PROCEEDINGS OF SCIENCE

Enhancing the Sensitivity of RNO-G Using a Machine-learning Based Trigger

Alan Coleman[a,*] **Christian Glaser**[a] **for the RNO-G Collaboration**
(a complete list of authors can be found at the end of the proceedings)

[a] *Dept. of Physics and Astronomy, Uppsala University, Box 516, S-75120 Uppsala, Sweden*
E-mail: alan.coleman@physics.uu.se

The Radio Neutrino Observatory in Greenland (RNO-G) is an array of radio detector stations which has been designed to study ultra-high energy ($E \gtrsim 10^{18}$ eV) neutrinos. The experiment, when completed, will have the best sensitivity in this energy range and will yield a major advancement in our understanding of the sources and propagation of the highest energy cosmic rays. While RNO-G will be sensitive to primarily $E \gtrsim 100$ PeV neutrinos, the optical-based detectors only have a large enough exposure to study up to ~ 1–10 PeV, leaving a gap in the energy range between the two detection methods. For RNO-G, the energy threshold is set by our ability to distinguish the Askaryan pulses, created from neutrino interactions, from the irreducible background of thermal noise. Using modern machine learning techniques, an online trigger can be implemented to identify small-amplitude pulses from in-ice cascades and thereby decrease the energy threshold of RNO-G. Such an advancement will increase the expected amount of observed neutrinos as well as close the gap between radio- and optical-based observatories. We present a convolutional neural network for classification of neutrino events that can be run as a second-stage trigger.

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

*Speaker

PoS(ICRC2023)1100

1. Introduction

With the discovery of astrophysical neutrinos (E$\gtrsim 10^{14}$ eV) by the IceCube Observatory [1, 2], the era of neutrino astronomy has begun. As a cosmic messenger, neutrinos have the benefits of being uncharged and can thus be used to identify their sources. Since they interact only via the weak force, individual particles can escape from relatively dense source environments, such as accretion disks. However, the small cross-section also requires that extremely large detectors need to be built in order to observe the small flux with a steep power-law spectrum ($\propto E^{-\gamma}$; $\gamma \simeq 2.5$). For the current generation of astrophysical neutrino detectors, comprised of optical sensors in liquid/frozen water, the observatories are $O(1\,\mathrm{km}^3)$ in size and observe several neutrinos per year with an energy of $O(1\,\mathrm{PeV})$.

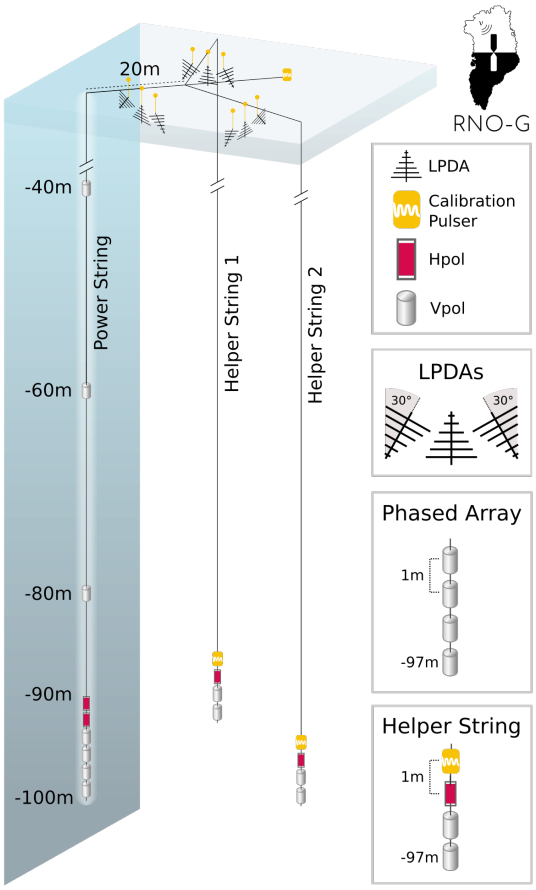

Figure 1: Schematic overview of an RNO-G station. The three cabled boreholes with antennas are shown, along with the depths and types of antennas that are placed along each string. The phased-array trigger is composed of the four Vpol antennas at the bottom-left.

The next generation of observatories, such as the Radio Neutrino Observatory in Greenland (RNO-G) [3], target an even higher-energy class of cosmic messengers, cosmogenic neutrinos. These neutrinos are expected to be produced by ultra-high energy (UHE) cosmic rays (E $\gtrsim 10^{18}$ eV) interacting with the cosmic microwave background [4, 5]. Cosmogenic neutrinos would be produced primarily with energies of 10^{17}–10^{19} eV, depending on the specifics of the sources and the distribution of energy/mass of the resultant UHE cosmic rays, see [6–8]. The observation of UHE neutrinos requires extremely large detectors, $O(100\,\mathrm{km}^3)$.

Such a large volume can be observed by placing radio antennas in glaciers, such as in Greenland [9]. The antennas observe the radio-frequency emission that is produced by the electromagnetic cascades caused by neutrinos interacting in the ice. RNO-G is based on this design and includes stations comprising 15 *deep* antennas at depths of 40–100 m below the surface and 9 *shallow* antennas a few meters below the surface, see Fig. 1. The radio technique becomes a viable method for detection starting at around $E \gtrsim 10^{17}$ eV where the radio emission is large enough, with respect to the radio-frequency noise, such that a station can observe in-ice showers up to a distance of ~ 1 km.

This leaves a notable gap in energy between where the current (and future) generation of optical detectors run out of statistics and where radio-based detectors begin to be sensitive. Instrumenting orders of magnitude more volume with optical sensors to reach 100 PeV is both difficult and expensive due to the nonlinear scaling of the flux and logistical considerations. Instead, the

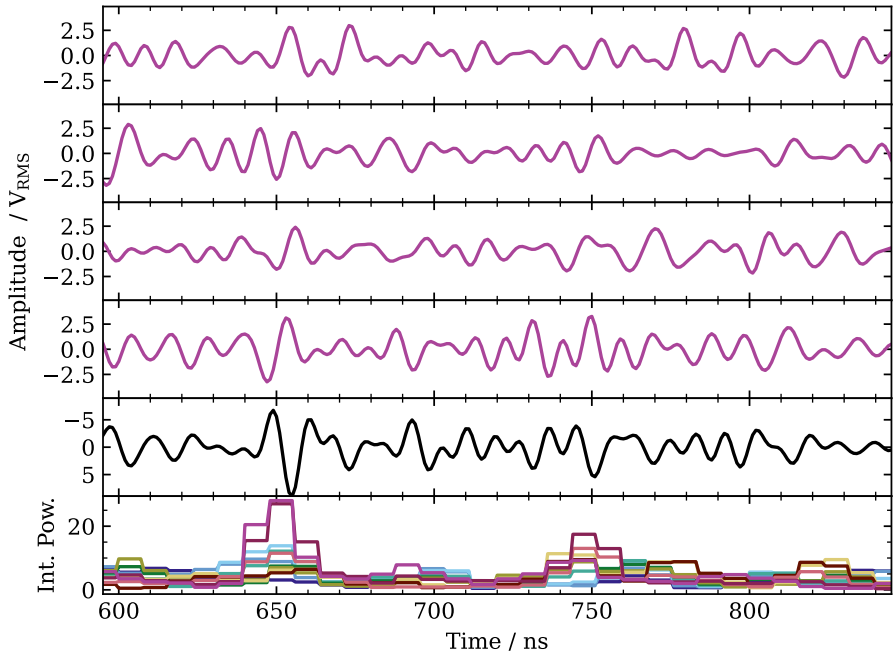

Figure 2: Example waveform that would pass a 10 kHz phased-array trigger, but not a 1 Hz trigger. The top four waveforms are the normalized voltages in the four vertically-polarized (Vpol) antennas. The fifth waveform is the coherent sum on the triggered beam. The bottom waveform shows the integrated-power amplitudes in the eleven beams. The color of the Vpol voltages match that of the triggering beam.

sensitivity of the radio technique can be improved by implementing a more intelligent trigger that is able to detect smaller amplitude waveforms within the noise. This will directly improve the expected signal rate at all energies and will have a particularly large influence for showers below $10^{17.5}$ eV, directly closing the gap with the optical array. In this proceeding, we discuss ongoing work to design a convolutional neural network (CNN) to be included in the trigger decision. This will directly increase the effective volume of the observatory at little-to-no additional cost.

2. Extending the current trigger design for RNO-G

The particle cascades, resulting from neutrino interactions in the ice, produce impulsive, coherent signals. The electric fields refract in the approximately-exponential refractive-index gradients of the Greenland ice [10, 11] and thus follow curved paths on the way to the antenna. To detect neutrino interactions in the ice, the received bipolar pulse must be distinguished from noise.

For in-ice antennas, the radio-frequency noise emitted by the galaxy is mostly shielded by the ice. Only emission originating from a few tens of degrees from the zenith is received by the dipoles. However, this is where the sensitivity of the antennas is the smallest. Additionally, the remote sites that are needed to instrument such large volumes of ice minimize the typical anthropogenic noise that might also cause impulsive signals, see [12, 13]. This leaves thermal noise as the dominant source of background when designing a trigger system.

The current RNO-G stations have two sets of triggers for the deep and shallow antennas. The shallow LPDA antennas are triggered using a Schottky diode and require two of the six antennas to pass a threshold to initiate data-readout. Neural network triggers have been studied in the context of

these types of antennas before [14]. This work focuses on the deep antennas for which the trigger is based on the four deepest antennas shown in Fig. 1. These antennas are known as the phased-array and trigger by coherently summing the four waveforms with respect to each other based on eleven different arrival directions (so-called *beams*), see Fig. 2. The power within overlapping windows is calculated, and a threshold is applied to this integral quantity, for more see [3, 15].

The current bandwidth for data-transfer from the RNO-G stations allows for a trigger rate of 1 Hz. In this work, we extend this single-stage trigger to a two-stage trigger. The general scheme is to decrease the power-integration threshold such that triggers are instead produced at a 10 kHz rate. Waveforms that pass this first stage are then given to a CNN, which performs filtering with a rejection rate of 10^4 such that the end bandwidth remains constant at 1 Hz.

3. Neutrino and electronics simulation

Simulations of the neutrino interactions, the propagation of the radio-frequency emission in the ice, and the station response were performed using NuRadioMC [16]. Neutrinos of all three flavors as well as both charged/neutral current interactions were generated and the radio-frequency emission was propagated to the antennas. For time and logistical reasons, only the four phased-array antennas were simulated. The gain pattern of the Vpol antennas was convolved with the received electric fields. The readout hardware was approximated by high- and low-pass Chebyshev filters with cutoff frequencies of 96 and 220 MHz, respectively.

The phased-array trigger is run on waveforms that are digitized at 500 MS/s and up-sampled to allow for 1 ns precision when forming the beams. For the simulations, the electric fields are first generated at 5 GS/s, folded with the antenna response and down-sampled to 500 MS/s. Since the current FPGA board already performs up-sampling for calculating the power integration and thus would already be in memory on the FPGA, we also use the up-sampled waveforms when calculating the neural network trigger.

Three simulation sets were made. The first is used for training and validation and included events that passed the 10 kHz trigger and *not* a 100 mHz trigger. This focuses the training set on the events that are the closest to the trigger threshold. Equal amounts of such triggered events were made for $\lg(E_{\text{show}}) = 0.25$ energy bins for each flavor/interaction combination, where E_{show} is the shower energy[1], for energies between 10^{16} and 10^{19} eV. For the testing data set, events were simulated at fixed *neutrino* energies between 10^{16} and 10^{19} eV and with $\lg(E_\nu) = 0.25$ spacing. A third data set was made of pure noise waveforms. These were produced by making random instances of 300 K thermal noise at 500 MS/s applying the phased-array algorithm. Waveforms were kept if they passed the 10 kHz trigger condition. In total, 711k neutrino simulations and 1M background events were made for training and validation and 780k neutrino simulations were made for testing.

4. Network design and training

When considering the architecture and size of the CNN, the most important consideration is that the network must run on the Cyclone® V 5CGXC7 field-programmable gate-array (FPGA) [17].

[1]This is defined to be the amount of energy deposited by the hadronic part of the shower. For charged current interactions from ν_e, the neutrino energy is equal to the shower energy.

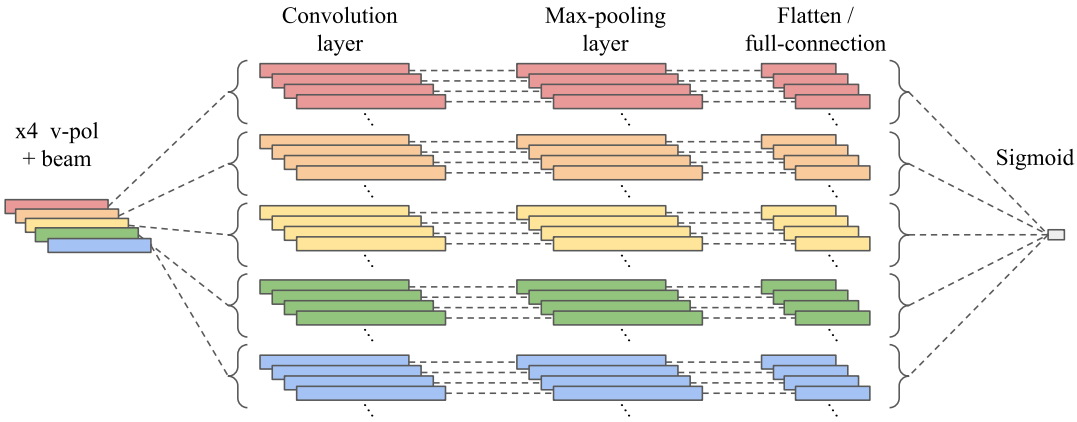

Figure 3: The schematic of the network architecture used in this study. The three parts of the network include a/some convolutional layer(s), max-pooling, and a fully connected layer. The 4+1 waveforms are indicated by color.

The space of meta-parameters for the network is thus limited by the constraints that the network must fit onto the FPGA and that the computation of the CNN must not introduce significant additional dead-time. In the proposed two-stage design, the phased-array will trigger at 10 kHz, making a trigger every 100 μs, on average. So the CNN that is included will need to be run after and thus should be executable on the FPGA in 1–10 μs. Given the 440 onboard multipliers and the 500 MHz clock, the network is limited to $O(10k-100k)$ floating-point operations (FLOPs), depending on how parallelizeable the chosen architecture is. This also limits the structure to be feed-forward and with few branch-points, as recursive elements and decision points significantly reduce parallelizability.

The design of the CNN design is shown in Fig. 3. The architecture includes three main components, a/some convolutional block(s) with a relu activation, max-pooling, and a fully connected layer with a sigmoid activation function. An important design consideration is that the 4+1 waveforms are essentially treated independently until the final full-connection step. Mixing the information across antennas firstly did not yield an improvement in the network response. Additionally, the separation of the information implicitly makes the network calculation more parallel.

Networks of various depths were tested to explore the meta-parameter space for architectures that require up to $O(10k)$ FLOPs. In this work, we highlight three networks that produced the best results for a given number of calculations. The designs of the three networks are given in Tab. 1. These networks increase by a factor of three in total number of computations per inference. However, it should be pointed out that the total amount of *time* to perform an inference may not scale linearly with the number of FLOPs due to bottlenecks in the computation of the algorithm, reducing the parallelizability. The subset of the waveform ("waveform length" in Tab. 1) given to the network was determined by the time at which the 10 kHz trigger was fulfilled. Each waveform was trimmed such that the phased-array trigger time is in the center of the waveform, mimicking what would be possible in the FPGA.

A binary-cross-entropy loss-function was used to train the models to classify the simulated neutrino events waveforms and noise. The metrics for training are shown for the medium network in Fig. 4. The left figure shows how well the data is separated based on the network output value. When implementing the trigger, the threshold applied to the network output value is the final

	CNN Size		
Network parameter	**Small**	**Medium**	**Large**
Waveform length (ns)	64	80	128
Convolution layers	1	1	4
Convolution filters	5	5	9
Convolution kernel	5	3	3
Convolution stride	3	1	1
Max-pool kernel	2	2	2
FLOPs	**5k**	**14k**	**47k**

Table 1: The parameters describing the architecture of three CNNs are given. Shown, schematically in Fig. 3, the three parts are a convolutional layer, a max-pooling layer, and a fully-connected layer. The total number of FLOPs needed to make an inference are also given.

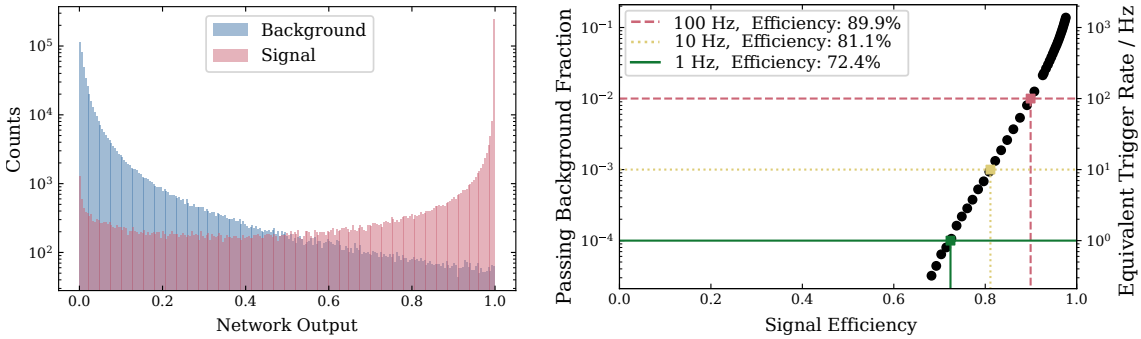

Figure 4: The training quality is shown for the medium-sized network (see Tab. 1). Left: distributions of the CNN network output for the 10 kHz pre-triggered signal and noise waveforms. Right: False-positive rate vs the true-positive for several threshold values on the network output (black circles). The efficiencies for several equivalent trigger rates are also shown.

tunable parameter. The right panel shows the results of a scan in the threshold parameter. The fraction of background waveforms that produce a network output above a given threshold is plotted against the fraction of the neutrino signals that are above the threshold. Several lines corresponding to second-stage trigger rates of 100 Hz, 10 Hz, and 1 Hz are indicated by the colored lines. For the medium-sized network, 72% of the signals are recovered. Similar values, 71% and 76%, are recovered for the small and large networks, respectively.

5. Enhanced effective volume using the neural network trigger

We applied the networks to the testing data sets as a way to determine the impact of the second-stage trigger. For each of the events, the waveform was trimmed, again with the time that the 10 kHz trigger occurred being half-way into the waveform. The CNN was applied to the 4+1 waveforms and the network output was checked against the threshold corresponding to a 1 Hz trigger level. Events that passed this selection were marked as detected.

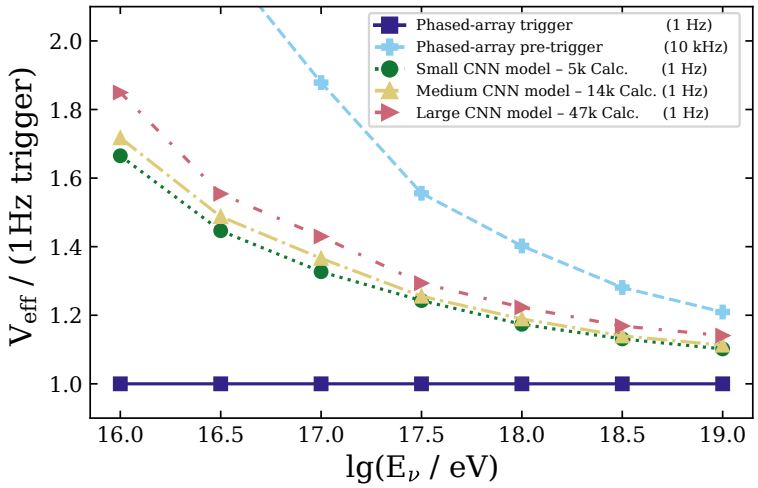

Figure 5: The increase in effective volume is shown with respect to the current 1 Hz phased-array trigger. The 10 kHz phased-array trigger is shown in dashed blue. The small, medium, and large CCN-based triggers, tuned to give a 10^4 background reduction, i.e. 1 Hz background rate, are also shown. The number of FLOPs per inference of the three networks are given in the legend.

The increase in effective volume is shown in Fig. 5. This plot shows the ratio of the effective volume for a given trigger type with respect to the standard 1 Hz phased-array trigger. The 10 kHz line is shown and indicates what would be possible for an ideal second-stage filter. The small, medium, and large CNN-based triggers fall in the middle. The increase in effective volume is $\simeq 45\%$ at 10^{17} eV and $\simeq 20\%$ at 10^{18} eV, and would directly translate to a proportional increase in the number of detected events.

That the increase of the sensitivity is not flat in energy is mostly related to the way that the effective volume scales with energy. At low energies, the viewable ice depends mostly on the distance of the emission to the antennas. Most events are very close to the trigger threshold, and a small increase in threshold results in a large increase in effective volume. At higher energies, the average signal-to-noise ratio is larger, and most events that pass a 10 kHz trigger also pass a 1 Hz trigger. The main limitation for detection is simply if the beamed emission is pointed towards the antenna. The increase in sensitivity equates to an increase in the viewing angle. However, since the emission falls sharply off the Cherenkov cone, the impact of using a CNN-based trigger is reduced.

6. Conclusion

One of the main science goals of RNO-G is the discovery of the UHE-neutrino flux. Even for the most optimistic models, the experiment may only detect $O(1)$ event per year. For such rare signals, the effective volume of the observatory is the most critical factor. We presented a method to directly enhance the sensitivity of the RNO-G stations by developing a CNN that can be run as a filter along with the current phased-array trigger.

We studied three light-weight networks that can be implemented on the current FPGA that is included in the RNO-G readout hardware. These networks range from 5k to 45k FLOPs to make an inference. Applying these networks to simulated events shows an increase in the effective volume by 80% to 15% for neutrinos with energies of 10^{16} eV to 10^{19} eV. The differences between the largest

and smallest networks is at the level of 5%–10%, indicating that in the conservative path, a very small dead-time of $\lesssim 1\,\mu$s ($\lesssim 1\%$) would need to be introduced for a much larger gain in sensitivity.

Further work will include lab tests of these types of networks, where neutrino pulses can be directly injected into the hardware using an arbitrary waveform generator. This will allow for an end-to-end test of the trigger to confirm its stability and effective background rejection. Other systematic effects, common to many simulation-based training-schemes, will also be studied including to determine where the simplifications in the simulation code may have an effect on the network response such as the quasi-exponential index of refraction [10, 11] and birefringent properties [18, 19] of the ice. Additionally, a trigger scheme where only a neural network is run in real-time could potentially boost the sensitivity even further, as it would not rely on an ad-hoc pre-trigger.

Acknowledgement:

We acknowledge support from the Swedish Research Council (VR2021-00158). The computations and data handling were enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC) at UPPMAX partially funded by the Swedish Research Council through grant agreement no. 2018-05973.

References

[1] **IceCube** Collaboration, M. G. Aartsen *et al. Science* **342** (2013) 1242856.
[2] **IceCube** Collaboration, M. G. Aartsen *et al. Phys. Rev. Lett.* **113** (2014) 101101.
[3] **RNO-G** Collaboration, J. A. Aguilar *et al. JINST* **16** no. 03, (2021) P03025. [Erratum: JINST 18, E03001 (2023)].
[4] K. Greisen *Phys. Rev. Lett.* **16** (1966) 748–750.
[5] G. T. Zatsepin and V. A. Kuzmin *JETP Lett.* **4** (1966) 78–80.
[6] A. Romero-Wolf and M. Ave *JCAP* **07** (2018) 025.
[7] K. Kotera, D. Allard, and A. V. Olinto *JCAP* **10** (2010) 013.
[8] K.-H. Kampert and M. Unger *Astropart. Phys.* **35** (2012) 660–678.
[9] S. Barwick and C. Glaser *to be published in Neutrino Physics and Astrophysics* . arXiv:2208.04971.
[10] **RNO-G** Collaboration, J. A. Aguilar *et al.* arXiv:2304.06181.
[11] **RNO-G** Collaboration, B. Oeyen *et al. PoS* **ICRC2023** (2023) 1042.
[12] **RNO-G** Collaboration, T. Glusenkamp *et al. PoS* **ICRC2023** (2023) 1056.
[13] **RNO-G** Collaboration, Z. Meyers *et al. PoS* **ICRC2023** (2023) 1142.
[14] **Arianna** Collaboration, A. Anker *et al. JINST* **17** no. 03, (2022) P03007.
[15] P. Allison *et al. Nucl. Instrum. Meth. A* **930** (2019) 112–125.
[16] C. Glaser *et al. Eur. Phys. J. C* **80** no. 2, (2020) 77.
[17] "Cyclone® V 5CGXC7 FPGA." `https://ark.intel.com/content/www/us/en/ark/products/210452/cyclone-v-5cgxc7-fpga.html`. Accessed: 2023-06-01.
[18] A. Connolly *Phys. Rev. D* **105** no. 12, (2022) 123012.
[19] N. Heyer and C. Glaser *Eur. Phys. J. C* **83** no. 2, (2023) 124.

Full Author List: RNO-G Collaboration

J. A. Aguilar[1], P. Allison[2], D. Besson[3], A. Bishop[10], O. Botner[4], S. Bouma[5], S. Buitink[6], W. Castiglioni[8], M. Cataldo[5], B. A. Clark[7], A. Coleman[4], K. Couberly[3], P. Dasgupta[1], S. de Kockere[9], K. D. de Vries[9], C. Deaconu[8], M. A. DuVernois[10], A. Eimer[5], C. Glaser[4], T. Glüsenkamp[4], A. Hallgren[4], S. Hallmann[11], J. C. Hanson[12], B. Hendricks[14], J. Henrichs[11,5], N. Heyer[4], C. Hornhuber[3], K. Hughes[8], T. Karg[11], A. Karle[10], J. L. Kelley[10], M. Korntheuer[1], M. Kowalski[11,15], I. Kravchenko[16], R. Krebs[14], R. Lahmann[5], P. Lehmann[5], U. Latif[9], P. Laub[5], C.-H. Liu[16], J. Mammo[16], M. J. Marsee[17], Z. S. Meyers[11,5], M. Mikhailova[3], K. Michaels[8], K. Mulrey[13], M. Muzio[14], A. Nelles[11,5], A. Novikov[19], A. Nozdrina[3], E. Oberla[8], B. Oeyen[18], I. Plaisier[5,11], N. Punsuebsay[19], L. Pyras[11,5], D. Ryckbosch[18], F. Schlüter[1], O. Scholten[9,20], D. Seckel[19], M. F. H. Seikh[3], D. Smith[8], J. Stoffels[9], D. Southall[8], K. Terveer[5], S. Toscano[1], D. Tosi[10], D. J. Van Den Broeck[9,6], N. van Eijndhoven[9], A. G. Vieregg[8], J. Z. Vischer[5], C. Welling[8], D. R. Williams[17], S. Wissel[14], R. Young[3], A. Zink[5]

[1] Université Libre de Bruxelles, Science Faculty CP230, B-1050 Brussels, Belgium
[2] Dept. of Physics, Center for Cosmology and AstroParticle Physics, Ohio State University, Columbus, OH 43210, USA
[3] University of Kansas, Dept. of Physics and Astronomy, Lawrence, KS 66045, USA
[4] Uppsala University, Dept. of Physics and Astronomy, Uppsala, SE-752 37, Sweden
[5] Erlangen Center for Astroparticle Physics (ECAP), Friedrich-Alexander-University Erlangen-Nürnberg, 91058 Erlangen, Germany
[6] Vrije Universiteit Brussel, Astrophysical Institute, Pleinlaan 2, 1050 Brussels, Belgium
[7] Department of Physics, University of Maryland, College Park, MD 20742, USA
[8] Dept. of Physics, Enrico Fermi Inst., Kavli Inst. for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA
[9] Vrije Universiteit Brussel, Dienst ELEM, B-1050 Brussels, Belgium
[10] Wisconsin IceCube Particle Astrophysics Center (WIPAC) and Dept. of Physics, University of Wisconsin-Madison, Madison, WI 53703, USA
[11] Deutsches Elektronen-Synchrotron DESY, Platanenallee 6, 15738 Zeuthen, Germany
[12] Whittier College, Whittier, CA 90602, USA
[13] Dept. of Astrophysics/IMAPP, Radboud University, PO Box 9010, 6500 GL, The Netherlands
[14] Dept. of Physics, Dept. of Astronomy & Astrophysics, Penn State University, University Park, PA 16801, USA
[15] Institut für Physik, Humboldt-Universität zu Berlin, 12489 Berlin, Germany
[16] Dept. of Physics and Astronomy, Univ. of Nebraska-Lincoln, NE, 68588, USA
[17] Dept. of Physics and Astronomy, University of Alabama, Tuscaloosa, AL 35487, USA
[18] Ghent University, Dept. of Physics and Astronomy, B-9000 Gent, Belgium
[19] Dept. of Physics and Astronomy, University of Delaware, Newark, DE 19716, USA
[20] Kapteyn Institute, University of Groningen, Groningen, The Netherlands

Acknowledgments

We are thankful to the staff at Summit Station for supporting our deployment work in every way possible. We also acknowledge our colleagues from the British Antarctic Survey for embarking on the journey of building and operating the BigRAID drill for our project. We would like to acknowledge our home institutions and funding agencies for supporting the RNO-G work; in particular the Belgian Funds for Scientific Research (FRS-FNRS and FWO) and the FWO programme for International Research Infrastructure (IRI), the National Science Foundation (NSF Award IDs 2118315, 2112352, 211232, 2111410) and the IceCube EPSCoR Initiative (Award ID 2019597), the German research foundation (DFG, Grant NE 2031/2-1), the Helmholtz Association (Initiative and Networking Fund, W2/W3 Program), the University of Chicago Research Computing Center, and the European Research Council under the European Unions Horizon 2020 research and innovation programme (grant agreement No 805486).