}}{\mathrm{TL}} \overset{!}{<} 0.5, \tag{8.4}$$

where TL denotes the total tracklength observed along the longitudinal shower axis by the HeCo telescopes. This design profits from the strengths of the two component quantities for assessing the reconstruction quality of the event, without having the same shortcomings. The threshold value corresponds to the idea of the 1σ-interval of X_{max} not extending beyond the measurement range (cf. Fig. 8.10, where the event does not fulfill this criterion).

Fig. 8.11 illustrates the energy specificity associated with the introduced threshold. As discernible from the visualization, the ratio $\frac{\Delta X_{\mathrm{max}}}{\mathrm{TL}}$ exhibits pronounced enhancement particularly for energy levels below $10^{17.5}$ eV. Conversely, at higher energies, the simulated events are concentrated at lower magnitudes. In addition, air-shower simulations triggered by photons behave similarly to such induced by proton primaries, which limits the level of inherent bias introduced by this additional threshold. A more thorough examination of the bias is contained in the next subsections.

When the specified threshold is imposed on the events contained within the burn sample, five out of 175 events above 1×10^{17} eV are filtered out, while for the previous energy limit $E > 2 \times 10^{17}$ eV two out of 104 events are rejected. None of the rejected events would have been considered photon candidates by the original analysis. Tab. 8.2 list the first events in the burn sample sorted by their value of $\frac{\Delta X_{\mathrm{max}}}{\mathrm{TL}}$ in descending order, the first five of which are rejected by the proposed cut. High values in the cut parameter $\Delta X_{\mathrm{max}}/\mathrm{TL}$ seem to be mainly caused by low observed tracklengths. The values of the numerator, however, are not particularly noticeable in this respect. Such outliers in ΔX_{max} are probably already filtered out indirectly by the other profile cuts.

SDId	T_{GPS}	Date	E_γ [10^{17} eV]	θ [°]	X_{max} [g cm^{-2}]	S_b [VEM]	N_{stations}	TL [g cm^{-2}]	$\frac{\Delta X_{\mathrm{max}}}{\mathrm{TL}}$
21528120	1050038680	2013-04-15	(1.22 ± 0.13)	(28.411 ± 0.029)	(717 ± 149)	4.67	5	54.47	2.74
33484805	1119246700	2015-06-25	(2.04 ± 0.21)	(35.846 ± 0.027)	(588 ± 118)	2.32	3	95.88	1.24
21653456	1051749860	2013-05-05	(1.18 ± 0.13)	(19.589 ± 0.028)	(723 ± 88)	3.59	4	124.81	0.70
13430123	1008474300	2011-12-21	(1.79 ± 0.14)	(47.968 ± 0.010)	(893 ± 64)	3.28	5	117.52	0.54
15666309	1026440580	2012-07-16	(1.50 ± 0.11)	(22.292 ± 0.023)	(706 ± 76)	6.46	5	140.11	0.54
22579661	1059276720	2013-07-31	(1.65 ± 0.14)	(18.413 ± 0.011)	(670 ± 50)	9.16	5	111.79	0.44
33361366	1118618000	2015-06-17	(1.61 ± 0.16)	(14.069 ± 0.033)	(632 ± 75)	6.92	5	171.83	0.43
25355159	1074928080	2014-01-28	(1.05 ± 0.11)	(22.538 ± 0.019)	(674 ± 49)	7.47	4	128.49	0.38
35524334	1131597420	2015-11-15	(1.71 ± 0.14)	(28.831 ± 0.013)	(753 ± 61)	8.36	5	166.66	0.37
12067440	993610080	2011-07-02	(1.04 ± 0.10)	(44.086 ± 0.010)	(732 ± 61)	1.27	3	170.88	0.35

Tab. 8.2 Events in the burn sample above 1×10^{17} eV with the highest values of the potential cut parameter $\Delta X_{\mathrm{max}}/\mathrm{TL}$. Events exceeding the proposed threshold are marked red. For each event, together with identifying quantities (SDId, T_{GPS}, Date), several reconstructed observables are listed, in part with the corresponding uncertainty.

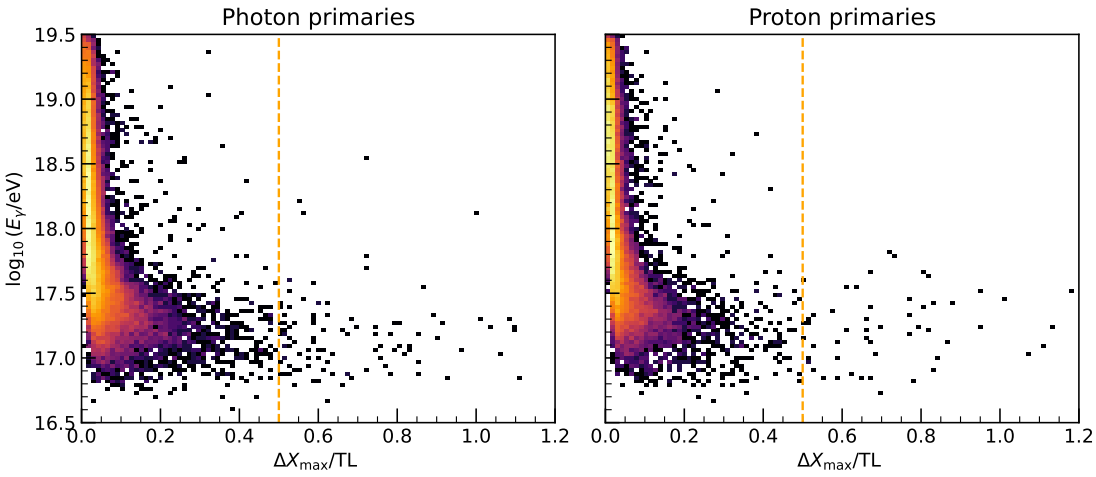

Fig. 8.11 Specificity of the proposed cut. The two scatter plots show the distribution of events in the two simulation classes in regard to their energy and $\Delta X_{\max}/\mathrm{TL}$. The colors correspond to the density of events in the 2d bins (pixels). The color grading follows a logarithmic scale in order to be able to characterize the parts with only a few events despite the high total number. Since the figure only serves illustrative purposes here, a uniform color scale and a corresponding key were omitted.

The characteristics of the cut can be recognized particularly well if the affected events are considered with reference to their overall measured population. For this purpose, the distribution of several physical key quantities of the rejected events were compared with the overall sample for both simulation categories in addition to the real events of the burn sample; Fig. 8.12 depicts a compilation of the results. The specificity towards lower energies is again evident for all three samples examined. However, no dependence on the zenith angle θ can be determined, as the affected events are distributed similarly to the populations as a whole. Regarding the atmospheric depth of the shower maximum $X_{\max}$ a slight enhancement at the lower values seems predominant, which comes as no surprise in view of the energy dependency. Since the cut is constructed from the ratio of $\Delta X_{\max}$ and observed tracklength, extreme values of both quantities are efficiently filtered out, whereby this property, as already noted for the preceding table, is significantly sharper for the tracklength.

It should also be noted that the properties listed here are pronounced very similarly in the two simulation classes as well as for the real data from the burn sample. This suggests that the cut actually filters out a certain type of event that occurs regardless of the nature of the primary particle.

8.2.4. Estimating the Efficiency for a Full-Sized Sample

The decision process for the numerical level of the cut is largely based on the expected efficiency on a sufficiently large data set. However, due to the statistical limitations of the burn sample, the realistic rejection rate for measured events can only be predicted to

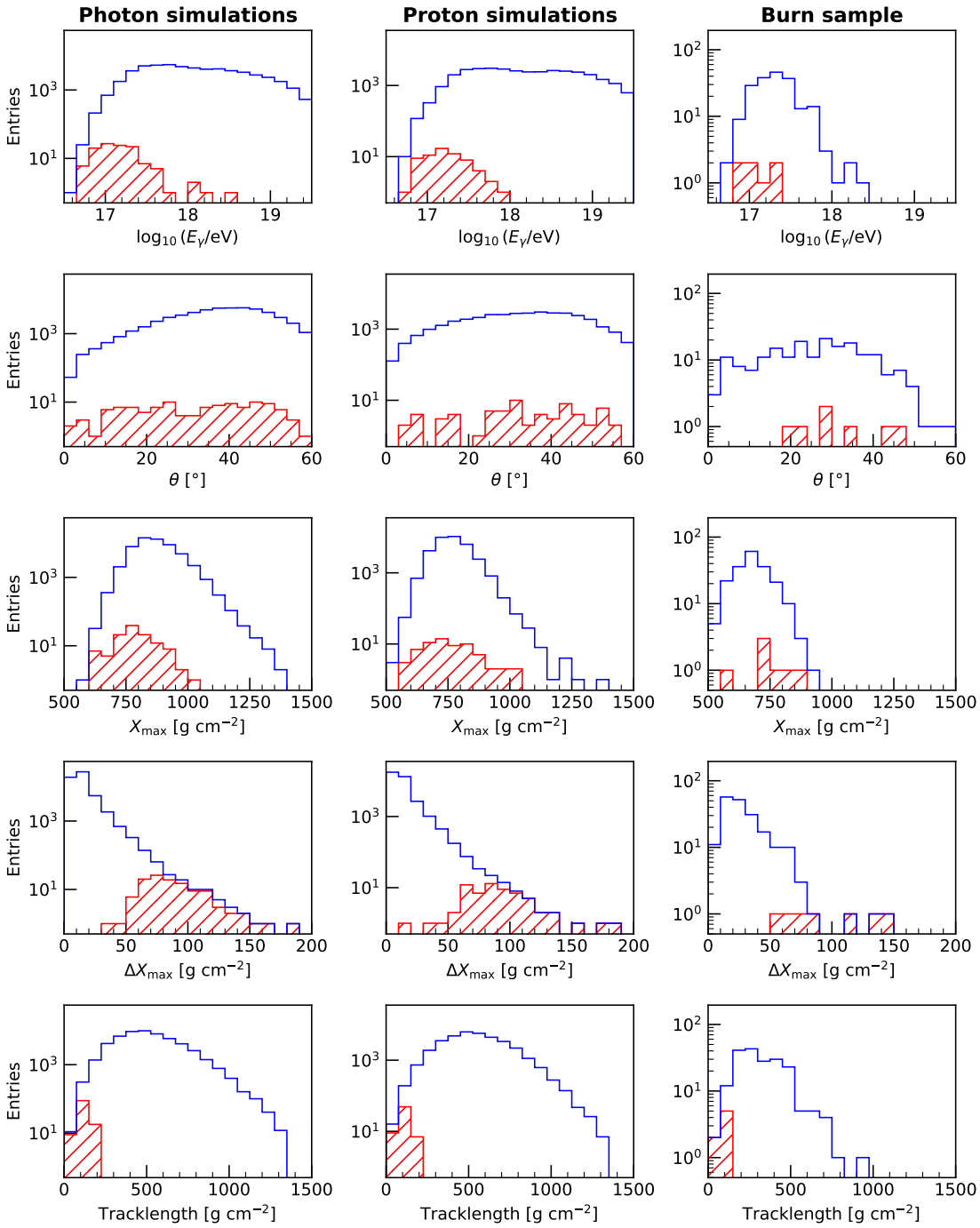

Fig. 8.12 Distributions of the events affected by the cut (red hatched) within the total distributions (blue). Columns correspond to the different samples (photon/proton simulations and burn sample), while rows correspond to various physical parameters. In contrast to the other plots presented in this section, there have not been any additional constrains on the energy of events in the samples. Furthermore, note the semi-logarithmic axis scaling.

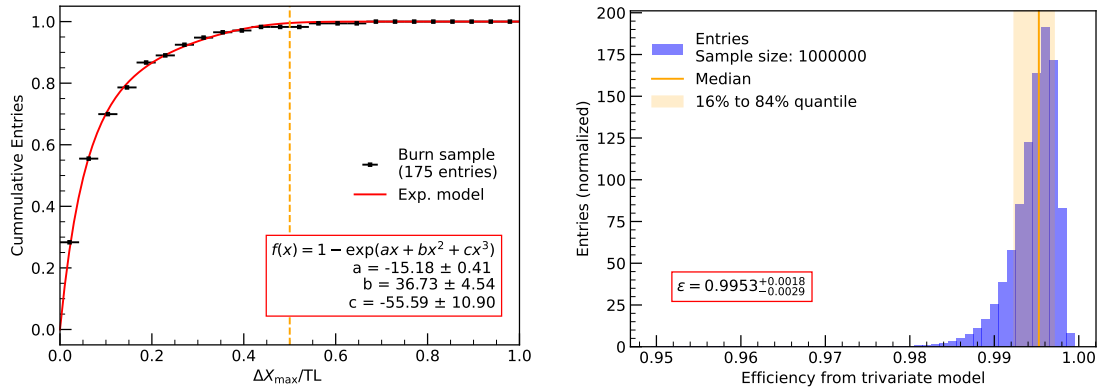

Fig. 8.13 Efficiency of the proposed cut in the burn sample above 1×10^{17} eV. (a) Efficiency of the cut (percentage of the events kept) in several bins of cut values. A trivariate exponential model (red) is fit to the cumulative data distribution to allow for a precise estimate of the performance on a larger sample. The parameter estimates are provided in the red box. (b) Estimate on the efficiency and its uncertainty for the proposed cut level $\Delta X_{\mathrm{max}}/\mathrm{TL} = 0.5$ on the trivariate model obtained from one million MC samples distributed accordingly to the fit in (a).

a limited accuracy. To obtain a realistic estimation of the efficiency, when deployed on a full-scale data set, the efficiency was modeled in the burn sample as a function of the cut threshold. For this, the efficiency, i.e. the ratio of events below the threshold, was calculated at 25 equidistant points between 0.0 and 1.0. Then a generic exponential model $f(x)$, where x denotes the cut quantity, that satisfies $f(0)$ and $f \to 1$ for $x \to \infty$, was fit to the data. In order to achieve sufficient accuracy of the model, but at the same time adequate generalization of the results, a total of three parameters were chosen to encode the regular behavior of the integrated event distribution. The modeling is depicted in Fig. 8.13a, where, in addition, the resulting values for the parameters with their individual uncertainties are given.

Finally, using a numerical MC approach to estimate the error propagation, the model $f(x)$ was evaluated one million times at $x = 0.5$ according to the parameters and their covariance matrix obtained in the fit. The resulting distributions of the efficiency estimates is shown in Fig. 8.13b. Due to its asymmetric nature, the final estimate was obtained as the median of this distribution, with the uncertainties given by the 16%, respectively 84% quantile. This procedure estimates the efficiency level, i.e. the share of events not affected by the cut, for the proposed level of $\Delta X_{\mathrm{max}}/\mathrm{TL} = 0.5$ above 1×10^{17} eV at

$$\varepsilon = 0.9953^{+0.0018}_{-0.0029}. \tag{8.5}$$

The here detailed method, however, can be employed to quickly obtain a sufficiently accurate efficiency estimate for any threshold value between 0.0 and 1.0, and may thus be considered for the decision process regarding the final cut level, if it is assumed that the events in the burn sample used here are representative of their population.

8.2.5. Studying the Intrinsic Bias

The efficiency level of the proposed cut fundamentally differs for the two simulation classes from the burn sample. This is primarily due to the fact that the simulations feature a different energy distribution than the actual data (cf. Fig. 8.12), as affected events are concentrated at the lower energies. Nonetheless, a comparison between photon and proton simulations allows for an estimation of the inherent bias, i.e. the inclination for the selection criterion to disfavor photon events relative to the nuclear background. It is important to note that from an overall analysis perspective, multiple factors may contribute to such bias against photon events. However, as the focus here is on evaluating absolute photon levels with upper limits set for the photon flux, the photon rejection rate is already considered through the overall photon efficiency factor in Eq. (5.4).

Nevertheless, it is imperative to exercise caution to avoid excessive photon rejection, especially relative to the background, to prevent further deterioration of the signal-to-noise ratio, potentially compromising the classification of individual potential photon events. For this reason, at least a brief consideration of the bias should be undertaken for this cut.

As a metric of the bias, the difference between the efficiencies of the cut for photons and protons,

$$f_{\text{bias}} = \varepsilon_\gamma - \varepsilon_p, \tag{8.6}$$

was utilized. This metric was chosen for its simpler uncertainty propagation, especially compared to the more conventional $\varepsilon_\gamma/\varepsilon_p$. The values for the bias parameter were determined for 50 equidistant points between the cut thresholds 0.0 and 1.0, using the events contained in both simulation samples above 1×10^{17} eV. The values for the individual efficiencies are shown in Fig. 8.14 and are accompanied by the corresponding bias parameter in the lower subplot.

For each value of the cut threshold, the calculated bias parameter was tested against the hypothesis of no inherent bias ($f_{\text{bias}} = 0$). Details on the calculation of the corresponding uncertainty bands can be found in the Appendix in Sec. C.2. For the proposed threshold $\Delta X_{\text{max}}/\text{TL} = 0.5$, the bias parameter was determined as

$$f_{\text{bias}} = (-0.93 \pm 2.34) \times 10^{-4}, \tag{8.7}$$

which is not regarded as significant against the null hypothesis of no bias. The effect of an inherent bias is therefore negligible when considering a threshold of $\Delta X_{\text{max}}/\text{TL} = 0.5$. It should, however, be noted that for decreasing threshold value, in particular below $\Delta X_{\text{max}}/\text{TL} = 0.2$, the inherent bias increases rapidly by more than one order of magnitude, crossing the significance level of 2σ with an effect strength of approximately $f_{\text{bias}} = -0.002$, which is probably caused by the difference in mean atmospheric depth for different primary particles and subsequent deviations in observed tracklength, due to geometrical constraints. This potential bias should be considered when lowering the cut threshold further than the here proposed level.

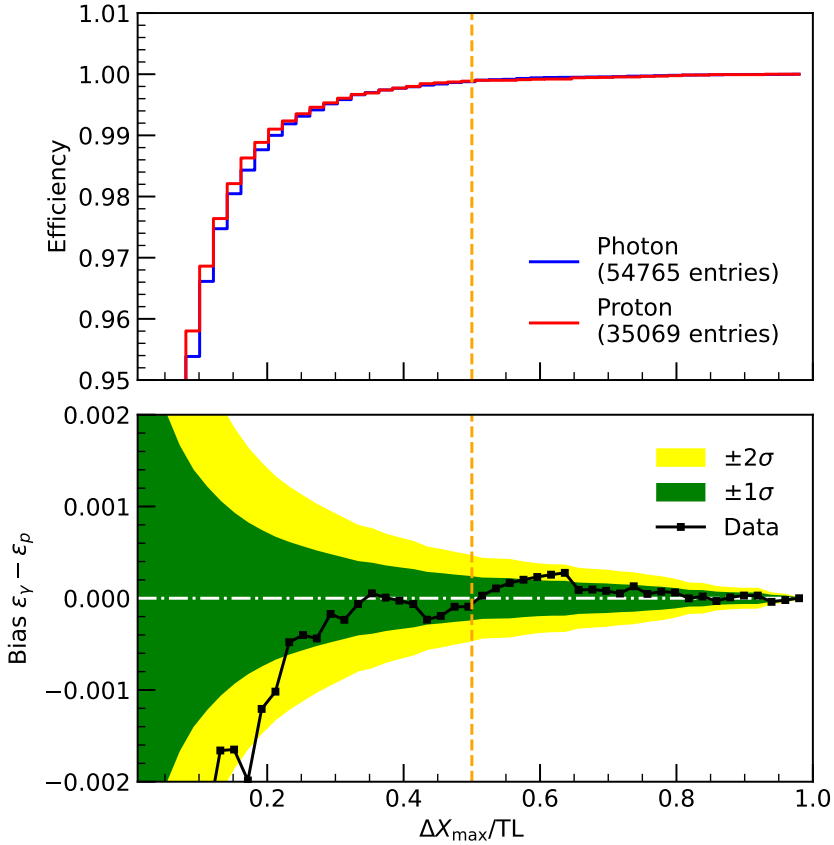

Fig. 8.14 Efficiency of the proposed cut in the two simulation samples as a function of the cut level $\Delta X_{\max}/\mathrm{TL}$ (top), from which the inherent bias $\varepsilon_\gamma - \varepsilon_p$ was calculated (bottom). The two colored bands represent different significance levels for the hypothesis of no inherent bias existing for a particular threshold. For $\Delta X_{\max}/\mathrm{TL} = 0.5$, the bias is not significant. Details on the value of the efficiencies for the simulations samples and the level of the calculated bias for the proposed threshold are given in the text.

8.2.6. Further Considerations

Cloud coverage during the time of observation can limit the measured tracklength of an air shower, which is an effect not included in the simulation environment. This was theorized to cause the considerable deviations of the efficiency observed between the simulation samples and the measured data (cf. Figs. 8.13a, 8.14). However, periods with attenuation effects due to clouds or in general unclear atmospheric conditions should be rejected by the initial selection cuts based on the data from the Lidar system [145] installed at the Coihueco FD site and infrared data from the GOES-12 satellite [146], and thus the effect should probably be attributed to the different energy distributions as stated earlier.

In addition to the isolated systematic studies detailed here, the proposed additional cut was tested within the original analysis context in application. Employing the merit factor as the metric, no significant reduction of BDT separation performance was determined. When looking at the background rejection based on the behavior of the proton sample, an initial increase, when implementing the cut, in wrongly identified proton simulation events (i.e., photon candidates) was recorded. Those events exhibit low S_b and N_{stations} values, which lead to the misidentification, while their attributed X_{max} values behave similar to the overall population. Furthermore, the background contamination events are uniformly distributed in the energy range $17 < \log_{10}(E\text{eV}) < 18$. While initially concerning, this increase was eventually deemed not statistically significant as well. It can therefore be said that the previously established basic condition that the performance of the analysis is not deteriorated by the additional cut, particularly at the higher energies, is fully met. However, if, connected with a decreased threshold value, higher rejection rates are accepted, this has to be evaluated again.

8.2.7. Summary

The study contained in this subsection explored the possibility and potential design of an additional selection cut regarding the quality of the longitudinal reconstruction, applied to both measured data and simulation samples. With the concept of specificity as a guideline, the quantity $\Delta X_{\text{max}}/\text{TL}$ was chosen as the cut parameter, together with a proposed upper threshold of 0.5. Individual consideration of the rejected events revealed the cut to effectively filter out the issue observed with low-energy events. The general distribution of the events contained in the burn sample regarding the cut parameter was modeled to allow for a quick determination of the efficiency as a function of the threshold value on a full-scale data set. In addition, the inherent bias of the cut was determined for the proposed cut level and deemed not significant. Finally, the application in the analysis context was tested, and, here too, no effects were found to restrict the potential introduction of this cut.

9. Application to Extended Data Sample

This section describes the application of the analysis to the extended data set. However, owing to the time constraints inherent in a master's thesis, only an initial preparatory step

was possible here in the form of the creation of an updated, comprehensive simulation set. Using this, and prospectively incorporating a new burn sample, further studies may then be conducted to finalize key aspects of the analysis adjustment. Following the principle of a blind analysis, to preclude a-posteriori adjustments to the data set, the actual Phase-1 data should only be accessed and analyzed in a conclusive step.

9.1. Production of New Simulation Samples

Similar to the setup of the original analysis, a set of simulations for photon-induced events and a smaller set of proton-induced events were created, with the latter providing a conservative representation of the nuclear background. Once again, prefabricated air-shower simulations from the *Napoli+Praha* library [133] were utilized for this step. They have been produced using CORSIKA [134] with EPOS-LHC [135, 136] employed as the high-energy hadronic interaction model. Assorted repositories of air shower simulations based on different releases of CORSIKA are available. A later subsection is devoted to the selection of the CORSIKA version. The selection of air showers that have been used here covers the energy range $16.5 < \log_{10}(E/\mathrm{eV}) < 19.5$, where the simulations are logarithmic-uniformly distributed in six energy bins of equal width.

The production aimed to match, at a minimum, the numerical quantity of simulations associated with the original analysis, allowing for the potential expansion of statistical data at a later stage. Consequently, 10 000 simulated photon-triggered air showers were generated for each of the six energy bins, while the nuclear background was addressed through 5000 proton air showers per bin. During the simulation production process, each shower underwent resampling to generate five distinct events. Notably, only the ground position and event time were shifted between these events. This approach resulted in a total of 300 000 simulation events with photon primaries and 150 000 with proton primaries. In order to avoid wasting computing capacity by disproportionately reducing the number of events due to the selection cuts of the geometry level, the simulation set was run with the shower core at ground level positioned within the infill array incorporating a 1.5 km margin. The event time was again drawn randomly within the now extended time interval and only accepted if the detector logs indicate a simultaneous uptime fraction of all HeCo telescopes above 95%. Both measures were implemented to ensure a close resemblance of the simulations to the actual data, aligning with the principles of *RealMC* simulations.

Other quantities inherently connected to the air shower, among others arrival direction, primary energy and height of first interaction, were kept the same. The individual showers are superposed on the virtual Observatory and the detection und reconstruction process is then simulated on local hardware. The virtual detection environment includes the baseline hybrid detector together with the low-energy extensions utilized in this analysis. Separate dedicated radio and muon detection instruments, as deployed during the AMIGA and AugerPrime upgrades, are occluded from the simulation framework for performance.

On the software side, $\overline{\text{Off}}\underline{\text{line}}$ version v3r99p3, ICRC '23 production release 1, [137] was used as the simulation framework. Apart from a few minor changes due to the new $\overline{\text{Off}}\underline{\text{line}}$

version, the module sequence and general configuration used was left analogous to that used in the original analysis and still largely resembles the `HdSimulationReconstruction` example provided within the Offline documentation. The simulations were performed in parallel on the local computation infrastructure of the experimental particle physics research group in Siegen, with a batch system ensuring the optimal and fair capacity utilization of the multiprocessor cluster. On average, provided that the resources were shared with other end users, about 200 simulation jobs could be run simultaneously, with some day-night modulation. With this setup, the simulation batch contained in a single energy bin of width 0.5 in $\log_{10}(E/\mathrm{eV})$ could be processed within less than one up to several days.

Initial issues in regard of the status database of the FD instruments and subsequent failure of simulation jobs lead to the additional exclusion of a single week of initially available event times in January of 2021, that lay in the time period added with this simulation and analysis update. However, the problems could not be reproduced outside this narrow time period, which is why a more in-depth investigation was waived.

Furthermore, simulation jobs occasionally failed due to exceeding the individually allocated time and resource limits. Since every restart of the simulation jobs leads to the new assignment of event time and core position, excessive overshooting of limits would lead to systematic underrepresention of processing intensive shower and detector configurations, and was thus aimed to be avoided with careful assignment of somewhat generous quotas, while at the same time respecting the needs of other end users and therefore not blocking excessive computational resources.

The following subsections are dedicated each to various aspects of the simulations, further elaborating several key measures in the production process.

9.1.1. Choice of Pre-Simulated Air Showers

The original simulation set was based on extensive air showers simulated with CORSIKA in version 7.6400. In the meantime, however, such ready-made simulations based on the more recent CORSIKA 7.7420 had become available within the *Napoli+Praha* library [133]. With the production of new simulations, the question arose on which version the updated set should be built.

To address this point, only isolated energy bins were initially downloaded on-site and run through the simulation process, whereupon some fundamental quantities inherently connected to the air-shower simulation set were checked. Since prospective photon upper limits are going to be calculated from the efficiency-weighted exposure, which in turn fundamentally depends on an exact determination of the efficiency based on a realistic distribution of simulated photon events, it is important to ensure that some essential assumptions for the photon flux are adhered to in the air-shower and subsequent complete simulation set. The accurate distribution of photon events also benefits the BDT in its training and validation step.

One of these properties is the assumed power-law energy spectrum of the photon flux, which was ensured with an additional reweighting step (cf. Sec. 9.1.4), which provides individual event weights. Another fundamental property of the photon flux is its presumed isotropy – at least there is no known direction that is currently expected to be unique in this energy domain, especially considering the diminishing statistics of the, at the best, isolated photon observations. In order not to disfavor any arrival direction for potential photon-induced air showers in the statistical considerations, especially in view of the proper motion of the Earth and thus of the Observatory, which would blur most if not all dependencies on galactic coordinates anyway, it should be ensured that the distribution of simulated photon arrival directions is also isotropic.

In light of the significance of the aforementioned variables, care was taken to examine the initial simulation bins in both aspects, whereby events of the same primary particle type and identical energy subrange were compared. The value of $\sin^2(\theta_{\mathrm{MC}})$ was utilized as a proxy for the investigation of the isotropy, which, if satisfied, would yield a uniform distribution in this parameter. The ensuing distributions are compared in Fig. 9.1. In order to enhance deviations, the cumulative distribution of events was used here. As apparent from this, the distributions closely approximate a uniform distribution. However, notable deviations from the theoretical distribution are evident, particularly for the sample based on CORSIKA 7.6400 within the midrange of the observed parameter spectrum. Nonetheless, due to the finite statistical nature of this study, certain deviations are expected.

Furthermore, owing to the different sample sizes, the deviants cannot be directly compared. To ascertain the significance of the observed deviations, the one-sample *Kolmogorov-Smirnov test*, as implemented in SciPy [147], version 1.11.2, was utilized. This test is designed to assess the hypothesis that a sample conforms to a specified probability density function. The test statistic is represented by the maximum deviation of the normalized cumulative distribution function (CDF) from the specified model within the prescribed parameter range. The lower subplot illustrates this comparison by presenting the CDFs subtracted by their corresponding theorized uniform distributions. Based on this maximum distance and considering the sample size at hand, a p-value is calculated for each sample. According to these p-values, the deviations of the distribution in the sample based on the showers generated with CORSIKA 7.6400 are much more pronounced.

A parallel analysis was performed for the distribution of the variable $\log_{10}(E_{\mathrm{MC}}/\mathrm{eV})$, with the corresponding results shown in Fig. D.6 in the appendix. Here again, the more recent air shower simulations based on CORSIKA 7.7420 adhere closer to the desired distribution. However, the allocation of primary energies is not an essential criterion for the choice between the different pre-simulated air-shower samples as the reweighting process already takes care of the correct distribution. This fact, though, is a severe hindrance to any further individual weighting of events and has the effect that the isotropy issue cannot be solved in this way. Due to this restriction, the pre-simulated air-shower sample based on the more recent CORSIKA version (7.7420) was chosen for the creation of the new simulations sets, as the deviations from the desired distributions do not appear as severe here, and any further production of detector simulations based on the older version was ceased.

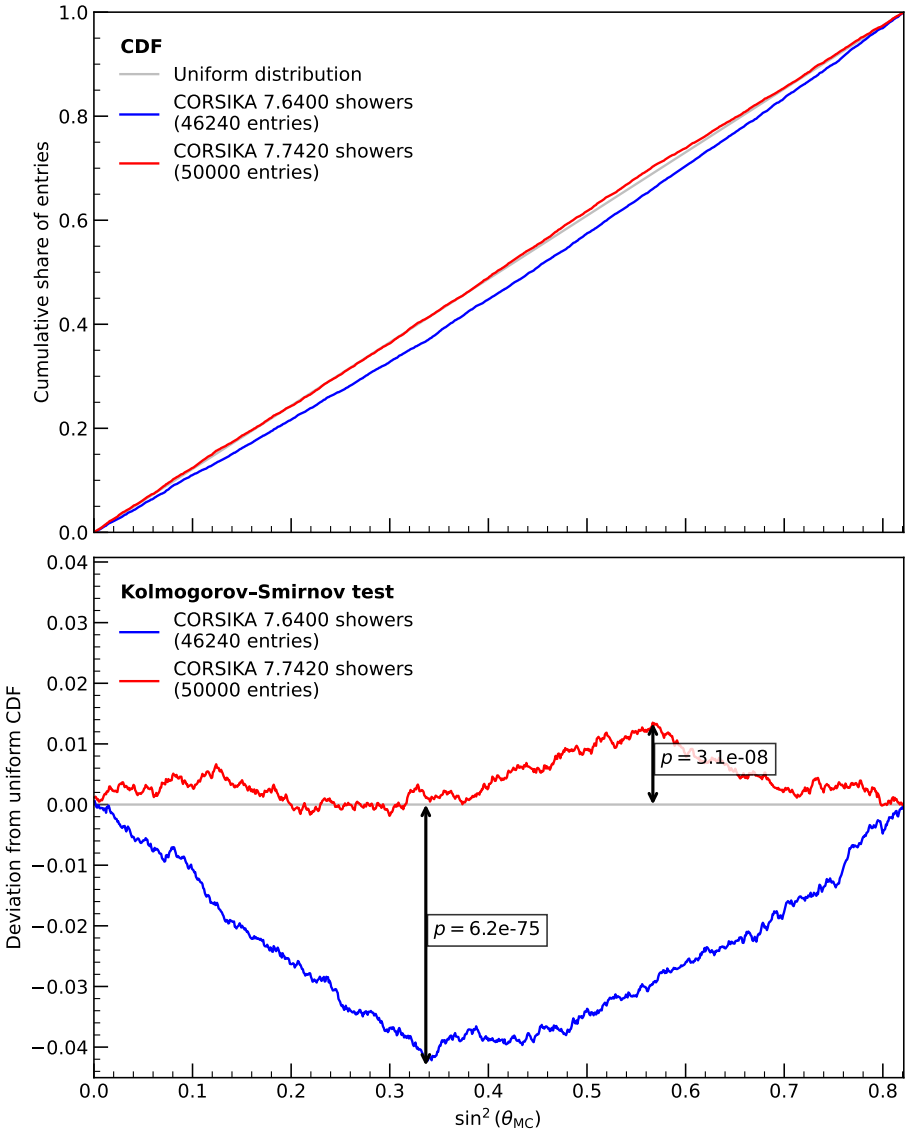

Fig. 9.1 Distribution analysis of $\sin^2(\theta)$ for the presimulated photon air showers between $10^{17.5}$ eV and $10^{18.0}$ eV. On the premise of isotropic arrival directions, the probability density function would equal a uniform distribution (up to the enforced cutoff). The cumulative distributions (CDFs) are compared within the upper subplot. The dataset containing the simulations based on CORSIKA version 7.6400 shows a considerable deviation from this theorized distribution in the center of the considered parameter region. The bottom subplot employs the Kolomogorov-Smirnov test to quantitatively evaluate these deviations. A p-value is calculated under the null hypothesis that the dataset conforms to the theoretical distribution, considering the maximal deviation within the specified range.

9.1.2. Final Status of the Simulation Production

Throughout the extensive simulation production process, frequent checks were conducted to proactively identify potential issues and ensure the growing data set adhered to the predetermined high-quality standards. The fundamental strategy involved continuous monitoring at the file system level. For instance, the number and sizes of output files in ROOT format were continuously tracked to promptly detect any systematic failures in simulation jobs. Additionally, real-time runtimes were extracted from log files using a dedicated script, facilitating the early detection of deviations from expected levels. This approach also provided an estimation of the overall remaining runtime for an entire simulation batch within a specific energy bin. Notably, during the development of detector simulations, a unique bug associated with the provided shower files, leading to the failure of affected simulation jobs, was uncovered using this method. The issue was promptly reported to the relevant individuals affiliated with the Napoli+Praha library. Subsequently, the problem was resolved through targeted re-simulation of the affected air showers, followed by local simulation of detection and reconstruction.

As of writing this thesis, the planned simulation sets have recently been completed across the entire energy range $16.5 < \log_{10}(E/\mathrm{eV}) < 19.5$ with the intended numerical extent. In Tab. 9.1 some statistics are listed related to the simulation batches for each primary particle and energy bin. For both the photon and proton case, file size and runtime scale like E^{α} with $\alpha \approx 2.3$. A power law is to be expected, as the total number of particles for an extended air shower grows proportional with the primary energy, and this is a major factor affecting run time and file size. For photons this increase is a bit steeper. This is partly due to the different nature of the underlying air showers, but also to slightly different detector acceptances at a fixed energy.

Photon simulations	$\log_{10}(E/\mathrm{eV})$					
Properties	16.5–17.0	17.0–17.5	17.5–18.0	18.0–18.5	18.5–19.0	19.0–19.5
status	Completed	Completed	Completed	Completed	Completed	Completed
total size of dir. [GB]	60.91	98.72	175.06	313.22	577.00	1017.05
total number of jobs run	10 000	10 000	10 000	10 000	10 000	10 000
avg. runtime per job [h:mm:ss]	0:17:50	0:24:46	0:39:12	1:01:30	1:43:24	2:44:01
total number of events simulated	50 000	50 000	50 000	50 000	50 000	50 000
avg. size of output file [MB]	0.97	1.72	3.27	6.09	11.49	20.49

Proton simulations	$\log_{10}(E/\mathrm{eV})$					
Properties	16.5–17.0	17.0–17.5	17.5–18.0	18.0–18.5	18.5–19.0	19.0–19.5
status	Completed	Completed	Completed	Completed	Completed	Completed
total size of dir. [GB]	28.37	45.20	79.14	138.81	267.02	491.20
total number of jobs run	5000	5000	5000	5000	5000	5000
avg. runtime per job [h:mm:ss]	0:19:55	0:29:34	0:46:55	1:15:44	2:16:18	3:40:47
total number of events simulated	25 000	25 000	25 000	25 000	25 000	25 000
avg. size of output file [MB]	0.88	1.54	2.92	5.36	10.60	19.77

Tab. 9.1 Final status of the production process of the updated simulation set together with some properties associated to the simulations within an energy bin.

9.1.3. Reconstruction Control

This subsection describes the checks carried out after finishing the production process. Once completed parts of the simulation sets were available, verification was performed on two levels: Firstly, isolated, randomly selected events were manually viewed in the EventBrowser included in the Offline framework and examined for inconsistencies and a coherent reconstruction. No abnormalities were detected during this process.

The final level of simulation checks consisted of the systematic investigation of the reconstruction quality in the simulation framework. The conducted procedure will be demonstrated in the following on the basis of the photon simulation sample:

For this, the three key quantities primary (photon) energy estimate E_γ, zenith angle θ, and depth of shower maximum $X_{\max}$ were compared in their reconstructed form to the "true value" MC input. Events where the hybrid reconstruction failed, for example due to geometrical reasons, were excluded from this process. In contrast, reconstructed quantities based on the HeCo hybrid reconstruction were available for 180 666 of the 300 000 simulated events. For the energy estimate, the photon energy estimate was used, while the true value for $X_{\max}$ was determined from an interpolation based on the energy deposit at individual air-shower simulation steps. Fig. 9.2 depicts separate scatter plots for the three quantities with the MC value on the horizontal axis and the reconstructed value on the vertical axis.

Considering these plots, the accuracy of reconstruction can be evaluated visually. A large spread around the ideal diagonal axis corresponds to a worse detection and reconstruction quality, while a more concentrated distribution indicates a precise assignment. However, this type of presentation neglects the fact that an uncertainty value is also assigned to each

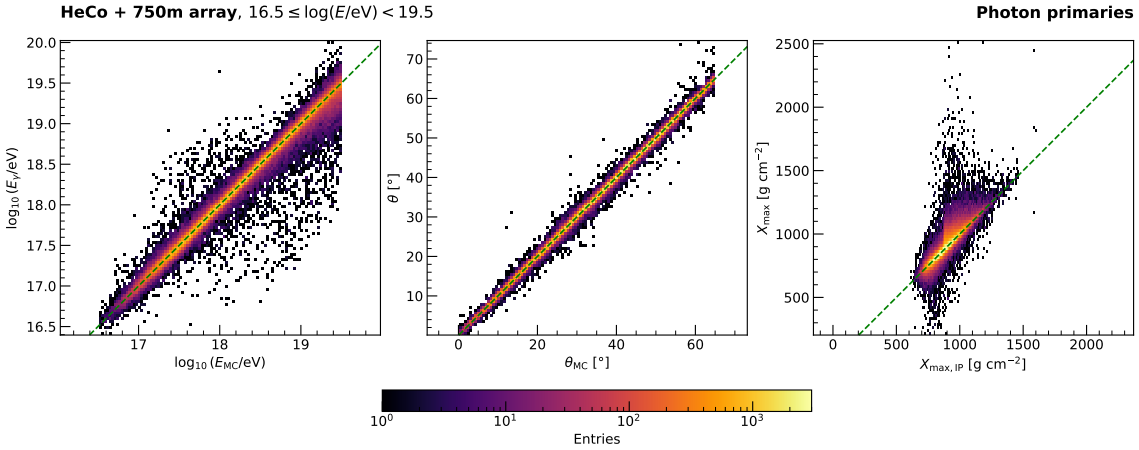

Fig. 9.2 Reconstructed key quantities of the photon simulations in relation to the input provided by the presimulated air shower data for the complete energy range of the simulations. The dispersion along the diagonal axis reflects the resolution of the simulated detector configuration. In an ideal detector, events would align precisely along the diagonal (dashed green line), forming a distinct line. Additionally, any deviations in the reconstruction process would manifest as shifts or tilts along the major axis.

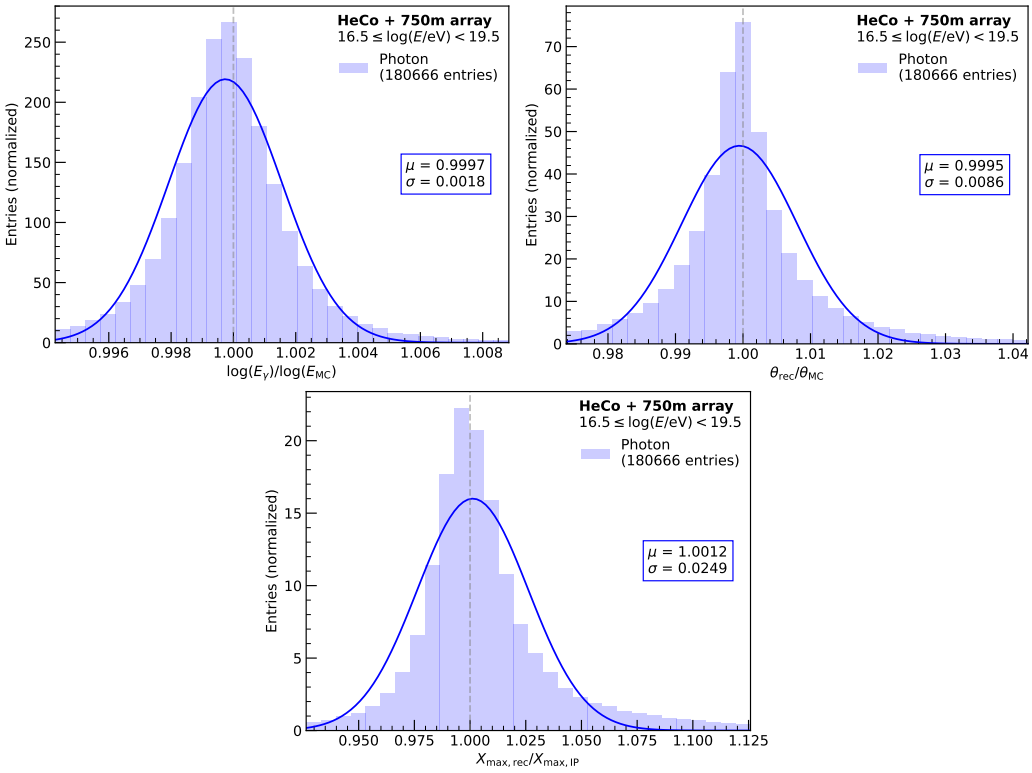

Fig. 9.3 Performance of the reconstruction for the three key quantities E_γ, θ, and $X_{\max}$, accompanied by a rough approximation using a normal distribution. The mean and width of the distribution are derived from the median and the 15.9% and 84.1% percentiles to enhance robustness against outliers in the dataset. However, it is evident that the data only partially conforms to a normal distribution. For simplicity, only the first two moments of the distributions are relied upon to define the resolution.

measurement in the reconstruction process. If a preselection is performed accordingly, the resolution can be improved considerably. Nevertheless, a certain degree of uncertainty and the associated spread of measured values is to be expected for technical and statistical reasons. If the reconstructed values are too accurate, this would also correspond to an undesirable deviation from a realistic detection behavior. Furthermore, a potential inherent bias in the reconstruction would be apparent from the shift or tilt of the reconstruction mapping axis.

Two conspicuous features of the photon simulation set could be derived from this illustration: Firstly, for large energies, i.e. $E \gtrsim 10^{18.5}$ eV, the detector has the slight tendency to underestimate the primary energy. Note, that here the photon optimized energy reconstruction was used, so this deviation cannot be attributed to issues with the missing energy correction. A second abnormality is found in the scatter plot for $X_{\max}$. For values above approximately $900\,\mathrm{g\,cm^{-2}}$ there seems to be a systematic smearing toward higher reconstructed values. Before an attempt at an explanation can be made here, a somewhat more systematic consideration including energy decomposition should follow first.

In this subsequent analysis step, the spread in the individual energy bins was quantified with the reconstructed value divided by the MC value. The corresponding distributions are depicted in Fig. 9.3, together with a normal distribution model. Mean and standard deviation of the model are determined from the median, respectively the 15.9% and 84.1% percentiles for robustness against outliers in the raw data, that might have been removed by quality cuts. The detector resolution is usually defined as the here depicted standard deviation of the reconstructed relative value, even if consideration of the data here suggests that higher moments of the distribution ought to be taken into account. It is evident that the detector can provide more precise estimates of the basic quantities of energy and arrival direction than for example for X_{max}. However, it should be noted that, when considering both nuclear primaries and photons, the accuracy of the energy reconstruction significantly suffers due to the a-priori indeterminacy of the primary particle type.

A detailed breakdown for these considerations into individual energy bins is available in the appendix under Figs. D.7, D.8, and D.9, in which it becomes apparent, without having to consider the relationship between depth of shower maximum and primary energy, that the tendency for the observed positive reconstruction bias of X_{max} is significantly more pronounced for $E \gtrsim 10^{18}$ eV. This energy bound suggests, since the reconstruction procedure is not adapted to photons behavior at higher energies, that the preshowering process included in the simulation framework may be a possible culprit of this deviation for a share of the events, in which case a geometrical dependence in regard to the geomagnetic field may be incorporated in the finalized simulation sample. In addition, this could perhaps also explain the deviation in the energy estimate.

Finally, in Fig. 9.4, the results of the resolution (and bias) measurement in the individual

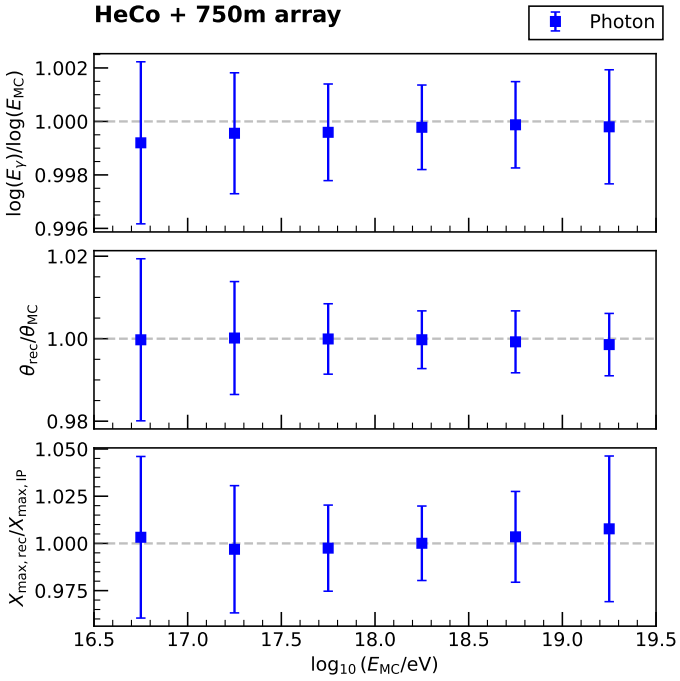

Fig. 9.4 Mean and standard variation of the three key quantities E_γ, θ, and X_{max} as reconstructed by the simulated detector relative to the input values as a function of energy. It is essential to emphasize that the bars do not correspond to uncertainties for the bias measurements. Instead they merely reflect the spread of the underlying distributions.

energy bins are compiled. This illustration shows that the resolution of the detector for all three quantities is best at the baseline design sweet spot of $E \sim 10^{18}$ eV. While the geometrical reconstruction retains its accuracy even for higher energies, the aforementioned spread of the photon sample in reconstructed energy and depth of shower maximum lead a degradation of the resolution for higher energies.

9.1.4. Reweighting of Events

After the simulation batches pass the previously detailed check levels, the events of both particle classes were reweighted individually to follow an assumed E^{-2} power-law spectrum for the photon flux. Fig. 9.5 depicts the distribution of the photons events in primary energy in an unweighted (left) and a weighted (right) histogram. The final event weights are calculated in two distinct steps. First, the individual energy bins of width 0.5 in $\log_{10}(E/\mathrm{eV})$ are normalized based on the total number of events contained in them. Then, afterwards, relative weights are calculated based on the factor E^{-2} for each event. The final weights are then obtained from the product of these two contributions.

This procedure is limited to logarithmic-uniform event distributions in the individual energy bins, otherwise prevailing structures in the spectrum would be reflected in the final spectrum as well. That is one of the reasons, why for the logarithmic energy spectrum of the provided air-shower simulations a uniform distribution was required earlier. In addition, this normalization method also allows the numerical scopes of individual energy bins to be extended, provided that they remain evenly distributed. This was done, for example, in the simulation set of the original analysis, where the particularly interesting low-energy region was supplemented by completely locally compiled simulations.

Event weights are stored separately in a dedicated ROOT file, from which they can be retrieved using the SDId of the events as a unique identifier. The proton simulations were

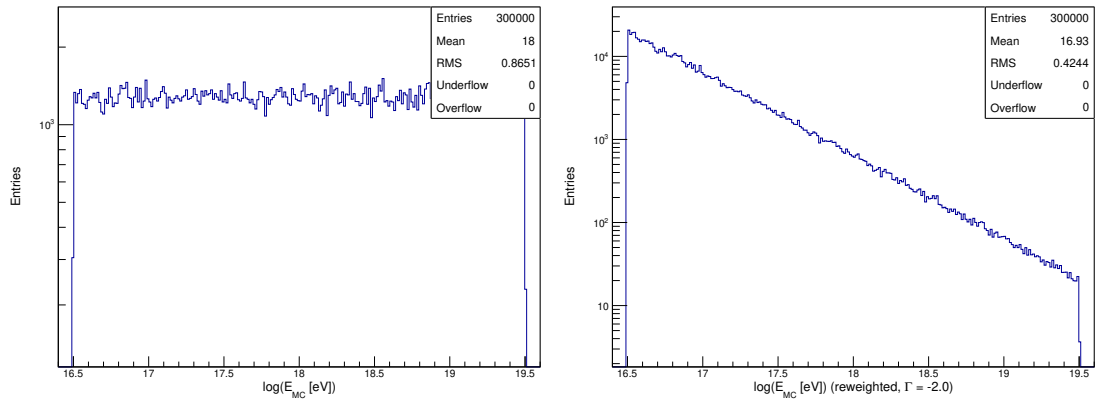

Fig. 9.5 Result of the reweighting process: The left subplot shows the distribution of simulated photon events in E_γ in its original form. On the right, the corresponding distribution is portrayed post reweighting, where events have been individually weighted in accordance with the power-law spectrum $\propto E^{-2}$ – consistent with the anticipated photon flux.

also reweighted in this scheme, to enable consistency between the simulation samples in the training and validation process of the BDT. In addition, the spectrum of nuclear cosmic radiation follows a very similar progression, thus also enhancing the realistic depiction of the background flux.

10. Outlook

The preceding sections discussed the progress achieved in updating the photon search in [2]. However, several essential steps are still forthcoming in the near future. In Sec. 9.1.2, the conducted steps for completing and reweighting the updated simulation set, have been detailed. Now, the immediate next step would be the renewed and updated calculation of the efficiency-weighted exposure based on the newly acquired simulation samples. After which the extended burn sample of real data events may be accessed and utilized for further self-contained inquiries into the analysis strategy. However, a rough skeleton should already have been drawn up from the findings of the simulation studies beforehand.

Not all ideas for updated aspects of the analysis could be investigated and realized in the scope of this thesis. For this reason, a handful of ideas are mentioned here that could provide the basis for future research work.

The statistical volume of the measurement data is limited, among other things, by the quality cuts initially applied. These are essential for any analysis, not only in high-energy physics, in order to ensure an appropriate separation of signal and background and, consequently, a certain conclusiveness of the results. Nonetheless, the cuts of the geometry level in particular do reduce the data set quite substantially. Only around 3.7% of the initial data events are accepted in the original constellation of geometry cuts [2]. Granted, since this level involves the very first cuts to the raw data, a significant relative reduction is to be expected; later cuts, on the other hand, are then applied to events that have already successfully passed the first checks and therefore have a certain quality. Of particular interest in this set is the `coreInInfillRegion` cut, that was newly introduced for the original analysis after issues with the analysis were identified upon considering the burn sample [144]. It requires that the core of the recorded air-shower associated with a data event has to be contained within the boundaries of the 750 m spaced infill array, which, if not fulfilled, has previously led to an event type with incomplete SD measurement. Fig. 10.1 illustrates this requirement. The deducted margin of half the row spacing has been picked more or less arbitrarily and was subsequently validated by a systematic study. The question now is whether this margin can be reduced slightly without significantly compromising the separation power of the multivariate analysis at hand. If, for example, it were possible to extend the contour to the edge of the sub-array, the instantaneous exposure could be increased by about 31.9% and the photon upper limits reduced accordingly. A corresponding redefinition of the cut would therefore be appealing.

Another aspect of the analysis that may be considered is the definition of the a-priori photon candidate cut. Following the philosophy of [129], it was fixed by the median of the final discriminant β in the photon distribution above 2×10^{17} eV (cf. Sec. 6). However,

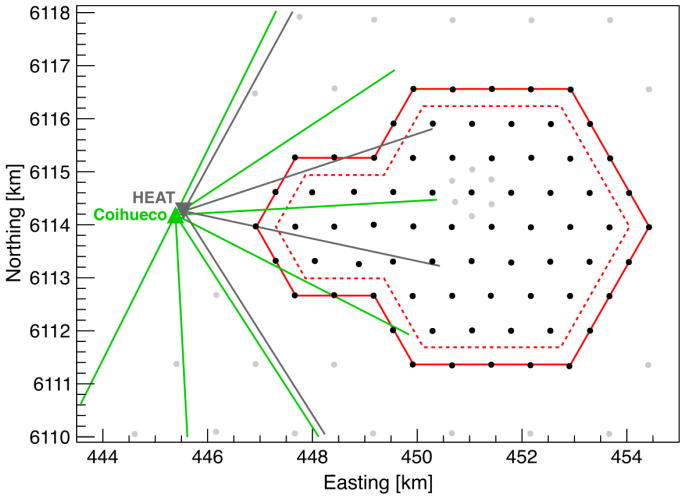

Fig. 10.1 Detector geometry in the 750 m spaced infill array [2]. The red line marks the outer border of this sub-array. For the initial `coreInInfillRegion` cut to be satisfied, the shower core on the ground has to be contained within the boundary defined by the red dashed contour, which is located at half distance between the first and second outer SD station rows of the sub-array.

it is also conceivable, for example, to set the critical value for each energy threshold separately. Given that the mass composition of UHE cosmic rays is most likely mixed and thus the measured events are distributed even further in β from the threshold value than was conservatively approximated with the proton simulations (cf. Fig. 6.2), it does appear sensible to allow an energy-threshold dependent photon candidate cut, thereby increasing the sensitivity while on the other hand also enabling a higher rate of background contamination. Since the photon efficiency is already included in the calculation of the upper limits, such a change would indeed have little effect there, but would nonetheless enhance the potential for the measurement of photon candidates, while at the same time increasing the probability that these were nuclear primaries. Ultimately, this is a fundamental design decision which can only be addressed to a lesser extent by gathering simulation data.

One final aspect remains, the scope of which could be the subject of a thesis in its own right: While the studies of Sec. 8.1 dealt with the discriminatory quantity S_b, the applicability of different observables or combinations thereof in the lower energy sector above 10^{17} eV remains to be inquired. Furthermore, the deployment of different MVA methods, such as deep neural networks [148], may prove beneficial for the discriminatory performance of the analysis, adapting to the specialties of this energy domain even better. While still quite novel in the field of astroparticle physics, they will certainly establish themselves in the upcoming years with the expected improvements in computational capacity.

11. Summary and Conclusions

In this master's thesis, the initial groundwork was laid for updating the photon search of [2]. Together with both the near doubling of the considered date range for the measurements and the extension in energy to a new lowest threshold of 1×10^{17} eV, some further adaptions to the low-energy instrumentation should be included. This updated photon search is planned to be compiled into a self-contained publication upon finishing.

For the work presented here, the original analysis was reproduced and its results validated,

before two pivotal aspects were addressed within separate simulation studies. S_b, one of the key observables for classification of the primary particle, was examined for potential improvements in the low-energy domain between 10^{17} eV and 10^{18} eV. Previously published results above this region were able to be extended downward and the application of the logarithmic variant $\log(S_b)$ was validated for this photon search. The variable exponent b was numerically optimized in this energy range, albeit yielding no significant enhancement in the separation power of the employed MVA. Future studies may further explore this endeavor.

The second study focused on the potential introduction of a novel quality cut, tied to the fit of the longitudinal energy deposit of an air shower and, consequently, the depth of the maximum shower development – an essential element in the MVA of the original publication. Several variations were tested, upon which the quantity $^{\Delta X_{\mathrm{max}}}/_{\mathrm{TL}}$ emerged as the most promising. Here, ΔX_{max} represents the uncertainty in the reconstruction of the shower maximum, while TL refers to the total recorded tracklength. The potential threshold value 0.5 was tested for its efficiency in a full-size data sample and checked for an intrinsic bias against photon events, which was found to be insignificant. Upon individual examination of the rejected events in the burn sample, this cut appears to effectively address the issues articulated in the motivation.

Finally, the production process of an updated and extended simulation set was illustrated. The corresponding section detailed the various checks performed to ensure integrity of the simulated data. With this newly acquired set of simulations, further systematic studies are now feasible, as indicated in the various subsections.

The updated photon search is anticipated to bring about significant enhancements in integrated exposure. This improvement can be attributed, among other factors, to the consideration of the nearly doubled measurement date range, now spanning the full Phase-1 data set. As a result, considerable improvements in the subsequently formulated upper limits on the photon flux are expected. Moreover, the increasing adaptation to the peculiarities of the low-energy extensions of the Pierre Auger Observatory is poised to elevate the sensitivity to photon candidates.

Appendix

A. References

[1] NASA, *Auroral beads seen from the International Space Station on Sept. 17, 2011.* `https://www.nasa.gov/sites/default/files/thumbnails/image/iss029-e-6012.jpg` (visited on 07/25/2023).

[2] P. Abreu et al., Pierre Auger Collaboration, "A Search for Photons with Energies Above 2×10^{17} eV Using Hybrid Data from the Low-Energy Extensions of the Pierre Auger Observatory", ApJ **933**, 125 (2022).

[3] C.-A. de Coulomb, "Troisième Mémoire sur l'Electricité et le Magnétisme", Mém. Académie R. Sci. Paris, 612–638 (1785).

[4] H. Becquerel, "Sur les radiations invisibles émises par les corps phosphorescents", Comptes Rendus Hebd. Séances L'Académie Sci. **122**, 501–503 (1896).

[5] A. Eve, "On the ionization of the atmosphere due to radioactive matter", Lond. Edinb. Dublin Philos. Mag. J. Sci. **21**, 26–40 (1911).

[6] T. Wulf, "Ein neues Elektrometer für statische Ladungen", Phys. Z. **8**, 246–248 (1907).

[7] T. Wulf, "Über den Ursprung der in der Atmosphäre vorhandenen γ-Strahlung", Phys. Z. **10**, 997–1003 (1909).

[8] T. Wulf, "Beobachtungen über Strahlung hoher Durchdringungsfähigkeit auf dem Eiffelturm", Phys. Z. **11**, 811–813 (1910).

[9] J. W. Cronin, "From spontaneous ionization to subatomic physics: Some vignettes from cosmic ray history", Astroparticle Physics, Centenary of cosmic ray discovery **53**, 6–18 (2014).

[10] D. Pacini, "La radiation pénétrante sur la mer", Radium (Paris) **8**, 307–312 (1911).

[11] F. Linke, "Luftelektrische Messungen bei zwölf Ballonfahrten", Abh. Ges. Wiss. Gött. Math.-Phys. Kl. **3**, 1–90 (1904).

[12] A. Gockel and T. Wulf, "Beobachtungen über die Radioaktivität der Atmosphäre im Hochgebirge", Phys. Z. **9**, 907–911 (1908).

[13] A. Gockel, "Luftelektrische Beobachtungen bei einer Ballonfahrt", Phys. Z. **11**, 280–282 (1910).

[14] W. Kohlhörster, "Messungen der durchdringenden Strahlung bis in Höhen von 9300 m", Verhandlungen Dtsch. Phys. Ges. **16**, 719–721 (1914).

[15] V. F. Hess, "Über Neuerungen und Erfahrungen an den Radiummessungen nach der γ-Strahlenmethode", Phys. Z. **14**, 1135–1141 (1913).

[16] V. F. Hess, "Über Beobachtungen der durchdringenden Strahlung bei sieben Freiballonfahrten", Phys. Z. **12**, 1090 (1912).

[17] The Nobel Foundation, "Nobel Lectures, Physics 1922-1941", Elsevier Publ. Co. Amst. (1965).

[18]R. A. Millikan and I. S. Bowen, "High Frequency Rays of Cosmic Origin I. Sounding Balloon Observations at Extreme Altitudes", Phys. Rev. **27**, 353–361 (1926).

[19]R. A. Millikan and R. M. Otis, "High Frequency Rays of Cosmic Origin II. Mountain Peak and Airplane Observations", Phys. Rev. **27**, 645–658 (1926).

[20]R. A. Millikan and G. H. Cameron, "High Frequency Rays of Cosmic Origin III. Measurements in Snow-Fed Lakes at High Altitudes", Phys. Rev. **28**, 851–868 (1926).

[21]E. Regener and G. Pfotzer, "Vertical Intensity of Cosmic Rays by Threefold Coincidences in the Stratosphere", Nature **136**, 718–719 (1935).

[22]B. Rossi, "Interaction between Cosmic Rays and Matter", Nature **132**, 173–174 (1933).

[23]A. Letessier-Selvon and T. Stanev, "Ultrahigh energy cosmic rays", Rev. Mod. Phys. **83**, 907–942 (2011).

[24]H. J. Bhabha, W. Heitler, and N. F. Mott, "The passage of fast electrons and the theory of cosmic showers", Proc. R. Soc. Lond. A **159**, 432–458 (1937).

[25]P. V. Auger, R. Maze, and T. Grivet-Meyer, "Grandes gerbes cosmiques atmosphériques contenant des corpuscules ultra-pénétrantes", Comptes Rendus, Académie des Sciences **206**, 1721–1723 (1938).

[26]P. V. Auger and R. Maze, "Les grandes gerbes cosmiques de l'atmosphère", Comptes Rendus, Académie des Sciences **207**, 228–229 (1938).

[27]P. V. Auger et al., "Extensive Cosmic-Ray Showers", Rev. Mod. Phys. **11**, 288–291 (1939).

[28]C. T. R. Wilson, "On an expansion apparatus for making visible the tracks of ionising particles in gases and some results obtained by its use", Proc. R. Soc. Lond. A **87**, 277–292 (1912).

[29]C. D. Anderson, "The Positive Electron", Phys. Rev. **43**, 491–494 (1933).

[30]C. D. Anderson, "The Apparent Existence of Easily Deflectable Positives", Science **76**, 238–239 (1932).

[31]S. H. Neddermeyer and C. D. Anderson, "Note on the Nature of Cosmic-Ray Particles", Phys. Rev. **51**, 884–886 (1937).

[32]G. D. Rochester and C. C. Butler, "Evidence for the Existence of New Unstable Elementary Particles", Nature **160**, 855–857 (1947).

[33]V. D. Hopper and S. Biswas, "Evidence Concerning the Existence of the New Unstable Elementary Neutral Particle", Phys. Rev. **80**, 1099–1100 (1950).

[34]R Armenteros et al., "LVI. The properties of charged V-particles", Lond. Edinb. Dublin Philos. Mag. J. Sci. **43**, 597–611 (1952).

[35]A. Bonetti et al., "On the existence of unstable charged particles of hyperprotonic mass", Nuovo Cim **10**, 1736–1743 (1953).

[36]V. Verzi, Pierre Auger Collaboration, "Measurement of the energy spectrum of ultra-high energy cosmic rays using the Pierre Auger Observatory", PoS(ICRC2019) **358**, 450 (2019).

[37]M. G. Aartsen et al., IceCube Collaboration, "Measurement of the cosmic ray energy spectrum with IceTop-73", Phys. Rev. D **88**, 042004 (2013).

[38]V. V. Prosin et al., "Tunka-133: Results of 3 year operation", Nucl. Instrum. & Methods Phys. Res. Sect. A **756**, 94–101 (2014).

[39]G. Giacinti and D. V. Semikoz, "Modeling of the Cosmic Ray Flux at the Knee", PoS(ICRC2023), 420 (2023).

[40]M. Kachelrieß and D. V. Semikoz, "Cosmic ray models", Progress in Particle and Nuclear Physics **109**, 103710 (2019).

[41]J. R. Hörandel, "On the knee in the energy spectrum of cosmic rays", Astroparticle Physics **19**, 193–220 (2003).

[42]D. Allard et al., "UHE nuclei propagation and the interpretation of the ankle in the cosmic-ray spectrum", A&A **443**, L29–L32 (2005).

[43]V. Berezinsky, A. Gazizov, and S. Grigorieva, "On astrophysical solution to ultrahigh energy cosmic rays", Phys. Rev. D **74**, 043005 (2006).

[44]A. Aab et al., Pierre Auger Collaboration, "Evidence for a mixed mass composition at the 'ankle' in the cosmic-ray spectrum", Physics Letters B **762**, 288–295 (2016).

[45]A. Aab et al., Pierre Auger Collaboration, "Features of the Energy Spectrum of Cosmic Rays above 2.5×10^{18} eV Using the Pierre Auger Observatory", Phys. Rev. Lett. **125**, 121106 (2020).

[46]J. Abraham et al., Pierre Auger Collaboration, "Observation of the Suppression of the Flux of Cosmic Rays above 4×10^{19} eV", Phys. Rev. Lett. **101**, 061101 (2008).

[47]J. Abraham et al., Pierre Auger Collaboration, "Measurement of the energy spectrum of cosmic rays above 10^{18} eV using the Pierre Auger Observatory", Physics Letters B **685**, 239–246 (2010).

[48]K. Greisen, "End to the Cosmic-Ray Spectrum?", Phys. Rev. Lett. **16**, 748–750 (1966).

[49]G. T. Zatsepin and V. A. Kuz'min, "Upper limit of the spectrum of cosmic rays", Sov. J. Exp. Theor. Phys. Lett. **4**, 78 (1966).

[50]M. Kachelrieß, E. Parizot, and D. V. Semikoz, "The GZK horizon and constraints on the cosmic ray source spectrum from observations in the GZK regime", Jetp Lett. **88**, 553–557 (2008).

[51]R. Aloisio, V. Berezinsky, and A. Gazizov, "Ultra high energy cosmic rays: The disappointing model", Astroparticle Physics **34**, 620–626 (2011).

[52]A. H. Compton, "A Geographic Study of Cosmic Rays", Phys. Rev. **43**, 387–403 (1933).

[53]L. Alvarez and A. H. Compton, "A Positively Charged Component of Cosmic Rays", Phys. Rev. **43**, 835–836 (1933).

[54]C. Grupen, *Astroparticle Physics* (Springer International Publishing, Cham, 2020).

[55]A. Aab et al., Pierre Auger Collaboration, "Depth of maximum of air-shower profiles at the Pierre Auger Observatory. I. Measurements at energies above $10^{17.8}$ eV", Phys. Rev. D **90**, 122005 (2014).

[56]A. Aab et al., Pierre Auger Collaboration, "Depth of maximum of air-shower profiles at the Pierre Auger Observatory. II. Composition implications", Phys. Rev. D **90**, 122006 (2014).

[57]J. Abraham et al., Pierre Auger Collaboration, "Anisotropy studies around the galactic centre at EeV energies with the Auger Observatory", Astroparticle Physics **27**, 244–253 (2007).

[58]P. Abreu et al., Pierre Auger Collaboration, "Cosmological implications of photon-flux upper limits at ultrahigh energies in scenarios of Planckian-interacting massive particles for dark matter", Phys. Rev. D **107**, 042002 (2023).

[59]P. Abreu et al., Pierre Auger Collaboration, "Limits to Gauge Coupling in the Dark Sector Set by the Nonobservation of Instanton-Induced Decay of Super-Heavy Dark Matter in the Pierre Auger Observatory Data", Phys. Rev. Lett. **130**, 061001 (2023).

[60]E. Fermi, "On the Origin of the Cosmic Radiation", Phys. Rev. **75**, 1169–1174 (1949).

[61]R. D. Blandford and J. P. Ostriker, "Particle acceleration by astrophysical shocks", ApJ **221**, L29–L32 (1978).

[62]A. M. Hillas, "Can diffusive shock acceleration in supernova remnants account for high-energy galactic cosmic rays?", J. Phys. G: Nucl. Part. Phys. **31**, R95 (2005).

[63]M. Ahlers, "Deciphering the Dipole Anisotropy of Galactic Cosmic Rays", Phys. Rev. Lett. **117**, 151103 (2016).

[64]M. G. Aartsen et al., IceCube Collaboration, "Anisotropy in cosmic-ray arrival directions in the southern hemisphere based on six years of data from the IceCube detector", ApJ **826**, 220 (2016).

[65]A. H. Compton and I. A. Getting, "An Apparent Effect of Galactic Rotation on the Intensity of Cosmic Rays", Phys. Rev. **47**, 817–821 (1935).

[66]J. Candia, S. Mollerach, and E. Roulet, "Cosmic ray drift, the second knee and galactic anisotropies", J. High Energy Phys. **2002**, 032 (2002).

[67]A. Aab et al., Pierre Auger Collaboration, "Large-scale Cosmic-Ray Anisotropies above 4 EeV Measured by the Pierre Auger Observatory", ApJ **868**, 4 (2018).

[68]A. Aab et al., Pierre Auger Collaboration, "Cosmic-Ray Anisotropies in Right Ascension Measured by the Pierre Auger Observatory", ApJ **891**, 142 (2020).

[69]M. Kachelrieß and P. D. Serpico, "The Compton–Getting effect on ultra-high energy cosmic rays of cosmological origin", Physics Letters B **640**, 225–229 (2006).

[70]D. Harari, S. Mollerach, and E. Roulet, "Anisotropies of ultrahigh energy cosmic ray nuclei diffusing from extragalactic sources", Phys. Rev. D **92**, 063014 (2015).

[71]A. Aab et al., Pierre Auger Collaboration, "Search for magnetically-induced signatures in the arrival directions of ultra-high-energy cosmic rays measured at the Pierre Auger Observatory", JCAP **2020**, 017 (2020).

[72]S. Mollerach and E. Roulet, "Progress in high-energy cosmic ray physics", Progress in Particle and Nuclear Physics **98**, 85–118 (2018).

[73]A. M. Hillas, "The Origin of Ultra-High-Energy Cosmic Rays", Annu. Rev. Astron. Astrophys. **22**, 425–444 (1984).

[74]T. Stanev, *High Energy Cosmic Rays*, 3rd ed., Astrophysics and Space Science Library (Springer International Publishing, Cham, 2021).

[75]P. K. F. Grieder, *Extensive Air Showers: High Energy Phenomena and Astrophysical Aspects - A Tutorial, Reference Manual and Data Book* (Springer, Berlin, Heidelberg, 2010).

[76]F. Schmidt, University of Leeds, UK, *CORSIKA Shower Images: Proton Showers*, `https://www-zeuthen.desy.de/~jknapp/fs/proton-showers.html` (visited on 09/28/2023).

[77]J. Matthews, "A Heitler model of extensive air showers", Astroparticle Physics **22**, 387–397 (2005).

[78]R. Engel and D. Schmidt, "Indirect Detection of Cosmic Rays", in *Handbook of Particle Detection and Imaging*, edited by I. Fleck et al. (Springer International Publishing, Cham, 2021), pp. 801–849.

[79]T. K. Gaisser and A. M. Hillas, "Reliability of the Method of Constant Intensity Cuts for Reconstructing the Average Development of Vertical Showers", PoS(ICRC1977) **8**, 353 (1977).

[80]K. Kamata and J. Nishimura, "The Lateral and the Angular Structure Functions of Electron Showers", Prog. Theor. Phys. Suppl. **6**, 93–155 (1958).

[81]K. Greisen, "Cosmic Ray Showers", Annu. Rev. Nucl. Part. Sci. **10**, 63–108 (1960).

[82]M. Ave et al., "Precise measurement of the absolute fluorescence yield of the 337nm band in atmospheric gases", Astroparticle Physics **42**, 90–102 (2013).

[83]H. M. J. Barbosa et al., "Determination of the calorimetric energy in extensive air showers", Astroparticle Physics **22**, 159–166 (2004).

[84]M. de Naurois and D. Mazin, "Ground-based detectors in very-high-energy gamma-ray astronomy", Comptes Rendus Physique **16**, 610–627 (2015).

[85]Pierre Auger Collaboration, "The Pierre Auger Cosmic Ray Observatory", Nucl. Instrum. & Methods Phys. Res. Sect. A **798**, 172–213 (2015).

[86]S. Fliescher, Pierre Auger Collaboration, "Radio detection of cosmic ray induced air showers at the Pierre Auger Observatory", Nucl. Instrum. & Methods Phys. Res. Sect. A **662**, S124–S129 (2012).

[87]T. Huege et al., "The LOPES experiment—Recent results, status and perspectives", Nucl. Instrum. & Methods Phys. Res. Sect. A **662**, S72–S79 (2012).

[88]B. Fuchs, Pierre Auger Collaboration, "The Auger Engineering Radio Array", Nucl. Instrum. & Methods Phys. Res. Sect. A **692**, 93–97 (2012).

[89]F. G. Schröder, "Radio detection of cosmic-ray air showers and high-energy neutrinos", Progress in Particle and Nuclear Physics **93**, 1–68 (2017).

[90]D. Veberič, *Maps of the Piere Auger Observatory*, (2020) `https://web.iap.kit.edu/darko/auger/auger-array/auger_array-pdf/` (visited on 10/26/2023).

[91]Pierre Auger Collaboration, *The Pierre Auger Observatory @ flickr*, (July 2015) `https://www.flickr.com/people/134252569@N07/` (visited on 10/26/2023).

[92]J. Abraham et al., Pierre Auger Collaboration, "Trigger and aperture of the surface detector array of the Pierre Auger Observatory", Nucl. Instrum. & Methods Phys. Res. Sect. A **613**, 29–39 (2010).

[93]X. Bertou et al., "Calibration of the surface array of the Pierre Auger Observatory", Nucl. Instrum. & Methods Phys. Res. Sect. A **568**, 839–846 (2006).

[94]G. Cataldi, Pierre Auger Collaboration, "Towards AugerPrime: the upgrade of the Pierre Auger Observatory", Nuclear and Particle Physics Proceedings **291–293**, 96–101 (2017).

[95]T. Huege, "The Radio Detector of the Pierre Auger Observatory – status and expected performance", EPJ Web Conf. **283**, 06002 (2023), arXiv:2305.10104 [astro-ph].

[96]A. Aab et al., Pierre Auger Collaboration, *The Pierre Auger Observatory Upgrade - Preliminary Design Report*, (Apr. 12, 2016) arXiv:1604.03637 [astro-ph], preprint.

[97]J. Abraham et al., Pierre Auger Collaboration, "The fluorescence detector of the Pierre Auger Observatory", Nucl. Instrum. & Methods Phys. Res. Sect. A **620**, 227–251 (2010).

[98]S. Eickhoff, M. Niechciol, and M. Risse, "Extending the search for primary photons with the hybrid detector to energies below 1 EeV", GAP **2019**, 010 (2019).

[99]P. Abreu et al., Pierre Auger Collaboration, "The energy spectrum of cosmic rays beyond the turn-down around 10^{17} eV as measured with the surface detector of the Pierre Auger Observatory", Eur. Phys. J. C **81**, 966 (2021).

[100]D. Kuempel, K. H. Kampert, and M. Risse, "Geometry reconstruction of fluorescence detectors revisited", Astroparticle Physics **30**, 167–174 (2008).

[101]Pierre Auger Collaboration, *Open Data*, version 2, 10.5281/zenodo.6867688, Dec. 22, 2022.

[102]A. Abdul Halim et al., Pierre Auger Collaboration, *The Pierre Auger Observatory Open Data*, (Sept. 28, 2023) arXiv:2309.16294 [astro-ph], preprint.

[103]Z. Cao et al., "Ultrahigh-energy photons up to 1.4 petaelectronvolts from 12 γ-ray Galactic sources", Nature **594**, 33–36 (2021).

[104]M. Risse and P. Homola, "Search for ultra-high-energy photons using air-showers", Mod. Phys. Lett. A **22**, 749–766 (2007).

[105]D. Fargion, B. Mele, and A. Salis, "Ultra-High-Energy Neutrino Scattering onto Relic Light Neutrinos in the Galactic Halo as a Possible Source of the Highest Energy Extragalactic Cosmic Rays", ApJ **517**, 725 (1999).

[106]P. Bhattacharjee and G. Sigl, "Origin and propagation of extremely high-energy cosmic rays", Physics Reports **327**, 109–247 (2000).

[107]Pierre Auger Collaboration, "Searches for Ultra-High-Energy Photons at the Pierre Auger Observatory", Universe **8**, 579 (2022).

[108]G. Ros et al., "A new composition-sensitive parameter for ultra-high energy cosmic rays", Astroparticle Physics **35**, 140–151 (2011).

[109]G. Ros et al., "Improving photon-hadron discrimination based on cosmic ray surface detector data", Astroparticle Physics **47**, 10–17 (2013).

[110]C. Bleve and Pierre Auger Collaboration, "Updates on the neutrino and photon limits from the Pierre Auger Observatory", PoS(ICRC2015) **34**, 1103 (2015).

[111]P. Abreu et al., Pierre Auger Collaboration, "Search for photons above 10^{19} eV with the surface detector of the Pierre Auger Observatory", JCAP **2023**, 021 (2023).

[112]J. K. Daugherty and A. K. Harding, "Compton Scattering in Strong Magnetic Fields", ApJ **309**, 362 (1986).

[113]P. Homola et al., "Simulation of ultra-high energy photon propagation in the geomagnetic field", Computer Physics Communications **173**, 71–90 (2005).

[114]L. D. Landau and I. J. Pomeranchuk, "The limits of applicability of the theory of Bremsstrahlung by electrons and of the creation of pairs at large energies", Dokl. Akad. Nauk SSSR **92**, 535 (1953).

[115]L. D. Landau and I. J. Pomeranchuk, "Electron cascade process at very high energies", Dokl. Akad. Nauk SSSR **92**, 735 (1953).

[116]H Bethe and W Heitler, "On the stopping of fast particles and on the creation of positive electrons", Proc. R. Soc. Lond. A **146**, 83–112 (1934).

[117]A. B. Migdal, "Bremsstrahlung and Pair Production in Condensed Media at High Energies", Phys. Rev. **103**, 1811–1820 (1956).

[118]H. D. Hansen et al., "Is the Electron Radiation Length Constant at High Energies?", Phys. Rev. Lett. **91**, 014801 (2003).

[119]G. J. Feldman and R. D. Cousins, "Unified approach to the classical statistical analysis of small signals", Phys. Rev. D **57**, 3873–3889 (1998).

[120]N. González et al., Pierre Auger Collaboration, "Search for primary photons at tens of PeV with the Pierre Auger Observatory", PoS(ICRC2023) **444**, 238 (2023).

[121]P. Savina et al., Pierre Auger Collaboration, "A search for ultra-high-energy photons at the Pierre Auger Observatory exploiting air-shower universality", PoS(ICRC2021), 373 (2021).

[122]R. U. Abbasi et al., Telescope Array Collaboration, "Constraints on the diffuse photon flux with energies above 1018 eV using the surface detector of the Telescope Array experiment", Astroparticle Physics **110**, 8–14 (2019).

[123]O. E. Kalashev, I. V. Kharuk, and M. Y. Kuznetsov, Telescope Array Collaboration, "Telescope Array search for EeV photons", PoS(ICRC2021) **395**, 864 (2021).

[124]W. D. Apel et al., KASCADE-Grande Collaboration, "KASCADE-Grande Limits on the Isotropic Diffuse Gamma-Ray Flux between 100 TeV and 1 EeV", ApJ **848**, 1 (2017).

[125]Y. A. Fomin et al., "Constraints on the flux of $\sim (10^{16} - 10^{17.5})$ eV cosmic photons from the EAS–MSU muon data", Phys. Rev. D **95**, 123011 (2017).

[126]M. Niechciol, Pierre Auger Collaboration, "Latest results from the searches for ultra-high-energy photons and neutrinos at the Pierre Auger Observatory", PoS(ICRC2023) **444**, 1488 (2023).

[127]A. Aab et al., Pierre Auger Collaboration, "A search for point sources of EeV photons", ApJ **789**, 160 (2014).

[128]A. Aab et al., Pierre Auger Collaboration, "A Targeted Search for Point Sources of EeV Photons with the Pierre Auger Observatory", ApJL **837**, L25 (2017).

[129]A. Aab et al., Pierre Auger Collaboration, "Search for photons with energies above 10^{18} eV using the hybrid detector of the Pierre Auger Observatory", JCAP **2017**, 009–009 (2017).

[130]A. Hoecker et al., *TMVA - Toolkit for Multivariate Data Analysis*, (July 7, 2009) arXiv:physics/0703039, preprint.

[131]B. P. Roe et al., "Boosted decision trees as an alternative to artificial neural networks for particle identification", Nucl. Instrum. & Methods Phys. Res. Sect. A **543**, 577–584 (2005).

[132]H.-J. Yang, B. P. Roe, and J. Zhu, "Studies of stability and robustness for artificial neural networks and boosted decision trees", Nucl. Instrum. & Methods Phys. Res. Sect. A **574**, 342–349 (2007).

[133]*Napoli + Praha library - AugerWiki*, https://www.auger.unam.mx/AugerWiki/NapoliLibrary (visited on 12/23/2023).

[134]D. Heck et al., "CORSIKA: A Monte Carlo Code to Simulate Extensive Air Showers", Forschungszentrum Karlsr. Rep. FZKA **6019** (1998).

[135]K. Werner, F.-M. Liu, and T. Pierog, "Parton ladder splitting and the rapidity dependence of transverse momentum spectra in deuteron-gold collisions at the BNL Relativistic Heavy Ion Collider", Phys. Rev. C **74**, 044902 (2006).

[136]T. Pierog et al., "EPOS LHC: Test of collective hadronization with data measured at the CERN Large Hadron Collider", Phys. Rev. C **92**, 034906 (2015).

[137]S. Argirò et al., "The offline software framework of the Pierre Auger Observatory", Nucl. Instrum. & Methods Phys. Res. Sect. A **580**, 1485–1496 (2007).

[138]R. Brun and F. Rademakers, "ROOT — An object oriented data analysis framework", Nucl. Instrum. & Methods Phys. Res. Sect. A **389**, 81–86 (1997).

[139]J. Pivarski et al., *Uproot*, version v5.0.11, 10.5281/zenodo.8239801, Aug. 11, 2023.

[140]J. D. Hunter, "Matplotlib: A 2D Graphics Environment", Comput. Sci. Eng. **9**, 90–95 (2007).

[141]N. González et al., "A muon-based observable for a photon search at 30–300 PeV", Astroparticle Physics **114**, 48–59 (2020).

[142]I. D. Vergara Quispe et al., "S_b for gamma/hadron separation in SWGO", PoS(ICRC2023), 952 (2023).

[143]T. Fehler and M. Niechciol, "Studying the S_b observable for photon searches in the energy range 10^{17} to 10^{18} eV", GAP **2023**, 020 (2023).

[144]M. Niechciol and M. Risse, "Search for primary photons with the hybrid detector at energies above 2^{17} eV: application to data", GAP **2020**, 044 (2020).

[145]S. Y. BenZvi et al., "The Lidar system of the Pierre Auger Observatory", Nucl. Instrum. & Methods Phys. Res. Sect. A **574**, 171–184 (2007).

[146]P. Abreu et al., Pierre Auger Collaboration, "Identifying clouds over the Pierre Auger Observatory using infrared satellite data", Astroparticle Physics **50–52**, 92–101 (2013).

[147]P. Virtanen et al., "SciPy 1.0: fundamental algorithms for scientific computing in Python", Nat. Methods **17**, 261–272 (2020).

[148]E. Guido, M. Niechciol, and M. Risse, "Towards searching for ultra-high-energy photons with deep learning techniques", PoS(ICRC2023), 191 (2023).

[149] R. L. Workman et al., Particle Data Group, "Review of Particle Physics", PTEP **2022**, 083C01 (2022).

B. List of Acronyms and Abbreviations

a.s.l.	above sea level
AERA	Auger Engineering Radio Array
AGN	Active Galactic Nucleus
AMIGA	Auger Muon and Infilled Ground Array
BDT	Boosted Decision Tree
BH	Black Hole
BLF	Balloon Launching Facility
C.L.	Confidence Level
CDAS	Central Data Acquisition System
CDF	Cumulative Distribution Function
CERN	Conseil européen pour la recherche nucléaire (European Organization for Nuclear Research)
CLF	Central Laser Facility
CMB	Cosmic Microwave Background
CORSIKA	Cosmic Ray Simulations for KASCADE
EAS-MSU	Extensive Air Shower Array of Moscow State University
EPOS	Energy conserving quantum mechanical multiple scattering approach, based on Partons (parton ladders) Off-shell remnants and Splitting of parton ladders
FADC	Flash Analog-to-Digital Converter
FD	Fluorescence Detector
GPS	Global Positioning System
GRB	Gamma-Ray Burst
GZK	Greisen, Zatsepin, Kuz'min
HEAT	High Elevation Auger Telescopes
HeCo	HEAT and Coihueco combined FD system
ICRC	International Cosmic-Ray Conference
KASCADE	Karlsruhe Shower Core and Array Detector
LDF	Lateral Distribution Function
LED	light emitting diode
LHC	Large Hadron Collider
LPM	Landau, Pomeranchuk, Migdal
MC	Monte Carlo
MVA	Multivariate Analysis
ndf	number of degrees of freedom
NKG	Nishimura, Kamata, Greisen
PMT	Photomultiplier Tube

QED	Quantum Electrodynamics
RG	Radio Galaxies
SD	Surface Detector
SDP	Shower-Detector Plane
SM	Standard Model (of Particle Physics)
SNR	Supernova Remnants
SSD	Surface Scintillation Detector
SWGO	Southern Wide-field Gamma-ray Observatory
TMVA	Toolkit for Multivariate Analysis
ToT	Time over threshold
UHE	Ultra-High-Energy
UHECR	Ultra-High-Energy Cosmic Ray
UMD	Underground Muon Detector
URB	Universal Radio Background
UV	Ultraviolet (light)
VEM	Vertical equivalent muon
WCD	Water Cherenkov Detector
WIMP	Weakly Interacting Massive Particle
XLF	eXtreme Laser Facility

C. Auxiliary Calculations

C.1. Decay Length of Energetic Muons

$$\begin{aligned} E &= \gamma m_\mu c^2 = 1\,\mathrm{GeV} \\ \Rightarrow \gamma &= \frac{E}{m_\mu c^2} = 9.464 \\ \ell &\approx \gamma c \tau_\mu = 6.24\,\mathrm{km} \quad \text{since} \quad \beta \approx 1 \end{aligned} \tag{C.1}$$

With:

Quantity	Notation	Value	Reference
Invariant mass of muon	m_μ	$(105.658\,375\,5 \pm 0.000\,002\,3)\,\mathrm{MeV\,c^{-2}}$	[149]
Mean lifetime of muon	τ_μ	$(2.196\,981\,1 \pm 0.000\,002\,2) \times 10^{-6}\,\mathrm{s}$	[149]
Vacuum speed of light	c	$299\,792\,458\,\mathrm{m\,s^{-1}}$	per def.
Lorentz factor	γ	-	-
Relativistic velocity	β	$\sqrt{1-\gamma^{-2}}$	-

C.2. Estimating the Uncertainty of the Bias $r_\gamma - r_p$

Since the two ratios r_i $(i = \gamma, p)$ are given by the unaffected number of events divided by their total count n_i, one can assume them to follow a binomial distribution, with the

standard deviation scaled by their total number n_i:

$$\sigma_{r_i} = \frac{1}{n_i}\sqrt{n_i\, p\,(1-p)} = \sqrt{\frac{p\,(1-p)}{n_i}}, \tag{C.2}$$

with p being the probability for an individual event to be unaffected. As this probability is unknown, p will be replaced with its best estimate r_i.

For an arbitrary linear combination of normal distributed variables x_i

$$f = \sum_i a_i\, x_i \tag{C.3}$$

the combined uncertainty is calculated from the individual standard deviations σ_i and the correlation coefficients ρ_{ij} as

$$\sigma_f^2 = \sum_i a_i^2\, \sigma_i^2 + \sum_i \sum_{j \neq i} a_i a_j \rho_{ij} \sigma_i \sigma_j. \tag{C.4}$$

For $f = r_\gamma - r_p$ and under the reasonable approximation that deviations from r_γ and r_p are not correlated ($\rho_{ij} \approx 0$), this results in the following expression for the uncertainty of the bias estimator:

$$\sigma_{\text{bias}}^2 \approx \sigma_{r_\gamma}^2 + \sigma_{r_p}^2 \tag{C.5}$$

$$\Rightarrow \sigma_{\text{bias}} = \sqrt{\sigma_{r_\gamma}^2 + \sigma_{r_p}^2} = \sqrt{\frac{r_\gamma(1-r_\gamma)}{n_\gamma} - \frac{r_p(1-r_p)}{n_p}} \tag{C.6}$$

This was used to estimate the uncertainty of the bias as function of the cut threshold and the corresponding ratios r_i.

D. Additional Figures

D.1. Additional Figures for the Optimization of the S_b Exponent

D.1.1. S_b

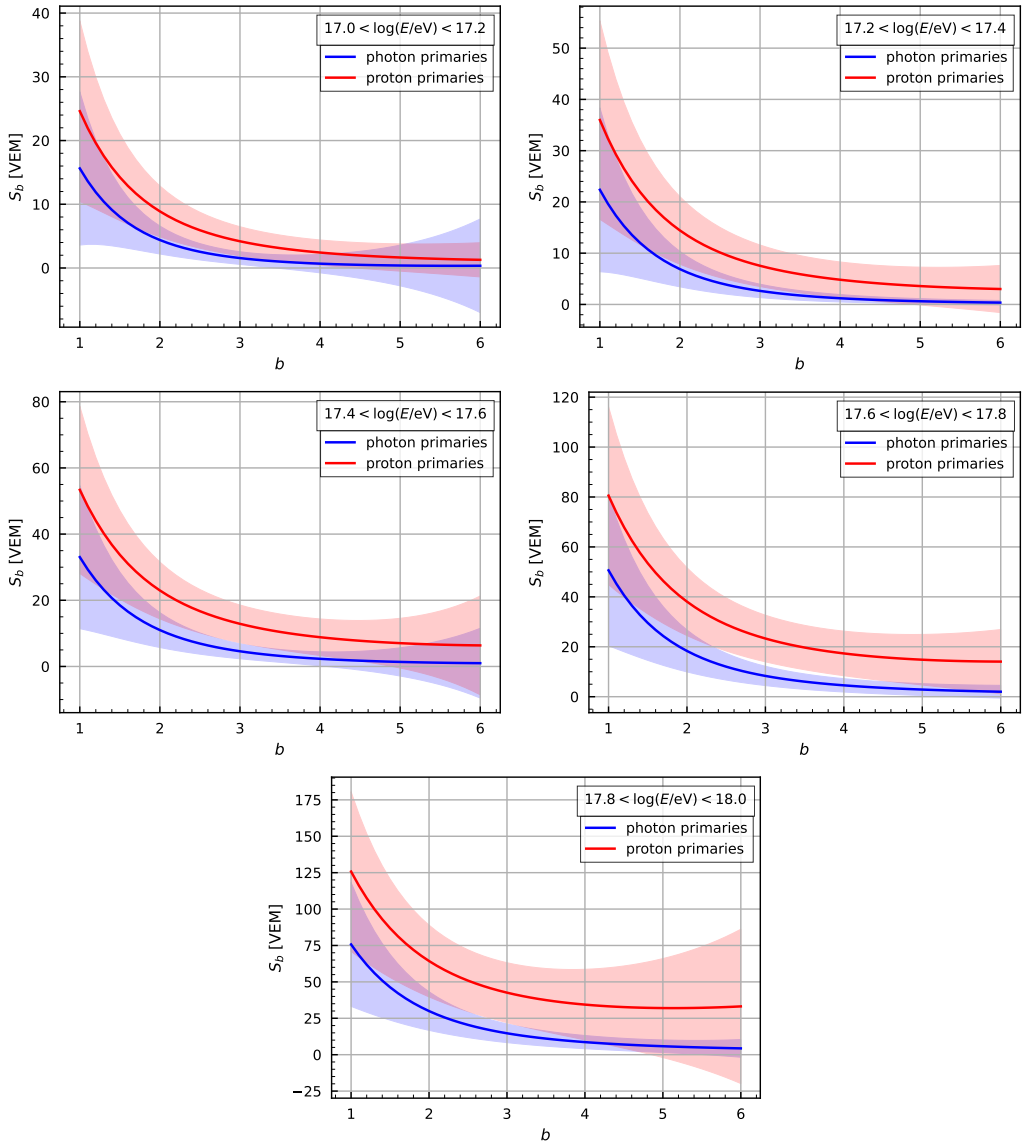

Fig. D.1 Mean values and standard deviations of the simulated S_b distributions for the two primaries in all five energy bins between 10^{17} and 10^{18} eV.

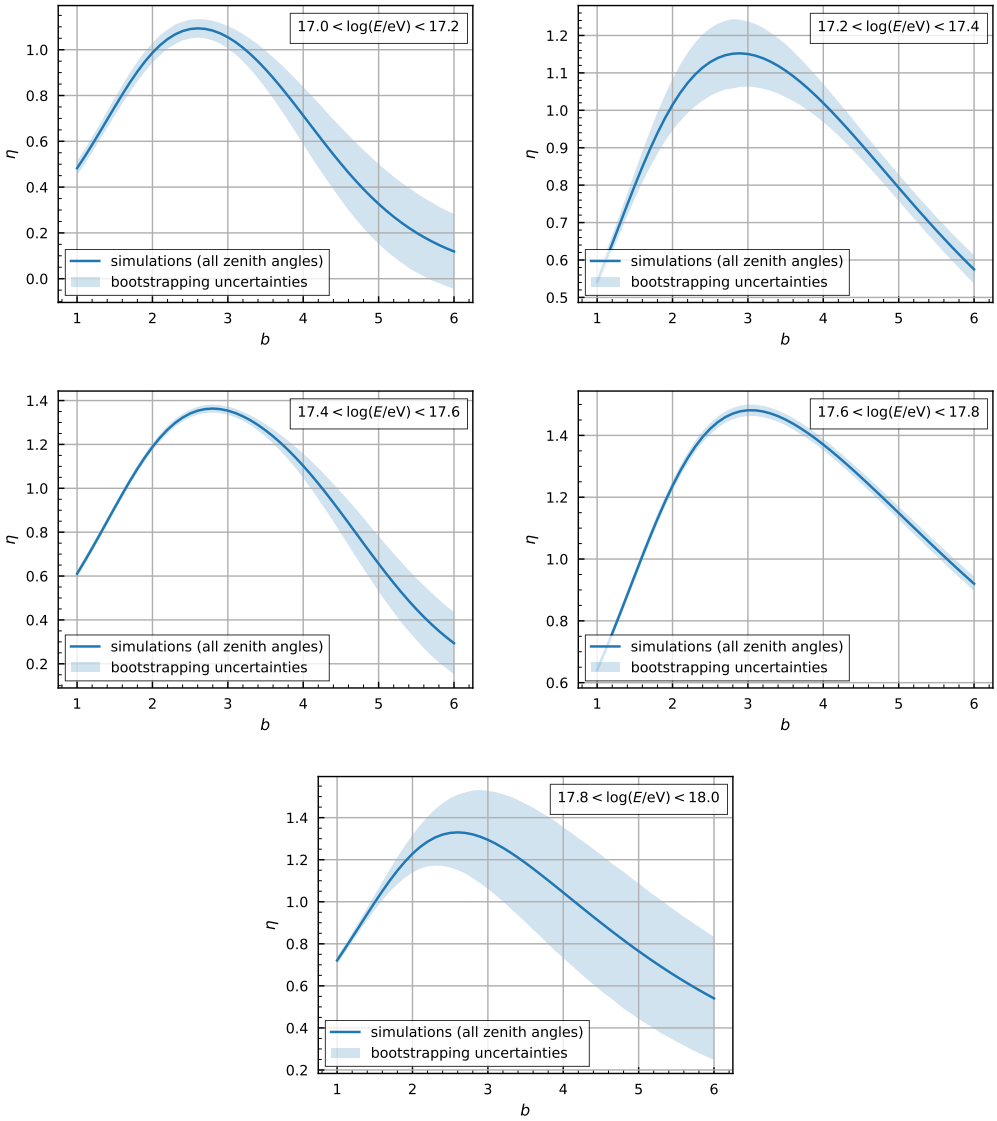

Fig. D.2 Merit factor of the simulated S_b distributions together with bootstrapping uncertainties in all five energy bins between 10^{17} and 10^{18} eV.

D.1.2. $\log(S_b)$

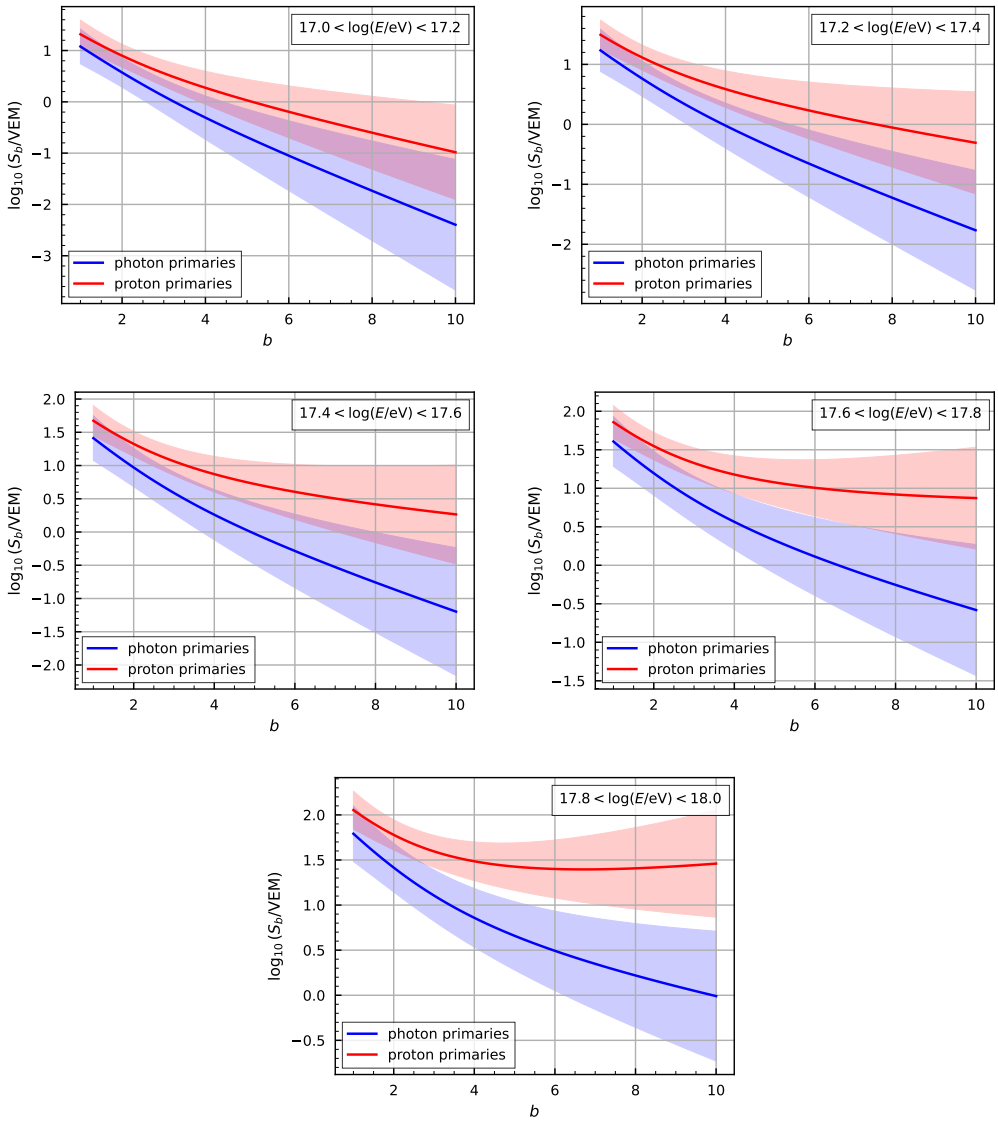

Fig. D.3 Mean values and standard deviations of the simulated $\log(S_b)$ distributions for the two primaries in all five energy bins between 10^{17} and 10^{18} eV.

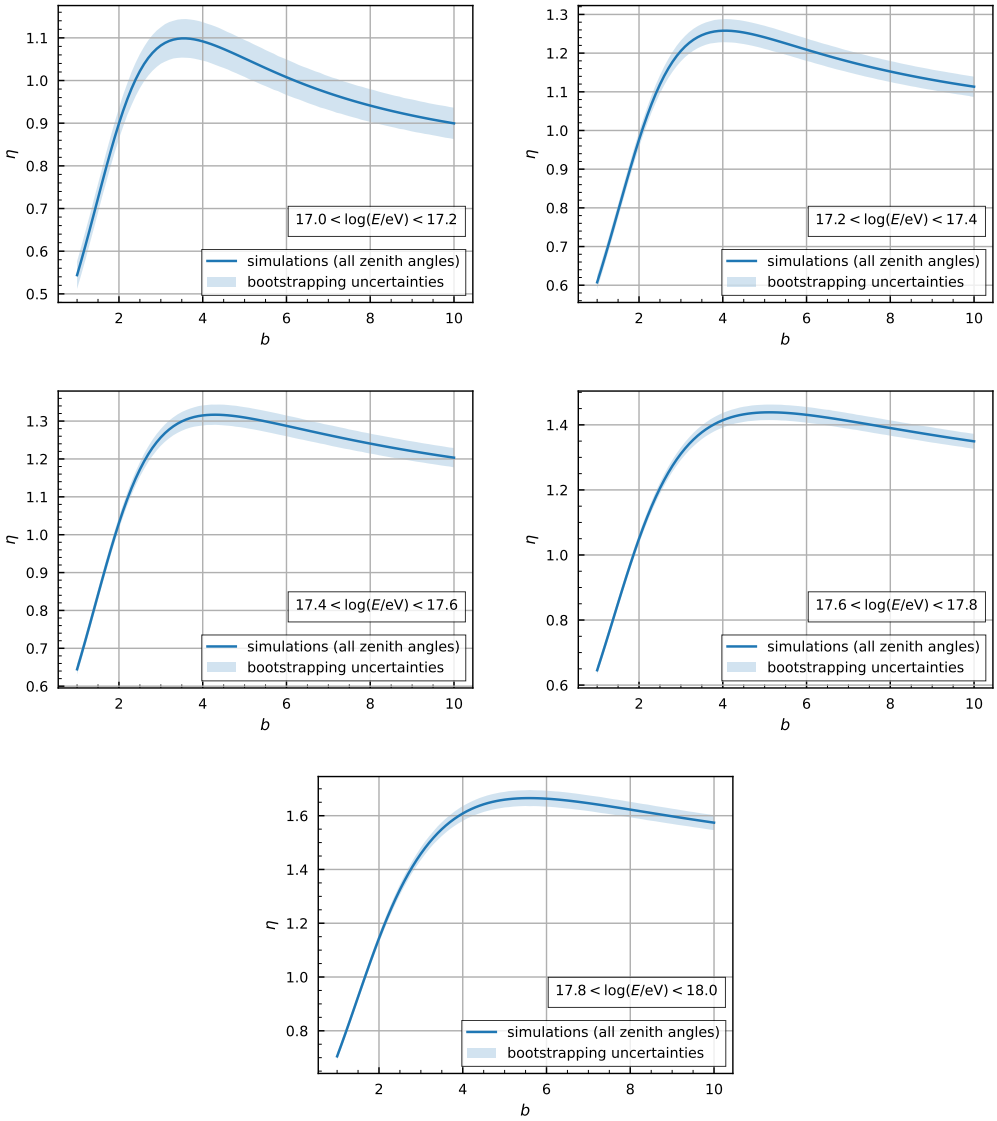

Fig. D.4 Merit factor of the simulated $\log(S_b)$ distributions together with bootstrapping uncertainties in all five energy bins between 10^{17} and 10^{18} eV.

D.1.3. Application in Multivariate Analysis

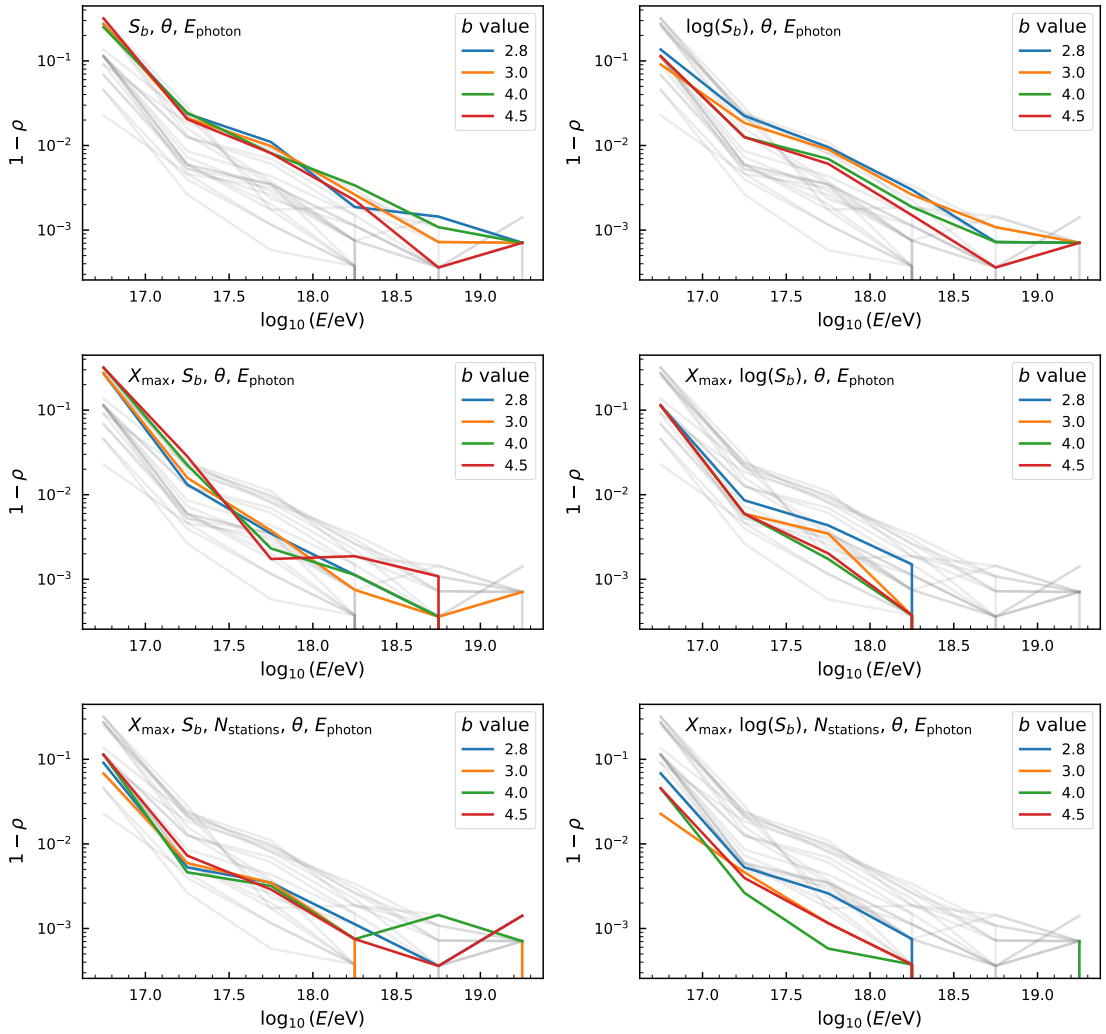

Fig. D.5 Background rejection ρ at 50% signal efficiency of BDT discriminant in various observable configurations measured in energy bins of width $\log_{10}(E/\text{eV}) = 0.5$. For presentation reasons $1-\rho$ was chosen for the vertical axis. The left column depicts combinations containing S_b, while in the right column instead of the former $\log(S_b)$ is used. Results of the other combinations are indicted by gray lines. Vertical lines are caused be the discrete counting nature of the metric. When all background events are rejected, the semi-logarithmic spacing is not able to indicate the value $\rho = 1$.

D.2. Additional Figures Regarding the Simulation Production

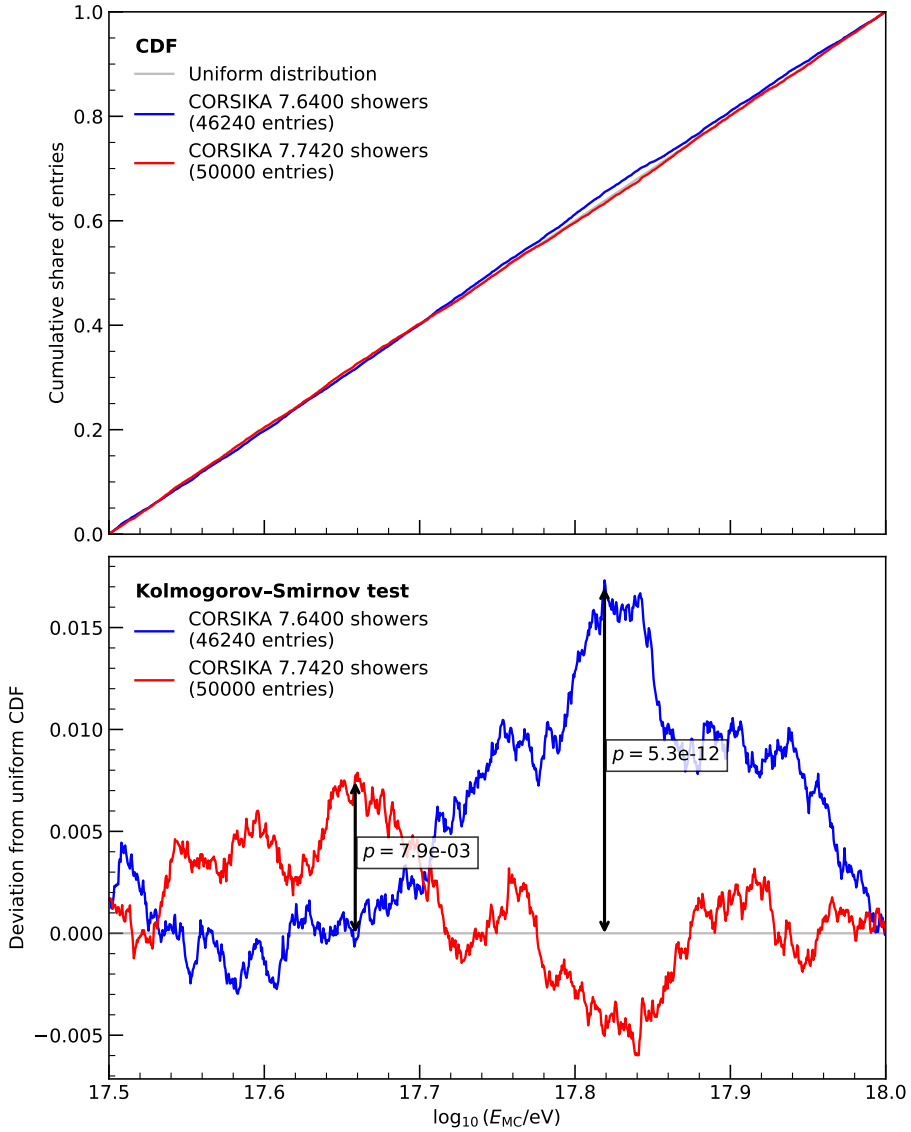

Fig. D.6 Distribution analysis of $\log_{10}(E_{\mathrm{MC}}/\mathrm{eV})$ for the presimulated photon air showers. The cumulative distributions are compared in the subplots in the top. The bottom subplot employs the Kolomogorov-Smirnov test to quantitatively evaluate these deviations. A p-value is calculated under the null hypothesis that the dataset conforms to the theoretical distribution, considering the maximal deviation within the specified range.

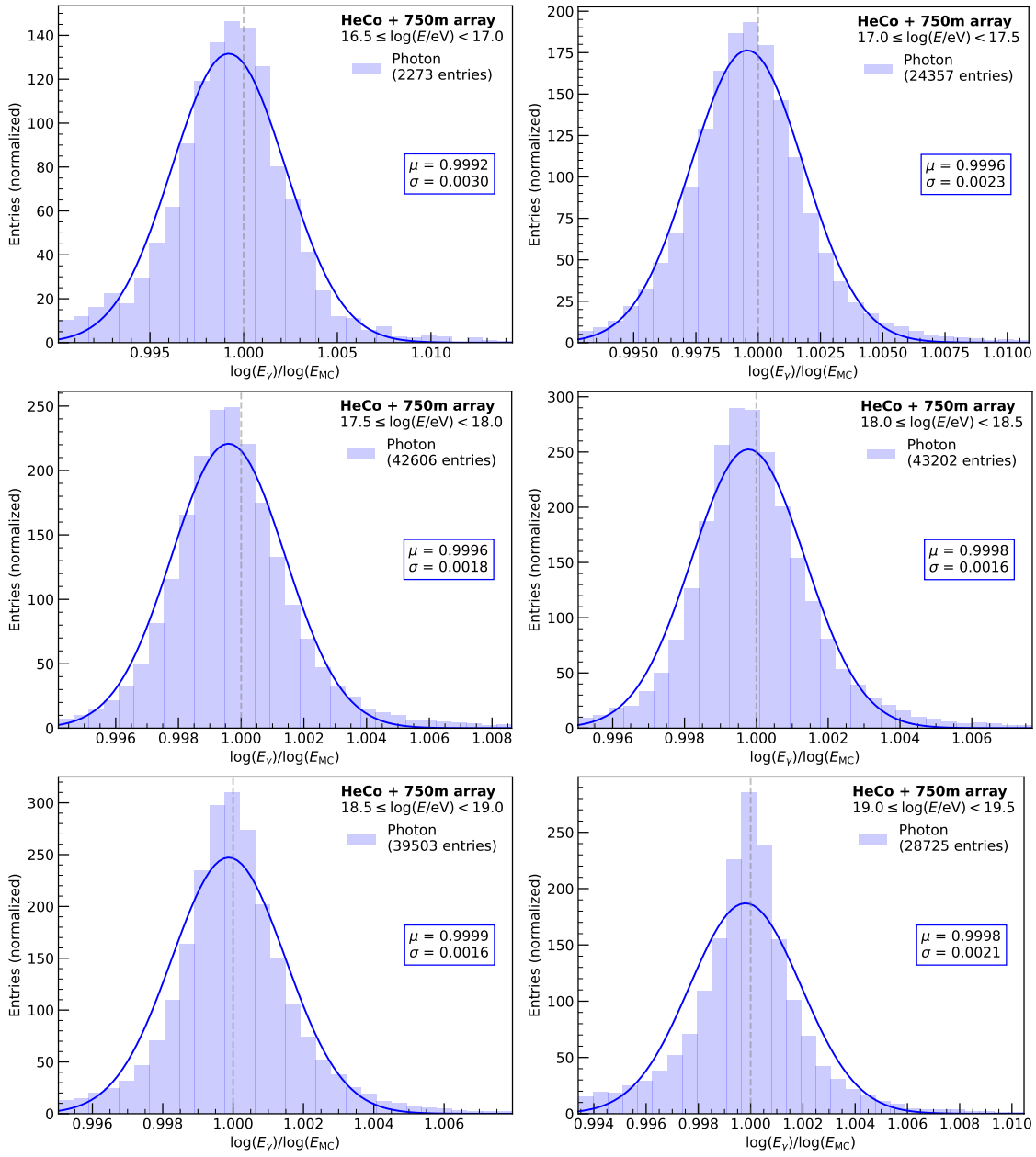

Fig. D.7 Performance of the reconstruction for the key quantity E_γ in the individual energy bins together with a rough approximation based on a normal distribution. Mean and width of the normal distribution are estimated from the median and the 15.9% resp. 84.1% percentile for robustness against outliers in the dataset.

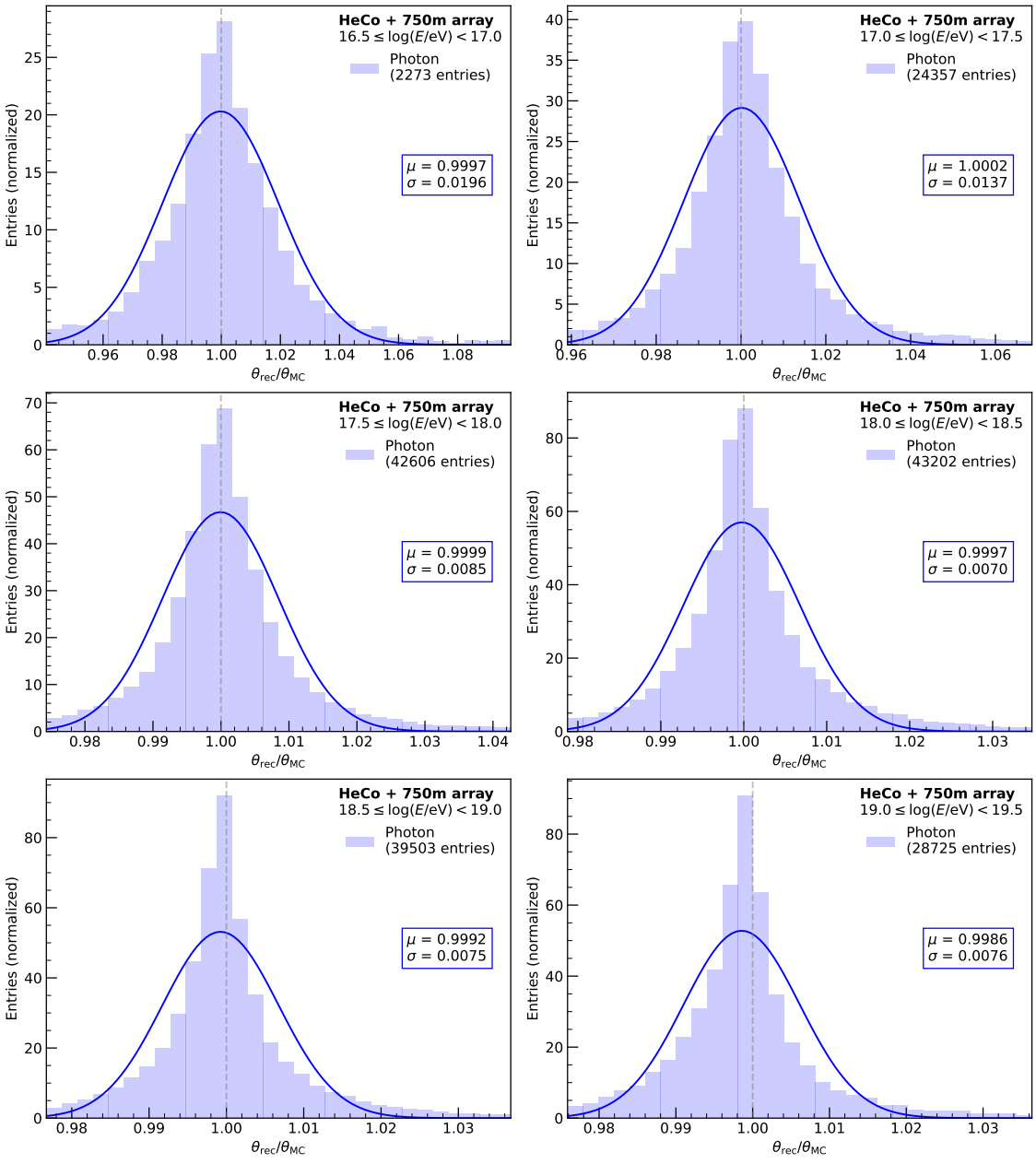

Fig. D.8 Performance of the reconstruction for the key quantity θ in the individual energy bins together with a rough approximation based on a normal distribution. Mean and width of the normal distribution are estimated from the median and the 15.9% resp. 84.1% percentile for robustness against outliers in the dataset.

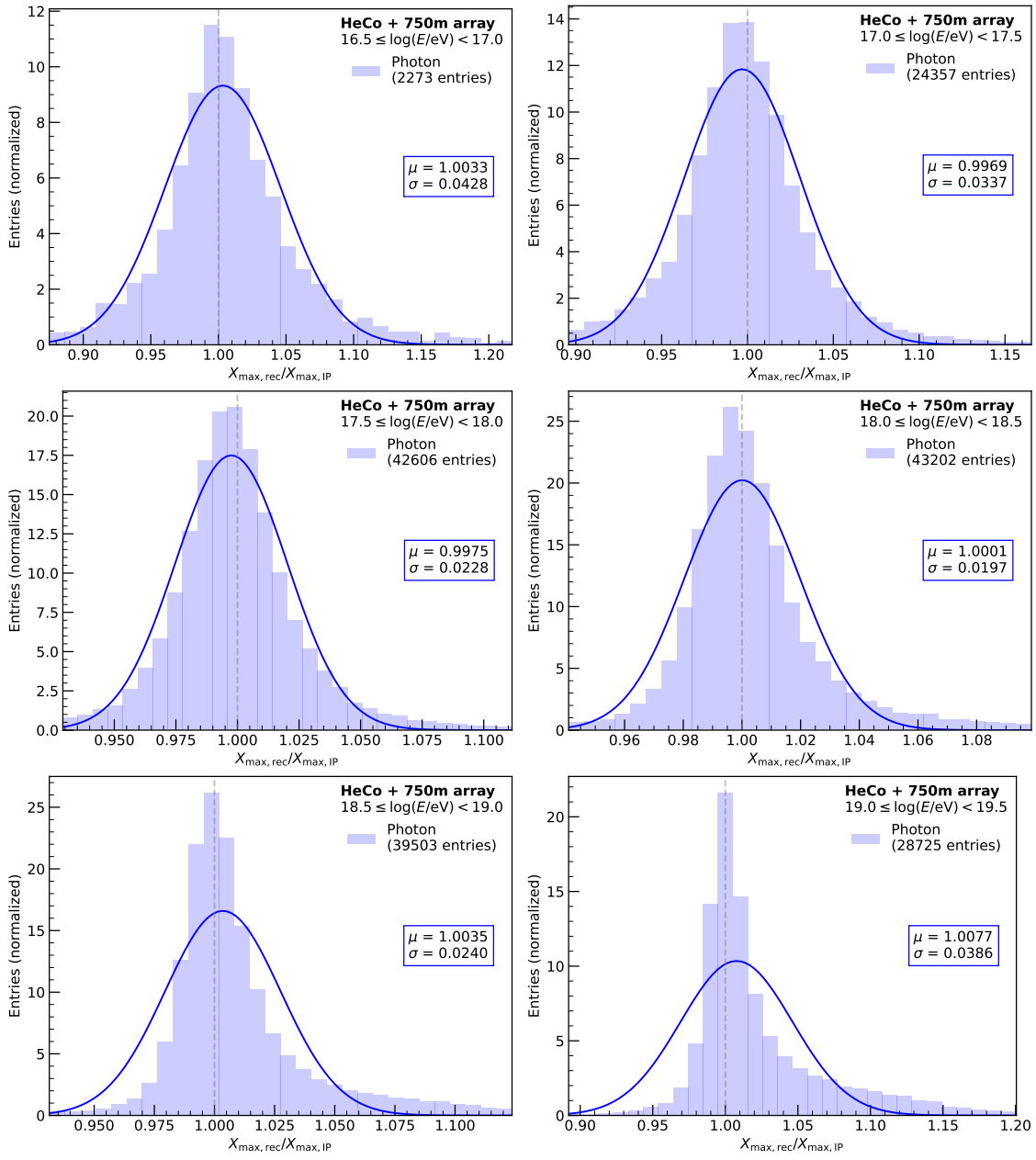

Fig. D.9 Performance of the reconstruction for the key quantity $X_{\max}$ in the individual energy bins together with a rough approximation based on a normal distribution. Mean and width of the normal distribution are estimated from the median and the 15.9% resp. 84.1% percentile for robustness against outliers in the dataset.

E. Acknowledgments

Mein Dank gilt einer Reihe von Personen, die die Erstellung dieser Masterarbeit überhaupt erst ermöglicht haben:

An erster Stelle ist hier *Prof. Dr. Markus Risse* zu nennen, der mich nach meiner Bachelorarbeit erneut in der Arbeitsgruppe Experimentelle Astroteilchenphysik aufgenommen hat, die notwendigen Ressourcen bereitstellte und sich immer wieder Zeit für Feedback und Anregungen nahm. Ich danke *Prof. Dr. Markus Cristinziani* dafür, dass er sich bereit erklärt hat die vermeintlich undankbare Aufgabe des zweiten Gutachters zu übernehmen.

Darüber hinaus gilt mein Dank *Dr. Marcus Niechciol*, der mir bei technischen Fragen und Problemen stets zur Seite stand und mit seinen Ideen, Anregungen und unseren regelmäßigen Besprechungen – auch über Kontinente hinweg – maßgeblich zum stetigen Fortschritt dieser Arbeit beigetragen hat. Ich bin dankbar für die vielen Einblicke in die Arbeit der Pierre-Auger-Kollaboration und den wissenschaftlichen Alltag jenseits des Physikstudiums, die ich in dieser Zeit von ihm erhalten habe.

Praktisch alle der hier aufgeführten Ergebnisse basieren auf Berechnungen und Simulationen, die auf der Recheninfrastruktur des lokalen SiMPLE-Clusters durchgeführt wurden. Ich danke *Dr. Wolfgang Walkowiak*, *Fabian Dünkel* und *Dr. Philip Rühl* für ihre Arbeit in der Bereitstellung und Administration, der sie sich mit großer Hingabe gewidmet haben.

Ferner danke ich allen Mitgliedern der Experimentellen Teilchen- und Astroteilchenphysik in Siegen, insbesondere denen mit welchen ich in dieser Zeit mein Büro geteilt habe, für die herzliche Aufnahme in ihre Mitte und die stets angenehme Arbeitsatmosphäre.

Weiterhin möchte ich mich bei den Personen bedanken, die mit ihrer gründlichen Kontrolle des Textes einiges zu dessen Qualität beigetragen haben.

Zu guter Letzt sind die Personen zu nennen, die mich auf ihre eigene Art und Weise im privaten Umfeld unterstützt haben: Dazu gehören meine Eltern, ohne deren Rückhalt und finanzielle Unterstützung ein Physikstudium in dieser Form sicherlich nicht möglich gewesen wäre, meine Freunde, die für die notwendige Ablenkung und Zerstreuung abseits der Physik sorgten, und schließlich auch *Marie*, die mit außerordentlicher Geduld meine vielen Stunden am heimischen Schreibtisch ertrug und nicht müde wurde, mich durch diese anstrengende Zeit zu begleiten.

F. Declaration of Authorship

Hiermit erkläre ich, dass ich die vorliegende Master-Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt, sowie Zitate und Ergebnisse Anderer kenntlich gemacht habe. Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

Ort, Datum

Unterschrift