Electric Field Reconstruction with Information Field Theory

Simon Strähnz,[a,*] Tim Huege,[a,b] Philipp Frank[c] and Torsten Enßlin[c]

[a]*Karlsruhe Institute of Technology,*
 Kaiserstraße 12, 76131 Karlsruhe, Germany

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

[c]*Max-Planck-Institute for Astrophysics,*
 Karl-Schwarzschild-Str. 1, 85748 Garching b. München, Germany
 E-mail: simon.straehnz@kit.edu, tim.huege@kit.edu,
 philipp@MPA-Garching.MPG.DE, ensslin@MPA-Garching.MPG.DE

Reconstructing the electric field from the measured voltages in an antenna, unfolding the antenna response, comes with several problems. Due to the noisiness of the signal it is often necessary to disregard part of the bandwidth of the antenna. It is also not guaranteed, that this system of equations can be inverted at all. In any case, the noise of the measurement will be converted into noise on the electric field.

This could be solved by Bayesian inference, however, the electric field is continuous, which would lead to an infinite-dimensional latent space. Information field theory (IFT) has been developed to deal with this problem and allow for Bayesian reasoning on fields. It provides a theoretical backbone and effective tools to approach the inference as a discrete problem in the continuum limit, taking the continuous nature of fields into account.

We will present a first working signal model that can be used with IFT-based inference algorithms, which can successfully reconstruct the electric field. The model is based on the current understanding of air shower emission physics, modelling geomagnetic and charge-excess emission and their respective polarisation and spectra separately. Since Bayesian inference provides the posterior distribution, this method also gives an estimate on the uncertainty of the measured field. The performance of this method will be demonstrated with Monte-Carlo simulations of air shower radio signals.

---

*Speaker

1. Introduction

Measuring extensive air showers with radio detectors is a very attractive option. Radio detectors are cost-effective, so it is easy to cover a large area with them. They are also able to directly probe only the electromagnetic component of air showers, so they can be used in conjunction with particle detector arrays to measure the ratio of the electromagnetic to the muonic component, a crucial indicator to determine the type of the primary particle. The one big drawback to radio detection is its susceptibility to anthropogenic noise. The most power from air shower signals is expected in the range of 1 to 200 MHz, a frequency range not only very commonly used for radio communications, but also where the galactic background emission is very high. [1]

Reconstructing the extensive air shower from the measured voltages in radio antennas first requires the reconstruction of the electric field generated by the air shower at the location of the antenna. This is traditionally done by simply trying to invert the antenna response. In most detectors, antennas with only two polarisations are used, which comes with the limitation that it is impossible to infer a three-dimensional field from only a two-dimensional measurement. This problem can be circumvented by employing the far-field approximation (i.e., the electric field component in the direction of propagation of the wave vanishes). This still does not guarantee that a reconstruction is possible: If the direction of propagation is parallel to one polarisation of the antenna, it will not have any signal and the problem of dimensionality reappears. This method also does not take measurement noise into account. The noise on the measurement of the voltage in the antenna is simply propagated to noise on the measurement of the electric field.

A promising method to handle this reconstruction step is Bayesian inference. However, applying Bayesian inference to continuous fields is in general an ill-posed problem: A continuous field will have an infinite number of degrees of freedom, while the number of our measurements will always be finite. This can be solved with Information Field Theory.

2. Information Field Theory

Information Field Theory (IFT) studies, as the name suggests, information theory applied to fields. Through IFT it has been shown, that these ill-posed problems can be approached by the careful setting of prior probabilities using our current understanding of physics combined with Gaussian processes. In this way, it is possible to capture the physical and probabilistic constraints imposed on the field of interest and use them to handle infinite-dimensional latent spaces. Even though this requires discretisation at some point, it has been shown that this is allowable when working in the continuum limit. [2]

In practice, working with IFT requires defining a suitable, probabilistic forward model, which includes the current understanding of the physical processes leading to our measurement, of the measurement process itself and of the prior probabilities associated with both. This model can then be used to apply Bayesian inference. Traditionally, there are two approaches to perform the inference: Markov-Chain Monte-Carlo sampling (MCMC) and Variational Inference (VI). The first approach (MCMC) directly samples the posterior distribution and is thus guaranteed to reconstruct the true posterior in the limit of infinite samples. However, it is also very computationally expensive

and thus mostly unsuitable for IFT approaches, which work in the continuum limit and thus have a high number of degrees of freedom.

The more useful approach is to approximate the true posterior with a simpler but more easily calculable family of distributions, which is the basis of VI. Two VI methods have been developed for usage with IFT: Metric Gaussian VI (MGVI) [3] uses a Gaussian distribution as the approximate posterior and estimates its covariance with the inverse Fisher metric. In an iterative process, the covariance matrix is estimated and the Kullback-Leibler divergence is minimised, until a satisfactory agreement is reached. In a further development of this method, Geometric VI (geoVI) [4] can also handle non-Gaussian posteriors by way of applying a coordinate transform to the latent space to transform it into the space where the posterior is Gaussian.

3. Forward Model

The forward model used in this work for the variational inference was developed on the basis of the model used by C. Welling et al. [5]. In general, the same approach of modelling the electric field in the frequency domain as an amplitude ($\mathcal{E}$) and phase (ϕ) with a certain linear polarisation defined by a polarisation angle (ϕ_{pol}) is applied. The entire problem is considered under the far field approximation ($E_r = 0$):

$$\vec{E}(f) = \mathcal{E}(f)e^{i\phi(f)}(\cos\phi_{\text{pol}}\hat{\theta} + \sin\phi_{\text{pol}}\hat{\varphi}) \tag{1}$$

Here, geomagnetic emission and charge-excess emission have been modelled separately with their own separate amplitudes, phases, and polarisations. Narrowband RFI, as it is part of the measured electric field, has also been included in the model, however, this is modelled separately for every antenna polarisation.

$$\vec{E}(f) = \vec{E}_{\text{geo}}(f) + \vec{E}_{\text{CE}}(f) + \vec{E}_{\text{RFI}}(f) \tag{2}$$

$$= [\mathcal{E}_{\text{geo}}(f)e^{i\phi_{\text{geo}}(f)}\cos\phi_{\text{pol,geo}} + \quad \mathcal{E}_{\text{CE}}(f)e^{i\phi_{\text{CE}}(f)}\cos\phi_{\text{pol,CE}} + \mathcal{E}_{\text{RFI},\theta}(f)e^{i\phi_{\text{RFI},\theta}(f)}]\hat{\theta} \tag{3}$$

$$+ [\mathcal{E}_{\text{geo}}(f)e^{i\phi_{\text{geo}}(f)}\sin\phi_{\text{pol,geo}} + \quad \mathcal{E}_{\text{CE}}(f)e^{i\phi_{\text{CE}}(f)}\sin\phi_{\text{pol,CE}} + \mathcal{E}_{\text{RFI},\varphi}(f)e^{i\phi_{\text{RFI},\varphi}(f)}]\hat{\varphi} \tag{4}$$

3.1 Detector model

The measured data is assumed to be described by the equation

$$\vec{d}_i = \vec{V}(t_i) + \vec{n}_i \tag{5}$$

i.e., the measured voltage in bin i is the voltage $\vec{V}$ induced by the electric field in the antenna at the corresponding time t_i plus some measurement noise $\vec{n}_i$. The voltage is obtained by applying the instrument response to the electric field in frequency space and then transformed into real space with the inverse discrete Fourier transform:

$$\vec{V}(t) = \text{DFT}^{-1}\left(\vec{V}(f)\right) = \text{DFT}^{-1}\left(\tilde{H}_f\vec{E}(f)\right) \tag{6}$$

Hereby is $\tilde{H}_f$ the frequency dependent complex instrument response operator which is taken to be linear and given by

$$\tilde{H}_f = \begin{pmatrix} H_\theta^1 & H_\varphi^1 & H_r^1 \\ H_\theta^2 & H_\varphi^2 & H_r^2 \\ H_\theta^3 & H_\varphi^3 & H_r^3 \end{pmatrix}_f \tag{7}$$

where H_x^i is the response of the ith channel to the electric field in the $\hat{x}$ direction. Since all H_x^i are complex, this encodes the amplitude as well as the phase shift of the induced voltage.

The measurement equation 5 can be rearranged as:

$$n_i = \vec{d}_i - \vec{V}(t_i) \tag{8}$$

Assuming that the noise in each bin n_i is independently identically distributed according to some distribution $\mathcal{N}$, it follows that the likelihood of the measurement (i.e., the probability of measuring $\vec{d}_i$ given the signal $\vec{V}_i$ is given simply by:

$$\mathcal{L} = P(\vec{d}_i | \vec{V}(t_i)) = \mathcal{N}(\vec{d}_i - \vec{V}(t_i)) \tag{9}$$

3.2 Signal model

Due to the limited bandwidth in many radio detectors, the amplitude ($\mathcal{E}$) is modelled semi-parametrically: S. Martinelli et al. [6] have shown that the amplitude of the spectrum of radio pulses from air showers can be described by linear (L) or quadratic (Q) exponentials

$$L = A \cdot 10^{m_f(f - f_0)} \tag{10}$$

$$Q = A \cdot 10^{m_f(f - f_0) + m_{f2}(f - f_0)^2} \tag{11}$$

whereby the parameters m_f, m_{f2} can be parameterised as a function the distance from the shower axis in units of the Cherenkov radius (r/r_c) and the geometric distance to shower maximum ($d_{\max}$). They have shown that the spectrum of the geomagnetic emission is best described using the quadratic exponential, whereas the charge-excess emission is already described well by the linear exponential. These parametrisations have been used as the parametric part of this forward model. To allow for deviations from this parametrisation, it is multiplied by an integrated Ornstein–Uhlenbeck process. This Gaussian process can be described as a biased, or bound, random walk. For this use case, it has been configured to create variations on the order of 10% around 1.

3.3 Noise model

The noise is described separately as noise from narrowband RFI, which is then incorporated into the signal model and the remaining instrument and galactic noise, which is modelled as independently identically distributed variables according to the noise distribution $\mathcal{N}$.

The narrowband RFI is modelled in frequency space and independently for each frequency bin. The amplitude is modelled as a random variable drawn from an inverse gamma distribution. Since this distribution falls off very quickly, this means that the amplitude is close to zero for almost all frequency bins, while there is still a reasonable possibility of very high peaks. The phase of the RFI is modelled as drawn from a wide Gaussian, making the prior practically agnostic on the phase of each RFI contribution

The remaining noise corresponding to the n in equation 5 is modelled as a multivariate Gaussian with a diagonal covariance matrix based on the standard deviation of the measured trace. This is derived for each measurement individually by calculating the sample standard deviation over a noise trace.

$$\mathcal{N} = \mathcal{G}(0, \Sigma) \tag{12}$$

$$\Sigma^2 = \text{diag}(\sigma_s^2) \tag{13}$$

This is by far the weakest link of the model and needs to be improved upon by better analysis of the noise distribution, which is currently ongoing: Since all detectors have a defined pass band, the noise is expected to be at least partly correlated. It has also already been shown that the real measurement noise is not purely thermal, i.e., it does not follow a normal distribution but has more pronounced tails. Thus, the current model severely underestimates the possibility of large deviations between signal and measurement due to noise.

4. Performance on simulations

This model has been tested on 800 CoREAS [7] simulations of inclined extensive air showers for proton, helium, nitrogen and iron primaries. These simulations range in energy from $10^{18.4}$ eV to $10^{20.1}$ eV, in zenith angle from 65° to 85° and over the entire azimuth angle range. The simulations were then passed through the detector simulation for the radio detector of the Pierre Auger Observatory [8, 9]. As a last step before reconstruction, measured noise from the Pierre Auger Observatory site was added to create realistic simulations of measurements. These were then reconstructed with the NIFTy framework [10] using the geoVI algorithm and the model as described above. For comparison, the full library (4000 showers) has been reconstructed using the current methods [11] and treated similarly. In the standard method, the antenna model is simply unfolded by inverting the antenna response matrix (eq. 7) and applying it to the measured voltage. Further analysis in then usually performed on the energy fluence, the amount of energy deposited per unit area. This can be calculated directly from the measured E-Field when using the model described in this work. For the standard method, noise is handled in this step: The energy fluence is calculated by subtracting the power of the noise (estimated from a noise window) from the power from the pulse (calculated in a signal window).

### 4.1 Reconstruction of the electric field

An example of the improvement of the reconstruction of the electric field can be seen in figure 1. It shows the simulated cosmic ray pulse in the time domain and frequency domain and the reconstruction, including a 1σ credible interval. This example is of a very-low signal radio detector station, which would usually have to be disregarded in further reconstruction. It can clearly be seen, that the IFT reconstruction is in very good agreement with the simulated ground truth.

To test the overall performance of the reconstruction of the electric field, a modified $\chi^2 = \sum (x - \bar{x})^2 / \sigma_x^2$ has been calculated for all traces. For a normal least-squares fit, a χ^2 score would be divided by the number of degrees of freedom, i.e., the number of data points minus the number of parameters. Since this method uses Bayesian inference and thus can have many more parameters than data points, this is not a reasonable approach.[1] Instead, the quality of fit can be judged by dividing χ^2 by the number of data points N. The distribution of this quantity should then be a χ^2 distribution with a mean of 1 and a mode of $1 - 2/N$ if the errors are assumed to be the standard deviation of a normal distribution. The distributions for this reconstruction and the standard method

---

[1]In fact, this method employs roughly 4 times more parameters than data points, which are however strictly constrained by their prior distribution.
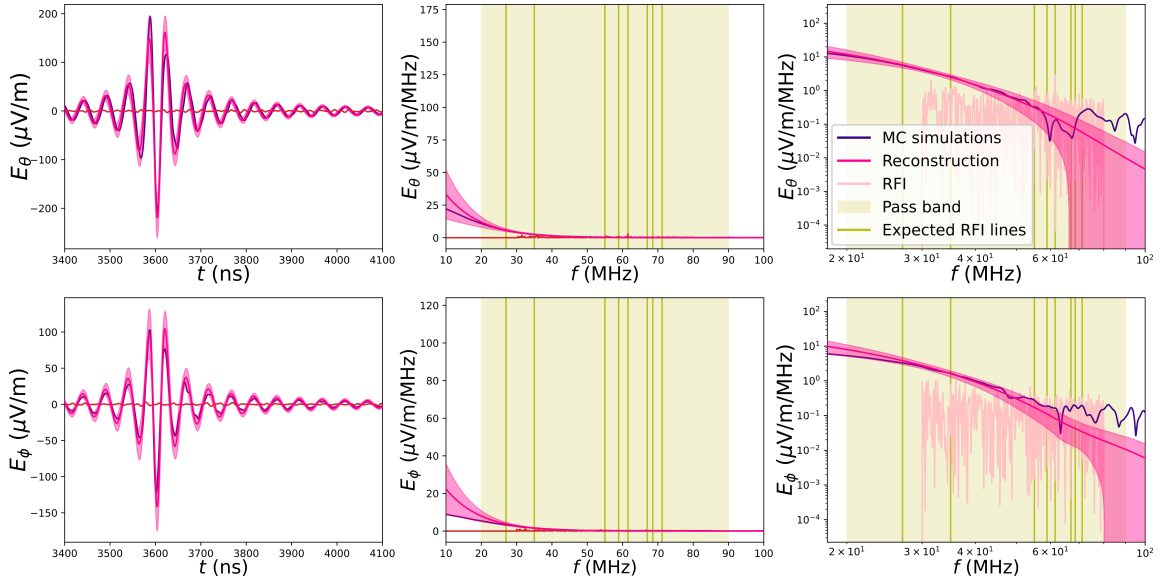
Figure 1: Example of the reconstruction of the electric field with the method presented in this work. Each row shows one polarisation. The first column shows the reconstruction in the time domain, zoomed in on the pulse. The middle and right columns show the amplitude spectrum in a linear and fully logarithmic plot, respectively. Expected RFI lines and reconstructed narrowband RFI are also shown (the reconstructed RFI is best seen in log-log coordinates, because of its intended small magnitudes)

(a) IFT   **(b)** Standard
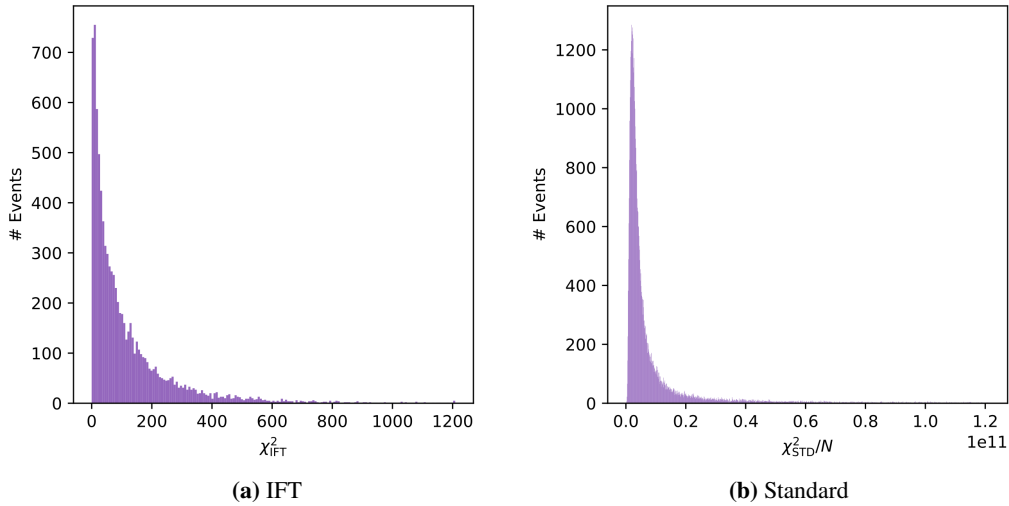
Figure 2: χ^2/N distribution for the comparison between simulations and reconstruction for the IFT method presented in this work (a) and the standard method of unfolding (b). Note that the axis for the standard method ranges to $1.2 \cdot 10^{11}$. The mean of the distribution for the IFT method is 163, the mode is 6.5 ± 0.9. For the standard method, the mean is $1.8 \cdot 10^{10}$ and the mode is $1.7 \cdot 10^9$.
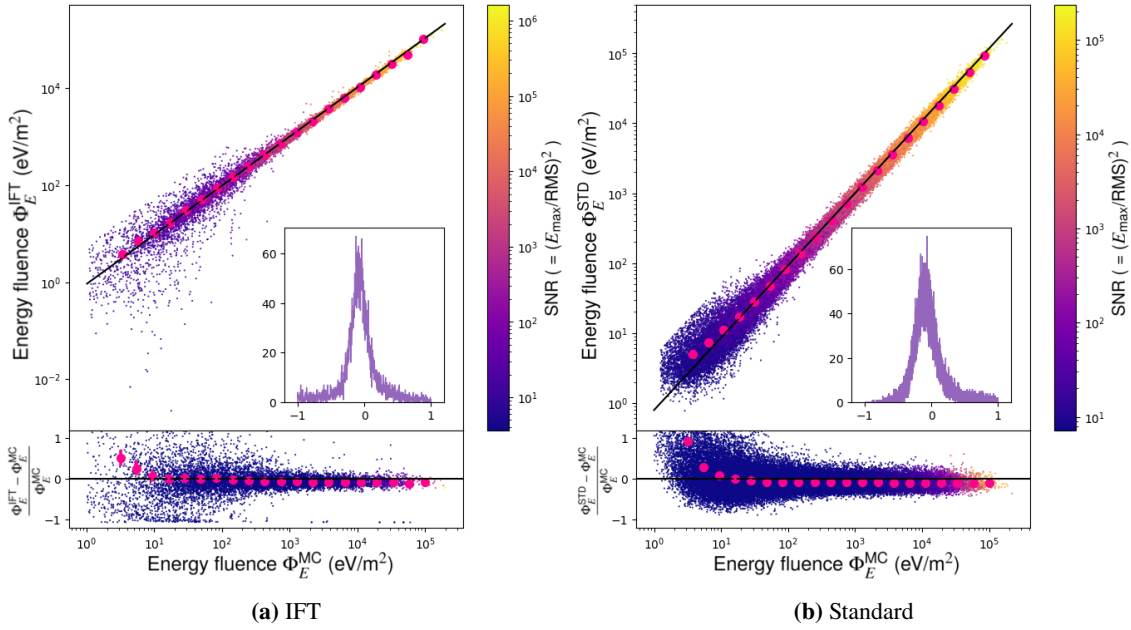
Figure 3: Comparison of the reconstructed to the simulated energy fluence for the IFT and the standard reconstruction method. The inset shows the distribution of the residuals. The profile dots and errors bars (mostly invisible) show the mean and standard deviation of the datapoints.

can be seen in figure 2. For the standard method, the naive assumption has been made, that the sample standard deviation of the noise can be used as an approximate error on the electric field reconstruction. From the distributions, it is immediately obvious, that the IFT-based reconstruction outperforms the standard reconstruction by a lot. This is easily explained by the fact that the new method can deal with measurement noise, while the old method simply propagates this noise to the electric field. However, the statistics of the new method (the mean of the distribution is ≈ 163, the mode is ≈ 6.5) indicate that the errors are still not well understood and underestimated. The algorithm used in reconstruction (geoVI [4]), although being able to handle non-Gaussian posterior distributions, is still only an approximation and can underestimate the variance if the posterior is unfavourably shaped. Especially in cases like this, where the noise model is simplistic and underestimates the tails of the noise distribution. That said, it is still a massive improvement compared to the old method.

4.2 Reconstruction of the energy fluence

A comparison of the reconstructed energy fluence (the amount of energy deposited per unit area, proportional to the time-integral of the squared electric field) can be seen in figure 3. Both methods of handling the noise seem to be working equally well, at least for large signals. For smaller signals, the IFT method introduces slightly less bias than the standard method. This is easily explained by the reduced influence of noise in the IFT method. While the total measured power becomes dominated by noise at some point for the standard method, the method presented here does either fail to perform inference, or find the signal pulse without noise. Still, there is no large benefit of performing the computationally heavy IFT reconstruction if the further analysis is

based on the energy fluence alone. The radio detector stations for which the calculation is less biased have signals so small that they would not be used in a reconstruction, and even if they were used, their contribution to the reconstruction would be negligible.

5. Conclusion

A forward model for the radio emission of extensive air showers, based on our current understanding of emission processes, has been developed and successfully implemented in an Information Field Theory based reconstruction algorithm to infer the electric field from noisy measurements. The reconstruction method has been tested on Monte-Carlo simulations and compared to the standard unfolding method. Based on the χ^2/N distributions, it has been shown, that the reconstruction of the electric field is greatly improved. When comparing the energy fluence, a quantity which is the basis of most further analysis methods currently in use, the IFT-based method does not offer a great improvement except for antennas with very small signals. This means, that current analysis methods can not benefit from the improved e-field reconstruction. However, there are further reconstruction methods under development, which depend on the pulse shape rather than the power alone (see e.g. [12]), which are relatively sensitive to noise and could be used to better effect with the electric field reconstruction detailed in this work. In the future, this method shall be developed further by using a single event-level model of the electric field, reconstructing the signal from all antennas at once.

Acknowledgments

References

[1] T. Huege, *Physics Reports* **620** (2016) 1.

[2] T. Enßlin in *AIP Conference Proceedings (MaxEnt 2023)*, vol. 1553.1, pp. 184–191, 08, 2013, DOI.

[3] J. Knollmüller and T.A. Enßlin, 1901.11033.

[4] P. Frank, R. Leike and T.A. Enßlin, *Entropy* **23** (2021) .

[5] C. Welling, P. Frank, T. Enßlin and A. Nelles, *JCAP* (2021) .

[6] S. Martinelli, F. Schlüter and T. Huege in *PoS(ARENA2022)*, vol. 424, p. 036, 2023, DOI.

[7] T. Huege, M. Ludwig and C.W. James in *AIP Conf. Proc. (ARENA2012)*, vol. 1535, pp. 128–132, 05, 2013, DOI.

[8] Huege, Tim for the Pierre Auger collaboration in *EPJ Web Conf. (UHECR 2022)*, vol. 283, p. 06002, 2023, DOI.

[9] The Pierre Auger collaboration, *Nucl Inst A* **635** (2011) 92.

[10] G. Edenhofer, P. Frank, J. Roth, R.H. Leike et al., *Journal of Open Source Software* **9** (2024) 6593.

[11] F.A. Schlüter, Ph.D. thesis, Karlsruher Institut für Technologie (KIT), 2022. 10.5445/IR/1000149113.

[12] N. Karastathis, S. Buitink, A. Corstanje, M. Desmet et al. in *PoS(ICRC2023)*, vol. 444, p. 487, 2023, DOI.