Estimation of the proton energy spectrum in knee region by analog read-out of ARGO-YBJ experiment*

QU Xiao-Bo(曲晓波)[1] CHEN Song-Zhan(陈松战)[2;1)] ZHA Min(查敏)[2]
ZHANG Xue-Yao(张学尧)[1] FENG Cun-Feng(冯存峰)[1]

1 (School of Physics, Shandong University, Ji'nan 250100, China)
2 (Institute of High Energy Physics, CAS, Beijing 100049, China)

Abstract Based on the six months data set of ARGO-YBJ experiment with analog read-out and its Monte Carlo simulation, we study the difference between different primaries induced showers by using the space-time information of the charged particles in Extensive Air Showers. With five parameters which can efficiently pick out primary proton induced showers as inputs of an artificial neural network, the proton spectrum from 100 TeV to 10 PeV can be obtained.

Key words analog read-out, "knee" region, proton energy spectrum

PACS 95.85.Ry, 02.50.Sk, 07.05.Mh

1 Introduction

The all-particle energy spectrum of cosmic rays shows a distinctive feature around several PeV, known as the "knee", where the spectral index of the power-law dependence changes from −2.7 to approximately −3.1[1]. Existence of this feature has been well established experimentally, but there still remain controversial arguments on its origin. To explain this feature, several mechanisms have been proposed[2]. In some of these theoretical models, it is believed that the knee is an intrinsic property of the energy spectrum, related to the acceleration and propagation of the cosmic ray. While in other models, the knee is explained as a new type of interaction at very high energy. All of these theoretical models are still under discussion due to lack of detailed knowledge about the chemical composition around the knee. Several ground based experiments have measured the energy spectrum of some components of the cosmic ray in the knee region, but due to the limited ability of identifying the primary particles and the limited statistics, the experimental results are still inconsistent:the Tibet AS-γ experiment shows that the knee is caused by the steepening of heavy nuclei spectrum in this region and proton spectrum should steepen at about 100 TeV[3]. But according to the KASCADE experimental result, it should be caused by light elements[4]. So a clear proton spectrum in the knee region is the key to a definite conclusion of this problem and it can also give a limitation for theoretical models.

To identify the individual components of the primary cosmic rays, the ground based experiments are required to give sufficient information on the Extensive Air Shower (EAS) produced by these high energy particles. The ARGO-YBJ experiment, located at Yangbajing, Tibet (4370m.a.s.l), utilizes a full coverage detector array to detect the EAS. Its good time resolution and fine space granularity enable it to get enough information on air showers, which can be used to discriminate between different primaries. In addition, at the observation level of this experiment, the EAS induced by cosmic ray particles with energies in the knee region reaches a maximum development, irrespective of the primary mass, so the shower size is less fluctuated and the energy determination is more precise and less dependent upon the unknown composition.

Received 18 January 2008, Revised 16 June 2008
* Supported by National Natural Science Foundation of China (10120130794)
1) E-mail: chensz@mail.ihep.ac.cn

In this work we study the feasibility of identifying the primary cosmic ray particles with energies in the knee region by using the Monte Carlo generated data. By inspecting the features of the lateral particle distribution in EAS produced by different primary particles, we find 5 characteristic parameters which could be used to discriminate between different primaries. With these parameters as inputs, multivariate analysis is performed by using the artificial neural network (ANN) method. The influence of hadronic interaction model is estimated via a detailed comparison between the results of two hadronic interaction models: QGSJET-II and SIBYLL. Through these analyses, we find that, with weak model dependence, the ANN method can efficiently pick out the proton induced events from others.

2 The ARGO-YBJ detector

The ARGO-YBJ detector[5] is located in Tibet (P.R. China) at the Yangbajing Cosmic Ray Observatory. It consists of a full coverage array of dimension 74 m×78 m, realized with a single layer of RPC (Resistive Plate Chamber), 280 cm × 125 cm each. The area surrounding this central carpet, up to 110 m×100 m, will be partially instrumented with RPCs (the guard ring). The RPCs are grouped into clusters (Each cluster consists of 12 contiguous RPCs). Each RPC is divided into 10 basic detection units (called PADs, each with 8 digital readout strips). In order to extend the dynamic range, a charge read-out layer has been implemented by instrumenting each RPC with two large size pads, 140 cm×125 cm each (the so-called Big Pad). In this paper, only the central carpet will be simulated 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). See Fig. 1.

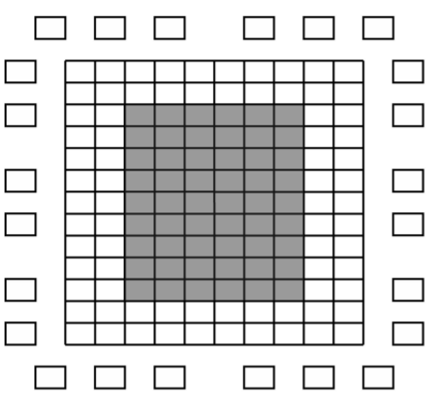

Fig. 1. ARGO-YBJ carpet detector. (one pane represents one cluster. The 6×9 clusters in the center(marked) are the internal detectors. For details, see the text).

3 Monte Carlo simulation and data selection

In order to estimate the influence of hadronic interaction model, the CORSIKA code[6] with QGSJET-II and SIBYLL as the hadronic interaction model is used to generate the EAS respectively. Following the result of Ref. [7], we have initiated the EAS with high energy protons (P,1), helium (He,4), light nuclei (CNO,7), median nuclei (Mg-Si,13) and iron nuclei (Fe,56) with energies ranging from 100 TeV to 10 PeV. The incident zenith angles of primary particles are isotropically sampled within 20°. The secondary particles in the EAS are traced until their energies are below 300 MeV for hadrons and 3 MeV for electromagnetic particles. The energy spectrum of each primary type is given by the Poly-gonato model[8] with rigidity dependent on bending point $Z\times4.49$ PeV and constant difference $\delta_\gamma=2.10$ between the spectral indices below and above the knee. From this model, the component fractions in the energy range of 100 TeV to 10 PeV are: P: 24.6%, He: 32.2%, CNO: 20.3%, Mg-Si: 9.8% and Fe: 13.1% respectively. The number of each component is generated according to these fractions.

The detector response is simulated using a GEANT3-based ARGO detector simulation program. The effects of geometry and the material of the detector are taken into account in the simulation. The charged particles reaching the Big Pad are recorded, and the lateral charged particle distribution is then gotten for each EAS event. The shower core position is randomly selected in a sampling area of 100 m × 100 m, which is big enough for the following data selection criteria. We get about 3.78×10^6 QGSJET-II events and 3.78×10^6 SIBYLL events.

In order to get enough information around the shower core and reduce the error caused by the shower core reconstruction, only events with the core located inside the internal detector are used for further analysis. These internal events are selected by using the following 3 criteria:

(1) The particle density of the internal 54 clusters is 1.11 times higher than the external 76 clusters;

(2) The module (consists of 3 RPCs) with largest number of charged particles is internal;

(3) The reconstructed core is inside a fiducial area of 50 m × 55 m around the center of the detector array.

The efficiencies of internal event selection for proton and iron nuclei induced showers are 97.7% and 92.3% respectively. We do a test simulation for those showers with core located beyond 100 m × 100 m but in 200 m × 200 m, after the selection, only 0.02% can come through. So 100 m × 100 m is big enough for the internal data selection.

4 Analysis and results

4.1 Identification of proton induced showers

After examining the behaviors of the lateral charged particle distribution of the EAS induced by different primaries, we found the following 5 parameters can be used to characterize the difference between proton and other nuclei induced showers:

1) $N_{\rm Hits}$: The total number of charged particles recorded by the big pad.

2) $\langle R\rangle$: $(=\sum(N_i\times R_i)/\sum N_i)$ Mean lateral spread radius of particle flow from the shower core position. N_i is the number of charged particles detected by the ith big pad. R_i is the distance between the ith big pad and the shower core.

3) R_{80}: The radius of the minimum circle which contains 80 percent of the total charged particles detected by the central carpet.

4) $S_{\rm front}$: The slope of the conical shower time front refers to the planar fit.

5) $\rm Ratio_{80}$: The ratio of core region's particle density to R_{80} region's. The core region is defined as a circle area centered at the shower core with a radius of 2.5 meters. The shower image is divided into concentric rings centered at the shower core with a width of 5 meters in radius. The ring with R_{80} in it is defined as R_{80} region.

The distributions of $\langle R\rangle$, R_{80}, $S_{\rm front}$ and $\rm Ratio_{80}$ are shown in Fig. 2. From this figure we can see that the proton induced events have relatively smaller $\langle R\rangle$ and R_{80}, larger $S_{\rm front}$ and $\rm Ratio_{80}$ compared with the events produced by other nuclei.

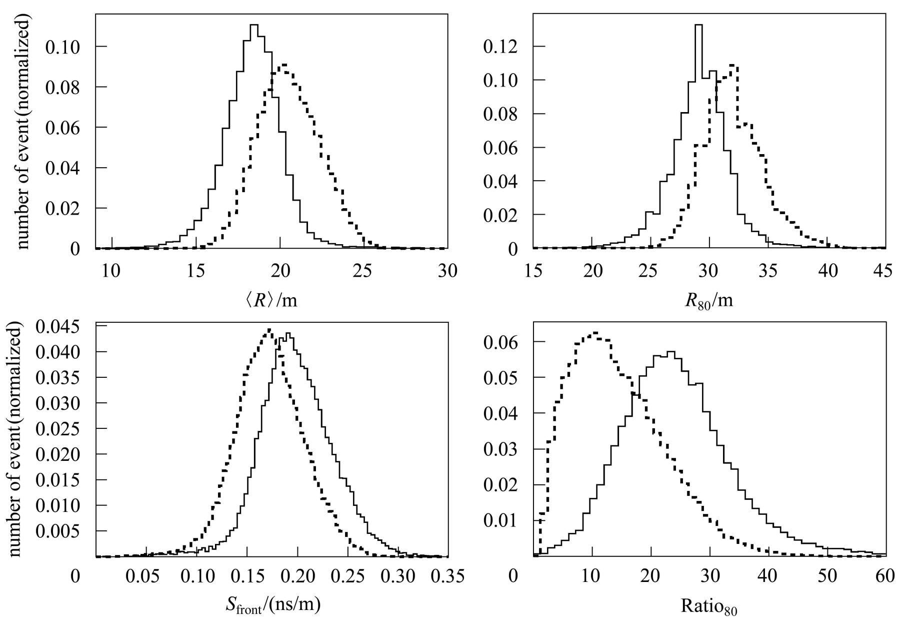

Fig. 2. Parameters distribution. The solid line for proton induced events, the dash line for other nuclei induced events. Up: The left is $\langle R\rangle$, the right is R_{80}. Down: The left is $S_{\rm front}$, the right is $\rm Ratio_{80}$.

To discriminate the proton induced showers from those induced by other nuclei, multivariate analysis was performed by using artificial neuron network (ANN) method. Feed forward ANN is created with 3 perceptron layers. The input layers consist of 5 neurons corresponding to the 5 parameters introduced above. The hidden layer has 15 neurons and the output layer has one neuron. The ANN is implemented and optimized by using the TMVA toolkit[9]. The selected Monte Carlo event samples are divided into two halves randomly: one for training the network and the other for testing the network. The target values for protons and others are put as unity and zero, respectively. After 700 epochs during the training process, the network becomes very stable. Then this network is used to select the proton induced events from the other half data sample. For the purpose of checking the influence of hadronic interaction model, two ANN are trained and tested by QGSJET-II data and SIBYLL data respectively, and the cross-examination is also done (test the QGSJET-II ANN with SIBYLL data and vice versa). The ANN outputs are shown in

Fig. 3.

The cut value of ANN output is set to 0.75, it means that the showers with ANN outputs greater than 0.75 are considered to be proton events. The ANN efficiency in this cut situation is shown in Table 1. As seen in Table 1, the dependence of the efficiency on hadronic interaction models is weak, and the primary proton events can be effectively identified from other events by using the ANN method.

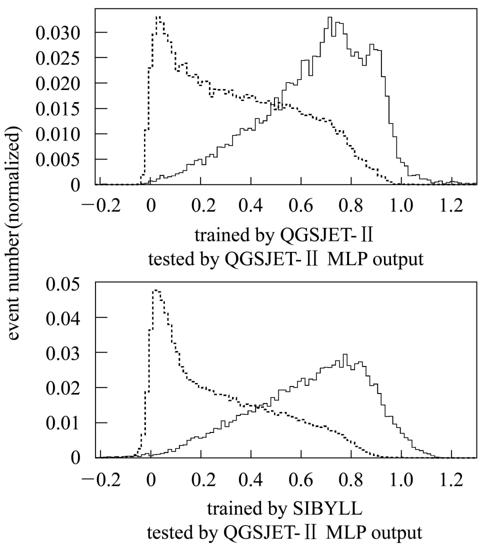

Fig. 3. ANN output distribution. Up: The result of the ANN trained and tested by QGSJET-II data. Down: The result of the ANN trained by SIBYLL data and tested by QGSJET-II data. The solid line for proton induced events(signal), the dash line for other nuclei induced events(background).

Table 1. The ANN efficiency.

data for train	data for test	fraction of picked proton	rejection ratio of other nuclei
QGSJET-II	QGSJET-II	41.1%	93.5%
	SIBYLL	40.3%	89.8%
SIBYLL	QGSJET-II	39.3%	95.6%
	SIBYLL	39.8%	93.1%

4.2 Primary energy reconstruction

The number of charged particles recorded by the charge read-out layer ($N_{\rm hits}$) can be used to estimate the primary energy of each EAS event. Fig. 4. shows the scatter plots of the primary energy E_0 versus $N_{\rm hits}$. The relationship between $N_{\rm hits}$ and E_0 can be represented by the following equation:

$$\log_{10}(E_0)=1.28+0.88\log_{10}(N_{\rm hits}) \ . \qquad (1)$$

The uncertainty in primary energy estimation using this equation is about 27%.

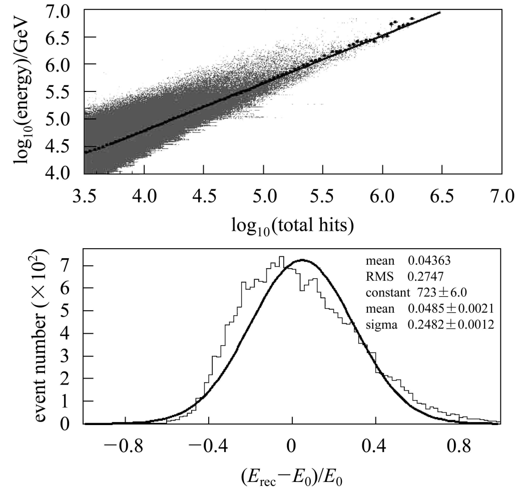

Fig. 4. Up: The relationship between primary energy and the number of total charged particles. The solid line is a fit using equation (1). Down: The energy resolution. ($E_{\rm rec}$ is reconstructed energy, and E_0 is primary energy.).

4.3 The estimation of proton energy spectrum

After estimating the showers' energy, we can get the integral flux between E_1 and E_2 with this formula:

$$J_{E_1-E_2}=N_{\rm events}/(T_{\rm eff}\Omega A_{\rm eff}) \ .$$

$N_{\rm events}$ is the number of events with the energy between E_1 and E_2 and comes through the effective area $A_{\rm eff}$ within the solid angle Ω during the effective time $T_{\rm eff}$. In this work $T_{\rm eff}$ is six months,and we can get the effective area $A_{\rm eff}$ with this formula:

$$A_{\rm eff}=\frac{N_{\rm internal}}{N_{\rm all}}A_{\rm g} \ .$$

$A_{\rm g}$ is the sampling area of 100 m × 100 m, $N_{\rm all}$ is the total number of events with the shower core located in the sampling area, $N_{\rm internal}$ is the number of internal events which are selected by the criteria mentioned in Section 3. The effective area is calculated by using 1.35×10^5 Monte Carlo simulation data at several fixed energies. The result is shown in Fig. 5.

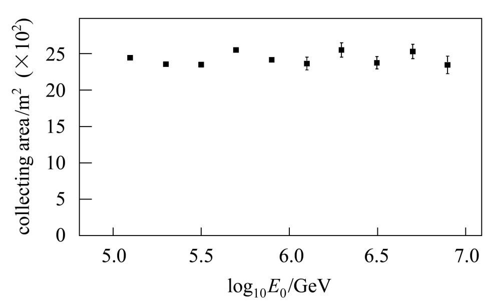

Fig. 5. The effective area as a function of energy.

After counting the ANN efficiency in selecting proton, energy reconstruction and effective collecting area, the spectra of proton in knee region can be obtained and the result is shown in Fig. 6. In this work, two ANNs trained by QGSJET-II data and SIBYLL data are used to pick out the proton events in QGSJET-II data sample respectively. Good agreement between the assumed and two estimated proton energy spectra is seen in the energy region from 100 TeV to 10 PeV. Fitting with two spectral index combination: Below the "knee", the spectral index of the power-law dependence in simulation is -2.74, and the estimated spectral index is -2.74 ± 0.013. Above the "knee", the spectral index of the power-law dependence in simulation is -4.84, and the estimated spectral index is -4.98 ± 0.72, and $\chi^2/ndf=33.41/8$; Fitting with one spectral index: the estimated spectral index is -2.81 ± 0.012, and $\chi^2/ndf=57.68/8$. So we fit the spectrum by using the former method (with two spectral index combination).

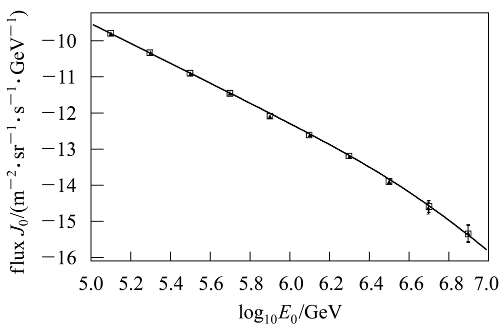

Fig. 6. Comparison between the assumed and two estimated spectra of proton. The solid line for assumed spectrum, the hollow squares for spectrum estimated by QGSJET-II ANN, the solid triangle for spectrum estimated by SIBYLL ANN.

5 Summary and discussion

ARGO-YBJ experiment with analog read-out can give detailed measurement of the lateral charged particle distributions in EAS induced by primary particles with the energy in the knee region. This experimental information can be used to separate the proton induced showers from nuclei induced ones. Through careful study of the behavior of the lateral distribution of the secondary particles, we get five parameters which can be used to characterize the difference between proton induced showers and the others. A multivariate analysis is done by using ANN method and the influence of hadronic interaction model is estimated. Via the ANN method, proton induced showers can be effectively selected and a good agreement between the assumed and estimated proton energy spectrum is seen in the energy region from 100 TeV to 10 PeV. Below the "knee", the spectral index of the power-law dependence in simulation is -2.74, and the estimated spectral index is -2.74 ± 0.013. Above the "knee", the spectral index of the power-law dependence in simulation is -4.84, and the estimated spectral index is -4.98 ± 0.72. In principle, the spectrum of other nuclei in the knee region, such as iron and helium, could be obtained with this method.

There are still a lot of factors which need to be taken into account while measuring the proton energy spectrum in the knee region, such as more practical simulations for big pad, more models in the simulation of hadronic interaction and more assumptions of mixed composition of cosmic ray in knee region. Checking out these factors' effect is our future plan. Every situation needs a lot of simulations. There is still a long way to go.

References

1 Haungs A et al. Acta Phys. Polon B, 2004, **35**: 331—339
2 Horandel J R. Astroparticle Physics, 2004, **21**: 241—265
3 Amenomori M et al. Physics Letters B, 2006, **632**: 58—64
4 Antoni T et al. Astroparticle Physics, 2005, **24**: 1—25
5 Aielli G et al. Nuclear Physics B, 2007, **166**(Proc. Suppl.): 96—102
6 Heck D et al. Report FZKA 6019, Forschungszentrum Karlsruhe, 1998
7 Gaisser T K, Honda M. Ann. Rev. Nucl. Part. Sci., 2002, **52**: 153—199
8 Horandel J R. Astroparticle Physics, 2003, **19**: 193—220
9 Hocker A et al. E-print arXiv: physics/0703039