Studying Cosmic Ray Composition around the knee region with the ANTARES Telescope.

CHING-CHENG HSU [1]
[1]*NIKHEF*

cchsu@nikhef.nl
DOI: 10.7529/ICRC2011/V04/0679

Abstract: The composition of the cosmic rays in the "knee" region ($\approx 10^4$ TeV/nucleus) of the all particle spectrum is considered to be the result of the particle acceleration and propagation from the astrophysical sources. The steeply falling CR spectrum makes a direct measurement of the composition difficult, but it can be inferred from the measurements of the showers generated by the interaction of the primary cosmic ray with the Earth atmosphere. In particular the characteristics of the muon bundles produced in the showers depend on the primary CR nature. The ANTARES telescope is situated 2.5 km under the Mediterranean Sea off the coast of Toulon, France. It is taking data in its complete configuration since May 2008 with nearly 900 photomultipliers installed on 12 lines. The trigger rate is a few Hz dominated by atmospheric muons. A method using a multiple layered neural network as a classifier was developed to estimate the relative contribution of proton and iron showers to the CR spectrum from the energy and multiplicity distribution of the muon tracks reaching the ANTARES detector. The performance of the method estimated from simulation will be discussed.

Keywords: Cosmic Ray, Knee, Composition, Antares

1 Introduction

Although the main goal of ANTARES telescope is to look for high energy neutrinos coming from the deep space, it also provides us opportunities to study cosmic ray physics. One of the most important topics is to distinguish the different chemical compositions around knee region of its spectrum. The cosmic ray spectrum is known as a power law with power index about -2.7 up to few PeV. Then the slope changes to -3.1 until the energy around 4×10^{18} eV [1]. The *knee* was first observed by the MSU group in 1970s, then has been confirmed by many groups afterward. The origin of the *knee* could be generally summarized into either astrophysical origin or particle physics origin. Nevertheless it is still a puzzle and generally believed to be a key issue to the problem of the origin of galactic cosmic rays.

Most people attributed the knee to the sudden reduction in Galactic trapping efficiency. A popular explanation is that the knee is associated with an upper limit of acceleration energy by galactic supernovae. Another popular scenario is the leakage of particles from the Galaxy, since the Larmor radius of a proton in the galactic magnetic field increases with its energy and finally exceeds the thickness of the galactic disk. Additionally, there are a minority of theorists who proposed that the knee is due to a single, recent and local supernova remnant (SNR) or a rapidly rotating pulsar interacting with radiation from its parent SNR.[2]

If the knee is caused by the maximum energy attained during the acceleration process or it is due to leakage from the Galaxy, the energy spectra for individual elements with charge Z would exhibit a cut-off at an energy E_c^Z = Z $\times$ E_c^p , where E_c^p is the cut-off energy of protons. The sum of the flux of all elements with their individual cut-off makes up the all-particle spectrum. In this picture the knee is related to the proton cut-off and the steeper spectrum above the knee is a consequence of the subsequent cut-off of heavier elements, resulting in a relatively smooth spectrum above the knee [3].

2 Composition Model

In the so-called polygonato model, the general form for the flux of primary nuclei of charge Z and energy E_0 is

$$\frac{d\phi_Z}{dE_0} = \phi_Z^0[1 + (\frac{E_0}{E_{trans}})^{\epsilon_c}]^{\frac{-\Delta\gamma}{\epsilon_c}}$$

where the transition energy E_{trans} could be determined according to three different scenarios. The parameter ϵ_c determines the smoothness of the transition, and γ_c is the hypothetical slope beyond the knee. E_{trans} is the cut-off energy. Under the polygonato model: three different scenarios are proposed :

- Rigidity dependent: From the astrophysical point of view, a rigidity dependent cut-off E_{trans} = $\hat{E}p \cdot$ Z is the most likely description if we take into account the acceleration and propagation of the cosmic rays.

- Mass dependent: This model predicts that the change in the power law index depends on the mass E_{trans} = $\hat{E}p \times$ A instead of the charge. This scenario leads to a steeper energy spectrum after the cut-off. The sharp cut-off would be hard to explain on astrophysical reasons. Maybe a nearby source or a new type of interaction in the atmosphere could yield such cut-off [3].

- Constant Composition: E_{trans} = $\hat{E}p$. The knee is explained by a common steepening in the energy spectrum; it occurs for all the particles at the same energy.

So far, the best measurement of the compositions around the knee region was done by KASCADE [4]. The measured primary energy spectra show that the knee in the all particle spectrum is due to a steepening of the light elements spectra.

3 Analysis Strategies

The ANTARES detector is located at 40 km off the coast of Toulon, France, at a depth of 2475 m in the Mediterranean Sea. It consists of 12 exible strings, each with a total height of 450 m, separated by about 60 m. They are anchored to the sea bed and kept near vertical by buoys at the top of the strings. Each string carries a total of 75 10-inch Hamamatsu photo multipliers (PMTs) housed in glass spheres, the so-called optical modules (OM) [5]. The OMs are arranged in 25 storeys (three optical modules per storey) separated by 14.5 m. The detector was starting taking data in 2007 and was fully completed in May 2009.
Since ANTARES is deeply buried under the sea, only the muon components from the air shower will survive at detector level. The muons will emit Cherenkov radiation only when passing through the sea water, which can be detected using photomultiplier tubes. To be registered by the ANTARES detector, muons have to travel at least 2.5 km of sea water and still be energetic enough to trigger the detector. The energy threshold for vertical down-going muon is around 500 GeV. At large zenith angles, the threshold increases because of the increasing depth of the sea water. The muon bundles properties (such as multiplicity) is strongly related to the primary energy and species of the nuclei. However, the ANTARES detector cannot resolve individual muons from a muon bundle. Hopefully, the topology of the hit distributions in space and time could give pertinent information about the properties of the muons in the bundles. Two useful analysis methods are combined and then applied on MC samples in this analysis. The first one is a cluster finding algorithm, and the second one is an electromagnetic shower searching algorithm[6].
In both analysis methods, we rotate our coordinate system such that the z axis is along the reconstructed muon track axis. We define the plane which is passing through the detector center and perpendicular to the reconstructed track

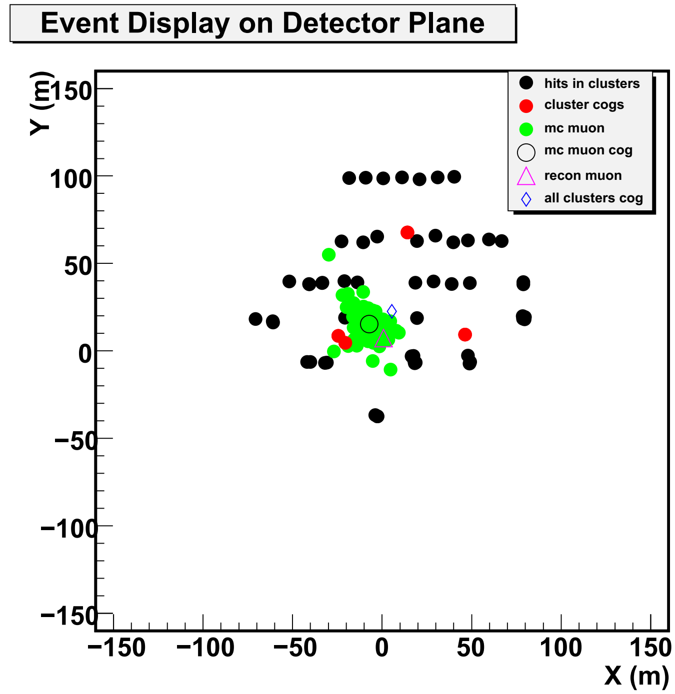

Figure 1: Typical patterns produced by muons bundles originating from primary protons(left) and iron nuclei with the same energy of muon bundle pattern on the detector plane.

as *detection plane*. Fig. 1 shows the snapshot of muon bundle patterns on the detector plane from the same primary energy of proton and iron nuclei. The projections of all the hits positions on the detection plane are also calculated, including the time correction information.
In the cluster finding algorithm, the clusters are formed if the hits fulfill the following three conditions:

- $\mid T_i$ - $T_j \mid \leq \mid r_i$ - $r_j \mid$ /C/n_g where T_i and T_j are the time for any two hit pairs in the hit cluster. r_i and r_j are the associated positions of these two hits.

- Any two hits within clusters should be in the same or neighboring strings or floors.

- $\mid T_i$ - $T_j \mid \leq \mid (z_i$ - $z_j) \mid$/C + d $\times$ $tan\theta_c$ /c + T_{ext}, where T_{ext} are maximum extra time, here we set 20 ns; T_i and T_j are again time informations of two hits. z_i and z_j are the rotated Z positions of two hits.

In order to quantify the cluster patterns from cosmic showers initiated by different groups of elements, we try to parametrize the hit patterns. There are two kinds of parameters in our analysis:

- Cluster-wise parameters: The parameters which are related to each individual clusters, for instances, the

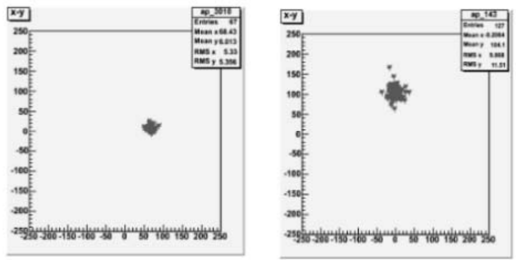

Figure 2: The hit patterns projection on the detector plane and then parametrized by cluster algorithm.

"Size" (total number of photoelectrons in one cluster), the Area: the area which contains 90 % of the hits...

- Event-wise parameters: The parameters which are related to all different clusters within one event, such as the Center of Gravity(C.O.G.) of all the clusters, X_{COG}, total number of clusters $N_{cluster}$ ...

In total, we have about 25 event-wise parameters and 12 cluster-wise parameters. One example of pattern after parametrization superimposed with reconstruction position of the muon tracks(bundles) is presented in Fig 2.

On the other hand, the ANTARES electromagnetic shower searching method gives us additional information about the muon bundles. The high energy muons suffer from catastrophic energy losses. Once it happens, electromagnetic showers are initiated either by γ or e^+ and e^- pairs. The radiation length of the electron is about 30 cm/gm^2. The spacing between optical modules and strings is large compared to the shower extent. Instead of projection hits information on the detector plane, we project all the hits on the reconstructed axis and search for the peak using the T-Spectrum function in ROOT package. The details of the algorithm could be found in [6]. Two more parameters are obtained by this methods. They are the number of showers N_{shower} and the amplitude of the baseline $N_{baseline}$ from the algorithm.

Combining the cluster-finding and shower-finding algorithms, we have in total 37 parameters. Each parameter gives different discrimination powers for the chemical species. To achieve multi-dimensional comparisons, we use the existing package "TMVA" (Toolkit for Multivariate Data Analysis with ROOT) for this analysis [8]. TMVA is a toolkit which hosts a large variety of multivariate classification algorithms. Training, testing, performance evaluation and application of all available classifiers is carried out simultaneously.

Several methods are implemented inside TMVA packages. In order to cross check the results, four different methods were chosen for this analysis. They are Multilayer Perception (MLP), (MLPBNN), (TMlpANN) and k-Nearest Neighbour (k-NN). The performances of multi-variate analysis are sensitive to correlation of the training parameters. Thus, to reduce the un-important parameters is necessary. We input each parameter into the analysis individually and then we kept only the top five rank of the parameters which gave us the most distinguished power. These five parameters are N_{hit}, $N_{cluster}$ and N_{npe}, representing number of hits, clusters, and photoelectrons from cluster-finding algorithm. N_{shower} and Base are the numbers of showers and the number of photoelectrons of the baseline from EM shower-finding algorithm.

Efficiency	P Purity	Fe Purity	P eff.	Fe eff.
30%-70%	0.7	0.66	0.68	0.77
1.8%-92.8%	0.76	0.62	0.61	0.81

Table 1: The estimated purity and rejecting efficiency for two different combinations of pseudo-data sets.

4 Analysis on MC samples

A full MC simulation was adopted in this analysis. The air showers induced by the primary nuclei with energy ranging from 1 to 10^5 TeV /nucleon and zenith angle between 0° and 85° using the CORSIKA software (Version 6.2) [7] and the hadronic interaction model QGSJET.01c. All muons reaching the sea level, with energies larger than threshold energy, are propagated through sea water to the detector. At last, muons are transported through the ANTARES sensitive volume, Cherenkov light is produced and the detector response is simulated. Background noises were added afterward. The trigger is done by standard ANTARES trigger. The muon direction and position are reconstructed using a multi-stage fitting procedure, which basically maximizes the likelihood of the observed hit times as a function of the muon direction and position[9].

We further divide the MC samples into 3 independent subsets for training, testing and evaluating in TMVA analysis. The relative sizes of the three subsets are 40%, 30% and 30%. Each event is with proper weight according to different spectrum models. The iron component is selected as source of signal and the rest of the four groups of elements are tagged as sources of background. The "pure" signal distributions are obtained, if we assume that the all coming cosmic rays are iron. On the other hand, the souces of the background samples are coming from the superposition of the rest of four groups of elemetns assuming all cosmic rays are from individual element. The distributions of N_{shower} and $N_{cluster}$ assuming the all particle spectrum are proton and iron are shown in Fig 3.

We carefully checked the output of neural network from test samples in order to avoid the over-training effects.

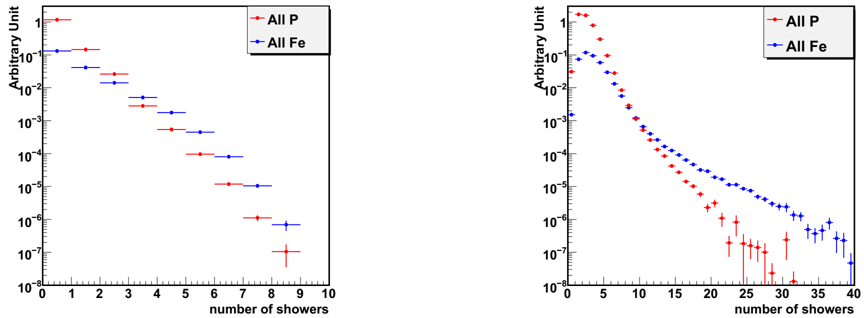

Figure 3: The relative distribution of N_{shower} and $N_{cluster}$ parameters after the reconstruction quality cut assuming that all the particles from cosmic rays are proton and iron.

5 Results and Conclusions

The training and test events fed into the neural network were subjected to a series of cuts. The main quality cut is the so-called Λ cut [9] in order to keep good quality on reconstructed events. For testing the method, we mixed proton and iron components from our pseudo-data set with $\approx$ 30%-70% and 1.8%-98.2%, individually. All the training, testing and psedo-data are applied on the same analysis chains and cuts. The output of neural network training on the proton and iron components is shown on the Fig 4. The green curve region (left) corresponds to the true proton events, while the red curve region (right) corresponds to the true iron events. The blue curve in the plot shows the psedo-data events with the 1.8%-98.2% configuration. The purity and rejecting efficiency for proton and iron are listed in the Table 1 if we cut on the MLP output value at 0.45, where the S/$\sqrt{S+B}$ is maximized. In addition, instead of applying a cut on the MLP output value. we adjust proton and iron distibutions as a whole to fit the psedo-data distribution ("Template fitting" in roofit). We found that the χ^2/ndf are 23.6/21 and 18.1/21 for the two configurations, respectively. The fitted numbers from the two-component model are satisfatory. In the future, we are planing to add more chemical compositions in the backgrounds and add magnesium in the source category. We have developed a method to estimate the ratio of the heavy elements in the triggered cosmic ray events based on the information from the muon tracks and electromagnetic showers in ANTARES detector. To combine all the discrimination powers from established multi-parameters needs the help from multi-variate analysis (neural network). The estimated ratio from heavy elements will be used as an input for calculating the true ratios between elements in original cosmic ray spectra. Further calculating the efficiencies from triggers, combinations of different cuts and effective areas are necessary and will be done. The analysis of ANTARES real data with the goal of deriving ratio between different groups of elements in the cosmic ray spectra is ongoing and will yield results in the near future.

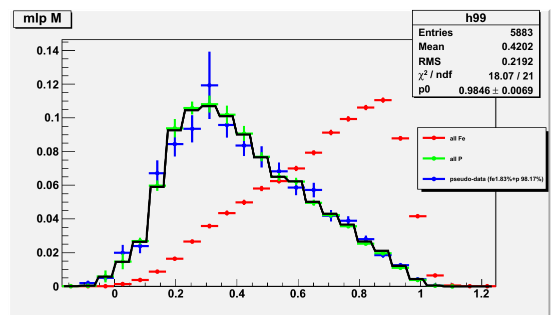

Figure 4: Output computed by the neural network in the presence of a mixture of proton and iron components from the pseudo-data sets. The green curve region (left) corresponds to the true proton events, while the red curve region (right) corresponds to the true iron events. The artificial mixture of pseudo-data is in blue.

References

[1] Bluemer, J. and Engel, R. and Hoerandel, J. R. Prog. Part. Nucl. Phys. 2009 **63**: 293-338

[2] Wibig, Tadeusz and Wolfendale, Arnold W. astro-ph/1012.4562 2010

[3] Hoerandel, J. R. Astropart. Phys. 2004 **21**: 241-265

[4] T. Antoni *et al.* [The KASCADE Collaboration] Astropart. Phys. **24** (2005) 1

[5] ANTARES Collaboration 2011 **astro-ph/1104.1607** :

[6] Mangano, Salvatore Nucl. Instrum. Meth. A **588** (2008) 107

[7] D. Heck et al. Forschungszentrum Karlsruhe Report FZKA, **6019** 2998

[8] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, and H. Voss, PoS ACAT **040** (2007), arXiv:physics/0703039

[9] A Heijboer. PhD Thesis, available from http://antares.in2p3.fr/Publications/index.html

[10] First Author et. al.: year, Book Name, ed. Editor, Publisher