# Machine learning methods for helium flux analysis with DAMPE experiment

**M. Stolpovskiy,[a,*] D. Droz,[a] A. Ruina,[a] A. Tykhonov[a] and X. Wu[a] on behalf of the DAMPE[†] Collaboration**
(a complete list of authors can be found at the end of the proceedings)

[a]*University of Geneva,*
*Switzerland*

*E-mail:* mikhail.stolpovskiy@unige.ch

DAMPE is a space-borne experiment for the measurement of the cosmic-ray fluxes at energies up to around 100 TeV per nucleon. At energies above several tens of TeV, the electronics of DAMPE calorimeter would saturate, leaving certain bars with no energy recorded. It is also observed that at high energies the tracker and the scintillator detector that serve for the charge identification become heavily populated with back-splash tracks. Both effects interfere in precise measurements of the helium flux at highest energies. In the present contribution we discuss the application of machine learning techniques for the treatment of DAMPE data , to compensate the calorimeter energy lost by saturation and to identify helium events.

*37*th *International Cosmic Ray Conference (ICRC 2021)*
*July 12th – 23rd, 2021*
*Online – Berlin, Germany*

---

[*]Presenter

https://pos.sissa.it/

## 1. Introduction

Machine learning (ML) methods have proven to be very efficient in different scientific domains. In the specific case of cosmic ray (CR) physics, ML plays a key role for different applications as, for example, for event reconstruction in ground-based experiments [1], source recognition [2] etc. In the current work, we focus on the application of ML to the analysis of the helium energy spectrum for the DAMPE experiment, aiming to avoid the effects due to instrumental issues of our detector at high energies.

DAMPE is a space-borne particle detector installed on-board a satellite orbiting around the Earth in a Sun-synchronous way at 500 km of altitude since December 17th, 2015. The instrument is composed bt four major parts, here described in order of passage of CRs entering from the top of the satellite:

- PSD – double layer (for *x* and *y* measurements) Plastic Scintillator Detector composed by 82 scintillator bars $884 \times 28 \times 10$ mm$^3$ each. The main purpose of PSD is to measure the charge information of the incomming particles with Z from 1 to 26 [3].

- STK – STK - Silicon-Tungsten tracKer converter composed by 6 layers of strip tracking detectors providing an excellent track reconstruction and by 3 thin tungsten plates inserted in order to enhance the photon conversion in electron-positron pairs. STK provides an excellent track reconstruction and also enhances the gamma-ray conversion [4].

- BGO – BGO – a deep electromagnetic calorimeter ( 31.5 radiation lengths, 1.6 nuclear interaction lenghts). It is made of 14 layers of Bismuth-Germanium-Oxygen crystal bars, 22 bars ($25 \times 25 \times 600$ mm$^3$) in each layer [5, 6]. Each consecutive layer is oriented orthogonally to the previous one, allowing to measure the CR showers in three dimensions.

- NUD – NeUtron Detector to discriminate between nuclei and electron/photon events. [7].

The DAMPE Collaboration already published the results about the measurements of proton [8], helium [9] and electron/positron [10] energy spectra, and some preliminary results about gamma-ray observations were also presented [11]. These studies show interesting spectral features which help to push our understanding of the CR astrophysics and constrain the dark matter models [12–14].

## 2. BGO saturation

The DAMPE BGO calorimeter has broad and highly linear dynamic range going up to ~10 TeV of energy deposit per crystal. However, at even higher energy deposits the read-out electronics of BGO would saturate, therefore the bar will have a null recorded energy deposit. An example of event with two saturated bars is shown in figure 1. Such saturated bars are easily recognised as they are always next to the bars with energy deposit larger than 25GeV. However, we look for a method that allows to recover the information about the missing energy. A recent paper [16] describes an analytical method to do it. The method relies on analytical function dependent on the energy deposits in the bars adjacent to a saturated bars.
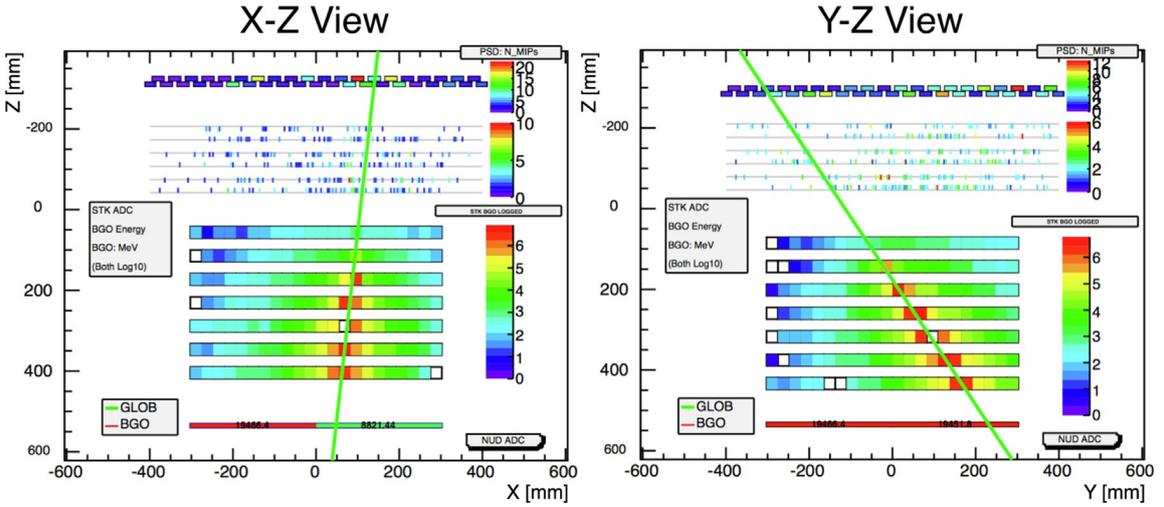
**Figure 1:** An illustration of a helium event with BGO readout saturation. The pre-correction total energy deposit is 49.4 TeV. The two empty BGO crystals on the shower axis are saturated, while the other empty crystals on the edge of shower are the ones without any deposited energy (or, the energy deposit is smaller than the noise threshold).

We developed an alternative approach using convolutional neural networks (CNN). This class of ML methods is widely used for image recognition, so it is very appropriate to be used for the BGO signals, combined in a single 14×22 matrix. As additional information we provide the neural network with the inclination values on X-Z and Y-Z projections (the coordinate system of DAMPE is illustrated on figure 1). The target of training is the energy missing due saturation. The training was performed on the Monte-Carlo sample of helium from 1 to 100 TeV.

The last layer of BGO is different from the upper layers: the optical filters on the ends of the BGO bars are adjusted to have more sensitivity at low energy deposited. This feature is important for the better electron-proton discrimination, which relies mainly on the distribution of signals in the last layer of BGO. However, this also means that the saturation happens at lower energies in the last layer. Figure 2 illustrates the fact that the missing energy for the last layer bars is substantially lower than that for the middle layers. This means also that the saturation in the last layer happens much more often, especially at lower energies. We split the problem of the saturation energy reconstruction in two. First, we focus on the reconstruction for the isolated saturated bars (i.e. not adjacent to any other saturated bar) in the BGO last layer. Second, we recover the adjacent saturated bars and the bars in the middle of BGO. The second step includes also the adjacent saturated bars where some of those bars are situated in the last layer. In the text below we will call the first class of the saturated bars as last layer (LL) saturation and the second class as the middle layer (MidL) saturation. Note that some events may have saturated bars of both classes. The frequencies of appearance of the saturated bars of these two classes are shown on figure 3.

We train two distinct CNN models with the same architecture as described above for reconstructing first the missing energy of the LL saturated bars and second the missing energy of the MidL saturated bars. Note that the paper [16] describes only the reconstruction of isolated saturated bars. The performance of the model is demonstrated in figure 4 for the events with a single isolated

3
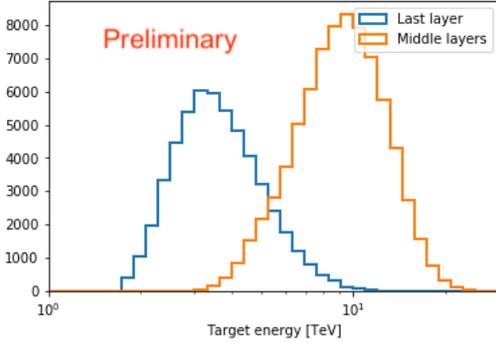
**Figure 2:** Missing energy due saturation in the last layer of BGO (blue histogram) and upper layers (orange histogram). Distributions build using helium MC.
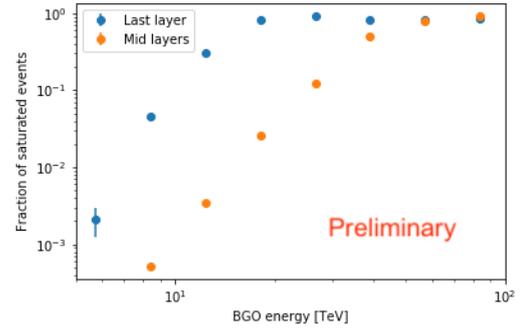


**Figure 3:** Frequency of the saturation event classes. Note that at highest energies the MidL start appearing more often than the LL saturated bars. This is not because the last layer bars are less often saturated at high energy, but because they start being adjacent to other saturated bars and so are classified as MidL ones.

saturated bar and 5 for all events.

Comparing with the performance reported in [16] paper we obtain approximately twice more precise reconstruction. Another advantage of the ML approach is that the CNN model does not have any problem of reconstructing the missing energy in events with multiple adjacent saturated bars. The analytical method described in [16] uses the signals from the bars adjacent to the saturated one, so it is not applicable for the cases of heavily saturated events. Such events, as suggested by the MC simulations and the observed flight data, are quite common at kinetic energy close to 100 TeV and above (that is close to the sensitivity limit of DAMPE), see figure 3. Thus the ML approach is preferential for the reconstruction of the saturated bars in the BGO calorimeter. It should be noted that the CNN model trained to reconstruct the missing energy for the helium incident CR, when being applied to different species adds some bias. Thus, for example, for the carbon flux analysis one needs to train a separate model. This is not specific for the ML model as the same is true for the analytical model too.

## 3. Helium charge recognition

The measurement of the helium energy spectrum measured by DAMPE [9] has been obtained by following the analysis steps here listed and applied both on on-orbit data and MC simulations of helium events performed with the GEANT4 toolkit [17]:

1. Pre-selection of well reconstructed events fully contained inside the whole DAMPE detector.

2. Charge selection of the events performed by using the charge measurements provided by the PSD and the first layer of the STK.

3. Template fitting of the PSD charge distributions with templates from the MC samples of proton and helium events, in order to estimate the proton pollution inside the selected helium candidate event sample.
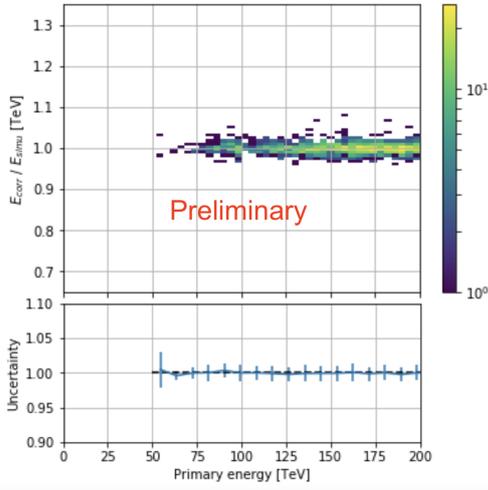
**Figure 4:** Top panel: Ratio between the reconstructed BGO energy ($E_{corr}$) and the true deposited energy ($E_{simu}$) as function of the kinetic energy of a particle for the events with a single isolated saturated bar. Bottom panel: uncertainty of the reconstruction.
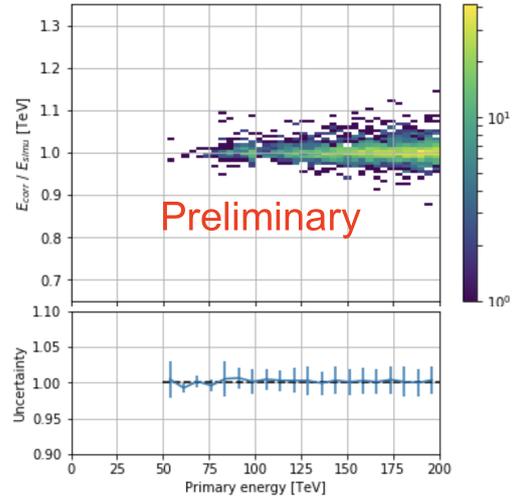
**Figure 5:** Top panel: Ratio between the reconstructed BGO energy ($E_{corr}$) and the true deposited energy ($E_{simu}$) as function of the kinetic energy of a particle for all events with saturated bars. Bottom panel: uncertainty of the reconstruction.

4.  Unfolding procedure applied in order to obtain the final helium energy spectrum.

The advantage of this approach is that the PSD charge distribution obtained after the second step shows very clean and well separated peaks of protons and helium nuclei. However, this comes on a price of relatively low effective acceptance. Many events with first inelastic interaction early in the particle path through DAMPE get rejected by the imposed selections.

The alternative ML approach relies on the classification model which is trained to separate protons and helium nuclei. In this case we apply very loose charge selections, which allows to have larger effective acceptance. This is crucially important for the highest energy bins where the statistics is very low. The template fit from the step three is then applied not to the PSD charge distributions but to the distributions of the ML model probability output (classification ML models have the possibility to estimate the chance that the input sample belongs to one class of events or another).

We use the following charge selections: we accept all events with energy in both PSD layers being below 10 MeV and events with maximal PSD signal over two layers above 10 MeV and average STK signal over the first layer below $1000./(S_{PSD} - 10.)^2 + 500$. ADC where $S_{PSD}$ is the average PSD signal over both layers. The cut line and the distribution of the events on the PSD-STK plane is shown in figure 6.

We used a boosted decision tree classifier with XGBoost open-source library [18]. We applied the following training hyperparameters: maximal tree depth 4, at least 100 entries in the leaves, number of estimators 850. An example of template fit of ML score for one of the energy bins is shown on figure 7. To select the helium candidates and estimate the background contamination

5
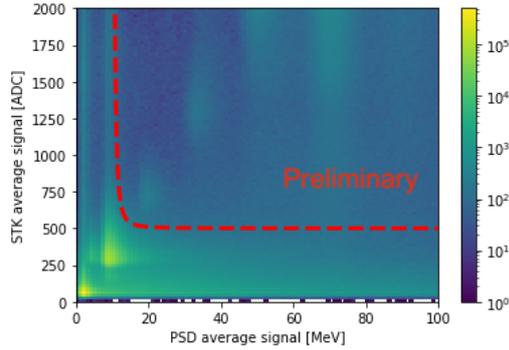
**Figure 6:** Two-dimensional distribution of STK mean signal over first layer versus PSD mean signal over two layers. The events below the red-dashed line are selected for the analysis. The peaks, visible on the diagonal are (starting from the bottom left corner): p, He, Li, Be etc.
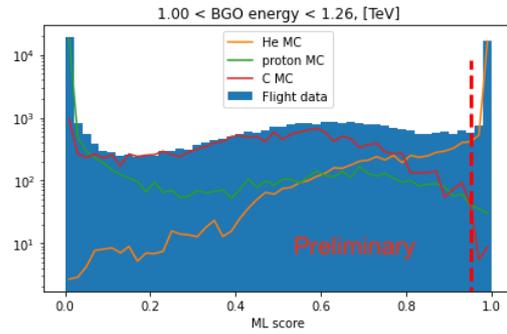


**Figure 7:** Example of the template fitting of flight data distribution of ML score (blue histogram) with a sum of helium, proton and carbon MC contributions for specific recorded energy bin (on the title of the figure). The red vertical dashed line shows the suggested limit of 0.95 above which the helium candidates are selected and the background contamination is estimated.

we cut at ML score 0.95. The helium signal peak is situated to the right from this limit. The background contamination is about 1% at incident energy around 1TeV.

The ML analysis is promising as it can help to extend the energy range of the reconstructed fluxes. For example, in the standard analysis [9] the number of selected helium candidate events in the deposited energy bin from 1 to 1.26 TeV is 13463, whereas the ML approach allows selecting 15924 helium candidates, or ~18% more. Such an increase can be indispensable for the highest energy bins. Currently, we are starting from the study of the helium flux with the ML method in order to apply the same procedure to the analysis of carbon energy spectrum.

## 4. Conclusions

Machine learning is a powerful tool that allows to recover the detector imperfections and analyse the data efficiently. In this contribution we presented the ML models for the reconstruction of the energy lost due saturation in the read out electronics of DAMPE calorimeter BGO, and for helium recognition. The model for BGO saturation correction is not only more precise than the concurrent analytical model, but it is also applicable for the most energetic events. The helium selection model is interesting due to its increased effective acceptance, which can potentially extend the flux measurement. The group is currently working on the tuning the models and extending their application to the heavier elements.

## References

[1] M. Erdmann, et al., "A deep learning-based reconstruction of cosmic ray-induced air showers." Astroparticle Physics 97 (2018): 46-53.

[2] O. Kalashev, et al., "Identifying nearby sources of ultra-high-energy cosmic rays with deep learning." Journal of Cosmology and Astroparticle Physics 2020.11 (2020): 005.

[3] Y. Yu, et al., The plastic scintillator detector for DAMPE, Astropart. Phys. 94 (2017) 1–10.

[4] P. Azzarello, et al., The DAMPE silicon–tungsten tracker, Nucl. Instrum. Methods A 831 (2016) 378–384.

[5] Y.-L. Zhang, et al., A high dynamic range readout unit for a calorimeter, Chin. Phys. C 36 (2012) 71–73.

[6] Z. Zhang, et al., The calibration and electron energy reconstruction of the BGO ECAL of the DAMPE detector, Nucl. Instrum. Methods A 836 (2016) 98–104.

[7] Y.-Y. Huang, et al., Calibration and performance of the neutron detector onboard of the DAMPE mission, Res. Astron. Astrophys. 20 (9) (2020) 153.

[8] Q. An, et al., (DAMPE Collaboration), Measurement of the cosmic-ray proton spectrum from 40 GeV to 100 TeV with the DAMPE satellite, Sci. Adv. 5 (9) (2019) eaax3793.

[9] F. Alemanno, et al. "Measurement of the cosmic ray helium energy spectrum from 70 GeV to 80 TeV with the DAMPE space mission." Physical Review Letters 126.20 (2021): 201102.

[10] G. Ambrosi, et al., (DAMPE Collaboration), Direct detection of a break in the teraelectronvolt cosmic-ray spectrum of electrons and positrons, Nature 552 (2017) 63–66.

[11] X. Li, et al., Recent gamma-ray results from DAMPE, in: 36th International Cosmic Ray Conference, ICRC2019, in: International Cosmic Ray Conference, vol. 36, 2019, p. 576.

[12] Q. Yuan, et al., Interpretations of the DAMPE electron data, 2017, preprint, arXiv:1711.10989.

[13] X. Pan, et al., Interpretation of the DAMPE 1.4 TeV peak according to the decaying dark matter model, Sci. China Phys. Mech. Astron. 61 (10) (2018) 101006.

[14] Q. Yuan, et al., Dark matter particle explorer observations of high-energy cosmic ray electrons plus positrons and their physical implications, Sci. China Phys. Mech. Astron. 61 (10) (2018) 101002.

[15] C. Yue, et al., Implications on the origin of cosmic rays in light of 10 TV spectral softenings, Front. Phys. (Beijing) 15 (2) (2020) 24601.

[16] C. Yue, et al. "Correction method for the readout saturation of the DAMPE calorimeter." Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 984 (2020): 164645.

[17] J. Allison, et al. "Recent developments in Geant4." Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 835 (2016): 186-225.

[18] T. Chen, et al., (2016). XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). New York, NY, USA: ACM.

## Full Authors List:   DAMPE[2] Collaboration

F. Alemanno[1,2], Q. An[3,4], P. Azzarello[5], F. C. T. Barbato[1,2], P. Bernardini[6,7], X. J. Bi[8,9], M. S. Cai[10,11], E. Casilli[6,7], E. Catanzani[12], J. Chang[10,11], D. Y. Chen[10], J. L. Chen[13], Z. F. Chen[10,11], M. Y. Cui[10], T. S. Cui[14], Y. X. Cui[10,11], H. T. Dai[3,4], A. De Benedittis[6,7], I. De Mitri[1,2], F. de Palma[6,7], M. Deliyergiyev[5], M. Di Santo[1,2], Q. Ding[10,11], T. K. Dong[10], Z. X. Dong[14], G. Donvito[15], D. Droz[5], J. L. Duan[13], K. K. Duan[10], D. D'Urso[12,1], R. R. Fan[8], Y. Z. Fan[10,11], K. Fang[8], F. Fang[13], C. Q. Feng[3,4], L. Feng[10], P. Fusco[15,16], M. Gao[8], F. Gargano[15], K. Gong[8], Y. Z. Gong[10], D. Y. Guo[8], J. H. Guo[10,11], S. X. Han[14], Y. M. Hu[10], G. S. Huang[3,4], X. Y. Huang[10,11], Y. Y. Huang[10], M. Ionica[12], W. Jiang[10], J. Kong[13], A. Kotenko[5], D. Kyratzis[1,2], S. J. Lei[10], W. H. Li[10,11], W. L. Li[14], X. Li[10], X. Q. Li[14], Y. M. Liang[14], C. M. Liu[3,4], H. Liu[10], J. Liu[13], S. B. Liu[3,4], Y. Liu[10], F. Loparco[15,16], C. N. Luo[10,11], M. Ma[14], P. X. Ma[10], T. Ma[10], X. Y. Ma[14], G. Marsella[6,7,2], M. N. Mazziotta[15], D. Mo[13], X. Y. Niu[13], X. Pan[10,11], A. Parenti[1,2], W. X. Peng[8], X. Y. Peng[10], C. Perrina[5,3], R. Qiao[8], J. N. Rao[14], A. Ruina[5], M. M. Salinas[5], G. Z. Shang[14], W. H. Shen[14], Z. Q. Shen[10], Z. T. Shen[3,4], L. Silveri[1,2], J. X. Song[14], M. Stolpovskiy[5], H. Su[13], M. Su[17], H. R. Sun[3,4], Z. Y. Sun[13], A. Surdo[7], X. J. Teng[14], A. Tykhonov[5], H. Wang[14], J. Z. Wang[8], L. G. Wang[14], S. Wang[10], S. X. Wang[10], X. L. Wang[3,4], Y. Wang[3,4], Y. F. Wang[3,4], Y. Z. Wang[10], D. M. Wei[10,11], J. J. Wei[10], Y. F. Wei[3,4], D. Wu[8], J. Wu[10,11], L. B. Wu[1,2], S. S. Wu[14], X. Wu[5], Z. Q. Xia[10], H. T. Xu[14], Z. H. Xu[10,11], Z. L. Xu[10], E. H. Xu[3,4], Z. Z. Xu[3,4], G. F. Xue[14], H. B. Yang[13], P. Yang[13], Y. Q. Yang[13], H. J. Yao[13], Y. H. Yu[13], G. W. Yuan[10,11], Q. Yuan[10,11], C. Yue[10], J. J. Zang[10,4] S. X. Zhang[13], W. Z. Zhang[14], Yan Zhang[10], Yi Zhang[10,11], Y. J. Zhang[13], Y. L. Zhang[3,4], Y. P. Zhang[13], Y. Q. Zhang[10], Z. Zhang[10], Z. Y. Zhang[3,4], C. Zhao[3,4], H. Y. Zhao[13], X. F. Zhao[14], C. Y. Zhou[14], and Y. Zhu[14]

[1]Gran Sasso Science Institute (GSSI), Via Iacobucci 2, I-67100 L'Aquila, Italy

[2]Istituto Nazionale di Fisica Nucleare (INFN) -Laboratori Nazionali del Gran Sasso, I-67100 Assergi, L'Aquila, Italy

[3]State Key Laboratory of Particle Detection and Electronics, University of Science and Technology of China, Hefei 230026, China

[4]Department of Modern Physics, University of Science and Technology of China, Hefei 230026, China

[5]Department of Nuclear and Particle Physics, University of Geneva, CH-1211, Switzerland

[6]Dipartimento di Matematica e Fisica E. De Giorgi, Universita' del Salento, I-73100, Lecce, Italy

[7]Istituto Nazionale di Fisica Nucleare (INFN) - Sezione di Lecce, I-73100, Lecce, Italy

[8]Institute of High Energy Physics, Chinese Academy of Sciences, Yuquan Road 19B, Beijing 100049, China

[9]University of Chinese Academy of Sciences, Yuquan Road 19A, Beijing 100049, China

[10]Key Laboratory of Dark Matter and Space Astronomy, Purple Mountain Observatory, Chinese Academy of Sciences, Nanjing 210023, China

[11]School of Astronomy and Space Science, University of Science and Technology of China, Hefei 230026, China

[12]Istituto Nazionale di Fisica Nucleare (INFN) - Sezione di Perugia, I-06123 Perugia, Italy

[13]Institute of Modern Physics, Chinese Academy of Sciences, Nanchang Road 509, Lanzhou 730000, China

[14]National Space Science Center, Chinese Academy of Sciences, Nanertiao 1, Zhongguancun, Haidian district, Beijing 100190, China

[15]Istituto Nazionale di Fisica Nucleare (INFN) - Sezione di Bari, I-70125, Bari, Italy

[16]Dipartimento di Fisica "M. Merlin" dell'Universita' e del Politecnico di Bari, I-70126, Bari, Italy

[17]Department of Physics and Laboratory for Space Research, the University of Hong Kong, Pok Fu Lam, Hong Kong SAR, China

[1]Now at Dipartimento di Chimica e Farmacia, Universita' di Sassari, I-07100, Sassari, Italy.

[2]Now at Dipartimento di Fisica e Chimica "E. Segre'", Universita' degli Studi di Palermo, via delle Scienze ed. 17, I-90128 Palermo, Italy.

[3]Also at Institute of Physics, Ecole Polytechnique Federale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland.

[4]Also at School of Physics and Electronic Engineering, Linyi University, Linyi 276000, China.