Measurement of the Depth of Shower Maximum in the Transition Region between Galactic and Extragalactic Cosmic Rays with the Pierre Auger Observatory

Zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften
an der Fakultät für Physik des
Karlsruher Institut für Technologie (KIT)

genehmigte

Dissertation

von

M.Sc. Alessio Porcelli

aus Pietrasanta (Italien)

Tag der mündlichen Prüfung: 06.06.2014
Referent: Prof. Dr. J. Blümer
Korreferent: Prof. Dr. U. Husemann
Betreuer: Dr. R. Engel, Dr. M. Unger

"There is a theory which states that if ever anyone discovers exactly what the Universe is for and why it is here, it will instantly disappear and be replaced by something even more bizarre and inexplicable. There is another theory which states that this has already happened."

Douglas N. Adams

Abstract

The mass composition of cosmic rays is one of the key ingredients for determining the sources of these high-energy particles. At energies above 10^{15} eV the mass cannot be measured directly and one has to resort to the study of the secondary cascades (air showers) produced in the Earth's atmosphere. The depth of which the number of secondary particles reaches a maximum is called X_{max}.

The subject of this thesis is the measurement of the average and standard deviation of the X_{max} distribution: $\langle X_{\mathrm{max}} \rangle$ and $\widehat{\sigma}(X_{\mathrm{max}})$. Data from the fluorescence detector of the Pierre Auger Observatory are used to achieve this goal. The *High Elevation Auger Telescopes* (HEAT) are used to extend the current Pierre Auger Observatory results to lower energies down to 10^{17} eV. The data sample analyzed in this thesis is selected to ensure its high quality and minimal bias. Moreover cross checks on the acquired data and reconstructed shower parameters are performed. To give a proper significance of the measurements, also systematic uncertainties and resolution are studied in detail.

$\langle X_{\mathrm{max}} \rangle$ and $\widehat{\sigma}(X_{\mathrm{max}})$ are presented for almost three order of magnitude of the energy. The results at energies above 10^{18} eV confirm the previous measurements of the Pierre Auger Collaboration. Below this energy, an unexpected feature at $10^{17.5}$ eV is found in the analysis of the $\langle X_{\mathrm{max}} \rangle$ distribution. At this energy the *elongation rate*, which represent the evolution of the composition as a function of $\log_{10}(E/\mathrm{eV})$, changes.

The measured moments of the X_{max} distribution are closely related to the nuclear composition of the cosmic ray flux. Thus, the results can be interpreted in terms of atomic mass composition as a function of the energy. The results of this work suggest a heavy dominated composition at low energy that becomes lighter and lighter up to $\sim 10^{18.5}$ eV, with a mixture of nuclei with very different atomic masses. Above this energy, the composition becomes heavier again, and the mixture is dominated by nuclei with very similar atomic masses.

The presented measurements cover the region of the transition between galactic and extragalactic cosmic rays expected by source models, where the data from other experiments are rather sparse. Therefore, the results provide important input to studies of cosmic ray sources and propagation in the universe.

Zusammenfassung

Die Messung der Massenzusammensetzung der kosmischen Strahlung ist ein wichtiger Bestandteil bei der Suche nach den Quellen der ultra-hochenergetischen Teilchen. Bei Energien oberhalb von 10^{15} eV kann die Primärmasse nicht direkt gemessen werden. Man muss deshalb auf indirekte Messungen ausweichen und die Kaskaden aus Sekundärteilchen (auch Luftschauer genannt) untersuchen, die in der Atmosphäre der Erde erzeugt werden. Die atmosphärische Tiefe, wo die Teilchenkaskade ihr Maximum erreicht heißt $X_{\max}$.
Thema dieser Arbeit ist die Messung von Mittelwert und Streuung der $X_{\max}$-Verteilung: $\langle X_{\max} \rangle$ und $\widehat{\sigma}(X_{\max})$. Zu diesem Zweck werden Daten des Fluoreszenzdetektors des Pierre Auger-Observatoriums verwendet. Die Erweiterung *High Elevation Auger Telescopes* (HEAT) dient dazu, die bisherigen Ergebnisse des Pierre Auger-Observatoriums zu niedrigeren Energien bis hin zu 10^{17} eV auszudehnen.
Die in dieser Arbeit verwendeten Daten wurden selektiert um größtmögliche Datenqualität zu gewährleisten und Verzerrungen zu minimieren. Weitere Überprüfungen der Messdaten und der rekonstruierten Schauerparameter wurden vorgenommen. Um die Signifikanz der Messung zu beurteilen werden systematische Unsicherheiten und Auflösung detailliert untersucht.
Die Verteilungen von $\langle X_{\max} \rangle$ und $\widehat{\sigma}(X_{\max})$ werden in einem Energiebereich vorgestellt, der fast drei Größenordnungen überspannt. Die Ergebnisse oberhalb von 10^{18} eV bestätigen das bisherige Ergebnis der Pierre Auger-Kollaboration. Bei kleineren Energien wurde im Bereich von $10^{17.5}$ eV ein unerwartetes Merkmal in der Verteilung von $\langle X_{\max} \rangle$ gefunden. Bei dieser Energie ändert sich die Steigung von $\langle X_{\max} \rangle$ als Funktion der Energie (*elongation rate*).
Die Momente der gemessenen $X_{\max}$-Verteilung sind eng mit der Massenzusammensetzung der kosmischen Strahlung verknüpft. Somit können die Ergebnisse hinsichtlich des Verlaufs der Massenzusammensetzung als Funktion der Energie interpretiert werden. Bei niedrigen Energien wird die Massenverteilung von schweren Primärteilchen dominiert. Mit ansteigender Energie nimmt die Primärmasse bis hin zu $\sim 10^{18.5}$ eV ab. Die mittlere Masse ist dabei Resultat einer Mischung sehr unterschiedlicher Atommassen. Bei noch höheren Energien nimmt die Primärmasse wieder zu und die Massenzusammensetzung wird von Kernen mit sehr ähnlicher Masse dominiert.
Die hier vorgestellten Messungen überspannen den Energiebereich in dem sich nach gängiger Modellvorstellung der Übergang von galaktischer zu extragalaktischer kosmischer Strahlung vollzieht. Bei dieser Energie liegen nur wenige Ergebnisse anderer Experimente vor. Deshalb sind die Ergebnisse dieser Arbeit bedeutsam zur Untersuchung der Quellen der kosmischen Strahlung und deren Ausbreitung im Universum.

Sommario

La composizione atomica dei raggi cosmici è una degli ingredienti chiave per determinare le sorgenti di queste particelle ad alta energia. Ad energie al di sopra di 10^{15} eV la massa atomica non può essere determinata direttamente e si deve ricorrere allo studio degli sciami secondari prodotti nell'atmosfera terrestre. La profondità alla quale il numero di particelle secondarie raggiunge il suo massimo si chiama $X_{\rm max}$

Il soggetto trattato in questa tesi è la misura della media e della deviazione standard della distribuzione di $X_{\rm max}$: $\langle X_{\rm max}\rangle$ e $\widehat{\sigma}(X_{\rm max})$. Per questo scopo sono stati usati i dati dai detector a fluorescenza di Pierre Auger Observatory. *High Elevation Auger Telescopes* (HEAT) è usato per estendere l'attuale limite inferiore in energia delle misure di Pierre Auger Observatory a 10^{17} eV. Il campione di dati analizzati in questa tesi è stato selezionato per assicurare la loro alta qualità con il minimo bias. Inoltre sono stati eseguiti controlli sui dati acquisiti e sui parametri degli sciami ricostruiti. Per dare la dovuta significatività alle misure, sono state studiate in dettaglio le incertezze sistematiche e la risoluzione

$\langle X_{\rm max}\rangle$ and $\widehat{\sigma}(X_{\rm max})$ sono presentate per circa tre ordini di magnitudine in energia. I risultati sopra 10^{18} eV confermano le precedenti misure della Pierre Auger Collaboration. Sotto tale energia è stata trovata una caratteristica inattesa a $10^{17.5}$ eV nell'analisi della distribuzione di $\langle X_{\rm max}\rangle$. Questa è un nuovo punto in cui la *elongation rate*, che rappresenta l'evoluzione della composizione in funzione dell'energia, cambia.

Le misure sui momenti della distribuzione di $X_{\rm max}$ sono strettamente correlate alla composizione nucleare del flusso di raggi cosmici. Quindi i risultati possono essere interpretati in termini di composizione di masse atomiche in funzione dell'energia. I risultati di questo lavoro suggeriscono un composizione dominata da nuclei pesanti alle basse energie che diventa via via più leggera fino a $\sim 10^{18.5}$ eV, con una miscela di nuclei con masse atomiche molto diverse fra loro. Oltre questa energia, la composizione diventa nuovamente pesante, e la miscela è dominata da nuclei con masse atomiche simili.

Le misure presentate coprono la regione di transizione tra raggi cosmici galattici ed extra-galattici attesa dai modelli sulla loro origine, dove i dati dagli altri esperimenti sono alquanto sparsi. Perciò i risultati forniscono un'importante input per studiare l'origine dei raggi cosmici e la loro propagazione nell'universo.

Contents

CHAPTER 1

Introduction

From the cosmos, particles come to Earth with energies up to $\sim 10^{20}$ eV. They are the messengers of astrophysical sources, carrying fundamental information of the universe and are called *Cosmic Rays*.

Since the first indirect detection of radiation from the Universe by Victor Hess in 1912 [1], cosmic rays assumed a leading role in astro- and particle physics. Through the studies of the reactions induced by them, for instance, physicists were able to discover new particles such as the positron, muon and pion. Nowadays the cosmic rays are the only way to study physics at the most extreme energies. The particle accelerators are a controlled environment, but they can only reach center of mass energies equivalent to 10^{17} eV, in contrast to $\sim 10^{20}$ eV achievable by the cosmic ray flux. Moreover cosmic rays are the key to understand astrophysical and cosmological scenarios still on debate.

Today much progress in understanding cosmic rays has been achieved. But a fundamental question still remains open: "where do they come from?". The known acceleration mechanism of the particles is through the shock front of the supernova explosions. This mechanism is able to explain galactic sources up to $\sim 10^{17}$ eV. At the highest energies the particles are assumed to come from extragalactic sources, but the astrophysical processes that can produce such energies are not yet clear. Different groups of models of acceleration and propagation in the universe exist, with different predictions on the chemical composition of the flux.

The simplest scenarios consider the extragalactic component of the cosmic ray flux being strongly dominated by protons. More complicate scenarios with mixed composition of the extragalactic particle flux exist, although they also predict a galactic component extended farther in energy than achievable by the supernova shock front mechanism. The transition between galactic and extragalactic sources depends on the model and occurs between 10^{17} and 10^{19} eV. Thus, the mass composition of the particle flux above 10^{17} eV is an important information for discriminating the models, especially in the $0.3 \div 1 \cdot 10^{18}$ eV range where the data from the experiments are rather sparse.

At these energies, the cosmic ray flux becomes so low that direct measurements are impossible

and the atmosphere has to be used as a calorimeter to produce air showers of secondary particles from incoming cosmic rays. Electromagnetic and muonic components of the shower can be measured at the ground. Shower tracks can be detected through fluorescence light. This light is produced by the molecules of the atmosphere. They are excited by the energy loss of the secondary particles in a shower due to ionization. When the molecules return to the ground state of energy, light is emitted at UV frequency.

In this work, data from the Pierre Auger Observatory were analyzed. The observatory is the largest ground-based cosmic ray detector and covers an area of $3000\,\mathrm{km}^2$. It is a hybrid detector that combines the detection of both fluorescence light and particles at the ground. The former are detected through the *Fluorescence Detector*, a set of telescopes that measure the energy released from the shower along its track. The latter are measured through the *Surface Detector Array*, which consists of water Cherenkov detectors that count the particle energy density at the ground. Combining both detectors, a high quality geometry of the shower can be reconstructed.

The aim of this work is to analyze the mass composition with the Fluorescence Detector of the Pierre Auger Observatory in the largest energy range as possible. Measuring directly the mass is impossible, however the particle flux composition is related to the distribution of the measured depth where the air shower development reaches its maximum (X_{max}). The observatory was originally designed to measure cosmic rays above $10^{18}\,\mathrm{eV}$ and beyond the GZK cut-off. With the Fluorescence Detector enhancement *High Elevation Auger Telescopes* (HEAT) it is able to measure showers also below this limit. Thus including HEAT data in the analysis, the transition region ($10^{17} \div 10^{19}\,\mathrm{eV}$) and beyond can be covered, and new measurements to enhance the understanding of the cosmic ray sources can be provided.

An introduction to the flux of cosmic rays and how air showers can be use to understand this flux will be briefly presented in Chapter 2. Chapter 3 will introduce the Pierre Auger Observatory experiment focussing on the fluorescence telescopes, especially the new enhancement HEAT. The quality selection of the data will be described in Chapter 4, discussing in particular:

- the optimal reconstructed data set;
- the selection of high quality data set to ensure a good X_{max} and energy resolution to resolve the true shower maximum distribution;
- the application of fiducial volume cuts to minimize possible biases given by a limited filed of view.

After the data selection, remaining shortcoming of the reconstruction and a small residual bias on the tails of the X_{max} distribution due to the limited field of view will be corrected based on simulations (Chapter 5). In Chapter 6 the study of the systematic uncertainties and the resolution of the measurements will be discussed. High quality measurements with minimal bias will be finally used to present a proper measurement of the moments of the X_{max} distribution in Chapter 7. In this chapter, also the interpretation of such moments will be provided in terms of the distribution of the mass composition of the cosmic ray flux. In Chapter 8 the results will be summarized.

CHAPTER 2

Cosmic rays at high energy

The energy-dependent flux of *Cosmic Rays* (CRs) is known over 11 decades in energy, starting from few hundred MeV with several thousand particles per square meter per second to less than one particle per square kilometer per century at 10^{20} eV . The spectrum is rather featureless with a few exceptions, following the power law

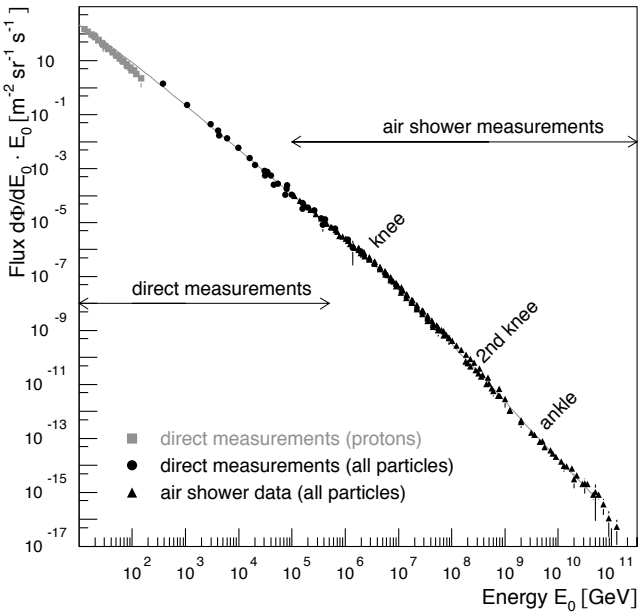
$$\mathrm{d}N/\mathrm{d}E \propto E^{\gamma}. \tag{2.1}$$

These exceptions are the first and second *knee*, the *ankle* and a strong suppression at the highest energies, often for historical reason referred to as the *GZK cut-off*. In Fig. 2.1 a compilation of flux measurement is shown. In Fig. 2.1(b), the flux shown in Fig. 2.1(a) is scaled by a factor $E^{2.5}$ to better observe the features.

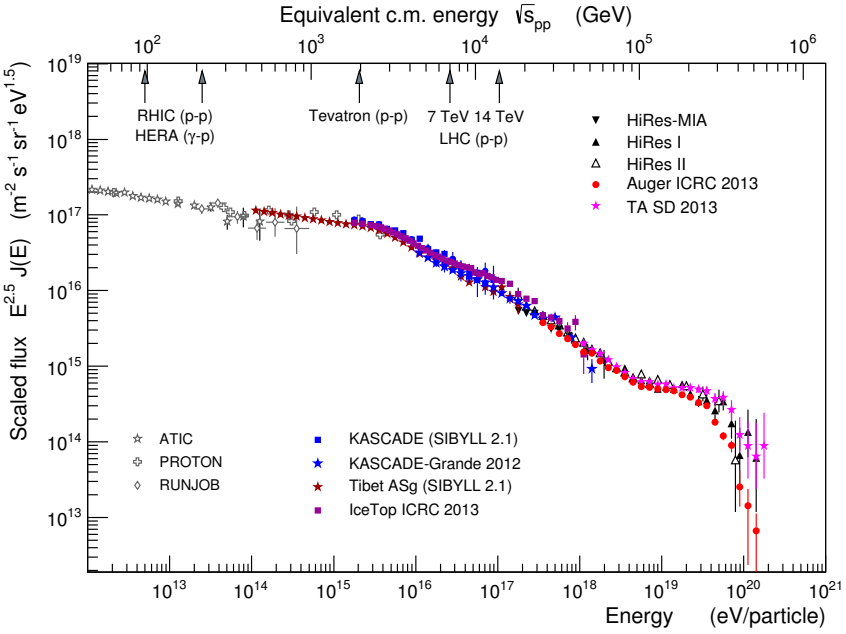
At the lowest energies, the source is the sun (often referred to as solar energetic particles). Increasing the energy, nuclei from the entire Galaxy are detected. Around $3 \div 5 \cdot 10^{15}$ eV the first knee is observed and the γ index becomes steeper. Around $0.8 \div 3 \cdot 10^{17}$ eV, the second knee is measured by the experiments [4–7]: a more slight decreasing of the index occurs. These features are associated with the cosmic rays from the Galaxy and their acceleration and propagation mechanisms.

At higher energies, the index γ of the energy spectrum increases for a small energy interval before decreasing again. This energy region is the so-called ankle. In Fig. 2.1(b) it can be observed as a flat region. The change of index occurs at $0.7 \div 1 \cdot 10^{19}$ eV. The interpretation of this change strongly depends on the scenario considered for the transition between galactic and extragalactic CR.

Above $3 \cdot 10^{19}$ eV, a strong suppression is observed [8]. Roughly at this energy, the Greisen-Zatsepin-Kuzmin suppression (GZK) [9, 10] was predicted. This effect is the energy loss due the interaction of protons with photons from the *Cosmic Microwave Background* (CMB). But also other effects are predicted at this energy, such as the *photo-disintegration* of the nuclei. Thus



(a) Flux not scaled.



(b) Flux measured by several ground experiments, scaled by $E^{2.5}$.

Figure 2.1: Energy flux of the cosmic rays. Pictures from [2, 3] respectively.

the origin of this suppression is still debated[1]. However, at these energies the CRs are from extragalactic sources.

Up to 10^{14} eV the CR flux allows direct measurement of the mass, charge and energy of individual nuclei. Balloon- or satellite-borne detectors are used for direct detection to avoid the interaction with the atmosphere.

At higher energies, the flux becomes so low that direct measurements are impossible on a reasonable time scale. However indirect measurement can be performed. The atmosphere can be used as a calorimeter to produce "air showers" of "secondary" particles from incoming CRs ("primary" particles). Electromagnetic ("soft") and muonic ("hard") components of the shower can be measured at the ground. Shower tracks can be detected through fluorescence light. This light is produced by the molecules of the atmosphere. They are excited by the energy loss of the secondary particles in a shower due to ionization. When they return to the ground state of energy, this light is emitted at UV frequency.

So-called "ground experiments" are designed to measure the shower information and reconstruct the characteristics of the original cosmic ray particle: energy, mass and arrival direction. The **Pierre Auger Observatory** is a ground experiment built to measure simultaneously both fluorescence light tracks and hard/soft components at the ground. It will be described in the next chapter.

Models for galactic (Section 2.1) and extragalactic (Section 2.2) CRs are necessary to explain all the features of the energy spectrum, as well their transition (Section 2.3). Several possible scenarios must be discriminated by the measurements. To study the cosmic rays with the ground experiments, the physics of the extensive air showers must be understood (Section 2.4). Thus it is possible to relate the observations with the evolution of the nuclear abundance in the cosmic ray flux (Section 2.5).

2.1 Galactic cosmic rays

To describe the observed energy spectrum of the experiments, models of CR sources and particle propagation in the galaxy are developed, for example see [11]. Most of them are based on the *Supernova Remnants* (SNR) paradigm. Supernovae are expected to be the main sources of the CR, accelerated by a statistical process and confined by the galactic magnetic field.

The acceleration process is the so-called Fermi acceleration mechanism. In a moving shock front with a strong magnetic field, a particle can go in the opposite direction to the moving front (from upstream to downstream) in a stochastic process. Due to the magnetic field it can be reflected back through the shock (downstream to upstream) with an acceleration. If a similar process occurs again upstream, the particle will gain energy one more time. This multiple reflection leads to an increase of the particle energy. As the energy is increased, so does the probability to escape from the shock front since this is a "statistical acceleration" process. Thus the energy is described by a distribution with a steeply falling power law as in Eq. (2.1), that is suited for the energy spectrum description.

The higher the charge of a particle, the more the magnetic field of the acceleration region confines it. Thus the heavier the nucleus is, the more energy it can gain. This behavior is confirmed by the KASCADE experiment. It showed a significant decrease in the flux of light nuclei (proton and helium) after the first knee [12]. Eventually, at very high particle energy, the magnetic field of the acceleration region is not capable of confining the heavier nuclei for further acceleration anymore. This could be the reason for the second knee in the cosmic ray spectrum. The historical second knee is claimed at $\approx 3 \cdot 10^{17}$ eV, however KASCADE has found

[1] This suppression may simply be the maximum energy that extragalactic sources can accelerate the particles.

a new feature, a knee at $0.8 \cdot 10^{17}$ eV. The first and this latter knees are also called "proton" and "iron" knees, respectively.

Supernova remnants are very good candidates for the acceleration of cosmic rays up to the knee region. The expanding shell of plasma forms the shock front because its velocity is much larger than the velocity of sound in the interstellar medium. Thus it accelerates the nuclei from the medium through Fermi acceleration. The lifetime of the shell ($\approx 10^5$ years), its size and the magnetic field strength determine the maximum energy attainable.

Once a particle leaves the acceleration site, it is deflected by the galactic magnetic field. The radius of the circular motion of a particle with mass m and charge q in a magnetic field with field strength B is the Larmor radius r_L

$$r_L = \frac{m v_\perp}{|q|\, B}. \tag{2.2}$$

In our Galaxy, it is expected that nuclei with energies below 10^{18} eV are confined by its magnetic fields. At energies above 10^{18} eV, no known sources of CRs within our galaxy exist.

The maximum energy for proton and iron primary particles predicted from this model are in agreement with the measurement of the knees from KASCADE-Grande [5, 11]. The experiment measured the spectra separately for electron-rich and electron-poor showers. The electron-rich showers are produced by proton and light elements. Their measurement estimates the energy of the first knee and it corresponds to the maximum energy estimated for protons or helium by the SNR model: $\sim 10^{15}$ eV. The electron-poor showers are produced by heavy elements. With them, the iron knee is estimated to be at $\sim 10^{17}$ eV, as predicted for irons by the model. The knee measurements are confirmed by IceTop [7]. A consequence of this model is that the flux of the galactic CR is expected to disappear quickly after the iron knee.

This basic model has possible extensions, from nonlinear amplification of the magnetic field in the acceleration region driven by cosmic ray streaming [13], to CRs generated by the nebulas of pulsars [14].

2.2 Extragalactic cosmic rays

When a nucleus has enough energy, the Larmor radius (Eq. (2.2)) is larger than the halo of the Galaxy and it is no longer confined. Since there is almost no anisotropy of the arrival direction of the CRs observed above the iron knee [15], the sources of these cosmic rays are presumably located outside of our Galaxy.

In Fig. 2.2, the so-called Hillas plot, astrophysical sources are shown [17]. Every kind of source (gray areas) is shown as a function of the size of the acceleration region in the Fermi process (x axis) and the magnetic field in the source (y axis). The relation of the axis is

$$E_{\max} \sim \beta_s \cdot z \cdot B \cdot L \cdot 10^{21}\,\text{eV} \tag{2.3}$$

B, L and β_s are the magnetic field in G, the size in pc and the speed of the moving shock front (accelerating region) relative to c, z is the charge of the nucleus.

A proton accelerated to an energy of 10^{20} eV through a relativistic plasma ($\beta_s \simeq 1$) corresponds to the dashed line. The solid line indicates less extreme shock front ($\beta_s \simeq 1/300$). $E_{\max}$ scales with the charge, thus heavier nuclei are represented by lines laying below the proton limits. Only a few sources remain for protons with such energies. Possible candidates are Active Galactic Nuclei (AGN), Gamma Ray Bursts (GRB), radio lobes of FR II radio galaxies and young neutron stars [18, 19].

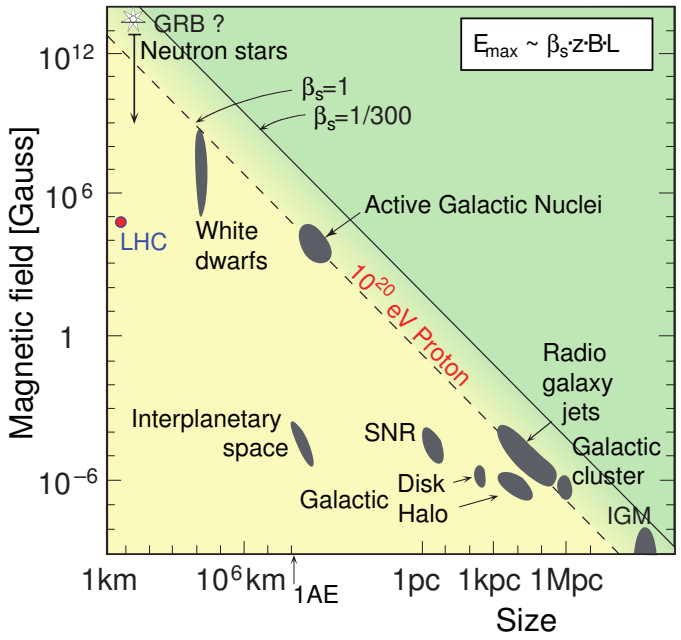


Figure 2.2: Hillas plot. On the x axis, size of the acceleration region in the Fermi process is displayed. On the y axis its magnetic field. Diagonal lines are the limits for the protons at 10^{20} eV (see Section 2.2 for details). From [16].

The mechanisms capable of producing nuclei with the observed energies are categorized into two distinct scenarios [20]. One is the "bottom-up" acceleration scenario, where particles increase their energy through some mechanism. The Fermi acceleration discussed until now is such a mechanism. The second is so-called "top-down" decay.

The top-down models describe the observed CR as secondary nuclei with large kinetic energies generated by the decay of exotic super-heavy particles (at least $10^{23 \div 24}$ eV). Candidates are Z-Burst [21], super-heavy dark matter (SHDM) or topological defects (TD) [22]. In these models a large number of ultra-high energy photons and neutrinos are predicted.

Air shower experiments are typically not optimized for photon detection, still limits on their integrated flux can be derived from shower data. In Fig. 2.3 integral photon fluxes from model predictions and the experimental upper limits from measurements are shown [23]. The top-down models are strongly disfavored.

Independent of the sources, a strong suppression of the particle flux is observed above $10^{19.5}$ eV. This feature can be explained by the interaction of the protons and nuclei with the CMB. Protons can interact via

$$p + \gamma_{\mathrm{CMB}} \to \Delta^+ \to p + \pi^0 \ \to p + \gamma\gamma, \tag{2.4}$$

$$p + \gamma_{\mathrm{CMB}} \to \Delta^+ \to n + \pi^+ \to n + \mu^+ + \nu_\mu, \tag{2.5}$$

where $m(\Delta^+) = 1232\,\mathrm{MeV}$. The kinetic energy of protons is reduced by the energy needed to form the pion. This is the so-called GZK effect [9, 10].

Heavier nucleus with atomic mass number A is fragmented to lighter nuclei (with n nucleons N) due to its excitation for giant dipole resonance. This effect is named *photo-disintegration*

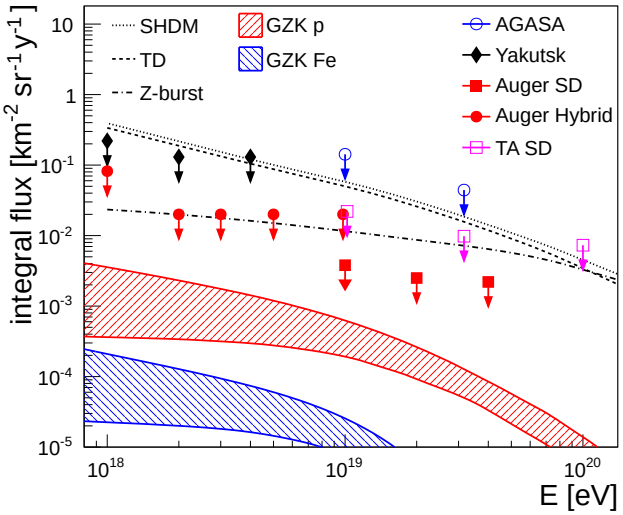


Figure 2.3: Integral limits on the flux of photons at 95 % CL from air shower experiments, compared to the predictions from top-down models and GZK effects. Experimental data are shown from AGASA, Yakutsk, Auger and TA. Top-down scenarios are from the super-heavy DM (SHDM), topological defect (TD) and Z-Bursts. Photons from the GZK effect (Eq. (2.4)) are predicted for the energy loss from protons and irons. From [23].

[9, 10] and can occur both with the CMB and the *Extragalactic Background Light* (EBL):

$$A + \gamma_{\mathrm{CMB,EBL}} \rightarrow (A - nN) + nN. \tag{2.6}$$

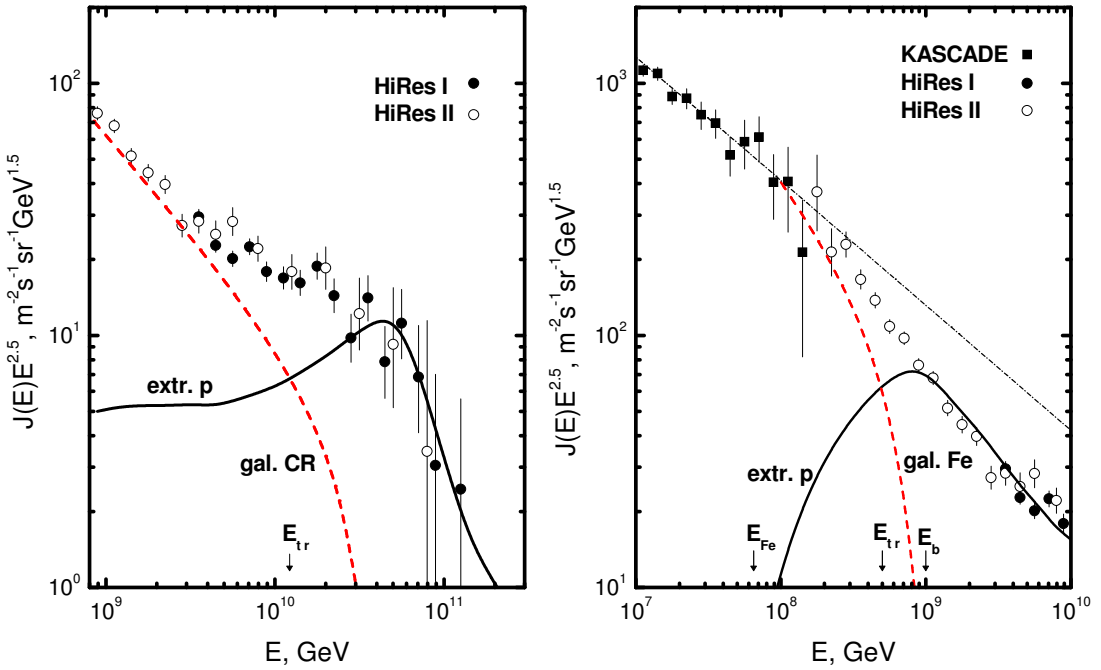
This effect can produce secondary protons that may lose energy with the GZK effect. In Fig. 2.3, also the prediction of the photon flux from Eq. (2.4) is shown, in case of proton and iron CRs. The sensitivity of the experiments is not sufficient to exclude models that predict photons created by the GZK effect.

Mainly two scenarios are studied to explain the observed suppression at the highest energies ($> 5 \cdot 10^{19}$ eV) of the spectrum. In case of a spectrum dominated by light nuclei (H, He), the GZK effect suppresses the CR flux. However, in case of heavy nuclei (Fe or nuclei with similar atomic mass) the energy lost is dominated by the fragmentation of the nuclei. The abundance of intermediate nuclei (say between He and Si) is expected to be suppressed due to the photo-disintegration [24].

Until now the experiments are not able to directly discriminate the subtle differences in the flux in the suppression region between the two scenarios. Composition measurements of the highest energies will be needed. Furthermore, to give an interpretation, the overall effects on the flux composition and arrival direction distribution must be considered. Understanding the transition region between the galactic and the extragalactic cosmic rays will be one key input for making progress.

2.3 Galactic-Extragalactic transition

The transition between galactic and extra-galactic sources of cosmic rays is expected to happen between $\sim 10^{17}$ eV and 10^{19} eV. This range includes a change of the flux structure around



(a) Example of an ankle scenario.

(b) Example of a dip scenario.

Figure 2.4: Examples of ankle and dip scenarios. In both, the extragalactic component is an injected pure proton flux ("extr. p", black solid lines) and the galactic one is obtained subtracting from the total flux the latter ("gal. CR/Fe", red dashed lines). $E_{\rm tr}$ is the energy where the transition occurs, $E_{\rm Fe}$ is the energy of the iron knee and E_b is the energy where the transition is complete. The two plots have different axes, centered in the transition region. From [11].

$\sim 10^{18.7}$ eV, also seen in the data [25] of the Pierre Auger Observatory: the ankle. The explanation of this peculiar feature is a central building block of the models described in the following. Three scenarios are typically considered: the *ankle*, *dip* and *mixed composition* models.

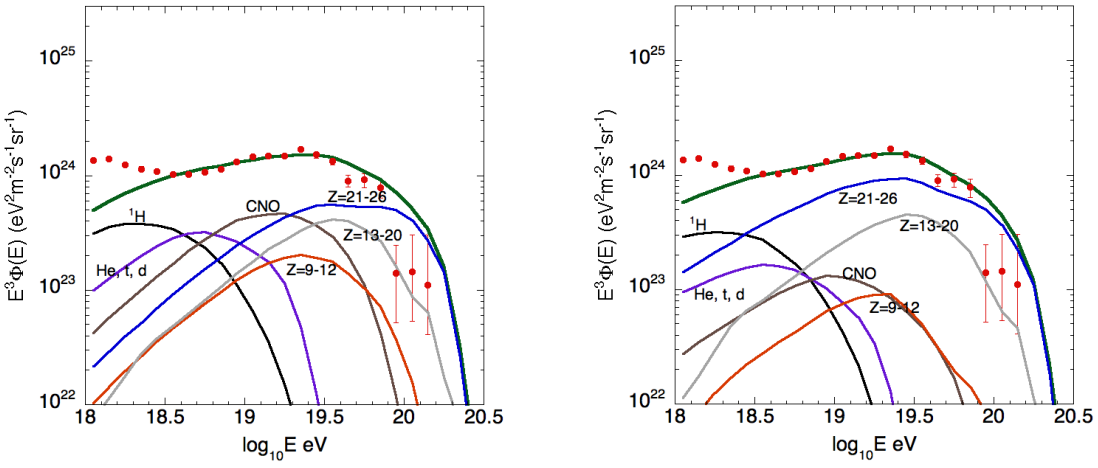
Ankle scenario

The ankle models, for example [26], are the first scenario that appeared in literature. It was proposed in the 1960s to explain the ankle feature. This peculiar region is generated, in this paradigm, by the different slope for galactic and extra galactic part in their transition region. Extragalactic sources start to dominate over the galactic around an $E_{\rm tr} \sim 10^{19}$ eV.

Assuming an iron predominance for the galactic sources after the iron knee, the majority of ankle models assume a pure proton composition for the extragalactic component. In Fig. 2.4(a) an example of this scenario is shown. In this example the extragalactic component is shown with a flux in the ankle region of E^{-2}, typical of non-relativistic shock acceleration. This spectrum is unmodified by the energy loss until the GZK effect occurs above $4 \cdot 10^{19}$ eV.

Such an extragalactic injected spectrum implies a galactic component that extends further than the iron knee. Thus an additional component[2] is needed, accelerated up to 30-40 times of the

[2]The extra component is usually assumed heavy-like, to address the difficulties to model the acceleration and the lack of a strong dipole anisotropy [15].



(a) Different element contribution is shown.

(b) As 2.5(a), but enriched of 30 % in heavy elements.

Figure 2.5: Examples of models from the mixed composition scenario. Data point (red dots) are from the Pierre Auger Observatory in 2011 [28]. From [29].

iron maximum energy. Hence the ankle is expected be dominated by heavy composition before the transition energy ($\sim 10^{18.5} \div \sim 10^{19}$ eV) and then the composition will be proton dominated.

Dip scenario

In the dip models [27], the ankle is explained by a dip in the extragalactic flux. This dip is given by an energy loss similar to the GZK mechanism. Protons at such energies can interact with photons from the CMB or from the EBL and produce electron-positron pairs:

$$p + \gamma_{\mathrm{CMB,EBL}} \rightarrow p + \mathrm{e}^+\mathrm{e}^-. \tag{2.7}$$

In Fig. 2.4(b) an example of this scenario is shown, assuming only protons are injected by the extragalactic sources. The proton flux starts at the iron knee ($E_{\mathrm{Fe}} \approx 10^{17}$ eV) and becomes dominant above $E_{\mathrm{tr}} \approx 5 \cdot 10^{17}$ eV. The dip starts at $E_{\mathrm{b}} \approx 10^{18}$ eV, immediately below the ankle region. With this extragalactic spectrum, the galactic flux has to have a strong suppression after the second knee, as expected for the SNR paradigm (Section 2.1).

With such fluxes, a heavy composition is expected between the iron knee and the transition (E_{tr}), becoming lighter and lighter. When the dip starts (E_{b}), the flux must be strongly proton-dominated. Thus the ankle too is expected to be proton dominated. Moreover the end of the spectrum is driven by the GZK suppression mechanism. However, the shape of the dip allows for a mixture of proton and light nuclei for the extragalactic CRs, but not more than $15 \div 20$ %.

Mixed composition scenario

The mixed composition scenario [29, 30] is the most complex of the three. It assumes that the extragalactic component consists of nuclei of various types. As first attempt, a mixed composition as for galactic cosmic-rays was injected as a possible mixture. Although there is no reasons to believe this hypothesis, it could be reasonable to preserve the hierarchy of the relative abundances between the different elements. Most of the models use this paradigm. In this scenario, the energy

where the transition occurs strongly depends on the model used, but it will be in-between the dip and ankle cases ($5 \cdot 10^{17} \div 10^{19}$ eV).

During the propagation of the cosmic rays, several energy loss processes can occur through the interaction nucleus-$\gamma_{\mathrm{CMB,EBL}}$. In case of protons, pair production (Eq. (2.7)) and the GZK effect (Eqs. (2.4) and (2.5)) are the processes of biggest importance. Nuclei loose energy through the photo-disintegration (Eq. (2.6)). Production of secondary protons may occur and be eventually involved in pair production and GZK processes as well. The heaviest nuclei (Fe or nuclei with similar atomic mass) have suppression similar to the protons, meanwhile the intermediate ones will be strongly suppressed by photo-disintegration.

In Fig. 2.5 two different examples of models are shown [29]. In Fig. 2.5(b) the flux has the same components as in Fig. 2.5(a) but different composition weights (enriched of 30 % in heavy elements). In both cases the injection of the composition is tuned to fit the flux measured by the Pierre Auger Observatory in 2011 [28]. Only extragalactic fluxes and their sum are shown.

2.4 Air Shower Phenomenology

All the scenarios to understand the sources of cosmic rays strongly depend on the abundances of nuclei in the flux. Without the knowledge of the mass composition, no firm conclusions on possible source models can be drawn. The mass of the incoming particles cannot be measured directly at energies above 10^{14} eV, but must be estimated from indirect measurements.

Particles and nuclei coming from outside the Earth interact with the atmosphere, which behaves as calorimeter. There are several different interactions that can occur in the air:

- *Bremsstrahlung*: emission of radiation (a γ particle) at the deceleration of the particles (mainly $e^{\pm}$).
- *Pair production*: a photon splits into 2 particles: matter/antimatter pairs (mainly $e^{\pm}$). This can occur only if the γ has enough energy to create the particle masses.
- *Fluorescence light*: a particle (mainly $e^{\pm}$) excites an air molecules (mainly N_2) losing energy for ionization; when these excited states decay, a photon in the UV range is emitted.
- *Cherenkov light*: light emitted when the speed of a particle in a medium is faster that the speed of light in the given medium.
- *Weak interaction of the particles and electromagnetic decay.*
- *Strong interaction*: interaction between quarks of the CR and the air nuclei.

Through these interactions, a primary CR generates several new particles that can interact with the atmosphere as well. These secondary particles form the so-called air shower. In ground-experiments only these particles are measurable.

A shower develops along its "shower axis", the prolonged track of the primary particle. The profile of the number of particles of this development as a function of atmospheric depth is called *longitudinal profile*. When the "shower front" reach the ground, the "shower core" can be defined as the point where the axis meets the ground. From this core, the density perpendicular to the axis is called *lateral profile*. Distances traveled across the atmosphere are usually expressed in g/cm^2 as traversed column density (slant depth) $X = \int \rho(l)\, \mathrm{d}l$. ρ is the density of the air and the integral is taken along the trajectory across the atmosphere.

The showers have a soft, or electromagnetic, component ($e^{\pm}$ and γ, easily absorbed by the atmosphere) and a hard component (deeply penetrating particles as $\mu^{\pm}$ and neutrinos). The

soft part is produced by electromagnetic cascades (Section 2.4.1), fed by photons from $\pi^0 \to 2\gamma$ decay and the decay of muons in electrons. The hard part is fed by the hadronic cascades. The concept of *Extensive Air Showers* (EAS), introduced by Pierre Auger in 1938/39, is necessary to explain the the muonic component (Section 2.4.2). For hadronic shower component, a detailed description of the hadronic interaction models (Section 2.4.3) is needed.

2.4.1 Electromagnetic cascades

The most simple shower development in the atmosphere is given by an electromagnetic cascade. This can occur when $e^{\pm}$ or γ interact with the atmosphere. A simple scaling model of this shower, called *Heitler model* [31, 32], can be explained with the Fig. 2.6.

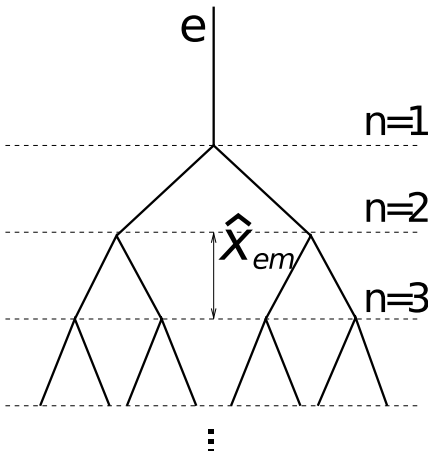


Figure 2.6: Heitler model of an electromagnetic cascade. From [3]

When primaries enter the atmosphere, they start a sequence of bremsstrahlung and pair production processes. Every radiation length electrons and positrons emit photons and photons create e^+e^- pairs. In this model, the radiation length $\hat{X}_{em}$ (about $37\,\mathrm{g/cm^2}$ in air) is the average distance that a particle travels in a medium before interacting. At every step (called *generation*), each particle, in essence, splits in two, sharing the energy equally. This process stops when the energy of the particle is so low that ionization losses become more important than radiative losses:

$$\left.\frac{\mathrm{d}E}{\mathrm{d}X}\right|^{\mathrm{ioniz}}_{E=E_{\mathrm{crit}}} \approx \left.\frac{\mathrm{d}E}{\mathrm{d}X}\right|^{\mathrm{brems}}_{E=E_{\mathrm{crit}}}, \tag{2.8}$$

where X, expressed in $\mathrm{g/cm^2}$, is the atmospheric depth. E_{crit} is the critical energy at which the two loss process are equally important (86 MeV in air). In this approximate model, a primary ($e^{\pm}$ or γ) with an initial energy E_0 will have secondaries ($e^{\pm}$ and γ) at the n-th step with an energy of $E_n = E_0/2^n$. If $E_n = E_{\mathrm{crit}}$ it is easy to estimate the number of generations before the condition in Eq. (2.8) is verified:

$$n_{\mathrm{max}} = \frac{1}{\ln 2} \ln\left(\frac{E_0}{E_{\mathrm{crit}}}\right). \tag{2.9}$$

Using this equation, the point of maximum shower development of the longitudinal profile, X_{max},

and the maximum number of particles, $N_{\max}$, can be defined as

$$X_{\max} = \widehat{X}_{em} \cdot n_{\max} = \frac{\widehat{X}_{em}}{\ln 2} \ln\left(\frac{E_0}{E_{\text{crit}}}\right), \tag{2.10}$$

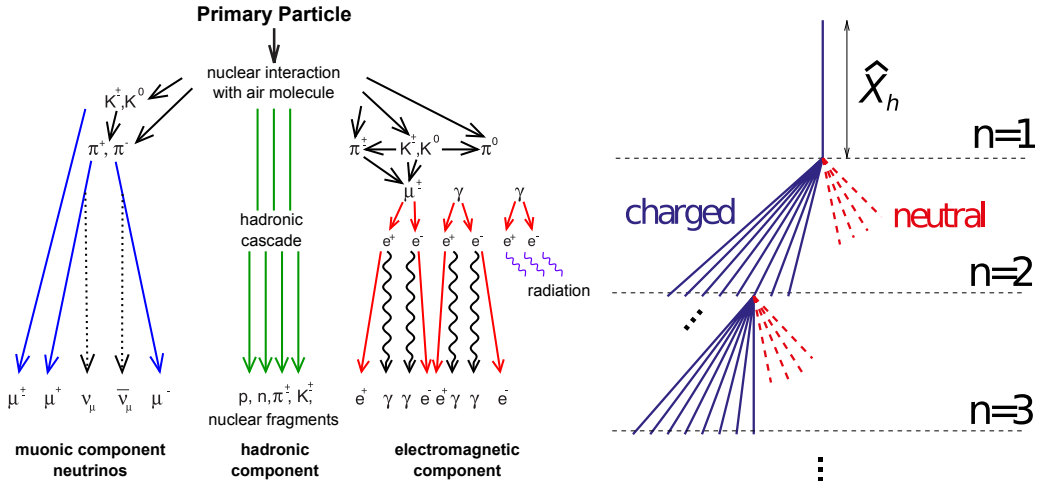
$$N_{\max} = \frac{E_0}{E_{\text{crit}}}. \tag{2.11}$$

As soon as the ionization losses become important, absorption starts to dominate over particle creation. Thus, after the $X_{\max}$ point, the number of particles will decrease.

Due to multiple Coulomb scattering of electrons with the atoms in the air, a lateral spread of the shower particles occurs [33]. The length scale of the lateral distribution of low-energy particles in a shower is characterized by the *Molière radius* $R_M = (21\,\text{MeV}/E_{\text{crit}})\widehat{X}_{em} \approx 9.3\,\text{g/cm}^2$ ($\sim 80\,\text{m}$ at sea level). This is the radius of a cylinder that, on average, contains 90% of the deposited energy from the shower particles.

2.4.2 Hadronic cascades

After Rossi's discovery of coincidences in rather distant counters during his measurement of CRs [34], Pierre Auger introduced the notion of EAS. He had performed measurements of coincidence rates of detectors placed both at sea level and at high altitude (3500 m and 2900 m) with a separation of up to several hundreds of meters [35]. His conclusion was the existence of primary particles with an energy above $10^{15}\,\text{eV}$ creating a multitude of secondaries when interacting with the atmosphere. Moreover he noted that these secondaries cover an extended area on the ground. These are the EAS, Air showers are produced by a cascades, more complicated than the electromagnetic cascade discussed previously.



(a) EAS Scheme (b) Heitler modoel

Figure 2.7: Simple example of the hadronic cascade. From [36] and [3] respectively.

When a nucleus enters the atmosphere a first hadronic interaction occurs, creating kaons, pions and several other hadrons. The initial energy is split among these secondaries. The exact height of the first interaction point is subject to a large fluctuation. This happens because the air has a low density at those heights and the particle interaction is stochastic in nature.

Because of the energies involved, some of these hadrons interact again before decaying. A chain of several strong interactions starts, creating the hadronic cascade. The dominant components in the cascade are pions and kaons. These will decay mostly in photons, muons and neutrinos. Generated $e^{\pm}$ and γ will start also electromagnetic cascades and Cherenkov radiations.

The most frequent secondaries are neutral and charged pions. The neutral pions decay almost immediately ($c\tau \simeq 25\,\text{nm}$) in two photons, forming the main source of the electromagnetic shower component. The charged pions live longer ($c\tau \simeq 7.8\,\text{m}$) and they can interact again before decaying to muons and neutrinos. Also kaons, with a lifetime shorter but comparable to the $\pi^{\pm}$ ($c\tau \simeq 3.7\,\text{m}$), can interact before decaying to muons, neutrinos or pions. Finally the soft and the hard shower components can be observed at the ground level. These hadrons form the *hadronic shower core* (pions, kaons and baryons).

A scheme of a simple Heitler model of an EAS, called Heitler-Matthews [37], is shown in Fig. 2.7. In this simplified model, several assumptions are made. Only pions are produced as secondaries from a single nucleon as primary. The π^0 initiate electromagnetic cascades, as already described. The $\pi^{\pm}$ start the hadronic one. The principle is the same as discussed in Section 2.4.1, but the hadronic interaction length $\widehat{X}_h$ (about $120\,\text{g/cm}^2$ in the air) must be introduced. This length takes into account the inelastic hadronic processes. Moreover a primary particle is not split in only two secondaries anymore. To account for the *multiplicity* of the secondary particles, it is, on average, split into a n_{mult} new particles, each one with the energy E_0/n_{mult}. One third are π^0 (dashed lines in Fig. 2.7(b)): $n_{\text{neut}} = \frac{1}{3} n_{\text{mult}}$. Two thirds are charged π: $n_{\text{ch}} = \frac{2}{3} n_{\text{mult}}$. In this approximation, at the n-th generation the energy is distributed as

$$E_{\text{had}} = \left(\frac{2}{3}\right)^n E_0 \qquad E_{\text{em}} = \left[1 - \left(\frac{2}{3}\right)^n\right] E_0. \tag{2.12}$$

With $n \approx 6$, about 90% of the initial shower energy is carried by the electromagnetic particles and eventually deposited as ionization energy in the atmosphere. The last parameter of this model is the decay critical energy $E_{\text{crit}}^{\text{decay}} (\approx 20\,\text{GeV})$. This is the energy at which the charged pion decay length becomes comparable to the interaction length, in other words, the energy when the decay of pions starts to dominate.

In Fig. 2.8 examples of longitudinal and lateral profiles of a simulated shower is shown. It is evident that the maximum of the shower profile is dominated by the electromagnetic component, which is initialized by the neutral pions. Only the first generation matters, because the energy of the subsequent particles is reduced very fast by the factor $1/n_{\text{mult}}$.

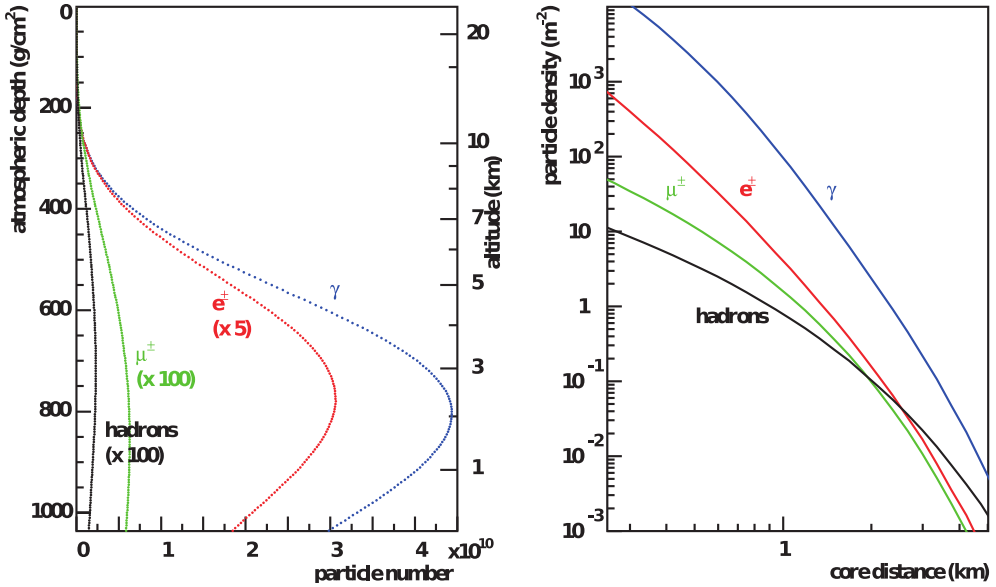
n_{mult} pions are produced at the first interaction point $X_0 \simeq \widehat{X}_h$. Each of them decays almost immediately in two γ. Thus the energy of a photon at the start of the cascade is $E_0/2n_{\text{mult}}$. Therefore the depth of maximum shower development, approximately given by Eq. (2.10), is:

$$X_{\max} \approx \widehat{X}_h + X_{\max}^{\text{em}}(E_0/2n_{\text{mult}}) = \widehat{X}_h + \frac{\widehat{X}_{em}}{\ln 2} \ln\left(\frac{E_0}{2n_{\text{mult}} E_{\text{crit}}^{\text{decay}}}\right). \tag{2.13}$$

With this model it is possible to estimate the number of electrons and muons:

$$\text{using Eqs. (2.11) and (2.12)} \quad \Rightarrow \quad N_{\text{e}} = \frac{E_0}{E_{\text{crit}}} \left[1 - \left(\frac{2}{3}\right)^n\right], \tag{2.14}$$

$$\left.\begin{aligned} E_{\text{crit}}^{\text{decay}} &= \frac{E_0}{n_{\text{mult}}^n} \\ N_\mu &= n_{\text{ch}}^n \end{aligned}\right\} \quad \Rightarrow \quad N_\mu = \left(\frac{E_0}{E_{\text{crit}}^{\text{decay}}}\right)^{\alpha} \quad \text{with } \alpha = \frac{\ln n_{\text{ch}}}{\ln n_{\text{mult}}}. \tag{2.15}$$



(a) Average longitudinal shower profile

(b) Average lateral shower profile

Figure 2.8: Longitudinal and lateral profiles of a simulated proton air showers of $10^{19}\,\mathrm{eV}$. Hadrons, $\mu^{\pm}$, $\mathrm{e}^{\pm}$ and γ are shown separately. The lateral distribution of the particles at ground is calculated for $870\,\mathrm{g/cm^2}$, the depth of the Pierre Auger Observatory. From [3].

The lateral profile of the hadronic cascade is broader than the electromagnetic one. Secondary hadrons are produced at a typical transverse momentum of $p_{\perp} \sim 350 - 400\,\mathrm{MeV}$. This momentum is almost energy independent. Thus the low-energy hadrons are generated with a larger angle relative to the shower axis. In Fig. 2.8(b), an example of the particle density expected at the ground as a function of the distance from the shower core is shown.

To treat nuclei, another model can be taken into account: the *superposition model* [38]. Every nucleus has a binding energy of $\sim 5\,\mathrm{MeV}$ per nucleon, which is much smaller that the typical interaction energy. If it has an atomic mass A and an initial energy of E_0, the shower can be treated as the superposition of A independent air showers from single nucleons, each with a starting energy of E_0/A. Applying this model in Eqs. (2.13) to (2.15) yields:

$$X^A_{\max} = X_{\max}(E_0/A) \approx \widehat{X}_h + \frac{\widehat{X}_{em}}{\ln 2}\left[\ln\left(\frac{E_0}{2n_{\mathrm{mult}}E^{\mathrm{decay}}_{\mathrm{crit}}}\right) - \ln A\right], \tag{2.16}$$

$$N^A_{\mathrm{e}} = A \cdot \frac{E_0/A}{E_{\mathrm{crit}}}\left[1 - \left(\frac{2}{3}\right)^n\right] = \frac{E_0}{E_{\mathrm{crit}}}\left[1 - \left(\frac{2}{3}\right)^n\right], \tag{2.17}$$

$$N^A_{\mu} = A \cdot \left(\frac{E_0/A}{E^{\mathrm{decay}}_{\mathrm{crit}}}\right)^{\alpha} = A^{1-\alpha} \cdot \left(\frac{E_0}{E^{\mathrm{decay}}_{\mathrm{crit}}}\right)^{\alpha}. \tag{2.18}$$

As consequence of this model (Eq. (2.16)), it is possible to use the measurement of $X_{\max}$ to determine the mass of primary particles (Section 2.5).

2.4.3 Hadronic interaction models

The simple models introduced until now are approximations. To get a better description of the hadronic cascades, phenomenological models are built based on the experiments at accelerators and the theory of *Quantum Chromodynamics* (QCD).

Models considered for this work are called **Sibyll-2.1** [39], **QGSJetII-04** [40] and **EPOS-LHC** [41], where the last two were updated recently. In all of them the interaction between the primary particles and air is described with a detailed scattering theory, including different versions of the Glauber model [42]. To treat hadron-hadron collisions and hadronic multi-particle production, a perturbative expansion of QCD is used, together with Regge phenomenology [43]. A hadronic interaction occurs when there is an exchange of gluons between hadrons. In some of these exchanges, a color-connection between partons (gluons and quarks), called *string*, is formed. The energy of the color field leads to the fragmentation of these strings and to the hadronic production of particles [44]. At high energies, the fragmentation of scattered partons at large angles creates *jets*.

A useful model to calculate hadronic cross sections is considering multi-gluon diagrams, named *pomeron*. Multiple interactions correspond to multiple pomeron exchanges, treating them as quasi-particles. This is known as Gribov's Reggeon field theory [43] and leads to good description of collider measurements.

Sibyll-2.1

This is the simplest model of those considered here [39]. It contains a minimum number of assumptions. It is optimized to describe those features of interactions needed for air shower simulation. But it is not suited for the simulation of heavy ion collisions at colliders. Interactions between nuclei are described by the semi-superposition model [45]. It gives only a description of high-energy secondaries in one beam direction. This model is not based on an explicit implementation of pomeron interactions, even though the final states and cross sections are similar.

QGSJetII-04

This is the model with the least number of free parameters [40]. A microscopic model of multi-pomeron interactions and exchanges is implemented and leaves only few free parameters. The version of the model used here has been tuned by the author to describe LHC data.

EPOS-LHC

This model is tuned to describe optimally the accelerator data, especially the new measurements obtained by the LHC experiments [41]. The microscopic pomeron model is explicitly calculated. Moreover, nuclear effects are included, thus EPOS-LHC is suited also for heavy ion physics. For a better description of data, the authors introduced additional parameters to modify the baseline predictions of the model where needed.

2.5 Mass composition

From the air shower phenomenology, the best estimator of the mass of the nuclei is the $X_{\max}$, as shown in Eq. (2.16)[3]. The ideal case would be to measure $X_{\max}$ shower by shower and then, from known parameters, obtain the estimation of A. However the nature of the parameters is

[3] Also the number of muons (Eq. (2.18)) are a good estimator. However this work is focussed on the $X_{\max}$.

stochastic. On average, a particle interacts at $\widehat{X}_i$, or $\widehat{X}_e$ in case of electromagnetic component. Moreover n_{mult} varies shower by shower. Thus it is more realistic to consider $X_{\max}$ distributions.

Equation (2.13) expresses the mean depth of the maximum of the shower profile for a proton ($X^p_{\max}$) and can be considered as a function of $\log_{10}(E/\text{eV})$. Thus, the Eq. (2.16) can be written, for any atomic mass number A_i, as

$$X^i_{\max} = X^p_{\max}(\log_{10}(E/\text{eV})) - f_E \ln A_i, \tag{2.19}$$

where the parameter f_E is approximately constant. However, in a more detailed model [46] this value is actually a linear function of $\log_{10}(E/\text{eV})$ as well.

At a given energy, every nucleus will have an $X^i_{\max}$ distribution with the mean value $\langle X^i_{\max}\rangle$ and the fluctuation $\widehat{\sigma}_i$. The heavier and larger a particle is, the larger is the cross section. Thus the probability to interact is higher and, on average, shallower showers are produced than for light nuclei. Moreover, from the superposition argument, heavy nuclei are characterized by smaller fluctuations (i.e. smaller $\widehat{\sigma}_i$). Therefore, the first two moments of the $X_{\max}$ distribution carry information about the mass composition.

Averaging Eq. (2.19) over the nuclei in the mixture, each with a different fractions w_i in the total flux for a given energy value, the mean of the $X_{\max}$ distribution is

$$\langle X_{\max}\rangle = \sum w_i X^i_{\max} = \langle X^p_{\max}\rangle - f_E \sum w_i \ln A_i = \langle X^p_{\max}\rangle - f_E \langle \ln A\rangle. \tag{2.20}$$

For a mixture of nuclei, each with its fractions w_i, the fluctuation is larger than the sum of the $\widehat{\sigma}_i$ because differences between the averages $\langle X^i_{\max}\rangle$ must also be considered:

$$\widehat{\sigma}^2(X_{\max}) = \sum w_i \widehat{\sigma}^2_i + \left(\sum w_i \langle X^i_{\max}\rangle^2 - \langle X_{\max}\rangle^2\right) = \langle \widehat{\sigma}^2_A\rangle + \left(\langle \langle X^i_{\max}\rangle^2\rangle - \langle X_{\max}\rangle^2\right). \tag{2.21}$$

Thus, using Eqs. (2.19) and (2.20)

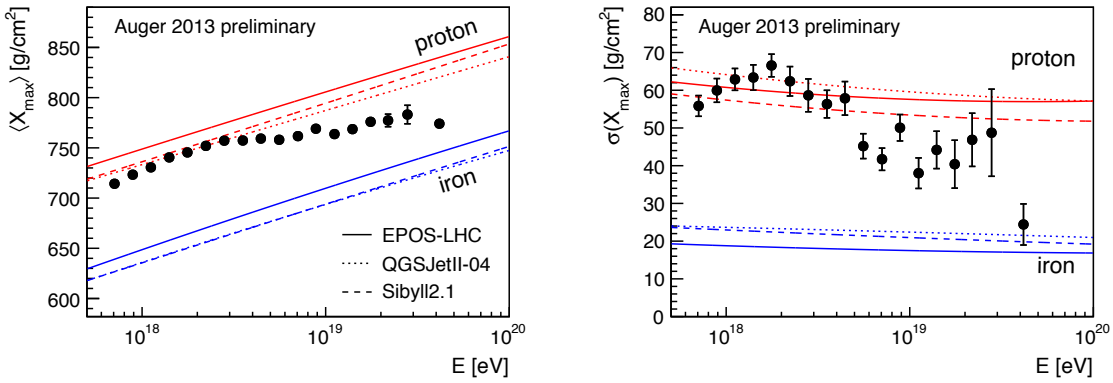
$$\widehat{\sigma}^2(X_{\max}) = \langle \widehat{\sigma}^2_A\rangle + f^2_E \left(\langle \ln^2 A\rangle - \langle \ln A\rangle^2\right) = \langle \widehat{\sigma}^2_A\rangle + f^2_E\, \sigma^2(\ln A). \tag{2.22}$$

The use of Eqs. (2.20) and (2.22) leads to an estimation of the moments of $\ln A$. These moments are a way to interpret the average composition and the degree of mixture in the flux. However, the parameters $\langle X^p_{\max}\rangle$, $\langle \widehat{\sigma}^2_A\rangle$ and f_E, which depend on the $\log_{10}(E/\text{eV})$, can be only known through the use of the hadronic interaction model. Thus, every model may give a slightly different interpretation of the measurements. Details on these parameters and their dependence on the energy are illustrated in Appendix G.

In Fig. 2.9, the status of the $\langle X_{\max}\rangle$ and $\widehat{\sigma}$ measurements (Figs. 2.9(a) and 2.9(b) respectively) are shown, from the preliminary results of the Pierre Auger Observatory presented at the 33rd International Cosmic Ray Conference [47]. The predictions from the most updated hadronic interaction models (summarized in Section 2.4.3) are displayed in case of proton and iron primaries.

These results show a strong dominance of light nuclei at the ankle region ($10^{18.5} \div 10^{19}$ eV). Hence the ankle scenario for the galactic-extragalactic transition is disfavored, because it predicts a heavy composition dominance. Moreover, at the highest energies the rates shown in the figures evolve in favor of an almost pure heavy composition. Thus also the dip scenario is disfavored, because it requires a pure proton flux at such energies.

A mixed composition scenario is the most favored from this measurements. However the data are not yet enough to give a solution to this problem. Moreover a possible discrepancy in the results from the Pierre Auger Observatory and HiRes/TA exists at these energies, as shown in Fig. 2.10. Several attempts to explain the observations are currently ongoing, from a



(a) $\langle X_{\max} \rangle$ measurement

(b) $\widehat{\sigma}$ measurement

Figure 2.9: Preliminary version of the $\langle X_{\max} \rangle$ and $\widehat{\sigma}$ measurements from the Pierre Auger Observatory shown at the 33rd International Cosmic Ray Conference. Also the values from the most updated hadronic interaction models (EPOS-LHC, QGSJetII-04, Sibyll2.1) are shown for pure proton and pure iron compositions. From [47].

project of a joint collaboration to reconcile the discrepancy in the data [48], to the development of composition models able to explain the results in Fig. 2.9. One of the new models predicts the absence of the GZK mechanisms, therefore GZK cut-off and cosmogenic neutrinos produced by the interaction with the CMB (Eq. (2.5)) will not be observed [49]. Another model [50, 51] is based on the injection of a new component in the galactic flux from still unknown sources [14] to extend it toward $5 \cdot 10^{18}$ eV.

Figure 2.10 is the collection of the $\langle X_{\max} \rangle$ measurements until the year 2012 from several experiments [23]: TA, HiRes, HiRes/MIA, CASA-BLANCA, Yakutsk, Tunka and Auger. Also the previous version of the hadronic interaction models are shown. The prediction of the most updated models are more in agreement with each other than their previous versions, especially at high energies. In this figure, the most important issue is the lack of measurements between $10^{17.5}$ eV and 10^{18} eV. Thus, an important piece of information is missing to discriminate the possible models for the cosmic rays sources.

Filling this gap and connecting the data below $10^{17.5}$ eV and above 10^{18} eV is the main purpose of this work. For this, data collected by the Pierre Auger Observatory will be used.

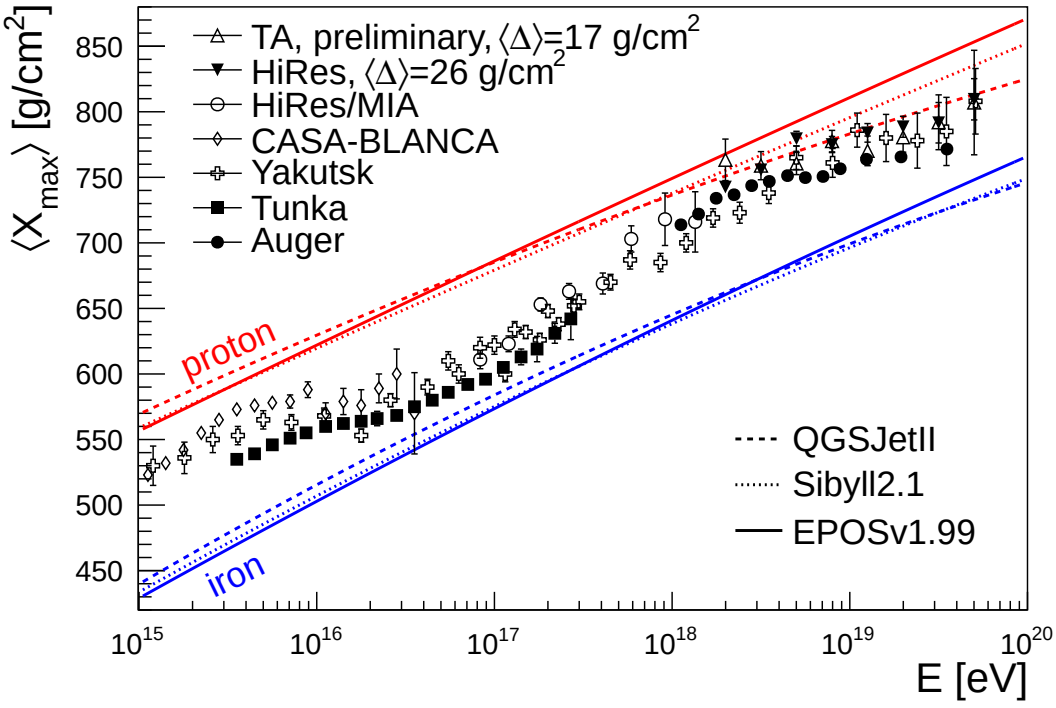


Figure 2.10: Collection of $\langle X_{\max} \rangle$ measurements from 2012. Older versions of the hadronic interaction models, compared to Fig. 2.9, are shown (EPOSv1.99, QGSJetII, Sibyll2.1). From [23].

CHAPTER 3

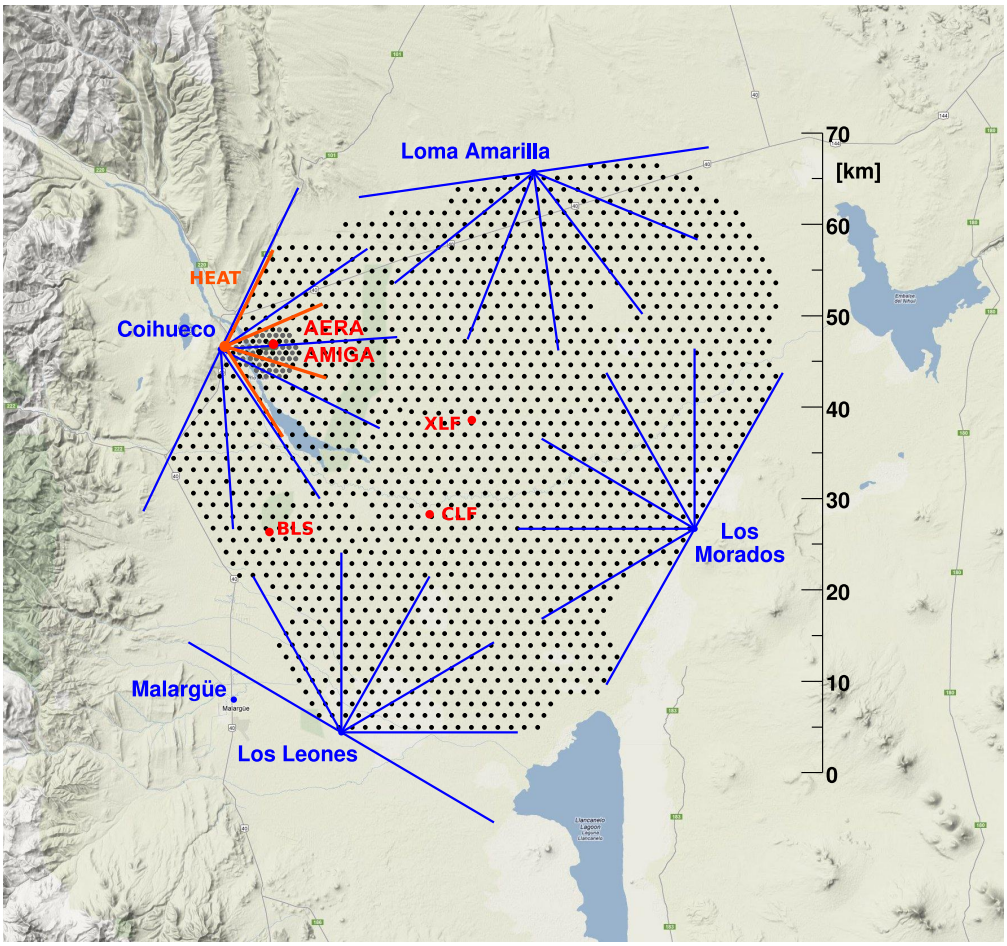
The Pierre Auger Observatory

The Pierre Auger Observatory is located in the Argentinian Pampas. Originally it was designed to measure the energy spectrum, the distribution of arrival directions and the chemical composition of *Ultra High Energy Cosmic Rays* (UHECR): from 10^{18} eV to beyond the GZK limit [52]. At these energies the rate of Cosmic Rays (CRs) is low. A large area of detection is needed to measure them with high statistical significance. Thus, it was decided to build the observatory on a 3000 km^2 large area in the Southern hemisphere to have the galactic center in the field of view.

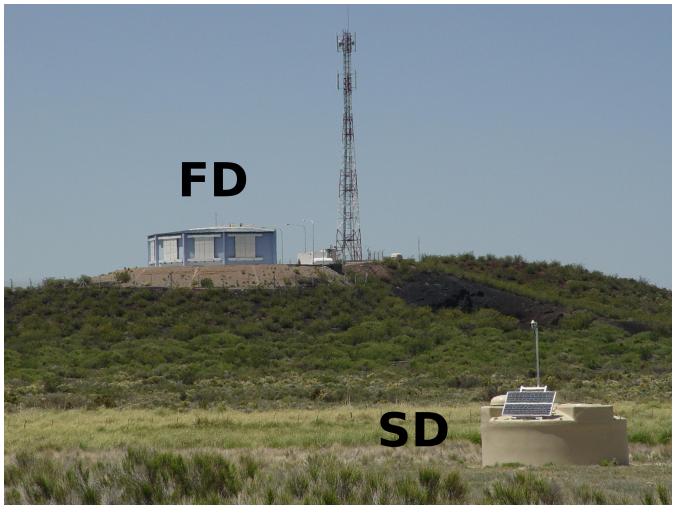
To detect UHECR, the Pierre Auger Observatory was designed as "hybrid detector". Two different kind of detectors were built: the *Surface Detector Array* (SD) and the *Fluorescence Detector* (FD) (see Fig. 3.1). The SD is an array of tanks filled with water and arranged on a triangular grid with a 1.5 km spacing. They sample the density of secondary particles on ground. The FD is a set of fluorescence telescopes grouped in four sites at the border of the experiment area, which is covered by their field of views. They measure the fluorescence light released in the atmosphere during the development of air showers. In the shower reconstruction information from both the detectors can be combined to better measure the geometry and the energy. These combined events are called *hybrid events*.

The atmosphere is an important variable in this experiment. It must be monitored in density, cloud coverage and aerosol content. There are two stations called *central* and *extreme laser facility* (CLF and XLF) dedicated to this task. They are equipped with high-power lasers at two different frequencies to measure the aerosol content of the atmosphere. Radiosonde balloons launched from the *Balloon Launch Station* (BLS) probe the humidity column density. Furthermore there are *LIDAR* stations and cloud cameras in each of the four FD installations. The first probe the clouds with lasers and the second by imaging. Finally, the Pierre Auger Observatory uses the *Global Data Assimilation System* (GDAS), that combines measurements and forecasts from numerical weather prediction to provide data for the whole of the globe.

The Pierre Auger Observatory is taking data since 2004, but was fully completed in 2007. Since then, several enhancements have been installed. SD and FD were improved to detected



(a) Auger map.



(b) A view of an FD site (LL) and an SD station.

Figure 3.1: The Pierre Auger experiment map (from [53]) and a picture of a FD and a SD stations. In the map, the water Cherenkov detectors are shown as black dots, in grey the Infill and HEATlet extensions (AMIGA). The telescopes of the FD and their field of view are shown in blue, in red the HEAT extension. CLF, XLF, BLS and AREA are also shown (red dots).

CRs at lower energies. These upgrades are the *High Elevation Auger Telescopes* (HEAT) for FD, and *Infill* and *HEATlet* for SD.

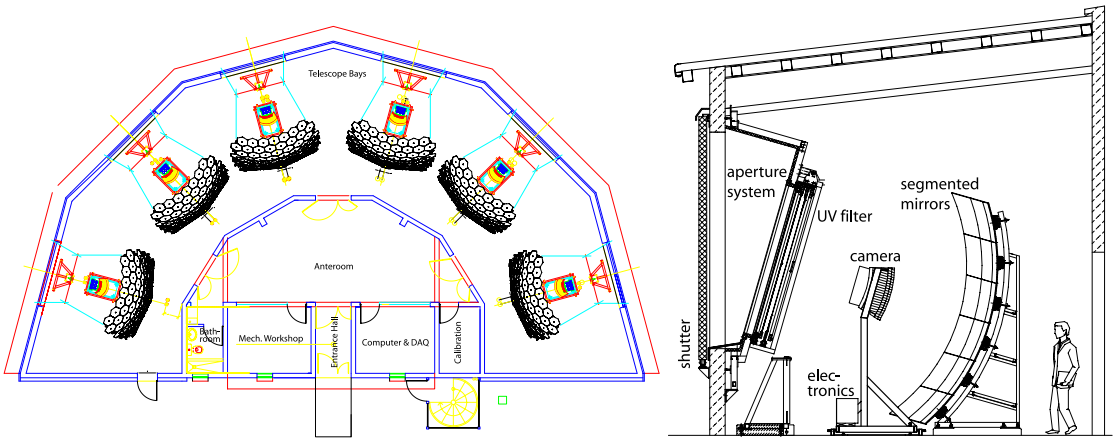
Further upgrades were installed after the construction of the observatory was finished. They are meant to extend the possibilities to study cosmic rays and extensive air showers: *Auger Engineering Radio Astronomy* (AERA) to detect showers with radio signals, and *Auger Muons and Infill for the Ground Array* (AMIGA), which includes Infill and HEATlet stations, to detect underground muons.

Important for this work is the fluorescence light, detected by the FD and HEAT upgrade. In Section 3.1 these telescopes will be discussed and the atmospheric influence on the fluorescence light detection will be reviewed. In Section 3.2 a brief description of the surface detectors and their extensions will follow. Finally, in the Section 3.3, the reconstruction of hybrid events will be introduced.

3.1 Fluorescence Detector

3.1.1 Telescopes

The fluorescence detector consists of 27 fluorescence telescopes. 24 of them, termed also "standard telescopes", are part of the original design of the Pierre Auger Observatory [54]. They are grouped at four different sites, around the area covered by SD stations. The locations are called *Los Leones* (LL), *Los Morados* (LM), *Loma Amarilla* (LA) and *Coihueco* (CO), as shown in Fig. 3.1(a). In each location there is a building with 6 bays. Each bay hosts one of the telescope system (see Fig. 3.1(b) for a picture of an FD building and Fig. 3.2 for the schemes).



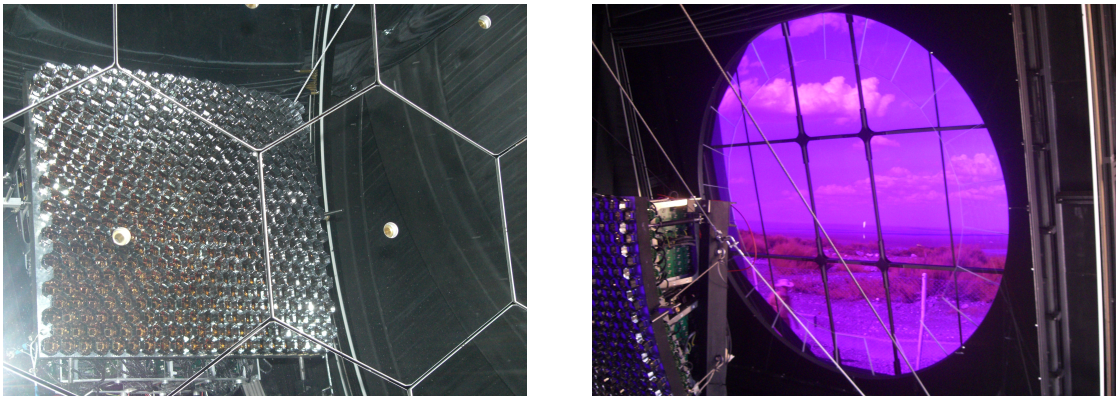
(a) Scheme of an FD building.

(b) Scheme of an FD bay and its telescope.

Figure 3.2: Illustration of a fluorescence detector installation (from [54]).

Each telescope system is a Schmidt optic system, with a segmented spherical mirror of $10\,\mathrm{m}^2$ surface and a camera of 440 *Photomultiplier Tubes* (PMTs) with an hexagonal surface. The PMTs are arranged in a grid of 20 rows by 22 columns. Each one is a *pixel* of the camera (Fig. 3.3(a)). Particles produced by an air shower can deposit a small part of their energy by exciting the air molecules in the atmosphere[1]. When the molecules de-excite, they emit photons

[1]The secondary particles that can excite the air nuclei, are mainly $e^{\pm}$. The air molecules of relevance for the



(a) Mirror with the camera reflection.

(b) UV filter and corrector ring at the bay aperture.

Figure 3.3: Pictures of telescope system details.

in the UV range, so-called "fluorescence light". This light detected by the telescopes is focussed into the cameras by the mirrors. The pixels digitize the light pulses every 100 ns.

The light is received through an aperture in a bay with a diameter of 2.2 m. The aperture is equipped with a "corrector ring", to correct for spherical aberrations at the edge of the aperture, and an UV filter (Fig. 3.3(b)). This filter suppresses the large background of visible light that is present even in good conditions of observation. Moreover, each bay is equipped with a shutter system to close the aperture during strong light periods (mainly daylight and full moon) or unfavorable weather conditions.

If a PMT receives too much light, they might broke. Also strong wind or rain can damage the detector. So a remote control is used to activate the shutter electronically. The remote controlled operation, so-called "slow control", refers to a "Central Campus" (i.e. the observatory building in Malargüe). Under normal conditions, the slow control system is used to power up the cameras, open the shutters and send commands to the calibration systems. If the light background of a camera or a single pixel is too high or the weather changes so that FD operations are not possible anymore, the shutters are closed. In case the shutter fails, a curtain can be dropped inside the aperture.

The *Field of View* (FoV) of each pixel is about 1.5°×1.5°. A single telescope covers about 30°×30° with a range from 1.5° to 31.5° in elevation. A single building has 180° of azimuthal coverage. The combined FoV of the telescopes viewing an air shower track defines the "geometrical FoV".

The FD is only operated in clear, moonless nights to not overwhelm the PMTs with the background light. To operate the FD safely, wind speeds have to be moderate at most and no rain is required. These conditions can differ across the large array. So one site might not take data due to unsuitable conditions while the others operate normally. The duty cycle of all the FD amounts to ≈ 13 % [55]. See Section 3.1.2 for further details.

HEAT

The latest three installed fluorescence telescopes of the Pierre Auger Observatory are the "High Elevation Auger Telescopes" [56]. They are an FD extension that started in 2008 and are stably

fluorescence light emission are mainly N_2.

(a) HEAT in downward position. (b) HEAT in upward position.

Figure 3.4: Pictures of HEAT in both positional mode.

taking data since June 2010[2]. Thanks to the HEAT electronic system, its pixels digitize the light pulses every 50 ns, i.e. at twice the speed the standard telescopes are read out.

HEAT is built about 160 m away from CO. The bays are not enclosed in one building as in case of the standard FD (shown in Fig. 3.2), but they are independent as shown in Fig. 3.4. Every single bay can be moved independently in two positions: "downward" (Fig. 3.4(a)) or "upward" (Fig. 3.4(b)).

In downward mode their FoV is equal to the standard FD. In this mode, the FoV approximately overlaps the FoV of the Coihueco FD[3]. This mode can be used as cross check of CO measurements and has proven to be very useful.

In upward mode, the FoV covers the elevation range from 31.5° to 61.5°. This mode is the main reason of the HEAT upgrade. Low energy cosmic rays have a faster shower development of the shower and the fluorescence light is less bright. To detect them, they must be near to the telescopes. Therefore their detection is not favored with the standard telescopes, because for close showers their FoV covers a reduced area. After the selection of high quality showers, the lower official energy limit measurable with the FD is $10^{17.8}$ eV [47]. Hence the HEAT upgrade was commissioned to "look upward" and extend the FoV to efficiently detect nearby showers. At lower energies the flux of CRs is expected to be high, therefore only three telescopes in one location were deemed sufficient.

In Fig. 3.5 both the downward and upward FoV are shown. Also two examples of shower profiles are shown. They are the same profile only with two different positions and inclinations. When the shower is far away, the standard FD FoV is sufficient to measure the longitudinal profile. However, when the shower is closer, only part of the shower development will be inside the downward field of view. Only the trailing edge is observed, thus it will be a poor quality detection. The second part of the profile can be detected by the upward FoV. Also in this case the event may be of poor quality because the profile is not complete. The solution is combining both

[2]The HEAT first run was in the end of January 2009. But several upgrades and cross checks of the system were done. The data acquisition is considered officially stable from June 2010.

[3]Referring to Fig. 3.2(a), the bays are numbered $1 \div 6$ from right to left. With the same counting orientation, the bay 1 of HEAT corresponds to bay 4 of CO, the HEAT-2 is in-between of bay 4 and 5 of CO, and the HEAT-3 overlaps with CO-6. See also Fig. 3.1(a).

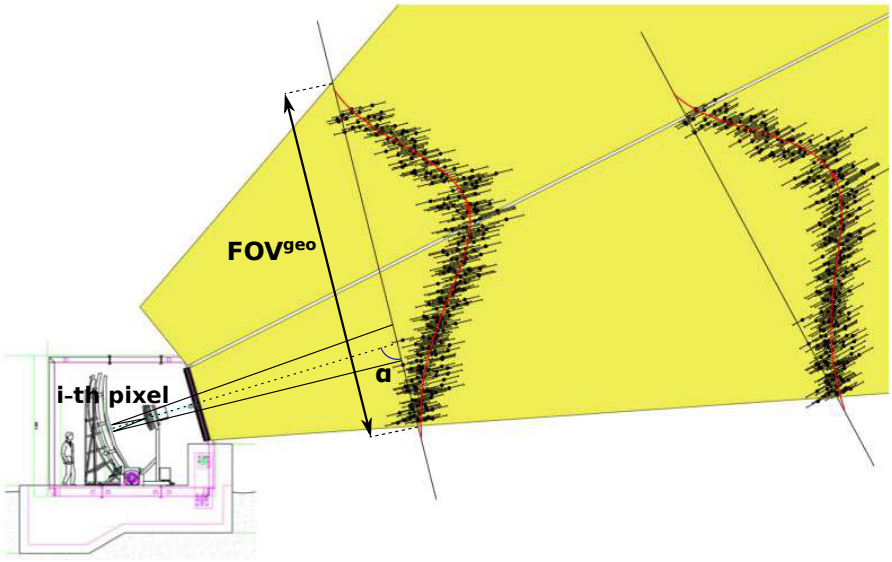


Figure 3.5: Downward and upward fields of view that compound the virtual location HeCo, and how longitudinal shower profiles are seen from the different FoV. $\mathrm{FOV}^{\mathrm{geo}}$ is the geometrical FoV of HeCo. Picture modified from [57]

fields of view into an extended geometrical FoV in elevation (referred in the figure as $\mathrm{FOV}^{\mathrm{geo}}$). This is done during the reconstruction process. A virtual FD is created merging HEAT and CO measurements: the HeCo telescopes. This virtual location will be further explained in Section 3.3.

3.1.2 Data acquisition system and trigger levels

The *Data Acquisition system* (DAQ) of a camera has a front-end containing one analog board for each column of photomultipliers. This board constantly reads out the 22 PMTs. It digitizes the signals from each channel with the following sampling rates: 100 ns (10 MHz) for the standard FD and 50 ns (20 MHz) for HEAT. When a pixel has a signal higher than a certain threshold, it is recorded as triggered. This is the *First Level Trigger* (FLT). During the data taking, this threshold is adjusted to obtain an FLT trigger rate of 100 ns.

Each FLT sends the signals of marked pixels to a *Second Level Trigger* (SLT) board. This trigger searches patterns of 5 pixels similar to those shown in Fig. 3.6. During the data acquisition, some tracks will not pass through every pixels center. Thus some PMTs along the track may not record enough light to trigger. Taking this into account, the minimum number of triggered pixels to pass the SLT is four, and the possible number of patterns to check becomes 108. Passing this trigger means to have a track. The FLT information and the GPS time stamp will be stored on a dedicated PC, the "Mirror PC".

The stored information will be tested by an algorithm. This algorithm is designed to identify tracks generated by lightnings, muons hitting the camera and random coincidences. This is done using pattern recognition based on a shower library obtained from real data. If such tracks are found, the event will be discarded. This is the *Third Level Trigger* (TLT). Approximately 94 % of all background events are correctly identified by the TLT. The fraction of true showers rejected by the trigger is below 0.7 % [54].

Every installation has a common dedicated PC, the "Eye PC". It is connected to all the Mirror

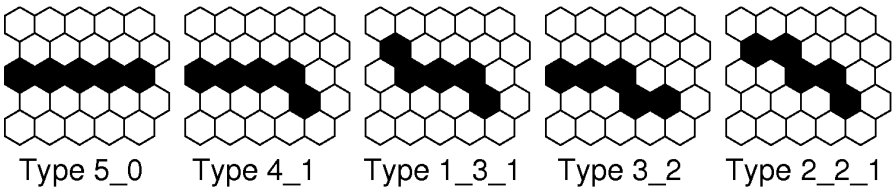


Figure 3.6: Patterns searched by the SLT (from [54]).

PCs from the same FD location. If the TLT is passed, the Eye PC will merge the information from each Mirror PC. This process forms the "FD Event". Moreover one more trigger is formed: the "Hybrid Trigger", also termed T3.

The T3 algorithm does a preliminary reconstruction. It calculates approximately the shower direction and the expected time of the shower front at the ground. This trigger is sent via the wireless network communication to the *Central Data Acquisition System* (CDAS) on the Central Campus and will act as an external trigger for the SD. From the CDAS, a request is sent to the SD for signals recorded within a window of 300 µs to this expected time. If an SD station is found near the FD building[4] that formed the T3, its signal is read out. The FD and SD data are considered to belong to the same Hybrid Event. They will be merged as a Hybrid Event offline.

The wireless communication system of the SD has a limited bandwidth. Thus a superfluous T3 rate from any FD site will cause loss of data packets in the SD communication network. To prevent this, a T3-veto is implemented, which suppresses excessively high rates of T3 triggers to be sent to CDAS and used as SD trigger. The main conditions that veto the T3 are:

- LIDAR activities (preventing a T3 formed accidentally by a LIDAR shot);
- too old event ($> 10\,\mathrm{s}$);
- high rate input limitation from the Eye PC to CDAS;
- high rate input remotely rejection from CDAS.

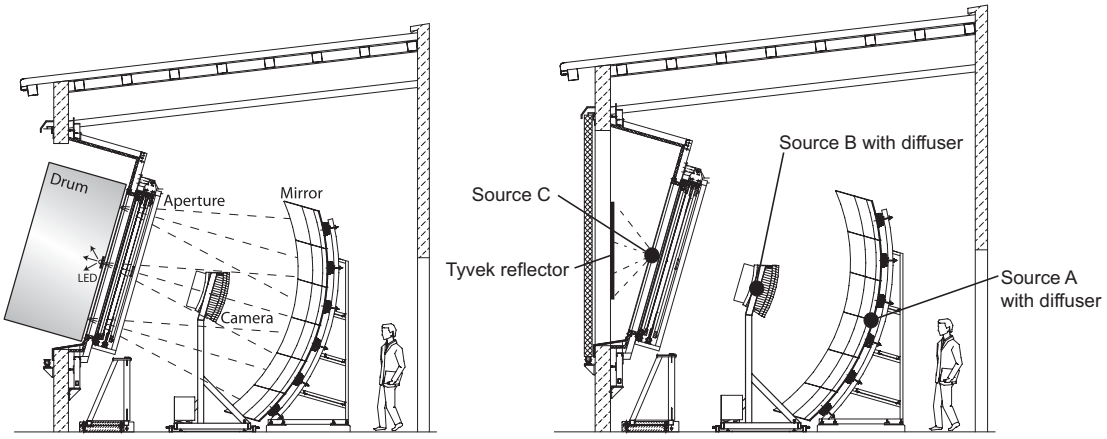
Implemented for HEAT, one further FD trigger exists: the T4. Its algorithm is designed to identify the Cherenkov light fraction, which is strong in the upward mode. Events with too much Cherenkov light are rejected.

As already explained, the FD duty cycle is $\approx 13\,\%$. This is because the DAQ is not always capable of acquiring data. As already anticipated, the telescopes are in dead time during the day and during bad weather conditions. Other sources of dead time are moon passages, full moon nights, activity of atmosphere monitoring instruments, FD calibrations, readout of the electronics, and any hardware or software problem. Most of these are usually a whole night long period. But some of them, the readout of the electronics for example, are short dead times. Thus, it is useful for the reconstruction (Section 3.3) and the analysis (Section 4.2) to define the "uptime". This is the fraction of the time in which a camera is in DAQ in a given period. The uptime is sampled every 10 minutes, camera by camera.

3.1.3 Calibration

To convert the ADC counts from the PMT to light flux at the telescope aperture, several effects must be taken into account for each pixel. For example, geometries acceptance, optical system

[4] An FD location covers approximately 1/4 of the whole sensitive area. Stations are considered "near" to a given FD, if they belong to its covered region.



(a) Absolute calibration with the Drum.

(b) Relative calibration with A, B and C sources.

Figure 3.7: Scheme of the calibrations techniques for FD (from [54]).

properties and electronic response. These contributions were studied on prototypes, but cumulative factors during the experiment can change the response of a pixel. Thus, calibrations are constantly needed. For long term changes, the absolute calibration is done year by year. However during the night the pixels can be influenced by different local factors (PMT heating, external temperature, etc.). Therefore, also relative calibrations are done every night, before and after the data taking [54]. In Fig. 3.7, schemes of these two calibrations are shown.

The absolute calibration uses a calibrated light source (known as the "drum") of 2.5 m of diameter at the telescope aperture (Fig. 3.7(a)). This source provides the same flux of light to each pixel simultaneously. The known flux from the light source and the response of the acquisition system give the required calibration for each PMT. Each telescope is drum-calibrated about once per year. The drum calibrator can also test the response of the PMTs to different wavelengths. As cross check, an independent absolute calibration is done with a roving laser. It shoots a beam of light from a distance of 4 km from the telescope vertically into the atmosphere. The response of some of the camera pixels can be studied.

The relative calibration is done with three light sources every night, before and after the data taking. They are permanently installed in three different positions of the telescope and called A, B and C (Fig. 3.7(b)). The A source is an LED with a diffusor, mounted in the center of the mirror and directly pointed at the camera. The source B is a Xenon flash lamp with a diffusor, mounted on the side of the camera and directed to the mirror. The calibration C uses also a Xenon flash lamp, but with an interference filter that allows to test different wavelengths. Each shutter has a Tyvek sheet, which is illuminated by the source C when the shutter is closed. This sheet reflects the light to the mirror.

The campaign of the drum calibration made in January 2010 has a significant reduction of the uncertainties and systematics of the calibration of the drum [58, 59]. To propagate this new feature back and forward in time, the calibration A is used. For this work, a cross check on this calibration with data is done in Section 4.1.2.

3.1.4 Atmospheric monitoring

When fluorescence light is emitted by the air molecules, it propagates through the atmosphere. As the light travels to the telescopes, scatterings and absorptions can occur. They must be considered properly in order to estimate the correct energy of the shower. For this reason, all ground-based cosmic ray observatories that use fluorescence telescopes have atmospheric monitoring programs of some kind. The Pierre Auger Observatory employs a unique and very extensive monitoring program.

Molecular and aerosol scattering, and cloud coverage are important characteristics properties of the atmosphere. The influence on the light transmission in the near UV of the scattering due aerosol particle is constantly monitored. It is done by the *Central Laser Facility* (CLF) and the recently installed *Extreme Laser Facility* (XLF). The CLF is situated approximately equidistant from LL, LM and CO. The XLF is more dedicated to LA. The cloud coverage is monitored by LIDAR stations and "cloud cameras", both mounted in each of the four FD sites. The LIDAR stations shoot a laser beam periodically above the FD field of view, probing clouds. The cloud cameras monitor clouds with infrared imaging. Information about the cloud coverage are also provided by the *Geostationary Operational Environmental Satellites* (GOES) as infrared imaging [60]. Moreover, temperature, pressure and humidity vertical profiles of the atmosphere are constantly monitored on the site of the Pierre Auger Observatory. These parameters are measured combining data from five weather stations, radiosonde balloons and the *Global Data Assimilation System* (GDAS). Balloons are launched from the *Balloon Launch Station* (BLS). GDAS combines measurements and forecasts from numerical weather prediction to provide data for the whole of the globe every three hours [61, 62]. In Fig. 3.1(a), of CLF, XLF and BLS locations are shown.

The light in the atmosphere is attenuated by molecular and aerosol scattering[5]. The molecular attenuation T_{mol} is given by Rayleigh scattering depending on the wavelength. This factor can be determined analytically once the temperature, pressure and humidity are known.

The aerosol attenuation T_{aer} is based on the Mie scattering theory. This theory relies on the assumption of a spherical scatterers. However this condition is not always fulfilled. Therefore, the knowledge of the aerosol transmission factor depends on the field measurement of the *vertical aerosol optical depth* VAOD(h), a function of the vertical hight h:

$$T_{\text{aer}}(h, \theta) = \mathrm{e}^{-\text{VAOD}(h)/\sin\theta}, \tag{3.1}$$

where θ is the angle between the telescope and the point where the scattering took place. The measurements are performed hourly using the CLF and XLF in each FD site and normalized with "reference clear nights". In such nights, also referred as "Rayleigh nights", the attenuation due to aerosols is minimal and the scattering is dominated by the molecular part.

The study of the reference night is a careful analysis that requires time. Thus the VAOD measurement are not always immediately available for each night. The reconstruction algorithm uses these measurements, so-called "measured Mie model", whenever possible. When such measurements are not available, a parametrization of the measured average of the aerosol content profile is used, termed "parametric Mie model". The most updated VAOD study is found in [63].

Thanks to the calibrations, the amount of light received by the cameras is measured from PMT signals. This must be corrected by the factor $1/T_{\text{mol}}$ and $1/T_{\text{aer}}$. Finally, the fluorescence yield Y_{air} converts light to deposited energy in the atmosphere. The light yield also depends on atmospheric conditions. Its measurement and dependences on temperature, pressure and humidity are from the work of the AIRFLY collaboration [64, 65]. Several experiments have

[5] Photons scattered so much as to not be able to contribute to the fluorescence signal are considered "absorbed" by the atmosphere.

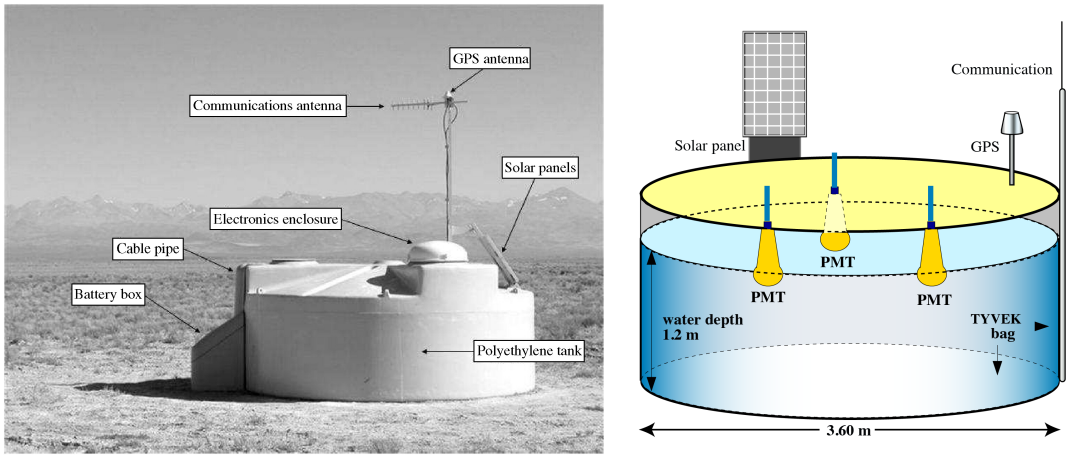
measured Y_{air}, but the Pierre Auger Collaboration considers the one from AIRFLY the most robust and precise [66, 67].

3.2 Surface Detector

3.2.1 Water Cherenkov Detector

More than 1600 "water Cherenkov detectors" (or tanks) form distributed uniformly on a triangular grid with a spacing of 1.5 km over the whole of area of the experiment. They are the Surface Detector array [68].

Each tank contains ~ 12000 L of highly purified water. The water is contained in a reflective shell (in Tyvek), which is surrounded by a plastic container. This external layer grants mechanical stability and insulation from the outside light. On top of the reflective shell, three photomultipliers are mounted, pointing downward in the water. The electronic equipment is powered by a solar panel and a battery. This grants the autonomy of the system. The read-out of the data is performed via a wireless network. Refer to Fig. 3.8 for a schematic structure.



(a) External view.

(b) Internal scheme.

Figure 3.8: Structure of the Auger water Cherenkov detector. External view (from [68]) and internal structure (from [69]).

The tanks are called "water Cherenkov detector" because they use water as a medium for Cherenkov light emission from the secondary particles that travel through the detector. The particle speed exceeds the speed of light in the water and the Cherenkov effect occurs, producing light. The Tyvek coating reflects the light. After few reflections, the light is isotropized and collected by the PMT. The light collection can be described as an exponential decay with a constant of ≈ 65 ns [70].

The amount of light is proportional to the energy of $\gamma/e^{\pm}$ and track length of $\mu^{\pm}$ in the tank. Electromagnetic particles are usually completely absorbed and their contribution to the signal is independent from the zenith angle of their incoming direction. But the muons are able to pass through the tank regardless to the direction, so the amount of light depends of their zenith angle. Each station is calibrated internally by ambient atmospheric muons. The mean signal of such muons defines the unit of 1 VEM, *Vertical Equivalent Muon*.

The main advantage of the SD is a duty cycle of almost 100 %. Measurement time is not limited by external factors but only by the shut down of single stations. Reasons can be power outages, electronic failures, various damages, decommissioning, upgrade periods, etc.

A drawback of the SD is that it is not possible to measure the longitudinal profile and its $X_{\max}$ directly. Moreover the energy of the primary particle cannot be determined accurately. The stations only sample the air shower at ground level, which highly depends on the development of the shower in the atmosphere. However, the Pierre Auger Observatory is a hybrid detector. The SD energy scale can be fixed using shower events that were recorded by the SD and the FD as hybrid events.

SD upgrades

In the 2007 the first SD upgrade started: the Infill area. The array was supplemented with 42 more tanks to fill an hexagonal area near the Coihueco/HEAT site. They are placed halfway between two other tanks, reducing the distance between tanks to 750 m. This improvement lets SD measure showers with lower energy, down to about $10^{17.5}$ eV [71]. The Infill area was completed and fully operational since the 2010.

In 2011, as second upgrade, 7 more stations similarly spaced as Infill were installed to fill a region even nearer to HEAT. This area is called HEATlet and is fully operational since 2012.

The total area Infill+HEATlet is 23.5 km^2 and it is shown in Fig. 3.9. The color scale shows the number of event detected by the HeCo in hybrid with the relative tank. Stations of the Infill and HEATlet have less counts compared to standard SD because the events are more distributed among the stations.

In the picture the third upgrade is also shown: AREAlet[6]. This upgrade results in an even more dense area of six tanks: 5.9 km^2 with a spacing of 433 m between the tanks. This area is meant to improve the data for the radio measurement from the AERA stations [72].

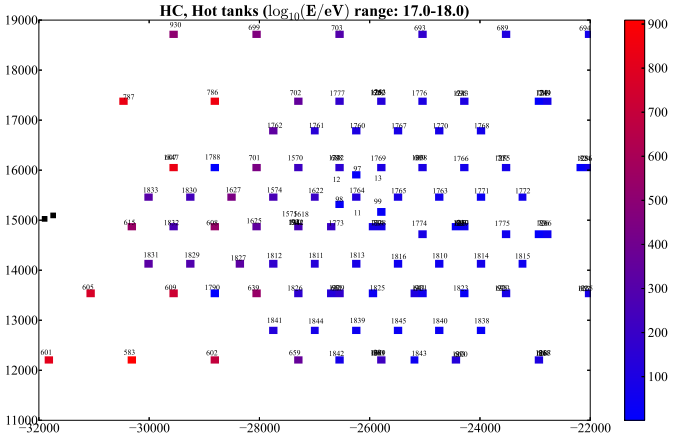


Figure 3.9: Map of the SD in the infill+HEATlet region and their event counts in hybrid with HeCo (color scale). An event is counted only for the stations closest to the shower core. Black squares are Coihueco and HEAT locations.

[6]In Fig. 3.9 only three AERAlet stations are visible, because, at the time of this figure, they were the only operative detectors of this upgrade.

3.2.2 Trigger levels

When data are acquired, they are sent to the CDAS via wireless network. However the limited network bandwidth does not allow to read out each detector in real time. The network allows to transmit the data of roughly one event per hour and tank.

Not all events are air showers, but many low-lever trigger are caused by atmospheric muons. This background must be suppressed. Meanwhile as many real air shower events as possible must be kept. To do so, the SD electronics and software provide a hierarchical trigger system [73, 74].

In the first step, shower candidates are selected by the local trigger, done autonomously by each station. Two different triggers are used for this step. The first is called *Threshold trigger* (TH). It is designed to catch short, sharp peaks created in the *Flash Analogical Digital Converter* (FADC) trace mainly by muons. The TH has two levels: TH-T1 and TH-T2. The first level catches muons from air showers as well as atmospheric muons and stores the information in a local buffer for 10 s. The T2 is applied in the station controller to reduce the rate of events per detector and cope with the bandwidth of the communication system between the detectors and the "Central Campus".

The second trigger is called *Time over Threshold trigger* (ToT). It is designed to catch the signal from the electromagnetic part of the shower, which is more spread out in time. The ToT is very efficient to reject the signals from single muons of the background.

In the second step, both the TH-T2 and ToT are passed to the CDAS. They are combined to form the the third level of trigger T3, requiring spatial and temporal coincidence. The neighborhood of the station with ToT or TH-T2 passed is checked. If there are nearby stations with ToT or TH-T2 passed too, their data are read out by CDAS and stored at the Central Campus. The SD Event is formed.

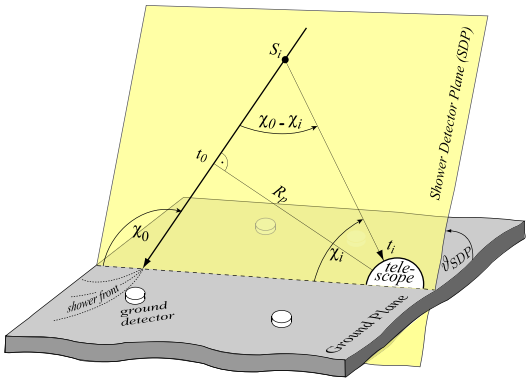
However, the array is not fully efficient for events at energies below $3 \cdot 10^{18}$ eV. Thus, they do not often generate an independent T3 trigger. Hybrid events at these energies occur within 20 km of the FD buildings and usually do not trigger more than one or two SD stations. For these events the external T3 generated by the FD Events is used. They will also form the Hybrid Events.

For the analysis of the SD, a fourth (T4) and a fifth (T5) level of triggers exist. T4 is also called the *physical trigger*. It is designed to select real showers and rejects background muons still present in large numbers at this stage. The T5 is also called the *quality trigger*. It is designed to select showers with great accuracy in the reconstruction of the core position and the energy. These two levels are not important for this work, based on the FD measurement reconstructed as hybrids.

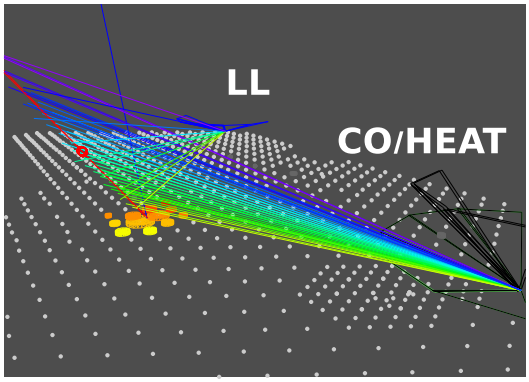
3.3 Hybrid reconstruction

In the CDAS, thanks to the T3, a hybrid event is recognized from the trigger point of view. As a single event, it will be formed offline from the data acquisition, merging FD Events and SD Event from the same shower. As explained in Section 3.2.2, the SD generate T3 only if at least three stations are triggered. The hybrid events with this independent SD Event are called "Golden Hybrid". If the SD do not generate the T3, the event can still be recovered as a hybrid event thanks to the external trigger generated by a FD Event. Hybrid events with at least one SD station are called "Brass Hybrid". For these definitions, all the hybrids are brass and some special candidates are also golden. In this work the term "brass" will refer to events formed with the FD dependent T3, to distinguish them from the golden.

Because their FD-independent nature, the golden hybrid events are also used to calibrate the energy of the SD. For this work, the entire hybrid data set (brass+golden) will be used.

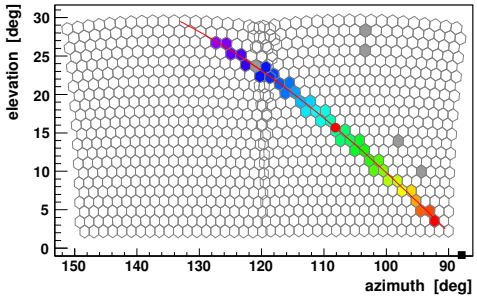


(a) Scheme of the shower geometry. Index i referes to the i-th pixel.

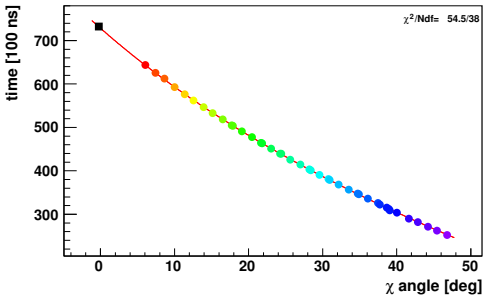


(b) Geometry reconstructed for a stereo event (AugerID 201217007204).

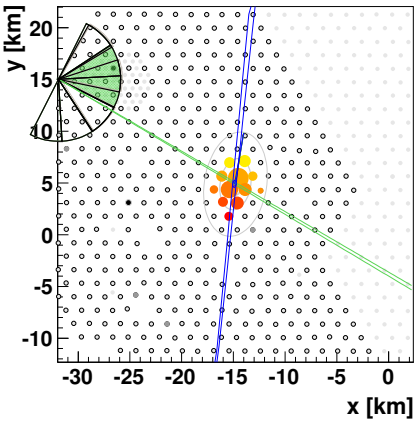
Figure 3.10: Geometry of a shower.



(a) Track of a FD Event.



(b) $\chi - t$ plane.



(c) Distribution of an SD Event.

Figure 3.11: FD and SD Event measurement and their projections in $\chi - t$ plane. Details in Section 3.3.1 (AugerID 201217007204).

The reason is the poor efficiency in the independent T3 formation for events at energies below $3 \cdot 10^{18}$ eV. The statistics for golden events at low energies is strongly reduced.

The merged hybrid events are not yet useful for the analysis. They must be reconstructed to obtain quantities related to the physics of the air shower such the geometry and the profile.

3.3.1 Geometry reconstruction

Through the FD track on the camera, the *Shower Detector Plane* (SDP) is determined. It is defined by the location of the FD building and the projection of the shower axis on the camera. In Fig. 3.10(a) a scheme of the SDP and the variable used is shown. From the pixel track alone, the position of the shower axis within this plane can not be determined. It has to be calculated from the viewing angles (χ_i) and detection times (t_i) of single pixels.

In Fig. 3.11(a) the shower track on the camera of the FD event is shown. The color scale of the triggered pixels corresponds to the time projection (purple to red). The fitted track is the red line, and the red dot is the measured position of $X_{\max}$. From the FD, the time t_i and angle χ_i are measured for each pixel i. Their projection in the $\chi - t$ plane is shown in Fig. 3.11(b). The fit of these data point through Eq. (3.2) leads to the geometry reconstruction.

$$t_i = t_0 + \frac{R_p}{c} \tan\left(\frac{\chi_0 - \chi_i}{2}\right). \tag{3.2}$$

If only FD data are used, the uncertainty in the arrival direction and core position is very large. Thus, the distance of the axis to the telescope is not known precisely. Therefore the attenuation of the fluorescence light cannot be estimated correctly. This leads to a large uncertainty in the calorimetric energy. However, the shower core is measured from the SD. In Fig. 3.11(c) the signals on the stations are shown. The size of the dots is proportional to the signal in the respective station and the color scale to the arrival time of the shower front (yellow to red). From this event the core position is determined as the station with the highest signal: the "hottest tank". This tank is projected in Figs. 3.11(a) and 3.11(b) as a black square.

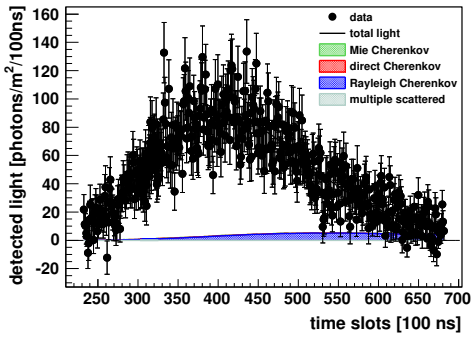
The start time of the signal in the hottest tank is enough to place a strong constraint on the geometry of the shower. Thanks to the position of this tank, the accuracy of the hybrid geometry reconstruction is 50 m on the core position and 0.6° on the shower axis [54].

In Fig. 3.10(b) an example of the reconstructed shower geometry is shown. This event is also detected simultaneously by CO and LL (a "stereo event" between CO and LL). The colored rays represent the light detected by the pixels. Figure 3.11 is produced by the measurement from CO of this event.

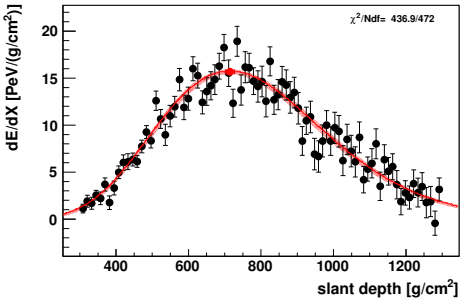
3.3.2 Shower profile reconstruction

The fluorescence light emitted by the de-excitation of the air molecules, excited by the energy loss of the shower development, is measured by the cameras of the telescopes. The calibration values are used to convert the signals to photon counts at the aperture. Once the geometry is known, the number of photons is converted in deposited energy on the detector as a function of slant depth. The measured fluorescence light at the shower is correlated to $\mathrm{d}E/\mathrm{d}X$ through the factor $Y_{\mathrm{air}}/(T_{\mathrm{mol}}T_{\mathrm{aer}})$. T_{mol}, T_{aer} and Y_{air} are estimated using the correct parameters of the atmosphere for the known shower geometry.

The light detected from the telescopes is not only fluorescence light. During the shower development in the atmosphere, several reactions occur (Section 2.4). Cherenkov emission and multiple scattered light must be disentangled to get the fluorescence light [75]. Due to the universality of electron distributions in UHECR [76], multiple scattered light is subtracted using



(a) Light measurement and its composition.



(b) $\mathrm{d}E/\mathrm{d}X$ measurement with Gaisser-Hillas fit (red line) and measured $X_{\max}$ (red dot).

Figure 3.12: Profile measurement as a function of slant depth, in light counts and $\mathrm{d}E/\mathrm{d}X$. The background is subtracted. This data are from the same shower as in Figs. 3.10(b) and 3.11 (AugerID 201217007204): the reconstructed calorimetric energy in CO is $(1.04 \pm 0.07) \cdot 10^{19}\,\mathrm{eV}$.

the parametrization from [77]. Cherenkov light is produced directly from the EAS or from scattered particles (both Mie and Rayleigh), and can be related to $\mathrm{d}E/\mathrm{d}X$ too. Therefore the energy deposit at the shower can be determined from the sum of Cherenkov and fluorescence light.

In Fig. 3.12 the conversion from the light (3.12(a)) to deposited energy (3.12(b)) is shown. In both figures, the average night sky background in the pixels during the shower measurement is subtracted.

Once the profile is measured, it is possible to reconstruct the energy and $X_{\max}$ by fitting it with a *Gaisser-Hillas* (G-H) function [78]:

$$f_{\mathrm{GH}}(X) = (\mathrm{d}E/\mathrm{d}X)_{\max} \cdot \left(\frac{X - X_0}{X_{\max} - X_0}\right)^{(X_{\max}-X_0)/\lambda} \mathrm{e}^{(X_{\max}-X)/\lambda}. \tag{3.3}$$

In Fig. 3.12(b), the red line is the G-H fit. The maximum of the profile (the red dot) is obtained directly from the parametrization of the Eq. (3.3). The total deposited energy, also called *Calorimetric energy* (E_{cal}), is the integral of f_{GH}:

$$E_{\mathrm{cal}} = \int_0^\infty f_{\mathrm{GH}}(X)\,\mathrm{d}X. \tag{3.4}$$

However part of the primary particle energy is missing, carried away by neutrinos and high energy muons. The total energy E of the shower is obtained by correcting for the "invisible energy" factor f_{invis} [79–81]:

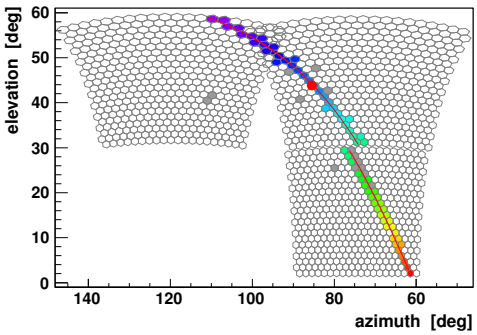
$$E = E_{\mathrm{cal}} \cdot f_{\mathrm{invis}}. \tag{3.5}$$

The statistical uncertainty of the energy is given by the propagation of the parameter errors from the fit, the geometry uncertainty and atmosphere statistical error. The systematics are given by several factors [67]. Most of the contribution comes from the calibrations (9.9 %). The effects from the fluorescence light yield (3.6 %), atmosphere ($3.4 \div 6.2$ %), point spread function [82] due the optical FD properties ($6.5 \div 5.6$ %), invisible energy ($3.0 \div 1.5$ %) and others ($3.0 \div 2.4$ %) sum up to a total of ≈ 14 %.

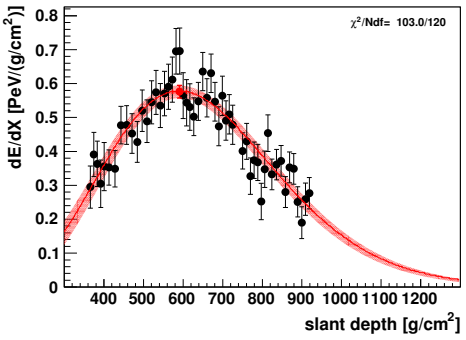
The measured uncertainties $\sigma_{X_{\max}}(X_{\max})$ of the maximum profile are due to two different factors. The first, statistical, is given by the fit parametrization error of $X_{\max}$ ($\sigma_{\text{stat}}(X_{\max})$) and the propagation of the geometry uncertainty. The second ($\sigma_{\text{atm}}(X_{\max})$) is from the statistical error of the VAOD. The systematic uncertainties will be discussed in Chapter 6.

3.3.3 Combined HEAT and CO: HeCo

HeCo are the telescopes of a virtual location that combine both HEAT and CO telescopes. They are a set of nine telescopes: $1 \div 6$ from CO and $7 \div 9$ from HEAT. In case of HEAT in downward position, overlapping occurs. This case leads to an increased number of $\mathrm{d}E/\mathrm{d}X$ data points for the profile determination. But the real power of HeCo is when HEAT is pointing upward, extending the field of view and being able to measure shallow showers. Moreover, also in case of CO events, the measurement of the longitudinal profile will be extended, for a better performance of the G-H fit. In the example in Fig. 3.13, a shower with low energy ($3.71 \cdot 10^{17}$ eV) is shown: its track on the camera (Fig. 3.13(a)) and its longitudinal profile with the Gaisser-Hillas fit (Fig. 3.13(b)). In this example is clear that without reconstruct the event as HeCo, the reconstruction of the depth of the shower profile maximum (red dot) would be of poor quality because not directly observed.



(a) Camera view



(b) Longitudinal profile. Gaisser-Hillas fit in red: the dot is $X_{\max}$, the area is the statistical uncertainty.

Figure 3.13: Example of a HeCo event with energy of $3.71 \cdot 10^{17}$ eV. The red dot is the depth of the shower profile maximum.

To reconstruct HeCo an accurate algorithm is used that takes into account the position, the geometry and the various differences between the two set of telescopes [83]. The different uptime and "T3 veto" are two features that must be particularly taken into account for the data analysis. They will be discussed in Section 4.2.

CHAPTER 4

Data Analysis

The data collected from the detectors of the Pierre Auger Observatory are analyzed to measure the distribution of the maximum of the shower development. To achieve a precise estimation of the composition of primary particle from $X_{\max}$, a high quality unbiased data sample must be assured. Therefore data are selected, event by event, taking care of:

1. *Data quality*: a good $X_{\max}$ and energy resolution is needed to resolve the true shower maximum distribution.

2. *Field of view bias*: due the limited field of view from the Fluorescence Detector (FD), the $X_{\max}$ distribution can be deformed.

Issue 1 is solved with a high quality selection of the events. Issue 2 requires a study of the field of view. Only air showers with a geometry for which a large enough FoV can be observed should be accepted. "Large enough" means that such showers have a proper sampling of $X_{\max}$. This sampling is necessary to have a correct interpretation of the mass of the primary cosmic ray particles. The study must be performed after the quality selection, which modifies the field of view bias.

After a choice of a proper data set (Section 4.1), issues 1 and 2 will be treated along the data selection and the field of view analysis (Sections 4.2 and 4.3). Checks on the latter will be treated to ensure the validity of the method (Section 4.4).

4.1 HEAT Data set

4.1.1 Hybrid performance

The Pierre Auger Observatory has been acquiring data since 2004 and, at the time of this work, the data are available to be reconstructed until the end of March 2014. However not all of them

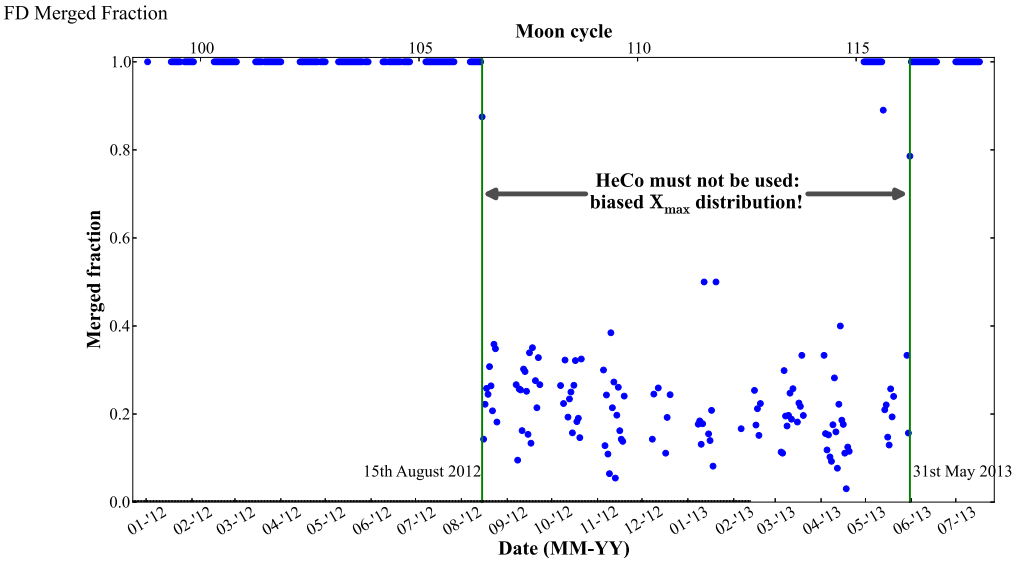


Figure 4.1: Fraction of the merged CO-HEAT FD Events that belong to the same shower. Every dot is the average of a single day.

can be used. Checks are done to ensure the quality of their acquisition. One of these checks points out a problem in the Hybrid Event formation with HEAT.

When a Hybrid Event is formed offline, FD Events and an SD Event identified as the same hybrid by the T3 are merged together. In recent data FD Events were found from HEAT that were not merged with the correct SD Events. They were merged with SD clearly not triggered by the same shower but from coincidental muons. In Fig. 4.1 the fraction of correctly merged Hybrid Events is shown. This ratio is estimated using showers that triggered both CO and HEAT. They are identified by their time stamp difference: less then 200 µs. Their total $N_{\mathrm{timeStamp}}$ is compared to the number of events correctly merged as a single Hybrid Event N_{merged}. In the figure, the average ratio $\langle N_{\mathrm{merged}}/N_{\mathrm{timeStamp}}\rangle$ per day is shown. To identify the same shower, the time stamp of the estimated shower core at the ground by the T3 is used.

Before the 15th August 2012, the efficiency is 1, as expected. The gaps are the full moon periods where FD data were not acquired. After this date, the efficiency is low and erratic. The correctly merged events are FD Events from HEAT without any SD Event associated (i.e. not tagged as hybrid by the triggers) and recovered as Hybrid Event offline. The recovering process is similar to this test: it compares the time stamps formed by the T3. The time stamp of an event identified as hybrid by the trigger is compared to the ones from FD Events not tagged as hybrid for some reason (i.e. by T3-veto). However, this process is possible only if the shower has already triggered some telescopes as hybrid, otherwise the event is lost.

The origin of this inefficiency was found to be a bug in a new software installation on CDAS. HEAT coordinates were set wrongly and CDAS sent a request from its T3 to SD stations triggered 20 ms earlier. The bug was fixed the 31st of May 2013 and since then FD Events from HEAT are correctly merged. Considering the data set from 2004 to the end of 2012[1], the fraction of HEAT events lost for this inefficiency are $\approx 15\,\%$.

The data from HeCo in this period cannot be used because they bias the sample. Hybrid

[1]The reason of considering only this subset is related to the selection cut **Measured Mie model available**, properly explained in Section 4.2.3.

Events formed from HEAT are recovered only if CO was also triggered as hybrid. Thus deeper showers are favored to be detected and HeCo[2] is partially blind. Even if for different reasons, this bias is similar to the one from the T3-veto[3], that will be explained in details in Section 4.2.4.

4.1.2 CO-HEAT cross calibration

The drum calibration campaign performed in January 2010 leads to a significant reduction of the uncertainties and systematics of the calibration (Section 3.1.3). Thus, also the reconstructed energy has smaller uncertainties (Section 3.3.2). A further check on the data was performed and is illustrated in the following.

This cross check uses events observed in two locations simultaneously ("stereo events"). The energy asymmetry between pairs of FD events is given by

$$\Delta E_{1-2} = \frac{E_1 - E_2}{\frac{E_1+E_2}{2}}, \tag{4.1}$$

where 1 and 2 are the first and second location taken into account in the stereo. The result, expressed in %, is expected to be compatible with 0 % inside the claimed total systematic uncertainty of the calibrations, which is 5.7 % [58, 59]. It is not clear which part of the uncertainty is uncorrelated. For sure, the calibration procedure of the drum is a common uncertainty to every telescope, which is $\approx 4.3\,\%$. Thus the uncorrelated systematic uncertainty will be considered to be $\leq 3.7\,\%$.

High quality[4] stereo events are analyzed. The mean values and uncertainty of Eq. (4.1) are shown in Fig. 4.2 pair by pair. The stereo events between LA-LM are not shown due the low statistics: only 5 events were found.

An incompatibility with $\Delta E_{1-2} = 0$ is evident when CO is involved. CO and HEAT in downward mode should be completely compatible. In the figure their difference is $5.3 \pm 0.6\,\%$. This value can be treated as the measurement of this incompatibility. If the CO energy is shifted down by this amount, the pairs LL-CO, CO-LA and CO-LM have values $-5 \pm 1.1\,\%$, $-0.8 \pm 1.2\,\%$ and $1.4 \pm 2.0\,\%$ respectively. These new values ΔE_{1-2} are compatible with 0 inside $1.1\,\sigma$ of the claimed systematics.

Checks on subsets of data in time and with different reconstruction algorithms were done, but no correlations with this incompatibility were found. Thus, the reason for this difference is not yet understood and still under study. There is no physical reason to believe that CO is measuring higher energies. The values in the figure are relative differences. Therefore, the same compatibility is obtained if all the other telescopes have the energy shifted up by the same amount.

CO and HEAT are combined to form HeCo. This energy difference influences the $X_{\max}$ measurement. The G-H fit will parametrize the maximum of the shower differently if parts of $\mathrm{d}E/\mathrm{d}X$ measurement are shifted up or down due to a wrong calibration. This happens whenever a shower track crosses a HEAT and a CO telescope.

In Fig. 4.3 the $X_{\max}$ difference between HeCo events reconstructed with the observed 5.3 % energy difference shown in Fig. 4.2 (labelled as "normal") and events with the calibration corrected by this amount is shown. The correction can be applied to CO (blue markers, labelled

[2] Reminder: HeCo is a virtual FD obtained merging HEAT and CO telescopes and located in the CO site.

[3] To be more precise, the situation is equivalent to the third condition in Table 4.2

[4] The event selection done is not the same used for this work and discussed in Section 4.2. It is softer to keep the number of the showers high. The results are asymmetries between stereo events with the same field of view. Thus the fiducial analysis (Section 4.3.2) is not needed because the FoV bias is the same for each telescope. For the CO-HEAT pairs the atmospheric factors are the same, thus atmospheric cuts are not needed as well. For reproducibility, the cut list is reported in appendix H.2.

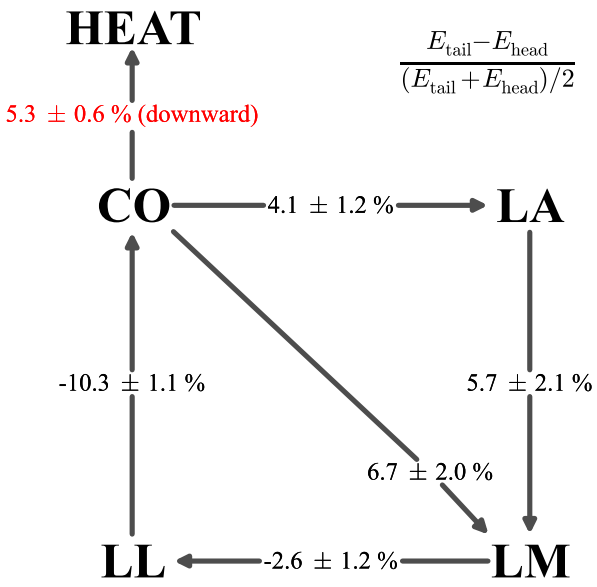


Figure 4.2: Summary scheme of the energy asymmetries between stereo events. Mean and uncertainty of Eq. (4.1) are reported, stereo by stereo. Using this equation as reference, the tails and the heads of the arrows are E_1 and E_2 respectively: $\Delta E_{1-2} = \Delta E_{\text{tail}-\text{head}}$. In red, the measured energy difference between CO and HEAT.

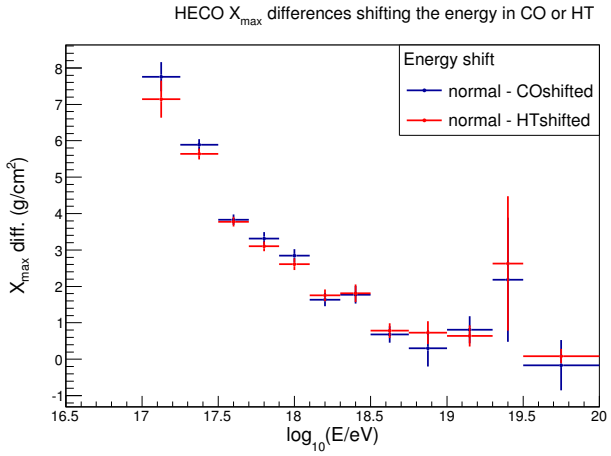


Figure 4.3: Difference between the $X_{\max}$ measured from HeCo without (labelled as "normal") and with energy shift. In blue, CO energy is shifted down by $5.3\,\%$, in red, HEAT is shifted up.

as "COshifted") or HEAT (red markers, labelled as "HTshifted"). Both cases are in agreement because the source of the gap is the relative difference between the telescopes.

As shown in the figure, $X_{\max}$ is overestimated by up to $\sim 8\,\mathrm{g/cm^2}$ at low energies because the track crosses CO and HEAT with higher probability. At high energies HeCo is dominated by CO telescopes and the signal from HEAT contributes only to the tail of the shower profile detection. In this case the difference between $X_{\max}$ is negligible.

As a conclusion, this energy correction must be applied and a decision must be taken: which telescopes must be corrected? Lowering CO or raising the others? HEAT has a newer and more advanced DAQ system with a faster sampling. Thus, in this work, it is assumed to have the most correct energy measurement. Moreover, lowering a single location (CO) will have a lesser impact than raising all the others. Hence, FD Events from CO are reconstructed with a correction to the calibration of 5.3 % that causes a downward shift of the CO energy by the same amount. All data analyzed in this work already include this correction.

As consequence of this choice, the systematic uncertainty of the reconstructed energy (Section 3.3.2) must be increased by the asymmetric error $^{+5.3}_{-0.0}$ %. Moreover, this value is obtained through a measurement with an uncertainty. Therefore, the measurement of the $X_{\max}$ moments must consider the propagation of this uncertainty (0.6 %) as systematic error (Chapter 6).

4.2 Data selection

The data selection is crucial to unbias the data sample and have high quality showers. This selection is done applying the cuts summarized[5] in Table 4.1.

4.2.1 Data acquisition

This group of selection criteria is meant to reject events recorded during periods with unstable detector conditions.

A first example of this selection is the **Bad FD periods** cut. It rejects the events acquired during periods when some telescopes were in non-optimal data acquisition conditions: no calibration available, glitches from the GPS, unstable baseline, wrong alignment, etc...

In this selection, also issues related to the photomultipliers are taken into account. Pixels can be temporarily considered "bad" for malfunctioning, bad calibration or other reasons. The **No bad pixels in the track** cut rejects those events with triggered bad pixels along the track. The **Skip saturated PMT** cut, instead, takes into account when the photomultipliers are saturated.

One more rejection related to the PMT is the **Shutter status** cut. During the closure of a shutter, events can be still acquired. The mean signal between the overall pixels in this case will be lower than the normal DAQ period with a smaller standard deviation (RMS). To be accepted, the event must have an RMS higher than a threshold ($17\,\gamma/100\,\mathrm{ns}$ at aperture). This official value is optimized for the standard FD. HEAT has half of the time bin size (Section 3.1.2), so also the average of the signal will be halved. The cut considers this in a proper way, rescaling the measurement to be comparable to the standard telescopes. Moreover, the HeCo situation is more complicated, because the mean signal is the average of the triggered CO and/or HEAT telescopes. For a proper selection, it will be considered the minimum average of the corresponding CO or HEAT event. In Appendix H.1 there are two different cuts: one for HeCo (`minBackgroundRMSMergedEyes`) and one for the other telescopes (`minBackgroundRMSSimpleEyes`).

[5] For reproducibility, the cut list to use in Offline is reported in Appendix H.1.

Table 4.1: Summary of the cuts. The efficiencies are relative to the previous cut. Details are explained in Section 4.2.

Cut	Condition	Efficiency
Data acquisition		
Bad FD periods	-	82.3 %
Skip saturated PMT	-	99.3 %
No bad pixels in the track	-	98.7 %
Shutter status	-	94.4 %
Good 10 MHz Correction	-	99.3 %
Hybrid		
Has hybrid geometry	-	36.4 %
ToT station	-	85.3 %
Maximum zenith angle	85°	85.1 %
Energy reconstructed	-	92.2 %
Maximum distance Core-Tank	1500 m	99.8 %
Ambiguous Hybrid Rejection	-	$\sim$ 100.0 %
Atmospheric		
Measured Mie model available	-	73.2 %
Maximum VAOD	0.1	94.8 %
Cut on the clouds	-	77.3 %
HeCo		
Minimum brass probability	90 %	84.9 %
Maximum brass probability difference p-Fe	5 %	94.2 %
Event not in T3veto	-	97.1 %
Reconstruction		
Minimum track length	200 g/cm^2	64.4 %
Correct X_{max} error estimation	-	98.9 %
Maximum gap in the slant depth	30 %	99.3 %
χ^2-cut	6	99.3 %

The last cut in this selection group is related to the GPS measurement. The timing is based on a precise "1 PPS" (pulse per second), i.e. a reference signal at each full second. Moreover a fast oscillator measures fraction of microseconds. This oscillator has a frequency of 10 MHz, thus it defines units of 100 ns. However this oscillator is of normal industry quality and such 100 ns-unit is not a precise subunit of the second. After accumulating a certain amount of counts, it is possible to have a slightly larger or shorter time than 1 s. This imprecision can be stored as data information and corrected [84]. The **Good 10 MHz Correction** cut rejects events without such a correction, therefore with an imprecise GPS timing at ns precision.

4.2.2 Hybrid

This group of selection criteria is meant to reject non-hybrid events or events with a failed or bad hybrid reconstruction.

First of all, not all the events are hybrids. Brass hybrids[6] can be formed from Surface Detector Array (SD) stations with only Threshold trigger (TH) passed (Section 3.2.2). Thus, FD events with SD tanks triggered by coincidental atmospheric muons are also considered Hybrid Events from the trigger point of view. During the reconstruction these SD Events are recognized as background signal and not reconstructed. The **Has hybrid geometry** cut rejects events without reconstructed stations, i.e. the accepted are real hybrid events. Moreover, the **ToT station** cut guarantees that the tank used as constraint to the geometry reconstruction has the Time over Threshold trigger (ToT). This trigger grants to have not used a tank triggered by coincidental atmospheric muons to reconstruct the geometry of the shower. Hence events selected by these cuts have a "real hybrid geometry": they are the 31 % of the data that have already passed the data acquisition selection.

However not all of them are reconstructable air showers useful for this work. Firstly, events too horizontal or coming from the ground[7] are rejected by **Maximum zenith angle** cut. Secondly, the **Energy reconstructed** cut rejects those events with a failed reconstruction. Reconstructed events with a real hybrid geometry are 24.4 % of the data that have already passed the data acquisition selection.

The **Maximum distance Core-Tank** cut is used as geometry selection. Those events that have used a SD station too far from the shower core (> 1.5 km) as reference to reconstruct the shower geometry (Section 3.3) are rejected by this cut.

Checking the events, few of them appear to be "ambiguous". This ambiguity is given by a SD station that is triggered as the one with the highest signal[8], but it does not belong to the core of the shower. This can happen for signals due to coincidental muons or coincidental showers with the same core. 183 events are rejected by the **Ambiguous Hybrid Rejection** cut through a list of identified showers.

4.2.3 Atmospheric

This group of selection criteria is meant to reject events with unfavorable atmospheric conditions. Two effects can bias the measurement: the aerosol scattering and the absorption and reflection due to clouds.

Firstly, events are only accepted if the aerosol content in the atmosphere had been measured by the CLF/XLF (i.e. the Measured Mie model was used during the reconstruction, see Section 3.1.4). The **Measured Mie model available** cut rejects events without such information

[6] Hybrid Event formed by less than 3 SD stations, with a T3 from a FD Event. See Section 3.3.
[7] For example from Central Laser Facility (CLF) and Extreme Laser Facility (XLF) activities
[8] It is expected to be closer to the shower core.

in the DB. The reconstructable data are acquired since 2004 until March 2014, but the Mie database has information only until the end of 2012. Events since 2013 are rejected.

The **Maximum VAOD** cut rejects events with a too large VAOD value at the reference height of 3 km from the ground. A value too large means too low T_{aer} because the atmosphere is too dense (Eq. (3.1)). In this case the reconstruction can lead to larger systematic uncertainties on the energy. In Fig. 4.4 the distribution of the VAOD values is shown before this selection. A cut at VAOD < 0.1 rejects events with a too high VAOD and keeps the number of the showers high.

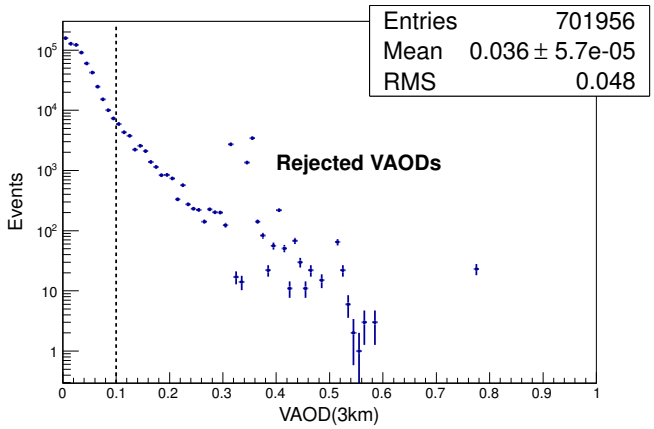


Figure 4.4: Distribution of VAOD values at the reference height of 3 km immediately before the **Maximum VAOD** is applied. The dashed line is the level of the **Maximum VAOD** cut. Logarithmic scale on y axis.

Clouds, if present, absorb and scatter the fluorescence light from an air shower. Thus the reconstructed energy will be unpredictably smaller than in reality. Through the cloud camera, the information from the LIDAR stations and the Geostationary Operational Environmental Satellites (GOES) database (Section 3.1.4), the **Cut on the clouds** cut rejects those events with a too large cloud coverage.

4.2.4 HeCo

This group of selection criteria is meant to reject events with bad conditions regarding the HeCo reconstruction.

At low energies, the probability to get at least one ToT trigger in the SD for a given FD event is not 1 [85]. It also differs from regions with different SD spacing (regular or Infill). However it is possible to calculate the probability for a single event to trigger at least one surface detector station. The algorithm uses the real set of active stations at the time of the event. This probability is estimated for both proton and iron primaries as a function of zenith angle, energy, and core position of the reconstructed shower. Hence a fiducial area close to the SD can be determined using the **Minimum brass probability** cut. It requires that both these two probabilities are larger than 90 % respectively ($P_{\text{proton}} > 0.9$ and $P_{\text{iron}} > 0.9$). Yet these two values should not be very different ($< 5\,\%$). If the difference is too large, one nucleus has a larger probability of being detected compared to the other. This privilege favors the detection of one of the masses, creating a bias. **Maximum brass probability difference p-Fe** cut requires that $|P_{\text{proton}} - P_{\text{iron}}| < 0.05$ [83].

As already introduced in Section 3.1.2, the T3 is the trigger used to combine the SD and FD events together, producing the Hybrid Event. The one formed from FD has a veto to deal with the limited bandwidth of the SD communication system and the possibility of a too high T3 rate from any FD site. If T3 is in veto status the request to the SD stations is not sent. Thus the brass hybrid cannot be created. HEAT and CO do not have a synchronized T3. Therefore it can happen that some of the HeCo brass hybrids are lost.

Giving an example, if HEAT has T3 in veto and CO has not, a shower that triggers only HEAT does not send the request to SD. Therefore, the potential brass event is lost. If the shower triggers also CO, the hybrid trigger is formed. The request is sent to SD and the whole of the event is recovered.

Because this veto, HeCo can be partially blind in one of its telescope sets. Hence deep and shallow events with the same energy are treated differently. This creates a bias in the data sample for the purpose of this work.

For the X_{max} distribution sake, the consequence is a measurement bias sometimes towards shallower showers, sometimes towards deeper and only sometimes the sample is unbiased. In Table 4.2 a summary of the T3 veto cases is shown, where also the blindness status of HeCo is exemplified.

Table 4.2: Summary of the T3-veto cases.

Veto Condition		HeCo Hybrid	HeCo is blind to
Co	HEAT		
Y	Y	Brass hybrids lost	No HeCo hybrids
Y	N	CO can be recovered if HEAT sent T3	Deep showers
N	Y	HEAT can be recovered if CO sent T3	Shallow shower
N	N	T3 sent in any case	Unbiased

The **Event not in T3veto** cut deals with the different T3 vetoes between CO and HEAT. The cut rejects HeCo events when at least one of the two locations has T3 in veto status. In other words, it accepts only events that respect the fourth condition of Table 4.2. After the selection done up until now 26.5 % of the HeCo events are potentially biased due to T3-veto.

4.2.5 Reconstruction

This group of selection criteria is meant to reject events reconstructed with poor quality.

The first cut, **Minimum track length**, is meant to reject events with a too short measured length of the track along the shower axis. In such cases it is not granted a G-H fit with an high quality parametrization. $200\,\mathrm{g/cm^2}$ is chosen as minimum value to have a measured longitudinal profile long enough without losing too many events at low energy. In Fig. 4.5 the scatter plot of the track length as a function of $\log_{10}(E/\mathrm{eV})$ is shown.

Sometimes the error propagation during the X_{max} reconstruction fails. When this happens, Offline sets the value of the X_{max} error as X_{max} itself. The **Correct X_{max} error estimation** cut rejects these events.

One more inefficiency in the quality of the sample can be given by the gaps between cameras. When the reconstruction combines the measurement of the profile along the cameras, the spacing between them is a gap in the track. The cut **Maximum gap in the slant depth** requires that this spacing is not larger than 30 % of the track.

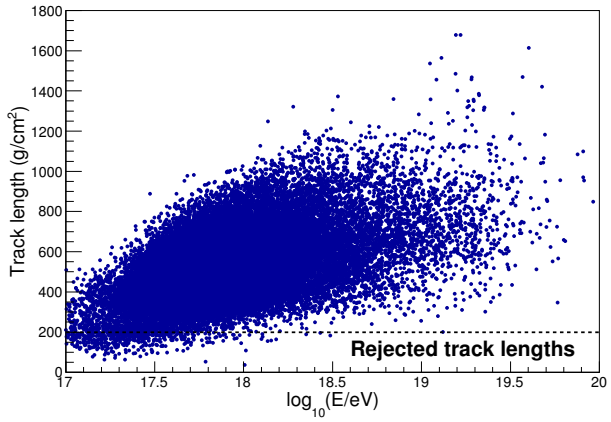
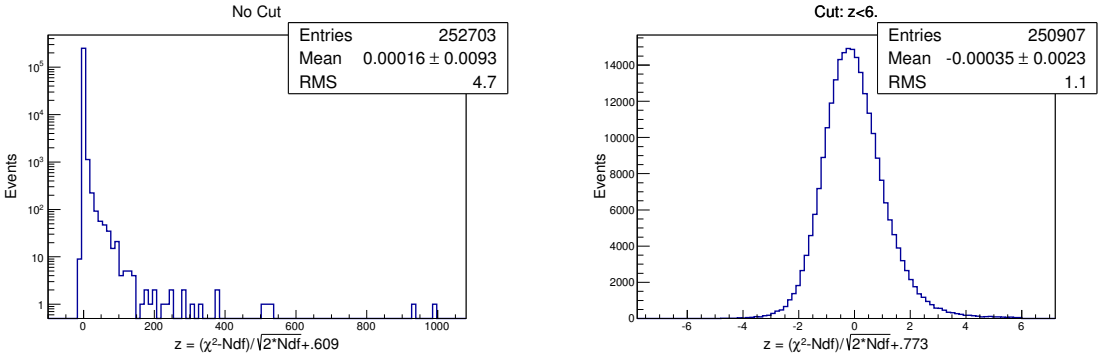


Figure 4.5: Scatter plot of the track length as a function of $\log_{10}(E/\text{eV})$. The dashed line is the level of the **Minimum track length** cut.



(a) Distribution without cut. The y axis is in logarithmic scale.

(b) Distribution with the cut $z < 6$.

Figure 4.6: Distribution of the normalized χ^2 (Eq. (4.2)), shifted by the proper $\overline{x}$ to have the variable z with a mean value $= 0$.

The last cut, **χ^2-cut**, deals with the quality of the fit, using χ^2. The idea to extract high quality showers is to have outliers rejected. They can be identified using the normalized value[9]

$$x = \frac{\chi^2 - \text{Ndf}}{\sqrt{2 \cdot \text{Ndf}}}, \tag{4.2}$$

where Ndf is the number of degree of freedom. According to the central limit theorem, x is shaped like a normal distribution N. It is just approximately distributed as N, because the Gaisser-Hillas fit is done using the maximum likelihood method. Thus x will be not centered at 0 and will not be symmetric. In Table 4.1, the maximum number of standard deviations (σ or RMS) is reported. The rejected events are those with x far away from the mean $\overline{x}$ more than this value. In Appendix H.1, a second value on the "profileChi2Sigma" cut is reported. This is the

[9]The mean value of a χ^2 distribution is $\mu = \text{Ndf}$ and its variance $\sigma^2 = 2 \cdot \text{Ndf}$

shift $\overline{x}$ needed to center the mean value to the 0: $z = x - \overline{x}$. In Fig. 4.6 z distributions are shown. Before the cut on the maximum number of σ (Fig. 4.6(a)): the RMS of the distribution is 4.7. After a cut of 6 standard deviations (Fig. 4.6(b)), it is 1.1, similar to the normal distribution.

4.3 Field of view analysis

As already anticipated, the field of view is limited and can bias the data sample used to measure $X_{\max}$. To unbias it, one more selection of the data is necessary, extending the Table 4.1 with the Table 4.3.

Table 4.3: Field of view cuts. The efficiencies are relative to the previous cut. Details are explained in Section 4.3

Cut	Condition	Efficiency
Field of view		
$\boldsymbol{X_{\max}}$ **observed in the expected FoV**	$\xi \leq 40\,\mathrm{g/cm^2}$; mva $\geq 20°$	58.2 %
Fiducial FoV	-	26.2 %

In this table, two more cuts are reported. The first one requires that the observed $X_{\max}$ is inside the so called "Expected field of view" (Section 4.3.1). After this cut, the quality bias is removed. The second deals with the FoV bias, estimating a "Fiducial field of view" (Section 4.3.2). After the entire data selection, only the 0.74 % of the hybrid events are selected for the analysis.

4.3.1 Expected field of view

Beside the geometrical field of view, a different FoV can be defined, termed *expected field of view* (expFoV). Here the data are expected to be detected with high quality measurements. In Fig. 4.7, an example of this region is shown for energies between $10^{18.5}$ eV and $10^{18.7}$ eV [86]. The axes are the distance from the telescopes to the showers (ground distance, height). Red lines represent the geometrical FoV. The color scale indicates this efficiency region. It is the selection efficiency of the $\boldsymbol{X_{\max}}$ **observed in the expected FoV** cut. Red dots are the $X_{\max}$ positions of the events until 2011.

To calculate the expFoV, virtual showers are considered as follow. Given a measured shower, the virtual ones have the same energy but a different $X_{\max}$. The maximum of the profile $\widetilde{X}_{\max}$ is moved along the slant depth inside the geometrical FoV[10]. The precise geometrical FoV depends on which telescopes are capable of data acquisition. As mentioned in Section 3.1.2, the fraction of the time in which a single telescope is capable of data acquisition during ten minutes is called "uptime". For the standard telescopes this ratio is not really important. If an event does not trigger a single camera, the camera is considered not in uptime. However, in HeCo case, some telescopes are looking upward in the sky (HEAT) and some in downward (CO). Thus the uptime can be fundamental to determine the elevation of the geometrical FoV. If, for example, a deep shower does not trigger one HEAT but it was capable of data acquisition, the geometrical field of view can be considered shorter then what really is.

Each new artificial shower has predicted signals and number of pixels with First Level Trigger (FLT). These estimations are possible thanks to the knowledge of the light yield and the light

[10] The geometrical FoV is the combined FoV of the telescopes viewing the air shower track (Section 3.1.1).

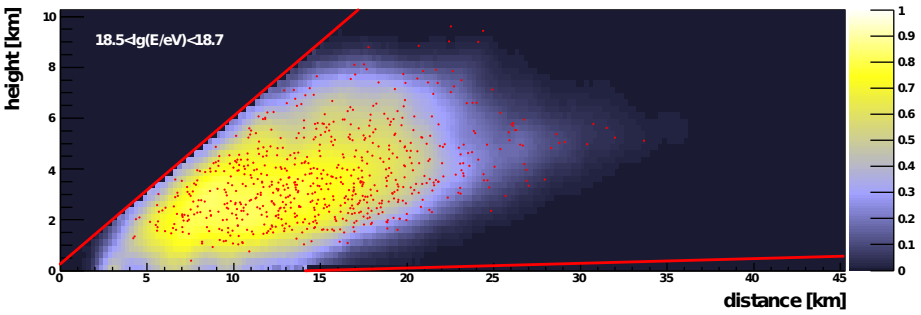
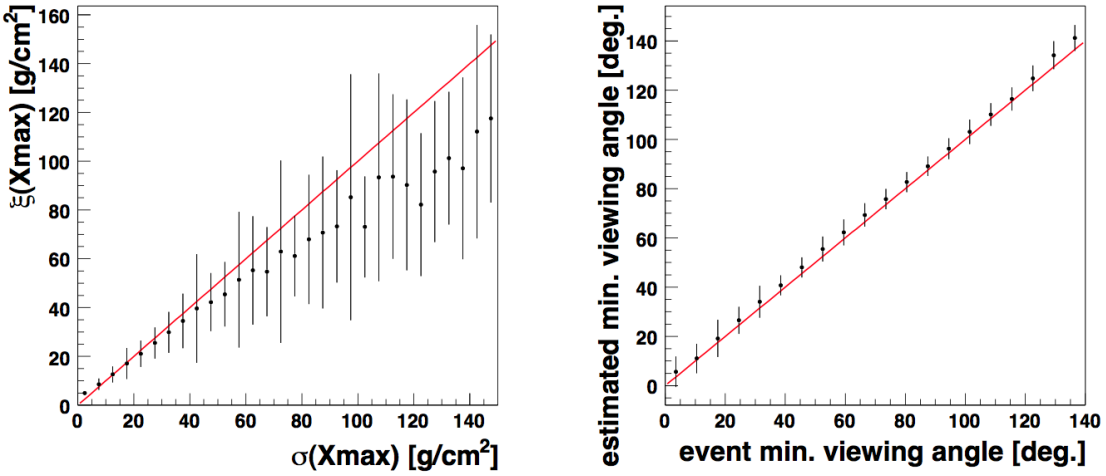


Figure 4.7: Example of expected FoV region for primary particle energies between $10^{18.5}$ eV and $10^{18.7}$ eV. The color scale is the density of high quality shower selected as a function of the shower distance from the telescope. The red lines are the geometrical FoV boundaries. The red dots are the position of $X_{\max}$ of the showers. From [86]



(a) $\xi(X_{\max})$ vs. $\sigma_{X_{\max}}(X_{\max})$.

(b) Estimated vs. real event mva($X_{\max}$).

Figure 4.8: Performance of ξ and mva determination [87]. The error bars are the standard deviation around the mean, the red lines show the expectation in case of a perfect correlation.

transmission through the atmosphere. For each of artificial showers, two variables are calculated: ξ and *minimum Viewing Angle* (mva).

ξ is the estimated resolution of $\widetilde{X}_{\max}$. Using error propagation, an *expected* uncertainty ξ^i of $\widetilde{X}^i_{\max}$ is given for the i-th artificial shower.

The angle between each pixel FoV direction and the shower axis (referred to as α in Fig. 3.5) is termed *Viewing Angle* (VA). The mva is the minimum VA below which there is a too large ($\gtrsim$ 20%) contamination of Cherenkov light. This light also carries information about the longitudinal development of a shower. Events with a large Cherenkov fraction are currently not used in the data analysis. These events are too susceptible to systematic errors in the determination of the geometry [87]. So it is necessary to set a minimum angle above which the contribution of the

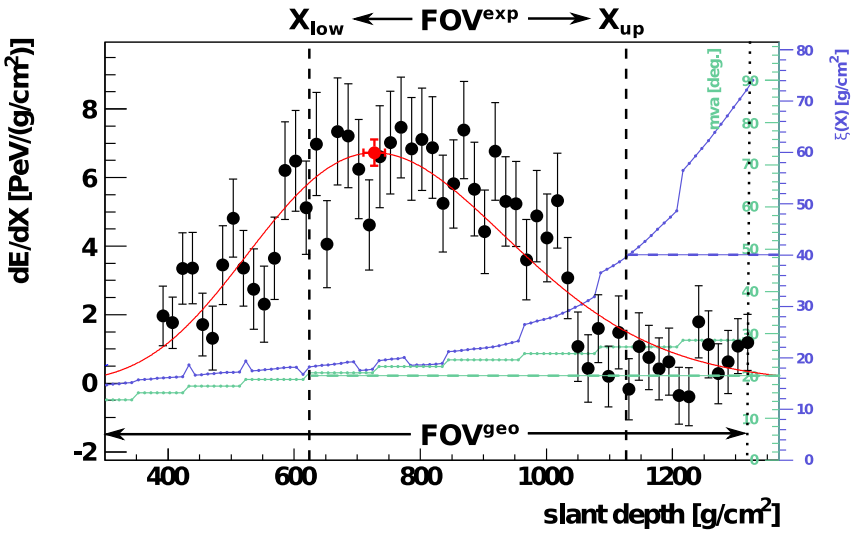


Figure 4.9: Example of geometrical ($\mathrm{FOV}^{\mathrm{geo}}$) and Expected FoV ($\mathrm{FOV}^{\mathrm{exp}}$) delimited by ξ (blue markers) and estimated mva (green markers) limits. The black markers are the deposited energy $\mathrm{d}E/\mathrm{d}X$ in the telescope. The red line is the Gaisser-Hillas profile and the red dot $X_{\max}$.

Cherenkov light is not relevant anymore. This angle is estimated by the reconstruction of $X_{\max}$ for each shower. Therefore, for each virtual $\widetilde{X}^i_{\max}$ the expected mva^i value is also extracted.

A check on the performance of these estimations can be done. The expected values for the measured $X_{\max}$ ($\xi(\widetilde{X}^i_{\max} = X_{\max})$ and $\mathrm{mva}(\widetilde{X}^i_{\max} = X_{\max})$) are compared to the estimated by the reconstruction. In Fig. 4.8 is shown a good correlation between the quantities. For large uncertainties, Fig. 4.8(a), $\xi(X_{\max})$ is systematically lower. However, the official selection sets $\xi \leq 40\,\mathrm{g/cm^2}$ and $\mathrm{mva} \geq 20°$, regions where the performances are good.

In Fig. 4.9 an example is shown of how the **$X_{\max}$ observed in the expected FoV** cut works. The black markers are the deposited energy in the telescope, the red line is the Gaisser-Hillas profile and the red dot the reconstructed $X_{\max}$. Each bluish dot has coordinates $(\widetilde{X}^i_{\max}, \xi^i)$, one for each artificial shower; each green dot, $(\widetilde{X}^i_{\max}, \mathrm{mva}^i)$. Where both the conditions $\xi \leq 40\,\mathrm{g/cm^2}$ and $\mathrm{mva} \geq 20°$ in the slant depth X are met, two boundaries are defined. These are named X_{up}, the upper limit near to the ground, and X_{low}, the lower high in the atmosphere. If $X_{\max}$ is observed and inside these boundaries, the event is accepted. It belongs to a region where the showers are expected to be reconstructed with a good resolution and a low Cherenkov light contamination. In other words, an high quality shower.

In this example X_{low} ($= 625\,\mathrm{g/cm^2}$) has $\xi = 18\,\mathrm{g/cm^2}$ and $\mathrm{mva} = 20°$, X_{up} ($= 1125\,\mathrm{g/cm^2}$) has $\xi = 40\,\mathrm{g/cm^2}$ and $\mathrm{mva} = 28°$. The event is accepted because $X_{\mathrm{low}} < X_{\max} < X_{\mathrm{up}}$.

After the data selection (Section 4.2) and this cut, high quality events are chosen.

4.3.2 Fiducial field of view

Different nuclei have different $X_{\max}$ and different distribution spread. Light nuclei penetrate, on average, deeper in the atmosphere with larger distribution width compare to heavier ones. Because of this, a limited field of view treats light and heavy cosmic rays differently. Considering, for example, a deep and a shallow penetrating primary particle with the same energy and the same vertical geometry. The deeper one can be easier detected from the FoV of the telescopes,

compared to the heavier. Thus the measured $X_{\max}$ distribution will be influenced: this is called the "field of view bias". Moreover the same two showers have the longitudinal profiles detected with different track length. This implies that also the data selection contributes to this bias. This is the reason why the fiducial field of view study is done after the quality selection. In Fig. 4.10 the illustration the deep/shallow example explained before is shown. For simplicity the same longitudinal profile is assumed and the difference between the two is just the shift in $X_{\max}$. The maximum (the dot) of the blue profile is not detected, thus a shallower $X_{\max}$ is lost.

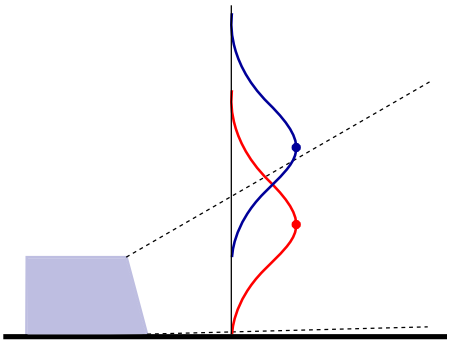


Figure 4.10: Simple illustration of two showers, identical but for the $X_{\max}$ position: deeply penetrating nucleus in red and shallow in blue. For simplicity the same longitudinal profile is assumed. Dots represent the $X_{\max}$

The solution is to study the field of view itself and define a fiducial volume. This region is where deep and shallow primaries can be detected with equal probability. In other words, where none of the possible masses are favored. The showers can have a lot of different possible geometries and physical characteristics. Thus, to save the statistics, this study must be optimized.

The idea is to study how the mean of the $X_{\max}$ distribution changes when the boundaries of the FoV are moved. A "large enough" region must be set to have an unbiased[11] sample keeping the number of the showers high. For this process, it is necessary to ensure the high quality of the events.

In first approximation, the $X_{\max}$ distribution is described by the convolution of two distributions: $G \otimes E$, a Gaussian and an exponential. The first interaction point is defined by the exponential distribution $E(X)$. The Gaussian $G(X)$ describes the shower development. The complete mathematical expression is given in Eq. (E.1).

In an ideal world, where the FoV is infinite, the mean of the distribution is the real unbiased mean $X_{\max}$:

$$\langle X^{\infty}_{\max} \rangle = \frac{\int_0^{\infty} x\, G \otimes E\,(x)\,\mathrm{d}x}{\int_0^{\infty} G \otimes E\,(x)\,\mathrm{d}x}. \tag{4.3}$$

Shrinking the limits of integration to (x_1, x_2) defines the truncated mean $\mu_{\text{trunk}}(x_1, x_2)$ (see

[11] Unless specified, in this section the bias considered is the filed of view bias.

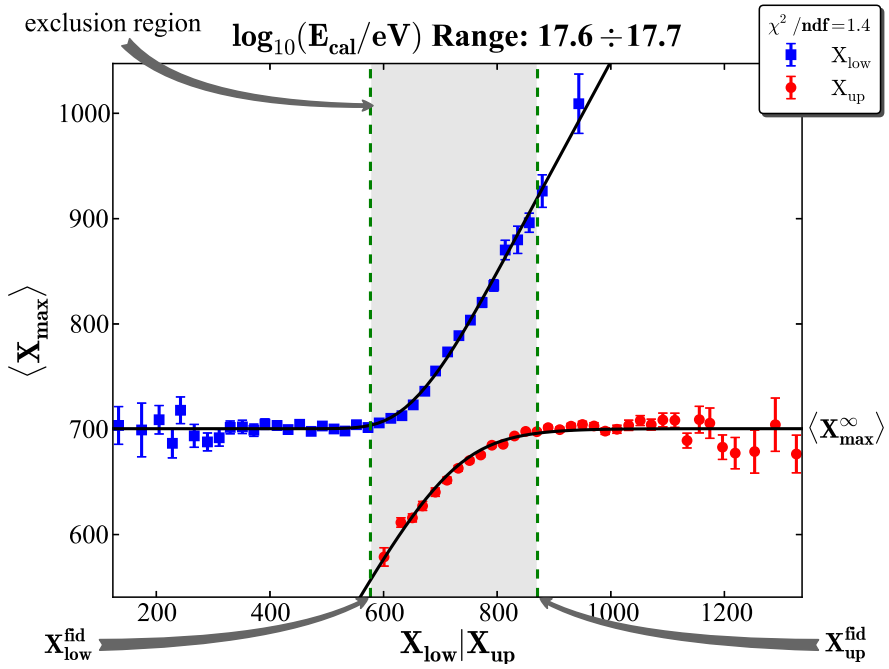


Figure 4.11: Example of the calculation of the fiducial FoV boundaries in the energy bin $17.6 \leq \log_{10}(E_{\mathrm{cal}}/\mathrm{eV}) < 17.7$. $\langle X_{\mathrm{max}} \rangle$ as a function of X_{up} and X_{low} are the two data sets represented by red square and blue dots respectively. Black lines are the simultaneous fit of both the data set with μ_{trunk} (Eq. (4.4)).

Appendix E for mathematical details):

$$\langle X_{\mathrm{max}}^{\mathrm{trunc}} \rangle = \mu_{\mathrm{trunk}}(x_1, x_2) = \frac{\int_{x_1}^{x_2} x\, G \otimes E\,(x)\,\mathrm{d}x}{\int_{x_1}^{x_2} G \otimes E\,(x)\,\mathrm{d}x}. \tag{4.4}$$

As already specified, the expected FoV range is the one that must be large enough to have unbiased distribution. Thus, it is natural to consider x_1 a variable along the X_{low} values and x_2 along X_{up}. Moreover if the sample is unbiased in a certain integral region, $\langle X_{\mathrm{max}}^{\mathrm{trunc}} \rangle \approx \langle X_{\mathrm{max}}^{\infty} \rangle$ must be true. As the bias increases, $\langle X_{\mathrm{max}}^{\mathrm{trunc}} \rangle$ deviates from the ideal mean. A systematic uncertainty is defined, termed Δ, as the maximum acceptable deviation from that mean. The standard value is set as $\Delta = 5\,\mathrm{g/cm^2}$ [87].

With the condition

$$\left| \langle X_{\mathrm{max}}^{\mathrm{trunc}} \rangle - \langle X_{\mathrm{max}}^{\infty} \rangle \right| \leq \Delta, \tag{4.5}$$

the two boundaries $X_{\mathrm{low}}^{\mathrm{fid}}$ and $X_{\mathrm{up}}^{\mathrm{fid}}$ of the fiducial volume can be determined. This boundaries are the limits of X_{low} and X_{up} to define if a shower has an expFoV large enough or not. In the end the **Fiducial FoV** cut accepts events only with an expected FoV range beyond these boundaries. In other words, every event with $X_{\mathrm{low}} < X_{\mathrm{low}}^{\mathrm{fid}}$ and $X_{\mathrm{up}} > X_{\mathrm{up}}^{\mathrm{fid}}$ is accepted.

Suppose to already know $X_{\mathrm{up}}^{\mathrm{fid}}$ and consider only the events with $X_{\mathrm{up}} > X_{\mathrm{up}}^{\mathrm{fid}}$. In Fig. 4.11 the

Table 4.4: Parameters of Eq. (4.6) to describe the $X^{\mathrm{fid}}_{\mathrm{low}}$ and $X^{\mathrm{fid}}_{\mathrm{up}}$ data in Fig. 4.12.

$\mathbf{FOV(\log E_{cal})}$	$\log E_0$	FOV_0	K	ω
$\boldsymbol{X^{\mathrm{fid}}_{\mathrm{up}}(\log E_{\mathrm{cal}})}$	17.63	871.47	45.94	0.48
$\boldsymbol{X^{\mathrm{fid}}_{\mathrm{low}}(\log E_{\mathrm{cal}})}$	16.14	673.08	-337.00	1.27

blue squares are the X_{low} of such events, sampled in bins[12]. In both the plots the $\langle X^{\mathrm{trunc}}_{\mathrm{max}}\rangle$ flat region and its deviation from $\langle X^{\infty}_{\mathrm{max}}\rangle$ are observable. While X_{low} increases, the expected FoV is moving from up the atmosphere toward the ground. This favors deeper showers when the FoV bias starts. Thus $\langle X^{\mathrm{trunc}}_{\mathrm{max}}\rangle$ increases.

The same procedure must be done for the X_{up} values. Suppose an $X^{\mathrm{fid}}_{\mathrm{low}}$, consider only events with $X_{\mathrm{low}} < X^{\mathrm{fid}}_{\mathrm{low}}$ and sample in bins their X_{up} values (red dots). The flat region and the deviation from $\langle X^{\infty}_{\mathrm{max}}\rangle$ are observable, also in this case. Meanwhile X_{up} decreases, the expFoV is moving from the ground upwards the atmosphere. This favors shallower showers when the FoV bias starts because the deepest ones are lost. Thus $\langle X^{\mathrm{trunc}}_{\mathrm{max}}\rangle$ is decreasing as well.

The condition $X_{\mathrm{up}} > X^{\mathrm{fid}}_{\mathrm{up}}$ implies that $\langle X^{\mathrm{trunc}}_{\mathrm{max}}\rangle$ is described by $\mu_{\mathrm{trunk}}(X_{\mathrm{low}}, \infty)$, meanwhile $X_{\mathrm{low}} < X^{\mathrm{fid}}_{\mathrm{low}}$ implies $\langle X^{\mathrm{trunc}}_{\mathrm{max}}\rangle = \mu_{\mathrm{trunk}}(0, X_{\mathrm{up}})$. For the same X_{max} distribution, the function μ_{trunk} must be the same regardless of the limits x_1 and x_2. Thus, for an initial guess $(X^{\mathrm{fid}}_{\mathrm{low}})^0$ and $(X^{\mathrm{fid}}_{\mathrm{up}})^0$, the two data sets previously described must be fitted simultaneously. From the fit, $\langle X^{\infty}_{\mathrm{max}}\rangle$ is estimated, and through Eq. (4.5) the new boundaries $(X^{\mathrm{fid}}_{\mathrm{low}})'$ and $(X^{\mathrm{fid}}_{\mathrm{up}})'$ are determined.

Because the initial $(X^{\mathrm{fid}}_{\mathrm{low}})^0$ and $(X^{\mathrm{fid}}_{\mathrm{up}})^0$ are guessed, the expected precision of $(X^{\mathrm{fid}}_{\mathrm{low}})'$ and $(X^{\mathrm{fid}}_{\mathrm{up}})'$ is not accurate. However it is now possible to iterate the process using the newly found fiducial boundaries as new initial guess. After few iterations, the fiducial FoV is stable[13] and the correct $X^{\mathrm{fid}}_{\mathrm{low}}$ and $X^{\mathrm{fid}}_{\mathrm{up}}$ are defined. In Fig. 4.11 the simultaneous fit of the two sets of values is represented by the black lines and the fiducial boundaries are shown as green dashed lines. The exclusion region to which the expected FoV boundaries are too small to allow for an unbiased sample is represented by the gray area.

The fidFoV limits are estimated sampling the entire dataset in energy bins (shown in Appendix A). Here the measured E_{cal} is used to be independent from the invisible energy correction (see Eqs. (3.4) and (3.5)). $X^{\mathrm{fid}}_{\mathrm{low}}$ and $X^{\mathrm{fid}}_{\mathrm{up}}$ pairs can be estimated for every $\log_{10}(E_{\mathrm{cal}}/\mathrm{eV})$ bins (denoted $\log E_{\mathrm{cal}}$). In Fig. 4.11, an example of these bins is shown. All the fiducial boundaries found with this procedure are shown in Fig. 4.12: blue squares for the lower and red dots for the upper bounds. To apply the **Fiducial FoV** cut as a function of $\log E_{\mathrm{cal}}$, the data points are fitted with a smooth function:

$$\mathrm{FOV}(\log E_{\mathrm{cal}}) = \mathrm{FOV}_0 + K \cdot \frac{\log E_{\mathrm{cal}} - \log E_0}{(\log E_{\mathrm{cal}} - \log E_0)^4 + \omega}. \tag{4.6}$$

This function is not based on a physical model, but rather an empirical function in order to have smooth cut. The parameters used in Eq. (4.6) for blue and red points are shown in Table 4.4. After applying this cut, the field of view bias is removed.

[12] Each bin has a minimum of 20 events and minimum width of $20\,\mathrm{g/cm^2}$.

[13] The fidFoV is considered stable when both the boundaries do not change more than $1\,\sigma$ from one iteration to the other.

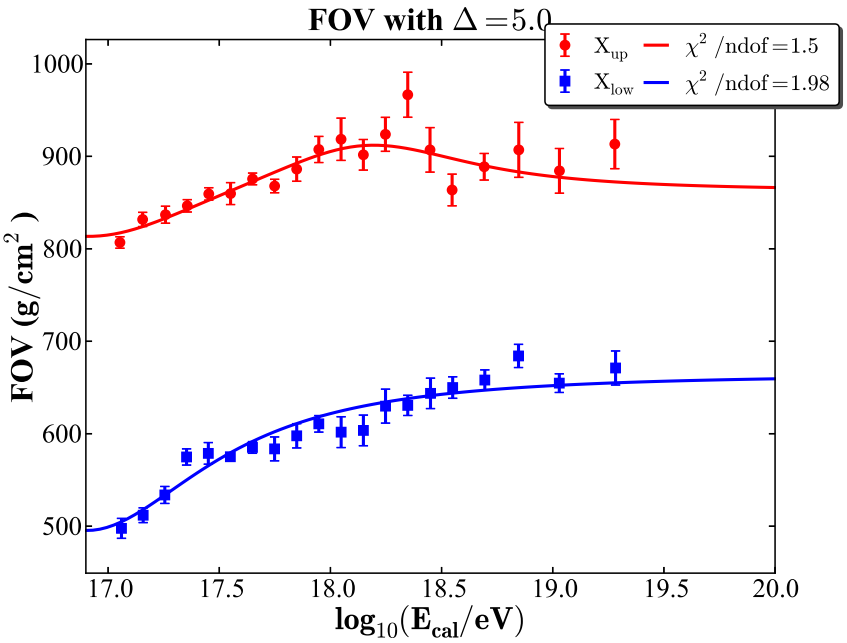


Figure 4.12: Fiducial FoV boundaries as a function of $\log_{10}(E_{\mathrm{cal}}/\mathrm{eV})$. The blue squares are the lower boundaries, the red circles the upper ones. The lines are fits with Eq. (4.6).

4.4 Check of the field of view analysis

4.4.1 Fiducial FoV limit

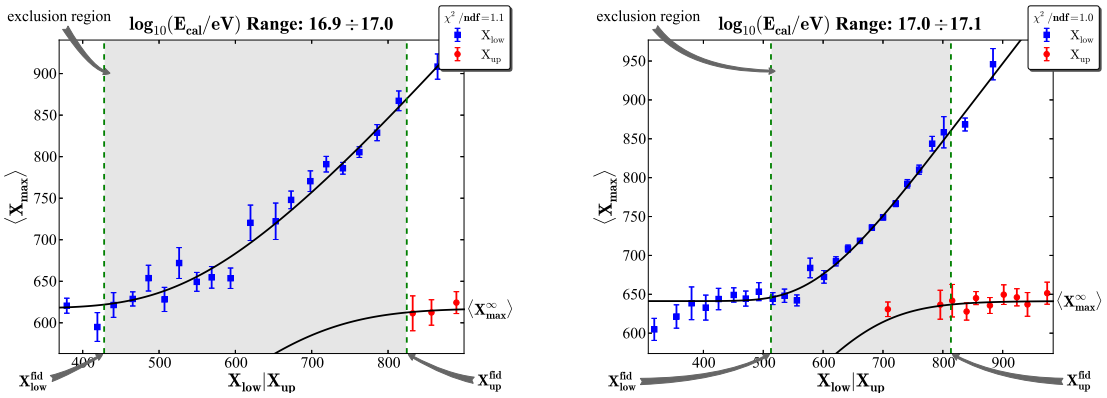
The fiducial FoV analysis minimizes the bias generated by a limited field of view. At low energies the showers must be close to the detector to be measured. But the closer they are, the more limited the field of view is. When the FoV is so small that unbiased events are not observed, the method described in Section 4.3.2 fails.

In Fig. 4.13 the calculation of the fiducial FoV boundaries in the two lowest available energy bins is shown. In Fig. 4.13(a) ($16.9 \leq \log_{10}(E_{\mathrm{cal}}/\mathrm{eV}) < 17.0$) the flat region has too few data points. This region determines the unbiased X_{max} ($\langle X^{\infty}_{\mathrm{max}} \rangle$). Thus, the result from this bin is not trustworthy enough to have an unbiassed sample. The result shown in Fig. 4.13(b) ($17.0 \leq \log_{10}(E_{\mathrm{cal}}/\mathrm{eV}) < 17.1$), however, has an evidently flat region. This determines the lowest energy bin in Fig. 4.12. For $\log_{10}(E_{\mathrm{cal}}/\mathrm{eV}) < 16.9$, not enough events are available to apply this method.

4.4.2 Field of view bias in standard telescopes and HeCo

Splitting the data set in the standard telescopes (CO+LL+LM+LA) and HeCo, the magnitude of the field of view bias can be studied and compared. In Fig. 4.14, $\langle X_{\mathrm{max}} \rangle$ of this two sets is shown: red squares and blue dots, respectively. They are shown both before (full markers) and after (open symbols) the fiducial cut.

A limited FoV leads to more field of view bias at low energies, as already explained in Section 4.4.1. This behavior is clearly observed in both the telescope sets. The standard telescopes



(a) $16.9 \leq \log_{10}(E_{\mathrm{cal}}/\mathrm{eV}) < 17.0$. (b) $17.0 \leq \log_{10}(E_{\mathrm{cal}}/\mathrm{eV}) < 17.1$.

Figure 4.13: Calculation of the fiducial FoV boundaries in the two lowest available energy bins. $\langle X_{\mathrm{max}} \rangle$ as a function of X_{up} and X_{low} are the two data sets represented by red square and blue dots respectively. Black lines are the simultaneous fit of both the data set with μ_{trunk} (Eq. (4.4)).

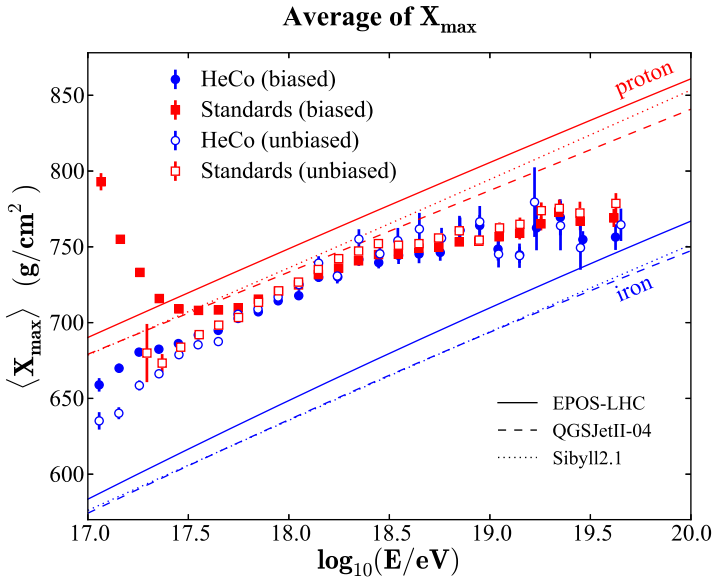


Figure 4.14: Comparisons of $\langle X_{\mathrm{max}} \rangle$ measurements between HeCo and the standard telescopes discussed in 4.4.2, before (full markers) and after (open symbols) the fidFoV cut. HeCo is represented by blue dots and standard telescopes by red squares.

have a more limited FoV compared to HeCo. Thus they have a larger bias. However, at these energies, the bias clearly dominates over the data in both cases. After the selection, both elongation rates agree.

The $\langle X_{\mathrm{max}} \rangle$ measurements shown in Fig. 4.14 as open symbols are without any further corrections. The overall first moment with all the proper adjustment from the analysis will be shown in Section 7.2.

4.4.3 Expected vs. fiducial FoV

One further cross check on the fiducial FoV method is the rejection of an event expected to be measured with a poor quality. In a high quality event, X_{max} belongs to the expect field of view region. Meanwhile, as already explained in Section 4.3.2, the fiducial volume cut rejects those events in which the expected FoV is not large enough to have an unbiased sample. Thus, if an event is accepted by the fiducial cut, it has a very low probability to have a large expected FoV without X_{max} inside. However, the $\boldsymbol{X_{\mathrm{max}}}$ **observed in the expected FoV** cut rejects also events without the maximum of the profile directly measured. This value can be still reconstructed through the G-H fit, without any measured signals in its region.

In the end, if the fiducial FoV is valid, switching the order in Table 4.3 should show a strong rejection caused by the **Fiducial FoV** cut and only a tiny additional rejection caused by the other cut. From Table 4.5, this behavior is confirmed. The cut related to the expected FoV rejects only 2.4 % of the events not already rejected by the fiducial volume.

Table 4.5: Field of view cuts with switched order. The efficiencies are relative to the previous cut.

Cut	Condition	Efficiency
Field of view		
Fiducial FoV	-	15.6 %
$\boldsymbol{X_{\mathrm{max}}}$ **observed in the expected FoV**	$\xi \leq 40\,\mathrm{g/cm^2}$; mva $\geq 20°$	97.6 %

CHAPTER 5

Analysis of simulated data

Although the analysis of the data, the data set is not yet with minimal bias. Approximation of the analysis method and shortcoming of the reconstruction chain may still be sources of biases. With the analysis of simulated showers, the results from the data set can be corrected from these sources.

The Auger simulations are based on the Monte Carlo (MC) technique. They are meant to reproduce the real situation as much as possible[1] for a given set of primary particles and hadronic interaction model (Section 2.4.3). Thus, an extensive use of the databases with real data is required. Information about telescope calibrations, uptime periods and atmospheric conditions are stored in Databases (DBs) day by day. Therefore, beside the geometry and the energy of the primary, the event is generated also with a random date. With that date it is possible to access at the correct information from the databases. This is the so-called REALMC.

For this work a variant of this Monte Carlo is used, in which $X_{\max}$ is also generated randomly with an uniform distribution. This distribution can be used in two possible ways:

1. "flat $X_{\max}$": the flatness can be used to study inefficiency factors that distort the measurement of $X_{\max}$ along the distribution range.

2. "weighted MC": the uniform distribution can be weighted event by event to reproduce a different $X_{\max}$ distribution.

The fiducial field of view analysis is an approximation. Thus a tiny bias in the tails of the $X_{\max}$ distribution can be found. The entire selection chain (Tables 4.1 and 4.3) will deform the flat distribution by rejecting more events in the tails. By using the "flat $X_{\max}$" mode it is possible to study the *acceptance of the selection* (Section 5.1).

The distribution of the mass composition of the CR is not known: it is a complicated mixture of primary particles that evolves with energy. With the standard Monte Carlo it is not possible to

[1] $\overline{\text{Off}}\underline{\text{line}}$ is also able to generate an ideal Monte Carlo, with ideal detectors and atmospheric conditions. But for this work it is not considered.

simulate the mass composition. But it is possible to weight the simulations as in the "weighted MC" mode with a known $X_{\max}$ distribution from the data, to reproduce reality as close as possible. In Appendix B such distributions are shown, energy bin by energy bin. They are fitted with the $G \otimes E$ function, which is used to weight the simulations. For the same reason, the simulation is weighted to reproduce the measured energy flux from the Pierre Auger Observatory [25].

This method is used to study the reconstruction biases of the energy and the $X_{\max}$ distribution close to the reality, and the detector resolution as well (Section 5.2). However the shower particles at the ground still depends on the primary particles chosen for the simulations. In this work a mixture 50 % p and 50 % Fe is simulated.

5.1 Acceptance of the selection

Through the simulations with the flat $X_{\max}$, it is possible to study the effect of the selection on the generated showers. The ratio between selected and unselected showers is the so-called *acceptance of the selection* [88]. This acceptance is expected to not be uniform, but to have a central region where the fiducial selection is efficient, and extreme regions where this efficiency drops. It can be described by the empirical function

$$A(x) = \begin{cases} \mathrm{e}^{\frac{x-X_1}{\lambda_1}} & \text{if } x < X_1 \\ 1 & \text{if } X_1 \leq x < X_2 \\ \mathrm{e}^{-\frac{x-X_2}{\lambda_2}} & \text{if } x \geq X_2 \end{cases} . \tag{5.1}$$

The fiducial FoV selection dominates the field of view. Thus, the limits x_1 and x_2 are expected to be approximately the boundaries of this cut. By splitting the data set in energy bins, it is possible to parameterize the acceptance A bin by bin. In Fig. 5.1, the fits obtained with the parametrization in Eq. (5.1) are shown as red lines. The y axis of each figure is multiplied by a normalization factor such that the maximum of the fit comes out as 1.

For each energy bin, a set of X_1, X_2, λ_1 and λ_2 is determined. All these parameters are shown in Fig. 5.2: X_1 and λ_1 as blue squares, X_2 and λ_2 as red dots. To have a smooth representation of Eq. (5.1), they can be parametrized as a function of $\log E_{\mathrm{cal}}$ by the functions

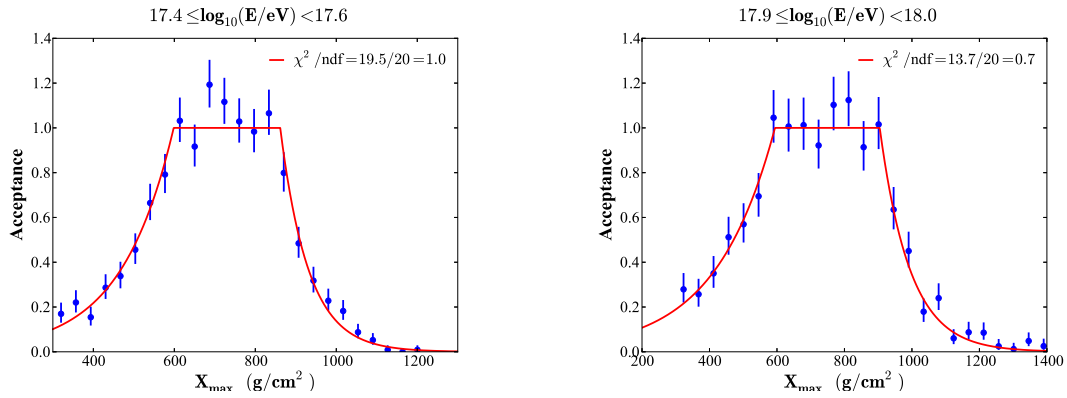
$$X_1 (\log E_{\mathrm{cal}}) = -68.87 \cdot (\log E_{\mathrm{cal}} - 18.06)^2 + 613.05 , \tag{5.2}$$

$$X_2 (\log E_{\mathrm{cal}}) = \begin{cases} 92.95 \cdot (\log E_{\mathrm{cal}} - 18.36) + 119.40 & \text{if } \log E_{\mathrm{cal}} < 18.36 \\ -13.92 \cdot (\log E_{\mathrm{cal}} - 18.36) + 940.05 & \text{otherwise} \end{cases} , \tag{5.3}$$

$$\lambda_1 (\log E_{\mathrm{cal}}) = \begin{cases} -98.64 \cdot (\log E_{\mathrm{cal}} - 17.61) + 119.40 & \text{if } \log E_{\mathrm{cal}} < 17.61 \\ 150.17 \cdot (\log E_{\mathrm{cal}} - 17.61) + 119.40 & \text{otherwise} \end{cases} , \tag{5.4}$$

$$\lambda_1 (\log E_{\mathrm{cal}}) = 54.22 \cdot (\log E_{\mathrm{cal}} - 18.) + 101.96 . \tag{5.5}$$

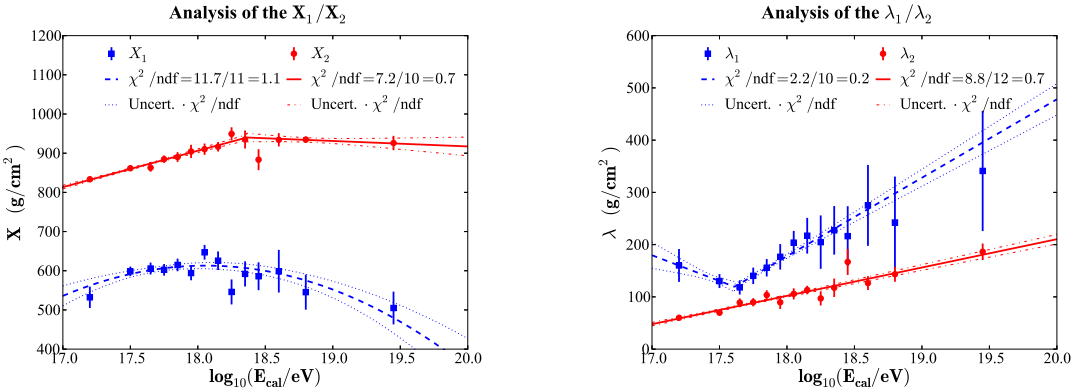
The parametrizations are shown in Fig. 5.2(a) (Eqs. (5.2) and (5.3)) and Fig. 5.2(b) (Eqs. (5.4) and (5.5)). X_1 and λ_1 are represented as blue dashed lines with the error propagation as dotted lines. X_2 and λ_2 are indicated as red solid lines and their error is shown as dash-dotted lines.



(a) $17.4 \leq \log_{10}(E_{\mathrm{cal}}/\mathrm{eV}) < 17.6$

(b) $17.9 \leq \log_{10}(E_{\mathrm{cal}}/\mathrm{eV}) < 18.0$.

Figure 5.1: Examples of the acceptance distribution and their fits (red lines) in two different $\log_{10}(E_{\mathrm{cal}}/\mathrm{eV})$ bins. The y axes are normalized to the the relative fits.



(a) X_1 (blue dashed), X_2 (red solid) and they propagated uncertainties (dotted, dash-dotted respectively).

(b) λ_1 (blue dashed), λ_2 (red solid) and they propagated uncertainties (dotted, dash-dotted respectively).

Figure 5.2: Parametrization of X_1, X_2, λ_1 and λ_2 from Eq. (5.1) as a function of $\log_{10}(E_{\mathrm{cal}}/\mathrm{eV})$. The propagated uncertainties are multiplied by χ^2/ndf from the respective fits.

5.2 Reconstruction biases and the detector resolution

Shortcoming in the reconstruction procedure can bias the variables. To study these biases and correct the results, the weighted MC is used. The variables of interest in this analysis are E and X_{max}. Their reconstruction biases will be studied in the following. Moreover, a correct simulation of the detectors leads to the determination of the resolution for the X_{max} measurement.

5.2.1 Energy bias

The relative difference between the reconstructed (rec) and generated (gen) calorimetric energy

$$\langle E_{\text{bias}} \rangle = \left\langle \frac{E_{\text{cal}}^{\text{rec}} - E_{\text{cal}}^{\text{gen}}}{E_{\text{cal}}^{\text{rec}}} \right\rangle \tag{5.6}$$

gives the magnitude of the energy bias. The bias is shown in Fig. 5.3, bin by bin, as a function of $\log_{10}(E_{\text{cal}}^{\text{rec}}/\text{eV})$ (denoted $\log E_{\text{cal}}^{\text{rec}}$). The fit (red line) is the parametrization of a logarithmic function:

$$\langle E_{\text{bias}}(\log E_{\text{cal}}^{\text{rec}}) \rangle = -14.95 + 22.19 \cdot \log_{10}(\log E_{\text{cal}}^{\text{rec}} - 15.69) \ . \tag{5.7}$$

From Eq. (3.5), the invisible energy is parametrized as a factor f_{invis} that multiplies the calorimetric energy. Thus Eq. (5.7) can be used directly to correct the total observed energy (obs) to get the true one:

$$E^{\text{true}} = E_{\text{cal}}^{\text{obs}} \cdot f_{\text{invis}} \cdot f_{\text{bias}}(E_{\text{cal}}^{\text{obs}}) = E^{\text{obs}} \cdot f_{\text{bias}}(E_{\text{cal}}^{\text{obs}}), \tag{5.8}$$

$$\tag{5.9}$$

where

$$f_{\text{bias}}(x) = 1 - \langle E_{\text{bias}}(\log_{10}(x)) \rangle. \tag{5.10}$$

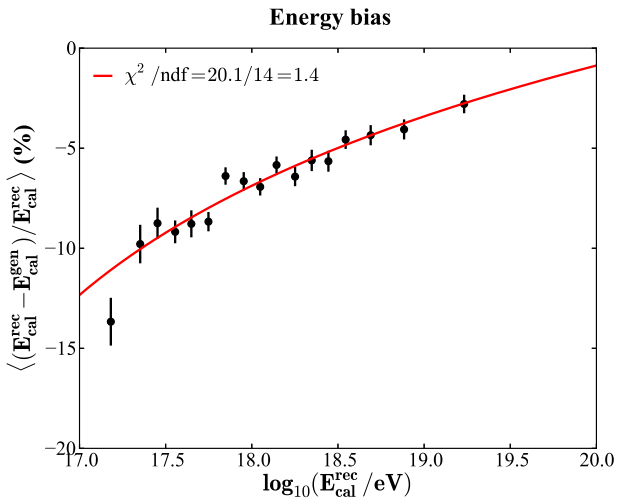


Figure 5.3: Energy bias as a function of $\log_{10}(E_{\text{cal}}/\text{eV})$, expressed in %.

The bias shown in Fig. 5.3 is always negative. Hence the observed energy is always smaller than the true one. Thus, events in the border of the energy bins will migrate toward their next higher bin after Eq. (5.8) is applied. Moreover the bias at low energy ($\log_{10}(E_{\text{cal}}/\text{eV}) < 17.5$) is $\sim -10\,\%$. In this work, the variable of interest is $\log E$. Thus, a bias of $\sim -10\,\%$ is translated to $\log E^{\text{true}} \simeq \log E^{\text{obs}} + 0.04$.

Compared to the official results from the Pierre Auger Collaboration, this bias is larger by $\sim 4\,\%$. The source of this difference is known. To be able to reconstruct HeCo, $\overline{\text{Off}}\underline{\text{line}}$ provides a slightly different reconstruction chain from the official one. Recent studies have also shown

a better performance of the chain used for HeCo compared to the official one, especially when reproducing the universal shower profile.

One of the differences is the approach to the integration of the light collected by the pixels along the shower track [83]. For the official chain, the energy scale [66,67] was optimized recently. To be consistent, the HeCo chain was forced to use the same integration. However the latter is optimized for its own integration approach. It is this incompatibility that causes the larger bias.

5.2.2 X_{max} bias and detector resolution

The difference between the reconstructed (rec) and generated (gen) X_{max}

$$X_{\mathrm{bias}} = X_{\mathrm{max}}^{\mathrm{rec}} - X_{\mathrm{max}}^{\mathrm{gen}} \tag{5.11}$$

leads to the distribution of the bias correction. The average of this distribution gives the magnitude of the reconstruction bias on X_{max}. Moreover, it describes the precision of the detectors to resolve the differences between the real and the observed maximum profiles. This is the detector resolution. The magnitude of this resolution is given by the standard deviation of the distribution from Eq. (5.11).

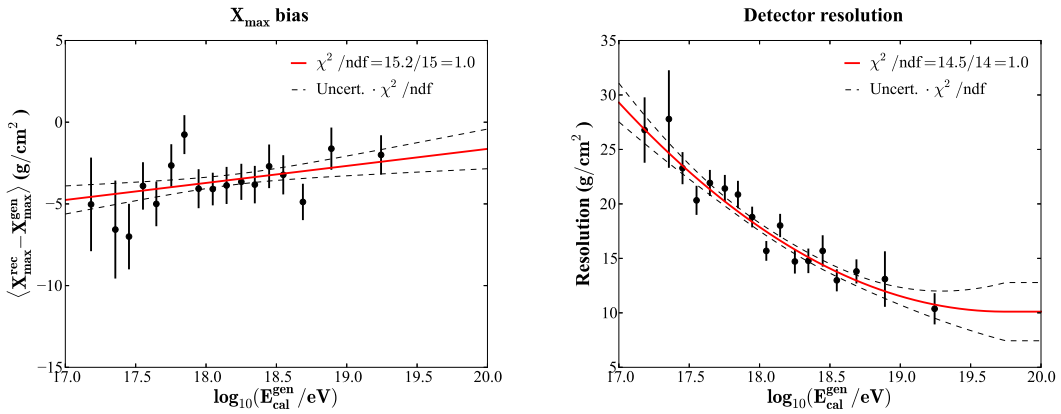
Bias and resolution are shown in Fig. 5.4, bin by bin, as a function of $\log_{10}(E_{\mathrm{cal}}^{\mathrm{gen}}/\mathrm{eV})$ (simply $\log E_{\mathrm{cal}}^{\mathrm{gen}}$). The average X_{bias} and the resolution can be parametrized by

$$\langle X_{\mathrm{bias}} \rangle (\log E_{\mathrm{cal}}^{\mathrm{gen}}) = -3.72 + 1.04 \cdot (\log E_{\mathrm{cal}}^{\mathrm{gen}} - 18.)\,, \tag{5.12}$$

$$\mathrm{Res}_{\mathrm{det}} (\log E_{\mathrm{cal}}^{\mathrm{gen}}) = \begin{cases} 10.11 + 2.54 \cdot (\log E_{\mathrm{cal}}^{\mathrm{gen}} - 19.75)^2 & \text{if } \log E < 19.75 \\ 10.11 & \text{otherwise} \end{cases}, \tag{5.13}$$

as shown Fig. 5.4(a) and Fig. 5.4(b) (red lines). In the figure, the solid dashed lines represent the error propagation of the parametrization.

With these results it is possible to correct the observed $\langle X_{\mathrm{max}} \rangle$ and $\widehat{\sigma}(X_{\mathrm{max}})$ and get the true values. Bias, resolution and fits are shown as a function of the $\log E_{\mathrm{cal}}^{\mathrm{gen}}$ and not $\log E_{\mathrm{cal}}^{\mathrm{rec}}$. The reason is that these correction will be applied after the energy bias correction.



(a) X_{max} bias.

(b) Detector resolution.

Figure 5.4: X_{max} bias and detector resolution as a function of $\log_{10}(E_{\mathrm{cal}}/\mathrm{eV})$. The red lines represent the fits and the dashed black lines their error propagation.

5.3 Validation of the detector simulation

5.3.1 Coihueco vs HEAT in downward mode

CO and HEAT in downward mode should be in agreement when measuring the energy and $X_{\max}$. In Fig. 5.5 the difference in these two variables is shown, in both simulations and data. The differences shown are:

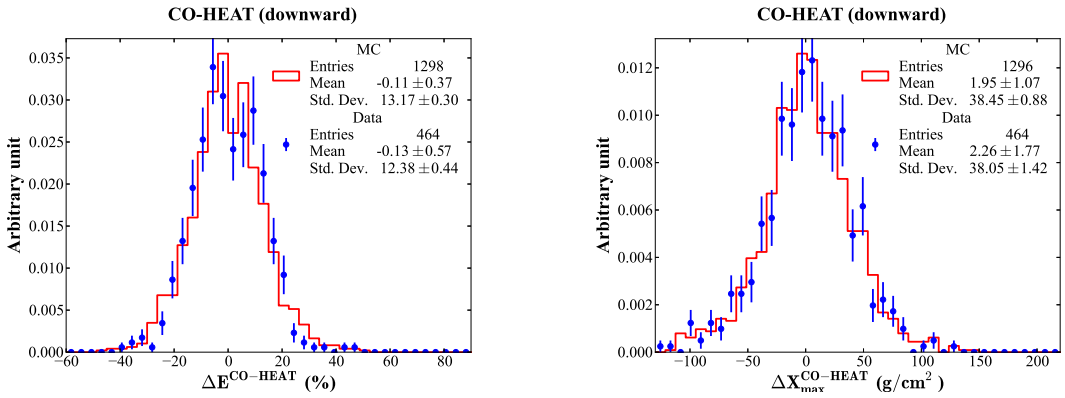
$$\Delta E^{\mathrm{CO-HEAT}} = \frac{E^{\mathrm{CO}} - E^{\mathrm{HEAT}}}{\frac{E^{\mathrm{CO}}+E^{\mathrm{HEAT}}}{2}} \tag{5.14}$$

$$\Delta X_{\max}^{\mathrm{CO-HEAT}} = X_{\max}^{\mathrm{CO}} - X_{\max}^{\mathrm{HEAT}}. \tag{5.15}$$

In case of the energy asymmetry with the MC events (red histograms in Fig. 5.5(a)), the mean difference is ~ 0. In case of the $X_{\max}$ difference (Fig. 5.5(b), red histogram) is not 0. This value is smaller than $\langle X_{\mathrm{bias}} \rangle$, which average is $\approx 3.2\,\mathrm{g/cm^2}$ thus this difference can be considered in agreement within the systematics related to the reconstruction bias (see Section 6.2.1 for further details).

The data (blue dots in Fig. 5.5) show similar results. They are also compatible with the simulations in both mean and standard deviation. This proves that the simulations can correctly resolve the detectors differences. Thus, the estimated resolution (Eq. (5.13)) can be trusted.

According to the above test on Monte Carlo simulation, HeCo can be safely used to do this analysis. The energy is in agreement and $X_{\max}$ has a tiny bias, already considered by the overall reconstruction bias. Moreover the detector resolution can be determined with the use of the simulation.



(a) Energy difference (in %).

(b) $X_{\max}$ difference.

Figure 5.5: Differences between CO and HEAT (in downward) in energy and $X_{\max}$, in both MC simulation (red histogram) and data (blue dots).

5.3.2 R_p test

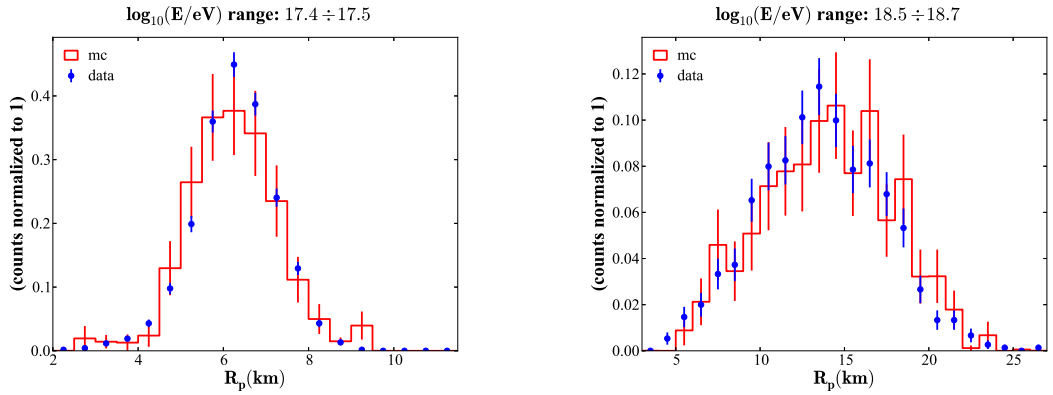
If the simulations correspond to reality, they should be able to reproduce every variable of the showers. These variables must be affected by the data selection process in the same way as the data. To reproduce reality, the weighted MC method is used, i.e. the simulated $X_{\max}$ uniform

distribution and the energy flux are weighted with the $X_{\max}$ distribution and flux obtained from the data.

An independent variable to test the full chain of reconstruction and selection is the distribution of the distance between the telescopes and the shower axis R_p (Fig. 3.10(a)). It is estimated through the geometry reconstruction that uses the light transmission in the atmosphere.

In Fig. 5.6 two examples of this distribution are shown for two different energy ranges. The red histograms with relative error bars are the simulations and the blue dots are the data. In all cases, both distributions are in agreement. Thus, this test is an independent confirmation that the MC is able to reproduce the real events. Hence the flat $X_{\max}$ Monte Carlo can be used to study the biases, the resolution and the acceptance of the selection.

All the R_p comparisons can be found in Appendix D.



(a) $17.4 \leq \log_{10}(E_{\mathrm{cal}}/\mathrm{eV}) < 17.5$

(b) $18.5 \leq \log_{10}(E_{\mathrm{cal}}/\mathrm{eV}) < 18.7$.

Figure 5.6: Examples of the R_p comparison between data (blue) and MC (red) in two different $\log_{10}(E_{\mathrm{cal}}/\mathrm{eV})$ bins.

CHAPTER 6

Systematic uncertainties and resolution

For a proper estimation of the systematic uncertainties of the energy (Section 6.1) and $X_{\max}$ moments (Section 6.2), several sources must be taken into account. In Fig. 6.1 and Table 6.1 (Section 6.3) these sources are summarized, subdivided into groups explained in the following subsections.

Whenever it is not possible to directly propagate the uncertainty of a given source to the variables of interest[1], the usual procedure to estimate the systematic uncertainties is

1. re-reconstruct the data set with the parameters which are considered a physical source of systematics changed;

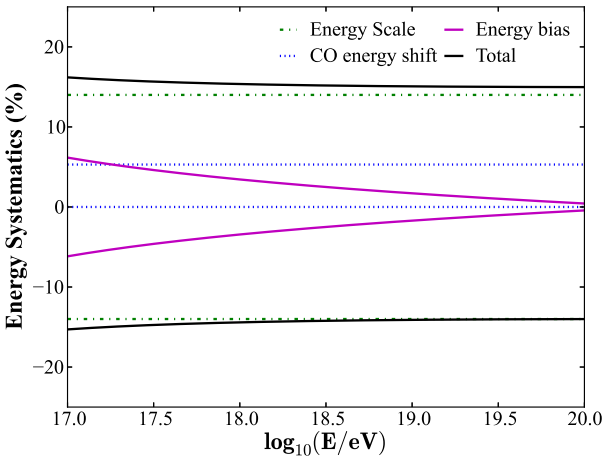
2. calculate the difference between the modified and unmodified reconstruction in the variable of interest.

The average of this difference will lead to the systematic uncertainty in the averaged variable considered. The average energy and $\langle X_{\max} \rangle$ systematics will be indicated as δE and δX respectively.

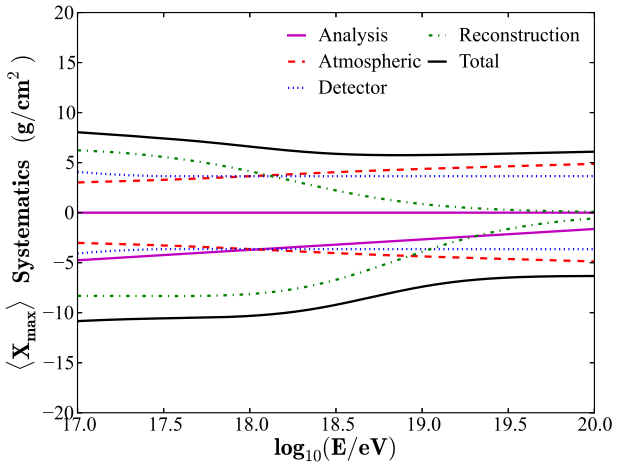
The standard deviation of the difference will lead to a correction of the resolution. In this work, only the resolution of $X_{\max}$ is considered. Thus the variable σ will denote the standard deviation of this difference between shower maxima. Through σ it is possible to estimate the resolution correction and eventually its systematic uncertainty.

All these variables are functions of $\log_{10}(E/\mathrm{eV})$ (denoted $\log E$), unless specified differently.

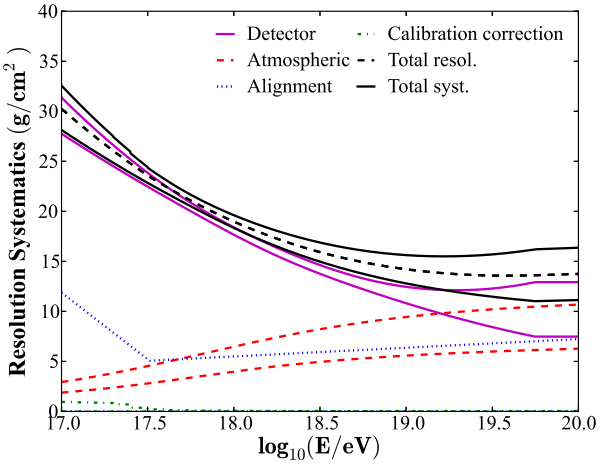
[1] For example, the calibration uncertainty can be propagated directly to $X_{\max}$ systematics



(a) Energy systematic.



(b) X_{max} systematic.



(c) Resolutions and its systematic.

Figure 6.1: Summary of the systematic uncertainties and the resolution corrections. See text for details.

6.1 Energy uncertainty

In Section 3.3.2 the energy uncertainty was introduced, given by the quadratic sum of several factors, up to a total of $\delta \mathrm{E}_{\mathrm{Escale}} \approx \pm 14\,\%$ [67]. In Table 6.1 it is termed as "Energy scale uncertainties" and dominates the overall systematic uncertainties (Fig. 6.1(a), dot-dashed green lines).

However two additional corrections to the energy are done in this work. The first is the choice to shift down the energy of CO due the discrepancy with the HEAT measurement in downward mode (Section 4.1.2). There are no strong reasons to shift down CO, instead of shifting up the other telescopes, thus a $\delta E_{\mathrm{Eshift}} = ^{+5.3\,\%}_{-0\,\%}$ is assigned (Fig. 6.1(a), dotted blue lines) as the systematics of this choice.

The second correction is the reconstruction bias (Section 5.2). To measure the true energy, a factor must be applied to correct the observed one. This factor may not be a real reconstruction shortcoming, but an incompatibility of simulation and reconstruction. Thus the same value must be considered as one-side systematic uncertainty. For this work, instead of applying the full factor, only half of it is used, leading to a symmetric systematic uncertainty. Hence Eq. (5.8) becomes:

$$E^{\mathrm{true}} = E^{\mathrm{obs}} \cdot \frac{1}{2} f_{\mathrm{bias}} \, ^{+f_{\mathrm{bias}}/2}_{-f_{\mathrm{bias}}/2} = E^{\mathrm{obs}} \cdot \frac{1}{2} f_{\mathrm{bias}} \, \delta E_{\mathrm{Ebias}}, \tag{6.1}$$

where f_{bias} (Eq. (5.10)) is a function of $E^{\mathrm{obs}}_{\mathrm{cal}}$ and $\delta E_{\mathrm{Ebias}}$ represents the uncertainty (Fig. 6.1(a), solid magenta lines).

Thus the total systematic uncertainty of the energy is

$$\delta E = \sqrt{\delta E^2_{\mathrm{Escale}} + \delta E^2_{\mathrm{Eshift}} + \delta E^2_{\mathrm{Ebias}}}, \tag{6.2}$$

where the terms can be considered uncorrelated because they are caused by different sources of correction on the energy measurement.

6.2 X_{max} systematic uncertainties and resolution corrections

6.2.1 Analysis

In Section 5.2 a reconstruction bias for X_{max} was estimated too. Since also in this case the bias may not be a shortcoming but an artifact from the simulation/reconstruction incompatibility, a one-side systematic uncertainty is assigned:

$$\delta X_{\mathrm{bias}} = ^{+0}_{-\langle X_{\mathrm{bias}} \rangle} \ \mathrm{g/cm^2}, \tag{6.3}$$

where $\langle X_{\mathrm{bias}} \rangle$ (Eq. (5.12)) is a function of $E^{\mathrm{true}}_{\mathrm{cal}} = E_{\mathrm{cal}} \cdot f_{\mathrm{bias}}/2$.

However, the detector resolution estimated in the same section is well described by the simulations, as discussed in Section 5.3. Thus the uncertainties are due to the error propagation of the detector resolution estimated through the parametrization of Eq. (5.13) (dashed lines in Fig. 5.4(b)). Hence, the resolution is

$$\mathrm{Res}_{\mathrm{det}} = \mathrm{Res} \, ^{+\varepsilon \, \mathrm{Res}_{\mathrm{up}}}_{-\varepsilon \, \mathrm{Res}_{\mathrm{low}}} \ \mathrm{g/cm^2}, \tag{6.4}$$

where Res is a function of $E^{\mathrm{true}}_{\mathrm{cal}}$ and $\varepsilon \, \mathrm{Res}_{\mathrm{up/low}}$ is the upper/lower uncertainty.

Finally also the acceptance correction (Section 5.1) must be taken into account. The acceptance of the selection (Eq. (5.1)) is a function of parametrized values (λ_1, λ_2, X_1 and X_2; see Eqs. (5.2) to (5.5)). Thus, the error propagation of their estimation leads to a systematic uncertainty, propagated directly to the systematics of Eqs. (6.3) and (6.4).

In Figs. 6.1(b) and 6.1(c) the systematic uncertainties from Eqs. (6.3) and (6.4) are represented as solid magenta lines. The acceptance weights are not only functions of the energy as the other estimated systematics, but also of $X_{\max}^{\text{true}}$. Thus it is not possible to give a proper representation of this source of uncertainty in these figures. However, on average they modify the resolution and the uncertainties of $\approx 2\,\%$, which is taken into account in the represented lines.

6.2.2 Atmospheric uncertainty

The atmosphere is constantly monitored due to its importance in the data reconstruction, as explained in Section 3.1.4. Uncertainties on such measurements lead to systematics in $X_{\max}$ reconstruction. The overall of this sources on δX and Res are represented as a dashed red lines in Figs. 6.1(b) and 6.1(c) respectively.

VAOD statistical uncertainty

The Vertical Aerosol Optical Depth (VAOD) is measured by the Central Laser Facility (CLF) and Extreme Laser Facility (XLF) and averaged every hour. This $\langle \text{VAOD} \rangle$ and its variance are stored into the Databases (DBs) [63]. The average is used later to reconstruct the showers and the uncertainty is propagated in the observables of the longitudinal profile.

For low energy showers their close distance limits the Mie scattering, thus a smaller fluctuation is expected. For high energies the limited area covered by the Surface Detector Array (SD) saturates the fluctuation. Therefore, from the study on [89], the average variance is described by the Fermi-function

$$\langle \sigma^2_{\text{atm}} \rangle = 12. \cdot \left[\mathrm{e}^{(17.9 - \log E)/0.28} + 1 \right]^{-1} \left(\mathrm{g/cm^2} \right)^2 . \tag{6.5}$$

As explained in Section 3.3.2, σ^2_{atm} is already accounted in the $X_{\max}$ statistical uncertainty of the data set.

VAOD uniformity

The VAOD measurements are based on the assumption of horizontal uniformity of the aerosol concentration along the distance telescope-CLF/XLF. Within this assumption it is possible to extend the calculation of the Mie attenuation to regions not monitored by the laser facilities. However this hypothesis may be correct only on first approximation.

Re-reconstructing the data switching the VAOD profiles among FD locations leads to an estimation of the systematics given by the uniformity assumption. From [89], the worst cases are used to estimate the systematic uncertainties:

$$\delta X_{\text{VAOD,unif}} = \pm 2.8 + 0.58 \cdot (\log E - 18) \ \ \mathrm{g/cm^2}, \tag{6.6}$$

$$\sigma_{\text{VAOD,unif}} = \sqrt{\frac{\left(14 \cdot \left[\mathrm{e}^{(17.8 - \log E)/0.65} + 1 \right]^{-1} \right)^2 - 2 \cdot \langle \sigma^2_{\text{atm}} \rangle}{2}} \ \mathrm{g/cm^2}. \tag{6.7}$$

However, $\sigma_{\text{VAOD,unif}}$ could be over-estimated in this analysis, because using a re-reconstructed shower with aerosol measurement from other sites is less representative of using the correct one.

Thus it is reasonable to consider half of this standard deviation as downward variation to take into account this hypothesis: $\mathrm{Res}_{\mathrm{VAOD,unif}} = \sigma_{\mathrm{VAOD,unif}}\,^{+0}_{-\sigma_{\mathrm{VAOD,unif}}/2}$ g/cm^2. In order to have a symmetric error considering the uniformity assumption, the resolution correction will be

$$\mathrm{Res}_{\mathrm{VAOD,unif}} = 0.75 \cdot \sigma_{\mathrm{VAOD,unif}}\,^{+\sigma_{\mathrm{VAOD,unif}}/4}_{-\sigma_{\mathrm{VAOD,unif}}/4}\ \mathrm{g/cm^2}. \tag{6.8}$$

VAOD systematics

The VAOD measurement also takes into account uncertainties from correlated systematics of the laser energy, FD calibration and the choice of clear reference nights. To propagate them into the reconstruction, the VAOD measurements are shifted by 1σ of these correlated uncertainties. From [89]

$$\delta X_{\mathrm{VAOD,syst}} = \pm 2. \cdot \left[\mathrm{e}^{(17.9-\log E)/0.4} + 1\right]^{-1}\ \mathrm{g/cm^2}, \tag{6.9}$$

$$\sigma_{\mathrm{VAOD,syst}} = 2.7 \cdot \left[\mathrm{e}^{(17.4-\log E)/0.6} + 1\right]^{-1}\ \mathrm{g/cm^2}. \tag{6.10}$$

A slightly different factor of the aerosol attenuation can lead to a non-negligible statistical fluctuation of the reconstruction. Thus, it is not clear how much of the estimated $\sigma_{\mathrm{VAOD,syst}}$ is given by the VAOD uncertainties propagation and how much by the statistical fluctuation. Therefore, half of it will be applied to correct the resolution, with a symmetric systematic error:

$$\mathrm{Res}_{\mathrm{VAOD,syst}} = \frac{1}{2}\,\sigma_{\mathrm{VAOD,syst}}\,^{+\sigma_{\mathrm{VAOD,syst}}/2}_{-\sigma_{\mathrm{VAOD,syst}}/2}\ \mathrm{g/cm^2}. \tag{6.11}$$

Light yield

As already explained in Section 3.1.4, the Pierre Auger Observatory uses the fluorescence light yield from the measurement of the AIRFLY collaboration. The recent update of this measurement, leads to a propagated uncertainty in $\langle X_{\max}\rangle$ [89,90]:

$$\delta X_{\mathrm{LY}} = \pm 0.4\ \mathrm{g/cm^2}. \tag{6.12}$$

Molecular atmopshere

The measurement of the atmospheric characteristic for the molecular attenuation (temperature, pressure, density), is a combination of the data with the GDAS. However the attenuation can be also calculated using only the measurements from the balloon soundings. The fluctuation of $X_{\max}$ difference between showers reconstructed with these two different attenuation factors leads to a resolution correction [62]:

$$\mathrm{Res}_{\mathrm{molAtm}} = 2.8 + 0.75 \cdot (\log E - 18)\ \ \mathrm{g/cm^2}. \tag{6.13}$$

Multiple scattering

The multiple scattering is treated using the Roberts model [77]. Alternatively the Pekala model [91] can be used. The difference between this two models leads to the systematic uncertainty [92] and the resolution correction [93]:

$$|\delta X_{\mathrm{ms}}| \leq 2.\ \mathrm{g/cm^2} \quad \Rightarrow \quad \delta X_{\mathrm{ms}} = \pm 2\ \mathrm{g/cm^2}, \tag{6.14}$$

$$\mathrm{Res}_{\mathrm{ms}} \leq 1.\ \mathrm{g/cm^2} \quad \Rightarrow \quad \mathrm{Res}_{\mathrm{ms}} = 1\ \mathrm{g/cm^2}. \tag{6.15}$$

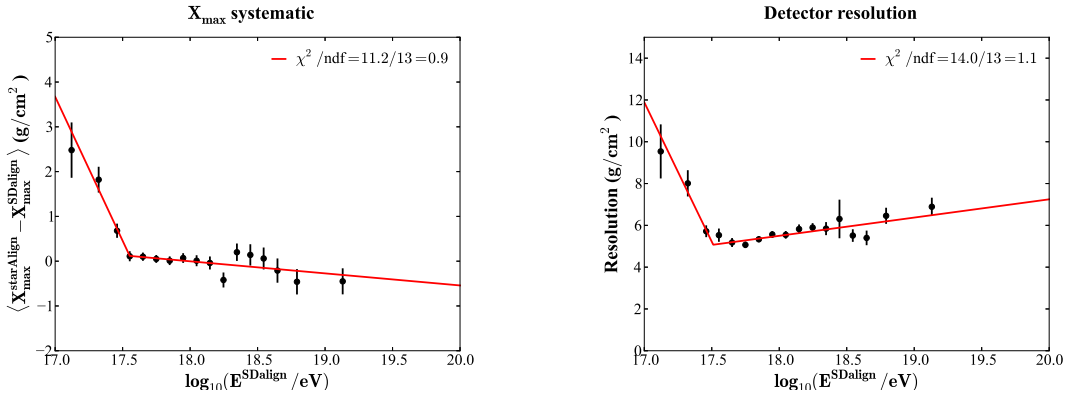
6.2.3 Detector features

The alignment, calibration and time-offsets of the detectors are important to data reconstruction. Slight changes can lead to different results. They are high quality measurements, but they may have uncertainties as well. In Figs. 6.1(b) and 6.1(c), the corresponding overall uncertainty is represented by dotted blue lines.

Telescope alignment

A misalignment of the telescopes leads to a different measurement of $X_{\max}$. For instance if a vertical shower is detected by a telescope with an elevation assumed to be higher than what it really is, $X_{\max}$ will be measured deeper. Near showers will have smaller biases than farther ones because an angular shift has a smaller projection on the shower axis at short distances.

For HeCo the situation is far more complicated. Misalignment between CO and HEAT in upward mode may lead to a different Shower Detector Plane (SDP) identification for shower viewed by both the telescopes. The lower the energy, the shorter the track length is in the cameras. Thus with fewer points to determine the shower detector plane, an azimuthal shift in opposite direction of the CO and HEAT telescopes leads to a different inclination of the SDP. Hence the systematic uncertainty is expected to be larger for low energy showers.



(a) $X_{\max}$ systematic. (b) Resolution.

Figure 6.2: Study of the differences in $X_{\max}$ measurements given by a different alignment procedures: "star alignment" (*starAlign*) − "SD alignment" (*SDalign*). See text for details.

The alignments in the Pierre Auger Observatory are derived from SD-FD cross calibration [94]. Previously they were determined from star tracks and laser shots [95–98]. In Fig. 6.2, they are labelled *SDalign* and *starAlign* respectively. The average of the difference and the resolution are shown in Figs. 6.2(a) and 6.2(b). In Fig. 6.2(b), the previously described behavior is visible. In Fig. 6.2(a) the features are also visible, but above $10^{17.5}$ eV the alignment difference enlarges towards negative values.

However, when HeCo starts to dominate over the standard telescopes (below $10^{17.5}$ eV), the difference increases again. In fact, switching the alignment method leads to an azimuthal difference between CO and HEAT of at most 1.60° in opposite directions.

The parametrization of these differences in $X_{\max}$ can be described by

$$\Delta X_{\text{align}} = \begin{cases} 0.12 - 6.43 \cdot (\log E - 17.55) & \text{if } \log E < 17.55 \\ 0.12 - 0.27 \cdot (\log E - 17.55) & \text{otherwise} \end{cases}, \tag{6.16}$$

$$\sigma_{\text{align}} = \begin{cases} 5.07 - 13.34 \cdot (\log E - 17.51) & \text{if } \log E < 17.51 \\ 5.07 + 0.87 \cdot (\log E - 17.51) & \text{otherwise} \end{cases}. \tag{6.17}$$

$$\tag{6.18}$$

The official alignment is derived from the SDalign. However it is not clear which, between the SDalign and the starAlign, is the true setting. Thus, half of Eq. (6.16) will be added to $\langle X_{\max} \rangle$. Therefore, the systematics and the resolution will be

$$\delta X_{\text{Align}} = ^{+|\Delta X_{\text{align}}|/2}_{-|\Delta X_{\text{align}}|/2} \text{ g/cm}^2, \tag{6.19}$$

$$\text{Res}_{\text{align}} = \frac{1}{2}\sigma_{\text{align}} \, ^{+\sigma_{\text{align}}/2}_{-\sigma_{\text{align}}/2} \text{ g/cm}^2. \tag{6.20}$$

Since uncertainties are added in quadrature, the absolute values of ΔX_{align} can be considered.

Calibration

As noted in [87], an elevation-dependent bias in the calA[2] constants induces uncertainties of the pixel calibration that affect $X_{\max}$. The various updates of the calA have not induced significant changes in the measurement of the maximum of the shower profile. Thus the contribution to the systematic uncertainty has been estimated to be

$$|\delta X_{\text{calib}}| \leq 1 \text{ g/cm}^2 \quad \Rightarrow \quad \delta X_{\text{calib}} = \pm 1 \text{ g/cm}^2. \tag{6.21}$$

FD-SD time offset

Slight changes in the time synchronization between FD and SD lead to different geometry reconstructions, hence to a slightly different $X_{\max}$. In [87] an estimation of this systematic was done for the standard telescopes. More recently, an estimate for HEAT and HeCo was given in [99]. The largest possible value is assumed:

$$|\delta X_{\text{timeOffset}}| \leq 3.5 \text{ g/cm}^2 \quad \Rightarrow \quad \delta X_{\text{timeOffset}} = \pm 3.5 \text{ g/cm}^2. \tag{6.22}$$

6.2.4 Corrections

During the reconstruction and analysis, several corrections are applied to $X_{\max}$. In Figs. 6.1(b) and 6.1(c), the corresponding overall uncertainty is represented by dot-dashed green lines.

Lateral width correction

The "lateral width correction" [66] is a phenomenological parametrization of the light outside the light track on the pixels in the camera described by

$$\text{LW} = 14.8 \cdot \left[e^{(\log E - 18.68)/0.43} + 1 \right]^{-1} \text{ g/cm}^2. \tag{6.23}$$

[2]The light source for calA source is an LED with a diffusor, mounted in the center of the mirror and directly pointed at the camera, as explained in Section 3.1.3.

An attempt to understand the cause for this correction and its energy dependence is still ongoing. Recent studies [100] have shown that a simple implementation of the measured point spread function of the FD telescopes in the simulation is not enough. Thus, conservatively a one-sided systematic uncertainty must be considered:

$$\delta X_{\mathrm{LW}} = ^{+0}_{-\,\mathrm{LW}} \,.$$

However the algorithm to measure the LW correction has a bias when applied to simulated showers [90]. Its parametrization can be described by [89]

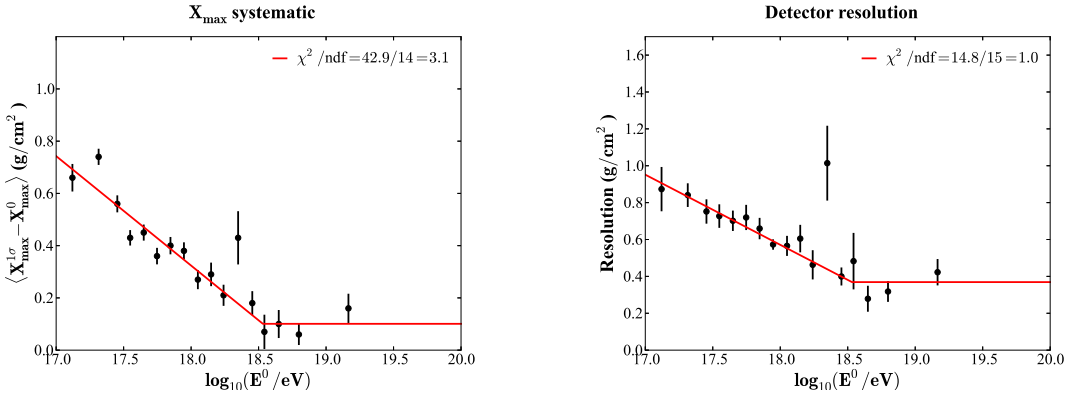
$$\mathrm{b_{LW}} = 6.5 \cdot \left[\mathrm{e}^{(\log E - 18.23)/0.41} + 1\right]^{-1} \ \mathrm{g/cm^2}. \tag{6.24}$$

This bias will be applied to correct $\langle X_{\mathrm{max}} \rangle$. Thus the systematic uncertainties for the lateral width correction will be

$$\delta X_{\mathrm{LW}} = ^{+\,\mathrm{b_{LW}}}_{-(\mathrm{LW}\,-\,\mathrm{b_{LW}})} \,. \tag{6.25}$$

CO calibration adjustment

The energy shift applied in CO to correct the calibration is measured with the discrepancy on the energy between CO and HEAT in downward mode (Section 4.1.2). The uncertainty of this measurement is 0.6 % and its propagation leads to the systematic uncertainty of this correction.



(a) X_{max} systematic. (b) Resolution.

Figure 6.3: Changes of X_{max} measured in HeCo due to the propagation of the uncertainty of the estimated energy shift from Section 4.1.2. The X_{max} reconstructed with the furthermore energy shift of $1\,\sigma$ is labelled $X^{1\sigma}_{\mathrm{max}}$. The red line is the fit.

In Fig. 6.3 the average difference of X_{max} in HeCo and its resolution with and without this further shift of $1\,\sigma$ are shown. Shifting HEAT up or CO down the calibration leads to the same result because what generates this bias in X_{max} is the discrepancy between CO and HEAT. The

parametrization of these differences is given by

$$\text{calAdj} = \pm \begin{cases} 0.1 - 0.42 \cdot (\log E - 18.53) & \text{if } \log E < 18.53 \\ 0.1 & \text{otherwise} \end{cases}, \tag{6.26}$$

$$\sigma_{\text{calAdj}} = \pm \begin{cases} 0.37 - 0.38 \cdot (\log E - 18.53) & \text{if } \log E < 18.53 \\ 0.37 & \text{otherwise} \end{cases}. \tag{6.27}$$

$$\tag{6.28}$$

The behavior shown in Fig. 6.3 is alike the one shown on Fig. 4.3. Whenever the CO part starts to dominate in HeCo ($\log E > 18.5$), this bias saturates.

The calibration shift can be assumed trustworthy because it forces the difference between the telescopes to be exactly 0, validated by Fig. 5.5(a). Thus, Eqs. (6.26) and (6.27) are used as one-side uncertainties:

$$\delta X_{\text{calAdj}} = ^{+|\text{calAdj}|}_{-0} \ \text{g/cm}^2, \tag{6.29}$$

$$\text{Res}_{\text{calAdj}} = 0.^{+\sigma_{\text{calAdj}}}_{-0} \ \text{g/cm}^2. \tag{6.30}$$

Moreover, this systematic uncertainty affects only HeCo, thus will be furthermore reduced in the average between the different telescope sets (Section 7.1.1).

6.3 Summary

The several contributions to the systematic uncertainties of the energy and X_{max}, and to the resolution corrections as well, are summarized in Table 6.1. To quantify these contributions, their values at 1 EeV are also reported. However, as shown in Fig. 6.1, they are not generally constant, but functions of the $\log_{10}(E/\text{eV})$.

Table 6.1: Summary of the systematic uncertainties and resolution corrections.

Type (referring to legends in Fig. 6.1)	Sources	@ 10^{18} eV
Energy scale (Section 6.1)	Energy scale uncertainties	14 %
Energy shift (Section 6.1)	Energy shift of CO	$^{+5.3\,\%}_{-0\,\%}$
Energy bias (Section 6.1)	Energy bias correction	$\approx 3.5\,\%$
	Total	$\approx^{+15.4\,\%}_{-14.4\,\%}$
Analysis (Section 6.2.1)	Reconstruction Bias correction	$\approx^{+0\,\mathrm{g/cm^2}}_{-3.7\,\mathrm{g/cm^2}}$
	Acceptance correction	$\lesssim 0.7\,\mathrm{g/cm^2}$
Atmospheric (Section 6.2.2)	VAOD uniformity	$\approx 2.8\,\mathrm{g/cm^2}$
	VAOD systematics	$\approx 5\,\mathrm{g/cm^2}$
	Multiple scattering	$\leq 2\,\mathrm{g/cm^2}$
	Light yield uncertainties	$0.4\,\mathrm{g/cm^2}$
Detector (Section 6.2.3)	Telescope Alignment	$\approx 0\,\mathrm{g/cm^2}$
	Calibration	$\leq 1\,\mathrm{g/cm^2}$
	FD-SD time offset	$\leq 3.5\,\mathrm{g/cm^2}$
Reconstruction (Section 6.2.4)	Lateral width correction	$4.6\,\mathrm{g/cm^2}$
	CO calibration adjustment	$\leq 0.1\,\mathrm{g/cm^2}$
	Total	$\approx^{+8.5\,\mathrm{g/cm^2}}_{-9.3\,\mathrm{g/cm^2}}$
Detector (Section 6.2.1)	Detector resolution	$\approx 18\,\mathrm{g/cm^2}$
	Acceptance correction	$\approx 4\,\mathrm{g/cm^2}$
Atmospheric (Section 6.2.2)	VAOD statistical uncertainty	$\approx 2.7\,\mathrm{g/cm^2}$
	VAOD uniformity	$\approx 5\,\mathrm{g/cm^2}$
	VAOD systematics	$\approx 2\,\mathrm{g/cm^2}$
	Molecular atmosphere	$2.8\,\mathrm{g/cm^2}$
	Multiple scattering	$\leq 1\,\mathrm{g/cm^2}$
Alignment (Section 6.2.3)	Telescope Alignment	$4.6\,\mathrm{g/cm^2}$
Calibration correction (Section 6.2.4)	CO calibration adjustment	$\leq 0.1\,\mathrm{g/cm^2}$
	Total	$\approx 20.2\,\mathrm{g/cm^2}$

Energy systematics

$\langle X_{\mathrm{max}} \rangle$ systematics

Resolution correction

CHAPTER 7

Results

The analysis described in the previous chapters has the purpose to obtain $X_{\max}$ measurements of high quality and with minimal bias. Together with the knowledge of the systematic uncertainties, this gives the foundation for the measurement of the moments of the $X_{\max}$ distribution.

In Section 7.1 the derivation of the first two moments of the distribution will be illustrated taking into account all the corrections studied with the analysis. These moments will then be discussed in Section 7.2 and compared to previous results from other experiments. An interpretation in terms of the mass composition of cosmic rays will be finally described in Section 7.3.

7.1 Calculation of $\langle X_{\max} \rangle$ and $\widehat{\sigma}(X_{\max})$

7.1.1 Averaging of events detected by multiple locations

Each air shower can be detected by multiple FD locations, especially at high energies for which showers have a large amount of light produced in the atmosphere. Each FD Event from the same shower has a different quality of the measured profile. Thus, these events must be averaged with weights to have single measurement of the observable for each shower. These weights are

inversely proportional to the variance of the measurement itself:

$$\sigma^2_{\mathrm{E}} = \left(\sum_i \frac{1}{\sigma^2_{\mathrm{E^i}}} \right)^{-1} \tag{7.1}$$

$$\sigma^2_{\mathrm{X_{max}}} = \left(\sum_i \frac{1}{\sigma^2_{\mathrm{X^i_{max}}}} \right)^{-1} \tag{7.2}$$

$$E^{\mathrm{obs}} = \sigma^2_{\mathrm{E}} \cdot \sum_i \frac{E^i}{\sigma^2_{\mathrm{E^i}}} \tag{7.3}$$

$$X^{\mathrm{obs}}_{\mathrm{max}} = \sigma^2_{\mathrm{X_{max}}} \cdot \sum_i \frac{X^i_{\mathrm{max}}}{\sigma^2_{\mathrm{X^i_{max}}}}, \tag{7.4}$$

where $\sigma_{\mathrm{E^i}}$ and $\sigma_{\mathrm{X^i_{max}}}$ are the estimated uncertainties of E^i and X^i_{max}, and the sum is over the FD Events.

The possible FD Events usable in these averages are only from the four locations, situated at the border of the experimental area: Los Leones (LL), Los Morados (LM), Loma Amarilla (LA), and Coihueco (CO). HeCo, the merged telescope set from High Elevation Auger Telescopes (HEAT) and CO telescopes, is important for the low energy measurement and is also located in the CO site. Thus, the events measured from this virtual location will be used, replacing CO events whenever is possible, i.e. data acquired between June 2010 and the 16th August 2012.

7.1.2 Correction of the events

The observable from Eqs. (7.3) and (7.4), however, are not yet the "true" measurements. They must be corrected for the reconstruction biases, the detector resolution and the acceptance of the selection.

First the energy is corrected by its bias factor (see Eq. (5.10)) to get the true energy E. After this correction, all the energies in the functions can be considered already correct, i.e. "true".

For a better understanding of how the true and measured X_{max} distributions are related, it is useful to consider again the approximate description of the true X_{max} distribution by the $G \otimes E$ function. E and G are an exponential and a Gaussian distributions. As explained in Section 5.1, even after applied the selection criteria of the fiducial field of view (Section 4.3.2), the distribution are still distorted in the tails because of the non-flat acceptance A. The reconstruction contribution to the $X^{\mathrm{obs}}_{\mathrm{max}}$ can be described as a Gaussian too, with the central value as $\widehat{X}_{\mathrm{bias}}$ and the spread represented by the resolution: $G_{\mathrm{reco}}(\mu = \widehat{X}_{\mathrm{bias}}, \sigma = \widehat{\mathrm{Res}})$. Thus the observed X_{max} distribution can be approximately described as

$$\begin{aligned} f\left(X^{\mathrm{obs}}_{\mathrm{max}}\right) &= \left\{ \left[f\left(X^{\mathrm{true}}_{\mathrm{max}}\right) \cdot A \right] \otimes G_{\mathrm{reco}} \right\} \approx \left[(G \otimes E \cdot A) \otimes G_{\mathrm{reco}} \right] \\ &\approx (G \otimes E \otimes G_{\mathrm{reco}}) \cdot A = \left[(G \otimes G_{\mathrm{reco}}) \otimes E \right] \cdot A. \end{aligned} \tag{7.5}$$

The approximation in Eq. (7.5) holds true in case of $\mathrm{Var}(G \otimes E) \gtrsim \mathrm{Var}(G_{\mathrm{reco}})$ and an acceptance $A = 1$ for a large portion of the depth, i.e. small part of the tails distorted, which is the case for this analysis.

The values $\widehat{X}_{\mathrm{bias}}$ and $\widehat{\mathrm{Res}}$ are the overall bias and resolution. They have been extensively discussed in Sections 5.2.2 and 6.2 and can be summarized as

$$\widehat{X}_{\mathrm{bias}} = \langle X_{\mathrm{bias}} \rangle + \mathrm{b_{LW}} - \frac{1}{2}\Delta X_{\mathrm{align}}, \tag{7.6}$$

$$\widehat{\mathrm{Res}}^2 = \mathrm{Res}^2_{\mathrm{det}} + \sigma^2_{\mathrm{atm}} + \mathrm{Res}^2_{\mathrm{VAOD,unif}} + \mathrm{Res}^2_{\mathrm{VAOD,syst}} + \mathrm{Res}^2_{\mathrm{molAtm}} + \mathrm{Res}^2_{\mathrm{ms}} + \mathrm{Res}^2_{\mathrm{align}}\,. \tag{7.7}$$

The total bias is given by the reconstruction and the lateral width correction biases ($\langle X_{\text{bias}} \rangle$ and b_{LW}), and the telescope alignment adjustment $\Delta X_{\text{align}}/2$. The total resolution is the sum in quadrature of the detector (Res_{det}), atmospheric and telescope alignment resolutions.

The lateral width correction is a phenomenological parametrization of the light outside the light track on the pixels in the camera. Its analysis method has a tiny bias b_{LW} (Eq. (6.24)), that is considered properly. The detector resolution Res_{det} is the variance of the difference between true and reconstructed $X_{\max}$ distribution, estimated through simulations (Section 5.2 and Eq. (5.13)). The average of such a difference is the reconstruction bias $\langle X_{\text{bias}} \rangle$ (Eq. (5.12)). The atmospheric resolution considers several effects given by the Vertical Aerosol Optical Depth (VAOD) studies (σ_{atm}[1], $\text{Res}_{\text{VAOD,unif}}$ and $\text{Res}_{\text{VAOD,syst}}$), as well molecular attenuation ($\text{Res}_{\text{molAtm}}$) and multiple scattering (Res_{ms}). All of them are obtained by the variance of the comparison between the measurements and re-reconstructed data set with the parameters which are considered a physical source of systematics changed (Section 6.2.2). The same procedure is used to estimated the telescope alignment resolutions $\text{Res}_{\text{align}}$ (Section 6.2.3). The Pierre Auger Observatory provides two different alignment settings. The choice of one of them is properly considered with the $\Delta X_{\text{align}}/2$ adjustment (Eq. (6.16)). All comparisons are approximately Gaussian distributed. Thus their variances can be summed up, as shown in Eq. (7.7).

$X^{\text{obs}}_{\max} - \widehat{X}_{\text{bias}} = X^{\text{corr}}_{\max}$ gives the estimator of the true $X_{\max}$. From the properties of the convolution, the two Gaussians can be combined together, summing up in quadrature their standard deviations. The variance of $G \otimes E$ distribution is the sum of the variances of the exponential and Gaussian distributions. Thus, once the acceptance is treated properly, the resolution can be simply subtracted in quadrature to obtain the standard deviation of the $X_{\max}$ distribution.

The acceptance A is a function of $\log_{10}(E_{\text{cal}}/\text{eV})$ and $X^{\text{true}}_{\max} \simeq X^{\text{corr}}_{\max}$. Therefore each event will have its own factor, which behaves as a weight that reduce the tails of the $X^{\text{obs}}_{\max}$ distribution. Hence, to calculate the moments of the $X_{\max}$ distribution, each $X_{\max}$ value must be weighted by

$$w(\log_{10}(E_{\text{cal}}/\text{eV}), X^{\text{corr}}_{\max}) = \frac{1}{A(\log_{10}(E_{\text{cal}}/\text{eV}), X^{\text{corr}}_{\max})}. \tag{7.8}$$

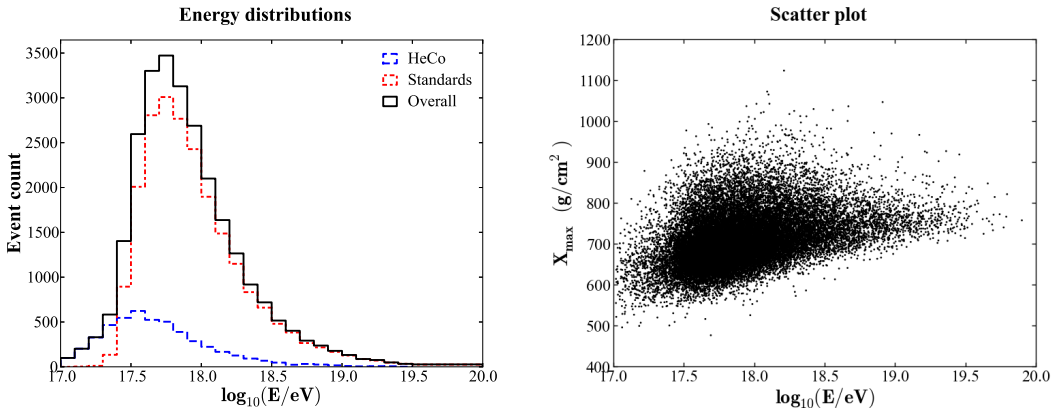
The acceptance is a small effect and it modifies $\langle X_{\max} \rangle$ and $\widehat{\sigma}$, on average, less than $1\,\text{g/cm}^2$ and $2.5\,\text{g/cm}^2$ respectively, and no more than $3\,\text{g/cm}^2$ and $7\,\text{g/cm}^2$. This small magnitude of the correction is also confirmed by an different method of analysis [101].

Applying the weights, the first two moments of the $X_{\max}$ distribution are

$$\langle X_{\max} \rangle = \frac{\sum w \left(X^{\text{obs}}_{\max} - \widehat{X}_{\text{bias}} \right)}{\sum w} = \frac{\sum w\, X^{\text{corr}}_{\max}}{\sum w}, \tag{7.9}$$

$$\widehat{\sigma}^2 = \frac{(\sum w)^2}{(\sum w)^2 - \sum w^2} \cdot \frac{\sum w\, (X^{\text{corr}}_{\max} - \langle X_{\max} \rangle)^2}{\sum w} - \widehat{\text{Res}}^2 = \widehat{\sigma}'^2 - \widehat{\text{Res}}^2, \tag{7.10}$$

where $\widehat{\sigma}'^2$ is the unbiased estimator of the variance. The estimation of statistical and systematic uncertainties of $\langle X_{\max} \rangle$ and $\widehat{\sigma}$ are discussed in detail in Appendix F.



(a) $\log_{10}(E/\mathrm{eV})$ distribution

(b) Scatter plot of $X_{\max}$

Figure 7.1: Distribution of the events in energy and $X_{\max}$.

7.2 Moments of the $X_{\max}$ distribution

The distribution of the bias-corrected energy and $X_{\max}$, i.e. $X_{\max}^{\mathrm{obs}} - \widehat{X}_{\mathrm{bias}}$, are shown in Fig. 7.1. The energy distribution (Fig. 7.1(a)) is shown as a function of $\log_{10}(E/\mathrm{eV})$. The black solid line represents the overall distribution of the events used in this work. The red dot-dashed line represents the statistics of the standard telescopes: Lx[2] and CO, when HeCo is not available. The blue dashed line represents the HeCo distribution. The latter has a data sample of only two years, compare to the eight years for the other telescopes. Nevertheless, as shown in the figure, below $10^{17.5}$ eV HeCo dominates the sample. It is worthwhile noting that the overall distribution is not the sum of the two distributions, but the event are combined according to Eq. (7.3).

In Fig. 7.1(b) the the scatter plot of $X_{\max}$ as a function of $\log_{10}(E/\mathrm{eV})$ is shown. The sharp edge at low $X_{\max}$ values can be explained as the edge of the heaviest composition in the CR mixture in a given energy. The upper edge is correlated with the change of the mass composition. The larger the width of the distribution, the more mixed the composition is. As can be seen, deeply penetrating showers are less frequent at low and high energies. By measuring the average and the standard deviation of such a distribution, it is possible to gain a better understanding of the sample.

Subdividing the data set into bins of $\log_{10}(E/\mathrm{eV})$ (denoted $\log E$), the $\langle X_{\max}\rangle$ and $\widehat{\sigma}$ measurements are shown in Fig. 7.2. The average value of the energy in each bin is

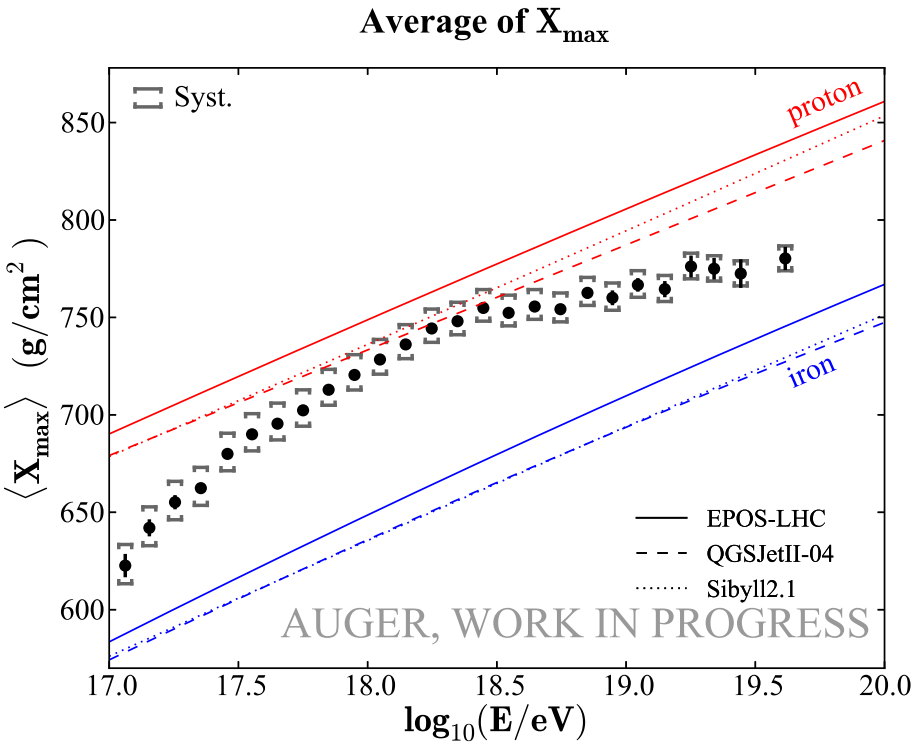
$$\langle \log_{10}(E/\mathrm{eV})\rangle = \frac{\sum w \, \log E}{\sum w}. \tag{7.11}$$

The binning used has a step of 0.1 in logarithmic scale below $10^{19.5}$ eV. Above this value one last bin is considered, as the integral of the events with energy $> 10^{19.5}$ eV, which are 30 showers. The most energetic event observed with the data set used in this work has an energy of $8.7 \cdot 10^{19}$ eV.

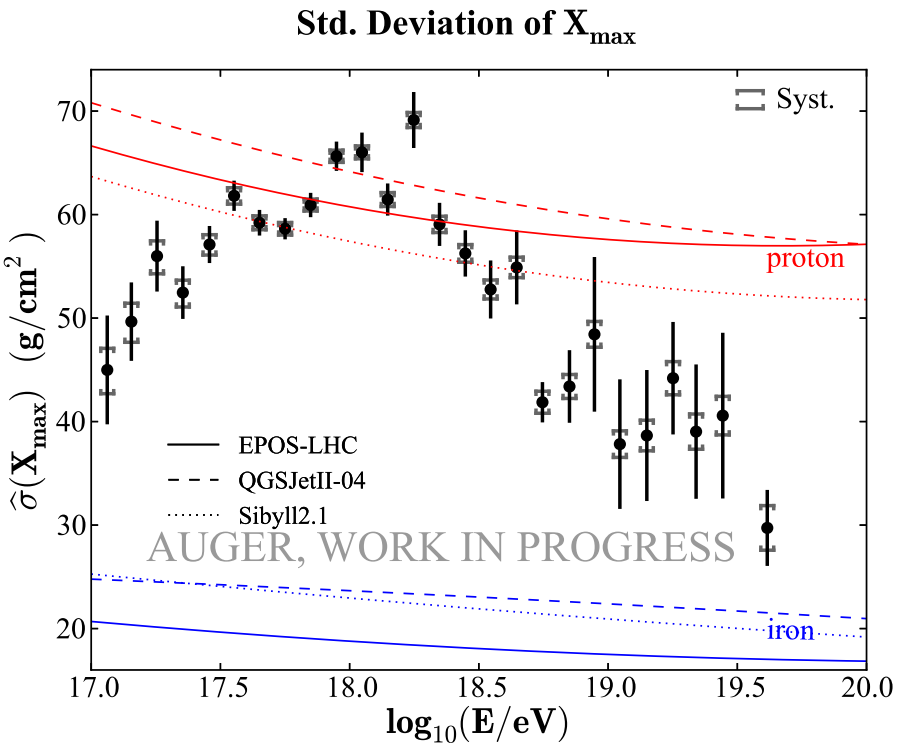
The average of the logarithm of E is used instead the logarithm of $\langle E\rangle$, because $\langle X_{\max}\rangle$ is expected to be linear in $\log E$ (Sections 2.4 and 2.5). The measurements are compared with

[1]The value σ_{atm} is the average statistical uncertainty of the VAOD measurement and is calculated event by event. As mentioned in Section 3.3.2, this term is summed up with the geometry uncertainty of the reconstruction process to form the $\sigma_{X_{\max}^{i}}$ in Eqs. (7.2) and (7.4).

[2]This is a compact notation that indicates the whole set of LL, LM and LA.



(a) $\langle X_{\max} \rangle$ measurement.



(b) $\widehat{\sigma}$ measurement.

Figure 7.2: Results for the moments of the $X_{\max}$ distribution. The values from air shower simulations using the hadronic interaction models (EPOS-LHC, QGSJetII-04, Sibyll2.1) are shown for pure proton and pure iron compositions.

the expected values from air shower simulations using the models (EPOS-LHC, QGSJetII-04 and Sibyll2.1) for pure proton and iron composition. In the $\langle X_{\mathrm{max}} \rangle$ measurement (Fig. 7.2(a)), a linear evolution with structures can be seen. Between $10^{17.5}$ and $10^{18.5}$ eV the data are compatible with the expectations for a light composition. After the structure point at $10^{18.5}$ eV, slightly below the ankle, the composition becomes heavier. One more structure point is observed around $10^{17.5}$ eV, where some experiments have reported a second knee in the flux [4].

The fits of the data (red solid lines) in Fig. 7.2(a) with a single-broken and double-broken lines are shown in Fig. 7.3. Considering the χ^2, the single-broken hypothesis is disfavored: $P(\chi^2 = 48.2, \mathrm{Ndf} = 22) = 1.02 \cdot 10^{-3}$, whereas the double-broken hypothesis has a $P(\chi^2 = 18.3, \mathrm{Ndf} = 20) = 0.57$.

The observed slopes are the so-called *Elongation Rate* (ER) [102, 103]

$$D = D_{10} = \frac{\mathrm{d}\langle X_{\mathrm{max}} \rangle}{\mathrm{d} \log_{10} E}, \tag{7.12}$$

which is a measure of the change of the shower maximum per logarithm of energy. A change in D means a change in the composition evolution, as a function of $\log E$, of the cosmic rays. Defining the double-broken line as

$$\langle X_{\mathrm{max}} \rangle_{\mathrm{fit}}(x) = \langle X_{\mathrm{max}} \rangle_2 + \begin{cases} D_1 \cdot (x - \log E_1) + \langle X_{\mathrm{max}} \rangle_1 & \text{if } x < \log E_1 \\ D_2 \cdot (x - \log E_2) & \text{if } \log E_1 \leq x < \log E_2 \\ D_3 \cdot (x - \log E_2) & \text{if } x \geq \log E_2 \end{cases}, \tag{7.13}$$

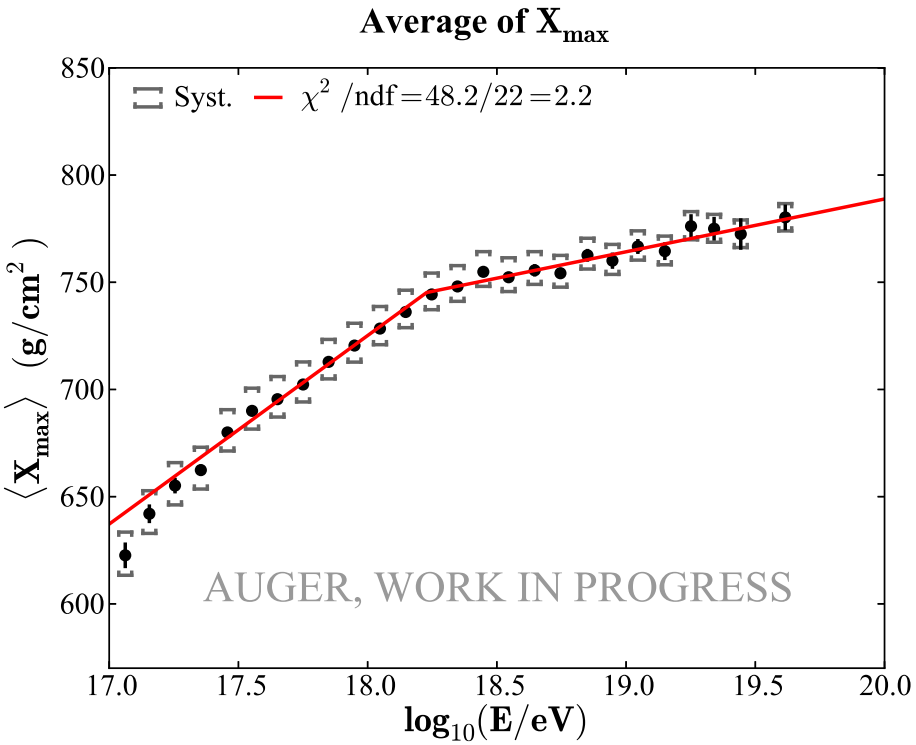
where x is $\log_{10}(E/\mathrm{eV})$ and $\langle X_{\mathrm{max}} \rangle_1 = \langle X_{\mathrm{max}} \rangle_{\mathrm{fit}}(\log E_1)$. $\log E_1$ and $\log E_2$ denote $\log_{10}(E_1/\mathrm{eV})$ and $\log_{10}(E_2/\mathrm{eV})$. From the fit, the parameters are

$$\begin{aligned} \langle X_{\mathrm{max}} \rangle_2 &= 747.19 \pm 2.04|_{\mathrm{stat}} \, {}^{+6.98}_{-9.52}|_{\mathrm{syst}} \\ D_1 &= 131.89 \pm 11.36|_{\mathrm{stat}} \, {}^{+0.31}_{-1.95}|_{\mathrm{syst}} \\ D_2 &= 80.48 \pm 2.33|_{\mathrm{stat}} \, {}^{+1.72}_{-1.08}|_{\mathrm{syst}} \\ D_3 &= 23.87 \pm 2.71|_{\mathrm{stat}} \, {}^{+1.03}_{-2.66}|_{\mathrm{syst}} \\ \log E_1 &= 17.49 \pm 0.04|_{\mathrm{stat}} \pm 0.07|_{\mathrm{syst}} \\ \log E_2 &= 18.27 \pm 0.03|_{\mathrm{stat}} \pm 0.07|_{\mathrm{syst}} \end{aligned}$$

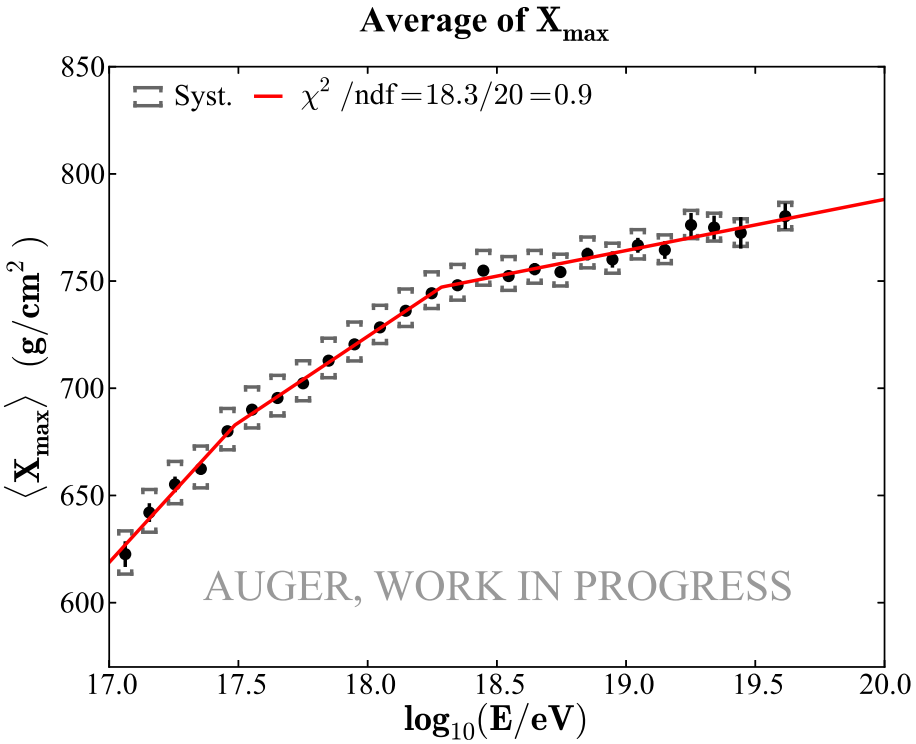
The systematic uncertainties are estimated numerically, shifting the data up and down by their systematic uncertainties. In case of $\log E_1$ and $\log E_2$, the energy uncertainty is also summed up in quadrature to their estimated uncertainty. It is worth nothing that a proper study of the systematics must still be done shifting the individual components of the overall systematic uncertainties to give a proper significance at these measurements.

In Fig. 7.4 the comparison between $\langle X_{\mathrm{max}} \rangle$ measurements from several experiment and this work is shown. Figure 7.4(a) is shown with the energy range used for this work, while Fig. 7.4(b) is shown with a broader overview (i.e. the update of Fig. 2.10).

The experimental results taken into account are from Tunka [104, 105], Yakutsk [106, 107], CASA-BLANCA [108], HiRes/MIA [109], HiRes [110] and Telescope Array (TA) [111]. The latter two do not measure the $\langle X^{\mathrm{true}}_{\mathrm{max}} \rangle$, but $\langle X^{\mathrm{obs}}_{\mathrm{max}} \rangle$. To give the interpretation of the results, these experiments do not unbias the measurements as done in this work. They compare the observed sample with the air shower simulations folded with their detector responses. Therefore, to compare their results to the other experiments and this work, applying a shift Δ of $\langle X_{\mathrm{max}} \rangle$ is



(a) Fit with a single-broken line.



(b) Fit with a double-broken line.

Figure 7.3: $\langle X_{\max} \rangle$ measurement and the fit (red solid line) of its elongation rate.

necessary in their data [23]. In the figures also the results presented at the International Cosmic Ray Conference (ICRC) 2013 by the Pierre Auger Collaboration are shown.

In Fig. 7.4(a) a general agreement between the experiments is visible. Both the Pierre Auger Observatory results (this work and the preliminary results presented in the ICRC13) are in good agreement. Only a known small difference of few g/cm^2 is observable: in case of the preliminary results, the corrections due to b_{LW} and ΔX_{align} were not applied. TA and HiRes, shifted as already discussed, agree each other, and they agree with this work inside the error bars. A proper attempt to compare the Pierre Auger Observatory and TA results is an ongoing work with the project of a joint working group [48]. While HiRes and TA can be compared for energy above $10^{18.2}$ eV, HiRes/MIA has published results below that energy. Its measurements have very few data point, however they agree with this work, except for the $\langle X_{\text{max}} \rangle$ value at $10^{17.4}$ eV. The Yakutsk results, where the X_{max} measurements are obtained indirectly via non-imaging Cherenkov detection, agree only for energies above $10^{18.5}$ eV. Below $10^{17.5}$ eV, its measurements are scattered: some of them agree with this work, some with Tunka. The results from Tunka (also a non-imaging Cherenkov detector) do not agree with this work.

Looking at Fig. 7.4(b), thanks to the measurements from this work, the visible lack of data shown in Fig. 2.10 between $3 \cdot 10^{17}$ eV and 10^{18} eV is covered. A general trend can be observed. From an intermediate composition on average (probably very mixed) at 10^{15} eV, the flux of the Cosmic Rays (CRs) is more and more dominated by heavy elements. Above the iron knee ($\approx 10^{16.9}$ eV), it becomes gradually lighter to then again heavier above $\approx 10^{18.3}$ eV.

The fluctuations shown in the $\widehat{\sigma}$ measurement (Fig. 7.2(b)) are less directly comparable with the models, which are for pure compositions. Proton-induced showers have a larger fluctuation than iron showers for superposition arguments. When a mixture is involved, the difference of the $\langle X_{\text{max}} \rangle$ further enlarge the $\widehat{\sigma}$. Thus, the values in-between the iron and proton lines can be interpreted as a mixture of nuclei with different atomic masses dominated by heavy elements, or a mixture of nuclei with similar atomic masses and the X_{max} fluctuations in-between the expected for proton and iron primary.

A more detailed interpretation in terms of mass composition is discussed in the next section.

7.3 Interpretation in terms of mass composition

As already introduced in Section 2.5, the X_{max} moments can be interpreted as moments of $\ln A$, where A is the atomic mass number of the primary, according Eqs. (2.20) and (2.22):

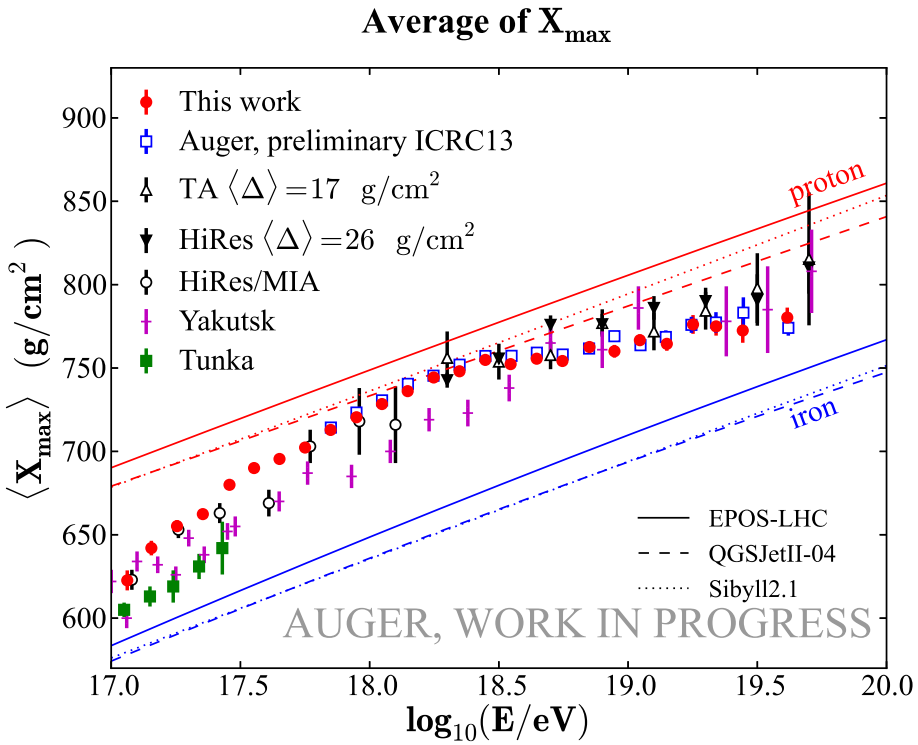
$$\langle X_{\text{max}} \rangle = \langle X^p_{\text{max}} \rangle - f_E \langle \ln A \rangle$$
$$\widehat{\sigma}^2(X_{\text{max}}) = \langle \widehat{\sigma}^2_A \rangle + f_E^2 \, \sigma^2(\ln A).$$

The $\langle X^p_{\text{max}} \rangle$, $\langle \widehat{\sigma}^2_A \rangle$ and f_E parameters are all functions of $\log_{10}(E/\text{eV})$ [46]. $\langle \widehat{\sigma}^2_A \rangle$ is the average of the X_{max} fluctuation over all the masses in a mixture with the relative fractions used as weights. It can be written as $\langle \widehat{\sigma}^2_A \rangle = \langle \widehat{\sigma}^2_{\text{sh}} \rangle + \widehat{\sigma}^2_p \, b \, \sigma^2(\ln A)$, where $\langle \widehat{\sigma}^2_{\text{sh}} \rangle$ is a quadratic function of $\langle \ln A \rangle$.

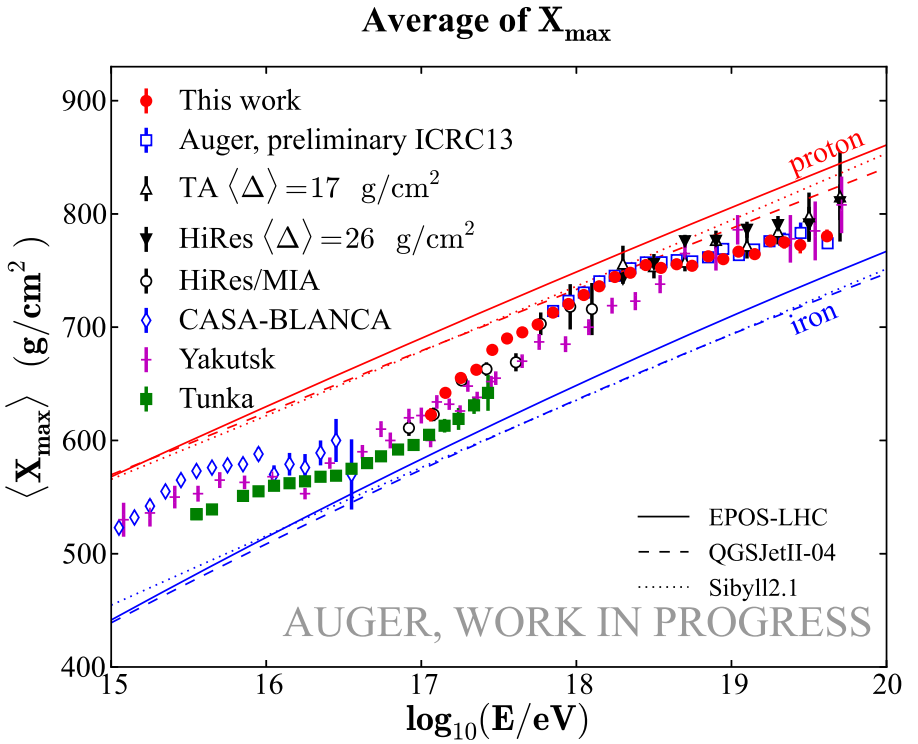
The conversion from the moments of the X_{max} distribution to the moments of $\ln A$ distribution is performed through

$$\langle \ln A \rangle = \frac{\langle X^p_{\text{max}} \rangle - \langle X_{\text{max}} \rangle}{f_E}, \tag{7.14}$$

$$\sigma^2_{\ln A} = \frac{\widehat{\sigma}^2(X_{\text{max}}) - \langle \widehat{\sigma}^2_{\text{sh}} \rangle}{f_E^2 + \widehat{\sigma}^2_p \, b}. \tag{7.15}$$



(a) Comparisons in the energy range of this work.



(b) Update of the collection shown in Fig. 2.10.

Figure 7.4: Comparison of $\langle X_{\max} \rangle$ measurements from several experiment (data from 2012, taken from [23]) and this work. The values from the hadronic interaction models (EPOS-LHC, QGSJetII-04, Sibyll2.1) are shown for pure proton and pure iron compositions. See text for details.

The derivation of this formula, all parameters and the error propagations are discussed in detail in Appendix G.

In Figs. 7.5 to 7.7, the interpretations of $X_{\max}$ moments (Fig. 7.2) in terms of $\ln A$ moments are shown for the hadronic interaction models EPOS-LHC, Sibyll2.1 and QGSJetII-04 respectively. In these figures the grey areas are the regions with unphysical values. The values in case of pure proton (p) or pure iron (Fe) composition, a mixture of 50 % p and 50 % Fe and a mixture of 85 % p and 15 % He are reported as well. The latter is approximately the maximum mixture that is allowed for the dip scenario, mentioned in Section 2.3. All pure compositions have $\sigma_{\ln A} = 0$ (indicated as "100 % A" in Figs. 7.5(b), 7.6(b) and 7.7(b)). The 50 % p and 50 % Fe mixture has the largest fluctuation considering only nuclei up to Fe, because it corresponds to the maximal possible difference in atomic masses in that range.

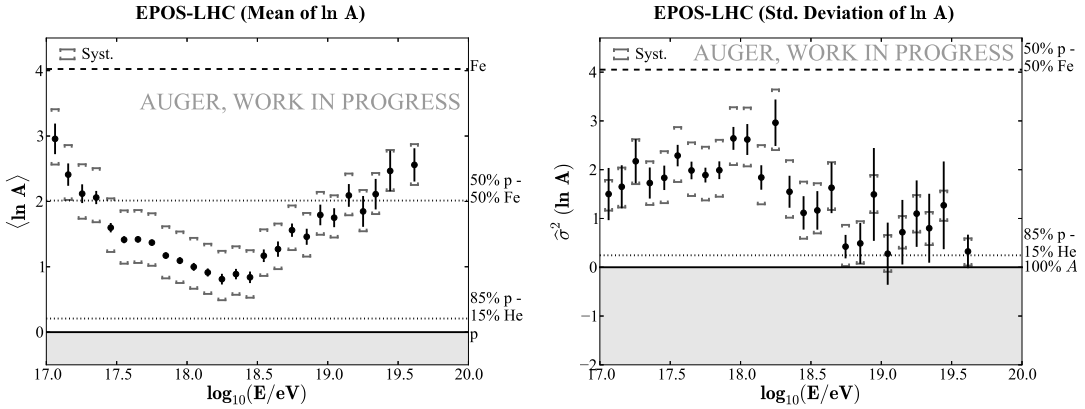
All the interpretations agree in the overall features. A heavy composition with a mixture of nuclei with very different atomic masses at 10^{17} eV becomes lighter until $\sim 10^{18.3}$ eV. Beyond $\sim 10^{18.5}$ eV (the visible structure point in Fig. 7.2(a)) the composition evolves towards a mixture with a heavier average and nuclei less different in atomic mass (i.e. purer). At $10^{17.5}$ eV a change of slope might confirm the second structure point observed in Fig. 7.2(a).

However the interpretation of the measurements using different models agrees only inside the systematic uncertainties. Using EPOS-LHC to interpret the data shows the heaviest scenario. With Sibyll2.1 as hadronic interaction model, the results show a very similar structure compared to EPOS-LHC, but shifted towards lighter mixture with nuclei less different in atomic mass. The results observed with QGSJetII-04 are shifted too, but in addition they appear to be tilted as well: below $\sim 10^{18.3}$ eV their $\langle \ln A \rangle$ is slightly heavier than Sibyll2.1 and above slightly lighter. While in the other two models below $\sim 10^{18.3}$ eV the $\sigma^2(\ln A)$ increases with the energy, in QGSJetII-04 it is almost constant. Moreover, above this energy, the results interpreted by the interaction model QGSJetII-04 are physical only inside $2\,\sigma$ of uncertainties.

A complete interpretation of the measurement in term of mass composition is beyond the scope of this work. However a mixed composition scenario (Section 2.3) seems to be favored. The galactic composition seems to be dominated by heavy elements and extended beyond the iron knee. The extragalactic composition is increasingly heavier and purer with the energy. Finally a possible transition between galactic and extragalactic sources might be occuring between 10^{18} eV and $10^{18.5}$ eV.

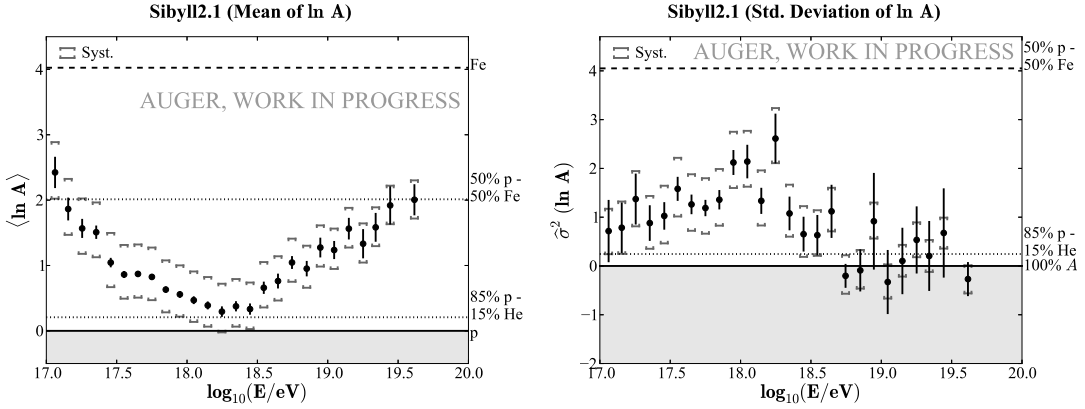
The measurements from [15] are of particular interest and may confirm the galactic composition behavior. In this article large-scale anisotropy searches are performed by the Pierre Auger Collaboration for energies above 10^{18} eV. According to the results shown in the paper, the fraction of protons should not exceed ≈ 10 % for sources densely distributed in the galactic disk. Hence the light component observed in this results is most likely from extragalactic sources.

Attempts to interpret the extragalactic component with the Pierre Auger Collaboration measurement above 10^{18} eV are performed in [14, 50, 51, 112]. Looking forward, better interpretation for a more complete scenario can be done with the measurement extension at lower energies provided by this work.



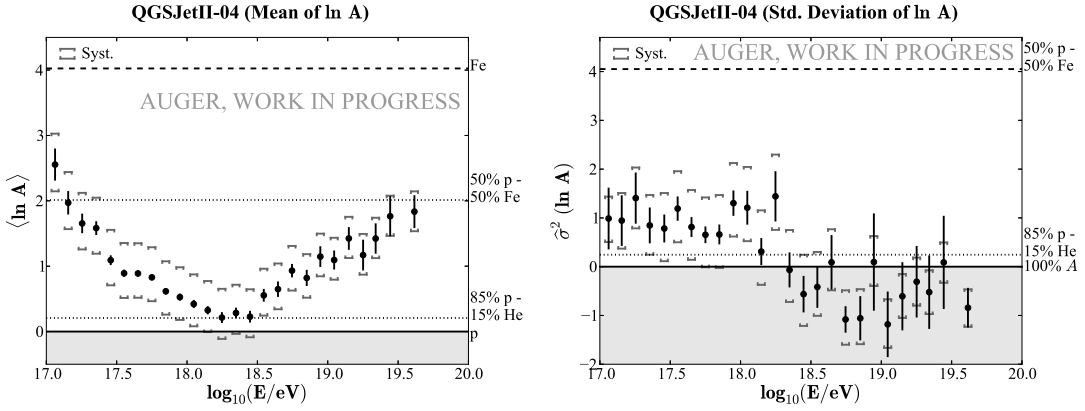
(a) $\langle \ln A \rangle$ (b) $\sigma^2(\ln A)$

Figure 7.5: $\ln A$ moments with EPOS-LHC as hadronic interaction model. Values of pure and mixed compositions are also shown.



(a) $\langle \ln A \rangle$ (b) $\sigma^2(\ln A)$

Figure 7.6: $\ln A$ moments with Sibyll2.1 as hadronic interaction model. Values of pure and mixed compositions are also shown.



(a) $\langle \ln A \rangle$ (b) $\sigma^2(\ln A)$

Figure 7.7: $\ln A$ moments with QGSJetII-04 as hadronic interaction model. Values of pure and mixed compositions are also shown.

CHAPTER 8

Summary and outlook

The aim of this work was the measurement of the depth of the shower maximum in the energy region of the expected transition from galactic to extragalactic cosmic rays. Knowing the average depth shower maximum ($\langle X_{\max} \rangle$) and the corresponding shower-to-shower fluctuation ($\widehat{\sigma}(X_{\max})$) allows to estimate the mass composition of cosmic rays in this energy region. The mass composition is one of the most important constraints on models of sources and propagation of high-energy and ultra-high-energy cosmic rays.

In this work, data from the Pierre Auger Observatory were analyzed. Using the new enhancement HEAT, the energy region that covers the different model predictions for the transition between galactic and extragalactic cosmic rays sources was possible to be explored. To measure the moments of the $X_{\max}$ distribution ($\langle X_{\max} \rangle$ and $\widehat{\sigma}(X_{\max})$), data and simulations were analyzed.

- A reliably reconstructed data set was obtained:
  - A period of unstable data acquisition of HEAT that may bias the measurements was found and discarded.
  - A test on the calibration was performed using the measurements from pair of telescopes that detected the same shower. An incompatibility of $5.3 \pm 0.6\,\%$ on the energy between one telescope building and the others was found and the calibration was corrected accordingly.
- A high quality data set was selected, through application of cuts, to ensure a good $X_{\max}$ and energy resolution to resolve the true shower maximum distribution
- A data set corrected for detection and reconstruction biases was developed:
  - The limited field of view influences the event acceptance. An analysis was performed with the so-called "fiducial field of view" method to unbias the data sample. This also set the lower energy limit at $10^{17}\,\mathrm{eV}$ to have a data sample trustworthy enough to be not biased.

  - As explained in the thesis, the fiducial volume cuts can only approximately remove the influence of the limited field of view of the telescopes. Thus the tails of the X_{max} distribution may still be biased. Corrections were estimated through simulations. The magnitude of these corrections to $\langle X_{\mathrm{max}} \rangle$ and $\widehat{\sigma}(X_{\mathrm{max}})$ are on average less than $1\,\mathrm{g/cm^2}$ and $2.5\,\mathrm{g/cm^2}$ respectively, and no more than $3\,\mathrm{g/cm^2}$ and $7\,\mathrm{g/cm^2}$, while the $\langle X_{\mathrm{max}} \rangle$ and $\widehat{\sigma}(X_{\mathrm{max}})$ measurements are around $600 \div 800\,\mathrm{g/cm^2}$ and $30 \div 70\,\mathrm{g/cm^2}$.
  - Shortcomings in the reconstruction can produce biases. Their magnitude on the energy and X_{max} reconstruction were estimated by comparing the known true observable with its reconstructed one in simulations. The correction applied on the data at $10^{18}\,\mathrm{eV}$ are $\approx 3.5\,\%$ and $\approx 3.7\,\mathrm{g/cm^2}$ respectively.

- The X_{max} resolution and systematic uncertainties were estimated through simulations, error propagation and the study of the parameters considered a physical source of systematics. At $10^{18}\,\mathrm{eV}$ the energy and X_{max} systematic uncertainties are $\approx {}^{+15.3}_{-14.6}\,\%$ and $\approx {}^{+8.5}_{-9.2}\,\mathrm{g/cm^2}$ respectively. The X_{max} resolution at this energy is $\approx 20\,\mathrm{g/cm^2}$.

In this thesis almost three order of energy magnitudes, from $10^{17}\,\mathrm{eV}$ to $\leq 10^{20}\,\mathrm{eV}$ were explored: an unprecedented energy span of measurements for experiments with fluorescence detectors. Moreover, the energy region $0.3 \div 1 \cdot 10^{18}\,\mathrm{eV}$, where the $\langle X_{\mathrm{max}} \rangle$ data from various experiments were rather sparse, is now well covered by this results.

From the $\langle X_{\mathrm{max}} \rangle$ measurement as a function of the energy, three different elongation rates,

$$\begin{aligned} D_1 &= 131.89 \pm 11.36|_{\mathrm{stat}}\, {}^{+0.31}_{-1.95}|_{\mathrm{syst}}, \\ D_2 &= 80.48 \pm 2.33|_{\mathrm{stat}}\, {}^{+1.72}_{-1.08}|_{\mathrm{syst}}, \\ D_3 &= 23.87 \pm 2.71|_{\mathrm{stat}}\, {}^{+1.03}_{-2.66}|_{\mathrm{syst}}, \end{aligned}$$

and two break points,

$$\begin{aligned} \log_{10}(E_1/\mathrm{eV}) &= 17.49 \pm 0.04|_{\mathrm{stat}} \pm 0.07|_{\mathrm{syst}}, \\ \log_{10}(E_2/\mathrm{eV}) &= 18.27 \pm 0.03|_{\mathrm{stat}} \pm 0.07|_{\mathrm{syst}}, \end{aligned}$$

were found.

A change in the elongation rate indicates a change in the evolution of the mixture of nuclei. The break point at $10^{18.27}\,\mathrm{eV}$ and the last two elongation rates are in agreement with other experiments and the previous results from the Pierre Auger Observatory. The break point at $10^{17.49}\,\mathrm{eV}$ and the first elongation rate are new features observed for the first time in this work. Independent studies of this break are needed to establish it beyond any doubt.

The X_{max} moments can be interpreted as $\ln A$ moments, where A is the atomic mass of the primaries. The first moment is the average composition of the flux. The second moment represents how different the masses involved in the mixture are from each other. For example in case of pure elements $\sigma^2(\ln A) = 0$; in case of $50\,\%$ protons and $50\,\%$ irons $\sigma^2(\ln A) \approx 4$.

The measurements show heavy composition at $10^{17}\,\mathrm{eV}$ with a mixture composed of nuclei with very different atomic masses that becomes lighter until $\sim 10^{18.3}\,\mathrm{eV}$. Beyond $\sim 10^{18.5}\,\mathrm{eV}$ the composition evolves towards a heavier nuclear mass and a mixture with nuclei with atomic masses less different from each other (i.e. purer). From these results, the ankle and the dip scenarios of the transition from galactic to extragalactic cosmic rays are disfavored because they require a proton dominance in the extragalactic flux. A mixed composition scenario seems to be favored due to the heavier composition at the lowest and highest energies.

The present results can be improved by:

- a dedicated study of the systematic uncertainty propagation on energy and significance of the break point at $10^{17.49}\,\mathrm{eV}$, to obtain a more significant measurement of the latter and the elongation rate below it;
- a proper simulation of the hypothesis of a four component mixture (H, He, CNO and Fe) to fit energy spectrum, $\langle X_{\mathrm{max}} \rangle$ and $\widehat{\sigma}(X_{\mathrm{max}})$ [113–116];
- further extension of the low energy limit through a dedicated study of events with a high Cherenkov light contamination, which are not used in this work.

These possible improvements are beyond the scope of this work and left for future analysis.

APPENDIX A

Fiducial FoV

As explained in Section 4.3.2, the boundaries of the fiducial field of view are calculated by simultaneously fitting the two data sets $\langle X_{\mathrm{max}} \rangle (X_{\mathrm{up}})$ and $\langle X_{\mathrm{max}} \rangle (X_{\mathrm{up}})$ with the μ_{trunk} function. Here the plot collection for this calculation is shown. $\langle X_{\mathrm{max}} \rangle$ as a function of X_{up} and X_{low} is represented by red squares and blue dots respectively. Black lines are the μ_{trunk} fits.

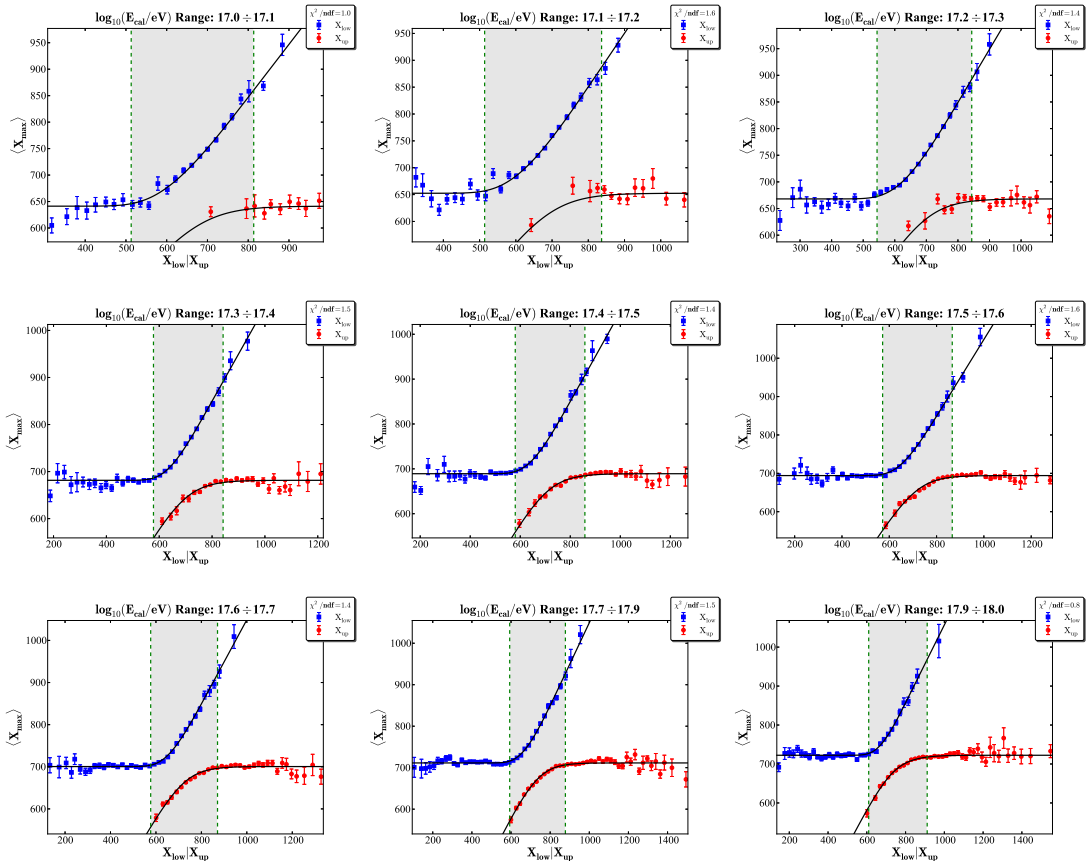


Figure A.1: Collection of the calculation of the boundaries of the fiducial field of view for different energy bins below $10^{18.0}\,\mathrm{eV}$. $\langle X_{\mathrm{max}} \rangle$ as a function of X_{up} and X_{low} is represented by red squares and blue dots respectively. Black lines are the simultaneous fit of both the data set with μ_{trunk} function.

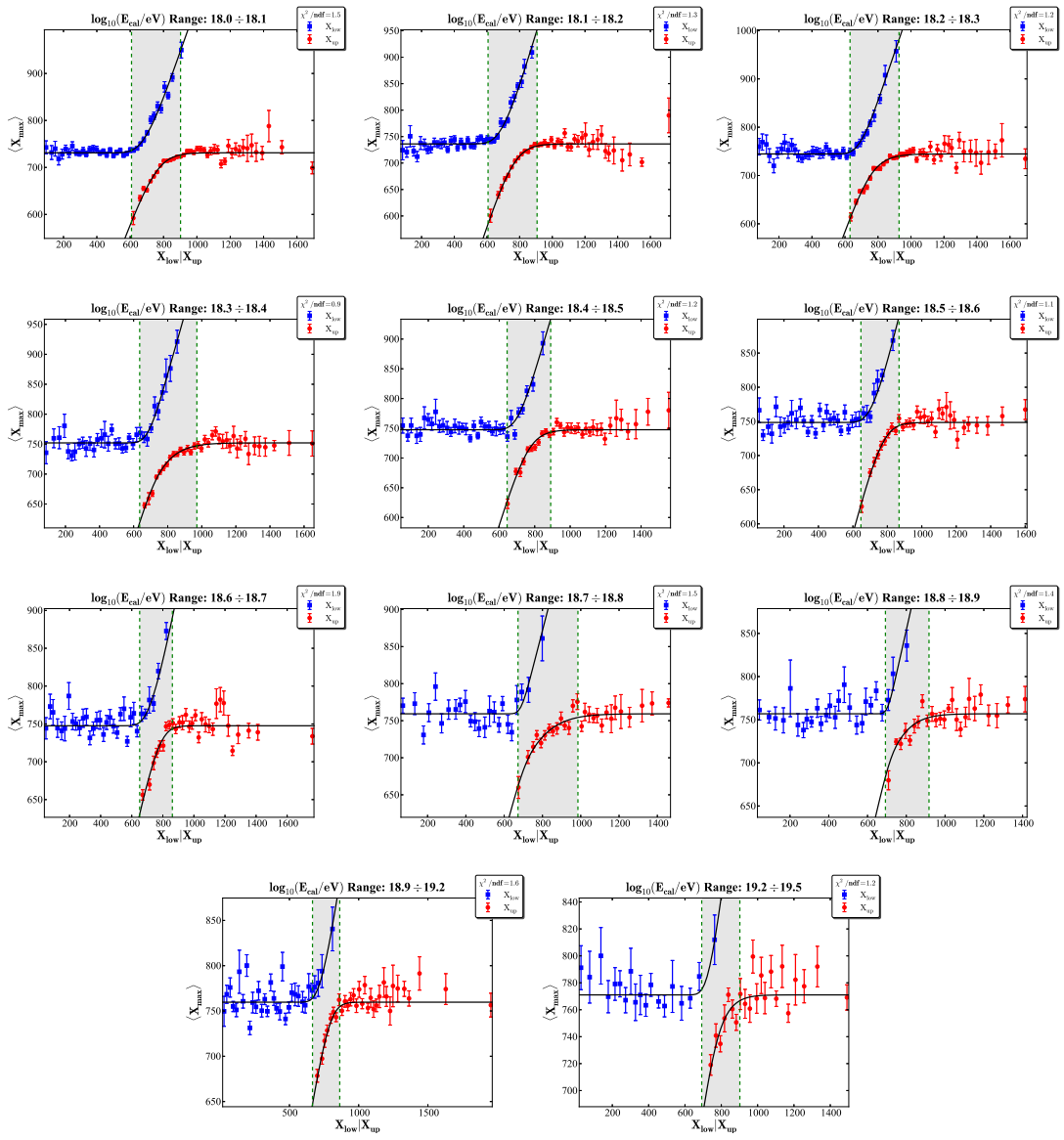


Figure A.2: Collection of the calculation of the boundaries of the fiducial field of view for different energy bins above $10^{18.0}\,\mathrm{eV}$. $\langle X_{\mathrm{max}} \rangle$ as a function of X_{up} and X_{low} is represented by red squares and blue dots respectively. Black lines are the simultaneous fit of both the data set with μ_{trunk} function.

APPENDIX B

$X_{\mathbf{max}}$ distributions

The collection of the $X_{\max}$ distributions for different energy bins is shown. The red lines represent the fit with the $G \otimes E$ function. The parametrization of this function is used to weight the $X_{\max}$ uniform distribution in the simulations to reproduce the true maximum shower distributions.

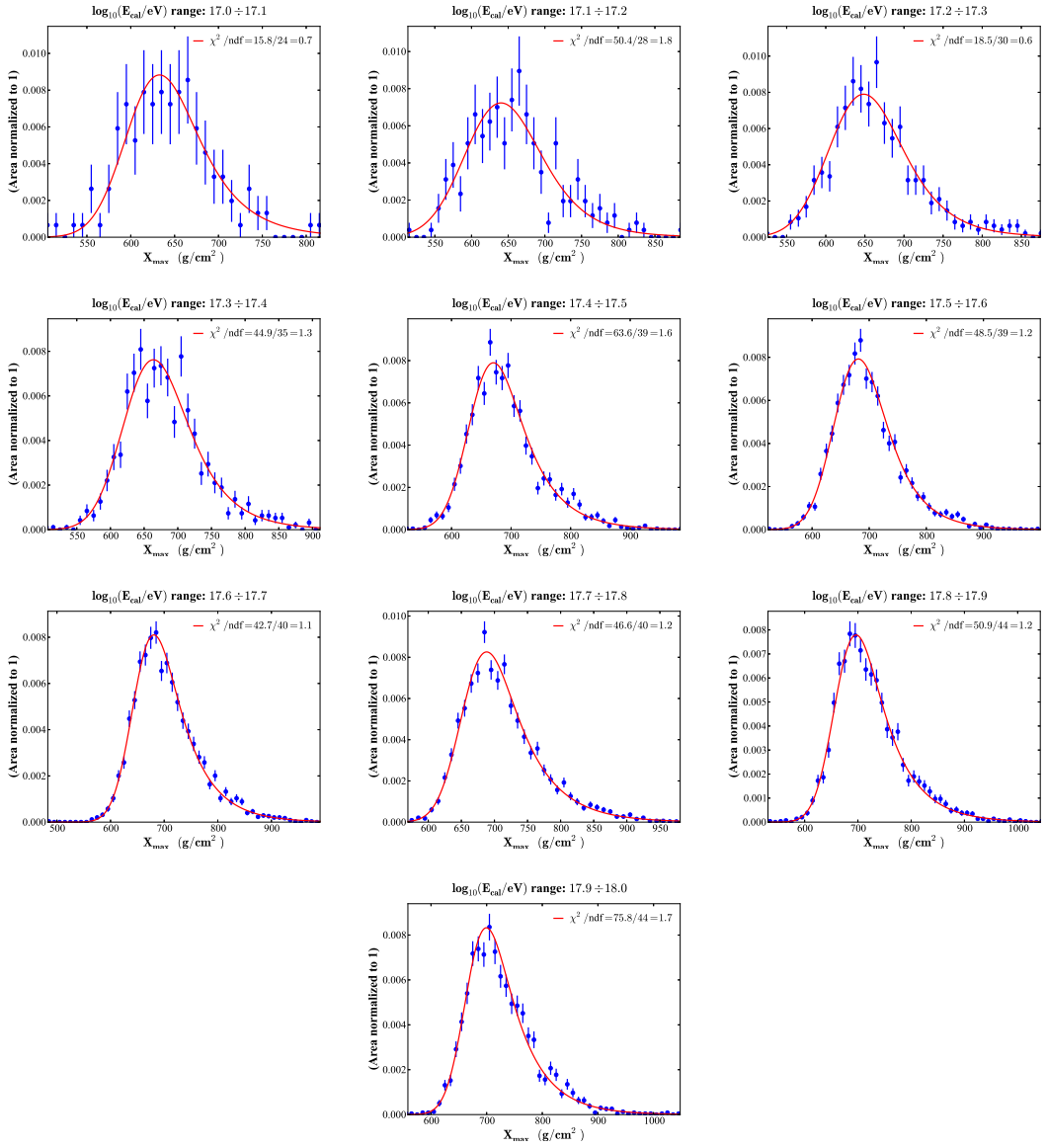


Figure B.1: $X_{\max}$ distributions in different energy bins below $10^{18.0}\,\mathrm{eV}$. The red lines represent the fit with the $G \otimes E$ function.

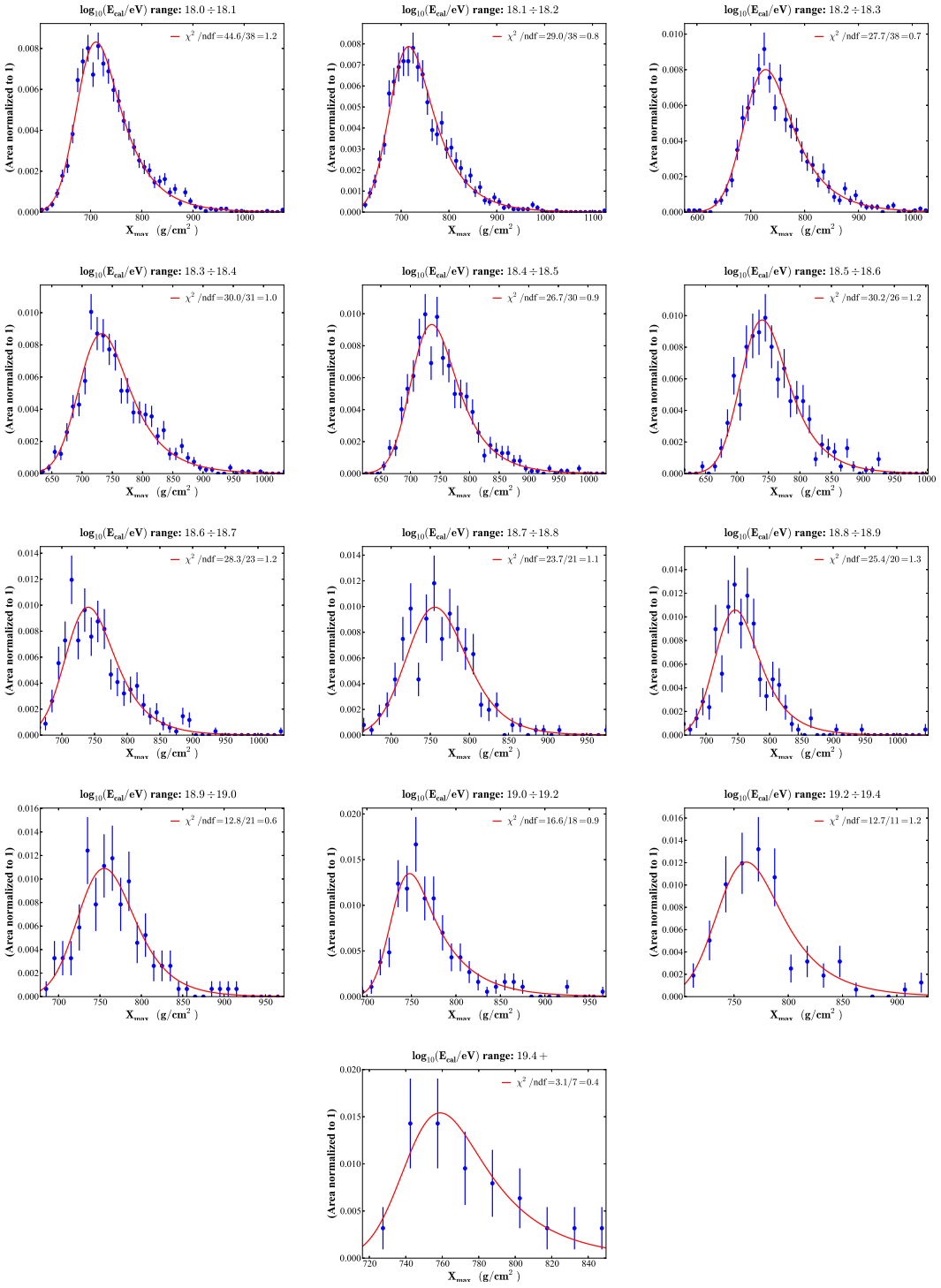


Figure B.2: X_{max} distributions in different energy bins above $10^{18.0}\,\mathrm{eV}$. The red lines represent the fit with the $G \otimes E$ function.

APPENDIX C

$X_{\mathbf{max}}$ Acceptance

In Section 5.1 the acceptance correction was explained. Here the collection of all the efficiencies is shown. The efficiencies are fitted using Eq. (5.1) and normalized to this function. λ_1, λ_2, X_1 and X_2 parameterized through this procedure were shown in Fig. 5.2.

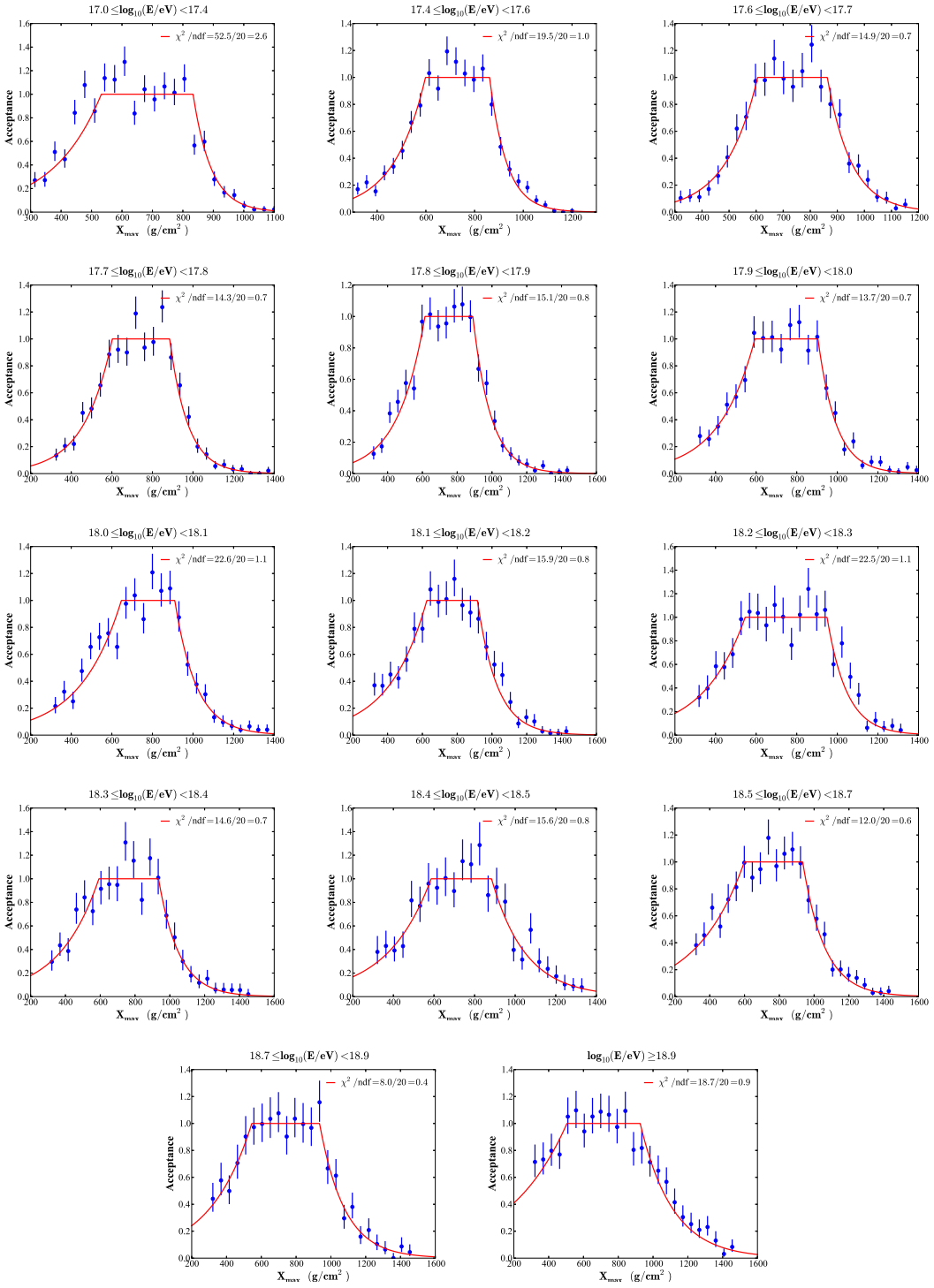


Figure C.1: Distributions of the acceptance and their fits (red lines) for different energy bins. y axes are normalized to the the relative fits with the Eq. (5.1)

APPENDIX D

R_p distributions data/MC

In Section 5.3.2 the R_p test was discussed. Here the collection of the full comparison between data (blue dots) and MC (red histograms and relative error bars) is shown. The y axes are normalized to unit area. Agreement inside error bars is found.

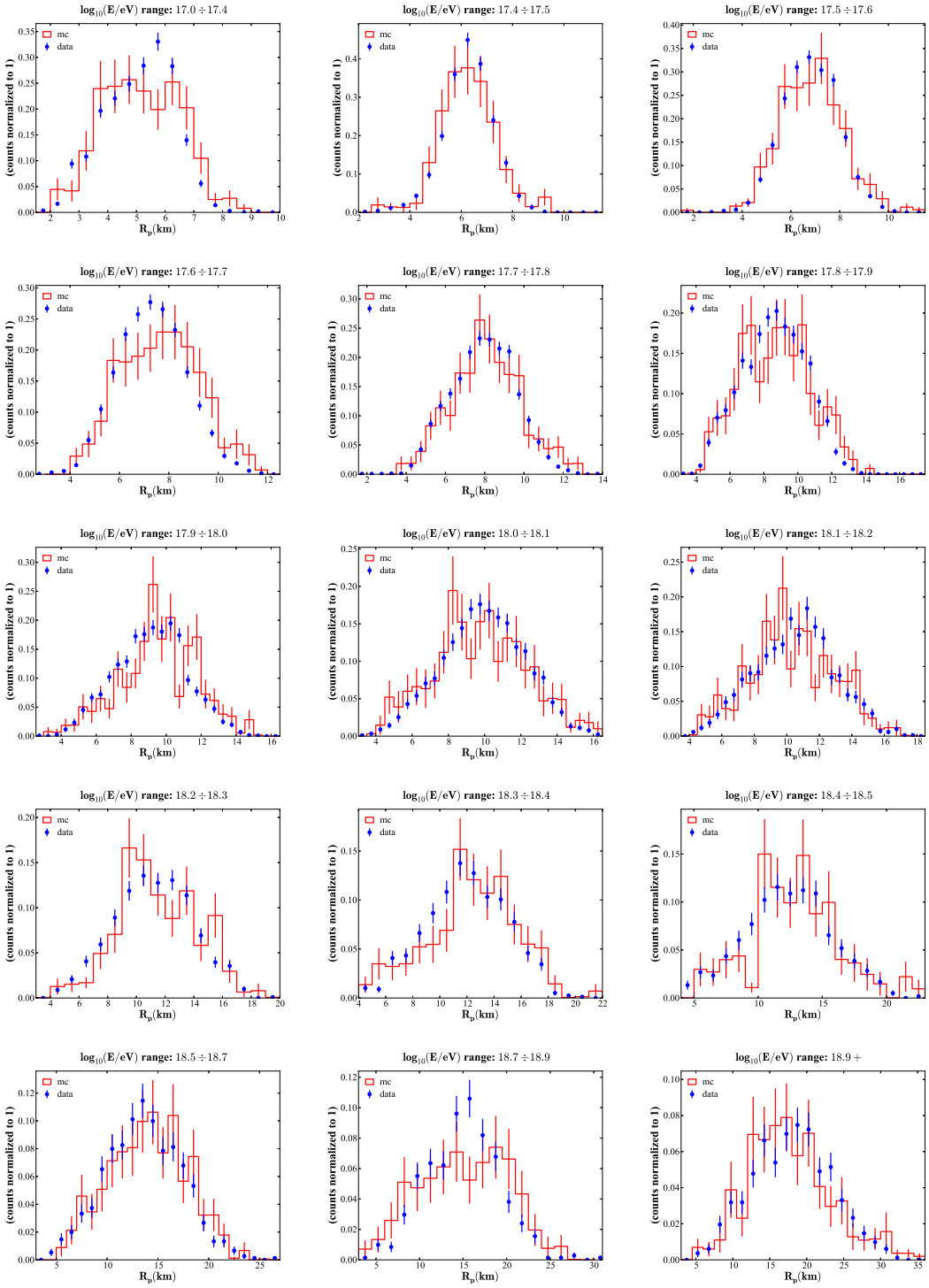


Figure D.1: Plot collection of the R_p test. Blue dots are data, red histograms and relative error bars are simulation. y axes are normalized.

APPENDIX E

Calculation of μ_{trunk}

As already explained in Section 4.3.2, μ_{trunc} is the truncated mean value of the $X_{\max}$ distribution. This appendix is going to show details of its calculation.

First of all, an heuristic function to describe the distribution should be defined. A very satisfying approximation is given by a Gaussian ($G(X)$) convoluted with an exponential ($E(X)$):

$$f(X) = G \otimes E\,(X) = \frac{1}{2\lambda} \exp\left(\frac{X_0 - X}{\lambda} + \frac{\sigma^2}{2\lambda^2}\right) \operatorname{erfc}\left(\frac{X_0 - X}{\sqrt{2}\sigma} + \frac{\sigma}{\sqrt{2}\lambda}\right), \tag{E.1}$$

where λ, X_0, σ and the complementary error function erfc are defined as

$$E(X) = \quad E(X, \lambda) = \frac{1}{\lambda} \mathrm{e}^{-X/\lambda}, \tag{E.2}$$

$$G(X) = N(X, X_0, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-(X-X_0)^2/2\sigma^2}, \tag{E.3}$$

$$\operatorname{erfc}(x) = \quad 1 - \operatorname{erf}(x) = 1 - \frac{2}{\sqrt{\pi}} \int_0^x \mathrm{e}^{-t^2}\,\mathrm{d}t = \frac{2}{\sqrt{\pi}} \int_x^\infty \mathrm{e}^{-t^2}\,\mathrm{d}t. \tag{E.4}$$

f can be interpreted as the Gaussian distribution of the shower development after the first interaction point defined by the exponential.

Using f, the truncated mean between the limits of the integral x_1 and x_2 is

$$\mu_{\text{trunc}}(x_1, x_2) = \frac{\displaystyle\int_{x_1}^{x_2} x f(x)\,\mathrm{d}x}{\displaystyle\int_{x_1}^{x_2} f(x)\,\mathrm{d}x} = \frac{\displaystyle\int_{x_1}^{x_2} x\, G \otimes E\,\mathrm{d}x}{\displaystyle\int_{x_1}^{x_2} G \otimes E\,\mathrm{d}x}. \tag{E.5}$$

To calculate the integrals, the derivate and the integral of the functions $\operatorname{erf}(x)$ and $\operatorname{erfc}(x)$ must are needed:

$$\begin{aligned} \frac{\mathrm{d}\,\mathrm{erf}(x)}{\mathrm{d}x} &= \frac{2}{\sqrt{\pi}}\mathrm{e}^{-x^2} & \int \mathrm{erf}(x)\,\mathrm{d}x &= x\,\mathrm{erf}(x) + \frac{1}{\sqrt{\pi}}\mathrm{e}^{-x^2} \\ \frac{\mathrm{d}\,\mathrm{erfc}(x)}{\mathrm{d}x} &= -\frac{2}{\sqrt{\pi}}\mathrm{e}^{-x^2} & \int \mathrm{erfc}(x)\,\mathrm{d}x &= x - x\,\mathrm{erf}(x) - \frac{1}{\sqrt{\pi}}\mathrm{e}^{-x^2}. \end{aligned} \tag{E.6}$$

Let

$$\begin{aligned} a &= X_0 + \sigma^2/\lambda, \\ b &= \sqrt{2}\sigma, \\ c &= \frac{1}{2\lambda}\exp\left(\frac{X_0}{\lambda} + \frac{\sigma^2}{2\lambda^2}\right) = \frac{1}{2\lambda}\exp\left(\frac{a}{\lambda} - \frac{b^2}{4\lambda^2}\right), \end{aligned}$$

then the Eq. (E.1) can be simply rewritten as $f(x) = G \otimes E\,(x) = c\exp\left(-x/\lambda\right)\mathrm{erfc}\left(\frac{a-x}{b}\right)$. Its integral is

$$\begin{aligned} \mathrm{I}(x) = \int f(x)\,\mathrm{d}x = \int G \otimes E\,(x)\,\mathrm{d}x &= \int c\,\exp\left(-\frac{x}{\lambda}\right)\mathrm{erfc}\left(\frac{a-x}{b}\right)\,\mathrm{d}x & = \\ = -c\lambda\,\exp\left(-\frac{x}{\lambda}\right)\mathrm{erfc}\left(\frac{a-x}{b}\right) + c\lambda\int\exp\left(-\frac{x}{\lambda}\right)\frac{-2}{\sqrt{\pi}}\exp\left[-\frac{(a-x)^2}{b^2}\right]\frac{-1}{b}\,\mathrm{d}x & = \\ = -\lambda\left\{G\otimes E - \frac{2c}{\sqrt{\pi}b}\int\exp\left[-\frac{x^2 - 2x\left(a - \frac{b^2}{2\lambda}\right) + a^2}{b^2}\right]\,\mathrm{d}x\right\} & = \\ = -\lambda\left\{f(x) - \frac{c}{b}\exp\left(\frac{b^2}{4\lambda^2} - \frac{a}{\lambda}\right)\int\frac{2}{\sqrt{\pi}}\exp\left[-\left(a - \frac{b^2}{2\lambda} - x\right)^2\frac{1}{b^2}\right]\,\mathrm{d}x\right\} & = \\ = -\lambda\left\{f(x) + \frac{c}{b}\frac{1}{2\lambda c}\,\mathrm{erf}\left(\frac{a}{b} - \frac{b}{2\lambda} - \frac{x}{b}\right)b\right\} = -\lambda f(x) - \frac{1}{2}\,\mathrm{erf}\left(\frac{X_0 - x}{\sqrt{2}\sigma}\right), & \end{aligned} \tag{E.7}$$

and the integral of $xf(x) = x\,G \otimes E\,(x)$ is

$$\begin{aligned} \mathrm{I}_x(x) = \int x\,G\otimes E\,(x)\,\mathrm{d}x = \int x f(x)\,\mathrm{d}x & = \\ = x\int f\,\mathrm{d}x - \iint f\,\mathrm{d}x^2 & = \\ = x\,\mathrm{I}(x) - \int \mathrm{I}(x)\,\mathrm{d}x & = \\ = x\,\mathrm{I}(x) + \lambda\int f(x)\,\mathrm{d}x + \frac{1}{2}\int\mathrm{erf}\left(\frac{X_0 - x}{\sqrt{2}\sigma}\right)\,\mathrm{d}x & = \\ = (x+\lambda)\,\mathrm{I}(x) - \frac{\sqrt{2}\sigma}{2}\left[\left(\frac{X_0 - x}{\sqrt{2}\sigma}\right)\mathrm{erf}\left(\frac{X_0 - x}{\sqrt{2}\sigma}\right) - \frac{1}{\sqrt{\pi}}\mathrm{e}^{-\left(\frac{X_0 - x}{\sqrt{2}\sigma}\right)^2}\right] & = \\ = (x+\lambda)\,\mathrm{I}(x) - \frac{\sigma}{\sqrt{2}}\left(\frac{X_0 - x}{\sqrt{2}\sigma}\right)\mathrm{erf}\left(\frac{X_0 - x}{\sqrt{2}\sigma}\right) - \sigma^2 G(x). & \end{aligned} \tag{E.8}$$

Hence Eq. (E.5) can be rewritten as

$$\mu_{\text{trunc}}(x_1, x_2) = \frac{\mathrm{I}_x(x_2) - \mathrm{I}_x(x_1)}{\mathrm{I}(x_2) - \mathrm{I}(x_1)}. \tag{E.9}$$

APPENDIX F

Error propagation for the $X_{\max}$ moments

The moments of the $X_{\max}$ distribution, $\langle X_{\max} \rangle$ and $\widehat{\sigma}$, were calculated in Section 7.1.2. The estimation of their statistical and systematic uncertainties are discussed in the Appendix.

F.1 Error propagation

In case of unweighted moments, the statistical uncertainty of $\langle X_{\max} \rangle$ is given by

$$\varepsilon_{\langle \mathrm{X_{max}} \rangle} = \frac{\widehat{\sigma}'}{\sqrt{N}}, \tag{F.1}$$

where N is the number of events. The variance of $\widehat{\sigma}'^2$, which is a variance as well, is given by the formula

$$\mathrm{Var}\left(\widehat{\sigma}'^2\right) = \frac{1}{N}\left[m_4 - \frac{(N-3)(N-1)}{N^2} \cdot \widehat{\sigma}'^4\right], \tag{F.2}$$

where m_4 is the fourth central moment (see appendix F.2 for details). In case of weighted moments with weights w, the follow substitutions must be applied:

$$N = \frac{\left(\sum w\right)^2}{\sum w^2}, \tag{F.3}$$

$$m_4 = \frac{\sum w \, \left(X_{\max}^{\mathrm{corr}} - \langle X_{\max} \rangle\right)^4}{\sum w}. \tag{F.4}$$

However the observable of interest is $\widehat{\sigma}'$, thus the error propagation yields

$$\mathrm{Var}\left(\widehat{\sigma}'\right) = \mathrm{Var}\left(\sqrt{\widehat{\sigma}'^2}\right) = \frac{\mathrm{Var}\left(\widehat{\sigma}'^2\right)}{4 \cdot \widehat{\sigma}'^2}. \tag{F.5}$$

Finally, to obtain $\widehat{\sigma}$ from $\widehat{\sigma}'$, the resolution is subtracted and the uncertainty of the latter must be propagated. For a general function $c(a, b)$ with:

$$c = \sqrt{a^2 - b} \quad \Rightarrow \quad \varepsilon_c = \sqrt{\frac{4 \cdot a^2 \cdot \mathrm{Var}\left(a\right) + \mathrm{Var}\left(b\right)}{4 \cdot \left(a^2 - b\right)}} = \frac{\sqrt{a^2 \cdot \mathrm{Var}\left(a\right) + \mathrm{Var}\left(b\right)/4}}{c}, \tag{F.6}$$

the statistical uncertainty of $\widehat{\sigma}$ will be

$$\varepsilon_{\widehat{\sigma}} = \frac{\sqrt{\widehat{\sigma}'^2 \cdot \mathrm{Var}\left(\widehat{\sigma}'\right) + \mathrm{Var}\left(\widehat{\mathrm{Res}}^2\right)/4}}{\widehat{\sigma}}. \tag{F.7}$$

The systematic uncertainty of $\langle X_{\max} \rangle$ is estimated by summing of all the contribution δX from Chapter 6 in quadrature:

$$\delta\langle X_{\max} \rangle = \sqrt{\sum_i \delta X_i^2}. \tag{F.8}$$

The resolutions Res_i in Eq. (7.7) are summed in quadrature. Thus the error propagation of the contributions follows the formula

$$c = \sqrt{a^2 + b^2} \quad \Rightarrow \quad \delta c = \sqrt{\frac{4 \cdot a^2 \cdot \left(\delta a\right)^2 + 4 \cdot b^2 \cdot \left(\delta b\right)^2}{4 \cdot \left(a^2 + b^2\right)}} = \sqrt{\frac{a^2 \cdot \left(\delta a\right)^2 + b^2 \cdot \left(\delta b\right)^2}{a^2 + b^2}}. \tag{F.9}$$

Therefore the systematic uncertainty of the resolution is

$$\delta\,\mathrm{Res} = \sqrt{\frac{\sum_i \mathrm{Res}_i^2 \cdot \left(\delta\,\mathrm{Res}_i\right)^2}{\sum_i \mathrm{Res}_i^2}}, \tag{F.10}$$

where Res_i considered in the sum are only the resolution corrections with an uncertainty: $\mathrm{Res}_{\mathrm{VAOD,unif}}$, $\mathrm{Res}_{\mathrm{VAOD,syst}}$, $\mathrm{Res}_{\mathrm{align}}$ and $\mathrm{Res}_{\mathrm{det}}$. The last one is weighted by the acceptance propagated uncertainties. This propagation defines upper and lower weights w_{up} and w_{low}, which are used to average the upper and lower uncertainties of $\mathrm{Res}_{\mathrm{det}}$. Finally, $\mathrm{Res}_{\mathrm{calAdj}}$ has not a resolution to correct, only an uncertainty, thus Eq. (F.10) is not valid. However it can be simply summed in quadrature. Therefore, the systematic uncertainty of $\widehat{\sigma}$ is

$$\delta\widehat{\sigma} = \sqrt{\delta\,\mathrm{Res}^2 + \delta\,\mathrm{Res}_{\mathrm{calAdj}}^2}. \tag{F.11}$$

F.2 Variance of a variance

In Eq. (F.2) the variance of a variance was used. The formula reported will be proven in the following.

Given a sample X of N elements, the following notation will be used:

$$\mu = \mathrm{E}(X) \qquad \text{(expectation value of } X\text{)};$$

$$m_k = \sum_i^N \frac{(X_i - \mu)^k}{N} \qquad (k\text{-th central moment});$$

$$\langle X \rangle = \sum_i^N \frac{X_i}{N} \qquad \text{(average)};$$

$$\widehat{s}^{\,2} = \frac{N}{N-1} \sum_i^N \frac{(X_i - \langle X \rangle)^2}{N} = \frac{N}{N-1} m_2 \qquad \text{(unbiased estimator of } m_2\text{)}.$$

The variance of $\widehat{s}^{\,2}$ is defined as

$$\mathrm{Var}(\widehat{s}^{\,2}) = E(\widehat{s}^{\,4}) - m_2^2. \tag{F.12}$$

Given $\widehat{s}^{\,2}$, only the term $E(\widehat{s}^{\,4})$ must be calculated. Defining the sample $Z = X - \mu$, for which the expectation value is $\mathrm{E}(Z) = 0$, $\widehat{s}^{\,2}$ can be written as

$$\widehat{s}^{\,2} = \frac{N}{N-1} \sum_i^N \frac{(Z_i - \langle Z \rangle)^2}{N} = \frac{N \sum_i^N Z_i^2 - \left(\sum_i^N Z_i\right)^2}{N(N-1)}, \tag{F.13}$$

and squaring

$$\widehat{s}^{\,4} = \frac{N^2 \left(\sum_i^N Z_i^2\right)^2 - 2N \sum_i^N Z_i^2 \left(\sum_j^N Z_j\right)^2 + \left(\sum_i^N Z_i\right)^4}{N^2 (N-1)^2}. \tag{F.14}$$

Considering that the variables Z are independent, the expected values of the terms in Eq. (F.14)

are:

$$\begin{aligned}
\mathrm{E}\left[\left(\sum_i^N Z_i^2\right)^2\right] &= \sum \mathrm{E}\left(Z_i^4\right) + \sum_{i\neq j} \mathrm{E}\left(Z_i^2 Z_j^2\right) = \\
&= N\,\mathrm{E}\left(Z_i^4\right) + N(N-1)\,\mathrm{E}\left(Z_i^2\right)\mathrm{E}\left(Z_j^2\right) = \\
&= Nm_4 + N(N-1)m_2^2, \qquad \text{(F.15)}
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}\left[\sum_i^N Z_i^2 \left(\sum_j^N Z_j\right)^2\right] &= \sum \mathrm{E}\left(Z_i^4\right) + \sum_{i\neq j} \mathrm{E}\left(Z_i^2 Z_j^2\right) + 2\sum_{i\neq j} \mathrm{E}\left(Z_i^3 Z_j\right) + \sum_{i\neq j\neq k} \mathrm{E}\left(Z_i^2 Z_j Z_k\right) = \\
&= N\,\mathrm{E}\left(Z_i^4\right) + N(N-1)\,\mathrm{E}\left(Z_i^2\right)\mathrm{E}\left(Z_j^2\right) + \\
&\quad + 2N(N-1)\,\mathrm{E}\left(Z_i^3\right)\cancelto{0}{\mathrm{E}(Z_j)} + \\
&\quad + N(N-1)(N-2)\,\mathrm{E}\left(Z_i^2\right)\cancelto{0}{\mathrm{E}(Z_j)}\cancelto{0}{\mathrm{E}(Z_k)} = \\
&= Nm_4 + N(N-1)m_2^2, \qquad \text{(F.16)}
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}\left[\left(\sum_i^N Z_i\right)^4\right] &= \sum \mathrm{E}\left(Z_i^4\right) + 3\sum_{i\neq j} \mathrm{E}\left(Z_i^2 Z_j^2\right) + 4\sum_{i\neq j} \mathrm{E}\left(Z_i^3 Z_j\right) + \\
&\quad + 6\sum_{i\neq j\neq k} \mathrm{E}\left(Z_i^2 Z_j Z_k\right) + \sum_{i\neq j\neq k\neq h} \mathrm{E}\left(Z_i Z_j Z_k Z_h\right) = \\
&= N\,\mathrm{E}\left(Z_i^4\right) + 3N(N-1)\,\mathrm{E}\left(Z_i^2\right)\mathrm{E}\left(Z_j^2\right) + \\
&\quad + 4N(N-1)\,\mathrm{E}\left(Z_i^3\right)\cancelto{0}{\mathrm{E}(Z_j)} + \\
&\quad + 6N(N-1)(N-2)\,\mathrm{E}\left(Z_i^2\right)\cancelto{0}{\mathrm{E}(Z_j)}\cancelto{0}{\mathrm{E}(Z_k)} + \\
&\quad + N(N-1)(N-2)(N-3)\cancelto{0}{\mathrm{E}(Z_i)}\cancelto{0}{\mathrm{E}(Z_j)}\cancelto{0}{\mathrm{E}(Z_k)}\cancelto{0}{\mathrm{E}(Z_h)} = \\
&= Nm_4 + 3N(N-1)m_2^2. \qquad \text{(F.17)}
\end{aligned}$$

The simplifications were possible due to $\mathrm{E}(Z) = 0$. Put together, Eq. (F.12) becomes

$$\begin{aligned}
\mathrm{Var}(\widehat{s}^{\,2}) &= \frac{(N^2-2N+1)Nm_4 + [N^2(N-1) - 2N(N-1) + 3(N-1)]Nm_2^2}{N^2\,(N-1)^2} - m_2^2 = \\
&= \frac{(N^2-2N+1)m_4 - (N^2-4N+3)m_2^2}{N\,(N-1)^2} = \frac{(N-1)^2 m_4 - (N-1)(N-3)m_2^2}{N\,(N-1)^2} = \\
&= \frac{1}{N}\left(m_4 - \frac{N-3}{N-1}m_2^2\right). \qquad \text{(F.18)}
\end{aligned}$$

Using $\widehat{s}^{\,2}$ instead of m_2, Eq. (F.18) finally becomes

$$\mathrm{Var}(\widehat{s}^{\,2}) = \frac{1}{N}\left[m_4 - \frac{N-3}{N-1}\left(\frac{N-1}{N}\right)^2 \widehat{s}^{\,4}\right] = \frac{1}{N}\left[m_4 - \frac{(N-3)(N-1)}{N^2}\widehat{s}^{\,4}\right]. \qquad \text{(F.19)}$$

APPENDIX G

Calculation of the ln A moments

As already introduced in Section 2.5, the $X_{\max}$ moments can be interpreted as moments of $\ln A$ according Eqs. (2.20) and (2.22):

$$\langle X_{\max} \rangle = \langle X^{p}_{\max} \rangle - f_E \langle \ln A \rangle,$$
$$\widehat{\sigma}^2(X_{\max}) = \langle \widehat{\sigma}^2_A \rangle + f_E^2 \, \sigma^2(\ln A).$$

The average of Eq. (2.16) in the superposition model, based on the simple Heitler-Matthews description, can be written as

$$\langle X^{A}_{\max} \rangle = \langle \widehat{X}_h \rangle + D \log_{10} \left(\frac{E}{A} \right), \tag{G.1}$$

where $\langle \widehat{X}_h \rangle$ is the average depth of the first interaction point, E is the energy of the primary, A its nuclear mass and D the elongation rate. This equation can be generalized to consider the deviations from this simple model introduced by the hadronic interactions [46] as

$$\langle X^{A}_{\max} \rangle = \langle \widehat{X}_h \rangle + D \log_{10} \left(\frac{E}{A} \right) + \xi \ln A + \eta \ln A \log_{10} (E) \, . \tag{G.2}$$

Thus the parameters of Eq. (2.20) can be expressed as

$$\langle X^{p}_{\max} \rangle = \langle \widehat{X}_h \rangle + D \log E, \tag{G.3}$$
$$f_E = \xi - \frac{D}{\ln 10} + \eta \log E, \tag{G.4}$$

where $\log E$ is $\log_{10} (E)$.

The fluctuation for a shower produced by a primary with atomic mass A is parametrize as a function of $\ln A$ [46] with

$$\widehat{\sigma}_A^2 = \widehat{\sigma}_p^2 \left[1 + a \ln A + b \ln^2 A\right], \tag{G.5}$$

where

$$\widehat{\sigma}_p^2 = p_0 + p_1 \log E + p_2 \log E^2, \tag{G.6}$$

$$a = a_0 + a_1 \log E. \tag{G.7}$$

Average of Eq. (G.5) over all the masses in a mixture with the relative fractions used as weights leads to

$$\begin{aligned}\langle\widehat{\sigma}_A^2\rangle &= \widehat{\sigma}_p^2 \left[1 + a\langle\ln A\rangle + b\langle\ln^2 A\rangle\right] = \\ &= \widehat{\sigma}_p^2 \left[1 + a\langle\ln A\rangle + b\langle\ln A\rangle^2\right] + \widehat{\sigma}_p^2\, b\, \sigma^2(\ln A) = \\ &= \langle\widehat{\sigma}_{\mathrm{sh}}^2\rangle + \widehat{\sigma}_p^2\, b\, \sigma^2(\ln A),\end{aligned} \tag{G.8}$$

where

$$\langle\widehat{\sigma}_{\mathrm{sh}}^2\rangle = \langle\widehat{\sigma}_{\mathrm{sh}}^2\rangle(\log E, \langle\ln A\rangle) = \widehat{\sigma}_p^2 \left[1 + a\,(\log E)\,\langle\ln A\rangle + b\langle\ln A\rangle^2\right] \tag{G.9}$$

is a function of $\log E$ and $\langle\ln A\rangle$.

Finally the moments of $\ln A$ can be calculated from Eqs. (2.20) and (2.22) as

$$\langle\ln A\rangle = \frac{\langle X_{\max}^p\rangle - \langle X_{\max}\rangle}{f_E}, \tag{G.10}$$

$$\sigma_{\ln A}^2 = \frac{\widehat{\sigma}^2(X_{\max}) - \langle\widehat{\sigma}_{\mathrm{sh}}^2\rangle}{f_E^2 + \widehat{\sigma}_p^2\, b}, \tag{G.11}$$

where $\langle\widehat{\sigma}_A^2\rangle$ has been replaced using Eq. (G.8), and $\sigma_{\ln A}^2 = \sigma^2(\ln A)$. All the parameters are functions of $\log E$, and $\langle\widehat{\sigma}_{\mathrm{sh}}^2\rangle$ is also a function of $\langle\ln A\rangle$. The errors in the measured variables affect these parameters, thus the uncertainties must be properly propagated.

The statistical uncertainty will be

$$\varepsilon_{\langle\ln A\rangle} = \frac{\varepsilon_{\langle X_{\max}\rangle}}{f_E}, \tag{G.12}$$

$$\begin{aligned}\varepsilon_{\sigma_{\ln A}} &= \frac{\sqrt{\varepsilon_{\widehat{\sigma}^2}^2 + \varepsilon_{\langle\widehat{\sigma}_{\mathrm{sh}}^2\rangle}^2}}{f_E^2 + \widehat{\sigma}_p^2\, b} = \\ &= \frac{\sqrt{4\,\widehat{\sigma}^2 \cdot \varepsilon_{\widehat{\sigma}}^2 + \left[\widehat{\sigma}_p^2\,(a + 2\,b\,\langle\ln A\rangle)\right]^2 \cdot \varepsilon_{\langle\ln A\rangle}^2}}{f_E^2 + \widehat{\sigma}_p^2\, b},\end{aligned} \tag{G.13}$$

where $\varepsilon_{\langle X_{\max}\rangle}$ and $\varepsilon_{\widehat{\sigma}}$ are the statistical uncertainty of $\langle X_{\max}\rangle$ and $\widehat{\sigma}$ described in Eqs. (F.1) and (F.7). The propagated uncertainty $\varepsilon_{\langle\widehat{\sigma}_{\mathrm{sh}}^2\rangle}$ is derived from Eq. (G.9).

The systematic uncertainty is obtained by numerically propagating the systematics of the variables. $\langle\ln A\rangle = \langle\ln A\rangle(\langle X_{\max}\rangle, \log E)$ is a function of $\langle X_{\max}\rangle$ and $\log E$. $\sigma_{\ln A}^2$ is a function of $\widehat{\sigma}$, $\log E$ and $\langle\ln A\rangle$, thus $\sigma_{\ln A}^2 = \sigma_{\ln A}^2(\widehat{\sigma}, \log E, \langle X_{\max}\rangle)$. The complete propagation of the systematics yields

$$\delta\langle\ln A\rangle = \sqrt{\left(\langle\ln A\rangle_{\delta\langle X_{\max}\rangle} - \langle\ln A\rangle\right)^2 + \left(\langle\ln A\rangle_{\delta E} - \langle\ln A\rangle\right)^2}, \tag{G.14}$$

$$\delta\sigma_{\ln A}^2 = \sqrt{\left(\sigma_{\delta\langle X_{\max}\rangle}^2 - \sigma_{\ln A}^2\right)^2 + \left(\sigma_{\delta E}^2 - \sigma_{\ln A}^2\right)^2 + \left(\sigma_{\delta\widehat{\sigma}}^2 - \sigma_{\ln A}^2\right)^2}, \tag{G.15}$$

where

$$\langle \ln A \rangle_{\delta\langle X_{\max}\rangle} = \langle \ln A \rangle \left(\langle X_{\max} \rangle + \delta \langle X_{\max} \rangle \right), \tag{G.16}$$

$$\langle \ln A \rangle_{\delta E} = \langle \ln A \rangle \left(\log E + \log_{10}(1 + \delta E) \right), \tag{G.17}$$

$$\sigma^2_{\delta\langle X_{\max}\rangle} = \sigma^2_{\ln A} \left(\langle X_{\max} \rangle + \delta \langle X_{\max} \rangle \right), \tag{G.18}$$

$$\sigma^2_{\delta E} = \sigma^2_{\ln A} \left(\log E + \log_{10}(1 + \delta E) \right), \tag{G.19}$$

$$\sigma^2_{\delta\widehat{\sigma}} = \sigma^2_{\ln A} \left(\widehat{\sigma} + \delta\widehat{\sigma} \right) \tag{G.20}$$

are the $\ln A$ moments with one of their variablea shifted by its uncertainty.

The parameters $\langle \widehat{X}_h \rangle$, D, ξ, η, a_0, a_1, b, p_0, p_1 and p_2 strongly depends on the hadronic interaction model used to parametrize them. A summary of their values are reported in Table G.1. These values are valid for energies rescaled by $E_0 = 10^{19}$ eV, i.e. $\log E = \log_{10}(E/E_0)$.

Table G.1: Summary of the parameters in Eqs. (G.3) to (G.7) for different hadronic interaction models. From [46, 117]

Model	$\langle \widehat{X}_h \rangle$	D	ξ	η	a_0	a_1	b	p_0	p_1	p_2
EPOS-LHC	806.1	55.6	0.15	0.83	-0.462	-0.0008	0.059	3284	-260	132
Sibyll2.1	795.1	57.7	-0.04	-0.04	-0.368	-0.0049	0.039	2785	-364	152
QGSJetII-04	790.4	54.4	-0.31	0.24	-0.397	0.0008	0.046	3738	-375	-21

APPENDIX H

Cut lists

These configuration files and lists of cuts are meant to be used with the executable "selectEvents" in the *ADST* package from Offline.

H.1 Standard analysis

The cuts are equivalent to the selections in Tables 4.1 and 4.3 and discussed in Sections 4.2 and 4.3.

Configuration file:

```
FDCutFile              DataAcquisition.cuts      # Data acquisition selection
FDCutFile              Hybrid.cuts               # Hybrid selection
FDCutFile              Atmosphere.cuts           # Atmosphere selection
FDCutFile              HeCo.cuts                 # HeCo selection
FDCutFile              Reconstruction.cuts       # Reconstruction selection
FDCutFile              FOV.cuts                  # FoV selection
T3TimesFileCo          t3TimeStampsCo.txt        # For RejectCDASVetoPeriods,
T3TimesFileHEAT        t3TimeStampsHEAT.txt      # RejectFDASVetoPeriods and
                                                 # RejectT3VetoPeriods cuts
CDASVetoFileCo         cdasVetoTimesCO.txt       # For RejectCDASVetoPeriods cut
CDASVetoFileHEAT       cdasVetoTimesHEAT.txt     # For RejectCDASVetoPeriods cut
FDASVetoFileCo         fdasVetoTimesCo.txt       # For RejectFDASVetoPeriods cut
FDASVetoFileHEAT       fdasVetoTimesHEAT.txt     # For RejectFDASVetoPeriods cut
T3VetoFileHEAT         t3ReductionTimesHEAT.txt  # For RejectT3VetoPeriods cut
FDAmbiguousHybridList  ambiguousHybrids.txt      # For ambiguousHybridRejection
FDCloudData            cloudCamHECOv9r5.txt.gz   # For cloudCut cut
```

DataAcquisition.cuts

```
adst cuts version: 1.0
badFDPeriodRejection
skipSaturated
noBadPixelsInPulse
minBackgroundRMSMergedEyes 17 6 110000
minBackgroundRMSSimpleEyes 17 11111
good10MHzCorrection
```

Hybrid.cuts

```
adst cuts version: 1.0
maxCoreTankDist            1.e20
hybridTankTrigger          2
maxZenithFD                85
minLgEnergyFD              1.e-20
maxCoreTankDist            1500
ambiguousHybridRejection
```

Atmosphere.cuts

```
adst cuts version: 1.0
hasMieDatabase
maxVAOD                    { params:  .1
                             nMinusOne: 100 0. 1. }
cloudCut
```

HeCo.cuts

```
adst cuts version: 1.0
HeCoHasUpTime
minPBrass                  0.9
maxPBrassProtonIronDiff    0.05
RejectCDASVetoPeriods      100000 4
RejectFDASVetoPeriods      100000 4
RejectT3VetoPeriods        100000 4
```

Reconstruction.cuts

```
adst cuts version: 1.0
depthTrackLength           200
XmaxErrorLessThenXmax
maxDepthHole               30
profileChi2Sigma           4. -0.742
```

FOV.cuts

```
adst cuts version: 1.0
xMaxObsInExpectedFOV       40 20
fidFoV
```

H.2 Stereo analysis

The cuts are used for the study on the compatibility of the stereos discussed in Section 4.1.2. They are used to produce the results in Fig. 4.2.

Configuration file used for stereos between standard telescopes:

```
FDCutFile               CommonSelection.cuts
FDCutFile               StdStereoSelection.cuts
```

Configuration file used for stereos between CO and HEAT in downward:

```
FDCutFile               CommonSelection.cuts
FDCutFile               CO-HEATStereoSelection.cuts
```

CommonSelection.cuts

```
adst cuts version: 1.0

!minDate                    120815 1
badFDPeriodRejection
skipSaturated
noBadPixelsInPulse
minLgEnergyFD               1.e-20
maxCoreTankDist             1500.
maxZenithFD                 85
maxVAOD                     { params: .1
                              nMinusOne: 100 0. 1. }
profileChi2Sigma            4. -1.2
maxDepthHole                30.
XmaxErrorLessThenXmax
xMaxObsInExpectedFOV        40 20
```

StdStereoSelection.cuts

```
adst cuts version: 1.0

eyeCut                      1111
hasMieDatabase
LidarCloudRemoval           { params: 25
                              nMinusOne: 101 -1. 100. }
```

CO-HEATStereoSelection.cuts

```
adst cuts version: 1.0

eyeCut                      11000
!heatOrientationUp
minPBrass                   0.9
maxPBrassProtonIronDiff     0.05
```

APPENDIX I

Glossary

Acronyms and names

AERA	**A**uger **E**ngineering **R**adio **A**stronomy
AMIGA	**A**uger **M**uons and **I**nfill for the **G**round **A**rray
BLS	**B**alloon **L**aunch **S**tation
Brass Hybrid	Hybrid Event formed by less than 3 SD stations, with a T3 from a FD Event
CDAS	**C**entral **D**ata **A**cquisition **C**enter
CLF	**C**entral **L**aser **F**acility
CMB	**C**osmic **M**icrowave **B**ackground
CO	**Co**ihueco telescopes
CR	**C**osmic **R**ay
DAQ	**D**ata **Ac**quisition system
DB	**D**ata**b**ase
EAS	**E**xtensive **A**ir **S**howers
EBL	**E**xtragalactic **B**ackground **L**ight
EPOS-LHC	Mixed glauber-pomeron hadronic model. Suited also to simulate heavy ion collisions
ER	**E**longation **R**ate
expFoV	**exp**ected **FoV**
FADC	**F**lash **A**nalogical **D**igital **C**onverter
FD	**F**luorescence **D**etector
FD Event	Event formed by the data from the telescopes of the same FD location
fidFoV	**fid**ucial **FoV**
FLT	**F**irst **L**evel **T**rigger. FD trigger
FoV	**F**ield **o**f **V**iew
GDAS	**G**lobal **D**ata **A**ssimilation **S**ystem

G-H	**G**aisser-**H**illas
GOES	**G**eostationary **O**perational **E**nvironmental **S**atellites
Golden Hybrid	Hybrid Event formed by at least 3 SD stations with an independent T3
GZK	**G**reisen-**Z**atsepin-**K**uz'min limit
HEAT	**H**igh **E**elevation **A**uger **T**elescopes
HEATlet	Infill-like SD stations nearest to HEAT
HeCo	**HE**AT+**Co**ihueco merged telescopes
Hybrid Event	Event formed merging (offline) FD Events and SD Event from the same air shower
Infill	Stations placed halfway between two other tanks, 750 m of distance each one
LA	**L**oma **A**marilla telescopes
LL	**L**os **L**eones telescopes
LM	**L**os **M**orados telescopes
Lx	LL+LM+LA telescopes
MC	Simulation using the **M**note **C**arlo method
Measured Mie Model	T_{aer} calculated with measured VAOD
Parametric Mie Model	T_{aer} calculated with Mie scattering theory on spherical scatterers. Used only if VAOD is note measured
Pixel	PMT of a camera
PMT	**P**hoto**m**ultiplier **T**ube
QCD	**Q**uantum **C**hromo**d**ynamics
QGSJet-II	Mixed glauber-pomeron hadronic model with the lest number of free parameters
SD	**S**urface **D**etector array
SD Event	Event from the SD data acquisition
SDP	**S**hower **D**etector **P**lane
Sibyl-2.1	Pure Glauber hadronic model. It is the simplest
SLT	**S**econd **L**evel **T**rigger. FD trigger
SNR	**S**uper**n**ova **R**emnant. Main paradigm for the galactic flux models
Stereo Event	2 FD Events generated by the same air shower
T1	First level of TH
T2	Second level of TH
T3	Third level of trigger or Hybrid trigger. SD trigger. Can be externally generated by FD Events
TH	**Th**reshold trigger. SD trigger
TLT	**T**hird **L**evel **T**rigger. FD trigger
ToT	**T**ime **o**ver **T**hreshold trigger. SD trigger
UHECR	**U**ltra **H**igh **E**nergy **CR**
Uptime	Fraction of the time in which a single telescope is in DAQ in the arc of ten minutes
VA	**V**iewing **A**ngle
VAOD	**V**ertical **A**erosol **O**ptical **D**epth
VEM	**V**ertical **E**quivalent **M**uon
XLF	**E**xtreme **L**aser **F**acility

Mathematical Variables

$\boldsymbol{A(\log E_{cal}^{true}, X_{max}^{true})}$	Function of the acceptance
$\mathbf{b_{LW}}$	Bias of the lateral width correction
$\boldsymbol{w(\log E_{cal}^{true}, X_{max}^{true})}$	Function of the acceptance correction weights (A^{-1})
$\boldsymbol{\Delta}$	Maximum acceptable deviation of $\langle X_{max}^{trunc} \rangle$ from X_{max}^{∞}, expressed in g/cm^2
$\boldsymbol{\Delta X_{align}}$	Telescope alignment adjustment, expressed in g/cm^2
$\boldsymbol{\delta \langle X_{max} \rangle}$	Systematics uncertainty of the $\langle X_{max} \rangle$, expressed in g/cm^2
$\boldsymbol{\delta}\,\mathbf{Res}$	Systematics uncertainty of the Res, expressed in g/cm^2
$\boldsymbol{\delta \widehat{\sigma}}$	Systematics uncertainty of the $\widehat{\sigma}(X_{max})$, expressed in g/cm^2
$\boldsymbol{\varepsilon_{\widehat{\sigma}}}$	Statistical uncertainty of the $\widehat{\sigma}(X_{max})$, expressed in g/cm^2
$\boldsymbol{\varepsilon_{\langle X_{max} \rangle}}$	Statistical uncertainty of the $\langle X_{max} \rangle$, expressed in g/cm^2
$\boldsymbol{E}$	Total energy (in eV unit): measured, scaled by the invisible energy factor
$\boldsymbol{\langle E_{bias} \rangle (\log E_{cal}^{obs})}$	Energy bias function
$\boldsymbol{E_{cal}}$	Calorimetric energy, expressed in eV: the measured one
$\boldsymbol{E_{cal}^{gen}}$	Generated calorimetric energy from the MC, expressed in eV
$\boldsymbol{E^{obs}}$	Observed energy from the data, expressed in eV
$\boldsymbol{E_{cal}^{obs}}$	Measured calorimetric energy from the data, expressed in eV
$\boldsymbol{E_{cal}^{rec}}$	Reconstructed calorimetric energy from the MC, expressed in eV
$\boldsymbol{E^{true}}$	True energy of the data, expressed in eV: the observed corrected by f_{bias}
$\boldsymbol{E_{cal}^{true}}$	True calorimetric energy from the data, expressed in eV
$\mathbf{erf}(\boldsymbol{x})$	Error function
$\mathbf{erfc}(\boldsymbol{x})$	Complementary error function: $1 - \mathrm{erf}(x)$
$\boldsymbol{f_{bias}(x)}$	Function of the factor for the energy bias correction
$\boldsymbol{f_{invis}(x)}$	Function of the invisible energy factor
$\boldsymbol{G_{reco}(X)}$	Gaussian distribution of the reconstruction folded with the real distribution (bring information about biases and resolutions)
$\boldsymbol{G \otimes E(X)}$	Gaussian folded exponential distribution
$\boldsymbol{\lambda_1 / \lambda_2}$	Parameters of the left ad right exponential part of the acceptance A, expressed in g/cm^2
$\log \boldsymbol{E}$	Contracted form of the variable $\log_{10}(E/\mathrm{eV})$
$\log \boldsymbol{E^{obs}}$	Measured $\log_{10}(E/\mathrm{eV})$ from the data
$\log \boldsymbol{E^{true}}$	True $\log_{10}(E/\mathrm{eV})$ of the data
$\log \boldsymbol{E_{cal}}$	Contracted form of the variable $\log_{10}(E_{cal}/\mathrm{eV})$
$\log \boldsymbol{E_{cal}^{gen}}$	Generated $\log_{10}(E_{cal}/\mathrm{eV})$ from the MC
$\log \boldsymbol{E_{cal}^{obs}}$	Measured $\log_{10}(E_{cal}/\mathrm{eV})$ from the data
$\log \boldsymbol{E_{cal}^{rec}}$	Reconstructed $\log_{10}(E_{cal}/\mathrm{eV})$ from the MC
$\log \boldsymbol{E_{cal}^{true}}$	True $\log_{10}(E_{cal}/\mathrm{eV})$ from the data
$\boldsymbol{\mu_{trunk}(x_1, x_2)}$	Truncated mean of he X_{max} distribution between x_1 and x_2, expressed in g/cm^2
mva	**m**inimum **V**iewing **A**ngle
$\widehat{\mathbf{Res}}$	Total resolution, expressed in g/cm^2
$\mathbf{Res_{det}}(\log \boldsymbol{E_{cal}^{true}})$	Function of the detector resolution
$\boldsymbol{\sigma_{atm}}\,(\boldsymbol{X_{max}})$	Uncertainty from the choice of the atmospheric description, expressed in g/cm^2
$\boldsymbol{\sigma_{stat}}\,(\boldsymbol{X_{max}})$	Statistical uncertainty from the X_{max} parametrization of the G-H fit of the profile, expressed in g/cm^2

$\boldsymbol{\sigma_{X_{\max}}(X_{\max})}$	Uncertainty of the measured $X_{\max}$ of each shower, expressed in g/cm^2: $\sigma_{X_{\max}}(X_{\max}) = \sqrt{\sigma^2_{\text{stat}}(X_{\max}) + \sigma^2_{\text{atm}}(X_{\max})}$
$\boldsymbol{\widehat{\sigma}(X_{\max})}$	Standard deviation of the $X_{\max}$ distribution (unbiased estimator), expressed in g/cm^2
$\boldsymbol{T_{\text{aer}}}$	Aerosol transmission coefficient based on the scattering from measured/parametric Mie model
$\boldsymbol{T_{\text{mol}}}$	Molecular trasmission coefficient based on Rayleigh scattering
$\boldsymbol{X_1/X_2}$	Limits of the range in which the acceptance A is flat, expressed in g/cm^2
$\boldsymbol{\langle X_{\text{bias}}\rangle(\log E^{\text{true}}_{\text{cal}})}$	Function of the average of the reconstruction bias
$\boldsymbol{\widehat{X}_{\text{bias}}}$	Total bias of the $X_{\max}$ (in g/cm^2 unit): $\langle X_{\text{bias}}\rangle + \text{b}_{\text{LW}} - \Delta X_{\text{align}}/2$
$\boldsymbol{\widehat{X}_{em}}$	Radiation length, expressed in g/cm^2
$\boldsymbol{\widehat{X}_{h}}$	Hadronic interaction length, expressed in g/cm^2
$\boldsymbol{X_{\text{low}}}$	Lower limit of the expFoV, located up in the atmosphere and expressed in g/cm^2
$\boldsymbol{X^{\text{fid}}_{\text{low}}}$	Lower limit of the fidFoV, located up in the atmosphere and expressed in g/cm^2
$\boldsymbol{X_{\max}}$	Depth of maximum of the shower development, expressed in g/cm^2
$\boldsymbol{\langle X_{\max}\rangle}$	Mean value of the $X_{\max}$ distribution, expressed in g/cm^2
$\boldsymbol{\langle X^{\infty}_{\max}\rangle}$	Asymptotic $\langle X_{\max}\rangle$ in case of unbiased sample, expressed in g/cm^2
$\boldsymbol{X^{\text{corr}}_{\max}}$	Corrected $X_{\max}$ from the data ($X_{\max} - \widehat{X}_{\text{bias}}$), expressed in g/cm^2
$\boldsymbol{X^{\text{gen}}_{\max}}$	Generated $X_{\max}$ from the MC, expressed in g/cm^2
$\boldsymbol{X^{\text{obs}}_{\max}}$	Observed $X_{\max}$ from the data, expressed in g/cm^2
$\boldsymbol{X^{\text{rec}}_{\max}}$	Reconstructed $X_{\max}$ from the MC, expressed in g/cm^2
$\boldsymbol{X^{\text{true}}_{\max}}$	True $X_{\max}$ of the showers, expressed in g/cm^2
$\boldsymbol{\langle X^{\text{trunc}}_{\max}\rangle}$	Average truncated $X_{\max}$ calculated by μ_{trunk}, expressed in g/cm^2
$\boldsymbol{X_{\text{up}}}$	Upper limit of the expFoV, located near the ground and expressed in g/cm^2
$\boldsymbol{X^{\text{fid}}_{\text{up}}}$	Upper limit of the fidFoV, located near the ground and expressed in g/cm^2
$\boldsymbol{\xi}$	Estimated resolution of a guessed $X_{\max}$, expressed in g/cm^2
$\boldsymbol{Y_{\text{air}}}$	Fluorescence yield on air

List of Figures

List of Tables

Bibliography

[1] V. F. Hess, "Über Beobachtungen der durchdringenden Strahlung bei sieben Freiballonfahrten", *Physik.Zeitschr.*, **vol. 13** (1084), 1912

[2] J. Blümer, R. Engel, J. R. Hörandel, "Cosmic rays from the knee to the highest energies", *Prog.Part.Nucl.Phys.*, **vol. 63** (2):pp. 293–338, 2009

[3] R. Engel, D. Heck, T. Pierog, "Extensive Air Showers and Hadronic Interactions at High Energy", *Annu.Rev.Nucl.Part.Sci.*, **vol. 61**:pp. 467–489, 2011

[4] D. R. Bergman, J. W. Belz, "Cosmic rays: the Second Knee and beyond", *J.Phys.G:Nucl.Part.Phys.*, **vol. 34**:pp. 359–400, 2007

[5] KASCADE-Grande Collaboration, "Kneelike structure in the spectrum of the heavy component of cosmic rays observed with KASCADE-Grande", *Phys.Rev.Lett.*, **vol. 107**, 2011

[6] S. P. Knurenko, *et al.*, "Cosmic ray spectrum in the energy range $10^{15}-10^{18}$ eV and the second knee according to the small Cherenkov setup at the Yakutsk EAS array", *Proc. 33rd International Cosmic Ray Conference*, 2013

[7] The IceTop Collaboration, "Measurement of the cosmic ray energy spectrum with IceTop-73", *Phys.Rev*, **vol. D88** (4):p. 042004, 2013

[8] Pierre Auger Collaboration, "Observation of the suppression of the flux of cosmic rays above $4 \cdot 10^{19}$ eV", *Phys.Rev.Lett.*, **vol. 101**, 2008

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

[10] G. T. Zatsepin, V. A. Kuz'min, "Upper Limit of the Spectrum of Cosmic Rays", *JETP Lett.*, **vol. 4**, 1966

[11] R. Aloisio, V. Berezinsky, A. Gazizov, "Transition from galactic to extragalactic cosmic rays", *Astropart.Phys.*, **vol. 39-40**:pp. 129–143, 2012

[12] The KASCADE Collaboration, "KASCADE measurements of energy spectra for elemental groups of cosmic rays: Results and open problems", *Astropart.Phys.*, **vol. 24**:pp. 1–25, 2005

[13] S. G. Lucek, A. R. Bell, "Non-linear amplification of a magnetic field driven by cosmic ray streaming", *Mon.Not.R.Astron.Soc.*, **vol. 314**:pp. 65–74, 2000

[14] P. Blasi, "Origin of very high and ultra high energy cosmic rays", Solicited Review Paper to appear in "Comptes Rendus Physique"

[15] The Pierre Auger Collaboration, "Constraints on the origin of cosmic rays above 10^{18} eV from large scale anisotropy searches in data of the Pierre Auger Observatory", *ApJL*, **vol. L13**:p. 762, 2013

[16] H. Blümer, K.-H. Kampert, "Die Suche nach den Quellen der kosmischen Strahlung", *Physikalische Blätter*, **vol. 56** (3):pp. 39–45, 2000

[17] A. M. Hillas, "The origin of Ultra-High-Energy Cosmic Rays", *Ann.Rev.Astron.Astrophys.*, **vol. 22**:pp. 425–444, 1984

[18] D. F. Torres, L. A. Anchordoqui, "Astrophysical Origins of Ultrahigh Energy Cosmic Rays", *Rept.Prog.Phys.*, **vol. 67**:pp. 1663–1730, 2004

[19] P. Blasi, R. I. Epstein, A. V. Olinto, "Ultra-High-Energy Cosmic Rays from Young Neutron Star Winds", *Astropart.Phys.*, **vol. 533** (2):p. 123, 2000

[20] P. Bhattacharjee, G. Sigl, "Origin and Propagation of Extremely High Energy Cosmic Rays", *Phys.Rept.*, **vol. 327**:pp. 109–247, 2000

[21] G. Gelmini, O. E. Kalashev, D. V. Semikoz, "GZK photons as ultra high energy cosmic rays", *J.Exp.Theor.Phys.*, **vol. 106**:pp. 1061–1082, 2008

[22] J. R. Ellis, V. Mayes, D. Nanopoulos, "UHECR particle spectra from crypton decays", *Phys.Rev*, **vol. D74**:p. 115003, 2006

[23] K.-H. Kampert, M. Unger, "Measurements of the Cosmic Ray Composition with Air Shower Experiments", *Astropart.Phys.*, **vol. 35**:pp. 660–678, 2012

[24] D. Allard, N. G. Busca, G. Decerprit, A. V. Olinto, E. Parizot, "Implications of the cosmic ray spectrum for the mass composition at the highest energies", *JCAP*, **vol. 0810**:p. 033, 2008

[25] Pierre Auger Collaboration, "The measurement of the energy spectrum of cosmic rays above $3 \cdot 10^{17}$ eV with the Pierre Auger Observatory", *Proc. 33rd International Cosmic Ray Conference*, 2013

[26] T. Wibig, A. W. Wolfendale, "At what particle energy do extragalactic cosmic rays start to predominate?", *J.Phys.*, **vol. G31**:pp. 255–264, 2005

[27] V. Berezinsky, A. Z. Gazizov, S. I. Grigorieva, "Dip in UHECR spectrum as signature of proton interaction with CMB", *Phys.Lett.*, **vol. B612**:pp. 147–153, 2005

[28] The Pierre Auger Collaboration, "Update on the measurement of the CR energy spectrum above 10^{18} eV made using the Pierre Auger Observatory", *Proc. 32nd International Cosmic Ray Conference*, 2011

[29] D. Allard, "Extragalactic propagation of ultrahigh energy cosmic-rays", *Astropart.Phys.*, **vol. 39-40**:pp. 33–43, 2011

[30] D. Allard, E. Parizot, E. Khan, S. Goriely, A. Olinto, "UHE nuclei propagation and the interpretation of the ankle in the cosmic-ray spectrum", *Astron.Astrophys.*, **vol. 443**:pp. L29–L32, 2005

[31] J. F. Carlson, J. R. Oppenheimer, "On Multiplicative Showers", *Phys.Rev.*, **vol. 51**:pp. 220–231, 1937

[32] W. Heitler, *Quantum Theory of Radiation* (Oxford University Press, 1954)

[33] Paul Friederich Gaspard Gert Molière, "Theory of the scattering of fast charged particles. 2. Repeated and multiple scattering", *Z.Naturforsch.*, **vol. A3** (78-97), 1948

[34] B. B. Rossi, "Misure della distribuzione angolare di intensità della radiazione penetrante all'Asmara", *Ric.Sci.Suppl.*, **vol. 1**:p. 579, 1934

[35] P. V. Auger, P. Ehrenfest, R. Maze, J. Daudin, R. A. Fréon, "Extensive Cosmic-Ray Showers", *Review of Modern Physics*, **vol. 11** (288-291), 1939

[36] A. Haungs, H. Rebel, M. Roth, "Energy spectrum and mass composition of high-energy cosmic rays", *Rept.Prog.Phys.*, **vol. 66**:pp. 1145–1206, 2003

[37] J. Matthewes, "A Heitler model of extensive air showers", *Astropart.Phys.*, **vol. 22** (287-297), 2005

[38] G. Battistoni, C. Forti, J. Ranft, R. Roesler, "Deviations from the superposition model in a Dual Parton Model with formation zone cascade in both projectile and target nuclei", *Astropart.Phys.*, **vol. 7**:p. 49, 1997

[39] E.-J. Ahn, R. Engel, T. K. Gaisser, P. Lipari, T. Stanev, "Cosmic ray interaction event generator SIBYLL 2.1", *Phys.Rev.*, **vol. D80** (094003), 2009

[40] S. Ostapchenko, "Monte Carlo treatment of hadronic interactions in enhanced Pomeron scheme: QGSJET-II model", *Phys.Rev.*, **vol. D83** (014018), 2011

[41] T. Pierog, I. Karpenko, J. Katzy, E. Yatsenko, K. Werner, "EPOS LHC : test of collective hadronization with LHC data", Prepared for submission to JHEP

[42] R. J. Glauber, G. Matthiae, "High-Energy Scattering of Protons by Nuclei", *Nucl.Phys.*, **vol. B21**:pp. 135–157, 1970

[43] V. N. Gribov, "A Reggeon Diagram Technique", *Sov.Phys.JETP*, **vol. 26** (2):pp. 414–423, 1968

[44] T. Sjöstrand, "Status of Fragmentation Models", *Int.J.Mod.Phys.*, **vol. A3** (751), 1988

[45] J. Engel, T. K. Gaisser, P. Lipari, T. Stanev, "Nucleus-Nucleus Collisions and Interpretation of Cosmic-ray Cascades", *Phys.Rev.*, **vol. D46** (5013), 1992

[46] The Pierre Auger Collaboration, "Interpretation of the Depths of Maximum of Extensive Air Showers Measured by the Pierre Auger Observatory", *JCAP*, **vol. 1302**:p. 026, 2013

[47] The Pierre Auger Collaboration, "Highlights from the Pierre Auger Observatory", *Proc. 33rd International Cosmic Ray Conference*, 2013

[48] The Telescope Array Collaboration, The Pierre Auger Collaboration, "Pierre Auger Observatory and Telescope Array: Joint Contributions to the 33rd International Cosmic Ray Conference (ICRC 2013)", *Proc. 33rd International Cosmic Ray Conference*, 2013

[49] R. Aloisio, V. Berezinsky, A. Gazizov, "Ultra High Energy Cosmic Rays: The disappointing model", *Astropart.Phys.*, **vol. 34** (620-626), 2011

[50] R. Aloisio, V. Berezinsky, P. Blasi, "Ultra high energy cosmic rays: implications of Auger data for source spectra and chemical composition", 2013. Prepared for submission to JCAP

[51] O. Deligny, "Cosmic rays around 10^{18} eV: implications of contemporary measurements on the origin of the ankle feature", 2014. Solicited review paper to appear in "Comptes Rendus: Physique"

[52] Pierre Auger Collaboration, "Properties and performance of the prototype instrument for the Pierre Auger Observatory", *Nucl.Instrum.Meth.*, **vol. A 523**:pp. 50–95, 2004

[53] D. Veberič, "Maps of the Pierre Auger Observatory", Webpage

[54] The Pierre Auger Collaboration, "The fluorescence detector of the Pierre Auger Observatory", *Nucl.Instrum.Meth.*, **vol. A 620**:pp. 227–251, 2010

[55] The Pierre Auger Collaboration, "The exposure of the hybrid detector of the Pierre Auger Observatory", *Astropart.Phys.*, **vol. 34**:pp. 368–381, 2011

[56] The Pierre Auger Collaboration, "The Pierre Auger Observatory V: Enhancements", *Proc. 32nd International Cosmic Ray Conference*, 2011

[57] S. Müller, "Die High Elevation Auger Telescopes (HEAT) Erweiterung des südlichen Pierre Auger Observatoriums", Talk at the Tagung der Deutschen Physikalischen Gesellschaft, 2010

[58] J. Brack, *et al.*, "Fluorescence Detector Absolute Calibration January and June 2010", **GAP-2010-118**

[59] B. Gookin, *et al.*, "November Absolute Calibration, Addendum to GAP 2010-118", **GAP-2011-037**

[60] The Pierre Auger Collaboration, "Identifying Clouds over the Pierre Auger Observatory using Infrared Satellite Data", *Astropart.Phys.*, **vol. 50-52**:pp. 92–101, 2013

[61] Pierre Auger Collaboration, "Techniques for Measuring Aerosol Attenuation using the Central Laser Facility at the Pierre Auger Observatory", *JINST*, 2013

[62] The Pierre Auger Collaboration, "Description of Atmospheric Conditions at the Pierre Auger Observatory using the Global Data Assimilation System (GDAS)", *Astropart.Phys.*, **vol. 35**:pp. 591–607, 2012

[63] M. Buscemi, *et al.*, "Update of the Aerosol Database May 2013", **GAP-2013-048**

[64] The AIRFLY Collaboration, "Temperature and Humidity Dependence of Air Fluorescence Yield measured by AIRFLY", *Nucl.Instrum.Meth.*, **vol. A597**:pp. 50–54, 2008

[65] The AIRFLY Collaboration, "Precise measurement of the absolute fluorescence yield of the 337 nm band in atmospheric gases", *Astropart.Phys*, **vol. 42**:pp. 90–102, 2013

[66] B. Dawson, *et al.*, "Proposal for an update of the Auger Energy Scale", **GAP-2012-124**

[67] Pierre Auger Collaboration, "The Energy Scale of the Pierre Auger Observatory", *Proc. 33th International Cosmic Ray Conference*, 2013

[68] I. Allekotte, *et al.*, "The surface detector system of the Pierre Auger Observatory", *Nucl.Instrum.Meth.*, **vol. A 586**:pp. 409–420, 2008

[69] Private comunication with J. Blümer

[70] R. Bardenet, B. Kégl, D. Veberič, "Single muon response: The signal model", **GAP-2010-110**

[71] The Pierre Auger Collaboration, "Measurement of the energy spectrum of cosmic rays above 3×10^{17} eV using the AMIGA 750 m surface detector array of the Pierre Auger Observatory", *Proc. 33rd International Cosmic Ray Conference*, 2013

[72] S. Messina, A. M. van den Berg, "Reconstruction comparison between Infill and AERAlet array", **GAP-2013-049**

[73] Pierre Auger Collaboration, "Trigger and aperture of the surface detector array of the Pierre Auger Observatory", *Nucl.Instrum.Meth.*, **vol. A 613**:pp. 29–39, 2010

[74] P. Billoir, "Top-down Selection of Events and Stations in Surface Detector Triggers", **GAP-2006-072**

[75] M. Unger, B. Dawson, R. Engel, F. Schüssler, R. Ulrich, "Reconstruction of longitudinal profiles of ultra-high energy cosmic ray showers from fluorescence and Cherenkov light measurements", *Nucl.Instrum.Meth.*, **vol. 588 A** (3):pp. 433–441, 2008

[76] F. Nerling, J. Blümer, R. Engel, M. Risse, "Universality of electron distributions in high-energy air showers - Description of Cherenkov light production", *Astropart.Phys.*, **vol. 24**:pp. 421–437, 2006

[77] M. D. Roberts, "The role of atmospheric multiple scattering in the transmission of fluorescence light from extensive air showers", *J.Phys.G:Nucl.Part.Phys.*, **vol. 31** (11):p. 1291, 2005

[78] T. K. Gaisser, A. M. Hillas, "Reliability of the method of constant intensity cuts for reconstructing the average development of vertical showers", *Proc. 15th International Cosmic Ray Conference*, 1977

[79] H. M. J. Barbosa, F. Catalani, J. Chinellato, C. Dobrigkeit, "Determination of the calorimetric energy in extensive air showers", *Astropart.Phys.*, **vol. 22** (2):pp. 159–166, 2004

[80] Pierre Auger Collaboration, "A new method for determining the primary energy from the calorimetric energy of showers observed in hybrid mode on a shower-by-shower basis", *Proc. 32th International Cosmic Ray Conference*, 2011

[81] Pierre Auger Collaboration, "Estimate of the non-calorimetric energy of showers observed with the fluorescence and surface detectors of the Pierre Auger Observatory", *Proc. 33th International Cosmic Ray Conference*, 2013

[82] Pierre Auger Collaboration, "Measurement of the Optical Properties of the Auger Fluorescence Telescopes", *Proc. 33th International Cosmic Ray Conference*, 2013

[83] S. Falk, *et al.*, "A First Look at HEAT Data", **GAP-2010-123**

[84] F. Meyer, F. Vernotte, "Time Tagging Board Tests at Besançon Observatory", **GAP-2001-050**

[85] M. Settimo, L. Perrone, G. Cataldi, I. D. Mitri, D. Martello, "Parameterisation of the "Lateral Trigger Probability" functions at low energies", **GAP-2007-069**

[86] Private comunication with M. Unger

[87] E. J. Ahn, *et al.*, "Measurement of the Depth of Shower Maximum of Cosmic Rays above 10^{18} eV", **GAP-2009-078**

[88] M. Unger, J. Bellido, F. Montanet, P. Privitera, R. Ulrich, "Status of $X_{\max}$ Paper", Talk at the Pierre Auger Observatory collaboration meeting

[89] M. Unger, J. Bellido, "Supplementary Material for the 'long $X_{\max}$ Paper'", 2014

[90] V. Verzi. Private communication

[91] J. Pekala, P. Homola, B. Wilczynska, H. Wilczynski, "Atmospheric multiple scattering of fluorescence and Cherenkov light emitted by extensive air showers", *Nucl.Instrum.Meth.*, **vol. A605**:pp. 388–398, 2009

[92] M. Unger, R. Engel, F. Schüssler, R. Ulrich, "Lateral shower light distributions in the Cherenkov-Fluorescence-Matrix Profile Reconstruction", **GAP-2008-052**

[93] K. Loudec *et al.*, "Effect of the aerosol size on multiple scattering - Proposal for an updated systematic uncertainty on energy and depth of maximum ", **GAP-2013-116**

[94] S. Falk, R. Engel, R. Ulrich, M. Unger, "Telescope Alignment Studies", **GAP-2011-123**

[95] M. Prouza *et al.*, "Star tracking using background data of FD telescopes", **GAP-2005-041**

[96] C. De Donato et al., "Using star tracks to determine the absolute pointing of the FD telescopes", **GAP-2005-008**

[97] P. Younk, B. Fick, "FD Telescope alignment using CLF laser shots", **GAP-2006-087**

[98] M. Scuderi, *et al.*, "Alignment method of the Pierre Auger Observatory fluorescence telescopes based on the central laser facility events", **GAP-2006-061**

[99] E. Varela, H. Salazar, J. Bellido, "HEAT-Co Systematics due to the SD/FD time offset", Talk at the Pierre Auger Observatory Analysis meeting, 2013

[100] M. Unger, J. Bäuml, R. Engel, R. Ulrich, "Simulation of the Point Spread Function of the FD Telescopes and its Effect on the Reconstructed Energy and $X_{\max}$", **GAP-2013-069**

[101] R. R. Prado, V. de Souza, "Unfolded Xmax distributions without fiducial volume cuts", **GAP-2011-106**

[102] J. Linsley, "Structure of Large Air Showers at Depth 834 g/cm^2 Applications", *Proc. 15th International Cosmic Ray Conference*, **vol. 12**:p. 89, 1977

[103] J. Linsley, A. A. Watson, "Validity of Scaling to 10^{20} eV and High-Energy Cosmic-Ray Composition", *Phys.Rev.Lett.*, **vol. 46**:p. 459, 1981

[104] The Tunka Collaboration, "The Cosmic Ray Mass Composition in the Energy Range $10^{15} - 10^{18}$ eV measured with the Tunka Array: Results and Perspectives", *Nucl.Phys.Proc.Suppl.*, **vol. 190**:pp. 247–252, 2009

[105] The Tunka Collaboration, "Tunka-133: Primary Cosmic Ray Mass Composition in the Energy Range $6 \cdot 10^{15} - 10^{18}$ eV", *Proc. 32nd International Cosmic Ray Conference*, **vol. Vol1**:pp. 197–200, 2011

[106] S. Knurenko, A. Sabourov, "The depth of maximum shower development and its fluctuations: cosmic ray mass composition at $E_0 \geq 10^{17}$ eV", *Proc. XVI ISVHECRI*, 2010

[107] S. P. Knurenko, A. Sabourov, "Study of cosmic rays at the Yakutsk EAS array: Energy spectrum and mass composition", *Nucl.Phys.Proc.Suppl.*, **vol. 212-213**:pp. 241–251, 2011

[108] J. Fowler, *et al.*, "A Measurement of the cosmic ray spectrum and composition at the knee", *Astropart.Phys.*, **vol. 15**:pp. 49–64, 2001

[109] The HiRes-MIA Collaboration, "Measurement of the cosmic ray energy spectrum and composition from 10^{17} to $10^{18.3}$ eV using a hybrid fluorescence technique", *Astrophys.J.*, **vol. 557**:pp. 686–699, 2001

[110] The HiRes Collaboration, "Indications of Proton-Dominated Cosmic-Ray Composition above 1.6 EeV", *Phys.Rev.Lett.*, **vol. 104**:p. 161101, 2010

[111] The Telesope Array Collaboration, "Cosmic Ray in the Northern Hemisphere: Results from the Telescope Array Experiment", *J.Phys.Conf.Ser.*, **vol. 404**:p. 012037, 2012

[112] A. M. Taylor, "UHECR Composition Models", *Astropart.Phys*, **vol. 54**:pp. 48–53, 2013

[113] S.Riggi, D.Allard, M.Unger, "Combined spectrum-X_{max} fit with a likelihood approach", **GAP-2011-101**

[114] D. Boncioli, A. F. Grillo, S. Petrera, F. Salamida, "Combined spectrum-composition fit", **GAP-2012-129**

[115] D. Boncioli, A. D. Matteo, A. F. Grillo, S. Petrera, "The "galactic" composition in mono-dimensional propagation", **GAP-2013-014**

[116] M. Erdmann, D. Walz, "Constraining UHECR source scenarios with spectrum and Xmax distributions", **GAP-2014-034**

[117] The Pierre Auger Collaboration, "Inferences about the mass composition of cosmic rays from data on the depth of maximum at the Auger Observatory", *Proc. 33rd International Cosmic Ray Conference*, 2013

Acknowledgment

First of all, I want to thank Johannes Blümer for agreeing to accept me as a Ph. D. student and the Institut für Kernphysik of Karlsruher Institut für Technologie to host me. I am grateful to Ulrich Husemann for accepting my thesis as a co-referee and for revising it with helpful advises.

My thanks go also to Ralph Engel and Michael Unger for supervising my work during these three years. I had the great opportunity to learn how to be a physicist. Special thanks to Michael which, with his competence and patience, helped me a lot to advance in my work.

I want to thank also Felix Riehn and Alexander Hervé, with which I share the office, for reading my thesis carefully and improving my "italienglish". My thanks also to Detlef Maurel for the same reason and for the help with the german language. I take the opportunity to thank also Hans Dembinski (who taught me also a lot about programming!), Alexander Schultz, Stefanie Falk and the countless colleagues at the Karlsruher Institut für Technologie that make it a very enjoyable working atmosphere. Many thanks go to Sabine Bucher for all her organizational support at the institute (without her, I would have been lost in the paperwork many times!), and Klaus Bekk for the IT support.

I thank also the colleagues of the Pierre Auger collaboration, with which I share the great opportunity of this thesis. Martin Will, with whom I shared my first shift in Argentina, for his help to understand the atmospheric part useful to this thesis. I also thank Daniel Kruppke-Hansen for maintaining and providing the uptime files, supervised by Julian Rautenberg. Ricardo Sato for fixing the CDAS bug after I found the merging inefficiency in HEAT. Jeff Brack for supporting my asymmetry studies in the energies between stereo events. Jose Bellido for giving importance to my results in the mass composition task group. And many more...

I would like to thank also Bruce Dawson and Alessandro Baldini to have accepted to write reference letters for my sake as external referees.

Thanks goes also to colleagues, friends and professors met and acquainted during the summer schools, conferences and the several international opportunities that this Ph. D. formation period offered to me.

I need to thank also the open source community and the software ecosystem that let me develop the programs for this analysis: ROOT, NLopt, SciPy/NumPy and Matplotlib. It must not be forgotten LaTeX, with which I wrote this thesis and prepared all my talks and posters.

Finally, my sincere thanks to my family and my friends, here and far away but still in contact through social network groups, e-mails and all the tools that internet can provide. Their support was the best panacea in the worst moments.

Moreover I thank the Universe for being so amazing...