

# **First Search for Heavy Neutral Leptons with IceCube DeepCore**

**Dissertation**  
zur Erlangung des akademischen Grades  
doctor rerum naturalium  
(Dr. rer. nat.)

im Fach: Physik  
Spezialisierung: Experimentalphysik

eingereicht an der  
Mathematisch-Naturwissenschaftlichen Fakultät  
der Humboldt-Universität zu Berlin

von  
**Leander Fischer M. Sc.**  
geboren am 24. Oktober 1992  
in Heidelberg

Präsidentin der Humboldt-Universität zu Berlin  
Prof. Dr. Julia von Blumenthal

Dekanin der Mathematisch-Naturwissenschaftlichen Fakultät  
Prof. Dr. Caren Tischendorf

### **Copyright Notice**

This book is released into the public domain using the CC-BY-4.0 code.

To view a copy of the CC-BY-4.0 code, visit:

<https://creativecommons.org/licenses/by/4.0/>

### **Colophon**

This document was typeset with the help of KOMA-Script and L<sup>A</sup>T<sub>E</sub>X using the open-source kaobook template class.

The source code of this thesis is available at:

[https://github.com/LeanderFischer/phd\\_thesis](https://github.com/LeanderFischer/phd_thesis)

# Todo list

Include some low level plots like the trigger efficiency for the HNL simulation (RED) . . . . .	2
get trigger efficiency for HNL (somehow..) . . . . .	2
mention how this affects the HNL rate (in a general way) (RED) . . . . .	3
add table with rates per level (split in flavor) - maybe better in analysis chapter to also show signal? (RED)	4
get estimate of HNL efficiency across the levels.. (RED) . . . . .	4
blow up figures to make better visible? (ORANGE) . . . . .	8
remake plot: no bin counts, blow up labels etc. margin possible? (RED) . . . . .	9
fix caption (RED) . . . . .	9
re-make normalized? (ORANGE) . . . . .	10
fix caption (RED) . . . . .	10
re-make normalized? (ORANGE) . . . . .	10
fix caption (RED) . . . . .	10
fix caption (RED) . . . . .	11
fix caption of this figure (RED) . . . . .	12
Make a single plot, showing S/B or so? (RED) . . . . .	12
fix caption of this figure (RED) . . . . .	13
fix caption of this figure (RED) . . . . .	14
Make/add relevant plots here (RED) . . . . .	15
Plot energy (true total) and true decay length across the different levels (RED) . . . . .	15
Make table with the rates across the different levels for benchmark mass/mixing (RED) . . . . .	15
fix caption of this figure (RED) . . . . .	15
fix caption of this figure (RED) . . . . .	16
fix caption of this figure (RED) . . . . .	16
add a statement about overtraining using the plots showing the FLERCNN muon classifier (RED) . .	18
add some performance plots of the FLERCNN reconstruction (RED) . . . . .	18



# Contents

<b>Contents</b>	<b>v</b>
<b>1 Event Processing and Reconstruction</b>	<b>1</b>
1.1 Processing . . . . .	1
1.1.1 Trigger and Filter . . . . .	1
1.1.2 Event Selection . . . . .	2
1.2 Double Cascade Reconstruction . . . . .	4
1.2.1 Table-Based Minimum Likelihood Algorithms . . . . .	5
1.2.2 Optimization for Low Energies . . . . .	6
1.2.3 Performance . . . . .	7
1.3 Double Cascade Classification . . . . .	11
1.4 Generalized Double Cascade Performance . . . . .	13
1.4.1 Idealistic Events . . . . .	14
1.4.2 Realistic Events . . . . .	15
1.5 Analysis Reconstruction . . . . .	17
1.5.1 Fast Low-Energy Reconstruction using Convolutional Neural Networks . . . . .	17
1.5.2 Analysis Selection . . . . .	18
<b>Figures</b>	<b>21</b>
<b>Tables</b>	<b>23</b>
<b>Bibliography</b>	<b>25</b>



# Event Processing and Reconstruction

1

The analysis presented in this thesis is highly dependent on an efficient filtering and event selection to reduce the raw IceCube trigger data to a usable atmospheric neutrino sample. Based on this selection, a precise estimation of both expected SM background and expected BSM signal events can be made using MC simulations. Starting from the PMT output, both real data and simulation are processed through the in-ice trigger, the online filter and processing, and the low-energy event selection to produce a neutrino dominated sample. Once the sample is small enough for more sophisticated reconstruction techniques to be feasible to run, the events can be reconstructed with the existing IceCube reconstruction algorithms. At this level it is also possible to test and develop new reconstruction algorithms, without worrying about the large amount of background events from atmospheric muons and noise that are present before the filtering.

After describing the processing and filtering chain in Section 1.1, the development and performance of a dedicated low energy double cascade reconstruction algorithm is presented in Section 1.2. Based on the results from this reconstruction, the ability of the detector to observe and identify double cascades is discussed in Section 1.3. Finally the state of the art SM neutrino event reconstruction is presented in Section 1.5, which is used to perform the analysis in this thesis.

## 1.1 Processing

After the detector simulation is performed, all MC and data are processed in exactly the same way. This section explains the trigger and event selection that is applied starting from the raw voltage measured by the PMTs. It is split in different steps that are run inside the ice, at the South Pole, and after the data was transferred to the North. The complexity and computational cost of the processing increases with each step, while the total number of events reduces, making it feasible and reducing the use of computational resources on events that are not of interest for the analysis.

### 1.1.1 Trigger and Filter

Before the data can be sent to the North, the initial signal coming from the PMT is a voltage waveform that is digitized (for data) and then information of photon hits are extracted (also for the MC coming from the detector response simulation). The trigger and filter explained here are tailored to select events that pass through the DeepCore volume, while rejecting background events (either from atmospheric muons or from random noise). There are other filters used in IceCube which will not be explained here, since they are not relevant for this work. A full description of the instrumentation and the online systems can be found in [1].

1.1	Processing	1
1.2	Double Cascade Reconstruction	4
1.3	Double Cascade Classification	11
1.4	Generalized Double Cascade Performance	13
1.5	Analysis Reconstruction	17

[1]: Aartsen et al. (2017), "The IceCube Neutrino Observatory: Instrumentation and Online Systems"

Include some low level plots like the trigger efficiency for the HNL simulation (RED)

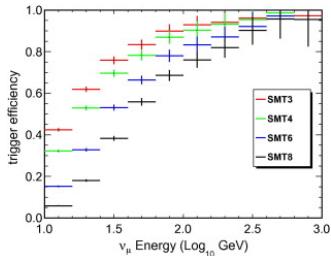


Figure 1.1: Efficiencies of different IceCube and DeepCore triggers, taken from [2].

[3]: Abbasi et al. (2009), "The IceCube data acquisition system: Signal capture, digitization, and timestamping"

[4]: Aartsen et al. (2017), "The IceCube Neutrino Observatory: instrumentation and online systems"

[2]: Abbasi et al. (2012), "The design and performance of IceCube DeepCore"

get trigger efficiency for HNL (somehow..)

1: Where *online* means running on hardware at the South Pole as opposed to *offline* at the IceCube institutions in the Northern Hemisphere.

### In-ice Trigger

The trigger is applied inside the DOM in the ice before sending the information to the ICL on the surface. The time dependent voltage curves are captured if a pre-defined threshold value is exceeded. Once the threshold, set to the equivalent of 0.25 PE, is crossed, 6.4  $\mu$ s of the waveform are coarsely digitized by a *Fast Analog-to-Digital Converter* (FADC) with a sampling rate of 40 MHz. Additionally, the first 427 ns are digitized using an *Analog Transient Waveform Recorder* (ATWD) with a sampling rate of 300 MHz [3], but only if some trigger condition is met, because this readout frequency is too high to be sampled directly and requires some buffering. For DeepCore, the HLC condition already mentioned in Section ?? has to be met for three DOMs inside the fiducial volume within a time window of 5  $\mu$ s. If this is the case, all waveforms that crossed the threshold within a 20  $\mu$ s time window around the trigger are digitized and sent to the ICL for further processing. This trigger is called DeepCore *Simple Multiplicity Trigger 3* (SMT-3). The DOM hits that are read out in this process, but do not meet the HLC condition, are called *soft local coincidence* (SLC) hits. The rate of the DeepCore SMT-3 trigger is  $\sim$ 250 Hz [4], accepting  $\sim$ 70 % of  $\nu_\mu$ -CC events at 10 GeV and  $\sim$ 90 % at 100 GeV [2]. The trigger efficiencies for different SMT triggers, including the DeepCore SMT-3, are shown in Figure 1.1.

### Online Filter

The digitized waveforms are sent to the ICL, where a further filter is applied *online*<sup>1</sup>. First, the WaveDeform algorithm is run to extract photon arrival times and charge from the waveforms. Next, the DeepCore filter is applied, which is an iterative hit cleaning starting from HLC hits and removing any hits outside a 125 m radius and a 500 ns time window (called *radius-time cleaning (RT-cleaning)*) of the initial hit. This mainly rejects unphysical SLC hits, which are potentially caused by random noise. All following selection steps are done using the resulting cleaned pulses.

An additional cut is applied to reject events that are likely to be caused by atmospheric muons. This is done by splitting the hits depending on whether they were inside the DeepCore fiducial volume or outside and then calculating the speed of each hit outside the fiducial volume towards the *center of gravity* (COG) of the hits inside. If one of them has a speed close to the speed of light, the whole event is rejected, because this is a strong indication for a muon event.

As input for the further selection levels, several event properties, such as vertex position and direction, are determined using fast and simple event reconstructions. After the DeepCore online filter is applied, the data rate is about 15 Hz, which can be sent to the North via satellite for further processing.

### 1.1.2 Event Selection

After the data was sent to the North, the *offline* filters and selections are applied to further reduce the background of atmospheric muons and noise. The selection is split into three levels referred to as *Level 3-5*

(L3-L5), which bring down the neutrino and muon rate to  $\sim 1$  mHz, while the remaining fraction of random noise is below 1 %.

mention how this affects the HNL rate (in a general way) (RED)

### Level 3

At the first offline filtering level, Level 3, one-dimensional cuts are used to reduce atmospheric muons, pure noise, and coincident muons. These cuts are targeting regions where the data/MC agreement is poor, so that more sophisticated *machine learning* (ML) techniques can be applied at later levels. The cuts are made using 12 control variables, that are inexpensive to compute for the very large sample at this stage. The variables are related to position, time, and overall number of hits in the event.

Pure noise hits, that are temporally uncorrelated, are cleaned by applying a 300 ns sliding window, requiring the containment of more than 2 hits at its maximum. Additionally, an algorithm is run to check whether the hits show some directionality, accepting them only if they do.

To reduce the amount of muons a series of cuts is applied using spatial and temporal information. Events that have more than 9 hits observed above  $-200$  m or the first HLC hit above  $-120$  m are rejected as well as events where the fraction of hits in the first 600 ns of the event is above 0.37, ignoring the first two hit DOMs. Additionally, the ratio between hits in the veto region and the DeepCore fiducial volume is required to be below 1.5.

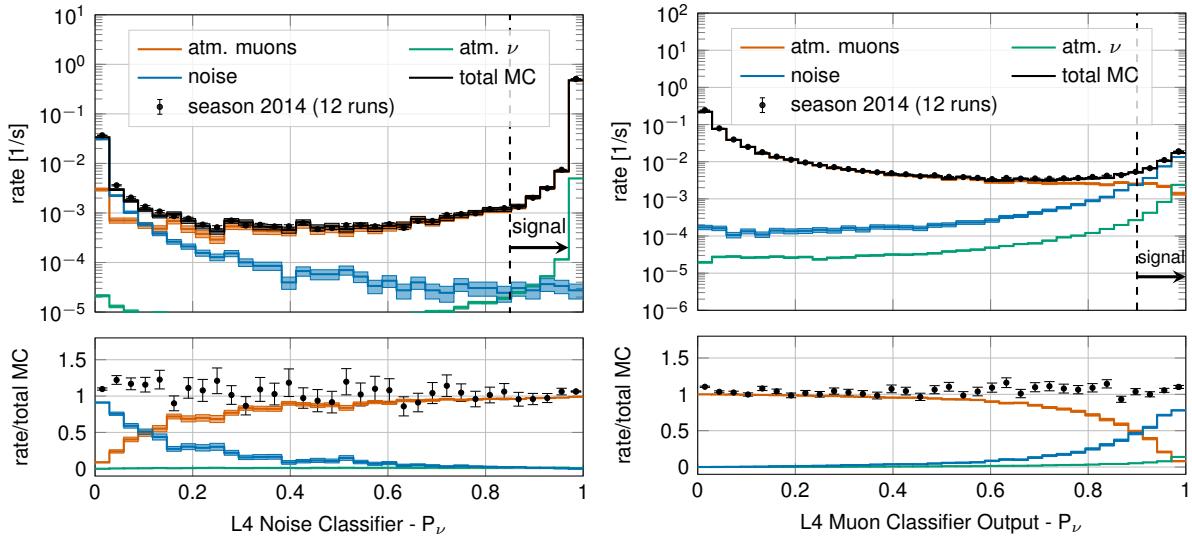
If a muon enters the detector after the data acquisition was already triggered, it causes events that span over a much larger time range. To reduce those coincident events, the time difference between first and last pulse cannot be above 5000 ns. This cut mainly affects a region of very poor data to MC agreement, because coincident events are not simulated at all.

The L3 cuts remove 95 % of the atmospheric muons and >99 % of pure noise hits, while keeping >60 % of the neutrino events. The sample now roughly contains muons/neutrinos/noise at a ratio of 100:10:1 with a total rate of  $\sim 0.5$  Hz.

### Level 4

After the total rate was reduced by the simple cuts of L3 and the overall agreement between data and MC is established, ML techniques can be applied to further reduce the background. For Level 4, two *Boosted Decision Trees* (BDTs) [5] classifier are trained to separate neutrino events from atmospheric muons and noise hits, separately. The output of each classifier, a probability score, can be seen in Figure 1.2. The noise filter is applied first and an event passes the score if it is larger than 0.7, reducing the noise hits by a factor of 100, while keeping 96 % of neutrinos. Then the second BDT classifier is applied to reject muons. It was trained partly on unfiltered data, which consists of >99 % atmospheric muons, to reject the data and keeping the neutrinos from the simulation. Rejecting events with a score smaller than 0.65 removes 94 % of atmospheric muons while keeping 87 % of neutrinos. This fraction varies depending on the flavor and interaction type,  $\nu_\mu$ -CC events for example, which have a muon in the final state, are therefore reduced to 82.5 %. After applying the L4 cuts based on the BDT classifier

[5]: Friedman (2002), "Stochastic gradient boosting"



**Figure 1.2:** Distributions of Level 4 noise classifier output (left) and muon classifier output (right), where larger values indicate more neutrino-like and lower values more noise-like/muon-like. Taken from [6].

outputs, the sample is still dominated by atmospheric muons, while the noise rate dropped to below most neutrino types.

### Level 5

Level 5 is the final selection level, before event reconstructions are applied. This level aims to reduce the remaining atmospheric muon rate below the rate of neutrinos. Muons not rejected by the earlier levels are those that produced little or no light in the veto regions. One possible reason is that they passed through one of the uninstrumented regions between the strings called *corridors*. To reject those, special corridor cuts are applied, which are based on the number of hits the event produced close to a potential corridor it passed through. The potential corridor in question is identified based on a simple infinite track fit. In addition to the corridor cuts, starting containment cuts are applied to reject events that start at the edge of the fiducial volume. Events with more than seven hits in the outermost strings of the detector or those that have a down-going direction in the uppermost region are rejected. This further reduces the fraction of muons by 96 % while keeping 48 % of neutrinos. The rates after this level are 1 mHz and 2 mHz for neutrinos and muons, respectively, making it a neutrino dominated sample.

add table with rates per level (split in flavor) - maybe better in analysis chapter to also show signal? (RED)

get estimate of HNL efficiency across the levels.. (RED)

## 1.2 Double Cascade Reconstruction

In the energy range relevant for this work, around 10s of GeV, the light deposition is very low and only a few DOMs detect light, making event reconstructions difficult. Existing reconstruction algorithms applied for low energy atmospheric neutrino events are either assuming a single cascade hypothesis or a track and cascade hypothesis, which are the two SM morphologies observable at these energies, as was described in Section ???. A HNL being produced and decaying inside the IceCube detector however, will produce two cascade like light depositions. The morphology, spatial

separation between the cascades, and their individual properties depend on the model parameters discussed in Section ???. To investigate the performance of the detector to observe and identify these events, a low energy double cascade reconstruction algorithm was developed. It is based on a pre-existing algorithm used to search for double cascades produced from high energy astrophysical tau neutrinos [7] that was established in [8], but first mentioned in [9].

### 1.2.1 Table-Based Minimum Likelihood Algorithms

The aforementioned reconstruction is relying on a minimum likelihood algorithm, which is the *classical* approach to IceCube event reconstructions, as opposed to ML based methods. It compares the observed light depositions in the detector to the expected light depositions from a given event hypothesis, where the event hypothesis can be constructed from building blocks of single cascade and track segment expectations. Varying the energies of the track segments and cascade components, will change the expected light and can be used to find the best fit to the observed light. A Poissonian likelihood is constructed, which compares the observed photon numbers,  $n$ , with their arrival times to the expected light depositions,  $\mu$ , for a given even hypothesis as

$$\ln(L) = \sum_j \sum_t n_{j,t} \cdot \ln(\mu_{j,t}(\Theta) + \rho_{j,t}) - (\mu_{j,t}(\Theta) + \rho_{j,t}) - \ln(n_{j,t}!) , \quad (1.1)$$

where  $\rho$  are the number of expected photons from noise,  $\Theta$  are the parameters governing the source hypothesis, and the likelihood is calculated summing over all DOMs  $j$  splitting observed photons into time bins  $t$ . The light expectations are calculated using look-up tables [10] that contain the results from MC simulations of cascade events or track segments. By varying the parameters defining the event hypothesis, the likelihood of describing the observed light pattern by the expected light depositions is minimized to find the reconstructed event. Algorithms of this kind used in IceCube are described in great detail in [11]. For the table production a specific choice of ice model has to be made, while the calibrated DOM information is taken from the measurement itself.

Based on the tabulated light expectations for cascades and track segments, various event hypothesis can be constructed, like the common cascade only or the track and cascade hypotheses. The hypothesis describing the double cascade signature of the HNL is using two cascades that are separated by a certain distance. The whole hypothesis is defined by 9 parameters and assumes that the two cascades are aligned with each other, which is a safe assumption for strongly forward boosted interactions. The parameters are the position of the first cascade,  $x, y, z$ , the direction of both cascades,  $\phi, \theta$ , and its time,  $t$ , as well as the decay length,  $L$ , between the two cascades. Assuming the speed of the HNL to be the speed of light,  $c$ , this already defines the full hypothesis, because the time and position of the second cascade are then fully determined by properties of the first cascade and the decay length. Note here, that the HNL particle does not produce any light while traveling, as it is electrically neutral. Since the likelihood only sums over DOMs that have observed photons, the non-observation of light is used as information and will exclude hypotheses with light expectation in those DOMs. The full 9 parameters describing the event are  $\Theta = (x, y, z, t, \theta, \phi, E_0, E_1, L)$ . To

[7]: Abbasi et al. (2020), “Measurement of Astrophysical Tau Neutrinos in IceCube’s High-Energy Starting Events”

[8]: Usner (2018), “Search for Astrophysical Tau-Neutrinos in Six Years of High-Energy Starting Events in the IceCube Detector”

[9]: Hallen (2013), “On the Measurement of High-Energy Tau Neutrinos with IceCube”

[10]: Whitehorn et al. (2013), “Penalized splines for smooth representation of high-dimensional Monte Carlo datasets”

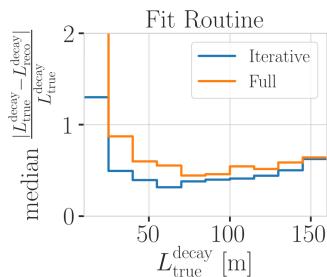
[11]: Aartsen et al. (2014), “Energy Reconstruction Methods in the IceCube Neutrino Telescope”

compute the full likelihood, the term in Equation 1.1 defined for a single event hypothesis, is summed over both cascade contributions, as  $\sum_i \ln(L_i)$ , with  $i$  being the cascade index.

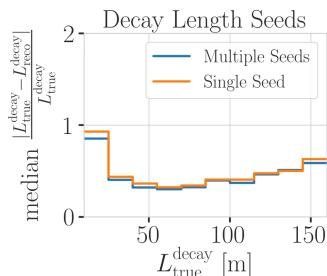
### 1.2.2 Optimization for Low Energies

Optimizing the double cascade reconstruction for low energy events was done in parallel to the development of the model dependent simulation generator introduced in Section ???. A preliminary sample of HNL events from the model dependent simulation was used, containing a continuum of masses between 0.1 GeV and 1.0 GeV and lab frame decay lengths sampled uniformly in the range from 5 m to 500 m. Even though this sample is not representative of a physically correct model and therefore not useful to predict the event expectation, it can still be used to optimize the reconstruction. The double cascade nature of the individual events is present and the evenly spaced decay length distribution is especially useful for the purpose of optimizing the reconstruction.

[12]: Abbasi et al. (2022), “Low energy event reconstruction in IceCube DeepCore”



**Figure 1.3:** Decay length resolution as a function of the true decay length, comparing a full 9 parameters fit to an iterative approach where first the energies and the decay length are fit, while fixing the other 7 parameters and then the full fit is performed.



**Figure 1.4:** Decay length resolution as a function of the true decay length, comparing the same fit routine seeded with just the seed decay length and seeded with a decay length of 5 m, 25 m, 50 m, 100 m, and 200 m on the left.

The simulation is processed up to Level 5 of the selection chain described in Section 1.1.2 and one of the reconstructions from [12] is applied to the events, fitting a cascade and a track and cascade hypothesis. The results from this reconstruction are used as an input for the double cascade reconstruction, where the position of the vertex, the direction of the event, and its interaction time are used as the input quantities for the first cascade, and the length of the track reconstruction is used as a seed for the distance between the two cascades.

#### Fit Routine

The full 9 dimensional likelihood space is very complex and can have many local minima, depending on the specific event and its location in the detector. For this reason, a more sophisticated fit routine than fitting all 9 parameters at once was tested. In a first fit iteration, some parameters are fixed and the resulting best fit point is used to fit all 9 parameters in a second iteration. The effect is shown in Figure 1.4, which shows the median of the absolute, fractional error with respect to the true decay length, as a function of the true decay length for a single length seed and multiple length seeds. It can be seen how a fit split into two consecutive steps, where the first step fits only both cascade energies and the decay length and the second step fits the full 9 parameters, performs better as compared to a single, full 9 parameter fit. The initial seed remains identical for both the routines.

#### Decay Length Seeds

From the seed values of the reconstruction, especially the length between the two cascades was found to have a very strong impact on whether the global minimum was found during the minimization. To mitigate this effect, multiple fits are performed, seeding with variations of the input length at different orders of magnitude. The best result is used, selected based on the total likelihood value of the best fit parameter set. A small improvement in the decay length resolution can be found by using this approach as compared to a single length seed.

## Minimizer Settings

To investigate the effect of the minimizer used to find the best fit parameters, the reconstruction was performed using three different minimizers, which were easily accessible within the reconstruction framework. The minimizers used were Minuit1 Simplex, Minuit2 Simplex, and Minuit2 Migrad [13, 14]. The initial idea was to test a global minimizer, or a routine that can find the rough position of the global minimum first and then a local minimizer to find the exact minimum, but unfortunately this was not possible with the minimizers available in the framework. As can be seen in Figure 1.5, Minuit1 Simplex performed best and was chosen as the default for the reconstruction.

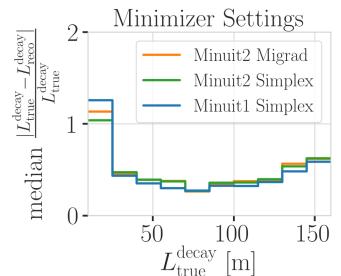
### 1.2.3 Performance

The chosen reconstruction chain used to test the performance of the detector to observe low energy double cascades is the following; Minuit1 Simplex is used as the minimizer, the decay length is seeded with 3 different values, 0.5x, 1.0x, and 1.5x the length of the preceding track reconstruction, and the fit routine is split into two steps, where the first step fits the energies and the decay length and the second step fits the full 9 parameters. In the first step, the number of time bins in Equation 1.1 is set to 1, so just the number of photons and their spatial information is used. The second step is seeded with the best results from the first step, and here the number of time bins is chosen such that each photon falls into a separate time bin, which means all time information is used. The average runtime per event is  $\sim 16$  s on a single CPU core, but is very dependent on the number of photons observed in the event, since the likelihood calculation in the second step scales with this number and a table lookup has to be performed for each photon.

To get a more realistic estimate of the reconstruction performance, it is run on a second preliminary sample of HNL events from the model dependent simulation, containing masses between 0.1 GeV and 3.0 GeV and the lab frame decay length is sampled from an inverse distribution in the range from 1 m to 1000 m, which is a better approximation of the expected exponential decay distribution of the HNL. The performance is shown for events where the reconstruction chain was successfully run, the event selection criteria up to the final selection level of low energy analyses are fulfilled, and the reconstructed energy of both cascades is above 3 GeV.

## Energy Resolutions

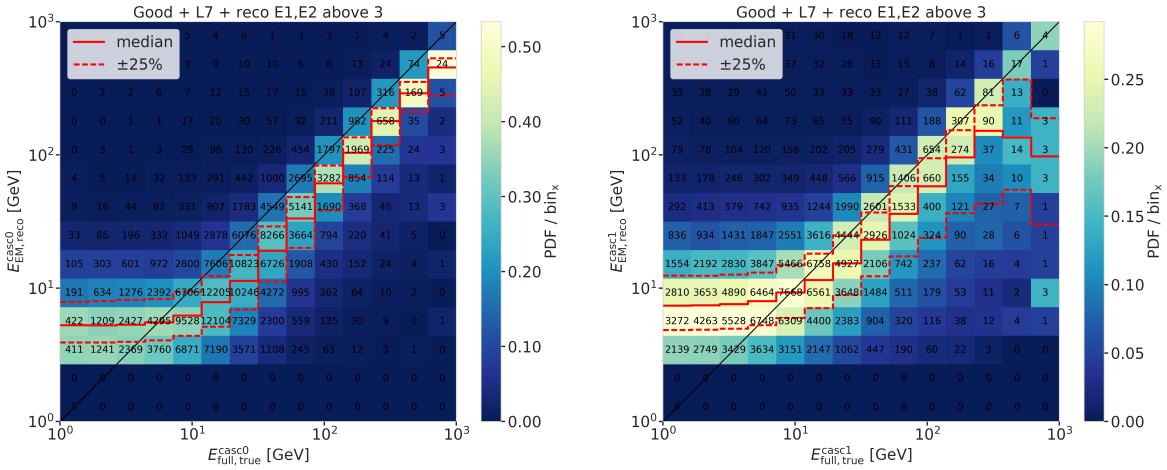
The energy resolution is inspected by looking at the two-dimensional distribution of reconstructed energy versus the true energy as shown in Figure 1.6. The bin entries are shown as well as the median and  $\pm 25\%$  calculated per vertical column, to get an idea of the distribution for a given energy slice. The color scale is showing the PDF along each true energy slice, which is the full information highlighted by the median  $\pm 25\%$  quantile lines. The reconstructed energy is only the energy that is observable from photons, while the true energy is the total cascade energy, including the parts that go into EM neutral particles that do not produce light. It is therefore expected that the reconstructed energy is lower than the true and the median therefore does not line up with the axis diagonal.



**Figure 1.5:** Decay length resolution as a function of the true decay length, comparing the same fit routine performed with different minimizers.

[13]: James et al. (1975), “Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations”

[14]: Dembinski et al. (2022), *scikit-hep/minuit*: v2.17.0



**Figure 1.6:** Reconstructed (EM) energy versus true energy (full) energy for the first cascade (left) and second cascade (right). The color scale is according to the PDF in each vertical true energy slice, with the solid and dashed lines showing the median  $\pm 25\%$  quantiles. The bin entries are shown as numbers.

The histogram for the first cascade energy is shown on the left and above an energy of  $\sim 10$  GeV the reconstruction performs well, with the median being parallel to the diagonal and the spread being small. Below this energy the reconstruction is over-estimating the true energy, because events that enter the sample are events with an over fluctuation in their light deposition, which makes them pass into the selection and being reconstructible in the first place.

For the second cascade the overall behavior is similar, only that the energy where the reconstruction starts to perform well is higher around  $\sim 20$  GeV. The spread around the median is also larger and starts to expand a lot above 200 GeV, where the statistics are lower as can be seen from the bin counts. It is also very apparent that the majority of events have a lower true energy in the second cascade, peaking between 1 GeV and 20 GeV. This can be seen by the indicated bin counts in the right part of Figure 1.6.

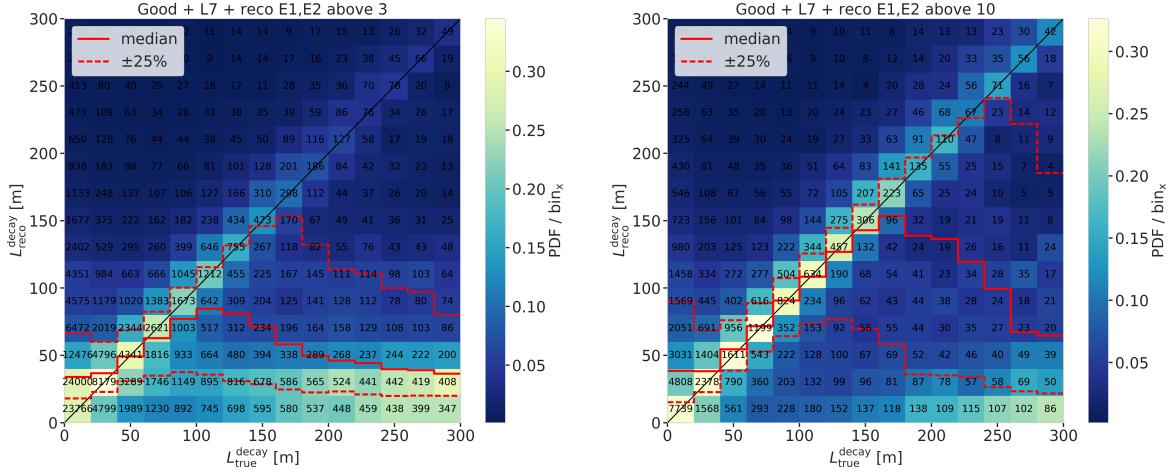
For both cascade resolutions the effect of the reconstruction being biased towards lower values can be seen. This is due to the comparison of the full true energy to the reconstructed EM energy as mentioned before.

### Length Resolutions

The decay length resolution is also investigated by looking at the two-dimensional histogram, where the reconstructed decay length is plotted versus the true decay length. The left part of Figure 1.7 shows the distributions after the same selection criteria from Section 1.2.3 are applied. It can be observed that for short true lengths the reconstruction is over-estimating the length, while for long true lengths the reconstruction is strongly under-estimating the length. There is a region between true lengths of 20 m and 80 m where the median reconstruction is almost unbiased, but the 50 % interquartile range is large and increasing from  $\sim 50$  m to  $\sim 70$  m with true decay lengths.

blow up figures to make better visible? (ORANGE)

The over-estimation at small true lengths can be explained by multiple factors, one being that the shortest DOM spacing is  $\sim 7$  m, vertically for



**Figure 1.7:** Reconstructed decay length versus true decay length for  $\sim 3$  GeV (left) and  $\sim 10$  GeV (right) minimum reconstructed cascade energies. The color scale is according to the PDF in each vertical true length slice, with the solid and dashed lines showing the median  $\pm 25\%$  quantiles. The bin entries are shown as numbers.

DeepCore strings, but mostly larger than that, so resolving lengths below this is very complicated, and the reconstruction tends to be biased towards estimating the length around where the light was observed. Additionally, approaching a length of 0.0, the reconstructed length will of course always be a one-sided distribution, because the lengths have to be positive.

The under-estimation at large true lengths is more puzzling, and it seems like the distribution becomes bimodal in the reconstructed lengths, with one population around the diagonal, meaning that they are properly reconstructed, and another population at very short reconstructed lengths, which are badly reconstructed. Above 150 m the badly reconstructed population starts to dominate, and the median resolution drops off strongly. The assumption is that for these events, only one cascade was observed with enough light to be reconstructed, and the reconstruction describes the one observed cascade in two parts, separated by a short distance, driven by similar factors as mentioned before. A quick check to confirm whether this is the case, was to increase the selection criteria to minimum reconstructed cascade energies of 10 GeV, which is shown in the right part of Figure 1.7. It can be seen that the median resolution is already much better, aligning with the expectation between 40 m and 160 m. Judging from the median resolution and the spread in this range, there are very few events with an over-expectation in the energy, since both of them are aligning with the diagonal. Towards lower reconstructed lengths on the other hand, the spread is still very large, and above 200 m the badly reconstructed population starts to dominate again.

### Badly Reconstructed Cascade Population

To investigate the badly reconstructed population further, a rough separation was made to find out what the cause of the difference is. It was already established that a larger reconstructed energy in both cascades, which is related to a larger true energy in form of more deposited light, leads to a better reconstruction in more events. To select the two populations, only events with true decay length larger than 80 m are used as shown in Figure

remake plot: no bin counts, blow up labels etc. margin possible? (RED)

fix caption (RED)

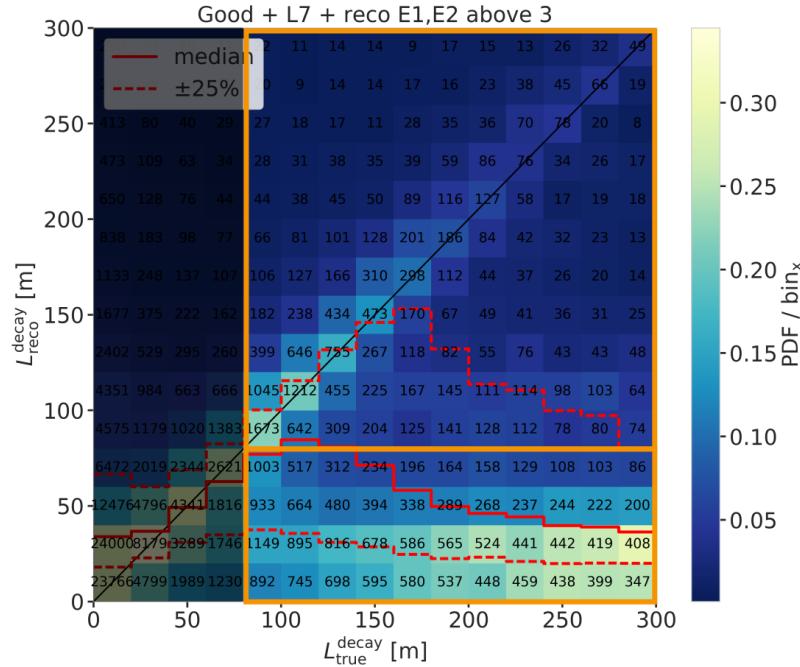


Figure 1.8

1.8, and the populations are split by the reconstructed decay length being larger or smaller than 80 m. To investigate the difference between the two populations, several variables were compared to find the reason(s) for the bad reconstruction.

The left part of Figure 1.9 shows the true horizontal distance of the second cascade from string 36. The distance is denoted as  $\rho_{36}$  and is a very good proxy for the distance to the center of the detector, because string 36 is almost at the center. While the distributions looks very similar for the first cascade (not shown), for the second cascade the badly reconstructed population extends to larger values. Considering that the DeepCore strings are roughly inside a 70 m radius from the center, and the next layer of IceCube strings is at a radius of 125 m, this is a plausible explanation for a worse reconstruction, because for the badly reconstructed population the second cascades are more often in regions without DOMs, so less or no light is observed from them.

Another possible reason why the reconstruction underperforms could be that the initial seed direction itself was off and therefore one of the cascades cannot be found properly. Looking at the error of the cosine of the reconstructed zenith angle shown in the right of Figure 1.9, we see that the badly reconstructed population has a larger error, and is less peaked around 0.0. This could be a hint that the direction is worse for the badly reconstructed population, which could be due to a bad seed direction, or just the result of one cascade not depositing enough light to be observed.

The true energies of both cascades are shown in Figure 1.10, where it can be observed that the first cascade energy is generally much larger than the second, peaking between 10 GeV and 20 GeV, while the second cascade peaks below 10 GeV. For the first cascade there is no significant difference between the two populations, but for the second cascade the badly reconstructed population has a larger fraction of events with lower energies and the distribution is almost uniform in the range of 2 GeV to 10 GeV, while the

re-make normalized?  
(ORANGE)

fix caption (RED)

re-make normalized?  
(ORANGE)

fix caption (RED)

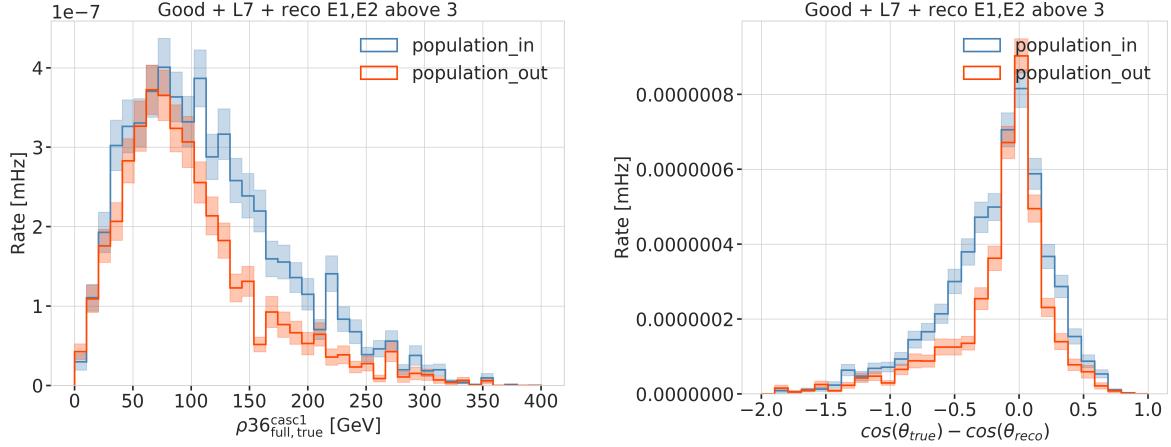


Figure 1.9

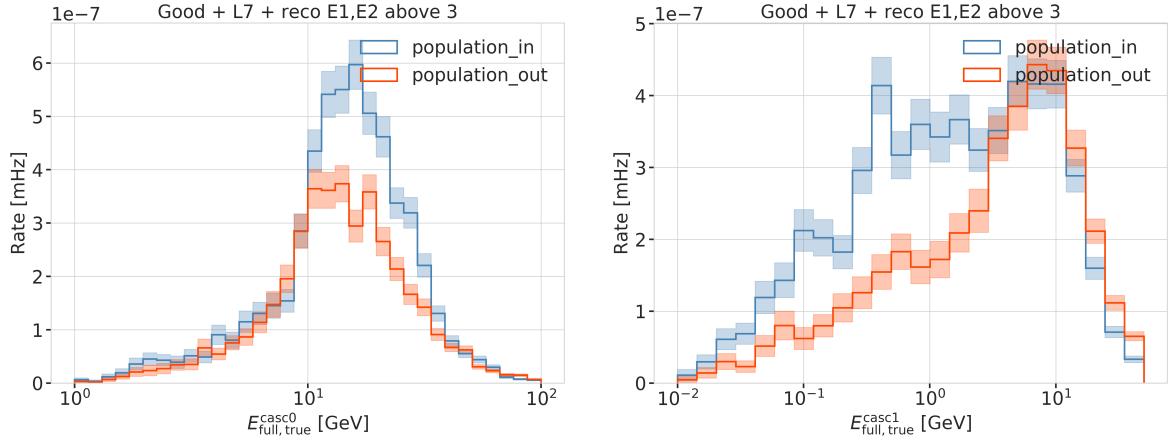


Figure 1.10

well reconstructed population has a peak around 10 GeV and falls off faster towards lower energies. This is a strong indication that the main reason for the bad reconstruction is the low energy of the second cascade.

### 1.3 Double Cascade Classification

Even though the performance results show that it is very complicated to reconstruct these low energy double cascade events, the attempt to identify them in the background of SM neutrino events was made. For this purpose a classifier was trained to distinguish between HNL *signal* events and SM neutrino *background* events, using the same preliminary sample of HNL events as was used to assess the reconstruction performance. To mitigate the effect of the bad reconstruction, a set of cuts was applied to make sure the classifier is trained on well reconstructed events. The cuts are a minimum reconstructed energy of both cascades of 5 GeV and a minimum reconstructed decay length of 40 m. and they are applied to both signal and background events.

fix caption (RED)

Additionally, some cuts on the true energies and decay length were applied

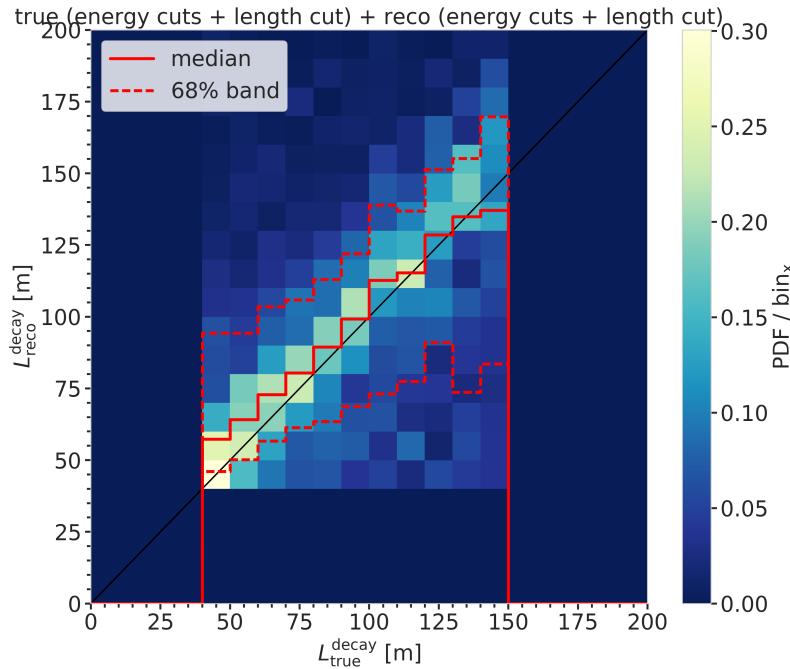


Figure 1.11

for the signal, which are a minimum true energy of both cascades of 5 GeV, and a true decay length between 40 m and 150 m. These were chosen to make sure the HNL events were theoretically double cascade like and at a sensible length scale inside DeepCore. Figure 1.11 shows the decay length two-dimensional histogram after the cuts were applied.

[15]: Pedregosa et al. (2011), “Scikit-learn: Machine Learning in Python”

The classifier used was a *Boosted Decision Tree (BDT)* from the *SCIKIT-LEARN* (*sklearn*) package [15] and the input features are taken from the double cascade reconstruction explained in Section 1.2 as well as some additional variables from earlier levels of the processing explained in Section 1.1. Figure 1.12 shows the distributions of two example input features, where the left plot shows the output probability of the classifier trained to distinguish track from cascade like events, which is used in the oscillation analysis, and the right plot shows the reconstructed decay length from the double cascade reconstruction. Shown are the distributions for the HNL signal, the individual background components, and the total background.

A single classifier and a combination of two classifiers were tested. The single classifier was trained to distinguish between HNL signal events and all SM background events at once. The two classifiers were trained separately, one to distinguish signal from track like background, and the other to distinguish signal from cascade like background. Since the SM neutrino events at these energies are either track like or cascade like, the latter approach was expected to perform better. Despite the fact that several combinations of features and classifier hyperparameters were tested, it was not possible to identify a pure double cascade region with a single classifier.

By applying the two classifiers trained to distinguish signal from track and signal from cascade, it is possible to select a region with only signal events. This is visualized in Figure 1.13, where the probabilities of 1 implies very signal like, and only the regions close to 1 are shown for both outputs, to highlight the interesting region, where a pure HNL sub-sample can be selected. When physical weights are applied to those signal events however,

fix caption of this figure  
(RED)

Make a single plot, showing S/B or so? (RED)

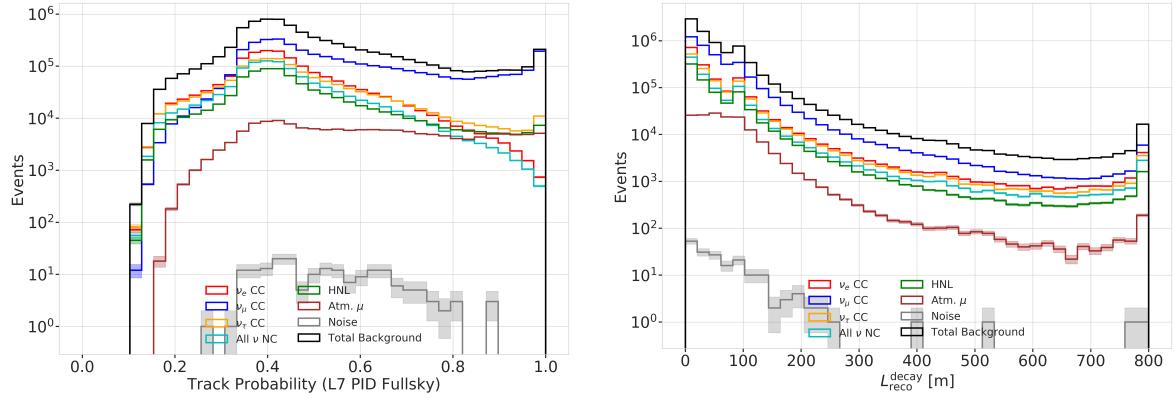


Figure 1.12

the expected event rate is very low, and even by assuming a highly optimistic mixing of 1, it would take more than 20 years of data taking to observe a single event. Additionally, with this low simulation statistics the prediction is not very reliable, either. Making a weaker cut to select a signal like region will contain a large amount of background events, which dominate over the signal at  $\sim 2$  orders of magnitude for a mixing of 0.1. The conclusion from this is, that with the current selection and reconstruction chain and a classical BDT, it is not possible to distinguish signal events at a level feasible to perform an analysis.

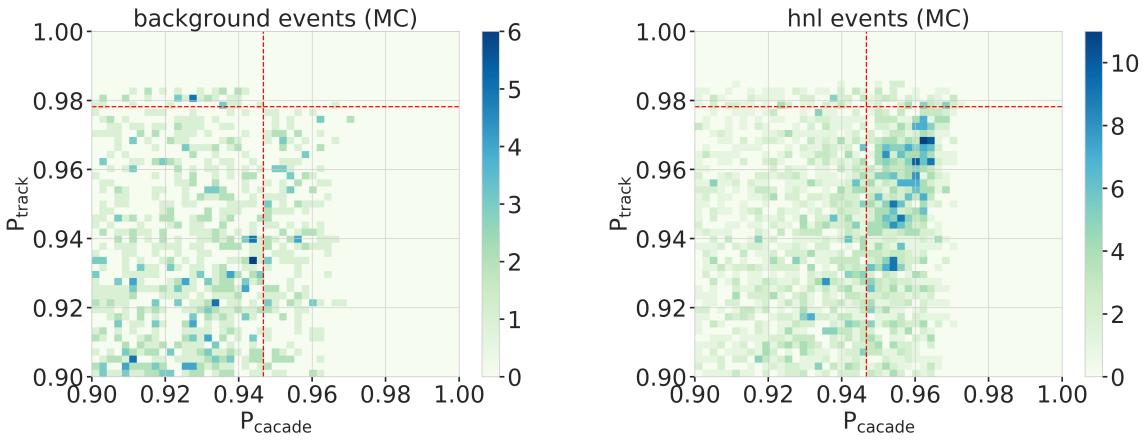


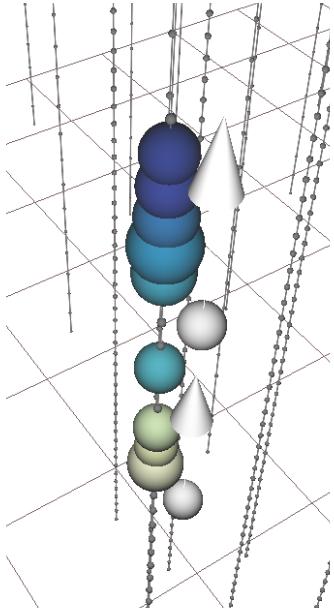
Figure 1.13

fix caption of this figure  
(RED)

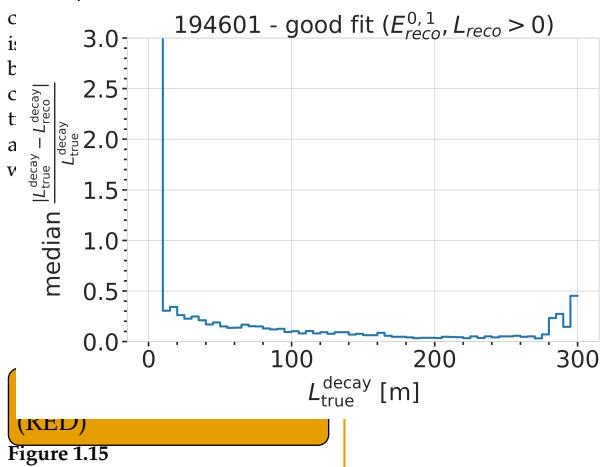
## 1.4 Generalized Double Cascade Performance

All the above results were obtained using preliminary development versions of the model dependent HNL simulation. To investigate the effect of the low energy event selection and the double cascade reconstruction performance in a more generic way, the model independent simulation introduced in Section ?? is used to repeat the performance checks and to run a series of additional checks. The important advantage of the model independent samples is the controllable parameter space, especially in cascade energies and decay length, because the event kinematics are not coupled to the underlying HNL

model, but can be chosen freely. This means that some benchmark edge cases can be investigated, and the performance can also be assessed for a realistic scenario in addition to mapping out the effects of the event selection and where the reconstruction breaks down.



**Figure 1.14:** Event view of an idealistic double cascade event, with cascade energies of 2.4 GeV and 4.9 GeV, and a decay length of 65.8 m. The colored spheres show the DOMs that have observed light, where the size is proportional to the number of observed photons and the color indicates the energy level.

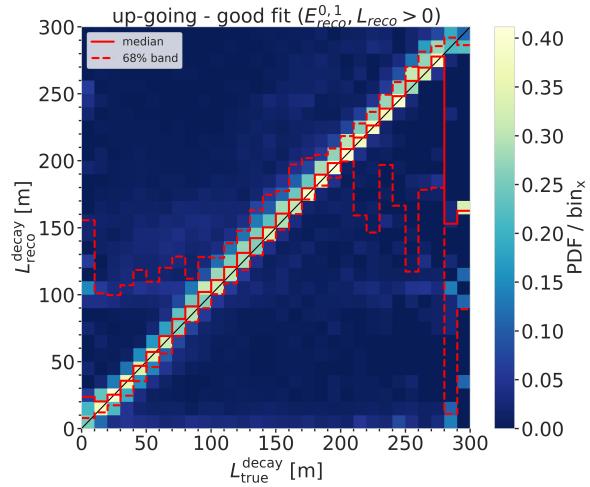


**Figure 1.15**

### 1.4.1 Idealistic Events

The *best case* scenario to observe an event is to be directly on top of a string with a straight up-going direction. Using the simulation sample introduced in Section ?? and running the double cascade reconstruction from Section 1.2 on these events, it is possible to estimate the performance limit of the reconstruction. Figure 1.14 shows one example event view from that sample, where the cascade energies are 2.4 GeV and 4.9 GeV, and the decay length is 144.5 m. It can be seen that despite the low energies, both cascades deposit light in the DOMs and the reconstruction is expected to work.

The performance of the length reconstruction is shown in Figure 1.15, where the median of the absolute, fractional decay length resolution is shown on the left and the two-dimensional histogram of the reconstructed versus the true decay length is shown on the right. For these results and the following, all events that were reconstructed with non-zero cascade energies and non-zero decay length are used, and the events are unweighted. The length is very well reconstructed, with the median resolution being below 30 % above a true decay length of ~10 m, and falling off with increasing true length, down to ~10 % at 100 m.



The two-dimensional histogram shows that there is no under-estimation of the length up to a true decay length of ~210 m, which shows that if there are DOMs in the region between the two cascades that have not observed any light, the reconstruction is very stable. Considering the underlying Poisson likelihood in Equation 1.1 used for the reconstruction, this makes sense, since DOMs being present, but not observing any light is affecting the light expectation that goes into the likelihood and therefore makes these hypotheses unlikely and therefore incompatible with the data.

### 1.4.2 Realistic Events

The sample of HNL events introduced in Section ??, which is a more realistic representation of the expected HNL events, but still offers more controlled energy and length distributions, is used to investigate the selection efficiency, to cross check the reconstruction performance, and to benchmark the limits where the reconstruction breaks down. An example event view is shown in Figure 1.16, for cascade energies of 30.8 GeV and 25.3 GeV, and a decay length of 144.5 m. Since the size of the colored spheres is proportional to the number of photons observed in the DOMs, it can be seen from the event view that even for these higher energies, only individual or few photons are observed. This makes detecting and reconstructing them significantly more challenging and is purely due to the larger distance of the cascades from the DOMs.

To assess the efficiency of the low energy event selection introduced in Section 1.1, the energy and length distributions are shown across the different selection levels in Figure ?? . Table ?? shows the total efficiency of the selection, where it can be seen that at level xx it is reduced the most and only xx% of the events pass the selection to level 5.

The energy distributions in Figure 1.17 show a similar behavior to the results discussed in Section 1.2.3. The difference is that now, there is no bias in the reconstructed energy, because the events are simulated as EM cascades, which means all energy is deposited in light and can be reconstructed. Above around 5 GeV to 6 GeV the median is very stable, and the 1-sigma resolution band is 50 % narrow and decreasing with energy down to 20 % at 100 GeV. Interestingly, the second cascade energy reconstruction performs slightly worse, although they have the same energy ranges for this sample. This could hint at an asymmetry in the reconstruction process, which might relate to how the two cascades are parameterized, or be due to the different positions and the dominantly up-going direction used in the sampling combined with the DOMs looking down. The total energy resolution shown in the left part of Figure 1.18 is very good, above 10 GeV it is unbiased and the 1-sigma resolution band is below 20 %.

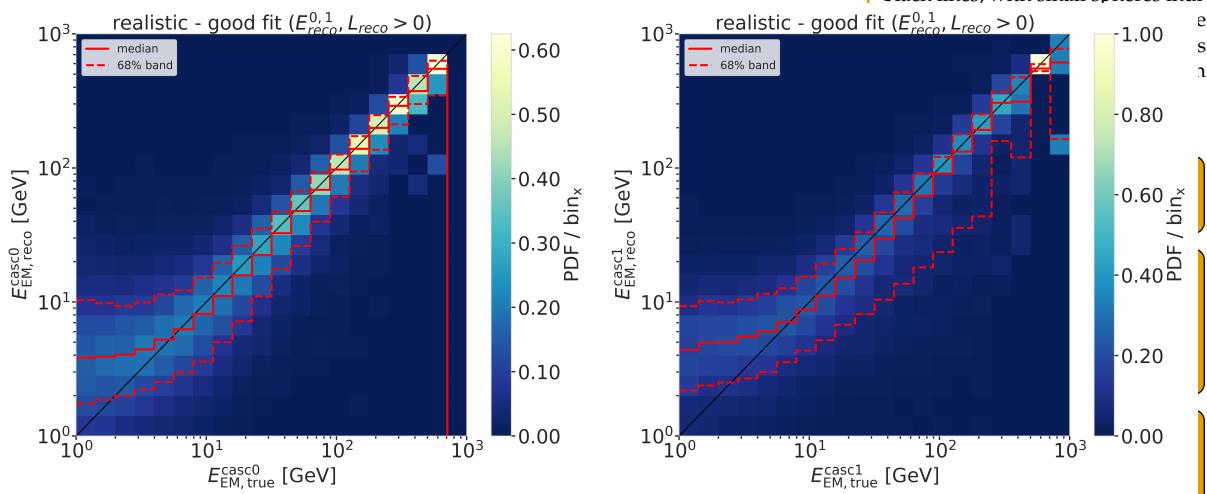


Figure 1.17

The decay length resolution shown in the right part of Figure 1.18 looks similarly bad to the results discussed in Section 1.2.3 and shows the same

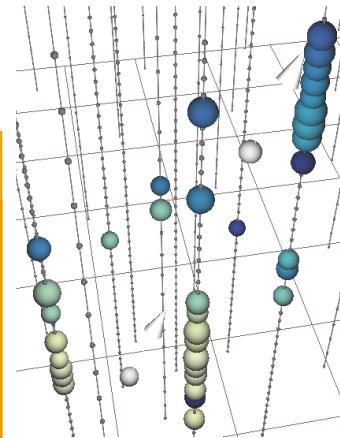
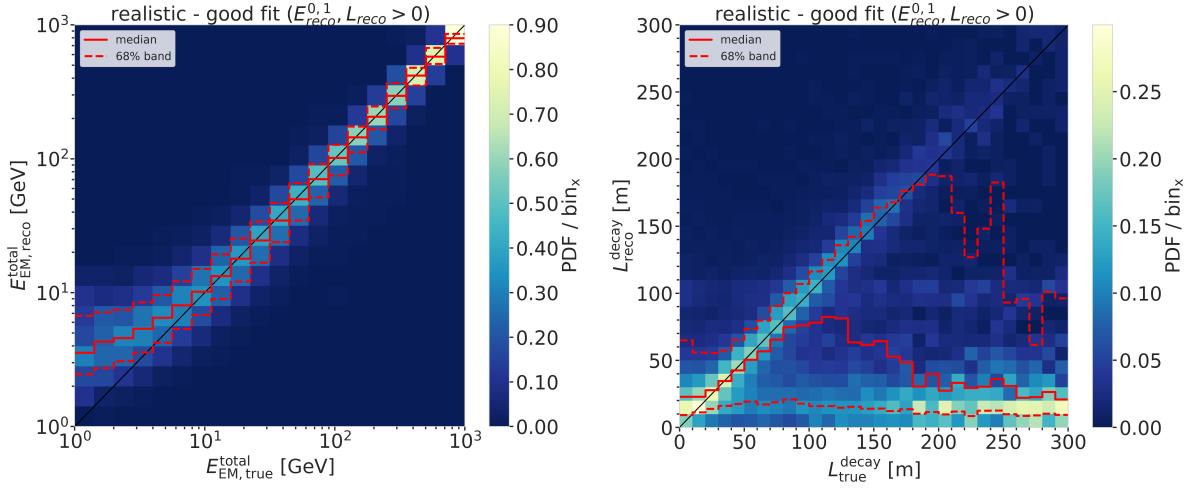


Figure 1.16: Event view of a realistic double cascade event, with cascade energies of 30.8 GeV and 25.3 GeV, and a decay length of 144.5 m. The colored spheres show the DOMs that have observed light, where the size is proportional to the number of observed photons and the color indicates the time (yellow is early, blue is late). The strings are shown as black lines, with small spheres indicating

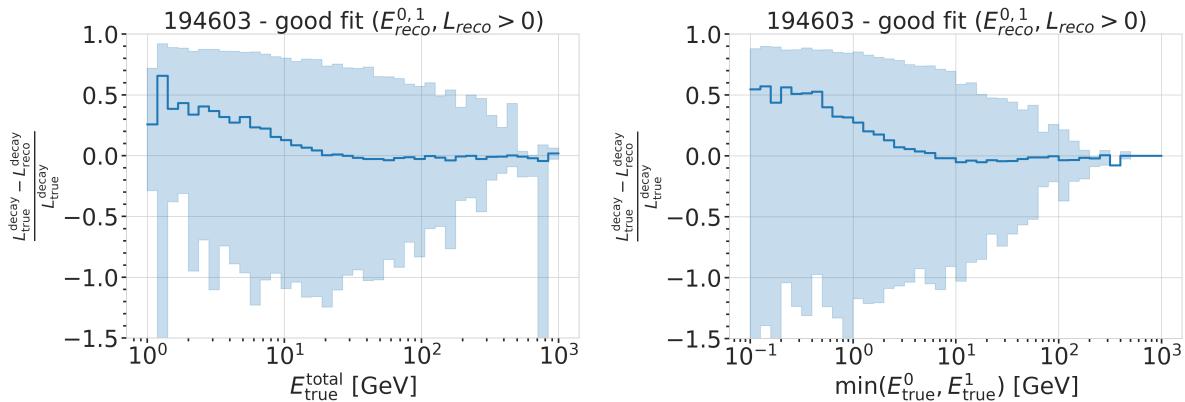
TOP BENCHMARKING (RED)  
fix caption of this figure (RED)

features with a region between 20 m and 80 m where it is roughly unbiased, but the 1-sigma resolution band is wide with a lot of outliers towards short reconstructed lengths. Below 65.8 m the reconstructed lengths are always over-estimating the true and above 80 m a population of events start to dominate where the decay lengths is not getting reconstructed at all, as investigated before.



**Figure 1.18**  
fix caption of this figure  
(RED)

To get an estimate of what minimum energies are necessary for the reconstruction to perform reasonably well, the fractional decay length resolution is shown as a function of the total true energy and the minimum energy of both individual cascades in Figure 1.19. In the left part it can be seen that the median of the decay length resolution stabilizes around 0 for a total energy above 20 GeV, but the spread of the distribution is still quite large with a 1-sigma band of 80 % to 100 %, decreasing down to ~60 % at 100 GeV. Based on the right part of the figure, the decay length resolution starts to be unbiased for a minimum energy of any cascade of 7 GeV, with an equivalently large spread. A rough takeaway from this is that the decay length reconstruction is not reliable for events with one cascade energy below 7 GeV and with a total energy below 20 GeV. Above these values the median resolution is roughly unbiased, but the spread is still large, decreasing with increasing energy.



**Figure 1.19**  
fix caption of this figure  
(RED)

From the investigation of the good and badly reconstructed events, it can be

concluded that the main reason for the bad reconstruction is the low energy of the second cascade. Despite the fact that the split into the two populations was very rudimentary, it is clear that this is the dominant cause, while other factors, like the position of the second cascade, or the potentially bad input seed direction are also contributing. For a thorough investigation, a more sophisticated separation would be needed.

## 1.5 Analysis Reconstruction

In contrast to the classical reconstruction methods described [12], which were applied in one recent IceCube atmospheric neutrino oscillation measurement using a sub-sample of the DeepCore sample [6], the reconstruction algorithm used in this work is a method that applies a *convolutional neural network* (CNN). It is both used to reconstruct the events properties and to determine some discriminating quantities. The latest muon neutrino disappearance result from IceCube [16] is based on this reconstruction.

### 1.5.1 Fast Low-Energy Reconstruction using Convolutional Neural Networks

As the name *Fast Low-Energy Reconstruction using Convolutional Neural Networks* (FLERCNN) already indicates, the FLERCNN reconstruction [17] [18] is a CNN optimized to reconstruct IceCube events at low energies (<100 GeV) in a fast and efficient manner, by leveraging the approximate translational invariance of event patterns within the detector. The architecture of the network is very similar to the preexisting IceCube CNN event reconstruction [19], but optimized on low-energy events and specifically tailored to include the DeepCore sub-array. Only the eight DeepCore strings and the central 19 IceCube strings are used for the reconstruction (compare to Figure ??). Because of the different z-positions of the DeepCore and IceCube DOMs, they are divided into two networks that are combined in the final layer of the network. The full architecture is shown in Figure 1.20. The first dimension of the network is the string index, while the second dimension is the order of the DOMs along the vertical axis. The horizontal position of the DOMs is not used, since the strings are arranged in an irregular pattern. The information from the DOM hits is summarized into five charge and time variables, which make up the last dimension of the input layer. The variables are the total summed charge, the time of the first hit, the charge weighted mean time of the hits, the time of the last hit, and the charge weighted standard deviation of the hit times.

Five different networks are trained using this architecture. Three networks do the regression of the events' energy, cosine of the zenith angle, and the starting vertex ( $x, y, z$  position), while two of them are used for classification. One is trained to predict the probability of the event being a  $\nu_\mu$ -CC event and the other to predict the probability of the event being an atmospheric muon. Each network is trained with an MC sample modified to have a flat distribution in the target variable, to be unbiased for that variable and ideally extending outside the target reconstruction region. For the classification tasks the loss function is the *binary cross entropy* and the activation function is a *sigmoid*. To perform the regression of zenith and vertex position, the loss

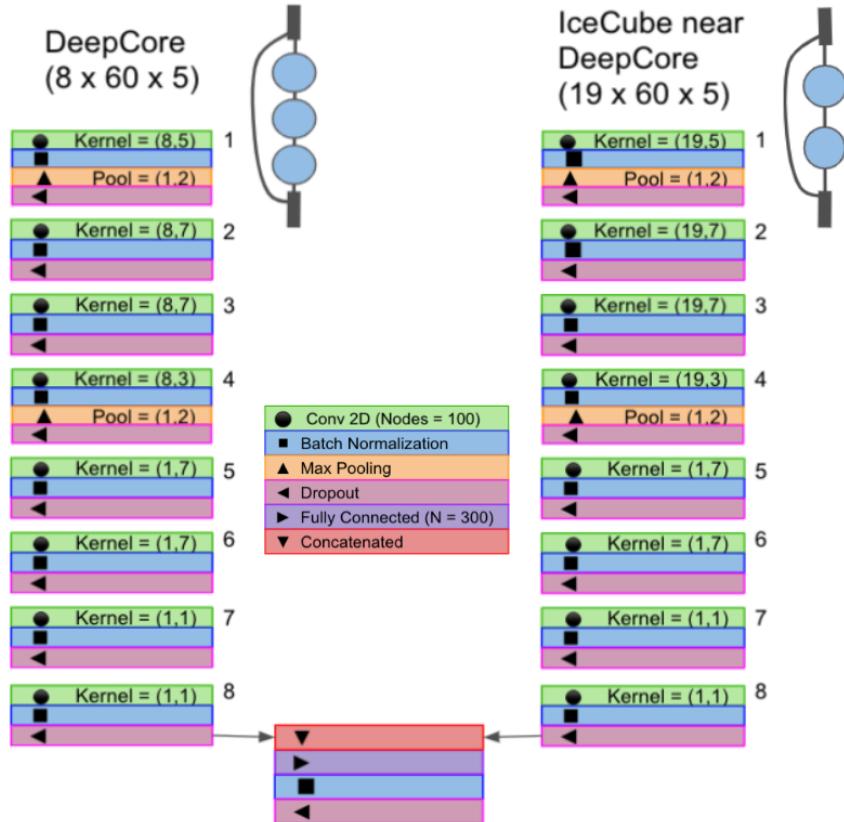
[12]: Abbasi et al. (2022), "Low energy event reconstruction in IceCube DeepCore"

[6]: Abbasi et al. (2023), "Measurement of atmospheric neutrino mixing with improved IceCube DeepCore calibration and data processing"

[16]: Yu et al. (2023), "Recent neutrino oscillation result with the IceCube experiment"

[17]: Yu et al. (2021), "Direction reconstruction using a CNN for GeV-scale neutrinos in IceCube"

[19]: Huenefeld (2017), "Deep Learning in Physics exemplified by the Reconstruction of Muon-Neutrino Events in IceCube"



**Figure 1.20:** Architecture of the FLERCNN neural networks, taken from [17].

add a statement about overtraining using the plots showing the FLERCNN muon classifier (RED)

add some performance plots of the FLERCNN reconstruction (RED)

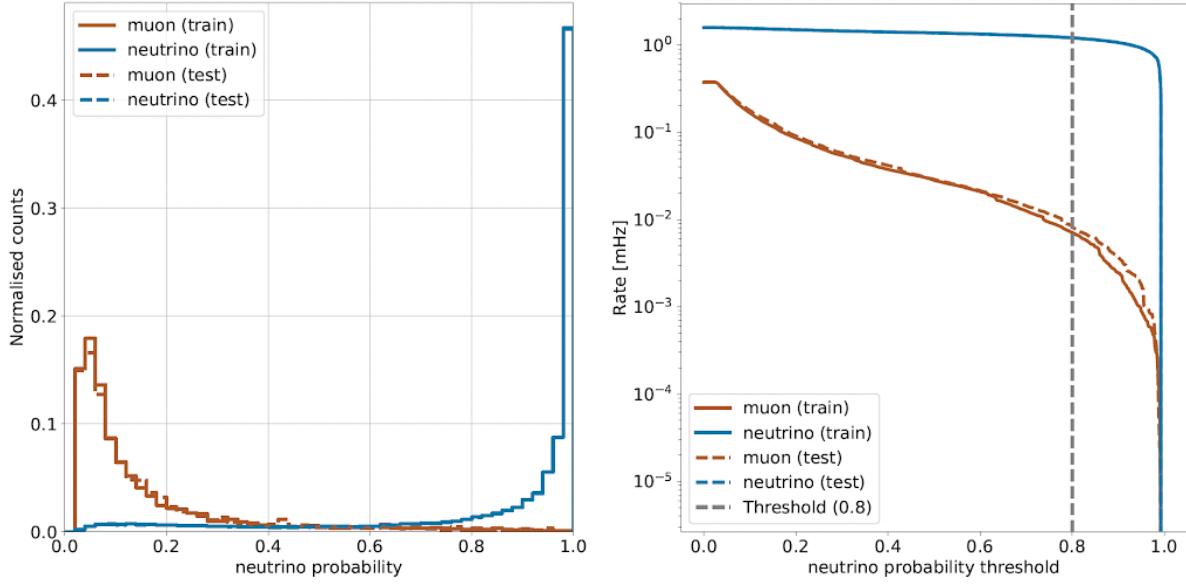
2: A radial variable that is often used in IceCube, is the horizontal distance to string 36 called  $\rho_{36}$ , which is basically the distance to the center of IceCube.

function is the *mean squared error (MSE)*, while for the energy it is the *mean absolute percentage error*. The activation for all regression tasks is *linear*.

### 1.5.2 Analysis Selection

After the FLERCNN reconstruction is applied, a BDT classifier is used to further reduce the muon background for the final sample. The BDT is trained on five high level variables, where three are FLERCNN reconstruction variables (vertex  $z$ ,  $\rho_{36}^2$  and muon probability), and two are lower level variables (L4 muon classifier output and L5 corridor cut variable). To train the BDT, the FLERCNN nominal simulation set is used, only using events with  $\cos(\theta_{\text{zenith}}) \leq 0.3$ . The output of the BDT is the neutrino probability and a cut at 0.8 is applied to reject events with a high probability of being a muon. Figure 1.21 shows the output of the BDT classifier, where the neutrinos in both training and testing sets are gathered at 1 and muons are around 0, which shows great classification power.

To get the final, pure sample of well reconstructed neutrinos another set of cuts is applied. The first cuts are meant to reject events with poor reconstruction quality, by requiring the events to fall into the DeepCore volume, where the denser, better instrumented detector leads to enhanced resolution. The cuts are applied on the vertex  $z$  and  $\rho_{36}$  and are listed in Table 1.1. The FLERCNN reconstruction was optimized for atmospheric neutrino analyses which are mainly in the region below 100 GeV and there are very few events with energies below 5 GeV, so the reconstructed energy is required to be in that range. Additionally, rejecting events with fewer than



**Figure 1.21:** FLERCNN muon classifier output score (left) and rate of neutrinos and muons as function of muon classifier cut (right).

Variable	Threshold	Removed
Number of hit DOMs	$\geq 7$	1.05 %
Radial distance	$< 200 \text{ m}$	0.09 %
Vertical position	$-495 \text{ m} < z < -225 \text{ m}$	5.48 %
Energy	$5 \text{ GeV} < E < 100 \text{ GeV}$	20.70 %
Cosine of zenith angle	$< 0.04$	19.66 %
Number of direct hits	$> 2.5$	10.50 %
Number of hits in top layers	$< 0.5$	0.03 %
Number of hits in outer layer	$< 7.5$	0.001 %
Muon classifier score	$\geq 0.8$	23.90 %

**Table 1.1:** Cuts performed to select the final analysis sample. Parts of the cuts are meant to increase the data/MC agreement, while others are meant to reject events with poor reconstruction quality.

seven hits in the selected DOMs used for FLERCNN showed to increase the resolution.

Another set of cuts is applied to make sure the agreement between data and MC is good. To remove coincident muon and neutrino events, cuts are applied to the number of hits in the top 15 layers of IceCube DOMs and the number of hits in the outermost IceCube strings. Coincident random noise events are removed by requiring more than three hit DOMs from direct photons<sup>3</sup> [12]. Neither of the two coincident event types are simulated, which can be seen as bad agreement between data and MC. The last cut is on the reconstructed cosine zenith, which is required to be smaller than 0.04 to reject down-going muons.

3: *Direct photons* are photons that were not scattered on their way from the interaction vertex to the DOM.  
[12]: Abbasi et al. (2022), “Low energy event reconstruction in IceCube DeepCore”



# List of Figures

1.1	IceCube trigger efficiencies . . . . .	2
1.2	Level 4 classifier outputs (muon and noise) . . . . .	4
1.3	Decay length resolution to optimize fit routine . . . . .	6
1.4	Decay length resolutions to optimize length seeding . . . . .	6
1.5	Decay length resolution to optimize minimizer settings . . . . .	7
1.6	Preliminary two-dimensional reconstructed versus true cascade energy resolutions . . . . .	8
1.7	Preliminary two-dimensional reconstructed versus true decay length resolutions . . . . .	9
1.14	Event view of an idealistic double cascade event . . . . .	14
1.16	Event view of a realistic double cascade event . . . . .	15
1.20	FLERCNN architecture . . . . .	18
1.21	FLERCNN muon classifier probability distributions . . . . .	19



# List of Tables

1.1 Final analysis cuts . . . . .	19
-----------------------------------	----



# Bibliography

Here are the references in citation order.

- [1] M. G. Aartsen et al. "The IceCube Neutrino Observatory: Instrumentation and Online Systems". In: *JINST* 12.03 (2017), P03012. doi: [10.1088/1748-0221/12/03/P03012](https://doi.org/10.1088/1748-0221/12/03/P03012) (cited on page 1).
- [2] R. Abbasi et al. "The design and performance of IceCube DeepCore". In: *Astropart. Phys.* 35.10 (2012), pp. 615–624. doi: [10.1016/j.astropartphys.2012.01.004](https://doi.org/10.1016/j.astropartphys.2012.01.004) (cited on page 2).
- [3] R. Abbasi et al. "The IceCube data acquisition system: Signal capture, digitization, and timestamping". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 601.3 (2009), pp. 294–316. doi: <https://doi.org/10.1016/j.nima.2009.01.001> (cited on page 2).
- [4] M. G. Aartsen et al. "The IceCube Neutrino Observatory: instrumentation and online systems". In: *Journal of Instrumentation* 12.3 (Mar. 2017), P03012. doi: [10.1088/1748-0221/12/03/P03012](https://doi.org/10.1088/1748-0221/12/03/P03012) (cited on page 2).
- [5] J. H. Friedman. "Stochastic gradient boosting". In: *Computational Statistics & Data Analysis* 38 (2002), pp. 367–378 (cited on page 3).
- [6] R. Abbasi et al. "Measurement of atmospheric neutrino mixing with improved IceCube DeepCore calibration and data processing". In: *Phys. Rev. D* 108 (1 July 2023), p. 012014. doi: [10.1103/PhysRevD.108.012014](https://doi.org/10.1103/PhysRevD.108.012014) (cited on pages 4, 17).
- [7] R. Abbasi et al. "Measurement of Astrophysical Tau Neutrinos in IceCube's High-Energy Starting Events". In: *arXiv e-prints* (Nov. 2020) (cited on page 5).
- [8] M. Usner. "Search for Astrophysical Tau-Neutrinos in Six Years of High-Energy Starting Events in the IceCube Detector". PhD thesis. Berlin, Germany: Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät, 2018. doi: [10.18452/19458](https://doi.org/10.18452/19458) (cited on page 5).
- [9] P. Hallen. "On the Measurement of High-Energy Tau Neutrinos with IceCube". MA thesis. Aachen, Germany: RWTH Aachen University, 2013 (cited on page 5).
- [10] N. Whitehorn, J. van Santen, and S. Lafebre. "Penalized splines for smooth representation of high-dimensional Monte Carlo datasets". In: *Computer Physics Communications* 184.9 (2013), pp. 2214–2220. doi: <https://doi.org/10.1016/j.cpc.2013.04.008> (cited on page 5).
- [11] M. G. Aartsen et al. "Energy Reconstruction Methods in the IceCube Neutrino Telescope". In: *JINST* 9 (2014), P03009. doi: [10.1088/1748-0221/9/03/P03009](https://doi.org/10.1088/1748-0221/9/03/P03009) (cited on page 5).
- [12] R. Abbasi et al. "Low energy event reconstruction in IceCube DeepCore". In: *Eur. Phys. J. C* 82.9 (2022), p. 807. doi: [10.1140/epjc/s10052-022-10721-2](https://doi.org/10.1140/epjc/s10052-022-10721-2) (cited on pages 6, 17, 19).
- [13] F. James and M. Roos. "Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations". In: *Comput. Phys. Commun.* 10 (1975), pp. 343–367. doi: [10.1016/0010-4655\(75\)90039-9](https://doi.org/10.1016/0010-4655(75)90039-9) (cited on page 7).
- [14] H. Dembinski et al. *scikit-hep/iminuit: v2.17.0*. Version v2.17.0. Sept. 2022. doi: [10.5281/zenodo.7115916](https://doi.org/10.5281/zenodo.7115916) (cited on page 7).
- [15] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830 (cited on page 12).
- [16] S. Yu and J. Micallef. "Recent neutrino oscillation result with the IceCube experiment". In: *38th International Cosmic Ray Conference*. July 2023 (cited on page 17).
- [17] S. Yu and on behalf of the IceCube collaboration. "Direction reconstruction using a CNN for GeV-scale neutrinos in IceCube". In: *Journal of Instrumentation* 16.11 (Nov. 2021), p. C11001. doi: [10.1088/1748-0221/16/11/C11001](https://doi.org/10.1088/1748-0221/16/11/C11001) (cited on pages 17, 18).

- [18] J. Micallef. <https://github.com/jessimic/LowEnergyNeuralNetwork> (cited on page 17).
- [19] M. Huennefeld. "Deep Learning in Physics exemplified by the Reconstruction of Muon-Neutrino Events in IceCube". In: *PoS ICRC2017* (2017), p. 1057. doi: [10.22323/1.301.1057](https://doi.org/10.22323/1.301.1057) (cited on page 17).