

Search for Astrophysical Tau Neutrinos with the IceCube Waveforms

The IceCube Collaboration

http://icecube.wisc.edu/collaboration/authors/icrc17_icecube

E-mail: lwille@icecube.wisc.edu

The IceCube Neutrino Observatory has detected a diffuse astrophysical neutrino flux which is expected to have a close to equal ratio of neutrino flavors due to thorough mixing over astronomical baselines. However, IceCube has yet to detect astrophysical tau neutrinos. A tau neutrino undergoing charged current interaction will have two subsequent energy losses, the first from the neutrino interaction and the second from the decay of the secondary tau lepton. At PeV neutrino energies, IceCube can resolve these depositions as two separated cascades. At energies near hundreds of TeV the cascades are not well separated but can be observed as a double pulse waveform in individual IceCube sensors. Here we will present three techniques to improve tau double pulse waveform identification. One technique utilizes neighboring IceCube light sensors to observe a double pulse event, another incorporates machine learning algorithms to boost detection of double pulses, and the third explores the possibility of stacking waveforms to increase the double pulse signal. The first two approaches show a promising increase of signal rates by at least 50% while keeping similar or lower backgrounds at the double pulse waveform identification stage.

Corresponding authors: Sandro Kopper¹, Maximilian Meier², Logan Wille^{*3}, and Donglian Xu³

¹ *Department of Physics and Astronomy, University of Alabama, Tuscaloosa, AL 35487, USA*

² *Lehrstuhl für Experimentelle Physik V, Technische Universität Dortmund, 44227 Dortmund, Germany*

³ *Wisconsin IceCube Particle Astrophysics Center and Department of Physics, University of Wisconsin, Madison, WI 53703, USA*

*35th International Cosmic Ray Conference - ICRC217 -
10-20 July, 2017
Bexco, Busan, Korea*

*Speaker.

1. Introduction

IceCube has discovered a flux of high energy neutrinos which is currently consistent with an origin of unknown astrophysical sources [1]. These astrophysical neutrinos stream to IceCube from extragalactic and galactic sources, covering astronomical distances during which they oscillate flavors. Regardless of the flavor composition at production, the neutrino flux oscillates to a nearly equal 1:1:1 flavor composition [2] assuming standard neutrino oscillations. The direct observation of a ν_τ will be an additional confirmation of astrophysical origin of high energy neutrinos observed by IceCube due to a substantially smaller atmospheric background [3]. A recent global analysis from IceCube found the flavor ratio to be consistent with 1:1:1 but with large uncertainties [4, 5]. This is partly due to IceCube not yet identifying a ν_τ [6, 7], which is consistent with expectations. A precise flavor ratio measurement will help constrain new physics in previously unconstrained neutrino oscillation sectors [2, 8]. We will present three improved techniques to identify ν_τ interacting inside of IceCube.

IceCube is a cubic kilometer detector at the geographic South Pole designed to detect GeV-PeV neutrinos. The detector consists of 86 strings, 78 spaced at 125 m and 8 strings in the center with closer spacing, each holding 60 digital optical modules (DOMs). The 78 strings spaced at 125 m hold the DOMs equally spaced at depths between 1450 m to 2450 m below the surface and the 8 closer spaced strings hold 50 DOMs between 2100 m to 2450 m and 10 DOMs between 1900 m to 2000 m [9]. This array of DOMs encompasses 1 gigaton of ultra-transparent glacial ice, making the largest calorimetric detector in the world. Each DOM consists of a 10 inch PMT connected to digitizing electronics. The waveforms collected by the PMTs are digitized before being sent to the surface. There are two types of waveform digitization, the Analog Transient Waveform Digitizer (ATWD) and the Fast Analog to Digital Converter (FADC). Two of the techniques presented here use the ATWD digitization which provides the most detailed description of the photon arrival time available. The ATWD uses three gain factors, 16, 2, 0.25, over a nominal gain of 10^7 . During digitization the highest gain without saturation is used to capture the finest details of the waveform. The ATWD captures the waveform in 3.3 ns bins with 128 samples.

2. Tau Neutrinos in IceCube

IceCube observes the Cherenkov light produced by the secondary particles created in the interaction of the neutrino with the ice nuclei. A ν_τ undergoing a charged-current interaction can have a unique topology in IceCube. The interaction produces a τ lepton which travels a distance about 50 m per PeV of the τ lepton energy away from the interaction vertex and subsequently decays. If the τ decays to hadrons or an electron there will be exactly two high energy depositions following the neutrino interaction as shown in Fig. 1 (left). However, IceCube can only distinguish the two depositions if the τ has a sufficiently long life time, which corresponds roughly to a τ that has traveled 5 m and above. If the tau decay length is long enough (>50 m), the two depositions can be resolved and reconstructed separately. This topology is referred to as a double bang [6, 10]. At more intermediate lengths (5-50 m) the two losses cannot be well separated, but may create two pulses in an optical module waveform, referred to as a double pulse. A search for these double pulse events in 3 years of IceCube was conducted and found zero double pulse events which was

consistent with the expectation [7]. In these proceedings we present three techniques that improve the algorithm previously used to identify double pulses in waveforms and remove single cascade event backgrounds, Fig. 1 (right) shows exemplary signal and background waveforms. Not discussed in these proceedings is the necessary separation of track-like background events generated by ν_μ charged current interactions and atmospheric muons. These track-like events can create a double pulse signal due to subsequent stochastic losses near DOMs but are straight forward to remove due to their event shape.

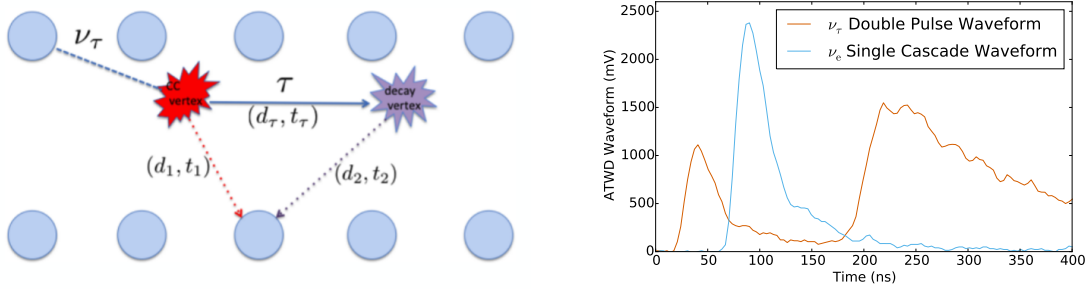


Figure 1: Left: Diagram (not to scale) showing a ν_τ undergoing a charged current interaction inside of IceCube. The blue spheres depict IceCube DOMs not drawn to scale. The two energy depositions of the charged current interaction and subsequent τ decay can be observed by a single DOM which creates a double pulse waveform. Right: Simulated ATWD waveforms recorded by individual DOMs. While a charged current ν_τ interaction creates a double pulse, a ν_e charged current interaction is observed as a single pulse. The goal of the double pulse identification is to separate these two types of waveforms.

3. Double Pulse Waveform Identification Techniques

3.1 Local Coincident Double Pulse Waveforms

Double pulse events are only observable in energetic, bright events, and in DOMs near the interaction vertex. Therefore, events with less than 2000 detected photoelectrons (PE) are rejected, and only the ATWD waveforms of individual DOMs which detected at least 432 PE are analyzed for a double pulse waveform. The original double pulse identification (DPI) was designed to find double pulse waveforms in individual DOMs while rejecting single pulse waveforms. The necessary features to identify a double pulse are the rising and falling edge of the first pulse and the rising edge of the second pulse. A second pulse falling edge is not necessary to search for as it is a guaranteed feature and offers no discriminating power. First a sliding time window is used to find the start of the waveform pulse. Then the rising and falling edges are found by calculating the time derivative of the waveform smoothed over 4 ATWD time bins, a 13.2 ns period. Once the necessary edges are found, their duration and steepness are determined and used to decide if the waveform is a double pulse. High thresholds for the rising and falling edges were necessary to reduce the single cascade background waveforms which can have additional structure due to scattering light and statistically varying light. These high thresholds also reduce the signal rate of ν_τ events.

Table 1: Comparison of the threshold values from the improved local coincidence method to the method used in the previous analysis. These are threshold values for declaring a waveform with two rising edges and a falling edge a double pulse.

	LC Method	PRD 93.022001
1st Rising Edge Steepness	>1 mV/ns	>10 mV/ns
1st Rising Edge Duration	>13.2 ns	>26.4 ns
Falling Edge Steepness	<-0.5 mV/ns	<-17 mV/ns
Falling Edge Duration	>26.4 ns	>26.4 ns
2nd Rising Edge Steepness	>12 mV/ns	>18 mV/ns
2nd Rising Edge Duration	>39.6 ns	>39.6 ns

The improved method, referred to as Local Coincidence (LC) method, adds a requirement for the double pulse waveform to be observed in pairs of nearest or next-to-nearest DOMs. A single cascade event tends to not create irregular waveforms in multiple DOMs while a double cascade tends to create double pulses in multiple DOMs. This allows a lower threshold widths and steepness of the rising and falling edges of the double pulse waveform without increasing the background rate. The threshold values of the original DPI and the LC modification are shown in Table 1. These thresholds were found by scanning over the six parameters and picking the set with the highest signal rate while keeping a 10:1 signal to single cascade background rate. One notable difference between the two thresholds is the significantly smaller rising edge threshold of the LC method. A first rising edge is a feature in both single cascade and double cascade events and so is not a good discriminator between background and signal. However, in the original method the first rising edge threshold was large to reduce prepulsing single cascade events. In the improved method this background is not a concern because prepulsing is a random process and rarely occurring in multiple DOMs in an event.

ν_τ interactions tend to have low inelasticity because of mass threshold effects of τ production, which manifests in IceCube as a small first energy deposition and more energetic second deposition. Decreasing the first pulse threshold makes this low inelastic phase space observable. Additionally the overall decrease in the double pulse thresholds makes lower energy ν_τ events observable. Both of these effects increase the effective area, as shown in Fig. 2. Overall the LC technique has a 50% increase in signal rate over the previous technique while maintaining a similar signal purity. A Monte Carlo event that passes the improved selection that would have been previously rejected is shown in Fig. 2.

3.2 Machine Learning Based Double Pulse Selection

A further next improvement under consideration is to employ multivariate machine learning algorithms. The event selection used in this approach consists of two major steps. First, quality cuts identical to those discussed in section 3.1 are applied. Following the quality cuts, samples of signal waveforms and background waveforms are selected which are used to train a Random Forest [11]. As in the LC method, the goal is to reject waveforms from events with single cascade topologies while retaining as many double pulse waveforms as possible.

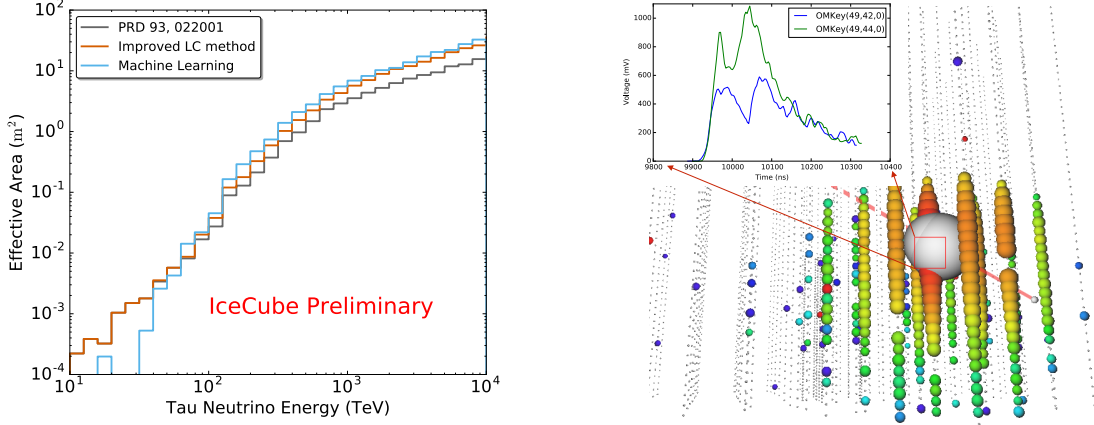


Figure 2: Left: Effective areas for the previous 3 year analysis, denoted by PRD 93.022001, the improved local coincident method discussed in section 3.1, and the machine learning technique discussed in section 3.2. Right: Monte Carlo event that passes the re-optimized LC double pulse algorithm that would have been rejected by the previous analysis. This is a simulated down-going ν_τ event with $E_\nu = 385$ TeV undergoing CC interaction in the detector. The two overlapping spheres in grey indicate the interaction and decay vertices with deposited energies of 111 TeV and 273 TeV respectively. The tau decay length is 12 m.

All waveforms that pass the quality cuts, are characterized by the same features calculated on their derivative as explained in section 3.1. Here, feature denotes an observable quantity of a particular waveform. Signal waveforms for training the Random Forest are selected by a set of cuts on these features which is described in [12]. These cuts were optimized to retain as many double pulse waveforms from a hand-selected sample as possible by applying rather loose cuts to the steepness of the first and second rising edge with no restrictions on the falling edge steepness. They were later re-optimized to reject single cascade waveforms more efficiently.

Background waveforms for training are sampled uniformly at random from all single cascade background components (all flavor neutral current neutrino events and charged current electron neutrino events) according to their relative occurrence of expected waveforms from an assumed $E^{-2.13}$ astrophysical flux [13]. In addition to the already described derivative features, eight new features are added. The new features are primarily focused on characterizing the statistical properties of the waveforms, their smoothness and their consistency with a single cascade waveform. The waveform mean, an energy proxy, allows the model to easily vary cuts depending on the energy deposited in the DOM. Another important feature describes the agreement of the waveform with a fit to an exponential function starting at the maximum of the first pulse. A good agreement is consistent with single cascade waveforms.

These features are used in a Random Forest [11] to distinguish double pulse waveforms from single cascade waveforms. The forest is trained and tested in a 10-fold cross validation with 200 trees, using ~ 6500 signal and $1.8 \cdot 10^6$ background waveforms. The resulting Random Forest score distributions for signal and background waveforms are shown on the left side of Fig. 3. The output of the Random Forest is a score from 0 to 1, spanning waveforms that are more background-like to

POS (ICRC2017) 1009

those that are more double pulse-like.

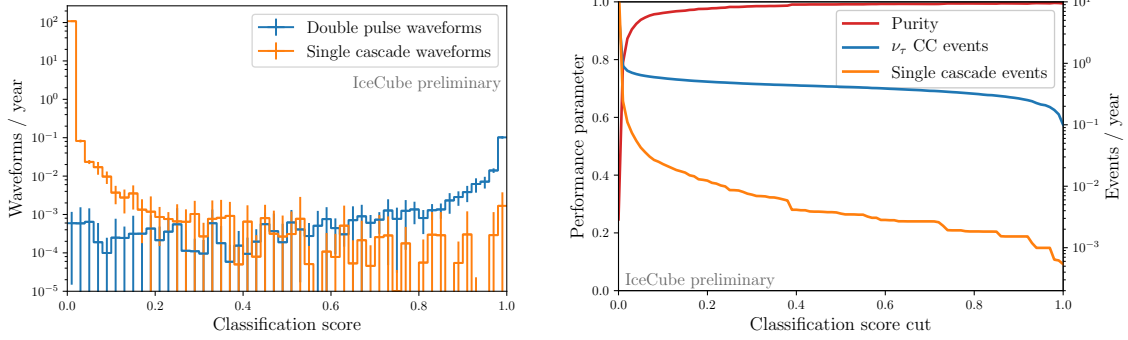


Figure 3: Results of the double pulse waveform selection. Left: Distribution of the classification score of the waveforms. The standard deviation of each bin is determined within the cross validation. Right: Purity as well as the expected remaining event rates per year as a function of the classification score cut.

This selection requires at least one waveform per event passing the Random Forest score cut. The right side of Fig. 3 shows the resulting purity as well as expected event rates from ν_τ CC interactions and all single cascade background components. The score cut was chosen at 0.2 to retain as much signal as possible, while still reaching a purity of $\sim 97\%$. The assumed astrophysical flux is an unbroken power law with a per flavor normalization of $0.9 \cdot 10^{-18} \text{ GeV}^{-1} \text{ cm}^{-2} \text{ s}^{-1} \text{ sr}^{-1}$ at 100 TeV and a spectral index of 2.13 [13]. Expected remaining event rates for signal and single cascade background components can be found in Tab. 2. The resulting effective area of this selection in comparison to [7] at the same analysis stage is depicted on the left in Fig. 2. It shows an improvement of the effective area with respect to the original DPI at energies above $\sim 70 \text{ TeV}$ which leads an increase in expected signal rate of around 50 %.

3.3 Charge Stacking Technique

Another way to extend the double pulse analysis currently under investigation is to stack the information gained from multiple DOMs. Instead of using the raw waveforms, a common frame of time and charge amplitude across all DOMs based on unfolded charge-timing information is produced and analyzed [14].

The stacked charge information is ideally the recovered arrival time of all the photoelectrons produced within the DOMs and so not a continuous function. It is much coarser than the original waveforms. Thus, the double pulse identification algorithm had to be adjusted. The extracted charge of each DOM being stacked is combined in a single charge-time distribution with 12 ns time bins starting at the time that 1% of the total charge has been deposited within the whole

Table 2: Preliminary event rates for signal and single cascade background components of the machine learning based double pulse selection. The assumed astrophysical flux can be found in [13]. Uncertainties are statistical only.

Event Type	Events per year
Astrophysical ν_τ CC	0.486 ± 0.005
Astrophysical ν_e CC	0.010 ± 0.001
Astrophysical ν NC	0.004 ± 0.001

detector. The distribution is further smoothed using a three bin moving average ("box smoothing") algorithm. For an example of such a histogram compare Fig. 4 (left). An upper time bound of the distribution of 500 ns is chosen to avoid issues with the known "afterpulsing" behavior of DOMs [9] that might mimic a second pulse. To ensure a proper reconstruction quality and to meet minimum energy requirements a total charge of 1000 PE within the whole detector and a charge of 100 PE deposited in the time bin with the highest charge is demanded to pass the selection. The requirements for the pulse identification algorithm differ for the first and second pulses. Since the beginning of the first pulse often falls outside the previously defined start time, the first pulse is defined as either a region of positive "slope" followed by a region of at least three bins of negative "slope" or simply a region of at least three bins of negative "slope" starting from the beginning of the distribution. The "slope" is being defined by the use of derivative proxies D_i at bin n . The three proxies D_1, D_2, D_3 used are based on the charge Q_n of neighboring bins: $D_i \propto (Q_{n+i} - Q_{n-1})$.

If any of the proxies has the same sign as the "slope" in the previous bin, it is considered unchanged. The additional requirement for the second pulse is an amplitude of at least half that of the highest charge bin in the distribution and a minimum start time of 100 ns. Unlike for the first pulses, both a rising and falling edge are now required. As before the falling (rising) edge is defined by at least three bins of negative (positive) derivative D using any of the three proxies.

Depending on the exact event geometry stacking the charge of multiple neighboring DOMs can lead to very sharp, short time frame charge pulses. This is caused by arrival time difference of light corresponding to the distance of the DOMs instead of the geometric difference between the ν_τ interaction and τ decay. The algorithm removes these "false" double pulses by requiring that none of the bins defining the start of the falling edge have a charge smaller than half the amplitude of the pulse. For an example of this charge structure compare Figure 4 (right).

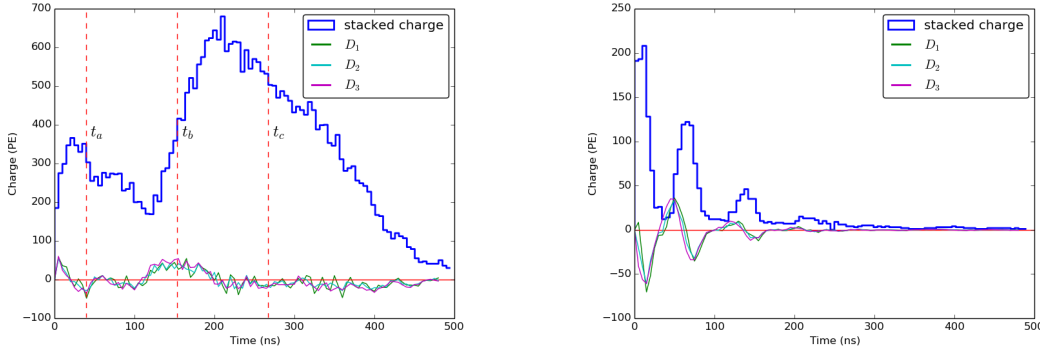


Figure 4: Left: Example of a stacked charge distribution of a ν_τ event for the brightest string. At times t_a , t_b and t_c the slope shows a falling, a rising and another falling behavior respectively. Right: Stacking the charge information can lead to multiple pulses for single cascade events. The time structure corresponds to the DOM separation. Lower curves: Derivative proxy D_i behavior.

The final open issue remaining for this technique is how to choose which DOMs to stack. The first naïve choice is to use them all. This, however, introduces a similar false positive problem as mentioned before due to the large separation between the IceCube strings. This time though, the distance is of the same order of magnitude as the actual separation of the interaction and decay

vertices of the ν_τ signal and so cannot easily be removed. A way around this is to only consider DOMs on the string with the highest total charge. Still, charge information of too many DOMs far from the event vertex "wash out" the double pulse signal and lead to a rather low passing rate of signal events. A promising approach seems to be to start with a selection of a combination of any two adjacent DOMs on the string and see if the event passes the selection. If not, we take three, four, etc. adjacent DOMs and repeat the algorithm until either the algorithm finds a double pulse or all combinations have been checked. If no double pulses are found the event is considered a background event. For a benchmark E^{-2} spectrum this leads to a signal passing rate of ≈ 0.24 events per year with a ν_e event background contribution that is $< 5\%$. The overlap fraction with the original double pulse method [7] is smaller than $1/3$ showing that this method adds additional information.

In addition to members of the IceCube collaboration we'd like to acknowledge the work of James Parkes who was directly involved in developing this technique.

4. Conclusion and Outlook

The direct observation of a ν_τ interaction in IceCube will be another confirmation of astrophysical neutrinos and can shed light on potential new physics. IceCube has yet to identify a ν_τ interaction but becomes more likely with increased observation time and improved methods as presented here. The improved double pulse identification methods presented here use more sophisticated techniques and additional information than the previous method that allows IceCube to access double cascade ν_τ events at lower energies and previously unexplored phase spaces.

References

- [1] **IceCube** Collaboration, [PoS \(ICRC2015\) 1081](#) (2016).
- [2] C. A. Argüelles, T. Katori, and J. Salvado, *Phys. Rev. Lett.* **115** (2015) 161303.
- [3] A. Fedynitch, R. Engel, T. K. Gaisser, F. Riehn, and T. Stanev, *EPJ Web Conf.* **99** (2015) 08001.
- [4] **IceCube** Collaboration, M. G. Aartsen et al., *Astrophys. J.* **809** (2015) 98.
- [5] **IceCube** Collaboration, M. G. Aartsen et al., *Phys. Rev. Lett.* **114** (2015) 171102.
- [6] **IceCube** Collaboration, [PoS \(ICRC2017\) 974](#) (these proceedings).
- [7] **IceCube** Collaboration, M. G. Aartsen et al., *Phys. Rev. D* **93** (Jan, 2016) 022001.
- [8] M. Bustamante, J. F. Beacom, and W. Winter, *Phys. Rev. Lett.* **115** (2015) 161302.
- [9] **IceCube** Collaboration, M. G. Aartsen et al., *JINST* **12** (2017) P03012.
- [10] **IceCube** Collaboration, [PoS \(ICRC2017\) 973](#) (these proceedings).
- [11] L. Breiman, *Machine Learning* **45** (2001) 5–32.
- [12] **IceCube** Collaboration in *Proceedings, 33rd International Cosmic Ray Conference (ICRC2013): Rio de Janeiro, Brazil, July 2-9, 2013*, p. 0643.
- [13] **IceCube** Collaboration, M. G. Aartsen et al., *Astrophys. J.* **833** (2016) 3.
- [14] **IceCube** Collaboration, M. G. Aartsen et al., *Journal of Instrumentation* **9** (2014) P03009.