

Cosmic ray detection with the LOFAR radio telescope

**K. Terveer,^{a,*} S. Bouma,^a S. Buitink,^{b,c} A. Corstanje,^{b,c} M. Desmet,^c H. Falcke,^{b,d,e}
B.M. Hare,^e J.R. Hörandel,^{b,d,a} T. Huege,^{f,a} N. Karastathis,^f P. Laub,^a
K. Mulrey,^{b,d} A. Nelles,^{g,a} O. Scholten,^{h,i} P. Turekova,^{e,h} S. Thoudam,^j G. Trinh^k
and S. ter Veen^e**

^aErlangen Centre for Astroparticle Physics, Friedrich-Alexander-University Erlangen-Nürnberg, 91058 Erlangen, Germany

^bDepartment of Astrophysics/IMAPP, Radboud University Nijmegen, P.O. Box 9010, 6500 GL Nijmegen, The Netherlands

^cVrije Universiteit Brussel, Astrophysical Institute, Pleinlaan 2, 1050 Brussels, Belgium

^dNikhef, Science Park Amsterdam, 1098 XG Amsterdam, The Netherlands

^eNetherlands Institute for Radio Astronomy (ASTRON), Postbus 2, 7990 AA Dwingeloo, The Netherlands

^fInstitut für Astroteilchenphysik, Karlsruhe Institute of Technology (KIT), P.O. Box 3640, 76021 Karlsruhe, Germany

^gDeutsches Elektronen-Synchrotron DESY, Platanenallee 6, 15738 Zeuthen, Germany

^hUniversity of Groningen, Kapteyn Astronomical Institute, Groningen, 9747 AD, Netherlands

ⁱInteruniversity Institute for High-Energy, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium

^jDepartment of Physics, Khalifa University, P.O. Box 127788, Abu Dhabi, United Arab Emirates

^kDepartment of Physics, School of Education, Can Tho University Campus II, 3/2 Street, Ninh Kieu District, Can Tho City, Vietnam

E-mail: karen.terveer@fau.de

The LOw Frequency ARray (LOFAR) has successfully measured cosmic rays for over a decade now. With its dense core of antenna fields in the Netherlands, it is an ideal tool for studying the radio emission from extensive air showers in the 10^{16} eV to $10^{18.5}$ eV range. Every air shower is measured with a small particle detector array and hundreds of antennas, which sets LOFAR apart from other air shower arrays. We present our current achievements and progress in reconstruction, interpolation, and software development during the final phases of measurement of LOFAR 1.0, before the LOFAR array gets a significant upgrade, including also plans for the final data release and refined analyses.

*10th International Workshop on Acoustic and Radio EeV Neutrino Detection Activities (ARENA2024)
11-14 June 2024*

The Kavli Institute for Cosmological Physics, Chicago, IL, USA

*Speaker

1. LOFAR

The LOw Frequency ARray (LOFAR) is the worlds largest radio telescope, consisting of numerous antenna fields (stations) distributed across Europe. There are 52 stations in total, 24 of which form a dense core in the Netherlands [1]. It is a shared telescope that is primarily used for astronomy observations.

For cosmic ray detection the core stations are utilized, each equipped with 96 Low Band Antennas (LBA) that consist of two perpendicularly arranged dipole antennas operating in the 30-80 MHz band. In addition, the core is equipped with 20 scintillator units (LORA) which are used for triggering and auxiliary air shower reconstruction [2]. Once LORA detects muons from a cosmic ray induced air shower, the buffered data from the antennas are read out.

LOFAR measures the radio emission stemming from the dominant geomagnetic effect and the sub-dominant Askaryan effect. In the LOFAR frequency range (30 MHz - 80 MHz) the two emission mechanisms lead to the characteristic asymmetric bean-shaped footprint on the ground, which the antennas sample and which can be used for reconstructing shower properties such as the shower maximum X_{\max} or the cosmic-ray energy. At the LOFAR site, the radio footprint of a shower can range between a few hundred meters to a kilometer in diameter, hence the dense instrumentation of antennas plays out its power when studying the features of the footprint.

The LOFAR-method for reconstruction of the X_{\max} has established itself as the current state-of-the-art, relying on a series of simulations for each measured shower and performing a χ^2 -minimization, yielding a systematic uncertainty of 7-9 g cm⁻² [3]. The method was also used by the Pierre Auger Observatory [4].

LOFAR targets an energy range of 10¹⁶ eV to 10^{18.5} eV which is precisely the expected transition region from galactic to extragalactic sources.

2. Current developments

Recent developments in the LOFAR Key Science Project Cosmic Rays involve, among others, the switch to a new reconstruction software, the usage of interpolation of air shower simulations and new, simulation-independent approaches to reconstruction. In this section we will highlight some of the work done in the past year.

2.1 Software

A new standardized software to use for all analyses has been in development for a while, motivated by no longer maintainable custom reconstruction solutions on continuously upgraded server infrastructure. Instead of developing something stand-alone for cosmic ray detection, there are significant advantages of joining an already existing software used by other radio experiments. This ultimately led the choice of NuRadioReco [5], a python framework for detector simulation and reconstruction of radio detectors for neutrinos and cosmic rays. Advantages of NRR include its modular structure, existing documentation and active developers as well as the many pre-existing modules that can be conveniently used once a detector description exists. In the past year we have integrated the full LOFAR reconstruction pipeline [6] into NuRadioReco: This required both the full detector description, geometry and antenna and hardware responses, and all modules that are

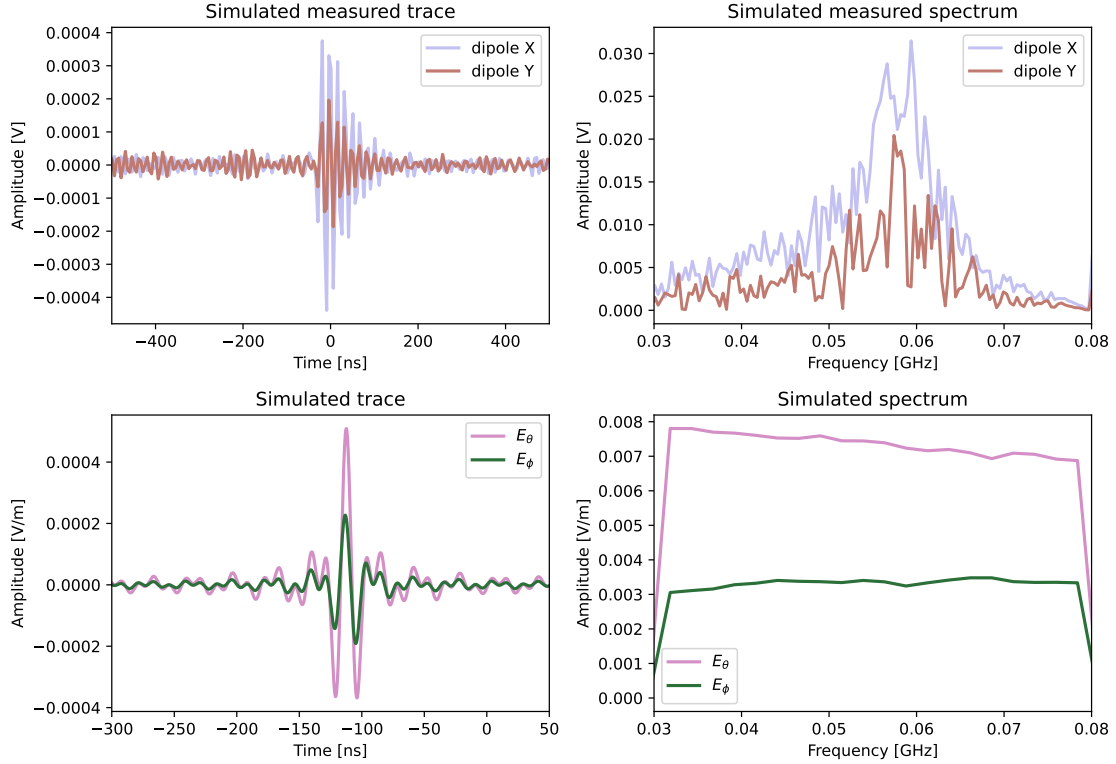


Figure 1: *Top:* Simulated trace forwarded through the LOFAR signal chain (antenna and hardware) in NuRadioReco, with added Galaxy noise. *Bottom:* Original Electric field trace from CoREAS star-shape simulation fed into NuRadioReco and interpolated to antenna location with the interpolator described in 2.2.

related to data processing. The latter includes the read-in of events and the subsequent pulse-finding, RFI filtering, Galactic calibration and direction fitting of the data.

In addition to reading in data for analysis, we can now also create realistic data from simulations by applying our signal chain to simulated pulses and adding noise. One example of that can be seen in Figure 1, which shows a simulated "measured trace" and the corresponding simulation truth. This will enable us to test our analyses on synthetic data for verification. This full circle, i.e. simulating and reconstructing with the same software had never been available for LOFAR data before.

Not only do we use existing modules, but we ourselves are also contributing to NuRadioReco by adding our own analyses, modules and improvements to code. One example is the module that simulates the radio emission from the Galaxy, to which we added various sky maps that all users can choose from. The sky models had been compiled in [7]. They rely on various reference measurements at different latitudes and frequencies, which makes some sky models more suitable for a given experiment than others. We will make use of this not only when creating synthetic data, but also for the absolute calibration of the antennas [8] which we are currently reviewing in the process of finding a possible cause for the tension between the mass composition reported by LOFAR and AERA [3]. As the energy reconstruction is dependent on the fluence of the signals which in turn is dependent on the absolute antenna calibration, a shift in calibration might cause a shift in reconstructed energy which could affect the compatibility of LOFAR and AERA. In Figure

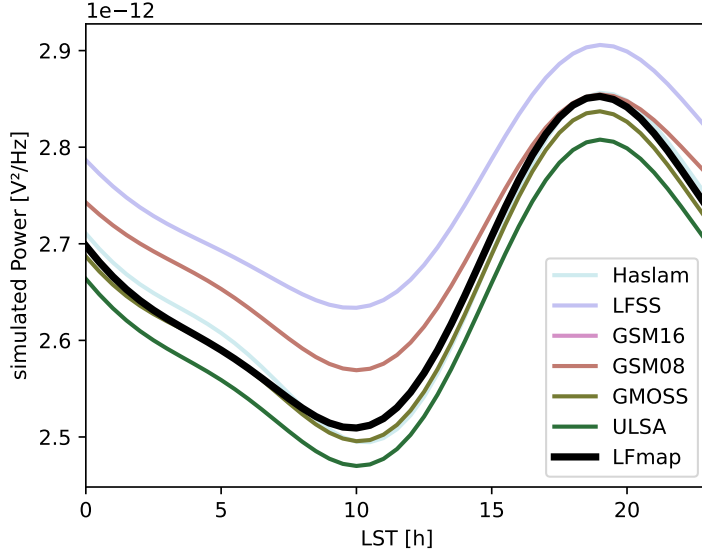


Figure 2: Simulated average integrated power of the Galaxy for one LOFAR antenna over the sidereal day as predicted by different sky models. The sky model used in the previous absolute calibration, LFmap, is marked as a bold black line.

2 we observe that for different sky maps the integrated power differs slightly from the sky map that was originally used in the calibration (LFmap, [9]). Here, the integrated power was calculated for each LST via

$$P = \int_{30\text{MHz}}^{80\text{MHz}} |F(\nu)| d\nu, \quad (1)$$

where $F(\nu)$ is the frequency component of the galaxy noise after propagating through the signal chain. A Fourier series was fit to the power values, as done in the original calibration. It is yet to be determined what impact the choice of sky model has on the calibration, however so far it seems the variation introduced by the different sky models alone is not large enough to account for the needed energy shift. This however is still under active investigation, along with the question of which the best sky model for the calibration would be. In any case, a completely new re-implementation of the calibration will be a good cross-check for potential issues.

Another contribution we made to NuRadioReco was implementing a new CoREAS interpolator, described in the following subsection. This interpolator can in the future be used by all users of NuRadioReco and significantly reduces the computational demand of air shower simulations for large arrays with many antennas while retaining the needed accuracy.

2.2 Interpolation

With growing detector arrays, air shower simulations with software such as CoREAS [10] get more and more computationally intensive. For a LOFAR sized detector, one shower simulation takes up 1 to 3 days of CPU time with a dedicated grid that allows a simple interpolation of the scales of LOFAR. With regard to next generation array such as the Square Kilometre Array (SKA) [11] we are looking at arrays the size of 60,000 antennas. For the LOFAR-method of reconstructing the X_{max} , around 30 simulations are run for each detected shower. In the SKA era, this would become too cost and energy intensive. With the interpolation algorithm we present in [12], it

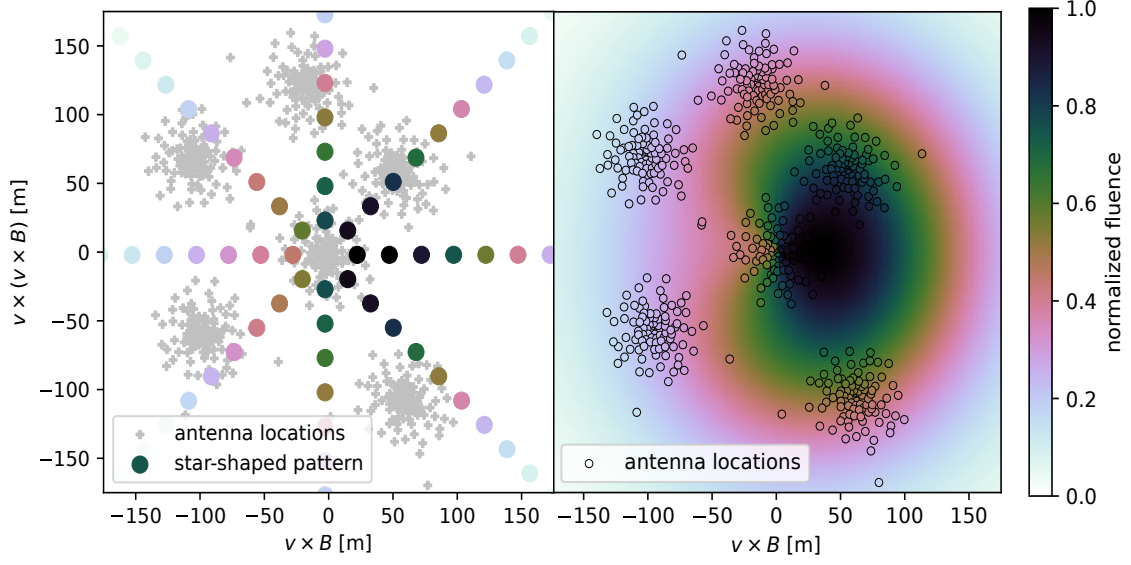


Figure 3: *Left:* the simulated star-shaped pattern, the color indicating the normalized fluence at that position, and the grey crosses in the background representing the LOFAR antenna positions that the simulation will be interpolated to. *Right:* The fluence interpolation shown as gradient in the background and the corresponding fluence values calculated from the interpolated pulses shown for each antenna location. Each antenna dot is filled with the to the fluence corresponding color.

suffices to simulate a shower for approximately 200 antennas while retaining enough accuracy for high-precision analysis. Also, previous LOFAR analyses have only used the signal fluence and its interpolated footprint to determine shower parameters. With the advent of SKA, one would like to use a full waveform interpolation (e.g. for interferometry), which is also part of the implemented interpolator.

The interpolation method is based on Fourier series and cubic splines. After simulating a radio shower with CoREAS for a star shaped antenna array, we use an FFT along a circle at each radius r from the shower core, to obtain Fourier series components for the fluence:

$$\hat{f}(r, \theta) = \sum_{k=0}^{n/2} c_k(r) \cos(k\theta) + s_k(r) \sin(k\theta) \quad (2)$$

These coefficients are then radially interpolated using cubic splines, allowing to get the fluence at any desired antenna position. The time series is modeled as an amplitude spectrum with a phase:

$$F(v) = \mathcal{F}(E(t)) = |F(v)| \exp(i\phi(v)), \quad (3)$$

and by applying the same interpolation method to the spectral amplitudes and phase factors, the full pulse time series can be obtained at any position.

Compared to the previous method of interpolation based on radial basis functions, with this method the fluence interpolation error is reduced by a factor of above 3 in the LOFAR frequency range. Figure 3 shows an exemplary result of using the interpolator as implemented in NuRadioReco, showing the simulated star-shape antennas, the interpolated LOFAR antennas of six stations and

the full interpolated radio footprint. Additionally, Figure 1 (bottom) shows interpolated pulses and their frequency spectrum for one antenna location. For the 30 to 80 MHz band, the typical error is only 0.3 %, while the timing error is around 0.04 ns.

2.3 Machine learning based reconstruction

With the current standard reconstruction being so heavily based on simulations, we have explored the possibility of using machine learning instead. For that, a model was trained on the already existing 300,000 air shower simulations with a simple noise model added to it.

First, the radio pulse was characterized by features: Peak time, total power (integral over the Hilbert envelope of the time trace), the parameters of a quadratic fit performed on the Amplitude spectrum in the frequency domain, and the atmosphere information of the simulation. The model, a Fourier Convolution Neural Network (FCNN), was then trained on these features to predict targets: X_{\max} , L , R , E and the core position. R and L are quantities that relate to the rise and fall of the lateral shower distribution, they can be introduced in a reparametrisation of the better known Gaisser-Hillas formula, the R, L formula [13]:

$$N_c^{R-L}(X_z) = N_{\max} \times \left(1 - \frac{R}{L}(X_{\max} - X_z)\right)^{R-2} e^{\frac{X_{\max} - X_z}{RL}} \quad (4)$$

The FCNN contains seven hidden layers, four million trainable parameters and was given 3 hours of training time. It yielded a mean absolute error on X_{\max} of 12.2 g cm^{-2} or 2.4 %, which is visualised in Fig. 4. For the energy reconstruction, the MAE of the $\log E$ was 0.0305, the shower core was resolved within $3.3 \pm 2.4 \text{ m}$ while L and R were the least accurate, with L yielding an MAE of 6.31 g cm^{-2} (13 %) and for R the MAE was 0.022 or 11 %. Caution needs to be taken when interpreting the MAE for the primary energy, as the model learned a correlation between atmosphere information and the energy from the simulations it was trained on. This is something to look into in future extensions of this work.

The MAE values suggest that there was not enough information on L and R in the features that are given as inputs to the model. As first attempt on a ML based reconstruction, these results are, however, promising. In the future, this work could be extended by adding realistic noise to the simulation data and investigating the changes in performance. Additionally, the reconstruction of X_{\max} , L and R would benefit strongly from including data from the High Band Antennas (HBA), which currently are challenging to use for air shower detection due to the data being already beamformed upon read-out. This will change in the next phase of the telescope, LOFAR2.0, which is currently at the start of being implemented, as discussed in [14]. The full ML analysis can be found in [15].

2.4 Holistic Air shower reconstruction

The ideal reconstruction algorithm would be a physics-informed one that does not rely on unique simulations to compare to for each event. Additionally, the classical methods of reconstruction (e.g. the LOFAR-method) do not use all information that is available in our data – for the X_{\max} reconstruction, only the footprint shape is used, while for the energy reconstruction [16] the fluence is used. Neither method uses the pulse shape or the wavefront timing.

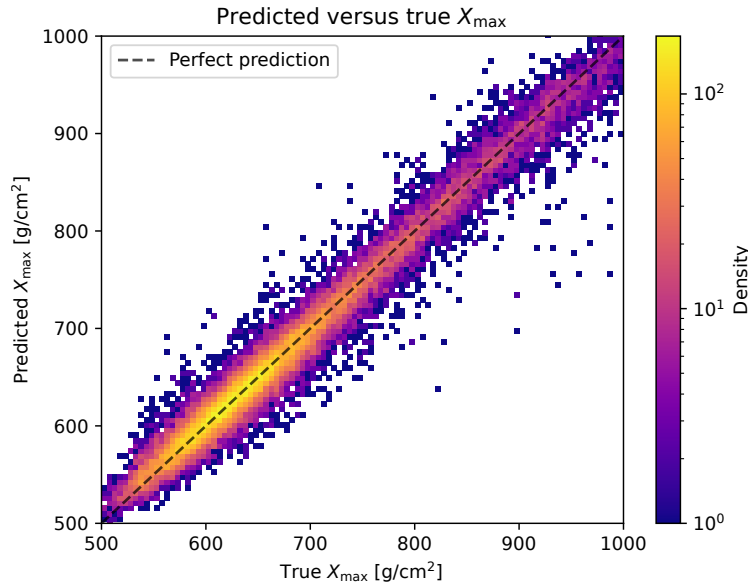


Figure 4: Reconstruction results for the X_{\max} value, showing the predicted values from the model versus the MC truth for each simulation. For most simulated cosmic rays, the values align nicely around the perfect prediction axis.

We have started efforts towards a holistic reconstruction that takes in all available information from the data: timing, polarization, signal strength, footprint, and also in addition the data available from the LORA particle detectors. This poses a complex challenge of ensuring the model is physics informed, i.e. finding a parametrization that relates measured quantities to physical ones such as X_{\max} , L , R or the energy, but leaving enough room for deviations from the parametrization due to unknowns in the data and imperfect parametrizations. It also requires a modeling of the noise. A method that has been well tested in the field of astronomy is a promising candidate for these prerequisites: Information Field Theory [17] can be used to reconstruct signals from noisy data given prior information. It has in the past already been successfully used for reconstructing radio pulses for the use cases of both in-ice detection and air showers [18], which we are currently working on applying to LOFAR as a first reconstruction step.

A further advantage of this method will be that it could be used on all candidate events, even those with low signal-to-noise ratio (SNR), while simultaneously providing reliable uncertainties on all reconstructed parameters.

3. Outlook

After more than a decade of continuous observation, the first phase of the Low Frequency Array, LOFAR1.0 is coming to an end, and observation with the core stations is temporarily interrupted while the hardware upgrades for the next phase, LOFAR2.0, are installed [14]. During this downtime, we will (re-)analyse the full set of LOFAR1.0 data and expect to subsequently release the results. The work that has been done on software in the previous year serves as the ideal basis for the upcoming analyses. This will also be the opportunity to test new reconstruction approaches along with the standard method, which will all serve as preparation for following observations with LOFAR2.0 and also SKA-low.

References

- [1] M. P. van Haarlem et al., *Astronomy and Astrophysics* 556 (2013) A2
- [2] S. Thoudam et al., *Nucl. Instrum. Meth. A* 767 (2014) 339–346
- [3] A. Corstanje et al., *Phys. Rev. D* 103, 102006 (2006)
- [4] A. Abdul Halim et al., *Phys. Rev. D* 109 (2024) 022002
- [5] C. Glaser, A. Nelles, I. Plaisier, C. Welling et al., *Eur. Phys. J. C* 79, 464 (2019)
- [6] P. Schellart et al. *Astronomy and Astrophysics* 560 (2013) A98
- [7] M. Büsken et al., *Astronomy and Astrophysics* 679 (2023) A50
- [8] K. Mulrey et al., *Astroparticle Physics* 111 (2019) 1-11
- [9] E. Polisensky, *LFmap: A low frequency sky map generating program* (2007)
- [10] T. Huege, M. Ludwig, C. W. James, *AIP Conf. Proc.* 1535, 128–132 (2013)
- [11] Gie Han Tan, T. J. Cornwell, P. E. Dewdney and M. Waterson, *URSI Atlantic Radio Science Conference (URSI AT-RASC)* (2015)
- [12] A. Corstanje et al., *Journal of Instrumentation* 18 P09005 (2023)
- [13] S. Andringa, R. Conceição, M. Pimenta. *Astroparticle Physics*, 34, 6 (2010) pp.360
- [14] K. Mulrey et al., *PoS (ARENA2024)*
- [15] L. van Zuijlen, Master Thesis (2024)
- [16] K. Mulrey et al., *Journal of Cosmology and Astroparticle Physics* 11 (2020) 017
- [17] T. A. Enßlin, *ANNALEN DER PHYSIK* 2019, 531, 1800127
- [18] C. Welling et al. *Journal of Cosmology and Astroparticle Physics* 04 (2021) 071