Available online at www.sciencedirect.com
SCIENCE DIRECT®
Astroparticle Physics 22 (2004) 81–94
Astroparticle Physics
www.elsevier.com/locate/astropart

The primary proton spectrum in the TeV region as seen by a next generation all-sky telescope

S. Bussino *, E. De Marinis, S.M. Mari

Dipartimento di Fisica, Università di Roma Tre and INFN, Sezione di Roma Tre, Via della Vasca Navale 84, I-00146 Rome, Italy

Received 4 May 2002; received in revised form 3 December 2003; accepted 15 May 2004
Available online 19 June 2004

Abstract

The next-generation of all-sky EAS detector will be able to lower the threshold below the TeV region and will provide a huge data flux, due to the large detecting area and the high duty cycle. In this paper we discuss in detail the possibility of using this detector in order to study the cosmic ray spectrum in the TeV region. The peculiarities of this class of next-generation detectors will make it possible to study the energy profile in the region where direct measurements are available with a large statistical significance. In such a way the measurement obtained by an indirect technique can be fully overlapped with direct measurement data, in order not only to study the cosmic ray spectrum but also to investigate the reliability of the full Monte Carlo simulation process. The sensitivity of such a measurement is also discussed and the data taking time required for a significative measurement is estimated. A standard chi-square fit and a Bayesian unfolding procedure are applied in order to obtain a measurement of the spectral index of the incoming primaries.

Keywords: EAS; Cosmic rays; All nucleon spectrum; Unfolding technique

1. Introduction

The study of the properties of cosmic rays is universally recognized as an intriguing problem, in particular concerning their origin and the energetic behavior in the region of the "knee". The spectrum of the cosmic rays spans an huge interval up to 10^{20} GeV or more [1,2]. Many experimental efforts were devoted to studying the properties of cosmic rays. The observation techniques can be grouped into two broad classes: direct and indirect measurements. Direct measurements can access only the low-energy region, up to few 100 TeV, because of the limited dimensions and the short exposure time of the detector. The indirect techniques can investigate higher and higher energies, up to 10^{20} eV, but they need a Monte Carlo simulation in order to unfold the relevant information from the measured data. Actually, indirect measurements in the region of the TeV seem not to agree completely with the data collected by space-based experiments [3], and a complete

* Corresponding author.
E-mail addresses: bussino@fis.uniroma3.it (S. Bussino), demarinis@fis.uniroma3.it (E. De Marinis), mari@fis.uniroma3.it (S.M. Mari).

0927-6505/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.astropartphys.2004.05.008

understanding of this topic can be obtained only by extending the indirect measurements in the low energy region covered by balloons and satellite detectors [4]. The scientific relevance for this analysis is the possibility of a better understanding of this disagreement, in order to investigate its origin and to check the consistency of the simulation model through which the indirect data are unfolded. The next generation of all-sky EAS detectors [5,6], with low energy threshold and high duty cycle, will be able to collect a lot of data and to fully overlap their measurement to a wide region covered by direct measurements. In the following sections of this paper the main characteristics of a general detector will be described briefly. The method of analysis will then be discussed with particular attention to the physics and computational tools required to reach the final result. Later the Monte Carlo data sample will be briefly described and finally, in the last two sections, the sensitivity of the detector will be evaluated and two different techniques to unfold the experimental data will be discussed.

2. An ideal detector

The detector we have in mind is a next generation Extensive Air Shower apparatus [7–9], with an high duty cycle and a full coverage. This family of detectors, especially devoted to study gamma ray sources, is able to furnish the electronic image of each event. We refer, as discussed in a previous paper [10], to a pixel-like detector, with each active element (scintillator, streamer tubes, RPC, and so on) of the order 50×50 cm^2. We have considered in our computation a continuous square made of 20,000 logical pixels (corresponding to about 70×70 m^2) and with a dead space of the order of 5%. We suppose that the time resolution of the single detector is of the order of 1 ns, which corresponds, for an apparatus of these dimensions, to an angular resolution better than half degree. A majority trigger has been chosen as the trigger of this detector. As we stated before, these requirements fit very well with the conceptual design of a new generation shower array detector [7]; however, the results can be modified or scaled over the more realistic structure of existing [6] or planned detectors [11,12].

3. Fundamental relations

As is well known, each shower is dominated by fluctuations, mainly due to the fluctuations in the point of the first interaction and on each following interaction. For this reason, the energy distribution of the incoming particles cannot be obtained event by event, but must be evaluated by means of an unfolding procedure. The observables quantity in each pixel-like detector is $M(K)$, which represents the number of events with a given number K of fired-pixels, collected in a unit period $\Delta T = 1$ s and accepted within a solid angle Ω. The number of primary cosmic rays can be expressed in the usual way as a power law:

$$\frac{\mathrm{d}N(E)}{\mathrm{d}E} = N_0 E^{-\gamma}. \tag{1}$$

The number of detected event is given by an integral of the previous equation. The integration has to be performed on the energy spectrum of the primary particles and on the overall acceptance of the detector:

$$M(K) = \int_{E_1}^{E_2} \int_{\Omega} A_{\mathrm{eff}}(E, K, \Omega') N_0 E^{-\gamma} \mathrm{d}\Omega' \, \mathrm{d}E, \tag{2}$$

where $A_{\mathrm{eff}}(E, K, \Omega)$ is the effective area of the detector for a given energy E of the primary particle and for a well defined event with K fired-pixels and detected around the solid angle Ω. The dependency of the effective area A_{eff} on the solid angle Ω can be hidden by considering the mean value within the observation solid angle Ω:

$$A_{\mathrm{eff}}(E, K) = \frac{1}{\Omega} \int_{\Omega} A_{\mathrm{eff}}(E, K, \Omega') \, \mathrm{d}\Omega'. \tag{3}$$

This equation makes it possible to express the effective area in a compact way, taking into account all the geometrical dependencies. Inserting it in Eq. 2 we obtain:

$$M(K) = \Omega \int_{E_1}^{E_2} A_{\mathrm{eff}}(E, K) N_0 E^{-\gamma} \, \mathrm{d}E, \tag{4}$$

this equation represents the connection between the observables quantity $M(K)$ and the value to be measured, N_0 and γ.

4. The method

Looking at Eq. 4, the left side of the equation is the measured quantity, related to good events selected requiring a set of fiducial cuts, for example: zenith angle less than 30°, reconstructed core position within a fiducial area in the detector, good chi-square value on the angular reconstruction, minimum number of fired-pixel required to accept an event, and so on. The right side of this equation is a convolution of known and unknown quantities:

(i) the $A_{\text{eff}}(E,K)$, which can be obtained from a Monte Carlo simulation;
(ii) the two values N_0 and γ which are the relevant quantities to be extracted from data.

The method of analysis requires a Monte Carlo simulation to compute the $A_{\text{eff}}(E,K)$ and a mathematical procedure in order to obtain the two values N_0 and γ. The Monte Carlo simulation must be structured around a standard procedure: a shower code generator and a simulation of the interaction of the particles with the detector. An analysis program will then be used in order to reconstruct the relevant parameters from the simulated data. The Monte Carlo generation will be described in detail in the following paragraph. One half of the Monte Carlo data sample will be used to compute the effective area $A_{\text{eff}}(E,K)$, while the other half will be used to test the sensitivity of the full analysis chain, by back-obtaining the value of the index γ and the absolute normalization N_0 used to simulate the events. Two different mathematical procedures will be discussed to deal with the second point:

(i) a classical fitting procedure, in which N_0 and γ are the two quantities to be obtained by fitting the experimental $M(K)$ by means of the function described by equation 4;
(ii) a Bayesian unfolding procedure, in order to obtain the energy distribution of the primary particle.

Both methods are well-known in literature: the first one, for example, was used by underground experiment to obtain the all-nucleon-spectrum [13], while the second one can count on a lot of physical applications [14–18].

5. The Monte Carlo data sample

The events were generated by using Corsika [19] code 5.62 which provides a complete simulation of the shower development in the earth's atmosphere. The electromagnetic part of the shower simulation is implemented by mean of the EGS4 code [20,21], while for the hadronic component several options are available: the Monte Carlo data generated for this paper were processed by QGSJET code [22] for the high energy hadronic interactions and by GHEISHA code [23] for the low energy hadronic interactions [24]. The data were generated in the energy range 1–100 TeV with an energy distribution given by

$$N(E)\mathrm{d}E = N_0 E^{-1}\,\mathrm{d}E \tag{5}$$

and in a cone of 15° around the vertical direction: the same geometrical cut will be applied to real data. The energy spectrum of Eq. (5) has a constant number of events in energy bins of the same width in a logarithmic scale and reduce the statistical error in the high-energy bins, avoiding an accumulation of simulated data in the low-energy bins. During the analysis, each event will be weighted with the correct factor in order to obtain the required energy spectrum at the production. The number of generated event is reported in Table 1, for three different energy intervals. The same table also shows the weighted number of events corresponding to three different values of the spectral index γ, chosen around the mean value 2.80 ± 0.4 measured by the JACEE collaboration [25].

The use of Eq. (5) to simulate a sample of events makes it possible a reduction of the statistical errors in the higher energy regions: from Table 1 it follows that in order to obtain the same statistical errors in the energy range 10–30 TeV with the ordinary $\gamma = 2.80$ spectral index, a more than 10 times larger simulation would be required. The

Table 1
The number of simulated events is shown for different energy ranges

Spectral index	Total number	1–3 TeV	3–10 TeV	10–30 TeV	30–100 TeV
$\gamma = 1$	540,000	$128{,}823 \pm 360$	$141{,}177 \pm 375$	$128{,}823 \pm 360$	$141{,}177 \pm 375$
$\gamma = 2.76$	66,603	$56{,}988 \pm 180$	8478 ± 26	990 ± 3	147 ± 1
$\gamma = 2.80$	65,128	$56{,}127 \pm 178$	7984 ± 25	890 ± 3	127 ± 1
$\gamma = 2.84$	63,714	$55{,}286 \pm 177$	7520 ± 24	799 ± 3	109 ± 1

The total number of weighted events is also shown, assuming three different values of the spectral index γ.

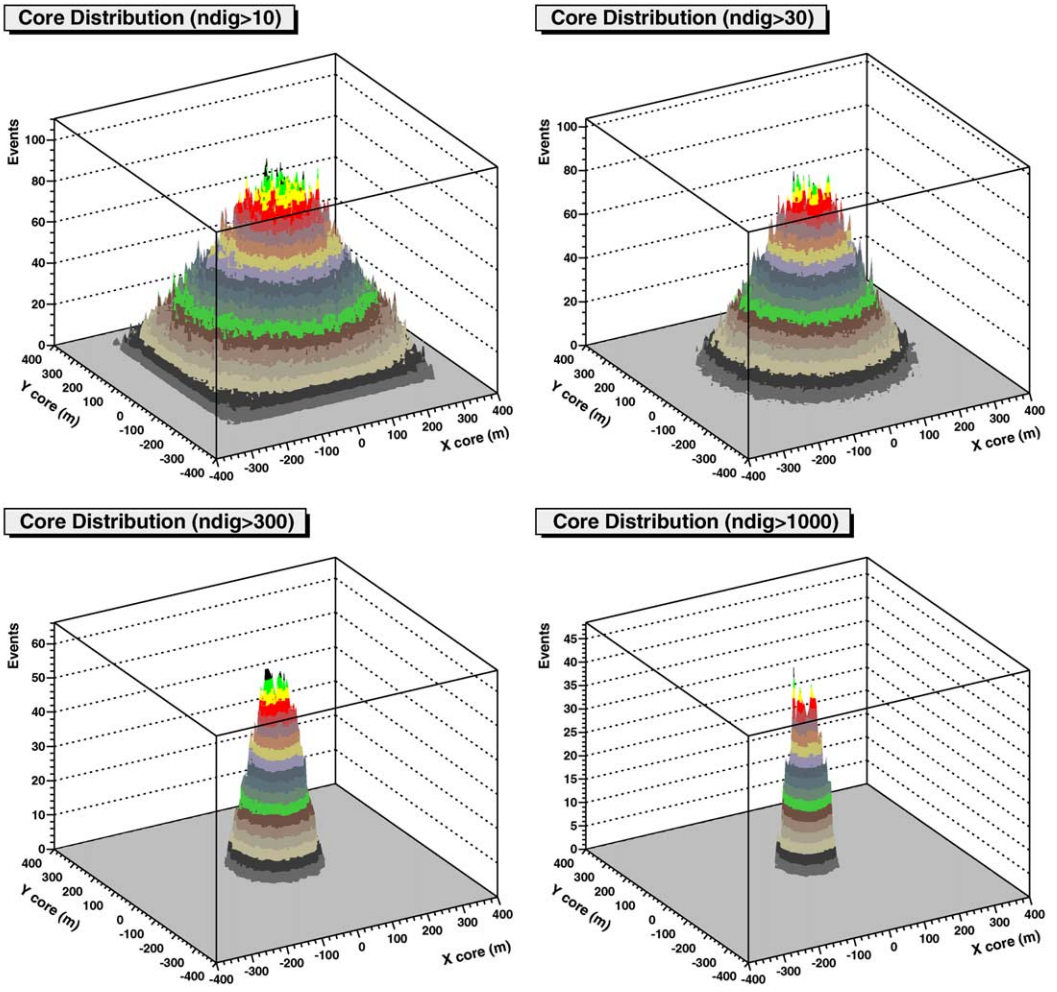

Fig. 1. The core position distribution for different multiplicity (*ndig*) ranges is shown in the upper panel, for the data sample generated according with Eq. 5. For the same sample, the distribution of the core position in one coordinate, for events with multiplicity greater than 30 fired-pixels, is also shown in the lower panel: the left scale refers to the total number of events, while the right scale shows the fraction of events with the core within a given distance from the detector center.

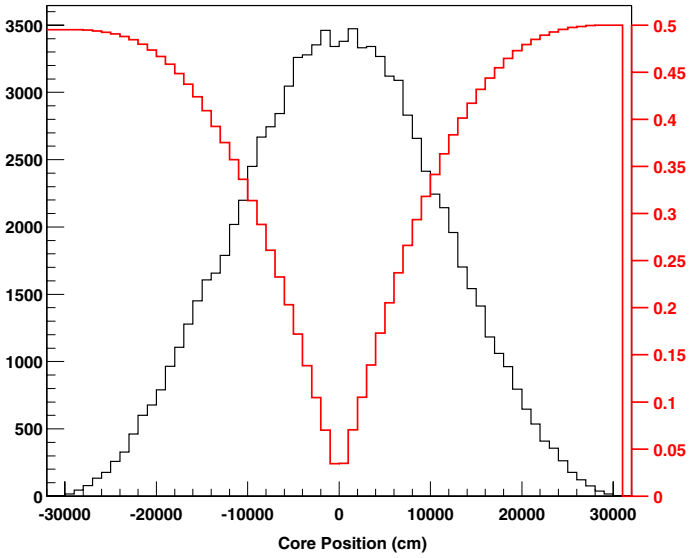

Fig. 1 (*continued*)

image of each shower was then sampled at the altitude of 600 g cm^{-2} and was recorded and prepared for the detector simulation. Only protons were generated, because we wish to focus specifically on the method of the analysis. A study of the contribution of He nuclei is discussed in Section 10, where an estimation of the influence of the presence of Helium nuclei on the evaluation of the energy spectrum is also given. However, at energy around the knee and at high altitude, the behavior of the hadronic shower is roughly independent of the atomic number, as it was reported by the *Asγ* [26] collaboration.

The Monte Carlo data were processed through a code simulating the expected performance of a detector which meets the general requirements (high duty cycle, large effective area, high altitude, fast timing) and the geometrical setup (20,000 pixels covering about 70×70 m^2) discussed in Section 2. The background generated by each logical pixel has been simulated and added and a simple majority trigger was applied. The core of the shower was uniformly distributed on a sampling area of 600×600 m^2. In order to check that the chosen value for the sampling area is large enough and does not imply a bias for the following evaluations, a parallel computation was performed by using a sampling area of 1000×1000 m^2. A plot of core position of the triggered events is shown in Fig. 1, for different multiplicity ranges. The chosen value of the sampling area (600×600 m^2) is large enough if only multiplicity greater than 30 pixels are accepted. The effective area computed by means of this procedure takes into account all the same geometrical cuts that will be applied to the experimental data set.

6. Data analysis

6.1. The pixel-multiplicity distribution

The first step in this work is to evaluate whether the detector described in the previous sections is able to discriminate between different cosmic ray spectra. In Fig 2 the differential pixel-multiplicity distribution is shown for three different values of the spectral index: $\gamma = 2.76$, $\gamma = 2.80$ and $\gamma = 2.84$. In order to reduce the effect of fluctuation, it is possible to refer to the integral pixel-multiplicity distribution, as it is shown in Fig. 3, for three different values of the spectral index: $\gamma = 2.76$, $\gamma = 2.80$ and $\gamma = 2.84$. The distributions shown in Fig. 3 are clearly separable, and this is a first

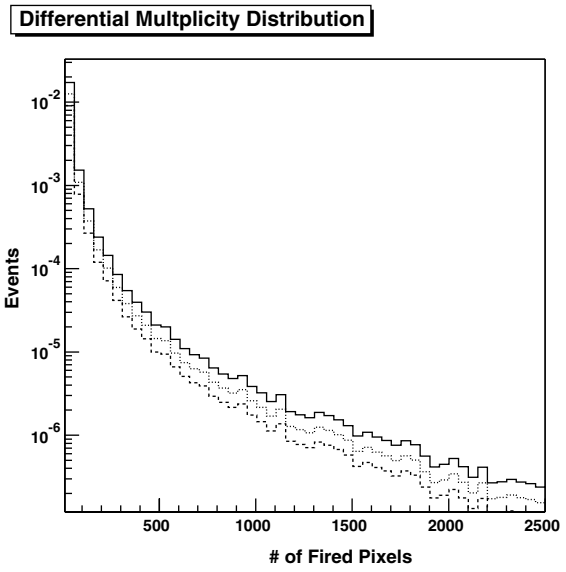

Fig. 2. The differential pixel-multiplicity distribution $M(P)$ is shown for three different values of the spectral index, $\gamma = 2.76$ (full line), $\gamma = 2.80$ (dotted line) and $\gamma = 2.84$ (dashed line).

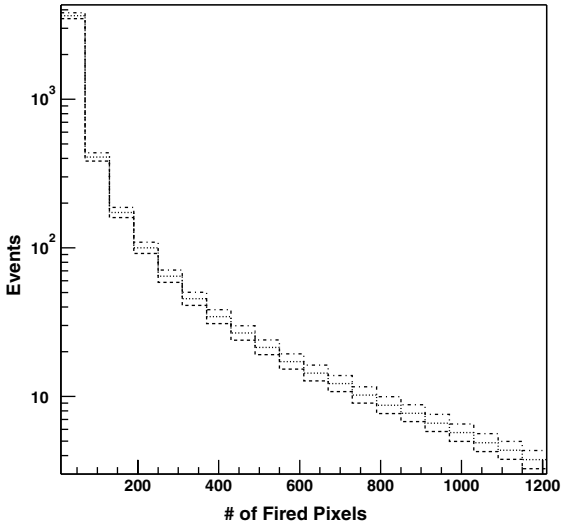

Fig. 3. The integral pixel-multiplicity distribution $M(> P)$ is shown for three different values of the spectral index, $\gamma = 2.76$ (dash-dotted line), $\gamma = 2.80$ (dotted line) and $\gamma = 2.84$ (dashed line). The three distributions are normalized to the value of the first bin of the differential distribution obtained assuming $\gamma = 2.76$.

indication of the possibility to use this observable to perform the required measurement. A more quantitative estimation can be obtained by looking at the ratio between different values of the spectral index: for pixel-multiplicity greater than 1000, the differences in flux corresponding to the two values $\gamma = 2.76$ and $\gamma = 2.84$ is larger than 20%, as is shown in Fig. 4. In order to evaluate how much the differences in flux corresponding to the two values $\gamma = 2.76$ and $\gamma = 2.84$ are affected by noisy channels in the detector, the same computation was repeated assuming two different values for the signal to noise ratio: the standard value of the order of 2% and an enlarged one of the order of 20%. Also in this second situation, with a very large value of the noise level, the two curves corresponding to $\gamma = 2.76$ and $\gamma = 2.84$ are yet different, as shown in Fig. 5.

6.2. The detector sensitivity

By using a standard value for primary cosmic ray flux given by [25]

$$\phi \simeq (1.1^{+0.08}_{-0.06}) \times 10^{-1} E^{-2.80\pm0.04}\ \mathrm{m^{-2}\,s^{-1}\,sr^{-1}\,TeV^{-1}}, \tag{6}$$

it is possible to estimate the data-taking period required to obtain a statistical error of the order of one tenth of the differences between the integral

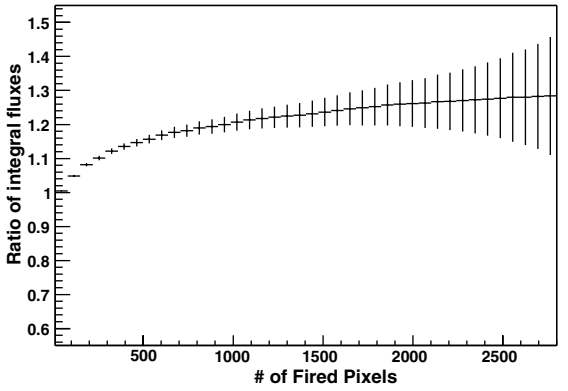

Fig. 4. The ratio of the integral pixel-multiplicity distributions for two different values of the spectral index, $\gamma = 2.76$ and $\gamma = 2.84$ is shown as a function of the pixel-multiplicity itself. The reported errors are statistical only and are obtained assuming a data taking period of about 3 h. The two integral distributions are previously normalized assuming equal to one the value of the content of the first bin.

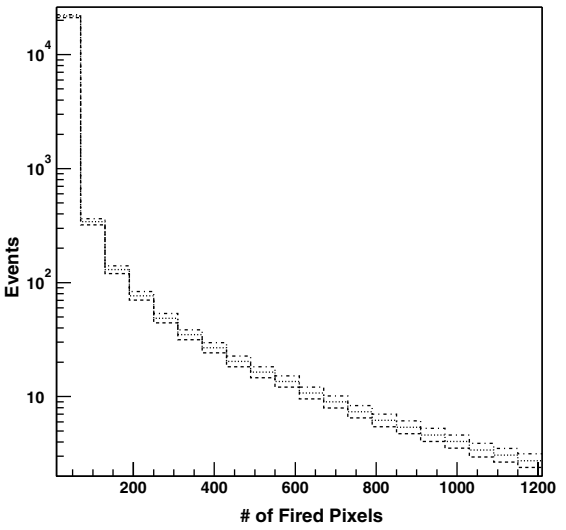

Fig. 5. The same quantity shown in Fig. 3 is now evaluated assuming a larger value of the noise level, as discussed in the text. See Fig. 3 for details.

fluxes corresponding at two values of the spectral index, $\gamma = 2.76$ and $\gamma = 2.84$. This ratio is a function of the pixel-multiplicity, and the value at 1000 pixels can be chosen as a reference point. We can estimate the time period required to obtain a number of data corresponding to the Monte Carlo sample that we simulated. The integrated effective area corresponding to a generation area A_g, i.e. the area over which the MC sample was thrown, can be obtained as follows integrating Eq. 3:

$$A_{\text{eff}}^{(g)}(E,K) = \frac{1}{\Omega}\int_{\Omega} A_{\text{eff}}(E,K,\Omega)\mathrm{d}\Omega = \frac{1}{\Omega}\int_{\Omega} A_g \cos(\theta)\mathrm{d}\Omega = \frac{\pi \sin^2\theta_\Omega}{\Omega} A_g, \quad (7)$$

where θ_Ω is the opening angle corresponding to a total solid angle Ω around the vertical. So that, if n_g is the number of events hitting the generation area A_g, the data taking period T can be evaluated by using Eq. 4 taking into account that $M(K)$ is by definition related to a unit data taking period, and integrating the primary cosmic ray spectrum above 1 TeV:

$$T = \frac{(\gamma-1)n_g}{\pi \sin^2(\theta_\Omega) A_g N_0}. \quad (8)$$

Considering events with 1000 fired-pixels, the total number n_g of events can simply be estimated requiring that the statistical error on the number of detected events with 1000 fired-pixels is one fifth of the difference between the expected value for a spectral index $\gamma = 2.80$ and $\gamma = 2.84$

$$\frac{\sqrt{N_{\gamma=2.80} + N_{\gamma=2.84}}}{N_{\gamma=2.80} - N_{\gamma=2.84}} \sim \frac{1}{5}. \quad (9)$$

Roughly speaking, this level of precision can be referred to as a "5σ signal", by using a notation very usual for gamma-ray detectors. By combining these two equations and taking into account the flux of Eq. 6 and the data in Table 1 and in Fig. 2, the required time can be computed:

$$T_{5\sigma} \sim 3 \text{ h}. \quad (10)$$

The Kolmogorov–Smirnov test was also applied to the integral distributions of Fig. 3. If the probability that the two pixel distributions obtained with $\gamma = 2.80$ and $\gamma = 2.84$ come from the same parent distribution is required to be lower than 1%, the data taking must be greater than about 20 min. A value of the same order was evaluated comparing the two curves obtained assuming $\gamma = 2.76$ and $\gamma = 2.80$. So that, a next generation of EAS detector will be able to discriminate between different values of the spectral index γ by collecting data for only few hours. This result could be scaled to the geometrical dimensions of other realistic detectors, taking into account that in a first approximation, the sensitivity scales with the root square of the geometrical area.

7. The computation of the effective area

Another data sample, different from the one used to obtain the distributions shown in Fig. 2, was generated in order to compute the effective area $A_{\text{eff}}(E,K)$, i.e. the yield of the event production. The value of the effective area $A_{\text{eff}}(E,K)$, as defined in Eq. (3), can be obtained by differentiating Eq. (4):

$$A_{\mathrm{eff}}(E,K) = \frac{1}{\Omega}\frac{1}{\phi(E)}\frac{\mathrm{d}M(K)}{\mathrm{d}E} = \frac{1}{\Omega}\frac{1}{\phi(E)}\cdot n(E,K), \tag{11}$$

where $n(E,K)$ is the number of event, induced by a primary particle of energy E and detected with a pixel-multiplicity K. This equation can be written one time referring to the events $n_{\mathrm{tr}}(E,K)$ triggered in the detector with a pixel-multiplicity K and another time by using the number n_{g} of the generated events onto the geometrical area A_{g}: in this second case, the effective area $A_{\mathrm{eff}}(E,K)$ can be analytically computed by integrating Eq. (3) and is given by Eq. (7). Dropping the unknown value of the flux between these two equations, the required result can be expressed in the form:

$$A_{\mathrm{eff}}(E,K) = \frac{\sin^2\theta_\Omega}{\Omega}\cdot\frac{n_{\mathrm{tr}}(E,K)}{n_{\mathrm{g}}(E)}\cdot A_{\mathrm{g}} = G(\Omega)\cdot\frac{n_{\mathrm{tr}}(E,K)}{n_{\mathrm{g}}(E)}\cdot A_{\mathrm{g}} \tag{12}$$

$$G(\Omega) = \frac{\sin^2\theta_\Omega}{2(1-\cos\theta_\Omega)},$$

where the opening angle θ_Ω and the generation area A_{g} are defined as in Eq. (7). This algorithms were implemented in a C++ specific code and the effective area $A_{\mathrm{eff}}(E,K)$ was computed.

8. The fit of the spectral index

The spectral index γ and the normalization factor N_0 can now be obtained from Eq. 3 by mean of a standard fitting procedure. We try to estimate the sensitivity of the pixel-like detector by computing back the spectral index γ and the N_0 for three different data samples, generated assuming three different values of the spectral index γ. All the three distributions are normalized at the same number of event at 1 TeV. The minimization package Minuit [27] was used for the computation. The results are shown in Table 2, for the three different values of the spectral index of the simulated distribution.

9. A Bayesian unfolding

9.1. Relevant relations

The problem of extracting the physics parameter of the cosmic ray flux from the experimental data is a classical unfolding problem [14] and can be dealt with by means of Bayesian techniques [15]. We briefly recall here the contents of the Bayes theorem in the case of n independent causes $C_i, i = 1,2,\ldots,n_C$ responsible for the effect $(E_{\mathrm{ff}})_j$, $j = 1,2,\ldots,n_E$. The Bayes theorem states that:

$$P(C_i) = \frac{\sum_{j=1}^{n_E} P(C_i|(E_{\mathrm{ff}})_j)\cdot P(E_{\mathrm{ff}})_j}{\sum_{l=1}^{n_C}\sum_{j=1}^{n_E} P((E_{\mathrm{ff}})_j|C_l)\cdot P(E_{\mathrm{ff}})_j},$$
$$P(C_i|(E_{\mathrm{ff}})_j) = \frac{P((E_{\mathrm{ff}})_j|C_i)\cdot P(C_i)}{\sum_{l=1}^{n_C} P((E_{\mathrm{ff}})_j|C_l)\cdot P(C_l)}, \tag{13}$$

where $P(C_i)$ and $P((E_{\mathrm{ff}})_j)$ are the probability distributions of the causes and of the effects, while $P(C_i|(E_{\mathrm{ff}})_j)$ and $P((E_{\mathrm{ff}})_j|C_i)$ are the conditioned probabilities. In the Bayesian unfolding schema, the $P((E_{\mathrm{ff}})_j|C_i)$ must be computed by a Monte Carlo program, the $P(E_{\mathrm{ff}})_j$ are the experimental data and the $P(C_i)$ are the quantity to be obtained. The problem is fully resolved by means of an iterative procedure: the $P(C_i|(E_{\mathrm{ff}})_j)$ are computed by the second of Eq. (13) by assuming a starting

Table 2
The table shows the *Monte Carlo* and *reconstructed* values for the normalization constant K and the spectral index γ, for three different values of the spectral index γ

γ_{MC}	K_{MC} ($\mathrm{m^{-2}\,sr^{-1}\,TeV^{\gamma-1}}$)	γ_{recons}	K_{recons} ($\mathrm{m^{-2}\,sr^{-1}\,TeV^{\gamma-1}}$)	χ^2/dof
2.76	1.548	2.763 ± 0.009	1.55 ± 0.09	52/48
2.80		2.814 ± 0.009	1.56 ± 0.07	57/48
2.84		2.852 ± 0.009	1.56 ± 0.11	59/48

The reconstructed values and the related errors are obtained by means of a standard fitting procedure: see text for details. The quoted errors arise from the number of Monte Carlo data used in the simulation.

value for the cause distribution $P(C_i)$, then from the first of Eq. (13) a new value for the $P(C_i)$ is computed. Then the $P(C_i|(E_{\rm ff})_j)$ can be evaluated again, and a new and more accurate value of the $P(C_i)$ is obtained. The iterative procedure ended when further variations on the value of the $P(C_i)$, computed by means of a chi-square procedure, are evaluated as negligible. The chi-square is computed using the values of the unfolded data corresponding to two contiguous steps of the unfolding procedure and the errors on the unfolded data sets. When this chi-square is lower than a cut chosen by the user the iterative procedure is stopped. During the iterative procedure, a smoothing algorithm can be applied to the unfolded data set that will be used in the following step of the iteration. This smoothing ensures a more fast convergence of the whole unfolding procedure and does not affect the final result because it is not applied at the last step of the iteration procedure. More details on the stability of this procedure and to the coupling with a smoothing procedure can be found in literature [15]. In our case, the "cause" is the incoming particle with an energy E, so that the cause distribution is given by the primary spectrum $P(E)$, while the "effect" is an event detected with pixel-multiplicity K, so that the "effect" distribution is linked to the $M(K)$ discussed above (see Eq. (4)). Finally, the conditioned probability $P((E_{\rm ff})_j|C_i)$ is linked, except for a normalization factor, to the effective area $A_{\rm eff}(E,K)$ discussed in the previous sections and computed by mean of the Monte Carlo code. The general formula must now be tuned in order to take into account that we are dealing with fluxes, and not just with numbers. The relevant quantities concerning the normalized distributions $P(E)$ and $P(K)$ can be coherently defined as follows:

$$P(E) = \frac{N(E)}{\int_{E_{\rm min}}^{E_{\rm max}} N(E'){\rm d}E'}, \tag{14}$$

$$P(K) = \frac{M(K)}{n_{\rm trigger}^{\rm tot}}, \tag{15}$$

where $N(E)$ is the differential energy distribution of the primary, $n_{\rm trigger}^{\rm tot}$ is the total number of particles detected in the apparatus and $M(K)$ is the number of events detected with pixel-multiplicity equal to K. Concerning the $P(K|E)$, computed by the Monte Carlo, their value can be expressed as a function of the effective area, considering Eq. (11):

$$P(K|E) = \frac{n_{\rm tr}(E,K)}{n_{\rm tr}(E)} = \frac{A_{\rm eff}(E,K)}{A_{\rm eff}(E)}, \tag{16}$$

where $A_{\rm eff}(E)$ is the effective area summed over all the multiplicity, i.e. the global yield for the detection of an event generated by a primary particle with energy E. Now it is possible to apply the Bayes theorem in order to obtain the required distribution $P(E)$, by means of a typical iterative procedure. Concerning the normalization, we can obtain the flux integrating over the effective area. If we use the generation area $A_{\rm g}$, the flux $\phi(E)$, according to Eq. (11) and referring to a data taking period T, is given by

$$\phi(E) = \frac{n_{\rm g}}{T}\frac{1}{\Omega}\frac{1}{G(\Omega)A_{\rm g}}\cdot P(E). \tag{17}$$

This equation can be used to compute the flux $\phi(E)$ by means of the Monte Carlo data, for which the generation area $A_{\rm g}$ is known. In the case of real data, it is better to express the flux $\phi(E)$ as a function of experimental observables and of quantities which parameterize the overall performance of the detector, like the effective areas $A_{\rm eff}(E,K)$ and $A_{\rm eff}^{\rm tot}(E)$, as follows:

$$\begin{aligned} A_{\rm eff}^{\rm tot}(E) &= \sum_K A_{\rm eff}(E,K) = G(\Omega)\frac{n_{\rm tr}^{\rm tot}}{n_{\rm g}}A_{\rm g}, \\ \phi(E) &= \frac{n_{\rm tr}^{\rm tot}}{T}\frac{1}{\Omega}\frac{1}{A_{\rm eff}^{\rm tot}}\cdot P(E). \end{aligned} \tag{18}$$

This equation completely solves the problems and makes it possible to obtain, from the experimental data, the energy of the primary particles, without any assumption, as required in the case of a standard fitting procedure, on the shape of the distribution itself.

9.2. The analysis method

The relevant equations, discussed in the previous section, were implemented in a Fortran code.

As known in literature [14,15], the Bayesian iterative procedure benefits from a smoothing procedure, that can be carried out by means of a simple linear function or by using a possible shape if the distribution to unfold is known. The authors stress that this procedure does not introduce a bias, but simply regularizes the convergences, because the last unfolding step is performed without fitting and it is determined by experimental data only. This combined procedure of fitting and iterative Bayesian unfolding is known as the most effective tool to obtain the parameter of a distribution of a known shape. This means that we can obtain the numerical behavior of the spectra of the incoming cosmic ray, without any assumption about its shape, or we can obtain the spectral index and the normalization if the power law is assumed. In order to test these assumptions in our specific problem, different chains of unfolding iterations were implemented, with or without the smoothing procedure.

9.3. The results on the unfolding of a power-law spectrum

The Bayesian unfolding was performed by using a power law or a flat distribution as the initial value of the $P(E)$. In both cases, a soft smoothing was applied to the nth value of the $P(E)$, during the recursive procedure, in order to ensure a stable convergence. The results are summarized in Tables 3–6. In all cases the results show the stability of the unfolding procedure, without any correlation with some variations in the iteration strategies. In order to investigate the dependency of the quality of the reconstructed parameters on the number of simulated data, the same procedure was repeated by using only a fraction of the simulated data. This investigation shows that the errors quoted in Tables 3–6 are of the same order of the errors resulting from unfolding half the simulated data set.

Table 3
The values obtained for K and the spectral index γ are shown

γ_{MC}	K_{MC} ($m^{-2}\,sr^{-1}\,TeV^{\gamma-1}$)	γ_{recons}	K_{recons} ($m^{-2}\,sr^{-1}\,TeV^{\gamma-1}$)
2.76	1.548	2.72 ± 0.02	1.51 ± 0.05
2.80		2.78 ± 0.03	1.53 ± 0.05
2.84		2.82 ± 0.03	1.53 ± 0.05

The data were obtained with the initial value of the $P(E)$ distributed as a power law and without smoothing.

Table 4
The values obtained for K and the spectral index γ are shown

γ_{MC}	K_{MC} ($m^{-2}\,sr^{-1}\,TeV^{\gamma-1}$)	γ_{recons}	K_{recons} ($m^{-2}\,sr^{-1}\,TeV^{\gamma-1}$)
2.76	1.548	2.74 ± 0.02	1.53 ± 0.05
2.80		2.78 ± 0.03	1.53 ± 0.05
2.84		2.82 ± 0.03	1.53 ± 0.05

The data were obtained with the initial value of the $P(E)$ distributed as a power law and applying a smoothing procedure.

Table 5
The values obtained for K and the spectral index γ are shown

γ_{MC}	K_{MC} ($m^{-2}\,sr^{-1}\,TeV^{\gamma-1}$)	γ_{recons}	K_{recons} ($m^{-2}\,sr^{-1}\,TeV^{\gamma-1}$)
2.76	1.548	2.73 ± 0.02	1.52 ± 0.05
2.80		2.78 ± 0.03	1.53 ± 0.05
2.84		2.82 ± 0.03	1.53 ± 0.05

The data were obtained with a flat starting value of the $P(E)$ distribution and without smoothing.

Table 6
The values obtained for K and the spectral index γ are shown

γ_{MC}	K_{MC} ($m^{-2}\,sr^{-1}\,TeV^{\gamma-1}$)	γ_{recons}	K_{recons} ($m^{-2}\,sr^{-1}\,TeV^{\gamma-1}$)
2.76	1.548	2.73 ± 0.02	1.53 ± 0.05
2.80		2.77 ± 0.02	1.52 ± 0.05
2.84		2.85 ± 0.03	1.56 ± 0.05

The data were obtained with a flat starting value of the $P(E)$ distribution and applying a smoothing procedure.

9.4. The unfolding of a Linsley-like energy spectrum

One of the major advantage of the use of the Bayes theorem is the possibility to unfold the primary particle spectrum without any assumption on the shape of the spectrum itself. In order to test the effectiveness of this feature, a Monte Carlo data sample was simulated according to a Linsley-like distribution given by

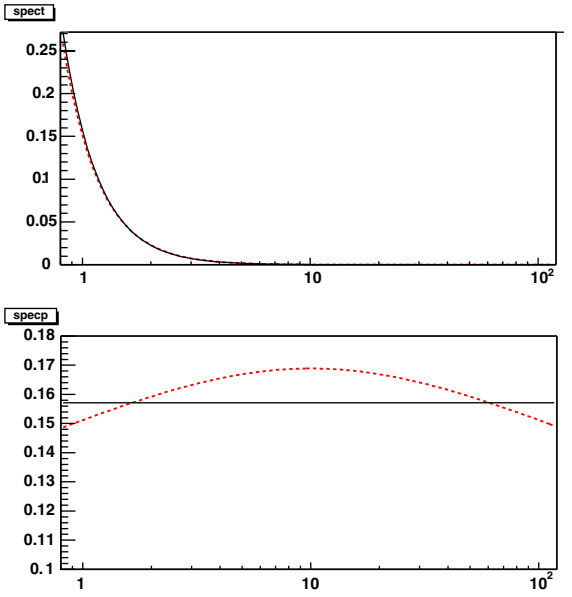

Fig. 6. The Linsley-like energy distribution described by Eq. 19 is shown (dashed line) and compared with a power law distribution (full line) with $\gamma = 2.8$ and with the same integrated flux between 1 and 100 TeV. In the plot on the lower panel the flux is multiplied by $E^{2.8}$ in order to enhance the differences between the two behaviors. The energies are expressed in TeV and the units of the flux are arbitrary.

$$\frac{\mathrm{d}N(E)}{\mathrm{d}E} = N_0 \frac{E^{-\gamma_1}}{\left(1+\frac{E}{E_0}\right)^{\gamma_2-\gamma_1}}. \tag{19}$$

The shape parameters in Eq. 19 was chosen as follows:

$$\gamma_1 = 2.6, \quad \gamma_2 = 3.0, \quad E_0 = 10 \text{ TeV}. \tag{20}$$

In Fig. 6 the Linsley-like spectrum described b y (19) is compared with a power law distribution with $\gamma = 2.8$ and with the same integrated flux between 1 and 100 TeV. The Bayesian unfolding procedure was applied to the Monte Carlo data produced by using this Linsley-like spectrum and the results are shown in Fig. 7. A fit of γ_1, γ_1 and E_0 was also performed and the results are summarized in Table 7.

10. The contribution of the helium nuclei

In the previous sections it was shown that a fitting and the Bayesian unfolding procedure can

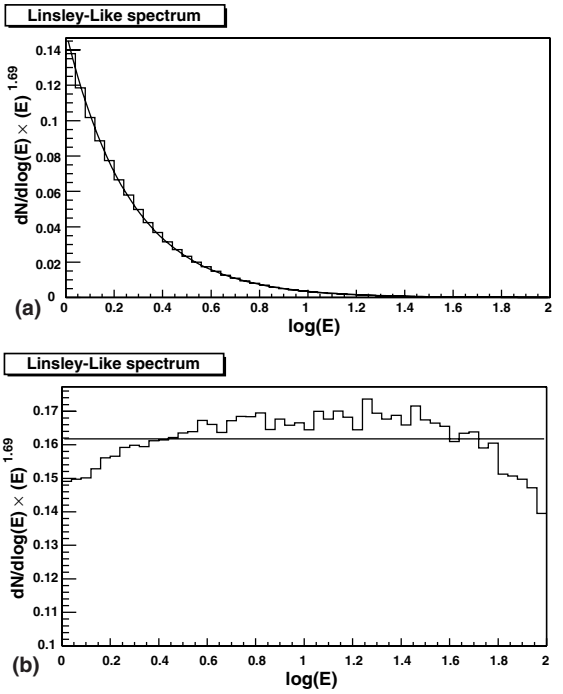

Fig. 7. The multiplicity distribution obtained by using the Linsley-like distribution is shown on upper panel. The energy distribution obtained by means of the Bayesian unfolding is shown on the plot on the lower panel: the superimposed horizontal line is the result of a fit with a power law and the flux is multiply by the value given by this power law, in order to enhance small differences in the energy behavior (cf. Fig. 6).

Table 7
The values of the parameters describing the Linsley-like distribution of Eq. 19 obtained by means of the Bayesan unfolding are shown

Type	γ_1	$\gamma_2 - \gamma_1$	E_0 (TeV)
MC	2.60	0.40	10
Reconstructed	2.62 ± 0.02	0.15 ± 0.05	11 ± 3

be successfully applied to the measurement of the spectral index of the primary particle, with a particular interest on the methodological aspects. The aim of this paragraph is now to evaluate the contribution of the helium nuclei to the shape of the pixel-multiplicity distribution and to the unfolding procedure itself. The following JACEE values [25] for the proton and helium fluxes were used:

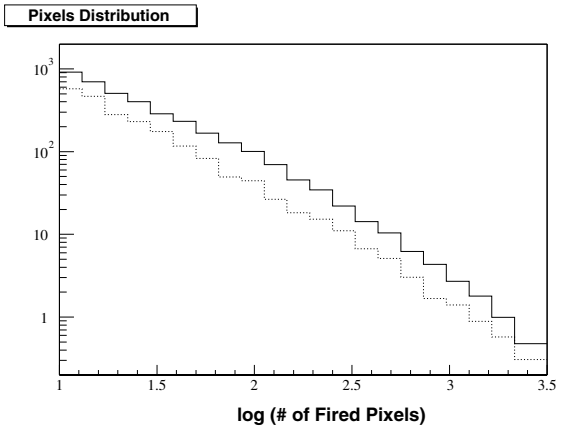

Fig. 8. The contribution of the proton (dashed line) and of the helium (dotted line) to the pixel-multiplicity distribution (full line) obtained assuming the primary particle fluxes described by Eq. (21).

$$\begin{aligned}\phi_{\mathrm{p}} &= (1.11^{+0.08}_{-0.06}) \\ &\quad \times 10^{-1} \cdot E^{-2.80\pm0.04}\ \mathrm{m^{-2}\,s^{-1}\,sr^{-1}\,TeV^{-1}}, \\ \phi_{\mathrm{He}} &= (7.86 \pm 0.24) \\ &\quad \times 10^{-3} \cdot E^{-2.68^{+0.04}_{-0.06}}\ \mathrm{m^{-2}\,s^{-1}\,sr^{-1}\,TeV^{-1}}.\end{aligned} \tag{21}$$

Fig. 8 shows the pixel-multiplicity distribution obtained assuming the fluxes described by Eq. 21. All the unfolding procedure was then repeated, computing again the effective area for the proton and the helium nuclei, and the two contributions from proton and for helium were separately unfolded by means of the Bayesian procedure. The unfolded primary spectra of proton and helium are shown in Fig. 9 and the reconstructed values of the parameters describing the two contributions are summarized in Table 8. The separately unfolding of the two contribution is relevant as a consistency check, but cannot be applied to real data, because in the pixel distribution the two contributions are

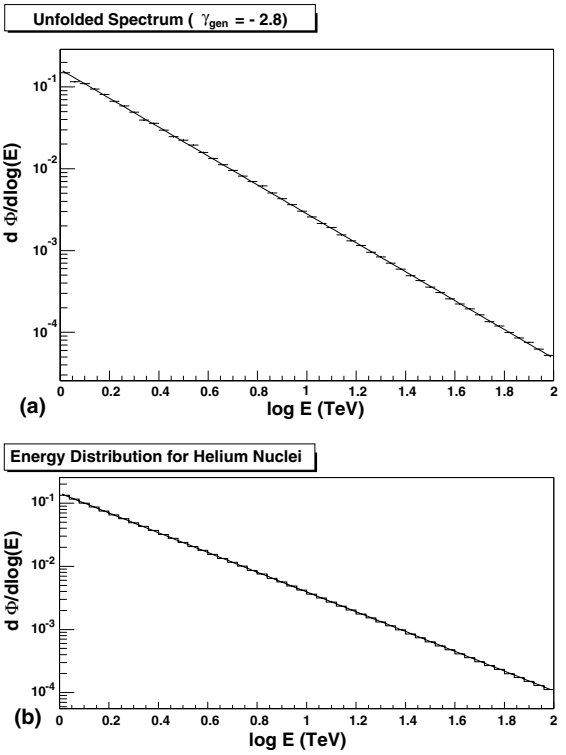

Fig. 9. The unfolded contribution of the proton (upper panel) and of the helium (lower panel) are shown.

mixed. The possibility to evaluate the light sector of the primary cosmic ray composition by means of the measurement of the pixel-multiplicity distribution is beyond the aim of this paper and it is at present under study. In the aim of the present work, the possibility to unfold the all nucleum spectrum was explored. The pixel-multiplicity distribution obtained by assuming the fluxes of Eq. 21 was then unfolded by means of the Bayesian procedure, using only the effective area computed for proton-induced shower, in the aim of the "superposition principle". The results of this

Table 8
The unfolded values for the helium and proton contribution to the primary particle energy distribution

Nucleus type	γ_{MC}	K_{MC} ($\mathrm{m^{-2}\,sr^{-1}\,TeV^{\gamma-1}}$)	γ_{recons}	K_{recons} ($\mathrm{m^{-2}\,sr^{-1}\,TeV^{\gamma-1}}$)
p	2.80	1.548	2.78 ± 0.03	1.53 ± 0.05
He	2.68	0.491	2.55 ± 0.04	0.45 ± 0.07

The two contributions are separately unfolded, as described in the text. The values for the proton component are the same quoted in Table 4.

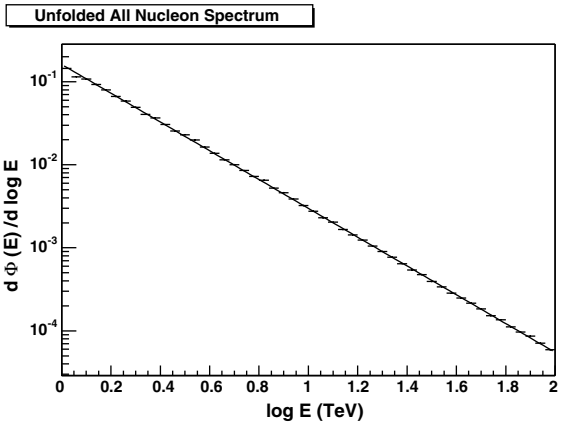

Fig. 10. The unfolded distribution of the all nucleon spectrum (full line) is compared with the distribution used to generate the events (histogram) and obtained by using the fluxes described by Eq. 21. The effective power law description of the original all nucleon distribution is also shown (dashed line).

Table 9
The unfolded values for the all nucleon spectrum obtained assuming the primary particle fluxes described by Eq. 21

$(\gamma_{MC})_{eff}$	$(K_{MC})_{eff}$ $(m^{-2}\,sr^{-1}\,TeV^{\gamma-1})$	γ_{recons}	K_{recons} $(m^{-2}\,sr^{-1}\,TeV^{\gamma-1})$
2.77	1.987	2.73 ± 0.03	1.94 ± 0.05

The reconstructed parameters are compared with an effective single power law description of the original all nucleon spectrum itself. See text for details on the unfolding procedure.

unfolding procedure are plotted in Fig. 10, which shows also the description of the all nucleon spectrum by means of a single power law. The numerical values of the unfolded parameters are shown in Table 9, compared with the effective slope γ_{eff} and the normalization $(N_0)_{eff}$ of the single power law description of the all nucleon spectrum obtained assuming the fluxes of Eq. 21. These results show that the unfolding procedure based on the Bayes's theorem can be successfully applied, in the energy range taken into account in this work, in order to determine the effective spectral index of the "all-nucleon spectrum".

11. Toward the knee of the energy spectrum

Besides the detailed study of the primary particle spectrum in the region of some TeV and more, the understanding of the feature of the "knee" region remains one of the most relevant topic in the field of the cosmic rays. The new generation of all sky detector, taken into account in this work, are planned in order to reach a very low threshold, because they are mainly devoted to the study of the gamma astronomy. For this reason it is difficult to evaluate in which way they should be used to reach the knee energy region, without knowing some details on the hardware setup. The study of this topic requires a devoted simulation, which will be the matter for a following paper.

12. Conclusion

This work aims to demonstrate the power of the next generation of gamma-ray detectors in order to clarify the puzzle of the behavior of primary cosmic rays around some TeV. The question is: why indirect data results seem different from direct measurement data? This work focuses on the problem and has suggested the way to solve it. A test was performed on a Monte Carlo data sample in order to quantify the requirements on the data sample used to perform the analysis and to evaluate the data taking period needed to obtain a significant measurement. A standard fit procedure and an iterative Bayesian unfolding procedure especially arranged to deal with this problem were used. The two procedures show comparable power to obtain the physics parameters and the Bayesian unfolding method has the additional peculiarity of not requiring any assumption about the shape of the primary spectrum. The helium contribution to the pixel-multiplicity distribution was also studied, and it was shown that the "all nucleon spectrum" can be successfully unfolded in the energy region of interest for our work. Further studies on the exploration of the knee region or on the possibility to extract informations from an huge amount of data, into which statistical uncertainties are negligible, are beyond the aim of the present work. The exploration of the "knee" region requires a devoted Monte Carlo simulation and a detailed study on the peculiarities of its sensitivity to high energy showers. On the other hand, the analysis of

a large data sample is possible only if the systematic uncertainties of the detector and of the simulation are known, because systematics dominate when statistical uncertainties are negligible.

Acknowledgements

We thank Giulio D'Agostini for having furnished us part of the Fortran code used in the Bayesian unfolding.

References

[1] M.S. Longair, in: High Energy Astrophysics, vol. 1, Cambridge University Press, Oxford, 1992.
[2] T. Gaisser, Cosmic Rays and Particle Physics, Cambridge University Press, Oxford, 1990.
[3] S. Petrera, Cosmic-ray spectrum and composition: ground observation, Il Nuovo Cimento 19C (1996) 737–754.
[4] S.J. Stochaj, Direct measurements of cosmic rays, in: Proceedings of 27th ICRC Invited, Rapporteur and Highlight Papers, 2001, pp. 136–146.
[5] C. Bacci et al., The use of RPC in the ARGO-YBJ project, Nucl. Phys. B (Proc. Suppl.) 78 (1999) 38–43.
[6] B.C. Shen et al., High energy gamma ray astroparticle physics with Milagro, Nucl. Phys. B (Proc. Suppl.) 71 (1999) 470–477.
[7] R.S. Miller, S. Westerhoff, Conceptual design of a next generation all-sky γ-ray telescope operating at TeV energies, Astropart. Phys. 11 (1999) 379–393.
[8] R.A. Ong, Very high-energy gamma-ray astronomy, Phys. Rep. 305 (1998) 93–202.
[9] C.M. Hoffman et al., Gamma-ray at high energies, Rev. Mod. Phys. 71 (1999) 897–936.
[10] S. Bussino, S.M. Mari, Gamma-hadron discrimination in extensive air shower using a neural network, Astropart. Phys. 15 (2001) 65–77.
[11] M. Abbrescia et al., ARGO proposal, 1996, unpublished.
[12] C. Bacci et al., Addendum to the Argo proposal, 1998, unpublished.
[13] M. Ambrosio et al., Vertical muon intensity measured with MACRO at the Gran Sasso laboratory, Phys. Rev. D 52 (1995) 3793–3802.
[14] V. Blobel, Unfolding Methods in High Energy Physics Experiments, CERN School of Computing Aiuguablava, Catalonia, Spain, 1984.
[15] G. D'Agostini, A multidimensional unfolding method based on Bayes' theorem, Nucl. Instrum. Methods A 362 (1995) 487–498.
[16] M. Ambrosio et al., Measurement of the residual energy of muons in the Gran Sasso underground laboratories, Astropart. Phys. 19 (2003) 313–328.
[17] G. Miller et al., Bayesian prior probability distribution for internal dosimetry, Radiat. Prot. Dosim. 94 (2001) 347–352.
[18] M. Roth, et al., Energy spectrum and elemental composition in the PeV region, in: Proceedings of 28th ICRC, vol. 1, 2003, pp. 139–142.
[19] D. Heck et al., Corsika: a Monte Carlo code to simulate extensive air showers, Forschungszentrum Karlsruhe Wissenhaltliche Berichte—FZKA 6019 (1998).
[20] W.R. Nelson et al., SLAC Rep. 265 (1985).
[21] A.F. Bielajew et al., SLAC Rep. 6499 (1994).
[22] N.N. Kalmykov, S.S. Ostapchenko, The nucleus–nucleus interaction, nuclear showers fragmentation, and fluctuations of extensive air, Phys. Atom. Nucl. 56 (1993) 346–353.
[23] H. Fesefeldt, Report PITHA (RWTH Aachen 1985), 85-02.
[24] J. Knapp et al., Comparison of hadronic interaction models used in air showers simulations and of their influence on shower development and observables, Forschungszentrum Karlsruhe Wissenhaltliche Berichte—FZKA 5828 (1996).
[25] K. Asakimori et al., Cosmic-ray proton and helium spectra: results from the JACEE experiment, Astrophys. J. 502 (1998) 278–283.
[26] M. Amenomori et al., The cosmic ray spectrum between $10^{14.5}$ e $10^{16.3}$ eV covering the "Knee" region, Astrophys. J. 461 (1996) 408–414.
[27] F. James, CERN Program Library Long Writeup, D506.