Development of Machine Learning Based Triggering for Liquid Argon Time Projection Chambers
REU Program at Columbia University - Nevis Labs

Josiah Dubovi[1]

[1]Pennsylvania Western University

August 21, 2023

Contents

1 Background

1.1 Neutrinos and Physics Beyond the Standard Model

The Standard Model of particle physics is a highly effective theory for predicting the behavior and interactions of the universe's most fundamental components. According to the standard model, neutrinos are massless, uncharged leptons that only interact via the weak nuclear force. They come in three flavors: electron neutrino (ν_e), muon neutrino (ν_μ), and tau neutrino(ν_τ). These particles exhibit unique behaviors that make them of particular interest to the future of particle physics. However, because they rarely interact with matter, they are very difficult to detect. Modern neutrino experiments must employ indirect methods of detection that take advantage of the presence of charged particles, which are comparably much easier to detect, in neutrino interactions.

Experimental discoveries over the last several decades have challenged the Standard Model. It is well known that, despite its high accuracy in predicting many particle behaviors, the Standard Model is incomplete. Many theoretical explanations for physics beyond the Standard Model have been suggested, and modern particle physics experiments aim to test these theories.

One significant piece of evidence suggesting the Standard Model's incompleteness is neutrino oscillation. As neutrinos travel, they have been observed to spontaneously switch flavors. This implies that neutrinos must have nonzero mass. This directly contradicts the predictions of the Standard Model, and developing a better understanding of neutrinos can help in forming a more complete and accurate model of the universe.

1.2 DUNE and ICEBERG

The Deep Underground Neutrino Experiment (DUNE) is an upcoming international project hosted at Fermi National Accelerator Laboratory (Fermilab) that will explore many new areas for experimental particle physics. Among the many objectives of DUNE, some of the most significant include an explanation for the imbalance between matter and antimatter in the universe, the observation of a proton decay, and the ability to detect supernovae before their light reaches Earth. The achievement of any one of these goals would be a major step forward in our understanding of the universe. As the name implies, DUNE will pursue these goals through the investigation of neutrinos and their interactions. Like several other recent cutting-edge neutrino experiments, it will make use of Liquid Argon Time Projection Chambers (LArTPCs) for the detection of charged particles. A beam of neutrinos will be produced at Fermilab's accelerators. This beam will then travel underground over 1300 km across the United States. Along the beam's path, there will be two sets of detectors: the near detector, located at Fermilab near the origin of the neutrino beam, and the far detector, located 1.5 km underground at Sanford Underground Research Facility in South Dakota. The far detector will consist of four of the largest LArTPCs ever constructed, with a total mass of 70 kilotons of liquid argon. At such a great depth, background noise, which is primarily due to cosmic rays, will be reduced significantly in this detector.

As a very costly and ambitious experiment, DUNE requires prototype detectors to be constructed and tested to ensure that the hardware is functional. Among these prototypes is the Integrated Cryostat and Electronics Built for Experimental Research Goals (ICEBERG) experiment. ICEBERG is a small-scale LArTPC designed to test the technology that will eventually be used in DUNE.

ICEBERG began operation in 2019. As a small prototype detector only containing approximately 4.5 tons of liquid argon, ICEBERG's detection capabilities are extremely limited compared to DUNE. Furthermore, while the DUNE detectors will be located over a kilometer underground, ICEBERG is on the surface; it is therefore less isolated from signals resulting from cosmic rays and other particle sources that DUNE is shielded from due to its underground location. This, however, is not a problem for a project such as ICEBERG because it is not attempting to discover new physics. In fact, cosmic ray signals are particularly beneficial to ICEBERG as they produce a well-known interaction that can be used to test and calibrate the detector's hardware and software.

1.3 Liquid Argon Time Projection Chamber (LArTPC)

ICEBERG and the DUNE far detectors are LArTPCs, shown in Figure 1, which are particle detectors consisting of a large volume of liquid argon within a strong, uniform electric field. As charged particles travel through the liquid argon medium, they ionize argon atoms, freeing electrons from their nuclei. Because of the electric field, these ionization electrons then drift across the chamber until being collected by a plane of parallel wires at one side. The current on the wires resulting from these electrons is read as an ADC waveform. This method of detection preserves the shape of the charged particles' paths, which can be reconstructed using the ADC waveforms collected by each channel (wire) at a particular time. The reconstructed event display (Figure 2) is therefore a plot of location (channel number) vs time vs ADC count on the x, y, and z axes respectively.

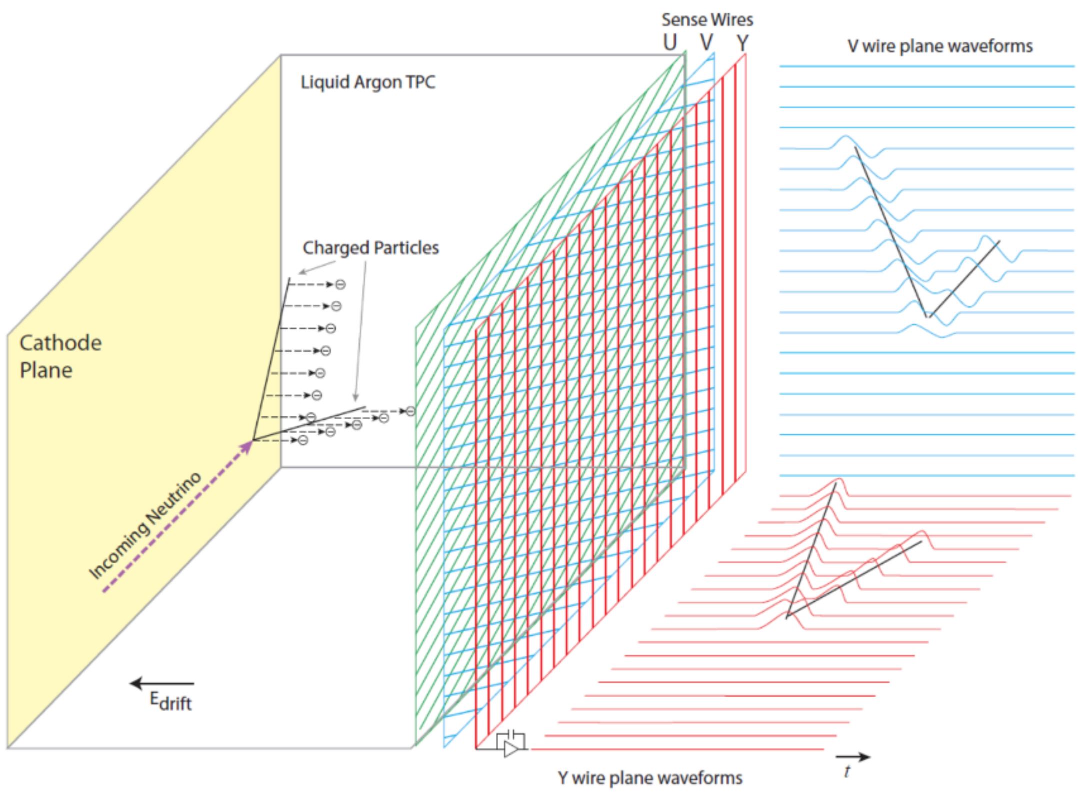


Figure 1: *Diagram of LArTPC.*

A typical LArTPC consists of three planes of parallel wires: two induction planes (named "U" and "V") and one collection plane (named "Z"). The induction planes have a relatively low potential, allowing the negatively charged ionization electrons to drift past. In doing so, the electrons induce a current within the wires, which can then be converted to a digital signal. The electrons then reach the collection plane, which has a relatively high potential, causing them to gather on the nearest wire within the plane. To allow the maximum spatial information to be gathered from the wire planes, the two induction planes are each oriented 60° from the vertical collection plane.

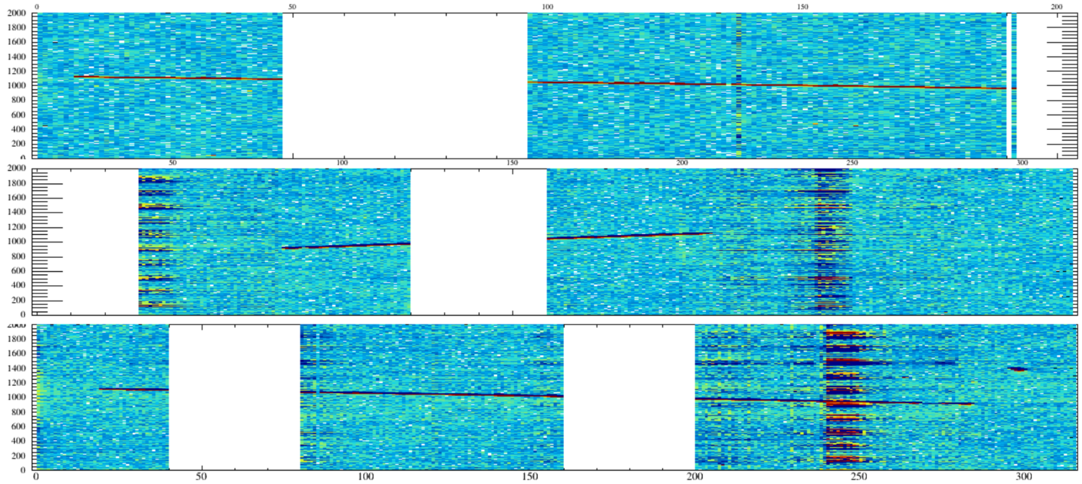


Figure 2: *ICEBERG event display.* ***GET BETTER IMAGE***

In addition to ionization electrons, charged particles produce scintillation light as they traverse the liquid argon medium. This light is collected by photomultiplier tubes (PMTs) at the sides of the chamber. Although the data gathered by the PMTs does not provide as much spatial information as the ionization electrons, it does give precise timing information.

It is necessary to note that a LArTPC does not give any information on the charge of the particles that it detects. Both positive and negative particles behave the same within a LArTPC. Many particle detectors use magnetic fields to curve the path of particles, and then use the direction of this curvature to determine the particles' charge. However, this cannot be done in a LArTPC because the ionization electrons that are used to reconstruct the particle's path would also be curved by the magnetic field. LArTPCs prioritize the preservation of the particle's path rather than its electric charge. However, because the properties of most elementary particles are relatively well understood, the particle's path is often enough to classify it.

Neutrinos are neutral particles which therefore cannot be detected by a LArTPC. However, many interactions involving neutrinos and charged particles are well understood; the laws of conservation of energy, conservation of momentum, and conservation of lepton number often indicate that a neutrino is present even though it is not directly observed.

High energy photons, electrons, and positrons create a unique path within a LArTPC which can be very useful for characterizing neutrino interactions. When a high energy electron is decelerated in the detector, it emits a photon as bremsstrahlung (braking radiation) and is deflected as momentum is conserved. If this photon has high enough energy, it will pair-produce an electron and positron. Assuming this electron has enough energy, it too will produce bremsstrahlung. The result is a cascade of electrons, positrons, and photons. However, as a neutral particle, the photons can not be detected, so the event display will only depict the electron and positron; the photon will appear as empty space separating the path of one charged particle and the next. An electromagnetic shower can be seen in Figure 3. These particle cascades are known as electromagnetic showers, and can originate with either an electron or a photon (the number of detectable particles at its origin can be used to determined which it is).

1.4 Stopping Muon Decay

In the Standard Model, muons are unstable, negatively charged leptons with a mass approximately two-hundred times the mass of an electron. They decay via the weak force, and their most common

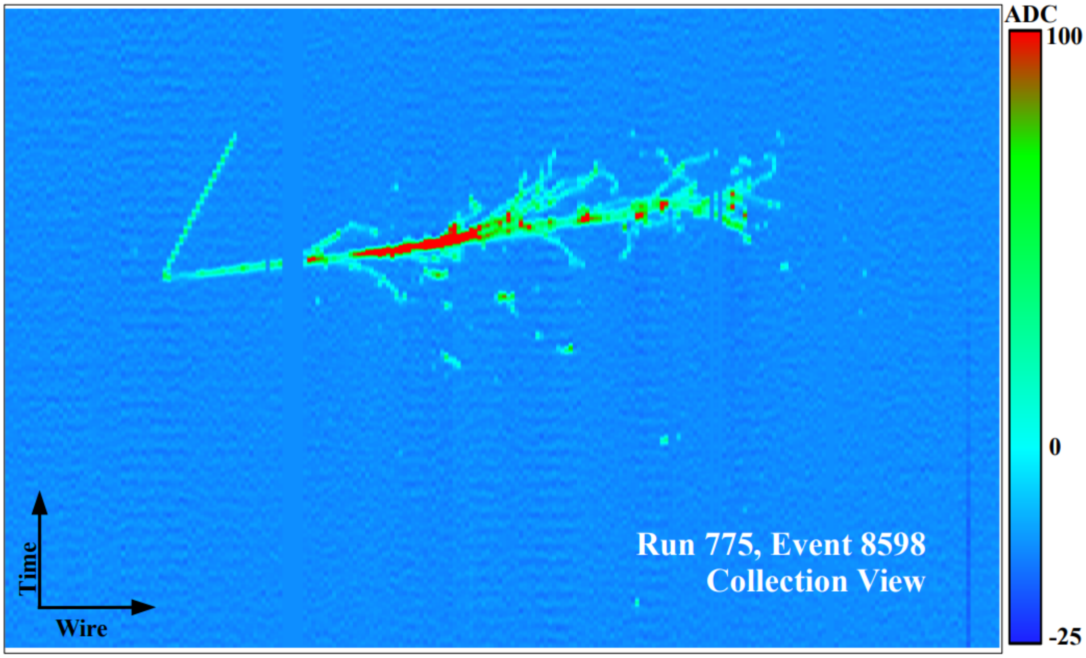


Figure 3: *Electromagnetic shower in ArgoNeuT LArTPC.*

mode of decay is called Michel Decay, which produces a muon neutrino, electron antineutrino, and an electron (called a Michel Electron):

$$\mu \rightarrow \nu_\mu + \overline{\nu_e} + e$$

Muons are produced when cosmic rays (atomic nuclei traversing the universe at relativistic speeds) collide with Earth's atmospheric particles. Because they are moving at relativistic speeds, these muons do not immediately decay from the reference frame of an Earth observer. If one of these particles collides with an atom in the LArTPC and comes to rest, it will decay within microseconds, producing an electromagnetic shower. Since muons and electrons are both charged particles, they can be detected by the LArTPC. The general shapes and sizes of their paths are well known.

1.5 Artificial Intelligence in Event Triggering

DUNE will collect several terabytes of data every second, but the majority of this data cannot be stored. Many events of interest, such as a proton decay and supernova neutrinos, are extremely rare. So an algorithm must be developed to reject a vast majority of background noise while also detecting low-energy rare events with very high accuracy. In the past, over-the-threshold methods have been used to detect signals of interest by only saving events with ADC counts above a set value [1]. However, Artificial Intelligence (AI) algorithms excellently meet these criteria and have been shown to be more effective than classical algorithms [2]. AI is therefore very promising in the future of event triggering. Because DUNE collects data at a rate of thousands of time-ticks per second and processes this data in real time, a triggering neural network must be able to select signals quickly while also fitting in the detector's limited hardware space. Field programmable gate arrays (FPGAs) have been demonstrated to be more efficient than CPUs or GPUs to this end [3]. The primary goal of the present research is to develop a neural network that can differentiate between background noise and stopping muon decays for ICEBERG data. This project acts as a proof-of-concept upon which future models can be built, so the strict hardware limits are not taken into account. However, steps have been taken to reduce the computational complexity of the neural network when possible.

1.6 Convolutional Neural Network

Because the LArTPC event display creates a two-dimensional image, a two-dimensional convolutional neural network (2D CNN) can effectively be applied in this context. For several years, 2D CNNs have been among the most effective image recognition algorithms available[4]. The architecture of different 2D CNNs can vary significantly depending on their intended applications and computational complexity limitations. The general structure of a 2D CNN is as follows:

The input is a tensor (array) of dimensions $L \times W \times C$, where L and W are the image's length and width respectively in pixels, and C is the number of color channels. In a LArTPC event display, the length and width dimensions translate to channel and time respectively. The event display only has one color channel, which corresponds to ADC count. The value of each entry in the tensor is determined by that pixel's color.

Padding layers may be used to change the dimensions of input tensor by adding entries (pixels) of value zero. This may be necessary depending on the dimensions of the convolution matrix and pooling size in the following layers. Since the additional entries are zero in value, they have minimal impact of the computational load.

Pooling layers are used to compress the image to reduce the computational complexity of later layers. Pixels are grouped together with nearby pixels and a mathematical operation is used to determine a numerical value to represent them all. In max-pooling, one of the most commonly used pooling layers, the entry with the value in each channel is used to represent every pixel in the group. Figure 4 shows an example of max-pooling on a single-channel tensor.

Convolutional layers are then used to process the image. A convolution matrix, which has dimensions less than that of the layer's input tensor but equal channel count, is used to extract features from the image. The exact dimensions of the convolution matrix are predetermined, and its entries are set during training. Mathematical convolutions, an operation used to show how one function is modified by another, are calculated between the input tensor and the convolution matrix. The process of padding, pooling, and convolving may be repeated several times depending on the network's architecture. Additional layers may be included to perform different types of matrix operations on the input tensor, but these are less common and less closely associated with a CNN.

The network typically ends with a fully-connected-layer. Here, each entry of the layer's input tensor is connected to each entry of the layer's output tensor. A weights matrix, whose entries are determined from training and whose dimensions are determined from the dimensions of the input tensor and the number of desired output classes, is used to perform a preset activation function on the input tensor. The tensor is then passed on to the output layer, which has dimensions determined by the number of possible classes that can be assigned to the input. It is also common for a network to record its confidence level in its output.

The CNN must be trained on a set of images that already have their outputs labeled. Typically, a portion of this dataset is designated as the training set, and the remaining portion is assigned as the validation set. The network is trained of the training set, and then it is tested on the validation set to ensure that it has not been overfitted to the training set. Because all of these images are pre-labeled, the network's accuracy on this dataset is known. Ideally, the network's accuracy on the training set and validation set will be almost equal (and close to 100%), indicating that the network has not been overfitted to the training set. In cases with a limited training set, additional measures can be taken to avoid overfitting. These include dropout layers, in which some of the training images are randomly removed from training in each epoch (cycle in which the network is trained on all of the data), and data augmentation layers, in which the training images are all randomly rotated, flipped, and/or zoomed to teach the network to recognize patterns and features rather than specific images in the training set.

2 ICEBERG Noise Simulation

In order to train the neural network to distinguish between stopping muon decays and background noise, large quantities of labeled data are necessary. A dataset of this size must be simulated. This additionally allows the events to be automatically labeled as either background noise or muon decay.

While muon decays have previously been simulated in ICEBERG, a data-driven noise model needs to be created to accurately simulate the background. This is done by deconstructing the ADC waveforms on each channel of a real ICEBERG event using a Fast Fourier Transform (FFT). The waveforms for noise can then be simulated using the frequencies and RMS on each channel.

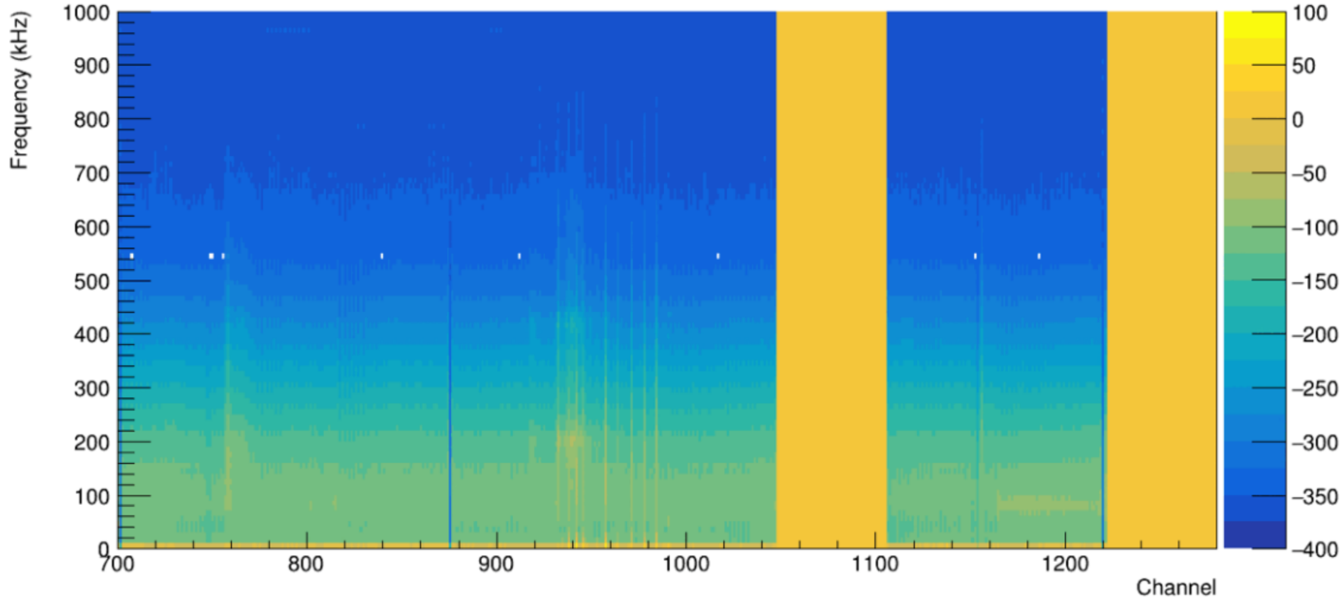
Each wire in the LArTPC may be subject to different levels of electronics noise due to their differing sizes and positions within the detector. Therefore, it is necessary to analyze the noise of each channel individually. Additionally, due to other external factors, ICEBERG's noise has changed over time. It is therefore necessary to consider data taken of the course of many days to avoid overfitting for one particular time. To this end, data was taken from ICEBERG Run 5, spanning every day from 3/13/21 to 4/4/21. The number of data files collected varies with each day, and the number of events varies with each file. To ensure consistency, at least 80 events were analyzed each day, with the exception of 3/26/21, which had no data collected. The FFT for each wire was calculated for every event, then averaged for each day. Figure 4 shows a plot of the average FFT for 3/13/21 (the first day of data that was used) compared with 4/4/21 (the last day of data that was used). Channel number, frequency, and amplitude are on the x, y, and z axes respectively.

The mean FFT of every channel on the Z plane for every event within the time-frame was taken (Figure 5). Each FFT was then divided into three equal frequency regions: Low (1 - 333 kHz), Medium (334-666 kHz), and High (667 - 999 kHz). The simulated waveform is a sum of three frequencies, one from each region. To determine which frequencies are selected, a continuous function was fitted to each region via polynomial interpolation (Figure 6). These three polynomials represent the frequency probability density function for their respective regions. Three frequency values, one from each region, are selected randomly based on these probability density functions. Their amplitudes are then weighted according to the original FFT spectrum. Figure 7 shows a simulated wave. This process is repeated for every wire. The waveforms for each channel are then plotted together in a simulated two-dimensional event display (Figure 8).

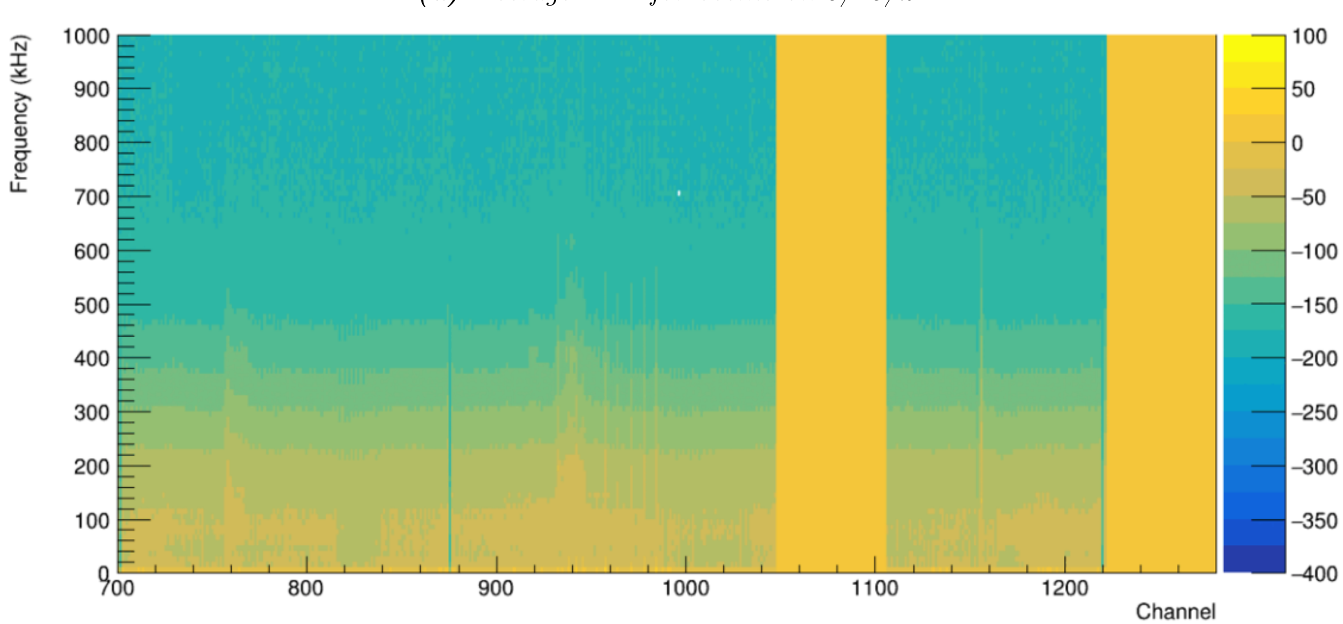
3 Convolutional Neural Network

The 2D CNN used in this project is adapted from a neural network developed for the detection of low energy events in DUNE. It is a variant of the popular image classification algorithm, "You Only Look Once" (YOLO). This architecture's original purposes effectively translate to the needs of the present research; the original YOLO architecture used several hundreds of thousands of parameters. This is both unnecessary and computationally excessive for a network that is intended to be implemented on an FPGA within a LArTPC. The architecture in [5] addressed this issue by reducing the number of convolutional layers, thus decreasing the parameter count significantly. This provided an effective foundation upon which to construct a neural network for the specific purpose of detecting Michel decay events. Because computational simplicity is not of primary concern in the present paper, further changes have been made to improve the networks performance at the cost of adding parameters and therefore increasing the computational load. Figure 9 shows the architecture of the network used in the present research.

Simulated Michel Decay events have not yet been created in a format readable to this neural network. It therefore cannot be trained and optimized at this time.



(a) *Average FFT for events on 3/13/21*



(b) *Average FFT for events on 4/4/21*

Figure 4: *Average FFTs for every wire. Color bar shows amplitude*

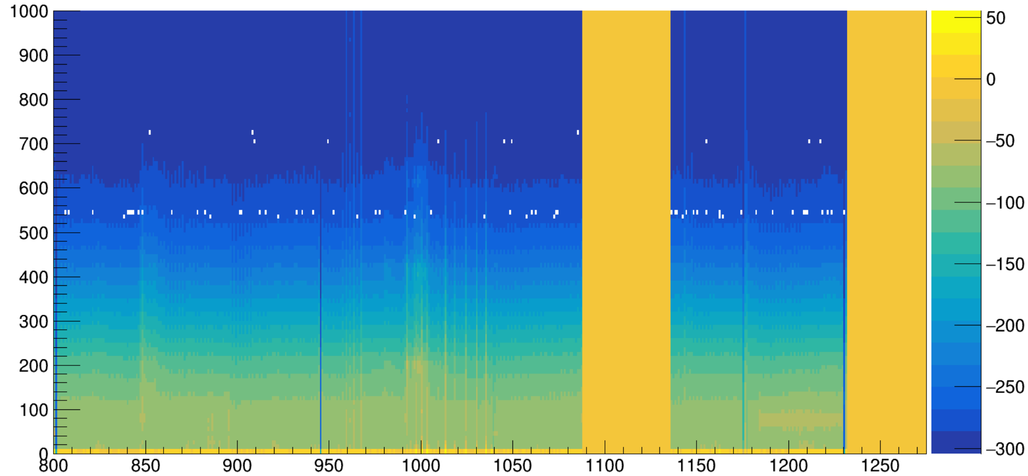


Figure 5: *Mean FFT spectra over all days for each wire*

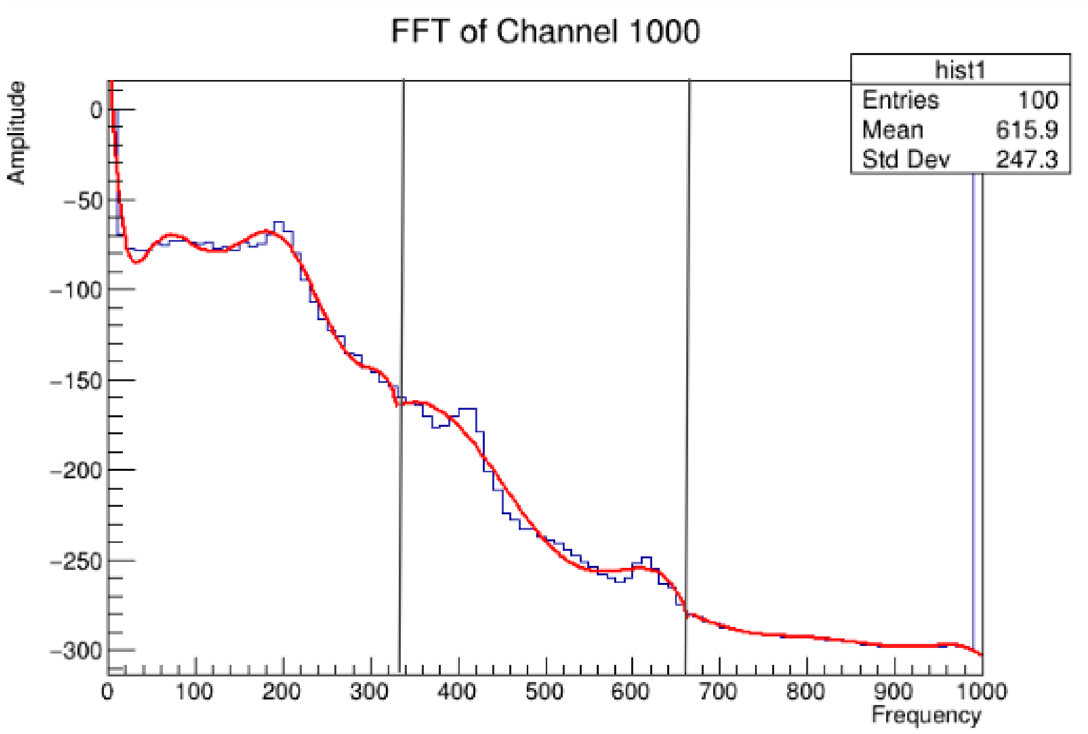


Figure 6: *Channel 1000 FFT with interpolation polynomial fit to each region*

4 Summary and Conclusions

A data-driven noise model has successfully been developed for ICEBERG,

A 2D CNN is a promising model for Michel Decay triggering in ICEBERG. At the present, the algorithm's

It has been shown that the noise present in ICEBERG is variable with time, which presents additional complications when trying to precisely develop a data-driven noise model. Only a small portion of the data collected by ICEBERG was used for this project; future work will have higher statistics, enabling a more accurate model. It is anticipated that significant improvements to this

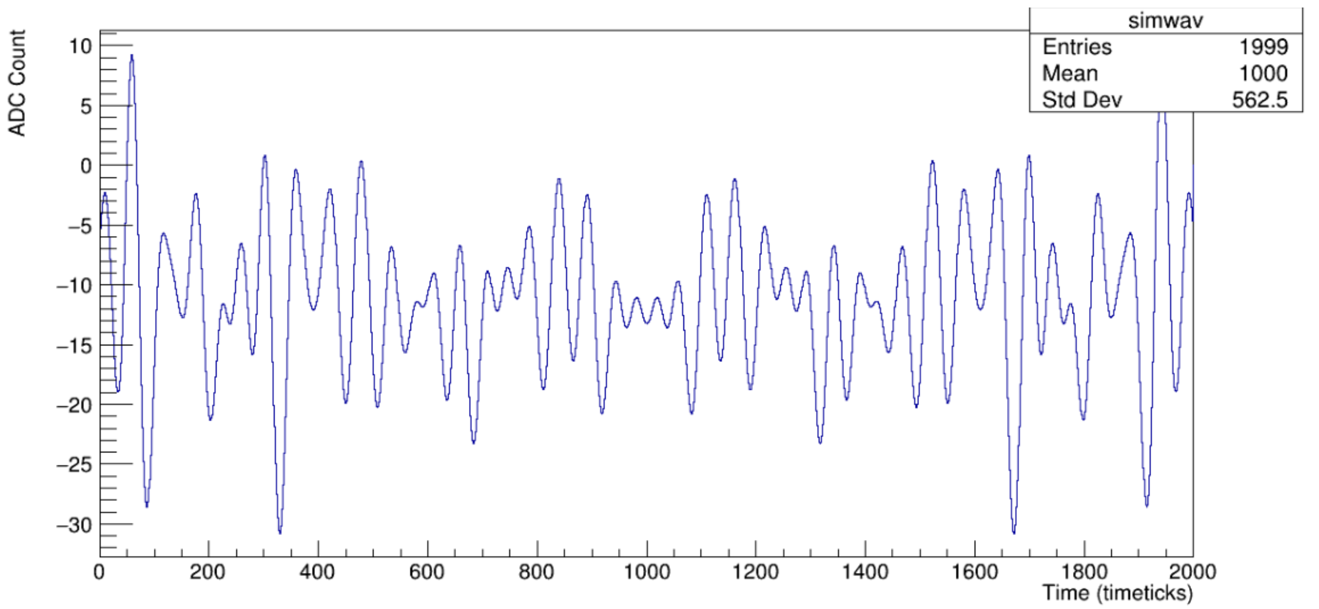


Figure 7: *Channel 1000 simulated wave.*
Frequencies: 70 kHz, 430 kHz, 700 kHz

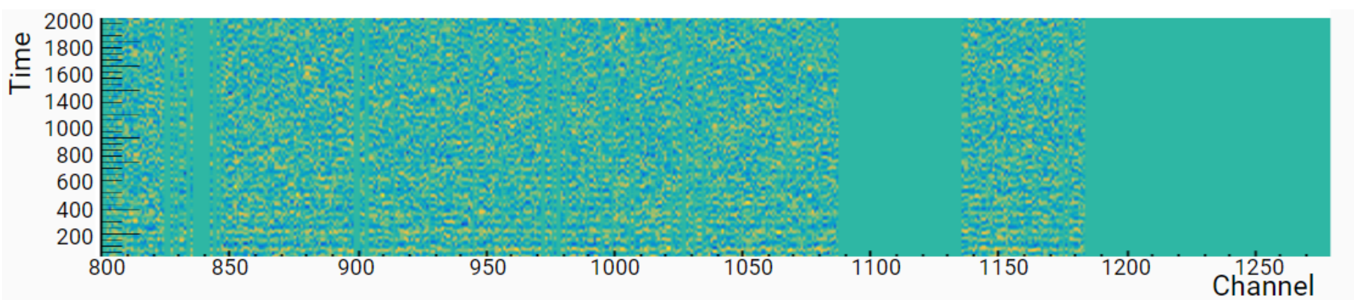


Figure 8: *Simulated Event Display (Collection Plane)*

noise model are possible, and it is likely that more precise noise simulation will subsequently result in a better dataset on which to train a neural network.

In addition to improving upon the present work, future projects will explore the development of triggering neural networks for other signals of interest in ICEBERG. While experimentally significant observations, such as a proton decay or supernova neutrino event, are beyond the scope of ICEBERG, there are other interactions observable in ICEBERG that may necessitate the application of a triggering neural network. Similar research is already being done to develop a similar algorithm for the selection of the decay of Ar-39, a radioactive isotope found within LArTPCs. While events such as Michel decay and Ar-39 decay are not of particular interest to DUNE, the ability to isolate and classify these signals will be important in avoiding the storing of unnecessary data.

5 Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. PHY/1950431.

```
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 resizing_28 (Resizing)      (None, 64, 480, 3)        0

 zero_padding2d_73 (ZeroPadd  (None, 64, 512, 3)       0
 ing2D)

 max_pooling2d_92 (MaxPoolin  (None, 32, 64, 3)        0
 g2D)

 zero_padding2d_74 (ZeroPadd  (None, 34, 66, 3)        0
 ing2D)

 conv2d_70 (Conv2D)          (None, 32, 64, 2)         56

 re_lu_60 (ReLU)             (None, 32, 64, 2)         0

 max_pooling2d_93 (MaxPoolin  (None, 16, 16, 2)        0
 g2D)

 zero_padding2d_75 (ZeroPadd  (None, 18, 18, 2)        0
 ing2D)

 conv2d_71 (Conv2D)          (None, 16, 16, 2)         38

 re_lu_61 (ReLU)             (None, 16, 16, 2)         0

 max_pooling2d_94 (MaxPoolin  (None, 4, 4, 2)          0
 g2D)

 reshape_13 (Reshape)        (None, 1, 1, 32)          0

 dense_46 (Dense)            (None, 1, 1, 12)          396

 re_lu_62 (ReLU)             (None, 1, 1, 12)          0

 dense_47 (Dense)            (None, 1, 1, 3)           39

=================================================================
Total params: 529
Trainable params: 529
Non-trainable params: 0
_________________________________________________________________
```

Figure 9: *Architecture of 2D CNN.*

References

[1] R. Acciarri, et. al. (2022) "A deep-learning based raw waveform region-of-interest" finder for the liquid argon time projection chamber. https://iopscience.iop.org/article/10.1088/1748-

0221/17/01/P01018/pdf

[2] J. Shen, et. al. (2018) "Towards a Uniform Template-based Architecture for Accelerating 2D and 3D CNNs on FPGA" $https://dl.acm.org/doi/abs/10.1145/3174243.3174257?casa_token = 0kFKOYEZuDYAAAAA : gb9c8FIkr3WaBXrc3ZeQybiAPsR_ODW6yn5iaFPk82Mp1uF_Ju_orHiwIQnMcntdxFj8lYWJdUnA$

[3] N. Sharma, et. al. (2018) "An Analysis Of Convolutional Neural Networks For Image Classification" https://www.sciencedirect.com/science/article/pii/S1877050918309335

[4] J. Clair, "Real Time Detection of Low-Energy Events for the DUNE Data Selection System", DUNE internal note, available upon request.