An upgrade to the NuRadioReco IFT electric field reconstruction

Master's Thesis

Presented by

Paul Lehmann

27.10.2023

Physikalisches Institut
Erlangen Centre for Astroparticle Physics
Friedrich-Alexander-Universität Erlangen-Nürnberg

Supervisor: Prof. Dr. Anna Nelles

Contents

1 Introduction

The existence of neutrinos was proposed by Wolfgang Pauli in 1930 to explain how the beta decay might comply with the conservation of energy, momentum, and spin. Finally detected in 1956, these fermions carry no electric charge and a mass of less than 10×10^{-36} kg - hence also their name, meaning "small neutral one" - and posses a property called flavor state.

Furthermore, neutrinos interact purely via weak force interactions, because they carry no electric charge. Due to this, neutrinos have very small interaction cross-sections compared to other particles such as electrons. This means that an astrophysical neutrino traveling through the universe is not deflected by magnetic fields and very likely does not interact with intervening matter. Thus, neutrino trajectories point back towards their origins, and their flavor and information can contain information about their sources.

This remains true even at high energies and ultra-high energies. Potential sources of cosmic neutrinos are astrophysical objects such as Gamma Ray Bursts, Blazars, and supernovae, [3] [4] and interactions between ultra-high energy cosmic rays and the cosmic photon background [5] [6]. Some astrophysical objects are expected to generate neutrinos with energies up to the EeV scale. Such ultra-high energy neutrinos might not only provide information about astrophysical processes but also be used in beyond-Standard-Model and dark matter searches[7]. All this makes neutrinos a highly promising probe.

However, the low interaction cross-sections also represent a big challenge for neutrino astronomy, as a detector must interact with a neutrino to detect it. The very low cross sections combined with declining flux at increasing energies means that large effective detector volumes are needed in order to collect enough event statistics. To overcome this, neutrino detector design often aims to maximize interaction rates or instrumented volume, by observing water[8] [9] or ice [10] for signals created by neutrino interactions. By detecting Cherenkov light generated by neutrino interactions, the first high energy neutrino sources have already been discovered [11] [12].

However, due to their short propagation lengths [13] and scattering [7] of optical light respectively, water and ice limit the spacing of detectors observing light in the visible spectrum to some 100 m, since with a larger spacing events could not be properly reconstructed. That, in turn, makes increasing the detector volume impractical due to the costs and efforts required. This makes the optical spectrum unfeasible to observe neutrinos above 100 TeV, where even larger detector volumes are required.

The solution to this is to observe radio signals from neutrino interactions instead. These signals are created by the particle shower caused by the neutrino via the Askaryan effect, and can travel up to a kilometer in ice[14][15]. The initial energy and flavor of the neutrino can be reconstructed from the signal properties.

This is the approach followed by the Radio Neutrino Observatory in Greenland RNO-G[16], which is currently under construction in the Greenland ice shelf. RNO-G aims to detect neutrinos ranging between several PeV up to EeV energies, while observing a volume of circa $100 \, \mathrm{km}^2$.The detector consists of a grid of radio antenna arrays embedded in the Greenland ice shelf.

To gain insights into the properties of the initial neutrino from the radio waveforms recorded by these antennas, sophisticated event reconstruction methods are necessary. As no in-ice radio signature of an neutrino has been identified up to this point, such a reconstruction should ideally rely on few theoretical assumptions about the exact shape of the signal. Performing this reconstruction without strong assumptions is possible via Information Field Theory[17][18], which provides statistical methods for the signal recovery. A computational implementation of IFT exists in the `NIFTy` [19] software package. A reconstruction module utilizing `NIFTy` for for the reconstruction of the electric fields of the neutrino radio signals exists as part of the `NuRadioReco` [20] [21] project.

This thesis covers an attempt at expanding this existing program:

Currently, one electric field is reconstructed for all antennae observing the signal. A more differentiated approach would allow each group of antennae to reconstruct distinct but similar electric fields. This might for example make it possible to become sensitive to the neutrino flavour. This thesis covers an attempt at expanding the existing program to the multi-field reconstruction. This reconstruction is tested using simulated data from `NuRadioMC`. [22]

In section 2, the processes of neutrino interaction, signal creation and propagation are discussed. In section 3, a description of the RNO-G detector hardware is given. In section 4, Information Field Theory and the mathematical foundation of the reconstruction algorithms are introduced. In section 5, an update of the pre-existing reconstruction is tested for viability as a basis for the work in this thesis. In section 6, the additions to the existing program are presented. section 7 discusses diagnostics performed as part of development and presents results of the expanded software. In section 8, a summary of the work performed and an assessment of the results are given.

2   Physical background

In this chapter, the interaction between neutrinos and ice, the resulting signal and its propagation to the detector are discussed to motivate and contextualize the subject of this work.

2.1   Neutrino sources

The RNO-G detector is expected to be sensitive to neutrinos of $10\,\text{PeV}$ and above. Two categories of processes are predicted to generate the neutrino flux at these energies: hadronic interactions with matter or photons near accelerators of Ultra-High Energy Cosmic Rays (UHECR), and interaction of UHECRs with background photon fields:

$$p + \text{nucleus} \rightarrow \begin{cases} \pi^0 + X \\ \pi^\pm + X \end{cases} \tag{1}$$

$$p + \gamma \rightarrow \Delta^+ \rightarrow \begin{cases} \pi^0 + p \\ \pi^+ + n \end{cases} \tag{2}$$

Both of these processes produce pions, which decay via

$$\pi^0 \rightarrow \gamma\gamma \tag{3}$$

$$\pi^\pm \rightarrow \begin{cases} \mu^+ + \nu_\mu \text{ and } \mu^+ \rightarrow e^+ + \overline{\nu_\mu} + \nu_e \\ \mu^- + \overline{\nu}_\mu \text{ and } \mu^- \rightarrow e^- + \overline{\nu_e} + \nu_\mu \,. \end{cases} \tag{4}$$

Therefore, a flavor ration of $2:1:0$ $\nu_e : \nu_\mu : \nu_\tau$ is expected at the neutrino origin. However, neutrino oscillations the ratio will have changed to $1:1:1$ by the time the neutrinos reach Earth. [23]

2.2   In-ice neutrino interaction

At energies above $40\,\text{TeV}$, the earth becomes opaque for neutrinos since the interaction cross-section increases with energy. Therefore, signals are expected mainly from downgoing or skimming neutrinos.[24]

High energy neutrinos above circa $10\,\text{GeV}$ interact with matter mainly via deep inelastic scattering off a quark within a nucleus. These interactions can be split into two categories: charged current interactions (CC) and neutral current interactions (NC) interaction[25]. Figure 1 depicts the relevant Feynman diagrams.
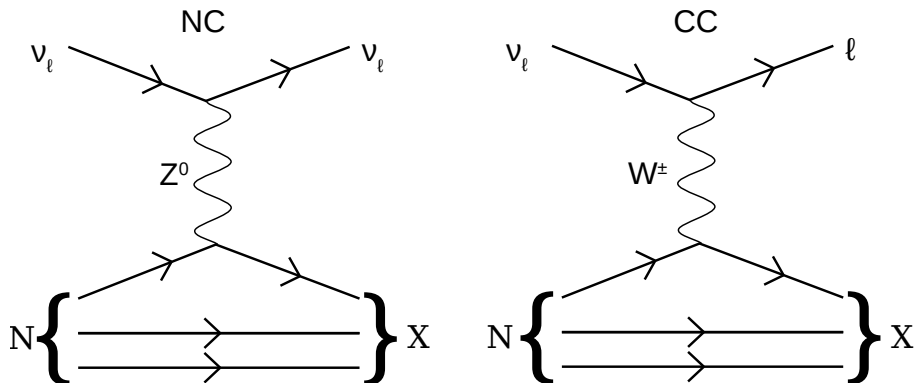
Figure 1: *Feynman diagram of deep inelastic scattering between a neutrino and a nucleus N for neutral current and charged current interaction. X represents the accelerated nucleus. ℓ represents any one lepton flavour.*

In a neutral current interaction, the neutrino and the nucleon interact by exchanging a $W^\pm$ boson, and the outgoing lepton is a neutrino once again. Kinetic energy transferred to the quark then causes the nucleus to produce secondary particles, which kicks off a hadronic cascade composed of a majority of pions. Produced π^0 quickly decay into photons and then electron-positron pairs, which then initiate secondary electromagnetic showers. The neutrino is still present in the final state, and its flavor does not change, and still possesses some kinetic energy. The fraction of energy transferred by the neutrino to the nucleus and thus the shower is called *inelasticity y* of the interaction. As the distribution of y is skewed towards low values, this creates an unresolvable uncertainty on reconstruction of the initial neutrino energy[26].

$$\nu_\ell\, N \to \nu_\ell\, X \qquad\qquad \overline{\nu}_\ell\, N \to \overline{\nu}_\ell\, X$$

In a charged current interaction, the exchange particle is a Z^0 boson, the final state consists of a hadronic system and a charged lepton, the flavor of which corresponds to the initial neutrino.

$$\nu_\ell\, N \to \ell\, X \qquad\qquad \overline{\nu}_\ell\, N \to \overline{\ell}\, X$$

If an electron is produced by a CC interaction, it generates an electromagnetic shower close to the hadronic shower originating at the nucleus. This means that the entire initial energy is ultimately deposited into measur-

able showers, which makes these type of events prime candidates for energy measurements, if they can be distinguished from a hadronic shower. [27]

A CC interaction can also produce a μ or τ instead of an electron. The heavier leptons continue through the medium until their decay, potentially producing secondary cascades, which in turn can emit detectable radio signals. The timing and position of these showers can serve as an identifying feature for the incoming neutrino's flavor. [28]

Because particle production mostly preserves the total electric charge of the system, the net charge of a particle shower developing in vacuum would remain close to zero. Such a neutral shower would not emit any coherent radio signal. A shower becomes electrically charged only due to interaction with external factors, such as the medium through which it travels.

When positive and negative charges present in the shower travel through a magnetic field such as the earth's magnetic field, they are separated by the Lorentz force. This polarization then causes radio emission. Additionally, as the shower progresses through the medium, positrons in the shower front annihilate with surrounding matter, while electrons from the medium are accelerated via Compton scattering off shower particles. These electrons get swept along by the shower, which leads to the shower developing a net negative charge relative to the surrounding ice. This effect is called Askaryan effect.

Due to the density of ice, particle cascades in ice are relatively small with a size in the order of ~10 m, which limits the contribution of the geomagnetic effect. The Askaryan effect is thus dominant, and the radio signal is therefore also called Askaryan radiation. [28]

As the shower particles travel faster than the speed of light in the medium, and close to the Cherenkov angle (circa 56° in ice) the emitted radiation is coherent over all wavelengths. The signal itself has a duration of a few ns, ranges from MHz to GHz, and is linearly polarized.[22] The coherency on the Cherenkov cone allows for constructive interference of the signal with itself, which is crucial to reach a signal strength that is detectable. See Figure 2 for a schematic of the overall process.

When deviating from the Cherenkov angle, high frequencies lose coherence first. Low frequencies retain their coherency long enough, so that a signal is recoverable even a few degrees off the Cherenkov cone. An exemplary reconstruction of signals at different opening angles but identical distance is given in Figure 3. Due to this angular dependence, reconstructing the signal spectra accurately is important for a correct energy reconstruction, while the reconstruction of the electric field in turn requires an adequate understanding
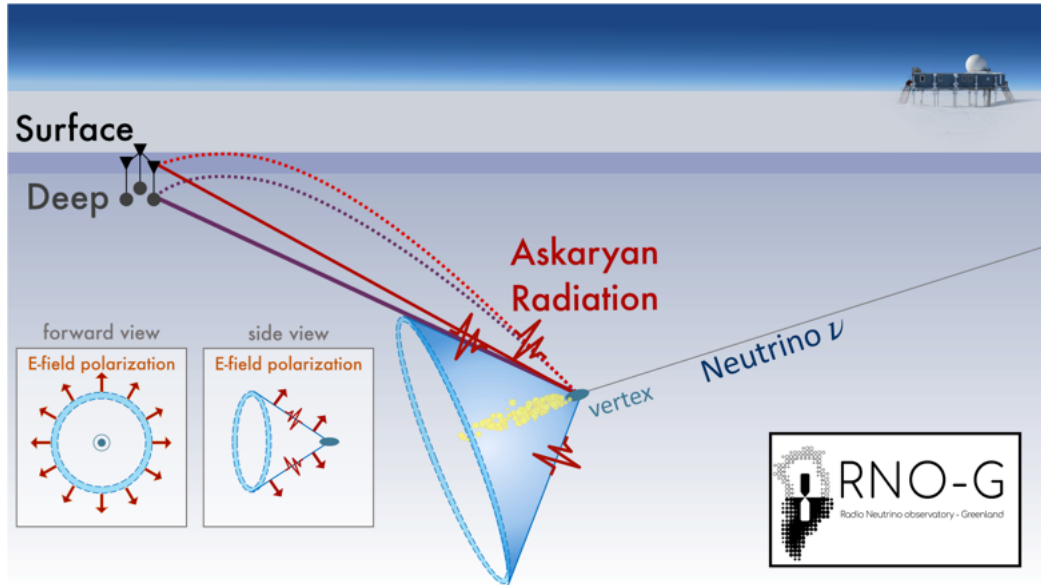
Figure 2: *Exemplary illustration of a particle shower creating a radio signal (not to scale). The neutrino (dark blue) interacts and causes a particle shower (yellow dots). These emit radiation (purple, red), which is strongest on the Cherenkov cone (light blue). This radiation then propagates straight or indirectly to the detector (black). From [16]*

of event geometry.

The form of the spectra is influenced by another factor: while the hadronic shower has a relatively simple shape, electromagnetic cascades can produce a more complex frequency distribution, see Figure 4. This is caused by the LPM effect. This effect is named after physicists Lev Landau, Isaak Pomeranchuk, who discovered it in 1963, and Arkady Migdal who developed on in it in 1956. As high energy electrons travel through a dense medium, it is subject to multiple small scattering interactions. If the scatterings occur close enough to each other, quantum interference between between the events will suppress cross sections for bremsstrahlung and pair-production, thus greatly increasing the attenuation length of the shower from ~10 m to over 100 m. The radiation emitted from this elongated shower will show multiple, smaller maxima, and interference with the primary hadronic signal will further increase the complexity of the spectrum.

2.3 Signal propagation

The radio signals in ice have an attenuation length of up to 1 km, depending on ice temperature and depth, with the attenuation length being the distance
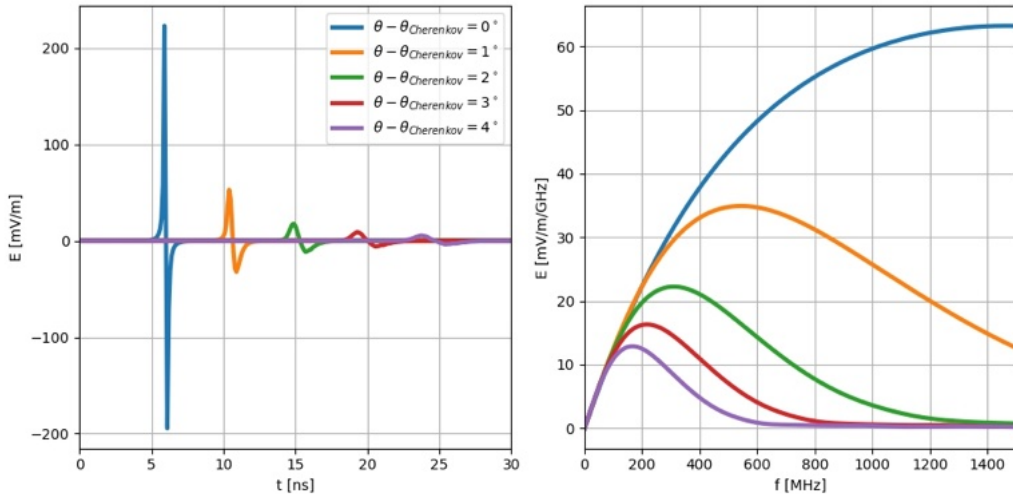
Figure 3: *Electric field waveform (left) and spectra (right) of an Askaryan signal dependent on deviation from the Cherenkov cone, as created by a hadronic cascade with $1\,\mathrm{EeV}$ deposited into it, without detector or propagation effects. Waveforms are offset in time to improve legibility. From [16]*

after which the signal strength has decreased to 1/e of its initial value. However, propagation of the radio signal through ice exhibits more complex behavior than propagation through e.g. air. Total loss of field amplitude is described by the formula

$$E = E_0 \cdot \exp\left(-\frac{l}{l_{\mathrm{att}}}\right) \cdot \frac{l_{\mathrm{ref}}}{l}\,, \tag{5}$$

with E_0 the initial field strength, l the distance traveled, l_{att} the attenuation length, and l_{ref} the reference distance.

Firstly, the refraction index of the ice is not uniform with depth. The top one hundred or more meters of ice are the area where snow is gradually compressed into ice through the weight of overlying snow, the firn layer. The refraction index of the ice is roughly linear with its density, leading to gradient ranging from $n \approx 1.35$ under the surface to $n \approx 1.78$ for bulk ice. Furthermore, this gradient is not perfectly smooth, but can exhibit a layered structure due climatic variation over the course of a year. The refraction index depending on depth can be approximated with

$$n(z) = n_{\mathrm{ice}} - \Delta n \cdot \exp\left(\frac{z}{z_0}\right)\,, \tag{6}$$

with $\Delta n = n_{\mathrm{ice}} - n_{\mathrm{snow}}$, and z_0 the inverse product of gravitational acceleration, snow density, and snow compressibility. z_0 in general has to be
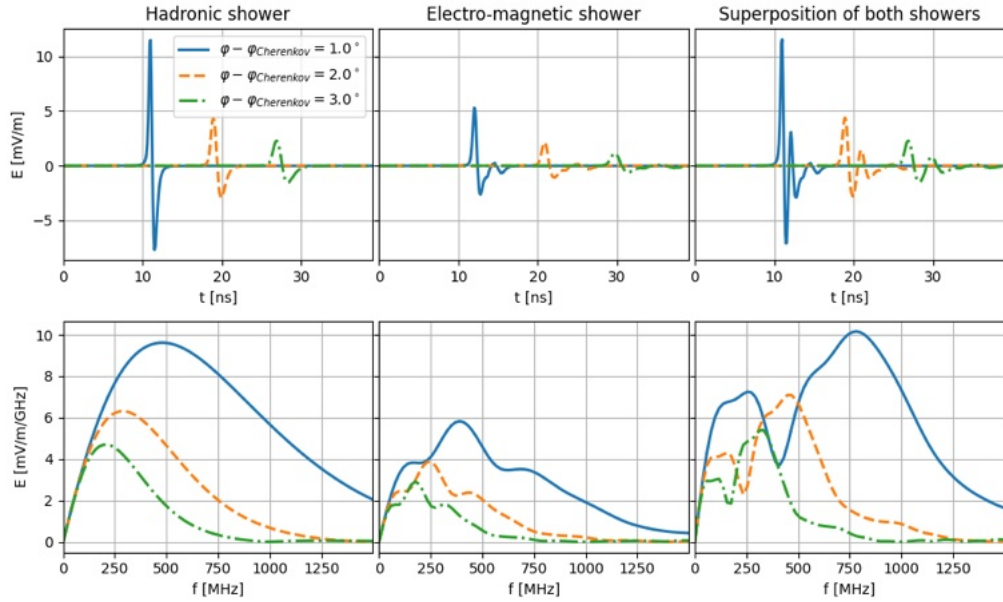
Figure 4: *Examples of radio waveforms (top) and spectra (bottom) from a 1 EeV particle shower, at 2 km distance, and for three different viewing angles (colors). From left to right: the hadronic shower signal, the electro-magnetic shower signal, and both signals superposed. The pulses for different angles or offset in time for better readability. From [29]*

determined via measurement and varies by at least single digit factors between different locations. [30]

Due to Fermat's principle, this leads to the trajectories of radio signals bending downwards towards the denser medium. This creates a "shadow zone": any signal created in this zone will be diffracted away from the antenna entirely. [31]

Secondly, this diffraction and deflection off the ice surface can offer a second trajectory along which a signal from an event might reach the detector, in addition to a direct trajectory. The two signals need not be identical, as they are emitted under different angles and thus can exhibit very different spectra. Indeed, often only one of the two signals will be above detection threshold, but both can be detected if the detector is close to the surface, or the event is far enough away, such that the viewing angle is sufficiently small.

A third effect possibly altering propagation of the signal is birefringence. This effect refers to the refractional index depending on polarization, and in ice, may result from a difference in the refraction indices along the vertical and horizontal direction, caused by gravitational effects and the horizontal shifting of the ice called ice flow. This produces a delay between the arrival

of horizontally polarized and vertically polarized electric fields, thus splitting the signal in two. This poses an additional challenge for trigger selection, due to the two pulses containing only part of the total signal strength. On the other hand, the timing can also yield information about the travel time of the signal, if precise information of local ice conditions are available. [32] Due to the relatively slow speed of ice flow at the detector location, however, birefringence effects are expected to not be strongly pronounced. [33] [34]

3   The RNO-G detector

The enhancements to the event reconstruction attempted in this work are strongly motivated by the geometry of the RNO-G antenna arrays. In this chapter, the detector is presented.

The Radio Neutrino Observatory in Greenland (RNO-G) is a radio neutrino detector being constructed at Summit station in Greenland since 2021. The project aims to detect neutrinos ranging from $10\,\text{PeV}$ up to the EeV range.

The completed detector is planned to consist of 35 detector stations arranged on a square grid with a spacing between station of $1.25\,\text{km}$, as shown on the left in Figure 5. This spacing was chosen to maximize detector volume, with the trade-off that most events will only be detected by a single detector station. Any reconstruction therefore operates under the constraint of using only data collected by a single station.

Each detector station consists of two parts: a shallow component consisting of radio antennae embedded in the snow layer a few meters under the surface, and a deep component, for which boreholes with a depth of $100\,\text{m}$ are drilled into the ice shelf, with the antennae deployed in these holes. The hardware for data acquisition and communication is located on the surface at the center of the station, and a number of solar panels there supply the station with electricity. A schematic of one station is presented on the right in Figure 5.

The shallow component is made up of logarithmic-periodic dipole antennae (LPDAs). They are deployed on the surface as their large size makes drilling boreholes that accommodate them impractical. Some of the antennae are angled upwards to detect air showers, while the rest are angled downwards to detect in-ice events. However, due to the refraction effects in the ice as described in subsection 2.3, it is predicted that only a small number of in-ice events will be visible to the shallow component.

The deep component is the component with which this work concerns itself. It consists of one so called string in each of the three boreholes. One of the strings is designated the Power String, the other two are called Helper Strings 1 and 2. The Power String has a more extensive instrumentation: At the bottom of the string, it holds a phased array of 4 vertically polarized (VPOL) antennae, and directly above it two horizontally polarized (HPOL) antennae. Further up the Power String, three more VPOL antennae are located at $80\,\text{m}$, $60\,\text{m}$, and $20\,\text{m}$ depth.

The arrangement of the Power String VPOL antennae as a phased array greatly reduces the signal-to-noise ratio on which the detector can trigger, down to values as low as 2. This is achieved by using interferometry to
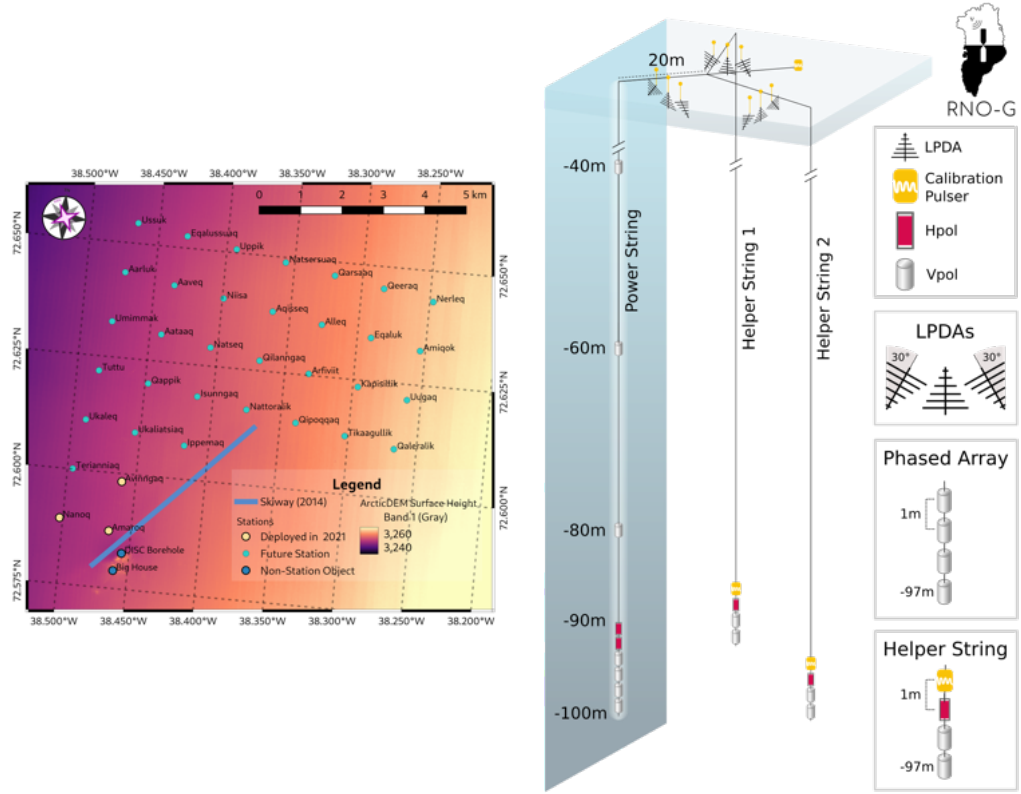
Figure 5: *Left: Map of the detector layout of RNO-G as of June 2022. Yellow dots are deployed stations, teal dots are stations which are not yet installed. The background color gradient indicates elevation. From [35] Right: design of one detector station. Grey cylinders represent VPOL antennae, red cylinders represent HPOL antennae. From [16]*

combine the signals of the four VPOL antennae, a technique that has been already been tested successfully in other detectors.

The helper strings respectively posses two VPOLs and one HPOL antenna towards the bottom of the string, as well as a radio emitter for calibration purposes.

RNO-G uses fat dipole style antenna as VPOLs, each of which are cylindrical with a height of 60 cm and a diameter of circa 13 cm. As shown in Figure 6, they are sensitive in a range from 50 MHz to circa 600 MHz, and their response is symmetrical in the azimuth.

The HPOLs are quadslot antennae with a height of 60 cm and a diameter of 20 cm. Their diameter is limited strongly by the size of the borehole, which restricts their sensitivity to a frequency band of about 200 MHz to
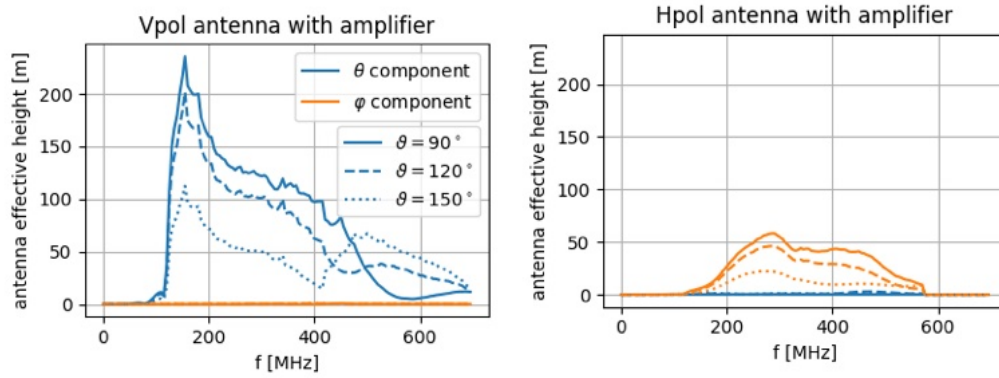
Figure 6: *Response of a VPOL antenna (left) and a HPOL (right) antenna. Blue are azimuthal contributions, orange are polar contributions. From [29].*

approximately 400 MHz. The HPOLs also exhibit a weaker overall response than the VPOLs, again due to the size limits imposed.

As Askaryan emissions from neutrinos are typically stronger in the vertical polarization as well, the VPOL antennae will likely be the more important observers for most signals. [16] [27]

4 Information Field Theory and NIFTy

In this chapter, Information Field Theory and its implementation in the NIFTy software package are given a short introduction to provide context for the mathematical approach of the reconstruction and its computational implementation.

4.1 Information Field Theory

The electric field which is to be reconstructed is a function over a continuous space. As such, it is defined at every point in that space, and thus infinite dimensional. As any data gathered about the electric field, and any data in general, is necessarily finite-dimensional, inference of the field from the data is generally an under-constrained problem. Therefore, additional information about the field is necessary to be able to make a meaningful selection from the infinite number of possible field realizations.

This additional information can be provided by properties of the field such as symmetry or physical restrictions. Commonly, it takes the form of knowledge about the self-correlation of the field being investigated. Indeed, most physical fields can expected to be "smooth", as jagged field realizations with steep differentials between points in close proximity are often suppressed, as they bring with them a high potential difference which is either outright unable to form, or quickly evened out as the field evolves over time. This smoothness can be defined as the correlation between two points of the field

$$\Phi(x, y) = \langle \varphi(x)\varphi(y) \rangle_{(\varphi)}, \tag{7}$$

at coordinates x and y for a field with vanishing expectation value $\langle \varphi \rangle_{\langle \varphi \rangle} = 0$. These correlations establish a relation between all pairs of points of the field, thus generating an infinite amount of equations which can be exploited to put sufficient restraints on the field to make statistically meaningful statements about the field realization.

To bring together this additional information with the measured data to infer about the field, Bayes' theorem is used. Consider a measurement equation

$$d = R(\varphi) + n \tag{8}$$

where d is measured data, φ is a field, $R(\varphi)$ is a response function of the instrument used to observe the field, and $n = d - R(\varphi)$ is the noise, a stochastic contribution to the observed data with no relation to the actual field. From

this, the likelihood to measure the data d given the field realization φ can be expressed as

$$
\begin{aligned}
\mathcal{P}(d|\varphi) &= \int \mathcal{D}n\, \delta(d - R(\varphi) - n)\, \mathcal{P}(n|\varphi) \\
&= \mathcal{P}(n = d - R(\varphi)|\varphi)\,,
\end{aligned}
\tag{9}
$$

where $\int \mathcal{D}n$ denotes the path integral over n, and $\mathcal{P}(n|\varphi)$ is the noise statistic, i.e. the probability density for a noise value given a certain field. This likelihood represents all information gained via measurement.

As the likelihood is limited by the data and thus necessarily finite dimensional, the additional information of the field structure have to be included. This allows formulating the posterior probability density $\mathcal{P}(\varphi|d)$, the probability density of a field under the condition of the specific data that was observed. [17] [18]

Bayes' theorem now connects this posterior, which is the information about the field gained through measurement, with the prior, the pre-existing knowledge about the field. Bayes' theorem states

$$
\mathcal{P}(\varphi|d) = \frac{\mathcal{P}(d|\varphi)\,\mathcal{P}(\varphi)}{\mathcal{P}(d)}\,,
\tag{10}
$$

with the evidence or partition function

$$
\mathcal{P}(d) = \int \mathcal{D}\varphi\, \mathcal{P}(d|\varphi)\, \mathcal{P}(d) = \mathcal{Z}(d)\,,
\tag{11}
$$

which normalizes the product of prior and likelihood to obtain some data given a field. Now, by defining an information Hamiltonian

$$
\mathcal{H} = -\ln \mathcal{P}(d|\varphi)
\tag{12}
$$

and with the partition function from Equation 11, Bayes' theorem in Equation 10 can be rewritten as the signal field posterior

$$
\mathcal{P}(\varphi|d) = \frac{\exp(-\mathcal{H}(d,\varphi))}{\mathcal{Z}(d)}.
\tag{13}
$$

The information Hamiltonians, which are also called information or surprises, have the very useful property of being additive over multiple independent measurements. For a set of data $d = (d_1, d_2)$, it holds that [17] [18]

$$
\mathcal{H}(d|\varphi) = \mathcal{H}(d_1|\varphi) + \mathcal{H}(d_2|\varphi).
\tag{14}
$$

This rewriting of Bayes' theorem now enables the usage of various methods of field theories to construct statistical statements, such as the Kullback-Leibler convergence, or KL convergence

$$\mathrm{KL}[\mathcal{P}'_\eta(\varphi)\|\mathcal{P}(\varphi|d)] = \int D\varphi\, \mathcal{P}'_\eta(\varphi) \ln \frac{\mathcal{P}'_\eta(\varphi)}{\mathcal{P}(\varphi|d)}\,, \tag{15}$$

which is used to calculate the statistical distance between two probability distributions $\mathcal{P}'_\eta(\varphi)$ and $\mathcal{P}(\varphi|d)$. If $\mathcal{P}'_\eta(\varphi)$ is an approximation of the posterior $\mathcal{P}'_\eta(\varphi)$ with variational parameters η. [36].

4.2   Numerical Information Field Theory NIFTy

NIFTy (Numerical Information Field Theory) is a python library which implements the methods of IFT to facilitate the development of signal inference algorithms, while allowing these algorithms to remain independent of their chosen discretization. Furthermore, it has been designed such, that its code is explicitly similar to the mathematical notations they implement to allow it to be parsed easily.

The NIFTy package is centered around three classes, all of which are analogues to their mathematical namesakes [37] [38]:

- Spaces, which are grid-based approximations of some geometrical space. Harmonic transformations such as a Fourier transformation are possible, which allows exploitation of symmetries present in a system.

- Fields, which are always defined with a space object as domain and target, and store values for that space. Fields and their functions implicitly gain the geometric properties of their underlying space, such as the proper inner product.

- Operators, which also always possess a space as their domain and target. Other than that, they can contain any operation. Importantly, operations which can not be implemented explicitly can approximated via sample averaging.

NIFTy also contains pre-implemented inference algorithms. The method used in this work is based on Metric Gaussian Variational Inference. As complex posteriors can generally not be calculated analytically, this approach instead approximates a true posterior with a process to produce random variables with a zero-mean Gaussian distribution and dependence on some parameters.

This approximation allows calculation of the KL divergence in Equation 15. The location of the distribution parameters can then be shifted towards a lower KL divergence by finding the gradient of the parameters. [39] [29].

5   Testing the NIFTy8 upgrade

As of the writing of this thesis, the IFT reconstruction module of `NuRadioReco` is based on `NIFTy` 5. A migration to the currently latest version `NIFTy` 8 exists at [1]. This latter module was tested as a potential foundation for this work to build on.

The reconstruction was tested using a Monte Carlo generated data set for RNO-G, see Appendix A.

The running time of these reconstructions was considerably longer than expected, taking over 4 hours to reconstruct just a single event purely for the phased array and the two power string HPOL antennae on a commercial notebook. This is over 200 time longer than the than the time required by the `NIFTy` 5 module to do the same. This points towards the reconstruction not converging properly.

The reconstructed electric field spectra in Fourier space of one neutrino event are shown in Figure 7, the corresponding voltage traces in the time domain are shown in Figure 8. Each channel corresponds to one antenna. In the spectra, it can be seen that especially channels 0 and 1 do deviate from the MC truth. Additionally, gray lines of the minimization samples still show a large variance in all channels. These are stochastically generated from the posterior model, the mean of which is the reconstruction, and can help visualize the variance remaining of the reconstructed result. This high spread of samples implies a weak convergence.

The pulse reconstruction in the time domain shows very little agreement with the MC truth at all. The general timing of the pulse is identified, but the reconstruction seems phase shifted with regards to the true pulse timing, which might indicate a problem with the phase reconstruction.

Reconstruction was then tested with varied parameters for the power spectrum prior generating the E-field amplitude. In `NIFTy` 8, the shape of the power spectrum prior is defined primarily by these parameters and their respective variances:

- `offset`, which sets the intercept with the y-axis.

- `fluctuations`, which defines the amplitude of field fluctuations.

- `loglogavgslope`, which is the exponent of the power law spectrum component of the prior.

- `flexibility`, which is the amplitude of the integrated Wiener process power spectrum component.

23

- `asperity`, which defines how ragged the generated spectra will be.

[37]

These parameters were varied individually and in combinations. Focus was set on the offset, the fluctuation, and the loglogavgslope, as these were deemed to have the highest potential to improve reconstruction. However, the reconstruction was very sensitive to even small changes to the prior. Even small variations cause the reconstruction to produce results not conforming to the shape of the E-field at all, and instead reconstruct a sawtooth-like spectrum instead. A greater robustness of the reconstruction towards small changes in prior is expected, this volatility indicates some problem of the model used for the reconstruction.

As diagnosing the problems with the `NIFTy` 8 based module was deemed to be beyond the scope of this thesis, the decision was made to build upon a version of `NuRadioReco` using `NIFTy` 5 for the rest of the work instead.
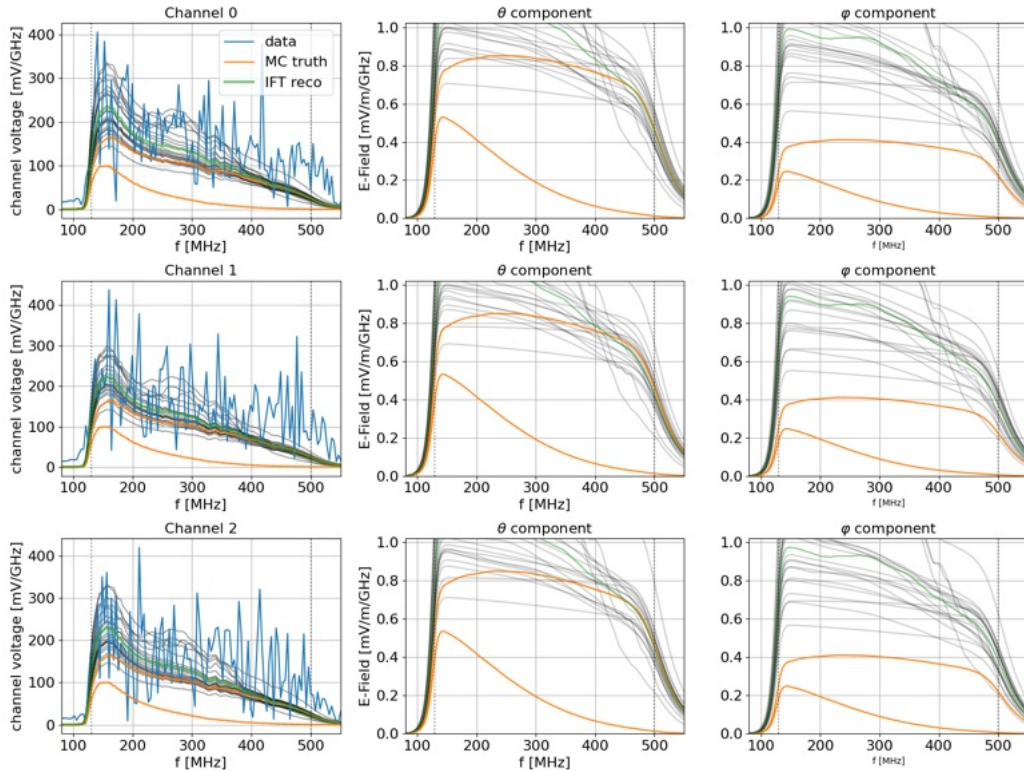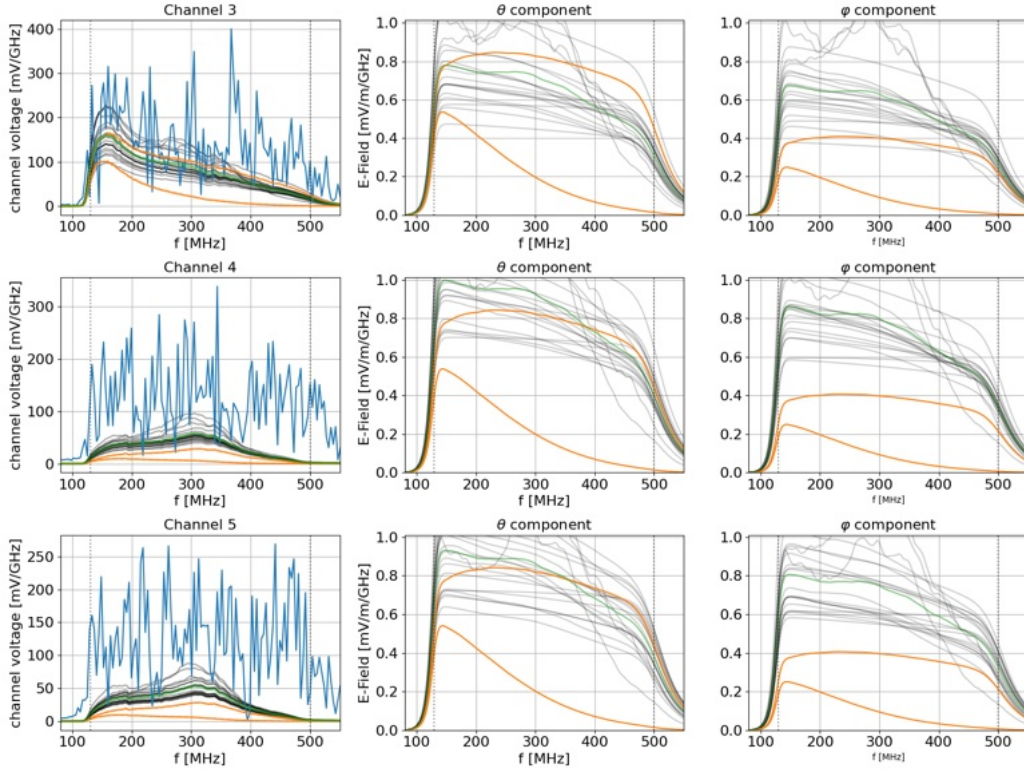

Figure 7

Figure 7: *(Continued) Example of an E-field reconstruction for one channel using* `NIFTy8`*, without any new functionality. From left to right, the plots are the the channel spectra in Fourier space, the θ component of the E-field in Fourier space, and the φ component of the E-field in Fourier space. Orange lines are the MC truth, blue lines are simulated data with noise. Green lines are the reconstructed electric fields. Gray lines are samples from the KL minimization, and provide insight into how the variance of the reconstruction.*
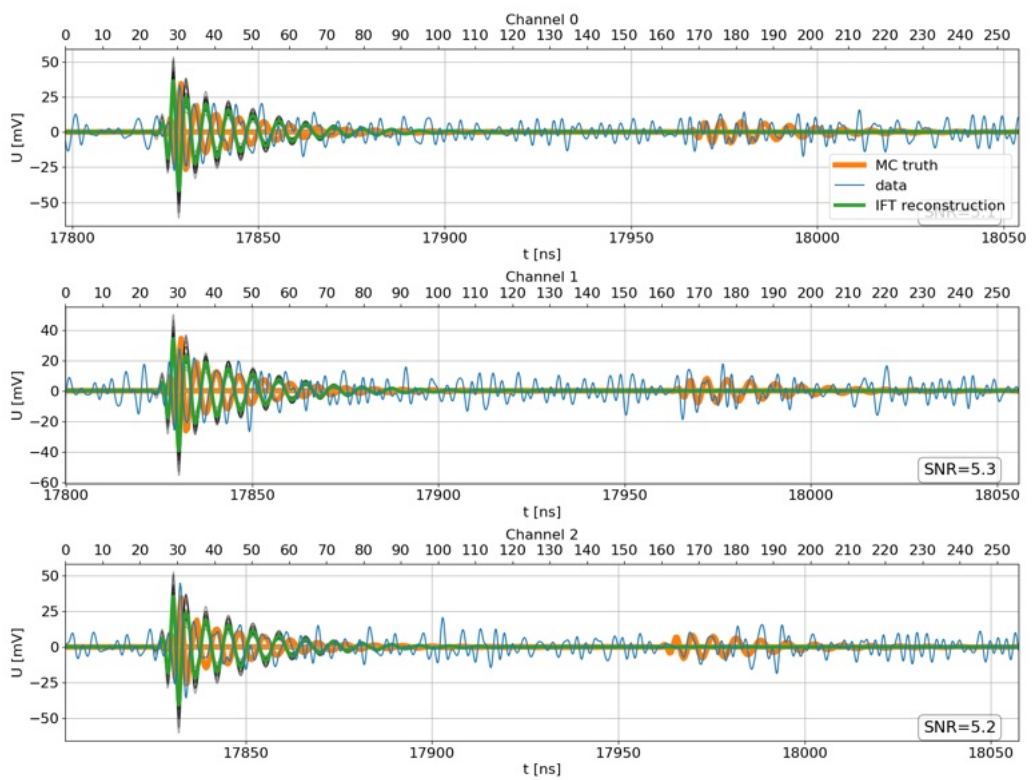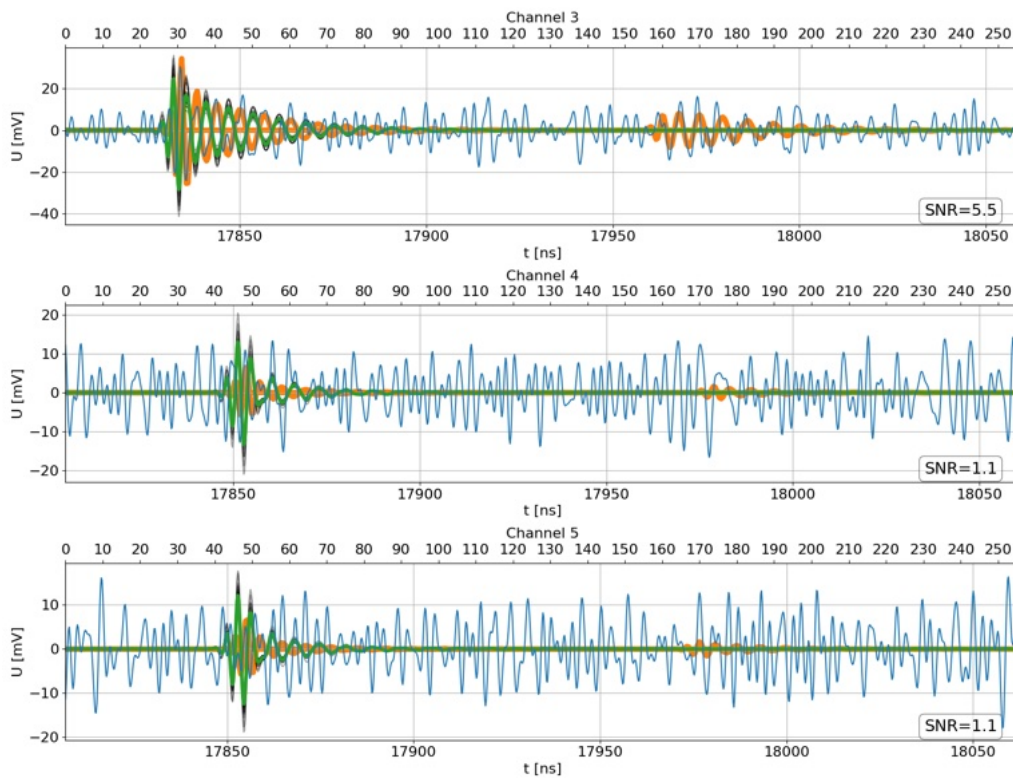
Figure 8

Figure 8: *(Continued) Example of a reconstruction for one channel using* `NIFTy8`*, without any new functionality. Shown are the channel voltage traces in the over time. Orange lines are the MC generated true signal, blue lines are the simulated data with noise. Green lines are the reconstructed electric fields. Grey lines are iterative samples for the reconstruction algorithm, and provide some insight into how accurately the reconstruction converges.*

27

6   Expanded reconstruction model

This chapter presents the enhancements made to the pre-existing reconstruction program. A branch using NIFTy 5 was used as foundation for the upgrade. The first part of this chapter explains the concept of the new reconstruction, and the second part details its technical implementation.

The code developed as part of this thesis can be found on GitHub at [2].

6.1   Improved reconstruction concept

Figure 9: *Channel numbering and group arrangement used in this work. Each black box represents one antenna, vertical boxes are VPOLs, horizontal boxes are HPOLs. Each gray shaded box encloses one channel group. Calibration VPOLs are not considered in this work. Note the greater distance between the three uppermost channels on the power string.*

The previous reconstruction module models the electric field as an amplitude and a phase

$$\mathcal{E}_n(f) = (E_0(f)) \cdot \exp(\mathrm{i} \cdot \varphi_n(f)), \tag{16}$$

where $\mathcal{E}_n$ is the complete electric field for the channel n, E_0 is the electric field amplitude in frequency space observed by all channels, and φ_n is the phase specific to channel n.

For the pre-existing reconstruction, the assumption is made that the shape and thus the amplitude of the electric field is identical for all channels, while each channel's phase is reconstructed individually. This assumption holds well for channels which are close together, such as the power string group which forms a chain with a $1\,\mathrm{m}$ spacing. However, for antennae with greater distance between them, the shape of the E-field can very well differ, for example because of different angles to the Cherenkov cone. To account for these possible differences, the reconstruction is expanded to allow for variation in the E-field amplitude between distant antennae. The antenna are organized into groups based on proximity as shown in Figure 9. The model for each group except the first group is extended by respectively one single E-field-like element that is shared among the group.

Expressed as an equation, the model from which the likelihoods are derived is modified from

$$\begin{aligned}
\mathcal{E}_0(f) &= (E_0(f)) \cdot \exp(\mathrm{i} \cdot \varphi_0(f)) \\
\mathcal{E}_1(f) &= (E_0(f)) \cdot \exp(\mathrm{i} \cdot \varphi_1(f)) \\
\mathcal{E}_2(f) &= (E_0(f)) \cdot \exp(\mathrm{i} \cdot \varphi_2(f)) \\
&\quad \cdots \\
\mathcal{E}_{23}(f) &= (E_0(f) \cdot \exp(\mathrm{i} \cdot \varphi_{23}(f))
\end{aligned} \tag{17}$$

to

$$
\begin{aligned}
\text{group } 0 &\left[\begin{array}{ll}
\mathcal{E}_0(f) = & (E_0(f)) \cdot \exp(\mathrm{i} \cdot \varphi_0(f)) \\
\mathcal{E}_1(f) = & (E_0(f)) \cdot \exp(\mathrm{i} \cdot \varphi_1(f)) \\
\ldots &
\end{array}\right. \\
\text{group } 1 &\left[\begin{array}{ll} \mathcal{E}_6(f) = & (E_0(f) + \Delta E_1(f)) \cdot \exp(\mathrm{i} \cdot \varphi_6(f)) \end{array}\right. \\
\text{group } 2 &\left[\begin{array}{ll} \mathcal{E}_7(f) = & (E_0(f) + \Delta E_2(f)) \cdot \exp(\mathrm{i} \cdot \varphi_7(f)) \end{array}\right. \\
\text{group } 3 &\left[\begin{array}{ll} \mathcal{E}_8(f) = & (E_0(f) + \Delta E_3(f)) \cdot \exp(\mathrm{i} \cdot \varphi_8(f)) \end{array}\right. \\
\text{group } 4 &\left[\begin{array}{ll} \mathcal{E}_9(f) = & (E_0(f) + \Delta E_4(f)) \cdot \exp(\mathrm{i} \cdot \varphi_9(f)) \\ \ldots & \end{array}\right. \\
\text{group } 5 &\left[\begin{array}{ll} \mathcal{E}_{21}(f) = & (E_0(f) + \Delta E_5(f)) \cdot \exp(\mathrm{i} \cdot \varphi_{21}(f)) \\ \ldots & \end{array}\right.
\end{aligned}
\tag{18}
$$

with ΔE_n the E-field difference for group n. This E-field-delta element is added to the term of the general E-field amplitude. For the first group, enumerated `group 0`, no difference is added. Any group may be chosen to as first group, but the power string group is a sensible choice as it contains the most antennae, thus maximizing the accuracy of the shared E-field amplitude.

6.2   Technical implementation

The channel grouping is implemented as a standard 2D python list: every sub-group represents a channel group and contains its respective channel IDs. A flattened 1D version of target channels is generated from that list as well, with the grouped list only being passed where the group structure is relevant, so compatibility with the previous code remains high.

The additional terms are implemented in the `get_likelihood_operator` function of

```
1    /NuRadioReco/modules/iftElectricFieldReconstructor/
2    iftElectricFieldReconstructor.py
```

which serves to build the likelihood operators which are later used for minimization.

First, a dict containing the parameters for the prior model of the delta E-field has been added.

In the previous module, one loop iterated over all channels constructing a likelihood term containing the electric field term and a phase, as described in Equation 17. This structure has been replaced with a nested loop iterating over all channel groups, and then over all channels within the respective group.

In that loop, an NIFTy Smooth Linear Amplitude is built using the parameter dict for every group after the first one, forming is the actual prior for the generative model.

For the first channel group, a NullOperator has been instead initialized. This operator is the equivalent of a matrix of all zeros. It has been assigned to the data structure to which the delta E-field operators are then appended later.

This SLAmplitude is then used as the amplitude of a Correlated field. This field, which currently exists in the frequency space, is zero padded, i.e. entries with 0s are added at the edges of the field. This is done because the field is constructed with periodic boundary conditions, but the edges of the field represent the low and high ends of the frequency spectrum. Field amplitudes at these very different locations in the spectrum are generally not identical, which would create a sudden jump in field amplitude. That in turn would clash with the smoothness conditions to be imposed on the field. Therefore, empty entries are added around the field, which provide space in which the field can bridge the difference between its two edges. The amplitudes in these regions added by the padding do not affect the reconstruction either, as these correspond to frequency regions which are filtered anyway.

Next, the field's domains are flipped, it is symmetrized, and realized.

At this point, a simple scaling factor has been introduced to the model, which is multiplied to the field operator to provide additional fine control in shaping the delta E-field. With this, the delta E-field operator has been constructed.

The construction of the spectral operators for the two polarization has consequently been modified.

To construct the polarization operators, the operator for the shared E-field amplitude is now added with the delta E-field operator of the current group. This sum is then multiplied with the complex phase and a polarization field, and a filter operator is applied.

These spectra operators are then processed as in the pre-existing module.

Additional modifications were made in the `run` function.

The `iftElectricFieldReconstructor` in its previous version is passed a 1D list of channels to use for reconstruction. These channels are then queried for their signal ray type, which describes whether the trajectory of the radio signal was direct, reflected, or refracted. As different ray types are reconstructed separately, this loop filters all channels with the ray type matching the one currently being reconstructed into another list, which then gets passed to the rest of the module.

This querying loop has been modified to a nested loop, which iterates over all groups and then over all channels in a group, while saving channels corresponding to the ray type currently being processed in group-wise fashion. This is done to preserve the group structure of the channels through the ray type sorting.

Lastly, changes have been made to

```
/NuRadioReco/examples/RNO_energy_reconstruction/
T03_electric_field_reco.py
```

This is a script that runs the electric field reconstruction module for Monte Carlo data. In this script, the actual 2D array of channel groups is defined. Additionally, some changes were made to how the channels are passed to preparatory functions to accommodate that structure.

7 Diagnostics and Results

This chapter presents the results produced with the expanded reconstruction module. As the module initially failed to produce satisfying reconstructions, a series of checks was run in an attempt to identify the cause of the errors. The following checks were run:

- Channel groups with only one channel were excluded.

- Channel group arrangement was rotated.

- Many different parameters for the prior model were tested.

Eventually, a prior model was found that does produce a promising reconstruction. The contribution of the expanded model to those results was then investigated.

Lastly, a possible cause of the unsatisfactory reconstruction behavior is proposed.

7.1 Initial results

During testing, all events in the MC data set were reconstructed. For comparability, all plots shown in this chapter are reconstructed from the same event. Channel 8, the topmost VPOL on the power string, did not observe this event and thus is not pictured in the following plots. This run was chosen as its incidental neutrino interaction vertex lies in in circa 1 km distance to the detector, creating a event with a overall low complexity as the signals on the antennae will only vary slightly. The prior for the delta E-field was initially chosen similar to the E-field amplitude, with a slightly less steep power law to account for greater possible variations at shorter wavelengths, with the following settings:

prior	sm	sv	im	iv	scaling
E-field amplitude	-4.9	0.5	2	0.5	N/A
delta E-field	-4	0.5	-4	2	10^{-2}

Table 1: *Prior parameters.*

An in depth description of the parameters of the prior model is given in subsection 7.4.

One reconstructed channel without the new functionality is shown in Figure 10. A similar results is expected for the new reconstructions, however
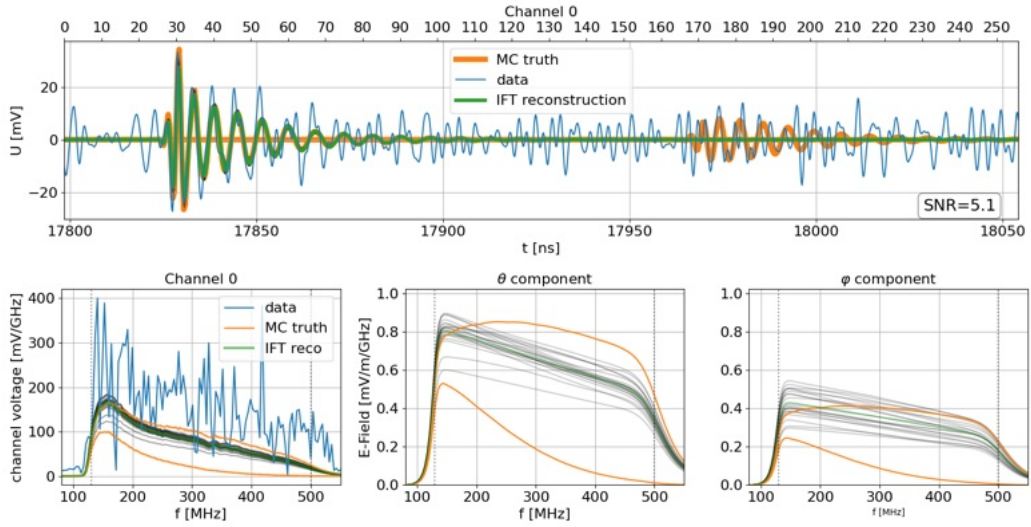
Figure 10: *Example of a reconstruction for one channel, generated by the pre-existing module. Top: The trace of the Antenna voltage over time. Bottom, from left to right: The reconstructed electric field in Fourier space, the it's Θ and φ components. Blue is the noisy data, orange represents the noiseless MC-generated truth, and green is the reconstruction. Grey lines in the lower plot are reconstruction candidates and can serve as an estimator for the uncertainty of the reconstruction.*

initial reconstructions failed to accurately identify the Monte Carlo signals. Each reconstruction produced on of two different types of faulty results:

The first type of result exhibits obvious problems. As can be seen in Figure 11, the reconstructed frequency spectrum fails to reproduce the shape of the E-field spectrum entirely. The channel voltages, shown in Figure 12, in the time domain do indeed identify the signal locations, but fail to accurately reconstruct the shapes of the signal pulses. This type of error is likely caused by the MGVI optimization algorithm settling in a local maximum, and already occasionally occurred using the previous version of the reconstruction. With the upgraded version, this behavior is more common in general, but less common than the second type of faulty reconstruction.

In the second type of result, as shown in Figure 13, the first channel group reconstructs a diminutively small E-field in the frequency domain. The rest of the channel groups match the peak of the frequency spectrum more closely, but also loses in accuracy at higher frequencies. Firstly, apparently the reconstructed E-field for the groups 2 to 6 seems to consist entirely of the delta E-field, while the actual E-field amplitude has been almost entirely minimized. Secondly, the E-field spectrum then fails to to follow the reduction

36

at high frequencies that the signal exhibits, although the delta E-field prior was chosen such that it generates more high frequency admixture to the delta E-field, thus in theory allowing it to be more flexible.

The channel voltages exhibit similar behavior. For the group of the phased array, a voltage very close to 0 is reconstructed. The VPOLs of the other groups do actually reconstruct the true signal to a moderate degree, even though the HPOLs in channels 11 and 23 also exhibit greater inaccuracy. This type of problem occurs most of the time compared to the first type.

After the initial failed reconstructions, the code was examined for problems responsible. After no such problems were identified, a series of checks was run with the goal of identifying the cause of the unexpected behavior.

7.2   Excluding the single-channel groups

As a first check, the channel groups containing only one VPOL antenna were excluded from the reconstruction. Because the individual VPOLs on the power string represented by channels 6, 7, and 8 each only add one channel of data to the reconstruction, but also each add a degree of freedom in the form of one additional delta E-field, they carry the potential to add considerable complexity to the likelihood which is being maximized, and thus possibly throw off the reconstruction.

Additionally, as can be seen in Figure 12 in channel 7, the true Monte Carlo pulses overlap in timing for this channel. This points towards an event geometry where refracted and directly traveling signals overlap in arrival timing, which is not handled in the IFT model currently. This can cause the minimization to fail to converge, a behavior that was already observed for previous reconstructions [29].

Excluding channels 6, 7, and 8 from reconstruction does prevent the first type of faulty result to a degree, depending on respective event geometry. For the shown event, the complete disagreement with the true signal disappears almost entirely if especially channel 7 is excluded. This confirms that the pulse overlap was responsible for that type of error. However, it does not affect the second type of error at all. If the individual channel groups are excluded, all results appear similar to Figure 13 and Figure 14.
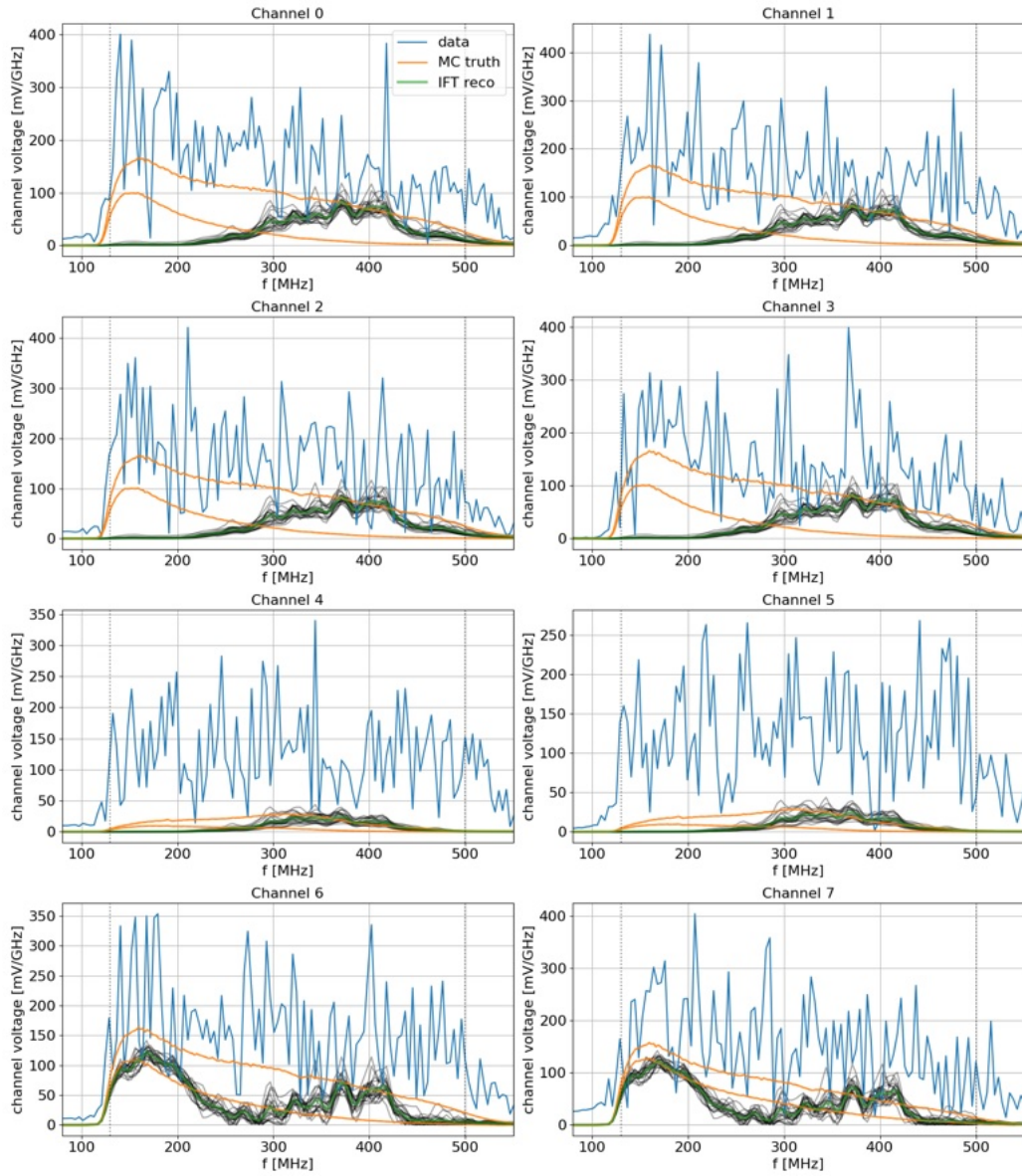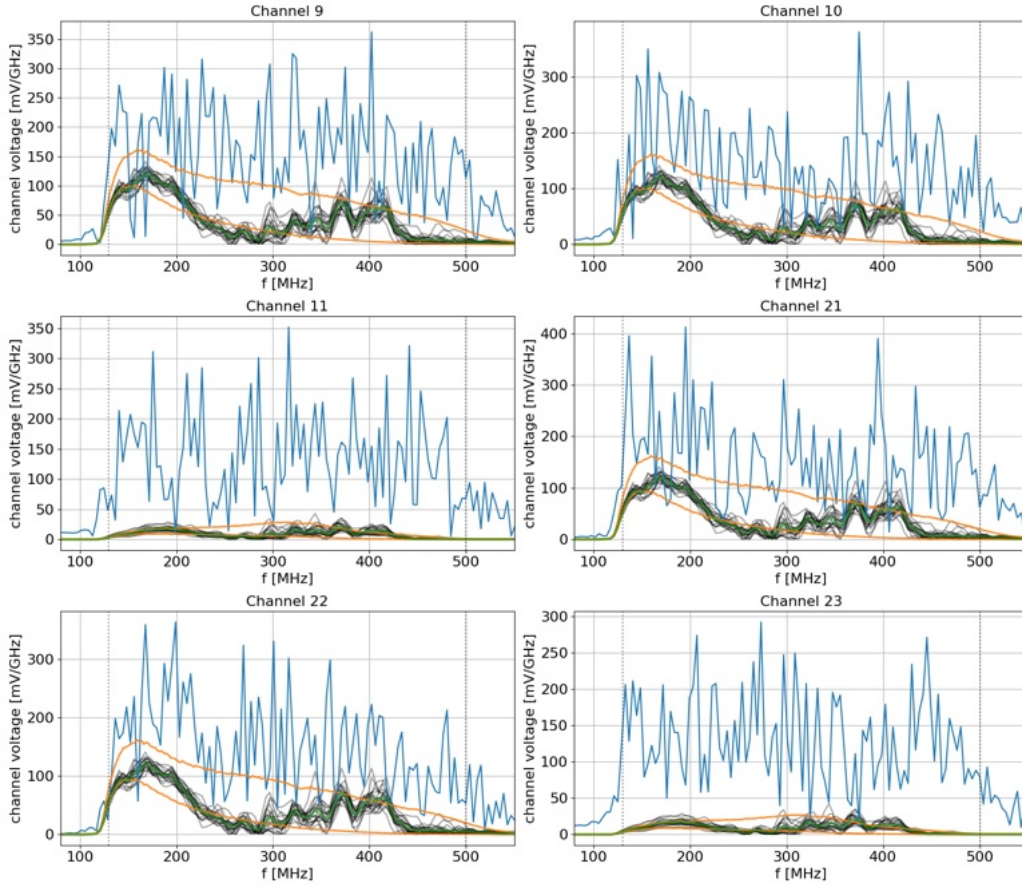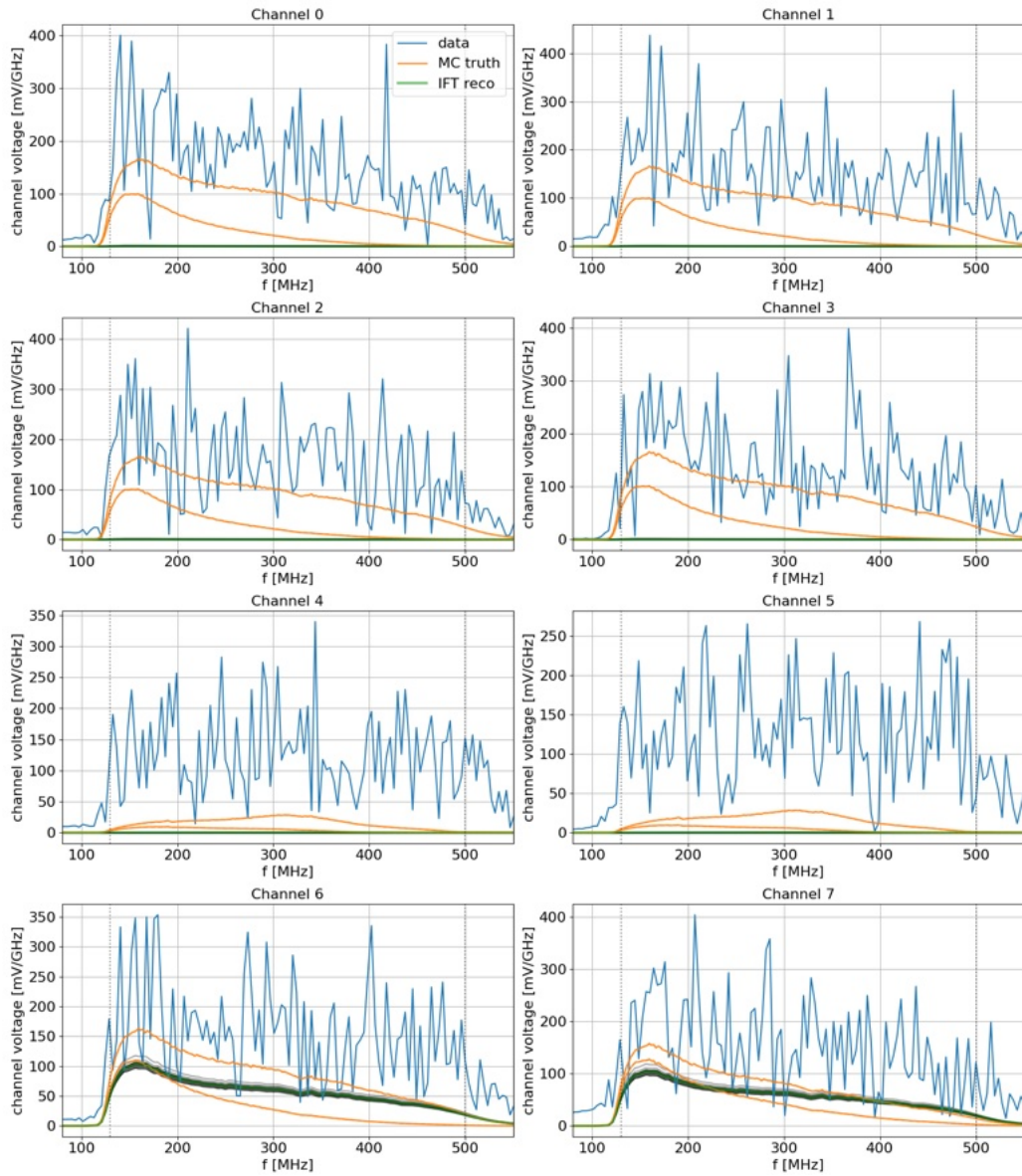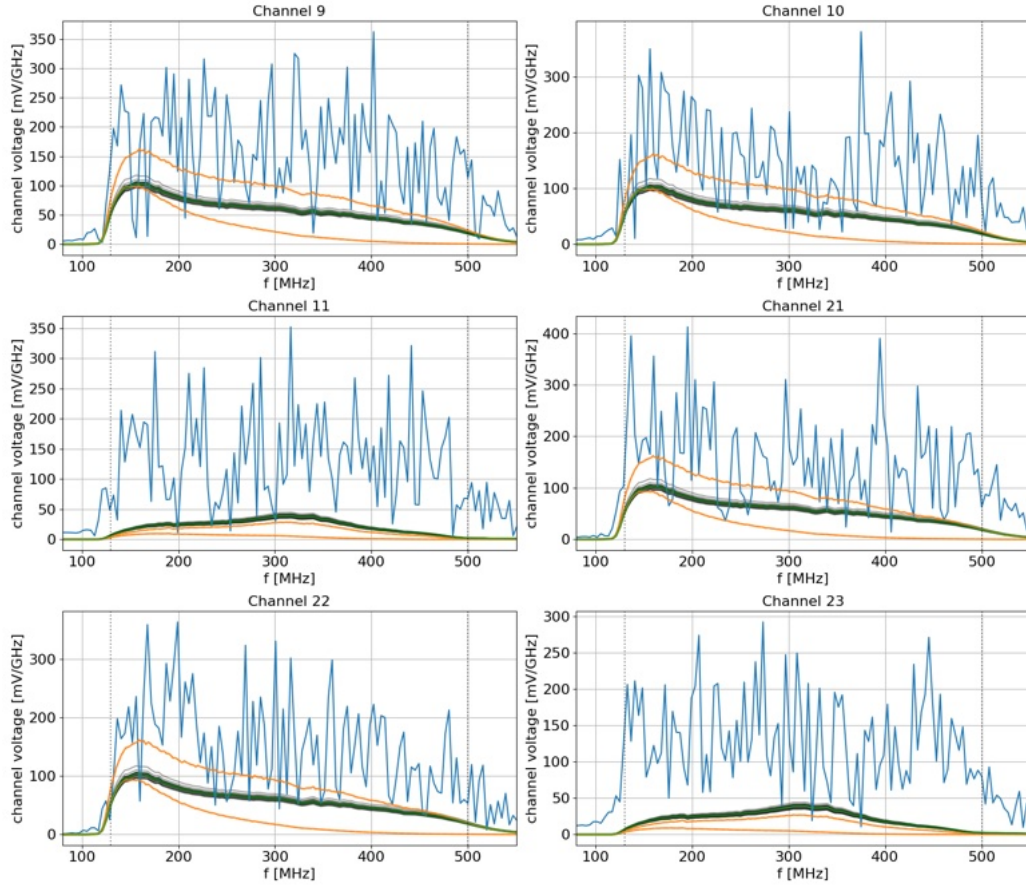
Figure 11

Figure 11: *(Continued) Initial electric field reconstruction, using all channels. Each row shows one channel, which corresponds to one antenna. From left to right, each row shows: the total electric field, its θ component, and its φ component. All plots are in the frequency domain. Orange is MC truth, blue is the noisy data, green is the reconstructed electric fields, gray are samples from the KL minimization process. This results shows the first type of error with strong disagreement with the MC truth in all channels.*
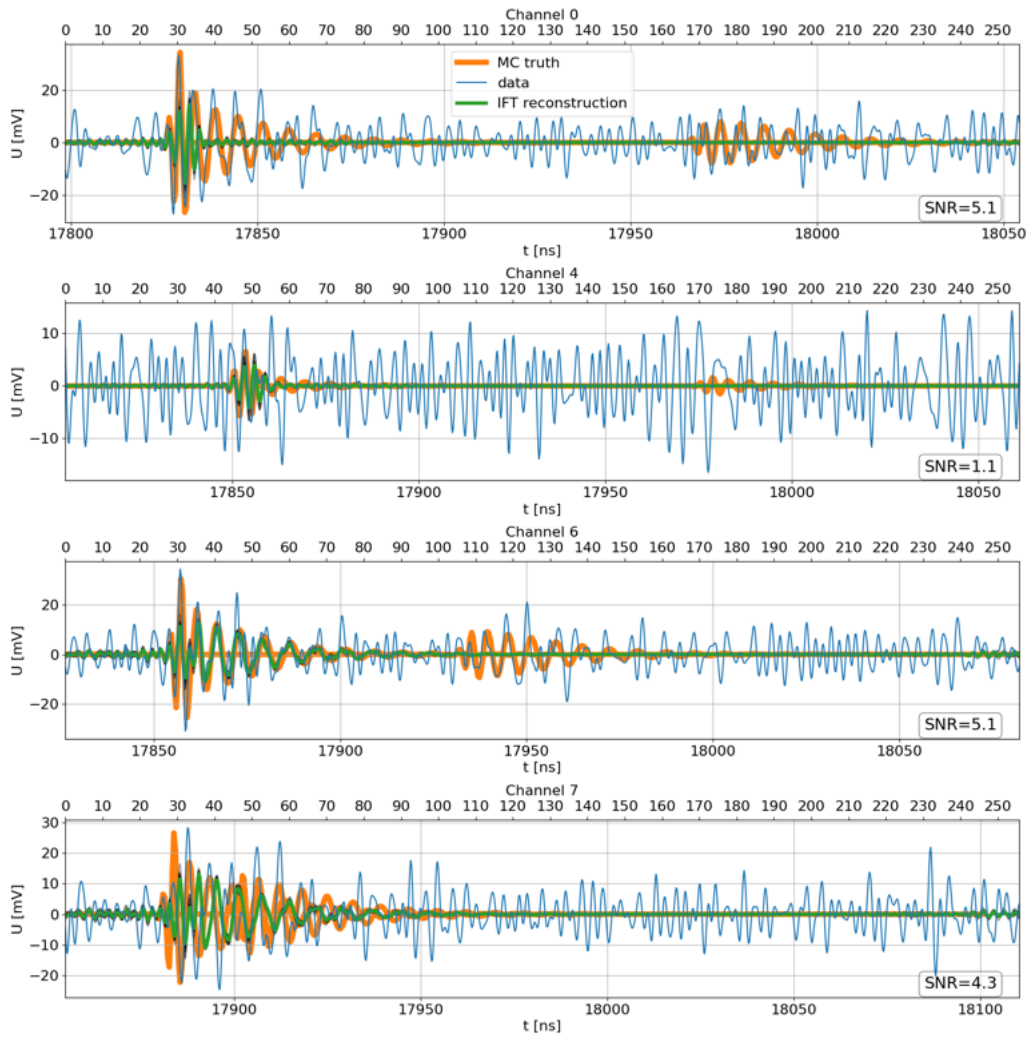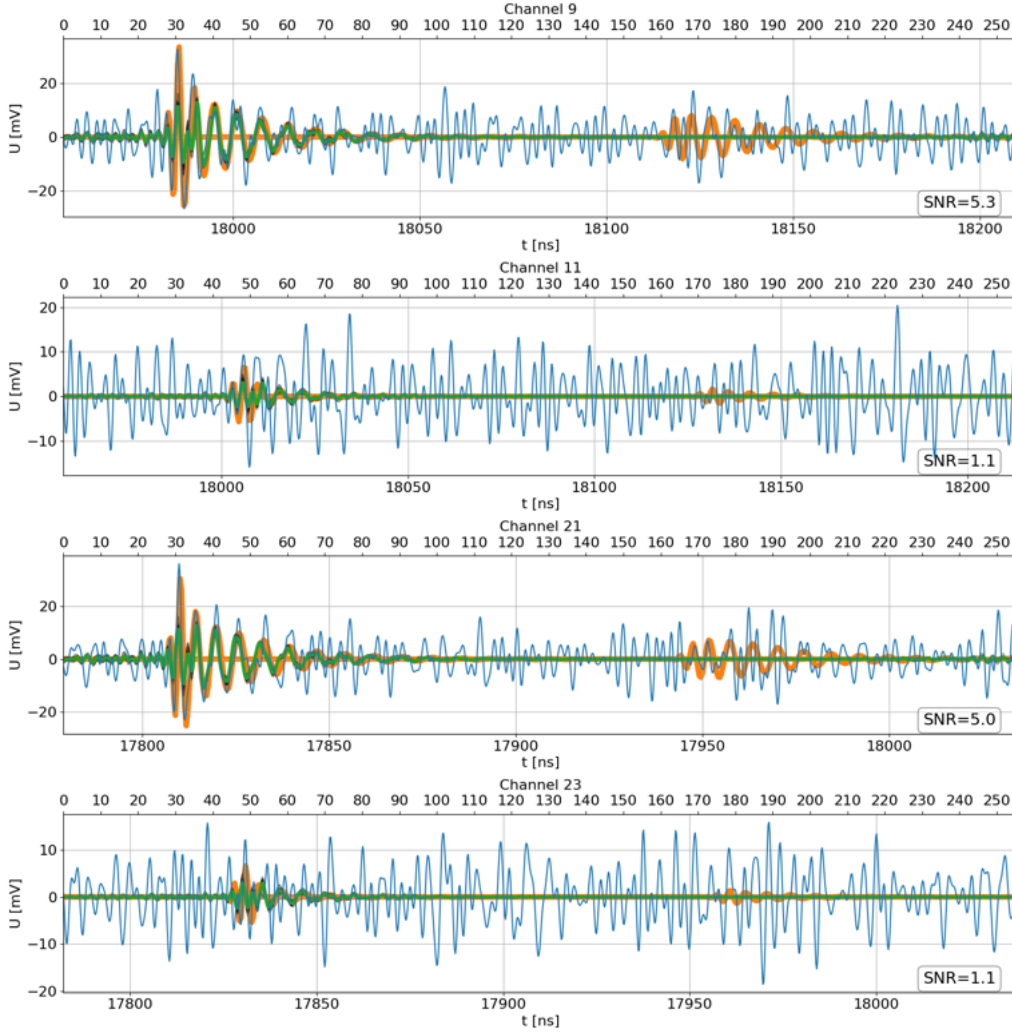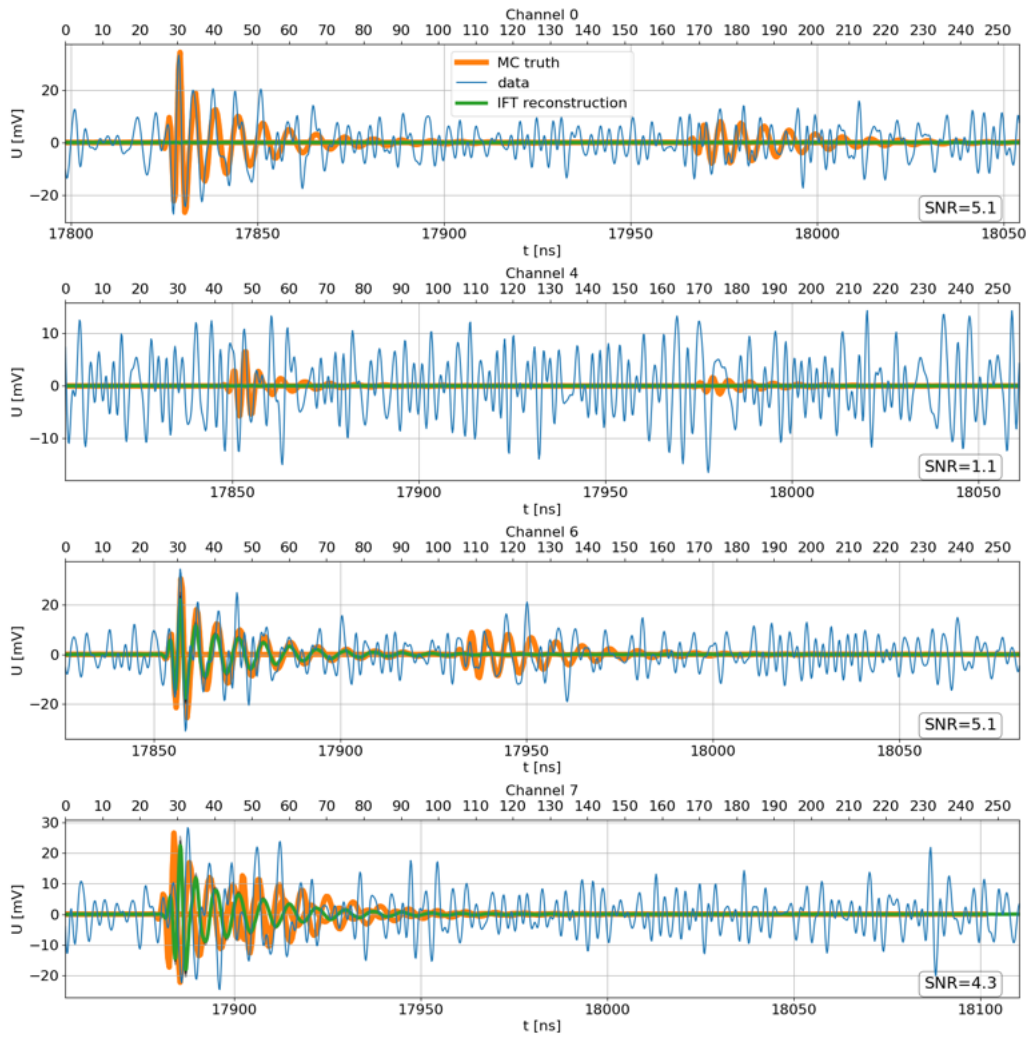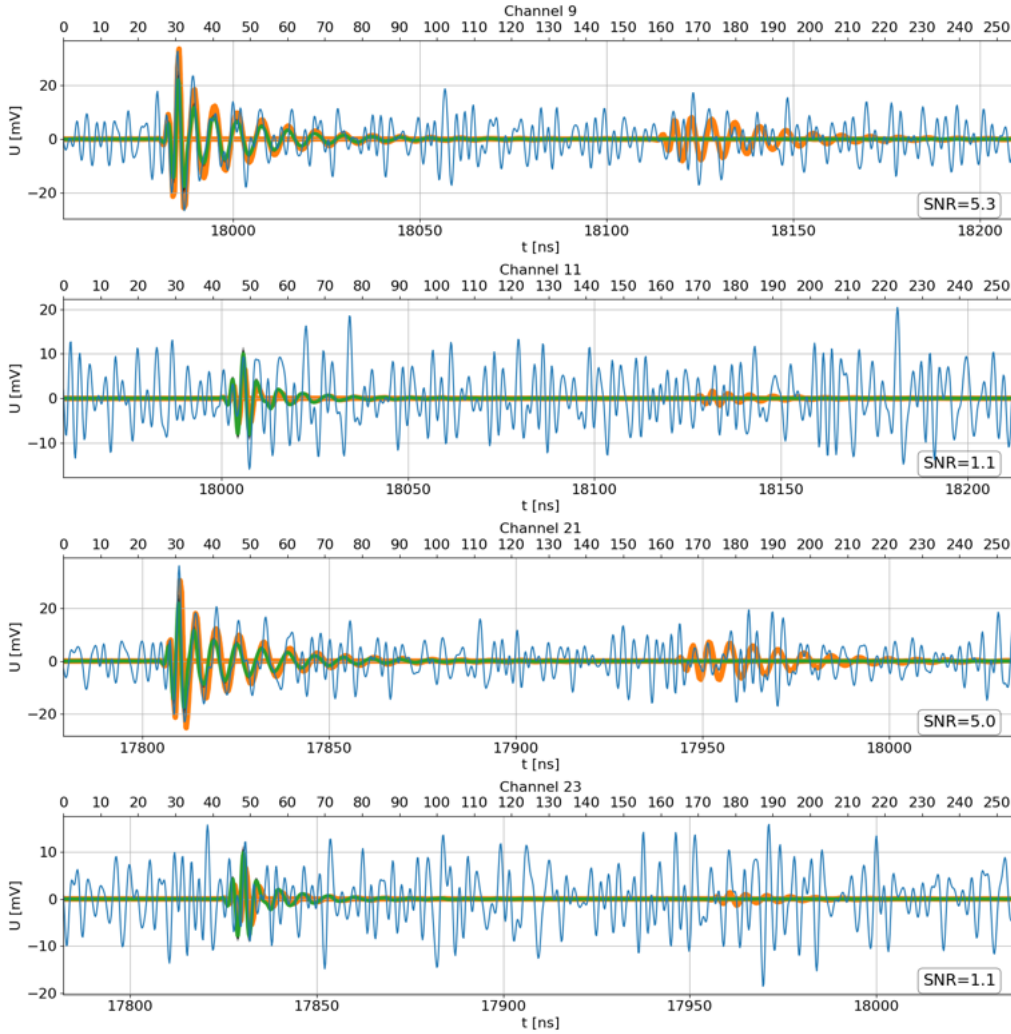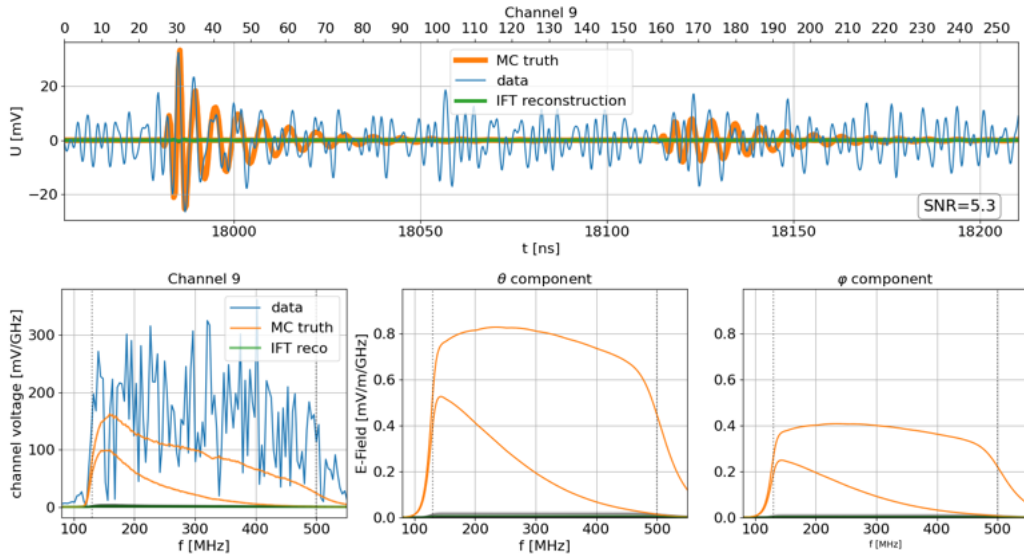
Figure 12

Figure 12: *(Continued) Initial reconstruction of the channel voltage in the time domain. Only one VPOL and HPOL per channel group are shown for space reasons, but all channels were used in the reconstruction. Orange is MC truth, blue is the MC generated noisy data, green is the reconstructed pulse. This results shows the first type of error with strong disagreement with the MC truth in all channels.*

41

Figure 13

Figure 13: *(Continued) Initial electric field reconstruction, using all channels. Each row shows one channel, which corresponds to one antenna. From left to right, each row shows: the total electric field, its θ component, and its φ component. All plots are in the frequency domain. Orange is MC truth, blue is the noisy data, green is the reconstructed electric fields, gray are samples from the KL minimization process. This results shows the second type of error, where the first channel group is not reconstructed at all.*

Figure 14

Figure 14: *(Continued) Initial reconstruction of the channel voltage in the time domain. Only one VPOL and HPOL per channel group are shown for space reasons, but all channels were used in the reconstruction. Orange is MC truth, blue is the MC generated noisy data, green is the reconstructed pulse. This results shows the second type of error, where the first channel group is not reconstructed at all.*

45

7.3   Shifting group arrangement

To confirm that the reconstructed E-field consists of mostly the delta E-field while the E-field amplitude is minimal, rather than some sort of effect related to the power string group of antennae causing the obtained results, the order of the channel groups was rotated. Each channel group was tested as the "first" group.

Two exemplary reconstructed channels are shown in Figure 15 and Figure 16, for a group constellation in which the channel group of Helper string 1 consisting of channels 9, 10, and 11 is the "first" group. As can be seen, now the channel 9 reconstructs its E-field to near zero, while channel 0 reconstructs better. This behavior was observed for every arrangement of channel groups, which confirms that only the delta E-field is reconstructed.

Figure 15: *Reconstruction of channel 9, which is in the first group in this configuration. Top: the channel voltage in time domain. Bottom, from left to right: the total E-field spectrum, its φ and its θ components. Green is the reconstruction, orange the MC generated true signal, blue the simulated noisy data, and gray random KL samples.*
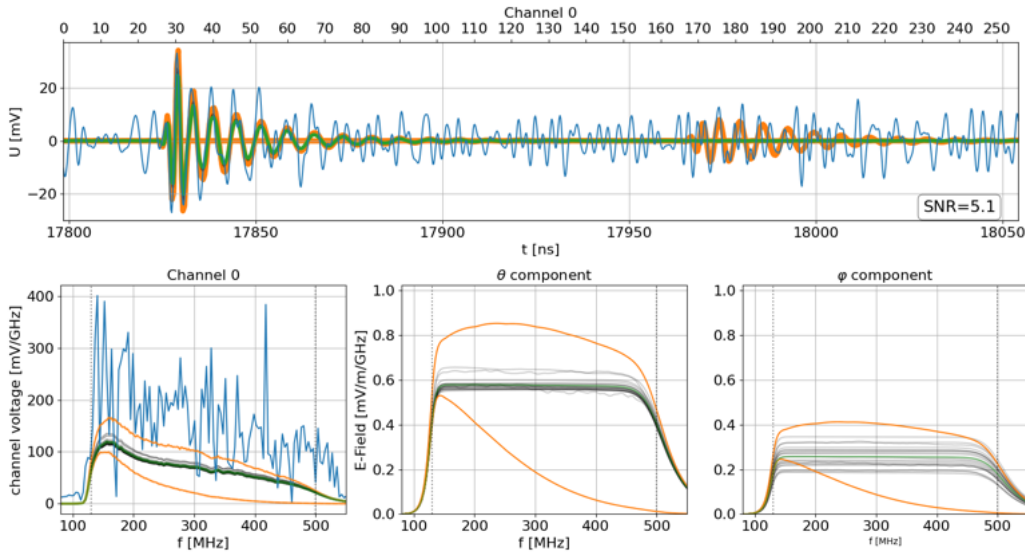
Figure 16: *Reconstruction of channel 0, which is in the second group in this configuration. Top: the channel voltage in time domain. Bottom, from left to right: the total E-field spectrum, its φ and its θ components. Green is the reconstruction, orange the MC generated true signal, blue the simulated noisy data, and gray random KL samples.*

7.4 Varying the delta E-field prior

As the prior has a great influence on the accuracy of the reconstruction and the potential to throw it off, prior selection for the delta E-field was tested extensively. An example of samples drawn from a prior for the E-field amplitude are given in Figure 17. It shows the power laws distributions which defines the E-field spectra, and the spectra and traces generated from those power laws without and with detector effects. The generative process which forms the delta E-field prior has the parameters:

- `sm` and `sv`: these parameters define the mean and variance of the slope of the power law in Fourier space which defines the spectral density of the signal. This parameter is the primary influence on the overall shape of the reconstructed field.

- `im` and `iv`: these parameters set the mean and variance of the intercept with the y-axis in Fourier space of spectral density function. In practice, this is an offset defining how strong the lowest and thus strongest frequencies are present, and can be understood as a sort of scaling of the overall delta E-field amplitude.

47
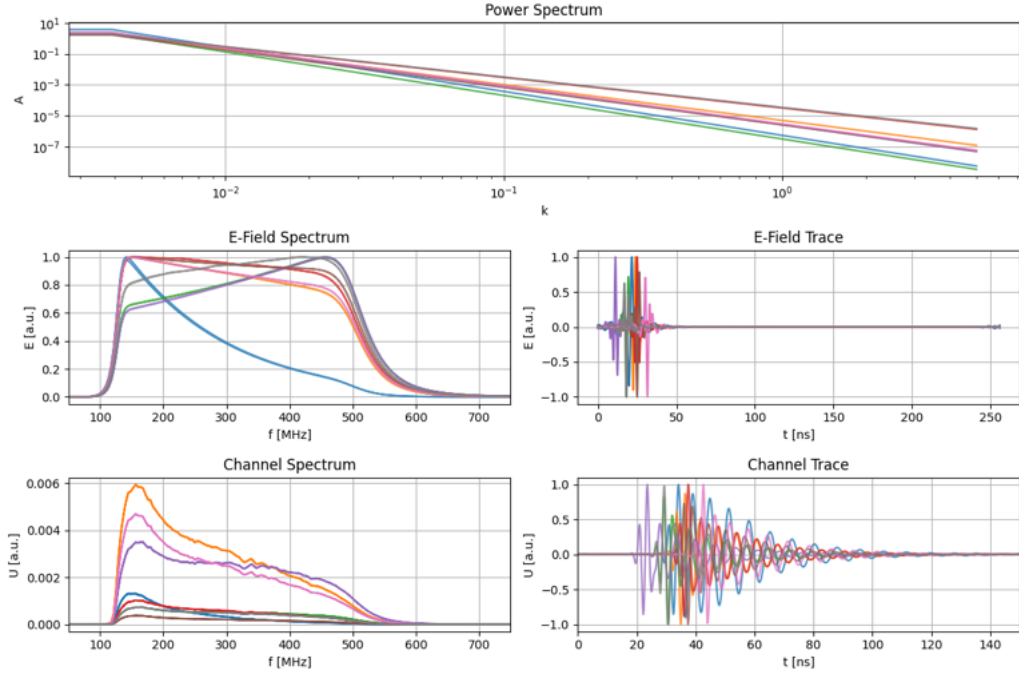
Figure 17: *Example of a prior of the E-field amplitude. Top: power spectrum from which the E-field is drawn. Middle, from left to right: sample E-fields in Fourier space and time space before detector effects and noise. Bottom, left to right: sample E-fields with noise and detector effects. Middle and Bottom row are normalized for comparability.*

- `a` and `k0`: these parameters define the smoothness and the low frequency cutoff of the prior, and are relevant for the Gaussian process part of the generative process.

- An additional scaling factor is introduced, which is scalar-multiplied with the prior operator after shaping, symmetrizing, and realizing to further fine tune the amplitude of the delta E-field.

`sm`, `im`, and the scaling factor were varied individually and in combinations to investigate the behavior of the reconstruction depending on prior choice. Initial values used for the results up to this point are shown in Table 1. As the scaling factor scales the entire prior field operator, it was generally set to 1 to investigate `sm` and `im`, as any change in the latter parameters would be scaled down otherwise as well.

`sv` and `iv` were set to a higher variance than for the general E-field amplitude prior, and kept static otherwise, as they have a lesser impact on the final result of the reconstruction than the means of the prior parameters.

The prior of the E-field amplitude was not varied.

If the delta E-field is suppressed by choosing a very low `im` for the prior, the reconstruction eventually fails to approximate the signal even for the second and subsequent channel groups. Some prior samples are shown in Figure 18. As also shown there, the spectra generated are no flexible enough anymore properly conform to either the spectrum in Fourier space, or the pulse in the time domain. The plotted results used an `im` of −50 and a scale of 0.01. For `im` between circa -50 and -4, the reconstruction is similar to the case in Figure 13, with almost purely the delta E-field reconstructing.

Increasing `im` has very little effect on the reconstruction, and is of lesser interest as the delta E-field is expected to posses a smaller amplitude rather than a greater one.

If `sm` is increased while keeping `im` fixed, the slope of the power spectrum becomes less negative and more high frequencies are generated for the delta E-field spectrum. Samples from the prior distribution and one exemplary channel reconstruction are shown in Figure 19. As can be seen there, the spectrum in Fourier space becomes very jagged and starts conforming with noise. The channel voltage in the time domain still reconstructs the pulse reasonably well. However, the first channel group still reconstructs to a very small E-field, which implies that the E-field visible is still almost entirely the delta E-field. The shown plot was generated with an `sm` of −3.5 and the scaling of 0.01, showing that the reconstruction is in fact very sensible to a too low slope of the prior's power law.

For lower `sm`, the reconstruction initially does not behave differently. All groups except group 0 reconstruct via the delta E-field, while the general E-field amplitude remains minimal. The reconstruction is indeed very insensitive towards lowering `sm`, as reconstruction shape did not change noticeably even for values of -1000 and a scaling factor of 1.

Small reductions of the scale parameter show little effect on the reconstruction. However, at very low values of 10^{-5} or even lower, the reconstruction improves considerably. As shown in Figure 21, the reconstruction will not only accurately approximate all radio pulses in the time domain reasonably well, the electric fields are approached by the reconstruction as well, both the sum and in their φ and θ components. This behavior is consistent across the entire Monte Carlo data set as well, and smaller inaccuracies in the results do appear together with problematic pulse shapes as discussed in subsection 7.2.

This behavior is in principle not surprising. The previous reconstruction implemented the option to scale each E-field amplitude individually with a

small multiplicative factor to allow at least some flexibility between different channels. As the entire delta E-field is scaled down enough, this also flattens it and allows it to act as an offset similar to that earlier scaling functionality. Nonetheless, this effect occurs at scaling factors $\lessapprox -10^{-5}$, which is considerably lower than comparing the Monte Carlo spectra would suggest.

This strong divergence from the assumptions made during model building would warrant further investigation, however due to time constraints it was not possible to test these results extensively.

7.5   Investigating contribution of the delta E-field

To understand the contributions of the delta E-field and check that the reconstruction does indeed function properly, the differences between the E-field amplitudes in Fourier space are examined. As each of the spectra should be close to identical except for a small contribution from hardware operators, taking the difference between a spectrum from group 0 and another group provides that group's delta E-field, and differences between two groups other than the first can demonstrate variance of the delta E-fields among each other.

Figure 22 shows that the reconstructed electric fields are very similar, even for different groups. This is consistent with the fact that the delta E-field is scaled with a very low factor.
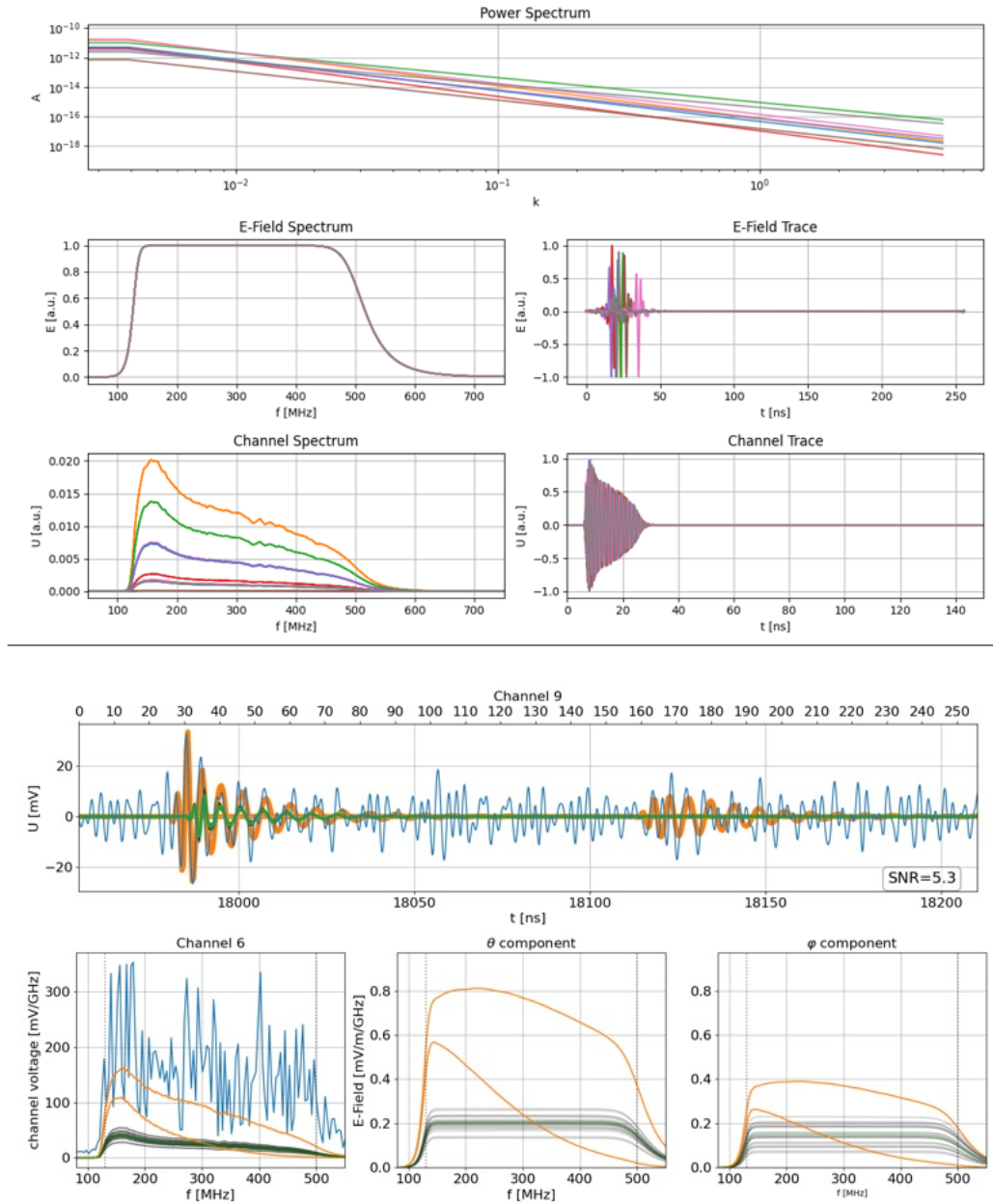
Figure 18: *Prior and reconstruction using a prior with a very low y-axis intercept. Row 1: Sample power spectra generated by the prior. Row 2: Pure E-field spectrum and channel trace generated from the power spectrum. Row 3: Spectrum and E-field trace taking after the application of hardware effects. Row 4: Channel voltage reconstruction for one VPOL in the time domain. Row 5: The corresponding electric field reconstruction, from left to right: the total electric field, its θ and its φ components. Orange are MC truths, green are reconstructed results, blue are simulated noisy data, gray are KL samples.*
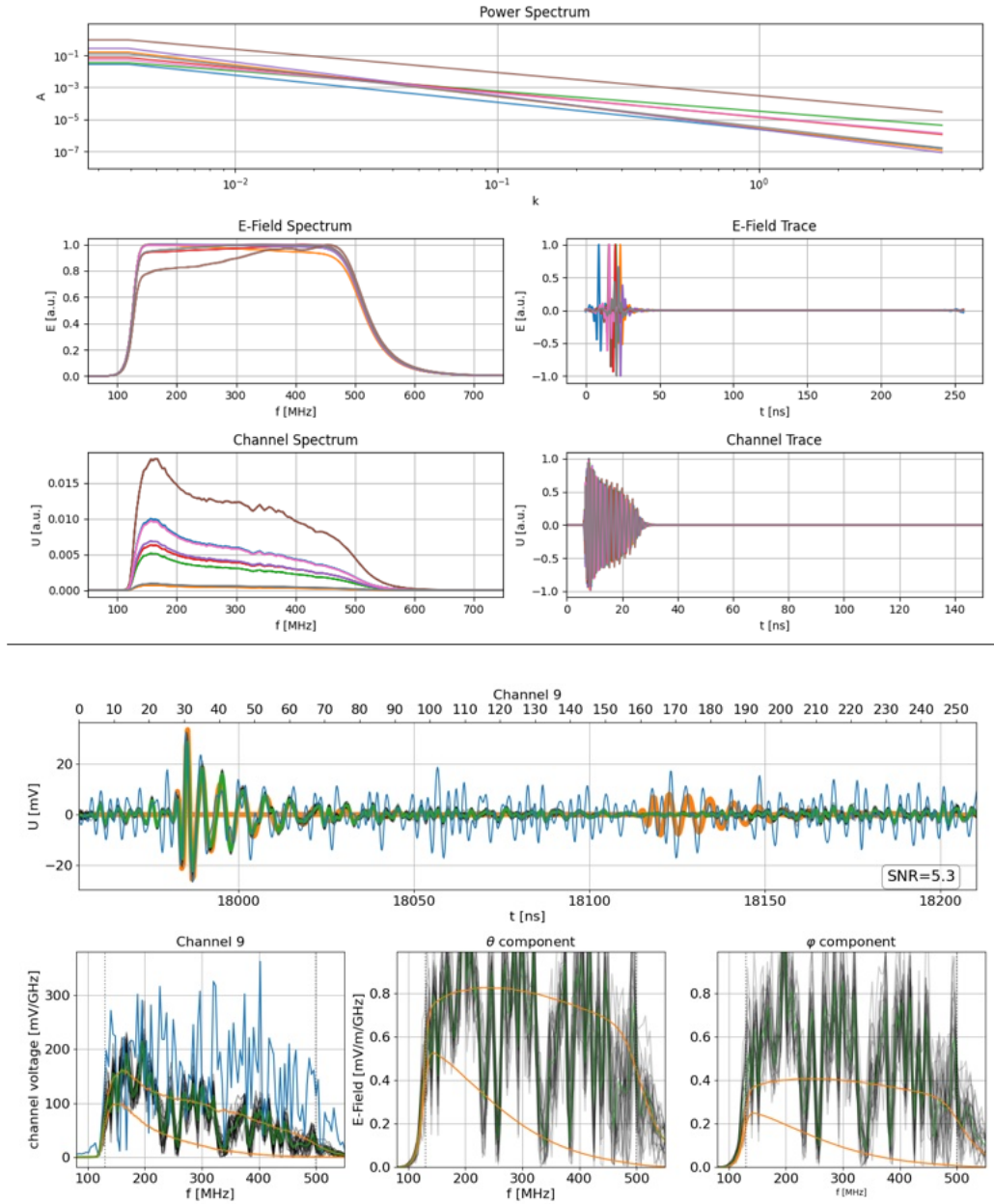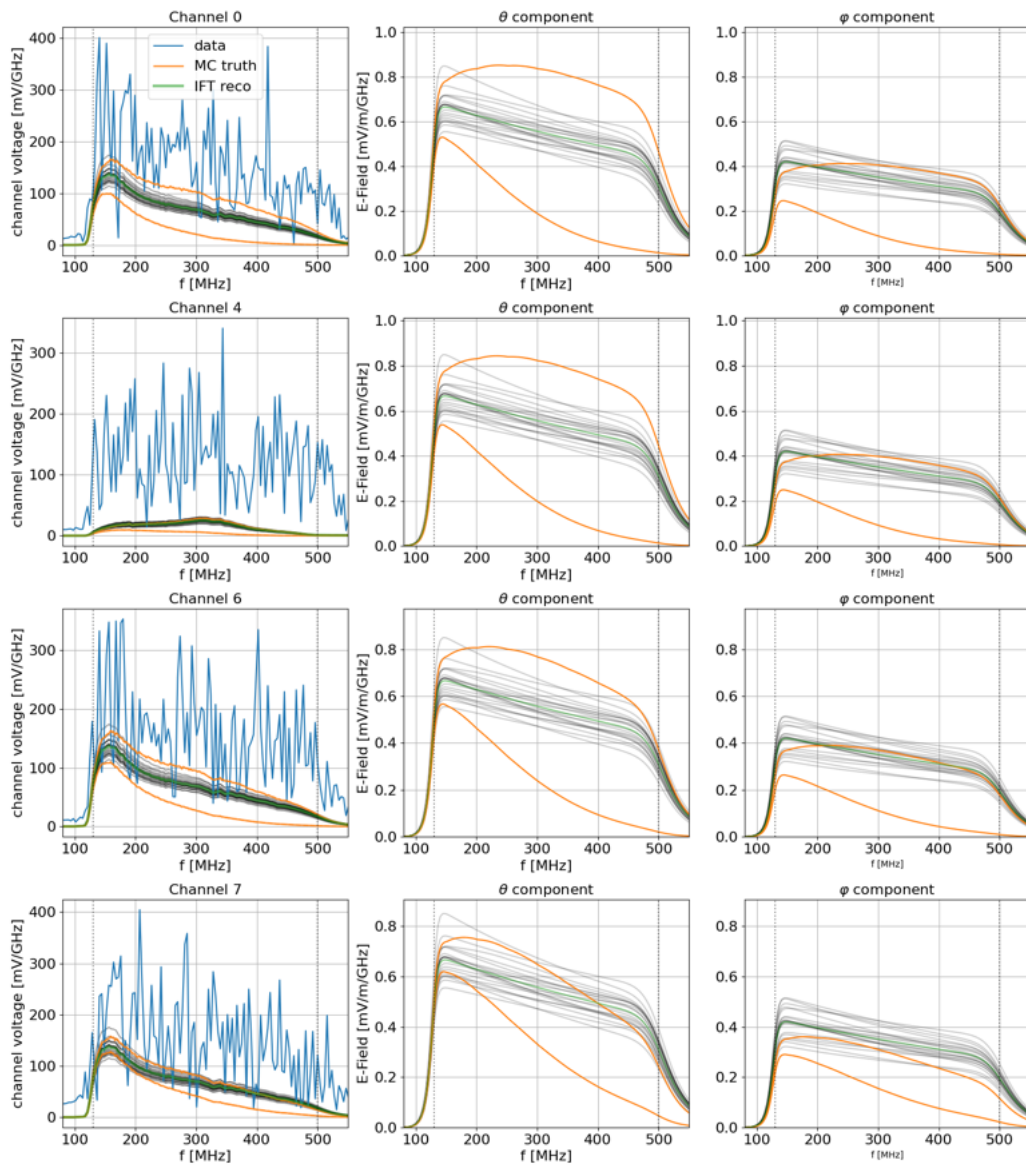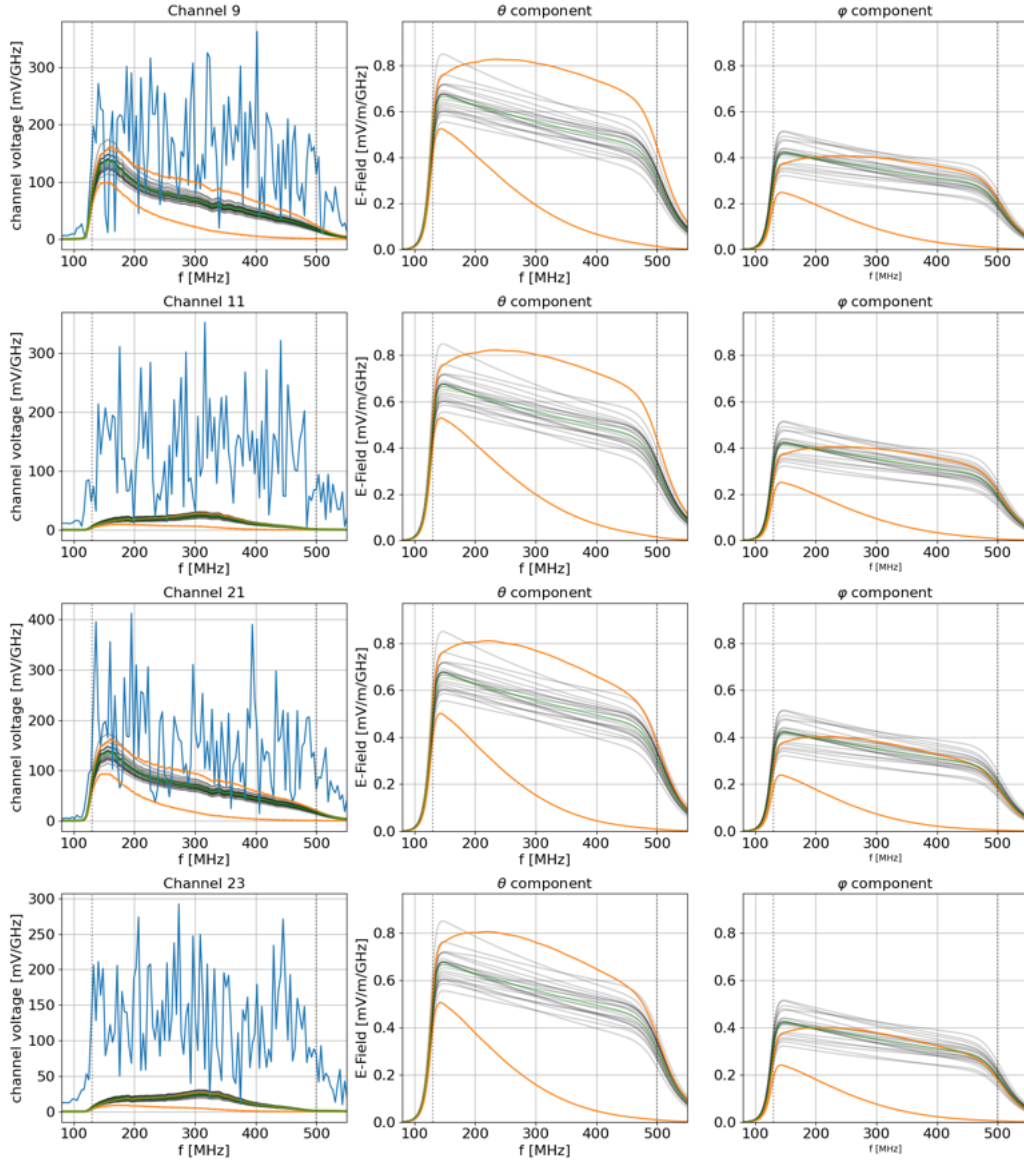
Figure 19: *Prior and reconstruction using a prior with a slightly higher spectral index. Row 1: Sample power spectra generated by the prior. Row 2: Pure E-field spectrum and channel trace generated from the power spectrum. Row 3: Spectrum and E-field trace taking after the application of hardware effects. Row 4: Channel voltage reconstruction for one VPOL in the time domain. Row 5: The corresponding electric field reconstruction, from left to right: the total electric field, its θ and its φ components. Orange are MC truths, green are reconstructed results, blue are simulated noisy data, gray are KL samples.*

Figure 20

Figure 20: *(Continued) Electric field reconstruction using the strongly scaled down delta E-field prior, using all channels. Each row shows one channel, with, from left to right: the total electric field, its θ component, and its φ component. Orange is MC truth, blue is the noisy data, green is the reconstructed electric fields, gray are samples from the KL minimization process.*
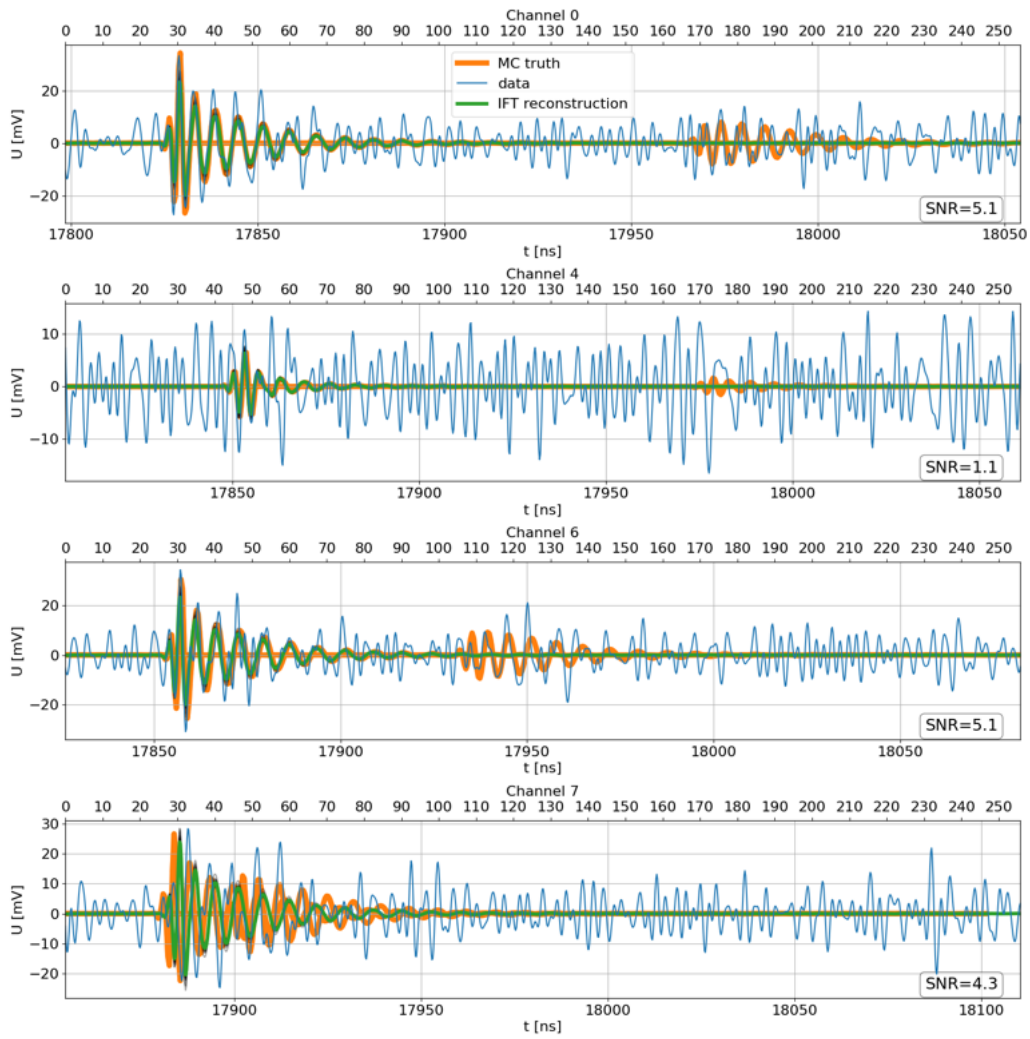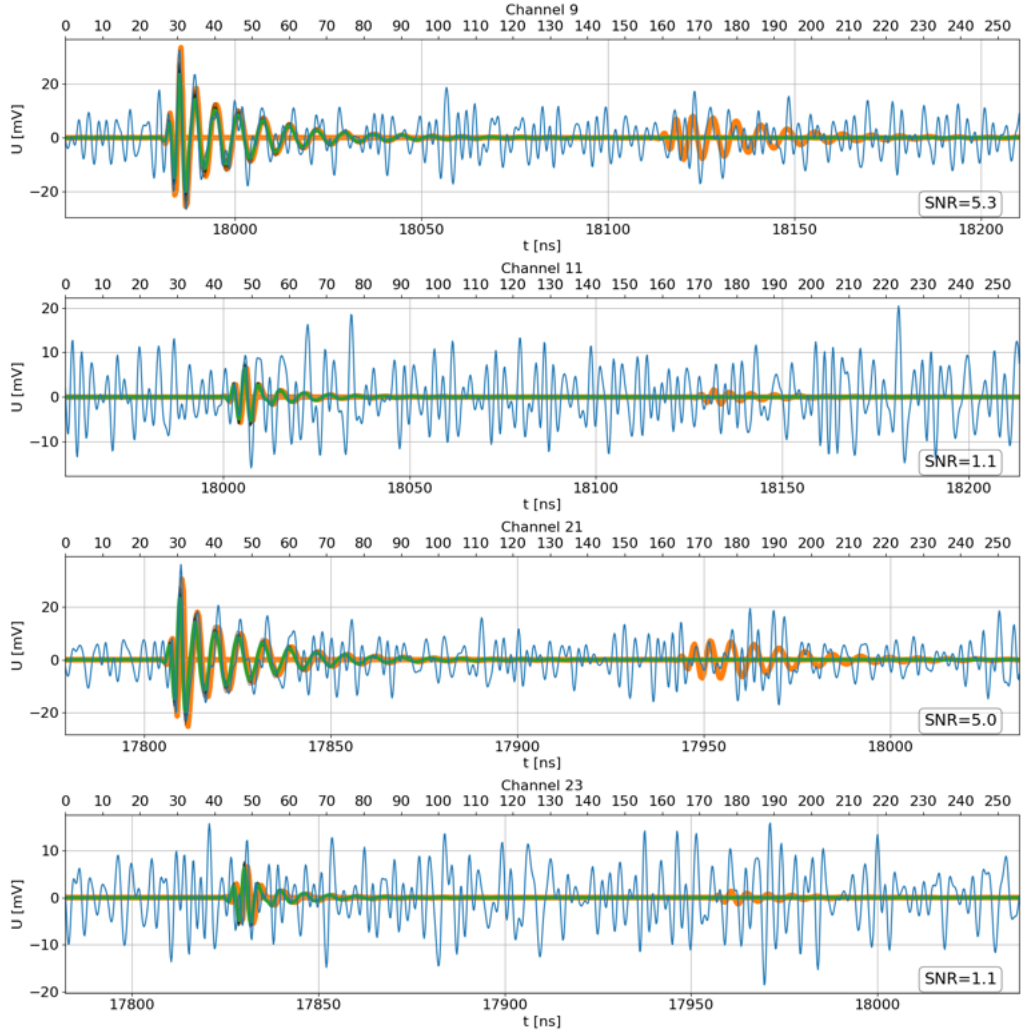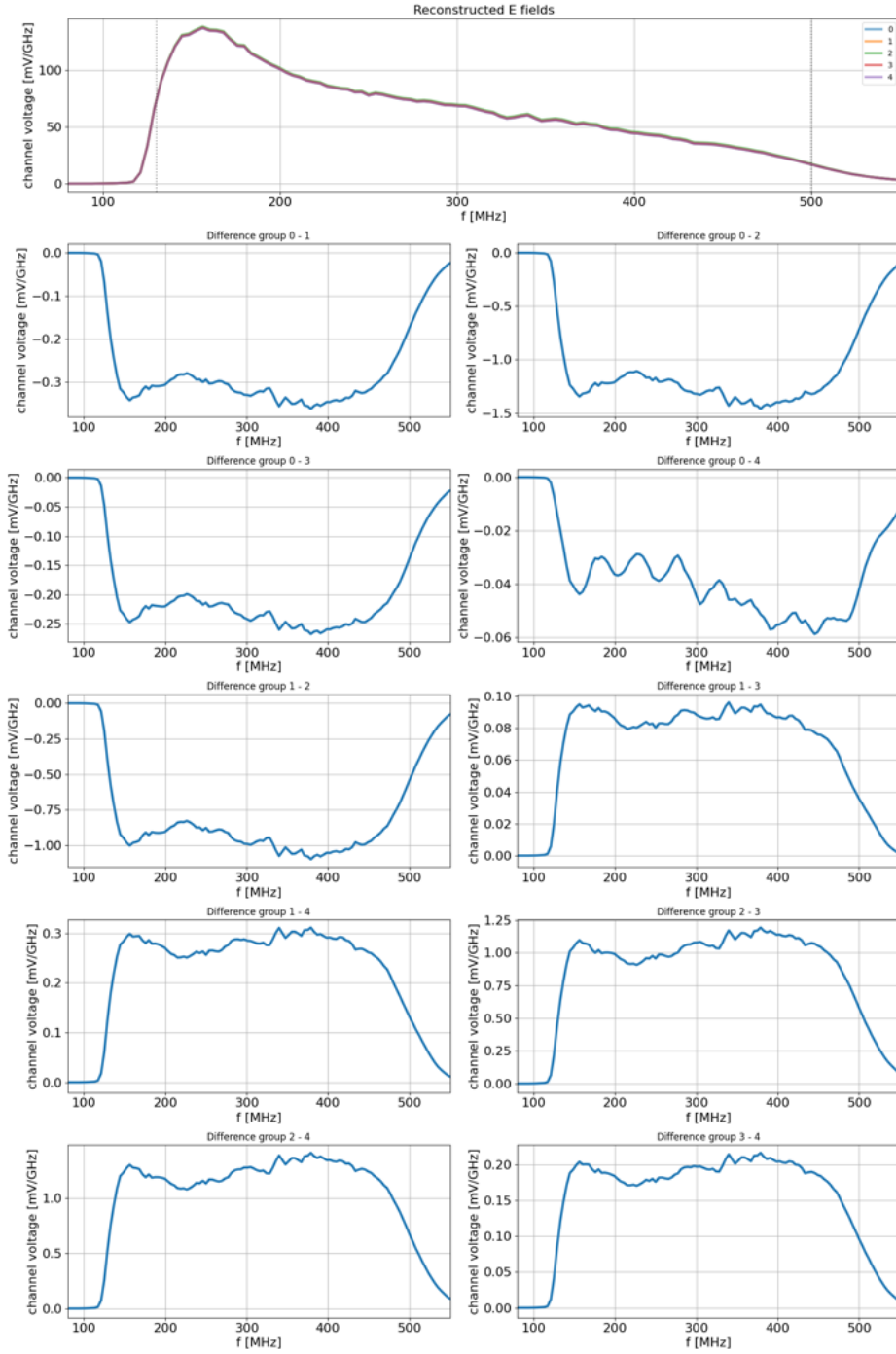
Figure 21

Figure 21: *(Continued) Reconstruction of the channel voltage in the time domain using the strongly scaled down delta E-field prior. Only one VPOL and HPOL per channel group are shown for space reasons, but all channels were used in the reconstruction. Orange is MC truth, blue is the MC generated noisy data, green is the reconstructed pulse.*

Figure 22: *Row 1: The reconstructed total E-field spectra in Fourier space, for the first VPOL channel of each group. Rows 2-5: differences between those E-field spectra for every pair of groups. The differences in the topmost plot are so small that all lines overlap.*

57

7.6   Investigating pulse time offset

One possible reason for the unexpected influence of the prior model is the pulse time offset. In preparation for the reconstruction, the radio pulse in the channel voltage trace is identified by its highest peak, the offset of all pulses to a reference pulse is calculated, and all traces are shifted such that the pulses occur at the same time mark. This is done as the reconstruction can be sensitive to time differences between the pulses, which can cause the minimization to converge on non-physical results.

The simulated channel traces for one event and all channels are shown in Figure 23. The plot shows a magnification of the primary peak of the trace. It is immediately apparent that the main peaks of the pulses do not align properly, but are staggered over circa 2 ns.

This difference in peak timing could potentially cause the convergence to be less distinct. As the delta E-fields added not only a number of additional degrees of freedom to the optimization, but the delta E-fields are also less restricted in their shape than the E-field amplitude is, this added uncertainty could be sufficient to mislead the reconstruction.

This would also be consistent with the behavior for the very low scaling factor. If the delta E-fields are sufficiently restricted, their degrees of freedom carry less weight and the reconstruction becomes dominated by the E-field amplitude once more, as is intendend. Therefore, improving the time offset calculation to produce a more accurate adjustment to the voltage traces might allow the scaling factor to be set to more reasonable values while still producing a good reconstruction.
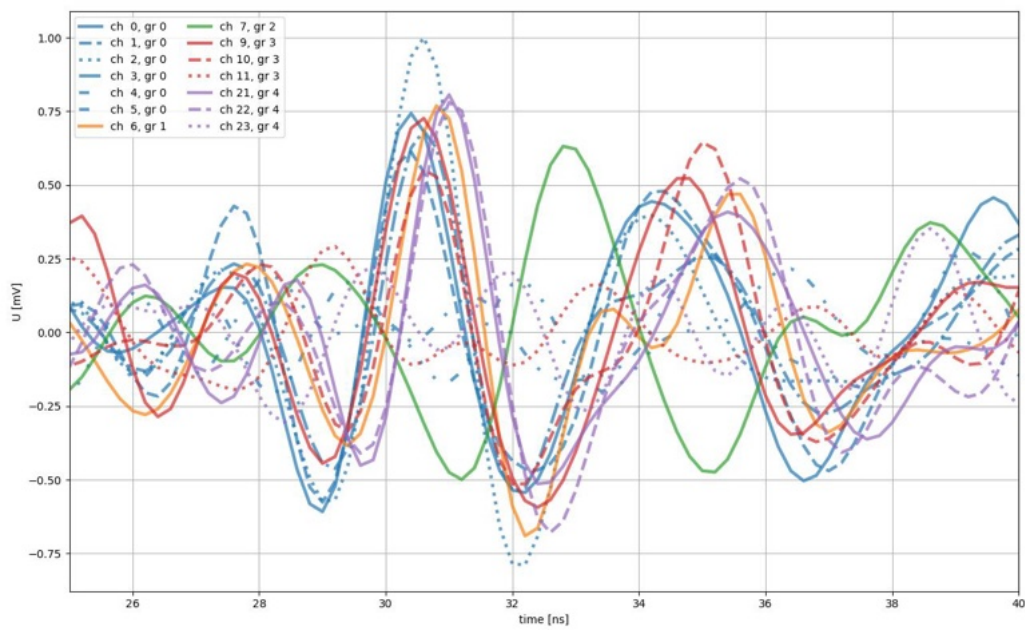
Figure 23: *Simulated channel voltage traces for all channels in the time domain. Each color represents a group. Each style of line represents a different channel. The primary peak of the pulses lies at circa 31 ns.*

8   Conclusion

This subject of this thesis was the electric field reconstruction using Information Field Theory of the `NuRadioReco` software package for the RNO-G radio neutrino detector. Research goal was to expand upon the existing reconstruction module by introducing an additional likelihood term to the electric field model to accommodate small differences in measured signal between radio antennae with non-trivial gaps between them.

As predicted in section 2, the propagation of neutrino-created radio signals in the Greenland ice-shelf has influence on the radio pulse and thus the electric field at an in-ice radio antenna. This results in the electric field being different at different antennae, a fact that is also represented in Monte-Carlo simulations of detector events.

The mathematical approach to reconstruct these electric fields is called Information Field Theory, and is shortly introduced in section 4. It combines Bayesian statistics with field theoretical methods to allow for inference. Notably, it produces statistical measures which can be optimized numerically to compound with systems which can not be formulated analytically, and it is independent of discretization beyond a minimum resolution. This theory is implemented computationally in the `NIFTy` software package used in this thesis.

An updated `NuRadioReco` reconstructor based upon `NIFTy` 8, the most recent version of the package, was tested in section 5. While the updated module would in general be viable, the presence of technical problems was identified which prevented a robust reconstruction. Due to this, the decision was made to implement the added functionality based on a stable branch of the reconstructor which uses `NIFTy` 5.

Implementation of the expanded statistical electric field model is described in section 6. The pre-existing program was expanded with a structure dividing the antennae into groups, with all antennae physically close to each other in a group. One additional likelihood term for each group except the first one was introduced. This term acts as an addition to the general amplitude of the electric field, in theory allowing each channel group to reconstruct a slightly different electric field.

section 7 presents the results obtained with the new module. Initial reconstructions proved unable to accurately identify simulated events. To find the cause of the problem, different diagnostics were performed. In the course

of that, the prior model for the electric field delta is investigated extensively. A prior configuration that produces a promising reconstruction was identified, but the configuration deviates strongly from expectations for the model. The deltas of the electric fields were investigated for the successful results. Lastly, one possible source of problems for the module is identified in the preparatory time-shifting of the voltage traces recorded by the antennae.

While encouraging results were ultimately produced with the new module, time constraints prevented more thorough testing. However, a possible improvement is already proposed, and the approach followed in this thesis has been demonstrated to be a viable one.

Bibliography

Code

[1] *NuRadioMC NuRadioReco nifty_update*. NuRadioReco reconstruction module based on NIFTy 8. This module was tested in this work. URL: https://github.com/nu-radio/NuRadioMC/tree/nifty_update.

[2] *NuRadioMC NuRadioReco RNO_G_energy_reconstruction_multiple_hpols*. NuRadioReco branch with the code developed as part of this thesis. URL: https://github.com/nu-radio/NuRadioMC/tree/RNO_G_energy_reconstruction_multiple_hpols.

References

[3] Denise Boncioli, Daniel Biehl, and Walter Winter. "On the Common Origin of Cosmic Rays across the Ankle and Diffuse Neutrinos at the Highest Energies from Low-luminosity Gamma-Ray Bursts". In: *The Astrophysical Journal* 872.1 (Feb. 2019), p. 110. DOI: 10.3847/1538-4357/aafda7. URL: https://dx.doi.org/10.3847/1538-4357/aafda7.

[4] Kohta Murase, Yoshiyuki Inoue, and Charles D. Dermer. "Diffuse neutrino intensity from the inner jets of active galactic nuclei: Impacts of external photon fields and the blazar sequence". In: *Physical Review D* 90.2 (July 2014). DOI: 10.1103/physrevd.90.023007. URL: https://doi.org/10.1103%2Fphysrevd.90.023007.

[5] Kenneth Greisen. "End to the Cosmic-Ray Spectrum?" In: *Phys. Rev. Lett.* 16 (17 Apr. 1966), pp. 748–750. DOI: 10.1103/PhysRevLett.16.748. URL: https://link.aps.org/doi/10.1103/PhysRevLett.16.748.

[6] V.S. Beresinsky and G.T. Zatsepin. "Cosmic rays at ultra high energies (neutrino?)" In: *Physics Letters B* 28.6 (1969), pp. 423–424. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(69)90341-4. URL: https://www.sciencedirect.com/science/article/pii/0370269369903414.

[7] Markus Ackermann et al. "High-Energy and Ultra-High-Energy Neutrinos". In: (2022). arXiv: 2203.08096 [hep-ph].

[8]   S Adrián-Martínez et al. "Letter of intent for KM3NeT 2.0". In: *Journal of Physics G: Nuclear and Particle Physics* 43.8 (June 2016), p. 084001. DOI: 10.1088/0954-3899/43/8/084001. URL: https://dx.doi.org/10.1088/0954-3899/43/8/084001.

[9]   A.D. et al Avrorin. "Baikal-GVD: status and prospects". In: *EPJ Web Conf.* 191 (2018), p. 01006. DOI: 10.1051/epjconf/201819101006. URL: https://doi.org/10.1051/epjconf/201819101006.

[10]   M. Ackermann et al. "Search for neutrino-induced cascades with AMANDA". In: *Astroparticle Physics* 22.2 (2004), pp. 127–138. ISSN: 0927-6505. DOI: https://doi.org/10.1016/j.astropartphys.2004.06.003. URL: https://www.sciencedirect.com/science/article/pii/S0927650504001197.

[11]   The IceCube Collaboration et al. "Multimessenger observations of a flaring blazar coincident with high-energy neutrino IceCube-170922A". In: *Science* 361.6398 (2018), eaat1378. DOI: 10.1126/science.aat1378. eprint: https://www.science.org/doi/pdf/10.1126/science.aat1378. URL: https://www.science.org/doi/abs/10.1126/science.aat1378.

[12]   R. Abbasi et al. "Observation of high-energy neutrinos from the Galactic plane". In: *Science* 380.6652 (June 2023), pp. 1338–1343. DOI: 10.1126/science.adc9818. URL: https://doi.org/10.1126%2Fscience.adc9818.

[13]   J.A. Aguilar et al. "Transmission of light in deep sea water at the site of the Antares neutrino telescope". In: *Astroparticle Physics* 23.1 (2005), pp. 131–155. ISSN: 0927-6505. DOI: https://doi.org/10.1016/j.astropartphys.2004.11.006. URL: https://www.sciencedirect.com/science/article/pii/S0927650504001902.

[14]   Jessica Avva et al. "An in situ measurement of the radio-frequency attenuation in ice at Summit Station, Greenland". In: *Journal of Glaciology* 61.229 (2015), pp. 1005–1011. DOI: 10.3189/2015jog15j057. URL: https://doi.org/10.3189%2F2015jog15j057.

[15]   S. Barwick et al. "South Polar in situ radio-frequency ice attenuation". In: *Journal of Glaciology* 51.173 (2005), pp. 231–238. DOI: 10.3189/172756505781829467.

[16]   J.A. Aguilar et al. "Design and sensitivity of the Radio Neutrino Observatory in Greenland (RNO-G)". In: *Journal of Instrumentation* 16.03 (Mar. 2021), P03025. DOI: 10.1088/1748-0221/16/03/p03025. URL: https://doi.org/10.1088%2F1748-0221%2F16%2F03%2Fp03025.

[17] Torsten Enßlin. "Information field theory". In: *AIP Conference Proceedings* 1553.1 (Aug. 2013), pp. 184–191. ISSN: 0094-243X. DOI: `10.1063/1.4819999`. eprint: `https://pubs.aip.org/aip/acp/article-pdf/1553/1/184/11454193/184\_1\_online.pdf`. URL: `https://doi.org/10.1063/1.4819999`.

[18] Torsten A. Enßlin. "Information Theory for Fields". In: *Annalen der Physik* 531.3 (2019), p. 1800127. DOI: `https://doi.org/10.1002/andp.201800127`. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1002/andp.201800127`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/andp.201800127`.

[19] *NIFTy software repository*. accessed 23-Oct-2023. URL: `https://gitlab.mpcdf.mpg.de/ift/nifty`.

[20] Christian Glaser et al. "NuRadioReco: a reconstruction framework for radio neutrino detectors". In: *The European Physical Journal C* 79.6 (June 2019). DOI: `10.1140/epjc/s10052-019-6971-5`. URL: `https://doi.org/10.1140%2Fepjc%2Fs10052-019-6971-5`.

[21] *NuRadioReco software repository*. accessed 23-Oct-2023. URL: `https://github.com/nu-radio/NuRadioMC/tree/develop/NuRadioReco`.

[22] C. Glaser et al. "NuRadioMC: simulating the radio emission of neutrinos from interaction to detector". In: *The European Physical Journal C* 80.2 (Jan. 2020). DOI: `10.1140/epjc/s10052-020-7612-8`. URL: `https://doi.org/10.1140%2Fepjc%2Fs10052-020-7612-8`.

[23] U.F. Katz and Ch. Spiering. "High-energy neutrino astrophysics: Status and perspectives". In: *Progress in Particle and Nuclear Physics* 67.3 (2012), pp. 651–704. ISSN: 0146-6410. DOI: `https://doi.org/10.1016/j.ppnp.2011.12.001`. URL: `https://www.sciencedirect.com/science/article/pii/S0146641011001189`.

[24] "Measurement of the multi-TeV neutrino interaction cross-section with IceCube using Earth absorption". In: *Nature* 551.7682 (Nov. 2017), pp. 596–600. DOI: `10.1038/nature24459`. URL: `https://doi.org/10.1038%2Fnature24459`.

[25] J. A. Formaggio and G. P. Zeller. "From eV to EeV: Neutrino cross sections across energy scales". In: *Reviews of Modern Physics* 84.3 (Sept. 2012), pp. 1307–1341. DOI: `10.1103/revmodphys.84.1307`. URL: `https://doi.org/10.1103%2Frevmodphys.84.1307`.

[26] Amy Connolly, Robert S. Thorne, and David Waters. "Calculation of high energy neutrino-nucleon cross sections and uncertainties using the Martin-Stirling-Thorne-Watt parton distribution functions and implications for future experiments". In: *Physical Review D* 83.11 (June 2011). DOI: 10.1103/physrevd.83.113009. URL: https://doi.org/10.1103%2Fphysrevd.83.113009.

[27] J. A. Aguilar et al. "Reconstructing the neutrino energy for in-ice radio detectors". In: *The European Physical Journal C* 82.2 (Feb. 2022). DOI: 10.1140/epjc/s10052-022-10034-4. URL: https://doi.org/10.1140%2Fepjc%2Fs10052-022-10034-4.

[28] D. García-Fernández, A. Nelles, and C. Glaser. "Signatures of secondary leptons in radio-neutrino detectors in ice". In: *Physical Review D* 102.8 (Oct. 2020). DOI: 10.1103/physrevd.102.083011. URL: https://doi.org/10.1103%2Fphysrevd.102.083011.

[29] C. Welling et al. "Reconstructing non-repeating radio pulses with Information Field Theory". In: *Journal of Cosmology and Astroparticle Physics* 2021.04 (Apr. 2021), p. 071. DOI: 10.1088/1475-7516/2021/04/071. URL: https://doi.org/10.1088%2F1475-7516%2F2021%2F04%2F071.

[30] S.W. Barwick et al. "Observation of classically 'forbidden' electromagnetic wave propagation and implications for neutrino detection." In: *Journal of Cosmology and Astroparticle Physics* 2018.07 (July 2018), pp. 055–055. DOI: 10.1088/1475-7516/2018/07/055. URL: https://doi.org/10.1088%2F1475-7516%2F2018%2F07%2F055.

[31] C. Deaconu et al. "Measurements and modeling of near-surface radio propagation in glacial ice and implications for neutrino experiments". In: *Physical Review D* 98.4 (Aug. 2018). DOI: 10.1103/physrevd.98.043010. URL: https://doi.org/10.1103%2Fphysrevd.98.043010.

[32] S. Barwick et al. "South Polar in situ radio-frequency ice attenuation". In: *Journal of Glaciology* 51.173 (2005), pp. 231–238. DOI: 10.3189/172756505781829467.

[33] Robert L. Hawley et al. "Greenland Ice Sheet Elevation Change: Direct Observation of Process and Attribution at Summit". In: *Geophysical Research Letters* 47.22 (2020), e2020GL088864. DOI: https://doi.org/10.1029/2020GL088864. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2020GL088864. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020GL088864.

[34] Christoph Welling. *Energy Reconstruction for Radio Neutrino Detectors.* PhD thesis. Available at `https://ecap.nat.fau.de/wp-content/uploads/2022/07/PhD_Thesis__Welling_final.pdf`, accessed 10-Oct-2023. 2022.

[35] RNO-G collaboration. *Wiki for the RNO-G collaboration.* internal wiki page, accessed 23-Oct-2023.

[36] Torsten A. Enßlin. *Lecture on Information Theory & Information Field Theory.* accessed 23-Oct-2023. URL: `https://wwwmpa.mpa-garching.mpg.de/~ensslin/lectures/lectures.html`.

[37] "Showcasing the Correlated Field model". In: (). Accessed: 2023-Oct-18. URL: `https://ift.pages.mpcdf.de/nifty/user/getting_started_4_CorrelatedFields.html`.

[38] M. Selig et al. "NIFTY - Numerical Information Field Theory. A versatile PYTHON library for signal inference". In: *Astronomy & Astrophysics* 554, A26 (June 2013), A26. DOI: `10.1051/0004-6361/201321236`. arXiv: `1301.4499 [astro-ph.IM]`.

[39] Jakob Knollmüller and Torsten A. Enßlin. *Metric Gaussian Variational Inference.* 2020. arXiv: `1901.11033 [stat.ML]`.

A   Parameters of the MC data set

The reconstructions in this work were tested using MC generated neutrino events and detector responses.

For the neutrino event generation, `NuRadioMC.EvtGen.generator generate_eventlist_cylinder()` was used to generate 10 000 events with a cylindrical geometry and an initial neutrino energy of 1 EeV. Fiducial detector volume was assumed as a cylinder with a height of 2.7 km, and a radius of 4 km. The neutrinos were generated following a log uniform energy spectrum, and with a uniform distribution on all three neutrino flavor states and all three anti-neutrino flavor states, with both charged current and neutral current interactions are considered.

For the event simulation, default detector descriptions for RNO-G were used.

Lookup tables for the event vertex reconstruction were generated with `NuRadioReco/modules/neutrinoVertexReconstructor/ create_lookup_table.py`.

69

B   Danksagung

An dieser Stelle möchte ich meine Dankbarkeit an all jene ausdrücken, die mich in der Erstellung dieser Arbeit unterstützt haben. Mein besonderer Dank gilt

Prof. Dr. Anna Nelles, dafür mit wie viel Engagement sie mich während dieser Arbeit betreute.

Dr. Christoph Welling, dafür dass er jederzeit bereit stand, meine Fragen zu beantworten und mir bei der Lösung technischer Probleme zu helfen.

Nadja Lessing, für ihre ausdauernde Unterstützung und ihre Kommentare und Korrekturen.

C   Erklärung zur ordnungsgemäßen Abfassung der vorliegenden Arbeit

Hiermit versichere ich, dass ich die vorliegende Arbeit selbständig und ohne unerlaubte Hilfe verfaßt habe.

Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt und alle wörtlich oder dem Sinn nach aus anderen Texten entnommenen Stellen als solche kenntlich gemacht.

Dies gilt für gedruckte Texte ebenso wie für Texte aus dem Internet.

Die Arbeit wurde in keiner anderen Lehrveranstaltung (weder an der FAU noch an einer anderen Hochschule) in der vorliegenden oder in einer modifizierten Form vorgelegt.

Mir ist bewusst, dass jeder Verstoß gegen diese Erklärung zu einer Benotung der Arbeit mit „nicht ausreichend" führt.

Ort, Datum   Unterschrift