Synthesis of radio signals from extensive air showers using previously computed microscopic simulations

Matias Tueros[1,2,3] **Anne Zilles**[3]

[1]*IFLP - CCT La Plata - CONICET, Casilla de Correo 727 (1900) La Plata, Argentina*

[2]*Depto. de Fisica, Fac. de Cs. Ex., Universidad Nacional de La Plata, Casilla de Coreo 67 (1900) La Plata, Argentina*

[3]*Institut d'Astrophysique de Paris, CNRS-Sorbonne Université, 98 bis boulevard Arago, F-75014 Paris, France*

E-mail: tueros@fisica.unlp.edu.ar

ABSTRACT: The detection of extensive air showers (EAS) through their radio signal is becoming one of the most promising techniques for the study of Neutrinos and Cosmic rays at the highest energies. For the design, optimization and characterization of radio arrays, and of their associated reconstruction algorithms, tens of thousands of Monte Carlo simulations are needed. Current available simulation codes can take several days to compute the signals produced by a single shower, making it impossible to produce the required simulations in a reasonable amount of time, in a cost-effective and environmental-conscious way. In this article we present a method to synthesize the expected signals (the full time trace, not just the peak amplitude) at any point around the shower core, given a set of signals simulated in a finite number of antennas strategically located in a pattern that exploits the signature features of the radio wavefront. The method can be applied indistinctly to the electric field or to the antenna response to the electric field, in the three polarization directions. The synthesized signal can be used to evaluate trigger conditions, compute the fluence or reconstruct the shower incoming direction, allowing for the production of one single library of simulations that can be used and re-used for the characterization and optimization of radio arrays and their associated reconstruction methods, for a thousandth part of the otherwise required CPU time.

arXiv:2008.06454v1 [astro-ph.IM] 14 Aug 2020

Contents

1 Introduction

When a high-energy cosmic particle interacts with the molecules in the atmosphere, an extensive shower of secondary particles is generated. Moving close to the speed of light, the charged particles in the shower are deflected by the Earth's magnetic field, collectively producing coherent electromagnetic pulses that can be detected at large distances from the shower in the VHF (30-300 MHz) frequency range with a relatively simple and cheap radio detector. This makes the radio detection of cosmic particles a very promising technique for instrumenting the large surface areas required by the next generation of cosmic rays and neutrino detectors [1, 2].

To assess the capabilities of a given detector array, the signals expected to be produced by the particle showers needs to be computed covering all the possible primary particle energies, the possible incoming directions and the possible particle types (from now on, the incoming particle phase-space) with enough statistics to account for the variability from shower-to-shower fluctuations, requiring thus the simulation of several tens of thousands of particle showers.

Furthermore, the optimization of the detection capabilities of a given array can require changing the separation between detectors, the array total size and general layout, the possibility to include

in-fills or voids in the array, the location of the detectors on the terrain, etc. A change in any of these parameters (from now on, the detector phase-space) will impact the performance of reconstruction algorithms, a factor that also needs to be studied and taken into account to make decisions on the final detector design. Covering the detector phase-space requires re-evaluating the signals produced by the showers in each different configuration.

This problem is common to all EAS detection experiments, regardless of their detection technique. However, the simulation of radio signals presents an additional challenge. In particle detectors for example, the generated signal depends only on the particles arriving at the detector. This means that if the type, position, energy and arrival direction of the particles reaching the surface are stored, the output of the shower simulation can then be re-used to compute the signal at any position, in any particle detector.

For radio detectors, the electric field produced at a given location is the sum of the contributions of *all* the charged particles in the shower, and depends on the relative position of the detector to the trajectories of the particles. This means that to separate the simulation of the shower from the simulation of the signals it produces, the trajectories of all the charged particles needs to be stored. To preserve the coherence information of the radio emission, the trajectories need to be detailed to better than the characteristic wavelength, and their timing to better than fraction of their period. Even using statistical thinning methods to keep the number of simulated particles at bay, this means tracking charge, position, time and energy for typically 10^6 simultaneous trajectories along 10^5 m. This would require, for a single shower, storage space several TBytes, which is at the limit of current technology and totally impractical when libraries of several tens of thousand of showers are required.

One way to circumvent this problem is to use a "macroscopic" approach. In this approach, the spatial and temporal distributions of the charged particles in the shower, or a parameterization of these, is used to solve the electro-dynamical equations and compute the resulting electric field at each detector location. This removes the need for storage of all the particle trajectories, replacing them by histograms or a relatively small set of parameterizations that tries to describe the overall characteristics of the particle cascade [3–6].

This simplification makes macroscopic approaches very fast, and very useful to gain insights in the processes responsible for the features of the radio signal. However, the assumptions used to simplify the problem and represent the individual particles with charge distributions erase the fine temporal and spatial structures in the signal, and require the fine-tuning of free parameters that affect the results and limit their predictive power [3].

If the limitations of macroscopic approaches are to be avoided, and the storage of the particle trajectories is impractical, the simulation of the radio signals needs to be done in parallel with the simulation of the particle shower. Consequently, each time the location of the detectors is changed, the simulation needs to be repeated.

The simulation of the electric field following the contribution of each individual particle (the "microscopic" approach [7, 8]) is CPU intensive. The computation of the electric field produced at a single position takes nearly as much CPU time as the computation of the whole particle shower itself. Since the computations at each antenna position are almost independent of each other, the required CPU time scales with the number of desired positions. The simulation of a single event involving a few dozens of antennas can easily take one CPU-day. Repeating the simulation of

thousands of events to cover the detector phase-space can easily require millennia of CPU time. While achievable in large supercomputing centers, when available, this is costly and requires lots of CO_2-emitting electric power and hardware. Looking for a way to re-utilize microscopic simulations, while keeping the accuracy of its results, is thus paramount.

Attempts at re-utilizing the results from microscopic simulations to synthesize the radio signal already exist, in different stages of development. One method [9] computes separately the signals at a given detector position originating from different slices of the longitudinal development of the shower, so that the total signal is the sum of the signals from all the slices. Next, the spectral features of the signal in each slice are fitted to produce a signal model that can be applied to synthesize the signals of a shower with a different longitudinal profile but with the same geometry, at the same antenna position. While promising, this method is still not suited for its general use, as the fit of the model parameters require human supervision on a case by case basis. If the fit procedure were to be automated, and the method proves to produce very accurate synthesized signals, an interpolation strategy similar to the one proposed later in this article could be used to get the signal at different antenna positions.

Another attempt at re-utilizing the results from microscopic simulations is the method known as radio-morphing [10]. This method tries not only to re-utilize the simulations, but also to drastically reduce the number of simulations needed to cover the incoming primary phase-space by re-scaling or "morphing" the signals simulated on a reference shower to the signals expected for a different shower. The morphing is achieved by using the known variation of the signal amplitude with primary energy, geomagnetic angle and air density, as well as taking into account the variation in the angular opening of the Cherenkov cone. Simulating the signals in a wisely chosen set of positions, the signal at any point in space can then be synthesized by interpolation, allowing the re-use of the reference simulation as needed. The method was successfully used for the synthesis of signals from showers initiated by neutrinos at the GRAND array [11], with a mean bias in the estimation of the peak amplitude of the signals of less than 10% and a standard deviation of 25% (albeit some significant outliers), which is impressive considering that a single shower was used to cover all the incoming phase-space.

Radio-morphing is bound to provide better results if more showers are used as reference, reducing the amount of "morphing" that needs to be done for going from the reference shower to a desired shower. In the limit, the error from "morphing" the shower can be completely removed by simulating a reference shower for every desired incoming configuration. This would negate completely the method's capability to reduce the number of simulations required to cover the incoming primary phase-space, but would keep the ability to re-use the simulations to synthesize the signals required to study the detector phase-space with improved accuracy.

The method presented in this paper is an implementation of this extreme approach, optimized for the synthesis of signals on the ground by a more efficient choice of simulated antenna positions and an improved phase interpolation algorithm, with the important addition of an accurate prediction of the synthesized signal arrival time, which was unavailable in the radio-morphing method and is crucial for the study of the performance of geometry reconstruction algorithms.

In this article we will present the basic aspects of radio emission from extensive air showers and their role in the motivation of the developed interpolation strategy, along with the interpolation procedure, in section 2. A general description of the performance of the method using the cross-

correlation of the synthesized and simulated signals will be presented in section 3. The obtained precision on several signal observables relevant to triggering and reconstruction algorithms, both on a flat surface and for complex topographies, will be shown in section 4. We finish by giving some concluding remarks in the last section.

2 Underlying model and strategy for the interpolation

The radio emission of particle showers can be described as a superposition of two distinct emerging phenomena. The most important one, accounting for 80 - 95 % of the emission, is the appearance of a charge dipole that changes with time as the positive and negative charges in the shower are deflected in opposite directions by the Earth's magnetic field $\vec{B}$, the so-called Geomagnetic effect [1, 15].

The second effect, known as the Askaryan effect [13, 14], is the build-up of a net negative charge in the shower front as electrons ripped-off from the molecules in the air are incorporated into the shower front, while positrons annihilate with electrons in the atmosphere.

The Geomagnetic emission is linearly polarized in the direction perpendicular to the direction of $\vec{B}$ and the shower direction $\vec{v}$. Since the dipole increases as more and more charge is separated and then decreases as the particles loose their energy and are removed from the shower, the resulting emission is bipolar. The Askaryan emission, which is akin to a current pulse on the shower axis, is also linearly polarized but in this case the polarization vectors are radially oriented perpendicular to the shower direction.

The emission region, the shower front, is only some meters thick and produces coherent radiation up to several MHz. The particles in the shower front travel almost at the speed of light in vacuum c, giving rise to a time compression of the received signal in the direction of the Cherenkov angle α_c (roughly 1 degree in air), pushing the signal bandwidth up to the GHz range around that particular direction [16–18].

The emission pattern of particle showers can be modeled, as a first approximation, as the emission of a point like source moving at c along the direction of $\vec{v}$, with an amplitude that is proportional to the number of particles in the shower [12]. This simple model implies that the bulk of the emission originates close to X_{max}, the position where the shower reaches its maximum development.

Considering a constant index of refraction n, and ignoring the contribution from the Askaryan effect, the amplitude of the radiation pattern in a plane perpendicular to $\vec{v}$ will be rotationally symmetric around the shower core. The radiation will be polarized in the direction of $\vec{v} \times \vec{B}$, and propagating as a spherical wavefront at the speed of light in the medium, $v = c/n$.

In this simplified model, due to the rotational symmetry, the signal at any point in a plane perpendicular to the propagation direction of the shower can be described solely as a function of α, the angle between the point and the shower axis, measured from X_{max} (Fig. 1). We will express α in units of the Cherenkov angle so that $\alpha_c = 1$, regardless of the value of n.

In a plane oblique to the propagation direction of the shower, such as the ground plane, the rotational symmetry is broken due to the difference in distance to the emission point for points at the same α. Consequently, to describe the signal in the ground plane its necessary to introduce a second coordinate, the polar angle ϕ, whose origin is chosen along the projection of $\vec{v} \times (\vec{v} \times \vec{B})$ on

the ground. The amplitude pattern varies more abruptly in α, in particular close to α_c, while the variations in ϕ are smoother.

We have ignored so far the Askarian emission, as it contributes on average with less than 10% [19] of the total emission. The amplitude of the Askarian emission is radially symmetric, but its polarization is pointing towards the shower axis, and will add constructively to the Geomagnetic emission for some values of ϕ and destructively for others, further disrupting the rotational symmetry.

2.1 Interpolation Strategy

The interpolation strategy that we will use to synthesize the signal $S_i(\alpha_i, \phi_i)$ at a point on the ground plane with coordinates (α_i, ϕ_i) is to choose four signals, $S_I(\alpha_I, \phi_I)$, $S_{II}(\alpha_{II}, \phi_{II})$, $S_{III}(\alpha_{III}, \phi_{III})$ and $S_{IV}(\alpha_{IV}, \phi_{IV})$, such that

1. $\alpha_I > \alpha_i > \alpha_{II}$

2. $\phi_I = \phi_{II} > \phi_i$

3. $\alpha_{IV} > \alpha_i > \alpha_{III}$

4. $\phi_{IV} = \phi_{III} < \phi_i$

5. $\alpha_I = \alpha_{IV}$

6. $\alpha_{II} = \alpha_{III}$

and interpolate first along the ϕ coordinate between S_I and S_{IV} to get $S_a(\alpha_I, \phi_i)$ and between S_{II} and S_{III} to get $S_b(\alpha_{II}, \phi_i)$, and finally interpolate along the α coordinate between S_a and S_b to get $S_i(\alpha_i, \phi_i)$ (see Fig. 1). When using the interpolation procedure described in 2.2, this is equivalent to a bilinear interpolation [20] of the signals in I,II,III and IV in (α, ϕ).

To be able to synthesize the signal at any desired position on the ground, the signal needs to be simulated in an array of positions in (α, ϕ), covering the full range of ϕ and the relevant range in α up to a given α_{max}. Since the variation of the signal is faster in α, this variable needs to be sampled more densely, specially close to the Cherenkov angle. In this work we use a quadratic spacing in α to sample more densely the region where $\alpha < 1$.

A denser array of sampled positions gives smaller interpolation errors, but requires more CPU time to perform the signal simulations, reducing the gains from synthesizing the signals. A good compromise was found to be the use of 8 equally spaced points to sample ϕ, and 20 points to sample α, as shown in fig. 2.

When using 8 values of ϕ, the resulting pattern on the ground resembles the "star-shape" array frequently used in the characterization of the shower front, but projected on the ground plane along a cone with it's vertex in X_{max}. We will refer to this array as the "conical star-shape", to differentiate it from the more frequently used parallel projection of the star-shape along the shower axis, which does not preserve α.

When index of refraction gradients are introduced in the description, the spherical wavefront is distorted by the differences in optical path to each observer position, and the rotational symmetry
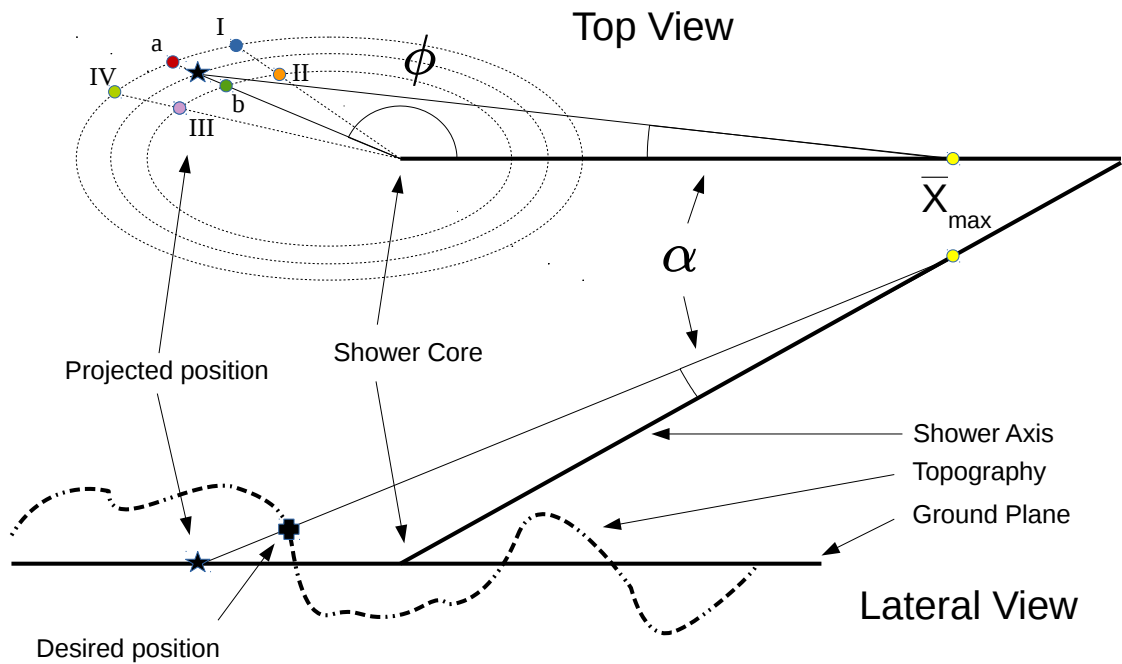
Figure 1. Shower geometry and interpolation strategy. To get the signal at a given position (cross) on the terrain, the signal at the projection of this position into the ground plane (star) is synthesized. This is done using the signals at points I and IV to get the signal at **a** and the signals at II and III to get signal at **b**. The signals at **a** and **b** are then interpolated to get the synthesized signal (star), whose amplitude is then scaled to take into account its true distance to X_{max}, and get the signal at the desired position (cross)
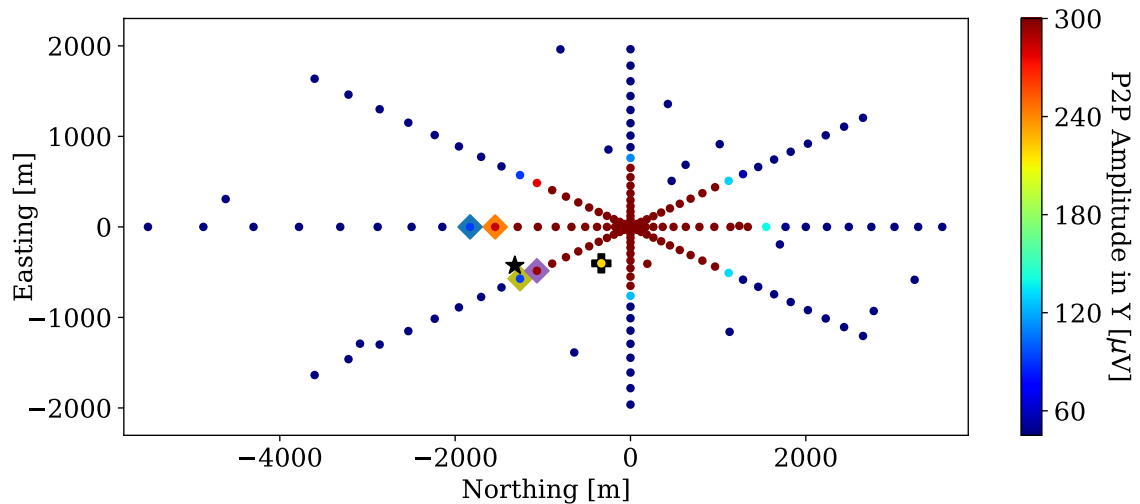

Figure 2. Example conical star-shape plus 16 random test locations from an 4 EeV proton cascade at 63° zenith incoming from the North, plotted on equal axis scales to emphasize the elliptical shape. For the synthesis of the signal at a given position (cross), which in this case is 450m above the ground plane, the projection to the ground plane is computed (star) and the signal is obtained from the signals in the highlighted postions, as shown in fig 4

is further disrupted. The determination of α_c is no longer straightforward, as the optical path depends on ϕ. We use $\alpha_c = \arccos(n_{X_{max}})$, where $n_{X_{max}}$ is the average value of n along the shower axis from X_{max} to the core position on the ground. In the case of an index of refraction varying exponentially with altitude, as is the one used in our simulations, we found that this prescription slightly over-estimates the value of α_c, ensuring that the Cherenkov peak remains inside the better sampled area of $\alpha < 1$, while retaining the scaling benefits of using a unit for α that changes with the position of X_{max}.

Since α_c is slightly different for each value of ϕ, the interpolation using a constant α will be less accurate. This will be one of the sources of systematic errors in the method, and could be reduced by a denser sampling of the ϕ coordinate at the expense of more CPU time, or using the correct value of α_c on each branch of the star-shape, at the expense of more complexity in the method.

The optimal value of α_{max} depends on the objectives of the study to be done with the synthesized signals. Signal amplitudes at large values of α are usually very small and/or out of the bandwidth of the detectors, except for high energy showers with the maximum very close to the ground. A parameterization of α_{max} with the shower energy and the distance to X_{max} is recommended to avoid wasting CPU time sampling regions of the shower front that will later be discarded by quality cuts.

When taking into account the local topography over a given site or when looking for a suitable site for the detector array along a given geographic region, the signals might need to be synthesized at different altitudes above sea level. To accommodate for this, the star-shape pattern is located on a ground plane at a conveniently chosen intermediate altitude (as shown in Fig. 1). When the signal at a location with a different altitude is desired, the signal is synthesized at the projection of the desired position into this ground plane. To compensate for the fact that the synthesized signal is being computed closer or farther away from X_{max} than the desired position, the signal amplitude is scaled with the ratio of the distance to X_{max} of the desired and synthesized positions.

2.2 Signal Interpolation

The interpolation of radio signals relies on the fact that the electric field produced by a particle shower (or the response of an antenna to it) changes smoothly and continuously between nearby positions. We will use an interpolation method in Fourier space based on the one presented in [21] and later used in [10], with a different treatment of the phase interpolation and introducing time alignment of the signals to minimize phase differences and to provide the correct time of arrival.

A signal of length T sampled regularly at intervals T/N_s can be written as a Fourier series

$$S_j(t) = \frac{1}{N_s} \sum_{k=0}^{N_s-1} s_j(k) e^{2i\pi kt/T}, \tag{2.1}$$

with N_s the number of time samples and the complex Fourier components

$$s_j(k) = |s_j(k)| e^{i\theta_j(k)}. \tag{2.2}$$

Given the signal at two reference points S_1 and S_2 the signal at a desired position between them S_i has interpolated Fourier components
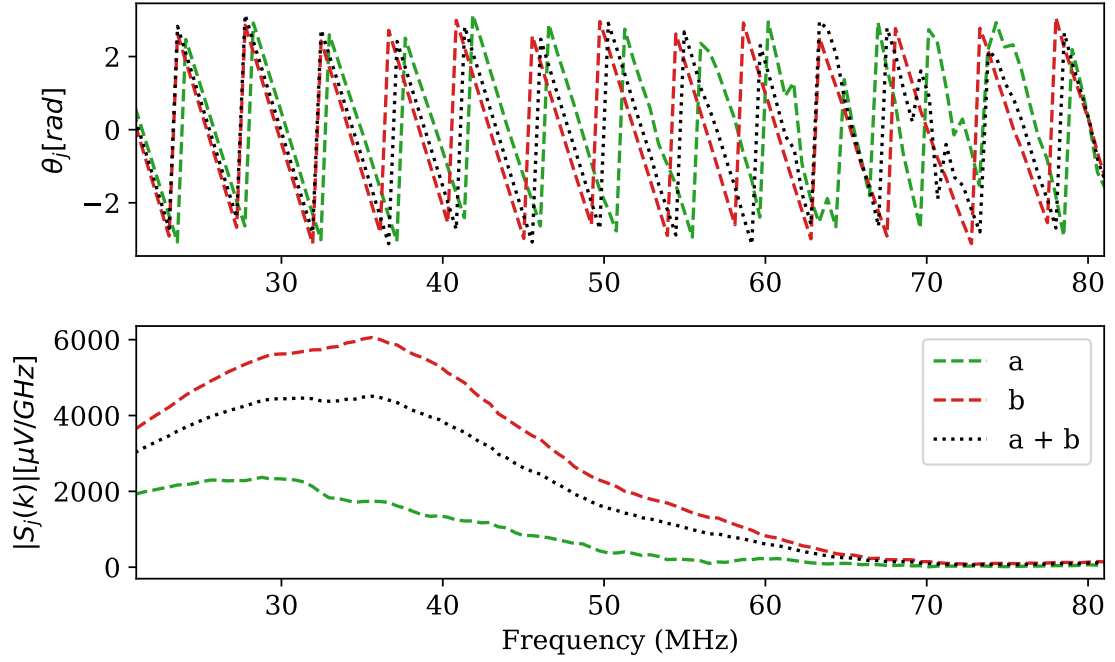
Figure 3. Example of the interpolation in Fourier space of signals in **a** and **b** (see fig. 11), for the selected antenna of fig. 2. The signals in time domain can be seen in the bottom left of fig. 4.

$$|s_i(k)| = \frac{d_1}{d_1 + d_2}|s_1(k)| + \frac{d_2}{d_1 + d_2}|s_2(k)|, \tag{2.3}$$

$$\theta_i(k) = \frac{d_1}{d_1 + d_2}\theta_1(k) + \frac{d_2}{d_1 + d_2}\theta_2(k), \tag{2.4}$$

where d_1 and d_2 are the distances from the desired position to the reference points.

Since θ is cyclic in $[0, 2\pi)$, there is an indetermination in the absolute phase difference of the Fourier components of the signals being interpolated. To reduce the effects of this indetermination, the times of the reference signals are measured relative to the expected time of arrival of the signal at the reference position.

The phase interpolation is performed minimizing the phase difference at each k, adding 2π to θ_1 or θ_2 so that $|\theta_1 - \theta_2| < \pi$, and then warping the resulting phase in the $[0, 2\pi)$ range. An example result from this procedure can be seen in fig. 3 for the for the selected antenna of fig. 2. An example instance of the θ shifting procedure can be seen at 59MHz, where the phase of **b** was shifted by -2π before the interpolation. This prevented an unnecessary jump in the interpolated phase.

According to our simplified model, the emission of the shower originates from a point source at X_{max}. Considering a non-dispersive medium and straight line propagation, the expected arrival time at a given location is $n_{eff}.d/c$, where n_{eff} is the average value of the index of refraction along the path between X_{max} and the given location and d is the distance along this path. To recover the absolute time at the interpolated position, the interpolated signal is referred to the time of arrival at the interpolated position, using the corresponding n_{eff}. Deviations from this simple model, that
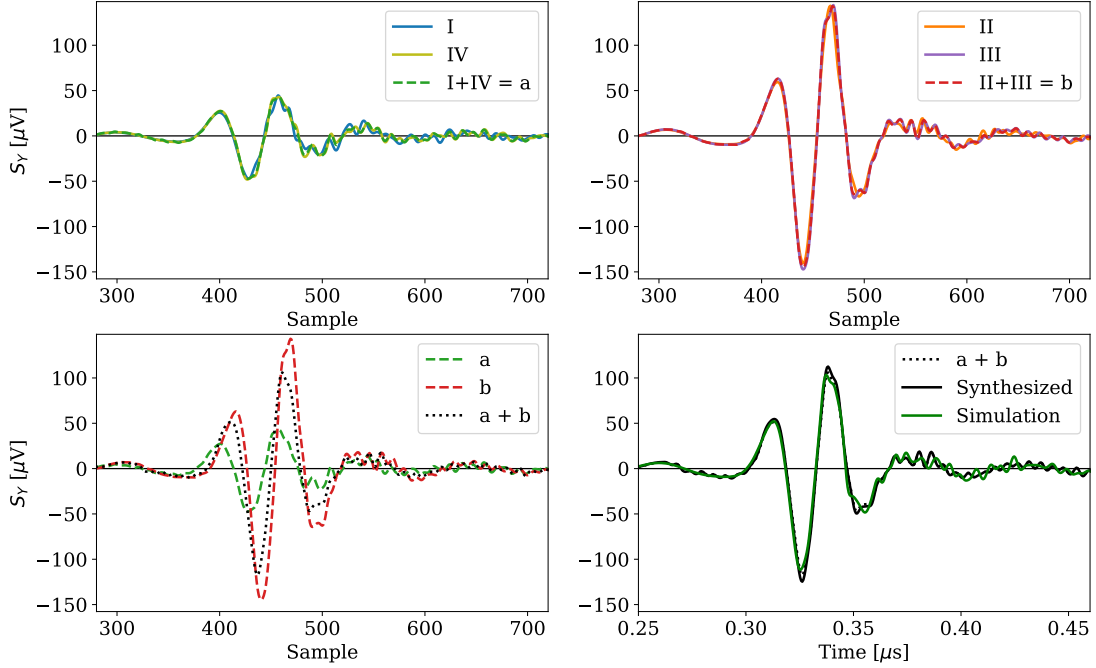
Figure 4. Example of the signal synthesis procedure. Signals in neighboring antennas I,II,III,IV (see fig. 1), for the example event of fig. 2, are interpolated into signals **a** and **b**, which are then interpolated, time aligned and scaled for the distance to get the synthesized signal. The true signal simulated at the desired antenna position is shown for comparison. The difference between the simulated and the synthesized signal is shown in the center of fig. 5.

might make the signal start a little before or after, will automatically be included in the reference signals, and propagated to the interpolation.

An example of the interpolation procedure, for the Y (East-West) polarization of the antenna marked with a cross in fig. 2 is shown in fig. 4. Note how similar signals I,IV (and its interpolation a) and signals II,III (and its interpolation b) are, showing the smoothness of the variations in ϕ. The resulting a and b are more dissimilar, showing the faster signal variation with α.

3 Performance of the method

3.1 Simulations used

To characterize the performance of the method, we produced a library of 3456 showers using ZHAireS 1.0.28 [22], with statistical thinning starting at $E/E_{primary} = 10^{-5}$ and the Sybill 2.3c [23] hadronic interactions model. The showers covered the energy range between $10^{16.3}$eV and $10^{18.6}$eV in 24 logarithmic steps of size 0.1, a 25% energy increase per step. The zenith angle θ ranged from 38.23 to 87.1 deg in 16 steps in $\log_{10}(1/\cos(\theta))$, giving each step approximately a 20% increase in footprint size. Gammas, Protons and Iron nuclei incoming from the North, South and East were used as primary particles. The ground level was set at 1086m above sea level. The

refractivity was modeled as an exponential with a value of 325 at sea level and a scale height of 8km.

For each shower, a conical star-shape of 160 antennas (8 arms, 20 antennas per arm) plus an additional 16 randomly distributed test antennas at ground level were generated. For each position, the simulated electric field was used to compute the response of the HorizonAntenna [11] to get the signal in each antenna channel, corresponding to each polarization direction: S_X (North-South), S_Y (East-West) and S_Z (Up-Down).

Each conical star-shape had its aperture angle (α_{max}) adjusted so that the peak amplitude of the the voltage signal falls below $45\mu V$ (3 times the expected galactic noise, σ_{noise}) around antenna number 10-12 of the arm, to guarantee good footprint coverage even if aggressive trigger scenarios are being studied.

At each test antenna the signal was synthesized following the prescription presented in this work, giving a total of 55296 synthesized signals in each channel to be compared with the simulation.

3.2 Zero Normalized Cross Correlation

To assess the similarity of the synthesized (S_i) and simulated (S_s) signals, we will study their Zero Normalized Cross-Correlation r(τ), defined for the zero-normalized signals $S'_j = S_j - \mu_j$ as

$$r(\tau) = \int \frac{(S'_s(t))(S'_i(t - \tau))}{\sigma_s \sigma_i} \, dt \tag{3.1}$$

where μ_j and σ_j are the mean and standard deviations of the corresponding signals. This is basically the Pearson's correlation coefficient usually used in correlation analyses [26], as a function of a signal lag τ. The value of τ that maximizes the modulus of the correlation function gives the error in time alignment of the signals Δt, while the value of r gives a measure of the similarity of the signals, with 1 denoting perfect correlation and -1 perfect anti-correlation.

An example of the evolution of the correlation function with τ for the 3 channels of the antenna selected in fig 2, scaled with the signal amplitude, is presented in fig. 5 (right column). In this particular case, the signal on the X channel is very small and incoherent, and although qualitatively the synthesized signal looks similar to the simulated one, the sample-to-sample difference is of the order of the signal itself. In this case r=0.4, and was very close to actually being classified as an anti-correlation with lag 12ns, looking at the negative maximum that corresponds to r=-0.34. The signal in the Z channel is 4 times larger (but still of the order of the expected σ_{noise}) and the correlation improves significantly, with r=0.81. The Y channel is where most of the signal power is received, the sample-to-sample difference is of the order of 10%, and r=0.98. In this case, the difference comes mostly from a high frequency component in the interpolated signal that is not present in the reference simulation.

The maximum cross-correlation r of all the signals in the simulation library in each polarization component is shown in fig. 6, as a function of the signal peak-to peak amplitude in units of the trigger amplitude, defined for this study to be 5 times σ_{noise} (akin to a signal over noise). The lines corresponding to 1,2,3 and 4 σ_{noise} are displayed to aid in the interpretation at other trigger levels. To give a measure of the overall agreement of the signal we also present in the top-left of this plot an average r, weighted by the amplitude on each channel, as a function of the maximum amplitude of the vectorial sum of the signals, defined as
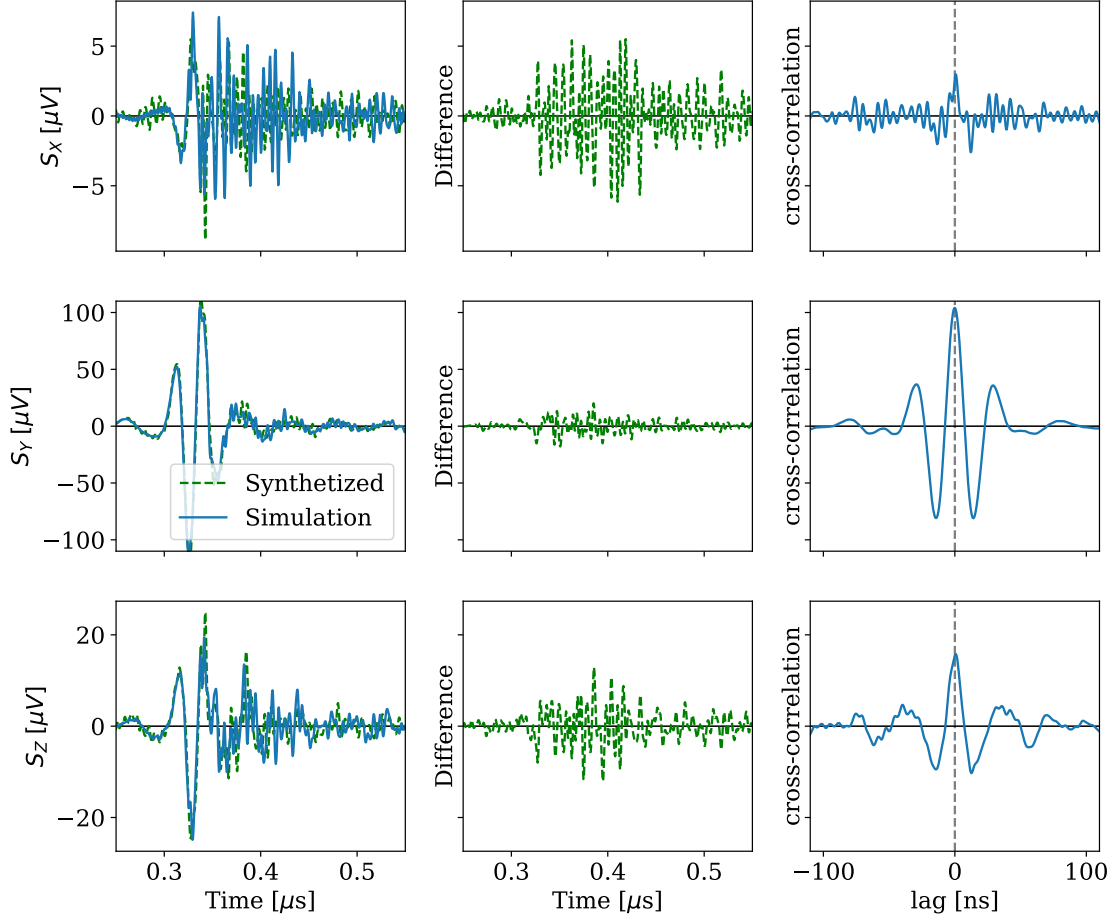
Figure 5. Example signals on the three polarization channels (left) for the selected antena of fig. 2, showing the difference between the simulated and synthesized signal (middle) along with the correlation function r(τ) multiplied by the signal maximum (right). Note that all the rows share the same scale for the ordinate. The maximum of the correlation is achieved close to lag 0, where the maximum correlation r=0.4 for S_X, 0.98 for S_Y and 0.81 for S_Z

$$|S| = \left(\sqrt{S_X(t)^2 + S_Y(t)^2 + S_Z(t)^2} \right) \tag{3.2}$$

The figure shows that the interpolation performs exceptionally well once the signals are above 3-5 σ_{noise} for antennas located farther than 3-5 km away from X_{max}.

Looking at the values of τ that maximize the correlation, shown in fig. 7, we can see that once the threshold signal amplitude is reached, the timing of the synthesized signal is almost perfect, with most events requiring shifts of less than 1ns. This is a testament to the accuracy of the model proposed to align the timing of the interpolated signals.
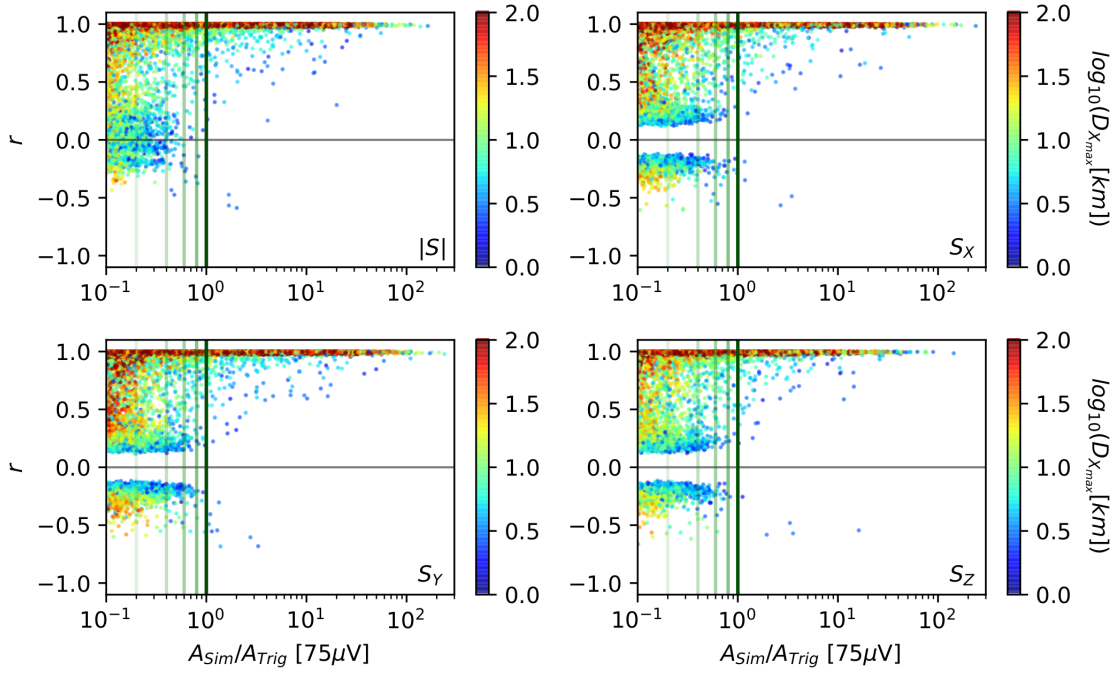
Figure 6. Correlation coefficient on each polarization and their weighted average (top-right, see text) for all the signals on the simulation library, as a function of the signal amplitude. Color indicates the distance from the antenna to X_{max}
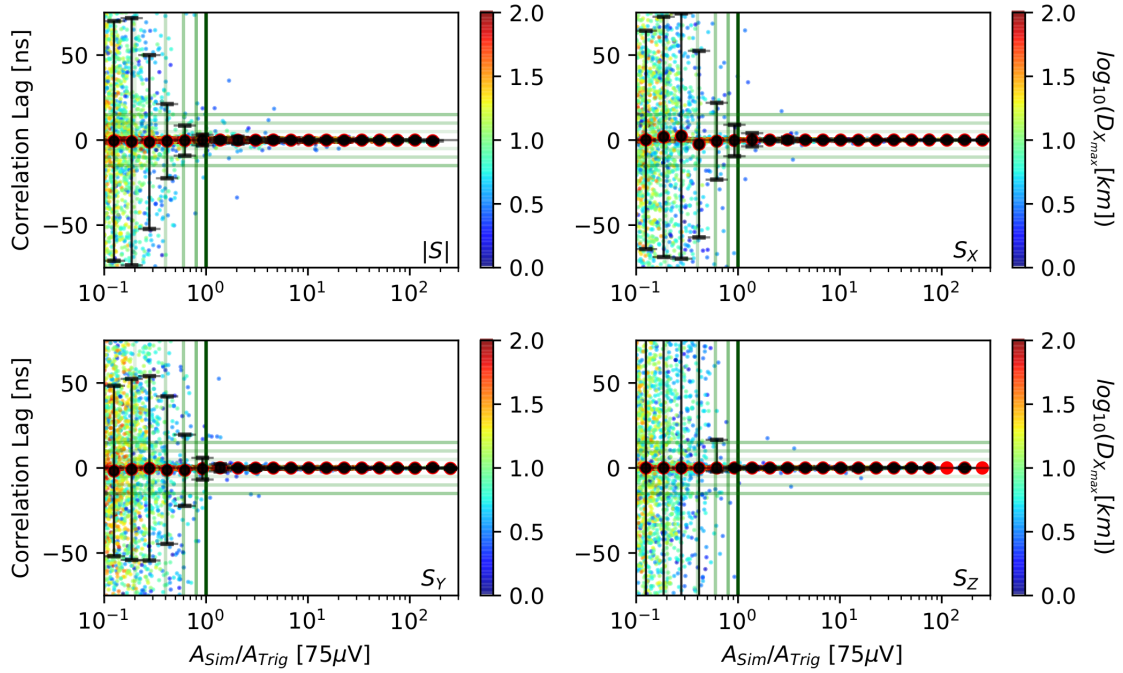

Figure 7. Correlation lag (with the weighted average on top-right, see text) for all signals on the simulation library, as a function of the signal amplitude for each polarization. Color indicates the distance from the antenna to X_{max}

The example shown in fig. 5 illustrates that while a value of r close to 1 assures a good correlation, a low value of r is more difficult to interpret. For many applications, the signal synthesized for the X channel would be considered an acceptable rendition of what to expect at that antenna position, even if the relative sample-to-sample difference with the reference simulation is big. The peak-to-peak amplitude is within 8% and the signal start time is accurate. For a larger signal, on the other hand, a value of r of 0.4 would be unacceptable.

3.3 Reduction in CPU time

ZHAireS computes the electric field of the shower by adding the contribution from every advancement step (a track) of every electron and positron in the shower. For every track in the simulation, there is one call to the field calculation routine per antenna, so the total CPU time is proportional to the number of required antennas. Inclined showers cover more distance in the atmosphere before arriving to ground level, requiring more tracks to be computed. Consequently, the CPU time scales approximately as $1/\cos(\theta)$.

The total CPU time required to compute the 16 test antennas of each event was on average 3.2 hr, varying from 1 hr at a 38 deg zenith angle to 10hr at 87 deg. In contrast, synthesizing the signal for 16 antennas with the method presented in this article takes on average 10 seconds, and its independent of the zenith angle. This is an improvement between 2.5 and 3.5 orders of magnitude in the required CPU time. To put it into perspective, making the microscopic simulation of the 55296 signals used for the comparisons took 1.3 years of CPU time and had to be run at a supercomputing center, while the synthesis of the same signals took only 7.5 hr of CPU time on a laptop. The interpolation method was implemented in python and was not optimized for speed, its performance can still be significantly improved.

The gain comes of course at the expense of having to previously compute the 160 signals in the star-shape for each event, which took 13 years of CPU time. The gains from this method are capitalized only when the produced star-shapes simulations are to be re-used several times, as is the case in array optimization studies.

4 Performance of the method on signal observables

We will focus our attention on the accuracy of the prediction for the peak-to-peak amplitude, which is generally used to evaluate the trigger probability when studying different array configurations. To asses the feasibility of using the method to characterize event reconstruction algorithms, we will also pay attention to the precision of the timing of the peak of the synthesized signals, used in the reconstruction of the incoming direction and position of X_{max}, and to the integrated power of the peak, used as a proxy for the reconstruction of the fluence, and thus the primary energy.

We have seen in the previous section that the interpolation has problems with signals with very low amplitudes and with antennas very close to the shower maximum, where the antennas are "embedded" in the shower and the signals lacks sufficient coherence. For this reason we set a quality cut removing antennas closer than 4 km to the shower maximum for the remainder of the discussion.

4.1   Maximum amplitude

The peak-to-peak amplitude (from now on, the amplitude for simplicity) is regularly used to estimate the probability of detection, usually applying a simple signal-over-threshold trigger condition. If the amplitude exceeds the trigger condition, the antenna is considered to have detected the event. The condition can be applied on each polarization channel separately, or on a combination of channels to form an antenna-level trigger, as could be the vectorial sum of the signal defined in eq. 3.2.

The accuracy of our method to give the maximum of the signal is illustrated in fig. 8, where the relative error in the amplitude on each channel is shown as a function of the amplitude. Like in fig 5, vertical green lines show the amplitudes corresponding 1 to 4 times σ_{noise}, to aid in the interpretation of the figure for different threshold values.

Since the distribution of the errors is not Gaussian, we use the mean and the median to characterize the accuracy of the method. To characterize the precision of the method, we use the interval containing 34.15 % of the population at each side of the median, akin to one standard deviation, and the interval containing 47.8 % of the population at each side of the median, akin to 2 standard deviations. We will call this precision estimates the 68% and 95% tolerance intervals (t.i.)

The close proximity between the mean and the median shows that the bulk of the events are symmetrically distributed around the median, and only the outliers are skewed. This is also evident from the increase in the asymmetry of the external error bars. Across all the channels, we see that the maximum amplitude is synthesized with an average accuracy well within 2%, and a precision above threshold well within -4/+7.5% at the 68% t.i., and within -8/+15% at the 95% t.i., diminishing with increasing amplitude. Furthermore, almost no outliers are found above a 25% relative error, which can be considered as an absolute upper limit.

Studies working with the signal amplitude face bigger systematic uncertainties from other sources, like the choice of the simulation program (20% [1]) or the experimental error from the calibration of the antennas (15-20 % [2]). Its safe to say that using synthesized signals instead of complete simulations won't be the major source of uncertainties.

4.2   Triggering errors

A finite precision in the determination of the amplitude will induce false positives and false negatives in the discrimination of the trigger condition. Events with an under-estimation of the amplitude, falling within the lower shaded area in Fig. 8 will give false-negatives, while events with an over estimation of the amplitude falling within the upper shaded area will give false-positives.

The chance of miss identification close to the threshold is shown in Fig. 9, were it can be seen to quickly drop within 8% of the trigger value, in accordance to the accuracy quoted in the determination of the amplitude. In our test library, less than 0.5% of the signals where miss-identified as triggering and less than 0.25% as non-triggering, regardless of the polarization channel. Note that when building event trigger conditions that involve multiple antennas, the miss-identification chance will be reduced even further.

4.3   Integrated Power

The integral of the fluence of the electric field at the ground plane can be used to reconstruct the energy of the particle that originated the cascade [24]. The fluence at each antenna location
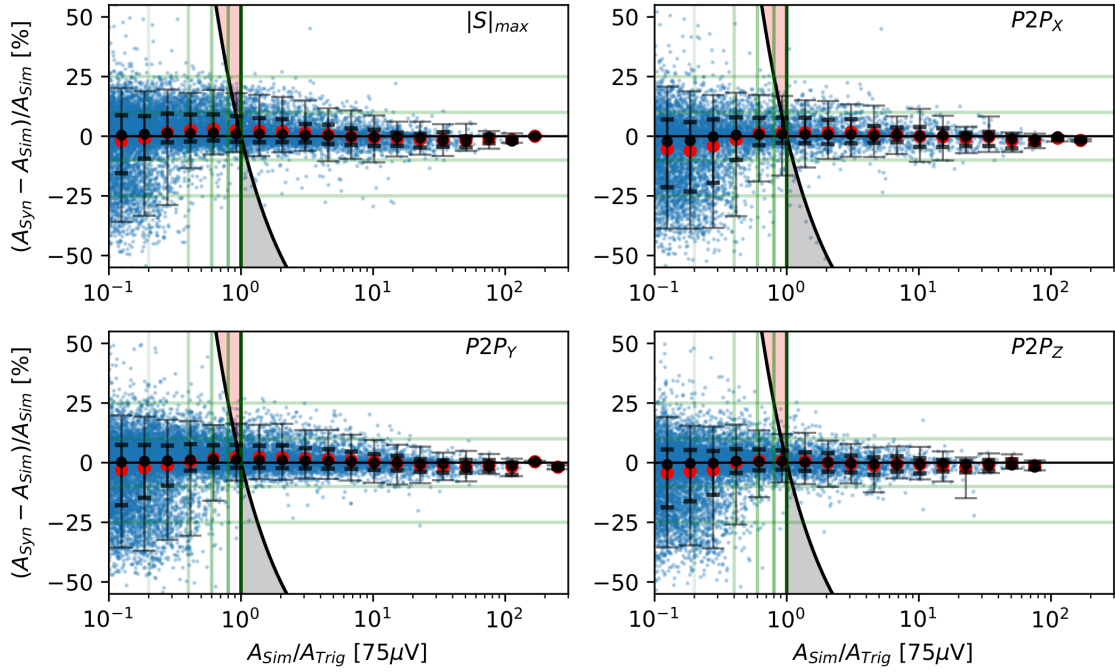
Figure 8. Relative errors in the amplitude of the synthesized antenna response, for the vectorial sum and for each antenna channel, as a function of the amplitude. Shaded areas indicate signals that would be miss-identified as false positives (red) or false negatives (gray). Vertical lines indicate 1, 2, 3, 4 and 5 (the trigger level) times the galactic noise. Horizontal lines indicate a 10 and 25% relative error. Black dots indicate the median of the amplitude bin, while red dots indicate the average. Thick (thin) error bars contain 68.3% (95%) of the events around the median

is calculated by integrating the absolute value of the Poynting vector in a time window of 200ns around the signal maximum. As a proxy to this quantity we use the integrated power received by the antenna, in the same time window.

The error in the determination of the integrated power is shown in Fig. 10, following the considerations presented in section 4.1. The average synthesized integrated power has a precision and accuracy comparable to the one found for the signal amplitude, well below the expected systematic errors coming from the antenna calibration, the reconstruction methods and the signal simulation itself, showing that this method can be safely used for studying the energy reconstruction algorithms. There is a slight trend in the average difference, implying that the method over-estimates the integrated power at signal amplitudes close to the threshold by 2% and underestimates it by the same amount at high integrated power, a trend that is also visible for the signal amplitude. We will get some insight on this by looking at the performance of the method as a function of α.

4.4   Linear interpolation effects

Despite having an almost unbiased estimation of the amplitude when considering the the full simulations set, some dependence of the quality of the synthesis of the signal as a function of α is expected to exist, as we are using linear interpolation for functions that are clearly non-linear.
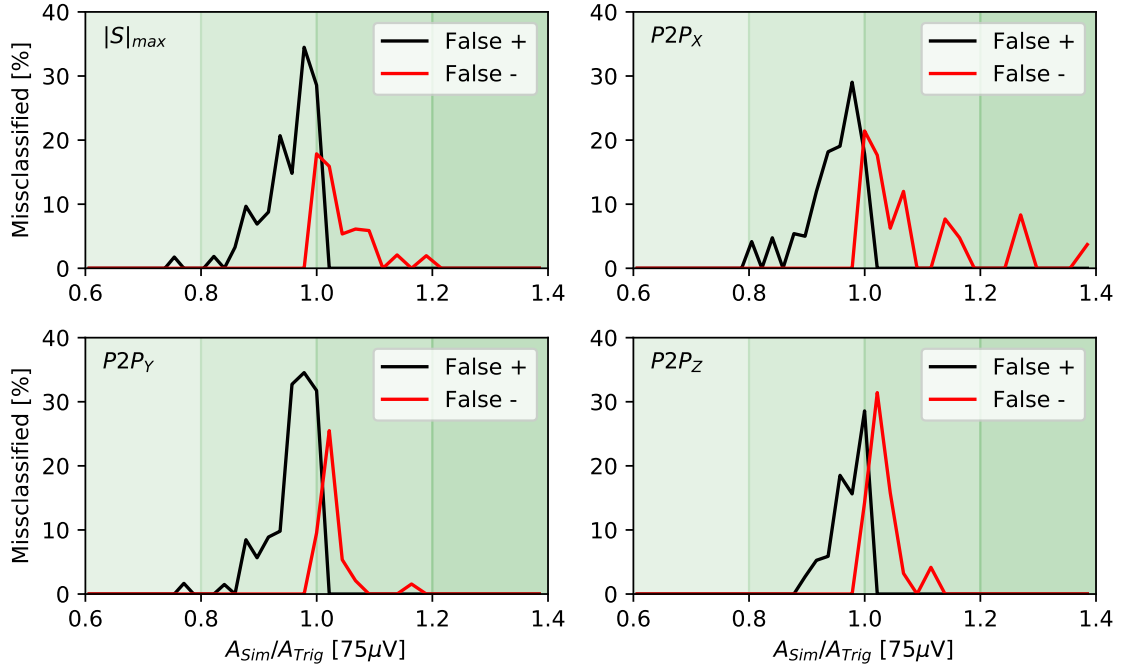
Figure 9. Miss-identification chance close to the threshold as a function of signal amplitude. Vertical lines indicate 4, 5 (the trigger condition) and 6 times the expected galactic noise.
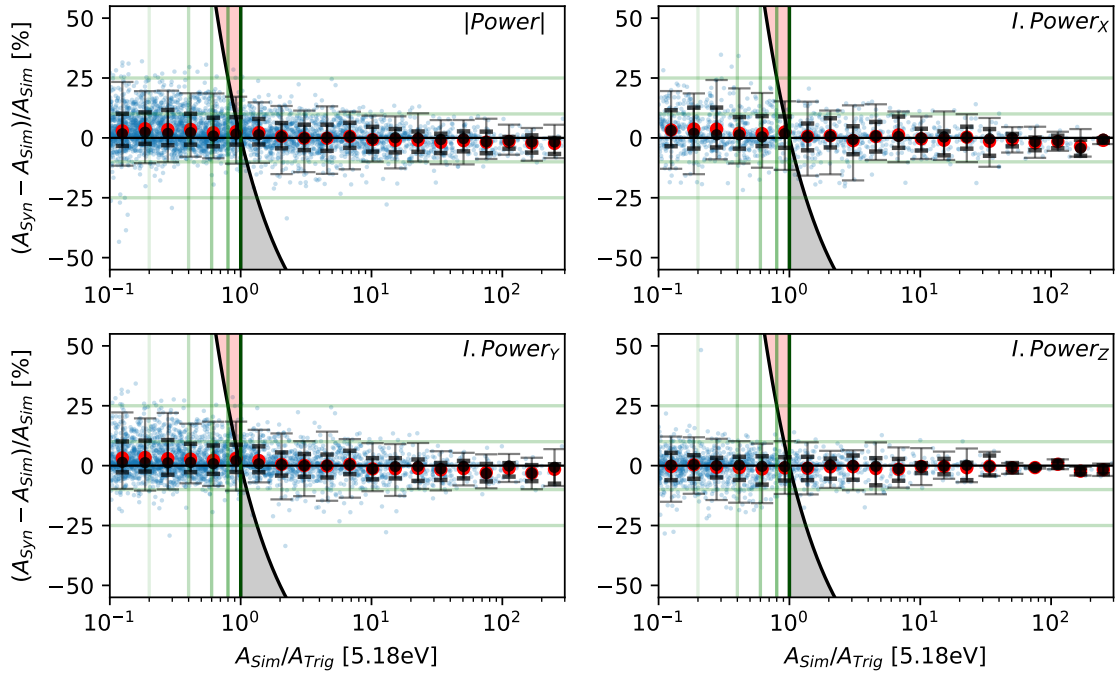

Figure 10. Relative errors in the integrated power of the synthesized antenna response for the Hilbert envelope of the complete signal and for the Peak-to-Peak amplitude in each polarization direction. Elements of the figure are described in fig 8
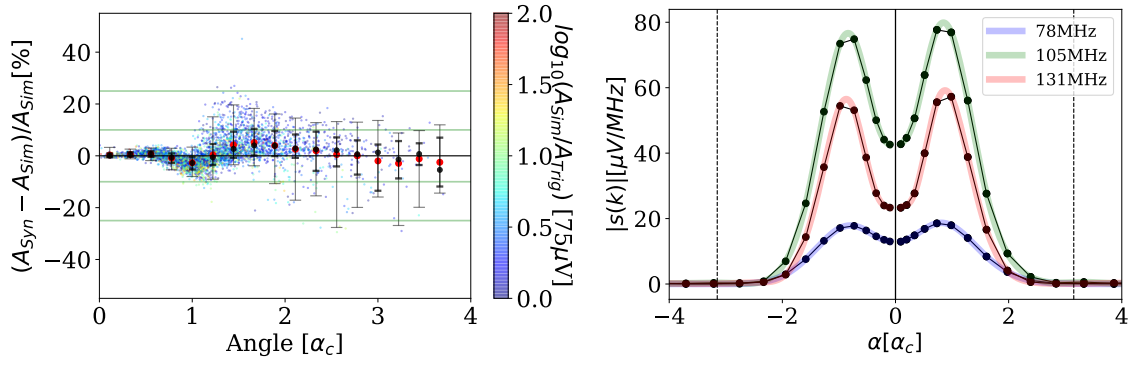
Figure 11. Relative error of the peak-to-peak amplitude of the sum of the antenna voltage (left). Horizontal lines indicate absolute errors of 10 and 25%. The amplitude of three different frequency components, and their linear interpolation are shown (right), to illustrates the angles at which the linear interpolation over/underestimates the amplitude. The vertical dashed lines indicate the angle after which the signals drop below the trigger threshold. See text for details.

To illustrate how this affects the signal synthesis we show in fig 11 (right), for our example event of fig. 2, the lateral distribution function (LDF) of the amplitudes of the Fourier components of the Y channel at 3 different frequencies. The South arm of the star-shape is displayed at negative values of α and the North arm at positive values of α. Looking closely at the linear interpolation between the antennas, we can see that when the curvature of the LDF is positive, a linear interpolation results in an overestimation of the signal, and when the curvature is negative in an underestimation. The errors are bigger at the places of maximum curvature, and the relative error is bigger at the inflection of the tail.

The α at which the curvature changes varies with frequency and the relative amplitudes of the frequency components of a signal vary with α, making the final effect on the signal amplitude difficult to grasp. All frequency components will be underestimated close to α_c, and all will be over estimated for α sufficiently big, with a different relative error for each frequency component.

Figure 11 shows the net effect on the vectorial sum of the antenna signals $|S|$. Individual components display the same behavior. The mean bias is of the order of -5% at α_c, negative as expected, and turns positive for bigger α. Notice that outliers are caused by low amplitude signals, and that the precision is reduced with increasing α. This also affects the integrated power, and is responsible for the trend observed in the average bias mentioned in the preceding section. Low amplitude signals, usually located at high α, will tend to be over estimated while high amplitude signals, usually located at α_c, will be under estimated.

These biases could be corrected with an ad-hoc parameterization, or reduced using a star-shape denser in α. A non-linear interpolation, using the expected functional form of the LDF at each frequency, is a possible future optimization for the method.

4.5 Peak Timing

For the reconstruction of the shower incoming direction, and of the position of X_{max}, its necessary to have the correct time of arrival of the signal at the antenna. Since the signal will be always
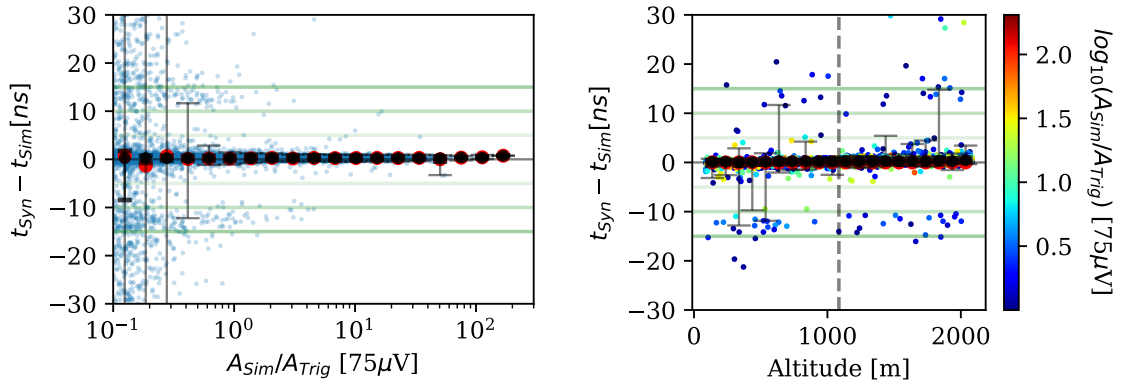
Figure 12. Hilbert envelope peak time error as a function of peak amplitude in the synthesized antenna voltage. Vertical lines indicate the 1, 2, 3, 4 and 5 (the trigger condition) times the expected galactic noise. Horizontal lines indicate absolute errors of 5, 10 and 15 ns.

recorded in a noisy environment, the time of arrival is determined using the timing of the maximum of the Hilbert envelope of the signal. We have seen in section 3 that the lag between the synthesized and the simulated signal is close to 0, but differences in the shape of the signals could introduce shifts in the position of the maximum. Fig. 12 (left) shows the error in the timing of the maximum of the Hilbert envelope for the antenna response to the electric field. For signals below the trigger threshold it is sometimes difficult to identify a clear single peak structure in the signal, giving rise to outliers. A more robust algorithm than just looking for the maximum amplitude could be used to improve on this, if deemed necessary. Barring these outliers by considering the 95% t.i., we can see that the timing accuracy is better than 1ns above 5 σ_{noise}, and is good to 2ns down to a more agressive 3 σ_{noise} trigger threshold. This is comparable to the 2ns time stamping accuracy that can be achieved with current GPS technology [25], indicating that signals synthesized with the method presented in this article can be used to develop and characterize reconstruction algorithms based on signal timing information without resulting in significant systematic errors.

4.6 Local topography

To test the performance of the method in dealing with different topographies, we generated an additional library of 1152 events in which the 16 test antennas were given an additional random vertical offset between -1000m and 1000m. For this library of events we used the same set of configurations as described in section 3.1, but at a constant energy of $10^{18.6}$eV to maximize the statistics of signals above trigger.

Figure 13 shows that there is no bias and no change in the accuracy of the maximum amplitude of the synthesized signal associated to the topography correction introduced in section 2. The timing accuracy is also not affected as seen in Fig. 12 (right). Note that in this case the outliers correspond to the lower amplitude signals, which suffer from miss-identification of the true position of the signal peak.
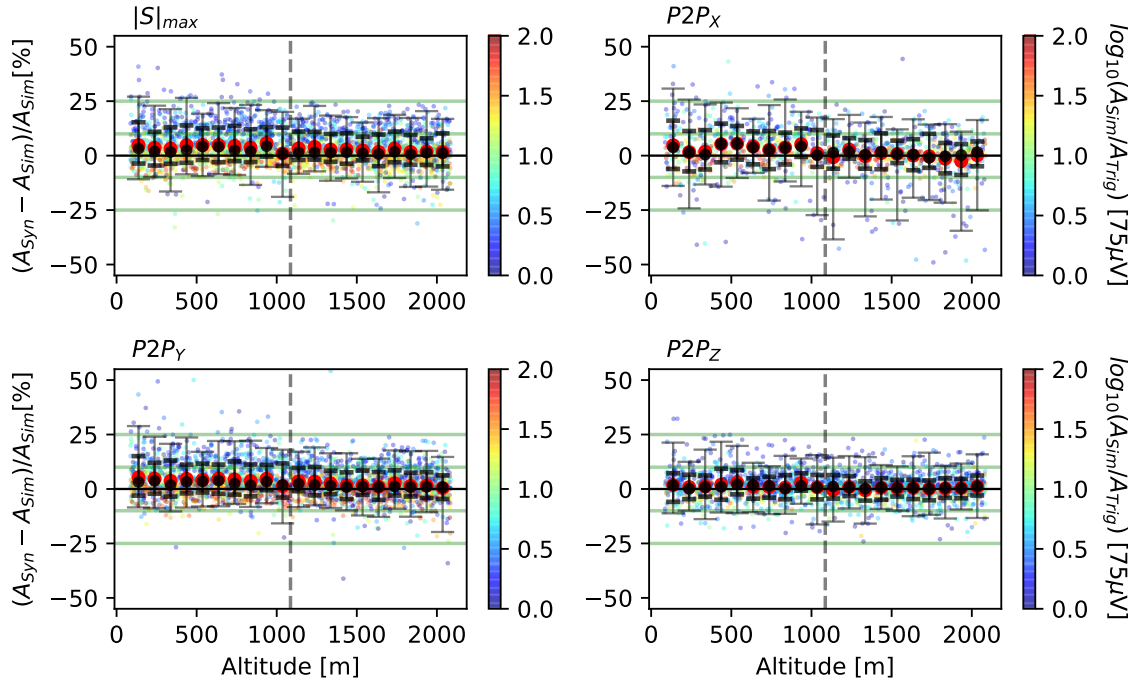
Figure 13. Relative errors in the amplitude of the synthesized antenna response for the vectorial sum and for each antenna channel, as a function of the altitude of the antenna. Horizontal lines indicate a 10% and 25% relative error. The vertical line indicates the altitude of the ground plane.

5   Conclusions

We have developed a method for synthesizing signals with the precision of microscopic simulations but with a speed comparable to macroscopic methods, that opens the possibility to make accurate characterizations and optimizations of radio arrays for the detection of extensive air showers.

The method has a precision on the estimation of the signal amplitude, usually used in triggering and reconstruction algorithms, of -4/+7% at a 68% t.i and of -8/+15% at a 95% t.i., well below other sources of systematic errors in the simulation chain. The method can handle differences in ground altitude related to the changes in local topography , without loosing its accuracy. The timing of the synthesized signals is accurate within to 2ns, enabling the use of the method for the study of geometry reconstruction algorithms. Since the full time trace of the signal is synthesized with good accuracy, the method could even be used for the production of datasets for the training of reconstruction algorithms based on neural networks, especially if noise is to be added later in the simulation chain.

Improvements are possible, at the expense of increasing complexity or a narrowing scope of application. Fine tuning the antenna spacing to the expected position of the Cherenkov peak, and using a non linear interpolation of the amplitude are two clear lines of improvement. Since we know that the method starts having problems for low amplitude signals (below 3 σ_{noise}), setting a tighter maximum angle in the star-shape would already make the method more precise and efficient.

We focused the study on the performance for the response of the HorizonAntenna, but it would

be equally applicable to other antennas or for the synthesis of the electric field filtered in the 30 to 300 Mhz range. The method could also be used for synthesizing the unfiltered electric field, if the problems of the linear interpolation at the Cherenkov angle is addressed properly.

Once a sufficiently big library of events with the star-shape patterns is available, this method provides a 3 orders of magnitude reduction in the CPU time required to produce the simulations for any given study. Since computing time consumption is one of the biggest environmental impacts of building a radio array for the detection of extensive air showers, early adoption of this method could significantly reduce the CO_2 footprint of such projects.

6 Aknowledgments

References

[1] T. Huege., Physics Reports 620 (2016), pp. 1-52.

[2] F. Schröder, Progress in Part. and Nuc. Phys. 93 (2017) 1-68 doi:10.1016/j.ppnp.2016.12.002

[3] O. Scholten, K. Werner, F. Rusydi, Astroparticle Physics 29 (2008) 94-103., doi:10.1016/j.astropartphys.2007.11.012

[4] O. Scholten, T. N. G. Trinh, K. D. de Vries, B. M. Hare, Phys. Rev. D 97 (2) (2018) 023005, doi:10.1103/PhysRevD.97.023005

[5] V. Marin, B. Revenu, Astroparticle Physics 35 (2012) 733-741., doi:10.1016/j.astropartphys.2012.03.007

[6] K. D. de Vries, et al., AIP Conference Proceedings, Vol. 1535, 2013, pp. 133-137.

[7] J. Alvarez-Muñiz, W. Rodrigues, A. Romero-Wolf, M. Tueros and E. Zas, Astropart. Phys., 35, 325, (2012)

[8] T. Huege, M. Ludwig, and C. W. James., AIP Conf. Proc. 1535 (2013), p. 128

[9] D. Butler, T. Huege and O Scholten, 35th International Cosmic Ray Conference (2017) PoS(ICRC2017) 307.

[10] A. Zilles et al., Astroparticle Physics 114 (2020) 10-21, doi:10.1016/j.astropartphys.2019.06.001

[11] J. Alvarez-Muñiz et al. (The GRAND Collaboration), Science China Physics, Mechanics & Astronomy volume 63 (2020), 219501. doi:10.1007/s11433-018-9385-7

[12] J.Alvarez-Muñiz, W.R. Carvalho Jr., H. Schoorlemmer, E. Zas, Astroparticle Physics 59, (2014) 29-38. doi:10.1016/j.astropartphys.2014.04.004

[13] G. Askaryan, Sov. Phys. JETP 14, 441 (1962)

[14] G. Askaryan, Sov. Phys. JETP 21, 658 (1965)

[15] H. Falcke, P. Gorham, Astroparticle Physics 19 (2003) 477-494, doi:10.1016/S0927-6505(02)00245-1

[16] K. D. de Vries, A. M. van den Berg, O. Scholten, K. Werner, Phys. Rev. Letters 107 (6) (2011) 061101, doi:10.1103/PhysRevLett.107.061101

[17] J. Alvarez-Muñiz, W. R. Carvalho, Jr., A. Romero-Wolf,M. Tueros, E. Zas, Phys. Rev. D 86 (2012) 123007. doi:10.1103/PhysRevD.86.123007

[18] A. Nelles, et al.,Astroparticle Physics 65 (2015) 11-21. doi:10.1016/j.astropartphys.2014.11.006

[19] P. Schellart et al JCAP10(2014)014, doi:10.1088/1475-7516/2014/10/014

[20] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery. Numerical Recipes in C-The Art of ScientificComputing, 2nd ed.; Cambridge University Press: New York, NY, USA, 1992

[21] E. Holt, Simulationsstudie für ein großskaliges antennenfeld zur detektion von radioemission ausgedehnter luftschauer, Master's thesis, Karlsruhe Institute of Technology (KIT) (2013). URL https://web.ikp.kit.edu/huege/downloads/theses/2013-Holt.pdf

[22] http://aires.fisica.unlp.edu.ar/zhaires

[23] F. Riehn, R. Engel, A. Fedynitch, T.K. Gaisser and T. Stanev, arXiv:1912.03300

[24] A. Aab et al., Phys. Rev. D 93 (2016) 122005 doi:10.1103/PhysRevD.93.122005

[25] A. Aab et al, J. of Instrumentation 11 (2016) P01018 doi: 10.1088/1748-0221/11/01/p01018

[26] R. Falk, A.D. Well, Journal of Statistics Education (1997) 5:3 doi: 10.1080/10691898.1997.11910597