FACHBEREICH C – MATHEMATIK UND NATURWISSENSCHAFTEN
BERGISCHE UNIVERSITÄT WUPPERTAL

Multivariate Search for a Directional Excess of EeV Photons with the Pierre Auger Observatory

Daniel Kümpel

DISSERTATION

zur Erlangung des Doktorgrades (Dr. rer. nat.)
des Fachbereichs C - Mathematik und Naturwissenschaften
der Bergischen Universität Wuppertal

März 2011

Diese Dissertation kann wie folgt zitiert werden:

urn:nbn:de:hbz:468-20110512-121703-5
[http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:hbz:468-20110512-121703-5]

1. Gutachter	2. Gutachter
Prof. Dr. Karl-Heinz Kampert (Bergische Universität Wuppertal)	Prof. Dr. Markus Risse (Universität Siegen)

Contents

Appendices

List of Figures

List of Tables

Chapter 1

Introduction

Even though you neither feel them nor see them, every second we are bombarded by thousands of ionized cosmic ray particles [1]. Most of them are protons, but also α-particles and heavier nuclei are among them. Fundamental questions arise:

- “Where do they come from?” and in particular
- “What is the acceleration mechanism to such high energies which have already been observed?”

Even today the answers to these questions are not fully understood. The measurement of the particle flux, elemental composition, arrival direction distribution and temporal variations are of central importance to get a clue of an answer. More insight to these questions would make a major break-trough in understanding the high energy universe and has already opened an entirely new field of research on its own.

The story of “astroparticle physics” started almost a century ago, when the Austrian physicist Victor Franz Hess discovered cosmic rays, charged particles that hit our atmosphere like a steady rain from space. Astrophysics together with particle physics has fundamentally changed our view of the universe. Although the term “astroparticle physics” has been widely accepted since only 15-20 years, the first triumph of the relatively new scientific field dates back to the seventies: the detection of solar neutrinos. Together with the detection of neutrinos from a supernovae in 1987, it marks the birth of neutrino astrophysics, acknowledged with the Nobel prize of physics in 2002. The enormous discovery potential of the field stems from the fact that attainable sensitivities are strongly improving in the previous two decades. But not this alone is arguably enough to raise expectations. We are entering territories with a high discovery potential, as predicted by theoretical models. For the first time we are able to tackle the aforementioned questions with the required sensitivity. One backbone of astroparticle physics are particle detectors, telescopes and antennas. The size of these instruments are generally large due to the scarcity of the signals that are to be detected and are instrumented in “open” media like water, ocean, ice or rarely populated area. They are operating e.g. at high altitudes and locations with small background from artificial light sources.

The present flagship in the search for ultra-high energy cosmic rays is the Southern Pierre Auger Observatory located in the Argentinean Pampa Amarilla. For the first time it combines two independent detection techniques. Surface detectors on the ground cover a huge area in order to detect and study secondary particles of extensive air showers. Another complementary technique utilizes the fact that shower particles excite nitrogen molecules on their passage through the atmosphere. The de-excitation proceeds partially through the emission of fluorescence light, which can be detected by telescopes at the ground. The synergy of these techniques is able to reduce systematic uncertainties, improves the event reconstruction and provides important cross-check information.

The search for photons at ultra-high energies is the main topic of this thesis. No detection has been reported so far. An observation would extend the already measured photon energies to several orders of magnitudes and open a new window in cosmic-ray research with significant impact on related fields. Current experimental results already set diffuse (i.e. with no pointing information) upper limits on the photon fraction referring either to ground array data or to fluorescence telescope information. The combination of both detection systems, and hence a synergy of surface array and fluorescence telescope observables in a multivariate analysis, is performed for the first time in this thesis. Furthermore, the directional information is utilized to search for photon sources and to set directional upper limits on the photon fraction.

The thesis is divided as follows. A brief introduction to cosmic ray physics and current experimental results with special attention to the highest energy part is given in Chapter 2. Since direct detection techniques can not be utilized at ultra-high energies, primary particles can only be measured indirectly via particles cascades. The development of these cascades in the atmosphere, the main components as well as prominent detection systems are introduced in Chapter 3. Chapter 4 gives an overview of the Pierre Auger Observatory and its main software framework. Special attention to ultra-high energy photons is given in Chapter 5. Here, production and propagation scenarios are discussed as well as distinguishing features of photon and hadron induced particle cascades. An important role in understanding the origin of cosmic rays is a detailed knowledge of particle propagation through the universe. In Chapter 6 the software tool CRPropa is used in a simulation study of particle propagation. It is shown that experimental limits on the photon fraction can be used to constrain Centaurus A as a strong proton source. Chapter 7 introduces fluorescence and surface detector observables to discriminate photon from hadron induced cascades. These observables are applied in a multivariate analysis described in Chapter 8. In Chapter 9 the obtained multivariate response output is adopted to real data measured at the Pierre Auger Observatory. A search for a localized excess is accomplished and directional upper limits on the photon fraction for sources are given. Finally, the most important results of this thesis are summarized in Chapter 10 together with an outlook for future analyses.

Chapter 2

Cosmic Rays

Since almost 100 years cosmic rays are exploited by thousands of physicists. Although much progress was made, the remaining issues are still numerous. This chapter summarizes the present status of cosmic ray research with special attention to the highest energy part. Each section recapitulates the topic very brief while giving references for a more profound understanding. The discovery of cosmic rays is sketched in Sec. 2.1 following observational results of the energy spectrum (Sec. 2.2), mass composition (Sec. 2.3) and anisotropy (Sec. 2.4). In Sec. 2.5 possible sources and acceleration scenarios are introduced and propagation of cosmic rays is covered in Sec. 2.6.

2.1 Introduction

At the end of the 19th century some scientists came to the conclusion that there was little more to do in physics than filling in a few more figures after the decimal point of various fundamental constants. They could not have been more wrong. Small variations in the expectation turned out to be crucial enough to roll up fundamental physics.
At this time it was already known that even perfectly insulated electrostatic devices would discharge themselves. It was realized that the gradual discharging of bodies could be explained if the air contained ionized particles. But where do those ions come from? The British physicist Charles Wilson carried out an, at this time, baffling experiment. He measured how quickly charge leaked away from a gold leaf electroscope and tried to find out the reasons for the discharge, but neither day/night variations nor different atmospheric conditions could cause any differences. He was forced to conclude that, in some way or another, ions were actually formed within the air in a sealed container at a rate that he could measure with equal amounts of positive and negative charge. It became

known as “spontaneous” ionization.
This spontaneous ionization had properties very similar to radiation from radioactive substances. In 1901, Wilson wondered whether the cause of the ionization might be radioactive rays from outside the Earth’s atmosphere, so he went into a Scottish railway tunnel to see if the ionization attenuates. Unfortunately, he did not realize that the discharge effect is affected not only by rays of particles penetrating the atmosphere but also by radioactivity in the Earth. His apparatus was not sufficient enough to separate these effects. He concluded that the source of ionization must be something in the air itself.

Figure 2.1: Victor Franz Hess

The crucial experiment started 10 years later at six o’clock in the morning of August 7, 1912, when the Austrian physicist Victor Franz Hess started a remarkable balloon ascent. In order to measure the ionization as a function of height he made his last trip of a series of seven balloon ascents. At that time, still, most of the ionization had been traced to radioactive impurities and deposits. Hess wanted to demonstrate with an improved electroscope, that the ionization in a hermetically sealed vessel reduces with increasing height due to the reduced affect of radioactive substances of the Earth [2], but he discovered a baffling result. Up to a height of about 1000 m the ionization decreased almost as expected, but then it increased and in roughly 3000 m height the ionization is as strong as it is on the Earth’s surface. He concluded, that the cause of that boost in ionization might be attributed to the penetration of the Earth’s atmosphere from outer space by hitherto unknown radiation of high penetrating capacity [3]. He discovered the *Cosmic Radiation*. 24 years later Hess shared[1] the Nobel price in physics “for his discovery of cosmic radiation”.

The current state of knowledge with special attention to the highest energy part is briefly summarized in the following sections.

2.2 Energy spectrum

Today we know that our galaxy and accordingly our own solar system is permanently exposed to a flux of highly energetic particles — the cosmic rays. Their energies were measured over a remarkably large energy range starting from the MeV region to at least 10^{20} eV. The amount of cosmic rays in a certain energy band is represented by the energy spectrum. It can be approximated by an inverse power low in energy with an differential flux given by

$$\frac{\mathrm{d}N}{\mathrm{d}E} \propto E^{-\gamma} \; , \tag{2.1}$$

indicating non-thermal acceleration processes (cf. Sec. 2.5.2). Most regions of the spectrum are rather featureless with a constant spectral index γ but, however, small structures become clearly visible when multiplying the flux by some power of the particle energy as

---

[1] together with Carl David Anderson for the discovery of the positron

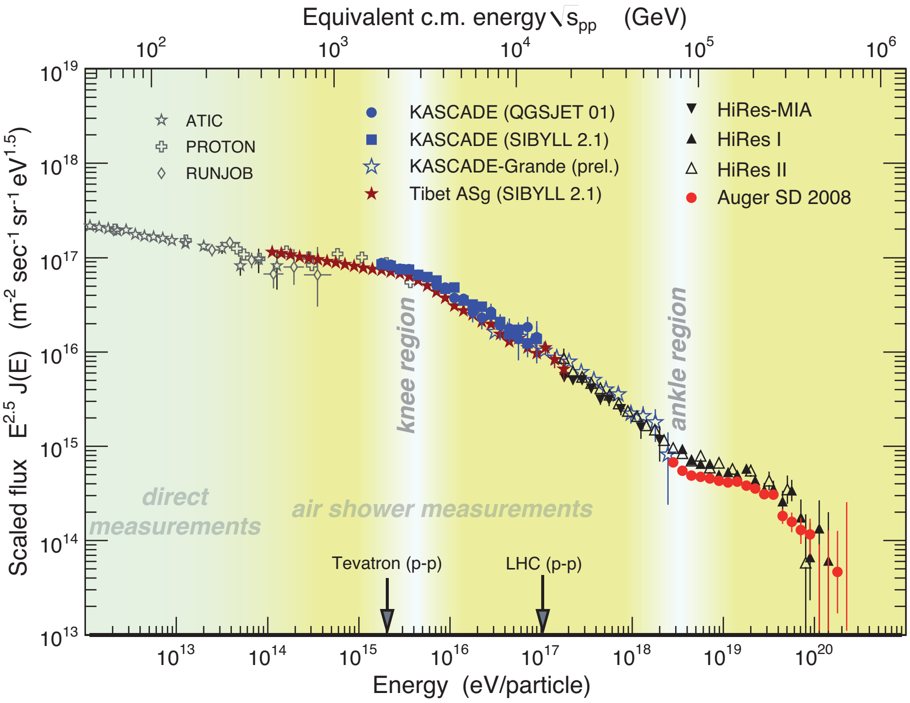


Figure 2.2: Primary cosmic ray flux multiplied by $E^{2.5}$ measured by different experiments, cf. [4]. Results from direct measurements above the atmosphere are shown from the ATIC [5], PROTON [6] and RUNJOB [7] experiment. The observed flux of selected air shower experiments (KASCADE [8], KASCADE-Grande [9], Tibet ASγ [10], HiRes-MIA [11], HiRes [12] and Pierre Auger [13]) is superimposed. The equivalent center of mass energy as well as the maximum energy of Tevatron and Large Hadron Collider (LHC) are indicated.

illustrated in Fig. 2.2. Up to an energy of $E \cong 3 \cdot 10^{15}$ eV the spectral index is $\gamma \approx 2.7$ whereas above this energy a steepening to $\gamma \approx 3.1$ is observed. This region is called the "*knee*" and was first deduced from observations made by Kulikov and Khristianson *et al.* in 1958 [14]. The position of the knee is dependent on the particle type and is interpreted as the spectral region where galactic sources fail to accelerate lighter elements to higher energies and only do so for heavier elements [15, 16, 17]. First discovered 1963 by Linsley, the particle spectrum reveals also an additional structure at about $\sim 4 \times 10^{18}$ eV known as "*ankle*" [18, 19, 20, 21, 22]. The origin of the ankle has been traditionally attributed to the transition from the galactic component of the cosmic ray flux to a flux dominated by extragalactic sources. A model for that is proposed e.g. by Hillas [23]. However, in recent years it became clear that a similar feature could also be explained by the propagation of protons from extragalactic sources placing the transition from galactic to extragalactic rays at much lower energies. Here the ankle is produced by the modification of the source spectrum of primary protons caused by $e^{\pm}$ pair production of protons with the photons of the cosmic microwave background radiation (cf. Sec. 2.6.2) during propagation through interstellar space. This idea is also known as "dip"-model and was first proposed by Berezinsky *et al.* [24] assuming an almost pure proton composition.

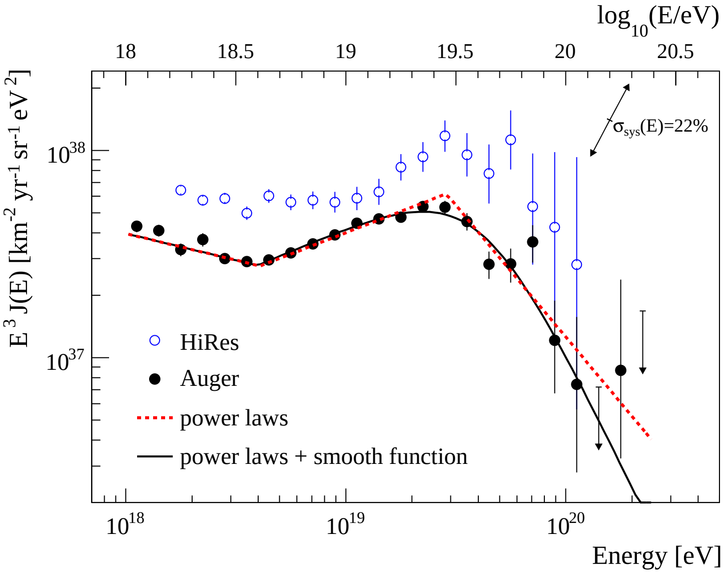


Figure 2.3: Combined energy spectrum at highest energies scaled with E^3 (cf. [25]). The measured flux of the Pierre Auger Observatory [25] is compared to the date from the HiRes Experiment [26].

At the highest energies the cosmic ray flux decreases from $\sim 10^3$ $m^{-2}s^{-1}$ at few GeV to ~ 1 km^{-2} per century at 10^{20} eV. While below the ankle a spectral index of $\gamma \approx 3.3$ is observed a flattening to $\gamma \approx 2.6$ is found above the ankle [22]. However, when comparing the power law extrapolation to actual measurements it is found that the spectrum is suppressed by a factor of two at $\log_{10}(E_{1/2}/\text{eV}) = 19.61 \pm 0.03$, cf. Fig. 2.3. The significance of this suppression is more than 20σ [25] and is similar to what is expected from the GZK effect which will be described in more detail in Sec. 2.6.2, but could also partly be related to a change of the shape of the average injection spectrum at the sources [27].

2.3 Mass composition

Certainly, one scientifically most relevant piece of information are precise data on the chemical composition of the primary cosmic ray flux as a function of energy. At energies below 10^{14} eV the abundance of individual elements has been measured with detectors above the atmosphere. In comparison to the composition of stellar material in our solar system the differences are quite small as shown in Fig. 2.4. However, some elements indicate larger differences which are very important. For light elements there is an overabundance of Hydrogen and Helium for solar system abundances. Lithium, Beryllium and Boron are overabundant in cosmic rays. Iron agrees quite well with solar system composition, but there is an excess of elements slightly lighter than Iron. One way to understand the overabundances of cosmic rays is to assume that cosmic rays have the same composition as solar matter at their origin. Propagating through the interstellar space they can interact with gas and dust particles, which results in heavier nuclei spallating

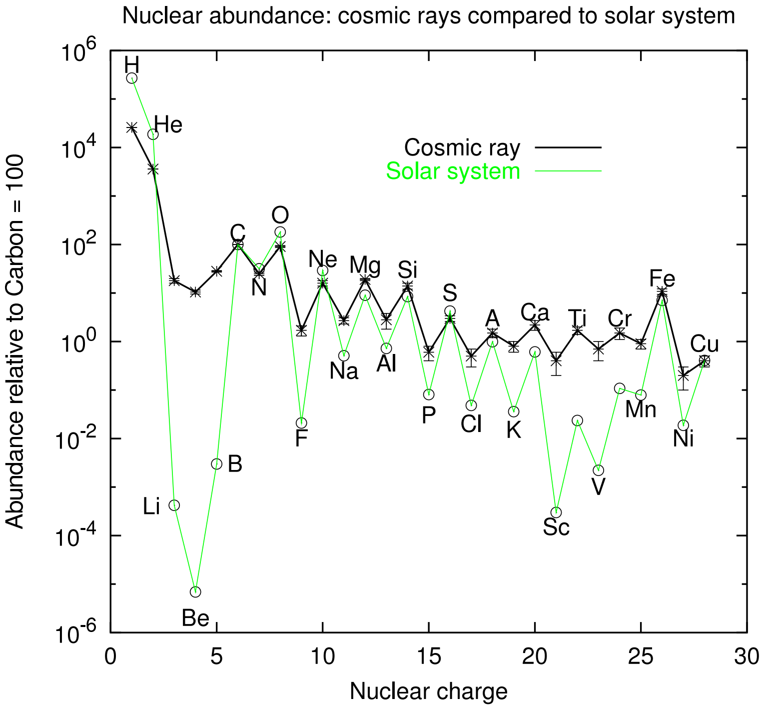


Figure 2.4: Relative abundances of solar [28] and cosmic ray material [29, 30, 31, 32] for low energy cosmic rays (cf. [33]) normalized to Carbon.

into lighter nuclei.

From the knowledge of the spallation cross sections obtained in accelerator experiments one can learn something about the amount of matter traversed by cosmic rays between production and observation. For the bulk of cosmic rays the average amount[2] of matter traversed is of the order $X = 5\ \mathrm{g/cm}^2$ to $X = 10\ \mathrm{g/cm}^2$ (cf. [34]). Furthermore, the density ρ_N of the galactic disc can be approximated to one proton per cm^3. With the proton mass $m_p = 1.67 \cdot 10^{-24}$ g one can calculate the corresponding thickness L of the material to

$$L = \frac{X}{m_p \rho_\mathrm{N}} = 3 \times 10^{24}\ \mathrm{cm} \cong 1\ \mathrm{Mpc}. \tag{2.2}$$

The diameter of the galactic plane is ≈ 30 kpc so one could conclude, that low energy cosmic rays propagate on a very winding way through our galaxy. The resulting lifetime τ is

$$\tau = \frac{L}{v} \approx 3 \times 10^6\ \text{years}. \tag{2.3}$$

Methods of radioactive dating [33] indicate $\tau \approx 2 \times 10^7$ years. This relative large discrepancy implies that cosmic ray nuclei spend also significant time diffusing in low density galactic halo regions before escaping into intergalactic space.

---

[2] Note that the amount is energy dependent

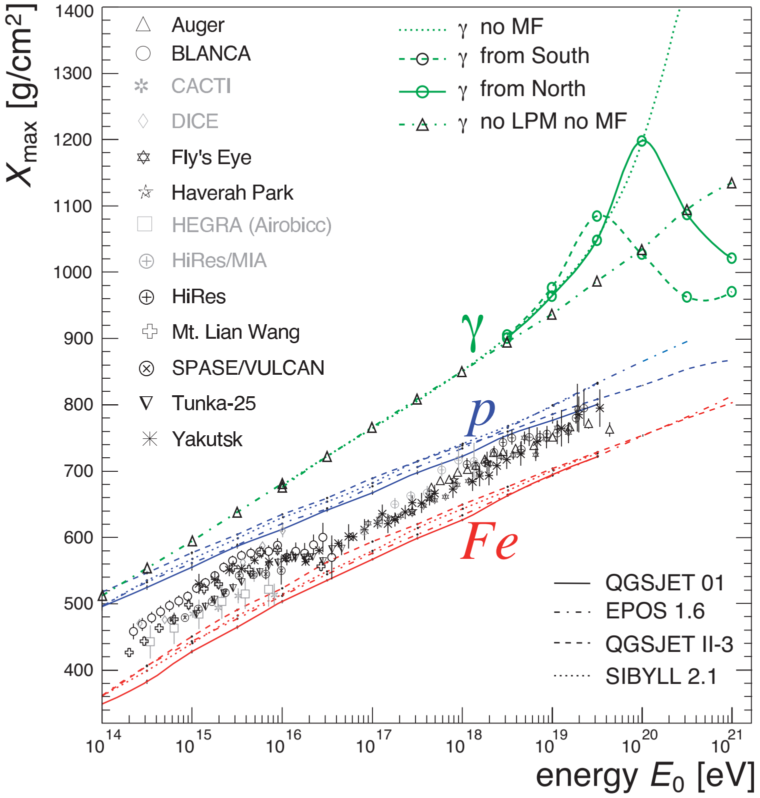


Figure 2.5: Average depth of shower maximum X_{max} as a function of primary energy E_0 modified from [4]. Experimental results from Auger [35], BLANCA [36], CACTI [37], DICE [38], Fly's Eye [39], Haverah Park [40], HEGRA [41], HiRes/MIA [42], HiRes [43], Mt. Lian Wang [44], SPASE/VOLCAN [45], Tunka-25 [46] and Yakutsk [47]. The results are compared to predictions of the average depth of shower maximum for primary photons (green), protons (blue) and iron (red). Different interaction models were used namely QGSJET 01 [48], EPOS 1.6 [49], QGSJET II-3 [50] and SIBYLL 2.1 [51] as well as modifications in the magnetic field (MF) and conversion processes for primary photons (cf. Sec. 5.4).

The decreasing cosmic ray flux at energies above 10^{14} eV makes it inevitable to measure properties of primary cosmic rays via secondary particles produced when interacting with the atmosphere of the Earth. The physics and detection techniques of so-called Extensive Air Showers (EAS) is described in more detail in Chapter 3. As a result of large fluctuations in the shower development an often-used quantity to characterize the composition is the mean logarithmic mass, defined as

$$\langle \ln A \rangle = \sum r_i \ln A_i \ ,$$

where r_i is the relative fraction of nuclei i with atomic mass number A_i. In an air shower experiment $\langle \ln A \rangle$ is obtained applying two methods:

1. The quantity is proportional to the ratio of the number of electrons and muons

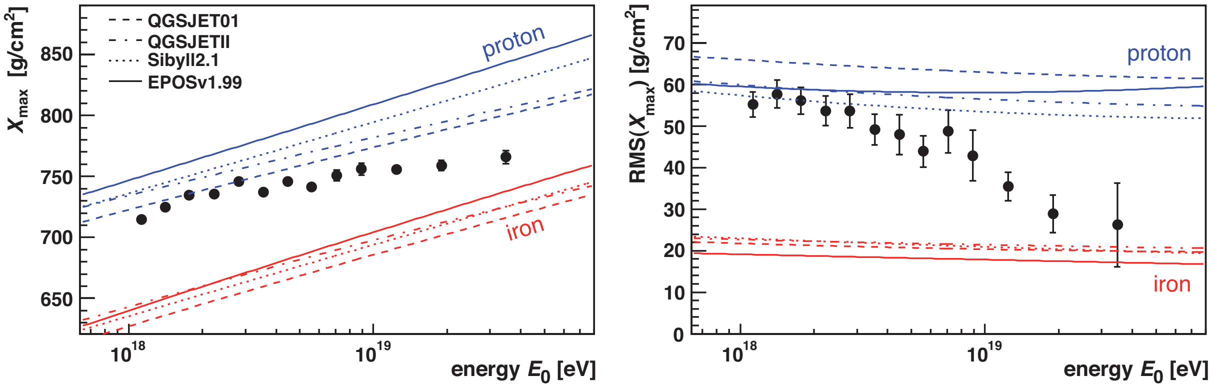


Figure 2.6: Mean X_{max} and RMS(X_{max}) as a function of primary energy E_0 [53]. The simulation predictions of primary protons and iron using different interaction models [54, 55, 56, 57] are indicated.

registered at ground level

$$\langle \ln A \rangle \propto \log_{10}\left(\frac{N_\mu}{N_e}\right)$$

2. $\langle \ln A \rangle$ is inverse proportional to the observed depth of shower maximum $\langle \ln A \rangle \propto 1/X_{\mathrm{max}}$ (cf. Sec. 3.1).

Experimental results of the average X_{max} as a function of primary energy E_0 are shown in Fig. 2.5. Superimposed are model predictions of the average depth of shower maximum for primary photons (green), protons (blue) and iron (red) using different interaction models and magnetic field assumptions. Although significant differences between models are visible for the absolute value, common trends can be observed. Below $\sim 4 \cdot 10^{15}$ eV the experimentally measured elongation rate[3] is larger than the one expected from simulations, which implies that the average composition would become lighter as a function of energy. Above the knee at about $E \gtrsim 4 \cdot 10^{15}$ eV up to $\sim 4 \cdot 10^{16}$ eV the elongation rate is ~ 0 indicating an increase of mass which could be explained from sequential breaks in the energy spectra for individual elements from light to heavy nuclei, cf. [8, 52]. Above $\sim 4 \cdot 10^{17}$ eV the elongation rate exhibits a rather constant value slightly larger than the one predicted from simulations indicating a gradual chance in composition towards lighter elements.

The composition at highest energies still remains a mystery. The Pierre Auger Observatory revealed a correlation between the arrival directions of ultra-high energy cosmic rays (UHECR) and positions of active galactic nuclei (AGN) [58, 59] (cf. Sec. 2.4). This perhaps indicates a lighter composition since heavier nuclei are more affected by magnetic fields (cf. 2.6.1) which is consistent with other experimental results, e.g. from HiRes [60]. However, recent measurements of the depth of shower maximum X_{max} of air showers seem

---

[3] The difference in $\langle X_{\mathrm{max}} \rangle$ when changing the primary energy E_0 by a factor of 10: $D_{10} = \langle X_{\mathrm{max}}(10 \cdot E_0) \rangle - \langle X_{\mathrm{max}}(E_0) \rangle$.

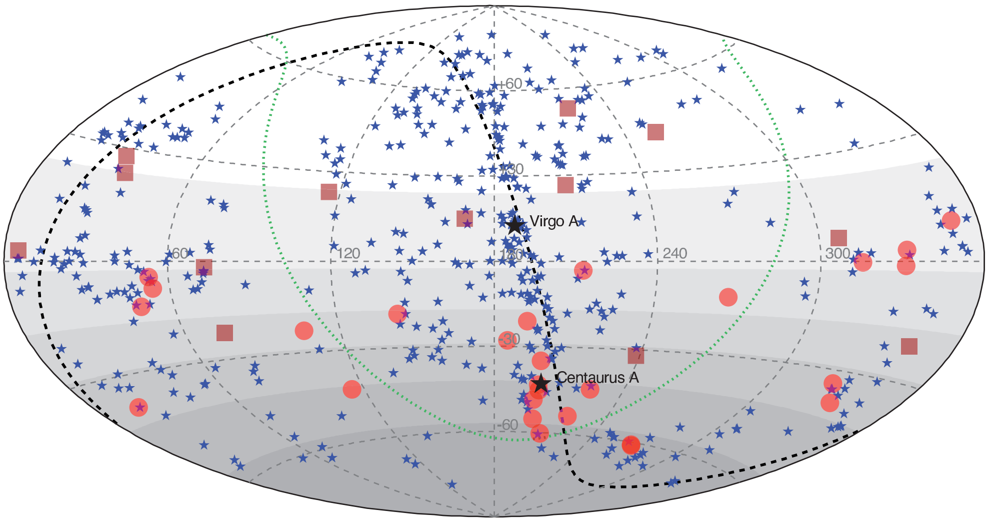


Figure 2.7: Sky map of nearby ($D_{\mathrm{max}} < 75$ Mpc) AGN (blue asterisks) from the Véron-Cetty & Véron catalog [61] in equatorial coordinates. Superimposed are the arrival directions of the highest energy cosmic rays observed by the Pierre Auger Observatory (red circles) [58] and the HiRes telescopes (red squares) [62]. The shaded area indicates the relative exposure of the Auger data set. The dashed black and dotted green line represents the supergalactic plane and the galactic disk, respectively. The position of Virgo A and Centaurus A is indicated, modified from [4].

to indicate a gradual increase of the average mass of cosmic rays assuming a realistic description of hadronic interactions at ultra-high energy as shown in Fig. 2.6 [53].

2.4 Anisotropy

Anisotropies in the arrival direction of cosmic rays are clearly of great interest to identify possible source regions or point sources. However, charged particle astronomy is, due to large uncertainties in the galactic and extragalactic magnetic field structure, still a challenge. Considering a particle with charge Z and energy E in PeV the Larmor radius r_L in pc can be approximated by

$$r_L[\mathrm{pc}] = 1.1[\mathrm{Am}^{-1}]\frac{E[\mathrm{PeV}]}{Z \cdot B[\mu\mathrm{G}]} \, , \tag{2.4}$$

with the magnetic field B in μG. A primary proton of energy 1 PeV in a galactic field of 3 μG has a Larmor radius of ~ 0.4 pc. With a diameter of the Milky Way of ~ 30 kpc it is not expected to find any point sources of charged cosmic rays. The situation changes for highest energies. In 1998 the AGASA collaboration found an excess of showers around 10^{18} eV coming from the Galactic Center and the Cygnus region [63]. Also SUGAR data confirmed an excess from a similar region [64]. An analysis of Auger data, however, does not support the previous findings of localized excesses [65].

A major breakthrough was published in 2007 when the Pierre Auger Collaboration introduced a correlation analysis between active galactic nuclei (AGN) and cosmic rays [58, 59, 66]. A prescription was set up to verify or reject the correlation hypothesis using an independent data set. The arrival directions of the highest energy cosmic rays above $E_{\mathrm{th}} = 5.7 \cdot 10^{19}$ eV are found to be correlated with the positions of nearby AGN ($z_{\max} < 0.017 \approx D_{\max} < 75$ Mpc) from the Véron-Cetty & Véron catalog [61] within an angular window of $\psi = 3.2°$. A sky map of the measured arrival directions as well as the position of AGN are shown in Fig. 2.7. Given the limited statistics one can not conclude from the found correlation that AGN are indeed sources of UHECR. AGN could also act as tracers for large scale distributions of luminous matter as described e.g. in [67]. Unequivocal source identification requires a larger data set in particular exploiting the fact that angular departures of the events from an individual source due to magnetic deflections should decrease in inverse proportion to the energy of the cosmic ray [59]. Although presently still only a hope, the prospects for charged particle astronomy are bright [68].

2.5 Origin of ultra-high energy cosmic rays

A major puzzle ever since the discovery of cosmic rays almost 100 years ago has been their exact origin. Particles with energies exceeding 10^{20} eV have already been observed [69, 70], which shows that there have to exist very powerful sites of UHECR creation in the universe. Since the magnetic fields of the Milky Way are not strong enough to confine particles above the knee ($\sim 10^{16}$ eV) it is plausible that their origin is outside the galaxy, whereas galactic sources are responsible for the lower-energy part. There are basically two fundamental approaches which explain how UHECR gain their energy:

- **Top-down models**: These models investigate the possibility of the decay of super-massive or high energy particles into UHECR. As they contain only a fraction of the energy corresponding to the energy of the primary particle these scenarios are called top-down models, cf. Sec. 2.5.1.

- **Bottom-up models**: These are theories where particles are accelerated from low energies to high energies by some external process. The most prominent models are briefly discussed in Sec. 2.5.2.

2.5.1 Top-down models

Top-down scenarios are motivated by theoretical models that introduce super-massive relict particles, generally called X particles with mass $m_{\mathrm{X}} > 10^{20}$ eV originating from high energy processes in the early universe. These particles typically decay to quarks and leptons. The quarks produce jets of hadrons containing mainly light mesons (pions) with a small percentage of baryons (mainly nucleons) [71]. The pions decay into photons (γ), neutrinos ($\nu, \overline{\nu}$) and electrons ($e^{\pm}$). Thus, energetic photons, neutrinos and charged leptons, together with a small fraction of nucleons, are produced directly with energies

up to $\sim m_{\mathrm{X}}$ without any acceleration mechanism. Prominent models are e.g. topological defects (TD) [72, 73], super heavy dark matter (SHDM) e.g. [74, 75], QCD fragmentation e.g. [76] or the Z-burst model (ZB) [77, 78, 79].
However, all these models predict a relative large fraction of ultra-high energy photons which is already strongly constrained by experimental limits on the photon fraction [80, 81, 82], cf. Sec. 5.6. In this regard top-down models are disfavored in comparison to bottom-up scenarios.

2.5.2 Bottom-up models

There are basically two types of mechanisms for bottom-up cosmic ray production:

1. The particles are directly accelerated to high energy by an extended **electric field**. This theory goes back to 1933 when W. F. Swann made the first plausible suggestion of how cosmic ray energies might be attained [83]. The acceleration is induced by changing magnetic fields near the surface of the sun and stars. It has been known that magnetic fields of up to several kilo-Gauss are associated with sunspots, which may appear and disperse over a period of days or weeks at the sun's surface. So so called "one-shot" mechanisms have been worked out in great detail and the electric field is now generally associated with the rapid rotation of small, highly magnetized objects such as pulsars or active galactic nuclei (AGN). Although it is quite fast, this mechanism is not widely favored these days, because it suffers from the circumstances, that the acceleration occurs in astrophysical sites of very high energy density, where the cross section for energy loss processes are high. Another reason is, that the theory can not explain the observed power law spectrum.

2. The particles are accelerated in a **stochastic** way. These models go back to E. Fermi in 1949 when he proposed an acceleration mechanism, in which particles gain energy gradually by numerous encounters with moving magnetized plasma [84]. However, this mechanism is slow compared to the electric field acceleration, and it is hard to keep the particles confined within the Fermi engine. More details on Fermis idea and possible enhancements can be found in the next section. Neglecting the details of the exact stochastic acceleration mechanism one can estimate an upper limit for the energy to which these sources can accelerate particles by the probability to escape from the acceleration region. If the Larmor radius r_L of the particle is of the order of the size of the acceleration site, particles can hardly be confined in the acceleration region. The maximum energy $E_{\max}$ is related to the field strength B in the source and the size R of the source region by [23]

   $$E_{\max} \approx \beta_s c \cdot Ze \cdot B \cdot R \,, \tag{2.5}$$

   where $\beta_s c$ is the shock velocity and Ze the particle charge. A. M. Hillas [23] was the first one who presented a graphical interpretation of that correlaton for various astrophysical sites as shown in Fig. 2.8.

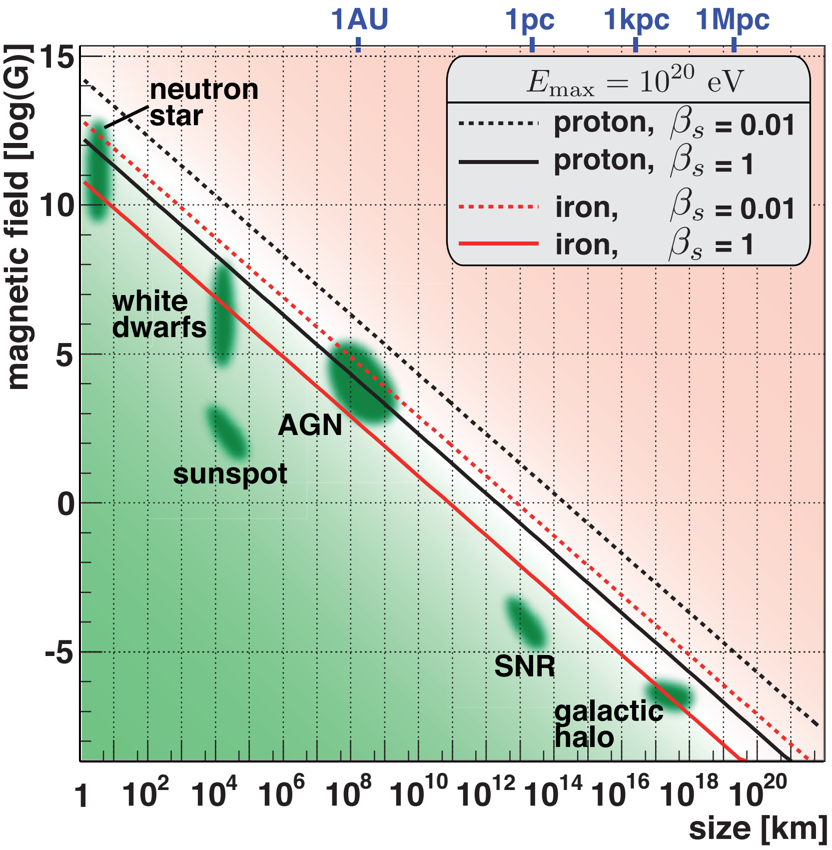


Figure 2.8: Size and magnetic field of possible sites of particle acceleration. Objects below the diagonal line can not accelerate protons (black) and iron (red) to $E_{\max} = 10^{20}$ eV. Two different shock velocities β_s (in units of c) are indicated.

Fermi mechanism

The basic idea is that cosmic ray particles traverse interstellar space and collide with large objects (like magnetized clouds), which move with random velocity and direction. Depending on the exact relative motion between particle and cloud, the cosmic ray can either lose or gain energy.

Consider a test particle which increases its energy E by an amount $\Delta E = \xi E$ proportional to its energy per "encounter" with a magnetic cloud. Let E_0 be the energy of injection. After n encounters the energy E_n is

$$\begin{aligned} E_n &= E_0(1+\xi)^n \\ n &= \frac{\ln(E_n/E_0)}{\ln(1+\xi)} \; . \end{aligned}$$

Let P_{esc} be the probability for a particle to escape from the region, that is occupied by magnetic clouds, after one encounter. The probability for a particle to remain in the acceleration region after n encounters is $(1-P_{\text{esc}})^n$. Clearly, the number of particles that remain longer in the cloud (and gain more energy) is proportional to the number of particles that remain in the acceleration region for more than n encounters. The proportion of particles accelerated to energies greater than E_n is

$$N(>E_n) = N_0 \sum_{m=n}^{\infty} (1 - P_{\mathrm{esc}})^m \propto \frac{1}{P_{\mathrm{esc}}} \left(\frac{E_n}{E_0} \right)^{-\gamma}$$

with

$$\gamma = \frac{\ln\left(1/(1-P_{\mathrm{esc}})\right)}{\ln(1+\xi)} \approx \frac{P_{\mathrm{esc}}}{\xi} \ .$$

The result is that stochastic acceleration leads to power law energy spectra.

- **Second order Fermi acceleration**
  The basic idea dates back to 1949, when Enrico Fermi proposed an acceleration mechanism for cosmic rays [84]. The acceleration relates to the amount of energy gained during the motion of a charged particle in the presence of randomly moving magnetized clouds ("magnetic mirrors"). Fermi argued, that the probability for a head-on collision is greater than a head-tail collision, so particles would, on average, be accelerated. Assuming a cosmic ray particle entering into a single cloud with energy E_i and incident angle θ_i with the cloud's direction, it undergoes diffuse scattering on the irregularities in the magnetic field. The energy gain of the particle, which emerges at an angle θ_f with energy E_f, can be obtained by applying Lorentz transformations between laboratory frame (unprimed) and cloud frame (primed):

  $$E'_i = \Gamma E_i (1 - \beta \cos\theta_i) \qquad (2.6)$$

  $$E_f = \Gamma E'_f (1 - \beta \cos\theta_f) \ , \qquad (2.7)$$

  where Γ and $\beta = V/c$ are the Lorentz factor and the velocity of the magnetic cloud in units of the speed of light, respectively. The fractional energy change is then

  $$\xi = \frac{\Delta E}{E} = \frac{E_f - E_i}{E_i} \ . \qquad (2.8)$$

  By averaging over $\cos\theta_i$ (depending on the relative velocity between the cloud and the particle) it can be shown (e.g. in [34]) that the fractional energy change is proportional[4] to $\frac{4}{3}\beta^2$:

  $$\xi \propto \frac{4}{3}\beta^2 \ . \qquad (2.9)$$

- **First order Fermi acceleration**
  The big disadvantage of the second order Fermi acceleration is the very slow acceleration process. During the late 70's a more efficient acceleration mechanism was proposed, realized for cosmic ray encounters with plane shock fronts [85]. Assume

---

[4] assuming a non-relativistic speed of the magnetic cloud

a large shock wave propagating with velocity $-\vec{u}_1$. Relative to the shock front, the downstream shocked gas is receding with velocity $\vec{u}_2$, where $|u_2| < |u_1|$, and thus in the laboratory frame it is moving in the direction of the front with velocity $\vec{V} = \vec{u}_2 - \vec{u}_1$. To find the energy gain per crossing, one can identify the magnetic irregularities on either side of the shock as the clouds of magnetized plasma and proceed similar to Fermi's original idea. For the rate at which cosmic rays cross the shock from downstream to upstream, and upstream to downstream, one finds $\langle \cos\theta_i \rangle = -2/3$ and $\langle \cos\theta'_f \rangle = 2/3$, cf. [34]. The fractional energy change ξ (cf. Eqn. (2.9)) can be written as [34]

$$\xi \propto \frac{4}{3}\beta \ . \tag{2.10}$$

The term "first order" stems from the fact that the energy gain per shock crossing is proportional to β, the velocity of the shock divided by the speed of light, and therefore more efficient than Fermi's original mechanism. This is because of the converging flow - it does not matter on which side of the plasma you are, if you are moving with the plasma, the plasma on the other side is approaching you.
Note that in the first order mechanism the spectral index γ is independent of the absolute magnitude of the velocity of the plasma. It depends only on the ratio of the upstream and downstream velocities. For strong shocks the acceleration mechanism leads in a natural way to an E^{-2} spectrum [86].

2.5.3 Possible sources of ultra-high energy cosmic rays

Altough up to now no astrophysical object has been identified as source of UHECR, several acceleration models at specific astrophysical objects have been developed. In this section only a brief description is given. For a more complete discussion see e.g. [87].

- **Active Galactic Nuclei**: AGN are so far the brightest sources in the universe. Highly relativistic and confined jets of particles are a common feature of these objects. The acceleration takes place in the jets of AGN with the advantage that acceleration on the jet frame could have maximum energies smaller than these of the observed UHECR. The main problem is the adiabatic deceleration of the particles when the jet velocity starts slowing down. [88]

- **Gamma Ray Bursts**: GRBs are flashes of gamma rays associated with extremely energetic explosions in distant galaxies. They are the most luminous electromagnetic events known to occur in the universe. Different models put the acceleration site at the inner [89] or the outer [90] GRB shock. However, to explain the observed UHECR one needs a high GRB activity since most of the GRBs with determined redshifts are at $z > 1$.

- **Clusters of Galaxies**: With magnetic fields of several μG and length-scales up to 500 kpc acceleration up to almost 10^{20} eV is possible. Most of the lower energy cosmic rays would be confined and only the highest energy particles could escape [91].

- **Pulsars**: Young magnetized neutron stars can have large magnetic fields of 10^{13} G and accelerate iron nuclei up to energies of 10^{20} eV [92]. In contrast to many other models, this acceleration process is magnetohydrodynamic, rather than stochastic resulting in a spectrum proportional to $1/E$ [93].
- **Giant Radio Galaxies**: One model is that UHECR are accelerated at the termination shocks of the jets that extend to more than 100 kpc [94]. The magnetic fields inside seem to be sufficient for acceleration up to 10^{20} eV. Since the shocks are already inside the extragalactic space no adiabatic deceleration is expected. Possible candidates include Centaurus A and M87 in the Virgo cluster.

2.6 Cosmic ray propagation

2.6.1 Magnetic fields

During the propagation from source to Earth charged cosmic rays are deflected by galactic and extragalactic magnetic fields according to Eqn. (2.4). Unfortunately the extragalactic magnetic field strength is poorly known. The existence is confirmed by diffuse radio emission as well as by observations of Faraday Rotation Measures (FRM) towards polarized radio sources within or behind the magnetized medium. The estimate of the average strength of these field is of the order of 10^{-9} G (1 nG) [95] whereat even larger magnetic fields have been observed in clusters of galaxies. Even fields with nG strength would affect the propagation of ultra-high energy (UHE) cosmic rays resulting in deviations from the source direction and an increase of the path length from the source to the observer. Assuming magnetic fields of strength exceeding 10 nG were present on 10 Mpc coherence length they would lead to significant biases in the propagated spectra [96]. Only particles of energy 10^{20} eV would be able to propagate through the magnetic field lines.

Galactic magnetic fields are known to have a regular large scale structure with a typical value of a few μG approximately uniform over scales of the order of a few kpc. From FRM it is known that the magnetic lines follow the spiral arms of the galaxy. Also the largest deviation should occur when charged cosmic rays propagate in the vicinity of the Galactic Center region. Excluding the Galactic Center region, the average deflection angle for 10^{20} eV protons is between 3.1° and 4.5° in different galactic field models and for $2 \cdot 10^{18}$ eV protons even 17.7°–25.9° [97].

2.6.2 Energy loss processes

Here, the main energy loss processes for UHE nucleons are summarized. A more detailed characterization of UHE photon propagation can be found in Sec. 5.2. A graphical illustration of various energy loss processes for UHE nucleons is shown in Fig. 2.9. The energy attenuation as a function of propagated distance for various source energies E_0 is shown in Fig. 2.10.

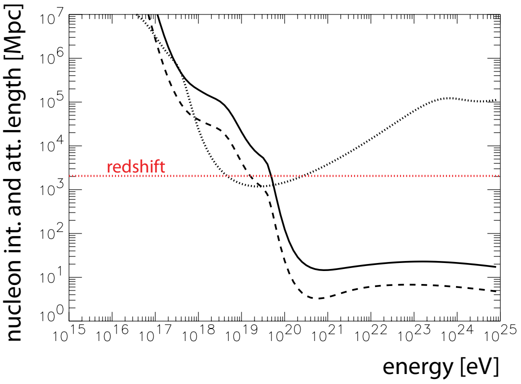


Figure 2.9: Nucleon interaction (dashed line) and attenuation (solid line) length for photo-pion production. The dotted line represents the proton attenuation length for proton pair production. The horizontal dotted red line indicates the expected attenuation due to Hubble expansion. Modified from [71, 98].

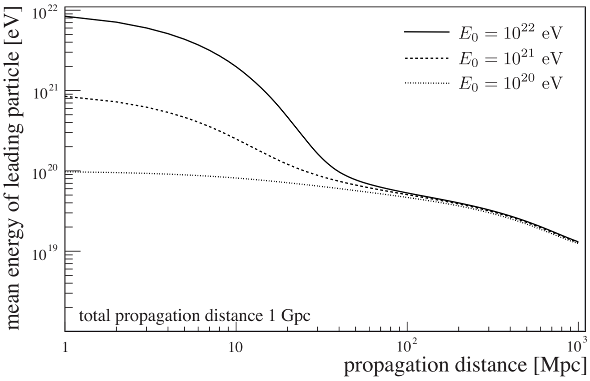


Figure 2.10: CRPropa [99] simulation of the mean energy of protons as a function of propagation distance through the extra-galactic background light for different indicated source energies E_0 (cf. Chapter 6). For propagation distances $\gtrsim$ 100 Mpc the primary energy is attenuated to almost the same value.

Pion production and GZK-effect

In 1965 Penzias and Wilson made a serendipitous discovery which was rewarded by the Nobel prize in 1978. They worked at the Crawford Hill location of Bell Telephone Laboratories and had built a horn antenna which they intended to use for radio astronomy and satellite communication experiments. Surprisingly their instrument had an excess 3.5 K antenna temperature which they could not account for [100]. It turned out that this puzzling antenna temperature is caused by a very uniform background radiation today known as cosmic microwave background (CMB) radiation. This radiation can well be explained as radiation left over from an early stage in the creation of the universe, and its discovery is considered a landmark confirmation of the Big Bang model of the universe.
The discovery, however, lead Greisen [101] 1965 and independently Zatsepin and Kuzmin [102] 1966 to the point, that this radiation would make the universe opaque to high energy protons, today known as the *GZK-suppression*[5]. They found that, above a few 10^{19} eV, thermal photons are seen highly blue-shifted by the protons in their rest frames. Here the energy of the microwave background photons γ_{CMB} is sufficient to excite baryon resonances and thus draining the high energy of the proton via pion production. The cutoff energy is a result of the threshold of pion production in the interaction of cosmic ray protons with cosmic background photons. The cross section is strongly increasing at the $\Delta^+(1232)$ resonance at a few tens of EeV making this process the dominant one in this energy range leading to the so-called GZK-suppression. The interaction can be described

---

[5] In literature this effect is also known as the GZK-Cutoff, although it is not a real cutoff.

as

$$p + \gamma_{\mathrm{CMB}} \rightarrow \Delta^{+}(1232) \rightarrow n + \pi^{+} \tag{2.11}$$

$$\rightarrow p + \pi^{0} \ . \tag{2.12}$$

In addition also other baryon resonances can occur with increasing energy:

$$p + \gamma_{\mathrm{CMB}} \rightarrow \Delta^{++} + \pi^{-} \rightarrow p + \pi^{+} + \pi^{-} \ ,$$

where Δ^{++} indicates e.g. $\Delta(1620)$ or $\Delta(1700)$ resonances. Assuming head-on collision the corresponding threshold energy E^{π}_{th} for the nucleon to produce pions can be described as

$$E^{\pi}_{\mathrm{th}} = \frac{m_\pi(2m_p + m_\pi)}{4\epsilon} \simeq 6.8 \cdot 10^{19} \ \mathrm{eV} \ ,$$

where ϵ is the typical CMB photon energy $\epsilon \simeq 10^{-3}$ eV [71] and m_π, m_p the pion and nucleon mass, respectively. Note that even lower threshold energies are possible since the CMB photon energy is represented by a Planck distribution and that also other extragalactic background light (e.g. infrared or radio background) affects the propagation of particles, cf. Fig. 2.9.

However, in spite of the prediction of the GZK-suppression, a number of experiments claimed to have observed events with $E > 10^{20}$ eV. Even before the suppression was proposed in 1966, Volcano Ranch [103] observed one event. Later on, SUGAR [104] and Haverah Park [19] observed high energy events as well, but the interpretation is still disputed. Recently, both, the Yakutsk Array [105] and AGASA [106] have claimed to measure events above 10^{20} eV. The Yakutsk Array result seems to be in accordance with the GZK-suppression, but AGASA has claimed the opposite. In 2006 the High Resolution Fly's Eye (HiRes) experiment claimed to observe the GZK-suppression [107]. HiRes observed two features in the ultra-high energy cosmic ray flux spectrum: The ankle at $4{\cdot}10^{18}$ eV and a high energy break in the spectrum at the energy of the GZK-suppression around $6 \cdot 10^{19}$ eV with a significance of about 4σ. Recent results from the Pierre Auger Observatory stress an flux suppression of more than 20σ above $\sim 4 \cdot 10^{19}$ eV [25] which is expected from the GZK-effect but could also be related to a change of the shape of the average injection spectrum at the sources, cf. Sec. 2.2 and Fig. 2.3 [27].

Pair production

Another important energy loss is the $e^{\pm}$ pair production, also known as Bethe-Heitler process which becomes important at energies below the GZK-suppression. It can be described as

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

The threshold energy E^{e}_{th} for the case of face to face collision is

$$E^{e}_{\mathrm{th}} = \frac{m_e(m_p + m_e)}{\epsilon} \simeq 4.8 \cdot 10^{17} \ \mathrm{eV} \ ,$$

where m_e denotes the electron mass. At energies around and above the GZK-suppression the characteristic time for $e^{\pm}$ production is $t \approx 5 \times 10^{9}$ yr [108]. At this energy photo-pion production is the main contribution to the proton energy loss, cf. Fig. 2.9.

Adiabatic fractional energy loss

The last important mechanism which dominates near and below the pair production threshold is redshifting due to the expansion of the universe. This adiabatic fractional energy loss can be described as

$$-\frac{1}{E}\left(\frac{\mathrm{d}E}{\mathrm{d}t}\right)_{\text{adiabatic}} = H_0 \ , \tag{2.13}$$

where H_0 is the present Hubble constant. An estimate of this effect on the energy loss is shown in Fig. 2.9. All other loss processes are negligible, except possibly in very dense central regions of galaxies.

Chapter 3

Extensive Air Showers

Measurement of extensive air showers are currently the only practicable method to explore cosmic rays above $\sim 10^{14}$ eV. This chapter introduces the physics of extensive air showers and its main properties. After an introduction on the discovery of these air showers by Pierre Auger, the development of individual components of an air shower are addressed in Sec. 3.1. Finally current and possible future detection techniques are discussed in Sec. 3.2.

An Extensive Air Shower (EAS) is a cascade of particles generated by the interaction of an initial high energy primary particle near the top of the atmosphere. The number of generated particles at first multiplies, then reaches a maximum before it attenuates more and more as particles fall below the threshold for further particle production. The measurement of EAS provides the only basis of cosmic ray observation above a primary energy of $\sim 10^{14}$ eV.

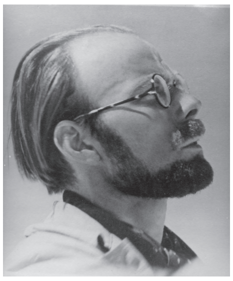


Figure 3.1: Pierre Auger

The history of EAS dates back to the late 1930s when the French physicist Pierre Auger first introduced the notation of *extensive cosmic-ray shower* [109]. He and his colleagues could show the existence of EAS with coincidence studies with counters and Wilson chambers partly at sea level and partly in two high altitude laboratories, Jungfraujoch (3500 m) and Pic du Midi (2900 m). With an arrangement of two parallel and horizontal counters placed at progressively increasing distances up to 300 m they searched for coincidences and concluded the existence of primary particles with energies around 10^{15} eV. What is happening in these showers is that nuclear cascades are initiated by cosmic rays of very high energy and many of the products reach the ground before losing all their

energy.
EAS can be studied at sea level, at various mountain elevations or even beneath the Earth. The experimentally determined quantities are:

- **Lateral distribution function (LDF)**
  This expresses the particle density as a function of distance from the shower axis. One differentiates between:
  - Lateral distribution of charged particles in the EAS ($e^{\pm} + \mu^{\pm}$)
  - Lateral distribution of Čerenkov light produced by EAS
  - Lateral distribution of muons generated by pion and kaon decays in the EAS ($\mu^{\pm}$)
- **Longitudinal development**
  This can be determined indirectly by studying the lateral distribution or directly by observing the atmospheric fluorescence and/or Čerenkov light associated with the passage of particles through the atmosphere.
- **Time distribution of particles arriving at ground**
- **Čerenkov light pulse rise time and width**
  This carries information about the longitudinal development of the shower.
- **Hadronic component**
  This component is concentrated very near the axis and is therefore difficult to study at high energies.

3.1 Development of extensive air showers

The first interaction of the primary cosmic ray with the atmosphere typically occurs at a height of 20–30 km, depending on the energy and mass of the primary particle. Assuming a primary cosmic ray nucleon, mostly pions and kaons together with a leading baryon are produced sharing the primary energy. Due to the large primary energy these secondary particles can again interact with other nuclei and produce new particles. The resulting air shower is composed of three main components as shown in Fig. 3.2:

- Hadronic component
- Muonic component
- Electromagnetic component

One important parameter of the longitudinal shower development is the matter traversed by the shower particles. Known as *slant depth X* it is measured in g/cm^2 from the

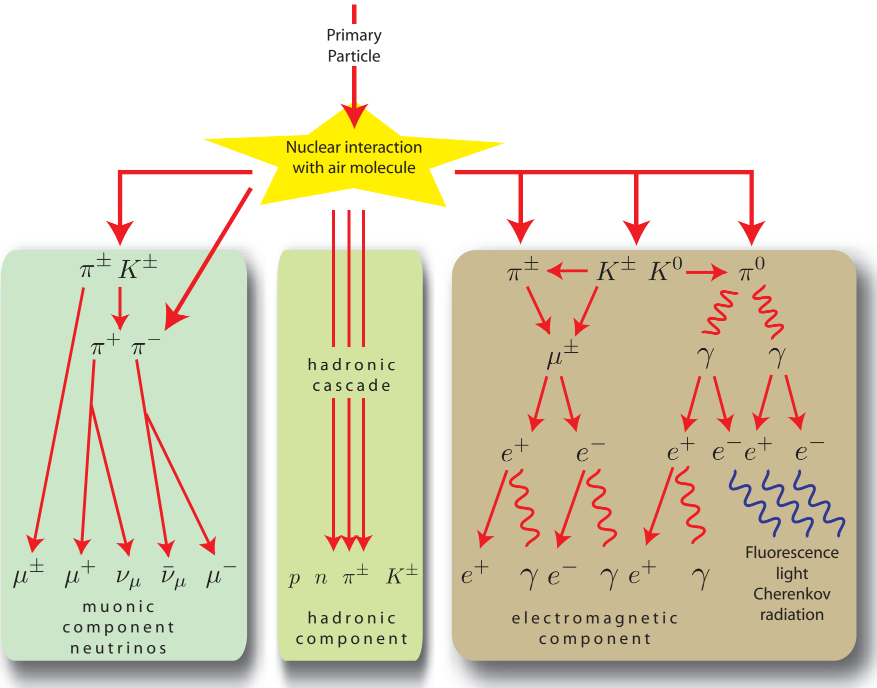


Figure 3.2: Schematic view of the development and interaction processes of an extensive air shower. Three main components are indicated.

top of the atmosphere along the direction of the incident nucleon and is related in good approximation[1] to the density profile $\rho(h)$ of the atmosphere by

$$X = \frac{X_v}{\cos\theta} ,$$

where X_v refers to the vertical atmospheric depth and is given by

$$X_v = \int_h^\infty \rho(h')\, dh' .$$

Cascade equations describe the propagation of particles through the atmosphere. They depend on the properties of the particles, their interactions and on the structure of the atmosphere [34]. In matrix notation one has:

$$\frac{dN_i(E_i,X)}{dX} = \underbrace{-\left(\frac{1}{\lambda_i}+\frac{1}{d_i}\right) N_i(E_i,X)}_{I} + \underbrace{\sum_j \int \frac{F_{ji}(E_i,E_j)}{E_i}\frac{N_j(E_j)}{\lambda_j}\, dE_j}_{II} . \tag{3.1}$$

Eqn. (3.1) describes the change in the number of particles of type i and energy E_i in an atmosphere at slant depth X. There are basically two parts:

---

[1] for $\theta \lesssim 60$ deg

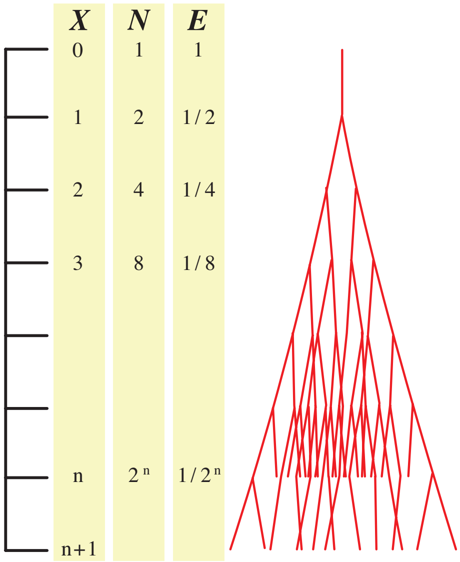


Figure 3.3: Heitler's toy model of cascade development. E symbolizes the energy, N the number of particles and $X = N\lambda$ the slant depth.

- **Part I**:
  This term describes the possibility that a particle i disappears into other types either through interaction with other particles having an interaction length λ_i or through decay with decay length d_i in g/cm^2. It can be understood as a *loss-term*.

- **Part II**:
  This term describes the possibility for creation of a particle of type i through interaction or decay of a particle j. The function $F_{ji}(E_i, E_j)$ is the dimensionless *inclusive cross section* and describes the probability of converting a particle of type j and energy E_j into the desired type i and energy E_i. It can be understood as a *creation-term*.

However, since all possible particle types are described with a cascade equation a set of coupled transport equations is needed. A numerical solution is possible and is implemented for instance in CONEX [110].

A simplified way to understand the most important features of cascades has been introduced by Heitler [111]. He describes a cascade of particles of the same type. After an interaction length λ two new particles are created, each carrying half of the primary particle energy $E = E_0/2$ as shown in Fig. 3.3. In each interaction process the number of particles doubles and the energy is shared among them. This sequence continues until the particle energy reaches a critical energy E_c for the splitting process. Below E_c the particles only lose energy, get absorbed or decay. The maximum number of particles is given by

$$N_{\max} = E_0/E_c \;, \tag{3.2}$$

while the depth of maximum is given by

$$X_{\max} = \lambda \frac{\ln(E_0/E_c)}{\ln 2} \ . \tag{3.3}$$

Although the Heitler toy model is extremely simple, it qualitatively correctly describes the shower development up to the maximum of shower development. The basic features of Eqn. (3.2) and Eqn. (3.3) hold for high energy electromagnetic cascades and also, approximately, for hadronic cascades, namely

$$X_{\max} \propto \ln(E_0) \tag{3.4}$$

$$N_{\max} \propto E_0 \ . \tag{3.5}$$

Still a central issue of air shower physics is the determination of the chemical composition of the primary cosmic ray nuclei above 10^{14} eV. The low flux does not allow direct measurements and one has to use measured properties of EAS to determine the composition. To use air showers for this purpose one first needs to know how showers initiated by heavy nuclei differ from those generated by light elements like protons or photons. The distribution of points where the nucleus first interacts inelastically with a target nucleon is crucial for the development of an air shower. The *superposition model* adequates for many purposes. Here one assumes that a nucleus of mass A and total energy E_0 is equivalent to A independent nucleons, each of energy E_0/A and that the distribution of first interactions is the same as if the nucleon had separately entered the atmosphere. Eqn. (3.4) then becomes

$$X_{\max} \propto \ln\left(\frac{E_0}{A \cdot E_c}\right) \ . \tag{3.6}$$

The dependence on A implies that on average showers generated by heavy primaries develop more rapidly than proton showers having the same energy as shown in Fig. 3.4. Unfortunately, there is only a logarithmic dependency on the mass, which makes it difficult to distinguish between masses.

Another distinguishing feature are the fluctuations in their longitudinal development. Heavy nucleons tend to have smaller fluctuations since each nucleus can be described as a beam of many incident nucleons.

3.1.1 Hadronic component

If the primary cosmic ray particle is a nucleon or nucleus, the cascade begins with a hadronic interaction, and the number of hadrons increases through subsequent generations of particle interactions. The depth of first interaction depends on the hadronic interaction length which is ~ 70 g/cm^2 for protons and ~ 15 g/cm^2 for iron nuclei. For a primary proton roughly half of the initial energy is lost in the first interaction for secondary particle production. The position of first interaction strongly influences the subsequent position of the shower maximum $X_{\max}$, which is therefore an important parameter to determine the composition of the primary particle. Since protons have a much larger interaction length

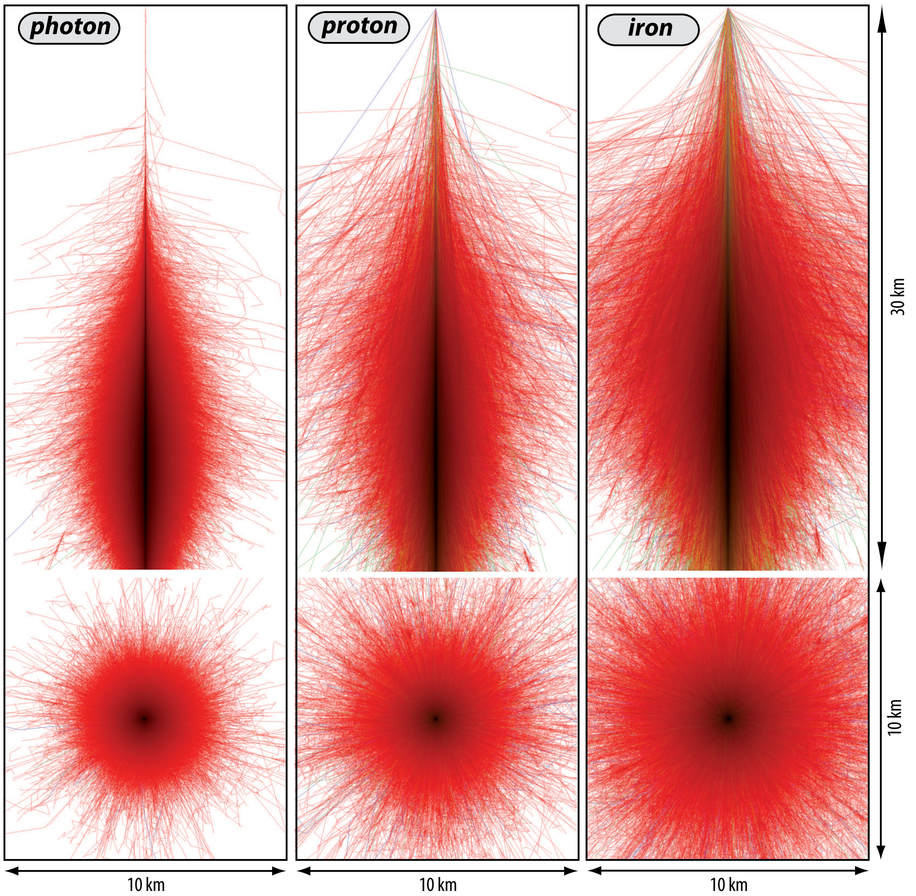


Figure 3.4: *Top pictures*: Longitudinal shower development of photon, proton and iron induced showers of energy 10^{14} eV using CORSIKA [112, 113] simulations (modified from [114]). The particle type is color coded with red ($e^{\pm}$, γ), green ($\mu^{\pm}$) and blue (hadrons). Colors can also be mixed whereat dark color corresponds to high track density. As can be seen the more heavier the primary particle the higher is the development in the atmosphere, (cf. Eqn. (3.6)). *Bottom pictures*: *xy*-projection of the above shower. The lateral extension is broader for heavy primaries.

than heavy nuclei, they will have larger fluctuations in the depth of the first interaction and develop deeper in the atmosphere.

Gaisser and Hillas [115] have parameterized the longitudinal development of hadronic showers as a function of first interaction X_0, depth $X_{\max}$ and size $N_{\max}$ at maximum and the mean free path λ:

$$N(X) = N_{\max} \left(\frac{X - X_0}{X_{\max} - \lambda} \right)^{\frac{X_{\max} - \lambda}{\lambda}} \exp\left(- \frac{X - X_0}{\lambda} \right) \ . \tag{3.7}$$

Eqn. (3.7) is used as a standard fit for the shower longitudinal development and is usually called the *Gaisser-Hillas* function (cf. Fig. 3.5 and Fig. 3.6).

The basic components in hadron showers are mainly pions and kaons, produced ei-

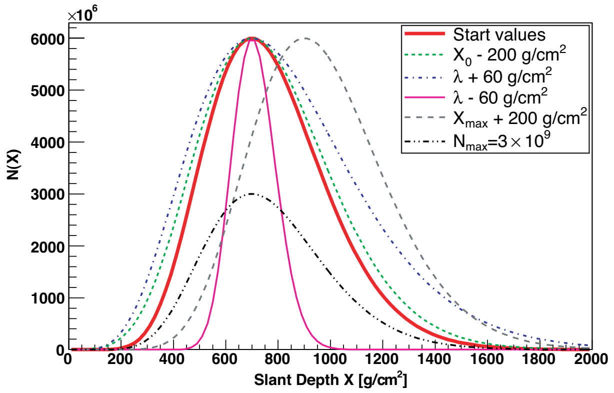


Figure 3.5: Example of a Gaisser-Hillas function with different parameters. The start parameters are $X_0 = 0$, $\lambda = 70$ g/cm^2, $X_{\max} = 700$ g/cm^2 and $N_{\max} = 6 \cdot 10^9$.

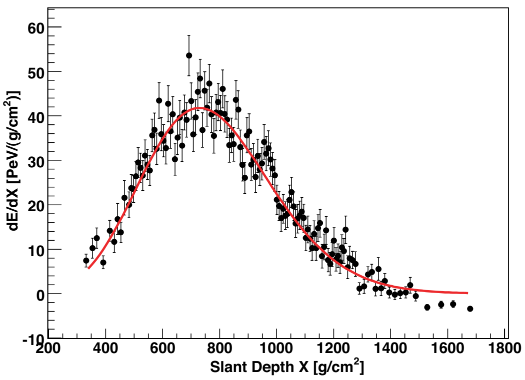


Figure 3.6: Example of a longitudinal profile of the Auger Golden Hybrid event 931431. The red line is the result of a Gaisser-Hillas fit.

ther directly in collisions or as decay products of short living resonances. This shower component is also called *shower core*, because it feeds all other components.

3.1.2 Electromagnetic component

The electromagnetic component of a hadron induced EAS essentially originates from the decay of neutral mesons, mainly pions

$$\begin{array}{lcll} \pi^0 & \longrightarrow & \gamma + \gamma & (\sim 98.8\%) \\ \pi^0 & \longrightarrow & \gamma + e^+ + e^- & (\sim 1.2\%) \ , \end{array}$$

Electromagnetic cascades can also be initiated directly by high energy photons or electrons. During an interplay between pair production and bremsstrahlung an electromagnetic cascade can develop. In an electromagnetic field of a nucleus N the pair production process can be described as

$$\gamma + N \rightarrow N + e^- + e^+ \ ,$$

whereas bremsstrahlung leads to

$$e^\pm + N \rightarrow N + e^\pm + \gamma \ .$$

The emission of further photons may produce additional $e^\pm$-pairs. This reaction chain proceeds until a threshold energy (critical energy) $E_c \approx 81$ MeV in air is reached. For $E < E_c$ the ionization energy loss starts to dominate the bremsstrahlung process and the electron is attenuated within one radiation length X_r.

An approximate formula for the longitudinal shower profile of electromagnetic air showers has been derived from cascade theory by Rossi and Greisen [116]. Considering only particles with energy E the depth of shower maximum $X_{\max}$ can be written by

$$X_{\max} \approx X_r \ln\left(\frac{E_0}{E}\right) \ . \tag{3.8}$$

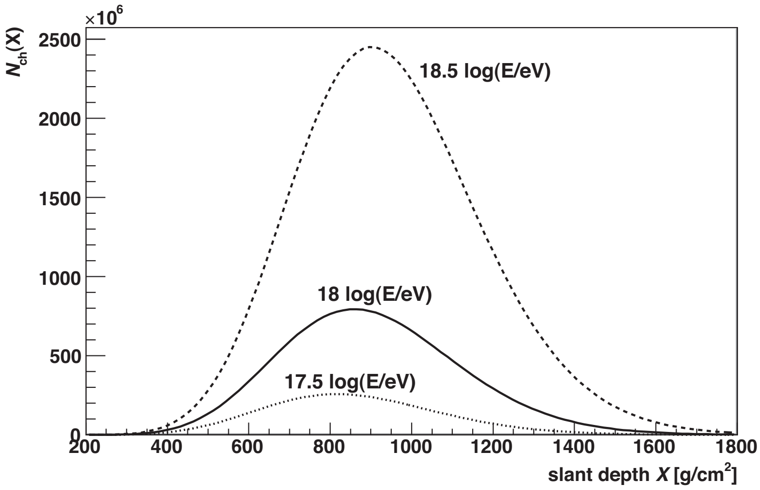


Figure 3.7: Illustration of the Greisen function for three different primary energies as indicated. The radiation length was set to $X_r = 37$ g/cm^2 and the critical energy to $E_c = 81$ MeV.

Greisen developed a compact parametrization of the mean number of charged particles N_{ch} as a function of slant depth X based on the solution of the one-dimensional cascade equations [117] today known as Greisen function:

$$N_{\text{ch}}(X, E) = \frac{0.31}{\sqrt{\ln(E/E_c)}} \exp\left(\left[1 - \frac{3}{2}\ln s\right]\frac{X}{X_r}\right) . \tag{3.9}$$

Here E_c denotes the critical energy, X_r the radiation length and s the shower age phenomenologically defined as

$$s = \frac{3X}{(X + 2X_{\text{max}})} . \tag{3.10}$$

Combining Eqn. (3.8), (3.9) and (3.10) the Greisen function can be rewritten as

$$N_{\text{ch}}(X, E) = \frac{0.31}{\sqrt{\ln(E/E_c)}} e^{\frac{X}{X_r}} \left(\frac{3X}{X + 2X_r \ln(E/E_c)}\right)^{-\frac{3X}{2X_r}} . \tag{3.11}$$

Since the Greisen function was derived from purely electromagnetic cascade theory EAS initiated by photon primaries should fit well to the profile in contrast to hadron induced showers, cf. Sec. 7.3.2. The primary energy as the only free parameter is another advantage. An illustration of the shape of the Greisen function for three different energies is shown in Fig. 3.7.

3.1.3 Muonic component

The muonic component of an EAS emerges from the decay of secondary pions and kaons of the hadronic component:

$$\begin{array}{lcll} \pi^{\pm} & \longrightarrow & \mu^{\pm} + \nu_{\mu}(\bar{\nu_{\mu}}) & (\sim 99.99\%) \\ K^{\pm} & \longrightarrow & \mu^{\pm} + \nu_{\mu}(\bar{\nu_{\mu}}) & (\sim 63.51\%) \end{array}$$

Indeed, the daughter muons are also unstable with typical lifetimes of $\tau_{\mu} \sim 2.2$ μs but taken their experienced time dilatation into account, they mostly reach the ground, unless their energy is smaller than a few GeV. Therefore, the muonic component is also called the hard component of cosmic radiation. On their way to the ground muons are not much deflected by multiple scattering. Their path through the atmosphere is almost rectilinear and makes detection on the ground very helpful for reconstructing the early stage of the shower development. Since the highest energy muons result from high energy pions and kaons, they carry important information about the hadronic interaction at those energies which can be used to test theoretical interaction models. Moreover, the muonic component is a very important parameter to examine the type of the primary particle: Most of the differences between photon and hadron–initiated showers are related to the fact that hadron induced showers develop a significant muon component whereas there are very few muons in photon induced showers.

3.2 Detection techniques

There are several detection techniques for EAS each utilizing special features of air showers ranging from direct sampling of particles in the shower to measurements associated with the emission of fluorescence light, Čerenkov light or radio emission. The most common approach is the direct detection of shower particles in an array of sensors spread over a large area (to account for the low cosmic ray flux) to sample particle densities as the shower arrives at the Earth's surface as described in Sec. 3.2.1. Another well-established method involves measurements of the longitudinal development of the EAS using fluorescence light produced via interactions of charged particles in the atmosphere, introduced in Sec. 3.2.2. There are also some recently rediscovered techniques like radio, radar and acoustic detection of EAS, explained in more detail in Sec. 3.2.3.

3.2.1 Surface arrays

The surface array is comprised of particle detectors, such as Čerenkov radiators or plastic scintillators, distributed with approximately regular spacing. The aim is to measure the energy deposited by particles of the EAS as a function of time. With the energy density measured at the ground and the relative timing of hits in the different detectors one can estimate the energy and direction of the primary cosmic ray. Furthermore, the shape of the measured traces from each individual detector provides additional information on the shower content.

Reconstructing air shower properties involves fitting the lateral distribution function of particle densities at the ground (cf. Fig. 3.8). Clearly, the lateral distribution function has to be determined for each experiment individually. At Haverah Park a good fit to

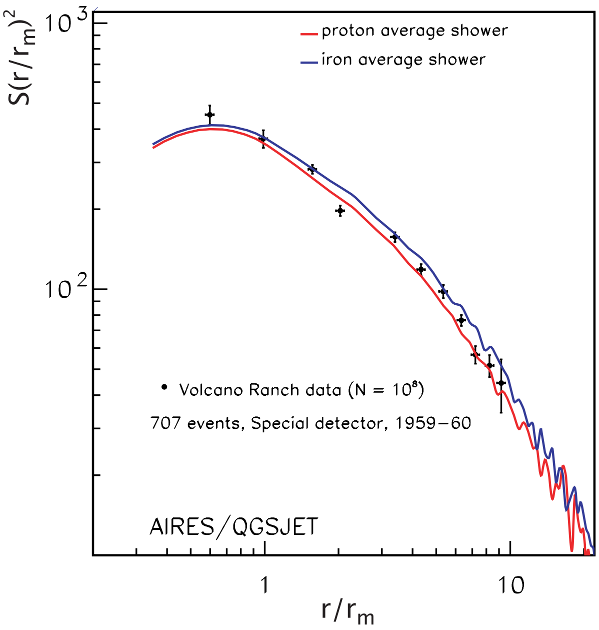


Figure 3.8: Example of an averaged lateral distribution function simulated with AIRES/QGSJET [118] compared to measurements from Volcano Ranch [119] of about 10^{18} eV. r/r_m refers to the distance to the shower axis and S is the lateral distribution of particles at ground (from [120]).

the water Čerenkov lateral distribution was found to be the modified power law function valid for core distances 50 m $< r <$ 700 m, zenith angles $\theta < 45°$ and energies $2 \cdot 10^{17}$ eV$<$ $E < 4 \cdot 10^{18}$ eV [121]

$$\rho(r) = kr^{-(\eta + \frac{r}{4000})} \ , \tag{3.12}$$

where k is the normalization parameter and η is given by

$$\eta = 2.20 - 1.29 \sec\theta + 0.165 \log\left(\frac{E}{10^{17}\ \mathrm{eV}}\right)$$

As already mentioned, the muon content at ground level depends on the composition of the primary cosmic ray. Surface arrays with the ability to distinguish muons from electrons and photons are therefore able to give some hints about the composition of the primary cosmic ray. Another way to gauge the muon content arises from the signal rise time, since the muon content tends to be compressed in time compared to the electromagnetic component.

3.2.2 Fluorescence detectors

Almost 50 years ago Chudakov in the Soviet Union and Suga in Japan realized that nitrogen fluorescence might be used to detect EAS. First measurements of temperature and pressure dependencies of the fluorescence efficiency were made by Greisen and his student Bunner at their Cornell group. They were also the first to build an air shower detector using Fresnel lenses [122], but no air showers were detected in an unambiguous way, because electronic devices were too slow at that time. In 1976 the technique was first successfully demonstrated by the Utah group which was the starting point for founding the Fly's Eye fluorescence detector [123].

During the propagation of an EAS through the atmosphere much of the energy is dissipated by exciting and ionizing air molecules (mainly nitrogen) along its path. During the de-excitation process ultraviolet radiation ($\lambda \sim 300 - 400$ nm) is emitted isotropically[2]. This allows detectors to view showers from the side, even at large distances. Although fluorescence light has a very low production efficiency, of the order of 4 photons per meter of electron track, it is possible to detect them over a very large distance. The shower development appears as a rapidly moving spot of light across a night-sky background of starlight, atmospheric air-glow, and man made light pollution. The observed angular motion of the spot depends on both, the orientation of the shower axis and the distance. The measured brightness of the spot indicates the instantaneous number of charged particles present in the shower, but is also affected by Čerenkov light contamination and atmospheric scattering. Since the ratio of energy emitted as fluorescence light to the total energy deposited is less than 1%, low energy showers ($< 10^{17}$ eV) are difficult to detect. Another interference arises from moonlight and therefore observations are only possible during clear moon-less nights, resulting in an average 10% duty cycle (cf. [124]).
A fluorescence telescope consists of several light collectors, which image different regions in the sky onto clusters of light sensing and amplification devices. The fluorescence light is collected by photomultiplier tubes (PMTs) positioned approximately on the mirror focal surface. The shower development can then be seen as a long, rather narrow sequence of hit PMTs. With this information the geometry of the shower is determined. Once the geometry is known the longitudinal profile can be determined. This usually involves a three parameter fit to the Gaisser-Hillas function (Eqn. (3.7)). The integral of the longitudinal profile is a calorimetric measure of the total electromagnetic shower energy

$$\begin{aligned} E_{\mathrm{em}} &= \alpha_{\mathrm{loss}} \int N(X)\,\mathrm{d}X \qquad (3.13) \\ &= \int \frac{\mathrm{d}E}{\mathrm{d}X}\,\mathrm{d}X\ , \qquad (3.14) \end{aligned}$$

where α_{loss} is the mean energy loss rate for relativistic electrons in the atmosphere which can be approximated as $\alpha_{\mathrm{loss}} \sim 2.2$ MeV g^{-1} cm^2 [125].
The largest cosmic ray event reported so far was detected by a fluorescence telescope of the Fly's Eye experiment with an estimated energy of $3.2 \cdot 10^{20}$ eV and maximum size near a depth of 815 g/cm^2 [126].

---

[2] unlike the very intense Čerenkov light produced by shower particles in air.

3.2.3 Other detection mechanism

- **Radio:**
  A more recent technique to detect air showers utilizes the effect that EAS also emit radio frequency (RF) energy. These radio pulses are produced by several mechanisms, though it is thought that from about 20-100 MHz, the dominant process can be described as coherent synchrotron emission by the electron and positron pairs propagating in the Earth's magnetic field [127]. In the early 1960s RF pulses coincident with EAS were already measured [128] but the promising results from surface arrays and fluorescence eyes abandoned this technique. In the context of next generation digital telescopes more ambitious possibilities have been described, e.g. by LOFAR [129]. The great potential of a large scale application has been reported by e.g. LOPES [130, 131, 132] and CODALEMA [133]. They also confirmed that the emission is coherent and of geomagnetic origin, as expected by the geosynchrotron mechanism [134]. The strategy to combine radio signals from EAS in coincidence with sophisticated surface arrays was successfully demonstrated by the LOPES experiment in combination with KASCADE-Grande [135]. To extend the energy range to above $\sim 10^{18}$ eV and to combine radio and fluorescence light detection for the first time, the radio test set-up AERA is being installed at the Pierre Auger Observatory site [136, 137, 138].

- **Radar:**
  Another re-explored radio technique may be the detection of radar reflections of the ionization columns produced by EASs [139]. This can be used as an independent technique to detect EASs or as a compliment to existing surface detectors or fluorescence telescopes.

- **Microwave:**
  The detection of EAS utilizing microwaves is a possible new technique which relies on detection of expected continuum radiation in the microwave range, caused by free electron collisions with neutrals in the tenuous plasma left after the passage of the shower. This microwave molecular bremsstrahlung radiation (MBR) has been investigated in first test facilities (AMBER and MIDAS [140]) and efforts are underway to deploy an AMBER test bed array within the Pierre Auger Observatory. More details on this techniques and first results are given in [141].

- **Acoustic:**
  The possibility of using acoustic detection of EAS was already described in the fifties by Askaryan [142]. During the late seventies Askaryan *et al.* [143] and Learned [144] developed the thermo-acoustic model, but similar to radio detection a revival was initiated $\sim$ 10 years ago motivated by the GZK-suppression and corresponding cosmogenic neutrino production at ultra-high energies [145, 146]. The acoustic detection is based on the reconstruction of characteristic sound pulses that are generated by (neutrino-induced) particle cascades in water or ice. Here the energy deposition of cascade particles is connected to a local heating accompanied by an

expansion of the medium. The resulting thermo-acoustic signal is bipolar in time with a corresponding spectral energy density peaked at about 10 kHz [147]. Several experiments have build R&D set-ups to investigate the feasibility of the technique, e.g. AMADEUS [148] for ANTARES [149] or Baikal [150]. In ice, however, recent measurements on the acoustic attenuation length show an unexpected small attenuation length placing questions on the feasibility on acoustic detection with large scale experiments like IceCube [151].

Chapter 4

Pierre Auger Observatory

In this chapter the currently largest cosmic ray experiment, the Pierre Auger Observatory, is introduced. After a brief introduction on the history of the observatory the current status of the Southern and the Northern site is illustrated. Sec. 4.1 and following focus on the Southern site and familiarize the surface detectors and fluorescence telescopes. Data acquisition and trigger logic as well as calibration measurements are briefly discussed. Sec. 4.3 sketches the current Auger software framework $\overline{\textit{Off}}\underline{\textit{line}}$ where Sec. 4.4 and following are addressed to geometry reconstruction and energy determination of measured extensive air showers.

Currently, the world's largest detecting system for ultra-high energy cosmic rays is located in the Argentinean Pampa Amarilla. Named after the French physicist, the Pierre Auger Observatory was designed to study the upper ($\gtrsim 10^{18}$ eV) end of the cosmic ray spectrum [152, 153]. The detectors are optimized to measure the energy spectrum, arrival directions and the chemical composition of cosmic rays utilizing two complementary techniques: detecting the nitrogen fluorescence in the atmosphere caused by an extensive air shower and measuring the lateral distribution function of particles that reach the ground. This so-called "hybrid" technique is unique, enhances the resolution and is valuable in determining systematic errors inherent in both techniques as well as providing more information to determine the particle type and check hadronic interaction models.

The history of the observatory dates back to the early 90s when J. Cronin and A. Watson came together at the ICRC conference in Dublin, Ireland. They realized that a giant air shower detector in the northern as well in the southern hemisphere is needed to solve current issues of the mysterious ultra-high energy cosmic rays. Previous experiments have brought some light in certain topics but they have added more open questions too. Within

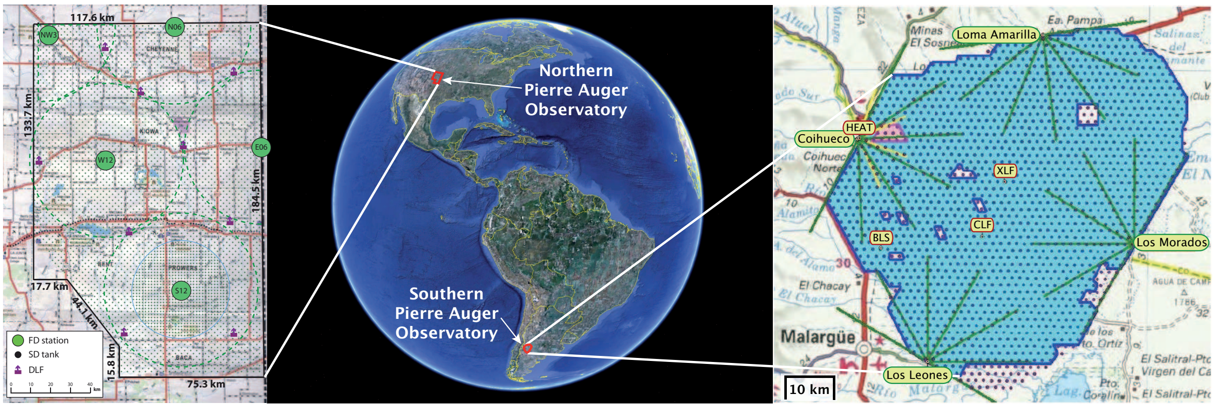


Figure 4.1: *Left:* Planned Northern site of the Pierre Auger Observatory. Each dot represents one water Čerenkov tank. The proposed positions of the fluorescence telescopes are indicated with green circles. *Right:* Status of the Southern site of the Pierre Auger Observatory in July 2010. The shaded area represents the already equipped surface stations. Small holes with missing tanks are due ongoing landowner arrangements or hardly accessible areas. The position of four fluorescence telescopes stations are indicated as well as the HEAT enhancement, the CLF, BLF and XLF. Modified from [154, 155, 156].

the next years the initial idea for the southern hemisphere observatory developed through a sequence of workshops in Paris (1992), Adelaide (1993), Tokyo (1993) and finally at Fermilab in 1995. The resulting design report [152] containing a reference design and a cost estimate for the detector became the basis for funding proposals and finally the guide for building the Southern Pierre Auger Observatory near the small town Malargüe in Mendoza Province in Argentina just east of the Andes Mountains.

Although not completely finished and still growing, data taking started in 2004 and first mayor publications were already made in 2007 [58]. The inauguration of the completion of the Southern Pierre Auger Observatory could be celebrated in November 2008. Today more than 280 physicists from more than 70 institutions in 18 countries around the world (as shown in Fig. 4.2) are collaborating in an joint effort to bring the nature of cosmic rays forward. The Southern site consists of ~1660 water Čerenkov tanks (surface detector - SD) with a spacing of 1.5 km at an altitude of ~1400 m above sea level (a.s.l.) covering an area of 3000 m^2. This area is overseen by 27 fluorescence telescopes (fluorescence detector - FD) located in five different telescope stations at the edge of the array (cf. Fig. 4.1). Given the very low particle flux at highest energies this size will collect a couple of events above 10^{20} eV per year.

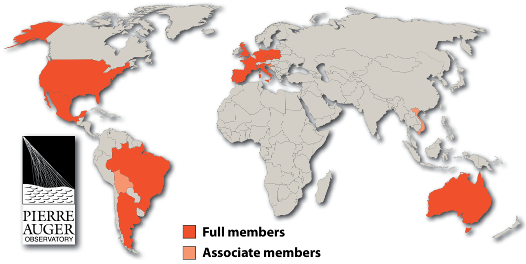


Figure 4.2: 18 participating countries. Modified from [157].

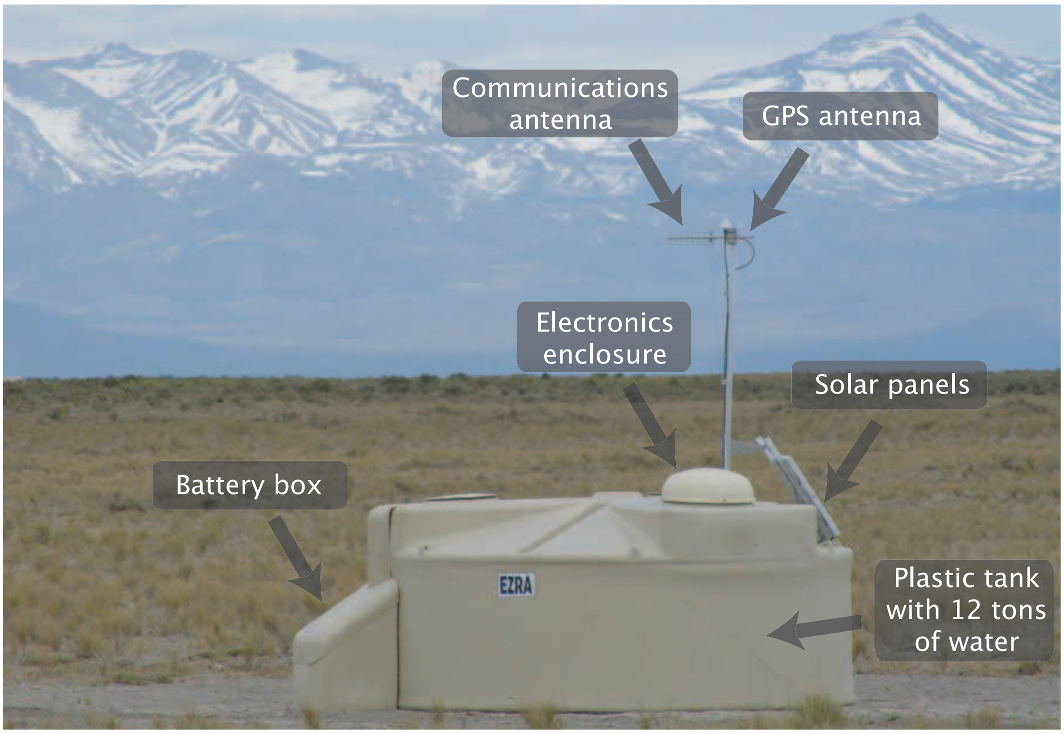


Figure 4.3: View of surface detector “Ezra” within the Argentinean pampa.

The Northern site of the Pierre Auger Observatory is currently in its planning phase. In 2005 the location was selected to the South-East corner of the State of Colorado (USA). At an altitude of about 1300 m a.s.l. the SD will consist of 4000 water Čerenkov tanks covering a total area of 20.000 km^2, more than six times larger than the Southern array. Almost full coverage of the SD system will be achieved with 39 fluorescence telescopes located in five different stations as shown in Fig. 4.1. Technically, the construction of Auger North could begin in 2011, but financial support from funding agencies is an issue. Recent developments indicate that the major funding agency, the National Science Foundation (NSF), will not support development and construction of the Northern site near-term, challenging the possibility for completion. A more detailed description of the Northern site can be found in [154]. In the following sections, however, the focus will be on the Southern Pierre Auger Observatory.

4.1 Surface detector

The surface detector (SD) of the Southern array is a ground array covering an area of 3000 km^2 with ∼1660 water-Čerenkov stations set on a regular triangular grid, with 1.5 km separation between them [152] yielding full efficiency for EAS detection above $5 \cdot 10^{18}$ eV. The communication to the central base station is accomplished through a radio link.
An example of a surface detector is shown in Fig. 4.3. Each station is a cylindrical tank,

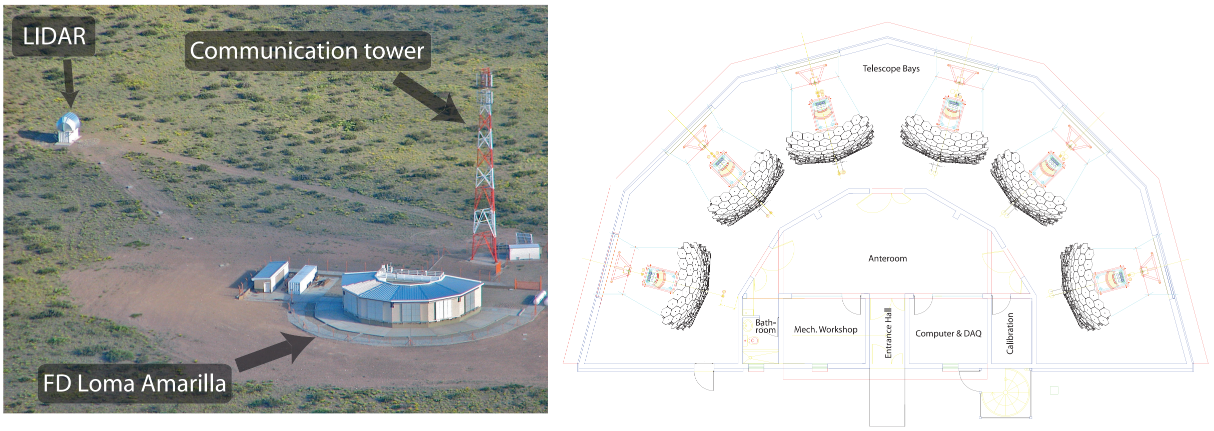


Figure 4.4: *Left panel:* Picture of fluorescence detector Loma Amarilla *Right panel:* Schematic view of a telescope station and its individual configuration [22].

filled with 12000 liter of purified water, operating as a Čerenkov light detector. The water is contained within a bag that has a high diffuse reflectivity in the wavelength range of Čerenkov light production and photocathode sensitivity. Three windows are placed on top of the bag where three 9″ PMTs are placed detecting Čerenkov light when particles propagate through the detector. The signals are then passed through filters and read out by a flash analog digital converter (FADC) that samples at a rate of 40 MHz. The digitized data are stored in ring buffer memories and processed by a programmable logic device (FPGA) to implement various trigger conditions [158, 159]. The timing information for each station is received from a GPS system located on each tank with timing resolution < 20 ns [160]. Local electronics as well as the GPS system are powered by two solar panels, combined with buffer batteries.

In order to cope with large amounts of data, the recorded signals are transferred to the Central Data Acquisition System (CDAS) only if a shower trigger has been detected in three adjacent tanks simultaneously. Since the trigger thresholds may change with time, calibration quantities are continuously monitored for each station in the array. The calibration is performed with single cosmic muons by adjusting the trigger rates. This is done with an accuracy of 5% for the PMT gains. For convenience, the number of particles in each tank is defined in units of Vertical Equivalent Muons (VEM) defined as the average charge signal produced by a penetrating down going muon in the vertical direction.

4.2 Fluorescence detector

The fluorescence detector (FD) of the Southern array is conceived to detect fluorescence light, emitted by de-excitation processes of nitrogen molecules. The fluorescence yield is very low[1], but large imaging telescopes are able to detect this light during clear new to half moon nights, resulting in a duty cycle of $\approx 10 - 15\%$.

---

[1] Approximately 4 photons per meter of electron track [161]

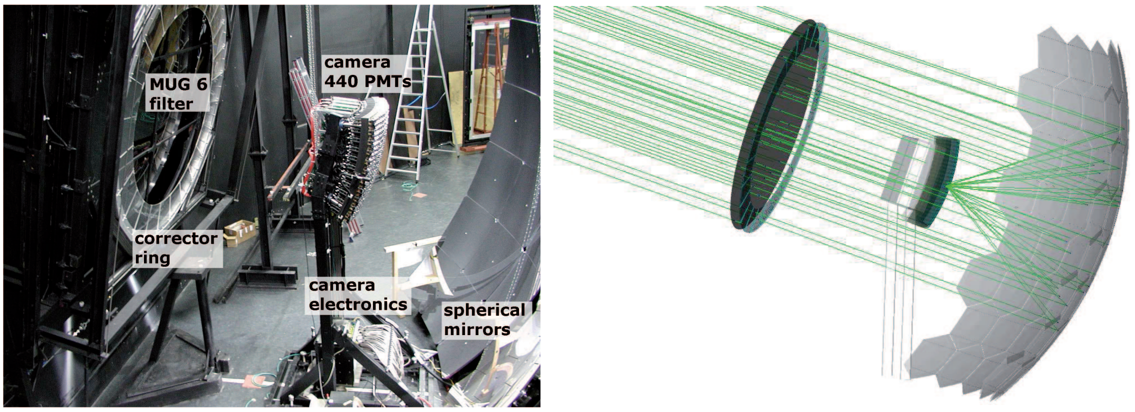


Figure 4.5: *Left panel:* Design of the fluorescence telescope [164]. *Right panel:* Ray tracing simulation of the optical system of the telescope using Geant4 [22, 165, 166].

The FD is composed of 4 different eyes (named Los Leones, Los Morados, Loma Amarilla and Coihueco) as shown in Fig. 4.1 located at the perimeter of the SD, which enables detection of EAS simultaneously by SD and FD ("hybrid detection"). Each eye consists of 6 independent Schmidt telescopes (bays) each made of a 440 pixel camera, which achieves a covering area of 1.5° × 1.5°. They are arranged in a 22 × 20 matrix to give a field of view of 30° in azimuth and 28.6° in elevation, adding to a 180° view inwards the array of one eye (cf. Fig. 4.1). The fluorescence light is collected by a 12 m^2 mirror with a radius of 3.4 m and reflected to the camera located at the focal surface of the mirror. The telescopes use a Schmidt optics design to avoid coma aberration, with a diaphragm, at the center of curvature of the mirror. The radius of the diaphragm is 1.1 m including a corrector lens with an inner radius of 0.85 m and outer radius of 1.10 m. The effect of the lens is to increase the light collection area by a factor of two while maintaining an optical spot size of 0.5° [162]. To avoid interfering background light each diaphragm has a UV transparent filter that restricts the incoming light to the wavelength range between 300 and 420 nm, which is where the main fluorescence emission lines can be found. To reduce signal losses when fluorescence light crosses PMT boundaries, small light reflectors ("mercedes stars") are placed between the PMTs [163]. The setup of the telescope as well as a ray tracing simulation of the optical system is illustrated in Fig. 4.5.

4.2.1 Data acquisition and trigger

The PMT signals are continuously digitized at 10 MHz sampling rate with a dynamic range of 15 bit in total. In order to filter traces out of a random background, a FPGA (Field Programmable Gate Array) based multi-level trigger system is used:

- **First level trigger (FLT)**:
  Each telescope consists of 20 FLT boards. One FLT board processes the data from one 22-channel column and fires if the signal of a PMT excesses a given threshold

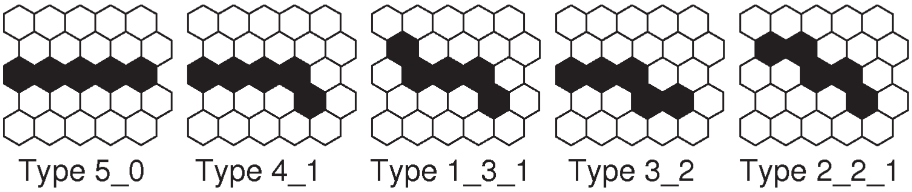


Figure 4.6: SLT recognition pattern regarded as straight track segments [22].

which is continuously adjusted to maintain a pixel trigger rate of 100 Hz.

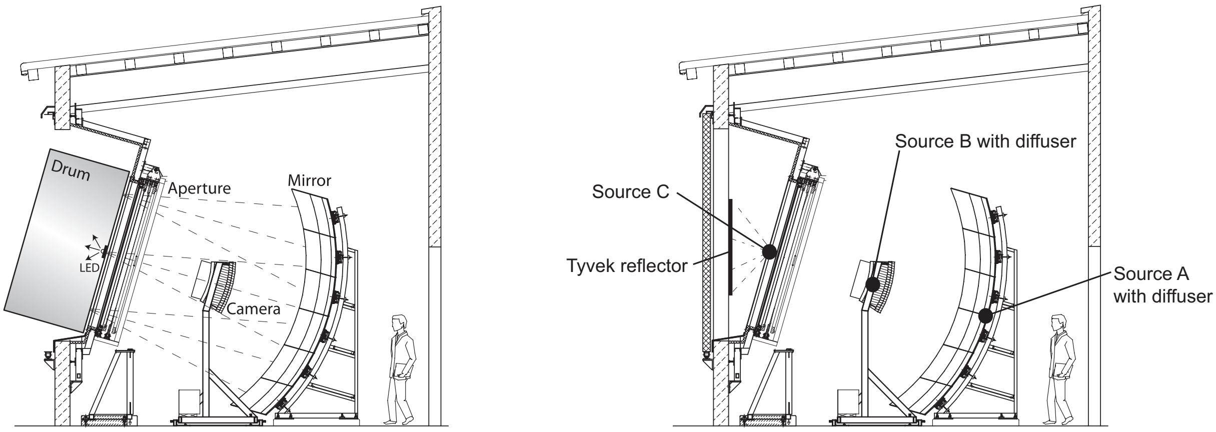
- **Second level trigger (SLT)**:
  To discriminate if the FLT was induced from a shower track or from noise a SLT was implemented [167]. Therefore all 20 FLT boards from one camera are read out by the SLT which is also implemented in the FPGA logic. The algorithm searches for pattern of at least five connected triggered pixels in length as shown in Fig. 4.6. Since some tracks may not pass through every pixel center and hence do not record enough light to fire the trigger and to be fault-tolerant against defective PMTs, the algorithm requires only four triggered pixels out of five. However, there may still be some fraction of unphysical events like small tracks induced by cosmic muons in the camera or noise from lightning or stars in the field of view.

- **Third level trigger (TLT)**:
  The aforementioned drawback is solved by the TLT which is, in contrast to the FLT and SLT, a software algorithm. In a first step lightning events are filtered out basically by the time development and integral of FLT multiplicities resulting that $\sim 99\%$ lightning events are correctly rejected [22]. In a second step the remaining noise events with smaller number of triggered pixels are filtered taking into account the correlation between the spatial arrangement and peak signal times of triggered pixels. The TLT performance was validated with simulated showers finding that $\sim 94\%$ of all noise events are rejected correctly whereat the fraction of true showers rejected by the trigger is below 0.7% [22].

- **Hybrid trigger (T3)**:
  For events passing the TLT a hybrid trigger (T3) is sent to the CDAS which acts as an external trigger for the surface array. The main purpose is to record hybrid events at low energies ($E < 3 \cdot 10^{18}$ eV) where the surface array is not fully efficient. After a simple online reconstruction only tanks near the FD are read out close to the calculated impact time.

4.2.2 Calibration

To measure air shower energies correctly the fluorescence detectors have to be calibrated and monitored. The basic principle is to convert ADC counts to a light flux at the telescope aperture for each channel that receives a portion of the shower signal.



(a) Schematic view of the absolute calibration using a drum shaped source [22].

(b) Schematic view of the relative calibration [22]. The light source positions A, B and C are indicated.

Figure 4.7: Schematic view of the calibration system [22].

- **Absolute calibration**:
  The absolute calibration provides the conversion between the digitized signal (in ADC units) and the photon flux incident on the telescope aperture. During the calibration a large homogeneous diffuse light source was constructed which can be mounted in front of the telescope diaphragm, as shown in Fig. 4.7 (a). This drum shaped source has a diameter of 2.5 m and the emitted light is known from laboratory measurements [168]. The ratio of the drum intensity to the observed signal for each PMT gives the required calibration. The advantage is that the complete light collection and detection system can be taken as a black box. However, this calibration method is very time- and work-intensive and can only be performed on a non-regular base approximately three or four times a year.

- **Relative calibration**:
  The main goal of the relative calibration is to monitor short term and long term changes between successive absolute calibration measurements and to check the overall stability of the FD [169]. The system is used before and after each night of operation, cf. Fig. 4.7 (b).

  - *Calibration A*: The light source for this calibration is a 470 mm LED located at the center of each mirror in the FD building. The light pulses are directed over a diffusor into the camera. With this method the PMT long-time stability is measured.
  - *Calibration B*: The light sources fibers are xenon flash lamps and split near each camera and terminate a thin Teflon diffusor located at the sides of the camera with the light directed at the mirror.
  - *Calibration C*: Also using xenon flash lamps the fibers end outside the aperture

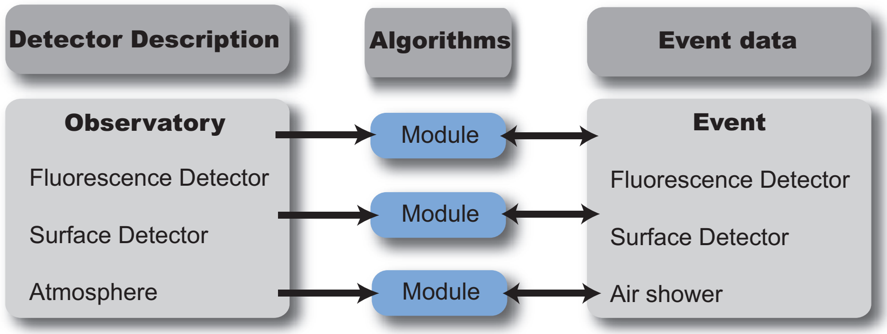


Figure 4.8: General structure of the Offline framework. Simulation and reconstruction are accomplished in different modules and each module is able to read information from the detector description and/or event data, process information, and write the results back into the event (cf. [172]).

> with light directed outwards. The light is reflected by Tyvek sheets and the diffuse light enters the aperture and the camera to give the opportunity to measure the whole light collecting system including optics and filters.

Also the atmospheric conditions must be monitored closely since attenuation of the light from the EAS to the telescope due to molecular (Rayleigh) and aerosol (Mie) scattering has to be corrected. Several methods are currently used to determine the effects in the air at any given time during data taking. The relevant parameters are determined by a Horizontal Attenuation Monitor (HAM), Aerosol Phase Function monitors (APF) and a Laser Illuminated Detection And Ranging system (LIDAR) located at each eye (cf. [170, 171]). There are also cloud and star monitors to detect clouds and track stars and any changes in their intensity caused by changing atmospheric conditions.

4.3 Offline framework

Within the Pierre Auger collaboration, the general purpose software framework Offline has been designed in order to provide an infrastructure to support a variety of distinct computational tasks necessary to analyze data gathered by the observatory [172, 173]. The requirements of this project place strong demands on the software framework underlying data analysis. Therefore, it is implemented in `C++` taking advantage of object-oriented design and common open source tools.

The general body comprises three principal parts as shown in Fig. 4.8:

1. **Processing modules**:
   Most tasks of interest can be reasonably factorized into sequences of self contained processing steps. These steps are realized in *modules*, which can be inserted into the framework via a registration macro. The advantage is to exchange code, compare algorithms and build up a wide variety of applications by combining modules in various sequences. In order to steer different modules, a *XML*-based run controller was

constructed for specifying sequencing instructions. This user friendly environment allows to choose which modules to use and to implement new modified modules. *XML* files are also used to store parameters and configuration instructions used by modules or by the framework itself. A central directory points modules to their configuration files which is created from a *bootstrap* file whose name is passed on the command line at run time.

2. **Event structure**:
   The event data structure acts as the principal backbone for communication between modules. It contains all raw, calibrated, reconstructed and Monte Carlo data changing for every event. Therefore, the event structure is build up dynamically, and is instrumented with a protocol allowing modules to interrogate the event at any point to discover its current constituents.

3. **Detector description**:
   In contrast to the event structure the detector description is a read-only information. It provides a unified interface from which module authors can retrieve static (stored in *XML* files) or relatively slowly varying information (stored in *MySQL* databases) about detector configuration and performance at a particular time. The requested data is passed to a registry of *managers*, each capable of extracting a particular sort of information from a particular data source. The detector description machinery is illustrated in Fig. 4.9.

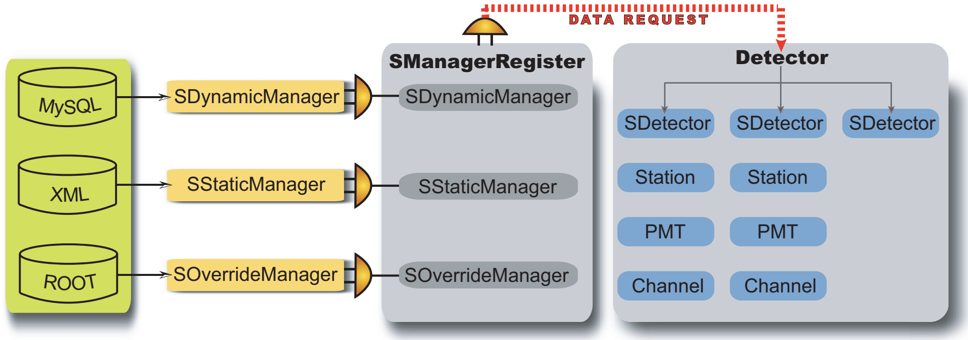
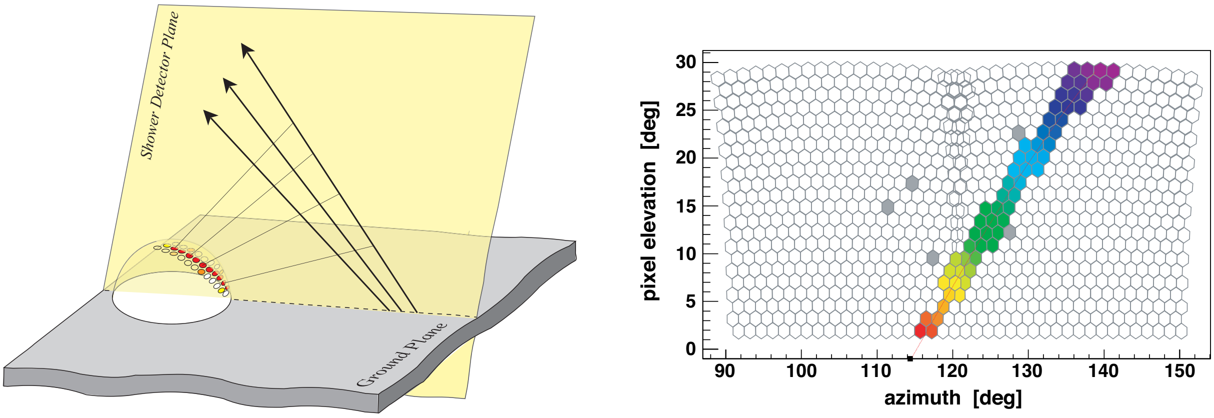


Figure 4.9: Detector description machinery of the Offline framework. An example of SD implementation is illustrated (cf. [172]).

4.4 Fluorescence geometry reconstruction

The geometry reconstruction of the shower axis, utilizing fluorescence light of EAS, was first successfully applied at the Fly's Eye experiment [123]. The basic principle did not change much over the years. The emitted fluorescence light along the shower axis appears as a sequential light track propagating across the night sky background starlight, man made civilization light and atmospheric air glow. The "hit pattern" of PMTs determines



(a) Illustration of the SDP. The geometry is reduced to a planar problem. Different showers still fit to one SDP.

(b) Light track of event 3308259 as seen by two adjacent fluorescence cameras (Los Morados). Different colors indicate the arrival time at the telescope. The filled black square at the bottom of the telescope denotes the position of the station used within the reconstruction.

Figure 4.10: Determination of the SDP.

a plane in space in which the trajectory of an EAS lies (cf. Fig. 4.10). This "shower detector plane" (SDP) is defined as the plane, containing the shower axis and the center of the eye. The reconstruction procedure mainly uses the trace of triggered pixels where high signal PMTs are expected to be more reliable than noisy ones. The orientation of the SDP is specified by a unit normal vector $\vec{n}$ referred to as the "SDP vector". Since every plane has two normal vectors, one opposite to each other, a convention is used to remove this ambiguity. The common definition is that the cross product of the SDP vector with the local vertical of the detector points in the direction of the core [174]. For this convention only, the core is defined as the intersection of the shower axis and the detector's horizontal plane. The direction of the shower is not taken into account, i.e. a vertical up-going laser shot and a vertical down-going shower at the same core location will have the same SDP. Within a χ^2 minimization the plane that best describes the triggered pixels is determined. The normal vector $\vec{n}$ is obtained using the pointing direction $\vec{r}_i$ of the i^{th} triggered phototube:

$$\chi^2 = \sum_i |\vec{n} \cdot \vec{r}_i|^2 \, w_i \ , \tag{4.1}$$

where w_i is basically[2] the sum of the signal found in pixel i.

Next, the geometry of the shower within this SDP is reconstructed based on the correlation between arrival time of the signals and viewing angle of the pixels projected into the SDP. The corresponding fit function is derived as follows. Assuming the fluorescence light to be emitted by a point-like object moving at c_{vac} along the shower axis vector, the

---

[2] There are also some corrections from studies on laser shots [175].

shower propagation time $\tau_{\mathrm{shower},i}$ from point S_i to the point at reference time t_0 on the shower axis (for angle definitions cf. Fig. 4.11 (a)) can be expressed as

$$\tau_{\mathrm{shower},i} = \frac{R_p}{c_{\mathrm{vac}} \cdot \tan(\chi_0 - \chi_i)} \ . \tag{4.2}$$

Next, assuming the fluorescence photons to propagate on straight lines with c_{vac}, the light propagation time $\tau_{\mathrm{light},i}$ from S_i to the telescope is

$$\tau_{\mathrm{light},i} = \frac{R_p}{c_{\mathrm{vac}} \cdot \sin(\chi_0 - \chi_i)} \ . \tag{4.3}$$

With Eqs. (4.2) and (4.3), and assuming an instantaneous emission of the fluorescence photons at S_i, the expected arrival time t_i (relative to the time t_0 of closest approach of the shower to the telescope) of fluorescence photons at a pixel viewing at an angle χ_i becomes

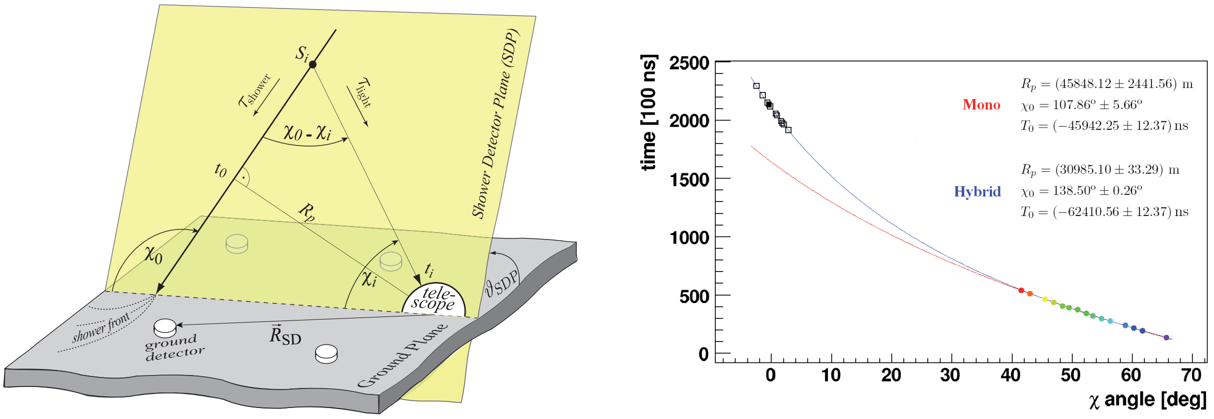
$$\begin{aligned} t_i &= t_0 - \tau_{\mathrm{shower},i} + \tau_{\mathrm{light},i} \\ &= t_0 + \frac{R_p}{c_{\mathrm{vac}}} \left(\frac{1}{\sin(\chi_0 - \chi_i)} - \frac{1}{\tan(\chi_0 - \chi_i)} \right) \\ &= t_0 + \frac{R_p}{c_{\mathrm{vac}}} \tan\left(\frac{\chi_0 - \chi_i}{2} \right) \end{aligned} \tag{4.4}$$

However, this commonly used Eqn. (4.4) for calculating the expected time-angle correlation is based on several simplifications like instantaneously fluorescence light production and propagation on straight lines at speed of c_{vac}. The validity of these assumptions were investigated in [176] finding typical corrections of 0.03–0.05° in arrival direction $\simeq$ 0.5–1% in energy and 2–3 g/cm^2 in $X_{\max}$ [176, 177].

Since the SDP can be reconstructed with high accuracy, the uncertainty mainly arises from the determination of the shower geometry within the SDP (FD-mono). The uncertainty of the fit parameters depends on the particular geometry as well as on the observed track length, e.g. for short track length there may be only insignificant curvature in the tangent function resulting in an ambiguity in the set of fit parameters χ_0, R_p and t_0. This translates directly into an uncertainty of the primary energy E_{prim} since to a good approximation the primary energy is proportional to

$$E_{\mathrm{prim}} \propto L_{\mathrm{fluor}} \propto L_{\mathrm{FD}} \cdot R^2_{X_{\max}} \cdot e^{R_{X_{\max}}/\lambda_{\mathrm{att}}} \ , \tag{4.5}$$

where L_{fluor} denotes the amount of light per unit length produced at shower maximum, L_{FD} the actually received light at the telescope, $R_{X_{\max}}$ the distance to shower maximum from the telescope and λ_{att} the attenuation length of fluorescence light. The quantity $R_{X_{\max}}$ is affected by changes in the parameters χ_0, R_p and t_0. The resulting asymmetric uncertainties are a drawback within mono fluorescence reconstruction.



(a) Illustration of the reconstruction of hybrid events.

(b) Viewing angle χ as a function of time and corresponding best fit for mono (red) and hybrid (blue) reconstruction of the same event, respectively. The corresponding fit parameters are indicated. Colored dots indicate triggered FD pixels and black squares SD timing information [22].

Figure 4.11: Illustration of the geometry reconstruction method and the advantage for hybrid events.

4.5 Hybrid geometry reconstruction

One of the key features of the Pierre Auger Observatory is the ability to detect high energy cosmic rays simultaneously by fluorescence telescopes and ground array. This hybrid detection can avoid the aforementioned ambiguities (mono-mode) and provides important cross checks and measurement redundancy. Much of the hybrid capability stems from the accurate geometrical reconstruction, better than either the ground array detectors or a single telescope. The synergy between both techniques can be seen in several examples:

- **Energy spectrum**:
  Due to the 100% duty cycle of the surface detectors together with a huge collecting area the energy parameter $S(1000)$ (which is the time-integrated water Čerenkov signal that would be measured by a tank 1000 m from the core [179]) can easily be calculated for the events. In order to convert $S(1000)$ into cosmic ray primary energy FD data is used, since it uses a near-calorimetric technique for determining energy. This has the advantage of being almost independent of the high energy hadronic interaction models used in simulations.

- **Mass composition**:
  The depth of shower maximum $X_{\max}$ is so far the most important parameter for mass composition studies. Hybrid data can therefore be used to calibrate and cross-check the search for new promising mass sensitive parameters measured by the SD alone [180].

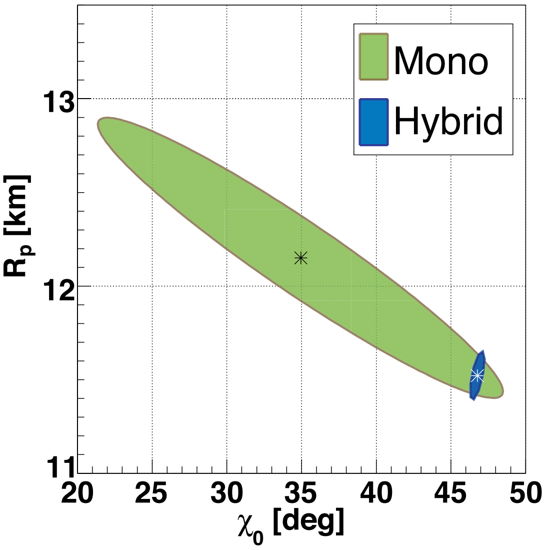


Figure 4.12: Uncertainty of the axis for mono and hybrid reconstruction (1σ accuracy) of a sample event. The large uncertainty of the monocular reconstruction is broken using the timing information from the surface detector. The stars indicate the solution that minimize the χ^2 for the axis reconstruction (cf. [178]).

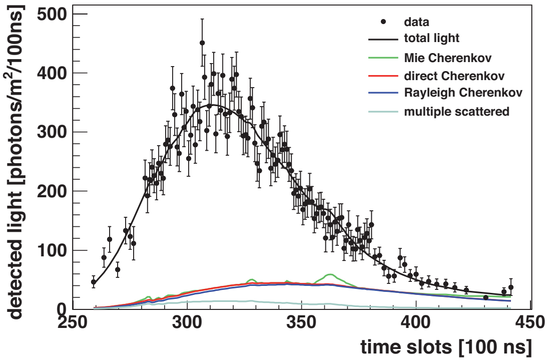


Figure 4.13: Different light contributions as a function of time. Data points and total light received are indicated in black. Other light contributions are shown as well.

- **Anisotropy studies**:
  Also in anisotropy studies hybrid data can be used to provide a sub-sample of high-precision shower arrival directions which, again, can be used to cross-check SD arrival directions and estimate SD angular resolution.

In order to achieve an improved geometry reconstruction, the SD information of the "hottest" tank is regarded as shown in Fig. 4.11 (b). The expected timing information from a hit ground station $t_{\mathrm{SD}}^{\mathrm{exp}}$ can be related to the reference time t_0, at which the shower passes the closest point to the telescope, by

$$t_{\mathrm{SD}}^{\mathrm{exp}} = t_0 - \frac{\vec{R}_{\mathrm{SD}} \cdot \hat{S}}{c} \,, \tag{4.6}$$

where $\vec{R}_{\mathrm{SD}}$ is the vector pointing from the telescope to the hottest SD tank and $\hat{S}$ the unit vector of the shower axis pointing towards the origin (cf. Fig. 4.11 (a)). In this expression it is assumed that the shower front is planar. In real situations the shower front curvature must be taken into account.
This additional information can be used as a supplemental data point for the timing fit as shown in Fig. 4.11 (b). The improvement is, that the data point is usually "far away"

in viewing angle from triggered FD pixels offering a long lever arm for the timing fit. The curvature can be expressed more accurately resulting in a better resolution, cf. Fig. 4.12.

The directional resolution can hereby be improved to be better than 0.5 deg making sensitive anisotropy searches possible as well as cross-checks of SD direction assignments [181].

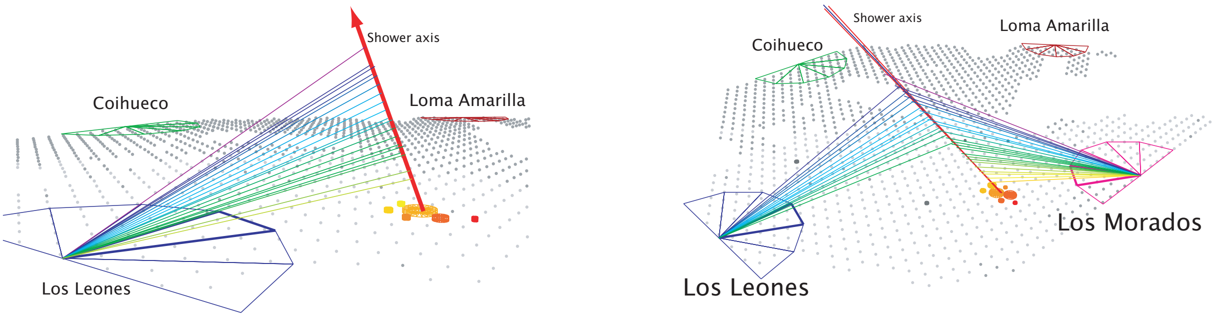
Showers that are triggered by the FD and one SD station are called “brass hybrid” events. Showers with at least three SD stations are called “golden hybrid” events. The geometry of those events can be independently reconstructed by either only the SD or FD information, or by combining this information using a hybrid technique. Also multiple eye events are possible and observed regularly as shown in Fig. 4.14.

4.6 Energy determination

Once the geometry is fixed, the light collected at the aperture is converted to energy deposit at the shower as a function of slant depth. For this conversion it is crucial to estimate the light attenuation from the shower to the telescope and all contributing light sources, like direct and scattered Čerenkov light or multiple scattered light, have to be disentangled as shown in Fig. 4.13. To estimate the calorimetric energy of the shower a Gaisser–Hillas function (Eqn. (3.7)) is fitted to the profile and integrated (Eqn. (3.14)). Since there is still some “invisible energy” carried away by neutrinos and high energy muons, a correction to the energy is applied. Finally the energy resolution[3] of the fluorescence detector is $\leq 10\%$ [181].

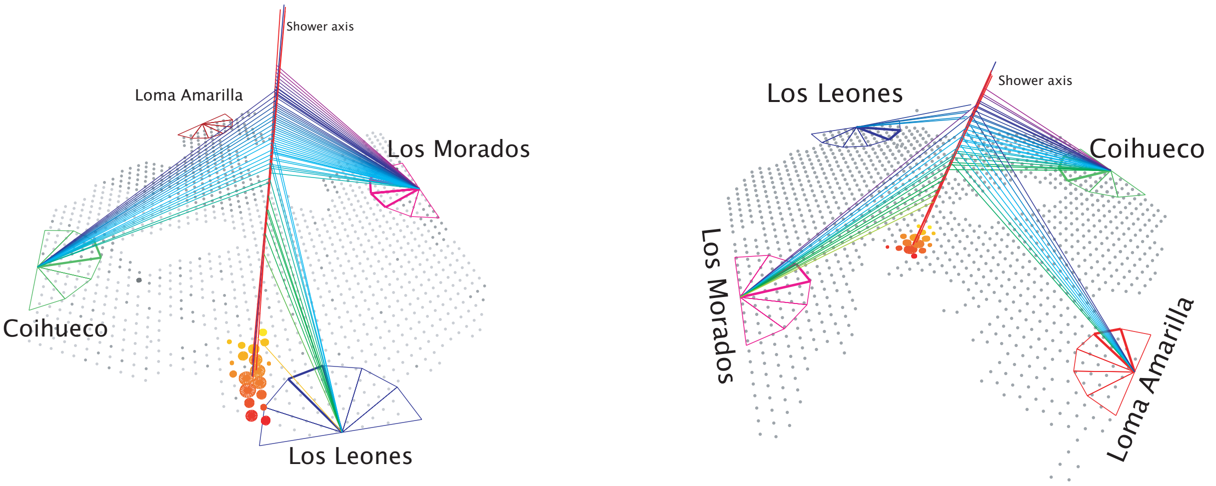
---

[3] Defined as event-to-event statistical uncertainty



(a) Single eye event indexed 3308445.

(b) Double eye event indexed 3031623.



(c) Triple eye event indexed 3336808.

(d) First quadruple event occurred on May 21st 2007 at 09:47:21 UTC with an energy of about 10^{19} eV.

Figure 4.14: Illustration of the geometry reconstruction method and the advantage for hybrid events.

Chapter 5

Ultra-High Energy Photons

Ultra-high energy photons are the main subject of this thesis. This chapter focusses on the highest energy part of the electromagnetic spectrum. First, a motivation for a search of high-energy photons is given before covering the possible life of ultra-high energy photons starting from production and origin explained in Sec. 5.1. Subsequently propagation of photons in interstellar space is addressed in Sec. 5.2 following flux expectations (Sec. 5.3) and features of ultra-high energy photon showers (Sec. 5.4). Detection and observables are sketched in Sec. 5.5. Finally, the experimental status and prospects are expressed in Sec. 5.6.

OBserving the universe is one of the oldest sciences. Already thousands of years ago ancient civilizations performed methodical observations of the night sky. The invention of the telescope was required before astronomy was able to develop into a modern science. However, a common feature is the observation of light or rather electromagnetic radiation coming from outer space reaching the Earth, making light the main messenger particle for exploring the universe. For thousands of years the observation of light was limited to the visible energy spectrum, but within the last century new technologies developed and the observed energy spectrum extends these days over a remarkable wavelength range:

- **Radio astronomy**:
  The wavelength for radio astronomy studies is greater than about one millimeter. In contrast to high energy observations, radio waves can still be treated as waves rather than discrete photons making it more easy to measure both the amplitude and phase. Most of the observed radio emission is seen in form of synchrotron radiation generated by charged particles in magnetic fields but also thermal radio waves produced by astrophysical objects contribute. Today a variety of objects is observed at radio wavelength including pulsars, interstellar gas, supernovae and

AGN, but also the hydrogen spectral line at 21 cm is observed in the radio frequency range.

- **Infrared astronomy**:
  The wavelength of infrared light ranges from about 0.75 to 300 microns. This radiation is heavily absorbed by the atmosphere making observations only possible at high, dry places or in space. The advantage of this energy range is the possibility to detect objects such as planets or circumstellar clouds which are too cold for optical astronomy. Also the observation of young stars in molecular clouds is possible since longer infrared wavelength can penetrate clouds and dust which block visible light.

- **Optical astronomy**:
  In the range of visible light and wavelengths between approximately 400 nm to 700 nm, optical astronomy is the oldest form of astronomy. Today images are made using digital detectors like charge-coupled devices (CCDs).

- **Ultraviolet astronomy**:
  At wavelengths between 10 to 320 nm ultraviolet radiation is mostly absorbed by the atmosphere making measurements on top or above the atmosphere necessary. This energy region is best suited to the study of thermal radiation and spectral emission lines. Common candidates for ultraviolet astronomy are planetary nebulae, supernovae remnants (SNR) or AGN. However, strong absorption by interstellar dust necessitates appropriate adjustments.

- **X-ray astronomy**:
  X-rays start at $\sim$ 8 pm and extend up to $\sim$ 8 nm. Observations are only possible at high-altitudes or in space. Typically the production mechanism is synchrotron emission by electrons oscillating in magnetic fields or thermal emission from very hot gases. There are a lot of X-ray sources identified including AGN, pulsars, X-ray binaries, SNR or clusters of galaxies.

- **Gamma-ray astronomy**:
  Gamma-rays have very small wavelengths of 10 pm and below. Up to now the current maximum energy of photons observed is $\sim 10^{14}$ eV [182]. They can be observed either directly by satellites or indirectly by atmospheric Čerenkov telescopes. These telescopes do not actually detect gamma-rays but rather Čerenkov light which is produced by secondary particles originally initiated by a gamma-ray interacting with the atmosphere. Gamma-ray emitters include pulsars, neutron stars and AGN. However, also flashes of short[1] gamma-rays have been observed which are even today not fully understood. These gamma-ray bursts are the most luminous electromagnetic events known to occur in the universe.

---

[1] from milliseconds to several minutes

It should be noted that all observation windows have their own features and discovery potential. Exploring a new window was always accompanied with new insights and astonishments. This chapter focusses on the extension of the already observed electromagnetic spectrum to photons with much larger energies at about 1 EeV (10^{18} eV) with a variety of prospects. A detection of EeV photons would again open a new window in cosmic-ray research with significant impact on astrophysics, particle physics, cosmology and fundamental physics. A more detailed summary on ultra-high energy (UHE) photons, from production to detection is given in [183].

5.1 Production and origin

The most prominent production mechanism of UHE photons is the decay of neutral pions produced previously by a "primary process" which leaves some wiggle room open. In conventional acceleration scenarios nuclear primaries are accelerated at suitable astrophysical sites to energies above 10^{18} eV before interacting in a primary process, cf. Sec. 2.5.2.

A more exotic production mechanism occurs in non-acceleration models where the primary process is given by the decay of primordial relics as already discussed in Sec. 2.5.1. Since these models predict a relatively large photon fraction observation or non-observation of UHE photons is a key feature to confirm or constrain so-called top-down models, cf. Sec. 5.6.

Some theoretical production processes are:

- Production of UHE photons in a **GZK-process** [101, 102, 184]. This scenario was already introduced in Eqn. (2.11) and the primary process is given by resonant photo-pion production:

  $$p + \gamma_{\text{CMB}} \rightarrow \Delta^{+}(1232) \rightarrow n + \pi^{0} \ ,$$

  producing UHE photons via $\pi^0 \rightarrow \gamma + \gamma$. The energy of these GZK photons is typically a factor ~ 10 below the primary nucleon energy [183].

- In [185] an enhanced photon flux $> 10^{18}$ eV from the **Galactic Center** region is predicted by nuclei or protons interacting with starlight and infrared photons.

- A diffuse anisotropic photon flux with about 10% increase in the direction of the **Galactic Center** region could come from decaying superheavy relic particles in the galactic halo, e.g. [74, 75, 186].

- A diffuse gamma-ray emission from the local supercluster is reported in [187]. The production scenario is motivated by the confinement of cosmic rays up to energies of 10^{19} eV in the intra-cluster medium. Subsequent photo-pion production and pair production / inverse Compton cascades, lead to a large-scale anisotropy at TeV and EeV energies along the **supergalactic plane**.

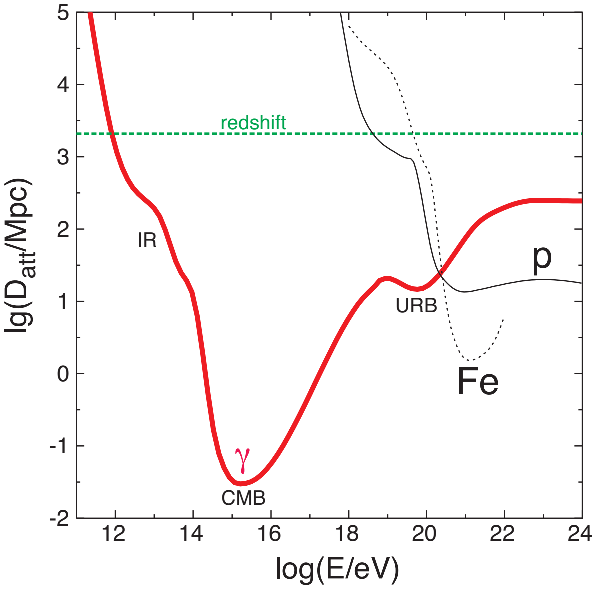


Figure 5.1: Attenuation length of UHE photons as a function of energy (thick red line) for interactions with cosmic microwave (CMB), infrared (IR) and radio (URB) background fields. The attenuation length for redshift evolution is indicated as well as the attenuation of primary protons (thin solid line) and iron (dotted line) nuclei. Modified from [183] and references therein.

- Is has also been suggested that a major UHECR flux may arise from just a few nearby AGN such as **Centaurus A** [188]. At a distance of 3.4 Mpc [189] Centaurus A is by far the nearest AGN. The importance of Centaurus A as well as UHE photon production scenarios is reported e.g. in [190].

5.2 Propagation

Since UHE photons have no charge they are not deflected by magnetic fields and thus point towards the production place. However, the existing cosmic photon background exacerbates the unhindered expansion of photons. The dominant interaction process is the attenuation of UHE photons γ_{UHE} due to pair production on background photons γ_{b} described as

$$\gamma_{\mathrm{UHE}} + \gamma_{\mathrm{b}} \rightarrow e^{+} + e^{-} \ .$$

The produced UHE $e^{\pm}$ can again interact with background photons via inverse Compton scattering resulting in an electromagnetic cascade that ends at GeV–TeV energies where the universe becomes increasingly transparent for photons as shown in Fig. 5.1. At EeV energies the most important photon fields are the cosmic microwave (CMB) and universal radio (URB) background. Typical energy loss length for UHE photons are assumed to be between 7–15 Mpc at 10^{19} eV and 5–30 Mpc at 10^{20} eV [183]. It should be noted, however, that large uncertainties exist for the low-frequency radio background because it is difficult to disentangle the galactic and extragalactic components [71]. Also the extragalactic magnetic field is poorly known and important for $e^{\pm}$ propagation [191, 184].

Another energy loss process is the adiabatic fractional energy loss due to the expansion of the universe, cf. Sec. 2.6.2 and Eqn. (2.13).

5.3 Flux expectations

Flux predictions are a dedicated task keeping in mind large uncertainties described in Sec. 5.2. Not only uncertainties in source and propagation models contribute, but also the absolute cosmic-ray flux differs between experiments by a factor ~ 2 [183]. Even when just regarding the fraction of photons, the shape of the assumed energy spectrum affects also the predicted fraction of photons [192]. Assuming a spectrum with flux suppression and nucleon sources, the predicted photon fluxes from photo-pion production are typically of the order of $\sim 0.1\%$ [193]. Larger photon fractions may occur in specific regions in the sky particularly when modifying source features and its environment.

5.4 Features of UHE photon showers

Since the flux of UHE photons is way to low for direct detection, UHE photons can only be detected by EAS. Giant air shower arrays such as the Pierre Auger Observatory are unique tools to explore this photon energy range. A common feature for all photon induced air showers is the almost purely electromagnetic cascade via pair production and bremsstrahlung. The production of muon pairs is suppressed by $(m_e/m_\mu)^2$. Since hadronic interaction dynamics are poorly known at highest energies[2] photon induced showers tend to be more predictable, e.g. differences in shower maximum X_{max} are only ~ 5 g/cm^2, compared to 30–40 g/cm^2 for UHE protons [183] between SIBYLL and QGSJET, cf. Fig. 2.5.

However, besides the already discussed features additional processes for photon induced EAS have to be considered.

- **Preshower effect**: Very high energy ($\sim 10^{20}$ eV) photons can convert in the geomagnetic field of the Earth to an $e^{\pm}$-pair emitting synchrotron radiation resulting in an earlier shower development (smaller X_{max}) as discussed e.g. in [195, 196, 194, 183]. The subsequent bunch of lower energy electromagnetic particles enters the atmosphere with significant implications on the shower development. The local differential conversion probability χ certainly does not only depend on energy of the parent particle E, but also on the local magnetic field component transverse to the direction of the particle's motion $B_\perp$ and thus on the specific trajectory of the particle through the atmosphere. The probability is given by

$$\chi = \frac{E}{mc^2}\frac{B_\perp}{B_c} \qquad (B_c \approx 4.414 \cdot 10^{13}\ \text{G}) \tag{5.1}$$

---

[2] this is a major limitation for conclusions on nuclear composition

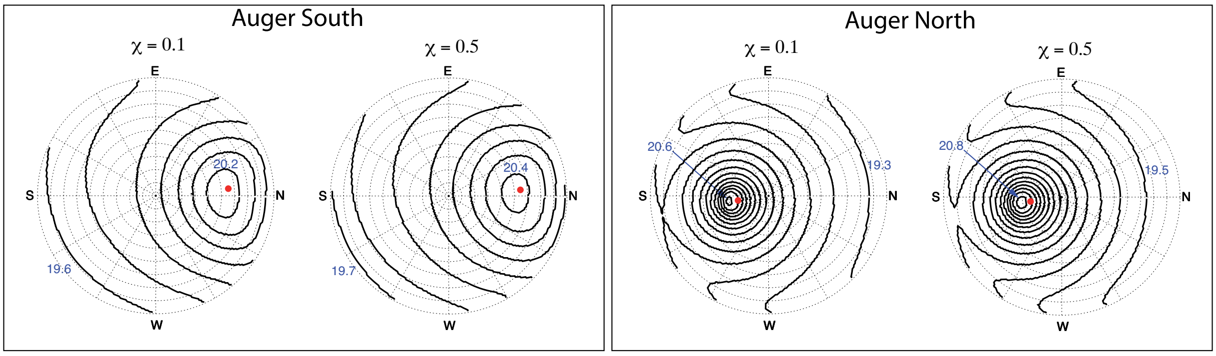
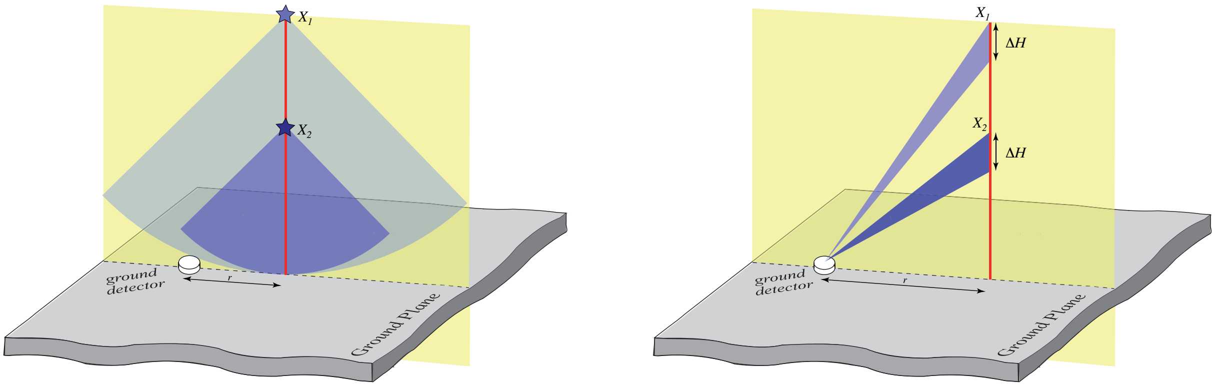


Figure 5.2: Photon energies at which the conversion probability is $\chi = 0.1$ and $\chi = 0.5$ as a function of arrival direction for the Southern (left panel) and Northern (right panel) Pierre Auger Observatory, respectively. The pointing direction of the local magnetic field is given by a red dot. Contour lines indicate a constant energy with a step-size of 0.1 $\log(E/\mathrm{eV})$. Concentric circles indicate constant zenith angles (step-size 10°) with the zenith in the center. At EeV energies the conversion is negligible, independent of the local direction; modified from [183, 194].

and illustrated for two sites (Auger North and Auger South) in Fig. 5.2. The final probability P_{conv} of a photon to convert in the geomagnetic field is given by integration on the particles trajectory. Negligible probabilities $P_{\mathrm{conv}} < 10\%$ are usually obtained if values do not exceed $\chi < 0.5$ along the trajectory, corresponding to photon energies below 2–3·10^{19} eV. It should be stressed that the conversion probability at EeV energies is negligible as can be inferred from Fig. 2.5 and Fig. 5.2.

- **LPM effect**: Lev Landau and Isaak Pomeranchuk showed in the early fifties that the formulas for bremsstrahlung and pair creation in matter which had been formulated by Hans Bethe and Walter Heitler [197] (Bethe-Heitler formula) were inapplicable at high energy or high matter density [198, 199]. A couple of years later Arkady Migdal developed a formula applicable at high energies or high matter densities which accounted for these effects [200]. Today this effect is known as Landau-Pomeranchuk-Migdal effect or LPM effect. The basic principle is that the Bethe-Heitler cross-section σ_{BH} for pair production by photons can be reduced due to destructive interference from several scattering centers. Ultra-relativistic electromagnetic interactions involve very small longitudinal momentum transfers. However, reactions occur gradually, spread over long distances. During this time, even relatively weak factors can accumulate enough to disrupt the interaction resulting in a reduction of σ_{BH} [201]. As a consequence the LPM effect delays the shower development giving larger X_{max}. As can bee seen in Fig. 2.5, the LPM effects starts to be important at a few times 10^{18} eV. Since there is a positive correlation $\sigma_{\mathrm{LPM}}(X_2) < \sigma_{\mathrm{LPM}}(X_1)$ for depth $X_1 < X_2$ fluctuations can be very large.



(a) Illustration of radius of curvature. Shower particles arrive more delayed at distance r from the shower axis originating from a deeper atmospheric depth X_2 compared to particles produced at depth X_1 with $X_1 < X_2$. Correspondingly, the radius of curvature is smaller for deep developing photon primaries.

(b) Illustration of rise time. The spread of arrival times at distance r from the shower axis, produced over a path length ΔH increases with deeper production depth $X_2 > X_1$. Correspondingly, the spread of arrival times for photon induced showers is increased.

Figure 5.3: Illustration radius of curvature and rise time.

5.5 Detection and observables

To differentiate between photon and hadron induced EAS a detailed knowledge on shower development and experimental feasibility is needed. Individual characteristics may lead to powerful discriminating observables which can be used to separate UHE photon from background data. The two most important features of EAS induced by a primary photon is the delayed shower development resulting in a deeper X_{max} and the lack of muons due to the smaller photonuclear cross-section. The longitudinal shower development and thus X_{max} is a direct observable of fluorescence telescopes whereas the number of muons is typically measured by ground arrays. The combination of the two techniques in a hybrid approach – as it is realized with the Pierre Auger Observatory – is therefore an excellent way to search for UHE photons. Commonly used observables utilizing the aforementioned characteristics are e.g.:

- **Depth of shower maximum**: The depth of shower maximum X_{max} is measured directly by *fluorescence telescopes*.

- **Radius of curvature**: This observable is derived by geometrical reasons exploiting the fact that the photon showers develop deeper in the atmosphere compared to nuclei primaries. Showers that develop deeper in the atmosphere (light particles) will have larger time delays t – and hence smaller radius of curvature – at fixed distance r to the shower axis compared to showers where the registered particles originated from larger heights (heavy particles) as shown in Fig. 5.3 (a). The effect is even amplified by the lack of muons in photon induced showers. This stems from the fact that shower muons can reach the ground from still higher altitudes further reducing the time delay. This quantity is measured by *surface detectors*.

- **Rise time of the signal**: Not only the curvature of the shower front but also the spread in time of the signal can be used to discriminate between photon and hadron induced showers. Assuming a fixed distance to the shower axis a larger spread of the signal intensity is expected in case of deep developing photon primaries. This can be understood when regarding the particle production at a fixed path length ΔH arriving at distance r from the shower axis at two different production depths $X_1 < X_2$ as shown in Fig. 5.3 (b). Consequently the rise time of the signal is increased for photon primaries developing deeper in the atmosphere. This parameter was used e.g. in [81] and the rise time of one *surface detector* at distance r_i from the shower axis is defined as the time $t_{1/2}^{\mathrm{meas}}(r_i)$ it takes to increase from 10% to 50% of the total signal. However, it should be noted that, in general, the situation is more complex since there are other dependencies on the specific shower geometry and on details of the previous shower development.

- **Steepness of the lateral distribution**: The measured lateral expansion of triggered *surface detectors* is correlated to its longitudinal development. Until the core remains active[3] the spread of the electromagnetic component increases. Since muons are produced early and their space distribution flattens continuously independent of the electromagnetic development, photon showers give a steeper distribution with, however, larger fluctuations [202].

5.6 Experimental status and prospects

No UHE photon detection has been reported so far. However, upper limits on the fraction of UHE photons have been set by various experiments as shown in Fig. 5.4. The prediction of the photon fraction for different non-acceleration models are indicated. Current limits already exceed the predicted photon flux placing stringent limits on top-down models. The expected photon fraction from GZK processes is shown as a purple band. All cuurent experimental limits refer *either* to ground array data *or* to fluorescence telescopes. A combination of both detection systems and hence a synergy of surface array and fluorescence telescope observables in a multivariate approach is performed for the first time in this thesis.

An overall estimate of current and future sensitivities of the Southern Pierre Auger Observatory is shown in Fig. 5.5. As can be seen, if the GZK process is real, the Pierre Auger Observatory will detect UHE photons within the next years if the primary composition consists mainly of protons. Details in the calculation are given in [183].

A detection of UHE photons would extend the already measured electromagnetic spectrum to several orders of magnitude with significant impact on astrophysics, particle physics, cosmology and fundamental physics. They can be used as a diagnostic tool for

---

[3] this can be even beyond X_{max}

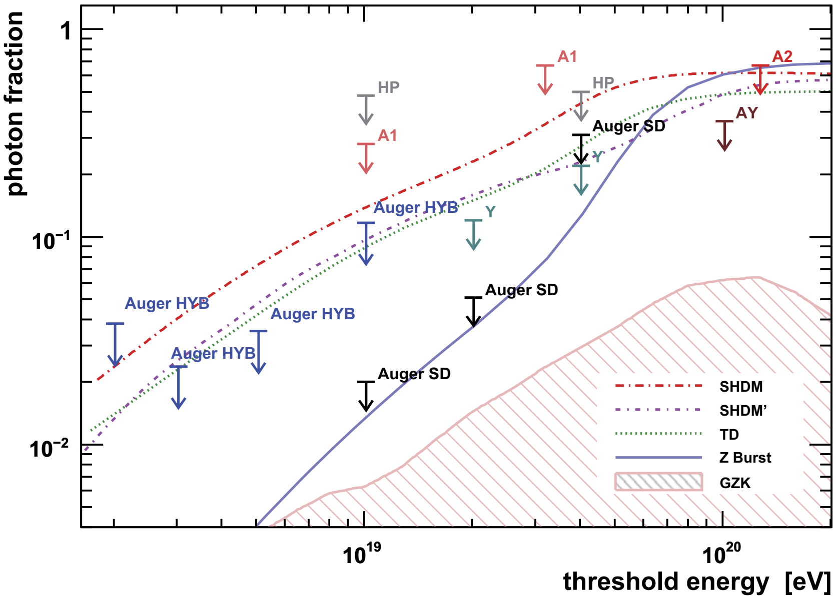


Figure 5.4: Integrated fraction of photons as a function of threshold energy. The predicted photon fraction of various non-acceleration models is indicated. Additionally the expected GZK photon fraction is illustrated as a shaded region calculated from [203]. Experimental limits are shown from Auger hybrid (Auger HYB) [82], Auger SD (Auger SD) [81], AGASA (A1, A2) [204, 205], AGASA-Yakutsk (AY) [206], Yakutsk (Y) [207] and Haverah Park (HP) [208, 209] observations. Modified from [82]

sources *accelerating nuclear primaries* since the production of photons is connected to source features [184]. Photons point back to their location of production making *astronomy* at UHE possible. The photon flux at Earth is also sensitive to the structure and strength of *extragalactic fields* [211]. The detection of UHE photons would give valuable insights of aspects to *QED and QCD at ultra-high energy* via preshower processes and photonuclear interactions. Current extrapolations to highest energies could be checked [212, 213].

Not only UHE photon detection but also setting upper limits on the photon flux already have significant impact on fundamental physics. According to theories as suggested e.g. in [214] there might be a departure from strict Lorentz invariance at highest energies. The basic principle is that *Lorentz invariance violation* (LIV) would lead to photon dispersion and suppress the GZK photon cascading process at energies around 10^{19} eV

$$\pi^0 \rightarrow \gamma + \gamma \nrightarrow e^+ + e^- \quad , \tag{5.2}$$

resulting in an enhancement of the photon fraction. However, current limits already constrain the expected photon flux improving the limit on LIV by seven orders of magnitude [215, 216, 217, 218].

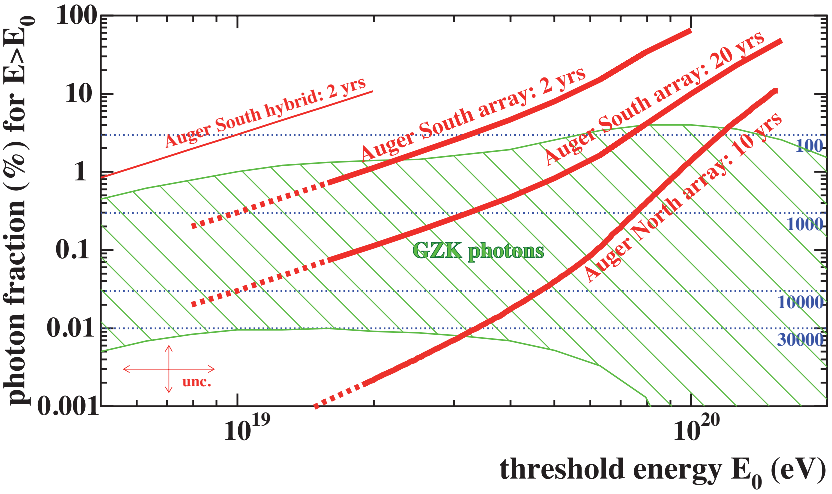


Figure 5.5: Pierre Auger sensitivity [183] on the integrated photon fraction as a function of threshold energy. The predicted fraction of GZK photons assuming a spectrum with flux suppression is indicated and obtained from [184]. Several sensitivity scenarios from the Southern and Northern Pierre Auger Observatory are shown. The uncertainty of the sensitivity estimate is indicated in the lower left corner. At lower threshold energies at about 10^{19} eV, additional effects may become increasingly important for the array (dashed lines) [183]. To place a limit to a certain photon fraction at 95% c. l., horizontal dotted lines with the minimum number of events required are shown. Modified from [183, 210].

Chapter 6

Simulation of UHECR Particle Propagation

In this chapter particle propagation at ultra-high energy is revisited. The utilized software package CRPropa *is introduced in Sec. 6.1. Particle interactions en route to Earth are discussed in Sec. 6.2 and the effects on the observed spectra in Sec. 6.3. Important for anisotropy studies, the GZK-horizon is analyzed in Sec. 6.4 and the dependence on source and propagation properties is illustrated. The question if the current photon flux limit already constrains scenarios in which the total flux above 57 EeV comes from Centaurus A is addressed in Sec. 6.5. Finally a brief summary is given in Sec. 6.6.*

The origin, composition and acceleration mechanism of ultra-high energy cosmic rays (UHECR) are still unknown. Understanding UHECR is a dedicated task requiring to model in a realistic way their propagation in the universe. To get a clue of an answer to the raised questions it is therefore desirable to expand the knowledge about particle propagation through the local universe. The photon background is a key ingredient in understanding the properties of particle propagation as discussed in Sec. 2.6. The produced electron/positron pairs as well as pions give rise to subsequent neutrinos and electromagnetic cascades extending down to MeV energies and opening up multi-messenger observations. The interplay between different astroparticle physics experiments has become very important. Neutrinos, γ-rays, cosmic ray physics and magnetic fields are strongly linked subjects and should be used together to extract maximal information from existing data. Current and planned projects range from UHECR observations like the Pierre Auger Observatory, to neutrino telescopes [219, 220], as well as ground and space based γ-ray detectors operating at TeV and GeV energies, respectively [221]. Even if a putative source were to produce exclusively UHECR up to the GZK-domain, photo-pion and pair production by protons on the photon background would lead to guaranteed fluxes of secondary photons and neutrinos that could be detectable.

6.1 CRPropa framework

With the aforementioned motivation a numerical tool called CRPropa [99] has been developed that can treat the interplay between UHECR, γ-ray and neutrino astrophysics, and large scale magnetic fields. CRPropa is able to follow the propagation of cosmic nucleons in magnetic fields, their interactions on photon backgrounds, and the propagation of secondary neutrinos and electromagnetic cascades. The publicly-available numerical package[1] allows to compute the observable properties of UHECRs and their secondaries for a variety of models for the sources and propagation of these particles. CRPropa inherits from various codes that have been previously developed like a Fortran code for the propagation in magnetic fields, by M. Lemoine and G. Sigl [222], a `C` code for the development of electromagnetic cascades, by S. Lee [223] and the public SOPHIA event generator for pion production [224]. Simulation parameters are defined in XML configuration files. A brief summary on steering options as well as output files is given in App. A.

Motivated by recent experimental results suggesting a heavier composition at the highest energies (e.g. [58]), the existing CRPropa framework is currently extended to allow for propagation of nuclei up to iron [225]. The most important features of the framework are summarized below.

6.1.1 Nucleon interactions

Pion production is modeled by using the event generator SOPHIA [224] that has been explicitly designed to study this phenomenon and is augmented in CRPropa for interactions with a low energy extra-galactic background light (EBL). Unlike pion production, pair production by protons is taken into account as a continuous energy loss due to the low inelasticity. More details on the specific spectrum of the pairs and applied approximations are given in [99]. The energy threshold down to which nucleons can be propagated is 10^{17} eV. Interactions at lower energies are negligible.

6.1.2 Secondary electromagnetic cascades

The electromagnetic (EM) cascade code is based on [223]. All relevant interactions with a background photon are taken into account and implemented in CRPropa including single pair production, double pair production, inverse Compton scattering and triplet pair production. If magnetic fields are selected, synchrotron losses of electrons are taken into account as well and the resulting lower energy synchrotron photons which are also followed in the subsequent EM cascade. The cascade stops until either the energy drops below 100 MeV or they reach an observer.

---

[1] http://apcauger.in2p3.fr/CRPropa/index.php

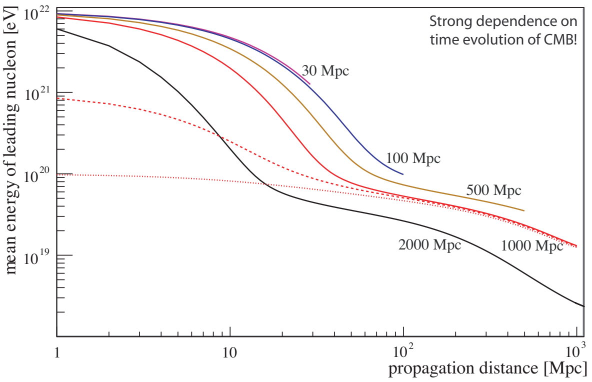


Figure 6.1: Mean energy of the leading nucleon as a function of propagated distance. The dashed and dotted line represents a primary energy of 10^{21} and 10^{20} eV, respectively. Since the time evolution of the background photon spectra is taken into account, attenuation is stronger for distant (earlier in time) sources, as shown for 10^{22} eV initial energy events.

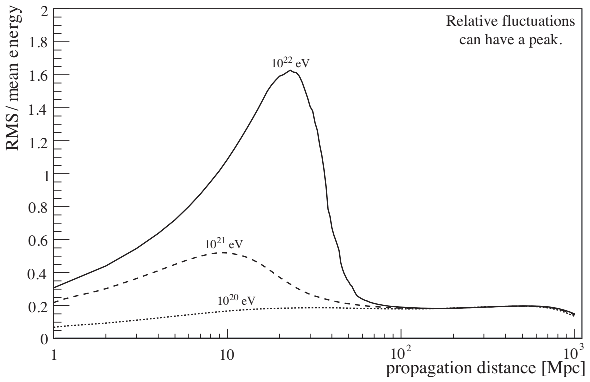


Figure 6.2: Ratio of RMS fluctuations of energies to mean energy as a function of propagation distance (time) for the indicated initial energies (corresponding to the lines labeled 1000 Mpc in Fig. 6.1.

6.1.3 Background photon spectra and their evolution

There are three different photon backgrounds implemented in CRPropa. The most important is the CMB with a well known redshift evolution. Three different infrared background (IRB) distributions can be chosen which are all consistent with recent limits from blazar observations in TeV by H.E.S.S. [226]. They become important for EM cascades around the threshold for pair production and are less significant in the UHE region. Above $\simeq 10^{18}$ eV interactions with the universal radio background (URB) become more important where it can inhibit cascade development due to the resulting small pair production length, cf. Fig. 5.1. The redshift evolution of the IRB and URB is more complicated and described in [99].

6.2 Interactions en route to Earth

In the following a one dimensional simulation is used to calculate the attenuation of a primary proton when propagating through the intergalactic background light (cf. [227]). All relevant energy losses (see above) are implemented. At a fixed distance from the observer, 60000 individual protons are injected and their energy loss is monitored with a stepsize of 1 Mpc. Three different primary energies of 10^{20}, 10^{21} and 10^{22} eV are simulated. The mean energy of the leading nucleon as a function of propagation distance is shown in Fig. 6.1. After a distance of ∼100 Mpc the mean energy is essentially independent of the initial energy of the protons and that energy is less than 10^{20} eV. However, as

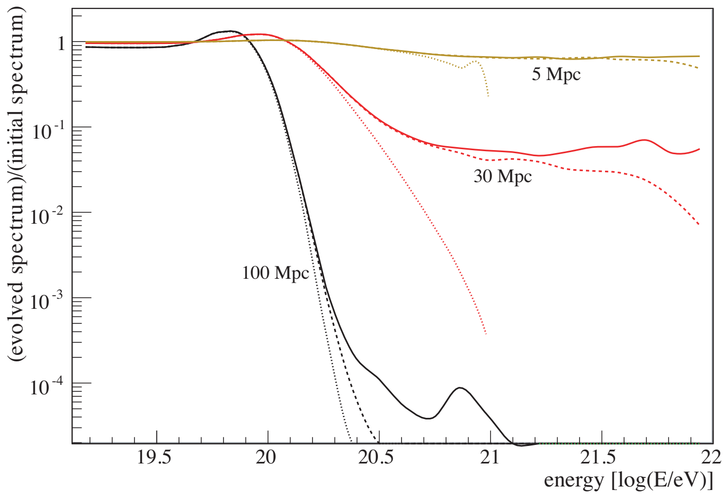


Figure 6.3: Modification factor $f(E)$ of proton sources located at 100, 30 and 5 Mpc, respectively. The maximum simulated energy E_{max} is 10^{23} (solid line), 10^{22} (dashed line) and 10^{21} eV (dotted line). The "bump" preceding the GZK-suppression is more pronounced for distant sources. A plateau beyond the GZK-suppression becomes visible for larger E_{max} or closer distances. Note that the "wiggles" are indicative of the Monte Carlo statistics.

a consequence of the time evolution of the CMB, the attenuation is stronger earlier in time with respect to a nearby source. This is shown for initial energies of 10^{22} eV where several distances (points in time) are illustrated. This effect starts to be significant at distances above 100 Mpc. The ratio of the RMS energy fluctuations to the mean energy as a function of propagation distance is shown in Fig. 6.2. For propagated distances in the range between $\sim 5-40$ Mpc, these fluctuations are very significant. This can significantly alter an energy spectrum based on a low number of events (cf. Fig. 6.3), and should be taken into account when interpreting the observed spectrum.

6.3 Effect on observed spectra

The observed energy spectrum depends on the spatial distribution and the input spectrum of the sources. For the highest-energy part of the spectrum, the bulk of particles originates from relatively nearby sources (<100 Mpc) and hence the redshift evolution of the CMB and the sources becomes negligible. In Fig. 6.3 the modification factor for different sources is illustrated. The modification factor $f(E)$ is given by

$$f(E) = \frac{I_p(E)}{I_0(E)} \, , \tag{6.1}$$

where $I_0(E)$ is the injected spectrum and $I_p(E)$ is the spectrum as modified by the background light. As can be seen, if the observed particles have an extragalactic origin, the

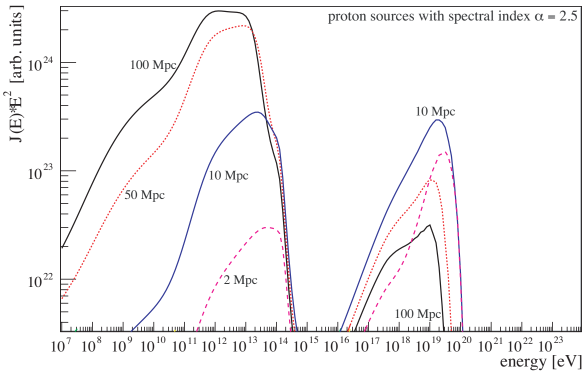


Figure 6.4: Spectrum of secondary photons generated by pion and pair production from a single UHECR *proton* source at a given distance. Here a one-dimensional model is considered, with an injection spectral index α=2.5 and maximum energy of $10^{20.5}$ eV. No magnetic fields were taken into account. In this example, GZK-photons are mainly observed within a propagation distance of up to $25 - 50$ Mpc.

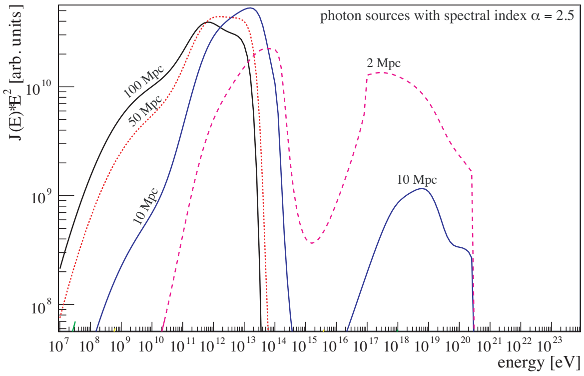


Figure 6.5: Spectrum of secondary photons from a single UHECR *photon* source at a given distance. Here a one-dimensional model is considered, with an injection spectral index $\alpha = 2.5$ and maximum energy of $10^{20.5}$ eV. No magnetic fields were taken into account.

interaction with the background light can dramatically change the original spectral distribution of accelerated particles injected into the intergalactic medium. By measuring the spectrum at highest energies the shape gives constrains on the maximum energy of sources at a given distance (cf. Fig. 6.3). Moreover, the "bump" preceding the GZK-suppression is more pronounced for distant sources. A plateau beyond the GZK threshold becomes visible for larger $E_{\max}$ or closer distances.

As described in Sec. 6.1.2 electromagnetic cascades are evolved and propagated to the observer. Fig. 6.4 illustrates the resulting spectra for a UHECR proton source at a given distance to the observer (with spectral index $\alpha = 2.5$). In this example, GZK-photons are mainly observed within a propagation distance of up to $25 - 50$ Mpc. Most of the contribution arises from nearby (< 25 Mpc) sources with a peak at about 10 Mpc. More distant sources have the main contribution in the TeV range. In Fig. 6.5 a photon source is assumed (also with spectral index $\alpha = 2.5$). At source distances close to the observer the largest EeV photon flux is expected.

6.4 GZK-horizon

Given the directional correlation of UHECR with relatively nearby AGN observed by the Pierre Auger Observatory [58, 59], it is interesting to investigate the "GZK-horizon". The GZK-horizon reflects that distance, within which the major part of the observed events should be produced and is therefore an important parameter for anisotropy studies (cf.

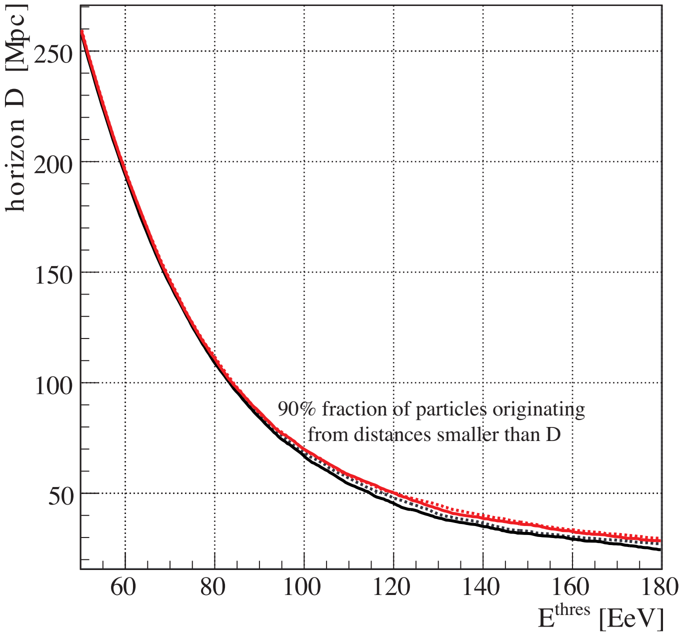


Figure 6.6: GZK-horizon as a function of threshold energy for varying maximum energies of 500 EeV (solid, black), 1000 EeV (dotted, black), 1500 EeV (solid, red) and 10000 EeV (dotted, red).

[58]). Within the present analysis the horizon is defined as the distance within which 90% of the observed events above a certain energy threshold E^{thres} were originally produced. In this simulation sources are distributed uniformly up to a distance of 800 Mpc. Unless stated otherwise, default values are $E_{\mathrm{max}} = 10^{21}$ eV, $\alpha = 2.7$ and $H_0 = 71$ km s^{-1} Mpc^{-1}. The GZK-horizon as a function of threshold energy is shown in Fig. 6.6 for varying maximum energies.

The calculated GZK-horizon at a threshold energy of $6 \cdot 10^{19}$ eV, where the correlation has maximum significance, is about 190 Mpc. This is to be compared with a value of ~ 210 Mpc from [228]. Compared to the distance $D_{\mathrm{max}} \simeq 75$ Mpc, where the correlation is most significant, a deviation of more than a factor 2 is observed. If these numbers were to be taken at face value, an upward shift in the energy calibration of $\sim$ 30%, as suggested in some simulations of the reconstruction of the shower energies [229], would lead to a better agreement between the maximum AGN distance D_{max} that maximizes the correlation signal and the theoretical expectations based on the idealized GZK attenuation [58]. However, as also noted in [58], D_{max} may not directly be comparable to the GZK-horizon (for instance, an accidental correlation with foreground AGN different from the actual source may induce some bias in the value of D_{max} toward smaller maximum source distances).

The effect on different input parameters is shown in Fig. 6.7. Fig. 6.7 *(left panel)* indicates the effect on the maximum energy. Larger E_{max} result in a more distant horizon for a growing energy threshold. Differences are of the order of $\sim$ 5% for $E^{\mathrm{thres}} > 120$ EeV

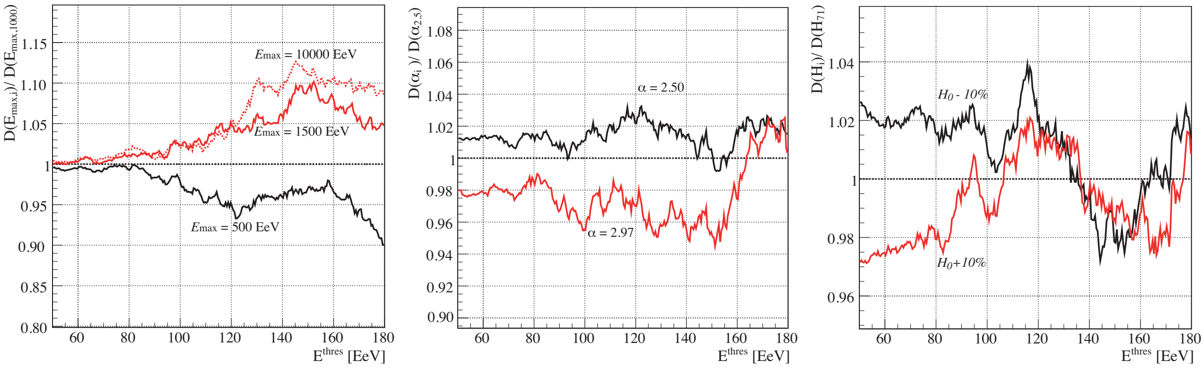


Figure 6.7: ***Left panel:*** Ratios of GZK-horizons shown in Fig. 6.6 normalized to $E_{\max} = 1000$ EeV. Larger $E_{\max}$ produce a more distant horizon for a growing energy threshold. ***Middle panel:*** Ratios of the GZK-horizon for varying spectral indices normalized to $\alpha = 2.7$. A spectral index of 2.5 (black) and 2.97 (red) is shown. Varying the spectral index induce a more constant (energy independent) offset in the horizon. ***Right panel:*** Ratios of the GZK-horizon for a varying Hubble parameter normalized to $H_0 = 71$ km s^{-1} Mpc^{-1}. A 10% variation H−10% (black) and H_0+10% (red) is shown. For lower energy thresholds the variations seems to dominate.

compared to the default values. Fig. 6.7 *(middle panel)* illustrates the dependence of the horizon to the spectral index at the source. A more constant offset (energy independent) of about 2% is induced. The effect on the Hubble parameter H_0 is shown in Fig. 6.7 *(right panel)*. For lower energy thresholds, the effect seems to be of the order of $\sim 2\%$.

6.5 GZK photon fluxes from Centaurus A

As shown in Fig. 6.4 a region around the source exists where the UHE photon flux is maximal. For closer distances (e.g. 2 Mpc) the GZK effect does not yet efficiently produce UHE photons, whereas for larger distances (e.g. 50 Mpc) the UHE photon population may go into the development of a full electromagnetic cascade with the main flux arriving at GeV-TeV energies. UHE photons can therefore provide information on local UHECR sources.

It has also been suggested that a major UHECR flux may arise from just a few nearby AGN such as Centaurus A [188]. At a distance of 3.4 Mpc [189] Centaurus A is by far the nearest active radio Galaxy. The importance for Centaurus A in high energy astrophysics as a nearby object with many of the properties expected of a major source of very high energy cosmic rays and gamma-rays is summarized e.g. in [190]. Given the current limits on the UHE photon fraction shown in Fig. 5.4 one might wonder if existing limits can already constrain Centaurus A as a pure proton source [230].

In this study a one-dimensional CRPropa simulation is performed for distances of 3.4 Mpc i.e. the distance to Centaurus A. Here we assume a proton source with spectral

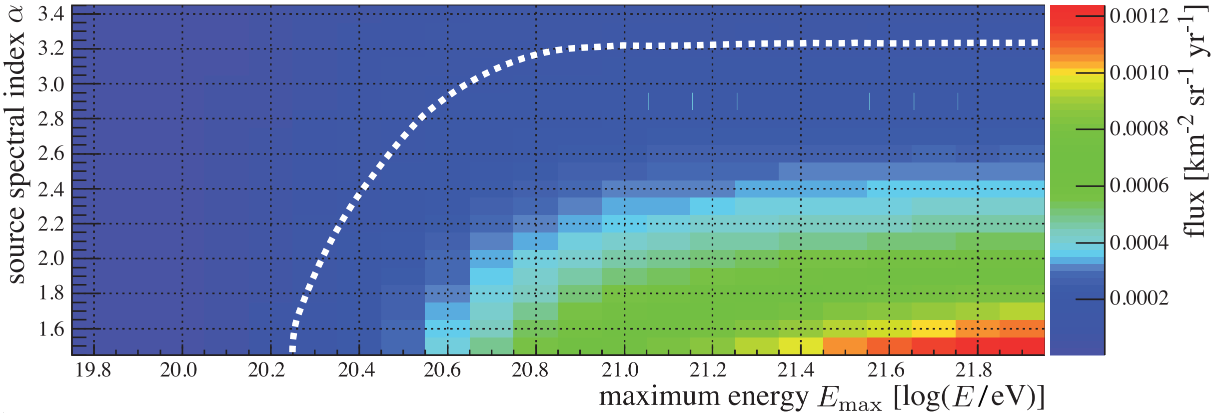


Figure 6.8: Simulation matrix of varying source parameters of Centaurus A (i.e. at a distance of 3.4 Mpc). A changing spectral index is shown on the y-axis in combination with a varying maximum energy of the source. Color coded is the expected photon flux $F_\gamma^{10\mathrm{EeV}}$ above a threshold energy of 10 EeV. The upper limit on the flux of photons above 10 EeV derived in [81] is $F_{10\mathrm{EeV}}^{\mathrm{limit}} = 3.8 \cdot 10^{-3}$ km^{-2} sr^{-1} yr^{-1}. The white dashed line indicates a photon flux level of $F_{10\mathrm{EeV}}^{\mathrm{limit}}/30$, where the factor 30 is a rough estimate of how much the final total exposure of Auger South exceeds the exposure used for $F_{10\mathrm{EeV}}^{\mathrm{limit}}$. One sees that interesting combinations of α and $E_{\max}$ can be tested by searching for the expected photon flux.

index α and maximum energy $E_{\max}$. Protons are injected and propagated (assuming no magnetic field) towards the observer. The resulting EM spectra are recorded and weighted according to the Auger flux spectrum [13] as follows: we assume in a first simplified step that the total flux above 57 EeV originates from a source at a distance of 3.4 Mpc, i.e. in this case 27 hadron events above 57 EeV with 7000 km^2 sr yr exposure (as observed in [13]). The ratio β_{7000} of the observed 27 events to the simulated total number of arriving particles above 57 EeV is then also used to scale the simulated photon flux above a certain energy E (here we take $E = 10$ EeV), i.e.

$$N_{\gamma,7000}^{10\mathrm{EeV}} = N_{\gamma,\mathrm{sim}}^{10\mathrm{EeV}} \cdot \beta_{7000} \ , \tag{6.2}$$

where $N_{\gamma,\mathrm{sim}}^{10\mathrm{EeV}}$ is the simulated number of photons above 10 EeV and $N_{\gamma,7000}^{10\mathrm{EeV}}$ the number of photons that are expected to be observed above 10 EeV with an exposure of 7000 km^2 sr yr [13].
The expected integrated γ-flux above 10 EeV, $F_\gamma^{10\mathrm{EeV}}$, is then calculated via

$$F_\gamma^{10\mathrm{EeV}} = \frac{N_{\gamma,7000}^{10\mathrm{EeV}}}{7000\ \mathrm{km}^2\ \mathrm{sr}\ \mathrm{yr}} \ . \tag{6.3}$$

In Fig. 6.8 the expected photon flux $F_\gamma^{10\mathrm{EeV}}$ is shown for varying source parameters α and $E_{\max}$. Compared to the upper limit on the flux of photons above 10 EeV derived in [81] of $F_{10\mathrm{EeV}}^{\mathrm{limit}} = 3.8 \cdot 10^{-3}$ km^{-2} sr^{-1} yr^{-1} all simulated source parameter combinations

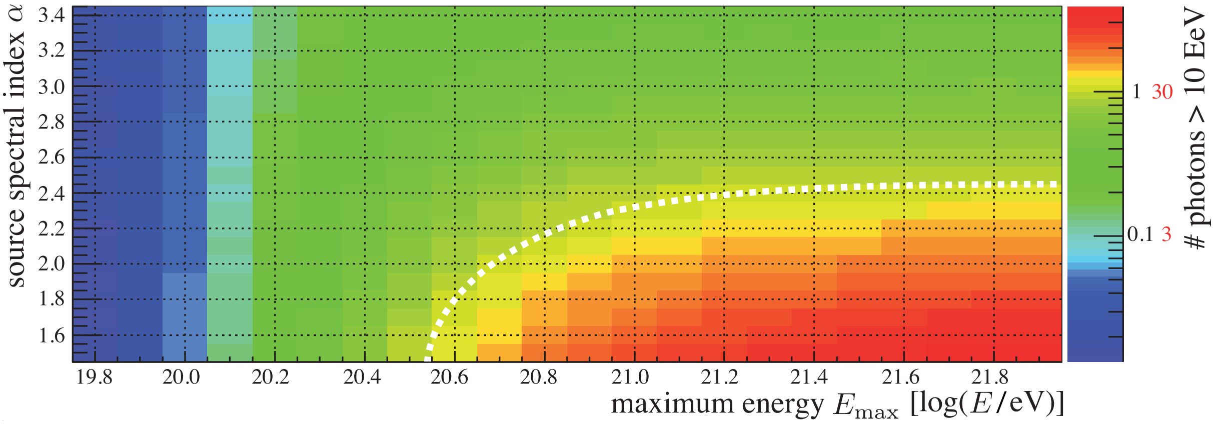


Figure 6.9: Simulation matrix of varying source parameters. A changing spectral index is shown on the y-axis in combination with a varying maximum energy of the source on the x-axis. Color coded are the expected number of secondary photons above a threshold energy of 10 EeV for an exposure of 3130 km^2 sr yr, corresponding to [81]. The white dashed line indicates one photon above 10 EeV. The red numbers indicate an increased exposure by a factor of 30, corresponding to the expected sensitivity of Auger South.

are compatible with the current upper limit. That is, the present upper limit on the photon flux does not yet constrain Centaurus A as a strong source of UHE protons. It was checked that a constant transverse magnetic field of 100 pG has just a marginal effect on the UHE photon flux in this scenario. The white dashed line in Fig. 6.8 indicates an improved photon flux level of $F_{10\mathrm{EeV}}^{\mathrm{limit}}/30$, corresponding to a rough estimate of the expected sensitivity of Auger South. Here, astrophysically relevant parameter combinations of α and $E_{\max}$ produce a larger photon flux and can thus be tested.

The number of arriving photons above 10 EeV for varying source parameters that are expected to be observed with an exposure of 3130 km^2 sr yr (corresponding to the exposure used for $F_{10\mathrm{EeV}}^{\mathrm{limit}}$ [81])

$$N_{\gamma,3130}^{10\mathrm{EeV}} = F_{\gamma}^{10\mathrm{EeV}} \cdot 3130\ \mathrm{km}^2\ \mathrm{sr\ yr} \tag{6.4}$$

is shown in Fig. 6.9. For large $E_{\max}$ and small α the expected number increases up to a few photons. For an increased exposure of a factor of 30 the number of arriving photons also scales by a factor of 30 as indicated by the red numbers in Fig. 6.9. As can be seen, depending on the source parameters, up to several 10 events could be expected.

So far, the simplified assumption was made that the total flux above 57 EeV is produced by the source – irregardless of the question whether the observed shape of the spectrum or the distribution of arrival directions were reproduced. Relaxing now this assumption one can ask for the fraction

$$f = \frac{F_{10\mathrm{EeV}}^{\mathrm{limit}}/X}{F_{\gamma}^{10\mathrm{EeV}}(\alpha, E_{\max})}\ , \tag{6.5}$$

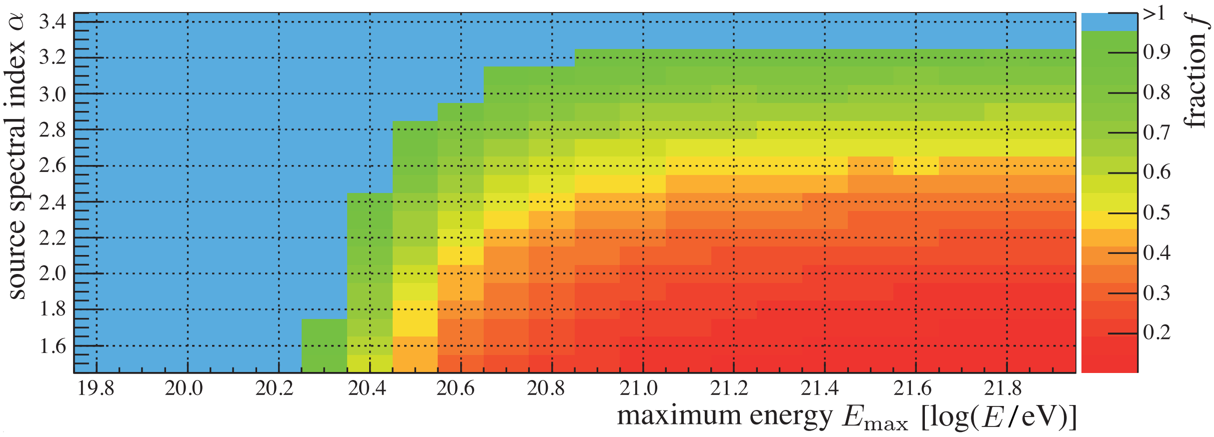


Figure 6.10: Simulation matrix of varying source parameters. Color coded is the fraction $f = (F^{\text{limit}}_{10\text{EeV}}/30)/F^{10\text{EeV}}_{\gamma}(\alpha, E_{\max})$ (cf. caption of Fig. 6.8).

where $X = 1$ refers to the exposure used for $F^{\text{limit}}_{10\text{EeV}}$ and $X \simeq 30$ to the Auger South sensitivity. Assuming now an observed photon upper limit $F^{\text{limit}}_{10\text{EeV}}/X$, for $f > 1$, no constrain on source parameters is possible, while values of $f < 1$ indicate which fraction of the total flux would still be allowed from the source for a given combination of α and $E_{\max}$. As is clear from Fig. 6.8, $f > 1$ for $X = 1$, for all simulated combinations of α and $E_{\max}$.

In Fig. 6.10, the case of $X = 30$ is shown, and the corresponding fractions of the total flux still allowed by the source can be extracted. For instance, in case of $\alpha \simeq 2$ and $E_{\max} \simeq 10^{21}$ eV, not more than $\sim 30\%$ of the total cosmic ray flux could be due to protons from Centaurus A.

6.6 Summary

The interactions en route to Earth of a primary proton/photon have been simulated using the Monte Carlo based propagation code CRPropa. There is a strong dependence on the evolution of the extragalactic background light in particular the CMB for distant sources ($>$ 100 Mpc). Relative fluctuations of energies to the mean energy are dominant in the range between $5 - 40$ Mpc and are important for a spectrum derived from a low number of events.

Due to the competition between GZK photon production and attenuation (both of which are increasing with travel distance), the expected photon fluxes show a non-trivial dependence on the source distance (Fig. 6.4, see also e.g. [231]). GZK photon fluxes from a proton source are mainly observed within a propagation distance of up to $25 - 50$ Mpc with a peak at about 10 Mpc.

The effect of different input parameters on the GZK-horizon is investigated. For the maximum energy produced in the source, differences are of the order of $\sim 5\%$ above an energy threshold of $E^{\mathrm{thres}} \simeq 120$ EeV (cf. Fig. 6.7). There is a more constant (energy independent) offset for varying spectral indices. A variation of ±10% induces an offset of $\sim 2\%$ in the horizon. Furthermore, a variation of ±10% of the Hubble parameter modifies the horizon of the order of $\sim 2\%$ for lower energy thresholds (< 80 EeV).

Regarding the specific case of Centaurus A (see also [232]), the current photon flux limit [81] does not yet constrain Centaurus A as a strong source of UHE protons for the investigated range of spectral indices α and maximum energies $E_{\max}$ (Fig. 6.8). However, the sensitivity that will be accumulated by Auger South will allow interesting constrains for a broad range of α and $E_{\max}$ (Fig. 6.8 and Fig. 6.10).

Depending on source parameters, the number of GZK photons above 10 EeV may reach several 10 over the lifetime of Auger South (Fig. 6.9). One can conclude that the search for UHE photons helps to provide significant clues about the characteristics of potential astrophysical sources.

Note that only GZK photons are regarded, i.e. photons produced during the propagation. The photon fluxes may be enhanced in case of interactions *at* the source.

Chapter 7

Simulation Study for Photon/Hadron Discrimination

The essential feature in photon searches is the knowledge of the trace of photons in experimental data compared to background traces of non-photons. Since there are no photon probes at highest energies, simulation studies of photon and background particles are necessary to discriminate between them. This chapter introduces the creation of a simulation sample used for photon/hadron discrimination. Extensive air showers are simulated with CORSIKA *described in Sec. 7.1. The corresponding detector response of the Pierre Auger Observatory is introduced in Sec. 7.2 using the Offline framework. Adopted observables for photon/hadron discrimination are explained in Sec. 7.3 utilizing information from the surface as well as the fluorescence detector.*

When searching for ultra-high energy photons, a detailed knowledge on extensive air shower development and detection techniques is essential. Differences in shower development and hence in detector response should be utilized to discriminate between photon and hadron air showers.

To study the evolution and properties of extensive air showers in the atmosphere a detailed Monte Carlo (MC) program has been developed to perform simulations for the KASCADE [233] experiment. The CORSIKA (*CO*smic *R*ay *SI*mulations for *KA*scade) [112, 113] code allows to simulate interactions up the energies of some 10^{20} eV and is therefore also applicable to experiments in the highest energy regime like the Pierre Auger Observatory. All secondary particles that are created in an air shower[1] are recorded including type, location, energy, direction and arrival times. CORSIKA consists of basically four parts:

---

[1] depending on thinning level, cf. App. B

1. **General program frame**: In- and output is handled as well as decay of unstable particles including ionization energy loss and deflection at the Earth's magnetic field.

2. **Hadronic interactions at higher energies**: Different models can be selected. In this simulation study QGSJET 01C [234] is used.

3. **Hadronic interactions at lower energies**: Different models can be selected and in this study the well approved GHEISHA [235] code was chosen.

4. **Electromagnetic processes**: The transport and interaction of electrons, positrons and photons is calculated. The electromagnetic component is simulated in this study with the analytic NKG formulae [236, 237] to obtain electron densities at selected locations and the total number of electrons.

7.1 CORSIKA production

An extensive CORSIKA simulation study is performed to examine differences in photon and hadron induced air showers. CORSIKA version 6.735 is used[2] to create a data sample of fixed energy and theta angles applying a thinning level of $\epsilon = 10^{-6}$. The energies ranging from $10^{17.75}$ eV to $10^{18.75}$ eV in discrete steps of 0.5 $\log(E/\mathrm{eV})$. Also discrete theta angles of 0°, 26° and 60° are used. Each energy-theta bin consists of 1000 proton and 1000 photon primaries. Table 7.1 summarizes the simulation sample. More details on simulation parameters can be found in App. B.

Energy [$\log(E/\mathrm{eV})$] \ Theta [°]	0	26	60
17.75	1000	1000	1000
18.25	1000	1000	1000
18.75	1000	1000	1000

Table 7.1: CORSIKA simulation sample for fixed energy and theta angles for proton and photon primaries. In total 9000 primary protons and 9000 primary photons were simulated.

In addition to fixed energy and theta MC simulations, a more realistic data sample is created using CORSIKA version 6.900. To be most conservative the discrimination power between photon and proton primaries is investigated. The MC data sample follows a spectral index of $\gamma = 2.7$ (cf. Eqn. (2.1)) between $10^{17.2}$ eV and $10^{18.5}$ eV and thinning level $\epsilon = 10^{-6}$. Theta angles are distributed between 0° and 60° in a way that respects equal particle fluxes from all solid angle elements of the sky assuming a horizontal flat detector arrangement. A sample CORSIKA input card is given in Fig. B.2. In total 60000 proton, 20000 photon and 10000 iron primaries are simulated. The iron sample

---

[2] http://www-ik.fzk.de/corsika/

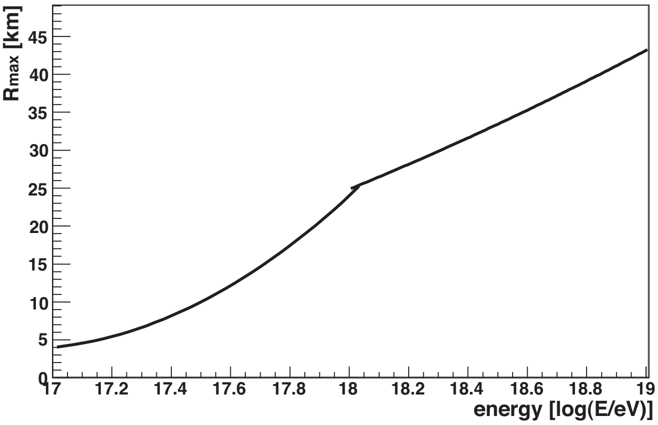


Figure 7.1: Maximal distance $R_{\max}$ between telescope and core as a function of primary energy as used in this analysis. Above 10^{18} eV a standard formula is used [238]. For energies $< 10^{18}$ eV a modified version is applied to account for small trigger propabilities.

is created to study the response of various multivariate techniques to different primary particles. More details can be found in Sec. 8.6.3.

7.2 $\overline{\text{Off}}\underline{\text{line}}$ simulation and reconstruction

The detector response of the Pierre Auger Observatory surface and fluorescence detector has been simulated using the $\overline{\text{Off}}\underline{\text{line}}$ framework version v2r7p0-svn_trunk which was already introduced in Sec. 4.3. The shower core is chosen randomly in an eye centric system in front of bay 4 of the Los Leones fluorescence telescope with an opening angle of 30°. The maximum distance $R_{\max}$ between telescope and core is 16 km in case of the realistic data sample (with spectral index). For fixed energies the miximum distance $R_{\max}$ is dependent of energy in a way that the fraction of triggered events above $R_{\max}$ is less than 1% [238], i.e.

$$\frac{\int_{R_{\max}}^{\infty} \epsilon(R,E) R^2 \mathrm{d}R}{\int_0^{\infty} \epsilon(R,E) R^2 \mathrm{d}R} < 0.01 \tag{7.1}$$

with the trigger efficiency $\epsilon(R,E)$ adopted from [239]. $R_{\max}$[m] is parameterized as

$$R_{\max} = p_2 \cdot E^2 + p_1 \cdot E + p_0 \tag{7.2}$$

with

$$p_0 = 4.86 \cdot 10^5 \text{ m} \quad p_1 = \frac{-6.72 \cdot 10^4 \text{ m}}{\log(\text{E/eV})} \quad p_2 = \frac{2.31 \cdot 10^3 \text{ m}}{\log(\text{E/eV})^2} \quad \text{for} \quad E \geqslant 18 \log(\text{E/eV})$$

$$p_0 = 45.6 \cdot 10^5 \text{ m} \quad p_1 = \frac{-54.0 \cdot 10^4 \text{ m}}{\log(\text{E/eV})} \quad p_2 = \frac{16.0 \cdot 10^3 \text{ m}}{\log(\text{E/eV})^2} \quad \text{for} \quad E < 18 \log(\text{E/eV}) .$$

$R_{\max}$ as function of energy is also shown in Fig. 7.1. To enhance the statistics each CORSIKA shower is used five times with random core locations.

After a successful detector simulation the events are reconstructed using an ideal detector configuration as well as standard atmospheric parameters. All important simulated and reconstructed parameters are stored into a data summary structure called "Advanced Data Summary Tree" (`ADST`) [240]. This `ROOT` [241] based file format has been developed within the Pierre Auger Collaboration to contain all high level physics variables needed for physics analysis including (if desired) also a fair amount of low level data to facilitate the development of new data selection cuts.

7.3 Observables for photon/hadron discrimination

To discriminate between photon and other background primaries at ultra-high energies is a challenge. Each experiment and detector system has its own characteristics. Is this section observables are introduced[3] that are optimized for photon/hadron discrimination at energies around 10^{18} eV and used in this analysis. After a physical motivation a more technical description on calculation and implementation is given. Finally, the performance of the observable for discrete energies and inclination angles is briefly discussed. The discrimination power of the observable p is estimated using the so-called merit factor defined as

$$\eta = \left| \frac{\mathrm{E}[p_\gamma] - \mathrm{E}[p_p]}{\sqrt{\mathrm{Var}[p_\gamma] + \mathrm{Var}[p_p]}} \right| . \tag{7.3}$$

$\mathrm{E}[p_\gamma]$ and $\mathrm{E}[p_p]$ are the expectation values (mean values) of the photon and proton distribution, respectively. $\mathrm{Var}[p_\gamma]$ and $\mathrm{Var}[p_p]$ are the corresponding variances (σ^2) of the distribution. In this thesis RMS is defined as the standard deviation σ of the distribution. It is calculated as $\sqrt{1/N \cdot \sum_i (x_i - \overline{x})^2}$. This convention was introduced many years ago in `ROOT` and is kept for continuity. Note that large η indicate good separation power while an η of 0 denotes no separation.

Fluorescence detector as well as surface detector information is utilized. Possible correlations among the observables are examined in Sec. 8.5.

7.3.1 FD observable: Depth of shower maximum

Motivation

The depth of shower maximum X_{max} is defined as the atmospheric depth at which the longitudinal development of a shower reaches its maximum in terms of secondary particles. It is correlated with the mass of the incident cosmic ray particle as already shown in Eqn. (3.6). On average, photon primaries develop later in the atmosphere compared to hadron primaries resulting in larger X_{max} values for photons.

---

[3] based on CORSIKA and $\overline{\mathrm{Off}}\underline{\mathrm{line}}$ simulations introduced in previous sections.

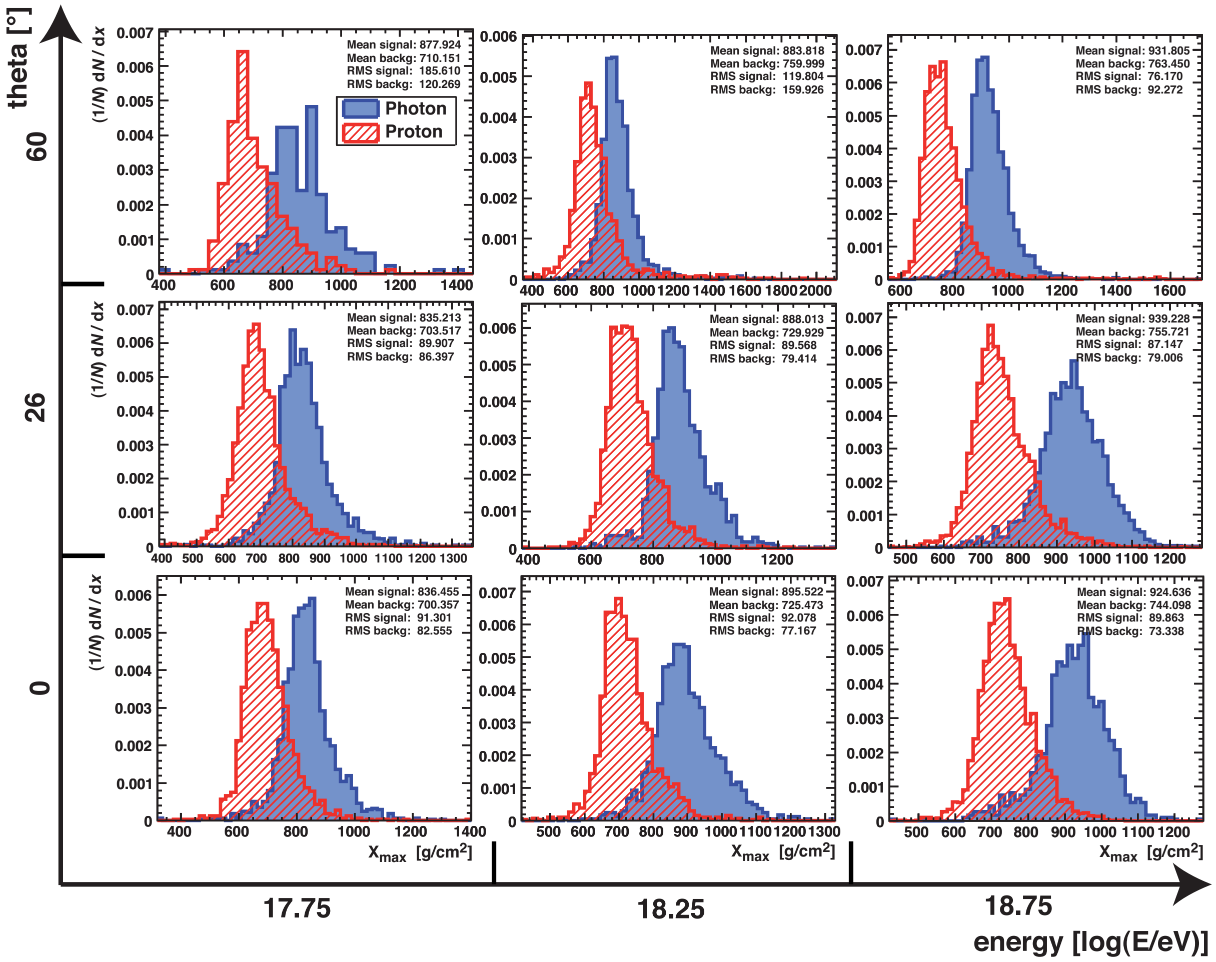


Figure 7.2: Distribution of $X_{\max}$ for various energy / θ bins. Primary photons and protons are shown in blue and red shades, respectively. Mean and RMS values for individual energy / θ combinations are indicated, cf. Fig. 7.3. Note that "signal" always refers to primary photons and "background" to primary protons.

Description

Here only events detected in hybrid mode are considered. The longitudinal profile of the energy deposit is reconstructed from the light recorded at the FD using Čerenkov and fluorescence yields. $X_{\max}$ is determined by fitting the longitudinal profile with a Gaisser-Hillas function, cf. Eqn. (3.7). The maximum of the function represents the depth of shower maximum, cf. Fig. 3.5.

Performance

As can be seen in Fig. 7.2 the $X_{\max}$ distributions show a clear separation for photon and proton primaries with typical differences of ~ 150 g/cm^2, dependent of energy and inclination angle for all θ / energy combinations. The merit factor as well as the depen-

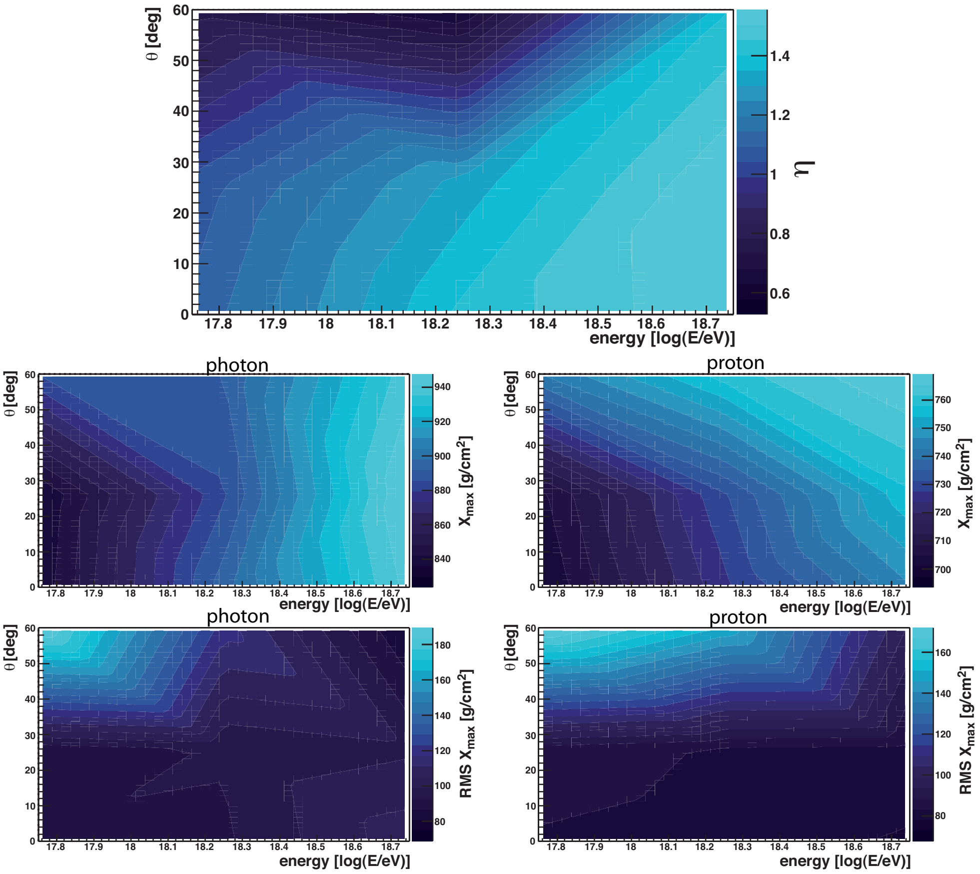


Figure 7.3: *Top panel*: Merit factor η as a function of energy and inclination angle for the $X_{\max}$ distribution. *Middle panel*: Energy and θ dependencies of $X_{\max}$ for photon (left) and proton (right) primaries. *Bottom panel*: Energy and θ dependencies of RMS fluctuations in $X_{\max}$ for photon (left) and proton (right) primaries.

dency of the mean $X_{\max}$ and its fluctuations on energy and inclination angle is shown in Fig. 7.3. Best separation results are achieved at high energy and small inclination angle. With increasing energy a clear increase in $\langle X_{\max} \rangle$ is observed. While there is no strong correlation with the inclination angle θ for photon primaries, protons tend to have a positive correlation on θ. Also shown in Fig. 7.3 is the evolution of RMS fluctuations with energy and inclination angle. As can be seen, fluctuations are stronger for low energies and large inclination angles.

7.3.2 FD observable: Greisen χ^2

Motivation

A common feature for all photon induced air showers is the almost purely electromagnetic cascade via pair production and bremsstrahlung. The production of muon pairs is suppressed by $(m_e/m_\mu)^2$ as discussed in Sec. 5.4. The parameterization of the longitudinal

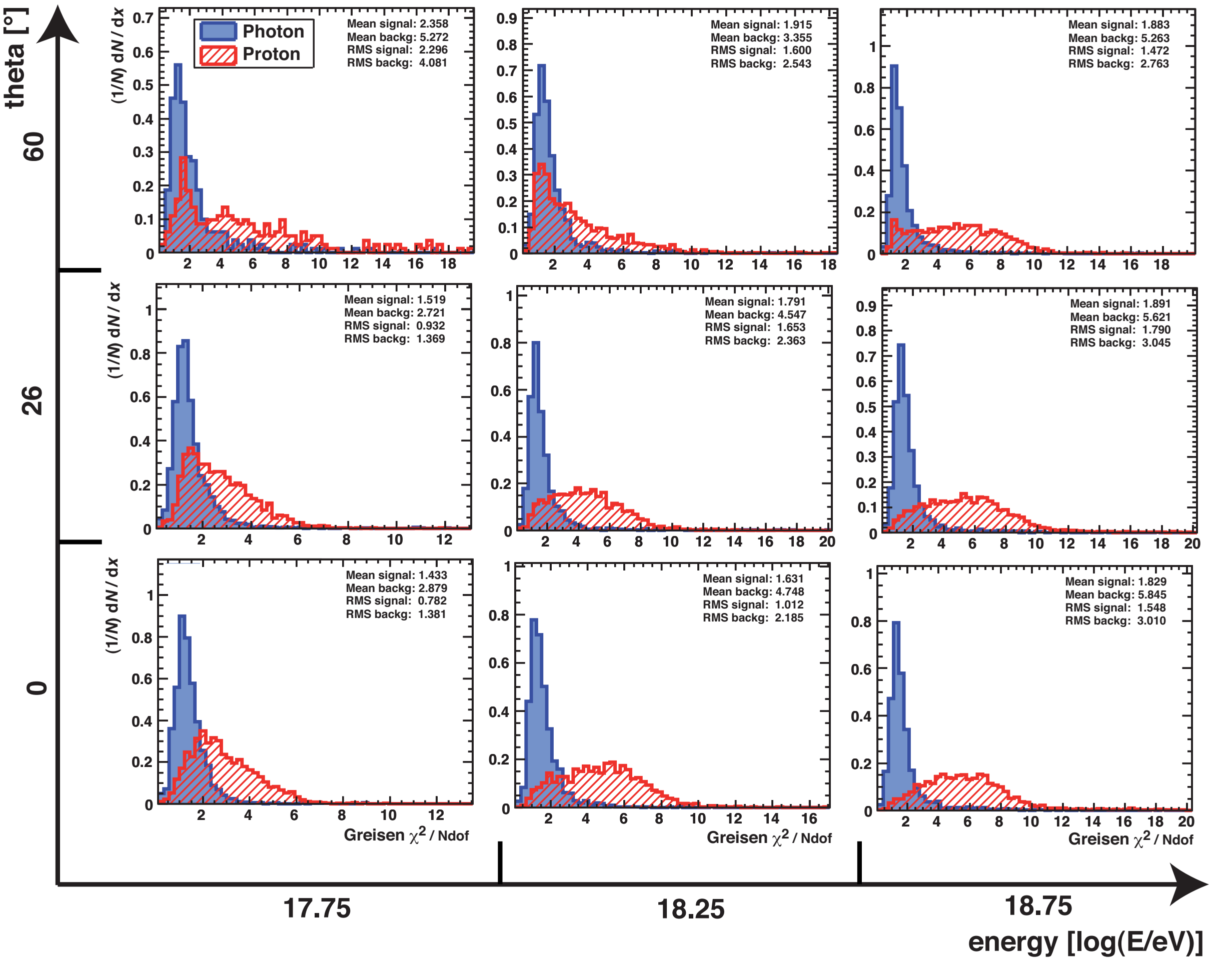


Figure 7.4: Distribution of the reduced χ^2_{Greisen} for various energy / θ bins. Primary photons and protons are shown in blue and red shades, respectively. Mean and RMS values for individual energy / θ combinations are indicated, cf. Fig. 7.5. Note that "signal" always refers to primary photons and "background" to primary protons.

shower development with the Greisen function (cf. Sec. 3.1.2) just covers the electromagnetic component of the shower and neglects muonic or hadronic elements. A fit to the longitudinal profile with the Greisen function for photon primaries should therefore give more accurate results than for proton primaries.

Description

The Greisen function was already introduced in Sec. 3.1.2 Eqn. (3.11) giving the mean number of charged particles $N_{\text{ch}}(X, E)$ as a function of atmospheric depth X and energy of the primary particle E. However, the energy deposit in the shower as a function of

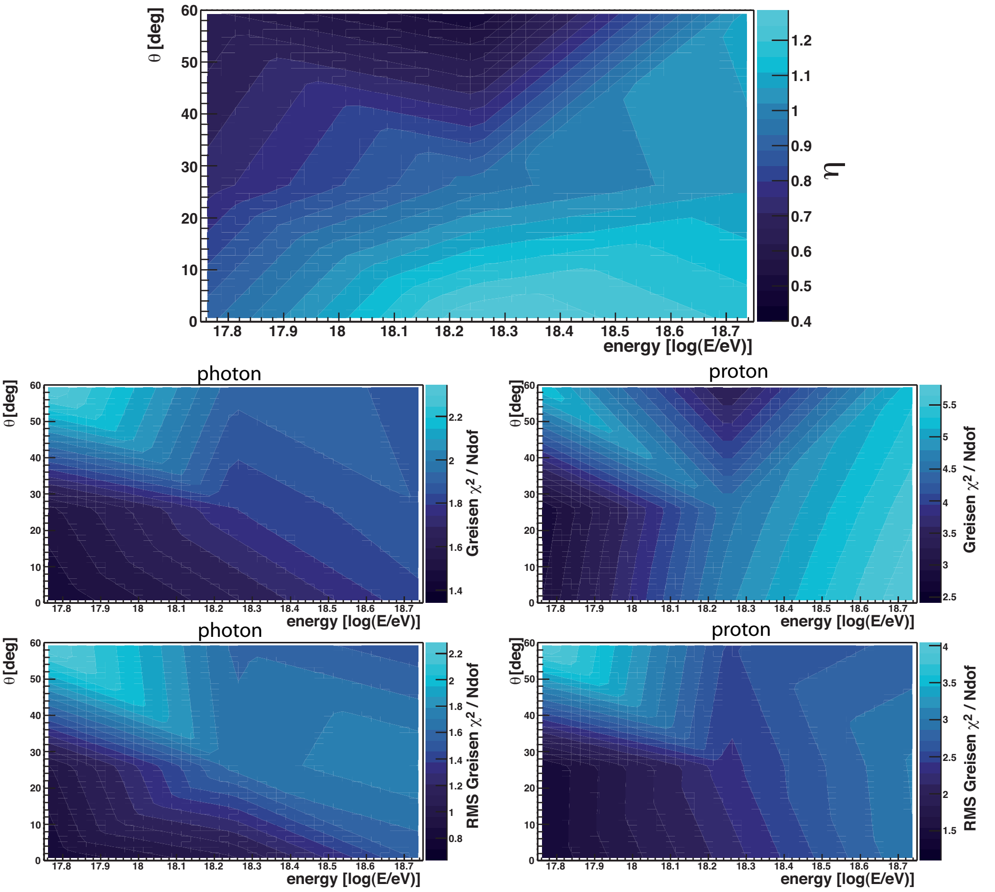


Figure 7.5: *Top panel*: Merit factor η as a function of energy and inclination angle for the reduced $\chi^2_{\mathrm{Greisen}}$ distribution. *Middle panel*: Energy and θ dependencies of the reduced $\chi^2_{\mathrm{Greisen}}$ for photon (left) and proton (right) primaries. *Bottom panel*: Corresponding energy and θ dependencies of RMS fluctuations for photon (left) and proton (right) primaries.

slant depth is measured in $\mathrm{d}E/\mathrm{d}X$ and is converted according to (cf. Eqn. (3.14))

$$E_{\mathrm{em}} = \int \alpha_{\mathrm{loss}}(X, E > E_{\mathrm{cut}}) \cdot N(X)\, \mathrm{d}X = \int \frac{\mathrm{d}E}{\mathrm{d}X}\, \mathrm{d}X \ , \tag{7.4}$$

where $\alpha_{\mathrm{loss}}(X, E > E_{\mathrm{cut}})$ is a parametrization of the mean energy loss rate and E_{cut} a low-energy threshold, which has to be applied in case of shower simulations [242]. The currently used Greisen fit function is a slightly modified version of Eqn. (3.11) as it is multiplied by a factor 0.91 to achieve better energy determination within the Greisen fit:

$$N_{\mathrm{ch}}(X, E) = 0.91 \cdot \frac{0.31}{\sqrt{\ln(E/E_c)}} e^{\frac{X}{X_r}} \left(\frac{3X}{X + 2X_r \ln(E/E_c)} \right)^{-\frac{3X}{2X_r}} \ . \tag{7.5}$$

Note, however, that the factor 0.91 does not affect the performance of photon / hadron discrimination. Finally, a χ^2 fit on the longitudinal profile of the Greisen function introduced in Eqn. (7.5) is performed and the resulting $\chi^2_{\mathrm{Greisen}}$ and the number of degrees of freedom is written out.

Performance

The impact of the reduced Greisen χ^2 for different energies and zenith angles is shown in Fig. 7.4. As expected, longitudinal profiles of photon primaries are much better described by the Greisen function than proton primaries independent of energy and zenith angle. RMS fluctuations are in general a factor of ~ 2 larger for proton compared to photon primaries which is also visible in the much broader χ^2 distribution. The merit factor, the dependency of the reduced Greisen χ^2 and its fluctuations on energy and inclination angle is shown in Fig. 7.5. Small θ-angles and larger energies indicate better separation power.

7.3.3 FD observable: Greisen energy / FD energy

Motivation

As already discussed in the previous section the Greisen function has been derived from purely electromagnetic cascade theory appropriate for photon but not for hadron induced longitudinal profiles. As the only free fit parameter the reconstructed Greisen energy E_{Greisen} should therefore differ for photon and hadron induced air showers.

Description

The applied fit function was introduced in Eqn. (7.5). As can be seen, the only free parameter is the energy of the primary particle that is determined when fitting the function to the longitudinal shower profile. The final observable is the ratio of the Greisen energy E_{Greisen} and the standard reconstructed FD energy E_{FD} using the Gaisser-Hillas fit function (Eqn. (3.7)) to determine the energy.

Performance

The ratio of the Greisen energy and standard FD energy for various primary energies and inclination angles is shown in Fig. 7.6. As can be seen, for photon showers one obtains $E_{\mathrm{Greisen}} < E_{\mathrm{FD}}$ independent of energy or inclination angle. On the contrary, proton induced showers show for large inclination angles and small energies $E_{\mathrm{Greisen}} < E_{\mathrm{FD}}$. With increasing energy and decreasing inclination angle the relation reverse to $E_{\mathrm{Greisen}} > E_{\mathrm{FD}}$. Fig. 7.7 stresses the aforementioned behavior and illustrates the RMS fluctuations as a function of energy and inclination angle as well. As for the previous observables the merit factor η is maximized for large energy and near vertical air showers.

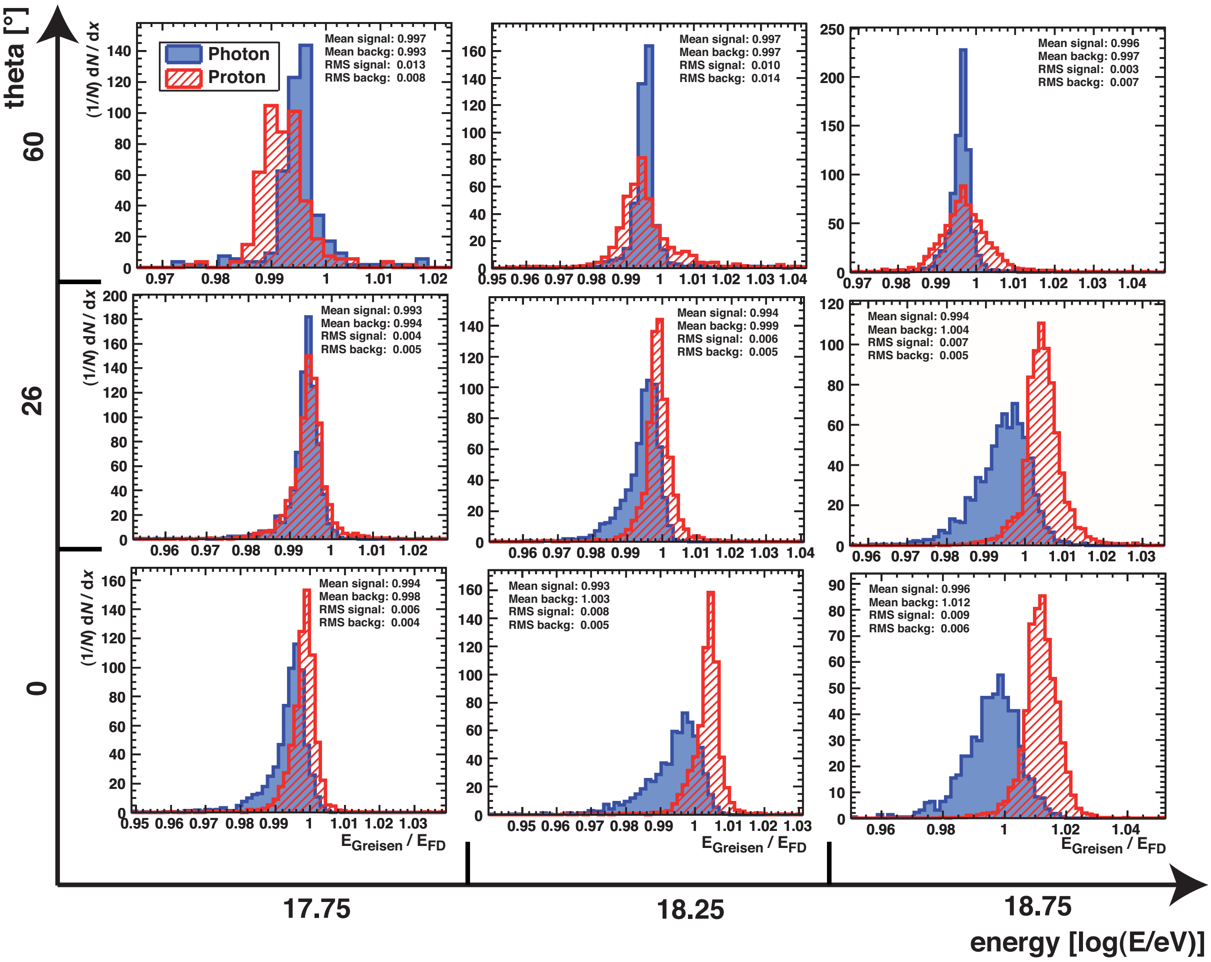


Figure 7.6: Distribution of the ratio of Greisen energy and FD energy for various energy / θ bins. Primary photons and protons are shown in blue and red shades, respectively. Mean and RMS values for individual energy / θ combinations are indicated, cf. Fig. 7.7. Note that "signal" always refers to primary photons and "background" to primary protons.

7.3.4 SD observable: S_b

Motivation

Compared to fluorescence detectors, surface detectors sample the lateral distribution of an air shower at discrete points facilitating to utilize different features of photon induced showers. The particle composition as a function of distance to the shower axis consists of three parts. The main *hadronic component* $(p, n, \pi^{\pm}, K^{\pm})$ is typically just within a few tens of meters from the shower core. More distant parts of the shower are dominated by the remaining *electromagnetic* $(e^{\pm}, \gamma)$ and *muonic component* $(\mu^{\pm})$. The development of these two components proceed in a different way. While the electromagnetic component propagates diffusively, muons propagate radially from their parent mesons. The final shower front can therefore thought to be a superposition of a more extended later arriving

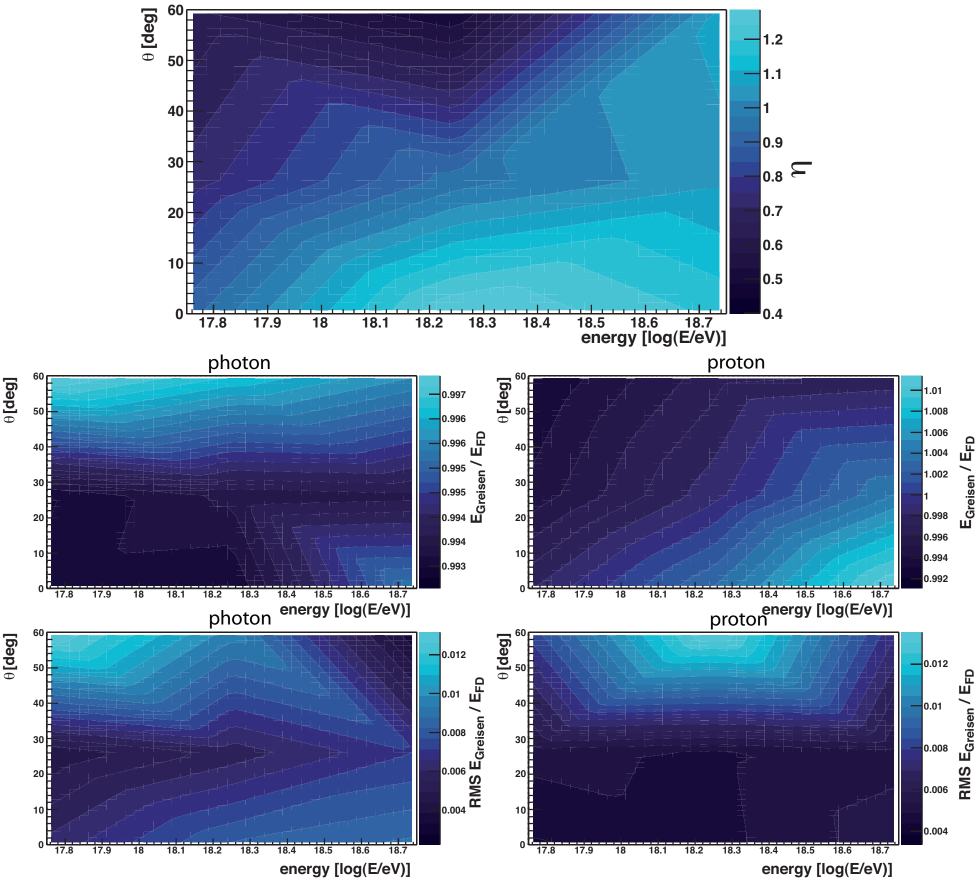


Figure 7.7: *Top panel*: Merit factor η as a function of energy and inclination angle for the ratio of Greisen energy and FD energy. *Middle panel*: Energy and θ dependencies of the ratio of Greisen energy and FD energy for photon (left) and proton (right) primaries. *Bottom panel*: Corresponding energy and θ dependencies of RMS fluctuations for photon (left) and proton (right) primaries.

electromagnetic component and a temporally thin first arriving muonic component. Since the muonic component is suppressed in case of photon induced air showers, information of the primary particle should also be distributed in a non-trivial behavior in the slope of the the lateral distribution, shower curvature and structure of the shower front. The proposed surface detector observable S_b tries to cover the aforementioned characteristics in a single parameter which is described in more detail below.

Description

The parameter was first introduced by G. Ros in [243] and is sensitive to the difference in the muon component and in the slope of the lateral distribution. It is defined as

$$S_b = \sum_{i=1}^{N} \left[S_i \cdot \left(\frac{r_i}{r_0} \right)^b \right] \quad [\text{VEM}], \tag{7.6}$$

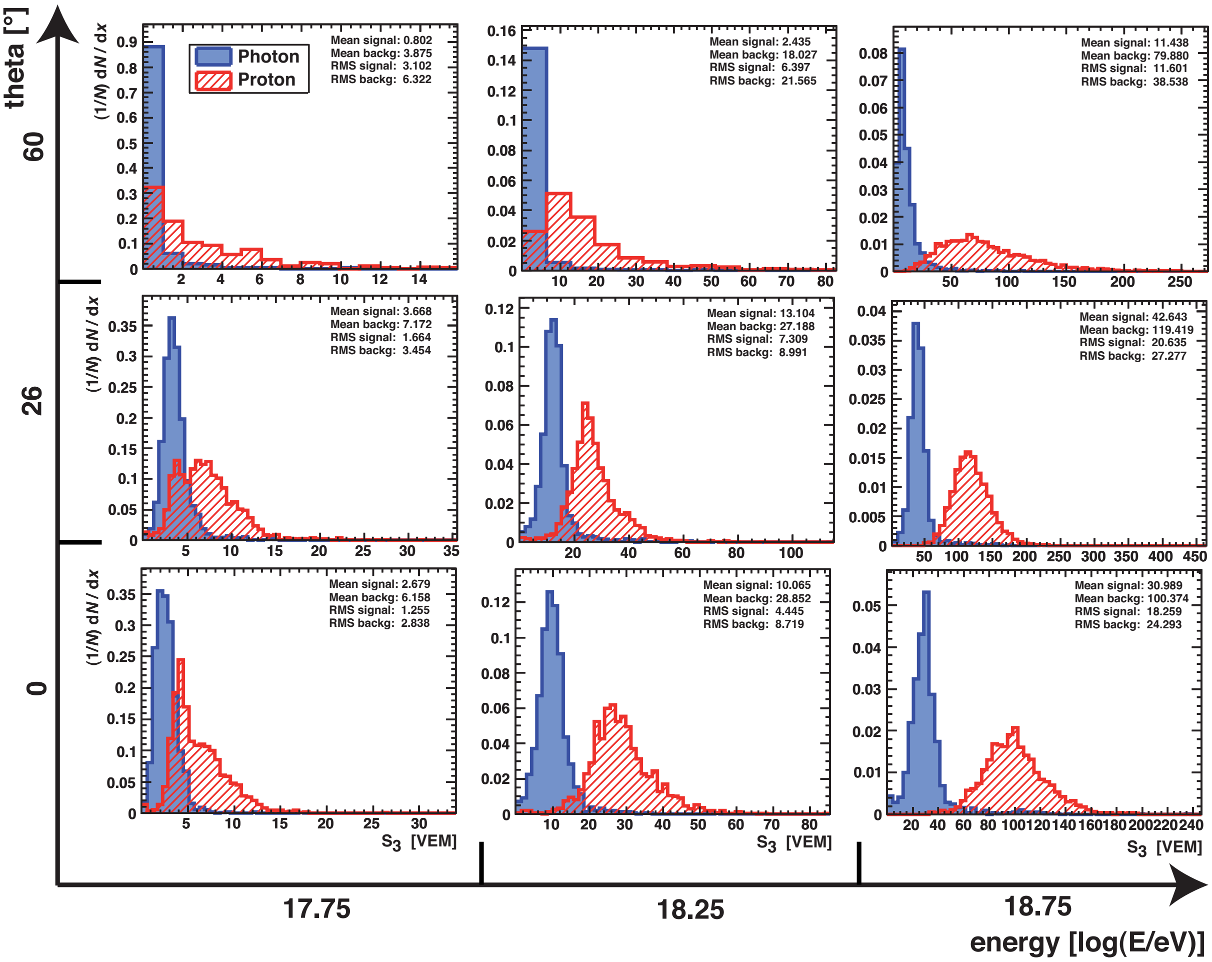


Figure 7.8: Distribution of S_3 for various energy / θ bins. Primary photons and protons are shown in blue and red shades, respectively. Mean and RMS values for individual energy / θ combinations are indicated, cf. Fig. 7.9. Note that "signal" always refers to primary photons and "background" to primary protons.

where the sum extends over all triggered stations N, S_i expresses the signal strength of the i–th surface detector station in VEM, and r_i the distance of this station to the shower axis in meters normalized to a reference distance $r_0 = 1000$ m. The exponent b is a free parameter and it is argued that the primary identity discrimination power goes through a maximum around $b = 3$ [243]. For photon/hadron discrimination, however, $b = 4$ is suggested [244] but own studies at EeV energies did not show a significant improvement. In this thesis $b = 3$ is applied and implemented in the analysis software referred to as S_3. Accidentally triggered stations are excluded as well as stations > 3 km away from the shower axis.

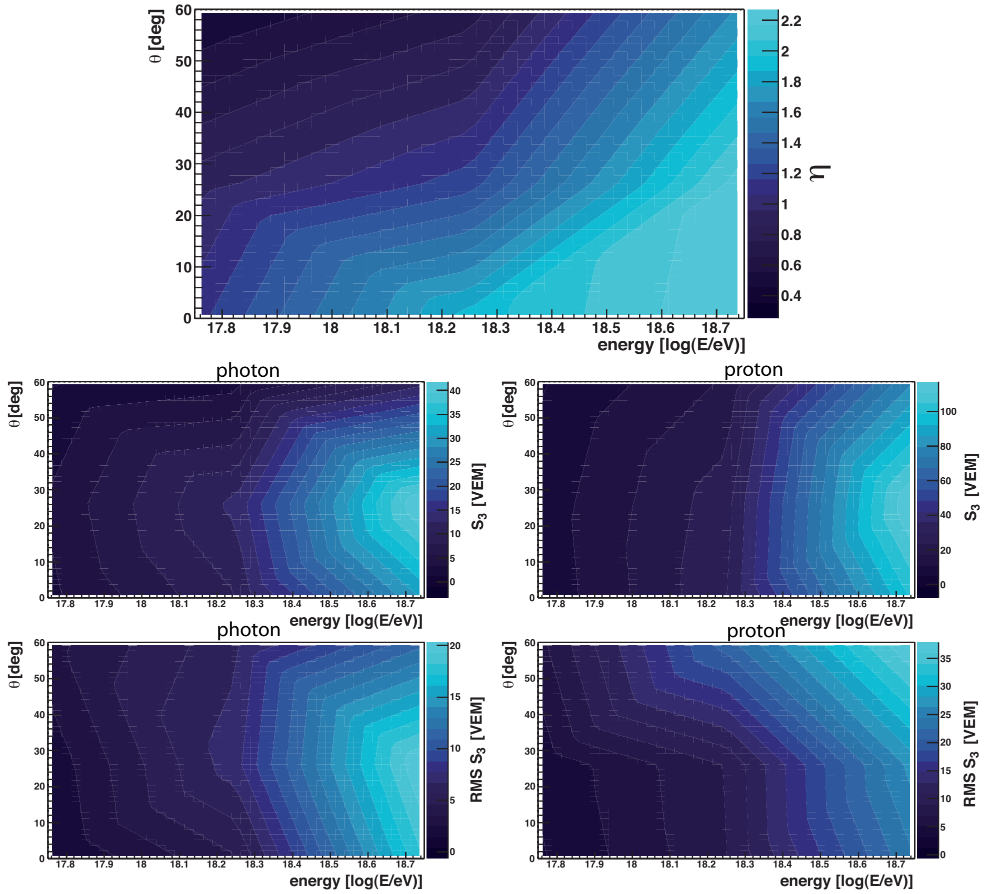


Figure 7.9: *Top panel*: Merit factor η as a function of energy and inclination angle for S_3. *Middle panel*: Energy and θ dependencies of S_3 for photon (left) and proton (right) primaries. *Bottom panel*: Corresponding energy and θ dependencies of RMS fluctuations for photon (left) and proton (right) primaries.

Performance

The S_3 parameter is shown in Fig. 7.8 for different energies and zenith angles. In general, proton distributions are broader with larger mean values, compared to photon induced air showers. Especially at higher energy and low inclination angles a clear separation is visible. In Fig. 7.9 the merit factor for S_3 is shown. As already visible by eye, discrimination power starts to increase for higher energy and more vertical showers. Also shown is the dependence of the mean and RMS of S_3 as a function of energy and inclination angle. With increasing energy, larger S_3 mean values are observed as expected. The related RMS values show a similar behavior.

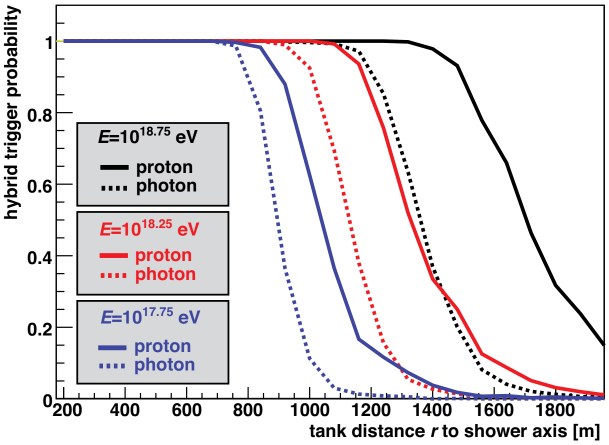


Figure 7.10: Hybrid trigger efficiency as a function of tank distance r to the shower axis for different primary energies as indicated. As expected for a fixed energy, photon induced showers lose full trigger efficiency at smaller r compared to proton induced showers.

7.3.5 SD observable: Number of candidate stations

Motivation

The lack of muons and the almost pure electromagnetic cascading in photon induced air showers has the effect that photon showers are in general more slim compared to proton or iron primaries. In other words, the lateral distribution function of photons has a steeper slope resulting in different trigger probabilities at a given distance r from the shower axis to a surface detector station. The trigger probability for photon and proton induced showers at fixed energies as a function of r is illustrated in Fig. 7.10. As expected, full trigger efficiency is retained for larger r in case of proton compared to photon induced showers of the same energy. In summary, photon showers should have in general less triggered SD stations than proton showers for a given energy and zenith angle.

Description

The parameter used to cover the effect of different trigger probabilities is the number of candidate stations of an air shower. Candidate stations are SD stations that are not rejected during the reconstruction process by e.g. no trigger, no GPS data, bad calibration or lonely stations. Furthermore it is checked if the current station is flagged as an accidental triggered station, i.e. stations that are originally not triggered by the current air shower. To tighten the condition even more, only tanks with a tank distance $r < 3$ km to the shower axis are considered. Note that this observable is an integer number.

Performance

The number of candidate stations for various energy and theta combinations is shown in Fig. 7.11. As can be seen, photon showers have in general less candidate stations com-

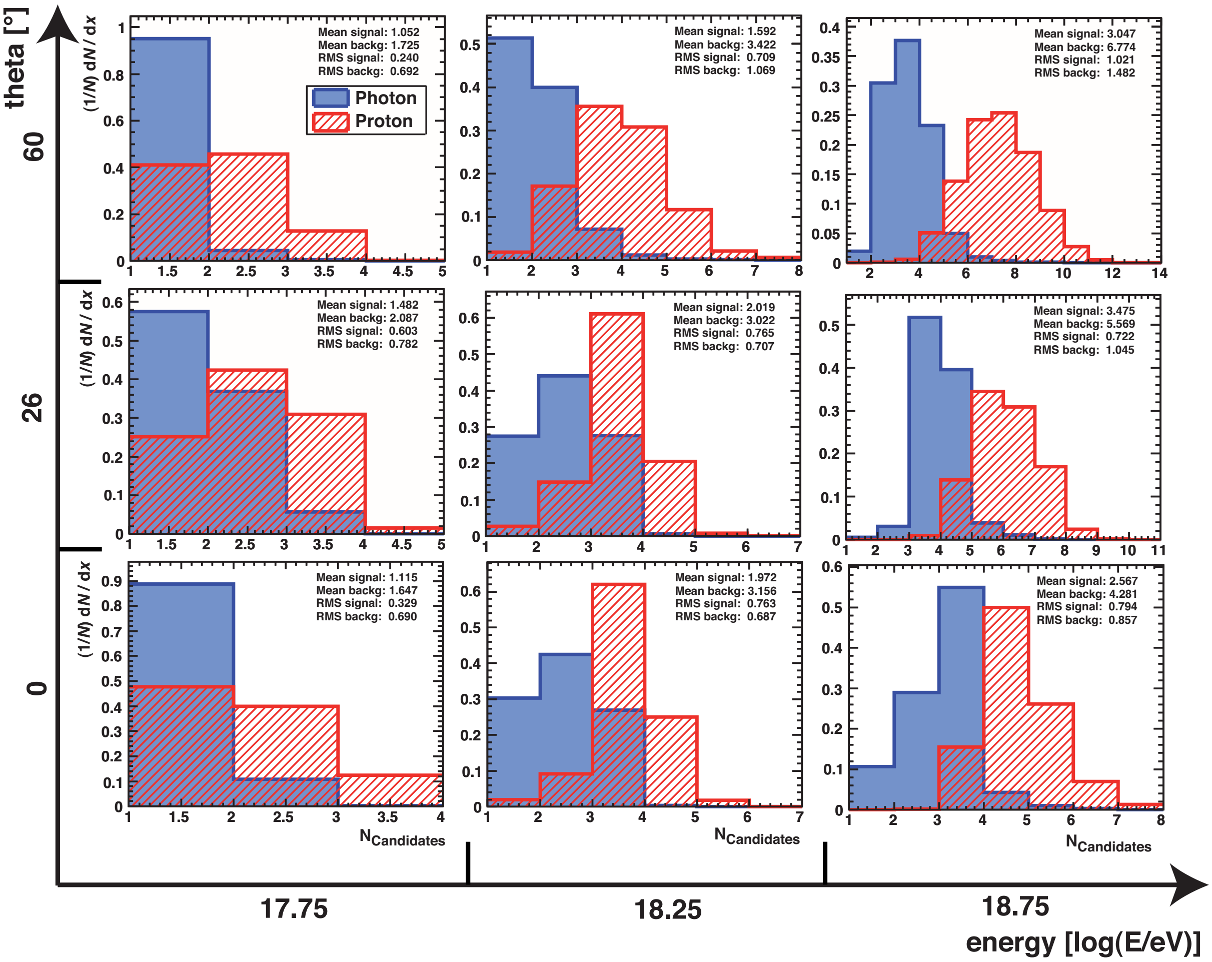


Figure 7.11: Distribution of the number of candidate stations for various energy / θ bins. Primary photons and protons are shown in blue and red shades, respectively. Mean and RMS values for individual energy / θ combinations are indicated, cf. Fig. 7.12. Note that "signal" always refers to primary photons and "background" to primary protons.

pared to protons at the same energy and inclination angle as justified in the motivation. The discrimination power is quantified in Fig. 7.12 using the merit factor. At high inclination and large energies best results are achieved. This can be explained by geometrical considerations. The closest distance from a SD station to the shower axis is, in general, smaller for inclined showers compared to a more vertical development. This advantage has the drawback that the shower development is at a later stage when triggering the tanks ($X = X_v \cdot \cos^{-1}\theta$, where X_v refers to the vertical atmospheric depth) making observations especially at low energies difficult. This downside is compensated at higher energies. The development of the mean and RMS value as a function of energy and inclination angle is also shown in Fig. 7.12.

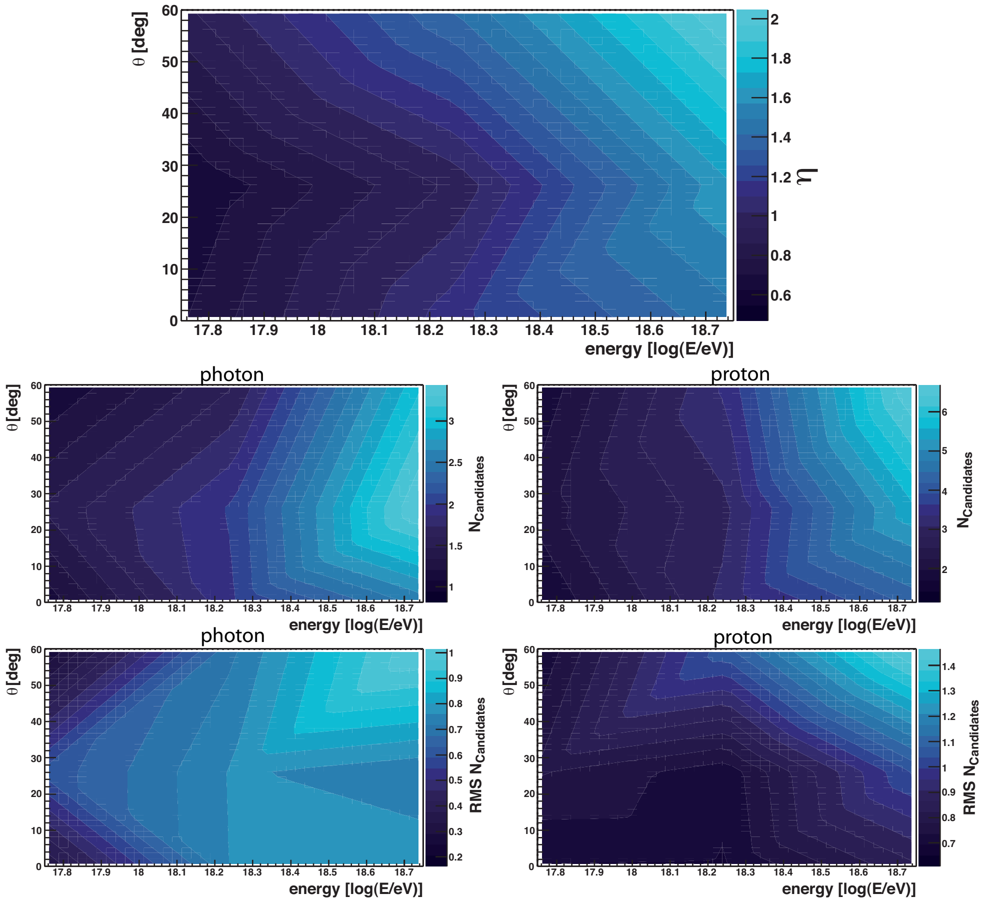


Figure 7.12: *Top panel*: Merit factor η as a function of energy and inclination angle for the number of candidate stations. *Middle panel*: Energy and θ dependencies of the number of candidates for photon (left) and proton (right) primaries. *Bottom panel*: Corresponding energy and θ dependencies of RMS fluctuations for photon (left) and proton (right) primaries.

7.3.6 SD observable: Tank energy / FD energy

Motivation

As already discussed in the motivation of the S_b observable, the slope of the lateral distribution function differs for photon and proton induced air showers at a given energy. This is used for composition studies and is utilized when estimating the shower energy using just the reconstructed hybrid geometry and information from individual tanks as described below.

Description

Originally developed to increase statistics, J. A. Bellido developed a way to determine the shower energy from the hybrid geometry, individual tank information and parametrization

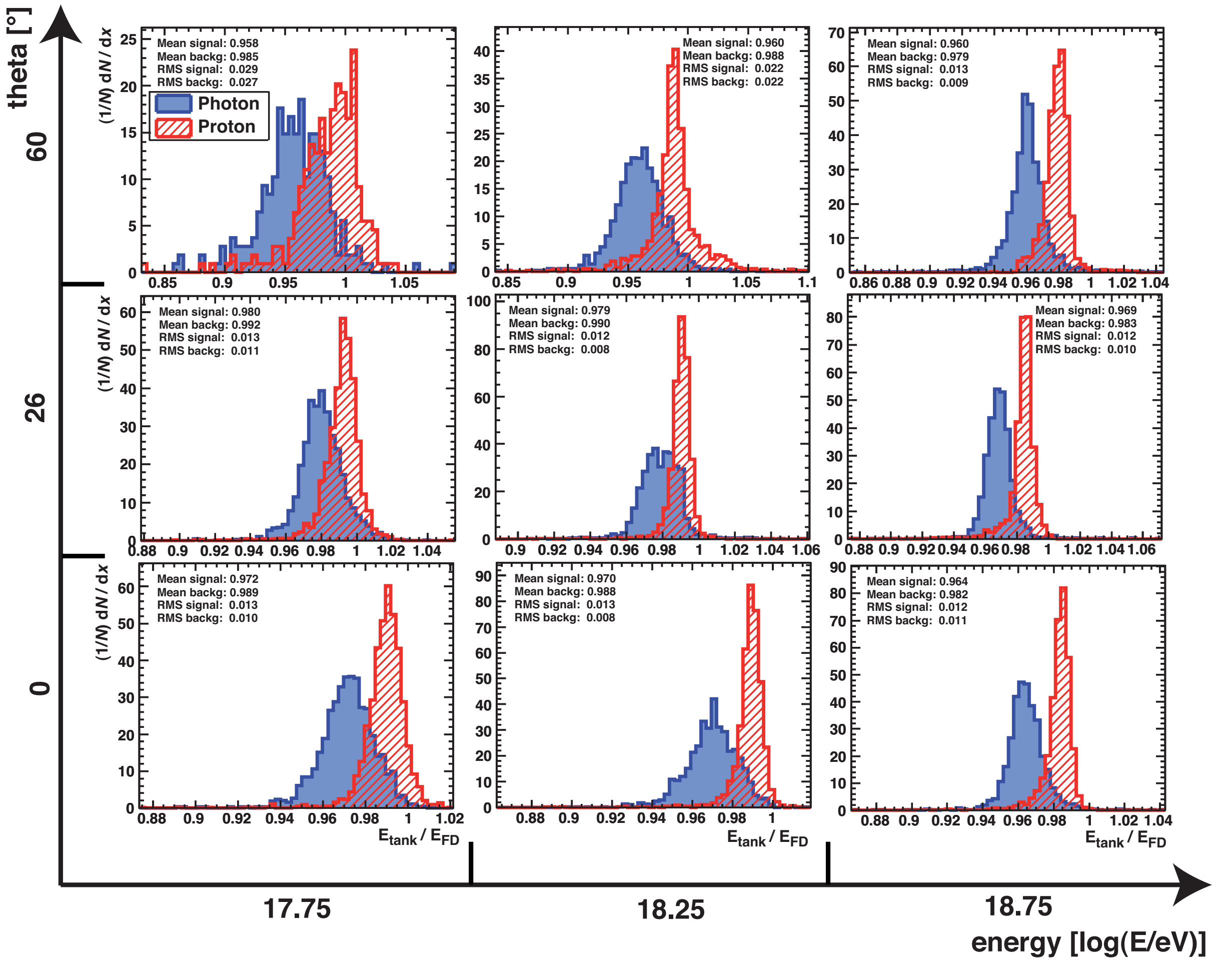


Figure 7.13: Distribution of the ratio of tank energy and FD energy for various energy / θ bins. Primary photons and protons are shown in blue and red shades, respectively. Mean and RMS values for individual energy / θ combinations are indicated, cf. Fig. 7.14. Note that “signal” always refers to primary photons and “background” to primary protons.

of the lateral distribution function (LDF) as a function of the shower energy and zenith angle [245]. The LDF is obtained using hybrid data itself.

The average signal at a given shower plane distance $S(r)$ changes as a function of zenith angle θ. In a first step $S(r)$ is normalized using the expected signal $S_{38}(S, r, \theta)$ at core distance[4] r and an inclination angle of 38° of the shower

$$S_{38}(S, r, \theta) = \frac{S(r)}{10^{-B(r)\cdot\cos^{-1}(38^\circ)+B(r)\cdot\cos^{-1}(\theta)}} \,. \tag{7.7}$$

The relation between $\log(S(r)/S_{38}(r))$ and $\cos^{-1}(\theta)$ is parameterized using a linear function[5]. $B(r)$ is a parametrization for the slope of the relation between $S(r)$ and $\cos^{-1}(\theta)$

---

[4] closest distance of the SD station to the shower axis.

[5] Note that $S_{38}(S, r, \theta)$ should not be confused with the common S_{38} which is the expected signal of a 38° shower at 1 km from the core.

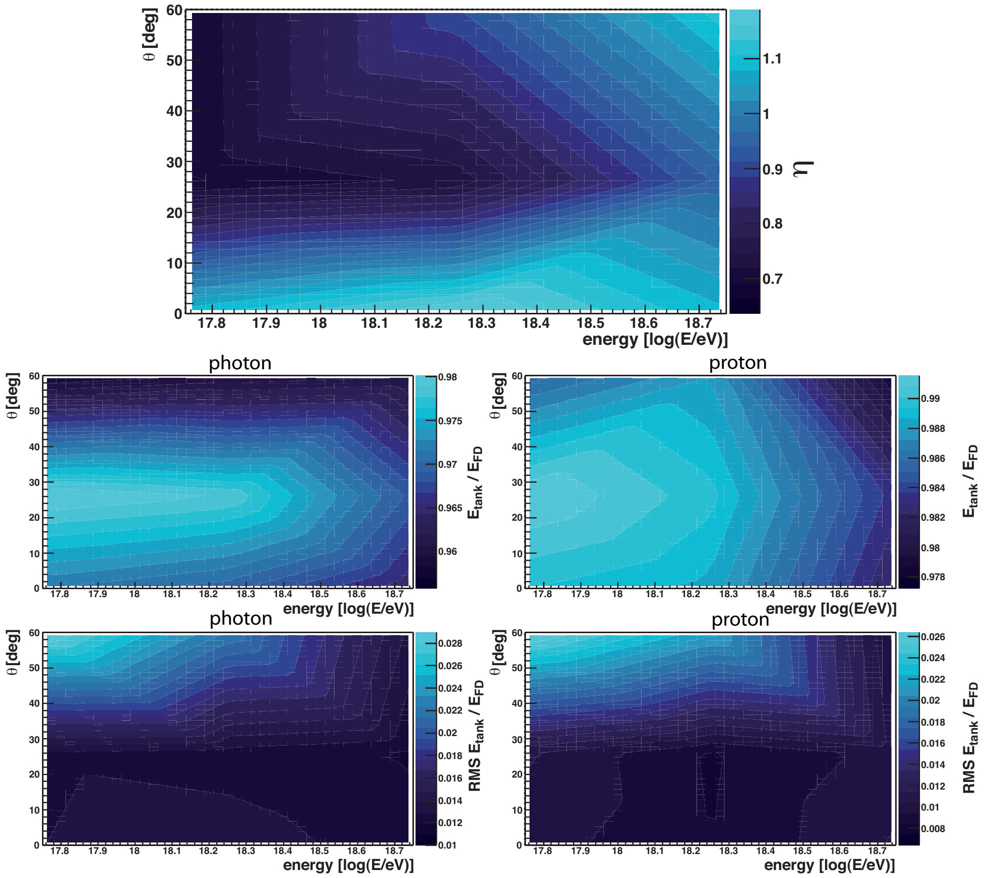


Figure 7.14: *Top panel*: Merit factor η as a function of energy and inclination angle for the ratio of tank energy and FD energy. *Middle panel*: Energy and θ dependencies of the ratio of tank energy and FD energy for photon (left) and proton (right) primaries. *Bottom panel*: Corresponding energy and θ dependencies of RMS fluctuations for photon (left) and proton (right) primaries.

and given by

$$B(r) = -(a + b \cdot r + c \cdot r^2)\ . \tag{7.8}$$

The free parameters are $a = 1.219$, $b = 0.001121$ and $c = 2.544 \cdot 10^{-7}$ (cf. [245]). Once $S_{38}(S, r, \theta)$ is determined, hybrid events are used to parameterize $S_{38}(S, r, \theta)/E^{1/\alpha}$ as a function of r with $\alpha = 1.078$. The parametrization is a NKG-type function and can be expressed as

$$\log\left(\frac{S_{38}(S, r, \theta)}{E^{1/\alpha}}\right) = p_1 + \beta(38^\circ, E) \cdot \log\left(\frac{r}{r_{\text{ref}}} \cdot \frac{(r + r_1)}{(r_1 + r_{\text{ref}})}\right) = F_{\text{LDF}}(r, E)\ , \tag{7.9}$$

with $p_1 = -15.9$, $r_1 = 700$ m and reference distance $r_{\text{ref}} = 1000$ m. The term $\beta(38^\circ, E)$ is the slope of the LDF for 38° inclined showers and energy E. The dependence of $\beta(38^\circ, E)$ with energy is found empirically by fitting β to the average LDF using stations from

$\theta = 38°$ showers at different energies

$$\beta(38°, E) = -(2 + 0.8 \cdot (\log(E) - 18)). \tag{7.10}$$

Using Eqn. (7.7) and Eqn. (7.9) the energy of the tank is given by

$$E_{\text{tank}} = \left(S_{38}(S, r, \theta) \cdot 10^{-F_{\text{LDF}}(r,E)}\right)^{\alpha} \tag{7.11}$$

Note that $\beta(38°, E)$ is determined in an iterative process since the energy is not known beforehand. The final value varies between $-2.25 < \beta(38°, E) < -2$.

The observable used in this thesis is in the end the ratio of tank energy E_{tank} and standard reconstructed FD energy E_{FD}.

Performance

The ratio of tank energy E_{tank} and FD energy E_{FD} is shown in Fig. 7.13 for various energy and theta combinations. For any given energy–θ combination, the mean ratio $E_{\text{tank}}/E_{\text{FD}}$ is less in case of photon induced showers. Noticeable is also the broader distribution of photons compared to protons. The discrimination power is analyzed in Fig. 7.14 using the merit factor. Better separation is achieved at low inclination angles where at higher energies the θ dependence starts to be less important. Also shown in Fig. 7.14 are mean and RMS values as a function of energy and inclination angle.

7.3.7 SD observable: Shape parameter

Motivation

As already stressed in the discussion of the S_b observable, the thickness of the local shower disk can be used for composition studies. The spread of arrival times at fixed shower axis distance increases for smaller production heights. Consequently, a large spread is expected in case of deep developing photon primaries (larger X_{max}) compared to proton primaries (smaller X_{max}). Note that the effect is superimposed also by geometrical effects in the relation between spread and primary composition. Also the competition between the signals from electromagnetic and muonic shower components are important. In general, however, PMT traces from surface detectors of photon induced showers should have a broader distribution compared to proton showers of the same energy, inclination and distance from the shower axis (cf. also Fig. 5.3).

Description

The so-called shape parameter is taking advantage of the aforementioned properties by analyzing the shape of the PMT trace. The original idea has been introduced by J. W. Cronin in 2003 [246] and was refined by J. A. Bellido. In principle the shape parameter is the ratio of the early arriving to the late arriving signal

$$\text{shape parameter}(r, \theta) = \frac{S_{\text{early}}(r, \theta)}{S_{\text{late}}(r, \theta)}\ . \tag{7.12}$$

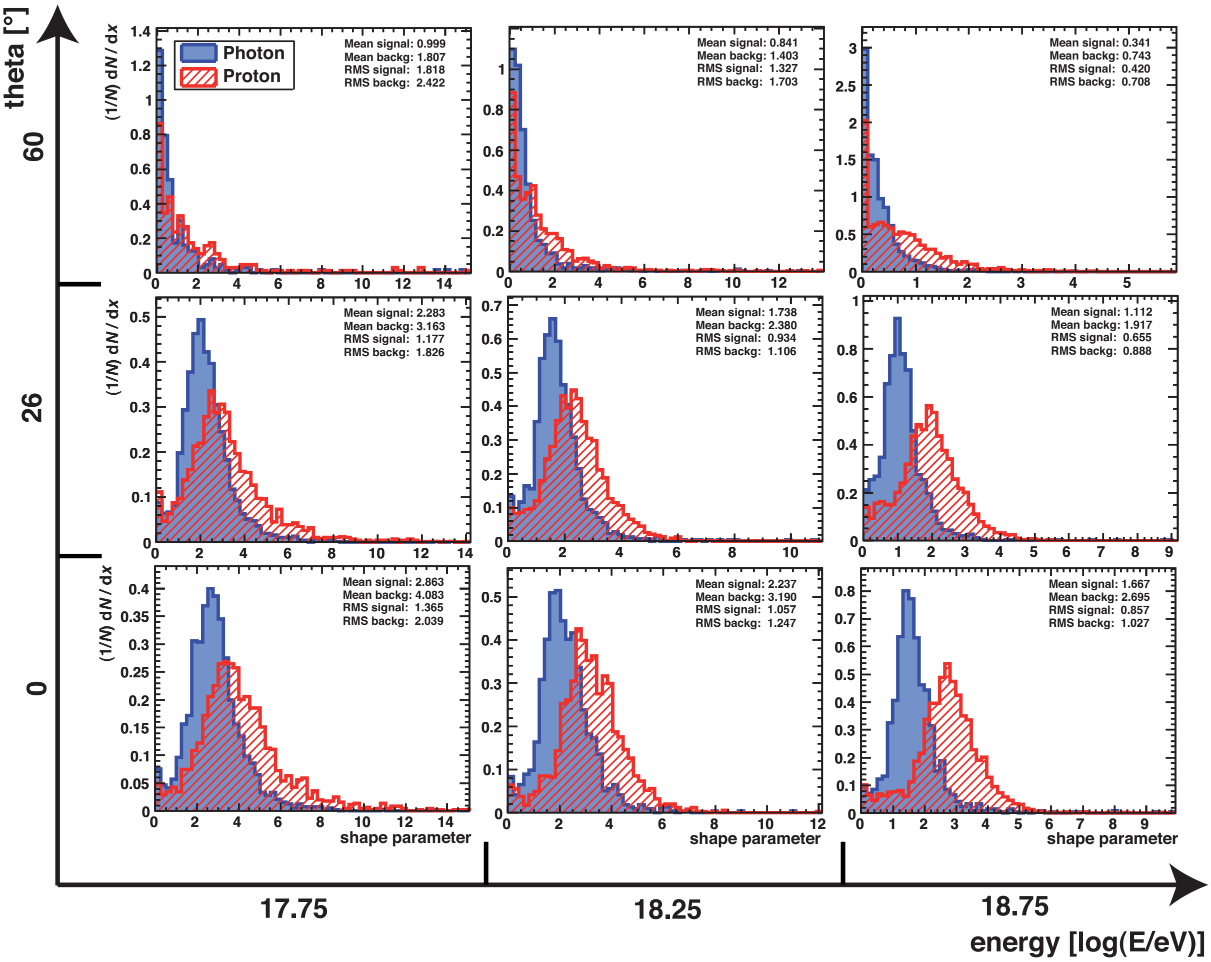


Figure 7.15: Distribution of the shape parameter for various energy / θ bins. Primary photons and protons are shown in blue and red shades, respectively. Mean and RMS values for individual energy / θ combinations are indicated, cf. Fig. 7.16. Note that "signal" always refers to primary photons and "background" to primary protons.

The early signal S_{early} is defined to be the integrated signal less than 0.6 μs beginning from the signal start slot. Correspondingly, the late signal S_{late} is the integrated signal greater than 0.6 μs until signal end. Note that the time applied here is an averaged time scaled with the inclination angle θ and distance to shower axis r and defined as

$$t_i^{\text{scaled}} = t_i \cdot \frac{r_0}{r} \cdot \frac{1}{c_1 + c_2 \cdot \cos(\theta)} \,, \tag{7.13}$$

where t_i is the real time of FADC bin i, $r_0 = 1000$ m a reference distance and $c_1 = -0.6$, $c_2 = 1.9$ scaling parameters to average traces for different zenith angles.

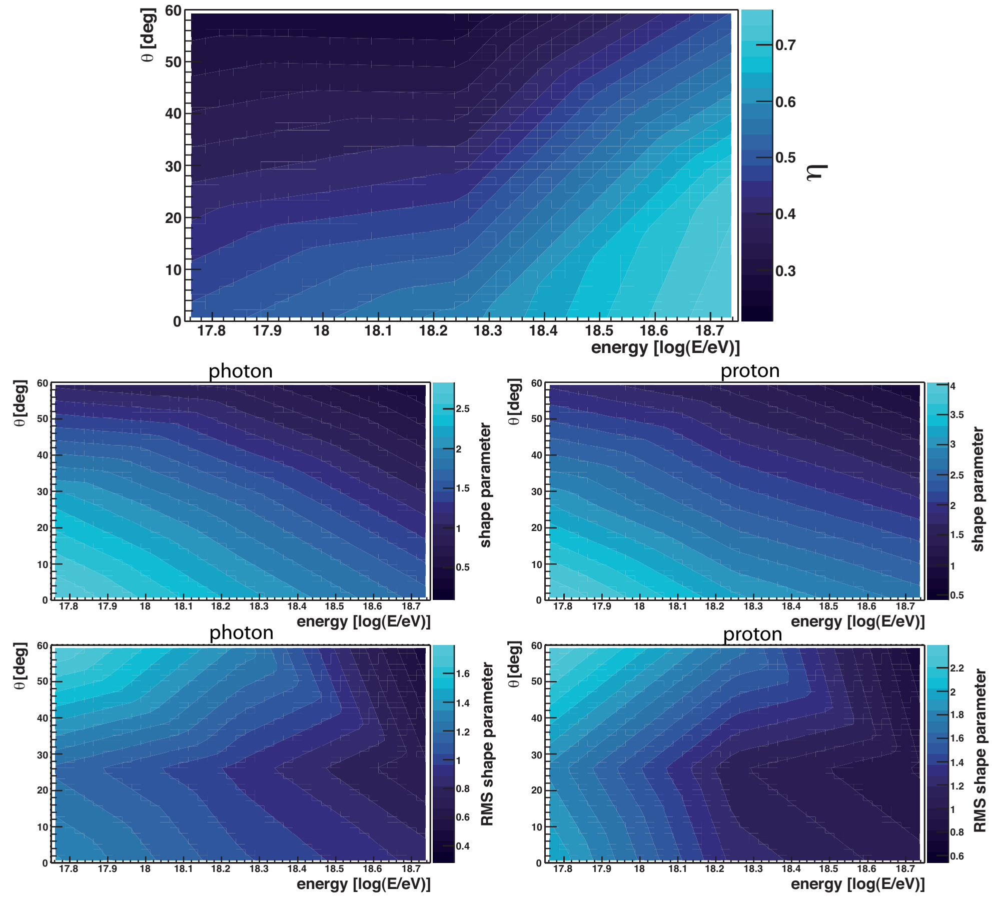


Figure 7.16: *Top panel*: Merit factor η as a function of energy and inclination angle for the shape parameter. *Middle panel*: Energy and θ dependencies of the shape parameter for photon (left) and proton (right) primaries. *Bottom panel*: Corresponding energy and θ dependencies of RMS fluctuations for photon (left) and proton (right) primaries.

Performance

The distribution of the shape parameter for photon and proton primaries is illustrated in Fig. 7.15 for several energy and θ combinations. Since broader trace distributions are expected for photon primaries corresponding mean values are shifted towards smaller numbers compared to proton induced showers. As can be seen by eye, the discrimination power is negligible for large inclinations and low energies. This is also supported using the merit factor as an indicator illustrated in Fig. 7.16. Only large primary energies and almost vertical showers have some discrimination power, but still less compared to other introduced observables. The development of mean and RMS values as a function of energy and inclination angle is also shown in Fig. 7.16.

Chapter 8

Particle Classification in a Multivariate Analysis

The purpose of this chapter is to perform particle classification in a multivariate analysis. Various observables for photon–hadron discrimination have already been introduced in Chapter 7. The combination of these parameters into one powerful response function is known as multivariate analysis and subject to many techniques that have been developed during the last decades. Some of them are addressed in this chapter. First, a brief introduction on the applied methods is given in Sec. 8.1, 8.2 and 8.3. The next step is to setup the multivariate analysis machinery (Sec. 8.4) and to examine the input data set (Sec. 8.5). The discrimination performance for photon/proton separation of boosted decision trees and artificial neural networks is analyzed and compared in Sec. 8.6. Finally the outcome is summarized in Sec. 8.7.

MOst experiments in high energy physics are multivariate in nature. That is, for each event several attributes or quantities are measured simultaneously. Analyzing and interpreting data of these experiments is a challenge for physicists and a problem in multivariate data analysis. Furthermore, multivariate analysis refers to all statistical techniques that simultaneously analyze multiple measurements on individuals or objects under investigation.

When searching for photons at ultra-high energy, only a very small photon fraction is expected against the large background of primary hadrons. As a natural consequence all available information on differences should be utilized and combined in a powerful way to enhance the discrimination power. In Chapter 7 several observables were introduced, each with individual characteristics, advantages and disadvantages. While e.g. the number of candidate tanks are especially adequate at large inclination angles, discrimination at low inclination angles is much more difficult. Here other observables like the Greisen χ^2 show better performance. Since in the univariate case the information of only one discriminating observable is exploited multivariate analysis maximizes the discrimination power using individual attributes at the same time.

The main part of the analysis is the **variate**, a linear combination of variables with empirically determined weights. *Variables* (or *observables*) are specified by the researcher, whereas the *weights* are determined by a multivariate technique to meet a specific objective. The value of a variate of n weighted variables $X_1, X_2, ..., X_n$ can be stated mathematically as:

$$\text{Variate Value} = w_1 X_1 + w_2 X_2 + \cdots + w_n X_n \ , \tag{8.1}$$

where w_n is the weight determined by the multivariate technique for variable X_n. As can be seen, the result is a single value representing a combination of the entire set of variables that best achieves the objective of the specific multivariate analysis. In the case of discriminant analysis between photons and hadrons, the variate value is formed so as to create scores for each observation that maximally differentiates between photon and hadron primaries.

Certainly, the determination of the individual weights is a dedicated task and plenty of techniques have been developed over the last couple of decades to improve the separation power in discriminant analysis. Each technique has its own advantages, drawbacks and philosophy. Simplicity is a virtue if it is not at the expense of significant discrimination power. Also robustness with respect to overtraining (cf. Sec. 8.1.2) could become an issue if the training sample is scarce.

The following sections summarize some common multivariate techniques used in this thesis with special attention to boosted decision trees and artificial neural networks. After a brief introduction on the general idea, a more detailed description is given. Finally, the performance in general is discussed including strengths and shortcomings. The implementation of these techniques is orientated on the Toolkit for Multivariate Analysis (TMVA) [247, 248] version 4.0.3 that provides a `ROOT`–integrated machine learning environment. It is specially designed to the needs of high-energy physics applications and consists of object-orientated implementations in `C++`. More details on the TMVA workflow can be found in Sec. 8.4.

8.1 Boosted decision trees

8.1.1 Introduction

A *decision tree* is a binary tree structured classifier where decisions (true – false) are taken on one single variable at a time until a certain stopping criterion is fulfilled. Each decision divides the phase space into two distinct regions. The full tree consists of many regions eventually classified as signal or background *leaf*, depending on the majority of signal or background events in the leaf. A schematic view of a decision tree is shown in Fig. 8.1. However, one single decision tree is generally quite unstable and the concept is

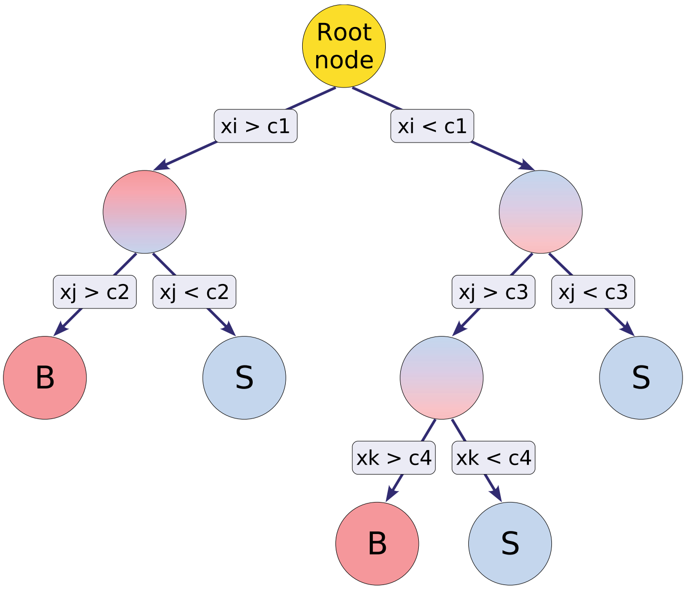


Figure 8.1: Illustration of a decision tree adopted from [247]. A series of binary splits using the discriminating observable x_i is applied to the data starting from the root node. For each split the best separating variable is used resulting that the same variable may be used at several nodes while others might not be used at all. Final leaf nodes are labeled "S" and "B" if the majority are signal and background events, respectively.

extended by *boosting* a decision tree to get several reweighted trees which form a forest. The final response output is then a combination of each individual tree. This multivariate technique is called boosted decision tree (BDT).

8.1.2 Description and implementation

While cut-based analysis is able to select only one hypercube as a region of phase space, a decision tree is able to split the phase space into a large number of hypercubes, each identified as "signal-like" or "background-like". The workflow for boosted decision trees is described below:

- **Building a decision tree**:
  As shown in Fig. 8.1 the training starts with the root node, where an initial splitting criterion for the full training sample is determined searching for the best separating observable. A variety of selection criteria exists[1] but tests have revealed no significant performance disparity [247]. In this thesis the G_{ini} index as selection criteria is used defined as

$$G_{\text{ini}} = \mathcal{P} \cdot (1 - \mathcal{P}) \ , \tag{8.2}$$

---

[1] e.g. G_{ini} index, cross entropy, misclassification error, statistical significance or average squared error

where $\mathcal{P}$ is the purity of the signal for a given cut defined as

$$\mathcal{P} = \frac{\sum_S w_S}{\sum_S w_S + \sum_B w_B} \ . \tag{8.3}$$

Here $\sum_S$ and $\sum_B$ are the number of signal and background events above a given cut with the corresponding event weights w_S and w_B, respectively. Note that $G_{\text{ini}} = 0$ ($\mathcal{P} = 1$) indicates if the sample is pure signal whereas $G_{\text{ini}} = 0.25$ ($\mathcal{P} = 0.5$) denotes a fully mixed sample. After the first splitting using the best separating variable, the two subsets are again examined and each split is determined by finding the observable and corresponding cut value that provides best separation between signal and background. This process continues until a minimum number of events in each subset is achieved. The final subset is called a signal leaf if $\mathcal{P} > 0.5$ and background leaf if $\mathcal{P} < 0.5$.

- **Boosting a decision tree**:
  Although decision trees have been known for some time [249] the idea of *boosting* decision trees is a relatively new technique [250, 251] which turned out to be one of the most powerful learning techniques introduced during the past decade [252]. The shortcoming of decision trees is their instability with respect to statistical fluctuations in the training sample. Consider two input variables A and B exhibit similar separation power. A fluctuation in the training sample could cause the algorithm to split using A while B could have been selected without that fluctuation. Below this node the tree structure is altered and possibly resulting in a substantially different classifier response. This problem is overcome by the boosting technique as a way of enhancing the classification of typically weak multivariate methods by reweighting (boosting) versions of the training data and finally taking a weighted majority vote. Three methods are briefly discussed. A more detailed description can be found in [247]:

  1. *Adaptive Boost (AdaBoost)*: This is currently the most popular boosting algorithm. The basic concept is that events that were misclassified during the training of a decision tree are given a higher event weight[2] α before training the following tree, expecting a better separation of misclassified events in the subsequent decision tree. Finally the event weights are normalized such that the sum of weights remains constant. Let $h(\vec{x})$ (with $\vec{x}$ being the tuple of input observables) be the result of the classifier encoded for signal and background as $h(\vec{x}) = 1$ and -1, respectively. The boosted classifier response is then given by

  $$y_{\text{Boost}}(\vec{x}) = \frac{1}{N_{\text{forest}}} \cdot \sum_{i}^{N_{\text{forest}}} \ln(\alpha_i) \cdot h_i(\vec{x}) \ , \tag{8.4}$$

     with the total number of decision trees N_{forest}.

---

[2] The misclassified events are multiplied by a common boost weight α derived from the misclassification rate, err, of the previous tree: $\alpha = (1 - \text{err})/\text{err}$.

2. *Gradient Boost (Grad)* [253, 254]: Gradient boosting is an amelioration of the already discussed adaptive boost technique. The general problem in function estimating is as follows: Consider a set of random input variables $\vec{x} = \{x_1, x_2, ..., x_n\}$ and a random output or response variable y. Given a training sample $\{y_i, \vec{x}_i\}_1^N$ of known $(y, \vec{x})$-values, the goal is to find a function $F(\vec{x})$ that maps $\vec{x}$ to y in such a way that a specified loss function $L(F, y)$ (which is a quantity for the deviation to the true value) is minimized, cf. [254]. Adaptive boost is based on an exponential loss function $L^{\text{AdaBoost}}(F, y) = e^{-F(\vec{x})y}$ which leads to the weighting algorithm described above. However, exponential loss has the disadvantage that it lacks robustness in presence of outliers or mislabeled data points with the consequence that the performance degrades in noisy settings. The gradient boosting algorithm attempts to cure this weakness by introducing a binomial log-likelihood loss
$$L^{\text{Grad}}(F, y) = \ln\left(1 + e^{-2F(\vec{x})y}\right) \quad , \tag{8.5}$$
for classification as a more robust loss function without giving up on the good out-of-the-box performance of AdaBoost. The minimization is done by an iterative procedure in a steepest-descent approach which has some drawbacks in computation time. In general gradient boosting is less susceptible to overtraining. The combination of gradient boosting and a bagging-like resampling procedure (see next item) is called *stochastic gradient boosting* which may further enhance the separation power.
3. *Bagging*: Although not a genuine boosting algorithm, bagging (or bootstrap aggregating) is discussed in this context. It is a way of smearing over statistical representations of the training data and suited to stabilize the response of a classifier often accompanied by a performance increase. The idea is to repeatly resample the training sample in a way that the same event is allowed to be picked several times from the parent sample making the training sample a representation of the probability density distribution of the parent sample. Training the classifiers with different parent distributions and combining them into a collection is finally more stable with respect to statistical fluctuations in the training sample.

- **Pruning a (boosted) decision tree**:
  By definition decision trees are susceptible for *overtraining*. This is when a machine learning program has too few degrees of freedom, because too many model parameters of a classifier were adjusted to too few data points. For clarification suppose the decision tree is build until only one event is left in each leaf suggesting that perfect discrimination is achievable since each leaf consists of either pure signal or pure background events. However, applying these rules to an independent data set (test sample) would cause a significant performance decrease. A convenient way to detect overtraining and to measure its impact is therefore to compare the classification results between training and test samples, cf. also Sec. 8.4. To counteract

overtraining in decision trees, several pruning methods have been developed with the objective to remove insignificant nodes. Here only the *cost complexity pruning* [249] is discussed. The idea is to relate the number of nodes in a subtree below a node to the gain in terms of misclassified events by the subtree compared to the node itself with no further pruning. The misclassification rate of a node is given by $R = 1 - \max(\mathcal{P}, 1 - \mathcal{P})$ and the corresponding cost complexity of this node is

$$\rho = \frac{R(\text{node}) - R(\text{subtree below that node})}{\#\text{nodes}(\text{subtree below that node}) - 1} \ . \tag{8.6}$$

By scanning over all nodes, the node with smallest ρ is recursively pruned away until a user defined prune strength.

8.1.3 Performance

Since boosted decision trees are a relatively new technique only limited experience has been gained so far. However, they are often referred as the best "out-of-the-box" method and little tuning is required to obtain reasonably good results. Since each step involves only a one-dimensional cut optimization the method is simple and not a "black box". Another advantage is the insensitivity to non-discriminating variables as they are simply not used (in contrast to e.g. neural networks discussed in Sec. 8.2). The simplicity, however, has the drawback that their theoretically best performance on a given problem is generally inferior to other techniques.

8.2 Artificial neural networks

8.2.1 Introduction

Inspired by the structure of biological neural networks, artificial neural networks (ANN) were first formulated by Warren McCulloch and Walter Pitts in 1943 [255]. During the next decades ANN were continuously refined and are today an inherent part in multivariate analysis.

Generally speaking ANN are a simulated collection of interconnected neurons, while each neuron produces a certain response at a given set of input signals. In case of signal and background discrimination, a neural network maps from a space of input variables $\vec{x} = \{x_1, x_2, ..., x_n\}$ onto a one-dimensional space of output variable y_{ANN}, cf. Fig. 8.2.

8.2.2 Description and implementation

There are a variety of different implementations of artificial neural networks. Besides the Clermont-Ferrand and a `ROOT`-integrated network, a newly developed multilayer perceptron (MLP) network is mostly used in this thesis. Note, however, that all neural networks are feed-forward multilayer perceptrons (cf. next item). For more details see also [247].

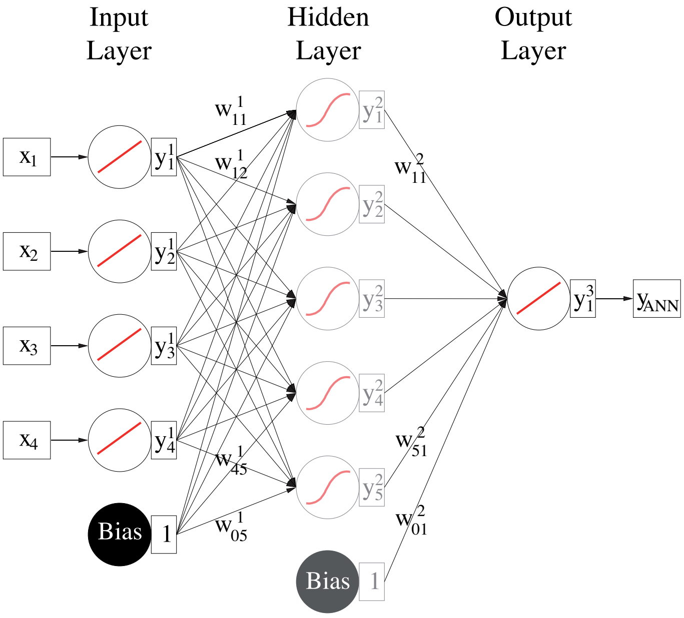


Figure 8.2: Schematic view of an artificial neural network with four input variables x_1, x_2, x_3 and x_4, one hidden layer with five neurons and output layer with output y_{ANN}. Each neuron is connected with all subsequent neurons whereupon each bonding has its own weight w as indicated. The neuron activation function is symbolized by the red lines. Two bias nodes are also introduced. They are connected with nodes in a higher layer node, but no node regardless of type connects to any bias node. They account for some special cases that cause the algorithm to break, namely when all node weights approach zero.

The main components of an artificial neural network are the layout of the neurons, the weights of the inter-neuron connections and the response of the neuron to an input signal denoted by the neuron response function ρ:

- *Layout of the neurons*: Theoretically a neural network with n neurons could have n^2 directional connections. This complexity can be reduced by organizing neurons into layers while each layer can only interact to the following layer (cf. Fig. 8.2) which is called multi-layer perceptron. In classification problems, the first layer is the input layer that holds the input variables $\{x_1, x_2, ..., x_{n_{\text{var}}}\}$. Subsequent layers are hidden layers while the last layer holds the output variable, the neural net estimator y_{ANN}. The optimal configuration of hidden layers and number of neurons has to be determined for each classification problem individually.

- *Neuron response function* ρ: This function maps the neuron input onto the neuron output and is often separated into a $\mathbb{R}^n \to \mathbb{R}$ synapse function κ, and a $\mathbb{R} \to \mathbb{R}$ neuron activation function α in a way that $\rho = \alpha \circ \kappa$.

- *Weights of the inter-neuron connections*: The most common way for adjusting the weights that optimize the classification performance is the so-called *back propagation*

(BP). In a first phase N training events are propagated through the neural network to generate the propagation's output activations. This output is compared to the desired output $\widehat{y}_a \in \{1, 0\}$ (1 for signal events and 0 for background events). The error function E is a quantity for the deviation to the expected value and given by

$$E(\vec{x}_1, ..., \vec{x}_n|\vec{w}) = \sum_{a=1}^{N} \frac{1}{2}(y_{\mathrm{Ann},a} - \widehat{y}_a)^2 \ , \tag{8.7}$$

where $\vec{w}$ denotes the ensemble of adjustable weights. Then a back propagation of the propagation's output activations through the neural network is performed using the training pattern's target in order to generate the error of all output and hidden neurons. In a second phase the individual weights are updated using the method of steepest or gradient decent. The method is repeated several times to find the minimum of Eqn. (8.7).

An alternative to the introduced back propagation is the Broyden-Fletcher-Goldfarb-Shannon (BFGS) method [256]. It differs from back propagation by the use of second derivatives of the error function to adapt the synapse weight by a varied algorithm. More details can be found in [247].

8.2.3 Performance

Neural networks are a current method in high energy physics. They are stable and facilitate a large variety of linear and non-linear classification problems. However, in contrast to e.g. boosted decision trees (BDT), the user is advised to reduce the number of input variables that have only little discrimination power. Neural networks are also susceptible for overtraining and also the transparency of the method is poor and more a black box.

8.3 Fisher discriminants

8.3.1 Introduction

Fisher discrimination [257] is a method for discrimination of up to linear correlations among the input variables. The idea is to determine an axis in the (correlated) hyperspace of input variables such that, when projecting the true classification (signal or background) upon the axis, they are pushed as far as possible from each other, while events of the same class are confined in close vicinity.

8.3.2 Description and implementation

The classification relies on the following ingredients: the overall sample means $\bar{x}_k$ for each input variable $k = 1, ..., n_{\mathrm{var}}$, the class specific sample means $\bar{x}_{\mathrm{S(B)},k}$ and the total covariance matrix C of the sample which can be decomposed into a within- (W) and between-class matrix (B). The final Fisher discriminant y_{Fi} for event i is given by

$$y_{\mathrm{Fi}}(i) = F_0 + \sum_{k=1}^{n_{\mathrm{var}}} F_k x_k(i) \ . \tag{8.8}$$

The Fisher coefficients F_k are given by

$$F_k = \frac{\sqrt{N_S N_B}}{N_S + N_B} \sum_{l=1}^{n_{\mathrm{var}}} W_{kl}^{-1} (\bar{x}_{S,l} - \bar{x}_{B,l}) \ , \tag{8.9}$$

with the number of signal and background events in the training sample $N_{S(B)}$, respectively. The offset F_0 centers the sample mean $\bar{y}_{\mathrm{Fi}}$ of all N_S and N_B events to zero.

8.3.3 Performance

In certain cases Fisher discriminants can be competitive with likelihood and non-linear discriminants. In particular they are optimal for Gaussian distributed variables with linear correlations. However, no discrimination at all is achieved when a variable has the same mean for signal and background even if the shape is different. They are transparent and can hardly be overtrained.

8.4 Multivariate analysis workflow

TMVA provides a full workflow for multivariate analysis and it can be divided into two phases, namely the *training* and the *application phase*. The training phase comprises different steps which are summarized below:

1. *Providing a training sample*: In this step the complete training sample with known composition is provided.

2. *Booking observables*: The training sample should consist of all necessary information for discrimination. Favored observables are selected and passed to the analysis procedure.

3. *Prepare training and test sample*: It is advisable to split the complete training sample into a learning (or training) and a test sample. The learning sample is passed to individual multivariate analysis (MVA) methods to “learn” how to discriminate between signal (photon) and background (proton) events. Once the learning phase is accomplished, the results are applied on an independent test sample to see the performance and possible overtraining effects. In this analysis the full data sample is divided randomly into two equally sized subsamples for training and testing. However, it is also advisable to create a third test sample which is untouched until the final configuration of the MVA method is found to avoid a bias introduced when

Observable	**Abbreviation**	**Unit**	**Description**
Depth of shower maximum	X_{max}	g/cm^2	FD observable, cf. Sec. 7.3.1
Reduced Greisen χ^2	χ^2_{Greisen}		FD observable, cf. Sec. 7.3.2
Greisen energy / FD energy	$E_{\text{Greisen}}/E_{\text{FD}}$		FD observable, cf. Sec. 7.3.3
S_b	S_3	VEM	SD observable, cf. Sec. 7.3.4
Number of candidate stations	$N_{\text{Candidates}}$		SD observable, cf. Sec. 7.3.5
Tank energy / FD energy	$E_{\text{Tank}}/E_{\text{FD}}$		SD observable, cf. Sec. 7.3.6
Shape parameter	ShapeP		SD observable, cf. Sec. 7.3.7
FD energy	E_{FD}	$\log(E/\text{eV})$	For energy correlations
Inclination angle	θ	deg	For theta correlations

Table 8.1: Summary of the observables used in multivariate analysis. Besides three FD and four SD observables also the reconstructed energy and the inclination angle is included to account for possible correlations.

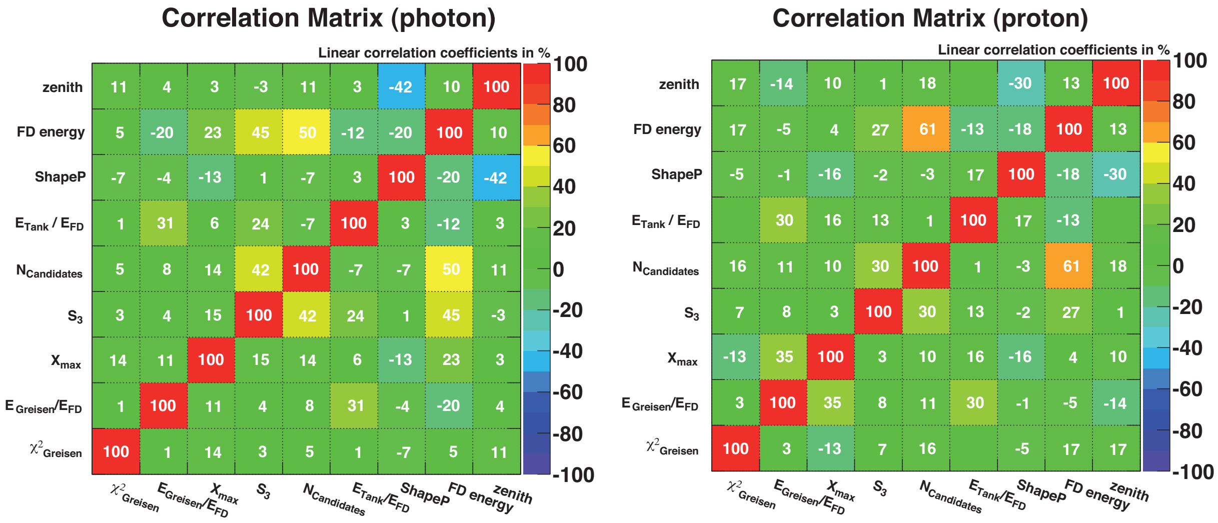
repeating the learning and testing phase several times [258].
If necessary, individual events are reweighted to adopt different input spectra or to emphasize specific event features. Also additional quality cuts may be applied.

4. *Multivariate Analysis*: Here one or more MVA techniques are applied on the training sample and the performance is checked using the test sample. Once an optimal configuration is found, the configuration is saved.

After the training phase is accomplished the trained method can be applied to “real data” where signal and background events are not known beforehand. In this application phase each event is flagged with the output response of the classifier and it is up to the user where to cut on the response variable for signal selection. More details can be found in Sec. 9.2.

8.5 Examination of the input data set

Before performing a multivariate analysis, a detailed examination of the input data is essential. The simulated data should be the most accurate copy of measured hybrid events and consideration of correlations among input variables is relevant when maximizing the performance of the method. As already shown in Sec. 7.3 the discrimination power of most observables is dependent of the primary energy and zenith angle. To account for this coherence, the reconstructed energy E_{FD} as well as the zenith angle θ is included in the set of observables for photon hadron discrimination. Table 8.1 summarizes the applied observables. As a measure of the strength and direction of a linear relationship between two variables, the linear correlation coefficients in % are shown in Fig. 8.3 for photon and proton primaries individually. The correlation coefficient between variable x



Correlation Matrix (photon)

Linear correlation coefficients in %

	$\chi^2_{Greisen}$	$E_{Greisen}/E_{FD}$	X_{max}	S_3	$N_{Candidates}$	E_{Tank}/E_{FD}	ShapeP	FD energy	zenith
zenith	11	4	3	-3	11	3	-42	10	100
FD energy	5	-20	23	45	50	-12	-20	100	10
ShapeP	-7	-4	-13	1	-7	3	100	-20	-42
E_{Tank} / E_{FD}	1	31	6	24	-7	100	3	-12	3
$N_{Candidates}$	5	8	14	42	100	-7	-7	50	11
S_3	3	4	15	100	42	24	1	45	-3
X_{max}	14	11	100	15	14	6	-13	23	3
$E_{Greisen}/E_{FD}$	1	100	11	4	8	31	-4	-20	4
$\chi^2_{Greisen}$	100	1	14	3	5	1	-7	5	11

Correlation Matrix (proton)

Linear correlation coefficients in %

	$\chi^2_{Greisen}$	$E_{Greisen}/E_{FD}$	X_{max}	S_3	$N_{Candidates}$	E_{Tank}/E_{FD}	ShapeP	FD energy	zenith
zenith	17	-14	10	1	18		-30	13	100
FD energy	17	-5	4	27	61	-13	-18	100	13
ShapeP	-5	-1	-16	-2	-3	17	100	-18	-30
E_{Tank} / E_{FD}		30	16	13	1	100	17	-13	
$N_{Candidates}$	16	11	10	30	100	1	-3	61	18
S_3	7	8	3	100	30	13	-2	27	1
X_{max}	-13	35	100	3	10	16	-16	4	10
$E_{Greisen}/E_{FD}$	3	100	35	8	11	30	-1	-5	-14
$\chi^2_{Greisen}$	100	3	-13	7	16		-5	17	17

Figure 8.3: Linear correlation matrix for photon (left) and proton (right) primaries. A positive value indicates a positive correlation where 100% indicates a full positive correlation, zero no correlation and -100% a full anti-correlation. Differences between proton and photon correlations may already suggest features that can be utilized in the multivariate analysis.

and y is given by

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)\sigma_x \sigma_y} \ , \tag{8.10}$$

where $\bar{x}$ and $\bar{y}$ are the sample means of x and y with the corresponding standard deviation σ_x and σ_y. As can be seen in Fig. 8.3, some observables are strongly correlated, e.g. E_{FD} and the number of candidate tanks which is expected, since N_{Tank} grows with increasing energy. It should be emphasized that only linear correlations are computed. Higher dimensional correlations can not be detected using Eqn. (8.10). However, sophisticated multivariate techniques, e.g. BDT or ANN, can also cope higher dimensional correlations. To visualize higher dimensional correlations, scatter plots of all possible combinations of input observables are shown in Figs. C.1, C.2 and C.3 in the appendix.

8.6 Testing multivariate analysis techniques for photon/hadron separation

In this section multivariate analysis techniques are applied for the discrimination of photon and proton induced extensive air showers. Special attention is given to boosted decision trees (cf. Sec. 8.6.1) and artificial neural networks (cf. Sec. 8.6.2), but also other techniques are discussed. A comparison between individual techniques is given in Sec. 8.6.3.

All methods are trained and tested with MC data. The raw data consists of 250000 protons and 75000 photons following a spectral index of $\gamma = 2.7$ and simulated with the $\overline{\text{Off}}\underline{\text{line}}$ framework, cf. Sec. 7.1. Basic selection cuts are applied on triggered events as

Basic selection cuts
Number of telescope stations > 0: Assure at least one triggered eye
Successful energy reconstruction
Flagged hybrid event
Require SD trace information

Table 8.2: List of basic selection cuts before applying sanity cuts listed in Table 8.3. The intention of these cuts is to create a data sample where all necessary information for multivariate analysis is included, e.g. successful energy reconstruction and SD trace information.

Sanity cuts
$\chi^2_{\text{Greisen}} < 20$
$X_{\text{max}} < 3000$ g/cm^2
$0 < N_{\text{Candidates}} < 8$
Shape parameter < 15
$S_3 < 5000$ [VEM]
$17.2 < E_{\text{FD}}$ [log(E/eV)] < 18.5
$\theta < 60°$

Table 8.3: Sanity cuts applied in this thesis after basic selection cuts, listed in Table 8.2.

listed in Table 8.2 with a photon (proton) cut efficiency of $\gamma_{\text{eff}} = 0.22$ ($p_{\text{eff}} = 0.19$). Note that the spectral index is reweighted to $\gamma = 3.0$ during the multivariate analysis to come up with a more realistic data sample. Also some basic sanity cuts are applied that are summarized in Table 8.3. These cuts aim at rejecting a few outlier events. Tests have also been performed with more stringent quality cuts, e.g. requiring X_{max} in the field of view or $\chi^2_{\text{linear}} - \chi^2_{\text{GH}} > 4$ (χ^2_{linear} and χ^2_{GH} denotes the χ^2 of a linear and a Gaisser-Hillas fit to the longitudinal profile, respectively). However, besides a slight enhancement in separation power the approach has some severe drawbacks. Searching for photons at EeV energies has the advantage of exploiting a large number of events[3], but the reconstruction accuracy of these events are in general worse compared to higher energy events. Applying stringent quality cuts would have significant effects on statistics as shown below.

Sanity cuts, cf. Table 8.3	**X_{max} in field of view**	$\chi^2_{\text{linear}} - \chi^2_{\text{GH}} > 4$
$\gamma_{\text{eff}} = 0.97$	$\gamma_{\text{eff}} = 0.68$	$\gamma_{\text{eff}} = 0.31$
$p_{\text{eff}} = 0.93$	$p_{\text{eff}} = 0.65$	$p_{\text{eff}} = 0.31$

Note that the basic selection cuts (Table 8.2) are applied to all data samples beforehand. Using just sanity cuts keeps most of the data while additional cuts (after applying sanity cuts) remove 1/3 or more of the original data set. It is decided to use just sanity cuts accepting a slight decrease in separation power.

8.6.1 Boosted decision trees

The multivariate method of boosted decision trees introduced in Sec. 8.1 is applied to the MC input data set described above. Several parameter combinations have been tested and a selection is given in Table 8.4.

The resulting classifier output response of two methods (in this case `BDTStandard` and `BDTGBagShrink`) is shown in Fig. 8.4. Training and test sample are superimposed to detect possible overtraining. The shape of the two response functions looks different

---

[3] Besides some other advantages, e.g. the absence of the LPM and preshower effect (cf. Sec. 5.4).

Method	Option	Description

`BDTStandard`*	`NTrees=400,`
	`nEventsMin=400,`
	`MaxDepth=3,`
	`BoostType=AdaBoost,`
	`SeparationType=GiniIndex,`
	`nCuts=20,`
	`PruneMethod=NoPruning`
		Standard options for BDT using adaptive boost algorithm `AdaBoost` (cf. Sec. 8.1.2). Other options refer to the number of trees (`NTrees`), minimum number of events required in a leaf node (`nEventsMin`), maximum depth of allowed decision tree (`MaxDepth`), G_{ini} as selection criteria (`SeparationType`), number of steps during node cut optimization (`nCuts`) and no pruning.

`BDTNTrees600`	`NTrees=600`	Variation in the number of trees.
`BDTDCost`	`PruneMethod=CostComplexity,` `MaxDepth=30`	Cost complexity pruning and an increased maximum depth.
`BDTNCuts30`	`nCuts=30`	Variation in `nCuts`.

`BDTG`*	`NTrees=1000,`
	`BoostType=Grad,`
	`Shrinkage=0.30,`
	`UseBaggedGrad,`
	`GradBaggingFraction=0.6,`
	`SeparationType=GiniIndex,`
	`nCuts=20, NNodesMax=5`
		Standard options for BDT using gradient boost algorithm `Grad` (cf. Sec. 8.1.2). Other options refer to the number of trees (`NTrees`), learning rate for the gradient boosting (`Shrinkage`), using a random subsample for growing the trees (`UseBaggedGrad`), the fraction of events to be used in each iteration (`GradBaggingFraction`), G_{ini} as selection criteria (`SeparationType`), number of steps during node cut optimization (`nCuts`) and maximum number of nodes in a tree (`NNodesMax`).

`BDTGNMax10`	`NNodesMax=10`	Variation in `NNodesMax`.
`BDTGNCuts10`	`nCuts=10`	Variation in `nCuts`.
`BDTGBagFrac`	`GradBaggingFraction=0.8,` `NNodesMax=10`	Variation in `GradBaggingFraction` and `NNodesMax`.

`BDTGBagShrink`	`Shrinkage=0.10,`
	`GradBaggingFraction=0.7,`
	`NNodesMax=10`
		Variation in `Shrinkage`, `GradBaggingFraction` and `NNodesMax`.

`BDTGBagShrinkTrees`	`NTrees=2000,`
	`Shrinkage=0.10,`
	`GradBaggingFraction=0.7,`
	`NNodesMax=10`
		Variation in `NTreesShrinkage`, `GradBaggingFraction` and `NNodesMax`.

`BDTB`*	`NTrees=400,`
	`BoostType=Bagging,`
	`SeparationType=GiniIndex,`
	`nCuts=20,`
	`PruneMethod=NoPruning`
		Using a bagging boosting algorithm.

Table 8.4: Summary of various BDT settings used in this thesis. Three different boosting algorithm were tested and are separated by several horizontal lines. Methods marked with a '*' indicate the default parameters for a particular boosting algorithm while the options of other methods just indicate deviations from the default settings. A brief description is also given. More details on individual steering parameters are given in [247].

since different boosting algorithm were applied. In case of `AdaBoost` (`BDTStandard`) a more Gaussian shape is achieved while using `Grad` boosting (`BDTGBagShrink`), photon

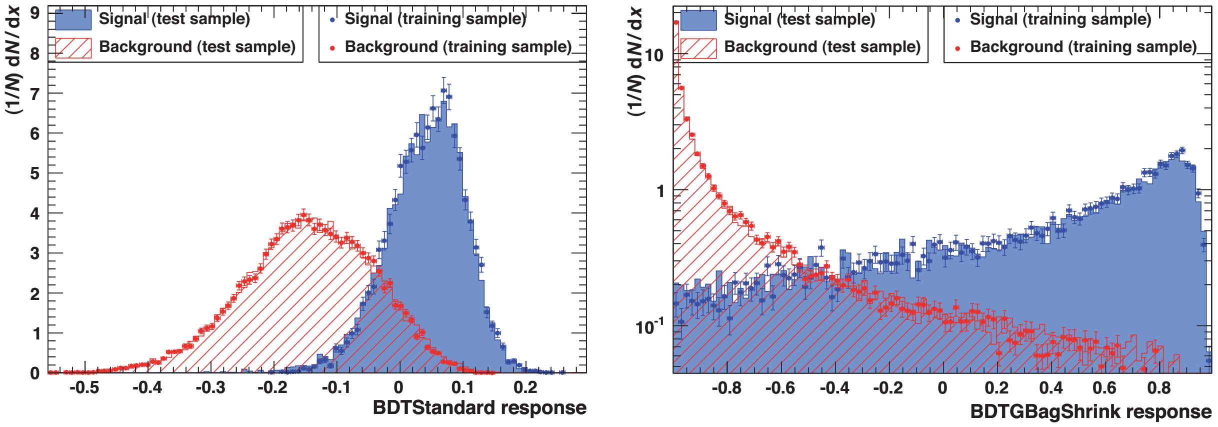


Figure 8.4: Classifier response value of the two methods `BDTStandard` (*left panel*) and `BDTGBagShrink` (*right panel*). "Signal" and "Background" always refers to primary photons and protons, respectively. The shaded area histograms correspond to the test sample while dots represent the training sample. A good agreement between training and test distribution indicates that overtraining is under control. The shape of the two distributions look rather different since different boosting algorithm were applied.

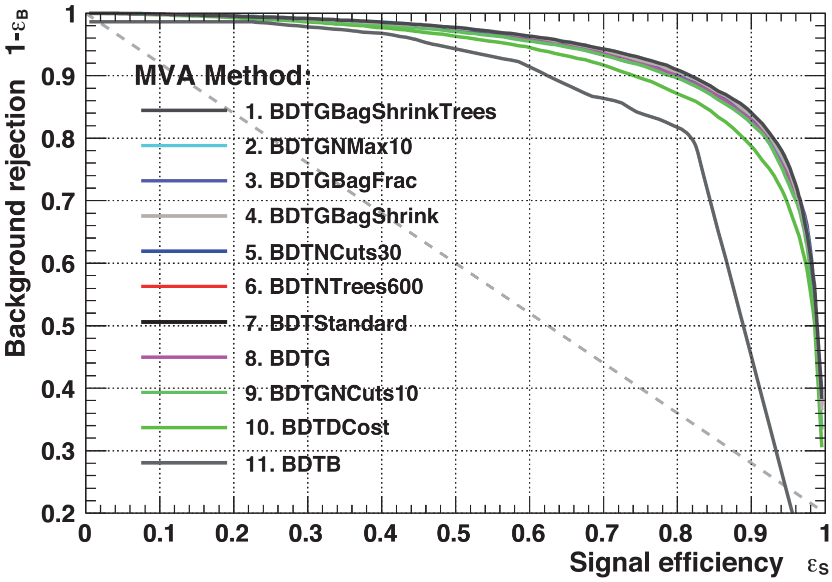


Figure 8.5: Receiver operating characteristic (ROC) diagram of various BDT methods described in Table 8.4. The methods are ranked by the integral of the ROC diagram showing the best method on top. Except `BDTB` and `BDTDCost` all methods show similar performance. The faint dashed diagonal line symbolizes a "random guessing" method.

and proton peaks are dispersed with long tails. At this stage it is rather difficult to decide the most powerful method for classification. There is no unique way to express the performance of a classifier, but some benchmark quantities can be computed. A prominent way to determine the performance of a method is to examine the *receiver*

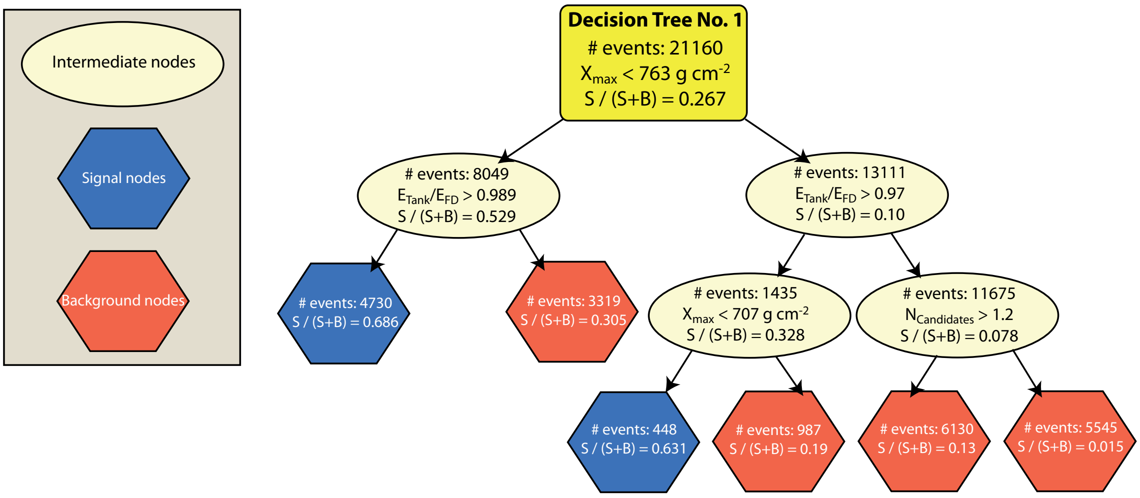


Figure 8.6: Example of the first decision tree using the `BDTGBagShrink` method. Split decision and the surviving number of events are given as well as the resulting signal purity $S/(S+B)$. Finally all events are classified as either signal (blue boxes) or background (red boxes).

Rank	Observable	Variable Importance
1	X_{max}	$2.836 \cdot 10^{-1}$
2	$E_{\text{Tank}}/E_{\text{FD}}$	$1.701 \cdot 10^{-1}$
3	$E_{\text{Greisen}}/E_{\text{FD}}$	$1.197 \cdot 10^{-1}$
4	E_{FD}	$1.102 \cdot 10^{-1}$
5	θ	$9.981 \cdot 10^{-2}$
6	ShapeP	$9.836 \cdot 10^{-2}$
7	S_3	$4.405 \cdot 10^{-2}$
8	χ^2_{Greisen}	$3.996 \cdot 10^{-2}$
9	$N_{\text{Candidates}}$	$3.429 \cdot 10^{-2}$

Table 8.5: Importance ranking of the input observables in case of `BDTGBagShrink`.

operating characteristic (ROC) diagram here defined as the signal efficiency of the method vs. background rejection. The signal (background) efficiency $\epsilon_{\text{S(B)}}$ of a certain MVA method is defined as

$$\epsilon_{\text{S(B)}}(i) = \frac{N_{\text{S(B)}}(i)}{N^{\text{tot}}_{\text{S(B)}}} \,, \tag{8.11}$$

where $N^{\text{tot}}_{\text{S(B)}}$ is the total number of signal (background) events and $N_{\text{S(B)}}(i)$ the number of events above[4] a certain cut i. The background rejection is then simply defined as $1 - \epsilon_{\text{B}}$. The ROC diagram for individual methods introduced in Table 8.4 is shown in Fig. 8.5. The area under the ROC diagram is an indicator for the performance of the

---

[4] By convention signal (background) events accumulate at large (small) classifier output values.

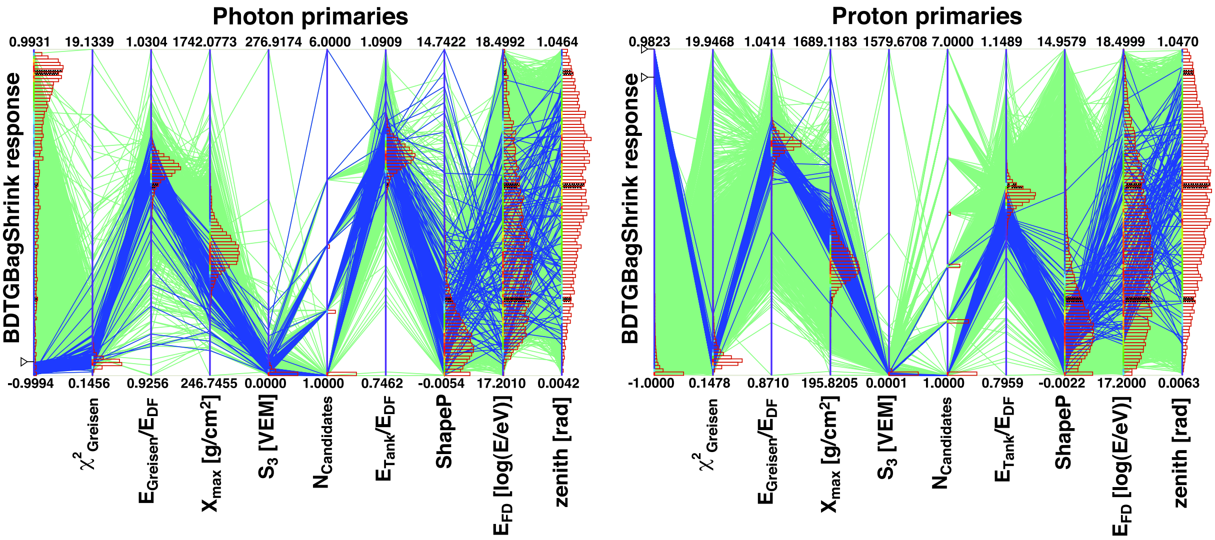


Figure 8.7: Parallel coordinates of the `BDTGBagShrink` method for photon (*left*) and proton (*right*) primaries. Each event is drawn by a green polyline connecting parallel vertical axis representing all input observables and the classifier response. The position of the line at the vertical axis corresponds to the individual value of that event. The frequency scale of each observable is shown by a red histogram superimposed to the vertical axis. Blue polylines highlight events that are misclassified by the classifier response.

method. As can be seen in Fig. 8.5 most methods show similar performance except `BDTB` and `BDTDCost`. In general the gradient boosting algorithm outperform the adaptive boost, but differences are small.

An example decision tree of the `BDTGBagShrink` method is shown in Fig. 8.6. Split decisions are given as well as the surviving number of events and the resulting signal purity defined as $\mathcal{P} = S/(S+B)$, where S and B are the number of signal and background events in the data sample, respectively. The importance of each input observable is shown in Table 8.5. Note that other methods do not differ much from these values. When using BDT the importance of a variable is defined by counting how often the variable is used to split decision tree nodes and by weighting each split occurrence by the separation gain-squared it has achieved and by the number of events in the node. In this case X_{max} is the most important variable.

Another way to study high-dimensional geometry and visualizing multivariate problems is the multidimensional system of parallel coordinates [259, 260]. Observables are represented by equally spaced vertical lines. A point in a n–dimensional space (representing n observables), is drawn as a polyline with vertices on the parallel axis. The position of the vertex on the i-th axis corresponds to the i-th coordinate of the point. The MC input data set for photon and proton primaries using the `BDTGBagShrink` method is shown in Fig. 8.7. Each event is drawn by a green polyline while a blue polyline indicates a misclassified event, i.e. a photon primary with a small response value and vice versa. As can be seen, misclassified proton primaries have a photon-like shape with large X_{max} and

Method	Option	Description

`MLPStandard`*	`NeuronType=tanh,`
	`VarTransform=N,`
	`NCycles=600,`
	`HiddenLayers=N+5,`
	`TestRate=5,`
	`TrainingMethod=BP`
		Standard options for the ANN using an MLP implementation. Options refer to the neuron activation type function `NeuronType`, the number of training cycles `NCycles`, the hidden layer architecture `HiddenLayers` ($N + 5$ indicates one hidden layer with $N + 5$ neurons, where N is the number of input variables), the number of epochs after an overtraining check is performed `TestRate` and the training method `TrainingMethod`. Additionally the observables are normalized to the interval $[-1, 1]$ via `VarTransform`.

`MLPBFGS`	`TrainingMethod=BFGS`	Here back propagation is performed by the Broyden-Fletcher-Goldfarb-Shannon (BFGS) method.
`MLPBFGSHL2`	`HiddenLayers=N+5,N,` `TrainingMethod=BFGS`	Here back propagation is performed by the BFGS method using two hidden layers with $N+5$ and N nodes.
`MLPNCy400`	`NCycles=400`	Reduced number of training cycles.
`MLPNCy800`	`NCycles=800`	Increased number of training cycles.
`MLPNHL-5`	`HiddenLayers=N-5`	Reduced number of nodes in the hidden layer.
`MLPNHLN+1-N`	`NCycles=800,` `HiddenLayers=N+5,N-5`	Using two hidden layers, the first with $N+5$ and the second with $N-5$ nodes.

Table 8.6: Summary of various artificial neural network settings used in this thesis. The `MLPStandard` method marked with a '*' indicate the default parameters for a particular neural network while the options of other methods just indicate deviations from the default settings. A brief description is also given. More details on individual steering parameters are given in [247].

small $E_{\mathrm{Tank}}/E_{\mathrm{FD}}$ values compared to the average distribution.

Although the `BDTGBagShrink` algorithm is not the top ranked method in the ROC diagram, the classifier is chosen to be the best performing boosted decision tree method. Using the gradient boost is typically less susceptible for overtraining. The robustness is also enhanced by a reduced learning rate of the algorithm while keeping the number of trees manageable. Another advantage is the use of a bagging-like resampling.

8.6.2 Artificial neural networks

Several differently configured artificial neural networks are tested for the separation between photon and proton primaries. The basic concept was introduced in Sec. 8.2 and the tested settings are listed in Table 8.6. An example of the `MLPBFGSHL2` response function is given in Fig. 8.8. The training and test sample are superimposed and show good agreement. Also illustrated in Fig. 8.8 is the corresponding network architecture. Individual connections are illuminated by arrows while the strength of the connections are

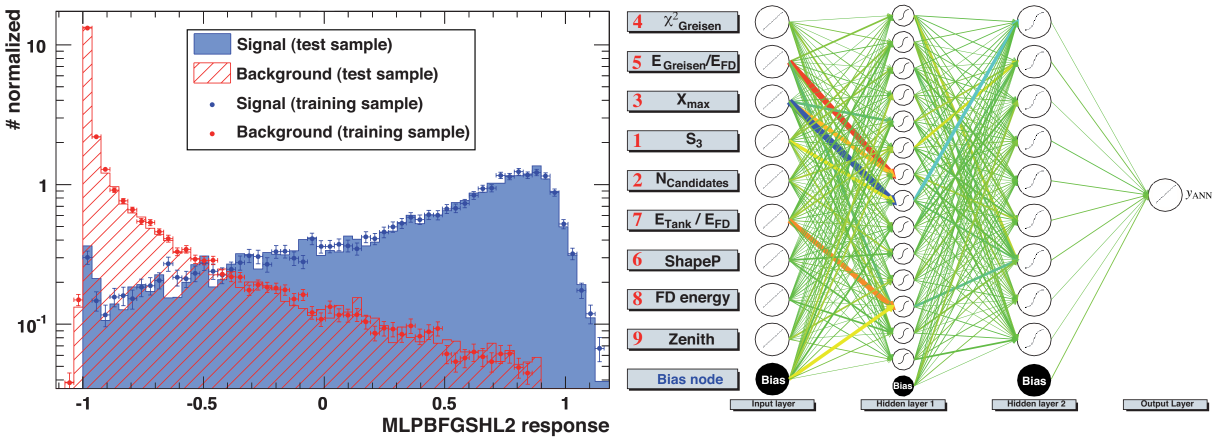


Figure 8.8: *Left:* Classifier response value of the `MLPBFGSHL2` method. “Signal” and “Background” always refers to primary photons and protons, respectively. The shaded area histograms correspond to the test sample while dots represent the training sample. A good agreement between training and test distribution indicates that overtraining is under control. *Right:* Corresponding network architecture of the `MLPBFGSHL2` method. Input observables are shown and the rank of the variable importance is indicated by a red number. Internode connections are shown by green arrows while stronger connections are drawn thicker with different colors.

highlighted by thicker lines and different colors. The variable importance of the input observable is also shown. It uses the sum of weights-squared of the connections between the variable's neuron in the input layer and the first hidden layer. The importance I_i of the input observable i is given by

$$I_i = \bar{x}_i^2 \sum_{j=1}^{n_h} \left(w_{ij}^{(1)}\right)^2 , \qquad i = 1, ..., n_{\rm var} \ . \tag{8.12}$$

Here $\bar{x}_i$ is the samples mean of input variable i, n_h the number of nodes in the first hidden layer and $n_{\rm var}$ the number of input observables.

The representation of the dataset in parallel coordinates applying the `MLPBFGSHL2` method is shown in Fig. 8.9. Each event is drawn as a green polyline. For photon primaries events with small $X_{\rm max}$ (i.e. proton-like) are highlighted in purple resulting in generally small output response values. This supports the assumption that general proton features are propagated correctly towards the classifier response value. This is also shown for proton primaries where misclassified events are highlighted in blue color.

The ROC diagram of all methods listed in Table 8.6 are shown in Fig. 8.10. The performance of all tested neural networks show similar behavior whereat the use of two hidden layers seem to increase the separation power. The method `MLPBFGSHL2` is chosen to be the best performing neural network. The BFGS algorithm has the advantage of using second derivatives with the effect that smaller number of iterations are necessary. The larger amount of computing time does carry no weight in this analysis.

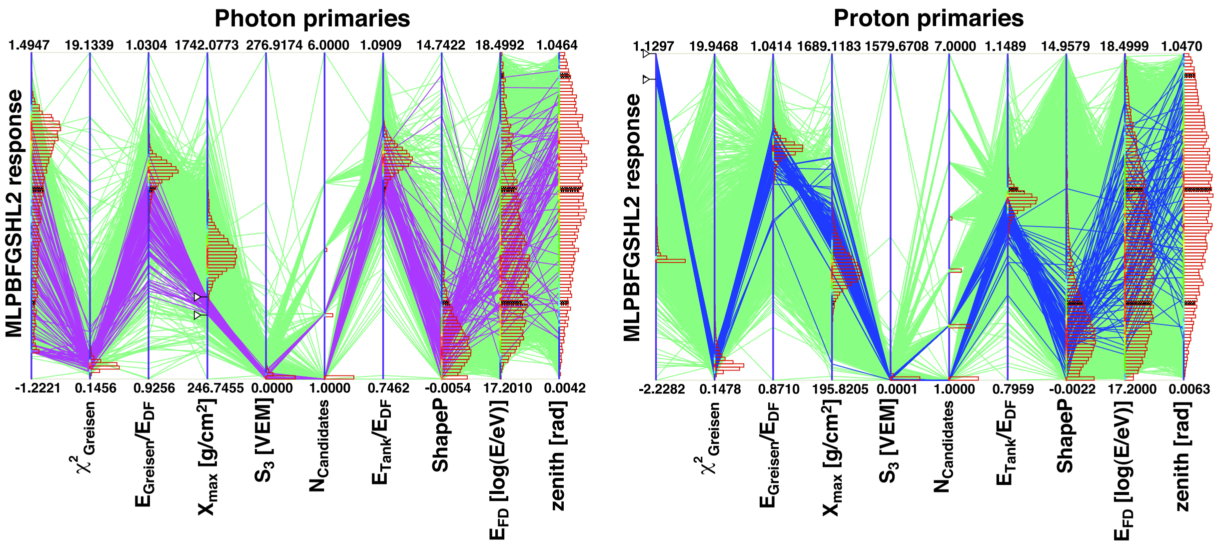


Figure 8.9: Parallel coordinates of the `MLPBFGSHL2` method for photon (*left*) and proton (*right*) primaries. Each event is drawn by a green polyline connecting parallel vertical axis representing all input observables and the classifier response. The position of the line at the vertical axis corresponds to the individual value of that event. The frequency scale of each observable is shown by a red histogram superimposed to the vertical axis. Blue polylines highlight events for proton primaries that are misclassified by the classifier response, i.e. with large classifier output value. For photon primaries, purple polylines indicate events that have low X_{max} in common. As can be seen most events are classified towards smaller `MLPBFGSHL2` response values, i.e. more proton-like behavior.

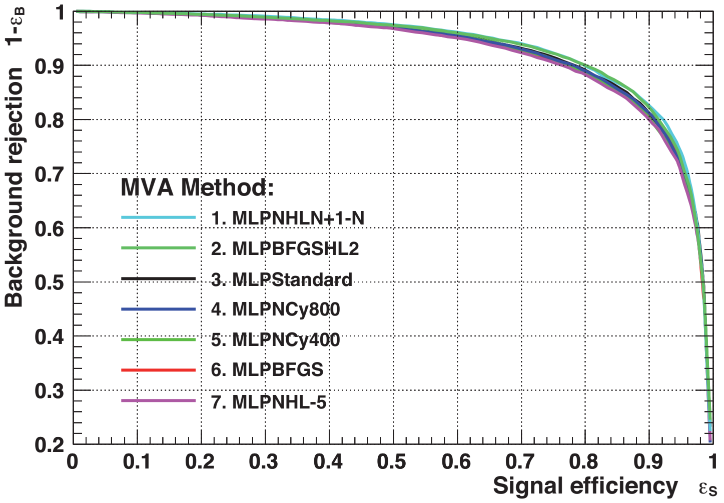


Figure 8.10: Receiver operating characteristic (ROC) diagram of various ANN methods described in Table 8.6. The methods are ranked by the integral of the ROC diagram showing the best method on top. All methods show similar performance.

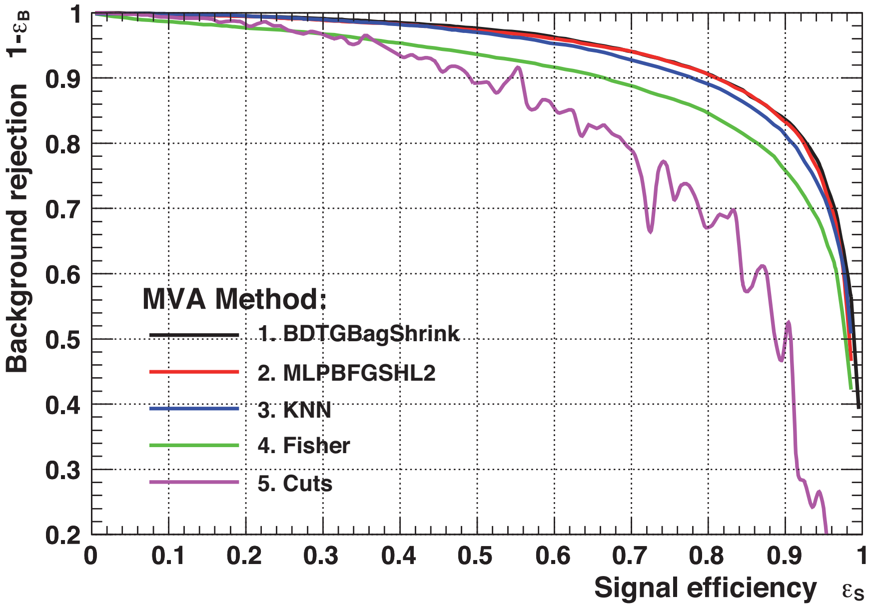
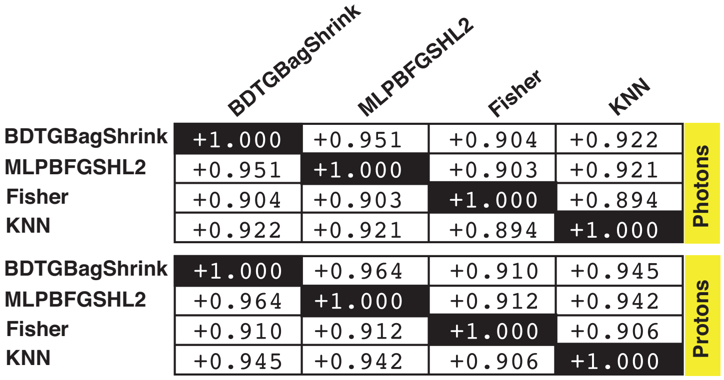


Figure 8.11: Receiver operating characteristic (ROC) diagram of various multivariate methods. The methods are ranked by the integral of the ROC diagram showing the best method on top. In addition to the already introduced ANN and BDT architectures, k-Nearest Neighbor (`KNN`) classification , Fisher discrimination (`Fisher`) and simple rectangular cuts (`Cuts`) are added. Note that the wiggles in the `Cuts` curve are a result of a rather simple MC approximation when finding the best set of cuts at a given signal efficiency. However, even when improving the algorithm the performance of ANN and BDT architectures can not be achieved.



	BDTGBagShrink	MLPBFGSHL2	Fisher	KNN	
BDTGBagShrink	+1.000	+0.951	+0.904	+0.922	Photons
MLPBFGSHL2	+0.951	+1.000	+0.903	+0.921	
Fisher	+0.904	+0.903	+1.000	+0.894	
KNN	+0.922	+0.921	+0.894	+1.000	
BDTGBagShrink	+1.000	+0.964	+0.910	+0.945	Protons
MLPBFGSHL2	+0.964	+1.000	+0.912	+0.942	
Fisher	+0.910	+0.912	+1.000	+0.906	
KNN	+0.945	+0.942	+0.906	+1.000	

Table 8.7: "Overlap" matrices of photon (top) and proton primaries (bottom). These matrices contain the fraction of events for which the multivariate method i and j have returned conform answers about "signal-likeness". Here signal-likeness is defined as if its classifier output exceeds the following value: `BDTGBagShrink`$\rightarrow -0.380$, `MLPBFGSHL2`$\rightarrow -0.351$, `Fisher`$\rightarrow +0.006$ and `KNN`$\rightarrow +0.347$, corresponding to a working point $\epsilon_S = 1 - \epsilon_B$.

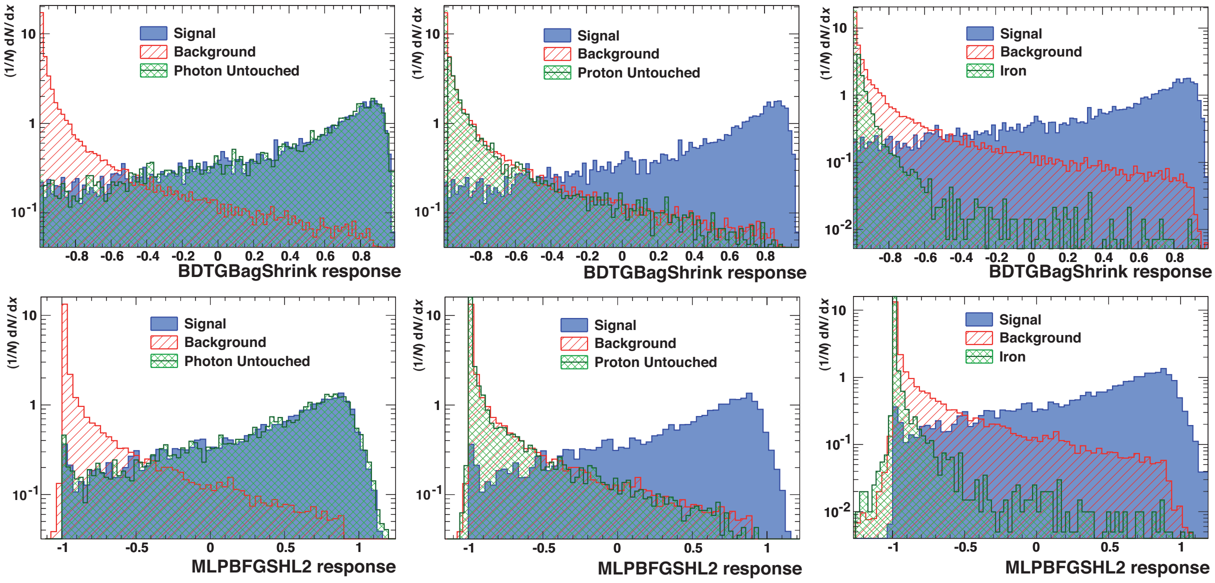


Figure 8.12: *Top row*: Comparison of the `BDTGBagShrink` output response of the test sample with an untouched data set of photon (left), proton (middle) and iron primaries (right). *Bottom row*: Output response of the test sample using the `MLPBFGSHL2` method compared with an untouched data set of photon (left), proton (middle) and iron primaries (right).

8.6.3 Searching for the best classifier

In this section different multivariate methods are compared to search for the best discriminating classifier response function. As already described in the previous section the BDT method `BDTGBagShrink` and neural network `MLPBFGSHL2` are selected to be compared. Furthermore a Fisher discrimination (cf. Sec. 8.3) is performed as well as simple rectangular cut optimization `Cuts` and a k-Nearest Neighbor (`KNN`) classification. Note that more details about these techniques can be found in [247]. The purpose here is just to compare the performance with ANN and BDT.

The corresponding ROC diagram is shown in Fig. 8.11. As can be seen, the BDT method `BDTGBagShrink` maximizes the integrated ROC diagram while `MLPBFGSHL2` shows a similar behavior. The other three methods have notably smaller integrated ROC values. To compare the fraction of events for which multivariate method i and j have returned conform answers about “signal-likeliness”, the “overlap” matrices for photon and proton primaries are illustrated in Table 8.7. Here “signal-likeliness” is given if the classifier response exceeds a certain cut value, namely $\epsilon_S = 1 - \epsilon_B$ (signal efficiency = background rejection). As can be seen both methods `BDTGBagShrink` and `MLPBFGSHL2` return conform answers about signal-likeliness in more than 95% above the working point.

An important condition for multivariate methods is that they are robust in case of unknown input events. General features of photon and proton induced EAS should be applied and no individual event characteristics. To this end the best performing BDT and ANN methods are tested with an untouched dataset which was not used during various

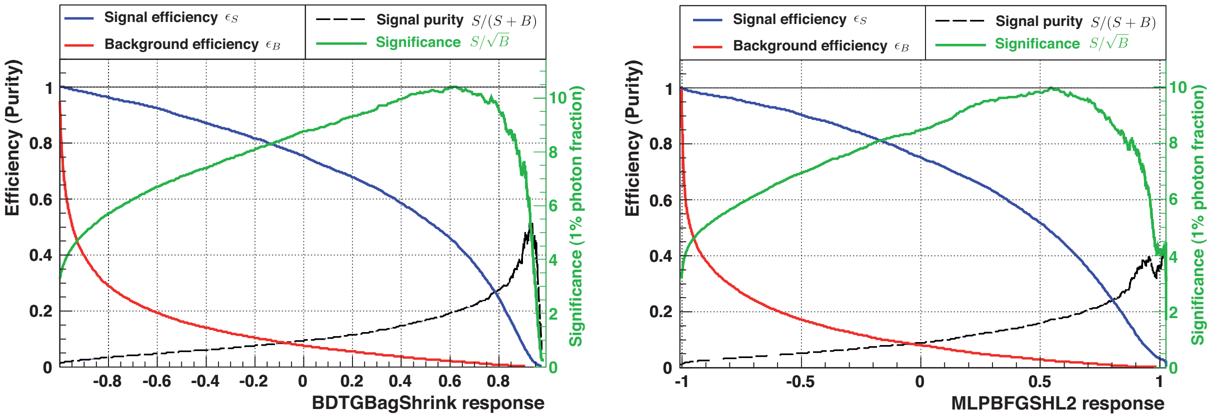


Figure 8.13: Cut efficiencies for the `BDTGBagShrink` (*left panel*) and `MLPBFGSHL2` method (*right panel*). The blue and red line indicates the photon (ϵ_S) and proton (ϵ_B) efficiency above a certain classifier response value, respectively. The dashed black line corresponds to the purity $\mathcal{P}$ of the signal. The significance of a cut at a certain output response value defined as $S/\sqrt{B}$ is illustrated as solid green line. Note that the axis scale depends on the assumed photon fraction and total number of events. Here 100000 proton and 1000 photon primaries are taken corresponding to a photon fraction of 1%.

training and testing sequences [258]. As part of the total simulated dataset, introduced in Sec. 7.1, 50000 proton and 25000 photon showers were simulated using the $\overline{\text{Off}}\underline{\text{line}}$ framework (see Sec. 7.1 and Sec. 7.2 for more details on steering parameters) and feed into the methods marked as unknown composition. The resulting classifier output is illustrated in Fig. 8.12. As can be seen, the untouched data sample fits very well with data of the test sample.
The general composition of cosmic rays at EeV energies is still unknown (cf. Sec. 2.3) and there are indications that a large fraction diverges from a pure proton composition towards heavier nuclei masses. For this purpose, the classifier response of 50000 simulated iron primaries compared to protons and photons is also shown in Fig. 8.12. As expected, the discrimination power between photon and iron is enhanced compared to proton primaries. Keeping a pure proton composition as background reflects therefore a conservative way for discrimination.

The cut efficiencies for the `BDTGBagShrink` and `MLPBFGSHL2` method are given in Fig. 8.13. Superimposed are the signal efficiency ϵ_S, background efficiency ϵ_B and the signal purity $\mathcal{P} = S/(S+B)$ at a certain classifier output value. Also given as a green solid line is the significance of a cut at a certain response value. The best working point is typically the maximized significance where the definition of significance may vary, dependent on the actual problem. Generally, when measuring a signal cross section $S/\sqrt{S+B}$ is a good choice while precision measurements require high purity $S/(S+B)$. However, for discovery of signal where $S \ll B$, a good choice for significance is $S/\sqrt{B}$ as shown in Fig. 8.13. Note that the assumed primary composition may affect the shape of the significance diagram. If defined for signal discovery as $S/\sqrt{B}$, however, the shape is independent of the photon fraction, i.e. the maximum significance is always at the same response value, but certainly the scale of the axis is changed. In Fig. 8.13 a photon fraction of 1% is assumed, more precise 100000 proton and 1000 photon primaries. Below is a roundup of the best working point for the `BDTGBagShrink` and `MLPBFGSHL2` method.

MVA method	**Optimal cut** (max. significance)	**# Sig** (after cut)	**# Bkg** (after cut)	ϵ_S	ϵ_B	$S/\sqrt{B}$
`BDTGBagShrink`	0.60	461	1957	0.461	0.010	10.41
`MLPBFGSHL2`	0.55	483	2347	0.483	0.023	9.97

8.7 Summary

In this chapter a multivariate analysis for photon/hadron discrimination was performed using an extensive MC shower library. Special attention was turned to boosted decision trees and artificial neural networks. As an advanced classifier they are able to take into account high-dimensional correlations using different separation philosophies. The best performing boosted decision tree (`BDTGBagShrink`) was found to be using the gradient boost algorithm in combination with a small shrinkage parameter (to reduce the learning rate) making the method less susceptible for overtraining. It turned out that the introduction of a bagging-like resampling procedure using random subsamples of the training events for growing the trees can further enhance the separation power.

When using artificial neural networks the back propagation by the Broyden-Fletcher-Goldfarb-Shannon algorithm indicated better performance than the standard back propagation. The introduction of two hidden layers could further increase the separation power and is realized in the `MLPBFGSHL2` method.

Differences between neural networks and boosted decision trees are small. Using the optimal working point for selection, both methods reduce the proton fraction by about $\sim 98\%$ while keeping $\sim 50\%$ of photon primaries. It was found that boosted decision trees are more robust with a slightly better performance. The ability to view individual trees makes the method more transparent and manageable. It is decided to continue with the `BDTGBagShrink` method as the main algorithm to apply on real unknown data from the Pierre Auger Observatory addressed in Sec. 9.2 of the next chapter.

Chapter 9

Search for EeV Photons

The search for ultra-high energy photons is a key topic of this thesis. This chapter describes the search using hybrid data observed at the Pierre Auger Observatory. Sec. 9.1 describes the preparation of the full hybrid data set. This data is used in a multivariate analysis, introduced in the previous chapter, to create a new, smaller dataset containing air showers similar to what is expected from photon primaries (Sec. 9.2). The effect on the reconstructed primary energy is discussed in Sec. 9.3. Sec. 9.4 and Sec. 9.5 are introducing important steps to produce sky maps when taking individual event uncertainties into account. Signal maps are shown in Sec. 9.6 and a way to calculate the expected background expectation for an isotropic sky is explained in Sec. 9.7. Sec. 9.8 up to Sec. 9.10 highlight the search for EeV photons using two-point correlation functions and Li-Ma statistics. Special attention to the Galactic Center region and Centaurus A is given in Sec. 9.11. Celestial limits on the photon fraction for sources are calculated in Sec. 9.12. Finally, a brief summary is given in Sec. 9.13.

THe previous chapters paved the way for searching for primary photons in the EeV energy range. A MC simulation study has been performed to find differences between photon and hadron induced extensive air showers. It was shown that FD as well as SD observables can be used for discrimination, each accompanied by its own characteristics and features. In a multivariate analysis all selected observables were combined to gain one powerful output response function. The trained multivariate method is now applied to real data collected at the Pierre Auger Observatory in the Argentinean Pampa Amarilla to search for EeV photons and to place directional limits on the photon fraction.

9.1 Real data reconstruction

The search for EeV photons has been performed with hybrid data of the Pierre Auger Observatory collected during the period of five years, to be specific 12/2004 – 12/2009. The detector response of the Pierre Auger Observatory surface and fluorescence detector has been simulated using the $\overline{\text{Off}}\underline{\text{line}}$ framework version v2r7p0-svn_trunk which was already introduced in Sec. 4.3. All important simulated and reconstructed parameters are stored using the `ADST` data structure. The reconstruction settings correspond to the official Auger event reconstruction [261].

The event selection is slightly different to official procedures and illustrated in Table 8.3. Real hybrid event selection usually starts with basic criteria concerning the validity of the analyzed data. Here, the main criterion is the selection of time periods with valid calibration constants. For fluorescence detector data only time periods after the first absolute calibration campaign are taken into account, namely the starting date

- Los Morados: December 1st 2004
- Los Morados: June 2nd 2005
- Loma Amarilla: May 1st 2007
- Coihueco: December 1st 2004

In addition, some periods affected by GPS-clock glitches and periods of unstable baselines have been excluded. Since this study utilizes not only FD data, also SD bad periods have to be excluded from the analysis. A summary of SD bad periods is provided by the Pierre Auger trigger and acceptance group [262]. As a suggestion of the acceptance working group, data in the time period between April 16 2009 and November 15 2009 has also been excluded since there were major communication problems within the SD [263].

9.2 Hadron reduced data set

After applying basic selection and sanity cuts (cf. Table 8.2 and Table 8.3) and excluding bad FD and SD periods the final data set consists of 240924 events between 17.2 and 18.5 $\log(E/\mathrm{eV})$. The resulting energy and $X_{\max}$ distribution is shown in Fig. 9.1. Using the trained boosted decision tree discussed in Sec. 8.6.3, the classifier output response compared to proton and iron primaries is shown in Fig. 9.2. Measured hybrid events are illustrated by the blue histogram. Obviously, the histogram is neither compatible with proton nor with pure iron primaries. The major part of observed cosmic rays seems to be in between suggesting a mixed composition in the energy range 17.2 – 18.5 $\log(E/\mathrm{eV})$. This is also supported by Fig. 2.5 where the average $X_{\max}$ as a function of primary energy is compared to air shower simulations.

In Fig. 9.3 the shape of a pure proton output response is compared to the response of samples with various photon fractions. At large photon fractions, e.g. 50% (green),

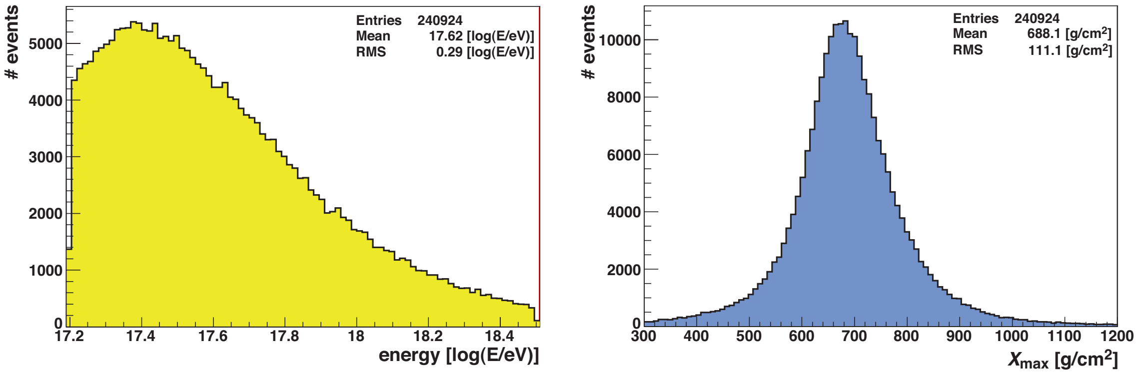


Figure 9.1: Energy (*left panel*) and X_{max} (*right panel*) distribution of real hybrid events after sanity cuts and bad period rejection.

the response shape is certainly significantly affected, but fractions $\lesssim 1\%$ are already compatible to a pure proton composition. Compared to the actual hybrid data shown in Fig. 9.2 no clear bump at large response values is visible suggesting a rather small photon fraction.

In contrast to protons, or more general to charged particles, photons are not deflected by magnetic fields. If a putative source would produce a certain fraction of photons or generate secondary photons in the vicinity, the arrival direction of UHE photons should point back to the location of the source resulting in an anisotropic arrival direction of photons[1]. Since the expected photon fraction is very small, such an anisotropy can hardly be seen in a dataset dominated by hadrons. For this reason a new dataset is created, henceforth called **hadron reduced data set**, in which the hadron fraction is significantly reduced while keeping most of the photons. This is accomplished by taking only events with `BDT` response > 0.60, cf. Sec. 8.6.3 and Fig. 8.13, i.e. the value where the maximum significance $S/\sqrt{B}$ is reached. Applying this cut to the hybrid data set with unknown composition the hadron reduced data set consists of 1950 events (4715 events are expected in case of a pure proton composition).

9.3 Energy dependence of the BDT response value

Fig. 9.4 and Fig. 9.5 illustrates how the BDT output response value is affected by an increase (decrease) of the reconstructed hybrid FD energy of 20%. Note that just the energy of the real data set is changed while keeping the MVA analysis untouched. This corresponds to the case that the real data is not characterized by MC simulations by an energy shift of ±20%. If this energy drift is also included in MC simulations, the MVA methods is trained for this new situation and differences of the response value would be only marginal. However, as shown in Table 8.1 several observables are influenced by the

---

[1] Here it is assumed that the majority of photons is produced in the vicinity of the source and not by propagation effects such as the decay of neutral pions of the GZK-effect.

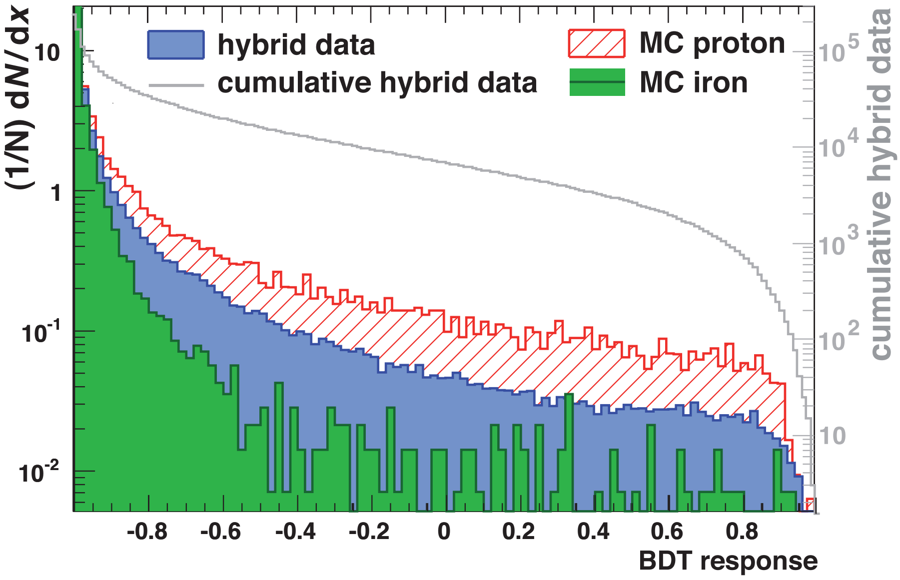


Figure 9.2: BDT classifier output response for real hybrid data (blue), MC proton primaries (red) and MC iron primaries (green). As can be seen the actual composition seems to be in between proton and iron suggesting a mixed composition in the energy range 17.2 – 18.5 log(E/eV). No dominant photon bump is visible at large response values. Additionally a cumulative histogram is shown in grey representing hybrid data.

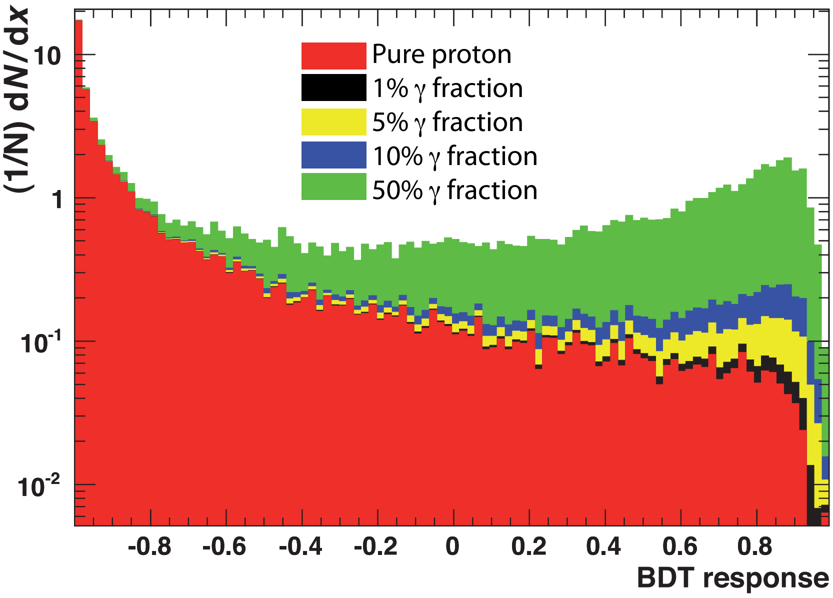


Figure 9.3: BDT output response value assuming various photon fractions. The red histogram indicates a pure proton composition. Subsequent histograms illustrate a photon fraction of 1% (black), 5% (yellow), 10% (blue) and 50% (green). By examining the true output response some photon fractions can already be constrained.

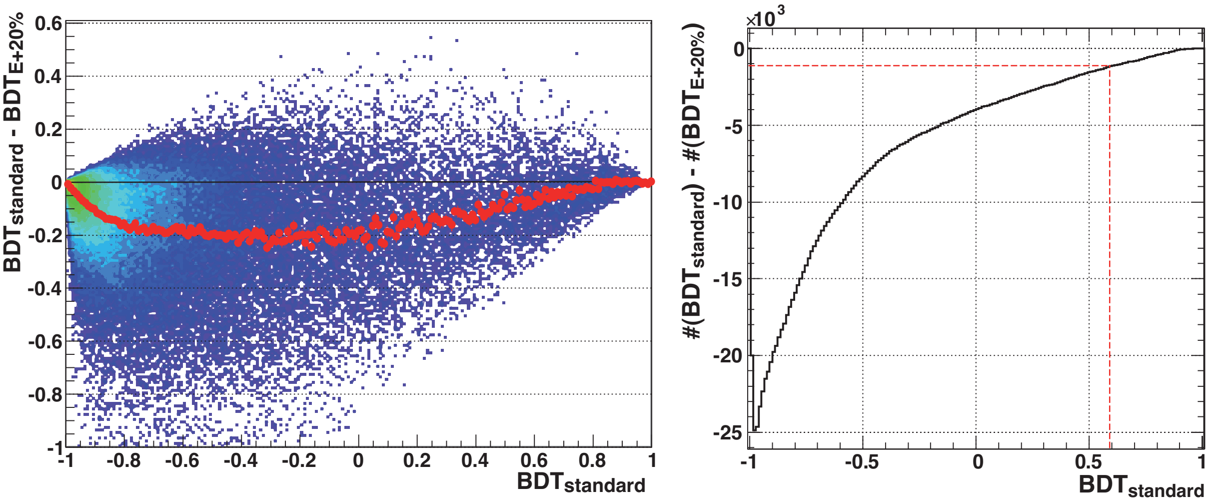


Figure 9.4: Effect of increasing (underestimating) the reconstructed hybrid FD energy by 20%. *Left panel*: Difference of the BDT output response function. $\mathrm{BDT}_{\mathrm{standard}}$ and $\mathrm{BDT}_{E+20\%}$ corresponds to the untouched and modified output response, respectively. The profile of the distribution is shown as red dots. In general an increase of energy relative to the simulated energy corresponds to an increase (more photon-like) of the response value. *Right panel*: Difference of the number of events above a certain BDT cut value when increasing the energy by 20%. The red dashed line denotes the applied cut value. The hadron reduced data set would increase by about 1000 events when increasing the energy.

FD energy and are accordingly taken into account when changing the energy artificially. Increasing the energy by 20% relative to the simulated energy results in a general increase of the BDT output response value (Fig. 9.4–*left*), or in other words a more photon-like behavior. This is also supported by Fig. 7.13 and Fig. 7.6 where an increase of E_{FD} corresponds to a more photon-like distribution. The impact on the number of events of the hadron reduced data set for a given cut is shown in Fig. 9.4–*right*. Applying the previously introduced cut, the hadron reduced data set is increased by ~1000 events when increasing the energy artificially. Similarly to Fig. 9.4 the energy is artificially decreased by 20% in Fig. 9.5 with the reciprocal effect. The hadron reduced data set would be reduced by ~500 events when decreasing the energy by 20%. Some consequences on significant maps and upper limits on the photon fraction are briefly discussed in Sec. 9.10 and 9.12.

9.4 Probability density distribution of events

After the geometrical reconstruction a cosmic ray detector assigns a direction of origin to each detected air shower. The measurement of arrival directions of extensive air showers is, however, always accompanied by some angular uncertainty. It is therefore advisable that the arrival direction is represented not by a single point in the sky, but by a probability distribution which has its maximum value at the most likely direction of origin. It can be

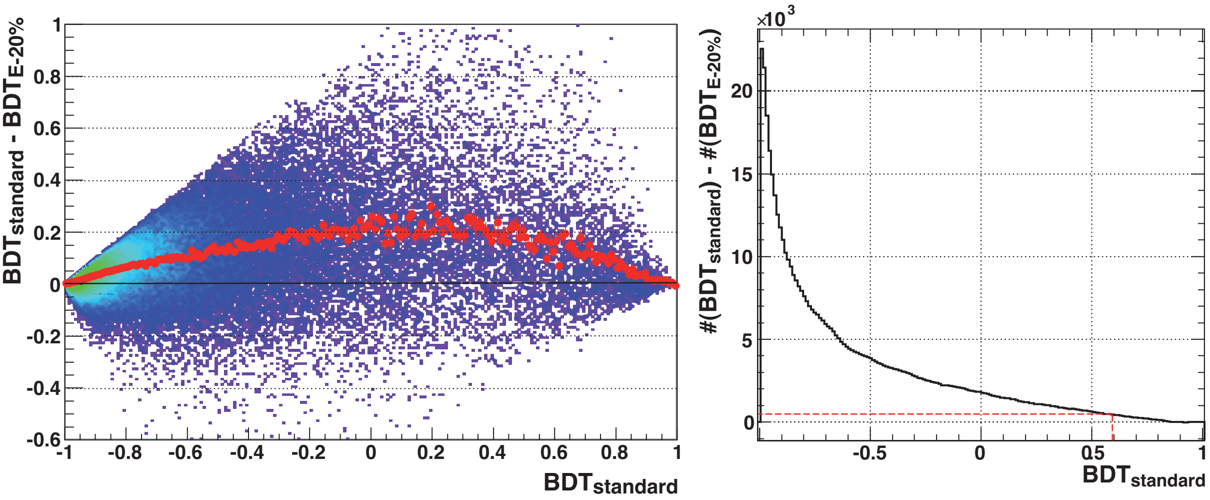


Figure 9.5: Effect of decreasing (overestimating) the reconstructed hybrid FD energy by 20%. *Left panel*: Difference of the BDT output response function. BDT$_{\text{standard}}$ and BDT$_{E-20\%}$ corresponds to the untouched and modified output response, respectively. The profile of the distribution is shown as red dots. In general a decrease of energy relative to the simulated energy corresponds to an decrease (less photon-like) of the response value. *Right panel*: Difference of the number of events above a certain BDT cut value when decreasing the energy by 20%. The red dashed line denotes the applied cut value. The hadron reduced data set would decrease by about 500 events when decreasing the energy.

regarded as a density function whose integral over the hole sky is one shower. The sum of all density functions, each representing an individual air shower, will be called the actual density[2].

The challenge is to calculate the actual density starting from basic geometric considerations and propagating the errors into sky maps. In this thesis it is performed in several steps. As a result of the hybrid reconstruction, the most natural geometrical uncertainties are given by the uncertainties of the SDP $\Delta\vartheta_{\text{SDP}}$ and $\Delta\phi_{\text{SDP}}$ (and its correlation), and the inclination angle within the SDP $\Delta\chi_0$, cf. Fig. 4.11 (a). By calculating the Jacobian-matrix, these uncertainties are propagated into uncertainties of the local θ and ϕ direction as well as their correlation coefficient $\rho_{\theta,\phi} \equiv \rho$. The idea is now to determine a probability density distribution using θ, ϕ and $\rho_{\theta,\phi}$ that has the form of a two-dimensional Gaussian. The general probability density can be stated as (cf. [264])

$$\begin{aligned} f(\phi,\theta)\mathrm{d}\phi\ \mathrm{d}\theta \quad &= \quad \frac{1}{2\pi\sigma_\phi\sigma_\theta\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\right. \\ &\cdot \left.\left[\left(\frac{\phi-\zeta}{\sigma_\phi}\right)^2 - 2\rho\left(\frac{\phi-\zeta}{\sigma_\phi}\right)\left(\frac{\theta-\eta}{\sigma_\theta}\right) + \left(\frac{\theta-\eta}{\sigma_\theta}\right)^2\right]\right\} \mathrm{d}\phi\ \mathrm{d}\theta\ . \quad (9.1) \end{aligned}$$

Here σ_ϕ and σ_θ are the corresponding standard deviations and ρ the correlation coefficient

---

[2] Its integral over the hole sky is the total number of detected air showers.

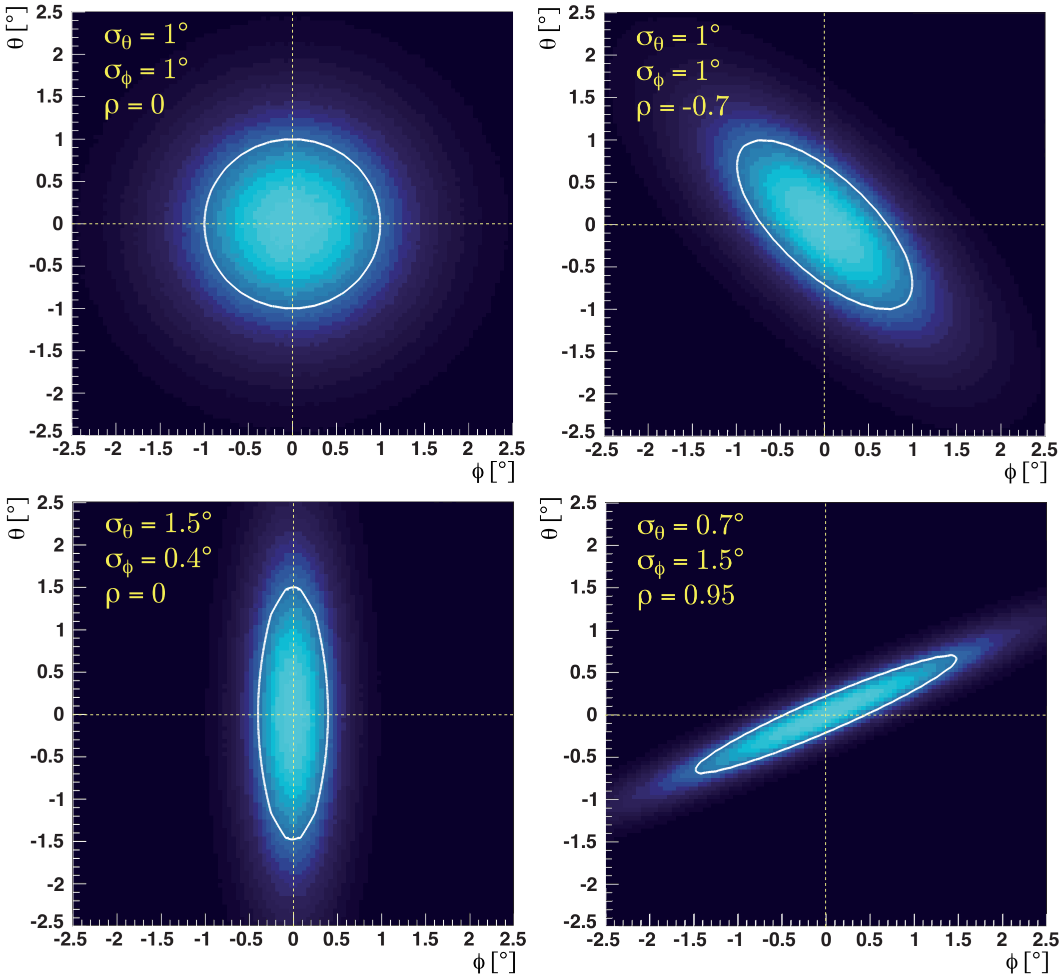


Figure 9.6: Illustration of individual event shapes centered at the origin. Color coded is the event probability as a function of θ and ϕ. The applied standard deviations σ_ϕ and σ_θ as well as their correlation coefficient ρ are given. The white ellipse denotes the $1-\sigma$ contour defined as the region where the probability density $f(\phi, \theta)$ is dropped to $1/\sqrt{e}$ of its maximum value (Eqn. (9.2)).

between ϕ and θ. The expectation values of ϕ and θ, i.e. the reconstructed directions, are expressed as $\langle\phi\rangle = \zeta$ and $\langle\theta\rangle = \eta$. By generating random directions according to the probability density given in Eqn. (9.1) [265] and normalizing the integral of the hole sky to one yields the expected density distribution of a single event in local coordinates. Some examples of event shapes are given in Fig. 9.6. Without loss of generality all sample events are centered around the origin, i.e. $\zeta = \eta = 0$. The standard deviations in θ and ϕ are given as well as their correlation coefficient. Also shown in Fig. 9.6 as a white ellipse is the $1-\sigma$ contour defined as the region where the probability density $f(\phi, \theta)$ is dropped to $1/\sqrt{e}$ of its maximum value and given by

$$\frac{\phi^2}{\sigma_\phi^2} - \frac{2\rho \cdot \phi \cdot \theta}{\sigma_\phi \sigma_\theta} + \frac{\theta^2}{\sigma_\theta^2} = 1 - \rho^2 \quad . \tag{9.2}$$

The ellipse is located in the rectangle $|\phi| \leq \sigma_\phi$ and $|\theta| \leq \sigma_\theta$. The probability that a combination of ϕ and θ is within the $1 - \sigma$ ellipse is $1 - 1/\sqrt{e} \approx 39\%$.

The representation of the probability density in local coordinates is then transformed into e.g. a galactic coordinate system. The conversion utilizes the Julian day of a specific event as timing information and the individual geographic longitude and latitude of the core impact point. Due to precession the right ascension and declination of stars are constantly changing. Therefore, the J2000 equinox/epoch is used as a celestial reference frame.

9.5 Towards sky maps

The representation of event densities on a sky map and subsequent anisotropy analysis requires a sophisticated pixelization of the sphere. The mathematical structure should support a suitable discretization of functions on a sphere at sufficiently high resolution, and facilitate fast and accurate statistical and astrophysical analysis of massive full-sky data sets. This is realized by the HEALPix software framework [266, 267]. It is an acronym for *H*ierarchical *E*qual *A*rea iso*L*atitude *Pix*elization of a sphere. The requirements are satisfied by the software because it possesses the following three essential properties:

1. The sphere is hierarchically tessellated into curvilinear quadrilaterals, cf. Fig. 9.7.

2. The area of all pixels at a given resolution are identical.

3. Pixels are distributed along lines of constant latitude.

The resolution of the grid is steered by the parameter N_{side} which defines the number of devisions along the side of a base-resolution pixel that is needed to reach a desired high-resolution partition. All pixel centers are located on $N_{\mathrm{ring}} = 4 \cdot N_{\mathrm{side}} - 1$ rings of constant latitude. The total number of pixels is $N_{\mathrm{pix}} = 12 \cdot N_{\mathrm{side}}^2$. Unless stated otherwise $N_{\mathrm{side}} = 128$ is used resulting in an average pixel radius of $\sim 0.26°$.

For the actual analysis a modified version of the *Coverage & Anisotropy Toolkit* [268] is used. This toolkit has been developed for the Pierre Auger Collaboration in order to perform small and large scale anisotropy analysis on the sky. In addition to the implementation of the HEALPix and `ROOT` framework it comprises of a library of `C` and Fortran routines for reading and writing `FITS` files (CFITSIO [269]) and basic space and time coordinate transformations.

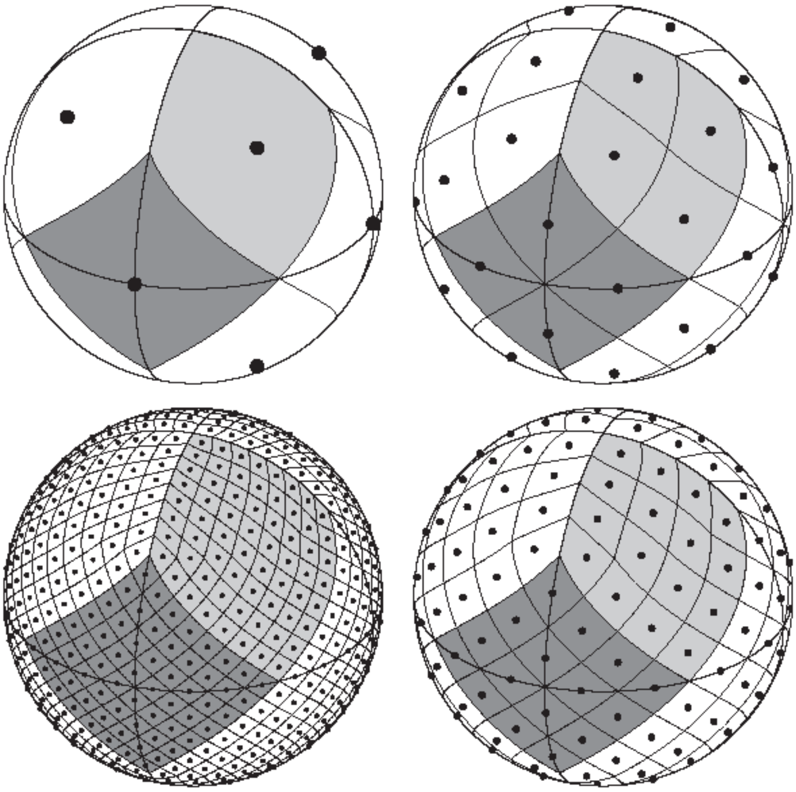


Figure 9.7: Illustration of the HEALPix partition of the sphere [266]. To express the octahedral symmetry equator and meridians are over-plotted. Starting from the upper left panel moving clockwise, the grid is subdivided with the N_{side} parameter equal to $N_{\text{side}} = 1, 2, 4, 8$. Within each panel the area of each pixel are identical. The light-gray shaded area denotes one of the eight (four north, and four south) polar base-resolution pixels. Dark-grey shading shows one of the four identical equatorial base-resolution pixels. Pixel centers are marked by black dots.

9.6 Signal density map

If there are any EeV photons in the dataset, the photon fraction should be enhanced in the hadron reduced data set. A directional correlation of arrival directions to galactic or extra-galactic objects, regions or just an abnormal clumping of events could give rise that this is caused by a certain fraction of photons since they are not deviated by magnetic fields. For this reason a sky map of the hadron reduced data set, introduced in Sec. 9.2, is shown in Fig. 9.8 using galactic coordinates. The color scale corresponds to the number of events per bin (cf. Sec. 9.5) in the HEALPix map.

Given Fig. 9.8 it is hard to say if there are significant regions of over-densities. Besides the necessity of calculating the background expectation discussed in Sec. 9.7, the flux from a candidate source has to be analyzed. There are mainly two alternatives for the analysis which are briefly introduced. More information can be found in Sec. 9.9.

1. *"Top-hat" counting*: The idea is to count the number of events within a prescribed

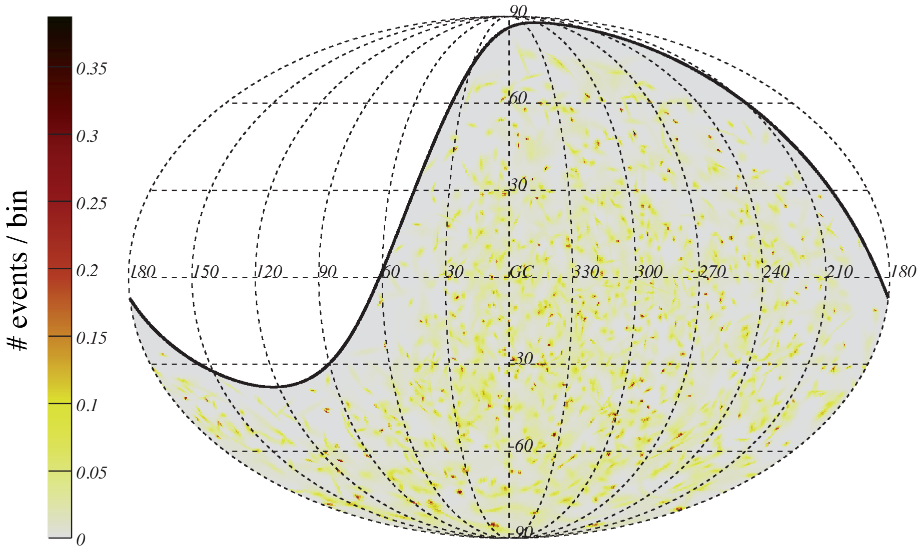


Figure 9.8: Sky map of events from the hadron reduced data set using the BDT algorithm. Shown are event densities in galactic coordinates using the Mollweide-projection (see e.g. [270]). The color scale refers to the number of events per bin. The declination limit of 25° is illustrated corresponding to a local zenith of 60°.

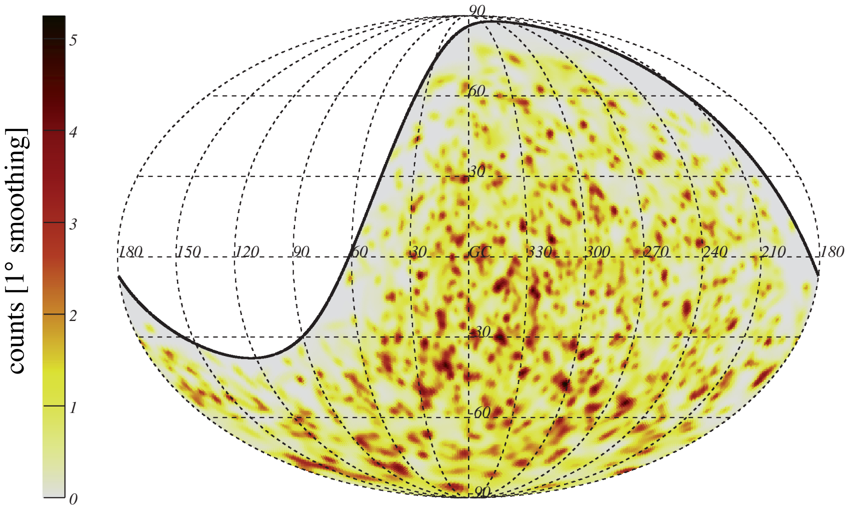


Figure 9.9: Same as Fig. 9.8 but now using a Gaussian weight with $\sigma = 1°$ up to an maximum angular distance of 25°.

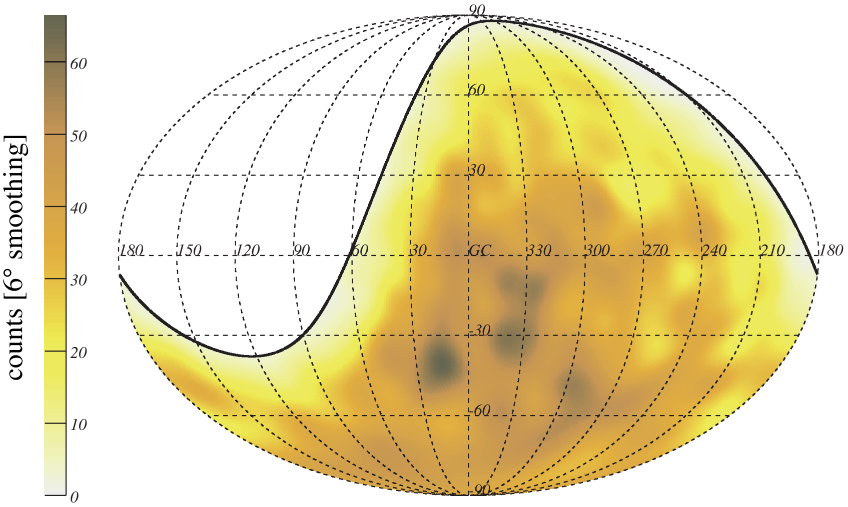


Figure 9.10: Same as Fig. 9.8 but now using a Gaussian weight with $\sigma = 6°$ up to an maximum angular distance of 25°.

radius of the source position relative to the number expected within that radius from an isotropic background. Assuming a Gaussian point spread function characterized by the width σ, the signal-to-noise ratio for a point source is optimized for a top-hat radius given by $r = 1.59 \cdot \sigma$ (see e.g. [271]).

2. *Weighted counting*: The expected flux from a point source is expected to produce a Gaussian distribution of arrival directions, centered on the source location, due to random intergalactic magnetic fields. When testing a certain direction for a point source, it is therefore sensible to give more weight to arrival directions that are near the center of the distribution and little weight to arrival directions that are far from it. Assuming a Gaussian distribution with width σ the probability distribution for offset θ_x is $P(\theta_x) = 1/(\sqrt{2\pi}\sigma) \cdot \exp\left(-\theta_x^2/(2\sigma^2)\right)$. In two dimensions the (space angle) offset θ becomes $\theta = \sqrt{\theta_x^2 + \theta_y^2}$ with the probability distribution

$$P(\theta) = P(\theta_x) \cdot P(\theta_y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{\theta^2}{2\sigma^2}\right) . \tag{9.3}$$

A clever definition of an appropriate weighting function w is [271, 272]

$$w = 4\pi\sigma^2 \cdot P(\theta) = 2 \cdot \exp\left(-\frac{\theta^2}{2\sigma^2}\right) , \tag{9.4}$$

where the term $4\pi\sigma^2$ denotes the solid angle over which the signal events arrive from a particular source. This definition of the weighting function w implicates several characteristics. On the one hand the expected summed weight from the background

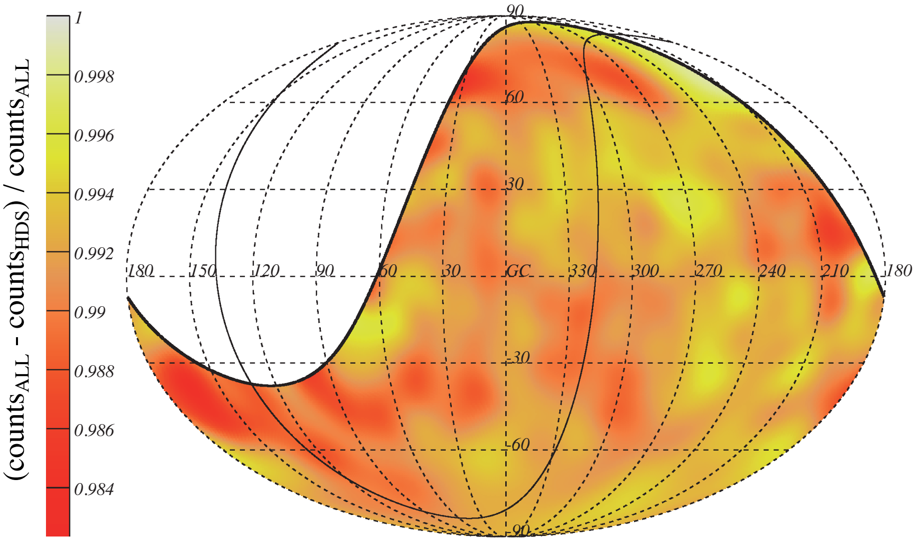


Figure 9.11: Directional selection efficiency of the hadron reduced data set. Shown is the ratio of two sky maps namely all events apart from the hadron reduced data set and all events using a 6° Gaussian weight. The field of view as well as the supergalactic plane are indicated by a black solid line.

density ρ is $4\pi\sigma^2\rho$ and the RMS fluctuation in that weighted background integral is its square root, provided a sufficient number of events ($\gtrsim 10$). On the other hand n_0 events from the source direction smeared out by the point spread function given by $P(\theta)$ will, in first approximation, produce a combined Gaussian weight equal to n_0.

In this thesis the weighted counting is used. Although the two methods are rather similar, the Gaussian method may be slightly more accurate since it emphasizes the central region. Fig. 9.9 illustrates a sky map of the hadron reduced data set applying a Gaussian weight with $\sigma = 1°$. Additionally to search for more extended objects a weighted map with $\sigma = 6°$ is shown in Fig. 9.10. In all cases the weighting is applied up to a maximal solid angle distance of $\theta_{\max} = 25°$. However, the significance of over-densities can not be defined until the background expectation is considered. This is discussed in the next section.

Fig. 9.11 indicates the directional selection efficiency of the hadron reduced data set. Here two sky maps are created, one using all events except the hadron reduced data set and the other one using all events. A 6° Gaussian weighting is applied and the ratio is formed. Note that larger fluctuations are expected at the edge of the field of view due to a decrease of statistics.

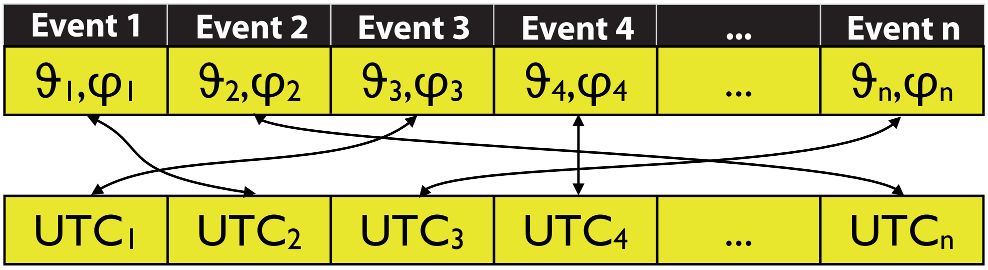


Figure 9.12: Illustration of the scrambling technique. Each event i carries information about local zenith ϑ_i, local azimuth φ_i and impact time on ground UTC_i. By randomly allocating each event a new impact time from the other actual UTC times a new right ascension is constructed while keeping the declination constant.

9.7 Calculation of the expected background density

In the previous sections a density map of signal events has been derived. To search for point sources or interesting regions, an essential condition is to compare the actual density with the density the detector is expected to have recorded if the particle intensity is isotropic. The simulation of the expected background for an isotropic sky is a dedicated task since all detector efficiencies and aperture features have to be taken into account. Various techniques have been proposed in the past to calculate the background expectation. Whatever scheme is adopted to calculate the expected density, it should provide a means for evaluating the probability that the particle intensity fluctuates such as to yield a density greater than or equal to the actual density.

The method used in this thesis is the so-called *shuffling* or *scrambling* method, first introduced by P. Sommers *et al.* at the end of the 80s [273] which is not controlled by complex detector simulations. For a better understanding it may be helpful to imagine that the detector's history of operations could be repeated thousands of times. In that case each repetition would include identical sidereal run times and identical histories of detector efficiency. Combining all data sets would result in a distribution of values for the density at each point in the sky. The mean and width of that distribution is therefore an accurate value of the expected density with its statistical fluctuations, dependent on the true particle density and on the detector's unique history of operation. The idea is that, under the assumption of an *isotropic flux*, such an ensemble data set can be constructed artificially from the measured data set. To clarify the idea, consider a small fixed region in the sky relative to the detector given by a zenith and azimuth range in local coordinates. As the Earth rotates through different sidereal times the region passes different right ascensions while keeping the declination band constant. An isotropic particle flux would present a time-independent flux to that detector direction. Only the detector's history of operations would produce a time-dependent variability.

The procedure of constructing a simulation data set and later the expected background density is as follows and also illustrated in Fig. 9.12. One uses the local zenith and

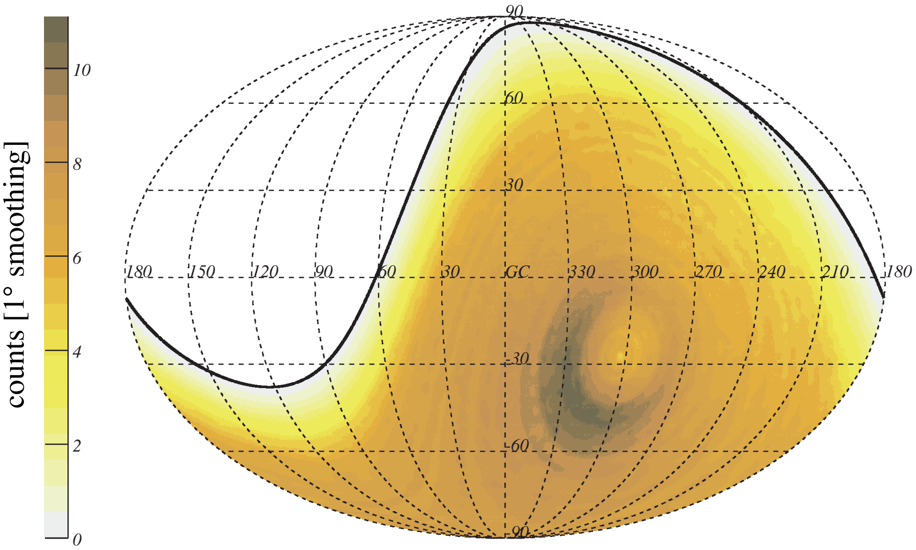


Figure 9.13: Background expectation of an enlarged (cf. Sec. 9.10) hadron reduced data set using a Gaussian weight with $\sigma = 1°$ up to an maximum angular distance of 25° using galactic coordinates. As a consequence of the scrambling technique, a band structure following constant declination is visible. Also apparent is the right ascension dependency of the expected density for fixed declination bands, cf. Fig. 9.14. This can be explained by a summer/winter effect.

azimuth but its UTC time of detection is selected randomly[3] from the other actual shower UTC times. Since the new assigned UTC time was taken from a measured event it is assured that the detector was operational and that the event could have been triggered at that particular time. In case of fluorescence detector data there are some more things to consider when using the scrambling technique. As each detector eye and even each telescope bay has different azimuthal trigger probabilities there may be the situation that a certain telescope triggered an event while a different telescope (from which the UTC time is exchanged) was not sensitive to that particular event. To avoid that ambiguity all events are binned telescope-wise (24 bins) before scrambling the data, resulting that only events from the same telescope are exchanged. Once the new simulation data set is created, an ensemble of typically a couple of thousand data sets is created (here ~5000) yielding a distribution of densities for each point of the sky and the mean value of a distribution is the expected density at that sky location[4].

The background expectation of an enlarged (cf. Sec. 9.10) hadron reduced data set is shown in Fig. 9.13 in galactic coordinates. Similarly to the signal sky map, a Gaussian weight with $\sigma = 1°$ is applied. As can be seen, arrival directions are smoothed but several

---

[3] UTC times from the original event can also be allocated and keeping the right ascension and declination

[4] Based on the hypothesis that the detector is responding to an isotropic flux

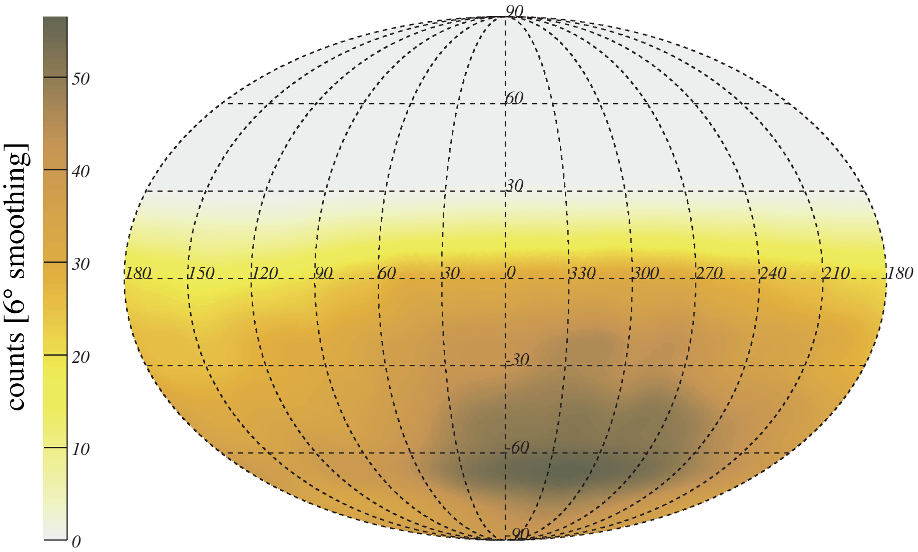


Figure 9.14: Background expectation for the hadron reduced data set using a Gaussian weight with $\sigma = 6^\circ$ up to an maximum angular distance of 25° displayed in equatorial coordinates. As a consequence of the scrambling technique, a band structure following constant declination (horizontal layers) is visible. Also apparent is right ascension dependency of the expected density for fixed declination bands. This can be explained by a summer/winter effect.

structures are still visible. A band structure along lines of constant declination is a result of the scrambling technique since events are allocated a new right ascension while the declination remains constant. It is also shown in e.g. Fig. 9.13 that the density varies as a function of right ascension while keeping the declination constant. This can be explained by a summer/winter effect. The duty cycle of FD observations is limited between dusk and dawn. This time period is maximized in the austral winter where certain parts of the sky are in the field of view of the detector. These parts are more exposed than regions only visible in the austral summer. Fig. 9.14 illustrates the background expectation of the hadron reduced data set with a Gaussian weighting of $\sigma = 6^\circ$ using equatorial coordinates. The aforementioned structures of constant declination can be seen now as horizontal lines since the sky map is represented in equatorial coordinates.

All these individual structures indicate that the calculation of the background expectation, using the scrambling technique, can not be applied to an arbitrary dataset and is restricted to events used for calculating the background expectation which is here the (enlarged) hadron reduced data set.

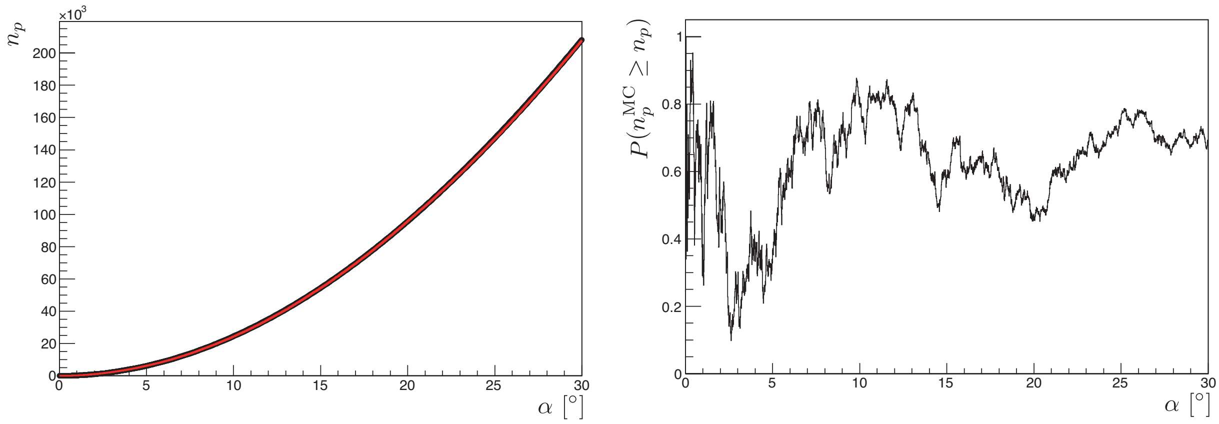


Figure 9.15: *Left panel*: Number of pairs n_p within a certain angular separation α for the hadron reduced data set. The red line indicates the result of the observed data set while the black line denotes the averaged simulation sample. No significant deviations from the background expectation are visible. *Right panel*: Fraction P of isotropic simulations with larger or equal number of pairs at a given angular separation.

9.8 Two-point correlation analysis

In the previous two sections signal and background maps have been created to search for possible anisotropies in the arrival direction of cosmic rays, or more specific, in a dataset where the photon fraction is enhanced as a result of a multivariate analysis. To find potential anisotropies a number of tests for particular anisotropic patterns are available such as the two-point correlation function. For this test, the angular separation distance α between pairs of points are calculated and compared to what is expected from an isotropic distribution. The expected number of pairs within a given angular radius is obtained by simulating a large number of MC simulations (here 1000) with the same number of events using the scrambling technique (cf. Sec. 9.7).

In Fig. 9.15 (*left*) the number of pairs n_p within an angular separation distance α is shown for the hadron reduced data set (red line) as well as for an averaged simulated data set (thick black line) representing the background expectation. As can be seen, no obvious deviation from the background expectation is visible. However, given the relatively large data set of 1950 events large differences would be unlikely. Another way to illustrate possible deviations from the background expectation is to measure the probability that an observed excess of pairs arise by chance from an isotropic distribution of events. To this end the fraction P of simulations with larger or equal number of pairs are counted as a function of separation distance. This is shown Fig. 9.15 (*right*). At angular distances of $\sim 3°$ a minimum for the chance probability P is given by $P \simeq 0.1$, suggesting an excess of clustering in the present data set. This is, however, only marginally significant with the present statistics and it remains to be seen if the correlation is real.

9.9 Combining signal and background maps

The search for anisotropies in cosmic ray arrival directions implies to utilize signal and background expectation for a given point in the celestial map. A common method is to calculate the signal-to-noise ratio $\mathcal{S}/\mathcal{N}$ of a sky location. Here the signal $\mathcal{S}$ is the deviation from the background expectation in a target region. Unless stated otherwise this is the difference of the weighted integrals (cf. Eqn. (9.4)) of the measured signal and the background expectation. The noise $\mathcal{N}$ is the amount of fluctuation in the expected background count. For high statistics this is the square root of the background $\sqrt{N_B}$. The final value using the measured signal N_S and the expected background N_B

$$S_{\text{simple}} = \frac{\mathcal{S}}{\mathcal{N}} = \frac{N_S - N_B}{\text{RMS}} = \frac{N_S - N_B}{\sqrt{N_B}} \ , \tag{9.5}$$

can be regarded as a significance of the signal from a certain sky location. However, Eqn. (9.5) is considering only the statistical fluctuation of the background counts while fluctuations in the signal are not treated. Obviously they underestimate the statistical error of the signal and, therefore, overestimate its significance. A more elaborated way of estimating statistical significance has been proposed by Li and Ma in 1983 [274], today known as Li-Ma method. It is a likelihood ratio method and the significance is estimated by use of the method of hypotheses test in mathematical statistics. The final definition of significance S_{LiMa} is given by [274]

$$S_{\text{LiMa}} = \sqrt{2} \cdot \left\{ N_{\text{on}} \ln \left[\frac{1+\alpha}{\alpha} \left(\frac{N_{\text{on}}}{N_{\text{on}} + N_{\text{off}}} \right) \right] + N_{\text{off}} \ln \left[(1+\alpha) \left(\frac{N_{\text{off}}}{N_{\text{on}} + N_{\text{off}}} \right) \right] \right\}^{1/2} . \tag{9.6}$$

Here, $\alpha = 1/N_{\text{sim}}$ is the inverse of the number of background simulations N_{sim}, $N_{\text{on}} = N_S$ the number of measured counts and $N_{\text{off}} = N_B/\alpha$ the sum of all simulated background counts.

A sky map of Li-Ma significances using the hadron reduced data set and Gaussian weighting with $\sigma = 6°$ is shown in Fig. 9.16. Resulting maximum significances are $S_{\text{LiMa}} \simeq 3$. This rather large σ emphasizes broader structures and is not optimized for finding point sources. However, decreasing the width of the Gaussian weighting requires larger event statistics to have a sufficient number of events as signal and background expectation. The enlargement of the hadron reduced data set is therefore discussed in Sec. 9.10.

Another way to search for possible excesses from isotropy is to measure how often the counts in the simulated data set are equal or exceed the counts given by the actual signal data set of the same region. In other words, for each grid point, the probability is determined by the fraction of simulations giving a weighted sum that is greater or equal to the sum that was obtained using the actual data set. That fraction can be used as the probability P that the excess occurred by chance from an isotropic cosmic ray intensity. The corresponding sky map of chance probabilities is shown in Fig. 9.17. Additionally a log-log plot of the probability distribution is shown to emphasize the behavior in the interesting low end. The solid red line denotes a uniform probability distribution. As can be seen, no significant excess is visible in the tail of the distribution.

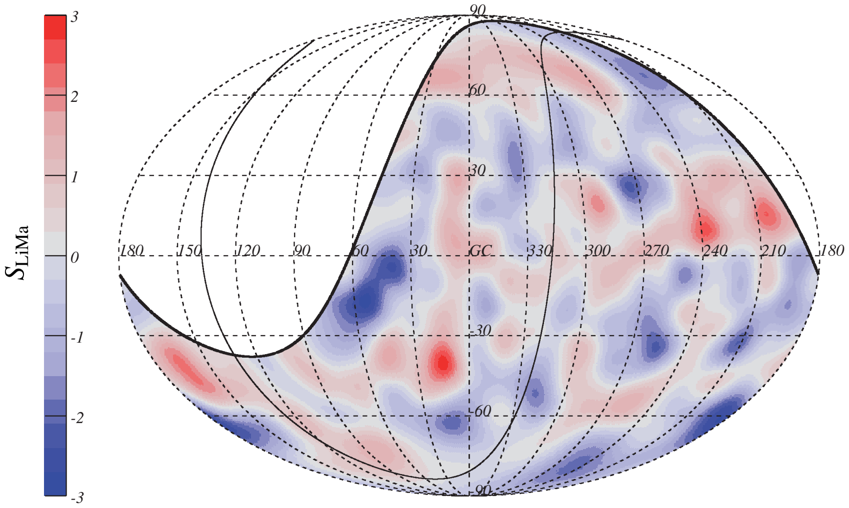


Figure 9.16: Sky map of Li-Ma significances using a Gaussian weight with $\sigma = 6°$ in galactic coordinates. The maximum field of view is indicated by a thick black line while the thin solid line denotes the position of the supergalactic plane. Resulting maximum significances are $S_{\rm LiMa} \simeq 3$.

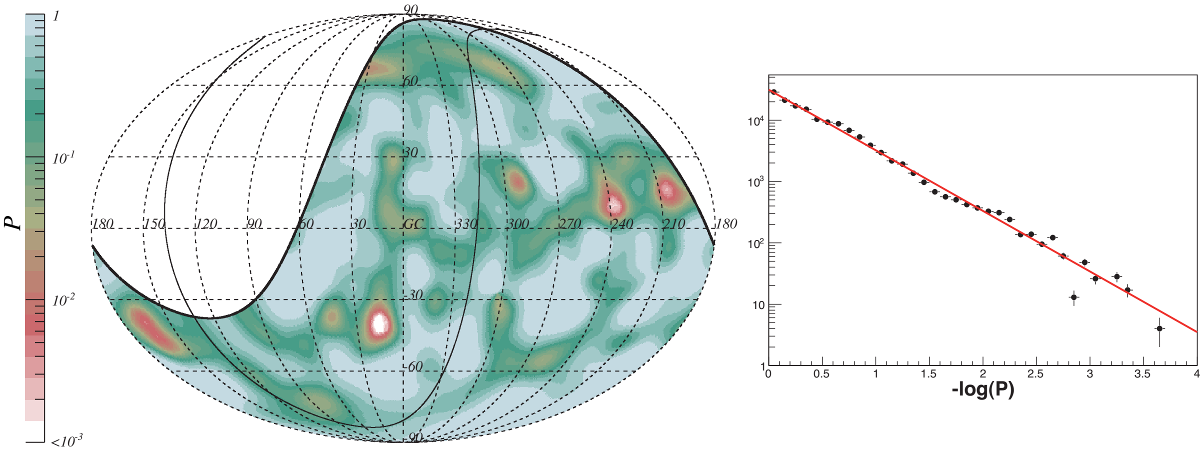


Figure 9.17: *Left panel*: Sky map of chance probabilities using $\sigma = 6°$ in galactic coordinates. Color coded is the fraction of simulations giving a weighted sum that is greater or equal to the sum that was obtained using the actual data set. *Right panel*: Corresponding distribution of probabilities. The log-log plot emphasizes the behavior in the interesting low end. The solid red line denotes a uniform probability distribution.

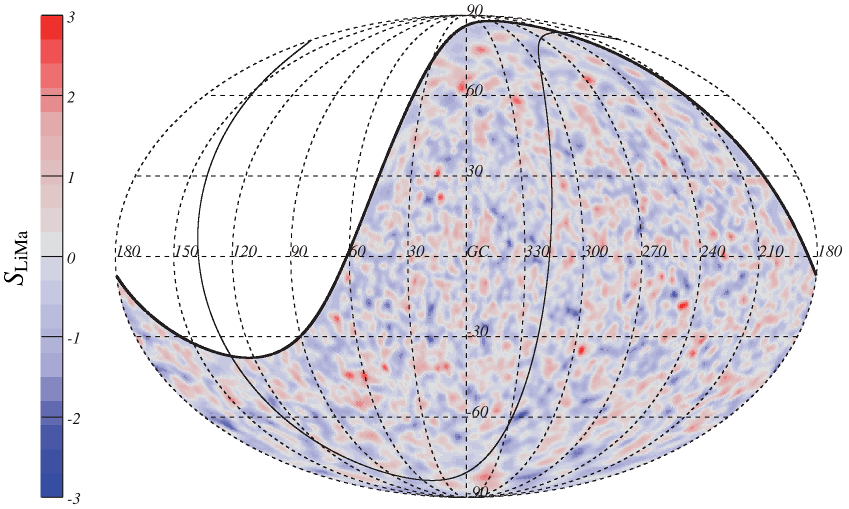


Figure 9.18: Sky map of Li-Ma significances using a Gaussian weight with $\sigma = 1°$ in galactic coordinates. The maximum field of view is indicated by a thick black line while the thin solid line denotes the position of the supergalactic plane.

9.10 Enlargement of the hadron reduced data set

As already discussed in the previous section a minimum number of events in a target region are required to calculate e.g. significance maps. With the current hadron reduced data set of 1950 events it was possible to look only for broader regions using a Gaussian spread of 6°. Now the concept is extended to a Gaussian spread of 1°. To increase statistics, the hadron reduced data set is enlarged by relaxing the final cut value of the BDT output response to -0.38. This corresponds to a working point where the signal efficiency is equal to the background rejection $\epsilon_S = 1 - \epsilon_B = 0.865$ (cf. Table 8.7). In case of a 1% photon fraction (100000 protons and 1000 photons) the significance $S/\sqrt{B}$ is ~ 7.5, cf. Fig. 8.13. The final enlarged data set consists of 12963 events. Unless stated otherwise, the new enlarged data set is used in the subsequent analysis.

A sky map of Li-Ma significances using the new data set and a Gaussian weighting of $\sigma = 1°$ is shown in Fig. 9.18. The supergalactic plane as well as the field of view is illustrated. There are some evidences for overdensities, but no obvious hot spot regions. The corresponding map of chance probabilities is shown in Fig. 9.19. A slight deviation from isotropy is visible at the lower end of the log-log plot. However, given the relatively low number of 5000 background maps $-\log(P)$ is always $\lesssim 3.7$. Larger values are accumulated in the last bin increasing that bin artificially.

Increasing the reconstructed energy artificially by 20%, as described in Sec. 9.3, does not affect the calculated Li-Ma significance substantially. However, a trend towards lower significances is visible.

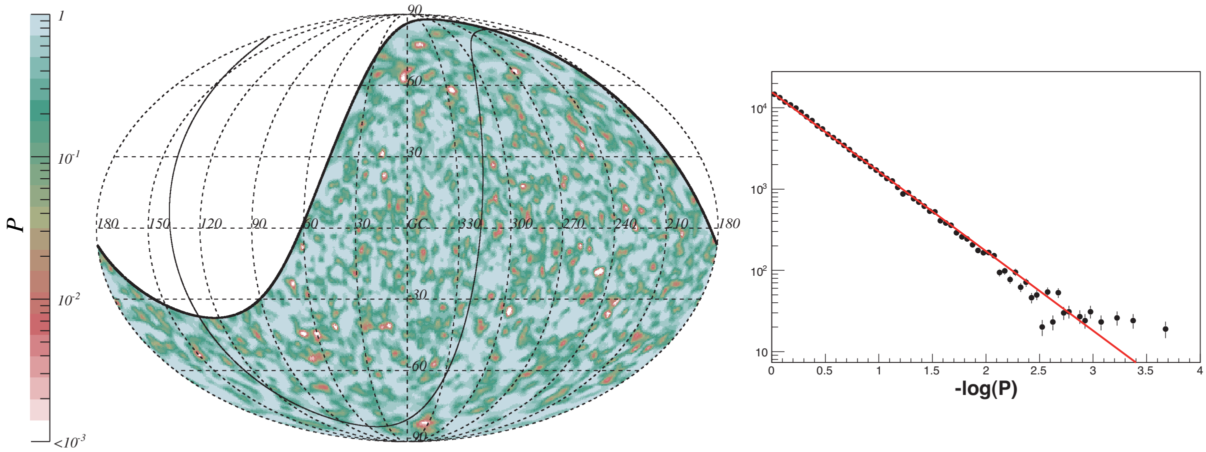


Figure 9.19: *Left panel*: Sky map of chance probabilities using $\sigma = 1°$ in galactic coordinates. Color coded is the fraction of simulations giving a weighted sum that is greater or equal to the sum that was obtained using the actual data set. *Right panel*: Corresponding distribution of probabilities. The log-log plot emphasizes the behavior in the interesting low end. The solid red line denotes a uniform probability distribution.

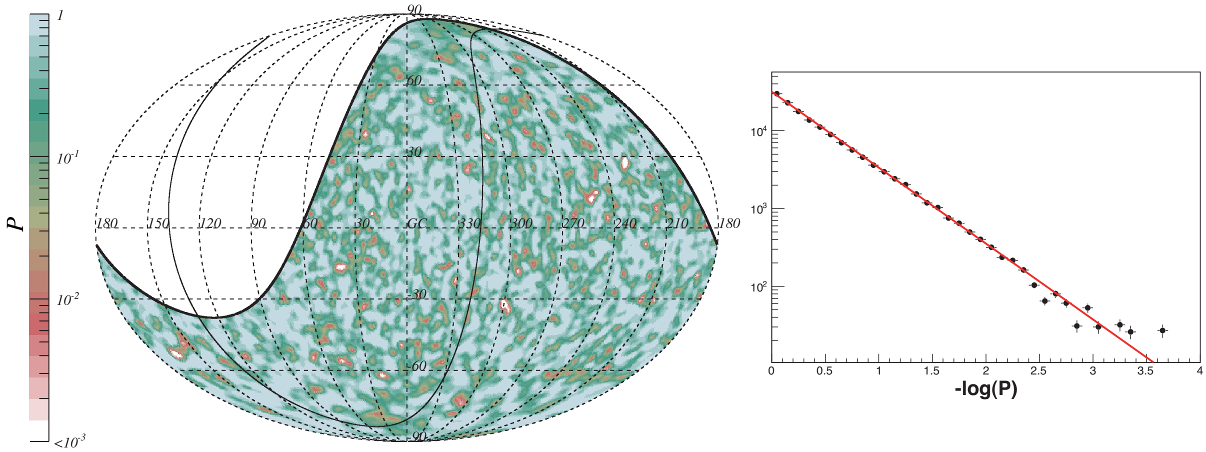


Figure 9.20: *Left panel*: Sky map of chance probabilities with $\sigma = 1°$ in galactic coordinates using **all** events, i.e. with no specific cut on photon-like events. Color coded is the fraction of simulations giving a weighted sum that is greater or equal to the sum that was obtained using the actual data set. *Right panel*: Corresponding distribution of probabilities. The log-log plot emphasizes the behavior in the interesting low end. The solid red line denotes a uniform probability distribution.

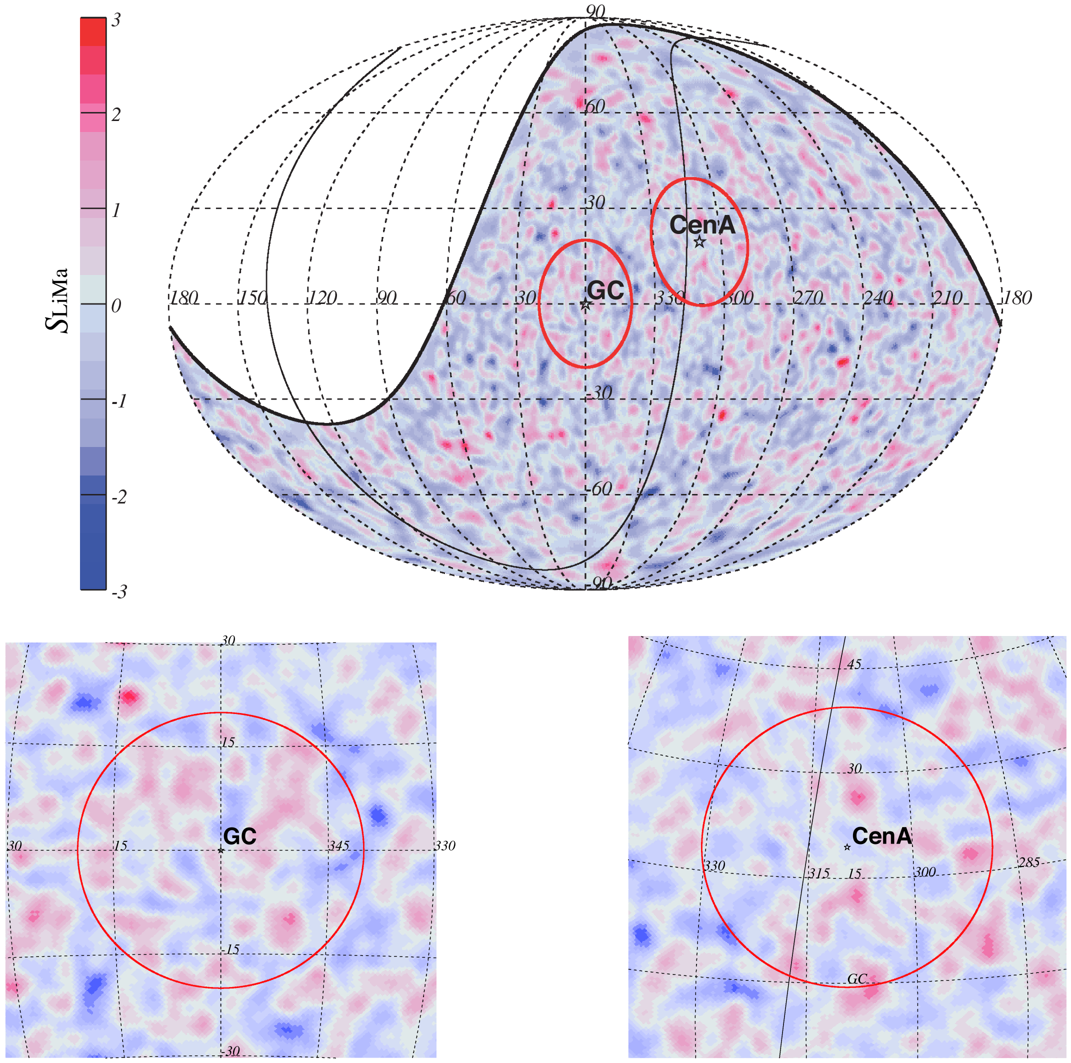


Figure 9.21: *Top*: Li-Ma significance map (cf. Fig. 9.18) emphasizing the position of the Galactic Center ($l = 0°$, $b = 0°$) and Centaurus A. ($l = 309.5°$, $b = 19.42°$) region. A radius of 20° around the Galactic Center and Centaurus A is denoted by a red circle. *Bottom*: Zoom of the two regions illustrated in a Lambert azimutal projection [270].

A data set without photon selection, i.e. no cut on a BDT response output, is illustrated in Fig. 9.20. This has to be compared with Fig. 9.19. It can be seen that, in general, overdensities in the complete data set do not point out in the enlarged hadron reduced data set and vice versa. Observed overdensities can be explained by statistical fluctuations and no statistically significant excess has been found.

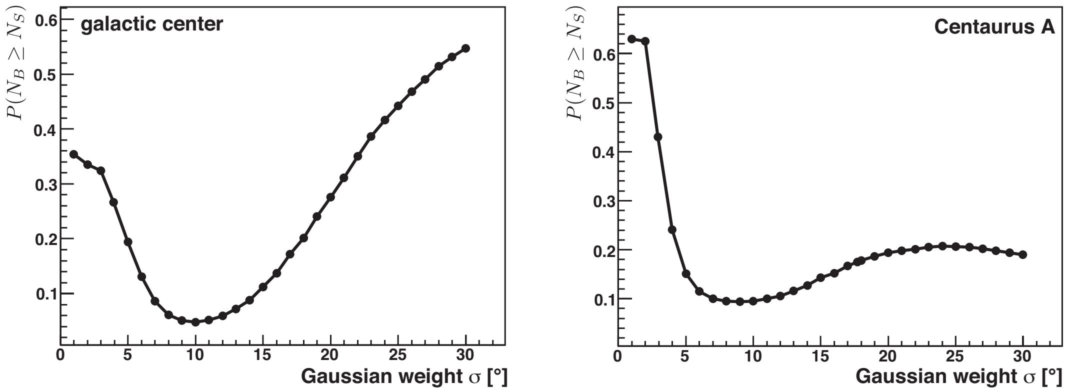


Figure 9.22: Probability that the background count (N_B) is equal or exceeds the signal count (N_S) for a given Gaussian weight σ. Two different arrival directions are shown, namely the Galactic Center (*left panel*) and Centaurus A (*right panel*). Note that the maximum angular integration distance has also been increased to 90°.

9.11 Galactic Center and Centaurus A region

Sec. 5.1 introduces different scenarios in which UHE photons are produced. GZK-processes would lead to a guaranteed production via decay processes of neutral pions ($\pi^0 \rightarrow \gamma + \gamma$). Also specific sources or regions may produce an enhanced photon flux such as the Galactic Center region or the nearest AGN Centaurus A at a distance of ∼3.4 Mpc.

The position of these two candidates is illustrated in Fig. 9.21. The red circles indicate an angular separation of 20° from Centaurus A as well as from the Galactic Center. As can be seen, no dominant excess is visible at their source position. When increasing the Gaussian smoothing σ, however, the sensitivity for broader extended source regions is enhanced while losing sensitivity for point like sources. The effect of an increased σ on the Galactic Center region and Centaurus A is shown in Fig. 9.22. Similar to e.g. Fig. 9.19 the probability is calculated that the background count (N_B) is equal or exceeds the signal count (N_S). Small probabilities indicate a large deviation from the isotropic expectation. A minimum is visible using a σ of about 10° with probabilities of 5% and 10% for the Galactic Center region and Centaurus A, respectively. An additional analysis using top-hat counting indicates similar results. These outcomes are in agreement with statistical fluctuations.

9.12 Source upper limit on the photon fraction

The introduced search for EeV photons does not show any strong hot spots indicating that the fraction of photons over a large hadron background is small. It remains to be seen with increasing statistic and enhanced analysis techniques if actual photon source candidates (i.e. regions with large significance and low chance probability) will be confirmed to be real. It is therefore natural to derive an upper limit on the photon fraction for candidate sources, or more generally, for each grid point in the sky [275]. The upper limit will

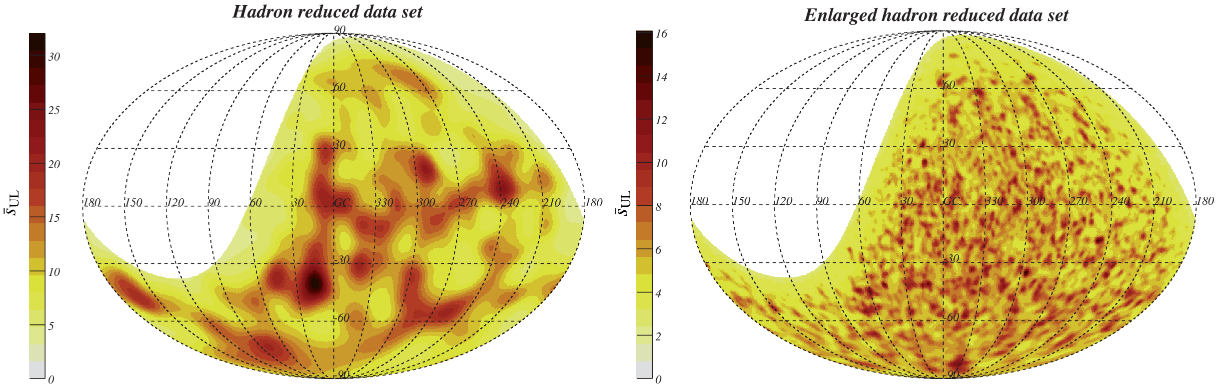


Figure 9.23: *Left panel*: Upper limit on the number of photons $\bar{s}_{\mathrm{UL}}$ in the hadron reduced data set from a putative source at 90% confidence level using a 6° Gaussian weight in galactic coordinates (cf. Eqn. (9.8)). *Right panel*: Upper limit on the number of photons $\bar{s}_{\mathrm{UL}}$ in the enlarged hadron reduced data set from a putative source at 90% confidence level using a 1° Gaussian weight in galactic coordinates (cf. Eqn. (9.8)).

be influenced by the exposure variation, but also by fluctuations in the observed data set. Upper limits are simple observations without any associated claim of discovery and consequently there is no need for a prescribed test. They are just based on the observed and expected sky densities.

An important issue is the introduction of an appropriate confidence level (CL). This is a particular kind of interval estimate of a population parameter and is used to indicate the reliability of an estimate. A common choice is CL $= 0.9$ which means that the result will meet the expectation in 90% of hypothetically repeated trials. In this analysis also the complementary small fraction $\alpha \equiv 1 - \mathrm{CL} = 0.1$ is used.

One way to find an upper limit for the number of signal events $\bar{s}_{\mathrm{CL}}$ from a certain sky direction at a given confidence level is to find the probability P such that

$$P(\leqslant n|\bar{b} + \bar{s}_{\mathrm{UL}}) = \alpha \ . \tag{9.7}$$

Here n is the observed number and $\bar{b}$ the expected background contribution. $\bar{s}_{\mathrm{CL}}$ is treated as a variable and is varied such that a Gaussian distribution centered on $\bar{b} + \bar{s}_{\mathrm{UL}}$ (and standard deviation given by its square root) has only the small fraction α of its area below the measured value n [271]. The interpretation is that if $\bar{s}_{\mathrm{UL}}$ is the expected number from a source, the measured number (consisting of signal and background) will be equal or less than n only in the fraction α of hypothetical repeated trials. The definition given in Eqn. (9.7) has, however, the disadvantage that it leads to negative upper limits in case the observed number n is significantly smaller than $\bar{b}$ caused by e.g. a downward fluctuation in the background. A way to overcome the problem is by using a posteriori distribution for the background. These are obtained using the restrictions given by experimental data and was proposed by G. Zech in the late 80s [276]. In this thesis the following equation is adopted, cf. [271]:

$$P(\leqslant n|\bar{b} + \bar{s}_{\mathrm{UL}}) = \alpha \cdot P(\leqslant n|\bar{b}) \ . \tag{9.8}$$

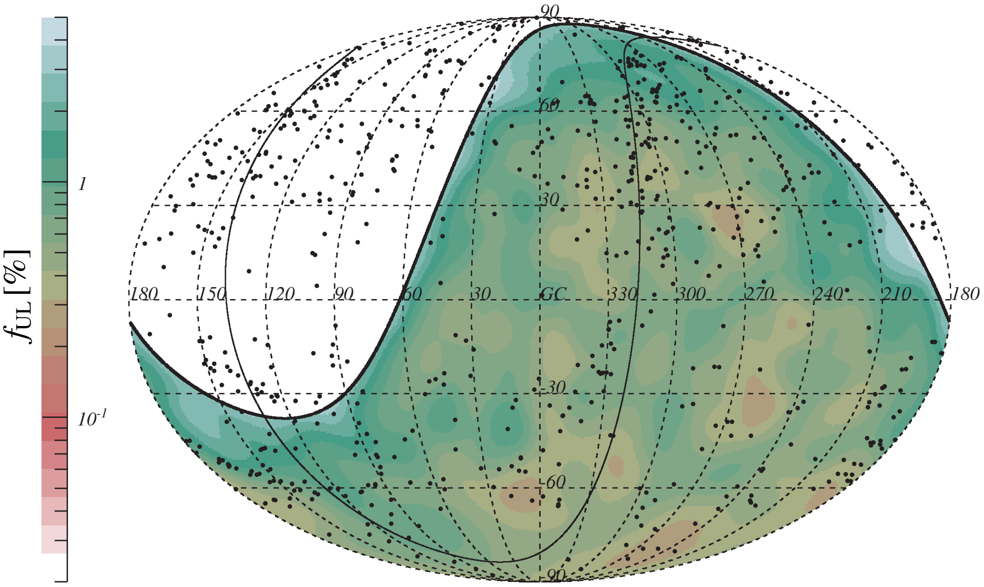


Figure 9.24: Upper limit on the photon fraction f_{UL} in % from a putative source at 90% confidence level using a 6° Gaussian weight in galactic coordinates using the hadron reduced data set. AGN from the VCV catalog within a distance of 100 Mpc are shown as black dots. The solid thick line indicates the maximum field of view while the thin black line denotes the supergalactic plane.

For $n \gg \bar{b}$ this agrees with Eqn. (9.7), but in case of $n \ll \bar{b}$ this does not become negative but respects the possibility that fluctuations in the background may have masked a positive expected signal. The new definition is not a true frequentist confidence or probability, but is similar [277] to the CLs technique introduced in [278]. Since the new condition on the background scales up the limit, it can be regarded as a conservative and very robust frequentist-motivated extension when making statements about the signal only [278]. The limit obtained by exploiting Eqn. (9.8) can be stated in the frequentist interpretation as follows: "For an infinitely large number of repeated experiments looking for a signal with expectation $\bar{s}_{\mathrm{UL}}$ and background with mean $\bar{b}$ – *where the background is restricted to values equal or less than* n – the frequency of observing n or fewer events is α". The Bayesian interpretation with flat prior on the expected signal (i.e. all values of $\bar{s}_{\mathrm{UL}}$ are equally likely) is in agreement[5] with Eqn. (9.8) and can be written as

$$\int_{\bar{s}_{\mathrm{UL}}}^{\infty} P(n|\bar{b}+\bar{s})\;\mathrm{d}\bar{s} = \alpha \cdot \int_{0}^{\infty} P(n|\bar{b}+\bar{s})\;\mathrm{d}\bar{s}\;. \tag{9.9}$$

Applying Eqn. (9.8) to the (enlarged) hadron reduced data set (and background maps obtained using the scrambling method), the maximum number of expected signal counts from a source $\bar{s}_{\mathrm{UL}}$ within the used data set can be calculated. Fig. 9.23 illustrates $\bar{s}_{\mathrm{UL}}$ for a given sky direction. The left sky map denotes the result using the hadron reduced

---

[5] Assuming Gaussian or Poisson probability distributions

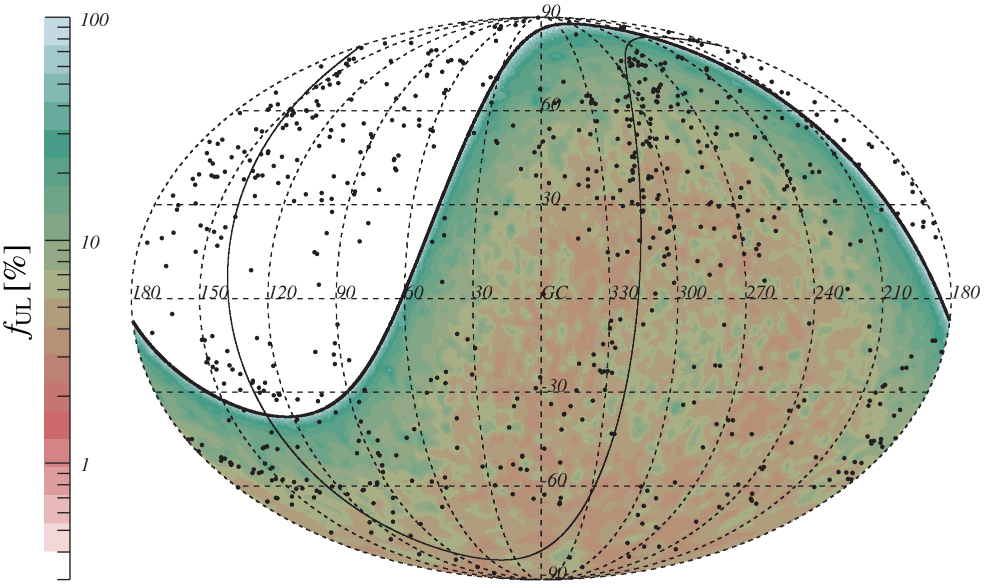


Figure 9.25: Upper limit on the photon fraction f_{UL} in % from a putative source at 90% confidence level applying a 1° Gaussian weight in galactic coordinates using the enlarged hadron reduced data set. AGN from the VCV catalog within a distance of 100 Mpc are shown as black dots. The solid thick line indicates the maximum field of view while the thin black line denotes the supergalactic plane.

data set applying a 6° Gaussian smoothing to emphasize the behavior of more extended sources. The right map shows the enlarged hadron reduced data set applying a 1° Gaussian smoothing to stress small scale structures and point sources. At this stage the number of photon counts only refers to the (enlarged) hadron reduced data set. To calculate the upper limit on the photon fraction for a given sky direction and accordingly a source location, the results have to be compared to the complete data set, i.e. the data set using no multivariate cut. Let ϵ_S be the photon efficiency at a given cut. In case of the (enlarged) hadron reduced data set the efficiency is given by $\epsilon_S^{\mathrm{HRD}} = 0.46$ ($\epsilon_S^{\mathrm{enlarged}} = 0.86$). The upper limit on the photon fraction f_{UL} for a certain arrival direction is calculated as follows:

$$f_{\mathrm{UL}} = \frac{\bar{s}_{\mathrm{UL}}}{n_{\mathrm{ALL}} \cdot \epsilon_S} \; . \tag{9.10}$$

Here n_{ALL} denotes the measured number of counts of a given sky direction using the full data set. Since there is no full trigger efficiency at EeV energies and below, differences in the trigger probability for varying primary particles have to be investigated. As a consequence of the later development of photon primaries (deeper X_{max}) they tend to be more efficient (up to $\sim 30\%$ compared to proton primaries) at energies $< 10^{18}$ eV [279]. This yields the upper limit to be more conservative since, in general, more photons are triggered compared to hadron primaries.

The upper limit of the photon fraction f_{UL} using the hadron reduced data set and a Gaussian weight of $\sigma = 6°$ is shown in Fig. 9.24. Superimposed as black dots are the positions of AGN within a distance of 100 Mpc obtained from the Véron-Cetty & Véron catalog [61]. Also shown as solid thin line is the position of the supergalactic plane. Typical upper limits are of the order of 1% but can decrease down to 0.1%. To enhance the sensitivity to point-like sources the Gaussian weight is modified to $\sigma = 1^\circ$ while using the enlarged hadron reduced data set. The upper limit on the photon fraction applying the new configuration is illustrated in Fig. 9.25. As can be seen stronger upper limits can generally be set in regions with high exposure while at the edge of the field of view no conclusion can be made. Typical directional upper limits are a few % in well exposed regions.

Increasing the reconstructed energy artificially by 20%, as described in Sec. 9.3, increases slightly the derived upper limit since more photon-like events are included. The typical upper limit on the photon fraction increases from 4.9% to 5.5% using a 1° Gaussian weight.

9.13 Summary

A search for EeV photons utilizing data from the Pierre Auger Observatory has been performed in this chapter. The fair amount of low energy hybrid events was used to create a hadron reduced data set based on a multivariate analysis for photon/hadron separation. Another feature was the inclusion of individual event uncertainties when creating event density maps. The background expectation for isotropic arrival directions of cosmic rays was determined using the scrambling technique by shuffling UTC arrival times of individual events of the same triggered telescope. The resulting significance maps do not yet show an obvious hot spots and deviations from the background expectation are in agreement with statistical fluctuations. It remains to be seen with increasing statistic and enhanced analysis techniques if actual photon source candidates will be confirmed to be real. Furthermore an analysis of the Galactic Center region and Centaurus A could not conclude a significant excess.

For the first time directional upper limits on the photon fraction for sources have been derived. On average the limit is of the order of a few % in well exposed regions using a Gaussian weight of 1°. Increasing the Gaussian weight to 6° (and thereby the collection area and statistics) reduces the average upper limits by one order of magnitude while losing sensitivity for point sources.

Chapter 10

Summary and Outlook

Photons are still the main messengers when exploring the universe. Large telescopes and observatories have been build to collect light that was emitted thousands of years ago from distant stars. Within the recent century new technologies have been developed extending already observed photon energies up to $\sim 10^{14}$ eV. This thesis has been dedicated to photons with energies 10000 times larger. Although they have never been observed, theoretical and experimental results predict a certain photon fraction with a variety of prospects. A unique tool to study photons at these extreme energies is the Pierre Auger Observatory located in the Argentinean Pampa Amarilla. The worldwide largest observatory to study cosmic rays utilizes a hybrid design comprising of surface detectors and fluorescence telescopes. The combination of these two techniques to search for a directional excess of ultra-high energy photons has been performed for the first time in this thesis. The most important outcomes are summarized below.

- **Production** and **propagation** of UHE photons (Chapter 6):
  The propagation of ultra-high energy protons and photons have been simulated using the Monte Carlo based propagation code CRPropa. There is a strong dependence on the evolution of the extra-galactic background light in particular the CMB for distant sources (> 100 Mpc). Relative fluctuations of energies to the mean energy are dominant in the range between $\sim 5-40$ Mpc and are important for a spectrum derived from a low number of events.

  Expected photon fluxes show a non-trivial dependence on the source distance cf. Fig. 6.4 due to the competition between GZK photon production and attenuation (both of which are increasing with travel distance). GZK photon fluxes from a proton source are mainly observed within a propagation distance of up to $25-50$ Mpc with a peak at about 10 Mpc.

  The effect of different input parameters on the GZK-horizon has been investigated. For the maximum energy produced in the source, differences are of the order of $\sim 5\%$ above an energy threshold of $E^{\text{thres}} \simeq 120$ EeV (cf. Fig. 6.7). There is a more constant (energy independent) offset for varying spectral indices. A variation of ±10% induces an offset of $\sim 2\%$ in the horizon. Furthermore, a variation of ±10% of the Hubble parameter modifies the horizon of the order of $\sim 2\%$ for lower energy thresholds (< 80 EeV).

Regarding the specific case of Centaurus A, the current photon flux limit [81] does not yet constrain Centaurus A as a strong source of UHE protons for the investigated range of spectral indices α and maximum energies E_{max} (Fig. 6.8). However, the sensitivity that will be accumulated by Auger South will allow interesting constrains for a broad range of α and E_{max} (Fig. 6.8 and Fig. 6.10).

Depending on source parameters, the number of GZK photons above 10 EeV may reach several 10 over the lifetime of Auger South (Fig. 6.9). One can conclude that the search for UHE photons helps to provide significant clues about the characteristics of potential astrophysical sources.

- Multivariate photon/hadron **discrimination** (Chapter 7 and 8):
  A full CORSIKA and $\overline{\text{Off}}\underline{\text{line}}$ simulation has been performed to study standard as well as newly developed surface detector and fluorescence telescope observables for discrimination of photon and proton induced air showers. These observables have been combined in a multivariate analysis while special attention was turned to boosted decision trees and artificial neural networks. As an advanced classifier they are able to take into account high-dimensional correlations using different separation philosophies. The best performing boosted decision tree was found to be using the gradient boost algorithm in combination with a small shrinkage parameter (to reduce the learning rate) making the method less susceptible for overtraining (cf. Fig. 8.5). It turned out that the introduction of a bagging-like resampling procedure using random subsamples of the training events for growing the trees can further enhance the separation power.

  When using artificial neural networks the back propagation by the Broyden-Fletcher-Goldfarb-Shannon algorithm indicated better performance than the standard back propagation. The introduction of two hidden layers could further increase the separation power (cf. Fig. 8.10).

  Differences between neural networks and boosted decision trees are small. Using the optimal working point for selection, both methods reduce the proton fraction by about $\sim 98\%$ while keeping $\sim 50\%$ of photon primaries. It was found that boosted decision trees are more robust with a slightly better performance. The ability to view individual trees makes the method more transparent and manageable (cf. Fig. 8.11).

- **Search** for photons (Chapter 9):
  The multivariate analysis has been applied to real data measured with the Pierre Auger Observatory resulting in a hadron reduced data set containing photon-like air showers. Individual event uncertainties have been taken into account to create signal maps as well as background maps using the scrambling method. A two-point

correlation analysis indicates a minimum chance probability P of $\sim$ 10% at an angular separation of 2.5°.

Directional excesses have been investigated in the energy range between $10^{17.2}$ eV and $10^{18.5}$ eV using the significance defined by Li and Ma. Independent of the applied Gaussian weight, significances are $\lesssim 3\sigma$. Increasing the reconstructed energy artificially by 20% does not affect the calculated Li-Ma significance substantially. However, a trend towards lower significances is visible. In addition also directional chance probabilities have been calculated finding no hot spot that stands out above the expected statistical fluctuations. Furthermore, deviations from the background expectation are in agreement with statistical fluctuations. It remains to be seen with increasing statistic and enhanced analysis techniques if actual photon source candidates will be confirmed to be real.

An analysis of the Galactic Center region and Centaurus A could not conclude a significant excess. A minimum chance probability of 5% and 10% using a Gaussian weighting of $\sigma \sim 10°$ has been found in the direction of the Galactic Center and Centaurus A, respectively (cf. Fig. 9.22).

- Directional **upper limits** on the photon fraction (Chapter 9):
  For the first time directional upper limits on the photon fraction for sources in the energy range between $10^{17.2}$ eV and $10^{18.5}$ eV have been derived. The limit has been calculated by a method similar to the CLs technique at 90% confidence level. On average the limit is of the order of a few % in well exposed regions using a Gaussian weight of 1° which is sensitive to point-like sources, cf. Fig. 9.25. Increasing the reconstructed energy artificially by 20% increases slightly the derived upper limit since more photon-like events are included. The typical upper limit on the photon fraction increases from 4.9% to 5.5% using a 1° Gaussian weight.
  Increasing the Gaussian weight to 6° (and thereby the collection area and statistics) reduces the average upper limits by one order of magnitude with the drawback of losing sensitivity for point sources, cf. Fig. 9.24.

Potential for further analysis

Beyond the results presented in this thesis, there is potential for further analysis in the future. A selection is given below:

→ **Increasing statistics**: Besides the continuously increasing statistics of hybrid data using the "standard" telescopes, the HEAT and AMIGA extension (cf. e.g. [280]) of the observatory will offer a unique tool to study cosmic rays around 10^{17} eV. Inclusion of this data would enhance the separation power, increase the sensitivity for photon sources and improve the quality of low energy data in general.

→ **Flux upper limits**: Not only photon fractions, but also upper limits on the photon flux are of particular interest. The results on the expected number of photons have to be divided by the acceptance to get limits on the flux. Time dependent detector simulations have been performed e.g. in [281] and could be applied.

→ **Interaction models**: A detailed modeling and extrapolation of hadronic interaction processes is an essential feature in understanding cosmic rays at ultra-high energies. New developments, especially when entering the LHC era, will enhance current models making MC simulations more realistic, cf. e.g. [282, 283, 284].

→ **Energy range**: This analysis is not restricted to the applied energy range between $10^{17.2}$ eV and $10^{18.5}$ eV. Extending this range to higher energies improves the event quality while losing statistics. In addition, new effects start to be dominant, e.g. preshower conversion and the LPM effect, cf. Sec. 5.4.

→ **Surface detector**: As a natural next step this analysis can be extended to surface data only. With its nearly 100% duty cycle a large amount of statistics is already available. Furthermore, acceptance and aperture effects could be taken into account more easily to calculate photon fluxes or photon flux upper limits. Already existing as well as new observables can be used in a multivariate approach.

→ **Constraining source parameters**: Directional upper limits on the photon fraction could be used to constrain source parameters (such as the maximum energy E_{max} and the spectral index α) of possible point sources, cf. Sec. 6.5.

→ **New features in TMVA**: The continuous development of the TMVA software framework enables new features to improve the separation and stability of the discrimination analysis [285]. To reduce the correlation between observables a *category* extension has recently been developed. The phase space is divided into disjunct regions while each region is trained and tested separately. Energy and zenith angle dependencies could be reduced while improving the classification performance. Utilizing a *multi-class classification* opens up the possibility not only to classify between signal and background but also between more than two classes, e.g. between photon, proton, helium and iron particles.

Final remarks

The origin and composition of ultra-high energy cosmic rays remains a mystery. However, small steps towards a deeper understanding have been made but it is too early to draw any decisive conclusion. Giant air shower experiments are a unique tool to continue these steps. The Pierre Auger Observatory is currently the most promising observatory to investigate pressing issues in ultra-high energy cosmic ray physics. It is just a question of time until the existence of ultra-high energy photons will be confirmed or disproved. A detection would open a new way to see our universe while a non-detection is not necessarily uninterestingly. It could guide the way to possible new physics and turn adopted principles upside down.

Appendix A

Steering CRPropa

The steering of the simulation is controlled by XML steering cards. A brief description is give below. More detailed explanations can be found in [286].

- **General parameters**:
  - `<TrajNumber value=10000000 />`
    The number of particles to be followed including possible secondaries.
  - `<MinEnergy_EeV value=10 />`
    Threshold energy in EeV below which charged particles are abandoned. The minimum energy is 0.1 EeV.
  - `<MaxTime_Mpc value=10 />`
    Maximum propagation time for charged particles before they are abandoned. Note that the time is given in units of Mpc.
  - `<RandomSeed value=136097 />`
    Seed used for generating random numbers.
  -
    ```
    <Output type="Events">
       <File type="ROOT" option="force"> Output.root </File>
    </Output>
    ```
    Specifying the output of the simulation. "*Events*" denotes that only particles and secondaries that reach the observer are recorded. The output is written into a root file named "*Output.root*".
  -
    ```
    <OmegaM value=0.3 />
    <OmegaLambda value=0.7 />
    <H0_km_s_Mpc value=71.  />
    ```
    Specifying cosmological parameters Ω_{M}, Ω_{λ} and the Hubble constant H_0 in km/s/Mpc.
- **Environment parameters**:
  `<Environment type="One Dimension" />`
  CRPropa is able to simulate in a one and three dimensional environment.

- **Interaction parameters**:
  ```
  <Interactions type="Sophia">
     <MaxStep_Mpc value=0.1 />
     <SecondaryPhotons />
     <SecondaryPairProdPhotons />
  </Interactions>
  ```
  Three interaction models are currently implemented in CRPropa. The standard interaction type is “*Sophia*”. Here proton pair production is treated as a continuous energy loss using tables derived from the DINT package. At each propagation step defined by “*MaxStep_Mpc*” pion production is checked using SOPHIA. “SecondaryPhotons” and “SecondaryPairProdPhotons” allows for electromagnetic cascades generated by pion and pair production, respectively.

- **Source parameters**:
  ```
  <Sources type="Discrete">
     <Number value=1 />
     <PointSource>
         <CoordX_Mpc value=3.4 />
     </PointSource>
     <Spectrum type = "Power Law" >
         <Alpha value=2.8 />
         <Ecut_EeV value=500 />
     </Spectrum>
  </Sources>
  ```
  Sources can either be “*Continuous*” or “*Discrete*”. While continuous sources must be defined with a space density, discrete sources can directly defined with a list of coordinates. In the one dimensional case the distance of the source to the observer is given by “*CoordX_Mpc*”. The source energy can either be fixed or following a spectral index. In the latter case the spectral index “*Alpha*” and the maximum energy “*Ecut_EeV*” has to be set.

An example input card is given in Fig. A.1.

```
+++++++++++++++ Example input card CRPropa +++++++++++++++

<?xml version="1.0" standalone=no>

<TrajNumber value=10000000 />
<MinEnergy_EeV value=10 />
<MaxTime_Mpc value=10 />
<RandomSeed value=136097 />

<Output type="Events">
    <File type="ROOT" option="force"> Output.root </File>
</Output>

<Environment type="One Dimension" />

<OmegaM value=0.3 />
<OmegaLambda value=0.7 />
<H0_km_s_Mpc value=71.  />

<Interactions type="Sophia">
    <MaxStep_Mpc value=0.1 />
    <SecondaryPhotons />
    <SecondaryPairProdPhotons />
</Interactions>

<Sources type="Discrete">
    <Number value=1 />
    <PointSource>
        <CoordX_Mpc value=3.4 />
    </PointSource>
    <Spectrum type = "Power Law" >
        <Alpha value=2.8 />
        <Ecut_EeV value=500 />
    </Spectrum>
</Sources>

++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
```

Figure A.1: Example of a CRPropa input card for a one dimensional simulation. An explanation of the different steering parameters is given in App. A.

Appendix B

Steering CORSIKA

The steering of a CORSIKA simulation is controlled by `ascii` steering files. A sample steering card is shown in Fig. B.1. Important steering parameters are:

- `ECUTS`:
  Defines energy cut-offs for different particles. The low-energy cut-off in GeV may be chosen differently for hadrons. The first number refers to the cut-off energy for hadrons[1], second number for muons, third for electrons and fourth for photons including π^0s.

- `ERANGE`:
  Defines the energy in GeV of the primary particle. The first and second number is the lower and the upper limit, respectively. The primary energy is chosen randomly out of the given interval using a differential energy spectrum with slope `ESLOPE`. If the first and second number is equal, a fixed energy at this value is used.

- `OBSLEV`:
  Ovservation level for particle output above sea level in cm.

- `PHIP`:
  Azimuth angle definition. The azimuth is selected randomly out of the interval given by the first and second number (in [°]).

- `PRMPAR`:
  Definition of the primary particle. E.g. '1' refers to photon and '14' to proton primaries.

- `THETAP`:
  Zenith angle definition given in [°]. First and last value refers to the low and high edge value, respectively. The zenith angle is selected randomly out of this interval in a manner which respects equal particle fluxes from all solid angle elements of the sky.

---

[1] without π^0s

- `THIN`:
  Thinning definition. The first value defines the energy fraction of the primary energy below which the thinning algorithm becomes active. The second parameter is the weight limit for thinning. If the weight of the particle exceeds that value, no further thinning is performed. The third parameter is the maximum radius at observation level in cm within which all particles are subject to inner radius thinning.

- `THINH`:
  Option for hadronic thinning. The first parameter defines the hadronic thinning limit differing from electromagnetic thinning. The second parameter defines the hadronic weight limit differing from the electromagnetic weight limit.

More detailed explanations on steering parameters and algorithm can be found in [113].

```
+++++++++++++++ Example input card CORSIKA +++++++++++++++
DATBAS T
DIRECT ’data/’
ECTMAP 25000
ECUTS 0.1 0.1 0.00025 0.00025
ELMFLG T T
ERANGE 5623413251.90349 5623413251.90349
ESLOPE 0
EVTNR 12
HOST acui
LONGI T 5 T T
MAGNET 20.1 -14.2
MAXPRT 1
MUADDI T
MUMULT T
NSHOW 1
OBSLEV 145200
PAROUT T T
PHIP 0 360
PRMPAR 1
RADNKG 500000
RUNNR 1
SEED 23787986 0 0
STEPFC 1
THETAP 0 0
THIN 1e-06 5623.41325190349 30000
THINH 1 100
USER kuempel
SEED 23787987 0 0
EXIT
++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
```

Figure B.1: Example CORSIKA input card used for a photon primary and fixed energy and theta angle. An explanation of some steering parameters is given in App. B. More detailed explanations can be found in [113].

```
+++++++++++++++ Example input card CORSIKA +++++++++++++++
DATBAS T
DIRECT 'data/'
ECTMAP 25000
ECUTS 0.1 0.1 0.00025 0.00025
ELMFLG T T
ERANGE 158489319.246111 3162277660.16838
ESLOPE -2.7
EVTNR 2
HOST acui
LONGI T 5 T T
MAGNET 20.1 -14.2
MAXPRT 1
MUADDI T
MUMULT T
NSHOW 1
OBSLEV 145200
PAROUT T T
PHIP 0 360
PRMPAR 14
RADNKG 500000
RUNNR 57520
SEED 30444765 0 0
STEPFC 1
THETAP 0 60
THIN 1e-06 158.489319246111 30000
THINH 1 100
USER kuempel
SEED 30444766 0 0
EXIT
+++++++++++++++++++++++++++++++++++++++++++++++++++++++++
```

Figure B.2: Example CORSIKA input card of a primary proton with spectral energy and realistic theta distribution. An explanation of some steering parameters is given in App. B. More detailed explanations can be found in [113].

Appendix C

Miscellaneous

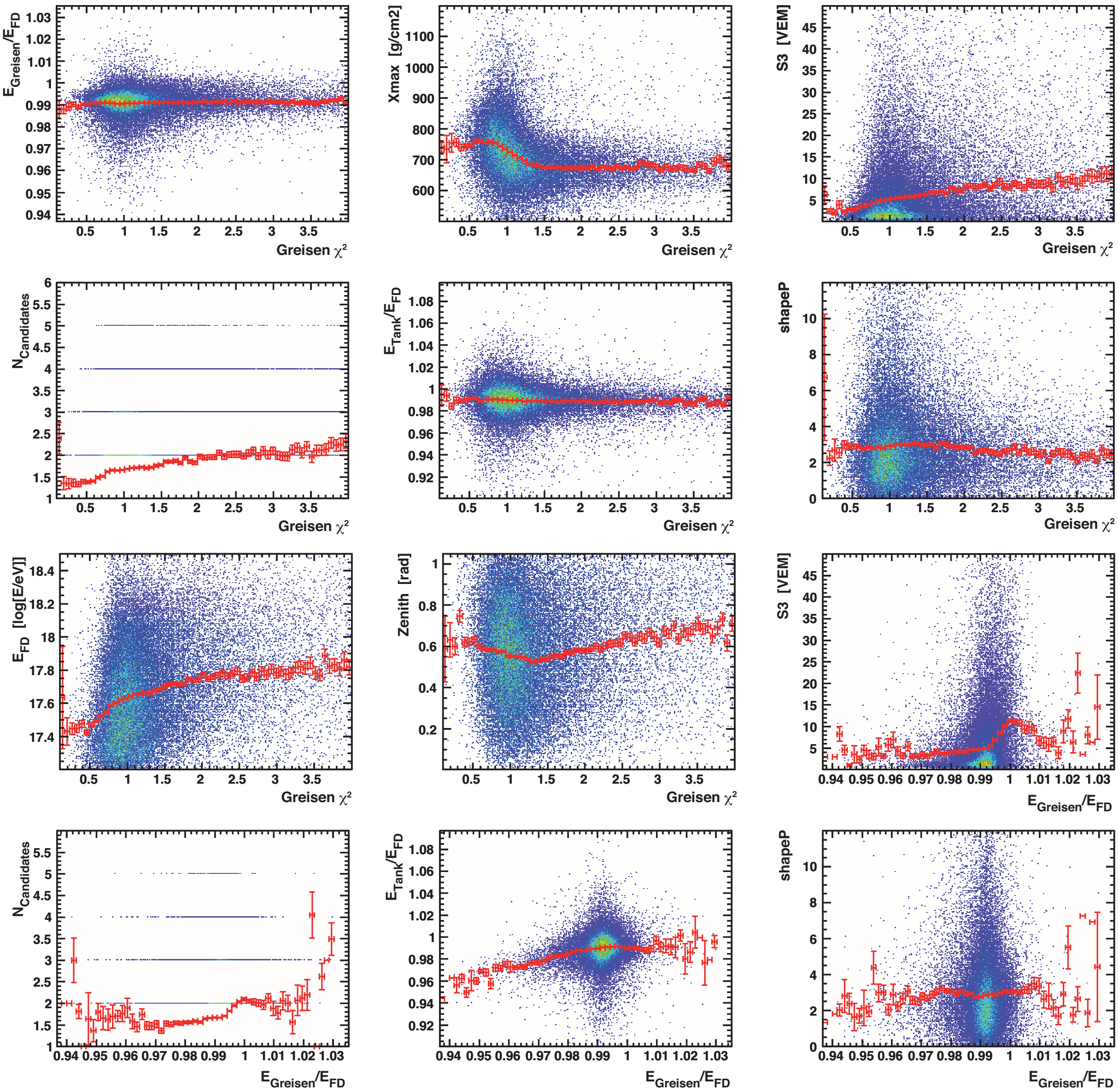


Figure C.1: Scatter plot of various observables to identify possible correlations. Each dot represents one event while the red markers denote the profile including statistical errors. Remaining combinations can be found in Fig. C.2 and Fig. C.3.

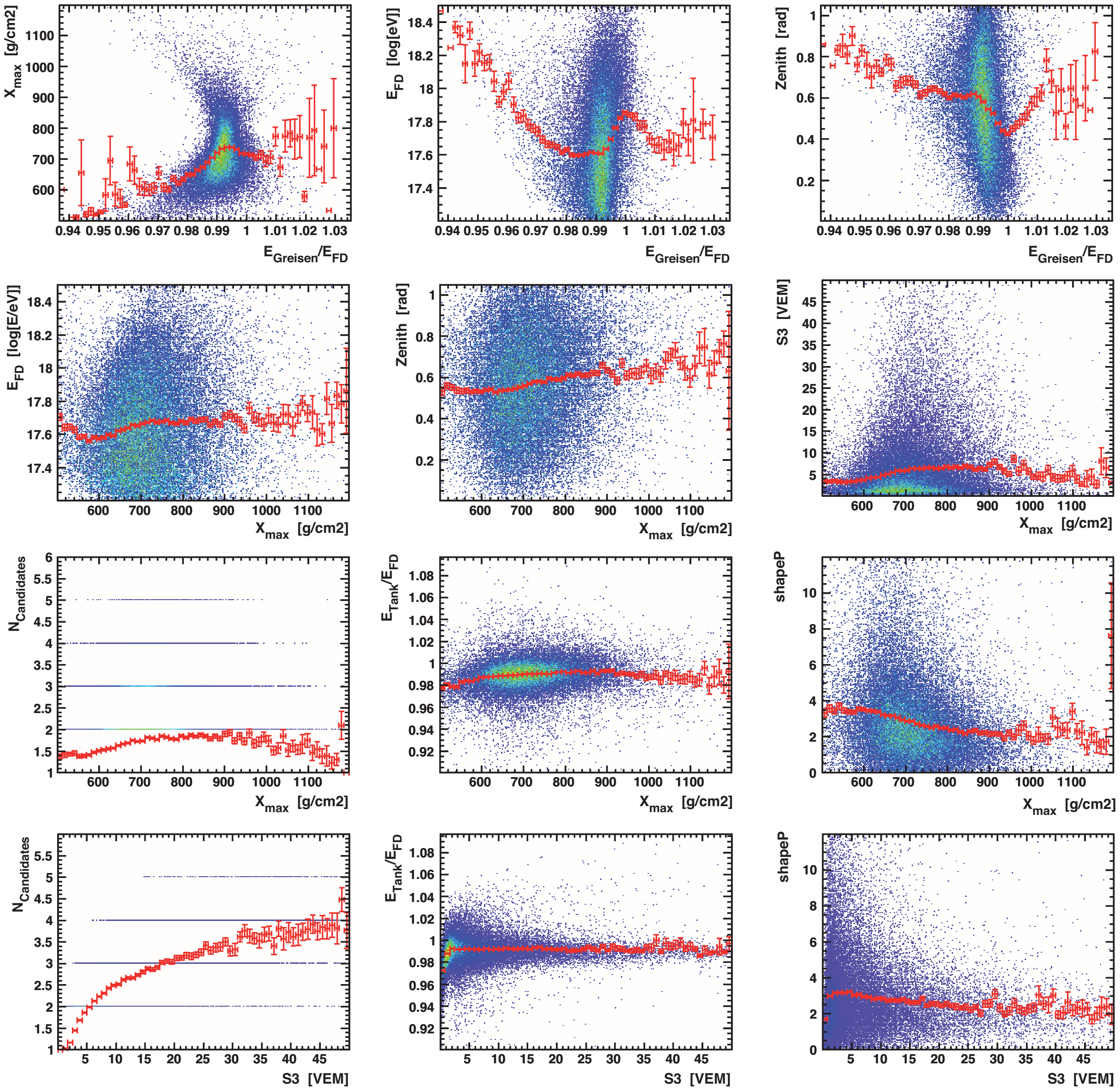


Figure C.2: Scatter plot of various observables to identify possible correlations. Each dot represents one event while the red markers denote the profile including statistical errors. Remaining combinations can be found in Fig. C.1 and Fig. C.3.

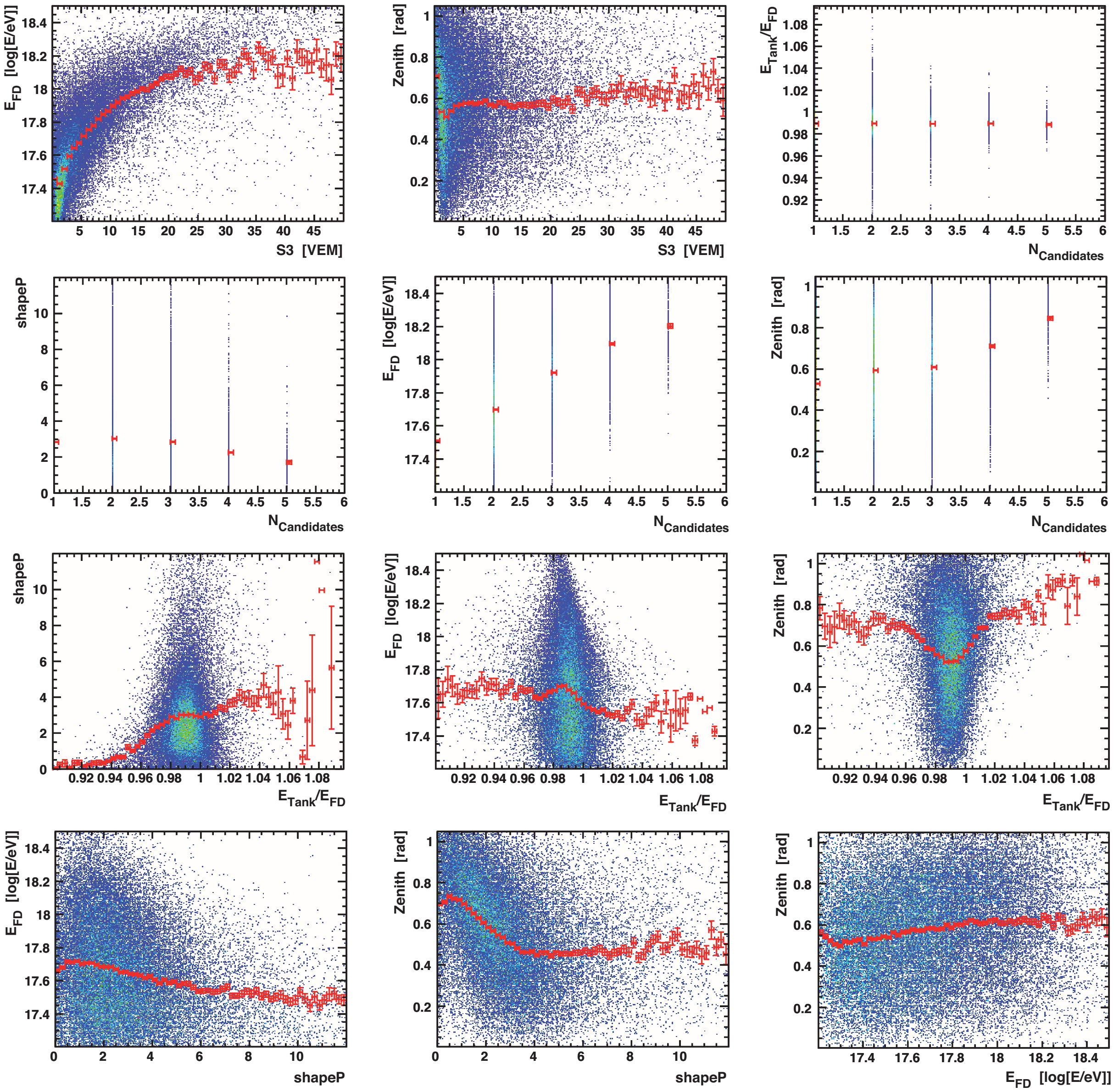


Figure C.3: Scatter plot of various observables to identify possible correlations. Each dot represents one event while the red markers denote the profile including statistical errors. Remaining combinations can be found in Fig. C.1 and Fig. C.2.

Glossary and Abbreviations

ADC
Analog to Digital Converter

ADST
Advanced Data Summary Tree

AERA
Auger Engineering Radio Array

AGN
Active Galactic Nuclei

AMADEUS
Antares Modules for Acoustic Detection Under the Sea

AMBER
Air-shower Microwave Bremsstrahlung Experimental Radiometer

AMIGA
Auger Muons and Infill for the Ground Array

ANN
Artificial Neural Networks

ANTARES
Astronomy with a Neutrino Telescope and Abyss environmental Research

APF
Aerosol Phase Function monitor

BDT
Boosted Decision Tree

BFGS
Broyden-Fletcher-Goldfarb-Shannon method

BLF
Ballon Launching Facility

BP
Back Propagation

Brass hybrid event
Showers that are triggered by the FD and one SD station are called "brass hybrid" events.

CCD
Charge-Coupled Devices

CDAS
Central Data Acquisition System

CFITSIO
CFITSIO is a machine-independent library of routines for reading and writing data files in the FITS data format. This library is written in `ANSI C` and provides a powerful interface for accessing FITS files.

CL
The confidence level (CL) is a particular kind of interval estimate of a population parameter and is used to indicate the reliability of an estimate. A common choice is CL = 0.9 which means that the result will meet the expectation in 90% of hypothetically repeated trials.

CLF
Central Laser Facility

CMB
Cosmic Microwave Background

CORSIKA
*CO*smic *R*ay *SI*mulations for *KA*scade

CR
Cosmic Ray

EAS
Extensive Air Shower

EBL
Extra-galactic Background Light

electron Volt (eV)
Equal to the amount of kinetic energy gained by a single unbound electron when it

accelerates through an electric potential difference of one volt. The energy is equal to approximately $1.602 \cdot 10^{-19}$ J. Commonly used abbreviations:

$$\begin{array}{rcl}
\mathrm{keV} & = & 10^{3}\ \mathrm{eV} \\
\mathrm{MeV} & = & 10^{6}\ \mathrm{eV} \\
\mathrm{GeV} & = & 10^{9}\ \mathrm{eV} \\
\mathrm{TeV} & = & 10^{12}\ \mathrm{eV} \\
\mathrm{PeV} & = & 10^{15}\ \mathrm{eV} \\
\mathrm{EeV} & = & 10^{18}\ \mathrm{eV} \\
\mathrm{ZeV} & = & 10^{21}\ \mathrm{eV}
\end{array}$$

EM

Electromagnetic

FADC

Flash Analog Digital Converter

Faraday Rotation Measures (FRM)

Discovered in 1845 by M. Faraday the Faraday effect was the first experimental evidence that light and electromagnetism are related. The basic principle is that the rotation of the plane of polarization is proportional to the intensity of the component of the applied magnetic field in the direction of the beam of light. In astronomy the effect is imposed on light over the course of its propagation from its origin to the Earth, through the interstellar medium.

FD

Fluorescence Detector

FITS

FITS or Flexible Image Transport System is a digital file format used to store, transmit, and manipulate scientific and other images. FITS is the most commonly used digital file format in astronomy. An advantage compared to other file formats is that FITS is designed specifically for scientific data and hence includes many provisions for describing photometric and spatial calibration information, together with image origin metadata.

FLT

First Level Trigger

FPGA

Field Programmable Gate Array

Golden hybrid event

Showers that are triggered by the FD and at least three SD stations are called “golden hybrid” events.

GPS
Global Positioning System

GRB
Gamma-Ray Burst

GZK effect
Greisen-Zatsepin-Kuzmin effect. A theory that galactic and extra-galactic photon fields make the Universe opaque for high energy protons.

HAM
Horizontal Attenuation Monitor

HEALPix
*H*ierarchical *E*qual *A*rea iso*L*atitude *Pix*elization

HEAT
High Elevation Auger Telescopes

HiRes
High Resolution Fly's Eye

ICRC
International Cosmic Ray Conference

IRB
Infrared Background

KNN
k-Nearest Neighbor classification

LDF
Lateral Distribution Function

LHC
Large Hadron Collider

LIDAR
Laser Illuminated Detection And Ranging system

LIV
Lorentz Invariance Violation

LPM effect
Landau-Pomeranchuk-Migdal effect. The basic principle is that the Bethe-Heitler cross-section σ_{BH} for pair production by photons can be reduced due to destructive interference from several scattering centers resulting in a reduction of σ_{BH}.

MBR
Molecular Bremsstrahlung Radiation

MC
Monte Carlo

MIDAS
Microwave Detection of Air Showers

MLP
Multilayer Perceptron

MVA
Multivariate Analysis

NSF
National Science Foundation

parsec (pc)
Parallax of one arcsecond is a unit of length. It is defined as the length of the adjacent side of an imaginary right triangle in space. The two dimensions that specify this triangle are the parallax angle (defined as 1 arcsecond) and the opposite side defined as 1 astronomical unit (AU), the average distance from the Earth to the Sun. One parsec is $\approx 30.857 \cdot 10^{15}$ m.

PMT
Photomultiplier Tube

RF
Radio Frequency

RMS
Root mean square. In this thesis RMS is defined as the standard deviation σ of the distribution. It is calculated as $\sqrt{1/N \cdot \sum_i (x_i - \overline{x})^2}$. This convention was introduced many years ago in `ROOT` and is kept for continuity.

ROC
Receiver Operating Characteristic diagram

ROOT
The ROOT system provides a set of object-orientated frameworks with all the functionality needed to handle and analyze large amounts of data in a very efficient way. Having the data defined as a set of objects, specialized storage methods are used to get direct access to the separate attributes of the selected objects, without having to touch the bulk of the data. These days ROOT has become a standard analysis tool in high energy astrophysics.

SD
Surface Detector

SDP
Abbreviation of Shower Detector Plane defined defined as the plane, containing the shower axis and the center of the eye.

SHDM
Super Heavy Dark Matter

SLT
Second Level Trigger

SNR
Supernova Remnant

TD
Topological Defects

TLT
Third Level Trigger

TMVA
Toolkit for Multivariate Analysis

UHE
Ultra-High Energy

UHECR
Abbreviation of Ultra-High Energy Cosmic Rays, generally cosmic rays above $\sim 10^{18}$ eV.

URB
Universal Radio Background

UTC
Coordinated Universal Time

VEM
Abbreviation of Vertical Equivalent Muons. Unit within the Pierre Auger Collaboration defined as the average charge signal produced in a surface detector tank by a penetrating down going muon in the vertical direction.

XLF
eXtreme Laser Facility

ZB
Z-burst model

Bibliography

[1] R. Clay and B. R. Dawson, *Cosmic bullets : high energy particles in astrophysics.* Allen & Unwin, St Leonards, N.S.W. 1997.

[2] V. F. Hess and W. Kolhoerster, "Über Beobachtungen der durchdringenden Strahlung bei sieben Freiballonfahrten," *Physikalische Zeitschrift* (1912).

[3] A. M. Hillas, *Cosmic rays.* The Commonwealth and International Library. Selected Readings in Physics, Oxford: Pergamon Press, 1972.

[4] J. Bluemer, R. Engel, and J. R. Hoerandel, "Cosmic Rays from the Knee to the Highest Energies," *Prog. Part. Nucl. Phys.* **63** (2009) 293–338, `0904.0725`.

[5] **ATIC-2** Collaboration, H. S. Ahn *et al.*, "ATIC experiment: Preliminary results from the flight in 2002,". Prepared for 28th International Cosmic Ray Conferences (ICRC 2003), Tsukuba, Japan, 31 Jul - 7 Aug 2003.

[6] N. L. Grigorov, V. E. Nesterov, I. D. Rapoport, I. A. Savenko, and G. A. Skuridin, "Investigation of energy spectrum of primary cosmic particles with high and superhigh energies of space stations *proton*," *Yad. Fiz.* **11** (1970) 1058–1069.

[7] **RUNJOB** Collaboration, V. A. Derbina *et al.*, "Cosmic-ray spectra and composition in the energy range of 10-TeV - 1000-TeV per particle obtained by the RUNJOB experiment," *Astrophys. J.* **628** (2005) L41–L44.

[8] **KASCADE** Collaboration, T. Antoni *et al.*, "KASCADE measurements of energy spectra for elemental groups of cosmic rays: Results and open problems," *Astropart. Phys.* **24** (2005) 1–25, `astro-ph/0505413`.

[9] **KASCADE-Grande** Collaboration, J. C. Arteaga-Velazquez *et al.*, "The constant intensity cut method applied to the KASCADE–Grande muon data," *Nucl. Phys. Proc. Suppl.* **196** (2009) 183–186.

[10] **Tibet ASγ** Collaboration, M. Amenomori *et al.*, "The all-particle spectrum of primary cosmic rays in the wide energy range from 10^{14} eV to 10^{17} eV observed with the Tibet-III air-shower array," *Astrophys. J.* **678** (2008) 1165–1179, `0801.1803`.

[11] **HiRes-MIA** Collaboration, T. Abu-Zayyad *et al.*, "Measurement of the cosmic ray energy spectrum and composition from 10^{17} eV to $10^{18.3}$ eV using a hybrid fluorescence technique," *Astrophys. J.* **557** (2001) 686–699, astro-ph/0010652.

[12] **HiRes** Collaboration, R. Abbasi *et al.*, "Observation of the GZK cutoff by the HiRes experiment," *Phys. Rev. Lett.* **100** (2008) 101101, astro-ph/0703099.

[13] **Pierre Auger** Collaboration, J. Abraham *et al.*, "Observation of the suppression of the flux of cosmic rays above 4×10^{19} eV," *Phys. Rev. Lett.* **101** (2008) 061101, 0806.4302.

[14] G. V. Kulikov and G. B. Khristiansen *JETP 35* (1958), no. 635,.

[15] **EAS-TOP** Collaboration, M. Aglietta *et al.*, "The cosmic ray primary composition in the "knee" region through the EAS electromagnetic and muon measurements at EAS-TOP," *Astropart. Phys.* **21** (2004) 583–596.

[16] **KASCADE** Collaboration, H. Ulrich *et al.*, "Indirect measurements around the knee: Recent results from KASCADE," *Int. J. Mod. Phys.* **A20** (2005) 6774–6777.

[17] **KASCADE-Grande** Collaboration, K.-H. Kampert *et al.*, "Cosmic Rays in the 'Knee'-Region — Recent Results from KASCADE —," *Acta Physica Polonica B* **35** (June, 2004) 1799–1812, arXiv:astro-ph/0405608.

[18] J. Linsley. Proceedings of 8th International Cosmic Ray Conference (1963), Jaipur (1963) 77.

[19] M. A. Lawrence, R. J. O. Reid, and A. A. Watson, "The Cosmic ray energy spectrum above $4{\times}10^{17}$ eV as measured by the Haverah Park array," *J. Phys.* **G17** (1991) 733–757.

[20] M. Nagano *et al.*, "Energy spectrum of primary cosmic rays above 10^{17} eV determined from the extensive air shower experiment at Akeno," *J. Phys.* **G18** (1992) 423–442.

[21] **HiRes** Collaboration, D. J. Bird *et al.*, "Evidence for correlated changes in the spectrum and composition of cosmic rays at extremely high-energies," *Phys. Rev. Lett.* **71** (1993) 3401–3404.

[22] **Pierre Auger** Collaboration, J. Abraham *et al.*, "The Fluorescence Detector of the Pierre Auger Observatory," 0907.4282.

[23] A. M. Hillas, "The origin of ultrahigh-energy cosmic rays," *Ann. Rev. Astron. Astrophys.* **22** (1984) 425–444.

[24] V. Berezinsky, A. Gazizov, and S. Grigorieva, "Propagation and Signatures of Ultra High Energy Cosmic Rays," *Nuclear Physics B Proceedings Supplements* **136** (Nov., 2004) 147–158, astro-ph/0410650.

[25] **Pierre Auger** Collaboration, J. Abraham *et al.*, "Measurement of the energy spectrum of cosmic rays above 10^{18} eV using the Pierre Auger Observatory," *Phys. Lett.* **B685** (2010) 239–246, 1002.1975.

[26] R. U. Abbasi *et al.*, "Measurement of the Flux of Ultra High Energy Cosmic Rays by the Stereo Technique," *Astropart. Phys.* **32** (2009) 53–60, 0904.4500.

[27] R. Aloisio, V. Berezinsky, and A. Gazizov, "Ultra High Energy Cosmic Rays: The disappointing model," *Astropart. Phys.* **34** (2011) 620626, 0907.5194.

[28] M. V. Zobeck, *Handbook of Space Astronomy and Astrophysics.* 2nd edition, Cambridge University Press, 1990.

[29] J. J. Engelmann, P. Ferrando, A. Soutoul, P. Goret, and E. Juliusson, "Charge composition and energy spectra of cosmic-ray for elements from Be to NI - Results from HEAO-3-C2," *Astron. Astrophys.* **233** (1990) 96–111.

[30] **AMS** Collaboration, J. Alcaraz *et al.*, "Cosmic protons," *Phys. Lett.* **B490** (2000) 27–35.

[31] **AMS** Collaboration, J. Alcaraz *et al.*, "Helium in near Earth orbit," *Phys. Lett.* **B494** (2000) 193–202.

[32] T. Sanuki *et al.*, "Precise measurement of cosmic-ray proton and helium spectra with the BESS spectrometer," *Astrophys. J.* **545** (2000) 1135, astro-ph/0002481.

[33] T. K. Gaisser and T. Stanev, "High-energy cosmic rays," *Nuclear Physics A* **777** (Oct., 2006) 98–110, astro-ph/0510321.

[34] T. K. Gaisser, *Cosmic rays and particle physics.* Cambridge and New York, Cambridge University Press, 292 p., 1990.

[35] **Pierre Auger** Collaboration, M. Unger, "Study of the Cosmic Ray Composition above 0.4 EeV using the Longitudinal Profiles of Showers observed at the Pierre Auger Observatory," *Proc of 30th Int. Cosmic Ray Conf., Merida* (2007) 0706.1495.

[36] J. W. Fowler *et al.*, "A measurement of the cosmic ray spectrum and composition at the knee," *Astropart. Phys.* **15** (2001) 49–64, astro-ph/0003190.

[37] **CACTI** Collaboration, S. Paling *et al. Proc of 25th Int. Cosmic Ray Conf., Durban* (1997) 253.

[38] S. P. Swordy and D. B. Kieda, "Elemental composition of cosmic rays near the knee by multiparameter measurements of air showers," *Astropart. Phys.* **13** (2000) 137–150, astro-ph/9909381.

[39] **HIRES** Collaboration, D. J. Bird *et al.*, “The Cosmic ray energy spectrum observed by the Fly’s Eye,” *Astrophys. J.* **424** (1994) 491–502.

[40] A. A. Watson, “Ultra-high-energy cosmic rays: The experimental situation,” *Phys. Rept.* **333** (2000) 309–327.

[41] **HEGRA** Collaboration, F. Arqueros *et al.*, “Energy spectrum and chemical composition of cosmic rays between 0.3 PeV and 10 PeV determined from the Cherenkov- light and charged-particle distributions in air showers,” *Astron. Astrophys.* **359** (2000) 682–694, `astro-ph/9908202`.

[42] T. Abu-Zayyad *et al.*, “Evidence for Changing of Cosmic Ray Composition between 10^{17} and 10^{18} eV from Multicomponent Measurements,” *Phys. Rev. Lett.* **84** (2000) 4276–4279.

[43] **HiRes** Collaboration, R. U. Abbasi *et al.*, “A Study of the Composition of Ultra High Energy Cosmic Rays Using the High Resolution Fly’s Eye,” *Astrophys. J.* **622** (2005) 910–926, `astro-ph/0407622`.

[44] M. Cha *et al. Proc of 27th Int. Cosmic Ray Conf., Hamburg* (2001) 132.

[45] J. Dickinson *et al. Proc of 26th Int. Cosmic Ray Conf., Salt Lake City* (1999) 136.

[46] D. Chernov *et al.*, “Primary energy spectrum and mass composition determined with the Tunka EAS Cherenkov Array,” *Int. J. Mod. Phys.* **A20** (2006) 6799–6801, `astro-ph/0411139`.

[47] S. Knurenko *et al. Proc of 27th Int. Cosmic Ray Conf., Hamburg* (2001) 177.

[48] N. N. Kalmykov and S. S. Ostapchenko, “Comparison of characteristics of the nucleus nucleus interaction in the model of quark-gluon strings and in the superposition model,” *Sov. J. Nucl. Phys.* **50** (1989) 315–318.

[49] T. Pierog and K. Werner, “The hadronic interaction model EPOS and air shower simulations: New results on muon production,”. Prepared for 30th International Cosmic Ray Conference (ICRC 2007), Merida, Yucatan, Mexico, 3-11 Jul 2007.

[50] S. Ostapchenko, “Non-linear screening effects in high energy hadronic interactions,” *Phys. Rev.* **D74** (2006) 014026, `hep-ph/0505259`.

[51] R. Engel, T. K. Gaisser, T. Stanev, and P. Lipari, “Air shower calculations with the new version of SIBYLL,”. Prepared for 26th International Cosmic Ray Conference (ICRC 99), Salt Lake City, Utah, 17-25 Aug 1999.

[52] **KASCADE** Collaboration, W. D. Apel *et al.*, “Energy Spectra of Elemental Groups of Cosmic Rays: Update on the KASCADE Unfolding Analysis,” *Astropart. Phys.* **31** (2009) 86–91, `0812.0322`.

[53] **Pierre Auger** Collaboration, J. Abraham *et al.*, "Measurement of the Depth of Maximum of Extensive Air Showers above 10^{18} eV," *Phys. Rev. Lett.* **104** (2010) 091101, `1002.0699`.

[54] N. N. Kalmykov and S. S. Ostapchenko, "The Nucleus-nucleus interaction, nuclear fragmentation, and fluctuations of extensive air showers," *Phys. Atom. Nucl.* **56** (1993) 346–353.

[55] S. Ostapchenko, "QGSJET-II: Towards reliable description of very high energy hadronic interactions," *Nucl. Phys. Proc. Suppl.* **151** (2006) 143–146, `hep-ph/0412332`.

[56] T. Pierog and K. Werner, "Muon Production in Extended Air Shower Simulations," *Phys. Rev. Lett.* **101** (2008) 171101, `astro-ph/0611311`.

[57] E.-J. Ahn, R. Engel, T. K. Gaisser, P. Lipari, and T. Stanev, "Cosmic ray interaction event generator SIBYLL 2.1," *Phys. Rev.* **D80** (2009) 094003, `0906.4113`.

[58] **Pierre Auger** Collaboration, J. Abraham *et al.*, "Correlation of the highest energy cosmic rays with nearby extragalactic objects," *Science* **318** (2007) 938–943, `0711.2256`.

[59] **Pierre Auger** Collaboration, J. Abraham *et al.*, "Correlation of the highest-energy cosmic rays with the positions of nearby active galactic nuclei," *Astropart. Phys.* **29** (2008) 188–204, `0712.2843`.

[60] **HiRes** Collaboration, R. U. Abbasi *et al.*, "Indications of Proton-Dominated Cosmic Ray Composition above 1.6 EeV," *Phys. Rev. Lett.* **104** (2010) 161101, `0910.4184`.

[61] M. P. Veron-Cetty and P. Veron, "A catalogue of quasars and active nuclei: 12th edition," *Astron. Astrophys.* **455** (2006) 773–777.

[62] R. U. Abbasi *et al.*, "Search for Correlations between HiRes Stereo Events and Active Galactic Nuclei," *Astropart. Phys.* **30** (2008) 175–179, `0804.0382`.

[63] **AGASA** Collaboration, N. Hayashida *et al.*, "The anisotropy of cosmic ray arrival directions around 10^{18} eV," *Astropart. Phys.* **10** (1999) 303–311, `astro-ph/9807045`.

[64] J. A. Bellido, R. W. Clay, B. R. Dawson, and M. Johnston-Hollitt, "Southern hemisphere observations of a 10^{18} eV cosmic ray source near the direction of the galactic centre," *Astropart. Phys.* **15** (2001) 167–175, `astro-ph/0009039`.

[65] **Pierre Auger** Collaboration, M. Aglietta *et al.*, "Anisotropy studies around the galactic centre at EeV energies with the Auger observatory," *Astropart. Phys.* **27** (2007) 244–253, `astro-ph/0607382`.

[66] **Pierre Auger** Collaboration, P. Abreu *et al.*, "Update on the correlation of the highest energy cosmic rays with nearby extragalactic matter," `1009.1855`.

[67] T. Kashti and E. Waxman, "Searching for a Correlation Between Cosmic-Ray Sources Above 10^{19} eV and Large-Scale Structure," *JCAP* **0805** (2008) 006, `0801.4516`.

[68] P. Sommers and S. Westerhoff, "Cosmic Ray Astronomy," *New J. Phys.* **11** (2009) 055004, `0802.1267`.

[69] J. Linsley, "Evidence for a primary cosmic-ray particle with energy 10^{20} eV," *Phys. Rev. Lett.* **10** (1963) 146–148.

[70] D. J. Bird *et al.*, "Detection of a cosmic ray with measured energy well beyond the expected spectral cutoff due to cosmic microwave radiation," *Astrophys. J.* **441** (1995) 144–150.

[71] P. Bhattacharjee and G. Sigl, "Origin and propagation of extremely high energy cosmic rays," *Phys. Rept.* **327** (2000) 109–247, `astro-ph/9811011`.

[72] C. T. Hill, "Monopolonium," *Nucl. Phys.* **B224** (1983) 469.

[73] M. B. Hindmarsh and T. W. B. Kibble, "Cosmic strings," *Rept. Prog. Phys.* **58** (1995) 477–562, `hep-ph/9411342`.

[74] V. Berezinsky, M. Kachelriess, and A. Vilenkin, "Ultra-high energy cosmic rays without GZK cutoff," *Phys. Rev. Lett.* **79** (1997) 4302–4305, `astro-ph/9708217`.

[75] M. Birkel and S. Sarkar, "Extremely high energy cosmic rays from relic particle decays," *Astropart. Phys.* **9** (1998) 297–309, `hep-ph/9804285`.

[76] Z. Fodor and S. D. Katz, "Grand unification signal from ultrahigh-energy cosmic rays?," *Phys. Rev. Lett.* **86** (2001) 3224–3227, `hep-ph/0008204`.

[77] T. J. Weiler, "Resonant Absorption of Cosmic Ray Neutrinos by the Relic Neutrino Background," *Phys. Rev. Lett.* **49** (1982) 234.

[78] T. J. Weiler, "Cosmic ray neutrino annihilation on relic neutrinos revisited: A mechanism for generating air showers above the Greisen-Zatsepin-Kuzmin cut-off," *Astropart. Phys.* **11** (1999) 303–316, `hep-ph/9710431`.

[79] D. Fargion, B. Mele, and A. Salis, "Ultrahigh energy neutrino scattering onto relic light neutrinos in galactic halo as a possible source of highest energy extragalactic cosmic rays," *Astrophys. J.* **517** (1999) 725–733, `astro-ph/9710029`.

[80] **Pierre Auger** Collaboration, J. Abraham *et al.*, "An upper limit to the photon fraction in cosmic rays above 10^{19} eV from the Pierre Auger Observatory," *Astropart. Phys.* **27** (2007) 155–168, `astro-ph/0606619`.

[81] **Pierre Auger** Collaboration, J. Abraham *et al.*, "Upper limit on the cosmic-ray photon flux above 10^{19} eV using the surface detector of the Pierre Auger Observatory," *Astropart. Phys.* **29** (2008) 243–256, `0712.1147`.

[82] **The Pierre Auger** Collaboration, J. Abraham *et al.*, "Upper limit on the cosmic-ray photon fraction at EeV energies from the Pierre Auger Observatory," *Astropart. Phys.* **31** (2009) 399–406, `0903.1127`.

[83] W. F. Swann, "A Mechanism of Acquirement of Cosmic-Ray Energies by Electrons," *Phys. Rev.* **43** (Feb., 1933) 217–220.

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

[85] R. D. Blandford and J. P. Ostriker, "Particle Acceleration by Astrophysical Shocks," *Astrophys. J.* **221** (1978) L29–L32.

[86] P. Sokolsky, "Introduction to ultrahigh-energy Cosmic Ray physics,". Redwood City, USA: Addison-Wesley (1989) 208p.

[87] D. F. Torres and L. A. Anchordoqui, "Astrophysical origins of ultrahigh energy cosmic rays," *Rept. Prog. Phys.* **67** (2004) 1663–1730, `astro-ph/0402371`.

[88] F. Halzen and E. Zas, "Neutrino fluxes from active galaxies: A model-independent estimate," *Astrophys. J.* **488** (1997) 669–674, `astro-ph/9702193`.

[89] E. Waxman, "Cosmological origin for cosmic rays above 10^{19} eV," *Astrophys. J.* **452** (1995) L1–L4, `astro-ph/9508037`.

[90] M. Vietri, "On the acceleration of ultrahigh-energy cosmic rays in gamma-ray bursts," *Astrophys. J.* **453** (1995) 883–889, `astro-ph/9506081`.

[91] H. Kang, D. Ryu, and T. W. Jones, "Cluster accretion shocks as possible acceleration sites for ultrahigh-energy protons below the Greisen cutoff," *Astrophys. J.* **456** (1996) 422–427, `astro-ph/9507113`.

[92] P. Blasi, R. I. Epstein, and A. V. Olinto, "Ultra-high energy cosmic rays from young neutron star winds," *Astrophys. J.* **533** (2000) L123, `astro-ph/9912240`.

[93] T. Stanev, "Ultra High Energy Cosmic Rays: origin and propagation," `0711.2282`.

[94] J. P. Rachen and P. L. Biermann, "Extragalactic ultrahigh-energy cosmic rays. 1. Contribution from hot spots in FR-II radio galaxies," *Astron. Astrophys.* **272** (1993) 161–175, `astro-ph/9301010`.

[95] P. P. Kronberg, "Extragalactic magnetic fields," *Rept. Prog. Phys.* **57** (1994) 325–382.

[96] T. Stanev, D. Seckel, and R. Engel, "Propagation of ultra-high energy protons in regular extragalactic magnetic fields," *Phys. Rev.* **D68** (2003) 103004, `astro-ph/0108338`.

[97] T. Stanev, "Ultra high energy cosmic rays and the large scale structure of the galactic magnetic field," *Astrophys. J.* **479** (1997) 290, `astro-ph/9607086`.

[98] S. Lee, "Propagation of extragalactic high energy cosmic and γ rays," *Phys. Rev. D* **58** (Jul, 1998) 043004.

[99] E. Armengaud, G. Sigl, T. Beau, and F. Miniati, "CRPropa : A numerical tool for the propagation of UHE cosmic rays, gamma-rays and neutrinos," *Astropart. Phys.* **28** (2007) 463–471, `astro-ph/0603675`.

[100] A. A. Penzias and R. W. Wilson, "A Measurement of excess antenna temperature at 4080-Mc/s," *Astrophys. J.* **142** (1965) 419–421.

[101] K. Greisen, "End to the cosmic ray spectrum?," *Phys. Rev. Lett.* **16** (1966) 748–750.

[102] G. T. Zatsepin and V. A. Kuzmin, "Upper limit of the spectrum of cosmic rays," *JETP Lett.* **4** (1966) 78–80.

[103] J. Linsley, "Evidence for a primary cosmic-ray particle with energy 10^{20}-eV," *Phys. Rev. Lett.* **10** (1963) 146–148.

[104] M. M. Winn, J. Ulrichs, L. S. Peak, C. B. A. Mccusker, and L. Horton, "The Cosmic Ray Energy Spectrum Above 10^{17} eV," *J. Phys.* **G12** (1986) 653–674.

[105] V. P. Egorova *et al.*, "The spectrum features of UHECRs below and surrounding GZK," *Nucl. Phys. Proc. Suppl.* **136** (2004) 3–11, `astro-ph/0408493`.

[106] M. Takeda *et al.*, "Energy determination in the Akeno Giant Air Shower Array experiment,". Prepared for 28th International Cosmic Ray Conferences (ICRC 2003), Tsukuba, Japan, 31 Jul - 7 Aug 2003.

[107] **HiRes** Collaboration, D. R. Bergman, "Observation of the GZK cutoff using the HiRes detector," `astro-ph/0609453`.

[108] G. R. Blumenthal, "Energy loss of high-energy cosmic rays in pair-producing collisions with ambient photons," *Phys. Rev.* **D1** (1970) 1596–1602.

[109] P. Auger, P. Ehrenfest, R. Maze, J. Daudin, and A. F. Robley, "Extensive cosmic-ray showers," *Rev. Mod. Phys.* **11** (1939) 288–291.

[110] T. Bergmann *et al.*, "One-dimensional hybrid approach to extensive air shower simulation," *Astropart. Phys.* **26** (2007) 420–432, `astro-ph/0606564`.

[111] W. Heitler, *Quantum Theory of Radiation.* Oxford University Press, 1944.

[112] D. Heck, G. Schatz, T. Thouw, J. Knapp, and J. N. Capdevielle, "CORSIKA: A Monte Carlo code to simulate extensive air showers,". FZKA-6019, Forschungszentrum Karlsruhe. 1998.

[113] D. Heck and T. Pierog, "Extensive Air Shower Simulation with CORSIKA: A User's Guide,". 2010.

[114] F. Schmidt, "CORSIKA Shower Images,". http://www.ast.leeds.ac.uk/∼fs/showerimages.html, July 2010.

[115] T. Gaisser and A. M. Hillas *Proc. 15th ICRC, Plovdiv, Bulgaria* **8** (1977) 353.

[116] B. Rossi and K. Greisen, "Cosmic-ray theory," *Rev. Mod. Phys.* **13** (1941) 240–309.

[117] K. Greisen *Progress in Cosmic Ray Physics* **3** (1965) 17. Ed: Wilson, J.G., North-Holland, Amsterdam.

[118] S. J. Sciutto, "Air shower simulations with the AIRES system," `astro-ph/9905185`.

[119] J. Linsley, "The structure function of EAS measured at Volcano Ranch. (Talk),". To appear in the proceedings of 13th International Cosmic Ray Conference (ICRC 1973), Denver, Colorado, 17-30 Aug 1973.

[120] M. T. Dova, M. E. Mancenido, A. G. Mariazzi, T. P. McCauley, and A. A. Watson, "A reinterpretation of Volcano Ranch lateral distribution measurements to infer the mass composition of cosmic rays," *Nucl. Phys. Proc. Suppl.* **122** (2003) 235–238, `astro-ph/0210464`.

[121] R. N. Coy, G. Cunningham, C. L. Pryke, and A. A. Watson, "The lateral distribution of extensive air showers produced by cosmic rays above 10^{19} eV as measured by water-Čerenkov detectors," *Astroparticle Physics* **6** (Mar., 1997) 263–270.

[122] A. N. Bunner *Sky and Telescope* **34** (1967) 204.

[123] R. M. Baltrusaitis *et al.*, "The Utah Fly's Eye Detector," *Nucl. Instrum. Meth.* **A240** (1985) 410–428.

[124] D. Kuempel, "Geometry Reconstruction of Fluorescence Detectors Revisited,". Diploma thesis, WU D 07-13, Wuppertal, August 2007.

[125] M. S. Longair, *High energy astrophysics. an informal introduction for students of physics and astronomy.* Cambridge: Cambridge University Press, 1981.

[126] D. J. Bird *et al.*, “Detection of a cosmic ray with measured energy well beyond the expected spectral cutoff due to cosmic microwave radiation,” *Astrophys. J.* **441** (1995) 144–150.

[127] F. D. Kahn and I. Lerche, “Radiation from Cosmic Ray Air Showers,” *Royal Society of London Proceedings Series A* **289** (Jan., 1966) 206–213.

[128] J. Jelley *et al.*, “Radio Pulses from Extensive Cosmic-Ray Air Showers,” *Nature* **205** (1965) 327.

[129] H. Falcke and P. Gorham, “Detecting radio emission from cosmic ray air showers and neutrinos with a digital radio telescope,” *Astropart. Phys.* **19** (2003) 477–494, `astro-ph/0207226`.

[130] T. Huege and H. Falcke, “Radio-emission from cosmic ray air-showers: A theoretical perspective for LOPES,” `astro-ph/0207647`.

[131] **LOPES** Collaboration, H. Falcke *et al.*, “Detection and imaging of atmospheric radio flashes from cosmic ray air showers,” *Nature* **435** (2005) 313–316, `astro-ph/0505383`.

[132] **LOPES** Collaboration, T. Huege *et al.*, “The LOPES experiment - recent results, status and perspectives,” `1009.0345`.

[133] D. Ardouin *et al.*, “Radio-detection signature of high-energy cosmic rays by the CODALEMA experiment,” *Nucl. Instrum. Meth.* **A555** (2005) 148–163.

[134] **LOPES** Collaboration, T. Huege *et al.*, “Radio detection of cosmic ray air showers with LOPES,” *Braz. J. Phys.* **36** (2006) 1157–1164, `astro-ph/0609427`.

[135] W. D. Apel *et al.*, “Lateral Distribution of the Radio Signal in Extensive Air Showers Measured with LOPES,” *Astropart. Phys.* **32** (2010) 294–303, `0910.4866`.

[136] AERA, “AERA proposal for the construction of the 20 km^2 Auger Engineering Radio Array at the Southern Auger Observatory,” *Internal Auger GAP-note*. GAP-2009-172.

[137] T. Huege *et al.*, “Radio detection of cosmic rays in the Pierre Auger Observatory,” *Nucl. Instrum. Meth.* **A617** (2010) 484–487, `0906.4970`.

[138] S. Fliescher *et al.*, “The radio extension of Auger $\overline{\text{Off}}\underline{\text{line}}$,” *Internal Auger GAP-note*. GAP-2010-056.

[139] P. W. Gorham, “On the possibility of radar echo detection of ultra-high energy cosmic ray and neutrino induced extensive air showers,” *Astropart. Phys.* **15** (2001) 177–202, `hep-ex/0001041`.

[140] C. Williams *et al.*, "The MIDAS Experiment: A New Technique for the Detection of Extensive Air Showers," *Presented at the XVI International Symposium on Very High Energy Cosmic Ray Interactions (ISVHECRI 2010), Batavia, IL, USA* (2010) `1010.2734`.

[141] P. W. Gorham *et al.*, "Observations of Microwave Continuum Emission from Air Shower Plasmas," *Phys. Rev.* **D78** (2008) 032007, `0705.2589`.

[142] G. A. Askarian *Atomnaya Energiya.* **3** (1957) 152.

[143] G. A. Askarian, B. A. Dolgoshein, A. N. Kalinovsky, and N. V. Mokhov, "Acoustic detection of high-energy particle showers in water," *Nucl. Instrum. Meth.* **164** (1979) 267–278.

[144] J. G. Learned, "Acoustic radiation by charged atomic particles in liquids: An analysis," *Phys. Rev.* **D19** (1979) 3293.

[145] A. V. Butkevich *et al.*, "Prospects for radio-wave and acoustic detection of ultra- and superhigh-energy cosmic neutrinos (cross sections, signals, thresholds)," *Phys. Part. Nucl.* **29** (1998) 266–272.

[146] G. Sigl, "Probing physics and astrophysics at extreme energies with ultra high energy cosmic radiation," *AIP Conf. Proc.* **579** (2001) 32–42.

[147] **ANTARES** Collaboration, K. Graf *et al.*, "Studies of Acoustic Neutrino Detection Methods with ANTARES," `1005.3951`.

[148] **ANTARES** Collaboration, J. A. Aguilar *et al.*, "AMADEUS - The Acoustic Neutrino Detection Test System of the ANTARES Deep-Sea Neutrino Telescope," `1009.4179`.

[149] **ANTARES** Collaboration, R. Lahmann *et al.*, "Status and First Results of the Acoustic Detection Test System AMADEUS," *Nucl. Instrum. Meth.* **A604** (2009) S158–S163, `0901.0321`.

[150] **Baikal** Collaboration, V. Aynutdinov *et al.*, "Acoustic search for high-energy neutrinos in Lake Baikal: status and perspectives," `0910.0678`.

[151] **IceCube** Collaboration, R. Abbasi *et al.*, "Measurement of Acoustic Attenuation in South Pole Ice," `1004.1694`.

[152] **Pierre Auger** Collaboration, J. Abraham *et al.*, "The Pierre Auger Observatory design report,". Second Edition, 1997, (http://www.auger.org/technical_info/design_report.html).

[153] **Pierre Auger** Collaboration, J. Abraham *et al.*, "The Pierre Auger project design report,". FERMILAB-PUB-96-024.

[154] **Pierre Auger** Collaboration, J. Bluemer *et al.*, “The northern site of the Pierre Auger Observatory,” *New Journal of Physics* **12** (2010), no. 3, 035001.

[155] K.-H. Kampert. private communication, July 2010.

[156] Google Earth, July 2010.

[157] J. Bluemer, “The Pierre Auger Observatory: Cosmic accelerators and the most energetic particles in the Universe,”. Talk given at the IPAC10 conference, Kyoto, Japan (May 2010).

[158] Z. Szadkowski, K.-H. Becker, and K.-H. Kampert, “Development of a new first level trigger for the surface array in the Pierre Auger Observatory based on the Cyclone Altera FPGA,” *Nucl. Instrum. Meth.* **A545** (2005) 793–802.

[159] **Pierre Auger** Collaboration, T. Suomijarvi, “The surface detectors of the Pierre Auger Observatory,” *Nucl. Phys. Proc. Suppl.* **136** (2004) 393–398.

[160] **Pierre Auger** Collaboration, X. Bertou, “Performance of the Pierre Auger Observatory surface array,” `astro-ph/0508466`.

[161] F. Kakimoto *et al.*, “A Measurement of the air fluorescence yield,” *Nucl. Instrum. Meth.* **A372** (1996) 527–533.

[162] **Pierre Auger** Collaboration, R. Sato and C. O. Escobar, “The performance of the corrector lenses for the Auger fluorescence detector,”. Presented at 29th International Cosmic Ray Conference (ICRC 2005), Pune, India, 3-11 Aug 2005.

[163] **Pierre Auger** Collaboration, J. A. Bellido, “Performance of the fluorescence detectors of the Pierre Auger Observatory,” `astro-ph/0508389`.

[164] K.-H. Kampert, “The Pierre Auger observatory: Status and prospects,” *Nucl. Phys. Proc. Suppl.* **151** (2006) 393–400, `astro-ph/0501074`.

[165] **GEANT4** Collaboration, S. Agostinelli *et al.*, “GEANT4: A simulation toolkit,” *Nucl. Instrum. Meth.* **A506** (2003) 250–303.

[166] J. Allison *et al.*, “Geant4 developments and applications,” *IEEE Trans. Nucl. Sci.* **53** (2006) 270.

[167] Z. Szadkowski, “Second-level trigger in the Pierre Auger fluorescence detector,” *Nucl. Instrum. Meth.* **A465** (2001) 540–549.

[168] **Pierre Auger** Collaboration, P. Bauleo *et al.*, “Absolute calibration of the Auger fluorescence detectors,” `astro-ph/0507347`.

[169] **Pierre Auger** Collaboration, R. Knapik *et al.*, “The Absolute, Relative and Multi-Wavelength Calibration of the Pierre Auger Observatory Fluorescence Detectors,” `0708.1924`.

[170] **Pierre Auger** Collaboration, C. Aramo *et al.*, “Optical relative calibration and stability monitoring for the Auger fluorescence detector,” `astro-ph/0507577`.

[171] **Pierre Auger** Collaboration, M. A. Mostafa, “Atmospheric monitoring for the Pierre Auger fluorescence detector,” `astro-ph/0308442`.

[172] S. Argiro *et al.*, “The Offline Software Framework of the Pierre Auger Observatory,” *Nucl. Instrum. Meth.* **A580** (2007) 1485–1496, `0707.1652`.

[173] L. Prado *et al.*, “Simulation of the fluorescence detector of the Pierre Auger Observatory,” *Nucl. Instrum. Meth.* **A545** (2005) 632–642.

[174] J. A. Bellido and V. Souza, “HybridGeometryFinder.cc,”. Offline version v2r6p4-branch_AsterixObelix, line 205.

[175] OFFLINE, “FdSDPFinder.cc,” *2.2p4-mojo-svn_trunk, Line 848.*

[176] D. Kuempel, K.-H. Kampert, and M. Risse, “Geometry reconstruction of fluorescence detectors revisited,” *Astropart. Phys.* **30** (2008) 167–174, `0806.4523`.

[177] D. Kuempel, K.-H. Kampert, and M. Risse, “Geometry reconstruction of fluorescence detectors revisited,”. Proceedings of the 21st European Cosmic Ray Symposium (ECRS), Kosice (Slovakia), 9-12 September 2008, P6.22.

[178] M. Mostafa, “Hybrid activities of the Pierre Auger Observatory,” *Nucl. Phys. Proc. Suppl.* **165** (2007) 50–58, `astro-ph/0608670`.

[179] D. Newton, J. Knapp, and A. A. Watson, “The optimum distance at which to determine the size of a giant air shower,” *Astropart. Phys.* **26** (2007) 414–419, `astro-ph/0608118`.

[180] **Pierre Auger** Collaboration, M. D. Healy *et al.*, “Composition-sensitive parameters measured with the surface detector of the Pierre Auger Observatory,” `arXiv:0706.1569 [astro-ph]`.

[181] **Pierre Auger** Collaboration, B. R. Dawson *et al.*, “Hybrid Performance of the Pierre Auger Observatory,” `arXiv:0706.1105 [astro-ph]`.

[182] T. C. Weekes, “Revealing the dark TeV sky: The atmospheric Cherenkov imaging technique for very high energy gamma-ray astronomy,” `astro-ph/0606130`.

[183] M. Risse and P. Homola, “Search for ultra-high energy photons using air showers,” *Mod. Phys. Lett.* **A22** (2007) 749–766, `astro-ph/0702632`.

[184] G. Gelmini, O. Kalashev, and D. V. Semikoz, “GZK Photons as Ultra High Energy Cosmic Rays,” *J. Exp. Theor. Phys.* **106** (2008) 1061–1082, `astro-ph/0506128`.

[185] A. Kusenko, J. Schissel, and F. W. Stecker, "Interactions of ultrahigh-energy cosmic rays with photons in the galactic center," *Astropart. Phys.* **25** (2006) 242–245, `astro-ph/0508142`.

[186] G. Gelmini and A. Kusenko, "Unstable superheavy relic particles as a source of neutrinos responsible for the ultrahigh-energy cosmic rays," *Phys. Rev. Lett.* **84** (2000) 1378, `hep-ph/9908276`.

[187] T. M. Kneiske, J. Kulbartz, D. Horns, and G. Sigl, "Diffuse Gamma-Ray and Neutrino Emission from the Local Supercluster,". Proceedings of the 31st International Cosmic Ray Conference (ICRC 2009), Lodz, Poland, 7-15 Jul. 2009.

[188] I. V. Moskalenko, L. Stawarz, T. A. Porter, and C. C. Cheung, "On the Possible Association of Ultra High Energy Cosmic Rays with Nearby Active Galaxies," *Astrophys. J.* **693** (2009) 1261–1274, `0805.1260`.

[189] F. P. Israel, "Centaurus A - NGC 5128," `astro-ph/9811051`.

[190] R. W. Clay, B. J. Whelan, and P. G. Edwards, "Centaurus A at Ultra-High Energies," `1001.0813`.

[191] S. Sarkar, "New physics from ultrahigh energy cosmic rays," *Acta Phys. Polon.* **B35** (2004) 351–364, `hep-ph/0312223`.

[192] R. Aloisio, V. Berezinsky, and M. Kachelriess, "On the status of superheavy dark matter," *Phys. Rev.* **D74** (2006) 023516, `astro-ph/0604311`.

[193] G. Sigl, "High energy astroparticle physics," `astro-ph/0612240`.

[194] P. Homola *et al.*, "Characteristics of geomagnetic cascading of ultra-high energy photons at the southern and northern sites of the Pierre Auger Observatory," *Astropart. Phys.* **27** (2007) 174–184, `astro-ph/0608101`.

[195] T. Erber, "High-energy electromagnetic conversion processes in intense magnetic fields," *Rev. Mod. Phys.* **38** (1966) 626–659.

[196] B. Mcbreen and C. J. Lambert, "Interactions of high-energy ($E > 5 \cdot 10^{19}$ eV) photons in the Earth's magnetic field," *Phys. Rev.* **D24** (1981) 2536–2538.

[197] H. Bethe and W. Heitler, "On the Stopping of fast particles and on the creation of positive electrons," *Proc. Roy. Soc. Lond.* **A146** (1934) 83–112.

[198] L. D. Landau and I. Pomeranchuk, "Limits of applicability of the theory of bremsstrahlung electrons and pair production at high-energies," *Dokl. Akad. Nauk Ser. Fiz.* **92** (1953) 535–536.

[199] L. D. Landau and I. Pomeranchuk, "Electron cascade process at very high-energies," *Dokl. Akad. Nauk Ser. Fiz.* **92** (1953) 735–738.

[200] A. B. Migdal, “Bremsstrahlung and pair production in condensed media at high-energies,” *Phys. Rev.* **103** (1956) 1811–1820.

[201] S. Klein, “Suppression of bremsstrahlung and pair production due to environmental factors,” *Rev. Mod. Phys.* **71** (1999) 1501–1538, `hep-ph/9802442`.

[202] X. Bertou, P. Billoir, and S. Dagoret-Campagne, “LPM effect and pair production in the geomagnetic field: a signature of ultra-high energy photons in the Pierre Auger Observatory,” *Astropart. Phys.* **14** (2000) 121–130.

[203] G. B. Gelmini, O. E. Kalashev, and D. V. Semikoz, “GZK Photons Above 10 EeV,” *JCAP* **0711** (2007) 002, `0706.2181`.

[204] K. Shinozaki *et al.*, “Upper limit on gamma-ray flux above 10**19-eV estimated by the Akeno Giant Air Shower Array experiment,” *Astrophys. J.* **571** (2002) L117–L120.

[205] M. Risse *et al.*, “Upper limit on the photon fraction in highest-energy cosmic rays from AGASA data,” *Phys. Rev. Lett.* **95** (2005) 171102, `astro-ph/0502418`.

[206] G. I. Rubtsov *et al.*, “Upper limit on the ultra-high-energy photon flux from AGASA and Yakutsk data,” *Phys. Rev.* **D73** (2006) 063009, `astro-ph/0601449`.

[207] A. V. Glushkov *et al.*, “Constraining the fraction of primary gamma rays at ultra-high energies from the muon data of the Yakutsk extensive- air-shower array,” *JETP Lett.* **85** (2007) 131–135, `astro-ph/0701245`.

[208] M. Ave, J. A. Hinton, R. A. Vazquez, A. A. Watson, and E. Zas, “New constraints from Haverah Park data on the photon and iron fluxes of UHE cosmic rays,” *Phys. Rev. Lett.* **85** (2000) 2244–2247, `astro-ph/0007386`.

[209] M. Ave, J. A. Hinton, R. A. Vazquez, A. A. Watson, and E. Zas, “Constraints on the ultra high energy photon flux using inclined showers from the Haverah Park array,” *Phys. Rev.* **D65** (2002) 063007, `astro-ph/0110613`.

[210] **Pierre Auger** Collaboration, J. Abraham *et al.*, “The Pierre Auger Northern Observatory Design Report 2009,”. draft version rev. 286, (2009).

[211] S. Lee, A. Olinto, and G. Sigl, “Extragalactic magnetic field and the highest energy cosmic rays,” *Astrophys. J.* **455** (1995) L21, `astro-ph/9508088`.

[212] M. Risse *et al.*, “Photon air showers at ultra-high energy and the photonuclear cross-section,” *Czech. J. Phys.* **56** (2006) A327–A336, `astro-ph/0512434`.

[213] T. C. Rogers and M. I. Strikman, “Hadronic interactions of ultra-high energy photons with protons and light nuclei in the dipole picture,” *J. Phys.* **G32** (2006) 2041–2063, `hep-ph/0512311`.

[214] S. R. Coleman and S. L. Glashow, “Evading the GZK cosmic-ray cutoff,” hep-ph/9808446.

[215] M. Galaverni and G. Sigl, “Lorentz Violation in the Photon Sector and Ultra-High Energy Cosmic Rays,” *Phys. Rev. Lett.* **100** (2008) 021102, 0708.1737.

[216] L. Maccione and S. Liberati, “GZK photon constraints on Planck scale Lorentz violation in QED,” *JCAP* **0808** (2008) 027, 0805.2548.

[217] M. Galaverni and G. Sigl, “Lorentz Violation and Ultrahigh-Energy Photons,” *Phys. Rev.* **D78** (2008) 063003, 0807.1210.

[218] F. R. Klinkhamer and M. Risse, “Ultra-high-energy cosmic-ray bounds on nonbirefringent modified-Maxwell theory,” *Phys. Rev.* **D77** (2008) 016002, 0709.2502.

[219] F. Halzen and D. Hooper, “High-energy neutrino astronomy: The cosmic ray connection,” *Rept. Prog. Phys.* **65** (2002) 1025–1078, astro-ph/0204527.

[220] A. B. McDonald, C. Spiering, S. Schonert, E. T. Kearns, and T. Kajita, “Astrophysical neutrino telescopes,” *Rev. Sci. Instrum.* **75** (2004) 293–316, astro-ph/0311343.

[221] H. J. Volk, “TeV gamma-ray observations and the origin of cosmic rays. III,” astro-ph/0312585.

[222] G. Sigl, M. Lemoine, and A. V. Olinto, “Maximum likelihood analysis of clusters of ultra-high energy cosmic rays,” *Phys. Rev.* **D56** (1997) 4470–4479, astro-ph/9704204.

[223] S. Lee, “On the propagation of extragalactic high-energy cosmic and gamma-rays,” *Phys. Rev.* **D58** (1998) 043004, astro-ph/9604098.

[224] A. Mucke, R. Engel, J. P. Rachen, R. J. Protheroe, and T. Stanev, “SOPHIA: Monte Carlo simulations of photohadronic processes in astrophysics,” *Comput. Phys. Commun.* **124** (2000) 290–314, astro-ph/9903478.

[225] K.-H. Kampert, J. Kulbartz, N. Nierstenhoefer, M. Risse, and G. Sigl, “Propagation of Ultra-High Energy Nuclei with CRPropa,”. Proceedings of the 31st International Cosmic Ray Conference (ICRC 2009), Lodz, Poland, 7-15 Jul. 2009.

[226] **H.E.S.S.** Collaboration, F. Aharonian *et al.*, “A Low level of extragalactic background light as revealed by gamma-rays from blazars,” *Nature* **440** (2006) 1018–1021, astro-ph/0508073.

[227] D. Kuempel, K.-H. Kampert, and M. Risse, “Simulation study of UHECR particle propagation with CRPropa,”. Proceedings of the 21^{st} European Cosmic Ray Symposium (ECRS), Kosice (Slovakia), 9-12 September 2008, P6.21.

[228] D. Harari, S. Mollerach, and E. Roulet, “On the ultra-high energy cosmic ray horizon,” *JCAP* **0611** (2006) 012, `astro-ph/0609294`.

[229] **Pierre Auger** Collaboration, R. Engel, “Test of hadronic interaction models with data from the Pierre Auger Observatory,” `0706.1921`. Contribution to the 30th International Cosmic Ray Conference (ICRC 2007), Merida, Yucatan, Mexico, 3-11 Jul 2007.

[230] D. Kuempel, K.-H. Kampert, and M. Risse, “Simulation study of GZK photon fluxes,” `0906.3099`. Proceedings of the 31st International Cosmic Ray Conference (ICRC 2009), Lodz, Poland, 7-15 Jul. 2009.

[231] A. M. Taylor and F. A. Aharonian, “The Spectral Shape and Photon Fraction as Signatures of the GZK-Cutoff,” *Phys. Rev.* **D79** (2009) 083010, `0811.0396`.

[232] A. M. Taylor, J. A. Hinton, P. Blasi, and M. Ave, “Identifying Nearby UHECR Accelerators using UHE (and VHE) Photons,” *Phys. Rev. Lett.* **103** (2009) 051102, `0904.3903`.

[233] **KASCADE** Collaboration, T. Antoni *et al.*, “The Cosmic ray experiment KASCADE,” *Nucl. Instrum. Meth.* **A513** (2003) 490–510.

[234] N. N. Kalmykov, S. S. Ostapchenko, and A. I. Pavlov, “Quark-gluon string model and EAS simulation problems at ultra-high energies,” *Nucl. Phys. Proc. Suppl.* **52B** (1997) 17–28.

[235] H. Fesefeldt, “The Simulation of Hadronic Showers: Physics and Applications,”. PITHA-85-02.

[236] A. A. Lagutin, A. V. Plyasheshnikov, and V. V. Uchaikin. Proceedings, 16th International Cosmic Ray Conference, Kyoto, 1979, Vol. 7.

[237] **KASCADE** Collaboration, J. N. Capdevielle *et al. Proc. of 22nd International Cosmic Ray Conference, Dublin, Ireland* (1991) 405.

[238] M. Unger, “EventGenerator.cc,”. $\overline{\text{Off}}\underline{\text{line}}$ version v2r6p4-branch_AsterixObelix, line 804.

[239] S. Petrera, “Towards a full simulation of the FD aperture,” *Internal Auger GAP-note*. GAP-2004-015.

[240] I. Maris, F. Schuessler, R. Ulrich, and U. M., “Data Summary Trees and Shower Visualization for Reconstructed Auger Events,” *ADST and EventBrowser Reference Manual, Internal Auger GAP-note*. GAP-2006-081.

[241] R. Brun and F. Rademakers, "ROOT: An object oriented data analysis framework," *Nucl. Instrum. Meth.* **A389** (1997) 81–86.

[242] F. Nerling, J. Bluemer, R. Engel, and M. Risse, "Description of Cherenkov light production in high-energy air showers," *Astropart. Phys.* **24** (2006) 421–437, `astro-ph/0506729`.

[243] G. Ros, A. D. Supanitsky, G. A. Medina-Tanco, L. del Peral, J. C. D'Olivo, and M. D. Rodríguez-Frías, "A new parameter for composition discrimination of Ultra-High Energy Cosmic Rays," *Internal Auger GAP-note.* GAP-2009-103.

[244] G. Ros, G. A. Medina-Tanco, A. D. Supanitsky, L. del Peral, and M. D. Rodríguez-Frías, "Sb for photon-hadron discrimination," *Internal Auger GAP-note.* GAP-2010-052.

[245] J. Bellido, "Estimating the Shower Energy Using the Information of a Single Tank (recommended for brass hybrid events)," *Internal Auger GAP-note.* GAP-2008-019.

[246] J. W. Cronin, "Particle Discrimination using the FADC traces from the Auger Observatory surface detectors," *Internal Auger GAP-note.* GAP-2003-076.

[247] A. Hocker *et al.*, "TMVA: Toolkit for multivariate data analysis," *PoS* **ACAT** (2007) 040, `physics/0703039`.

[248] P. Speckmayer, A. Hocker, J. Stelzer, and H. Voss, "The toolkit for multivariate data analysis, TMVA 4," *J. Phys. Conf. Ser.* **219** (2010) 032057.

[249] L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees.* Wadsworth and Brooks, Monterey, CA, 1984.

[250] R. E. Schapire, "The strength of weak learnability," *Mach. Learn.* **5** (1990), no. 2, 197–227.

[251] Y. Freund, "Boosting a weak learning algorithm by majority," *Inf. Comput.* **121** (1995), no. 2, 256–285.

[252] B. P. Roe, H. Yang, J. Zhu, Y. Liu, I. Stancu, and G. McGregor, "Boosted decision trees as an alternative to artificial neural networks for particle identification," *Nucl. Instrum. Meth. in Physics Research A* **543** (May, 2005) 577–584, `arXiv:physics/0408124`.

[253] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Annals of Statistics* **29** (2000) 1189–1232.

[254] J. H. Friedman, "Stochastic gradient boosting," *Computational Statistics and Data Analysis* **38** (1999) 367–378.

[255] W. McCulloch and W. Pitts, “A logical calculus of the ideas immanent in nervous activity,” *Bulletin of Mathematical Biology* **5** (1943) 115–133. 10.1007/BF02478259.

[256] C. G. Broyden, “The Convergence of a Class of Double-rank Minimization Algorithms 1. General Considerations,” *IMA Journal of Applied Mathematics* **6** (1970), no. 1, 76–90, `http://imamat.oxfordjournals.org/content/6/1/76.full.pdf+html`.

[257] R. A. Fisher, “The use of multiple measurements in taxonomic problems,” *Annals of Eugenics* **7** (1936), no. 7, 179–188.

[258] M. Erdmann. Private communication, 2010.

[259] A. Inselberg, “N-dimensional graphics, part 1 – lines and hyperplanes,” tech. rep., in IBM LASC, 1981.

[260] A. Inselberg and B. Dimsdale, “Parallel coordinates: a tool for visualizing multi-dimensional geometry,” in *VIS ’90: Proceedings of the 1st conference on Visualization ’90*, pp. 361–378. IEEE Computer Society Press, Los Alamitos, CA, USA, 1990.

[261] H. Dembinski *et al.* The Auger Observer, http://augerobserver.fzk.de/.

[262] C. Bonifazi and I. Lhenry-Yvon. SD event selection/aperture and trigger, http://ipnweb.in2p3.fr/∼auger/AugerProtected/AcceptMain.html.

[263] C. Bonifazi and I. Lhenry-Yvon. SD event selection/aperture and trigger, http://ipnweb.in2p3.fr/∼auger/AugerProtected/AcceptCrisis.html.

[264] V. Blobel and E. Lohrmann, *Statistische und numerische Methoden der Datenanalyse.* Teubner Verlag, 1 ed., 1998.

[265] J. Rautenberg. Private communication, 2010.

[266] K. M. Gorski, B. D. Wandelt, F. K. Hansen, E. Hivon, and A. J. Banday, “The HEALPix Primer,” `astro-ph/9905275`.

[267] K. M. Gorski *et al.*, “HEALPix – a Framework for High Resolution Discretization, and Fast Analysis of Data Distributed on the Sphere,” *Astrophys. J.* **622** (2005) 759–771, `astro-ph/0409513`.

[268] B. Rouillé d’Orfeuil, J.-C. Hamilton, and B. Revenu. Internet: http://apcauger.in2p3.fr/Protected/Toolkit/ , November 2010.

[269] W. Pence, “CFITSIO, v2.0: A New Full-Featured Data Interface,” in *Astronomical Data Analysis Software and Systems VIII*, D. M. Mehringer, R. L. Plante, & D. A. Roberts, ed., vol. 172 of *Astronomical Society of the Pacific Conference Series*, pp. 487–+. 1999.

[270] L. M. Bugayevskiy and J. P. Snyder, *Map projections : A Reference Manual.* Taylor & Francis, London ; Bristol, PA, 1995.

[271] S. Cheng, S. Coutu, A. Criss, P. Sommers, and R. Ulrich, "Methods for Neutron Point Source Upper Limits," *Internal Auger GAP-note.* GAP-2010-112.

[272] P. Sommers, "Prospects for Charged Particle Astronomy," 0708.2122.

[273] G. L. Cassiday, R. Cooper, S. C. Corbat, B. R. Dawson, J. W. Elbert, B. E. Fick, K. D. Green, D. B. Kieda, S. Ko, E. C. Loh, M. H. Salamon, J. D. Smith, P. Sokolsky, P. Sommers, S. B. Thomas, and B. Wheeler, "Mapping the U.H.E. sky in search of point sources," *Nuclear Physics B - Proceedings Supplements* **14** (1990), no. 1, 291 – 298.

[274] T. P. Li and Y. Q. Ma, "Analysis methods for results in gamma-ray astronomy," *Astrophys. J.* **272** (1983) 317–324.

[275] P. Sommers, "Galactic Neutron Astronomy," *Internal Auger GAP-note.* GAP-2010-024.

[276] G. Zech, "Upper limits in experiments with background or measurement errors," *Nucl. Instrum. Meth. in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **277** (1989), no. 2-3, 608 – 610.

[277] G. Zech. Private communication, 2011.

[278] A. L. Read, "Presentation of search results: the CLs technique," *Journal of Physics G: Nuclear and Particle Physics* **28** (2002), no. 10, 2693.

[279] L. Perrone, V. Scherini, and M. Settimo, "Relative Hybrid Trigger Efficiency for photon primary at low energies," *Internal Auger GAP-note.* GAP-2008-027.

[280] **Pierre Auger** Collaboration, D. Kruppke-Hansen, "Exploring the Ankle of the Cosmic Ray Spectrum with the Pierre Auger Observatory," *Submitted to Advances in Space Research.* Proceedings of the COSPAR conference, Bremen, Germany, 18-25 Jul. 2010.

[281] M. Settimo, "Hybrid detection of Ultra High Energy Cosmic Rays with the Pierre Auger Observatory," *PhD Thesis - Internal Auger GAP-note.* GAP-2010-130.

[282] R. Ulrich, R. Engel, and M. Unger, "Hadronic Multiparticle Production at Ultra-High Energies and Extensive Air Showers," 1010.4310.

[283] D. d'Enterria, R. Engel, T. Pierog, S. Ostapchenko, and K. Werner, "Constraints from the first LHC data on hadronic event generators for ultra-high energy cosmic-ray physics," 1101.5596.

[284] R. D. Parsons, C. Bleve, S. S. Ostapchenko, and J. Knapp, "Systematic uncertainties in air shower measurements from high-energy hadronic interaction models," 1102.4603.

[285] Internet: http://tmva.sourceforge.net/, 02/21/11.

[286] E. Armengaud and T. Beau, "The CRPropa framework: A numerical tool to study propagation effects on UHECRs and their secondaries in the Local Universe,". CRPropa manual version v1r3 – http://apcauger.in2p3.fr/CRPropa/index.php.

Acknowledgement

I Am indebted to many people who supported me during the completion of this thesis. I am deeply thankful to Prof. Karl-Heinz Kampert who gave me the opportunity to work on this interesting subject and for his continuous support, proofreading and supervision in all issues even in very busy times.

Very special thanks to Prof. Markus Risse. He was an important support in Wuppertal as well as in Siegen, never got tired in answering my questions and gave important initial ideas for this work. Thanks also for proofreading and comments.

Several research visits abroad turned out to be crucial steps in completing this thesis. I acknowledge the collaboration with Prof. Paul Sommers at Penn State University and his numerous ideas. The continuous connection to the University of Adelaide turned out to be very fruitful. Large parts of the scrambling technique and creation of significant maps have been developed together in Adelaide. In particular I want to thank Dr. Jose A. Bellido, Prof. Bruce Dawson and Philip Wahrlich for helpful annotations and comments. I also want to thank Prof. Günter Sigl for an introduction into the CRPropa code in Hamburg.

All these research visits would not have been possible without financial support. In particular I express my gratitude to the DAAD (Deutscher Akademischer Austausch Dienst) and the DFG (Deutsche Forschungsgemeinschaft) for various scholarships.

I appreciate the very welcoming and friendly atmosphere of the Astroparticle Group in Wuppertal. Thanks to all members of this group for fruitful discussions and backing throughout of this work. Special thanks to Dr. Julian – alias Juliano – Rautenberg for being a reliable contact person in all technical issues and for numerous travel adventures abroad. Thanks also for proofreading and helpful comments. I am indebted to Ms. Schaarwächter for assisting me in all secretarial work. Although not in Wuppertal anymore, I appreciate proofreading and important remarks of Anna Franckowiak throughout this thesis.

In the end I want to express my gratitude to my friends and family who always supported me during my studies with patience and encouragement. Finally, special thanks to Julia for her continuous support and being a major backing in completion of this thesis.

Selbständigkeitserklärung

Hiermit versichere ich, dass ich diese Arbeit nur unter Zuhilfenahme der angegebenen Quellen und Hilfsmittel selbständig angefertigt habe.

Ort, Datum	Daniel Kümpel

Index