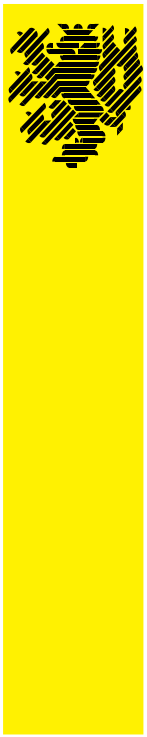


Fachbereich C - Physik
Bergische Universität
Wuppertal

# Reconstruction of the Primary Energy Spectrum from Fluorescence Telescope Data of the Pierre Auger Observatory

Heiko Geenen

DISSERTATION

zur Erlangung des Doktorgrades
des Fachbereiches C - Physik
der Universität Wuppertal

Juli 2007
WUB-DIS 2007-06

Fachbereich C - Physik
Bergische Universität
Wuppertal

# Reconstruction of the Primary Energy Spectrum from Fluorescence Telescope Data of the Pierre Auger Observatory

Heiko Geenen

DISSERTATION

zur Erlangung des Doktorgrades
des Fachbereiches C - Physik
der Universität Wuppertal

Juli 2007
WUB-DIS 2007-06

*Für meine Eltern.*

## Abstract

The Pierre Auger Observatory is the largest extensive air-shower (EAS) experiment in operation. It is still being constructed, and the final configuration will have detectors at the two sites Argentina and USA observing both celestial hemispheres. The aim of the experiment is to determine the energy, composition and origin of ultra-high energy cosmic-rays (UHECR) using two complementary detection techniques. The detector at the southern site presently contains more than 1400 (Jul. 2007) water-Cherenkov detectors at ground level ($870\,\mathrm{gcm}^{-2}$). Completion of the $3000\,\mathrm{km}^2$ large detector array is expected by the end of 2007 with finally more than 1600 tanks. The atmosphere above the site is observed by 24 fluorescence telescopes located in four buildings at the boundary of the array. During clear moon-less nights, this configuration permits *hybrid* measurement of both longitudinal development of an EAS and lateral particle density at ground. All fluorescence telescopes are fully operational since February 2007.

The aim of this work is to reconstruct the cosmic ray energy spectrum between a few $10^{17}\,\mathrm{eV}$ up to $10^{20}\,\mathrm{eV}$. This would provide an overlap to spectral results from other experiments at lower energies. The hybrid detection provides an accurate geometry determination and thereby a good energy resolution. However, the energy threshold is limited to the threshold of the surface array: larger than a few $10^{18}\,\mathrm{eV}$. The advantage of FD-monocular events (FD-mono) is a lower energy threshold in the aimed $10^{17}\,\mathrm{eV}$ regime. In addition, the present FD-mono exposure is about 1.5 times larger than the hybrid one. However, the energy resolution of FD-mono events is worse compared to hybrid, and the detector acceptance is strongly energy dependent. Therefore, the determination of the energy spectrum requires an unfolding procedure, which considers both the limited acceptance and the limited resolution.

In this analysis the FD-mono data are reconstructed. The reconstruction performance is validated and the simulation accuracy is quantified in terms of data-MC agreement of characteristic observables. The performance of different unfolding approaches are compared and the Gold algorithm is selected. For the final unfolding a dedicated kernel-MC is generated using the importance sampling. The uncertainties of the final unfolded energy spectrum are estimated using a frequentist approach. The absolute energy scale of the measured energy spectrum agrees with the KASCADE-Grande experiment within 5 % at $10^{17.5}\,\mathrm{eV}$. The agreement with the HiRes experiment below $10^{19}\,\mathrm{eV}$ is 7 %. An ankle structure is observed at about $10^{18.5}\,\mathrm{eV}$. Results are discussed.

## Zusammenfassung

Mit dem Pierre Auger Experiment wird zur Zeit das größte Luftschauerexperiment betrieben. Es befindet sich noch im Aufbau und die finale Konzeption sieht zwei Detektionsstandorte in den USA und Argentinien vor, um somit volle Himmelsabdeckung zu erlangen. Ziel des Experimentes ist es, die Energie, Art und Herkunft der ultrahoch-energetischen kosmische Strahlung mit zwei komplementäre Nachweistechniken zu untersuchen. Die Detektoranordnung auf der Südhemisphäre besteht zur Zeit (Stand Jul. 2007) aus mehr als 1400 Wasser-Cherenkovdetektoren am Boden ($870\,\mathrm{gcm}^{-2}$). Die Fertigstellung des $3000\,\mathrm{km}^2$ großen Oberflächendetektorkomplexes ist für Ende 2007 anvisiert mit einer finalen Konfiguration von mehr als 1600 ausgebrachten Tanks. Die Atmosphäre darüber kann mit 24 Fluoreszenz-Teleskopen beobachtet werden, die sich in vier Gebäuden am Rand des Oberflächenkomplexes befinden. In klaren mondlosen Nächten ist es mit dieser Anordnung möglich *hybride* Messungen von longitudinaler Schauerentwicklung und lateraler Teilchen-Dichteverteilung eines Schauers am Boden durchzuführen. Seit Februar 2007 sind alle Fluoreszenzteleskope in Betrieb.
Ziel dieser Arbeit ist, das Energiespektrum der kosmischen Strahlen im Bereich von einigen $10^{17}$ eV bis hin zu $10^{20}$ eV zu rekonstruieren. Dies erlaubt einen Überlapp mit spektralen Resultaten anderer Experimente zu niedrigen Energien hin. Die Hybriddetektion liefert eine sehr genaue Geometrierekonstruktion und dadurch bedingt, eine gute Energieauflösung. Die Energieschwelle ist jedoch durch den Bodenkomplex auf $> 10^{18}$ eV limitiert. Der Vorteil von monokularen Fluoreszenzdetektor Ereignissen (FD-mono) liegt darin, dass die Energieschwelle niedriger ist und im interessanten Bereich von $10^{17}$ eV liegt. Zusätzlich ist die momentane Statistik an FD-mono Ereignissen etwa anderthalb mal grösser als von Hybridereignisse. Die Energieauflösung der FD-mono Ereignisse ist schlechter im Vergleich zu Hybridereignissen und die Detektorakzeptanz ist stark energieabhängig. Deshalb ist es für die Bestimmung des Energiespektrums nötig, Entfaltungsprozeduren anzuwenden, die sowohl die Effekte der limitierten Akzeptanz als auch der limitierten Auflösung berücksichtigen. Für diese Analyse wurden die FD-mono Ereignisse rekonstruiert. Das Rekonstruktionsverhalten wurde validiert und die Genauigkeit der Simulation durch Daten-MC Vergleiche in charakteristischen Observablen quantifiziert. Die Qualität verschiedener Entfaltungsansätze wurde verglichen und der Gold-Algorithmus ausgewählt. Zur finalen Entfaltung wurde ein besonderes MC-Sample mit Hilfe der Methode des *importance samplings* produziert. Die Unsicherheiten des finalen entfalteten Energiespektrums wurden mit Hilfe einer frequentistischen Betrachtungsweise abgeschätzt.
Die absolute Energieskala des rekonstruierten Energiespektrums stimmt mit dem KASCADE-Grande Experiments bei $10^{17.5}$ eV innerhalb von 5 % überein. Unterhalb von $10^{19}$ eV ist die Übereinstimmung mit dem HiRes Experiment innerhalb von 7 %. Eine Knöchelstruktur im Spektrum der kosmischen Strahlung wird bei $10^{18.5}$ eV beobachtet. Die Ergebnisse werden diskutiert.

*Physics is like sex: sure, it may give some practical results, but that's not why we do it.*

Richard Feynman

# Contents

# Chapter 1

# Introduction

*The story so far: In the beginning the Universe was created. This has made a lot of people very angry and has been widely regarded as a bad move.*

Douglas Adams

In 1912, V. F. Hess [90] discovered, in a manned balloon flight, that the intensity of the ionising radiation in the atmosphere as a function of the altitude did not conform with the idea that this ionisation was caused by radioactive elements in the surface of the earth but rather pointed to an extraterrestrial origin. In the following decades, before the advent of accelerator physics, this radiation, which was first called *'ultra-radiation'* and later baptised *'cosmic rays'* by R. A. Millikan [132], proved to be one of the most fruitful means of doing particle physic experiments. In 1938, P. Auger pioneered at an experimental site at the Jungfernjoch, that cosmic rays at high energies induce an avalanche of secondary particles, that can be measured at ground in coincidence and opened up a new way of exploring an energy regime far beyond direct measurement methods [20, 19]. At the end of the first half of the 20th century such experiments had lead to the discovery of the positron [12], the pion [117] and the muon [174]. Also the first particles with strangeness were found using cosmic rays[160]. Eventually, in the beginning of the fifties, proton beams from accelerators replaced the cosmic proton flux as an experimental tool. However, after many years of accelerator physics dominating particle physics, at the end of the 20th century cosmic rays again play a major role. Nowadays, cosmic ray physics is facing a big puzzle. The cosmic radiation spectrum spreads over 10 decades in energy with great accuracy. However, the sources and the nature of highest energy particles still remains a mystery.
The Pierre Auger Observatory utilises Auger's discovery to detect cosmic rays at highest energies. For the first time an experiment has been installed, able to measure the longitudinal shower development in the atmosphere and the lateral distribution of the shower on ground instantaneously

with two complementary detection techniques. Fluorescence telescopes observe the atmosphere in clear moon-less nights while a giant surface array of water-Cherenkov tanks allows simultaneous measurement of the shower's footprint on ground. The experiment is designed to investigate the energy regime beyond a few $10^{17}$ eV. This opens up a very exciting possibility to solve some of the longstanding miracles in cosmic ray physics, namely the question about the origin of the ultra-high energetic cosmic rays (UHECR). The Pierre Auger Observatory will be capable to search for the astrophysical origin of UHECR and determine their composition and energy spectrum. The investigated energy regime allows to trace back direction of the primary particles to their origin, because their deflection in galactic and inter-galactic magnetic fields become small for energies larger than $10^{19}$ eV. This opens up a new window for astro-particle astronomy beyond our host galaxy.
Other experiments have already measured in that energy regime, like the Akeno Giant Air Shower Array (AGASA ) [202] or HiRes experiment [176], however results are still inconclusive. Therefore, the Pierre Auger Observatory has been designed to solve many current puzzles like:

- The structure of the ultra-high energy spectrum of cosmic rays and the question of a possible cutoff due to interaction with the cosmic microwave background.
- The composition of primary cosmic rays.
- The question of the sources of the cosmic ray particles. These can be related to astrophysical sources (*bottom-up* acceleration models) or to particle physical processes (*top-down* decay-scenarios of super-heavy relic particles).

In the context of the present work we focus on the determination of the energy spectrum with monocular data from the fluorescence detectors. This study is considered as an overlap with measurements at lower energies, e.g. like done with the KASCADE-Grande experiment [188]. This data-set has the two major advantages:

Statistics: The investigated data has been taken while the detector has been still under construction. Many events detected by the fluorescence telescopes, have been measured in an area where no water-Cherenkov tanks have been deployed. Restricting on events with both, informations from the surface array and the fluorescence detector (hybrid events), limits the statistics significantly.

Lower energy threshold: The fluorescence monocular events have a lower trigger threshold energy compared to the surface array. Therefore, the energy spectrum can be measured below $10^{18}$ eV.

The energy resolution of fluorescence-detector only reconstruction is worse compared to the one when using the additional information from the surface array. The limited resolution is treated in a statistical way by the method of spectral unfolding. We analyse the resulting energy spectrum for features like the ankle and compare the absolute flux with results from other experiments.

The thesis is structured in the following way:
Chapter 2 gives an overview on cosmic ray physics. Possible detection techniques are described. The wide spectrum of potential astroparticle- and particle-physics source models are sketched.
Chapter 3 gives an introduction to the experimental setup of the Pierre Auger Observatory . Both relevant detection techniques, water Cherenkov tanks and fluorescence light detection are described, and an outlook to possible future enhancements of the experiment is given. The fluorescence telescopes, their construction, calibration and monitoring are described in detail.
Chapter 4 gives insights into the data processing and reconstruction techniques of the experiment. The reconstruction of data from the fluorescence telescopes only (FD-only) is described and compared to data taken in hybrid mode. A full simulation- and reconstruction-chain has been set up on the ALiCENext computing facility [4] to generate the large Monte-Carlo (MC) statistics needed for this analysis. The architecture of this chain and the underlying MySQL database are described.
Chapter 5 describes a dedicated MC study by which the reconstruction and trigger performance is evaluated. The basic features of FD mono reconstruction are quantified and compared to the reconstruction performance when adding information from the surface array.
The disadvantage of MC based validation is that intrinsic systematics are difficult to estimate and depend on the accuracy the simulation represents reality. Especially, measures like trigger efficiency cannot be verified. Therefore, The assumptions made are described and the model dependency is discussed. The data-MC comparison shows that the detector response and trigger-simulation manifest with high precision. It is even possible to observe composition dependencies and features of the spectrum at this level.
After validation of the simulation- and reconstruction-chain in Chapter 7 the MC is used to unfold the detector-response kernel and reconstruct the physical spectrum. The method of unfolding is introduced and unfolding strategies are discussed. Deconvolution algorithms are presented and compared. Finally, the Gold unfolding algorithm is applied to unfold the primary spectrum. The implementation and the error propagation is presented. The interpretation of the unfolded FD-mono energy spectrum on physics is discussed.
Finally, Chapter 8 gives a summary of the present work and implications of the findings are discussed.

# Chapter 2

# Cosmic Rays and Extensive Air Showers

*The ability to reduce everything to simple fundamental laws does not imply the ability to start from those laws and reconstruct the universe.*

Philip W. Anderson

The Earth's atmosphere is continuously bombarded by extraterrestrial particles. This radiation is commonly known as cosmic rays. The bulk of this radiation is of hadronic nature, nevertheless, it also contains high energy $\gamma$-rays as well as neutrinos.[1] Most of the particles seem to stem from galactic origin, but the ultra-high energy tail of the spectrum with energies exceeding $10^{18}$ eV is probably of extra-galactic origin. Cosmic rays form a major component of the energy density of the interstellar medium. They consist of such a wide range of particles and energy that it is difficult to believe that they can be the result of a single type of source. Analysing the energy spectrum and the composition as well as the origin of the cosmic radiation could unveil hints on the tremendous processes that accelerate particles to energies several decades higher then man made accelerators.

[1] Sometimes in literature the definition of cosmic rays only refers to charged particles. Nevertheless, neutrino- and photon-component plays an important multi-messenger role in the understanding of the astrophysical source interpretation.

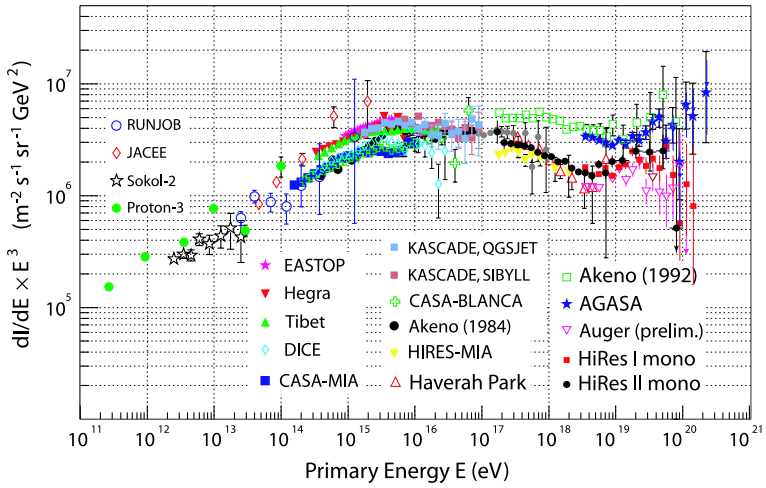



**Figure 2.1:** The all-particle spectrum, taken from [99].

## 2.1 Cosmic Rays

### 2.1.1 The Energy Spectrum of Cosmic Rays

In the following context we consider cosmic rays just as the charged particle component. In literature, one can find many ways to describe the flux, here we always note the energy to be the total energy per particle regardless the amount of nuclei.[2]
Although the main focus of the presented work is the high energy tail of the cosmic ray spectrum above $10^{17}$ eV, we here discuss the complete cosmic ray spectrum. The information collected so far gives a picture of the cosmic rays energy distribution between a few $10^9$ eV up to $10^{20}$ eV. The cosmic ray energy regime spreads over 10 decades in energy and about 30 decades in intensity. The total flux of particles over all energies amounts to about $1000\,\mathrm{m}^{-2}\mathrm{s}^{-1}$.

The differential all-particle spectrum is shown in Fig. 2.1. Obviously, a general feature of the flux is that it follows a broken power-law:

$$\frac{dN_{\mathrm{prim}}}{dE} \propto E^{-\gamma}, \text{with } \gamma \sim \begin{cases} 2.7 & \text{for} \quad E \leq 4 \cdot 10^{15}\,\mathrm{eV} \\ 3.0 - 3.1 & \text{for} \quad 4 \cdot 10^{15}\,\mathrm{eV} \leq E \leq 3 \cdot 10^{19}\,\mathrm{eV} \\ 2.7 - 3.0 & \text{for} \quad E \geq 3 \cdot 10^{19}\,\mathrm{eV} \end{cases} . \tag{2.1}$$

The energy distribution is not exponential, this indicates that the particles originate from non-thermal processes. (see Chap. 2.2.1). The spectrum itself can be separated into the following parts:

**Below $10^8$ eV,** where the solar wind shields protons coming from outside the solar system, the Sun gives the dominant contribution to the proton flux. The low energy *'solar'* cosmic rays

[2] This is the common way in air shower physics as one measures the total amount of energy deposited in the atmosphere per event, as described below.

fade more or less smoothly into the galactic ones. Their intensity strongly depends on the solar activity and solar events such as flares.

**Below (2-5)**$\cdot 10^{10}$ **eV** particles are modulated by the solar flux. There is a strong anti-correlation between solar activity and intensity of the cosmic rays.

**The bulk of cosmic rays from** $10^{10}$ **eV up to** $\sim 4 \cdot 10^{15}$ **eV** is believed to originate from within our galaxy as their energy is not large enough to allow them to escape the magnetic field of our galaxy. The most plausible acceleration process is shock acceleration in supernovae remnants.

**Around** $10^{15} - 10^{16}$ **eV** the spectrum steepens and the spectral index changes from 2.7 to 3.0. This kink in the energy spectrum is called knee and might indicate a change in the source mechanism and thus acceleration processes of the cosmic rays. This is consistent with the idea that the galactic cosmic rays are accelerated by supernovae and the acceleration process becomes less efficient above $10^{15}$ eV.x The knee can be interpreted as superposition of the energy spectra of the different elements, that cutoff at different energies. It is still under investigation, whether the exact shape of the knee can be described by a rigidity- or mass-dependent cutoff energy [186].

**Between** $10^{16}$ **eV and** $10^{18}$ **eV** the spectrum is again a power-law, with spectral index 3.0-3.1. The existence of a second knee at around $4 \cdot 10^{17}$ eV has been reported by [175, 33], but could not be confirmed yet.

**At around** $10^{18.5}$ **eV** the spectral index changes again and the spectrum becomes flatter. This kink is known as ankle and may again indicate a change in the source mechanisms[3].

**Above** $10^{19.5}$ **eV** the experimental data give a contradictory picture of the spectrum. Air-shower detector-arrays like a AGASA indicate a continuation of the spectrum far beyond $10^{20}$ eV[202], whereas the fluorescence-detector experiment HiRes shows evidence of a cutoff, that could be understood by an interaction of cosmic rays with the cosmic microwave background[183]. This so called GZK cutoff is discussed in Chap. 2.1.4.

### 2.1.2 The Composition of Cosmic Rays

The primary cosmic radiation is mainly composed of hadrons by 98 % and electrons by about 2 %. The amount of anti-particles, neutrinos and photons is rather small and not in the scope of this work. The hadronic component of the cosmic rays is dominated by protons (87 %), helium nuclei (12 %) and heavier nuclei [123]. The exact composition is well known in the low energy regime, where the flux allows to measure the particles directly via balloon or satellite based experiments. Below $10^{10}$ eV the cosmic radiation is dominated by protons and helium nuclei, due to the fact that the Earth's magnetic field and the solar wind shield heavy nuclei. Below $10^{10}$ eV the flux opens up the possibility to separate the isotopes, like $Be^9/Be^{10}$ [60]. That allows to estimate the escape time from the galaxy to be about 15 million years. Hydrogen and helium nuclei are much less abundant

[3] In this context, we do not distinguish if the acceleration or propagation process change or the kind of source itself.

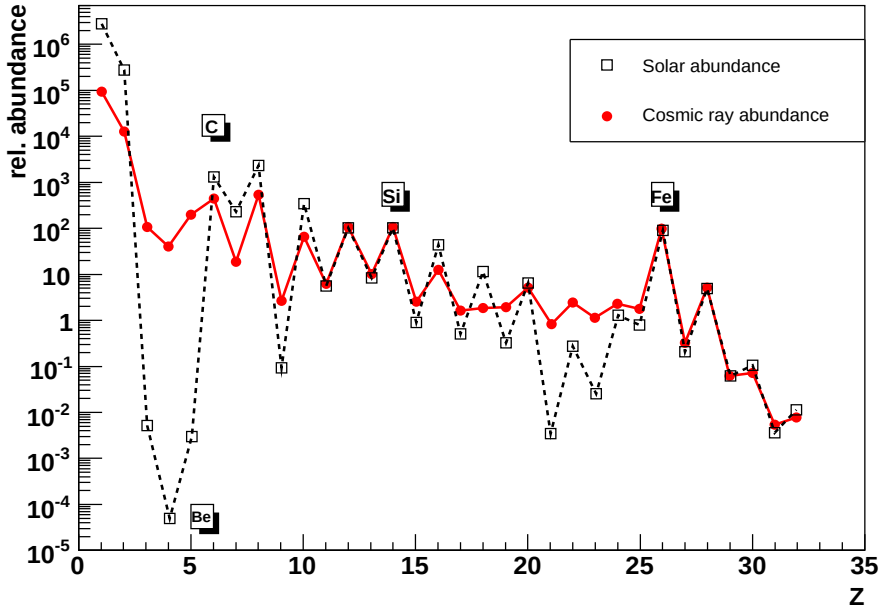



**Figure 2.2:** The abundances of various nuclei in the cosmic radiation compared to the abundances in the solar system. The nuclear charge scale refers to the elementary group H-Ni. The abundances are normalised to Si=1. Compilation is based on [197] . Further details on the light element composition of cosmic rays can be found in [198, 82]. Details on heavy element cosmic rays compositions can be found in [35, 36, 49, 68].

in the cosmic rays than in our solar system material. This is not fully understood, but it might be due the fact that hydrogen is relatively hard to ionise for injection into the acceleration process, or it could reflect a genuine difference in composition at the source [69]. The abundances of heavier elements are similar to those in our solar system, except a higher abundance in the Li-Be-B group and iron group (compare Fig. 2.2). The enhancement can be understood in terms of the spallation of heavy nuclei due to collisions during their propagation from the distant sources to the solar system. Thus, e. g., the Li-Be-B group is expected to stem from the C-N-O group.
In the past years it has become obvious that the knee in Fig. 2.1 is not a sharp kink but more a continuous change of the slope and extends over a wider energy range between $3{\cdot}10^{15}$ eV and $10^{17}$ eV. It can be understood in terms of the different elementary components having similar energy spectra, but different cutoff energies. This agrees with the observation of a tendency in the mass composition to become heavier with increasing energy [186]. There are investigations if the ankle in the cosmic rays spectrum is coincidence with change in the composition from heavy to light primaries at highest energies [187].

### 2.1.3 Arrival Directions of Cosmic Rays

The answer to the question, where the cosmic rays stem from, is difficult and not yet fully solved, depending on the energy regime.

**'Solar' cosmic rays ($<10^9$ eV):** Strictly, the solar cosmic rays are no cosmic rays, as *'cosmic'* implies an extra-solar origin. Nevertheless, they play an important role for the low energy cosmic rays regime. Furthermore, an increase in the intensity of solar cosmic rays is followed by a decrease in all other cosmic rays, due to the fact that the solar wind is sweeping the cosmic rays further outward. This is called the Forbush effect (see e.g., [3]).

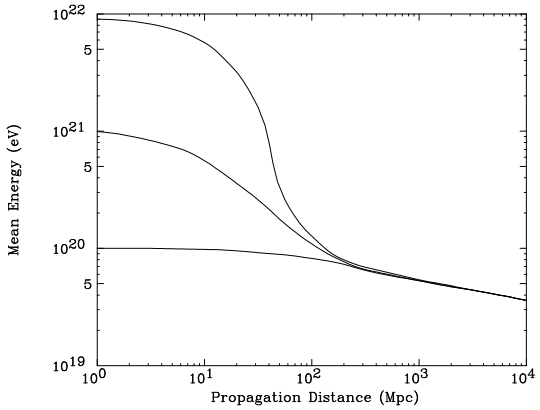
**Cosmic rays with energies between $10^9$ eV and $10^{17}$ eV:** Above $10^9$ eV the majority of detected particles originates from outside the heliosphere. The energies are small enough so that these particles can be trapped by galactic magnetic fields. These fields isotropise the cosmic rays in this energy regime. Measurements, e.g., from the KASCADE experiment in the PeV-energy regime [98, 13] or L3-Cosmics in the multi GeV-energy regime [97] indicate that anisotropies larger than 1 % can be excluded. Nevertheless, in the multi-TeV energy regime a small scale anisotropy has been observed by the Super-Kamiokande experiment [147, 84, 85] and Tibet experiment [98, 185, 6]. The Tibet results are of the order of 0.01 % and can be explained by the rotation of the Sun within the galaxy and the Earth's motion along its orbit [6]. One expects a sidereal-diurnal effect of amplitudes. This effect is known as Compton-Getting effect and manifests as dipole. Additionally, the Super-Kamiokande experiment has measured additional higher harmonic small scale anisotropies [85].

**Cosmic rays above a few $10^{18}$ eV:** Above $10^{18}$ eV the deflection of the magnetic field in our galaxy becomes smaller. An anisotropy is expected and the cosmic rays should point to their sources. Currently, there is an issue of plausibility of experimental data. Experiments like AGASA [87] and SUGAR [29] have reported an excess from the galactic centre, as well as spacial multiplets. The Pierre Auger Observatory and HiRes could not confirm these results. In case of the excess from the galactic centre the results from AGASA and SUGAR can be excluded by the Pierre Auger Observatory [179, 182].

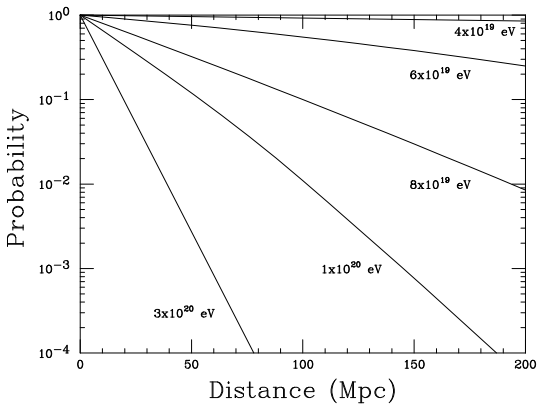
**Cosmic rays above a few $10^{20}$ eV:** The angular deflection of a particle from a straight line trajectory in a magnetic field can be estimated by [177]:

$$\Delta\Psi \approx 0.3^\circ Z \frac{L_{\mathrm{kpc}} \cdot B_{\mu\mathrm{G}}}{E_{20}} = 0.3^\circ Z \frac{L_{\mathrm{Mpc}} \cdot B_{\mathrm{nG}}}{E_{20}}$$

Here, the particle energy $E_{20}$ is in units of $10^{20}$ eV, the magnetic field strength $B$ is in $\mu G$, when the source distance $L$ is in units of kpc and $nG$ when $L$ is in Mpc. The former case is typical for Galactic dimensions and fields, while the latter one is for extra-galactic paths. The exact strength of intergalactic field is rather vague, but it is plausible to assume that the fields are coherent on scales of about 1 Mpc. Interpreting the consecutive deflections along the regions of coherent fields in terms of multiple scattering, the average angular deviation




**(a)** Energy loss as function of propagation distance through the CMBR for different source energies. Plot is taken from [55].




**(b)** Probability that an observed event has a source distance larger than indicated. This plot is taken from [55], the original plot was made by P. Sommer and assumes an $E^{-2.5}$ differential energy spectrum.

**Figure 2.3:** The GZK cutoff.

from a straight line path can be approximated as $\sqrt{L_{Mpc}}$ times the angular deflection over 1 Mpc. However, extra-galactic galactic deflection are likely to exceed galactic deflection, even though pion photon-production limits the path to about 50 Mpc (see Chap. 2.1.4). This means that the arrival direction should be dispersed at maximum by $\sqrt{50} \cdot 0.3^\circ \approx 2.1^\circ$. This allows to trace back single sources although up to now no source has been identified.

### 2.1.4 The GZK Cutoff

It is expected that cosmic ray interactions with the cosmic microwave background radiation (CMBR) produce a sharp cutoff in the observed spectrum [32]. The reason is that nucleons above $5 \cdot 10^{19}$ eV strongly loose energy during their propagation from the source to the Earth by photo-pion production in nucleon collisions with the photons of the CMBR. This so called GZK-cutoff[4] in the cosmic ray spectrum at about $5{\cdot}10^{19}$ eV is characteristic for models with extra-galactic sources homogeneously distributed on the Hubble scale $c/H_0 \sim 3 \cdot 10^3 \cdot h^{-1}$ Mpc. The mean free path for $10^{20}$ eV protons moving through the 2.7 K microwave radiation is estimated to be about 5-7 Mpc with a typical energy loss of 20 %. No cosmic accelerator has been detected at less than 50 Mpc distance, yet, which can provide such extreme physics to accelerate primary particles to $10^{20}$ eV. Thus, the observation of a cutoff in the cosmic ray spectrum is expected[5].

[4] Greisen, 1966; Zatsepin & Kuzmin, 1966

[5] More precisely, in terms of a continuous spectrum reaching to trans-GZK energies the GZK effect would manifest as dip at the resonance in the spectrum

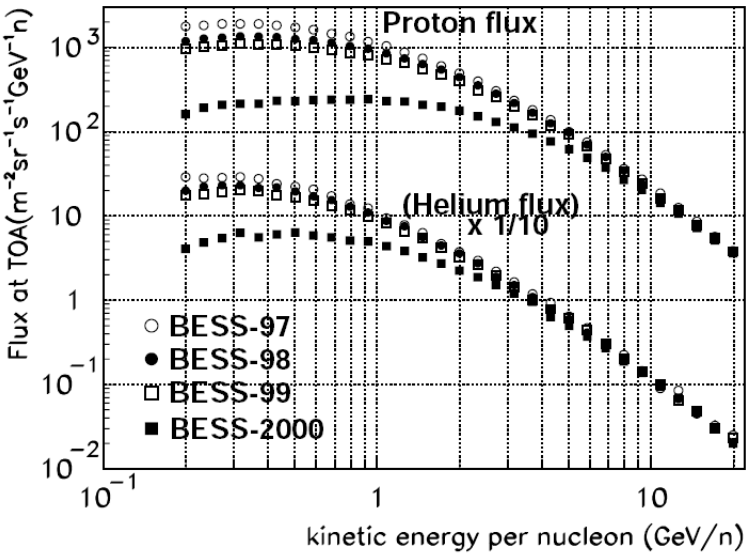



**Figure 2.4:** Solar Modulation: Proton and helium fluxes at the top of the atmosphere from 1997 to 2000 measured by the BESS experiment [169].

## 2.2 Source and Propagation Models of UHE Cosmic Rays

As previously mentioned, the flux of CR at a few GeV is modulated by the Sun's activity. Particles with $10^8 - 10^{11}$ eV show an anti-correlation with the solar activity, indicating an extra-solar origin (compare Fig. 2.4). Between $10^{10}$ eV and the knee at $3 \cdot 10^{15}$ eV it is widely accepted that the bulk of particles stem from galactic sources. This is also motivated by their total energy content and the fact, that the composition is proton dominated. Recently, multi-wavelength observations of X-ray and TeV $\gamma$-rays favour that most of the CR particles are accelerated in supernovae remnants [2]. The region between knee and ankle is still matter of discussion. The knee itself is usually interpreted as transition to sources of different processes. There are indications, that the sources of these particles are more astrophysical like scenarios. E.g., predict particle-physically motivated scenarios a large fraction of photons in the cosmic rays, which is strongly disfavoured by measurements of the Pierre Auger Observatory for the UHECR [88, 158]. In case of astrophysical scenarios, the knee is interpreted as transition from galactic to extra galactic sources, like AGN-jet or GRBs[6]. Assuming a galactic magnetic field in the order of $10^{-10}$ T particles with higher momenta than $4.6 \cdot 10^{15}$ eV/c are expected to have a gyro-radius larger than a few pc and should start to escape from the galaxy. The investigated energy regime of this work is above $10^{17}$ eV. The origin of these UHECR remains

[6] AGN are active galactic nuclei and GRB are gamma ray bursts, the astrophysical objects are explained below

a mystery and in the following, plausible scenarios are classified and described.

### 2.2.1 Models of the Origin of Ultra-high Energy Cosmic Rays

The scenarios, that explain how UHECR can gain their energy, are usually classified into two fundamentally different approaches:

**i)** Theories describing particles being accelerated from low energies to higher energies by some external process are commonly refered to as **bottom-up models**.

**ii)** The other group of scenarios investigates the assumed decay of super-massive or high energy particles. The decay products would manifest as UHECR. As they contain only a fraction of the energy corresponding to the mass of the primary particle, these scenarios are classified as **top-down models**.

### 2.2.2 Bottom-up Origin - Astro-Physical Scenarios

Bottom-up scenarios can be classified into models that assume a single-shock acceleration of particles in an electromagnetic field or statistical acceleration via Fermi acceleration in magnetised plasma. Before discussing the two scenarios in detail it is helpful to plot the possible sources in a Hillas diagram. Neglecting the details of the exact acceleration mechanism one can estimate an upper limit for the energy to which these sources can accelerate particles by the probability to escape the acceleration region. If the Larmor radius of the particle is of the order of the size of the acceleration size, it is difficult to confine the particle. Hillas described in [91] that the maximum energy which a particle of the charge $Ze$ can reach within a region of size $R$ is given by

$$E_{max} \sim \beta Z \left(\frac{B}{1\mu G}\right) \left(\frac{R}{1kpc}\right) 10^{18}\,\mathrm{eV}\ . \tag{2.2}$$

This so called Hillas criterion considers $\beta$ the velocity of shock waves in case of stochastic acceleration or the efficiency of the acceleration in case of a single-shock acceleration scenario. The Hillas plot is a nice representation (Fig. 2.5) for classifying possible sources. The abscissa usually gives the region size and the ordinate the magnetic field strength. For a fixed energy and particle type the possible sources have to be above a straight line.

Usually, these objects are not expected to achieve the maximum energy motivated by Eqn. 2.2 due to many processes of energy loss. Considering synchrotron radiation loss or photon-pion production allows to draw exclusion regions in the B-$\beta$ plane as represented in Fig. 2.6:

**Synchrotron radiation loss:** When accelerating a charged particle, it emits synchrotron photons. The size of the acceleration region has to be large enough to make energy loss due to synchrotron radiation negligible. The energy loss is proportional to $E^4/R^2$, that yields the general trend:

$$-\frac{dE_{\mathrm{sync}}}{dt} \propto B^2 E^2\ . \tag{2.3}$$

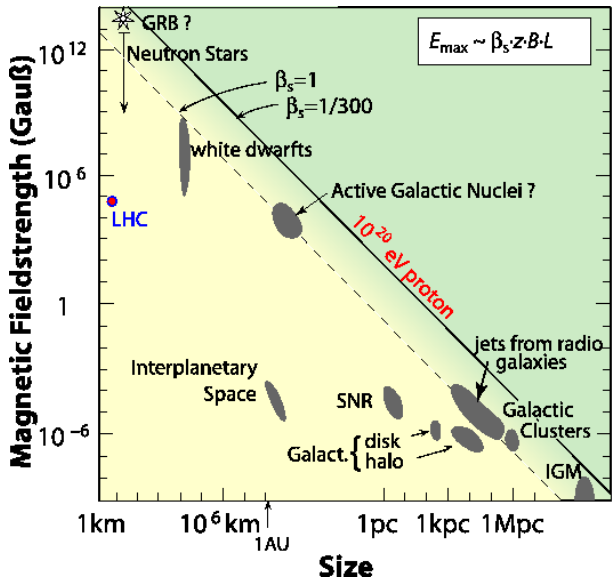



**Figure 2.5:** Hillas plot, compilation by K.-H. Kampert [100].

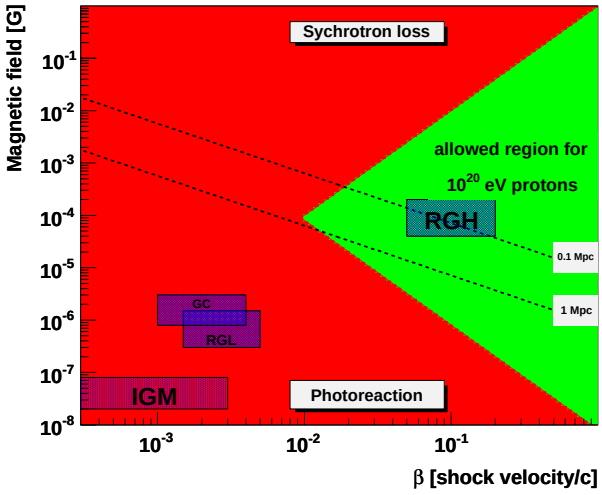



**Figure 2.6:** Additional limits can be drawn from synchrotron loss processes and photo-pion dissipation. Magnetic field strength and shock velocity of possible sites of acceleration are drawn. Only the green region allows acceleration of protons to $10^{20}$ eV. Candidate accelerators must also lie above the dashed lines appropriate to their characteristic size. Compilation from [177]. RGH: Radio Galaxy hot spots, RGL: Radio Galaxy Lobes, GC: Galactic Cluster, IGM: intergalactic medium.

On the other hand, assuming a particle being accelerated for a time $T_a = R/\beta$ the rate of energy gain can be expressed by:

$$\frac{dE_{\text{gain}}}{dt} \sim \frac{E_{\text{max}}}{T_a} \propto \beta^2 B \; . \tag{2.4}$$

Acceleration ceases when the rate given in Eqn. 2.3 excesses the energy gain given in Eqn. 2.4. This condition leads to the upper exclusion region in Fig. 2.6.

**Photon-pion production:** The acceleration region should be smaller than the interaction length for the particle for loosing energy via photon-pion production with the CMBR or more realistic with local radiation at the source.

$$\frac{dE_{\text{pion}}}{dt} \sim \text{constant;} \, . \tag{2.5}$$

Equating this loss with the rate of energy gain given in Eqn. 2.4 leads to the lower exclusion region in Fig. 2.6. Parameters used are for interaction with the CMBR and it is likely that the photon density at in the source region is much larger. This would restrict further the allowed region in Fig. 2.6.

### 2.2.3 Single-Shock Acceleration

To motivate a direct acceleration along an electric field, it is necessary to explain the origin of a constant strong field. Therefore, these models focus basically on rotating magnetised neutron stars, called pulsars. Other models focus on accretion discs around black holes, that are threaded by magnetic fields. One of the major disadvantages of direct acceleration models is, that most models do not reproduce the observed power-law spectrum. It is clear that direct acceleration can be excluded to explain the origin of the UHECR. Pulsars or stellar black holes cannot produce a field strong enough to accelerate particles up to $10^{20}$ eV. Galactic black holes like the core of a rotating radio galaxy would give rise to an electromagnetic field large enough to accelerate the particles up to $10^{20}$ eV, but these objects are shielded by dense pair plasma and intense radiation, which would cause energy losses. The crab pulsar might have the capacity to accelerate a proton up to a few $10^{16}$ eV and iron into the $10^{18}$ eV range, but it is not expected, that the complete electric field is available to accelerate particles. Due to pair production in the electromagnetic field the potential should be smaller.

### 2.2.4 Fermi Acceleration

Most source models discussed are based on particle acceleration in repeating encounters, which can be described by the Fermi mechanism. In the following we sketch the basic principles of Fermi acceleration, mainly based on the argumentation in [69, 124, 172, 173]. The original idea of E. Fermi in 1949 ([67]) is based on the diffuse acceleration of cosmic rays by random collisions with magnetised interstellar clouds. It can be shown that this *cloud acceleration* is not effective enough

to explain the cosmic ray spectrum, more likely, it is believed that particles are accelerated in the vicinity of shocks. Here, a charged particle is reflected from so called magnetic mirrors which are associated with irregularities in the magnetic field of each rest frame, before and behind the shock front. In that way the particle repeatedly crosses the shock-front.
The energy gain $\Delta E = \alpha E$ is assumed to be proportional to the energy $E$ of the particle before each specific encounter. Hence, the energy after $n$ encounters can be calculated as:

$$E_n = E_0(1+\alpha)^n \ .$$

The particle does not stay in the vicinity of the shock wave *ad infinitum* . More realistically, the particle is considered to have a probability $P_{\text{esc}}$ of escaping from the shock in each encounter. Thus, the probability that the particle is still in the shock wave after $n$ encounters is $(1-P_{\text{esc}})^n$. Therefore, the fraction of particles accelerated to energies higher than $E_f$ can be estimated as:

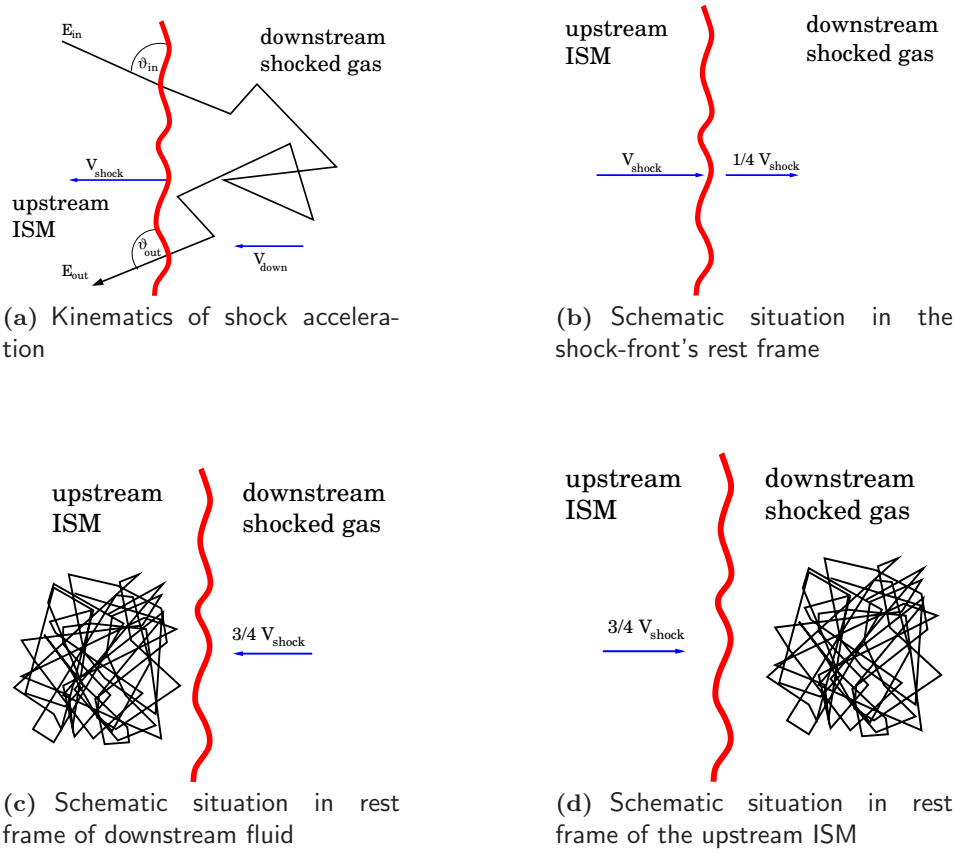
$$N(E \geq E_f) \propto \frac{1}{P_{\text{esc}}} \left(\frac{E_f}{E_0}\right)^{\ln(1-P_{\text{esc}})\cdot\ln(1+\alpha)} \ .$$

Thus, the integral spectrum can be described as power-law $N(\geq E) \propto E^{-\gamma}$ with

$$\gamma = -\ln(1-P_{\text{esc}}) \cdot \ln(1+\alpha) \approx \frac{P_{\text{esc}}}{\alpha} = \alpha^{-1}\frac{T_{\text{cycl}}}{T_{\text{esc}}} \ .$$

Here, $T_{\text{cycl}}$ is the characteristic time, that the particle needs for each encounter and $T_{\text{esc}}$ is the characteristic time for the particle to leave. Independently of the scenario, clouds or shocks, the advantage is that a power-law spectrum with spectral index $\gamma \sim 2$ at the source is predicted. It depends on the fraction of energy gain $\alpha$, how efficient the process can be.
The original cloud acceleration scenario should happen in the interstellar medium and $\alpha$ turns out to scale with $(v_{\text{cloud}}/c)^2$ [124, 69]. Therefore, it is classified as second-order Fermi-acceleration. In the following the more efficient acceleration scenario in the vicinity of shock-fronts is investigated:
A shock-front[7] moves with the velocity $v_{\text{shock}}$ into the interstellar medium. The existence of high energy particles in front of the shock-front and behind is assumed. The shock-front's size has to be much smaller than the gyro-radius of the particles, so that they hardly notice the traverse. The shocked gas is expected to flow away from the shock-front downstream slower than the shock-front $|v_{\text{down}}| < |v_{\text{shock}}|$. At both sides of the shock-front magnetic turbulences are expected to be able to deflect elastically a charged particle. Downstream the turbulences are believed to be produced by the shock itself, upstream they are induced by the CR. In contrast to the cloud acceleration the CR are assumed to be isotropic in both the ISM reference frame as well as the rest frame of the downstream fluids. Fig. 2.7 sketches the shock acceleration scenario. Starting in the shock-front's rest frame, the upstream fluid is approaching the shock-front with $v_{\text{shock}}$ and flowing away with

[7] For sake of illustration one might think of supernovae. Here, the ejected material can reach velocities up to $10^7\,\text{ms}^{-1}$. This can be compared to the Alfvén speed of the interstellar medium of about $10^5\,\text{ms}^{-1}$. A strong shock-front is a realistic scenario.




**(a)** Kinematics of shock acceleration

**(b)** Schematic situation in the shock-front's rest frame

**(c)** Schematic situation in rest frame of downstream fluid

**(d)** Schematic situation in rest frame of the upstream ISM

**Figure 2.7:** Schematic illustration of the 1st order Fermi acceleration in strong shock winds.

$v_{\rm down}$. Using the continuity equation one can derive:

$$\frac{v_{\rm shock}}{v_{\rm down}} = \frac{\rho_{\rm down}}{\rho_{\rm ISM}} \; .$$

For a mono-atomic or fully ionised gas using the adiabatic constant for specific heats of an ideal gas $\gamma = 5/3$ one can derive

$$\frac{v_{\rm shock}}{v_{\rm down}} = 4 \; .$$

The idea is, that a particle in the upstream rest frame sees the downstream gas approaching with $3/4\, v_{\rm shock}$. When crossing the shock-front a small amount of energy is transfered to the particle. The turbulences in the downstream region isotropise the particles again and in the downstream rest frame the interstellar medium approaches with $3/4\, v_{\rm shock}$. When crossing the shock-front again,

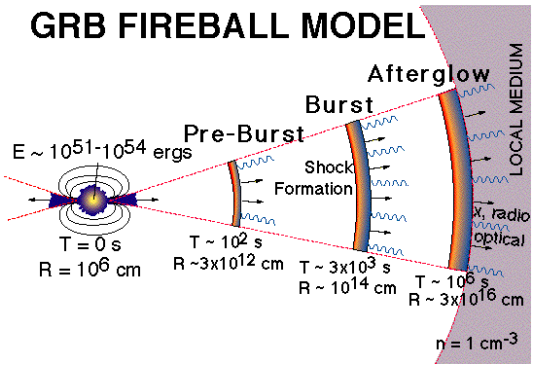



**Figure 2.8:** Schematic drawing of the Fireball scenario of a GRB.

they encounter gas moving towards the shock front with $3/4 \cdot v_{\mathrm{shock}}$. Therefore, the particles are expected to be accelerated by exactly the same process when crossing from the upstream to the downstream region and vice versa. This makes particle acceleration more efficient, compared to the second order Fermi acceleration as for each encounter they gain energy, while in the cloud acceleration scenario they can only gain energy in head-on collisions [69]. Calculating the full kinematics it turns out that non-relativistic shock acceleration scales with $\alpha \propto (v_{\mathrm{shock}}/c)$ and thus is refered to as first-order Fermi process.
Relativistic shock-flow brings substantial changes to this picture, because the assumptions made in the derivation of the spatial diffusion equation are no longer valid [144, 145, 118]. Thus, when $v_{shock}$ is of the order of the velocity of light, the quantity $v_{\mathrm{shock}}/v_{\mathrm{particle}}$ can no longer be small. The Lorentz boost linking the two frames significantly deforms the angular distribution of the particles, and it cannot be assumed isotropic in both the upstream and the downstream rest frames. Instead of assuming spatial diffusion, the angular dependence of the distribution function must be computed explicitly. It turns out that in the upstream region the particle is quickly caught back by the shock-front (timescale $\sim t_L/\Gamma$), which allows to obtain larger maximum energies for acceleration and opens up the possibility to explain the UHECR. Relativistic winds are common in high energy astrophysics. Especially, the effect of ultra-relativistic shocks has been investigated in the context with jets from Active Galactic Nuclei ($\Gamma \sim 10$) or Gamma Ray Bursts ($\Gamma \sim 100 - 1000$) [172, 173].

### Plausible Bottom-Up Scenarios for Ultra-High Energy Cosmic Rays

Until now, no clear astro-physical source could be identified to be the source of the UHECR, nevertheless the list of possible candidate sites is long [9], and the spectrum of plausible sources contains e.g.: SN explosions[8] [116, 194], Pulsars [9], large scale wind termination shocks from galaxies [96]

[8]Hereby, the most energetic scenarios are supernova remnant shocks in stellar wind cavities are, that are commonly considered to contribute to the energy regime up to at maximum $10^{16}$ eV, thus cannot explain the origin of the UHECR *in sensu stricto*

| Acceleration model | $E_{max}[eV]$ | Ref | primary | possible reducing due to |
|---|---|---|---|---|
| Shock acceleration in GRBs | FB: $5 \cdot 10^{15} Z$<br>CB:$4 \cdot 10^{20} Z$ | [73, 72]<br>[59] | p,Fe | B field structure |
| Relativistic jets from AGN or micro-blazars | $\lesssim 10^{20} ZeV$ | [92, 153] | p | structure of B field |
| dead quasars | $\lesssim 4.410^{20} Z$ | [44] | p,O,Fe | maintenance of B field |
| Super galactic accretion shocks | $10^{17} - 10^{18}$ | [146, 143] | p,O,Fe | maintenance of B field |
| Wind and shock acceleration in rotating neutron stars | $3 \cdot 10^{19} Z$ | [39] | Fe | injection efficiency |
| direct acceleration in neutron stars | $\ll 10^{20} Z$eV<br>$\sim 10^{17} Z$ | [143, 191] | p, Fe | pair production in magnetosphere reduces voltage by few orders of magnitude |

**Table 2.1:** List of possible UHECR acceleration scenarios discussed in the text. Column 2 gives the maximum energy that is attainable in the acceleration. This energy can differ from the expected energies by more than one order of magnitude depending on the model. Details can be read in the references given in Column 3.

or galaxy clusters [140, 101, 133], acceleration in AGN [151, 184, 81, 153, 154, 11, 141, 142], dead quasars [44, 45], magnetars[9] [109, 111, 110, 34, 15], direct acceleration in polarisation fields, which are generated by plasmoids produced in planetoid impacts onto neutron stars [121], acceleration in star-burst galaxies [65, 10, 7], magneto-hydrodynamic winds driven by newly formed strongly magnetised neutron stars [39], GRBs [196, 192, 131, 135], stranglets accelerated in astrophysical environment [126] or hostile aliens with a big CR gun [8]. Here is just a focus on the most seriously discussed candidate sources, a summary is given in Tab.2.1

**Shock-acceleration in GRBs:** About once a day a strong gamma ray burst is observed in the universe. Although the exact understanding of the source mechanism is still a mystery, it is possible that these events contribute to the UHECR-flux. A strong motivation for the GRB-UHECR association is that the GRB red shifts indicate an extra-galactic origin, and thus could explain the level of isotropy of the UHECR. Moreover, the injection rate of CR above the ankle is about $3 \cdot 10^{44}\,\mathrm{erg\,Mpc^{-3}\,yr^{-1}}$, this is similar to the $\gamma$-ray emission rate between 0.1-1 MeV of local GRBs. Presently, two models are widely discussed: the 'Cannon-ball model' (CB) and the 'Fireball model' (FB).

The FB model assumes an ultra-relativistic wind as source for the photons (compare Fig. 2.8).

[9] pulsar with dipole and a very strong magnetic field of about $10^{15}$G

However, the astrophysical object that could be the engine for the wind is not clearly defined. E.g., there are hyper-novae[10] or binary merger under consideration as initiating events. The GRB $\gamma$-pulses can be produced by synchrotron radiation in collisions between the highly relativistic shells. The afterglow can be explained via synchrotron radiation or electron cooling produced when the merged shells collide with the ISM. This should drive a blast wave in the ISM and a reverse shock into the merged shell. A simplified model, developed by Waxman [196], predicts the acceleration of UHECR in the shocks. Later modifications by Gallant [72] and Ostrowski [27] showed that the process should be gradual due to the extreme particle anisotropy, and provide a gain of $\Delta E/E \sim 1$. The maximum energy is expected to be in the order of $10^{18}$ eV.

The CB model describes the long-duration GRB and their afterglows by bipolar jets containing bunches of ordinary atomic matter, so called cannon balls [152]. The astrophysical event is an ordinary core collapse supernova explosion. The remaining object forms an accretion disc or torus, containing either stellar material originally close to the surface of the imploding core, left behind by the explosion generating out-gone shock or more distant stellar matter falling back after its passage. It is assumed that spontaneously parts of the accretion torus fall onto the compact object, producing the cannon-balls, emitted with high bulk Lorenz factors in opposite direction along the objects rotation axis, where matter already been absorbed by the object due to missing rotational momentum. The $\gamma$-rays are produced by the CB passing through the SN light scattered by the SN and pre-scattered ejecta. The electrons in CB Compton-up scatter the photons to GRB energies. A Fermi-accelerated particle in a CB can theoretically be accelerated up to an energy when the Larmor radius exceeds the CB size. A nuclei from ISM trapped in a cannonball can be accelerated within it and exit the cannonball with energies extending up to $(2-6)\cdot 10^{20}$ eV [59, 61].

**Relativistic jets in AGN or micro-blazars:** Active galactic nuclei (AGN) are very promising candidates to accelerate particles to high energies. These are the most powerful single objects in the cosmos. About 1 % of all bright galaxies possess an active nucleus. This nucleus is of the size of our solar-system, but the emitted radiation is as luminous as a normal galaxy. Based on thermal emission, a Wien peak in the ultraviolet can be reconstructed. The only process imaginable that can provide this luminosity for at least $10^8$ years, are super massive black holes ($m_{\mathrm{BH}} \sim 10^8 - 10^9 M_\odot$) at the centre of these galaxies. Such a massive black hole is accreting huge amounts of matter forming an accretion disc. The maximum luminosity for an assumed spherical accretion disk around a typical AGN can be estimated by equalising the radiation and the acceleration force to the Eddington luminosity of $L_E = m/10^8 M_\odot \cdot 1.3 \cdot 10^{46}$ erg/s.

Along the rotation axis highly relativistic jets are formed. These jets emit a strong non-thermal radiation. The kinematics of the jets are highly forward boosted. The generation and acceleration of cosmic radiation is subdivided into two classes of models [86].

[10] This is a hypothetical rare class of very energetic supernovae (SNe) type Ic, generated by a direct collapse of a massive progenitor star to a black hole. This collapse produces a highly conical and ultra-relativistic fireball [27].

The so called **AGN-core model** describes the production and acceleration of particles in shocks in the accretion flow close to the black hole. The second type of model assumes that the production and acceleration of particles takes place in shocks in the jets. The latter effect can sometimes take place several tens of light-years away from the AGN core. These models are collected under the concept **jet models** [86, 70]. These models focus on acceleration in AGN that point with the jet cone towards the Earth. These AGN are called blazars. Among the jet models it is useful to distinguish between those producing $\gamma$-emission via synchrotron and inverse Compton scattering only and proton blazars which are assumed to produce the $\gamma$-emission from neutral pion decays. The pions are generated by proton-proton or proton-$\gamma$ interactions (only in high pion multiplicity processes). The proton blazars are assumed to describe a non-negligible contribution to the CR spectrum at ultra-high energies. The hadronic primaries are expected to be accelerated up to $10^{18} - 10^{20}$ eV [143].

**Micro-quasars:** The morphology of a micro quasar is similar to those of AGN (compare Fig.2.9). The engine of a micro-blazar is a spinning black-hole, accreting a torus or disk heated by viscous dissipation and ejects collimated jets of relativistic particles in the ISM. In contrast to AGNs the mass of the black-hole is only a few tens of the solar mass (compared to up to few $10^6 M_\odot$ in AGNs) [134]. Thus, the accretion disk has a much smaller radius ($10^3$ km, compared to $10^9$ km in AGNs). Therefore, the temperature in the accretion disk is in the order of a few million degrees (few 1000 ° in AGN), the Wien peak is shifted and the thermal radiation from the accretion disk is visible in the X-ray regime (UV for blazars). The jet is expected to contain, similar to the AGN model, collimated magnetised plasma at relativistic speed, but these particles should travel only a few light-years, instead of several million light years like the jets of giant radio galaxies can travel. Presently, it seems to be evident that the group of micro-quasars might contribute to the galactic component of the CR spectrum in the few TeV region, but could not contribute to the UHECR [120, 50]. There are attempts to describe GRBs, Blazars and Micro-blazars in a unified approach as a universal process with different sources [62, 134] (Fig.2.9). The cannonball model would support this concept.

**Dead quasars:** When scanning the universe at cosmological red shifts ($z \sim 2$) it becomes obvious that in the early universe the density of quasars was much higher. This can be interpreted that many of the present time galaxies host a rapidly rotating super-massive black hole in their centre. This black hole does not appear as AGN as it has accreted most of the matter in its neighbourhood. Due to missing fresh gas it has become an inactive 'dead quasar'. Nevertheless, assuming a small accretion rate, this can bring magnetised plasma near the rotating black hole and form a stressed magnetosphere near the event horizon [44]. For black-holes with $m = 10^9 M_\odot$ and a strong field at the horizon of $B = 10^4 G$ the effective electro-motive force can be estimated to be $cBR \sim 4.4 \cdot 10^{20}$ V [44] and thus would be candidate site for UHECR acceleration. But up to now, there are no indications that such small accretion rates can maintain a giant magnetic field strength in the order of $10^4$G.

**Super-galactic accretion shocks:** Formation of large super-galactic structures could produce

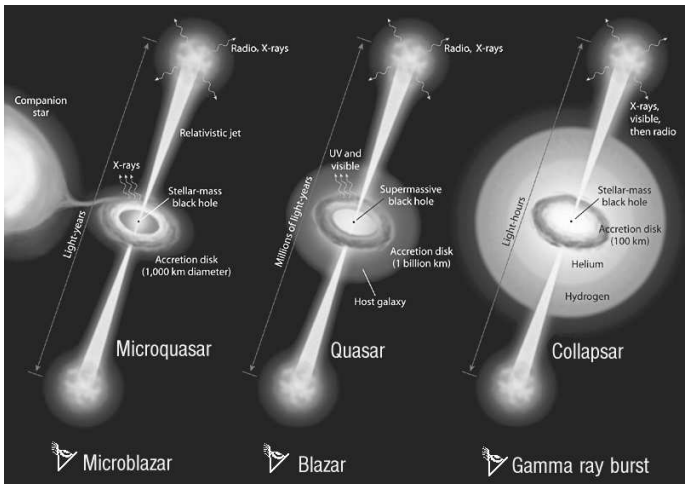



**Figure 2.9:** Unified approach to describe GRBs, Blazars and Micro-blazars as a universal process with different sources [134].

plasma accretion in the gravitational potential wells reaching up velocities of a few $10^3$ km/s. There are indications ([146, 164]), that these accretions could form strong shock waves on Mpc-scales. The problem with diffusive shock acceleration in these structures arises from the fact, that the magnetic field in the accretion region is not larger than 100 nG, that would predict a gyro-radius $r_g \sim 1$ Mpc for a $10^{20}$ eV proton. For a magnetic field perpendicular to the shock-front the mean free path is expected to be larger than the gyro-radius. This would require unreasonable large diffusive regions and an escape time in the order of the age of the universe. As way out some authors [146, 164, 95] suggest, that particles could be accelerated in these shocks by Jokipii diffusion. Here the magnetic field is perpendicular to the shock-front's normal and should decrease the mean free path in that way that it becomes much smaller than the gyro-radius. The maximum energy can be larger than $10^{17}$ eV but should be smaller than a few $10^{18}$ eV [146]. The presented idea can *mutatis mutandis* be applied to processes taking place in colliding galaxies. However, it is very unlikely to reach the same maximum energy as in large super galactic structures [143].

**Wind- and shock-acceleration of rotating neutron stars:** Pulsars[11] have been discussed as galactic sources of the UHECR. These objects begin their lifes as fast rotating neutron star ($\omega \sim 3000\,Hz$) [9] with large magnetic fields ($B_{NS} > 10^{13}\,G$). Inside the light cylinder a magnetosphere of density

$$n_{GJ}(r) \sim \frac{B(r)\omega}{4\pi Zec}$$

co-rotates with a magnetic dipole field component scaling with $B(r) \sim B_{NS}(r_{NS}/r)^3$. Here $r$

[11] fast rotating neutron stars, which emit detectable synchrotron radiation.

is the radius in the light cylinder, e.g. $r < c/\omega$ and $r_{NS}$ the neutron star radius. With increasing distance from the neutron star surface the dipole field cannot be maintained and beyond the light cylinder a mostly azimuthal field can be assumed. Thus, one expects that from the light cylinder a relativistic plasma with Alfvén speed expands as magneto-hydrodynamic wind. As the neutron star is produced in a supernova one considers the surface of the young neutron star enriched with elements up to the iron peak. Those ions can strip of and be shock-accelerated in the winds to maximum energies of [39]

$$E_{max} = \frac{ZeB_{lc}R_{lc}}{c} \sim 8 \cdot 10^{20} Z_{26} B_{13} \Omega_{3k}^2 \,\mathrm{eV} \,.$$

**Strong magnetic field acceleration in rotating neutron stars:** The rotating magnetosphere of neutron stars is expected to induce a strong electric field of the order of $c^{-1}|(\Omega \times R) \times B|$. Here $R$ is the stellar radius of the neutron star, $B$ the field-strength and $\Omega$ the opening angle of the light cone. This would lead to a potential of the order of ([143]):

$$\Delta\Phi \sim 10^{20} B_{13} R_6^2 / T_{-3} \,\mathrm{V} \,.$$

This potential looks promising to accelerate particles to highest energies, nevertheless due to pair production in the pulsar's magnetosphere the $e^+e^-$ pairs short the circuit and reduce the effective voltage by a few orders of magnitude.

### 2.2.5 Top-down models - Particle Physics Scenarios

The basic idea of top-down scenarios is to introduce a super-massive relict particles. This can be: topological and non-topological solitons, X-particles, cosmic defects, microscopic black-holes, fundamental strings. To explain the UHECR one commonly classifies between two cases:

**Decay Top-Down Scenarii:** In case the particles are unstable they decay and among their decay products the UHECR are produced An unified formula for the quantum decay rate of all these objects was provided in [63]. For all relics one has to fine tune the lifetime of these objects to be the age of the universe and the value of their mass that must be adjusted to be larger than $10^{20}$ eV. Heavy relics could have been formed by the end of inflation at typical GUT's energy scales.

**Annihilation Top-Down Scenarii:** It has been further proposed that stable heavy relics can produce UHECR through annihilation by pairs [53]. In this scenario, the relics are assumed to have masses around $M_X \sim 10^{12} GeV$ and produce EHECR through annihilation when they collide. Here, the lifetime free parameter is replaced by the annihilation cross section. These super-heavy particles are assumed to be produced during reheating.

The decay or annihilation of the relict particles would produce, among other fragments, quarks and leptons. While the quarks hadronise in jets the leptons decay. The hadronisation of the quarks

generates many pions. Their decay produce photons, neutrinos and light leptons via:

$$\pi^0 \rightarrow 2\gamma \tag{2.6}$$

$$\pi^\pm \rightarrow \mu^\pm + \nu_\mu(\bar{\nu_\mu}) \rightarrow e^\pm + \nu_\mu + \bar{\nu_\mu} + \nu_e(\bar{\nu_e}) \tag{2.7}$$

One would expect a certain photon fraction in the EHECR[12] composition. This signature makes the hypothesis testable. Although, current analysis are not able to rule out these scenarios completely the parameter space for possible scenarios is strongly constrained by current results from the Auger Collaboration [156, 157, 158].
Another signature of top-down scenarios is that the spectra of the particles generated in the top-down models are typically flatter than the bottom-up ones. Contrary, to the acceleration mechanisms, the top-down generated spectra do not follow a power-law.

### 2.2.6 Propagation Models

Up to $10^{18}$ eV the propagation of cosmic rays is supposed to be related to the structure of our galaxy. Once accelerated several physical processes take place. These are the transportation along magnetic field lines, diffusion on the irregularities of the magnetic field, spallative reactions with the interstellar gas[13] and radioactive decay. Following the diffusive transport equations ([79]) which govern both, acceleration and propagation of cosmic rays can be written as:

$$\frac{dN_i(E)}{dt} = -\underbrace{c_i N_i(E)}_{\text{loss}} + \underbrace{\sum_i \int_E^\infty dE_j N_j(E_j) f_{ji}(E, E_j)}_{\text{gain}} - \underbrace{\nabla(D_i \nabla N_i)}_{\text{convection}} + \underbrace{\nabla v_C N_i(E)}_{\text{diffusion}} - \underbrace{\frac{\partial}{\partial E}(N_i a_i)}_{\text{acceleration}} + \underbrace{Q_i}_{\text{source}}$$

Here, $c_i$ gives the collision and decay probability and $f_{ij}$ gives the probability that a particle j with energy $E_j$ is converted to a particle i with energy $E_i$. $v_C$ is the convection velocity and $D_i$ the convection coefficient for a particle i. $a_i$ describes the acceleration of particle i.
A common simplification of this propagation description is the leaky box model. Here it is assumed that the particles propagate freely along the field lines in a containment volume[14] and at each encounter with the boundaries they have an energy dependent probability to escape to the extragalactic space. The density is fixed to a constant value within the confinement volume. The escape time, $T_{\text{esc}}$ is related to the escape length $\lambda_{\text{esc}}$ via $\lambda_{\text{esc}} = \rho\beta c T_{\text{esc}}$. The energy dependence of the ratio of secondary to primary particles can then be fitted to high energy data with $\lambda_{\text{esc}}$ as free parameter. This model predicts that the energy spectrum steepens by propagating from the source to the observer. To quantify the effect, an additional small value $\delta$ is introduced, in the way that

$$\Phi = \Phi_{\text{source}} \cdot E^{-\delta} = \Phi_0 \cdot E^{-(\gamma+\delta)} \ .$$

[12] Extremely high energy cosmic rays, that are considered in the scope of this work events with energies above $10^{18}$ eV
[13] This is discussed in context with Fig. 2.2
[14] our galaxy

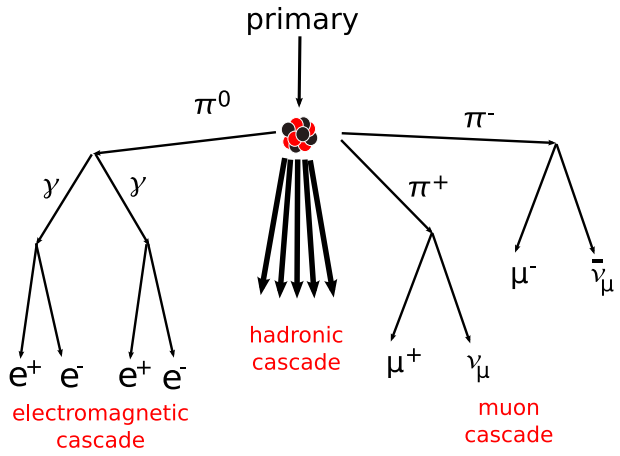



**Figure 2.10:** The three components of an extensive air shower

With a usually source spectral index of $\gamma_{\text{source}}$, $\delta$ has to be around 0.6. The situation changes around the knee, in this energy regime the gyro-radius of the particles exceeds the size of the galactic disk and it is assumed that the galactic fields cannot longer bind the particles. Here, it is supposed, that the diffusion of the galaxy becomes dominant and one would expect to see a rapidity dependent cutoff in the energy spectra of the different components. However the question of rapidity or mass dependence is still matter of discussion, first results from the KASCADE experiment encourage the rapidity dependent picture [186].

## 2.3 Air Shower Physics

The flux of CR beyond $10^{14}$ eV is not large enough to permit direct measurements of the primaries via balloon or satellite borne devices. CR particles at these energies are detected indirectly using experimental devices located on ground. These experiments detect the secondary particles of a cascade that is initiated by the primary particle interacting with the Earth's atmosphere. These cascades are called extensive air showers. After their discovery by Pierre Auger 1938 the understanding of the physics of air showers has reached a fairly detailed level.

### 2.3.1 Air Shower Development

An air shower starts developing when a primary particle enters the Earth's atmosphere. Usually, the first interaction is an inelastic nuclear collision producing several secondary particles, which interact and also produce secondaries similar to their predecessors.
Usually one distinguishes three different components of the cascade, connected to different interaction processes (as sketched in Fig. 2.10):

**Hadronic component:** Inelastic collision of hadrons with air nuclei of the atmosphere produce usually high energetic hadronic secondaries, which again interact or decay. This component is mostly, dominated by mesons, like $\pi$ or $K$.

**Electromagnetic component:** Simplified, about 1/3 of the hadronic secondaries are $\pi^0$. The photons from the decay $\pi^0 \to \gamma + \gamma$ initiate an electromagnetic cascade[15]. The photons produce lepton pairs, most dominantly via pair-production. These electrons emit bremsstrahlung[16]. These photons then produce new $e^+e^-$-pairs and the process continues. This process ceases when the electrons have reached the critical energy $E_{\text{crit}}$ at which the ionisation loss in air dominates over the loss due to bremsstrahlung [64, 161],

$$E_{\text{crit}} = \frac{710}{Z + 0.92}\ \text{MeV} \sim 86\,\text{MeV}.$$

Here $Z = 7.3$ is the effective charge number of the atmosphere. Via ionisation the electromagnetic cascade dissipates about 90 % of the primary energy.

**Muon component:** Muons are produced in the decay of $\pi^{\pm}$ and $K$-mesons from hadronic interactions. These processes are also connected to neutrino production. Neutrinos are not detectable[17] and result in an invisible energy loss. Muons contribute little to the further development of the cascade and remove energy from the cascade development. The majority reaches ground or decays.

### 2.3.2 The Longitudinal Shower-Profile

The longitudinal shower profile describes the number of particles $N(X)$ as function of the atmospheric depth $X$. The longitudinal evolution of the shower is dominated by the electromagnetic cascade[18]. The number of particles increases with the shower depth and reaches its maximum at a characteristic depth $X_{max}$, when the electromagnetic component starts to attenuate. The total integrated energy deposit is a good calorimetric measure [19], while $X_{max}$ is a measure of the primary particle mass. The mass dependency of $X_{max}$ for a fix energy can be understood in terms of the so called superposition model. Here, a primary of mass $A$ and energy $E$ is considered as bunch of $A$ independent protons of the energy $E/A$. One would expect that in the average this shower profile reaches its maximum higher in the atmosphere as a proton induced shower of energy $E$. More quantitatively, the shower development can be investigated by the Heitler model [89]. This toy model neglects multiple scattering and Compton scattering effects and characterises the electromagnetic shower development by bremsstrahlung and pair production. It is assumed, that

[15] One should mention, that the pion can also decay via $\pi^0 \to \gamma + e^+ + e^-$ or $\pi^0 \to e^+ + e^- + e^+ + e^-$. But the rate is suppressed, and for the description of the electromagnetic component these processes can be neglected.
[16] Additionally, deflected in the geomagnetic field they can emit synchrotron photons
[17] The Pierre Auger Observatory cannot detect neutrinos. Experiments, like AMANDA/ICECUBE, are able to detect the neutrino component.
[18] even more for photon induced showers, as they can only produce muons via photon-air interactions
[19] 90 % of the primary energy is dissipated by the electromagnetic cascade, compare previous section.

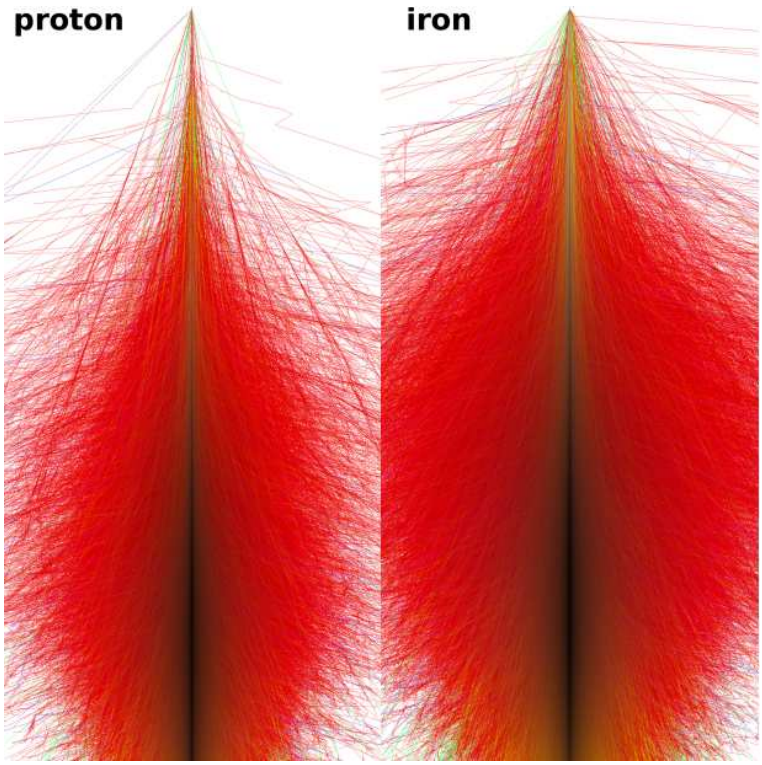


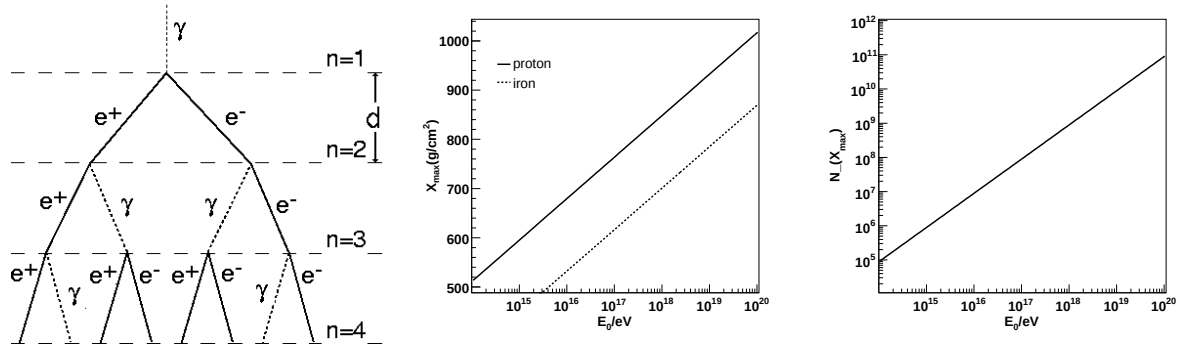

**Figure 2.11:** The longitudinal shower development for a $10^{15}$ eV vertical proton and iron induced shower simulated with CORSIKA. Taken from [166]. The red tracks indicate the electrons and positrons, the green ones represent the muon traces, while the blue ones show the hadronic component. It is visible that the iron induced shower develops higher in the atmosphere than the proton induced one.

the photon energy in pair processes is shared between both leptons. Furthermore, the energy of an electron in a bremsstrahlungs-processes is shared between the photon and outgoing electron. A scheme of this procedure is shown in Fig. 2.12(a). An electron is supposed to emit a photon after travelling a splitting length $d = ln2X_0$, where $X_0$ is the radiation length of air. This photon produces a $e^+e^-$ pair after travelling another $d$. Thus, the shower expands like a tree, that bifurcates every grammage $d$. Therefore, the number of particles at a certain depth is

$$N(X) = 2^{d/X_0} \ .$$

The average energy per particle is assumed to be

$$E(X) = \frac{E_0}{N(X)} \ .$$




**(a)** A schematic sketch of the Heitler formalism for the electromagnetic cascade.

**(b)** Predicted $X_{max}$

**(c)** Predicted $N_{max}$

**Figure 2.12:** Predictions for a pure electromagnetic cascade in the Heitler toy model.

This bifurcation continues until the critical energy is reached. Therefore

$$X_{max} = X_0 ln(E_0/E_C) \ .$$

Reasoning with the superposition principle, the following trend becomes obvious:

$$X_{max} \propto ln(E_0/A) \ .$$

Although this model is a very simple picture, it reproduces the average trends in the longitudinal shower development:

- Fig. 2.12(b)) shows $X_{max}$ versus the primary energy. This representation is commonly referred to as elongation rate and allows to draw conclusions on the average mass composition. Generally, one can see, that

  $$X_{max} \propto \left[\ln \frac{E_0}{E_C} - \ln < A >\right]$$

  This corresponds with detector data and full EAS[20] Monte-Carlo (MC) simulations [150, 128].

- The number of particles in the shower maximum scales linearly with the energy. Nevertheless, this toy model overestimates the number of electrons compared to photons that converge in this simple ansatz to a ratio of 2:1. This is not surprising as the electrons are expected to produce often multiple bremsstrahlungs photons, and many electrons and positrons range out in the air.

[20] extensive air-shower

For complete and precise modelling of the shower development, full Monte-Carlo simulation of all interaction processes is needed, including also Compton scattering, photoelectric effects and electron-positron annihilation, as shown in Fig. 2.11
Commonly, a semi-empirical parametrisation is fitted to the shower development: the so called Gaisser-Hillas parameterisation [71],

$$N(X) = N_{max} \left( \frac{X - X_0}{X_{max} - X_0} \right)^{(X_{max} - X_0)/\lambda} \exp\left( \frac{X_{max} - X_0}{\lambda} \right) . \tag{2.8}$$

Here, $N_{max}$, $X_0$, $X_{max}$, $\lambda$ are treated as free fit parameters. $X_0$ is the starting point of the Gaisser-Hillas curve, $X_{max}$ is the depth of its maximum, $\lambda$ is the interaction length of the primary particle and $N_{max}$ the maximum shower size.

### 2.3.3 The Shower Footprint on Ground

While propagating through the atmosphere the shower-front evolves conically around the primary trajectory. This is known as leading particle effect and allows to track the primary. The particle density in the perpendicular plane of the shower-axis is expected to be radial symmetric and steeply falls with the distance from the shower impact point. The lateral density at a predefined distance can be used as measure for the primary energy. If it is possible to distinguish between the muonic and electromagnetic component of the particles at ground[21] this information can be used to unfold the primary mass compositions of the showers.

At ground, the shower front consists mostly of photons, electrons and muons. The majority of relativistic hadrons is concentrated around the core and can be neglected 50 m away. Usually, the soft component thins out and the electrons arriving on ground have much lower energies than the muons[22]. The most dominant energy loss is due to bremsstrahlung. The interaction cross-section scales with $m^{-4}$. Therefore, the heavy muons can penetrate much deeper into the atmosphere than electrons (see Fig. 2.13). Therefore, muons are also called the hard component of a shower. For low energy primaries only the hard component reaches ground. The electromagnetic and hadronic component is also referred to as soft component of a cascade as these particles tend to deposit far more energy per atmospheric depth than the muons and neutrinos. Thus, from the footprint one can distinguish a young shower with a electron dominated shower-front far away from the core and an old shower, that shows a muon dominated component far away.

## 2.4 Detection Techniques

Large detection volumes are needed to detect particles at reasonable rate above $10^{14}$ eV (see Chap. 2.3). The only chance in the UHECR regime is to detect a cosmic ray particle indirectly

[21] e.g. using muon counters

[22] E.g., for a $10^{19}$ eV proton induced shower 900 m away from the shower core electron energies range in average around a few 10 MeV while muons have energies usually around a few GeV.

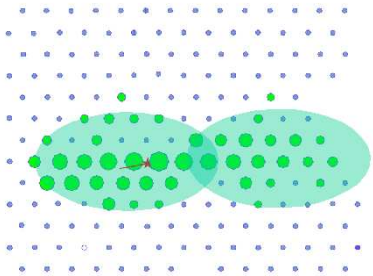


**Figure 2.13:** The muonic component especially in horizontal showers travels long enough to make a separation of the positive and negative charged muons visible in the footprint

by its induced air shower. Several detection techniques allow to complementary measure both the longitudinal development of the shower traversing the atmosphere and the lateral distribution and composition of the footprint, that is produced by the shower front on ground. While fluorescence telescopes can see the longitudinal shower evolution, water Cherenkov detectors and muon counter arranged in coincidence in a large arrays can see the lateral footprint. A relatively new technique allows also the detection of EAS by their radio emission.

### 2.4.1 Air Fluorescence Detection

Most of the particles' energy is lost via ionisation and excitation of the atmospheric molecules by the electromagnetic cascade (see Chap. 2.3). Using the atmosphere as calorimeter one can obtain the longitudinal profile of an EAS, visible in UV-fluorescence light of excited molecules. The requirement to detect an EAS by its fluorescence light is the capability to detect a very weak signal of a few photons on a noisy night sky background within a few microseconds.[23] Thus, this detection technique is limited to operate only during clear nights. Furthermore, the pure fluorescence signal from the EAS is contaminated by direct and scattered Cherenkov light. In the following the fluorescence light emission, the Cherenkov contribution and the light propagation through the atmosphere are described, to make the reconstruction-procedure of fluorescence telescope data more transparent.

#### Fluorescence Light Emission

The electrons of the electromagnetic cascade loose their energy via continuous ionisation and excitation of the air molecules, while the excited molecules itself dissipate the energy gain via non-radiative collisions or internal quenching[24]. The fluorescence emission spectrum of the exited ni-

[23] The need to detect a $10^{18}$ eV EAS 30 km away from the detector is comparable to detect a 60 W bulb falling through the atmosphere with speed of light.

[24] Here, quenching is meant to be the transfer of electron excitation to a higher vibration mode in connection with IR emission.

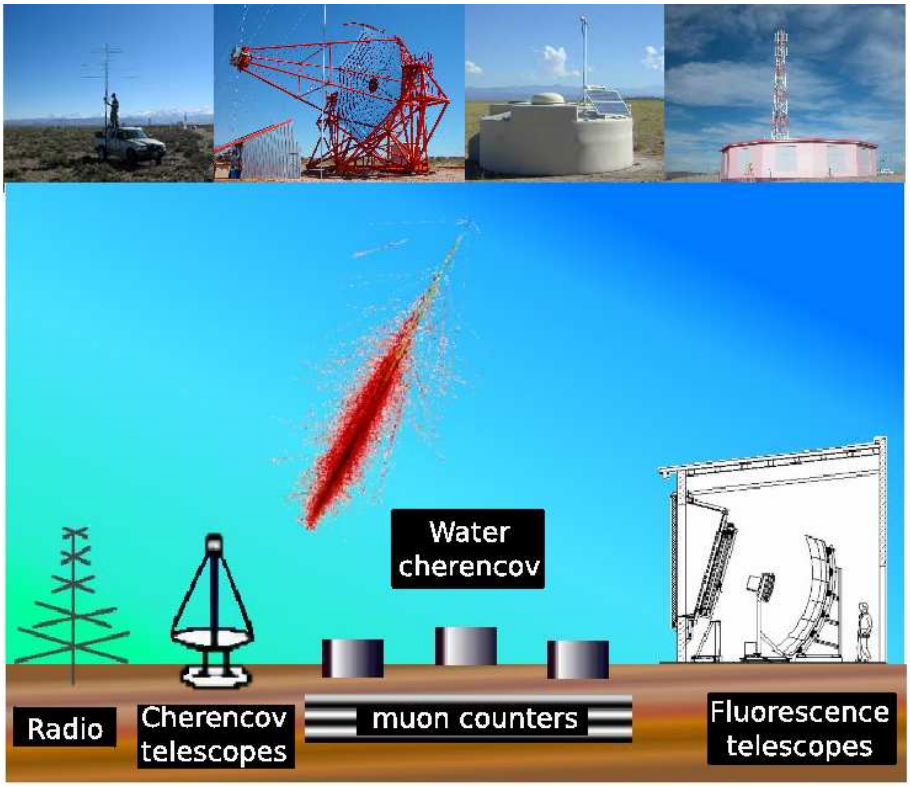



**Figure 2.14:** Indirect measurement techniques to observe a cosmic ray air shower.

trogen molecules is shown in Fig. 2.15 and covers the UV range between 300 nm up to 400 nm. The efficiency of radiating energy via this process compared to other processes is in the order of 0.005 % for UHECR showers. However, this emission becomes rather significant. The emission is isotropic and proportional to the number of charged particles in the shower. The $N_2$-molecules have a band spectrum, in which up to now 19 bands have been resolved. 18 of them are in the second positive system (2P) and on in the first negative system (1N) [103]. In the following we focus on the effect of electron excitation of the $N_2$-molecules, as this is the dominant process in air-showers. Taking the mean lifetime for a radiative transition to a ground state $\tau_0$ and taking also the mean lifetime for non radiative de-excitation via quenching $\tau_c$ into account, the quantum efficiency for the fluorescence emission can be written as

$$\frac{\text{radiative de-excitations}}{\text{all de-excitations}} = \frac{\tau_0}{\tau_0 + \tau_c} \,.$$

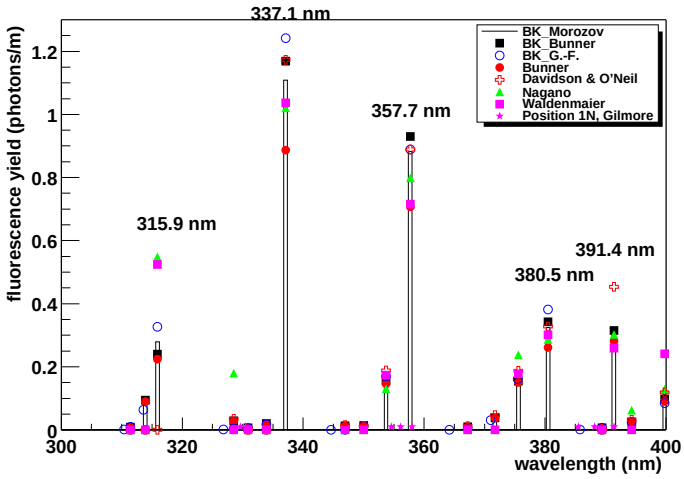



**Figure 2.15:** Fluorescence yield spectra, measured and calculated for 0.85 MeV electrons as excited particles. Taken from [102], see also [103]

Quenching enhances with the rate of molecule collisions, and thus is depending on the mean velocity of molecules $\langle v \rangle$. Using the ideal gas approximation $\langle v \rangle = \sqrt{8kT\pi^{-1}M^{-1}}$ one obtains

$$\tau_c = \frac{\sqrt{kT\pi^{-1}M^{-1}}}{4\rho_n \sigma_{\mathrm{NN}}} \,.$$

Here, $\rho_n$ is the particle number density, $\sigma_{\mathrm{NN}}$ the cross section for nitrogen-nitrogen collisions, and $M$ the mean molecular mass. By this, the efficiency for the emission of fluorescence-energy can be expressed:

$$\epsilon_\lambda(p,T) = \frac{\text{energy deposit into radiation}}{\text{total energy deposit}} = \frac{\epsilon_\lambda(p \to 0)}{1 + (p/p'_\nu(T))} = \frac{nE_\gamma}{E_{\mathrm{dep}}} \,.$$

$\epsilon_\lambda(p \to 0)$ is the efficiency to radiate fluorescence light of wavelength $\lambda$ in absence of collisional quenching. $n$ is the number of fluorescence photons. $p$ is the pressure of air, while $p'$ is a reference pressure. In a realistic scenario, assuming at least a two-component gas (N and O), the pressure relation for a certain excitation level $\nu$ can be expressed via the lifetimes [105]:

$$\frac{p}{p'} = \tau_\nu \cdot \left( \frac{1}{\tau_{\mathrm{NN},\nu}} + \frac{1}{\tau_{\mathrm{ON},\nu}} \right)$$

$\tau_{\mathrm{NN},\nu}$ is the mean lifetime of collisional de-excitation between two nitrogen molecules and $\tau_{\mathrm{NO},\nu}$ the corresponding one for nitrogen-oxygen. These lifetimes are functions of the cross-sections for the collisions, the mean molecular mass and atmospheric compositions. An important number is

the fluorescence yield $N_f$ which defines the number of fluorescence photons emitted per unit length and number of charged particles.

$$\frac{dN_\gamma}{dl} = N_f(p,T)N_e \ .$$

The fluorescence yield can now be described by the fluorescence efficiency [105]:

$$N_f = \epsilon_\lambda(p,T)\frac{\lambda}{hc}\frac{dE_{\text{ion}}}{dN}\rho_{\text{air}} \ .$$

The ionisation energy loss $\frac{dE_{\text{ion}}}{dN}$ is about $0.1677\,\text{MeV}\,\text{kg}^{-1}\,\text{cm}^{-2}$.

### Cherenkov Radiation

Nearly all charged particles in an EAS carry enough energy to emit Cherenkov light. In a region between the shower axis and 25° off-axis the direct Cherenkov light emission dominates above the fluorescence light emission. The number of Cherenkov photons can be estimated to follow [22]:

$$\frac{dN_\gamma}{dl} \sim 33N_eF(1.57E_s)\exp(-h/H_0)\,\text{m}^{-1} \ .$$

Here $N_e$ is the number of electrons, $F$ the fraction of electrons with an energy exceeding the threshold to emit Cherenkov radiation. $H_0$ is an atmospheric scale factor and $h$ is the production height. More than 25° away from the shower axis the Cherenkov contamination can be estimated by folding the Cherenkov production rate with the angular distributions of the electrons:

$$\frac{d^2N_\gamma}{dld\Omega} = \frac{dN_\gamma}{dl}\frac{\exp(-\theta/\theta_0)}{2\pi\sin(\theta)} \ .$$

The characteristic angle $\theta_0$ is a function of the threshold energy for Cherenkov radiation and therefore scales with the primary energy:

$$\theta_0 \sim 0.83/E^{0.67} \ .$$

### Light Propagation Through the Atmosphere

While propagating through the atmosphere the emitted light undergoes multiple scattering processes and is attenuated. A key role of air fluorescence detectors is to understand each of these processes and to adjust the parameters on nightly changing atmospheric conditions, because the light emission at the shower has to be counted back precisely from the light detected at the diaphragm.

**Rayleigh scattering:** Rayleigh scattering describes the scattering of light at the molecules of the air, and can be extended to scattering at particles up to about a tenth of the wavelength of the light. For the energy regime of Cherenkov and fluorescence light ($\sim 400\,\text{nm}$) the interaction length of UV photons in air at sea level has been measured to be $2.3\cdot10^4\,\text{m}$. This corresponds

to a grammage[25] of nearly $3000\,\mathrm{gcm}^{-2}$. Assuming a beam with $N_\gamma$ photons the loss due to Rayleigh scattering can be calculated via

$$\frac{d^2N_\gamma}{dld\Omega} = -\rho\frac{N_\gamma}{3000\,\mathrm{gcm}^{-2}}\frac{400}{\lambda}\frac{3}{16\pi}(1+cos^2\theta)\,.$$

Rayleigh scattering can be considered elastic since the energy of the molecules is not changed. Scattering in which the scattered photons have either a higher or lower photon energy is called Raman scattering. Usually, this kind of scattering involves the excitation of some vibrational modes of the molecules. This results to a lower energy for the scattered photons, or scattering at an excited vibrational state, which adds its vibrational energy to the incident photon.

**Mie scattering:** In contrast to Rayleigh scattering, Mie scattering parameterises processes of light being scattered at objects in the order of the wavelength. This happens in the atmosphere while the light propagates through aerosols. A good approach[26] is to describe the amount of Mie scattered photons by a two parameter model assuming an exponential decrement with the altitude.

$$\frac{d^2N_\gamma}{dld\Omega} = -0.8\frac{N_\gamma}{L_{\mathrm{Mie}}}\exp(-h/H_{\mathrm{Mie}})\exp(\theta/26.7^\circ)\,.$$

The two parameters are referred to scale height $H_{\mathrm{Mie}}$ and horizontal attenuation length $L_{\mathrm{Mie}}$ and have to be adjusted to the nightly conditions.

**Attenuation:** The detected light has to be corrected for attenuated photons including the transmission factors of both effects, Mie and Rayleigh scattering. The light intensity at a detector observing an angular interval of $\Delta\Omega$ can be estimated by using the transmission coefficients of both processes, $T_{Ray}$ and $T_{Mie}$:

$$I = I_0T_{\mathrm{Ray}}T_{\mathrm{Mie}}(1+O_{\mathrm{multi}})\frac{\Delta\Omega}{4\pi}\,.$$

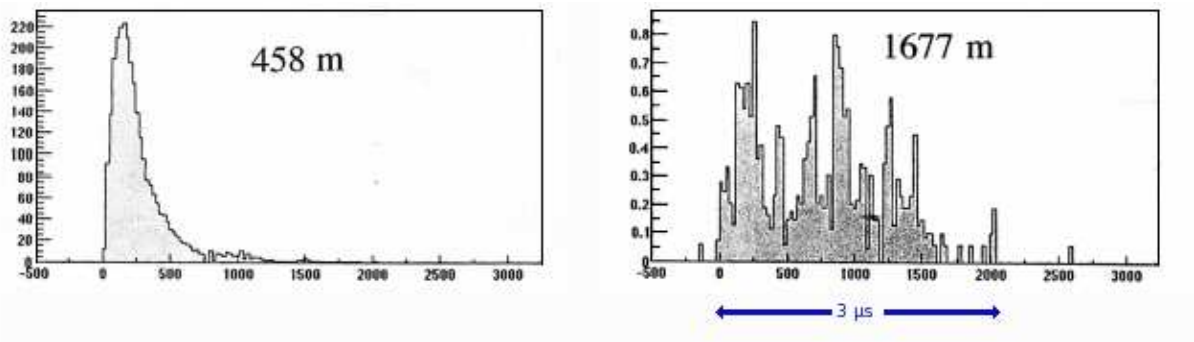
$O_{\mathrm{multi}}$ are higher order corrections due to multiple scattering.

Cherenkov photons undergo also multiple scattering and it is common to distinguish between the direct and indirect Cherenkov contribution. The indirect Cherenkov light is emitted at higher altitudes and can significantly dominate over the fluorescence light that is emitted more deep in the atmosphere, where the electromagnetic component start attenuating.

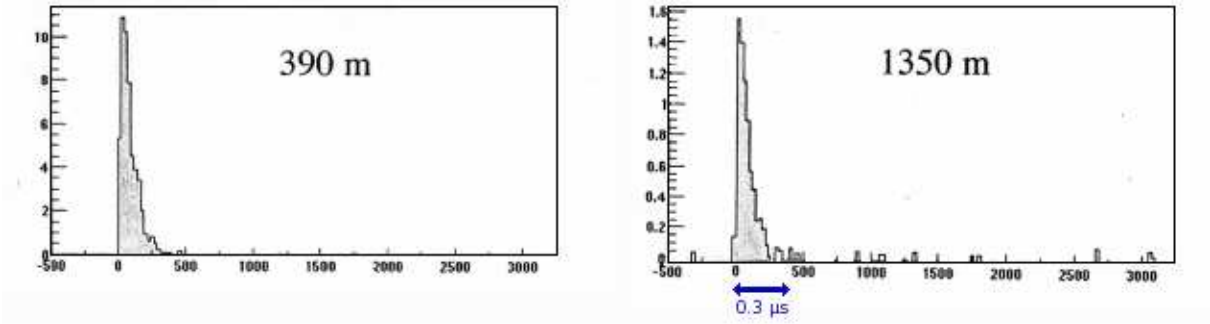
### 2.4.2 Water Cherenkov Tanks

Complementary to the longitudinal information obtained by air fluorescence telescopes, the electromagnetic and muon component can be detected, when the EAS strikes ground level. A very

[25] using US standard atmosphere
[26] at least in the range between 5° and 60° zenith




(a) Young, vertical shower




(b) Old, horizontal shower

**Figure 2.16:** While the density decreases by a factor 134 with the distance increasing by a factor 3.7 for young showers, the signal decreases by a factor 7.5 for distance increasing by a factor 3.5 for old showers.

probed technique is to use water Cherenkov detectors. These devices consist of a large volume of pure water. Particles of the shower front can produce Cherenkov light when traversing the water, as most of them travel much faster than the speed of light in water. Monitoring the water allows to observe the signal. Usually, many of these devices are arranged to a surface array. A clear signature of an EAS is a trigger of coincident signals. The technique itself does not directly allow to distinguish the muon component from the electron component, that passes through the water Cherenkov detector. As discussed in Chap. 2.3.3, this information is a measure for the primary composition. However, the time structure and rise-time of the signals can be used to distinguish between large signals from penetrating muons and small signals from absorbed electrons (see Fig. 2.16).
The shower plane can be considered slightly curved like a sphere. Particles far away from the core arrive delayed because of geometry reasons. Thus, particles arriving early at the detector originate to higher altitudes in the atmosphere due to a slightly smaller path. The electromagnetic compo-

nent diffuses far away off the shower axis. The time spread of the signal far off-axis is proportional to the distance from the shower axis. In this picture it becomes obvious, that the time-spread becomes larger with the shower maximum being deeper in the atmosphere. In addition, muons tend to arrive earlier than electrons as they are not affected so much by scattering. This allows to measure the primary mass composition by water Cherenkov information:
Iron induced showers develop higher in the atmosphere and therefore contain more muons compared to proton induced showers. Thus the signal far away from the core is expected to be shorter.

### 2.4.3 Muon Counters

Additionally, some UHECR experiments use muon counters. E.g., the KASCADE experiment uses a large central detector with a hadronic calorimeter and muon detectors with different threshold energies, as well as a muon tracking detector in an underground tunnel. This allows to quantify the number of electrons and muons separately for most showers and therefore unfold the primary composition spectra.

### 2.4.4 Radio

A very young experimental technique to detect air shower properties is radio detection of EAS. The idea of radio emission was originally predicted by Askaryan [17, 18] in 1962 and experimentally discovered by Jelley [94] in 1965 for dense transparent media like ice. The idea for coherent radio emission of air showers is the following:
The electrons and positrons of the electromagnetic cascade are deflected by the Earth's magnetic field and emit synchrotron radiation, often referred to as geo-synchrotron radiation.
The radio signal can be detected by an array of dipole antennas and the air shower can be reconstructed. The frequency belt is between a few up to 100 MHz and can be measured by standard RF techniques. The advantage of this technique is that the radio antennas can operate with 100 % duty cycle, similar to the water Cherenkov detectors.

### 2.4.5 The Concept of Hybrid Detection

Hybrid detection refers to a measurement of an EAS using two different detection techniques. E.g., a configuration of a surface array of water Cherenkov tanks and fluorescence detectors permits the measurement of the lateral as well as the longitudinal development of an EAS. The capability to detect the shower independently with complementary detectors has the advantage to include well established detection techniques and use the coincident measures to reduce the individual systematic disadvantages of each single technique. The ground array technique using water-Cherenkov tanks was first successfully applied in the Haverah Park experiment and operated there for over 20 years [21]. The energy calibration of the lateral density profile is still based on MC simulations and thus model dependent. The fluorescence technique was first successfully adapted by the Fly's Eye experiment [38]. As the fluorescence technique allows to measure directly the energy deposition in

the atmosphere, the energy reconstruction is almost model independent and mostly biased by uncertainties in the fluorescence yield. Fig. 2.17 schematically demonstrates the hybrid measurement of the Pierre Auger Observatory (see Chap. 3). This concrete concept aims the following issues:

**Cross-validation of the reconstruction schemas:** Independent measurements by both subsystems allow the reconstruction of the shower characteristics independently for a subset of showers and a better understanding of the systematic effects of each reconstruction scheme.

**Model-independent energy calibration of the SD:** The ground array can trigger EAS with almost 100 % duty cycle. As described in Chap. 2.3.3, the energy measure of the shower's footprint is given by the extrapolated signal at a fixed distance from the shower core. Simulations of the lateral shower development are needed to scale the energy. Thus, the energy scale depends on the hadronic interaction model used for these Monte Carlo simulations. The subset of hybrid events can be used to scale the energy measure of the events detected by the surface array directly with experimental data. Furthermore, the fluorescence telescope energy reconstruction benefits from a better geometry reconstruction, due to the additional information from the ground array.

**Higher sensitivity to composition analysis:** As explained in Chap. 2.3.2 and Chap. 2.4.2, the atmospheric depth of the shower maximum and the rise-time of the signals in the water Cherenkov tanks can be used as measures of the primary composition. The first observable is measurable with the fluorescence telescopes, while the second one is detected by the water Cherenkov tanks. The subset of hybrid events offers the opportunity to correlate both variables in coincidence, and to study the UHECR composition with improved accuracy.

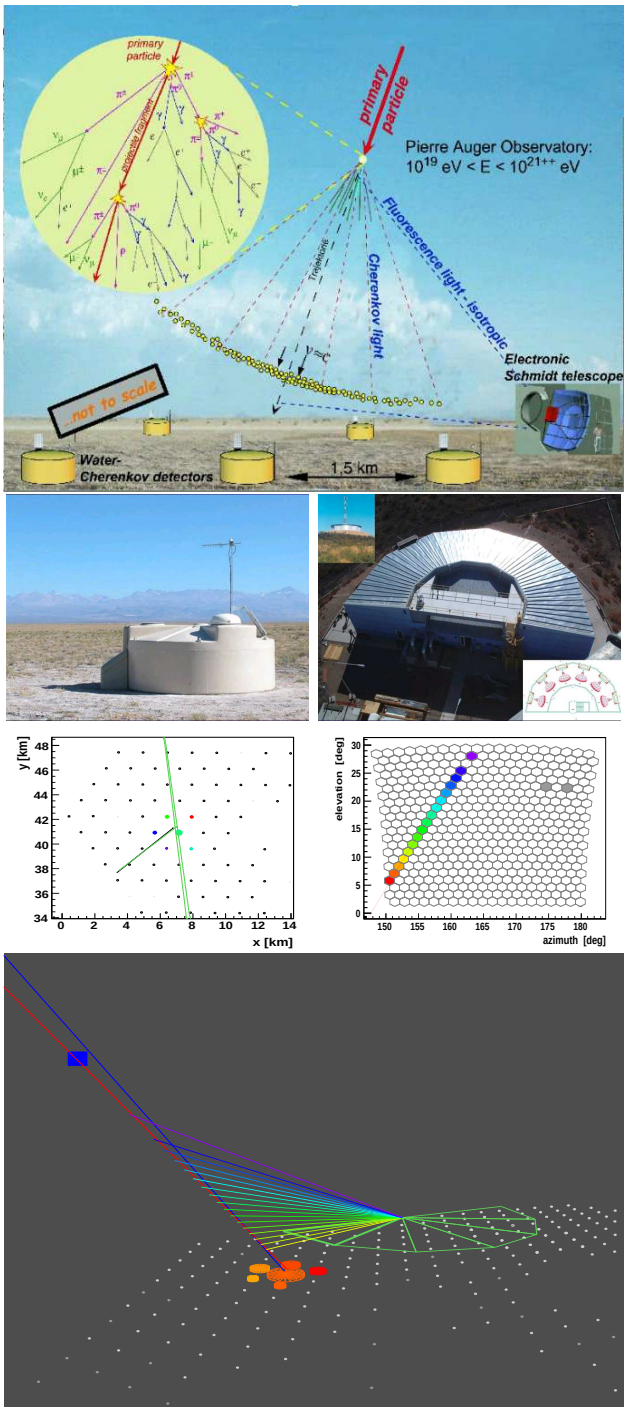



**Figure 2.17:** Hybrid detection techniques: The hybrid concept of the Pierre Auger Observatory

# Chapter 3

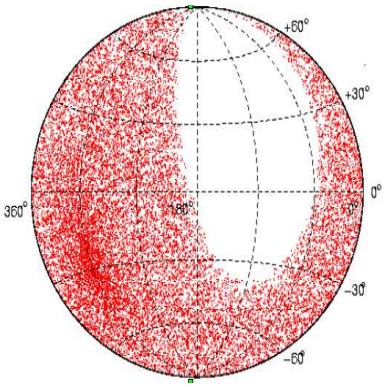
# The Pierre Auger Observatory

*The greatest lesson, however, is a moral and philosophical one. It is truly awesome to discover, by example, that we humans can come to comprehend Nature's s deepest principles, even when they are hidden in remote and alien realms. Our minds were not created for this task, nor were appropriate tools ready at hand. Understanding was achieved through a vast international effort involving thousands of people working hard for decades, competing in the small but cooperating in the large, abiding by rules of openness and honesty. Using these methods which do not come to us effortlessly, but require nurture and vigilance we can accomplish wonders.*

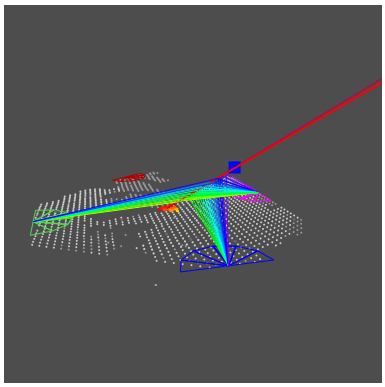
Frank Wilzcek

The Pierre Auger Observatory is a multi-national project for research on the cosmic rays at highest energies. The aim of the experiment is to determine the energy, composition and origin of the UHE cosmic-rays above $10^{18}$ eV using two complementary detection techniques: water Cherenkov detectors and fluorescence telescopes.

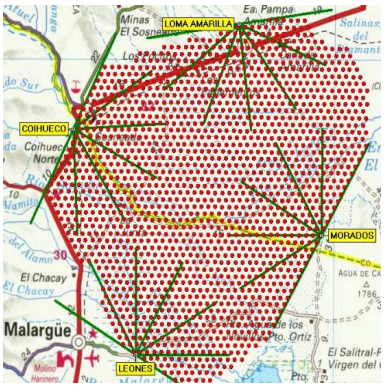
In clear nights, this configuration permits the measurement of the lateral as well as the longitudinal development of an EAS (hybrid event). The fluorescence detector (FD) aims not only for the calibration of the energy scale of the experiment, but also in the understanding of the systematics of the experiment by using the data itself.




(a) Current sky coverage using the southern hemisphere array



(b) Hybrid events allow to calibrate the energy scale of the surface array




(c) Using a large instrumented area of 3600 $km^2$ and four fluorescence telescopes to provide large aperture.

**Figure 3.1:** The main aims of the Pierre Auger Observatory .

The concept of the experiment was guided by the following requirements ([177], see Fig. 3.1):

**Full sky coverage:** To achieve this requirement the design foresees two experimental sites, one on the Northern hemisphere in Colorado and one on the Southern hemisphere in the Argentinian Pampa. While the Northern site is facing its build up in 2009[1] [139], the Southern observatory will be completed by February 2008. Experimental data are taken since 2002 and the data analysed in the present work are based on the Argentinian site.

**Large aperture:** As previously mentioned the flux of cosmic rays in the GZK regime is extremely low. The rate of particles with an energy larger than $10^{19}$ eV is about $1\,\mathrm{km}^{-2}\,\mathrm{yr}^{-1}$. Thus, it is necessary to instrument a sufficiently large area to achieve significant statistics in a reasonable measurement time. The Southern site will instrument an area of about $3000\,\mathrm{km}^2$ with surface detectors, while the Northern site is planned for $10000\,\mathrm{km}^2$. For the Southern array this would allow to detect up to 3000 cumulative events per year above the energy region of the ankle. About 30 cumulative events per year are expected in the region beyond the predicted GZK cutoff. This is enough to solve the GZK-puzzle with reasonable significance.

**Hybrid concept:** On each site the ground area will be equipped with water Cherenkov detectors, while the atmosphere above the surface array is observed in clear nights by fluorescence detectors. Thus, the FD can operate with about 10 % duty cycle compared to the surface array. This allows to gain a subset of events with hybrid information (see Chap. 2.4.5).

[1] expected completion 2012

In addition, the collaboration is testing various enhancements to the original design [107, 130], like radio detection techniques [190], an in-fill area with finer granularity of tanks in a sub-array to achieve low energy extension [66] as well as muon counters in this array. These techniques can provide additional information about showers. Especially, the radio upgrade could provide a cheap way to measure EAS complementary to the SD with 100 % duty cycle.

## 3.1 The Southern Experiment

The Southern Experiment is build in the Argentinian Pampa Amarilla about 300 km south of Mendoza (35° latitude) in the uplands at an altitude between 1300 m and 1500 m above sea level. This corresponds to a vertical atmospheric depth of about 870 $\mathrm{gcm}^{-2}$, needed to study air showers at highest energies with the fluorescence telescopes, as the shower maximum should be above ground. The landscape provides an area of more than 3000 $\mathrm{km}^2$ for instrumentation, that is mostly flat (average slope of 0.5 %). The surface area was formed by an ancient river. Furthermore, vegetation and topology allow feasible deployment of tanks as well as construction and maintenance of infrastructure. The fluorescence eyes are situated at the edges of the area, optimised to observe the atmosphere above the ground array. The cloud coverage less than 15 %, which provides good measuring conditions. The central campus site, located in Malargue, is housing the central data acquisition system as well as an assembly hall for the water Cherenkov tank units. The only larger city, San Rafael, is located about 100 km north. The population in the region of the experimental site is small and thus, light levels due to civilisation are small.
In the following we focus on two main sub-detector systems, the fluorescence telescopes and the surface detectors array. While all fluorescence telescopes are fully operational since Feb. 2007, the completion of the surface array is expected for begin of 2008. Nevertheless, data is continuously taken and first physical results on anisotropy [179, 182], CR energy spectrum [170, 201, 201] and photon fraction [158, 159, 88] have been published.

## 3.2 The Surface Array in Argentina

The surface array consists of 1600 water-Cherenkov tanks that cover a total area of 3000 $\mathrm{km}^2$ (see Fig. 3.2(a)). The main motivations to chose water Cherenkov tanks as detection technique was led by the following considerations:

**Costs:** The production and assembly of the water tanks is relatively cheap, the tanks can be brought out into the fields without de-ionised water even to parts of the array difficult to reach. This fact allows to install the tanks rather handy and fill them afterwards.

**Muon/electron component separation feasibility:** Using the time structure and rise time it is possible to distinguish - at least in a statistical way - the muon and electron component of the air showers footprint. As explained above, this is essential to draw conclusions on the

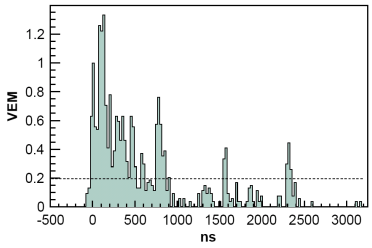
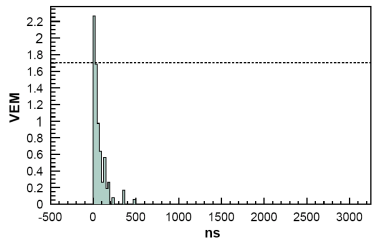
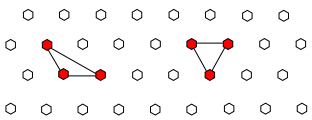
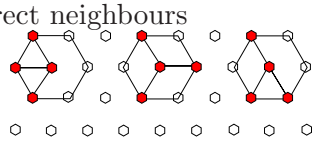
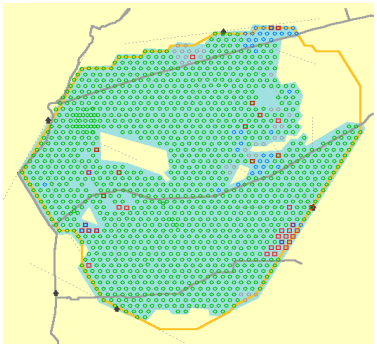
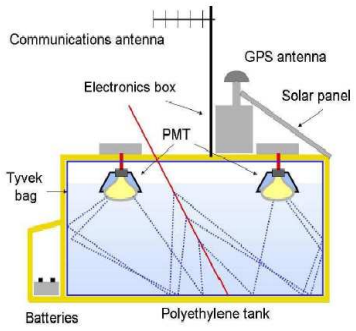
<table>
<tr><th colspan="2">T1 - hardware trigger</th></tr>
<tr><td>T1 - TOT trigger<br>  <br><br>• 2-fold PMT coincidence<br>• 13 bins in 120 bin window above 0.2 VEM</td><td>T1 - single bin trigger<br>  <br><br>• 3-fold PMT coincidence<br>• single bin above 1.75 VEM</td></tr>
<tr><td>1.6 Hz</td><td>100 Hz</td></tr>
<tr><th colspan="2">T2 - Trigger<br>performed in station controller</th></tr>
<tr><td>T2 - TOT trigger<br>• T1 TOT are directly promoted</td><td>T2 - single bin trigger<br>• 3-fold PMT coincidence<br>• single bin above 3.2 VEM</td></tr>
<tr><td>1.6 Hz</td><td>20 Hz</td></tr>
<tr><td colspan="2">data transfer to central station</td></tr>
<tr><th colspan="2">T3 - performed at campus site<br>tank coincidence trigger</th></tr>
<tr><td>mode 1<br>• 3-fold coincidence of TOT-T2 trigger tanks<br>• minimal compactness (at least one tank is next and one is second next neighbour)</td><td>mode 2<br>• 4-fold coincidence of any T2 triggered tanks<br>• moderate compactness (one tank can be up to 6 km away)</td></tr>
<tr><td>90 % purity</td><td>2 % purity</td></tr>
<tr><td colspan="2">0.025 Hz</td></tr>
<tr><th colspan="2">T4 - physics trigger<br>officially applied for non-incline studies ($\theta < 60°$)</th></tr>
<tr><td>Mode 1: T4 - 3TOT<br>• 3-compact TOT T3<br>  </td><td>Mode 2: T4 - 4C1<br>• at least one station has 3 triggered tanks as direct neighbours<br>  </td></tr>
<tr><th colspan="2">T5 - quality trigger<br>choose events for physics and aperture analysis</th></tr>
<tr><td colspan="2">• tank with highest signal has 5 direct neighbours triggered<br>• core lies in equilateral triangle of working stations</td></tr>
</table>

**Table 3.1:** The SD trigger schema.



(a) Status of the ground array at June 2007.

(b) Tanks aligned in field.




(c) Schematic drawing of the Cherenkov tanks, taken from [168].

**Figure 3.2:** The surface array of the Pierre Auger Observatory in Argentina.

primaries composition. An alternative usage of scintillator technique in combination with muon counters would be far too complicated to install with such fine grid-size.

The tanks are arranged in a hexagonal grid and have a spacing of 1.5 km. This spacing is optimised to observe showers above $10^{18}$ eV. The trigger efficiency is supposed to be 100 % above $10^{19}$ eV. The optimisation of the spacing was determined by simulations requiring a trigger of at least 5 tanks detecting a signal. Each tank unit can be described as a cylindrical volume of 3.4 m in diameter and 1.2 m height filled with pure water acting as interaction medium. The sensitive volume is viewed by three photomultipliers (PMT). The water converts also the gammas in the electromagnetic cascade. The height corresponds to three radiation lengths in water to guarantee the absorption of at least 90 % of all incident photons and electrons. The inner cylinder is mantled by a Tyvek liner to reflect the Cherenkov light and shield the effective volume to the outside light. The PMTs look directly downward to obtain optimum uniformity. This arrangement guarantees a good ratio between Cherenkov light and photo-electron yield in the PMTs. The tank itself consists of rotational moulded polyethylene. The material and final design was motivated to resist large temperature differences and occasional damages by cattle.
Furthermore, each unit consists of an autonomous power supply (solar panel with a battery unit), a GPS receiver for independent absolute timing and a GSM-like transceiver unit for wireless communication and data transfer (see Fig. 3.2(b) and Fig. 3.2(c)).

### 3.2.1 Calibration

The signal detected by each tank is usually expressed in units of vertical muon equivalents (VEM). This concept allows to have an absolute measure which is independent of the specific features of each tank. One VEM corresponds to the charge being collected by all three PMTs produced by

a muon passing vertically through the 1.2 m of water[2]. The tank calibration can be performed by using atmospheric muons. The flux is known as well uniform background. The signal is directly proportional to the path length of the particle within the active volume of the tank. The collaboration used a test tank to measure the ratio between down-going vertical muons (VEM) and the peak of the signal distribution for an omni-directional atmospheric muon distribution. This calibration is performed for each tank installed in the field. The gain of each PMT is adjusted to give an expected trigger rate over a given threshold of VEM. The calibration precision lies within 5 %.

### 3.2.2 Data Acquisition and Trigger

The readout electronics have to accomplish the requirement of low power consumption on the one hand, and a large dynamic range on the other hand. The capabilities of the implemented system are quite challenging: The readout is performed using flash ADCs. The sampling rate is 40MHz with a dynamic range of 17 bits. The electronics can be divided in the following subunits:

- Front end electronics
- Communication interface and RF Front end
- GPS receiver and timing unit
- power unit
- station controller

The trigger scheme is given in Tab. 3.1. The trigger hierarchy is made of PMT-wise triggers (T1), tank-wise triggers (T2) and tank-coincidence triggers (T3) as well as physics triggers (T4 and T5). The front end electronics is supposed to form T1 trigger on board. The station-controller applies the T2 trigger based on all T1 triggers. If a T2 trigger condition is satisfied the signal is transmitted to the Central-Station, where the T3 trigger is performed. If this trigger is satisfied, data of all involved tanks are transfered to the Central-Campus site. T4 is a software trigger, that can be considered to be more like a cut, triggering physical EAS events. The T5 trigger is a quality cut to reject events with unreliable reconstruction. The different levels are defined and motivated in the following way:

**SD-T1 trigger:** The first level trigger can fire in two different modes:

The **T1-TOT trigger** is a simple time over threshold trigger for each PMT, triggering with about 1.6 Hz. The condition required is, that at least 13 bins in a 120 bin interval are above a threshold of $0.2\,I_{VEM}$[5, 1] in coincidence with at least two of the tank's PMTs. Aim of this mode is to select a spread signal far off the shower axis from very high energy EAS or for low energy showers to allow rejection from muon background.

[2] using the concept of atmospheric muons, this corresponds to 0.1 particle $\mathrm{m}^{-2}$

The **T1-single-bin trigger** requires a coincidence of all three PMTs having a signal above 1.75 $I_{VEM}$ in just one single bin. The rate of this condition is about 100 Hz[3].

**SD-T2 trigger:** The T2 trigger is a software trigger analysing the full FADC trace of each PMT. All T1-TOT fire directly the T2 trigger, while the T1-single-bin triggers are required to have a signal peak above 3.2 $I_{VEM}$ for one bin in all 3 PMTs. The rate is about 20 Hz. The concept of not directly requiring a threshold of 3.2 $I_{VEM}$ and combine T1 and T2, is to trigger T1 with high rate to monitor the tank and calibration stability[4].

**SD-T3 trigger:** Trigger formed at the Central-Station taking into account coincidences with other tanks in two different modes:
The **T3-3TOT** mode requires at least 3 tanks satisfying the T2-TOT condition, with at least two of the tanks being in direct neighbourhood and one being at least second neighbouring[5]. This mode is firing for 90 % of all T3 events. The T3-TOT trigger is very sensitive on vertical showers.
The **T3-4C1** mode is a trigger on 4-fold coincidence of tanks having any T2 trigger. The required compactness is moderate, that means, that one triggered tank of four can be up to 6 km far away from the others in a certain time window. Although only 2 % of these triggers are real events the T3-4C1 trigger is of importance for inclined showers, that produce wide spread patterns.

**SD-T4 trigger:** The aim of this trigger is to select physical EAS events. The current implementation is applied to all data used for analyses and studies of events below 60° zenith angle. The requirements are based on the compactness of the fired tanks and the spread of the FADC of TOT triggered tanks. The final implementation is still matter of study, currently, T4 requires either a compact 3TOT T3 trigger[6] or a compact configuration of any local trigger. This so called 4C1 requires that one tank unit has at least three triggered tanks in direct neighbourhood. The 3TOT-T4 trigger looses 5 % of all events below 60°. The 4C1-T4 rejects about 98 % of all previous triggers.

**SD-T5 trigger:** The T5 trigger is a quality cut on data, requiring that at least five of the six neighbouring tanks of the tank with the hottest signal have triggered as well and that the reconstructed core is located in an equilateral triangle of working stations. The bias of one missing tank is estimated to be below 8 % on the final event sampling rate, but this cut leaves a reliable sample of events to calculate aperture or energy spectra.

**Hybrid trigger** In addition the SD-array can be triggered by the FD-telescopes with the FD-T3 trigger. This is an external trigger from the Fluorescence Data Acquisition System (FDAS).

[3]The high trigger rate cannot be reduced further, as it allows to be sensitive on the fast signals from the muon component of old showers
[4]e.g. the T1 rate is checked as function of temperature
[5]this condition is often refered to as minimum compactness
[6]This ensures a purity of 99 %

The aim is to look for single tanks that could have been triggered below the coincidence requirement of the general SD-trigger scheme. This tank information can be used to reconstruct the event in a hybrid mode. At the energy threshold of the FD-telescopes around a few $10^{17}$ eV the probability of the shower front to fire a T2 is still about 100 %, while these showers would have never been triggered by the SD. Technically, the FD-T3 estimates the landing line and landing time of the EAS using a fast reconstruction algorithm and the Central Data Acquisition System (CDAS) looks for compatible T2-tanks in this region with a 20 $\mu$s time window around the estimated landing time. The T3-triggers are required to fire with at least 5 s time difference to not degrade the SD trigger performance.

## 3.3 The Fluorescence Telescope Sites in Argentina

The fluorescence telescope detection is provided by 24 telescopes. The complete fluorescence detection system has been finished in February 2007 by Loma Amarilla telescope site detecting first light. Currently, the design implemented looks the following way:

**Infrastructure:** The atmosphere above the ground array in the Pampa Amarilla is observed by four telescope stations. Each station hosts 6 telescopes (see Fig. 3.4). The telescopes are mounted in the same housing, separated by large canvases dividing the telescope buildings in so called bays. Each telescope covers a field of view of 1° to 31° in elevation and 30° in azimuth, so each telescope building covers an azimuthal range of 180°. Therefore, the fluorescence detectors cover the complete ground array. To guarantee a clear view on the atmosphere the exact locations of the buildings have been chosen at exposed places, like small hills. Fig. 3.3 shows pictures of all four stations. The first completed station was Los Leones, going into operation in May 2001, followed by Coihueco in April 2002, Los Morados in November 2004 and finally Loma Amarilla in February 2007. With the telescopes placed at the edges of the ground array pointing inside the surface detection field, it is also possible to trigger EAS by more than one telescope. These so called multi-fold events form a high quality subset of data, that can be used to cross check FD-performance with real data. Data transfer goes via RF-technique.

**Light collection:** The telescope is arranged as Schmidt optic (see Fig. 3.5). A shutter of each telescope-bay opens a window of 3 m width and 3.5 m height. Within this window the aperture box is mounted. This box carries the aperture system with an outer radius of 1.1 m plus additional area with a corrector ring to compensate spherical aberration. Additionally, the optical filters are installed in the aperture box. These filters are only transmissive in the UV regime to reduce the noise of the night-sky background. Beyond the aperture box a failsafe curtain is mounted. This system allows to darken the bay within less than a second in an emergency case[7]. The segmented mirror has a diameter of 3.4 m. A corrector ring at the centre of curvature eliminates coma aberration.

[7] e.g., the shutters have not shut at dawn

(a) Los Leones, in operation since May 2001

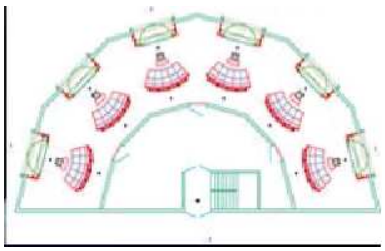
(b) Coihueco, in operation since April 2002

(c) Los Morados, in operation since November 2004

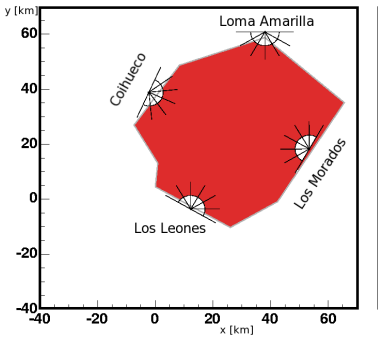
(d) Loma Amarilla, in operation since February 2007

**Figure 3.3:** The fluorescence telescopes of the Pierre Auger Observatory in Argentina.

(a) Central Site at main campus

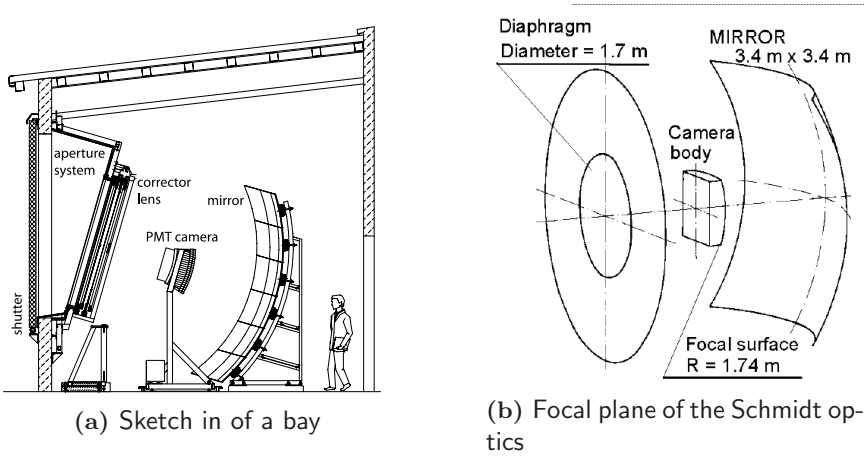


(b) Sketch of a FD-site building




(c) The location of the FD sites

**Figure 3.4:** Main campus and FD infrastructure.

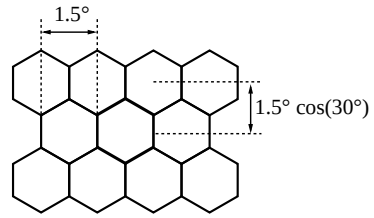



(a) Sketch in of a bay

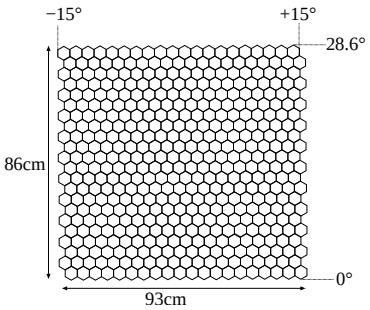
(b) Focal plane of the Schmidt optics

**Figure 3.5:** The Schmidt optics arrangement of the FD telescopes.

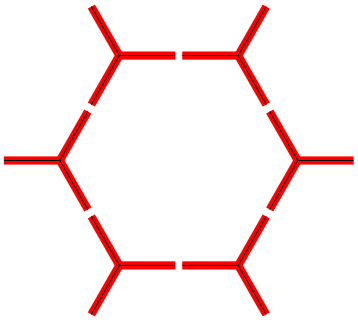
**Light detection:** The light is detected by a fluorescence camera (see Fig. 3.6), that is mounted in the focal surface of the mirror. Each camera is composed of a 440 PMT matrix mounted in a 20×22 pixel mask. The camera body covers only 0.1 $m^2$ in the field of view of the mirror. The matrix was optimised in order to find a good compromise between resolution and minimum confusion, that is reached for a PMT distance of 45.6 mm. To achieve a good covering a hexagonal structure was chosen. A significant amount of insensitive area is still present. On the one hand there is need of space between the PMTs for safe mechanical mounting in the mask, on the other hand, the effective cathode is smaller than the area of the glass envelope. Each pixel corresponds to a space angle of 1.5°. Additionally, light collectors are mounted




(a) Hexagonal structure of the camera.

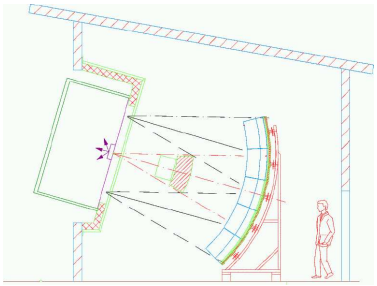



(b) Dimension of the PMT camera.



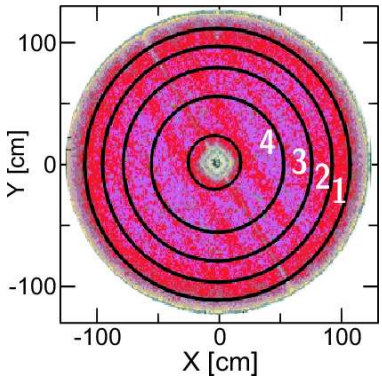
(c) The Mercedes light guide.

**Figure 3.6:** The PMT camera of the FD-telescopes.



(a) Schematic viewing of drum mounted in front of a bay.

(b) The mounted drum.




(c) Intensity map of the drum showing the uniformity of light emission measured with a CCD.

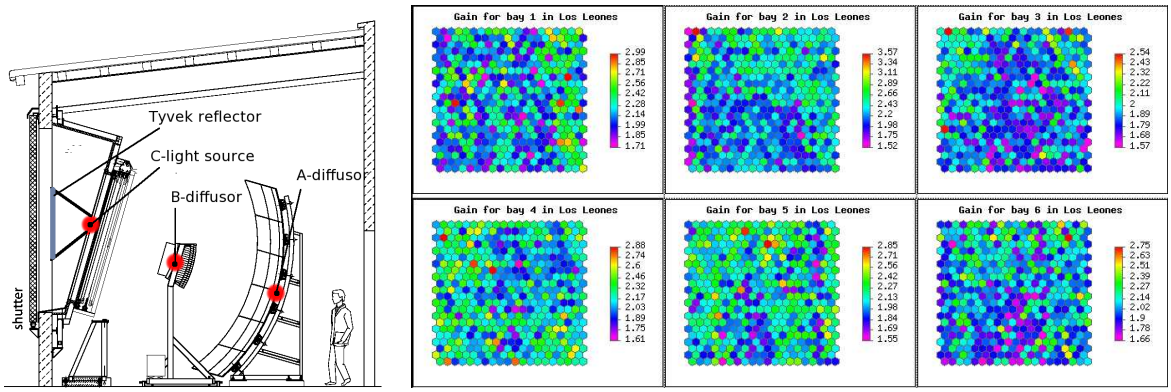
**Figure 3.7:** Drum calibration, figures taken from [24].

between the PMTs to maximize the light collection efficiency and obtain a sharp transition between the pixels. These light collectors are composed out of basic elements called Mercedes stars as they are shaped like a star with three arms oriented in 120° rotational symmetry. Six Mercedes co-rotate one PMT and guarantee a collection efficiency of 94 %. The PMT readout is performed by a FADC to digitise the signal. The sampling rate of 10 MHz and wide dynamic range allows good geometry and profile reconstruction of the detected showers. In fact a sample duration of 100 ns corresponds to a profile sampling of $4\,g/cm^2$. Thus, the event reconstruction is limited by the signal to noise ratio.

### 3.3.1 Calibration

A precise knowledge of the calibration constants for each PMT is needed to convert the final FADC pulse into the number of photons reaching the diaphragm and calculate the light emission at the shower. Two general calibration schemas are applied to guarantee high accuracy: *Relative calibration* is done in three steps, including direct response of the camera, response of the camera mirror system and the full light collection system. *Absolute calibration* is performed by using a standardised homogeneous light source that can be mounted in front of the shutter (*drum calibration*) or using calibrated laser shot in the field (*Rayleigh calibration*). Additionally, an analytical method can be applied to estimate the absolute calibration by comparing the pixel-by-pixel response including ray-tracing (*piece-by-piece simulation*). All three relative calibrations are performed on a nightly base, while the drum calibration is applied in non-regular intervals. In detail, the calibration schema looks like this:

**Drum calibration:** To obtain absolute end-to-end calibration for the fluorescence telescope a calibrated light source (*'drum'*) can be mounted in front of a bay. The advantage of this cali-




(a) Schematic drawing of relative calibrations

(b) Typical gain distributions from calibration A in each bay of telescope Los Leones

**Figure 3.8:** Relative calibration of the FD telescopes.

bration is that the complete light collection and detection system can be taken as black box, and the PMTs' response is measured under normal data taking conditions. The disadvantage is that this calibration can only be performed on a non-regular base, as the mounting of the drum is very work- and time-intensive. The idea of the drum is to provide a known source of photons simulating a far away point source simultaneously in all PMTs (see Fig. 3.7). This requires a high uniformity of the light source. The device is working with a pair of 375 nm LEDs embedded in a Teflon diffuser mounted near the front of a drum shaped cylindrical box[8].

A ray-tracing program using the uniformity and angular intensity information from the CCD images shows less than 1 % variation in total flux seen by the pixels, and corrections are applied for these variations [24]. The drum itself is calibrated by using NIST-calibrated photodiodes, that are capable of measuring the absolute light flux within 7 % accuracy.

**Calibration A:** This calibration is performed at the beginning and end of every data taking night. A light diffuser mounted in the mirror pulses directly into each camera. As this calibration is performed on a nightly base, it can be used to monitor the PMT stability.

**Calibration B:** A diffusor mounted in the camera is firing on a nightly base into the mirror. The response of the reflected light is measured to monitor the mirror-camera system.

**Calibration C:** Here, light sources are mounted in the ports of the aperture box. The light pulses are fired on a Tyvek foil mounted on the inner doors of the shutters. These foils reflect and

[8] A new drum device has recently been used with multi-wavelength sources, that allow to investigate also the wavelength dependency of the absolute calibration[46].

diffuse the light and give the opportunity to monitor the relative response of the whole light collection system including optics and filters.

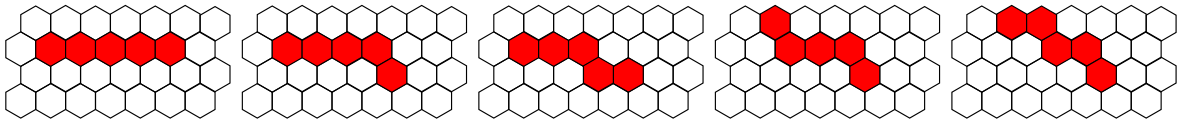
Monitoring of *Calibration A* shows, that the PMT response is stable on a longterm and monthly base within a few percent.

### 3.3.2 Data Acquisition and Trigger

The FD trigger scheme consists of two hardware triggers, the first level trigger (FLT) and second level trigger (SLT) followed by two software triggers, the third level trigger (TLT). In addition the so called T3-trigger acts as external trigger for the ground array to enhance the hybrid event rate.

**The First Level Trigger (FLT):** This trigger fires if the signal of a PMT passes over a certain threshold. The threshold is adjusted dynamically. A feedback system monitors the trigger rate and changes the signal threshold in such a manner that the trigger rate is constantly at 100 Hz.

**The Second Level Trigger (SLT):** The aim of the SLT is to discriminate if the FLT was induced by a shower track or from noise. Therefore, the 22×20 matrix of PMTs in each camera is scanned for certain patterns. Showers manifest as trace in the camera and thus collinear patterns can be found. The probability that such collinear patterns are induced by background is small. Possible background sources can be random noise, automobile headlights or airplanes. The optimised trigger sees for a 4-fold coincidence in a 5-fold collinear pattern. The elementary patterns, which are geometrically possible on the grid of the camera are:



Rotating according to the symmetry axes by ±60° and ±120° and flipping yields 39 different patterns. Additionally, including the odd-even asymmetry of the rows in the hexagonal mask there are in total 78 patterns. The requirement to get 4-fold coincidence in a pattern leaves a final number of 108 possible configurations. Technically, the grid of the camera is analysed column-wise 5×20 sub-matrices using lookup-tables. Each scan takes about 1 $\mu$s. SLT events still contain a small fraction of unphysical background, like small tracks induced by cosmic muons in the camera, coherent noise from lightening or stars in the field of view.

**The Third Level Trigger (TLT):** The TLT is designed as software trigger to discriminate non-physical events. The TLT uses a fast pre-reconstruction algorithm including the T2-triggered pixel list and their closest neighbours. An eye-event builder checks if more than one bay has triggered and allows to merge the different bay triggers into one event, storing FLT- and SLT-trigger information as well as all relevant FADC traces. The trigger rate is about 100 events per hour.

**T3:** This trigger selects shower candidates from accidental background and transmits these triggers to the CDAS. The pulse-width is analysed to check for pulses with finite transit time[9]. Additionally, the distribution of pixels is analysed by calculating the 'Hough' parameter. This allows to distinguish between track-like events and 'blob'-like noise from lightenings that illuminate almost the whole camera. Cosmic muons manifest by high and short pulses and negligible transit time between the pixels. Shower candidates are then selected with about 5-10 events per hour. In that case the fast pulse finder and noise subtraction is applied. A fast reconstruction estimates the shower-detector plane and its projection on ground, the so called *'landing line'*. A preliminary time fit extrapolates the time when the shower hits the ground, the so called *'landing time'*. The T3 trigger sends the landing line and time to the CDAS to give an external trigger to the ground array looking for matching SD-T2 stations.

### 3.3.3 Performance Tests and Calibration of the Phototubes

The fluorescence cameras are instrumented with PHOTONIS XP-3062 PMTs. These are eight-stage PMTs with flying leads and a 40 mm edge-to-edge hexagonal shaped photo-cathode of bialcali type. Each PMT is soldered to small printed circuit boards, the Head Electronics unit (HE) [108]. The HE implements a voltage divider for the PMT and delivers the PMT signal, after low noise amplification, with twisted pair differential output. The main characteristics of a PM unit are:

**Uniform response:** non-uniformity less than 15 %.

**Nominal gain** between $5 \cdot 10^4 - 10^5$.

**Uniform spectral response:** quantum efficiency is 0.25 in the wavelength regime of interest with spectral deviation less than 10 %.

**Linear response** is better than 3 % over a dynamic range of $10^4$ of 1 $\mu$s signals.

**Longevity:** integrated maximum anode charge corresponds to a half life of about 50 years.

For the verification of these quality requirements all 11000 PMTs have been tested and classified by their nominal high voltage $H_0$ corresponding to a gain of $10^5$. This classification is needed, because 44 PMTs share one HV supply. In order to equalise the response and to minimise calibration corrections, PMTs of similar gain characteristics are connected to the same HV supply. The experimental requirements are described in [25, 177]. Because of the large number of PMTs to be tested the measurement procedure has to be fast and focused on the main requirements.

- Identification of dead and unstable PMTs.
- Measurement of the gain as function of the HV.
- Classification of the PMTs into gain classes

[9] E.g., lightening or planes produce usually pulses that are longer than the sampling time.

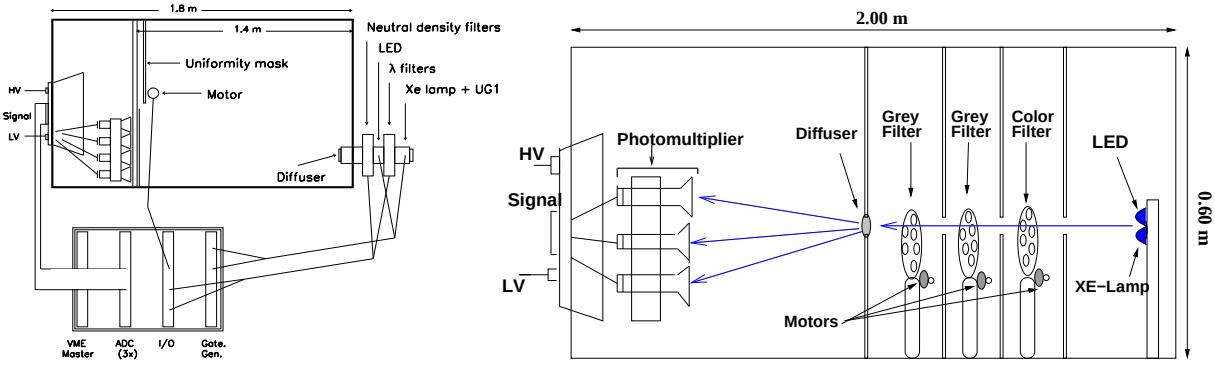



**Figure 3.9:** Sketches of the test setups in Rome (left) and Wuppertal (right).

- Test of the linearity of the PMT response for different light intensities.
- Test of the wavelength dependence of the PMT response, especially for UV wavelengths.

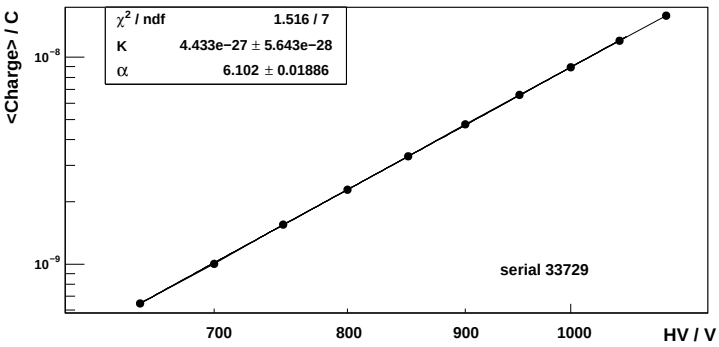
Therefore, dedicated test stands in Rome and Wuppertal[25, 26] have been set up (see Fig. 3.9). About 6100 out of a total of about 10800 PMTs plus additional spares have been tested in Wuppertal with a dedicated setup, the remaining ones have been tested in Rome with a similar procedure, described in [25, 148, 129].

In the Wuppertal setup all components are mounted on an optical bench inside a 2 m long wooden box. Optical filters, located on motor driven wheels, can be inserted into the light path by the remote control software. A total of 37 PMTs can be mounted in parallel. The test procedure takes about 2 hours. A speciality of the setup in Wuppertal is the use of a more complex DAQ system in which the full PMT waveforms are digitised by FADCs; the system was adopted from [195]. The digitisation with FADCs provides the opportunity to implement a highly automated online monitoring and analysis of the PMT's charge response, checking for saturation, and implementing dynamic integration gates and pedestal subtraction.

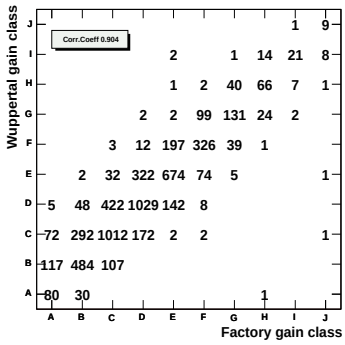
The measurements were done using a blue LED with the exception of the spectral measurement, where a Xe-lamp was used. The width of the pulses was set to about 1-2 $\mu$s, which corresponds to the expected duration of fluorescence signals in the experiment. With different neutral density attenuation filters the light intensity can be changed by more than 3 orders of magnitude. Filters for 337 nm and 390 nm are used to select specific wavelength-intervals.

For each measurement batch, the PMTs are mounted together with 4 absolutely calibrated *reference* PMTs. The different measurements are then performed in a sequential order. According to the results of the data analysis the PMTs are then classified and packed in portions optimised for the construction procedures in Argentina. The results of the data analysis are stored in a MySQL based database, which provides access to each individual result for the Pierre Auger Collaboration.

The same reference PMTs are used throughout the entire measurement campaign to monitor changes of the setup; the test PMTs are calibrated with respect to these PMTs. Deviations in the homogeneity of the light at positions of the PMTs are about a few percent; this is frequently




(a) Typical result of the calibration of the gain as function of the applied high voltage. [26]



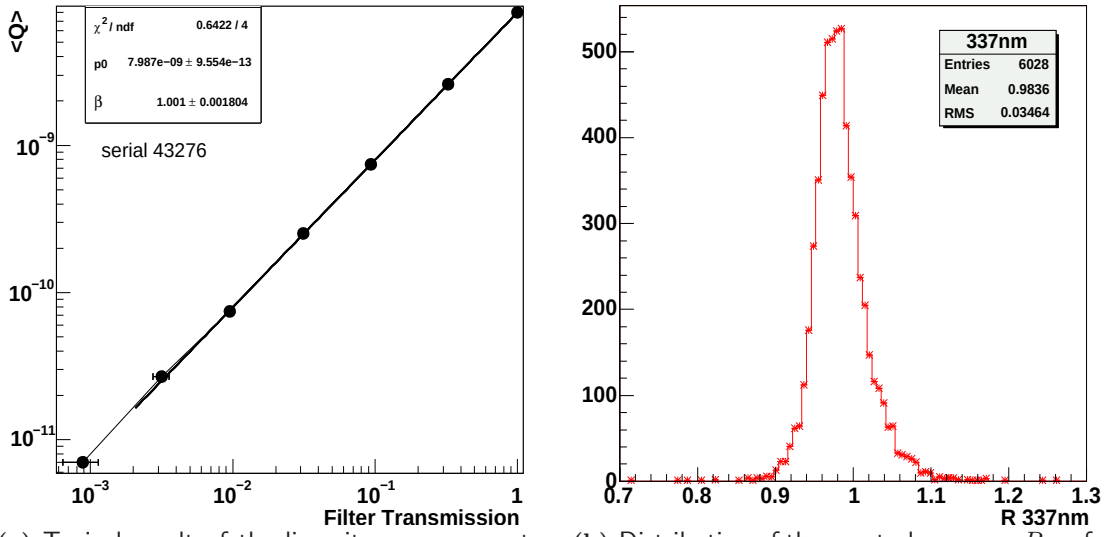

(b) Distribution of the classified gain classes. The numbers correspond to the number of PMTs in each bin. [26]

**Figure 3.10:** Results of the PMT classification tests.

monitored and corrected for in the data analysis. The results for the complete set of tested PMTs are summarised below:

**Gain calibration:** A typical measurement is shown in Fig. 3.10(a). The collected charge as function of the high voltage is well described by a power-law $Q = K \cdot (\mathrm{HV})^{\alpha}$. The fit values for $\alpha$ are found to be Gaussian distributed with a mean of $\simeq 6.1$ and variance of $\simeq 0.1$. This is consistent with the typical expectation of $\alpha \approx 6$ for an 8-stage PMT. The obtained fit parameters $K$ and $\alpha$ are used to calculate the nominal high voltage, $\mathrm{HV}_0$. The PMTs are then classified according to $\mathrm{HV}_0$ into 10 gain classes $A \rightarrow J$. The distribution of gain-classes can be seen in Fig. 3.10(b). Each entry in the scatter plot corresponds to a PMT with its gain class as classified by *PHOTONIS* and by the Wuppertal test setup, respectively. Apart from the generally good correlation, we have observed an apparent migration from the initial factory classification to our final value. It should be noted that our measurement includes the mounted head-electronics and conditions closer to the operation *in situ* than for the initial factory calibration. Using our estimate on the total error $\delta(HV_0) \approx 3.5\,\mathrm{V}$ and the natural spread of $HV_0$ within one gain class we expect gain variations within one high voltage channel of the FD camera to be $\simeq 5\%$. This can be corrected in the front-end electronics of the camera.

**Linearity:** The linearity of the PMT's charge response is measured at the nominal $\mathrm{HV}_0$ using a blue LED. The light intensity is varied by neutral density transmission filters and the corresponding response is fitted with a power-law $< Q >= p_0 \cdot T^{\beta}$ as function of the relative transmission $T$. Figure 3.11(a) shows the excellent linear behaviour for a typical single measurement. The values for the power-index $\beta$ are consistent with $\beta = 1$ and the fits have a reasonable $\chi^2$ distribution.




(a) Typical result of the linearity measurement. [26]

(b) Distribution of the spectral response $R_{337}$ for the tested PMTs. [26]

**Figure 3.11:** Performance of the tested PMTs with respect to linearity and spectral response.

**Spectral response:** A Xe-light source is used and the charge response with narrow band transmission filters and without filters ("*white*") is measured. In order to account for fluctuations in the intensity and wavelength spectrum of the Xe-lamp the response with respect to our reference PMTs is investigated. Therefore, only the double ratio

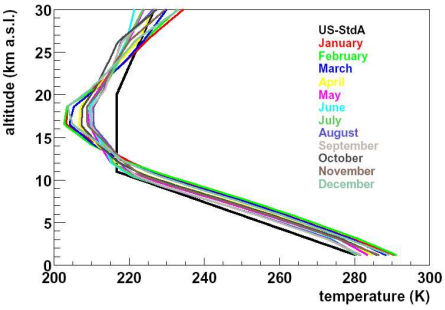
$$< R_\lambda > \equiv \frac{< Q_\lambda > / < Q_\lambda^{ref} >}{< Q_{white} > / < Q_{white}^{ref} >} \ ,$$

is inspected, in which these effects are largely canceled out. As an example, Fig. 3.11(b) shows the result for $\lambda = 337\,\mathrm{nm}$. The deviations from unity have an RMS of only $\sim 3.5\,\%$.

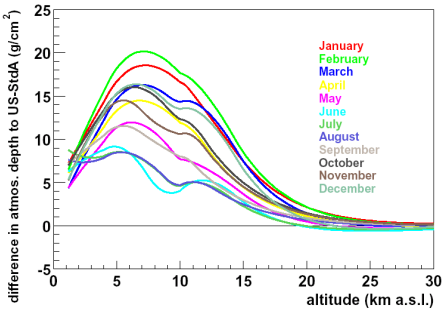
### 3.3.4 Monitoring the Experimental Site

The fluorescence telescopes can be regarded as calorimeter using the atmosphere above the ground array as huge interaction medium. Therefore, it is essential to monitor the properties of the atmosphere (see Chap. 2.4.1). The Pierre Auger collaboration institutes a broad program of nightly atmospheric monitoring. The aim of the program is to characterize the absorption and scattering of ultraviolet light as it propagates from an air shower to the fluorescence detectors. The aim of the monitoring is to correct fluorescence observations for changing atmospheric conditions, and thereby contribute to the absolute calibration of the observatory's energy measurements. Currently, the Pierre Auger Observatory uses the following devices to guarantee good reconstruction quality:

(a) A balloon launch




(b) Atmospheric profile ([104])




(c) Deviation from the US standard atmosphere ([104])

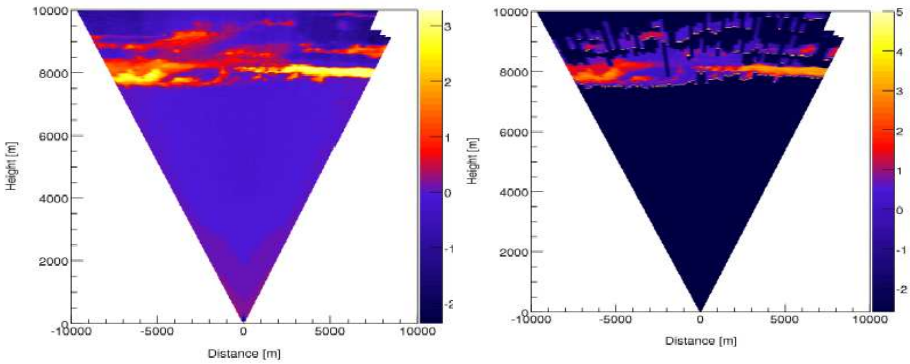
**Figure 3.12:** Balloon based radio soundings.

**Balloon Based Radio Soundings:** To reconstruct an EAS, the specific density profile of the traversed atmosphere is crucial [104, 106], as it affects the fluorescence yield (see Chap. 2.4.1) and propagation characteristics (see Chap. 2.4.1). Furthermore, as explained in Chap. 2.4.1 the fluorescence yield does not only depend on the pressure, but also on the temperature, thus one needs to monitor also the temperature profile of the atmosphere. Therefore, the collaboration performs regularly measurement campaigns launching sondes fixed at helium balloons (see Fig. 3.12). In the meantime, data from these measurements for all seasons are available for the southern site. Depending on the launching site and wind directions these measurements represent various slices of the atmosphere above the experiment. The sondes are standard meteorological instruments to measure pressure, temperature and humidity. A GPS receiver on board allows to determine the geometrical height. Data transfer is provided by an RF transmitter. In connection with local weather stations located at each telescope site a three dimensional model has been developed representing the 'Malargue'-monthly atmospheric profiles above the surface array. The atmospheric model is stored in a MySQL database and can be accessed by the reconstruction software to guarantee time dependent atmospheric parameterisation. The alternative model is the US Standard Atmosphere from the NASA going back to original publications from 1976. This model fixes wind speed and relative humidity to zero and does not distinguish between seasonal variations. Furthermore, this model is not very useful for event reconstruction, because no error to the applied parameters for the south side can be attached. The US standard atmosphere is implemented in the Pierre Auger Observatory atmospheric database, but it has turned out that the Malargue monthly profiles are more suitable and offer a more reliable reconstruction.

**LIght Detection and Ranging (LIDAR):** The atmospheric monitoring campaign includes also a system of four elastic LIDAR stations[10] located next to each FD-site. Actually three out of

[10] the term elastic refers to the backscattered light, as these LIDARs measure the light in the same frequency range

(a) Coihueco LIDAR-station.




(b) LIDAR cloud detection. Left: Azimuth continuous scan taken by Coihueco LIDAR station, right: detected clouds in the same scan (bright points indicate the beginning of the cloud, colour scale in arbitrary units). [137]

**Figure 3.13:** The LIDAR system.

four stations are fully operational and the fourth is expected to be completed by 2007[31]. In each LIDAR station a high repetition UV laser[11] fires short light pulses in the atmosphere. The backscattered light is detected by PMTs mounted in the foci of three parabolic mirrors. Both laser and mirror system are mounted on a steering frame and allow full sky pointing feasibility. The LIDAR operates during FD data taking periods. Each hour the LIDAR scans the sky above each FD site in north-south and east-west direction. These data are used for monitoring the local aerosol scattering and absorption properties of the atmosphere and reconstruct the atmospheric homogeneity between the FD sites. Furthermore, it is possible to detect height and optical depth of clouds above each site and estimate a local cloud coverage (see Fig. 3.13). In a second mode it is possible to use the LIDAR for real-time monitoring of a dedicated part of the atmosphere. For an interesting sub-selection of triggered events, e.g. high energetic hybrid events, the LIDAR standard scan mode is interrupted and 2 to 4 minutes after the events the LIDAR scans the atmosphere in the vicinity of the reconstructed air shower location. This mode is refered to as *'shoot-the-shower'* mode, and aims to increase the reconstruction quality of these dedicated events.

**Horizontal Attenuation- and Aerosol Phase Function-Monitors:** A *Horizontal Attenuation Monitor system* (HAM) is located at a telescope site and consists of a DC light source pointing to a receiver located at another telescope site. Currently, only one HAM is operating between the Los Leones and the Coihueco telescope site, two more systems are under investigation. The intention of these systems is to measure the attenuation length near ground level. Therefore, the light source emits a broad spectrum in the UV regime (300-400 nm).

as the laser shot
[11] 300 Hz

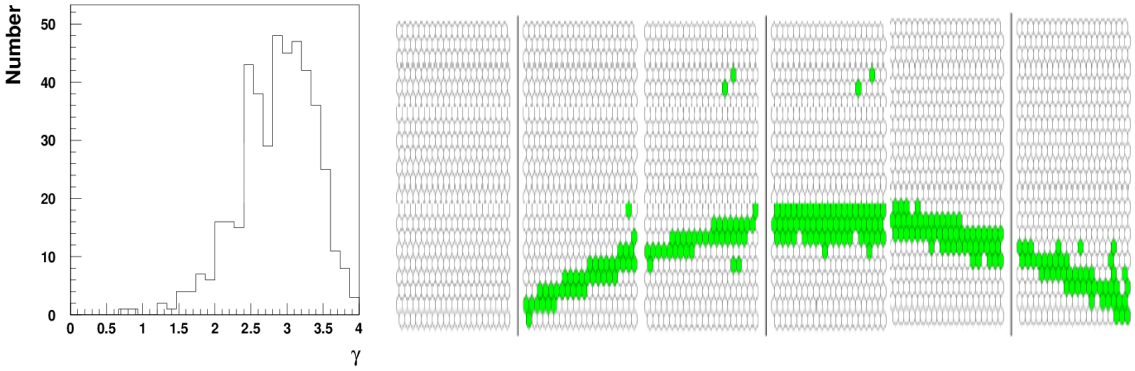



**Figure 3.14:** Left: measured power-law indices $\gamma$ from HAM, a pure molecular atmosphere would predict $\gamma = 4$, right: APF pulse as seen by Coihueco FD, light travels from the left to the right. [193].

The light is detected by UV enhanced CCD arrays[12] in the focus of 15 cm diameter mirrors. A filter wheel in front of the CCD arrays allows to distinguish light detection at different wavelengths[13]. The HAM system operates hourly during fluorescence data taking periods. The principle measurement made by HAMs is the wavelength dependency of the horizontal attenuation that should follow a power-law with index $\gamma$. By comparing the measured power law index to the expected index for a pure molecular atmosphere ($\gamma$=4) it is possible to extrapolate the aerosol contribution at ground (see Fig. 3.14).
The *Aerosol Phase-Function Monitor system* (APF) consists of a Xenon light flash, generating a horizontal collimated beam of light. This beam is fired across the front of a telescope-side. The detected photons in the trace of the cameras stem of a broad range of scattering angles from the beam (30°-150°) and allow to measure the aerosol differential cross section $d\sigma/d\Omega$. This property is important for the reconstruction as it allows to estimate the Cherenkov contamination of the signal detected (see Chap. 2.4.1). [52].

**Central Laser Facility (CLF):** The central laser facility is located in the middle of the surface array, 26-39 km away from the telescope-sites. In the CLF a calibrated 335 nm laser is mounted, that can be pointed to any part of the night-sky with an accuracy of 0.2°. The pulse intensity can be set up to 7 mJ and has a width of about 7 ns. The scattered photons appear in the fluorescence telescopes as up-going track. The CLF fires various scans during data taking periods in 15 minutes-intervals. Power supply is provides by a solar panel fuelled battery system. The CLF system also uses an optical fibre to extract a fraction from the laser beam to pulse a nearby tank (Celeste) to simulate a hybrid event. The detected CLF shots in SD and FD can be used for a variety of systematic investigations [52, 16]:

[12] Starlight Xpress model MX5-16
[13] 365 nm, 404 nm, 436 nm, 542 nm

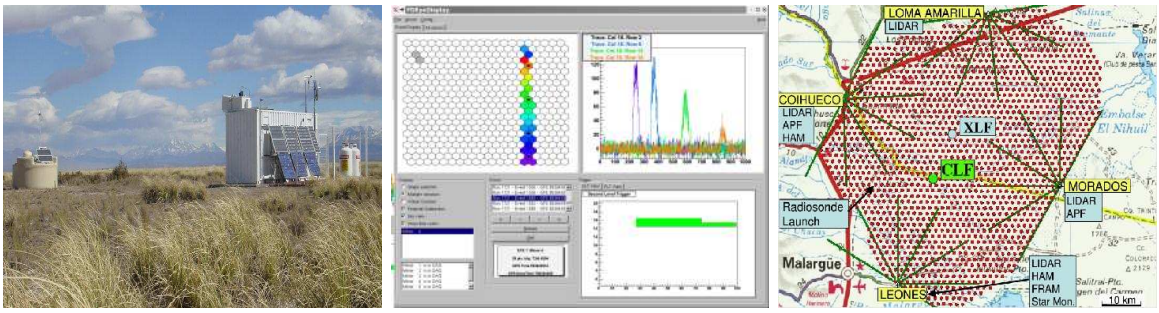



**Figure 3.15:** CLF and tank Celeste. [136, 16, 200]

- Geometry and energy reconstruction accuracy and trigger efficiency of fluorescence data
- Mounting accuracy of the fluorescence cameras.
- FD-SD time offset by comparing the Celeste signal to the fluorescence ones.
- Aerosol measurement.
- Cloud fraction measurement (clouds act like diffusers and reflect parts of the laser beam, that manifests as blob shaped track in the camera, so called *nukes*).

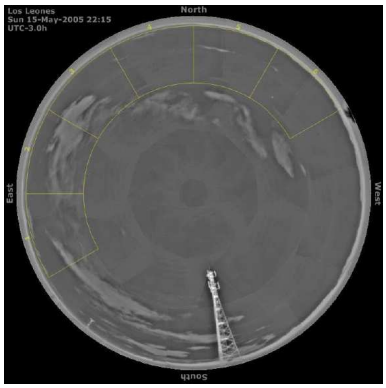
A second Central Laser Facility called the XLF equidistant from Coihueco, Los Morados and Loma Amarilla has recently produced the first vertical beam. The test beam was initiated from CDAS and recorded.

**Cloud Monitor:** Clouds can indirectly be detected by the CLF and LIDAR. The cloud monitor is a dedicated device to measure the cloud fraction directly with an all-sky map. The information is important as clouds can have large optical depths and thus have great influence on atmospheric scattering and transmission properties. Therefore, a finely pixelated infrared camera[14] is mounted on the roof of each telescope site (see Fig. 3.16(a)). As the presence of a cloud manifests as temperature difference between cloud and night sky, clouds become visible on the infrared photos. The camera itself has a field of view of $46° \times 35°$ and is mounted on a steerable frame to observe the full sky around each site. During data-taking these cameras provide a full sky image every 15 minutes (see Fig. 3.16(b)). The development of a fully automated analysis of these images using pattern recognition techniques is in progress. Furthermore, the images can be directly used by the FD shift crew to examine remotely the weather conditions at the telescope site. [52]

**Star Monitor:** On top of the Los Leones telescope building the Auger Star Monitor (ASM) is installed (see Fig. 3.16(c)). The system consists of a CCD, that continuously images the night sky through a wide-field lens (105°) and a Johnson U-band filter (close to the FD). The

[14] Raytheon, 2000B, 7 $\mu$m-14 $\mu$m

(a) Cloud camera mounted on top of Los Leones.




(b) A typical cloud image.


(c) Star monitor (left) mounted on the roof of Los Leones, right beside the cloud camera.

**Figure 3.16:** Devices on top of each telescope site: Star monitor and cloud cameras.

ASM is designed to take simultaneous photometric measurements of the stars. Variations between the expected apparent brightness of the stars and their observed brightness as a function of their instantaneous zenith angles give the total extinction of the atmospheric overburden. Also information about the cloud cover can be obtained from the images. The ASM is working remotely and fully automated all nights during the operation time of the fluorescence detectors.[51, 52]

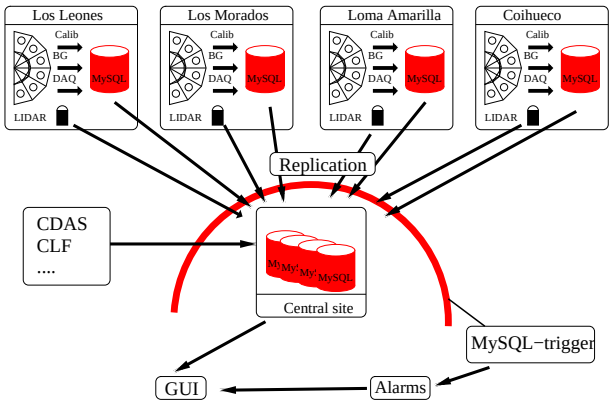
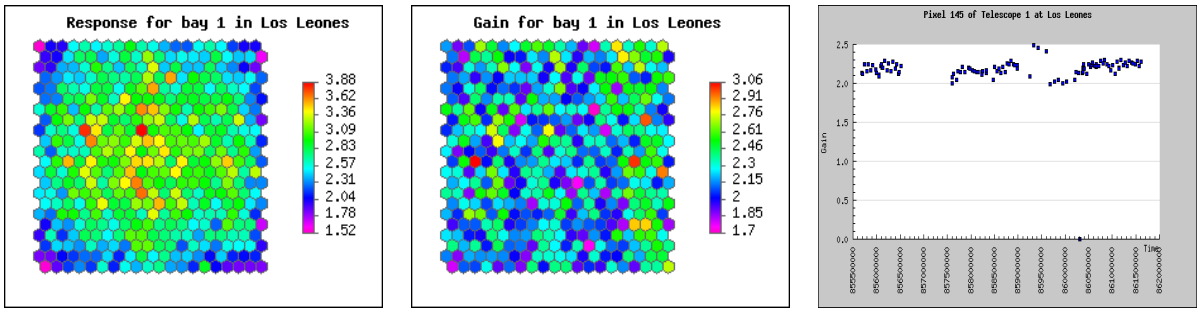



**Figure 3.17:** Monitoring scheme.




**(a)** 2-d representation of the response of all camera pixel in calibration A in bay 1 Los Leones

**(b)** 2-d representation of the gain of all camera pixel in calibration A in bay 1 Los Leones

**(c)** Detailed information can be accessed for single PMTs, here the history of PMT 145 in bay 1 Los Leones

**Figure 3.18:** Monitoring Calibration.

### 3.3.5 Data Taking and Monitoring the Data Quality

The fluorescence telescopes can be used in clear moon-less nights (see Chap. 2.4.1). At the same time, a shift crew on site is supervising the data-taking and steering of the 24 fluorescence telescopes to allow smooth operation. It has to be monitored constantly to assure the quality of the recorded data as well as guaranteeing the safe operation of all detector components, electronics and data-acquisition. The sensitive cameras can only be operated in dark nights without strong wind and without rain. This makes the operation a full task for the shifters, who have to judge the operation-mode on the bases of the information given. Many important cases that can be identified in a unique way have been automated by a slow control system. E.g. there are light sensors in each bay, to force the failsafe curtain to close immediately and close the shutter if outside light is reaching a critical level. This can be due to lightenings in the field of view of that bay or the spot light of a far away jeep. The task of the shift crew is to identify the reason and unlock the bay as soon as it is possible. In the beginning of continuous data taking in 2002 there has been only Los Leones in the data acquisition and it was possible for the crew to check the status of all components and processes continuously by hand. Meanwhile, more telescopes have come into operation, and thus the situation is more complex for the shift crew. Therefore, a monitoring tool has been developed to support the shifters in judging and supervising the status of the detector components, the electronics and the data-acquisition. Since the beginning of the development the Wuppertal group is involved in the technical implementation, design of the schema and database, visualisation and installation on site.

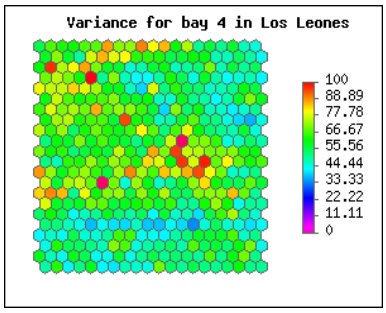
### 3.3.6 Concept of the Monitoring Structure

In the meantime, the SD and FD monitoring tools have been merged to access information of the complete experiment. During operation there are processes running that take monitoring-data in

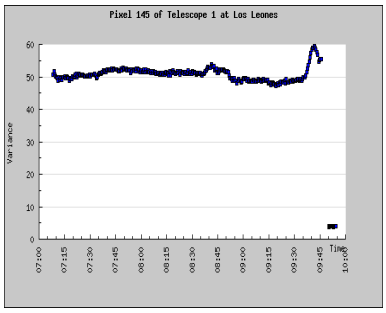
BG Loop summary at Los Leones

| | Tel 1 | Tel 2 | Tel 3 | Tel 4 | Tel 5 | Tel 6 |
|---|---|---|---|---|---|---|
| Variance [ADC$^2$] | 3.5 | 3.4 | 3.7 | 4.1 | 4.2 | 3.7 |
| Threshold [ADC] | 1261 | 1217 | 1235 | 1321 | 1428 | 1274 |
| Hitrate [Hz] | 109 | 118 | 114 | 109 | 124 | 123 |
| Pedestal [ADC] | 150.0 | 150.0 | 150.0 | 150.0 | 150.0 | 150.0 |

(a) Background page starts with a summery survey




(b) Variances measured in background loop as camera plot




(c) Detailed history of pixel is also accessible

**Figure 3.19:** Monitoring background loop.

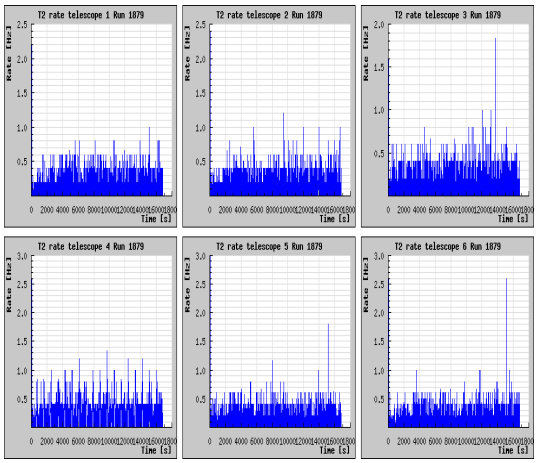
a continuous mode. While for some components, like SD, this information is directly transmitted to CDAS, the monitoring data from each FD-site are locally stored to a MySQL database, that replicates with a common database at the central station. This replication has to be done via the same wireless long distance link the regular data are transfered. This requires the monitoring data to be limited. Basically, for each eye the following information is monitored

**Calibration:** The data from all calibrations are added to a dedicated database. On the one hand, this allows the crew to access calibration data and on the other hand, the crew can directly verify that calibration A and B, that should be done before each data-taking, have completed smoothly.

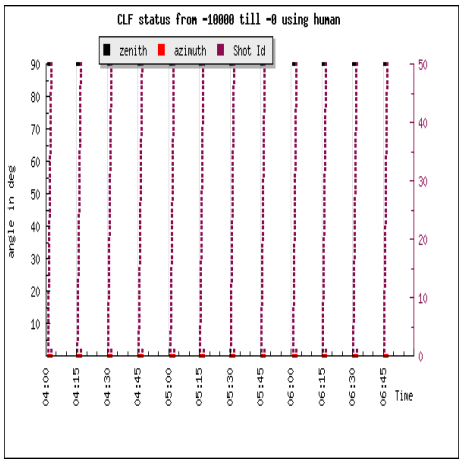
**Night-sky background:** Every 30 s each camera is read out for background noise estimation. This information is valuable as non biased observation of background. From a monitoring point of view, it allows to check if the FLT threshold works in a correct manner. The rate should be of the order of 10 Hz. Additionally, the variances allow to distinguish independently from the slow control system whether the shutter is opened or closed[15], and even more what the weather conditions are in the field of view[16]. Furthermore, it can be used as a cross-check for verifying the uptime as besides the shutter and failsafe curtain status it is possible to verify that HV are ramped up and readout electronics work smoothly for all pixels. The original intention for taking background data is to follow up stars as noise in the night-sky view.

**LIDAR:** It was considered convenient to include the LIDARs in the telescope site structure since they are organised in combination with the telescope-buildings. They provide information when and in which mode the LIDAR has been fired. An online analysis tool is under development, that allows to provide cloud height and depth observed in the vertical scans. This

[15] a closed shutter leads to well identifiable pattern of low variances
[16] e.g. lightening manifests in a high variances and large threshold values




(a) The monitored T2 trigger rate in telescope 1 in Los Leones



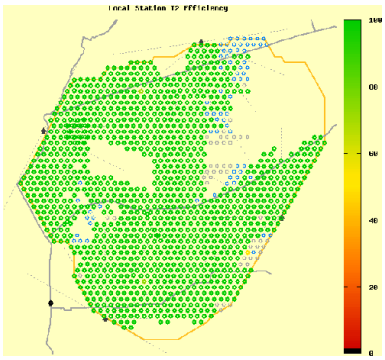

(b) The visibility of the CLF shoots every 15 minutes in the corresponding bays is an independent crosscheck of the DAQ.

**Figure 3.20:** Monitoring DAQ and CLF.

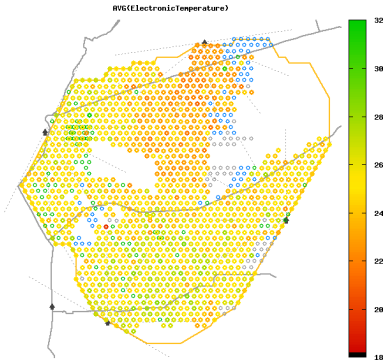
information can be very useful to decide, besides the cloud camera information, if data-taking should be stopped, e.g., if the clouds are too deep.

**Data acquisition:** The frequency of fired FLT and TLT indicates the status of the telescopes at an advanced stage. This signals show not only low-level telescope failures, but also errors in the data acquisition. Furthermore, the T3 trigger provides first reconstruction results, like distribution of the landing lines. An important feature is the monitoring of the T3 triggers send from FDAS to CDAS and compare it to the processed T3 trigger in CDAS, that allows to verify smooth operation up to the final trigger level.

Besides the calibration database, that is filled usually before and at the end of each shift, the other databases are filled continuously. The schema of data transfer is given in Fig. 3.17. The data storage is organised locally on each eye-PC. The motivation for a site-to-campus replication comes with wireless connection and the idea is copied for the data-flow of the FD-monitoring. The data-transport is organised via data-base internal replication mechanism. This is necessary because in case of communication problems from the buildings to the central campus the data are stored in the buildings and in case the connection is reestablished the replication process tries to catch up with submitted database changes. This guarantees the completeness of the data-set on the central server side, even if they are not always available online immediately because of network failures. The locally stored data and monitoring-data have to be deleted on a long term base. A nice feature of the data-base concept is furthermore that data from various detector components are accessible from one MySQL server in the end. Additionally to the FD-monitoring data, also data from the



(a) The monitored T2-rate of the SD array



(b) Technical information like SD electronics temperature is accessible

**Figure 3.21:** Monitoring the surface array.

SD and other sources like CLF[17] are streamed onto the main server.

### 3.3.7 Visualisation

As front-end to the shift crew a Php based web interface is provided. The shifters do not need to know the details of the database structure. The web pages are organised in a way that people can jump quickly between the subcomponent pages and always obtain a general survey of the components status. It is always possible to access information at a more detailed level, like variances of the pixels. Technically, an interface has been developed for the visualisations. Currently only the usage of Gnuplot is implemented. The second option used for the generation of visualizations is the JPGraph package, which is implemented for direct Php-calls on an object-oriented base with the interface defined in the inherited classes. Beside Php, also CSS and Java scripts can be used.

### 3.3.8 Alarms

Alarms, occurrences of states that require immediate action, are first filled into a specified table of the data-base. The web-front-end checks this table for new entries and indicates them on the web-page. Therefore, MySQL trigger are used, that check each incoming value. The check is performed using MySQL user-defined functions (UDF), that allow to validate if the values are within predefined ranges, set in an alarm table. This concept leaves the alarm design very flexible and extendable. If an alarm is triggered, the shifter is supposed to notice and acknowledge the alarm. After solving the problem the shifter can declare this alarm as resolved.

[17] The implementation of the CLF in the monitoring is a good example for flexibility of the monitoring schema. The operation of the CLF produces a plain text log-file, that is tunnelled by a cron job to the central database server, where a perl parser is filling the data-base. This rather simple mechanism has proven to be robust and reliable.

# Chapter 4

# Reconstruction and Data Processing

*In order to make further progress, particularly in the field of cosmic rays, it will be necessary to apply all our resources and apparatus simultaneously and side-by-side; an effort which has not yet been made, or at least, only to a limited extent.*

Viktor F. Hess

## 4.1 Data Reconstruction

### 4.1.1 Event Topologies

In the previous chapter, the trigger scheme of the Pierre Auger Observatory was introduced. The surface array and each telescope site use independent trigger schemes, except the FD-T3 trigger. In periods when both, the fluorescence and the surface array, are in operation it is possible to trigger an EAS in more than one sub-detector. Tab. 4.1 shows an overview on possible topologies. Generally the following classification can be applied:

**Normal SD:** Events triggered by the surface detector (at least SD-T3) and can be fully reconstructed by using only the information from the water-Cherenkov tanks.

**(Normal) FD-mono:** Events triggered by a fluorescence telescope (at least FD-SLT) and are fully reconstructed by using only the information from a single telescope site (*'monocular reconstruction'*). Furthermore, one distinguishes between single- or multi-mirror events. The data investigated in the present work are already at the level, where all the mirror-wise information per telescope site is merged into one event.

**Multi-fold FD:** It might by possible that an air shower is observed in more than one telescope site. Using the additional information for reconstruction enhances the quality of reconstruction. According to the number of telescopes involved in the reconstruction, one sub-classifies: FD-stereo, FD-3-fold and FD-4-fold events.

**Brass-hybrid:** Events that have been triggered by one fluorescence telescope (FD-SLT) and have additional information from at least one SD-tank (T3). These events are triggered with the FD-T3 trigger.

**Golden hybrid:** Events that have been triggered by one fluorescence telescope and the surface array (SD-T3 or FD-T3). Golden hybrid events require that they can be reconstructed independently using the fluorescence or surface array information only. Therefore these events can be used for the energy calibration of the SD (see Chap. 2.4.5).

**Platinum hybrid:** Events that are multi-fold FD, with informations from the surface array.

In the following, the different reconstruction strategies are sketched.

### 4.1.2 SD Event Reconstruction

As the focus of this work is on fluorescence data the event reconstruction is just given in a schematic overview[1]:

**Signal times:** In a first step the PMT signals for all T2 triggers are analysed and the signal times $t_i$ as well as the signal intensities $S_i$ in VEM are evaluated.

**Impact point:** The impact point of the shower core on ground $\vec{b}$ can then be found by the signal weighted mean of all triggered tank positions $\vec{x}_i$:

$$\vec{b} = \frac{\sum w_i \vec{x}_i}{\sum w_i} \quad \text{with} \quad w_i = \sqrt{S_i} \; .$$

**Full geometry:** The shower axis $\vec{a}$ is determined with a $\chi^2$-fit, assuming a flat shower front moving with speed of light. A shower passing through $\vec{x}$ at time $t$ provides the following relation by geometric reasoning:

$$\vec{x}(t) - \vec{b} = -c(t - T_0)\vec{a} \; .$$

Here, $T_0$ is the associated impact time of the shower at the core. Using the tank positions $\vec{x}_i$ and pulse times $t_i$ allows to form a $\chi^2$:

$$\chi^2 = \frac{1}{\sigma_t^2} \sum_i (ct_i - cT_0 + \vec{x}_i\vec{a})^2 \; .$$

Minimising the $\chi^2$ allows to reconstruct the full geometry of the shower.

[1] A detailed description can be found e.g. in [162, 181, 178, 155, 5]

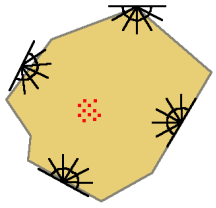
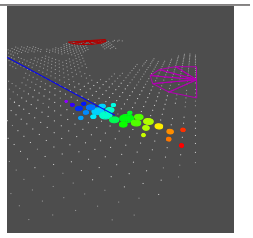
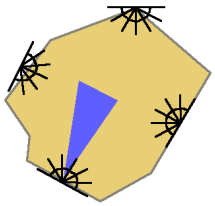
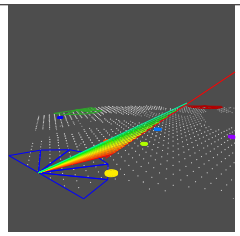
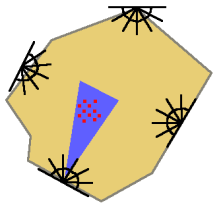
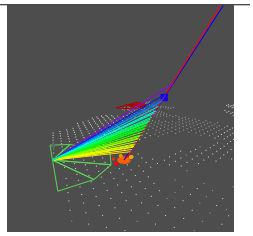
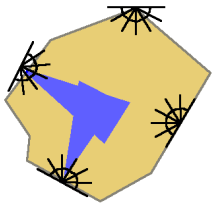
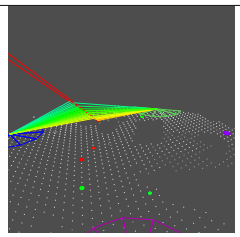
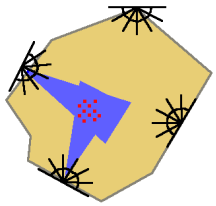
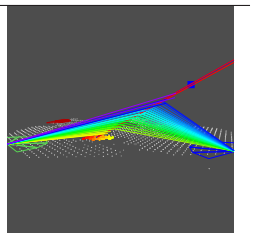
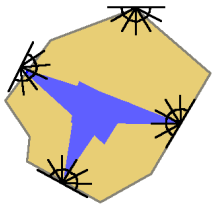
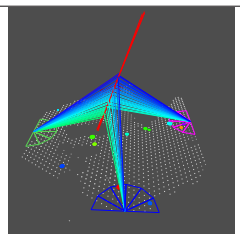
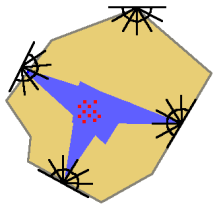
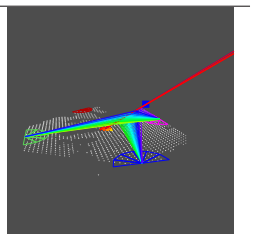
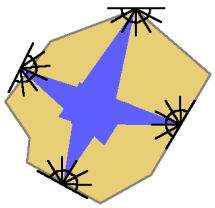
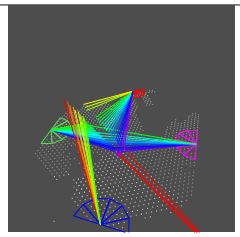
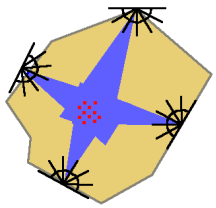
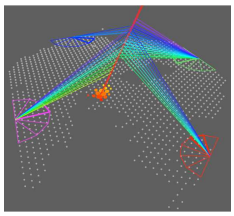
| FD-only | | Hybrid | |
|---|---|---|---|
| sketch | typical event (data) | sketch | typical event (data) |
| | |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**Table 4.1:** Possible topologies and example events observed by the AUGER experiment.

**Energy reconstruction:** The shower energy is estimated by the determination of the lateral density function (LDF) that describes the signal strength $S_i$ as function of the distance $d$ to the core $\vec{b}$ at ground. The LDF depends on the zenith angle, energy and primary composition. It has been found by Monte Carlo simulations, that at a distance of about 1000 m from the shower core the event-by-event systematics and mass-dependence are minimal. This so called $S_{1000}$-parameter scales with the energy according to:

$$E = a(S_{1000}/\mathrm{VEM})^b \,\mathrm{EeV} \ .$$

Here, the signal strength is in VEM (compare also Fig. 4.1(c)) and the zenith angle have to be taken into account. This is parametrised as [64]:

$$\begin{aligned} a &= 0.37 - 0.51\sec(\theta) + 0.30\sec^2(\theta) \\ ,b &= 1.27 - 0.27\sec(\theta) + 0.08\sec^2(\theta) \ . \end{aligned}$$

The Pierre Auger Observatory has the unique opportunity to avoid the dependence of the SD energy scale on the hadronic interaction model by using hybrid events (see Chap. 2.4.5).

**Composition** The primary composition is measured, using the rise-time of the signals far off the core (see Chap. 2.4.2).

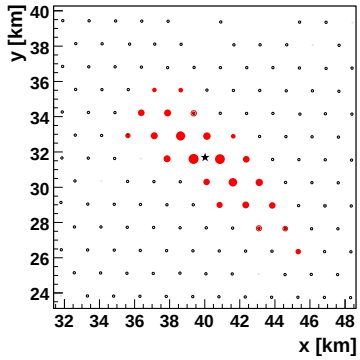
### 4.1.3 Fluorescence Data Reconstruction

The electromagnetic cascade of an EAS appears in a distant fluorescence telescope as rapidly moving spot of ultraviolet light, describing a long path across the night-sky background of starlight, atmospheric air-glow and man made civilisation light (see Chap. 2.4.1). The trace of the spot in connection with the detector location defines the so called *'shower detector plane'* (SDP) in which the shower axis has to be reconstructed. The shower axis can be fixed by investigating the angular motion of the spot, as it directly depends on both, the distance and orientation of the shower axis. With fixed the shower geometry it is possible count back the energy deposit at the shower axis. This longitudinal profile contains measures on the primary energy and composition. As this work focuses on FD-mono events, the procedure is explained in more detail.

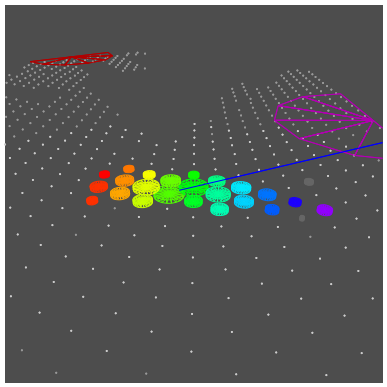
#### Extracting the Shower-Detector Plane

The SDP is defined by the shower axis and the location of the observing telescope (see Fig. 4.2(a) and Fig. 4.2(b)). Therefore, the photons, induced by the shower, should lie on a straight line in the camera (see Fig. 4.2(c)). The normal vector of the SDP, $\vec{n}$, can be obtained with a signal-weighted $\chi^2$-fit using the pointing directions $\vec{r}_i$ of the triggered photo-tubes:

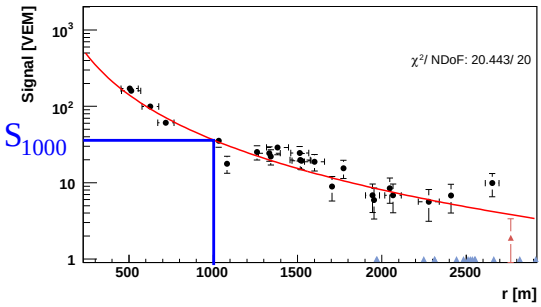
$$\chi^2 = \sum_i (\vec{n} \cdot \vec{r}_i)\, w_i \ .$$




(a) The core position (black star) as estimated by the weighted mean of all tank amplitudes.



(b) The Shower axis is fitted using the tank pulse times.



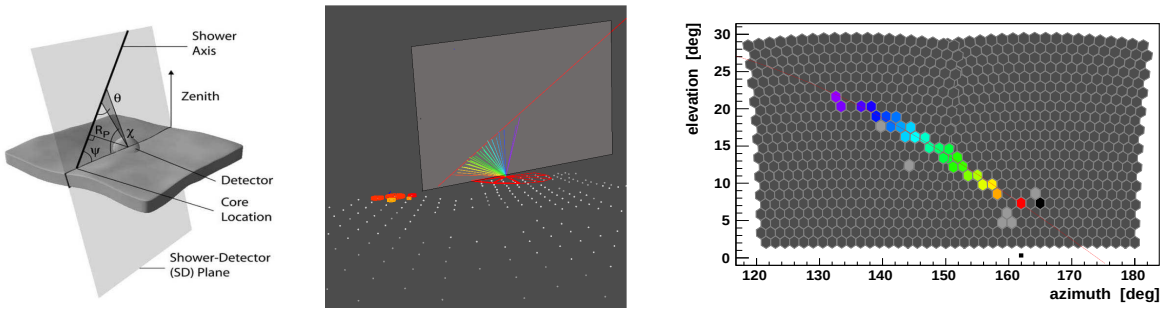

(c) The energy can be estimated via fitting the lateral density function. The interpolated signal at 1000m away from the core is a measure for the primary energy.

**Figure 4.1:** Schematic procedure of reconstructing an event from ground array data.

The weight includes the signal amplitude $w_i \propto \sqrt{S_i}$, as pixels with large signals are considered to be more reliable than noisy ones. It is obvious that the fit quality depends on the pixel selection. For this work the following constrain is used: A '*first guess*' fit was performed using all triggered pixels. For the final SDP fit only those pixels have been used that are within an interval of 5° around the '*first guess*' SDP. Reconstruction is just performed if the number of pixels with a pulse is larger than five.

### Reconstruction of the Shower Axis in the SDP by a Time-Fit

Within the SDP the shower geometry reduces to a planar problem, and can be reconstructed by finding the closest distance of the shower axis to the observer $r_p$ and the shower inclination in the




**(a)** The scheme of the shower geometry.

**(b)** A real shower observed by the FD.

**(c)** The corresponding trace seen in the camera.

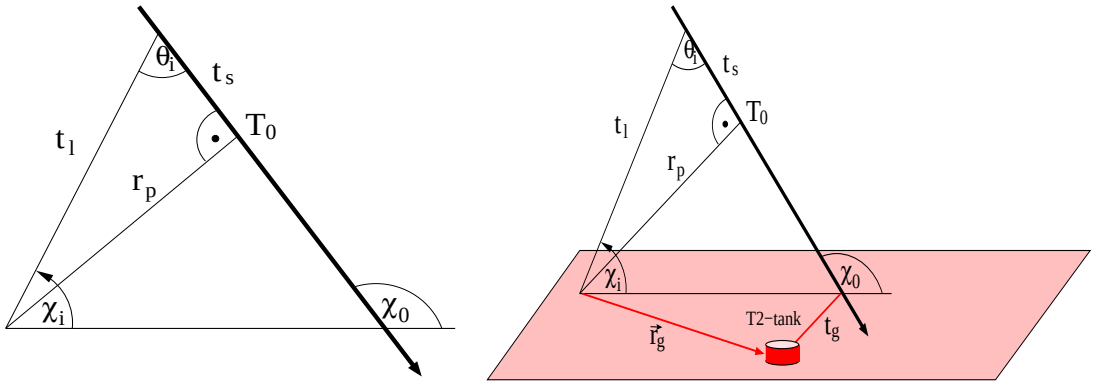
**Figure 4.2:** Reconstructing the SDP for an event triggered by Loma Amarilla FD.

SDP $\chi_0$ (see Fig. 4.3(a)). It is useful to define as reference time $t_0$ the instant, when the shower is observed at $r_p$. With simple geometry it is possible to express the time difference of a photon arriving at the $i$-th PMT by taking into account the time the shower has moved through the atmosphere $t_s$ and the propagation time of the photon to the telescope. Thus,

$$t_s = \frac{r_p}{c_0 \tan(\Theta_i)} \ .$$

In first order, one can focus on direct photons and express the propagation time delay:

$$t_l = \frac{r_p}{c_{\text{air}} \sin(\Theta_i)} \ .$$




**(a)** The FD-mono geometry scheme.

**(b)** The FD-hybrid geometry scheme.

**Figure 4.3:** The shower geometry regarded in the SDP. The shower axis is defined in space and time by $r_p$, $T_0$ and $\chi_0$.

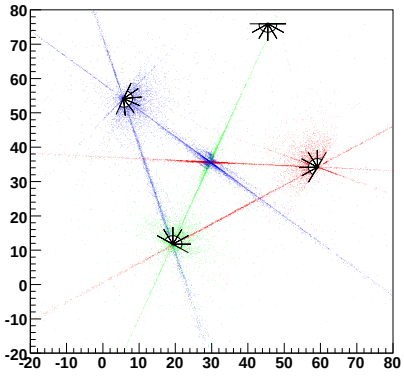
The shower can be assumed to pass with the speed of light in vacuum $c_0$, while the fluorescence photons propagate with the corresponding speed of light in the atmospheric layer $c_{\text{air}}$. Usually, the simplification is applied that the fluorescence photons propagate to the detector with the speed of light in vacuum, $c_{\text{air}} \approx c_0$. The effect of this approximation has been extensively studied in [112, 114, 113, 115]. It has been shown that for a worst case scenario the effect is at maximum 30 ns.[2] The assumption $c_{\text{air}} \approx c_0$, simplifies the expected time $t_i$ of fluorescence photons reaching the $i$-th tube:

$$\delta t = t_l + t_s = \frac{r_p}{c_{\text{air}} \sin(\Theta_i)} - \frac{r_p}{c_0 \tan(\Theta_i)} \sim \frac{r_p}{c} \tan\left(\frac{\Theta_i}{2}\right) \approx \frac{r_p}{c} \tan\left(\frac{\chi_0 - \chi_i}{2}\right) .$$

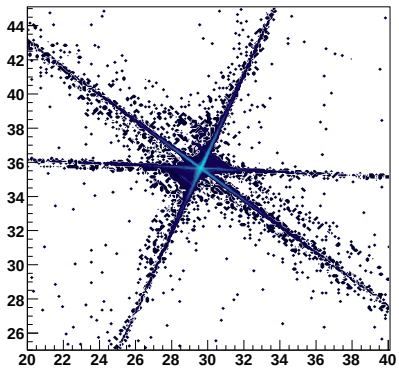
A $\chi^2$ for the arrival times each PMT can be derived:

$$\chi^2 = \sum_i w_i \left[ T_0 + \frac{r_p}{c} \tan\left(\frac{\chi_0 - \chi_i}{2}\right) - t_i \right]^2 .$$

Here, $w_i$ is a weight, that includes the error of the time measurement. Minimising the $\chi^2$ as function of $r_p$, $\chi_0$ and $T_0$ determines the complete shower geometry. Again, the result varies with the choice of PMT pulses used for the minimisation. The following constrains have been applied for pixels used in this work, to optimise the result by rejecting noise, that manifests as isolated hits in space and time:



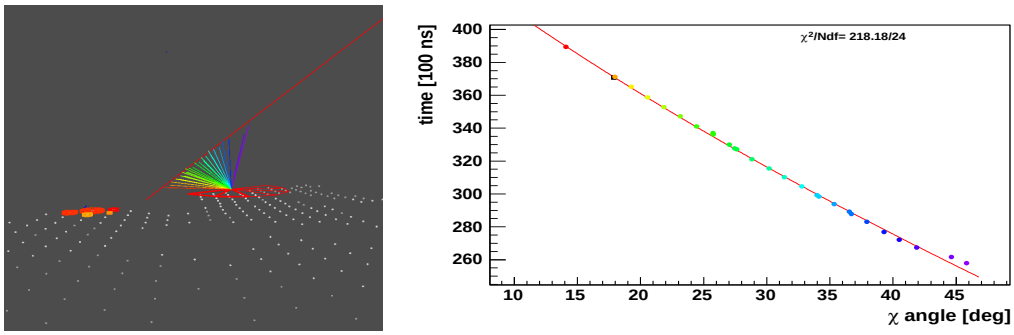

**(a)** Triggered core positions: green seen by Los Leones, red seen by Los Morados, blue seen by Coihueco.



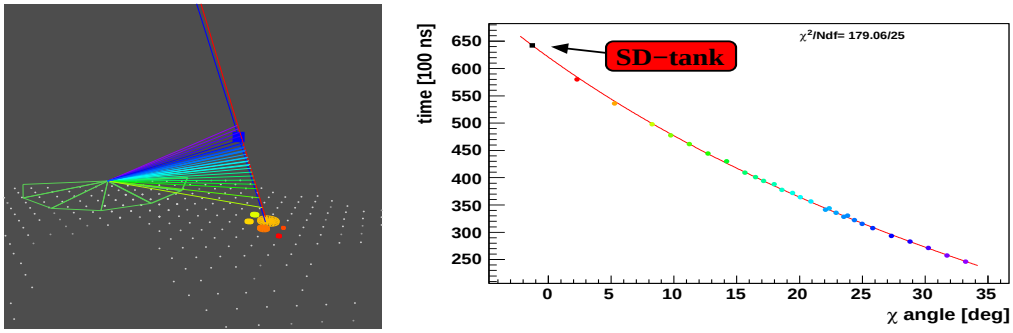

**(b)** Zoomed into the crossing point. Contour plot using logarithmic z-scale for the number of events.

**Figure 4.4:** While the SDP can be reconstructed quite accurately the crucial part is to determine the exact position of the shower axis within the SDP. This becomes obvious when looking at CLF data. The mono-reconstructed core position of CLF shots is shown here.

[2] This refers to very incline showers 40 km away, a usual delay is in the order of 20 ns.




(a) Time-Fit for an FD-Mono Event.




(b) Hybrid geometry fit: Here, the hottest tank provides a time-stamp at ground-level. One can see the improvement to the fit obtained by the long lever-arm of the SD data-point.

**Figure 4.5:** Time-fit for an FD mono and a *'hybrid'* event.

**Isolated hits in space:** Only those pixel take part in the time-fit, that are contained in an interval 2° around the SDP.

**Isolated hits in time:** Only those pixel take part in the time-fit, that have a time residual smaller than 10 $\mu$s with respect to all the others.

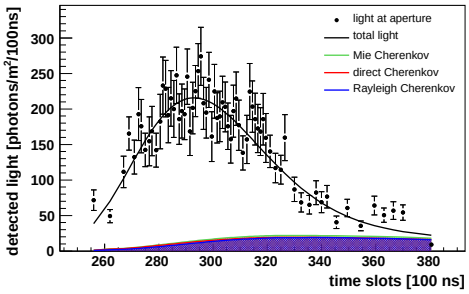
While the SDP can be reconstructed with high accuracy, the deficit of the FD-mono analysis is the exact determination of the shower in the SDP, as e.g. demonstrated in Fig. 4.4 for CLF induced events.
An improvement in resolution can be achieved by hybrid reconstruction. In this case the shower's footprint on ground is known from at least one tank information. This can be used as additional data-point for the time-fit. The schematic situation is shown in Fig. 4.3(b). The expected arrival time of the shower front at the triggered tank can be expressed via:

$$t_{\text{theo}} = T_0 + \frac{\vec{r}_g \vec{a}}{c} \, .$$

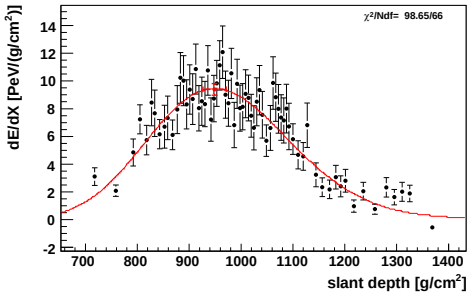
$\vec{r}_g$ is the vector pointing from the FD observer to the tank and $\vec{a}$ is the shower axis vector. This is included in the $\chi^2$ fit. Since the tank data-point is far off the FD data-points it offers a long lever-arm to determine the curvature of the time distribution. It is obvious that the exact time offset between the SD and FD detector has to be taken into account to achieve correct results. This time offset can e.g., be estimated by analysing CLF shots (see Chap. 3.3.4).
Although hybrid-data offer a great improvement in geometrical resolution the sample of hybrid events is smaller than the FD-mono due to two reasons:

- The ground array is nearly completed and will be finished in Feb. 2008, but for the analysed data[3] the ground array has regions, that were not instrumented, and thus it was not possible to obtain SD information there.
- The trigger threshold of the surface array is at about $10^{18}$ eV. Therefore, low energy showers landing near the fluorescence detector can trigger the fluorescence telescopes, but are only seen in the surface array, if the core is close to a surface tank. The tank-spacing of 1.5 km induces a large fraction of low energy events without tank information, even if the area is instrumented.

### Reconstruction of the Longitudinal Profile




(a) Light at aperture, reconstructed contribution to the detected light from Rayleigh scattering and from direct as well as multiple scattered Cherenkov contribution.




(b) By subtracting the Rayleigh photons and Cherenkov contribution one can calculate the fluorescence photons emitted by the shower at a given depth. The energy deposit gives the reconstructed longitudinal profile. A Gaisser-Hillas fit is done.

**Figure 4.6:** Reconstructing the longitudinal profile. The event was detected in Loma Amarilla. The energy was reconstructed to be $3.55 \pm 0.14$ EeV and $X_{max} = 949 \pm 22\,\mathrm{gcm}^{-2}$.

Once the full geometry is reconstructed, it is possible to obtain the longitudinal profile of the shower. This is done in several steps.

[3] since 2004

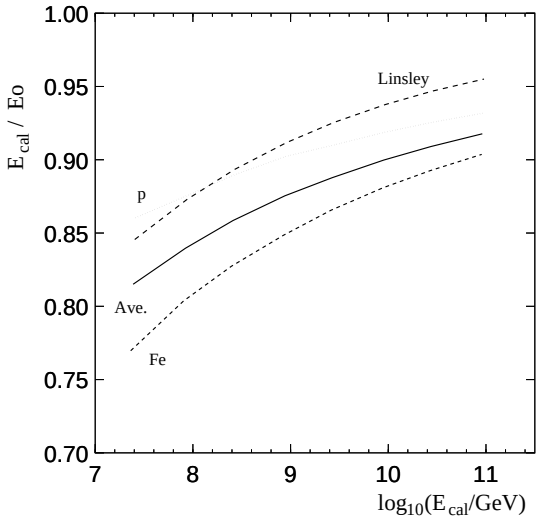



**Figure 4.7:** Conversion of electromagnetic energy deposit and primary energy taken from [171]. The curves show the correction of the calorimetric energy to the primary energy as function of primary energy for proton induced (dotted line) and iron induced (lower dashed line) EAS. Additionally, the Linsley function (upper dashed line) is plotted for comparison. The average is given as solid line.

for a known shower geometry, the longitudinal profile can be obtained by the following steps:

- Firstly, the FADC pulses in each pixels have to be associated with photons arriving at the diaphragm. Therefore, the triggered time window is binned in time slots. For each pixel and time slot the number of detected photons is calculated using the individual calibration constants. Fig. 4.6(a) shows a typical reconstructed aperture light curve.

- Inverting the time-fit relation, it is possible to find the corresponding inclinations $\chi_i$ of each time-slot in the SDP:

$$\chi_i = \chi_0 - 2 \arctan\left[\frac{r_p}{c}(t_i - T_0)\right] \ .$$

  Using the reconstructed light flux at the detector, the photons can be propagated back through the atmosphere to their origin at the shower axis. In this step, the specific atmospheric conditions have to be taken into account to correct for attenuation. This procedure provides the number of photons at the shower axis as function of slant depth, which can be converted into the energy deposit per slant depth.

- The shower geometry and energy deposit per slant depth allow to estimate the Cherenkov photons fraction. In an iterative procedure the Cherenkov photons can be tracked forward to

the telescope, to obtain an estimate of how many photons in the aperture light flux are due to direct and indirect Cherenkov, as well as Rayleigh scattered light. They are subtracted from the aperture light and the procedure of back-tracing is repeated. Usually this iterative procedure converges, and finally the energy deposit from the electromagnetic cascade into the fluorescence light is obtained.

- This final longitudinal profile is fitted using a Gaisser-Hillas parameterisation (Eqn. 2.8), which allows to extrapolate the shower size.

Finally, the energy of the electromagnetic cascade can be obtained directly by integration on the atmospheric depth:

$$E_{\mathrm{elm}} = \int dX \frac{dE}{dX} \,.$$

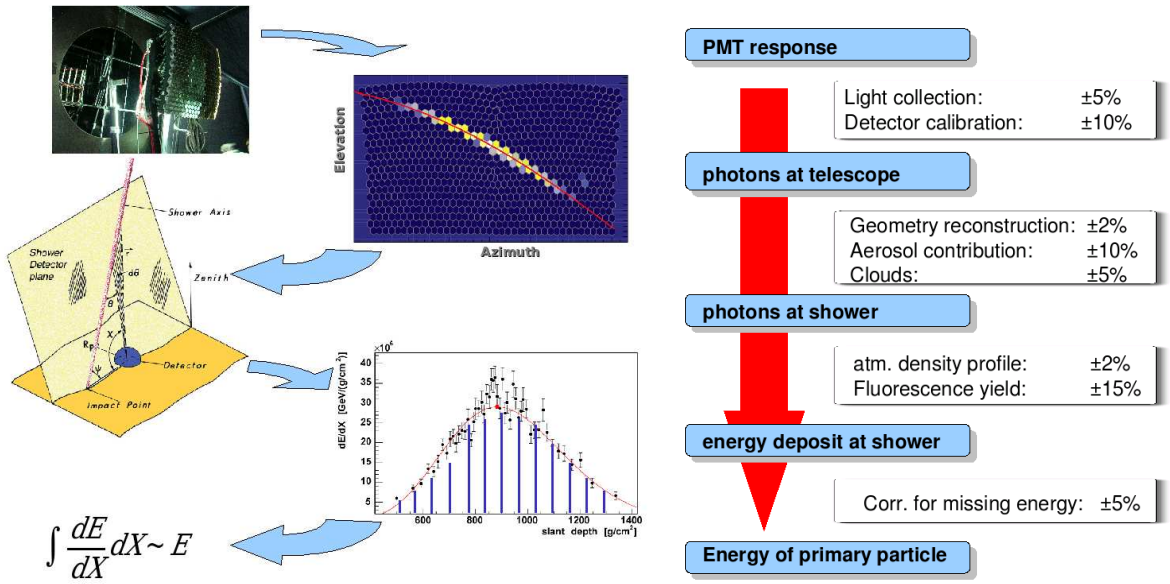
This energy of the electromagnetic cascade can be converted into the primary energy. One has to correct for the amount of energy that is not detected by the fluorescence telescope. This so called unseen energy depends on the primary composition and is therefore source for uncertainties in the total energy estimation. An average correction for proton and iron induced showers is given by [171]:

$$\frac{E_{\mathrm{elm}}}{E_{\mathrm{tot}}} = 0.959 - 0.082\ E_{\mathrm{elm}}^{-0.15}$$

The two extreme cases are shown in Fig. 4.7. Obviously, the systematic uncertainties due to unseen energy can be estimated to be smaller than 5 %. Nevertheless, it is important to keep in mind that the Gaisser-Hillas fit is a semi-empiric model that is not valid to describe pure hadronic or pure electromagnetic cascades. Therefore, one has to check if the Gaisser-Hillas fit really describes the reconstructed longitudinal profile.

## 4.2 Systematic Uncertainties

The described reconstruction scheme using the FD information alone is shown in Fig. 4.8(a). Once the geometry is determined the transformation of measured photo-electrons (PE) into charged particles at shower depth has to be done. This process includes light propagation properties of the atmosphere as well as understanding of the PMT response. The major systematic uncertainties stem from this conversion process and are listed in Fig. 4.8. First, the PE are translated into photons at the camera using the calibration constants. The light collection efficiency of the PMTs is known within 5 % uncertainty, whereas the calibration constants can be measured with an uncertainty of 10 %. The back-propagation of the photons from the camera to the shower axis depends on the shower geometry. The influence of the geometry reconstruction accuracy on the conversion is less than 2 %. Weather conditions play a major role. Clouds can absorb and scatter light. This makes it important to monitor the cloud coverage. Their systematic influence on the final energy is about 5 %. The absorption of aerosols can be determined within 10 %.
The number of emitted photons at shower axis has to be converted into the energy deposit of




(a) Reconstruction scheme FD only

(b) Systematics FD only

**Figure 4.8:** Table of major sources for systematic uncertainties in the FD event reconstruction. Adding the uncertainties in quadrature results in a 25 % uncertainty in the reconstruction of the primary energy. Numbers are based on [102, 28, 180].

charged particles at a certain shower depth. Hereby, it is important to know the exact atmospheric density profile and the fluorescence yield. The density can be assumed to be known by 2 %, whereas the major impact on the systematics of 15 % comes from the fluorescence yield. The amount of invisible energy which goes into muons and neutrinos depends on the shower development and is subject to fluctuations. For protons this correction is about 5 %, for iron induced showers the mean correction is about 15 %. Uncertainties from missing energy corrections can be assumed to be smaller than 10 %. A detailed summary and discussion of the FD uncertainties can be found in [102, 28, 180].

## 4.3 The Offline Framework

The Offline framework [14] provides the core machinery to develop and apply SD, FD and hybrid simulation- and reconstruction-code. Furthermore, it is designed according to the requirements of all various simulation and reconstruction tasks, and thus, to make it easier to share code and collaborate on the physics problems. Briefly, the Offline framework offers the following services:

- Machinery for modularising physics algorithms, so that different algorithms can be easily compared and/or arranged in different ways to carry out various tasks.

- Structures for building up simulated and reconstructed events and handling persistency (storing on disk), including the ability to read from and write to a variety of existing file formats[4].
- Access to information of the detector configuration and performance as a function of time, including data stored in MySQL databases.
- Machinery for managing configuration files.
- A collection of utilities to assist in geometrical manipulation, parsing files, logging error messages, and converting units, as well as a number of useful classes representing traces, time intervals and time stamps, tabulated functions, particles, etc..
- A simple-to-use and robust built system based on the GNU auto-tools.

## 4.4 The Simulation and Reconstruction on the ALiCENext Cluster

A full simulation reconstruction chain has been setup using the Wuppertal computer farming machinery on the ALiCENext cluster.

### 4.4.1 The ALiCENext Cluster

In June 2004, the new computer cluster ALiCENext (Advanced Linux Cluster Engine, next generation) has been installed at the University of Wuppertal. The machinery consists of 1024 AMD-Opteron 1.8 GHz processors. The processors are distributed on 512 blades. Each blade is equipped with two 160 GB hard-discs, two times 1024 MByte RAM and six Gigabit-Ethernet connections. The blades itself are mounted in eleven towers. Each tower keeps 48 blades arranged in four rows with twelve blades each. The network is a combination of Gigabit-Ethernet and 2-dimensional torus topologies. Linux is used as operating system and processes get launched by the Torque batch system.

### 4.4.2 The Mass-Production Scheme

The full mass-production can be divided into two steps: i) The simulation of EAS with CORSIKA and ii) the detector simulation-reconstruction chain using the Offline framework.

**CORSIKA production:** CORSIKA v6.2 is used as shower generator, thinning level is $\epsilon = 10^{-5}$ and QGSJET/FLUKA are applied as high, resp. low energy hadronic interaction model. The CORSIKA package has been installed on each node[5]. A perl script generates executable scripts, that launch CORSIKA and copy back the result-files to a common storage. A queueing system distributes the jobs to the nodes. In case the production is finished a booking script checks

[4] like aires, CORSIKA, CDAS, FDEventLib

[5] We set up also a watchdog daemon checking on a nightly base the package status of all nodes. This daemon reinstalls CORSIKA, in case a node has been off at the time of the last software-update.

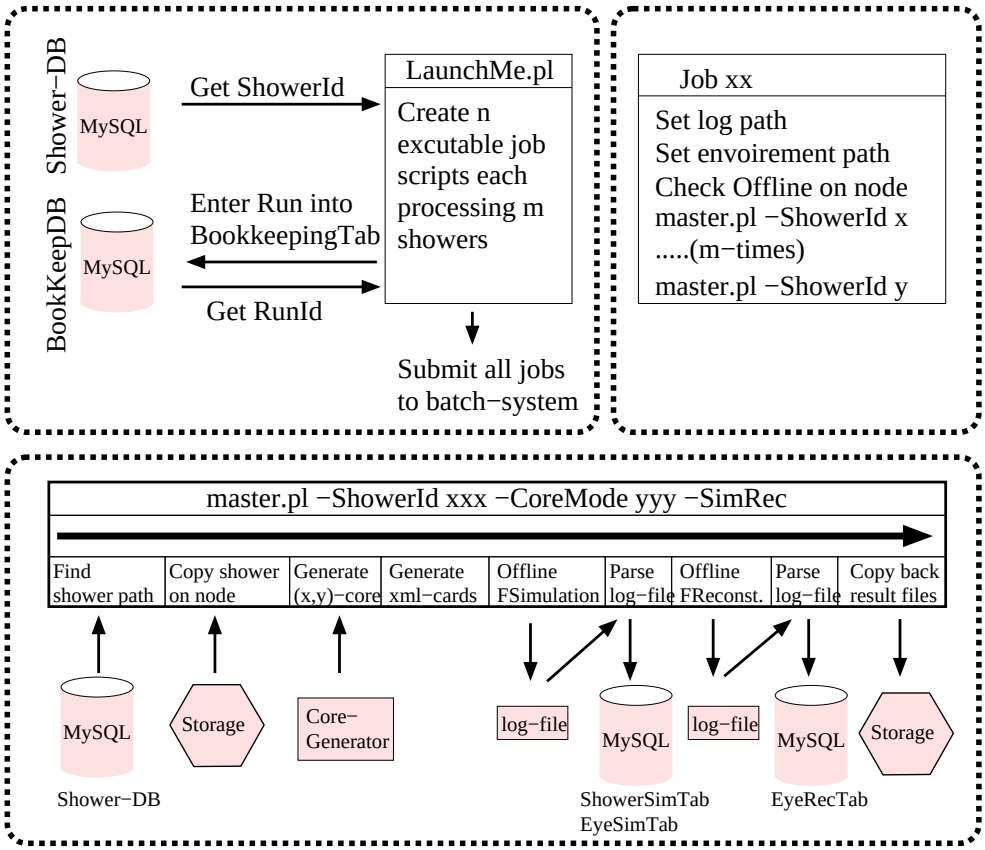



**Figure 4.9:** The MassProd chain installed on ALiCENext.

the produced CORSIKA files for their completeness and stores the file locations, and other meta-data into a MySQL database.

**Simulation-Reconstruction chain:** A schematic sketch of the simulation-reconstruction chain is given in Fig. 4.9. For the presented analysis is based on the Offline version 'DrEvil', that has been installed on each node. A cron job is checking on a nightly base the installation status on each node and guarantees that all nodes in the queueing system provide the same Offline installation. The backbone of the chain is the LaunchMe.pl script. This script looks up which shower sample is specified by the user (e.g., proton induced CORSIKA showers with $E^{-2}$ energy spectrum). The specific features of the simulation-reconstruction run are stored to a global bookkeeping database. From this database the script queries for the current RunId identifier of the production. Finally, the script creates executable job-files that are submitted to the batch system.

Each job processes a set of CORSIKA-showers. The job executable sets the log paths and

verifies, if all needed programs and libraries are installed on the node. Then, the job launches the master.pl script, which controls the simulation-reconstruction sequence.
The master.pl script retrieves from the shower database the location of the CORSIKA files on the storage. These files are copied locally to the node. Then, the needed XML-cards for the specific runs are created. The core position is generated by an external program called CoreGenerator. This tool was implemented to allow dedicated core simulations (e.g., homogeneous on ground in the field of view of a certain telescope or fixed core position.) Then, master.pl starts Offline using the module sequence for FSimulation. The final module of this sequence writes out a set of generated observables to a log file. master.pl parses this log file and inserts the exact variables to a MySQL table. In the same way the reconstruction is performed parsing the corresponding reconstructed parameters to a dedicated database. Finally, the result files are copied back to the storage.
The chain can also be used in a *reconstruction-only* mode[6]. In this case only the FReconstruction is executed by the master.pl script. Instead of a ShowerId identifier a DataRunId identifier is promoted. A MySQL table allows to lookup the location of the data-run on the storage. Reconstructed parameters are inserted into a database.

## 4.5 MySQL as DST

For detailed studies of reconstruction performance, it is needed to investigate low-level variables, like FADC traces. For this reason, an additional root file is stored for each event, which contains the full information. But for validation and general performance studies usually only high-level variables like number of triggered pixels or reconstructed and true geometry are compared for large statistics. In this case it is suitable to have a database. The variables stored in the database are only high-level observables. The backbone of the mass-production is a MySQL database used as DST[7]. A schematic view of the full database layout is given in App. E.1. The idea to use a relational database was guided by the following motivations:

- A database allows to access and correlate information during analysis flexible.
- For running the mass-production many lookup tables have to be accessible at a central place, e.g., the BookOfEvil table, that serves as bookkeeping table to store the specific run characteristics of each mass-production and generates a unique RunId identifier. This task can also be achieved using plain ASCII files, but the advantage of MySQL is to allow concurrent access, that guarantees, that two clients cannot get the same RunId. This is a critical requirement when running the production on a large cluster with typically 300 nodes running in parallel.
- In the same manner, it is possible to store generated data and reconstructed data by many clients onto a server instantaneously.

[6] e.g., to reconstruct experimental data
[7] In reminiscence to the '*Data summary tapes*' once commonly used in high energy physics. DST is considered in this context more like '*Data storage tree*'.

- A database is located at the front-end of ALiCENext. Using MySQL replication provides simultaneous back-up to a user database.
- A database allows many users to fill and select data at the same time.

# Chapter 5

# Validation of the MC Simulation and Reconstruction

*A computer terminal is not some clunky old television with a typewriter in front of it. It is an interface where the mind and body can connect with the universe and move bits of it about.*

Douglas Adam

In general, the reconstruction is the part of the data analysis, that transforms data to measures. The knowledge of the trigger efficiency is essential in the final interpretation of the results. Therefore, we investigate the trigger response of the fluorescence telescopes based on MC and validate the performance of the reconstruction algorithm itself.

## 5.1 The MC Sample

A dedicated MC sample has been generated to evaluate the performance of the simulation-reconstruction chain. The sample consists of mono-energetic proton and iron induced CORSIKA-showers[1]. The energies are ranging from $10^{17}$ eV to $10^{21.5}$ eV in discrete steps of 0.5 in a $\log_{10}$-scale. Each energy bin consists of about 750 proton and 750 iron induced showers. For this study the core locations of the showers have been placed in the middle of the field of view of telescope 5 in Los Leones at various distances[2]. This allows to investigate in detail the dependency of the reconstruction performance as function of energy and distance from the telescope. The angular

[1] CORSIKA v6.2, thinning level is $\epsilon = 10^{-5}$ and QGSJET/FLUKA for high, resp. low energy hadronic interaction
[2] 10 km, 15 km, 20 km, 25 km, 30 km, 35 km, 40 km

distribution follows:

$$dN \propto \sin(\theta)\cos(\theta)d\theta d\phi \ . \tag{5.1}$$

This corresponds to an isotropic flux[3]. The characteristics of the generated MC sample are summarised in Fig. 5.1.

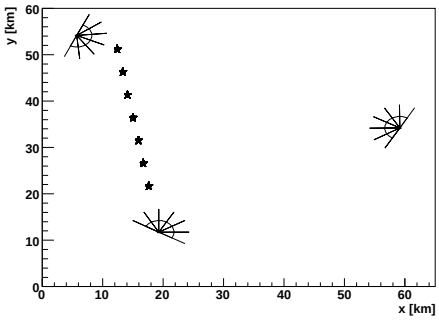
## 5.2 Fluorescence Data Simulation

The simulation was performed using the Offline framework[4] of the Pierre Auger collaboration (see Chap. 4.3) including full detector and trigger response. The simulated detector response is provided in the same format as real data, before being reconstructed. The reconstruction procedure is the same as the one applied to experimental data. In detail, the detector response is simulated by the following procedure within the framework (Fig. 5.2):
An EventFileReader provides the framework access to the CORSIKA longitudinal profile. Afterwards, the CoreGenerator fixes the shower on ground. The longitudinal shower profile produced by the CORSIKA code contains informations about the energy deposit as a function of atmospheric depth. This has to be converted into numbers of fluorescence photons emitted along the shower axis. This step is performed by the ShowerLightSimulator. In addition this module provides the produced Cherenkov light. Subsequently, these photons are propagated through the atmosphere to the diaphragm. This is done by the LightAtDiaphragmSimulator module. Afterwards, the ray-tracing through each individual telescope is performed using the TelescopeSimulator module. An additional background module simulates additional background noise from night sky photons. The final FADC traces are calculated by the FdElectronicsSimulator by generating the electronics response to the light of each PMT. The trigger schema as described in Chap. 3.3.2, is simulated by the FdTriggerSimulator module. At this step the complete simulated detector response information is available and has to be promoted to its final output format. This is done by the EventBuilder module. In a final step a UserModule establishes the connection to the MySQL database to write out the basic simulated observables. Reconstruction is performed as described in Chap. 4.1.3.

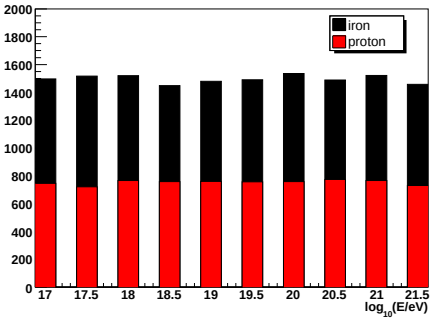
## 5.3 Trigger Performance

The MC sample is suited for investigations of the trigger performance of FD-mono events (see Chap. 4.1.1). The SLT efficiency is studied as function of energy and distance (see Fig. 5.3(a)). Its accurate knowledge is essential to evaluate the final aperture. For the final spectral unfolding, the aperture is directly folded in the end into the response matrix of an unfolding problem. However, we emphasise that based on MC it is hard to quantify the correctness of the trigger-response simulation in terms of how good it represents reality. This can be done by comparison to experimental data (see

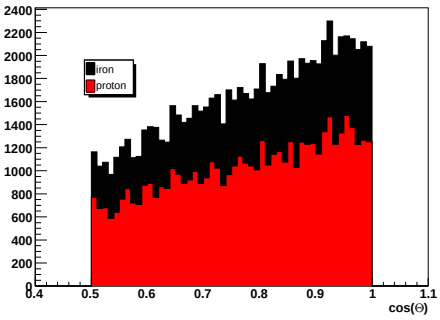
[3]The $\cos(\theta)$ term reflects that the detector observes just the projection of the flux. The formulation is motivated in more detail in the next chapter.
[4]Dr.Evil-version




(a) Core positions are chosen in the middle of the FoV of bay 5, Los Leones at discrete distances.




(b) Mono-energetic sample allows to study purely the performance as function of energy.




(c) The zenith distribution follows an isotropic flux.

**Figure 5.1:** Features of the MC sample, that is used to validate the simulation-reconstruction chain.

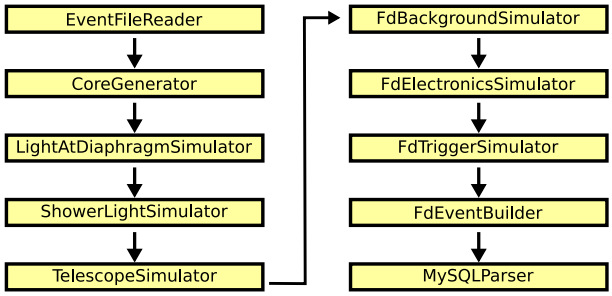
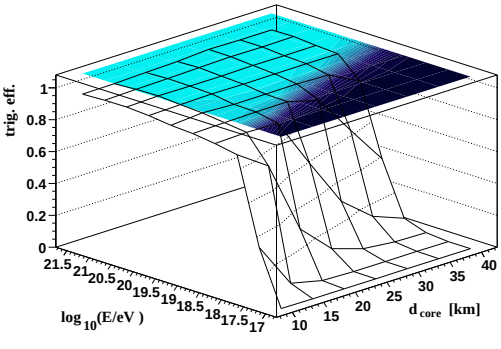
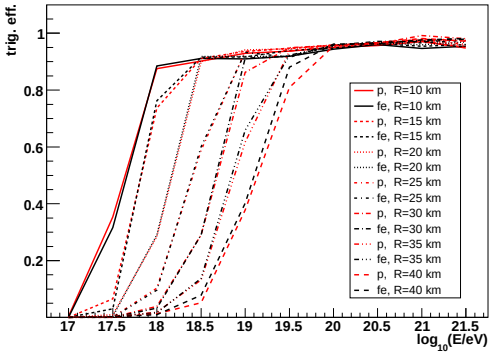


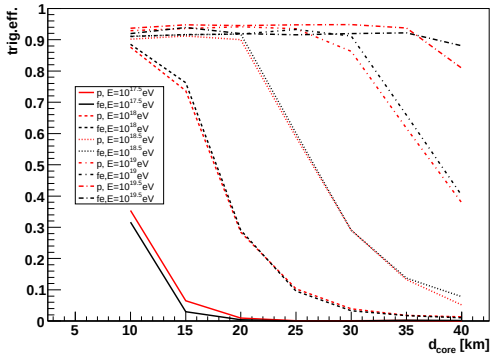

**Figure 5.2:** The module sequence used for FSimulation



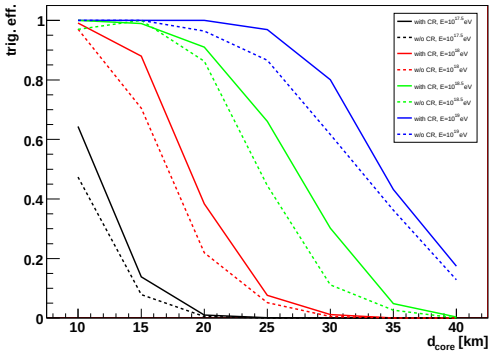

(a) Trigger efficiency as function of the core distance and primary energy.




(b) Trigger efficiency as function of the primary energy.




(c) Trigger efficiency as function of the core distance.




(d) Influence of the corrector rings.

**Figure 5.3:** The SLT performance.

Chap. 6). Moreover, it is important to check if the trigger performance is reliable and reasonable. Additionally, it allows to investigate systematic uncertainties coming from composition and detector configuration. Therefore, the trigger-rate of FD-mono is given for protons and irons individually. For the spectral unfolding (see Chap. 7) the MC is composed of proton and iron induced showers (the exact mixture is motivated in Chap. 6). It is essential to quantify the systematics arising from the composition. The trigger efficiency as function of the shower energy for various core distances is given in Fig. 5.3(b). The vice versa projection, namely the efficiency as function of distance for various energies, is plotted in Fig. 5.3(c). The influence of the primary composition can be estimated to be less than 3 % on the total trigger efficiency in the investigated parameter space. This study is performed in telescope 5 for a time period when the corresponding aperture is mounted without corrector ring lens. Hence, the simulation and reconstruction settings have been adjusted to a telescope without corrector ring. Telescopes with different corrector lens configurations are used for the spectral unfolding (see Chap. 7) and data-MC comparison (see Chap. 6). Therefore, we also investigate the influence of the corrector lenses. In Fig. 5.3(d), the trigger efficiency with and without corrector lens for a 50 %proton and 50 % iron composition is plotted as function of distance for various energies. This plot is based on a different MC sample using homogeneous distributed core positions on ground and a continuous $E^2$ spectrum. Nevertheless, the effect becomes visible. The corrector rings allow to measure showers out to larger distances. A more detailed investigation of trigger efficiencies for various event topologies and the influence of the corrector rings has been performed and the results are presented in [30], but the details of this presentations are beyond the scope of the present work.

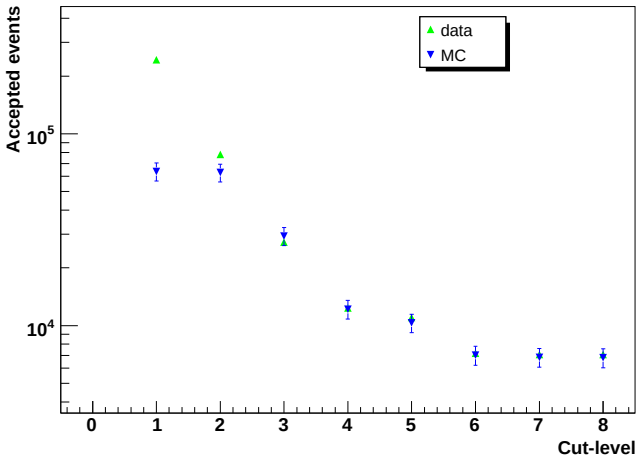
## 5.4 Reconstruction and Performance Cuts

For the present work, MC and data are restricted by eight successive cuts. Some of these cuts can be considered as *sine qua non* for reconstructing the energy spectrum, like geometry and profile feasible to reconstruct. Furthermore, there are quality cuts to guarantee reliable reconstruction parameters. The effects of the cuts on data and MC prediction are slightly different and represent the fact of non-physics events in the data that are not treated in the simulation. On the other hand, data and MC predictions[5] converge at a fairly good level, as shown in Fig. 5.4. In the following context, we would like to motivate the cuts in detail:

**Cut 1 (SLT triggered):** The second level trigger is described in Chap. 3.3.2. Only events with a collinear pattern are promoted to reduce noise events.

**Cut 2 Geometry found):** Only events with a reconstructible geometry are investigated in this study. As explained in Chap. 4.1.3 the geometry reconstruction for FD-mono events consists of two steps: SDP determination and time-fit. In the present work, the fits are optimised by rejecting isolated pixels in time and space to increase the reconstruction accuracy (see

[5]The calculation of the MC prediction is explained in Chap. 6 in detail. The result is anticipated to illustrate the cut performance.




| cut | description | passing rate [% of prev. cut] | |
|---|---|---|---|
| | | MC | data |
| 1 | SLT triggered | $\sim 5$ | - |
| 2 | geometry found | 98.6 | 32.2 |
| 3 | GH fit possible | 46.7 | 34.9 |
| 4 | $X_{max}$observed | 41.6 | 45.3 |
| 5 | profile$> 200\,\mathrm{gcm}^{-2}$ | 84.6 | 89.0 |
| 6 | no MINUIT warnings | 68.0 | 65.1 |
| 7 | $\chi^2_{\mathrm{TF}}/NDF_{\mathrm{TF}} < 5$ | 97.5 | 98.6 |
| 8 | $\chi^2_{\mathrm{GH}}/NDF_{\mathrm{GH}} < 5$ | 99.5 | 99.9 |

**Figure 5.4:** Cut-acceptance for MC and real data.

Chap. 4.1.3). The remaining pixels are refered to as 'time-fit' pixels. As the time-fit is a three-parameter fit[6], the geometry can be only reconstructed if the event has more than three time-fit pixels.

**Cut 3 (Gaisser-Hillas fit possible):** The main purpose of this work is the unfold of the energy spectrum. The energy estimation requires a longitudinal profile. Thus, it is essential, that enough pulses are available, to fit a Gaisser-Hillas curve. This Gaisser-Hillas parametrisation is a 4-parameter fit, thus at least four pulses are required. Furthermore, the minimisation has to converge to obtain reliable parameters.

**Cut 4: $X_{max}$ observed):** A characteristic feature of the longitudinal profile is the depth of the shower maximum. This cut requires that the reconstructed $X_{max}$ is located between the minimum and maximum observed depth of the shower trace. Without the $X_{max}$ being observed it is hard to fit a reliable normalisation to the GH-curve.

**Cut 5 (profile $> 200\,\mathrm{gcm}^{-2}$):** This quality cut allows in connection with Cut 4 to accept only events where at least $200\,\mathrm{gcm}^{-2}$ around the observed $X_{max}$ are covered by data. The GH-curve for those events becomes distinct.

**Cut 6 (no MINUIT warnings):** This is a technical cut. For the complete FD-mono reconstruction three fits have to be performed: SDP-fit, time-fit and GH-fit. Each fit is implemented as $\chi^2$-minimisation using MINUIT [93]. MINUIT has at anytime its own current estimates of the parameter uncertainties, which it prints out on request, no matter how accurate they are at

[6] $R_p$, $t_0$, $\chi_0$

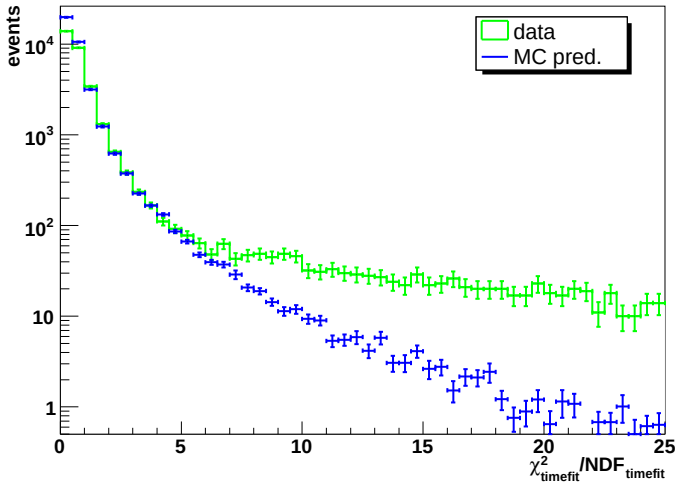



**Figure 5.5:** $\chi^2/NDF$ behaviour for MC and data. Obviously, in real data more events with bad fit performances are contained that are not simulated by MC. Therefore, a cut at $\chi^2/NDF < 5$ for the geometry fit performance is reasonable.

any given point in the execution. E. g., at initialisation, these estimates are just the starting step sizes as specified by the user. After a MIGRAD or HESSE process step, the uncertainties are usually quite accurately estimated, unless there has been a problem. Usually, warning messages are produced during the minimisation or error analysis. Although the spectrum of reasons is wide, some warnings indicate that the final fit result or at least its uncertainties are not reliable. E.g., if the Hesse-matrix is not positive-definite. Even at the minimum, this may mean that the solution is not well-defined. Typical reasons are more unknowns than there are data points, or that the parameterisation of the fit contains a linear dependence. If this is the case, then MINUIT (or any other program) cannot solve the problem uniquely, and the error matrix is necessarily meaningless, so the user must remove the under-determinedness by reformulating the parameterisation. MINUIT cannot do this itself. It is possible that the apparent lack of positive-definiteness is in fact only due to excessive roundoff errors in numerical calculations in the user function or not enough numerical precision. This is unlikely in general, but becomes more likely if the number of free parameters is very large, or if the parameters are badly scaled (as it is e.g., the case for the GH-fit), and correlations are large, too. In any case, whether the non-positive definiteness is real or only numerically, is largely irrelevant, the error matrix is unreliable and the minimum suspicious. Therefore, for the final data-set, we require, that all events had no MINUIT warning during their minimisation. Of course, some of the warnings can be neglected and the result is still trustable, but it is technically very complex to investigate individually the different flavours of warnings. As

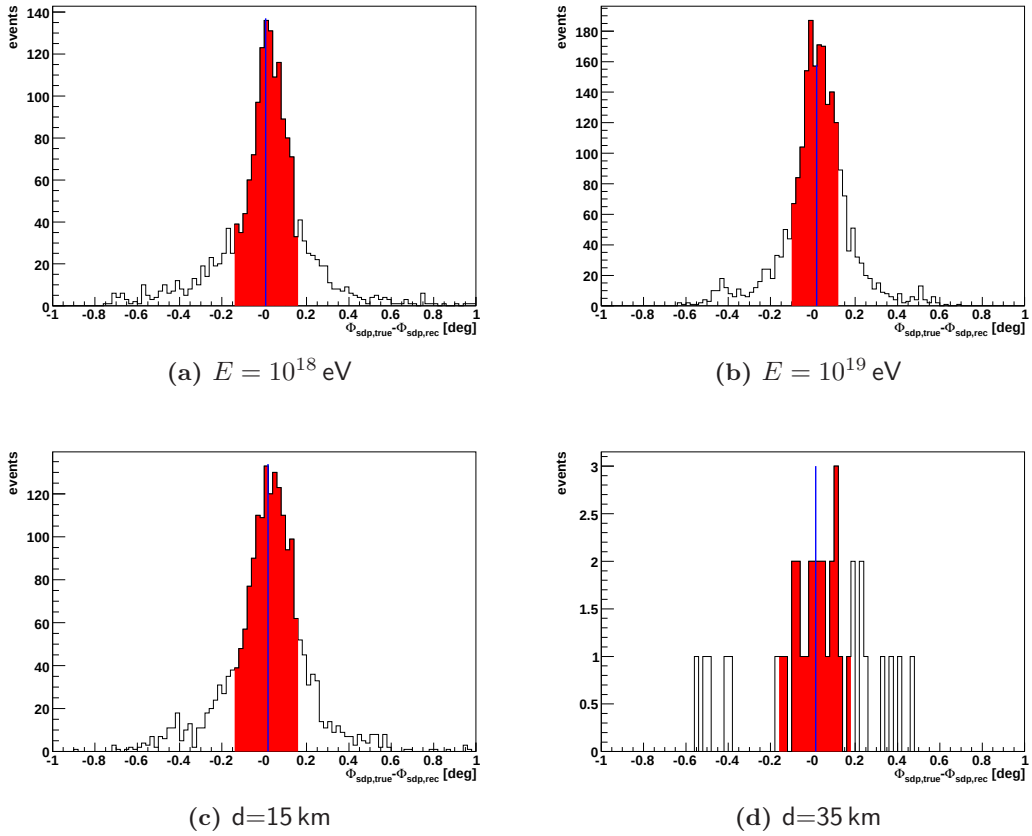
shown in Fig. 5.4, this cut rejects about one third of all events. The gain from investigating each warning individually would be about 10 %.[7] The gain is not worth while a very complex analysis of all MINUIT warning. By restricting on only those with no warnings a very pure sample with reliable results is obtained.

**Cut 7 ($\chi^2_{\mathrm{TF}}/NDF_{\mathrm{TF}} < 5$):** This is a quality cut on the reconstructed geometry. As a reliable reconstruction is essential for a good energy determination we focus only on those events that we understand from the geometry point of view rather accurate. Although included in the simulation the night sky noise results in a much larger fraction of bad reconstructible event geometries in real data than in MC. This is demonstrated in Fig. 5.5. In this figure the MC has been simulated according to a physical assumption and compared to data. The exact calculation of the MC prediction is described in Chap. 6 in detail. The result is anticipated to motivate the upper bound of $\chi^2_{\mathrm{TF}}/NDF_{\mathrm{TF}} < 5$. Obviously, the $\chi^2_{\mathrm{TF}}/NDF_{\mathrm{TF}}$ distribution of MC prediction and data agree up to $\chi^2_{\mathrm{TF}}/NDF_{\mathrm{TF}} = 5$ fairly good in shape and normalisation, while data show an excess in bad reconstructible geometries above.

**Cut 8 ($\chi^2_{\mathrm{GH}}/NDF_{\mathrm{GH}} < 5$):** This quality cut guarantees reliable energy reconstruction. A consequence of this cut is a suppression of events triggered at highest energies. This effect becomes important above $10^{20}$ eV and is mainly due to the pixelization of the camera. The signal in each time slot is rather large and thus the relative statistical uncertainty becomes small. The reconstructed longitudinal profile shows a kind of wavy undulation on its shape, that is due to the Mercedes light guides. The light suppression when the spot is crossing between two pixels has to be corrected. This effect is under investigation. Currently, this effect leads to a worse $\chi^2_{\mathrm{GH}}/NDF_{\mathrm{GH}}$ for high energy events. For reconstructing the energy spectrum around the ankle the effect is negligible and the cut is affecting both data and MC in the same manner. Thus, reliable results are guaranteed.

## 5.5 Performance of the FD-Reconstruction Using a Full Simulation-Reconstruction Chain

The reconstructed MC sample can be used to compare observable at generator and reconstruction level. This allows to validate the reconstruction. Hereby, the difference between reconstructed and true variable is histogramed as function of energy, core distance and zenith. The mean of this distribution can be interpreted as systematic shift in the reconstruction chain. Around the mean the spread in which 68 % of all events are contained is calculated. This spread $\sigma_{68\%}$ is interpreted as reconstruction uncertainty.




**(a)** $E = 10^{18}$ eV **(b)** $E = 10^{19}$ eV

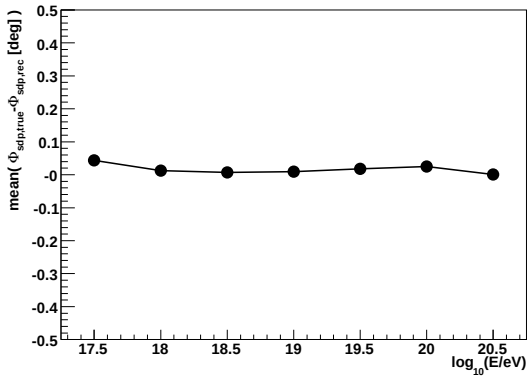
**(c)** d=15 km **(d)** d=35 km

**Figure 5.6:** Resolution of the SDP vector. Here as example the $\phi$-component of the orthonormal vector for some energies and distances. The blue line indicates the mean of the distributions and the red area shows the region in which 68 % of all events fall.

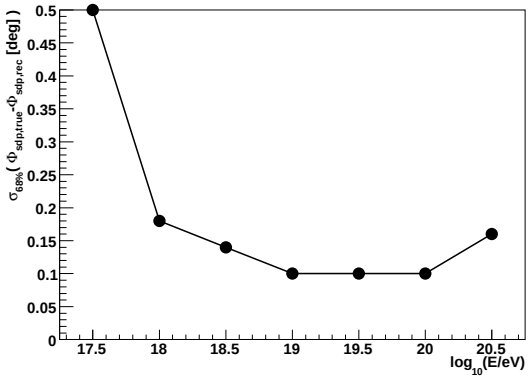
### 5.5.1 Reconstruction Performance of the SDP

The trace on the FD camera allows a very accurate reconstruction of the shower detector plane (see Chap. 4.1.3). That performance is represented in Fig. 5.6 for the orthonormal vector defining the SDP. Usually, this vector is expressed in spherical coordinates. Fig. 5.6 investigates the reconstruction of the $\phi$-component of this vector for some exemplary energies and distances. The trends of the mean and spread as function of the primary energy and distance are summarised in Fig. 5.7. Obviously, the reconstruction of the SDP is not biased by a constant shift. The mean is centred around zero, while the resolution is around 0.2°. Investigating the $\theta_{SDP}$ distributions a similar behaviour is found. There is no systematic bias and the resolution is between 0.3° and 0.45°. The

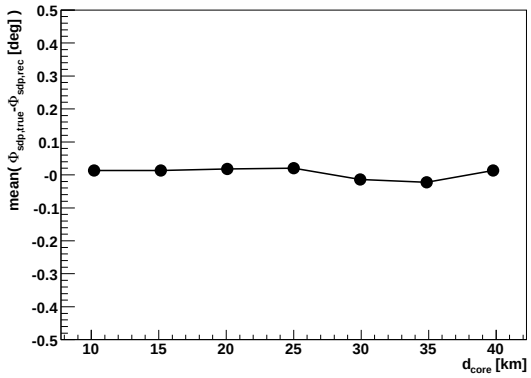
[7] The number was roughly constrained via a by-eye-check of warning-labelled events



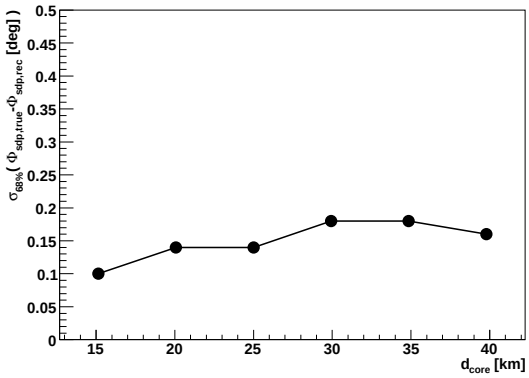

(a) Mean of $\phi_{\mathrm{SDP}}$ vs. energy.




(b) Resolution of $\phi_{\mathrm{SDP}}$ vs. core distance.




(c) Mean of $\phi_{\mathrm{SDP}}$ vs. energy.




(d) Resolution of $\phi_{\mathrm{SDP}}$ vs. core distance.

**Figure 5.7:** The performance of reconstructing $\phi_{\mathrm{SDP}}$. Mean and spread as function of energy, distance and zenith.

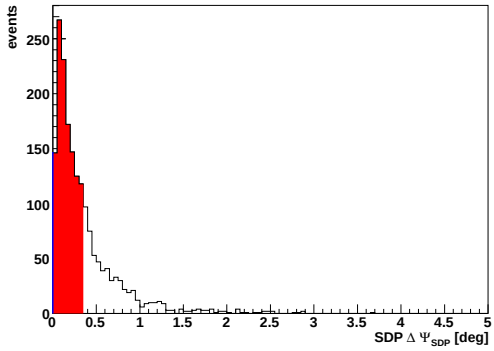
absence of a systematic bias motivates to summarise the performance of the SDP fit in terms of total angular difference between true and reconstructed orthonormal vector.

$$\Delta\Psi := \arccos(\vec{n}_{\mathrm{true}} \cdot \vec{n}_{\mathrm{rec}}) \ .$$

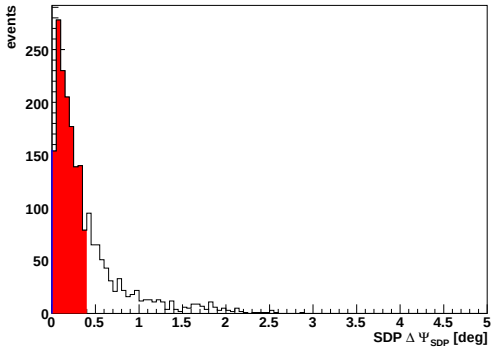
This distribution is defined to be always positive. As there is no bias in $\theta_{SDP}$ and $\phi_{SDP}$, the resolution can be estimated by calculating the positive 68 % confidence interval. This is shown in Fig. 5.8, leading to a final resolution of

$$\Delta\Psi_{SDP} \sim 0.15^{\circ} \text{ - } 0.5^{\circ}$$

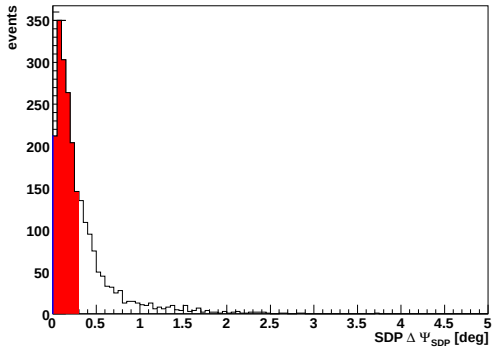
for reconstructing the SDP using FD-mono data.



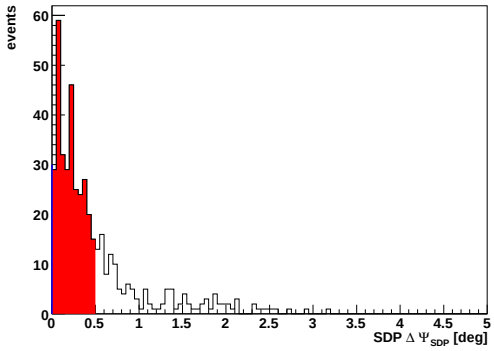

(a) $E = 10^{18}$ eV



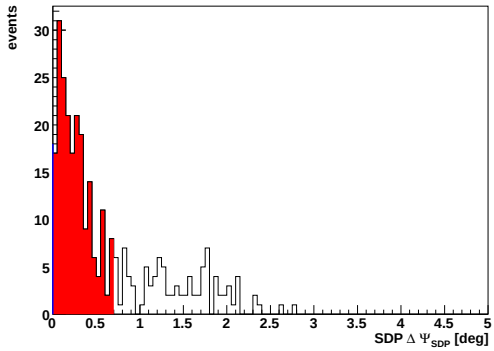

(b) d=15 km



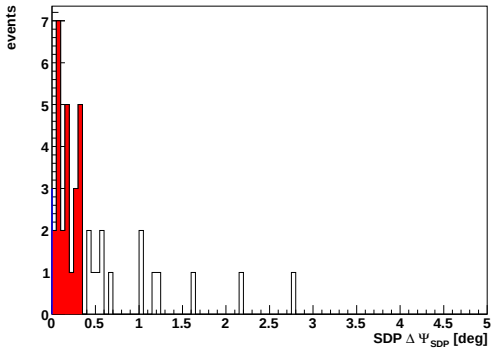

(c) $E = 10^{19}$ eV



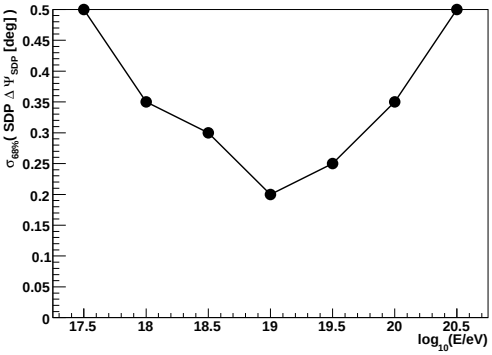

(d) d=25 km



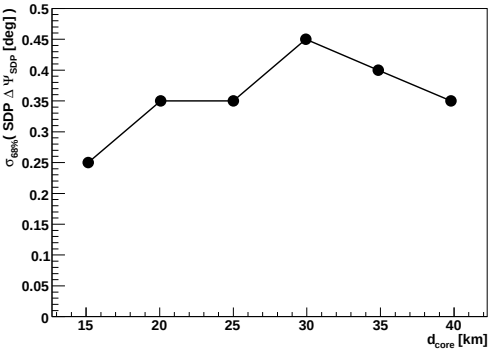

(e) $E = 10^{20}$ eV




(f) d=35 km

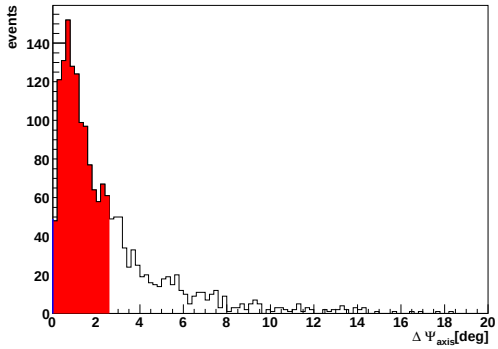



(g) SDP resolution $\sigma(\Psi_{\mathrm{SDP}})$ vs energy.

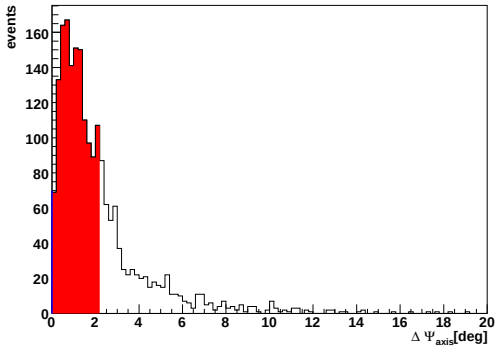



(h) SDP resolution $\sigma(\Psi_{\mathrm{SDP}})$ vs core distance.

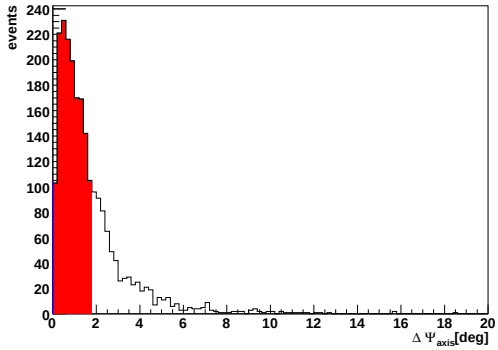
**Figure 5.8:** Angular resolution of the SDP.



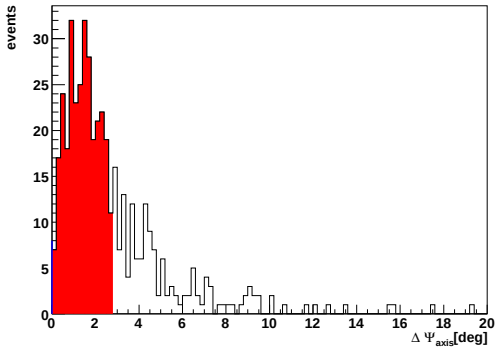

(a) $E = 10^{18}$ eV



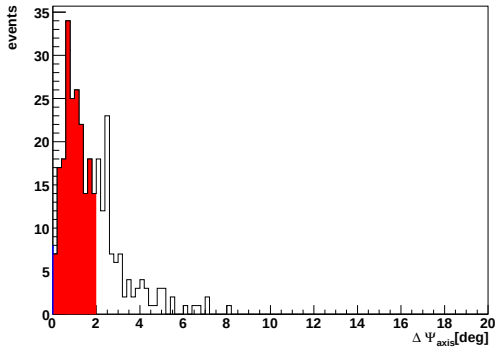

(b) d=15 km



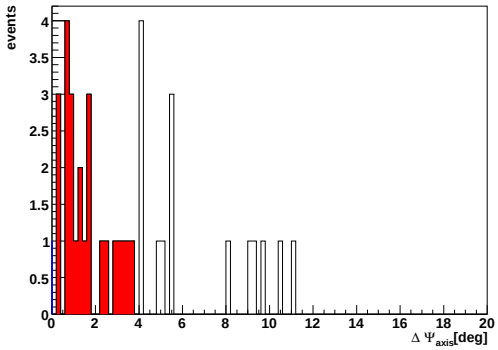

(c) $E = 10^{19}$ eV



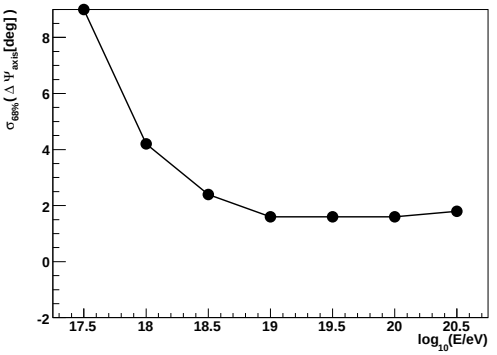

(d) d=25 km



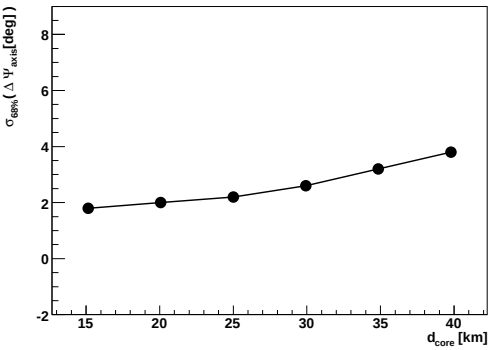

(e) $E = 10^{20}$ eV




(f) d=35 km

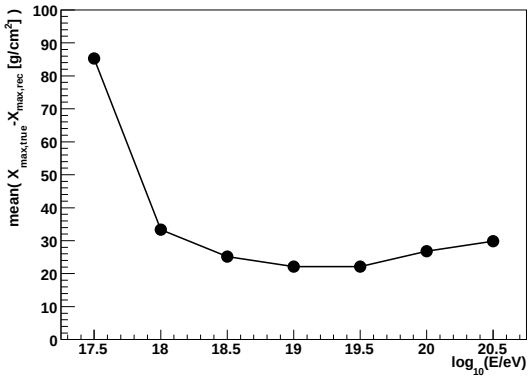



(g) Axis resolution $\sigma(\Psi_{\mathrm{axis}})$ vs. energy.

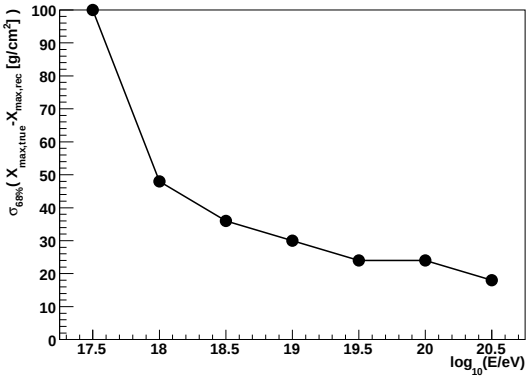



(h) Axis resolution $\sigma(\Psi_{\mathrm{axis}})$ vs. core distance.

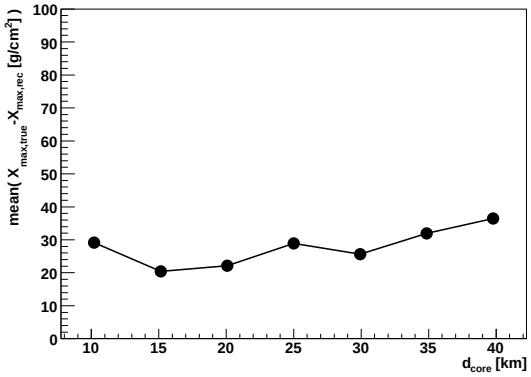
**Figure 5.9:** Angular resolution of FD-mono.



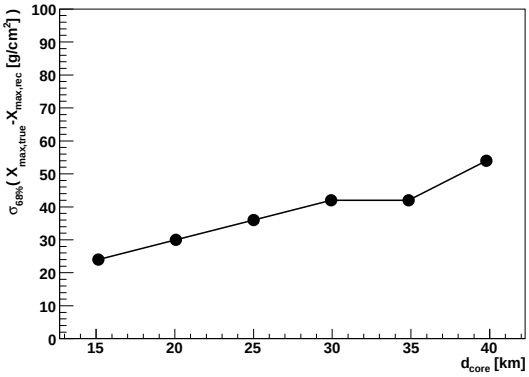

(a) Bias of $X_{max}$ vs. energy.




(b) Resolution of $X_{max}$ vs. energy.




(c) Bias of $X_{max}$ vs. core distance.

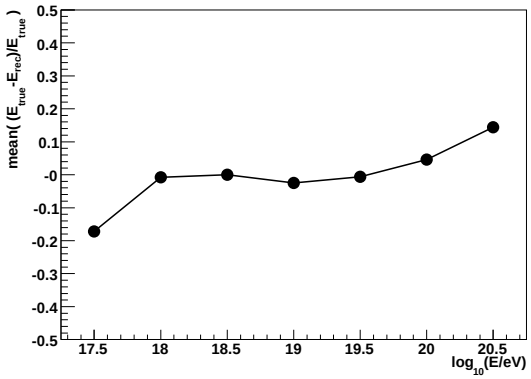



(d) Resolution of $X_{max}$ vs. core distance.

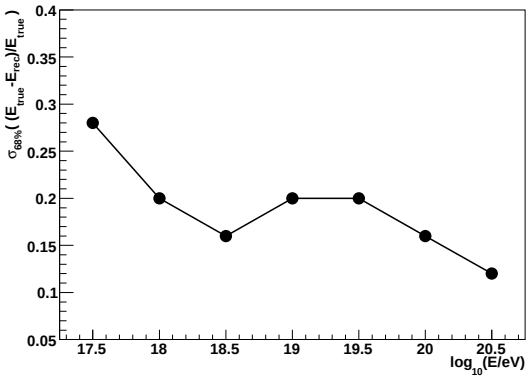
**Figure 5.10:** Performance of reconstructing $X_{max}$.

### 5.5.2 Geometry Reconstruction

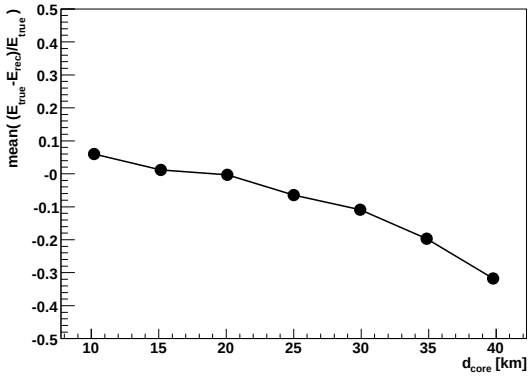
While the SDP can be reconstructed very accurately, the uncertainties of the shower geometry are dominated by the exact position of the axis within the SDP (see Chap. 4.1.3). The performance of reconstructing the $R_p$-parameter is summarised in App. A. In general, the time-fit tends to reconstruct the shower axis farther. The constant bias is around 500 m and 1000 m overestimation, while the resolution as function of the energy is quite constant around 1000 m. The resolution degrades with increasing distance, far away showers have uncertainties up to 2500 m. The performance to reconstruct the shower axis $\vec{a}$ is investigated in terms of zenith and azimuth angular resolution. The mean of $\theta$ and $\phi$ is negligible compared to the spread. Therefore, it makes sense to calculate similar to the SDP the total angular resolution. Fig. 5.9 shows that the axis can be reconstructed



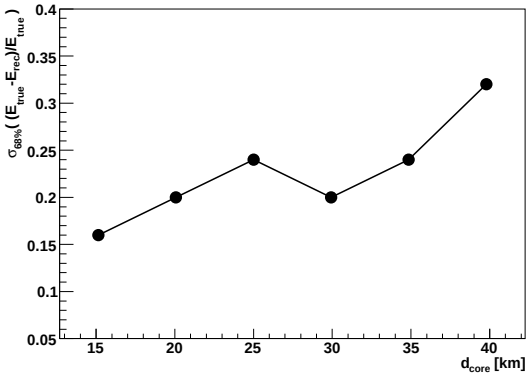

(a) Bias of the energy reconstruction vs. energy.




(b) Resolution of the energy reconstruction vs. energy.




(c) Bias of the energy reconstruction vs. core distance.




(d) Resolution of the energy reconstruction vs. core distance.

**Figure 5.11:** Performance of reconstructing energy.

within an accuracy of 2° to 4°, while the error is dominated by the resolution of the axis azimuth angle.

### 5.5.3 Profile Reconstructing

The reconstructed Gaisser-Hillas profile provides the depth of shower maximum $X_{max}$ and the final energy. Obviously, $X_{max}$ is overestimated by $30\,\mathrm{gcm}^{-2}$, while the resolution is in the same order. Proton and iron induced shower differ about $\Delta X_{max} \sim 70 - 100\,\mathrm{gcm}^{-2}$. Hence, FD-mono events are not an appropriate sample to study composition, hybrid data are more suitable for these kind

| Variable | Performance as function of... | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Energy | | distance | | zenith | | Overall | |
| | mean | $\sigma$ | mean | $\sigma$ | mean | $\sigma$ | mean | $\sigma$ |
| $\theta_{SDP}[^\circ]$ | 0-0.1 | 0.3-0.5 | -0.1-0 | 0.2-0.35 | -0.1-0.1 | 0.15-0.35 | $< \pm 0.1$ | $\sim 0.35$ |
| $\phi_{SDP}[^\circ]$ | $< \pm 0.05$ | 0.1-0.25 | $< \pm 0.05$ | 0.1-0.2 | $< \pm 0.05$ | 0.05-0.2 | $< \pm 0.05$ | $\sim 0.2$ |
| $\Delta\psi_{SDP}[^\circ]$ | - | 0.1-0.5 | - | 0.2-0.4 | - | 0.1-0.4 | - | $\sim 0.3$ |
| $R_p[km]$ | 0.5-1.3 | $\sim 1$ | 0.5-1.3 | 1-2.5 | 0.5-1.3 | 0.5-1.5 | 0.5-1.3 | $\sim 1$ |
| $\phi_{axis}[^\circ]$ | 0-0.2 | 2-4.5 | 0-0.5 | 2.5-3.5 | 0-0.5 | 2.5-3.5 | $< 0.25$ | $\sim 3.5$ |
| $\theta_{axis}[^\circ]$ | 0.2-0.4 | 0.5-1.5 | 0.2-0.4 | 0.7-1 | 0.2-0.4 | 1.7-1.3 | 0.2-0.4 | $\sim 1.3$ |
| $\Delta\psi_{axis}[^\circ]$ | - | 2-4 | - | 2-4 | - | 2 | - | $\sim 4$ |
| $X_{max}[g/cm^2]$ | $\sim 30$ | 20-50 | $\sim 30$ | 25-50 | 10-30 | 20-40 | 30 | $\sim 40$ |
| energy [%] | -5-10 | 10-20 | -20-5 | 15-30 | -30-20 | 5-20 | $\sim -5$ | $\sim 20$ |

**Table 5.1:** Performance of the FD-mono reconstruction based on the described sample of mono-energetic and fixed showers.

of investigations instead. The energy resolution is calculated on a linear scale:

$$\frac{E_{\text{true}} - E_{\text{rec}}}{E_{\text{true}}} \, .$$

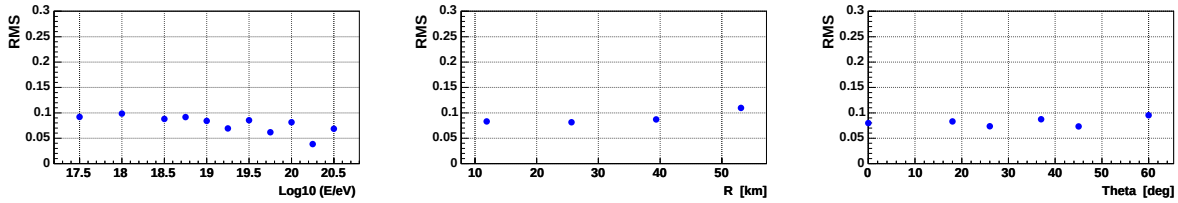
The energy reconstruction as a function of the primary energy indicates no significant bias, while the energy reconstruction is overestimated for farther showers. This issue has to be taken into account for the reconstruction of the energy spectrum. The energy resolution for FD-mono data is about 20 % to 25 % on a linear scale. This result emphasises the necessity to unfold the FD-mono data, as the bin-by-bin migration becomes relevant.

### 5.5.4 Summary

The study performed on this dedicated sample of mono-energetic and fixed showers is summarised in Tab. 5.1. While the geometry reconstruction is rather accurate, the final energy reconstruction suffers from the uncertainties in the time-fit.

## 5.6 Performance of the Hybrid-Like Reconstruction

In an independent study we investigate the improvement when performing a hybrid like reconstruction. The details have been published in [149]. Here, we provide a sketch of the analysis and summarise the results as hybrid data is not in the main focus of this work. Nevertheless, the results of this study demonstrates the general difference between FD-mono and hybrid data and, in addition, motivates the application of unfolding methods for FD-mono data (see Chap. 7).
Here, dedicated hybrid performance is investigated by emulating hybrid-like data in the following




(a) (b) (c)

**Figure 5.12:** Performance of the energy reconstruction for hybrid-like MC data, [165].

way: The simulation-reconstruction chain is applied including full atmospheric propagation, detector and trigger simulation. The properties at generator and reconstruction level are compared. In contrast, to the previous FD-mono reconstruction validation, in this study the energy resolution has been estimated for the case of known (*'true'*) geometry. Therefore, the SDP and time-fit has been skipped from the reconstruction sequence and the shower geometry is set to the true parameters from generator level. Then the profile is reconstructed. This procedure provides an estimate of the energy resolution in hybrid mode, as the hybrid reconstruction benefits from a precise shower core location due to the tank information (see Chap. 4.1.3). The final performance is summarised in Fig. 5.12. Here, the RMS of the relative energy reconstruction performance

$$\frac{\text{rec. observable -true observable}}{\text{true observable}}$$

is plotted as function of distance, primary energy and zenith. The RMS is interpreted as reconstruction uncertainty and is comparable to the 68 % confidence interval of the mono-study[8]. The resolution is analysed as function of the energy and distance of the EAS to the telescope. More details, especially on the high energy performance can be found in [165]. Obviously, the energy resolution in case of true geometry is smoothly below 10 % and independent of distance, axis geometry and energy. Compared to the 20 % energy resolution of FD-mono, the big improvement of hybrid-data becomes visible. Furthermore, we have shown in this study that the passing rate of events with fixed geometry up to cut-level 8 is nearly twice the one off mono events. This is clearly due to the fact, that the Gaisser-Hillas fit performs much more reliable for precise geometry. Nevertheless, FD-mono data is more suitable for the reconstruction of the energy spectrum down to $10^{17}$ eV, due to two reasons.

**Statistics:** One has to keep in mind that the hybrid-data exposure is still smaller than the FD-mono exposure due to a continuously instrumented SD array.

**Energy threshold:** Furthermore, the hybrid-data requires at least on hit tank, therefore aims a slightly different energy regime, starting from $10^{18}$ eV.

The lack of energy resolution for FD-mono data requires spectral unfolding.

[8] RMS and the 68 %-confidence interval differ in most cases on the order of 5 %

# Chapter 6

# Data Analysis at Detector Level

*It doesn't matter how beautiful your theory is, it doesn't matter how smart you are. If it doesn't agree with experiment, it's wrong.*

Richard Feynman

In Chap. 5 the performance of the FD-mono reconstruction is validated based on simulated data. It is shown that due the energy resolution it is necessary to unfold the reconstructed energy spectrum. The procedure, that is explained in Chap. 7 in detail, requires large statistics of MC data to calculate the response matrix. An essential condition to follow this approach is the agreement between simulation and experimental data. This can be validated by simulating and reconstructing a large set of MC events. These events are distributed according to an isotropic flux assumption. Their energy spectrum follows an assumed physical flux. Then, the basic variables can be compared at detector level. This procedure provides a more realistic validation than the validation based on MC only, which is described in Chap. 5. It folds in the validation of the complete simulation, too. Thus, it allows e.g., to check the accuracy of the trigger simulation, that is not possible with the MC-based validation. Instead of generating the physically flux event-by-event, we try to cover the phase-space in a more efficient way and re-weight the events using importance sampling.

## 6.1 Importance Sampling

Often, the quantities of physical interest and their distribution e.g., as function of the energy are not directly measurable. To compare data to MC simulated events one has in principle two possibilities depending on what is the final aim:

- Compare data and MC *at final trigger level*.
- Compare data and simulation *at generation level*.

For comparisons at *final trigger level* one has to generate the events according to an *a priori* spectrum. Usually, this is a challenging task, as one has to face an enormous parameter space. Using the true physical spectrum, it is very ineffective and not always wanted. The volume to be sampled is large, but is characterised by small probabilities to be triggered over most parts. In addition, one aims to achieve the same statistical significance on the whole phase-space. That means one should try to simulate the events, which are distributed in such a way that they have the same statistical significance after folding with the detector acceptance. Finally, one has to re-weight the events to obtain an assumed physical distribution.
Importance sampling is a variance-reducing techniques in Monte Carlo methods. A key issue in order to achieve small errors on the obtained result (for a given number of samplings) is a suitable strategy of sampling the available multidimensional parameter space. Importance sampling ensures that important regions are reached by redistributing the number of points sampled in various parts of the domain, but in a continuous manner. The goal of importance sampling is to re-weight the distribution of the random variable used in the Monte Carlo estimate, so that it is more likely to get a random sampling-point in regions, where the integrand is larger.

**Re-Weighting**

Given the assumed physics distribution of real events passing a detector in an observable $x$:

$$\frac{dN_{\text{true}}}{dx}(x) \ .$$

Therefore, the total amount of expected events passing the detector is:

$$N_{\text{true}} = \int_{x_{\min}}^{x_{\max}} dx \ \frac{dN_{\text{true}}}{dx}(x) \ .$$

The physics distribution might be very non-uniform[1] as the detector has usually not the same sensitivity on the whole parameter space.[2] Simulating events according to the physics spectrum could be very CPU extensive and might not lead to extra information as most events are not triggered. It is practical to follow another approach: Instead of simulating $N_{\text{true}}$ events according to $\frac{dN_{\text{true}}}{dx}(x)$ one might tend to find a more suitable distribution $\frac{dN_{\text{sim}}}{dx}(x)$ and re-weight events afterwards. Here, each simulated event can no longer be interpreted as integer real event but as part of a real event. The weight $w(x)$ can be calculated via:

$$\frac{dN_{\text{true}}}{dx}(x) = \underbrace{\frac{\frac{dN_{\text{true}}}{dx}(x)}{\frac{dN_{\text{sim}}}{dx}(x)}}_{:=w(x)} \frac{dN_{\text{sim}}}{dx}(x') = w(x)\frac{dN_{\text{sim}}}{dx}(x') \ .$$

[1] As it is e.g., in case of the steeply falling energy spectrum of cosmic rays.
[2] As it is the case for the fluorescence trigger around the threshold-energy.

This guarantees when histogramming the weighted events:

$$\sum_{N_{\rm sim}} w_i \rightarrow \int_{x_{\rm min}}^{x_{\rm max}} dx'\; w(x') \frac{dN_{\rm sim}}{dx}(x') = \int_{x_{\rm min}}^{x_{\rm max}} dx'\; \frac{dN_{\rm true}}{dx}(x') = N_{\rm true}\;. \tag{6.1}$$

The variance of an ensemble of $N$ *'integer'*-events can be calculated with the central-limit theorem to be $\sigma = \sqrt{N}$. Instead, using weighted events the bin content of an ensemble of weighted event is different from the number of events. The uncertainties of the bin-content turn out to be

$$\sigma = \sqrt{\sum_{i=1}^{N} w_i^2}\;.$$

In some cases, it makes sense to calculate the number of equivalent events $\tilde{n}$. This number represents the number of unweighted events which would have the same relative error as the weighted sum:

$$\tilde{n} = \left( \sum_{i=1}^{N} w_i \right)^2 / \sum_{i=1}^{N} w_i^2\;.$$

The concept of equivalent events is e.g., very useful for the likelihood analysis of low event rates.

### 6.1.1 Example: A Power-Law Spectrum

The merit of importance sampling is demonstrated by investigating a steeply descending power-law energy spectrum with spectral index $\alpha = -3$. In arbitrary units we investigate the energy regime between 10 and 10000. Hence, the corresponding probability density function (pdf) is

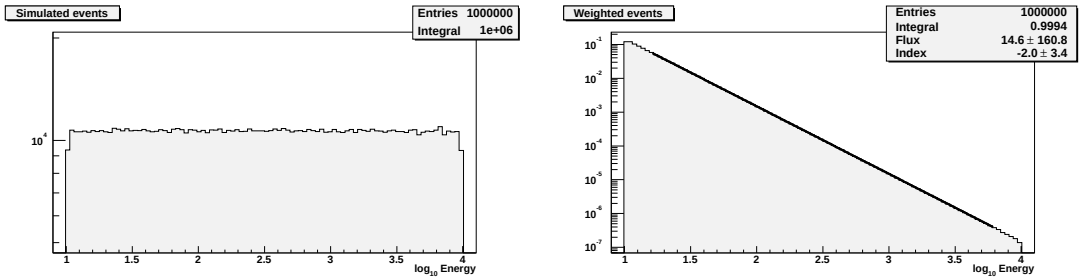
$$\Phi_0 = \left[ \int_{10}^{10^4} dE\; E^{-3} \right]^{-1}\;.$$

Obviously, the low energy regime dominates the sample. E.g., simulating the this distribution means to simulate one high energy event one has to generate 100 events in the energy regime one order of magnitude below. This is a serious point as usually the region of interest covers orders of magnitude in the energy. Thus, a toy-MC was launched simulating $N_{\rm sim} = 10^6$ events logarithmically flat in energy and re-weight the events.

$$\frac{dN_{\rm sim}}{dE} = \Phi_{\rm sim} E^{-1} = N_{\rm sim} \left[ \int_{10}^{10^4} dE\; E^{-1} \right]^{-1} E^{-1}\;.$$

The distribution of simulated events is shown in Fig. 6.1(a). A physical distribution can be obtained by weighting the events via

$$w(E) = \frac{dN_{\rm true}}{dE} / \frac{dN_{\rm sim}}{dE} = \frac{\Phi_0}{N_{\rm sim}} \left[ \int_{10}^{10^4} dE\; E^{-1} \right] E^{-2}\;.$$




(a) Distribution of the $10^6$ simulated events

(b) Distribution of the $10^6$ weighted events, a power law is fitted

**Figure 6.1:** Illustration how to derive from a logarithmical flat simulation to a power-law with the differential spectral index 3

The weighted distribution is shown in Fig. 6.1(b). This example underlines two major advantages of importance sampling:

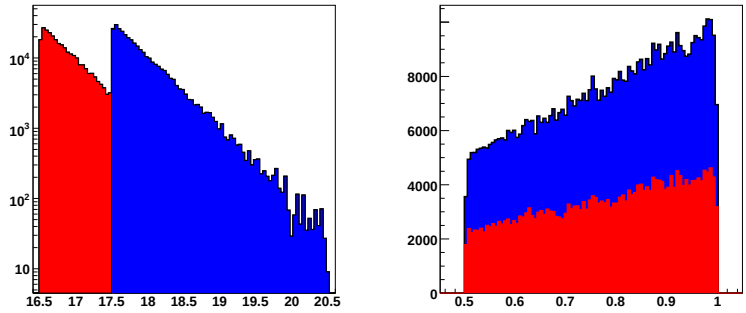
1. By simulating a logarithmical flat spectrum it is possible to have uniform statistics. E.g., without weighting, it would have been necessary to simulate about $10^4$ times more events below 100 toy-units to obtain one above 1000 toy-units.

2. Once simulated it is possible to re-weight the MC and use the same MC to obtain any other spectrum in a straight forward way.

## 6.2 The Simulation Sample

### 6.2.1 The Energy Spectrum

Generally, it makes sense to distinguish between mono-energetic and continuous generated energy spectra. Mono-energetic MC samples have the advantage of testing the performance of the reconstruction in detail under well defined constrains (see Chap. 5). A continuous energy distributed sample is more suitable to study a realistic detector response. In Chap. 7, continuous energy spectrum are used to estimate the detector response matrix.
Here, we focus the data-MC comparison at detector level. Thus, it is necessary to simulate continuous spectrum and re-weight it to follow a physical flux. The energy regime of interest is between $10^{17.5}$ eV and $10^{20.5}$ eV. In very special cases, e.g., near a fluorescence telescope, showers at the threshold-energy, below $10^{17.5}$ eV, can still trigger the detector and can be miss-reconstructed to higher energies. Therefore, it is necessary to extend the energy range and simulate also events below the threshold-energy. Moreover, the events should be simulated according to a power-law spectrum. Two effects have to be considered, when choosing the spectral index $\alpha$ of the simulation distribution:




**(a)** Energy distribution **(b)** $\cos\theta$ distribution

**Figure 6.2:** Energy and Zenith distribution

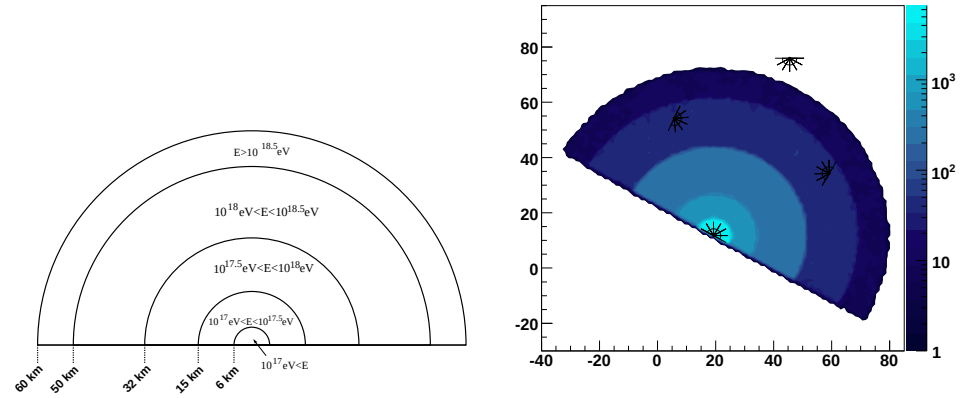
**$\alpha < 3$:** In the energy range of interest all experiment indicate a steeply falling energy spectrum with a spectral index of about $\alpha = 3.3$ up to the ankle and about $\alpha = 3.$ beyond. Neglecting for the moment the question of a possible cutoff at highest energies, the generation of showers according to the true physics distribution appears to be hopeless. One would waste CPU, when trying cover the high energy part with enough statistics by simulating to many low energy events that would never be triggered at all. Quantitatively, one needs to simulate about $10^6$ events in the energy range between $10^{17} - 10^{17.5}$ eV to generate one event in the energy regime between $10^{20} - 10^{20.5}$ eV.

**$\alpha > 1$:** A logarithmical flat spectrum might appear suitable as it covers the whole energy range homogeneously. On the other hand, a logarithmical uniform spectrum at generator level, would not provide enough statistics of triggered events in the low energy range. Due to the steeply falling spectrum most events are found around the threshold energy of about $10^{17.5}$ eV, this should be also represented by the MC statistics.

The most efficient spectral index is hard to find, basically it depends on the width of the fluctuations relatively to the steepness of the spectrum. This is not a homogeneous function over the whole parameter range. Therefore, the most efficient way is most probably not following a strict power-law behaviour but a rather more complicated function. Technically, the gain in performance to parametrise this function is compensated by the amount of complexity.
A more pragmatic way is the following:

- The energy spectrum of the simulated MC event is according to an $E^{-2}$ power-law. This seems to be a quite reasonable compromise between the extreme cases discussed above.

- The event generation is done in two sub-samples: A high energy sample at $10^{17.5} - 10^{20.5}$ eV and low energy sample at $10^{16.5} - 10^{17.5}$ eV. This considers that at highest energies the detector




**(a)** Schematic idea of generating core positions

**(b)** Generated core positions

**Figure 6.3:** Realisation of an energy dependent core position area.

acceptance is fairly good and the statistics of triggered events is limited due to a steeply falling power law spectrum. Whereas, at low energies the event statistics is limited due to the detector acceptance. The attempt to cover the low energy part with significantly high statistics using one continuous spectrum only, would mean to loose statistics in the physical relevant part of the spectrum or to generate an impossible amount of showers. Therefore, we chose the high energy sample to cover the physical region with enough statistics and the low energy sample the low-acceptance region.

Fig. 6.2(a) shows the final MC statistics as function of the energy.

### 6.2.2 The Core Position

The core positions are expected to be uniformly distributed on the surface. Obviously, the more bright a shower appears in the camera the more likely it is triggered. Thus, it does not make sense to simulate the core positions of low energy showers far away from the detector, as they are not triggered. The idea was to simulate the events in a half-circle around the telescope building of Los Leones in the field of view of telescope 1-6. The radius of the circle depends on the primary energy and is chosen between 6 km for low energy events up to 60 km for the highest energy ones (see Fig. 6.3). The areas can be summarised:

$$A_{ground} = A_{proj} = \begin{cases} \frac{\pi}{2}(6\,\mathrm{km})2 & E < 10^{17}\,\mathrm{eV} \\ \frac{\pi}{2}(15\,\mathrm{km})^2 & 10^{17}\,\mathrm{eV} \leq E < 10^{17.5}\,\mathrm{eV} \\ \frac{\pi}{2}(32\,\mathrm{km})^2 & 10^{17.5}\,\mathrm{eV} \leq E < 10^{18}\,\mathrm{eV} \\ \frac{\pi}{2}(50\,\mathrm{km})^2 & 10^{18}\,\mathrm{eV} \leq E < 10^{18.5}\,\mathrm{eV} \\ \frac{\pi}{2}(60\,\mathrm{km})^2 & 10^{18.5}\,\mathrm{eV} \leq E \end{cases}$$

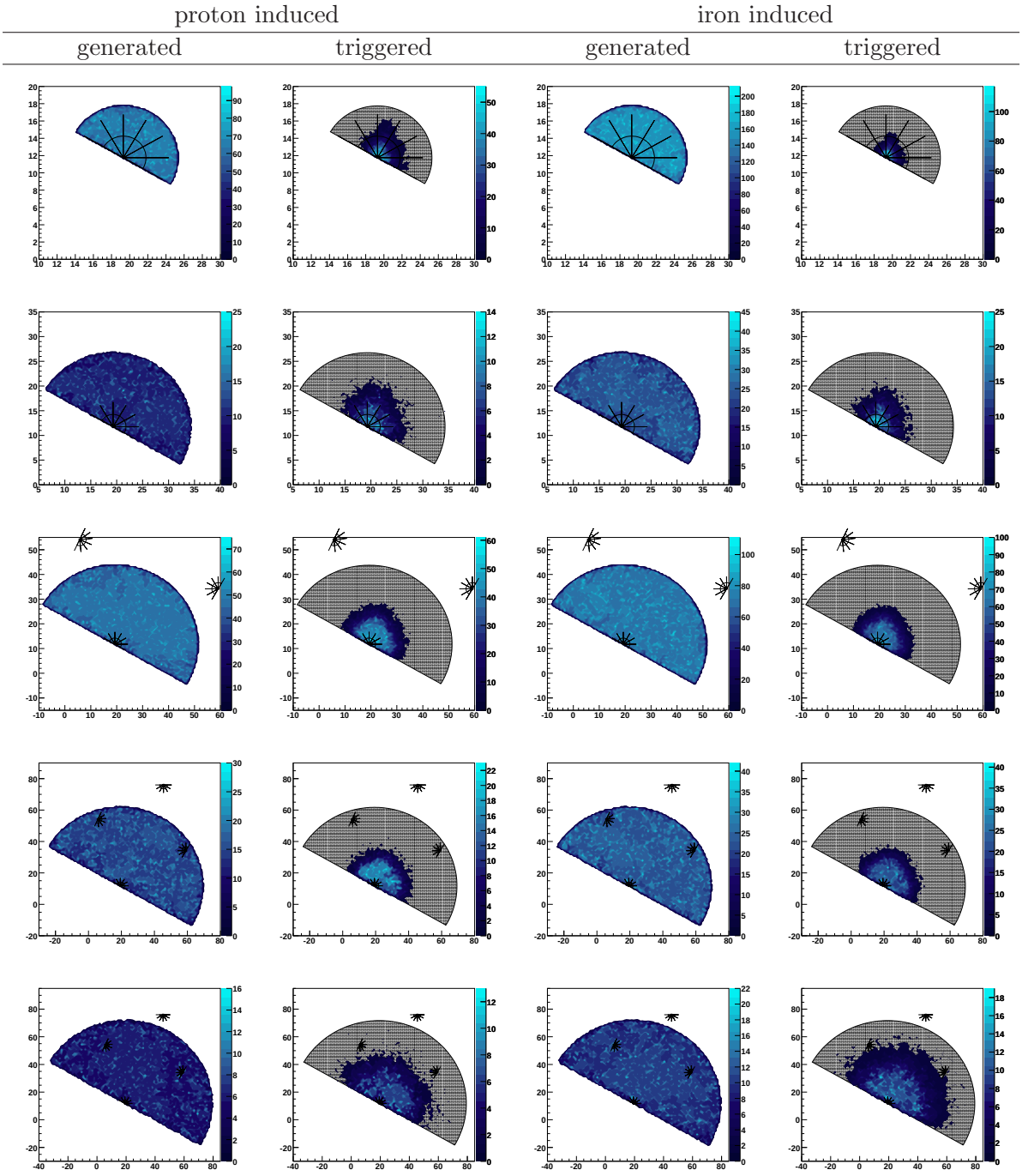
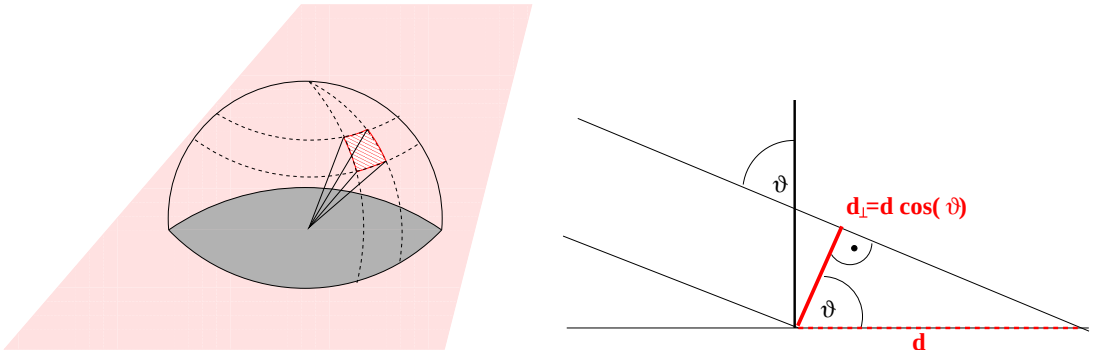



**Table 6.1:** Generated and triggered core positions. Investigating iron and proton induced showers in the five different energy regimes of the core generator. From top to bottom: $E < 10^{17}$ eV, $10^{17}$ eV $\leq E < 10^{17.5}$ eV, $10^{17.5}$ eV $\leq E < 10^{18}$ eV, $10^{18}$ eV $\leq E < 10^{18.5}$ eV,$10^{18}$ eV $\leq E$.




(a) Space angle effect

(b) Relation between projected and orthogonal area

**Figure 6.4:** Two effects have to be considered when simulating an isotropic flux: the space angle and the projected area.

Tab. 6.1 demonstrates, that this methods guarantees uniform distributed core positions on ground. Moreover, the positions of triggered events are plotted to indicate that the choice of radii is conservative.

### 6.2.3 The Shower Direction

The shower axis is simulated according to an isotropic flux, which implies that two effects have to be taken into account:

- The flux from each steradian space angle $d\Omega$ should be constant:

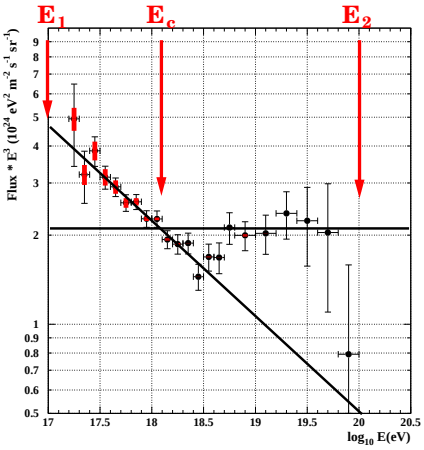
$$\frac{dN}{d\Omega} = \frac{dN}{d\phi \sin\theta d\theta} = \frac{dN}{d\phi d\cos\theta} = 1 \; .$$

  This represents that the events are distributed homogeneously in azimuth and that the space angle for vertical events is smaller than for inclined ones. (see Fig. 6.4(a))

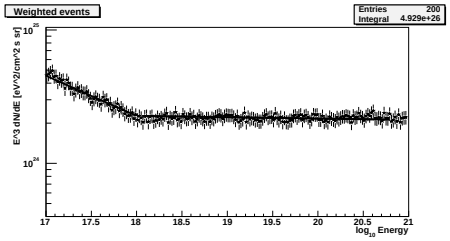
- The flux assumes that the number of particles through any orthogonal area is constant, but the core positions are placed on ground. Hence, the generation area is the projected area. This area is larger than the orthogonal area by a factor $1/\cos\theta$ (see Fig. 6.4(b)).

### 6.2.4 Physics Expectation and Re-Weighting

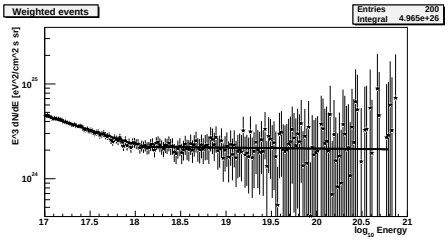
In order to compare data and MC at trigger level all events are promoted through the full simulation-reconstruction chain. Afterwards, the observable are re-weighted according to a physics flux assumption (see Fig. 6.5). The aim of this study is to compare MC-reconstructed observable distributions (*MC-prediction*) to the ones of experimental data. Thereby, the full FD response and trigger performance can be validated based on experimental data. This allows to validate intrinsic



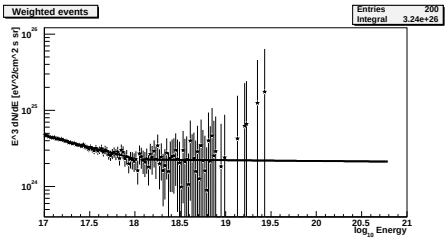

(a) Parametrisation of the assumed flux




(b) Re-weighted $E^{-1}$ spectrum




(c) Re-weighted $E^{-2}$ spectrum




(d) Re-weighted $E^{-3}$ spectrum

**Figure 6.5:** Influence of the generated spectral shape on the final statistics after re-weighting using a toy MC

effects which cannot be found by a MC based validation (see Chap. 5).[3]. The bulk of triggered and reconstructed events stem from energies, where the energy spectrum of cosmic primaries is known reasonably well. Therefore, the shape of the distributions that should be compared is dominated by events below a few $10^{18}$ eV. The existence of a possible cutoff can only change slightly the final

[3] like the accuracy of the trigger simulation

shape. Therefore, we assume for the data-MC comparison a broken power-law fit to experimental data, describing the ankle feature, but neglecting any cutoff-feature.

Following this approach, the measured spectrum by HiRes experiment [37] can be used to fit a spectral behaviour:

**'Flat assumption':** A very rough approach would be to simulate an featureless $E^{-3}$-power-law spectrum. The normalisation would be around

$$\frac{dN}{dE} \sim 2.2 \cdot 10^{24} E^{-3} \frac{\mathrm{eV}^2}{\mathrm{cm}^2 \mathrm{ssr}} \, .$$

**'Ankle-hypothesis':** The spectrum in the part below the ankle is a little bit steeper. This can be respected by using a broken power-law:

$$\frac{dN}{dE} = \begin{cases} \Phi_1 E^{-3+x} & E < E_c \\ \Phi_2 E^{-3} & E < E_c \end{cases} \tag{6.2}$$

## 6.3 The Data Sample and Filtering

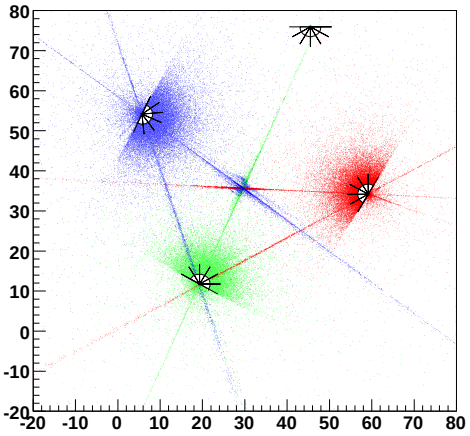
We want to compare the MC prediction with experimental data from the Los Leones telescope station. The analysed data have been taken in the period between Jan. 2004 and Oct. 2006[4]. In contrast to MC, experimental data additionally contain artificial events, e.g., induced by CLF or LIDAR laser shots. Usually the data acquisition of the telescope site of the firing LIDAR is vetoed for the time of the LIDAR scan. Nevertheless, there is still a non-negligible amount of events that are clearly LIDARs. These are supposed to be shot by a LIDAR station into the field of view of another station. These artificial events become visible when plotting the core position of all events (see Fig. 6.6(a)). As described in Chap 4.1.3, the drawback of fluorescence reconstruction is the exact location of the shower axis within the SDP. CLF and LIDAR events appear as up-going events in the time-fit, while the time-fit assumes a down-going event. Besides this effect the LIDAR events appear as very faint noisy events. In total this leads to the effect that LIDAR or CLF induced events do not manifest accumulated around the exact position of their device, but more as accumulation of core positions along the field of sight in which the device is visible in the corresponding telescope. This structure is clearly visible in Fig. 6.6(a). The CLF is located in the crossing point of the three lines in the centre of the array.

For the analysed data-taking period the Los Leones LIDAR has been equipped with a 10 Hz laser. LIDAR induced events from Los Leones are clearly visible in Los Morados and Coihueco. The Coihueco LIDAR was equipped with a high repetition laser, that is mostly visible in Los Leones and only poor in Los Morados[5].

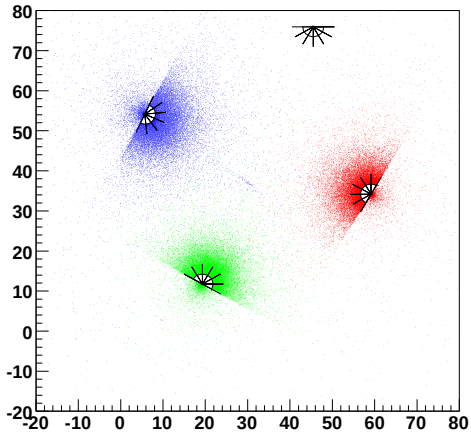
These artificial events have been removed from the data-sample by filtering out certain GPS

[4] In GPS time stamps this corresponds to the time period between 758254649 and 843468808

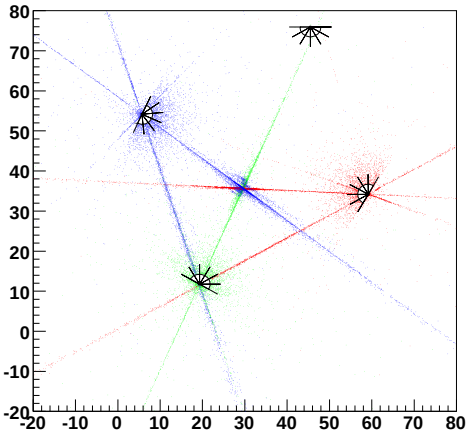
[5] due to the large distance of about 60 km most light is attenuated



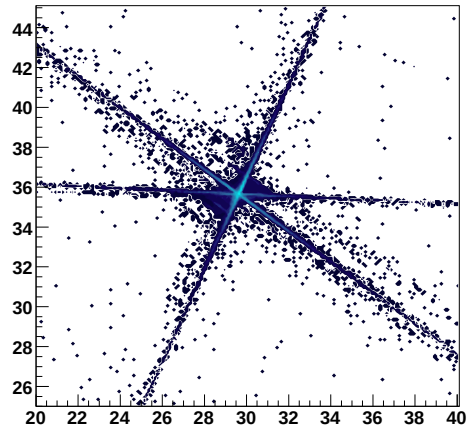

(a) Reconstructed core positions of all events with geometry No cut is applied. The structure visible is due to mis-reconstructed LIDAR and CLF events.




(b) Core position after rejecting time intervals coincident with LIDAR and CLF operation.

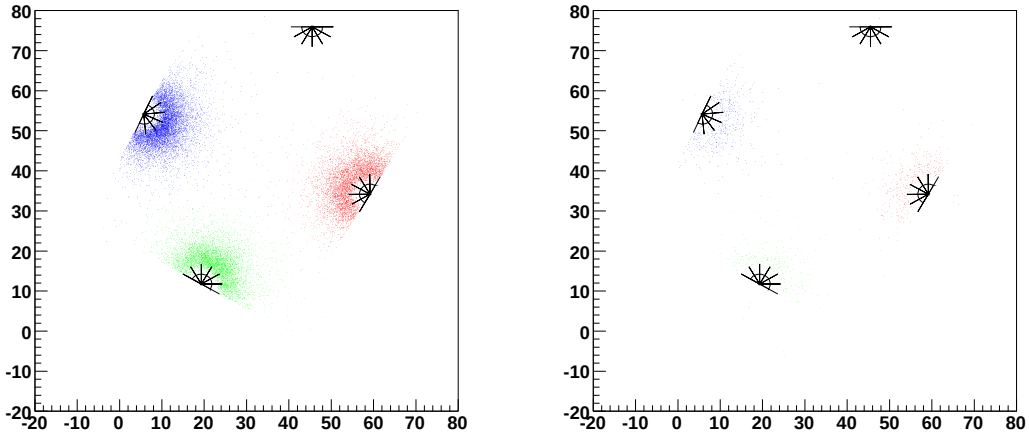



(c) Core position of rejected time intervals shows the power of LIDAR and CLF filter.




(d) Zoomed position of rejected time intervals. The performance of the filter algorithm is capable to resolve the CLF position with a few meters.

**Figure 6.6:** Before applying the quality and performance cuts it is necessary to filter out time periods that coincident with LIDAR or CLF operation. The performance of this filter becomes visible at the level of SLT events with geometry. The colour code indicates the telescope that has triggered the event: green=Los Leones, red=Los Morados, blue=Coihueco. Loma Amarilla has not been operational a that period.




(a) Reconstructed core positions at final cut level after CLF and LIDAR filter.

(b) Core positions of events at final cut level that have been rejected due to CLF and LIDAR filter. No structure is visible.

**Figure 6.7:** The data at final cut level. The core positions of events that have been rejected by CLF and LIDAR appear homogeneous. An indication that LIDAR or CLF induced events are mostly rejected by the standard quality cuts.

nanosecond time-windows[6]. The dedicated filter algorithm has been developed by N. Nierstenhöfer and is described in detail in [138]. It is optimised to minimise the dead-time correction due to leaving off certain nanoseconds. Therefore, hot phases have been identified in the data, in which a LIDAR has been in operation. The advantage of identifying artificial events by their GPS ns-structure is that no additional cuts on both MC and data have to be applied except a lifetime correction of the MC weight. The filter is physically motivated and guarantees complete rejection of CLF and LIDAR events as they just trigger in these certain time windows. On the other hand, the final lifetime correction of

$$\Delta t_{laser} = 0.3\,\% t_{up}$$

is smaller than the systematic uncertainties of the uptime itself (see Chap. 6.4). The filter performance is demonstrated in Fig. 6.6(b), 6.6(c) and 6.6(d). The filtered core positions show a homogeneous distribution on ground (Fig. 6.6(b)), the pattern of the artificial event positions disappeared, while plotting the core position of rejected events the characteristics of the pattern is enhanced (see Fig. 6.6(c)). Moreover, the laser core positions allows to determine the position of the CLF within a few meters. This is demonstrated in Fig. 6.6(d). It underlines the accuracy of

[6] E.g. the CLF is always firing on GPS ns 250000000 and 500000000

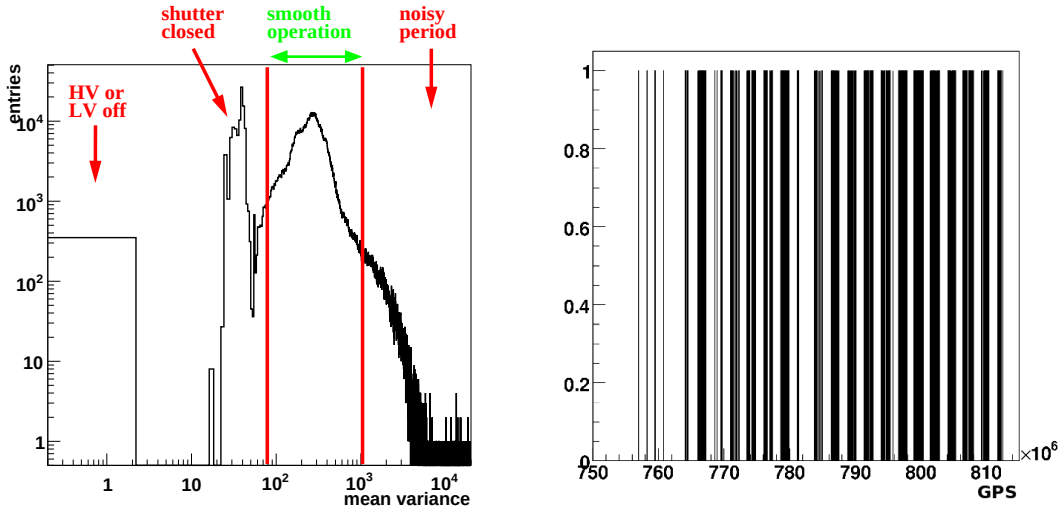
| | events at raw SLT level | | | final cut level | | |
|---|---|---|---|---|---|---|
| | total | rejected | accepted | total | rejected | accepted |
| Los Leones | 249525 | 210243 | 39282 | 5873 | 252 | 5621 |
| Los Morados | 123937 | 93000 | 30937 | 5261 | 348 | 4913 |
| Loma Amarilla | - | - | - | - | - | - |
| Coihueco | 277664 | 222960 | 54704 | 8590 | 477 | 8113 |

**Table 6.2:** Performance of the laser filter applied to data. Details of the implementation can be found in [138].

the SDP reconstruction and the performance of the filter algorithm. At final cute level, the core positions of the filtered data set are shown in Fig. 6.7(a). No structure is visible, as most laser events are already rejected by the standard cuts. Fig. 6.7(b) shows the core positions of rejected events that would survive all cuts and would appear in the final data set. Their distribution does not show any pattern as it is visible at raw SLT level, but is widely distributed homogeneously on ground, as expected for real events. The effect of the laser filter can be summarised in Tab. 6.2.

## 6.4 Uptime

The definition of the detector uptime is rather complex. Here, the uptime of a telescope is defined by the state, when the high voltage (HV) for the PMTs and low voltage (LV) for the electronics is ramped up, data acquisition (DAQ) is operating in a smooth mode, as well as the failsafe-curtains are up and the shutter is open. For the data period analysed, the monitoring of each single component was still in progress. Direct monitoring of the uptime within the slow control system (SCS) is not yet fully possible. Thus, an indirect way to estimate the uptime is applied: for this issue the night-sky background data is investigated. Every 30 seconds the background noise of each telescope is read out (see Chap. 3.3.4). In case, that the HV is off, there is no noise visible. If the HV is ramped up and electronics are working the noise-level due to the night sky with open shutters differs significantly from the situation with closed shutters. Thus, time bins with background-data indicate HV and LV being ramped up. Fig. 6.8(a) shows the distribution of variances collected from all eyes and telescopes. This allows to define four running modi. A zero mean variance is indicates that the background DAQ has collected data, but for that period either HV or LV have been off. A second peak around 50 indicates that in this period HV and LV had been ramped up but the very low variance indicates missing night-sky background noise. Hence, the shutter have been closed or the failsafe curtains must have been down. Variances higher that 1000 are rather atypical periods. Such high noise can only be induced by lightenings or artificial light like air-planes. The data for these few periods are excluded for the final analysis as we can not guarantee reliable results. The region of interest is for variances between 80 and 1000, indicating smooth operation with ramped up LV and HV, as well as open shutter and failsafe curtains.




(a) Distribution of the mean variance and classification of smooth operation

(b) Uptime bins for telescope 4 in Los Leones

**Figure 6.8:** Procedure to estimate the uptime.

In addition, the DAQ is monitored directly by the SLT data. Time bins with triggered events indicate that the DAQ has been working smoothly.
Corresponding time bins with correct background variances and SLT data are considered as period that fulfils the uptime criteria (*uptime-bin*). The total uptimes can be obtained by telescope-wise integration of all uptime-bins.
A time binning 10 minutes intervals shows best performance. E.g., the uptime bins for telescope 4 in Los Leones are shown in Fig. 6.8(b). Filled bins indicate smooth operation while zero bin content is defined as 'detector off'. The periodic off-phases are due to monthly periods when the moon is too bright to allow operation of the fluorescence telescopes. moreover, zooming into this periods shows the day-night periodicity. The total uptimes are summarised in Tab. 6.3. The statistical uncertainties derived by this approached can be estimated to be less than 10 %. The data are taken over a total time period lasting $8.5 \cdot 10^7$ s. Since the SD has been fully operational for this period, the duty-cycles of the telescopes in Los Leones range between 7.5 % and 9.5 % with respect to the surface array. The corresponding duty-cycles for the Coihueco ones are found to be between 8.5 % and 10.3 %. Los Morados has smaller uptimes due to the fact that the telescopes at this site have come into operation in Nov. 2004. Normalised, to the corresponding time since Nov. 2004 the Los Morados telescopes have duty-cycles between 7.5 % and 8.3 % with respect to the surface array.

| | Uptime in $10^6$ s | | | | | |
|---|---|---|---|---|---|---|
| Station | tel 1 | tel 2 | tel 3 | tel 4 | tel 5 | tel 6 |
| Los Leones | 6.32 | 6.44 | 7.42 | 8.16 | 6.78 | 6.11 |
| Los Morados | 4.64 | 4.48 | 4.86 | 4.56 | 4.42 | 4.36 |
| Loma Amarilla | - | - | - | - | - | - |
| Coihueco | 7.62 | 8.82 | 8.62 | 7.28 | 7.40 | 7.21 |

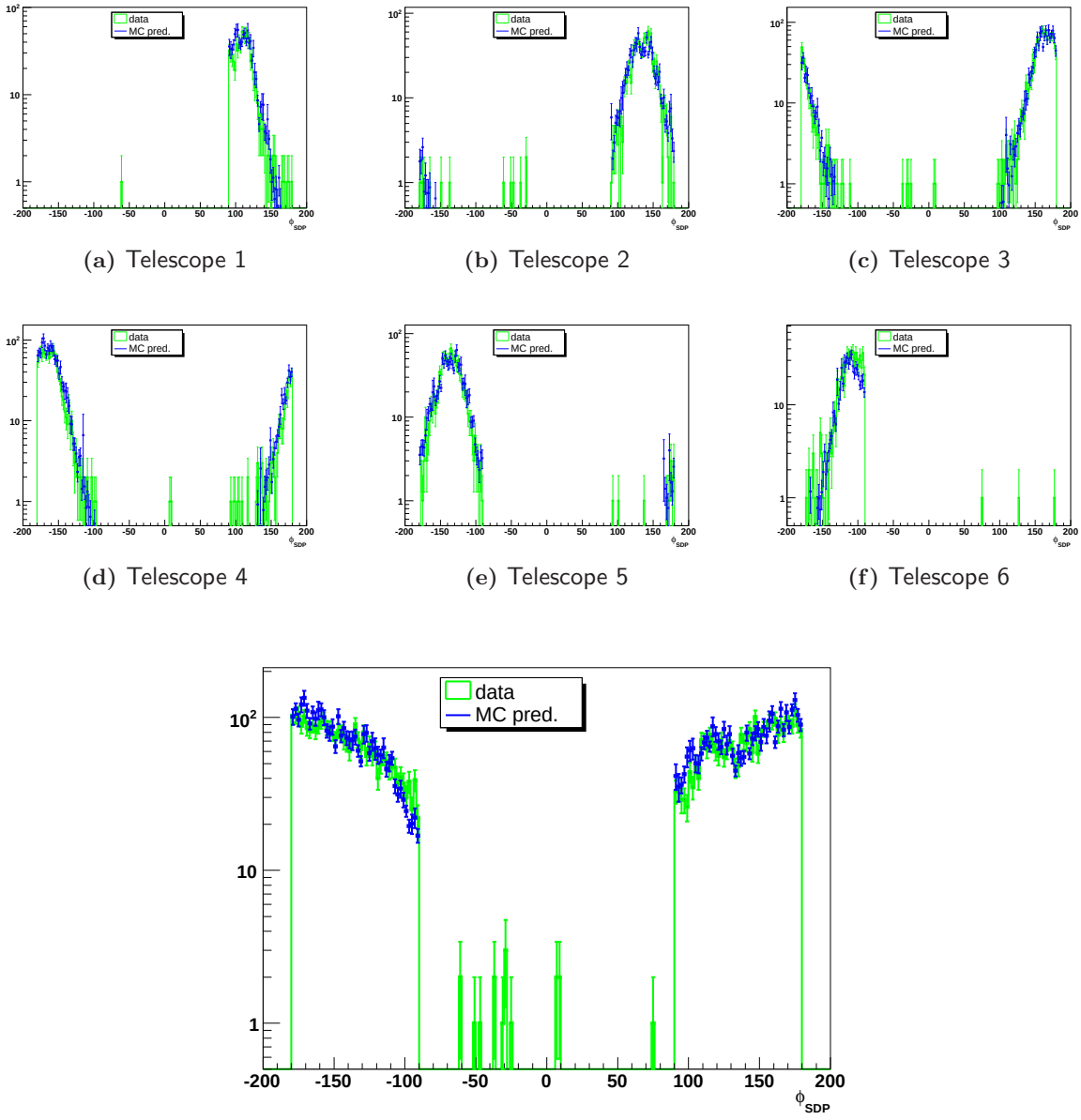
**Table 6.3:** Summary on estimated uptimes for the telescope stations Los Leones, Coihueco, Los Morados

## 6.5 Data-MC Comparison and Data Quality

The MC sample is weighted, regarding the individual lifetimes of each telescope to obtain the correct normalisation. Thereby, the MC predictions for the distribution of different observables can be calculated, which allow to validate the simulation reconstruction chain on experimental data. This is exemplarily demonstrated in Fig. 6.9. For each telescope the MC prediction for the $\Phi$ component of the SDP normal vector is plotted. This example, represent that each telescope has a different field of view and thus the reconstructed SDPs should follow in individual telescope-wise characteristic. The assignment of the observables to a certain telescope is based which telescopes have triggered pixels. Hence, the same event can appear in more than one MC telescope-prediction. This treatment is valid as we proceed in the same way for data and MC and the aim is to validate the shape and normalisation agreement of the individual observables.[7] The agreement is investigated for each telescope individually and for observables at different levels of reconstruction, starting with the 'number of pixel' distributions up to highly sophisticated variables like energy and $X_{max}$. Here, a representative subset of cumulative plots is presented to keep the overview transparent. In these plots the individual distributions of all six telescopes have been merged. This is exemplarily performed in Fig. 6.9(g) and demonstrates that if the individual distribution disagree it is rather unlikely that the cumulative distributions agree. Fig. 6.10 summarises the result of this data-MC validation for Los Leones data.

Obviously, experimental data and MC-prediction agree at the final cut level up to highly sophisticated variables like energy and $X_{max}$. The difference in normalisation is less than 7 % for the individual telescope-wise comparisons. The cumulative data and MC-predictions indicate a 5 % difference, that is still in agreement within the uncertainties due to the lifetime estimation (see Chap. 6.4). Additionally, the assumptions about the energy spectrum are rather idealised. The 'ankle hypothesis' is used (see Chap. 6.2.4), a simple broken power law with no further features. Slight modifications can affect the normalisation of the MC-prediction. Nevertheless, the accuracy of the detector simulation is demonstrated.

[7] This is to avoid confusion, when looking at the individual $\phi_{\mathrm{SDP}}$-distributions. A $\phi_{\mathrm{SDP}}$-entry can be assigned to a telescope, which has a field of view that does not cover $\phi_{\mathrm{SDP}}$. E.g., telescope 1 has a field of view between 0° and 30°. Thus, most $\phi_{\mathrm{SDP}}$-values range between 90° and 120°, while entries outside have been triggered additionally in the neighbouring telescope.




(a) Telescope 1 (b) Telescope 2 (c) Telescope 3

(d) Telescope 4 (e) Telescope 5 (f) Telescope 6

(g) Cumulative for Los Leones

**Figure 6.9:** Data-MC comparison of the orthonormal vector for the SDP. Here, is given the MC prediction of the $\phi$-angle and the observed data distribution

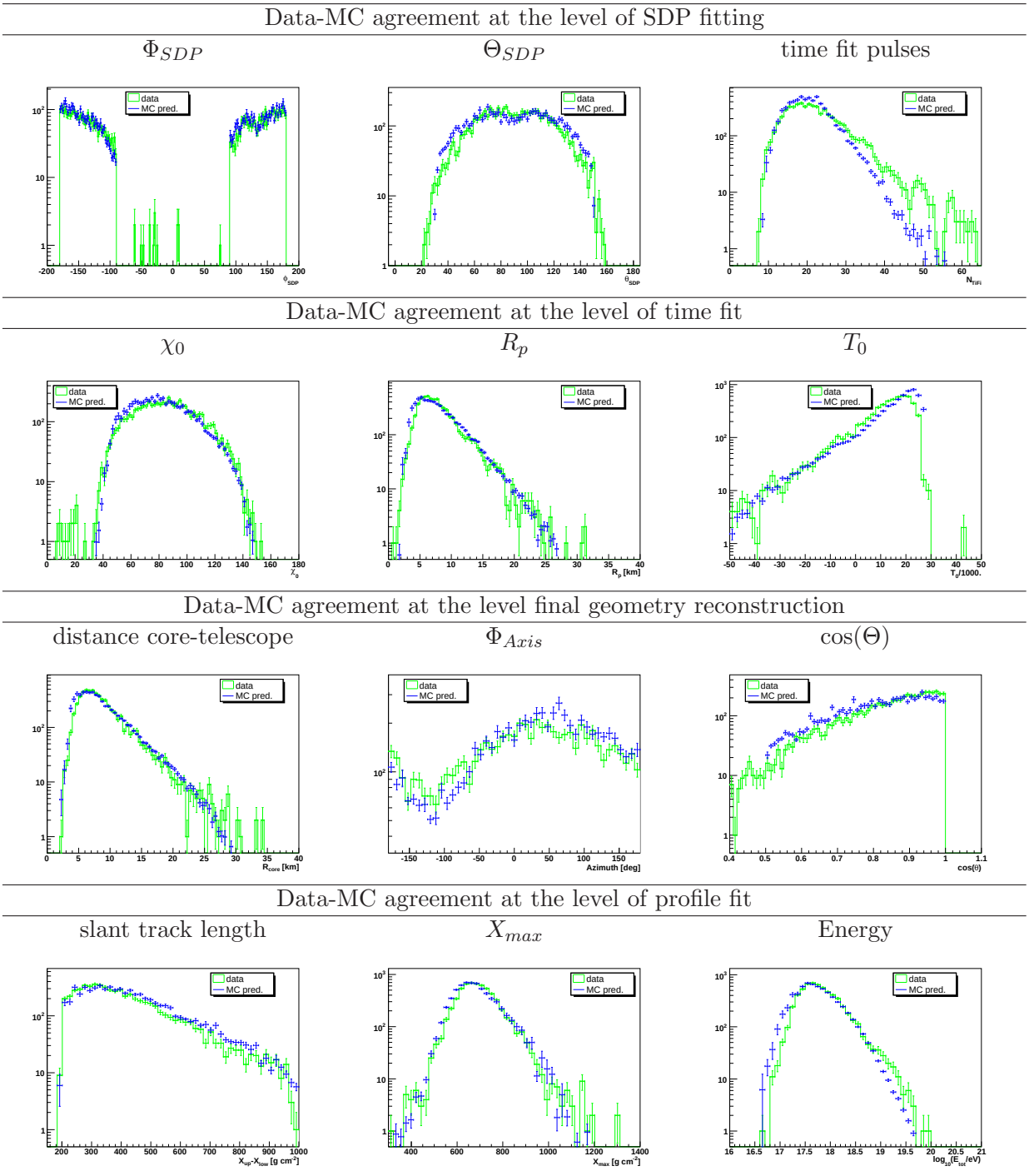



**Figure 6.10:** Data-MC comparison for different observables

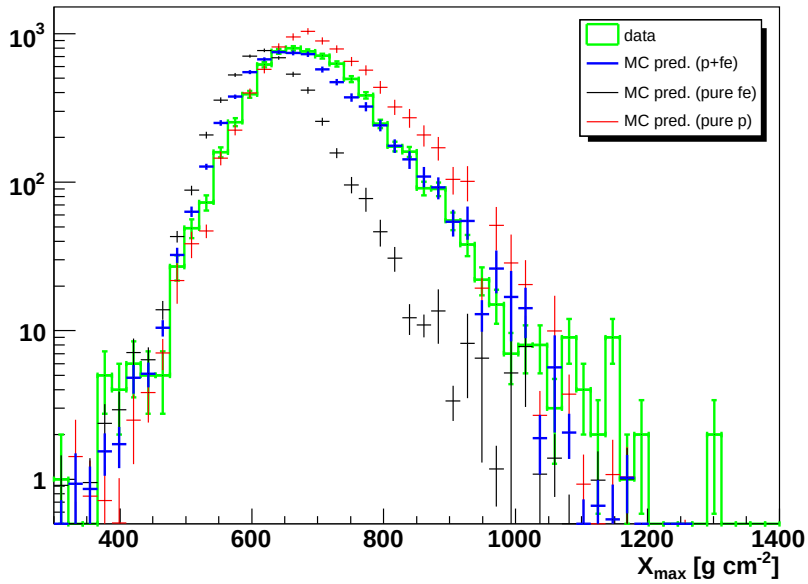
## 6.6 Effect of Input Spectrum and Composition

Besides validation of the detector simulation, the MC-prediction can be used to demonstrate the impact of primary composition and assumed energy spectrum on the final analysis. The effect of the primary composition becomes visible when investigating the $X_{max}$-distribution. Fig. 6.11(a) illustrates the MC prediction using just proton induced or just iron induced showers. The blue bar indicates the 50 % mixture that is used for this analysis. Obviously, the data look more proton like than iron, although at this level it is surely not possible to quantify this appearance. Nevertheless, the 50 % mixture is a valid assumption. For a known energy spectrum this plot could be a starting point for an unfolding based composition analysis, investigating the distributions at several energies and tune the composition to fit the data best. A more sophisticated approach would unfold in both events per energy bin (spectral unfolding) and proton-iron ratio as function of the energy (compositional unfolding).
Another effect becomes visible in the MC-prediction of the energy distribution using different spectral assumptions (see Fig. 6.11(b)). The MC prediction is calculated for the 'ankle hypothesis' and for the 'flat assumption' . Both input spectra are rather simple assumptions (see Chap. 6.2.4). Nevertheless, the 'ankle hypothesis' seems to describe reality much better than just the flat $E^{-3}$ spectrum. At this level it is not yet possible to quantify this effect, but it demonstrates the sensitivity of the MC simulation and gives an outlook to the unfolding procedures described in Chap.7, that allow to find the input spectrum that describes the data best.

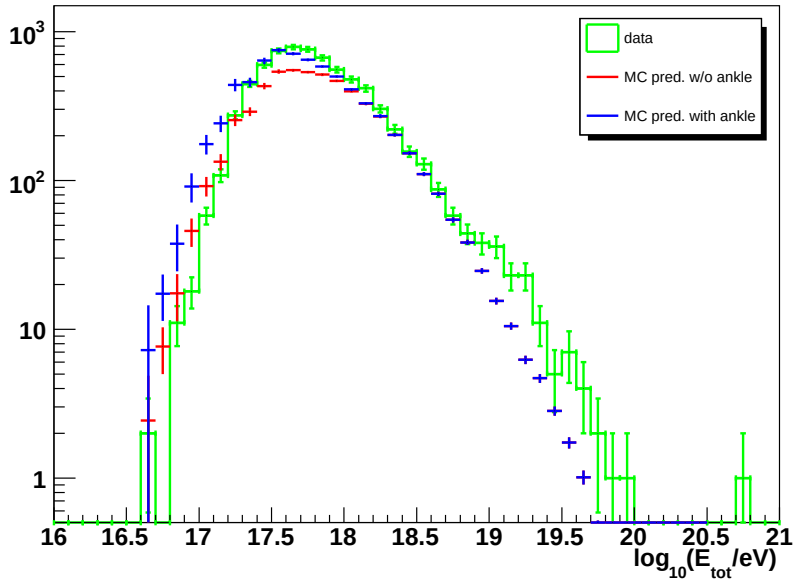
## 6.7 Extension to the Complete Data Set

The data-MC comparison presented in the previous parts is performed for events that have been triggered by Los Leones telescopes (see Chap. 6.2.2). The simulation to reach the needed level of significance has been very CPU-intensive, and even on the ALiCENext cluster, it takes several month to reach the presented statistics. Thus, there is no corresponding MC available for Los Morados and Coihueco data. Nevertheless, we can use the MC for Los Leones to emulate MC for the missing telescopes. Here, it makes just sense to compare observables in the eye coordinate system, like $R_p$ or observables that are independent from the frame, like the energy. The problem arising with this approach is, that the telescope stations are not configured in the same way. Each telescope station is equipped with a characteristic setting of telescopes with and without corrector ring. The time periods and configurations are given amongst others in the summary table Tab. 6.4. The specific effects that result from corrector rings can be studied now by both experimental data and MC. In case of MC, the input energy-spectrum is the only model dependency, and it is checked that it has no major influence on the result. Therefore, we compare the event rate using the *'ankle-hypothesis'* and the *'flat assumption'*. The following considerations have been studied and are taken into account for the final emulation.

**The effect of a corrector ring on the telescope itself:** This effect can be investigated by comparing the number of events per uptime for a telescope with corrector ring and the same num-




**(a)** $X_{max}$ prediction using a pure proton or a pure iron model and the used 50 % mixture compared to real data




**(b)** Reconstructed $Energy$ prediction using a monotone $E^{-3}$ spectrum and a kink power-law spectrum compared to data

**Figure 6.11:** MC prediction depending on input spectrum and composition

ber for a non-neighbouring telescope without corrector ring. Uptimes of data and MC have to be evaluated for the specific time period. The ratio between telescope 4 and telescope 2 in Los Leones has been investigated for the time period before 2004-08-01:

$$r_{\mathrm{cr}} := \frac{N_{\mathrm{tel4}}/t_{\mathrm{up,tel4}}}{N_{\mathrm{tel2}}/t_{\mathrm{up,tel2}}} \; .$$

At that time only telescope 4 was equipped with a corrector lens. The telescope-wise uptimes have been estimated and are given amongst others in the summary table, Tab. 6.4. Data indicate, that $r_{\mathrm{cr}}^{\mathrm{data}} = 2.02$, while MC shows that $r_{cr}^{MC,flat} = 1.80$ for the *flat assumption* and $r_{\mathrm{cr}}^{\mathrm{MC,ank}} = 1.90$ for the ankle hypothesis. For the following analysis, $r_{\mathrm{cr}} = 1.9 \pm 0.1$ is used, when correcting for non-simulated corrector lens or vice versa. This result is also in agreement with the studies presented in Chap. 5.3.

**The effect of a corrector ring on the neighbouring telescopes:** The existence of a telescope with corrector lens leads to an enlargement of the trigger rate in the neighbouring telescope due to multi-mirror events. This effect can be investigated by comparing telescope 3 and telescope 4 in Los Leones for the mentioned period by defining

$$r_{\mathrm{nb}} := \frac{N_{\mathrm{tel4}}/t_{\mathrm{up,tel4}}}{N_{\mathrm{tel3}}/t_{\mathrm{up,tel3}}} \; .$$

Data indicate $r_{\mathrm{nb}}^{\mathrm{data}} = 1.10$, while MC shows $r_{\mathrm{nb}}^{\mathrm{MC,flat}} = 1.13$, resp. $r_{\mathrm{nb}}^{\mathrm{MC,ank}} = 1.15$. We use a final setting of $r_{\mathrm{nb}} = 1.13 \pm 0.02$.

**The effect of Malargue on telescope 5 and telescope 6 in Los Leones:** The city of Malargue is about 15 km away from Los Leones. The civilisation light has influence on the trigger rate. The night sky noise is significantly higher in telescope 5 and telescope 6. The PMT thresholds of these telescopes are higher compared to other telescopes due to the dynamical adjustment by the FLT. Thus, the trigger rate at final quality cut level is slightly different between telescopes facing Malargue and those facing pure Pampa. Of course for Los Leones simulation the corresponding background maps have been used and therefore the presented MC sample describes the data in a correct manner. Nevertheless, when emulating Los Morados or Coihueco MC this effect has to be subtracted. Therefore, we investigate:

$$\begin{aligned} r_{\mathrm{tel6}} &:= \frac{N_{\mathrm{tel1}}/t_{\mathrm{up,tel1}}}{N_{\mathrm{tel6}}/t_{\mathrm{up,tel6}}} \\ r_{\mathrm{tel5}} &:= \frac{N_{\mathrm{tel1}}/t_{\mathrm{up,tel1}}}{N_{\mathrm{tel5}}/t_{\mathrm{up,tel5}}} \; . \end{aligned}$$

Data indicate $r_{\mathrm{tel6}}^{\mathrm{data}} = 1.5$ and $r_{\mathrm{tel5}}^{\mathrm{data}} = 1.3$, while MC indicates $r_{\mathrm{tel6}}^{\mathrm{MC,flat}} = 1.49$ and $r_{\mathrm{tel6}}^{\mathrm{MC,ank}} = 1.59$, as well as $r_{\mathrm{tel5}}^{\mathrm{MC,flat}} = 1.15$ and $r_{\mathrm{tel5}}^{\mathrm{MC,ank}} = 1.15$. For the analysis we use $r_{\mathrm{tel6}}^{\mathrm{MC,flat}} = 1.55 \pm 0.05$ and $r_{\mathrm{tel5}}^{\mathrm{MC,flat}} = 1.2 \pm 0.05$.

**Effect of neighbouring telescope dropping off DAQ:** The influence of dropping out of DAQ for a telescope itself is treated in a correct manner by investigating the uptimes telescope-wise (see Tab. 6.3). The effect that a neighbouring telescope is not in DAQ on a telescope in DAQ can be neglected. Indeed, the effect that for a telescope in DAQ the neighbouring telescope drops out of DAQ happens in less than 12 % of the telescope-wise uptime. On the one hand, the influence on the neighbouring telescope can be estimated from the effect of a corrector ring telescope neighbouring a telescope without corrector ring, that is about 13 %. On the other, one can investigate a telescope at the border of the station (telescope 1 or telescope 6) and a telescope in the middle of the station (e.g., telescope 2 or telescope 5). Here, one can focus on configurations of bordering and non-bordering telescopes without corrector lens and compare the trigger-rates. Data indicate a 20 % depletion, while MC shows a 10 % (flat) and 30 % (ankle) depletion. As this effect only appears in less than 12 % of the uptime the total influence is less than 2 %. This does not justify the immense technically complexity to include also the telescope-neighbouring status into the analysis.

Unfortunately, the configurations have been modified for the data-period analysed. The experiment is still under construction and therefore the corrector lenses are installed step by step. This asks for investigating the uptime not only telescope-wise, but additionally for all possible configurations. The result is shown in Tab. 6.4. The emulation of MC-prediction for Coihueco and Los Morados is technically obtained by correcting the uptime by the corresponding correction factors. E.g., for the period when a corrector lens has been installed but in MC no corrector ring was assumed, the uptime in the weight is scaled by $r_{cr}$. Illustratively, this means that the number of predicted events are emulated by assuming a higher uptime instead of higher trigger rate.
The final results are given in Fig. 6.12 and shows still good agreement. The normalisations disagree at a 15 % level, that can be understood in terms of uncertainties of the uptime estimation and correction factors, as well as assumptions on the input spectrum.

## 6.8 Apertures of Different Event Topologies of the Pierre Auger Observatory FD

Besides FD-mono it is possible that a shower is seen and can be reconstructed by two or more telescope stations (see Chap. 4.1.1). The possibility to see a shower by more than one telescope is of course dependent on the brightness (energy) of the shower and the geometry. Another independent crosscheck of the simulation and reconstruction performance can be done by investigating how good the MC can reproduce the expected trigger rates for various topologies. Hence, a MC study was launched distributing events homogeneously over an area of 80 km × 80 km that enclose the surface array. The events have been generated according to an $E^{-2}$ power-law and have been re-weighted according to the *ankle hypothesis* (see Chap. 6.2.4). The influence of the primary composition on the aperture calculation is cross-check using the previous MC sample. The trigger rate as function of the energy is shown in Fig. 6.13. The trigger behaviour is studied for proton and iron induced showers on raw SLT level and after final quality cuts. Obviously, the influence of the primary

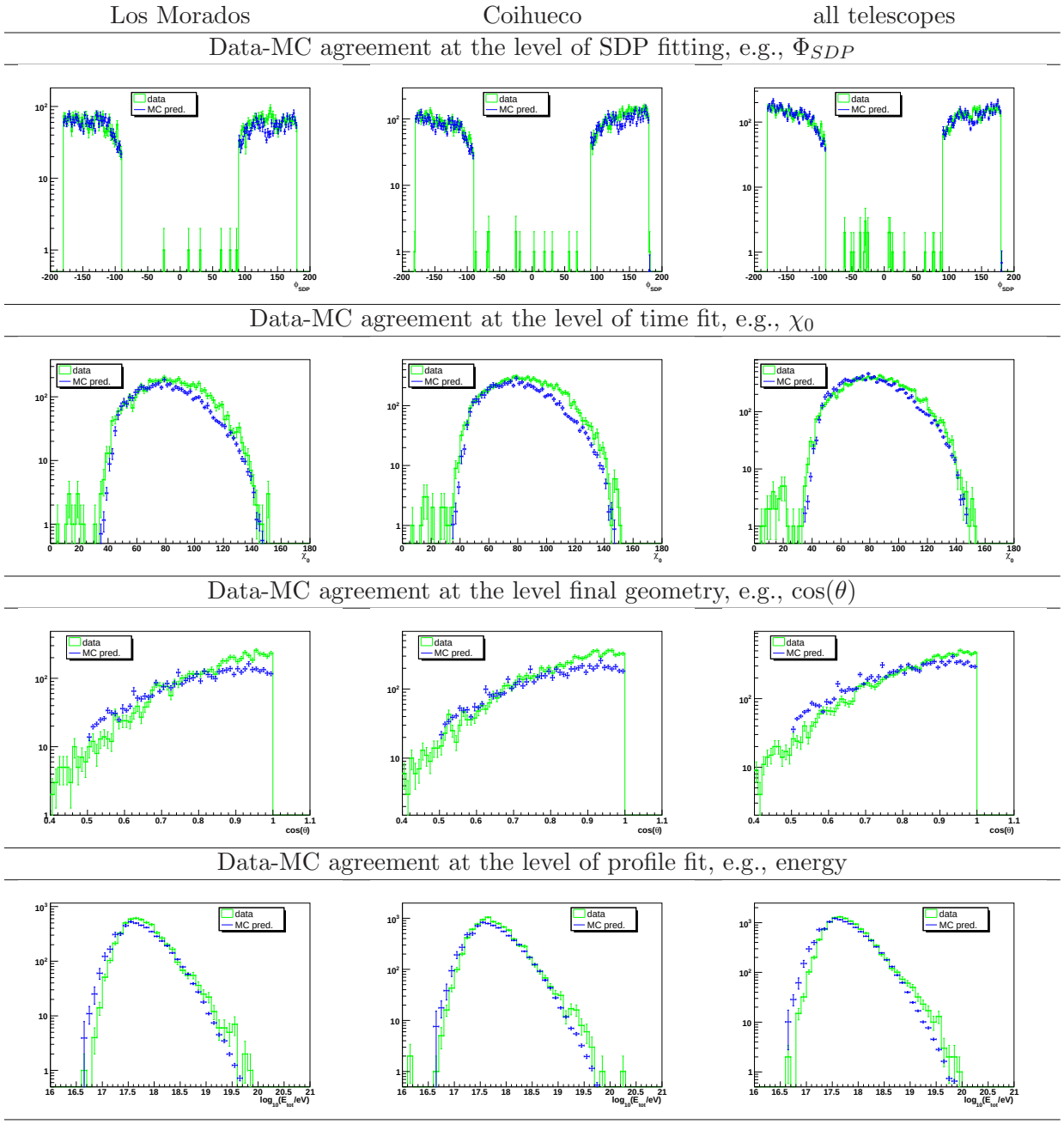



**Figure 6.12:** Data-MC comparison for different observables, emulating the whole array.

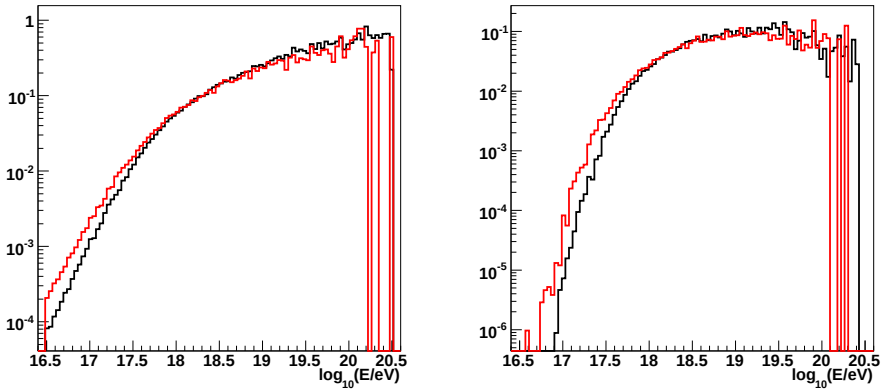
| Eye | Configuration | | Uptime in $10^6$s | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | tel 1 | tel 2 | tel 3 | tel 4 | tel 5 | tel 6 |
| Los Leones | 111211 | 2001-05-01 | 0.480 | 0.416 | 0.721 | 0.763 | 0.569 | 0.670 |
| | 112211 | 2004-08-01 | 1.009 | 1.150 | 1.344 | 1.442 | 1.232 | 1.048 |
| | 112221 | 2005-03-01 | 1.239 | 1.286 | 1.637 | 1.752 | 1.578 | 1.283 |
| | 122221 | 2005-08-26 | 0.001 | 0.015 | 0.001 | 0.001 | 0.001 | 0.000 |
| | 222222 | 2005-08-29 | 3.594 | 3.569 | 3.713 | 4.207 | 3.402 | 3.112 |
| Los Morados | 111111 | 2004-11-01 | 0.283 | 0.260 | 0.332 | 0.279 | 0.283 | 0.283 |
| | 111211 | 2005-05-31 | 0.010 | 0.012 | 0.016 | 0.014 | 0.010 | 0.008 |
| | 111221 | 2005-06-01 | 0.426 | 0.354 | 0.540 | 0.355 | 0.403 | 0.364 |
| | 112222 | 2005-08-27 | 0.470 | 0.492 | 0.670 | 0.507 | 0.406 | 0.399 |
| | 212222 | 2005-11-10 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 222222 | 2005-11-16 | 3.450 | 3.362 | 3.304 | 3.405 | 3.317 | 3.310 |
| Coihueco | 112111 | 2002-04-01 | 0.275 | 1.009 | 0.973 | 0.290 | 0.277 | 0.203 |
| | 112211 | 2004-08-01 | 1.222 | 1.407 | 1.312 | 1.230 | 1.274 | 1.234 |
| | 122211 | 2005-03-01 | 1.723 | 1.878 | 1.739 | 1.399 | 1.584 | 1.596 |
| | 222211 | 2005-09-01 | 0.940 | 1.032 | 1.026 | 0.959 | 0.840 | 0.908 |
| | 222222 | 2006-01-16 | 3.459 | 3.491 | 3.571 | 3.404 | 3.429 | 3.270 |

**Table 6.4:** Telescope-wise calculated uptimes for all possible detector configurations. A telescope with corrector lens is indicated by a 2, while telescopes without corrector ring are labelled by 1, e.g. 111211 means just telescope 4 has corrector lens.

composition can be neglected and manifests only in the very low energy regime. The interesting energy regime for multi-fold events starts at $10^{18}$ eV. This justifies to investigate only iron induced showers for this study instead of a mixed composition. The slight drop in the trigger performance above $10^{20}$ eV is due to the effect, that the pixelization of the PMT cameras becomes visible for very bright events , but does not affect this analysis (see Chap.5.4).
This full-array MC sample allows to investigate the trigger rate for various topologies. Fig. 6.14 shows the efficiency to trigger a 1-,2-,3- or 4-fold event at different core positions. The efficiency is studied for increasing energies.

This crosscheck shows that the trigger simulation provides a realistic picture of the detector. Stereoscopic events start to be triggered between the telescope stations, while the FD-mono events start triggering radial-symmetric around the telescope station (see also Fig. 6.1). Furthermore, this study allows to estimate the energy threshold to observe certain topologies. Above an energy threshold of $10^{18}$ eV it should be possible to observe events in at least two detector stations. 3-fold events should appear at energies above $10^{18.5}$ eV and the 4-fold topology is expected to be produced by events above $10^{19}$ eV.[8] The re-weighted MC events can be used to estimate the individual apertures and expected event rates for possible configurations. This is summarised in Fig. 6.15 and Fig. 6.16.

[8]The first 4-fold event detected in May 2007 has been reconstructed with an energy of approximately few $10^{19}$ eV.




(a) Trigger efficiency proton (red) and iron (black) induced showers at trigger level

(b) Trigger efficiency proton (red) and iron (black) induced showers at final cut level

**Figure 6.13:** Trigger efficiency at trigger level and final cut level for different primaries.

Fig. 6.15 shows the expected core positions and cumulative number of expected triggered stereo events per $10^6$ s of uptime. All 6 possible stereo combinations have been investigated. The next neighbour stereo topologies are predicted with a reasonable number of 45-65 events/$10^6$ s. Additionally, the diagonal topologies, *Coihueco-Los Morados* and *Loma Amarilla-Los Leones*, have rather small apertures, and reproduce nicely the feature that these events can only start at higher energies due to the larger attenuation. Furthermore, it was shown in this analysis that most events that trigger between diagonal facing stations should be 3-fold and 4-fold events. The number of expected exclusive[9] stereo events between Los Leones and Loma Amarilla is 0.02 events/$10^6$ s uptime compared to an inclusive number of 4.19 events/$10^6$ s uptime. The corresponding exclusive number for events between Los Morados and Coihueco is 0.8 compared to 14.34 inclusive events/$10^6$ s uptime. This is in agreement with observed data. Fig. 6.16 investigate the inclusive 3-fold apertures. As expected 3-fold topologies including the most distant telescope stations Loma Amarilla and Los Leones are suppressed by a factor 2. In addition, the exclusive numbers indicate that a non-negligible fraction of 3-fold events are 4-fold events. For Los Leones-Coihueco-Loma Amarilla we predict 0.7 exclusive events/$10^6$ s uptime (compared to about 3.4 inclusive events per $10^6$ s) and for *Los Morados-Loma Amarilla-Coihueco* 4.6 exclusive events/$10^6$ s uptime are predicted (compared to 7.4 inclusive events/$10^6$ s uptime). Finally, for a subset of data from the year 2004 the corresponding uptimes for all possible telescope combination have been estimated. The calculated exclusive apertures have been folded with the individual uptimes of the corresponding telescope

[9] Here, *'exclusive'* means that only the telescopes mentioned have been triggered and no additional one. On the other hand, *'inclusive'* counts all events of a certain topology, regardless if they where fragments of other topologies or not. E.g., an exclusive 3-fold event is composed by 3 inclusive stereo events.

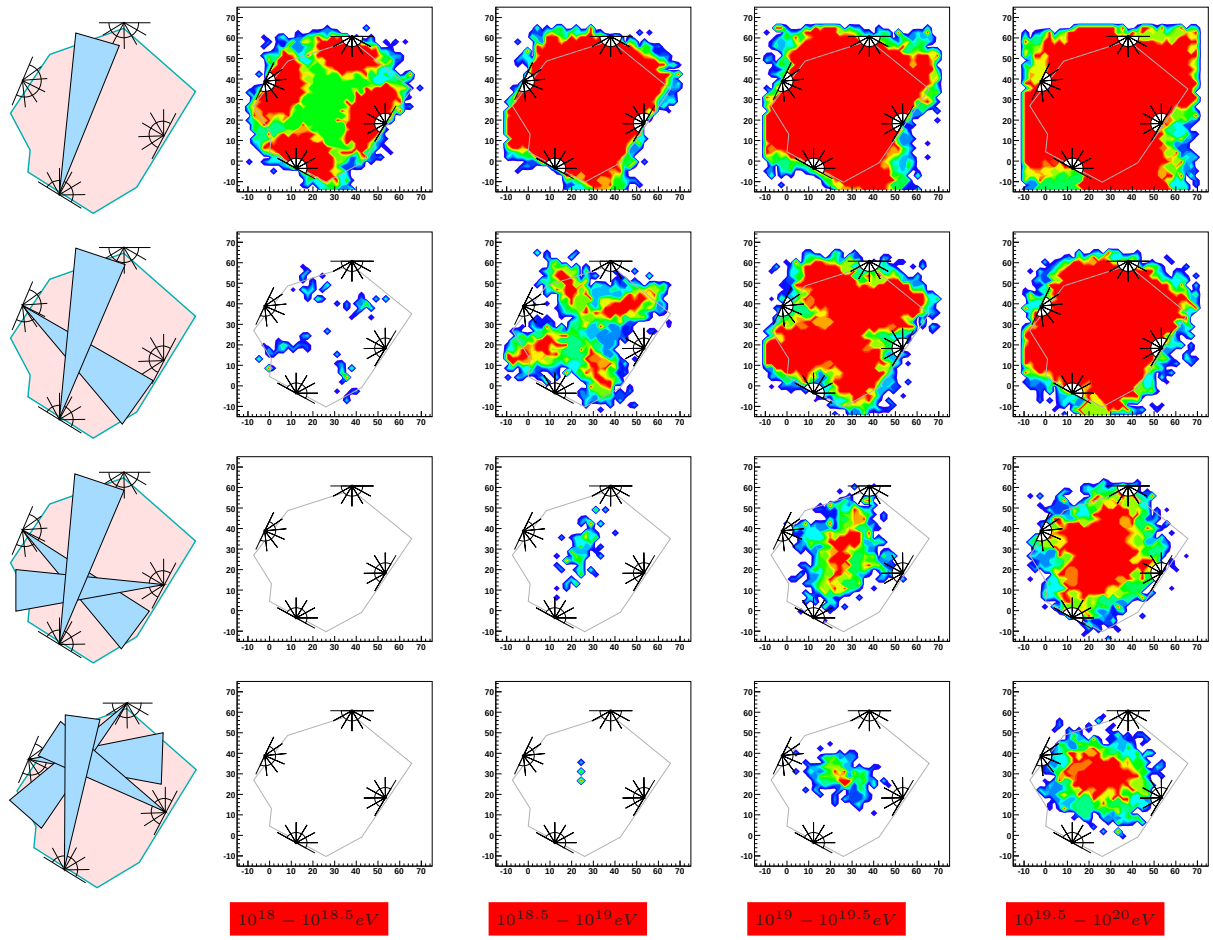



**Figure 6.14:** Trigger efficiency (in arb. units) for various event topologies as function of the core location on ground for increasing energies.

combinations to obtain the corresponding exposure. Therefore, the method of importance sampling is applied again to represent a physical flux. Fig. 6.17(a) shows the predicted exposure for stereo events. The distribution is shifted towards Coihueco due to the fact that Los Morados has come later into operation. The statistics do no permit a quantitative analysis. Nevertheless, the predicted locations correspond with the experimental data qualitatively. During the investigated period only three 3-fold events have been detected. Therefore, the calculated 3-fold exposure (see Fig. 6.17(b)) represents only the topology Los Leones-Coihueco-Los Morados, which has been operated during that period. The distribution indicates qualitatively, where most 3-fold events are expected and the three detected events might give the impression that the result is reasonable.
Since the Pierre Auger Observatory has detected its first 4-fold event it is worth while to investigate also the 4-fold exposure. Until now it has not been possible to calculate the exact uptime for the

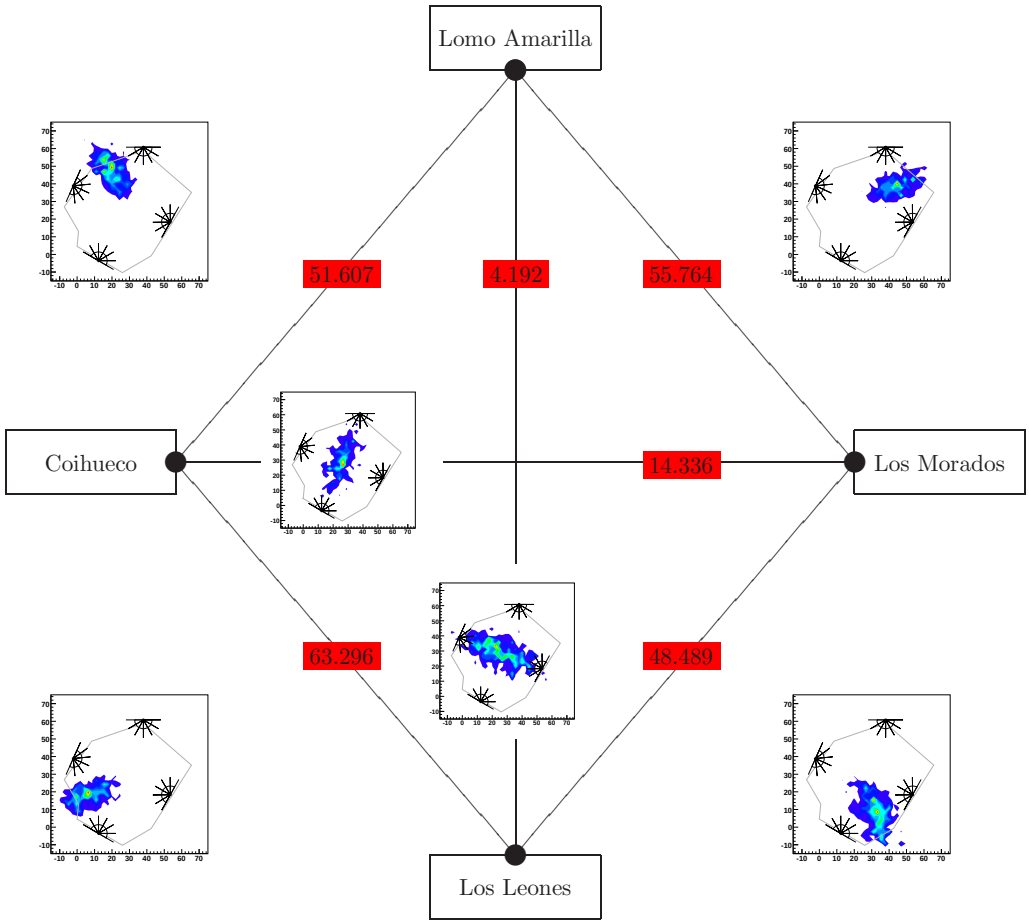



**Figure 6.15:** Aperture and predicted events per $10^6$ s uptime for all possible (inclusive) stereo configurations.

4-fold configuration. Nevertheless, the prediction for 4-fold events has been calculated roughly to be 2.6 events/$10^6$ s uptime. With a single event it is not possible to validate the 4-fold exposure quantitatively. However, the MC prediction seems to provide a reasonable number, as the following assumption shows: it has taken about $9.5{\cdot}10^6$ s (110 days) from the moment when Loma Amarilla has seen first light and the detection of the first 4-fold event. Assuming a 10 % duty cycle the expected number of 4-fold event for this period is between 1-2. Furthermore, the detected 4-fold event occurs in the expected region.
Summarising, re-weighted MC provides predictions on shape and normalisation of observables that can be compared to experimental data. This comparison shows a high level of reliability and ac-

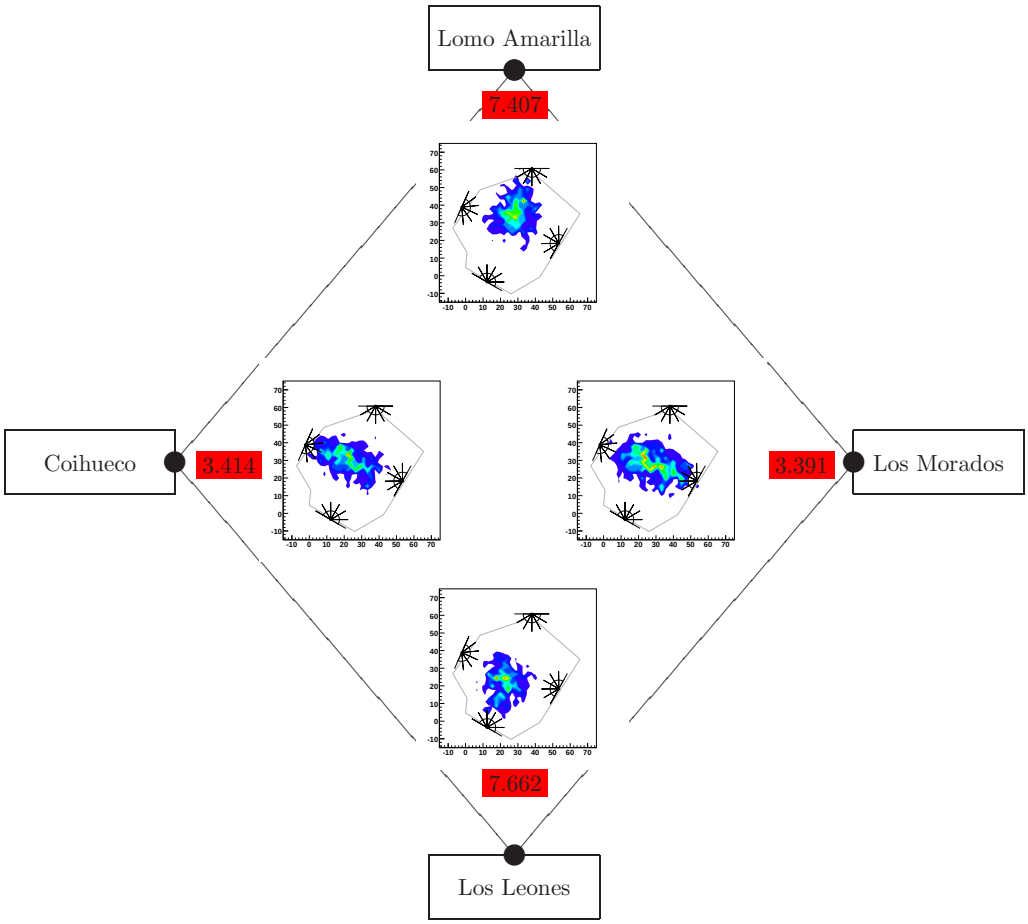
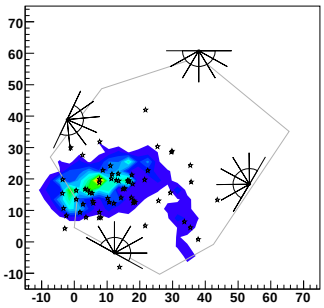
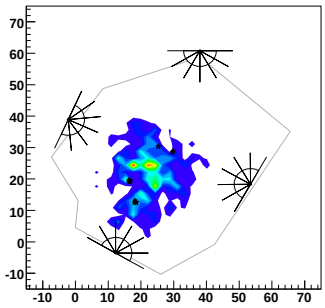



**Figure 6.16:** Aperture and predicted events per $10^6$ s uptime for all possible (inclusive) 3-fold configurations.

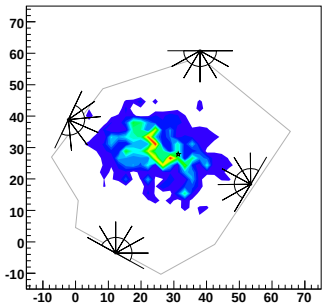
curacy of the simulation, and indicates the sensitivity to spectral assumption. Data and MC agree within shape and normalisation up to a level of 5 %-15 % (within uncertainties steming from the input spectrum hypotheses and 10 % uncertainties due to the uptime estimation). The potential to use this simulated MC for unfolding the energy spectrum has been demonstrated. In a more qualitative way, the overall trigger simulation has been investigated by comparing predictions for multi-fold events to experimental data.



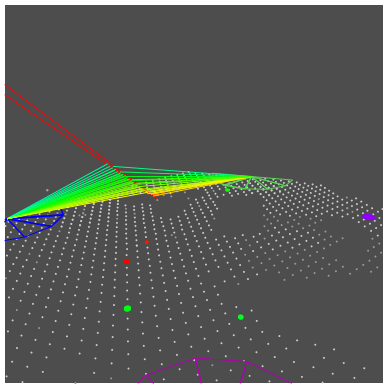

(a) Exposure for stereo events, compared to data



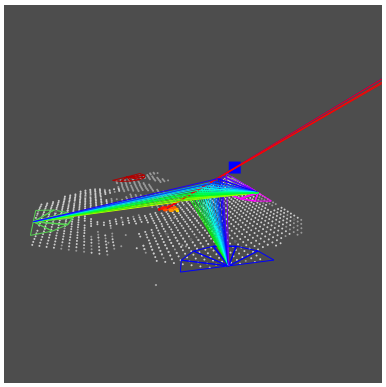

(b) 3-fold exposure compare to data



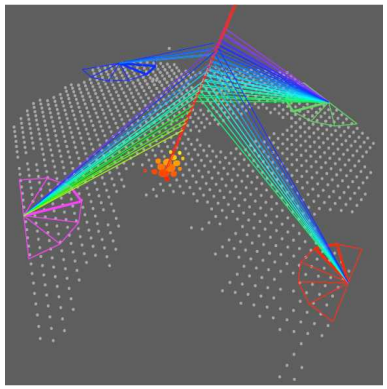

(c) Quadro aperture



(d) A 2-fold event



(e) A 3-fold event



(f) First quadro event

**Figure 6.17:** Exposure for event topologies and typical events.

# Chapter 7

# Physics Analysis

*The analysis of data inevitable involves some trafficking with the field of statistics, that gray area which is not quite a branch of mathematics - and just as surely not quite a branch of science.*

Numerical Recipes, p.614

In contrast to the previous investigations, this chapter introduces methods to compare data and simulation at physics level. The previous chapters have demonstrated that with the current status of simulation and reconstruction software it is possible to quantify the detector response and reconstruction in terms of limited resolution and limited acceptance with high accuracy. For known resolution and acceptance the determination of the primary flux is reduced to a Fredholm equation, that can be solved numerically by implementation of the Gold unfolding algorithm. The implementation has been developed by myself in collaboration with the Wuppertal group and is strongly motivated by H. Ulrich [186], sharing with me the insights of his implementation to unfold KASCADE data.

## 7.1 The Problem of Unfolding

The uncertainties of the reconstructed energy for a certain air-shower detected by the fluorescence telescopes are usually determined under the assumption of mono-energetic flux. This has the advantage, that no *a priori* constrains to the energy distribution have to be made. In order to reconstruct the true energy spectrum these resolutions have to be taken into account, without any *a priori* knowledge of the spectrum, as this would bias the final result. Thus, unfolding algorithms have to be applied in order to get the correct spectrum. The effect of bin-by-bin

migration is non-negligible, as the energy resolution is $\sim 20\,\%$ on a linear scale[1] (see Chap. 5). Basically, when interpreting the distribution of reconstructed energies at trigger-level in terms of the physical spectrum, two different statistical problems have to be faced: **limited acceptance** (see Chap. 7.1.1), a minor problem that is handled by bin-wise correction, and **limited resolution** (see Chap.7.1.2), which can lead to a deformation of the spectral shape. The effect of a limited resolution on a power-law spectrum is analysed (see Chap. 7.1.3) and a more general description is given by Fourier analysis (see Chap. 7.1.4). Unfolding is a technique for dealing with limited resolution (see Chap. 7.2). A Gold-unfolding algorithm based on [41] has been elaborated. The advantage of the Gold algorithm compared to other unfolding algorithms like regularised unfolding or Bayesian methods is discussed later (see Chap. 7.3).

### 7.1.1 Limited Acceptance

A measured, uncorrected energy spectrum $\hat{g}(x)$ differs from the true physical spectrum, because the detector sensitivity is a function of the energy. It has to be understood, how both spectra correspond to each other. Here, high energy air showers are much brighter than low energy ones, and below $10^{17}\,\mathrm{eV}$ the reconstruction efficiency is decreasing drastically. Hence, the ratio of the true amount of air showers compared to the detected ones for a low energy bin is much higher than the ratio for a high energy bin. These effects are collectively known as *limited acceptance* and for these ones has to be corrected. The convolution between the true spectrum of the CR flux ($\hat{f} = \Phi(E)$) and the reconstructed spectrum $\hat{g}$:

$$\hat{f}_i = A_{ij}\hat{g}_j$$

is very simple in this case as the response matrix $A_{ij}$ is diagonal. So, it is possible to correct the spectrum bin-wise by generating a Monte Carlo sample and weight each bin with

$$\hat{f}_i = \hat{g}_i \frac{f_i^{\mathrm{MC}}}{g_i^{\mathrm{MC}}} \,.$$

For this strategy the detector aperture and exposure has to be calculated. This is only valid if the detector has high calorimetric precision. This can be done e.g. for the hybrid-data sample, as is has three times better energy resolution (see Chap. 5.6). Thus, for these data, even for a small binning of 0.1 in logarithmical energy units, the bin-content is dominated by the aperture and not by neighbouring bins. This is different in case of FD-mono data.

### 7.1.2 Limited Resolution

The situation becomes more difficult when considering the effect of limited resolution in addition to that of limited acceptance. The motivation comes from the fact that there is a certain probability

[1] $20\,\%$ resolution on a linear scale corresponds to a logarithmical resolution of $log_{10}(E_{\mathrm{rec}}/E_{\mathrm{true}}) \sim 11\,\%$. The logarithmical value is more useful in the context of this chapter.

for reconstructing an event of energy $x$ with the reconstructed energy $y$. If this is taken into account, the response matrix $A_{ij}$ is no longer diagonal. The off-diagonal elements are responsible for events shifting from high-energy bins to low-energy ones and vice versa. This can deform the spectral shape. As an example a power-law spectrum can be considered. In this case, in total more low-energy events shift to higher energies than high-energy events shift to low energies. The effect can be studied analytically for the case of a pure power-law spectrum and a Gaussian resolution on a logarithmic scale. This demonstrates the effect of limited resolution and provides a rough quantitative estimate of the correction that has to be applied.

### 7.1.3 Convolution of a Power-Law Spectrum

The spectra in astro-particle physics often follow a power-law. Given a primary flux $f(x)$ according to a power-law of spectral index $\gamma$:

$$f(x) = \Phi_0 \cdot \exp(\gamma \cdot x) \ .$$

E.g., $x$ can be associated with $x = log(E/E_0)$. Furthermore, it can be assumed that the energy resolution is Gaussian in logarithmic scale. Usually, experiments measure energy distributions over many orders of magnitude. In this case the probability $P(x, y)$ of reconstructing the energy $y$ for an event of true energy $x$ can be expressed via:

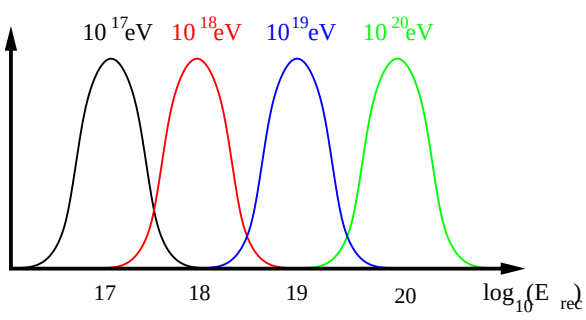
$$P(x, y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-y)^2}{2\sigma^2}\right] \ .$$

The observed, reconstructed spectrum $g(y)$ manifests as a convolution of the true spectrum with the resolution[2]

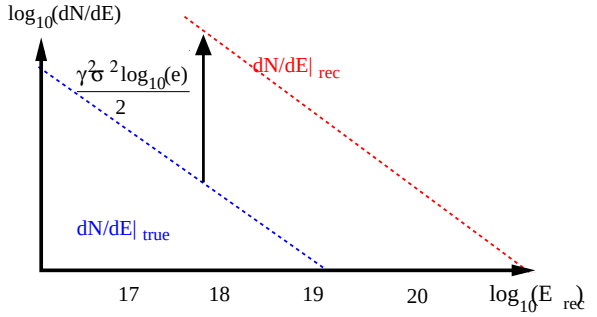
$$\begin{aligned} g(y) &= (f \otimes P)(y) = \int dx f(x) P(x, y) \\ &= \Phi_0 \exp\left\{\frac{\left[y + \sigma^2\gamma\right]^2 - y^2}{2\sigma^2}\right\} \\ &= f(y) \exp\left\{\frac{\sigma^2\gamma^2}{2}\right\} \ . \end{aligned} \tag{7.1}$$

This idealised scenario demonstrates quantitatively the effect of smearing: The absolute flux is scaled by $exp(\sigma^2\gamma^2/2)$. In the commonly used double-logarithmic representation, the shape remains the same. A power law flux with same spectral index is observed, but shifted by an offset of about $0.22 \cdot \sigma^2\gamma^2$. Obviously, the effect is insensitive on the sign of the spectral index. This is expected intuitively: The symmetric Gaussian resolution guarantees that for each energy the same number

[2] In order to point out the general effect of the convolution, effects due to limited acceptance should be neglected and the detector is assumed to have 100 % efficiency on the full energy range.




(a) Gaussian resolution.




(b) The reconstructed spectrum has the same spectral index but is shifted.

**Figure 7.1:** Schematic illustration of the power-law scenario in combination with a Gaussian resolution. The reconstructed spectrum is shifted by about $0.22\sigma^2\gamma^2$ in the double logarithmic representation.

of events is mis-reconstructed to both directions, higher and lower energies. In case of a steeply falling spectrum the following situation results: Via the left Gaussian tails of high-energy events are shifted to lower energies. Low-energy events are shifted via their right Gaussian tails to higher energies. But in total more events are shifted to high energies than vice versa, simply due the fact that there are less high-energy events. In case of a rising power-law the effect is vice versa but the result remains the same. $\sigma$ is the resolution of the energy with respect to the natural logarithm. In order to have a rough quantitative estimate of the effect on FD-mono data: The absolute flux would be shifted by about 26 %.[3] It is emphasised that the presented scenario cannot be applied to quantitative data analysis, as it is an ideal scenario. In case of the FD-mono spectrum, it is not possible to simply correct for the shift, due to:

*A priori* **knowledge of the spectrum:** The aim of the analysis is to reconstruct the spectrum. The spectral behaviour is not known *a priori*. Furthermore, the spectrum is not believed to follow a simple power-law, but more to have some features.

**Non-Gaussian resolution:** The energy resolution is not perfectly Gaussian. Moreover, the energy reconstruction is not fully symmetric, but shows a systematic bias.

**Limited range of detector sensitivity:** So far the integration is performed over all energies. In reality, the detector covers only a limited energy range. A correct treatment is to fold in also the acceptance. The drop of the efficiency at the trigger threshold usually smoothens the reconstructed spectrum.

[3]The final flux of cosmic rays is expected to follow a power-law of spectral index $\gamma \sim -3$, the energy reconstruction in natural logarithmic units is about $\sigma \sim 23\,\%$.

A more sophisticated approach is to understand the problem in terms of convolution and find an algorithm for inversion.

### 7.1.4 General Unfolding Formalism

In general, the task is to measure the physical (*'true'*) probability distribution function (p.d.f.) $f(x)$ of a certain variable $x$. Often, $f(x)$ cannot be measured directly due to imperfect detectors (*limited acceptance* and *limited resolution*). Thus, the variable $x$ is transformed by a measurement into an observable $y$ and the original distribution $f(x)$ is distorted by limited acceptance and resolution to a measured (*'reconstructed'*) distribution $g(y)$.[4] Depending on the transformation properties the variable $y$ may differ completely from $x$ and also the dimension of the observable space $y \in \mathbb{Y}$ may be larger than the physics-space $x \in \mathbb{X}$.[5]
Usually, the measurement can be expressed as a convolution of the true distribution $f(x)$ with the detector kernel $A(x, y)$.

$$g(y) = \int dx\ f(x)A(x, y)\ .$$

Thereby, the problem is expressed as Fredholm integral equation of 1st kind. The kernel $A(x, y)$ describes the response of the detector, which can be interpreted as the probability for an event $x$ to be triggered and reconstructed at $y$. Usually, the kernel is factorised in a trigger/reconstruction-efficiency $\epsilon(x)$ and a detector response s(x,y). In practice, also the observed distribution $g(y)$ is distorted by an additional statistical error background, $b(y)$. Thus, the complete problem is often formulated in this way:

$$g(y) = \int dx\ f(x)s(x, y)\epsilon(x) + b(y)\ .$$

Once the kernel is known, by simulation or direct measurement[6], the physical distribution $f(x)$ can be obtained by solving the integral equation. Hereby, an accurate estimate of $A(x, y)$ is essential for any meaningful analysis. Therefore, the validation of the simulation is needed. The detector response can now be calculated by using the produced MC sample with known input $f(x)$ and response $g(y)$. There is no standard way for inverting such an integral equation. Depending on the kernel in most cases the existence of a unique solution is not guaranteed. Usually, the integral equation is discretised and a matrix equation is obtained, that is solved numerically. However, this matrix equation does not provide a unique solution. The way it is solved and how the result is interpreted, strongly depends on the kernel function. Usually small fluctuations due to statistical uncertainties in the side-diagonal elements, can produce large distortions. In order to illustrate the

[4] From a physics point of view: The transformation includes both distortion due to measurement uncertainties as well as additional kinematic effects like smearing due to attenuation and scattering of the fluorescence light.
[5] E.g. when unfolding the energy distribution the most obvious energy correlated variable is the reconstructed energy. But one can think of experimental situations that provide more energy correlated variables. Also these problems are covered by unfolding.
[6] if possible

convolution problem, it is useful to decompose $f(x)$ as a Fourier sum:[7]

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos(nx) + b_n \sin(nx) \ ,$$

where $a_n = \frac{1}{\pi}\int_0^{2\pi} dy \ N(y)\cos(ny)$ and $b_n = \frac{1}{\pi}\int_0^{2\pi} dy \ N(y)\sin(ny)$. Hereby, the problem of convolution an arbitrary continuous function is reduced to the question of how periodic functions are effected by unfolding. For illustration, a Gaussian resolution is assumed once more:

$$p(x,y) = \frac{1}{\sqrt{2\pi}\sigma} exp\left\{-\frac{(x-y)^2}{2\sigma^2}\right\} \ .$$

The convolution integral for each term in the Fourier sum can be analysed by estimating:

$$\int_{-\infty}^{\infty} dx\, p(x,y)\cos(nx) = \exp\left(-\frac{n^2\sigma^2}{2}\right)\cos(ny)$$
$$\int_{-\infty}^{\infty} dx\, p(x,y)\sin(nx) = \exp\left(-\frac{n^2\sigma^2}{2}\right)\sin(ny) \ .$$

Hence, the measured distribution turns out to be:

$$\begin{aligned} g(y) &= \int_{-\infty}^{\infty} dx\, p(x,y) f(x) \\ &= \frac{a_0}{2} + \sum_{n=1}^{\infty} \underbrace{a_n \exp\left(-\frac{n^2\sigma^2}{2}\right)}_{=:\tilde{a}_n} \cos(nx) + \underbrace{b_n \exp\left(-\frac{n^2\sigma^2}{2}\right)}_{=:\tilde{b}_n} \sin(nx) \ . \end{aligned}$$

This demonstrates the difficulty of unfolding. $g(y)$ is directly measured and the Fourier coefficients $\tilde{a}_n$ and $\tilde{b}_n$ can be obtained with respective statistical errors. When calculating the Fourier coefficients of the physical (*'true'*) distribution, $a_n$ and $b_n$, the error scales exponentially. As $f(x)$ is considered to be piecewise continuous, the Fourier sum converges. Thus, both $a_n$ and $b_n$ should be a zero sequence. The problem arises that these coefficients are very soon dominated by the measurement uncertainties, so that the unfolded result is completely deformed by statistical fluctuations. This problem can be handled by dedicated unfolding algorithms.

### 7.1.5 Discretisation of the Problem

Here, the measure $x$ is associated with the distribution of primary energies:

$$x = \log_{10}(E_{\text{true}}) \ .$$

[7] Each spectrum $f(x)$ discussed from now on may be considered as a piecewise continuous function, and therefore, the Fourier sum converges absolutely at $f(x)$.

The observed distribution is the one of the reconstructed energies of the corresponding air-showers, that have been triggered and survived all quality cuts:[8]

$$y = \log_{10}(E_{\mathrm{rec}}) \ .$$

A direct functional behaviour for reconstructed and true energies is difficult to access. Moreover, discrete histograms of the distributions are obtained. Given $\nu_j$ as the bin content of the $j$-th measured bin and $\mu_i$ as the content of the $i$-th bin content in the true histogram:

$$\nu_j := \int_{\mathrm{bin}j} d\log_{10}(E_{\mathrm{rec}})\ g(\log_{10}(E_{\mathrm{rec}})) \tag{7.2}$$

$$\mu_i := \int_{\mathrm{bin}i} d\log_{10}(E_{\mathrm{true}})\ f(\log_{10}(E_{\mathrm{true}})) \ . \tag{7.3}$$

Furthermore, defined $\mu_{tot} = \sum \mu_i$ as the total sum of all events passing through the detective volume. In this case, the distribution of reconstructed energies can be expressed by:

$$\begin{aligned}
\nu_i &= \mu_{tot} P(\text{event reconstructed in bin } i) \\
&= \mu_{\mathrm{tot}} \int d\log_{10}(E_{\mathrm{true}})\ P\left( \begin{array}{c} \text{reconstructed and} \\ \text{observed in bin } i \end{array} \middle| \begin{array}{c} \text{event of energy } E_{\mathrm{true}} \\ \text{and survived all cuts} \end{array} \right) \epsilon(E_{\mathrm{true}}) f(\log_{10}(E_{\mathrm{true}})) \\
&= \mu_{\mathrm{tot}} \int_{\mathrm{bin}i} d\log_{10}(E_{\mathrm{rec}}) \int d\log_{10}(E_{\mathrm{true}})\ s(E_{\mathrm{rec}}|E_{\mathrm{true}}) \epsilon(E_{\mathrm{true}}) f(\log_{10}(E_{\mathrm{true}})) \\
&= \sum_j \underbrace{\frac{\int_{\mathrm{bin}i} d\log_{10}(E_{\mathrm{rec}}) \int_{\mathrm{bin}j} d\log_{10}(E_{\mathrm{true}})\ s(E_{\mathrm{rec}}|E_{\mathrm{true}}) \epsilon(E_{\mathrm{true}}) f(\log_{10}(E_{\mathrm{true}}))}{\mu_i/\mu_{\mathrm{tot}}}}_{R_{ij}} \mu_j \\
&= R_{ij}\mu_j \ .
\end{aligned} \tag{7.4}$$

At this point, the problem of solving the integral equation is reduced to a discrete matrix problem. Furthermore, a clear instruction is given to obtain the response matrix from the MC.

$$\begin{aligned}
R_{ij} &= \frac{\int_{\mathrm{bin}i} d\log_{10}(E_{\mathrm{rec}}) \int_{\mathrm{bin}j} d\log_{10}(E_{\mathrm{true}})\ s(E_{\mathrm{rec}}|E_{\mathrm{true}}) \epsilon(E_{\mathrm{true}}) f(\log_{10}(E_{\mathrm{true}}))}{\int_{\mathrm{bin}j} d\log_{10}(E_{\mathrm{true}})\ f(E_{\mathrm{true}})/\mu_{\mathrm{tot}}} \\
&= \frac{P(\text{reconstructed in bin } i \text{ and true value is in bin } j)}{P(\text{event has energy } j)} \\
&= P\left( \begin{array}{c} \text{event reconstructed} \\ \text{in bin } i \end{array} \middle| \begin{array}{c} \text{event has true} \\ \text{energy in bin } j \end{array} \right)
\end{aligned} \tag{7.5}$$

Thus, the response matrix element $R_{ij}$ is the conditional probability to reconstruct the energy to fall in energy bin $j$, for an air-shower associated with a true energy falling in bin $i$. The response

[8] It is notified, that at no point in the analysis it is assumed that the reconstructed energy is symmetric around the true energy, or that the reconstructed energy has no bias. In principle, it is also possible to regard two different measures like sum of all signals and true energy.

matrix weakly depends on the spectral distribution of the kernel-MC $f(E_{\text{true}})$. As long as the bins of the unfolded spectrum and the kernel are small enough the dependence on the input spectrum cancels out. For the further discussion of possible unfolding algorithms we have to assume that the uncertainties in the response matrix due to the spectral distribution of the kernel MC can be neglected. This has been tested to be a reasonable assumption (see App. B).
Equation 7.4 is the general starting point of all analysis that have to treat deconvolution. The question might arise, why not directly inverting the matrix to obtain the spectrum. Unfortunately, in most cases a direct inversion of the system of equations fails, due to statistical fluctuations as well as the structure of the response matrix itself:

**Statistics:** This raises the question, how to interpret the *'exact'* solution of the equation system. Usually, the 'exact' solution is not the one of physics interest. Due to statistical uncertainties in the observed distribution $\nu_i$, the system of equations is never exactly fulfilled. Thus, the existence of a solution that is reasonably close[9] to the expected values of the physical (*'true'*) distribution $\mu_i$ cannot be guaranteed. Especially, this happens in case of different binning in the observed and physics parameter space. In that case, the response matrix is no longer quadratic and the system in Eqn. 7.4 is either over- or under-determined. In case of an exact solution existing, it is mostly not usable. After forward folding, the exact solution includes all statistical uncertainties of $\nu_i$ and thus, it shows an oscillating-behaviour. As discussed for the Fredholm equation and motivated with the Fourier analysis (see Chap. 7.1.4) the amplitudes of the oscillation can be rather large. Hence, the exact solution is not the meaningful and needed one.

**Structure:** Neglecting statistical uncertainties and assuming a perfect measurement, still leaves the problem of the structure of the response matrix itself. Usually, the diagonal elements and their next neighbour elements differ significantly from zero. Furthermore, the matrix is nearly symmetric along the diagonal[10]. This leads to the fact that the response matrix is nearly singular and therefore,

$$\det R \sim 0 \; .$$

For a complete singular matrix no unique solution exists, for a nearly singular matrix the solution is mostly far off the truth. As one has implicitly to divide by $\det R$ during the matrix-inversion process, numerical uncertainties arise. Senseless high values and an oscillating solutions are obtained.

**Numerical problems:** Furthermore, one has to consider numerical uncertainties for estimating the response matrix from the kernel-MC.

All in all, this leads to the necessity to deconvolute Eqn. 7.4 in a more sophisticated way. Algorithms that deconvolute such problems are called unfolding algorithms.

[9] At least that can be considered as meaningful solution.
[10] The reconstructed energy should spread symmetric around the true energy.

## 7.2 Unfolding Algorithms

Here, some of the most prominent unfolding algorithms should be explained. We follow mostly the review in [54]. For a more detailed description it is recommended to have also a look into [40, 41, 42, 43, 23, 57, 58]. In Chap. 7.3, the performance of different algorithms are compared with respect to the present problem of unfolding the FD-mono spectrum.

### 7.2.1 Direct Algorithms

**Fit parametric function:** In case the functional progression of the demanded physics distribution $\mu_i$ is known one can try to parametrise the input distribution and look for that set of parameters $\alpha_i$ for which $f(x, \alpha_i)$ forward folded describes the data best. This can be done using a minimising algorithm, e.g., by minimising

$$\chi^2 = \sum \frac{\nu_i - \sum_j R_{ij} f(\alpha_k, E_{\mathrm{true,j}})}{\sigma^2(E_{\mathrm{rec,i}})} \; .$$

This approach has been discussed in context with the data-MC comparison in Chap. 6.6. In principle, one can assume as parametric function a broken power-law spectrum, with two different spectral indices and the position of the kink as well as absolute flux as free parameters. This can be fitted in such a way, that data and MC agree best at detector level. The disadvantage of this procedure is that an *a priori* knowledge of the spectrum is needed. Following this approach it is impossible to discover unknown features in the spectrum. Furthermore, the final result indicates that a broken power-law with one kink is not the best solution for high energy data.

**Regularised unfolding:** Following the approach of fitting a parametric function one can extend the idea to fit directly the bin contents of the demanded distribution $\nu_i$ using e.g. a $\chi^2$ minimisation. In that case, one would try to investigate:

$$\chi^2 = \sum \frac{\nu_i - \sum_j R_{ij} \mu_i}{\sigma^2(E_{\mathrm{rec,i}})} \; .$$

Unfortunately, the minimisation of this equation is nothing else than the inversion of equation 7.4. This fails due to statistics and structure of the response matrix (see Chap. 7.1.5). The solution found by direct minimisation would be as meaningless as the one obtained by direct matrix inversion. Therefore, one can try to restrict the solution space by introducing an additional constrain to the solution $\nu_i$. This constrain is refered to as *'regularisation'* [40, 41, 42, 43] and is implemented by appending an additional functional $S(\nu_0, ..., \nu_N)$. The regularisation effect can be tuned by the so called *'regularisation parameter'* $\tau$. Hence, e.g., the $\chi^2$-term transforms to

$$\chi^2 = \sum \frac{\nu_i - \sum_j R_{ij} \mu_i}{\sigma^2(E_{\mathrm{rec,i}})} + \tau S(\vec{\mu}) \; .$$

This approach has been firstly mentioned in [119] and has been refined by [127] and [40, 41, 42, 43]. We discuss two different implementation of the regularisation, the Tikhonov unfolding and the principle of maximum entropy.[11]

**Tihkonov unfolding:** For this approach one postulates a smooth behaviour of the solution. This can be achieved by a regularisation term that represent the behaviour of a derivative of the demanded solution:

$$S(x) = \int dx \left( \frac{d^k f(x)}{dx^k} \right)^2 .$$

This requires, that the minimisation of the functional the mean square of the derivatives is small. One can use several derivatives, but usually one is sufficient. Normally, the second derivative is used to minimise the curvature of the final solution. The advantage is that it is possible to represent the constrains as matrix problem. In this case the regularisation functional can be approximated by the difference between neighbouring elements of $\nu_i$. Hence, for a constant bin-width it is obtained:

$$S(\mu_0, ...\mu_N) = \frac{1}{\Delta^2_{\text{bin}}} \sum_{i=2}^{N-1} (\mu_{i-1} - 2\mu_i + \mu_{i+1}) .$$

The bin-width normalisation $\Delta^{-2}_{\text{bin}}$ is usually factorised in the regularisation parameter $\tau$. In that case, the regularisation manifests itself as a matrix operation:

$$S(\vec{\nu}) = \vec{\mu}^T B^T B \vec{\mu} ,$$

where

$$B = \begin{pmatrix} 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 \\ \vdots & & & \ddots & & \vdots \\ 0 & \cdots & 1 & -2 & 1 & 0 \\ 0 & \cdots & 0 & 1 & -2 & 1 \end{pmatrix} .$$

Introducing the error matrix $C_{ij} = \delta_{ij}/\sigma(\nu_i)$ a suitable representation for the $\chi^2$ functional can be found:

$$\chi^2 = [C(\vec{\nu} - R\vec{\mu})]^T [C(\vec{\nu} - R\vec{\mu})] + \tau \vec{\mu}^T B^T B \vec{\mu} .$$

Hereby, the $\chi^2$ minimisation becomes equivalent to the problem of solving the linear equation system:

$$\begin{pmatrix} CR \\ \sqrt{\tau} B \end{pmatrix} \vec{\mu} = \begin{pmatrix} C\vec{\nu} \\ 0 \end{pmatrix} .$$

This system is over-determined and there are standard procedures to minimise these kind of problems, like the *damped least square algorithm*. In case of spectral unfolding, one expects

[11] E.g., the RUN code is based on a refined Tikhonov regularisation (see Chap. 7.3).

a positive solution, meaning that all bin contents are positive[12]. In the original Tikhonov algorithm it is difficult to implement the constrain of positiveness. The algorithm tends to respond to small non-negative entries with zero or small negative elements. One has to give up the nice matrix formalism, in order to avoid this unphysical result of negative amplitudes. An improvement of the Tikhonov algorithm is implemented in the RUN algorithm from V. Blobel [40]. This approach is based on the Tikhonov algorithm, but interpolates the solution by B-splines. The system of parameters to define the B-spline representation is over-determined. An additional constrain to the B-spline is achieved by approximating a natural spline in the way that the total curvature

$$\int dx \left(S''(x)\right)^2$$

should minimise. Furthermore, one can fold in the requirement of positiveness.

**Principle of maximum entropy:** Another standard application in image- or pattern-restoring convolution problems is the principle of maximum entropy [54]. This methods introduces the so called entropy of information:

$$H = -\int dx f(x) \ln(f(x)) = -\sum_{i=1}^{N} \mu_i \ln(\mu_i) \ .$$

The negative entropy can be used as a regularisation term, that has to be minimised. In that case, a minimal $S(\vec{\mu}) = -H(\vec{\mu})$ represents maximum entropy. Interpreting the entropy as a measure for the absence of structure in f(x), maximum entropy gives the most smooth solution. The undulation features (see Chap. 7.1.5) are suppressed efficiently. Moreover, this formalism uses implicitly the positiveness of $\mu_i$. A general problem of these methods is that one ends up with a non linear minimisation functional which could be complex.

**First guess via method of correction factors:** The method of correction factors is a good way to find start values for any minimising approach. In this case, the diagonal matrix is assumed to be diagonal and thus an aperture correction analysis is performed, neglecting the effect of limited resolution. The matrix equation 7.4 then simplifies to

$$\nu_i = R_{ii}\mu_i \ .$$

Introducing the correction factor $C_i = 1/R_{ii}$ *'aperture'* the expectation for the solution is

$$E(\vec{\mu_i}) = C_i\nu_i \ .$$

This is only an approximation and just valid in case the ratio $\mu_i/\nu_i$ is the same in MC and data, as

$$E(\vec{\mu_i}) = \frac{\mu_i^{MC}}{\nu_i^{MC}}\nu_i = \left(\frac{\mu_i^{MC}}{\nu_i^{MC}} - \frac{\mu_i}{\nu_i}\right)\nu_i + \mu_i \ .$$

[12] A negative flux makes no sense.

The resulting solution does not treat statistics and systematics in a correct manner and is not representing effects of limited resolution. Nevertheless, it is a good first guess for any iterative or minimisation algorithm.

### 7.2.2 Iterative Algorithms

Iterative algorithms are procedures to approximate the solution by recursive series. The expectation values for the solution $\mu_i^k$ at iteration stage $k$ can be expressed by the solution at the previous iteration stage $\mu_i^{k-1}$. The advantage of most iterative procedures is that they approximate the most meaningful solution instead of the exact solution (see Chap. 7.1.5). One has to guarantee convergence of the iteration procedure. Due to numerical uncertainties arising in the context of iteration and the nearly singular structure of the response matrix it can happen that at a certain iteration depth fluctuations get amplified and the solution deviates from any demanded solution. Therefore, it is important to define a suitable cutoff criteria defining when to stop the iteration. We discuss this exemplary in detail for the Gold algorithm in Chap. 7.4.2.

**Van Citter procedure:** One of the most profound algorithms is the Van Citter-algorithm [48, 47, 189]. It constructs the expectation by correcting the predecessor solution with the residual of the forward folded predecessor solution and measurement:

$$\mu_i^{k+1} = \mu_i^k + \alpha \left( \nu_i - \sum_j R_{ij} \mu_j^k \right) \ .$$

Obviously, this procedure can only be applied when observational and solution space are of the same dimension. A way out can be obtained by the following substitutions:

$$R \Leftrightarrow R^T R \vec{\nu} = R^T \vec{\nu} \ .$$

The additional statistical uncertainties of the data can be propagated using the error matrix $C_{ij} = \delta_{ij} \sigma(\nu_i)$ and defining the modified response matrix and data vector

$$\tilde{R} = C R \vec{\tilde{\nu}} = C \vec{\nu} \ .$$

The series is converging for $0 < \alpha < 2/\lambda_{\max}$, with $\lambda_{\max}$ being the largest eigen-value of the response matrix. One disadvantage of the procedure is that the positiveness of the solution cannot be guaranteed, and under certain conditions, a strong amplification of fluctuations can happen.

**Gold algorithm:** The Gold algorithm is a refinement of the Van Citter algorithm. The algorithm converges stable and is more robust. The advantage is the implicit positiveness of the solution that is achieved by a further regularisation. The motivation of this approach is similar to typical eigenvalue problems. The essential requirement is that the response matrix is a

quadratic[13] non-singular matrix. All elements have to be positive[14] and the diagonal elements have to be non-zero. In that case the problem is an eigenvalue problem and there exists a matrix $D$ with only diagonal entries, so that

$$\vec{\mu} = D\vec{\nu}\,, \quad \text{with} \quad D_{ij} = \delta_{ij}\frac{\mu_i}{\nu_j}\,.$$

Note, that $D$ is the inverse of the response matrix, only in the special case if $R$ itself is a diagonal matrix. The idea of the Gold algorithm is now instead of iteratively approximating $\mu_i$, to find an iterative procedure to approximate the diagonal eigen-matrix $D$. The iteration procedure can be found by

$$D_{ii}^{k+1} = \frac{D_{ii}^{k}\nu_i}{\sum_j R_{ij}D_{jj}^{k}\nu_j}\,.$$

Multiplying the equation by $\nu_i$ leaves directly an iterative prescription for the solution vector, that also shows that for positive measurements in $\nu_i$ one can guarantee the positiveness of all $\mu_j$:

$$\mu_i^{k+1} = \frac{\mu_i^{k}\nu_i}{\sum_j R_{ij}\mu_j^{k}}\,.$$

In [80] it is shown that the series 7.2.2 converges to the exact solution of the equation system 7.4. The convergence is equivalent to the convergence of the following quadratic functional to zero:

$$F(\vec{\mu}^k) = (\vec{\mu}^k - \vec{\mu})^T R(\vec{\mu}^k - \vec{\mu}) \overset{k\to\infty}{\longrightarrow} 0\,.$$

The nice feature of the Gold algorithm is that the choice of the starting values can change the rate of convergence, but not the convergence itself. This makes the Gold algorithm very robust.

As already discussed the exact solution of Eqn. 7.4 is not of interest (see Chap. 7.1.5). Therefore, a good cutoff-criterium is needed to stop the iterative procedure at the point when the most meaningful answer to the matrix equation 7.4 is found. The exact implementation of such a criterium is discussed in Chap. 7.4.2.

As already noted the response matrix has to be quadratic and in case of different dimensions of measurement and solution space, the problem can be circumvented by a similar substitution as already mentioned for the Van Citter algorithm. In case of the Gold algorithm one has to guarantee that the $m \times n$ response matrix, with $m > n$ has rank $n$.[15] It has been shown in [80] that the quadratic $n \times n$ matrix $R_{\text{mod}} = R^T R$ has also rank $n$, is positive definite and thus equation 7.4 can be modified by multiplication with $R^T$:

$$\vec{\nu} = R\vec{\mu} \Leftrightarrow \underbrace{R^T\vec{\nu}}_{\vec{\nu}_{\text{mod}}} = \underbrace{R^T R}_{R_{\text{mod}}}\vec{\mu}\,. \tag{7.6}$$

[13] In case the observable and physics space have different dimensions, a similar approach as for the Van Citter algorithm can be performed.

[14] That can be guaranteed in case of the spectral unfolding as it contains only probabilities.

[15] Otherwise the existence of enough eigenvalues for the eigen-matrix is not given.

The matrix equation $\vec{\nu}_{\mathrm{mod}} = R_{\mathrm{mod}}\vec{\mu}$ fulfils the requirement for the Gold-convergence.
The statistical uncertainties can be propagated through the iteration, by modifying the response matrix with the error matrix $C_{ij} = \delta_{ij}/\sigma(\nu_i)$. Using $\tilde{R} = CR$ the following equation represents correct error propagation:

$$(CR)^T(CR)\vec{\mu} = (CR)^T C\vec{\nu} \Leftrightarrow \tilde{R}^T\tilde{R}\vec{\mu} = \tilde{R}^T C\vec{\nu} := \vec{\tilde{\nu}} \; . \tag{7.7}$$

As $C$ is a diagonal matrix it does not change the rank of the response matrix $R$ (or $R_{mod}$). The errors have to be positive. In that case $\tilde{R}$ is still positive definite. Using the modified matrix and data vector it is possible to show that the iterative Gold procedure is equivalent to a standard $\chi^2$ minimisation of equation 7.4:

$$F(\vec{\mu}) = \sum_i \frac{\left(\sum_j R_{ij}\mu_j^k - \nu_i\right)^2}{\sigma(\nu_i)^2} = \chi_k^2 \overset{k\to\infty}{\longrightarrow} 0 \; .$$

**Bayesian unfolding:** This approach makes use of the Bayesian theorem. The details of the implementation on unfolding problems can be found in [57, 58]. Basically, the idea is as follows: If there exists a solution for Eqn. 7.4, then there should be a matrix $P$ that describes the probability that a measurement $\nu_i$ contributes to the solution $\mu_j$. In that case, the solution can be written as a linear combination of the measurement with the probabilities as coefficients:

$$\mu_i = P_{ij}\nu_j = \sum_j P(\mu_i|\nu_j)\nu_j \; .$$

It is emphasised, that $P$ is not the inverse of the response matrix $R^{-1}$. Moreover, it should be interpreted it in a statistical sense. The interpretation of $P$ implies only positive elements and by this the positiveness of the solution is guaranteed. Similarly to the Gold algorithm one has converted the problem of finding the solution $\vec{\mu}$ to the problem of finding the matrix $P$.[16] The procedure can be summarised as follows:
A first guess for the solution is provided, e.g., this can be done by using the method of correction factors. In that case, an initial matrix $P^0$ can be found via

$$P_{ij}^0 = P(\mu_j^0|\nu_i) = \frac{R_{ij}\mu_j^0}{\sum_l R_{jl}\mu_l^0} \; .$$

By this, a better estimate for the solution can be obtained:

$$X_j^{k+1} = \sum_i P(\mu_j^k|\nu_i)\nu_i =: \sum_i P_{ij}^k\nu_i \; .$$

Sometimes it might be necessary to consider efficiencies $\epsilon_j$, arising from limited acceptance:

$$\epsilon_j := \sum_i P(\nu_i|\mu_j^k) = \sum_i R_{ij} \; .$$

[16] In the Gold algorithm the matrix $D$ is diagonal, here $P$ does not necessarily have to.

This has to be applied to correct in the $k$-th iteration the estimated solution element

$$X_j^{k+1} = \frac{1}{\epsilon_j} \sum_i P(\mu_j^k|\nu_i)\nu_i \ .$$

A cutoff criterium has to be found similar to the other iterative approaches. The Bayesian approach is rather similar to the Gold approach, but has more calculation steps as it does not converge in the eigen-space of the problem. Nevertheless, it is shown that both, Gold and Bayesian algorithms, perform similar for the problem of spectral unfolding (see Chap. 7.3). The unfolding implementation refined by G. D'AGOSTINI is applied as a representation for the Bayesian algorithms. Although, the performance is rather similar, the definition of a clear cutoff criteria is more complex for the Bayesian approach.

## 7.3 Comparison of Unfolding Algorithms

It is reasonable to test the performance of a subset of promising algorithms, to find the most suitable algorithm for the final data analysis. *A priory* , it makes sense to neglect the original Tihkonov algorithm and the Van Citter iteration, as they cannot guarantee the positiveness of the solution. The principle of maximum entropy is applied in pattern- and picture-restoring problems. These kind of problems are characterised by the feature that the equations are mostly under-determined. This is not the case in our unfolding problem. The method of correction factors is not really a kind of unfolding procedure, but just simply an acceptance correction and thus does not include smearing due to bad resolution. As shown, it is necessary to include effects due limited resolution of the detector response in case of FD-mono reconstruction. Thus, the method of correction factors is considered only as a first guess for the other algorithms.
Finally, we investigate the Gold iterative algorithm, a Bayesian unfolding algorithm that has been refined by D'AGOSTINI [57, 58] and a regularised unfolding algorithm using B-splines as regularisation functional. The latter implementation made use of the RUN-algorithm developed by V. BLOBEL [40, 41, 42, 43]. The requirements of a suitable algorithm should be that the solution is stable under the following conditions:

- The final solution allows for a power-law. That does not mean that we expect a featureless power-law. This should express that the final bin entries for the $\mu_i$ can differ by several orders of magnitude.
- The investigated energy range is at least three orders of magnitude, namely $10^{17}$ eV up to $10^{20}$ eV.
- The acceptance is a rapidly falling function at the threshold energy and cannot be compensated by any spectral index.
- The final kernel-MC statistics are based on the $E^{-2}$ compromise.

- The final solution is 'in the order' of an $E^{-3}$ spectrum.
- The resolution is about 0.1 in logarithmical energy units.

To test the three implementations a toy-MC has been set up. In total five pairs of toy kernel-MC and toy-data have been generated. They differ in the resolution assumption. The toy kernel-MC and toy-data have been produced in the following way.

**Kernel-MC:** The kernel MC is generated following an $E^{-2}$ power-law between 10 and 10000 in toy energy-units.

**Toy-data:** The data are generated following an $E^{-3}$ power-law between 10 and 10000 in toy energy-units.

**Detector-response** The detector response is approximated by a Gaussian in $\log(E)$. Like in the Pierre Auger Observatory , the energy is measured over many orders of magnitude. Therefore, the energy resolution is Gaussian in logarithmical energy units. This is emulated by calculating the measured value $y$ of a generated variable $x$ according to the p.d.f.:

$$g(x,y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\log_{10}(x) - \log_{10}(y))^2}{2\sigma^2}\right) .$$

The Gaussian is not cumulative invertible, therefore the measured value has to be evaluated by rejection method. The five data-MC pairs differ in their assumed resolution starting from 0 % (unlimited resolution) up to 40 % in logarithmical energy units.

**Trigger response:** A detector has always a threshold energy. This means that events below a certain energy are not registered, while nearly all events above that energy are triggered. The transition is described by a smooth differential function. A good approximation of this behaviour is given by the incomplete $\Gamma$ function, that is basically a convolution of an exponential decrement with a power-law and corresponding normalisation:

$$P(\alpha, x) = \frac{1}{\Gamma(\alpha)} \int_0^x dt \ \exp(-t) t^{\alpha-1} .$$

Applying this behaviour to our toy-MC means that the acceptance is parameterised as:

$$A(E) = P(\alpha, \log_{10}(E/E_{\min})) .$$

The general trends are plotted for various $\alpha$ in Fig. 7.2. The trigger flag itself has to be evaluated via rejection method. The nice feature of the $\Gamma$ function is that it drops steeper below the trigger threshold, than any power-law.

The toy kernel-MCs have been used to calculate the complete response matrix of the problem. The response matrix and the corresponding toy-data set have been unfolded using the three different algorithms. The input distributions and the unfolded results are given in Fig. 7.3. The final comparison shows the following features:

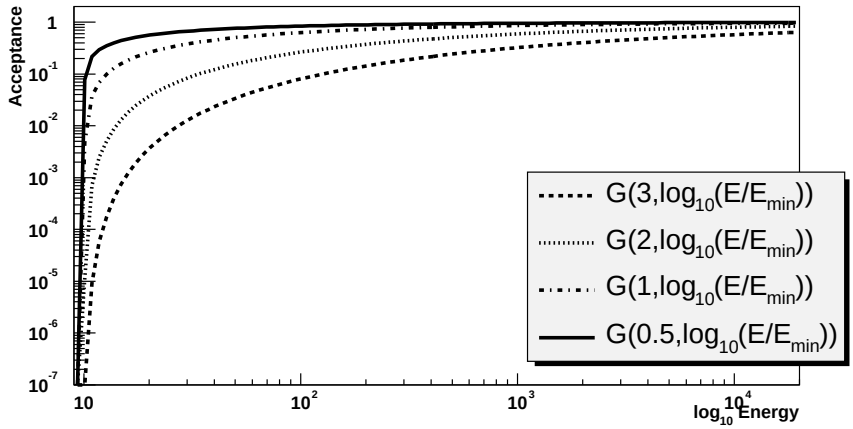



**Figure 7.2:** The incomplete $\Gamma$ function is used to emulated a trigger efficiency. The general trigger efficiency for different parameter of $\alpha$ is demonstrated.

**Convergence:** Although the kernel MC differs strongly from the final solution, all algorithms achieve rather good convergence. This shows that the influence of the exact input spectrum mostly cancels out in the final calculation. (Compare also App. B)

**Deficits of RUN:** Nevertheless, the regularised unfolding using the RUN algorithm shows strongly fluctuating solutions. A feature of all unfolding algorithms is an underestimation of the first and last bins, while the underestimation is extreme for the regularised unfolding. The regularised unfolding works quite robust for scenarios with a good resolution, however, the fluctuations are amplified for scenarios with a bad resolution. Additionally, in some scenarios the convergence has not been achieved.

**Performance of Gold and Bayesian algorithm:** The Gold algorithm and the Bayesian approach turned out to be very robust. Furthermore, their results are in rather good agreement. This is not surprising as they should be equivalent, albeit the technical implementation is different.

**Stability/robustness:** The best performance is achieved using the Gold algorithm. The performance is mostly independent from the spread of the response matrix. The convergence is stable.

**Effect of limited resolution:** The effect of limited resolution becomes rather obvious when regarding the true toy-data spectrum and measured toy-data spectrum for a resolution of 40 %

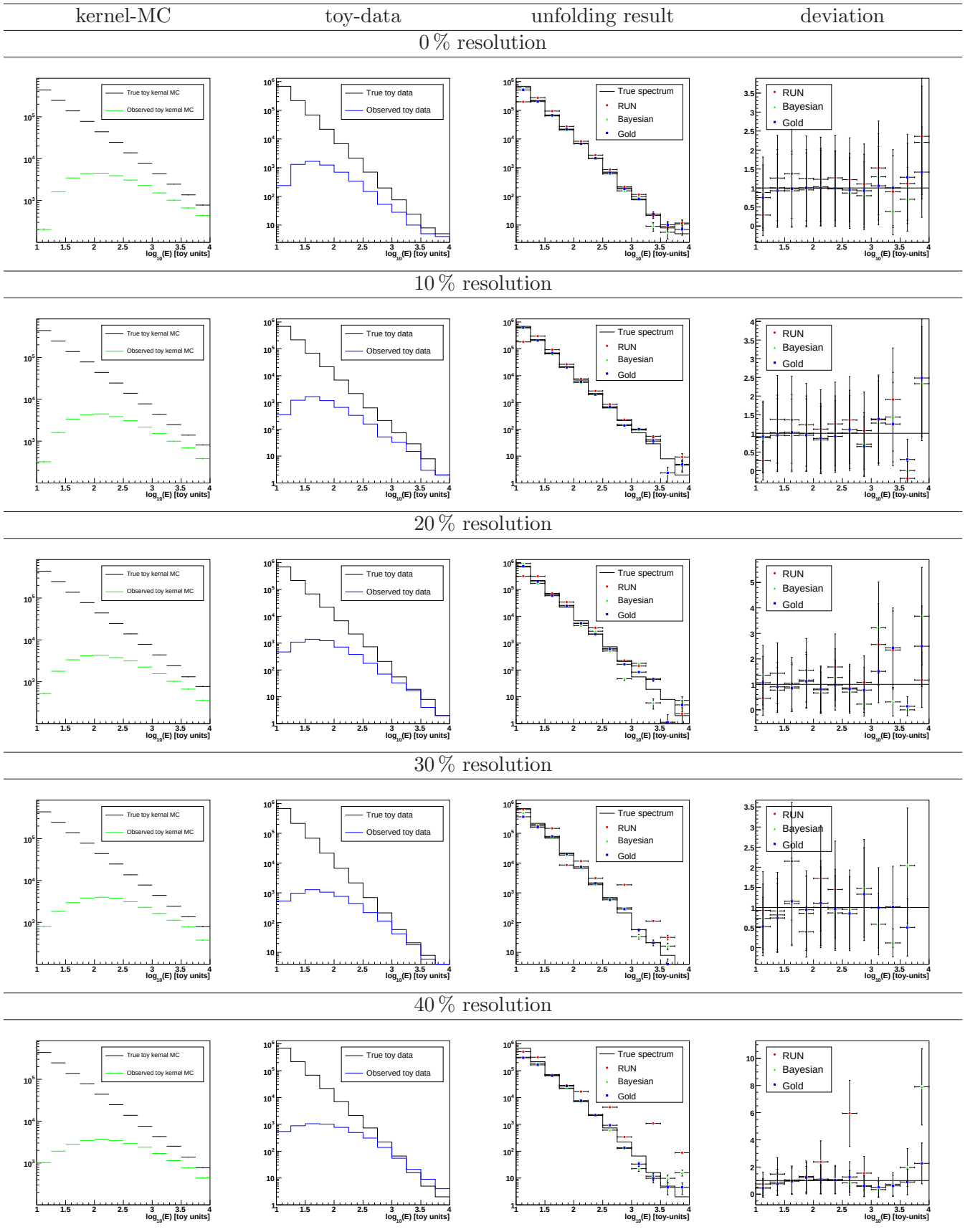



**Figure 7.3:** Comparing the performance of the Gold iterative unfolding algorithm, the Bayesian approach and the regularised unfolding.

in $\log_{10}(E)$. The smearing shuffles so many events to higher energies that even the acceptance cannot compete and the observed data in the high energy bins overshoots the generated amount of events. Simply correcting for limited acceptance would lead to erroneous interpretations.

This motivates using the Gold algorithm for final unfolding of the FD-mono data. In the following, we discuss the concrete implementation of the Gold algorithm in this study and additional refinements.

## 7.4 Implementation of the Gold Algorithm

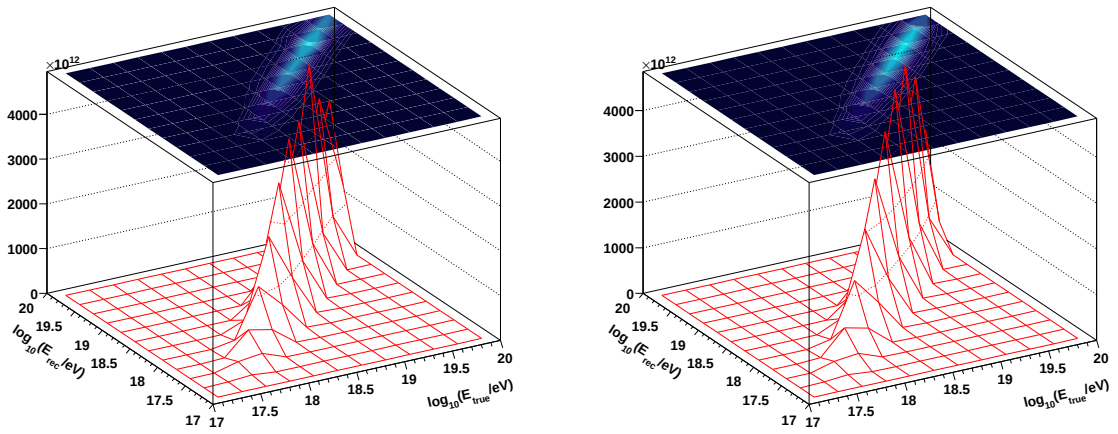
As a consequence of the performance test (see Chap. 7.3) the Gold algorithm is applied to unfold the FD-mono data. Here, we discuss the final implementation of the algorithm. An additional smoothing of the response matrix is implemented, that can be interpreted in terms of regularisation (see Chap. 7.4.1). The error propagation comprises three parts: estimation of the influence of the statistical uncertainties in the measurement on the final solution (*'statistical error'*), systematic uncertainties of $R$ arising from statistical uncertainties of the kernel MC when calculating the response matrix (*'systematic error'*) and systematic bias produced by the algorithm itself (*'bias'*). All three parts are included in the final analysis and discussed (see Chap.7.4.1). It is shown that the influence of smoothing the response-matrix is included in the systematic error. Finally, the difference of the result with and without smoothing is demonstrated. It turns out that the effect is negligible, nevertheless the algorithm converges more robust. Finally, we discuss the cutoff criteria (see Chap. 7.4.2).

### 7.4.1 Regularisation by Parametrisation

In order to compensate the effect, that the structure of the response matrix is amplifying strong fluctuations it is possible to smooth the matrix itself. The nearly singularity is produced by the fact that many next-to-next diagonal bins are zero and only a few can sometimes fluctuate a non zero-entry due to statistics in the kernel-MC. Therefore, we investigate to fit a two-dimensional parametrised function to the response matrix and interpolate a new response matrix by this function. The calculated response matrix is plotted in Fig. 7.4(a). Obviously, the reconstructed energy shapes in logarithmical scale are Gaussian around the true value, as already shown in Chap. 5. This motivates to fit a Gaussian

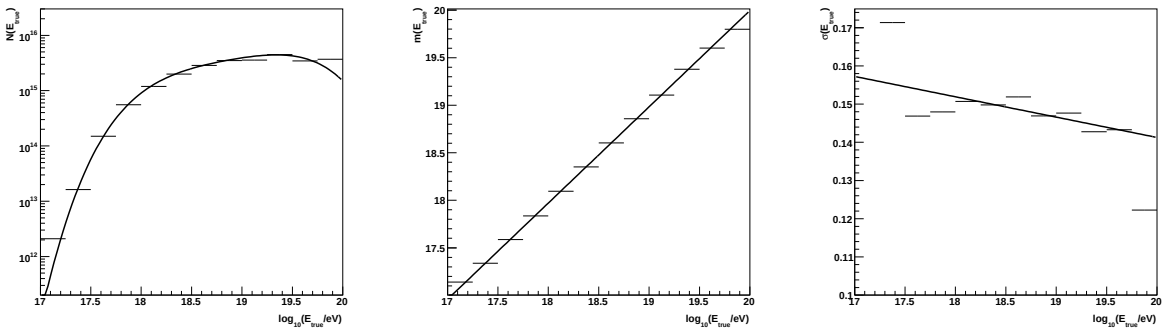
$$f(E_{\mathrm{true},i}, E_{\mathrm{rec}}) = N(E_{\mathrm{true},i}) \exp\left(-\frac{1}{2}\frac{\log_{10}(E_{\mathrm{rec}}) - m(E_{\mathrm{true},i})}{\sigma(E_{\mathrm{true},i})}\right) .$$

The fits for each slice in $E_{\mathrm{true}}$ are listed in App. D. The parameters $N(E_{\mathrm{true,i}})$, $m(E_{\mathrm{true,i}})$, $\sigma(E_{\mathrm{true,i}})$ have been parametrised as a function of $E_{\mathrm{true}}$. Fig. 7.5 shows that the normalisation can be nicely parametrised by using a polynomial of 4th order. The mean and $\sigma$ are parametrised by a linear fit. The obtained two-dimensional parametrisation is plotted in Fig. 7.4(b). The effect of




(a) The response matrix (b) The parametrisation

**Figure 7.4:** The response matrix calculated from the kernel MC and the applied parametrisation.




(a) $N(E_{\mathrm{true,i}})$ (b) $m(E_{\mathrm{true,i}})$ (c) $\sigma(E_{\mathrm{true,i}})$

**Figure 7.5:** Behaviour of the Gaussian fit parameters as function of $E_{\mathrm{true}}$, and fitted parametrisation. The fits for each slice in $E_{\mathrm{true}}$ are listed in App. D.

this parametrisation becomes visible in Fig. 7.6. The main diagonal and side diagonal entries are mostly unchanged, while the next-to-next entries are no longer fluctuating around zero, but show a smooth behaviour. In that sense, the parametrisation can be considered as a kind of additional regularisation of the problem. The uncertainties due to the fitted function have to be taken into

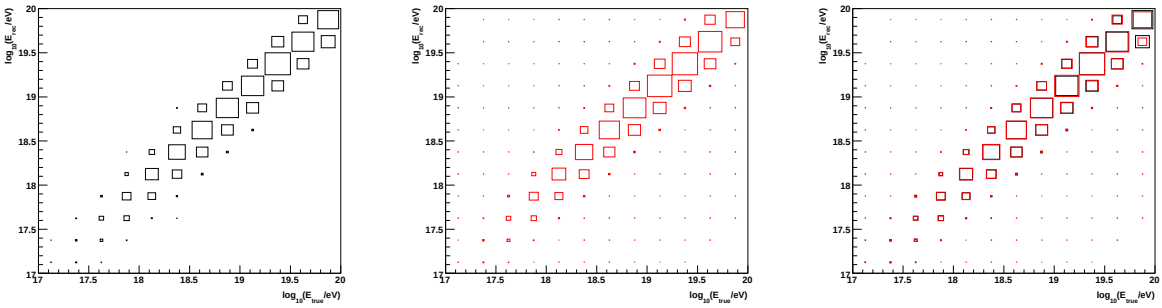
account as a systematic error. The final analysis is performed with and without parametrisation and it is shown that the deviation of both solutions agree within systematics. Fig. 7.13 shows that the effect of the parametrisation on the final result is negligible and correctly treated within the error propagation, while the robustness of the Gold convergence and reliability increase.

### Error Propagation

Three kinds of errors have to be taken into account for the Gold iteration and are discussed in detail now. We face all three categories of errors by implementing a frequentist approach.

**Statistical Uncertainties:** The measured data $\nu_i$ have uncertainties arising from the exposure (limited data taking time). The statistical fluctuations have to be taken into account and the final result has to represent the statistical error of the measurement. The basic idea is to simulate the statistical uncertainties by generating in a frequentist approach a set of MC experiments and unfolding them. We consider the bin-content of the measured data to behave in a Poissonian way. This means we would expect that we can generate bin-by-bin a new bin content that is Poissonian distributed around the measured value. A procedure is set up to do this several times to generate a sufficiently large statistics of $K$ MC-experiments ($K \sim 200$). By doing so we can consider the complete unfolding procedure as a black box and unfold all $K$ MC-experiments. The schematic procedure is visualised in Fig. 7.7. All $K$ unfolded spectra are compared and the variance of each bin content is considered as a measure for the statistical uncertainty of the final result,

$$\sigma^2_{stat,i} := \frac{1}{K} \sum_{k=1}^{K} (\bar{\mu}_i - \mu_{i,k})^2 \ .$$



(a) Response matrix. (b) Parametrisation. (c) Comparison between response matrix and parametrisation.

**Figure 7.6:** Original response matrix and parametrisation in comparison.

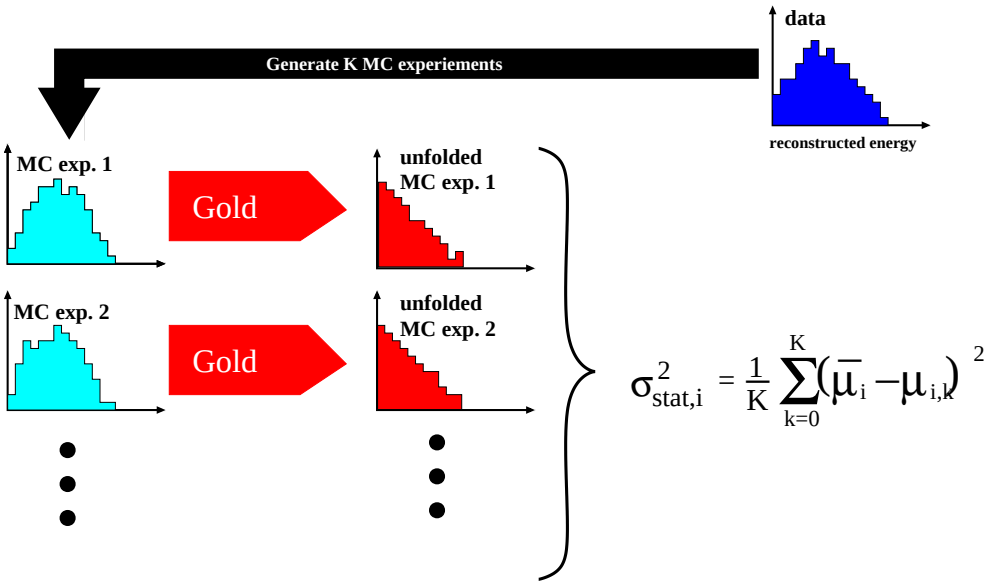



**Figure 7.7:** Schematic procedure of estimating the statistical uncertainty of the unfolded spectrum.

**Systematic bias:** It is possible that the whole iteration procedure has a systematic bias, which manifests itself in the following way: When forward-folding the derived solution, the estimated data may be shifted with respect to the true measured data. This shift is refered to as systematic bias of the algorithm itself.
As the exact solution of the unfolding problem is unknown we cannot directly compare the unfolded distribution to the exact solution to estimate the bias of the procedure. On the other hand, we expect a possible bias being an intrinsic systematic uncertainty from the unfolding procedure itself. Therefore, we can construct again a frequentist approach to estimate the bias. A sketch of the procedure is given in Fig. 7.8. The idea is again to produce significantly large statistics of MC-experiments with a random generator. These MC-experiments spread around the actual one by a Poissonian hypothesis. Unfolding these experiments allows us to find a mean unfolded solution to the experimental data. As described above the statistical error can be obtained by the variance. We can consider this ensemble of unfolded solutions as an exact spectrum and fold it forward again to generate a corresponding ensemble of measured data sets. By re-unfolding these data sets, a set of re-unfolded spectra is obtained. This allows us to compare the unfolded spectra with the re-unfolded spectra and to estimate the systematic bias of the unfolding procedure itself, via:

$$b_i = \frac{1}{K}\sum_{k=1}^{k}(\bar{\mu}_i - \mu_{i,k}) \; .$$

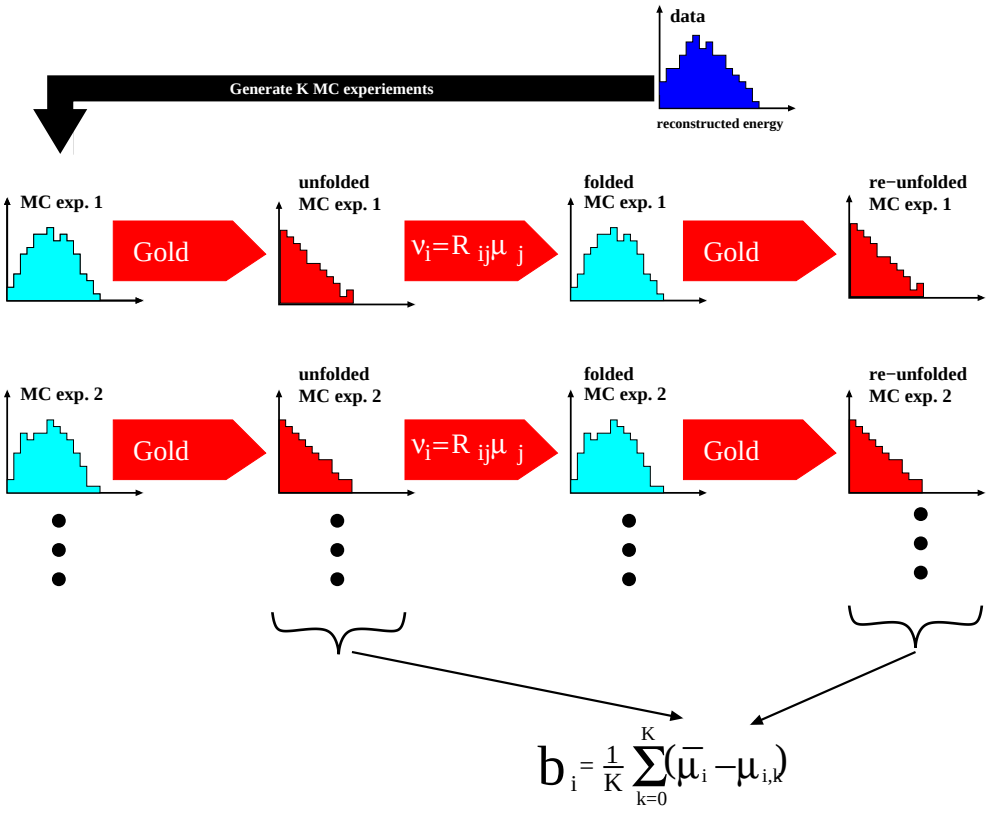



**Figure 7.8:** Schematic procedure of estimating the systematic bias of the unfolding algorithm.

Here, $\bar{\mu}_i$ is the mean of the unfolded MC-experiments, so to say the generated exact solution, and $\mu_{i,k}$ the bin-content in the $k$-th re-unfolded spectrum.

**Systematic Uncertainties:** Although, the kernel-MC is generated to cover the phase space very efficiently it is impossible to obtain a perfect response matrix. While the diagonal entries are always obtainable with high enough statistics the elements far off the diagonal are usually fluctuating around zero. Even with high statistics it is hard to determine e.g., the probability to mis-reconstruct a $10^{17}$ eV shower to $10^{20}$ eV. This happens rather unlikely. Therefore, these entries show large uncertainties. On the other hand, many events get reconstructed with energies between $10^{17}$ eV and $10^{18}$ eV, therefore a small fluctuation in the far off-diagonal elements in the response matrix can modulate strongly the final result. These uncertainties have to be taken into account. Again, a frequentist approach is chosen. Firstly, we describe the procedure performed to the parametrisation scenario. In this case the response matrix is fitted by a 2-dimensional function. The fit procedure determines a set of parameters and their uncertainties. The fit takes the uncertainties of each bin of the response matrix into account.

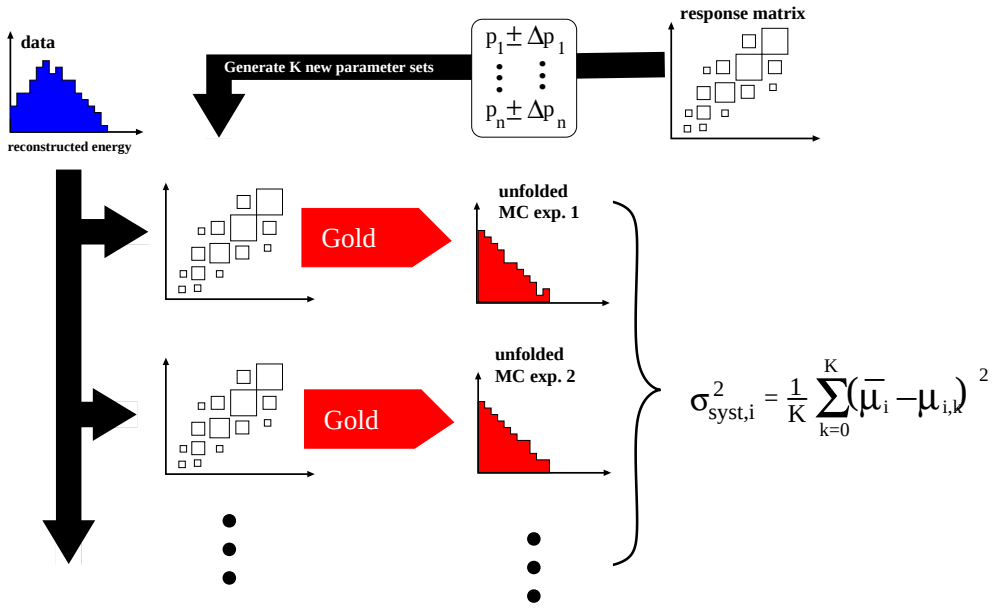


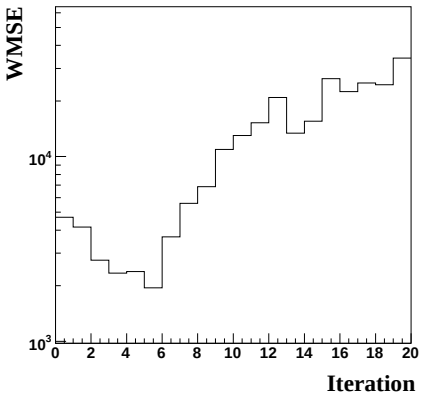

**Figure 7.9:** Schematic procedure of estimating the systematics due to uncertainties in the response matrix in case of parametrisation. The case without parametrisation is performed similarly by generating new response matrices by Poissonian bin modulation.

Thus, also each parameter and its error represents the uncertainty of the response matrix in a correct manner. New parameter sets within the given errors can be generated. Unfortunately, the distribution of the p.d.f. of the parameters within their errors is unknown. In case of a $\chi^2$-fit all parameters that are on the manifold between $\chi^2 = 1$ and $\chi^2 = 2$ have to be included. A very conservative approximation is in any case to assume each parameter uncorrelated to the others and being distributed in a Gaussian way with $\sigma_{\mathrm{Gauss},i} = \Delta p_i$, where $\Delta p_i$ is the uncertainty of the $i$-th parameter. This provides a rule to generate new parameter sets and calculate from them new response matrices. The observed data can be unfolded with each new generated response matrix, as schematically drawn in Fig. 7.9. The bin-wise variance of the unfolded results gives a measure for the systematic uncertainties arising from the kernel itself.

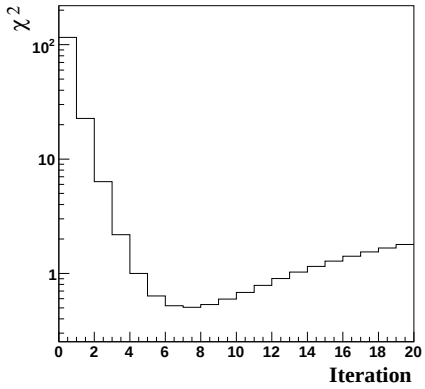
If a parametrisation is not used, the generation of the response matrices can be done directly by bin-wise modulation of the bin-content within its errors. The bin-content does not necessarily follow a Poissonian, as it is the result of many divisions. The p.d.f. of the bin content within its error is not trivial to find, but a Gaussian assumption of the p.d.f. is rather conservative.

### 7.4.2 The Iteration-Stopping Criterium

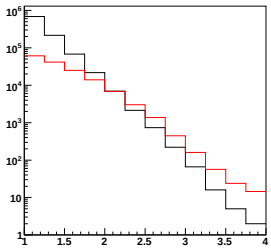
In case of iterative algorithms it is important to define a clear criterium when to stop the iteration. As already mentioned the iteration procedure is converging towards the exact solution of Eqn. 7.4. This solution is not the demanded one due to strong fluctuations (see Chap. 7.1.5). Therefore, the general behaviour of the iteration can be described as follows. In the beginning the result deviates mostly from the demanded solution due to systematic bias. The more iterations are performed the less the systematic deviation from the expected spectrum becomes. On the other hand, in the beginning the statistical error is rather small and the more iterations are performed the more dominant the statistical error becomes due to the fact that the unfolded spectrum becomes more and more sensitive to statistical fluctuations of the measured distribution. Obviously, $\chi^2$ appears



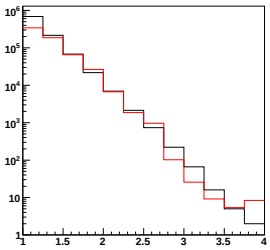

(a) WMSE for each iteration step.



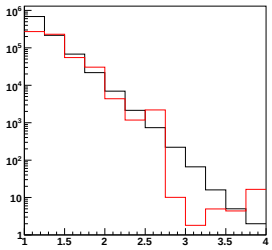

(b) $\chi^2$ for each iteration step.



(c) True and unfolded spectrum after one iteration.



(d) True and unfolded spectrum after six iterations.



(e) True and unfolded spectrum after 20 iterations.

**Figure 7.10:** $\chi^2$ and WMSE evolution with the iteration. The optimum depth to stop iteration is reached at $k = 6$. Below are given unfolded (red) and true (black) toy-data spectrum at the beginning in the optimum and for $k = 20$.

to be a good choice to judge when the optimum is reached.

$$\chi^2_k = \frac{1}{N}\sum_{j=1}^{N}\left(\frac{\nu_j - \sum_i R_{ij}\mu_{i,k}}{\sigma_j^2}\right)^2 .$$

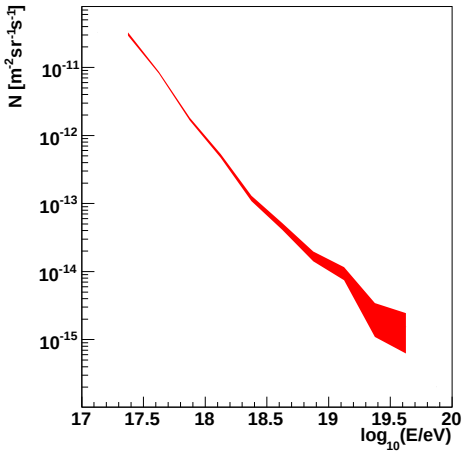
$\chi^2_k$ is here the $\chi^2$ at the $k$-th iteration step and $\mu_{i,k}$ the solution estimate at the $k$-th iteration. As shown above the Gold iteration is equivalent to a $\chi^2$ minimisation. $\chi^2_k$ can in this case be considered rather as a measure how good the measured data are reproduced by the estimated solution, and by this it strongly depends on the estimated statistical uncertainties. On the other hand, one is more interested in a measure that describes the quality and structure of the found solution. A measure, which is often used, is the *'weighted mean squared error' (WMSE)*. It is defined as:

$$\mathrm{WMSE}_k = \frac{1}{N}\sum_{i=1}^{N}\frac{\sigma^2_{stat,i,k} + b^2_{i,k}}{\mu_{i,k}} .$$

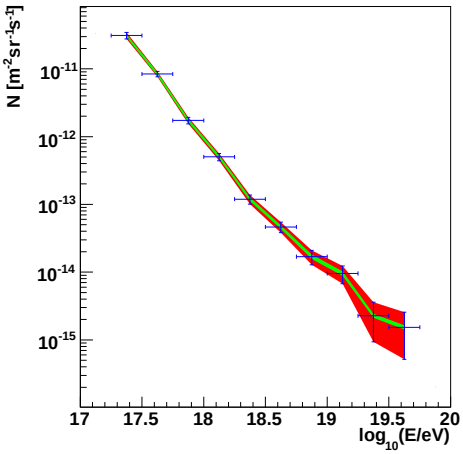
Here, $\sigma_{stat,i,k}$ is the statistical error at the $k$-th iteration and it can be estimated using the frequentistic approach as described in Chap. 7.4.1. $b_{i,k}$ describes the systematic bias of the unfolding iteration at the $k$-th iteration and is also available by the frequentist procedure. It is obvious that this measure should have a clear minimum defining the optimum solution. In the beginning the bias dominates the WMSE, the more iterations are performed the smaller the bias becomes and statistical uncertainties start to dominate. Hence, the minimum of the WMSE parameter seems to be a suitable compromise between both kind of uncertainties, the systematic bias and the statistical error.
Sometimes the *mean weighted systematic shift* is also used as parameter

$$\chi^2_{b,k} := \frac{1}{N}\sum_{i=1}^{N}\left(\frac{b_{i,k}}{\sigma_{b,i,k}}\right)^2 .$$

Here, $b_{i,k}$ is the systematic bias at the $k$-th iteration, while $\sigma_{b,i,k}$ describes the corresponding variance of the $b_{i,k}$. The more iterations are performed the smaller becomes the bias $b_{i,k}$. On the other hand, the statistical uncertainties increase. The idea is to stop iterating when $\chi^2_{b,k}$ is around 1. That is, when statistical uncertainties and systematic bias equalise. At that point it does not make sense to continue iteration with the aim to reduce the systematic bias, as the statistical uncertainties becomes larger than the gain by the decreased bias.
Both criteria $\chi^2_{b,k}$ and WMSE are equivalent and further on we use the WMSE criterium. Fig. 7.10 demonstrates the performance of the stopping criteria for the presented toy-MC. As example the unfolding of the $E^{-3}$ toy-data spectrum with a resolution of 40 % in logarithmical energy units is shown. Fig. 7.10(a) shows the evolution of the WMSE values versus iteration depth, while Fig. 7.10(b) shows the same evolution for the $\chi^2$ values. The optimum iteration depth is according to the WMSE criteria $k_{toy} = 6$. Indeed, the $\chi^2$-distribution is around 1 for that depth. In order to demonstrate the different behaviour of bias and statistical uncertainties Fig. 7.10(c), 7.10(d), 7.10(e)




(a) Unfolded integral spectrum showing statistical uncertainties.




(b) Unfolded integral spectrum showing statistical and systematic uncertainties.

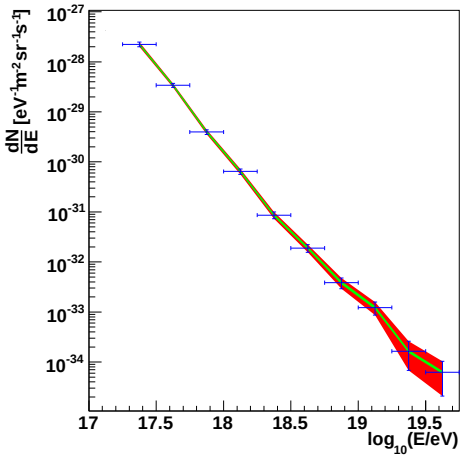
**Figure 7.11:** Integrated spectra obtained by the unfolding procedure.

show the unfolded result and true data spectrum in the beginning. The unfolded spectrum significantly deviates from the expectation. At low energies it is underestimated, while for higher energies there is an overestimation. The situation changes with the iteration depth. After 20 iteration no systematic bias is visible, but fluctuations due to the sensitivity of the reached solution estimation to statistical uncertainties of the measurement. These fluctuations are responsible for the dominant statistical error. The solution at the optimum iteration depth seems to be reasonable.

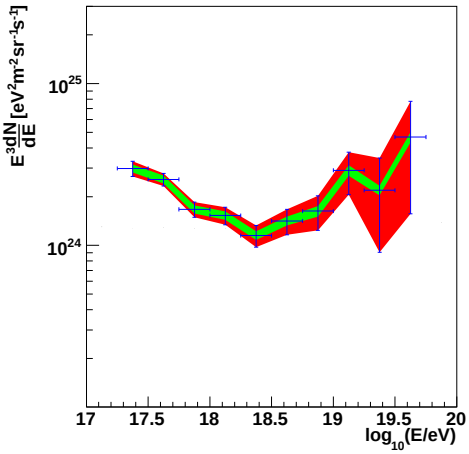
## 7.5 Data Analysis

The Gold algorithm as described above is applied to FD-mono data. As a stopping criterium for the iteration we use the WMSE-criterium (see Chap. 7.4.2). The optimum solution is already reached after five iterations. This is quite normal for a one-dimensional problem. The same algorithm applied to a two-dimensional problem like in the KASCADE unfolding analysis [186] needs about 70 iterations, that is reasonable as the complexity scales with the square of the dimension. The resulting integrated energy spectrum is shown in Fig. 7.11. The statistical uncertainties are labelled by a red belt while the systematic uncertainties are superimposed by a green belt. While for small energies the systematic uncertainties dominate, the statistical ones become dominant for high energies. Obviously, already at this stage obviously the spectrum seems to have features. The differential spectrum Fig. 7.12 is obtained by bin-wise dividing through the bin-width.

Usually, the spectrum is scaled by $E^3$ to make the feature more visible. We use this representation also to demonstrate the influence of the parametrisation to our final result. Fig. 7.13 shows

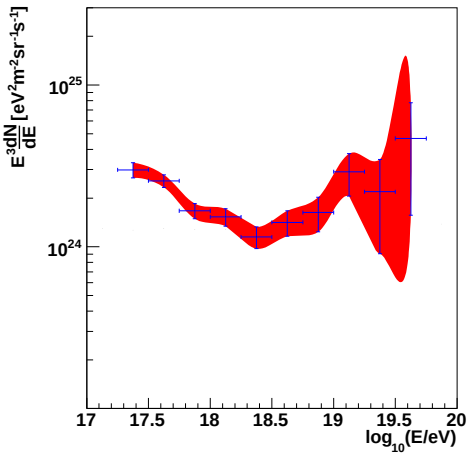



(a) Differential energy spectrum including statistical and systematic uncertainties.

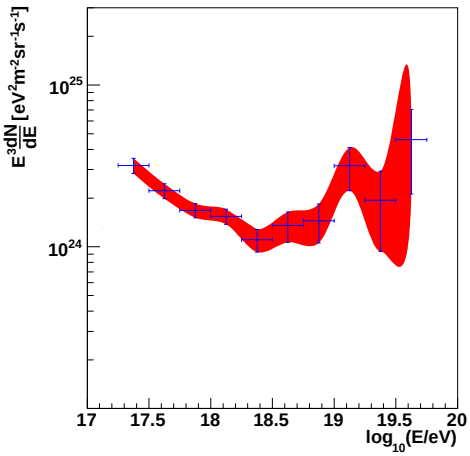



(b) The flux weighted with $E^3$ including statistical and systematic uncertainties.

**Figure 7.12:** Differential unfolded flux and spectrum weighted with $E^3$.




(a) Unfolded spectrum using parametrisation.




(b) Unfolded spectrum using the response matrix directly.

**Figure 7.13:** Effect of parametrisation of the response matrix.

the unfolded energy spectrum using a parametrisation and using directly the response matrix. The features and the flux normalisation is completely the same, both spectra differ only little in the steepening between $10^{18}$ eV and $10^{18.5}$ eV. Within their uncertainties they agree.
Fig. 7.12(b) shows the final energy spectrum weighted by $E^3$ including both systematics and statistical uncertainties. Above $10^{19.5}$ eV statistical uncertainties are large and it is not possible to make quantitative statements on the spectral shape in the region of possible cutoff scenarios. Nevertheless, some features are visible:

- The normalisation is in agreement with results from HiRes [176], KASCADE-Grande [188] and the SD-hybrid combined result from the Pierre Auger Observatory [201]. This is discussed in detail in Chap. 7.5.1.
- There is an ankle-structure at $\sim 10^{18.5}$ eV visible.
- There is no clear statement about a cutoff possible, yet. The data indicate a drop, but the statistical uncertainties are too large to draw a firm conclusion.

The final result shows the expected link between established experimental knowledge on the energy distribution of cosmic rays and the UHECR-data coming from the SD. The question of a possible cutoff will be solved soon by the Pierre Auger Observatory with more data becoming available. The presented SD spectrum is in agreement with the presented FD-mono spectrum and benefits from the validation using the fluorescence data as link to low energies.

### 7.5.1 Comparison to Other Experimental Data

Fig. 7.14 shows the final spectrum of unfolded FD-data in comparison with other experimental results. The combined data-points indicate the energy spectrum derived by the SD- and hybrid-data of the Pierre Auger Observatory [201]. The HiRes data plotted here are based on the recent results using the fluorescence yield parametrisation by Nagano [176]. This makes the data comparable with the FD data of the Pierre Auger Observatory as this analysis is using also the Nagano parametrisation. Recent results from KASCADE-Grande have extended the energy measurement to $10^{17.5}$ eV by unfolding the $\mu$-component detected in their array [188]. The unfolded FD-mono spectrum agrees with the KASCADE-Grande data within 5 % in the energy (compare App. C). Obviously, all high energy experimental data indicate an ankle position at around $10^{18.5}$ eV. The unfolded FD-data from the Pierre Auger Observatory drop harder towards the ankle. The discrepancy to HiRes data is in the order of 7 % as shown in Fig. 7.16. Here it is assumed that the energy in the HiRes experiment is overestimated by 7 %. Scaling the HiRes data agreement can be achieved. An effect of this order does not indicate a discrepancy as systematics can still be different between both experiments. In addition, Fig. 7.16 shows the effect of unfolding. The unfolding covers both resolution effects as well as a possible bias in the energy reconstruction. Therefore, one does not expect a constant shift between unfolded and non-unfolded FD-mono spectrum.
Fig. 7.15 investigates the effect due to different parametrisations of the fluorescence yield. While the first results of HiRes have been based on a parametrisation by Kakimoto[37], the recent result are

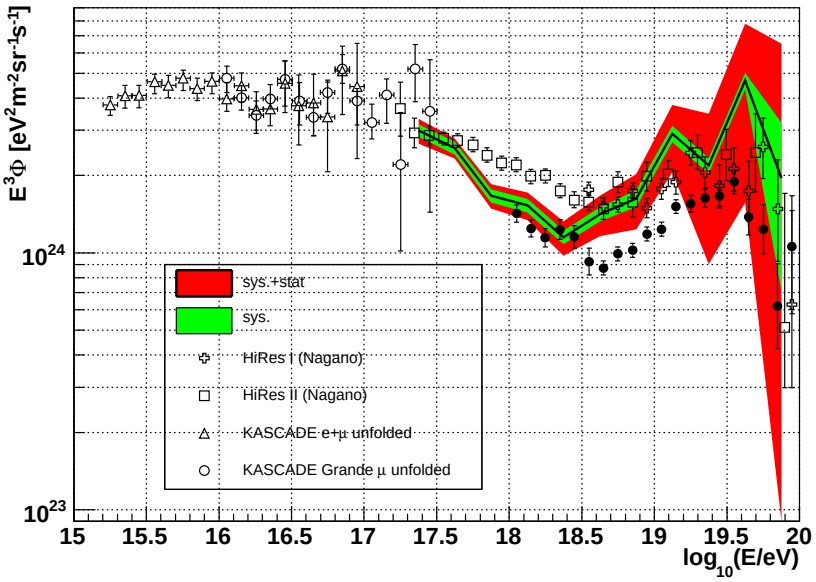



**Figure 7.14:** The final unfolded FD-spectrum compared to data from other experiments [176, 188, 201].

based on NAGANO parametrisation [176] like the analysed data from the Pierre Auger Observatory . The difference is quite small.
For the low energy regime of the spectrum, it is remarkable to see the influence of the primary composition. Fig. 7.17 shows the unfolded spectrum using a mixed composition in comparison to the unfolded spectra in case the kernel-MC is derived from pure proton MC or pure iron MC. The most significant influence is visible at energies below $10^{18}$ eV.

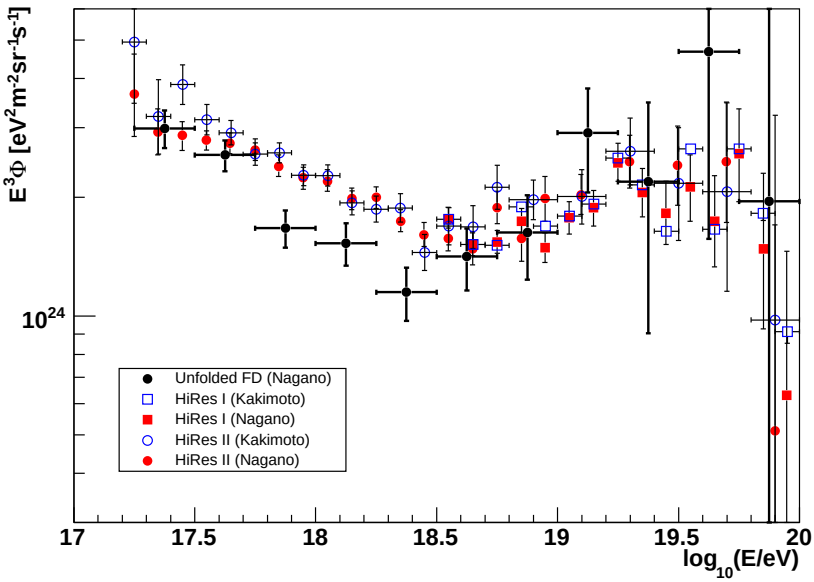



**Figure 7.15:** The influence of the fluorescence yield. HiRes data using Kakimoto [37] and Nagano [176] parametrisation compared to unfolded FD data using Nagano parametrisation.

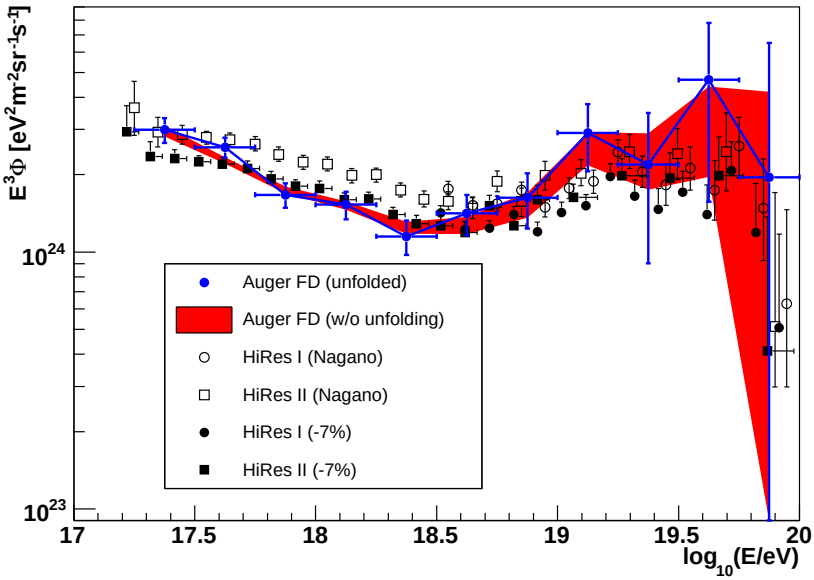



**Figure 7.16:** The effect of unfolding. Unfolded and non-unfolded FD spectrum compared to HiRes data[176]. The difference between results from the HiRes experiment and the analysed FD data from the Pierre Auger Observatory is about 7 % in the energy.

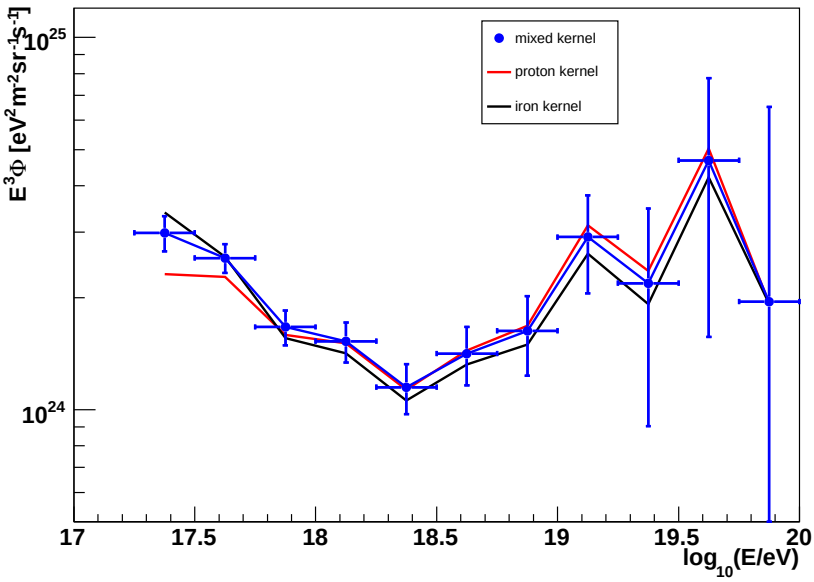



**Figure 7.17:** The influence of the composition on the final result. Unfolded spectra using pure proton kernel MC (red), pure iron kernel MC (black) and a mixture kernel MC (blue).

# Chapter 8

# Summary and outlook

*Count the stars in the sky. Measure the waters of the oceans with a teaspoon. Number the grains of sand on the sea shore.*
*Impossible, you say?*

Philip Glass
*Einstein on the Beach*

The reconstruction of the energy spectrum of UHECR using only data from the fluorescence detectors of the Pierre Auger Observatory is a challenging task: Difficulties arise from the complexity of the physical processes, which take place in the shower development. Then, the propagation of light from the shower to the telescope and the detection in the telescopes leads to significant fluctuations. In addition the detector is highly inhomogeneous, as the experiment is still under construction. Finally, the detector medium, is subject to changes due to varying atmospheric conditions. For the measurement of the energy spectrum, all uncertainties have to be taken into account carefully in a statistically correct way. The presented work has investigated these problems and yields the following results:

- The fluorescence signal is detected by 440 PMT pixel cameras. The robustness and stability of the PMT response is essential for the required data-quality level. Quality- and classification-tests have been performed in Wuppertal and the achieved results underline the high level of data-quality. ([25, 26])

- The FD-data are taken in shifters mode. In the course of the current work contributions to the development of a monitoring software have been accomplished. These accomplishments enable the shifter-crew to supervise the current status of the ongoing data-taking and give alarms in case of malfunctions. Moreover, it allows longterm monitoring of the detector stability. The presented sketch of the current implementation emphasises the reliability of

the data processed during the shift periods.

- Compared to hybrid reconstruction, FD-mono reconstruction suffers from a poorer geometrical resolution. In the context of the presented work the effects of this deficit on energy reconstruction are quantified. A dedicated validation and performance study has been carried out. The full MC simulation and reconstruction chain has been set up on the Wuppertal ALiCENext cluster. For this analysis, a MySQL database has been set up and linked to the mass-production. The present work describes the implemented production scheme. Detailed investigations of reconstructed and true parameters allow to determine the resolution of the energy reconstruction. A comparison to a hybrid-like study has been performed. FD-mono energy resolution is $\sim 25\,\%$ on a linear scale compared to $10\,\%$ for hybrid events. ([78, 30]) This was used as a basic motivation to start an unfolding analysis.

- To validate the FD-simulation itself a comparison between data and MC has been performed. The method of importance sampling is implemented to investigate the parameter space efficiently. Events have been re-weighted according to a physical flux assumption. In this context it has been necessary to develop a method to quantify the telescope-wise uptime. Background data turn out to be a reliable measure to decide on the detector status. At final cut level, data and MC proved to be in good agreement, both in shape and normalisation. ([75])

- In the context of the data-MC comparison the sensitivity of the spectral behaviour and composition becomes already visible at detector level. A pure iron or a pure proton composition hypothesis is disfavoured compared to recent simulation predictions. It has been shown that a featureless power-law spectrum does not describe data well. ([76])

- As a further cross-check of the reliability of simulations a study of multi-fold topologies has been performed. Simulations have shown that most triple events are hybrid events and most diagonal stereo events are triple. The 2-dimensional exposure for a stereo and 3-fold data-set was calculated and compared to real data. Although statistics are too poor to quantify the agreement, the comparison shows good qualitative agreement. Recently, the first 4-fold event was detected in a region that was predicted by this work. ([75])

- The non-negligible uncertainties in FD-mono reconstructions have to be folded into the analysis in a correct statistical manner. Therefore, the analysis is carried out in terms of a classic Fredholm integral equation. The deconvolution problem has been introduced and techniques have been described to unfold the final energy spectrum. The three most suitable algorithms have been compared using a toy MC. The Gold algorithm turns out to be the most robust algorithm.

- Iteration-stopping criteria have been developed for the Gold unfolding. A frequentist approach was implemented to reliably estimate the bias, the statistical uncertainties and systematic error of the procedure.

- The final energy spectrum is unfolded. The objective set in the preface was achieved to find a link between the spectra from the SD of the Pierre Auger Observatory at ultra-high energies and data from HiRes and KASCADE-Grande . The absolute energy scale of the measured energy spectrum agrees with the KASCADE-Grande experiment within 5 % at $10^{17.5}$ eV. The agreement with the HiRes experiment below $10^{19}$ eV is 7 %. The final spectrum clearly shows a kink at $10^{18.5}$ eV that can be interpreted as ankle. Although no quantitative statements about cutoff scenarios are possible at this level, data indicate a drop, but within uncertainties they are still in agreement with no cutoff hypotheses.

Beyond all achievements the presented work has opened up the potential for further analyses:

- The sensitivity of the data-MC comparison at detector level has demonstrated the feasibility of a two dimensional unfolding of energy and composition.
- With the detector reaching its final configuration the FD-mono data can gain significantly from a fixed setting and offer the possibility to analyse the energy regime below $10^{18}$ eV, in contrast to SD and hybrid.
- With increasing exposure it might even be possible to draw conclusions on cutoff scenarios that could be used as independent cross-checks on hybrid and SD results.

Although the FD-mono data have a degraded resolution than hybrid-data, they can still compete with hybrid analyses. This thesis has shown that currently the mono exposure exceeds the hybrid exposure by factor three. Furthermore, it has been shown that mono-data are the only key to fill the missing link to results from other experiments in the low energy part. It is a thrilling time to see in which form the energy spectrum will continue.

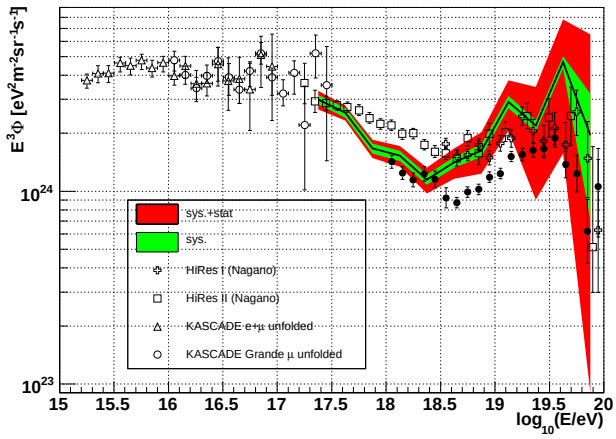
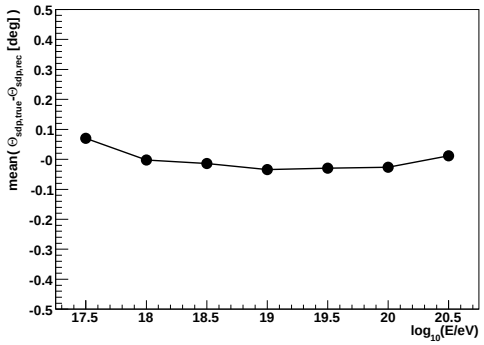



**Figure 8.1:** The final unfolded energy spectrum compared to data from other experiments

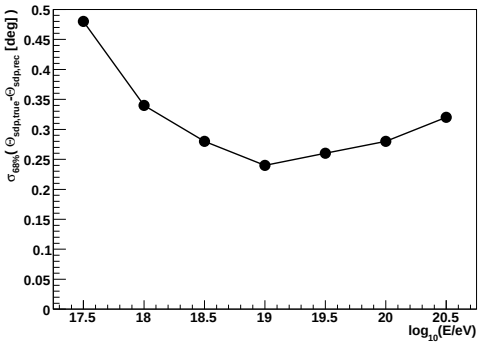
# Appendix A

# Simulation-Reconstruction Performance

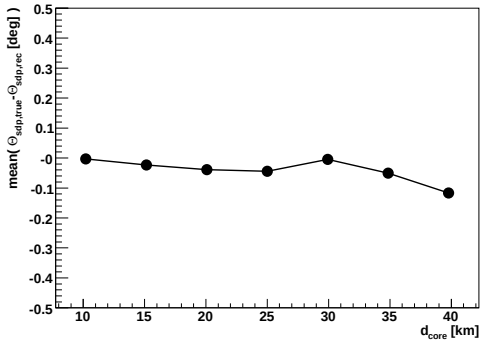
In the context of the validation study described in Chap. 5 the resolution of various observables has been investigated as function of energy, distance and zenith angle. For sake of completeness the following figures show the resolution of many observables connected to geometry reconstruction.



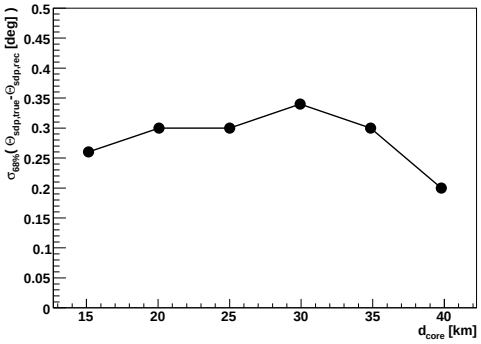

(a) Mean vs. energy.



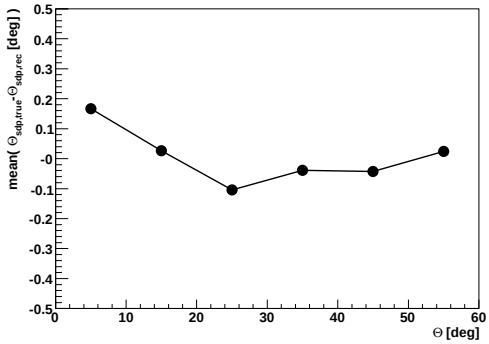

(b) Resolution vs. energy.



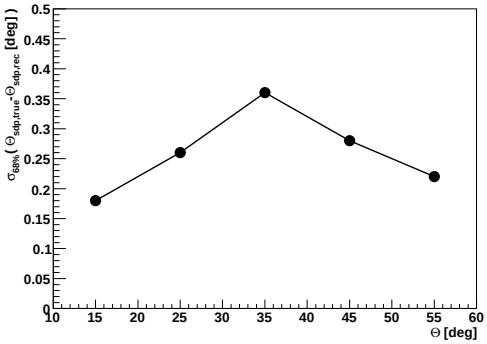

(c) Mean vs. distance.




(d) Resolution vs. distance.

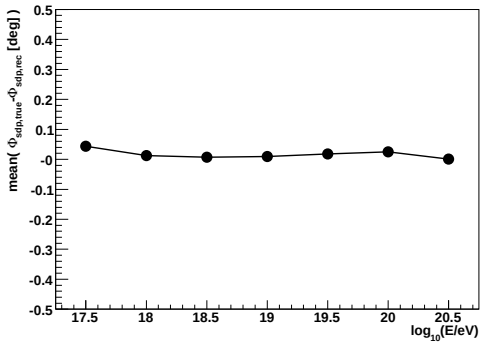



(e) Mean vs. zenith.

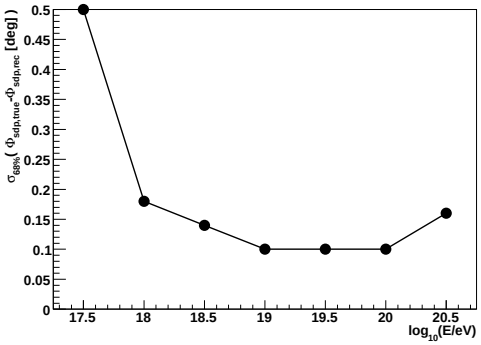



(f) Resolution vs. zenith.

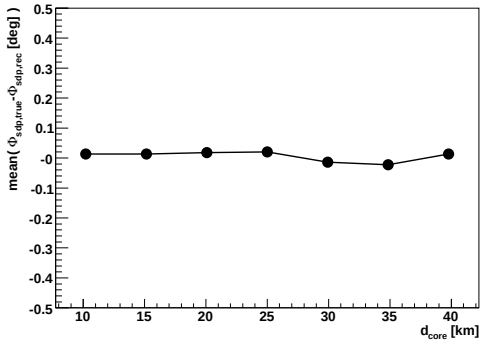
**Figure A.1:** Resolution of $\theta_{SDP}$.



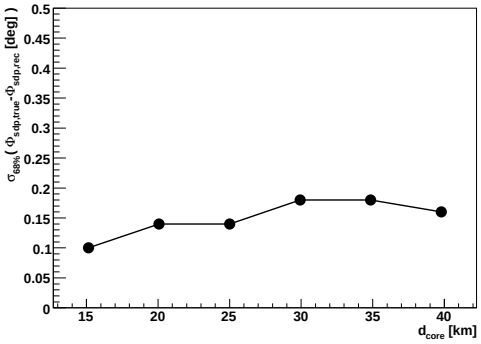

(a) Mean vs. energy.



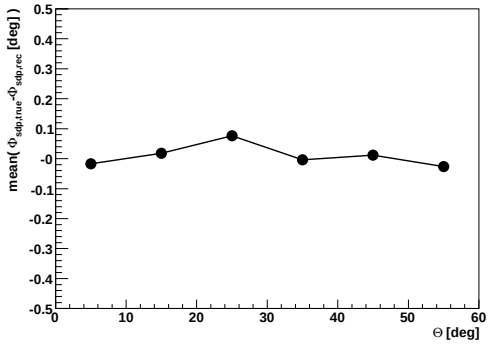

(b) Resolution vs. energy.



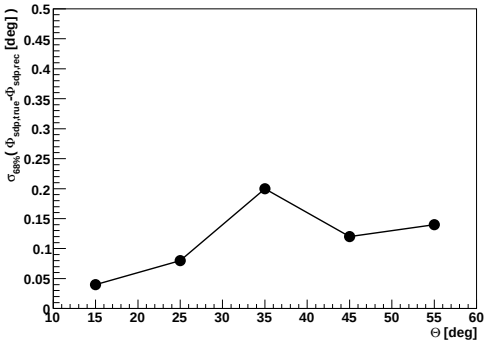

(c) Mean vs. distance.




(d) Resolution vs. distance.

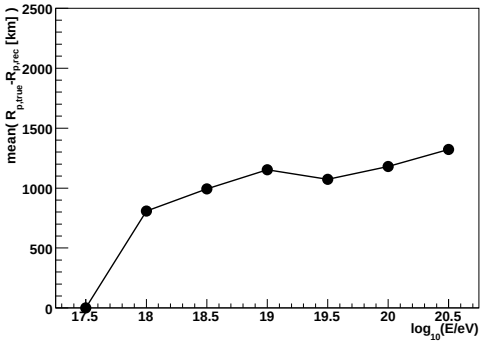



(e) Mean vs. zenith.

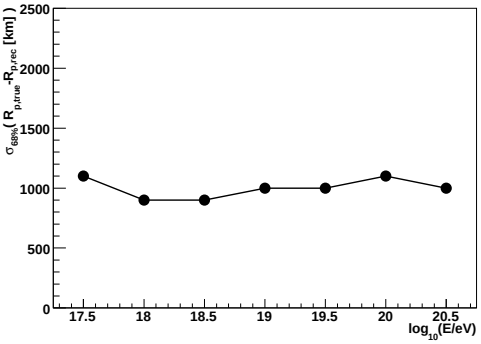



(f) Resolution vs. zenith.

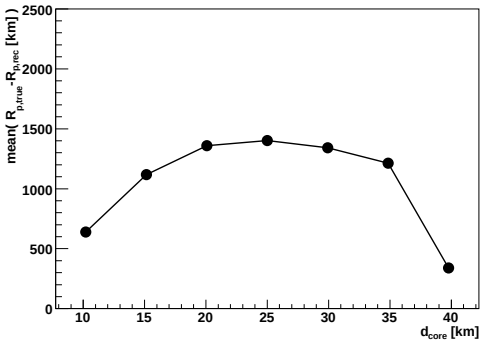
**Figure A.2:** Resolution of $\phi_{SDP}$.



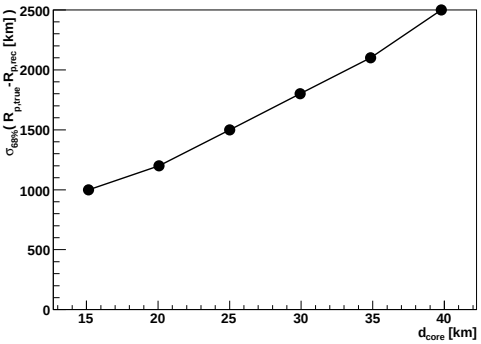

(a) Mean vs. energy.



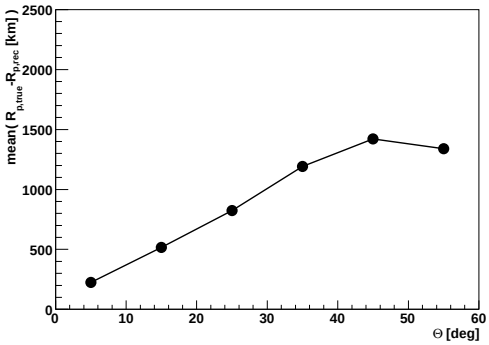

(b) Resolution vs. energy.



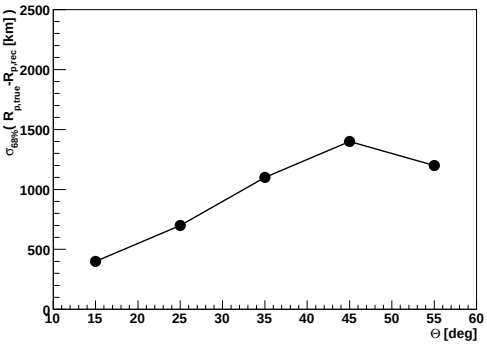

(c) Mean vs. distance.




(d) Resolution vs. distance.

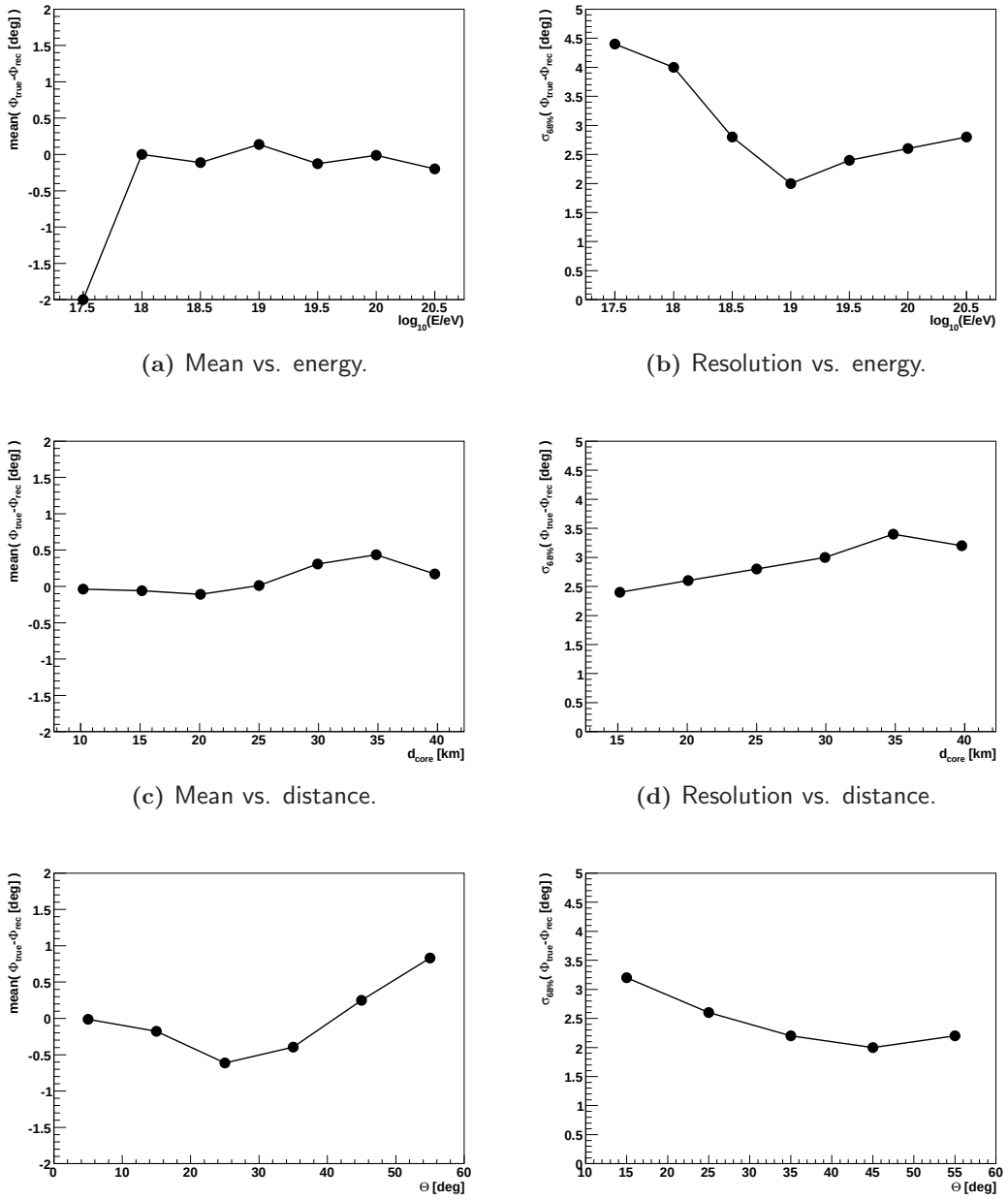



(e) Mean vs. zenith.




(f) Resolution vs. zenith.

**Figure A.3:** Resolution of $R_p$.




(a) Mean vs. energy.

(b) Resolution vs. energy.

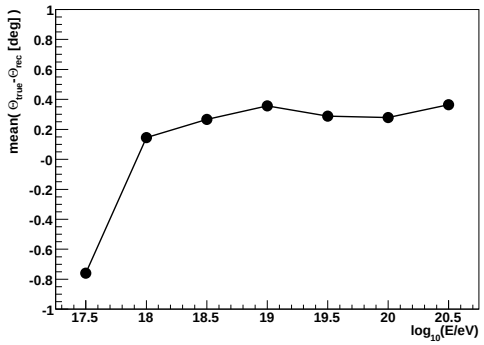
(c) Mean vs. distance.

(d) Resolution vs. distance.

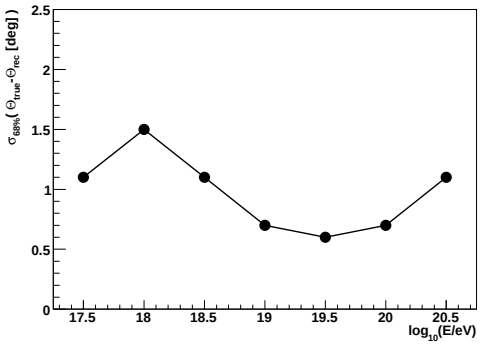
(e) Mean vs. zenith.

(f) Resolution vs. zenith.

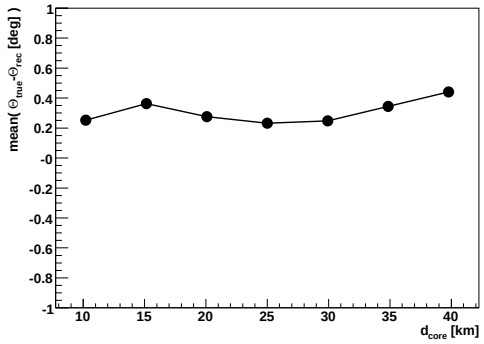
**Figure A.4:** Resolution of azimuth angle.



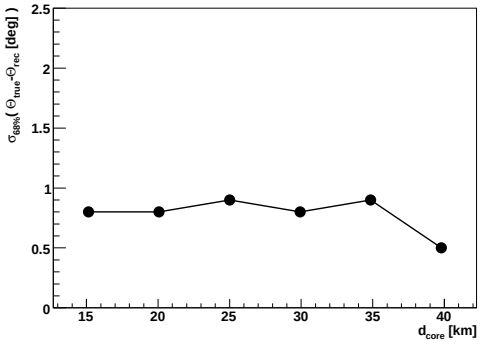

(a) Mean vs. energy.



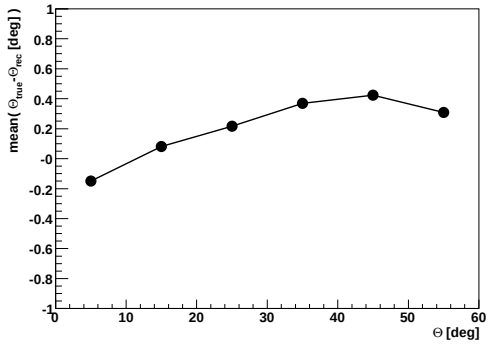

(b) Resolution vs. energy.



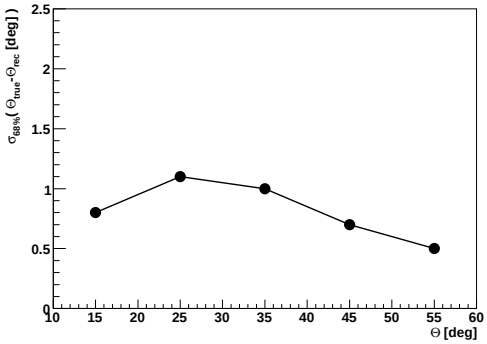

(c) Mean vs. distance.




(d) Resolution vs. distance.




(e) Mean vs. zenith.



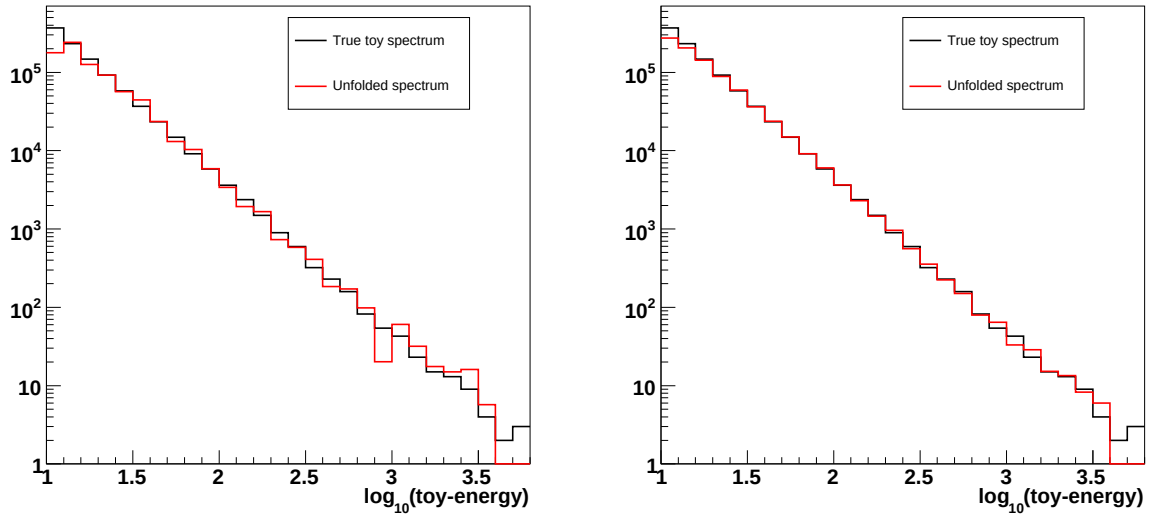

(f) Resolution vs. zenith.

**Figure A.5:** Resolution of zenith angle.

# Appendix B

# Dependence of the Unfolding result on the kernel MC

In Chap. 7.1.5 we require for possible unfolding algorithms, that the uncertainties in the response matrix due to the spectral distribution of the kernel MC can be neglected. Therefore, we use the toy MC to investigate the dependency of the final result on the kernel MC. An $E^{-3}$ spectrum is unfolded using on the one hand an $E^{-2}$ kernel MC, and on the other hand an $E^{-3}$ kernel MC. The results are shown in the plots below:




(a) Using an $E^{-2}$ kernel MC

(b) Using an $E^{-3}$ kernel MC

**Figure B.1:** Unfolding of an $E^{-3}$ power-law spectrum using two different kernel MC.

# Appendix C

# Scaling other Experimental Results

The final unfolded FD-mono spectrum can be compared to results from other experiments like HiRes and KASCADE-Grande . The agreement is expressed in terms of a mismatch in the absolute energy scale. To quantify the agreement the energy scale of KASCADE-Grande results has been shifted by 5 % and of HiRes results by 7 %.

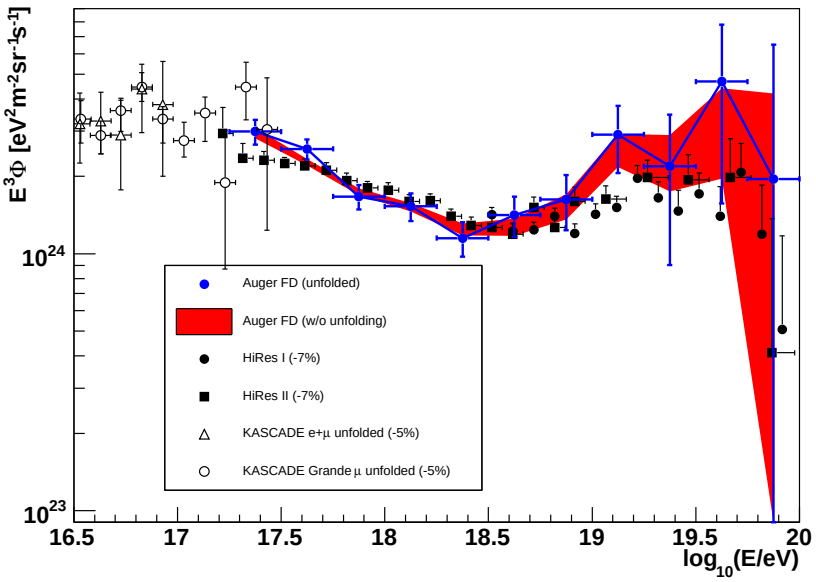



**Figure C.1:** Unfolded FD-mono spectrum compared to results from the HiRes and KASCADE experiment. The experimental results have been scaled in the absolute energy, by 5 % (KASCADE /KASCADE-Grande ) and 7 %(HiRes )

# Appendix D

# Parametrisation-Fits

The parametrisation of the response matrix (see Chap. 7) is done by fitting Gaussians to the $E_{\mathrm{rec}}$ values in slices of $E_{\mathrm{true}}$. The individual fits are given below:

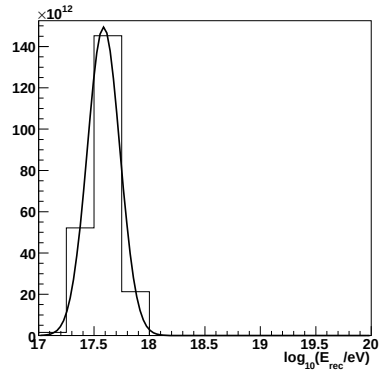
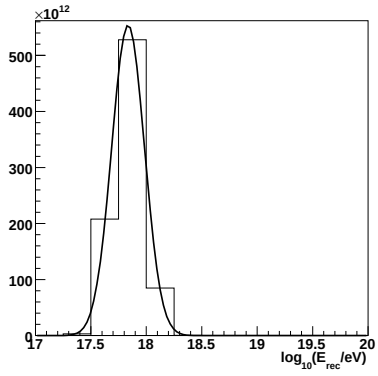
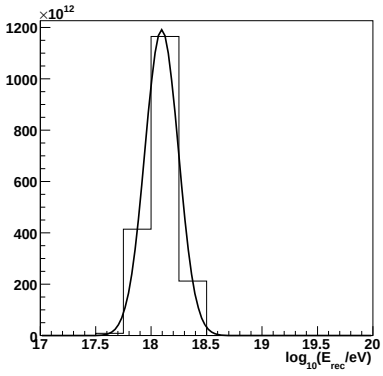
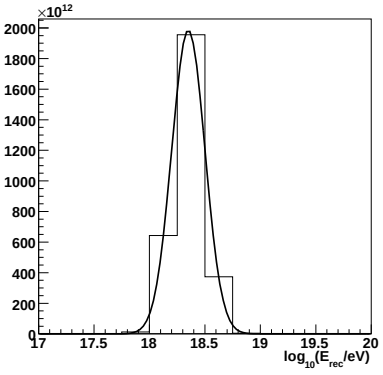
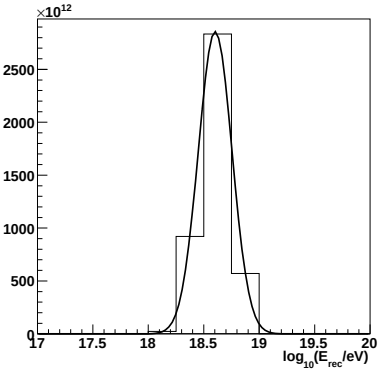
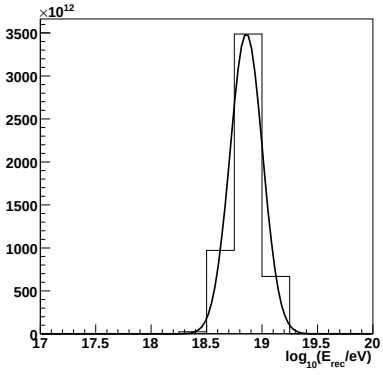
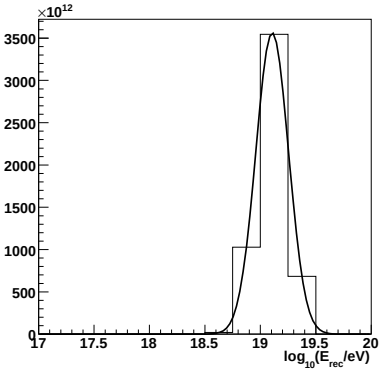
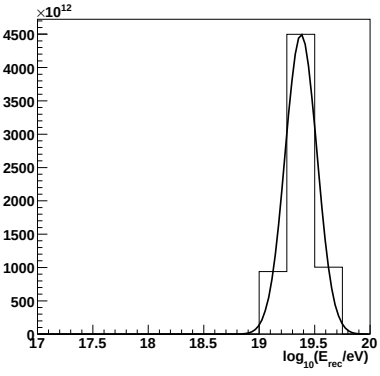
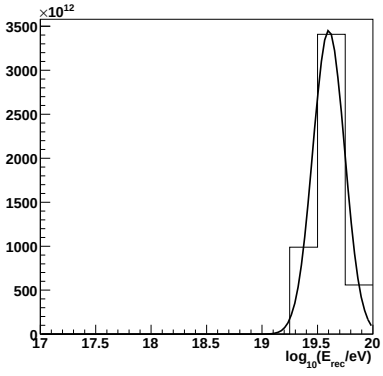
| $17.5 < \log_{10}(E_{\text{true}}/eV) < 17.75$ | $17.75 < \log_{10}(E_{\text{true}}/eV) < 18$ | $18 < \log_{10}(E_{\text{true}}/eV) < 18.25$ |
|---|---|---|
|  |  |  |
| $18.25 < \log_{10}(E_{\text{true}}/eV) < 18.5$ | $18.5 < \log_{10}(E_{\text{true}}/eV) < 18.75$ | $18.75 < \log_{10}(E_{\text{true}}/eV) < 19$ |
|  |  |  |
| $19 < \log_{10}(E_{\text{true}}/eV) < 19.25$ | $19.25 < \log_{10}(E_{\text{true}}/eV) < 19.5$ | $19.5 < \log_{10}(E_{\text{true}}/eV) < 19.75$ |
|  |  |  |

**Table D.1:** The Gaussian fits performed to parametrise the response matrix.

# Appendix E

# The MySQL database for the simulation-reconstruction chain

In Chap. 4.5 the use of a MySQL database as DST is motivated. The idea is to provide shower information from simulation and reconstruction. Therefore, high-level observables at generator-level and reconstruction level are stored into MySQL databases. The implemented MySQL layout is given below.

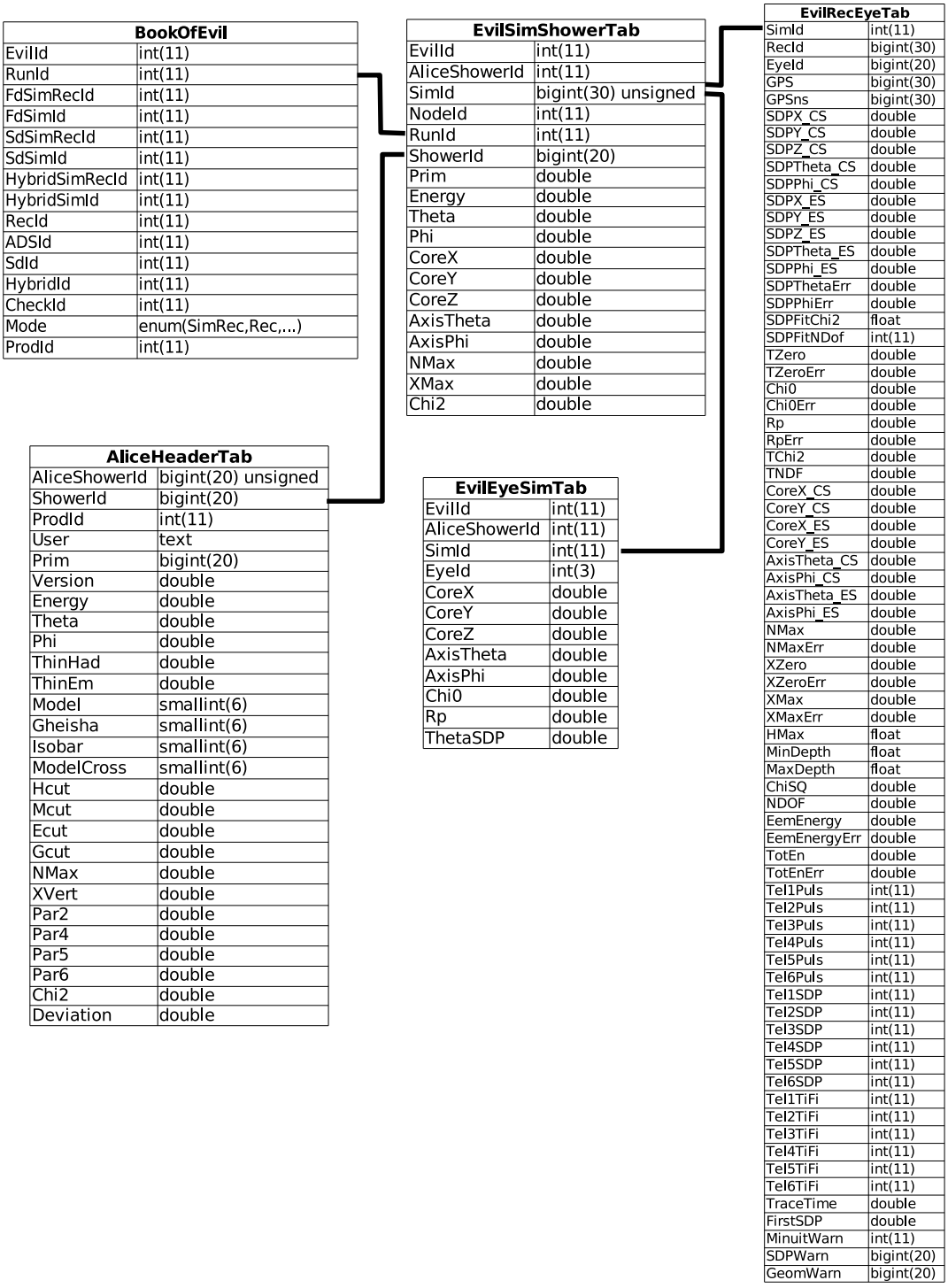



**Figure E.1:** The simulation-reconstruction storage schema.

# Bibliography


[1] M. Aglietta et al. Response of the Pierre Auger Observatory Water Cherenkov Detectors to Muons. In *International Cosmic Ray Conference*, volume 7 of *International Cosmic Ray Conference*, pages 83–+, 2005.

[2] F. Aharonian et al. The H.E.S.S. Survey of the Inner Galaxy in Very High Energy Gamma Rays. *ApJ*, 636:777–797, January 2006.

[3] M. V. Alania, J. Szabelski, and A. Wawrzynczak. Modeling and Experimental Study of Forbush Effects of Galactic Cosmic Rays. In *International Cosmic Ray Conference*, volume 6 of *International Cosmic Ray Conference*, pages 3585–+, July 2003.

[4] ALiCENext, 2006. http://www.alicenext.uni-wuppertal.de.

[5] D. Allard et al. The trigger system of the Pierre Auger Surface Detector: operation, efficiency and stability. In *International Cosmic Ray Conference*, volume 7 of *International Cosmic Ray Conference*, pages 287–+, 2005.

[6] M. Amenomori et al. Observation by an Air-Shower Array in Tibet of the Multi-TeV Cosmic-Ray Anisotropy due to Terrestrial Orbital Motion Around the Sun. *Physical Review Letters*, 93(6):061101–+, August 2004.

[7] L. Anchordoqui, H. Goldberg, S. Reucroft, and J. Swain. Extragalactic sources for ultrahigh energy cosmic ray nuclei. *Phys. Rev. D*, 64(12):123004–+, December 2001.

[8] L. Anchordoqui, T. Paul, S. Reucroft, and Swain J. movie-script in preparation.

[9] L. Anchordoqui, T. Paul, S. Reucroft, and J. Swain. Ultrahigh Energy Cosmic Rays. *International Journal of Modern Physics A*, 18:2229–2366, 2003.

[10] L. A. Anchordoqui, G. E. Romero, and J. A. Combi. Heavy nuclei at the end of the cosmic-ray spectrum? *Phys. Rev. D*, 60(10):103001–+, November 1999.

[11] L. A. Anchordoqui, G. E. Romero, J. A. Combi, and S. E. Perez Bergliaffa. High Energy Protons from PKS 1333-33. *Modern Physics Letters A*, 13:3039–3044, December 1998.

[12] C. D. Anderson. The Positive Electron. *Physical Review*, 43:491–494, March 1933.

[13] T. Antoni et al. Large-Scale Cosmic-Ray Anisotropy KASCADE. *ApJ*, 604:687–692, April 2004.

[14] S. Arigro et al. *Offline Software User's Guide*.

[15] J. Arons. Magnetars in the metagalaxy: An origin for ultra high energy cosmic rays in the nearby universe. *The Astrophysical Journal*, 589:871, 2003.

[16] F. Arqueros et al. The Central Laser Facility at the Pierre Auger Observatory. In *International Cosmic Ray Conference*, volume 8 of *International Cosmic Ray Conference*, pages 335–+, 2005.

[17] Askaryan. Exess negative charge of an electron-photon shower and its coherent radio emission. *Soviet Phys. JETP*, 14:441–443, 1962.

[18] Askaryan. Coherent radio emission from cosmic showers in air and in dense media. *Soviet Phys. JETP*, 21:658, 1965.

[19] P. Auger, P. Ehrenfest, R. Maze, J. Daudin, and R. A. Fréon. Extensive Cosmic-Ray Showers. *Reviews of Modern Physics*, 11:288–291, July 1939.

[20] P. Auger and T. Grivet. Analysis of deep rays. *Reviews of Modern Physics*, 11:232–234, July 1939.

[21] M. Ave et al. Cosmic-Ray Energy Spectrum above $3 \times 10^{17}$ eV measured with the Haverah Park Array. In *Proceedings of ICRC 2001*, pages 381–384, 2001.

[22] R. M. Baltrusaitis, R. Cady, G. L. Cassiday, R. Cooperv, J. W. Elbert, P. R. Gerhardy, S. Ko, E. C. Loh, M. Salamon, D. Steck, and P. Sokolsky. The Utah Fly's Eye detector. *Nuclear Instruments and Methods in Physics Research A*, 240:410–428, October 1985.

[23] R. Barlow. *Statistics-A Guide to the Use of Statistical Methods in Physical Science*. John Wiley & Sons, 1989.

[24] P. Bauleo, J. Brack, L. Garrard, J. Harton, R. Knapik, R. Meyhandan, A. C. Rovero, A. Tamashiro, D. Warner, and the Auger Collaboration". Absolute Calibration of the Auger Fluorescence Detectors. In *International Cosmic Ray Conference*, volume 8 of *International Cosmic Ray Conference*, pages 5–+, 2005.

[25] K.-H. Becker, A. Behrmann, F. Bracci, D. Delle Fratte, P. Facal San Luis, C. Fontaine, H. Geenen, D. Guerin, S. Hartmann, K.-H. Kampert, P. Lavoute, G. Matthiae, P. Petrinca, P. Privitera, J. Rodriguez Martino, G. Salina, V. Scherini, V. Verzi, U. Weischer, and C. Wiebusch. Qualification tests of the 11000 photomultipliers for the pierre auger observatory fluorescence detectors. *Nucl. Instr. Meth.*, A(576):301–311, 2007.

[26] K. H. Becker and et al. Photomultiplier Qualification Tests for the Fluorescence Detectors of the Pierre Auger Observatory. In *International Cosmic Ray Conference*, volume 8 of *International Cosmic Ray Conference*, pages 49–+, 2005.

[27] J. Bednarz and M. Ostrowski. Energy Spectra of Cosmic Rays Accelerated at Ultrarelativistic Shock Waves. *Physical Review Letters*, 80:3911–3914, May 1998.

[28] J. A. Bellido. Performance of the Fluorescence Detectors of the Pierre Auger Observatory. In *International Cosmic Ray Conference*, volume 7 of *International Cosmic Ray Conference*, pages 13–+, 2005.

[29] 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. *arXiv:astro-ph/0009039*, 15:167–175, April 2001.

[30] J. A. Bellido, D. D'Urso, H. Geenen, F. Guarino, L. Perrone, S. Petrera, L. Prado, and F. Salamida. Fluorescence and Hybrid Detection Aperture of the Pierre Auger Observatory. In *International Cosmic Ray Conference*, volume 8 of *International Cosmic Ray Conference*, pages 113–+, 2005. astro-ph/0205109.

[31] S. Y. BenZvi et al. The Lidar System of the Pierre Auger Observatory. *astro-ph/0609063*, September 2006.

[32] V. Berezinsky, A. Z. Gazizov, and S. I. Grigorieva. On astrophysical solution to ultra high energy cosmic rays. *Physical Review D*, 74:043005, 2006.

[33] D. R Bergman and J. W. Belz. Cosmic Rays: The Second Knee and Beyond. *astro-ph/0704.3721*, 704, April 2007.

[34] P. Bhattacharjee and G. Sigl. Origin and propagation of extremely high energy cosmic rays. *Physics Reports*, 327:109, 2000.

[35] W. R. Binns, T. L. Garrard, P. S. Gibner, M. H. Israel, M. P. Kertzman, J. Klarmann, B. J. Newport, E. C. Stone, and C. J. Waddington. Abundances of ultraheavy elements in the cosmic radiation - Results from HEAO 3. *ApJ*, 346:997–1009, November 1989.

[36] W. R. Binns, M. H. Israel, N. R. Brewster, D. J. Fixsen, and T. L. Garrard. Lead, platinum, and other heavy elements in the primary cosmic radiation - HEAO 3 results. *ApJ*, 297:111–118, October 1985.

[37] D. J. Bird. The HiRes Fly's Eye Detector. In J.P. Wefel M. M. Shapiro, R. Silberberg, editor, *Currents in High-Energy Astrophysics - Proceedings of the 9th International School of Cosmic Ray Astrophysics (Erice, Italy)*, pages 135–142. Kluwer Academic Publishers, 1994.

[38] D. J. Bird et al. Recent Results from the Fly's Eye Experiment. In J. P. Wefel, editor, *Currents in High-Energy Astrophysics - Proceedings of the 9th International School of Cosmic Ray Astrophysics (Erice, Italy)*, pages 123–134. Kluwer Academic Publishers, 1994.

[39] P. Blasi, R. I. Epstein, and A. V. Olinto. Ultra-High-Energy Cosmic Rays from Young Neutron Star Winds. *ApJ*, 533:L123–L126, April 2000.

[40] V. Blobel. Unfolding Methods in High-Energy Physics Experiments. In *Proceedings of the 1984 CERN School of Computing*. CERN, 1984.

[41] V. Blobel. The Run Manual - Regularized Unfolding for High-Energy Physics Experiments. Technical report, OPAL, 1996.

[42] V. Blobel. Unfolding in HEP Experiments. In *Proceedings of the ZEUS Seminar on Experiments at HERA*, 1997.

[43] V. Blobel. Unfolding in Data Analysis. In *Proceedings of the ZEUS Seminar*, 2000.

[44] E. Boldt and P. Ghosh. Cosmic rays from remnants of quasars? *MNRAS*, 307:491–494, August 1999.

[45] E. Boldt and M. Loewenstein. Cosmic ray generation by quasar remnants: constraints and implications. *MNRAS*, 316:L29–L33, August 2000.

[46] J. Brack, P. Bauleo, J. Harton, R. Knapik, and A. Rovero. Multi-wavelength calibration at los leones. Technical Report GAP-2007-024, University of Colorado, Boulder; Colorado State University, Ft. Collin; Instituto de Astronomia y Fisica del Espacio, Buenos Aires, 2007.

[47] H. C. Burger and P. H. van Cittert. Wahre und scheinbare Intensitätsverteilung in Spektrallinien. *Zeitschrift fur Physik*, 79:722–+, December 1932.

[48] H. C. Burger and P. H. van Cittert. Wahre und scheinbare Intensitätsverteilung in Spektrallinien. II. *Zeitschrift fur Physik*, 81:428–+, March 1933.

[49] B. Byrnak, N. Lund, L. I. Rasmussen, M. Rotenberg, J. Engelmann, P. Goret, and E. Júliusson. The Abundances of the Elements with Z $<$ 26 IN the Cosmic Radiation. In *International Cosmic Ray Conference*, pages 29–32, August 1983.

[50] S. Chaty. The role of microquasars in astroparticle physics, 2006.

[51] J. Chirin. Private communications.

[52] J. Chirin et al. Atmospheric aerosol monitoring at the Pierre Auger Observatory. In *International Cosmic Ray Conference*, volume 8 of *International Cosmic Ray Conference*, pages 347–+, 2005.

[53] D. J. H. Chung, E. W. Kolb, and A. Riotto. Production of massive particles during reheating. *Phys. Rev. D*, 60(6):063504–+, September 1999.

[54] G. Cowan. *Statistical Data Analysis*. Clarendon Press, Oxford, 1998.

[55] J. W. Cronin. The highest-energy cosmic rays. *Nuclear Physics B Proceedings Supplements*, 138:465–491, January 2005.

[56] G. D'Agostini. Bayesian Unfolding. *Nuclear Instruments and Methods in Physics Research A*, 362:487–499, 1995.

[57] G. D'Agostini. Bayesian reasoning versus conventional statistics in high-energy physics. *Talk at the XVIII International Workshop on Maximum Entropy and Bayesian Methods (Maxent98), Garching / München (Germany)*, 1998.

[58] G. D'Agostini. Bayesian Reasoning in High-Energy Physics. Principles and Applications. In *CERN Yellow Report 99-03*. CERN, 1999.

[59] A. Dar. The origin of cosmic rays —A 92-year-old puzzle solved? *Nuovo Cimento B Serie*, 120:767–+, June 2005.

[60] G. A. de Nolfo et al. Observations of the Li, Be, and B isotopes and constraints on cosmic-ray propagation. *Advances in Space Research*, 38:1558–1564, 2006.

[61] A. De Rujula. A cannonball model of cosmic rays, 2004.

[62] H. J. de Vega and N. Sanchez. Extreme Energy Cosmic Rays: Bottom-up vs. Top-down scenarii. *astro-ph/0301039*, January 2003.

[63] H. J. de Vega and N. G. Sánchez. Decay rates of unstable particles and the extreme energy cosmic rays top-down scenarios. *Phys. Rev. D*, 67(12):125019–+, June 2003.

[64] D. D'Urso. *Pierre Auger Observatory: Fluorescence Detector Event Reconstruction and Data Analysis*. PhD thesis, Universita Degli Studi Di Catania, Faculta' di scienze matematiche, fisiche e naturali, 2005.

[65] J. W. Elbert and P. Sommers. In search of a source for the 320 EeV Fly's Eye cosmic ray. *ApJ*, 441:151–161, March 1995.

[66] A. Etchegoyen for the Pierre Auger Collaboration. Amiga, auger muons and infill for the ground array. In *Proceedings of the 30th ICRC, Mexico*, 2007. icrc-1307.

[67] E. Fermi. On the Origin of the Cosmic Radiation. *Physical Review*, 75(8):1169, 1949.

[68] P. H. Fowler, R. N. F. Walker, M. R. W. Masheder, R. T. Moses, A. Worley, and A. M. Gay. Ariel 6 measurements of the fluxes of ultraheavy cosmic rays. *ApJ*, 314:739–746, March 1987.

[69] T. K. Gaisser. *Cosmic-Rays and Particle Physics*. Cambridge University Press, 1990.

[70] T. K. Gaisser et al. Particle Astrophsics with High Energy Neutrinos. *Physics Reports*, 258:173–185, 1995.

[71] T. K. Gaisser and A. M. Hillas. Reliability of the Method of Constant Intensity Cuts for Reconstructing the Average Development of Vertical Showers. In *International Cosmic Ray Conference*, volume 8 of *International Cosmic Ray Conference*, pages 353–+, 1977.

[72] Y. A. Gallant and A. Achterberg. Ultra-high-energy cosmic ray acceleration by relativistic blast waves. *MNRAS*, 305:L6–L10, May 1999.

[73] Y. A. Gallant, A. Achterberg, J. G. Kirk, and A. W. Guthmann. Particle Acceleration at Ultra-relativistic Shocks and the Spectra of Relativistic Fireballs. In R. M. Kippen, R. S. Mallozzi, and G. J. Fishman, editors, *Gamma-ray Bursts, 5th Huntsville Symposium*, volume 526 of *American Institute of Physics Conference Series*, pages 524–+, 2000.

[74] H. Geenen. Energy reconstruction and spectral unfolding of atmospheric leptons with the AMANDA detector, 2002. Diploma thesis, WU-D-2002-10.

[75] H. Geenen. Analysis of the pierre auger fluorescence data. In *Proceedings ECRS 2006, Lisbon*, 2006.

[76] H. Geenen. The fluorescence telescopes of the pierre auger observatory. In *Proceedings RICAP 2007, Rome*, 2007.

[77] H. Geenen et al. Simulation and reconstruction performance of fdsimpreprod and flores using large corsika shower statistics. Technical Report GAP-2004-006, University of Wuppertal, 2004.

[78] H. Geenen et al. Validation of the real and simulated data of the Pierre Auger fluorescence telescopes. In *Proceedings of the 29th ICRC, Pune*, volume 1, 2005.

[79] V. L. Ginzburg and S. I. Syrovatskii. *The Origin of Cosmic Rays*. The Origin of Cosmic Rays, New York: Macmillan, 1964, 1964.

[80] R. Gold. An iterative unfolding method for response matrices. Technical Report USAEC Report ANL-6984, Argonne National Laboratory, 1964.

[81] D. Gorbunov, P. Tinyakov, I. Tkachev, and S. Troitsky. Evidence for a connection between the gamma-ray and the highest energy cosmic-ray emissions by bl lacertae objects. *The Astrophysical Journal*, 577:L93, 2002.

[82] N. Grevesse and E. Anders. Solar-system abundances of the elements - A new table. In C. J. Waddington, editor, *AIP Conf. Proc. 183: Cosmic Abundances of Matter*, pages 1–8, 1989.

[83] D. E. Groom and S. R. Klein. Passage of Particles through Matter. In *The European Physical Journal C*, volume 15, chapter 23, pages 163–173. Springer Verlag, 2000.

[84] G. Guillian. Observation of the Anisotropy of 10 TeV Primary Cosmic Ray Nuclei Flux with the Super-Kamiokande-I Detector. In *International Cosmic Ray Conference*, volume 6 of *International Cosmic Ray Conference*, pages 85–+, 2005.

[85] G. Guillian et al. Observation of the anisotropy of 10TeV primary cosmic ray nuclei flux with the Super-Kamiokande-I detector. *Phys. Rev. D*, 75(6):062003–+, March 2007.

[86] F. Halzen. Lectures on Neutrino-Astronomy: Theory and Experiment. astro-ph/9810368, 1998.

[87] N. Hayashida et al. The anisotropy of cosmic ray arrival directions around 10ˆ18 eV. *arXiv:astro-ph/9906056*, 10:303–311, May 1999.

[88] M. Healy for the Pierre Auger Collaboration. Search for ultra-high energy photons with the pierre auger observatory. In *Proceedings of the 30th ICRC, Mexico*, 2007. icrc-602.

[89] W. Heitler. *Quantum theory of radiation*. International Series of Monographs on Physics, Oxford: Clarendon, 1954, 3rd ed., 1954.

[90] V. Hess. Über beobachtungen der durchdringenden strahlung bei sieben freiballonfahrten. *Physikalische Zeitschrift*, 22(21):1084–1091, November 1912.

[91] A. M. Hillas. The Origin of Ultra-High-Energy Cosmic Rays. *ARA&A*, 22:425–444, 1984.

[92] S. Inoue. Astrophysical Origins of the Highest Energy Cosmic Rays. *astro-ph/0701835*, January 2007.

[93] F. James. *MINUIT-Function Minimization and Error Analysis, Refernce Manual*. CERN, Geneva, 1994. Version 94.1, CERN Program Library entry D506.

[94] J. V. Jelley and J. H. Fruin. Radio Pulses from Extensive Cosmic-Ray Air Showers. *Nature*, 205:327–328, January 1965.

[95] J. R. Jokipii. Rate of energy gain and maximum energy in diffusive shock acceleration. *ApJ*, 313:842–846, February 1987.

[96] J. R. Jokipii and G. Morfill. Ultra-high-energy cosmic rays in a galactic wind and its termination shock. *ApJ*, 312:170–177, January 1987.

[97] L. W Jones. Cosmic ray studies with the l3-cosmics program at cern. *Czechoslovak Journal of Physics*, 56:A201–A210, 2006.

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

[99] K.-H. Kampert. Cosmic Rays from the Knee to the Ankle - Status and Prospects -. *Invited paper presented at the CRIS 2006 Cosmic Ray International Seminar*, November 2006.

[100] K.-H. Kampert. priv. communication, 2007.

[101] H. Kang, J. P. Rachen, and P. L. Biermann. Contributions to the Cosmic Ray Flux above the Ankle: Clusters of Galaxies. *MNRAS*, 286:257–267, April 1997.

[102] B. Keilhauer. The Fluorescence Detector of the Pierre Auger Observatory - a Calorimeter for UHECR. In *AIP Conf. Proc. 867: Calorimetry in High Energy Physics: XII*, pages 175–182, October 2006.

[103] B. Keilhauer, J. Blümer, R. Engel, and H. O. Klages. Impact of varying atmospheric profiles on extensive air shower observation: Fluorescence light emission and energy reconstruction. *astro-ph/0511153*, 25:259–268, May 2006.

[104] B. Keilhauer et al. Atmospheric profiles at the southern pierre auger observatory and their relevance to air shower measurement. In *Proc. 29th Int. Cosmic Ray Conf., Pune*, volume 7, page 123, 2005.

[105] B. Keilhauer et al. Nitrogen fluorescenceyield in dependence on atmospheric conditions. In *Proceedings of the 29th ICRC, Pune*, volume 1, pages 101–106, 2005.

[106] B. Keilhauer, H. O. Klages, and M. Risse. Results of the first balloon measurements above the pampa amarilla. Technical Report GAP-2003-009, Forschungszentrum Karlsruhe, 2003.

[107] H. O. Klages for the Pierre Auger Collaboration. Heat - enhancement telescopes for the pierre auger southern observatory. In *Proceedings of the 30th ICRC, Mexico*, 2007. icrc-065.

[108] M. Kleifges. The fluorescene detector electronics of the Pierre Auger Observatory. *Nuclear Instruments and Methods in Physics Research A*, 518:180–182, February 2004.

[109] C. Kouveliotou. Magnetars. *Proceedings of the National Academy of Science*, 96:5351–5352, May 1999.

[110] C. Kouveliotou, S. Dieters, T. Strohmayer, J. van Paradijs, G. J. Fishman, C. A. Meegan, K. Hurley, J. Kommers, I. Smith, D. Frail, and T. Murakami. An X-ray pulsar with a superstrong magnetic field in the soft $\gamma$-ray repeater SGR1806 - 20. *Nature*, 393:235–237, May 1998.

[111] C. Kouveliotou, T. Strohmayer, K. Hurley, J. van Paradijs, M. H. Finger, S. Dieters, P. Woods, C. Thompson, and R. C. Duncan. Discovery of a Magnetar Associated with the Soft Gamma Repeater SGR 1900+14. *ApJ*, 510:L115–L118, January 1999.

[112] D. Kümpel. Diploma thesis in preparation.

[113] D. Kümpel et al. Delayed fluorescence light within hybrid reconstruction. Technical Report GAP-2007-059, University of Wuppertal, 2007.

[114] D. Kümpel et al. Effect of realistic speed of light on hybrid reconstruction. Technical Report GAP-2007-036, University of Wuppertal, 2007.

[115] D. Kümpel et al. Geometry reconstruction of fluorescence detectors revisited. In *Proceedings of the 30th ICRC, Mexico*, 2007. icrc-082.

[116] P. O. Lagage and C. J. Cesarsky. Cosmic-ray shock acceleration in the presence of self-excited waves. *A&A*, 118:223–228, February 1983.

[117] C. M. G. Lattes, H. Muirhead, G. P. S. Occhialini, and C. F. Powell. Processes involving charged mesons. *Nature*, 159:694 – 697, 1947.

[118] M. Lemoine, G. Pelletier, and B. Revenu. On the Efficiency of Fermi Acceleration at Relativistic Shocks. *ApJ*, 645:L129–L132, July 2006.

[119] K. Levenberg. A method for the solution of certain problems in least square. *Quart. Appl. Math. 2*, pages 164–168, 1944.

[120] A. Levinson. High-energy aspects of astrophysical jets. *International Journal of Modern Physics A*, 21:6015, 2006.

[121] C. Litwin and R. Rosner. Plasmoid Impacts on Neutron Stars and Highest Energy Cosmic Rays. *Physical Review Letters*, 86:4745–4748, May 2001.

[122] W. Lohmann. Energy Loss of Muons in the Energy Range 1-10000 GeV. Technical report, CERN, Geneva, 1985. CERN-85-03.

[123] M. S. Longair. *High energy Astrophysics-Volume I.* Cambridge University Press, 2004.

[124] M. S. Longair. *High energy Astrophysics-Volume II.* Cambridge University Press, 2004.

[125] L. B. Lucy. An Iterative Technique for the Rectification of Observed Distributions. *The Astronomical Journal*, 70(6):745–754, 1974.

[126] J. Madsen and J. M. Larsen. Strangelets as cosmic rays beyond the greisen-zatsepin-kuzmin cutoff. *Physical Review Letters*, 90:121102, 2003.

[127] D. W. Marquardt. An algorithm for least-squares estimation of non-linear parameters. *. J. Soc. Indust. Appl. Math.*, 11:431–441, 1963.

[128] J. Matthews. A Heitler model of extensive air showers. *Astroparticle Physics*, 22:387–397, January 2005.

[129] G. Matthiae et al. Report on the tests of the pmt/he units of the fd. Technical Report ROM2F/2003/23, University Roma II, 2003.

[130] G. Medina Tanco for the Pierre Auger Collaboration. Astrophysics motivation behind the pierre auger southern observatory enhance- ments. In *Proceedings of the 30th ICRC, Mexico*, 2007. icrc-991.

[131] M. Milgrom and V. Usov. Gamma-ray bursters as sources of cosmic rays. *Astroparticle Physics*, 4:365–369, April 1996.

[132] R. A. Millikan and G. H. Cameron. High Frequency Rays of Cosmic Origin III. Measurements in Snow-Fed Lakes at High Altitudes. *Physical Review*, 28:851–868, November 1926.

[133] F. Miniati, D. Ryu, H. Kang, and T. W. Jones. Large cosmic shock waves as sites for particle acceleration, 1999.

[134] I. F. Mirabel. Microquasars as sources of high energy phenomena, 2002.

[135] J. Miralda-Escude and E. Waxman. Signatures of the Origin of High-Energy Cosmic Rays in Cosmological Gamma-Ray Bursts. *ApJ*, 462:L59+, May 1996.

[136] M. Mostafa. "checking the alignment of the fd cameras". Technical Report GAP-2005-102, University of New Mexico, Albuquerque, 2005.

[137] R. Mussa, G. Sequeiros, and A. Tonachini. Detecting clouds with elastic lidar telescopes. Technical Report GAP-2007-038, INFN Torino, University of Torino, 2007.

[138] N. Nierstenhöfer. *Untersuchung systematischer Rekonstruktionsunsicherheiten der Fluoreszenzdetektoren des Pierre Auger Observatoriums anhand experimenteller Daten.* PhD thesis, University of Wuppertal, 2007.

[139] D. Nitz for the Auger Collaboration. The northern site of the pierre auger observatory. In *Proceedings of the 30th ICRC, Mexico*, 2007. icrc-180.

[140] C. A. Norman, D. B. Melrose, and A. Achterberg. The Origin of Cosmic Rays above 10 18.5 eV. *ApJ*, 454:60–+, November 1995.

[141] M. Ostrowski. Acceleration of uhe cosmic ray particles at relativistic jets in extragalactic radio sources, 1998.

[142] M. Ostrowski. On possible 'cosmic ray cocoons' of relativistic jets. *MNRAS*, 312:579–584, March 2000.

[143] M. Ostrowski. Mechanisms and sites of ultra high energy cosmic ray origin. *astro-ph/0101053*, 18:229, 2002.

[144] M. Ostrowski and J. Bednarz. Some constraints for the first-order Fermi acceleration at ultra-relativistic shocks. In A. Gimenez, V. Reglero, and C. Winkler, editors, *Exploring the Gamma-Ray Universe*, volume 459 of *ESA Special Publication*, pages 175–178, September 2001.

[145] M. Ostrowski and J. Bednarz. Comment on the first-order Fermi acceleration at ultra-relativistic shocks. *A&A*, 394:1141–1144, November 2002.

[146] M. Ostrowski and G. Siemieniec-Ozieblo. Cosmic ray acceleration at supergalactic accretion shocks: a new upper energy limit due to a finite shock extension, 2002.

[147] Y. Oyama. Anisotropy of the primary cosmic-ray flux in Super-Kamiokande. *astro-ph/0605020*, May 2006.

[148] Facal San Luis P., E. Kemp, G. Matthiae, and P. Privitera. The test system for the photomultipliers of the fluorescence detector. Technical Report GAP-2001-042, University of Rome II, 2001.

[149] L. Perrone et al. Auger fd: Detector response to simulated showers and real event topologies. In *Proceedings CRIS, Catania*, 2004.

[150] S. Plewnia, T. Berghöfer, J. Blümer, P. Buchholz, J. Engler, N. O. Hashim, J. R. Hörandel, R. Lixandru, J. Milke, W. Walkowiak, and J. Wochele. A sampling calorimeter with warm-liquid ionization chambers. *Nuclear Instruments and Methods in Physics Research A*, 566:422–432, October 2006.

[151] R. J. Protheroe and A. P. Szabo. High energy cosmic rays from active galactic nuclei. *Physical Review Letters*, 69:2885–2888, November 1992.

[152] J. Pugliese. *An Exploded Jet for Gamma Ray Bursts*. PhD thesis, Rheinische Friederich-Wilhelms-Universität, Bonn, 2000.

[153] J. P. Rachen and P. L. Biermann. Extragalactic Ultra-High Energy Cosmic-Rays - Part One - Contribution from Hot Spots in Fr-II Radio Galaxies. *A&A*, 272:161–+, May 1993.

[154] J. P. Rachen, T. Stanev, and P. L. Biermann. Extragalactic ultra-high energy cosmic rays. II. Comparison with experimental data. *A&A*, 273:377–+, June 1993.

[155] J. Ridky. The Surface Detector of the Pierre Auger Observatory. *Nuclear Physics B Proceedings Supplements*, 165:45–49, March 2007.

[156] M. Risse. Upper limit on the primary photon fraction from the Pierre Auger Observatory. In *International Cosmic Ray Conference*, volume 7 of *International Cosmic Ray Conference*, pages 147–+, 2005.

[157] M. Risse. An upper limit to photons from first data taken by the Pierre Auger Observatory. *astro-ph/0701065*, January 2007.

[158] M. Risse et al. An upper limit to the photon fraction in cosmic rays above $10^{19}$ eV from the Pierre Auger Observatory. *arXiv:astro-ph/0606619*, 27:155–168, March 2007.

[159] M. Risse and P. Homola. Prospects of searches for ultra-high energy photon. In *Proceedings of the 30th ICRC, Mexico*, 2007. icrc-164.

[160] G. Rochester and C. Butler. Evidence for the existence of new unstable elementary particles. *Nature*, 160:855, 1947.

[161] B. Rossi. *High Energy Particles*. Englewood Cliffs, NJ, 1952.

[162] M. Roth for the Auger Collaboration. Measurement of the UHECR energy spectrum using data from the Surface Detector of the Pierre Auger Observatory. *astro-ph/0706.2096*, 706, June 2007.

[163] I. L. Rozental. Interaction of Cosmic Muons of High Energy. *Soviet Physics Uspekhi*, 11(1):49, 1968.

[164] D. Ryu, J. P. Ostriker, H. Kang, and R. Cen. A cosmological hydrodynamic code based on the total variation diminishing scheme. *ApJ*, 414:1–19, September 1993.

[165] V. Scherini. Performance of the pierre auger detector at the highest energies. In *Proceedings ECRS 2006, Lisbon*, 2006.

[166] F. Schmidt. Corsika shower images. http://www.ast.leeds.ac.uk/ fs/showerimages.html ).

[167] N. Schmitz. *Neutrinophysik*. Teubner Studienbücher, 1997.

[168] R. C. Shellard for the Pierre AUGER Collaboration. First Results from the Pierre Auger Observatory. *Brazilian Journal of Physics*, 36:1184–1193, December 2006.

[169] Y. Shikaze and BESS Collaboration. Solar Modulation Effect on the Cosmic-Ray Proton Spectra Measured by BESS. In *International Cosmic Ray Conference*, volume 7 of *International Cosmic Ray Conference*, pages 4027–+, July 2003.

[170] P. Sommers. First Estimate of the Primary Cosmic Ray Energy Spectrum above 3 EeV from the Pierre Auger Observatory. In *International Cosmic Ray Conference*, volume 7 of *International Cosmic Ray Conference*, pages 387–+, 2005.

[171] C. Song, Z. Cao, B. R. Dawson, B. E. Fick, P. Sokolsky, and X. Zhang. Energy estimation of UHE cosmic rays using the atmospheric fluorescence technique. *astro-ph/9910195*, 14:7–13, August 2000.

[172] T. S. Stanev. *Cosmic Ray Physics*. John Wiley and Sons Ltd, 1998.

[173] T. S. Stanev. *High Energy Cosmic Rays*. Springer, 2004.

[174] J. C. Street and E. C. Stevenson. New Evidence for the Existence of a Particle of Mass Intermediate Between the Proton and Electron. *Physical Review*, 52:1003–1004, November 1937.

[175] M. Takeda et al. Energy determination in the Akeno Giant Air Shower Array experiment. *astro-ph/0209422*, 19:447–462, July 2003.

[176] The HiRes Collaboration. Observation of the GZK Cutoff by the HiRes Experiment. *astro-ph/0703099*, March 2007.

[177] The Pierre Auger Collaboration. The pierre auger observatory design report. Technical report, Fermilab, 1997.

[178] The Pierre AUGER Collaboration. The Surface Detectors of the Pierre Auger Observatory. *Nuclear Physics B Proceedings Supplements*, 136:393–398, November 2004.

[179] The Pierre Auger Collaboration. Anisotropy Studies Around the Galactic Center at EeV Energies with Auger Data. *astro-ph/0507331*, July 2005.

[180] The Pierre Auger Collaboration. Performance of the Fluorescence Detectors of the Pierre Auger Observatory. *astro-ph/0508389*, August 2005.

[181] The Pierre Auger Collaboration. Statistical and systematic uncertainties in the event reconstruction and S(1000) determination by the Pierre Auger surface detector. *astro-ph/0507029*, July 2005.

[182] The Pierre AUGER Collaboration. Anisotropy studies around the galactic centre at EeV energies with the Auger Observatory. *Astroparticle Physics*, 27:244–253, April 2007.

[183] G. B. Thomson, f. t. High Resolution Fly's Eye, and Collaboration. Observation of the GZK Cutoff by the HiRes Experiment. *ArXiv Astrophysics e-prints*, September 2006.

[184] P. G. Tinyakov and I. I. Tkachev. BL Lacertae are Probable Sources of the Observed Ultrahigh Energy Cosmic Rays. *Soviet Journal of Experimental and Theoretical Physics Letters*, 74:445–+, November 2001.

[185] S. Udo and Tibet Asgamma Collaboration. Galactic Anisotropy of Multi-TeV Cosmic-Ray Intensity Observed by the Tibet III Air Shower Array. In *International Cosmic Ray Conference*, volume 7 of *International Cosmic Ray Conference*, pages 3917–+, July 2003.

[186] H. Ulrich. *Untersuchung zum primären Energiespektrum der kosmischen Strahlung im PeV-Bereich mit dem KASCADE-Experiment*. PhD thesis, Univierstität Karlsruhe (TH), 2003.

[187] M. Unger for the Pierre Auger Collaboration. Study of the cosmic ray composition above 0.2 eev using the longitudinal profiles of showers observed at the pierre auger observatory. In *Proceedings of the 30th ICRC, Mexico*, 2007. icrc-594.

[188] J. van Buren. *Untersuchung der myonischen Komponente ausgedehnter Luftschauer mit dem KASCADE-Grande Experiment*. PhD thesis, University of Karlsruhe, 2005.

[189] P. H. van Cittert. Zum Einfluß der Spaltbreite auf die Intensitätsverteilung in Spektrallinien. II. *Zeitschrift fur Physik*, 69:298–+, May 1931.

[190] A. van den Berg for the Pierre Auger Collaboration. Radio detection of high-energy cosmic rays at the pierre auger observatory. In *Proceedings of the 30th ICRC, Mexico*, 2007. icrc-176.

[191] A. Venkatesan, M. C. Miller, and A. V. Olinto. Constraints on the Production of Ultra–High-Energy Cosmic Rays by Isolated Neutron Stars. *ApJ*, 484:323–+, July 1997.

[192] M. Vietri. The Acceleration of Ultra–High-Energy Cosmic Rays in Gamma-Ray Bursts. *ApJ*, 453:883–+, November 1995.

[193] E. F. Visbal, S. Y. BenZvi, B. M. Connolly, J. Matthews, M. Prouza, and S. Westerhoff. First analysis of signals from the auger apf light sources. Technical Report GAP-2006-083, Columbia University, Carnegie Mellon University, University of New Mexico Albuquerque, 2006.

[194] H. J. Voelk and P. L. Biermann. Maximum energy of cosmic-ray particles accelerated by supernova remnant shocks in stellar wind cavities. *ApJ*, 333:L65–L68, October 1988.

[195] W. Wagner. *Design and Realisation of a new AMANDA Data Aquisition System with Transient Waveform Recorders*. PhD thesis, University of Dortmund, Oct. 2004.

[196] E. Waxman. Cosmological Gamma-Ray Bursts and the Highest Energy Cosmic Rays. *Physical Review Letters*, 75:386–389, July 1995.

[197] J. P. Wefel. The composition of the cosmic rays - an update. *NASA STI/Recon Technical Report A*, 93:51971–+, 1991.

[198] B. Wiebel-Sooth. *Measurement of the Allparticle Spectrum and Chemical Composition of Cosmic-Rays with the HEGRA Detector*. PhD thesis, Universität Wuppertal, 1998.

[199] C. Wiebusch et al. Muon Energy Reconstruction with AMANDA.

[200] L. Wiencke. Extracting first science measurements from the southern detector of the Pierre Auger observatory. *Nuclear Instruments and Methods in Physics Research A*, 572:508–510, March 2007.

[201] T. Yamamoto for the Pierre Auger Collaboration. The uhecr spectrum measured at the pierre auger observatory and its astrophysical implication. In *Proceedings of the 30th ICRC, Mexico*, 2007. icrc-318.

[202] S. Yoshida et al. The Cosmic-Ray Energy Spectrum above $3 \cdot 10^{18}$ eV measured by the Akeno Giant Air Shower Array. *Astroparticle Physics*, 3:105–123, 1995.

# Glossary

| | |
|---|---|
| **AGN** | Active Galactic Nuclei |
| **ALiCENext** | Advanced Linux Cluster Engine, next generation, farming cluster at the University of Wuppertal |
| **BH** | Black Hole |
| **CO** | Fluorescence Telescope Station Coihueco |
| **DAQ** | Data Acquisition System |
| **DST** | Data storage tree, in reminiscence to the once commonly used Data Summary Tapes. |
| **EAS** | Extensive Air Shower |
| **EHECR** | Extreme High Energy Cosmic Rays, in the context of this work we consider cosmic rays above the ankle to be EHECR. |
| **FADC** | Flash Analog To Digital Converter Pulse |
| **FD** | Fluorescence Detector |
| **FoV** | Field of View |
| **GRB** | Gamma Ray Burst |
| **GZK** | Greisen, Zatsepin, Kuzmin, see Chap. 2.1.4 |
| **HE** | Head electronics |
| **HV** | High voltage |
| **ISM** | InterStellar Medium |
| **LA** | Fluorescence Telescope Station Loma Amarilla |
| **LDF** | Lateral Density Function |
| **LL** | Fluorescence Telescope Station Los Leones |
| **LM** | Fluorescence Telescope Station Los Morados |
| **LV** | Low voltage |
| **pdf** | probability density function |
| **PMT** | Photo-multiplier |
| **SD** | Surface Detector |
| **SDP** | Surface Detector Plane |
| **SN** | SuperNova |
| **UHECR** | Ultra High Energy Cosmic Rays, in the context of this work we consider cosmic rays above the knee to be UHECR. |
| **VEM** | Vertical Equivalent Muon |

# Acknowledgement

Numerous people have contributed greatly both to my understanding of astrophysics in general and to this thesis in particular - too many to acknowledge with the hope of completeness.
Especially I would like to thank Prof. Dr. K.-H. Kampert, who introduced me to the Pierre Auger experiment and inspired me all along. His support and his ideas have an important influence on my work.
I am deeply indebted to Prof. Dr. C. Wiebusch for many constructive discussions and to Prof. Dr. L. Perrone for his cooperative way in guiding the progress of this work. I really enjoyed working with both of you.
Many people made the time in Wuppertal very pleasant. Especially, I would like to express my gratitude to V. Scherini, Dr. J. Rautenberg, N. Nierstenhöfer, J. Pouryamout and M. Risse for the many discussions we had and the experiences we shared in South America and anywhere else. I should mention the fortune to have Dr. S. Robbins around not only to train my English, but also to invest a lot of work and time in the progress of this work and sharing with me his insights of bug fixing. I am obliged to all other colleagues in the Astro-Particle group in Wuppertal:
Thanks to my office mates L. Linden and A. Franckowiak for their patience. Thanks to A. Tepe, B. Semburg, M. Gurtner, D. Kümpel, O. Tascau, D. Fuhrmann, A. Nasseri, S. Hartmann, J. Auffenberg, Prof. Dr. Helbing and T. Karg for the nice time and for many fruitful discussions. Thanks for your patience to listen to my problems on astro-particle physics and besides it. Especially, I would like to express my gratitude to I. Schaarwächter for having been around and helping me with all my paperwork. Many thanks also go to K.-H. Becker, T. Harenberg and D. Meder-Marouelli for having always listened to my IT-problems. I also remember Prof. B. Jacobi, who supported me and I regret that he has not got the chance to see my final work.
I am obliged to mention the *academia*: Dr. B. Breidenbach, M. Deeg, Dr. T. Messarius and K. Schinarakis for motivating me since the very beginning of our studies. It was great to have you around! Many thanks also to C. Markert for many discussions on physics and life besides it.
This work has been carried out in the framework of the Auger collaboration and I am much obliged to all former and present members as well as the local staff in Malargue, Argentina, for their tremendous work. *¡Gracias a todos ustedes por el gran trabajo, esfuerzo y dedicación! Gracias por las maravillosas vivencias y su hospitalidad durante mis estadías en la bella Argentina.*

I would also like to thank the Deutscher Forschungsgemeinschaft (DFG) for supporting my travels with grants.
I am deeply indebted to Dr. H. Ulrich, who contributed a lot to this work by having shared with me his insights of unfolding. Furthermore, I would like to thank Prof. Dr. S. Petrera, Dr. F. Salamida, Dr. M. Unger and for many fruitful hints and support. Special thanks go to Dr. J. Diaz for her friendship and many nice hours in Los Leones.
I am very grateful to Dr. M. Pycior, without whom I would not have reached this point. I would like to thank my parents and all my friends, who have always motivated and continuously supported me with incredible patience. Finally, thank you, Viola, for keeping my spirits up, especially during the last month of writing this thesis.