A higher-order interpolation method for simulated radio signals of air showers

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

^{a}*Vrije Universiteit Brussel, Astrophysical Institute, Pleinlaan 2, 1050 Brussels, Belgium*

^{b}*Department of Astrophysics/IMAPP, Radboud University Nijmegen, P.O. Box 9010, 6500 GL Nijmegen, The Netherlands*

^{c}*Department of Physics, Khalifa University, P.O. Box 127788, Abu Dhabi, United Arab Emirates*

^{d}*Netherlands Institute for Radio Astronomy (ASTRON), Postbus 2, 7990 AA Dwingeloo, The Netherlands*

^{e}*Nikhef, Science Park Amsterdam, 1098 XG Amsterdam, The Netherlands*

^{f}*Institut für Astroteilchenphysik, Karlsruhe Institute of Technology (KIT), P.O. Box 3640, 76021 Karlsruhe, Germany*

^{g}*Deutsches Elektronen-Synchrotron DESY, Platanenallee 6, 15738 Zeuthen, Germany*

^{h}*ECAP, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91058 Erlangen, Germany*

^{i}*Interuniversity Institute for High-Energy, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium*

^{j}*University of Groningen, Kapteyn Astronomical Institute, Groningen, 9747 AD, Netherlands*

^{k}*Department of Physics, School of Education, Can Tho University Campus II, 3/2 Street, Ninh Kieu District, Can Tho City, Vietnam*

E-mail: a.corstanje@astro.ru.nl

For high-precision cosmic-ray measurements using the radio emission of air showers such as done with LOFAR and envisioned for the Square Kilometre Array (SKA), simulating ensembles of air showers takes considerable CPU time. In particular, simulating every antenna position would become intractable for the 60,000 antennas at SKA. We present a higher-order algorithm for interpolating radio signals of air showers simulated on a radial (polar) grid. It allows to interpolate the full pulse signal in each antenna/polarization along any position in the footprint. It provides significant improvement over existing, linear methods. From a set of simulated showers analyzed from 30 to 500 MHz, we show that simulating about 200 antennas is sufficient to represent signals at any relevant position to high accuracy. Thus, by reducing simulation time for SKA by a factor ~ 300, it offers a relevant extension to radio simulation codes.

---

*Speaker

1.　The challenges of interpolating radio air shower footprints

The radio air shower community has been successfully studying air showers since the arrival of highly accurate simulations, such as CoREAS [1] and ZHAireS [2], and the operation of modern instruments such as LOFAR [3, 4] or AERA at the Pierre Auger Observatory [5].

While highly accurate, simulating the radio emission of air showers is computationally expensive. Simulating a single air shower in an instrument with 150 to 200 antennas takes on the order of 1 to 2 days in a single CPU thread, and even longer at the highest energies. In particular, the simulation effort scales almost linearly with the number of simulated antennas.

The community uses several parameterizations, based on simulations, to reconstruct the energy and the height of shower maximum of the shower (e.g. [6, 7]). However, it has been found that the highest accuracy in reconstruction can, at this point, only be used by a direct comparison of dedicated simulations (same arrival direction, different heights of shower maximum) and data. At LOFAR this requires, for example, at least 25 simulations per detected high-quality air shower [8]. Since the radio footprint is not axial symmetric around the shower axis and furthermore also strongly and non-monotonously varying as function of distance to the shower axis, one would also have to simulate different air showers per axis position on the ground, if one were to use the real detector lay-out. In order to avoid this, the radio community has converged on simulating so-called 'star-shaped' patterns, i.e. aligning 150 to 200 fictive detectors along the coordinate system that is spanned by the shower vector $\vec{v}$ and the magnetic field $\vec{B}$, with axes being $\vec{v} \times \vec{B}$ and $\vec{v} \times (\vec{v} \times \vec{B})$, and lines diagonal to this. From this star-shaped pattern, quantities were derived to an accuracy sufficient to not limit accuracy of the air shower reconstruction. Higher-level analyses up to this point only use the energy fluence in simulations and the detected pulses. The energy fluence is typically defined as integrated quantity over time and polarizations, which makes it a very robust quantity when it comes to the detailed instrument response and its uncertainties. Many current observatories work at low frequencies, 30 to 80 MHz; in this range, the energy fluence varies relatively slowly along the footprint. Interpolation methods provided in Scipy (based on radial basis functions [9]) using the star-shaped pattern have proven sufficient for this case.

At higher frequencies, e.g. up to 350 or 500 MHz this method was found to be insufficiently accurate. Thus, a dedicated interpolation method is called for, that makes use of known symmetries in the air shower footprints. This is even more urgent for analyses that attempt to use the individual signal polarizations, pulse shapes, and frequency content in a much wider frequency band as it will be the case for air shower detection with the Square Kilometre Array (SKA) [10]. Especially the SKA poses a challenge in accuracy that is not met by any other experiment, since a single air shower will be detected with up to 60,000 antennas. This means measurements of air showers can be done with unprecedented precision, yet simulating so many antennas directly is intractable with current 'microscopic' simulation codes.

2.　Novel approach to interpolation

We briefly describe our interpolation toolkit and its approach here. For more in-depth explanation we refer to the public code base[1] and to the recently published article accompanying the code

---

[1] https://github.com/nu-radio/cr-pulse-interpolator

[11].

All interpolations are done in the shower plane (perpendicular to the shower axis). Interpolations assume an existing radial grid, also known as star-shaped pattern. The performance was tested on random positions on the shower plane.

2.1 Interpolation of the energy fluence

To replace currently used methods, we first have developed an algorithm to interpolate the energy fluence, or other single-valued quantities that vary along the footprint. We make use of the geometry of the dominant contributions to the radio signal, being the geomagnetic and charge excess components. At a fixed radius, the amplitudes vary with angular position θ around the shower axis, expressing positions in the (2D) shower plane in polar coordinates (r, θ). Variations are proportional to $\cos(\theta)$ for the charge excess component, while the geomagnetic part is circular symmetric. A natural choice is to use a Fourier series along the angular direction, i.e. circles at constant distance from the axis, to describe variations. The Fourier series is a complete representation of a periodic signal, like the variation along a circle, provided the sampling is uniform and dense enough to describe the highest (spatial) frequencies that occur, i.e. the Nyquist sampling criterion. Taking the 8 antenna positions along a circle, we apply a Fast Fourier Transform (FFT) to obtain the Fourier components. At an arbitrary position on the circle, evaluating a Fourier sum using these components yields the interpolated value:

$$\hat{f}(r, \theta) = \sum_{k=0}^{n/2} c_k(r) \cos(k\,\theta) + s_k(r) \sin(k\,\theta), \tag{1}$$

where $\hat{f}(r, \theta)$ is the interpolated value at position (r, θ), and c_k and s_k are the Fourier coefficients expressed in cosine and sine modes enumerated by k. For the radial variations, these Fourier components are interpolated (radially) in between antenna positions using cubic splines, obtaining $c_k(r)$ and $s_k(r)$ in the equation.

2.2 Interpolation of the time-dependent electric field

For the full pulse interpolation, the electric field $E(t; x, y)$ is taken to the frequency domain using an FFT, giving an amplitude and a phase spectrum at each antenna position, and each needs to be interpolated.

At each position (x_i, y_i) one can separate (absolute) amplitude and phase spectrum as:

$$F(\nu) = \mathcal{F}(E(t)) = |F(\nu)| \exp\left(i\,\phi(\nu)\right), \tag{2}$$

For every frequency ν the value of the absolute amplitude spectrum can be interpolated like the energy fluence as described above.

The phase spectrum is more challenging, as it consists of phase factors $\exp(i\,\phi)$ which would identify ϕ with $\phi + 2\,\pi\,k$ for integer k. This poses difficulties for any straightforward interpolation method, as it can be unclear in which period-window the intermediate values should lie. In particular in the presence of noise, the choice becomes ambiguous, which leads to artifacts in the interpolation.

We have identified two algorithms, which have been found to perform similarly well, however, performing slightly different for different corner cases. Both methods have been implemented and are interchangeable for the user [12].

First, the generic pulse arrival time in all signals is determined, as well as the phase constant ('Hilbert phase'), after which one is left with the residual phase spectrum which needs to be interpolated.

One method estimates phases using pulse timing in a sliding frequency window, thus trying to break the 2π periodicity. It identifies the frequency up to which the signal still shows coherence and stops interpolating beyond this frequency. In particular for observers far from the Cherenkov angle, high frequencies are strongly suppressed and thus cannot be used for a signal interpolation, while simulation codes may still yield non-zero predictions due to artifacts in the simulations (see Fig. 4).

The other method expresses the phases (after subtracting the first-order contribution) as complex phasors $\exp(i\,\phi)$ and interpolates their real and imaginary part.

All individual interpolations, like the amplitude in each frequency channel, are performed using the same interpolation ansatz as for the energy fluence.

3. Performance of the interpolation

We show the performance for both the fluence interpolation, as well as the pulse shape interpolation. We note that the method performs even better for the restricted frequency range from 30 - 80 MHz, which is being used by most existing air shower arrays. The detailed results are shown in [11].

All results are shown for 250 random locations in the shower plane, using interpolations based on a star-shaped pattern of 208 antennas. Showers are simulated with CoREAS for 460 m above sea level, and a magnetic field vector of 55.6 μT at an inclination of 60.24 °, which corresponds to the SKA-Low site in Australia. The set has discrete arrival directions at zenith angles of 15°, 30°, 45°, and 80°, and azimuth angles of 0°, 15°, 30°, 45°, 90°, 135°, and 180°, as measured clockwise from North for the incoming direction. For each direction, 3 showers were simulated to avoid a bias for outlying heights of shower maximum. Each shower has an energy of 10^{17} eV with protons as primary particle.

In Fig. 1, we show two interpolated pulses as an example, comparing them to the 'true' simulated pulses. This demonstrates that for typical pulses, the residual (difference) is near zero; differences due to small mismatches in timing are included. As shown here, pulses with significant structure in the time domain are well reconstructed. This structure arises not only from the amplitude spectrum shown in the right-hand panels, but also from the phase spectrum which is therefore important to approximate accurately.

Fig. 2 shows the results for the energy fluence. The relative error both on energy fluence and on signal amplitude is better than 1% with only a few outliers with more than 5%. It should be noted that the interpolation performs worse for inclined showers using the current antenna layout. This is most likely for geometric reasons as the inclined showers at 80° zenith angle show a thin but pronounced Cherenkov ring near 750 m from the core, where the existing star-shaped patterns were not sampled finely enough. This illustrates the need of a bit of fine-tuning for the simulations
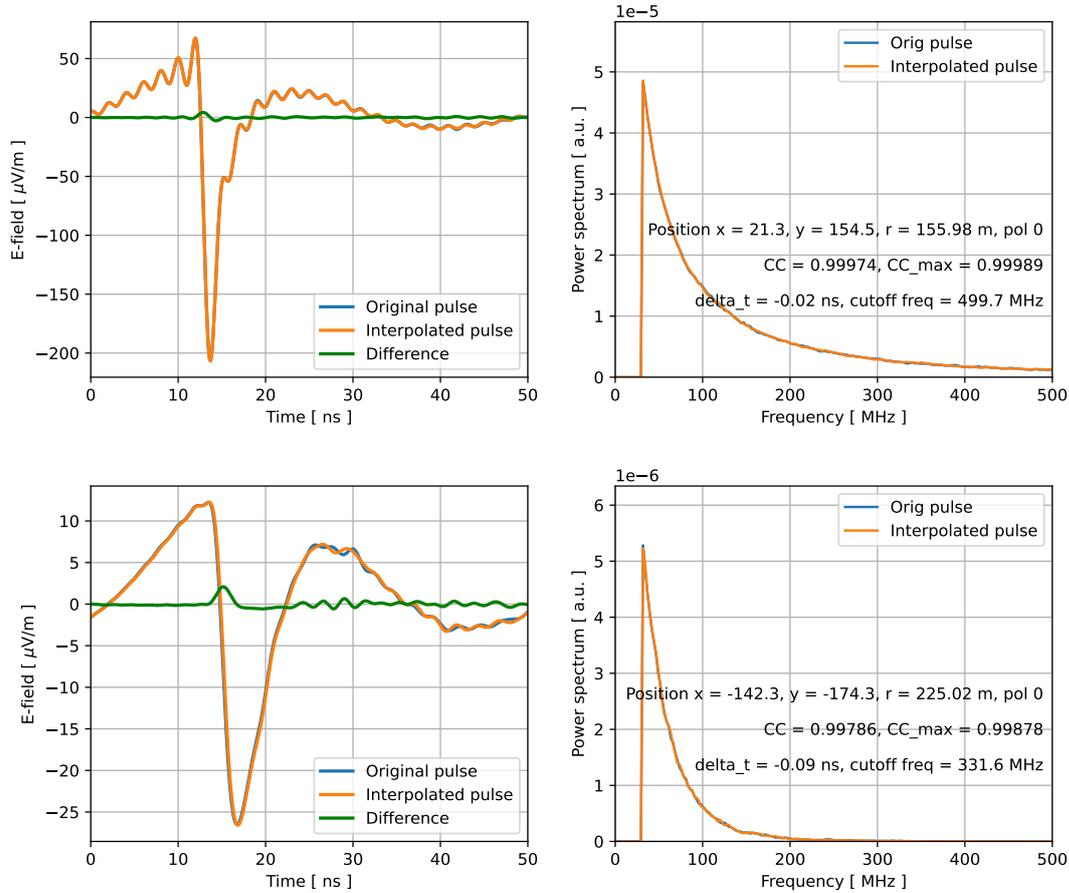
Figure 1: Two examples of an interpolated pulse, compared to the real simulated pulse at the same location. The top plots show the E-field and the spectrum (on a linear scale) for a relatively strong pulse. The residual is nearly zero. The bottom plots show a weaker pulse, further out from the core. The residual signal arises mostly from a small timing mismatch of 0.09 ns, still smaller than the time resolution in the simulation. Frequency range is 30 to 500 MHz.

for every particular experiment and use case. Also, relative deviations are stronger at small signal strengths as the signal quality degrades there, which will be further discussed when detailing the performance of the full pulse interpolation below.

Fig. 3 shows the performance of the full electric field interpolation. We use a normalized cross-correlation as performance indicator, which is a number between -1 and 1, with 1 denoting perfect agreement. For the vast majority of moderate or strong signals, the interpolation is better than 0.99 and averaging to 0.998, indicating near-perfect reconstruction. There are some outliers, at a level of 1 to 2 per 1000 tested pulses. Similarly, the pulse timing error, here defined as the shift required to maximize the cross-correlation, is smaller than the sampling resolution of 0.1 ns used in the simulations, for all reasonably strong signals. This is sufficient for interferometric reconstructions. It can be seen that, again, the performance is worse for low-amplitude signals. This has been found not to be a limitation of the method, but of the CoREAS simulations themselves.

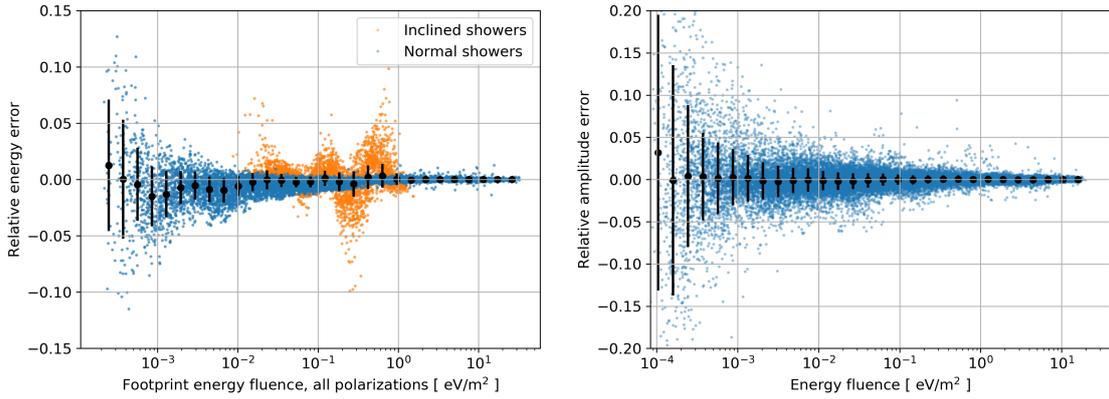As shown in Fig. 4 the performance is a clear function of the thinning level applied in the

Figure 2: The relative interpolation error in the total signal energy fluence is shown on the left. Each dot represents a test antenna on a random position in the shower plane. Black dots show the binned mean, and their error bars show the standard deviation. On the right, the same is shown for the relative interpolation error in the pulse amplitude, per polarization. Results are shown for 30 to 500 MHz.
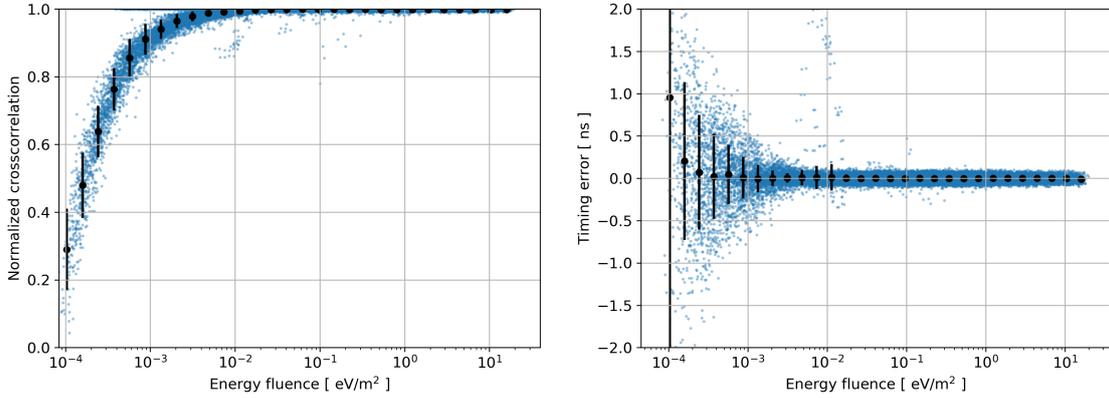

Figure 3: Left: Cross-correlation of interpolated and true pulse at the random test positions in the footprint, for all showers with geomagnetic angle above 10 degrees. Frequency range is 30 to 500 MHz. Right: Timing accuracy as a function of energy fluence, 30 to 500 MHz.

simulations. CORSIKA employs thinning to be computationally efficient. In this thinning particles are grouped, which introduces artificial coherence for the radio simulations and leads to what is typically referred to as *thinning noise*. This is most clearly seen as additional noise contribution that dominates at high(er) frequencies, where the true smooth air shower signal drops off. The noise also scales with shower energy.

The thinning level indicates how many particles can be grouped together. A low energy threshold, below which thinning becomes active, means that fewer particles are grouped. This makes the simulations significantly slower, but also more accurate. This challenge is in principle unrelated to the pulse interpolation introduced here, however, became prominent when testing for accuracy. It is noteworthy that the energy fluence level where signals degrade due to thinning is expected to shift roughly proportionally to the primary particle energy. Therefore, this should be taken into account when simulating air showers from very high-energy particles (say 10^{19} eV).
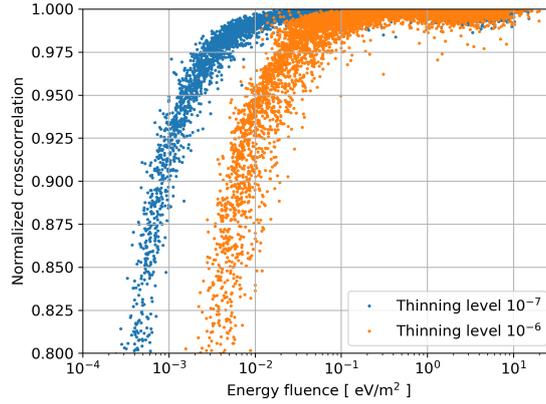
Figure 4: The effect of the thinning level on the cross-correlation between interpolated and simulated pulses. Frequency range is 30 to 500 MHz. Unsurprisingly, the accuracy of the pulse interpolation depends on the quality of the simulations that are used to interpolate the signals.

The interpolation breaks down in the presence of thinning noise, so it cannot be used for weak signals. The code establishes a cutoff frequency which provides an indication of whether one tries to interpolate noise-dominated signals and can be used as a cross-check for the frequency band in which the signal is reliable.

4. Conclusions

We have presented a pulse interpolation approach that is suitable for radio air shower simulations as required for the Square Kilometre Array. By using this interpolation, we will save a factor of about 300 in computing time to study air showers with SKA without significant concessions to accuracy, thus overcoming an important practical obstacle. This interpolation can be considered enabling for SKA air shower analyses, but would also be useful for other air shower experiments, in particular those using the full pulse information in their analysis, for example by using interferometry. The code is publicly available at [12].

References

[1] T. Huege, M. Ludwig, and C. W. James *ARENA 2012, AIP Conf. Proc.* **1535** no. 1, (2013) 128.

[2] J. Alvarez-Muniz, W. R. Carvalho, Jr., and E. Zas *Astropart. Phys.* **35** (2012) 325–341.

[3] P. Schellart, A. Nelles, *et al. Astronomy & Astrophysics* **560** (Dec., 2013) A98.

[4] K. Mulrey for the LOFAR Collaboration *PoS(ICRC2023)443* .

[5] R. de Almeida for the Pierre Auger Collaboration *PoS(ICRC2023)279* .

[6] A. Nelles *et al. JCAP* **05** (2015) 018.

[7] C. Glaser, M. Erdmann, J. R. Hörandel, T. Huege, and J. Schulz *JCAP* **09** (2016) 024.

[8] S. Buitink *et al. Phys. Rev. D* **90** (Oct, 2014) 082003.

[9] P. Virtanen *et al. Nature Methods* **17** (2020) 261–272.

[10] S. Buitink for the SKA High Energy Particle Science Working Group *PoS(ICRC2023)503* .

[11] A. Corstanje *et al. Submitted to JINST* (6, 2023) .

[12] A. Corstanje *et al.* `https://github.com/nu-radio/cr-pulse-interpolator` .