Deep-learning-based reconstruction of the neutrino direction and energy for in-ice radio detectors

C. Glaser [a], S. McAleer [b], S. Stjärnholm [a], P. Baldi [b], S.W. Barwick [c]

[a] *Uppsala University Department of Physics and Astronomy, Uppsala SE-75237, Sweden*
[b] *Department of Information and Computer Science, University of California, Irvine, CA 92697, USA*
[c] *Department of Physics and Astronomy, University of California, Irvine, CA 92697, USA*

ABSTRACT

Ultra-high-energy (UHE) neutrinos ($> 10^{16}$ 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) developed and tested on simulated data. The DNN determines the energy with a standard deviation of a factor of two around the true energy ($\sigma \approx 0.3$ in $\log_{10}(E)$), which meets the science requirements of UHE neutrino detectors. For the first time, we are able to predict the neutrino direction precisely for all event topologies including the complicated electron neutrino charged-current (ν_e-CC) interactions. The obtained angular resolution shows a narrow peak at $\mathcal{O}(1°)$ with extended tails that push the 68% quantile for non-ν_e-CC (resp. ν_e-CC interactions) to 4°(5°). 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 [3–5].

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 [6–12]: 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 low expected 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 [7,10]. With the Radio Neutrino Observatory in Greenland (RNO-G) a much larger detector (35 stations) is being constructed at the

moment [13] and an order-of-magnitude more sensitive radio detector is foreseen for IceCube-Gen2 [11,12].

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 a *deep* design (as explored by ARA [14]) antennas are placed into narrow boreholes down to a depth of up to 200 m. This increases the sensitivity to neutrinos per detector station but also increases the costs per station and limits the choice of available antennas due to the narrow borehole. The second design is a *shallow* detector station (as explored by ARIANNA [15]) 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 [12].

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 [10]. 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 [16]. The exact depth is a compromise between the fraction of events that show a *Direct and Reflected* (D'n'R) signature which decreases with depth (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 [16]. 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 $\mathcal{O}(100\,\text{m})$ of the ice sheet [17]. 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 [18–20] as well as by measuring the radio emissions of cosmic rays [21]. In addition, a novel method for determining the vertex distance was developed and tested in an in situ measurement on the Ross Ice Shelf [16]. 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 [22,23].

To determine the neutrino direction at low signal-to-noise ratios, the *forward folding* method was developed [24]. Here an analytic model of the radio signal is fitted directly to the antennas' observed voltage measurements. The method has been applied successfully to shallow and deep detector components but requires a reconstruction of the neutrino interaction vertex as input [18,20,25]. In a Monte Carlo (MC) study using NuRadioMC [26], which includes a full trigger simulation and modeling of the signal chain and mixes the expected signals with noise, the angular resolution was determined to be approx. 3° for all triggered events of a shallow detector station [18,20]. 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 reduced the fraction of usable events [25]. 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 (v_e-CC) interactions, the generated electron generates an electromagnetic shower next to the hadronic shower. The electromagnetic shower is subject to the LPM effect so that for neutrino energies of approx. 1×10^{18} eV and beyond, several spatially displaced subshowers are generated (see for instance the examples in [6,26]). At lower energies, often only the start of the shower is delayed, resulting in a displacement with respect to the hadronic shower. Furthermore, the electromagnetic shower(s) can interfere with the hadronic shower. This can lead to more complex radio signal shapes but, more importantly, the correlation between for instance the pulse width and the viewing angle changes, which further complicates the reconstruction. For v_e-CC interactions, no results on direction reconstruction have been reported so far because of the difficulty of modeling

the LPM effect in traditional reconstruction methods [18,25]. Because v_e-CC interactions dominate the number of observable neutrinos,[1] this paper addresses this significant caveat of current methods. We note that the signatures of v_e-CC interactions allow to distinguish them from non-nu_e-CC interactions. Preliminary results are promising [27] and will be studied thoroughly in a forthcoming publication. In this work, we assume that the event type (v_e-CC or not) is known prior to the direction and energy reconstructions, but we also estimate the additional uncertainty resulting from a wrong guess of the event type.

For a deep detector station, an energy reconstruction was developed recently [22] with a shower energy resolution of 30% for hadronic showers after moderate quality cuts. However, due to the differences in station design between a deep and a shallow detector station, these findings are not directly comparable to our work. Here we focus on a shallow detector station, for which no end-to-end reconstruction of the neutrino energy has been presented so far.

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 in a supervised manner [28,29] to extract the neutrino properties of interest directly from the simulated raw data. Deep-learning-based reconstructions have already produced promising results in closely related fields and have often outperformed existing methods, see e.g. [30,31]. We use the NuRadioMC code which simulates the expected signals from neutrinos [26] as well as the corresponding detector response [24] which allowed us to generate the needed training datasets. A preliminary version of the analysis was presented at the last International Cosmic Ray Conference (ICRC2021) [27,32].

The importance of the angular and energy resolution on the science output of in-ice radio detectors will depend strongly on the science case and also on the number of observed neutrinos which, in turn, depends on the largely unknown neutrino flux level. Different science cases and scenarios are currently under investigation. Recently, the potential to discover point sources [33] and to measure the neutrino–nucleon cross-section [4] was presented for the future IceCube-Gen2 radio detector. Both analyses find a strong dependence on the angular resolution but only a weak dependence on the energy resolution. For an angular resolution of $\sigma = 3°$[2] promising results were obtained. Increasing the uncertainty to $\sigma = 7°$ (14°) would increase the cross-section uncertainty by 50% (178%). For the point source study, increasing the angular resolution to $\sigma = 7°$ roughly doubles the size of the multiplets needed to claim discovery. The impact of energy resolution to distinguish different flux scenarios is currently under investigation.

2. Dataset generation

We use the NuRadioMC code [26] 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 [12]: The station comprises two parallel pairs of downward-facing LPDA antennas 2 m below the snow surface

---

[1] Assuming a 1:1:1 flavor ration, at 1×10^{17} eV neutrino energy roughly 60% of all triggered events stem from v_e-CC interactions. At neutrino energies of 1×10^{18} eV the fraction reduces to 35%.

[2] The authors quantify the angular resolution by quoting the uncertainty of the zenith angle only which is roughly $\sqrt{2}$ smaller than the space angle uncertainty quoted in this work if the zenith and azimuth uncertainties contribute equally to the total uncertainty.
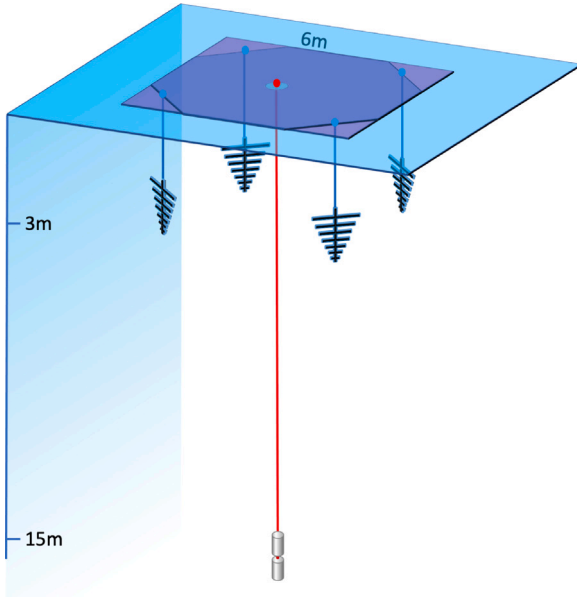
Fig. 1. Schematic illustration of a shallow radio detector station. *Source:* Fig. adapted from [34].
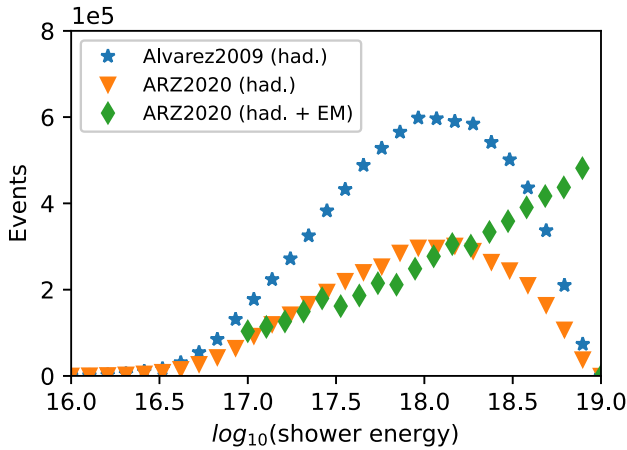

Fig. 2. Number of events in the 3 datasets as a function of the shower energy.

and an additional dipole antenna at a depth of 15 m below the snow surface (see. Fig. 1). 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 to increase the sensitivity to neutrinos [35] 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 [36].

We generate neutrino interactions with neutrino energies (E) from 1×10^{17} eV to 1×10^{19} eV. The simulations are run separately for small energy bins (0.1 in $\log_{10}(E)$) in which the neutrino energies are distributed uniformly on a logarithmic scale. More simulations at low energies are executed to compensate for the lower trigger probability (see discussion below). The neutrino interactions are distributed uniformly in the ice around the detector with random incoming directions (uniformly distributed in the azimuth angle and the cosine

of the zenith angle to assure a uniform sky coverage). 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 *Alvarez2009* [37,38] as implemented in NuRadioMC [26]. This setup was used in previous analyses that used the forward folding technique [18,25]. 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 [26], following the approach in [39,40], 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 (ν_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-ν_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 standard uncertainty of 0.3 in $\log_{10}(E)$, i.e., a factor of two on a linear scale [16]. This provides a target resolution for the shower energy. For ν_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. Hence, the shower energy equals the neutrino energy. We note that due to the LPM effect, the showers do not develop in phase (see discussion of the LPM effect above) 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 1×10^{17} eV to 1×10^{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 reconstruction at low energies due to an underrepresentation in the training dataset. In total the *Alvarez2009 (had.)* dataset consists of 8.2×10^6 events, the *ARZ2020 (had.)* dataset consists of 4.1×10^6 events, and the *ARZ2020 (had. + EM)* consists of 5×10^6 events.
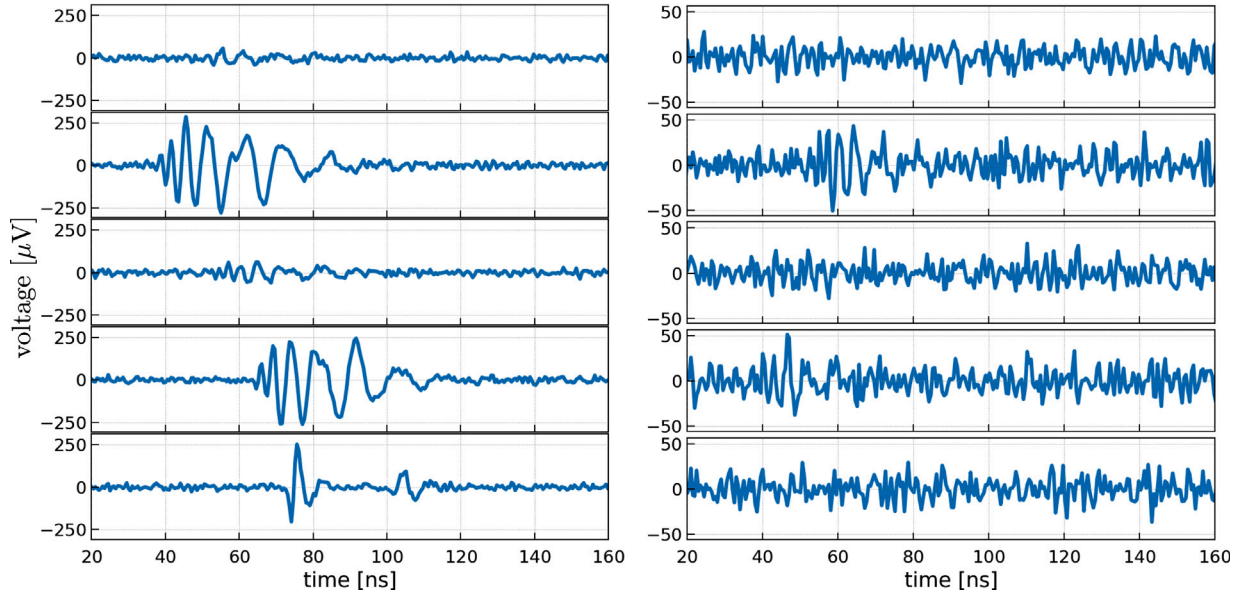
Fig. 3. Two examples of the training dataset. The traces show the recorded voltages in the 5 antennas as a function of time. The upper 4 panels are the signal in the 4 LPDAs, and the bottom panel shows the signal in the dipole antenna. Left: example at a particularly high signal-to-noise ratio, where the direct and reflected pulse is clearly visible in the dipole antenna. Right: a more typical example at a low signal-to-noise ratio.
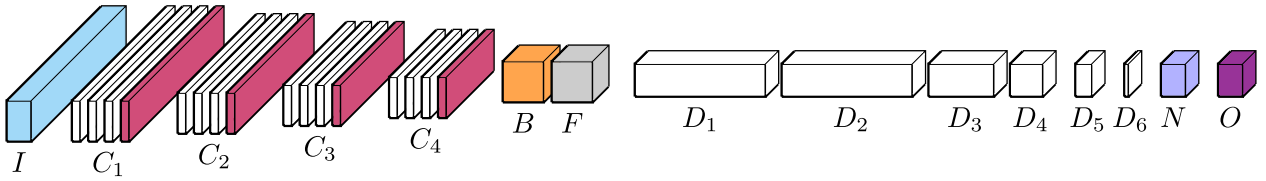

Fig. 4. Sketch of the structure of the developed neural network. The labels are specified in Table 1. The relative sizes of the convolutional layers C_i hint at the layers' output dimensions, while the relative sizes of the dense layers D_i hint at the neuron amount in each layer.

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 resulted in 512 samples per antenna. The shape of one event is $(5, 512)$. The data is normalized by using μV as the base unit which brings the RMS of the data to $\mathcal{O}(10)$. The mean of the dataset is already zero, as the voltages fluctuate around the zero baseline. Two examples are shown in Fig. 3.

3. Direction reconstruction

In this work, we use a neural network architecture that follows the VGG model [41]. The general idea of this architecture is the repeated application of convolutional layers followed by a single pooling, respectively. 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. For a detailed description of the different layers, the reader is referred to textbooks, e.g., [28,29]. The VGG model was developed for image recognition. Similarly, 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 [10]. Convolutional layers can be thought of 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 carry information, we apply the convolution filters independently to each

antenna: i.e., we use an input shape of (5, 512, 1) corresponding to 5 antennas and 512 samples. The last component is 1 as the only available information from each antenna is the voltage amplitude (unlike three color channels in the case of images). We share the filter weights over the five antennas because the signal shapes are expected to be similar — at least in the four LPDA antennas. A future improvement could be to have an individual set of filters for the dipole antenna, however, we expect the low-level features (e.g. sharp rising and falling amplitudes) to still be similar, independent of the antenna type.

An overview of the neural network architecture is presented in Fig. 4, with a more detailed description presented in Table 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 the same filter. The average pooling layer has a pool size 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 C_1, 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

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, kernel size (1,5))	$(5, 512, 32)$	192
	Conv2D (32 filters, kernel size (1,5))	$(5, 512, 32)$	5152
	Conv2D (32 filters, kernel size (1,5))	$(5, 512, 32)$	5152
	AveragePooling2D	$(5, 128, 32)$	0
C_2	Conv2D (64 filters, kernel size (1,5))	$(5, 128, 64)$	10 304
	Conv2D (64 filters, kernel size (1,5))	$(5, 128, 64)$	20 544
	Conv2D (64 filters, kernel size (1,5))	$(5, 128, 64)$	20 544
	AveragePooling2D	$(5, 32, 64)$	0
C_3	Conv2D (128 filters, kernel size (1,5))	$(5, 32, 128)$	41 088
	Conv2D (128 filters, kernel size (1,5))	$(5, 32, 128)$	82 048
	Conv2D (128 filters, kernel size (1,5))	$(5, 32, 128)$	82 048
	AveragePooling2D	$(5, 8, 128)$	0
C_4	Conv2D (256 filters, kernel size (1,5))	$(5, 8, 256)$	164 096
	Conv2D (256 filters, kernel size (1,5))	$(5, 8, 256)$	327 936
	Conv2D (256 filters, kernel size (1,5))	$(5, 8, 256)$	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

fully connected layers. The *dense* block consists of 6 fully connected layers. The size of the fully connected layers decreases for layers further along the network and reduces the network to the 3 output neurons. The fully connected 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 use tensorflow/Keras to train the neural network [42,43]. 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 8.2×10^6 events, 4.1×10^6 events, and 5×10^6 events respectively. Each event has a size of $20\,\text{kB}$ resulting in data volumes between $82\,\text{GB}$ and $160\,\text{GB}$ per dataset. The neural network is trained independently for each data set. 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 a batch size of 64 and ran the training on an NVIDIA Quadro RTX 6000 GPU resulting in a training time of roughly $10\,\text{min}$ per epoch. We always made sure that the training converged and that no overfitting took place which took a few hours of training time.

To train the model, we use the optimizer *Adam* [44] with a learning rate of $5 \cdot 10^{-5}$ 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 low signal quality to enable a good reconstruction. The mean-squared error (MSE) loss will put more weight on events that might intrinsically be harder to reconstruct. We do not want these events to dominate the training. In a future analysis, quality cuts will be developed to remove those events from the dataset. We found that for this particular task of direction reconstruction, the MAE loss performed better than the MSE loss.

Through many experiments, we tuned 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 fully connected 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, 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)*. All results are evaluated for the independent test data set of 300,000 events (cf. Section 3.1). Fig. 5 shows histograms
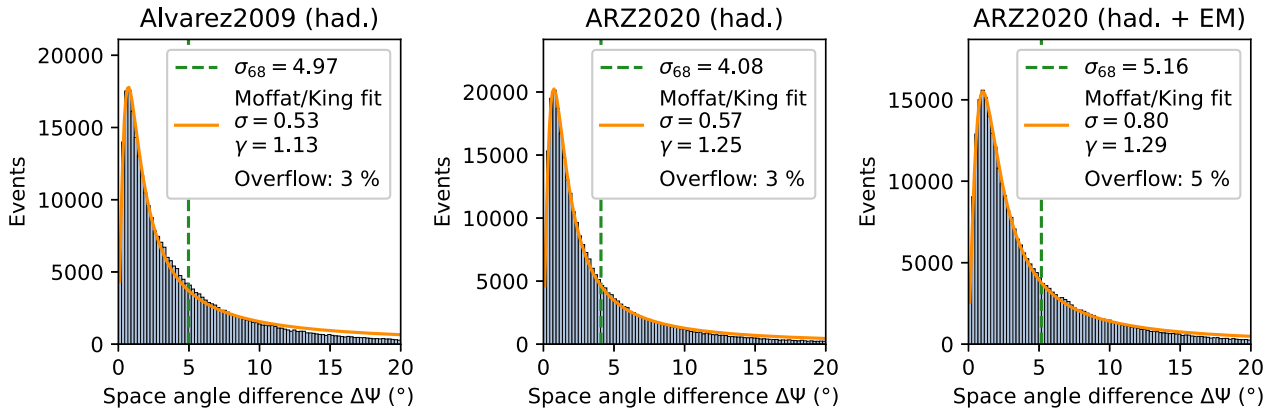
Fig. 5. Space angle difference for the three datasets. A Moffat/King function gives a good description of the point spread function. The vertical dashed line shows the position of the 68% quantile. The overflow in the legend specifies the fraction of events that have a space angle difference larger than 20°.

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 σ_{68} which is also specified in the graphs. The point spread function can be described with a Moffat/King function that can be thought of as a Gaussian with extended tails [45–47] and is given by:

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

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 ν_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 ν_e-CC interactions are learned well and are only slightly worse than for the simpler non-ν_e-CC interactions. This is a significant improvement over previous reconstruction methods that fail to reconstruct ν_e-CC interactions [18, 25].

The average resolution strongly depends on the event distribution, in particular the number of events as a function of energy (cf. Fig. 2). Therefore, the average resolution obtained on the test data set needs to be interpreted with caution. 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 µV. The maximum signal amplitude is determined from noisy traces which results in typical SNR values of 3.5 or higher. Smaller values are rare. We observe an increase in performance with increasing energy and SNR as expected. The SNR and energy is weakly correlated with higher-energy events having more often high SNRs than low-energy events. We show the correlation in Fig. 8. However, in each energy bin, the SNR distribution still peaks at low SNRs, i.e., most events are still measured with signals close to the trigger threshold independent of the neutrino energy. Therefore, the improvement with energy cannot be solely explained by an increasing SNR.

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 [18,25]. We speculate that the reason for the better performance of the *ARZ2020* emission model is that it generates more information than the neural network is able to learn. The *Alvarez2009* model corresponds to a frequency domain parameterization where the phase was assumed to be a constant 90° at all
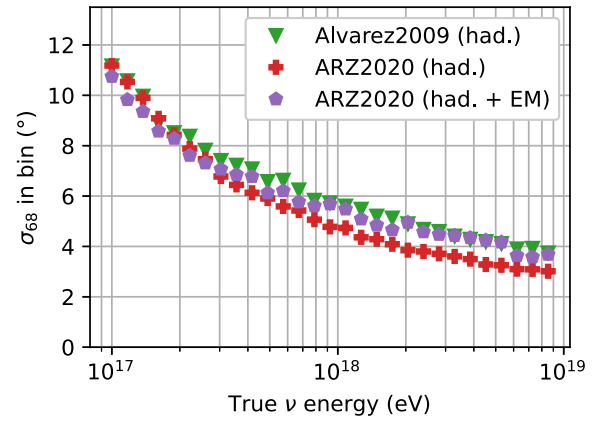

Fig. 6. Angular resolution σ_{68} as a function of neutrino energy for the three datasets.

frequencies leading to perfectly symmetric pulse forms. However, a more precise calculation of the radio emission in the time domain (as done in *ARZ2020*) shows additional structures in the pulse shape that depend on the even geometry (see. e.g. [26,39,40]). 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 [18,25]. 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.

So far we assumed that we know the event type, i.e., we reconstructed ν_e-CC events with the neural network trained on ν_e-CC events, and the same for non ν_e-CC events. The signature of ν_e-CC interactions allows to distinguish them from non-ν_e-CC interactions, [27], however, with uncertainties. Therefore, as we do not always know the event type,
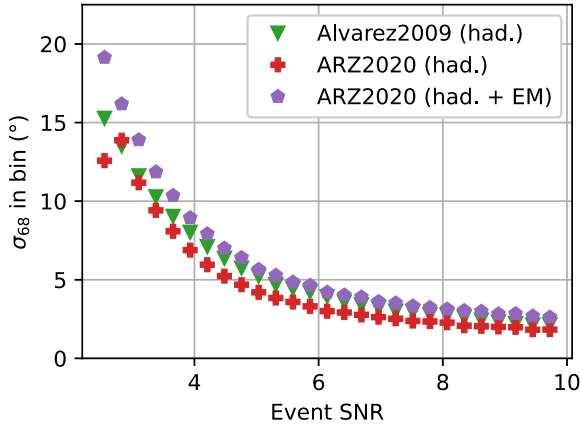
Fig. 7. Angular resolution σ_{68} as a function of signal-to-noise ratio for the three datasets.
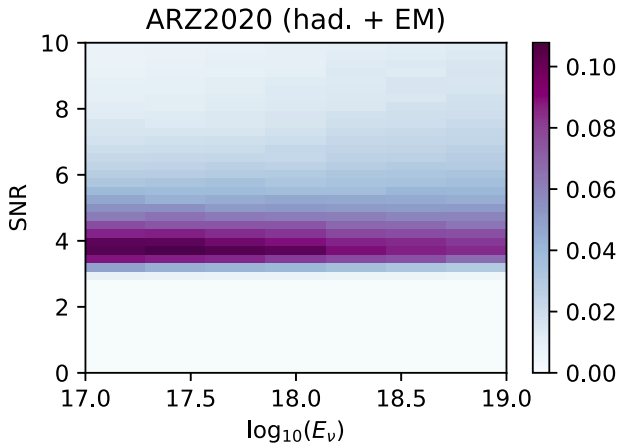

Fig. 8. Correlation between neutrino energy and signal-to-noise ratio. Each column is normalized to 1. Here, the test data set of the ARZ2020(had. + EM) model is shown that is also representative for the other datasets.

Fig. 9. Angular resolution σ_{68} as a function of neutrino energy for an incorrect guess of the event type, i.e., the neural network trained on ν_e-CC events is evaluated on the non-ν_e-CC dataset and vice versa. Filled markers show the correct combinations, empty markers the incorrect combinations.

Fig. 10. Angular resolution σ_{68} as a function of neutrino energy for an incorrect combination of the Askaryan emission model, i.e., the neural network trained on the *Alvarez2009 (had.)* dataset is evaluated with the *ARZ2020 (had.)* dataset and vice versa. Filled markers show the correct combinations, empty markers the incorrect combinations.

we also studied the deterioration in reconstruction performance when the *wrong* network is used for reconstruction. The results are presented in Fig. 9. We observe that the model trained on ν_e-CC events performs better on non-ν_e-CC events than the other way around. This is expected because the ν_e-CC dataset also contains events that are dominated by the hadronic shower and, thus, are similar to the non-ν_e-CC dataset. The *ARZ2020 (had. + EM)* network reconstructs the *ARZ2020 (had.)* dataset almost as good as the *ARZ2020 (had. + EM)* dataset it was trained on, with a similar performance above $E_\nu > 1 \times 10^{18}$ eV and a deterioration of approx. 1° for smaller energies. However, the *ARZ2020 (had.)* network still performs approx. 1° better on the *ARZ2020 (had.)* dataset over the full energy range. Thus, if the event type is now known, it is best to use the *ARZ2020 (had. + EM)* network but it will lead to a deterioration in angular resolution of approx. 1° for non-ν_e interactions.

We can also obtain an estimate of the systematic uncertainty of the Askaryan emission model by analyzing the *Alvarez2009 (had.)* dataset with the network trained on the *ARZ2020 (had.)* dataset and vice versa. Because we know that the *ARZ2020* model is more accurate, the deterioration of the angular resolution does not represent the systematic uncertainty of the emission model but rather sets an upper bound. The result is presented in Fig. 10. We find that the *Alvarez2009* network works better on the *ARZ2020* dataset than the other way around. A possible explanation for this behavior is that the *Alvarez2009* network only learned simple features that are also present in the *ARZ2020* dataset. In contrast, the *ARZ2020* network learned to use second order
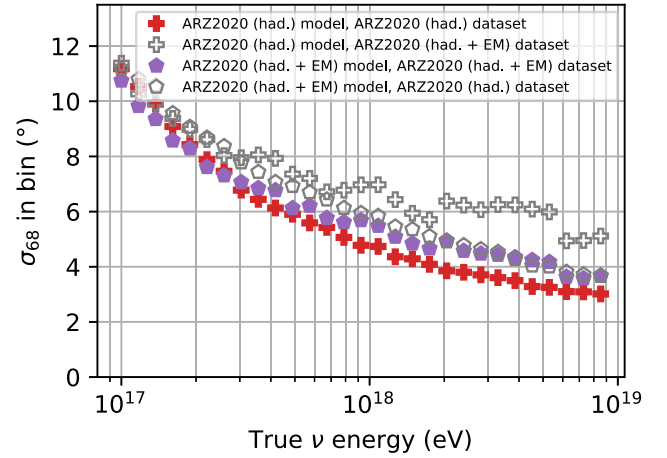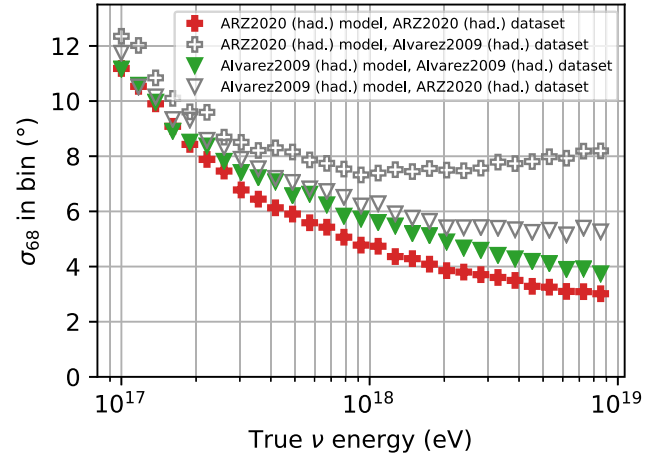
features that give it a better performance than the *Alvarez2009* network (see discussion above and Figs. 6 and 7) but as these features are not present in the *Alvarez2009* dataset it performs worse when applied to this dataset.

3.3. Comparison to previous results

A comparison with the work of [18,20] indicates that our DNN reconstruction still has room for improvement. The work of [18,20] used the forward folding technique [24] 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. For ν_e-CC interactions, no results on direction reconstruction have been reported so far because of the difficulty of modeling the LPM effect in traditional reconstruction methods [18,25].
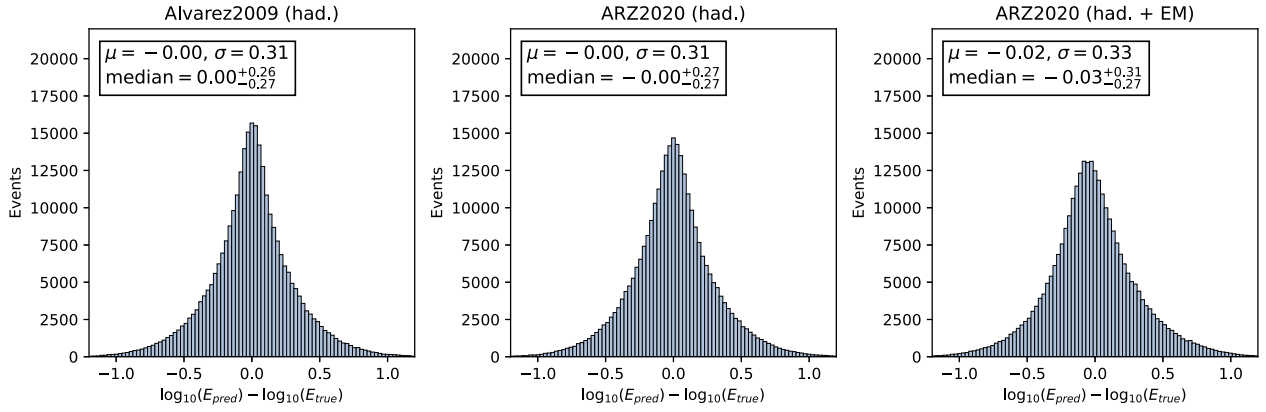
Fig. 11. Energy difference histograms for the 3 datasets.
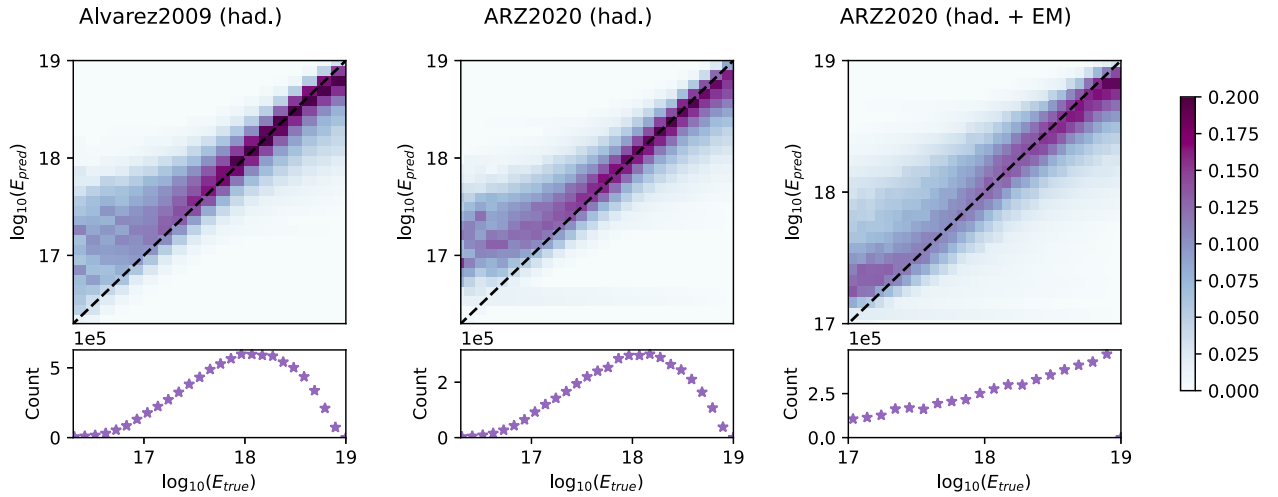

Fig. 12. Predicted energy versus true energy for the 3 datasets. Each column is normalized such that the values sum to unity. The lower panels show the number of events in each energy bin.

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 [48] 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. 11. 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. 11 also list the 68% quantiles around the median which are smaller than the standard deviation, indicating the presence of non-Gaussian tails. These events could potentially be identified and removed through quality cuts in future work. 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. 12 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. 13. 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. 12. 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
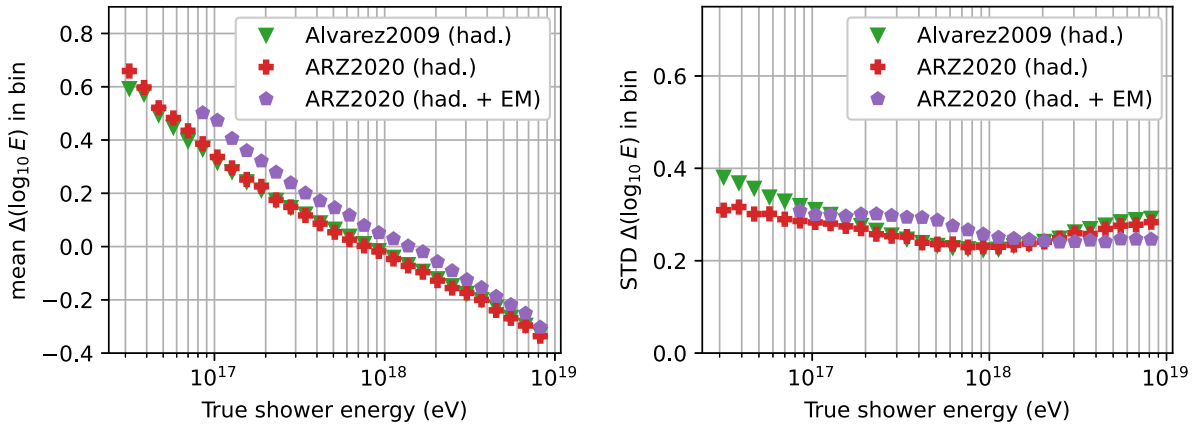
Fig. 13. Mean and STD of the predicted shower energy in bins of true shower energy.
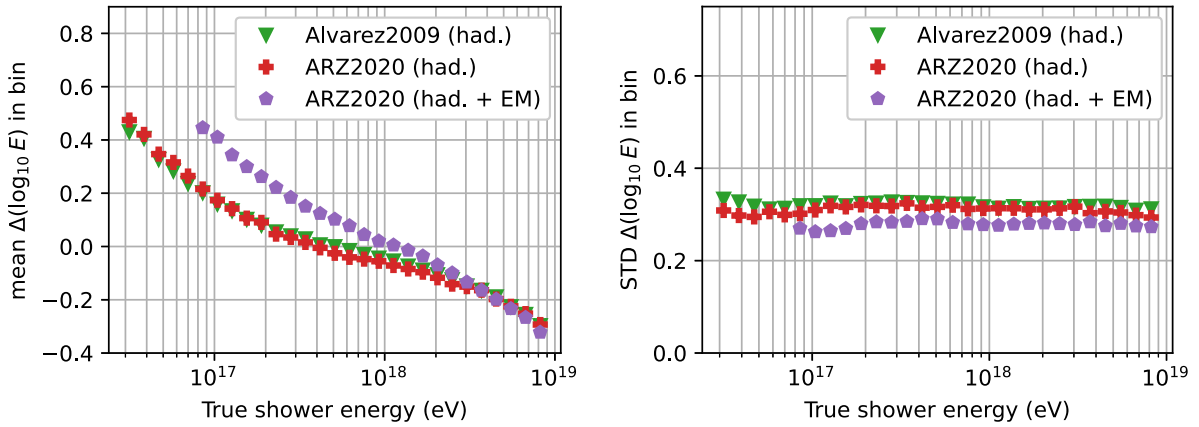

Fig. 14. Mean and STD of the predicted shower energy in bins of true shower energy after re-weighting of the loss function during training with a maximum weight of 10.

to 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. 14 shows equivalent plots to Fig. 13 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. 13 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. 15. 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.

As in the previous section, we also study how much the reconstruction performance deteriorates if we do not know the event type. The results are presented in Fig. 16. We find that the *ARZ2020 (had. + EM)* network performs equally well on the *ARZ2020 (had.)* dataset than on the *ARZ2020 (had. + EM)* dataset it was trained on, whereas the *ARZ2020 (had.)* network performs worse on the *ARZ2020 (had. + EM)* dataset. Above 3×10^{18} eV, the *ARZ2020 (had. + EM)* network performs even better on the *ARZ2020 (had.)* dataset than the *ARZ2020 (had.)* network itself which we attribute to the larger amount of training data at high energies in the *ARZ2020 (had. + EM)* dataset. At energies below 1×10^{18} eV, the *ARZ2020 (had.)* network still performs better (cf. Fig. 13). Hence, there is still a benefit in knowing the event type.

We also repeat our estimation for an upper bound of the systematic uncertainty of the Askaryan emission model by exchanging the *ARZ2020 (had.)* and *Alvarez2009 (had.)* networks/datasets. The results are presented in Fig. 17. As for the direction reconstruction, we find that the *Alvarez2009 (had.)* network works better on the *ARZ2020 (had.)* dataset than the other way around and attribute it to the same reason.

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
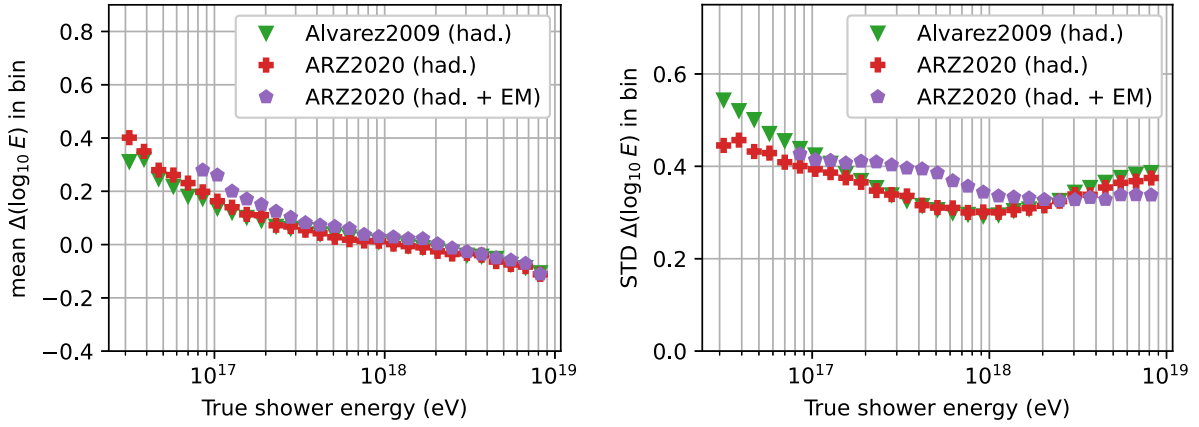
Fig. 15. Mean and STD of the predicted shower energy in bins of true shower energy after bias correction (see text for details).
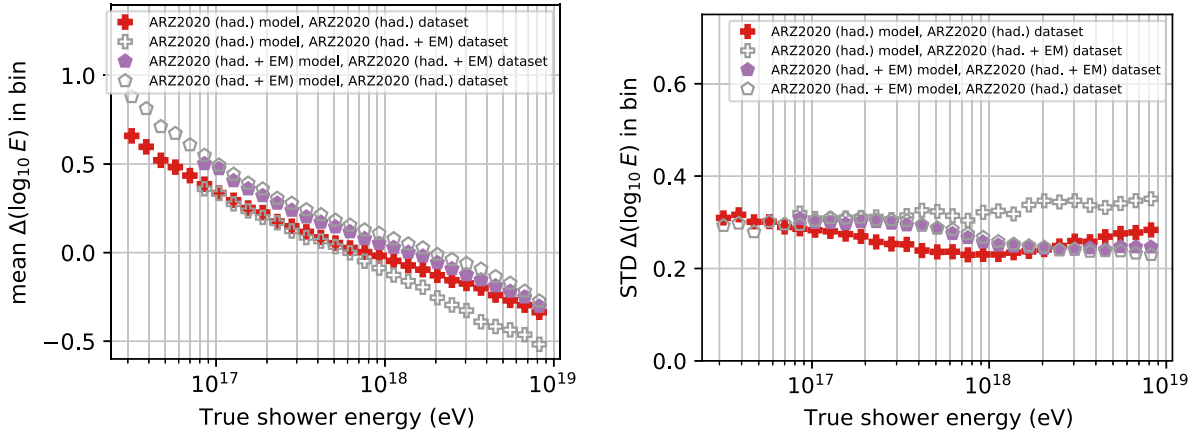

Fig. 16. Energy resolution (bias left, standard deviation right) as a function of neutrino energy for an incorrect guess of the event type, i.e., the neural network trained on ν_e-CC events is evaluated on the non-ν_e-CC dataset and vice versa. Filled markers show the correct combinations, empty markers the incorrect combinations.

DNN architecture was a combination of several convolution blocks followed by fully connected layers. The DNN was trained and tested with extensive datasets of several million events of expected radio signals from neutrino interactions with energies between 1×10^{17} eV and 1×10^{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 ν_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 ν_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 ν_e-CC interactions between neutrino direction and observed radio flashes. The performance is only slightly worse than for the simpler non-ν_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 ($\sigma \approx 0.3$ in $\log_{10}(E)$) 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 ν_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.

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
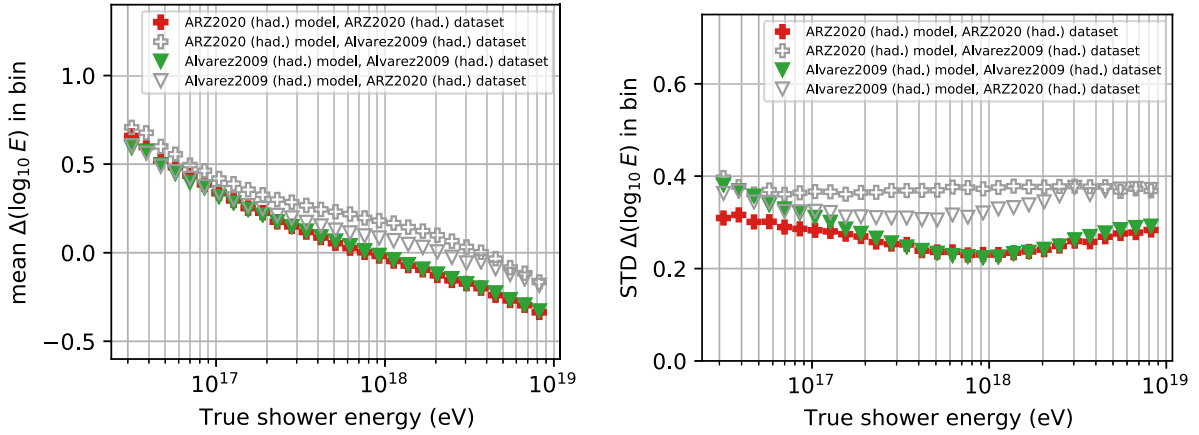
Fig. 17. Energy resolution (bias left, standard deviation right) as a function of neutrino energy for an incorrect combination of the Askaryan emission model, i.e., the neural network trained on the *Alvarez2009 (had.)* dataset is evaluated with the *ARZ2020 (had.)* dataset and vice versa. Filled markers show the correct combinations, empty markers the incorrect combinations.

uncertainty, the expected angular uncertainties are not Gaussian but banana-shaped regions in the sky [17]. In future work, we will use *normalizing flows* [49] to directly predict the event-by-event probability distribution instead of a single direction plus uncertainty [50].

Another critical aspect is the accuracy of the training data. With the development of NuRadioMC [26] and its ongoing improvements (see e.g., [51–54]), 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, the network architecture can be tuned once, and then retrained on different detector designs quickly to evaluate the 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 [55].

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The code to generate the MC data sets is already open source. The datasets itself are too big to be publically shared and are still being used for ongoing analyses.

Acknowledgments

References

[1] M. Ackermann, et al., Astrophysics uniquely enabled by observations of high-energy cosmic neutrinos, Bull. Am. Astron. Soc. 51 (2019) 185, [1903.04334].

[2] M. Ackermann, et al., High-energy and ultra-high-energy neutrinos, snowmass 2021 white paper, J. High Energy Astrophys. 36 (2022) 55–110, http://dx.doi.org/10.1016/j.jheap.2022.08.001, [2203.08096].

[3] M. Ackermann, M. Ahlers, L. Anchordoqui, M. Bustamante, A. Connolly, C. Deaconu, D. Grant, P. Gorham, F. Halzen, A. Karle, K. Kotera, M. Kowalski, M.A. Mostafa, K. Murase, A. Nelles, A. Olinto, A. Romero-Wolf, A. Vieregg, S. Wissel, Fundamental physics with high-energy cosmic neutrinos, Bull. Am. Astron. Soc. 51 (2019) 215, [1903.04333].

[4] V.B. Valera, M. Bustamante, C. Glaser, The ultra-high-energy neutrino-nucleon cross section: measurement forecasts for an era of cosmic EeV-neutrino discovery, J. High Energy Phys. 06 (2022) 105, http://dx.doi.org/10.1007/JHEP06(2022)105, [2204.04237].

[5] I. Esteban, S. Prohira, J.F. Beacom, Detector requirements for model-independent measurements of ultrahigh energy neutrino cross sections, Phys. Rev. D 106 (2022) 023021, http://dx.doi.org/10.1103/PhysRevD.106.023021.

[6] S. Barwick, C. Glaser, Radio detection of high energy neutrinos in ice, in: F.W. Stecker, G.G. Fazio (Eds.), Neutrino Physics and Astrophysics, in: The Encyclopedia of Cosmology II, World Scientific Publishing Company, Singapore, 2022, in press. [2208.04971].

[7] 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, http://dx.doi.org/10.1103/PhysRevD.102.043021, [1912.00987v2].

[8] ARIANNA Collaboration, A. Anker, et al., Targeting ultra-high energy neutrinos with the ARIANNA experiment, Adv. Space Res. 64 (2019) 2595–2609, http://dx.doi.org/10.1016/j.asr.2019.06.016, [1903.01609].

[9] IceCube Collaboration, M.G. Aartsen, et al., Constraints on ultrahigh-energy cosmic-ray sources from a search for neutrinos above 10 PeV with IceCube, Phys. Rev. Lett. 117 (2016) 241101, http://dx.doi.org/10.1103/PhysRevLett.117.241101.

[10] ARIANNA Collaboration, A. Anker, et al., A search for cosmogenic neutrinos with the ARIANNA test bed using 4.5 years of data, J. Cosmol. Astropart. Phys. 03 (2020) 053, http://dx.doi.org/10.1088/1475-7516/2020/03/053, [1909.00840].

[11] M.G. Aartsen, et al., IceCube-Gen2: the window to the extreme Universe, J. Phys. G: Nucl. Part. Phys. 48 (2021) 060501, http://dx.doi.org/10.1088/1361-6471/abbd48.

[12] S. Hallmann for the IceCube-Gen2 collaboration, Sensitivity studies for the IceCube-Gen2 radio array, PoS(ICRC2021)1183 http://dx.doi.org/10.22323/1.395.1183.

[13] RNO-G Collaboration, J.A. Aguilar, et al., Design and sensitivity of the radio neutrino observatory in greenland (RNO-G), JINST 16 (2021) P03025, http://dx.doi.org/10.1088/1748-0221/16/03/P03025, [2010.12279].

[14] P. Allison, et al., Performance of two Askaryan Radio Array stations and first results in the search for ultrahigh energy neutrinos, Phys. Rev. D 93 (2016) 082003, http://dx.doi.org/10.1103/PhysRevD.93.082003, [1507.08991].

[15] S. Barwick, et al., A first search for cosmogenic neutrinos with the ARIANNA Hexagonal Radio Array, Astropart. Phys. 70 (2015) 12–26, http://dx.doi.org/10.1016/j.astropartphys.2015.04.002.

[16] ARIANNA Collaboration, A. Anker, et al., Neutrino vertex reconstruction with in-ice radio detectors using surface reflections and implications for the neutrino energy resolution, J. Cosmol. Astropart. Phys. 11 (2019) 030, http://dx.doi.org/10.1088/1475-7516/2019/11/030, [1909.02677].

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

[18] G. Gaswint, University of California, Irvine (Ph.D. thesis), University of California, Irvine, 2021.

[19] ARIANNA Collaboration, A. Anker, et al., Probing the angular and polarization reconstruction of the ARIANNA detector at the South Pole, JINST 15 (2020) P09039, http://dx.doi.org/10.1088/1748-0221/15/09/P09039, [2006.03027].

[20] ARIANNA Collaboration, S. Barwick, et al., Capabilities of ARIANNA: Neutrino pointing resolution and implications for future ultra-high energy neutrino astronomy, PoS(ICRC2021)1151 http://dx.doi.org/10.22323/1.395.1151.

[21] ARIANNA Collaboration, A. Anker, et al., Measuring the polarization reconstruction resolution of the ARIANNA neutrino detector with cosmic rays, J. Cosmol. Astropart. Phys. 04 (2022) 022, http://dx.doi.org/10.1088/1475-7516/2022/04/022, [2112.01501].

[22] RNO-G Collaboration, J.A. Aguilar, et al., Reconstructing the neutrino energy for in-ice radio detectors, Eur. Phys. J. C 82 (2022) 147, http://dx.doi.org/10.1140/epjc/s10052-022-10034-4, [2107.02604].

[23] Y. Pan for the ARA collaboration, A neural network based UHE neutrino reconstruction method for the Askaryan Radio Array (ARA), PoS ICRC2021, 1183, http://dx.doi.org/10.22323/1.395.1157.

[24] C. Glaser, et al., NuRadioReco: A reconstruction framework for radio neutrino detectors, Eur. Phys. J. C 79 (2019) 464, http://dx.doi.org/10.1140/epjc/s10052-019-6971-5, [1903.07023].

[25] I. Plaisier for the RNO-G collaboration, Direction reconstruction for the radio neutrino observatory greenland (RNO-G), PoS ICRC2021, 1026, http://dx.doi.org/10.22323/1.395.1026.

[26] C. Glaser, et al., NuRadioMC: Simulating the radio emission of neutrinos from interaction to detector, Eur. Phys. J. C 80 (2020) 77, http://dx.doi.org/10.1140/epjc/s10052-020-7612-8, [1906.01670].

[27] S. Stjärnholm, O. Ericsson, C. Glaser, Neutrino direction and flavor reconstruction from radio detector data using deep convolutional neural networks, PoS(ICRC2021)1055 http://dx.doi.org/10.22323/1.395.1055.

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

[29] M. Erdmann, J. Glombitza, G. Kasieczka, U. Klemradt, Deep Learning for Physics Research, World Scientific, 2022, http://dx.doi.org/10.1142/12294.

[30] R. Abbasi, et al., A convolutional neural network based cascade reconstruction for the IceCube neutrino observatory, JINST 16 (2021) P07041, http://dx.doi.org/10.1088/1748-0221/16/07/P07041, [2101.11589].

[31] Pierre Auger Collaboration, A. Aab, et al., Deep-learning based reconstruction of the shower maximum X_{max} using the water-cherenkov detectors of the pierre auger observatory, JINST 16 (2021) P07019, http://dx.doi.org/10.1088/1748-0221/16/07/P07019, [2101.02946].

[32] C. Glaser, S. McAleer, P. Baldi, S. Barwick, Deep learning reconstruction of the neutrino energy with a shallow askaryan detector, PoS(ICRC2021)1051 http://dx.doi.org/10.22323/1.395.1051.

[33] D.F.G. Fiorillo, M. Bustamante, V.B. Valera, Near-future discovery of point sources of ultra-high-energy neutrinos, 2022, [2205.15985].

[34] ARIANNA Collaboration, A. Anker, et al., White paper: ARIANNA-200 high energy neutrino telescope, 2020, 2004.09841.

[35] C. Glaser, S.W. Barwick, An improved trigger for askaryan radio detectors, J. Instrum. 16 (2021) T05001, http://dx.doi.org/10.1088/1748-0221/16/05/t05001, [2011.12997].

[36] ARIANNA Collaboration, A. Anker, et al., Improving sensitivity of the ARIANNA detector by rejecting thermal noise with deep learning, JINST 17 (2022) P03007, http://dx.doi.org/10.1088/1748-0221/17/03/P03007, [2112.01031].

[37] J. Alvarez-Muñiz, C.W. James, R.J. Protheroe, E. Zas, Thinned simulations of extremely energetic showers in dense media for radio applications, Astropart. Phys. 32 (2009) 100–111, http://dx.doi.org/10.1016/j.astropartphys.2009.06.005.

[38] J. Alvarez-Muñiz, W.R. Carvalho, M. Tueros, E. Zas, Coherent Cherenkov radio pulses from hadronic showers up to EeV energies, Astropart. Phys. 35 (2012) 287–299, http://dx.doi.org/10.1016/j.astropartphys.2011.10.002.

[39] J. Alvarez-Muñiz, A. Romero-Wolf, E. Zas, Cerenkov radio pulses from electromagnetic showers in the time domain, Phys. Rev. D 81 (2010) 123009, http://dx.doi.org/10.1103/physrevd.81.123009, [1002.3873].

[40] J. Alvarez-Muñiz, P.M. Hansen, A. Romero-Wolf, E. Zas, Askaryan radiation from neutrino-induced showers in ice, Phys. Rev. D 101 (2020) 083005, http://dx.doi.org/10.1103/physrevd.101.083005, [2003.09705].

[41] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition. 1409.1556.

[42] M. Abadi, et al., TensorFlow: Large-scale machine learning on heterogeneous systems, 2015, 1603.04467.

[43] F. Chollet, et al., Keras, 2015, https://keras.io/.

[44] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2017, 1412.6980.

[45] J. van Santen, B.A. Clark, R. Halliday, S. Hallmann, A. Nelles, Toise: a framework to describe the performance of high-energy neutrino detectors, J. Instrum. 17 (2022) T08009, http://dx.doi.org/10.48550/ARXIV.2202.11120, [2202.11120].

[46] A.F.J. Moffat, A theoretical investigation of focal stellar images in the photographic emulsion and application to photographic photometry, Astron. Astrophys. 3 (1969) 455.

[47] I. King, The structure of star clusters. I. an empirical density law, Astron. J. 67 (1962) 471, http://dx.doi.org/10.1086/108756.

[48] https://keras.io/api/callbacks/reduce_lr_on_plateau/.

[49] G. Papamakarios, E.T. Nalisnick, D.J. Rezende, S. Mohamed, B. Lakshminarayanan, Normalizing flows for probabilistic modeling and inference., J. Mach. Learn. Res. 22 (2021) 1–64.

[50] T. Glüsenkamp, Unifying supervised learning and VAEs – automating statistical inference in high-energy physics, 2020, 2008.05825.

[51] D. García-Fernández, C. Glaser, A. Nelles, The signatures of secondary leptons in radio-neutrino detectors in ice, Phys. Rev. D 102 (2020) 083011, http://dx.doi.org/10.1103/PhysRevD.102.083011, [2003.13442].

[52] C. Glaser, D. García-Fernández, A. Nelles, Prospects for neutrino-flavor physics with in-ice radio detectors, PoS(ICRC2021)1231 http://dx.doi.org/10.22323/1.395.1231.

[53] B. Oeyen, I. Plaisier, C. Glaser, A. Nelles, T. Winchen, Effects of complex ice models on radio neutrino simulations using a RadioPropa ray tracer, PoS(ICRC2021)1027 http://dx.doi.org/10.22323/1.395.1027.

[54] N. Heyer, C. Glaser, First-principle calculation of birefringence effects for in-ice radio detection of neutrinos, 2022, 2205.06169.

[55] T. Dorigo, A. Giammanco, P. Vischia, et al., Toward the end-to-end optimization of particle physics instruments with differentiable programming: a white paper, 2022, 2203.13818.