DEUTSCHES ELEKTRONEN-SYNCHROTRON

DESY-THESIS-1998-014
June 1998

On the Rejection of Atmospheric Muons in the AMANDA Detector

by

A. Biron von Curland

ISSN 1435-8085

NOTKESTRASSE 85 - 22603 HAMBURG

To be sure that your reports and preprints are promptly included in the HEP literature database
send them to (if possible by air mail):

DESY
Zentralbibliothek
Notkestraße 85
22603 Hamburg
Germany

DESY
Bibliothek
Platanenallee 6
15738 Zeuthen
Germany

On the Rejection of Atmospheric Muons in the AMANDA Detector

Diplomarbeit

Humboldt–Universität zu Berlin
Mathematisch–Naturwissenschaftliche Fakultät I
Institut für Physik

eingereicht von

Alexander Prinz Biron von Curland
geb. am 18.09.1972 in München

Berlin, März 1998

Abstract

The young field of high energy neutrino astronomy gives the opportunity to open a new observational window to the universe. This provides a large variety of prospects to investigate new phenomena in astro and particle physics. One of the major challenges neutrino telescopes are facing is the rejection of background from atmospheric muons. One approach to achieve the required rejection rate is to cut on quality parameters obtained from the reconstruction of detected events. In this work an alternative approach, based on artificial neural networks, is attempted for the AMANDA experiment. A first aspect is the comparison between a neural network analysis and a "conventional" analysis, when both are restricted to use the same parameter set. This is performed with Monte Carlo data simulated for AMANDA-II, the year 2000 version of the AMANDA array. It is shown that the neural network analysis can obtain a $\sim$ 30 % better signal passing rate while having the same background rejection. If further input parameters are added, not only an almost 50 % increase in signal passing rate can be achieved, but also an improved background rejection. During the second main part of this work, another neural network analysis is performed for the AMANDA-B4 detector. The results of this analysis are subsequently applied to data taken with that detector in 1996. Finally the two most likely neutrino candidates from this analysis of one day of measured data are presented.

Contents

1 Introduction

Neutrinos, they are very small.
They have no charge and have no mass
And do not interact at all.
The earth is just a silly ball
To them, through which they simply pass,
Like dustmaids down a drafty hall
Or photons through a sheet of glass.

J. Updike [1]

One of the most fascinating pictures nature offers us is the star spangled sky during a dark night. This is probably the reason, why astronomy is one of the oldest sciences, which was always pursued from ancient cultures to the modern days. While it also served other purposes (e.g. the ancient Egypts used astronomy to forecast the regular floodings of the river Nile [66]), its main purpose always was to explain the phenomena seen in the sky. Despite the use of telescopes ever since Gallilei pointed this invention of a Dutch sailor into the sky, the underlying principle of astronomy has stayed the same for centuries: Detecting photons with the eye or (starting last century) with a photographic plate. Nowadays telescopes detect photons with energies between $\sim 10^{-9}$ eV (radio waves) and $\sim 10^{10}$ eV (gamma rays).

1911/12 Victor Hess discovered the cosmic rays during his balloon flights [25] and the aim to reveal their origin eventually became one of the major challenges in astrophysics. By now individual cosmic rays with energies up to $\sim 3 \cdot 10^{20}$ eV have been detected. But since they are deflected in cosmic or galactic magnetic fields, the question about the origin of these highest energy cosmic rays has yet to be answered. There are three main approaches to find the answer: The search for photons of similar energies, the search for particles of such high energies that they are not significantly deflected in magnetic fields and the search for high energy neutrinos. Interestingly, all three approaches have a common feature: Cosmic rays interact with the natural environment of a detector and produce secondary particles. All properties of the primary particle (energy, direction, particle identification, etc.) have to be deduced from signals induced by the secondaries.

1. Gamma rays can be detected in large air shower experiments. Among the various designs, Čerenkov telescopes are the (so far) most successful to locate gamma ray point sources. They resemble conventional telescopes, but do not register the primary gamma ray. Instead, they register radiation emitted by charged particles from the electro-magnetic cascade caused by the gamma ray's collision with the earth's atmosphere. Hadronic cascades caused likewise by incident protons or other charged particles can be rejected due to different signatures [69, 55]. The WHIPPLE telescope in Arizona [70], HEGRA in La Palma [32], the CANGAROO array in Australia [20] and others, pave(d) the way for this approach[2]. They have thresholds of 300 GeV

[1] *Cosmic Gall* in *Telephone Poles and other Poems*, André Deutsch, London, 1964

[2] The mentioned experiments are not dedicated gamma ray experiments. They are rather combined experiments for gamma rays as well as for charged cosmic rays.

and higher. New projects try to lower this number to some 20 GeV [50] to get an overlap with the highest energies gamma-ray satellites can measure. In the TeV region several point sources have been located [33], whereas searches for point sources in the PeV region have been inconclusive so far. This is not really surprising, since the 2.7 K background radiation opaques the universe for very energetic photons. Unambiguous gammas have only been detected up to $\sim$ 50 TeV so far [65].

2. Energetic charged nuclei can be detected with air shower arrays. Present arrays measure the energy and the chemical composition of the primary particle spectrum. An alternative aim is to look at such high energies, that the deflection due to magnetic fields is small. The energy scale where this deflection becomes less dominant can be estimated from the (measured) isotropy of the flux of the charged particles. This isotropy is high for energies below 10^{17} eV but starts decreasing heavily above these energies [30]. This is in accordance with cosmological calculations for galactic magnetic fields, leading to a deflection of only some degrees for energies above 10^{19} eV [61]. An example of running experiments is the Japanese AGASA experiment [3], which is a prototype of the future PIERRE AUGER project [8]. Particles with energies up to $\sim 3 \cdot 10^{20}$ eV have been detected [31], but no unambiguous point source was found yet [46].

3. The third attempt is to probe the universe via neutrinos. There are several advantages to this concept: Neutrinos are not deflected by galactic magnetic fields or absorbed in intergalactic dust clouds. Furthermore they are not scattered but carry direct information from the whole volume of cosmic ray sources rather than only surface information[3]. These advantages come at the price of difficult detection of the neutrinos. These characteristics all have a common reason: The tiny interaction cross section for only weakly interacting particles.

Until the 1990s cosmic neutrinos could only be investigated by experiments originally build for other purposes; like the FRÉJUS detector [58] which was designed to look for proton decays, or (SUPER-)KAMIOKANDE [42] which's main purpose now is the study of solar neutrinos. In the late 70s, research and development for a large neutrino telescope was started by the (now abandoned) DUMAND experiment [22]. DUMAND laid out the basic principle for detecting high energy cosmic neutrinos: A large grid of photomultiplier-tubes (PMTs) is deployed in a transparent medium. A high energy neutrino interacts with the environment in the vicinity of the detector. In case of a charged current interaction a massive lepton is produced, which in average receives 2/3 of the neutrino's energy. This $\tau^{\pm}$, $\mu^{\pm}$ or $e^{\pm}$ emits Čerenkov radiation, which is detected by the PMTs. In order to register enough events, a large volume has to be monitored. Due to the short lifetime (i.e. short distance traveled) of tauons and the high showering rate of electrons, the telescopes are optimized to detect muons and their neutrinos. In order to shield them from the enormous background of atmospheric muons (see figure 2.6), the detectors are mounted in great depths. Since transparent media are required throughout the volume, deep lakes, the

[3] In potential cosmic neutrino sources like supernova bursts, densities $\rho \gtrsim 10^{12}$ g cm^{-3} might exist. In this case the neutrino mean free path becomes smaller than the size of these objects [9]. The volume/surface argument then becomes less important, but is still true in comparison to photons.

deep sea or deep ice are the only possible places for such detectors. Today, there are four projects for neutrino telescopes:

The BAIKAL experiment was christened after Lake Baikal, where the first running neutrino telescope is located. Being the only such experiment in a lake its characteristics are also determined by the lake: Despite being the deepest lake on earth, Lake Baikal restricts the experiment to the rather shallow depth of $\sim$ 1.1 km. Its ice cover during the winter provides a stable deployment platform, which simplifies deployment severely compared to ocean detectors. To overcome the high photon absorption rate of water, the neutrino detector needs a relatively narrow PMT spacing. This in turn leads to a low energy threshold [10]. The experiment started with 36 PMTs ("NT-36") deployed in 1993, was extended step by step, has 144 PMTs at present and aims for 192 PMTs ("NT-200") in 1998 [12]. This is the only neutrino telescope having published unambiguous (atmospheric) neutrino events so far [11].

NESTOR is a Greek-Italian project headed by the university of Athens. Its 168 PMTs shall be deployed in a hexagonal structure of several floors in the Mediterranean sea, close to Pylos, Greece [49].

The ANTARES collaboration is also planning to build a deep ocean telescope in the Mediterranean sea. A fully equipped 3-dimensional test array is to be deployed within the next two years. The site for this so called *demonstrator* near Toulon is at a depth of 2.5 km, but the final site is meant to be deeper and might be located somewhere else [5]. ANTARES also plans to emphasize non-astronomical research like ocean science [6].

At present AMANDA is the biggest running experiment with 424 PMTs deployed after the 97/98 season [4]. It is located at the geographic south pole, using the infrastructure of the American *Amundsen Scott South Pole Station*. It does not use water as its transparent detection medium but very pure natural ice from the previous interglacial period. Compared to ocean water, ice has a low scattering length but a high absorption length. This simplifies energy reconstruction, but necessitates a more sophisticated track reconstruction.

One of the major challenges all of these projects are facing is background: Despite the depth, down-going atmospheric muons occur several orders of magnitude more frequent than cosmic neutrinos. The most important mean to suppress this background is to cut on the reconstructed zenith angle and only accept up-going muon events. However due to a remainder of mis-reconstructed atmospheric muon events, further criteria have to be applied.

A standard approach is to cut on various further parameters. An alternative is to use a neural network analysis. One main subject of this work is a comparison between these two approaches. This comparison is performed with simulated data for the AMANDA-II array. This detector is planned to be completed by the year 2000. The other major investigation concerns the application of the neural network analysis method to data. In this work data taken with the AMANDA-B4 array in 1996 is analyzed by neural networks. The two most likely neutrino candidates from this analysis are presented.

2 Neutrino Telescopes

2.1 High Energy Astrophysics

2.1.1 Gamma Ray Astronomy

Two very different efforts are undertaken to detect high energy cosmic gamma rays: Direct measurements are only possible outside the earth's atmosphere with satellites. The main restrictions here are cost, weight and size of the instrument. The flux of cosmic gamma rays from point sources in the energy range accessible to satellite based experiments decreases like $E^{-(1.9+\epsilon)}$, where $0 < \epsilon < 0.6$. This makes it very unlikely for the small instruments to be hit by a gamma ray of very high energy (VHE gamma ray). Therefore existing satellites are not even designed to measure $E > 30$ GeV photons accurately. Since there are uncertainties concerning the background, no diffuse flux in this energy range could be measured so far.

An alternative approach to detect cosmic gamma rays are large, ground based air shower Čerenkov experiments. Here, the main concern is to suppress the dominating Čerenkov background from cosmic ray protons. So far this suppression can best be achieved by the Čerenkov telescopes used within these experiments. As there has not been an all sky survey so far, no diffuse gamma ray flux in that very high energy range is known either. Since such air shower experiments only detect high energy gammas ($E \gtrsim 300$ GeV for gammas detected in existing air shower experiments), the two efforts are seen as complementary rather than rivalry.

High energy gamma ray astronomy has some fundamental problems: Even though photons might originate from the center of stellar objects, they are heavily scattered and absorbed on their way to the surface. Therefore it is difficult to deduce direct information on anything but surface properties. Gamma rays also have a high interaction probability with infrared or cosmic microwave background radiation (CMBR)

$$\gamma\gamma_{CMBR} \rightarrow e^+e^- \tag{2.1}$$

and cosmic or galactic dust. This leads to a mean free path for 10 TeV photons of the order of 10 Megaparsec [28], further decreasing with energy.

2.1.2 Charged Cosmic Rays

The spectrum of charged particles extends up to energies of a few 10^8 TeV, see figure 2.1. It is assumed that most of the very energetic particles originate from supernovae or AGNs [15]. The knee in the spectrum probably indicates an energy scale above which standard acceleration processes become unlikely. The existence of particles with energies beyond the ankle might hint at another, rarer but more powerful acceleration mechanism. The most energetic particles emitted from such accelerators are supposed to be heavily attenuated between source and potential detection due to their interaction with the microwave background radiation. This should lead to the so-called Greisen-Kuz'min-Zatsepin (GKZ) cutoff in the cosmic ray spectrum. Protons with energies beyond this cutoff, i.e. with energies $\gtrsim 5 \cdot 10^7$ TeV, registered on earth therefore have to originate from sources at distances of some 10 Megaparsec or less [28]. So these ultra high energy (UHE) cosmic

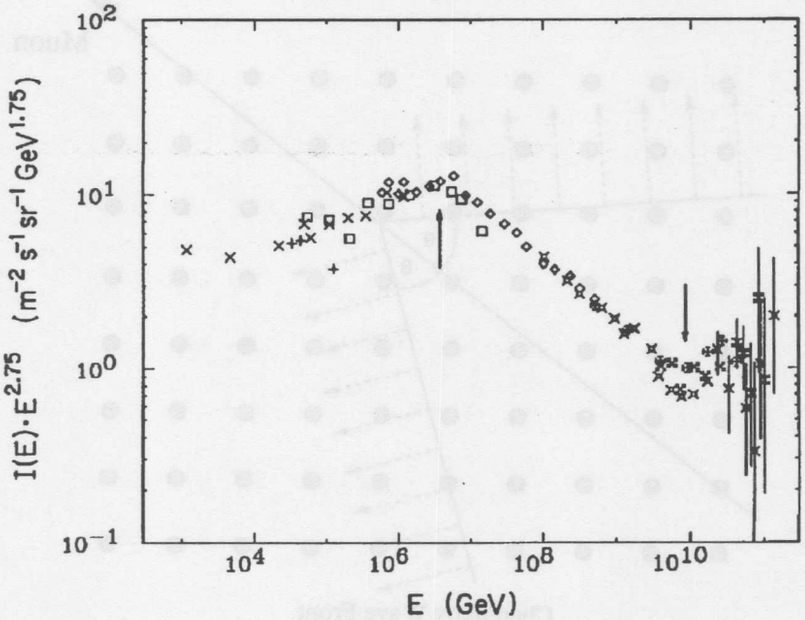


Figure 2.1: Flux of high energy cosmic rays, multiplied by $E^{2.75}$ for better visibility of structures like the knee near 1 PeV and the ankle near 10 EeV (both marked by arrows)

rays can only probe our local region of the universe. Cosmic rays of energies $\lesssim 10^{19}$ eV on the other hand loose their directional information due to deflections in galactic or intergalactic magnetic fields [30]. Compared to electrically neutral particles which are only slightly deflected via gravitational lenses, this is a serious disadvantage.

2.1.3 Neutrinos

Since they interact only weakly, neutrinos have neither the problem of absorption (like photons) nor that of deflection (like charged cosmic rays). They can travel over a Hubble radius still carrying the information from their point of origin. As always, such advantages come at a price: They are hard to detect, and in order to do so one has to monitor a large volume, i.e. build a large detector.

2.2 Detection of High Energy Cosmic Neutrinos

2.2.1 Čerenkov Light

Energetic particles with velocities $v > c/n$ (n being the refractive index of the medium) emit Čerenkov radiation [53]. For muons in ice or water (typical environments for neutrino telescopes with $n \sim 1.3$) this emission starts at $E_\mu \sim 300$ MeV. The photons are emitted under an angle $\theta = \arccos(\frac{1}{\beta n})$ and form a so-called Čerenkov cone. In the energy range interesting for neutrino telescopes, $\beta = 1$ and $n = const.$, i.e. the Čerenkov angle is independent of the muons' energy. The photons have a continuous energy spectrum. The

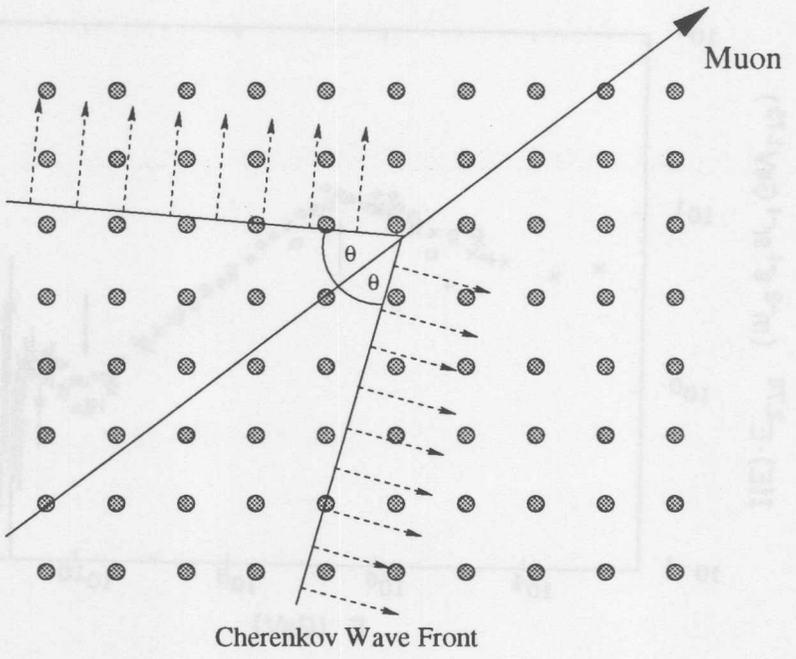


Figure 2.2: Čerenkov cone and track of emitting muon traveling through a PMT lattice.

energy loss (due to Čerenkov radiation) per unit length is given by

$$\frac{dE}{dx} = \frac{4\pi^2 e^2}{c^2} \int \left(1 - \frac{1}{\beta^2 n^2(\nu)}\right) \nu \, d\nu \; . \tag{2.2}$$

So the number of photons of a given frequency is proportional to $\nu \, d\nu$. This is why the blue light is dominant for Čerenkov radiation.

The emission of Čerenkov light by the muon itself is not the only photon source relevant. Muons also loose energy due to ionisation (continuous energy loss), production of δ-electrons, bremsstrahlung, e^+e^--pair production, muon nucleus interactions and $\mu^+\mu^-$-pair production (all stochastic processes). All of these processes lead to further Čerenkov photons from shower particles. The energy loss due to the stochastic processes is rising with energy. Thus an energetic muon emits more light per unit track length, making it easier to be detected and allowing a better energy estimation.

All emitted photons can interact with the environment before they hit a PMT. These interactions are stochastic processes only, namely scattering and absorption. Their probabilities are mainly determined by the optical properties of the water/ice, but (in a first approximation) they are independent of the photon 's frequency. The optical properties are normally parameterized by an effective scattering length λ_{eff} and an absorption length λ_{abs}, or by the attenuation length λ_{att}. In the case of diffuse scattering (like in AMANDA-A), there exists a simple relationship:

$$\lambda_{att} = \sqrt{\frac{\lambda_{eff}\lambda_a}{3}} \; . \tag{2.3}$$

	AMANDA-A 900 m	AMANDA-B 1800 m	Lake Baikal	Oceans
absorption length λ_{abs}	67–234 m	95 ± 5 m	21 ± 1 m	$40 - 55$ m
eff.scattering length λ_{eff}	≈ 60 cm	24 ± 2 m	$150 - 300$ m	$150 - 300$ m
attenuation length λ_{att}	$\lesssim 7$ m	≈ 30	≈ 20	≈ 50

Table 2.1: Comparison of important optical parameters for various possible detector sites. Data taken from [16]. The absorption length of the ice at AMANDA-A is wavelength dependent: $\lambda_{abs} = 67$ m for a photon wavelength of 500 nm, whereas $\lambda_{abs} = 234$ m for a wavelength of 410 nm.

Compared to water detectors, ice detectors have a large λ_{abs}, a similar λ_{att}, but a small λ_{eff} (see table 2.1). The reason is the low absorption in very pure ice crystals on the one hand, but higher scattering due to residual impurities (air bubbles in shallow ice, dust, etc.) on the other hand. The high scattering rate in ice complicates track reconstruction. Contrary to this, energy reconstruction can actually benefit from scattering: For ice detectors with very small λ_{eff} but large λ_{abs} (like AMANDA-A), scattering enhances the probability for a given photon to hit any PMT. The detector has thus better calorimetric properties.

2.2.2 The Detectors

Deep underwater/ice telescopes consist of a lattice of PMTs spread over a large transparent volume to detect Čerenkov light. Since the detectors are placed in great depths, one shields the PMTs with a pressure sphere made of very transparent boron-silicate glass. Other (e.g. electronic) devices might also be included into these housings. This equipped pressure sphere is then called an optical module (OM). Depending on the optical properties of the medium and the distances to the various OMs, a certain fraction of the Čerenkov photons emitted by a muon will hit a PMT. For AMANDA-B, a 1 TeV muon in average induces one photoelectron in a PMT at 25 m distance to the track [43]. The PMT's time and amplitude information is then recorded. If enough PMTs are hit, the muon's track and its energy can be reconstructed from the time and amplitude information gathered. With this approach, it is not possible to discriminate between μ^+ and μ^-, and therefore only the combined fluxes of particle and antiparticle are measured.

Several theoretical calculations indicate that a size of 1 km^3 is desirable for large neutrino detectors [26]. To estimate the required size of the detector, one can compare neutrino fluxes to the gamma ray fluxes known from air shower experiments: Galactic accelerators might produce high energy hadrons[1] (mostly protons), which interact with a matter or photon fields to produce pions. These are supposed to be the sources for high energy photons and neutrinos via their decays, e.g.:

[1] Some theories doubt such processes exist. They describe cosmic accelerators for electrons only. If this was true, there would be no cosmic point sources for neutrinos.

$$
\begin{array}{llll}
p+p \rightarrow N_1\pi^0 + N_2\pi^- & +N_3\pi^+ & +X & N_1 \approx N_2 \approx N_3 \\
\pi^0 & \rightarrow 2\gamma & & \\
\pi^- & \rightarrow \mu^- & +\overline{\nu}_\mu & \\
 & \mu^- & \rightarrow e^- + \overline{\nu}_e + \nu_\mu & \\
\pi^+ & \rightarrow \mu^+ & +\nu_\mu & \\
 & \mu^+ & \rightarrow e^+ + \nu_e + \overline{\nu}_\mu &
\end{array}
\tag{2.4}
$$

So the original muon neutrino flux Φ_ν is expected to be of the same order of magnitude as the gamma ray flux Φ_γ. Two main processes can alter this approximate 1:1 ratio: The gamma ray flux might be higher due to electro-magnetic showering within the source, producing more (albeit less energetic) photons out of the initial gammas. In contrast, if the source is dense enough, absorption of photons within the source might become relevant. In this case, the neutrino flux reaching earth might be larger.

Taking into account the absorption probability of gammas A_γ (within the source and on their way to earth), the detected flux of photons [65] and the interaction probability of neutrinos inside the earth [34], one can calculate the upward muon flux. For example for binary stars, which some years ago were considered as possible neutrino point sources, the upward muon flux can be estimated to

$$
\Phi(\uparrow \mu) \lesssim \frac{2 \text{ events}(E_\mu > 1\ TeV)}{10^5 m^2 yr} \times \frac{1}{(1 - A_\gamma)} \ , \tag{2.5}
$$

for an energy spectral index γ in the range 2.1 to 2.3[2] [24]. Similar estimations for various other potential sources justify the quest for effective detection areas > 1 km^2 [16]. In order to achieve such an "effective area" an equipped volume of ~ 1 km^3 is needed [26]. The task to equip such a large volume with PMTs and the aim to keep costs at a low level enforces an OM spacing of $\gtrapprox 10$ meters. Such a spacing in turn leads to a minimum threshold of $\gtrapprox 10$ GeV for reconstructible tracks. This compares to the MeV limits for solar neutrino experiments like IMB or KAMIOKANDE.

A key feature distinguishing neutrino telescopes from most other detectors in the field of high energy physics and astronomy is the inherent use of their natural environment as part of the detector. This leads to an interesting property of such detectors: The effective volume,

$$
V_{eff} = A_{eff} \cdot R_\mu \tag{2.6}
$$

becomes considerably larger than the actual detector, i.e. the detector *sees beyond its limits*, see figure 2.3. The effective area A_{eff} from equation 2.6 is the constant of proportionality incident particle flux and detection rate. R_μ, the mean muon range is given by

$$
R_\mu = \frac{1}{b} \cdot \ln\left(\frac{bE_\mu}{a} + 1\right) \ , \tag{2.7}
$$

where $a = 2$ MeV/cm and $b \sim 3.4 \cdot 10^{-6}$ cm^{-1} [72]. So $R_\mu \sim E_\mu$ up to $E_\mu \sim 1$ TeV and $R_\mu \sim \log E_\mu$ at higher energies. Since multi-TeV muons travel several kilometers, this effect ($V_{eff} > V_{detector}$) is not only relevant for present detectors, but still for km^3-sized extensions.

[2] The spectral index γ is defined via the flux. For example for photons: $\Phi_{photon} = const \times E_{photon}^{-\gamma_{photon}}$.

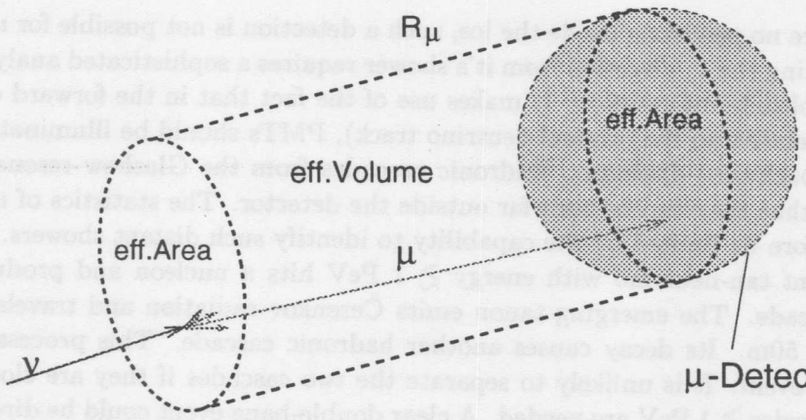


Figure 2.3: The effective volume, detector volume and the effective area of a neutrino telescope (≡ muon detector). Contrary to accelerator experiments the reaction ($\nu \rightarrow \mu$) is not required to be inside the detector.

2.2.3 Non-Muon Neutrinos

Neutrino telescopes are optimized to detect muon neutrinos, but there are signatures for charged current interactions of the two other flavors as well.

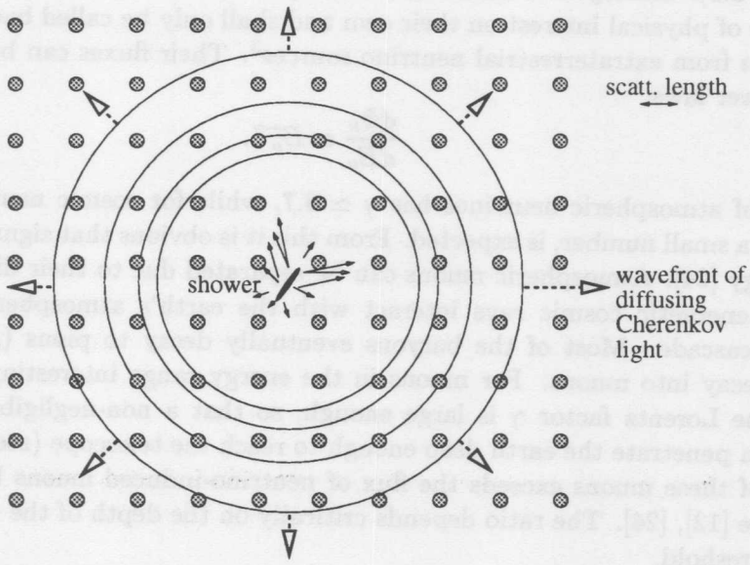


Figure 2.4: Point-like events like energetic electro-magnetic or hadronic cascades give spherically symmetric Čerenkov fronts. This symmetry concerns timing distributions only - not photon density distributions.

Single showers of energies 6.4 PeV (see figure 2.4) can be assigned to hadronic or electro-magnetic cascades from the Glashow resonance:

$$\overline{\nu}_e + e^- \rightarrow W^- \rightarrow shower \ . \tag{2.8}$$

Since there are no positrons inside the ice, such a detection is not possible for ν_e, but only for $\overline{\nu}_e$. Deducing the $\overline{\nu}_e$ direction from it's shower requires a sophisticated analysis, mainly based on amplitude information. It makes use of the fact that in the forward direction of the shower (relative to the original neutrino track), PMTs should be illuminated stronger than in the backward direction. Hadronic cascades from the Glashow resonance should be so bright that they can be seen far outside the detector. The statistics of ν_e detection should therefore be limited by the capability to identify such distant showers.

An incident tau-neutrino with energy $\gtrsim 1$ PeV hits a nucleon and produces a local hadronic cascade. The emerging tauon emits Čerenkov radiation and travels a distance $d = c\beta\gamma\tau \gtrsim 50\text{m}$. Its decay causes another hadronic cascade. This process is called a *double-bang event*. It is unlikely to separate the two cascades if they are close together, therefore energies $\gtrsim 1$ PeV are needed. A clear double-bang event could be direct evidence for the existence of the so far undetected tau-neutrinos. Source tracing is still possible via the tauon track. Again no distinction between particle and anti-particle is possible. Since the detection of ν_τ does not profit from the long range of a muon (see equation 2.7), the statistics for these processes will be limited.

2.3 Background

There are two different sources of background muons inherent to large neutrino detectors (see figure 2.5): atmospheric muons and atmospheric neutrinos. Both kind of background signals are of physical interest on their own and shall only be called background to distinguish them from extraterrestrial neutrino sources[3]. Their fluxes can be approximated by simple power laws:

$$\frac{d\Phi_\nu}{dE_\nu} \propto E_\nu^{-\gamma}. \tag{2.9}$$

The flux of atmospheric neutrinos has $\gamma \simeq 3.7$, while for cosmic neutrinos $\gamma \simeq (2 + \epsilon)$, where ϵ is a small number, is expected. From this it is obvious that signal to noise improves with energy [24]. Atmospheric muons can be separated due to their direction.

When energetic cosmic rays interact with the earth's atmosphere, they produce a hadronic cascade. Most of the baryons eventually decay to pions (*pionisation*), which in turn decay into muons. For muons in the energy range interesting for neutrino telescopes, the Lorentz factor γ is large enough, so that a non-negligible fraction of these muons can penetrate the earth deep enough to reach the telescope (see figure 2.6). In fact the flux of these muons exceeds the flux of neutrino-induced muons by several orders of magnitude [12], [24]. The ratio depends critically on the depth of the detector and on the energy threshold.

The background of atmospheric muons is suppressed severely by a simple cut on the zenith angle of the reconstructed track: Up-going muons are considered to be neutrino-induced signal, while down-going muons are all ignored as atmospheric muons. Theoretically, with a perfect track reconstruction, this single cut would be sufficient to suppress such background. In reality, some tracks are reconstructed into the wrong direction and

[3] A diffuse flux of neutrinos from the interaction of cosmic rays with interstellar or intergalactic dust, the sun's atmosphere or the CMBR exists [24]. Hope is that extraterrestrial point sources can also be discovered.

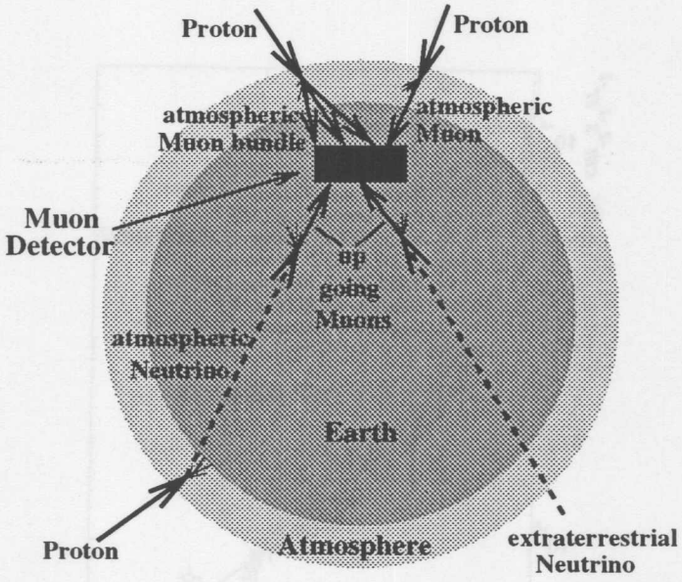


Figure 2.5: Signal comes from muons originating from cosmic neutrinos (bottom right). Indistinguishable background is due to muons induced by atmospheric neutrinos (bottom left). These, as well as atmospheric muons (top right) are due to cosmic rays interacting with the earth's atmosphere. Sometimes this interaction not only causes the occurrence of one, but of several coincident muons at the detector (top left).

additional cuts are needed. Due to the necessary zenith angle cut, it is said that neutrino telescopes can only *look down*, i.e. a telescope can only detect neutrinos from the opposite hemisphere (or, in best case, $10° - 20°$ above horizon). This implies that at least two such instruments (as antipodic as possible) are needed to scan the whole sky. In this sense AMANDA is actually needed as the complementary partner by any of the other planned neutrino telescopes since they are all located on the northern hemisphere.

The other important (though much rarer) source of background events, the *atmospheric neutrinos* also originate from the interaction between incident cosmic rays and the atmosphere: Due to individual lepton number conservation, one $\bar{\nu_\mu}$ (ν_μ) is produced for every μ^- (μ^+) produced and another ν_μ ($\bar{\nu_\mu}$) for every μ^- (μ^+) decaying. This is the same as in equation 2.4. Some of these neutrinos interact in or near the detector (exactly like cosmic neutrinos do) and produce a muon. These atmospheric neutrino-induced muons are indistinguishable from signal muons. Therefore individual events cannot be assigned to potential cosmic neutrino point sources. Instead, peaks within the homogeneous background have to be found. The higher the angular resolution of the detector is achieved, the more significant these peaks become.

Both kind of background events can be utilized however: First of all they are a proof of method: These sources are known to exist, so their detection indicates the capabilities of neutrino telescopes. Furthermore, fluxes are known to within 10-20% [52, 73]. They can thus be used to calibrate the telescope. In addition, atmospheric neutrinos can probe uncharted territory in neutrino oscillation parameter space [16]. However, since neutrino

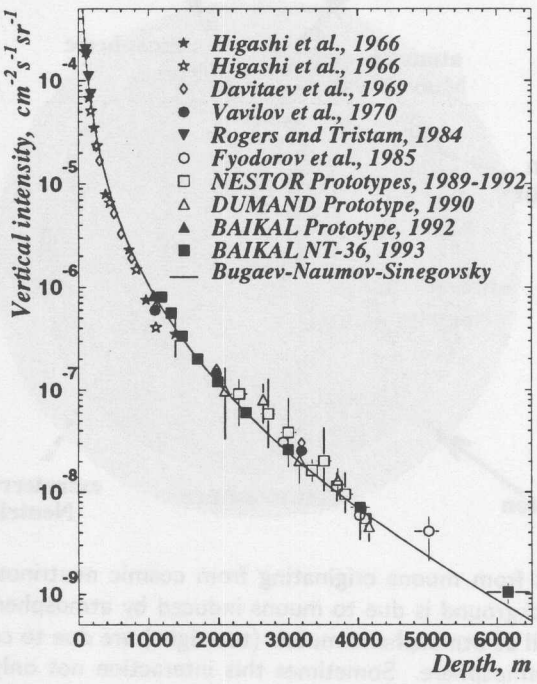


Figure 2.6: Vertical flux of atmospheric muons, measured at various sites [12]. This explains the quest for sites located as deep as possible in order to minimize the number of downgoing events.

telescopes are optimized to (high energy) muon neutrinos they are not expected to help solve the atmospheric neutrino anomaly.

Contrary to ice detectors in a sterile, very cold ($\sim -40°C$) environment, water detectors have to cope with a high level of noise originating mostly from bioluminescence. Bioluminescence is especially important since it may not only be random noise: Light emitting bacteria can stimulate other bacteria, leading to coincident hits in various PMTs [72]. Deep ocean detectors have one further major background: The natural radioactivity of ^{40}K. This potassium isotope, which is present in the dissolved salt, emits beta rays, which in turn can emit Čerenkov photons. The beta rays themselves are readily absorbed in water, but their Čerenkov photons can trigger PMTs. To cope with noise, BAIKAL uses and ANTARES plans to use a local coincidence trigger [10, 5]. That is to say a PMT signal is disregarded as long as no neighboring PMT signal is reported within a few nanoseconds.

2.4 Scientific Goals

Neutrino telescopes, like most other large detectors in the field of physics, are no single purpose instruments. Their primary goal is the detection of cosmic neutrinos. From directional and kinetic information, one aims to deduce information on point sources. All

standard models of such sources predict some acceleration mechanism producing energetic charged particles, which then hit a so-called "beam-dump", which is just dense (on interstellar scale) material or a dense photon field. There the charged particles interact with target material, producing pions which then decay to neutrinos, see equation 2.4. Some of the possible or expected sources are supernovae, AGNs and GRBs. Nowadays supernovae seem to be understood fairly well in the low energy region, but for AGNs and especially GRBs many questions remain unanswered. Neutrino based information on them would give a set of information complementary to the present knowledge, hopefully yielding new indications on what kind of theories are required to describe these objects. Even the non-detection of point-sources would be an important information.

But the astrophysical search for point sources is not the only task of neutrino telescopes. They also offer the opportunity for particle physics to detect, or set more stringent upper limit on the fluxes of new particles: Many cosmological and GUT theories describe new, so far undetected particles. For the detection of some of them (or their decay products), like neutralinos or magnetic monopoles, neutrino telescopes might be an appropriate device. Environmental studies are also performed at detector sites. Since their results are of vital importance to the experiments, these studies are sometimes considered as compulsory exercises, but in fact they are interesting studies by themselves and are a nice example of interdisciplinary science.

2.4.1 Low energy neutrinos from supernova bursts

Heavy stars which burned up their hydrogen, helium and other light elements (up to iron) start collapsing until the outer layer rebounds from a suddenly solid or expanding inner core. As proven by SN 1987A, supernovae emit a large flux of neutrinos [14]. In general

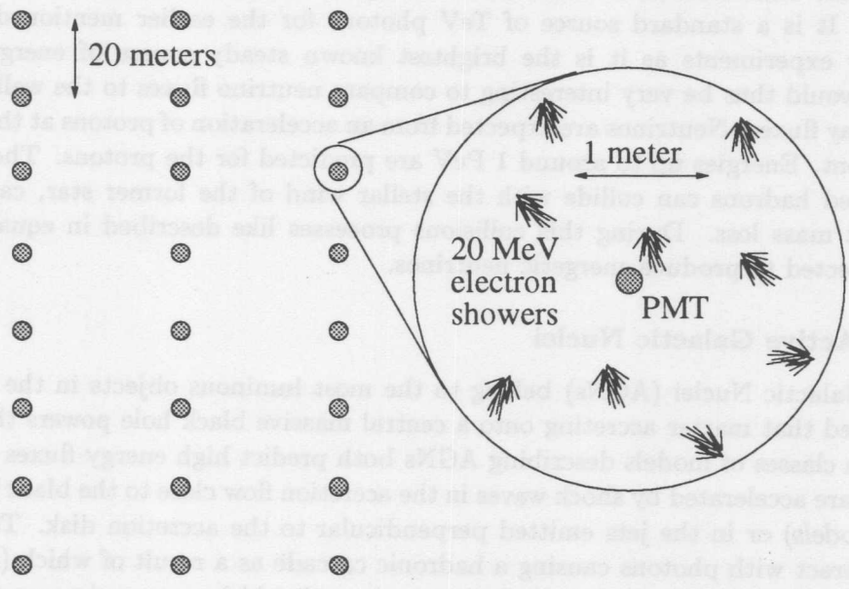


Figure 2.7: Burst-like events can be detected by a rise in the counting rate

these neutrinos have energies far below telescope thresholds. However disregarding the standard telescope operating mode, one can also use the detector as a simple rate counter: A sudden increase in the $\overline{\nu}_e$ flux causes small localized cascades (see figure 2.7) and thereby a sudden increase of the counting rate. During the initial collapse of a supernova, only electron neutrinos are produced, leading to electron showers in the detector. But in addition all three flavors of $\nu\ \overline{\nu}$ pairs are produced via the Z^0 exchange just ~ 1 second later, when there are enough positrons in the hot environment of the burst to annihilate with electrons [64]. But as muon and tau-neutrinos do not have enough energy to produce charged leptons, only ν_es and $\overline{\nu}_e$s ignite showers from this "delayed" burst. Depending on the average size of these showers (determined by the neutrino energy), a certain fraction of the showers will illuminate the closest PMT. These are independent hits, so no source tracking is possible with a single neutrino telescopes. But an overall increase in the counting rate can give a significant signal indicating the existence of a burst-like event, e.g. a supernova. Due to their large noise (especially in K-40 polluted ocean detectors) and a small absorption length, water detectors are less sensitive for such supernova searches than ice detectors. With the good timing resolution of neutrino telescopes, one obtains very precise information on the duration of the neutrino signal leading to limits on ν_e masses: After the supernova 1987A, underground experiments could set mass limits which are of similar quality as those of dedicated tritium decay experiments [1]. Furthermore, once at least three neutrino detectors running, their combined timing information can be used for triangulation.

2.4.2 Supernova Remnants

After the actual supernova burst, a so-called supernova remnant (SNR) remains. Such remnants are well known sources of electro-magnetic radiation. One of the most prominent SNRs is the Crab nebula, the remnant of a supernova reported by Chinese astronomers in 1054. It is a standard source of TeV photons for the earlier mentioned air shower Čerenkov experiments as it is the brightest known steady source of energetic gamma rays. It would thus be very interesting to compare neutrino fluxes to the well established gamma ray fluxes. Neutrinos are expected from an acceleration of protons at the expanding shock front. Energies up to around 1 PeV are predicted for the protons. These, or other accelerated hadrons can collide with the stellar wind of the former star, caused by the pre-burst mass loss. During this collisions processes like described in equation 2.4 are then expected to produce energetic neutrinos.

2.4.3 Active Galactic Nuclei

Active Galactic Nuclei (AGNs) belong to the most luminous objects in the universe. It is assumed that matter accreting onto a central massive black hole powers them. Two of the main classes of models describing AGNs both predict high energy fluxes from AGNs: Protons are accelerated by shock waves in the accretion flow close to the black hole (*generic AGN models*) or in the jets emitted perpendicular to the accretion disk. These protons then interact with photons causing a hadronic cascade as a result of which (among other particles) neutrinos are produced. Such neutrinos should have energies comparable to or higher than the known gamma rays from AGNs, i.e. far above detector thresholds [16].

2.4.4 Gamma Ray Bursts

Yet to be fully explained sudden high fluxes of gamma rays from seemingly point sources are called gamma ray bursts (GRBs). They can last for tens to hundreds of seconds and are isotropically scattered over the whole sky. Since 1996 counterparts in the X-ray, optical and radio spectrum have been reported for some GRBs [45]. Many models predict either "internal" or "external" shock acceleration[4]. Both classes assume some hadronic acceleration mechanisms, and therefore predict the emission of energetic neutrinos, like seen in equation 2.4. Neutrino energies up to 100 TeV are expected [28]. Detecting such high energy neutrinos in coincidence to the gamma ray bursts would make a strong point for hadronic models. So signatures for high-energy neutrinos from GRBs are the coincidence with detected gamma rays and the directional agreement.

If GRBs also lead to low-energy neutrino bursts, these can be detected by an increase in the counting rate (just like supernova bursts can be detected). Additional information obtained from photon detectors then leads to a distinction between supernova and gamma ray bursts.

Dedicated gamma ray observatories have timing accuracies of the order of milliseconds. Neutrino telescopes can easily monitor their counting rates with the same precision. The time of flight from source to earth would be longer for massive neutrinos than for massless particles like photons. Comparing the arrival time distributions from neutrino telescopes and gamma ray satellites could therefore lead to 10^{-4} eV information on neutrino masses for a source of 1000 Mpc distance [29].

2.4.5 Topological Defects

GUT theories tell us that at a very early stage, there was an exact symmetry between the electroweak and the strong force. This is not the case any more. In order to account for this broken symmetry, there must have been a phase transition in the past[47]. During this phase transition, topological defects might have been frozen out[68]. They could be zero-dimensional (monopoles), one-dimensional (strings) two-dimensional (domain walls) or even more bizarre objects. Such topological defects might accelerate hadrons, thereby perhaps accounting for the highest energy cosmic rays.

2.4.6 WIMPS

Cosmologists describing a flat universe postulate the existence of hot and/or cold dark matter. Cold dark matter refers to particles which have little kinetic energy (on the scale of their mass). A subclass of these are the so-called *weakly interacting massive particles* (WIMPS), which only interact via the weak force. The most prominent candidates are neutralinos postulated by super symmetric extensions of the standard model. This is a nice example how closely related the theories describing the biggest (cosmos) and the smallest (elementary particles) are related: They even propose similar new particles. After being created during the big bang, WIMPS started annihilating, but due to their small interaction cross section not all have vanished yet. They are expected to scatter off nuclei,

[4]However many questions remain unanswered, as no convincing model fully describes all aspects of gamma ray bursts so far.

2.4.9 The Unexpected

Every time a new "window" to the sky was opened, i.e. every time new photon frequencies or new particles were used, new phenomena were discovered: Quasars, pulsars and AGNs were discovered with radio waves during the fifties. Once satellites were available, X- and gamma rays became a new tool for astronomy, which lead to the discovery of black hole candidates, accretion discs and the still enigmatic gamma ray bursts [25]. None of these discoveries were expected. Likewise the atmospheric neutrino anomaly was discovered by chance while looking for proton decays. AMANDA-II, which's simulation is analyzed during this work will exceed present neutrino detectors by 1-2 orders of magnitude, a final 1 km^3 detector will exceed them by 3 orders of magnitude in effective area. Having this in mind, it would be very unlikely that nothing new will be discovered. As one might put it: *Astrophysics is a subject where observers generally lead, and theorists follow behind* [56].

3 The AMANDA Detector

3.1 History

The DUMAND collaboration had experienced serious problems regarding reliability, deployment techniques and access (ship) time. In the early 1980s Halzen and Learned suggested using ice rather than water as the natural environment for a neutrino telescope [27]. The premise was that inferior properties of ice with respect to water could be compensated by a greater number of PMTs. Advantages should arise from a solid deployment platform and low background. South Pole is the only place on earth to combine the two requirements for such a project: A constant, several kilometers thick ice shield and an excellent infrastructure with the possibility of year-round access for maintenance. This was the reason to locate AMANDA at the *Amundsen-Scott South Pole Station.*

In the austral summer 1991/92 preliminary site studies and drilling tests were performed. Two years later the first four strings[1] were deployed with 20 OMs each at depths between 810 and 1000 meter. This is nowadays referred to as the AMANDA-A array. Together with the OMs, several other devices (like artificial light sources) were deployed to calibrate the detector and to analyze the (esp. optical) properties of the ice at that depth. It was concluded that the small scattering length (mainly due to residual air bubbles) of $\sim$ 25 cm was insufficient for track reconstruction. However in combination with the surprisingly large absorption length AMANDA-A proved to be an useful tool for the observation of contained ν_e-induced electron showers and for supernova burst searches [54].

Glaciologists estimated that under the higher pressure, deeper ice (below $\sim$ 1200 – 1400 meters) should consist of purer ice crystals with significantly better optical properties. So 1995/96 four new strings were deployed (one in the center and three on a circle of 35 meter radius), but this time the OMs reached depths between 1600 and 1950 meters. It was shown that the ice around this so-called AMANDA-B4 detector had a scattering length of 24 m, sufficiently large to allow reconstruction [71]. During the next campaign in 1996/97 six additional strings were deployed on a second circle (concentric to the first one) of 60 meter radius. In 1997/98 three further (survey) strings were deployed, with OM locations from 1300 meter depth down to 2350 meters, yielding AMANDA-B13.

3.2 Status

3.2.1 Geometry

The present geometry of the AMANDA detector is sketched in figure 3.8. So far the geometry is close to a cylindrical symmetry. This will also remain the case for the next drilling season, when eight new strings are supposed to complete the 100 meter circle around AMANDA-B10, resulting in AMANDA-II. For ICECUBE, the planned 1 km^3 sized extension of the AMANDA detector, several possible geometries are under consideration. Based on Monte Carlo simulations as well as experience with the present detector, a

[1] The term *string* is sometimes confusing: Normally it stands for the whole set of cables from surface to the bottom of one hole. But sometimes it is also used to describe only that region of the cables where the OMs are attached.

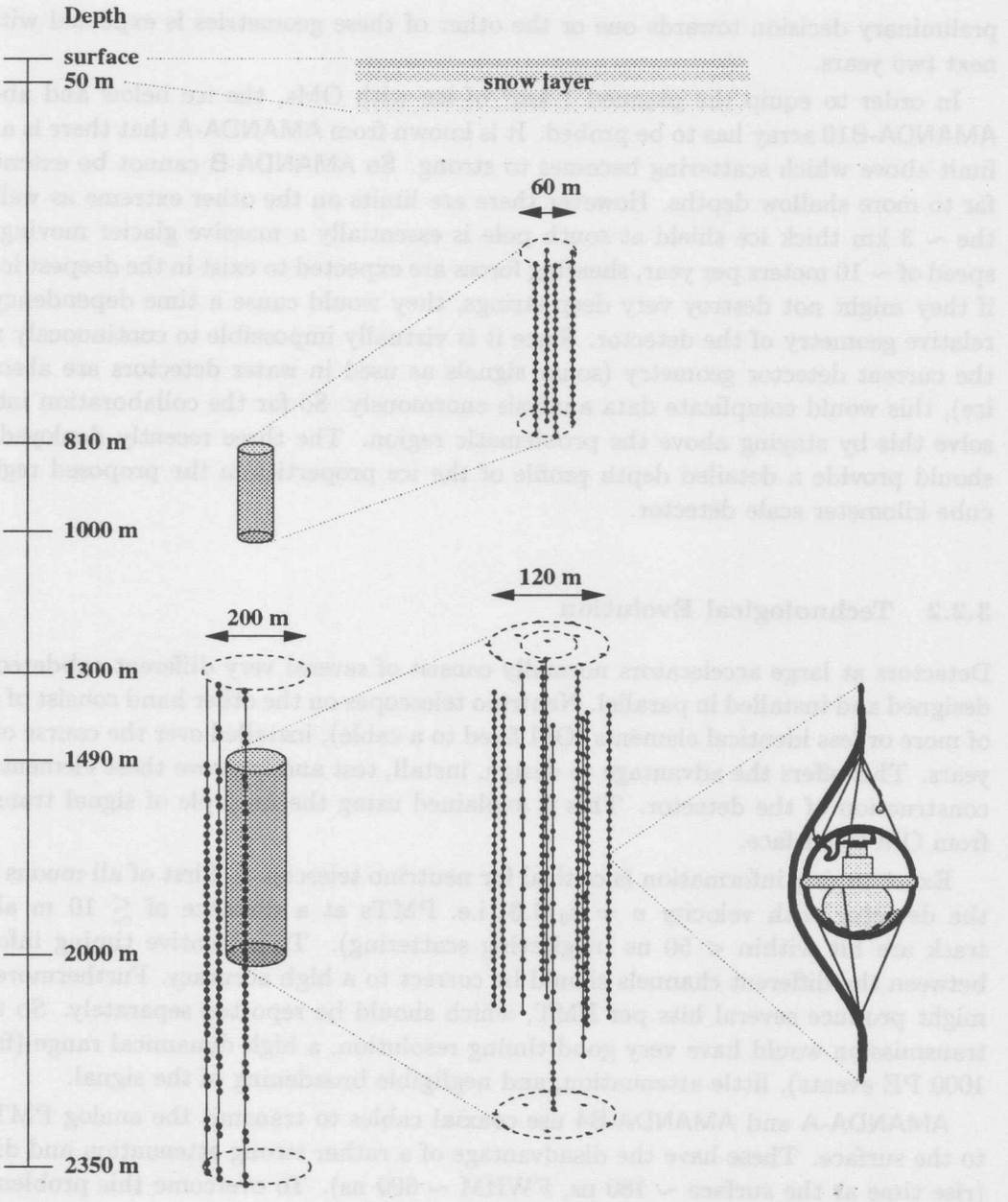


Figure 3.8: The present AMANDA detector after the 1997/98 deployment season. The shallow AMANDA-A detector consists of four strings, forming part of a 30m radius cylinder. AMANDA-B consists of 13 strings: One in the center, three on a cylinder of 35 m radius, six on a cylinder of 60 m radius and three on one with 100 meter radius. Up to 42 optical modules are attached to each string. The optical modules are looking down in order to increase efficiency for up-going muons and decrease efficiency for down-going muons.

preliminary decision towards one or the other of these geometries is expected within the next two years.

In order to equip the planned 1 km^3 of ice with OMs, the ice below and above the AMANDA-B10 array has to be probed. It is known from AMANDA-A that there is a certain limit above which scattering becomes to strong. So AMANDA-B cannot be extended too far to more shallow depths. However there are limits on the other extreme as well: Since the $\sim$ 3 km thick ice shield at south pole is essentially a massive glacier moving with a speed of $\sim$ 10 meters per year, shearing forces are expected to exist in the deepest ice. Even if they might not destroy very deep strings, they would cause a time dependency of the relative geometry of the detector. Since it is virtually impossible to continuously monitor the current detector geometry (sonar signals as used in water detectors are absorbed in ice), this would complicate data analysis enormously. So far the collaboration intends to solve this by staying above the problematic region. The three recently deployed strings should provide a detailed depth profile of the ice properties in the proposed region of a cube kilometer scale detector.

3.2.2 Technological Evolution

Detectors at large accelerators normally consist of several very different subdetectors, all designed and installed in parallel. Neutrino telescopes on the other hand consist of a lattice of more or less identical elements (OM fixed to a cable), installed over the course of several years. This offers the advantage to design, install, test and improve these elements during construction of the detector. This is explained using the example of signal transmission from OM to surface.

Exact timing information is critical for neutrino telescopes: First of all muons traverse the detector with velocity $v = c_0/1.3$, i.e. PMTs at a distance of $\lesssim$ 10 m along the track are hit within $<$ 50 ns (neglecting scattering). Thus relative timing information between the different channels should be correct to a high accuracy. Furthermore, muons might produce several hits per PMT, which should be reported separately. So the ideal transmission would have very good timing resolution, a high dynamical range (from 1 to 1000 PE events), little attenuation, and negligible broadening of the signal.

AMANDA-A and AMANDA-B4 use coaxial cables to transmit the analog PMT signals to the surface. These have the disadvantage of a rather strong attenuation and dispersion (rise time at the surface $\sim$ 180 ns, FWHM $\sim$ 600 ns). To overcome this problem the six strings deployed 1995/96 were mainly equipped with twisted pair cables. Rise time and attenuation improved by a factor three (rise time at the surface $\sim$ 60 ns, FWHM $\sim$ 210 ns). In this case the main challenge concerns the appearance of cross talk signals: When a non-coax cable transmits a signal over a long distance and a second cable is running parallel, induction will cause a pulse in this second cable. Such induced pulses mimic additional roughly coincident PMT signals. As a third attempt fiber optics are tested at present [44]: Inside the OM the PMT signal is used as input for a LED transmitter. The LED light output is then fed into an optical fiber which is connected to a photodiode receiver on the surface. Time resolutions are improved considerably (rise time at the surface $\sim$ 4 ns, FWHM $\sim$ 4 ns [16]), see figure 3.9.

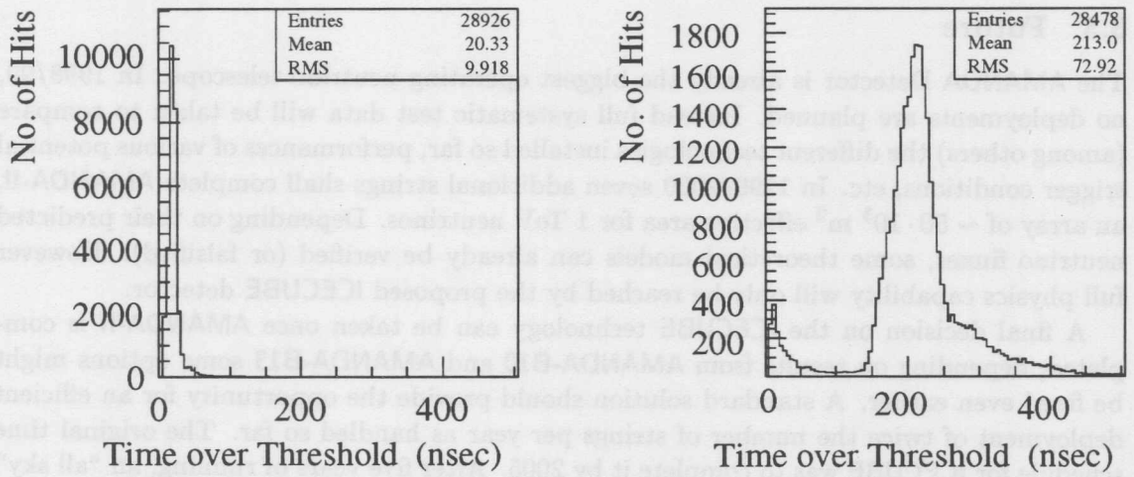


Figure 3.9: The time over threshold (TOT) distribution is shown for 6 fiber optic OM (left) and for 6 twisted pair OMs (right). 50000 events have been used. The mean TOT is 20 ns and 213 ns respectively. The TOT is a measure of the minimum time to separate single photoelectrons.

LEDs are the first non-passive electronic devices inside OMs. Contrary to the original AMANDA approach to keep as much electronics as possible accessible inside buildings at the surface, tests of digital OMs (DOMs)[21] are also being performed. These digitize the pulse shape of a signal inside the OM with a resolution of 1 ns and then send this information through the electrical cable. This would be the most flexible solution and would overcome the whole problem of attenuation in the $\sim$ 2 km long cables. As a second advantage it is not necessary to operate the PMTs at the present large gain of 10^9 to get the signal to the surface. This should reduce potential aging effects, e.g. on the last dynode inside the PMT. However reliability has not yet been proven satisfactorily. A large detector with many DOMs would need a hierarchical structure introducing higher risks of failure. This is especially problematic for an ice detector, where recovery and/or repair of deployed OMs (or other devices) is not possible. Furthermore costs are much higher than for the other mentioned solutions. But nevertheless, like all the other technologies mentioned, a few DOMs are already used in order to be tested under experiment conditions. Decisions on technologies used in the future are mainly taken on basis of these "in situ"-performances rather than purely on laboratory test or theoretical predictions.

Likewise other parts of the detector are likely to change: For example the deployment of several test-OMs with wavelength shifters coating the UV-opaque pressure spheres to yield higher light input for the PMTs is planned for the next drilling season. Other research areas include improvements of the OMs' penetrators and larger OMs. Obviously like in all high-energy experiments electronics and on-line software is continually upgraded as well. In 1997/98 for example a whole new DAQ was installed. It is capable of supporting very sophisticated trigger conditions and designed to cope with long-term AMANDA extension plans.

3.3 Future

The AMANDA Detector is already the biggest operating neutrino telescope. In 1998/99, no deployments are planned. Instead full systematic test data will be taken to compare (among others) the different technologies installed so far, performances of various potential trigger conditions, etc. In 1999/2000 seven additional strings shall complete AMANDA-II, an array of $\sim 50 \cdot 10^3$ m^2 effective area for 1 TeV neutrinos. Depending on their predicted neutrino fluxes, some theoretical models can already be verified (or falsified). However full physics capability will only be reached by the proposed ICECUBE detector.

A final decision on the ICECUBE technology can be taken once AMANDA-II is completed; depending on results from AMANDA-B10 and AMANDA-B13 some options might be fixed even earlier. A standard solution should provide the opportunity for an efficient deployment of twice the number of strings per year as handled so far. The original time schedule for ICECUBE was to complete it by 2005. After five years of running, an "all sky" map[2] of possibly discovered neutrino point sources is hoped to be sufficient accurate to start looking for time dependent effects and energy spectra. As it is certain by now that south pole station is being heavily upgraded during the next five years, transport capacities might not be sufficient to meet AMANDA's desires. So the completion of ICECUBE could be delayed slightly.

[2] As explained in section 2.3, due to background from atmospheric muons, neutrino telescopes can only look down through the earth. Such an "all sky" map for AMANDA is therefore of the northern hemisphere only.

4 Artificial Neural Networks

4.1 The Need for Filtering

In modern high-energy physics experiments it is a common challenge to uncover signals, hidden by an overwhelming background. This process is called *filtering*, the hardware or software tools used are *filters*. The main task for designing filters is to find information gathered in the detector, which allows some kind of distinction between signal and background. Mathematically spoken one aims for a manifold in the available parameter space, which separates a region of high signal to noise ratio from the rest. When deciding on filter specifications, one has to determine an optimal balance between large signal passing rate and high background rejection.

Usually filtering is applied to the data at several levels between data taking and final result. These may include the following:

1. A trigger decides on the readout and/or storage of the measured data. It is an important element of the data acquisition system (DAQ). Depending on the complexity of the experiment and the capacities of the data handling system, such a trigger can consist of several consecutive levels. For AMANDA-B there are actually several trigger conditions used in parallel: The standard trigger condition requires at least eight channels hit within a time window of $2\mu s$. Alternative "external" triggers are provided by events in nearby detectors like AMANDA-A[1], SPASE[2] or GASP[3].

2. Before starting an individual analysis it may be useful to filter the acquired data again, according to the properties relevant for that specific analysis. This reduces storage requirements and CPU time for the analysis. For example, in AMANDA an analysis of individual PMTs might be performed. Then only such events are used, where that particular PMT was hit.

3. Finally one might perform a quality analysis after the main analysis (for AMANDA: reconstruction) is done. This is often not considered as a filter any more but practically has the same characteristic: With such an analysis, signal shall be accumulated relative to background. As was explained in chapter 2, one very important aspect of the AMANDA analysis is the rejection of mis-reconstructed down-going muons. Such a quality analysis is the topic of this work.

Sometimes peculiarities of an experiment require special filters. AMANDA for example takes data during the austral winter, when there is no possibility to ship data to the collaboration laboratories for analysis. There exists the opportunity to use a small bandwidth satellite connection however. A filter determines the small fraction of data transmitted

[1] Coincidences between AMANDA-A and AMANDA-B are a very important calibration source, e.g. for the reconstruction: The direction of muons is known to be almost vertical, see figure 3.8.

[2] The South Pole Air Shower Experiment is located on the surface close to the AMANDA array. Since SPASE has a very good angular resolution, coincidence events between the two experiments can be used to calibrate the AMANDA array. This combination of air shower experiment and neutrino detector is not possible for ocean detectors.

[3] The Gamma Astronomy South Pole experiment was running until the end of Dec 1997.

via this satellite link [40]. The rest of the data is then shipped from the pole once the station is accessible again.

4.2 Strategies to Enrich Signal

Examples for the different separation techniques introduced are displayed in figure 4.10. The aim is to separate the darker ("signal") from the brighter ("background") sample, both shown in (A). One has only information about two observable called 1 and 2.

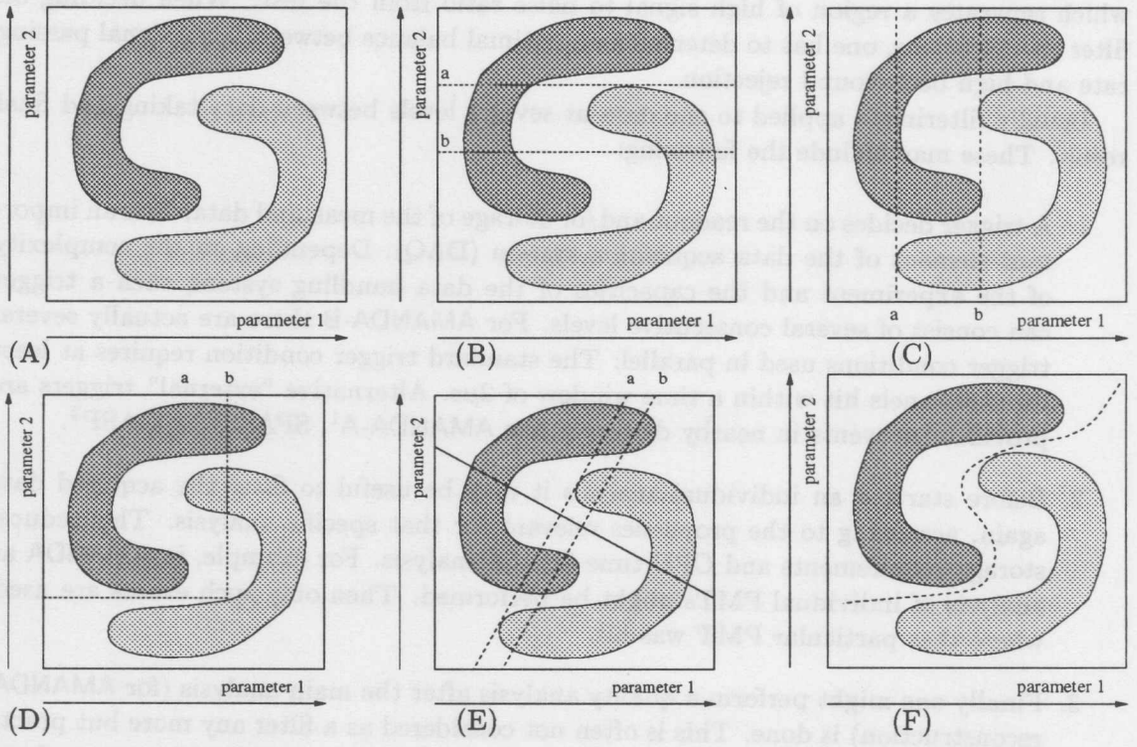


Figure 4.10: Example for various techniques to separate signal (dark) from background (bright). (A): Two sample classes that shall be separated; (B): Possible cuts on parameter 2; (C): Possible cuts on parameter 1; (D): A combination of individual cuts; (E): Linear combination of both parameters (solid line) and cuts on it (dashed lines) (F): Possible separation contour from an artificial neural net.
In general, both classes are overlapping distributions rather than compact regions in parameter space. A perfect separation like in (F) is then impossible. Furthermore parameter spaces have normally more than two dimensions. This figure thus only outlines general features of separation techniques.

The easiest way to tackle any separation is to look for a simple cut in one variable like the up/down cut on the muon's direction for the AMANDA experiment. In figure 4.10 (B) possible cuts on parameter 2 are shown, in (C) similar cuts on the other parameter are indicated. Generally not only the best cut observable is unknown, but also it's optimal cut value. Figure 4.10 (B) and (C) display two cases: The cuts named (a) leave a background-

free signal, while those named (b) try to balance background rejection and signal passing rate. They might be appropriate in cases where a higher signal passing rate than with cut (a) is required. One notices that for case (a) the better result (higher signal passing rate) is reached with a cut on parameter 2, while for case (b) a cut on parameter 1 seems more reasonable.

Often enough one parameter is not even sufficient and a set of parameters has to be found. Such a multi-parameter cut is shown in figure 4.10 (D): If an event has a large value with respect to parameter 2, and a low value with respect to 1, it is considered signal, otherwise background. This considerably improves the separation compared to one cut on a single observable. Since the potential cut parameters are often not independent from each other neither are the cut values. So the quality of a parameter set for filtering depends on its set of cut values which have to be optimized simultaneously. Keeping in mind that such cuts are restricted to produce hyper-rectangular parallelepipeds (multi-dimensional rectangles) in parameter space, i.e. that normally an ideal contour for separation cannot be found anyway, other separation techniques are often being employed.

The so called discriminant analysis, as described by Fischer [48], assumes a gaussian distribution with respect to each parameter. (So the samples in figure 4.10 is not an ideal example to explain this method.) It is nevertheless explained with the aid of figure 4.10 (E). The analysis consists of an analytic calculation of a so-called *orthogonal vector*, which basically points from the center of the signal distribution to the center of the background distribution in parameter space. The scalar product of this orthogonal vector with the position vector (in parameter space) of each event is then a new parameter, defined for every single event. It is a linear combination of all parameters used (the solid line in the figure). A cut on this new parameter (dashed lines in the figure are potential cuts) defines a hyperplane separating both distributions. So the separation problem in the multi-dimensional parameter space is reduced to a one-dimensional problem[4]. For the example of figure 4.10, the discriminant analysis has a similar quality as a combined cut on both parameters, but is considerably better than a single cut on one parameter. This method makes assumptions on the shape of the distributions (gaussian). Only such an assumption offers the opportunity for an analytic solution to the search for the appropriate linear combination of potential cut parameters in a quality filter. The before mentioned combined cuts on sets of individual parameters (figure 4.10 (D)) cannot be found by similar (fast) analytical methods - even with assumptions on the distributions.

The main disadvantage of the discriminant analysis and its extensions is the restriction to *linear* combinations of the original parameters. Lifting this restriction could lead to an optimized separation contour for the individual problem. Such a contour is sketched in figure 4.10 (E). This approach is the domain of neural networks. They are nowadays widely used, e.g. as standard tools for filtering problems, in high energy experiments. Details of such an analysis are explained in the next section.

[4]There are similar, more general tools, which can reduce the problem to any number of specified dimensions[23]. It is then hoped, that a human being might recognize some good separation contour within these reduced dimensions. Since an "eye-scan" in more than two dimensions is hardly applicable, the problems are normally reduced to one or two dimensions.

4.3 Artificial Neural Networks

The term artificial neural networks describes where they come from: The task to replicate the main feature of the human brain, i.e. the network of neurons. In the early stages of this approach, such a network design was very different from proven computing solutions: Standard computers were sequential machines performing one calculation after the other. During assembly, the individual elements were connected and neither the elements nor the connections could be changed afterwards. Computers could perform simple mathematical operations much faster than humans, but failed to perform complex tasks like pattern recognition. This failure could not be caused by using "dead" electronics, since it had been realized that the brain also basically works via electrical signals. Thus the idea was to combine the speed of signal transmission in computers with the complex "design" of the brain, into an artificial neural network. This then should unite the advantages of both systems.

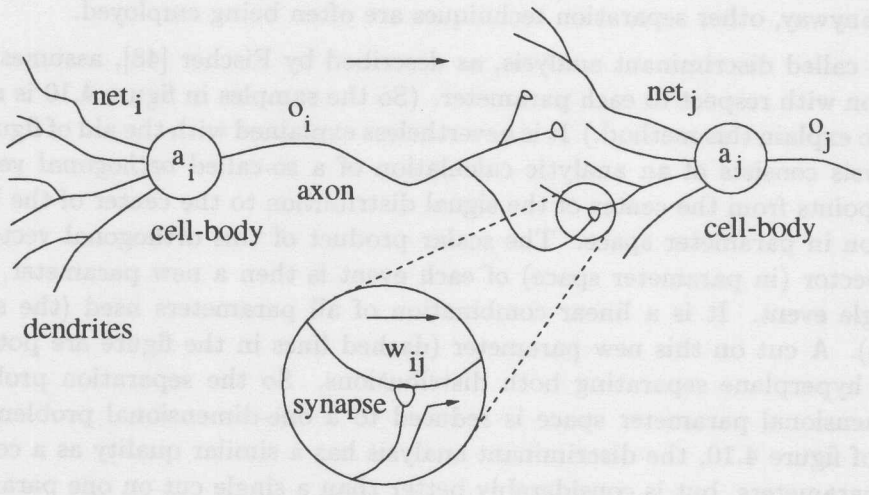


Figure 4.11: Sketch of a small fraction of a biological neural network (The information flow is from left to right). The dendrites deliver the network input net_i to their cell. This in turn reaches a new state of activation a_i and generates some output o_i in its axon. Via synapses of individual strength w_{ij}, branches of this axon are connected with the dendrites of neighboring cells.

The main biological knowledge was the following: The brain consists of individual, albeit connected nerve cells. A cell receives its stimulus (input net_i) from its *dendrites*. Depending on the stimulation strength, the cell turns into a certain stage of activity (a_i) and changes the electrical potential of its axon (o_i) accordingly. This axon is branched out and the various branches are connected via synapses to dendrites of neighboring cells. The synapse determines the strength of the influence (w_{ij}) an axon has on the attached dendrite. A model of such a biological network is shown in figure 4.11.

Compared to computers, one of the most remarkable features of the human brain is the ability to learn. More sophisticated software can provide improved results from the same computers. But the hardware always stays the same - unlike the brain which can adapt itself. As a rough approximation learning is performed by adapting the synapses. If they

disappear, are newly formed or change the strength of axon-dendrite couplings, then the communication between the cells is altered, i.e. the brain's architecture is reconfigured.

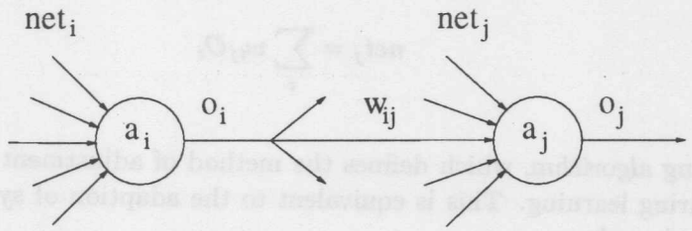


Figure 4.12: Idealized model of a neural network as found in biology or in artificial replicants: dendrites, synapses and axon branches are combined to weighted links. Apart from this the main ingredients are the same as in the more realistic biological model from figure 4.11.

In an effort to translate the various biological elements and their interplay as directly as possible, artificial neural networks were designed as simple copies of biological models (see figure 4.12) and consist of the following:

1. The *network architecture* which specifies the topology of layers of neurons and links between them. This is the equivalent to its biological counterpart.

2. The *units* (neurons, elements, cells). In an artificial net a unit has three attributes:

   - The *threshold* Θ_j applied to the input,
   - The *activation* a_j ($\equiv$ exciting voltage), which is calculated by an *activation function* f_{act} from the input net_j (for its definition see below), the present activation and the threshold Θ_j of this unit j:

   $$a_j(t + \delta t) = f_{act}(\Theta_j, a_j(t), net_j(t)) \ , \tag{4.10}$$

   where δt is equivalent to one update iteration for the units. Often, the calculation of an activation is independent from the previous state of activation. Then formula 4.10 reduces to

   $$a_j = f_{act}(\Theta_j, net_j). \tag{4.11}$$

   - The *output* $\mathcal{O}_j$ ($\equiv$ signal in axon), which is calculated by an *output function* f_{out} from the activation a_j of the unit j

   $$\mathcal{O}_i = f_{out}(a_i) \ . \tag{4.12}$$

   The output and functions are normalized as to produce activations with values between 0 and 1.

3. The *weights* of the links: A link connecting units i and j is equivalent to the axon branches of a biological cell. The weight w_{ij} of that link is equivalent to the synapses, i.e. it determines the importance of a unit (anti-)stimulus from one cell to a connected cell.

4. The *propagation function*, which computes the input of one unit j from the outputs $\mathcal{O}_i$ of its predecessor units and the weights w_{ij} of the links between them:

$$net_j = \sum_i w_{ij}\mathcal{O}_i \tag{4.13}$$

5. The *learning algorithm*, which defines the method of adjustment of thresholds and weights during learning. This is equivalent to the adaption of synapses in the biological learning phase.

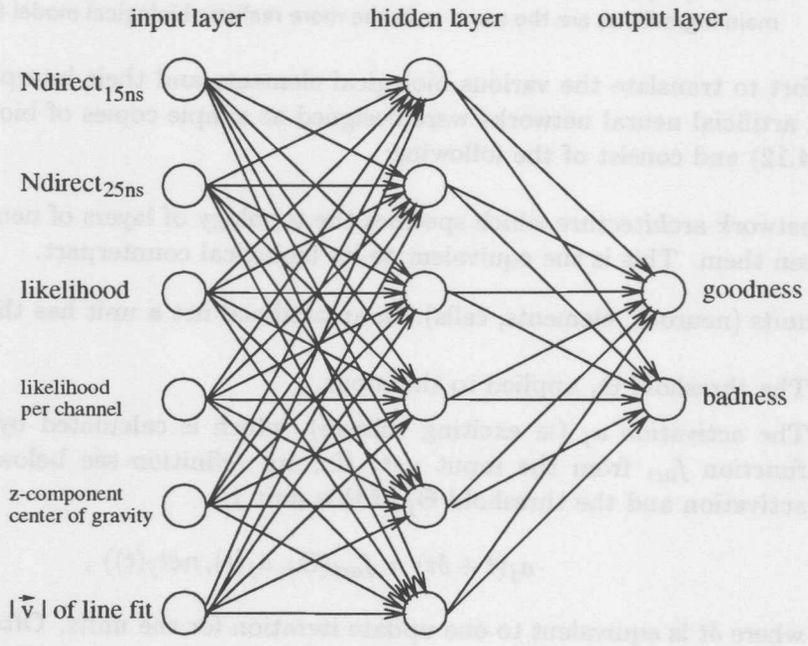


Figure 4.13: Typical feedforward net used for this work, with six input units, one hidden layers of six units and two output units. For a detailed description of the input parameters see section 5.3.

In short, an artificial neural network works in the following way: First an architecture is specified. During the learning period the neural network is given some data sets of known "demand" output. It then adjusts its weights, thresholds and possibly links according to its learning algorithm in order to reproduce the demand output as good as possible. Then, after learning these weights, thresholds and links are fixed and new data sets can be fed into the net. Each data set results in a certain pattern of activation and output of the individual units. This output pattern is the result of the analysis.

Some of the units are input units. For a certain application each of these receives its own specialized kind of information. For example, if one would use a neural network as trigger in the data acquisition, the first unit might always get the ADC information of the first channel, the second unit the ADC information of the second channel and so on.

Of similar importance are the output units[5]. After a set of information is processed by the net, the output of these units is the "answer" of the net. As a standard example, one might have trained a neural network to separate two classes of events. With one output unit, an output close to 0 of this unit indicates a likely member of one class, while an output close to 1 indicates a likely member of the other class. The units in between the input and the output units are called hidden units.

Before learning starts, the starting values of the activations of the various units, their thresholds and the weights of the links are set to initial values, most easily random. Then a set of information is fed into the input units of the net. From this the new activations of the input units (and in turn also their output) is determined. Their output, multiplied by the weights, is the input of the next layer of units. Now their activation is recalculated and so on. In case of back-coupling links[6] the ordering of these updates is crucial and several updates per unit per set of input information might be necessary. In the case of *feed-forward networks*, where no back-coupling is possible, update ordering is performed in layers and multiple updates per unit are not required. After every unit was finally updated, the whole network has reached a new state and especially the output units have new activations and outputs.

With a fully trained net, this output of the last layer is the result of the analysis. During the learning period however, this output is compared to a desired or *demand output*. Such learning, where the net is not only fed with input data, but also with the expected output is referred to as *supervised learning*. The difference between demand output and calculated output is considered as the error for this event (set of information). The error is an indicator for the quality of the analysis and the basis of the learning: According to the learning algorithm the weights of the links and the thresholds of the units are updated[7]. Likewise a large sample of representative information sets passes the net one at a time. Most learning algorithms are based on a descending gradient method. It is hoped that the global error (in parameter space, i.e. on a large sample of data sets) can be minimized by adjusting free parameters to get a smaller error on the individual (local) event. If the network specifications as well as the choice of the used input parameters fit the problem, the net adapts its free parameters automatically to this global error minimum. Once the error cannot be reduced any further (or is small enough for the particular purpose) learning is stopped. The adjusted parameters are now fixed to their current values. The analysis of data with unknown demand output is then nothing else but the calculation of a (more or less) complicated analytic function of the input parameters.

[5]There is a whole class of artificial neural networks called *cohonen nets*, without output units. But they were not used during this work and their theory is very different from those networks with output units, so they will not be treated here. An outline of their theory is given in most books on neural nets, like [75].

[6]Just like in networks of regular conductors, it is possible to construct neural networks (biological or artificial) with or without back-coupling. If the information flow is restricted to one direction, an artificial neural net is described as *feed-forward*. These are easier to implement, train and use. However networks with back-coupling potentials are more general in there output. One might even consider feed-forward nets as a subset of back-coupling nets with zero back-coupling, but this is normally not done. During this work only feed-forward nets were used.

[7]In some cases even new links might be formed or others abolished, but such dynamic topologies were not used during this work so again this is an effect which will be ignored from now on.

It is well known that the calculation of such a specific function can be done on a multi-purpose computer with the appropriate software or with some hardware optimized on the specific function. In general, hardware solutions result in single-purpose chips/cards which are more expensive, but faster than software solutions. For neural networks, there exist all kind of hard-, software or hybrid solutions. For example learning might be performed on a simulation platform and once the neural network's properties are specified, a hardware element is produced accordingly.

In general, the activation and/or output functions are restricted to yield values in the range [0,1]. Thereby they also set a natural scale for the input data. Depending on the specific net used, data sets that are not restricted to this range will prolong the learning period, yield worse results or might even not be accepted as input data. Therefore all input data should be (has to be) transformed to this range before it is fed into the network. Depending on the original distributions, this transformation can be performed by simple operations like a division by some maximum value or by more sophisticated processes. The major concern is not to loose relevant information during this transformation. The transformations used in this work are described at the beginning of chapters 6 and 7.

4.4 Application to AMANDA

So far, the standard analysis of AMANDA data has been performed without the aid of neural networks. However, preliminary studies on the potentialities of such networks at the trigger level or as a pre-reconstruction filter have been performed [57]. It was found that information on the OM-number and the leading edge of its signal were the most relevant information. During this work a complementary application, as a post-reconstruction filter, was investigated. The aim was to separate up-going, neutrino-induced muons from down-going atmospheric neutrinos.

During this work only feed-forward networks with one hidden layer were used. All units of one layer were connected with all units of adjacent layers, but no back-coupling links or links connecting units from next-to-neighboring were allowed. Two anti-symmetric output units "goodness" and "badness" were used: "goodness" should produce an output of 1 for up-going muon events and 0 otherwise, while "badness" should be 1 for mis-reconstructed atmospheric muons and 0 for signal events. There were several different networks tested, so that the specific number of units in each layer will be described when the individual analysis are presented in the chapters 6 and 7.

The activation function for a unit j used in this work is called *logistic activation function*[8]. It calculates the activation a_j of a unit j from its input net_j and its threshold Θ_j:

$$a_j = f_{act}(net_j, \Theta_j) = \frac{1}{1 + e^{-(\sum_i w_{ij}\mathcal{O}_i - \Theta_j)}} , \tag{4.14}$$

with $\mathcal{O}_i$ being the outputs of the unit i from the previous layer, w_{ij} being the weight between this unit j and unit i. i is running over all units in the previous layer. As one can see, it is normalized to values between 0 and 1.

[8]This "logistic activation function" is closely related to the "Fermi function", well known for example from thermodynamics: $f_{act}(\epsilon) \equiv f_{Fermi}(-\epsilon) \equiv 1 - f_{Fermi}(\epsilon)$

The output function used is the identity function:

$$\mathcal{O}_j = f_{out}(a_j) \equiv a_j \tag{4.15}$$

thus during this work the terms output and activation of a unit are identical.

Supervised learning is often based on the *Hebbian rule*, which states that the link between two units is strengthened if both units are active at the same time. In order to train the networks, the *backpropagation weight update rule*, a special form of the Hebbian rule, was used. The alteration Δw_{ij} of a weight between unit i and unit j can be calculated by

$$\Delta_{w_{ij}} = \eta \sum_p \mathcal{O}_{pi}\delta_{pj} \tag{4.16}$$

where p runs over all patterns (input sets / events), η is a constant factor called *learning rate*. If η is too big, the global error minimum might be missed during learning. If it is too small, learning might stop in a local minimum. So a common solution is to use several successive learning periods (each with up to a few hundred loops over the whole training sample) with decreasing values for η. In this work five steps with values from 0.2 to 0.001 were used. In the case of backpropagation with the logistic activation function and the identity as output function (as during this work), the error signal δ_{pj} is given by

$$\delta_{pj} = \begin{cases} \mathcal{O}_{pj}(1-\mathcal{O}_{pj}) \cdot (t_{pj} - \mathcal{O}_{pj}) & \text{if unit } j \text{ is an output unit.} \\ \mathcal{O}_{pj}(1-\mathcal{O}_{pj}) \cdot \sum_k (\delta_{pk} w_{jk}) & \text{if } j \text{ is a hidden unit.} \end{cases} \tag{4.17}$$

$\mathcal{O}_{pj}$ is the actual output of unit j for input pattern p and t_{pj} is the theoretical (demand) output value of unit j for input pattern p [75].

The simulation itself was performed on the free system SNNS 4.1 [74]. This package implements a large variety of neural network techniques and provides a X11 interface for visualization and a batch mode interface. Further details on specific nets used can be found in chapters 6 and 7, when results are presented.

5 Simulation and Reconstruction: Description of Methods and Input Parameters for Neural Networks

5.1 Monte Carlo

Like in any kind of experiment, the quality of a search for neutrino point sources in AMANDA is determined by two kind of errors: The statistical error is given by the number of events, i.e. basically by the geometry (size) and trigger efficiency of the detector and by the neutrino flux from point sources. It can be minimized by building big and efficient telescopes. The systematic error is governed by the quality of the reconstruction algorithms and the filtering tools. In order to optimize these, Monte Carlo simulations are needed: By simulating a muon with known direction traversing the detector and the detector response to this passage, one can simulate experimentally gathered data, which can be reconstructed and analyzed. Via a comparison between the generated muon direction and the reconstruction result, the quality of the reconstruction can be tested and in turn be improved. This is just one among many reasons why Monte Carlo simulations are important. Others include optimization of detector geometry, determination of ice-parameters, optimization of trigger conditions or estimations of trigger rates (thus determining the required data storage capacities), see for example [35].

Events in the detectors AMANDA-B4 and AMANDA-II studied in this work were simulated by the *AMASIM* program developed by Stefan Hundertmark [38]. A full discussion of the simulation tools including air shower generation and muon propagation can be found in [37, 39].

The generation of background events starts from isotropic cosmic ray protons with energies up to 1000 TeV. The atmospheric shower program *basiev* [18, 60] simulates the interaction of these protons with the earth's atmosphere. The resulting air shower is simulated to obtain the correct muon multiplicities and muon energy distributions. The propagation of the produced muons is then simulated with an averaged energy loss until they intercept the so-called *generation plane*. This is a large plane far outside the detector (see figure 5.14). The generation plane is at a distance of $\sim$ 800 m to the detector. Its orientation is always perpendicular to the mean shower direction.

For the case of neutrino-induced muons, single muons can be simulated with arbitrary angular and energy distributions. Tracks are generated on a *generation plane* or inside a *generation volume* (for low energy muons). For the AMANDA-II simulation used in this work, muons were generated isotropically on a generation surface at $\sim$ 800 m distance from the detector. In the case of the AMANDA-B4 simulation, muons were generated inside large volumes. The boundaries were so far away from the detector, that muons generated at the boundaries in general stopped before reaching the detector.

Starting from the generation plane, it's the same procedure for both kind of muons[1]: The muon propagation program *MUDEDX* simulates the energy losses due to continuous (ionisation) and stochastic (δ-electrons, bremsstrahlung, e^+e^-pair production, muon-nucleus interactions and $\mu^+\mu^-$pair production) processes [41]. Since the continuous emission of Čerenkov radiation by the muon itself causes only a negligible fraction of the energy

[1]Single muons generated inside a volume are treated as if they already had intercepted the generation plane

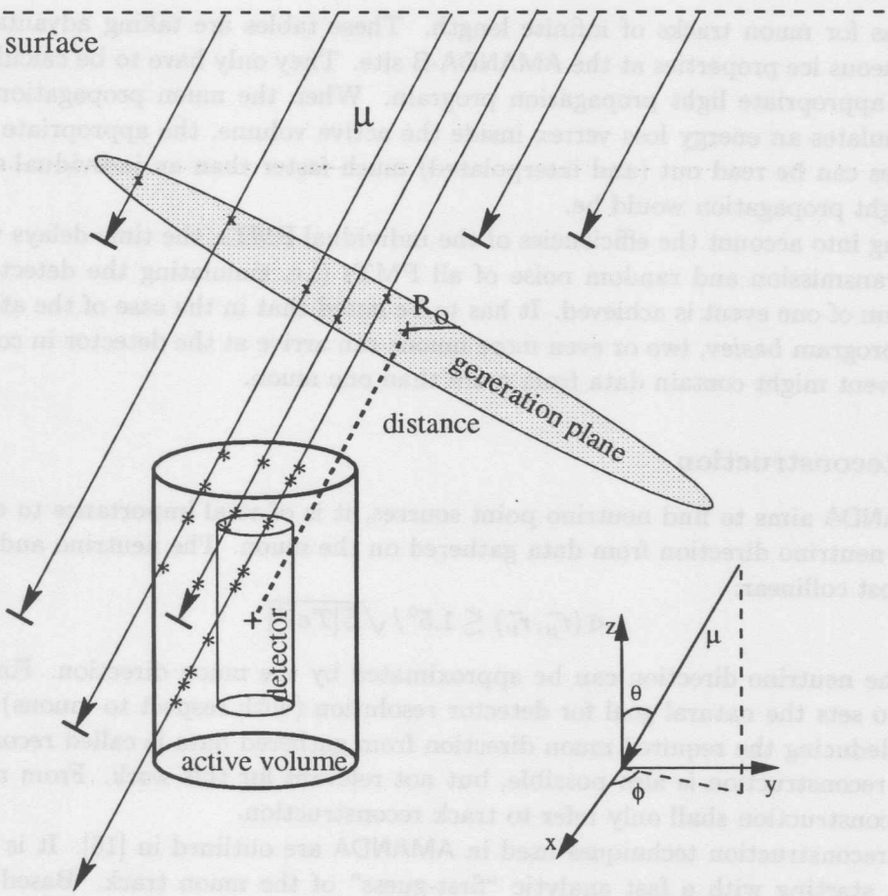


Figure 5.14: Sketch of the geometries relevant for the muon simulation. Only muons intercepting the generation plane (marked with an x) are passed on to the muon propagation program. Photons from secondary processes (marked with a $*$) are only generated inside the active volume around the detector.

loss, it is not simulated here. During this phase of the simulation the other mentioned processes are simulated in detail. One therefore obtains a correct simulation of the energy loss of the muon (rather than an averaged one). However, the propagation of photons only becomes interesting close enough to the actual detector, since photons normally travel less than 200 m in ice. Therefore an *active volume* is defined, consisting of the equipped volume plus a thick "layer" of 200 m extension in every direction.

Inside the active volume the same processes (plus Čerenkov radiation) are simulated. But this time the propagation of the photons, including time delays (due to scattering) and attenuation (due to absorption), is relevant. This propagation is not simulated for each individual photon, but performed with the aid of so called *scatter tables*. One of these large multi dimensional tables contains information on the mean number of photons as a function of the distance and relative orientation between a PMT and an energy loss vertex or a muon track. The other scatter table gives the corresponding time delay distribution relative to unscattered Čerenkov photons. There are tables for stopping muons

as well as for muon tracks of infinite length. These tables are taking advantage of the homogeneous ice properties at the AMANDA-B site. They only have to be calculated once with an appropriate light propagation program. When the muon propagation program then simulates an energy loss vertex inside the active volume, the appropriate entries of the tables can be read out (and interpolated) much faster than an individual simulation of the light propagation would be.

Taking into account the efficiencies of the individual PMTs, the time delays within the signal transmission and random noise of all PMTs (i.e. simulating the detector), a full simulation of one event is achieved. It has to be noted that in the case of the atmospheric shower program *basiev*, two or even more muons can arrive at the detector in coincidence, so one event might contain data from more than one muon.

5.2 Reconstruction

As AMANDA aims to find neutrino point sources, it is of vital importance to deduce the original neutrino direction from data gathered on the muon. The neutrino and the muon are almost collinear:

$$\sphericalangle(\vec{r_\mu}, \vec{r_\nu}) \lesssim 1.5^\circ / \sqrt{E[TeV]} \; . \tag{5.18}$$

Thus, the neutrino direction can be approximated by the muon direction. Formula 5.18 then also sets the natural goal for detector resolution (with respect to muons). The process of deducing the required muon direction from gathered data is called *reconstruction*. Energy reconstruction is also possible, but not relevant for this work. From now on the term reconstruction shall only refer to track reconstruction.

The reconstruction techniques used in AMANDA are outlined in [16]. It is a two-step process, starting with a fast analytic "first-guess" of the muon track. Based on this, a large fraction of background atmospheric muons can be rejected. Then, with the reduced remaining events, a likelihood reconstruction is performed. It uses the first-guess fit as starting value for its calculation.

Several algorithms can be used for the first-guess. At present a *line-fit* is preferred. It is described in [62] and ignores scattering of the photons or amplitude information, but produces a fit only on basis of the hit times. It assumes light traveling with a velocity $\vec{v}$ (which is not necessarily the speed of light within the ice) through a 1-dimensional projection of the detector: The i-th PMT is located at $\vec{r_i}$ and reached at t_i, so it is assumed that at $t = 0$ the light was at $\vec{r}$:

$$\vec{r} = \vec{r_i} - \vec{v} t_i \; . \tag{5.19}$$

The error is then minimized with respect to the components of $\vec{r}$ and $\vec{v}$. The error χ^2 is defined by:

$$\chi^2 = \sum_{i=1}^{N_{ch}} (\vec{r_i} - \vec{r} - \vec{v} t_i)^2 \; , \tag{5.20}$$

where N_{ch} is the number of hit channels (PMTs). Differentiation leads to

$$\vec{r} = < \vec{r_i} > - \vec{v} < t_i > \tag{5.21}$$

and (inserting equation 5.21 into the corresponding equation for the velocity):

$$\vec{v} = \frac{<\vec{r_i}t_i> - <\vec{r_i}><t_i>}{<t_i^2> - <t_i>^2} , \tag{5.22}$$

where $< x_i >$ indicates the mean of parameter x with respect to all hits i. The velocity $|\vec{v}|$ of the line-fit turns out to be an important cut parameter, see section 5.3. With a cut on the zenith angle after this first guess reconstruction, a high fraction of atmospheric muons can be rejected without loosing much signal.

For the likelihood reconstruction, a more sophisticated approach is needed. As stated in section 2.2.1, energetic particles with velocities $v > c/n$ (n being the refractive index of the medium) emit Čerenkov radiation under an angle θ, where $\cos\theta = 1/(\beta n)$. For energies of interest in AMANDA the Čerenkov angle is independent of energy. But the photons can be scattered between emission from the muon and detection by a PMT. This scattering enlarges the time of flight for the photon. The resulting delay, together with the time jitter of the PMT, is described by a term called *time-residual* t_{res}:

$$t_{res} = t_{hit} - t_{expected} , \tag{5.23}$$

where t_{hit} is the measured time when the PMT was hit and $t_{expected}$ is the time at which an unscattered photon would have hit a jitter-free PMT. So $t_{res} > 0$ for scattered photons, but $t_{res} \approx 0$ for unscattered ones. Having reconstructed a track, the $t_{expected}$ are simply calculated and compared to the measured t_{hit}. In this case negative values for t_{res} do occur – not indicating a violation of causality, but a noise hit, a hit due to a second muon, or an error in the reconstruction. The primary challenge of the reconstruction is to develop a probability function $P(t_{res,i})$ describing the probability of certain time-residuals for a given track and set of PMTs hit. For a monochromatic, point-like light source, an exact analytic probability function can be found [51]. However this function cannot simply be used for radiating muons: It is not even defined for negative time residuals. So, taking into account the different source, the inevitable reconstruction error and the PMT jitter, one has to adapt this analytic function non-analytically in the region $t_{res} \lesssim 20$ ns. Using an appropriately adapted function [16], the reconstruction tries to optimize the overall *likelihood* of the measured time-residuals and amplitudes. The starting track for this likelihood reconstruction is the result from the line-fit.

One challenge for the reconstruction is to identify and reject the noise hits within an event. Therefore a process called hit-cleaning is used to reject noise hits during reconstruction. Contrary to a normal filter which rejects or passes a full event, hit-cleaning rejects individual hits, but does not reject events. It is especially important for the analytic line-fit, while the likelihood reconstruction is less sensitive to individual noise hits. For each reconstruction step individual criteria determine which hits are ignored and which were accepted. So only a summary of potential criteria is given here, the numbers given just indicate approximate values. If not stated otherwise, the listed criteria only reject hits for an individual reconstruction step. In general rejected hits from the line-fit do enter the likelihood reconstruction. Details on this can be found in [16, 35].

1. Only hits within a given time-window ($\sim$ 5 μs) are accepted. Hits outside this window are rejected for all reconstruction steps.

2. Hits which are time isolated from all other hits by more than $\sim$ 700 ns are ignored.

3. Hits are ignored when no other closeby PMT was hit within a short time window.

4. Hits with amplitudes below a certain threshold are ignored.

5. In case of multi-hits on one PMT, all but the first hit (least scattered) are ignored.

Filtering during reconstruction is important to improve speed by rejecting down-going muons or events which are unlikely to be reconstructed correctly. A cut on the reconstructed zenith angle is performed after each step of the reconstruction. To reject "bad" events, which probably were not reconstructed correctly, further filters (described in the next two chapters) can be applied. Apart from events which do not leave enough information inside the detector, "bad" events can for example come from muon bundles, stopping muons, bremsstrahlung within the detector or muons traversing the ice below the detector. E.g. if an individual incident cosmic ray causes the occurrence of several muons, two or more muons can arrive simultaneously at the detector. In this case it is difficult to assign the individual PMT hits to the various muons.

5.3 Quality Criteria

In this section the input parameters used during the neural network analysis are described. For those parameters used in the AMANDA-B4 analysis, the distributions before filtering, after filtering and after the transformation process (of the filtered data) is shown. Points represent data taken on June the 23rd 1996, the solid line refers to simulated atmospheric muons and the dotted line shows the distributions for simulated and good reconstructed signal muons. In order to make them comparable, all distributions are normalized to percentages rather than absolute numbers. The statistics entering these plots are the following:

- Unfiltered (filtered) measured data events: $1.4 \cdot 10^6$ (4523)
- Unfiltered (filtered) MC background events: $8.1 \cdot 10^5$ (2872)
- Unfiltered (filtered) MC signal events: $3.6 \cdot 10^4$ (3332)

Further details on the samples which enter these plots can be found at the beginning of chapter 7. As is explained in that chapter, data (dominated by mis-reconstructed atmospheric muons) and Monte Carlo background should be compared after filtering, not before. The filters applied for this AMANDA-B4 analysis are explained in detail in chapter 7. They were:

- At least eight hits were required
- The reconstruction error has to be less than 15° for signal muons
- The zenith angle of the line-fit has to be bigger than 120°
- The reconstructed zenith angle has to be bigger than 120°

- Three hits with a time residual of less than 25 ns are demanded
- The reconstructed track must be more likely than a down-going one.

The mentioned transformations of the filtered events are necessary since the neural network simulator requires input data from the range [0..1]. Those used for the AMANDA-II analysis were slightly different from the transformations for the AMANDA-B4 analysis, so they are explained when the individual analysis is presented.

5.3.1 Direct Hits

As mentioned in section 5.2, time-residuals are zero for unscattered photons from the true muon track hitting jitter-free PMTs. Assuming that the reconstruction result was close to the true muon track, a calculated time-residual of approximately zero should indicate hits from undelayed photons. Such hits are called *direct hits* [35]. A large number of them indicates a good reconstruction. This is already known from conventional analysis [16, 35]. Three different definitions of direct hits were used:

- $N_{direct,\ 15}$: The number of direct hits with $-5ns \leq t_{res} \leq 15ns$
- $N_{direct,\ 25}$: The number of direct hits with $-5ns \leq t_{res} \leq 25ns$
- $N_{direct,\ 15-25}$: The number of direct hits with $15ns < t_{res} \leq 25ns$.

$N_{direct,\ 15}$ and $N_{direct,\ 15-25}$, which were used for the AMANDA-B4 analysis are shown in figure 5.15. It can be seen that signal events are accumulated at large N_{direct} values. It can also be seen that Monte Carlo and data do agree pretty well.

5.3.2 Likelihood

During reconstruction, different tracks are tested until the one with the best likelihood for this event is found. The absolute likelihood[2] of this final track is a measure for the quality of the reconstruction. Since there are events with very different numbers of channels, the likelihood itself may not always be strictly comparable. Therefore not only the likelihood $\mathcal{L}$, but also the likelihood per channel $\mathcal{L}/N_{ch}$ was used. N_{ch} is the number of PMTs hit in an event. Since PMTs can resolve more than one hit per event, this is in general smaller than the number of hits. These parameters are shown in figure 5.16. As expected signal events are accumulated at low values (large probabilities). MC and data agree pretty well. With a higher resolution, one actually sees that the data and MC background distributions contain equivalent structures at the same values, however high likelihood values seem to occur more frequent for data.

5.3.3 Velocity of Line-Fit

The line-fit is the first guess for the muon track. If the initial first guess is already good, a good reconstruction is likely. As was described above, the velocity (and therefore its

[2]Strictly speaking the parameter called likelihood in this work is not the likelihood itself, but twice the negative logarithm of it.

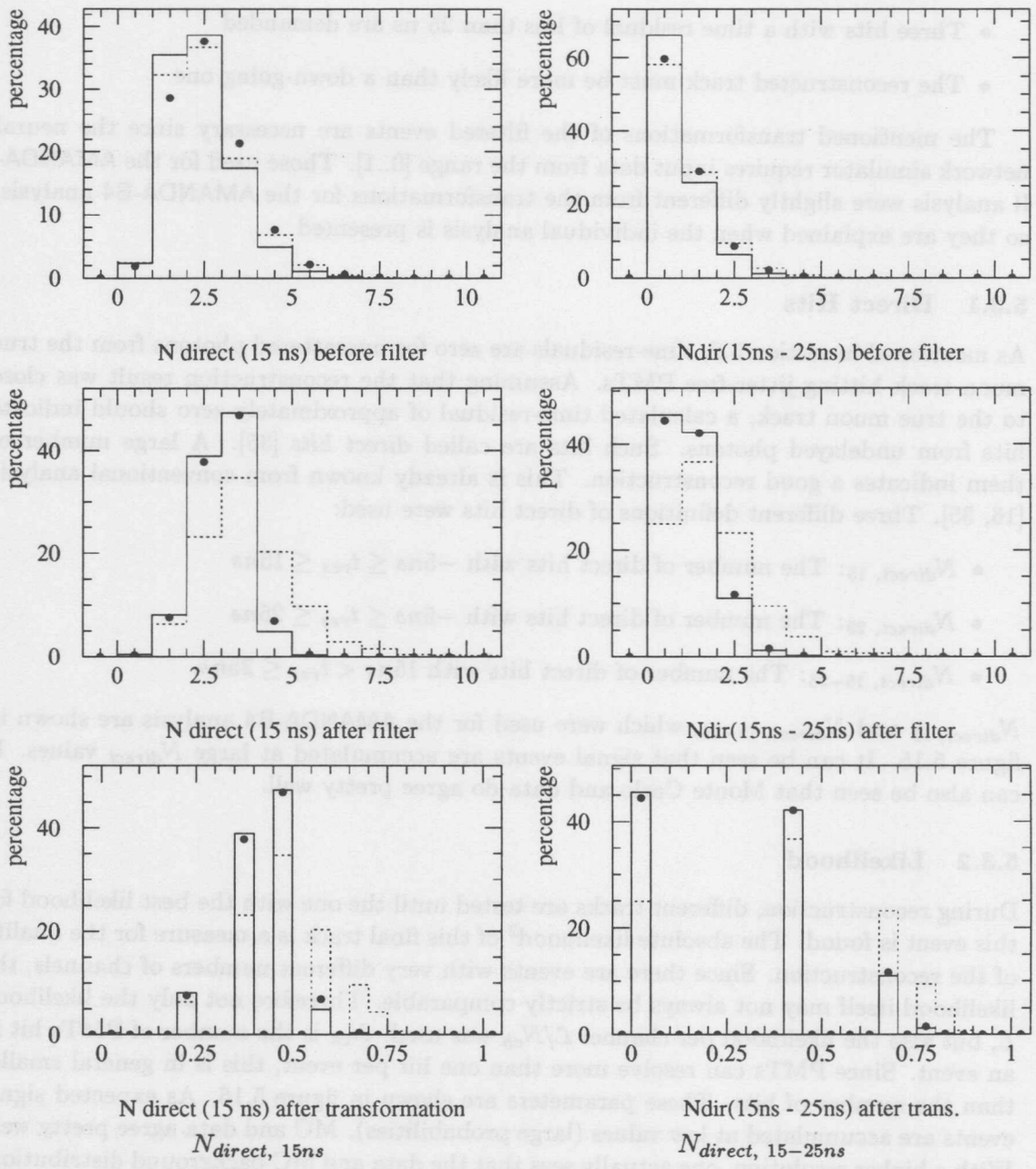


Figure 5.15: Distribution of direct hits. Left column: hits with time residuals between -5 and +15 ns, right column: hits with time residuals between +15 and +25 ns. First line: unfiltered; second line: filtered; third line: transformed distribution.

absolute value $|\vec{v}|$) of the line-fit is a free parameter, adjusted to minimize the error of the line-fit. Since this is no physical velocity, it may become larger than 0.3 m/ns, so a normal transformation has to be applied to this variable. It is assumed that the first guess results in a better track, when $|\vec{v}|$ is large. In an ideal case it is equal to the speed of light in ice.

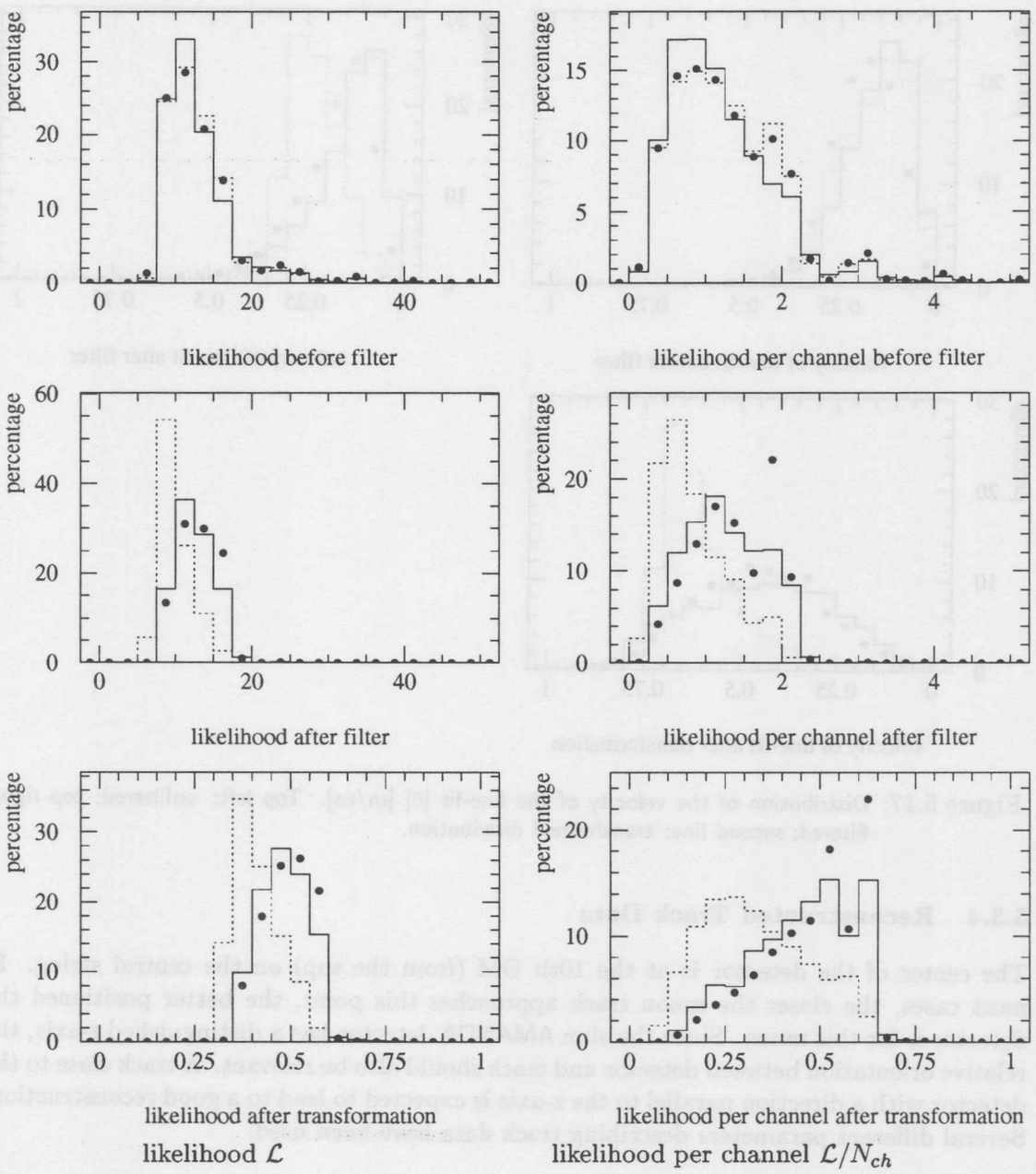


Figure 5.16: Distribution of likelihoods. Left column: likelihood itself; right column: likelihood per channel. First line: unfiltered; second line: filtered; third line: transformed distribution.

This parameter is shown in figure 5.17. Signal (containing good reconstructed events only) indeed has high values of $|\vec{v}|$. There is a slight shift to smaller values for MC compared to data (middle row). By correcting for this shift within the transformation, the agreement between the two distributions becomes good. Together with the transformation itself, this correction is described in section 7.3.

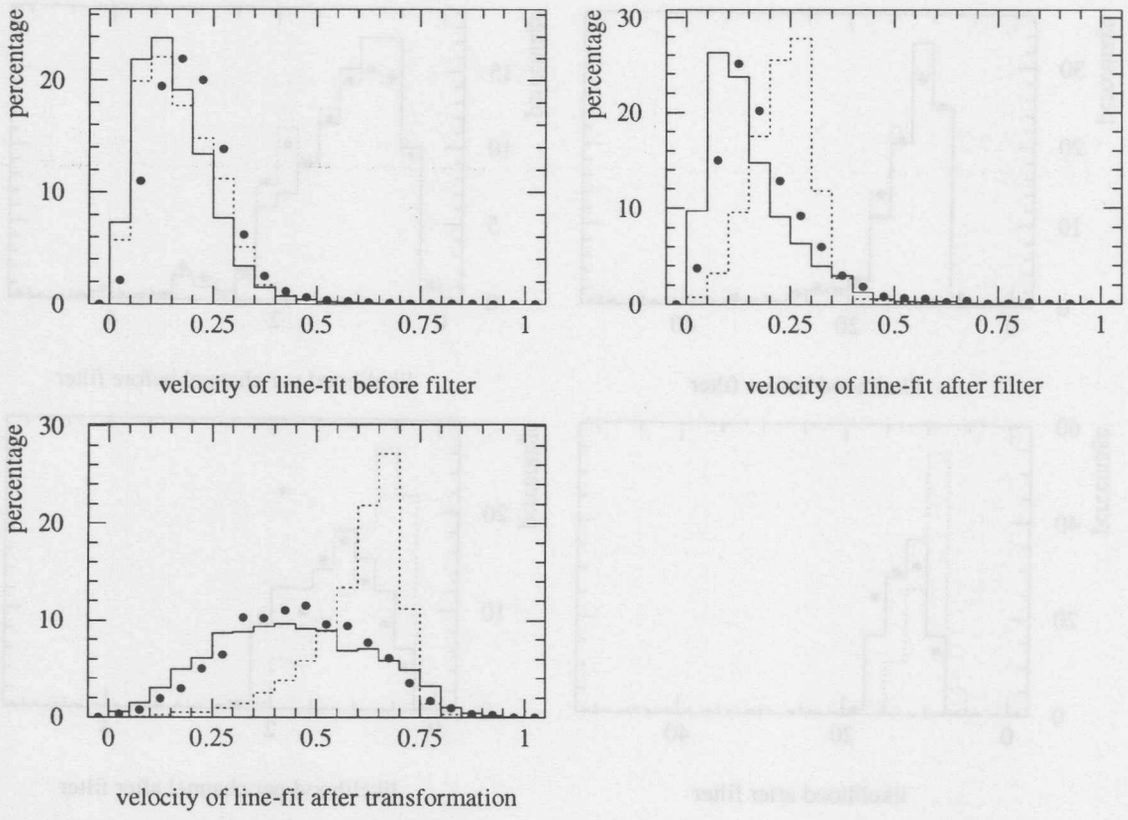


Figure 5.17: Distribution of the velocity of the line-fit $|\vec{v}|$ [m/ns]. Top left: unfiltered; top right: filtered; second line: transformed distribution.

5.3.4 Reconstructed Track Data

The center of the detector is at the 10th OM (from the top) on the central string. In most cases, the closer the muon track approaches this point, the better positioned the detector is for this muon. Since the slim AMANDA detector has a distinguished z-axis, the relative orientation between detector and track should also be relevant. A track close to the detector with a direction parallel to the z-axis is expected to lead to a good reconstruction. Several different parameters describing track data have been used:

- $\mathcal{R}_{rec}$ is the distance between the detector center and the reconstructed muon track.
- $\mathcal{Z}_{rec}$ is the difference between the z-coordinates of these two points.
- ρ_{rec} is the distance between the two points when projected onto the x-y plane.
- θ_{rec} is the reconstructed zenith angle. It is presented in figure 5.18. As mentioned in section 2.3, a cut on this parameter is required during reconstruction.

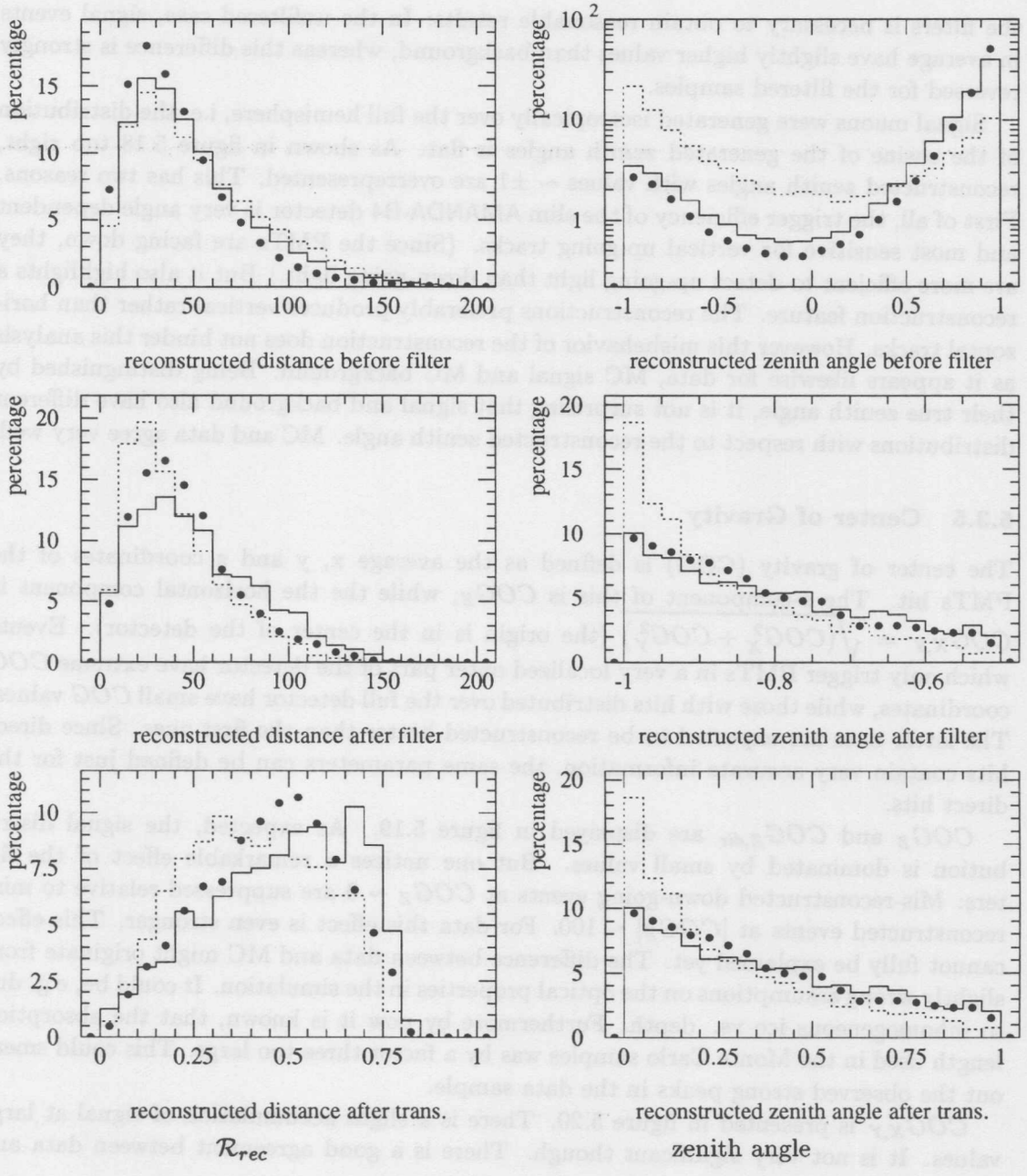


Figure 5.18: Distribution of the distance from detector center to the reconstructed muon track and the reconstructed zenith angle. Left column: distance between reconstructed track and detector center; right column: reconstructed zenith angle. First line: unfiltered; second line: filtered; third line: transformed distribution.

As can be seen in figure 5.18, $\mathcal{R}_{rec}$ does accumulate signal at low values. Data and MC both peak at $\mathcal{R}_{rec} \sim 30$m, but MC has too many events in the tail. This difference is not understood so far. The parameter $\mathcal{R}_{rec}$ is a nice example, where the application of

the filters is necessary to obtain reasonable results: In the unfiltered case, signal events in average have slightly higher values than background, whereas this difference is strongly reversed for the filtered samples.

Signal muons were generated isotropically over the full hemisphere, i.e. the distribution of the cosine of the generated zenith angles is flat. As shown in figure 5.18 top right, reconstructed zenith angles with values $\sim \pm 1$ are overrepresented. This has two reasons. First of all, the trigger efficiency of the slim AMANDA-B4 detector is very angle dependent and most sensitive for vertical up-going tracks. (Since the PMTs are facing down, they are more efficient to detect up-going light than down-going light.) But it also highlights a reconstruction feature. The reconstructions preferably produces vertical rather than horizontal tracks. However this misbehavior of the reconstruction does not hinder this analysis as it appears likewise for data, MC signal and MC background. Being distinguished by their true zenith angle, it is not surprising that signal and background also have different distributions with respect to the reconstructed zenith angle. MC and data agree very well.

5.3.5 Center of Gravity

The center of gravity (COG) is defined as the average x, y and z coordinates of the PMTs hit. The z-component of this is COG_Z, while the the horizontal component is $COG_{X,Y} = \sqrt{(COG_X^2 + COG_Y^2)}$ (the origin is in the center of the detector). Events which only trigger PMTs in a very localized outer part of the detector have extreme COG coordinates, while those with hits distributed over the full detector have small COG values. The latter ones are expected to be reconstructed better than the first ones. Since direct hits contain very accurate information, the same parameters can be defined just for the direct hits.

COG_Z and $COG_{Z,dir}$ are displayed in figure 5.19. As expected, the signal distribution is dominated by small values. But one notices a remarkable effect of the filters: Mis-reconstructed down-going events at $COG_Z \sim 0$ are suppressed relative to mis-reconstructed events at $|COG_Z| \sim 100$. For data this effect is even stronger. This effect cannot fully be explained yet. The difference between data and MC might originate from slightly wrong assumptions on the optical properties in the simulation. It could be, e.g. due to inhomogeneous ice vs. depth. Furthermore by now it is known, that the absorption length used in the Monte Carlo samples was by a factor three too large. This could smear out the observed strong peaks in the data sample.

$COG_{X,Y}$ is presented in figure 5.20. There is a slight accumulation of signal at large values. It is not very significant though. There is a good agreement between data and Monte Carlo.

5.3.6 Combined Reconstruction and Hit Data

Apart from $COG_Z(direct)$, several further combinations of reconstruction and hit data were used. One of them is the "length of the hits", $\mathcal{Z}_{leng}$. If one projects the locations of the PMTs hit onto the reconstructed track, one obtains an estimate for the length over which the muon's radiation entered the reconstruction. The reconstruction quality is expected to benefit from a long "lever arm". Again this can also be defined for the subsample of direct

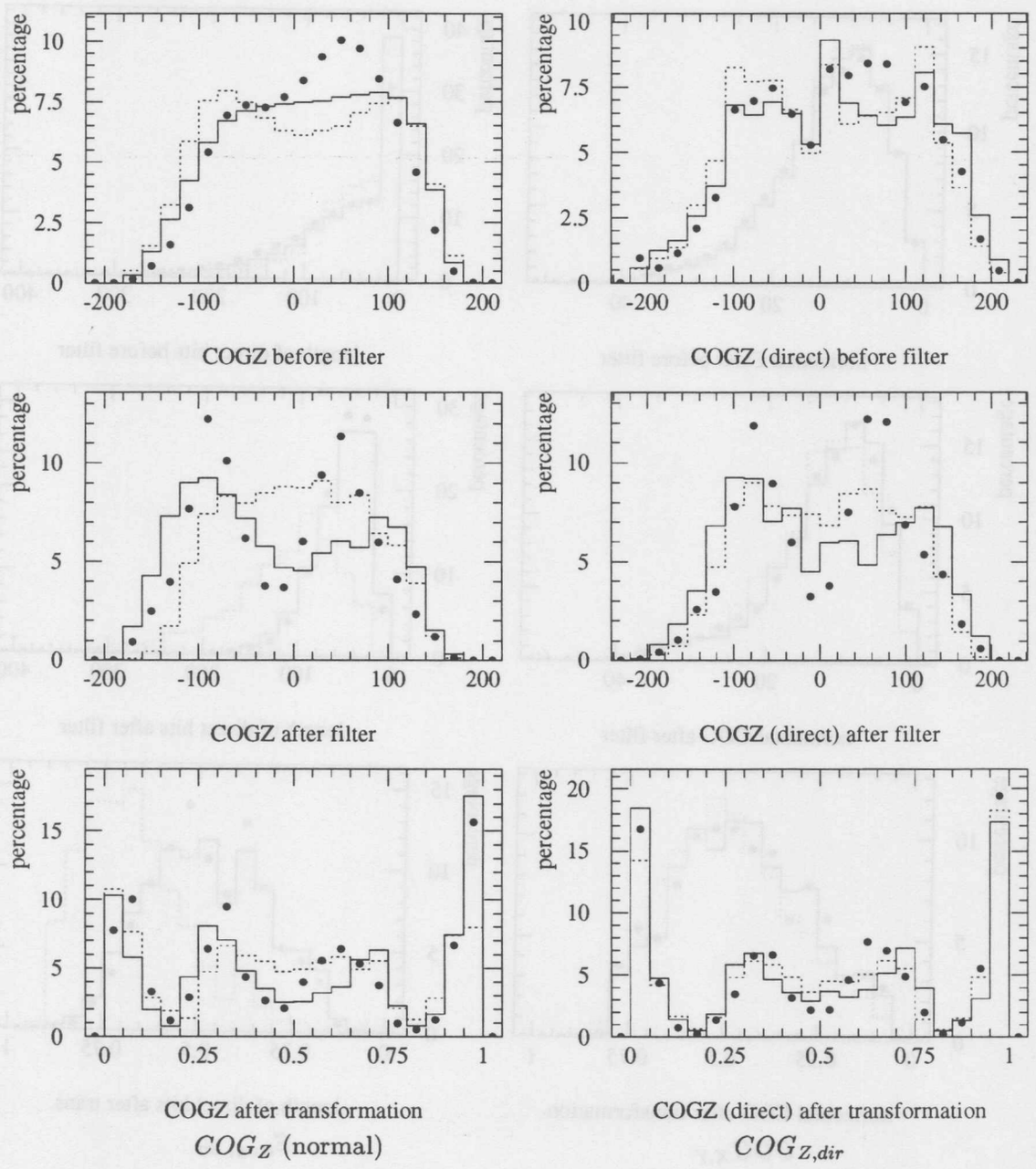


Figure 5.19: Distribution of COG_Z. Left column: normal COG_Z; right column: COG_Z just for direct hits (time residual smaller 25 ns). First line: unfiltered; second line: filtered; third line: transformed distribution.

hits only. This is then the length of direct hits $\mathcal{Z}_{leng,\,dir}$. It is displayed in figure 5.20. As expected, background events predominantly occur at small values, while signal events have a much broader distribution. One notices that there is a nice agreement between data and MC.

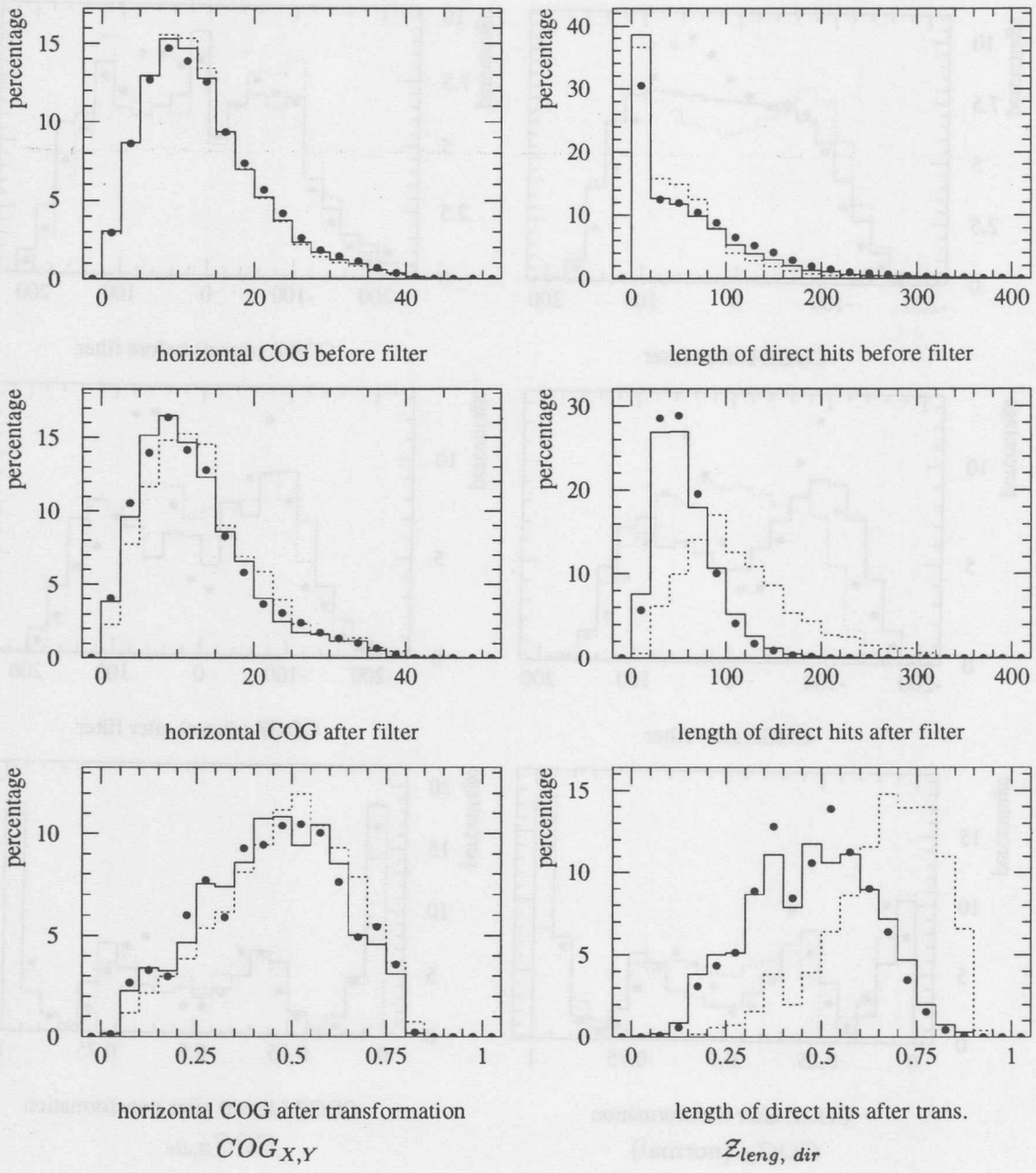


Figure 5.20: Distribution of $COG_{X,Y}$ and $\mathcal{Z}_{leng,\ dir}$. Left column: Horizontal component of center of gravity ($COG_{X,Y}$); right column: Length of direct hits (time residual smaller 25 ns). First line: unfiltered; second line: filtered; third line: transformed distribution.

The likelihood of the hit topology $\mathcal{L}_{hit}$ is the probability that the reconstructed track illuminates the hit PMTs and just those. This probability ignores timing or amplitude information, but nevertheless should be a parameter indicating the quality of the reconstruction. It is presented in figure 5.21. As hoped, good reconstructed signal events

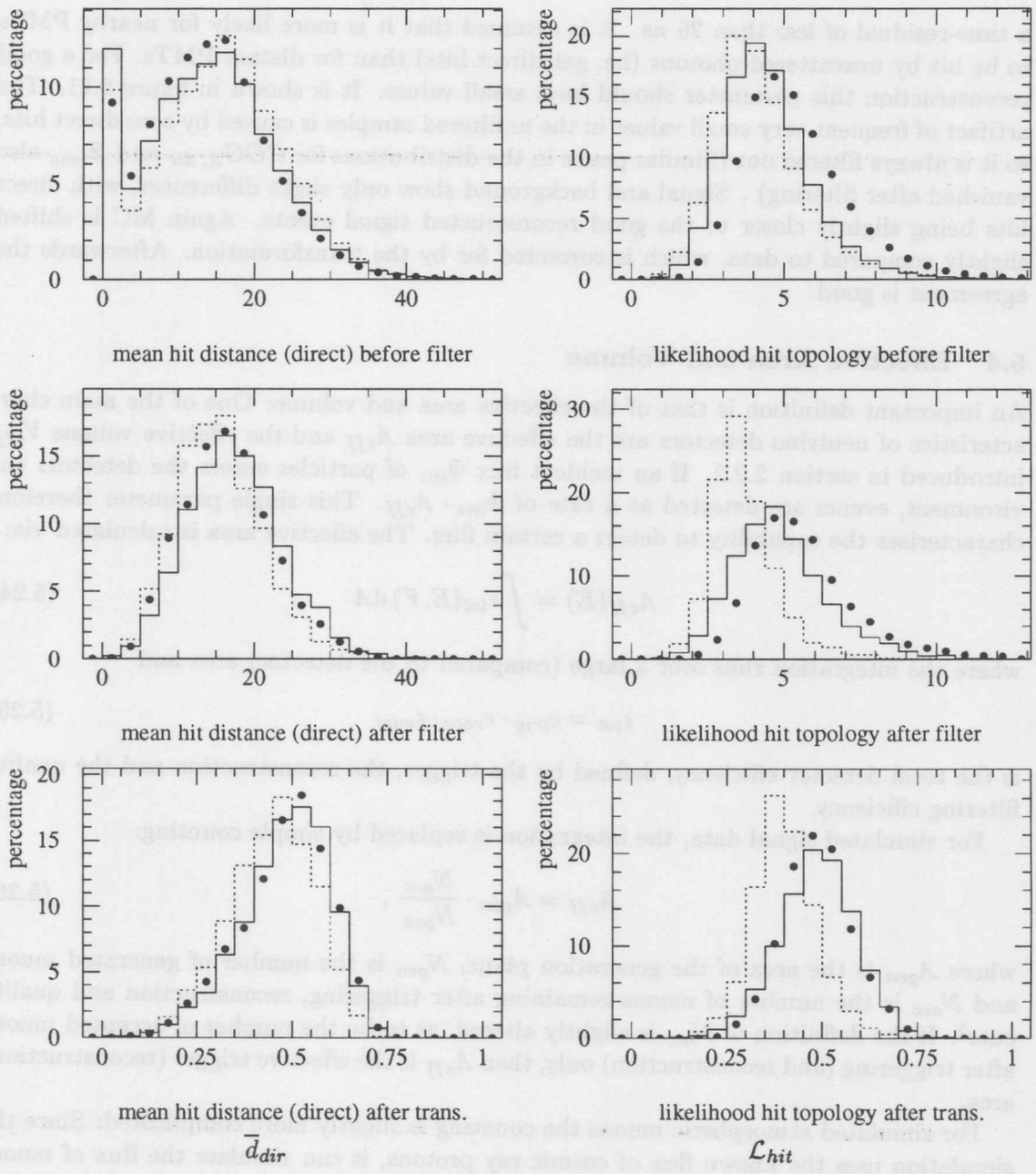


Figure 5.21: Average distance of direct hits (left column) and hit probability (right column). First line: unfiltered; second line: filtered; third line: transformed distribution.

are more likely (have smaller likelihood values) than background. Since MC is shifted to slightly smaller values than data, the transformation corrects for this difference. For details see section 7.3. After this correction, MC and data agree very well.

The *average distance of direct hits*, $\overline{d}_{dir}$ is the last of these parameters used. It is the average distance of PMTs to the reconstructed track, but just for those PMTs hit with

a time-residual of less than 25 ns. It is assumed that it is more likely for nearby PMTs to be hit by unscattered photons (i.e. get direct hits) than for distant PMTs. For a good reconstruction this parameter should have small values. It is shown in figure 5.21. The artifact of frequent very small values in the unfiltered samples is caused by zero direct hits, so it is always filtered out (Similar peaks in the distributions for $COG_{Z,\,dir}$ and $\mathcal{Z}_{leng}$ also vanished after filtering) . Signal and background show only slight differences, with direct hits being slightly closer to the good reconstructed signal events. Again MC is shifted slightly compared to data, which is corrected for by the transformation. Afterwards the agreement is good.

5.4 Effective Area and Volume

An important definition is that of the effective area and volume: One of the main characteristics of neutrino detectors are the effective area A_{eff} and the effective volume V_{eff} introduced in section 2.2.2. If an incident flux Φ_{inc} of particles enters the detectors environment, events are detected at a rate of $\Phi_{inc} \cdot A_{eff}$. This single parameter therefore characterizes the capability to detect a certain flux. The effective area is calculated via:

$$A_{eff}(E) = \int \epsilon_{tot}(E, \vec{r})\, dA \tag{5.24}$$

where the integration runs over a large (compared to the detector) area and

$$\epsilon_{tot} = \epsilon_{trig} \cdot \epsilon_{reco} \cdot \epsilon_{qual} \tag{5.25}$$

is the total detector efficiency, defined by the trigger, the reconstruction and the quality filtering efficiency.

For simulated signal data, the integration is replaced by simple counting:

$$A_{eff} = A_{gen} \cdot \frac{N_{acc}}{N_{gen}}\,, \tag{5.26}$$

where A_{gen} is the area of the generation plane, N_{gen} is the number of generated muons and N_{acc} is the number of muons remaining after triggering, reconstruction and quality cuts[3]. If the definition of N_{acc} is slightly altered, as to be the number of accepted muons after triggering (and reconstruction) only, then A_{eff} is the effective trigger (reconstruction) area.

For simulated atmospheric muons the counting is slightly more complicated: Since the simulation uses the known flux of cosmic ray protons, it can simulate the flux of muons through the detector Φ_μ and the resulting trigger rate of the detector R_{trig}. The product of both values is the effective trigger area. Now, a similar calculation as above yields:

$$A_{eff} = A_{eff,trig} \cdot \frac{N_{acc}}{N_{trig}} = \Phi_\mu \cdot R_{trig} \cdot \frac{N_{acc}}{N_{trig}}\,. \tag{5.27}$$

[3]This definition of A_{eff} is independent of the area of the generation plane or other features of the generation process: The effect of an increased area is canceled by an increased N_{gen}, so that only a generation density enters the formula. This density is proportional to the acceptance density, which in turn is proportional to the number of accepted muons. So all generation specific numbers cancel out and the remaining constants of proportionality give the effective area.

Likewise an effective volume can be defined:

$$V_{eff}(E) = \int \epsilon_{tot}(E, \vec{r})\, dV \ . \tag{5.28}$$

Again this integral can be replaced if muons were simulated inside a volume:

$$V_{eff} = V_{gen} \cdot \frac{N_{acc}}{N_{gen}} \ , \tag{5.29}$$

The relevance of A_{eff} was explained above; that of V_{eff} is to offer a different description of the detectors performance in cases where A_{eff} is not well defined, namely if the muon track is of finite length. Then an effective area can be calculated via

$$A_{eff}(E) = \frac{V_{eff}(E)}{< R_{\mu}(E) >} \ , \tag{5.30}$$

where $< R_{\mu} >$ is the mean muon range, introduced in equation 2.7.

6 AMANDA-II Analysis

AMANDA-II is the extended version of the present AMANDA-B detector, planned to be completed in the austral summer 1999/2000. In [16], it was shown that the simulated background of this detector can be rejected to a high degree by applying successive cuts on reconstruction parameters. In this chapter, a neural network analysis is compared to that "conventional" analysis. Furthermore an extension of the neural net analysis is performed by using additional input parameters.

6.1 Data used

For this work, a pair of one signal and one background Monte Carlo sample were used. For the signal sample, a large number of 1 TeV muons were generated isotropically over the full hemisphere. The generation planes were so large, that most generated muons missed the detector and therefore could not be triggered by the detector simulation. Due to energy losses, the $10 \cdot 10^3$ muons that did trigger had an average energy of about 600 GeV in the center of the detector. Out of these, 5291 were true "up-going" muons. A subsample of 5149 of these muons were also reconstructed as "up-going" and subsequently used as "up-going" or "signal" sample. The effective trigger area (integrated over all angles) for these muons was determined to be $\sim 1.0 \cdot 10^5$ m^2. For this analysis, the term "up-going" describes muons with zenith angles bigger than 80°, i.e. muons which come from within 10 degrees *above* the horizon ($\theta = 90°$) are still referred to as up-going.

For the background sample, the atmospheric shower program simulated $1.1 \cdot 10^6$ triggering events from atmospheric muons. In this case the effective trigger area was calculated to be $\sim 5.0 \cdot 10^4$ m^2. Out of the triggered events 19556 were mis-reconstructed as up-going (reconstructed zenith angle $\theta_{rec} > 80°$) and henceforth used as background sample.

During or after the reconstruction, cuts were also performed on the number of direct hits ($N_{direct,\ 25} \geq 3$), on the zenith angle of the line-fit reconstruction ($\theta_{line-fit} > 80°$) and on the reconstructed zenith angle ($\theta_{rec} > 80°$). Via equation 5.26 the effective area for background after filtering was calculated to be $0.89 \cdot 10^3$ m^2. For the signal sample, one is interested in effective areas from potential point sources, so equation 5.26 had to be adapted to compensate the effects of a reduced solid angle:

$$A_{eff,\ fil} = A_{eff,\ trig} \cdot \frac{N_{fil}}{N_{trig}} \cdot \frac{\Omega_{gen}}{\Omega_{fil}} , \tag{6.31}$$

where Ω_{gen} is the generated solid angle, Ω_{fil} is the solid angle accepted by the filter, N_{trig} and N_{fil} are the numbers of triggered events and of those left after filtering and $A_{eff,\ trig}$ is the effective trigger area. This calculation yielded an effective area of $90 \cdot 10^3$ m^2 after filtering.

In order to compare the neural net analysis to that from the AMANDA-II proposal, the parameters used in this analysis were the same as in the proposal. The extended analysis additionally used parameters describing the reconstructed track. All these are summarized in table 6.2.

	Cut value	description	Cut in [16]
(1)	$N_{direct,\ 15}$	Number of direct hits with $-5ns < t_{res} < 15ns$	> 4
(2)	$N_{direct,\ 25}$	Number of direct hits with $-5ns < t_{res} < 25ns$	> 6
(3)	$\mathcal{L}$	Likelihood of the reconstruction	< 15
(4)	$\mathcal{L}/N_{ch}$	Likelihood per hit channel	< 0.5
(5)	$\|\vec{v}\|$	velocity of line-fit (first guess)	$> 0.1\ m/ns$
(6)	COG_z	Center of gravity of hits (z-component)	$> -250\ m$
(7)	θ_{rec}	Reconstructed zenith angle	—
(8)	R_{rec}	reconstructed distance (to the detector center)	—
(9)	ρ_{rec}	reconstructed horizontal distance	—
(10)	Z_{rec}	reconstructed vertical distance	—

Table 6.2: Observables used as input to the neural networks investigated in this chapter. A more detailed description is given in section 5.3. The cuts used in the AMANDA-II proposal are stated in the fourth column.

6.2 Transformation

The transformations used in this analysis to limit input parameters to the range [0..1] were the following:

- For $\cos\theta$, which is restricted to values between -1 and 1, the transformation was:

$$\cos(\theta') = 0.5 \cdot \cos(\theta) + 0.5\ . \tag{6.32}$$

- For parameters x, which were restricted to positive values (like distances):

$$x' = \frac{2}{\pi} \cdot \arctan\left(\frac{x}{2 \cdot \overline{x}}\right)\ , \tag{6.33}$$

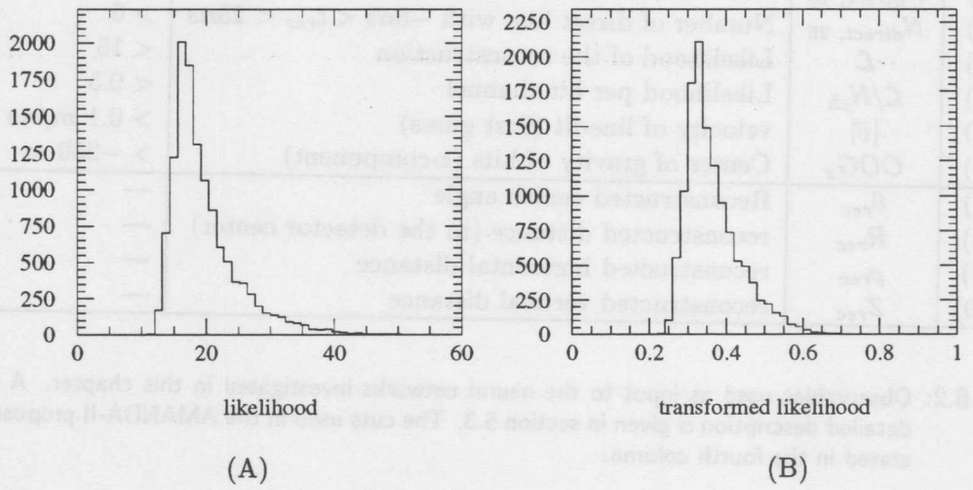
  where $\overline{x}$ is the mean of the distribution of x.

- For parameters x, which coould have positive and negative values (like coordinates), the means $\overline{x}_+$ and $\overline{x}_-$ were calculated separately for each of the two parts of the distribution. Then:

$$x' = \frac{1}{\pi} \cdot \arctan\left(\frac{x}{2 \cdot (\overline{x}_+ + |\overline{x}_-|)}\right) + 0.5\ . \tag{6.34}$$

The means were calculated from the Monte Carlo background distributions. As an example of the effect of such a transformation, the likelihood distributions for the untransformed and transformed simulated AMANDA-II background events are shown in figure 6.22.

In general, the transformations are a compromise between contradicting aims. First of all, all possible values had to be transformed to the range [0..1]. The arctan function is limited to the range $[-\frac{\pi}{2}..+\frac{\pi}{2}]$. Together with a simple linear transformation it can be



(A) (B)

Figure 6.22: Distribution of Likelihood values for the AMANDA-II simulation. (A) before and (B) after transformation to the input values of the network.

used to transform any distribution to [0..1]. Then the aim was to have a transformation "as linear as possible". Since $\arctan(x) = x - x^3/3 + ...$ for small x, it is roughly linear for $|x| < 0.7$. So the distribution was prescaled as to make sure that the main contributions got values smaller than 0.7. This was performed by a division by twice the mean. A division by bigger values was not desired, as then the main part of the distributions would be squeezed to a small range of values.

6.3 The Conventional Analysis

The analysis from this work cannot be compared directly to the analysis presented in the AMANDA-II proposal. Both were performed with the same data samples, but with slightly different definitions of the signal sample: In the AMANDA-II proposal, simulated up-going (neutrino-induced) muons which where mis-reconstructed as down-going were included in the signal sample. In this work all signal muons which were reconstructed as down-going were rejected. Only muons which were generated as up-going and reconstructed as up-going entered the signal sample. As a second difference to the proposal, for this neural network analysis, the MC samples had to be divided into one sample used for learning and one to analyze the resulting network. Therefore the analysis from the proposal was performed again, this time using only the control sample from the neural network analysis.

The rejection achieved with this reproduced analysis is shown in figure 6.23. In the left column, the reconstructed $\cos\theta$ distribution of the remaining mis-reconstructed atmospheric muons is shown. The right column presents the same distributions for the signal sample. In the first row, the full distributions without cuts are shown. Here the decrease in signal event numbers at large $\cos\theta$ values is due to edge effects during reconstruction:

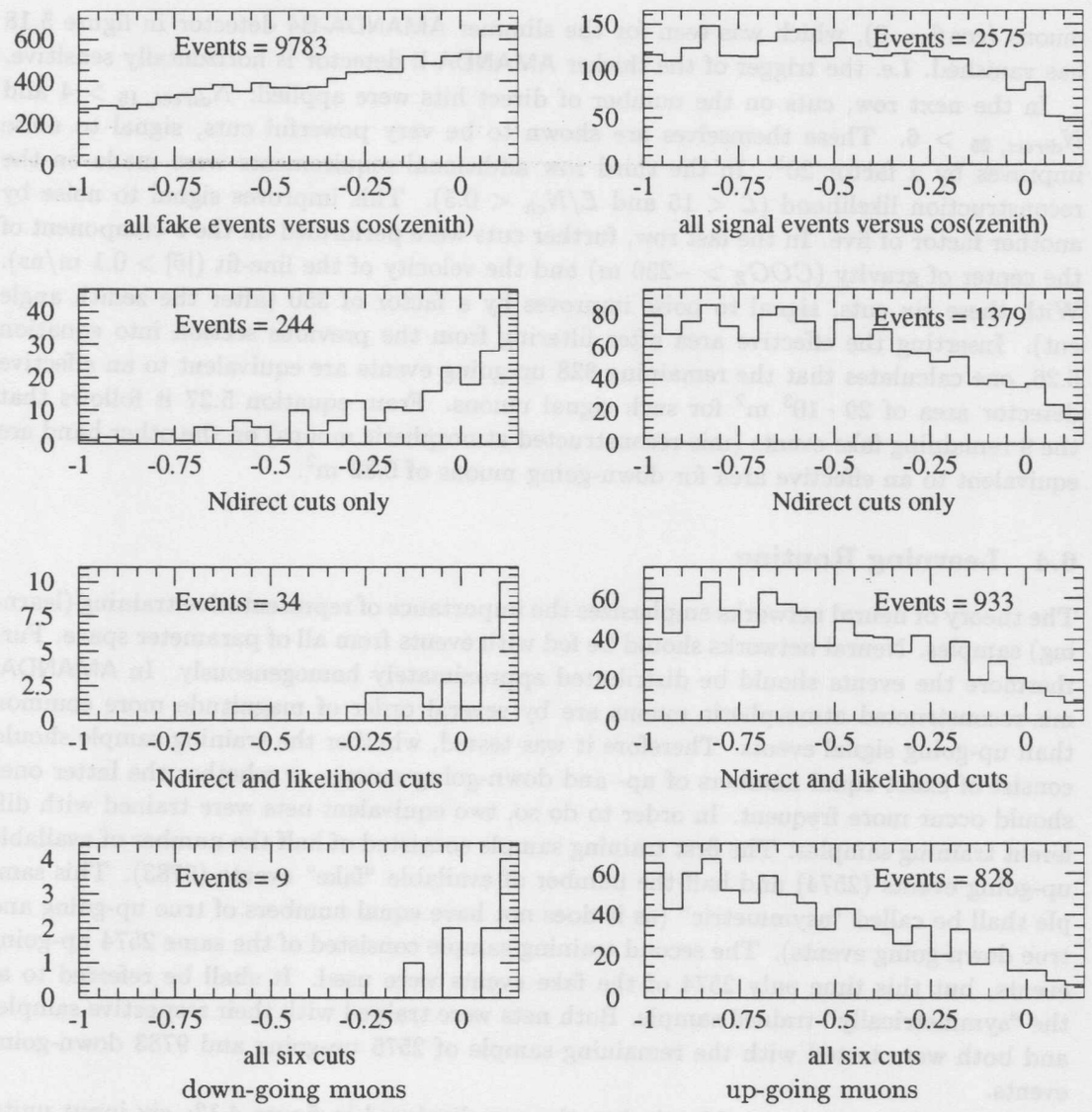


Figure 6.23: Stepwise rejection of mis-reconstructed atmospheric muons. The figure shows the reconstructed $\cos\theta$ distribution during various stages of the conventional quality analysis presented in [16]. Left column: fakes (mis-reconstructed muons); right column: up-going events.
First line: all events, second line: remaining events after cuts $N_{direct,\ 15} > 4$ and $N_{direct,\ 25} > 6$, third line: remaining events after additional cuts: $\mathcal{L} < 15$ and $\mathcal{L}/N_{ch} < 0.5$, fourth line: remaining events after additional cuts:$|\vec{v}| > 0.1$ m/ns and $COG_Z > -250$ m.

Compared to events with large $\cos\theta_{rec}$ values, events with $\cos\theta_{rec} \sim 80°$ are more likely to come from mis-reconstructed true down-going muons, which are not included in this sample. It is interesting to notice that the effect of very low trigger efficiencies for horizontal

muons ($\cos\theta \sim 0$), which was seen for the slimmer AMANDA-B4 detector in figure 5.18 has vanished. I.e. the trigger of the thicker AMANDA-II detector is horizontally sensitive.

In the next row, cuts on the number of direct hits were applied: $N_{direct,\ 15} > 4$ and $N_{direct,\ 25} > 6$. These themselves are shown to be very powerful cuts, signal to noise improves by a factor 20[1]. In the third row additional requirements were made on the reconstruction likelihood ($\mathcal{L} < 15$ and $\mathcal{L}/N_{ch} < 0.5$). This improves signal to noise by another factor of five. In the last row, further cuts were performed on the z-component of the center of gravity ($COG_Z > -250$ m) and the velocity of the line-fit ($|\vec{v}| > 0.1$ m/ns). With these six cuts, signal to noise improves by a factor of 350 (after the zenith angle cut). Inserting the effective area after filtering from the previous section into equation 5.26, one calculates that the remaining 828 up-going events are equivalent to an effective detector area of $29 \cdot 10^3$ m^2 for such signal muons. From equation 5.27 it follows that the 9 remaining fake events (mis-reconstructed atmospheric muons) on the other hand are equivalent to an effective area for down-going muons of 0.82 m^2.

6.4 Learning Routine

The theory of neural networks emphasizes the importance of representative training (learning) samples. Neural networks should be fed with events from all of parameter space. Furthermore the events should be distributed approximately homogeneously. In AMANDA, mis-reconstructed atmospheric muons are by several order of magnitude more common than up-going signal events. Therefore it was tested, whether the training sample should consist of exact equal numbers of up- and down-going muons, or whether the latter ones should occur more frequent. In order to do so, two equivalent nets were trained with different training samples: The first training sample consisted of half the number of available up-going events (2574) and half the number of available "fake" events (9783). This sample shall be called "asymmetric" (as it does not have equal numbers of true up-going and true down-going events). The second training sample consisted of the same 2574 up-going events, but this time only 2574 of the fake events were used. It shall be referred to as the "symmetrically" trained sample. Both nets were trained with their respective samples and both were tested with the remaining sample of 2575 up-going and 9783 down-going events.

The networks used were identical to the one displayed in figure 4.13: six input units, one layer of six hidden units and two output units. The input parameters were all six parameters known from the AMANDA-II proposal, summarized in the top of table 6.2. The result of the networks was given by the difference between the output units "goodness" and "badness". I.e. an output $\mathcal{O}$ close to +1 indicated a likely up-going event, while one close to -1 indicated a likely down-going event. The performances of the networks are displayed in figures 6.24 and 6.25. The first row of the plots displays the general shape of the output distributions. It looks reasonable in that sense, that the distributions are falling from the demand output to the other extreme. However the very tail of the up-going

[1] The term *signal to noise* in this circumstance is slightly misleading. This is obviously no real signal to noise ratio, as the signal sample chosen is unrealistic (only 1 TeV neutrinos interact exactly 800 m before the detector) and atmospheric neutrinos are not accounted for. In this work, the expression signal to noise is thus only used to describe the relative efficiencies for the MC samples under consideration.

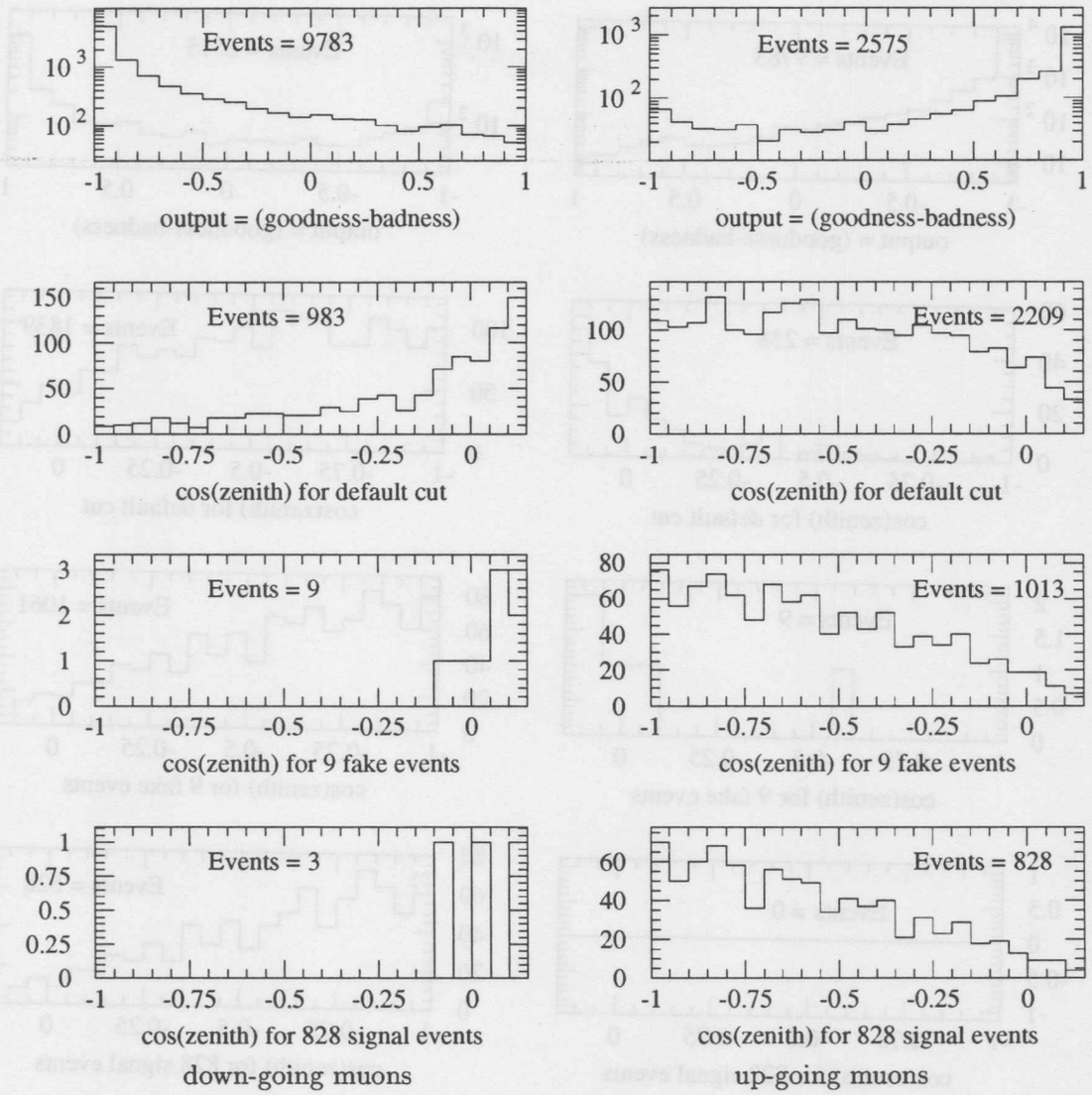


Figure 6.24: "Symmetrically" trained net: It was trained with 2574 fake and signal events and tested with 9783 fake and 2575 signal events. Left column: fakes, right column: up-going events.
First line: output ($\mathcal{O}$) distribution (notice the peak at $\mathcal{O} \sim -1$ for up-going muons), second line: $cos\theta_{rec}$ distribution for the "default cut" $\mathcal{O} > 0$, third line: $cos\theta_{rec}$ distribution for $\mathcal{O} > 0.961$ (leaves 9 fakes), last line: $cos\theta_{rec}$ distribution for stringent cut $\mathcal{O} > 0.98029$ (leaves 828 up-going events)

spectrum slightly rises again as $\mathcal{O}$ approaches -1. The reason for this is not understood. The events within this unexpected peak are not distinguished with respect to any of the six parameters used or with respect to their reconstructed or true zenith angle. There is no corresponding peak at $\mathcal{O} \sim +1$ for down-going muons. This peak effect is slightly stronger for the asymmetrically trained net.

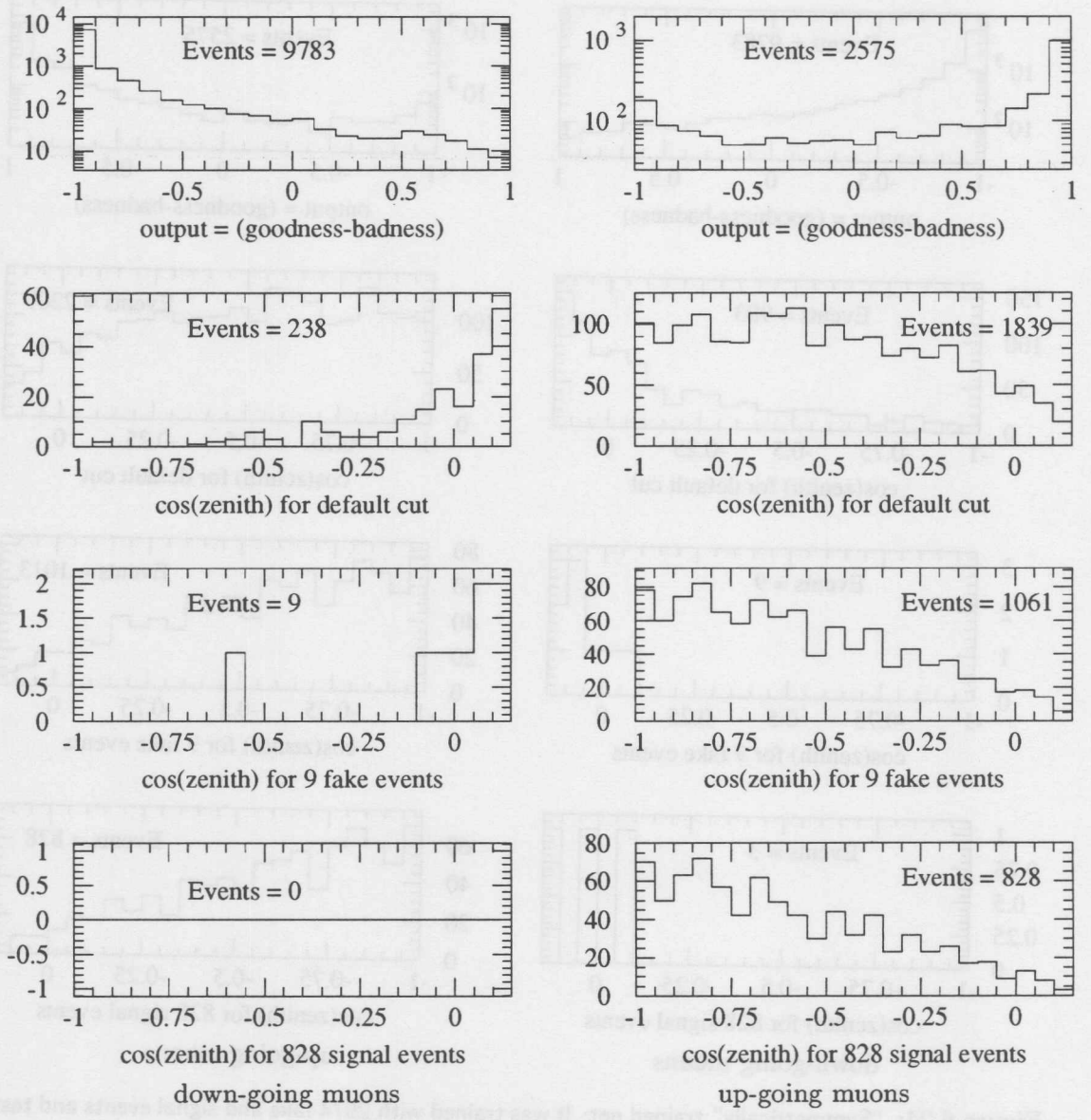


Figure 6.25: "Asymmetrically" trained net: It was trained with 9783 fake and 2574 signal events and tested with 9783 fake and 2575 signal events. Left column: fakes, right column: up-going events.
First line: output ($\mathcal{O}$) distribution (notice the peak at $\mathcal{O} \sim -1$ for up-going muons), second line: $cos\theta_{rec}$ distribution for the "default cut" $\mathcal{O} > 0$, third line: $cos\theta_{rec}$ distribution for $\mathcal{O} > 0.87$ (leaves 9 fakes), last line: $cos\theta_{rec}$ distribution for $\mathcal{O} > 0.93965$ (leaves 828 up-going events)

A comparison between neural networks is not trivial: Such networks are non-linear and therefore no natural cuts on their output exists. E.g. in the second row of the figures 6.24 and 6.25, a "default cut" at $\mathcal{O} = 0$ was performed. Here the comparison between the results is ambiguous: The asymmetrically trained net has a signal passing rate of 74 %

and a background rejection of 97.1 %. The numbers for the other net are 86 % and 90.0 % respectively. So this "default cut" turned out to be more stringent for the asymmetrically trained network. In order to compare the two learning methods, two other cuts were used: In the third row of the figures, a cut on the output was performed in such a way, that 9 fake events (like in the AMANDA-II proposal) remained. In this case 54 additional signal events ($\sim 1.5\ \sigma$) passed the asymmetrically trained net. In the last row a similar cut is set, as to leave 828 (again number taken from the proposal) signal events. Here three fake events passed the symmetrically trained net, while the asymmetrically trained net rejected 100 %. Comparing how many signal (fake) events remain in the third (fourth) row, one sees that the asymmetrically trained net is performing better than the symmetrically trained one. Therefore learning continued from now on with a sample containing almost four times more fake events than signal events. The results of this test are summarized in table 6.3. It also states the cuts required on the neural network output to achieve the demanded passing/rejection rates.

	proposal	symmetric	asymmetric
remaining up-going for 9 fakes left	828	1013 (0.961)	1061 (0.87)
remaining fakes for 828 up-going passing	9	3 (0.98029)	0 (0.93965)

Table 6.3: Results of the two different learning routines.
First row: resulting number of up-going events for a fixed number of 9 fake events. Second row: resulting number of fake events for a fixed number of 828 up-going events. Also given: the required cut on the neural net output to leave the desired number of events (in brackets) .

6.5 Hidden Units

In order to test the influence of the number of units in the hidden layer a "6-6-2net" was compared to a "6-3-2net". I.e. in both cases six input units and two output units were used, but once there was a layer of six hidden units and once a layer of three hidden units. The input parameters were the same six as in the previous section. The results are shown in figure 6.26.

In the left column, the results from the bigger net are plotted, in the right column those of the smaller net. In the first line the output distributions for all fakes are shown, in the next line the corresponding distributions for all signal events. No obvious differences appear. But using a cut criterion as to leave only 828 up-going events (third line) one sees that the 6-3-2net does not perform as good as the bigger net: It leaves two fakes rather than none. In the last line a cut criterion is set so that 9 fakes remain. Now the reconstructed $\cos\theta$ distribution is plotted for the remaining signal events. Again the larger net seems to have a better performance. It keeps 1061 signal events, while the other net only leaves 987. The improvement is a 2 σ effect.

This indicates, that using more hidden units gives a better network performance. A further increase in the number of hidden units did not improve the nets any more, while a decrease worsened the result. For the investigation of different numbers of input parame-

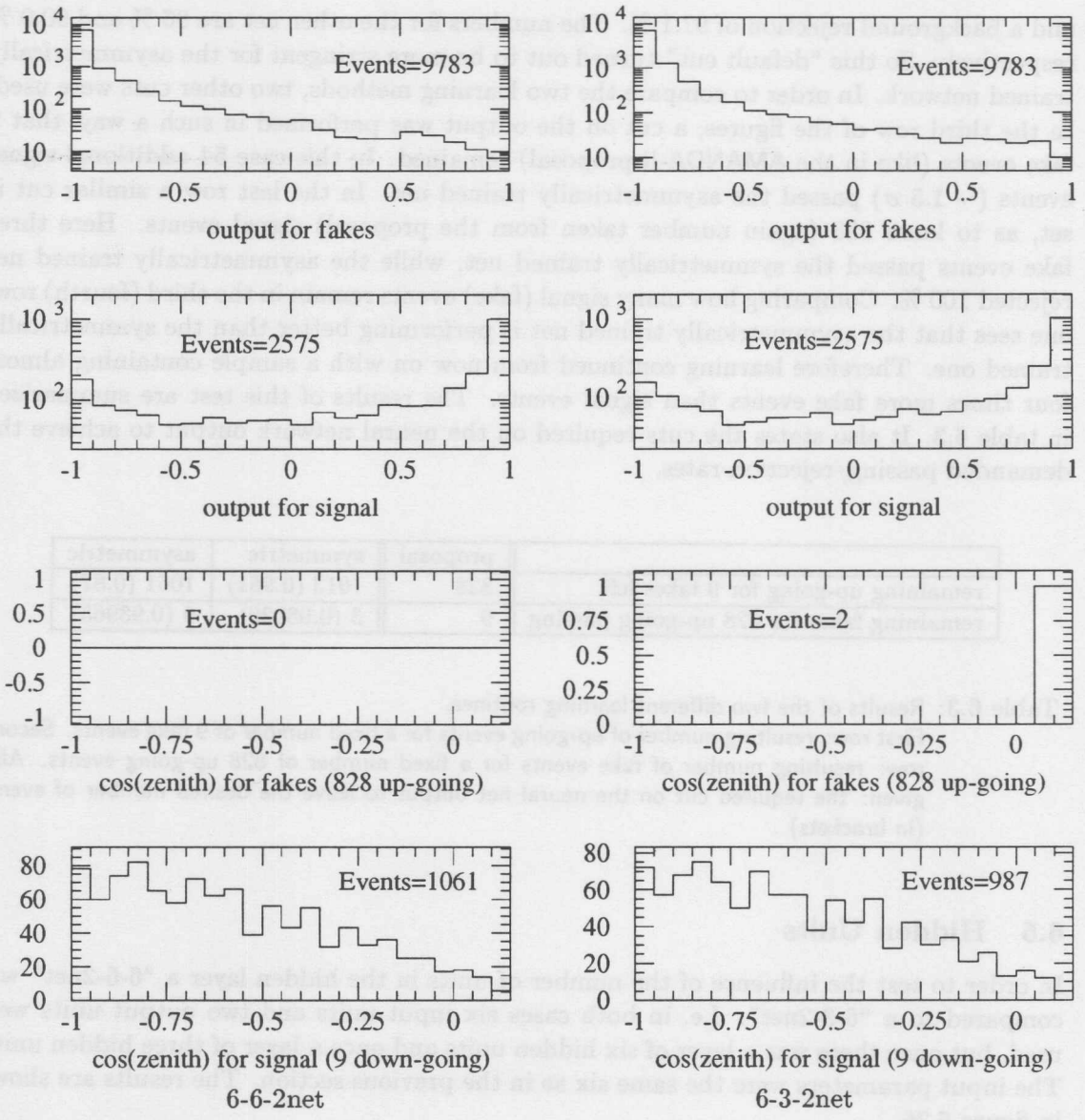


Figure 6.26: Test of number of units in the hidden layer: Left column: six hidden units; right column: three hidden units.
First (Second) line: fakes (signal) events versus output. Third line: fakes for 828 remaining signal events (Cuts: $\mathcal{O} > 0.93965$ and $\mathcal{O} > 0.939$). Bottom line: up-going events for 9 remaining down-going events (Cuts: $\mathcal{O} > 0.87$ and $\mathcal{O} > 0.908$).

ters it was decided to generally use networks with the same number of units in the hidden layer as are in the input layer. Once the set of input parameters of such a neural network analysis has converged to a fixed number, this test of the optimal number of hidden units should be re-performed. The rule of thumb obtained here might otherwise lead to nets with not enough hidden units (worse performance) or it could lead to too many hidden

units, which would yield slow networks: The speed of such a network is proportional to its number of links. With networks of the topologies used here, the speed is then simply proportional to the number of hidden units.

6.6 Comparison between a Conventional and a Neural Network Analysis

So far it was shown what kind of networks were chosen, how many hidden units were used and how the nets were trained. In the following, a comparison between the "conventional" cuts from the proposal and neural network results is presented.

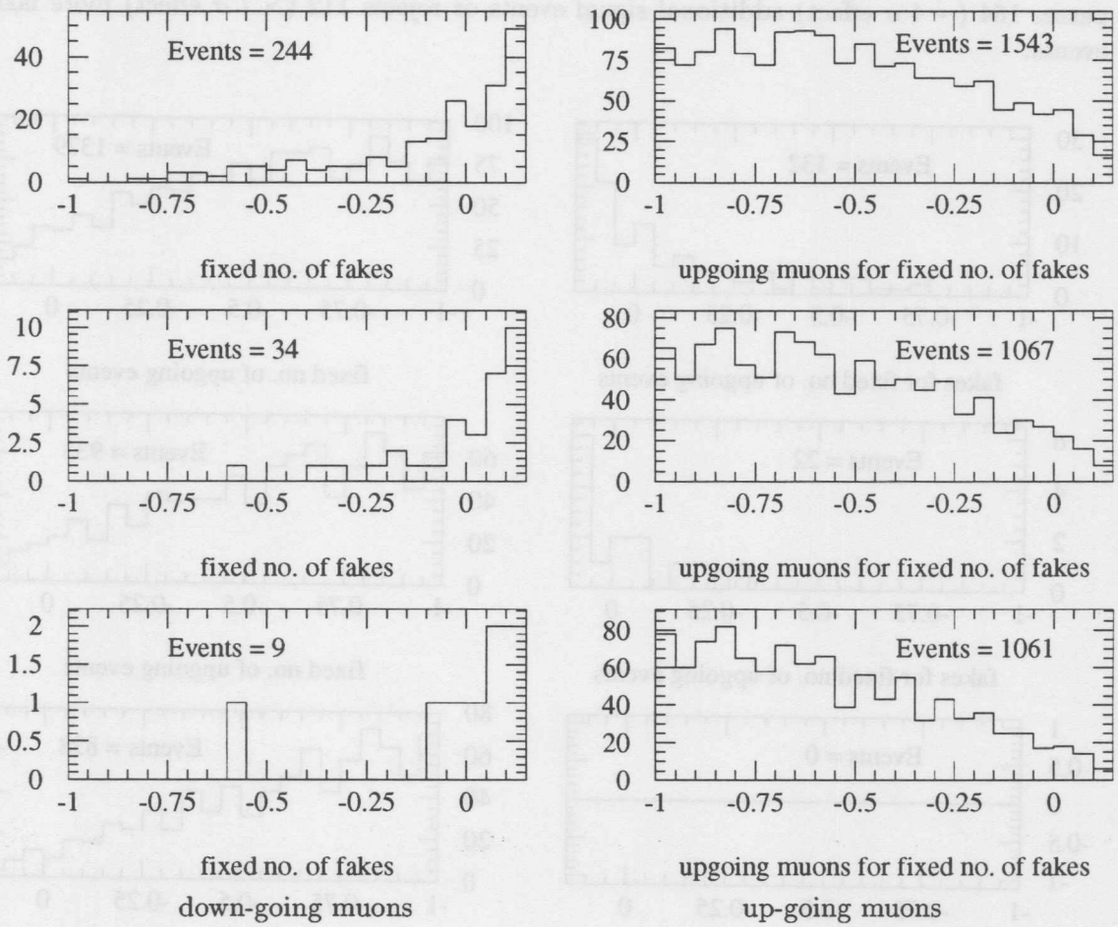


Figure 6.27: Comparison between a neural network analysis and conventional cuts; rejection adjusted to proposal. Left column: fakes; right column: up-going events;
First line: input $N_{direct,\ 15}$ and $N_{direct,\ 25}$, output $\mathcal{O} > 0.045$; second line: additional input: $\mathcal{L}$ and $\mathcal{L}/N_{ch}$, output $\mathcal{O} > 0.665$, third line: additional input:$|\vec{v}|$ and COG_Z, output $\mathcal{O} > 0.87$.

In order to do this comparison, analogous results to those given in the proposal were produced. There the rejection was improved by successively applying cuts on the quality

parameters, see figure 6.23. For this analysis, the neural network was fed with the same parameters. One single cut on its final result (output $\mathcal{O}$) is then needed. In order to compare the result with that of the conventional cuts the cut was adjusted as to reproduce the proposal's rejection rate (see figure 6.27) or the proposal's passing rate (see figure 6.28). This comparison was not only performed for the final result with all six parameters, but for the two intermediate steps from the proposal as well.

First of all it was tried to produce analogous results using information on direct hits only. Thus a 2-2-2net (2 input units, 2 hidden units, 2 output units) was fed with $N_{direct,\ 15}$ and $N_{direct,\ 25}$. The results are shown in the first lines in figures 6.27 and 6.28. Already on this level a significant improvement is achieved: Depending on the cut, the network passes 164 ($\sim 4\ \sigma$ effect) additional signal events or rejects 112 ($> 7\ \sigma$ effect) more fake events.

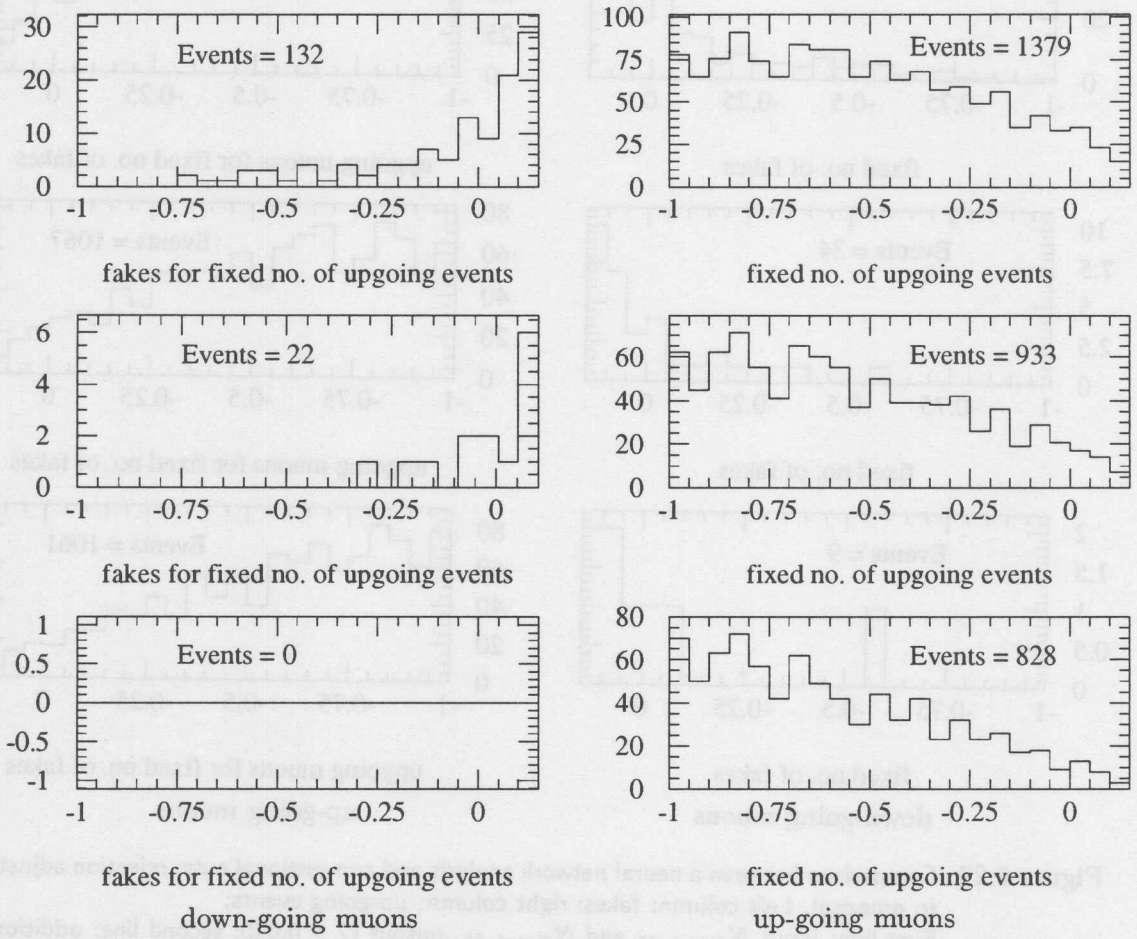


Figure 6.28: Comparison between a neural network analysis and conventional cuts; passing rate adjusted to proposal. Left column: fakes; right column: up-going events;
First line: input $N_{direct,\ 15}$ and $N_{direct,\ 25}, output > 0.3$; second line: additional input: $\mathcal{L}$ and $\mathcal{L}/N_{ch}, output > 0.773$, third line: additional input:$|\vec{v}|$ and COG_Z, output $\mathcal{O} > 0.93965$

Then additional information on the likelihood was included: A 4-4-2net was fed with $N_{direct,\,15}$, $N_{direct,\,25}$, $\mathcal{L}$ and $\mathcal{L}/N_{ch}$ information. The results are shown in the second line of figures 6.27 and 6.28. This time the improvement is not as large as before, but still significant: With the two slightly different cuts presented, one can choose between 134 more signal events ($\sim 4\ \sigma$ effect) or 12 less fake events.

Finally information on $|\vec{v}|$ and on COG_Z was included. This time a 6-6-2net was fed with information on all six parameters. Again the results were made comparable by fixing the passing rate and the rejection respectively (see last row of figures 6.27 and 6.28). This time the number of signal events can be increased by 233 ($\sim 7\ \sigma$ effect), or one might reject all 9 remaining fake events.

Cut on output	N_{fakes}	$A_{eff,down}$	$N_{acc,up}$	$A_{eff,up}$	S/N
$\mathcal{O} > 0.93965$	0	< 0.045 m^2	828	$29 \cdot 10^3$ m^2	$> 0.64 \cdot 10^6$
$\mathcal{O} > 0.87$	9	0.82 m^2	1061	$37 \cdot 10^3$ m^2	$0.045 \cdot 10^6$
Proposal	9	0.82 m^2	828	$29 \cdot 10^3$ m^2	$0.035 \cdot 10^6$

Table 6.4: Proposal results compared to results with the 6-6-2net ($\theta_{rec} > 80°$). N_{fakes} is the number of fakes accepted, $A_{eff,down}$ and $A_{eff,up}$ are the effective areas for down-going and up-going muons according to equation 6.35 and $S/N = A_{eff,up}/A_{eff,down}$ is signal to noise (for the two samples investigated)

The final result is summarized in table 6.4. The effective areas are calculated according to equations 5.26 and 5.27:

$$A_{eff} = \frac{N_{acc}}{N_{fil}} \cdot A_{eff,\,fil}\ , \tag{6.35}$$

where N_{acc} is the number of accepted events, $A_{eff,\,fil}$ is the effective area for the filtered sample (given in section 6.1) and N_{fil} is the number of events that passed the filter in the considered test sample only.

As can be seen from the plots 6.23, 6.27 and 6.28 as well as from the table 6.4, the "conventional" result could be improved on every level of input information. Using all six parameters, the passing rate could be increased by more than 25 % on the same rejection. Instead, with the same passing rate as in the proposal, one could reject all fakes (even those from the horizon). Relaxing the last cut slightly one could have passed between 835 and 861 events, if one last fake event is accepted. However, if trying to perform such a fine tuning to very high rejection rates, one reaches the statistical limits of the samples used.

6.7 Extension of the Neural Network Analysis

6.7.1 Reconstructed Zenith Angle

In order to investigate the influence of $\cos\theta_{rec}$ (the reconstructed zenith angle) as input unit, a 6-6-2net was tested versus a 7-7-2net. $\cos\theta_{rec}$ was the additional input parameter

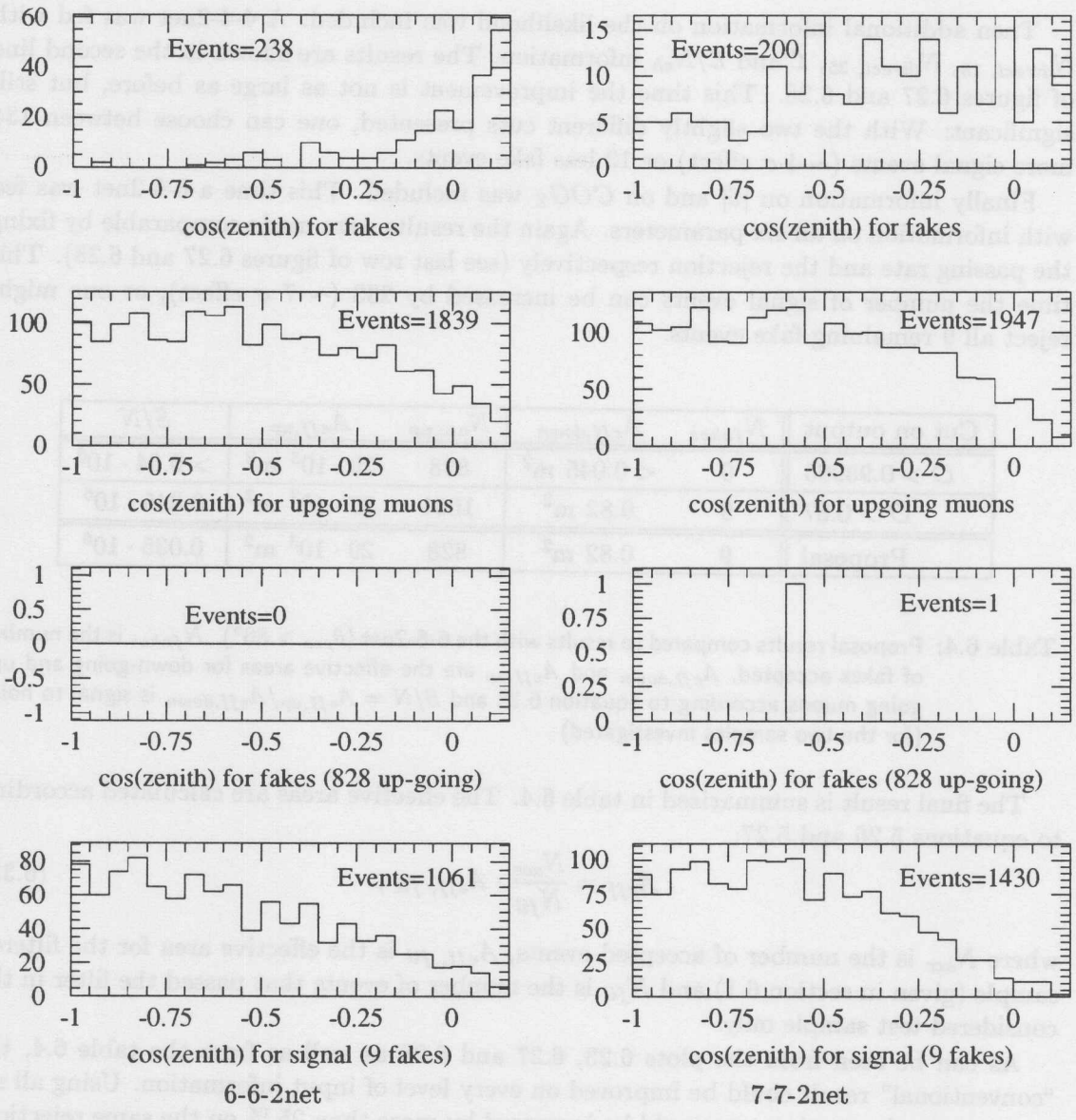


Figure 6.29: Influence of $\cos\theta_{rec}$ as input unit: Left column without $\cos\theta_{rec}$ information, right column with this information.
First line: fakes left after cut *output* > 0, second line: up-going events left after cut *output* > 0, third line: fakes left after cut leaving only 828 up-going ($\mathcal{O} > 0.93965$ and $\mathcal{O} > 0.98557$), fourth line: signal left after cut leaving only 9 fake events ($\mathcal{O} > 0.87$ and $\mathcal{O} > 0.8$)

in the second case. According to the results from section 6.5 this required a seventh hidden unit. The results are shown in figure 6.29.

The distribution of the smaller net is plotted against $\cos\theta_{rec}$ in the left column of figure 6.29, while the corresponding distribution for the bigger net is plotted in the right column. The cut criterion for the plots in the first two rows was *output* > 0. From the

plots one can see, that with the 7-7-2net there are 26 %($\equiv$ 38) less fake events than with the 6-6-2net. This difference is a 3 σ effect, so the result is significant. Furthermore one gets a higher passing rate (+108 up-going events), which is still a 2.5 σ effect. Signal to noise for the two samples is improved by a factor of 1.3. So here the "default cut" does give an unambiguous answer as to which net is performing better (contrary to the case in section 6.5).

The $\cos\theta_{rec}$ distribution of the fakes in the 7-7-2net have an interesting shape: It does not show the same peak near the horizon as the proposal analysis or the smaller net does, but rather has a flat shape. Actually the rejection of fakes is even worsened for events with $\cos\theta_{rec} \sim -1$ values (see figure 6.29 top right); but it considerably improved for events with $\cos\theta_{rec} \sim -0$. This phenomenon is not completely understood. One possible, yet heuristic explanation is based on the way the error of the network analysis is defined: Without $\cos\theta_{rec}$ information, the net can identify clear up-going events rather easy. Up-going events near the horizon however are topologically little different from fake events. As seen with the 6-6-2net one therefore expects a good fake rejection on $\cos\theta_{rec} \sim -1$ values, but a worse fake rejection at $\cos\theta_{rec} \sim 0$. Including the $\cos\theta_{rec}$ information through the additional input unit, the net optimizes its output angle-dependent. That fraction of the error due to events in the $cos\theta \sim -1$ region is smaller if it accepts some few fake events for the benefit of a large passing rate. At $cos\theta \sim 0$ on the other hand, the error is smallest when a large fraction of mis-reconstructed fake events is rejected.

In the next two rows of the figures, more stringent cuts are shown: In the third row the cut accepts 828 signal events and the remaining fake(s) are shown. One sees, that this cut is already too strong to give statistically significant results. So in the fourth row a different cut is set, which leaves 9 fake events – and the events which pass this cut are shown. This cut gives a very nice result indeed: 369 further signal events passed, a $\sim 10\ \sigma$ improvement. So information on $\cos\theta_{rec}$ is still very important for the analysis - although a "conventional" cut on it had already been performed ($\cos\theta_{rec} > 80°$) during filtering.

6.7.2 Reconstructed Track Distance

The AMANDA detector is close to being cylindrically symmetric. Inspired by this, possible dependences on the actual location of a muon track inside the detector are tested with the following "cylindrical" parameters: distance between track and detector center, z-component of this distance and horizontal component of it.

The effect of these new parameters is shown in figure 6.30. In the left column the results from the already introduced 7-7-2net are shown, in the right column, those obtained with the new 10-10-2net. Again requiring the output to be bigger than zero (first two rows), the rejection factor was virtually indifferent against the additional parameters and remained at 98 per cent. The passing rate improved slightly: from 76 to 80 per cent (2.5 σ). The overall shape of the distributions are similar for both nets. In the third line, the passing rate is adjusted to 828 events and both nets reject all but one fake event. In the last line, demanding a less stringent cut, which leaves 9 fake events ($\mathcal{O} > 0.98557$ for the smaller net; $\mathcal{O} > 0.99053$ for the bigger net), no benefit is achieved by the new parameters any more. Actually the performance of the larger net is marginally worse (less than 1 σ difference). This indicates, that the additional information on track distance is almost obsolete for

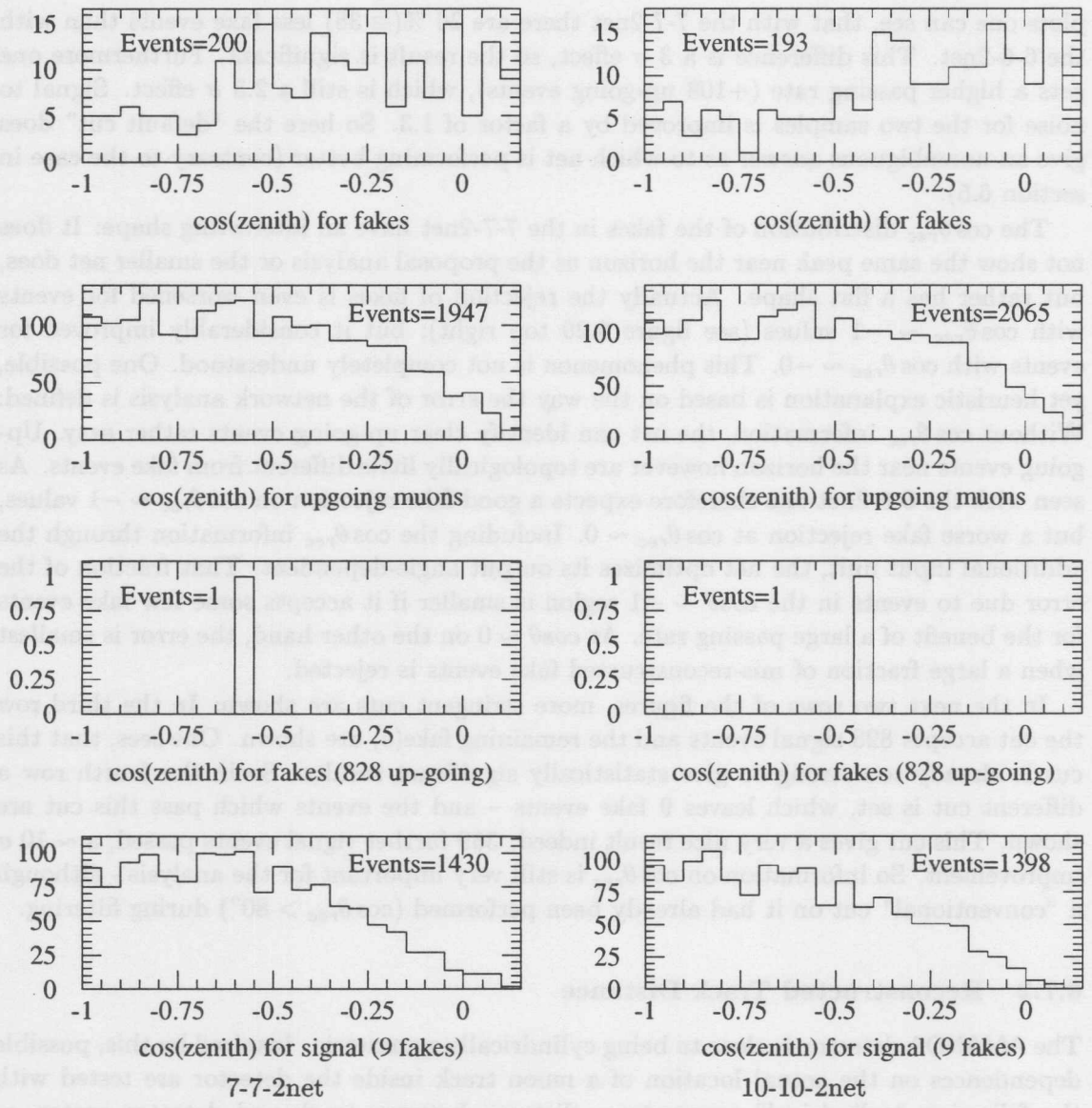


Figure 6.30: Results obtained by adding information on the reconstructed track distance: R_{rec}, ρ_{rec} and Z_{rec}. Left column: without reconstructed distance information, right column: with that information.
First line: remaining fakes versus $cos\theta_{rec}$ after an $output > 0$ cut, second line: remaining up-going events versus $cos\theta_{rec}$ after an $output > 0.9$ cut, third line: remaining fakes after a cut to leave 12 fakes, fourth line: remaining up-going events after a cut to leave 12 fakes.

this analysis. However, if the rejection is fixed as to leave less than 5 fakes, the bigger net performs slightly better than the smaller net again. So a decision which is the better net, cannot be finally drawn with the limited statistics available. The additional CPU time required for the bigger network might be an argument for the smaller network.

Cut on output	N_{fakes}	$A_{eff,down}$	$N_{acc,up}$	$A_{eff,up}$	S/N
$\mathcal{O} > 0.9613$	1	0.091 m^2	1202	$42 \cdot 10^3$ m^2	$0.47 \cdot 10^6$
$\mathcal{O} > 0.9222$	9	0.82 m^2	1398	$49 \cdot 10^3$ m^2	$0.060 \cdot 10^6$
$\mathcal{O} > 0.9$	13	1.2 m^2	1470	$52 \cdot 10^3$ m^2	$0.043 \cdot 10^6$
$\mathcal{O} > 0.5$	64	5.8 m^2	1880	$66 \cdot 10^3$ m^2	$0.011 \cdot 10^6$
$\mathcal{O} > 0$	235	21 m^2	1979	$70 \cdot 10^3$ m^2	$0.0033 \cdot 10^6$
Proposal	9	0.82 m^2	828	$29 \cdot 10^3$ m^2	$0.035 \cdot 10^6$

Table 6.5: Proposal results compared to results with the 10-10-2net ($\theta_{rec} > 80°$). N_{fakes} is the number of fakes accepted, $A_{eff,down}$ is the effective area for down-going muons according to equation 5.27, $A_{eff,up}$ is the effective area for up-going events according to equation 5.26 and $S/N = A_{eff,up}/A_{eff,down}$ is signal to noise for the two samples.

Table 6.5 summarizes the effects of the final 10-10-2net in comparison to the results obtained in the proposal. The strong correlation between performance of the neural net (measured via the effective areas) and the cut on the output is obvious. Compared to the analysis presented in the Zeuthen AMANDA-II proposal, an almost 50 % increase of signal passing rate could be achieved while background rejection was still improved considerably. Alternatively, with a less stringent cut, a 70 % increase in the effective area for signal muons can be achieved for the same background rejection.

7 AMANDA-B4 Analysis

7.1 Data Samples

The Monte Carlo background sample used for the AMANDA-B4 analysis was generated similar to the one used in the previous chapter. This time $8.1 \cdot 10^5$ of the generated atmospheric muons triggered the detector simulation. The effective area for this background sample was determined to be $6.6 \cdot 10^3$ m^2. This sample was reconstructed and subsequently divided into two equally sized subsamples – one for training the networks and one for testing.

The signal sample however was produced slightly different from the previous signal sample: This time 100 GeV, 1 TeV and 10 TeV muons (rather than just 1 TeV muons) were generated inside a volume (rather than on a generation plane). The dimensions of the generation volumes were larger than the maximal distance muons of the generated energy can travel. So, due to energy losses, muons reaching the detector had energies distributed between 0 and the generation energy. However no simulation of the atmospheric neutrino spectrum was performed. The effective trigger areas were determined to be $3.1 \cdot 10^3$ m^2, $7.2 \cdot 10^3$ m^2 and $1.6 \cdot 10^4$ m^2 respectively. Again the simulated samples were divided into a teaching and a testing sample.

The measured data used in this chapter consists of $1.1 \cdot 10^6$ triggered events, which were gathered with the AMANDA-B4 detector on June 23rd 1996. The day was chosen since a Swedish analysis had already found a neutrino candidate among the data measured on that day [17]. Apart from the standard AMANDA-B4 trigger (at least eight hits within 2μs), additional external triggers can initiate an event readout. The reconstruction performed prior to this analysis used all events regardless of their trigger. Since external triggers are responsible for $\lesssim 10$ % of all triggers, this introduced a small bias only. As no teaching is possible with measured data, the full sample was analyzed with the trained networks.

7.2 Initial Filtering

Some initial filtering had to be performed in order to reject down-going muons. The filters used for the AMANDA-B4 analysis were the following:

1. At least eight hits were required. This mimicked the AMANDA-B4 trigger. It is not equivalent since this filter was set after the reconstruction had rejected likely noise hits. Therefore it is slightly stronger than the AMANDA-B4 trigger. The aim of this filter was to correct for the different trigger conditions between measured data and MC simulation.

2. The reconstruction error had to be less than 15° for the simulated signal. Bad reconstructed events are unlikely to yield unambiguous signals, so they should not be used to train the neural network. This cut was not (could not be) applied to MC background (measured data).

3. The zenith angle of the line-fit had to be bigger than 120°.

4. The reconstructed zenith angle had to be bigger than 120°. This and the previous filter determine what is meant by the term "up-going" in this chapter: events with zenith angles larger than 120°.

5. Three hits with a time residual of less than 25 ns were demanded. Without some direct hits it is highly unlikely to get good reconstruction results.

6. The normal reconstruction started with the track from the initial line fit. A second reconstruction was also performed, which started from a vertical down-going track. If the zenith angle from this second reconstruction was less than 120° and its likelihood was better (i.e. smaller) than that of the normal reconstruction, the event was rejected.

Filter	Data (exp.)	MC background	MC signal 100 GeV	MC signal 1 TeV	MC signal 10 TeV
triggered events	$1.4 \cdot 10^6$	$8.1 \cdot 10^5$	$1.2 \cdot 10^4$	$1.1 \cdot 10^4$	$1.2 \cdot 10^4$
$A_{eff,\,trig}$ [m^2]	—	$6.6 \cdot 10^3$	$3.1 \cdot 10^3$	$7.2 \cdot 10^3$	$1.6 \cdot 10^4$
events with $N_{ch} > 7$	$1.1 \cdot 10^6$	$4.6 \cdot 10^5$	$9.4 \cdot 10^3$	$9.6 \cdot 10^3$	$9.7 \cdot 10^3$
$\lvert\theta_{rec} - \theta_{true}\rvert < 15°$	—	—	33%	31%	21%
$\theta_{line-fit} > 120°$	7.4%	12%	23%	18%	11%
$\theta_{rec} > 120°$	6.3%	9.2%	22%	16%	9.8%
$N_{direct,\,25ns} > 2$	1.1%	1.2%	16%	12%	6.0%
$\mathcal{L}_{down} > \mathcal{L}_{up}$	0.42%	0.42%	14%	12%	5.8%
passing events	4523	2872	1470	1195	667
$A_{eff,\,filter}$ [m^2]	—	28	$1.5 \cdot 10^3$	$3.1 \cdot 10^3$	$3.6 \cdot 10^3$

Table 7.6: Filters used between reconstruction and neural network analysis for the AMANDA-B4 simulation. Energies given are at generation points inside volume.
First filter: at least eight hits within the event; second filter (for MC signal only): reconstruction error smaller than 15 degrees; third filter: angle reconstructed with initial line-fit must be up-going; fourth filter: finally reconstructed angle must be up-going; fifth filter: At least three hits with a time residual of less than 25 ns; last filter: The up-going track must be more probable than a potential down-going alternative.

The filters and their respective passing rates are summarized in table 7.6. The effective areas in the last row were calculated according to equations 6.31 (signal samples) and 5.26 (background sample).

7.3 Parameter Transformation

Again three simple transformations $x \rightarrow x'$ were used to transform the parameter distributions to the range [0..1]:

- Since this time $\cos\theta$ was restricted to values between -1 and -0.5 (after the cut on the reconstructed zenith angle), its transformation was:

$$\cos(\theta') = 2 \cdot \cos(\theta) + 2 \; . \tag{7.36}$$

- For parameters x, which were restricted to positive values (like distances):

$$x' = \frac{2}{\pi} \cdot \arctan\left(\frac{x}{\overline{x}}\right) , \tag{7.37}$$

where $\overline{x}$ is the mean of the distribution of x.

- For parameters x, which could have positive and negative values (like coordinates), the means $\overline{x}_+$ and $\overline{x}_-$ were calculated separately for each of the two parts of the distribution. Then:

$$x' = \frac{1}{\pi} \cdot \arctan\left(\frac{x}{\overline{x}_+ + |\overline{x}_-|}\right) + 0.5 . \tag{7.38}$$

The means were calculated from the distributions of measured data and used for the transformation of data and Monte Carlo samples. The distributions of all parameters used for the networks of this chapter can be seen in the figures of section 5.3. Compared to the transformations of the previous chapter, the main difference is that this time the division was by the mean of the parameter – not by twice the mean. There were two main reasons for this:

1. Figure 6.22, the example of the transformation from the previous chapter, is typical in that sense that the transformed distributions there did not use the whole range [0..1]. The tails often almost vanished for values bigger than 0.6 – 0.8.

2. For more than half of the parameters investigated, signal events tended to have smaller values than mis-reconstructed atmospheric muons. This interesting range of small values could be expanded by using smaller numbers for the division. One might have thought of using two different transformations: One with a division by twice the mean for the minority of parameters which had signal accumulated at large values, and a second one (like the transformation used in this chapter) for parameters, which had signal accumulated at small values. This was discarded for the benefit of a common transformation formula for all parameters.

As a last subtlety, slight deviations between MC and Data were corrected for: In the case of the three parameters velocity of line-fit $|\vec{v}|$, average distance of direct hits $\overline{d}_{dir}$, and likelihood of the hit topology $\mathcal{L}_{hit}$, MC background was systematically shifted compared to data. This shift was corrected for by multiplying the MC distributions of these parameters with the ratio $\overline{x}_{data}/\overline{x}_{MC\ background}$ prior to the transformation. As can be seen in figures 5.17 and 5.21, where the corrections are included in the bottom line only, this improved the consistency between data and MC background. More complicated deviations that occurred between data and MC for other parameters could not easily be corrected for and remained.

7.4 Basic Network

The analysis presented in this chapter was performed after the analysis of the previous chapter. Therefore use was made of two results obtained there. The first result was that the network performance can be improved when all available statistics are used for

training – whether the resulting training sample is asymmetric or not. In the case of the AMANDA-II analysis, this resulted in a training sample which consisted of almost four times as many background as signal events. In this chapter's analysis, the mismatch was smaller and reversed (with 16 % more signal events). The second result from which this analysis benefited was that the number of hidden units in the network topology should be the same as the number of input units – therefore only "n-n-2nets" were used. The input parameters that were chosen for a first network are already known from the networks in chapter 6. They were:

- number of direct hits within 15 ns, $N_{direct,\ 15ns}$,
- number of direct hits between 15 and 25 ns, $N_{direct,\ 15-25ns}$ (This was used instead of $N_{direct,\ 25ns}$ as it is independent of the already used $N_{direct,\ 15ns}$),
- likelihood of the reconstruction, $\mathcal{L}$,
- likelihood per channel, $\mathcal{L}/N_{ch}$,
- velocity of the first guess line-fit, $|\vec{v}|$ and
- z-component of the center of gravity, COG_Z.

The results of the trained "6-6-2net" are presented in figure 7.31. In the first row, the output distribution of the network is given for data (left column; points), background (left column; line) and signal (right column). It can be seen that MC background and data agree well within the available statistics. As expected for a successful analysis, both decrease from $\mathcal{O} \sim -1$ to $\mathcal{O} \sim 1$. However there seems to be a slight but significant peak in the data for values at $\mathcal{O} \sim 1$, which is not as strongly developed in the MC background sample. The background sample on the other hand has a stronger peak at $\mathcal{O} \sim -1$. The signal distribution is flatter, but with a prominent (expected) peak at $\mathcal{O} \sim 1$.

In the second row, the cosine of the reconstructed zenith angle $\cos\theta_{rec}$ is shown for the full testing samples. As pointed out in section 5.3.4, the slight rise of the data and MC background distribution to negative values is due to a reconstruction feature.

This distribution can be compared to the same distribution after a cut $\mathcal{O} > 0.5$ on the output (in the third line). As can be seen, data and MC still agree within the available statistics. The cut gives a signal passing rate of 78 % and a background rejection of 95 %. Data is rejected slightly less – by 92 %. The reason for this difference could already be seen in the first row: Compared to MC background, there is a higher ratio of data events at large output values. Accordingly an even more stringent cut at $\mathcal{O} = 0.95$ (bottom row) again gives a higher rejection for MC background than for data (another 88 % rather than another 80 %). Only approximately further 20 % of signal events are rejected by this more stringent cut.

The passing rates of the trained "6-6-2net" are summarized in table 7.7. The effective areas are calculated according to equation 6.35 (background) and 6.31 (signal samples). Signal to noise (S/N) is given by the ratios of the effective areas of the specific signal sample to that of the background sample. One has to remember that the signal samples used here do not represent the atmospheric neutrino flux but hypothetical monoenergetic neutrino

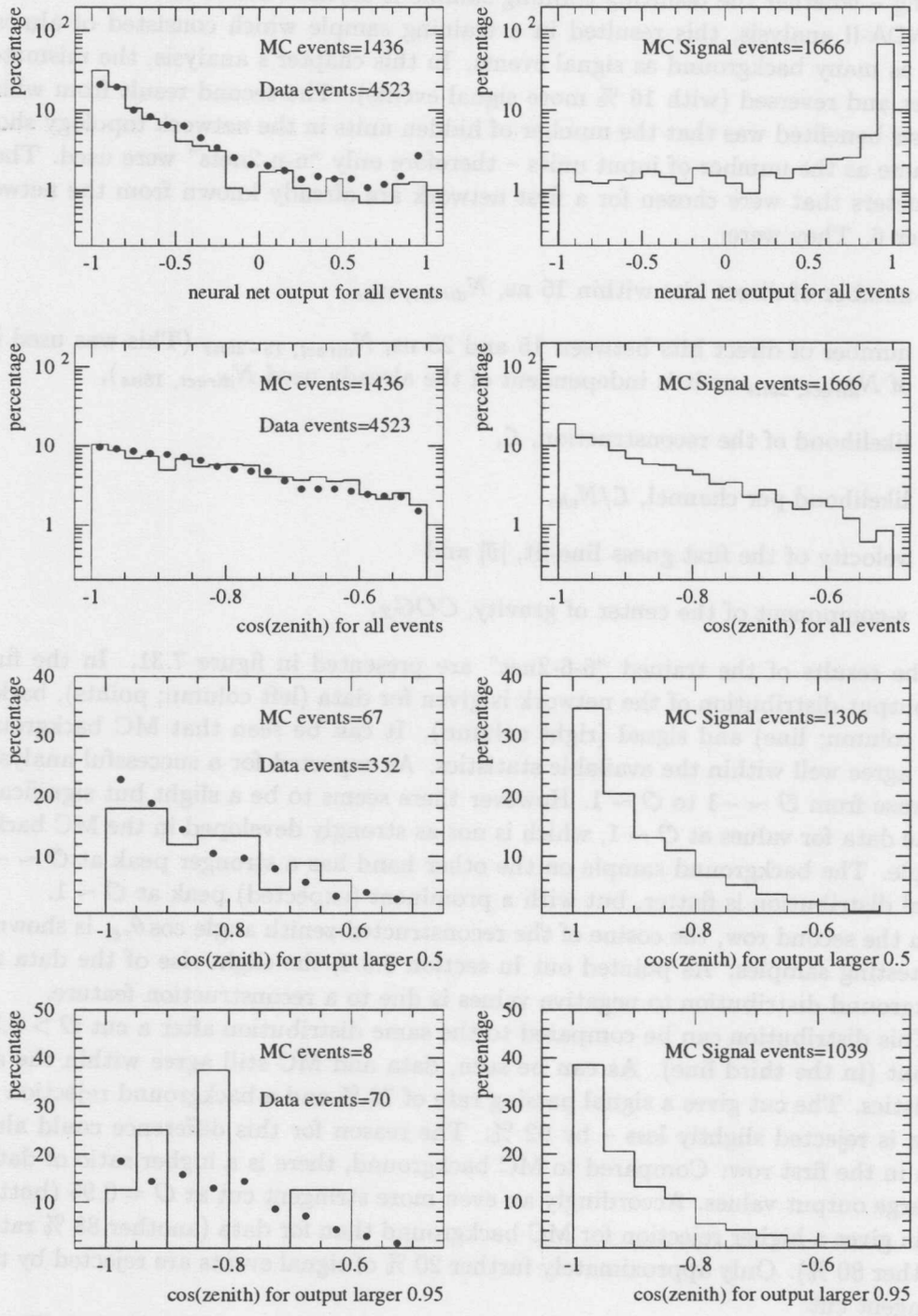


Figure 7.31: Result of the basic network: Left column: MC background (line) and data (points); right column: MC signal. First line: Distribution of neural network output $\mathcal{O}$ for all events. Second line: Distribution of $\cos\theta_{rec}$ for all events. Third line: $\cos\theta_{rec}$ distribution after a cut $\mathcal{O} > 0.5$. Fourth line: $\cos\theta_{rec}$ distribution after a cut $\mathcal{O} > 0.95$.

Cut value	Data	MC background	MC signal 100 GeV	MC signal 1 TeV	MC signal 10 TeV
all events	4523	1436	736	597	333
$A_{eff,\ filter}$ [m^2]	—	28	$1.5 \cdot 10^3$	$3.1 \cdot 10^3$	$3.5 \cdot 10^3$
$\mathcal{O} > 0$	711 (16%)	163 (11%)	581 (79%)	481 (81%)	280 (84%)
$\mathcal{O} > 0.5$	352 (7.8%)	67 (4.7%)	521 (71%)	444 (74%)	247 (74%)
$\mathcal{O} > 0.95$	70 (1.5%)	8 (0.56%)	393 (53%)	340 (57%)	178 (53%)
$A_{eff,\ filter}$ [m^2]	—	0.13	$8.1 \cdot 10^2$	$1.7 \cdot 10^3$	$1.9 \cdot 10^3$
S/N	—	—	$6.2 \cdot 10^3$	$1.4 \cdot 10^4$	$1.5 \cdot 10^4$
$\mathcal{O} \geq 0.99752$	13 (0.29%)	1/2 (0.03%)	259 (35%)	196 (33%)	73 (22%)
$A_{eff,\ filter}$ [m^2]	—	$8.1 \cdot 10^{-3}$	$5.4 \cdot 10^2$	$1.0 \cdot 10^3$	$7.8 \cdot 10^2$
S/N	—	—	$6.6 \cdot 10^4$	$1.3 \cdot 10^5$	$9.6 \cdot 10^4$

Table 7.7: Performance of the smallest network in the AMANDA-B4 analysis. First two lines: events entering the control sample and their respective effective areas. Next two lines: number of events passing cuts $\mathcal{O} > 0$ and $\mathcal{O} > 0.5$ respectively. Next three lines: Events passing cut $\mathcal{O} > 0.95$, the corresponding effective areas and the signal to noise ratios calculated as the ratio of the effective areas. Last three lines: Events after cut $\mathcal{O} > 0.99752$ (the output of the last fake event), the corresponding effective areas and the signal to noise ratios.

sources. Furthermore the passing rates of very stringent cuts on the neural network output differs for data and MC. The difference is too large to be explained solely on basis of a few neutrino events in the data sample. Therefore the signal to noise ratios given for the MC simulated events cannot be used to determine signal to noise ratios for the remaining data events. They can only give an estimation of the order of magnitude of the actual signal to noise ratios after such cuts.

7.5 Nine Input Parameters

After the results from the basic network proved that the method of a neural network analysis seems to be applicable to the AMANDA-B4 detector, further input parameters were included. They were:

- the likelihood of the hit topology $\mathcal{L}_{hit}$,
- the average distance of the direct hits to the muon track $\overline{d}_{dir}$ and
- the length of the direct hits $\mathcal{Z}_{leng,\ dir}$.

This time a "9-9-2net" was trained; the results are shown in figure 7.32. In the first row the output distribution is presented. Compared to the previous network, the shape of the distributions are similar. This time however the bin at $\mathcal{O} \sim -1$ is more prominent in the distribution for data and MC background. Furthermore, data and MC background agree better. This is especially true for the extreme values at $\mathcal{O} \sim \pm 1$. In the second row, the $\cos\theta_{rec}$ distribution is reproduced for comparison with the next two lines.

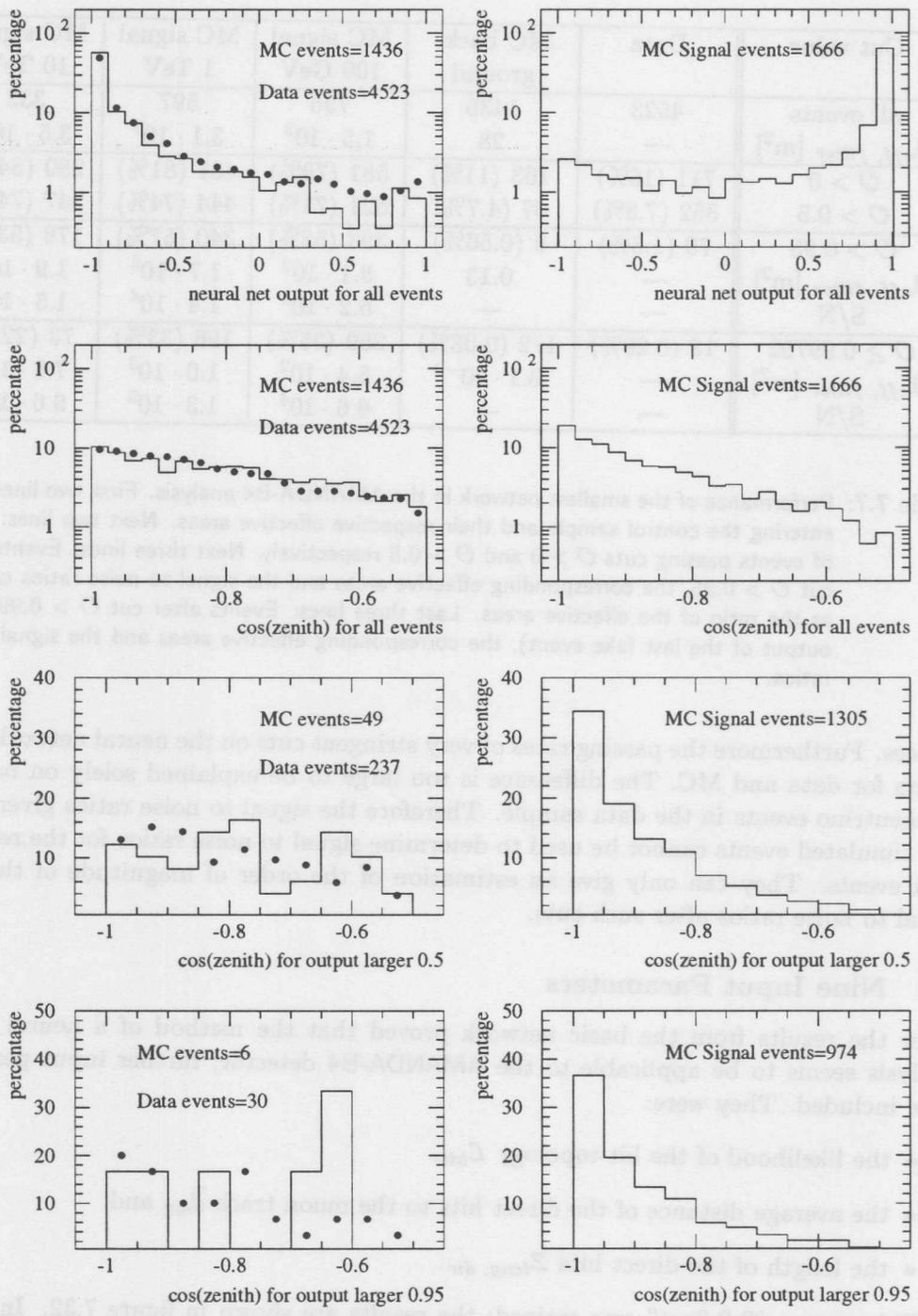


Figure 7.32: Result of the extended network: Left column: MC background (line) and data (points); right column: MC signal. First line: Distribution of neural network output $\mathcal{O}$ for all events. Second line: Distribution of $\cos\theta_{rec}$ for all events. Third line: $\cos\theta_{rec}$ distribution after a cut $\mathcal{O} > 0.5$. Fourth line: $\cos\theta_{rec}$ distribution after a cut $\mathcal{O} > 0.95$.

In the third line again a cut of $\mathcal{O} > 0.5$ is set on the output. The signal passing rate is virtually unchanged (compared to the previous network), namely 78 %, but background rejection improved to 97 % (from 95 %). Data is also rejected at a higher rate (95 % rather than 92 %). The feature of a stronger background than data rejection remains. However, the ratio of passing rates between data and MC background improved from 1.7 to 1.5.

In the bottom line of figure 7.32, the cut is set at $\mathcal{O} > 0.95$. This gives another factor of eight in background (and data) rejection, while it reduces signal by 25 % only. This time the cut does reject the same ratio of data and MC background events.

Cut value	Data	MC background	MC signal 100 GeV	MC signal 1 TeV	MC signal 10 TeV
all events	4521	1436	736	597	333
$A_{eff,\ filter}$ [m^2]	—	28	$1.5 \cdot 10^3$	$3.1 \cdot 10^3$	$3.6 \cdot 10^3$
$\mathcal{O} > 0$	554 (12.3%)	122 (8.5%)	611 (83%)	521 (87%)	282 (85%)
$\mathcal{O} > 0.5$	237 (5.2%)	49 (3.4%)	558 (76%)	477 (80%)	260 (78%)
$\mathcal{O} > 0.95$	30 (0.66%)	6 (0.42%)	393 (53%)	344 (58%)	198 (59%)
$A_{eff,\ filter}$ [m^2]	—	$9.8 \cdot 10^{-2}$	$8.1 \cdot 10^2$	$1.8 \cdot 10^3$	$2.1 \cdot 10^3$
S/N	—	—	$8.3 \cdot 10^3$	$1.8 \cdot 10^4$	$2.1 \cdot 10^4$
$\mathcal{O} \geq 0.98848$	12 (0.27%)	1/2 (0.03%)	292 (40%)	274 (46%)	152 (47%)
$A_{eff,\ filter}$ [m^2]	—	$8.1 \cdot 10^{-3}$	$6.0 \cdot 10^2$	$1.4 \cdot 10^3$	$1.6 \cdot 10^3$
S/N	—	—	$7.4 \cdot 10^4$	$1.8 \cdot 10^5$	$2.0 \cdot 10^5$

Table 7.8: Performance of the extended network in the AMANDA-B4 analysis. First two lines: events entering the test sample and their effective areas. Next two lines: number of events passing cuts $\mathcal{O} > 0$ and $\mathcal{O} > 0.5$ respectively. Next three lines: Events after cut $\mathcal{O} > 0.95$, the corresponding effective areas and the signal to noise ratios calculated as the ratio of the effective areas. Last three lines: Events passing cut $\mathcal{O} > 0.98848$ (the output of the last fake event), the corresponding effective areas and the signal to noise ratios.

The passing rates for various cuts on the output of this network are summarized in table 7.8. The effective areas are calculated from equations 6.31 (signal samples) and 6.35 (background sample). Again caution is needed when interpreting the S/N ratios given: These are no signal to noise ratios for the actual experiment, but only for the MC samples investigated.

7.6 Large Network

In a final analysis, four further input parameters were added. They were:

- The cosine of the reconstructed zenith angle $\cos\theta_{rec}$,
- the reconstructed track distance $\mathcal{R}_{rec}$,
- the horizontal component of the the center of gravity $COG_{X,Y}$ and
- the z-component of the center of gravity of direct hits $COG_{Z,\ dir}$.

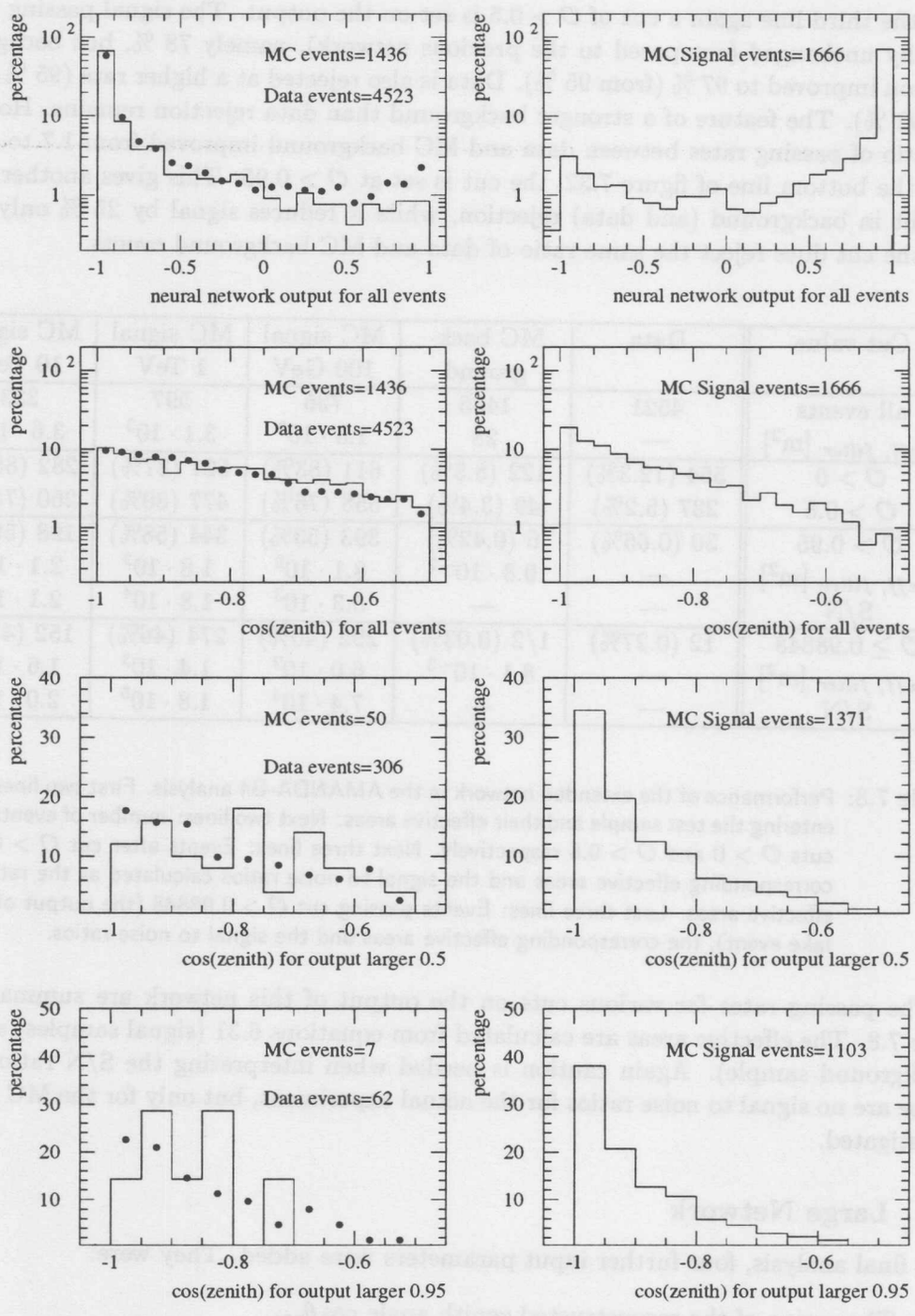


Figure 7.33: Result of the largest network: Left column: MC background (line) and data (points); right column: MC signal. First line: Distribution of neural network output $\mathcal{O}$ for all events. Second line: Distribution of $\cos\theta_{rec}$ for all events. Third line: $\cos\theta_{rec}$ distribution after a cut $\mathcal{O} > 0.5$. Fourth line: $\cos\theta_{rec}$ distribution after a cut $\mathcal{O} > 0.95$.

So a "13-13-2net" was trained and tested. The results are displayed in figure 7.33. In the first row, the output distributions for the full samples are shown. The peaks in the extreme bins at $\mathcal{O} \sim +1$ (signal) and $\mathcal{O} \sim -1$ (data and background) contain approximately seventy (data sixty) per cent of all events. Compared to the previous network, one notices that the little peak for data at $\mathcal{O} \sim +1$ has reappeared.

In the second row, the well known $\cos\theta_{rec}$ distribution is again reproduced for comparison. In the third row, the required cut $\mathcal{O} > 0.5$ is as stringent for MC background as the equivalent cut on the output of the previous network: It rejects 97 %. However it is less stringent for data (rejection of 93 % rather than 95 %) and MC signal (passing rate 82 % rather than 78 %). This proves again that cuts on the same output for different networks cannot strictly be compared.

In the last row, the cut $\mathcal{O} > 0.95$ confirms this: all three data samples get a higher passing rate than in the case of the previous network. This time 66 % of the data, 0.49 % of the background and 1.4 % of the data pass the filter. So as expected from the plot in the first line, the deviations between data and MC background become significant on such stringent cuts. Before these deviations are fully understood, results obtained from MC simulations (like signal to noise ratios) cannot be used to make accurate quantitative statements for data.

Cut value	Data	MC background	MC signal 100 GeV	MC signal 1 TeV	MC signal 10 TeV
all events	4521	1436	736	597	333
$A_{eff,\ filter}$ [m^2]	—	28	$1.5 \cdot 10^3$	$3.1 \cdot 10^3$	$3.6 \cdot 10^3$
$\mathcal{O} > 0$	578 (13%)	109 (7.6%)	635 (86%)	517 (87%)	298 (89%)
$\mathcal{O} > 0.5$	306 (6.8%)	50 (3.5%)	601 (82%)	498 (83%)	275 (83%)
$\mathcal{O} > 0.95$	62 (1.4%)	7 (0.49%)	494 (67%)	407 (68%)	227 (68%)
$A_{eff,\ filter}$ [m^2]	—	0.11	$1.0 \cdot 10^3$	$2.1 \cdot 10^3$	$2.4 \cdot 10^3$
S/N	—	—	$8.9 \cdot 10^3$	$1.9 \cdot 10^4$	$2.1 \cdot 10^4$
$\mathcal{O} \geq 0.99972$	6 (0.13%)	1/2 (0.03%)	277 (38%)	250 (42%)	128 (38%)
$A_{eff,\ filter}$ [m^2]	—	$8.1 \cdot 10^{-3}$	$5.7 \cdot 10^2$	$1.3 \cdot 10^3$	$1.4 \cdot 10^3$
S/N	—	—	$7.0 \cdot 10^4$	$1.6 \cdot 10^5$	$1.7 \cdot 10^5$

Table 7.9: Performance of the largest network in the AMANDA-B4 analysis. First two lines: events entering the control sample and their respective effective areas. Next two lines: number of events passing cuts $\mathcal{O} > 0$ and $\mathcal{O} > 0.5$ respectively. Next three lines: Events after cut $\mathcal{O} > 0.95$, the corresponding effective areas and the signal to noise ratios calculated as the ratio of the effective areas. Last three lines: Events passing cut $\mathcal{O} > 0.99972$ (the output of the last fake event), the corresponding effective areas and the signal to noise ratios.

The results from the analysis of this network are summarized in table 7.9. As stated before, the calculated signal to noise ratios are only for the signal and background samples used. As the atmospheric neutrinos were not simulated (they have different energy spectra from the MC signal used here) and due to the deviations between data and MC background, it is not valid to transfer these S/N ratios directly to data.

Comparing the second, third and fourth plot in the left column, for the results of all three networks, one notices that the ratio of events with reconstructed zenith angles close to 180° increases. With the stringent cuts required to get a high signal to noise ratio, almost no events remain at values close to $\cos\theta_{rec} \sim -0.5$. This justifies the filter condition $\cos\theta_{rec} < -0.5$, which had been chosen at the beginning of this analysis.

7.7 Comparison of the "Best" Data Events

Successively setting more and more stringent cuts on the output of the three neural networks until only very few data events remain, one notices an interesting result: The two last data events (i.e. the two most likely neutrino candidates) are always the same. This is especially impressive, as among the next three most likely data events, there is only one event commonly identified by two networks: The third most likely event according to the medium sized network was considered to be the fourth most likely candidate by the largest network.

One of the two neutrino candidates found is just the neutrino candidate which had been found by the earlier mentioned Swedish study. This is a very satisfying result, since this analysis was completely independent from the Swedish study: That study started with a different reconstruction algorithm and then applied very stringent "conventional" cuts on their reconstructed zenith angle, their number of direct hits and their z-component of the velocity of the line-fit [17]. The two neutrino candidates found with the analysis presented here, are shown in figures 7.34 and 7.35.

	small network	medium network	large network
output of first ν candidate	$\mathcal{O} = 0.99994$	$\mathcal{O} = 0.99988$	$\mathcal{O} = 1.0$
remaining 1 TeV signal events	107	122	147
$A_{eff,\ up}$ [m^2]	$5.6 \cdot 10^2$	$6.4 \cdot 10^2$	$7.7 \cdot 10^2$
S/N = $A_{eff,\ up}/A_{eff,\ down}$	$> 6.9 \cdot 10^4$	$> 7.8 \cdot 10^4$	$> 9.4 \cdot 10^4$
output of second ν candidate	$\mathcal{O} = 1.0$	$\mathcal{O} = 0.99953$	$\mathcal{O} = 1.0$
remaining 1 TeV signal events	27	151	147
$A_{eff,\ up}$ [m^2]	$1.4 \cdot 10^2$	$7.9 \cdot 10^2$	$7.7 \cdot 10^2$
S/N = $A_{eff,\ up}/A_{eff,\ down}$	$> 1.7 \cdot 10^4$	$> 9.7 \cdot 10^4$	$> 9.4 \cdot 10^4$
output of third best data events	$\mathcal{O} = 0.99992$	$\mathcal{O} = 0.99842$	$\mathcal{O} = 0.99996$
remaining 1 TeV signal events	116	190	176
$A_{eff,\ up}$ [m^2]	$6.1 \cdot 10^2$	$9.9 \cdot 10^2$	$9.2 \cdot 10^2$
S/N = $A_{eff,\ up}/A_{eff,\ down}$	$> 7.5 \cdot 10^4$	$> 1.2 \cdot 10^5$	$> 1.1 \cdot 10^5$

Table 7.10: Comparison of the best data events (according to this analysis). First column: Results from the 6-6-2net, second column: results from the 9-9-2net, last column: results from the 13-13-2net. First block of data: Results for the "first neutrino candidate", the candidate which was known from a Swedish study. Second block of data: Results for the "second neutrino candidate", the candidate which was newly found with this analysis. Third block of data: Results for the third most likely neutrino candidate. Only the results for the 1 TeV signal sample entered this table. Effective area for background: $8.1 \cdot 10^{-3}$ m^2.

One can compare the number of two ν-candidate events found within one day to expectations: AMANDA-B4 is expected to trigger $\sim$ 2000 atmospheric neutrino events per year [37]. The efficiency of the filters for the true energy spectrum is not known. Using the passing rates for the 1 TeV MC signal sample as an approximation, one would expect 0.6 neutrino events per day. However the day investigated was chosen as the Swedish study had already found a very likely neutrino candidate (the Swedish cuts were so stringent, that only two candidates in several months worth of data were found), so for this day the expectations are approximately 1.6 neutrino events.

As mentioned before, the true signal to noise ratio for these candidates can not be calculated from the present analysis. Table 7.10 nevertheless lists the output values for the two candidates and the third most likely candidates (which is a different event for each of the three networks) together with lower limits for MC signal to noise ratios. The candidate named "first ν candidate" is the one known from the Swedish study. As no MC background events remained after these stringent cuts, the effective area for background events was set at $8.1 \cdot 10^{-3}$. This is equivalent to half an event remaining. Since the expressiveness of these numbers is limited, only one signal sample (1 TeV) was used for calculating effective areas and S/N ratios.

Comparing the required cut values on the outputs of the neural networks in this chapter to the ones in the previous chapter, one notices that cuts very close to $\mathcal{O} = 1$ are needed for the AMANDA-B4 simulation. This indicates that the analysis of this chapter is not only limited by the available MC statistics, but also by the detector itself. AMANDA-II can already be considered to be a neutrino detector, whereas AMANDA-B4 was only one step in that direction. Clear neutrino candidates are expected however already from the AMANDA-B10 detector.

7.8 Event Display

On the following pages, a selection of data and MC events are shown. The displayed events can be explained as follows: The detector (strings are blue) is seen from the side. The little coordinate system indicated near the center of the plots defines the origin of the detector. The reconstructed muon track is shown in red. If a Monte Carlo event is displayed, the true generated muon directions are indicated in yellow. There can be more than one generated muon within an event. E.g. in figure 7.40 two coincident muons from one air shower reached the detector. In some cases the generated muon stops inside or close to the detector. An example of this can be seen in figure 7.38. The black circles (ellipses in the projection used) indicate the hit PMTs: The radius is proportional to the amplitude of the hit. The green lines finally are the assumed photon tracks from muon to PMT. They do not end at the PMT. The length of that part of the photon track which goes beyond the PMT is proportional to the calculated time residual of the track. Neglecting the PMT jitter, one can establish the total length of the green lines as the total distance traveled by the (in general scattered) photon.

The events were chosen as to represent a large variety of results: First, the two neutrino candidates found are presented. As contrast, one data event which was considered as background by all three networks is shown. The next event is a data event with output values above 0.98, but still in a region where the majority of data events is supposed

to be background. These four data events are presented in figures 7.34 to 7.37. The output values of the three networks for these events are presented in the first four lines of table 7.11.

The first simulated event shown is the last remaining MC background event when one sets very stringent cuts on the output of the smallest network. The next two events are its counterparts from the two other networks. It is interesting to notice that the respective other nets generally also have difficulties identifying these last fake events correctly. There is one exception to this rule however: The last background event from the largest network is clearly marked as a background event by the smallest network. Looking at this particular event in figure 7.40, one notices that this is a bundle of two muons. The last MC background event presented is one which is correctly identified by all networks. The four MC background events are displayed in figures 7.38 to 7.41. The corresponding output values are listed in the second block of table 7.11.

Finally some MC signal events are shown. The first one is a signal event mis-identified as background by all three networks. Looking at it in figure 7.42, one sees that this is a generated muon, which stops inside the detector. The second signal event is one which received the maximum output of 1.0 three times, i.e. is clearly marked as signal event by all three nets. The last event shown is a generated signal event which had rather high output values, but these were not as high as those of the (data) neutrino candidates. It therefore is not unambiguously identified. These three MC signal events are shown in figures 7.42 to 7.44. Their output values are given in the last three lines of table 7.11.

Table 7.11 summarizes the neural network results for these events.

	small network	medium network	large network
Neutrino candidate I	$\mathcal{O} = 0.99994$	$\mathcal{O} = 0.99988$	$\mathcal{O} = 1.0$
Neutrino candidate II	$\mathcal{O} = 1.0$	$\mathcal{O} = 0.99953$	$\mathcal{O} = 1.0$
Probable fake event	$\mathcal{O} = -0.87936$	$\mathcal{O} = -0.95914$	$\mathcal{O} = -0.88907$
Ambiguous measured event	$\mathcal{O} = 0.98378$	$\mathcal{O} = 0.99520$	$\mathcal{O} = 0.98215$
Last fake from the small net	$\mathcal{O} = 0.99752$	$\mathcal{O} = 0.98264$	$\mathcal{O} = 0.94195$
Last fake from the medium net	$\mathcal{O} = 0.97118$	$\mathcal{O} = 0.99948$	$\mathcal{O} = 0.97585$
Last fake from the large net	$\mathcal{O} = -0.95218$	$\mathcal{O} = 0.96289$	$\mathcal{O} = 0.99972$
Identified fake event	$\mathcal{O} = -0.95362$	$\mathcal{O} = -0.99874$	$\mathcal{O} = -0.99982$
Mis-identified signal event	$\mathcal{O} = -0.82418$	$\mathcal{O} = -0.99472$	$\mathcal{O} = -0.96747$
Good identified signal event	$\mathcal{O} = 1.0$	$\mathcal{O} = 1.0$	$\mathcal{O} = 1.0$
Ambiguous signal event	$\mathcal{O} = 0.99646$	$\mathcal{O} = 0.99360$	$\mathcal{O} = 1.0$

Table 7.11: The results of the neural networks for the events presented on the following pages.

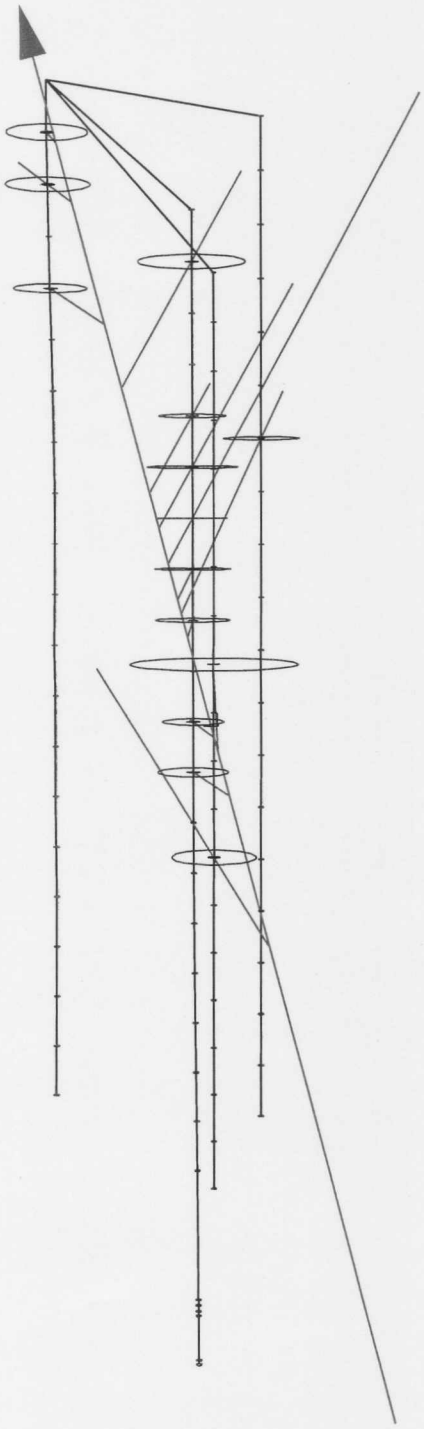


Figure 7.34: "Neutrino Candidate I", the first of the two data events which are potential neutrino candidates. This one was already discovered by a Swedish analysis.

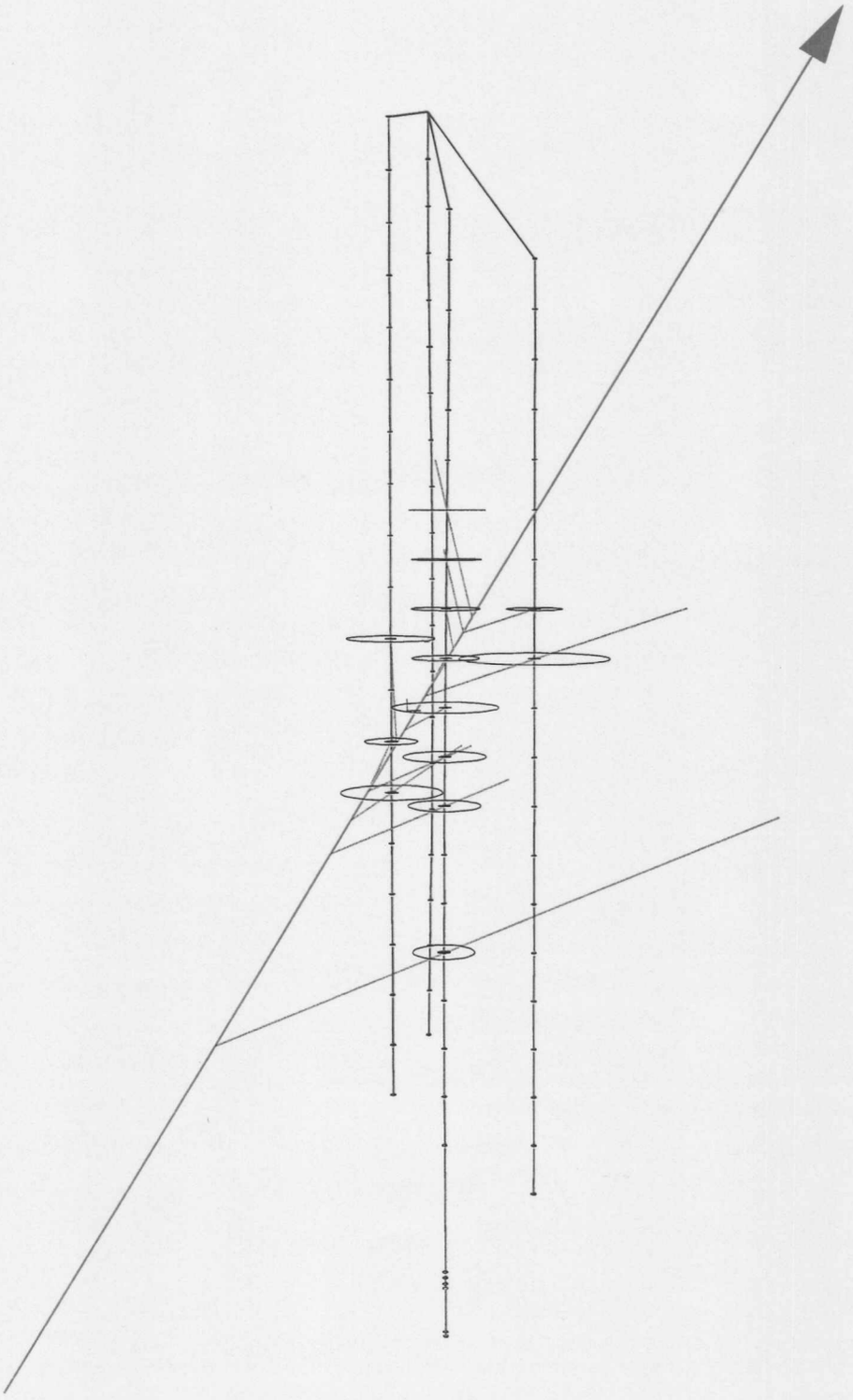


Figure 7.35: "Neutrino Candidate II", the second of the two data events which are potential neutrino candidates.

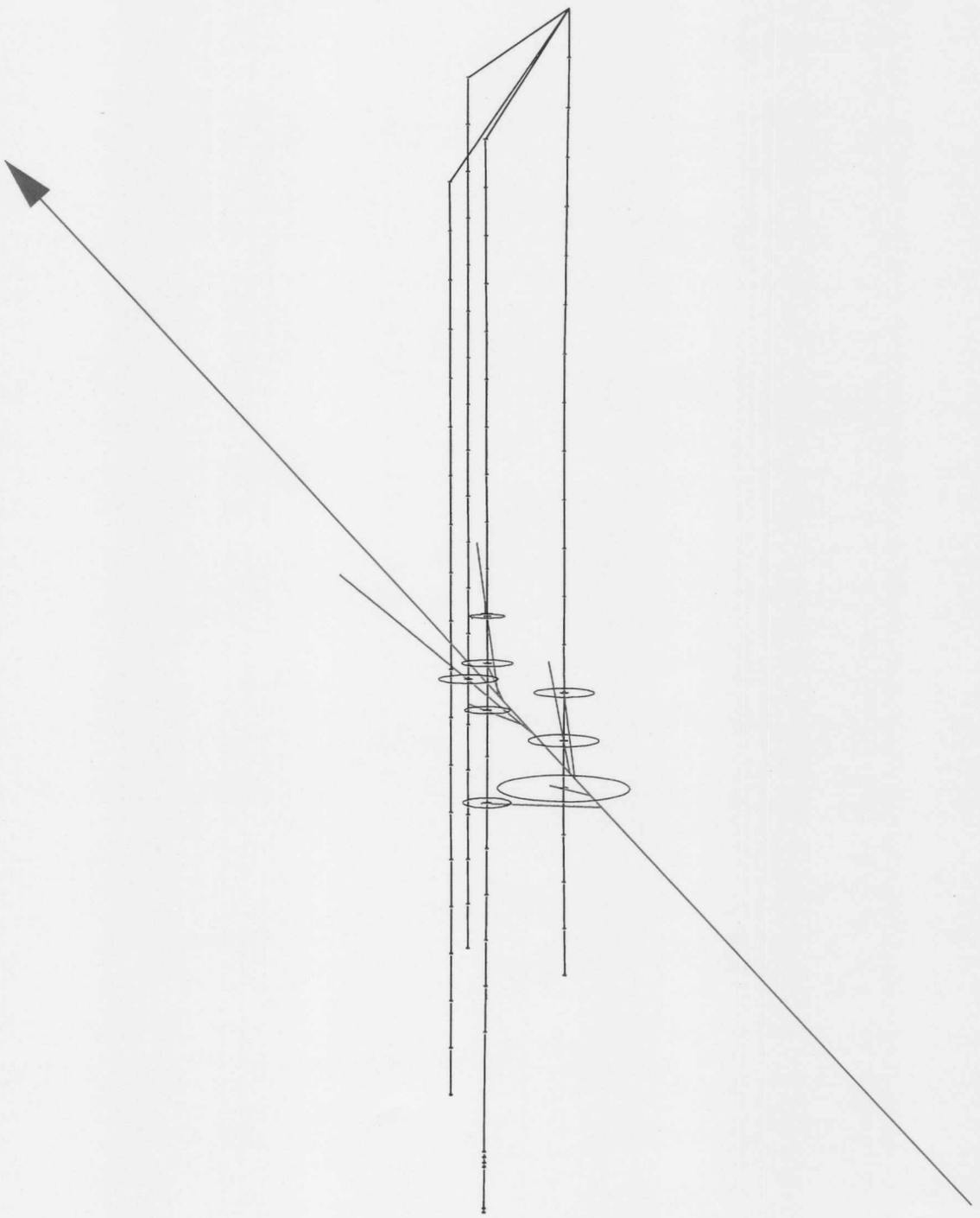


Figure 7.36: "Probable fake event", a measured event which all three networks considered to be a mis-reconstructed atmospheric muon event.

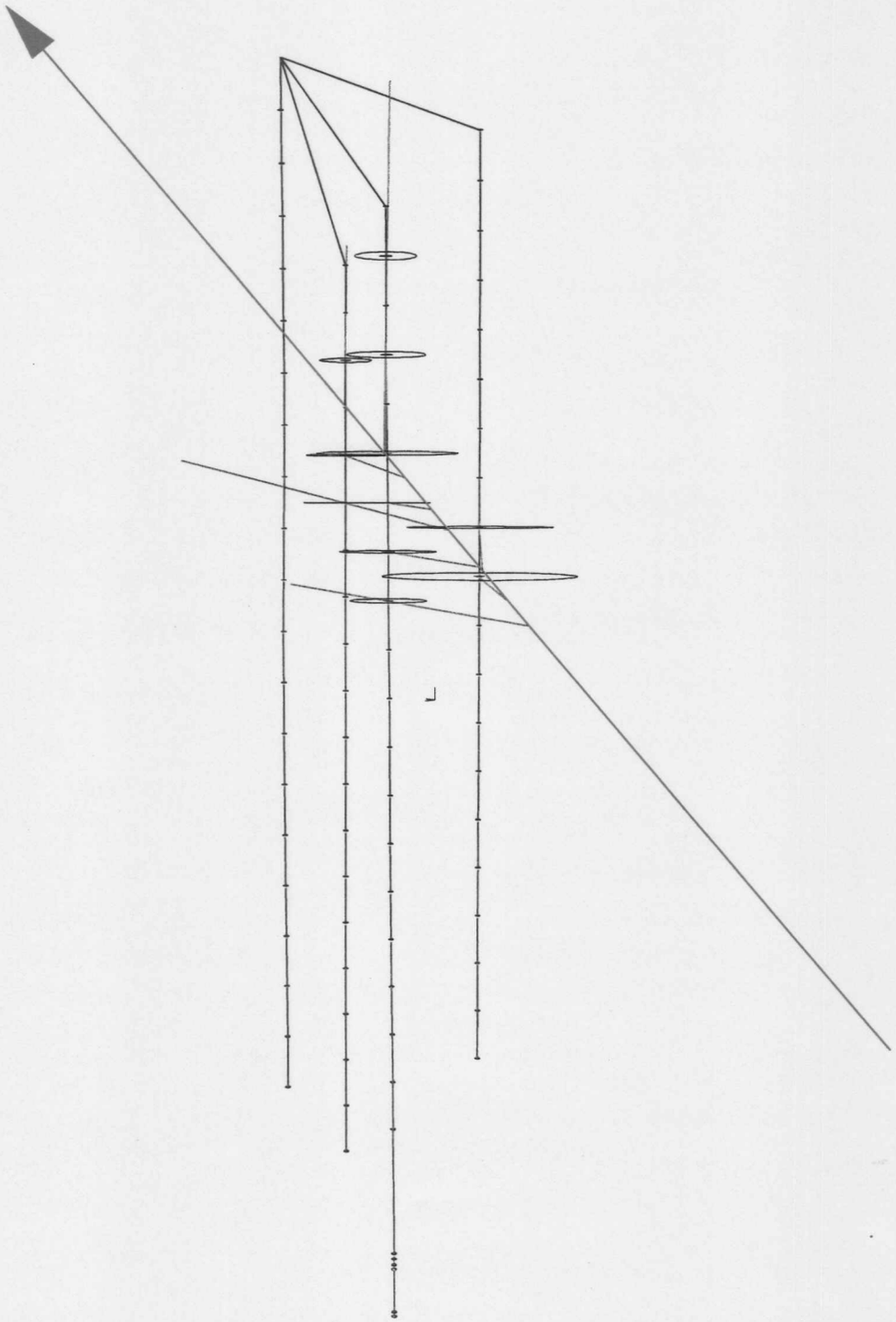


Figure 7.37: "Ambiguous measured event", a measured event which had an output $\mathcal{O} > 0.98$ for all three networks. Due to the expected neutrino flux, it is nevertheless probably a fake event.

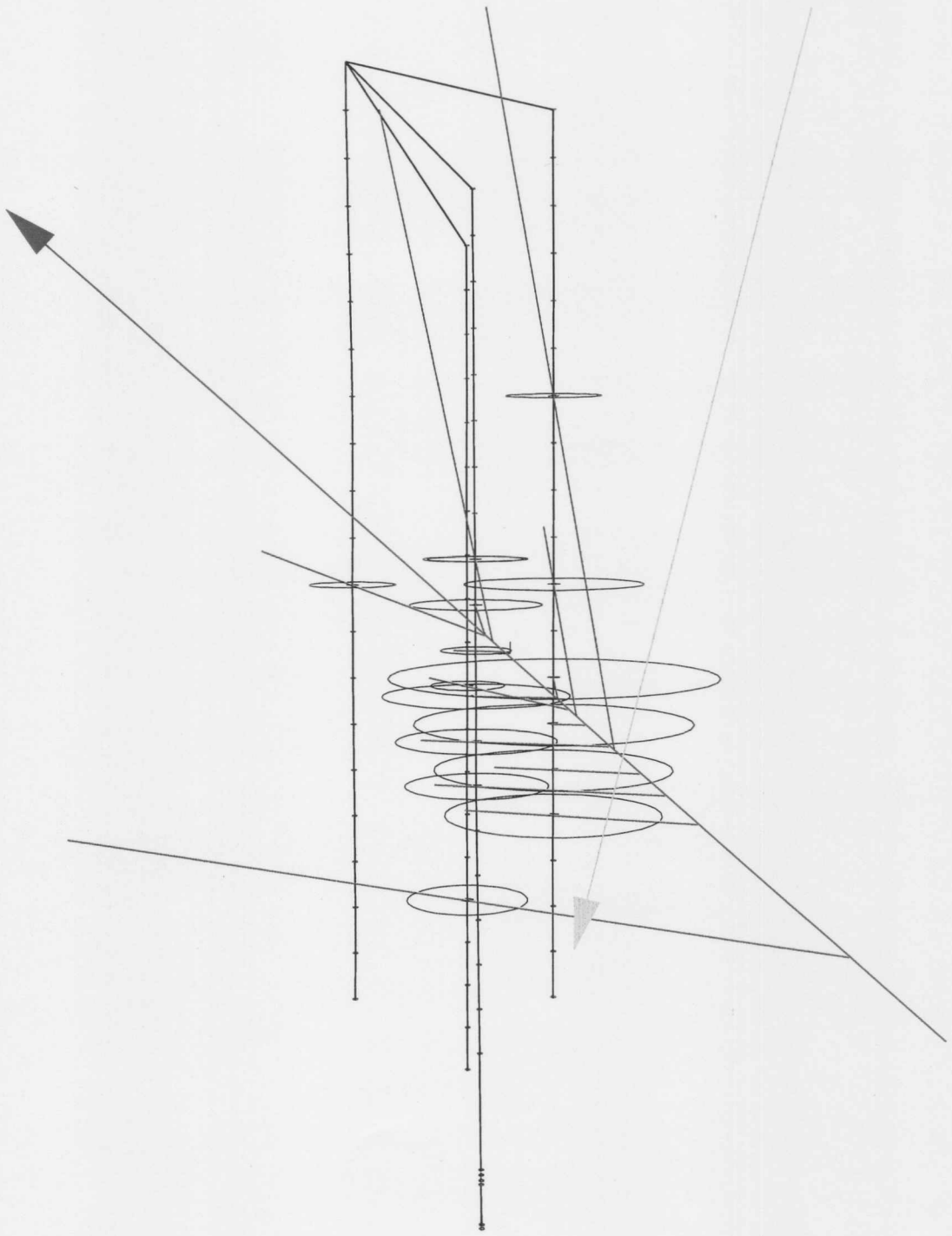


Figure 7.38: "Last fake from the small net", the mis-reconstructed MC background event which had the highest probability to be a neutrino event according to the analysis with the basic network. The generated muon is stopping near the detector.

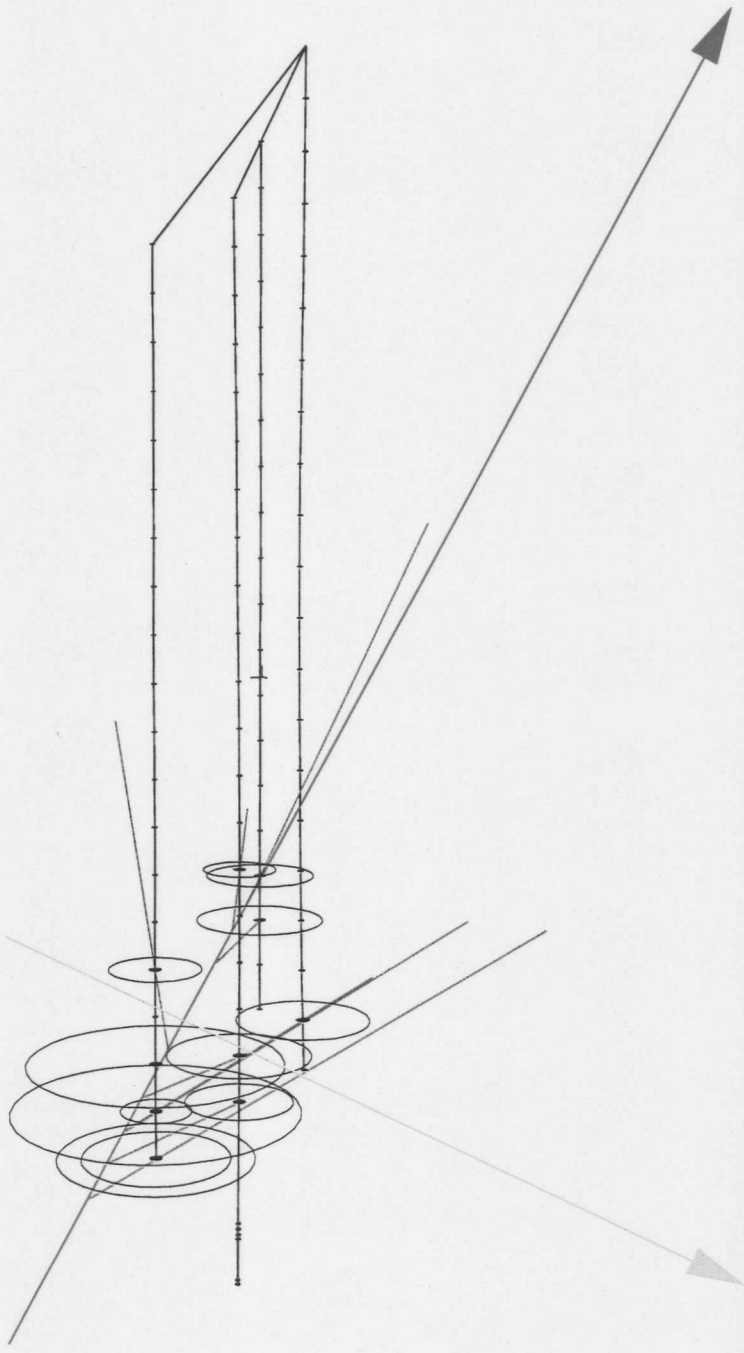


Figure 7.39: "Last fake from the medium net", the mis-reconstructed MC background event which had the highest probability to be a neutrino event according to the analysis with the extended network.

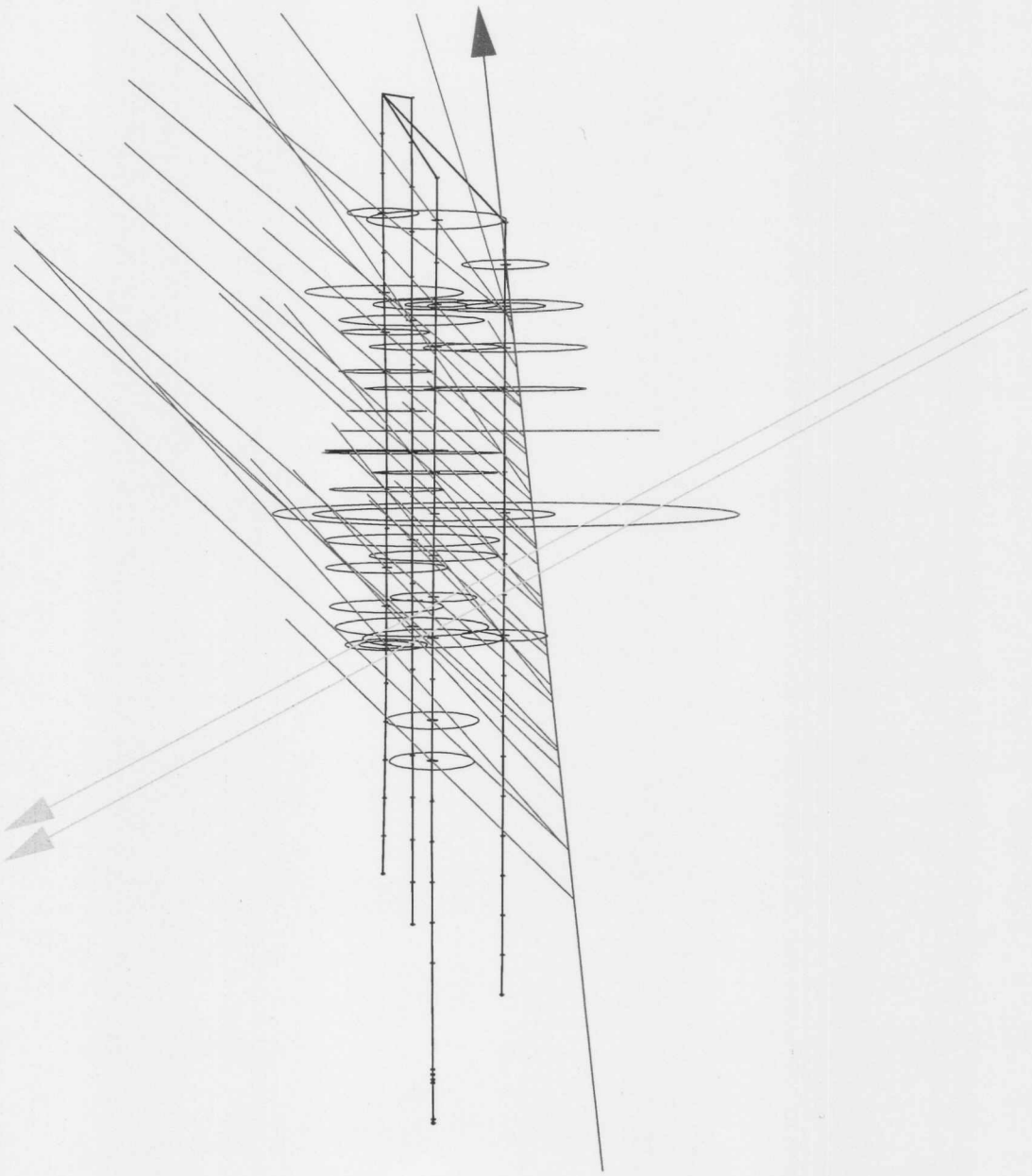


Figure 7.40: "Last fake from the large net", the mis-reconstructed MC background event which had the highest probability to be a neutrino event according to the analysis with the largest network. This is an example of an atmospheric muon bundle. A total light output of this event is especially large due to a 3.9 TeV bremsstrahl event in less than 100 m distance of the detector.

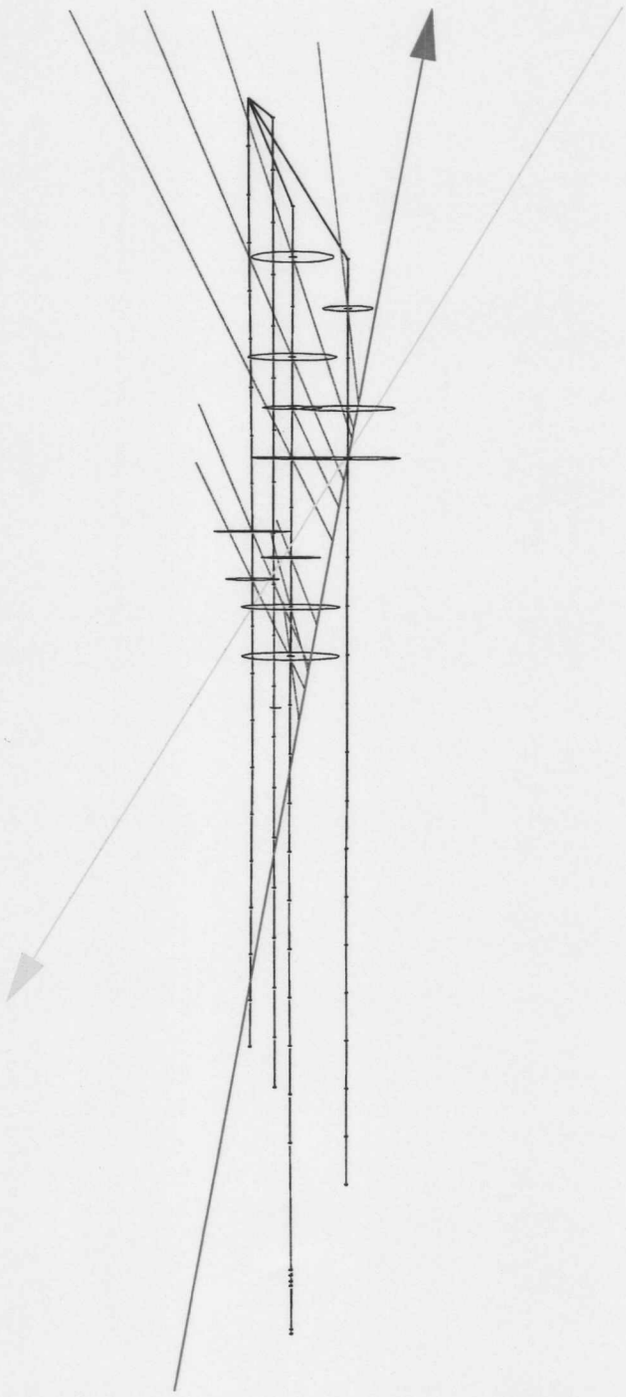


Figure 7.41: "Identified fake event", a mis-reconstructed MC background event, which was identified as mis-reconstructed by all three networks.

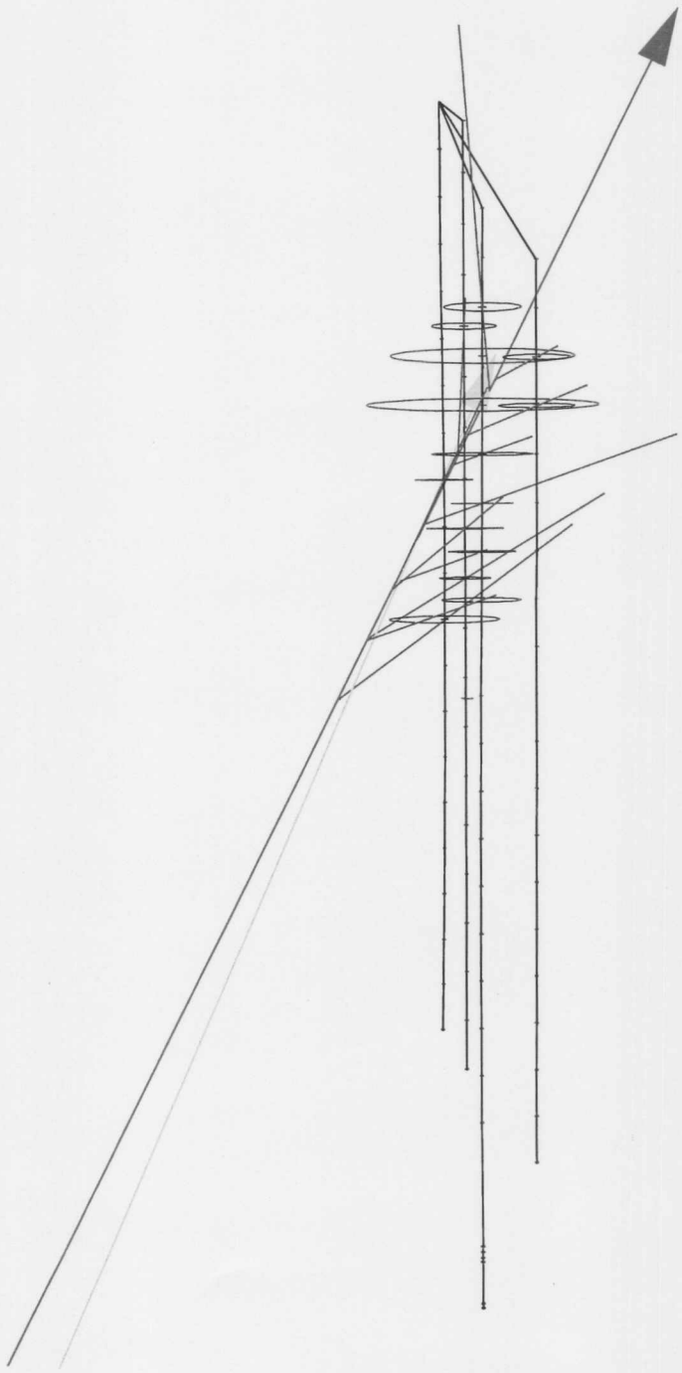


Figure 7.42: "Mis-identified signal event", a generated signal event, which was mis-identified as a fake event by all three networks. The generated muon is stopping behind the detector.

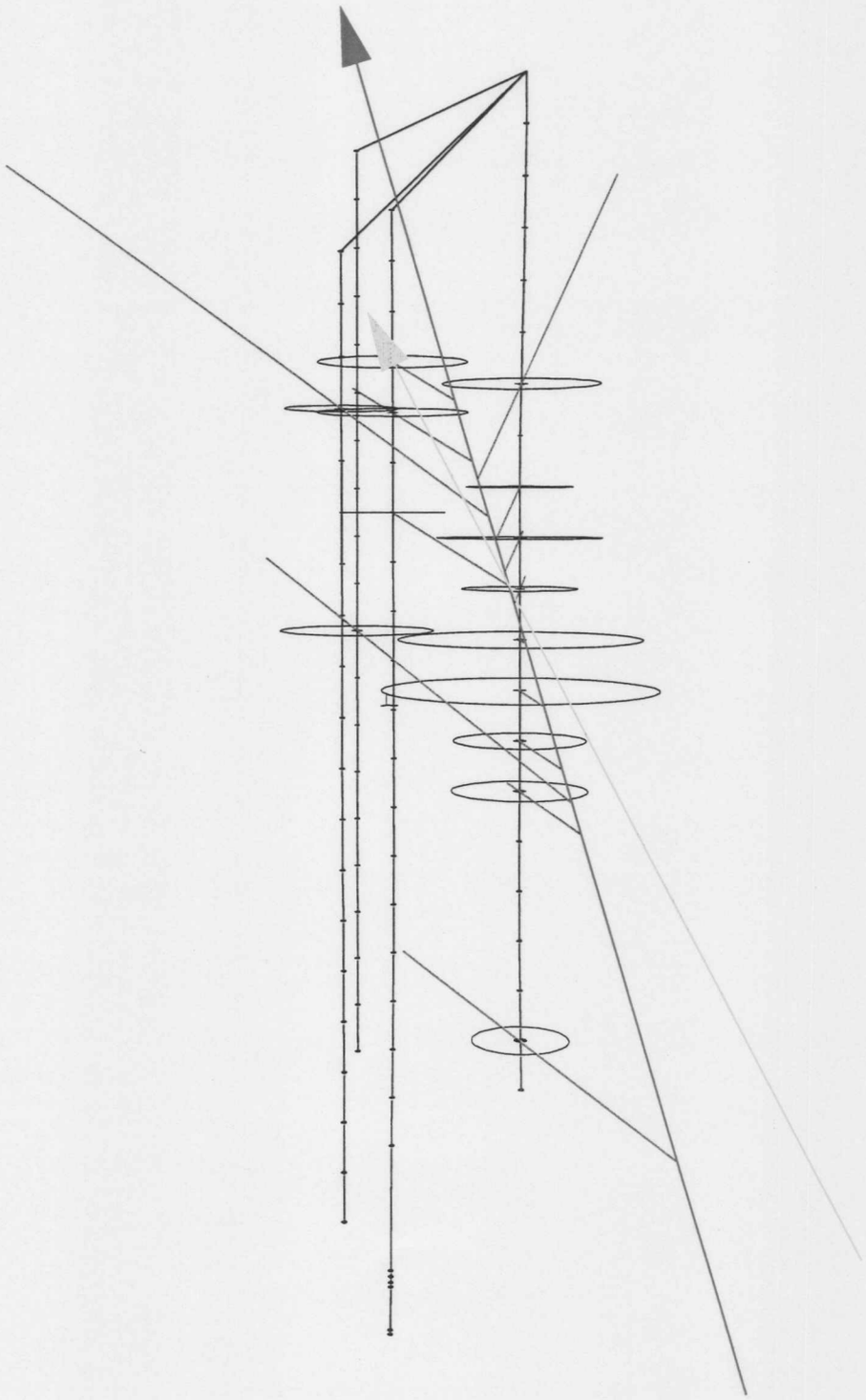


Figure 7.43: "Good identified signal event", a generated signal event which got assigned an output of 1.0 by all three networks.

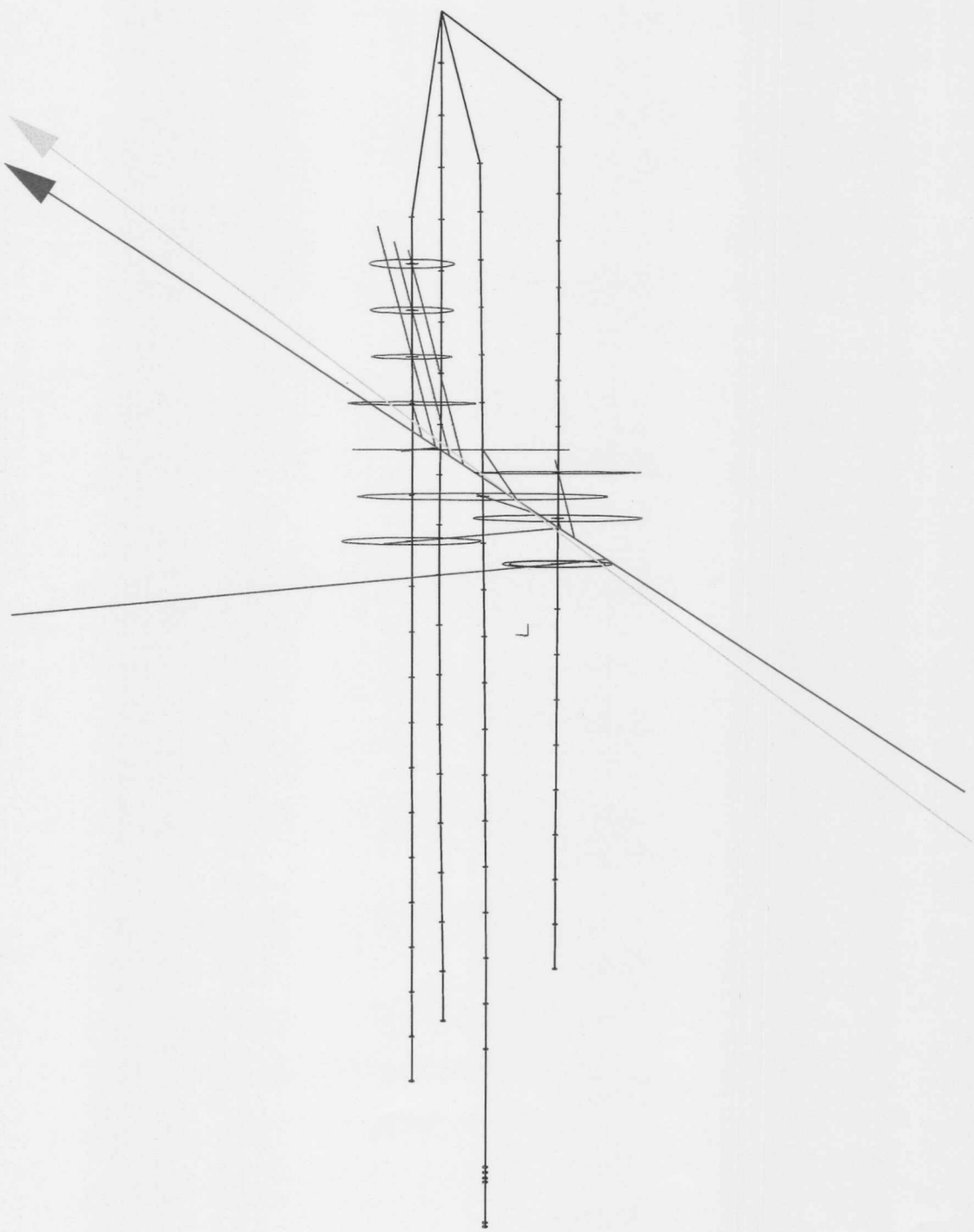


Figure 7.44: "Ambiguous signal event", an example of a generated signal event, which got assigned an output $\mathcal{O} > 0.99$ by all three networks.

8 Summary and Outlook

This work was a first approach to incorporate artificial neural networks into the quality analysis of the AMANDA detector. The main results can be summarized as follows:

1. Already standard artificial neural network designs can be used as a tool for the quality analysis in the AMANDA experiment.
2. As the detector is being constantly upgraded and expanded, only geometry independent parameters were used. These are either reconstruction results (like the reconstructed zenith angle) or averaged hit data (like the average z-component of all hit PMTs). It was shown that a set of input parameters which was appropriate for one detector version could also be used for a different geometry.
3. The results of a neural network analysis of the future AMANDA-II detector were compared to an established analysis which sets individual cuts on the same quality criteria. Both were performed with the same MC samples for atmospheric muons and an arbitrary signal of 1 TeV neutrinos. Demanding the same rejection for atmospheric muons, the neural network analysis achieved a ~ 30 % improved signal passing rate. With a few additional input parameters, the signal passing rate could be improved by almost ~ 50 %, while at the same time rejecting more fake events.
4. The method of a neural network analysis can be applied to data. Events measured with the AMANDA-B4 detector in 1996 were analyzed and two neutrino candidates were found. A signal to noise ratio for them cannot be given due to limited MC data. The two candidates are presented together with some generated MC and measured data events.
5. The neural network analysis gives results which are consistent with those from an independent analysis: One of the neutrino candidates had already been discovered by a Swedish study, which used a very different analysis technique.

This work was not intended to and could not be a complete analysis of the capabilities of neural networks as analysis tools for AMANDA. As an outlook for potential future research, several areas of interest for neural networks in AMANDA shall be given:

- True signal to noise ratios have to be calculated. In order to do so, large statistics of simulated MC background and MC signal have to be generated.
- Additional potential input parameters should be tested.
- Instead of separating signal events from mis-reconstructed atmospheric muons, neural networks might be trained to estimate the reconstruction error.
- New classes of networks can be tested: These could for example contain more than one hidden layer, backcoupling links, other activation or update functions, etc.

With the AMANDA-II detector, which shall be completed in the year 2000, first physics results are expected. By that time a neural network analysis should be established as a standard tool during the analysis of measured data events.

List of Figures

List of Tables

References

[1] Particles and Fields, Phys. Rev. D 54 (1996)

[2] A. Ącker, S. Pakavasa, Phys. Lett. B 397 (1997) 209
G. Conforto et al., Astropart. Phys. 5 (1996) 147

[3] The AGASA collaboration, presented by N. Sakaki, Proc. 25 ICRC, Durban, Aug 1997, Vol.5, 217
The AGASA homepage, http://www.icrr.u-tokyo.ac.jp/as/project/agasa.html

[4] The AMANDA collaboration, presented by S. Barwick, Proc. 25 ICRC, Durban, Aug 1997, Vol.7, 1
The AMANDA homepage, http://amanda.berkeley.edu/

[5] The ANTARES Collaboration, presented by N. de Botton, Proc. 25 ICRC, Durban, Aug 1997, Vol.7, 17
The ANTARES homepage, http://marcpl1.in2p3.fr/astro

[6] ANTARES Collaboration, proposal to the IN2P3 Scientific Committee, May 96

[7] C. Athanassopoulos et al., Phys. Rev. Lett. 75 (1995) 2650 (nucl-ex/9706006)

[8] The PIERRE AUGER collaboration, presented by M. Boratav, Proc. 25 ICRC, Durban, Aug 1997, Vol.5, 205
Auger collaboration, The PIERRE AUGER Design Report, Fermilab, March 1997
The PIERRE AUGER homepage, http://www-lpnhep.in2p3.fr/auger/

[9] J.N. Bahcall, Neutrino Astrophysics, Cambridge University Press, Cambridge, 1989

[10] The BAIKAL collaboration, presented by L. Kuzmichev, Proc. 25 ICRC, Durban, Aug 1997, Vol.7, 21
The BAIKAL homepage, http://www.ifh.de/baikal/baikalhome.html

[11] The BAIKAL collaboration, presented by C. Spiering, Talk at Int. School of Nuclear Physics, Erice, Sep 1997 (astro-ph/9801044)

[12] I.A. Belolaptikov et al., Astropart. Phys. 7 (1997) 263

[13] L. Bergström et al., Phys. Rev. D55 (1997) 1765

[14] R.M. Bianti et al., Phys. Rev. Let. 58 (1987) 1494
K.Hirata et al., Phys. Rev. Let. 58 (1987) 1490

[15] P.L. Biermann et al., Phys. Rev. D 51 (1995) 3450 (astro-ph/9501001)

[16] A. Biron et.al., Upgrade of AMANDA-B towards AMANDA-II, Proposal, DESY-Zeuthen, Jun 1997.

[17] A. Bouchta, Muon analysis with the AMANDA-B Four-String Detector, PhD thesis, Stockholm University, Mar 1998

[18] S.N. Boziev et al., INR-Preprint P-0630, 1989

[19] C.G. Callen, Phys. Rev. D26 (1982) 2058
V.A. Rubakov, JETP Lett. 33 (1981) 644

[20] The CANGAROO collaboration, presented by T. Tanimori, Proc. Towards a Major Atmospheric Cherenkov Detector V, Kruger, South Africa, Aug 1997
The CANGAROO homepage, http://icrhp9.icrr.u-tokyo.ac.jp/

[21] V. Chaloupka et al., LBL-38321, Lawrence Berkley Laboratory, Feb 1996

[22] Report of the Scientific Assessment Group for Experiments in Non Accelerator Physics, Feb. 96, DOE/USA
The DUMAND homepages, http://www.phys.hawaii.edu/dmnd/dumand.html, http://web.phys.washington.edu/local_web/dumand/aaa_dumand_home.html

[23] B.F.J. Manly, Multivariante statistical methods, Capman and Hall, 1986
M.S. Srivastava and E.M. Carter, An Intrduction to Applied Multivariate Statistics, North Holland, 1979
J.H. Friedmann et al., IEEE Transactions on computers, Vol. C-23, No. 9 (1974) 881

[24] T.K. Gaisser, F. Halzen and T. Stanev, Phys. Rep. 258 (1995) 173

[25] T.K. Gaisser, Cosmic Ray and Particle Physics, Cambridge University Press, Cambridge, 1991.

[26] F. Halzen, Talk given at the 5th International Workshop on Topics in Astroparticle and Underground Physics TAUP 97, Gran Sasso, Italy, Sep 97, MADPH-97-1026, Univ. of Wisconsin, Madison, Nov 1997
T.K. Gaisser, Talk given at the OECD MEGASCIENCE Workshop, Taormina, Italy, May 97 (astro-ph/9707283)
F. Halzen, MADPH-96-981, Univ. of Wisconsin, Madison, Dec 1996 (astro-ph/97/01029)
F. Halzen, Nuclear and Particle Astrophysics and Cosmology, Proc. Snowmass 94, eds: R. Kolb and R. Peccei

[27] F. Halzen, Ice Fishing for Neutrinos,
http://amanda.berkeley.edu/www/ice-fishing.html

[28] F. Halzen, MADPH-97-1007, Univ. of Wisconsin, Madison, Jul 1997 (astro-ps/9707289)

[29] F. Halzen, G. Jaczko, Phys. Rev. D 54 (1996) 2774

[30] N. Hayashida et al., Proc. 25th ICRC, Durban, Aug 1997, Vol.4, 177
V.A. Kolosov et al., Proc. 25th ICRC, Durban, Aug 1997, Vol.4, 169
A.M. Hillas, Ann. Rev. Astr. Astrophys., 22 (1984) 425

[31] N. Hayashida, Proc. 25 ICRC, Durban, Aug 1997, Vol.4, 145
G. Cocconi, Astropart. Phys.4 (1996) 281
N. Hayashida et al, Phys. Rev. Lett. 73 (1994) 3491

[32] The HEGRA collaboration, presented by A. Lindner, Proc. 25 ICRC, Durban, Aug 1997, Vol.5, 113
The HEGRA homepage, http://wpos6.uni-wuppertal.de:8080/

[33] Various authors, Proc. 25th ICRC, Durban, Aug 1997, Vol.3, 161 - 301

[34] G.C. Hill, Astropart. Phys.6 (1997) 215
R. Gandhi, C. Quigg, M.H. Reno, I. Sarcevic, Astropart. Phys.5 (1996) 81

[35] S. Hundertmark et.al., Optimisation of the AMANDA-B detector geometry including a first approach to muon reconstruction. AMANDA internal report, Zeuthen, Dec 1996

[36] S. Hundertmark and M. Leuthold, Calculation of the Effective Trigger Volume at 100, 1000, 10000 GeV for the AMANDA-B Detector, AMANDA internal report, Zeuthen, Feb 1998

[37] S. Hundertmark, PhD thesis, to appear at the Humboldt-Universität zu Berlin, Berlin 1998

[38] S. Hundertmark, AMASIM homepage,
http://www.ifh.de/~hundert/amanda/amasim/amasim.html

[39] J. Jacobsen, RAVEN homepage, http://alizarin.physics.wisc.edu/jacobsen/raven

[40] J. Jacobsen, Simulating the Detection of Muons and Neutrinos in Deep Antarctic Ice, PhD Thesis, University of Wisconsin in Madison, Dec 1996 (available at http://alizarin.physics.wisc.edu/jacobsen)

[41] W. Lohmann et al., CERN 85-03, Geneva 1995

[42] The SUPER-KAMIOKANDE collaboration, presented by J.G. Learned, Proc. 25 ICRC, Durban, Aug 1997, Vol.7, 73
The SUPER-KAMIOKANDE homepage, http://www-sk.icrr.u-tokyo.ac.jp/doc/sk/

[43] A. Karle, Monte Carlo simulation of photon scattering and absorption for AMANDA, AMANDA internal report, draft, Zeuthen, Feb 1997

[44] A. Karle et al., NIM A 387 (1997) 274

[45] J.I. Katz, T. Piran, Proc. 4th Huntsville GRB Symp., 1997 (astro-ph/9712242)

[46] G.B. Khristiansen et al., Proc. 25th ICRC, Durban, Aug 1997, Vol.4, 201

[47] A.R. Liddle, SUSSEX-AST 96/12-1, Univ. of Sussex, U.K., Dec 1996 (astro-ph/9612093)
E.W. Kolb and M.S. Turner, The Early Universe, Addison-Wesley, Redwood City, California (1990) [updated paperback 1994]

[48] R.A. Fisher, Annals of Eugenics, Band 7 (1936)

[49] The NESTOR collaboration, presented by A. Capone, Proc. 25 ICRC, Durban, Aug 1997, Vol.7, 49
The NESTOR homepage, http://www.roma1.infn.it/nestor/nestor.html,

[50] Proc. Padova Workshop on TeV Gamma-Ray Astrophysics, Padova, 1995

[51] D. Pandel, Measurement of water and detector parameters and reconstruction of muons to energies of 100 TeV with the BAIKAL Neutrino Telescope NT-72 (in German), diploma thesis, Humboldt-University, Berlin, Feb 1996

[52] L. Pasquali et al., Talk presented at the Fifth International Workshop on Topics in Astroparticle and Underground Physics, Sep 1997 (astro-ph/9711457)
T.K. Gaisser et al., Phys. Rev. D 54 (1996) 5578

[53] D.H. Perkins, Introduction to high energy physics, Addison-Wesley, New York 1987

[54] R. Porrata, Phd Thesis, Univeristy of California in Irvine, Sep 1997
(http://www.ps.uci.edu/~porrata/thesis.html)

[55] J. Prahl et al., Proc. 24th ICRC, Rome, 1995, Vol.2, 354

[56] M.J. Rees, astro-ph/9701162, Jan 1997

[57] A. Reinders, Use of an Artificial Neural Network in AMANDA, AMANDA internal report, Uppsala, June 1997,
http://www3.tsl.uu.se/thep/thunman/reports/isv_int/1997/isvint31997.html

[58] W. Rhode et al., Astropart. Phys.4 (1996) 217

[59] SAGE collaboration, J.N. Abdurashitov et al., Phys. Rev. Lett. 77 (1996) 4708
GALLEX collaboration, W. Hampel et al., Phys. Lett. B 388 (1996) 384
KAMIOKANDE collaboration, Y. Fukuda et al., Phys. Rev. Lett. 77 (1996) 1683
B.T. Cleveland et al., Nucl. Phys. B (Proc. Suppl.) 38 (1995) 47
J.N. Babcall , M.H. Pinsonncault , Rev. Mod. Phys. 67 (1995) 781

[60] The SiEGMuND homepage,
http://www.ifh.de/baikal/software/siegmund/siegmund.html

[61] T. Stanev et al., Phys. Rev. Lett. 75 (1995) 3056

[62] V.J. Stenger, Track fitting for the DUMAND Octagon, HDC-1-90, Hawaii, 1990

[63] O. Streicher, Nachweis von Myonen aus Wechselwirkungen atmosphärischer Neutrinos in Unterwasser-Neutrino-Teleskopen , diploma thesis (in German), Humboldt-University, Berlin, Oct 1994

[64] C. Sutton, Spaceship Neutrino, Cambridge University Press, Cambridge 1992

[65] T. Tanimori et al., Astrophys. J. Lett., 492, (1998) L33-36
X.Sarazin et al., Astropart. Phys 4 (1996) 227 and references therein

[66] K. Simonyi, Kulturgeschichte der Physik (in German) , Verlag Harri Deutsch, Thun/Frankfurt am Main, 1990
Hungarian original: A fizika kultúrtörténete, Gondolat Kiadó, Budapest 1986.

[67] The SOUDAN-2 collaboration, W.W.M. Allison, et al., Phys. Lett. B 391 (1997) 491
IMB collaboration, R. Becker-Szendy et al., Nucl. Phys. B (Proc. Suppl.), (1995) 331
Y. Fukuda et al., Phys. Lett. B 335 (1994) 237
KAMIOKANDE collaboration, K.S. Hirata et al., Phys. Lett. B 280 (1992) 146

[68] A.Vilenkin and E.P.S. Shellard, Cosmic Strings and Other Topological Defects, Cambridge University Press (1994)

[69] S. Westerhoff et al., Astropart. Phys.4 (1996) 119

[70] The WHIPPLE collaboratioon, presented by G. Sembroski, Proc. 24 ICRC, Rome, Sep 1995, Vol.3, 428
The WHIPPLE homepage, http://egret.sao.arizona.edu/index.html

[71] C. Wiebusch et al., Muon reconstruction with AMANDA. Proc. 25 ICRC, Durban, Aug 1997, Vol.7, 13

[72] C. Wiebusch, The Detection of Faint Light in Deep Underwater Neutrino Telescopes, PhD thesis, PITHA 95/37, RWTH-Aachen, 1995

[73] E. Zas, Talk given in "Very High Energy Phenomena in the Universe", XXXIInd Rencontres de Moriond, Apr 1997 (astro-ph/9704016)

[74] A. Zell A. et al., SNNS — Stuttgart Neural Network Simulator, user manual 4.1, University of Stuttgart, Jun 1995.

[75] A. Zell, Simulation Neuronaler Netze (in German), Addison Wesley, 1994.
R. Brause, Neuronale Netze (in German), Teubner, 1991.

D Acknowledgement

I would like to thank a lot of people.

First of all in Zeuthen, I have to thank Stefan Hundertmark (for providing the Monte Carlo samples as well as Fortran expertize), Sabine Schilling (for all the assistance throughout her six weeks in Zeuthen), Dr. Christian Spiering (for being my superviser and proof-reading), Dr. Christopher Wiebusch (for providing the reconstruction used, proof-reading and helping me throughout the year) and all the rest of the AMANDA group.

At the university, thanks go to several people, but especially to Prof. Dr. Thomas Lohse for long discussions prior to this work and for taking the responsiblity for my thesis.

I would like to thank my friends for providing the necessary distraction – the importance of which is often underestimated. I want to send particular thanks to Stefan Fischer Rivera for proof-reading.

But most of all I'd like to thank my family for being everything you wish your family to be – the perfect gift nobody else can replace.

E Deutsche Zusammenfassung

Die vorliegende Diplomarbeit beschäftigt sich mit der Unterdrückung atmosphärischer Myonen im AMANDA Experiment. Ziel dieses Experimentes ist der Nachweis hochenergetischer Neutrinos. Hierzu werden Photomultiplier in einer Gitteranordnung tief ins südpolare Eis versenkt. Vollzieht ein hochenergetisches Myonneutrino in der Umgebung des Detektors eine geladene schwache Wechselwirkung, so wandelt es sich in ein ebenfalls hochenergetisches Myon um, welches Čerenkovlicht abstrahlt. Dieses Licht kann von den einzelnen Photomultiplierzellen registriert werden. Sprechen hinreichend viele Zellen an, so ermöglichen die gewonnenen Zeit- und Amplitudeninformationen eine Rekonstruktion der Myonspur. Im für AMANDA interessanten Energiebereich oberhalb einiger zehn GeV sind Myon und Neutrino fast kolinear, weswegen sich die Neutrinospur durch die Myonspur annähern läßt. Man hofft hierdurch Punktquellen kosmischer Neutrinos entdecken zu können.

Der Hauptuntergrund bei solchen *Neutrinoteleskopen* sind atmosphärische Myonen. Falls geladene kosmische Strahlen auf die Erdatmosphäre treffen, so lösen sie eine hadronische Kaskade aus, in deren Verlauf eine große Zahl Myonen entstehen. Da dies minimalionisierende Teilchen mit einer relativ langen Lebensdauer ($\tau = 2,2 \cdot 10^{-6}$ s) sind, dringen einige von ihnen bis zum AMANDA-Detektor in das Eis ein. Im Vergleich zu den erwarteten Flüssen von neutrinoinduzierten Myonen, treten diese atmosphärischen Myonen um etwa fünf bis sechs Größenordnungen häufiger auf. Da Myonen im Gegensatz zu Neutrinos nicht durch die Erde durchfliegen können, werden im Experiment nur solche Ereignisse berücksichtigt, die als von unten kommend rekonstruiert wurden. Aufgrund von Fehlrekonstruktionen dominieren aber auch nach diesem Winkelschnitt atmosphärische Myonen jegliches gemessene Signal.

Ziel dieser Arbeit war es Möglichkeiten zur Unterdrückung dieser fehlrekonstruierten Untergrundmyonen mit Hilfe künstlicher neuronaler Netzwerke zu untersuchen. Die Untersuchungen beschränkten sich auf *feed-forward* Netzwerke mit je einer Eingabe-, einer verborgenen und einer Ausgabeebene. Als Eingabeparameter wurden insbesondere Qualitätsmerkmale des Rekonstruktionsverfahrens benutzt. Die Netzwerke wurden mit Hilfe einiger tausend Monte Carlo-simulierten Untergrund- und Signalereignissen trainiert. Als Ausgabe berechneten sie einen Wert zwischen -1 und $+1$, wobei ein Wert von etwa -1 ein wahrscheinliches Untergrundereignis, ein Wert nahe $+1$ jedoch ein wahrscheinliches Signalereignis darstellte.

Simulationen von zwei verschiedenen Ausbaustufen des AMANDA Detektors wurden untersucht. Begonnen wurde mit einer Analyse für den AMANDA-II Detektor, welcher im Jahre 2000 fertiggestellt werden soll. Hierbei wurde zunächst versucht dieselbe Unterdrückung, welche bereits mit einer „konventionellen" Analyse erzielt worden war zu erreichen. Hierfür wurden dieselben Monte Carlo Simulationen eines 600 GeV Signals (im Detektor), sowie des wahren Spektrums atmosphärischer Myonen benutzt, die bereits bei der anderen Analyse verwendet wurden. Bei identischer Untergrundunterdrückung wurde eine um fast 30 % erhöhte Neutrinoakzeptanz erzielt. Wurden zusätzliche Eingabeparameter verwendet, welche nicht für die „konventionelle" Analyse berücksichtigt worden waren, so konnte bei weit verbesserter Untergrundunterdrückung die Signalakzeptanz sogar um

annähernd 50 % verbessert werden. Das für diese beiden Ereignisklassen erzielte Verhältnis der Akzeptanzen betrug schließlich $5 \cdot 10^5$.
Anschließend wurde eine analoge Analyse über eine Simulation des seit 1996 bestehenden AMANDA-B4 Detektors unternommen. Als Signal standen kontinuierliche Spektren mit Energien zwischen 0 und 100 GeV, 1 TeV bzw. 10 TeV zur Verfügung. als Untergrund wurde wiederum eine Simulation des wahren atmosphärischen Myonspektrums verwendet. Erneut zeigte sich, daß eine hohe Unterdrückung des Untergrunds erreicht werden konnte. In diesem Teil der Arbeit wurden drei verschiedene Netzwerke trainiert. Wandte man diese auf die an einem Tag gemessene Daten an, so blieben in allen drei Netzwerken dieselben beiden Ereignisse als wahrscheinlichste gemessene Neutrinokandidaten dieses Tages zurück. Aufgrund fehlender passender Monte Carlo-Statistik und leichter Abweichungen zwischen Daten und Monte Carlo konnte für die beiden Kandidaten jedoch keine Neutrinowahrscheinlichkeit berechnet werden. Man erwartet pro Tag etwa 0,6 getriggerte atmosphärische Neutrinoereignisse. Da aus einer anderen Analyse bereits bekannt war, daß an dem untersuchten Tag ein wahrscheinlicher Neutrinokandidat gemessen worden war, erwartet man für diesen Tag etwa 1,6 Neutrinoereignisse. Somit liegt es im Bereich des möglichen, daß es sich bei beiden um tatsächliche Neutrinoereignisse handelt. Wenn man die Ergebnisse für das Signalspektrum bis 1 TeV zugrunde legt, so entsprechen die beiden Kandidaten Signal zu Untergrund Verhätnissen von $\gtrsim 1 \cdot 10^5$.
Einer der beiden gefundenen Kandidaten wurde bereits in einer vollständig unabhängigen schwedischen Studie gefunden. Somit wurde gezeigt, daß die Methode der neuronalen Netzwerkanalyse nicht nur auf simulierte Monte Carlo-Daten, sondern auch auf experimentell gemessene Daten anwendbar ist. Sie eignet sich somit als künftiges Standardwerkzeug der im AMANDA Experiment verwendeten Analysekette.

F Erklärung

Hiermit bestätige ich, daß ich die vorliegende Arbeit ohne unerlaubte fremde Hilfe angefertigt habe.
Ich bin mit der Auslage meiner Diplomarbeit in der Bibliothek der Humboldt–Universität zu Berlin einverstanden.

Berlin, den 31. März 1998

Alexander Prinz Biron von Curland