

Analysis of the First Data of the AugerPrime Detector Upgrade

For the attainment of the academic degree of
Doctorate in Science

from the
Karlsruher Institut für Technologie (KIT)
and the
Universidad Nacional de San Martín (UNSAM)

accepted

Dissertation

of

M.Sc. Álvaro Taboada Núñez
from Ourense, Galicia, Spain

Day of the oral examination: 29th of November 2019

Referee: Prof. Dr. Dr. h.c. Johannes Blümer

Co-referee: Dr. Alberto Etchegoyen

Advisor: Dr. Markus Roth

Analysis of the First Data of the AugerPrime Detector Upgrade

Zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften

von der KIT-Fakultät für Physik des
Karlsruher Instituts für Technologie (KIT)

und der

Universidad Nacional de San Martín (UNSAM)

genehmigte

Dissertation

von

M.Sc. Álvaro Taboada Núñez

aus Ourense, Galicien, Spanien

Tag der mündlichen Prüfung: 29. November 2019

Referent: Prof. Dr. Dr. h.c. Johannes Blümer

Korreferent: Dr. Alberto Etchegoyen

Betreuer: Dr. Markus Roth

Analysis of the First Data of the AugerPrime Detector Upgrade

Tesis presentada para optar por el título de

Doctor en Astrofísica

del Instituto de Tecnología "Prof. Jorge A. Sabato" de la
Universidad Nacional de San Martín (UNSAM)

y del

Karlsruher Institut für Technologie (KIT)

por

M.Sc. Álvaro Taboada Núñez

de Ourense, Galicia, España

Fecha de la defensa oral: 29 de noviembre de 2019

Director: Prof. Dr. Dr. h.c. Johannes Blümer

Co-director: Dr. Alberto Etchegoyen

Colaborador: Dr. Markus Roth

Abstract

Despite their discovery almost one hundred years ago, the physics of cosmic rays is still a mystery. Huge efforts have been employed over the last decades in order to understand their origin and propagation through the cosmos until they reach the Earth. Several astrophysical scenarios were developed aiming to explain how cosmic rays can be accelerated up to such high energies, which are orders of magnitude above the ones achieved by man-made accelerators. The different features observed in the measured energy spectrum such as the flux suppression at the highest energies or the region thought to be a transition between galactic and extra-galactic origin remain unclear. The need of models for the magnetic fields at the source environment and acting on the cosmic rays during propagation adds a further complication. The interpretation of measurements from extensive air showers induced by cosmic rays often requires the usage of simulations based on LHC-tuned models for the hadronic interactions. The observation of a deficit in the number of muons predicted by these models is also an unclear issue. Knowing the mass composition of cosmic rays at the top of the atmosphere is of key importance in order to answer these questions.

The Pierre Auger Observatory is currently the largest experiment dedicated to the measurement of ultra-high-energy cosmic rays. Covering more than 3000 km^2 with ground-based detectors, it employs a hybrid detection technique of air showers. The Fluorescence Detector uses telescopes for measuring the longitudinal profile of showers as they develop in the atmosphere. The Surface Detector measures the footprint of showers at the ground by sampling lateral distributions of particles with water-Cherenkov detectors (WCDs). This allows us to estimate the energy and the mass composition of the primary particle. However, the reduced duty cycle of the Fluorescence Detector limits the statistics of mass-sensitive measurements at the highest energies. In order to achieve the aforementioned goals, the observatory is currently undergoing a detector upgrade, named AugerPrime, which aims to enhance the mass-sensitivity of the Surface Detector by placing a plastic scintillator on top of each of the water-Cherenkov detectors. This Scintillator Surface Detector (SSD) provides with a complementary measurement thus allowing for the separation of the electromagnetic and muonic shower components. By the end of 2016, twelve Surface Detector stations were upgraded as part of an engineering phase. In addition, seventy-seven SSDs were deployed during March 2019 increasing the wealth of showers measured at high energies. The analysis of the data provided by these detectors is the final goal of this work.

The work presented in this dissertation includes methods developed for the reconstruction of air showers with the SSDs and the analysis of the first data recorded. 1. Studies of the calibration of the SSD by means of simulating the flux of secondary particles from low energy showers. 2. Inclusion of the SSD into the standard reconstruction methods with the Surface Detector. This implies the study of the lateral distribution of the SSD signals. 3. Development of algorithms for reconstructing the mass composition using information from both the WCD and the SSD reconstructions. 4. Analysis of first data delivered by upgraded detectors during the engineering phase. 5. First estimate of the mass composition using data from AugerPrime detectors.

The tools developed in this work will serve as a precedent for future analyses carried out in the context of mass composition with AugerPrime.

Zusammenfassung

Nach fast hundert Jahren ist die Physik der kosmischen Strahlung immer noch ein Geheimnis. In den letzten Jahrzehnten wurden große Anstrengungen unternommen, um ihren Ursprung und ihre Ausbreitung durch den Kosmos zu verstehen. Verschiedene astrophysikalische Szenarien wurden entwickelt, um zu erklären, wie kosmische Strahlung auf Energien beschleunigt werden kann, die um mehrere Größenordnungen über denen von künstlichen Beschleunigern liegen. Die verschiedenen Merkmale des gemessenen Energiespektrums wie beispielsweise die Flussunterdrückung bei den höchsten Energien oder der mögliche Übergangsbereich zwischen galaktischer und extra-galaktischer Strahlung sind bis heute unklar. Die Notwendigkeit von Modellen für die Beschreibung von Magnetfeldern in der Nähe von Quellen, sowie deren Auswirkung auf die Propagation der kosmischen Strahlung stellt eine weitere Komplikation dar. Die Interpretation der Messungen von ausgedehnten Luftschauern, die durch die kosmische Strahlung hervorgerufen werden, erfordert häufig den Einsatz von Simulationen deren Grundlage hadronische Wechselwirkungsmodelle bilden. Des Weiteren ist die Beobachtung eines Defizits in der Anzahl der Myonen, die von diesen Modellen vorhergesagt werden, ein ebenfalls ungeklärtes Thema. Um alle diese Fragen beantworten zu können, ist die Kenntnis der Massenzusammensetzung der kosmischen Strahlung von entscheidender Bedeutung.

Das Pierre Auger Observatorium ist derzeit das größte Experiment zur Messung von ultrahochenergetischer kosmischer Strahlung. Es deckt eine Fläche von mehr als 3000 km² mit bodengestützten Detektoren ab und verwendet eine hybride Messtechnik zur Detektion von Luftschauern. Der Fluoreszenzdetektor, bestehend aus 27 Teleskopen, beobachtet die longitudinalen Profile von Teilenschauern, die sich in der Atmosphäre entwickeln. Der Oberflächendetektor misst den Fußabdruck von Schauern am Boden, indem er die laterale Verteilung der Teilchen mit Hilfe von Wasser-Cherenkov-Detektoren (WCDs) erfasst. Diese kombinierte Analyse der Luftschauer ermöglicht eine Schätzung der Energien und der Massenzusammensetzung der ankommenden Primärteilchen. Die deutlich reduzierte Einsatzzeit des Fluoreszenzdetektors schränkt jedoch die Statistik der massenempfindlichen Messungen bei den höchsten Energien stark ein. Um die oben genannten Ziele zu erreichen, wird das Observatorium derzeit einer Aufrüstung der Detektoren unter dem Namen "AugerPrime" unterzogen, die darauf abzielt, die Massensensitivität des Oberflächendetektors zu erhöhen, indem auf jeden der Wasser-Cherenkov-Detektoren ein Kunststoff-Szintillator installiert wird. Dieser Szintillator-Oberflächendetektor (SSD) bietet dabei eine komplementäre Messung, die eine verbesserte Trennung der elektromagnetischen und muonischen Schauerkomponenten ermöglicht. Bis Ende 2016 wurden zwölf Oberflächendetektorstationen im Rahmen einer Test- und Entwicklungsphase modernisiert. Darüber hinaus wurden im März 2019 siebenundsiebzig SSDs in Betrieb genommen, die die Zahl an gemessenen Luftschauern bei den höchsten Energien deutlich erhöhen werden.

Die in dieser Dissertation vorgestellte Arbeit beinhaltet Methoden, die für die Rekonstruktion von Luftschauern mit Hilfe der neuen SSDs und für die Analyse der ersten aufgezeichneten Daten entwickelt wurden. 1. Studien zur Kalibrierung der SSDs mittels Simulation des Flusses von Sekundärteilchen aus niederenergetischen Schauern. 2. Aufnahme der SSDs in die Standard-Rekonstruktionsmethoden des Oberflächendetektors. Dies impliziert die Untersuchung der lateralen Verteilung der SSD-Signale. 3. Entwicklung von Algorithmen zur Analyse der Massenzusammensetzung unter Verwendung von Informationen aus den WCD- und SSD-Rekonstruktionen. 4. Analyse der ersten Daten der modernisierten Detektoren, die in der Testphase aufgenommen wurden. 5. Erste Schätzung der Massenzusammensetzung unter Verwendung der Daten der AugerPrime-Detektoren. Die in dieser Arbeit entwickelten Methoden werden als Präzedenzfall für zukünftige Analysen dienen, die im Rahmen der Bestimmung der Massenzusammensetzung mit AugerPrime durchgeführt werden.

Resumen

Los rayos cósmicos siguen entrañando misterios tras casi un siglo desde su descubrimiento. Cuestiones como su origen y propagación por el cosmos, han despertado el interés de generaciones de científicos y que a día de hoy todavía buscan respuesta. Numerosos marcos teóricos han sido desarrollados con el afán de entender cómo estas partículas alcanzan energías tan altas, las más elevadas conocidas en el Universo. Para explicar los diferentes aspectos de su espectro de energía, se precisa del entendimiento de los campos magnéticos que interfieren en su propagación y cuyo conocimiento es limitado.

La interpretación de las observaciones de las lluvias de partículas generadas por rayos cósmicos en la atmósfera, a menudo requiere de la utilización de simulaciones, las cuales dependen de modelos de interacciones hadrónicas. La observación de un déficit en el número de muones predicho por los modelos supone un grado de complejidad añadido que motiva la sinergia entre la física estudiada en los aceleradores de partículas y la astrofísica. El conocimiento de la composición química de los rayos cósmicos que llegan a la atmósfera es una pieza clave para dar respuesta a estos entresijos.

El Observatorio Pierre Auger es actualmente el experimento más grande del mundo dedicado a la observación de rayos cósmicos. Cubre un área de más de 3000 km², desplegada con detectores de partículas que emplean un método de detección híbrido. Por un lado, el Detector de Fluorescencia dispone de telescopios que miden el desarrollo longitudinal de las lluvias de partículas en la atmósfera. Por otro lado, el Detector de Superficie mide la densidad de partículas que llegan al suelo por medio de la utilización de detectores Cherenkov en agua (WCD). De este modo se pueden estimar la energía y la composición química de la partícula originaria de la lluvia. Sin embargo, el número de mediciones con información acerca de la composición a las más altas energías se ve limitado debido al reducido tiempo de operación del Detector de Fluorescencia. Es por ello que se está realizando una mejora del Detector de Superficie con el fin de incrementar su sensibilidad a la composición. Este proyecto recibe el nombre de AugerPrime y tiene como eje principal la instalación de un detector de centelleo (SSD) en cada uno de los detectores Cherenkov en agua. Los dos detectores proporcionan una medida complementaria, la cual facilita la separación entre las componentes muónica y electromagnética de la lluvia. A finales del año 2016, doce de estos detectores comenzaron la adquisición de datos. A mayores, durante marzo de 2019 setenta y siete más fueron desplegados, incrementando así el número de lluvias observadas a altas energías. Este trabajo tiene por objetivo principal el análisis de los datos de estos detectores.

El trabajo presentado en esta tesis doctoral incluye métodos desarrollados para la reconstrucción de lluvias medidas con los detectores de centelleo así como el análisis de los primeros datos obtenidos. Se abordan los siguientes puntos: 1. Estudio de la calibración del SSD en base a simulaciones. 2. Implementación del SSD en la cadena de reconstrucción del Detector de Superficie. 3. Desarrollo de algoritmos para el estudio de la composición química utilizando el WCD y el SSD. 4. Análisis de los primeros datos obtenidos por los detectores durante la fase de prueba. 5. Primera estimación de la composición química utilizando datos de los detectores de AugerPrime.

Las herramientas desarrolladas en el marco de este trabajo suponen un precedente para para los futuros análisis llevados a cabo en el contexto de AugerPrime.

Contents

1	Introduction	1
1.1	Cosmic rays	1
1.1.1	The flux of cosmic rays	1
1.1.2	Propagation and candidate sources	3
1.1.3	Extensive Air Showers	4
1.2	The Pierre Auger Observatory	6
1.2.1	Surface detector	7
1.2.2	Fluorescence detector	9
1.2.3	AMIGA muon detector	10
1.2.4	Radio detector	11
1.3	Surface detector reconstruction	11
1.3.1	Trigger chain	12
1.3.2	Shower geometry	13
1.3.3	Signal fluctuations	14
1.3.4	Fit of the Lateral Distribution	14
1.3.5	Energy	17
1.4	Selected results from the Pierre Auger Observatory	18
2	AugerPrime Upgrade	21
2.1	Physics motivation and goals for the upgrade	22
2.2	Components of the upgrade	25
2.2.1	The Scintillator Surface Detector	25
2.2.2	Small Photo-Multiplier Tube	33
2.2.3	Surface Detector Electronics Upgrade	34
2.2.4	Further upgrades	35
3	Event reconstruction and mass composition	37
3.1	Library of simulations	38
3.2	Towards a reconstruction with the SSD	38
3.2.1	Describing the SSD signals	38
3.2.2	Uncertainty model for the SSD signals	39
3.2.3	Obtaining the Lateral Distribution Function for the SSD	43
3.2.4	Reconstruction efficiency	47
3.3	Estimate of the energy	50

3.3.1	Attenuation correction to S_{1000}	51
3.3.2	Unbiased energy estimate	52
3.4	Sensitivity to mass composition	57
3.4.1	Separation power at a fixed energy	58
3.4.2	Results from Principal Component Analysis	61
4	Data analysis from AugerPrime detectors	67
4.1	Configurations of the detectors in the field	68
4.2	Calibration and performance of the upgraded stations	69
4.3	Update of the <u>Offline</u> framework	72
4.4	First measured showers	73
4.4.1	Station level evaluation	74
4.4.2	Event level evaluation	76
4.5	First estimate of mass composition with AugerPrime	82
4.5.1	Selection cuts	82
4.5.2	Mean composition	84
4.5.3	Results from previous measurements	84
4.5.4	Sources of systematic uncertainties	86
5	Summary	91
A	Proposals for the Detector Upgrade	103
B	Simulations and <u>Offline</u> Sequences	107
B.1	Simulations of low energy showers	107
B.1.1	Studies of the calibration of the Scintillator Surface Detector	107
B.2	Sequence for reconstruction of events with AugerPrime detectors	109
C	Analysis	111
C.1	Chapter 1	111
C.1.1	The Cherenkov radiation	111
C.2	Chapter 3	118
C.2.1	Signal uncertainty model	118
C.2.2	Distance cut for rejecting low signals	119
C.2.3	Parameterization of the LDF for the SSD	120
C.2.4	Validation of the parameterization	121
C.2.5	Non-linear ansatz for mass-independent energy bias	122
C.2.6	Uncertainties on the merit factor	124
C.3	Chapter 4	126
C.3.1	Mean composition	126
D	Acknowledgments	129

CHAPTER 1

Introduction

1.1 Cosmic rays

Cosmic rays are messengers of the universe, relativistic charged particles whose origin remains unclear. They are ionized nuclei, most of them protons and alpha particles, but also heavier elements, and they hit the Earth's atmosphere at a rate of about 1000 per square meter per second. Some of them reach energies up to 10^{20} eV (about 20 Joules), being by far the most energetic particles ever known. The answer to the question of how they are accelerated to such high energies is another challenge to be solved within the next years. In this section, we aim to give a general overview on cosmic rays. In particular we focus on those of energies above 10^{18} eV, the so-called ultra-high energy cosmic rays (UHECR), as these are of the interest of this thesis. Detailed descriptions and discussions may be found in [1].

1.1.1 The flux of cosmic rays

The flux of cosmic rays covers more than ten orders of magnitude in energy, from less than 10^9 eV to more than 10^{20} eV. The dependency of the flux with the energy can be written as a power-law, $E^{-\gamma}$, where γ is an spectral index of value close to 3. This means that the flux decreases by three orders of magnitude per decade of energy, i.e., from one particle per square centimeter per second at around 100 MeV to less than one particle per square kilometer per century at around 100 EeV ($1 \text{ EeV} = 10^{18} \text{ eV}$). In order to look from deviations of a single-index power law, it is common to show the flux scaled by the energy to some power as depicted in Fig. 1.1. Three different features of the spectrum are evident. A steepening of the spectrum at energies of about 3×10^{15} eV is known as the *knee*, followed by a flattening at around 5×10^{18} eV known as the *ankle*, and a strong *suppression* of the flux at the highest energies, above 4×10^{19} eV.

The knee is mostly attributed to galactic cosmic rays which may be originate from diffuse particle acceleration mechanisms taking place in supernova remnants (SNRs). The steepening at the knee can be associated with the maximum potential of galactic accelerators or a rigidity-dependent change in propagation. In both cases, a change from light toward a heavy composition is supported by the measurements of the KASCADE experiment [2]. Another steepening at energies of around 8×10^{16} eV, known as the *second knee*, was reported by measurements of KASCADE-Grande [3]. Whereas the knee

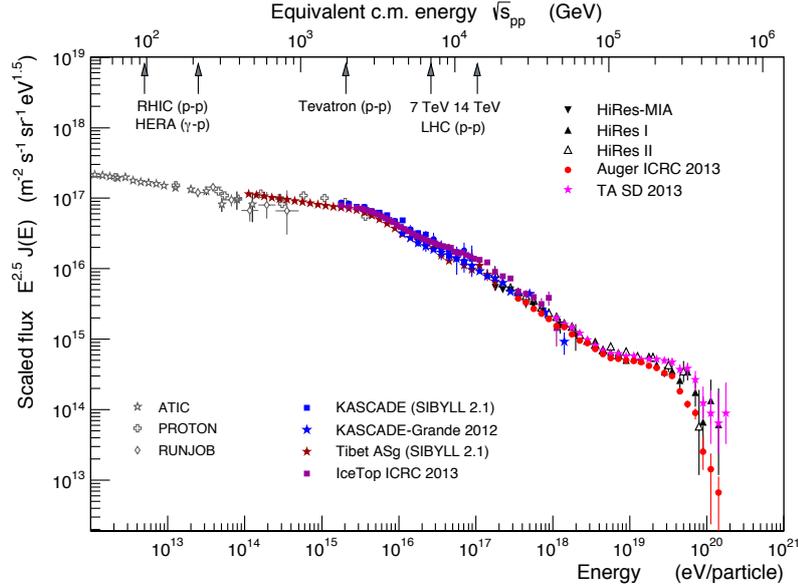
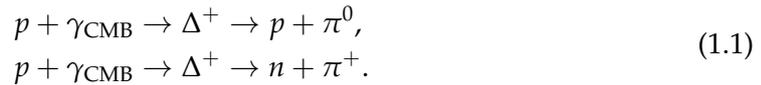


Figure 1.1: Energy spectrum of cosmic rays measured by different experiments. The flux is scaled by a factor $E^{2.5}$ to enhance the visibility of the spectral features. The upper axis shows the equivalent energy in the center of mass of different man-made accelerators. Courtesy of Ralph Engel.

is associated with light primary masses, the second knee is thought to be related with the decrease in the flux of heavy elements.

The ankle is thought to be a region of transition between galactic to extragalactic sources. Currently, there are mainly two competing models for the description of this feature. The first one, known as the *dip model* assumes that the ankle is a consequence of the flux of extragalactic protons which interact with photons of the cosmic microwave background (CMB) producing e^+e^- pairs. This model allows for a mixture of heavier nuclei of the order of 20%, but the main composition up to the region of the suppression must be due to protons. On the other hand, the *mixed composition model* assumes that the galactic component is dominant at energies before the ankle and a smooth transition to extragalactic elements occurs at the ankle. Regardless of the mechanism, it is assumed that extragalactic sources must be behind the flux of cosmic rays above the ankle due to the lack of any directional evidence for the production of these cosmic rays in the galactic plane. More information may be found in [4–7].

The reason for the suppression at the highest energies remains a mystery. It is still unclear whether it is mainly due to propagation effects or to extragalactic sources reaching their maximum for acceleration. The first case can be understood given the interaction of ultra-high energy protons with photons of the CMB. This interaction leads to a Δ^+ resonance through the following processes



In both cases, the primary protons lose about 20% of their original energy. This is known as the GZK¹ effect [8, 9]. For heavier nuclei, photo-disintegration processes occur resulting in a similar effect.

Alternatively, the flux suppression might be just a realization of the maximum energy of accelerators that depends on the rigidity, $R = E/Z$. Under this assumption, sources

¹Greisen-Zatsepin-Kuzmin.

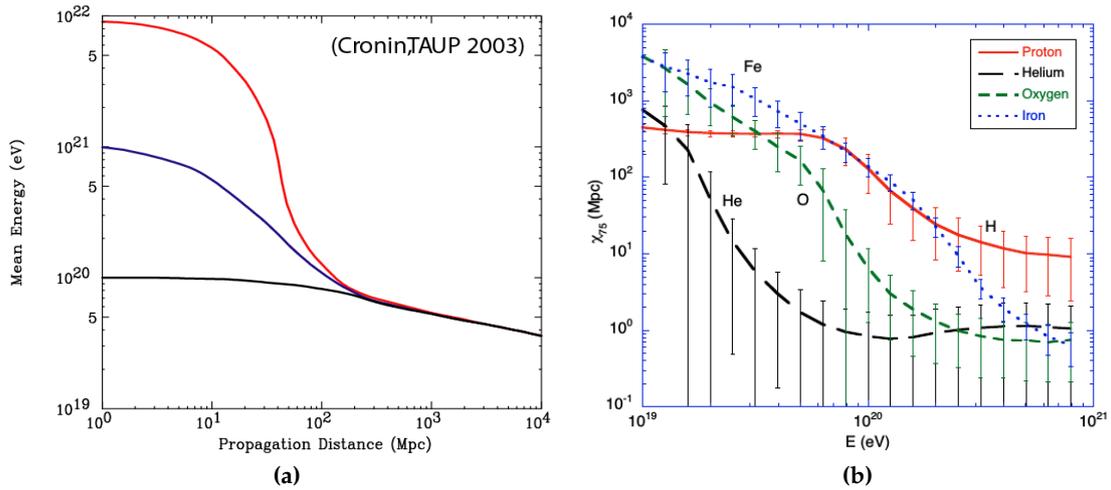


Figure 1.2: (a) Mean energy of protons as a function of propagation distance through CMB. Curves are for energy of 10^{20} eV, 10^{21} eV and 10^{22} eV at the source. Taken from [10]. (b) Energy loss length for different nuclei as a function of the energy. Taken from [11].

accelerate first lighter elements and then heavier ones. The observed suppression of the flux at the highest energies is then the result of a superposition of different energy cut-offs at the source spectrum. The knowledge of the mass composition of UHECR is crucial to determine the cause of the suppression. This questions will be addressed again in Chapter 2.

1.1.2 Propagation and candidate sources

The interaction of cosmic rays with the CMB and magnetic fields plays a key role in their propagation. It was already mentioned that a proton interacting with a CMB photon (see Eq. (1.1)) will get its initial energy reduced. The mean energy of protons as a function of the propagation distance through the CMB is shown in Fig. 1.2a. It can be seen that after about 100 Mpc, the proton energy drops considerably. Similarly, heavier nuclei at such high energies undergo photo-disintegration processes within the same radius as depicted in Fig. 1.2b. This limit is referred to as the “GZK horizon”, resulting in an opacity of the universe to UHECR.

As they are charged particles, cosmic rays suffer from the effect of magnetic fields present in the source environment as well as in the intergalactic medium and in our Galaxy. The deviation in their trajectories can be of a few degrees depending on their charge and energy [12], which turns into a complication when estimating their point of origin. Nevertheless, efforts have been made in order to classify potential astrophysical sources of cosmic rays. As first order, one could estimate the maximum energy to which a particle can be accelerated as

$$E_{\max} \approx Z \left(\frac{B}{1 \mu\text{G}} \right) \left(\frac{R}{1 \text{kpc}} \right) \text{EeV}, \quad (1.2)$$

where R is the extension of the acceleration region, B the strength of the magnetic field and Z the charge. According to this criterion different candidate sources are classified in the so-called “Hillas plot” shown in Fig. 1.3.

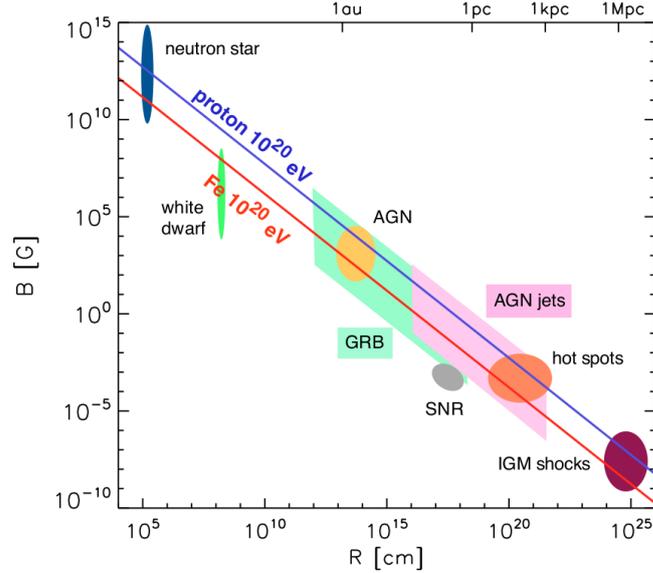


Figure 1.3: Hillas plot. It shows potential astrophysical sites for acceleration of protons and iron nuclei up to a maximum energy of 10^{20} eV as a function of their size (x -axis) and the strength of the magnetic field (y -axis). Taken from [12].

1.1.3 Extensive Air Showers

Due to their low flux, the detection of UHECR is performed indirectly by observing the *cascade* of particles produced when they interact with a molecule of nitrogen or oxygen in the atmosphere. These are called extensive air showers (EASs) and can be divided in three parts or components: the electromagnetic, the muonic and the hadronic component. An illustration is given in Fig. 1.4. The core of the shower consists of high-energy hadrons and nucleons that constantly feed the electromagnetic component, mostly from the decay of neutral pions and eta particles into photons. Each of this photons starts an electromagnetic sub-shower as a consequence of e^+e^- production and bremsstrahlung. In addition, low energy pions and kaons decay creating the muonic component of the shower. In this section each of this components are summarized. For a detailed description see, for example, chapter 16 of [1].

The electromagnetic component

A simple model was introduced by Heitler [14] in order to illustrate the development of an electromagnetic cascade. The basics can be understood mainly through two main processes: an electron (same holds for a positron) producing one photon via bremsstrahlung, and a photon producing an electron-positron pair. In this model, after one collision length, λ , a branching process occurs where the energy of the particle (electron or photon) is split in two outgoing particles. After $n = X/\lambda$ branchings the total number of particles (electrons and photons) is

$$N(X) = 2^{X/\lambda}, \quad (1.3)$$

where X is the slant depth and its units are given in g cm^{-2} . This splitting process continues until the energy of the particles drops below a critical value E_c , which is determined as the point where particles only lose energy via inelastic collisions, and no further particles are created. The value for the critical energy in air is $E_c \approx 87$ MeV. Considering a shower initiated by a single photon with energy E_0 , the shower reaches its maximum

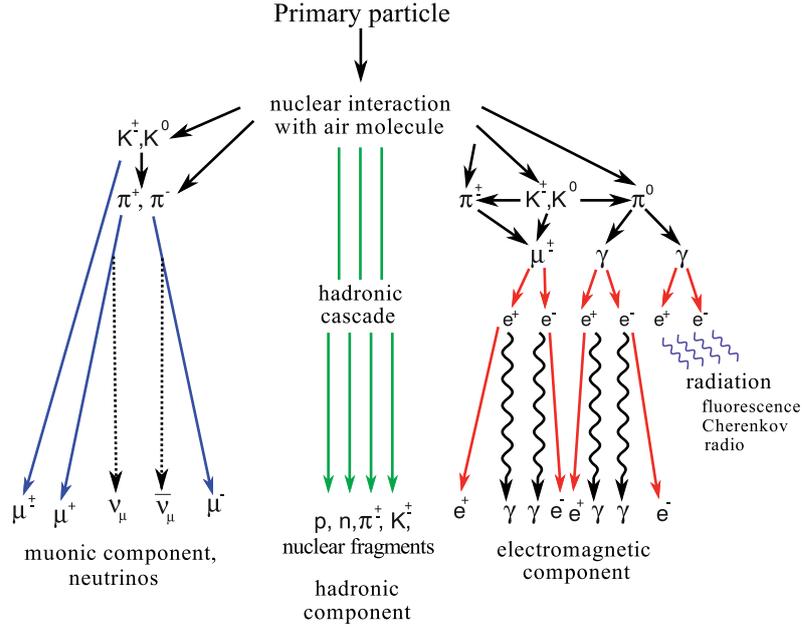


Figure 1.4: Schematic of the internal components of an EAS. Taken from [13].

when all particles have energy E_0 . Therefore the number of particles at the shower maximum is

$$N_{\max} = E_0/E_c, \quad (1.4)$$

and the depth at the maximum of the shower

$$X_{\max} = \lambda \frac{\ln(E_0/E_c)}{\ln(2)}. \quad (1.5)$$

The essential features of this model is that the number of particles at the maximum is proportional to E_0 and the depth of maximum increases logarithmically with the energy.

The hadronic component

A generalized version of the Heitler model was proposed by Matthews [15] to explain the physics of the hadronic shower. This can be described, in general, as processes of pion production. Two-thirds of particles are charged (π^\pm) and one-third are neutral (π^0). As neutral pions immediately decay ($\pi^0 \rightarrow \gamma + \gamma$), after n generations the initial energy E_0 is split into the hadronic and the electromagnetic components as

$$E_{\text{had}} = \left(\frac{2}{3}\right)^n E_0, \quad E_{\text{em}} = \left[1 - \left(\frac{2}{3}\right)^n\right] E_0. \quad (1.6)$$

Charged hadrons continue interacting until their energy is lower than the decay energy $E_{\text{dec}} \sim 30 \text{ GeV}$ [16]. Once the energy of hadrons is below E_{dec} they decay producing muons and neutrinos, e.g.,

$$\pi^- \rightarrow \mu^- + \bar{\nu}_\mu, \quad (1.7)$$

$$\pi^+ \rightarrow \mu^+ + \nu_\mu. \quad (1.8)$$

These decays feed the muonic component. The number of muons produced in the shower can be estimated as

$$N_\mu = \left(\frac{E_0}{E_{\text{dec}}} \right)^\alpha, \quad \alpha = \frac{\ln n_{\text{ch}}}{\ln n_{\text{tot}}} \approx 0.82 \dots 0.9, \quad (1.9)$$

where n_{ch} is the number of charged particles and n_{tot} the total number of particles. High-energy neutrinos and muons are responsible for the *invisible* energy carried to the ground, which represents approximately 10% of the primary energy.

Since the binding energy per nucleon (~ 5 MeV) is much smaller than the typical interaction energies, one can consider a nucleus as a bunch of independent nucleons. In the so-called superposition model, a nucleus of mass A and energy E_0 is considered as A independent nucleons of energy E_0/A . Under this assumption one can find that the number of muons produced in an air shower originated by a nucleus is

$$N_\mu^A = A \left(\frac{E_0/A}{E_{\text{dec}}} \right)^\alpha = A^{1-\alpha} \left(\frac{E_0}{E_{\text{dec}}} \right)^\alpha. \quad (1.10)$$

Therefore the heavier the nucleus, the more muons are produced. On average, iron showers contain about 40% more muons than proton showers.

Longitudinal development and mass composition

The measurement of the longitudinal development of the shower is specially important for determining the mass of the primary. The atmosphere acts as a huge calorimeter that drives the produced shower particles and allows for their detection via fluorescence or Cherenkov light emitted. The longitudinal profile is studied as a function of the traversed column density

$$X = \int \rho(l) dl, \quad (1.11)$$

being ρ the density of air. The depth in the atmosphere at which the shower reaches its maximum of development is called X_{max} . Using the aforementioned superposition model, one can find a relation between X_{max} and the primary energy and mass (see [17])

$$\langle X_{\text{max}} \rangle = X_0 + D \log \left(\frac{E}{E_0 A} \right) + \zeta \ln A + \delta \ln A \log \left(\frac{E}{E_0} \right), \quad (1.12)$$

where X_0 is the mean depth for proton showers at energy E_0 and D is the elongation rate (i.e., the change of X_{max} with energy). The parameters ζ and δ represent deviations from the ideal superposition model. This means that for the same energy, iron showers reach their maximum higher in the atmosphere, of about 100 g cm^{-2} [18].

1.2 The Pierre Auger Observatory

The Pierre Auger Observatory [19] is the world's largest observatory dedicated to the measurement of UHECR. It was originally designed to study cosmic rays of energies above 10^{18} eV [20] by employing a hybrid detection technique by means of a Surface Detector (SD) [21] and a Fluorescence Detector (FD) [22]. The SD consists of a vast array of water-Cherenkov Detectors (WCDs) which measure the energy deposited by particles from air showers at ground. The FD is composed by 27 telescopes deployed at four different locations overlooking the SD. The telescopes measure the fluorescence light emitted by particles as the shower develops in the atmosphere.

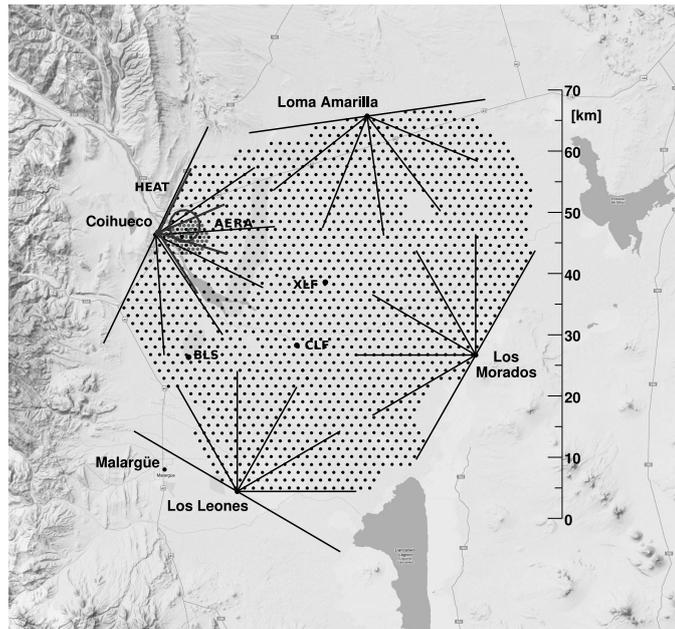


Figure 1.5: Layout of the Pierre Auger Observatory. Each dot corresponds to one SD station. Lines represent the field of view of the fluorescence telescopes, six on each location with the addition of 3 high-elevation telescopes for the observation of showers induced by lower energy primaries. A denser part of the array is indicated by gray dots. Each of these stations will be equipped with underground muon detectors for the direct measurement of the muonic component of the shower. An engineering array of radio antennas is also located in that area. Further details are given in the text. Courtesy of Darko Veberič.

The observatory is situated near the town of Malargüe, Argentina, next to the beautiful mountain range of the Andes. A layout of the observatory can be seen in Fig. 1.5. It is located at an elevation of approximately 1400 m or 875 g cm^{-2} in atmosphere overburden, which allows for the measurement of EAS at its maximum of development at the desired energies of observation. The appropriate atmospheric conditions such as low levels of aerosols and light pollution in addition to low cloud coverage, makes the Argentinian region of the Pampa Amarilla to be an excellent emplacement for the observatory.

1.2.1 Surface detector

The SD of the Pierre Auger Observatory is composed by more than 1660 WCDs (also referred to as stations) deployed in an array with 1500 m spacing (SD-1500) covering a total area of around 3000 km^2 . The deployment of the SD stations started in 2004 and finished in 2008. The energy threshold at which the SD-1500 becomes full efficient is $3 \times 10^{18} \text{ eV}$ [23]. The station consists of a cylindrical polyethylene tank of about 10 m^2 which is filled with 12 tons of purified water up to a height of 1.2 m. The water is contained in a highly-reflective liner inside the polyethylene enclosure. When relativistic charged particles from EASs pass through the water, they produce Cherenkov photons which are then collected by three 9-inch photo-multiplier tubes (PMTs) placed in transparent windows at the top of the liner (a detailed description of the WCD response is given in Appendix C.1.1). A picture and an schematic of a WCD can be seen in Fig. 1.6. Each PMT has two outputs, one in the anode (referred as to the low gain channel) and an amplified one in the last dynode (also called high gain channel). There is a factor of 32 in the charge

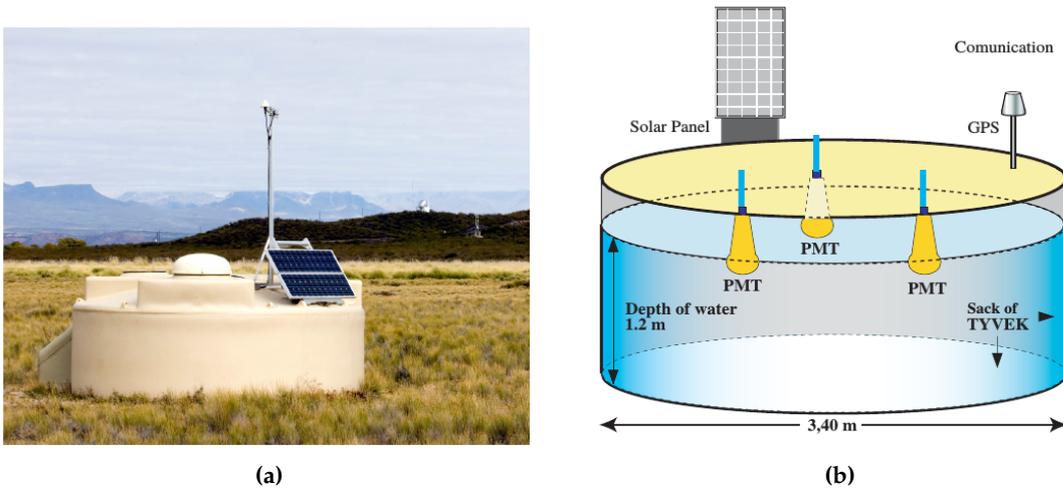


Figure 1.6: (a) Picture of a WCD. (b) Schematic of a WCD with different components. Taken from [24].

gain between the two channels enabling a higher dynamic range in order to measure signals close to the shower core.

The electronics of the surface detector (referred to as the Unified Board (UB)), sample the analog pulses produced at the base of the PMTs with two flash analog to digital converters (FADCs) of 10 bit at a frequency of 40 MHz. The digitized signals are stored in a cyclic memory and the electronics checks whether a trigger condition is fulfilled (see Section 1.3.1). If that is the case, a block of 768 time bins is transferred to the central data acquisition system (CDAS) for the subsequent analysis. This block is known as FADC (or time) trace and there are six of them (two per PMT). The size of the time bins is defined by the sampling rate of the electronics, namely

$$1 \text{ bin} \equiv \frac{1}{40 \text{ MHz}} = 25 \text{ ns}, \quad (1.13)$$

such that the total duration of the trace is $768 \times 25 \text{ ns} = 19.2 \mu\text{s}$. The amplitude of the signal is measured in terms of FADC counts, ranging from 0 to $2^{10} - 1 = 1023$ counts. The converter has an input range between 0 V to 2 V, therefore the following relation is given

$$1 \text{ FADC count} \equiv \frac{2 \text{ V}}{2^{10}} = 1.95 \text{ mV}. \quad (1.14)$$

An example of a FADC trace is given in Fig. 1.7a. Further details on the signal calibration are given below.

Calibration of the SD signals

The flux of particles that reach the ground is constantly measured and is used to provide a common reference for all signals recorded by the SD stations. Signals measured by the WCD are expressed in units of vertical-equivalent muon (VEM), which corresponds to the average signal resulting from a vertical and centered muon traversing the tank, measured as the integrated pulse of the PMT over time, thus a “charge”. An estimation of the VEM can be obtained from the spectrum of signals generated by atmospheric muons passing through the tank. An example of a charge distribution in a WCD can be seen in

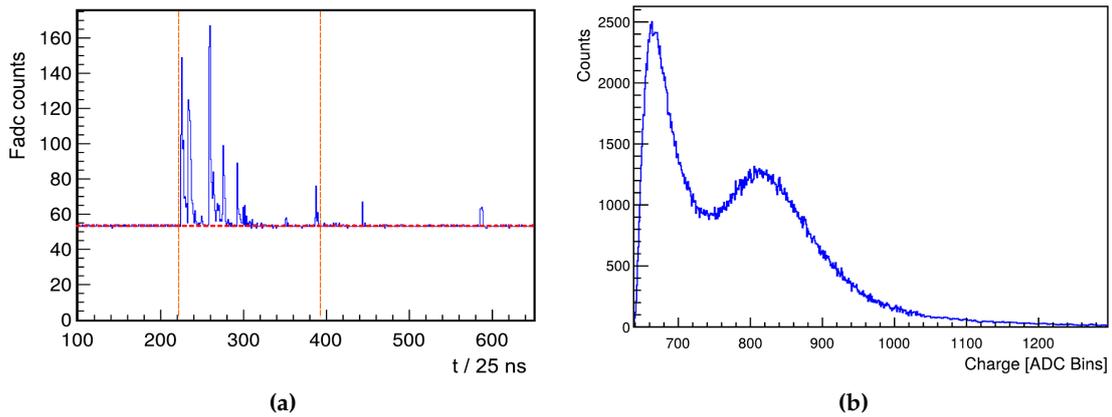


Figure 1.7: (a) Example of a FADC trace. The baseline noise is indicated by the horizontal dashed line. Vertical lines indicate the start and stop time bins. (b) Charge distribution measured by a SD station. The calibration constant is taken from the fit of the second peak in the charge distribution, in this case of the order of 800 ADC counts. Both taken from [25].

Fig. 1.7b. Such histograms exhibit two characteristic peaks. The first peak is due to small signals which are produced mainly by electromagnetic particles (e^\pm , γ) due to their low average energy deposited in water. The second peak is produced by atmospheric muons where the integrated signal is the result of a convolution of effects related to the muons' angular and energy spectra. By fitting the muon peak position, the value of the VEM charge (Q_{VEM}) is obtained [26]. The calibrated signal is obtained by the integral of the time trace normalized by the value of Q_{VEM} , namely

$$S = \sum_{t=t_s}^{t_f} (s(t) - b(t)) / Q_{\text{VEM}}, \quad (1.15)$$

where $s(t)$ and $b(t)$ refers to the signal and baseline in units of FADC counts at a given time interval t , and t_s and t_f denote the start and stop time of the integration window and are determined by the PMT with the largest signal.

Extensions to the SD array

In order to observe showers induced by lower energy primaries, the SD was extended by a nested array with 750 m spacing (SD-750) lowering the threshold to 3×10^{17} eV. The SD-750 consists of 71 stations covering an area of 27 km². The data acquisition started in September 2007 and the final deployment finished in September 2012. The implementation of additional software triggers in July 2013 lowered the threshold of full efficiency for the SD-750 to $10^{17.2}$ eV [27].

An hexagonal cell with 433 m spacing (SD-433) was deployed by January 2013, using the SD-750 station with ID 1764 as a central station. The aim of the SD-433 is to measure showers of energies below 10^{17} eV, in order to study the region of the second knee of the spectrum. First results have shown an energy extension down to 6×10^{16} eV [28].

1.2.2 Fluorescence detector

The standard FD consists of 24 telescopes which are situated in four locations overlooking the SD: Los Leones, Los Morados, Loma Amarilla and Coihueco. On each site, there is an FD building containing six fluorescence telescopes covering a total field of view (FoV)

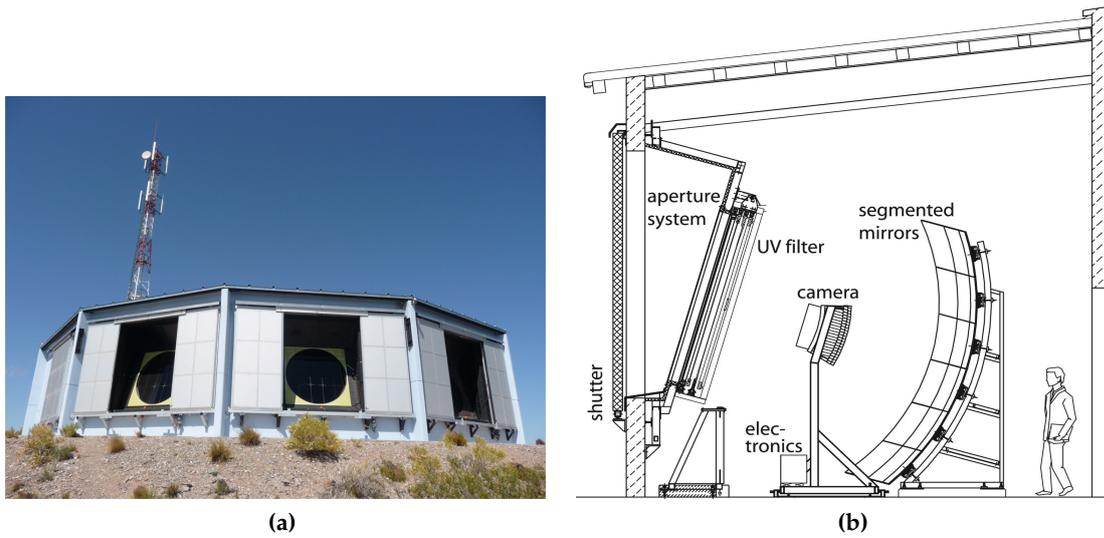


Figure 1.8: (a) Picture of the FD building at Los Leones. (b) Schematic of a telescope. Both taken from [19].

of 180° in azimuth and 30° in elevation². A picture of an FD building during the day can be seen in Fig. 1.8a. Three additional telescopes with elevated FoV were deployed near the FD site at Coihueco (see Fig. 1.5). The so-called High Elevation Auger Telescopes (HEAT) are similar to the standard FD telescopes but they can be tilted up to a maximum of 60° , allowing for the measurement of shallower showers initiated by lower energy primaries [29]. While the standard FD measures in combination with the SD-1500, HEAT forms a hybrid detector together with the SD-750.

During their propagation to the ground, charged particles from the EAS excite the nitrogen molecules present in the air, which then de-excite by emitting fluorescence light in the range of 300 nm to 400 nm. The light reaches the telescopes and, after passing a UV transmitting filter it is focused by a 10 m^2 mirror onto a camera. This camera is comprised by a grid of 22×20 hexagonal PMTs, where each PMT represents one pixel of the camera. A schematic of a telescope is depicted in Fig. 1.8b. Since the FD cameras are highly sensitive, the running time of the FD is limited to clear and almost moon-less nights. Another limitation is imposed by the weather conditions such as rain, snow or heavy winds that might damage the detector. All this results in an estimated duty cycle of around 15%.

The FD measures the longitudinal development of the shower in the atmosphere. The calorimetric energy can be estimated from the integral over the longitudinal energy deposit profile

$$E_{\text{cal}} = \int_0^\infty dE/dX(X)dX. \quad (1.16)$$

In addition, from the shower profile an estimate of X_{max} can be obtained. The energy of a shower can be determined with the FD with a precision of 8% above 10 EeV. For the shower maximum, the resolution is of 20 g cm^{-2} [30, 31].

1.2.3 AMIGA muon detector

The Auger Muon Detectors for the Infill Ground Array (AMIGA) enhancement was designed for the direct measurement of the muonic component of the shower. It consists

²Each of the telescopes has a FoV of $30^\circ \times 30^\circ$.



Figure 1.9: (a) Photograph of AMIGA detectors during their installation in the field. Modules of plastic scintillators are buried at around 2.3 m. The tubes are used for accessing the electronics. Taken from [34]. (b) Picture of an antenna of the AERA radio detector.

of modules of scintillator detectors (also known as Muon Detector (MD)) buried at a depth of around 2.3 m next to the stations of the SD-750 [32]. A picture of AMIGA modules being deployed can be seen in Fig. 1.9a. The overburden of earth above the AMIGA detectors serves as a natural shielding for the electromagnetic particles of the shower, and imposes a cut-off for vertical muons of 1 GeV [33]. The MDs have an area of 30 m² and are split into modules, each of them segmented in 64 bars of plastic scintillator with an embedded wave-length shifting (WLS) optical fiber intended to transport the scintillating light produced by passing muons, to optical sensors from which the signal is read.

A prototype phase (finished in 2017) equipped modules of 5 m² and 10 m² with silicon photomultipliers (SiPMs) and PMTs as optical devices. Results from the engineering phase can be found in [35–37], being the SiPMs selected as baseline design for the full deployment of AMIGA detectors. As it will be discussed in Chapter 2, each station of the SD-750 will be equipped with such detectors, allowing for the direct measurement of the muon content in showers induced by primary cosmic rays at energies close to the ankle.

1.2.4 Radio detector

The radio emission in air showers is also studied by the Pierre Auger Observatory. The Auger Engineering Radio Array (AERA) was built in 2009 inside the SD-750 enabling a complementary measurement of the shower in the radio regime. The array consists of 153 antennas of two different types, one of them is depicted in Fig. 1.9b, which operate in the frequency range of 30 MHz to 80 MHz. The detector can be operated in hybrid mode with the SD, the FD and the MD, and various results can be found in [38–40].

1.3 Surface detector reconstruction

The reconstruction of air shower events with the SD of the Pierre Auger Observatory is performed in different steps, from the event selection based on triggers of individual stations, to the reconstruction of the energy and arrival direction of the shower. In this section these steps are summarized. The information presented here was compiled from

different sources. Summarized descriptions can be found in [25, 41]. For more detailed descriptions on the SD reconstruction see, for example, [42, 43].

1.3.1 Trigger chain

First of all, a hardware-based trigger (T1) is applied at each station in order to reject the background flux of particles produced in low energy showers. The T1 trigger requires either a coincidence of the three PMTs having a signal above a threshold value of 1.75 VEM ($Thr1$) or a coincidence of two PMTs with more than 12 FADC bins above 0.2 VEM (after baseline subtracted) in a time window of 120 bins. If the T1 trigger condition is passed, the station software processes the T2 trigger which also has two modes:

1. Threshold trigger ($Thr2$): requires the coincidence of the three PMTs above 3.2 VEM. The $Thr2$ trigger has a frequency of 20 Hz and its main objective is to detect signals from muons in highly inclined showers.
2. Time over threshold trigger (ToT): is based on the coincidence of at least two PMTs having more than 13 bins above 0.2 VEM within a time window of 120 bins (similar to the T1). The rate of ToT triggers is between 1 Hz and 5 Hz.

If either of the T2 trigger conditions are fulfilled, the station notifies the CDAS to evaluate the possibility of an event trigger. The first trigger above the single station level is the T3. It searches for time coincidences in the signals of the T2 triggers previously sent to the CDAS. This is done to associate these stations with events from air showers. There are different allowed conditions:

1. 3-fold coincidence ($T3-3ToT$): requires the coincidence of 3 neighboring stations with ToT triggers.
2. 4-fold coincidence ($T3-4T2$): requires 4 tanks with T2 to be in coincidence.

The next level is the T4 trigger. It comprises a physics selection which also has two (more strict) variants:

1. $T4-3ToT$: 3 neighboring stations satisfying the ToT T2 trigger must form either an equilateral triangle (each station separated from another by the array spacing) or isosceles triangle. An schematic can be seen in Fig. 1.10a.
2. $T4-4C1$: 4 stations satisfying either of the T2 trigger conditions, where three of the stations are in the first crown of a fourth station. A depiction can be seen in Fig. 1.10b.

A further requirement implies that the difference between the start times of the selected stations for the T4 has to be smaller than d/c , where d is the distance between stations and c is the speed of light. Stations that do not fulfill this criteria are rejected. One example is the case of *accidental stations* which happens when a background muon triggers a distant station that is in time coincidence with the start times of the other stations in the event. Finally, a T5 trigger can be applied prior or post-reconstruction. A 6T5 trigger condition requires that the station with largest signal in the event (the *hottest* station) is surrounded by six functioning stations (not necessarily triggered). The purpose is to avoid events that landed in the edges of the array as they might bias the reconstruction. For spectrum analysis, the 6T5 condition on the events is employed. For anisotropy studies the trigger is relaxed to a 5T5 in order to increase the amount of events at the highest energies.

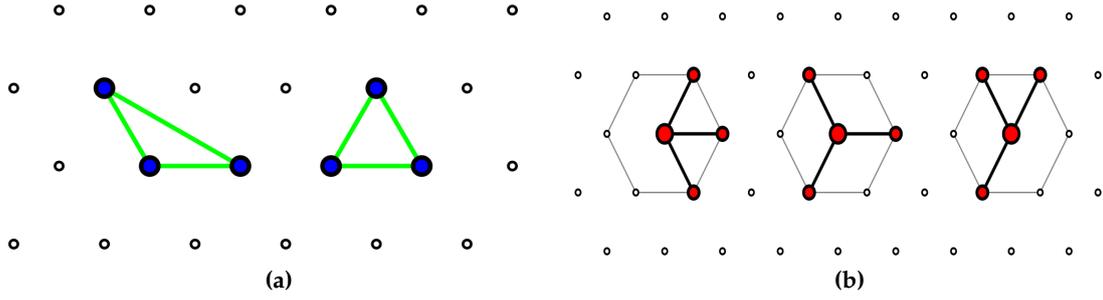


Figure 1.10: (a) Schematic depiction of the two possible configurations of the $T4-3ToT$ trigger. (b) Schematic depiction of the three possible configurations of the $T4-4C1$ trigger.

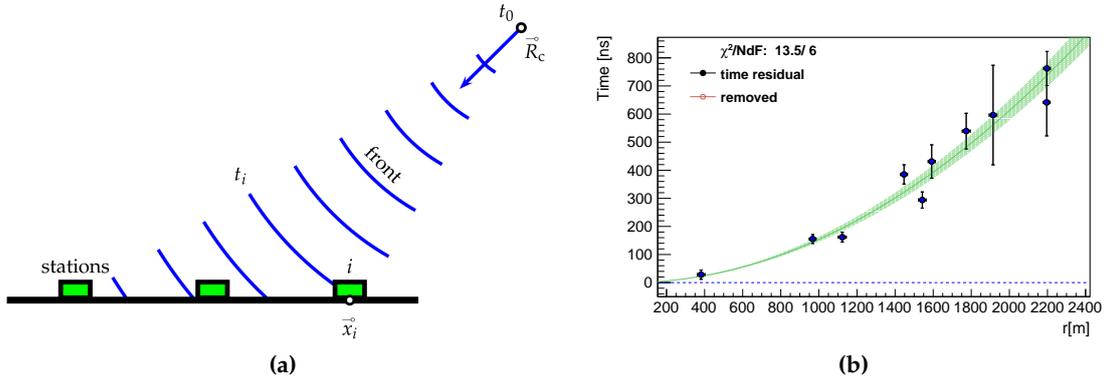


Figure 1.11: (a) Schematic depiction of a spherical shower front. (b) Example of a fit of station times to a curved shower front.

1.3.2 Shower geometry

In order to find the shower geometry with the SD a first approximation is done fitting station start times to a plane moving along the axis \hat{a} at the speed of light c

$$\vec{x}(t) - \vec{b} = -c(t - t_0)\hat{a}, \quad (1.17)$$

where the signal weighted barycenter of triggered stations, \vec{b} , is set as the spatial origin and the weighted time, t_0 , is used as time origin.

A more precise description of the shower is given by a spherical model. This description is used in a further step if there are at least four triggered stations and if the fit of the lateral distribution function was successful (see Section 1.3.4). The spherical approximation can be written as

$$c(t_i - t_0) = |\vec{R}_c - \vec{x}_i|, \quad (1.18)$$

with t_i representing the station trigger times at positions \vec{x}_i and \vec{R}_c is used as a virtual origin of the spherical front and can be related to the shower axis by

$$\vec{R}_c = \vec{c} + R_c\hat{a}, \quad (1.19)$$

being R_c the radius of curvature of the shower at the core position, \vec{c} , which has to be estimated beforehand. A depiction of the spherical front can be seen in Fig. 1.11a and a fit to the trigger start times in a real event is shown in Fig. 1.11b.

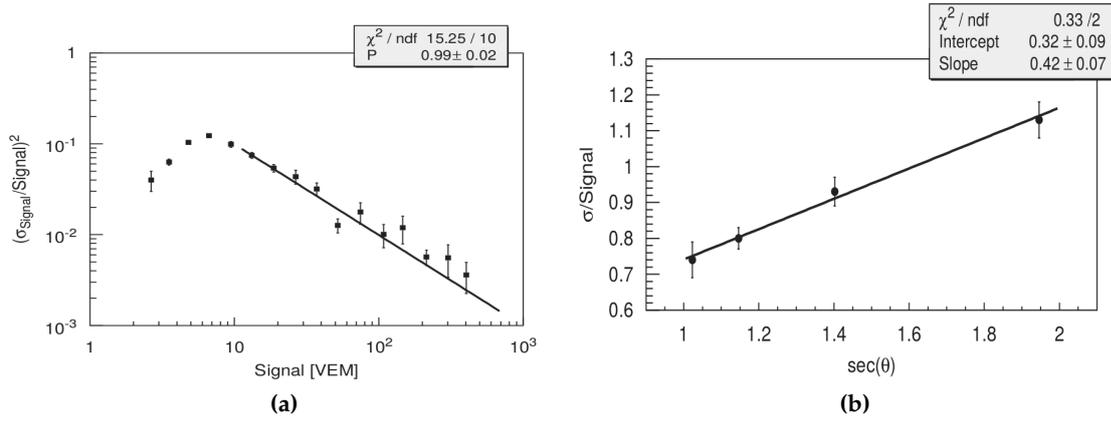


Figure 1.12: (a) Signal accuracy for the WCD. The linear fit (indicated by the line) corresponds to the assumption $\sigma/S = P/\sqrt{S}$. The turn at low signals is due to the observed trigger efficiency. (b) Dependency of the relative variance as a function of $\sec \theta$. Plots taken from [45].

1.3.3 Signal fluctuations

The understanding of any statistical and systematic effects underlying the measurements of air showers is crucial for a proper reconstruction of the primary properties. A dedicated study of the uncertainties is then needed in order to estimate the uncertainties of the reconstructed quantities. On one hand, the development of air showers is attached to statistical fluctuations which are given by the inherent nature of particle interactions. Often called *shower-to-shower* fluctuations, showers induced by identical primaries within identical atmospheric conditions, will produce different signals at ground. On the other side, signals of the SD stations are subject to uncertainties which are driven from the statistical sampling of particles. The relative uncertainty of the signals in the WCD scales with $1/\sqrt{S}$, reflecting the Poissonian behavior of a particle counter. Using measurements with twin stations (separated by a few meters) it was confirmed that the signal uncertainties can be described under the assumption of Poisson statistics. This is shown in Fig. 1.12a. An heuristic model was introduced to describe the dependency of the signal uncertainty with the zenith angle. This dependency is shown in Fig. 1.12b and arises from the increasing fraction of signal produced by muons in the WCD, as the electromagnetic component of the shower suffers from the attenuation in the atmosphere. A model for the signal uncertainty can be written as

$$\sigma_s(\theta, S) = f_s(\theta)\sqrt{S} = (a + b \sec \theta)\sqrt{S}, \quad (1.20)$$

values for the parameters can be found in [44].

1.3.4 Fit of the Lateral Distribution

The SD measures the footprint left by air showers at ground. More precisely, with the SD stations we measure distributions of particles at ground. The signal at a given distance from the sower axis³ is known as the lateral distribution function (LDF) and can be written as

$$S(r) = S(r_{\text{opt}})f_{\text{NKG}}(r), \quad (1.21)$$

³In the shower detector plane (SDP)

where the functional form of the LDF is a modified Nishimura-Kamata-Greisen (NKG) type function [46, 47]

$$f_{\text{NKG}} = \left(\frac{r}{r_{\text{opt}}} \right)^{\beta} \left(\frac{r + r_s}{r_{\text{opt}} + r_s} \right)^{\gamma + \beta}. \quad (1.22)$$

The parameter r_{opt} indicates the optimal distance at which the uncertainty in the shape of the LDF is minimized. It has been shown that this parameter depends on the array geometry and is approximately 1000 m for the SD-1500 [48]. The distance r_s and the parameter γ allow for more flexibility of the function at distances far from the core. The lateral distributions of the electromagnetic and muonic components depend on the radial distance, being the latter dominant at distances far from the core. The current estimation for r_s is 700 m.

An example of a lateral distribution of signals is given in Fig. 1.13a together with the fit to the LDF shown in Fig. 1.13b. This particular example shows a saturated station (the closest to the core) which signal has been successfully recovered [49, 50]. The recovered signal is used for the fit of the LDF as well as the non-saturated stations and the non-triggered stations, which may be observed at large distances.

Fit procedure

The fit of the LDF is performed using a maximum likelihood method. Before constructing the likelihood, each signal is converted to an effective particle numbers given by

$$n_{\text{eff}} = pS, \quad (1.23)$$

where p is the so-called ‘‘Poisson factor’’ and is defined by

$$p = \max(1, f_s^{-2}(\theta)), \quad (1.24)$$

and $f_s(\theta)$ is given by Eq. (1.20). The minimum constrain of 1 is based on the assumption of how much signal a particle produces on average. Once they are converted to effective particle numbers, signals are classified according to the following probability distributions:

1. Small signals

Signals corresponding to particle number less than 30 are described by a Poisson probability distribution

$$f_{\text{P}}(n_i, \mu_i) = \frac{\mu_i^{n_i} e^{-\mu_i}}{n_i!}, \quad (1.25)$$

where n_i is the effective particle number for a given station and μ_i its expectation given from the LDF.

2. Large signals

Signals with effective number of particles larger than 30 are described by a Gaussian distribution

$$f_{\text{G}}(n_i, \mu_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(n_i - \mu_i)^2}{2\sigma_i^2}\right), \quad (1.26)$$

where σ_i is given by the variance model of the signal (see Section 1.3.3). Recovered signals from saturated stations are also treated in this way.

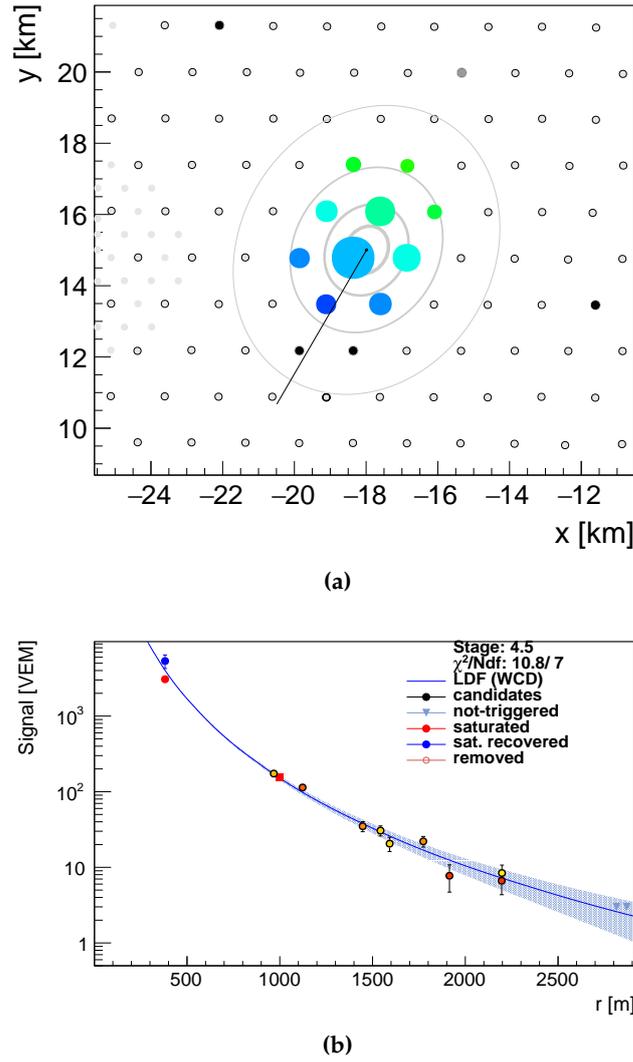


Figure 1.13: (a) Schematic of the footprint of an air shower measured by the SD. The arrival direction and the core position are indicated by the solid line. Color indicate the station trigger times, ranging from blue for early trigger times to green for late ones. Size of the markers changes according to the measured signal. The x and y -axes represent the easting and northing distances, respectively. (b) WCD signals fitted to the LDF (solid line). The signal of the saturated station was recovered and used for the fit of the LDF. Non-triggered stations are indicated by gray triangles and are used to estimate the position of the core. The red square indicates the signal at the optimal distance. The event was reconstructed with energy of 3.12×10^{19} eV and a zenith angle of 32.9° (Event time: 31. May 2019, Auger ID: 191508582800). Plots taken from the [Offline EventBrowser](#).

3. Saturated signals

Saturated signals that can not be recovered are treated as lower limit, therefore they can be described as the integral of a Gaussian probability over all values above n_i

$$F_{\text{sat}}(n_i, \mu_i) = \int_{n_i}^{\infty} f_G(n, \mu_i) dn = \frac{1}{2} \operatorname{erfc} \left(\frac{n_i - \mu_i}{\sqrt{2}\sigma_i} \right), \quad (1.27)$$

being $\operatorname{erfc}(z) = 2/\sqrt{\pi} \int_z^{\infty} e^{-t^2} dt$ the complementary error function.

4. Zero signals

Non-triggered stations are also taken into account for the fit. Zero signals are assigned the probability of obtaining an effective number of particles n_i less than the observed at threshold, n_{th} . This probability can be calculated by integrating the Poisson probability distribution from 0 to n_{th} , namely

$$F_{\text{zero}}(n_{\text{th}}, \mu_i) = \sum_{n=0}^{n_{\text{th}}} f_{\text{P}}(n, \mu_i). \quad (1.28)$$

The likelihood is written by adding each contribution

$$\mathcal{L} = \prod_i f_{\text{P}}(n_i, \mu_i) \prod_i f_{\text{G}}(n_i, \mu_i) \prod_i F_{\text{sat}}(n_i, \mu_i) \prod_i F_{\text{zero}}(n_i, \mu_i). \quad (1.29)$$

The fit of the LDF is performed maximizing the likelihood or, equivalently, minimizing the negative log-likelihood

$$-l = \sum_i \ln f_{\text{P}}(n_i, \mu_i) + \sum_i \ln f_{\text{G}}(n_i, \mu_i) + \sum_i \ln F_{\text{sat}}(n_i, \mu_i) + \sum_i \ln F_{\text{zero}}(n_i, \mu_i). \quad (1.30)$$

The LDF (given by Eq. (1.22)) by itself has three parameters: the shower size (S_{1000}) and the slopes β and γ . Now, adding the core position and the arrival direction the number of parameters increases up to eight. The current fit procedure implemented in Offline is performed in several stages in which some of the parameters are fixed to estimations from previous calculations, reducing the number of free parameters and improving the fit convergence. If there are less than four candidate stations, the parameters β and γ cannot be fitted and they are fixed to parameterizations obtained from large multiplicity events [51]. However, sometimes β and/or γ are set as free parameters of the fit if criteria based on the station multiplicity and station positioning are fulfilled.

1.3.5 Energy

A detailed description on the energy reconstruction with the SD may be found in chapter 4 of [43]. For a given energy, the value of S_{1000} depends on the zenith angle θ due to the attenuation of the shower in the atmosphere. A correction is then applied in order to get a zenith-independent estimator

$$S_{38} = S_{1000} / f_{\text{Att}}(\theta), \quad (1.31)$$

where f_{Att} is a third degree polynomial in the variable $x = \cos^2 \theta - \cos^2(38^\circ)$, namely

$$f_{\text{Att}}(x) = 1 + a x + b x^2 + c x^3. \quad (1.32)$$

The parameters a , b and c are obtained through a fit to the event intensities (number of events above a given threshold) at different zenith angles, assuming an isotropic flux in $\cos^2 \theta$. This method is referred to as the Constant Intensity Cut (CIC).

S_{38} can be interpreted as the S_{1000} that a shower would have produced, had it arrived with a zenith angle of 38° . After weather and atmospheric corrections are applied to S_{38} , the energy is calculated by

$$E_{\text{SD}} = A \left(\frac{S_{38}}{\text{VEM}} \right)^B \text{ EeV}. \quad (1.33)$$

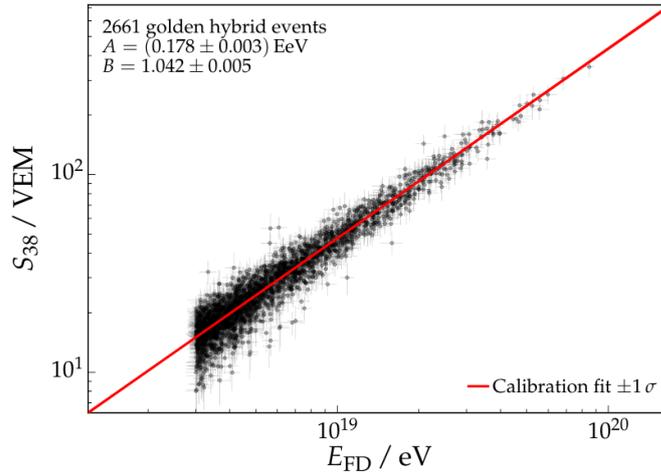


Figure 1.14: Energy calibration of the SD-1500. Taken from [43].

A calibration of S_{38} is performed using high quality events measured with both the SD and the FD, from which an energy estimate is obtained. This fit is shown in Fig. 1.14, resulting the energy estimated with the SD in

$$E_{SD} = 0.178 \left(\frac{S_{38}}{\text{VEM}} \right)^{1.042} \text{ EeV.} \quad (1.34)$$

With the FD energy resolution of about 8% (see Section 1.2.2) the resulting resolution in the energy estimated with the SD is of about 12% at the highest energies.

1.4 Selected results from the Pierre Auger Observatory

In this section we summarize some of the physics results obtained from data measured by the Pierre Auger Observatory. The selection is intended to show those results that have a relevant connection to the work presented in this thesis.

Auger has confirmed the suppression in the energy spectrum above 4×10^{19} eV with high significance [52]. The energy spectrum measured by Auger using ten years of data can be seen in Fig. 1.15. It can be seen the flattening above the ankle, $E_{\text{ankle}} = 4.8 \times 10^{18}$ eV, and a suppression of the flux at the highest energies which was established with a significance of more than 20σ [53].

Auger has also measured the energy evolution of the depth of the shower maximum. The first two central moments of the X_{max} distributions are shown in Fig. 1.16. The change in the evolution of $\langle X_{\text{max}} \rangle$ is evident at energies around $10^{18.3}$ eV. The evolution predicted by air shower simulations for a constant composition is not in agreement with this observations, therefore one can deduce a change in the composition of cosmic rays: towards lighter elements up to energies of $10^{18.3}$ eV then pointing towards heavier nuclei above that energy. This behavior is also seen for the width of the X_{max} distributions, $\sigma(X_{\text{max}})$.

The measurement of the muonic component of the shower is of a key importance for the task of determining the mass of the primary. Auger has developed techniques to perform both, direct and indirect measurements of the muon content. Some results are shown in Fig. 1.17. The muon density is estimated by AMIGA (see Section 1.2.3) by fitting the muon lateral distribution function, from which the density at 450 m to the shower

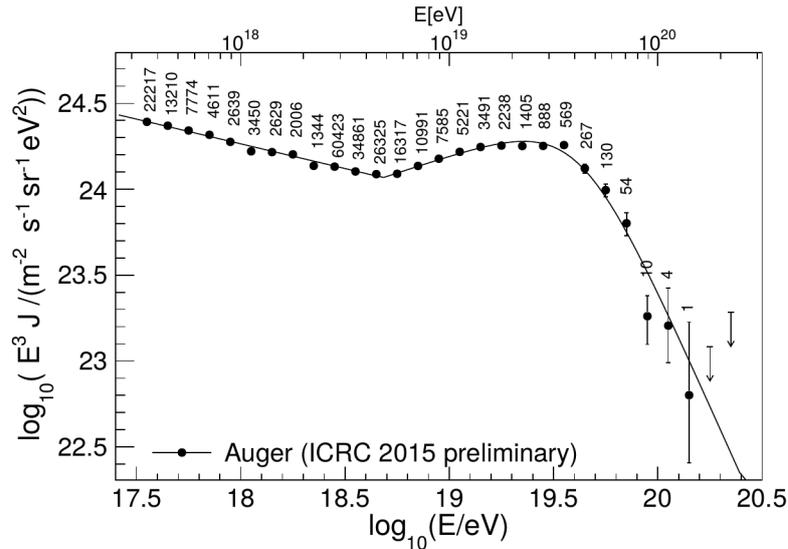


Figure 1.15: Combined energy spectrum of cosmic rays measured by Auger. Statistical uncertainties are shown by the error bars, the systematic uncertainty on the energy scale is 14%. The number of events in each bin is also given. Plot taken from [53].

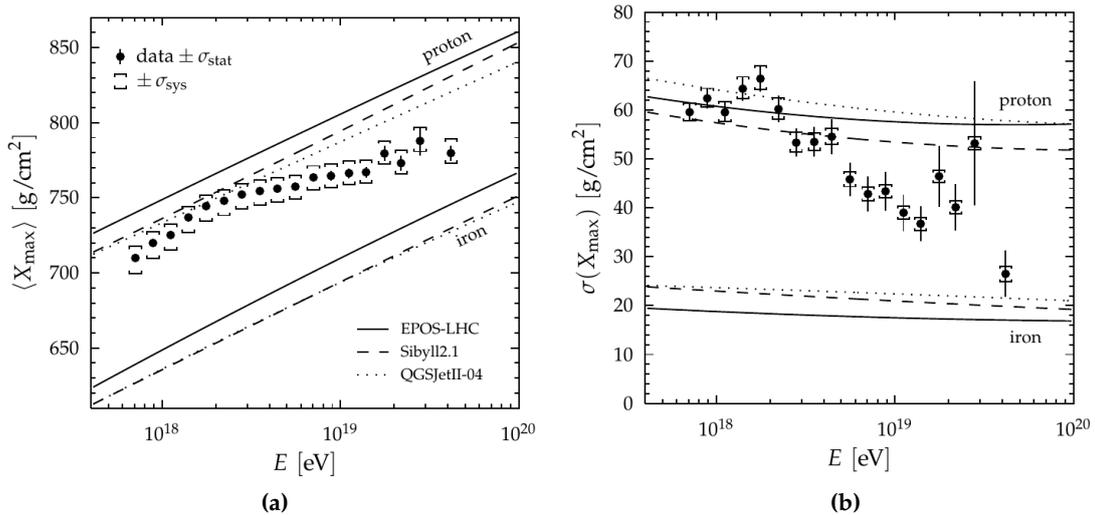


Figure 1.16: Moments of the X_{\max} distributions as a function of the energy measured by Auger. (a) The average of X_{\max} as a function of energy. (b) The second moment of the X_{\max} distributions as a function of the energy. Lines show the predictions from air shower simulations of different primaries. Plots taken from [54].

core is obtained. After corrections due to the attenuation in the atmosphere, a zenith-independent estimator of the muon density, ρ_{35} , can be derived [36, 55]. In Fig. 1.17a, the average, energy-normalized densities as estimated with AMIGA are shown as a function of energy. A comparison with the predicted muon densities using air shower simulations for proton and iron primaries are also shown. Comparing the slopes obtained with simulations and with data, one could conclude that observations made with AMIGA do not show a strong change of the primary mass in the range of $10^{17.5}$ eV to 10^{18} eV. Another important fact is the disagreement in the absolute scale between data and simulations, which suggests that simulations fail to reproduce the muonic content of the shower. This will be discussed below.

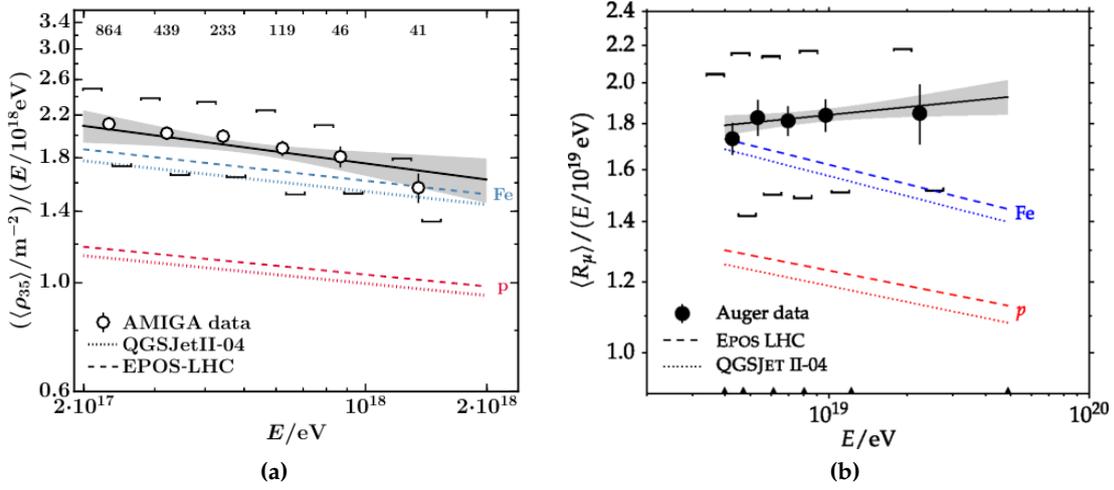


Figure 1.17: Measurements of the muonic content in air showers with Auger. (a) Average energy-normalized muon densities, $\langle \rho_{35} \rangle$, estimated with AMIGA as a function of energy. Plot taken from [55]. (b) The energy-normalized average of the muon content, $\langle R_\mu \rangle$, as obtained with the SD using very inclined showers is shown as a function of energy. Plot taken from [56]. In both measurements, the systematic uncertainty is represented by square brackets. For comparison, the predictions of proton and iron showers are also shown.

The results obtained with AMIGA at lower energies complement the previous analyses performed by Auger to determine the muonic component at energies above 4×10^{18} eV using highly inclined showers (above 62°) [56]. This result is shown in Fig. 1.17b. It is evident that data indicates a larger number of muons with respect to predictions given by simulations. The abundance of muons scales with energy as

$$N_\mu = A \left(\frac{E_0/A}{\zeta_c^\pi} \right)^\beta, \quad (1.35)$$

where ζ_c^π is the critical energy at which charged pions decay into muons, and the slope $\beta \approx 0.9$ depends on the model used to simulate the hadronic part of the shower and the details of the simulations [15].

The problem of the muon deficit in simulations was addressed by Auger with an analysis of vertical showers [57]. The results of this analysis showed that a rescaling of the hadronic part of the shower (of the order of 30% to 60%, depending on the model) is needed for the simulations in order to match the observations. This questions will be addressed again in Chapter 4.

Furthermore, Auger has also measured the proton-air and proton-proton cross sections at the highest energies [58], and has discovered a dipole in the arrival directions of cosmic rays with energies above 8×10^{18} eV with a post-trial significance above 5σ [59]. Last but not least, Auger has put limits for photon [60] and neutrino [61] searches at the highest energies. An overview on different topics covered by Auger over the last three years can be found in [62, 63].

CHAPTER 2

AugerPrime Upgrade

The results provided by the Pierre Auger Observatory are of relevant importance for the understanding of UHECR. Despite all these progresses, there is a number of enigmas in the field which still remain unclear. These can be summarized by the following questions: *what are the sources of cosmic rays?* and *how are they accelerated up to such high energies?* The knowledge of the mass composition of cosmic rays once they arrive to Earth is a strong argument in order to answer these questions.

In Chapter 1, we discussed how shower observables such as the depth at the shower maximum (X_{\max}), correlate with the mass of the primary particle. In addition, we saw that the FD of Auger measures the longitudinal development of the shower in the atmosphere from which X_{\max} can be inferred. The SD can also provide information about the primary mass. This information is encoded in the time traces produced by shower particles in the WCD: while the electromagnetic component produce a large number of relatively small signal pulses spread in time, muons produce a small number of large pulses. A limited muon identification is therefore possible using the WCD signals. The separation of electrons and muons is the basics of mass composition analyses carried out with ground-based detectors.

Due to the limited duty cycle of the FD, the number of measurements with composition sensitivity at the highest energies is not large enough. Auger addresses this point by means of the upgrade of the Observatory, AugerPrime, with the main objective of reconstructing the primary mass at the highest energies with 100% duty cycle. The main component of the upgrade is a plastic scintillator detector which will be placed on top of each of the existing WCDs. The idea of using a scintillator detector relies on the differing response of the two detectors to shower particles, therefore enhancing the capabilities of the SD for reconstructing the mass of the primary. In addition, a smaller PMT will be installed in the WCD together with upgraded electronics. These and further components of AugerPrime, as well as the physics motivation for the upgrade are the main focus of this chapter.

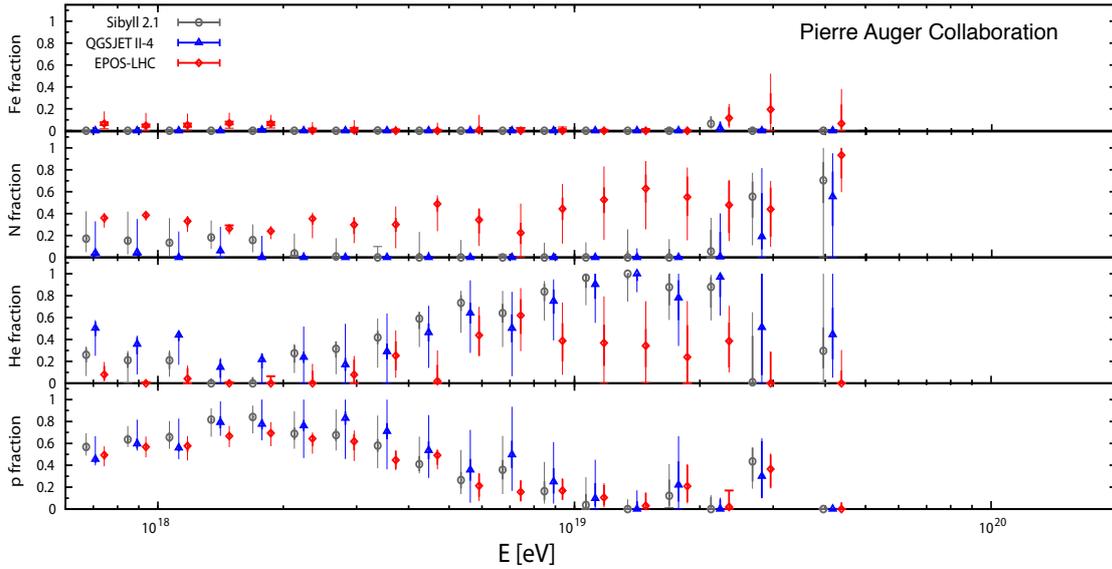


Figure 2.1: Fractions of UHECR composition as a function of the energy. The measured X_{\max} distributions were fitted to Monte Carlo templates using four mass groups (proton, helium, nitrogen and iron). For the simulations, three models for the hadronic interactions were used. Error bars show the combined statistical and systematic uncertainties except those related to the choice of the hadronic models. Plot taken from [68].

2.1 Physics motivation and goals for the upgrade

The composition of UHECR can be estimated using the distributions of the depth at the shower maximum, X_{\max} . Fig. 2.1 shows the result from fitting the distributions of X_{\max} as measured with the FD, to Monte Carlo templates obtained using four mass groups [64]. For the simulations, three hadronic models QGSJET-II.04 [65], EPOS-LHC [66] and SIBYLL-2.1 [67] were used.

The first observation is that the proton fraction varies in the whole energy range, rising above 60% near the ankle region ($\sim 10^{18.2}$ eV), then dropping to almost zero just above 10^{19} eV with a possible resurgence at higher energies. At the same time, the disappearance of the proton component is surprisingly correlated with the appearance of the helium component. There are indications of a transition to heavier cosmic rays with increasing energy, but statistics is not sufficient to be conclusive. Finally, there are indications for a possible re-appearance of a proton component at high energy.

Confirming the existence of a proton population at the highest energies could open a window to particle astronomy, as these protons would not be deflected by the magnetic fields and should point to the source. However, the data from the FD is not enough (from an statistical point of view) to confirm such a population. Therefore measurements with composition sensitivity above $10^{19.5}$ eV are needed to allow for a reliable interpretation of the observed changes of composition in terms of astrophysical models.

The particle spectrum and the distributions of X_{\max} derived from Auger data have been used by many authors to develop generic scenarios of UHECR models. In the following, some of these scenarios are summarized. Further details and references are given in [68].

In the *photo-disintegration scenario*, sources can accelerate nuclei up to a maximum energy above the energy threshold for disintegration by interaction with the CMB. Lighter elements observed at Earth could then be products of heavier nuclei that disintegrated

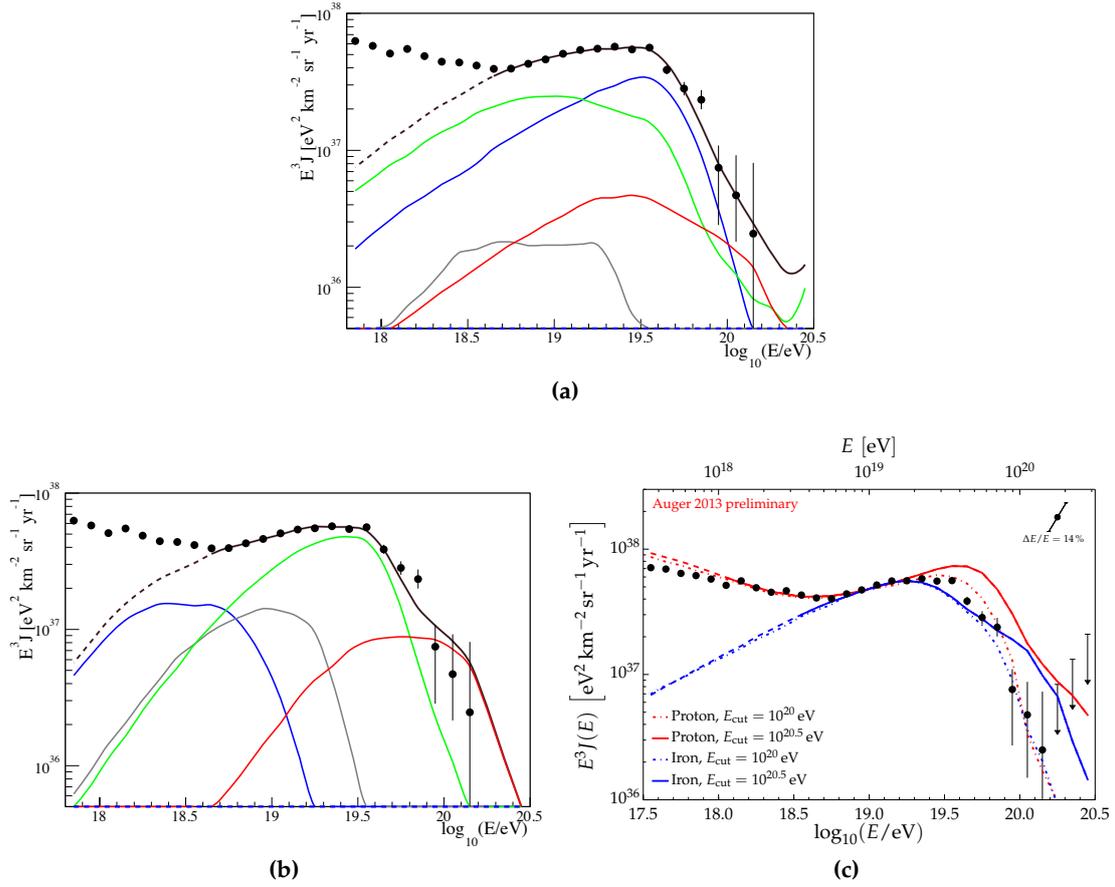


Figure 2.2: Auger spectrum as described using different astrophysical scenarios. (a) Photo-disintegration scenario and (b) maximum-rigidity scenario. In these two plots colors represent four groups of primaries: proton (blue), helium (gray), nitrogen (green), and iron (red). (c) Proton-dominance scenario. The energy spectrum is compared to predictions from sources injecting only protons (red) or iron primaries (blue). Plots taken from [68].

during propagation, and they would appear shifted in energy by the ratio of the daughter to parent mass numbers.

The *maximum-rigidity* scenario assumes that particles are accelerated to maximum energies proportional to their charge (same maximum rigidity). In this model, protons with energy of about $10^{18.5}$ eV would be related to iron nuclei shifted in energy by 26 times. In this sense, the upper part of the spectrum would be dominated by heavier elements and the suppression would be caused by the maximum acceleration power of the sources and not by energy loss processes (photo-disintegration).

As it can be seen in Fig. 2.2a and in Fig. 2.2b, both, the photo-disintegration and the maximum-rigidity scenarios do not fit the spectrum in the region of the ankle. An extra component would be needed to explain such feature in these models.

The *proton-dominance model*¹, provides an explanation of the ankle as a result of e^+e^- production. This model assumes that the spectrum is mainly dominated by extragalactic photons at energies above 10^{18} eV and the suppression is related to photo-pion production. The best fit of the “dip model” to Auger data is given in Fig. 2.2c.

¹This model is also referred as the *dip model*.

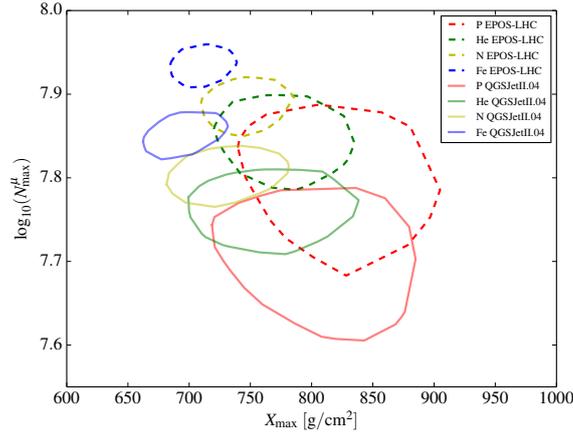


Figure 2.3: 1σ contours of N_{\max}^{μ} as a function of X_{\max} , for showers at 10^{19} eV and 38° . Four different masses were simulated using QGSJET-II.04 (solid) and EPOS-LHC (dashed) as hadronic interaction models. Taken from [68].

In this context of an open interpretation of Auger data, it was decided to undergo the observatory to a detector upgrade, named AugerPrime, which addresses the following objectives:

- (i) Elucidate the origin of the flux suppression at the highest energies is the primary goal of the upgrade of the observatory. This relates to the question of differentiating between the energy loss effects due to propagation and the maximum energy of particles injected by astrophysical sources.
- (ii) The search for a contribution of protons to the flux up to the highest energies. The aim is to reach a sensitivity to a contribution as small as 10% in the region of the flux suppression [69]. Measuring the fraction of protons is a key factor for a potential proton astronomy, and the future cosmic ray, neutrino, and gamma-ray detectors. Furthermore, the flux of secondary gamma-rays and neutrinos due to proton energy loss processes will be predicted.
- (iii) As it was shown in Section 1.4, a deficit in the number of muons predicted in simulations with respect to the observed in data was confirmed by Auger. The upgrade will also serve as probe for hadronic interaction models which may help us with the understanding of shower physics.

Importance of muons

In addition to X_{\max} , the measurement of the muon component with high accuracy is of a key importance for estimating the mass of the primary. This can be better understood with the help of Fig. 2.3. The number of muons at the maximum of the muon shower development, N_{\max}^{μ} , is shown as a function of X_{\max} for showers of four different primaries at 10^{19} eV and 38° . It can be seen that even the distributions of intermediate masses could be separated if the statistics were sufficient. As it was already mentioned, having composition sensitivity in the flux suppression region would allow us to distinguish different model scenarios for understanding the origin of the flux suppression. Furthermore, a direct access to the muonic component would be of a great help for the tuning of hadronic interaction models at energies unreachable by particle accelerators on Earth.

2.2 Components of the upgrade

Different proposals for the upgrade were studied in order to accomplish the physics objectives discussed in the previous section. A detailed description of some of them can be found in Appendix A. The final choice was to install a scintillator detector with an area of approximately 4 m^2 on top of each WCD. Furthermore, an additional smaller PMT to the WCD is meant to increase the dynamic range, allowing for measurements as close as 200 m from the shower core even for the most energetic showers. An upgrade of the SD electronics is needed to handle the signals from the upgraded stations, with a faster sampling frequency and higher resolution in the signal amplitude. These and other further enhancements are described below.

2.2.1 The Scintillator Surface Detector

The first prototypes of scintillator units, called, *Auger Scintillator for Composition - II*, or “ASCII” for short, were conceived and developed by the Bariloche group, starting data acquisition in 2010. A first prototype of 0.25 m^2 was operating for a year as a proof of concept. Later, a 2 m^2 detector was used in order to study the calibration and after completing these tests an array of $7 \times 2 \text{ m}^2$ detectors was deployed at the central hexagon of the SD-750. The outcome from the prototype phase was used to validate the detectors performance under the different weather conditions and to study signal comparisons with the WCD. The results showed that the signal of a 2 m^2 scintillator was roughly the half of the signal of the water-Cherenkov detector (in their respective units) [68]. Since the signal fluctuations are of the order of \sqrt{S} , an area of around 4 m^2 in the scintillator detector would imply similar accuracy in the measurement of both detectors.

With this in mind, the design of the detector named Scintillator Surface Detector (SSD) was developed and the first units were installed in September 2016. The analysis of these prototype detectors will be discussed in Chapter 4. In the following, a description of the detector design is presented as well as studies on the detector response and calibration.

Design

The SSD has two modules of an area $\approx 2 \text{ m}^2$, each filled with 24 plastic scintillator bars which are read out by wavelength-shifting (WLS) fibers that guide the light to a PMT.

The scintillator bars were produced by Fermi National Accelerator Laboratory and have dimensions of 160 cm long, 1 cm thick and 5 cm width. The bars are made of polystyrene (Polystyrene Dow Styron 663 W) mixed with two dopants, PPO (1%) and POPOP (0.03%) shifting the wavelength of the photons produced in the scintillator from the UV to the visible range of the electromagnetic spectrum. In Fig. 2.4, the spectrum of emission of the scintillator bars is shown.

The bars are covered with a layer of polystyrene mixed with 15% of titanium dioxide (TiO_2) of 0.25 mm thick. This layer has reflective properties and is intended to increase the number of photons that eventually will reach the fibers, and to protect the scintillating material from external agents like mechanical damages or light pollution.

The bars have two holes where the fibers are positioned following a configuration in “U”. This means that each fiber traverses two holes of the scintillator bars in such a way that the light yield is maximized and a better uniformity is obtained. The optical fibers chosen for the SSD are the Y-11 (300) MSY from Kuraray [71] with a 1.00 mm diameter. Fibers have a multi-cladding structure i.e., multiple surfaces with different refraction index are used to ensure photons get trapped inside the fiber resulting in more light yield.

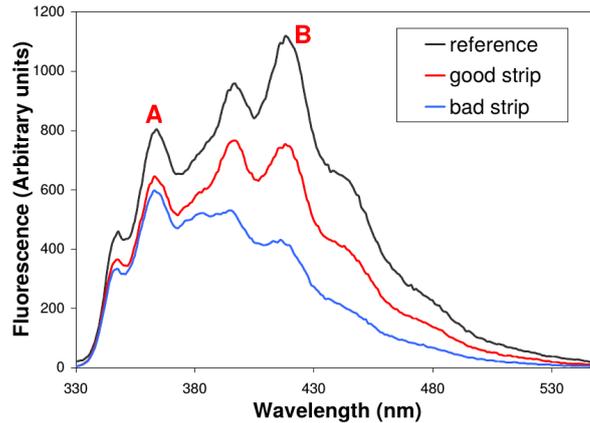


Figure 2.4: Spectrum of emission of the scintillator bar. Peaks A (365 nm) and B (420 nm) correspond to the emission peaks of the PPO and POPOP dopants, respectively (see text for details). The spectrum of scintillators with high and low response are represented by the red and blue curves respectively. Plot taken from [70].

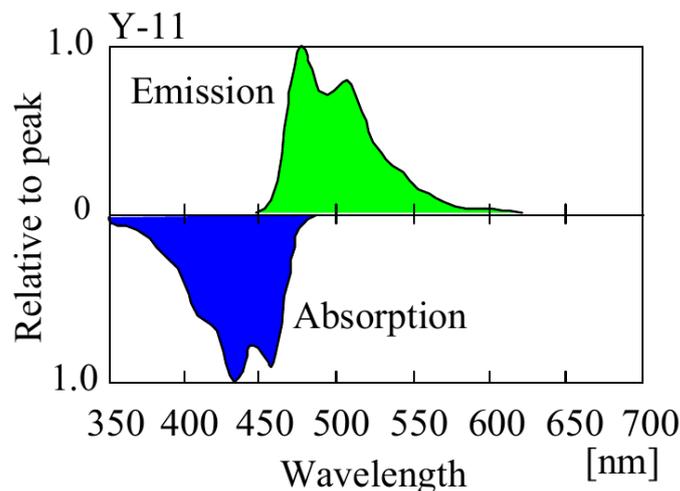


Figure 2.5: Spectrum of absorption and emission of the wavelength-shifting dye K-27 used as dopant for the fibers [72]. Nearly all photons are absorbed below 490 nm (blue light) and emitted at a second peak of around 500 nm (green light).

The WLS fibers are of S-type, allowing a shorter bending diameter of 10 mm and minimizing the risk of damages during the detector assembly [68]. The fibers are also doped with a wavelength shifting dye K-27. Below 490 nm (blue light) all photons are absorbed and then emitted at a peak of around 500 nm (green light) as shown in Fig. 2.5. The resulting effect is to increase the attenuation length of photons in the fiber to 3.5 m [71] which is of our benefit since photons in the SSD might travel several meters.

The chosen photo-sensor is a PMT that collects the scintillation light guided by the fibers. The PMT used in the SSD is the Hamamatsu R9420 [73]. It has a dynode system of 8 stages and a bialkali photocathode of 38 mm diameter. The maximum of the spectral response is at a wavelength of 420 nm, corresponding to a quantum efficiency of $\approx 25\%$. The optical fibers from the two modules of scintillator bars are bunched together using a cylinder called *cookie* as depicted in Fig. 2.6a. The cookie acts as the connection piece between the fibers and the PMT, which is inserted in a gray plastic tube with a silicone pad in front of the photocathode glass acting as optical coupling to the cookie window. A photograph of a PMT unit may be observed in Fig. 2.6b.

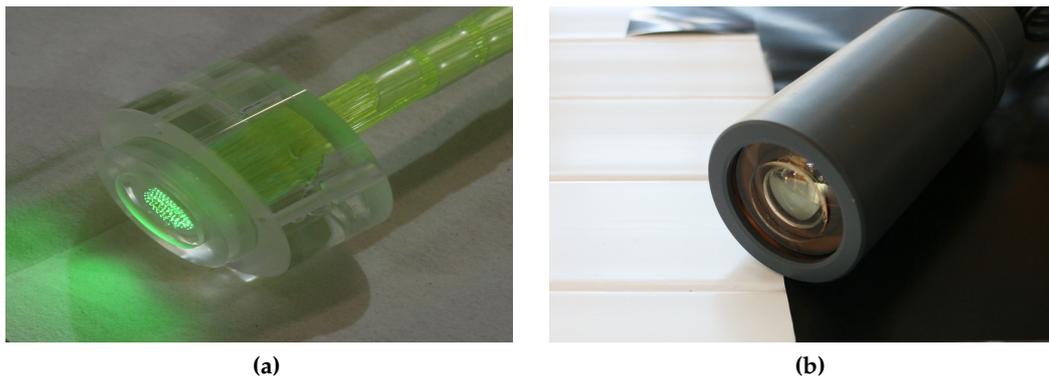


Figure 2.6: (a) Bundle of 96 fibers inside the cookie used to connect the fibers to the PMT. (b) PMT unit inside a gray plastic tube. Pictures taken from [74].

A more detailed description about the other materials, properties and methods involved in the construction of the SSD units may be found in [74]. Essentially, the two modules with the scintillator bars and fibers are placed inside an aluminum frame of rectangular cross section of $80 \text{ mm} \times 30 \text{ mm}$ and 2 mm thick. Forming a L-shape, the long sides have length of 3800 mm and the short sides are 1280 mm long, resulting a detector with total area of 4.86 m^2 . At the bottom, an aluminum composite panel is glued and riveted to the aluminum frame. This panel consist of two thin aluminum sheets of 1 mm thick and a 22 mm layer of extruded polystyrene placed between the aluminum sheets. This layer acts as a mattress for the 24 scintillator bars located in each module, reducing the air inside the detector. The two modules are separated by a gap of 384 mm. Fibers are cut in segments of 5.85 m length and each segment is routed from the central axis of the SSD through a hole in one bar, bent with a diameter of 100 mm and then routed back towards the central axis through the hole of another scintillator bar. The central axis of the SSD collects the routed fibers bunched in the cookie and then connected to the PMT which is housed aluminum tube with a total length of 231 mm and a diameter of 52 mm which separates the inner components from the readout electronics. The components of one module may be seen in Fig. 2.7a. Another layer of polystyrene-aluminum sheets is placed on top of the scintillator bars in each module to reducing any movement and damage of the inner components of the detector. The final enclosure is composed by a 1 mm aluminum sheet which is riveted and glued on top of the frame. In addition, a roof consisting of waved aluminum plates is placed on top to protect the detector from direct sun-light. A photograph of a SSD unit after construction may be observed in Fig. 2.7b.

Detection principle

The physics behind the detection principle of the SSD is basically the scintillation process which takes place when charged particles from air showers traverse the detector. Scintillator materials convert the kinetic energy deposited by charged particles into fluorescence light which eventually will be collected by a photo-detector. There is a wide amount of scintillator materials with different characteristics and used for different purposes, but we can group them typically into two types: organic and inorganic scintillators. Organic scintillators have a faster response but less light is produced, while inorganic have in general a bigger light yield but are relatively slow in their response time. For theoretical descriptions of the physics of scintillator detectors and its applications see for example [75, 76].



Figure 2.7: (a) Top-down view of the right module of a SSD. The 24 scintillator bars lay together and are interconnected with the WLS fibers. The “U” routine may be observed on the right side of the picture. Fibers are then routed back to the PMT which is placed at the bottom left corner in the picture. Photo courtesy of David Schmidt. (b) Roof and aluminum housing of the SSD (see text for details). Picture taken from [74].

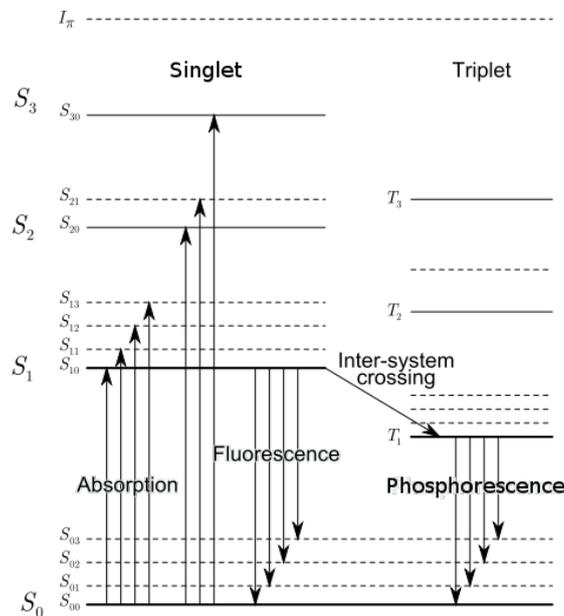


Figure 2.8: π – electron structure energy levels of an organic molecule. S_0 denotes the ground state, S_1 , S_2 , S_3 are the excited singlet states. T_1 , T_2 , T_3 are the excited triplet states and S_{ij} are the vibrational sublevels. Taken from [75].

The process of *fluorescence* is the prompt emission of visible light from a substance following its excitation by some means. We are particularly interested on how this process occurs in plastic (organic) scintillators. Fluorescence in organics arises from transitions in the energy level structure of single molecules with certain symmetry properties rising to what is known as a π – electron structure as illustrated in Fig. 2.8.

The kinetic energy of a charged particle passing nearby is absorbed by exciting the electron configuration into one of a number of excited states denoted as S for the *singlet* states (spin 0) and T for the *triplet* states (spin 1). For organic scintillators the energy gap between S_0 and S_1 is of the order of 3 – 4 eV. Each of these electronic configuration is further split into different levels with finer spacing (of around 0.15 eV) which correspond to vibrational states of the molecule. The higher singlet electronic states that are excited

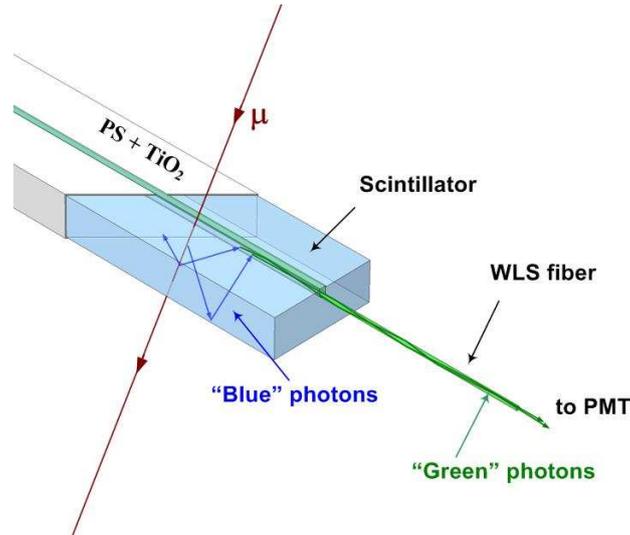


Figure 2.9: Sketch of the particle detection principle of a scintillator. Taken from [77].

are quickly de-excited to the S_1 state through internal conversion. Furthermore the states with excess vibrational energy (like S_{11} or S_{12}) are not in thermal equilibrium with their neighbors and also lose their vibrational energy. Therefore any net effect of excitation translates into a population of the S_{10} state.

The principal scintillation light (or prompt fluorescence) is emitted in transitions between this S_{10} state and one of the vibrational states of the ground electronic state. This is represented as down-going arrows in Fig. 2.8. The fluorescence intensity at a time t following excitation is given by

$$I = I_0 e^{-t/\tau}, \quad (2.1)$$

where τ represents the fluorescence decay time and in organic scintillators is of the order of a few nanoseconds [76]. Transitions to the triplet states are possible through what is called *inter-system crossing*. The lifetime of the T_1 is much longer than that of the singlet state S_1 . Therefore, de-excitations from the triple state result into delayed light emission.

As in many other cases, the scintillator material of the bars are mixed with dopants which act as wavelength shifters. They absorb the light produced in the primary scintillation process and then re-emit it at a longer wavelength. This is intended to match the emission spectrum of the fibers with the absorption spectrum of the photo-sensor. As depicted in the sketch of Fig. 2.9, when a charged particle from an air shower enters the scintillator, it will produce photons in the blue-UV part of the electromagnetic spectrum through the aforementioned processes. These photons are emitted by the scintillator in all directions forming a *cloud* of photons which, after some refractions in the sides of the bar eventually reach the fiber. The fiber increases the wavelength of these photons, transforming them into “green” green light which is driven to the PMT located at the fiber end.

The response of the SSD can be inferred from the energy that particles deposit in the scintillator bar. The energy loss per unit of traversed length (also called *stopping power*) in the scintillator is shown for electrons and muons in Fig. 2.10. In both cases, there is a *dip* in the energy loss which corresponds to the minimum energy needed to ionize the material. Another feature is that for a wide range of energies, the stopping power is nearly constant to a value of ≈ 2 MeV/cm, which is close to the value of the

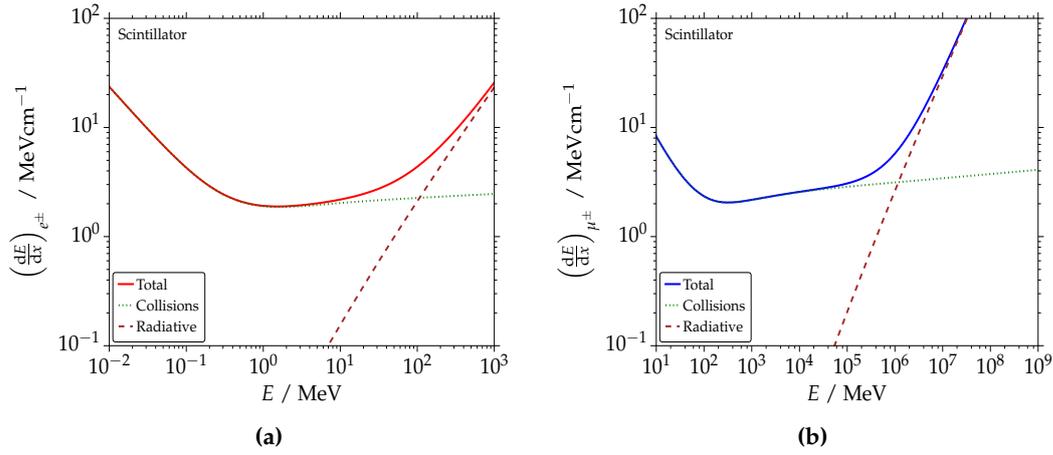


Figure 2.10: Stopping power of electrons (a) and muons (b) as a function of the particle energy. Losses by collisions (green - dotted) and by radiative processes (brown - dashed) are also shown. Values taken from tables in [78].

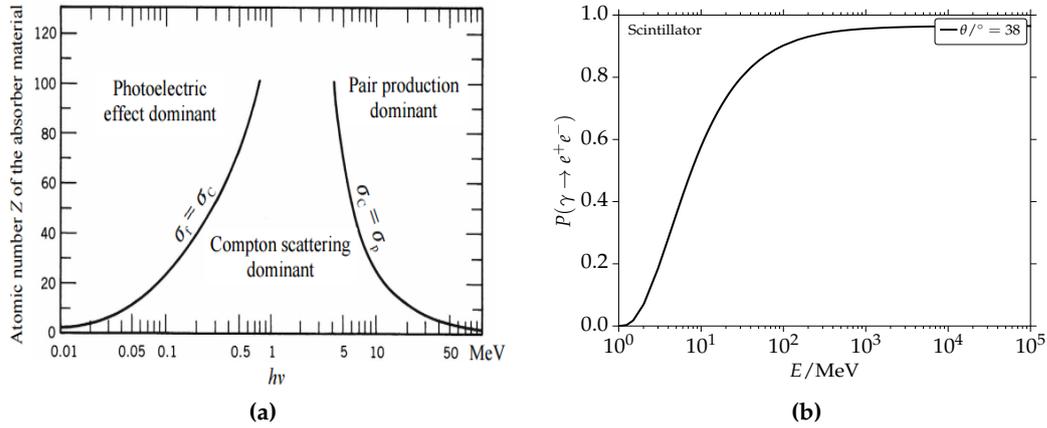


Figure 2.11: (a) Photon cross section as a function of energy and atomic number of the absorber material [79]. (b) Probability of photon conversion into e^+e^- pair as a function of energy in a plastic scintillator (polystyrene).

minimum. These particles are usually called minimum ionizing particles (MIPs). The signal produced by one MIP is taken as reference unit for the SSD signals.

Photons are the most numerous particles produced in air showers but they have zero charge. They will produce a signal via secondary processes implying charged particles (such as electrons) which eventually ionize the scintillator material. These are: photoelectric effect, Compton scattering and pair production. In Fig. 2.11a the cross section of each of these processes is shown as a function of the photon energy and on the electron density (or atomic number Z) of the absorber material.

For low energies and high atomic number the dominant process is the photoelectric effect. Compton scattering becomes dominant for energies of the order of a few MeV and intermediate Z number. Pair production has an energy threshold for the photon of $E_\gamma > 2 m_e$. The probability of pair production as a function of energy is given by

$$P_{e^\pm}(E_\gamma) = 1 - e^{-\frac{x}{\lambda(E_\gamma)}}, \quad (2.2)$$

where $x = h / \cos \theta_\gamma$ represents the tracklength of a photon traversing a material of thickness h with zenith angle θ_γ and λ is the mass attenuation length (or mean free path), which depends on energy. The probability for photons at $\theta = 38^\circ$ in the scintillator material as a function of energy is shown in Fig. 2.11b. The typical energies of photons produced in air showers are of the order of 10 MeV. Due to the aluminum housing, the effective atomic number in the SSD is between 10 and 20, then according to Fig. 2.11a, most of the photons will produce a signal via Compton scattering and a smaller fraction will convert into an electron-positron pair.

Calibration of the Scintillator Surface Detector

Analogous to the calibration of the signals which are produced in the WCD [26], one could think of a similar procedure for the calibration of the SSD signals. We have performed a study of the calibration procedure of the SSD using the signal distributions of secondary particles produced in simulated low energy showers. The showers were simulated using Cosmic Ray Simulations for Cascade (CORSIKA) and a description is given in Appendix B.1.

The energy spectrum of secondary particles arrived at ground obtained from the low energy showers is shown in Fig. 2.12a. Electrons and positrons have energies in the range of 10 MeV to 35 MeV. Photons have a flatter distribution around the maximum, with energies of around 6 MeV. As the shower evolves, the electromagnetic component is constantly growing, electrons and positrons produce photons through bremsstrahlung and high energetic photons produce electron-positron pairs. The peak observed in the spectrum of photons at low energies corresponds to the 511 keV from e^+e^- annihilation. Muons are more energetic having their energy spectrum a peak at around 3 GeV. The angular distribution of muons at ground can be approximated by

$$I(E, \theta) = I_0 \cos^\alpha \theta, \quad (2.3)$$

where the coefficient α was determined by experimental measurements being $\alpha = 1.96 \pm 0.02$ [80], this is why muons are known to have a $\cos^2 \theta$ distribution². Taking into account the projection of the muon flux over a flat detector at ground the distribution can be written as

$$I(E, \theta) \propto \cos^\alpha \theta \cos \theta \sin \theta, \quad (2.4)$$

where the additional $\cos \theta$ factor comes from the projection at ground and the $\sin \theta$ factor comes from the projection of the solid angle. The expected distribution for muons is expected to be

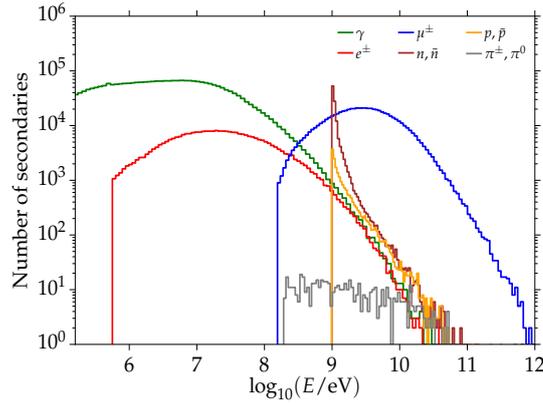
$$I(E, \theta) \propto \cos^3 \theta \sin \theta. \quad (2.5)$$

Fig. 2.12b shows the angular distribution for 663847 secondary muons produced in the simulation of 1 hour of background radiation and the fit to the function in Eq. (2.5) giving the following result

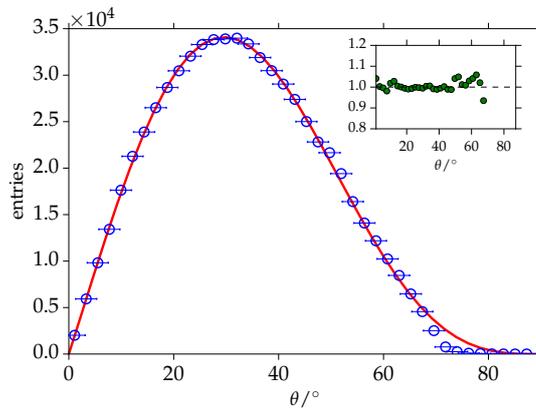
$$I(E, \theta) = 1.5 \times 10^5 \cos^{3.09} \theta \sin^{1.02} \theta. \quad (2.6)$$

The hadronic component is dominated by nucleons, produced by fragmentation during the collisions of the primary particle and nuclei of air. While for energies of the order of GeV the amount of neutrons is of about one order of magnitude the amount of protons, both species equal in number at energies of ~ 5 GeV. Pions represent the last component in terms of secondaries, with a flat energy distribution of the order of GeV.

²More precisely, α depends on the energy of the muons, but here is treated as a constant for simplicity.



(a)



(b)

Figure 2.12: Results from the simulation of 1 hour flux of low energy showers with CORSIKA over Malargüe (1440 m a.s.l.) for a flat detector. (a) Energy spectrum of the secondaries that reach the ground (see text for details). (b) Angular distribution of secondary muons produced in the simulation (blue circles) and fit to a function of the form $I \propto \cos^3 \theta \sin \theta$ (red line). The inset shows the relative difference between the data points and the fit result as a function of zenith angle θ . The difference is below 10% for angles up to 75° where the difference start to grow. This is due to the fact that primaries were simulated in a zenith range of $0^\circ \leq \theta \leq 75^\circ$, then almost no secondaries are produced with $\theta > 75^\circ$.

Each of the secondary particles was injected into a simulated upgraded station comprised by a WCD and a SSD using the `Offline` framework [81]. The distribution of the charge in the SSD obtained from each of the components is shown in Fig. 2.13. The charge is obtained from the time integral of the FADC trace which digitizes the total amount of electrons collected at the PMT anode, i.e., the integrated charge given by the time integral of the pulse

$$q = \int_0^\infty \frac{V(t)}{R} dt. \quad (2.7)$$

One feature in Fig. 2.13 is that electrons (and positrons) contribute to signal in the same way as muons do. This is related to what it was discussed earlier, electrons and muons deposit on average the same amount of energy which is close to the minimum value to ionize the material, i.e., they are MIPs. Although there are less muons than electrons (and positrons) at ground, the former produce a signal in the vast majority of the cases, while the latter are often scattered by aluminum housing and not always enter the detector.

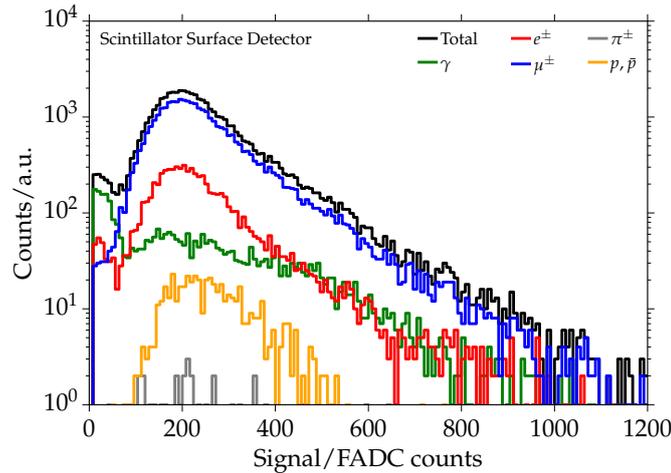


Figure 2.13: Distribution of charge produced in the SSD by different particles.

A peak in the distribution at around around 200 FADC counts is produced as an average of muons and electrons passing through the scintillator. The tail of the distribution is produced by particles impinging at high zenith angles resulting in longer tracklengths inside the bars (and therefore in higher signals). Photons have a significant contribution to high signals. This is due to the fact that photons are the most numerous particles produced in air showers and their energies allow them to produce signals by the aforementioned processes.

The signal in the scintillator is proportional to the tracklength, L , traveled by particles in the detector. Since the active material of the SSD is thin (in comparison to the water volume of the WCD), the average track-length of particles can be approximated by

$$L \approx h / \cos \theta, \quad (2.8)$$

where h is the thickness and θ the particle zenith angle. Then the signal follows the following relation

$$S \propto L \propto \sec \theta. \quad (2.9)$$

Since we know the particle direction, we can correct by a factor $\cos \theta$ to get the signal by a vertical equivalent through-going MIP or VMIP. The result of this analysis is shown in Fig. 2.14. The histograms of the signal produced by omnidirectional particles and the signal obtained after applying the cosine correction are fitted to a parabolic function $f(x) = ax^2 + bx + c$ obtaining the values of 35.4 ± 0.4 PE/MIP for the MIP and 31.3 ± 0.3 PE/VMIP for the VMIP (the units correspond to the number of PE produced at the photocathode of the PMT). This is in good agreement with the results obtained from real measurements of SSD signals triggered by the muon tracking detector from the KASCADE experiment at the Karlsruhe Institute of Technology, where the reported average values are 34.8 PE/MIP and 30.3 PE/VMIP [82]. For simplicity, the term MIP will be used to denote the signal produced in the SSD. However, the correction factor of $C = 0.88 \pm 0.01$ VMIP/MIP is used in the process of the signal calibration.

2.2.2 Small Photo-Multiplier Tube

In addition to the three PMTs, an extra PMT with significantly smaller active surface will be implemented in each WCD to extend the dynamic range. The so-called Small

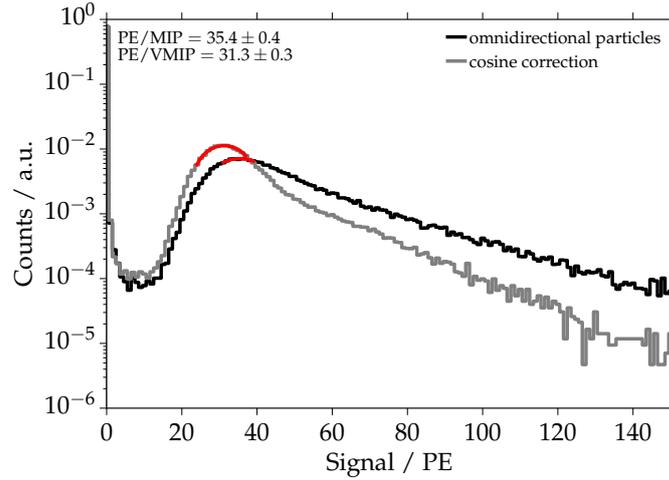


Figure 2.14: Calibration of the SSD. The black line shows the histogram of the signal (in photoelectron (PE)) deposited by omnidirectional particles. The gray line shows the histogram of the signal produced for vertical-equivalent particles. The signal is obtained after applying the correction accounting for their angle of incidence. Red curves indicate the fit to a quadratic function.

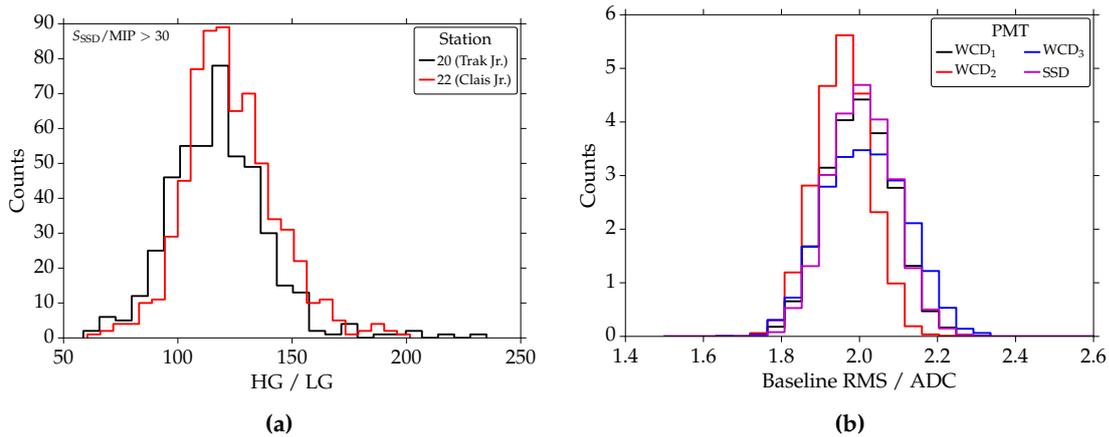


Figure 2.15: (a) Distribution of the high gain to low gain ratio in the SSD PMT from two upgraded detectors. The mean of the distributions is around 120, which is close to the design value. (b) Distribution of the baseline root mean square (RMS) in the high gain channel of the WCD and the SSD PMTs in upgraded stations. Mean value is of around 1.9 ADC counts.

PMT (SPMT), model Hamamatsu-R8619-22 of 30 mm of diameter will be placed at the center of the tank. Due to the smaller light collection combined with a suitable gain setting will extend the dynamic range of the WCD matching the saturation limits of the scintillator detector, allowing us to measure showers of 10^{20} eV at distances up to 200 m from the shower axis with both sub-detectors [83]. Further details on the performance of the SPMT will be given in Chapter 4.

2.2.3 Surface Detector Electronics Upgrade

The electronics of the SD needs to be upgraded in order to readout the signals of the SSD and the SPMT. The current electronics board will be replaced with an Upgraded Unified Board (UUB) which additionally to the signal processing, it provides a higher sampling

frequency (120 MHz instead of 40 MHz) and enhanced resolution in amplitude (12 bits instead of 10 bits) [84].

The gain of the SPMT is set in such a way that the dynamic range is extended to 20000 VEM, matching the dynamic range of the SSD of 20000 MIP [83]. Regarding the amplification in the UUB, two amplifiers with different gains are used to extract the high and low gain signals from the anode of the large and the SSD PMTs. The signals of the three large PMTs are split with a gain ratio of 32 into two ADC channels (as in the UB) while for the SSD PMT a gain ratio of 128 is used between the two channels. The distribution of high gain to low gain ratio of the SSD PMT in stations with upgraded electronics can be seen in Fig. 2.15a.

The noise in the UUB is estimated from the RMS of the distribution of baseline, which is calculated taking the first 500 bins of the time traces. The RMS noise on the high gain channel of the WCD and the SSD PMTs is shown in Fig. 2.15b. An average value of 1.9 ADC was obtained. On the other hand, the time resolution was estimated using two neighboring upgraded stations to be about 9.5 ns on a single station [84]. More details about the performance of upgraded stations will be given in Chapter 4.

2.2.4 Further upgrades

In addition to the Scintillator Surface Detector, the Small PMT and the corresponding upgrade of the electronics mentioned before, there are other components of the upgrade which are listed below:

- **Extension of the uptime of the Fluorescence Detector**

The gains of the PMTs of the FD will be reduced by around a factor 10. This allows the FD to measure earlier in the evenings and later in the mornings, thus increasing its uptime by around 50%. The total uptime of the FD will result in around 20%.

- **Underground Muon Detector (UMD)**

AMIGA detectors will be deployed in the SD-750, corresponding to 61 scintillator-based UMDs buried at 2.3 m depth. This upgrade of the SD in its denser area will significantly increase the composition sensitivity in the region of the ankle in the energy spectrum. Furthermore, the direct measurement of the muonic component by the UMD will serve as calibration of the methods for reconstructing the muon content with the SSD and the WCD.

- **Radio upgrade**

Each station of the SD will be equipped with a radio antenna in order to enhance the composition sensitivity with horizontal air showers, given their large footprint of radio emission at ground. Details on the radio upgrade and its expected performance are given in [85].

CHAPTER 3

Event reconstruction and mass composition

With the different components of a detector upgrade in hand, the next natural step is to include them in our reconstruction methods in order to achieve the goal of reconstructing the mass of cosmic rays at the highest energies. In the context of mass composition, we typically talk about two observables, which are the longitudinal development of a shower and the simultaneous measurement of the electromagnetic and muonic components at ground (for a good review on this topic see [18]).

In Chapter 1, it was explained that Auger measures the longitudinal profile of showers with the FD and it was found a direct relation between $\langle X_{\max} \rangle$ and the primary mass (see Eq. (1.12)). In addition, mass composition observables can be estimated from sampling particle distributions at ground. In general, reconstruction algorithms can be employed in order to separate the muonic and the electromagnetic signals. Despite being more numerous, electromagnetic particles (e^\pm , γ) stop after some centimeters in water. On the other hand muons, more energetic, traverse the whole volume of water. This leads to different signatures in the time traces corresponding to different particle types. The former produce a large number of small Cherenkov pulses whereas muons produce spiky structures in the traces. Different features can be found by looking at the lateral distribution functions of the different components, being steeper for electromagnetic particles, while muons have a flatter LDF as they are less attenuated in the atmosphere.

Current methods for determining the mass range from studies of the arrival times of particles at the SD stations [86] to studies based on the properties of shower universality [13, 42]. The addition of AugerPrime detectors will significantly increase the information about the shower, as different detectors are dedicated to give specific measurements independently.

The aim of this Chapter is to incorporate the SSD into the chain of the standard SD reconstruction. To accomplish this goal a parameterization of the SSD LDF will be obtained, which requires an understanding of the SSD signals and their fluctuations. With a parameterization of the LDF in hand, the potential for mass discrimination will be studied at the level of single events by means of using paired measurements of the WCD and the SSD reconstructions. All the methods developed in this Chapter were derived from Monte Carlo simulations and are expected to be applied to real data. This will be covered in Chapter 4.

3.1 Library of simulations

All the analyses presented in this Chapter were derived from Monte Carlo simulations of air showers. Two libraries of CORSIKA showers were used: a *fixed library* of showers with discrete values for the energy and for the zenith angle and a *continuous library* with energy and zenith continuously distributed. The characteristics of each set of simulated showers can be seen in Table 3.1 and Table 3.2, respectively.

Table 3.1: 120 events were used for each configuration consisting a total number of 8400 simulated showers.

Primaries	Proton, Iron
$\lg(E/\text{eV})$	18.0, 18.5, 19.0, 19.5, 20.0
$\theta/^\circ$	0, 12, 22, 32, 38, 48, 56
$\phi/^\circ$	0 – 360, uniformly distributed
Had. Int. Models	QGSJET-II.04

Table 3.2: Set of simulated showers with energy and zenith angle continuously distributed. The set consists of a total number of 4000 showers.

Primaries	Proton, Iron
$\lg(E/\text{eV})$	18.0 – 20.0, increment of $\Delta \lg(E/\text{eV}) = 0.1$
$\theta/^\circ$	0 – 70, increment of $\Delta\theta = 0.1^\circ$
$\phi/^\circ$	0 – 360, uniformly distributed
Had. Int. Models	QGSJET-II.04

3.2 Towards a reconstruction with the SSD

3.2.1 Describing the SSD signals

The first step on the way to achieve an event reconstruction including the scintillator detector is the description of the signals produced when particles from EAS hit the detector. This description should be encoded in a probability density function which will be used first, for the parameterization of the signal uncertainty model (see Section 3.2.2) and later for the fit of the LDF (see Section 3.2.3). Same as for the case of the WCD signals, this probability density function (p.d.f.) should describe large and small signals but with the caveat that the SSD operates in slave mode, i.e., it is triggered by the WCD and therefore its signals can be as low as zero (accounting for fluctuations of the baseline).

Let us assume a measured signal S with uncertainty σ ; the probability of observing S given its prediction μ can be formulated in terms of a positive-side normal distribution¹

$$p(S|\mu, \sigma) = \mathcal{N}_{\text{TR}}(S|\mu, \sigma) = \frac{\mathcal{N}(S|\mu, \sigma)}{\mathcal{N}^+(\mu, \sigma)}, \quad (3.1)$$

where

$$\mathcal{N}(S|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(S-\mu)^2}{2\sigma^2}\right), \quad (3.2)$$

¹Also referred to as “truncated” normal distribution.

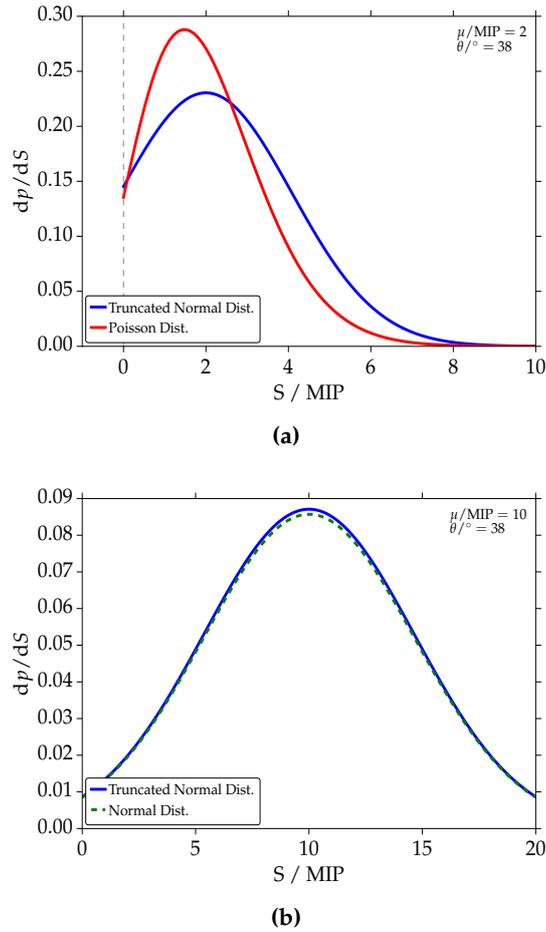


Figure 3.1: Distributions of SSD signals. (a) A truncated normal distribution (blue line) is compared to a Poisson distribution (red line). For this plot, samples of distributions with mean signal $\mu = 2$ MIP were produced. (b) The truncated normal distribution is compared to a normal distribution (green dashed line) with mean $\mu = 10$ MIP. See text for details.

is the normal distribution function and

$$\mathcal{N}^+(\mu, \sigma) = \int_0^\infty \mathcal{N}(S|\mu, \sigma) dS = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{\mu}{\sigma\sqrt{2}} \right) \right], \quad (3.3)$$

is the normalization factor. The definition given in Eq. (3.1) ensures that the probability of observing any measured signal S is always one irrespective of its prediction μ , i.e.,

$$\int_0^\infty p(S|\mu, \sigma) dS = 1 \quad \text{for any } \mu > 0. \quad (3.4)$$

A depiction of the signal distributions is given in Fig. 3.1. At small signals, the distribution of signals given by Eq. (3.1) is compared to the one obtained from a Poisson p.d.f.. This is shown in Fig. 3.1a. At large signals, the signal distribution given by Eq. (3.1) matches the one obtained from a Gaussian p.d.f.. This is because $\operatorname{erf} x \rightarrow 1$ when $x \rightarrow \infty$. An example is shown in Fig. 3.1b.

3.2.2 Uncertainty model for the SSD signals

The signal uncertainty model for the SSD was firstly derived in [87] by means of simulating doublet stations, which measure different particle samples of the shower. These

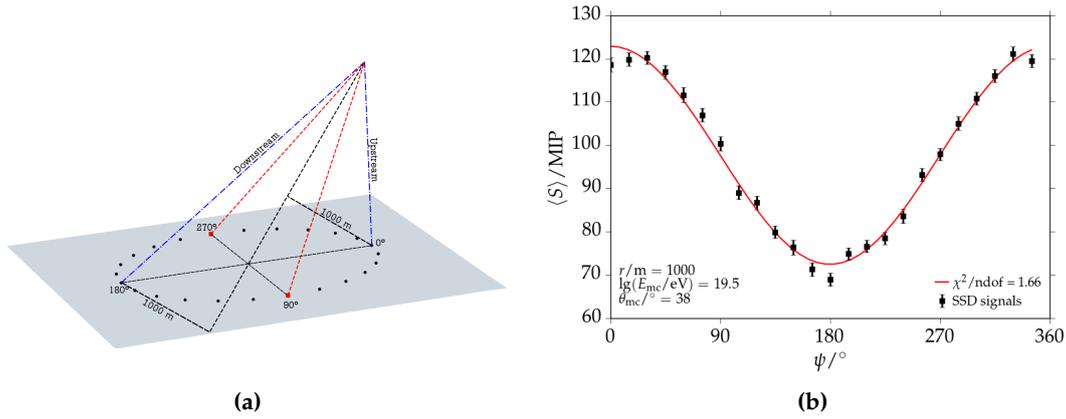


Figure 3.2: (a) Sketch of the ring of stations used in the simulations. The 24 stations are placed at 1000 m from the shower core in the SDP. Taken from [87]. (b) Average signal of the SSDs as a function of the azimuth angle ψ (with respect to the shower direction) for showers with an energy of $10^{19.5}$ eV and zenith angle of 38° . The red line shows the fit to Eq. (3.5) (see text for details).

were in fact called *pseudo doublets* as, contrary to the original measurements for the WCD, the simulated SSDs were not separated by 10 m (due to thinning algorithms applied in simulations) but the pair of stations were located at 1000 m from the shower core and at azimuthal angles of 90° and 270° in the SDP. This was done in order to ensure that particles arriving to each detector would traverse the same amount of atmosphere, reducing the biases due to attenuation effects. Such *pseudo doublets* would exhibit signal fluctuations since they measure particle samples that come from different shower developments higher in the atmosphere, which are mostly driven by initial interactions.

The signal uncertainty model for the SSD was revisited in this work using a slightly different approach than the one used in [87]. The main reason is due to the changes in the simulation framework, which include modifications in the material properties of the SSD and different energy threshold for particles interacting with the scintillator. Furthermore, here we make use of the probability density function described in Section 3.2.1 while in [87] a different likelihood for the signals was used.

In order to investigate the signal fluctuations in the SSD, each of the events on Table 3.1 was thrown over a ring of 24 stations placed at 1000 m to the shower core. Although all stations are located at the same distance, shower particles will travel different depths in atmosphere until they arrive to the different stations leading to asymmetries in the detected signals. The asymmetry may be visualized in Fig. 3.2. The schematic in Fig. 3.2a shows the position of the 24 stations at the ground for a non-vertical shower. It can be seen that stations below the *upstream* or “early” region of the shower will record higher signals than stations hit by the *downstream* or “late” region due to the different amount of atmosphere traversed by particles on each case.

In order to compare the signals from all stations, a correction for the asymmetry is needed. At first order, one could explain the signal variation as a function of the azimuth angle ψ between the station position and the shower axis as

$$S(\psi) = S_0(1 + S_1 \cos \psi). \quad (3.5)$$

Fig. 3.2b shows the signal variation as a function of the azimuth angle and the fit to Eq. (3.5). The fit is performed for each energy, zenith angle and for different distances. The evolution of the fit parameter S_0 as a function of the zenith angle is shown in Fig. 3.3.

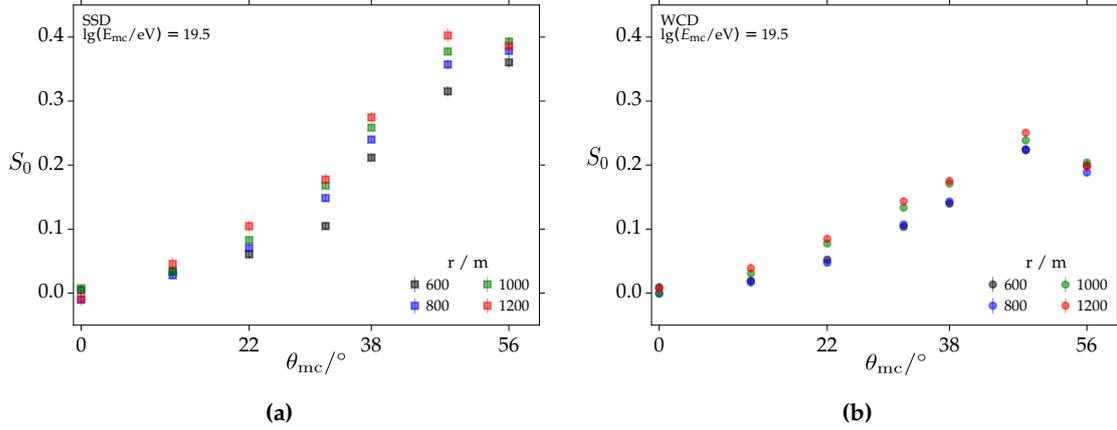


Figure 3.3: Result of the fit parameter S_0 from Eq. (3.5) as a function of the zenith angle and for different distances as predicted by the fit to SSD signals (a) and to WCD signals (b).

For each ring of stations (zenith angle and energy), each signal is corrected by subtracting the result from the fit to the original signal s_i . An estimator for the signal uncertainty was obtained as the variance of the corrected signals $S_i = s_i - S_i(\psi)$

$$\sigma_S = \sqrt{\frac{1}{N-2} \sum_{i=1}^N (S_i - \langle S \rangle)^2}, \quad (3.6)$$

where $\langle S \rangle$ is the average of the signal of all 24 stations and N is the number of stations in the ring. The 2 in the denominator arises from the two parameters used in the fit. The signal variance has often been expressed in terms of the expected signal and the zenith angle

$$\sigma_S = f(\theta) \sqrt{S}, \quad (3.7)$$

where the \sqrt{S} is related to the Poisson fluctuations of particles being sampled by the detectors and the so-called *spectral factor*, $f(\theta)$, is written as a linear relation on the secant of the zenith as²

$$f(\theta) = a (1 + b(\sec \theta - \sec 35^\circ)). \quad (3.8)$$

Estimates of the signal variance (Eq. (3.6)) and of the mean signal from the dense rings were used to obtain the parameters a and b in Eq. (3.8) by performing a χ^2 minimization. For the data uncertainties, the standard deviation from the distribution of $\sigma_S / \sqrt{\langle S \rangle}$ divided by \sqrt{N} on each zenith bin was used (where N is the number of entries inside the bin). The fit converges with a χ^2/n_{dof} of 251.86 and parameters

$$f(\theta) = (1.449 \pm 0.001) \cdot (1 + (0.175 \pm 0.002) \cdot (\sec \theta - \sec 35^\circ)). \quad (3.9)$$

The minimization was performed using the NLOpt package [89]. The reason for the high value of the reduced- χ^2 can be attributed to a wrong estimation of the errors or to a bad description of the data by the model. Since the error bars are underestimated, the quality of the fit will be sensitive to any deviation of the data points from linearity. These deviations can be related to the fact that a library of showers with discrete values for energy and zenith was used. Each zenith bin contains data from different energies and unexpected changes in signals, for example due to transitions between high and low

²It has been shown that the spectral factor also depends on the ratio between the electromagnetic and muonic components of the shower as well as on the distance to the shower core [88].

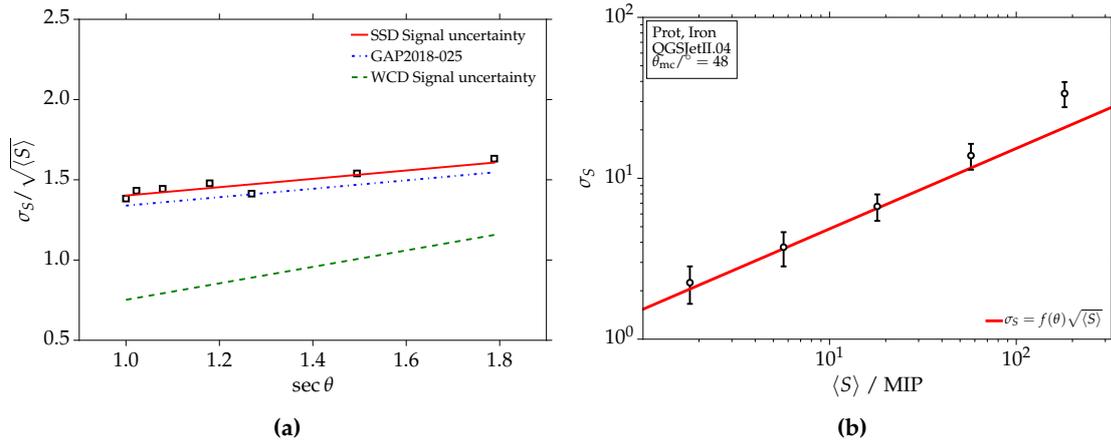


Figure 3.4: Signal uncertainty model for the SSD derived from simulated station rings. (a) Result of the fit as function of the secant of the zenith angle. All energies $< 10^{20}$ eV were used (see text for details). (b) Signal variance as a function of the average signal at a zenith angle of $\theta = 48^\circ$.

gain, might induce systematics which are not taken into account. The resulting signal uncertainty model for the SSD can be seen in Fig. 3.4. The model derived in this work predicts values for the uncertainty on a given signal slightly larger than that predicted by [87] (as it can be seen in Fig. 3.4a). This could be related to the fact that the thickness of the SSD and the energy threshold for simulated particles were different in the two studies. On the other hand, the deviation of the data at high signal sizes from the prediction of the fit (as can be seen in Fig. 3.4b) suggests that the assumption that the variance depends with the square root of the signal might be valid only for a certain range of signals. In this particular example, signals above 100 MIPs correspond to the average signal at 1000 m for showers with energies of 10^{20} eV. For this reason, these showers were not included in the derivation of the signal uncertainty model.

Low signals case

From Fig. 3.4b, it can be seen that the assumption of the \sqrt{S} for the signal dependency of σ_s seems to be valid for a certain range of signal sizes. The deviation at higher signals must be studied in more detail. Previous studies on the signal uncertainty model derived from simulations have shown that this feature also appears for the WCD for showers with energies $\geq 10^{19.5}$ eV [87].

At low signals, the situation is more complicated. In fact, the assumption of the square root of the signal only holds for large values i.e., when the number of sampled particles is large. Furthermore, since the SSD can be triggered by the WCD when no particle passed through the scintillator, the signal in the SSD may just be an integral of the baseline and the fluctuations around baseline should be taken into account for the total signal uncertainty. Assuming that the baseline for the SSD PMT has an RMS of 0.5 ADC and a conversion factor of 200 ADC/MIP, an estimate of the signal variance due to baseline fluctuations is found to be around $\sigma_b = 0.04$ MIP, as it can be seen in Fig. 3.5. The total uncertainty results in the sum in quadrature of the two terms

$$\sigma^2 = \sigma_s^2 + \sigma_b^2. \quad (3.10)$$

Now assuming an expected signal of $S_{\text{exp}} = 0.1$ MIP and a zenith angle of 38° , the uncertainty according to Eq. (3.10) is $\sigma = 0.45$ MIP, which turns out to be too small

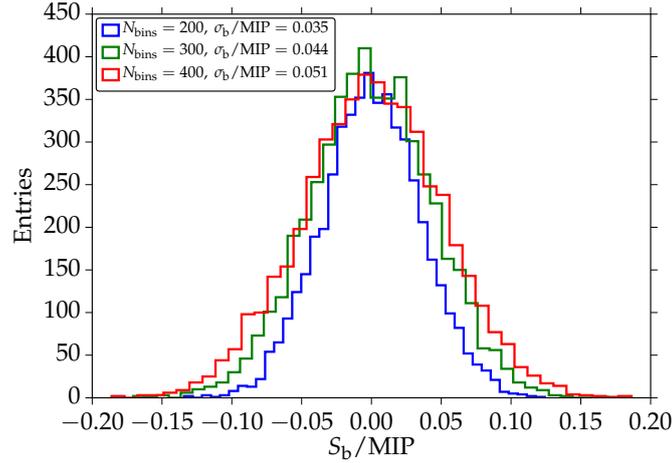


Figure 3.5: Signal distributions for Monte Carlo traces assuming an RMS of 0.5 ADC around baseline. Colors indicate different integration windows.

as one could have measured one, two or three particles which simply did not hit the detector. In other words, the current model for the uncertainty does not handle small signals properly. One solution could be to set a minimum value for the uncertainty at low signals and this value should be big enough in such a way that these signals do not contribute much to the likelihood fit of the LDF. This will be discussed in the next section.

3.2.3 Obtaining the Lateral Distribution Function for the SSD

With a p.d.f. for describing the signals and a model for their uncertainties in hand, the next step is to obtain the lateral distribution function of signals recorded by the SSD at the ground. The signal at any distance to the shower core can be expressed by

$$S(r) = S(r_{\text{opt}})f_{\text{NKG}}(r), \quad (3.11)$$

where f_{NKG} is a modified NKG-like function

$$f_{\text{NKG}}(r) = \left(\frac{r}{r_{\text{opt}}}\right)^{\beta} \left(\frac{r+r_s}{r_s+r_{\text{opt}}}\right)^{\beta+\gamma}. \quad (3.12)$$

Since the location of the SSDs is the same as for the WCDs, the values of 1000 m for r_{opt} and of 700 m for r_s were chosen (see Section 1.3.4). Obtaining the lateral distribution function means to derive a functional form for the shape parameters, β and γ , in terms of the shower size and the geometry. By fixing these parameters according to a certain parameterization we may fit LDFs in events with low multiplicity of SSDs. An ansatz for the functional form of β and γ was chosen to be linear in S_{1000} and quadratic in zenith angle, namely

$$\begin{aligned} \beta(S_{1000}, \theta) &= a_{\beta} + b_{\beta} s + (c_{\beta} + d_{\beta} s) \sec \theta + (e_{\beta} + f_{\beta} s) \sec^2 \theta \quad \text{and} \\ \gamma(S_{1000}, \theta) &= a_{\gamma} + b_{\gamma} s + (c_{\gamma} + d_{\gamma} s) \sec \theta + (e_{\gamma} + f_{\gamma} s) \sec^2 \theta, \end{aligned} \quad (3.13)$$

where $s = \lg S_{1000}$. This type of parameterization may be found in different parts of this thesis (as well as in many other related works). The dependency of the slope parameters on the zenith angle and on the shower size is due to the changes in the lateral distribution of shower particles with the shower age. The procedure for obtaining the values of the coefficients is detailed below.

Table 3.3: Coefficients of the β and γ parameterization of the SSD LDF.

Parameter	a	b	c	d	e	f
β	3.63	-0.55	-9.24	0.49	3.78	-0.19
γ	-19.59	3.30	25.16	-4.40	-9.28	1.59

Fit procedure

In order to obtain the parameterization of the slope parameters of the LDF, events from the simulation library in Table 3.1 were used. The goal of the parameterization is to model $f_{\text{NKG}}(r)$ based on the measured signals and distances to the core (S_i, r_i). To do so, the Monte Carlo values of the core position and geometry were used as well as an estimate of the signal at the optimum distance of 1000 m, which was obtained by averaging the signal of the 24 dense station SSDs placed at this distance.

For each event, a maximum-likelihood fit of the LDF is performed. The likelihood function can be written as a product of the probabilities of observing a signal S given its expectation \hat{S}

$$\mathcal{L} = \prod_{i=1}^N p(S_i | \hat{S}_i, \sigma_i), \quad (3.14)$$

where $p(S | \hat{S}, \sigma)$ is the p.d.f. described in Section 3.2.1 and σ_i is the uncertainty on the expected signal given by Eq. (3.10). Maximizing the probability in Eq. (3.14) is equivalent to minimizing the negative log-likelihood. One can then write the log-likelihood explicitly as

$$l = \sum_{i=1}^N \ln p = -\frac{1}{2} \ln(2\pi\sigma_i^2) - \frac{1}{2} \left(\frac{S_i - \hat{S}_i}{\sigma_i} \right)^2 - \ln \left[\frac{1}{2} \left(1 + \operatorname{erf} \left(\frac{\hat{S}_i}{\sqrt{2}\sigma_i} \right) \right) \right]. \quad (3.15)$$

The procedure for obtaining the parameterization involves the fit of individual events. Each event of the shower library was thrown two times in order to increase the multiplicity of stations for the fit of the shape parameters. As it was discussed at the end of Section 3.2.2, the signal uncertainty model does not describe low signals well. In order to minimize the impact of small signals in the likelihood fit, a cut on distance was introduced in such a way that 95% of the SSDs below that distance have a signal at least of 1 MIP. The distance was parameterized in terms of S_{1000} from the WCD and zenith angle (see Appendix C.2.2). On the other hand, low-gain saturated signals were not taken into account.

In the first step, each event was fit leaving β free and fixing γ to the initial guesses given by a parameterization obtained in previous works [87, 90]. After fitting all the events, β is parameterized based on Eq. (3.13) and first estimates of the coefficients are obtained. In the next step, β was fixed to the previous parameterization and γ was left free. Analogously, β is parameterized according to Eq. (3.13) and first estimates of the coefficients are found. This process is repeated iteratively until all the coefficients for β and γ converge (see Fig. C.9) to the values given in Table 3.3.

The dependency of these two parameters on $\sec\theta$ is shown in Fig. 3.6. The data points are obtained using the weighted averages \bar{x} with weights $\omega_i = 1/\sigma_i$, being σ_i the individual uncertainty on each data point. The uncertainty from the fit is used as no

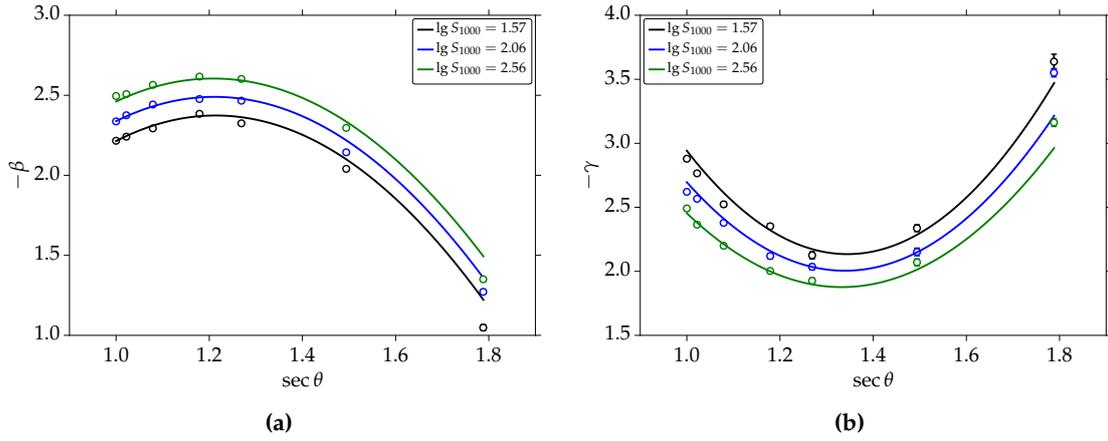


Figure 3.6: (a) The slope parameter β of the SSD LDF as a function of $\sec \theta$ for different bins in shower size. For comparison, the data points show the weighted average on each bin (see text for details). (b) Same as for the slope parameter γ .

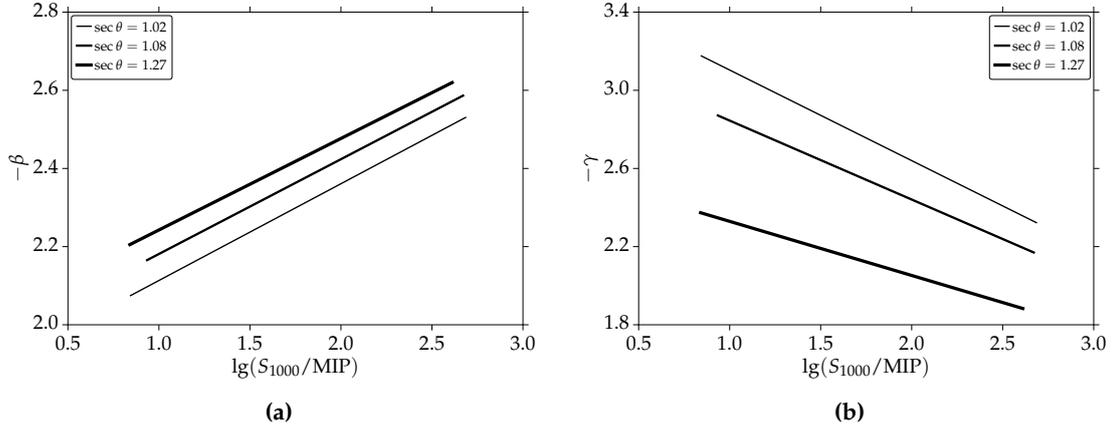


Figure 3.7: (a) β of the SSD LDF as a function of the logarithm of the shower size for different zenith angles. Line width increases with increasing zenith. (b) Same as for γ .

uncertainty model was derived for the slope parameters. The weighted average and its variance on a given bin with n entries is given by

$$\bar{x} = \frac{\sum_{i=1}^n x_i / \sigma_i^2}{\sum_{i=1}^n 1 / \sigma_i^2}, \quad \text{and} \quad (3.16)$$

$$\sigma^2[\bar{x}] = \frac{1}{n \sum_{i=1}^n 1 / \sigma_i^2} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma_i} \right)^2. \quad (3.17)$$

A larger deviation from the data at large zenith angles is observed. Furthermore, the model has limitations at higher zeniths where β turns into positive values, something which is non-physical as the signal must decrease with distances irrespective of the zenith angle. An alternative functional form for β to restrict its value to be always negative was proposed and can be seen in Fig. C.10. The dependency of the slope parameters with the logarithm of the shower size is shown in Fig. 3.7.

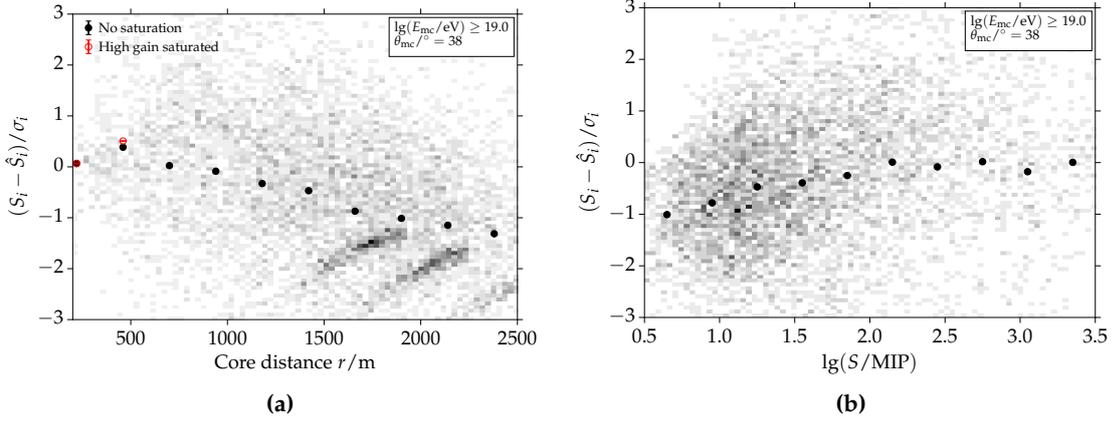


Figure 3.8: (a) Residuals of the LDF parameterization as a function of the distance to the shower core. (b) Residuals as a function of the logarithm of the signal. Proton and iron showers at 38° of zenith and energies above 10^{19} eV were used.

Validation of the parameterization

In order to evaluate the goodness of the parameterization of β and γ , the set of simulated events were reconstructed once again using the new parameterization. The residuals

$$\text{Res}(S_i) := \frac{S_i - \hat{S}_i}{\sigma[\hat{S}_i]}, \quad (3.18)$$

are given by the differences between the measured signals S_i and the expected signals \hat{S}_i relative to the uncertainties. Another useful quantity is the relative difference between the data and the predicted values from the model

$$\text{Rel}(S_i) := \frac{S_i - \hat{S}_i}{\hat{S}_i}. \quad (3.19)$$

As it can be seen in Fig. 3.8 and Fig. 3.9, the model gives an overall good description of the data over a wide range in distance and in signal size. The deviations become apparent at large distances (above 1300 m) and at small signals. The stripe-like structures observed in Fig. 3.8a might be due to signal deposited by individual particles.

The next important check is the bias in the reconstructed shower size. For each event, the value of the signal at the optimal distance obtained from the fit of the LDF (S_{1000}^{rec}) was compared to that obtained from averaging the signals of the dense ring of stations at a 1000 m ($\hat{S}_{1000}^{\text{mc}}$),

$$b = \frac{S_{1000}^{\text{rec}}}{\hat{S}_{1000}^{\text{mc}}} - 1. \quad (3.20)$$

An example for a fixed zenith is shown in Fig. 3.10a. The average bias for proton and iron showers is shown by the red and blue markers, respectively, and is within the $\pm 10\%$. On average the bias is almost zero over all energies, this is shown by the black markers. The average bias, $\langle b \rangle$, is shown as a function of the zenith and for different energies in Fig. 3.10b. It can be seen that the average bias is within $\pm 5\%$ for most of the energies and zeniths, and almost zero for $10^{19.5}$ eV. This can be seen in more detail in Fig. 3.11, where numbers are given. Fig. 3.11a shows the aforementioned average bias while Fig. 3.11b shows the difference between the bias in proton and iron showers. This difference is

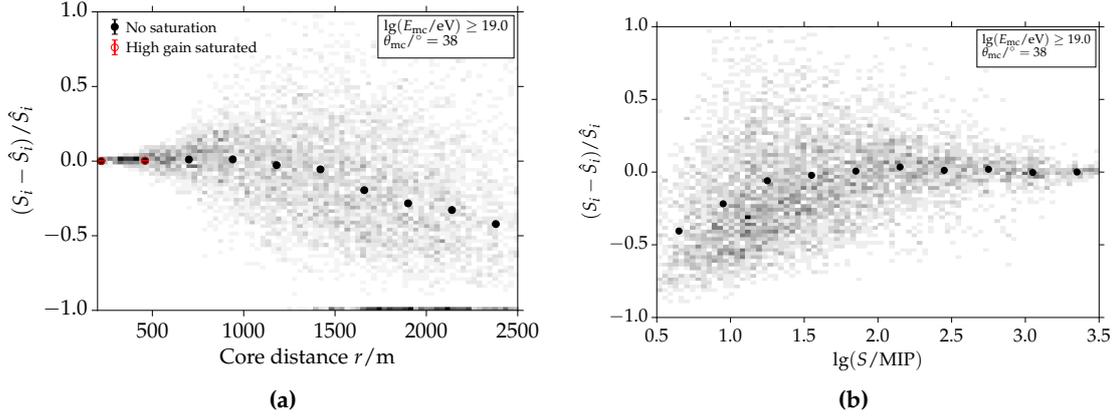


Figure 3.9: (a) Relative difference between data and model as a function of the distance to the shower core. (b) Relative difference as a function of the logarithm of the signal. The same showers were used as those in Fig. 3.8.

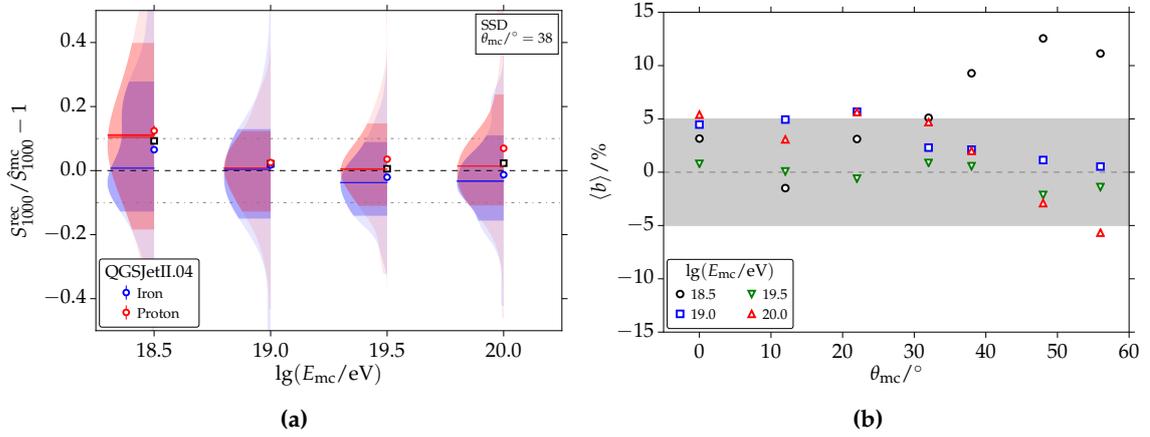


Figure 3.10: (a) Bias in S_{1000}^{ssd} as a function of the Monte Carlo energy for proton and iron primaries and showers at a zenith angle of 38° . The black squares show the average bias on each energy bin. The dotted lines indicate the $\pm 10\%$ region. (b) Average bias in S_{1000}^{ssd} as a function of the zenith angle and for different energies. The gray area shows the $\pm 5\%$ region. See text for details.

important as it gives information on how one primary is mis-reconstructed with respect to the other at a particular energy or zenith. Although there is no clear trend, it seems that S_{1000} from proton is underestimated with respect to iron at low energies and small zeniths and at the same time is overestimated in the case of iron. The differences between primaries become smaller at intermediate zeniths and energies. In general we can conclude that this configuration of the LDF gives a satisfactory representation of the SSD signals and will be used for the rest of the analyses presented in this thesis.

3.2.4 Reconstruction efficiency

In this section, we investigate the probability that a shower with a given energy and arrival direction can be reconstructed with the SSD. This should not be confused with the trigger efficiency of the SD, although both are related. The current implementation of the SSD reconstruction in Offline requires at least 3 candidate SSDs for the fit of the

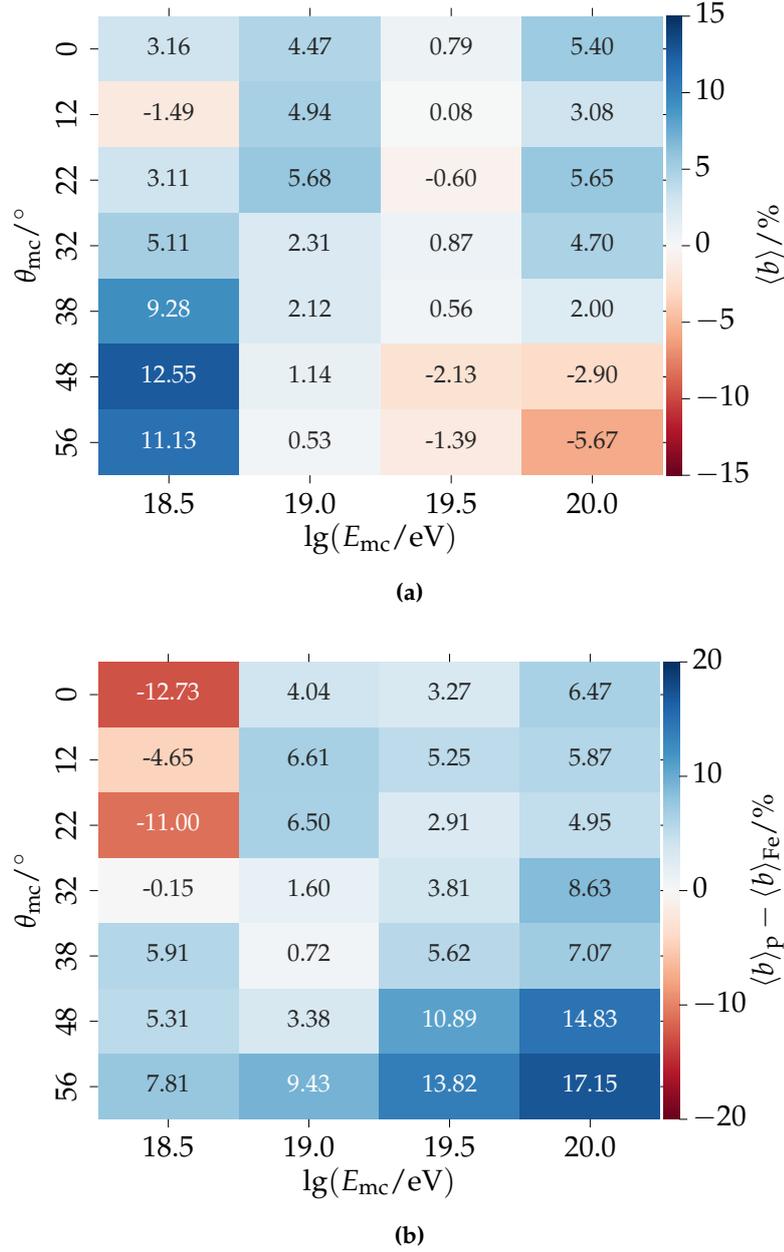


Figure 3.11: (a) Average bias in SSD S_{1000} for all zeniths and energies. (b) Relative bias between proton and iron for all zenith angles and energies used in the parameterization.

LDF, i.e., no saturation and within the distance cut applied for the parameterization of the LDF in Section 3.2.3.

The reconstruction efficiency is calculated as the ratio between the number of events for which the fit of the LDF was possible and the total number of events

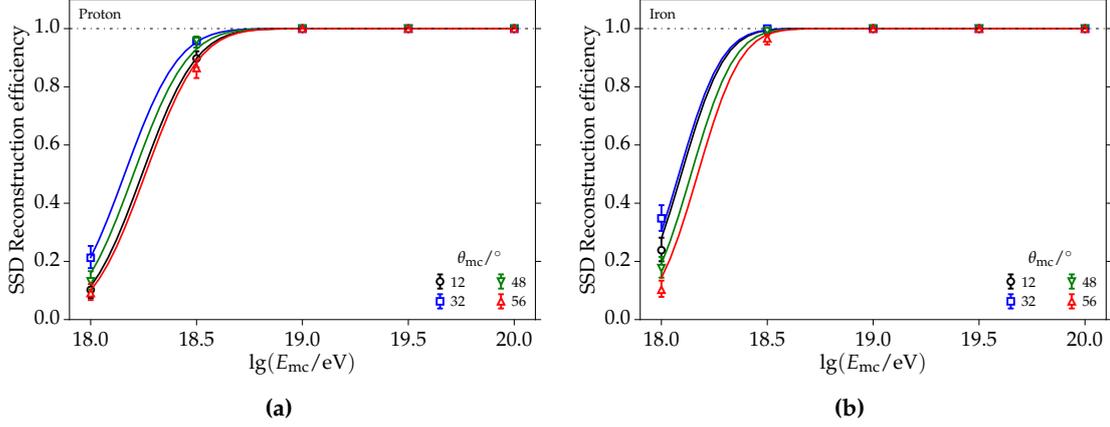
$$\epsilon = \frac{n_{\text{rec}}}{n_{\text{tot}}}. \quad (3.21)$$

In order to describe the reconstruction efficiency, we use the following empirical model

$$\epsilon(E, \theta) = \frac{1}{2} \left(1 + \operatorname{erf} \left(\frac{\lg E - a(\theta)}{b} \right) \right), \quad (3.22)$$

Table 3.4: Coefficients of the SSD reconstruction efficiency (Eq. (3.22)) for proton and iron showers.

Primary	a_0	a_1	a_2	a_3	b
Proton	18.42 ± 0.12	-0.53 ± 0.81	-0.16 ± 1.51	0.54 ± 0.84	0.29 ± 0.01
Iron	18.18 ± 0.13	0.22 ± 0.83	-1.00 ± 1.54	0.71 ± 0.86	0.23 ± 0.01

**Figure 3.12:** SSD reconstruction efficiency as a function of energy and for different zenith angles. (a) Proton and (b) iron showers.

with parameters a and b and the error function $\text{erf} := \frac{2}{\sqrt{\pi}} \int_0^x dy e^{-y^2}$. The b parameter is chosen to be constant while a has a dependency on the zenith angle as

$$a(\theta) = a_0 + a_1 \cos^2 \theta + a_2 \cos^4 \theta + a_3 \cos^6 \theta. \quad (3.23)$$

The parameterization of a is purely empirical and was taken from previous works that describe the trigger efficiency of the SD [42, 43]. The coefficients of the parameters are found by minimizing the following negative log-likelihood

$$-\ln \mathcal{L} = -\sum_i \ln p(n_{\text{rec}}; n_{\text{tot}}, \epsilon(E, \theta)), \quad (3.24)$$

where p denotes the probability mass function of the binomial distribution. The minimization is performed separately for proton and for iron showers and the results of the parameters are summarized in Table 3.4. The result of the parameterization of the reconstruction efficiency as a function of the Monte Carlo energy and for different zenith angles can be seen in Fig. 3.12. The values of the SSD reconstruction efficiency for each of energy and zenith are summarized in Fig. 3.13.

Above $10^{18.5}$ eV the SSD reconstruction is fully efficient for both primaries. At that energy (and below), the efficiency drops faster for proton than for iron. This is due to larger amount of muons in iron showers (around 40% more) making more likely the probability of producing a signal in the SSD, specially at larger distances from the core which are reached mostly by muons. The efficiency for protons is then the most conservative and will be taken as reference for future analyses.

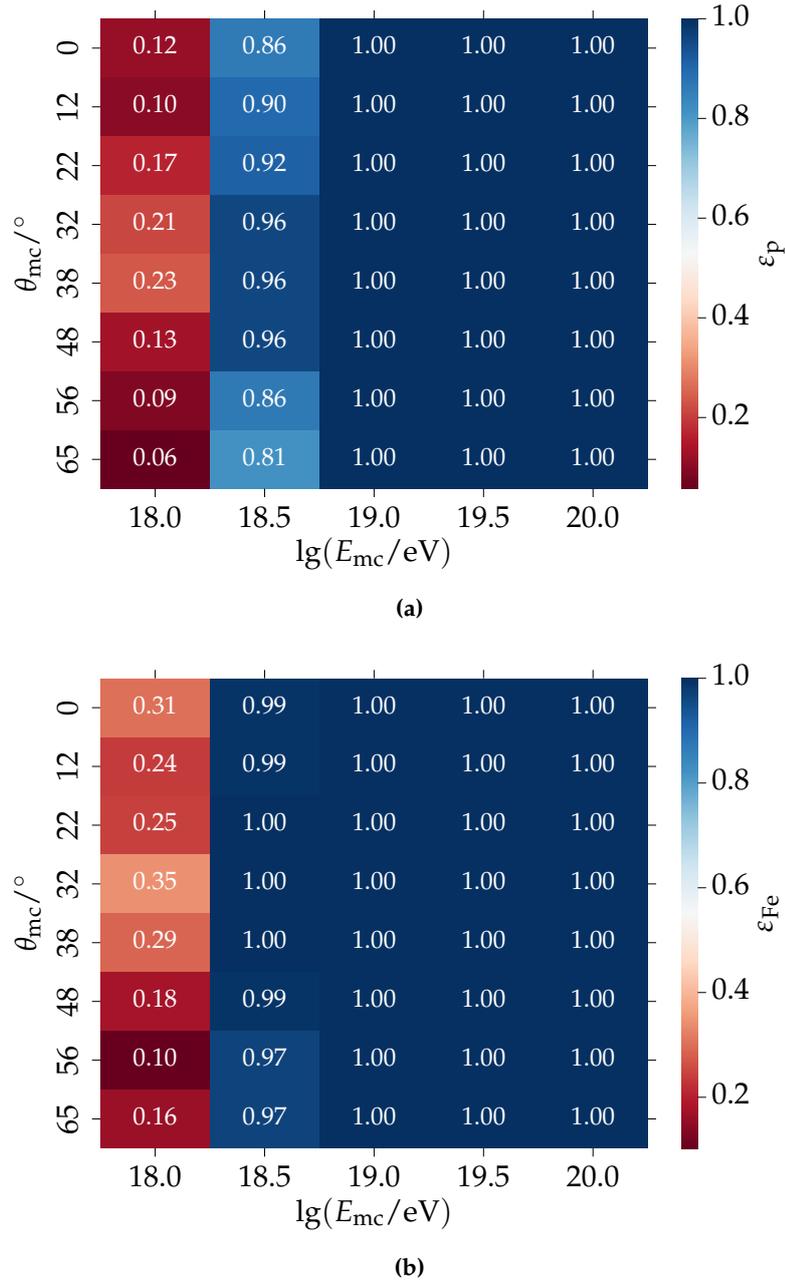


Figure 3.13: Values of the SSD reconstruction efficiency for discrete values of zenith angle and the logarithm of the energy. (a) Proton and (b) iron showers.

3.3 Estimate of the energy

The energy of the primary is estimated with the SD using the following expression

$$E_{SD} = A(S_{38}^{wcd})^B, \quad (3.25)$$

where $A = 0.178$ EeV and $B = 1.042$ are parameters obtained from a calibration using high-quality events measured with both the FD and the SD (see Section 1.3.5). In principle, one could write the energy calibration in terms of the SSD signal although this would have larger fluctuations with respect to the WCD as the SSD samples less amount

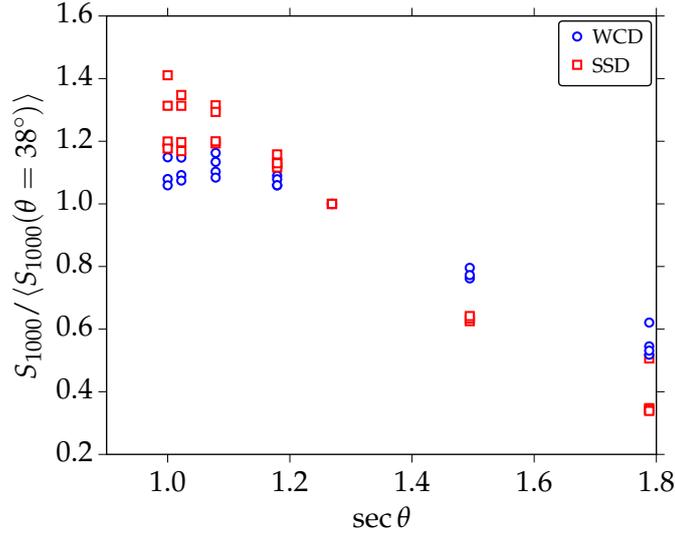


Figure 3.14: Attenuation with zenith in the reconstructed shower size with the WCD and the SSD. For each zenith and energy, the value of S_{1000} was normalized by its value at 38° at that energy.

of particles. In this section we will investigate how the additional information provided by the SSD might help in reconstructing the energy of the primary.

3.3.1 Attenuation correction to S_{1000}

For a given energy and primary, the shower size will depend on the zenith angle as the attenuation changes with the amount of atmosphere traversed by shower particles. The increase of the amount of traversed matter can be approximated by $1 / \cos \theta$, meaning that particles from a shower with zenith of 60° will traverse about twice the amount of matter as particles from a vertical shower. The attenuation in the atmosphere is mostly dominant in the electromagnetic component, as the probability of interaction for muons is smaller. This effect can be inferred by looking at Fig. 3.14, where the shower size measured by the SSD drops by about 70% between 0° and 56° while for the WCD the decrease is of about 45%.

In order to get an energy estimator which is zenith-independent, the value of the shower size must be corrected. This is done in a similar way to the correction obtained using the CIC method. In our case, the energy of the shower is known and, therefore no cut is needed. Assuming no dependency on primary mass and energy, the attenuation correction is obtained by comparing the average shower size (from all energies) to the size at a reference angle

$$S_{\theta_{\text{ref}}}^{\{\text{wcd,ssd}\}} = S_{1000}^{\{\text{wcd,ssd}\}} / f_{\text{Att}}(\theta), \quad (3.26)$$

where $\theta_{\text{ref}} = 38^\circ$. Similarly to the CIC, the functional form for the attenuation is chosen to be a third order polynomial in the variable $x = \cos^2 \theta - \cos^2(38^\circ)$, namely

$$f_{\text{Att}}(x) = 1 + a x + b x^2 + c x^3. \quad (3.27)$$

A χ^2 fit was performed and the results can be seen in Fig. 3.15. The resulting fit parameters and their uncertainties are summarized in Table 3.5.

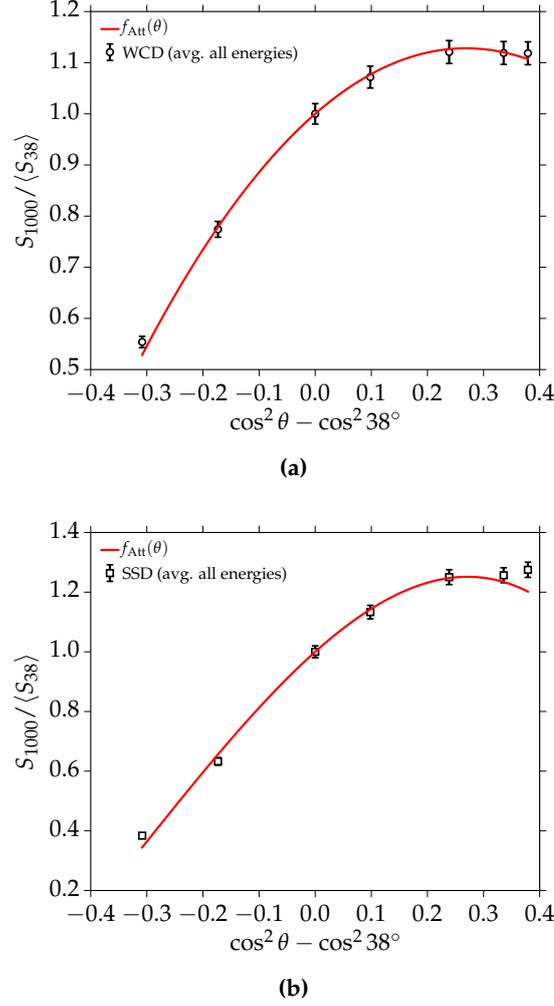


Figure 3.15: Parameterization of the attenuation function for the WCD and the SSD. The label $\langle S_{38} \rangle$ makes reference to the averaged S_{1000} at 38° .

Table 3.5: Coefficients for the parameterization of f_{Att} (Eq. (3.27)) for the WCD and the SSD.

Detector	a	b	c	χ_{red}^2
WCD	0.9600 ± 0.0006	-1.8202 ± 0.0008	0.0812 ± 0.0065	6.93
SSD	1.6817 ± 0.0013	-2.1558 ± 0.0013	-2.3105 ± 0.0125	6.74

3.3.2 Unbiased energy estimate

The energy of the primary is reconstructed with the SD using Eq. (3.25). Now, a proton of energy E_0 will produce a shower for which the value of S_{1000} is on average smaller than the one produced by an iron of the same energy. This is primarily due to the different number of muons produced in the corresponding air showers, which is $\sim 40\%$ more numerous for iron than for proton. From Eq. (3.25), it follows that the resulting energy for proton showers will be underestimated with respect to iron showers. In Fig. 3.16, the energy of the primary as reconstructed with the SD, E_{rec} , is compared to the “true” Monte Carlo energy. The absolute offset of around 20% is due to the lack of muons in simulations with respect to real data (see, for example, [57]), and could be corrected for. The differences between proton and iron showers are due to the aforementioned effect,

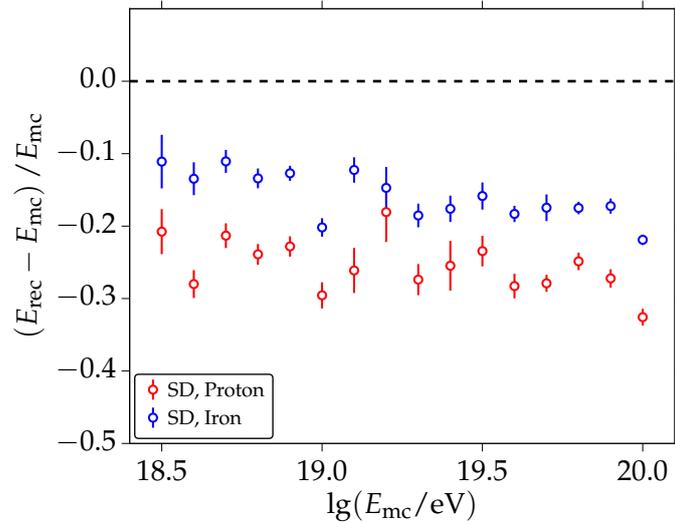


Figure 3.16: The reconstructed energy is compared to the “true” Monte Carlo energy for proton (red) and iron (blue) showers.

resulting in a mass-dependent bias in the reconstructed energy. The SSD is less sensitive to the muonic component of the shower. Therefore it might aid in reducing the bias in energy that is due to the mass of the primary. On the other hand, the SSD is subject to larger fluctuations (due to its smaller size). The values of S_{38} measured with the WCD and SSD for proton and iron showers are shown in Fig. 3.17. It can be seen that for a given energy, the values of S_{38} for proton and iron showers are similar in the case of the SSD.

Principal Component Analysis using SSD and WCD data

Principal Component Analysis (PCA) is a statistical method in which a given dataset of (possibly) correlated variables is transformed into a new subspace where the transformed variables are uncorrelated. This subspace is defined by the directions (also called principal components) which maximize the variance of the data. PCA is often used for dimensionality reduction, projecting a d -dimensional dataset onto a k -dimensional subspace (with $k < d$) retaining most of the information. The transformation is done in such a way that the first principal component has the largest variance of the data, the second component indicates the second greatest variance, and so on. One way to apply PCA is performing the decomposition in eigenvalues and eigenvectors of the covariance matrix, where the eigenvectors give the directions of the principal components and the eigenvalues the strength of the variance along its axis³.

Our dataset is composed of four pieces of information: the estimates of the shower size from the WCD and the SSD, the energy of the shower and the mass of the primary. It can be depicted as a 4-dimensional vector

$$\mathbf{x}^T = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} s_{\text{wcd}} \\ s_{\text{ssd}} \\ \mathcal{E} \\ a \end{pmatrix}, \quad (3.28)$$

³For a nice description of PCA and its implementation in Python see [91].

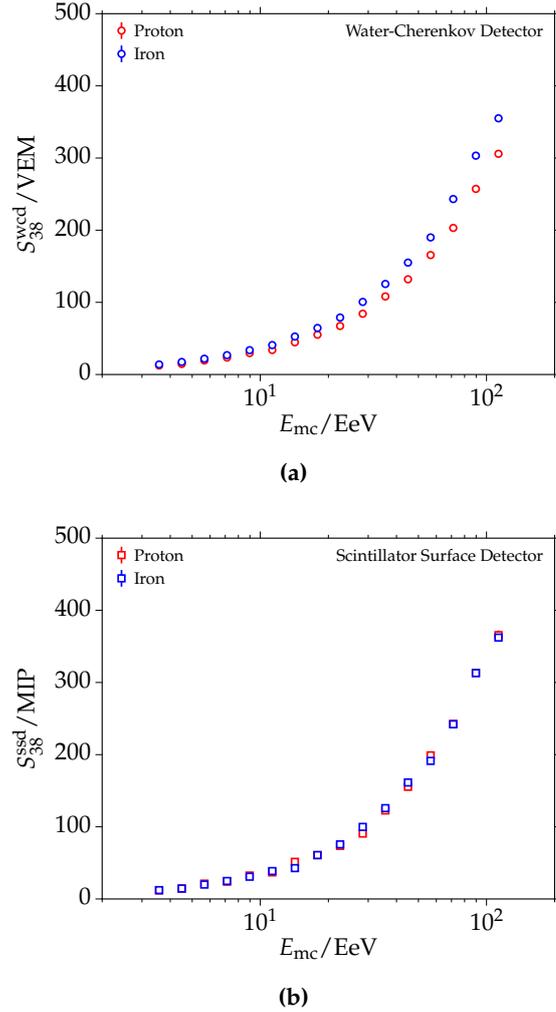


Figure 3.17: S_{38} as a function of the primary energy for proton and iron showers. (a) Measured with the WCD and (b) with the SSD.

where we have defined the following variables $s_{\text{wcd}} := \lg S_{38}^{\text{wcd}}$, $s_{\text{ssd}} := \lg S_{38}^{\text{ssd}}$ and $\mathcal{E} := \lg E_{\text{mc}}$ and $a := \ln A$.

The first step is to perform the *eigendecomposition* of the covariance matrix

$$\Sigma = \frac{1}{1-n} ((\mathbf{x} - \bar{\mathbf{x}})^{\top} (\mathbf{x} - \bar{\mathbf{x}})), \quad (3.29)$$

where $n = 4$ and $\bar{\mathbf{x}}$ is the mean vector, i.e., a d -dimensional vector containing the means of the dataset $\bar{\mathbf{x}} = (\langle x_1 \rangle, \langle x_2 \rangle, \langle x_3 \rangle, \langle x_4 \rangle)$. The eigenvectors and eigenvalues are obtained by diagonalizing the covariance matrix, i.e., if \mathbf{v} is a d -dimensional vector which satisfies

$$\Sigma \mathbf{v} = \lambda \mathbf{v} \quad (3.30)$$

then \mathbf{v} is an eigenvector of Σ with eigenvalue λ . The decomposition was performed using the Python package Scikit-learn [92] yielding to the following four eigenvectors

$$\mathbf{v}_1^\top = \begin{pmatrix} 0.0111 \\ -0.5536 \\ 0.7476 \\ 0.3666 \end{pmatrix}, \mathbf{v}_2^\top = \begin{pmatrix} -0.0102 \\ -0.5985 \\ -0.0510 \\ -0.7994 \end{pmatrix}, \mathbf{v}_3^\top = \begin{pmatrix} -0.0075 \\ -0.5790 \\ -0.6620 \\ 0.4758 \end{pmatrix}, \mathbf{v}_4^\top = \begin{pmatrix} 0.9999 \\ -0.0043 \\ -0.0138 \\ -0.0088 \end{pmatrix}, \quad (3.31)$$

with eigenvalues

$$\lambda_1 = 4.0522, \lambda_2 = 0.6022, \lambda_3 = 0.0016, \lambda_4 = 0.0075. \quad (3.32)$$

The eigenvectors constitute the new basis. However, they only give the direction of the new axes as, they all have the same unit length 1. The idea behind PCA is to choose those eigenvectors which represent best our data when projecting the data onto the new subspace. In order to select which eigenvectors should be dropped without losing too much information we have to look at the eigenvalues. The eigenvectors with lower eigenvalues carry the least information about the distribution of the data and can be dropped. From Eq. (3.32), it can be seen that the first two eigenvalues have the largest value. A useful quantity that can be computed in order to choose the number of principal components is the explained variance. This quantity tells us how much variance is related to each of the components. For each eigenvector the explained variance is calculated as

$$var_{\text{exp}}(\mathbf{v}_i) = \lambda_i / \sum_{j=1}^n \lambda_j,$$

yielding to the following values: 86.89%, 12.91%, 0.17% and 0.03% respectively. This means that the first two components, \mathbf{v}_1 and \mathbf{v}_2 contain 99.8% of the information and the other two can be neglected. We can express our data in terms of the two chosen vectors

$$\mathbf{x} \approx p \mathbf{v}_1 + q \mathbf{v}_2 + \dots, \quad (3.33)$$

where p and q are intrinsic quantities of the shower in the new subspace. Eq. (3.33) can be written explicitly as

$$\begin{pmatrix} s_{\text{wcd}} \\ s_{\text{ssd}} \end{pmatrix} \approx p \begin{pmatrix} v_{11} \\ v_{12} \end{pmatrix} + q \begin{pmatrix} v_{21} \\ v_{22} \end{pmatrix} \quad (3.34)$$

and

$$\begin{pmatrix} \mathcal{E} \\ a \end{pmatrix} \approx p \begin{pmatrix} v_{13} \\ v_{14} \end{pmatrix} + q \begin{pmatrix} v_{23} \\ v_{24} \end{pmatrix}. \quad (3.35)$$

This means that we can find an estimate for the reconstructed energy and an estimate for the primary mass using a linear combination of s_{wcd} and s_{ssd} . Solving for p and q in Eq. (3.34)

$$\begin{aligned} p &= \frac{1}{\mathcal{D}} (v_{22} s_{\text{wcd}} - v_{21} s_{\text{ssd}}) \quad \text{and} \\ q &= \frac{1}{\mathcal{D}} (v_{11} s_{\text{ssd}} - v_{12} s_{\text{wcd}}), \end{aligned} \quad (3.36)$$

where $\mathcal{D} = (v_{11} v_{22} - v_{21} v_{12})$. We can use the solutions in Eq. (3.36) to extract estimates of the energy and mass of the primary, namely

$$\mathcal{E} = \frac{1}{\mathcal{D}} [v_{13} (v_{22} s_{\text{wcd}} - v_{21} s_{\text{ssd}}) + v_{23} (v_{11} s_{\text{ssd}} - v_{12} s_{\text{wcd}})], \quad (3.37)$$

and

$$a = \frac{1}{\mathcal{D}} [v_{14} (v_{22} s_{\text{wcd}} - v_{21} s_{\text{ssd}}) + v_{24} (v_{11} s_{\text{ssd}} - v_{12} s_{\text{wcd}})]. \quad (3.38)$$

The solutions in Eq. (3.37) and Eq. (3.38) are estimates of the energy and of the logarithmic mass based on linear combinations of s_{wcd} and s_{ssd} through PCA. They are just approximations as we have used only two principal components to describe the data. Something to take into account is that before performing PCA, the data has to be standardized, which means transforming the data in such a way that its mean is 0 and the standard deviation is 1. The resulting variables are "mean-subtracted" therefore we need to add the mean of each variable ($\langle s_{\text{wcd}} \rangle$, $\langle s_{\text{ssd}} \rangle$, $\langle \mathcal{E} \rangle$ and $\langle a \rangle$) to obtain the correct result. We can write the estimates of the energy and the logarithm of the mass as follows

$$E_{\text{pca}} = 10^{\mathcal{E}} 10^{\langle \mathcal{E} \rangle}, \quad (3.39)$$

where

$$\mathcal{E} = \mathcal{E} (s_{\text{wcd}} - \langle s_{\text{wcd}} \rangle, s_{\text{ssd}} - \langle s_{\text{ssd}} \rangle)$$

and

$$\ln A_{\text{pca}} = a + \langle a \rangle. \quad (3.40)$$

The subindex "pca" in Eq. (3.39) and in Eq. (3.40) denotes that these are just estimates that correlate with the "true" energy and logarithmic mass, which are indeed properties of the primary cosmic ray. We will return to Eq. (3.40) and its implications for discriminating the mass of the primary at a later point (see Section 3.4). If we now expand Eq. (3.39) we get

$$E_{\text{pca}} = C \left(S_{38}^{\text{wcd}} \right)^{\alpha} \left(S_{38}^{\text{ssd}} \right)^{\beta}, \quad (3.41)$$

where C , α and β are constants involving the components of the eigenvectors

$$E_{\text{pca}} = 27.01 \left(\frac{S_{38}^{\text{wcd}}}{\text{VEM}} \right)^{0.12} \left(\frac{S_{38}^{\text{ssd}}}{\text{MIP}} \right)^{0.86} \text{ EeV}. \quad (3.42)$$

This solution is very similar to the current energy estimator with the SD, now including the SSD signal. In fact, if $S_{38}^{\text{wcd}} \sim S_{38}^{\text{ssd}}$ Eq. (3.42) reduces to

$$E_{\text{pca}} \sim A \left(S_{38}^{\text{ssd}} \right)^B, \quad (3.43)$$

where A and B would be of similar values as in Eq. (3.25). In Fig. 3.18 the difference between the energy as obtained using PCA, E_{pca} , and the Monte Carlo energy is shown for proton and iron showers. It can be seen that using the SSD information, the mass-dependent bias in the reconstructed energy can be reduced, in particular at the highest energies.

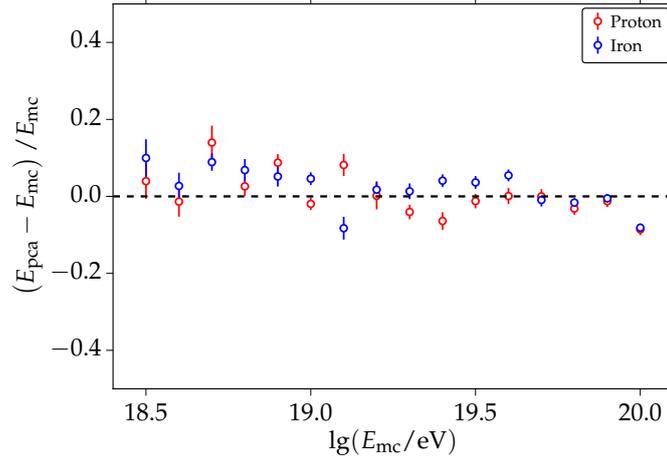


Figure 3.18: Bias in reconstructed energy for proton (red) and iron (blue) showers using WCD and SSD data through PCA. See text for the discussion.

The muonic and the electromagnetic components grow at different rates with energy. Given the different response of the WCD and the SSD to the shower components, there might be possibility to find a suitable combination of the two signals that helps to reduce the biases in energy. The WCD gives a more precise reconstruction of the signal at the optimal distance. On the other hand, the reconstructed signal at the optimal distance with the SSD has larger fluctuations (due to its smaller detection area) but also less variation between proton and iron showers.

The approach presented here results in a linear combination of $\lg S_{38}^{\text{wcd}}$ and $\lg S_{38}^{\text{ssd}}$. This linear ansatz is a first order solution, but it does not mean that is the correct one. Another attempt was performed by fitting the energy to the signals of the WCD and the SSD combined in a non-linear ansatz, namely

$$\lg E = A \left(\lg S_{38}^{\text{wcd}} \right)^{\alpha} + B \left(\lg S_{38}^{\text{ssd}} \right)^{\beta}. \quad (3.44)$$

The results can be found in Appendix C.2.5. On average the reconstructed energy is unbiased with respect to the Monte Carlo energy but the difference between proton and iron showers could not be corrected in this way.

The term “signal” actually denotes a convolution of the energy and geometry spectra of the different shower components with the response of two sub-detectors. This convolution should be hidden in the α and β parameters on Eq. (3.41), and should reflect the dependency of the muonic and electromagnetic components with energy. More dedicated analyses could be done in order to study any kind of subtleties, for example by looking the behavior of the detector responses to the different components as a function of the energy, zenith, mass, or any other observable which might have an impact on the energy reconstruction. Unfortunately such a dedicated studies are out of the scope of this thesis.

3.4 Sensitivity to mass composition

In this section, we explore the power of AugerPrime’s SSD for discriminating the mass of the primary cosmic ray. As it was discussed already, the main objective is the separate measurement of the number of muons and electrons, since their relationship with the primary mass is well understood.

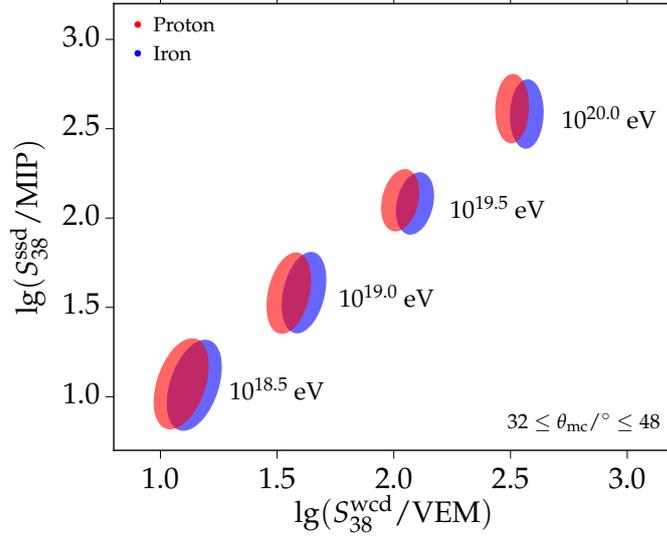


Figure 3.19: 2σ contours of the distributions of $\lg S_{38}^{\text{wcd}}$ and $\lg S_{38}^{\text{ssd}}$ for proton and iron showers at different energies and for zeniths in the interval $\theta_{\text{mc}} / ^\circ \in [32, 48]$.

One common approach is to combine the measurements of the two detectors to obtain an observable which is sensitive to the mass of the primary. Then, one can study how this sensitive depends on the energy, geometry and other shower properties. A useful tool to quantify the mass separation is the so-called *merit factor*. Assuming we have an observable \mathcal{O} obtained from some measurement with our detector(s) and we want to know the strength of this observable in order to discriminate between proton or iron showers, we can define the merit factor as

$$\text{MF} = \frac{|\langle \mathcal{O}_p \rangle - \langle \mathcal{O}_{\text{Fe}} \rangle|}{\sqrt{\sigma^2(\mathcal{O}_p) + \sigma^2(\mathcal{O}_{\text{Fe}})}}, \quad (3.45)$$

where the numerator is just the difference (in absolute value) between the means of the distributions and the denominator is the square root of the standard deviation of each distribution added in quadrature.

3.4.1 Separation power at a fixed energy

In this section, we will investigate the potential for mass discrimination by evaluating pairs of SSD and WCD measurements, for example the reconstructed signal at the optimal distance, at different zenith angles and energies. Fig. 3.19 shows the 2σ contours of the distributions of $\lg S_{38}^{\text{wcd}}$ and $\lg S_{38}^{\text{ssd}}$ for proton and iron showers in a certain zenith range. The different clouds correspond to different energies (as we used the fixed library). It can be seen how the distributions are wider for the SSD signal. On the other hand there is a variation on the slope of the cloud with energy which might be related to the different growing rates of the muonic and electromagnetic component. In the lines below we will explain how Linear Discriminant Analysis (LDA) can help to study the separability between the two masses.

Linear Discriminant Analysis using SSD and WCD data

Linear Discriminant Analysis (LDA) is a statistical method used for dimensionality reduction in different applications. The approach is very similar to a Principal Component

Analysis (see Section 3.3.2), but instead of finding the component axes that maximize the variance of our data (PCA) we are interested in the axes that maximize the separation between multiple classes. In LDA, data is projected onto a smaller subspace while keeping the class information. PCA is often described as an “unsupervised” method as it “ignores” the class labels whereas LDA is a “supervised” algorithm that computes the directions, or linear discriminants, that represent the axes which maximize the separation between the classes⁴. In this case our dataset will be composed just by the signal of the SSD and the WCD

$$\mathbf{x}^T = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} s_{\text{wcd}} \\ s_{\text{ssd}} \end{pmatrix}, \quad (3.46)$$

where again $s_{\text{wcd}} := \lg S_{38}^{\text{wcd}}$ and $s_{\text{ssd}} := \lg S_{38}^{\text{ssd}}$. In order to find the linear discriminants, we first need to compute the so-called scatter matrices. First of all, we compute the *within-class scatter* matrix S_W using the following expression

$$S_W = \sum_{i=1}^c S_i, \quad (3.47)$$

where

$$S_i = \sum_{j=1}^n (\mathbf{x}_j - \bar{\mathbf{x}}_i)(\mathbf{x}_j - \bar{\mathbf{x}}_i)^T. \quad (3.48)$$

This is the scatter matrix for every class c which in our case are the proton and the iron classes. $\bar{\mathbf{x}}$ is the mean vector which is computed as the mean of all the dataset. Eq. (3.48) is similar to the covariance matrix on each class. The second step is to compute the *between-class scatter* matrix S_B which is defined as

$$S_B = \sum_{i=1}^c N_i (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})^T, \quad (3.49)$$

where $\bar{\mathbf{x}}$ is the overall mean and $\bar{\mathbf{x}}_i$ and N_i are the mean of the dataset and the size of each class respectively. The next step is to diagonalize the matrix $S_W^{-1} S_B$ in order to obtain the linear discriminants. These will be the eigenvectors \mathbf{v} of the equation

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}, \quad (3.50)$$

with $\mathbf{A} = S_W^{-1} S_B$ and λ the corresponding eigenvalues. The eigenvector with highest eigenvalue represents the axis which maximizes the separation between the two classes.

We can use that eigenvector to project our data into a new subspace where the class separation is maximal. An example can be seen in Fig. 3.20. The distributions of $\lg S_{38}^{\text{wcd}}$ and $\lg S_{38}^{\text{ssd}}$ for proton and iron showers at an energy of $10^{19.5}$ eV are shown in Fig. 3.20a. The white line shows the best separation between the two clouds of data obtained using LDA as explained above. The projection of each of the data points onto the the new subspace is denoted by \mathcal{M} and its distribution for proton and iron data is shown in Fig. 3.20b. A merit factor of 1.32 ± 0.04 was obtained using Eq. (3.45), showing a good separation between the two classes at this energy. In Fig. 3.21a, the merit factor is shown as a function of the energy. The uncertainty on the merit factor is estimated using the bootstrap method (a detailed description is given in Appendix C.2.6). In Fig. 3.21b, a two dimensional representation of the merit factor for all energies and zenith angles is shown. It can be seen that the separation power decreases for high zeniths and low energies. This could be due to different reasons:

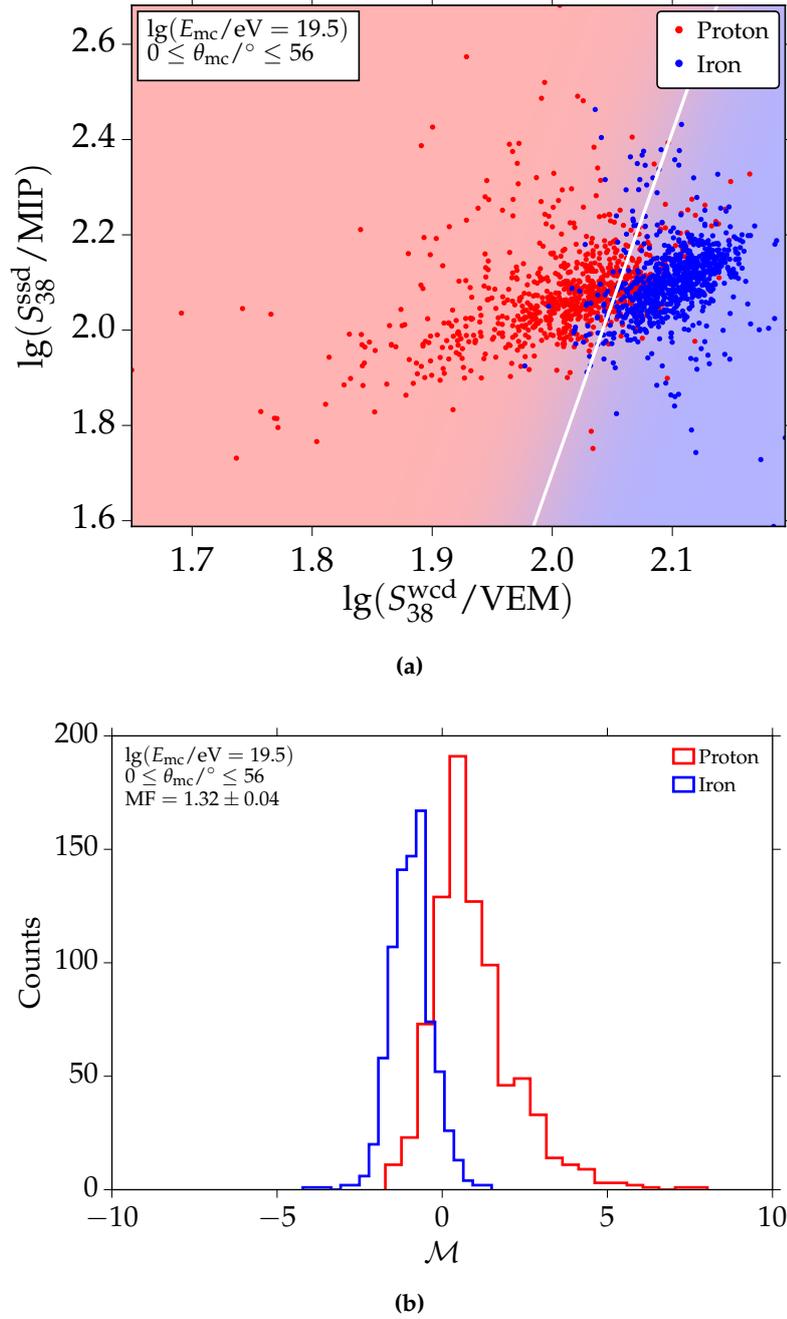


Figure 3.20: (a) Distributions of $\lg S_{38}^{\text{ssd}}$ and $\lg S_{38}^{\text{wcd}}$ for proton and iron showers at $10^{19.5}$ eV and all zenith angles. The white line shows the direction of the vector which maximizes the separation between the two classes. (b) Distribution of the observable \mathcal{M} for proton and iron showers at $10^{19.5}$ eV. \mathcal{M} is obtained after projecting the data onto the axis that separates the two classes (white line in (a)). A merit factor of 1.32 was obtained.

1. The fluctuations in signals, both for the WCD and for the SSD are larger at smaller signal sizes, i.e., large zeniths and low energies. This worsens the fit of the LDF and induces fluctuations in the value of S_{1000} (and therefore in S_{38}).

⁴A nice description of LDA and implementation in Python can be found in [93].

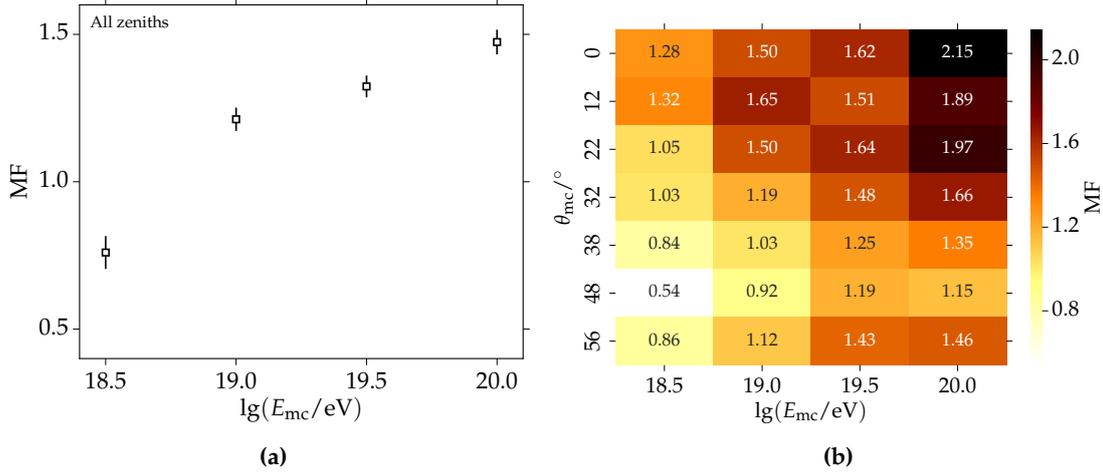


Figure 3.21: Merit factors showing the discrimination power between proton and iron showers when using Linear Discriminant Analysis on WCD and SSD data. (a) Merit factor as a function of the Monte Carlo energy using all zenith angles. (b) Summary of all merit factors obtained for each zenith and energy.

2. A systematic effect could be product of the dependency of S_{38} with the mass (and energy) of the primary. In Section 3.3.1 the attenuation is described as a purely geometric effect which is not true, since proton showers contain (relatively) more electromagnetic component and therefore will be more attenuated. This dependency could have an impact widening the signal distributions and thereby artificially decreasing the separation.
3. Another feature (which can be seen in the example in Fig. 3.20) is the presence of tails in the signal distributions. This could be due to mis-reconstructed events, e.g., a bad fit of the LDF, and would need a systematic examination.
4. Shower-to-shower fluctuations make the distributions broader as the shower size from identical primaries (and energies) will be affected by the statistical fluctuations in the shower development.

3.4.2 Results from Principal Component Analysis

Earlier in this Chapter we introduced the concept of Principal Component Analysis applied on WCD and SSD data, and we obtained expressions for the energy (Eq. (3.39)) and for the logarithmic mass (Eq. (3.40)) of the primary. The energy case was already discussed in Section 3.3.2. Here we will focus on the results regarding the mass estimator.

The estimate of the logarithmic mass as obtained from PCA is shown as a function of the Monte Carlo energy in Fig. 3.22. On average, the estimate of PCA for the logarithmic mass is very close to the true value but there is a strong dependency with the energy. The fluctuations of the data points are produced by the limited statistics of the fixed library. The bias in energy suggests the limitations of the PCA for reconstructing both the energy and the mass at the same time. It must be said that PCA gives just an approximation based in linear combinations of the input variables, and that we have selected only two principal components in order to describe the initial four-dimensional dataset. Some information is therefore lost in the process.

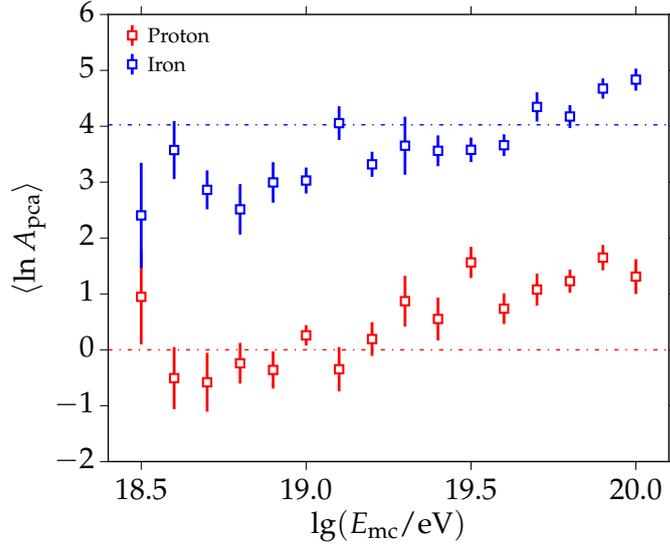


Figure 3.22: Average of the estimate of the logarithmic mass as obtained applying PCA with WCD and SSD data, as a function of energy. Markers show the mean on each energy bin and the error bars indicate the error on the mean. The dashed lines show the true values of $\ln A$ for proton (red) and iron (blue).

Table 3.6: Coefficients for the parameterization of $\ln A_{\text{rec}}$ with energy (Eq. (3.51)) for proton and iron showers.

Primary	m	n	χ^2/ndof
Proton	1.543 ± 0.194	19.00 ± 0.08	14.47/12
Iron	1.547 ± 0.173	17.05 ± 0.28	16.81/12

In order to correct for the bias in energy the values of $\ln A_{\text{pca}}$ were fitted to a linear function in energy

$$\ln A_{\text{pca}} = m (\lg E - n). \quad (3.51)$$

The parameters are summarized in Table 3.6 and the results of the fit can be seen in Fig. 3.23b. For the fit the first two bins in energy were not used as they might be affected by the efficiency threshold. The uncertainty of the fit is shown by the shaded area and is calculated using error propagation with correlated variables

$$\sigma_f^2 = \left(\frac{\partial f}{\partial m} \right)^2 \sigma_m^2 + \left(\frac{\partial f}{\partial n} \right)^2 \sigma_n^2 + 2 \frac{\partial f}{\partial m} \frac{\partial f}{\partial n} \text{cov}(m, n), \quad (3.52)$$

where f is given by Eq. (3.51) and $\text{cov}(m, n)$ denotes the covariance between the fit parameters m and n . The values of the slope for proton and iron are compatible within their uncertainties. This suggests that the deviation in energy is a systematic effect caused by the dimensionality reduction of the PCA and/or the assumptions taken. We could use the result of the fits to correct the bias in energy of $\ln A_{\text{pca}}$. This correction would be independent of the mass and the resulting estimate would still be bias with respect to the true $\ln A$ (since the intersecting point with the energy differs for both primaries). One could overcome this problem by defining the following correction

$$\ln A'_{\text{pca}} := \frac{\Delta \ln A_{\text{true}}}{f_{\text{Fe}} - f_{\text{p}}} (\ln A_{\text{pca}} - f_{\text{p}}), \quad (3.53)$$

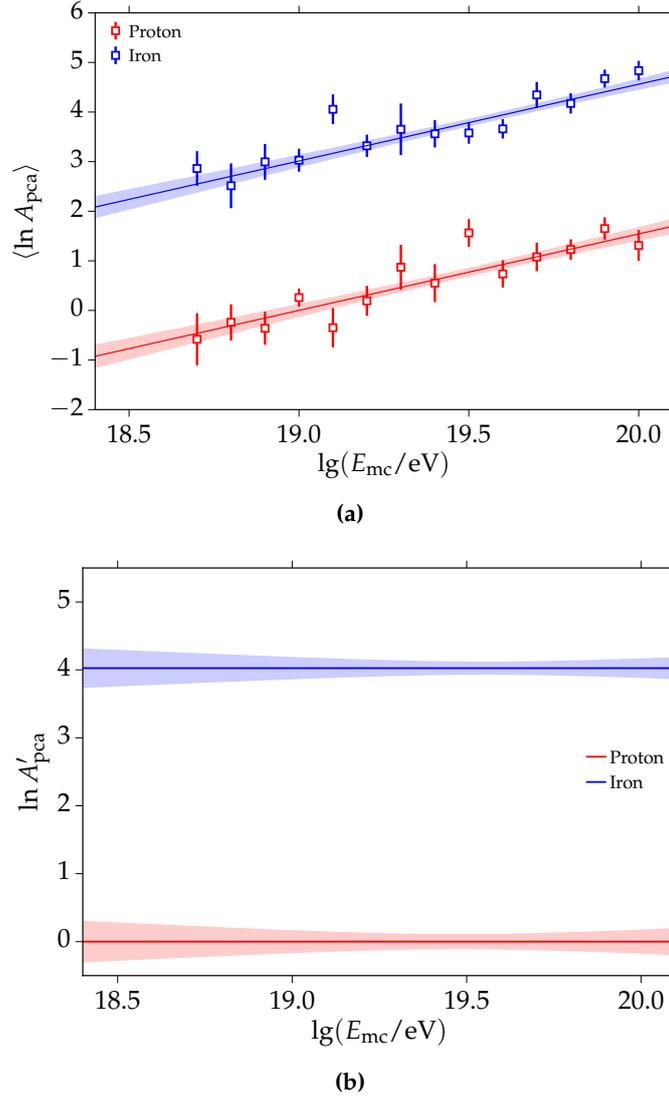


Figure 3.23: (a) Fit of $\ln A_{\text{pca}}$ to a linear function in energy as in Eq. (3.51) for proton and iron. The first two bins in energy (on both primaries) were not used for the fit. The bands show the error on the fit and were calculated using error propagation accounting for correlation in the parameters. (b) Result of $\ln A'_{\text{pca}}$ as a function of the energy after corrections were applied. See text for the discussion.

where $\Delta \ln A_{\text{true}} = \ln A_{\text{Fe}} - \ln A_{\text{p}} = 4.025$, and $f_{\text{p}} (f_{\text{Fe}})$ is the result of Eq. (3.51) for proton (iron). By construction, $\ln A'_{\text{pca}}$ is unbiased with respect to the energy and correlates with the true values of $\ln A$ for proton and iron. Results are shown in Fig. 3.23b. The uncertainties in $\ln A'_{\text{pca}}$ are represented by the shaded bands and were calculated propagating the error in Eq. (3.53).

The distributions of $\ln A'_{\text{pca}}$ obtained for proton and iron showers with energies above $10^{19.7}$ eV are shown in Fig. 3.24. The merit factor was computed according to Eq. (3.45) resulting in a value of 0.68 ± 0.04 . The results for different energies is shown in Fig. 3.25. The violin plot in Fig. 3.25a show the distributions of the mass estimator at specific energies. On each violin, the distribution is represented by the darker area and the mean and the median on each bin are shown by the markers and the solid lines, respectively. It can be seen that on average, the predicted value of the mass estimator is in agreement with the true value. On the other hand, the large spread of the distributions indicate a

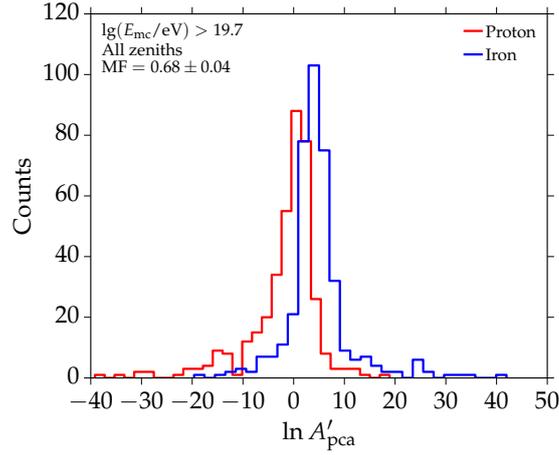


Figure 3.24: Distribution of $\ln A'_{\text{pca}}$ for proton and iron showers above $10^{19.7}$ eV using zenith angles in the range $0 \leq \theta_{\text{mc}}/\text{°} \leq 56$.

poor resolution which can also be seen in the merit factors shown in Fig. 3.25b, where in contrast to the case of the fixed library, the results from PCA indicate a smaller separation power. There could be different reasons to explain this effect. On one side, we have to take into account that the algorithm is using information of only two primaries, proton and iron. PCA is intended to work with the variance of the data, and having only two extreme cases for the variable $\ln A$ does not allow for much freedom in the optimization. In Section 3.3, we saw that the reconstructed energy, $\lg E_{\text{pca}}$, is in good agreement with the true energy, $\lg E_{\text{mc}}$. A good test would be to run PCA using more elements such as helium or nitrogen, and see if the resolution in the predicted $\ln A$ improves. Unfortunately sufficient simulations over a wide range of energies was bit available to perform such exercise.

Furthermore, the tails of the distributions might be affected (up to some extent) by the fact that we are combining showers with different zenith angles and also by the biases in the reconstruction. Regarding this last point, we have seen that biases in the reconstructed S_{1000}^{ssd} can be of the order of 15% between proton and iron showers (see Fig. 3.11). In addition, the assumption that the attenuation correction does not depend on the mass of the primary might also induce a systematic error. All source of systematics that might affect the resolution in our mass estimator must be studied in detail.

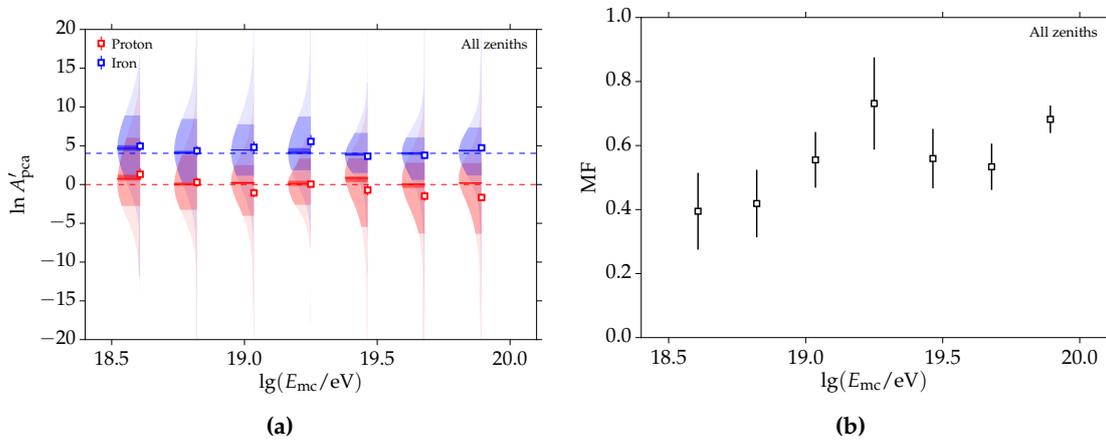


Figure 3.25: (a) Distributions of $\ln A'_{pca}$ for proton and iron showers at different energies. In each of the energy bin, the distribution of the data is shown by the shaded-lighter area. The inner line represents the median and the markers show the mean. Dashed lines show the true value of $\ln A$ for both primaries. (b) Merit factors computed for each of the distributions shown in Fig. 3.25a.

CHAPTER 4

Data analysis from AugerPrime detectors

During September of 2016, twelve prototype units of scintillator detectors and upgraded electronics were deployed in the field as part of the Engineering Array. The locations of these upgraded stations are indicated with yellow squares in Fig. 4.1a. A photograph of an upgraded station in the field can be seen in Fig. 4.1b. Although the production of SSDs was growing continuously in the different assembly sites [94], the new electronics board on its final version was not ready for production until mid of the year 2019 [84]. On March 2019, seventy-seven SSDs were deployed and connected to stations with non-upgraded electronics constituting the so-called SSD preproduction array. The locations of the preproduction array are indicated by blue circles in Fig. 4.1a. The aim of this chapter is the analysis of the data from these stations and its implications for the determination of the mass composition using the methods described in Chapter 3. Some of the results discussed in this chapter were presented in [95].

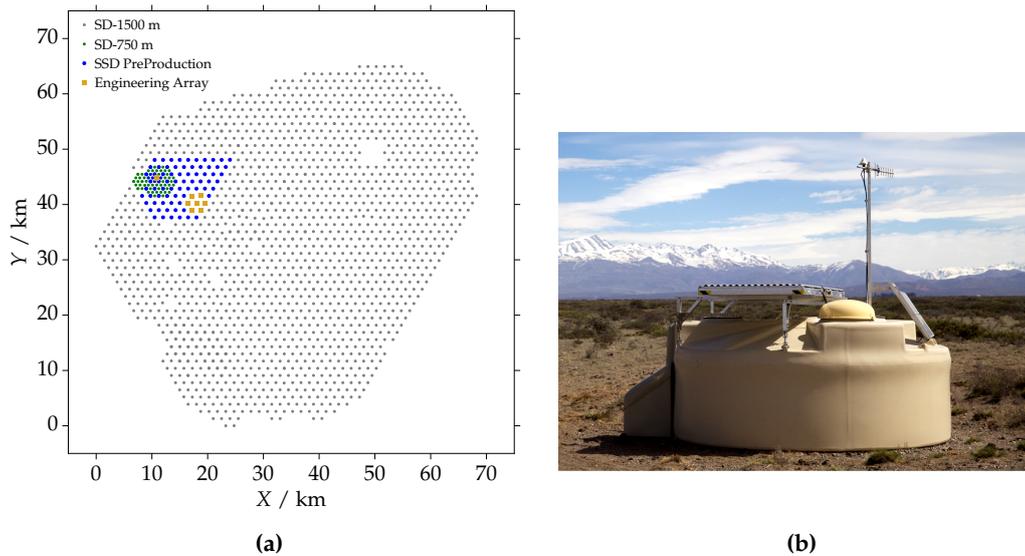


Figure 4.1: (a) Layout of the surface detector with the locations of the Engineering Array (golden squares) and of the SSD preproduction array (blue dots). (b) Photograph of a station of the AugerPrime upgrade.

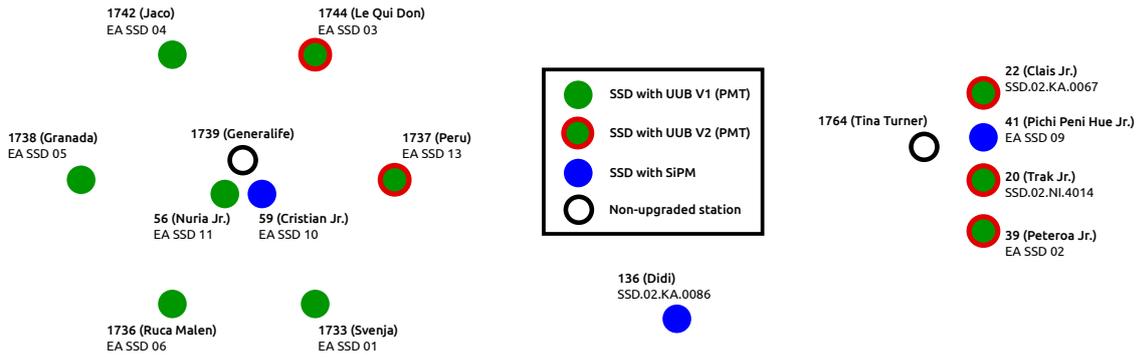


Figure 4.2: Sketch of the EA with the different configuration of the detectors by March 2019. The hexagon around station Generalife is located in the SD-1500 and the triplet next to station Tina Turner is located in the SD-750. The station Didi (136) is located at the assembly building in Malargüe.

4.1 Configurations of the detectors in the field

The Engineering Array

The goal for the Engineering Array (EA) is to test and validate the performance of a few prototype detectors in the field. This allows us to verify the hardware, the Local Station (LS) software, acquisition and monitoring, etc. The analysis of the data delivered by the first prototypes of upgraded stations is of a great importance for the production of AugerPrime detectors which will be deployed in the whole array.

Twelve stations with SSD, upgraded electronics and SPMT were deployed in the EA during September 2016. Part of these stations were located in the SD-1500 forming a hexagon around the non-upgraded station Generalife (with ID 1739). The remaining stations are located in the SD-750 near the AMIGA station Tina Turner (with ID 1764). A depiction of the EA can be seen in Fig. 4.2.

There were different configurations of the upgraded stations located in the EA since their deployment until the present date. For the SSD, two different power supplies were used to generate the required voltage at the base of the PMT. These are the CAEN [96], and the ISEG base [97]. In addition, some of the SSDs use a SiPM as optical device instead of a PMT. The data from those SSDs will not be analyzed in this thesis. The design of the UUB was also changing during the run time of the EA stations, having up to three different versions in the field [84]. More details about the UUB are given in Section 2.2.3.

The SSD preproduction array

The SSD preproduction array (PPA) was conceived to test the performance of the SSDs on a larger scale and to collect data from events at higher energies. As a final version of the electronics board was not yet available, in March 2019 seventy-seven SSDs were deployed and connected to the old electronics using an adapter box. The locations of the PPA stations are depicted by the blue dots in Fig. 4.1a.

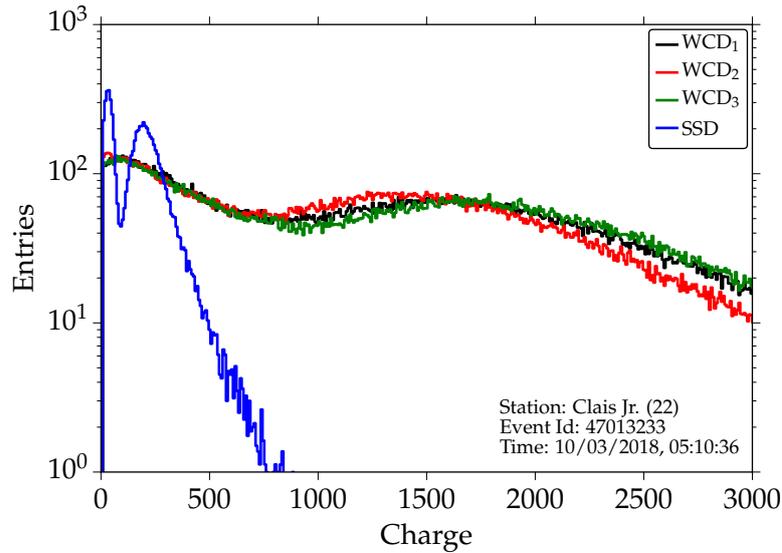


Figure 4.3: Example of a charge histogram of the SSD and WCD PMTs taken from the upgraded station Clais Jr. (22) for the event with SD ID: 47013233.

4.2 Calibration and performance of the upgraded stations

Calibration

The calibration of the signals in an upgraded station works in the same way as for the old electronics (see Section 1.2.1) but with the addition of the SSD. In Section 2.2.1, the calibration of the SSD signals was studied from the point of view of the detector response. The reference unit of the signals in the SSD is the MIP¹ due to the similar deposited energy of electrons and muons in the scintillator. Analogous to the WCD, the average charge deposited by a MIP can be seen in the second peak of the charge distribution and the value of the MIP charge, Q_{MIP} , can be inferred from the fit to that peak. An example of a charge histogram from an upgraded station in the EA is shown in Fig. 4.3. The position of the MIP peak is located at around 200 integrated FADC counts, as it can be seen in the blue curve in that figure.

It was already mentioned that the VEM is the reference unit of the WCD. The VEM charge, Q_{VEM} , is related to about 100 PE at the cathode of the PMT or around 1500 FADC counts in the upgraded electronics as it can be seen in the second peak of the distributions for the WCD-PMTs in Fig. 4.3. Due to its small area, the SPMT can not be calibrated using atmospheric muons. A cross-calibration with the WCD is performed by selecting small showers that allows us to overlap the signals of the SPMT and the WCD large PMTs [98]. Due to its smaller size, the SSD will produce a calibration trigger of about 40% of the calibration triggers of the WCD.

The values of the VEM and MIP charge vary from station to station due to differences in gain of the PMT, which might be affected by changes in the temperature or in the high-voltage, for example. Therefore, studying the time evolution of the charge might give us insights on the performance of the detectors. The violin plot in Fig. 4.4 shows the average behavior of Q_{VEM} and Q_{MIP} as obtained from a station of the EA over one year of acquisition. For each bin, the violin extends up to the 1σ standard deviations of the distribution. The mean and median of the distribution are represented by markers and

¹Although the conversion factor to vertical-equivalent MIP (VMIP) is applied, the term MIP is chosen for simplicity.

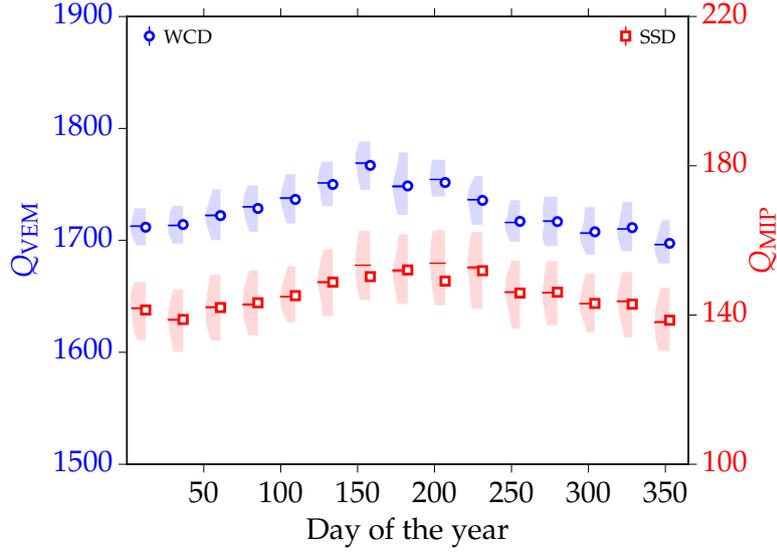


Figure 4.4: Profile of the variation with time of the VEM (blue circles) and MIP (red squares) charge. For each bin the spread of the distributions is indicated by the shaded bands.

lines, respectively. The variation of Q_{VEM} and Q_{MIP} with time suggests a variation due to seasonal modulations. At the same time, no aging effects seem to be present in the SSD in one year of acquisition.

Signals

The signal of the WCD (SSD) is obtained from the integral of the time trace normalized by the value of the VEM (MIP) charge corresponding to that specific measurement, i.e.,

$$S_{\text{wcd}} = \sum_{t=t_s}^{t_f} (s(t) - b(t)) / Q_{\text{VEM}},$$

$$S_{\text{ssd}} = \sum_{t=t_s}^{t_f} (s(t) - b(t)) / Q_{\text{MIP}},$$
(4.1)

where $s(t)$ and $b(t)$ refers to the signal and baseline in units of FADC counts at a given time interval t , and t_s and t_f denote the start and stop time of the integration window and are determined by the WCD. Examples of time traces from real events that triggered upgraded stations are shown in Fig. 4.5. In Fig. 4.5a, the trace of the SSD is shown together with the trace of one of the WCD-PMTs. The y -axis shows the signal in FADC counts after baseline subtraction and the x -axis shows the time range since the beginning of the signal in ns (lower axis) and the corresponding bin number with a sampling frequency of 120 MHz (upper axis). In Fig. 4.5b the traces of a large PMT and SPMT are shown for the case of an upgraded station where the large PMTs were all saturated. A comparison of the WCD signals between stations with old and upgraded electronics is shown as a function of the signal in Fig. 4.6. This comparison is performed using upgraded stations in the EA, which are part of a doublet with a non-upgraded station. A good agreement between the old and new electronics can be seen for a wide range of signals.

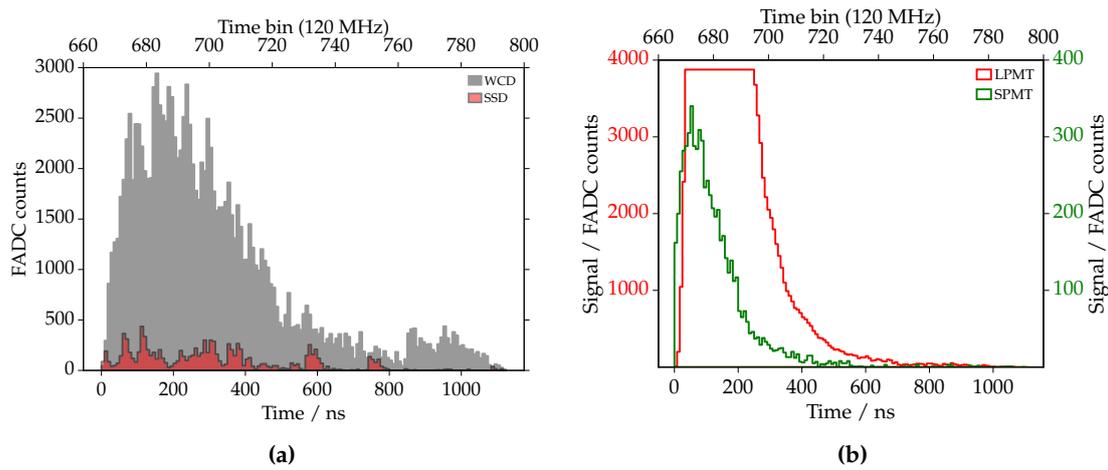


Figure 4.5: Examples of FADC traces in an upgraded station of the EA. (a) Traces of the WCD (gray) and SSD (red). (b) Traces of a saturated large PMT (red) and the unsaturated trace of the SPMT (green). That particular case corresponds to a station placed at 260 m to the shower core.

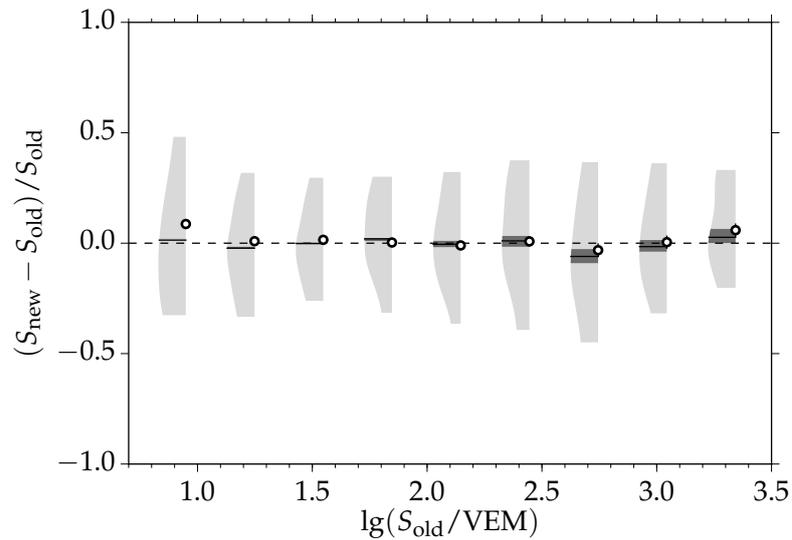


Figure 4.6: Signal comparison using measurements with doublets. A doublet is comprised by two stations, one with old electronics and the other with upgraded (new) electronics. For this analysis 19337 doublet measurements with unsaturated signals were used. The comparison is shown as a function of the signal in the station with old electronics.

Extended dynamic range

The SSD operates in slave mode to the WCD, meaning that when a shower triggers the WCD the information of the corresponding SSD is also read. The signal distribution of both detectors is shown in Fig. 4.7a. The physics trigger is visible for the WCD with a sharp cut-off at low signals while for the SSD the signals can go down to baseline fluctuations in cases where no particles passed through the scintillator. The shape of the signal distribution in the SSD can be understood as a transition between the signal distribution due to baseline integration (no particle passing through) and/or small energy deposits in the scintillator (e.g., photons, corner-clipping particles, etc), and eventually the signal distribution produced by one or many particles traversing the detector. With the addition

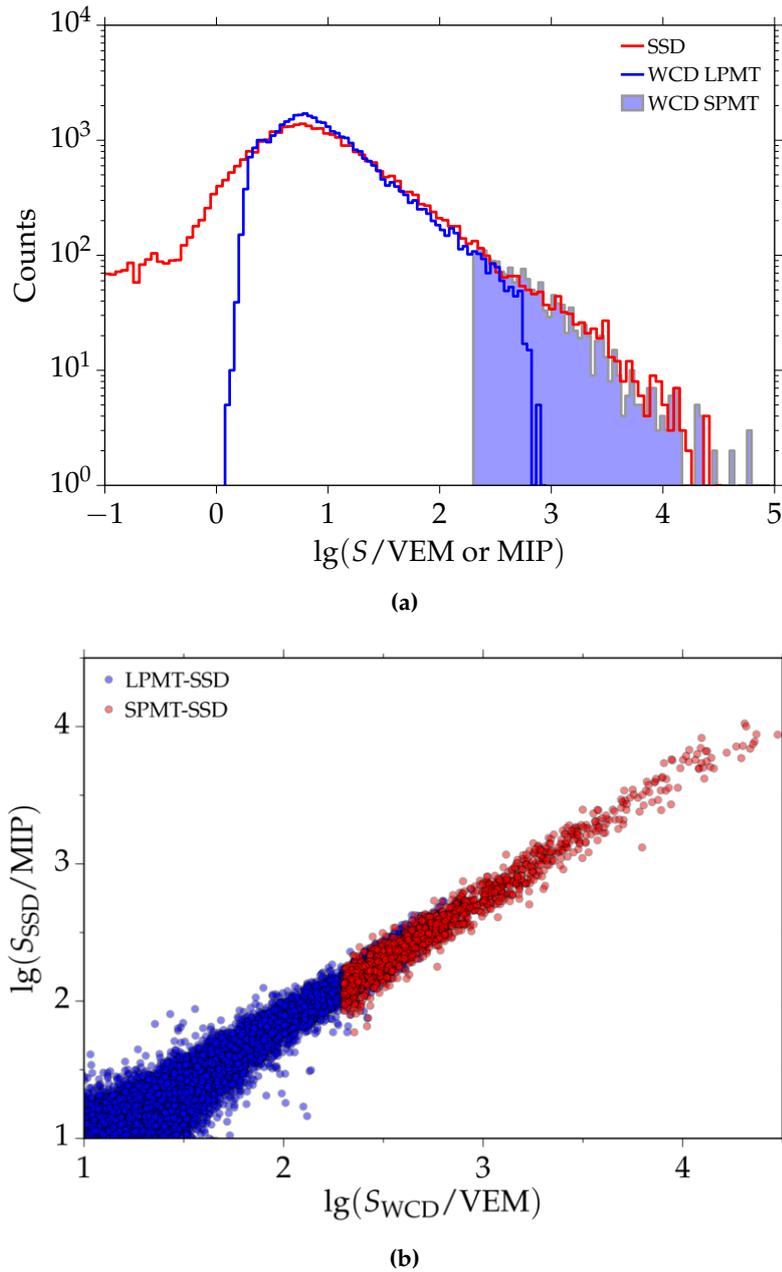


Figure 4.7: (a) Distribution of signals of the SSD (red) and of the WCD (blue) The blue line shows the distribution of signals using the large PMTs whereas the filled histogram shows the distribution of signals using the SPMT. (b) Correlation between signals of the SSD and of the WCD using data of the WCD large PMTs (blue) and of the SPMT (red).

of the SPMT, the non saturated range of the WCD extends up to ≈ 20000 VEM, as seen in Fig. 4.7b. The correlation between the signals measured by the SSD and the SPMT is also shown there.

4.3 Update of the Offline framework

The Offline software framework [81] has been used by the Auger collaboration for simulating and reconstructing events from air showers for over a decade. Major updates to the *detector description* and the *event data* were performed in order to place the information of

Table 4.1: Quantities used for the calibration of the WCD and the SSD signals. The values of the UUB quantities correspond to average values obtained from analysis of events recorded by the EA stations with IDs 20 and 22 in August 2019. Values from the UB are also shown.

Quantity / ADC	WCD (UB)	WCD (UUB)	SSD (UUB)
Charge	200	1425	145
Peak	50	180	63
HG baseline	50	250	257
σ (HG baseline)	0.5	2.0	2.0
LG baseline	25	262	251
σ (LG baseline)	0.5	0.6	0.7

AugerPrime detectors [90]. However, further updates were needed for a proper handle of the detectors' data upon its collection, allowing for the reconstruction of real events including both the standard and the upgraded stations. These changes are described below:

The first part of the software which had to be modified was the *converter* which translates the information of the event (firstly stored within the CDAS structures) into data readable by the *Offline*. The main changes are related to the readout of calibration information and of the PMT traces of the upgraded stations, as the CDAS stores this information in different places as for the non-upgraded stations.

Once the *converter* is updated, we are able to read information of the upgraded stations such as the charge histograms and the FADC traces from all PMTs, which in principle allows us to get the calibrated signal of the WCD and the SSD. However, there is still information missing which is required by the *modules* in the reconstruction chain. Examples are the value of the high-gain to low-gain ratio, estimates of the VEM and MIP charge, status of the PMTs... This information is given by the LS software in the non-upgraded stations and is used to check the performance conditions of the station before the off-line calibration procedure takes place. This situation is solved by means of an additional *module* that fills-in estimates of the missing information making it possible to proceed with the calibration process. More than 95% of the processed signals were successfully calibrated. The values of some of the quantities used for the calibration of the UUB signals are summarized in Table 4.1.

The last update needed occurs in the *module* for the calibration of the signal. Many of the algorithms used there must be accommodated for the settings of the UUB as they were designed for calibrating signals sampled at 40 MHz and with 10 bits of maximum amplitude. In addition, a *module* that fits the lateral distribution of the SSD was implemented.

The aforementioned updates enabled the reconstruction of events using detectors of the upgrade in the EA and the PPA in the *Offline* software framework. A module sequence for such a event reconstruction with AugerPrime detectors is given in Appendix B.2.

4.4 First measured showers

Since the first SSD units were deployed in the EA, more than 28000 showers have been measured including both upgraded and standard stations in the event reconstruction. Due to its higher trigger rate, the majority of these showers were measured by EA stations at the SD-750 and with energies $\gtrsim 10^{17.5}$ eV. Fewer showers (but more energetic) have

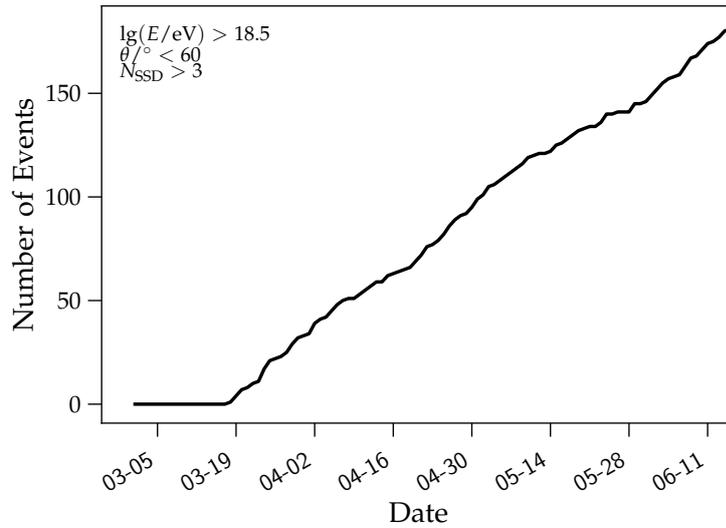


Figure 4.8: Rate of events above full efficiency ($E > 10^{18.5}$ eV) measured with the SSDs at the PPA. Only events with at least 3 SSDs and $\theta < 60^\circ$ were considered.

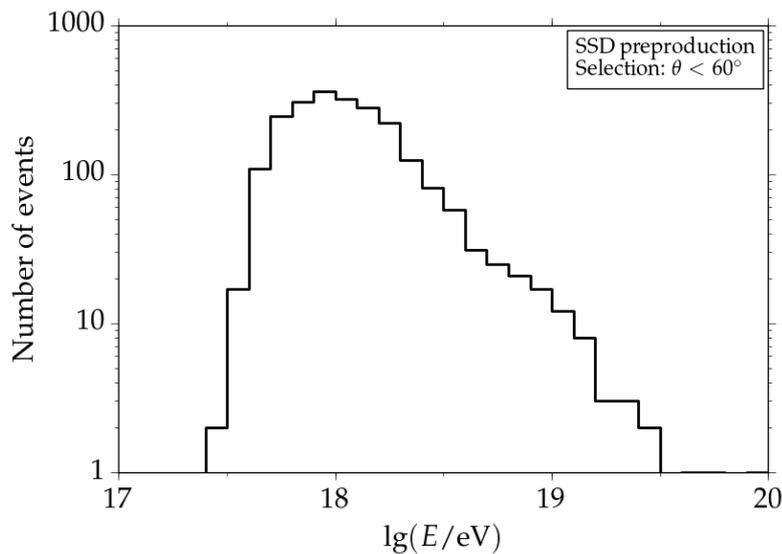


Figure 4.9: Histogram of the reconstructed energy of events measured with stations of the PPA. Events with zenith angle below 60° were selected for a better reconstructed energy.

also triggered the upgraded stations located at the SD-1500. The additional SSD units in the PPA increased the amount of events rapidly. The rate of events measured with stations of the PPA is shown in Fig. 4.8. In three months of acquisition more than 150 showers with energies above 3 EeV were measured. The distribution of the reconstructed energy of events measured with stations of the PPA can be seen in Fig. 4.9.

4.4.1 Station level evaluation

The idea of using a scintillator detector in order to enhance the sensitivity of the surface detector to the mass of UHECR relies on the differences in the SSD and WCD responses to shower components. While electromagnetic particles have smaller energy deposits

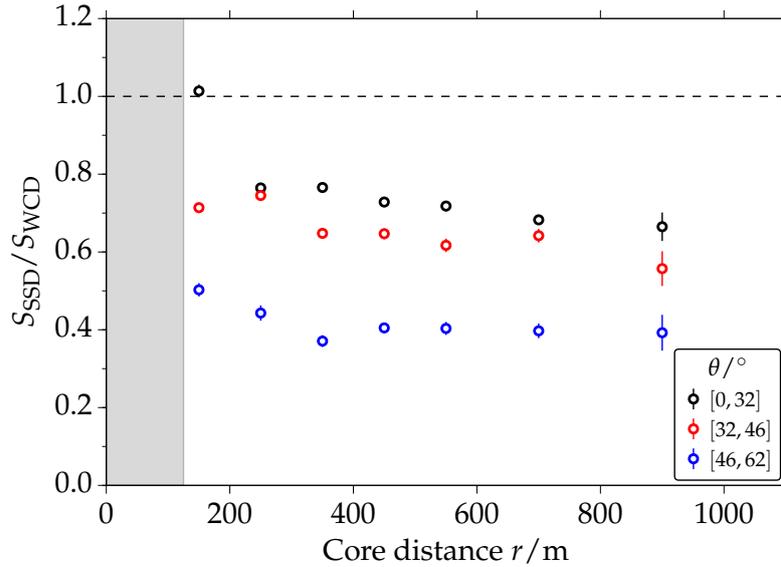


Figure 4.10: Ratio of SSD and WCD signals as a function of the core distance, for different zenith angles of incidence. The ranges in zenith are chosen according to equal bins in $\sin^2 \theta$. The gray area shows the region of saturated signals.

than muons in water, both components deposit on average the same amount of energy in the scintillator. One can see this in terms of the produced signal: an electron will produce, on average, a signal of 1 MIP in the SSD, and the same holds for a muon. In the WCD, however, an electron will produce about 1/20 of a signal produced by a muon, i.e., 1/20 VEM. In this sense, the SSD is more sensitive to the electromagnetic component of the shower.

The ratio of SSD signals to WCD signals is shown in Fig. 4.10 as a function of the distance to the shower core. The trend changes with the zenith angle. For vertical showers, the ratio S_{SSD}/S_{WCD} is closer to 1 at distances closer to the shower core and decreases with increasing distance. For inclined showers, the ratio becomes nearly constant for distances $r > 300$ m.

In Fig. 4.11, the same ratio is shown as a function of the reconstructed energy, using data from the EA stations located in the SD-750 at energies below $10^{18.1}$ eV and data from stations of the PPA above that energy. For this plot, signals from stations located at distances in the range $350 \text{ m} \leq r \leq 550 \text{ m}$ were used. The sensitivity of the SSD to the electromagnetic component, which has a faster increase with energy than muons, is shown by the up-going trend in this figure. The drop of the ratio S_{SSD}/S_{WCD} with zenith is very similar for all energies. The dip observed at around $10^{18.3}$ eV for the largest zeniths bin might be an statistical fluctuation as the detector is not fully efficient at those energies.

As showers elongate with energy of the primary, at higher energies the depth of shower maximum (X_{\max}) occurs at distances closer to the detector plane. Therefore a higher contribution of the electromagnetic component is expected, increasing the signal in the SSD with respect to the WCD. Of course the ratio must depend on the primary mass. Since lighter showers penetrate deeper in the atmosphere than heavier ones, the number of electrons at ground is expected to be larger for showers induced by light primaries. Correlations with X_{\max} measurements should be performed to clarify the features presented here.

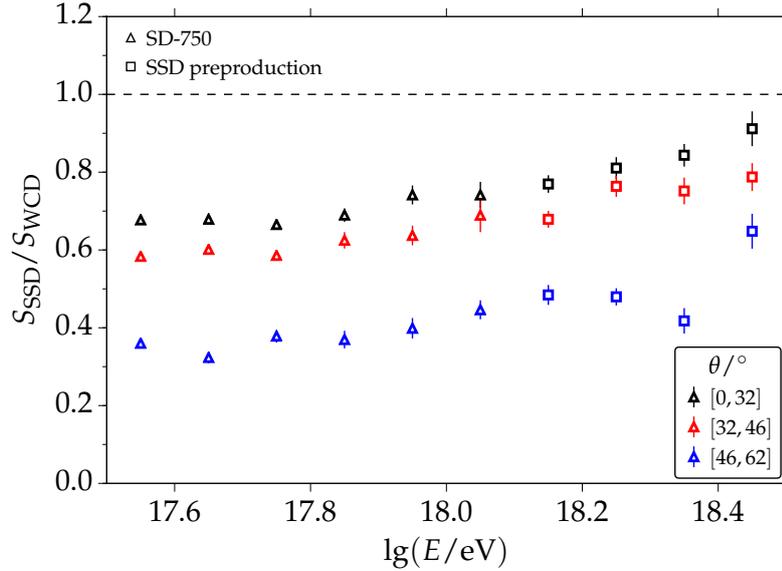


Figure 4.11: Ratio of signals as a function of the reconstructed energy. Data from EA stations at the SD-750 (triangles) and from the PPA (squares) are used. Same zenith ranges as in Fig. 4.10 were used.

The profile of the average LDF for the WCD and SSD signals can be seen in Fig. 4.12. Each signal was normalized by the corresponding shower size to eliminate any degeneracy due to showers with different energies. For this plot, 2468 events with zenith angles in the range $0^\circ \leq \theta \leq 32^\circ$ and with energies $E \geq 10^{17.5}$ eV were used. The average is performed for distances within 1200 m in order to avoid biases due to trigger effects on the WCD (a discussion can be found in chapter 5 of [99]).

As expected, the SSD has a steeper lateral distribution of signals than the WCD, with slightly larger signals close to the shower core and a faster fall off at large distances. This is due to the differing sensitivity of the two sub-detectors to the shower components and to the attenuation of the electromagnetic particles in the atmosphere which increases with the radial distance.

4.4.2 Event level evaluation

The results shown in the previous section may help us to understand the differences in the detector responses to shower particles by looking at how measured signals behave with the geometry or the energy of the shower. The next step is to analyze quantities which are inferred from the event reconstruction, such as the signal at the optimal distance S_{1000} , measured with both detectors.

An example event reconstructed using the WCD and the SSD is shown in Fig. 4.13. The shower arrived to the SSD preproduction array with a zenith angle of about 37° and energy of about 48 EeV. The core is located to the north-eastern side of the SD-750 as it can be seen in Fig. 4.13a. The WCD and SSD LDFs are shown in Fig. 4.13b. The energy and the arrival direction are estimated using the fit of the WCD LDF. Then this information is used for the fit of the SSD LDF. First, the zenith angle and an estimate of S_{1000}^{ssd} are used to fix the shape of the LDF using the parameterization of the slope parameters (see Section 3.2.3), then the fit is performed leaving S_{1000}^{ssd} as a free parameter. It can be seen that the SSD has a steeper LDF as well as larger fluctuations in the signals. The contribution of the different components to the total signal is encoded in the time

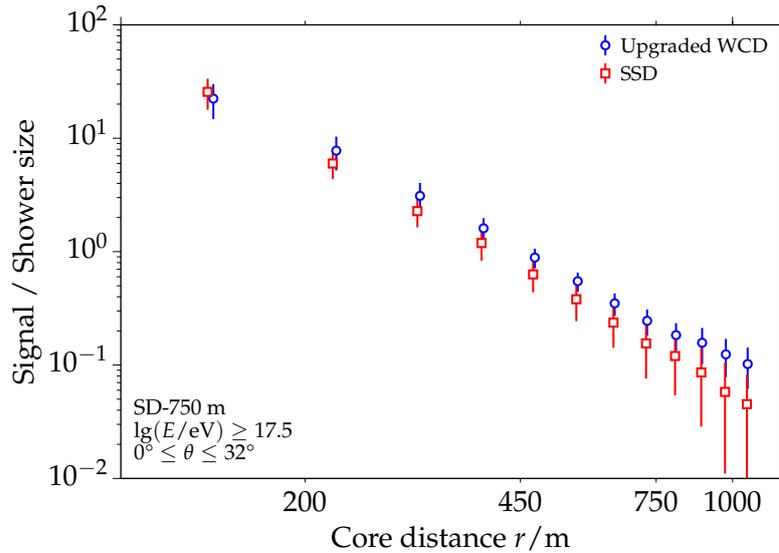


Figure 4.12: Average lateral distribution of signals of the WCD (blue circles) and the SSD (red squares). From the reconstruction of events of the SD-750, the shower size (signal at 450 m) can be obtained and used as normalization. Showers with energies $\geq 10^{17.5}$ eV and zeniths $\theta \in [0^\circ, 32^\circ]$ were used.

traces. The time structure is affected by the distance particles travel from their point of origin in the shower to the detector. In addition, both detectors have different time responses, for example, the signal in the WCD has a relatively larger spread due to the multiple reflections of Cherenkov light in the tank while in the SSD the time spread can be indicative of particles that were produced later in the shower development. Example of calibrated traces for the WCD and SSD are shown in Fig. 4.14, for stations located at different distances from the shower core. The features of the air showers may be understood by looking at the different time development of the signal in both detectors. At a closer distance to the core (as seen in Fig. 4.14a), the WCD and the SSD seem to have a similar time structure, as they sample the bulk of the electromagnetic part. However, at a larger distance (as shown in Fig. 4.14b) the differences become more apparent. At those distances one expects to measure mostly muons. The signal in the WCD has a larger spread which could be related to the sampling of some electromagnetic particles which are not measured by the SSD, while the signal in the SSD seems to be related to individual muons (higher spikes). Although it is difficult to conclude from the analysis of the individual events, the analysis of the time traces from both detectors might aid in disentangling the shower components.

The event shown in Fig. 4.13 was measured by 12 SSDs which allows for more flexibility of the fit of the LDF. The number of SSDs that have valuable information (i.e., a signal produced by real MIPs and not by baseline fluctuations) is a very important quantity for the event reconstruction, as it will have a direct impact on the quality of the fit of the LDF. The multiplicity of SSDs is shown as a function of the energy in Fig. 4.15. Only SSDs with signal above 1 MIP were considered in order to reject those that were not “triggered” by shower particles. Furthermore, a selection of events which were fully contained within the SSD PPA was made in order to eliminate border effects. The linear relation is expected. There are several events with only one SSD but those are at low energies (below full efficiency) as these showers are not energetic enough to produce at least one MIP in several stations. Above $10^{18.5}$ eV most of the events have at least 3 SSDs

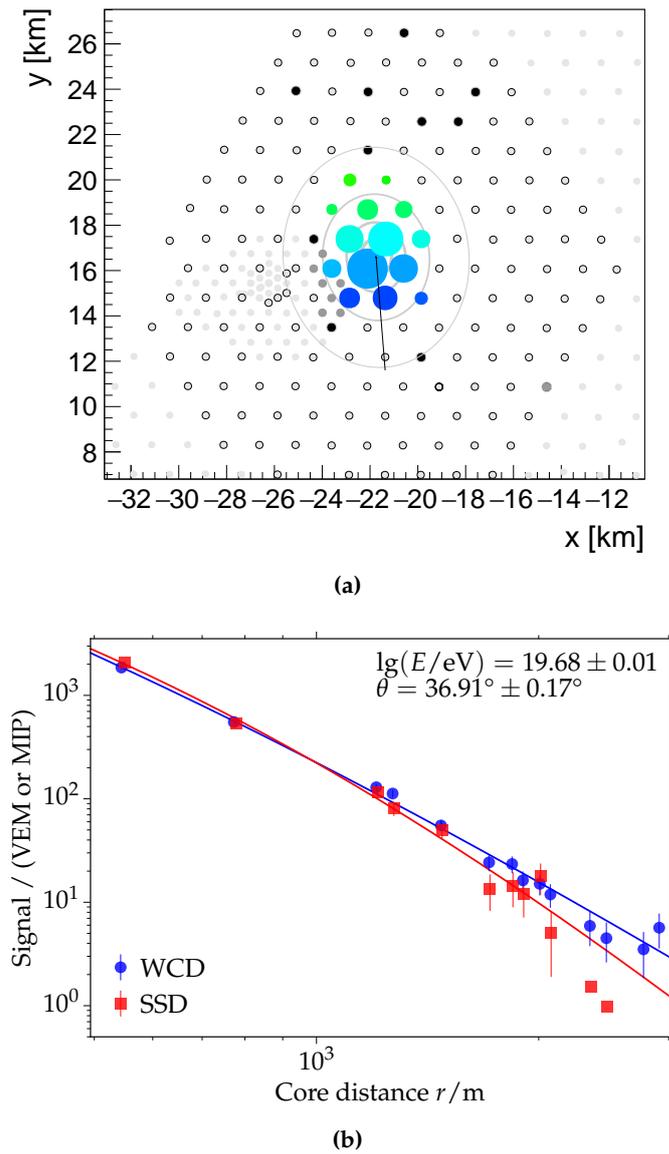


Figure 4.13: Reconstruction of an event with AugerPrime detectors. The event was reconstructed with an energy of 4.84×10^{19} eV and a zenith angle of $\theta = 36.9^\circ$ (Event time: 24. June 2019, Auger ID: 191753530400). (a) Footprint of the shower at ground. The arrival direction of the shower is indicated by the thin line. The color indicate the time information (green for early times and blue for late ones) and the size of the markers is proportional to the signal measured by each WCD (figure taken from Offline EventBrowser). (b) Fit of the lateral distribution function. Markers show the signals of the WCD (blue dots) and of the SSD (red squares) as a function of the distance to the shower core. The lines show the fit to the corresponding LDFs.

with signal above 1 MIP. Although there is not enough statistics to read of the number directly from the data, one can expect more than 10 SSDs at energies around $10^{19.5}$ eV.

The potential of the SPMT can be seen in Fig. 4.16. In this particular event, two upgraded stations from the EA have the WCD's large PMTs saturated. The LDF is poorly fitted with the stations close to the core, as seen in Fig. 4.16a. The same event was reconstructed using the signal of the SPMT improving the fit of the LDF as it is shown in Fig. 4.16b. The SPMT allows us to extend the unsaturated measure of the signal very

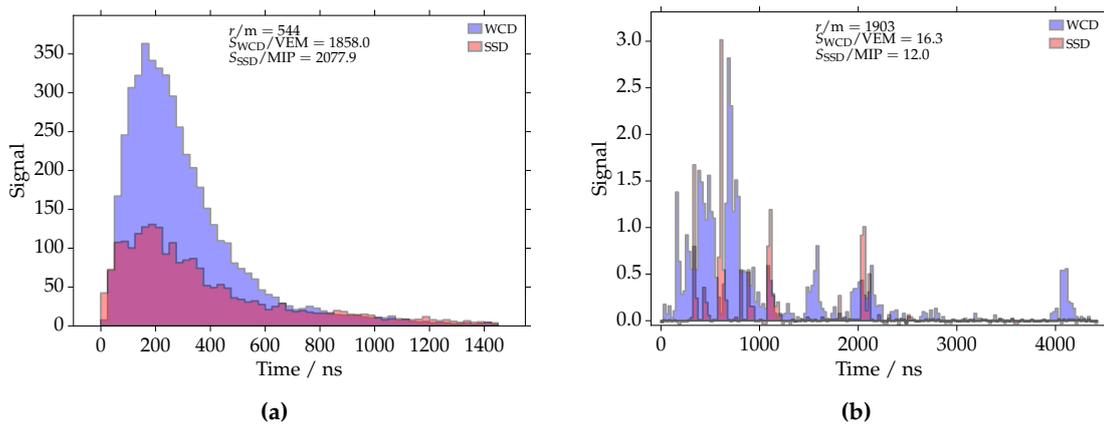


Figure 4.14: Calibrated traces of the WCD (blue) and of the SSD (red) from two stations of the event shown in Fig. 4.13. (a) Station located at 544 m to the shower core. (b) Station located at 1903 m to the shower core.

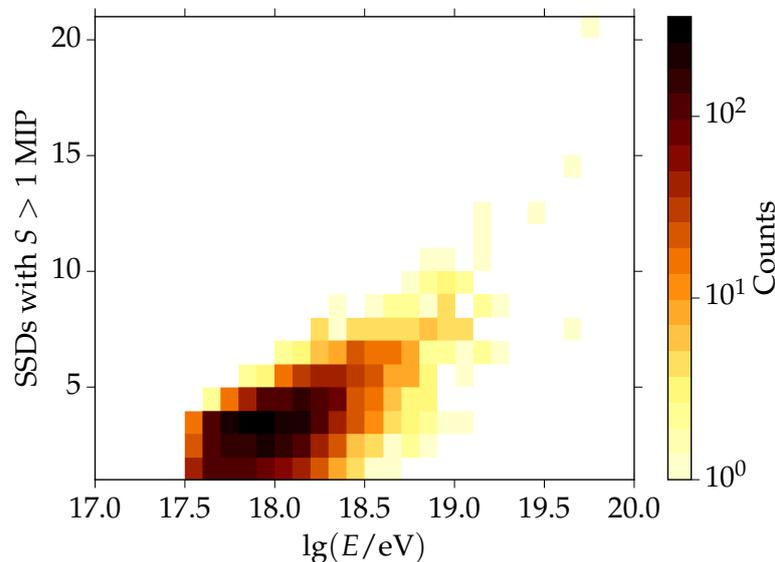


Figure 4.15: Number of SSDs with signal above 1 MIP as a function of the energy.

close to the core of the shower. It can be seen how signals of the saturated stations can be corrected from ≈ 3000 VEM to ≈ 7000 VEM. The saturated and unsaturated traces of one of these stations are shown in Fig. 4.5b.

The fit of the lateral distribution function of the SSD signals is performed using a parameterization derived from Monte Carlo studies with proton and iron showers (see Chapter 3). In Section 3.2.3, it was shown that the obtained parameterization works well for simulated showers. In order to see how well our simulations-based model describes the data from real measurements, the residuals of the SSD signals, i.e., the difference between the measured signal S_i and the expected signal from the fit of the LDF \hat{S}_i , were computed. The residuals can be seen in Fig. 4.17. For this analysis, all reconstructed events using stations of the PPA for energies above 10^{18} eV and zenith angle $0^\circ \leq \theta \leq 56^\circ$ were included. It can be seen that, on average, the LDF gives a good description of the data for a wide range of distances (Fig. 4.17a) and SSD signals (Fig. 4.17b). Their fluctuations at around 900 m and 1.2 in $\lg S$ can be due to the mixture of different energies and zeniths, as the statistics is limited to perform a detailed analysis. There is also a good agreement

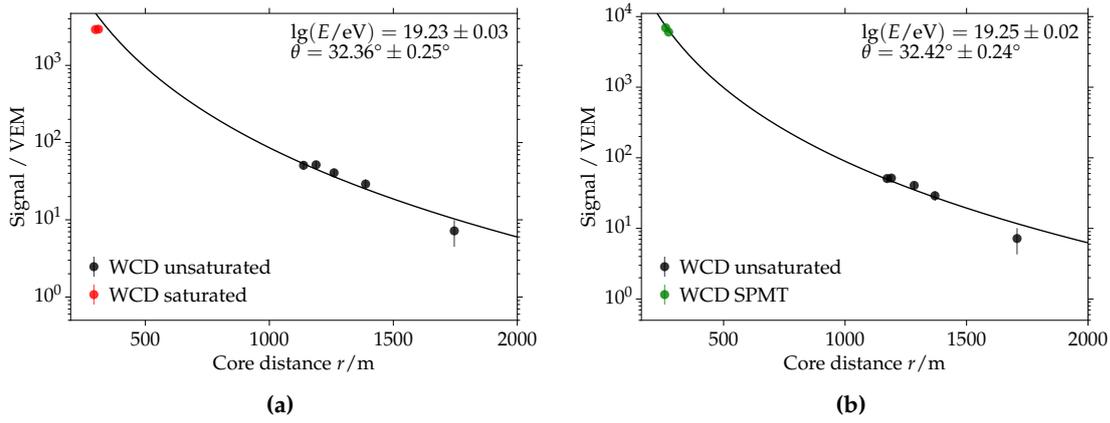


Figure 4.16: Reconstruction of the event with Sd ID: 46663172. The two stations closest to the shower core are upgraded stations of the EA. (a) Fit of the LDF using only the large PMTs (b) Fit of the same LDF but using the SPMT signals (see text for the discussion).

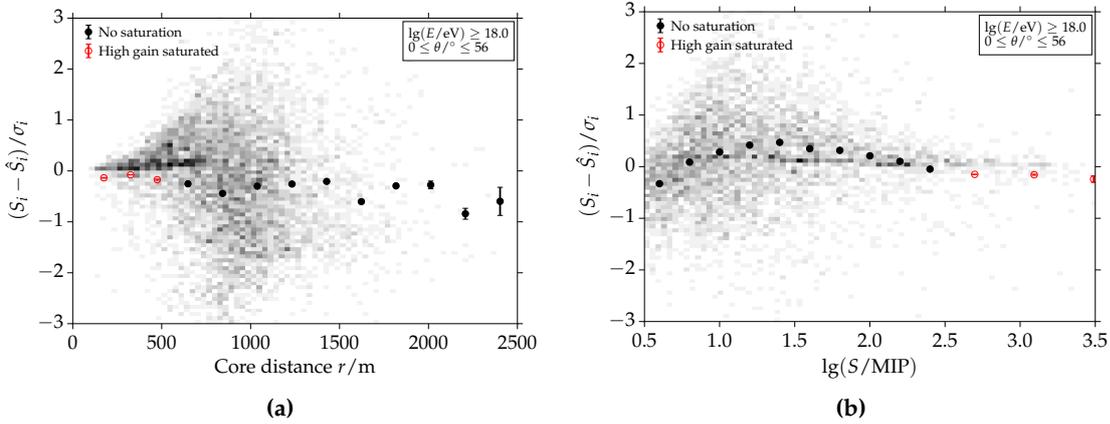


Figure 4.17: Residuals of the SSD LDF using data of stations from the PPA. (a) Residuals as a function of the distance to the shower core. (b) Residuals as a function of the logarithm of the signal. The dots show the profile of the distributions (shown by the histogram in the background). Data from non-saturated SSDs is shown with black dots and data from high-gain saturated SSDs is shown with red open dots. Reconstructed showers with energy above 10^{18} eV and zenith angle $0^\circ \leq \theta \leq 56^\circ$ were used.

between data from high-gain saturated and non-saturated SSDs. Information about the high-gain saturated SSDs can be found in Fig. 4.18. The distribution of distances to the core for which an SSD was high-gain saturated is shown in Fig. 4.18a. It can be seen that most of the SSDs saturate the high-gain channel at around 300 m. A few stations seem to have a high-gain saturated SSD above 1000 m, as it can be read from that distribution. This is an unlikely situation (due to the dynamic range of the SSD PMT). In fact, these particular cases correspond to showers with energies of about 3×10^{18} eV and signals in the SSDs between 26 MIP and 56 MIP. Therefore, they must be related to hardware issues.

Given that the parameterization of the SSD lateral distribution function gives a good description of the real data, we proceed to analyze quantities which can be extracted on an event basis, such as the shower size. In Fig. 4.19, the correlation between S_{1000} measured by the SSD and the WCD can be seen. For this plot, a selection of 1794 showers

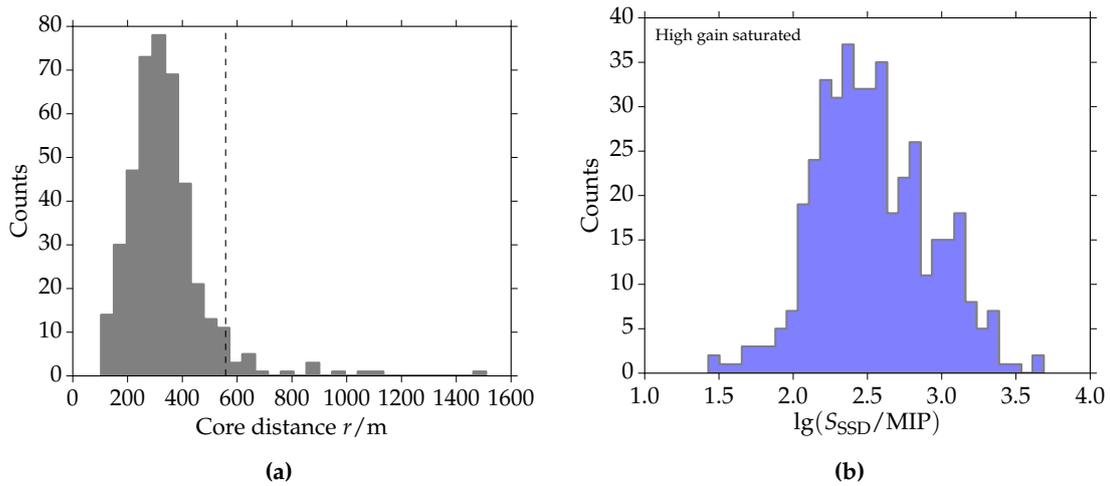


Figure 4.18: (a) Distribution of distance to the shower core for SSDs with high-gain saturation. The dashed line shows the 95% quantile (corresponding to about 550 m). (b) Distribution of signal of SSDs with high-gain saturation.

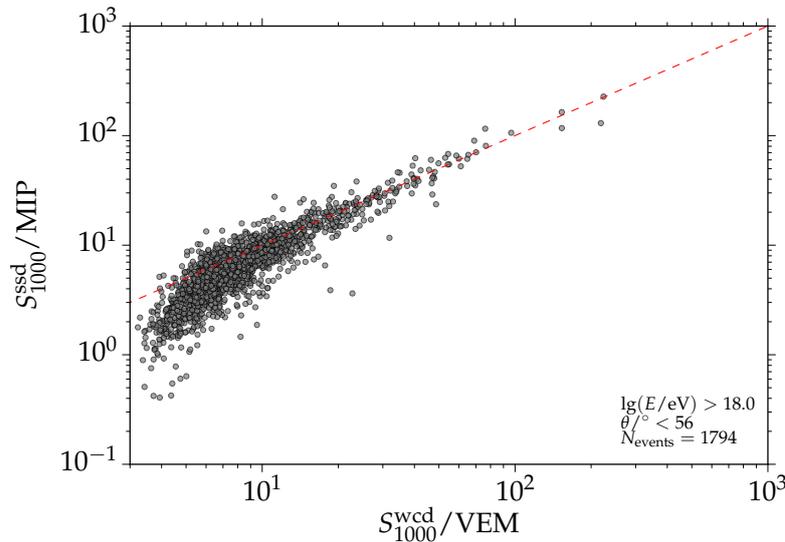


Figure 4.19: Correlation of S_{1000}^{ssd} and S_{1000}^{wcd} from events measured with the SSD preproduction array. 1794 showers of energy above 10^{18} eV and zenith below 56° were used.

reconstructed with stations from the PPA was made. These showers had energies above 10^{18} eV and zenith angles below 56° . A good correlation between both measurements is observed. The tail towards low values of S_{1000}^{ssd} due to the efficiency of the reconstruction (most of those events correspond to low energy showers).

The same information can be used to compute the ratio $S_{1000}^{ssd}/S_{1000}^{wcd}$ as a function of the energy. This is shown in Fig. 4.20. The scatter is in part due to the mixing of the zenith angles and also due to the mis-reconstructed events at low energies. However, the profile shows an up-going trend of the ratio with energy. Again, the differing in the detector responses might be the cause of this behavior.

This ratio, however, must depend on the primary mass and therefore could be indicative for mass composition. From the measurements of $\langle X_{max} \rangle$ [31], a transition to a lighter composition at energies around $10^{18.7}$ eV is expected. Unfortunately, the limited statistics

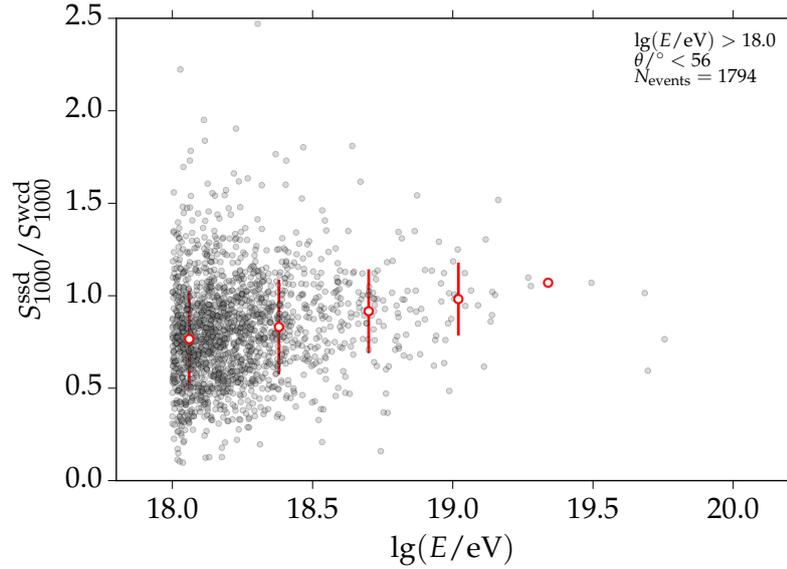


Figure 4.20: Ratio of S_{1000}^{ssd} and S_{1000}^{wcd} as a function of the energy. Same events as in Fig. 4.19 were used. The red circles show the profile of the data in the energy range $\lg(E/eV) \in [18.1, 19.5]$.

does not allow us to perform analysis of hybrid events with SSD data. With data from the FD, one should be able to look for correlations with X_{max} , as well as with the muon content of the shower, N_μ . The latter is measured at low energies ($\sim 10^{17.5}$ eV to 10^{18} eV) by AMIGA (see, for example, [36, 55]) which could be then used to correlate with the muon estimator obtained from WCD and SSD data at higher energies. Such estimator could be extracted on a single event level using the ratio $S_{1000}^{ssd}/S_{1000}^{wcd}$ or more sophisticated methods like the Matrix Formalism [90] and eventually Universality with the SSD. The crucial part is the handling of the systematic errors in our reconstruction methods in order to increase the resolution in the estimator of N_μ .

4.5 First estimate of mass composition with AugerPrime

The main goal for the AugerPrime upgrade is the estimation of the mass of UHECR. In Chapter 3, studies based on Monte Carlo simulations were performed in order to show the potential for discriminating the mass of the primary using information from the SSD reconstruction. In this section, we address the task of mass composition using data from real events measured with AugerPrime detectors.

4.5.1 Selection cuts

During the time period, starting on March 17, 2019 and ending on August 07, 2019, more than 4500 showers with energies above $10^{17.5}$ eV and zeniths below 60° were measured with with stations of the PPA (see Fig. 4.21). These events have a 6T5 trigger selection (see Section 1.3.1) and at least the WCD LDF was reconstructed (see Section 1.3.4). Another condition is that the hottest station² has an SSD. Thus ensuring that most of the events are contained in the PPA improving the fit of the SSD LDF. Only events with the hottest station on the edge of the array would be partially contained. The reconstructed core

²The hottest station is the one with highest WCD signal in the event.

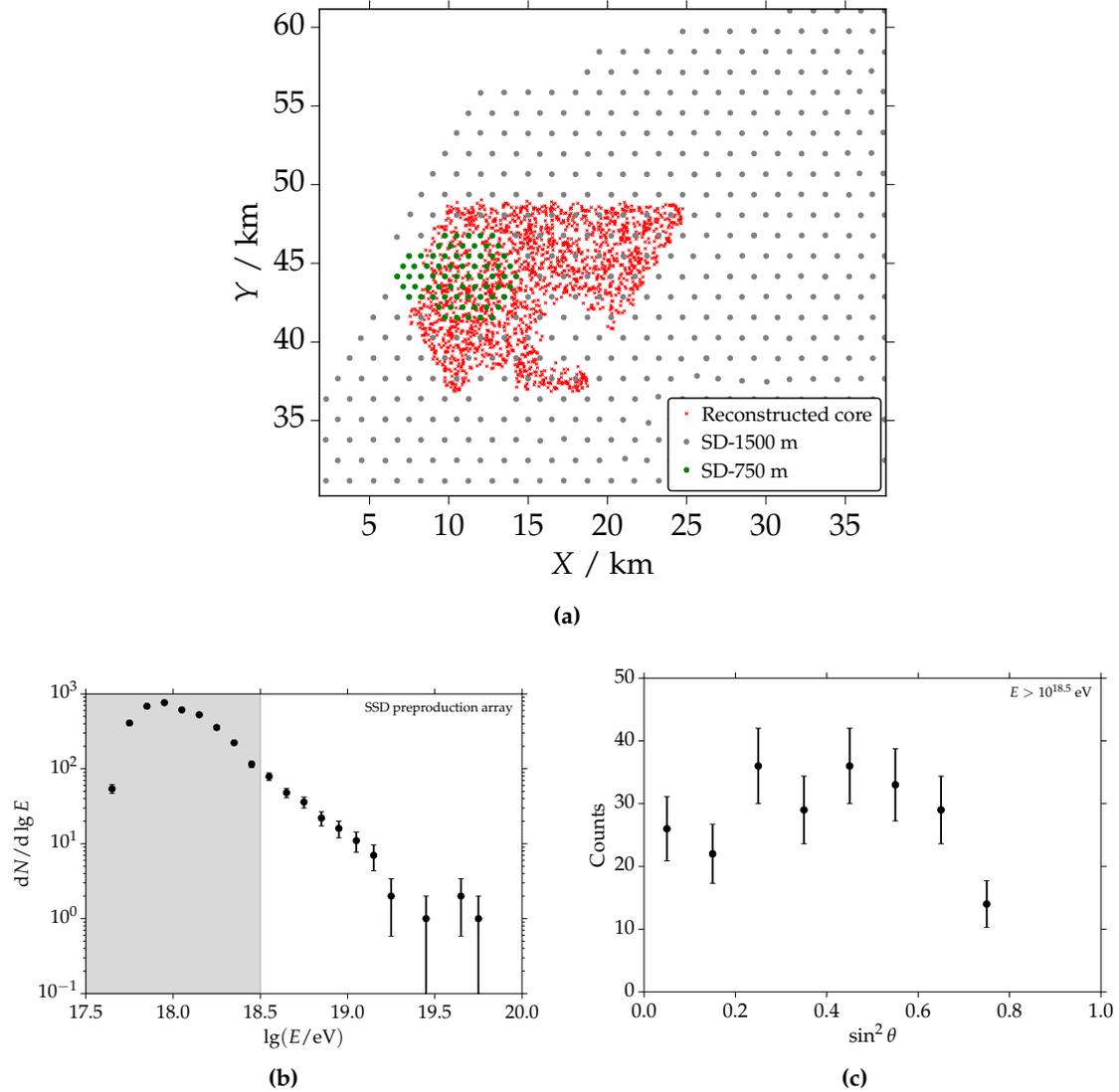


Figure 4.21: (a) Zoom of the array in the area of the PPA. The red crosses show the position of the reconstructed core. (b) Number of events per bin in $\lg E$ as a function of the energy. The shaded area shows the region below full efficiency. (c) Distribution of $\sin^2 \theta$ of reconstructed showers with energies above $10^{18.5} \text{ eV}$. See text for details.

Table 4.2: Number of reconstructed events with the PPA and quality cuts.

Selection cut	Number of events	Relative number (in %)
Total events (6T5 & LDF)	4516	—
SSD Rec. ($\epsilon_{\text{SSD}} > 95\%$)	209	4.6
$N_{\text{SSD}} \geq 3$	194	4.3

positions can be seen in Fig. 4.21a. The rate of events in terms of the logarithm of the energy can be seen in Fig. 4.21b. The drop in the number of events below 10^{18} eV is due to the trigger efficiency of the SD. In addition, the distribution in bins of $\sin^2 \theta$ is shown in Fig. 4.21c. It is compatible with a flat distribution (within the uncertainties), the drop in the number of events on the last bin is also due to trigger efficiency.

In order to improve the quality of our dataset for future analyses, a selection in energies above $10^{18.5} \text{ eV}$ and zeniths below 56° was performed based on the efficiency of the

SSD reconstruction (which was studied in Section 3.2.4). As the shape of the SSD LDF is fixed to the parameterizations (see Section 3.2.3), in principle only one SSD is needed for the fit of the LDF and therefore obtain an estimate of S_{1000}^{SSD} . An additional requirement that at least 3 SSDs are used for the fit of the LDF was imposed, thus increasing the quality of the fits. The number of events after the cuts are applied is summarized in Table 4.2. 194 events passed the selection cuts and will be used for the rest of the analysis presented in this chapter.

4.5.2 Mean composition

In Section 3.4.2, the potential for discriminating the mass of the primary was studied using pairs of SSD and WCD measurements of the shower. This evaluation was performed with the Principal Component Analysis using the reconstructed signal at the optimal distance by the two sub-detectors and the energy from simulated proton and iron showers. Here, we will apply this method to the selected data measured with stations of the PPA (see Section 4.5.1).

The average value of $\ln A'_{\text{pca}}$ as obtained using data from the PPA is shown as a function of the energy in Fig. 4.22. The data is distributed in energy bins of $\Delta \lg(E/\text{eV}) = 0.1$ extending from $10^{18.5}$ eV to $10^{18.9}$ eV. The last bin contains all data with energy above $10^{18.9}$ eV. In each energy bin, the average of $\ln A'_{\text{pca}}$ and its error were calculated using the median and the median absolute deviation, respectively. This distributions of $\ln A'_{\text{pca}}$ in each energy bin are shown in Appendix C.3.1. The position of data point on the energy axis is determined using the mean of $\lg E$ in each bin, which differs slightly from the bin center due to the fall of the spectrum. Systematic uncertainties are indicated by the square brackets (see Section 4.5.4). The lines show the predicted value for proton and iron showers using QGSJET-II.04 as hadronic interaction model (see Section 3.4.2). The uncertainties on the Monte Carlo predictions are not shown for simplicity.

The scatter of the data can be seen in Fig. 4.23a. Each dot corresponds to a single event measured with stations from the PPA above $10^{18.5}$ eV and with zenith below 56° . The gray arrows represent events for which the value of $\ln A'_{\text{pca}}$ lies out of the axis limits. The full distribution of the data is shown in Fig. 4.23b. The predicted values from Monte Carlo simulations of proton and iron are shown as a reference. It can be seen that the mean of the distribution of the data is close to the iron line. However the width of the distribution is too large. This means that the algorithm in its current state, has a poor resolution for reconstructing the mass of the primary on a single-event level. However, the average composition can be estimated. Using other type of information (such as the time information) and more comprehensive reconstruction methods, the resolution is expected to improve significantly.

4.5.3 Results from previous measurements

It is not the aim of this section to compare the result obtained in Section 4.5.2 with those obtained by previous analyses. The result presented in this work must be considered as a first estimate of the mass composition using the SSD and might serve for future analysis as reference. Below, a brief description of analyses that were performed before in Auger is given.

An estimation of the mass composition of cosmic rays at energies above $10^{17.8}$ eV was obtained by studying the energy dependence of the mean and standard deviation of the X_{max} distributions and comparing them to air shower simulations for different

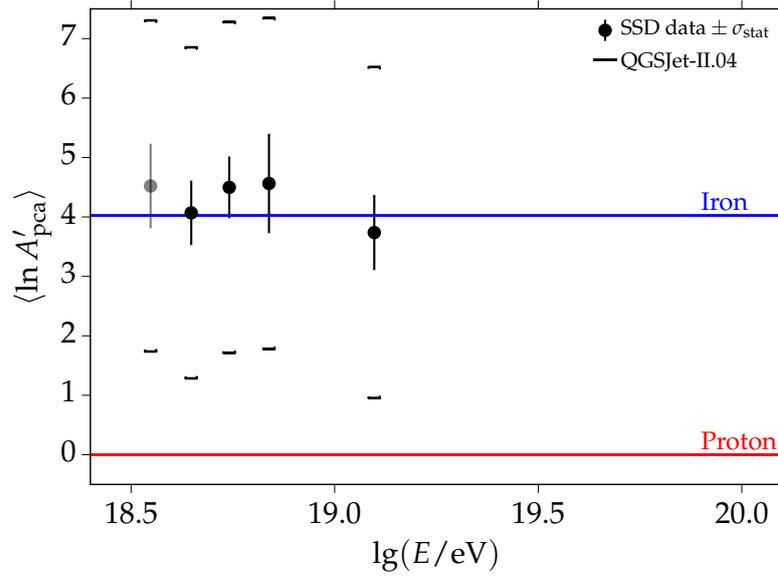


Figure 4.22: Estimate of the logarithmic mass as a function of the energy using data of stations of the PPA. Markers show the median of $\ln A'_{pca}$ in each energy bin. Error bars show the statistical uncertainty on each bin. The first data point is displayed with a transparent marker to indicate that events in that bin may be affected by the reconstruction efficiency. Systematic uncertainties are indicated by square brackets. For comparison, the results are shown for simulations with QGSJET-II.04 hadronic interaction model for proton (red) and iron (blue) primaries.

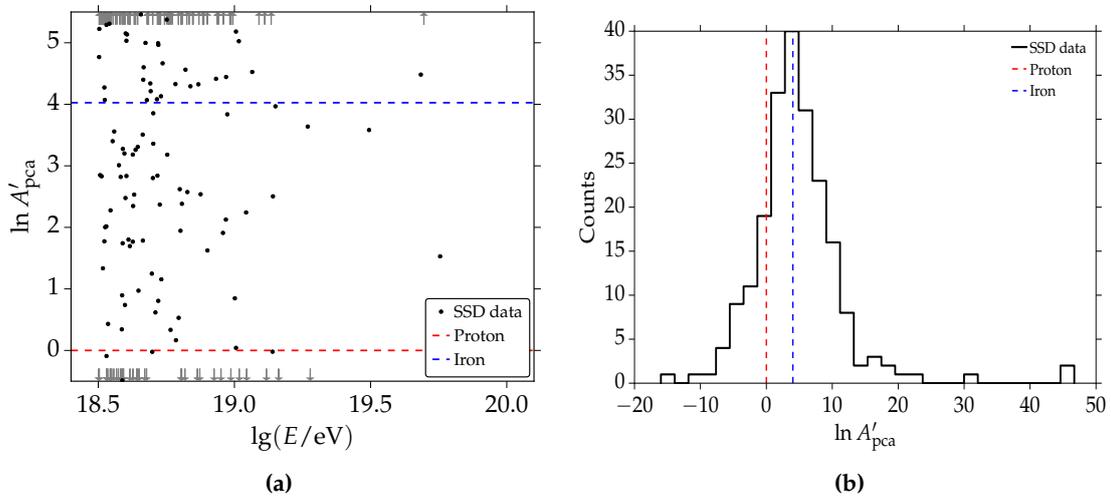


Figure 4.23: (a) Scatter of $\ln A'_{pca}$ (used in Fig. 4.22) as a function of energy. The arrows indicate the events for which the value of $\ln A'_{pca}$ is out of the limits. (b) Distribution of $\ln A'_{pca}$. Dashed lines show the mean value obtained for proton (red) and iron (blue) showers.

primaries [64]. The relation between the first moment of the X_{max} distributions and the mean logarithmic mass is given by

$$\langle X_{max} \rangle = \langle X_{max} \rangle_p + f_E \langle \ln A \rangle, \quad (4.2)$$

where $\langle X_{max} \rangle_p$ is the mean X_{max} for protons and f_E is a parameter that depends on details of the hadronic interactions [17].

The average muon number was measured by Auger by analyzing highly inclined events ($62^\circ \leq \theta \leq 80^\circ$) above energies of 4×10^{18} eV [100]. The estimation of the muon number for each shower was performed comparing the measured footprint at the ground with simulations for a reference hadronic interactions model. This study is based on the fact that at such high zenith angles the electromagnetic component is strongly attenuated and therefore the WCD signals are dominated by the muonic part of the shower. Given the relation between the number of muons in a the shower and the energy $N_{\mu,p} = (E/\zeta_c^\pi)^\beta$ (where Eq. (1.35) was applied for protons), an expression similar to Eq. (4.2) can be derived for the average logarithm of the muon number [100]

$$\langle \ln N_\mu \rangle = \langle \ln N_\mu \rangle_p + (1 - \beta) \langle \ln A \rangle \quad (4.3)$$

$$\beta = 1 - \frac{\langle \ln N_\mu \rangle_{\text{Fe}} - \langle \ln N_\mu \rangle_p}{\ln 56}. \quad (4.4)$$

More recently, Auger has published the results from the analysis of AMIGA data. AMIGA estimates the muon content of the shower by fitting the lateral distribution of muon densities on single events. The muon estimator, ρ_{35} , is obtained from the muon density at 450 m corrected for attenuation effects [36]. Introducing the z factor [101]

$$z = \frac{\ln N_\mu - \ln N_{\mu,p}}{\ln N_{\mu,\text{Fe}} - \ln N_{\mu,p}}, \quad (4.5)$$

a comparison between the results obtained with AMIGA and the ones with the inclined showers can be performed [55]. In Eq. (4.5), N_μ corresponds to the measured number of muons and $N_{\mu,p}$ ($N_{\mu,\text{Fe}}$) is the number of muons for proton (iron) showers. The relation between z and the logarithmic mass is simply $z = \frac{\langle \ln A \rangle}{\ln 56}$. Of course, simulations need to be used in order to obtain $\langle \ln A \rangle$ in the aforementioned analyses, and therefore a dependency on the model used for the hadronic interactions exists.

In Fig. 4.24, the mean logarithmic mass, as obtained from different observations in Auger, is shown as a function of the energy. The gray squares show the predictions from X_{max} measurements [64], the blue dots correspond to the analysis of inclined events³ [101], the white dots correspond to AMIGA measurements [55], and the red squares show the $\ln A$ as obtained using SSD data in this work. For this plot, the events corresponding to the bin at $10^{18.5}$ eV are not shown as the SSD reconstruction efficiency at this energy drops below 95% (see Section 3.2.4). The data of the other measurements corresponds to predictions using QGSJET-II.04 as hadronic interaction model.

The predicted $\ln A$ using SSD data up to $10^{18.9}$ eV seems to be in agreement with the measurements obtained from the analysis of inclined showers. However, systematic uncertainties are still too large to draw any conclusion regarding the energy dependence. What it is clear is the absolute offset with respect to the X_{max} measurements. This is expected since the deficit in the number of muons in simulated showers has a smaller effect on X_{max} than on ground signals.

4.5.4 Sources of systematic uncertainties

In this section, we will discuss the main sources of systematic uncertainties of the result presented in Section 4.5.2. In general, any kind of systematic uncertainties (or the most relevant) come from the fact that we are applying reconstruction methods, that were derived from Monte Carlo simulations, to real data. The main sources for systematics are discussed below.

³The acronym HAS stands for *Horizontal Air Showers*.

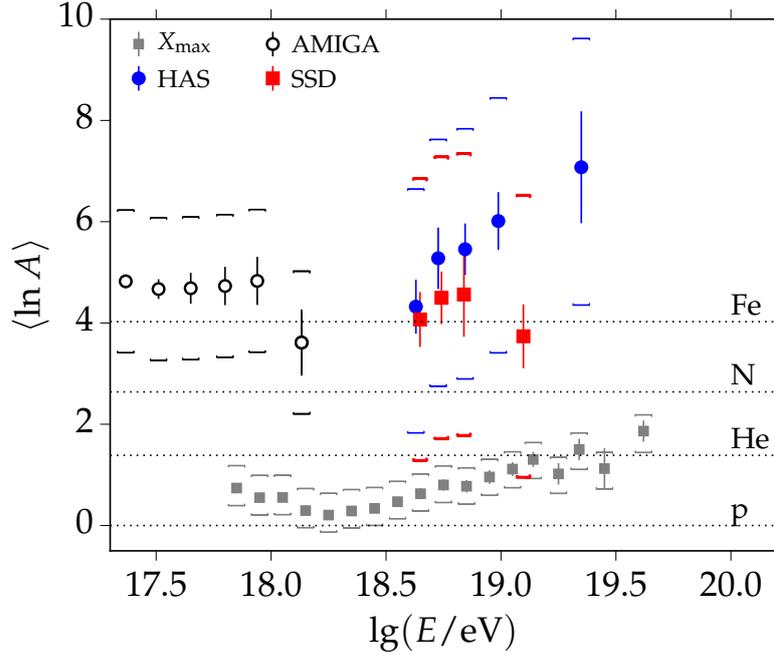


Figure 4.24: Average logarithmic mass as a function of the energy as estimated from different measurements and using QGSJET-II.04 as hadronic interaction model. Systematic uncertainties on each measurement are shown with square brackets (see text for details).

LDF parameterization

The fit of the lateral distribution function is a source of systematic errors as the reconstructed signal at the optimal distance might be affected by the assumptions taken for the parameterization of the LDF. In the case of the WCD, previous studies have reported an overall systematic uncertainty on S_{1000}^{wcd} of the order of 5% for proton and iron showers [43]. This uncertainty is probably underestimated as it does not take into account any mass-dependent systematics. The case of the SSD can be understood using what was already discussed in Section 3.2.3. In Fig. 3.10b, it can be seen that the average bias in the reconstructed S_{1000}^{ssd} is within $\pm 5\%$ for almost all energies and zenith angles.

Another systematic uncertainty is a result of applying the attenuation correction to the shower size

$$S_{38} = S_{1000}/f_{\text{Att}}(\theta), \quad (4.6)$$

which was discussed in Section 3.3.1. The systematic uncertainty can be calculated by error propagation

$$\begin{aligned} \sigma_{\text{sys}}(S_{38}) &= \sqrt{\left(\frac{\partial S_{38}}{\partial S_{1000}}\right)^2 \sigma_{\text{sys}}^2(S_{1000}) + \left(\frac{\partial S_{38}}{\partial f_{\text{Att}}}\right)^2 \sigma_{\text{sys}}^2(f_{\text{Att}})} \\ &= S_{38} \sqrt{\sigma_{\text{sys}}^2(S_{1000})/S_{1000}^2 + \sigma_{\text{sys}}^2(f_{\text{Att}})}, \end{aligned} \quad (4.7)$$

where $\sigma_{\text{sys}}^2(f_{\text{Att}})$ is the systematic uncertainty in the parameterization of f_{Att} and is caused by the statistical uncertainties of the parameters a , b and c in Table 3.5. These uncertainties are smaller (by one order of magnitude) than the systematic uncertainty in the shower size $\sigma_{\text{sys}}(S_{1000})$ and therefore can be neglected in the calculation of $\sigma_{\text{sys}}(S_{38})$. In order

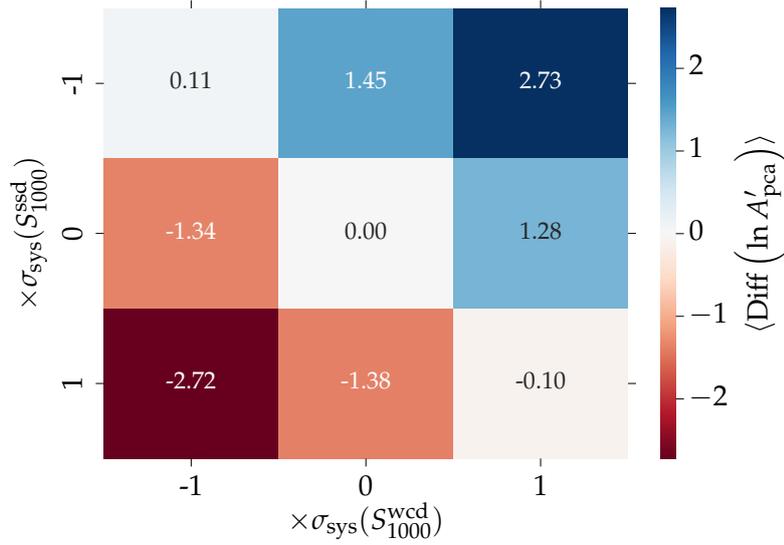


Figure 4.25: Average difference in $\ln A'_{\text{pca}}$ (as defined in Eq. (4.8)) accounting for different systematic shifts in S_{1000} of the WCD (x -axis) and of the SSD (y -axis). See text for details.

to estimate the impact of the systematics in the reconstructed shower size on the mass estimator, the following difference is calculated

$$\text{Diff}(\ln A'_{\text{pca}}) = \ln A'_{\text{pca}} - (\ln A'_{\text{pca}})^*, \quad (4.8)$$

where $(\ln A'_{\text{pca}})^*$ corresponds to the value of $\ln A'_{\text{pca}}$ obtained after $S_{1000}^{\{\text{wcd}, \text{ssd}\}}$ were shifted by $\pm \sigma_{\text{sys}}$. Since $\ln A'_{\text{pca}}$ is a combination of the two measurements, it is possible that a shift in S_{1000}^{wcd} is counterbalanced by the shift in S_{1000}^{ssd} , and vice-versa. Furthermore, we do not know the degree of correlation between the systematic uncertainties on both measurements. Therefore, we calculate the difference in $\ln A'_{\text{pca}}$ for all possible combinations. The result is shown in Fig. 4.25. The values of the absolute difference are obtained varying S_{1000}^{wcd} by $\pm 5\%$ (x -axis) and varying S_{1000}^{ssd} by $\pm 5\%$ (y -axis). As a cross-check, when both measurements are not shifted the result in the difference is zero. A maximum difference of -2.72 is obtained when S_{1000}^{wcd} and S_{1000}^{ssd} are shifted by -5% and $+5\%$, respectively.

Choice of the hadronic interaction model

The next source of uncertainty is related to the hadronic interaction model used in CORSIKA to simulate the showers that were used for the parameterization of the LDF and therefore to obtain our mass estimator. There are three models which are mostly used to describe the hadronic interactions in the shower and these are QGSJET-II.04 [102], EPOS-LHC [103] and SIBYLL-2.1 [67]. Although analyzing the differences between these models is not the scope of this thesis, it is well known that they predict different values for different shower observables (see, for example, [104]). For the analyses presented in this work we have used QGSJET-II.04 and this makes our parameterizations, in principle, “model-dependent”. One way to estimate the impact of using another hadronic interaction model is to apply the results for the predicted $\ln A'_{\text{pca}}$ to simulated data using other models and compare with the current result.

For this comparison, we have used a library of simulated showers at fixed energies with EPOS-LHC. The result on the predicted $\ln A'_{\text{pca}}$ as a function of energy is shown in Fig. 4.26. The discrepancy at $10^{18.5}$ eV is of the order of $+2$ in $\ln A'_{\text{pca}}$ for both, proton and

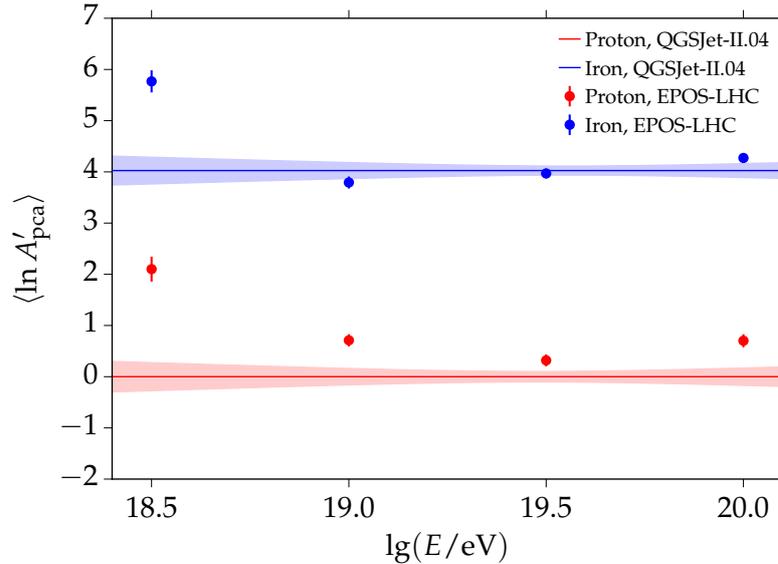


Figure 4.26: Mean of $\ln A'_{pca}$ as a function of energy estimated using simulated showers with EPOS-LHC as hadronic interaction model (circles). Lines correspond to $\ln A'_{pca}$ derived using QGSJET-II.04 as described in Section 3.4.2.

iron, but at this energy we are affected by the efficiency cut. Due to this cut, we select only events that were reconstructed (with a probability of 95%) and these correspond to showers that could have higher number of muons, depositing a higher signal at ground and causing the upward fluctuation (the same fluctuation is observed for simulations using QGSJET-II.04). Therefore the result obtained at this energy bin might not be a good representation of the differences between the models and will be omitted. Unfortunately, we do not have a library of simulations with a continuous range in energies using EPOS-LHC (this is the reason why QGSJET-II.04 was chosen for the analysis), and we can not estimate the systematic error at energies just above the efficiency threshold. We will consider the average of $\ln A'_{pca}$ as obtained with EPOS-LHC on proton showers above 10^{19} eV, giving a systematic error due to the choice of the hadronic interaction model of

$$\sigma_{\text{sys}}(\text{had}) / \ln A'_{pca} = +0.57. \quad (4.9)$$

Discrepancy in the number of muons between data and simulations

As it was mentioned at the beginning of this section, systematic errors are mainly driven from the fact that we are applying reconstruction algorithms derived using Monte Carlo simulations to real data. It is well known that simulations and data, in general, do not match. Perhaps one of the most notorious results is the discrepancy between the observed number of muons in simulations and in data (see for example the discussion given in [104] and references therein). Using Auger hybrid events with energies from $10^{18.8}$ to $10^{19.2}$ eV it was found that for a mixed composition, a rescaling in the hadronic shower of about 33% (61%) for EPOS-LHC (QGSJET-II.04) is needed [57]. More recently, direct muon measurements with AMIGA have confirmed this deficit at lower energies (from $10^{17.5}$ to 10^{18} eV) [36, 55].

The deficit of N_{μ} in simulations reported by Auger was determined by studying the ratio of S_{1000}^{wcd} in data and in simulations, using high quality hybrid events, selecting those simulated events that matched the longitudinal profile at a certain energy. As we do

not want to speculate how much this deficit affects the SSD signals at ground, we leave this as a future exercise to be solved as soon as the statistics of hybrid measurements including SSD reconstruction is sufficient.

The total systematic uncertainty in $\ln A'_{\text{pca}}$ is determined by adding in quadrature the systematic uncertainties previously discussed, namely

$$\sigma_{\text{sys}}(\text{tot}) = \sqrt{\sigma_{\text{sys}}^2(\text{LDF}) + \sigma_{\text{sys}}^2(\text{had})}, \quad (4.10)$$

where we have assumed no correlation between the systematics. The total systematic uncertainty in $\ln A'_{\text{pca}}$ is indicated with square brackets in Fig. 4.22.

CHAPTER 5

Summary

Determining the mass composition of UHECR at the top of the atmosphere is one of the key ingredients in elucidating their origin and propagation through the cosmos. With the aim to answer these questions, the Pierre Auger Observatory is undergoing a major detector upgrade, called AugerPrime. The main component of the upgrade is a scintillator detector which complements the water-Cherenkov detector in the task of measuring the footprint of air showers induced by cosmic rays. Their differing responses to the shower components aid in the reconstruction of the primary mass. The work presented in this dissertation is focused on the development of the necessary tools for the reconstruction of air showers with AugerPrime detectors and the analysis of the first measurements. The major contributions are summarized below.

Calibration of the Scintillator Surface Detector

Studies of the calibration of the SSD were performed using simulations of low energy showers. These simulations are intended to reproduce the flux of background particles which are constantly measured and used for the calibration of the SD stations. We confirmed that, contrary to the WCD, the peak in the charge distribution deposited by particles in the SSD is similar for all charged particles, i.e., for MIPs. From the fit to this distribution, the so-called MIP charge is obtained and then used for the calibration of the SSD signals. The results from this analysis may serve as reference for future measurements of the MIP charge in the field.

Event reconstruction and mass composition

Methods based on simulations were developed in order to incorporate the SSD into the standard reconstruction of air showers with the SD. To achieve this objective a model for the uncertainty of the SSD signals was derived, and which is valid over the vast majority of the dynamic range. In addition, a parameterization of the shape parameters of the SSD LDF was obtained. This is necessary for the reconstruction of events with low multiplicities of SSDs. From the fit of the LDF, the corresponding shower size of the SSD, S_{1000}^{ssd} , can be obtained. The average bias in S_{1000}^{ssd} was shown to be within $\pm 5\%$ for energies above $10^{18.5}$ eV and for zenith angles below 56° .

With a parameterization of the LDF in hand, the next step is to parameterize the zenith dependence of the shower size in order to account for the attenuation in the atmosphere. Using this parameterization, zenith-independent estimators S_{38}^{wcd} and S_{38}^{ssd} can be obtained. Statistical methods such as PCA were exploited in order to reconstruct properties of the primary. For this purpose, a library of simulated air showers was used. Proton and iron showers were simulated with energies in the range from 10^{17} eV to 10^{20} eV, and zenith angles below 70° . QGSJET-II.04 was used as model for the hadronic interactions.

PCA was applied to a dataset composed the reconstructed quantities, S_{38}^{wcd} and S_{38}^{ssd} , the energy and the logarithmic mass of the primary, in such a way that it was possible to find solutions for the energy and the mass as linear combinations of S_{38}^{wcd} and S_{38}^{ssd} . It was shown that adding the information provided by the SSD can help to reduce the mass-dependent bias in the reconstructed energy with the SD.

In addition, an energy-independent estimator of the logarithmic mass, $\ln A$, was obtained. It was shown that the distributions of $\ln A$ from proton and iron showers have large tails which difficult the mass separation on a single-event basis. However, mass composition at the highest energies can be studied on average with this method.

Analysis of data from AugerPrime detectors

The first analysis of air showers measured with AugerPrime detectors was presented. In September 2016, twelve stations of an EA were upgraded with an SSD, a new electronics board and an additional small PMT inside the WCD for measurements close to the core, where the large PMTs saturate. The analysis of these prototype stations is of key importance prior to the deployment of AugerPrime detectors in the full SD array. Studies on the calibration and performance of the upgraded stations were performed.

An update of the Offline software framework was also carried out in order to ease the data analysis and to incorporate the detectors of the upgrade in the standard reconstruction methods used by the Auger collaboration.

During March 2019, seventy-seven additional SSDs were deployed and connected to the existing electronics in the so-called PPA. In combination with the stations in the EA, the PPA increased the number of showers measured with an SSD at the highest energies. The analysis of the ratio between the signals, $S_{\text{SSD}}/S_{\text{WCD}}$, revealed the differing response to the shower components of the two detectors. The increase of the ratio with the energy as well as the steeper LDF of the SSD signals, confirmed the expected behavior of the SSD being more sensitive to the electromagnetic component of the shower.

The aforementioned parameterization of the LDF was applied to fit the lateral distributions of real events, thus allowing us to obtain correlations between S_{1000}^{wcd} and S_{1000}^{ssd} . The ratio of these two quantities as a function of the energy showed an interesting behavior which needs to be studied once a sufficient number of events have been measured, in particular with showers also detected with the FD.

First estimate of the mass composition of UHECR with AugerPrime

The first estimate of the mass composition above $10^{18.5}$ eV was obtained from real events measured with AugerPrime detectors. The studies performed using Monte Carlo simulations were applied to a selected number of events reconstructed with stations of the PPA. The selection requires that the station with largest signal in the event has an SSD. In

addition, the requirement of at least 3 SSDs used for the fit of the LDF was applied. This increases the quality of the fit and reduces the probability of a given event only being partially contained in the PPA. The final dataset consists of 194 events with energies above $10^{18.5}$ eV.

The estimate of the average logarithmic mass, $\langle \ln A \rangle$, was obtained using information from the SSD reconstruction. It was shown that the values of $\langle \ln A \rangle$ as estimated with SSD data indicate a heavy composition at the available energies. This is, however, expected given the deficit in the number of muons predicted in simulated air showers. In order to confirm this, we compared our estimate to those obtained from measurements of the muonic content of the shower such as those with AMIGA and the analysis of very inclined showers with the SD. It was shown that the values of $\langle \ln A \rangle$ estimated in this work between $10^{18.6}$ eV and $10^{18.9}$ eV seem to be in agreement with the measurements obtained with the analysis of very inclined showers. However, systematic uncertainties are still too large to draw any conclusion from this comparison. The main source of systematic uncertainties was determined to be driven by the fit of the LDF. The impact of the choice of the hadronic interaction model in the simulations was also studied as a potential source of systematic error. The prediction of $\ln A$ using simulations with EPOS-LHC resulted in small deviations at high energies with respect to the baseline model, QGSJET-II.04. More exhaustive analysis needs to be done in order to confirm such a small impact of the choice of hadronic model in the results.

As a final remark, a step forward needs to be taken in order to improve some of the methods presented in this work, which may help in developing new tools for the analysis of the data. The results presented in this thesis should be taken as preliminary, but may serve as a precedent for future analysis in the context of mass composition studies with the AugerPrime upgrade.

Acronyms

This is a list of acronyms used within this work sorted alphabetically according to the short version.

AERA	Auger Engineering Radio Array	11
AMIGA	Auger Muon Detectors for the Infill Ground Array	10
CDAS	central data acquisition system	8
CIC	Constant Intensity Cut	17
CORSIKA	Cosmic Ray Simulations for Cascade	31
CMB	cosmic microwave background	2
EAS	extensive air shower	4
EA	Engineering Array	68
FADC	flash analog to digital converter	8
FD	Fluorescence Detector	6
FoV	field of view	9
HEAT	High Elevation Auger Telescopes	10
LDA	Linear Discriminant Analysis	58
LDF	lateral distribution function	14
LS	Local Station	68
LSD	Layered surface detector	104
MD	Muon Detector	11
MIP	minimum ionizing particle	30
NKG	Nishimura-Kamata-Greisen	15
PCA	Principal Component Analysis	53
p.d.f.	probability density function	38
PE	photoelectron	34
PMT	photo-multiplier tube	7
PPA	preproduction array	68
RMS	root mean square	34
SD	Surface Detector	6
SD-1500	array with 1500 m spacing	7
SD-750	array with 750 m spacing	9
SD-433	433 m spacing	9
SNR	supernova remnant	1
SDP	shower detector plane	14
SiPM	silicon photomultiplier	11
SPMT	Small PMT	33
SSD	Scintillator Surface Detector	25
UB	Unified Board	8
UHECR	ultra-high energy cosmic rays	1
UMD	Underground Muon Detector	35
UUB	Upgraded Unified Board	34
VEM	vertical-equivalent muon	8
WCD	water-Cherenkov Detector	6
WLS	wave-length shifting	11

List of Figures

1.1	All-particle flux of cosmic rays	2
1.2	Interactions of protons and nuclei with CMB	3
1.3	Hillas diagram	4
1.4	Illustration of an Extensive Air Shower	5
1.5	Layout of the Pierre Auger Observatory	7
1.6	Picture and schematic of a WCD	8
1.7	FADC trace and calibration histogram in the WCD	9
1.8	Picture of a FD building and schematic of a telescope	10
1.9	Pictures of an AMIGA module and an AERA antenna	11
1.10	Configurations of the T4 triggers	13
1.11	Schematic of the shower front and time fit	13
1.12	Variance of the WCD signals	14
1.13	Example of a lateral distribution function	16
1.14	Energy calibration of the SD-1500	18
1.15	Energy spectrum measured by Auger	19
1.16	Moments of the X_{\max} distributions	19
1.17	Measurements of the muonic content by Auger	20
2.1	Evolution of the composition of UHECR with energy	22
2.2	Auger spectrum described with different astrophysical scenarios	23
2.3	N_{\max}^{μ} as a function of X_{\max}	24
2.4	Spectrum of emission of the scintillator bar	26
2.5	Spectrum of absorption and emission of the WLS fibers	26
2.6	Cookie and PMT of the SSD	27
2.7	Pictures of open and closed SSD	28
2.8	Energy levels of an organic molecule	28
2.9	Sketch of the particle detection principle of a scintillator	29
2.10	Stopping power of electrons and muons in polystyrene	30
2.11	Cross section of photons and probability of pair production	30
2.12	Results from background simulations	32
2.13	Distribution of charge produced in the SSD by different particles	33
2.14	Calibration fits of SSD simulated PE histograms	34
2.15	Distributions of the high gain to low gain ratio and RMS of baseline from upgraded detectors	34

3.1	Distributions of SSD signals	39
3.2	Sketch of stations ring and correction by azimuth dependency	40
3.3	Azimuthal dependency	41
3.4	Signal uncertainty model for the SSD	42
3.5	Baseline fluctuations	43
3.6	β and γ as function of $\sec \theta$	45
3.7	β and γ as function of $\lg S_{1000}$	45
3.8	Residuals of the SSD LDF	46
3.9	Relative difference between SSD LDF and data	47
3.10	Bias in S_{1000}^{ssd}	47
3.11	Average and relative bias of SSD S_{1000}	48
3.12	SSD reconstruction efficiency as a function of the energy	49
3.13	SSD reconstruction efficiency	50
3.14	Attenuation in shower size	51
3.15	Parameterization of f_{Att} for WCD and SSD	52
3.16	Comparison between reconstructed and Monte Carlo energy	53
3.17	S_{38} of WCD and SSD for proton and iron showers	54
3.18	Bias in reconstructed energy for proton and iron showers using the SSD	57
3.19	Contours of S_{38}^{wcd} and S_{38}^{ssd} for proton and iron showers	58
3.20	Mass discrimination for proton and iron showers at $10^{19.5}$ eV with SSD	60
3.21	Merit factors obtained applying LDA on fixed library	61
3.22	Estimate of $\ln A$ using PCA for proton and iron primaries	62
3.23	Corrected $\ln A_{pca}$ with energy	63
3.24	Distribution of $\ln A'_{pca}$ for proton and iron showers above $10^{19.7}$ eV	64
3.25	Distributions $\ln A'_{pca}$ for different energies	65
4.1	Map of the Surface Detector with locations of AugerPrime detectors	67
4.2	Sketch of Engineering Array	68
4.3	Charge histogram of an upgraded station	69
4.4	Variation of the VEM and MIP charge with time	70
4.5	Example of FADC traces in the upgraded station	71
4.6	Signal comparison between old and new electronics	71
4.7	SSD and WCD signal distribution and signal correlation	72
4.8	Rate of events measured with stations of the preproduction array	74
4.9	Reconstructed energy of events measured by the SSD preproduction array	74
4.10	Ratio of SSD and WCD signals as a function of radial distance	75
4.11	Ratio of SSD and WCD signals as a function of the reconstructed energy	76
4.12	SSD and WCD average lateral distributions	77
4.13	Event reconstruction with AugerPrime detectors	78
4.14	WCD and SSD traces	79
4.15	Multiplicity of SSDs as a function of energy	79
4.16	Event reconstruction using WCD large PMTs and SPMT	80
4.17	Residuals of the SSD LDF using data from the preproduction array	80
4.18	Distributions of distances to the core and $\lg S$ for saturated SSDs	81
4.19	Correlation of S_{1000}^{ssd} and S_{1000}^{wcd}	81
4.20	Ratio of S_{1000}^{ssd} and S_{1000}^{wcd} as a function of the energy	82
4.21	Reconstruction of events with the SSD preproduction array	83
4.22	Mean composition with data from the SSD preproduction array	85
4.23	Scatter and distribution of $\ln A'_{pca}$ estimated from SSD data	85
4.24	Average $\ln A$ estimated from different measurements	87

4.25	Difference in $\ln A'_{\text{pca}}$ due to systematics in S_{1000}	88
4.26	Mean of $\ln A'_{\text{pca}}$ estimated from simulations with EPOS-LHC	89
B.1	Number of simulated primaries	108
C.1	Index of refraction Cherenkov production in water	114
C.2	Stopping power of muons and electrons in water	115
C.3	Range for muons and electrons in water	116
C.4	Probability of pair production in water	117
C.5	Photoelectron production in the water-Cherenkov detector	119
C.6	SSD signal variance as a function of signal for different zenith angles . . .	120
C.7	SSD signals fraction	121
C.8	Parameterization of the distance cut	121
C.9	Coefficients of β and γ parameterization	122
C.10	Alternative functional form for β	122
C.11	Residuals of the SSD LDF as a function of the distance to the core	123
C.12	Residuals of the SSD LDF as a function of the SSD signal	124
C.13	Bias in reconstructed S_{1000}^{ssd} as a function of energy	125
C.14	Distributions of S_{38} from WCD and SSD	126
C.15	Residuals and average bias in energy using a Non-linear ansatz	126
C.16	Bias in energy for proton and iron showers using Non-linear ansatz	127
C.17	Bootstrap method for calculating the uncertainty on the Merit Factor	127
C.18	Distributions of $\ln A'_{\text{pca}}$ in different energy bins	128

List of Tables

3.1	Simulations of the <i>fixed</i> library	38
3.2	Simulations of the <i>continuous</i> library	38
3.3	Coefficients of the β and γ parameterizations of the SSD LDF	44
3.4	Coefficients of the function for the SSD reconstruction efficiency	49
3.5	Parameterization of f_{Att} for the WCD and SSD	52
3.6	Parameterization of $\ln A_{pca}$ with energy	62
4.1	Calibration parameters	73
4.2	Number of reconstructed events with the PPA and quality cuts	83
B.1	Parameters of the simulations of low energy primaries	108

APPENDIX A

Proposals for the Detector Upgrade

Many efforts have been made over the past years to upgrade the observatory and several upgrade proposals were studied in detail with the general goal of measuring composition sensitive observables over the whole available energy range. This implies both the study of the prototypes in the field and the study of detector simulations to quantify the expected performance in order to achieve the desired physics results. For example, determining the resolution in the measured number of muons at distances closer to the shower core. The different proposals for the detector upgrade that were taken into consideration are summarized below.

AMIGA-Grande

With the AMIGA muon detector system operating since the end of 2009, all the relevant features (mechanics, electronics and communication) had already been validated. AMIGA-Grande was suggested as a proposal for the upgrade of the Observatory based on the experience gained with AMIGA. The baseline of the project aims to cover the full array with single 10 m^2 muon counters deployed at 1500 m and 3000 m spacings [105]. With AMIGA and AMIGA-Grande the spectrum of cosmic rays would be covered from $\sim 10^{17}$ eV onwards providing a well understood data set. The scintillator counters are buried underneath the SD tanks at a chosen depth to guarantee an uniform shielding of the electromagnetic component of the shower. Preliminary studies based on simulations showed that the accuracy for reconstructing the muon density at 1000 m from the shower core is in the range 12% – 15% for proton and iron primaries of 10^{19} eV at all zenith angles [106].

ASC-II

The idea of using scintillators on the surface to get the muonic component from comparison with the WCD signals was proposed many years ago, and an implementation for the Auger observatory was proposed in 2009 by the Bariloche group [107]. The original prototype of 0.25 m^2 was built in 2010 and took data for more than a year. The ASC-II (Auger Scintillators for Composition-II) upgrade suggested adding a small 2 m^2 thin scintillator atop each tank of the Surface Detector. The combined analysis of the top scintillator (equally sensitive to the muonic and the electromagnetic components) and the bottom WCD (strongly sensitive to muons and less to electrons and photons) showed

that determining the muonic component of an air shower with good accuracy was possible, giving the absolute energy scale at 10% and the number of muons at 20% [108]. The original detector design consisted in build a low cost scintillator made by 27 scintillator bars of 1.8 m length with green wavelength shifting optic fiber. Each of the fibers guide the light to a 1/2" PMT collecting the total charge of all the bars. The detector is comprised by a single block of scintillator bars glued together with a PVC enclosure of 5 mm of thickness. To ensure light tightness the detector is enclosed within two steel sheets of 0.6 mm of thickness. A second steel roof is placed 2 cm on top to allow air flow and reduce temperature changes. It integrates easily in the Auger data acquisition scheme and its low cost, around 1500\$, allow to build and deploy one detector for every Auger WCD, both in the SD-1500 and in the SD-750, allowing to measure the muon contents in the whole Auger energy range.

LSD

The Layered surface detector (LSD) proposal aims to separate the electromagnetic and muonic components of the shower by dividing the water volume of the WCD in two layers. The top layer will collect a majority of Cherenkov photons produced by the electromagnetic component while the bottom one will collect mostly light from muons. The two water volumes of the LSD are separated by a reflective layer at a height of 80 cm from the tank bottom. The three standard PMTs of the Auger WCD are left in place and a central cylinder of about 10 inch diameter provides the structure and enclosure for an additional 9 inch PMT that collects the light from the bottom layer. Studies based on Monte Carlo simulations have shown that the reconstruction of shower parameters with optimal accuracy is possible with the LSD. In particular, the separation of the muonic and electromagnetic lateral distributions on an event by event basis provides an estimation of X_{\max} with resolution from 30 to 50 g/cm^2 and $< 20\%$ for N_{μ} [109].

MARTA

In the Muon Auger RPC for the Tank Array (MARTA) proposal, a direct measurement of the average value and RMS of the number of muons at ground is explored by deploying RPCs (Resistive Plate Chambers) under the WCD of the Auger SD array [110]. The tank act as shielding for the electromagnetic component and a concrete structure would be placed below it hosting four RPCs (covering a total area close to 8 m^2). MARTA units would be deployed with three different spacings (750 m, 1500 m and 2600 m) covering an effective area of about 2800 km^2 . The combined analysis of the MARTA and water-Cherenkov detector data would allow for the independent measurement of the electromagnetic and muonic components of the shower, as well as the development of cross-calibration methods that would improve the stability of the measurements and thus ensuring a better control of the systematic errors.

TOSCA

In strong synergy with AMIGA-Grande, The Observatory SCintillator Array (TOSCA) aimed to measure the muonic component of the shower in the ultra high energy region by covering the largest possible surface of the Auger SD array with simple and reliable buried detectors [111]. The TOSCA project upgrades the Auger Surface Detector providing a 10 m^2 muon detector made of plastic scintillator bars. Each detector is buried at 1.3 m aside each tank and is composed by four identical smaller modules hosting 2 meters

long plastic scintillator bars with WLS fibers. Each fiber bundles are collected together and connected to a single PMT located at the center of the active surface.

APPENDIX B

Simulations and Offline Sequences

B.1 Simulations of low energy showers

For the studies of the calibration of the SSD we generated a the flux of secondary particles using simulations of low energy showers over Malargüe. The simulations were performed with CORSIKA within the framework for simulation and analysis of the LAGO collaboration [112]. The flux of cosmic rays can be approximated as a power law

$$j(E) = j_0 E^{-\gamma},$$

where γ has a value close to 3. According to this, we simulated nuclei with atomic number in the range $1 \leq Z \leq 26$ and mass number in the range $1 \leq A \leq 56$ in the energy range from 10 GeV to 10^6 GeV. For this work we simulated a flux of 1 hour per square meter. Details are given in Table B.1. The Number of primaries as a function of the atomic number Z is shown in Fig. B.1.

B.1.1 Studies of the calibration of the Scintillator Surface Detector

The Offline sequence used to study the calibration of the SSD signals is shown below. The process can be described in three steps: (i) Injection of secondary particles produced in the low energy showers (see Appendix B.1) into the station comprised by an WCD and a SSD. (ii) Simulation of the detectors response, production of photoelectrons and time traces. (iii) Calculation of the total signal deposited by each particle by integrating the FADC traces.

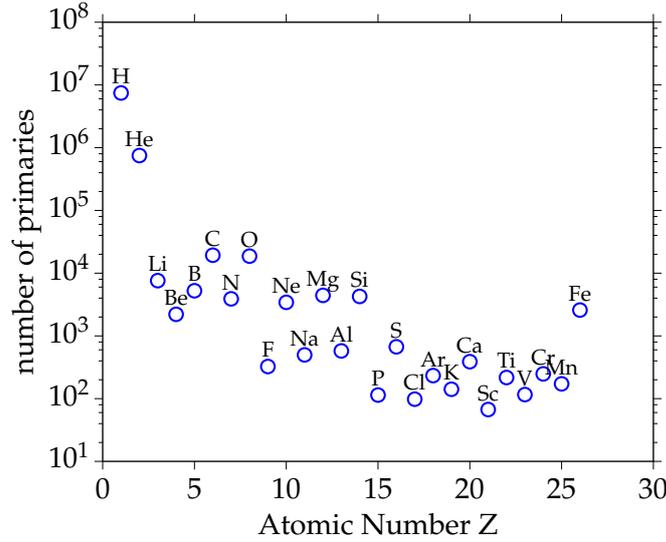
The secondary particles produced in the simulations of low energy showers are stored in ASCII files which serve as input of the module LEInjectorASCII. This module is intended to “inject” each individual particle in the station. This injection is done accordingly to the particle trajectory. From the components of the particle momentum $\vec{p} = (p_x, p_y, p_z)$, the zenith angle with respect to the vertical axis (axis perpendicular to the plane of the detector), and the azimuth angle can be computed

$$\begin{aligned}\theta &= \arccos(p_z/|\vec{p}|), \\ \phi &= \arctan(p_y/p_x).\end{aligned}$$

Particles are injected in a virtual cylinder which contains the WCD and the SSD. The position of the particles in the cylinder is found taking into account their zenith angle:

Table B.1: Parameters of the simulations of low energy primaries.

Energy	$5 \text{ GeV} \leq E \leq 1 \text{ PeV}$
Flux	$j(E) = j_0 E^{-2.7}$
Zenith	$0^\circ \leq \theta \leq 70^\circ$
Azimuth	$0^\circ \leq \phi \leq 360^\circ$
Time	3600 s
Number of primaries	8273227

**Figure B.1:** Number of low energy primary cosmic rays simulated for the studies of the calibration of the SSD.

particles with small zenith angles will have more chances to enter from the top area of the cylinder, while more inclined particles will enter from the side of the cylinder. The probability of hitting the top area can be calculated as

$$P_{\text{top}} = \frac{A_{\text{top}}}{A_{\text{top}} + A_{\text{side}}},$$

where A_{top} and A_{side} are the projected top and side areas of the cylinder

$$\begin{aligned} A_{\text{top}} &= \pi R^2 \cos \theta, \\ A_{\text{side}} &= 2 R h \sin \theta, \end{aligned}$$

being R and h the radius and height of the cylinder, respectively. Once a decision was made, the position of the particle is chosen by selecting random points from a circle of radius R , in the case of a top injection, or from a rectangle of area $R \times h$, in the case of a side injection.

The LEInjectorASCII module gives the information of the particle type, position and momentum (energy and direction) to the next module. The G4StationSimulator is a Geant4-based tool to simulate the response of the detectors to different particles. Time distributions of photoelectrons are simulated in the SdPMTsimulator module and the digitization of these pulses by the electronics is made in the SdFilterFADCSimulator module. We will not go through the details of the simulation, for the implementation of the SSD in the simulation chain see [90]. We can obtain the signal deposited by each

injected particle by integrating the FADC traces. This is done in the DataWriter module. This sequence was also used in other unrelated topics like the study of the detection of Forbush decrease with the WCD [113].

```
<sequenceFile>

  <enableTiming/>

  <moduleControl>
    <loop numTimes="1" pushEventToStack="yes">
      <module> EventGeneratorOG </module>

      <loop numTimes="unbounded" pushEventToStack="yes">

        <module> LEInjectorASCII </module>

        <module> G4StationSimulatorOG </module>
        <module> SdSimulationCalibrationFillerOG </module>
        <module> SdPMTSimulatorOG </module>
        <module> SdFilterFADCSimulatorMTU </module>
        <module> SdBaselineSimulatorOG </module>

        <module> DataWriter </module>
      </loop>

    </loop>

  </moduleControl>

</sequenceFile>
```

B.2 Sequence for reconstruction of events with AugerPrime detectors

An example *Offline application* dedicated for reconstructing events using stations from the Engineering Array and from the SSD preproduction array can be found in Documentation/ExampleApplications/SdSSDDataReconstruction, with the following sequence of *modules*:

```
<sequenceFile>

  <enableTiming/>

  <moduleControl>
    <loop numTimes="1" pushEventToStack="yes">
      <module> EventGeneratorOG </module>

      <loop numTimes="unbounded" pushEventToStack="yes">

        <module> EventFileReaderOG </module>
        <module> EventCheckerOG </module>
        <module> SelectEvents </module>
        <module> SdEACalibrationFillerKG </module>

        <module> SdPMTQualityCheckerKG </module>
        <module> TriggerTimeCorrection </module>
        <module> SdCalibratorOG </module>

        <module> SdStationPositionCorrection </module>

      </loop>

    </loop>

  </moduleControl>

</sequenceFile>
```

```
<module> SdBadStationRejectorKG </module>
<module> SdSignalRecoveryKLT </module>
<module> SdEventSelectorOG </module>

<module> SdPlaneFitOG </module>
<module> LDFFinderKG </module>
<try>
  <module> ScintillatorLDFFinderKG </module>
</try>
<module> EnergyCalculationPG </module>
<module> Risetime1000LLL </module>
<module> DLECorrectionGG </module>
<module> SdEventPosteriorSelectorOG </module>

<module> RecDataWriterNG </module>
</loop>

</loop>

</moduleControl>

</sequenceFile>
```

APPENDIX C

Analysis

This appendix contains additional information related to the analyses performed in different parts of the thesis.

C.1 Chapter 1

C.1.1 The Cherenkov radiation

This section was highly motivated by the nice description given in the dissertation of H. Asorey [114].

A charged particle moving through a dense medium will lose its energy due to the interaction with the medium. Let us assume b as the impact parameter of a charged particle over the line along the direction of motion in a dispersive medium of constant density ρ . The energy loss after traversing a distance dx , in a cylinder of radius a around the path of the particle is given by [115]

$$\left(\frac{dE}{dx}\right)_{b>a} = -ca \operatorname{Re} \int_0^\infty B_3^*(\omega) E_1(\omega) d\omega. \quad (\text{C.1})$$

This expression relates the differential energy loss for regions where $b > a$ after traversing an amount of matter x on the direction of motion, with the longitudinal component of the electric field E_1 , and the transversal component of the magnetic field B_3 , as a function of the frequency ω .

Assuming a charged particle moving in the longitudinal direction with speed $v = \beta c$ in a medium with dielectric constant $\epsilon(\omega)$ and atomic number Z . The wavelength of the radiation emitted by the particle will be modified by the presence of the dielectric medium

$$\lambda^2 = \frac{\omega^2}{v^2} - \frac{\omega^2}{c^2} \epsilon(\omega) = \frac{\omega^2}{v^2} (1 - \beta^2 \epsilon(\omega)). \quad (\text{C.2})$$

If we take the corresponding expressions for the fields E_1 and B_3 , the integrand in Eq. (C.1) results

$$B_3^*(\omega) E_1(\omega) = \left(\frac{Ze}{c}\right)^2 \left(-i\sqrt{\frac{\lambda^*}{\lambda}}\right) \omega \left(1 - \frac{1}{\beta^2 \epsilon(\omega)}\right) e^{-a(\lambda + \lambda^*)}. \quad (\text{C.3})$$

Taking the real part of this expression and integrating over frequencies, gives the energy deposited far from the path of the particle. If λ has a positive real part, the exponential factor in Eq. (C.3) will cause the expression to vanish rapidly at large distances. All the energy is deposited near the path. This is not true only when λ is purely imaginary, $\lambda^* = -\lambda$ and the exponential factor is equal to one.

Expression Eq. (C.1) becomes independent of a ; some of the energy escapes to infinity as radiation. This can be seen in Eq. (C.2) if $\epsilon(\omega)$ is real (no absorption) and $\beta^2\epsilon(\omega) > 1$. This condition can be written as

$$v > \frac{c}{\sqrt{\epsilon(\omega)}}. \quad (\text{C.4})$$

This shows that particles whose speed is larger than the phase velocity of the electromagnetic fields at frequency ω will emit electromagnetic radiation of that frequency, the so-called *Cherenkov radiation*. In the case of a slightly absorbing medium, it can be shown that

$$\lambda = -i |\lambda| \quad \text{for } \beta^2\epsilon > 1,$$

meaning that $\sqrt{\lambda^*/\lambda} = i$ and Eq. (C.3) is real and independent of a . Equation Eq. (C.1) represents the energy radiated as Cherenkov radiation per unit distance along the path of the particle

$$\left(\frac{dE}{dx}\right)_{\text{Cherenkov}} = \left(\frac{Ze}{c}\right)^2 \int_{\beta^2\epsilon(\omega) > 1} \omega \left(1 - \frac{1}{\beta^2\epsilon(\omega)}\right) d\omega. \quad (\text{C.5})$$

This result is known as the Frank-Tamm formula which was published in 1937 as a theoretical explanation of the radiation observed by Cherenkov in 1934. Eq. (C.5) shows the strong dependency of the emission of radiation with the frequency.

Another feature of Cherenkov radiation is the angle under which radiation is emitted. Since the direction of propagation of the electromagnetic radiation is given by the Poynting vector, i.e., the direction of $\mathbf{E} \times \mathbf{B}$, it can be inferred from the geometry of the problem that Cherenkov radiation is emitted in an angle defined by

$$\cos \theta_C = \frac{1}{\beta\sqrt{\epsilon(\omega)}}, \quad (\text{C.6})$$

which is only dependent on the medium and the particle speed (for a relativistic particle with $\beta \approx 1$, in water, $\theta_C \approx 41^\circ$). The criterion $\beta^2\epsilon > 1$ can now be rephrased as the requirement that θ_C be a physical angle with cosine less than unity. The emission angle can be interpreted qualitatively in terms of a "shock" wavefront in a similar way as the sonic boom produced by an aircraft in supersonic flight. The number of photons produced per unit path length of a particle with charge Ze and per unit energy interval of the photon is [116]

$$\frac{d^2N}{dEdx} = \frac{\alpha Z^2}{\hbar c} \sin^2 \theta_C = \frac{\alpha^2 Z^2}{r_e m_e c^2} \left(1 - \frac{1}{\beta^2 n^2(E)}\right), \quad (\text{C.7})$$

where $\alpha \equiv (e^2/\hbar c)$ is the fine structure constant, and β is related with the rest mass of the particle and its momentum using $p \equiv mv = \beta\gamma m_0 c$ or

$$\beta(p) = \frac{1}{\sqrt{\left(1 + \left(\frac{m_0 c}{p}\right)^2\right)}}. \quad (\text{C.8})$$

The index of refraction n is a function of photon energy $E = \hbar\omega$, as is the sensitivity of the photomultiplier used to detect the light. As we will discuss later, Eq. (C.7) must be multiplied by the transducer response function and integrated over the region for which the Cherenkov condition holds.

Water-Cherenkov Detector response

As we already know, the WCD uses the Cherenkov light emitted by the charged particles when passing through the water volume inside the tank. In order to understand its response to the different particles composing the EAS, we previously need to describe the concept of *energy loss*.

A relativistic charged particle traversing some amount of matter will loss its energy progressively due to the electronic interactions in single collisions with the surrounding material. The mean rate of energy loss, also called *stopping power* for heavy particles is well-described by the Bethe-Bloch equation [116]

$$\left\langle -\frac{dE}{dx} \right\rangle = K z^2 \frac{Z}{A} \frac{1}{\beta^2} \left[\frac{1}{2} \ln \frac{2m_e c^2 \beta^2 \gamma^2 W_{\max}}{I^2} - \beta^2 - \frac{\delta(\beta\gamma)}{2} \right], \quad (\text{C.9})$$

where $K = 4\pi N_A r_e^2 m_e c^2$, W_{\max} is the maximum energy transfer in single collisions, z is the charge of incident particle, Z and A are the charge number and the atomic mass of medium, respectively, I is the mean excitation energy of the medium and δ is the density correction.

It describes the mean energy rate of energy loss in the region $0.1 \lesssim \beta\gamma \lesssim 1000$ for Z -intermediate materials with an accuracy of a few percent. $\langle dE/dx \rangle$ as described in Eq. (C.9) is the basis of much of our understanding of energy loss by charged particles¹, and in practical cases, most relativistic particles (like muons in air showers) have mean energy loss rates close to the minimum ("minimum-ionizing particles")

In the case of our detector, we are concerning about the interaction of the charged particles composing the extensive air shower, with the water inside the tank. These particles are mostly muons, electrons (positrons) and photons. Looking at Eq. (C.7), we clearly see that the total number of Cherenkov photons is proportional to the track length L of the particle inside the detector

$$N \propto L \int_{E_{\min}}^{E_{\max}} \left(1 - \frac{1}{\beta^2 n^2(E)} \right) dE. \quad (\text{C.10})$$

As we anticipated on the previous section, the collection of light by the PMTs depends on the energy of the produced photons due to the quantum efficiency. For a PMT used on the tanks of the Pierre Auger Observatory, higher sensitivity is observed for the range $300 \text{ nm} \lesssim \lambda \lesssim 570 \text{ nm}$ [114]. This range of wavelength belongs to the visible-NUV part of the electromagnetic spectrum as shown in Fig. C.1a, where the index of refraction for water is nearly constant $n(\lambda) \approx 1.33$. In Fig. C.1b the number of Cherenkov photons produced by electrons and muons at different energies and at the range of interest in wavelengths after traversing 1 cm of water is shown. We observe two important features: the emission of light starts when the condition given by Eq. (C.4) is fulfilled. On the other hand, the number of photons rapidly increases and tends to a constant value of $N_{\text{Ch}} \simeq 315 \text{ cm}^{-1}$.

¹For heavy projectiles, like iron, additional terms are required to account for higher order photon coupling to the target. Also, for electrons and positrons, the stopping power differs from the stopping power for heavy particles because of the kinematics, spin, charge, and the identity of the incident electron with the electrons that it ionizes. See chapter 32 of [116] for more details.

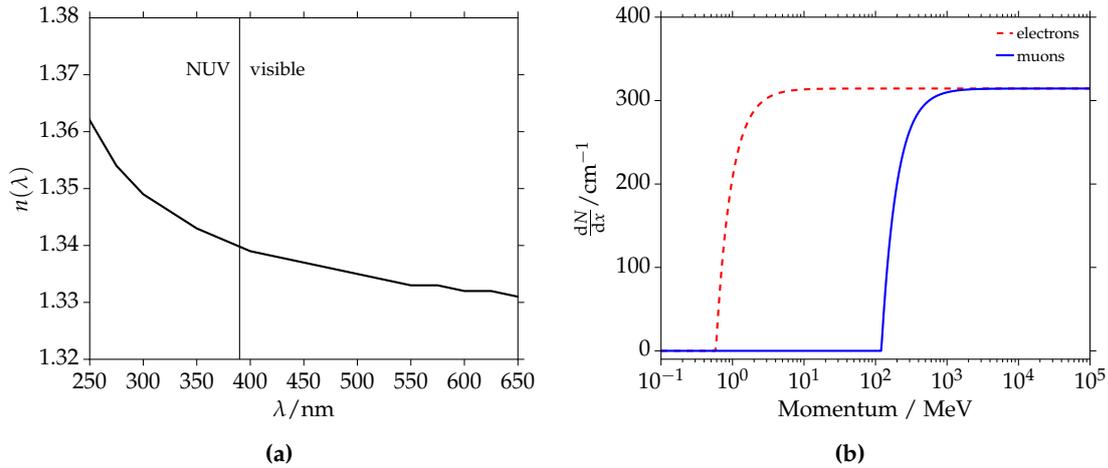


Figure C.1: (a) Refraction index of liquid water $n(\lambda)$ as a function of the wavelength. Data was taken from [117]. (b) Production of Cherenkov photons by electrons (dashed) and muons (solid) as a function of the momentum of the particle in the wavelength band $300 \text{ nm} \lesssim \lambda \lesssim 570 \text{ nm}$.

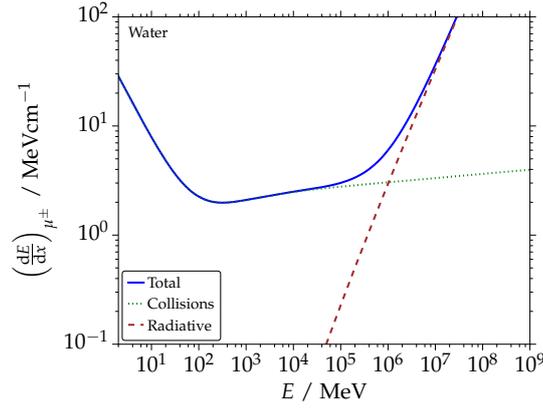
This is an important feature to understand the response of the water-Cherenkov detector. The signal arises from the number of photons produced in the tank, which at some point is no longer dependent on the particle energy but on the amount of water traversed by the particle. On the other hand, this is intrinsically related to the energy loss of particles in the medium. Fig. C.2 shows the mean rate for the energy loss of muons and electrons in liquid water.

Muons are the most penetrating component of the shower. This is because their stopping power is almost constant for a wide range of energies ($dE/dx \lesssim 10 \text{ MeV/cm}$), meaning that atmospheric muons, with energies about $E_\mu \sim 3 \text{ GeV}$ (Fig. 2.12a), are able to pass through several meters of water or even soil before being absorbed. The stopping power curve for muons (Fig. C.2a) shows that for energies between $\sim 100 \text{ MeV}$ and 150 GeV the mean energy loss has a nearly constant value between 2 and 3 MeV per centimeter. We can consider atmospheric muons as MIPs and therefore, their stopping power, or in other words, the deposited energy as constant.

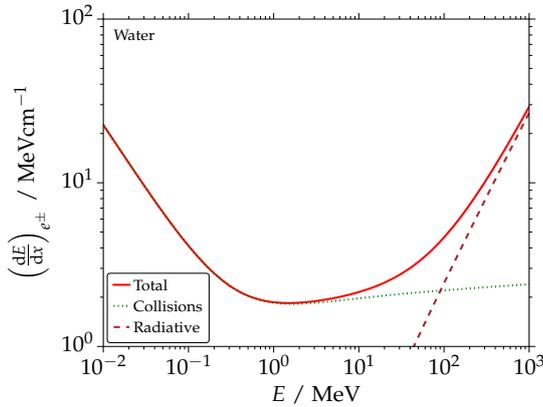
Taking the Frank-Tamm formula for the production of Cherenkov photons given by Eq. (C.5), we see that the number of produced photons rapidly saturates with the energy of the incoming particle and for energies below the typical for atmospheric muons. The number of Cherenkov photons only depends with the total track length of the muons inside the detector. This means that the deposited energy E_d by muons in the detector and therefore the detected signal, only depend on the trajectory of muons in water. We could conclude that the signal produced by muons in the water-Cherenkov detector is a convolution of the muon tracklength distribution with the muon zenith angle distribution.

For electrons, however, the situation is quite different. As shown in Fig. C.2b right, the stopping power for electrons in water is not small compared with the typical energies $E_e \sim 20 \text{ MeV}$ (Fig. 2.12a). Once the electron penetrates the water, it will lose its kinetic energy with the consequent slowing down of the velocity. Cherenkov photons will be produced while condition Eq. (C.4) holds.

Another interesting quantity is the range, R , or maximum distance a particle of a given energy can penetrate through a material before all kinetic energy is lost. It is a common way to parametrize particle interactions with materials and can be obtained by



(a)



(b)

Figure C.2: Stopping power dE/dx (solid line) for muons (Fig. C.2a) and electrons (Fig. C.2b) in liquid water as a function of the kinetic energy. Contributions due to collisions (dotted) and radiative losses (dashed) are also shown. Data taken from [118].

integrating the total stopping power over the full penetration depth in the continuous-slowing-down approximation (CSDA). In the CSDA, the rate of energy loss, dE/dx at every position along the track length is assumed constant; variations in energy-loss with energy, E , or penetration depth, x , are neglected. For a given incident energy E_0 the CSDA range, $R(E_0)$, is obtained such that

$$E_0 = \int_0^{R(E_0)} \left(\frac{dE}{dx} \right) dx. \quad (\text{C.11})$$

In Fig. C.3 the range for muons and electrons in water is shown as a function of the energy of the particle.

It can be seen that for electrons, the typical range inside the detector is around 10 cm, up to 80 cm for an energy of $E_e = 500$ MeV. This means that all electrons in the shower will be absorbed inside the tank.

The water-Cherenkov detector behaves as a calorimeter for the electromagnetic component of the shower, in the sense that all kinetic energy is absorbed. According to Fig. C.3a, it turns out that for muons with $T \lesssim 280$ MeV, ranges are shorter than the water depth (for vertical muons). These muons will deposit their energy and eventually will decay inside the tank, being possible the detection of the muon decay.

High energy photons are also detectable by the water-Cherenkov detector due to pair production $\gamma \rightarrow e^+e^-$. According to the relativistic formula $p = \beta\gamma m_0c$ and the condition

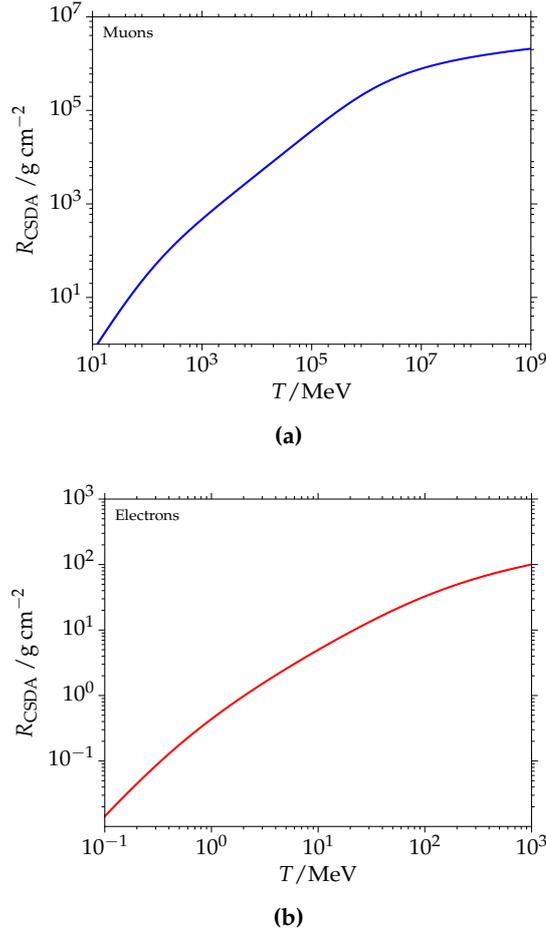


Figure C.3: Range in water using the continuous-slowing-down approximation as a function of the kinetic energy of (a) muons and (b) electrons.

for Cherenkov production $\beta > 1/n_w = 0.75$ ($\gamma \approx 1.51$), the minimum momentum for a particle with rest mass m_0 to produce Cherenkov radiation is

$$p_{\text{Ch}} = 1.13m_0c.$$

Then, the lower threshold in momentum for electrons is $p_{\text{Ch}}^e = 577 \text{ keV } c^{-1}$, which means an energy of $E_{\text{Ch}}^e = 0.770 \text{ MeV}$. Photons with energies $E_\gamma > 2E_{\text{Ch}}^e \simeq 1.6 \text{ MeV}$ will be able to produce Cherenkov radiation. The response of the WCD to photons is determined by the probability P that a photon interaction inside the tank will result in conversion to an e^+e^- pair.

The photon mass attenuation length (or mean free path) is given by $\lambda = \frac{1}{(\mu/\rho)}$ where μ/ρ is the mass attenuation coefficient and ρ the density. The intensity I of a beam of photons remaining after traversing an amount of matter x is

$$I(x, E_\gamma) = I(0, E_\gamma) \exp(-x/\lambda).$$

Then, the probability of a photon to convert into a e^+e^- pair is

$$P_{e^\pm}(E_\gamma) = 1 - e^{-\frac{x}{\lambda(E_\gamma)}}, \quad (\text{C.12})$$

where $x = h/\cos\theta_\gamma$, being θ_γ the zenith angle of the photon. The maximum value for θ_γ is determined by the diagonal of the cylinder $\theta_{\gamma_{\text{max}}} = \arctan(d/h) \approx 72^\circ$. This is shown in Fig. C.4.

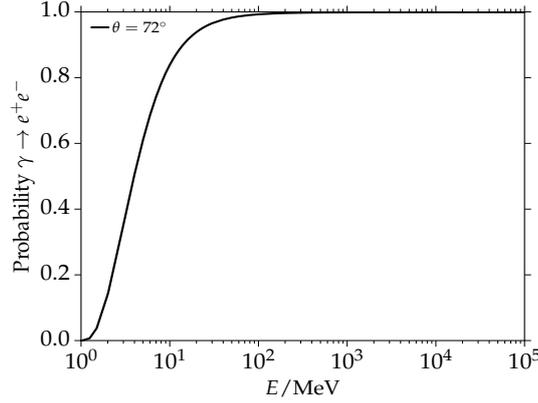


Figure C.4: Probability of a photon at $\theta_\gamma = 72^\circ$ to convert into a e^+e^- pair as a function of the energy.

Production of photoelectrons

The photomultiplier tubes collect the Cherenkov photons emitted by charged particles when passing through the water volume inside the tank. When one of this photons hits the surface of the PMT, it will interact by photoelectric effect with its sensitive part, the photocathode, resulting the release of an electron, called *photoelectron*. In this section we discuss an analytical approach for the production of photoelectrons.

The number of photoelectrons, N_{pe} , will be given by the number of Cherenkov photons that arrive to the PMT multiplied by the quantum efficiency for traducing this light into photoelectrons. This arises from the fact that not all of the arriving photons will release an electron in the photocathode. The quantum efficiency is then defined as

$$q(E) = \frac{\text{number of emitted photoelectrons}}{\text{number of arriving photons}}. \quad (\text{C.13})$$

Since it depends on the photon energy, we have to integrate over energies to get the number of photoelectrons

$$N_{pe} = \int_0^\infty q(E) \left(\frac{dN}{dE} \right)_a dE. \quad (\text{C.14})$$

When a photon is emitted in the tank, it will undergo different processes that will difficult its arrival to the photomultiplier surface. In particular, photons will bounce in the tank walls, where some of them will get absorbed since the Tyvek is not a perfect reflector, i.e., reflectivity R is not 100 percent. Additionally, during their journey to the PMT, some photons will get absorbed due to collisions in water. This is determined by the absorption length, λ .

The number of photons that arrive to the PMT is given by

$$\left(\frac{dN}{dE} \right)_a = \mathcal{L} \times \left(\frac{dN}{dE} \right)_e, \quad (\text{C.15})$$

where \mathcal{L} is the *loss* factor that accounts for the rate of absorption, and $\left(\frac{dN}{dE} \right)_e$ is the number of *emitted* photons per unit energy interval, given by Eq. (C.7). In order to determine an analytical expression for \mathcal{L} we will assume that the Cherenkov light is emitted as a gas of photons traveling in all directions in the water volume. Let us call $\langle l \rangle$ to the mean free path. After a number m of reflections, photons will have traveled a distance $l = m \langle l \rangle$,

while the absorption due to reflections and collisions will be given by $R^m \exp(-m \langle l \rangle / \lambda)$. The probability of a photon to hit the photomultiplier after m reflections on the walls will be determined by the ratio between the area occupied by the PMTs and the total area of the tank $p = 3A_{\text{PMT}} / A_{\text{tank}}$, and is given by a binomial distribution

$$P(m | p) = \binom{m}{1} p(1-p)^{m-1} = \frac{m!}{(m-1)!} p(1-p)^{m-1}. \quad (\text{C.16})$$

The loss factor can be computed as a weighted mean of the absorption with weights given by Eq. (C.16)

$$\mathfrak{L}(E) = \frac{\sum_m R^m(E) e^{-m \langle l \rangle / \lambda(E)} \times P(m | p)}{\sum_m P(m | p)}, \quad (\text{C.17})$$

where we have included the dependency of the reflectivity and the attenuation length on the energy of the emitted photons. This will give us the fraction of Cherenkov photons that survive the journey to the photomultiplier tube since they were emitted.

The number of photoelectrons produced in the PMTs of the water-Cherenkov detector is given by the following expression

$$N_{\text{pe}} = \frac{\alpha}{\hbar c} L \left(1 - \frac{1}{\beta^2 n^2} \right) \int q(E) \mathfrak{L}(E) dE, \quad (\text{C.18})$$

where we have assumed that the index of refraction does not depend on the photon energy.

The form of $q(E)$, $R(E)$ and $\lambda(E)$ complicate the integral in Eq. (C.18). We introduce a purely empirical model to fit the signal produced in the WCD as a function of the particle energy. This model consists in two parts, the first one describes the signal due to the emission of Cherenkov photons, which we know that saturates to a certain constant value above energies of ≈ 1 GeV. This value depends on the track length of the particle in water. The second part of the module accounts for the contribution of *knock-on electrons* (δ rays), which production is stochastic and might be energetic enough to emit Cherenkov light. The model is represented in Eq. (C.19) and it fits very well the data obtained from simulations, as we can see in Fig. C.5b.

$$N_{\text{pe}}(p) = \frac{A_0}{1 + \exp\left(\frac{p-p_0}{\alpha}\right)} + A_1 \arctan(p). \quad (\text{C.19})$$

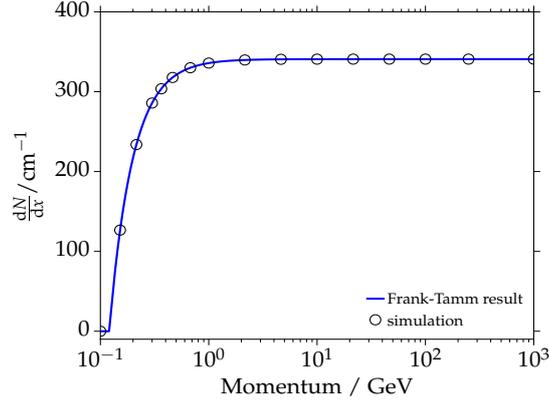
C.2 Chapter 3

C.2.1 Signal uncertainty model

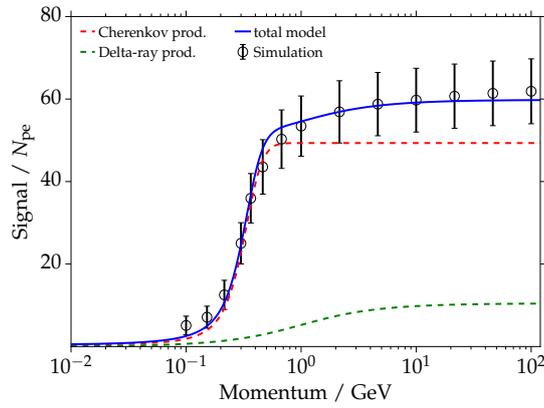
The signal variance of the SSD signals can be expressed as

$$\sigma_S = f(\theta) \sqrt{S},$$

where $f(\theta)$ is a linear function in $\sec \theta$ which takes in to account the attenuation of particles in the atmosphere. The variance of the SSD signals is shown as a function of the signal, for different zenith angles in Fig. C.6. The line corresponds to the parameterization obtained in Section 3.2.2.



(a)



(b)

Figure C.5: (a) Number of Cherenkov photons produced by muons in water as a function of the momentum of the muon. Data points indicate the results from the simulation and the line shows the prediction by the Frank-Tamm formula (Eq. (C.5)). (b) Production of photoelectrons in water-Cherenkov detector by muon as a function of the momentum of the muon. The data from simulations is compared to the model derived in Eq. (C.18).

a	b	c	d	e	f	g	h	i
303.2	-355.6	171.7	586.4	1309.5	-544.9	156.5	-558.2	236.2

C.2.2 Distance cut for rejecting low signals

A distance cut was introduced in the fit of the SSD LDF in order to keep all stations below a certain distance, r_{cut} , in such a way that 95% of the SSDs below that distance have at least a signal of 1 MIP. This criteria is shown in Fig. C.7

The distance cut was parameterized in terms of the shower-size and the zenith angle as

$$\begin{aligned}
 r_{\text{cut}} = & a + b \sec \theta + c \sec^2 \theta \\
 & + (d + e \sec \theta + f \sec^2 \theta) s \\
 & + (g + h \sec \theta + i \sec^2 \theta) s^2,
 \end{aligned}$$

where $s = \lg S_{1000}$ of the WCD. The optimized parameters are

Examples of the parameterization of the distance cut can be seen in Fig. C.8

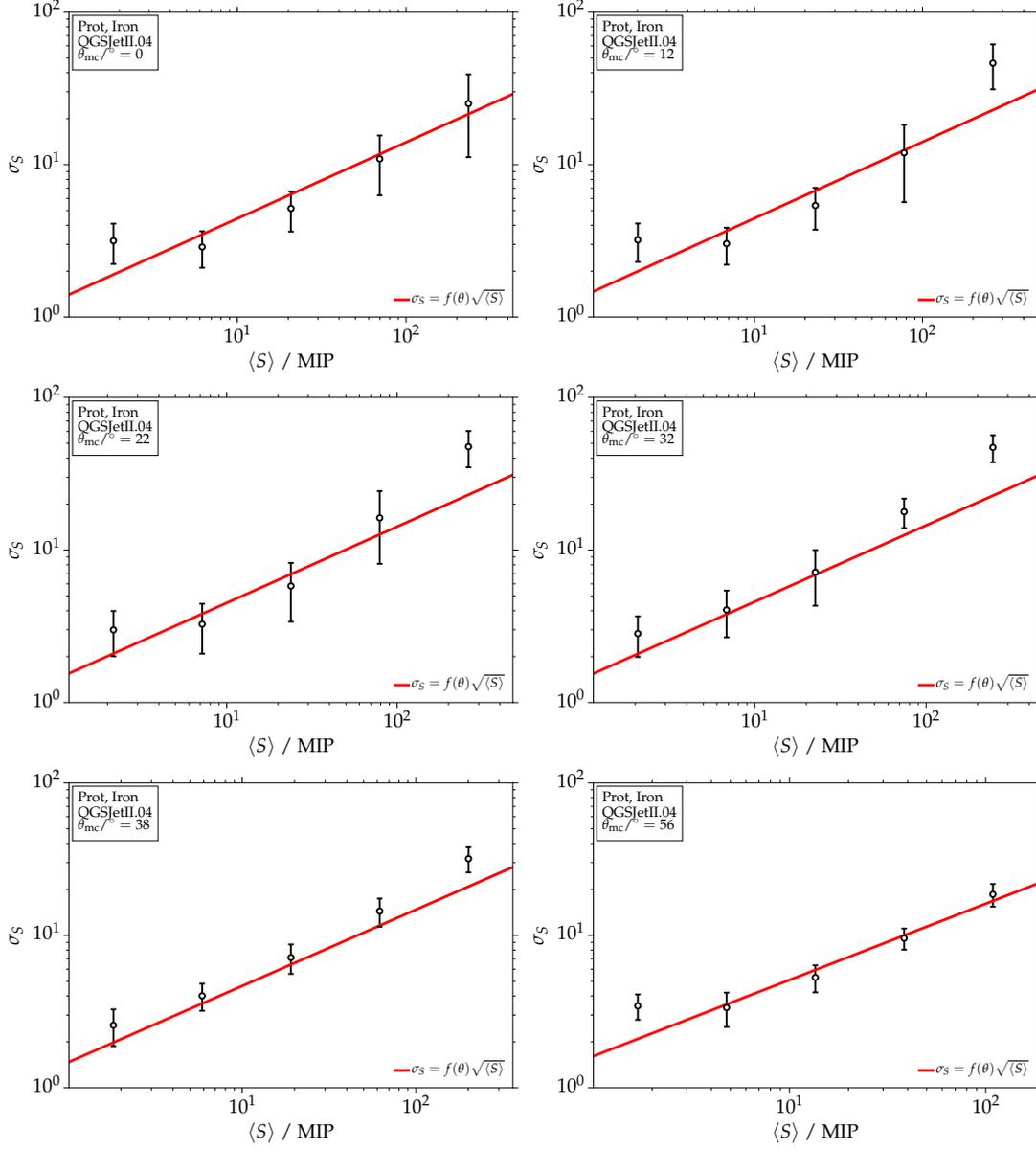


Figure C.6: SSD signal variance as a function of signal for different zenith angles.

C.2.3 Parameterization of the LDF for the SSD

With the current functional form given in Eq. (3.13), β turns into positive values for large zenith angles. In order to avoid this non-physical regime the following functional form was proposed

$$\beta = (\beta_0 + \beta_1 s) \exp(-\beta_2(\sec \theta - \sec \theta_0)^2). \quad (\text{C.20})$$

A comparison of the two models is shown in Fig. C.10

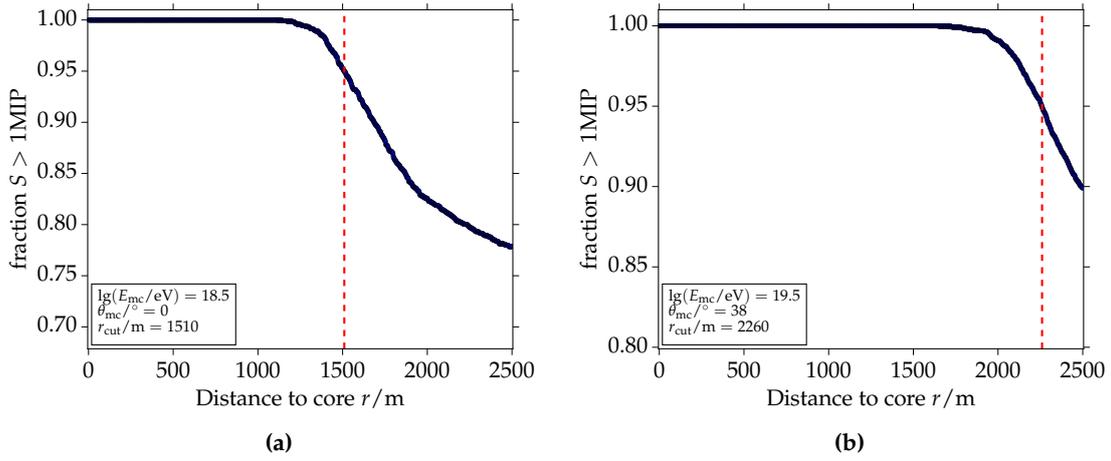


Figure C.7: Fraction of SSD signals above 1 MIP as a function of distance for showers of $10^{18.5}$ eV and 0° (Fig. C.7a) and of $10^{19.5}$ eV and 38° (Fig. C.7b). The cut in distance is shown by the red dashed line.

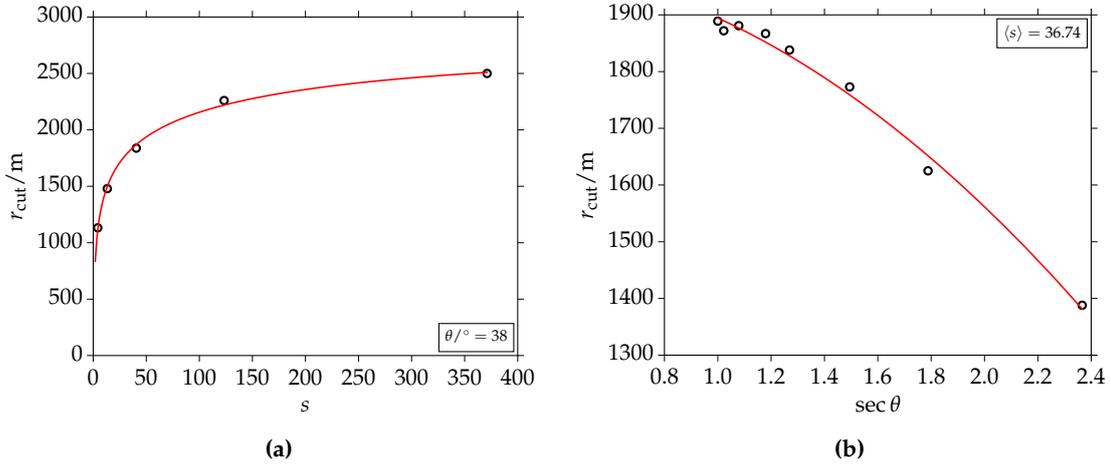


Figure C.8: Fig. C.8a: Distance cut as a function of the WCD shower size for showers with zenith angle of 38° . Fig. C.8b: Distance cut as a function of $\sec \theta$ for showers of 10^{19} eV of energy.

C.2.4 Validation of the parameterization

An essential check of the parameterization of the LDF derived in Section 3.2.3 is the residuals, i.e.,

$$\text{Res}(S_i) := \frac{S_i - \hat{S}_i}{\sigma[\hat{S}_i]}.$$

The residuals are shown as a function of the distance to the shower core in Fig. C.11 and as a function of the SSD signal in Fig. C.12, for all showers with energy above 10^{19} eV and for different zenith angles.

The following quantity we checked is the bias in the reconstructed shower size which is defined by

$$b = \frac{S_{1000}^{\text{rec}}}{\hat{S}_{1000}^{\text{mc}}} - 1,$$

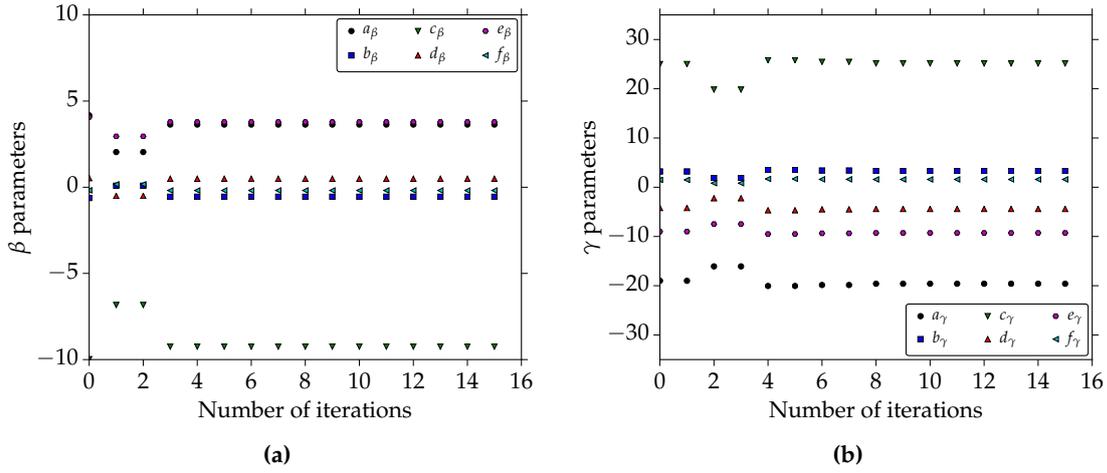


Figure C.9: Evolution of the coefficients of the β (Fig. C.9a) and γ (Fig. C.9b) parameterization with the number of iterations.

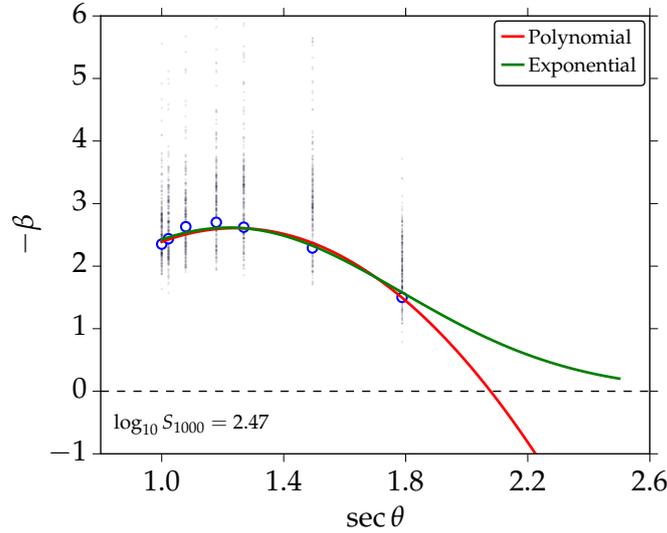


Figure C.10: Comparison between the current model use for the parameterization of β given by Eq. (3.13) (red line) and a model using an exponential form as in Eq. (C.20) (green line).

being $\hat{S}_{1000}^{\text{mc}}$ the SSD signal as obtained from the average of the dense ring of stations. The bias in S_{1000}^{rec} for proton and iron showers is shown as a function of the energy in Fig. C.13. Different plots correspond to different zenith angles.

C.2.5 Non-linear ansatz for mass-independent energy bias

In this section a different approach to the one used in Section 3.3 in order to determine an unbiased energy estimate is presented. The following non-linear ansatz for the energy was used

$$\lg E = A \left(\lg S_{38}^{\text{wcd}} \right)^\alpha + B \left(\lg S_{38}^{\text{ssd}} \right)^\beta. \quad (\text{C.21})$$

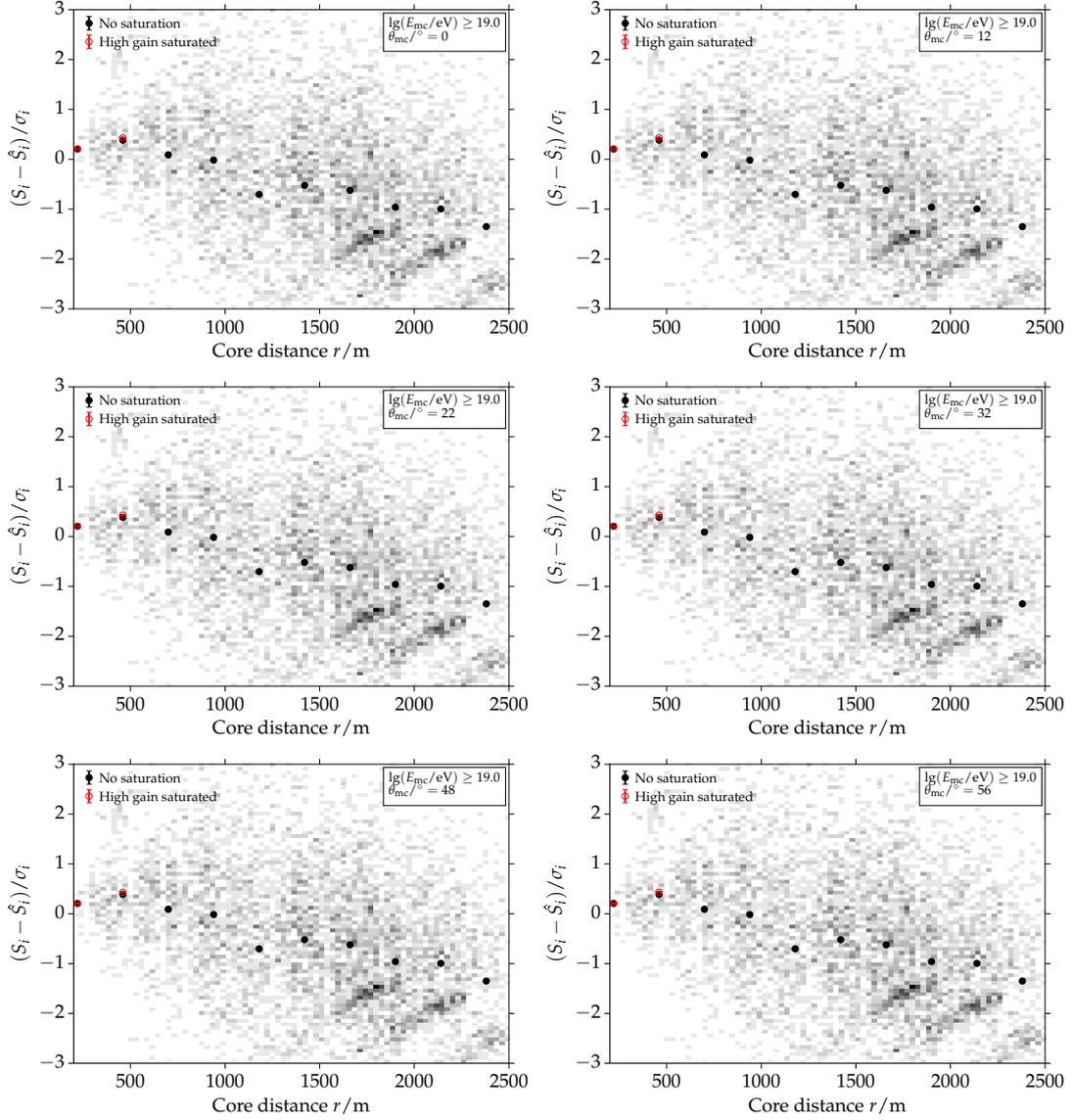


Figure C.11: Residuals of the SSD LDF as a function of distance to the shower core for all zenith angles.

The energy of each event (same proton and iron showers as in Section 3.3 were used) was fitted to Eq. (C.21) by minimizing the following χ^2

$$\chi^2 = \sum \frac{(\lg E_{mc} - \lg \hat{E})^2}{\sigma(\lg \hat{E})^2}, \quad (\text{C.22})$$

where $\lg \hat{E}$ represents the model in Eq. (C.21) and $\sigma(\lg \hat{E})$ is the uncertainty on the predicted logarithm of the energy and it was calculated using the formula for the error-propagation

$$\sigma(\lg E)^2 = \left(\frac{\partial(\lg E)}{\partial(S_{38}^{wcd})} \right)^2 (\sigma(S_{38}^{wcd}))^2 + \left(\frac{\partial(\lg E)}{\partial(S_{38}^{ssd})} \right)^2 (\sigma(S_{38}^{ssd}))^2 \quad (\text{C.23})$$

The data used for the fit can be seen in Fig. C.14.

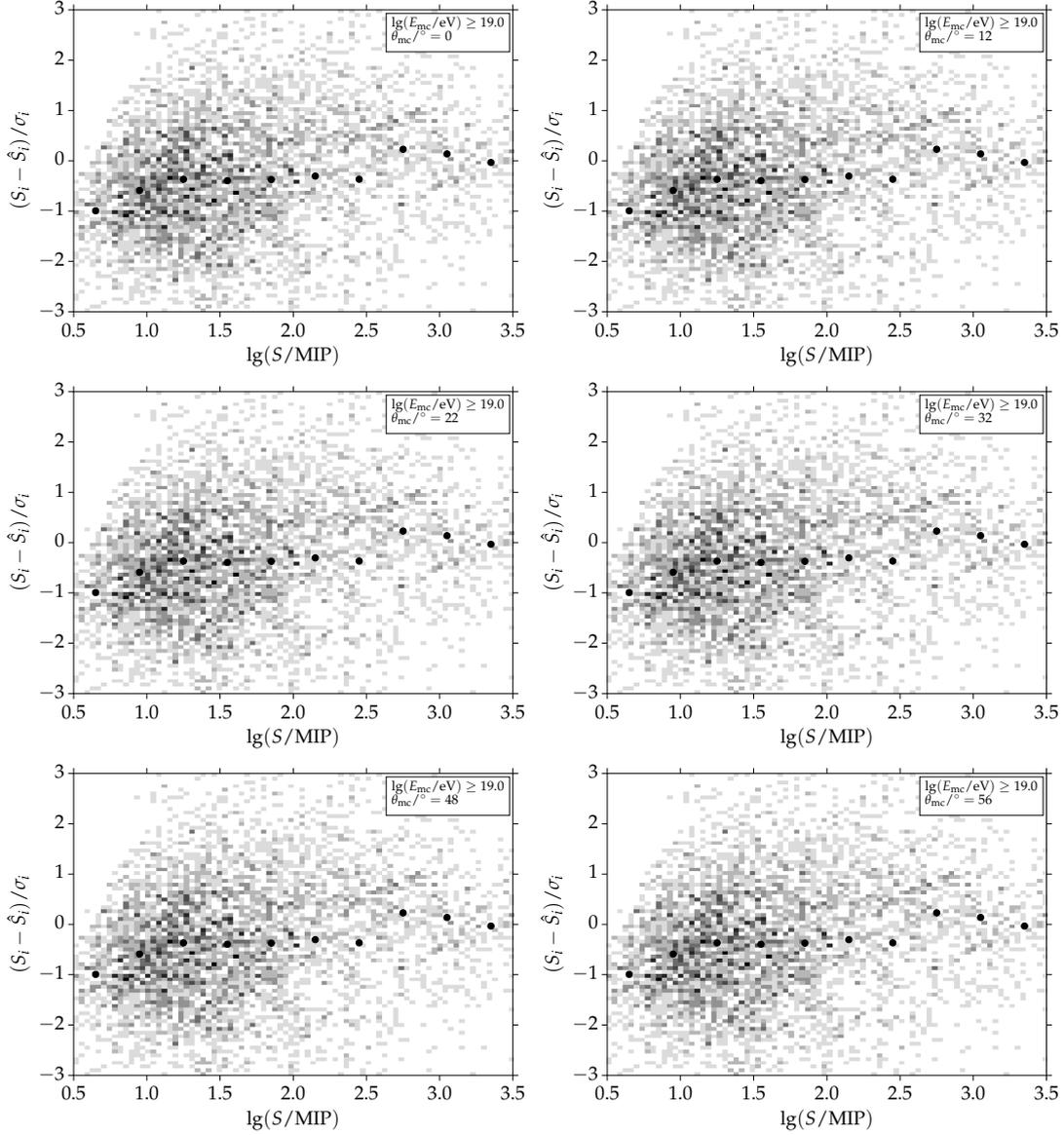


Figure C.12: Residuals of the SSD LDF as a function of the signal in the SSD for all zenith angles.

The minimization was performed using NLoft giving the following result:

$$\lg E = 0.2408 \left(\lg S_{38}^{\text{wcd}} \right)^{0.9825} - 0.0192 \left(\lg S_{38}^{\text{ssd}} \right)^{1.0213}. \quad (\text{C.24})$$

The fit residuals are computed as $(E_{\text{rec}} - E_{\text{mc}}) / E_{\text{mc}}$ and are shown in Fig. C.15. It can be seen that, on average, we are able to reduce the bias. However the bias due to the mass of the primary cannot be reduced. This can be seen in Fig. C.16.

C.2.6 Uncertainties on the merit factor

We used the bootstrap method (or *bootstrapping*) to calculate the uncertainties in the merit factor. This method is a statistical technique that allows us to estimate quantities of a dataset (e.g., its variance) by computing the average of small samples of the dataset. The process can be summarized in different steps

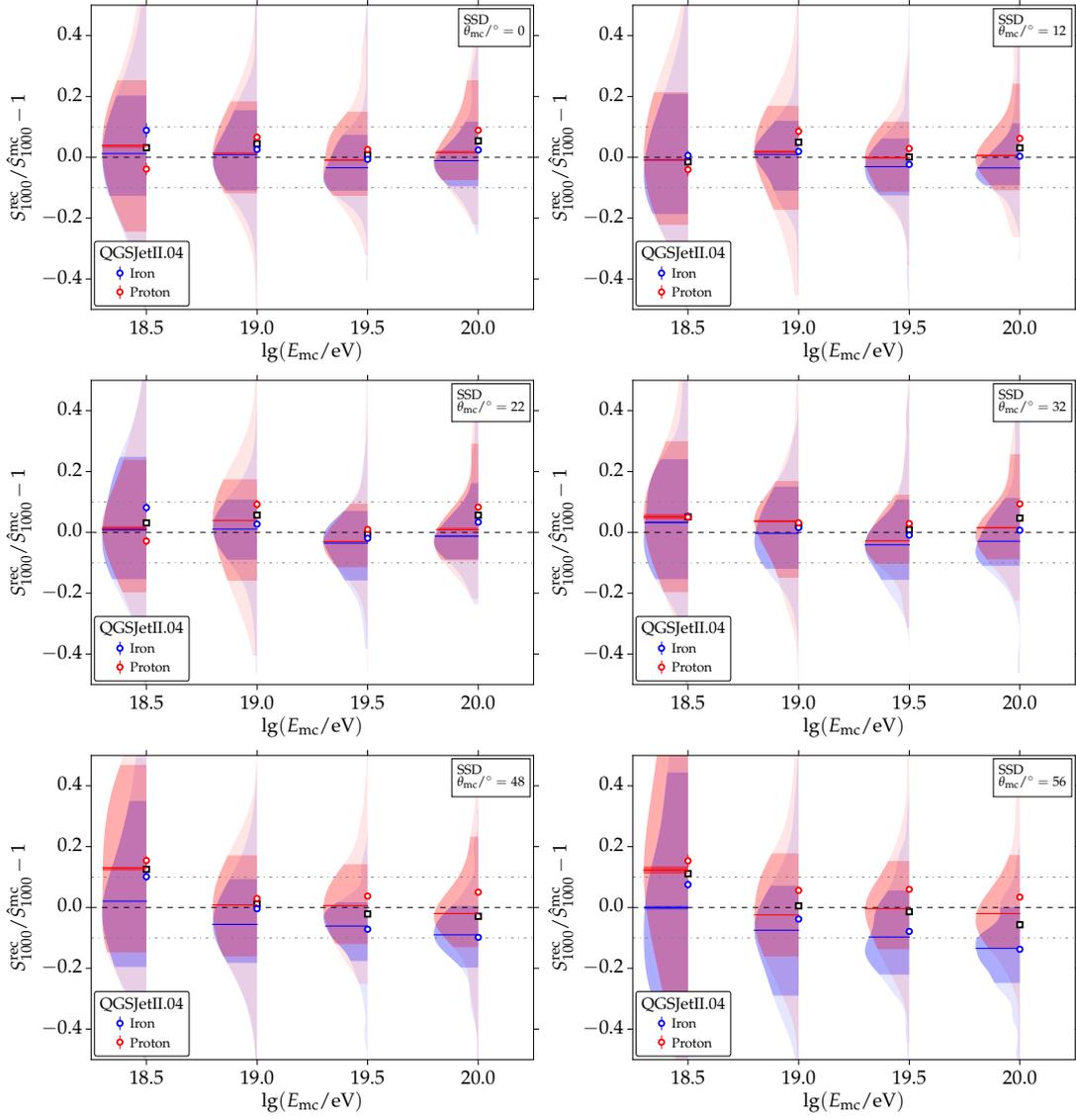


Figure C.13: Bias in S_{1000}^{ssd} as a function of energy for proton (red) and iron (blue) showers, for different zenith angles. Black markers show the average bias between both primaries.

1. Choose the size, n , of the sample ($n \leq N$ being N the size of the large dataset).
2. Draw a sample of size n by taking random observations of the large dataset.
3. Compute the mean of the sample.
4. Repeat the process k times.

An important fact is that samples are computed by taking random observations of the large dataset but it is allowed to include a given observation more than once. This is called sampling with replacement. The new dataset created from the mean of k samples resembles the original dataset. An example is shown in Fig. C.17a. The original distributions of proton and iron observables are indicated by solid lines while the distributions using bootstrapping are indicated by a dashed line. Now, this process can be applied several times in order to get “copies”, in this case, of the distributions of \mathcal{M} for proton and iron and compute a value of the merit factor on each iteration. An example distribution

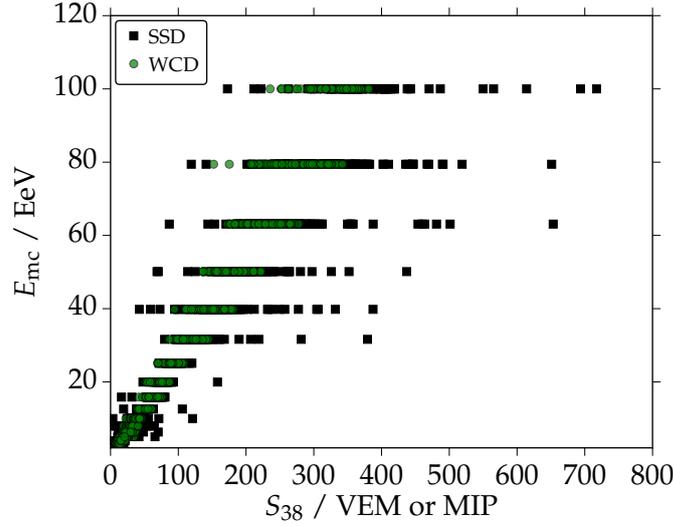


Figure C.14: Monte Carlo energy as a function of S_{38} from the WCD (green circles) and SSD (black squares) used for the fit of Eq. (C.21).

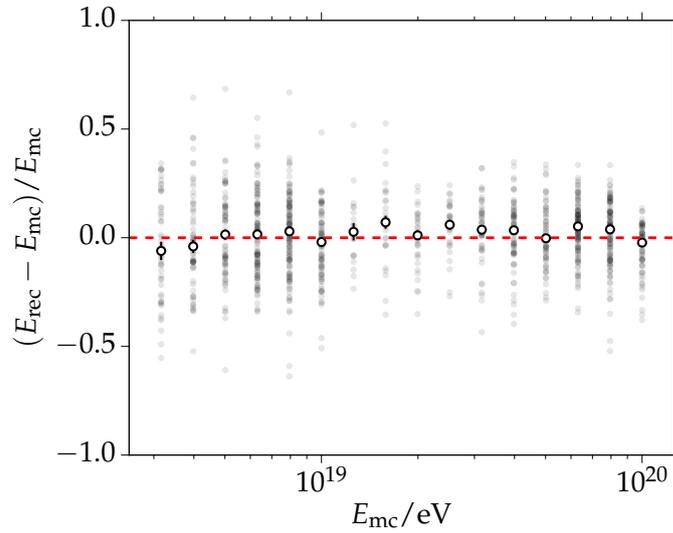


Figure C.15: Residuals of the fit to Eq. (C.21). The shaded scatter dots the individual residuals and the markers show the mean on each bin.

of merit factors as obtained from 500 iterations is shown in Fig. C.17b. The mean of the distribution is shown by the solid line giving a value of 1.31. This value is identical to the one obtained from the original distribution. Using the bootstrap method we obtain a distribution of the merit factor corresponding to a single observation. One way to estimate the statistical uncertainty in the merit factor is taking the RMS of the distribution.

C.3 Chapter 4

C.3.1 Mean composition

The distribution of $\ln A'_{\text{pca}}$ from each of the individual energy bins used to make Fig. 4.22 are shown in Fig. C.18.

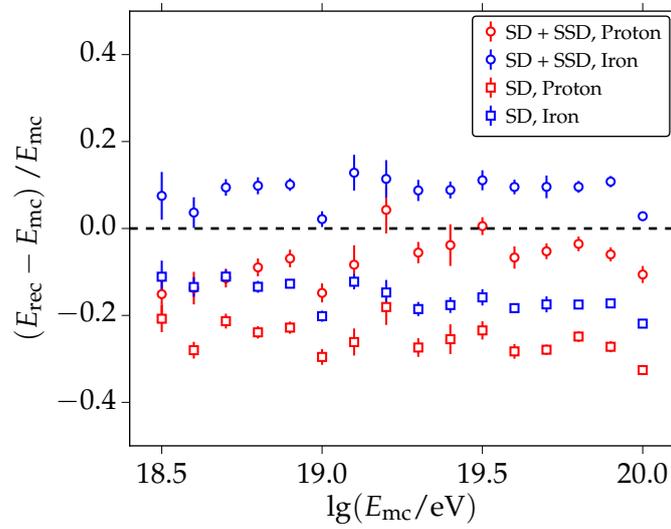


Figure C.16: Bias in the reconstructed energy for proton and iron showers using the Non-linear ansatz (circles). For comparison the same bias but using the reconstructed energy with SD (squares).

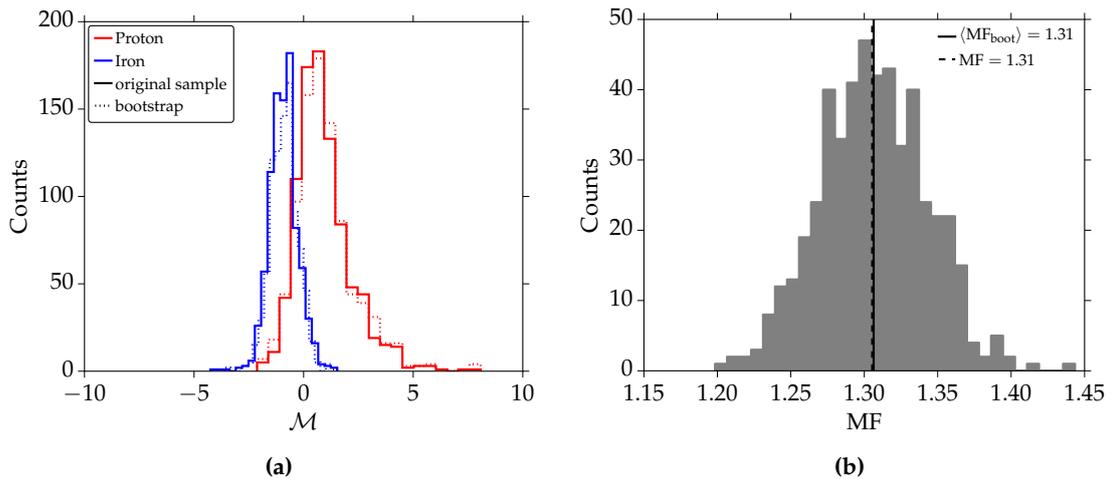


Figure C.17: Helpful figures to explain bootstrap method applied to merit factor. See text for details.

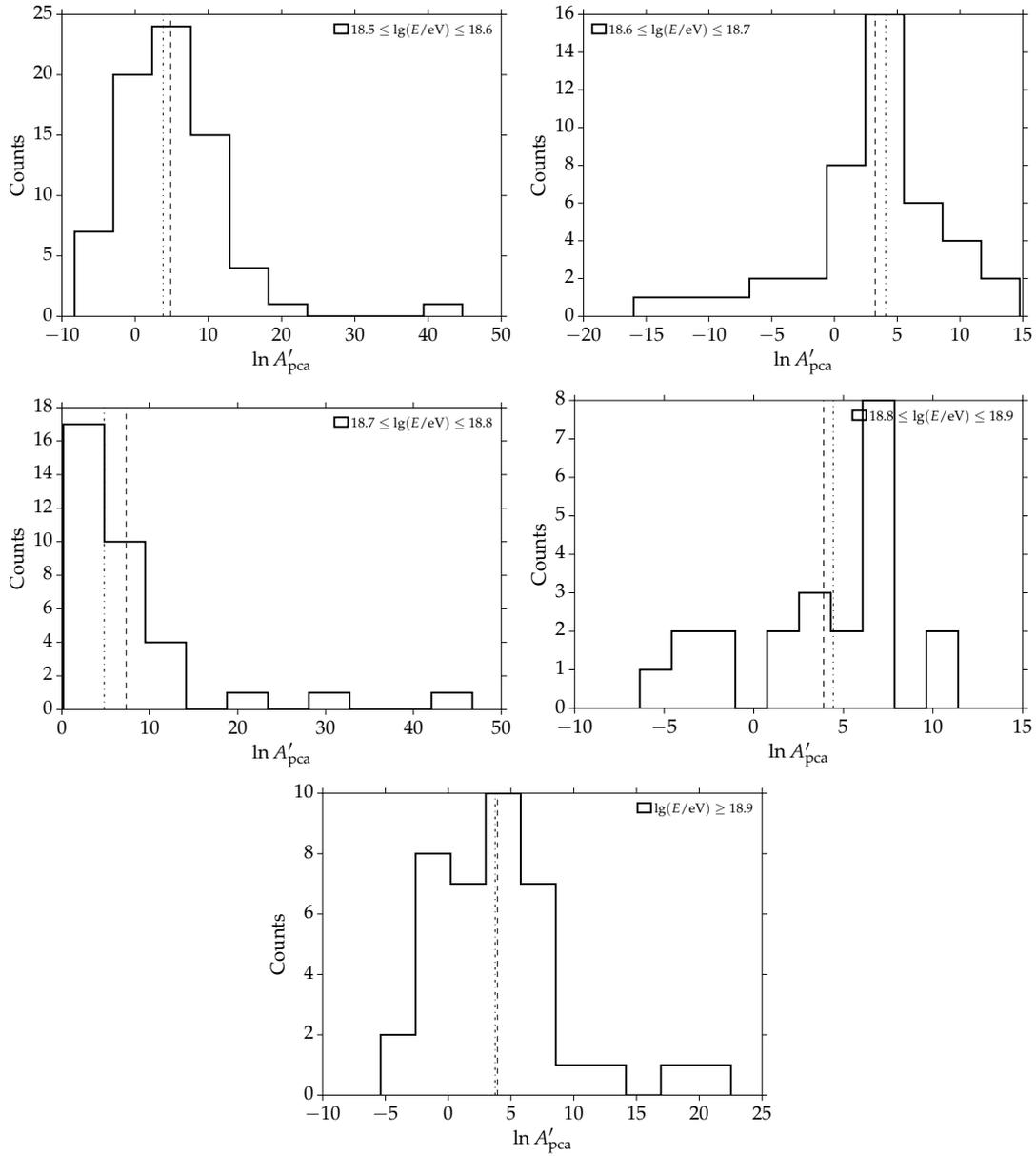


Figure C.18: Distributions of $\ln A'_{pca}$ in different energy bins.

APPENDIX D

Acknowledgments

I would like to thank Prof. Johannes Blümer and Prof. Alberto Etchegoyen, for reviewing this thesis and for founding the Double Doctoral Degree in Astrophysics between Germany and Argentina. Such a program has enriched me with many professional but also personal experiences. I want to thank my supervisor Dr. Markus Roth, for your constant support and guidance. Thanks also to Sabine Bucher and to Marie-Christine Mundt, for all the time you have spent helping me with all the logistics and bureaucratic stuff.

I would like to thank all my colleagues at IKP (also the ones in the “not-that-cool” side of the corridor), for creating such a nice working atmosphere. Thanks to Alex Schulz, Ariel and Daniela for helping me in the beginning of my thesis. To Ana Martina and Nico for the nice welcome and the first months in Germany. As I know I will forget people, I will thank in general, to all the mates I had the pleasure to share office with. In particular to Ana Laura, Johan, Max and Marcel. Thanks to Martin for all your help and fruitful discussions ;) *Humongous* thanks to Darko for all your help and support during these years. Same to Alex Streich, my longest-term officemate, for being such a nice company and willing to help. Thank you David, you know I am not good expressing things in English but you also know that this work wouldn't be possible without your help, strong support and advice. For all the many conversations we had over these years and the good moments. Finally, I would like to thank to my brothers from another mother, Álvaro and Arnote. Thanks for such a friendship, and all the good moments we spent together.

Argentina... En primer lugar me gustaría agradecer a todo el personal de ITeDA. He tenido el gusto de poder trabajar con muchos de vosotros, sin embargo me gustaría resaltar el grupo humano que formáis. Empezando por la gente de secretaría, gracias por haberme tratado siempre con tanta amabilidad. Agradecer a todo el cuerpo de investigadores, en especial a Federico Sánchez y a Hernán Asorey, por vuestros consejos y por estar siempre dispuestos a echarme una mano en lo que fuera necesario. A todos mis compañeros de oficina y colegas “físicos”, Ana, Nico, Belén, Perlin, Marina, Mauricio, Rolando, Gaia, Christian... y a muchos más que seguro me estoy olvidando. Gracias a toda la gente de la C-101, por toda la buena onda, en especial a Lucho, a Manu, a Juan, al Diegote y a Alan por tantas conversaciones sobre música y fútbol (River no es el más grande pero la experiencia es inolvidable :) Gracias también a Mati y a Alejandro, por ayudarme siempre con los temas de electrónica ;) Gracias a Carlos, a Sebastián y al genio Adrián por toda vuestra ayuda en mi entrañable relación con la computación... Gracias a

todos los compañeros de ITeDA fútbol y bueno, ya que estamos, gracias al fútbol. Mil gracias a Mariela, por tu forma de ser, por ayudarme siempre e integrarme desde el primer día! Gracias a mis compañeros de piso, con quienes guardo una gran amistad, Christian y Laura, Gaia y Sebastián. Todos hicisteis que este *gallego* se haya sentido siempre como en casa.

A tantas otras personas, que de algún modo u otro se fueron cruzando en mi camino, gracias por todo el cariño.

Agradecervos a todos os que estivestes sempre. Aos meus amigos da universidade, con quen de algunha maneira comencei esta traxectoria, en especial a Alexis, Pablo, Tato, Aida, Alejandro e Don Carlos; e aos meus amigos de Ourense, Casco, Juan, Diego y a *Los Canelitas*, Hugo e Fer. Finalmente, grazas de todo corazón á miña familia, en especial aos meus pais Elena e Rafa, e á miña irmá Belén, polo voso apoio incondicional durante todo este tempo. Sen vos esto non fora posible.

Bibliography

- [1] T. K. Gaisser, R. Engel, E. Resconi, *Cosmic Rays and Particle Physics*, 2nd Edition, Cambridge University Press, 2016. doi:[10.1017/CB09781139192194](https://doi.org/10.1017/CB09781139192194).
- [2] T. Antoni *et al.*, KASCADE measurements of energy spectra for elemental groups of cosmic rays: Results and open problems 24 (1-2) (2005) 1 – 25. doi:[10.1016/j.astropartphys.2005.04.001](https://doi.org/10.1016/j.astropartphys.2005.04.001).
- [3] W. D. Apel *et al.*, KASCADE-Grande Collaboration, Kneelike Structure in the Spectrum of the Heavy Component of Cosmic Rays Observed with KASCADE-Grande, *Phys. Rev. Lett.* 107 (2011) 171104. doi:[10.1103/PhysRevLett.107.171104](https://doi.org/10.1103/PhysRevLett.107.171104).
- [4] V. Berezhinsky, Transition from galactic to extragalactic cosmic rays, in: *Proc. 30th Int. Cosmic Ray Conf., Mérida, Mexico, 2007*, arXiv astro-ph/0710.2750.
- [5] A. M. Hillas, *Cosmic Rays: Recent Progress and some Current Questions*, arXiv:astro-ph/0607109 (2006). arXiv:[arXiv:astro-ph/0607109](https://arxiv.org/abs/astro-ph/0607109).
- [6] M. Unger, G. R. Farrar, L. A. Anchordoqui, Origin of the ankle in the ultra-high energy cosmic ray spectrum and of the extragalactic protons below it (2015) arXiv:[1505.02153](https://arxiv.org/abs/1505.02153).
- [7] D. Allard, E. Parizot, A. Olinto, On the transition from galactic to extragalactic cosmic-rays: spectral and composition features from two opposite scenarios, *Astropart. Phys.* 27 (2007) 61 – 75.
- [8] K. Greisen, End to the cosmic-ray spectrum?, *Phys. Rev. Lett.* 16 (17) (1966) 748–750. doi:[10.1103/PhysRevLett.16.748](https://doi.org/10.1103/PhysRevLett.16.748).
- [9] G. T. Zatsepin, V. A. Kuzmin, Upper limit of the spectrum of cosmic rays, *JETP Lett.* 4 (1966) 78–80, [*Pisma Zh. Eksp. Teor. Fiz.* 4,114(1966)].
- [10] J. W. Cronin, The highest-energy cosmic rays, *Nuclear Physics B - Proceedings Supplements* 138 (2005) 465 – 491. doi:<http://dx.doi.org/10.1016/j.nuclphysbps.2004.11.107>.
- [11] D. Allard, E. Parizot, A. V. Olinto, E. Khan, S. Goriely, UHE nuclei propagation and the interpretation of the ankle in the cosmic-ray spectrum, *Astroparticle Physics* 443 (2005) L29–L32. doi:[10.1051/0004-6361:200500199](https://doi.org/10.1051/0004-6361:200500199).

- [12] K. Kotera, A. V. Olinto, The astrophysics of ultrahigh-energy cosmic rays, *Annual Review of Astronomy and Astrophysics* 49 (1) (2011) 119–153. doi:10.1146/annurev-astro-081710-102620.
- [13] A. Bridgeman, Determining the Mass Composition of Ultra-high Energy Cosmic Rays Using Air Shower Universality, Dissertation, Karlsruhe Institute of Technology (2018).
- [14] W. Heitler, *The Quantum Theory of Radiation*, third ed. Edition, Oxford University Press, Oxford, 1954.
- [15] J. Matthews, A heitler model of extensive air showers, *Astropart. Phys.* 22 (5-6) (2005) 387–397.
- [16] R. Engel, D. Heck, T. Pierog, Extensive air showers and hadronic interactions at high energy, *Ann. Rev. Nucl. Part. Sci.* 61 (2011) 467–489. doi:10.1146/annurev.nucl.012809.104544.
- [17] P. Abreu, et al., Pierre Auger Collaboration, Interpretation of the Depths of Maximum of Extensive Air Showers Measured by the Pierre Auger Observatory, *JCAP* 1302 (2013) 026. doi:10.1088/1475-7516/2013/02/026.
- [18] K.-H. Kampert, M. Unger, Measurements of the cosmic ray composition with air shower measurements, arXiv:1201.0018 [astro-ph.HE] (2012).
- [19] A. Aab, et al., Pierre Auger, The Pierre Auger Cosmic Ray Observatory, submitted to *Nucl.Instrum.Meth. A* (2015). arXiv:1502.01323.
- [20] J. Abraham *et al.* (Pierre Auger Collab.), Properties and performance of the prototype instrument for the pierre auger observatory, *Nucl. Instrum. Meth. A* 523 (1-2) (2004) 50 – 95. doi:DOI:10.1016/j.nima.2003.12.012.
- [21] I. Allekotte, et al., Pierre Auger, The Surface Detector System of the Pierre Auger Observatory, *Nucl.Instrum.Meth. A* 586 (2008) 409–420. doi:10.1016/j.nima.2007.12.016.
- [22] J. Abraham *et al.* (Pierre Auger Collab.), The fluorescence detector of the Pierre Auger Observatory, *Nucl. Instrum. Meth. A* 620 (2010) 227 – 251.
- [23] J. Abraham *et al.* (Pierre Auger Collab.), Trigger and aperture of the surface detector array of the Pierre Auger Observatory, *Nucl. Instrum. Meth. A* 613 (2010) 29–39.
- [24] B. Keilhauer, Investigation of atmospheric effects on the development of extensive air showers and their detection with the pierre auger observatory, Ph.D. thesis, Karlsruhe University, Karlsruhe, Germany (2003).
- [25] D. Mockler *et al.* (Pierre Auger Collab.), Reconstruction of Vertical Events Recorded by the Surface Detector of the Pierre Auger Observatory, in: *Proc. 36th Int. Cosmic Ray Conf.*, Madison, Wisconsin, USA, 2019.
- [26] X. Bertou, P. Allison, C. Bonifazi, P. Bauleo, C. M. Grunfeld, M. Aglietta, F. Arneodo, D. Barnhill, J. Beatty, N. Busca, A. Creusot, D. Dornic, A. Etchegoyen, A. Filevitch, P. L. Ghia, I. Lhenry-Yvon, M. C. Medina, E. Moreno, D. Nitz, T. Ohnuki, S. Ranchon, H. Salazar, T. Suomijarvi, D. Supanitsky, A. Tripathi, M. Urban, L. Villasenor, Pierre Auger Collaboration, Calibration of the surface array of the Pierre Auger Observatory, *Nucl. Instrum. Meth. A* 568 (2006) 839–846.

- [27] D. Mockler, The first measurement of an energy spectrum at 0.1 EeV with the Surface Detector of the Pierre Auger Observatory, Master's thesis, Karlsruhe Institute of Technology, Karlsruhe (10 2014).
- [28] S. Messina, Extension to lower energies of the cosmic-ray energy window at the Pierre Auger Observatory, Dissertation, University of Groningen (2016).
- [29] H. O. Klages, Heat - enhancement telescopes for the pierre auger southern observatory, in: Proc. 30th Int. Cosmic Ray Conf., Mérida, Mexico, 2007.
- [30] B. Dawson *et al.* (Pierre Auger Collab.), The Energy Scale of the Pierre Auger Observatory, in: Proc. 36th Int. Cosmic Ray Conf., Madison, Wisconsin, USA, 2019.
- [31] Jose Bellido *et al.* (Pierre Auger Collab.), Depth of maximum of air-shower profiles at the Pierre Auger Observatory: Measurements above $10^{17.2}$ eV and composition implications, in: Proc. 35th ICRC, (2017) 506, arXiv:1708.06592, Busan, South Korea, 2017.
- [32] F. Suarez *et al.* (Pierre Auger Collab.), The AMIGA muon detectors of the Pierre Auger Observatory, in: Proc. 33rd Int. Cosmic Ray Conf., Rio de Janeiro, Brazil, 2013.
- [33] S. Müller, Measurement of the cosmic ray composition with air showers detected by the AMIGA extension at the pierre auger observatory, Dissertation, Karlsruhe Institute of Technology (2018).
- [34] A. Aab, Prototype muon detectors for the amiga component of the pierre auger observatory, Journal of Instrumentation 11 (02) (2016) P02012.
- [35] A. Taboada, Preliminary results of the amiga engineering array at the pierre auger observatory, EPJ Web of Conferences 210 (2019) 02016. doi:10.1051/epjconf/201921002016.
- [36] S. Müller, Direct measurement of the muon density in air showers with the pierre auger observatory, EPJ Web of Conferences 210 (2019) 02013. doi:10.1051/epjconf/201921002013.
- [37] A. M. Botti *et al.* (Pierre Auger Collab.), The AMIGA underground muon detector of the Pierre Auger Observatory - performance and event reconstruction, in: Proc. 36th Int. Cosmic Ray Conf., Madison, Wisconsin, USA, 2019.
- [38] A. Aab, et al., Pierre Auger Collaboration, Probing the radio emission from air showers with polarization measurements, Phys.Rev. D89 (2014) 052002. doi:10.1103/PhysRevD.89.052002.
- [39] A. Aab, et al., Pierre Auger, Energy Estimation of Cosmic Rays with the Engineering Radio Array of the Pierre Auger Observatory, Phys. Rev. D93 (12) (2016) 122005. doi:10.1103/PhysRevD.93.122005.
- [40] E. M. Holt, Combined detection of muons and radio emission of cosmic-ray air showers, Dissertation, Karlsruhe Institute of Technology (2018).
- [41] D. Veberič, M. Roth, Offline reference manual: Sd reconstruction, Auger internal note GAP-2005-035, an updated version is shipped with the Offline software distribution (2005).

- [42] A. Schulz, Measurement of the Energy Spectrum and Mass Composition of Ultra-high Energy Cosmic Rays, Dissertation, Karlsruhe Institute of Technology (2016).
- [43] D. Mockler, Measurement of the Cosmic Ray Spectrum with the Pierre Auger Observatory, Dissertation, Karlsruhe Institute of Technology (2018).
- [44] L. M. Bueno, P. Billoir, I. C. Maris, Signal variance for the TOTd and MoPS triggers, Auger internal note GAP-2014-035 (2014).
- [45] M. Ave, P. Bauleo, A. Castellina, A. Chou, J. Harton, R. Knapik, G. Navarra, The accuracy of signal measurement of the water Cherenkov detectors of the Pierre Auger Observatory, Nucl. Instrum. Meth. A578 (2007) 180 – 184.
- [46] Greisen, Progress in elementary particle and cosmic ray physics, Vol. 3, North-Holland Publishing, Amsterdam, 1956.
- [47] K. Kamata, J. Nishimura, The Lateral and the Angular Structure Functions of Electron Showers, Prog. Theoret. Phys. Suppl. 6 (1958) 93–155.
- [48] D. Newton, J. Knapp, A. A. Watson, The Optimum Distance at which to Determine the Size of a Giant Air Shower, Astropart. Phys. 26 (2007) 414–419.
- [49] M. Aglietta, I. De Mitri, S. Maglio, S. Maldera, I. C. Maris, D. Martello, G. Navarra, M. Roth, Recovery of Saturated Signals of the Surface Detector, Auger internal note GAP-2008-30 (2008).
- [50] D. Veberič *et al.* (Pierre Auger Collab.), Estimation of Signal in Saturated Stations of the Pierre Auger Surface Detector, in: Proc. 33rd Int. Cosmic Ray Conf., Rio de Janeiro, Brazil, 2013.
- [51] T. Schmidt, Measurement of the Flux of Ultra High Energy Cosmic Rays using very inclined Extensive Air Showers measured at the Pierre Auger Observatory, Auger internal note GAP-2010-046 (2010).
- [52] J. Abraham *et al.* (Pierre Auger Collab.), Measurement of the energy spectrum of cosmic rays above 10^{18} eV using the Pierre Auger Observatory, Physics Letters B 685 (4-5) (2010) 239 – 246. doi:DOI:10.1016/j.physletb.2010.02.013.
- [53] I. Valino *et al.* (Pierre Auger Collab.), The flux of ultra-high energy cosmic rays after ten years of operation of the Pierre Auger Observatory, in: Proc. 34th Int. Cosmic Ray Conf., The Hague, The Netherlands, 2015.
- [54] A. Aab, et al., Pierre Auger Collaboration, Depth of Maximum of Air-Shower Profiles at the Pierre Auger Observatory: Measurements at Energies above $10^{17.8}$ eV, Phys.Rev.D (2014)arXiv:1409.4809.
- [55] F. Sánchez *et al.* (Pierre Auger Collab.), The muon component of extensive air showers above $10^{17.5}$ eV measured with the Pierre Auger Observatory, in: Proc. 36th Int. Cosmic Ray Conf., Madison, Wisconsin, USA, 2019.
- [56] A. Aab, et al., Pierre Auger Collaboration, Muons in air showers at the Pierre Auger Observatory: Mean number in highly inclined events, Phys.Rev.D (2014)arXiv:1408.1421.

- [57] Pierre Auger Collaboration, Testing hadronic interactions at ultrahigh energies with air showers measured by the pierre auger observatory, *Phys. Rev. Lett.* 117 (2016) 192001. doi:[10.1103/PhysRevLett.117.192001](https://doi.org/10.1103/PhysRevLett.117.192001).
- [58] J. f. t. P. A. C. Ebr, Pierre Auger Collaboration, Measurement of the proton-air cross-section at $\sqrt{s} = 57$ TeV with the Pierre Auger Observatory (2012) 341–347 [arXiv:1212.4053](https://arxiv.org/abs/1212.4053).
- [59] A. Aab, et al., Observation of a large-scale anisotropy in the arrival directions of cosmic rays above 8×10^{18} eV, *Science* 357 (6357) (2017) 1266–1270. doi:[10.1126/science.aan4338](https://doi.org/10.1126/science.aan4338).
- [60] M. Settimo *et al.* (Pierre Auger Collab.), An update on a search for ultra-high energy photons using the Pierre Auger Observatory, in: Proc. 32nd Int. Cosmic Ray Conf., Beijing, China, 2011.
- [61] P. Abreu, et al., Pierre Auger Collaboration, Ultrahigh Energy Neutrinos at the Pierre Auger Observatory, *Adv.High Energy Phys.* 2013 (2013) 708680. [arXiv:1304.1630](https://arxiv.org/abs/1304.1630), doi:[10.1155/2013/708680](https://doi.org/10.1155/2013/708680).
- [62] The Pierre Auger Collaboration, The Pierre Auger Observatory: Contributions to the 35th International Cosmic Ray Conference (ICRC 2017), ICRC 2017, Bexco, Busan, Korea, 2017.
- [63] The Pierre Auger Collaboration, The Pierre Auger Observatory: Contributions to the 36th International Cosmic Ray Conference (ICRC 2019), ICRC 2019, Madison, Wisconsin, USA, 2019.
- [64] A. Aab, et al., Pierre Auger Collaboration, Depth of Maximum of Air-Shower Profiles at the Pierre Auger Observatory: Composition Implications, *Phys.Rev.D* (2014) [arXiv:1409.5083](https://arxiv.org/abs/1409.5083).
- [65] S. Ostapchenko, Total and diffractive cross sections in enhanced pomeron scheme, *Physical Review D* 81 (11) (2010). doi:[10.1103/physrevd.81.114028](https://doi.org/10.1103/physrevd.81.114028).
- [66] K. Werner, F.-M. Liu, T. Pierog, Parton ladder splitting and the rapidity dependence of transverse momentum spectra in deuteron-gold collisions at the BNL Relativistic Heavy Ion Collider, *Phys. Rev. C: Nucl. Phys.* 74 (4) (2006) 11. doi:[10.1103/PhysRevC.74.044902](https://doi.org/10.1103/PhysRevC.74.044902).
- [67] E.-J. Ahn, R. Engel, T. K. Gaisser, P. Lipari, T. Stanev, Cosmic ray interaction event generator SIBYLL 2.1, *PRD* 80 (9) (2009) 094003. doi:[10.1103/PhysRevD.80.094003](https://doi.org/10.1103/PhysRevD.80.094003).
- [68] The Pierre Auger Collaboration, A. Aab, P. Abreu, M. Aglietta, E. J. Ahn, I. A. Samarai, I. F. M. Albuquerque, I. Allekotte, P. Allison, A. Almela, et al., The Pierre Auger Observatory Upgrade - Preliminary Design Report, ArXiv e-prints (2016) [arXiv:1604.03637](https://arxiv.org/abs/1604.03637).
- [69] R. Engel *et al.* (Pierre Auger Collab.), Upgrade of the Pierre Auger Observatory (AugerPrime), in: Proc. 34rd Int. Cosmic Ray Conf., The Hague, The Netherlands, 2015.
- [70] A. Dyshkant, D. Beznosko, G. Blazey, D. Chakraborty, K. Francis, D. Kubik, J. G. Lima, V. Rykalin, v. Zutshi, E. Baldina, A. Bross, P. Deering, About NICADD extruded scintillating strips (2018).

- [71] Wavelength shifting fibers, <http://kuraraypsf.jp/psf/ws.html>.
- [72] NOvA Collaboration, The NOvA Technical Design Report, 2007, FERMILAB-DESIGN-2007-01.
- [73] Hamamatsu Photonics R9420 datasheet, https://www.hamamatsu.com/resources/pdf/etd/R9420_TPMH1296E.pdf.
- [74] A. Streich, Scintillator surface detector for the upgrade of the pierre auger observatory, Master's thesis, Karlsruhe Institute of Technology (May 2017).
- [75] J. Birks, The Theory and Practice of Scintillation Counting, Pergamon, 1964.
- [76] G. F. Knoll, Radiation detection and measurement, 3rd Edition, Wiley New York, 1989.
- [77] I. Yashin, N. Ampilogov, I. Astapov, N. Barbashina, V. V Borog, D. Chernov, A. Dmitrieva, K. Kompaniets, A. A Petrukhin, V. V Shutenko, D. A Timashkov, Scintillation hodoscope for muon diagnostics (2009).
- [78] PDG, Atomic and Nuclear Properties of Materials.
- [79] R. D. Evans, The Atomic Nucleus, McGraw-Hill, 1955.
- [80] P. F. K. Grieder, Cosmic Rays at Earth, 1st Edition, Elsevier, 2001.
- [81] S. Argiro, S. L. C. Barroso, J. Gonzalez, L. Nellen, T. Paul, T. A. Porter, L. Prado Jr., M. Roth, R. Ulrich, D. Veberič, The offline software framework of the pierre auger observatory, Nucl. Instrum. Meth. A580 (2007) 1485–1496.
- [82] N. Barenthien, H. Bolz, K. Daumiller, R. Engel, H. Kern, M. Kessler, J. Maintok, H.-J. Mathes, M. Riegel, M. Roth, A. Schulz, R. Smida, A. Streich, D. Veberic, G. Woerner, Scintillator Surface Detectors of the Engineering Array: Production and Validation, Auger internal note GAP-2016-040 (2016).
- [83] A. Castellina *et al.* (Pierre Auger Collab.), The dynamic range of the AugerPrime Surface Detector: technical solution and physics reach, in: Proc. 35th ICRC, (2017) 397, arXiv:1708.06592, Busan, South Korea, 2017.
- [84] D. Nitz *et al.* (Pierre Auger Collab.), New Electronics for the Surface Detectors of the Pierre Auger Observatory, in: Proc. 36th Int. Cosmic Ray Conf., Madison, Wisconsin, USA, 2019.
- [85] B. Pont *et al.* (Pierre Auger Collab.), A Large Radio Detector at the Pierre Auger Observatory - Measuring the Properties of Cosmic Rays up to the Highest Energies, in: Proc. 36th Int. Cosmic Ray Conf., Madison, Wisconsin, USA, 2019.
- [86] Patricia Sanchez-Lucas *et al.* (Pierre Auger Collab.), $\langle X_{\max} \rangle$ measurements and tests of hadronic models using the surface detector of the Pierre Auger Observatory, in: Proc. 35th ICRC, (2017) 506, arXiv:1708.06592, Busan, South Korea, 2017.
- [87] T. Schulz, Enhanced UHECR Event Reconstruction by means of Sampling Lateral Distributions with Multiple Surface Sub-Detectors, Auger internal note GAP-2018-025 (2018).
- [88] P. Billoir, Natural and artificial fluctuations in the auger surface detector, Auger internal note GAP-2001-005 (2001).

- [89] [NLOpt - library for non-linear optimization](http://ab-initio.mit.edu/wiki/), [http://ab-initio.mit.edu/wiki/](http://ab-initio.mit.edu/wiki/index.php/NLOpt_manual).
URL http://ab-initio.mit.edu/wiki/index.php/NLOpt_manual
- [90] D. Schmidt, Sensitivity of AugerPrime to the masses of ultra-high-energy cosmic rays, Auger internal note GAP-2019-030 (2019).
- [91] S. Raschka, Implementing a Principal Component Analysis (PCA) - in Python, step by step, http://sebastianraschka.com/Articles/2014_pca_step_by_step.html#%E2%80%93in-python-step-by-step.
- [92] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in Python, *Journal of Machine Learning Research* 12 (2011) 2825–2830.
- [93] S. Raschka, Linear Discriminant Analysis - Bit by Bit, https://sebastianraschka.com/Articles/2014_python_lda.html.
- [94] J. Pękala *et al.* (Pierre Auger Collab.), Production and Quality Control of the Scintillator Surface Detector for the AugerPrime Upgrade of the Pierre Auger Observatory, in: Proc. 36th Int. Cosmic Ray Conf., Madison, Wisconsin, USA, 2019.
- [95] A. Taboada *et al.* (Pierre Auger Collab.), Analysis of Data from Surface Detector Stations of the AugerPrime Upgrade, in: Proc. 36th Int. Cosmic Ray Conf., Madison, Wisconsin, USA, 2019.
- [96] A7501 Power Supply Module, <https://www.caen.it/products/a7501/#>.
- [97] PHQ, Integrated HV Supplies for PMT, <https://iseg-hv.com/en/products/detail/PHQ>.
- [98] A. Castellina, Augerprime: the pierre auger observatory upgrade, *EPJ Web of Conferences* 210 (2019) 06002. doi:10.1051/epjconf/201921006002.
- [99] A. Coleman, Measurement of the cosmic ray flux above 100 pev at the pierre auger observatory, Dissertation, The Pennsylvania State University (2018).
- [100] A. Aab, et al., Pierre Auger, Muons in air showers at the Pierre Auger Observatory: Mean number in highly inclined events, *Phys. Rev. D* 91 (3) (2015) 032003, [Erratum: *Phys. Rev. D* 91, no. 5, 059901 (2015)]. [arXiv:1408.1421](https://arxiv.org/abs/1408.1421), doi:10.1103/PhysRevD.91.032003, 10.1103/PhysRevD.91.032003.
- [101] Dembinski, H.P., Arteaga-Velázquez, J.C., Cazon, L., Conceição, R., Gonzalez, J., Itow, Y., Ivanov, D., Kalmykov, N.N., Karpikov, I., Müller, S., Pierog, T., Riehn, F., Roth, M., Sako, T., Soldin, D., Takeishi, R., Thompson, G., Troitsky, S., Yashin, I., Zadeba, E., Zhezher, Y., for the EAS-MSU, IceCube, KASCADE-Grande, NEVOD-DECOR, PierreAuger, SUGAR, Telescope Array, Yakutsk EAS Array, Report on tests and measurements of hadronic interaction properties with air showers, *EPJ Web Conf.* 210 (2019) 02004. doi:10.1051/epjconf/201921002004.
- [102] S. Ostapchenko, Monte Carlo treatment of hadronic interactions in enhanced Pomeron scheme: I. QGSJET-II model, *Phys. Rev. D* 83 (2011) 014018. doi:10.1103/PhysRevD.83.014018.

- [103] T. Pierog, I. Karpenko, J. M. Katzy, E. Yatsenko, K. Werner, Epos lhc: Test of collective hadronization with data measured at the cern large hadron collider, *Physical Review C* 92 (3) (2015). doi:[10.1103/physrevc.92.034906](https://doi.org/10.1103/physrevc.92.034906).
- [104] L. Cazon, Probing high-energy hadronic interactions with extensive air showers (2019). arXiv:[1909.02962](https://arxiv.org/abs/1909.02962).
- [105] The AMIGA Group, AMIGA-Grande: a beyond 2015 proposal with muon counters, Auger internal note GAP-2013-003 (2013).
- [106] J. M. Figueira, Private Communication about performance of AMIGA-Grande (April 2018).
- [107] The Bariloche Group, Proposal for Auger Beyond 2015: ASC-II: Auger Scintillators for Composition II, Auger Upgrade proposal note (2013).
- [108] M. Ave, Ground signal parameterization for water Cherenkov and scintillator detectors and performance of ground arrays that combine both techniques, Auger internal note GAP-2012-141 (2012).
- [109] A. Letessier-Selvon, P. Billoir, M. Blanco, I. C. Mariş, M. Settimo, Layered water Cherenkov detector for the study of ultra high energy cosmic rays, *Nuclear Instruments and Methods in Physics Research A* 767 (2014) 41–49. doi:[10.1016/j.nima.2014.08.029](https://doi.org/10.1016/j.nima.2014.08.029).
- [110] The MARTA Group, MARTA Muon Auger RPC for the Tank Array Design Report, Auger internal note GAP-2013-020 (2013).
- [111] The TOSCA Group, TOSCA The Observatory SCintillator Array Technical Project, Auger internal note GAP-2014-074 (2014).
- [112] R. Calderón-Adrila *et al.* (LAGO collaboration), Modeling the LAGO's detectors response to secondary particles at ground level from the Antarctic to Mexico, in: Proc. 36th Int. Cosmic Ray Conf., Madison, Wisconsin, USA, 2019.
- [113] M. Suárez-Durán, H. Asorey, A. Taboada, S. Dasso, L. A. Núñez, Studies of a Water Cherenkov Detector response to the Forbush Decrease of December 2006, Auger internal note GAP-2018-059 (2018).
- [114] H. Asorey, Los detectores Cherenkov del Observatorio Pierre Auger y su aplicación al estudio de fondos de radiación, Phd thesis, Instituto Balseiro, Universidad Nacional de Cuyo, CNEA (2012).
- [115] J. D. Jackson, *Classical Electrodynamics*, 3rd Edition, Wiley, 1998.
- [116] C. Patrignani, et al., Particle Data Group, Review of Particle Physics, *Chin. Phys.* C40 (10) (2016) 100001. doi:[10.1088/1674-1137/40/10/100001](https://doi.org/10.1088/1674-1137/40/10/100001).
- [117] G. M. Hale, M. R. Querry, Optical constants of water in the 200-nm to 200- μm wavelength region, *Appl. Opt.* 12 (3) (1973) 555–563. doi:[10.1364/AO.12.000555](https://doi.org/10.1364/AO.12.000555).
- [118] P. D. Group, Atomic and Nuclear Properties of Materials, <http://pdg.lbl.gov/2019/AtomicNuclearProperties/index.html>.