PROCEEDINGS OF SCIENCE

Probing hadronic cross sections in the TeV - PeV regime with DAMPE through machine learning techniques

Paul Coppin[a,*], **Andrii Tykhonov**[a], **Andrii Kotenko**[a], **Arshia Ruina**[a], **and Mikhail Stolpovskiy**[a] **for the DAMPE collaboration**

[a] *Département de physique nucléaire et corpusculaire, Université de Genève, École de physique, Quai Ernest-Ansermet 24, 1205 Genève*

E-mail: paulcppn@gmail.com

Thanks to its large calorimeter, the DArk Matter Particle Explorer (DAMPE) satellite experiment is ideally suited for the direct detection of cosmic rays (CRs) up to the knee. At these TeV to PeV energies, the main uncertainty on the CR flux measurements comes from the hadronic cross sections, which are largely experimentally unconstrained. We developed novel machine learning (ML) tools that are able to probe the depth at which CRs inelastically interact inside the DAMPE experiment. Applying these techniques to 7 years of DAMPE data, and comparing the results to predictions made by CR simulation frameworks such as Geant4 and FLUKA, we demonstrate how DAMPE data can be used to constrain the hadronic cross sections. Our results thus provide an important step towards reducing the uncertainties of CR flux measurements. Additionally, they form a pathfinder for similar studies with future experiments.

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

*Speaker

https://pos.sissa.it/

PoS(ICRC2023)142

1. Introduction

Motivation. Satellite-based direct detection experiments in the current era have measured the flux of cosmic-ray (CR) ions at sub-GeV energies up to several hundreds of TeV. Due to the rapidly falling spectrum, $dN/dE \propto E^{-2.7}$, flux measurements usually only become dominated by statistical uncertainties at the highest energies. Over the majority of the energy range, uncertainties on the measured flux are instead generally dominated by systematics. These systematic uncertainties originate, broadly speaking, from the complexities involved in simulating hadronic showers (see e.g. [1, 2]), along with the effect of poorly constrained nuisance parameters such as the inelastic hadronic cross section. For calorimetric direct detection experiments, such as CALET [3], NUCLEON [4], and DAMPE [5]; the uncertainty related to hadronic modelling forms the dominant systematic. In particular, cross sections play a vital role, as their uncertainty strongly affects quantities such as e.g. the detector trigger efficiency. These effects result in an uncertainty that can amount up to several tens of percent of the measured CR ion flux. Reducing systematics by constraining the ion-nucleon cross sections is therefore a vital requirement for obtaining more precise measurements. This will make it possible to better distinguish features in the CR flux, and allow for a more precise comparison between the results of different experiments.

Objective. In these proceedings, we present an analysis that uses CR ion events detected by the DArk Matter Particle Explorer (DAMPE) to perform a measurement of the ion-nucleon cross section. Given that proton and helium[1] are the two most abundant CR primaries below the knee, these are the two particle species considered in the current work. In particular, a measurement is performed of the inelastic hadronic cross section for proton and helium on $Bi_4Ge_3O_{12}$, which is the material that makes up the DAMPE calorimeter. It is worth noting that similar studies within DAMPE are currently also being performed for heavier ions, particularly carbon and oxygen [6].

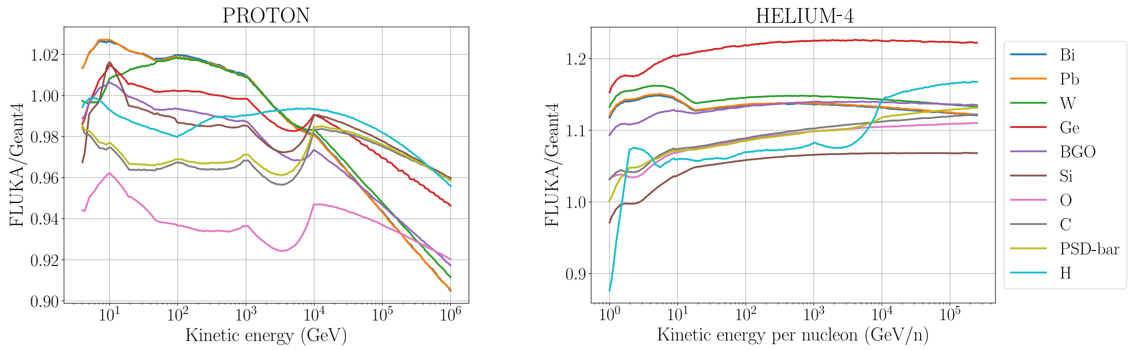
Content. Before going into the details of our analysis, we will first give a short discussion of the inelastic hadronic cross sections and particularly their differences in the FLUKA and Geant4 simulation frameworks (Section 2). This is followed by a description of the DAMPE experiment (Section 3), an overview of the analysis procedure used to measure the cross section (Section 4), and finally the results of our measurement followed by a short discussion (Section 5).

2. Inelastic hadronic cross sections in Geant4 and FLUKA

Ion-nucleon cross section measurements are generally not available at the high-energies reached by space-based CR direct detection experiments, except for a few scattering points at LHC energies. Phenomenological models are therefore used in simulation tools to predict the ion-nucleon cross section based on certain model parametrisations. A standard procedure is to use the Glauber [7, 8] or Gribov–Regge [9] framework, to convert the cross sections measured in proton-proton collisions to the inelastic hadronic cross section of heavy ions on arbitrary target materials.

Figure 1 shows the ratio between the cross section parametrisation in Geant4 [10–12] and FLUKA [13, 14] for p+A and ^{4}He+A interactions. Good agreement is observed for p+A reactions, with values for the cross section typically agreeing within a few percent. In contrast, a significant discrepancy is observed for ^{4}He+A , where the cross sections of FLUKA are ~10-15% higher than

[1]In this proceeding, helium specifically refers to the helium-4 isotope.



(a) **(b)**

Figure 1: Ratio of the inelastic hadronic cross section in FLUKA and Geant4. The left figure **(a)** shows that good agreement is observed for proton on a range of target materials (see colour legend). In contrast, the inelastic hadronic cross section of helium **(b)** is observed to be systematically higher in FLUKA than in Geant4. Target materials are indicated by their atomic symbol; PSD-bar and BGO correspond to EJ-200 plastic scintillator bars and $Bi_4Ge_3O_{12}$, respectively, both used in the DAMPE experiment.

in Geant4 on average. The close agreement for p+A is expected, given that the proton cross section has been measured for a range of target materials (see e.g. [15]). However, in the GeV to PeV energy range, little to no experimental constraints exists for helium-induced reactions. An exception is the helium-carbon cross section measurement by AMS [16]. This study found the inelastic hadronic cross section of helium-carbon to be ~10% higher than in Geant4 on average, and therefore to be more in line with FLUKA (see grey line in Fig. 1b). No similar studies have yet been performed for heavier targets. The aim our analysis is to test the model prediction of the inelastic hadronic cross section, by measuring this cross section of proton and helium on BGO.

3. The DAMPE Experiment

The DArk Matter Particle Explorer (DAMPE), also known as *Wu-Kong*, is a space-based experiment designed for the direct detection of CR ions, electrons/positrons, and photons [5]. DAMPE was launched into a Sun-synchronous orbit on December 2015 from the Juquan base in China. Since its launch, it has been stably taking data at an altitude of 500 km. On average, DAMPE detects $\sim 5 \cdot 10^6$ particles per day. DAMPE is effectively sensitive to CR electrons/positrons, and photons in the energy range from 5 GeV to 10 TeV. For these particles, it's effective acceptance corresponds to ~0.3 m^2 sr. For CR ions, DAMPE has an acceptance of ~0.1 m^2 sr, making it sensitive to particles in the energy range from about 10 GeV to 1 PeV.

DAMPE's distinguishing feature is that it has the largest calorimeter ($\sim$ 32 radiation lengths) of any current space-based CR experiment. This calorimeter, shown in Fig. 2, consists of 308 bars of $Bi_4Ge_3O_{12}$ (BGO) crystals, which are arranged in 14 layers, alternatingly oriented along the $x-$ and $y-$direction. Each BGO crystal is coupled to a PMT on either side, read-out at 3 dynode stages to allowing for a large dynamic gain.

Located above the calorimeter are a Plastic Scintillator strip Detector (PSD), and a Silicon-

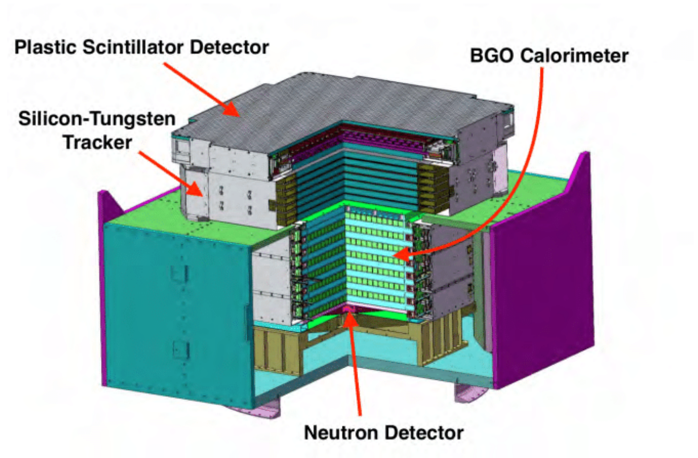


Figure 2: Layout of the DAMPE detector.

Tungsten Tracker (STK). The first subdetector[2] (PSD) serves to identify the electric charge of the CRs; while the second subdetector (STK) converts gamma-rays into electron-positron pairs, is used for tracking purposes (directional reconstruction), and can measure the electric charge of CRs. Finally, there is the neutron detector located below the calorimeter, which allows for additional[3] separation between electromagnetic and hadronic showers. For a detailed overview of the DAMPE experiment, we refer the reader to Ref. [5].

4. Analysis

Triggering and pre-selection. The first step in our analysis procedure is to construct a clean sample of proton or helium events. A series of pre-selections is first made, aimed at identifying clean down-going particles, before applying charge dependent cuts to select the desired ion species. Our preselection requires that events pass the MIP trigger[4], and that they have a reconstructed shower vector that is fiducially contained within the calorimeter. To ensure that the BGO shower vector can be reliably reconstructed, events are rejected if a single BGO layer contains more than 35% of the total deposited energy, or if in any of the first three layers, the BGO bar with the maximal energy is at the edge of the detector. Additionally, we require that events deposited at least 20 GeV in the calorimeter.

Following these pre-cuts, more advanced algorithms are applied to reconstruct the track of the primary particle. This track is generally hidden in a background of backscattered particles. A two-stage approach is taken as follows. First, the trajectory reconstruction is performed on the image of the BGO calorimeter. This BGO direction then acts as a seed for the more precise track reconstruction performed on STK. Both the BGO and STK track prediction are made by a convolutional neural network (CNN), as it has been shown that ML techniques strongly outperform

[2]In normal operation, DAMPE points towards the zenith (away from Earth). Due to Earth shielding, particles that traverse the detector vertically therefore always travel from top to bottom in Fig. 2.

[3]In addition to the separation power of the calorimeter based on the geometric shape of the shower.

[4]The MIP trigger requires that at least the signal of a minimum ionising charge 1 particle is observed in the top and bottom layers of the calorimeter. For more details on the trigger conditions, see [17, 18].

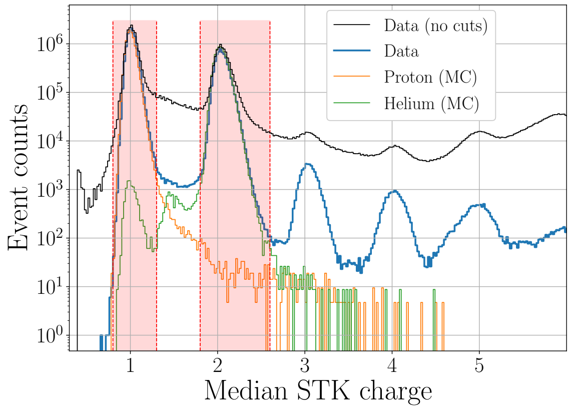


Figure 3: Median STK charge for data before (black) and after (blue) applying our analysis selection cuts. Red bands mark the charge window used to select proton and helium. Geant4 simulation for these two ions with the same cuts applied are shown in orange and green, demonstrating good data-MC agreement.

classical algorithms such as e.g. the Kalman-filter for particle tracking in DAMPE [19]. Given the final track parameters, the requirement is made that the track of the primary particle must be fiducially contained with the PSD, STK, and BGO.

Simulation. To evaluate the performance of the analysis, the passage of proton and helium events through the detector is simulated with both Geant4 (version 4.10.5) and FLUKA (version 2011.2X.7). In both frameworks, primary particles are sampled in a half-sphere around the detector[5] with an $dN/dE \propto E^{-1}$ spectrum. Events are then weighted to $\Phi \propto E^{-2.65}$ to resemble the CR proton and helium spectrum to first order. All simulated events are reconstructed and subjected to the same pre-cuts as data events.

Event selection. Proton and helium ions are selected based on the charge they deposit in the STK detector[6]. Deposited charges are corrected for the angle of incidence of the primary particle, and the position of impact on the silicon strips [22]. As STK consists of 12 consecutive planes, there are up to 12 charges measured per event. To combine these into a unified STK charge, we take the median of all non-zero charges[7]. Only events with a robust median are selected by requiring that at least 6 layers have a charge that deviates less than 0.3 of the median. The distribution of all data events that pass the MIP trigger, and those that satisfy our pre- and STK-charge selection are shown by the black and blue line in Fig.3, respectively. The contribution of proton and helium particles is clearly visible as the first and second peak. A charge window is then applied (red bands) to extract proton and helium ions. Simulations of Geant4 and FLUKA shows that the contamination of proton to helium is less than 0.2% (and vice versa); while keeping $\gtrsim$ 85% of signal events.

Vertex prediction. Given the selected sample of proton and helium events, we derive the ion-BGO cross section based on the fraction of ions that interact per layer of BGO material. For a

[5]As our analysis selects vertical events, Earth shielding implies that only down-going particles need to be simulated.

[6]Previous studies [20, 21] have shown that simulated proton and helium events generally overestimate the charge deposited in the PSD. This excess is observed for both Geant4 and FLUKA, and thought to be due to the mismodelling of backscattered particles. Due to this effect, we found that a selection based on the PSD signal causes the depth at which ions inelastically interact to be biased between data and MC. For this reason, PSD data is not used in our analysis.

[7]A zero charge is observed for layers in which the particles passes between the ladders with Si-strip detectors.

mono-energetic, collimated beam of particles, the survival fraction follows an exponential decay:

$$\frac{N(z)}{N(z=0)} = 1 - \exp(-\rho \cdot \sigma \cdot z), \tag{1}$$

where z is the distance covered by the particles, ρ the BGO density, and σ the inelastic hadronic cross section. Data observed with DAMPE, however, consists of a mixture of events at different incident angles, entry points, and energies. The underlying distribution of these variables is given by the source spectrum convoluted with the detector response (after applying our event selection). For this reason, event counts do not follow a simple exponential decay; and we instead use MC to evaluate how the survival fraction is affected by changes to the cross section.

The BGO calorimeter consists of 14 layers with 22 bars each. BGO data can thus be presented as a 22×14 image, in which each pixel represents the signal from a single BGO bar. We normalise these images such that the maximal pixel of the image corresponds to 1, and then decode the pixel values using 8 bit precision. A convolutional neural network is then applied to predict the layer in which the primary particle interacts inelastically. This CNN was trained on a mixture of Geant4 and FLUKA simulation. Aside from the output, the network is the same to that published in [19], and we hence refer readers to this publication for further details on the architecture and training of the CNN. We validated that our CNN can reproduce the true survival fraction with $\lesssim 1\%$ accuracy. Based on the CNN prediction, the fraction of events, α, that pass two BGO layers without interacting is calculated. Specifically, we consider the ratio between events that interact inelastically after the third and first BGO layer. This choice is motivated by three main considerations:

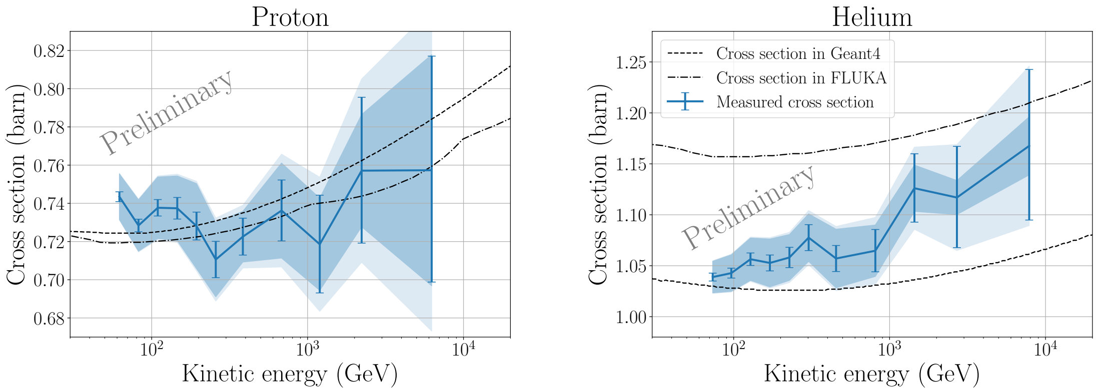
- After two layers, roughly half of the events have converted, giving us optimal statistics.
- The signal in the first layer can be used to ensure the particle did not convert before BGO.
- 11 layers of the calorimeter remain to accurately estimate the shower energy.

The ratio α is computed for data and then compared to that obtained for MC. If cross sections in MC are too low (high), it follows that α_{MC} will be higher (lower) than α_{data}.

Re-weighting MC. To parametrise the change of α_{MC} as a function of cross section, simulation sets are required in which the cross section is artificially increased or decreased. Simulating such events in large statistics, over an extended energy range, for different ion species, and for various cross sections would be very computationally expensive. We hence instead opt to re-weight our simulation. The main effect that comes from changing the cross section is that it induces a shift of the interacting vertex, z_i. Higher (lower) cross sections imply that more (less) particles interact early on. Events are weighted to reflect this change. Specifically, we bin events as a function of their angle of incidence, θ, and energy, E, such that Eq. (1) is applicable to each bin. We then note that under a change of the cross section, $\sigma' = (1+\epsilon)\cdot\sigma$, the cumulative distribution function of the interaction vertex, $F(z_i)$, changes in the following way:

$$F(z_i|\sigma') = 1 - [1 - F(z_i|\sigma)]^{1+\epsilon}\,. \tag{2}$$

Deriving $F(z_i)$ with respect to z_i to obtain probability density function, f, it then follows that the weighting factor corresponds to the ratio $w = f(z_i|\sigma')/f(z_i|\sigma)$. These weights are computed over a 3 dimensional grid $(\theta,\ E,\ z_i)$. A 3d linear interpolator is then applied for a more accurate and efficient evaluation, and to remove binning artefacts. It is worth noting that this weighting scheme is applied to different z-ranges of the detector individually. This allows modifying the cross



(a) **(b)**

Figure 4: Measurement of the inelastic hadronic cross section (XS) of proton and helium on $Bi_4Ge_3O_{12}$. Error bars show the statistical uncertainty on data. Dark and light colour bands show the systematic uncertainty of the analysis and the total (statistical + systematic) uncertainty, respectively.

section specifically in the calorimeter, without affecting the interaction probability in the previous sub-detectors. Additionally, we produced small statistics Geant4 simulation sets with increased and decreased cross sections to demonstrate that these reproduce the same shift in interaction vertex as our re-weighted samples, thus validating the procedure.

Cross section derivation. As cross sections vary smoothly as a function of energy, the same can be expected of the survival fraction $\alpha(E)$. To reduce the effect of fluctuations due to MC statistics on our results, we use a fourth order smoothing spline to obtain a continual parametrisation of α_{MC} as a function of energy. The weighting procedure described above is then used to determine the change in cross section required to make $\alpha_{\mathrm{MC}}(E)$ match the survival fraction in data.

5. Results

Figure 4a shows the inelastic hadronic cross section measured with our analysis for proton on $Bi_4Ge_3O_{12}$. Our result oscillates around the theoretical values (dashed lines), generally differing by only a few percent, and overlaps within the analysis uncertainty. Hence, we find that our measurement is in good agreement with the cross sections of Geant4 and FLUKA (dashed lines). As previously discussed, this close agreement is expected since cross sections for proton are well constrained [15, 23]. The proton cross section measurement thus serves to validate our method.

In the case of helium (Fig. 4b), measured cross sections agrees with the theoretical Geant4 cross section at low energies, after which a gradual increase is observed. Above 1 TeV, the measured cross section is still lower than that of FLUKA, but on average $\sim 10\%$ higher than that in Geant4.

6. Conclusion

Our results, along with those of Ref. [6], present the first measurement by DAMPE of inelastic hadronic cross sections. ML techniques were used, first to construct a pure sample of proton and helium events, and then to identify the interaction vertex of those events. By measuring the cross section of proton, which is well constrained, we demonstrated the validity of our method. In the case of helium, our analysis presents the first measurement of the BGO cross section at the considered energies. We found that while the measured cross section initially agrees with Geant4, an increase is observed of about $\sim 10\%$ above 1 TeV.

Acknowledgement The DAMPE mission was funded by the strategic priority science and technology projects in space science of Chinese Academy of Sciences (CAS). In China, the data analysis was supported by the Na- tional Key Research and Development Program of China (No. 2022YFF0503302) and the National Natural Sci- ence Foundation of China (Nos. 12220101003, 11921003, 11903084, 12003076 and 12022503), the CAS Project for Young Scientists in Basic Research (No. YSBR061), the Youth Innovation Promotion Association of CAS, the Young Elite Scientists Sponsorship Program by CAST (No. YESS20220197), and the Program for Innovative Talents and Entrepreneur in Jiangsu. In Europe, the activities and data analysis are supported by the Swiss National Science Foundation (SNSF), Switzerland, the National Institute for Nuclear Physics (INFN), Italy, and the European Research Council (ERC) under the Euro- pean Union's Horizon 2020 research and innovation pro- gramme (No. 851103).

References

[1] D. collaboration, *Chinese Phys. Lett.* **37** (2020) 119601.
[2] S. Banerjee and V. Ivanchenko, *EPJ Web Conf.* **251** (2021) 03010.
[3] S. Torii, *PoS* **ICRC2015** (2016) 581.
[4] E. Atkin et al., *EPJ Web Conf.* **105** (2015) 01002.
[5] DAMPE collaboration, *Astropart. Phys.* **95** (2017) 6–24.
[6] E. Xu, *PoS* **ICRC2023** (these proceedings) 161.
[7] R. J. Glauber, *Phys. Rev.* **100** (1955) 242–248.
[8] R. J. Glauber and G. Matthiae, *Nucl. Phys. B.* **21** (1970) 135–157.
[9] V. N. Gribov, *J. Exp. Theor. Phys.* **26** (1968) 414–423.
[10] S. Agostinelli et al., *Nucl. Instrum. Methods Phys. Res. A* **506** (2003) 250–303.
[11] J. Allison et al., *IEEE Trans. Nucl. Sci.* **53** (2006) 270–278.
[12] J. Allison et al., *Nucl. Instrum. Methods Phys. Res. A* **835** (2016) 186–225.
[13] T. Böhlen et al., *Nucl. Data Sheets* **120** (2014) 211–214.
[14] G. Battistoni et al., *Ann. Nucl. Energy* **82** (2015) 10–18.
[15] CMS Collaboration, *Phys. Lett. B* **759** (2016) 641–662.
[16] Q. Yan, V. Choutko, A. Oliva, and M. Paniccia, *Nucl. Phys. A* **996** (2020) 121712.
[17] Y. Zhang, Y. Liu, and J. Guo, *PoS* **ICRC2017** (2018) 232.
[18] Y.-Q. Zhang et al., *Res. Astron. Astrophys.* **19** (2019) 123.
[19] A. Tykhonov et al., *Astropart. Phys.* **146** (2023) 102795.
[20] DAMPE collaboration, *Sci. Adv.* **5** (2019) eaax3793.
[21] DAMPE collaboration, *Phys. Rev. Lett.* **126** (2021).
[22] A. Ruina et al., *PoS* **ICRC2021** (2022) 083.
[23] CMS collaboration, *J. High Energy Phys.* **2018** (July, 2018) 161.