The ANN method to measure proton spectra below the 'knee' region with the ARGO-YBJ experiment

S.Z Chen [1], M Zha[1], and H.H He[1] on behalf of ARGO-YBJ collaboration
(1) Key Laboratory of Particle Astrophysics, Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, Beijing, China

Abstract

Based on Monte Carlo Simulation, the space-time information of the shower is used to study the difference between different primaries collected by ARGO experiment. With one month exposure, the proton spectrum from 20 TeV to 200 TeV energy region can be obtained with the data sample of digital read-out selected by an artificial neural network. With the analog read-out data, the energy region could be extrapolated to 'knee' region.

1. Introduction

The energy of cosmic ray (CR) particles can exceed many orders of magnitude from GeV energies up to at least 10^{20} eV. The all particle energy spectrum has a steepening around several PeV, generally called the 'knee'. To interpret this feature, many approaches have been discussed from the origin, acceleration and propagation of CR [1]. But, the mechanism is still uncertain. The spectrum of single component is believed to be the key point to understand the 'knee' mechanism. But different experiments give different results. The Tibet $AS\gamma$ experiment thinks that the 'knee' is caused by the steepening of heavy nuclei and proton should steepen at about 100 TeV [2]. But according to KASCADE experiment unfolding result, it would be caused by light elements [3].

Operated at high altitude and full coverage, the ARGO-YBJ is unique in better understanding the 'knee' problem. Its low energy threshold provides a "bridge" between direct and indirect measurement. While, the detailed space-time information of shower can be used to discriminate different primaries. In this paper by using the artificial neural network(ANN) method we explore the feasibility of ARGO-YBJ to obtain the spectrum of different components.

2. The ARGO-YBJ detector

The ARGO-YBJ experiment [4] , located at the Yangbajing Tibet of China (4300 m a.s.l), consists of a full coverage of detector of area $78 \times 74\ m^2$ realized

pp. 1–4

with a single layer of RPCs. The guard ring enlarges the area up to 110×100 m^2. A cluster consists of 12 contiguous RPCs while one RPC is divided into 10 basic detection units, which are called PADs and each has 8 digital read-out strips. With the digital read-out ARGO reaches its energy upper limit at about 100 TeV, which can be extended to the 'knee' region by means of analog read-out (two BigPADs for each RPC). In the following, only the central carpet will be used and this central carpet is artificially divided into two parts: the internal detectors (the 6×9 clusters in the center) and the external detectors (the 76 clusters around).

3. Method

The general procedure for ANN method [5] is adopted here: with assuming of mixed composition of CR from some measurements, a large Monte Carlo simulation samples has been collected; we divide it to equally 2 parts, 50% of artificial samples are used to train the ANN, the other 50% is used to determine the energy spectrum.

3.1. Simulation

CORSIKA code [6] with QGSJET-II model is used to generate air showers induced by 5 primary chemical components: protons(P,1), helium(He,4), light nuclei(CNO, 7), median nuclei(Mg-Si, 13) and heavy nuclei(Fe,56). Their flux in the low energy region is got by fitting the direct measurement result. The best fit in poly-gonato model(rigidity dependent on bending point $Z \times 4.49$ PeV, constant difference $\delta\gamma = 2.10$ between the spectral indices below and above the knee) [7] is used here to extrapolation to higher energy region. After normalization at 10 TeV, the component fractions at 100 TeV are: P 25.5%, He 31.5%, CNO 19.5%, Mg-Si 10.4% and Fe 13.1%.

The incident zenith angles of primary particles are isotropically sampled within 20°. The energy region for all five groups is from 10 TeV to 10 PeV. The detector response has been simulated via a GEANT3-based code, and an array with 10×13 clusters, detector efficiency of 95%, noise 380 Hz/PAD and sample area 100×100 m^2 are taken into account.

3.2. Data analysis

In order to obtain a good data sample for the following data analysis, the events with core located inside the internal detector are selected by 2 conditions: (1)the particle density of the internal 54 Clusters is 1.1 times higher than that of the external 74 clusters; (2)the most fired Module(consists of 3 RPCs) is internal; (3) the core is inside a fiducial area 50×55 m^2 at the center. After these cuts the efficiency to select internal P and Fe events are 99.3% and 95.7% respectively. In

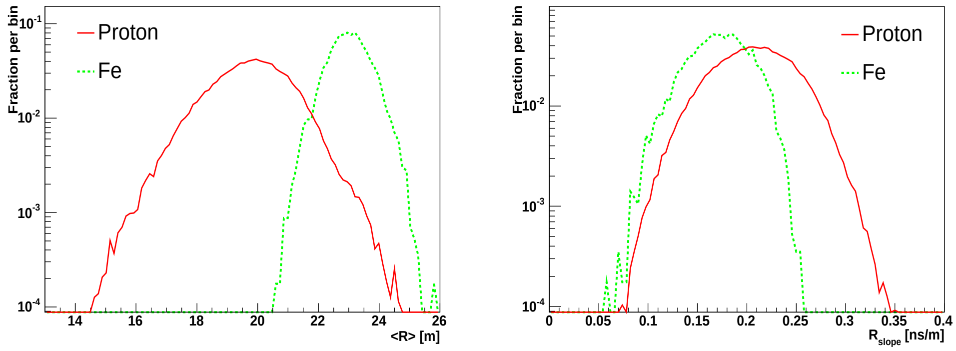

Fig. 1. The difference between proton and Fe in parameters $< R >$ and R_{slope}.

the end about 2.25 million simulation internal events are got, corresponding to a 24 days exposure of ARGO experiment.

Each event obtained above can be characterized by four parameters:

- The number of total fired strips N_{strip};
- $< R > (= \Sigma(N_i \times R_i)/\Sigma N_i)$: mean lateral spread radius of particle flow from the shower core position. N_i is the number of fired strips of the ith detector, and R_i is the distance to the core;
- N_{module}: the number of modules with fired strips more than 3;
- R_{slope}: the slope of the conical shower time front refer to the planar fit.

The proton-induced events can be characterized by small lateral spread and large conical slope, while those induced by heavy nuclei have the opposite character as their early development in the atmosphere. The difference between proton and Fe in parameters $< R >$ and R_{slope} are shown in figure 1. With these four input parameters, a $4 : 10 : 1$ network is trained to discriminate the signal (proton-induced events) from the background with a half data sample. After 300 epochs the learning of the network becomes very stable and the network is able to select proton induced from the other half data sample as shown in figure 2. When the cut value of ANN output is set to 0.2 the fraction of correct classification is 90%.

3.3. Result

The number of fired strips N_{strip} is in direct proportion to primary energy E_0 when the detectors is not saturated, thus the primary energy of each event can be estimated by the number of fired strips. The relationship between N_{strip} and E_0 got in simulation is:

$$log_{10}(E_0) = 0.448 + 1.058 log_{10}(N_{strip}). \quad (1)$$

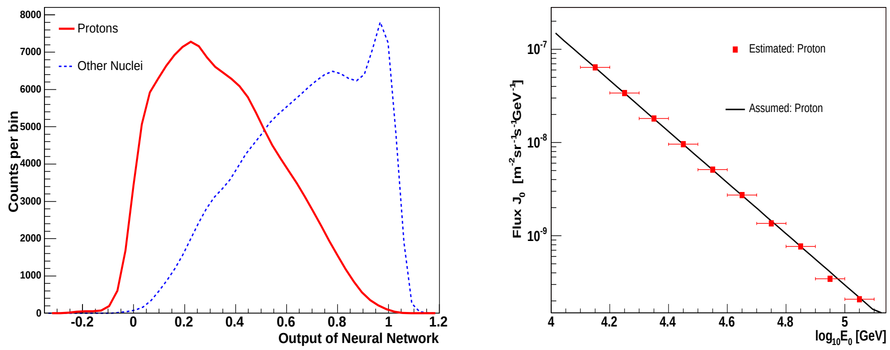

Fig. 2. Left: ANN output distribution of the proton-induced showers and other nuclei induced showers respectively. Right: Comparison between the assumed and estimated spectrum of protons. Only statistic error is taken into account here.

The uncertainty in this estimation is about 30%. With effective collecting area and ANN efficiency in selecting proton, the spectra of the proton can be determined and the result is shown in figure 2.

4. Discussion

Based on the result obtained above and assuming one month operation, a good agreement between the assumed and obtained spectra is seen in the energy region from 20 TeV to 200 TeV . This means that the ANN is effective to obtain the proton spectrum for ARGO experiment. The spectrum of other nuclei, like iron and helium, could be obtained with this method in principle. So, to get the spectrum of other nuclei and using the trained ANN to analyze the experiment data is our future plan. With the analog read-out data, the energy could be extrapolated to the 'knee' region.

References

[1] J.R Hörandel, Astropart. Phys. 21(2004) 241-265

[2] M. Amenomori, Phys Rev.D 62, 112002

[3] T.Antoni, W.D.Apel, Astropart. Phys. 24(2005) 1-25

[4] G.Aielli, Nucl. Phys.B (Proc. Suppl.) 166 (2007) 96-102

[5] X.W Xu, J. Phys.G 29(2003)719-735

[6] D. Heck et al., Report FZKA 6019, Forschungszentrum Karlsruhe, 1998.

[7] J.R Hörandel, Astropart. Phys. 19(2003) 193