A novel approach for air shower profile reconstruction with dense radio antenna arrays using Information Field Theory

K. Watanabe,[a,*] **S. Bouma,**[b] **J.D. Bray,**[c] **S. Buitink,**[d,e] **A. Corstanje,**[d,e] **V. De Henau,**[d] **M. Desmet,**[d] **E. Dickinson,**[f] **L. van Dongen,**[e] **T.A. Enßlin,**[g,h,i] **B. Hare,**[j,k] **H. He,**[l] **J.R. Hörandel,**[e] **T. Huege,**[a,d] **C.W. James,**[f] **M. Jetti,**[g,h] **P. Laub,**[b] **H.J. Mathes,**[a] **K. Mulrey,**[e,m] **A. Nelles,**[b,n] **S. Saha,**[a,o] **O. Scholten,**[j] **S. Sharma,**[b] **R.E. Spencer,**[c] **C. Sterpka,**[k] **S. ter Veen,**[k] **K. Terveer,**[b] **S. Thoudam,**[p] **T.N.G. Trinh,**[q] **P. Turekova,**[j,k] **D. Veberič,**[a] **M. Waterson,**[r] **C. Zhang,**[s,t] **P. Zhang**[u] **and Y. Zhang**[l,v]

E-mail: keito.watanabe@kit.edu

Reconstructing the longitudinal profile of extensive air showers, generated from the interaction of cosmic rays in the Earth's atmosphere, is crucial to understanding their mass composition, which in turn provides valuable insight on their possible sources of origin. Dense radio antenna arrays such as the LOw Frequency ARray (LOFAR) telescope as well as the upcoming Square Kilometre Array Observatory (SKAO) are ideal instruments to explore the potential of air shower profile reconstruction, as their high antenna density allows cosmic ray observations with unprecedented accuracy. However, current analysis approaches can only recover $X_{\max}$, the atmospheric depth at shower maximum, and heavily rely on computationally expensive simulations. As such, it is ever more crucial to develop new analysis approaches that can perform a full air shower profile reconstruction efficiently.

In this work, we develop a novel framework to reconstruct the longitudinal profile of air showers using measurements from radio detectors with Information Field Theory (IFT), a state-of-the-art reconstruction framework based on Bayesian inference. Through IFT, we are able to exploit all available information in the signal (amplitude, phase, and pulse shape) at each antenna position simultaneously and explicitly utilise models that are motivated through our current understanding of air shower physics. We verify our framework on simulated datasets prepared for LOFAR, showcasing that we can not only reconstruct the air shower profile with uncertainties in each atmospheric depth bin but also recover the reconstructed trace at each antenna position. Our framework demonstrates that radio measurements with dense antenna layouts such as LOFAR and SKAO have the capability to go beyond reconstruction of $X_{\max}$ and will thus aid in our understanding of the mass composition of cosmic rays.

ICRC 2025
The Astroparticle Physics Conference
Geneva July 15-24, 2025

---

*Speaker

1. Introduction

The origins of high-energy cosmic rays remain elusive, and understanding their primary mass composition is essential for uncovering the sources of these particles. Radio emission produced by cosmic ray air showers has proven to be an effective method for cosmic ray detection. Notably, the upcoming Square Kilometre Array Observatory (SKAO), with its SKA-Low array having over 60,000 antennas distributed across a $1\,\mathrm{km}^2$ core area, will greatly enhance our capacity to study cosmic rays [1]. Current reconstruction techniques for mass composition in dense antenna arrays such as the LOw Frequency ARray (LOFAR) and SKA-Low rely on energy fluence measurements from numerous CoREAS simulations [2] to estimate X_{max}, the atmospheric depth at shower maximum [3, 4]. However, additional insights into mass composition and the hadronic interaction models underlying the interpretation of measured data can be obtained from other parameters, including the width and asymmetry of the longitudinal profile of the air shower [5]. With its unprecedented precision in measuring cosmic ray air showers, SKA-Low has the capability to reconstruct more than just X_{max} [6, 7], making it imperative to develop analyses that can extract this additional information from the radio emission.

In this work, we showcase the development of a novel approach to reconstruct the longitudinal profile of air showers using Information Field Theory (IFT) [8]. This method allows us to not only reconstruct shower parameters more than just X_{max}, but also do so without explicitly relying on simulations. In our model, we parameterise the longitudinal profile as a Gaisser-Hillas function and utilise a fast forward model based on template synthesis [9] to simulate the radio emission from air showers. We include the antenna response for antennas used in SKA-Low for a realistic detector description. With IFT, we infer the shower parameters from realistically simulated data from CoREAS simulations and demonstrate the reconstruction efficiency of the longitudinal profile of air showers on the basis of an idealised star-shaped antenna layout. Through our method, we highlight the potential of air shower profile reconstruction for dense radio detectors such as SKA-Low.

2. Model

The model is constructed through IFT, a Bayesian reconstruction framework that can be applied to signal reconstruction problems [8]. IFT allows us to infer parameters of a model from noisy data by combining domain and prior knowledge and likelihood functions motivated by our current physical understanding of air showers. Furthermore, its inference procedure based on variational inference makes it possible to model parameters defined at each atmospheric depth bin, making it an ideal tool for reconstructing the entire profile. The Bayesian nature of IFT also allows us to reconstruct the signal through the inferred shower parameters. IFT has already shown its applicability for reconstructing the electric field signal from air showers [10, 11].

To utilise the IFT framework, the forward model (prior + likelihood function) needs to be fully differentiable. This is automatically encoded in JAX, the language framework that is used by the numerical framework for IFT (NIFTy, [12]). Another necessary requirement is for the forward model to be computationally performant, as hundreds of instances of the forward process will be executed per reconstruction.

The model used in this work consists of several components, which we highlight below:

Shower Profile : The longitudinal profile of the air shower, describing the evolution of the air shower at each atmospheric depth X (in units of grams per square centimetre), is parameterised by the Gaisser-Hillas function in the LR formalism, where the width and asymmetry parameters are described by L and R respectively [5]:

$$N(X) = N_{max} \exp\left(-\frac{X - X_{max}}{RL}\right) \left(1 + \frac{R}{L}(X - X_{max})\right)^{R^{-2}}. \tag{1}$$

Here N_{max} is the number of electrons and positrons at X_{max} which is proportional to the electromagnetic energy of the cosmic ray. The implications of L and R on the mass separation between primary particles have been thoroughly studied in the context of the Pierre Auger Observatory [5] and SKAO [6, 7]. Furthermore, their correlations are more uniform as compared to other standard parameterisations (such as with X_0 and λ). As such, it is a suitable choice for our model, and we can interpret the reconstructed shower parameters in a physically motivated manner. In order to capture fluctuations that cannot be expressed by the Gaisser-Hillas function, we also incorporate fluctuations based on the Ornstein-Uhlenbeck process at each atmospheric bin [13].
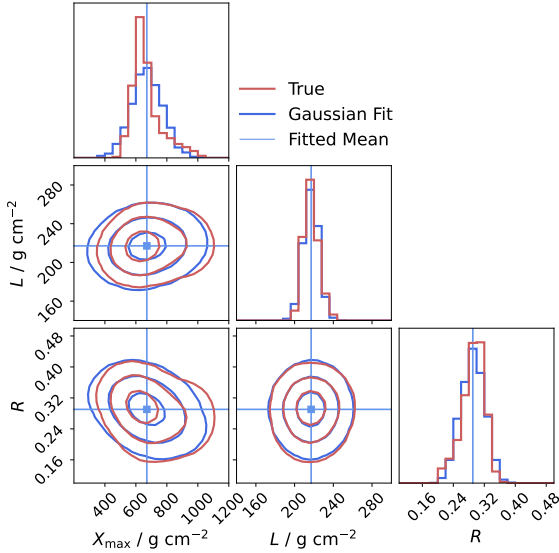

In our model, we place prior distributions on each shower parameter that encodes our physical understanding of each parameter. Previous studies have shown that X_{max}, L and R exhibit strong correlations between each other. As such, we describe the shower parameters by a multivariate Gaussian with mean and covariance matrix that encodes the correlations between the parameters. The mean and covariance are estimated from shower parameters obtained from 10,000 pre-generated CORSIKA simulations [15] with the hadronic interaction model QGSJETII-04 [16] specific to atmospheric conditions in the context of LOFAR [17]. Figure 1 shows a corner plot of the prior distribution of the shower parameters overlaid on the true distribution obtained from CORSIKA events. We also apply physical limits of $X_{max} \in [400, 1100]$ g/cm^2, $L \in [200, 240]$ g/cm^2, and $R \in [0.2, 0.4]$. As N_{max} is only weakly correlated with the other shower parameters, we set a normal prior

Figure 1: Prior distribution for X_{max}, L, and R used in this model. The true distribution obtained from CORSIKA simulations are also overlaid. The contours are plotted according to the 1, 2, and 3σ confidence level for a 2-D Gaussian distribution. Figure generated using corner.py [14].

$\log_{10}(N_{max}) \sim \text{Normal}(8, 0.5)$ independent of other parameters.

Radio Emission Model : In order to quickly but accurately model the radio emission generated from a given shower profile, we use SMIET [9], which allows us to synthesise the radio pulses from any longitudinal profile using pre-computed templates of the electric field signal at each shower geometry for a fixed antenna layout within a matter of seconds. The resulting electric fields are described by the geomagnetic and charge-excess emission that encapsulates the physical mechanisms
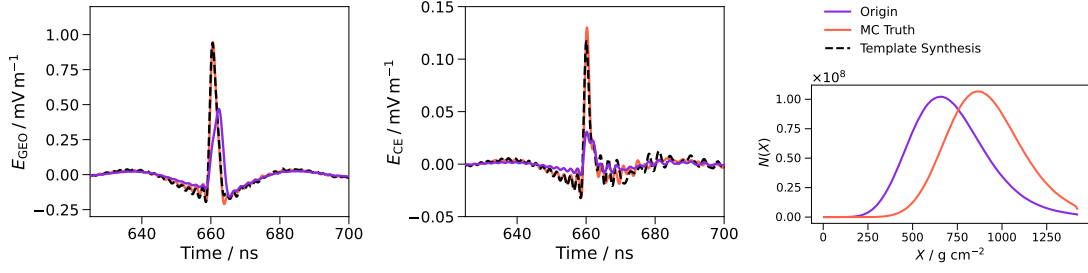
Figure 2: Left: An example of a synthesised trace using template synthesis (black, dashed) within a frequency bandwidth of $[30, 500]$ MHz at an antenna position $d_{\text{core}} = 52$ m from the shower core. We apply a template with an origin shower (overlaid in purple) with a zenith and azimuth angle of $\theta = 42°$, $\phi = 117°$ respectively. We also show the electric field traces generated from a CoREAS simulation with the same configuration (red) to verify its accuracy. Right: The longitudinal profile of the origin (purple) and synthesised shower (red), where we use the same profile from CoREAS for the synthesised shower.

of the radio emission. The template is based on an "origin shower", that is, a generated shower based on CoREAS simulations which simulate the electric field signal at fixed antenna positions from individual atmospheric slices. Using scaling relations previously and independently pre-computed from CoREAS simulations, we can re-scale the origin shower to generate the desired template. Applying this template to any given shower profile, we can synthesise the electric field pulse at each antenna position. This approach matches the accuracy of CoREAS simulations within 6% for $\Delta X_{\text{max}} \leq 100\,\text{g cm}^{-2}$ from the origin shower. Figure 2 shows an example of a synthesised trace with SMIET compared to the true electric field signal as obtained from a CoREAS simulation with the same configuration.

SMIET an ideal forward model for our reconstruction framework, and its development was motivated by the aim to apply it in an IFT reconstruction. Firstly, we can describe the radio emission from any profile without the need for extensive simulations for every possible shower geometry. Furthermore, each synthesis process can be performed on the order of seconds, vastly increasing the speed of the forward model compared to running CoREAS simulations. Lastly, its semi-analytical nature makes the model fully differentiable, which is a necessary requirement for the IFT framework. While template synthesis can be applied to other shower geometries through the interpolation of the phase spectrum, we opt to currently use only a fixed geometry for our model; we will extend our model to allow variations of the arrival direction in future works.

Antenna & Noise Model : The antenna response describes how the electric field signal is detected at each antenna, which depends on the antenna type and orientation. The electric fields can be transformed into the voltage traces observed at each antenna through the following response function H_{ij}:

$$\begin{bmatrix} V_X \\ V_Y \end{bmatrix} = \begin{bmatrix} H_{X,\theta} & H_{X,\phi} \\ H_{Y,\theta} & H_{Y,\phi} \end{bmatrix} \cdot \begin{bmatrix} E_\theta \\ E_\phi \end{bmatrix} \tag{2}$$

In our model, we use the response function modelled for the SKA-Low Log-periodic Antennas (SKALA4) [18], implemented through the NuRadioReco software framework [19]. We apply a frequency bandwidth of $[50, 200]$ MHz with a sampling rate of 500 MHz, which adequately captures the detector response of the SKA-low antennas. As SMIET synthesises pulses at fixed

ID	θ / deg	ϕ / deg	E / 10^{17} eV	X_{max}^{origin} / $g\,cm^{-2}$	N_{sims}
65490891	28.5	85.3	2.04	642	19
62804598	42.8	117.1	1.53	655	21
82321543	39.7	135.9	3.73	678	21
81409140	26.1	223.1	7.00	714	18
70360905	21.0	39.7	1.83	723	21
64021966	31.3	319.1	2.15	734	20
72094854	30.4	122.9	7.98	803	22

Table 1: Event sets (zenith, azimuth, energy, and X_{max} for the origin shower) used in this work. The number of simulations contained per event is also listed.

antenna positions, we only work with fixed star-shaped antenna layouts for now and plan to utilise a realistic antenna layout from SKA-Low through pulse interpolation [20] in future works.

The noise within the signal has a variety of sources, of which the most prominent are the Galactic radio emission and instrumental noise. While it is possible to model these noise sources separately, we characterise the combined broadband noise through a normal distribution with an RMS voltage of $\sigma_N^V = 7 \times 10^{-5}$ V such that all events used for our reconstruction have a signal-to-noise ratio (SNR), defined by the ratio of the maximum amplitude of the trace to the noise RMS, to be greater than one.

3. Results

To verify our model, we compare our results with simulated events used in [17] which have been generated using CoREAS with environmental conditions for the LOFAR site. From this dataset, we take seven event sets with varying energies and event geometries (zenith and azimuth angles) with protons as their primary particle. Each set consists of approximately 20 simulations that we use for our reconstruction. For each set, we then generate a single origin shower with $X_{max}^{origin} \geq 640\,g\,cm^{-2}$. Each origin shower is then used within our model through template synthesis. The event sets used in this work are listed in Table 1.

The simulated events are convolved with the antenna response matrix to generate the detected voltage pulse, to which we then add a random noise signal at each time bin generated from our noise model to synthetically generate data. In this way, we generate a mock signal that incorporates both broadband noise and the voltage signal. Using this data, generated for each event, we apply our model to infer the shower parameters and benchmark against the true values for each simulation.

Figure 3 shows an example of an event reconstruction, where we show the reconstructed longitudinal profile, energy fluence, electric field, and voltage traces for an event with $X_{max}^{truth} = 602\,g\,cm^{-2}$, taking event configuration 65490891 as listed in Table 1. For the electric field and voltage traces, the pulses at an antenna position with $d_{core} = 106$ m away from the shower core were taken (highlighted in green in the figure). We show that the longitudinal profile can be reconstructed within one standard deviation, except at atmospheric depths from approximately $400\,g\,cm^{-2}$ to $700\,g\,cm^{-2}$, where the profile is underestimated. We also observe that the errors increase near the edges of the atmospheric grid, which is expected as the number of particles decreases significantly
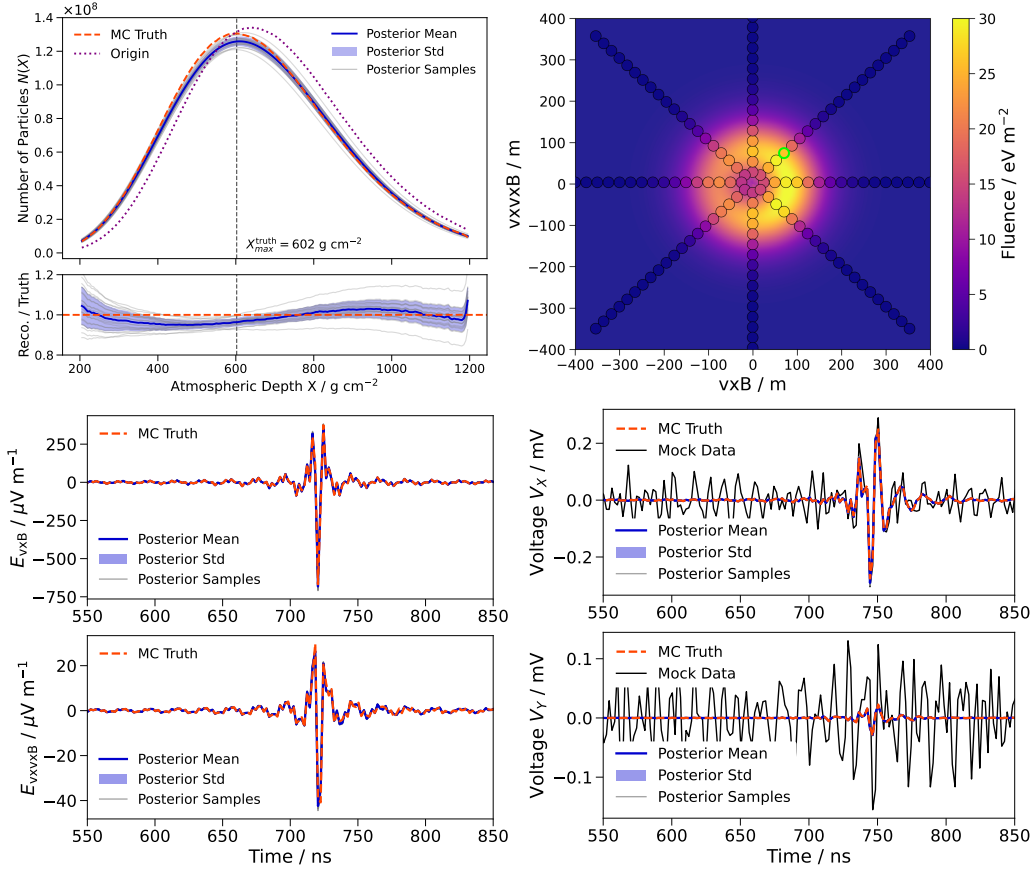
Figure 3: An example of an event reconstruction from this model for the event configuration 65490891 with $X_{max}^{truth} = 602\,\mathrm{g\,cm^{-2}}$. Top Left: The posterior mean and standard deviation (blue) of the longitudinal profile, compared to the true profile from the CoREAS simulation (red, dashed). Each posterior sample used for inference (gray) and the profile from the origin shower used in the template (purple, dotted) is also shown. The ratio between the true and reconstructed profiles are shown in the subplot. Top Right: The posterior mean of the energy fluence at each antenna (depicted by circles) in the shower plane. The interpolated true energy fluence is overlaid to show the qualitative agreement with the reconstructed results. Bottom Left: The reconstructed electric field traces in the shower plane at an antenna $d_{core} = 106\,\mathrm{m}$ away from the shower core (indicated by the green circle on the fluence map). Bottom Right: The reconstructed voltage traces at the same antenna position in the X/Y polarisation of the SKALA antenna. The mock data used for reconstruction is also shown.

in such regions. We highlight that, through our model, we can also reconstruct both electric field and voltage signals. In particular, we can recover the amplitude and shape of the pulse for most antenna positions. For antenna positions along the vxB axis, where we approximate the emission on the vxB plane in our model by averaging the emission from nearby arms, the reconstructed fluence is slightly underestimated.

To verify our model for all events, Figure 4 shows the distribution of the bias (difference between Monte-Carlo true parameter and reconstructed value) for each reconstructed shower parameter (X_{max}, $\log_{10} N_{max}$, L, and R). We also display the correlations between the biases of each parameter. We show that on average most reconstructed shower parameters are unbiased, however with a
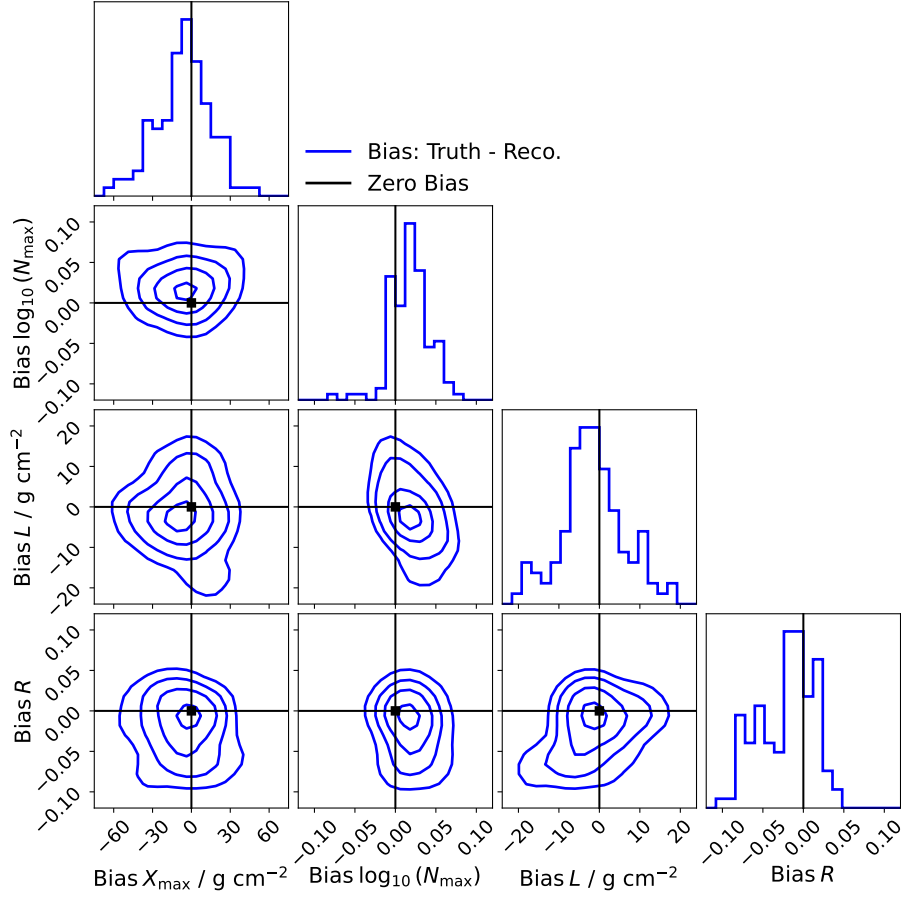
6

Figure 4: Distribution of the bias of the shower parameters X_{max}, $\log_{10}(N_{max})$, L and R for all events considered in this work. The contours show the 0.5, 1, 1.5, and 2σ values following a 2-D Gaussian distribution. The point where the bias is zero is indicated by the black lines and dots. Figure generated using `corner.py` [14].

somewhat large variation. In particular, the precision of X_{max} amounts to $25\,\mathrm{g\,cm^{-2}}$, which is still slightly larger than the reported precision from LOFAR [3] amounting to $19\,\mathrm{g\,cm^{-2}}$. Furthermore, we observe that our model currently underestimates N_{max}, which can also be observed from the underestimated profile reconstruction around X_{max}. The origin of this bias is not yet known and is under investigation. We also observe that some events overestimate the asymmetry parameter R, which is loosely correlated to the reconstruction efficiency of the shower width L. This correlation is expected, as both L and R are strongly dependent on each other [5].

4. Conclusion

In this work we showcase the development of a novel method to reconstruct the longitudinal profile of air showers using radio measurements. The reconstruction is performed via Information Field Theory, a Bayesian framework applicable for signal reconstruction. We parameterise the shower profile using the Gaisser-Hillas function in the LR formalism. The radio emission observed

at each antenna is modelled using template synthesis, where electric field pulses are synthesised for any profile with a given simulated shower. We generate voltage traces by convolving with the antenna response from the SKALA4 antennas and apply Gaussian noise to each time bin. We verified our reconstruction framework with seven event configurations with an ideal star-shaped antenna layout, showcasing that, while we currently underestimate the particle number, we are able to recover the longitudinal profile within the 1σ uncertainty band. The depth of shower maximum $X_{\max}$ is reconstructed without a bias and with a resolution of currently $25\,\mathrm{g\,cm^{-2}}$. Furthermore, we showcase that we also reconstruct the electric field and voltage traces, recovering both their shape and amplitudes.

In this work, we tested our model with approximately 100 simulations. To properly quantify the performance of our model, we plan to verify our reconstruction framework with all ~20 000 events within the same dataset. We also plan to extend this model by using pulse interpolation to apply our framework on realistic antenna layouts for LOFAR and SKA-Low. Furthermore, we hope to include the interpolation of the phase spectrum within template synthesis to perform arrival direction variation along with the shower profile, potentially lifting the restriction of single event geometry reconstructions.

References

[1] T. Huege et al., *EPJ Web Conf.* **135** (2017) 02003 [1608.08869].

[2] T. Huege et al., *AIP Conf. Proc.* **1535** (2013) 128 [1301.2132].

[3] S. Buitink et al., *Physical Review D* **90** (2014) .

[4] A. Corstanje et al., 2504.16873.

[5] S. Andringa et al., *Astropart. Phys.* **34** (2011) 360.

[6] S. Buitink et al., *PoS* **ARENA2022** (2023) 046 [2307.02907].

[7] A. Corstanje et al., *PoS* **ARENA2022** (2023) 024 [2303.09249].

[8] T.A. Enßlin, *Annalen Phys.* **531** (2019) 1800127 [1804.03350].

[9] M. Desmet et al., 2505.10459.

[10] C. Welling et al., *JCAP* **04** (2021) 071 [2102.00258].

[11] S. Strähnz et al., 2409.14970.

[12] G. Edenhofer et al., *Journal of Open Source Software* **9** (2024) 6593.

[13] G.E. Uhlenbeck and L.S. Ornstein, *Phys. Rev.* **36** (1930) 823.

[14] D. Foreman-Mackey, *The Journal of Open Source Software* **1** (2016) 24.

[15] D. Heck et al., *Forschungszentrum Karlsruhe Report* **FZKA-6019** (1998) .

[16] S. Ostapchenko, *EPJ Web Conf.* **52** (2013) 02001.

[17] A. Corstanje et al., *Phys. Rev. D* **103** (2021) 102006 [2103.12549].

[18] P. Bolli et al., *IEEE Open Journal of Antennas and Propagation* **1** (2020) 253.

[19] C. Glaser et al., *Eur. Phys. J. C* **79** (2019) 464 [1903.07023].

[20] A. Corstanje et al., *JINST* **18** (2023) P09005 [2306.13514].

Affiliations

[a] Institut für Astroteilchenphysik, Karlsruhe Institute of Technology (KIT), P.O. Box 3640, 76021 Karlsruhe, Germany

[b] Erlangen Centre for Astroparticle Physics, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91058 Erlangen, Germany

[c] Jodrell Bank Centre for Astrophysics, Department of Physics and Astronomy, University of Manchester, Manchester M13 9PL, UK

[d] Vrije Universiteit Brussel, Astrophysical Institute, Pleinlaan 2, 1050 Brussels, Belgium

[e] Department of Astrophysics/IMAPP, Radboud University Nijmegen, P.O. Box 9010, 6500 GL Nijmegen, The Netherlands

[f] International Centre for Radio Astronomy Research, Curtin University, Bentley, 6102, WA, Australia

[g] Max-Planck Institut für Astrophysik, Karl-Schwarzschild-Str. 1, 85748 Garching, Germany

[h] Ludwig-Maximilians-Universität München (LMU), Geschwister-Scholl-Platz 1, 80539 München, Germany

[i] Deutsches Zentrum für Astrophysik, Postplatz 1, 02826 Görlitz, Germany

[j] Kapteyn Astronomical Institute, University of Groningen, P.O. Box 72, 9700 AB Groningen, Netherlands

[k] Netherlands Institute for Radio Astronomy (ASTRON), Dwingeloo, The Netherlands

[l] Key Laboratory of Dark Matter and Space Astronomy, Purple Mountain Observatory, Chinese Academy of Sciences, No. 10 Yuanhua Road, Nanjing, China

[m] Nikhef, Science Park Amsterdam, 1098 XG Amsterdam, The Netherlands

[n] Deutsches Elektronen-Synchrotron DESY, Platanenallee 6, 15738 Zeuthen, Germany

[o] Department of Physics, Indian Institute of Technology Kanpur, Kanpur, UP-208016, India

[p] Department of Physics, Khalifa University, P.O. Box 127788, Abu Dhabi, United Arab Emirates

[q] Physics Education Department, School of Education, Can Tho University, Campus II, 3/2 Street, Ninh Kieu District, Can Tho City, Viet Nam

[r] SKA Observatory, Jodrell Bank, Lower Withington, Macclesfield, SK11 9FT, UK

[s] School of Astronomy and Space Science, Nanjing University, Nanjing 210023, China

[t] Key Laboratory of Modern Astronomy and Astrophysics, Nanjing University, Ministry of Education, Nanjing 210023, China

[u] School of Electronic Engineering, Xidian University, No.2 South Taibai Road, Xi'an, China

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

Acknowledgements