PROCEEDINGS OF SCIENCE

Measurement of the Proton and Helium spectrum with KM2A and WFCTA of the LHAASO experiment

Liping Wang,[a,b,*] **Ling Ling Ma,**[a,b] **Shoushan Zhang,**[a,b] **Cunfeng Feng**[b] **and LHAASO collaboration**

[a] *Key Laboratory of Particle Astrophysics, Institute of High Energy Physics Chinese Academy of Sciences, 100049 Beijing, China*

[b] *Key Laboratory of Particle Physics and Particle Irradiation (MOE), Institute of Frontier and Interdisciplinary Science Shandong University, Qingdao, Shandong 266237, China*

E-mail: wanglp@ihep.ac.cn, llma@ihep.ac.cn, zhangss@ihep.ac.cn, fengcf@sdu.edu.cn

We report on the measurement with high statistics of the energy spectrum of light component (Proton plus Helium nuclei) in cosmic rays by Large High Altitude Air Shower Observatory (LHAASO) around the knee region. LHAASO is a composite cosmic ray observatory, which consists of three detector arrays, including the square meter Kilometer Array (KM2A), the Water Cherenkov Detector Array (WCDA), and the Wide Field of View Cherenkov Telescope Array (WFCTA). The LHAASO experiment with multiple types of detectors can achieve the multi-parameter measurement of the cosmic ray air shower, the parameters including N_0, $Dist_0$, which are sensitive to the component of the cosmic ray were defined and can be used for the mass separation. The data used in this work were taken from Nov 1, 2020, to Mar 31, 2021. During that period the LHAASO consisted of the first six WFCTA telescopes, the first half KM2A array, and the first water pool of WCDA. The analysis was performed using only information from combined observations of WFCTA and KM2A. The energy spectrum of the light component measurement process based on simulated data is reported in this paper.

38th International Cosmic Ray Conference (ICRC2023)
26 July - 3 August, 2023
Nagoya, Japan

ICRC2023
38th International Cosmic Ray Conference
The Astroparticle Physics Conference

*Speaker

PoS(ICRC2023)413

1. Introduction

Although cosmic rays have been discovered more than 100 years, their origin, acceleration, and propagation mechanisms still remain unclear. It is generally believed that galactic cosmic rays originated from supernova remnants because they can provide the required energy for cosmic ray acceleration. Recently, the observation of ultra-high energy gamma-ray sources by Large High Altitude Air Shower Observatory (LHAASO) indicates that the pulsar wind Nebula, young massive stars could also be the candidates of cosmic rays [1]. If the cosmic rays originated from these sources, the knees' position of individual elemental cosmic ray energy spectra should appear Z dependence, namely, the rigidity dependence. So the study of the cosmic ray energy spectra of individual components near the "knee" region can help us to understand the origin and acceleration mechanism of cosmic rays.

The flux of the cosmic rays drops rapidly with increasing energy, the study of high-energy cosmic rays with energies around the "knee" region is difficult for space-born experiments due to their limited effective area and mainly relies on measurements of the extensive air shower (EAS) induced by cosmic rays on the ground. Measurement of spectra for individual elements or mass groups is limited by the large intrinsic fluctuations of EAS observables. So far, many ground-based experiments have measured the energy spectrum of proton and helium, including KASCADE [2], Tibet AS-γ [3] and ARGO-YBJ [4]. However, their results are inconsistent with each other. One reason is that there is no cosmic ray beam with a given energy and component to calibrate their detectors. Another reason is that the selection efficiency may affect the energy spectrum shape of proton and helium.

In this paper, we report the ability to measure the cosmic ray components and energy spectrum based on the first half of KM2A and six WFCTA telescopes. The experiment setup and detectors are briefly described in Section 2. The Monte Carlo (MC) simulation is described in Section 3. In Section 4, the air shower reconstruction is described. In Section 5, the component sensitive parameters are defined based on the features of EAS and the toolkit for multivariate analysis (TMVA) is used, and the influences of selection efficiency on the expectation of proton and helium spectrum measurement are presented. And the conclusions are presented in Section 6

2. Experimental setup

LHAASO was formally approved on December 31, 2015 and was completed the construction in 2021. The combined observation of LHAASO can be maily divided into two stages to measure the energy spectra of the cosmic rays. The first stage with the aim to measure the light nuclei component (proton and helium) spectrum or proton spectrum with energy from 100 TeV to 10 PeV is the combined observation of 6 telescopes and the first half of KM2A array (2020.11.01-2021.3.31), as shown in Figure 1. The second stage is the combined observation of 18 telescopes and the whole KM2A array. This work is based on the detector setup of the first stage.

3. Monte Carlo Simulations

EAS events are generated by the COsmic Ray SImulation for KAscade (CORSIKA) (7.4000 version) [7] with the high energy interaction models QGSJet-II-04 [8] and EPOS-LHC [9]. For the

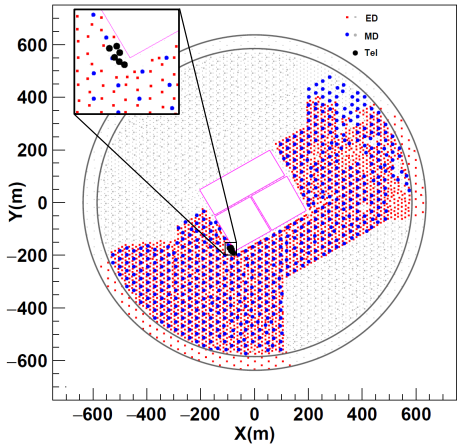

Figure 1: The layout of the detectors of the first half KM2A and the first six telescopes of WFCTA (see the black dots in the zoomed view). The solid squares and solid circles indicate the EDs and MDs in the half KM2A array. The central purple squares indicate the WCDA array region.

low-energy interactions, the FLUKA code is used. In order to simulate the combined observation of KM2A and WFCTA, the Cherenkov option is selected and both the Cherenkov data and particle data of each shower are recorded. In order to simulate the response of the telescope to photons with different wavelengths, the output of CORSIKA is modified so that the wavelength of each photon can be also recorded. To speed up the generation of the air shower, some of the Cherenkov photons are thrown away according to the quantum efficiency of the SiPM.

Five elements representing individual mass groups: proton (H), helium (He), nitrogen (CNO-group), aluminum (MgAlSi-group), and iron are generated to study the composition identification. All of the five composition groups are generated in three energy ranges 10 TeV - 100 TeV, 100 TeV - 1 PeV, and 1 PeV - 10 PeV, following a power-law function with a spectral index -1 to increase the statistics of high energy events. The simulated events are weighted to Gaisser (H3a) composition model [10]. Due to the first six WFCTA telescopes pointing to the 30° in the zenith direction, the zenith angle is sampled in the range from 20° to 40° and the azimuth angle is sampled in the range from 100° to 280°.

3.1 Selection cuts

In order to ensure a high quality of the reconstructed shower observables, the event selection criteria are applied:
(1) The events with cores located outside the KM2A array should be removed, so the events with the reconstructed core at the edge of KM2A are thrown away.
(2) The intersection angle (α) should be less than 10° to further rule out the events with erroneous reconstruction.
(3) R_p is selected from 70 m to 150 m to keep a full trigger efficiency of WFCTA for air showers with energies higher than 100 TeV.
(4) Considering the pointing of the telescopes, the showers with zenith angle range $|rec_\theta - tel_\theta| < 8°$, azimuth angle range $|rec_\phi - tel_\phi| < 13°$ are selected to make sure an integrity observation of the

Cherenkov images.
(5) The gravity center (MeanX, MeanY) of the image should be $|MeanX| < 6^\circ$ and $|MeanY| < 6^\circ$, the tail of the image should be (TailX, TailY) $|TailX| < 7^\circ$ and $|TailY| < 7^\circ$ to obtain complete Cherenkov images.
(6) The number of fired EDs should be larger than 20 and the hits number of EDs with filtering out noises should be larger than 10 to make sure a high-quality reconstruction of the shower core and arriving directions.
(7)The number of fired SiPMs is larger than 10 in the cleaned Cherenkov image.

4. Air shower reconstruction

4.1 The number of secondary particles

Considering the limitation of the area of the first half KM2A array, the new method to measure the number of secondary particles is used[11]. Taking the secondary particle in the ring of 40-200 m as an example, the ring of 40-200 m is divided into 8 sub-rings with 20 m width, and the number of secondary particles in the ring of 40-200 m can be calculated by

$$N_{se} = \sum \frac{N_{se,i}}{S_i^{eff}} \times S_i, \tag{1}$$

where $i = 1 \ldots 8$ means the i^{th} sub-ring with 20 m width in the ring of 40-200 m, $N_{se,i}$ and S_i^{eff} are the number of measured secondary particles and the sum of the area of all detectors in the i^{th} sub-ring, S_i is the geometric area of the i^{th} sub-ring. Following this definition, the number of secondary particles is independent of whether the ring $Ring_{40-200}$ is completely covered by detectors or not.

4.2 Energy reconstruction

We combine the Cherenkov lights and the number of muons to develop a new energy reconstruction method to measure the energy of air showers induced by the nucleus[11]. Based on the Heitler-Mattews model, a composite variable $N_{c\mu}$ is proposed by combining the normalized C light size and the number of muons as the following formula

$$N_{c\mu} = N_{pe} + 90N_{\mu}, \tag{2}$$

where N_{pe} is the Cherenkov lights, and the N_{μ} is the the number of muons within the ring of 40-200 m from the shower axis.

The relationship is similar between different cosmic ray components. The $N_{c\mu}$ is linear with primary energy. So the reconstructed energy (E_{rec}) can be proposed by

$$E_{rec} = kN_{c\mu}, \tag{3}$$

where the parameter k obtained by fitting the relationship between the primary energy and $N_{c\mu}$ of the light component.

According to the Eq. 3, the resolution and bias of the energy reconstruction are shown in the left plot of Figure 2. The resolution is better than 10% above 1 PeV, and the bias is less than 2%

at 1 PeV. This method of energy determination yields systematic differences between the proton and helium less than 1% above 300 TeV, as shown in Figure 2 (right). The reconstructed energy resolutions of the proton and helium of air showers improve as the energy increases. Compared with energy reconstruction only with Cherenkov light [4], this approach reduces the difference of the relative energy deviations between the proton and helium and improves the energy resolution as well.

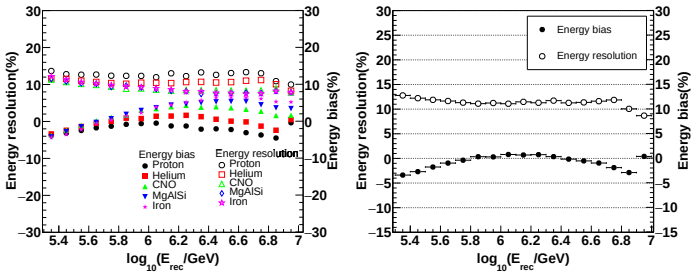

Figure 2: (Left): The energy resolution (hollowed circles) and energy bias (solid circles) versus the E_{rec} of the light components. (Right): The energy resolutions (hollowed shapes) and biases (solid shapes) of the five component versus the E_{rec}.

5. The component identification

5.1 The number of muons and electromagnetic particles

According to the Matthews-Heitler EAS model[12], The number of the muon in an air shower induced by a nucleus with atomic number A and energy E_0 can be described by the $N\mu^A_{total} \sim E_0^{0.85}A^{0.15}$. The primary energy of the air shower can be estimated by $E_0 \sim Ne^{-0.97}_{total}$. So, the parameter $N\mu^A_{total}/Ne^{0.82}_{total}$ is sensitive to the mass composition of cosmic rays. The $N\mu^A_{total}$ and Ne_{total} indicate the number of muons and the number of electromagnetic particles respectively. A is the atomic number of the primary cosmic ray. The KM2A array with ED and MD can measure the number of electromagnetic particles and the number of muons, simultaneously.

Considering the limitation of the area of the first half KM2A array, a new method to measure the number of secondary particles is used[11]. The N_e is the number of electromagnetic particles within 100m from the shower core, and the N_μ is the number of muons within 40 m - 200 m from the shower core. Based on the N_μ and N_e, the parameter N_0 (= $N_\mu/N_e^{0.82}$) is built to identify the component of cosmic rays. Figure 3(left) shows the variations of the mean value of the N_0 versus $log_{10}(E)$ for five composition groups, which indicates that the N_0 is almost energy independent for the showers with energy higher than 500 TeV. For showers with energy below 500 TeV, the N_0 is slightly energy dependent, due to the N_e measured is not the electromagnetic size at the shower maximum which is attenuated during the shower development, especially for iron showers.

5.2 The parameter related to shower maxima

According to the superposition model [12], in addition to the number of muons, the atmosphere depth of the air shower maximum ($X^A_{max} \sim \ln(E_0/A)$) is sensitive to the primary mass composition

of the air showers. Most of Cherenkov light in the gravity center of the Cherenkov images comes from the position of the shower maximum, so the $Dist$, the angular distance between the direction of the arriving direction of the air shower and the gravity center of the Cherenkov image is X_{max} related. However, for a given X_{max}, the Cherenkov images are elongated for showers with a larger R_p due to the geometric effects, so the $Dist$ is R_p dependent. Since X_{max} is increased with the primary energy of the air shower, the $Dist$ is also energy dependent. With the correction of R_p and the energy of the air shower, The mass sensitive parameter $Dist_0$ is defined as follows:

$$Dist_0 = Dist - k_p \times R_p - a_p \times N_{\mu e}^2 - b_p \times N_{\mu e}, \tag{4}$$

where k_p=0.014, a_p=-0.206, b_p=3.131 are obtained by light component and the $N_{\mu e} = N_e + 45N_\mu$ is energy estimator. In order to study the shower energy dependence of the $Dist_0$, the variations of the mean value of the $Dist_0$ versus $log_{10}(E)$ of the five groups are shown in Figure 3.

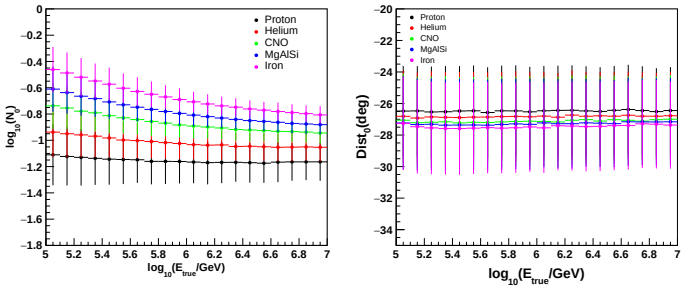

Figure 3: (Left):The variations of the mean values of the N_0 versus $log_{10}(E)$ of the five groups of cosmic rays. (Right): The variations of the mean values of the $Dist_0$ versus $log_{10}(E)$ of the five groups of cosmic rays. The MC simulation is normalized to the Gaisser model (H3a).

5.3 The Toolkit for Multivariate methods

In this paper, the parameters N_0 and $Dist_0$ are used to identify components of cosmic rays. The boundary of light components (proton and helium) and heavy components (CNO, MgAlSi, and Iron) can't be found by a regular formula, and the relation between the boundary and the energy of cosmic rays is also hard to find. In order to optimize boundary, the Toolkit for Multivariate methods (TMVA) is used in this work [13]. Compared with traditional cut-based analysis techniques, TMVA integrates multiple advanced algorithms, such as Boosted Decision Trees (BDT) and Artificial Neural Networks (ANN). The TMVA method can consider the nonlinear correlations between input parameters and find the optimal boundary.

In this work, Boosted Decision Trees with Gradient (BDTG) is employed for training. The light component and heavy component generated based on hadronic interaction model QGSJet-II-04 and weighted by Giasser composition model was input as the signal and background, respectively, more details can be found in [14]. With TMVA, a new variable so-called "score" can be obtained, which is sensitive to components of cosmic rays and almost energy independent, as shown in Figure 4.

5.4 Purity and identification efficiency

The reconstruction of shower energy is finished with all good samples of light components with a certain ratio of proton and helium. After light component selection by score cut, the energy

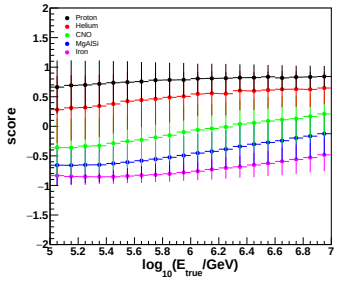

Figure 4: $score$ versus $log_{10}(E)$ of five components.

reconstruction resolution and bias of light component would be worse due to the changes in the ratio of proton and helium under different selection efficiency. The light components' efficiency eff and the purity of light components pur are defined as

$$eff(E) = \frac{N_{score}^{H+He}}{N^{H+He}}, \tag{5}$$

$$pur(E) = \frac{N_{score}^{H+He}}{N_{score}^{H+He} + N_{score}^{Heavy}}, \tag{6}$$

Where N^{H+He} is the number of pure proton and Helium events after event selection cuts, N_{score}^{H+He} is the number of pure proton and Helium events after event selection cuts and score cut. N_{score}^{Heavy} is the number of contamination events after event selection cuts and score cut. Figure 5 shows the variation of eff and pur with reconstruction energy based on MC simulations with the Gaisser (H3a) model, the purity can reach 95% with 80% selection efficiency around 1 PeV. And we observe that pur decreases at high energies, due to the increased relative abundance of the heavy components.

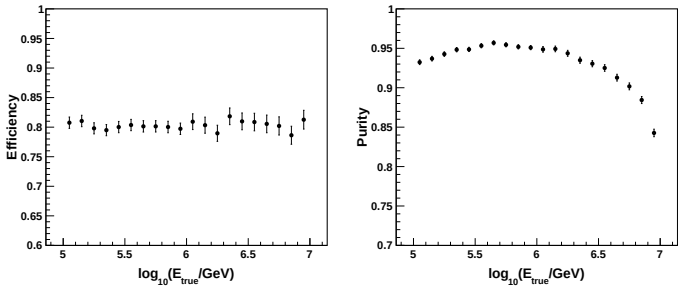

Figure 5: (Left): The selection efficiency of light component versus reconstructed energy. (Right) The purity of light component versus reconstructed energy. The error bar is statistical error. The MC simulations are generated using QGSJET-II-04.

6. Summary and discussion

The advantages and performance of the combined observations of the first half of KM2A and 6 telescopes of WFCTA are studied in detail. With the combined observation, the data quality is improved effectively. Total p.e. in the Cherenkov images and the number of moun are used to

reconstruct the energies of the air showers, the energy resolution of the light component can reach up to 13% with 3% bias above 300 TeV. The variables (N_0, $Dist_0$) are defined in this work, which is sensitive to the component of cosmic rays and shower energy independent. They are used as inputs for the BDTG classifier for the TMVA method. The new variable named score is obtained by the TMVA method. Comprehensive consideration of the effects of purity and selection efficiency of light components, under the current identification variable, the 80% selection efficiency of light components is fixed, and the purity of light components is about 95% at 1 PeV.

References

[1] Z. Cao *et al.* [LHAASO], Nature **594**, no.7861, 33-36 (2021) doi:10.1038/s41586-021-03498-z

[2] T. Antoni *et al.* [KASCADE], Astropart. Phys. **24**, 1-25 (2005) doi:10.1016/j.astropartphys.2005.04.001 [arXiv:astro-ph/0505413 [astro-ph]].

[3] M. Amenomori *et al.* [Tibet ASgamma], Adv. Space Res. **47**, 629-639 (2011) doi:10.1016/j.asr.2010.08.029

[4] B. Bartoli *et al.* [ARGO-YBJ and LHAASO], Phys. Rev. D **92**, no.9, 092005 (2015) doi:10.1103/PhysRevD.92.092005 [arXiv:1502.03164 [astro-ph.HE]].

[5] Huihai. He *et al.* [LHAASO], Radiation Detection Technology and Methods **2**,no.7 (2018) doi:https://doi.org/10.1007/s41605-018-0037-3

[6] X. H. Ma, Y. J. Bi, Z. Cao, M. J. Chen, S. Z. Chen, Y. D. Cheng, G. H. Gong, M. H. Gu, H. H. He and C. Hou, *et al.* Chin. Phys. C **46**, no.3, 030001 (2022) doi:10.1088/1674-1137/ac3fa6

[7] Heck, D. and Knapp, J. and Capdevielle, J. and Schatz, G. and Thouw, T., CORSIKA: A Monte Carlo code to simulate extensive air showers

[8] Ostapchenko, S., QGSJET-II: physics, recent improvements, and results for air showers, EPJ Web of Conferences [9]

[9] T. Pierog, I. Karpenko, J. M. Katzy, E. Yatsenko and K. Werner, Phys. Rev. C **92**, no.3, 034906 (2015) doi:10.1103/PhysRevC.92.034906 [arXiv:1306.0121 [hep-ph]].

[10] T. K. Gaisser, T. Stanev and S. Tilav, Front. Phys. (Beijing) **8**, 748-758 (2013) doi:10.1007/s11467-013-0319-7 [arXiv:1303.3565 [astro-ph.HE]].

[11] L. Wang, L. Ma, S. Zhang, C. Feng, Z. Cao, L. Yin, Y. Wang, J. Zhao, Z. Li and L. Wang, *et al.* Phys. Rev. D **107**, no.4, 043036 (2023) doi:10.1103/PhysRevD.107.043036

[12] J. Matthews, Astropart. Phys. **22**, 387-397 (2005) doi:10.1016/j.astropartphys.2004.09.003

[13] P. Speckmayer, A. Hocker, J. Stelzer and H. Voss, J. Phys. Conf. Ser. **219**, 032057 (2010) doi:10.1088/1742-6596/219/3/032057

[14] L. Q. Yin *et al.* [LHAASO], Chin. Phys. C **43**, no.7, 075001 (2019) doi:10.1088/1674-1137/43/7/075001