Cosmic ray spectrum of protons plus helium nuclei between 6 and 158 TeV from HAWC data
A. Albert et al.
Phys. Rev. D 105, 063021 — Published 25 March 2022
DOI: 10.1103/PhysRevD.105.063021
The cosmic ray spectrum of protons plus helium nuclei between 6 TeV and 158 TeV
from HAWC data

A. Albert, 1 R. Alfaro, 2 C. Alvarez, 3 J.R. Angeles Camacho, 2 J.C. Arteaga-Velázquez, 4, † K.P. Arunbabu, 5 D. Avila Rojas, 2 H.A. Ayala Solares, 6 E. Belmont-Moreno, 2 C. Brisbois, 7 K.S. Caballero-Mora, 7 T. Capistrán, 8 A. Carramiñana, 9 S. Casanova, 10 U. Cotti, 4 J. Cotzomi, 11 E. De la Fuente, 12, 13 R. Diaz Hernandez, 9 M.A. DuVernois, 14 M. Durocher, 1 J.C. Díaz-Vélez, 12 C. Espinoza, 2 N. Fraija, 8 J.A. García-González, 15 F. Garfías, 8 M.M. González, 8 J.A. Goodman, 7 J.P. Harding, 1 B. Hona, 16 D. Huang, 17 F. Hueyotl-Zahuantitla, 3 P. Hüntemeyer, 17 A. Iriarte, 8 V. Joshi, 18 D. Kieda, 16 G.J. Kunde, 1 A. Lara, 5 W.H. Lee, 8 H. León Vargas, 2 J.T. Limnemann, 19 A.L. Longinotti, 8 G. Luis-Raya, 20 K. Malone, 1 O. Martínez, 11 J. Martínez-Castro, 21 J.A. Matthews, 22 P. Miranda-Romagnoli, 23 J.A. Morales-Soto, 4 E. Moreno, 11 M. Mostafá, 6 A. Nayerhoda, 10 L. Nellen, 24 M. Newbold, 16 R. Noriega-Papaqui, 23 N. Omodei, 25 E.G. Pérez-Pérez, 20 C.D. Rho, 26 D. Rosa-González, 9 H. Salazar, 11 F. Salesa Greus, 10, 27 A. Sandoval, 2 J. Serna-Franco, 2 A.J. Smith, 7 R.W. Springer, 16 K. Tollefson, 19 I. Torres, 9 R. Torres-Escobedo, 28 F. Ureña-Mena, 9 L. Villaseñor, 11 X. Wang, 17 E. Willcox, 7 H. Zhou, 28 C. de León, 4 J.D. Álvarez, 4, ‡ G.B. Yodh, 29 and A. Zepeda 30

1Physics Division, Los Alamos National Laboratory, Los Alamos, NM, USA
2Instituto de Física, Universidad Nacional Autónoma de México, Ciudad de México, Mexico
3Universidad Autónoma de Chiapas, Tuxtla Gutiérrez, Chiapas, Mexico
4Universidad Michoacana de San Nicolás de Hidalgo, Morelia, Mexico
5Instituto de Geofísica, Universidad Nacional Autónoma de México, Ciudad de México, Mexico
6Department of Physics, Pennyslvania State University, University Park, PA, USA
7Department of Physics, University of Maryland, College Park, MD, USA
8Instituto de Astronomía, Universidad Nacional Autónoma de México, Ciudad de México, Mexico
9Instituto Nacional de Astrofísica, Óptica y Electrónica, Puebla, Mexico
10Institute of Nuclear Physics Polish Academy of Sciences, PL-31342 IFJ-PAN, Krakow, Poland
11Facultad de Ciencias Físico Matemáticas, Benemérita Universidad Autónoma de Puebla, Puebla, Mexico
12Departamento de Física, Centro Universitario de Ciencias Exactas e Ingenierías, Universidad de Guadalajara, Guadalajara, Mexico
13Institute for Cosmic Ray Research, University of Tokyo, Kashiwa, Kashiwanoha, Japan
14Department of Physics, University of Wisconsin-Madison, Madison, WI, USA
15Tecnológico de Monterrey, Escuela de Ingeniería y Ciencias, Ave. Eugenio Garza Sada 2501, Monterrey, N.L., Mexico, 64849
16Department of Physics and Astronomy, University of Utah, Salt Lake City, UT, USA
17Department of Physics, Michigan Technological University, Houghton, MI, USA
18Erlangen Center for Astroparticle Physics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany
19Department of Physics and Astronomy, Michigan State University, East Lansing, MI, USA
20Universidad Politécnica de Pachuca, Pachuca, Hgo, Mexico
21Centro de Investigación en Computación, Instituto Politécnico Nacional, México City, Mexico.
22Dept of Physics and Astronomy, University of New Mexico, Albuquerque, NM, USA
23Universidad Autónoma del Estado de Hidalgo, Pachuca, Mexico
24Instituto de Ciencias Nucleares, Universidad Nacional Autónoma de México, Ciudad de México, Mexico
25Department of Physics, Stanford University, Stanford, CA 94305-4060, USA
26University of Seoul, Seoul, Rep. of Korea
27Instituto de Física Corpuscular, CSIC, Universitat de València, E-46980, Paterna, Valencia, Spain
28Tsung-Dao Lee Institute & School of Physics and Astronomy, Shanghai Jiao Tong University, Shanghai, China
29Department of Physics and Astronomy, University of California, Irvine, Irvine, CA, USA
30Physics Department, Centro de Investigación y de Estudios Avanzados del IPN, Ciudad de México, Mexico

A measurement with high statistics of the differential energy spectrum of light elements in cosmic rays, in particular, of primary H plus He nuclei, is reported. The spectrum is presented in the energy range from 6 TeV to 158 TeV per nucleus. Data was collected with the High Altitude Water Cherenkov (HAWC) Observatory between June 2015 and June 2019. The analysis was based on a Bayesian unfolding procedure, which was applied on a subsample of vertical HAWC data that was enriched to 82% of events induced by light nuclei. To achieve the mass separation, a cut on the lateral age of air shower data was set guided by predictions of CORSIKA/QGSJET-II-04 simulations. The measured spectrum is consistent with a broken power-law spectrum and shows a knee-like feature at around $E = 24.0^{+3.3}_{-3.1}$ TeV, with a spectral index $\gamma = -2.51 \pm 0.02$ before the break and with $\gamma = -2.83 \pm 0.02$ above it. The feature has a statistical significance of 4.1 $\sigma$. Within systematic uncertainties, the significance of the spectral break is 0.8 $\sigma$. 

I. INTRODUCTION

Cosmic rays are mainly relativistic atomic nuclei that impinge nearly isotropically on Earth from outer space with energies that extend from a few MeV to some TeV [1–3]. Above $10^{13}$ eV, cosmic rays can be studied indirectly by means of air shower techniques [4–6], and below $10^{15}$ eV, with direct methods by using particle detectors on board of balloons, spaceships and satellites [7, 8].

The energy region between $10^{13}$ eV and $10^{15}$ eV is the frontier between the direct and indirect detection techniques of cosmic rays. Historically, data have been difficult to obtain in this energy interval due to limitations owing to both detection methods. In spite of that, early experiments have found out that the differential energy spectrum of cosmic rays in this energy regime can be roughly described by a power-law $E^\gamma$ with a spectral index $\gamma \sim -2.7$ and that it seems to be dominated by hydrogen and helium nuclei [1, 3] at least up to 700 TeV [9]. These observations appear to be consistent with theoretical models that assume the existence of a common type of galactic source for TeV and PeV cosmic rays [10–12], for example, supernova remnants [1–3]. Yet, they cannot rule out more complex astrophysical scenarios that involve, for instance, the presence of local cosmic ray sources [13, 14] or of a new population of cosmic ray accelerators with cutoff energies of TeV [15–17]. Such models usually predict fine structures in the energy spectrum of the all-particle and individual mass groups of cosmic rays at TeV energies, whose existence can only be tested with precise data and with high statistical power measurements on the energy and composition of cosmic rays.

In this regard, recent data provided by the satellites DAMPE [18, 19] and NUCLEON [20], as well as the extensive air shower (EAS) observatory HAWC [21] seem to reveal that, in fact, the energy spectra of cosmic rays in the TeV region show the presence of individual features which cannot be fitted by a single power law.

First hints about the existence of fine structure in the energy spectra of cosmic rays came from the balloon-borne ATIC-2 [22] and CREAM [23] experiments, and from early measurements carried out with the NUCLEON satellite observatory [24, 25]. The data from these instruments seemed to point out the presence of spectral breaks between 10 TeV and 40 TeV in the spectra of H and He nuclei. However, those results were not conclusive due to the lack of statistics. A clear indication of a feature in the 10 TeV – 100 TeV range was provided later by the HAWC observatory, which showed the existence of a break in the all-particle energy spectrum of cosmic rays at around 46 TeV [21]. Just recently, the NUCLEON experiment, with more statistics, gave further support to the existence of breaking features in the proton and helium spectra at energies of around $Z \times 10$ TeV [20], respectively, while the DAMPE satellite experiment [18, 19] provided significant evidence for individual knee-like structures in the spectra of protons and helium nuclei at $\sim 14$ TeV and $\sim 34$ TeV, respectively. In addition, the HAWC collaboration found a steepening in the energy spectrum of the light mass group ($\text{H} + \text{He}$) of cosmic rays close to 30 TeV [26]. The relation between these structures and the spectral break in the all-particle energy spectrum at 46 TeV is still not clear, but future studies on the different mass groups of cosmic rays, as in [20], may throw some light on the issue.

In the present paper, we have updated the analysis performed in [26] on the energy spectrum of light primaries at tens of TeV. Since the appearance of [26] further improvements have been included in the study, such as the employment of an updated set of Monte Carlo (MC) simulations of the HAWC detectors [27] and the usage of a bigger experimental data set, which spans the period of time from June 11th, 2015 to June 3rd, 2019. This reduced both statistical and PMT systematic uncertainties. As in [27], the analysis procedure in this work is based on an unfolding technique, which is applied on a large collection of data that has a high proportion of H and He nuclei induced events ($> 82\%$ abundance). Mass separation is done event-by-event using an energy dependent cut on the lateral shower age parameter, derived from predictions of the QGSJET-II-04 hadronic interaction model [28] for different primary nuclei.

We have targeted the mass group of light elements, as it is the most abundant component in the flux of cosmic rays in the energy region of interest [1–3] and because it is easier to separate with the present analysis technique. The paper is organized in the following way: In section II, we present the HAWC detector, the EAS reconstruction, and the methods for the calibration of the primary energy and estimation of the lateral shower age. In section III, we discuss the MC simulations used in our analysis. Section IV gives the event selection criteria. Section V describes the data, the mass separation and the reconstruction of the spectrum. The unfolded result and a comparison with measurements of other experiments come in section VI. Section VII discusses the result and section VIII gives our conclusions from the work. Appendix A contains a description of the composition models used in this analysis. Appendix B provides a detailed list of the statistical and systematic error sources. Appendix C describes systematic checks carried out to verify our result.

II. THE HAWC OBSERVATORY

A. Experimental set-up

HAWC is a high altitude air shower observatory optimized for studying the gamma-ray sky in the 500 GeV –
lays between detectors [30, 32]. The reconstruction software uses data from PMTs with $Q_{\text{eff}}$ below a maximum calibrated value of $\approx 10^4$ $PE$ to estimate various EAS observables of the event, such as arrival direction, shower core position, the lateral distribution of deposited charge, lateral shower age and primary energy [21, 27, 30]. In the following subsections, we detail the estimation of the lateral shower age and primary energy in HAWC.

1. Lateral shower age

The lateral shower age $s$ is related to the shape of the lateral distribution of an EAS and gives a measure of its steepness. It is an important parameter for the study of air showers as it depends on the distance from the shower maximum to the observation point and is sensitive to the primary mass. The lateral age was introduced through the Nishimura-Kamata-Greisen (NKG) lateral density distribution in the context of pure electromagnetic cascades [33–35].

According to their relative age value, EAS can be classified into old and young showers [36]. Old showers have large $s$ values and flatter lateral distributions. They are characterized by shower maxima at small atmospheric depths. Young showers possess small values of $s$ and steeper lateral distributions, and are associated to EAS which penetrate deeper in the atmosphere. On average, heavy primaries tend to produce older showers than light nuclei, while high energy primaries create younger EAS than low energy ones.

In HAWC, the lateral age of EAS is obtained event-by-event from a $\chi^2$ fit with a modified NKG function,

$$f(r) = A \left( \frac{r}{r_0} \right)^{s-3} \left( 1 + \frac{r}{r_0} \right)^{s-4.5},$$

(1)

to the lateral charge distribution measured by the PMTs $Q_{\text{eff}}(r)$ [27]. Here, $r$ is the radial distance to the EAS axis in the shower plane, $r_0 \approx 124$ m is the Molière radius at the HAWC site and $A$ is the amplitude of the function, which is also a fit parameter. This lateral distribution function, originally proposed for describing EAS initiated by gamma rays, also gives a reasonable description of the measured lateral distribution of hadron-induced showers [42]. This is illustrated in fig. 1, where we show a fit of equation (1) to the measured lateral distribution of a typical hadronic event that arrived with a zenith angle $\theta = 1.04^\circ$ and an azimuth $\phi = 202.24^\circ$, which had a reconstructed primary energy of $\log_{10}(E_{\text{rec}}/\text{GeV}) = 5.05$. The rather young shower age was $s = 1.41 \pm 0.02$. The result of the fit gave a reduced $\chi^2$ of 3.63 for $n_{\text{dof}} = 1018$ degrees of freedom. This is a large value for $\chi^2/n_{\text{dof}}$, which is due to the natural width of the lateral distribution of hadronic air showers, which is bigger than the experimental error on $Q_{\text{eff}}(r)$.

Note in fig. 1, the presence of outliers in the measured lateral distribution. These features are usually present in hadronic induced EAS and are mainly associated to...
large and localized charged depositions in the detectors from shower muons [43]. In general, gamma-rays create EAS with smoother lateral distributions than those from cosmic rays. This difference is employed in HAWC for gamma/hadron separation [27, 30]. The outliers produce a small bias on the fitting parameters of the order of a few percent. In particular, for the example presented in fig. 1, they induce an increment on \( s \) and \( \log_{10}(A) \) of 3\% and 2\%, respectively.

It is worth to point out that the reduced \( \chi^2 \) of the lateral distributions of the measured data is similar to the predictions of MC simulations up to \( \log_{10}(E_{\text{rec}}/\text{GeV}) = 4.2 \) for a mixed composition scenario using our reference composition model, which will be described in the next section, and QGSJET-II-04. Meanwhile, at higher energies the experimental mean of \( \chi^2/n_{\text{dof}} \) tends to be larger than the MC expectations for the mixed composition assumption. In particular, for \( \log_{10}(E_{\text{rec}}/\text{GeV}) > 5.3 \) the values of the reduced \( \chi^2 \) of the data are above the MC predictions for pure proton and iron nuclei, which implies that in this energy regime the width of the measured lateral distributions of hadronic EAS is larger than expected from QGSJET-II-04 simulations. Further studies are needed to understand the origin of such differences.

2. Primary energy

The primary energy of the shower event is estimated from a maximum log-likelihood procedure [21], which computes and compares the probabilities that the measured lateral distribution of PMT signals from a given shower with reconstructed zenith angle \( \theta \) is produced by proton primaries of different energies, \( E \). The calculation also includes the probability of observing active PMTs with no signals during the event. In the algorithm, the probability values of the operational PMTs are extracted from probability tables, which are generated using proton-induced EAS simulations with a number of hit PMTs (\( n_{\text{Hit}} \)) greater than 75 and with EAS cores and arrival directions successfully reconstructed. The tables are obtained from CORSIKA/QGSJET-II-04 simulations for \( \log_{10}(E/\text{GeV}) = [1.85, 6.15] \) and \( \theta \leq 60^\circ \).

III. MONTE CARLO SIMULATIONS

Air shower simulations initiated by cosmic rays in HAWC were carried out using CORSIKA v7.40 [44] without the thinning option and with the hadronic interaction models FLUKA [45] and QGSJet-II-04 [28]. FLUKA is employed for hadron energies of \( E_{\text{lab}} < 80 \text{ GeV} \), while QGSJet is used at higher energies.

Simulations were conducted for eight primary species, in particular, H, He, C, O, Ne, Mg, Si and Fe, using an energy spectrum \( E^{-2} \) for the energy interval 5 GeV \(-\) 2 PeV. The MC data cover the zenith angle range \( \theta = [0^\circ, 65^\circ] \) with a \( \cos \theta \sin \theta \) distribution. Shower cores are thrown flat in radius up to 1 km from the center of the array, but reweighted to simulate a distribution uniform in area.

The HAWC detector response was simulated using software based on GEANT4 [46]. Both MC and measured events were reconstructed with the same algorithm in order to study the influence of experimental systematic uncertainties on the estimated EAS parameters.

MC events were weighted to reproduce the nominal composition model introduced in [21]. This model gives a fair description of the cosmic ray elemental spectra measured by the direct experiments AMS-2 [37, 38], CREAM I-II [39, 40], and PAMELA [41] in the energy interval from 100 GeV to \( \sim 200 \text{ TeV} \). The data are fit with a broken power-law, which is extrapolated up to a few PeV. Fig. 2 illustrates the cosmic ray intensities in our nominal composition model, with the predictions for light (\( Z \leq 2 \)), intermediate (3 \( \leq Z \leq 14 \)) and heavy cosmic ray nuclei. The expressions and fit parameters\(^1\) are taken

\(^1\) There is a typo in the value of the normalization energy \( E_0 \) of the broken power-law functions in the nominal model of [21] that is corrected here. The parameter \( E_0 \) should have the values 1200, 1600, 2000, 2400, 2800 and 5600 in GeV units, for C, O, Ne, Mg, Si and Fe nuclei, respectively.
from [21]. The total number of simulated EAS in the full zenith angle range and the whole energy interval for protons and helium primaries were $3 \times 10^{10}$ and $1.3 \times 10^{10}$, respectively, while for the rest of elemental nuclei, we simulated $10^9$ MC events per mass group. For vertical events with $\theta \lesssim 16^\circ$ and primary energies greater than 10 TeV, the number of simulated events is reduced by a factor of $2.2 \times 10^4$. Appendix A gives other composition models used to estimate systematic errors.

### IV. SELECTION CUTS

A set of selection criteria were applied to both data and MC simulations for the reconstruction of the energy spectrum with the main purpose of reducing the influence of systematic uncertainties in the final result. The selection criteria were chosen after a detailed MC study of their effect on the core position, arrival angle, primary energy of air showers and on the HAWC effective area.

The first cut discards EAS events that have not successfully passed the core and arrival direction reconstruction or have less than 75 hit PMTs. To reduce the uncertainty on the core position, we selected events with at least 40 hit PMTs within a radius of 40 m from the reconstructed EAS core ($N_{\text{hit}} \geq 40$). According to MC simulations, this cut only leaves data with reconstructed shower cores on the array or within a distance of 20 m from the boundary of HAWC. A tighter selection would reduce the efficiency of HAWC for cosmic rays with energies close to 10 TeV and increase the uncertainties on both the effective area and the reconstructed spectrum: requiring event cores inside the array increases the uncertainty on the energy spectrum up to 50% around 10 TeV respect to the value with our standard cut.

To decrease systematic errors associated with inclined showers, we keep only near-vertical EAS with $\theta < 16.7^\circ$. This value is close to the upper limit of the zenith angle range corresponding to the table used for energy calibration of vertical EAS. We also removed showers with low efficiency by requiring an estimated shower energy $\log_{10}(E_{\text{rec}}/\text{GeV}) > 3.5$ and $f_{\text{hit}} \geq 0.2$, where $f_{\text{hit}}$ is the fraction of active PMTs with hits in the event [30]. Finally, we applied an upper cut of $\log_{10}(E_{\text{rec}}/\text{GeV}) < 5.5$ to focus our analysis in the region where the uncertainties in the composition studies of cosmic rays due to the PMT systematic effects are smaller.

The effects of the successive application of the selection cuts on the total number of events of both MC and experimental data sets are seen in table I. Large reductions in the total number of selected events are associated with constraints on $N_{\text{hit}}$, $\theta$ and $N_{\text{hit}}$, as in [21]. There is also an important decrease in the data and in MC simulations due to the cut on $f_{\text{hit}}$, which removes shower events below a few TeV. These are low energy events that trigger the detector much more frequently than EAS at higher energies, which is why such a cut strongly affects both the data and MC samples.

The selection criteria have almost the same effect on both MC and experimental data samples according to table I. There are, however, some differences between the selection efficiencies of simulations and measurements. One of the largest ones is found when applying the cut

| Selection cut                  | % of remaining events respect to previous cut | Measured rate (kHz) |
|--------------------------------|---------------------------------------------|--------------------|
| Trigger                       | 100.00                                      | 100.00             |
| Passed angle and core         |                                             | 24.61              |
| reconstruction                |                                             |                    |
| $N_{\text{hit}} \geq 75$      | 95.18                                       | 100.00             |
| $N_{\text{hit}} \geq 40$      | 23.65                                       | 5.54               |
| Zenith angle                   | 26.70                                       | 5.41               |
| Fraction hit                   | 27.82                                       | 4.11               |
| Primary energy                 | 36.02                                       | 0.15               |
| Secondary energy               | 93.48                                       | 0.14               |

| Table I: Effects of the selection criteria on the data sets. The cuts are shown on the left column. The central columns represent the fraction of events from the previous cut (in percent) which pass the cut. The second column was obtained for measured data, and the third column, for MC simulations in the framework of the nominal composition model used in this work. Calculations start with data sets which satisfy $N_{\text{hit}} > 10$, the minimum for which the reconstruction saves data. As in [21], the cosmic ray detection rates in HAWC are also computed. |
The bias and resolution of the primary energy are selected data are below core position and the arrival direction of EAS in the position model and the energy spectrum of cosmic rays consider the contributions from uncertainties in the composition model for cosmic rays is softer than the actual one below 20 TeV. An increment of $\Delta \gamma = 0.1$ in the magnitude of the spectral index in the simulations with the nominal composition model for $E < 20$ TeV increases the selection efficiency due to the cut on $f_{hit}$ up to approximately 35% in the MC sample, which is closer to the corresponding selection efficiency for HAWC data. The value of $\Delta \gamma$ used to perform the previous calculation was derived by comparing the histograms of $E_{rec}$ for the measured data and for the nominal composition model. On the other hand, the selection efficiency in MC data due to the $f_{hit}$ cut can be further incremented by $\sim 1\%$ taking also into account the observed difference between the nominal composition model and the measured data regarding the relative abundance of light primaries. It is worth to mention that the estimated systematic errors for the energy spectrum of H+He performed in this work consider the contributions from uncertainties in the composition model and the energy spectrum of cosmic rays (see appendix B).

According to MC simulations, for $\log_{10}(E_{rec}/\text{GeV}) \geq 3.8$ the mean systematic uncertainties of the shower core position and the arrival direction of EAS in the selected data are below 17 m and 0.5°, respectively. The bias and resolution of the primary energy are $|\Delta \log_{10}(E_{rec}/\text{GeV})| \leq 0.09$ and $\sigma \log_{10}(E_{rec}/\text{GeV}) \leq 0.3$, correspondingly, above $\log_{10}(E_{rec}/\text{GeV}) = 3.8$. As an example, the expected mean bias and the resolution of the primary energy are shown in fig. 3 as a function of $E_{rec}$. The energy estimation and the pointing accuracy of the detector have been verified independently in [21] using measurements of the position of the Moon shadow as a function of the reconstructed energy.

V. DESCRIPTION OF THE ANALYSIS

A. Experimental data set

In the present analysis, we have used data collected with the central detector of HAWC from June 11th, 2015 to June 3rd, 2019. The total effective time amounts to $T_{eff} = 3.74$ yr, which corresponds to an experimental livetime of 94%. The data sample contains $2.9 \times 10^{12}$ EAS. After applying the selection criteria, we kept $1.6 \times 10^{10}$ showers.

B. Analysis technique

The reconstruction of the energy spectrum of proton and helium primaries applies an unfolding analysis to a subsample of events enriched in light elemental nuclei by a cut on the shower age. We correct for contamination by heavy nuclei, and triggering and reconstruction efficiency. We give details of the reconstruction chain in the
following subsections.

1. Extraction of an enriched subsample of light elements

The lateral shower age is sensitive to the mass composition of cosmic rays in HAWC, as can be seen in fig. 4. The plot shows QGSJET-II-04 predictions for the mean \( s \) of EAS caused by different mass cosmic rays as a function of the estimated energy \( E_{\text{rec}} \). The age parameter defined in eq. (1) decreases for light nuclei and for high energy cosmic rays, since these primaries produce more penetrating EAS with shower maxima closer to HAWC. The age increases slightly above \( E_{\text{rec}} = 10^5 \) GeV, due to the maximum calibrated charge of the PMTs and the finite sampling area of the detector.

On the other hand, in fig. 4, we have also compared model predictions against the average HAWC shower age. The comparison shows overall agreement between data and expectations up to \( E = 3.2 \times 10^5 \) GeV for vertical EAS. The measured shower age lies between the predictions for H and Fe primaries. For \( E < 2 \times 10^4 \) GeV, the mean shower age of the data lies between the predictions for pure H and He nuclei, suggesting that light cosmic rays dominate in this energy range. At higher energies, the data is between the expectations for He and C primaries, which may indicate that heavier cosmic ray nuclei become more important from \( 2 \times 10^4 \) GeV to \( 3.2 \times 10^5 \) GeV. We will return to this point later, in section VII.

In order to extract our data subsample for the analysis, we apply an age cut at \( s_{\text{He} - \text{C}} \), which lies between He and C nuclei. The selection criterion keeps showers with low age which are most likely produced by protons and He primaries. We choose this simple cut as we looked for a separation criterion with minimal complications that allows to get a subsample dominated by light primaries and with large statistics. The current age cut does not maximize the purity of the sample. However, it provides an energy spectrum for H plus He nuclei with a similar shape (within 1% and 8%) to the one obtained with the criterion based on the maximization of the purity of the subsample (see appendix C). Besides its simplicity, the age cut \( s_{\text{He} - \text{C}} \) has the advantage that it provides an effective area that is flatter than the one derived from a maximum purity criterion. This is another reason of our preference for the cut \( s_{\text{He} - \text{C}} \). In any case, we have included the contribution to the systematic error of the spectrum due to variations in the purity of the subsample by moving upwards and downwards our age cut (see appendix B). In particular, we have put the selection cut at the curves for the mean shower age predictions of C and He, respectively.

We must point out that the spectrum of the light mass group of cosmic rays can also be estimated without applying a cut on the measured data, for example, by fitting the bi-dimensional histogram for the measured shower age and \( E_{\text{rec}} \) with MC distributions for the light and heavy cosmic ray nuclei using unfolding methods as those applied in [47]. These procedures have the advantage that they allow to estimate independently the background of heavy cosmic ray nuclei in the data sample but they introduce a larger correlation with the light cosmic ray spectrum than in the case of the simple approach with the age cut, where the influence of the heavy nuclei is expected to be reduced. A small dependence of the result on the composition model is, however, introduced in the simple approach with the age cut through the estimation of the contamination of the heavy primaries in the selected data subsample. Each procedure has its own systematic errors. Therefore it is important to confirm the results with different techniques. In this paper, we have adopted the analysis using the age cut, however, alternative analyses with unfolding methods like in [47] are under way. In this regard, preliminary results were presented in [48]. They are very encouraging, as they confirm the main findings in this paper about the existence of a break at TeV energies in the \( \text{He+He} \) energy spectrum of cosmic rays.

According to MC simulations with our nominal composition model, the fraction of light nuclei in the subsample selected with the shower age cut varies from roughly 97% at \( E_{\text{rec}} = 3.2 \times 10^5 \) TeV down to 82% at \( 3.2 \times 10^5 \) TeV. About \( \sim 64\% \) of hydrogen and helium primaries pass the cut, almost independent of the estimated energy. After using the age cut on the measured data, we retained 9.9 \times 10^4 events. The separation of light and heavy nuclei is imperfect, as fluctuations event by event of the shower age are comparable to the average separation of light and heavy nuclei, as shown in fig. 5.

On the other hand, MC simulations also predict that the systematic uncertainties in the energy interval \( \log_{10}(E_{\text{rec}}/\text{GeV}) > 3.8 \) for the arrival direction and the core position of the selected data subset are smaller than 0.44° and 13 m, respectively, and that the energy resolution \( \sigma_{\log_{10}(E_{\text{rec}}/\text{GeV})} \) is not larger than 0.26 and decreases with the reconstructed primary energy. In particular for \( \log_{10}(E_{\text{rec}}/\text{GeV}) = 4.0 \) and 5.0, \( \sigma_{\log_{10}(E_{\text{rec}}/\text{GeV})} \) is equal to 0.23 and 0.10, respectively.

2. Measured energy histogram

The next step in the reconstruction procedure is to build the energy histogram \( N(E_{\text{rec}}) \) for the selected subsample of EAS obtained after the shower age cut. This is shown in figure 6, where a bin size of \( \Delta \log_{10}(E_{\text{rec}}/\text{GeV}) = 0.2 \) has been used, which is of the order of magnitude of the energy resolution in the selected subsample.

In fig. 7, we have estimated the ratio between the measured rate of events and the predicted ones using MC simulations with our nominal composition model and alternative ones, described in appendix A, after applying the shower age cut of fig. 4. From fig. 7, we observe that the ratios have values between 1.6 and 0.5 and they vary...
with the reconstructed energy. All of them exhibit a maximum at around $\log_{10}(E_{\text{rec}}/\text{GeV}) = 4.3$. The measured rates are larger than the expectations with the nominal, Polygonato, JACEE and MUBEE models, but smaller than the predictions with the ATIC-2 model. Therefore, albeit of the individual differences between the data and the models, the experimental rates are within the expectations from the cosmic ray composition models. On the other hand, the energy evolution of the ratio curves implies that the energy distribution behind the measured data does not follow a single power-law like in MC simulations. The results of fig. 7 seem to hint the existence of a break in the measured energy distribution at around the position of the maximum in the ratio curves. We will come to this point later in section VI.

Finally, in fig. 8, we have calculated the efficiency of the shower age cut or the fraction of remaining events after applying the age cut over the selected HAWC data. The computation was carried out by dividing the contents of the energy histograms of fig. 6 for the subsample of young EAS and for the selected data sample that does not contain the shower age cut. The efficiency of the age cut in measured data is compared with the corresponding efficiency for QGSJET-II-04 simulations using different cosmic ray composition models, including the nominal one. From the plots of fig. 8, we see that the fraction of remaining events in the experimental subsample of young EAS is smaller than expected from MC simulations mainly at high energies. This discrepancy seems to point out that in the framework of QGSJE-II-04 the relative abundance of heavy nuclei in HAWC data is larger than predicted by the cosmic ray composition models used in this work. Such difference between data and MC simulations reduces the magnitude of the intensity of protons and helium nuclei estimated with the present procedure, but it does not change the main conclusions about its
shape. The effect of the discrepancy in the final result was estimated and included as a systematic error (see appendix B).

3. Unfolding procedure

Now, in order to correct the measured distribution for migration effects, we must apply an unfolding procedure. For this aim, we employed the Bayesian algorithm [49–51]. However, the final result has been verified using the Gold’s unfolding procedure (see appendix B) [52, 53]. In the Bayesian method, the unfolded distribution, $N(E)$, is found iteratively from the measured histogram by means of the calculation of a matrix $P(E|E_{\text{rec}})$ which provides the conditional probability that a given event with energy in the bin $E_{\text{rec}}$ is due to an EAS with true energy in the interval around $E$. The smearing matrix is computed using the Bayes’s theorem

$$P(E|E_{\text{rec}}) = \frac{P(E_{\text{rec}}|E) \cdot P(E)}{\sum_{E'} P(E_{\text{rec}}|E') \cdot P(E')} ,$$

(2)

where $P(E) = N(E)/\sum_{E} N(E)$ is the previous approximation to the probability of the unfolded distribution and $P(E_{\text{rec}}|E)$ is the response matrix of the detector. The response matrix is generally estimated from MC simulations. It represents the probability that a shower with a primary energy $E$ is reconstructed with an energy $E_{\text{rec}}$ in the experiment. The smearing matrix is then substituted into the equation

$$N(E) = \sum_{E_{\text{rec}}} P(E|E_{\text{rec}}) N(E_{\text{rec}}),$$

(3)

from which the unfolded distribution is obtained.

The unfolding procedure starts with a first guess at the probability $P(E)$. This is used to estimate a more accurate $N(E)$ by means of eqs. (2) and (3), which is employed to calculate $P(E)$ for the next iteration. The stopping criterion is described below.

To begin with, we computed the response matrix for the subset of selected events by using our nominal MC data set. The matrix was built in the $\log_{10}(E_{\text{rec}}/\text{GeV})$ vs $\log_{10}(E/\text{GeV})$ phase space for the ranges from 2.4 to 6.0, which were both divided in bins of width 0.2 as for the measured energy histogram. The resulting response matrix is shown in fig. 9. For the initial guess of $N(E)$, we used a power-law distribution with spectral index as predicted by the nominal composition model. In addition, to eliminate the propagation of statistical fluctuations from the response matrix, a smoothing procedure was applied to $N(E)$ at the end of each iteration, but not in the final result. The procedure was carried out by smoothing the unfolded distribution with a broken power-law function [54] inside the range from $E = 10^{3}$ GeV to $3.2 \times 10^{6}$ GeV. However, we have cross-checked the unfolded result using as a smoothing function a polynomial of degree 5 and the smoothing 353HQ-twice algorithm [55] as installed in the ROOT package [56]. Employing as a smoothing function a single power-law formula produces an unfolded energy distribution whose forward-folded histogram is flatter than the original distribution $N(E_{\text{rec}})$. For this reason, we avoided to use this approach in our unfolding analysis.

As the stopping criterion, we look for a minimum in the weighted mean squared error (WMSE) [47, 57] at which the squared sum of the average statistical uncertainties and the systematic biases of the result are smallest. The WMSE is defined as:

$$WMSE = \frac{1}{n} \sum_{i=1}^{n} \frac{\sigma_{\text{stat},i}^2 + \delta_{\text{bias},i}^2}{N(E_i)} ,$$

(4)

where $n$ is the number of energy bins in the unfolded distribution, $\sigma_{\text{stat},i}$ is the average statistical uncertainty of $N(E_i)$, while $\delta_{\text{bias},i}$ is the mean bias of $N(E_i)$ introduced by the unfolding algorithm. For the estimation of $\sigma_{\text{stat}}$ and $\delta_{\text{bias}}$, the bootstrap method [58] was implemented as described in [47, 59]. In particular, at a given iteration level, a set of $m = 60$ toy distributions are produced from $N(E_{\text{rec}})$ for the estimation of $\sigma_{\text{stat}}$ and $\delta_{\text{bias}}$.

4. Reconstruction of the energy spectrum

Once the unfolded spectrum $N(E)$ for the data sub-sample is obtained, the energy spectrum for protons and helium nuclei is calculated from the formula

$$\Phi(E) = \frac{N(E)}{A_{eff}(E) \Delta E T_{eff} \Delta \Omega},$$

(5)

where $\Delta E$ is the size of the bin at $E$, $T_{eff} = 1.18 \times 10^{6}$ s is the effective livetime for the collected data, $\Delta \Omega = 0.27$ sr.
is the solid angle interval covered by the measurements and $A_{\text{eff}}(E)$ is a corrected effective area defined as

$$A_{\text{eff}}(E) = f_{\text{corr}}(E) A_{\text{eff}}^{H+He}(E).$$  \hspace{1cm} (6)

In the above expression, $f_{\text{corr}}(E)$ is the factor introduced to correct the unfolded result for the contamination of heavy elements ($Z \geq 3$); $A_{\text{eff}}^{H+He}(E)$ is the effective area (defined below) of the instrument for detection of protons and helium nuclei in the enriched subsample of young EAS, which correct the unfolded result for the loss of light primaries after using the different selection cuts.

The factor $f_{\text{corr}}(E)$ is estimated as the inverse of the proportion of light primaries in the aforementioned subsample at the energy $E$, calculated from MC simulations with our nominal cosmic ray composition model. Specifically, $f_{\text{corr}}(E)$ is just the ratio $N_{H+He}^{\text{MC}}(E)/N_{H+He}^{\text{MC}}(E)$ between the number of selected events after using the age cut $N_{H+He}^{\text{MC}}(E)$ and the number of H and He events $N_{H+He}^{\text{MC}}(E)$ in this subsample at the true energy $E$. The result is shown in fig. 10 (left panel). From this plot, we observe that $f_{\text{corr}}(E)$ grows at high energies, due to the increasing relative abundance of the heavy mass nuclei in the subsample. It has a feature at around $\log_{10}(E/\text{GeV}) = 4.5$, because of the effect of the trigger and selection efficiency on the energy distribution of the heavy mass group in the MC data subset. Below this energy, the corresponding efficiency decreases rapidly producing a fast reduction in $f_{\text{corr}}(E)$ at low energies. At higher energies, the efficiency for heavy primaries starts to reach its maximum value, which reduces the rate of increment of $f_{\text{corr}}(E)$.

The shape of the correction factor is almost similar for the different cosmic ray composition models employed in this work, with some differences in the slopes and the magnitudes of the curves due to the distinct abundances of the heavy elements in each model. Since the light mass group of cosmic rays is the dominant component in the subsample of young EAS, the effect of the uncertainties of the relative abundance of the heavy component on $f_{\text{corr}}(E)$ is reduced and, hence, the corresponding error on the shape of the reconstructed energy spectrum. Therefore, we can use $f_{\text{corr}}(E)$ as estimated with the MC simulations for the present analysis independently of the shape of the experimental spectrum for H+He in this energy regime. This point is demonstrated in appendix C, where we have performed systematic checks with MC simulations and different composition models that show that our analysis method allows to reconstruct the shape of the spectrum of light primaries without previous knowledge about the existence of features in the spectrum under analysis. Even more, the study of systematic uncertainties performed in appendix B points out that the shape of the reconstructed spectrum for light primaries is the same whether we use our nominal composition model or the alternative models described in appendix A.

The effective area for the light primaries is defined by [21]:

$$A_{\text{eff}}^{H+He}(E) = A_{\text{thrown}} \frac{\cos \theta_{\text{max}} + \cos \theta_{\text{min}}}{2} \epsilon^{H+He}(E).$$  \hspace{1cm} (7)

Here $A_{\text{thrown}}$ is the total area at ground level where the core of the MC events were thrown, the cos term gives the projection of the area averaged on the solid angle within the zenith angle range from $0^\circ$ to $16.7^\circ$ and, $\epsilon^{H+He}$ is the probability that an EAS event induced by...
responding systematic and statistical errors. The result is presented in fig. 11 and table II for energies (as we will see in the next subsection). Fig. 11 shows the rapid increase of the systematic uncertainties at higher energies. The all-particle energy spectrum for cosmic rays measured by HAWC and presented in [21] is also shown (open squares). Statistical errors smaller than the data points are not shown.

A light primary (hydrogen or helium nuclei) triggers the detector and passes all the selection cuts for the young shower subsample. $A_{eff}^{H+He}(E)$ from our nominal MC simulations is plotted in fig. 10 (right panel) against the true primary energy compared to the effective area for pure hydrogen and helium nuclei. The maximum efficiency is achieved between $log_{10}(E/GeV) \sim 4$ and 5.4. At lower energies, the decrease is due to the trigger and the selection cuts, while above $log_{10}(E/GeV) = 5.4$, it is caused by the cut on the reconstructed energy. The effective areas for pure H and He nuclei are not equal. For $log_{10}(E/GeV) < 3.8$ GeV they differ by more than 30$\%$ with respect to the central value for $H+He$, for this reason and due to the reduction of the effective area as well as the increment of the correlations at lower energies, we report the spectrum only above 6 TeV.

VI. RESULTS

A. Energy spectrum of H plus He cosmic ray nuclei

The energy spectrum of light cosmic ray nuclei estimated from this analysis is presented in fig. 11 and table II for $log_{10}(E/GeV) = [3.8, 5.2]$ along with its corresponding systematic and statistical errors. The result has been constrained to $log_{10}(E/GeV) < 5.2$ due to a rapid increase of the systematic uncertainties at higher energies (as we will see in the next subsection). Fig. 11 seems to reveal a slope change around a few tens of TeV in the spectrum of $H+He$ primaries. We compared two fits to the data with a single power law

$$\Phi(E) = \Phi_0 E^{\gamma_1}, \quad (8)$$

and a broken power-law [60]

$$\Phi(E) = \Phi_0 E^{\gamma_1} \left[ 1 + \left( \frac{E}{E_0} \right)^{\frac{\gamma_2 - \gamma_1}{\varepsilon}} \right], \quad (9)$$

where $E_0$ is the energy position of the break, $\gamma_1$ and $\gamma_2$ are the spectral indexes before and after the break in the spectrum, while $\varepsilon$ measures the sharpness of the feature. The fits were done by chi-squared minimization for correlated data points [61], taking into account the correlation from the unfolding. The covariance matrix has the contributions from the statistics of MC and experimental data (see the next section and appendix B for details). The contributions were calculated according to [62, 63] and added to obtain the total covariance matrix, $V_{stat}$, used for the fit.

By fitting the spectrum with eq. (8), we obtained

$$\Phi_0 = 10^{4.32 \pm 0.02} \text{ m}^{-2} \text{s}^{-1} \text{sr}^{-1} \text{GeV}^{-1},$$

$$\gamma_1 = -2.66 \pm 0.01,$$

with $\chi^2_0 = 177.51$, for $v_0 = 5$ degrees of freedom. The fit with the broken power-law formula of eq. (9) yielded

$$\Phi_0 = 10^{3.71 \pm 0.09} \text{ m}^{-2} \text{s}^{-1} \text{sr}^{-1} \text{GeV}^{-1},$$

$$\gamma_1 = -2.51 \pm 0.02,$$

$$\gamma_2 = -2.83 \pm 0.02,$$

$$E_0 = 10^{4.38 \pm 0.06} \text{ GeV},$$

$$\varepsilon = 9.8 \pm 4.1.$$ The resulting chi-squared was $\chi^2_0 = 0.26$ and the number of degrees of freedom were $v_1 = 2$. The fitted functions are shown in fig. 12. We will use now the test statistic

$$TS = -\Delta \chi^2 = - (\chi^2_1 - \chi^2_0) \quad (10)$$

TABLE II. Values of the energy spectrum $\Phi(E)$ for the light mass group of cosmic rays as derived in this analysis using HAWC data calibrated with the QGSJET-II-04 model. The width of the energy bins employed in this study is $\Delta log_{10}(E/GeV) = 0.2$. The statistical ($\delta \Phi_{stat}$) and systematic ($\delta \Phi_{syst}$) errors of the spectrum are also given.

| $E$ [GeV] | $\Phi(E)$ | $\delta \Phi_{stat}$ | $\delta \Phi_{syst}$ |
|-----------|------------|----------------------|----------------------|
| 7.94 $\times$ 10^{3} | (8.44 $\pm$ 0.07) $\pm$ 0.45 $\pm$ 1.06 $\times$ 10^{-7} | |
| 1.26 $\times$ 10^{4} | (2.66 $\pm$ 0.03) $\pm$ 0.14 $\pm$ 0.38 $\times$ 10^{-7} | |
| 2.00 $\times$ 10^{4} | (8.34 $\pm$ 0.12) $\pm$ 0.46 $\pm$ 1.36 $\times$ 10^{-8} | |
| 3.16 $\times$ 10^{4} | (2.42 $\pm$ 0.05) $\pm$ 0.29 $\pm$ 0.45 $\times$ 10^{-8} | |
| 5.01 $\times$ 10^{4} | (6.55 $\pm$ 0.16) $\pm$ 1.11 $\pm$ 1.33 $\times$ 10^{-9} | |
| 7.94 $\times$ 10^{4} | (1.77 $\pm$ 0.05) $\pm$ 0.41 $\pm$ 0.39 $\times$ 10^{-9} | |
| 1.26 $\times$ 10^{5} | (4.95 $\pm$ 0.19) $\pm$ 1.43 $\pm$ 1.12 $\times$ 10^{-10} | |
to compare the scenarios. From the fits, we found $TS_{obs} = 177.25$. We translated this into a $p$-value using $49 \times 10^3$ toy MC spectra with correlated data points, assuming that the data is best described by the single power-law function. Following [64] we used a multivariate Gaussian as a probability distribution for the data and the covariance matrix $V_{stat}$. In the resulting $TS$ values, we found just one case with $TS \geq TS_{obs}$, which implies a $p$-value equal to $2 \times 10^{-5}$. We also observed that $0.5\%$ of the MC toy spectra have a $\chi^2$ smaller than $\chi^2 = 0.26$ when using formula (9) in the fits. Thus, from the test statistic, the broken power-law hypothesis is favored by the data with a significance of $4.1 \sigma$. We also performed several “sanity checks” (see appendix C) to rule out the kink being produced by systematic effects. Our result confirms the kink that HAWC [26] previously reported at tens of TeV in the cosmic ray energy spectrum for protons and helium nuclei and the hints found in fig. 7 in favor of a spectral break in the spectrum of this mass group.

In figure 13, this work is compared with other experiments. We have included measurements from the direct cosmic ray detectors ATIC-2 [22], CREAM I-III [23], NUCLEON [20], JACEE [65] and DAMPE [66] along with data from the air shower observatories ARGO-YBJ [67], Tibet AS-gamma [68] and EAS-TOP [69]. Close to $E = 10$ TeV, we see good agreement of HAWC data with ATIC-2 within systematic uncertainties. Between 20 TeV and 126 TeV, the HAWC measurement is in a fair agreement with the NUCLEON spectrum. In general, the HAWC result is higher than the CREAM I-III and ARGO-YBJ data below 80 TeV. However, close to 100 TeV, the CREAM I-III and ARGO-YBJ spectra are in agreement within systematic uncertainties with the HAWC spectrum. On the other hand, HAWC data is above JACEE and Tibet AS-gamma measurements.

The HAWC spectrum is not in agreement with the single power law behavior reported by ARGO-YBJ in this energy interval [67], while above 24 TeV the slope of the HAWC spectrum is harder than that of JACEE, but softer than the one of NUCLEON. Finally, at higher energies, our result is in agreement with the single data point from the EAS-TOP experiment at $\sim 80$ TeV.

B. Uncertainties in the magnitude of the spectrum

The total uncertainty (the sum in quadrature of the systematic and statistical uncertainties) of the unfolded spectrum is between $+29.1\%$ and $-22.9\%$ for energies $\log_{10}(E/\text{GeV}) < 5.2$. Figure 14 shows that the uncertainties decrease for low energies. The statistical uncertainty (dominated by MC statistics) rises from $\pm 0.8\%$ to $\pm 3.8\%$ between $\log_{10}(E/\text{GeV}) = 3.8$ and $\log_{10}(E/\text{GeV}) = 5.2$. The systematic uncertainties over the same range vary from $+5.3\%/-12.6\%$ to $+28.9\%/-22.6\%$; the systematic uncertainties dominate the total error. For energies $\log_{10}(E/\text{GeV}) > 5.2$, the systematic error grows rapidly to $+41.8\%/-25.1\%$ at $\log_{10}(E/\text{GeV}) \sim 5.4$, so we report the spectrum only up to $\log_{10}(E/\text{GeV}) = 5.2$.

We evaluated a number of sources of systematic uncertainty. The most important ones involve uncertainties on the PMT performance ($+28.5\%/-10.6\%$ effect on the deconvolved spectrum); the cosmic ray composition model ($+2.1\%/-17.2\%$); and the hadronic interaction model ($-10.9\%$ to $-3.7\%$). The rest of the systematic uncertainties, added in quadrature, contribute $+5.2\%/ -7.0\%$. Appendix B presents the systematic error evaluation in detail.

The feature observed in the energy spectrum of light primaries does not disappear under the effect of these systematic sources, although we observe some variations in the intensity of the spectrum and the value of the change of the spectral index $\Delta \gamma$ around the break. One of the dominant systematic sources in the spectrum is the uncertainty on the PMT performance, which is dominated by the PMT-late-light systematics (c.f. appendix B). The late light effect dominates the upper limit of the total systematic error and introduces an energy dependent variation in the spectrum, which grows from $+4.5\%$ up to $+28.3\%$ and reduces $\Delta \gamma$. If we apply this systematic shift on the energy spectrum and repeat the fit with formula (9) as well as the corresponding statistical analysis of the feature, we find that the significance of the break is reduced up to $3.8 \sigma$. Hence, in spite of the flattening of the spectrum, the feature is still significant.

On the other hand, we have also investigated, whether the measured spectrum is consistent with a break in the
FIG. 13. The spectrum for H\(^{+}\)+He cosmic ray nuclei as measured by HAWC (black circles) and calibrated with the post-LHC hadronic interaction model QGSJET-II-04 in comparison with similar measurements of the spectrum from direct and indirect experiments. In particular, the spectra from the direct cosmic ray detectors ATIC-2 (squares) [22], CREAM I-III (diamonds) [23], NUCLEON (downward solid triangles) [20], JACEE (upward triangles) [65] and DAMPE (crosses) [66] are presented. Indirect measurements are also shown from the EAS observatories ARGO-YBJ (downward hollowed triangles) [67], Tibet AS-gamma (open circles) [68] and EAS-TOP (hollowed star) [69]. The gray band around the HAWC data points represents systematic uncertainties. The statistical uncertainties of the HAWC measurements are shown with vertical error bars. The magnitude of the systematic uncertainty in the spectrum after varying the energy scale within systematic errors \(\delta E = \pm 16\%\) is shown in the upper right corner of the plot with arrows.

For this aim, we have fitted the HAWC spectrum below the break with a power-law function, like equation (8), including the systematic uncertainties due to the PMT late light in the procedure. Then, we extrapolated the data up to higher energies and use it to predict the number of expected EAS in the subsample of enriched EAS for the energy interval \(\log_{10}(E/\text{GeV}) = [4.4, 5.2]\). For the estimation we used the expression

\[
N(E_i, E_f) = A_{\text{eff}}(\bar{E}) \Delta \Omega \Phi_0 \frac{E_i^{\gamma_1 + 1} - E_f^{\gamma_1 + 1}}{\gamma_1 + 1},
\]

where \([E_i, E_f]\) is the energy interval of integration, \(\log_{10}(\bar{E}) = [\log_{10}(E_i) + \log_{10}(E_f)]/2\) and \(\Phi_0\) and \(\gamma_1\) are the parameters of the fitted power-law function used in the extrapolation. This calculation gave \((7.57^{+0.65}_{-0.30}) \times 10^8\) events. Next, we calculated the amount of events in the HAWC unfolded energy distribution inside the same energy interval within the systematic errors from the PMT-late-light simulations. This procedure resulted in \((6.15^{+0.65}_{-0.13}) \times 10^8\) events. To end, we compared the expected and the observed amount of events in the true primary energy range from \(\log_{10}(E/\text{GeV}) = 4.4\) up to 5.2, which gave a deficit of \(\sim 2.0 \sigma\) in the data. This result points out that the break can not be explained just

spectrum for light primaries at 24 TeV even after considering the systematic uncertainties due to the PMT-late-light effect in the analysis of the spectral feature.
presented in appendix B.

atic source to the total uncertainty in the energy scale is
detailed estimation of the contribution of each system-
log \text{− } E/\text{GeV}) = [4.4, 5.2] and \(5.1^{+0.95}_{−1.6}\) \times 10^7 events.
Therefore, the differences between the expectations and
the measurements are \(0.8\sigma\) and \(1.1\sigma\), correspondingly.
They are small, however, they seem to indicate a ten-

cision in the light component of cosmic

C. Uncertainties in the energy scale

Associated with the systematic uncertainties of the
spectrum there are uncertainties on the energy scale \(\delta E\),
which can be roughly estimated from the following re-
lation: \(\delta\Phi/\Phi = -(\gamma + 1)\delta E/E\) [70], where \(\gamma\) is the
local value of the spectral index of the energy spectrum.
This procedure gives a total systematic uncertainty in
the energy scale between \(-8.3\%\) and \(+3.5\%\) at the low
energy bin \((\log_{10}(E/\text{GeV}) \sim 3.9)\), which evolves up to
\(-12.4\%\) and \(+15.8\%\), respectively, at high energies (bin
\(\log_{10}(E/\text{GeV}) \sim 5.1\)), as it can be seen in fig. 15. A
detailed estimation of the contribution of each system-
atic source to the total uncertainty in the energy scale is
presented in appendix B.

VII. DISCUSSION

HAWC’s observation of a spectral break in the cosmic
ray spectrum of protons plus helium nuclei at 24 TeV
provides further support to previous results from ATIC-
2 [22], CREAM I-III [23] and NUCLEON [20, 24, 25] in
favor of fine structure in the spectra of light primaries in
the \(10 − 100\) TeV regime. The case is strengthened by
recent DAMPE [18, 19] measurements on the spectra of
H and He nuclei exhibiting significant kinks at energies
close to 14 TeV and 34 TeV, respectively. These results
imply that the break in the energy spectrum of H+He
presented in fig. 11 has its origin in breaks in the in-
dividual spectra of hydrogen and helium nuclei between
10 TeV and 40 TeV.

The TeV structure in the light component of cosmic
rays may be connected with the break in the all-particle
energy spectrum observed at approximately 46 TeV with
HAWC [21] and confirmed by NUCLEON in [20].
The presence of both features in the same energy interval
suggests that the referred feature in the intensity of H+He
could contribute to the structure observed at TeV ener-
gies in the all-particle spectrum. In fig. 11, we have com-
pared the total spectrum of HAWC [21] with our result
for H+He. There are two major differences between the
features: the all-particle spectrum feature is wider and it
is shifted to higher energies. Further research is needed
to find out the reasons. Nevertheless, these facts may
suggest an increasing influence of the heavy component
\((Z > 2)\) of cosmic rays close to \(100\) TeV, which seems
consistent with the heavy element data from NUCLEON
[20, 24, 25], the measurements of HAWC on the mean
shower age (c.f. fig. 4) and the analysis of the efficiency of
the age cut (see fig. 8). The ratio \(\Phi_{H+He}/\Phi_{Tot}\) between
the spectrum of the light nuclei and the total intensity of
cosmic rays measured with HAWC also seems to sup-
port such possibility. As observed in fig. 16, \(\Phi_{H+He}/\Phi_{Tot}\)
FIG. 16. The intensity of the light cosmic ray nuclei obtained in this work divided by the all-particle spectrum of cosmic rays obtained with HAWC in [21] plotted as a function of the primary energy. The vertical error bars represent the total uncertainty.

decreases from 10 TeV to 158 TeV, which suggests an increase in the relative abundance of heavy nuclei in the total spectrum of cosmic rays close to 100 TeV.

The physical interpretation of our result is not yet clear, but it seems to require non-conventional models of production, acceleration and propagation of galactic cosmic rays. In general, it is thought that cosmic rays with energies from TeV to PeV are of galactic origin and that their acceleration and transport in the galaxy occur through diffusive processes driven by magnetic fields. Acceleration up to PeV energies is assumed to take place through the first order Fermi acceleration mechanism in shocked astrophysical plasmas [71–74] of supernova remnants [75, 76] and propagation is believed to occur through scattering on random fluctuations in the interstellar magnetic field [77–79]. The presence of magnetic fields in these processes implies a maximum confinement energy either at the source or at the galaxy and hence, the presence of rigidity dependent cuts in the primary spectra at PeV energies [80], while diffusive shock acceleration predicts a power-law behaviour for the energy spectrum of cosmic ray nuclei from TeV to PeV [1, 76]. As a consequence, it is difficult to understand the HAWC result within this scenario.

Some non-conventional models predict features in the spectra of different nuclei in the TeV energy range, like the one observed by HAWC, and invoke the existence of new kinds of cosmic ray accelerators, nearby sources, or modifications to the standard mechanism of particle acceleration in astrophysical shocks. For instance, in [13, 14] an old supernova remnant (age \( \sim 2 – 3 \) Myr) located close to the Earth at a distance of \( \sim 0(100 \) pc) is postulated as the dominant source of measured 10 – 100 TeV cosmic rays. Its maximum achievable energy is assumed to be around \( Z \times 10 \) TeV. The model leads naturally to breaks in the spectra of cosmic ray nuclei at tens of TeV and could explain the observed feature in the spectrum of proton plus helium primaries. On the other hand, in [15] a phenomenological model is proposed based on the hypothesis that the measured all-particle cosmic ray spectrum from 100 GeV to 100 PeV could be described by assuming the contribution of three different types of sources, each characterized by a power-law cosmic ray spectrum but with distinct magnetic rigidity cutoffs: one at 200 GV, another at 50 TV and the other at 4 PV. The authors associate the first class of cosmic ray accelerators to nova explosions, the second to supernova remnants (SNRs) expanding in the interstellar medium, and the last one to superbubbles. In the model, the second population produce \( H \) and \( He \) spectra with a knee-like feature in the energy range explored in this work. In [81], the possibilities that the above structure is due to the existence of a new population of TeV accelerators or just to a single local source of TeV cosmic rays are explored. They favor a local source, following arguments from [82, 83], supported on data of the phase and dipole anisotropy of galactic cosmic rays. Finally, in [84], TeV features in the spectrum of light cosmic rays appear as a consequence of two new \( H \) and \( He \) contributions with hard spectra accelerated at reverse shocks of SNRs of type II and I, respectively. In this scenario, the new sources contribute to the hardening of the energy spectra of protons and helium at rigidities of 240 GV and to the increase in the \( \Phi_{He}/\Phi_H \) ratio between 100 GeV and 1 TeV as observed by the ATIC-2 [22], CREAM [40, 85] and PAMELA [41] detectors.

In order to distinguish among the predictions of different models that may explain the physical origin of the feature in question, we also must look at the details of the energy spectra of heavier cosmic ray elements in the interval 10 TeV – 1 PeV. NUCLEON, in a recent study [20], has also provided evidence in favor of the existence of rigidity dependent breaks at \( \sim 10 \) TV in the individual spectra of primary cosmic rays with \( Z \geq 3 \). Further research with CALET [86], DAMPE, HAWC, TAIGA-HiSCORE [87] and LHAASO [88] will soon test NUCLEON’s observations and, in turn, provide an opportunity to understand the systematic uncertainties inherent in the direct and indirect cosmic ray detection techniques.

VIII. CONCLUSIONS

The HAWC observatory has measured with high statistics and precision the cosmic ray energy spectrum of \( H \) plus \( He \) in the energy interval from 6 TeV to 158 TeV and confirmed previous hints from the ATIC-2, CREAM I-III and NUCLEON direct detectors that the energy spectrum of light primaries deviate from a plain power-law behaviour in the 10 – 100 TeV energy interval. HAWC result also agrees with DAMPE recent measurements that point out the existence of individual softening in the spec-
tra of protons and helium nuclei at tens of TeV. Hence, HAWC results does not support previous observations from the ARGO-YBJ air shower detector between 3 and 300 TeV, whose spectrum agrees with a simple power-law form [67]. HAWC results find a break in the H + He spectrum of cosmic rays close to 24.0^{+3.6}_{-3.1} TeV, which is produced by what it seems to be a smooth decrease in the spectral index from $\gamma = -2.51 \pm 0.02$ to $\gamma = -2.83 \pm 0.02$. Such a structure was previously reported by HAWC in [26]. Now, it has been confirmed using a larger EAS data set with improved MC simulations of the detector. The break is observed with a statistical significance of 4.1 $\sigma$. Under systematic uncertainties, the feature has a significance of 0.8 $\sigma$. This study demonstrates that research on the composition of cosmic rays is possible with the HAWC detector and opens the door to deeper investigations in the TeV range not only in HAWC but also in other present/future high altitude water Cherenkov observatories.

**ACKNOWLEDGMENTS**

We acknowledge the support from: the US National Science Foundation (NSF); the US Department of Energy Office of High-Energy Physics; the Laboratory Directed Research and Development (LDRD) program of Los Alamos National Laboratory; Consejo Nacional de Ciencia y Tecnología (CONACyT), México, grants 271051, 232656, 260378, 179588, 254964, 258865, 243290, 132197, A1-S-46288, A1-S-22784, cátedras 873, 1563, 341, 323, Red HAWC, México; DGAPA-UNAM grants IG101320, IN111315, IN111716-3, IN111419, IA102019, IN112218; VIEP-BUAP; PIFI 2012, 2013, PROFOCIE 2014, 2015; the University of Wisconsin Alumni Research Foundation; the Institute of Geophysics, Planetary Physics, and Signatures at Los Alamos National Laboratory; Polish Science Centre grant, DEC-2017/27/B/ST9/02272; Coordinación de la Investigación Científica de la Universidad Michoacana; Royal Society - Newton Advanced Fellowship 180385; Generalitat Valenciana, grant CIDEGENT/2018/034. We also acknowledge the significant contributions over many years of Stefan Westerhoff, Gaurang Yodh and Arnulfo Zepeida Domínguez, all deceased members of the HAWC collaboration. Thanks to Scott Delay, Luciano Díaz and Eduardo Murrieta for technical support.

**Appendix A: Composition models**

To study the impact of the uncertainties in the composition of cosmic rays, we used four additional composition models in our study. The first is the Polygonato model, described in [89]. The others were obtained from fits to the spectra of the different mass groups of cosmic rays measured by the ATIC-2 [90], JACEE [65] and MUBEE [91] collaborations. For the individual fits, we used a broken-power law expression similar to the one employed for the nominal composition model in [21]. This formula is

$$
\Phi(E/GeV) = \left\{ \begin{array}{ll}
\Phi_0(E/E_0)^{\gamma_1} & , \ E < E_b \\
\Phi_0(E_b/E_0)^{\gamma_1-\gamma_2}(E/E_0)^{\gamma_2} & , \ E \geq E_b.
\end{array} \right.
$$

(A1)

Here, $\Phi_0$ is a normalization factor at the reference energy $E_0$ and $E_b$ is the energy at which appears the break; $\gamma_1$ and $\gamma_2$ are the spectral indexes of the function before and after the kink. For all models, we chose $E_b(GeV) = 100$ for light primaries and 1200 GeV for heavy ones. The results are shown in table III. The first model, denoted as ATIC-2, was derived from ATIC-2 data [90] between $E = 49$ GeV and 31 TeV. In this case, to get the parameters of the C–Si heavy mass groups, a joint fit was performed to a single value for $\gamma_1$, $\gamma_2$ and $E_b$ in the spectra. The individual normalization factors for each of the heavy elements were treated as free parameters during the fit. To obtain the spectrum for Fe nuclei, the previously fitted $\gamma_2$ and $E_b$ values were substituted in the corresponding broken-power law formula and were fixed during the fit.

The second fitted model, named MUBEE, was obtained from a fit to combined ATIC-2 [90] and MUBEE [91] measurements. Below $E = 10$ TeV only data from ATIC-2 was used, and above, just the measurements from MUBEE. To obtain $\gamma_1$ and $E_b$ for the spectra of C–Si nuclei, we proceeded as before with the difference that to find $\gamma_2$ we decided to divide the heavy data in two mass groups: C–F and Ne–K, and to perform different fits for each of them. For the model called JACEE, we proceeded similarly, except that ATIC-2 data was kept up to $E = 25$ TeV and at higher energies, JACEE measurements were employed [65]. In addition, we also fitted $E_b$ separately for each distinct mass group.

To illustrate the results of the fits, in fig. 17 we display the predictions for the energy spectra of the light and heavy mass groups of cosmic rays and for the ratio of light to heavy cosmic ray nuclei $\Phi_{H+He/\Phi_{Z\geq3}}$ according to each of the composition models, including the nominal one and the Polygonato scenario.

**Appendix B: Statistical and systematic errors**

For the calculation of the statistical error we have considered the following sources of uncertainty:

Statistics of the data: The magnitude of this error is less than 0.03% due to the large number of events in the data. It was calculated by propagating the statistical error from $N(E_{rec})$ according to [62, 63]. We assume that the reconstructed energy bin contents are independent and Poisson-distributed. Let’s define

$$
N_{\mu} = N(E_{\mu}),
$$

(B1)

$$
N_{rec,j} = N(E_{rec,j}),
$$

(B2)

$$
M_{\mu j} = P(E_{\mu}|E_{rec,j}).
$$

(B3)
TABLE III. Values of the parameters of three composition models used in the present analysis. The models were derived from fits with expression (A1) to the ATIC-2 [90], JACEE [65] and MUBEE [91] measurements on the elemental spectra of cosmic rays.

| Model | $\Phi_0$ $[10^{-6} \text{ m}^{-2} \text{s}^{-1} \text{sr}^{-1} \text{GeV}^{-1}]$ | $\gamma_1$ | $\gamma_2$ | $E_b$ [GeV] |
|-------|---------------------------------|---------|---------|---------|
| ATIC-2 | | | | |
| H | $4.40 \times 10^4$ | $-2.86$ | $-2.60$ | $159.6$ |
| He | $2.59 \times 10^4$ | $-2.61$ | $-2.45$ | $1093.6$ |
| C | $6.61$ | $-2.64$ | $-2.48$ | $1125.5$ |
| O | $10.73$ | $-2.64$ | $-2.48$ | $1125.5$ |
| Ne | $2.78$ | $-2.64$ | $-2.48$ | $1125.5$ |
| Mg | $4.72$ | $-2.64$ | $-2.48$ | $1125.5$ |
| Si | $5.34$ | $-2.64$ | $-2.48$ | $1125.5$ |
| Fe | $13.10$ | $-2.61$ | $-2.48$ | $1125.5$ |
| MUBEE | | | | |
| H | $4.44 \times 10^4$ | $-2.72$ | $-2.72$ | $-72$ |
| He | $2.66 \times 10^4$ | $-2.60$ | $-2.63$ | $732.7$ |
| C | $6.41$ | $-2.64$ | $-2.56$ | $31693.5$ |
| O | $10.46$ | $-2.64$ | $-2.56$ | $31693.5$ |
| Ne | $2.46$ | $-2.64$ | $-2.00$ | $31693.5$ |
| Mg | $4.13$ | $-2.64$ | $-2.00$ | $31693.5$ |
| Si | $4.58$ | $-2.64$ | $-2.00$ | $31693.5$ |
| Fe | $13.10$ | $-2.61$ | $-3.00$ | $4283.8$ |
| JACEE | | | | |
| H | $4.39 \times 10^4$ | $-2.80$ | $-2.69$ | $1094.4$ |
| He | $2.67 \times 10^4$ | $-2.60$ | $-2.59$ | $1586.4$ |
| C | $6.35$ | $-2.64$ | $-2.24$ | $13106.9$ |
| O | $10.35$ | $-2.64$ | $-2.24$ | $13106.9$ |
| Ne | $2.46$ | $-2.64$ | $-2.48$ | $31693.5$ |
| Mg | $4.13$ | $-2.64$ | $-2.48$ | $31693.5$ |
| Si | $4.58$ | $-2.64$ | $-2.48$ | $31693.5$ |
| Fe | $13.10$ | $-2.61$ | $-2.51$ | $80717.9$ |

Limited statistics of the MC simulations: The finite size of the MC data contributes with a statistical error within $\pm 3.8\%$ through the response matrix. The uncertainty was computed by error propagation following [62, 63]. The covariance matrix $V_{\text{stat}}$ from which these errors are derived is calculated from the following expression:

$$V_{\text{stat}}[N_{\mu}^i, N_{\nu}^i] = \sum_{\lambda,j} \sum_{\rho,k} \frac{\partial N_{\mu}^i}{\partial P_{j\lambda}} \text{Cov}[P_{j\lambda}, P_{k\rho}] \frac{\partial N_{\nu}^i}{\partial P_{k\rho}},$$

where we have defined

$$P_{j\mu} = P(E_{\text{rec},j}|E_{\mu}).$$

and

$$\frac{\partial N_{\mu}^i}{\partial P_{j\lambda}} = \left[ N_{\mu}^{i-1} \frac{\partial}{\partial P_{j\lambda}} - M_{\mu,j}N_{\lambda}^{i-1} \right] \frac{N_{\text{rec},j}}{\sum_{\sigma} P_{j\sigma} N_{\sigma}^{i-1}} + \frac{N_{\mu}^{i-1}}{N_{\mu}^i} \frac{\partial N_{\nu}^{i-1}}{\partial P_{j\lambda}} - \sum_{\sigma,k} \left[ N_{\text{rec},k} M_{\mu,k} M_{\sigma,k} \frac{\partial N_{\nu}^{i-1}}{\partial P_{j\lambda}} \right].$$

In formula (B7), Cov[$P_{j\lambda}, P_{k\rho}$] represents the covariance matrix for the bins of the response matrix. It is different from zero in the following cases [51]:

$$\text{Cov}[P_{j\lambda}, P_{k\rho}] = \left\{ \begin{array}{cl} P_{j\lambda} [1 - P_{k\lambda}] / \bar{N}_{MC,\lambda} & : j = k \\ -P_{j\lambda} P_{k\lambda} / \bar{N}_{MC,\lambda} & : j \neq k \end{array} \right.,$$

where $\bar{N}_{MC,\lambda} = (\sum_{k=1} w_{k\lambda})^2 / (\sum_{k=1} w_{k\lambda}^2)$ is the equivalent number of unweighted events inside the bin $E_\lambda$ of the true MC energy distribution used in the construction of the response matrix. Here, the sum runs over the number of simulated events in the bin and $w_{k\lambda}$ denotes the weight of each of these events. In case that $i = 0$, we have $\partial N_{\mu}^0 / \partial P_{j\lambda} = 0$.

On the other hand, the systematic uncertainty was calculated by summing in quadrature the error sources below; in each case we cite the effect on the intensity of the $H + \text{He}$ spectrum, and took the systematic error for each source as the size of the observed effect:

Unfolding algorithm: Its uncertainty is found between $-1.1\%$ and $+1.2\%$. It was evaluated by comparing the experimental result obtained with the Bayesian procedure with that using the Gold’s algorithm [52] as implemented in [53].

In Gold’s procedure [52], a real diagonal matrix $D$ is found iteratively, which allows to estimate the unfolded histogram $N(E)$ from the data by using the equation

$$N(E) = DN(E_{\text{rec}}).$$

with

$$\text{diag}(D) = \frac{P(E)}{\sum_{E'} P(E_{\text{rec}}, E') P(E')},$$

where $N(E) = D N(E_{\text{rec}})$.
Here, $P(E)$ is the probability distribution of the unfolded histogram at the previous iteration level.

Now, in order to guarantee real positive solutions and to take into account the statistical uncertainties of the data [53], we replaced $N(E_{\text{rec}})$ by

$$N'(E_{\text{rec}}) = [P(E_{\text{rec}}|E)]^T C C N(E_{\text{rec}}), \quad (B12)$$

and $P(E_{\text{rec}}|E)$ by the matrix

$$P' = [P(E_{\text{rec}}|E)]^T C C P(E_{\text{rec}}|E), \quad (B13)$$

where $C$ is a diagonal matrix, whose matrix elements on the main diagonal are equal to the inverse of the statistical uncertainties of the measurements $N(E_{\text{rec}})$.

Seed for the unfolding method: To estimate the error of the unfolded result due to the initial energy distribution used in the Bayesian algorithm, we have repeated the unfolded procedure with two distinct priors: a uniform distribution, as suggested by [51], and an $E^{-1.5}$ distribution. The power-law choice that was employed matches the all-particle energy histogram measured with HAWC for $10 \text{ TeV} \leq E \lesssim 46 \text{ TeV}$ [21], which may be expected to be closer to the true distribution of H plus He primaries, since this region seems to be dominated by the light mass group of cosmic rays [20, 22, 23]. By comparing the resulting spectra with the one of reference obtained in this work (see fig. 11), we found a bias in the intensity within $-1.4\%$ and $+0.7\%$. We used that as the systematic error due to the seed in the unfolding algorithm.

Smoothing procedure in unfolding algorithm: The employment of a broken power-law in the smoothing procedure of the unfolding analysis was done to achieve a fast convergence of the result. However, this procedure may introduce a systematic error in the unfolded spectrum. To compute this uncertainty, we smoothed the unfolded distributions with two alternative functions: a fifth degree polynomial and the 353HQ-twice smoothing algorithm [55] as installed in ROOT [56]. The resulting spectra were then compared with the one of reference obtained with the Bayesian method. We found differences that range from $-2.5\%$ to $+3.7\%$, which we used as the corresponding systematic error.

Corrected effective area: The systematic uncertainty in $A_{\text{eff}}(E)$ was evaluated using MC simulations. By varying $A_{\text{eff}}(E)$ inside its allowed limits, we observed a variation in the energy spectrum from $-2.1\%$ to $+2.2\%$.

Position of the age cut: We moved the cut to the expected line for the mean age of He events and then, to the mean age of C events (see fig. 4). Using the cut on $s_C$, the spectrum is increased by at most $+3.7\%$, while by setting the cut on $s_{He}$, a decrease of up to $-6.6\%$ is observed.

Cosmic ray composition uncertainty: The dependence on the primary cosmic ray abundances enters through the response matrix and the corrected effective area, which are computed with MC simulations. We replaced our nominal model with the four alternative composition scenarios of appendix A and repeated the analysis in each case. In addition, we have also considered the uncertainty in the relative abundance of heavy nuclei observed in the analysis of the efficiency of the age cut (c.f fig. 8) presented in section V. To estimate its effect on the unfolded spectrum, we have multiplied by a factor of two the intensity of heavy elements in our cosmic ray composition models, since, in this case, model predictions for the efficiency of the shower age cut are closer to the measured...
value. Then we have repeated the unfolding analysis with the new models. The maximum and minimum differences of the results obtained with the different scenarios for the composition of cosmic rays with respect to the experimental spectrum of reference were recorded as systematic uncertainties. At the bin $\log_{10}(E/\text{GeV}) = 3.9$, the systematic error has a value of $-11.6\%$, while for the bin $\log_{10}(E/\text{GeV}) = 5.1$, it lies within $+2.1\%/-17.3\%$. In general, when using models with heavier (lighter) relative abundances than the nominal one, the magnitude of the spectrum decreases (increases) due to the larger (smaller) correction factors. In case of the ATIC-2 model, which have a lighter abundance, we observe a small decrement close to 10 TeV that is due to a compensation from the response matrix. This effect is related to the hard spectrum of the light component employed to construct the above matrix.

**PMT charge resolution:** In the nominal MC data set, a PMT charge uncertainty of 10% is used [27]. To evaluate the impact of this parameter on the final spectrum, the PMT charge resolution was varied within its allowed interval [0%, 15%] (see [27]). Changes of the H + He spectrum were between $-2.6\%$ and $+1.8\%$.

**PMT-late-light simulation:** Air showers have a broader time distribution than the laser pulses employed for calibration in HAWC [27, 30], consequently there appears a systematic error associated with the calibration of the effective charge produced by the late light during the EAS event, which is important for high values of $Q_{\text{eff}}$ (> 50 $\text{PEs}$). The effect of this systematic source leads to an overestimation of the charge, since broader pulses have a longer Time-over-Threshold. In simulations, to take into account the effect of the PMT late light, a linear correction is added in logarithmic space to the effective charge. The value of the correction is the same for all PMTs and it increases from zero at $\log_{10}(Q_{\text{eff}}) = 1.25$ up to 0.1 at around $\log_{10}(Q_{\text{eff}}) = 2.25$. To estimate the uncertainty of the PMT late light, the value of the correction at the upper limit of $\log_{10}(Q_{\text{eff}})$ is varied between 0.5 and 1.25. The limits are quite conservative and they are chosen in such a way that allow us to describe the measured charge distribution for triggered events. On the other hand, we have also included the case for zero correction in the study, assuming that the difference between the MC simulations and the data is due to deficiencies of the hadronic interaction model. A smaller value of the correction of the PMT-late-light effect in the simulation tends to flatten the energy spectrum, however, even with zero correction the spectral feature can be observed (c.f. section VI). We varied the late light effect within the above range and found that it produces a systematic effect from $-9.2\%$ to $+28.3\%$ in the energy spectrum.

**PMT threshold:** The impact of the uncertainty in the minimum detectable charge at the PMTs of HAWC has been also evaluated. For this aim, the nominal value (which is of the order of 0.2 $\text{PE}$) was varied within the error interval of $\pm0.05\text{PE}$, which was found from a HAWC calibration study based on vertical muon data [27]. Correspondingly, we observed a bias of $-1.96\%/+2.3\%$ in the magnitude of the spectrum.

**PMT efficiency and its temporal evolution:** We estimated this uncertainty by using distinct MC simulations,
FIG. 19. Energy dependence of the relative statistical and systematic uncertainties of the energy spectrum presented in this work. The total error is calculated as the sum in quadrature of the statistical and systematic uncertainties. For a description of each uncertainty source, see appendix B. On the plot of the right, systematic uncertainties due to smoothing in unfolding procedure, seed for the unfolding method and unfolding algorithm were added in quadrature and are shown as a single contribution with a dotted line.

FIG. 20. The total systematic uncertainties on the energy scale against the primary energy. The results are shown along with the contributions from the different systematic sources listed in appendix B.

which incorporate the measurements of the individual PMT efficiencies in HAWC at different moments during the data taking period [27] (in particular, on September 2015, April and July 2016, February and June 2017 and February 2018). This procedure allowed us to determine the uncertainties in the efficiency of the PMTs, which can change with the time due to possible aging effects. The detector layouts of the active PMTs in HAWC registered during these sampling epochs were also incorporated into these simulations in order to study the influence of PMTs that are removed for maintenance or are not active in HAWC during the data taking periods, because the MC simulations used in our nominal analysis were carried out for an ideal situation where all PMTs are working. Using these MC data sets, we found that the corresponding error in the energy spectrum ranges from $-4.1\%$ to $+5.0\%$.

Hadronic interaction model: To evaluate the influence of the uncertainties in the physics of the high energy hadronic collisions, we produced a small set of MC simulations using EPOS-LHC and repeated the analysis procedure. Following section III, we generated $4 \times 10^5$ MC events for each of the following nuclei: H, He and C, and $2 \times 10^5$ MC showers for each of the other primaries: O, Ne, Mg, Si and Fe. The mean lateral shower age as computed
with EPOS-LHC for some selected cosmic ray primaries is presented in fig. 18, left, as a function of $E_{\text{rec}}$ in comparison with the measured data. In this figure, we also display the line that represents the $s_{\text{He-C}}$ cut as calculated with the EPOS model. Fig. 18, right shows the energy spectrum of H+He as estimated using HAWC data calibrated with EPOS-LHC compared with QGSJET-II-04. The magnitude of the spectrum decreases when using the EPOS model. At $\log_{10}(E/\text{GeV}) = 3.9$ the error is $\sim -3.7\%$. The minimum value of the error (10.9%) was found at $\log_{10}(E/\text{GeV}) = 4.9$.

Finally, fig. 19 shows the relative uncertainty of different sources of statistical and systematic errors as a function of the primary energy, along with the total fractional uncertainty. The dominant contributions to the error come from the uncertainty in the PMT-late-light simulation, the hadronic interaction model and uncertainties in the cosmic ray composition. To end this section, we have estimated the systematic uncertainties in the energy scale associated with the different systematic sources listed in this section. We have employed formula $\delta \Phi/\Phi = -(\gamma + 1)\delta E/E$ [70], where $\gamma$ is the local value of the spectral index in the energy spectrum as obtained from the fit to the data with equation (9). The results are shown in the plots of fig. 20.

**Appendix C: Systematic checks**

We performed several checks to see whether systematic effects could produce the observed change in slope in our H + He spectrum.

First, we tested the reliability of the reconstruction method by applying it to MC simulations produced with QGSJET-II-04, which we treated as fake experimental data. The idea behind this test was to check that the analysis procedure does not introduce artificial features in the spectrum under study and that the reconstructed spectrum of light primaries is in agreement with the true one within systematic uncertainties. As input data, we have used the different composition models of appendix A, which predict distinct $\Phi_{\text{H+He}}/\Phi_{Z \geq 3}$ ratios. For each test, we used two alternative spectra of H+He: a single power-law formula or a broken power-law behavior, the latter with the same change in spectral index and position of the spectral feature as observed in the measured spectrum. In all cases, we found that the shape of the studied spectrum did not have any dramatic modifications due to the analysis method (see, for instance, fig. 21) and that the reconstructed spectrum agrees with the true one within the systematic errors. In each test, we quantified the systematic uncertainties listed in appendix B, but the one corresponding to the hadronic interaction model.

On the other hand, it is interesting to point out that the correction factor $f_{\text{corr}}$ has a strong feature at
puling the standard age cut

tions and the optimized age cut based on the maximization
estimated with eq. (6) for HAWC using QGSJET-II-04 simula-
FIG. 22. Predictions of the QGSJET-II-04 model for the en-
ergy dependence of the mean lateral age in vertical air showers
itated by four cosmic ray species at HAWC. From top to
bottom, the MC points correspond to Fe (solid triangles), C
(hollowed triangles), He (hollowed circles) and H (solid cir-
icles) primaries, respectively. For clarity, not all the elemental
nuclei simulated in this work were included in the plot. The
standard cut \( s_{He-C} \) employed to extract the enriched sub-
sample of light nuclei used for the main analysis of the paper
is plotted using a gray dotted line, while the optimized age

cut based on the maximization of the \( Q \) factor is shown with a
dashed black line.

log
\(_{10}(E/\text{GeV}) = 4.5\), close to the location of the break
in the measured spectrum of \( H+\text{He} \). The factor \( f_{corr} \)
does not contribute to the formation of the cutoff in the
spectrum, on the contrary it tends to flatten the feature,
which appears already in the unfolded distribution. If
we assume that the feature in the spectrum is due to an
underestimation of the relative abundance of the heavy
mass group and, thereby, of \( f_{corr} \) by the cosmic ray com-
position models, we would need an enormous amount of
heavy nuclei in the models around \( \log_{10}(E/\text{GeV}) = 4.5 \)
because of the small fraction of heavy primaries in the
selected data subset. This particular overabundance of
heavy primaries would be in contradiction with direct
measurements and with HAWC data on the shower age.
Hence, this possibility is discarded. In consequence,
the origin of the break in the spectrum of light primaries
is not connected to the feature of the correction factor ob-
served at \( \log_{10}(E/\text{GeV}) = 4.5 \).

We also performed other systematic checks, which are
not included as systematic errors, but that are impor-
tant to discard that the observed feature in the measured
spectrum is induced by the reconstruction method.

We started by studying the possibility that an un-
known spectral break in the intensity for heavy primaries
produces the above mentioned feature. To rule out that
scenario, we introduced a cut at \( E_c = 24\text{ TeV} \) with a
change of spectral index \( \Delta \gamma = -0.32 \) in the spectrum of
the heavy component of our MC composition models and
kept the single power-law behaviour for the intensity of
light elements unchanged. The unfolded results produced
a small bump in the light spectrum, but with a \( \Delta \gamma \) too
small (\( \lesssim -0.03 \)) to explain the observations. A sharp cut

FIG. 23. The corrected effective area for light primaries es-
timated with eq. (6) for HAWC using QGSJET-II-04 simu-
lations and the optimized age cut based on the maximization
of the \( Q \) factor (black data points). For comparison, also the
corrected effective area for light primaries obtained after ap-
plying the standard age cut \( s_{He-C} \) is displayed (gray open
squares). Statistical uncertainties are smaller than the size of
the data points.

FIG. 24. The energy spectrum of protons plus helium cosmic
ray nuclei as measured with HAWC and reconstructed within
the framework of QGSJET-II-04 using a subset of data se-
lected with the standard age cut of fig. 4 (gray open squares)
and another experimental subsample obtained by applying
the optimized cut of fig. 22 (black circles). Vertical error bars
represent the statistical uncertainties.
at $E_c$ has been discarded as it would be in contradiction with the measured shower age distributions, but when investigated anyway, we observed a bump in the light mass group spectrum with $\Delta \gamma \lesssim -0.01$. A recovery in the intensity of the heavy component of cosmic rays at $E_c = 24$ TeV was also investigated. We found that it can neither explain the observations, as it would produce a positive change in $\Delta \gamma$ in the spectrum of light primaries in contradiction with the HAWC measurement.

We also ruled out systematic effects from either the calibration of large induced signals or the cut $Q_{\text{eff}} \gtrsim 10^4$ PE as the reason behind the slope change. That conclusion was achieved by reconstructing the spectrum of light primaries for inclined air showers (in particular, with $\theta$ close to $45^\circ$) and by observing that the break in the spectrum is still present for events with large zenith angles. These EAS suffer a stronger attenuation in the atmosphere and, in consequence, have lower $Q_{\text{eff}}$ values than vertical EAS. Therefore the calibration errors are expected to be smaller and so decrease the effects from the $Q_{\text{eff}}$ cut on the composition analysis. This point offers the possibility of extending HAWC studies on the composition of cosmic rays up to 1 PeV using inclined EAS. A complete study with inclined air shower events is in progress and will be presented in an upcoming paper.

As a systematic check, we have also reconstructed the energy spectrum for H+He primaries from a data subset obtained by applying an age cut derived from the maximization of the purity of the subsample and by using the unfolding procedure described in section V. For maximizing the purity of the data subset, we define the $Q$ factor. This parameter is just the ratio between the number of remaining events of light nuclei (signal) obtained after applying the cut to the squared root of the number of heavy primaries (background) that passed the age selection. Then we found the maximum value of the $Q$ factor for each bin of reconstructed energy with the constraint that we keep a retention factor of at least 50% for H and 50% for He in each bin. The optimized shower age cut obtained with this procedure is presented in figure 22 in comparison with the selection cut $s_{He-C}$ used in our standard analysis and with the QGSJET-II-04 predictions for the mean shower age of different primary nuclei. From fig. 22, we observed that the optimized cut and the standard one have a similar tendency up to $10^5$ GeV, but at higher energies they have distinct behaviours. With the optimized cut the purity of the subsample for $\log_{10}(E_{\text{rec}}/\text{GeV}) \lesssim 5.2$ is larger than the one achieved with the standard age cut. For the optimized selection, it is $> 88.5\%$, while for the standard cut is $> 86\%$. At higher energies, the optimized age cut has a poorer performance. In this case, the purity of the selected data subset can decrease up to a value of 78%, while for the standard selection is reduced up to 82%.

Using the selected data set with the optimized cut and the unfolding procedure presented in section V, we have obtained the corresponding energy spectrum of light cosmic ray primaries, which is displayed in fig. 24 together with the standard result. The corrected effective area employed for the computation of the spectrum with the optimized cut is shown in fig. 23, in comparison with $A_{\text{eff}}(E)$ for the standard data subsample. We can see that the effective area corresponding to the optimized cut it is not flat. Despite of that the energy spectrum of H+He has the same shape whether we use the standard age cut or the optimized one, as we can see from fig. 24. Below 20 TeV, both spectra are in agreement. However, we observe that at higher energies the spectrum with the optimized cut is softer than the original one and that its intensity is smaller by at most 8%.

In addition, we have also investigated the impact of seasonal variations in the data. For this analysis, we have divided our data in four subsets corresponding to different periods of the year, which cover the following months: March-May, June-August, September-November and December-February. For each period, we reconstructed the energy spectra of H+He and compared them with our nominal result. We found negligible variations with regard to the spectrum of fig. 11. At low energies ($\log_{10}(E/\text{GeV}) = 4.9$), the changes in the cosmic ray intensity due to seasonal variations are of the order of $+0.7\%/-0.5\%$. They increase with the primary energy and reach values up to $+1.26\%/-0.85\%$ for the bin $\log_{10}(E/\text{GeV}) = 5.1$.

We also studied the influence of the uncertainties in the EAS core position on the energy spectrum by using only cores reconstructed inside the array; again the slope change did not disappear. Finally, we investigated the combined effect of fluctuations in the signals at the lateral distribution of EAS along with the effects of shower core resolution, but that is not able neither to explain the presence of the spectral break.

[1] S. Mollerach and E. Roulet, Prog. Part. Nucl. Phys. 98, 85 (2018).
[2] M. Kachelriess and D. V. Semikoz, Prog. Part. Nucl. Phys. 109, 103710 (2019).
[3] J. J. Beatty, J. Matthews, and S. P. Wakely, Phys. Rev. D 98, 030001 (chapter 29) (2018).
[4] T. Gaisser, EPJ Web of Conferences 145, 18003 (2017).
[5] J. Bluemer, R. Engel, and J. R. Hoerandel, Prog. Part. Nucl. Phys. 63, 293 (2009).
[6] A. Haungs, R. Heinigerd, and R. Markus, Rep. Prog. Phys. 66, 1145 (2003).
[7] P. Maestro, in Proceedings of the 34th ICRC (the Hague, Netherlands), PoS (ICRC2015) No. 016 (2015).
[8] P. Picozza, P. Spillantini, and L. Marcellia, Nucl. Part. Phys. Proc. 297-299, 207 (2018).
[9] B. Bartoli et al. (ARGO-YBJ Collaboration), Phys. Rev. D 92, 092005 (2015).
[10] J. R. Hoerandel, Astropart. Phys. 21, 241 (2004).
Phys. 21, 223 (2004).
[70] P. Abreu et al. (Pierre Auger Collaboration), Eur. Phys. J. C 81, 966 (2021).
[71] E. L. W. I. Axford and G. Skadron, in Proceedings of the 15th ICRC (Plovdiv, Bulgaria), Vol. 11 (1977) p. 132.
[72] G. F. Krymsky, Dokl. Akad. Nauk. SSSR 243, 1306 (1977).
[73] A. R. Bell, Mon. Not. R. Astron. Soc. 182, 147 (1978).
[74] R. D. Blandford and J. P. Ostriker, Astrophys. J. Lett. 221, L29 (1978).
[75] W. Baade and F. Zwicky, Phys. Rev. 46, 76 (1934).
[76] A. R. Bell, Astropart. Phys. 43, 56 (2013).
[77] C. Cesarsky, Ann. Rev. Astron. Astrophys. 18, 289 (1980).
[78] V. S. Ptuskin et al., Astron. Astroph. 268, 726 (1993).
[79] A. W. Strong, I. V. Moskalenko, and V. S. Ptuskin, Annu. Rev. Nucl. Part. Sci. 57, 285 (2007).
[80] B. Peters, Il Nuovo Cimento XXII, 800 (1961).
[81] C. Yue et al., Front. Phys. 15, 24601 (2020).
[82] V. Savchenko, M. Kachelriess, and D. V. Semikoz, Astrophys. J. Lett. 809, L23 (2015).
[83] W. Liu, Y. Q. Guo, and Q. Yuan, JCAP 10, 010.
[84] V. Ptuskin, V. Zirakashvili, and E. S. Seo, Astrophys. J. 763, 47 (2013).
[85] H. S. Ahn et al. (CREAM Collaboration), Astrophys. J. Lett. 714, 89 (2010).
[86] O. Adriani et al. (CALET Collaboration), Phys. Rev. Lett. 122, 181102 (2019).
[87] V. V. Prosin et al. (TAIGA Collaboration), Bull. Russ. Acad. Sci. Phys. 83, No. 8, 1016 (2019).
[88] X. Bai et al. (LHAASO Collaboration), e-print arXiv:astro-ph.HE/1905.02773 (2019).
[89] J. R. Hoerandel, Astropart. Phys. 19, 193 (2003).
[90] A. D. Panov et al. (ATIC-2 Collaboration), Bull. Russ. Acad. Sci. Phys. 71, 494 (2007).
[91] V. I. Zatsepin et al. (MUBEE Collaboration), in Proceedings of the 23rd ICRC (Calgary, Canada), Vol. 2, No. 13, edited by D. A. Leahy, R. B. Hickws, and D. Venkatesan (World Scientific, 1993).
[92] T. Adye, in Proceedings of the PHYSTAT 2011 Workshop (CERN, Geneva, Switzerland), CERN-2011-006 (2011) p. 313.
[93] T. Pierog, I. Karpenko, J. M. Katzy, E. Yatsenko, and K. Werner, Phys. Rev. C 92, 034906 (2015).