

Signal classification and event reconstruction for acoustic neutrino detection in sea water with KM3NeT

Dominik Kießling^{1,*} for the ECAP Acoustic Detection Group

¹Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen Centre for Astroparticle Physics, Erwin-Rommel-Str. 1, 91058 Erlangen, Germany

Abstract. The research infrastructure KM3NeT will comprise a multi cubic kilometer neutrino telescope that is currently being constructed in the Mediterranean Sea. Modules with optical and acoustic sensors are used in the detector. While the main purpose of the acoustic sensors is the position calibration of the detection units, they can be used as instruments for studies on acoustic neutrino detection, too. In this article, methods for signal classification and event reconstruction for acoustic neutrino detectors will be presented, which were developed using Monte Carlo simulations. For the signal classification the disk-like emission pattern of the acoustic neutrino signal is used. This approach improves the suppression of transient background by several orders of magnitude. Additionally, an event reconstruction is developed based on the signal classification. An overview of these algorithms will be presented and the efficiency of the classification will be discussed. The quality of the event reconstruction will also be presented.

1 Introduction

The acoustic neutrino detection technique is a promising approach to create a neutrino telescope with several cubic kilometers effective volume in the energy range beyond 10^{18} eV. Measuring the cosmic neutrino flux in this energy range allows testing the flux predicted by the GZK-effect [1–3]. The acoustic signal generated by neutrinos is produced via the thermoacoustic effect, which was predicted by Askaryan [4]. The neutrino interacts with the sea water, generating a hadronic cascade. Superposition of the emitted elementary sound waves from each point of the shower generates a bipolar pulse. The emission strength is strongly peaked in the plane perpendicular to the direction of the incoming neutrino. This characteristic shape of the sound emission is often called "pancake". The properties of the sound emission pattern by neutrino signals can be used to classify the recorded events. As experience with the experimental test set up AMADEUS [5] has shown, using only the bipolar shape of the waveform to distinguish the background and the signal causes too many false positives [6]. The initial rate of events recorded by the AMADEUS project, where a reconstruction of the source position is possible, will be shown in Table 1. Several classifiers remove background events, but after all currently implemented steps are applied, 63 k events/year remain. The effective volume of AMADEUS is very small compared to the size needed for an actual acoustic neutrino detector. The GZK neutrino flux would have to be several orders of magnitude higher than the most optimistic

*e-mail: dominik.kiessling@fau.de

prediction to produce that many events. Therefore it is assumed that all remaining events are caused by background sources. However, further improvement of the background suppression would require a cut on the "pancake" emission pattern. The size of the AMADEUS set up unfortunately is not suited for this task. The future detector of the KM3NeT collaboration does not have this size limitation.

2 The KM3NeT neutrino telescope

The KM3NeT Collaboration is currently constructing a neutrino telescope of the same name in the Mediterranean Sea [7]. In this study, possibilities of acoustic neutrino detection with the device are investigated. The telescope will be composed of several detection units anchored at the sea bed, which are kept taut vertically by a buoy. Each detection unit of KM3NeT/ARCA has a length of about 700 hundred meters. There are 18 modules on each unit with spacings of 36 m, each housing photomultiplier tubes and one acoustic sensor. The detection units are about 90 m apart from each other. While the primary purpose of the acoustic sensors is the position calibration of the detector, it can also be used to study the feasibility of the acoustic neutrino detection. Since the response function has not been fully measured for the acoustic sensors used in the KM3NeT-modules, it is assumed that they have the same characteristics as the hydrophones in AMADEUS [5]. A building block for the final detector comprises 115 detection units. The signals from transient background sources and neutrino signals produce different signatures in such a large detector. The amount of data produced by approx. 2000 sensors, which have to be sampled at a frequency of at least 100 kHz for acoustic neutrino detection, is very large. It is not feasible to store all produced data, so the following signal classification will be using only the peak to peak amplitude and the arrival time of the sound.

3 Simulating the event samples

Both the background and the signals were simulated using the software SeaTray that was developed internally. The background consists of spherically emitting sound sources, that are placed randomly in a cylindrical volume around the detector. The initial amplitude is also chosen randomly in the range of signals recorded by AMADEUS. Another class of background signals are emissions from the acoustic positioning system. These differ from the first kind, since their source is always located close to the detector and their amplitude is fixed. Finally, the third part of the background are random coincidences between signals in multiple hydrophones. The relative occurrence for the samples was chosen as 75:20:5 respectively, which keeps the order of the actual rates and guarantees a statistical significant number of events for each class. The neutrino signals are generated by code adapted from the ACoRNe project [8]. The interaction vertex is chosen randomly in the same cylinder as for the background samples. The direction of the neutrino is randomly chosen from an uniform distribution on the upper hemisphere. This can be done because the earth is opaque for neutrinos with energies exceeding 10^{18} eV, which is roughly the energy threshold for acoustic detection. The shower energy is determined from neutrino energy by using an estimation for the Bjorken- y value, while the neutrinos are generated according to a E^{-1} power spectrum. The amplitude is calculated by a parametrization from the shower energy. The arrival time of the pressure pulse at the sensors is determined by raytracing the path from the source. The ration of background to signal events in the samples is varied from 1 to 3, while the typical size for a training or testing sample is 20 k events.

4 Signal classification

The signals are classified by using multivariate analysis. A set of features is calculated from the signature in the detector, which is independent of the exact layout of the detector. This ensures that

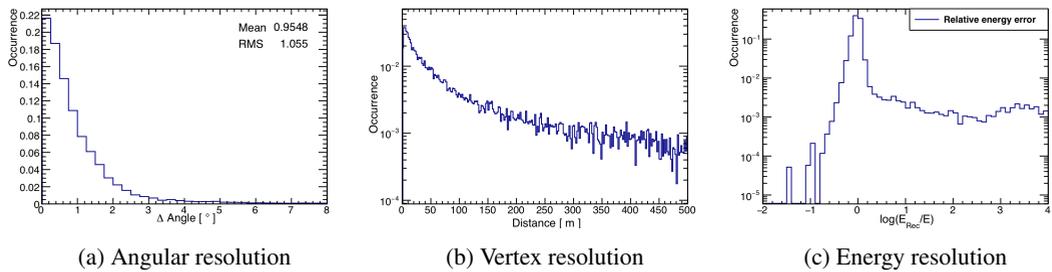


Figure 1: Resolution of the event reconstruction for 20k simulated neutrinos. Panel (c) shows the relative occurrence of the ratio of the reconstructed and the true shower energy. For 90% of the reconstructed events the error is smaller than 75% of the true energy.

the classification will still be functional if small changes to the detector occur, e.g. the failure of a detection unit. First, the position of the sound source is reconstructed. In order to calculate the other features, the positions of all sensors that triggered for that event are used to generate a 3-dimensional point cloud. The distances are normalized, so that the center of gravity of the point cloud is in the origin, and the variance of the distances from the origin is normalized to 1. A singular value decomposition is used to process the data. The singular values obtained this way are the first set of features. The singular vector corresponding to the smallest singular value is also the normal vector to the pancake plane. It represents the direction of the neutrino (if it is not a background event) with an accuracy of about 2.5° . The sound disk is reconstructed by using the latter vector as normal vector and the center of gravity of the point cloud in the detector as starting point. The distance between the reconstructed interaction vertex and this plane is another feature. Another value used is the sum of the squared distances of the sensors from the reconstructed plane. Since the amplitude for neutrino signals should decrease with the distance from the plane, the correlation coefficient between those values is also used as feature. Lastly, a "likelihood" L of the signature is calculated by summing up $L = \sum_i \left(\frac{\alpha_i}{\sigma}\right)^2$, where α is the angle between the pancake plane and the direction in which the sound is emitted to reach the sensor with index i . The value $\sigma = 3.2^\circ$ represents the width of the pancake, taking into account the error of the vertex and direction reconstruction.

This feature vector is then analyzed by using the machine learning tools from the OpenCV [9] library. Boosted decision trees, random decision trees and a simple decision tree are used here. The algorithms were trained on an individual sample of 20k events each, minimizing the cross entropy. The classification is carried out as a majority vote. In total, this removes 99.99% of background events (tested on 120k events), while 98% of neutrinos remain in the sample, see Table 1.

5 Energy reconstruction

Now only the energy of the neutrino is missing for the complete event reconstruction. It should be noted that there is no way to determine the Bjorken- y , so even a perfectly reconstructed shower energy would only impose a lower limit for the neutrino energy. In the previous section, the parameters of the event have been reconstructed in a step by step manner. The energy however is extremely dependent on the direction and vertex reconstruction, so a simple calculation using the previously reconstructed values will often produce wrong results due to small errors in the input values. One possible way to circumvent this is using a combined fit for all parameters. A toy Monte Carlo algorithm is used

Table 1: Remaining events after each step of the background reduction. The first three steps were already implemented for AMADEUS.

Background reduction method	Events per year
Source position reconstruction possible	10000000
Waveform classified as bipolar pulse	315000
Temporal and spatial clusters removed	63000
Pancake–signature detected	7

to create the arrival times and amplitudes at the sensors for a given set of vertex position, shower direction and shower energy. The MINUIT package from ROOT is used to minimize the following function f by varying the starting values, where the sum runs over:

$$f = \sum_i \left[\frac{\Delta t_i^2}{\sigma_t^2} + \left(H(A_i^{rec} - A_i) \frac{A_i - A_i^{rec}}{\sigma_A} + H(A_i - A_i^{rec}) \frac{A_i(A_i - A_i^{rec})}{\sigma_A A_i^{rec}} \right)^2 \right] \quad (1)$$

where Δt_i is the difference between the measured and reconstructed arrival time, $H(x)$ is the Heaviside step function, A_i the (reconstructed) amplitude at sensor i and the σ are the respective expected deviations. The resulting errors for this combined fit are much smaller than for the step by step reconstruction. The angular resolution for a KM3NeT Building Block is 1° and the distribution of the errors is shown in Fig. (1a). The average vertex error is 250 m, see Fig. (1b). The energy resolution for the shower energy is 30%, as shown in Fig. (1c).

6 Conclusions

The rate of transient background sources for acoustic neutrino detection can be reduced by 4 orders of magnitude when the sound emission pattern is taken into account. With the rates shown in Table 1, this would reduce the background events from 63 k events/year to 7 events/year. A further reduction of the background rate is still necessary, as the the expected rate of neutrinos is extremely small. This can be achieved by using quality cuts on the reconstruction of the combined fit of vertex position, shower direction and energy. Preliminary results show a possible further reduction of the background by 80%, while also removing about 20% of the neutrinos.

References

- [1] K. Greisen, Phys. Rev. Lett. **16**, 748 (1966)
- [2] G.T. Zatsepin, V.A. Kuzmin, JETP Lett. **4**, 78 (1966)
- [3] V.S. Berezinsky, G.T. Zatsepin, Phys. Lett. B **28**, 423 (1969)
- [4] G.A. Askariyan et al., Nucl. Instr. and Meth. **164**, 267 (1979)
- [5] J. Aguilar et al., Nucl. Instrum. Methods Phys. Res., Sect. A **626-627**, 128 (2011)
- [6] M. Neff, Dissertation, Friedrich-Alexander-Universität Erlangen-Nürnberg (2012)
- [7] S. Adrián-Martínez et al. (KM3NeT collab.), J. Phys. G: Nucl. Part. Phys. **43**, 130 pages (2016)
- [8] S. Bevan et al., Astrop. Phys. **28**, 366 (2007)
- [9] *Open source computer vision library* (2014), <http://opencv.org/>