PROCEEDINGS OF SCIENCE

Identifying Cosmic Ray Radio Emission in RNO-G's Deep Antennas

The RNO-G Collaboration
(a complete list of authors can be found at the end of the proceedings)

E-mail: blh5615@psu.edu

Ultra-high-energy neutrinos are an invaluable messenger for learning about the most energetic and distant astrophysical processes in the universe. Detecting them is challenging due to their extremely low flux and small cross section, necessitating immense detection volumes. The Radio Neutrino Observatory in Greenland (RNO-G) addresses this challenge by leveraging the Askaryan effect, using sparse radio instrumentation to detect neutrino-induced cascades in ice. However, cosmic rays represent an important background, as their induced showers produce radio emission that can mimic neutrino signals; both in-air emission, primarily through the geomagnetic effect, and in-ice Askaryan emission can act as backgrounds for a neutrino search. At the same time, these radio emissions from cosmic-ray induced air showers provide a crucial tool for detector calibration and validation. This work presents progress toward quantifying and separating these backgrounds in RNO-G's deep antennas using machine learning classification methods.

Corresponding authors: Bryan Hendricks[1*]
[1] *Dept. of Physics, Penn State University, University Park, PA 16802, USA*

[*] Presenter

39th International Cosmic Ray Conference (ICRC2025)
15–24 July 2025
Geneva, Switzerland

PoS(ICRC2025)1057

1. Introduction and Motivation

Ultra-high-energy (UHE) neutrinos can be produced through interactions of UHE cosmic rays (CRs) and serve as powerful messengers of information on the sources capable of accelerating UHECRs to such high energies [1]. As neutral, nearly massless particles, neutrinos travel undeflected and with minimal attenuation, preserving directional information, unlike high-energy CRs. However, this coupled with their low flux at UHE also makes them challenging to detect: volumes of $O(100\ \mathrm{km}^3)$ are needed to observe enough of them over practical timescales [2], making traditional detection methods prohibitively expensive. Radio detection offers a scalable alternative. Neutrino-induced cascades in dense, dielectric media produce coherent radio signals via the Askaryan effect [3]. In glacial ice, where radio attenuation lengths are $O(1\ \mathrm{km})$, detectors can be sparsely spaced, making large-scale arrays feasible and in-ice radio detection a highly promising approach.

The Radio Neutrino Observatory in Greenland (RNO-G) is being constructed to realize this potential [4]. RNO-G currently operates 8 of a planned 35 autonomous stations embedded in the Greenland ice, with the stations including two primary components: a shallow array of up- and down-facing high-gain log-periodic dipole antennas, and a deep array of dipoles deployed in 100 m deep boreholes optimized for Askaryan signal detection and reconstruction. See [5] for a full overview of the current status of the detector.

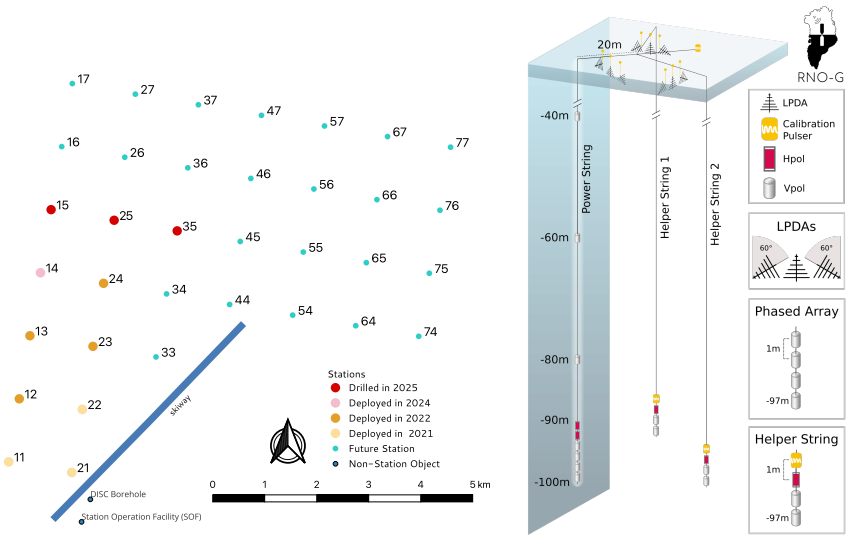

Figure 1: Left: RNO-G existing and future station map reflecting the progress as of mid-season 2025. Right: Layout of each individual RNO-G station. 24 antennas (Vpol, Hpol, and LPDA) are installed in trenches and in holes down to 100 m in depth [5].

Although RNO-G's primary goal is the discovery of UHE neutrinos, this requires detailed understanding of backgrounds that can mimic or obscure potential neutrino signals - chief among them, CR-induced radio signals. CR air showers can produce both geomagnetic and Askaryan radio signals, including from air-shower cores reaching the air-ice boundary, made possible by RNO-G's 3.2 km altitude. While surface signals can often be vetoed using the shallow component, deep in-ice signals are harder to reject, as studies suggest that cores producing emission detectable in

the deep channels deposit most of their energy in the ice and lack strong surface radio signals [6]. Compounding this challenge, simulations predict a detection rate of ~10 CRs/year/station for the in-ice signals alone [6], already exceeding the expected neutrino rate. Yet, this abundant CR population offers critical opportunities for essential calibration and validation. This work focuses on identifying the CR signals in the deep channels of RNO-G. A complementary template-based analysis of air-shower detection using the shallow channels appears in [7], with follow-up results presented in [8]. Alternative methods for the deep CR search are also being explored in [9].

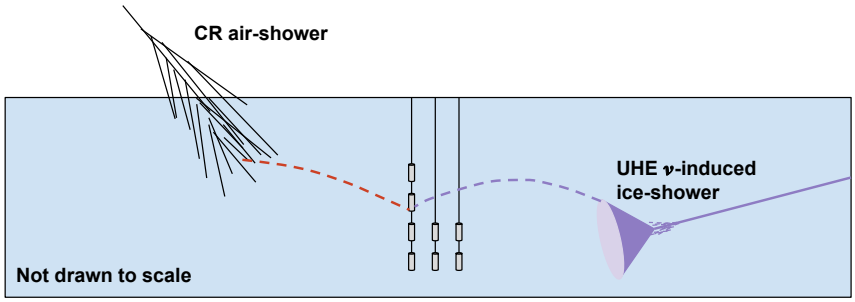

Figure 2: CR shower core entering the ice (on the left) and emitting radio signal which mimics radiation from neutrino-induced showers in the ice (shown on the right).

2. Analysis Pipeline

Initial livetime 2021–2023	
Station	**Days**
11	292.2
12	138.8
13	240.8
21	321.6
22	158.5
23	228.7
24	218.0

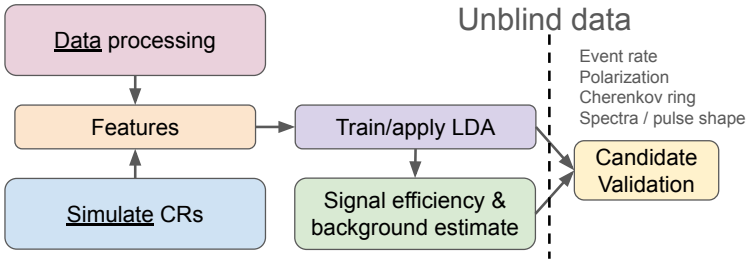

Figure 3: Left: Initial livetime for 2021-2023 by station. Right: Pipeline showing each analysis stage.

The pathway for the analysis can be seen in Fig. 3 (right), and is informed by similar analyses done by ARA [10], ANITA [11], and ARIANNA [12]. Given that RNO-G's trigger dynamically adjusts its threshold to maintain a trigger rate ~1 Hz (see [4] for more detail), the bulk of the recorded events are thermal noise. High-SNR signals are easy to identify above this background, but low-SNR events can closely resemble thermal noise, especially when viewed in individual channels. Fig. 4 (left) shows such a simulated CR, where the traces are visually difficult to separate from noise. Yet, with appropriate signal processing, the underlying structure can be recovered (right), demonstrating that even subtle signals can be extracted under favorable conditions. Still, there are cases where no single procedure provides sufficient separation power on its own. In such situations, a multivariate approach that simultaneously incorporates several observables offers a more robust strategy for distinguishing true CR signals from thermal backgrounds. We employ a machine learning classifier to perform this separation.

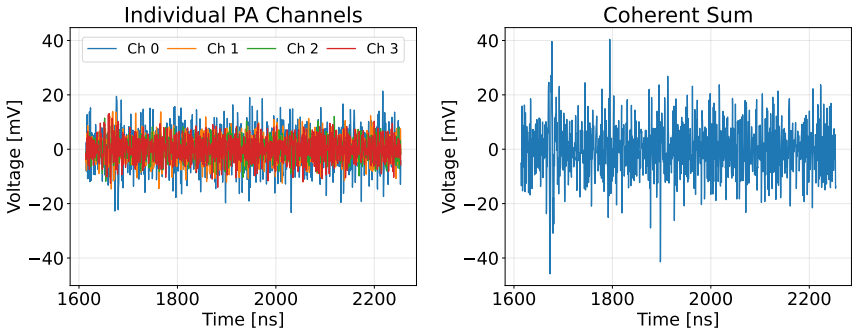

Figure 4: Traces illustrating how a **simulated** low-SNR CR-proxy event (left) closely resembles thermal noise when examined in individual PA channels, but emerges as a clear signal after coherently summing the PA channels (right).

Because this analysis in part targets a signal that has never been observed in nature, radio Askaryan emission from cosmic rays in ice, using a classification method with transparent decision-making is valuable for building confidence in candidate events and understanding classifier behavior. At the same time, the growing RNO-G dataset demands a computationally lightweight approach that can be iterated quickly during ongoing development. Linear Discriminant Analysis (LDA), while simple, offers interpretable feature weights and efficient training, making it well-suited to early stage searches. LDA finds the linear combination of features that best separates two or more classes by maximizing the ratio of between-class variance to within-class variance. In our case, the two classes are thermal noise and CRs. Before getting into which features were selected for the classification, we will go back to the start of the pipeline and first discuss the data processing.

2.1 Data processing

Currently, years 2021-2023 for all stations constructed in this period are under analysis for the deep CR search, encompassing multiple years of livetime (Fig. 3, left). Data recorded from 2024 onward contain a change in sample rate and timing offsets that significantly affects the analysis approach and is reserved for future work. In these proceedings, we present preliminary results from station 13's (see Fig. 1, left) 2022 data, which starts with 75.6 days of livetime. Before classifier training and application, several cleaning steps are applied. Only events triggered by the four deep adjacent Vpols denoted as the phased array (PA) are retained to ensure visibility in deep channels (despite the "phased" label, the events used here were recorded with a simple high-low coincident trigger). Periods of elevated noise are also systematically removed using a rolling 2 Hz trigger rate cut, over a 30 s window with 1 s steps, comfortably above the nominal 1 Hz rate. The analysis is performed in a blind fashion, with only 10% of the remaining livetime (burn sample) used for its development.

We develop further cuts on the burn sample to suppress impulsive backgrounds not associated with CRs. We remove events in periods with wind speeds above 10 m/s, as the triboelectric effect can be induced by high winds to produce impulsive radio signals strong enough to reach in-ice detectors and mimic CRs [13]. Airplane-correlated events and directions with persistent excesses are likewise excluded. These removals are necessary, as the classifier is built to distinguish only between thermal noise and CRs; constructing additional classes would require both reliable labeling

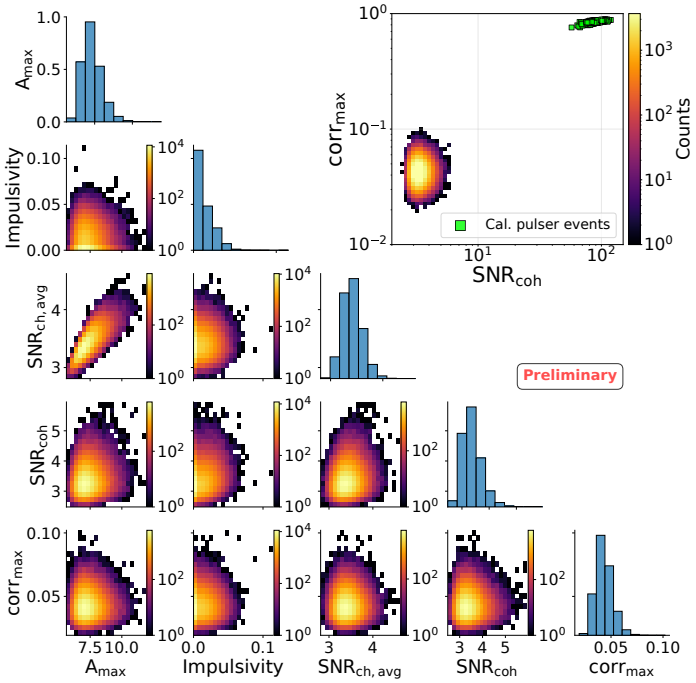

Figure 5: Station 13 burn sample histograms for a subset of features. Left to right and top to bottom: Max amplitude, impulsivity, PA-averaged SNR, coherent SNR, surface correlation, and max correlation. Top right: Example showing separation of impulsive calibration pulses from the burn sample.

and sufficient statistics, which are currently limited for other impulsive event types. With all cuts applied, the 10% burn sample for station 13 consists of 3.21 days of livetime. Much of the excluded livetime may be recoverable through future refinements, as current cuts are conservative.

2.2 Classifier

To effectively differentiate CRs from thermal noise, our classifier leverages several key features designed to maximize their separation. These include the mean signal-to-noise ratio (SNR) of the PA channels, the coherently summed waveform (CSW) SNR, impulsivity, max correlation, surface correlation, and max amplitude. The CSW is derived by aligning the PA traces until they are maximally correlated, then summing them (Fig. 4, right). Impulsivity quantifies how temporally concentrated the CSW trace's power is. Max correlation represents the peak correlation value extracted from a correlation map, which is based on a ray-traced interferometric direction reconstruction of the event using the PA channels. Surface correlation is the max correlation from that same map, but specifically within the first 10 m beneath the ice surface. This value should be high for CR events, as any CR showers producing detectable signals must originate from above, and the majority of their in-ice core development occurs at shallow depths.

Fig. 5 presents a corner plot of histograms for a subset of these features for the station 13 2022 burn sample. Each histogram is consistent with expectations for thermal noise. This is crucial as each of these events is given the thermal noise label in the training data, and a significant portion of non-thermal events with this label would contaminate the training and reduce performance. A few slight outliers remain and are being investigated. An additional plot comparing to impulsive events

Simulation Set Summary			
Qty	Azimuth	Zenith	Energy
Range	0-360°	8-60°	10^{16} – 10^{19} eV
Bin size	5°	0.25 in $\cos(\theta)$	Half decades
Sampling Dist.	Uniform	Uniform $\cos(\theta)$	~Uniform (triggered)

Table 1: Summary of CR-proxy simulation set developed for training.

is included, using an in-situ calibration pulse dataset. Training a classifier to distinguish CRs from thermal noise requires labeled examples of both classes. While the ideal training set would include CR radio events observed in RNO-G's in-ice environment, no confirmed events are yet available for training. As a result, we construct our signal class using simulated CRs.

2.3 Simulations

CR training data are currently generated using NuRadioMC, with neutrino-induced in-ice showers serving as proxies due to the similarity of their radio signals under appropriate conditions [6] and the lack of a dedicated genuine-CR simulation set. By forcing neutrino interactions to occur in the ice in regions and directions expected for CR showers, both the time- and frequency-domain signal characteristics become comparable. While this remains an approximation and excludes contributions from in-air signals, dedicated genuine-CR simulations are in progress.

For this analysis, over 600,000 triggered CR-proxy events were generated to span the expected CR feature space (see Table 1) and then downsampled to match the number of events in data for balanced training. Showers were generated to produce roughly equal numbers of triggered events per energy bin, and reweighting to match the true flux has not yet been implemented. This leads to biases in signal efficiency and background estimates. Future efforts will incorporate proper flux-weights.

In parallel, a simulation set using the most complete available framework, FAERIE [14], is being developed. FAERIE combines the CORSIKA7 + CoREAS and GEANT4 codes to model full-cascade development from first principles, including both in-air and in-ice components. However, its computational demands currently prevent the generation of sufficient event statistics for this study. Ongoing developments to address these limitations are presented in [15].

3. Preliminary Results and Outlook

The classifier is trained on 70% of the burn sample and an equal number of simulated CR proxies, with performance validated on a 20% set. Features and the decision boundary are refined iteratively, with a final 10% test set reserved for unbiased estimates after tuning iterations are complete. Each event receives an LDA score, calculated as a linear combination of input features weighted by coefficients learned during training. The distribution of LDA scores for the validation set can be seen in Fig. 6. Higher scores correspond to events being more CR-like. Overlap at low scores is expected, as very low-SNR CRs can still be confused for thermal noise. Nevertheless, the classifier successfully recovers a substantial portion of low-SNR events.

Signal efficiency is defined as the fraction of simulated CR-proxies passing above a given LDA score threshold value. Thermal background is estimated by fitting an exponential to the tail of the

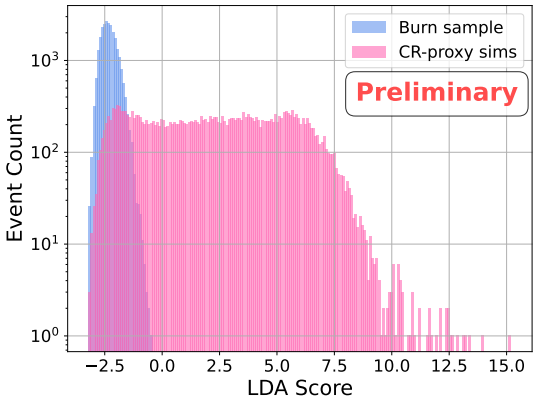

Figure 6: Histogram of LDA scores for the burn sample and simulated CR-proxies for the 20% validation set.

validation set burn sample LDA score distribution (Fig 7, left) and integrating from the threshold outward. To quantify uncertainties, 10,000 Monte Carlo samples of parameter pairs (a_i, b_i) are drawn from their joint probability distribution using the covariance matrix from the fit. Each parameter pair yields a different background estimate via integration, creating a distribution from which the median and 16th/84th percentiles provide the central value and 68% confidence intervals. This process is repeated across thresholds to yield the signal efficiency vs. background curve (Fig 7, right), with all background estimates scaled up for the full cleaned dataset livetime (~32 days in this case).

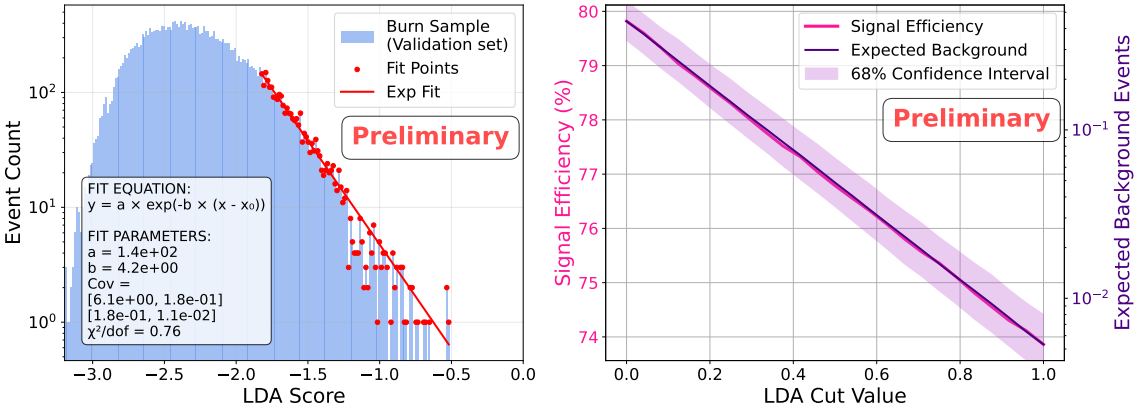

Figure 7: Left: Exponential fit to the tail of the burn sample (validation set) LDA score distribution. Right: Signal efficiency and expected background versus LDA score threshold. Background estimates are scaled to full dataset livetime (~32 days). Note the different y-axis scales.

Across a range of thresholds, both signal efficiency and background levels remain within acceptable bounds. A final LDA score cut will be optimized to ensure $\geq 5\sigma$ detection significance while maximizing overall signal efficiency. These preliminary results suggest that, with the years of livetime (post-cuts) currently under analysis, including station 13, and assuming the conservative ice-only estimate of ~10 CRs/year/station [6], detection at the desired significance is within reach. After unblinding, the classifier will be applied to the full dataset to identify candidate events, which

will then be validated using polarization, direction, pulse shape, spectral content, and Cherenkov ring structure. Though some backgrounds may mimic CRs, cross-checks across these observables will help isolate true CRs and confirm the analysis.

4. Conclusion

In this analysis, we explored progress towards identifying CR signals in RNO-G's deep antennas. Multiple years of cumulative station livetime from 2021-2023 spanning 7 stations are currently under advanced stages of analysis. Preliminary results were shown for station 13's 2022 burn sample. Results are promising, with candidate detection and validation within reach, though current estimates will change with simulation energy weighting. The full analysis results for the 2021-2023 dataset, including the additional stations, will be released soon following simulation updates, subsequent cut optimization, and unblinding. The final analysis will support detector calibration and demonstrate RNO-G's capability for detecting in-ice Askaryan radio signals.

References

[1] M. S. Muzio *The European Physical Journal Special Topics* (Feb., 2025) .

[2] V. Decoene, "Review of neutrino experiments searching for astrophysical neutrinos," 2023. https://arxiv.org/abs/2309.17139.

[3] G. Askaryan *J. Phys. Soc. Japan* **17** no. Suppl A, (1962) .

[4] **RNO-G** Collaboration, J. A. Aguilar *et al.*, "Instrument design and performance of the first seven stations of rno-g," 2025. https://arxiv.org/abs/2411.12922.

[5] **RNO-G** Collaboration, A. Nelles *et al. PoS* **ICRC** (2025) 1129.

[6] A. Coleman, C. Glaser, R. Rice-Smith, S. Barwick, and D. Besson *Astroparticle Physics* **172** (Oct., 2025) 103136.

[7] **RNO-G** Collaboration, J. Henrichs *et al. PoS* **ICRC2023** (2023) 259.

[8] **RNO-G** Collaboration, J. Henrichs *et al. PoS* **ICRC** (2025) 288.

[9] **RNO-G** Collaboration, C. Liu *et al. PoS* **ICRC** (2025) 319.

[10] **ARA** Collaboration, P. Allison *et al. Physical Review D* **105** no. 12, (June, 2022) .

[11] **ANITA** Collaboration, S. Hoover *et al. Physical Review Letters* **105** no. 15, (Oct., 2010) .

[12] **ARIANNA** Collaboration, S. Barwick *et al. Astroparticle Physics* **90** (Apr., 2017) 50–68.

[13] **RNO-G** Collaboration, J. A. Aguilar *et al.*, "Triboelectric backgrounds to radio-based uhe neutrino experiments," 2022. https://arxiv.org/abs/2103.06079.

[14] S. De Kockere, D. Van den Broeck, U. A. Latif, K. D. de Vries, N. van Eijndhoven, T. Huege, and S. Buitink *Physical Review D* **110** no. 2, (July, 2024) .

[15] **RNO-G** Collaboration, N. Punsuebsay *et al. PoS* **ICRC** (2025) 366.

Full Author List: RNO-G (June 30th, 2025)

S. Agarwal[1], J. A. Aguilar[2], N. Alden[3], S. Ali[1], P. Allison[4], M. Betts[5], D. Besson[1], A. Bishop[6], O. Botner[7], S. Bouma[8], S. Buitink[9,10], R. Camphyn[2], J. Chan[6], S. Chiche[2], B. A. Clark[11], A. Coleman[7], K. Couberly[1], S. de Kockere[12], K. D. de Vries[12], C. Deaconu[3], P. Giri[13], C. Glaser[7], T. Glüsenkamp[7], H. Gui[4], A. Hallgren[7], S. Hallmann[14,8], J. C. Hanson[15], K. Helbing[16], B. Hendricks[5], J. Henrichs[14,8], N. Heyer[7], C. Hornhuber[1], E. Huesca Santiago[14], K. Hughes[4], A. Jaitly[14,8], T. Karg[14], A. Karle[6], J. L. Kelley[6], C. Kopper[8], M. Korntheuer[2,12], M. Kowalski[14,17], I. Kravchenko[13], R. Krebs[5], M. Kugelmeier[6], R. Lahmann[8], C.-H. Liu[13], M. J. Marsee[18], K. Mulrey[10], M. Muzio[6,5], A. Nelles[14,8], A. Novikov[19], A. Nozdrina[4], E. Oberla[3], B. Oeyen[20], N. Punsuebsay[19], L. Pyras[14,21], M. Ravn[7], A. Rifaie[16], D. Ryckbosch[20], F. Schlüter[2], O. Scholten[12,22], D. Seckel[19], M. F. H. Seikh[1], Z. S. Selcuk[14,8], J. Stachurska[20], J. Stoffels[12], S. Toscano[2], D. Tosi[6], J. Tutt[5], D. J. Van Den Broeck[12,9], N. van Eijndhoven[12], A. G. Vieregg[3], A. Vijai[11], C. Welling[3], D. R. Williams[18], P. Windischhofer[3], S. Wissel[5], R. Young[1], A. Zink[8]

[1] University of Kansas, Dept. of Physics and Astronomy, Lawrence, KS 66045, USA
[2] Université Libre de Bruxelles, Science Faculty CP230, B-1050 Brussels, Belgium
[3] Dept. of Physics, Dept. of Astronomy & Astrophysics, Enrico Fermi Inst., Kavli Inst. for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA
[4] Dept. of Physics, Center for Cosmology and AstroParticle Physics, Ohio State University, Columbus, OH 43210, USA
[5] Dept. of Physics, Dept. of Astronomy & Astrophysics, Center for Multimessenger Astrophysics, Institute of Gravitation and the Cosmos, Pennsylvania State University, University Park, PA 16802, USA
[6] Wisconsin IceCube Particle Astrophysics Center (WIPAC) and Dept. of Physics, University of Wisconsin-Madison, Madison, WI 53703, USA
[7] Uppsala University, Dept. of Physics and Astronomy, Uppsala, SE-752 37, Sweden
[8] Erlangen Centre for Astroparticle Physics (ECAP), Friedrich-Alexander-University Erlangen-Nürnberg, 91058 Erlangen, Germany
[9] Vrije Universiteit Brussel, Astrophysical Institute, Pleinlaan 2, 1050 Brussels, Belgium
[10] Dept. of Astrophysics/IMAPP, Radboud University, PO Box 9010, 6500 GL, The Netherlands
[11] Department of Physics, University of Maryland, College Park, MD 20742, USA
[12] Vrije Universiteit Brussel, Dienst ELEM, B-1050 Brussels, Belgium
[13] Dept. of Physics and Astronomy, Univ. of Nebraska-Lincoln, NE, 68588, USA
[14] Deutsches Elektronen-Synchrotron DESY, Platanenallee 6, 15738 Zeuthen, Germany
[15] Whittier College, Whittier, CA 90602, USA
[16] Dept. of Physics, University of Wuppertal D-42119 Wuppertal, Germany
[17] Institut für Physik, Humboldt-Universität zu Berlin, 12489 Berlin, Germany
[18] Dept. of Physics and Astronomy, University of Alabama, Tuscaloosa, AL 35487, USA
[19] Dept. of Physics and Astronomy, University of Delaware, Newark, DE 19716, USA
[20] Ghent University, Dept. of Physics and Astronomy, B-9000 Gent, Belgium
[21] Department of Physics and Astronomy, University of Utah, Salt Lake City, UT 84112, USA
[22] Kapteyn Institute, University of Groningen, PO Box 800, 9700 AV, The Netherlands

Acknowledgments

We are thankful to the support staff at Summit Station for making RNO-G possible. We also acknowledge our colleagues from the British Antarctic Survey for 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 2112352, 2111232, 2111410, 2411590, and collaborative awards 2310122 through 2310129), and the IceCube EPSCoR Initiative (Award ID 2019597), the Helmholtz Association, the Swedish Research Council (VR, Grant 2021-05449 and 2021-00158), the University of Chicago Research Computing Center, and the European Union under the European Unions Horizon 2020 research and innovation programme (grant agreements No 805486), as well as (ERC, Pro-RNO-G No 101115122 and NuRadioOpt No 101116890).