Deep Learning for the HAWC Observatory

Ian James Watson[a,*] **for the HAWC collaboration**

[a]*University of Seoul,*
163 Seoulsiripdaero, Dongdaemun-gu, Seoul, South Korea

E-mail: ian.james.watson@cern.ch

We present the results of applying a transformer-based deep learning neural network to the data from the HAWC Gamma-Ray Observatory. HAWC observes the extensive air showers produced by very high energy gamma rays, and registers the Cherenkov radiation produced by the shower by photo-multiplier tubes (PMTs) instrumented in 300 large water Cherenkov detectors. The current HAWC method uses a staged parameterized fitting of the PMT information to find the shower center and incoming angular direction of the initiating gamma-ray, and produces variables which can be used for separating showers produced by gamma rays versus the overwhelming cosmic-ray background. The deep learning model, on the other hand, takes the charge and relative timing information of the PMTs as input and directly outputs an estimate of the incoming direction of the initiating gamma ray and a gamma-hadron discriminator. Both tasks are vital for source analysis. Better angular reconstruction allows for better source localization. Improved cosmic-ray rejection improves the signal-to-noise ratio. The deep learning network is found to perform better in simulation than the current methods at lower energies (around several hundred GeV gamma rays) where fewer PMTs are turned on by the shower, and therefore less information is available.
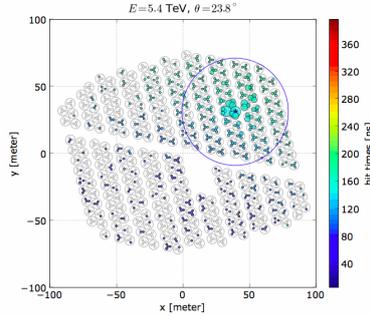
---

*Speaker

1. Introduction

Figure 1: A simulated HAWC detection of a gamma-ray shower is shown. The size of the circles corresponds to the effective charge measured by a PMT and the color indicates the relative time.

Deep Learning networks are showing performance on classification and regression tasks which outperform hand-built or classical machine learning techniques. Therefore, there is interest in applying these networks to the task of scientific data analyses. The present work introduces a Transformer-based network trained to analyse the data of the High Altitude Water Cherenkov (HAWC) gamma-ray observatory.

HAWC is located on the slope of the Sierra Negra volcano in the state of Puebla, Mexico and consists of 300 water cherenkov detectors, each one instrumented with four photomultiplier tubes (PMTs) which detect the extensive air showers produced by gamma-rays, with sensitivity to gamma-rays with energy from 300 GeV to 100 TeV. Fig. 1 shows the footprint of a gamma-ray shower on HAWC. HAWC has a trigger rate of 25 kHz and operates with a duty cycle greater than 95% and with a field-of-view of 2 sr [1]. The majority of the events are cosmic-ray showers so the gamma-hadron separation is of vital importance to study gamma-ray events. Additionally, in order to analyse gamma-ray sources, it is essential to reconstruct the direction and energy of the detected shower.

In this proceeding, we present deep learning networks which reconstruct the incoming direction and energy of the shower, and distinguish gamma ray showers from the overwhelming cosmic ray background. We then present the results on data, and show a method to finetune the network to overcome differences between the simulation and the real data from HAWC.

2. Deep Learning Network

Previously, work using machine learning in HAWC energy estimators [2] and gamma-hadron separation [3] were implemented using high-level input variables passed through a Multi-Layer Perceptron Neural Network (MLP). In this proceeding, we use Deep Learning techniques to apply a Transformer-based network direcly to the low-level PMT responses. Our network is based around the `Attention` module introduced by the *Attention is All You Need* paper [4], and is implemented in the PyTorch framework [5].

The data flow of the network is shown in Fig. 2. The input for each event is $\vec{x} \in \mathbb{R}^{1200 \times 2}$, where $x_i = (t_i, c_i)$ is the reconstructed time (relative to the trigger) and charge information in number of
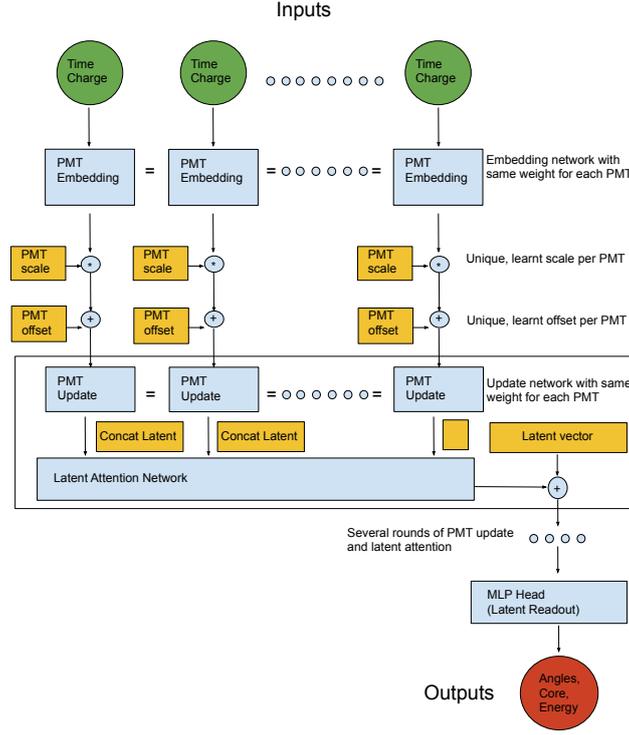
Figure 2: The data flow of the Deep Learning network used in this proceeding. Data flows from top to bottom.

photoelectrons (PEs), of the hit in the ith of 1200 PMTs (4 per tank in the 300 HAWC tanks). If there is no hit (a PMT response above a threshold equivalent to 1/4 PE) recorded in a PMT for the event, then $x_i = (0, 0)$. The PMT input array is passed through an MLP (with a single hidden layer) into an embedding space, $e_i = MLP_E(x_i) \in \mathbb{R}^C$, where C is the size of the embedding space, in our case $C = 16$. Then, a fixed vector s_i is multiplied and another vector o_i is added to the embedded vector, separately for each PMT, to produce the initial input which will be used by the `Attention` layers: $e_i^{(0)} = e_i \cdot s_i + o_i$. These vectors are optimized during the training and are the only place in the network where per PMT information is provided, so will encode information such as the relative PMT positions and the difference in PMT responses, in particular the central PMT and the 3 outer PMTs of HAWC are different sizes so have different responses to the same signal.

A latent vector $l^{(0)} \in \mathbb{R}^{256}$ is produced for each event (this initial value is optimized during training), which will be updated several times to become a shower representation by attending to the PMT embedding vectors as follows. The latent vector is used as the Query, and the PMTs are used as the Key and Value for the multi-head `Attention` and the output is passed through an MLP and added as a residual to the latent vector: $l^{(l+1)} = l^{(l)} + MLP_A(MHA(l^{(l)}, e^{(l)}, e^{(l)}))$, where MHA is the pytorch `MultiheadAttention` function. The PMT embedding is also updated residually by a concatenation the latent vector to each PMT vectorpassed through an MLP: $e^{(l+1)} = e^{(l)} + MLP_A(concat(l^{(l)}, e^{(l)}))$. This `Attention` update procedure is repeated in total six times.

The final latent vector is passed through a readout MLP with one hidden layer of size 2048 and the output $\hat{y}$ is returned: $\hat{y} = MLP_H(l^{(6)})$. The network is optimized so the output $\hat{y}$ matches

a target variable y. For the gamma-hadron classification network this is a single number per event representing how gamma-like the event is. The target is $y = 0$ for hadronic events, $y = 1$ for gamma-ray events. For the event reconstruction network, there are 5 targets $y = (\alpha, \zeta, x, y, E)$, where α is the azimuth and ζ the zenith of the initiating gamma-ray, x and y are the Cartesian coordinates of the core position of the shower on the HAWC detector array plane, and E is the predicted energy of the shower.

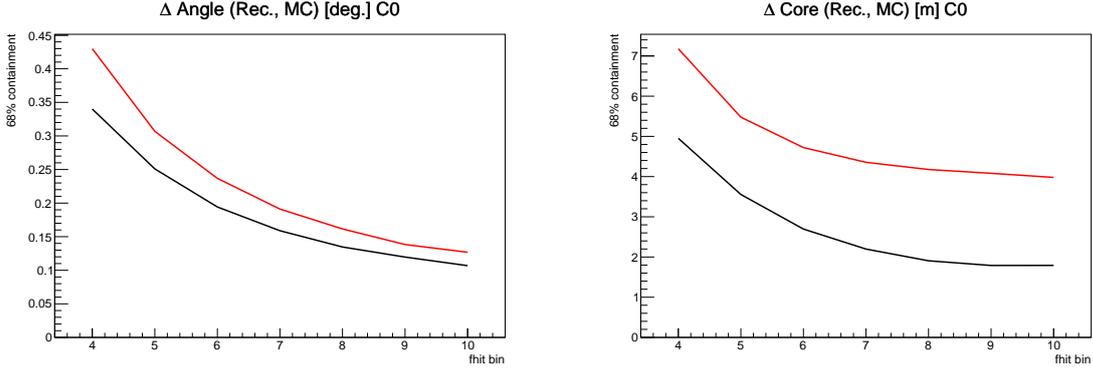## 3. Event Reconstruction Results in MC

Figure 3: The angular (left) and core (right) resolution in fHit bins 4-10. The points indicate the angular (distance) difference between the true and reconstructed angle (core position) which contains 68% of events as a function of fhit bin. The red line shows the current HAWC algorithms, the black line shows the deep learning network.

The simulation data is divided by using 80% of the total data for training the network and 20% of the data set aside to test the performance on an independent subset. The gamma-hadron network is trained using a binary cross-entropy loss, $L = -y \log \hat{y} - (1 - y) \log(1 - \hat{y})$, using Monte Carlo simulations (MC) of gamma ($y = 1$) and proton ($y = 0$) events. The event reconstruction network is trained using the L1 loss, $L = \sum_i |y_i - \hat{y}_i|$, using only gamma MC.

In the real data, the fraction of PMTs which are available for each event vary. For example, PMT hits are not used if they occur shortly after a high-charge hit due to afterpulsing contamination. Therefore, to adapt the network to these conditions, we applied a random erasing to the training data, in which we set a random selection of 0.25% of the PMTs to 0 charge and timing. A small Dropout is also applied as a regularization to the intermediate steps of the network. For both networks, we initially train with the AdamW optimizer [6] and then run an additional training with the SGD optimizer with a smaller learning rate, which was found to further improve the minimum loss. The trained networks are then applied to the testing subset to check the performance.

The angular resolution is a function of the number of PMTs with signal hits, more PMTs giving more data for the angular fit to use. This fractions of hits is also correlated with the gamma ray energy, with more PMT hits corresponding to events with higher energies. Therefore, the HAWC Collaboration typically divides data events into bins of fraction of available channels which have PMT hits, or fHit bins. Each bin is chosen so that the event rate is approximately half of the previous bin. The bins go from bin 0 where the fraction of PMTs with hits is 2.7% to 4.7%, bin 1 then goes

from 4.7% up to 6.8%, bin 2 up to 10.4%, 3 to 16.1%, 4 to 24.5%, 5 to 35.1%, 6 to 47.2%, 7 to 59.9%, 8 to 72.2%, 9 to 82.8% and 10 up to 100%.

The bins are also divided by core position. When the core of the shower is on the array, we call the event C0. When the core of the shower is between the edge of array and an additional 50% of the distance from the center to the edge of the array, we call the event C1. When the core of the shower is between an additional 50% to 100% of the distance from the center to the edge of the arrayon the array, we call the event C2. In standard HAWC analyses, only C0 and C1 events are used. Fig. 3 shows the angular and core reconstruction resolutions, compared to the offline reconstruction currently used by HAWC when the simulated core is on the array. The network acheives better performance over all fHit bins in the simulation.

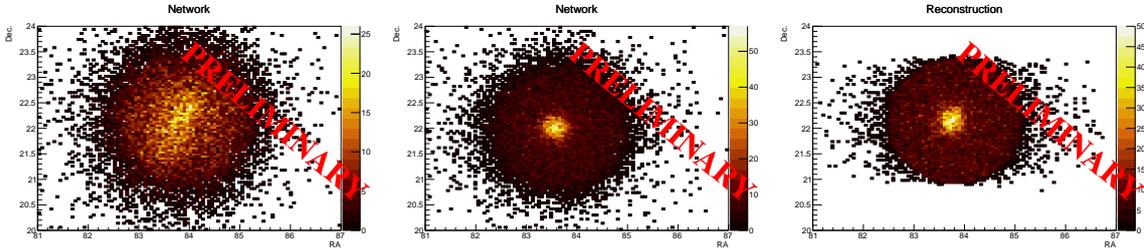## 4. Fine-tuning the Event Reconstruction Network for Use in Data

Figure 4: Data collected from nearby the Crab nebula. From left to right, the data is reconstructed by the simulation trained network, then the network fine-tuned to Crab data, and finally the standard HAWC reconstruction. In these plots, the data is from our testing sample, and were not used in finetuning the network.

The networks are applied to the real HAWC data after transforming the network for inference to be used by the ONNX Runtime [7]. For an initial study of the data performance, we used the standard HAWC reconstruction to analyse data from the Crab Nebula.

The Crab Nebula is the brightest gamma-ray object in the northern sky and is often used for calibration and performance studies by gamma-ray observatories. We therefore started performance studies by using a selection of HAWC data of events from fHit bins 7-10 which were reconstructed to be coming from near the Crab Nebula by the standard offline reconstruction algorithms. This data is passed through the gamma-hadron network and we select events with a high gamma-hadron score for further study. The selected events are passed through the event classifier network to reconstruct the core position, incoming angle and energy of the event.

The event reconstruction network as trained in Monte Carlo simulation was found to have worse angular reconstruction performance when applied to real data than in simulation. We hypothesized that this is due to the PMT response and so used this data to fine-tune the network by optimizing only the s_i and o_i (per PMT scale and offset) using the target $y = (\alpha, \zeta)$ of the known Crab position for each event (only the angles are used for this finetuning). Fig. 4 shows the number of events, binned right ascension and declination, for events from nearby the Crab nebula before and after finetuning, and the standard HAWC reconstruction for comparison. The event selection was done using an earlier reconstruction than the reconstruction we applied, therefore there are some events which are spread outside the circular region we accepted in right ascension (RA) and declination

(dec.). The Crab nebula in the pre fine-tuned network is spread across the selected region, while clearly visible as a large excess near the true position of the Crab nebula in events post fine-tuning.
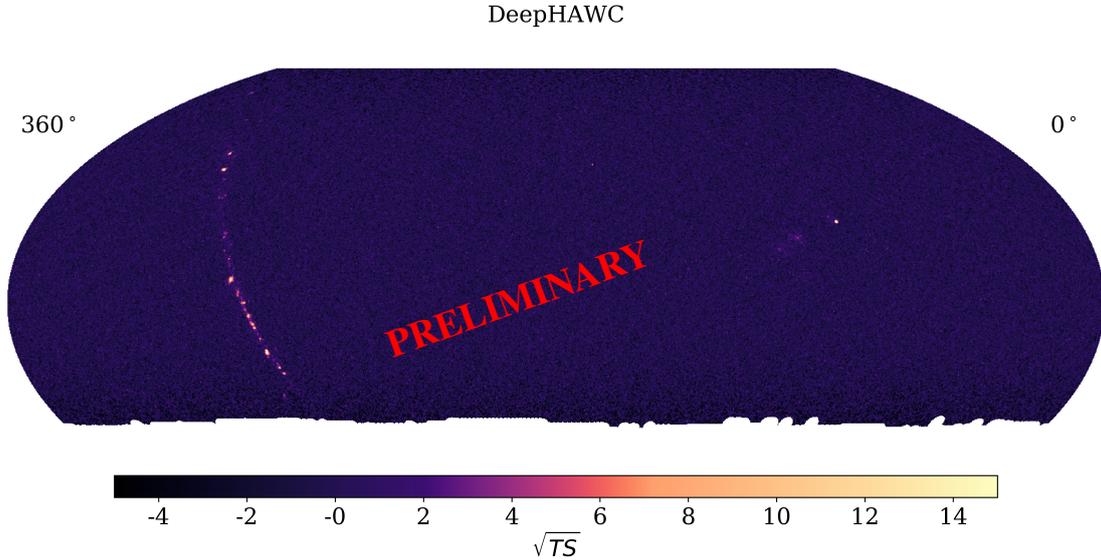
5. Results from HAWC Data

Figure 5: Preliminary skymap HAWC data from collected between June, 2015 and January, 2018, using the Deep Learning event reconstruction and gamma-hadron separation in fHit bins 7-10. The data and background estimation from the bins is combined into a significance assuming a point source with a power law spectrum of index 2.7.

After reconstructing the data, we divide the data into bins divided by fHit and core position as described above and apply a selection on the output of the gamma-hadron classification network and fill a Healpix-binned sky map with the reconstructed position. An additional small rotation was applied to account for a small bias on the Crab position compared to the known position. The direct integration method with region-of-interest masking is used to produce background estimations for each pixel on the map. The 12 fHit-core bin maps are then combined to make a significance map using the Li-Ma method [8]. For the combination of the bins, we assume a point-source response with power law spectrum for energy with an index of 2.7.

For checking if the performance translates to data across the whole sky, a fraction of high fHit bin data is used. The preliminary significance skymap of the HAWC data from June, 2015 to January 2018 using the Deep Learning event reconstruction and gamma-hadron separation in fHit bins 7-10 is shown in Fig. 5. These fHit bins account for less than 1% of the total collected data (higher fHit bins being rarer), and the HAWC observatory has been continuously running since June, 2015 until now, so this is a small fraction of the total data that could be analysed.

The significance map of the Crab Nebula using the HAWC data used in this study is shown in Fig. 6. Using the reconstruction and calibration in the previous section, the Crab is clearly seen at the correct position.
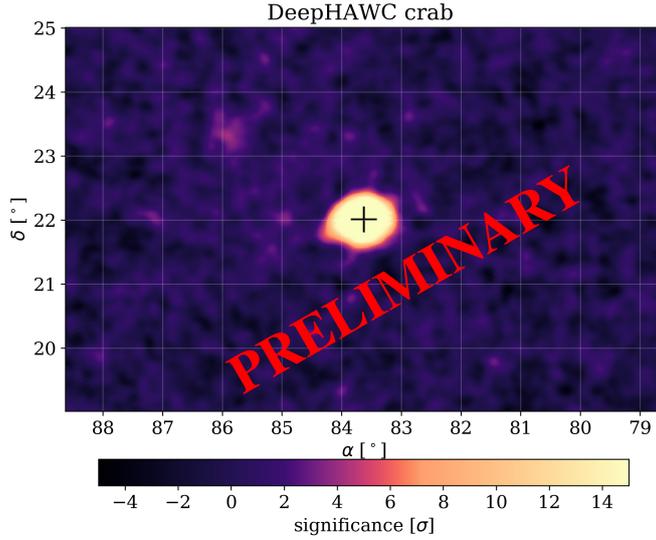
Figure 6: Signficance map at the Crab nebula using the Deep Learning reconstruction.
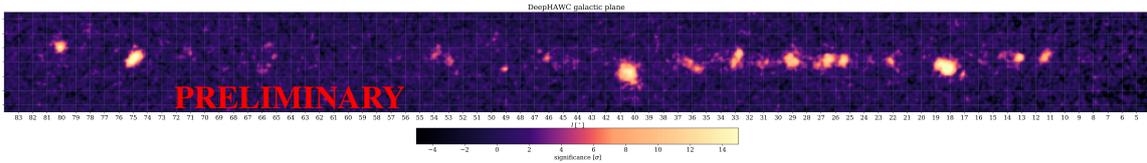

Figure 7: Significance map of the galactic plane using the Deep Learning reconstruction.

The galactic plane is rich in gamma-ray sources and the significance map for the Deep Learning networks is shown in Fig. 7. Many very high-energy galactic sources are seen in the map. This shows that the data-driven method works independent of the source declination across a wide declination range.

6. Conclusion

We have presented Deep Learning networks for HAWC event reconstruction and gamma-hadron separation. The event reconstruction MC-trained network was shown to perform well on the MC, but poorly in data, which we assumed was due to PMT differences in data and MC. We presented a method for data-driven finetuning of network parameters only related to PMT data-MC difference. After applying this technique to our network, we achieved good performance in the data subset used for this study. This model is now being applied to a larger subset of the data and in the future we plan to make a comparison with HAWC offline using the full dataset.

7. Acknowledgements

References

[1] A.U. Abeysekara et al., *Observation of the crab nebula with the hawc gamma-ray observatory*, *The Astrophysical Journal* **843** (2017) 39.

[2] A.U. Abeysekara et al., *Measurement of the Crab Nebula Spectrum Past 100 TeV with HAWC*, *The Astrophysical Journal* **881** (2019) 134 [1905.12518].

[3] R. Alfaro et al., *Gamma/hadron separation with the HAWC observatory*, *Nuclear Instruments and Methods in Physics Research A* **1039** (2022) 166984 [2205.12188].

[4] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez et al., *Attention is all you need*, *Advances in neural information processing systems* **30** (2017) .

[5] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan et al., *Pytorch: An imperative style, high-performance deep learning library*, in *Advances in Neural Information Processing Systems 32*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox and R. Garnett, eds., pp. 8024–8035, Curran Associates, Inc. (2019), http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

[6] I. Loshchilov and F. Hutter, *Decoupled weight decay regularization*, in *International Conference on Learning Representations*, 2019.

[7] ONNX Runtime developers, *ONNX Runtime*, Nov., 2018. https://github.com/microsoft/onnxruntime.

[8] T.-P. Li and Y.-Q. Ma, *Analysis methods for results in gamma-ray astronomy*, *Astrophysical Journal, Part 1 (ISSN 0004-637X), vol. 272, Sept. 1, 1983, p. 317-324.* **272** (1983) 317.

Full Authors List: HAWC Collaboration

A. Albert[1], R. Alfaro[2], C. Alvarez[3], A. Andrés[4], J.C. Arteaga-Velázquez[5], D. Avila Rojas[2], H.A. Ayala Solares[6], R. Babu[7], E. Belmont-Moreno[2], K.S. Caballero-Mora[3], T. Capistrán[4], S. Yun-Cárcamo[8], A. Carramiñana[9], F. Carreón[4], U. Cotti[5], J. Cotzomi[26], S. Coutiño de León[10], E. De la Fuente[11], D. Depaoli[12], C. de León[5], R. Diaz Hernandez[9], J.C. Díaz-Vélez[11], B.L. Dingus[1], M. Durocher[1], M.A. DuVernois[10], K. Engel[8], C. Espinoza[2], K.L. Fan[8], K. Fang[10], N.I. Fraija[4], J.A. García-González[13], F. Garfias[4], H. Goksu[12], M.M. González[4], J.A. Goodman[8], S. Groetsch[7], J.P. Harding[1], S. Hernandez[2], I. Herzog[14], J. Hinton[12], D. Huang[7], F. Hueyotl-Zahuantitla[3], P. Hüntemeyer[7], A. Iriarte[4], V. Joshi[28], S. Kaufmann[15], D. Kieda[16], A. Lara[17], J. Lee[18], W.H. Lee[4], H. León Vargas[2], J. Linnemann[14], A.L. Longinotti[4], G. Luis-Raya[15], K. Malone[19], J. Martínez-Castro[20], J.A.J. Matthews[21], P. Miranda-Romagnoli[22], J. Montes[4], J.A. Morales-Soto[5], M. Mostafá[6], L. Nellen[23], M.U. Nisa[14], R. Noriega-Papaqui[22], L. Olivera-Nieto[12], N. Omodei[24], Y. Pérez Araujo[4], E.G. Pérez-Pérez[15], A. Pratts[2], C.D. Rho[25], D. Rosa-Gonzalez[9], E. Ruiz-Velasco[12], H. Salazar[26], D. Salazar-Gallegos[14], A. Sandoval[2], M. Schneider[8], G. Schwefer[12], J. Serna-Franco[2], A.J. Smith[8], Y. Son[18], R.W. Springer[16], O. Tibolla[15], K. Tollefson[14], I. Torres[9], R. Torres-Escobedo[27], R. Turner[7], F. Ureña-Mena[9], E. Varela[26], L. Villaseñor[26], X. Wang[7], I.J. Watson[18], F. Werner[12], K. Whitaker[6], E. Willox[8], H. Wu[10], H. Zhou[27]

[1]Physics Division, Los Alamos National Laboratory, Los Alamos, NM, USA,

[2]Instituto de Física, Universidad Nacional Autónoma de México, Ciudad de México, México,

[3]Universidad Autónoma de Chiapas, Tuxtla Gutiérrez, Chiapas, México,

[4]Instituto de Astronomía, Universidad Nacional Autónoma de México, Ciudad de México, México,

[5]Instituto de Física y Matemáticas, Universidad Michoacana de San Nicolás de Hidalgo, Morelia, Michoacán, México,

[6]Department of Physics, Pennsylvania State University, University Park, PA, USA,

[7]Department of Physics, Michigan Technological University, Houghton, MI, USA,

[8]Department of Physics, University of Maryland, College Park, MD, USA,

[9]Instituto Nacional de Astrofísica, Óptica y Electrónica, Tonantzintla, Puebla, México,

[10]Department of Physics, University of Wisconsin-Madison, Madison, WI, USA,

[11]CUCEI, CUCEA, Universidad de Guadalajara, Guadalajara, Jalisco, México,

[12]Max-Planck Institute for Nuclear Physics, Heidelberg, Germany,

[13]Tecnologico de Monterrey, Escuela de Ingeniería y Ciencias, Ave. Eugenio Garza Sada 2501, Monterrey, N.L., 64849, México,

[14]Department of Physics and Astronomy, Michigan State University, East Lansing, MI, USA,

[15]Universidad Politécnica de Pachuca, Pachuca, Hgo, México,

[16]Department of Physics and Astronomy, University of Utah, Salt Lake City, UT, USA,

[17]Instituto de Geofísica, Universidad Nacional Autónoma de México, Ciudad de México, México,

[18]University of Seoul, Seoul, Rep. of Korea,

[19]Space Science and Applications Group, Los Alamos National Laboratory, Los Alamos, NM USA

[20]Centro de Investigación en Computación, Instituto Politécnico Nacional, Ciudad de México, México,

[21]Department of Physics and Astronomy, University of New Mexico, Albuquerque, NM, USA,

[22]Universidad Autónoma del Estado de Hidalgo, Pachuca, Hgo., México,

[23]Instituto de Ciencias Nucleares, Universidad Nacional Autónoma de México, Ciudad de México, México,

[24]Stanford University, Stanford, CA, USA,

[25]Department of Physics, Sungkyunkwan University, Suwon, South Korea,

[26]Facultad de Ciencias Físico Matemáticas, Benemérita Universidad Autónoma de Puebla, Puebla, México,

[27]Tsung-Dao Lee Institute and School of Physics and Astronomy, Shanghai Jiao Tong University, Shanghai, China,

[28]Erlangen Centre for Astroparticle Physics, Friedrich Alexander Universität, Erlangen, BY, Germany