Alma Mater Studiorum Bologna

Fermilab Pisa Summerschool 2018

Final Report

Cosmic Background Studies for the SBN Collaboration

Author:
Federico Battisti

Supervisor:
Minerba Betancourt

October 16, 2018

1 Neutrino Oscillations and SBN

The field of neutrino physics has rapidly evolved over the past two decades thanks to many successful experiments which have established that neutrinos have mass and oscillate between different flavors. More recently, many of the neutrino mixing parameters which govern oscillations have been precisely measured. There are still many ananswered questions regarding neutrinos behaviour though, such as: are there additional neutrino flavour states that are invisible to our detectors since they don't interact? How do neutrinos aquire mass? and so on.

Many experiments are being employed right now in order to answer some of these questions, most of which make use of the Liquid Argon Time Projection Chamber detectors, who have been chosen or their fine-grained spatial and calorimetric resolution which can be used to accurately identify particles produced in neutrino interactions with low detection thresholds. One of these experiments is the Short Baseline Neutrino (SBN) collaboration.

In order to understand what the goals and difficulties of this experiment are, in this section we present a review of the current status of neutrino oscillation physics, together with an overview of the anomalies that motivate the construction of a short baseline oscillation neutrino experiment.

1.1 Neutrino oscillation physics

Neutrino oscillations describe the phenomenon of neutrino flavor mixing which has now been observed by numerous experiments performing a variety of measurements, the first of which were the Super-K and SNO collaborations. The mixing of neutrino flavors is a quantum mechanical effect caused by the non-orthogonality of the neutrino flavor eigenstate basis (through which neutrinos are observed as they interact via the weak force) and the mass-eigenstate basis which appears in the Hamiltonian and governs neutrino propagation.

1.1.1 Two-flavor mixing

A simple approach to explaining neutrino flavor mixing is to assume the existence of only two flavor eigenstates and then generalize to the three real eigenstates dictated by the standard model and experimental observations [1]. Following this approach, let's assume the exisistance of two flavor states

$|\nu_\alpha\rangle$ and $|\nu_\beta\rangle$ and two mass states $|\nu_1\rangle$ and $|\nu_2\rangle$. The two eigenstate cathegories are related by a unitary matrix:

$$\begin{pmatrix}|\nu_\alpha\rangle & |\nu_\beta\rangle\end{pmatrix} = \begin{pmatrix} cos\theta & sin\theta \\ -sin\theta & cos\theta \end{pmatrix} \begin{pmatrix}|\nu_1\rangle \\ |\nu_2\rangle\end{pmatrix} \tag{1}$$

Neutrinos interact as well-defined flavor eigenstates, but they propagate as mass eigenstates of the Hamiltonian. A neutrino which was originally produced in the flavor eigenstate $|\nu_\alpha\rangle$ will evolve as:

$$|\nu_\alpha(t,\vec{r})\rangle = cos\theta\ e^{-ip_1 r}\,|\nu_1\rangle + sin\theta\ e^{-ip_2 r}\,|\nu_2\rangle \tag{2}$$

p_1 and p_2 are the mass eigenstates' momentums, which can be written, assuming the neutrino is ultra-relativistic, as:

$$p_i = \sqrt{E^2 - m_i^2} \approx E[1 - \frac{m_i^2}{2}] \tag{3}$$

$$\boldsymbol{p_i} \cdot \boldsymbol{r} = E \cdot t - \vec{p_i}\vec{r} \approx (E - p_i)r = \frac{m_i^2}{2E} r \tag{4}$$

r being the distance travelled by the neutrino. From this we get:

$$|\nu_\alpha(r)\rangle = \cos\theta\ e^{-i\frac{m_1^2}{2E}r}\,|\nu_1\rangle + sin\theta\ e^{-i\frac{m_2^2}{2E}r}\,|\nu_2\rangle \tag{5}$$

The probability of the neutrino interacting as the flavor eigenstate ν_α will be given by the amplitude:

$$P_{\alpha\alpha} = |\,\langle\nu_\alpha|\nu_\alpha(r)\rangle\,|^2 = 1 - sin^2(2\theta)sin^2(\frac{\Delta m_{12}^2 L}{4E}) \tag{6}$$

As it is clear by the formula the probability for a neutrino produced in a certain state to be detected in the same state (referred to as survival probability) depends on two fundamental quantities set by nature: the mixing angle θ, which dictates the oscillation amplitude, and the squared difference between the mass eigenstates $\Delta m_{12}^2 = m_2^2 - m_1^2$. The energy E and propagation distance of the neutrino L are the only two that can be directly manipulated, which contribute to the oscillation frequency.

1.1.2 Basic structure of a Neutrino Oscillation Experiment

Measuring the mixing angles and mass-splittings of equation 6 requires mapping out the oscillation probability as a function of L and E. In order to do so one must be able to infer the neutrino energy, propagation distance, and neutrino flavor. One possible source of neutrinos are accelerators, which have the advantage of a very well known L and an energy spectrum that can somewhat be tuned.

The experimental setup for an accelerator based oscillation experiment is shown in Figure 1 [2]. A near detector is placed close to the un-oscillated source, the neutrino flux can be accuratly scanned in order to exactly know its initial composition . The "far" detector is then placed at a baseline such that the oscillation feature being studied ($\Delta m^2 L/4E$)is enhanced; its task is to count neutrino interactions, tag their flavor, and measure the neutrino energy.
Oscillation experiments can be performed as either appearance or disappearance measurements. In the first case, one searches for the appearance of neutrinos of flavor β in a ν_α beam. For disappearance experiments the survival probability $P_{\alpha\alpha}$ of the state α is measured.

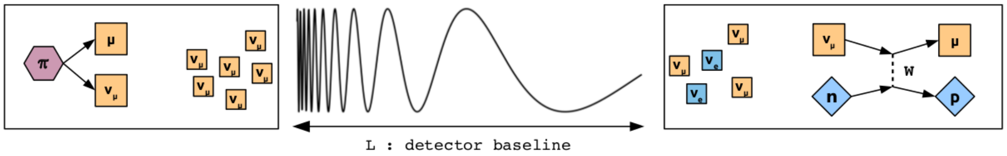


Figure 1: Layout of a basic Oscillation experiment

1.1.3 Three Flavor Neutrino Oscillations

The two-flavor formalism discussed in section 1.1.1 can be generalized to include other neutrino states. The EW unitary matrix, in the case of three neutrino flavors can be expressed as:

$$\boldsymbol{U} = \begin{pmatrix} U_{e1} & U_{e2} & Ue3 \\ U_{\mu 1} & U_{\mu 2} & U\mu 3 \\ U_{\tau 1} & U_{\tau 2} & U\tau 3 \end{pmatrix} \tag{7}$$

This matrix can be divided into three terms, one for each mixing angle θ_{12}, θ_{23} and θ_{13}:

$$\boldsymbol{U} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & cos\theta_{23} & sin\theta_{23} \\ 0 & -sin\theta_{23} & cos\theta_{23} \end{pmatrix} \times \begin{pmatrix} cos\theta_{13} & 0 & sin\theta_{13}e^{-i\delta} \\ 0 & 1 & 0 \\ -sin\theta_{13} & 0 & cos\theta_{13} \end{pmatrix} \times$$

$$\begin{pmatrix} cos\theta_{12} & sin\theta_{12} & 0 \\ -sin\theta_{12} & cos\theta_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \times \begin{pmatrix} e^{i\alpha_1/2} & 0 & 0 \\ 0 & e^{i\alpha_2/2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (8)$$

The matrix $\boldsymbol{U}$ is called Pontecorvo, Maki, Nakagawa, and Sakata (PMNS) matrix. Oscillation probabilities can be derived as for the two-neutrino simple example, but now will exhibit contributions from three mixing angles and mass-splittings which interfere with each other. The generic oscillation probability in a three-flavor scenario can be expressed as:

$$P_{\nu_\alpha \to \nu_\beta (\bar{\nu_\alpha} \to \bar{\nu_\beta})} = \sum_i |U_{\beta i}|^2 |U_{\alpha i}|^2 +$$

$$2 \sum_i |U_{\beta j} U^*_{\alpha j} U_{\alpha i} U^*_{\beta i}| cos\left(\frac{\Delta m_{ij}}{2E} L - (+) \ arg(U_{\beta j} U^*_{\alpha j} U_{\alpha i} U^*_{\beta i}) \right) \quad (9)$$

Note that the mixing angles of the PMNS matrix are accompanied by three phases: δ which measures the CP violating angle, and α_1 and α_2 which are the Majorana phases, important if neutrinos are Majorana particles.

1.2 Hints for Additional neutrino States

The absence of a right-handed neutrino in the SM EW theory causes neutrinos to be massless. Neutrinos (anti-neutrinos) are produced via weak interactions in a predominantly left-handed (right-handed) state. For non-zero neutrino masses, the neutrino field can be expressed as a linear superposition of a left-handed helicity -1/2 and right handed helicity +1/2 states, with the right-handed component heavily suppressed by a term proportional to m_ν/E. Right handed sterile neutrinos (which do not interact with the weak bosons W^{+-} and Z^0) can be added as an extension to the SM.
Hints for sterile neutrino states have emerged from several neutrino experiments performed with different sources and detectors [3], even though none of these measurements offers yet a conclusive evidence of their existence and of oscillation into and from them. All potential signal, if interpreted as due to oscillations, are compatible with a mass-splitting of order 1 eV^2.

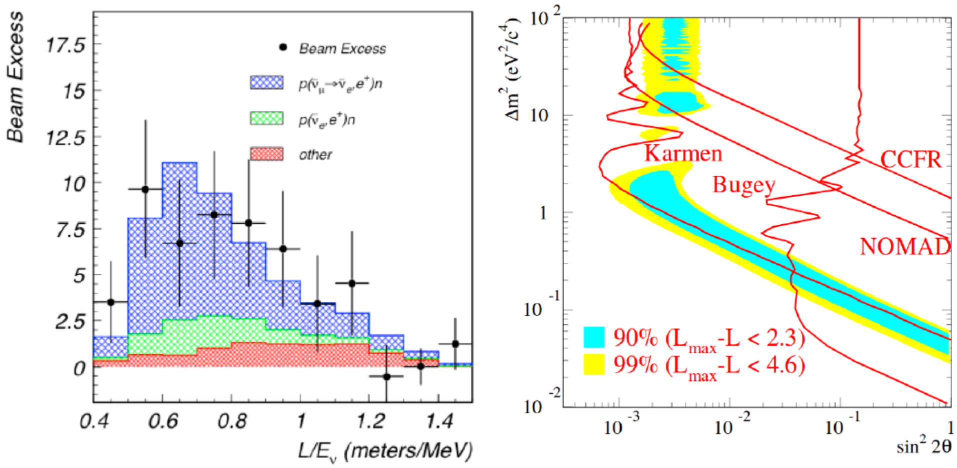
1.2.1 The Gallium Anomaly and the Reactor Anomaly

GALLEX and SAGE were solar neutrino detectors working in the spectrum of the few hundred keV. During calibration measurements with neutrino sources ^{51}Cr and ^{37}Ar these experiments measured a deficit of ν_μ from both of order $\sim 2-3\sigma$. Being the deficit from a low energy source (400-700 keV) 10 cm from the detector led to speculation that the cause of this anomaly could be due to oscillations to sterile neutrinos in the eV^2 range.
A similar anomaly has been widely observed in reactor neutrino experiments as a deficit of a few percentage points in $\sim 1-10\ MeV\ \bar{\nu}_e$ events,uniform in events. Such deficit has been hypothesised to also be linked to oscillations into a sterile neutrino.

1.2.2 The LSND Anomaly

The LSND (Liquid Scintillator Neutrino detector) operated recording the neutrino flux produced by the LANSCE neutrino beam, composed of ν_μ and ν_e from muon and pion decay at rest. The main goal of the experiment was to study neutrino oscillations at the $\Delta m^2 \sim 1eV^2$ at a time when there still was much uncertainty on the frequencies at which to expect oscillations. The analysis consisted of a measurement of ν_e from inverse β decay and recorded an excess of electron neutrino interactions with respect to predictions for intrinsic events from the beam [4]. The excess of events (Figure 2 (a)) was fit

to an oscillation appearance signal under a two-flavor hypothesis, leaving to the constrained parameter space shown in Figure 2 (b).
The $\sim 1eV^2$ mass splitting suggested by LSND's results interpretation as neutrino oscillation is non-conclusive, but it has led to an interest in the field for the search for potential sterile neutrinos in that range.



(a)

(b)

Figure 2: (a) Distribution of measured ν_e events, with the best-fit excess modeled as oscillations in blue(b) Constraints on parameter space under a two-flavor oscillation hypothesis

1.2.3 The MiniBooNe Low Energy Excess

MiniBooNE's goal was to test the 1 eV^2 oscillations hypothesis but with a different neutrino beam and detector in order to not be affected by the same systematics of LSND. The detector consists of an 800 ton circular tank filled with mineral oil surrounded by PMTs which observe Cherenkov light produced by leptons and other particles originating in neutrino interactions. The experiment was a ν_e appearence experiment in a largely ν_μ beam. The detector was able to distinguish the two topologies by the difference in PMT signature: electrons produced unfocused Cherenkov rings caused by the broadening elecro-magnetic showers, while muons produced a sharp Cherenkov ring.

The latest results of MiniBooNe's ν_e and $\bar{\nu}_e$ appearance searches present an excess of events in both channels that falls below 475 MeV of reconstructed neutrino energy. This excess falls below where an appearance signal consistent with that of LSND would have been expected, and is foer this reason referred to as Low Energy Excess (LEE). Sterile neutrino states of $\Delta m^2 \sim 1eV^2$ have been proposed as an explanation for this anomaly as well , even though the data would seem to suggest a more complex sterile neutrino model with two or three additional neutrino states.

A clear picture for what is the nature of the MiniBooNE LEE is yet to be identified. The most significant backgrounds to the MiniBooNE low-energy excess are of two types: intrinsic electron neutrinos in the muon neutrino beams, and interactions producing photons which are mis-identified as electrons. While the first background is irreducible, the second is present because MiniBooNE cannot distinguish between electrons and photons. This is one of the main arguements that led to the development of new experiments using the LArTPC technology.

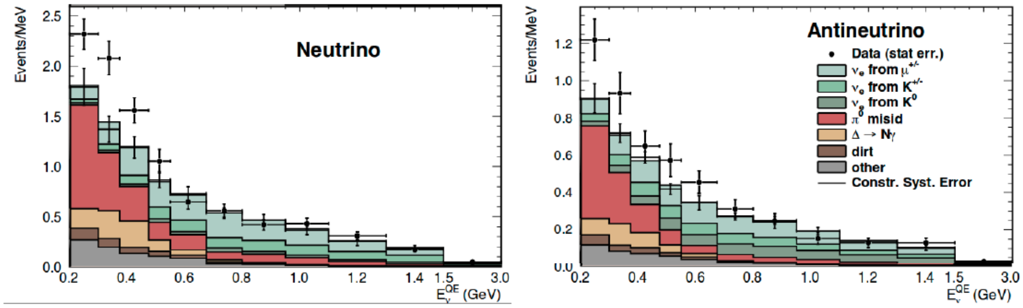


Figure 3: Neutrino mode (a) and anti-neutrino mode (b) results for a ν_e appearance search in the MiniBooNE detector. The excess of events at low energies (¡ 475 MeV) is referred to as the low energy excess (LEE). The same excess seen by LSND would show up in MiniBooNE at slightly higher energies. The dominant backgrounds to the selection are caused by single photon events from $\Delta \rightarrow N\gamma$ and neutral current π_0 production.

1.3 SBN

The Short Baseline Neutrino (SBN) Program is an extensive experimental edeavour set to explore neutrino properties and detector technology in the Fermilab Booster Neutrino Beam-line [5]. It consists of three Liquid Argon Time Projection Chambers(LArTPCs): SBND, MicroBooNe and ICARUS, placed at varying distances from the neautrino source, running in Booster Neutrino Beam-line(BNB) at Fermilab (Figure 4). The goals of the program are to followup on the MiniBooNE low energy excess, explore the phase space of short baseline neutrino oscillations, make precise measurements of of cross sections of neutrinos interacting in liquid Argon and further develop LArTPC technology.

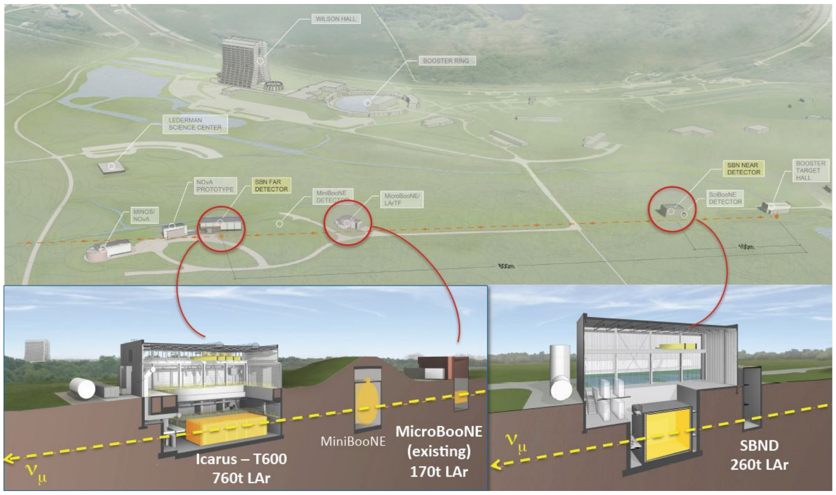


Figure 4: Layout of the three detectors composing the SBN Program.

The primary goal of SBN is to investigate the hints of the possible existence of one or more sterile neutrino states in the 1eV mass range, suggested by the LSND and MiniBooNe experiments (as discussed in Section 1.2). These new sterile states would cause an excess in oscillation towards electron neutrinos at short baselines ($< 1km$) in pion-decay in flight based neutrino beams such as BNB.

The size and placement of the three TPCs allows for a characterization of the neutrino beam before the oscillation in SBND and simoultaneous mea-

surements of ν_e appearence and ν_μ disappearence in the MicroBooNe and ICARUS detectors. The LArTPC technology also solves one of MiniBooNe problems, which was the fact that the Cherenkov light based detectors in use were not able to easily distinguish between electron and photon induced interactions. The new TPCs are able to distinguish between the two topologies by either observing a gap between a neutrino interaction vertex and a shower or by observing calorimetric differences in the beginning of a shower between a $\gamma \rightarrow e^+e^-$ pair production and a single ionazing electron.
The SBN configuration will enable measurements at the 5σ level all in a single experiment, obviating the need for aglobal fit over multiple experiments to reach sensitivity in the $\Delta m^2 \sim 1eV^2$ region.

1.3.1 The LArTPC technology

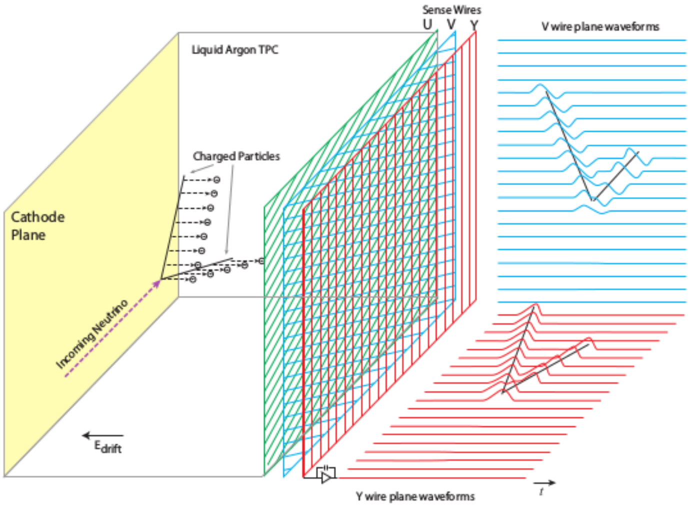


Figure 5: Working principle of a LArTPC. A constant electric field applied to the cathode (left-hand side) produces a uniform drift-field in the drift-coordinate direction. Charged particles traversing the TPC, will ionize the argon, depositing a trail of ionized electrons, which will then begin to drift towards the wire planes under the effect of the electric field

In a LArTPC [6] charged particles lose energy as they travel through the detector volume by exciting and ionizing the argon. Energy lost to ionization will manifest itself as a trail of electrons which faithfully maps the

three-dimensional trajectory of a charged particle. Excited argon atoms can form excimers which then decay leading to the production of scintillation light. The energy lost by the traversing charged particle is split roughly equally between ionization charge and scintillation photons.
Scintillation light will propagate isotropically through the detector volume and can be measured with photomultiplier tubes. Ionization charge, on the other hand, will remain stationary in the detector. In order to measure the ionization charge preserving the topological and calorimetric information it conveys an electric field is applied across the detector volume which drifts the ionization electrons towards a mesh or wires. Ionization electron clouds, by drifting past or being collected on the wires will induce a current which can be recorded. The pattern of charge deposited on the wires as a function of time can be used to produce an image of interactions in the detector of comparable quality to those produced by bubble-chamber detectors.

The basic setup of LArTPC is shown in Figure 5. The TPC has three main components: a cathode-plane, field cage, and anode plane. These are responsible for maintaining a uniform electric field in the detector volume allowing ionization electrons to drift towards the sense-wire. The anode plane houses the sense-wires, usually arranged in three planes, kept at a fixed electric potential in order to shape the electric field in a way that maximizes signal transparency.Drifting electrons will pass by two wire-planes closest to the cathode, producing an induction signal, and will be collected on the last plane, denoted as the collection plane. By allowing the same electron cloud to leave a time-coincident signature on multiple planes with different orientation one is able to produce images which show different two-dimensional projections of the same three- dimensional energy deposition pattern. These multiple view points allow to triangulate the exact 3D location of energy deposited in the detector.

1.3.2 Neutrino-Argon Interactions and LArTPC Detector Development

Future long-baseline neutrino experiments such as DuNe will require much improved neutrino interaction models in order to achieve the precision required for the neutrino oscillation measurements. The SBN Program will make the highest precision cross section measurements of $\nu_e - Ar$ and $\nu_\mu - Ar$ scattering in the few hundreds of MeV to few GeV range, using neutrinos both from the on-axis BNB and off-axis NuMI beams [7].

The SBND design also includes features that are similar to the ones planned for the Deep Underground Neutrino Experiment (DUNE) far detector and so it will be a critical proof of concept of these designs. Scaling up to these larger detector masses will benefit from the developments in, for example, LAr purifications, modular detector anode components, and scintillation-light detection systems used in the SBN program.

In addition simulation, reconstruction and event selection algorithms are being developed in a shared software platform for use by all LArTPC based experiments.

2 Cosmic Background rejection using Boosted Decision Trees

2.1 The Cosmic Background

The SBN LArTPC detectors lack appreciable shielding from cosmic rays (CR) since they are at the earth's surface and have little overburden[8]. Triggered events with a neutrino interaction typically have the products of up to 20 cosmic rays coincident with the beam spill in the event readout window (4.8 msec) contributing to a recorded event along with the products of the neutrino collision. The situation is particularly complicated with events containing neutrino interactions with an isolated μ being produced, which share the same topology with the most common single-muon CR configuration.

In order to separate between these two muon configurations one of the various options is to use a multivariate analysis technique which aims to use a certain number of the variables characterizing the muon tracks and to aggregate them into a single classifier. Some of these techniques, such as boosted tecision trees fall into the cathegory of machine learning.

2.2 Decision Trees and Machine learning

Decision trees are a machine learning technique used more and more commonly in high energy physics [9], first developed and formalized by Breiman with the proposal of the CART algorithm (Classification and Regression Trees). The basic idea behind decision trees consists in extending a simple cut-based analysis into a multivariate technique by continuing to analyse events that fail a particular criterion since most events do not have all characteristics of either signal or background.

Mathematically decision trees are rooted binary trees (Figure 6). A decision tree starts from a root node and then recursevely split into two daughters or branches, until some stopping condition is reached. The algorithm used to grow a tree from a root node can be simply described in a few passages. Consider a sample of signal s_i and background b_i events, each with weights w_i^s and w_i^b respectivelly, described by a set of x_i varibles.This sample constitutes the root node of the decision tree and starting from it this is the algorithm:

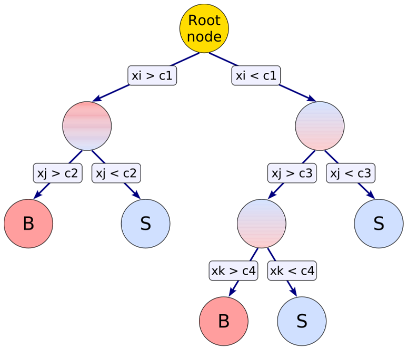


Figure 6: Schematic view of a decision tree. Starting from the root node, a sequence of binary splits using the discriminating variables x_i is applied to the data. Each split uses the variable that at this node gives the best separation between signal and background when being cut on. The same variable may thus be used at several nodes, while others might not be used at all. The leaf nodes at the bottom end of the tree are labeled ”S” for signal and ”B” for background depending on the majority of events that end up in the respective nodes.

1. If the node satisfies a stopping criterion, end the algorithm and lable the sample as a leaf of either signal or background
2. Sort all events according to each variable in x_i
3. For each variable, find the cut value that gives the best separation between one mostly signal branch and one mostly background. If no splitting gives any improvement exit the algorithm and lable as either signal or background leaf.
4. Select the variable and splitting value leading to the best separation and split into two new nodes.
5. Apply recursively from step 1 to each node.

At each node all variables can be considered, even if they have been used in previous iterations: this allows to find intervals of interest either then limiting oneself to simple cuts. The output for the decision tree is defined by the purity of the leaf to which the event will be associated after having passed all the different cuts, where purity is defined as:

$$p = \frac{\sum_{i \in signal} w_s^i}{\sum_{i \in signal} w_s^i + \sum_{i \in background} w_b^i} = \frac{s}{s+b} \tag{10}$$

Background purity will easily be defined as $1 - p$.
To choose what the ideal split for a certain node would be we want to maximize the decrease of impurity from one level to the next. To describe what impurity is we would need a function such that is:

- maximal for an equal mix of background and signal(no separation)
- minimal for nodes with either only signal or only background events
- symmetric in signal and background purities.

Common impurity functions that exhibit most or all these characteristics are:

- the misclassification error: $1 - max(1, 1-p)$
- the cross entropy: $1 - \sum_{1=s,b} p_i log p_i$
- the Gini index of diversity: $2sb/(s+b)^2$

These fuctions are plotted in Figure 7.

Regarding the stopping criterions that can be used to end the growth of the decision tree, the possibilities are various. Some examples of the most common conditions are:

- Reaching a minimum leaf size inquiring at least N_{min} training events in each node after splitting. This will ensure the statistical significance of the purity measurement. In the case of weighted events this criterion can be generalized using the effective number of events instead:

  $$N_{eff} = \frac{(\sum_{i=1}^{N} w_i)^2}{\sum_{i=1}^{N} w_i^2} \tag{11}$$

  for a node with N events with weiths w_i.
- Reaching perfect separation (each event in the node is either signal or background)

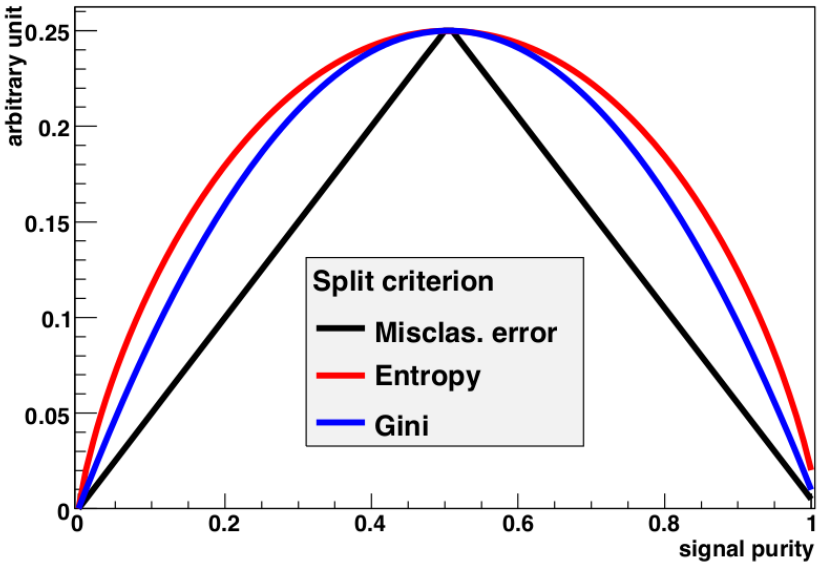


Figure 7: Various popular impurity measures as a function of signal purity.

- Having insufficient improvement with further splitting
- Reaching a maximal tree depth: the tree will have no more then a certain number of layers.

The last condition ensures that the tree structure will remain relatively simple, which is ideal in the case of Boosted Decision Trees.

2.2.1 Boosting

Despite all the nice features presented above, decision trees are known to be relatively unstable. If trees are too optimised for the training sample, they may not generalise very well to unknown events. One of the many techniques that one can use to circumvent this problem is using the technique known as Boosting. The boosting algorithm is not unique to decision tree and can in principle be applied to any classifier.

The basic idea behing the Boosting technique is that other then creating a single very complex multivariate discriminant, which is usually fairly difficult, one could instead build many classifiers that are just slightly better then a random guess . One could then combine the output of all these so called weak classifiers obtaining a new, more stable one with a smaller error rate and better performance.

The basic algorithm of any boosting technique can be streamlined in a few iterative steps. Consider a training sample T_k containing N_k events each associated with a weight w_i^k, a vector of carachterizing variables $\vec{x}$ and a class label $y = \pm 1$ for signal or background respectivelly. What you do then is:

- Initialize T_1 by assignining to each event a weight equal to $1/N_1$
- Train the classifier C_1 on T_1
- Reweight the sample T_1 by some criterion and rename it as T_2
- Assign a weight α_1 to C_1
- Repeat the process for the number of classifiers you want to train (N_{train})

The Boosted output will be some function $F(C_1, ..., C_{N_{train}})$, typically a weighted average:

$$F(i) = \sum_{k=1}^{N_{train}} \alpha_k T_k(\vec{x_i}) \tag{12}$$

2.2.2 AdaBoost

AdaBoost is one of the most succsessful and most used implementations of the boosting algorithm. The term stands for adaptive Boosting, referring to the fact that during the training procedure the classifier is adjusted to the data in order to better classify it. Its implementation works as follows. After the C_k tree has been built one checks for the events in the training sample T_k that have been misclassified and defines the misclassification rate $R(C_k)$ as the number of misclassified events over the total. Let's now define a boolean function $I : x \to I(X)$ such that $I(X) = 1)$ if X is true. One can then define a function that tells wether an event is misclassified by the tree. If the output of the tree is given by the purity of the leaf to which the event belongs to, with a critical purity of 0.5, then the fuction can be defined as:

$$M_k(i) = I(y_i \times T_k(i) \leq 0) \tag{13}$$

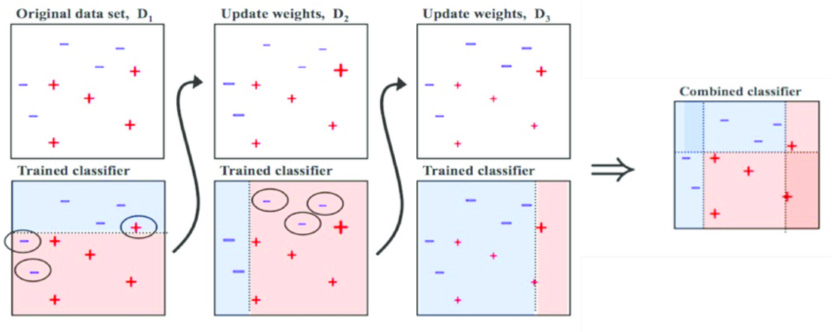


Figure 8: Schematic rappresentation of the AdaBoost algorithm. The first classifier is trained over a evenly weighted sample containing two classes of events: + and − . The weights of the misclassified events are then Boosted in order to be more heavily scrutinized by the next tree in the chain (the symbols are litteraly bigger in the picture). After this process has been repeated a set number of times the output of all the trees is combined in order to build the final Boosted classifier.

The misclassification rate can then be defined as:

$$R(T_k) = \epsilon_k = \frac{\sum_{i=1}^{N_k} w_i^k \times M_k(i)}{\sum_{i=1}^{N_k} w_i^k} \tag{14}$$

We can then assign to the tree T_k a weight that is inversely proportional to its ϵ_k:

$$\alpha_k = \beta \times ln \frac{1-\epsilon_k}{\epsilon_k} \tag{15}$$

where β is a free parameter that can be used to adjust the strength of Boosting.

The caracterizing step of the AdaBoost algorithm is the following: each event in the sample T_k is reweighted in order to create a new sample T_{k+1} such that:

$$w_i^k \to w_i^{k+1} = w_i^k \times e^{\alpha_k \cdot M_k(i)} \tag{16}$$

This means that properly classified events remain unchanged, while the misclassified ones increase their weight by a factor of e^{α_k}. The next tree is then trained on the new sample so that it will be more interested in correctly classifying the events that the previous itheration found more difficult and failed to identify correctly. This iterative process is repeated for the number of classifiers we want to train N_{train} and the final AdaBoost output is given by the weighted average:

$$T(i) = \frac{1}{\sum_{k=1}^{N_{tree}} \alpha_k} \sum_{k=1}^{N_{tree}} \alpha_k T_k(i) \tag{17}$$

It can be shown that the misclassification rate ϵ of the Boosted classifier on the training sample is bounded from above:

$$\epsilon \leq \prod_{k=1}^{N_train} 2\sqrt{\epsilon_k(1-\epsilon_k)} \tag{18}$$

This leads to the fact that for a sufficiently large number of N_{train}, if each tree is at least better then a random guess ($\epsilon_k \neq 0.5$) ϵ falls to 0. At the same time though the classifier goes into overfitting, so that one should find a good balance beetween ability of classification and specificity.

2.2.3 TMVA

Integrated into the analysis framework ROOT, TMVA is a toolkit which hosts a large variety of multivariate classification algorithms. Training, testing, performance evaluation and application of all available classifiers is carried out simultaneously via user-friendly interfaces [10].

A typical TMVA classification (in our case netween signal and cosmic background) analysis consists of two independent phases: the training phase, where the multivariate methods are trained, tested and evaluated, and an application phase,where the chosen methods are applied to the concrete classification problem they have been trained for. For this project we limited ourself to the first portion of this process since the goal was to evaluate how well a trained Boosted Decision Tree would perform over simulated muon data, and not to use it on any real sample, at least for the moment being.

In the training phase, the communication of the user with the data sets and the MVA methods is performed via a Factory object, created at the beginning of the program. The TMVA Factory provides member functions to specify the training and test data sets, to register the discriminating input and – in case of regression – target variables, and to book the multivariate methods. Subsequently the Factory calls for training, testing and the evaluation of the booked MVA methods. Specific result ("weight") files are created after the training phase by each booked MVA method.

In our case ROOT Trees were used to store the signal and background events with their defining variables. Data trees can be provided specifically for the purpose of either training or testing or for both purposes. In the latter case, which corresponds to our own, the factory then splits the tree into one part for training, the other for testing.

2.3 Preliminary Fiducial Cuts and Definition of the Classifier Variables

Using the functionalities of TMVA building and training any Multivariate machine learning classifiers is fairly easy; one only needs to choose which kind of classifier he wants to use and which variables he wants to implement, and provide a training/testing sample in the form of tree files. The kind of classifiers we decided to experiment with were, as previously mentioned, are Boosted Decision Trees.

The data we used was Montecarlo simulated data generated for SBND and reconstructed using Pandora and PMA algorithms [11]for Signal and Back-Ground samples respectivelly and contained 74035 signal tracks and 491317 cosmic tracks. As for the spatial coordinates the z axis corrisponded to the initial direction of the muon neutrino beam, while the x and y correspond to the horizzontal and vertical directions exactly. The dimensions of the detector are then:

- $x = [-200; 200]\ cm$
- $y = [-200; 200]\ cm$
- $z = [0; 500]\ cm$

The reconstruction trees contained all sorts of informations regarding the muon tracks we wanted to analyzed, so we decided to build new trees containing only the information directly relevant to our analysis. Such data consisted of basic spatial informations such as three-dimensional positions of the beginning and end of the track and the variables we used to build the Boosted Decision Tree Classifier.

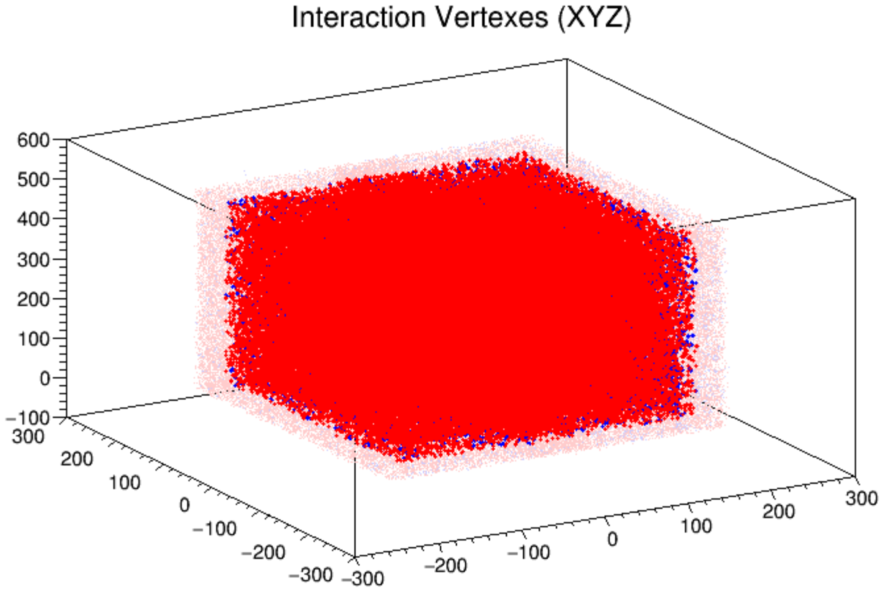


Figure 9: 3D distributions of the origins of the muon tracks from the Montecarlo reconstructed data, produced by ν_μ (in blue) and by cosmic rays (in red) inside the SBND detector. The faded dots correspond to the events located in the outer $20cm$ of the detector, that were eliminated by the fiducial cuts. The sample is mainly constituded by cosmic tracks, roughly respecting the real proportions that are expected for a surface LArTPC

The spatial information regarding the origin of the tracks was used to apply a basic fiducial cut over the two samples. A fiducial volume is defined as the volume of the detector outside of which we can expect that most signal is actually produced by background sources. In the case of muon tracks in a surface LArTPC this corresponds to the internal volume of the detector. This is justified intuitively by the fact that if the neutrino beam is well focused the muon tracks produced by it will not diverge much laterally, and that the ν_μ will take some time to interact in the Liquid Argon Medium. Removing the outer $20cm$ of the detector volume in each direction we obtained a sample that was reduced by 31% in signal and by 39.4% in background (Figure 9) with an overall increase in $signal/background$.

Having done these preliminary cuts, all that was left to do was to select which characterizing variables we would use. We decided to let ourself be guided in our choice by a previous similar work produced at NoVa, where many different observables where tried in the context of BDT classifiers [12]. The specific ones we decided on in the end where:

(a) The length of the muon track

(b) The cosine of the angle between the direction of the muon track and the average direction the neutrino flux estimated using the Montecarlo simulation truth information on the ν_μ

(c) The maximum value of the vertical coordinate y between the beginning and the end of the track

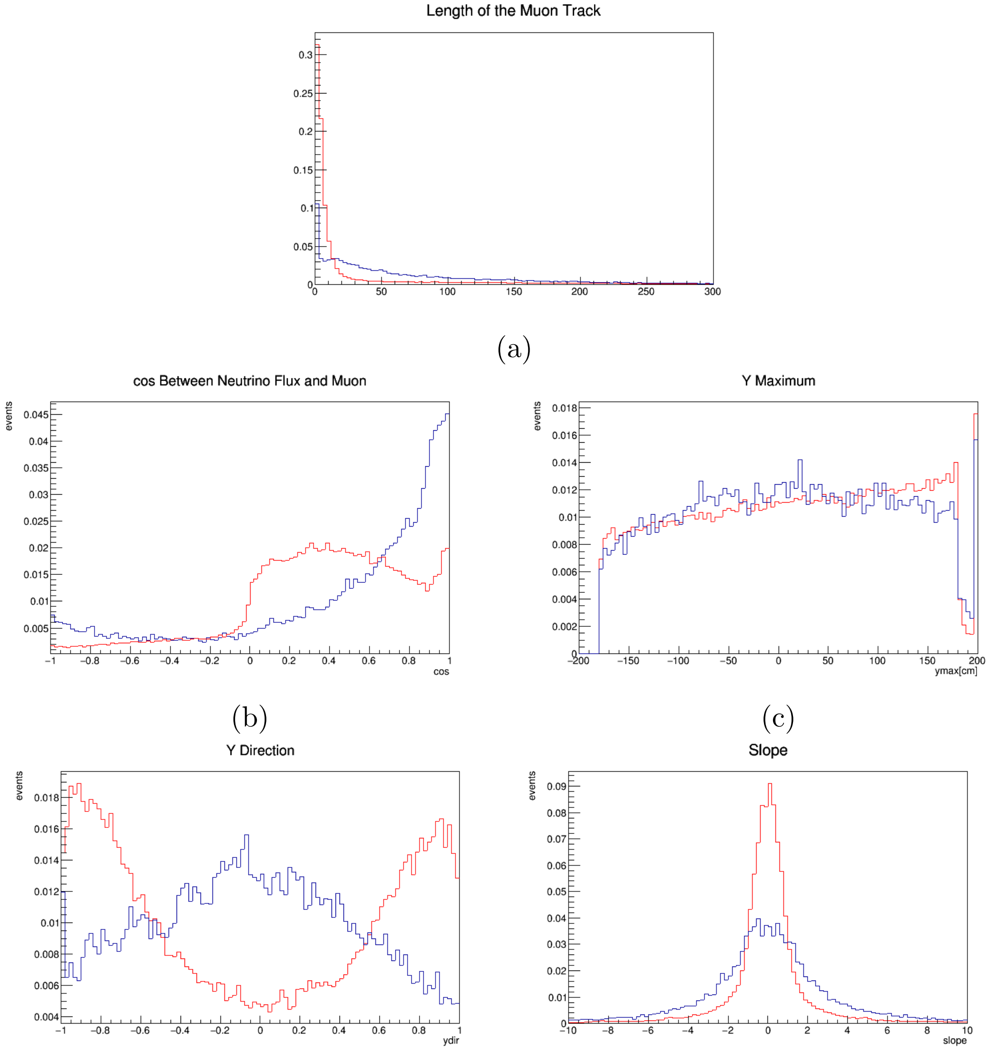
(d) The cosine between the direction of the muon track and the y axis

(e) The slope of the muon track defined as :

$$slope = \frac{z_2 - z_1}{y_2 - y_1} \tag{19}$$

where z_2 and z_1 are the values of the z coordinate at the end and beginning of the muon track and y_2 and y_1 are the equivalents for the y coordinate.

All the histograms of the distributions of the variables for the signal and background samples are reported in Figure 10.



(a)

(b)

(c)

(d)

(e)

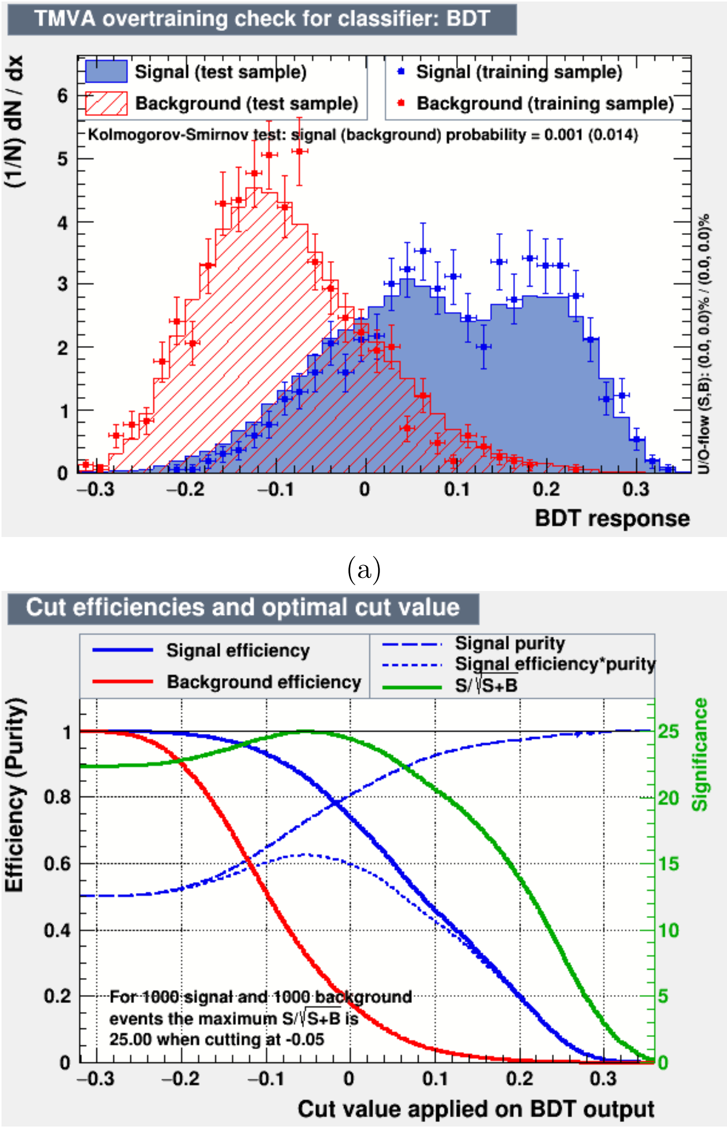
Figure 10: Histograms of signal (in blue) and background (in red) events for each variable used for the BDT (a) length of the muo track (b)cosine of the angle between the direction of the muon track and the average direction the neutrino flux (c) maximum value of the vertical coordinate y between the beginning and the end of the track (d) cosine between the direction of the muon track and the y axis (e) slope of the muon track

2.4 BDT Training and Testing and Results

Using the variables described in the previous section we wrote a root macro that, using the functionalities of TMVA, would let us build, train and test a BDT classifier over the neutrino and cosmic files mentioned in section.
In Figure (a) we can see the distribution of the trained BDT over the training and testing samples which were selected automatically by TMVA from the SBND files. Note that the performances over the two samples are very similar, which indicates that the classifier did not get into overtraining.
In Figure (b) we can see how the signal and background efficiency, which are defined as rhe fraction of events that are correctly classified as signal or background respectively, vary with respects to the possible cuts on the trained BDT output. To decide on where to apply the cut we used a well known factor of quality defined as:

$$f = \frac{s}{\sqrt{(s \cdot b)}} \tag{20}$$

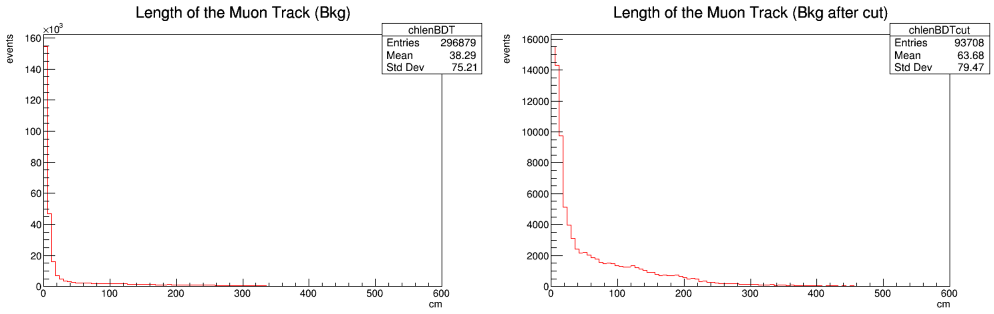
where s and b are the signal and background efficiencies respectivly. The value of f resulted maximal for a cut at $BDT = -0.05$. The results of this cut can be seen in Figure where we used the length variable as an example. After the cut the signal sample was reduced by 14%, while the background sample was reduced by 69% resulting in a clear improvement in signal purity. This demonstrates the viability of the technique and incoourages further improvement, i.e. over the quantity and quality of the variables utilized.



(a)

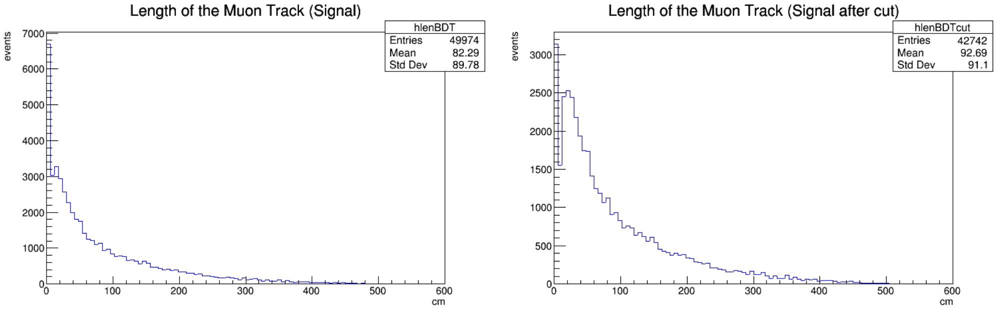
(b)

Figure 11: (a) Output of the trained BDT classifier over the testingand training samples (b) Signal efficiency s, Background efficiency b, Signal purity and $s/\sqrt{(s \cdot b)}$ as a function of the cut on the BDT output



(a) (b)

Figure 12: Distributions of the length of the muon track for the background events (a) before the cut on the BDT output (b) after the cut on the BDT output



(a) (b)

Figure 13: Distributions of the length of the muon track for the signal events (a) before the cut on the BDT output (b) after the cut on the BDT output

3 Study of the Michel electrons in LArTPC

Michel electrons are defined as electrons produced by the decay-at-rest of cosmic-ray muons that come to a stop in the LArTPC. Michel electrons have a well characterized energy spectrum with energies up to 50 MeV. We can use these electrons to study the response to neutrino interactions which produce EM activity in the same energy range.

What we tried to do in this study was to develop an algorithm that would let us identify those muons that are the most likely to have produced a Michel electron in their interaction with the liquid Argon medium.
The files used for this analysis are SBND Montecarlo produced reconstructed cosmic files analogous to the ones described in Section 2.3.

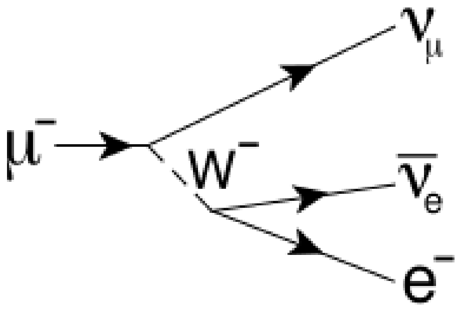


Figure 14: Feynmann diagramm of the decay at rest of a muon into an electron, a muon neutrino and an electron anti-neutrino, via weak interaction

3.1 Topology of the interaction

In order to develop an algorithm that selects the muons that are the most likely to stop in the detector and produce an electron, we looked at the energy deposition of the particle and the topology of the interaction. Muons travelling in the medium will use energy by ionizing the electrons of the liquid Argon medium, producing a charachteristic rise in energy deposition along the particle's trajectory, when they come to a stop, called the Bragg peak (Figure). Most Michel electrons produced will then propagate in a direction different than that of the stopping muon and will lose energy by either ionizing and exciting the atoms which make up the material (collision

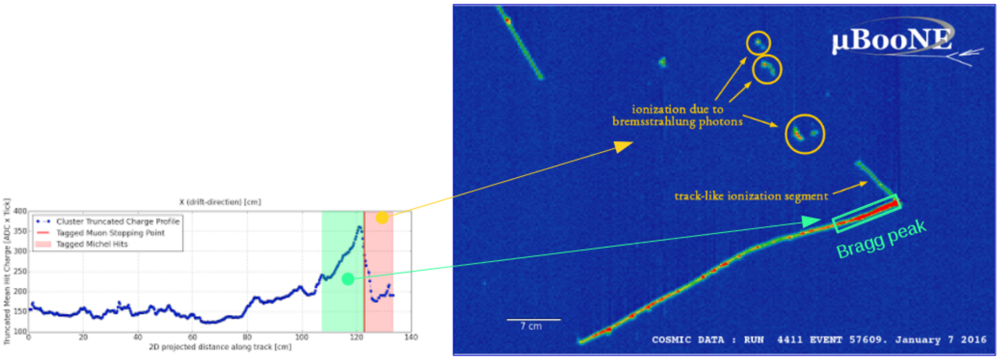


Figure 15: Visualization of the topology of an interaction of a muon producing a Michel electron: spatial 2d reconstruction on the right, energy deposition on the left. Reconstruction taken from a MicroBoone analysis [13]

stopping power) or by producing photons from a bremsstrahlung interaction (radiative stopping power).

Concentrating only on the energy deposition of the stopping muon we can develop an algorithm that lets us do a confrontation between the mean charge deposition between the beginning and the end of the muon track: the muons where this value is bigger at the end then at the beginning of the track will be more likely to have stopped in the medium and have produced a Michel electron.

Before we do that it is usefull to smooth out the hit charge data.Given a spatially sorted list of 3D hits we can substitute the charge deposition values with the median of the surrounding hits (4 on each side), eliminating many possible outliers. After having done that we can take the and last 10 cm of the muon track and calculate the median of the charge deposition in each region. If we plot the two values (figure) we will expect two main regions to be formed: one where the $dqdx_{beg}$ and $dqdx_{end}$ have about the same value and one where $dqdx_{end} > dqdx_{beg}$

3.2 Spacial and Energetic Cuts

Having access to the spatial and energetic information we can try to apply various selections criterions to the data, such as:

- The muon track must be fully contained in the volume of the SBND detector: the beginning of the track must be contained in the outer 20 cm of the detector and the end must be contained in the complementary volume.

- The median of the charge deposition in the last 10 cm of the muon track must be bigger then the one at the beginning

We can see the effects of these cuts on the data in Figure 17: we reduced the data sample from 123759 events in the original sample to 5561 events. In the new set it was easy to find a few that corresponded to the characteristic Bragg peak structure, such as the one in Figure 15.
To further develop the algorithm one could now look for example to the change in direction of the track after the production of the electron or the energy deposition of the electron itseld, as it was done in a previous work for the MicroBooNe experiment [13].

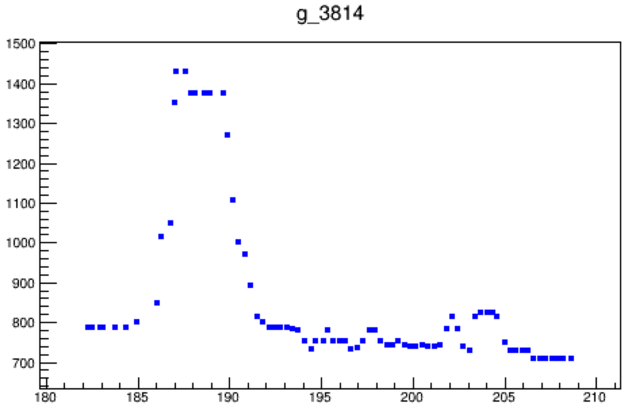
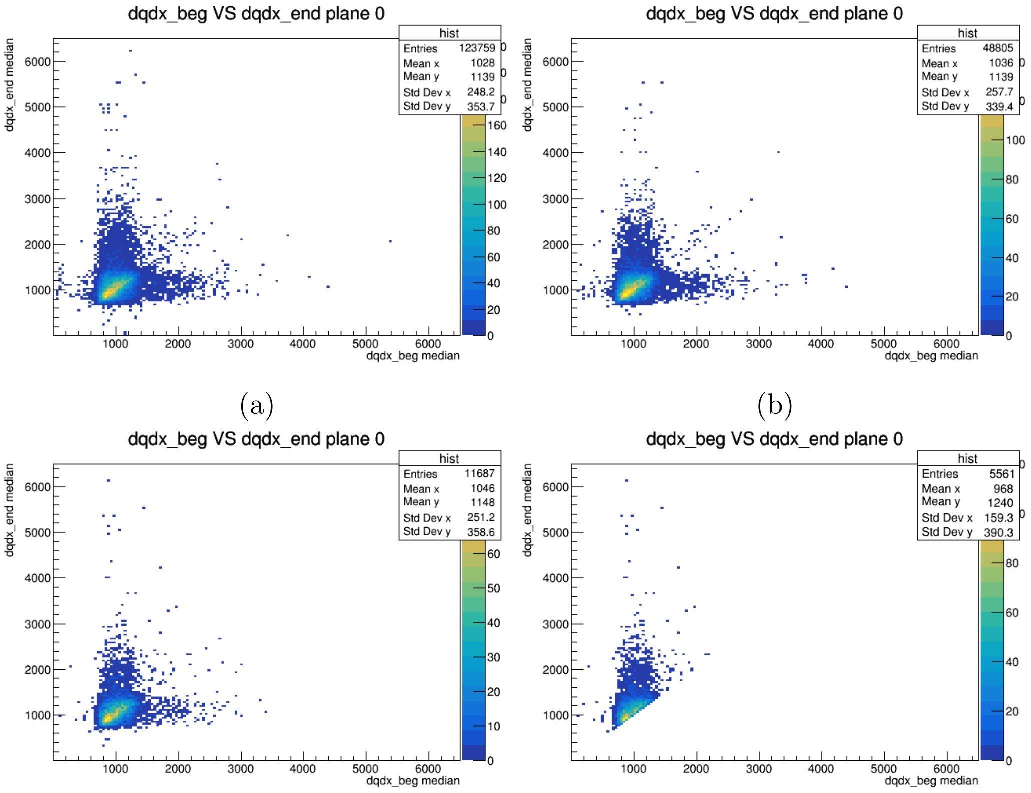


Figure 16: Example of an event presenting a Bragg peak in its energy deposition (keV) as a function of residual range (distance in cm from the end of the track)



(a) (b)

(c) (d)

Figure 17: Histograms of the $dqdx_{beg}$ vs $dqdx_{end}$ in keV (a) before any cut (b) after imposing that the beginning of the track is contained in the outer 20 cm of the SBND detector (c) after imposing that the end of the track is contained in the complementary volume (d) after imposing that $dqdx_{beg} > dqdx_{end}$

References

[1] Andre de Gouvea, *2004 TASI Lectures on Neutrino Physics*, NUHEP-TH/04-17, arXiv:0411274

[2] David Caratelli, *2018 Study of Electromagnetic Interactions in the MicroBooNE Liquid Argon Time Projection Chamber*,

[3] J. Conrad, and M. Shaevitz, *Sterile Neutrinos:An Introduction to Experiments,*, arXiv:1609.07803

[4] J. Conrad, W. Louis, M. Shaevitz, *The LSND and MiniBooNE Oscillation Searches at High* Δm^2*,*, Δm2 , Ann. Rev. Nucl. Part. Sci. 63 (2013) pp. 45-67.

[5] R. Acciarri et. al. *A Proposal for a Three Detector Short-Baseline Neutrino Oscillation Program in the Fermilab Booster Neutrino Beam*, arXiv:1503.01520

[6] C.Rubbia *The Liquid-Argon Time projection Chamber: a new concept for Neutrino Detector,*, CERN-EP/77-08, 1977

[7] Matthew Bass *The Short Baseline Neutrino Oscillation Program at Fermilab*, arXiv:1702.00990v1 [physics.ins-det] 3 Feb 2017

[8] The MicroBooNe collaboration *Comparison of* $\nu_\mu - AR$ *multiplicity distributions observed by MicroBooNE to GENIE model predictions*, arXiv:1805.06887v1 [hep-ex] 17 May 2018

[9] Yann COADOU *Boosted Decision Trees and Applications*, EPJ Web of Conferences 55 , 02004 (2013), DOI:10.1051/ epjconf / 20135502004

[10] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, H. Voss *TMVA 4 Users Guide*, arXiv:physics/0703039, [Data Analysis, Statistics and Probability], CERN-OPEN-2007-007, TMVA version 4.2.0, March 3, 2017

[11] MicroBooNE collaboration *The Pandora multi-algorithm approach to automated pattern recognition of cosmic-ray muon and neutrino events in the MicroBooNE detector*, arXiv:1708.03135 [hep-ex]

[12] NOvA collaboration *NOvA Cosmic Rejection*, NOVA-doc-11205-v5

[13] MicroBooNE collaboration *Michel Electron Reconstruction Using Cosmic-Ray Data from the MicroBooNE LArTPC*, arXiv:1704.02927v2 [physics.ins-det] 30 Aug 2017