A Monte Carlo exploration of methods to determine the UHECR composition with the Pierre Auger Observatory

Domenico D'Urso* for the Pierre Auger Collaboration†

*Dipartimento di Fisica dell'Università di Napoli and Sezione INFN, Via Cintia 2, 80123 Napoli, Italy
†Pierre Auger Observatory, Av. San Martin Norte 304, (5613) Malargüe, Prov. De Mendoza, Argentina

***Abstract*. Measuring the mass composition of ultra high energy comsic rays is crucial for understanding their origin. In this paper, we present three statistical methods for determining the mass composition. The methods compare observables measured with the Pierre Auger Observatory with corresponding Monte Carlo predictions for different mass groups obtained using different hadronic interaction models. The techniques make use of the mean and fluctuations of X_{max}, the log-likelihood fit of the X_{max} distributions and the multi-topological analysis of a selection of parameters describing the shower profile. We show their sensitivity to the input composition of simulated samples of known mixing and their ability to reproduce mass sensitive observables, like the average shower maximum as a function of the energy, measured at the Pierre Auger Observatory.**

***Keywords*: Cosmic Ray Mass, Monte Carlo studies**

I. Introduction

The understanding of the nature of the ultra high energy cosmic rays (UHECR) is a crucial point towards the determination of their origin, acceleration and propagation mechanisms. The evolution of the energy spectrum and any explanation of its features strongly depend on the cosmic chemical composition since the galactic confinement, the attenuation length of various energy loss mechanisms and the energy achievable at the sources depend on the primary particle type. Above 10^{19} eV all observed cosmic particles are presumed to have extragalactic origin, because there are no galactic sources able to produce particles up to such energies and they cannot be confined in our galaxy long enough to be accelerated. The energy at which the transition from galactic to extragalactic cosmic rays occurs is still unknown and only a detailed knowledge of the composition spectrum will allow to discriminate among different astrophysical models [1].

The *hybrid* design of the Auger Observatory, the integration of a surface detector array (SD) and a fluorescence detector (FD), exploits stability of experimental operation, a 100% duty cycle, and a simple determination of the effective aperture for the SD, calorimetric shower detection, direct observation of shower longitudinal profile and shower maximum for the FD. This hybrid design allows to simultaneously use the most sensitive parameters to the primary mass from both the SD and FD: the slant depth position X_{max} at which the maximum of the shower profile is reached and its fluctuations from the FD [2][3], the signal risetime in the Cherenkov stations, the curvature of the shower front, the muon-to-electromagnetic ratio and the azimuthal signal asymmetry from the SD [4]. In this paper, we present three statistical techniques for determining the mass of primary particles. These methods compare shower observables measured by the FD at the Pierre Auger Observatory with corresponding Monte Carlo predictions, including a full detector simulation.

II. Composition analysis with the moments of X_{max} distribution

The first two moments of X_{max} distribution, the mean and its variance, have been used as mass discriminators. We derive the mass composition from the best choice of primary fractions that reproduce experimental data using their expectation values (method of moments, MM). With this two observables, the cosmic ray flux can be modeled as a mixture of three primary masses (a, b and c) and define the two parameters describing the mixture P_1 and P_2. The relative abundances in terms of P_1 and P_2 are

$$\begin{aligned} P_a &= P_1 \\ P_b &= P_2(1-P_1) \\ P_c &= (1-P_1)(1-P_2) \end{aligned} \quad (1)$$

The expected mean shower maximum of the mixture is

$$\langle X_{exp}\rangle = P_a\langle X_a\rangle + P_b\langle X_b\rangle + P_c\langle X_c\rangle \quad (2)$$

where $\langle X_a\rangle$, $\langle X_b\rangle$and$\langle X_c\rangle$ are the mean X_{max} for simulated data sets of species a,b and c. The expected X_{max} fluctuations (i.e., the root mean square of the X_{max} distribution) ΔX_{exp} for a mixture of three masses can be written in an easier way defining first its value for two masses, b and c, and then considering its mixture with the species a:

$$\langle X_{b-c}\rangle = P_2\langle X_b\rangle + (1-P_2)\langle X_c\rangle; \quad (3)$$

$$(\Delta X_{b-c})^2 = P_2\Delta X_b^2 + (1-P_2)\Delta X_c^2 + P_2(1-P_2)(\langle X_b\rangle - \langle X_c\rangle)^2; \quad (4)$$

$$(\Delta X_{exp})^2 = P_1\Delta X_a^2 + (1-P_1)\Delta X_{b-c}^2 + P_1(1-P_1)(\langle X_a\rangle - \langle X_{b-c}\rangle)^2; \quad (5)$$

where ΔX_i, $\langle X_{i-j}\rangle$ and ΔX_{i-j} are the X_{max} fluctuations for the primary i and the mean and its fluctuations for the mixture i-j. Assuming that the data set is so large that $\langle X_{max} \rangle$ and ΔX_{max} are statistically independent, in each energy bin we can fit the data could be fitted solving equations 5 and 2 with two unknowns, P_1 and P_2.

III. MASS COMPOSITION FROM A LOGARITHMIC LIKELIHOOD FIT TO X_{max} DISTRIBUTION (LLF)

The method assumes that the observed events N_{data} are a mixture of N_m pure mass samples with unknown fractions p_j. The expected number of showers ν_i with X_{max} into i-th bin is therefore:

$$\nu_i(\mathbf{p}) = N_{data} \sum_{j=0}^{N} p_j \frac{a_{ij}}{N_j^{MC}} \quad , \; i = 1, \ldots, N \tag{6}$$

where a_{ij} are the number of Monte Carlo events from primary j in bin i, N is the total number of bins in the X_{max} distribution and $N_j^{MC} = \sum_{i=0}^{N} a_{ij}$. The probability $P(n_i)$ to observe n_i events in the i-th bin is given by the product of the Poisson distributions of mean ν_i

$$P(n_i) = \prod_{i=1}^{N} \frac{\nu_i^{n_i}}{n_i!} e^{-\nu_i}$$

The logarithm of $P(n_i)$ gives the log-likelihood function:

$$\log L = \sum_{i=0}^{N} [\, n_i \log \nu_i - \nu_i - \log n_i!] \tag{7}$$

Maximizing eq. 7 with respect to p_j, one finds the primary fractions in the measured data sample.

In Fig. 1, we show the X_{max} distribution for a sample of 70% proton-30% iron (black dots) between $10^{18.2}$ and $10^{18.3}$ eV fitted by the weighted sum of the expected X_{max} distribution for proton (dotted line) and iron (dashed line). The techniques searches for the best choice of primary fractions that optimize the fit.

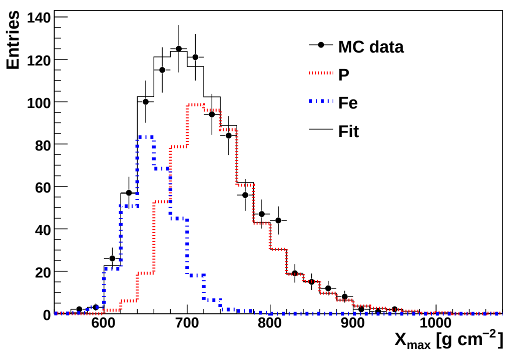

Fig. 1. X_{max} distribution for a sample of 70% proton-30% iron (black dots) between $10^{18.2}$ and $10^{18.3}$ eV fitted by the weighted sum of the expected X_{max} distribution for proton (dotted line) and iron (dashed line) with the best best choice of proton and iron fractions. Distributions are normalized to the number of events in the test sample.

IV. MULTIPARAMETRIC ANALYSIS FOR THE PRIMARY COMPOSITION

We apply the multiparametric topological analysis (MTA) described in [5] to classify observed showers using the correlations of their characteristics. Starting from a set of observables, it is possible to define a parameter space, which is divided in cells whose dimensions are related to the experimental accuracy. A wide set of simulated cascades produced by different primary nuclei is used to populate the parameter space. In each cell, that in the most general n-dimensional case is defined by $(h_1 \ldots h_n)$, one can define the total number of showers $N_{tot}^{(h_1 \ldots h_n)}$ and the total number of showers induced by the primary i $N_i^{(h_1 \ldots h_n)}$ populating the cell, and then derive the associated frequency:

$$p_i^{(h_1 \ldots h_n)} = N_i^{(h_1 \ldots h_n)} / N_{tot}^{(h_1 \ldots h_n)} \tag{8}$$

which can be interpreted as the probability for a shower falling into the cell $(h_1 \ldots h_n)$ to be initiated by a nucleus of mass i. Considering a sample of M showers, its fraction of primary j is given by

$$p_j = \sum_{m=1}^{M} p_j^{(h_1 \ldots h_n)_m} / M \tag{9}$$

with $(h_1 \ldots h_n)_m$ indicating the cell interested in by the m-th event.

A second set of showers is used to compute the mixing probabilities $P_{i \to j}$ that an event of mass i is identified as primary j. The mean $P_{i \to j}$ is obtained by computing p_j for samples of pure primary composition i. Assuming the measured sample as composed by N_m species, the mixing probabilities $P_{i \to j}$ can be used for the reconstruction of the primary mass composition as the coefficients in the system of linear equations:

$$\begin{aligned} N'_1 &= \sum_{i=1}^{N_m} N_i \cdot P_{i \to 1} \\ \vdots &= \vdots \\ N'_{N_m} &= \sum_{i=1}^{N_m} N_i \cdot P_{i \to N_m} \end{aligned} \tag{10}$$

where N_i are the true values, which are altered to N'_j due to misclassification. The solution of eq. 10 gives the mass composition of the measured sample in terms of N_m primary masses and, dividing by the total number of showers N_{data}, their fractions p_j.

MTA performances on CONEX [6] showers, fully simulated through the Auger apparatus, has been already presented [5]. Its performances has been also recently tested on a set of observables from the longitudinal profile and the lateral distribution of CORSIKA [7] simulated showers [8].

In this paper, the MTA application to only FD data using 2 parameters is described and the space is defined by X_{max} and X_0 of the Gaisser-Hillas function

$$\frac{dE}{dX} = \left(\frac{dE}{dX}_{max}\right)\left(\frac{X - X_0}{X_{max} - X_0}\right)^{\frac{X_{max}-X_0}{\lambda}} e^{\frac{X_{max}-X_0}{\lambda}}$$

where dE/dX and $(dE/dX)_{max}$ are the energy deposit at the depth X and at the shower maximum. In Fig. 2 the parameter space built for (X_{max}, X_0) between $10^{18.9}$ and $10^{19.1}$ eV. The space is divided in cells with dimensions 20 and 50 g cm^{-2} respectively and is populated with Conex simulated showers induced by proton (dots) and iron (triangles). Clearly the effective parameter is X_{max}. Despite that, a 2-parameter case is reported to show how the technique can include N parameters in a natural way. The on-going extention to quantities measured by the SD allows to a larger set of effective parameters and to better discriminate among different primaries.

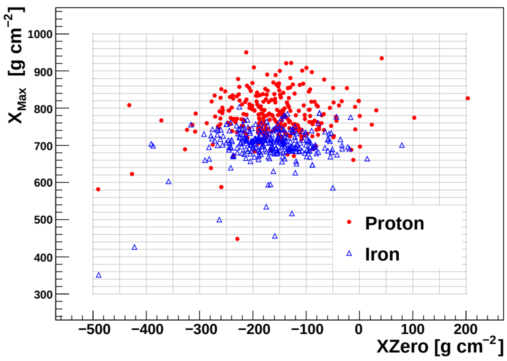

Fig. 2. Parameter space built for (X_{max}, X_0) between $10^{18.9}$ and $10^{19.1}$ eV. The space is divided in cells with dimensions 20 and 50 g cm^{-2} respectively and is populated with Conex simulated showers induced by proton (dots) and iron (triangles).

V. Influence of reconstruction efficiency

The reconstruction of event fractions in terms of N_m masses could be altered by different efficiencies with respect to each primary particle (possible trigger, reconstruction and selection effects). The field of view of the Auger FD, located at 1400 m over the sea level (870 g cm^{-2} of vertical depth), covers an elevation angle from 1.5 to 30. If we require to detect the shower maximum to ensure a good X_{max} resolution, we favour light primaries at lowest energies and heavy nuclei at highest energies. To have an unbiased measurement due to the FD field of view limits one should select at each energy only geometry ranges (zenith angle, etc..) at which this effect is negligible (see [2] and [3] for further details).

Such cuts can be avoided, retaining a larger statistics, if the obtained primary fractions are corrected taking into account the reconstructed efficiencies for each primary mass. The reconstruction efficiency can be determined as the ratio between the total number of accepted events over the total number of generated events for the mass j, $\epsilon_j^{tot} = N_j^{acc}/N_j^{gen}$, and for a specific observable in a particular bin i of its distribution the number of expected events should be weighted by $\epsilon_{ji} = N_{ji}^{acc}/N_{ji}^{gen}$. The corrected fractions p_j^{corr} are $(p_j/\epsilon_j^{tot})/(\sum_{j=1}^{N_m} p_j/\epsilon_j^{tot})$.

VI. Method Performances

The described techniques have been tested on simulated samples of known composition. For different proton-iron mixing, N events have been randomly selected from proton and iron Monte Carlo data and the resulting samples have been analyzed. The whole procedure have been repeated many times. In Fig. 3, the mean value and the root mean square of the distribution of the difference between the reconstructed input fractions and the expected ones are shown for different mixtures of protons and iron CONEX showers, using QGSJETII-03 [9], fully simulated through the Auger detector with the Auger analysis framework [10]. The input abundances are well reproduced by the methods in all cases.

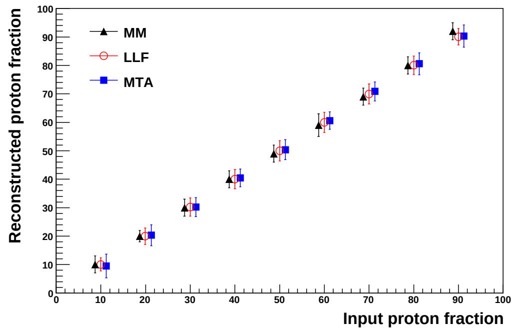

Fig. 3. Reconstructed primary fractions with the MM (full triangles), LLF (empty circles) and MTA (full squares) for different mixtures of protons and iron CONEX showers, with QGSJETII-03, fully simulated through the Auger detector.

VII. Mean X_{max} estimation from the mass composition analysis

The aim of the described techinques is to derive directly the primary composition of the observed cosmic ray flux. Of course, the obtained primary fractions depend on the hadronic interaction model adopted. Since the study of all the systematics introduced by the models and shower simulation and reconstruction are still under way, we don't report any composition results at this stage. We limit ourselves at checking the consistency of the composition obtained by the different approaches and their ability to reproduce mass sensitive observables.

The change of the mean X_{max} with energy (elongation rate) depends on the primary composition and it is measured directly from fluorescence detectors as at the Auger Observatory. From the primary fractions obtained by the mass composition methods, one can easily derive the mean X_{max} corresponding to the reconstructed mixture. The comparison allows to test if the mass analyses reproduce all the measured elongation rate structures and to have an independent cross-check of the effectiveness of the anti-bias cuts discussed in [2] and [3].

All the hybrid data collected by the Auger Observatory between 1st of December 2004 and the 30th of April 2007 (reported in [2]) have been analyzed with the described mass composition techniques. The studies have been done with a large sample of CONEX simulated showers of protons and iron nuclei, produced with QGSJETII-03. The set has been processed with the Auger analysis framework taking into account the detector evolution with time and the exact working conditions, as done by the Auger Collaboration to compute the hybrid exposure of the Auger Observatory [11]. The analysis has been done above 10^{18} eV as the hybrid detector trigger (an Fd event in coincidence with at least one SD station) becomes full efficient both for protons and irons primaries [11].

In each energy bin, the mean X_{max} is given by

$$\langle X_{max} \rangle = P_p \langle X_p \rangle + (1 - P_p)\langle X_{Fe} \rangle \qquad (11)$$

where P_p is the reconstructed proton fraction, while $\langle X_p \rangle$ and $\langle X_{Fe} \rangle$ are the expected mean X_{max} values for proton and iron nuclei. In Fig. 4 the mean X_{max} as a function of energy obtained from the composition results of LLF and MTA, in terms of proton and iron fraction, is shown along with the measured curve [2]. The elongation rates estimated with the two techniques are in agreement with the measured one. All the observed features are well reproduced.

The reconstructed primary fractions obtained by a Monte Carlo based composition analysis are model dependent. To test if the methods could describe the measured elongation rate independently from the hadronic model used, the hybrid data set has been analyzed with a second set of CONEX simulated showers of proton and iron nuclei produced with Sibyll2.1 [12]. The mean X_{max} as a function of energy derived from MTA, in terms of proton and iron fraction with Sibyll2.1 (empty squares) is shown in Fig. 5 along with that one obtained with QGSJETII-03 (empty triangles). The change in the reconstructed primary fractions, due to a different hadronic model, is completely compensated by a change in the expected mean X_{max}, giving a compatible curve.

VIII. Conclusions

Typical performances of the techniques have been evaluated on Monte Carlo data. The input composition abundances are very well reproduced in all the cases.

The techniques have the great advantage to be not biased by the set of analysis cuts applied for the efficiency correction effects, then we can avoid very strong cuts and exploit a larger statistics.

The mass composition methods give primary consistent fractions that allow to reproduce the measured elongation rate reported by the Auger Collaboration at the ICRC 2007, independently from the hadronic model and from the applied set of analysis cuts. The comparison confirms the published Auger results with independent Monte Carlo techniques.

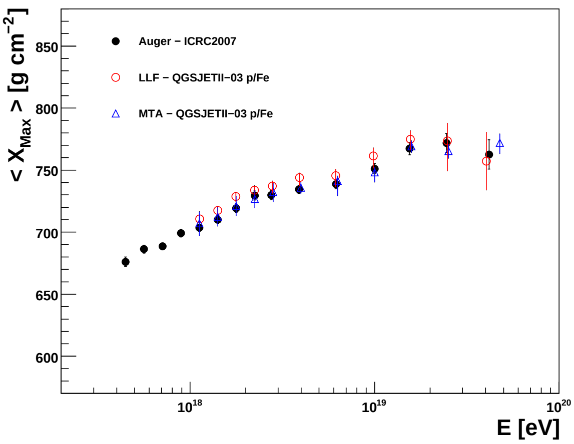

Fig. 4. Mean X_{max} as a function of the energy estimated from LLF (empty circles) and MTA (empty triangles) composition results, obtained using QGJETII-03, compared with that one measured by the Auger Collaboration [2].

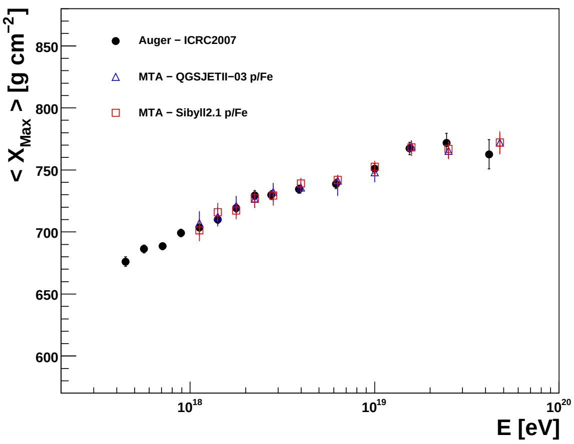

Fig. 5. Mean X_{max} as a function of the energy estimated from MTA primary fractions two different hadronic models: QGSJETII-03 (empty triangles) and Sibyll2.1 (empty squares).

References

[1] A.M. Hillas, arXiv:0607109 [astro-ph]; T. Stanev, Nucl. Phys. Proc. Suppl. **168**, (2007), 252; V. Berezinsky, Proc. 30th ICRC, Merida (2007), arXiv:0710.2750 [astro-ph]; C. De Donato and G.A. Medina-Tanco, arXiv:0807.4510 [astro-ph].
[2] M. Unger [Pierre Auger Collaboration], Proc. 30th ICRC, Merida (2007), arXiv:0706.1495 [astro-ph].
[3] J. Bellido [Pierre Auger Collaboration], these proceedings.
[4] M.D. Healy [Pierre Auger Collaboration], Proc. 30th ICRC, Merida (2007), arXiv:0706.1569 [astro-ph]; H. Wahlberg [Pierre Auger Collaboration], these proceedings.
[5] M. Ambrosio *et al.*, Astroparticle Physics **24**, (2005), 355; D. D'Urso *et al.*, Nucl. Instrum. Meth. **A588**, (2008) 171.
[6] T. Bergmann *et al.*, Astroparticle Physics **26**, (2007), 420.
[7] D. Heck *et al.*, Report FZKA 6019, 1998.
[8] A. Yushkov *et al.*, these proceedings.
[9] S. Ostapchenko, Nucl. Phys. B(Proc. Suppl.), **151**, (2006), 143.
[10] S. Argiro *et al.*, Nucl. Instr. Meth. **A580** (2007) 1485.
[11] F. Salamida [Pierre Auger Collaboration], these proceedings.
[12] R. Engel *et al.*, Proc. 26th ICRC, Salt Lake City (1999), 415.