Deep learning reconstruction of the neutrino direction and energy from in-ice radio detector data

C. Glaser\textsuperscript{a}, S. McAleer\textsuperscript{b}, S. Stjärnholm\textsuperscript{a}, P. Baldi\textsuperscript{b}, S. W. Barwick\textsuperscript{c}

\textsuperscript{a}Uppsala University Department of Physics and Astronomy, Uppsala SE-75220, Sweden.
\textsuperscript{b}Department of Information and Computer Science, University of California, Irvine, CA 92697, USA.
\textsuperscript{c}Department of Physics and Astronomy, University of California, Irvine, CA 92697, USA.

Abstract

Ultra-high-energy (UHE) neutrinos (>10\textsuperscript{16} eV) can be measured cost-effectively using in-ice radio detection, which has been explored successfully in pilot arrays. A large radio detector is currently being constructed in Greenland with the potential to measure the first UHE neutrino, and an order-of-magnitude more sensitive detector is being planned with IceCube-Gen2. For such shallow radio detector stations, we present an end-to-end reconstruction of the neutrino energy and direction using deep neural networks (DNNs). The DNN determines the energy with a standard deviation of a factor of two around the true energy, which meets the science requirements of UHE neutrino detectors. For the first time, we are able to predict the neutrino direction well for all event topologies including the complicated electron neutrino charged-current (ν\textsubscript{e}-CC) interactions, a significant improvement compared to previous approaches. The obtained angular resolution follows a Gaussian distribution with σ ≃ 0.6°(0.8°) with extended tails that push the 68% quantile to 4°(5°) for non-ν\textsubscript{e}-CC and ν\textsubscript{e}-CC interactions, respectively. This highlights the advantages of DNNs for modeling the complex correlations in radio detector data, thereby enabling measurement of neutrino energy and direction.

1. Introduction

The detection of ultra-high-energy (UHE) neutrinos is a key to solving the 100-year-old mystery of the origin of cosmic rays and is one of the crucial milestones for astroparticle physics [1, 2]. Their detection gives access to the most violent phenomena in the universe, those that happen in the vicinity of supermassive black holes (active galactic nuclei), in neutron star mergers, or in gamma-ray bursts. Furthermore, it allows for fundamental measurements of neutrino cross-sections and flavor ratios at energies beyond the reach of Earth-based accelerators like the LHC [3, 4].

A cost-efficient way to measure these UHE neutrinos above 30 PeV of energy is via a sparse array of radio antenna stations installed, for instance, in the Arctic or Antarctic ice [5, 6, 7, 8, 9, 10]: A neutrino interaction in the ice generates a few-nanoseconds-long radio flash that can be detected from kilometer-long distances due to the large attenuation length of radio signals in ice. Because of the small flux, no UHE neutrino has been observed yet, but the technology has already been shown to work reliably with small test-bed arrays such as ARA and ARIANNA [8, 5]. With the Radio Neutrino Observatory in Greenland (RNO-G) a much larger detector is being constructed at the moment [11] and an order-of-magnitude more sensitive radio detector is foreseen for IceCube-Gen2 [9, 10].

With the first detection of a UHE neutrino on the horizon for the next years, the development of reconstruction methods becomes increasingly important. In addition, a good estimation of the energy and pointing resolution for different detector designs is crucial for planning IceCube-Gen2, which is happening at the moment. Two different station designs have been established: In the first, a deep design (as explored by ARA [12]) antennas are placed into narrow boreholes down to a depth of up to 200 m, thereby increasing the sensitivity to neutrinos per detector station but also increasing the costs per station and limiting the choice of available antennas due to the narrow borehole. The second design is a shallow detector station (as explored by ARIANNA [13]) with high-gain LPDA antennas installed a few meters below the surface. The Radio Neutrino Observatory in Greenland (RNO-G) combines both designs into hybrid detector stations. The radio detector of IceCube-Gen2 foresees a hybrid array of shallow-only stations interspersed with hybrid stations [10].

This work focuses on a shallow station design as shown in Fig. 1, which has been explored by the ARIANNA test-bed detector on the Ross Ice Shelf and at the South Pole [8]. Each station consists of 4 LPDA antennas installed at a depth of just a few meters below the snow surface, and 1 dipole antenna installed at a depth of 10 m to 15 m in a narrow borehole. These antennas observe the ice below for neutrino interactions. The dipole antenna was added
to help with the reconstruction of the neutrino properties. At this depth, the antenna will observe two signals: one from a direct path to the antenna, and a second delayed signal from a reflection off the surface. The time difference between these two signals provides information about the distance to the neutrino interaction, which is important for estimating the neutrino energy [14]. The exact depth is a compromise between the fraction of events that show a Direct and Reflected (D’n’R) signature (the shallower the better), the resolution of those events where a D’n’R signature is visible (the deeper the better), and deployment effort (the shallower the better) and is still being optimized. In the following, we set the depth to 15 m; however, previous studies have indicated that 10 m also will lead to energy resolutions that match the science requirements [14]. In a future study, the work presented here can be used to study the impact of dipole depth on energy resolution in detail.

Several aspects of the reconstruction of the neutrino direction and energy have been studied already and partly probed with in situ measurements. The reconstruction of the neutrino direction and energy requires measurement of the distance to the neutrino vertex, the signal arrival direction, the viewing angle, and the signal polarization, as well as a good understanding of the ice to correct for the bending of signal trajectories due to the changing index-of-refraction in the upper O(100 m) of the ice sheet [15]. For a shallow detector station, the measurement of the low-level parameters signal arrival direction (including the correction of ice propagation effects) and polarization has been probed experimentally at the South Pole using a transmitter lowered up to 1.7 km deep into the ice [16, 17, 18] as well as by measuring the radio emissions of cosmic rays [19]. In addition, a novel method for determining the vertex distance was developed and tested in an in situ measurement on the Ross Ice Shelf [14]. This method uses the time difference between the direct and reflected-at-the-surface signals measured in the dipole. For a deep detector component, a similar method has been developed [20].

To determine the neutrino direction and energy at low signal-to-noise ratios, the forward folding method was developed [21], which fits an analytic model of the radio signal directly to the antennas’ observed voltage measurements. The method has been applied successfully to shallow and deep detector components [16, 18, 22]. In a MC study using NuRadioMC [23], which includes a full trigger simulation and modeling of the signal chain and noise interference, the angular resolution was determined to be approx. 3° for all triggered events of a shallow detector station [16, 18]. For a deep detector, a similar resolution could only be obtained after a quality cut on the signal strength of the horizontally polarized antennas which significantly reduced the fraction of usable events [22]. A caveat of these analyses is that so far only hadronic particle cascades have been considered; these have a predictable shape and corresponding radio emission. For electron neutrino charged-current (νe-CC) interactions, the generated electron generates an electromagnetic shower that is subject to the LPM effect. For neutrino energies of approx. $10^{18}$ eV and beyond, this leads to several spatially displaced sub-showers which in turn lead to more complex radio signal shapes [23] and complicates the reconstruction. Current reconstruction algorithms of the neutrino direction do not work for or fail to produce reasonable results for the event class of νe-CC interactions. Because νe-CC interactions dominate the number of observable neutrinos, this paper will address this significant caveat of current methods.

The reconstruction of the neutrino direction and energy from observable radio flashes is a complex problem: All information about the neutrino is compressed in a few-nanoseconds-long radio flash that is observed in just a few antennas. As a consequence, the development of traditional analysis approaches as described above is a very time-consuming process and the reconstruction algorithms often do not take into account all available information. In this article, we take a different approach: training deep neural networks [24, 25] to extract the neutrino properties of interest directly from the raw data. This approach was made possible by the development of the NuRadioMC code which precisely simulates the expected signals from neutrinos [23] as well as the corresponding detector response [21]; this development allowed us to generate the needed training datasets. A preliminary version of the analysis was presented at the last International Cosmic Ray Conference (ICRC2021) [26, 27].

2. Dataset generation

We use the NuRadioMC code [23] to create a large dataset of around 40 million recorded neutrino events that...
correspond to a data volume of 1.4 TB. We simulate a shallow radio detector station at the South Pole with the same configuration and trigger settings as foreseen for IceCube-Gen2 [10]: The station comprises two parallel pairs of downward-facing LPDA antennas 2 m below the snow surface and an additional dipole antenna at a depth of 15 m below the snow surface. The detector is triggered by requiring a high and low-amplitude threshold crossing of the signal with an additional requirement of a time coincidence of two triggers out of the four LPDA antennas. The bandwidth of the trigger channels is reduced to 80-150 MHz in order to increase the sensitivity to neutrinos [29] but the full frequency information within 80-800 MHz is recorded for offline analysis. The trigger threshold is adjusted to yield a trigger rate on thermal noise fluctuations of 100 Hz, as the data rate is the limiting factor in current and future experiments and thus sets the trigger threshold. We note that with a deep learning-based second-stage trigger that can run on the autonomous detector stations the data rate can be reduced significantly without impacting the neutrino efficiency [30].

We generate neutrino interactions with energies from $10^{17}$ eV to $10^{19}$ eV. The neutrino interactions are distributed uniformly in the ice around the detector with random incoming directions. We created three datasets with a comparable number of triggered events. The first dataset only contains hadronic showers – i.e., particle cascades initiated by the breakup of the nucleus – which are initiated by all neutrino interactions in the ice. The Askaryan signal is generated using the frequency domain parameterization of Alvarez2009 [31, 32] as implemented in NuRadioMC [23]. This setup was used in previous analyses that used the forward folding technique [33, 22]. This dataset is referred to as Alvarez2009 (had.) in the following.

The second dataset also contains only hadronic showers but uses a more precise calculation of the Askaryan radiation: The ARZ2020 model calculates the Askaryan radiation from a library of charge-excess profiles of the particle cascades in the time domain [23], following the approach in [34, 35], which matches a microscopic MC simulation at a level of ±3%. This dataset is referred to as ARZ2020 (had.) in the following.

The third dataset contains electron neutrino charged-current ($\nu_e$-CC) interactions that produce both a hadronic shower as well as an additional electromagnetic shower through the outgoing electron. We again use the ARZ2020 model with the shower library available in NuRadioMC, which allows a proper simulation of the LPM effect that leads to a spatial delay of the electromagnetic shower with respect to the hadronic shower, and for increasing energy also to multiple spatially displaced sub-showers. We expect that this will complicate the reconstruction performance. Due to this additional complication, electron charged-current interactions were so far ignored in previous analyses. This dataset is referred to as ARZ2020 (had. + EM) in the following.

The observed radio signals only depend on the shower properties. Therefore, we will train the deep neural network to determine the shower energy instead of the neutrino energy. The amount of energy transferred from the neutrino into the shower is a stochastic process and can be estimated from theory. For non-$\nu_e$-CC interactions, only a part of the neutrino energy is transferred to the particle shower. This stochastic process also introduces an irreducible uncertainty into the conversion from hadronic shower energy back to the neutrino energy which amounts to a factor of two [14]. This provides a target resolution for the shower energy. For $\nu_e$-CC interactions, the complete neutrino energy is deposited in the ice but the relative distribution into the hadronic and electromagnetic shower undergoes the same stochastic process. Here, we set the shower energy equal to the neutrino energy. We note that due to the LPM effect, the showers do not develop in phase which in practice often means that only a part of the shower energy is measured.

The likelihood of triggering the detector increases quickly with neutrino energy as the signal amplitude increases linearly with energy. We made an effort to simulate more low-energy neutrinos to obtain a similar number of triggered events across all energies, but our dataset is still biased toward high energies, as can be seen in Fig. 2. In this figure, we show the number of triggered events as a function of the shower energy which shows another interesting property. For the ARZ2020 (had. + EM) dataset the shower energy equals the neutrino energy, and the number of triggered events steadily increases from $10^{17}$ eV to $10^{19}$ eV, as expected. In contrast, for the other two datasets, the distribution extends to lower energies because the shower energy is always lower than the neutrino energy. For the same reason, the number of events decreases at the high-energy end because it is unlikely that the complete neutrino energy is deposited into the hadronic shower. We expect that this uneven event distribution has a negative impact on the energy reconstruc-

Figure 2: Number of events in the 3 datasets as a function of the shower energy.
tion at low energies due to an underrepresentation in the training dataset. In total the Alvarez2009 (had.) dataset consists of $8.2 \times 10^6$ events, the ARZ2020 (had.) dataset consists of $4.1 \times 10^6$ events, and the ARZ2020 (had. + EM) consists of $5 \times 10^6$ events.

The training data consists of the recorded voltages as a function of time of the 5 antennas. We simulated a length of the time trace of 256 ns with a time binning of 0.5 ns which results in 512 samples per antenna. The shape of one event is $(5, 512)$. Two examples are shown in Fig. 3.

2.1. Simplified datasets

To assess the performance of the deep learning reconstruction, we generate three simplified datasets to resemble a simple traditional analysis. We characterize the signal in each antenna by its maximum amplitude and time which we extract from the maximum of the Hilbert envelope of each antenna’s simulated signal trace. This is similar to a deep learning-based analysis performed for the ARA experiment [36].

The first simplified dataset extracts the parameters from the dataset that is used in the nominal analysis where noise interference impacts the precision of the extracted parameters. This results in 10 parameters per event; i.e., 5 amplitudes and 5 times.

The second simplified dataset is similar to the first one but uses noiseless waveforms to extract the signal time and amplitude. This dataset can be used to assess how well a reconstruction could perform with a perfect measurement of the signal time and amplitude.

The third simplified dataset analyzes the noiseless waveforms from both signal trajectories separately. Each antenna sees two signals: one signal from a direct trajectory to the antenna, and another one from a refracted or reflected-off-the-surface trajectory. In practice, the two signals often interfere with each other and cannot be easily separated. This dataset contains 20 values per event and allows us to assess how much information on the neutrino direction and energy is theoretically available in the data.

3. Direction Reconstruction

In this work, we use a neural network architecture that follows the VGG model [37]. The general idea of this architecture is the repeated application of convolution layers followed by pooling. After a few of these blocks, the data is flattened and several fully connected layers are added to shrink the information into the output nodes. The VGG model was developed for image recognition but one can interpret our data as one-dimensional images where the voltage as a function of time corresponds to the color amplitude as a function of spatial pixel position.

The usage of convolutional layers matches the physical properties of the data. The pulse shape contains a lot of information about the neutrino properties. Traditionally, template matching techniques have been used to identify neutrino-induced radio pulses [8]. Convolutional layers can be thought of as a more versatile and capable template where the subsequent application of small filters can match a multitude of different waveform shapes. This allows the neural network to identify and characterize the signal pulses that are invariant in time (cf. Fig. 3). Because the signals’ relative time and amplitude differences

Figure 3: Two examples of the training dataset. The traces show the recorded voltages in the 5 antennas as a function of time. The upper 4 panels are the signal in the 4 LPDAs, and the bottom panel shows the signal in the dipole antenna. Left: example at a particularly high signal-to-noise ratio, where the direct and reflected pulse is clearly visible in the dipole antenna. Right: a more typical example at a low signal-to-noise ratio.
and the singularities at the poles. Therefore, we add an ordinate due to the correlation between the two angles compared to the zenith and azimuth angle in spherical coordinates. For example, in the case of images.

An overview of the neural network architecture is presented in Fig. 4, with a more detailed description presented in Tab. 1. In the following, we first describe the best-performing model architecture to provide a good overview of the model before describing how we arrived at this particular model.

The convolutional blocks consist of three convolutional layers, followed by pooling layers. The convolutional layers are configured to use padding, such that the dimensionality does not change following a convolutional layer. The filter size of the convolutional layers is (1, 5), which will result in the model scanning each antenna trace separately, with a filter width of 5 samples. The average pooling layer has a stride of 4, meaning it will decrease the trace’s dimensions by a factor of 4 by taking the average of blocks of 4 samples. For example, in C1, the trace’s length is reduced from 512 to 128 samples following the pooling layer. The number of filters is doubled in each block compared to the previous block. This is a very common technique in developing convolutional neural network models with pooling layers and can be seen as a way to mitigate underfitting due to the risk of the model becoming too simple if the dimensions decrease a lot, following the pooling layers.

After four convolutional blocks, we apply a batch normalization and flatten the data tensor to a vector before applying a block of fully connected layers. The dense block consists of 6 fully connected layers. The size of the dense layers decreases for layers further along the network, which helps to reduce the number of parameters as well as allowing the model to narrow down to the 3 output neurons. The dense layers compile the information gathered by the convolutional layers, especially in regard to the time difference between the pulses in the 5 antenna traces.

We encode the neutrino direction as a three-dimensional cartesian unit vector that points in the direction of the neutrino origin. This is often the better choice compared to the zenith and azimuth angle in spherical coordinates due to the correlation between the two angles and the singularities at the poles. Therefore, we add an L2-normalization layer to make sure that the network prediction is a unit vector.

3.1. Training and Optimization of Network Architecture

We reserved 300,000 events for an independent test dataset. We used 87% of the remaining data for the training dataset and 13% for the validation dataset. The datasets Alvarez2009 (had.), ARZ2020 (had.), and ARZ2020 (had. + EM) contained $8.2 \times 10^6$ events, $4.1 \times 10^6$ events, and $5 \times 10^6$ events respectively. The training dataset is so large that it does not fit into memory. To avoid slowing down the training process due to I/O and data preprocessing times, we developed a data pipeline based on the tensorflow.data class, which preloads and preprocesses the data on several cores. We used an NVIDIA Quadro RTX 6000 GPU to perform the training.

To train the model, we use the optimizer Adam\cite{kingma2014adam} with a learning rate of $5 \times 10^{-3}$ and the mean absolute error (MAE) error as loss function. The latter choice is motivated by the expectation that a certain percentage of the events will have too poor a signal quality to allow for a good reconstruction. We don’t want these events to dominate the training. In a future analysis, quality cuts will be developed to remove those events from the dataset.

Through many experiments, we optimized the network inference performance via an iterative procedure by testing different hyperparameters but staying within the general VGG-like architecture of convolutional + pooling blocks followed by fully connected layers. The best-performing network was set as the one to be compared against for further experiments. During the optimizations, model parameters such as learning rate, number of convolutional and dense layers, activation functions, batch normalization use, loss functions, choice of the optimizer as well as convolutional layer stride length, filter size, and filter amount were tested in order to find the best-performing network structure and hyperparameters. We tested the parameters multiple times during the optimization to make sure that the values were not derived from a previous minimum that had moved when other parameters were tuned. When additional experimentation was performed with little or no improvements to the model performance, the optimization of the network was done. We considered other factors as well such as the number of trainable model parameters, where a model with a smaller number of trainable parameters was chosen when deciding between models that had equal performance. This made the model less complex.
Table 1: Detailed structure of the model that was developed. The labels correspond to the labels used in Fig. 4. The dimensions specify the output dimensions of each layer, and the parameters specify the number of trainable parameters for each layer. Note that $B$ has 512 trainable parameters, but also 512 non-trainable parameters.

| Label | Layer | Dimensions       | Parameters |
|-------|-------|------------------|------------|
| $I$   | Input layer | $(5, 512, 1)$ | 0          |
| $C_1$ | Conv2D (32 filters) | $(5, 512, 32)$ | 192        |
| $C_1$ | Conv2D (32 filters) | $(5, 512, 32)$ | 5 152      |
| $C_1$ | Conv2D (32 filters) | $(5, 512, 32)$ | 5 152      |
|       | AveragePooling2D | $(5, 128, 32)$ | 0          |
| $C_2$ | Conv2D (64 filters) | $(5, 128, 64)$ | 10 304     |
| $C_2$ | Conv2D (64 filters) | $(5, 128, 64)$ | 20 544     |
| $C_2$ | Conv2D (64 filters) | $(5, 128, 64)$ | 20 544     |
|       | AveragePooling2D | $(5, 32, 64)$  | 0          |
| $C_3$ | Conv2D (128 filters) | $(5, 32, 128)$ | 41 088     |
| $C_3$ | Conv2D (128 filters) | $(5, 32, 128)$ | 82 048     |
| $C_3$ | Conv2D (128 filters) | $(5, 32, 128)$ | 82 048     |
|       | AveragePooling2D | $(5, 8, 128)$  | 0          |
| $C_4$ | Conv2D (256 filters) | $(5, 8, 256)$  | 164 096    |
| $C_4$ | Conv2D (256 filters) | $(5, 8, 256)$  | 327 936    |
| $C_4$ | Conv2D (256 filters) | $(5, 8, 256)$  | 327 936    |
|       | AveragePooling2D | $(5, 2, 256)$  | 0          |
| $B$   | BatchNormalization | $(5, 2, 256)$  | 512        |
| $F$   | Flatten          | $(2560)$        | 0          |
| $D_1$ | Dense (1024 units) | $(1024)$        | 2 622 464  |
| $D_2$ | Dense (1024 units) | $(1024)$        | 1 049 600  |
| $D_3$ | Dense (512 units) | $(512)$         | 524 800    |
| $D_4$ | Dense (256 units) | $(256)$         | 131 328    |
| $D_5$ | Dense (128 units) | $(128)$         | 32 896     |
| $D_6$ | Dense (3 units)  | $(3)$           | 387        |
| $N$   | Normalization    | (3)             | 0          |
| $O$   | Output layer     | (3)             | 0          |

**Total:** 5 449 027
while also decreasing the required time for training and inference.

### 3.2. Reconstruction resolution and dependence on event properties

In the following, we show the results of the direction reconstruction for the three different datasets **Alvarez2009 (had.)**, **ARZ2020 (had.)**, and **ARZ2020 (had. + EM)**. Fig. 5 shows histograms of the space angle difference $\Delta \Psi$ between the true and reconstructed neutrino direction. We quantify the angular resolution by quoting the position of the 68% quantile $\sigma_{68}$ which is also specified in the graphs. The point spread function can be described well with a Moffat/King function that can be thought of as a Gaussian with extended tails [30] and is given by:

$$f(x; A, \sigma, \gamma) = A \frac{x}{2\pi \sigma^2} \left(1 - \frac{1}{\gamma}\right) \left[1 + \frac{1}{2\gamma} \frac{x^2}{\sigma^2}\right]^{-\gamma}$$

(1)

We find that the narrow part of the distribution is described with a sigma of 0.5° to 0.6° for the hadronic datasets which increases to 0.8° for the more challenging $\nu_e$-CC dataset. Due to the tails of the distribution, the 68% quantiles are at 4° to 5°. This finding highlights the advantage of a DNN for event reconstruction. The complex correlations between neutrino direction and observed radio flashes of $\nu_e$-CC interactions are learned well and are only slightly worse than for the simpler non-$\nu_e$-CC interactions. This is a significant improvement over previous reconstruction methods that fail to reconstruct $\nu_e$-CC interactions.

The average resolution of the training dataset strongly depends on the event distribution. A better way of showing the resolution is by quoting the resolution in bins of neutrino energy and/or signal-to-noise ratio (SNR) which is presented in Figs. 6 and 7. We define the SNR of an event as the maximum of the absolute value of the LPDA traces, divided by the noise floor, which is approximately 10 $\mu$V. We observe an increase in performance with increasing energy and SNR as expected.

The three datasets’ energy dependence gives interesting insights. Overall, the **ARZ2020 (had.)** dataset performs best, especially at higher energies. The **Alvarez2009 (had.)** dataset performs worse at almost all energies which was a surprise at first. Our initial expectation was a better performance because the simpler **Alvarez2009** emission model should be easier to model which is indeed the case for traditional reconstruction techniques [16, 22]. However, it seems that the more precise **ARZ2020** emission model generates more information that the neural network is able to learn. This highlights the advantages of deep neural networks for event reconstruction. The **ARZ2020 (had. + EM)** dataset performs slightly worse at larger neutrino energies where the LPM effect becomes relevant and the stochasticity in the shower development complicates the reconstruction. Nevertheless, the resolution is just one to two degrees worse than for the simpler hadronic showers which is encouraging as traditional reconstruction techniques failed to reconstruct this event class with good resolution [22, 16]. At low energies where the LPM effect is negligible, the performance is similar to the **ARZ2020 (had.)** dataset. The lowest energy bins are difficult to interpret because of low statistics in these bins. Another factor that influences that dependence on the energy is the different statistics in the training dataset. Part of the reason for the worse performance at lower energies might just be an underrepresentation in the training dataset.

The resolution as a function of the signal-to-noise ratio of Fig. 7 shows a similar trend. Ignoring the first bin which has poor statistics, the resolution improves with increasing SNR. The performance of the **ARZ2020 (had.)** dataset is the best over the entire range of SNR values, followed by the **Alvarez2009 (had.)** dataset and the **ARZ2020 (had. + EM)** dataset.
3.3. Comparison to previous results

A comparison with the work of [16, 18] indicates that our DNN reconstruction still has room for improvement. The work of [16, 18] used the forward folding technique [21] and found a 3° resolution for hadronic showers (Alvarez2009 (had.)) with little energy dependence. The analysis however assumed that the neutrino vertex position was known and only simulated uncertainties of the core position along the signal trajectory, whereas our DNN uses only the raw data without any additional information. But as this is a reasonable assumption, we expect that the limit of the DNN reconstruction is not yet reached and can be further improved in future work.

4. Energy Reconstruction

The structure of the neural network used for energy reconstruction is very similar to the one used for direction reconstruction, with the most notable difference of only one output node representing the reconstructed shower energy. In all other regards, the models used for direction and energy reconstruction are identical, which illustrates the benefits of using machine learning for these kinds of tasks. We perform the training in the logarithm of the shower energy, we use the mean squared error (MSE) as the objective function and otherwise use the same settings as before. In addition, we utilized the ReduceLROnPlateau callback in Keras to reduce the learning rate when the loss stagnated.

After the neural network was trained until convergence, we evaluate the performance on the independent test dataset. We show the resulting energy resolution for all three datasets in Fig. 8. We find similar results for all three datasets with an average standard deviation of approx. 0.3 in the $\log_{10}(E_{\text{shower}})$ which translates to a factor of two on a linear scale. This uncertainty is almost identical to the intrinsic uncertainty from inelasticity fluctuations that we defined as the target resolution. The stat boxes in Fig. 8 also list the 68% quantiles around the median which are smaller than the standard deviation, indicating the presence of non-Gaussian tails. In future work, these events could be identified and removed through quality cuts. The resolution of the ARZ2020 (had. + EM) dataset is only slightly worse (STD = 0.33 vs. 0.31), although the LPM effect and interference with the hadronic shower make this reconstruction more challenging. This result again highlights the usefulness of deep neural networks for such a complex reconstruction task.

In Fig. 9 we show the correlation between predicted and true shower energy. All three datasets show a clear correlation. For the had. datasets, a significant bias for shower energies below $10^{17.5}$ eV is visible. At low energies, the DNN overestimates the true energy. Also at the high-energy end, a systematic underestimation of the energy is visible. The scatter around the identity line for the ARZ2020 (had. + EM) dataset is a bit larger, and a similar bias toward low energies is present but less visible from the figure because of the more restricted shower energy range.

To gain further insight into the energy dependence, we show the mean and standard deviation in bins of the true shower energy in Fig. 10. We find only a weak energy dependence for the standard deviation but a large energy-dependent bias. At low shower energies, the shower energy is overpredicted and at the high-energy end it is underpredicted. In the following, we discuss two approaches to reduce the bias.

In the data used for training, validating, and testing the neural networks, the events are not distributed uniformly across the energy range, as seen in the lower panels of Fig. 9. This will affect the network such that it becomes good at reconstructing events at energies where there are many events in the training dataset, and it will perform worse for events where there are few events. This affects low-energy events the most, as these are the rarest. One possible solution to this bias against low-energy events is to use weights in the loss function, such that it puts higher weights on the events that are rarer when compared to
Figure 8: Energy difference histograms for the 3 datasets.

Figure 9: Predicted energy versus true energy for the 3 datasets. Each column is normalized such that the values sum to unity. The lower panels show the number of events in each energy bin.

more common events. This loss function weight is used for both training and validation. As an example, if a shower energy bin has half the number of events when compared to another shower energy bin in Fig. 2, then the network will weigh the loss function twice as much for events in that bin as compared to the other bin. This makes it so that the network’s weights also take in consideration the less common events, even though they represent a smaller proportion of the training events. We restrict the maximum weighting factors to a range between 1 and 10 to avoid single events dominating a batch.

Fig. 11 shows equivalent plots to Fig. 10 for the model that used re-weighting of the loss function during training. It is evident that using re-weighting of the loss function reduces the energy-dependent bias although it does not remove it completely. However, it also makes the network perform slightly worse. This is because the network is now adapting to events that are not as common instead of focusing on the most common events, which leads to the more common events not being reconstructed as well as previously. In future work, we recommend making a larger effort for an equal representation of all energies in the datasets to mitigate this effect further.

Another method to make the bias less significant is to adjust the predicted shower energies for the bias on an event-by-event basis. Once a prediction has been made using the ordinary model, the average bias values at the reconstructed energy from Fig. 10 are used as an estimate of the bias. This bias is then subtracted from the predicted energy, which yields a bias-adjusted value of the predicted energies. However, because the bias depends on the true shower energy which is unknown, and because the predicted shower energy has significant uncertainties, this bias correction will have significant uncertainties which we see in the resulting distributions shown in Fig. 12. As we expect, the bias is significantly reduced and smaller than
what we obtained using the re-weighting approach. However, the uncertainty (standard deviation) is also increased substantially which disfavors using this method.

4.1. Simplified datasets

To gain further insight into what performance we can expect from a DNN reconstruction, we investigate what performance can be reached using the simplified datasets described previously. In these datasets, the waveform information is reduced to the maximum amplitude and its time. For the first simplified dataset, the parameters are extracted from the noisy waveforms of the Alvarez2009 (had.) dataset, which has been used in the nominal analysis described above. For the second and third simplified datasets, the parameters are extracted from noiseless waveforms of the same Alvarez2009 (had.) dataset. This is of course unrealistic but tests if a perfect measurement would at all have enough information to determine the shower energy (see Sec. 2.1 for a detailed description).

We use the same network architecture as before, except for a smaller input size along with fewer and smaller layers. The network weights are trained until convergence and we made sure that no overfitting takes place. The resulting reconstruction performance is shown in Fig. 13. The first simplified dataset (with noise) does not work at all. The network always predicts the same average energy in the dataset. This shows that using low-level experimental data with the full waveform information has huge benefits.

The second simplified dataset (noiseless) shows a better performance. A clear correlation between true and reconstructed energy is visible but the performance is worse than what we obtain with the full model on the noisy dataset. This shows not only that the DNN is capable of dealing with noisy waveforms but also that it extracts additional information from the pulse shapes apart from maximum amplitudes and positions.

The third simplified dataset shows a good correlation between true and predicted energy with a small scatter and only a slight bias toward the highest studied energies. Our interpretation of this result is that a high-precision measurement of the amplitudes and arrival times of both signal trajectories provides enough information to determine the shower energy precisely. Again, we see a deterioration at the edges of the available energy range which we
attribute to the training dataset’s poor statistics at these energies.

This study shows the usefulness and necessity of an advanced reconstruction algorithm to reconstruct the shower energy from the measured radio data. Characterization of the waveform by its maximum amplitude and time is not sufficient. The DNN presented in this work is a suitable reconstruction algorithm for radio detector data that considers all available information.

5. Summary and Outlook

We presented the first end-to-end reconstruction of the neutrino direction and energy from shallow radio detector data using a deep neural network (DNN). The only information available to the DNN was the raw voltage output of the five antenna receivers. The best-performing DNN architecture was a combination of several convolution blocks followed by fully connected layers. The DNN was trained and tested with extensive datasets of expected radio signals from neutrino interactions with energies between $10^{17}$ eV and $10^{19}$ eV at the South Pole. Throughout this work we considered all triggered events: i.e., we did not apply any quality cuts with respect to trigger level.

We studied three datasets that differed by how the Askaryan radiation was calculated and by the type of neutrino interaction. To be comparable with previous work, one dataset was restricted to only hadronic showers and used a simple frequency domain parameterization of the Askaryan emission. The second dataset improved on that by using a detailed time-domain calculation of the Askaryan signal. The third dataset considered the challenging event class of electron neutrino charged current $\nu_e$-CC interactions where the shower development is impacted by the LPM effect leading to the interference of multiple particle cascades. Previous studies often ignored this case or found that the reconstruction performance deteriorated significantly. Our work solved this problem.

The resolution of the neutrino direction (the angular difference between the true and reconstructed direction) shows a narrow peak at $\mathcal{O}(1')$ with extended tails. The point spread function can be described well with a Moffat/King function that can be thought of as a Gaussian with extended tails. We find that the narrow part of the distribution is described with a sigma of 0.6° which increases to 0.8° for the more challenging $\nu_e$-CC dataset. Due to the tails of the distribution, the 68% quantiles are at 4° and 5° respectively. This finding highlights the advantage of a DNN for event reconstruction which is able to learn the complex correlations of $\nu_e$-CC interactions between neutrino direction and observed radio flashes. The performance is only slightly worse than for the simpler non-$\nu_e$-CC interactions which is a significant improvement over previous reconstruction methods. We find the angular resolution improves with both neutrino energy and signal-to-noise ratio. The energy dependence is partly due to an underrepresentation of low-energy events in the training dataset.

We also find promising results for the energy reconstruction. The DNN achieved an average resolution of a factor of 2 in shower energy for all triggered events: i.e., without applying any quality cuts with respect to trigger level. The reconstruction uncertainty matches the science requirements of UHE neutrino detectors and is below the intrinsic uncertainty from inelasticity fluctuations. The reconstruction also works surprisingly well for the more complicated $\nu_e$-CC interactions. We find a bias toward the low-energy end, where the DNN overpredicts the shower energy. This is again partly due to the underrepresentation of these events in the training dataset which we plan to address in future work. We compared the energy reconstruction to simplified datasets where the information is reduced to the signal amplitude and its position to mimic a simple traditional analysis and find a much better performance of our model even when the information is extracted from noiseless waveforms. This shows that the DNN exploits the additional information that is encoded in the
waveforms and highlights the usability of a deep learning based reconstruction to determine the neutrino properties from radio detector data.

Although these are promising results, this work is just the beginning of exploiting deep learning for event reconstruction of radio detector data. Additional network architectures can be exploited to potentially improve the reconstruction performance such as recurrent neural networks for encoding the time domain waveforms. Furthermore, we plan to extend the DNN with a prediction of the event-by-event uncertainty which is needed for many physics analyses. The simplest way to achieve that would be to assume a Gaussian distribution of the uncertainty and to let the network predict the corresponding sigma parameter. While Gaussian errors are an acceptable approximation of the energy uncertainty, the expected angular uncertainties are not Gaussian but banana-shaped regions in the sky \[15\]. In future work, we will use Normalizing Flows to directly predict the event-by-event probability distribution instead of a single direction plus uncertainty \[40\].

Another critical aspect is the accuracy of the training data. With the development of NuRadioMC \[23\] and its ongoing improvements (see e.g., \[41, 42, 43, 44\]), a lot of work has already been done, but more work is needed to ensure the dataset’s validity: i.e., that it correctly resembles real data. Our study of the impact of different Askaryan emission models already explored a part of potential systematic uncertainties of the MC dataset, but more work is needed to systematically study uncertainties of the simulation code as well as to test robustness against experimental uncertainties.

The work presented here also opens up new ways to optimize future detectors. Currently, optimization with respect to the reconstruction performance is often not possible as it takes too much time to adapt traditional reconstruction techniques to changing detector layouts. With deep learning based reconstruction, different detector designs can be quickly evaluated in terms of reconstruction performance as long as enough training data can be provided. Additional efforts to exchange the time-consuming MC simulation with fast surrogate models – e.g., by using generative adversarial networks in combination with the reconstruction methods developed here – would allow the construction of a fully-differentiable pipeline that enables an end-to-end optimization of the detector design: e.g., the position and orientations of the antennas of a shallow radio detector station \[45\].

6. Acknowledgments

We thank all developers of the NuRadioMC code for enabling the creation of the training datasets. We thank Jakob Beise and Beth Riley for the feedback on the manuscript. The computations and data handling were enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC) at UPPMAX partially funded by the Swedish Research Council through grant agreement no. 2018-05973. The work of SM and PB in part supported by NSF grant NRT 1633631 to PB.

References

[1] M. Ackermann et al., Astrophysics Uniquely Enabled by Observations of High-Energy Cosmic Neutrinos, Bull. Am. Astron. Soc. 51, 185 (2019) , \[1903.04334\].
[2] M. Ackermann, S. K. Agarwalla, J. Alvarez-Muniz, C. A. Arguelles, M. Bustamante, B. A. Clark et al., High-Energy and Ultra-High-Energy Neutrinos, Snowmass 2021 white paper (2022) , \[2203.08096\].
[3] M. Ackermann, M. Ahlers, L. Anchordoqui, M. Bustamante, A. Connolly, C. Deaconu et al., Fundamental physics with high-energy cosmic neutrinos, Bull. Am. Astron. Soc. 51 (2019) 215, \[1903.04333\].
[4] V. B. Valera, M. Bustamante and C. Glaser, The ultra-high-energy neutrino-nucleon cross section: measurement forecasts for an era of cosmic EeV-neutrino discovery, 2204.04237.
[5] ARA collaboration, P. Allison et al., Constraints on the diffuse flux of ultra-high energy neutrinos from four years of askaryan radio array data in two stations, Phys. Rev. D 102 (2020) 043021, \[1912.00987v2\].
[6] ARIANNA collaboration, A. Anker et al., Targeting ultra-high energy neutrinos with the ARIANNA experiment, Advances in Space Research 64 (2019) 2595 – 2609, \[1903.01609\].
S. Barwick et al., Constraints on Ultrahigh-Energy Cosmic-Ray Sources from a Search for Neutrons above 10 PeV with IceCube, Physical Review Letters 117 (2016) 241101.

ARIANNA collaboration, A. Anker et al., A search for cosmogenic neutrinos with the ARIANNA test bed using 4.5 years of data, JCAP 03 (2020) 053, [2009.00840].

M. G. Aartsen et al., IceCube-Gen2: the window to the extreme Universe, Journal of Physics G: Nuclear and Particle Physics 48 (2021) 060501.

S. Hallmann for the IceCube-Gen2 collaboration, Sensitivity studies for the IceCube-Gen2 radio array, PoS(ICRC2021)11183.

RNO-G collaboration, J. A. Aguilar et al., Design and sensitivity of the Radio Neutrino Observatory in Greenland (RNO-G), JINST 16 (2021) P03025, [2010.12279].

P. Allison et al., Performance of two Askaryan Radio Array stations and first results in the search for ultrahigh energy neutrinos, Physical Review D 93 (2016) 082003, [1507.08991].

S. Barwick et al., A first search for cosmogenic neutrinos with the ARIANNA Hexagonal Radio Array, Astroparticle Physics 70 (2015) 12–26.

ARIANNA collaboration, A. Anker et al., Neutrino vertex reconstruction with in-ice radio detectors using surface reflections and implications for the neutrino energy resolution, JCAP 11 (2019) 030, [1909.02677].

C. Glaser for the ARIANNA collaboration, Neutrino direction and energy resolution of Askaryan detectors, PoS(ICRC2019)899 (2019), [1911.02093].

G. Gaswint, University of California, Irvine. PhD thesis, University of California, Irvine, 2021.

ARIANNA collaboration, A. Anker et al., Probing the angular and polarization reconstruction of the ARIANNA detector at the South Pole, JINST 15 (2020) P09030, [2006.03027].

ARIANNA collaboration, S. Barwick et al., Capabilities of ARIANNA: Neutrino Pointing Resolution and Implications for Future Ultra-High Energy Neutrino Astronomy, PoS(ICRC2021)1151.

ARIANNA collaboration, A. Anker, P. Baldi, S. W. Barwick, J. Beise, D. Z. Besson, S. Bouma et al., Measuring the polarization reconstruction resolution of the arianna neutrino detector with cosmic rays, JCAP 04 (2022) 022, [2112.01501].

RNO-G collaboration, J. A. Aguilar et al., Reconstructing the neutrino energy for in-ice radio detectors, Eur. Phys. J. C 82 (2022) 147, [2107.02604].

C. Glaser et al., NuRadioReco: A reconstruction framework for radio neutrino detectors, European Physical Journal C 79 (2019) 461, [1903.07023].

I. Plaisier for the RNO-G collaboration, Direction Reconstruction for the Radio Neutrino Observatory Greenland (RNO-G), PoS ICRC2021 1026.

C. Glaser et al., NuRadioMC: Simulating the radio emission of neutrinos from interaction to detector, European Physical Journal C 80 (2020) 77, [1906.01670].

P. Baldi, Deep Learning in Science. Cambridge University Press, Cambridge, UK, 2021.

M. Erdmann, J. Glombitza, G. Kasieczka and U. Klemradt, Deep Learning for Physics Research, World Scientific, 2022, 10.1142/12294.

C. Glaser, S. McAleer, P. Baldi and S. Barwick, Deep learning reconstruction of the neutrino energy with a shallow Askaryan detector, PoS(ICRC2021)1051.

S. Stjärnholm, O. Ericsson and C. Glaser, Neutrino direction and flavor reconstruction from radio detector data using deep convolutional neural networks, PoS(ICRC2021)1055.

ARIANNA collaboration, A. Anker et al., White Paper: ARIANNA-200 high energy neutrino telescope, 2004.09841.

C. Glaser and S. W. Barwick, An improved trigger for Askaryan radio detectors, Journal of Instrumentation 16 (2021) T01001, [2011.12997].