# Neural networks as a statistic diagnostic tool for mass composition at the highest energies

Gustavo Medina Tanco,[1] Sergio Sciutto,[2] and Andre Tiba[1]
*(1) Instituto Astronomico e Geofisico, Univ. of Sao Paulo, CEP 05508-900, Brazil*
*(2) Departamento de Fisica, Universidad de La Plata, C. C. 67 - 1900 La Plata, Argentina.*

**Abstract**

The performance of a neural network based multivariate analysis for the determination of mass composition at the highest energies is studied. We use a simulation chain that includes the code AIRES plus a surface array simulator configured to emulate the Auger Southern Observatory. A very large set of more than 30,000 showers simulated with great detail is used in our analysis. A set of multiple observables is used as input for the neural network algorithm that is tuned for optimum determination of the primary composition. The most relevant characteristics of the method are analyzed, including its capability for hadronic model independent composition assignment.

## 1. Introduction

A wide range of theoretical models for the origin of Ultra High Energy Cosmic Rays have been explored in past years. From the more classical bottom-up models, in which nuclei are accelerated beyond $10^{19}$ eV, in different astrophysical environments to more audacious top-down scenarios in which either new particles or exotic phenomena are involved, such as decaying topological defects.

Testing this models requires an accurate determination of the energy spectrum, the distribution of arrival directions over the whole sky, and the identity of the particles.

The Auger Observatory, with its 3000 $km^2$ and hybrid capacity, has the potential to give appropriate answers to these questions.

The development of extensive air showers (EAS), as characterized by lateral distribution, curvature of the shock front, rising time, pulse shape, total number of photoelectrons, etc., carry information regarding the direction, energy and identity of the incoming primary. However, while direction and energy can be estimated rather easily from ground array data, the definition of a convenient and efficient diagnostic for primary identity discrimination remains a challenging issue.

The LPM effect modifies considerably the development of photons initiated

pp. 587–590 

EAS in some cases, and their muonic component is relatively small simplifying considerably the discrimination between photons from hadrons [5]. The separation among different hadrons is, however, much more difficult.

A recent effort by Ave and coworkers [6], for example, has described a new approach to estimate the mass composition beyond 1 EeV, based on measuring the ratio of vertical and inclined showers. Their conclusions are consistent with previous estimates in that above $10^{19}$ eV the spectrum is dominated by a lighter component.

Given the present uncertainties, results so far remain mostly qualitative and it is almost sure that such a complex problem will not be solved by the use of a single technique.

In this paper, a pragmatic approach is taken to the practical problem of statistically determining the identity of the primaries starting EAS at the top of the atmosphere with the ground array of the Auger observatory as the specific target. To this end, we employ neural networks (**NN**) trained to reproduce the simulated outputs of the Auger detector triggered by simulated showers for different nuclei ranging from protons to Fe.

## 2. Numerical method and discussion

A large sample of showers for several hadronic primaries is generated with the AIRES code and, transformed into ground array events of a model Auger observatory, used to trigger the surface detectors, simulated with the sample-sim code.

The AIRES system is a set of programs to produce simulations of air showers, and to analyze the corresponding data. All the relevant particles and interactions are taken into account during the simulations, and a number of observables are measured and recorded, among them, the longitudinal and lateral profiles of the showers, the arrival time distributions, and detailed lists of particles reaching ground that can be further processed by detector simulation programs. The AIRES system is explained in detail elsewhere [3,4].

The showers processed in this work were generated with the AIRES system, and consist in: (i) a series of 10,000 proton, He, C, O, Mg, Si, Ca, and Fe showers, with energies in the range $10^{17.5}$ eV to $10^{20.9}$ eV; (ii) a series of 25,000 proton and iron showers, with energies in the range $10^{17.5}$ eV to $10^{20.5}$. The zenith angles of the showers of both series range from 0 to 84 deg. All the simulations have been performed using high quality statistical sampling. Each shower is reused 20 times at different location in the array, and so the final number of available events is 700,000. The hadronic models used are QGSJET [2] and SIBYLL [1].

The surface detectors have been simulated using the "sample-sim" SD simulation program.

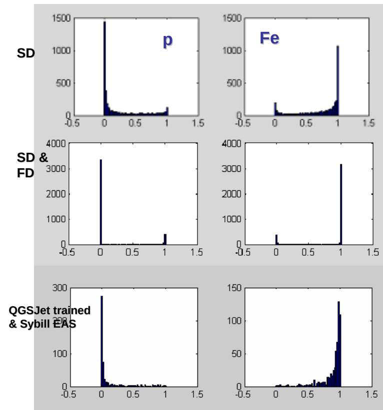


**Fig. 1.** Discrimination between protons and iron nuclei.

## 3. Results

The detector outputs have been used to build different parameter sets which include information on number of triggered stations, zenith angle, shower front shape, signal rise times (T10, T50 and T90), and lateral distribution. In some cases, $X_{max}$ has been included in order to stress the impact of hybrid events in the discrimination potential of the technique.

In figure 1, we show the much simpler case of discriminating between protons and hadrons. Different NN were trained to separate proton (output=0) from Fe (output=1). In the left panels only protons were injected, while Fe was injected at the right. The upper panel has only surface array information. The middle panels is for hybrid events, i.e., $X_{max}$ was included as a parameter. The lower panel corresponds to the same NN, trained with QGSJET and processing showers simulated with SIBYLL. It can be seen that NN can be very efficient at separating two nuclei even using only information from the surface array. The efficiency of the network is considerably increased, however, if fluorescence data is also included from hybrid events. Finally, the NN is reasonably independent of the hadronic interaction model used for training in this simpler case.

In figures 2.a and 2.b, we show results corresponding to two different NN trained to separate p, $He^4$, $C^{12}$, $O^{16}$, $Mg^{24}$, $Si^{28}$, $Ca^{40}$ and $Fe^{56}$.

No attempt has been made to separate individual events. We have not been able to design a NN able to do that. We have attempted only the statistical determination of the average atomic mass of different composition spectra. We arbitrary used simulated composition spectra of the form $dA/dN \propto A^{\nu}$, with $\nu = -0.5, 0.0, 0.5, 1.0 and 2.0$.

For each value of $\nu$, 1000 samples of 300 events each, different from those used for NN training, were used to calculate the distribution function of the

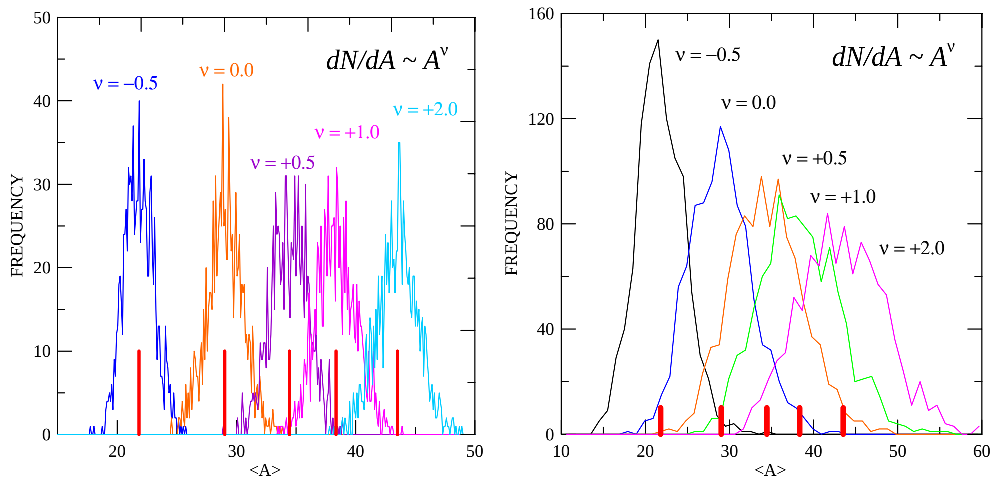


**Fig. 2.**

estimated average mass. The vertical thick lines are, the true value of the average mass for each $\nu$. Figure 2.a includes information on hybrid events. Figure 2.b relays only on the surface array.

In both cases several cuts had to be imposed for data preprocessing, which reduce the effective available number of events to roughly 25% of the available data set.

## 4. Conclusions

We show that NN can be a helpful tool for the discrimination of primary UHECR mass composition. The present results are exploratory and we don't make any hard claims about the bounties of the technique, specially because, at this stage, we have not employed events reconstructed with the actual codes that will be used during the run of the experiment.

GMT was partially supported by Brazilian agencies FAPESP and CNPq.

## 5. References

1. Engel R. Gaisser T. K. Stanev T. 1999, Proc. 26th ICRC(Utah) 1 415.
2. Kalmykov, N. et al. 1997, Nucl. Phys. B, **52B** 17.
3. Sciutto, S. J., Proc. 27th ICRC (Hamburg), **1**, 237, 2001.
4. Sciutto, S. J., preprint astro-ph/9911331, 1999.
5. Halzen et al. 1995, Astroparticle Physics **3** 151
6. M. Ave J.A. Hinton R.A. Vazquez A.A. Watson E. Zas 2003, Astropart.Phys. 18 367-375