Alma Mater Studiorum · University of Bologna

School of Science
Department of Physics and Astronomy
Master Degree in Physics

Cosmic Ray Tagging in Icarus: a combined analysis

Supervisor:
Dr. Gabriele Sirri

Co-supervisor:
Dr. Franceso Poppi

Submitted by:
Ricardo Andrés Jaimes Campos

Academic Year 2023/2024

Table of contents

Introduction

In the last decades the study of neutrino physics has been gaining momentum as a portal to access physics beyond the standard model. The measurement of the oscillation parameters of the three known families, ν_e, ν_μ, ν_τ, is the most relevant result in the field to date.

Neutrinos are assumed to be massless in the standard model of particle physics, propagating as wave packets at the speed of light in each of the three flavor eigenstates. However, inconsistencies in the measurements of solar and atmospheric neutrino fluxes lead to the discovery of neutrino oscillations implying that neutrinos are massive particles. The phenomenon was discovered in 1998 and since then it has been confirmed several times in experiments based on solar, atmospheric, and accelerator beam neutrinos. While most of the experimental data can be accommodated in the framework of three neutrino oscillations, the results reported by some experiments at relative short baselines do not fit in the global picture requiring (at least) an additional neutrino mass state that does not participate in weak interactions (namely *the sterile neutrino*).

The Short-Baseline Neutrino Program at Fermilab (SBN) aims to perform the most sensitive search for sterile neutrinos at the eV mass scale. The program includes a far detector and a near detector located along the Booster neutrino beamline (BNB), at about 600 m and 100 m, respectively, from the neutrino source. Both detectors exploit the Liquid Argon Time Projection Chamber (LArTPC) technique. The ICARUS detector which serves as SBN far detector, is also exposed, in off-axis position to the Neutrino Main Injector

Beam (NuMI).

This thesis addresses the issue of cosmic ray tagging in the ICARUS detector. ICARUS is located at a shallow depth and is exposed to a significant flux of cosmic rays during the neutrino beam spill, which lasts for 1.6 μs for the BNB, and even during the drift window.

To reduce the impact of the cosmic rays, a Cosmic Ray Tagger (CRT) system has been installed, complemented by a six m.w.e. layer of concrete overburden above the CRT. This setup is designed to suppress the soft electromagnetic and hadronic components of the cosmic rays and to tag muons that cross the detector.

The initial commissioning phase of ICARUS was completed in June 2022, and physics runs started with both the BNB and NuMI beams. Early analysis of the cosmic background rejection has been performed exploiting the time coincidence between the CRT and photon detection (PMT) systems. The CRT-PMT matching technique has demonstrated a conservative rejection of over 65% of both in-time and out-of-time cosmic rays.

This work begins with a brief overview of neutrino oscillations and the hints for sterile neutrinos. Then the SBN is presented and the specifics of the ICARUS detector. After a detailed examination of the ICARUS subsystems, the thesis explores how a combined analysis on them allows a more comprehensive and robust classification of out-of-time events for further analysis.

A multivariate analysis is presented and discussed in the final chapter, along with the development and implementation of a novel matching algorithm that utilizes all three sub-detectors within ICARUS to accurately assign a time to each track reconstructed in the Time Projection Chamber using the photon detection system and Cosmic Ray Tagger spatial and temporal resolutions.

After being trained on a Monte Carlo sample and tested over Data from the BNB Run2, the multivariate analysis, referred as triple-matching, performs a tagging of TPC reconstructed tracks with over 90 % efficiency and purity, incorporating new information on the events that occur inside the LAr volume that will be useful for future analysis.

Chapter1

Neutrino oscillations

Neutrino oscillations have been measured from a long interval of time since the first experimental hints of the phenomenon appeared in the 60s, reaching a convincing confirmation in 2001 when the SNO experiment measured via a compared analysis between charged current interactions ($\nu_e + d \longrightarrow p + p + e^-$) and neutral current interactions ($\nu_\alpha + d \longrightarrow p + n + \nu_\alpha$) that the amount of electronic-type neutrinos coming from the Sun corresponded to just one third of the expected ones while for the neutral current interactions [1], which are leptonic family independent, the number of interactions remained within the expected results.

The previous observations led then to a modification of the Standard Model, in which neutrinos are massless particles and therefore should not oscillate. In the Standard Model neutrinos are defined as singlets of the SU(3) symmetry, but Weyl doublets of $SU(2)_L$ that maximally violate parity. Without a right-handed component, neutrinos can't couple with the BEH mechanism, which couples as $\Psi_L \Phi \Psi_R$ and therefore can't acquire mass. However, neutrinos do couple weakly and interact with other leptons and quarks according to the

weak current Lagrangian states:

$$\frac{-g}{\sqrt{2}}\sum_{\alpha=e,\mu,\nu}\bar{\nu_\alpha}\not{W}\ell_\alpha - \frac{g}{2\cos\theta_w}\sum_{\alpha=e,\mu,\nu}\bar{\nu_{\alpha L}}\not{Z}\nu_{\alpha L} + h.c.$$

where g corresponds to the coupling constant, θ_W to the Weinberg mixing angle and $\not{W} = \gamma^\mu W_\mu$. The existence of additional active neutrinos is strictly forbidden by the results presented by LEP, in which the invisible width of the decay $Z \to \nu\bar{\nu}$ was compatible with 3 generations[2]

$$N_\nu = \frac{\Gamma_{inv}}{\Gamma_{\nu\bar{\nu}}} \approx 2.984 \tag{1-1}$$

Neutrinos that do not participate in any weak interaction, referred to as sterile, are still allowed, but the Standard Model has to be extended to include them as singlets under every interaction. Neutrino flavours are related by a unitary mixing matrix (Pontecorvo, Maki, Nakagawa and Sakata matrix, or PMNS matrix) as:

$$\nu_{\alpha L} = \sum_{i=1}^{3} U_{\alpha i}\nu_{iL}$$

where $\alpha = e, \mu, and \tau$ $i = 1, 2, 3$ enters the Lagrangian expressed in equation 1 by writing one of the bases in terms of a linear combination of the other. A common form used for the PMNS mixing matrix is:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \times \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta} & 0 & c_{13} \end{pmatrix} \times \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \times \begin{pmatrix} e^{i\alpha_1/2} & 0 & 0 \\ 0 & e^{i\alpha_2/2} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

where c_{ij} and s_{ij} represent $\cos\theta_{ij}$ and $\sin\theta_{ij}$ respectively and δ, α correspond to the Dirac CP violation and Majorana phases. This parametrization is widespread since it allows to isolate in different matrices the neutrino oscillation parameters, the first matrix contains the leading parameters for atmospheric neutrinos, the second for reactor neutrinos, the third for the solar neutrinos, and the last one isolates the Majorana phases,[3].

The oscillation probability can be calculated via a plane wave approximation in which the state ν_α is the initial state

$$|\nu, t=0\rangle = |\nu_\alpha\rangle = \sum_{i=1}^{3} U^*_{\alpha i} |\nu_i\rangle$$

where the $|\nu_i\rangle$ are eigenstates of the free Hamiltonian with eigenvalues $E_i = \sqrt{p^2 + m_i^2}$, so the time evolution of the neutrino state is obtained from the application of the time evolution operator to the initial state:

$$|\nu, t\rangle = \exp(-iHt) |\nu_a\rangle = \sum_{i=1}^{3} U^*_{\alpha i} \exp(-iHt) |\nu_i\rangle$$

from which a probability of transition from a given state ν_α to another ν_β is obtained projecting the state $|\nu, t\rangle$ into the ν_β direction as:

$$P(\nu_\alpha \to \nu_\beta) = |\langle \nu_\beta | \nu, t\rangle|^2 = |\sum_i U_{\beta i} U^*_{\alpha i} \exp(-iE_i t)|^2$$

neutrinos are highly relativistic so it is possible to approximate as:

$$E_i - E_j \approx \frac{m_i^2 - m_j^2}{2p}$$

and similarly approximate $L = t$, resulting in the general formula for the neutrino oscillations

$$P(\nu_\alpha \to \nu_\beta) = |\langle \nu_\beta | \nu, t\rangle|^2 = |\sum_i U_{\beta i} U^*_{\alpha i} \exp(-i\frac{\Delta m_{ij}^2 t}{2E})|^2 \qquad (1\text{-}2)$$

Where $E \approx p$ and $\Delta m_{ij}^2 = m_i^2 - m_j^2$. The case in which $\alpha = \beta$ is usually referred as survival probability or disappearance channel, while the case for which $\alpha \neq \beta$ is the transition probability or appearance channel. in its explicit form, the neutrino oscillation probability can be written as:

$$P(\nu_\alpha \to \nu_\beta) = \sum_{k,j} U^*_{k\alpha} U_{k\beta} U_{j\alpha} U^*_{j\beta} \exp(-i\frac{\Delta m_{kj}^2 L}{2E})$$

Where separating the real and imaginary parts becomes:

$$\Delta_{\alpha\beta} - 4\sum_{i>j} \mathrm{Re}(U^*_{i\alpha}U_{i\beta}U_{j\alpha})U^*_{j\beta}\sin^2\frac{\Delta m^2_{ij}L}{4E} + 2\sum_{i>j}\mathrm{Im}(U^*_{i\alpha}U_{i\beta}U_{j\alpha})U^*_{j\beta}\sin^2\frac{\Delta m^2_{ij}L}{4E}$$

In a simplified scenario with only two flavors involved, the oscillation probability becomes:

$$P(\nu_\alpha \to \nu_\beta) = \sin^2 2\theta \sin^2\frac{\Delta m^2 L}{4E}$$

Which expressing the mass difference of the neutrinos in eV^2 the distance in km and the neutrino energy in GeV can be written as:

$$P(\nu_\alpha \to \nu_\beta) = \sin^2 2\theta \sin^2 1.27\frac{\Delta m^2 L}{4E}$$

From equation 1 it is clear that neutrino oscillations are modulated by the L/E ratio, where L is the traveled distance and E is the energy of the neutrino.

Oscillation experiments are classified with respect to the source they study, being it *artificial* like in reactors or accelerators, or *natural* like solar or atmospheric neutrinos. Depending on the studied channel neutrino experiments can be distinguished into:

- Appearance: measuring the oscillation probability of detecting a given flavor β that originates from a source that emits neutrinos of a different flavour α. $P(\nu_\alpha \to \nu_\beta)$
- Disappearance: measuring the decrease in the flux of the neutrino flavor emitted at the source, $P(\nu_\alpha \to \nu_\alpha) = 1 - \sum_{\alpha\neq\beta} P(\nu_\alpha \to \nu_\beta)$

1.0.1 Solar neutrinos

Solar neutrinos are produced in the core of the Sun principally as products of the *pp* chain and the CNO cycle. Solar neutrinos have a ratio L/E of about $10^{12}\ \frac{m}{MeV}$ that allows mainly the mixing between ν_1 and ν_2, so oscillation terms proportional to Δm^2_{31} are subdominant. Experiments focused on solar neutrinos are able to characterize parameters as $\sin^2\theta_{12}$ and Δm^2_{12}.

Modern experiments are based on Cherenkov detectors and liquid scintillators, being of the most known ones Sudbury Neutrino Observatory (SNO), SuperKamiokande and Borexino. SNO exploited 1 kton of heavy water and used Cherenkov light to measure the charged current ($\nu_e d \rightarrow p + p + e^-$) and the neutral current ($\nu_x d \rightarrow p + n + \nu_x$) interactions, being the first interaction sensitive only to ν_e under the conditions of the experiment, in which the energies were low enough to not expect CC with ν_μ, while the other one was sensitive to all neutrino flavors. The result obtained was that the ratio between CC events and NC events was about 1/3, in correspondence with the 3 family hypothesis, and winning the Nobel prize for these results [4]. SuperK is a 50 kton pure water Cherenkov detector placed below the mountains of Kamioka, Japan [5] and the BOREXINO is a 300 tons organic scintillator situated at the Laboratori Nazionali del Gran Sasso (LNGS) in Italy [6]. Both experiments studied an energy range spanning from hundreds of KeV to some MeV. A region of energy accessible to reactor neutrinos too.

1.0.2 Atmospheric neutrinos

Atmospheric neutrinos are produced as decay products of hadronic showers resulting from collisions of cosmic rays with nuclei in the atmosphere. Pions and Kaons decay mainly as $\pi^+ \rightarrow \mu^+ \nu_\mu$ and $K^+ \rightarrow \mu^+ \nu_\mu$ while electron neutrinos are produced by the muon decay $\mu^+ \rightarrow e^+ \nu_e \bar{\nu}_\mu$; for atmospheric neutrinos the parameters θ_{23} and Δm^2_{23} are dominant.

The ratio between muon to electron neutrinos is known with a precision of around 5% [3]:

$$R = \frac{N_{\nu_\mu} + N_{\bar{\nu}_\mu}}{N_{\nu_e} + N_{\bar{\nu}_e}} \approx 2$$

So the oscillations from atmospheric neutrinos can be identified via the differences between R and the measured value.

Important experiments on atmospheric neutrinos were SuperK and MACRO. [7].

MACRO was located at LNGS, and it was composed of three sub-detectors:

liquid scintillation counters, limited streamer tubes, and nuclear track detectors. MACRO Measured a lower than expected ratio R and confirmed the hypothesis of $\nu_\mu \rightarrow \nu_\tau$, calculated a maximal mixing at $\Delta m^2_{23} = 2.3 \times 10^{-3}$ eV^2, ruling the non oscillation hypothesis [8].

1.0.3 Reactor neutrinos

Nuclear reactors emit $\bar{\nu}_e$ via fission reactions with an energy of the order of few MeV, as in solar experiments, this energy sets a boundary since they do not allow to produce charged leptons heavier than the electron in CC interactions. These experiments used as detection technique the inverse β-decay, measuring the rate of positrons emitted and comparing it to the predicted rate in terms of the reactor power and the emitter neutrinos. The parameter that is dominant in reactor experiments is θ_{13}; several experiments like Daya Bay (China), RENO (South Korea), and Double Chooz measured it [9–11] with a best fit for θ_{13} of around 8.60 ± 0.13 degrees.

Another reactor experiment is KamLAND (Kamioka Liquid scintillator Anti-Neutrino Detector), surrounded by 53 nuclear reactors at an average distance of 180 km. It was able to investigate values of L/E = 10^5m/MeV, typical for solar neutrino experiments.[12]

Two scenarios of hierarchy consistent with experimental measurements are possible: normal and inverted ordering. In the first one neutrino masses are ordered as $m_1 < m_2 << m_3$ while in inverted one as $m_3 << m_1 < m_2$ [13]. Expected to start in 2025, the Jiangmen Underground Neutrino Observatory (JUNO), will try to determine mass hierarchy from the most precise measurements of θ_{13} and its oscillation parameters[14]. The two hierarchy scenarios are presented in figure **1-1**.

1.0.4 Accelerator neutrinos

Neutrino beams are generated by accelerating protons up to the order of tens or hundreds GeV, and directing them into a target where secondary particles of the desired momentum are focused and left to decay, producing neutrinos (mainly ν_μ). These experiments present the advantage of allowing

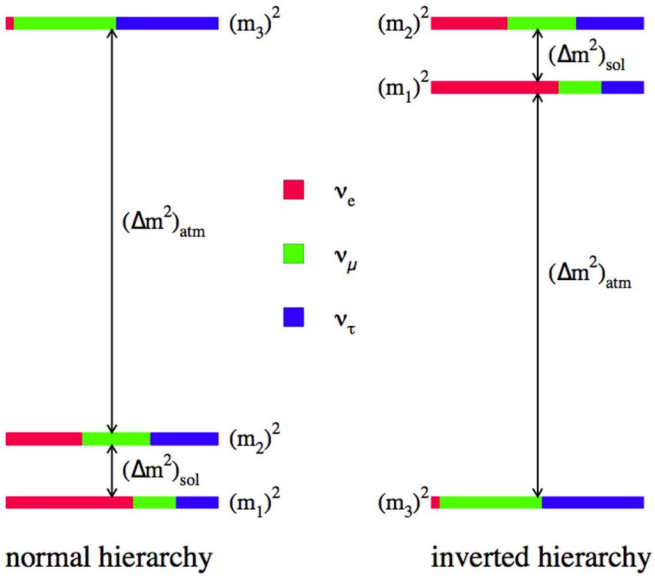

Figure 1-1: Neutrino ordering scenarios for which the flavor "share" of mass eigenstates is also known

the distance parameter to be modifiable, hence accessing different values of the L/E ratio. A diagram of how a flux of accelerator neutrinos is produced is shown in figure **1-2**, showing all the steps from the proton acceleration to the decay of the products.

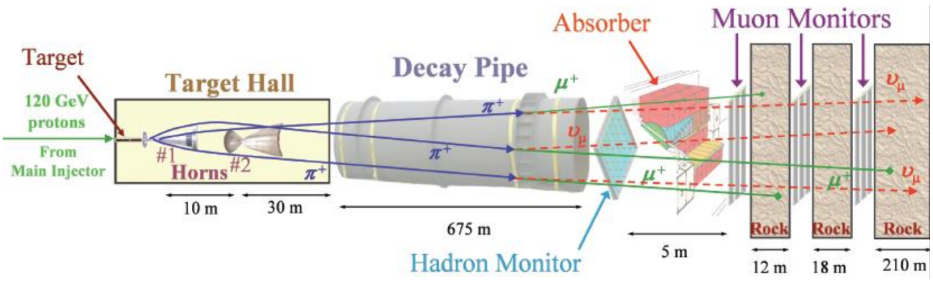

Figure 1-2: A neutrino production line for protons of 120 Gev, in addition to the proton collision and the decay pipe a set of absorbers is present to clean the neutrino flux from other decay products [15].

Long baseline experiments are characterized by a long length L and are de-

signed to measure with precision oscillation parameters as the ν_{12} and ν_{12} mixing angles. These experiments present a characteristic ratio of around

$$\frac{L}{E} \leq 10^3 \text{ km/GeV}$$

which corresponds to a mass difference of $\Delta m^2 \geq 10^{-3}$ eV^2.

Two experiments of this type are K2K and MINOS. K2K (KEK to Kamioka) directed a muon neutrino beam of around 1 GeV from the accelerator facilities of KEK to the Kamioka observatory. The near detector was situated around 300 meters from KEK and Superk was used as the far detector, with a distance of 250 km [16].

MINOS operated on the NuMI beam at the Fermi National Accelerator Laboratory (FNAL) towards the Soudan Mine facilities at 735 km with muon neutrinos of around 2 GeV confirming atmospheric oscillations with great sensitivity, excluding the $\theta_{13} = 0$ hypothesis with a 96 % confidence level. [17].

Other long baseline appearance experiments like OPERA, ICARUS-T600 at the LNGS and T2K from Tokai to Kamioka in Japan are noticeable examples. The first two studied the CERN neutrinos to Gran Sasso (CNGS) beam generated at CERN SPS and directed towards LNGS in a baseline of 732 km with a mean energy of 17 GeV.

OPERA was designed to search for ν_τ using nuclear emulsions with micrometric resolution, reporting the observation of 5 ν_τ candidates for a background of 0.25 events, establishing the appearance in the beam with a significance exceeding 5 σ [18].

Actual experiments like the NuMI Off-axis ν_e Appearance (NOVA), which measures the oscillation rates for both ν_{mu} and $\bar{\nu}_\mu$ with the aim of measuring the cp violating angles present in the PMNS matrix, or the future Deep Underground Neutrino Experiment (DUNE) show that accelerator experiment can be used for multiple purposes.

Short Baseline experiments are characterized by short oscillation lengths with

a characteristic ratio of around:

$$\frac{L}{E} \leq \ 1km/GeV$$

Which means $\Delta m^2 \geq 1\text{eV}^2$ Operating also in both appearance and disappearance channels. The first short baseline experiments were NOMAD and CHORUS at CERN, where both studied $\nu_\mu \rightarrow \nu_\tau$ oscillations, showing no oscillations as a result [19, 20]

1.1 The neutrino mass formalism

The origin of the neutrino masses is still an open question of modern Physics since the Standard Model does not predict them, making explicit then the need for an extension of the theoretical frame.

1.1.1 Dirac Masses

The most direct extension of the standard model is the assumption of the existence of right handed neutrinos in a way that the Higgs mechanism can be extended as

$$-L = y\bar{L}\tilde{H}\nu_R + h.c \tag{1-3}$$

where $L = \nu_L^T, \ell_L^T$ and $\tilde{H} = i\sigma_2 H^\star$, after symmetry breaking of the Higgs field, a mass term is obtained:

$$L_{Dirac} = \frac{v_H}{\sqrt{2}}\bar{\nu_L} y_\nu \nu_R + h.c.$$

where the coupling constant y_ν is estimated to be of the order of 10^{-12} for a neutrino mass of the sub-eV order, additionally as the lepton number is conserved, the conservation from accidental becomes fundamental in the standard model to justify the lack of a Majorana neutrino term[13].

1.1.2 Majorana Masses

Neutrinos are the only known particles for which a correspondence particle-antiparticle is possible, leading the way to a possible Majorana mass term,

for example via a singlet combination:

$$L_{Majorana} = \frac{\lambda}{\Lambda} L^T \cdot \tilde{H}^\star C^\dagger \tilde{H}^\dagger \cdot L + h.c.$$

forming a Weinberg operator, of dimension 5 and cut-off by a mass scale Λ; this operator would break lepton number by two units and generate a mass term:

$$L_{Majorana} = \frac{v_h^2 \lambda}{2\Lambda} \nu_L^T C^\dagger \nu_L + h.c.$$

1.1.3 The see-saw mechanism

The Lagrangian that describes the neutrino masses has terms that include both the Dirac and Majorana expressions, which can be written in terms of the column vector:

$$N_L = \begin{pmatrix} \nu_L \\ \nu_R^c \end{pmatrix} = \begin{pmatrix} \nu_L \\ C\overline{\nu_R}^T \end{pmatrix}$$

the Lagrangian is of the form:

$$L_{mass}^{D+M} = \frac{1}{2} N_L^T C^\dagger M N_L + h.c$$

where the matrix M is written as:

$$M = \begin{pmatrix} m_L & m_D \\ m_D & m_R \end{pmatrix}.$$

where m_D corresponds to the Dirac mass terms generated after symmetry breaking and m_R, m_L to the Majorana mass terms of a left-handed and eventual right-handed Majorana neutrino after the symmetry breaking [3]. The non-diagonal form of the mass matrix shows that flavor eigenstates correspond to linear combinations of the mass states and vice versa, describing neutrino oscillations. This process will be explained in detail in the following section [13].

The neutrino masses can be generated via extensions of the standard model in which a virtual massive particle is exchanged at tree level with a mass of the order Λ. Depending on the exchanged particle the three main models studied are:

- Fermion singlet exchange (Type I see-saw)
- Scalar triplet (Type II see-saw)
- Fermion Triplet (type III see-saw)

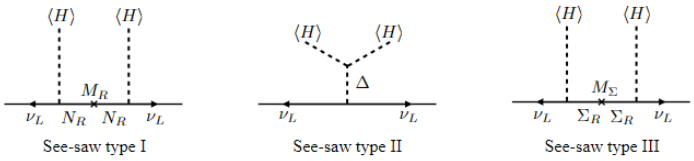

Figure 1-3: Feynman diagrams of the see-saw types for the singlets and triplet couplings.

The order of Λ is usually assumed to be the one of Grand Unified Theories(GUT), of around 10^{14} GeV, providing an explanation for the small neutrino masses and the baryon asymmetry of the universe. Smaller couplings can be directly tested with actual experiments, for example, assuming a Λ of the order of TeV, and therefore observing eventual signatures such as the scalar or fermion triplets proposed by the see-saw or, eventual sterile neutrinos. Further modifications of the Λ scale contemplate scenarios in which the heavy neutrinos could be found in scales of GeV or even MeV, magnitudes in which its signatures would induce peaks in the emission of electrons and muons in meson decays, as the ones searched by experiments as the NA62 [21]. At keV the signature expected corresponds to a kink in the electron spectrum of the beta decay [22], while at eV would introduce neutrino oscillations at short distances. For each one of the cases, the Yukawa coupling value changes from 10^{-12} for the eV mass case to 1 in the GUT one[13].

1.2 Neutrino anomalies in oscillations

. The oscillation probabilities described in equation 1-2 adjust to models in which there are 3 types of neutrinos, however, these oscillation probabilities do not correspond to the ones observed in some experiments. These so-called anomalies have been found in a variety of experiments such as MiniBooNE, LSND, 4-neutrino, and in reactor and gallium experiments.

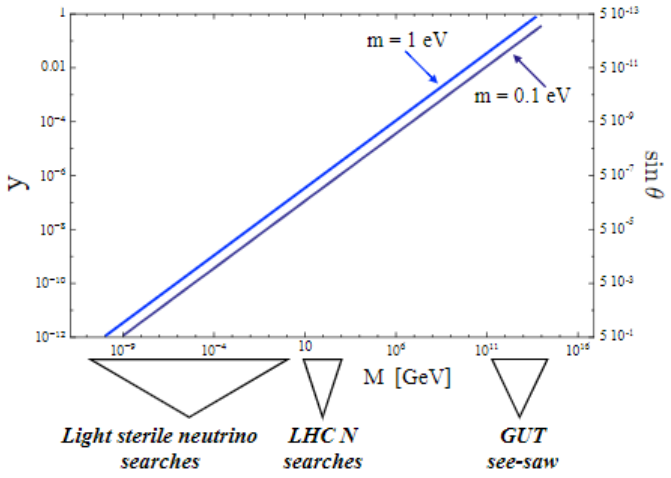

Figure 1-4: Range of the Yukawa coupling and the sterile neutrino mass required to generate a ≈ 0.1 eV light neutrino mass for each scenario [13].

From these anomalies the most common interpretation requires an oscillation frequency of $\Delta m^2 \approx 1\text{eV}^2$, which implies the existence of at least one additional neutrino that oscillates with the other three flavors. However, the interpretation remains incomplete since it doesn't explain all the anomalies observed.

1.2.1 Reactor Anomalies

reactor neutrino experiments reported measurements of $\bar{\nu}_e$ fluxes at short distances (10-100 m), lower than the ones expected by the 3 neutrino mixing theory, where the most precise source modelings on the inverse beta decay at reactors give a discrepancy of around 5-6% between the prediction and the actual measurements [23]. This result can be explained by errors in the flux modeling of uranium in reactors or by eventual sterile neutrinos mixing with active ones; leading to a fraction of the $\bar{\nu}_e$ that oscillates at the $\Delta m^2 \approx 1\text{eV}^2$.

In Gallium experiments the anomaly is present in the ratio of the Germa-

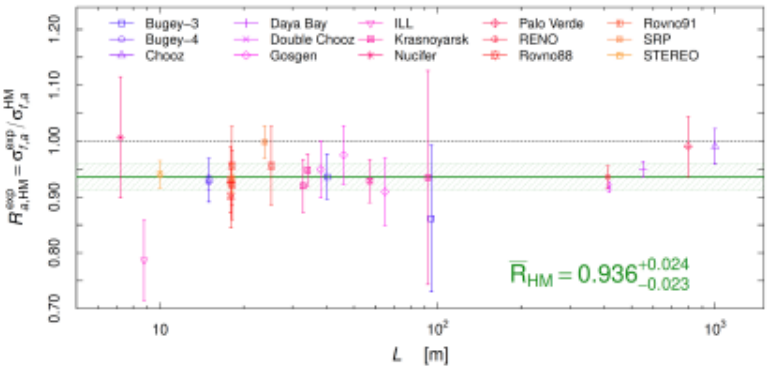

Figure 1-5: Ratio of the measured to predicted inverse beta decay as a function of the baseline. Each data point corresponds to an experiment, the green line and band show the average ratio and uncertainty respectively [23]

nium atoms produced via the inverse beta decay present in experiments like GALLEX and SAGE [24, 25] which use both 71 Ga and measure the reaction:

$$\nu_{\mathrm{e}} + 71\,\mathrm{Ga} \longrightarrow 71\,\mathrm{Ge} + \mathrm{e}^{-}$$

The Gallium anomaly refers then to a ratio of $R_{avg} = 0.86 \pm 0.05$ which is 3σ away from unity.

The Neutrino-4 experiment consisted of a 1.8 m^3 Gd-doped liquid scintillator detector located in Dimitrovgrad that is designed to detect $\bar{\nu_e}$ generated from a 57 MW_{Th} reactor with a distance from the target that can be moved from 6 to 12 m. The experiment performed an eventual sterile neutrino search in the L/E parameter space and observed with a 2.9 σ an oscillation effect with a best that corresponded to $\Delta m^2_{14} = 7.3 \pm 0.13(\mathrm{stat}) \pm 1.16(\mathrm{syst})$ and $\sin\theta_{14}{}^2 = 0.36 \pm 0.12$ [27].

1.2.2 Accelerator anomalies

The LSND (Los Alamos Neutrino Detector) and MiniBooNE experiments measured independently the inverse beta decay produced from a neutrino flux whose original particle was a pion ($\pi^+ \to \mu^+ + \nu_\mu$ and $\mu^+ \to e^+ + \nu_e + \bar{\nu}_\mu$),

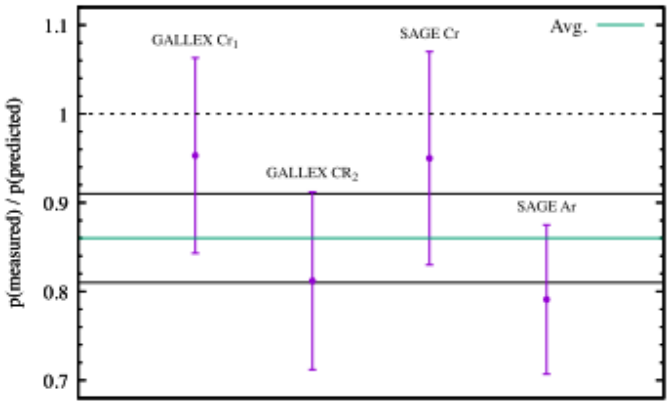

Figure 1-6: Ratio of the observed and predicted events for different sources in the GALLEX and SAGE experiments. The shadowed region corresponds to 1σ around the average [26].

being the π^+ of the first one decay-at-rest while the ones of the second decay-in-flight, offering a ν_μ dominated flux with higher mean energy, which meant a longer baseline for measuring the neutrino oscillations. LSND consisted of a target of 167 tons of mineral oils, doped with liquid scintillators for both Cherenkov and scintillation light, the light readout carried by 1220 photomultiplier tubes distributed on the inner surface of the tank, observing evidence for $\bar{\nu_\mu} \rightarrow \bar{\nu_e}$ oscillations that could be also described by a 2-neutrino oscillation model with $\Delta m^2 \approx 1\text{eV}^2$ and an oscillation amplitude ($\sin 2\theta_{e\mu}$) smaller than 1% [28], and in which the anomaly corresponds to a background-substracted excess of 87.9(stat) $\pm$ 6.0(sys) events respect to the expected number.

MiniBooNE had a muon neutrino beamline developed at the Fermi National Accelerator from protons extracted from the Booster accelerator, providing a mean energy of around 600 MeV [29] with an L/E ratio comparable to LSND on a longer baseline (540 meters). Structurally MiniBooNE consisted of a spherical tank, with 12.2 meters of diameter and 818 tons of mineral oil covered with 1520 PMTs in the inner surface. After its final results, MiniBooNE showed a low energy excess(LEE) of data over background prediction for both the neutrino and antineutrino modes possible for the beam of about

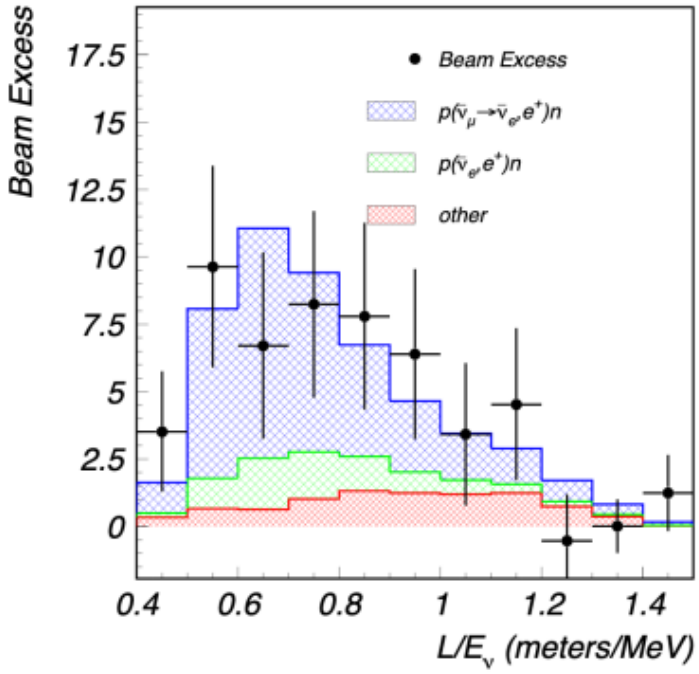

Figure 1-7: LSND anomaly as a function of L/E, the blue shaded region corresponds to best fit values for 2 neutrino oscillations $\sin^2 2\theta = 0.003$ and $\Delta m^2 = 1.2\text{eV}$ [28]

560 ± 119.6 and 77.4 ± 28.5 events for the neutrino mode and the antineutrino mode respectively with a best fit of $\Delta m^2 = 0.043\text{eV}^2$ and $\sin 2\theta^2 = 0.807$, which is an excess events consistent with the LSND best fit parameters but not able to describe the shape of the events excess, motivating further interpretations.

1.3 Sterile Neutrino global fit

All of the anomalies previously mentioned can be explained by the addition of at least one sterile neutrino state with a mass of the order of 1 to 10 eV. In the simplest 3+1 extension, a fourth neutrino state can be added to the PMNS matrix, the newly added neutrino would be sterile as it would

not actively participate in the weak interactions. Assuming that $\Delta m^2_{41} >> |\Delta m^2_{31}|$, Δm^2_{21}, it is possible to approximate Δm^2_{31} and Δm^2_{21} as degenerate and null, so in the limit where they are negligible a short baseline oscillation can be described depending on the channel by the following expressions:

- electronic neutrino disappearance channel

$$P(\nu_e \to \nu_e) \approx 1-4|U_{e4}|^2(1-|U_{e4}|^2)\sin\frac{\Delta m^2_{41}L}{4E_\nu} = 1-\sin 2\theta_{ee}\sin^2(\frac{\Delta m^2_{41}}{4E_\nu})$$

- muon neutrino disappearance channel:

$$P(\nu_\mu \to \nu_\mu) \approx 1-4|U_{\mu 4}|^2(1-|U_{\mu 4}|^2)\sin\frac{\Delta m^2_{41}L}{4E_\nu} = 1-\sin 2\theta_{\mu\mu}\sin^2(\frac{\Delta m^2_{41}}{4E_\nu})$$

- Electron neutrino appearance channel:

$$P(\nu_\mu \to \nu_e) \approx 4|U_{4\mu}||U_{4e}|\sin^2\frac{\Delta m^2_{41}L}{4E_\nu} = \sin 2\theta_{e\mu}{}^2\sin^2(\frac{\Delta m^2_{41}L}{4E_\nu})$$

These oscillations are relevant for $L/E \approx 1$m/MeV or km/GeV if Δm^2_{41} is of the order of 1eV^2. If both $|U_{4e}|^2$ and $|U_{4\mu}|^2$ are nonnull then electron-neutrino and muon-neutrino disappearance and appearance channel must occur at the same L/E. The amplitudes in this case are related by the electron-muon channel $\sin 2\theta_{e\mu}{}^2 \leq \frac{1}{4}\sin 2\theta_{\mu\mu}{}^2\sin 2\theta_{ee}{}^2$. Although anomalies have been observed in the ν_e disappearance and the $\nu_\mu \to \nu_e$ appearance one, there hasn't been any significative anomaly in the $\nu_\mu/\bar{\nu_\mu}$ disappearance channel. Strong limits to the latter channel are provided by several experiments like MINOS and IceCube [30]. These constraints provide a value for $|U_{4\mu}|^2 \leq 10^{-2}$ for a wide range Δm^2_{41} from 2×10^{-1}eV and 10 eV^2. Constraints given on $|U_{e4}|^2$ by reactor experiments require that for obtaining the values of $\sin 2\theta_{e\mu}{}^2 = 4|U_{e4}|^2|U_{4\mu}|^2$ required by LSND and MiniBooNE can only be reached by a large value of $|U_{4\mu}|^2$, which is clearly disfavored by the already mentioned $\nu_\mu/\bar{\nu_\mu}$ null results; being this the origin of a severe tension in the global fits between the appearance and disappearance data.

The Short Baseline Neutrino (SBN) program at Fermilab will lead the effort to provide an answer to the sterile neutrino puzzle, studying by means of

LArTPC detectors both ν_μ disappearance and ν_e appearance channels one the same neutrino beam studied by MiniBoonE.

Chapter2

The Short Baseline Neutrino Program

The Short baseline neutrino program (SBN) at Fermilab will carry out the most sensitive search with the aim to address the possible existence of 1eV mass-scale sterile neutrinos that are suggested by the anomalies; it consists of multiple liquid argon time projection chambers (LArTPCs) [31] located along the Booster Neutrino Beam (BNB) [32].

2.1 The SBN program

The multiple LArTPCs are located at different baselines, searching for high-Δm^2 neutrino oscillations in the appearance and disappearance channel. The Short-Baseline Near Detector has an active mass of 112 t of liquid argon and is located at 100 meters from the neutrino source and is planned to start data taking by the end of 2024. MicroBooNE has an active mass of 89 t of liquid argon and is located 470 meters from the BNB target; it started collecting data in October 2015 and completed its data taking in 2021 after 5 years of activity. The far detector is ICARUS-T600, refurbished and upgraded

from its original design at LNGS for optimal performance in SBN. ICARUS is located 600 m from the BNB source and has an active mass of 476 t of liquid argon and started data collection in June of 2022. The program will cover the parameters allowed by the previously mentioned anomalies with a significance of 5σ. Qualitatively the goals of the program are

- understand the nature of MiniBooNE low energy excess with MicroBooNE (Phase 1) ;
- search for sterile neutrinos both in appearance and disappearance channels, using SBND and ICARUS as near and far detectors respectively (Phase 2);
- provide verification or rejection of the Neutrino-4 experiment evidence for a large mixing angle sterile neutrino;
- study intensively the LArTPC technology and the statistics produced in this for measurements of $\nu - Ar$ cross sections in the few GeV region, both insights useful for the future long-baseline experiment DUNE.

Additionally, SBN will perform also a rich program of BSM searches like: neutrino tridents, Higgs portals for scalar, dark matter and heavy neutral leptons. A diagram that shows the structure of the SBN detectors is shown in figure **2-1**

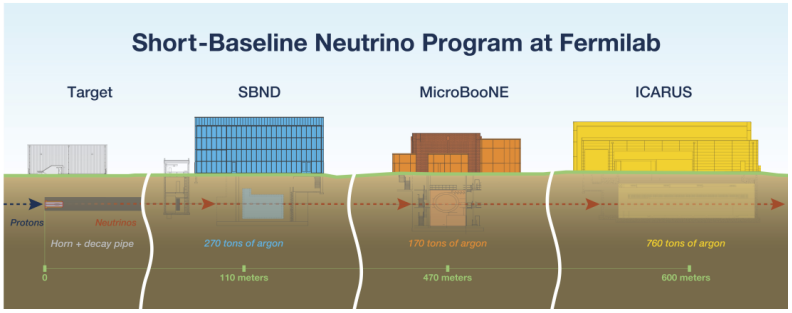

Figure 2-1: The SBN detectors positions along the Booster Neutrino Beam. The Short-Baseline Near Detector is the closest to the neutrino source, MicrooBooNE, the first to complete its operations sits in the middle, ICARUS, the largest of the three, is located at 600 m from the beam target.

The global ν_e appearance favors a mass splitting (Δm^2_{41}) between 0.3 eV^2 and 1.5 eV^2 and a mixing strength in the range $0.002 \leq \sin^2 2\theta_{e\mu} \leq 0.015$, the projected sensitivities (assumed for an exposure of 6.6×10^{20} protons on the BNB target) to the ν_e appearance and disappearance oscillation signals are estimated and shown in the figures **2-2** and **2-3**. BNB will operate until the LBNF (long neutrino beamline facility) long shutdown occurs, expected by early 2027, enabling both ICARUS as SBND to collect more than twice the statistic of the SBN proposal, while ICARUS alone is expected to achieve more than three times the proposed statistic. The ν_e disappearance channel will be studied by both SBND and ICARUS, providing a direct probe of $\sin^2\theta_{ee}$ for a neutrino beam. For an estimated of 6.6×10^{20} protons on target (POT), SBND will collect $\sim$ 35000 ν_e; ICARUS on the other hand can leverage its off-axis position along the NuMI beam (around 6 degrees) and observe a ν_e enriched flux.

2.2 The Booster Neutrino Beam

The Booster Neutrino Beam is generated from 8 GeV protons expected from the Booster accelerator and then fired on a 1.7λ beryllium target, which produces a secondary beam formed mainly by pions [33]; this particles are focused by a single toroidal aluminum alloy focusing horn surrounding the target, supplied with 175 kA in 143 μs pulses in coincidence with the proton delivery. The horn then defocuses pions based on their charge in order to steer the maximal amount of same-charge pions into a collimated beam. These mesons pass through a collimator and are allowed to propagate in the decay tunnel, a 50 meters long, 0.9 radius air-filled volume. In the decay tunnel, the majority of the mesons decay mostly in the channel

$$\pi^{+/-} \rightarrow \mu^{+/-} + \nu_\mu/\bar{\nu}_\mu \tag{2-1}$$

with a branching ratio of 99.99%. At the end of the decay tunnel the residual particles are stopped by a 25 m system of concrete and steel plates. The spill width is of about 1.6 μs with around 5×10^{12} protons delivered per spill to the beryllium target, with an extraction rate of up to 5Hz. The beam is extracted using a fast-rising kicker that steers the particles in a single

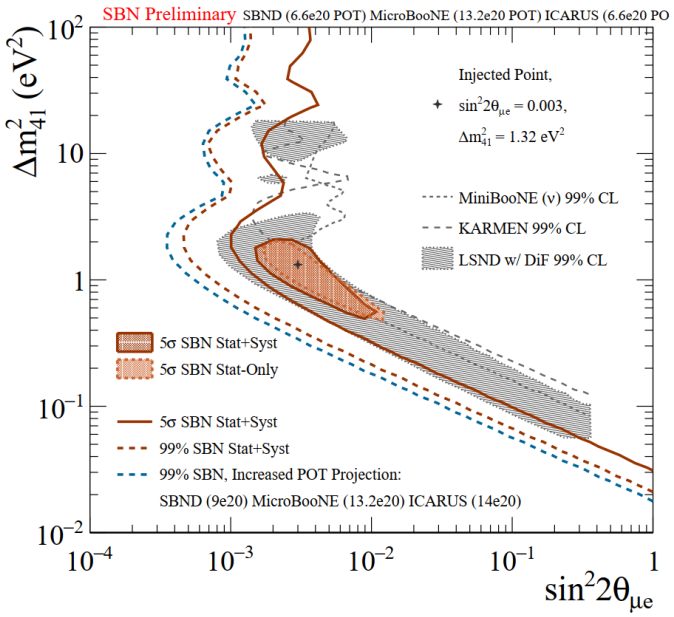

Figure 2-2: SBN 5σ expected sensitivity to a light sterile neutrino in the appearance channel. the LSND preferred region at 99% C.L. (shaded black) is also shown. The SBN sensitivities correspond to an integrated 6.6×10^{20} protons on the BNB target. The sensitivity for the expected POTs is also shown (blue).

turn, giving 81 bunches of protons of about 2 ns wide and 19 ns separation between each one. The Booster neutrino beam can produce positive mesons that produce a ν_μ dominant beam in forward mode, whereas in reverse one produces negative mesons that focused produced $\bar{\nu}_\mu$ dominated beam, and is expected to operate in forward horn current (neutrino mode) for the length of the SBN program.The fluxes of neutrinos observed in SBND and ICARUS peak around 0.7 GeV, consisting almost entirely of ν_μ, with only a 0.5% component of $\nu_e/\overline{\nu_e}$.

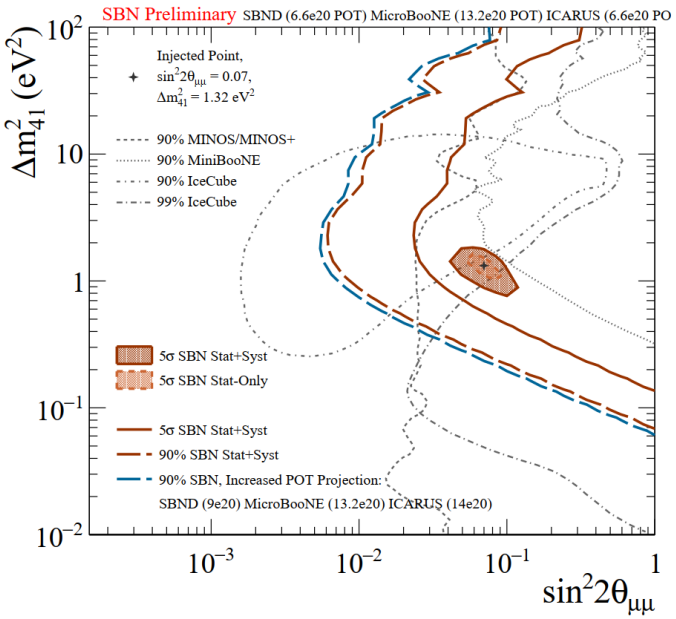

Figure 2-3: SBN 5σ expected sensitivity to a light sterile neutrino in the disappearance channel. The SBN sensitivities correspond to a integrated 6.6×10^{20} protons on the BNB target. The sensitivity for the expected POTs is also shown (blue)

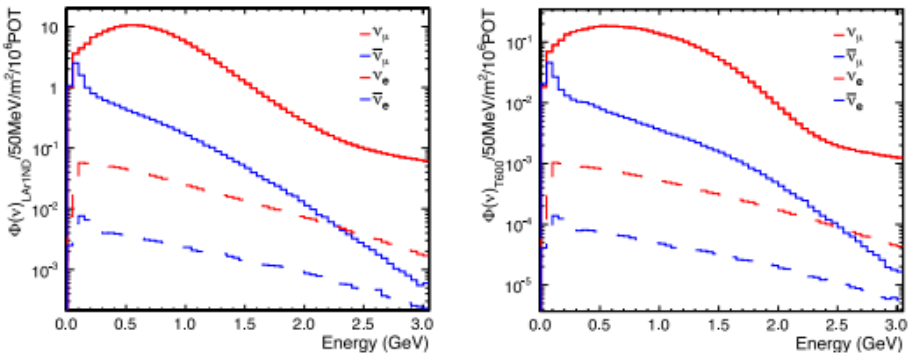

Figure 2-4: $\nu_\mu, \overline{\nu_\mu}, \nu_e$ and $\overline{\nu_e}$ components of the flux fo SBND (left) and ICARUS (right) in forward mode.

2.3 NuMI

The NuMi (Neutrinos at the Main Injector) beam is currently the most powerful beam avaliable, produced by extracting 120 GeV protons from the Main Injector [34]. The initial 8 GeV protons from the booster are steered in the main injector, that is 7 times the circumference of the Booster, allowing the storage of seven Booster batches of which six can be accelerated as the last slot is used for the pulse kicker rise time. The injector accelerates the protons to 120 GeV and the bunches for the NuMI beam line are extracted, bent downward to point at the NuMI target.

The target is made of grafite and the produced hadrons are focused by two magnetic horns into a 675 meters long decay volume. At the end of the tunnerl a hadron monitor is located, in front of a 5 m thick absorber that is followed by 240 m of rocks, interspersed with 4 muon monitors. The horns are pulsed with half a sine wave of 2.3 ms, producing magnetic fields up to 3T with a maximum current of 205 kA with a repetition rate of 1.87s. NuMI delivers up to 6.5×10^{13} protons per spill with a beam pulse of 9.6μs of width.

The ICARUS detector is located 795 m downstream the NuMI beam with an off-axis angle of 6 degrees. This off-axis angle provides a significant flux of both muon and electron neutrinos and antineutrinos in both Forward and Reverse configurations, corresponding to about 3×10^{20} POTs per year.

2.4 The Near Detector: SBND

The Short-Baseline Near Detector has an active LAr mass of about 112 t and is located 110 meters from the BNB target. SBDN is designed to measure the unoscillated BNB neutrino flux, and is composed by two TPCs with a drift region of 2 meters that form a volume of 4 m of height, 4 m of width and 5 meters long. The drift direction is perpendicular to the neutrino beam and the maximal drift length corresponds to about 1.3 ms of the drift time for a nominal voltage of 500 V/cm. The TPC is composed by 2 cathode planes and 4 anode ones (two adjacent planes per TPC). The cathode planes lie side-by-side and are located opposite to the cathode. Each anode plane

is composed of three wire planes angled at $\pm 60^{\circ}$ and 0° with respect to the vertical plane. The planes are 3mm apart with a wire pitch of also 3mm. The diameter of the wires is of 150 μm and are made of copper-beryllium. The SBND TPC is composed of 2816 wires;each of the two jointed cathode planes are formed by metallic subframes supporting a reflective foil coated in TPB wavelenght-shifter. The TPB covered foil maximises the light yield of the detector by reflecting scintillation photons to a Photo-Detection System (PDS) on the opposite wall. SBND front-end electronics for the TPC readout is placed inside the liquid argon, to maximize signal-to-noise ratio.

The PDS setup will test new technologies for future LArTPC experiments, it is divided in 24 modules, 12 behind each anode plane, each containing 5 PMTs (4 coated in TPB and one uncoated). In adition each module contains 4 X-ARAPUCA pairs. The ARAPUCA are devices to record scintillations photons via trapping them within a box with highly reflective internal surfaces [35, 36].

As SBND is located on surface level, on average 3 cosmic rays per neutrino event are expected to cross the active volume in coincidence with the neutrino beam. For this reason a cosmic Ray Tagger is placed surrounding the detector. The CRT is composed of 7 planes, one on each cryostat plus an additional plane above the top one, forming a telescopic tagging system with a resolution of < 2 cm.

The SBDN cryostat was filled in spring 2024 and its currently in its commissioning phase. It will be able to collect more than 5000 neutrino events per day; which means an exposure of $\sim 10^{21}$ POT for a 3 years run means that SBND will generate a dataset of ~ 6 million ν_{μ} and ~ 45000 ν_{e} CC events. A structure of the SBND TPC is shown in figure **2-5**.

2.5 MicroBooNE

MicrooBooNE is a single LArTPC with a total active mass of 89 t of liquid argon at 470 meters from the BNB source [37]. The cryostat has a cylindrical shape and the inner TPC has a volume of 2.3 meters of height, 2.6 meters of

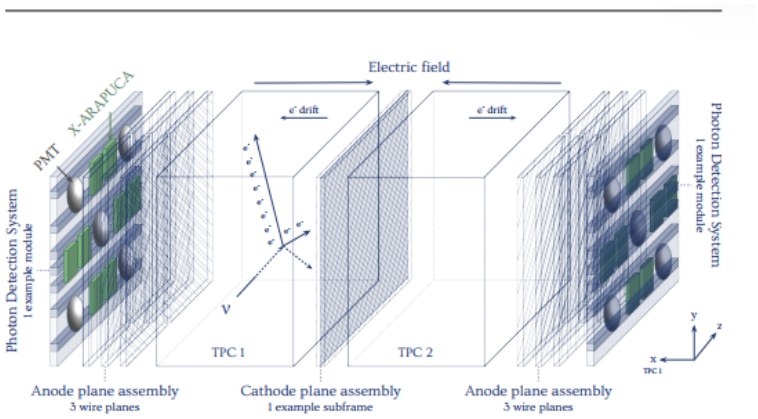

Figure 2-5: Representation of a subvolume of the SBND TPC. 2 PDS modules (1 per TPC) are shown.

width and 10.4 meters long; it is formed by a single cathode and anode planes operating at an electric field of 273 V/cm. The anode plane is composed of wire planes angled at $\pm 60^{\circ}$ and 0° degrees to the vertical plane. The light collection system is formed by 32 optical units consisting of PMTs located behind a wavelength-shifting plate. The secondary system consists of four light guide paddles, included as a $R\&D$. The electron drift time is 2.2 ms, which determines an increased rate of cosmic ray events per neutrino with respect to SBDN or ICARUS. MicrooBooNE is instrumented also with a CRT system composed of 4 planes that provide a coverage to the top, bottom and both long sides of the cryostat.

From late 2015 to the start of 2021 MicroBooNE collected ~ 0.5 million neutrino interactions aiming to clarify the nature of the mentioned MiniBoone low energy excess. The first set of results disfavoured the single-photon background as a sole source of the mentioned excess at 94.8% confidence level [38, 39].

The hypothesis of an energy-dependent enhancement of intrinsic ν_e events at low energy was also tested. An empirical model was derived from the MiniBooNE results and three independent searches were performed on the MicrooBooNE dataset targeting four different final states $1eNp0\pi, 1e0p0\pi, 1e1p, 1eX$.

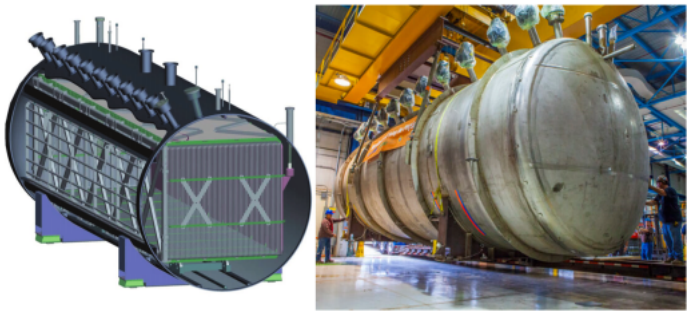

Figure 2-6: Schematic representation of MicrooBooNE along the actual cryostat of the detector.

Observing a number of ν_e candidates that is statistically consistent with the predicted background rates in the low energy excess region; the hyphothesis that ν_e events are fully responsible for the median MiniBooNE excess is rejected at 97% confidence level. Although the data obtained by MicrooBoNE results to be consistent with 3-flavor hypothesis, the existence of sterile neutrinos cannot be ruled out by the electronic low energy excess measured by MicrooBoNE [40]

2.6 The far detector: ICARUS-T600

ICARUS (Imaging cosmic and rare underground signals), with an active mass of 476 tons of liquid argon is the first large-scale operating LAr-TPC detector [41, 42]. It Consists of two adjacent modules of $3.6 \times 3.9 \times 19,9$ m^3 filled with a total mass of 760 tons of liquid argon, that was purified by removing electronegative impurities.

Each module is composed of two LAr-TPCs separated by a common cathode. Anode and cathode planes have a maximum drift distance of 1.5 m, corresponding to $\sim$ 0.96 ms drift time at a nominal 500 V/cm electric drift field. The anode is composed of three parallel wire planes 3 mm apart and oriented at different angles with respect to the horizontal: 0 degrees in induction 1, +60 degrees in induction 2 and -60 degrees in the collection one.

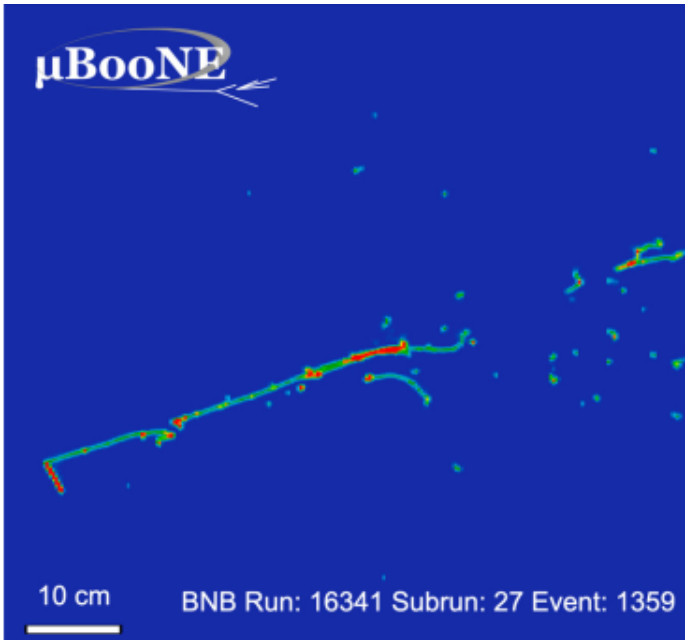

Figure 2-7: ν_e event in MicroBooNE from the event display

In total 53,248 wires with a 3 mm pitch and a maximum length of 9 m are installed in the detector. The anode plane is set such that induction 1 and induction 2 planes are set at a different bias voltages and provide a non destructive charge measurement, and the ionization charge is fully collected by the collection plane. The optical system is composed of an array of PMTs located behind the anodic wire planes that collect the scintillation light used to generate the global event trigger.

The LAr-TPC was first tested at the LNGS and then improved at the CERN, where it was refurnished with: the implementation of new cold vessels, a new cryogenic and LAr purification equipment, the flatting of the TPC cathode, implementation of new electronics and an upgrade of the LAr light detection system. During the overhaul, new electronics for the wire-channels were designed to be compatible with the higher data rates expected at FNAL. The new electronics consists of the same modularity and architecture, but the analogue and digital parts are integrated onto the same custom board mounted directly onto the flange.

The ICARUS detector arrived to Fermilab in July of 2017 and was installed in the SBN far detector building in August of 2018.

2.6.1 The Cosmic Background

The detector was originally designed to operate in the low muon cosmic background of the Gran Sasso laboratory, the shallow depth of the far detector at FNAL determines a large background of cosmic rays, with an expected rate of about $\sim 17\ kHz$ [43]. On average ~ 17 cosmic muons are expected to cross the volume during the 0.96 ms of the drift time. The cosmic rays fluxes were also simulated using the CORSIKA sofrtware, predicting a total flux for $\mu^{\pm}$ above 50 MeV of 129 Hz/m^2.

In order to mitigate the induced background, ICARUS is surrounded with an external Cosmic Ray Tagger system (CRT) below a 3 m concrete overburden (equivalent to 6 meters of water). The overburden is expected to reduce the dominant muon flux by ~ 20 %, stopping the muons with a $E_k \leq 1.5\ GeV$. The suppression is more effective for hadrons, with a reduction factor of $\sim$ 200 primary neutrons and ~ 500 for the protons, primary γ are almost fully suppressed.

The most important sources of background to the ν_e appearance the one due to electromagnetic showers induced by γ produced by the cosmic particles that cross the overburden and propagate through the detector and surrounding materials. A cosmogenic photon can mimic a genuine $1e0p$ ν_e CC interaction. This background can be produced or by all cosmogenic events in coincidence with the beam spill or by interaction occurring at anytime within the acquisition time, which corresponds to the already mentioned maximum electron drift time. The potential cosmogenic background can be also categorized on the basis of the presence "Topology I" or absence, "Topology II" of the parent particle in the TPC active volume. A diagram that shows the rates of cosmic particles is shown in figure **2-8**, showing that with the expection of muons, most cosmic particles are contained.

In order to obtain the proposed sensitivities to the $\nu_\mu \rightarrow \nu_e$ appearance channel, a 95% cosmic ray track identification is required, for which an efficient

Particle	without OB [Hz]	with OB [Hz]	without/with OB
$\mu^{\pm}$	17117	12761	1.34
p	54	0.10	> 500
γ	116	0.03	> 3500
n	1426	6.8	> 200

Figure 2-8: Rate (Hz) of secondary cosmic particles at ground level with $E > 50$ MeV entering the LAr active volume with and without the overburden [44].

cosmic ray tagging system is needed along a light collection system and a precise beam-related trigger system.

The overburden deployed on the top of the ICARUS CRT started in April of 2022 after the completion of the top CRT commissioning. The concrete blocks are installed 10 cm above the top CRT horizontal modules. The second and third layers of the overburden correspond to concrete blocks repurposed from previous Fermilab experiments, tested with a Geiger counter to verify their low radioactivity. The completion of the concrete layer set the start of the ICARUS Run 1.

2.7 ICARUS Data taking

The Run 1 officialy began on June 9 of 2022 and lasted until the beam summer shutdown of July 10 of 2022. The data acquisition was largely successful, with uninterrupted physics runs of ~ 1 day and extended up to 5 days. During Run 1 the experiment collected $\sim 6.8 \times 10^{19}$ POTs and $\sim 4.1 \times 10^{19}$ POTs for NuMI and BNB respectively. ~ 6000 ν_μ CC quasi-elastic events are expected from the data sample; events that will be used for the Neutrino-4 physics analysis, since an oscillatory pattern as the one observed in the mentioned experiment [27] should be directly observable in the ν_μ/ν_e survival probability as a function of L/E in ICARUS, averaged over the 50 meters of the pion decay tunnel in a way similar to the one in figure **2-9**.

After some detector improvements and calibration with cosmic rays, along a

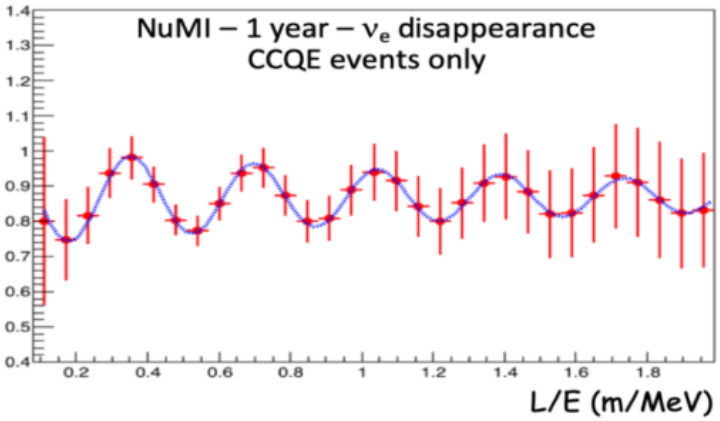

Figure 2-9: Predicted survival ν_e oscillation probability for the neutrino-4 anomaly(blue) for the best fit ($\Delta m^2_{N4} = 7.25\text{eV}^2$ and $\sin^2 2\theta = 0.26$) and the expected signature observed at ICARUS with 1 year of NuMI data (red)

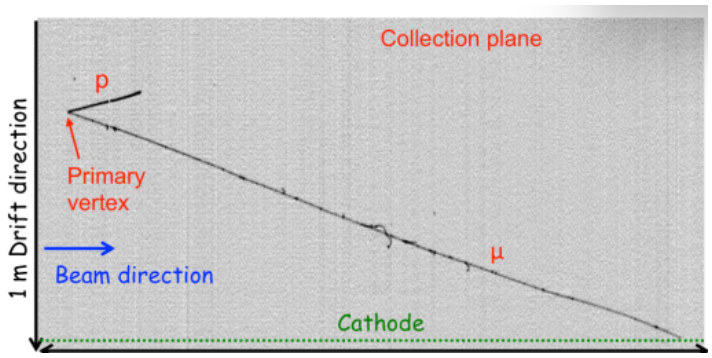

Figure 2-10: Visually selected ν_μ CC candidate from the BNB beam

LAr filter regeneration in the west cryostat, the Icarus Run 2 officially started on December 20 of 2022 and was completed in july 2022. The combined statistic of Run 1 and Run2 is $\sim 3.4 \times 10^{20}$ POTs for NuMI and $\sim 2.5 \times 10^{20}$ POTs for BNB. Corresponding to $\sim$ 240k/308k neutrino interactions for BNB/NuMI [45]. Starting from spring 2024 ICARUS began Run 3 data collection.

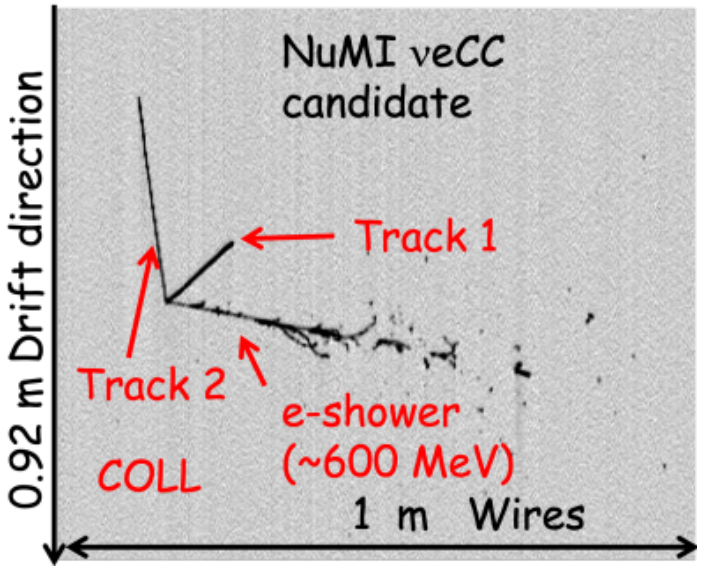

Figure 2-11: Visually selected ν_e CC candidate from the NuMI beam

Chapter3

The ICARUS detector

The ICARUS detector consists of three subsystems: a cosmic ray tagger (CRT), a time projection chamber (TPC) and a photomultiplier system (PMT) that collects the scintillation light from the interactions in the chamber. In this chapter the tree subsystems are described in more detail.

3.1 The liquid Argon time projection chamber

The time projection chamber is an evolution of the gaseous time projection chamber, in which a volume filled with noble gases that formed tracks due to ionization each time a charged particle passes through them. In LARTPC the volume is filled with liquid argon, in which the electrons produced by ionization of the material drift towards the anode plane, instrumented with electronic read-out systems that reconstruct the 3-dimensional coordinates of the event from the 2-dimensional charge collection and the drift time [46].

ICARUS consists of two identical adjacent modules with dimensions of $3.6 \times 3.9 \times 19.6$ m^3 filled with 760 tons of ultra purified liquid argon (continuously purified to prevent absorption of ionization electrons by electronegative im-

purities). Inside each module there are two TPCs separated by a central, vertical, common cathode, in which there is an uniform electric field of $\sim$ 500 V/cm going from the cathode to anode, with a maximum drift time of about 1 ms. Each TPC has 3 parallel read-out planes, 3 mm apart from each other: induction 1, induction 2 and collection. The first plane (induction 1) has horizontal parallel wires, while the other two are oriented $\pm 60^{\circ}$ with respect to it. Each of the TPCs consists of 13332 wires; the reconstruction of the image of the charged particle traversing the volume is obtained by combining the coordinates in the wires of each of the planes. In addition, a measurement of the time of the ionizing event is obtained via the array of photomultiplier tubes along the wall of the TPC [47]. An sketch of ICARUS is shown in figure **3-1** .

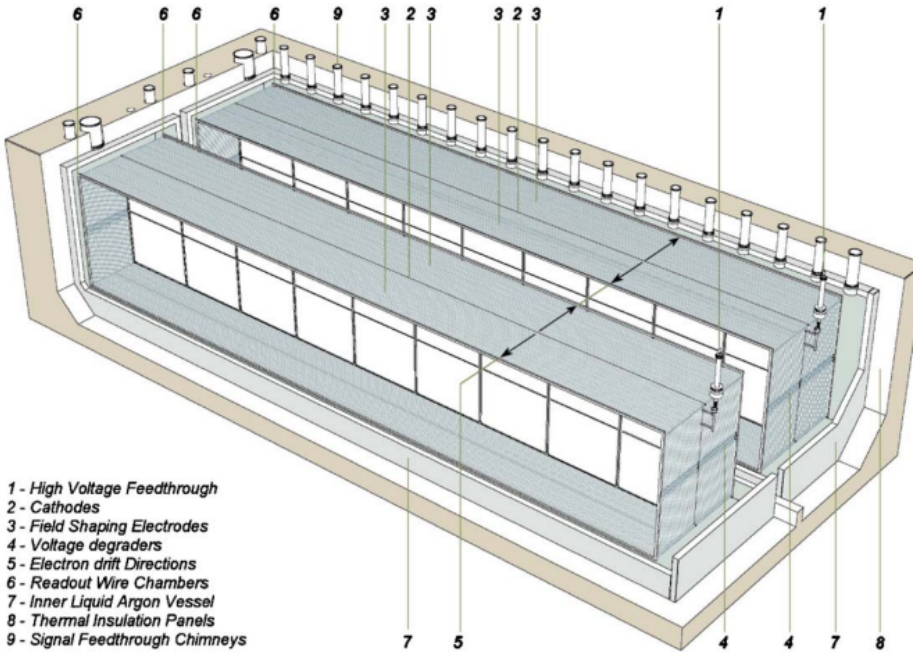

Figure 3-1: 3D diagram of ICARUS [48].

The electronics of the detector initially consisted of a readout chain organized in a 64 channel modality in which a decoupling and biasing board (DBB) passes the signals from the TPC to an analog board via decoupling

capacitors. The analog board performs a serial 12 bit ADC digitization with a 400 ns sampling per channel;with a serial bus architecture allowing Gb/s data transmission and an overall gain of 1000 electrons per ADC count [49]. Afer the overhaul at FNAl the digitazion

The signal from the wires is extracted from each cryostrat via ultra-high vacuum feed-through flanges on top of 96 chimmeys and then via twisted pair flat cables, grouped in sets of 18 cables, serving 576 channels. The digital boards also use custom programmable chips designed specifically for the ICARUS experiment called DAEDALUS, which implement the hit-finding algorithm; each board receives the multiplexed digital data via an external serial link cable.

Once the hit is found, its position, height and area are calculated, and following this process, neighboring hits within each wire plane are grouped together to form two-dimensional clusters. Finally, the 3D reconstruction of each hit is computed exploiting the fact that each wire plane contains two spatial coordinates, one common to all wires (the drift coordinate) and one specific to each wire (the wire coordinate). However, this approach has the disadvantage of being inefficient for reconstructing tracks that move parallel to the wire plane, where the small variation in the drift increases ambiguities in the individual hit assignment by plane [50]. Finally the total deposited energy of a shower or track is obtained by summing all of the deposited energies on each hit [51].

3.2 Reconstruction in the TPC

A hit represents a cluster of electrical charges arriving on a wire at a given time, the algorithm works by finding distributions above a certain threshold for deconvoluted gaussian shapes. In general the TPC electronic modules loop over the input ROIs from the waveforms looking for candidate pulses and fitting them into a gaussian shape; if the hit shape is not a single gaussian but a multiple charge distribution the pulse is divided into a certain number of hits. Assuming that the signal can be described as an overlap of multiple gaussian distributions, the number of distributions used to fit a given signal

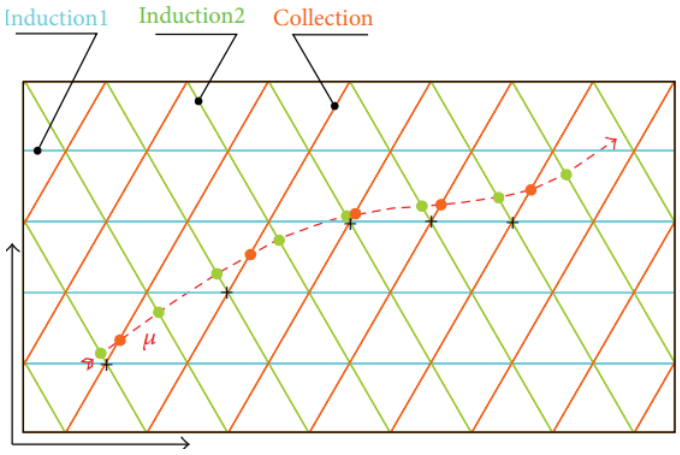

Figure 3-2: *ZY* projection of the TPC with marked intersections on the readout wires and the reconstructed track that can be obtained by associating wire signals from induction 2 and collection using drift time(black dot) [50].

is defined as multiplicity. The properties of the hits are the parameters of the gaussian fit: the area represents the total charge, the mean represents the peak time of the hit, the amplitude the height of the hit and the width the RMS of the hit. This method maximizes efficiency at the expense of purity as it assumes that further ambiguities will be resolved in the subsequent reconstruction step (pattern recognition).

The candidate 3D hits are found by combining the 2D hits on neighboring planes with compatible hits. Compatible hits are those that coincide in time, i.e. the difference between their peaks is within a predefined interval of 0.4 μs. In addition, their wires must intersect in the Y-Z wire plane coordinates projection.

Taking advantage of the fact that the drift coordinate for correlated hits is common to all three views, the third plane is searched by means of a compatible hit with the pre-projected 3D position in each plane. If a compatible hit is found, a 3D space point is formed and stored; finally, the sets of hits produced by the reconstruction algorithm on each wire are the input to the pattern recognition algorithm that reconstructs the tracks in the detector.

The PANDORA software development kit [52] is the pattern recognition tool used in the ICARUS experiment, as it is specifically designed to identify energy deposits from individual particles in fine granularity detectors, using more than 100 clustering algorithms. Each algorithm addresses a specific aspect of event reconstruction and together they provide a robust pattern recognition with the goal of reconstructing each particle as a single object that is both pure (containing only hits from the particle) and complete (containing all hits from the particle). PANDORA has been successfully tested in the reconstruction of cosmic ray muons and neutrino interactions in experiments such as MicrooBoonE and ProtoDUNE-SP [53] [54].

The main output of PANDORA is a list of 3D reconstructed particles, called particle flow particle (PFPs) in PANDORA nomenclature. Each PFP corresponds to a particular track or shower in the event, for which there are associated objects such as collections of 2D hits for each view (clusters), reconstructed 3D positions (spacepoints) and a vertex position defining its point of interaction or first energy deposit. The PFPs are defined in a hierarchical way that identifies particles in a parent-daughter relationship and describes the flow in the interaction. In addition, PANDORA does not reconstruct the identity of each particle, but rather identifies them as track-like or shower-like based on topological features. Track and shower objects contain additional features such as position and momentum values for tracks and principal axis information for showers.

3.2.1 Pandora reconstruction algorithms

PANDORA has two main chain algorithms for event reconstruction in neutrino detectors, PANDORACosmic and PANDORANu, targeting the reconstruction of interactions due to cosmic rays and neutrinos respectively. PANDORACosmic is strongly track-oriented, producing primary particles representing cosmic muons, while PANDORANu focuses on the identification of a neutrino event vertex, which becomes the fulcrum for the reconstruction of all particles originating from it. Vertex identification algorithms are specific to PANDORANu and there's a more sophisticated treatment of tracks and shower-like particles. To optimize the pattern recognition a consolidated

reconstruction approach is often adopted; this starts by running PANDORA-Cosmic on the entire collection of identified hits during the readout window.

PANDORACosmic performs a two dimensional hits clustering for each readout plane, producing a list of clusters corresponding to continuous and unambiguous lines of hits. Separate clusters are created every time a bifurcation or an ambiguity is found, providing high purity clusters on which a cluster merging algorithm is applied to identify possible associations to improve the completeness of the clusters without compromising their purity. The main approach to merging is based on the proximity of the clusters or pointing clusters. Hit selection is further improved by splitting individual clusters into two parts if a topological test indicates that they may contain hits from multiple particles.

Once the 2D clusters are defined and refined PANDORACosmic performs a 3D reconstruction based on these clusters by identifying consistent 2D cluster groups from the three readout planes that describe a single track-like particle. The suitability of the cluster combinations is calculated by another algorithm which stores the score value for further analysis. For each combination of clusters, a number of sample points are defined in the common drifted region, exploiting the redundancy provided by the three independent views. From there, given a sample drifted coordinate (x), the position for a pair of clusters (two different positions on two different wire views) can be extracted, and from the previous position the next cluster position is predicted and compared with the actual position of the reconstructed cluster. By considering all combinations of wire planes, a χ^2 quantity that provides information about the consistency of the new 3D clusters is computed, the connection between multiple clusters and their matching ambiguities. If an ambiguity occurs (a given 3D cluster contains more than one cluster for a single view), another set of specific algorithms is applied to resolve the problem.

When the 3D reconstruction is complete, the 2D clusters that are not associated with a reconstructed object are dissolved, as they are assumed to be fragments of delta ray showers used to seed and grow the shower particles reconstruction. The final task of the process performed by PANDORACos-

mic is to assign an initial position for the reconstructed cosmic muon at the highest vertical coordinate, as they are always assumed to be pointing downwards. Secondary shower-like particles are linked with the parent track by a hierarchical parent-daughter dependency representing Michel electrons and delta rays.

Since the reconstruction of the drift coordinate of the track depends only on the time at which the signal hit the wires, an accurate and robust time designation for each reconstructed object within the TPC is required to accurately locate the interaction within the drift volume. The time assigned to the collection of the charge by the wire (t_m) is defined as a function of the time at which the particle enters the detector (t_0) with respect to the trigger time, while the distance on the drift coordinate is calculated as $t_m = t_0 + x/\mathrm{v}_{\mathrm{drift}}$. It is possible to see that if the value of t_0 is not known exactly is not possible to precisely correlate t_m and x.

The initial time can be accurately measured if it is possible to assign an optical flash to a given reconstructed track, however, the high number of cosmic tracks crossing the drift volume during the readout window makes this difficult at the reconstruction level; therefore, interactions that occur in time are assigned $t_0 = 0$ and no ambiguity is present, while in all other cases the time assigned to the interaction by default corresponds to the trigger time, assigning a preliminary $t_0 = 0$, meaning that only the interactions that occur in time have an accurate position reconstruction in the drift coordinate. Particles arriving out of time present an ambiguity that makes it difficult to distinguish whether the particle arrived before the trigger time but far from the anode, or after the trigger time in a position closer to the anode (t_m would be the same in both cases).

For out-of-time tracks that cross the cathode common to the two drift volumes, it is possible to measure t_0 correctly using light-independent methods; since in these cases the reconstruction will generally produce two separate segments characterized by the same displacement with respect to the cathode, but shifted in opposite directions as the drift field directions alternate between adjacent TPCs. The displacement is directly proportional to the real time t_0 of the track and inversely proportional the drift velocity; while

the direction of the shift depends on whether the interaction occurred before or after the trigger time.

Geometric properties are exploited by PANDORA to search 3D clusters reconstructed in two different drift volumes that are consistent with a single continuous trajectory in both position and direction across the cathode boundary. The stitching of the separate components occurs by shifting the two clusters to the x coordinate of the cathode plane, and the time corrections done by the process allow to estimate the track time as:

$$t_0 = t_{max} - (L - \Delta x)/\mathrm{v_{drift}}$$

Where t_{max} is the time of the closest hit to the cathode, L is the maximum drift length and Δx is the distance between the cathode and the reconstructed x position at t_{max} before any correction. For ICARUS $E = 500\mathrm{V/cm}, v_{drift} = 1.6\frac{\mathrm{mm}}{\mu\mathrm{s}}$ and L= 1.5 m, which means that a track would be completely visible if, for a total of 4096 tics (each one of around 0.4 μs) of which $\sim$ 850 occur within the trigger time, the first hit is recorded at earliest at $\sim$ 340 μs before the trigger time and the last one no later than $\sim$ 1300 μs after the trigger time.

3.2.2 Classification of the tracks

Once the stitching algorithm is complete, PANDORA classifies a track as clear cosmic if:

- the reconstructed particle crosses the upper and lower boundaries of the detector
- a hit in the reconstructed particle (assuming $t_0 = 0$) appears outside the physical drift volume, indicating that the object was not associated with the beam spill gate.
- the time correction estimated by the stitching is too large to be compatible with a particle coming from the beam

The tagging of cosmic muons is the first output of the reconstruction performed by PANDORA. The sample of cosmic rays, in particular those whose

t_0 is assigned, forms a critical component of the detector calibration studies [55].

Once the *clear cosmics* have been removed from the initial hit collection, the new array of clusters serves as input to the PANDORANu chain algorithm. However, as there are cosmic ray remnants that have passed the selection, PANDORANu runs the previously described algorithms up to the 3D point creation, where the hits are then divided into *slices*. *slices* are separate lists of hits that are created on the basis of proximity and direction, each one intended to isolate a neutrino interaction and cosmic remnants in different *slices*. Each *slice* is individually processed by neutrino reconstruction algorithms and results in a neutrino candidate [52].

3.2.3 PANDORA reconstruction

The first step in the PANDORANu reconstruction is to take the track-oriented clustered hits and apply more complex topologies with the aim of describing more detailed interactions and the possible number of reinteractions. The list of 2D hits used to generate a list of possible 3D vertex candidates, and once the candidate vertex positions are identified, one is selected as the most likely neutrino interaction vertex after passing a quality cut that evaluates how much the vertex is close to the hit in all plane views; a score is assigned to each candidate and the candidate with the highest score is selected. The score algorithm produces three inputs: the first reflects the fact that particles produced in the interaction should point back towards the vertex of the true interaction, thus penalizing candidates for which the sum of the transverse energy over the whole cluster is not compensated. The second consists of an asymmetry check, where vertexes with a strong symmetry are penalized, since a true interaction point is expected to have a strong imbalance between downstream and upstream hits. The third factor corresponds to a check of the beam coordinate position; where candidates with a smaller z coordinate are privileged due to their proximity to the origin of the beam direction in ICARUS [55]. When the candidate with the highest highest score is selected, any 2D cluster crossing the vertex is split in two, creating a new cluster on each side of the interaction vertex.

Once the clusters have been generated the next step is the 3D track reconstruction, which proceeds almost exactly as in PANDORACosmic, with the difference that PANDORANu also reconstructs primary electromagnetic showers of electrons and protons by adding branches to any long cluster that might represent the central structure of a shower. Once the 3D shower has been reconstructed a second 3D reconstruction for tracks is applied to recover some of the inefficiencies of the previous ones. The final output of the process is a Boosted Decision Tree (BDT) that classifies the each particle as shower-like or track-like. Subsequent steps correspond to refinement and particle hierarchy reconstruction, the latter organizing the particle into a parent-daughter structure.

The final structure of the process starts from the 3D neutrino vertex and proceeds by adding leaves and branches to it. All the primary particles are branches directly related to the neutrino particle, while the leaves represent all of the particles produced by its primary parent;either new interactions or decays. The interaction grows as much as necessary until all of the particles of the *slice* are correctly associated, the final output of the reconstruction chain being a single neutrino particle for each *slice* with internal particle hierarchy, providing a flow of the neutrino interaction.

3.3 The photomultiplier system

Ionization in the LAr TPC is accompanied by a scintillation light emission, which is useful for a possible absolute time measurement of an interaction event. And for triggering the even acquisition, for this reason a large set of PMTs is also inmersed in the liquid argon, so UV photons from the scintillation provide a signal that allows a measurement of the drift time, and hence of the distance traveled by the drifted electrons. The scintillation light emission of the LAr as an active material has been extensively studied, experiencing a narrow intrinsic VUV light around 128 nm in transitions from the lowest excited molecular state to the ground one. The emitted light is characterized by a fast $\tau \sim 6$ ns and a slow $\tau \sim 1.6$ μs decay component; being their relative intensity given by dE/dx, ranging from roughly 1:3 for

minimun ionizing particles to 3:1 for alpha particles. The *PMT* system of ICARUS is coated with aa TPB compound that acts as a wavelength shifter from the above mentioned region to the *PMT* sensitive one.

The photomultipliers selected for each TPC were 90 Hamamatsu R5912-MOD of 8 diameter hemispheral glass mounted behind the wire chambers and adapted to operate at cryogenic temperatures in each TPC. These photomultipliers were chosen because of their good sensitivity to ionizing interactions inside the LAr down to an energy deposition of 100 MeV, a dynamic of scintillation light that allows the detection of fast scintillation light pulses and, at the same time, of single photons arriving from the slow de-excitation; a light collection system that is able to provide unambiguously the absolute time of each interaction and the absence of bursts, sparkling, lightening or other noise generating pulse effects above the single photon level [56].

The electronic readout system of the PMTs is designed to allow continuous readout, digitization and an independent waveform recording of the signals coming from each one of the *PMT*s of the light detection system. This operation is performed by 24 CAEN V1730B digitizers, from which a read-out based on discrimination and coincidence units has been designed to generate a *PMT* pulse that serves the detector trigger logic; this corresponds to at least 9 *PMT* pairs fired. Each channel is equipped with an amplifier that saves over 10μs of signal followed by a 14-bit ADC that samples the signal every 2 ns. [49].

When a board receives an external trigger request, the active buffers of the acquisition system are frozen and write operations are moved to the next available buffers. The stored data are available for download via optical links, and the reconstruction of the scintillation light associated with the specific event is based on the recorded waveforms sampled at 500 MHz. Events coincident with the beam spill receive all the digitized signals from all 360 *PMT*s signals recorded in the 28 μs of the beam window,while events out of the beam time within $\pm$ 1 ms that satisfy the trigger logic are recorded in a 10 μs window that considers only the cryostat where the activity took place [55].

An optical hit is defined as a signal above a pedestal baseline that defines the threshold, where the pedestal is a rolling average of the waveform, computed around a window of 40 clock ticks from the start of the waveform and 20 before the end of it, over consecutive 2 ns samples (where 1 tick is equal to 2 ns). The pedestal for the i-th tick is then calculated as the average between all the values between the (i-20)th and the (i+19)th tick. In addition to the averaged pedestal the RMS is also calculated for each sample, giving an estimate on how significant the signal is; this to avoid biases due to noisy regions or regions where signal is present but not possible to evaluate the pedestal.

Once a hit has been defined, a finding algorithm is applied to the waveform to determine the extent of it and determine its parameters. Three thresholds define the start, tail and end points of a hit, each specified in units of RMS or ADC counts, using the highest value of the two each time. The 3 optical thresholds are defined as follows

- 10 ADC or 3 x local RMS for the hit start,
- 6 ADC or 2 x local RMS for the hit tail,
- 2 ADC or 1 x local RMS for the hit end.

Each optical hit is characterized by the start time, the width of the hit, the maximal amplitude, the integral signal above the baseline and the collected PMT information. In addition, a hit time with relative to the trigger is stored since it is essential to associate light activity with trigger events.[55]

Once the optical hits have been reconstructed, they are grouped together into a parent object defined as an optical flash. In an optical flash, the time scale is sampled in 10 ns bins and each hit is placed in a bin that corresponds to its peak time. Time bins with less than 5 optical hits or less than 100 photoelectrons are discarded; and each time bin is considered as a possible candidate for a flash in terms of the number of photoelectrons they have. Finally an optical flash produced by another interaction can be identified if it occurs after 1 μs from the previous one. The temporal resolution of the photomultiplier system is subject to the one of the optical hit definition

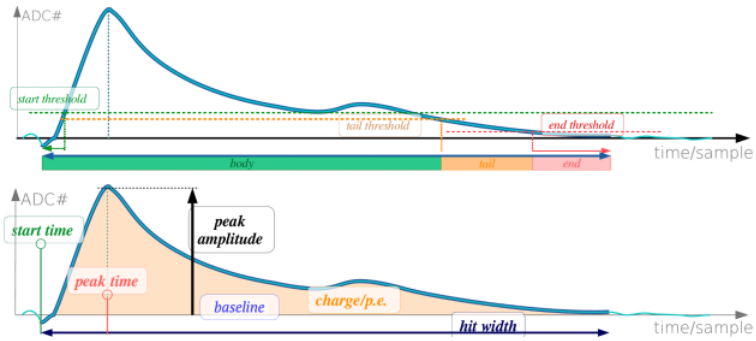

Figure 3-3: Waveform illustration of a *PMT*,on the top image the regions of the reconstruction algorithm are shown and in the bottom one the definition of the variables contained in the optical hit object. [55]

and an additional smearing due to the optical flash reconstruction algorithm; being for the ICARUS case below of 1 ns.

3.4 The Cosmic Ray Tagger

The CRT system is a subdetector external to the cryostats, designed identify charged particles coming from outside the beam and passing through or near the active volume of the TPC. It covers an area of $\sim$ 1100 m^2 and its divided into tree subsystems: top, side and bottom CRT. Each covering different regions of the TPC. All these subsystems have a time resolution of the order of ns, allowing a filtering of events to rejec those where the primary event trigger was tagged in the subsystem; The CRT is designed to tag $\sim$ 95% of the cosmic particles entering the TPCs.

3.4.1 The Bottom CRT

The bottom CRT consists of 14 modules divided into two daisy chains of 7 modules each installed under the TPC warm vessel. The modules consist of two layers of 32 polystyrene scintillator strips, parallel and each 5 cm wide each [57]. A wavelenght-shifting fiber is placed at one end and read by a Hamamatsu multianode *PMT*, the other end being mirrored. The overall

dimensions of each module are $4 \times 1.6 \times 3.2$ m^3

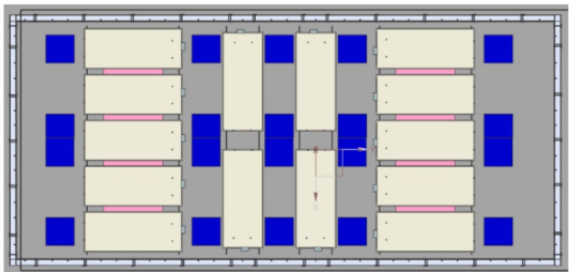

Figure 3-4: Map of the Bottom CRT layout from top-down wiev, each rectangle is a scintillation module, while the blue squares are the warm vessels support. [44]

3.4.2 The side CRT

The side CRT uses decommissioned scintillator modules from the MINOS experiment [58], and each module is formed by 20 adjacent strips of $800 \times \times 1$ cm^2 polyester scintillator contained in a metal sheath, and for which each strip has an embedded WLS fiber running through the center. The fibers are connected to the optical readout, which consists of an array of 10 Hamamatsu silicon photomultipliers. The side CRT is divided into 8 different regions, each formed from two layers of MINOS modules: north (downstream), south (upstream), west-north, west-center, west-south, east-north, east-center and east-south. The upstream wall is the only one in which the two layers are arranged orthogonally, while the others are modules with strips arranged horizontally and parallel to the cryostat. Finally, due to its proximity to the cryostat, the north wall is composed of cut MINOS modules of different lengths to maximize surface coverage.

3.4.3 The TOP CRT

The top CRT is composed of 123 modules, 84 installed on the upper horizontal plane and 39 that cover the upper perimeter of the cryostat (vertical

rims). Each module consist of a hodoscope made up of two orthogonal layers in which 8 scintillator bars of 23 cm each are enclosed in 1.86 × 1.86 m^2 aluminum boxes,the bars of the top layer are 10 mm thick and the ones of the bottom 15 mm thick . Each scintillator is instrumented with two WLS fibers embedded along the length of the bar and 6 cm from each side [44].The Read-out system is conformed by a Silicon Photomultiplier(SiPM) with a configurable discrimination threshold.

Both the top and side CRTs have a multichannel front end board on each module and a time-to-digital converter operating at a clock frequency of 250 MHz with and accuracy down to 1 ns. A map of top CRT layout is shown in **3-5**.

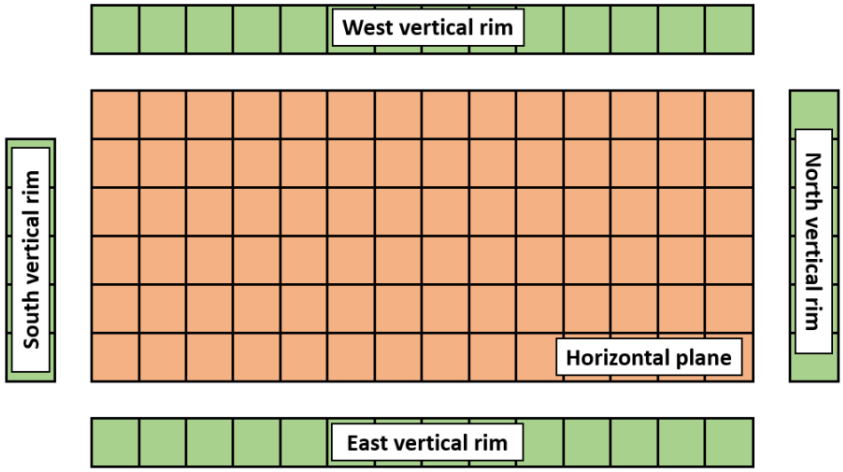

Figure 3-5: Map of the Top CRT layout from top-down [44].

3.5 Reconstruction in the CRT

The front end boards (FEB) of the CRT are sustained on a charge amplifier with configurable gain and for each one of the 32 channels that make up the CRT. The amplified signal passes through a fast shaper of 15 ns from which the digitized signal is the binarized by a discriminator. The discriminator threshold is provided by a 10-bits ADC on each of the 32 channels. The 32

trigger signals are sent to an FPGA where they are paired with AND logic to form coincidence signals for each of the two fibers from the same scintillation bar.

The event trigger is obtained by the OR logic of the even-odd coincidence channels of one scintillator layer in AND logic with the OR even-odd coincidence of the other layer, where even-odd coincidence means the coincidence of the ith and (i+1)th channel,i.e both on the same module. Each of the 32 discriminators can be enabled or disabled individually.

Unlike the *PMT* and TPC systems the CRT operates in auto trigger mode. Each data line is read out by the associated DAQ server, which polls every 80 ms and stores events within and enlarged 50 ms widown around the trigger time for further analysis. The pedestal and gain values from which the trigger is set are obtained by fitting the integrated ADC charge spectrum to each channel. The fit to the ADC spectrum that defines the pedestal is a gaussian distribution on the spectrum of a channel for which there is no hit coincidence, while the gain is obtained as the average distance between the photoelectron peaks in the cosmic muons charge spectrum.

After converting the raw data into a readable format the CRT hits are reconstructed for the side and top CRTs due to the difference in geometry and trigger. The difference is graphically shonw in **??**.

The first step of the reconstruction is the conversion from ADC to number of photoelectrons following the expression:

$$n_{p.e} = \frac{ADC_i - P_i}{G_i}$$

Where $n_{p.e}$ is the number of photoelectrons, ADC_i corresponds to the number of counts on the i-th channel and P_i, G_i are the pedestal and the gain evaluated from the previous calibration.

The top CRT hit reconstruction is performed by selecting the scintillator strip corresponding to the channel with a higher PE for the top (1 cm thick) and bottom (1.5 cm thick) layers. Due to the orthogonal configuration of the

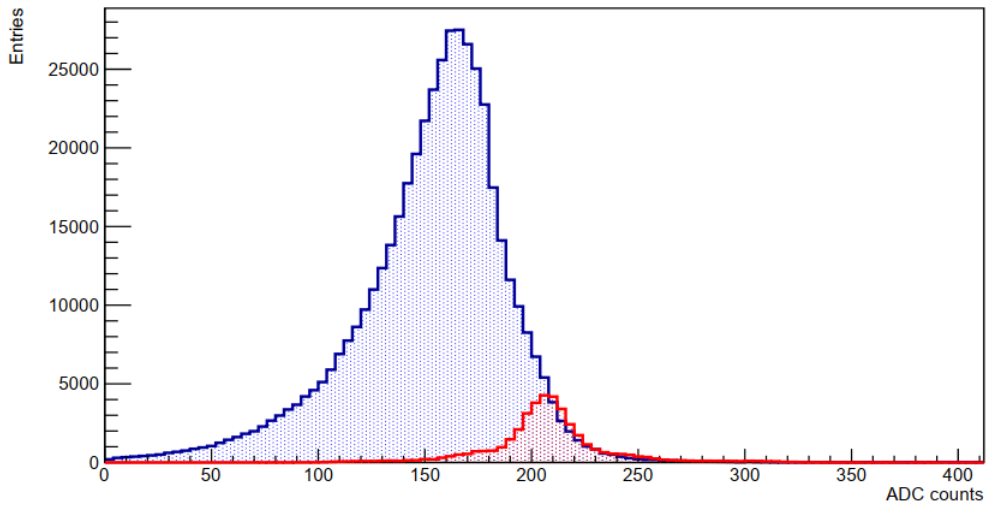

Figure 3-6: Pedestal distributions when the channel does not contirbute to the CRT trigger coincidence(blue) and when it does (red). Its possible to see a gaussian distribution centered at a higher number of counts for the second case.[44]

two layers, the position of the hit in the top CRT is univocally determined by the overlapping region of the two scintillator bars. The global module coordinates are defined with respect to the center of the module, so the local module coordinates are reconstructed with respect to the middle position along each scintillator strip.

The coincidence logic of the side CRT is performed offline in the reconstruction stage, as the inner and outer modules are read by multiple FEBs. Hits are identified by selecting in each FEB the channel that generated the trigger signal, i.e the one with the highest charge amplitude. Since most of the sidewalls have and x-x configuration different reconstruction scenarios are considered.

- If the strip is read on both ends by two different FEBs and both of the corresponding channels are above threshold, the longitudinal position with respect to the center position of the strip is reconstructed by

comparing the time stamps recorded by each FEB.

$$z = \frac{T_A - T_B}{2} \times v_{wls}$$

Where T_A, T_B are the timestamps recorded by the two FEBs, z is the hit position and v_{wls} is the light group velocity in the wavelenght shifter fiber, estimated to be of 0.062 cm/ns. The procedure is performed in both the inner and outer layers in order to obtain the final longitudinal coordinate as the average of the two reconstructed positions. x and y are obtained by averaging the respective transverse coordinates of each strip given by the geometry.

- If only one of the two channels has both channels above threshold, the CRT hit position is reconstructed using only the information from that strip
- If none of the layers has two channels above threshold coincidences with one channel are considered and the hit is reconstructed by default at the center of the scintillator bar and averaged over the two layers.

The south side wall is the only one of the side configurations that uses an x-y setup: the outer layer has vertical strips while the inner layer has horizontal ones, and is read only in one end. The east-west coordinate is given by the position of the vertical strip hit in the outer layer and the top-bottom coordinate by the horizontal strip hit on the inner layer [55].

The spatial resolution of the reconstructed hitsfor ach CRT subsystem. The top CRT assigns a hit within a volume of $23 \times 23 \times 2.5$ cm^3; hence the standard deviation for each coordinate, assuming the hits are uniformly distributed along each strip, is $\sigma = \frac{width}{\sqrt{12}}$, where the width is 2.5 cm for the z coordinate and 23cm for the other two. This results in a set of upper limits defined as σ_x ¡ 7 cm,σ_y ¡ 7cm,σ_z ¡ 1 cm [44].

The side CRT resolution needs to be treated more carefully, as the spatial resolution depends strongly on the timing resolution. In general, the error along the transverse direction is $\sigma_z \sim$ 1cm, while the resolution along the longitudinal directions has only been evaluated in simulations, with its values reported in table **3-1**.

Region	$\sigma_x[cm]$	$\sigma_z[cm]$
Side,lateral	16	229
Side,north	11	103
Side south	17	14

Table 3-1: Spatial resolutions of the side CRT hit reconstruction with x as the lateral and y as the longitudinal coordinates. [44]

3.6 The ICARUS global trigger

The data collection from all of the three ICARUS subdetectors is set up by the global trigger of ICARUS, which is generated when a given number of *PMT* pairs in the same TPC are above a threshold in coincidence with the expected beam gate. The requirement corresponds to $> 5 PMTs$ above a 400 ADC threshold each (on-beam majority trigger). When satisfied,the trigger sends a signal to the *PMT* digitizers which store and sample the corresponding waveforms. The *PMT* signals are recorded for $\pm$ 1 ms around the trigger time of the primary event, with the aim of recognizing and tagging cosmic rays crossing the detector during the TPC drift time. Since the the beam width is significantly different depending of the beam (1.6 μs for BNB and 9.6 μs for NuMI), only a small fraction of the 2 ms on the readout window is set up to the trigger requirements while the rest is called the *out-of-time* window. The logic of the out-of-time window presents more stringent requirements, as the threshold is defined by $>$ 10 *PMT*s over a 400 ADC threshold, with a sampling of the *PMT* waveforms that is also different: 28 μs for the on-beam and 10 μs for the off-beam ones.

Since the in-time and out-of-time windows differ in their threshold requirements the latter cannot be used as background for the neutrino beam window. For this reason, a fake beam gate window with the same on-beam time requirements of the trigger is opened between the different beam spills. The data stream is called off-beam and its frequency depends on the beam type. This means that ICARUS operates on 4 different data streams: BNB, NuMI, BNB OffBeam and NuMI OffBeam.

TH triggering logic is determined by an FPGA, which at the fulfillment of the programmed requirements generates a logical signal. This logical signal is referred to as the global trigger signal and is distributed via a fan-out to the CRT and *PMT* subsystems, while a similar signal is generated from two different FPGAs and distributed to the TPC board digitizers. The CRT FEBs digitize the global trigger signal and in the CRT hit reconstruction stage, the digitized trigger is used as the 0 reference of the CRT hit timing

Chapter4

Combined analysis on ICARUS

Neutrino interactions need to be distinguished from the cosmic induced background, this can be achieved by determining the time at which each particle enters in the LAr volume as out-of-time cosmics can be rejected since their absolute timing would not coincide with the beam spill. This work contributed in the development of an analysis tool that, combining all the reconstructed information from the three ICARUS sub-detectors (TPC, PMT and CRT), is able to provide a precise timing to the majority of the reconstructed tracks.

Each one of the detectors that constitute ICARUS reconstructs signals independently and with distinctive features that correspond to their principal function in the system.

The TPC is able to precisely reconstruct tracks in three dimensions, but its distance from the anode plane is not known, as it can be determined only from the particle drift time. By default, the track position is reconstructed assuming that the particle is interacting at the time of trigger, so in-time interactions are correctly reconstructed while out-of-time ones are spatially displaced. For a subset of tracks that cross the cathode, the event reconstruction is able to perform a stitching between the portions of tracks in the two opposite TPCs and, from the drifted distance at which the two objects

merge, to determine the track timing with respect to the trigger.

For interactions inside the LAr volume, an optical flash provides a very precise ns timing information, but lacks in spatial resolution due to the large volume in which the scintillation light propagates and the limited granularity of the PMTs.

The CRT detector, on the other hand, provides a ns time resolution and a spatial resolution of the order of cm, but it is located outside of the cryostat, therefore it cant being easily correlated with reconstructed interactions inside the LAr.

A precise econstruction and analysis of events combines the features of each of the sub detectors, enhancing background rejection and eventually allowing the analysis of out-of-time interactions for calibration and physic studies.

A figure that sketches the signals acquired on each detector and how they are related to each other in ICARUS is shown in **4-1**. A cosmic event triggers a CRT hit and after few nanoseconds fires a light that triggers the PMT systems while in parallel a reconstruction in the LArTPC by means of the drifted charged particles is done. The first step of the combined analysis approach started by coupling the coincident times of the PMT and CRT detectors, thus connecting the LArTPC with the CRT system and obtaining a decent classification of cosmic and neutrino events in the in-time window. This was an analysis tool developed last year and, as a next step, for the first time signal out of the TPC detector was included in the combined analysis, enabling the classification and rejection of out-of-time events.

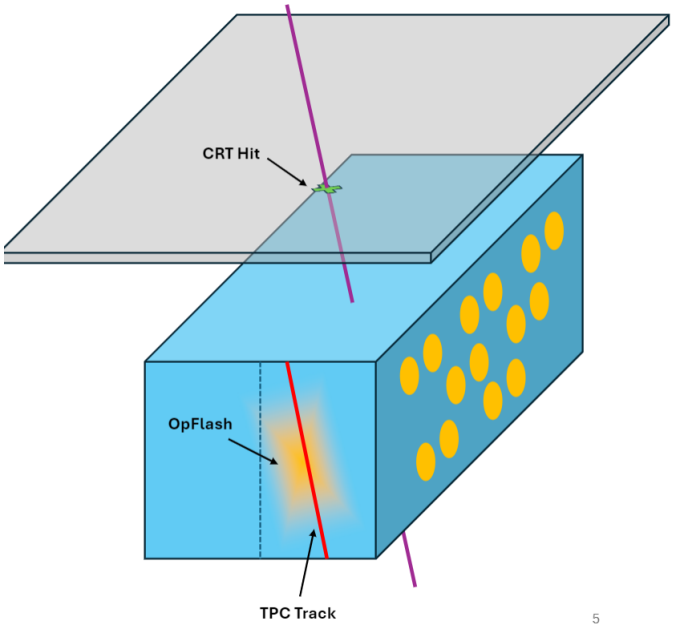

Figure 4-1: Signals of the three subdetectors of ICARUS for an out-of-time track.

4.1 CRT-PMT matching

4.1.1 CRT hits and optical flashes

The synchronization between the Cosmic Ray Tagger and the photo multiplier system by means of the global trigger signal can be used for event selection and, by comparing the time of the CRT hits and the optical flashes, it is able to determine the direction of a cosmic particle and the origin of the primary event trigger, whether or not it was matched with a CRT hit.

The selection is done taking advantage of time distribution for both the CRT hits and the flashes with respect to the global trigger signal, due to the time of flight between the particle crossing the CRT plane and the scintillation light reaching the PMTs. This shift is visible in figure **4-2**, where the similarities on the shape of the distributions are seen along the time shift of one with respect to the other.

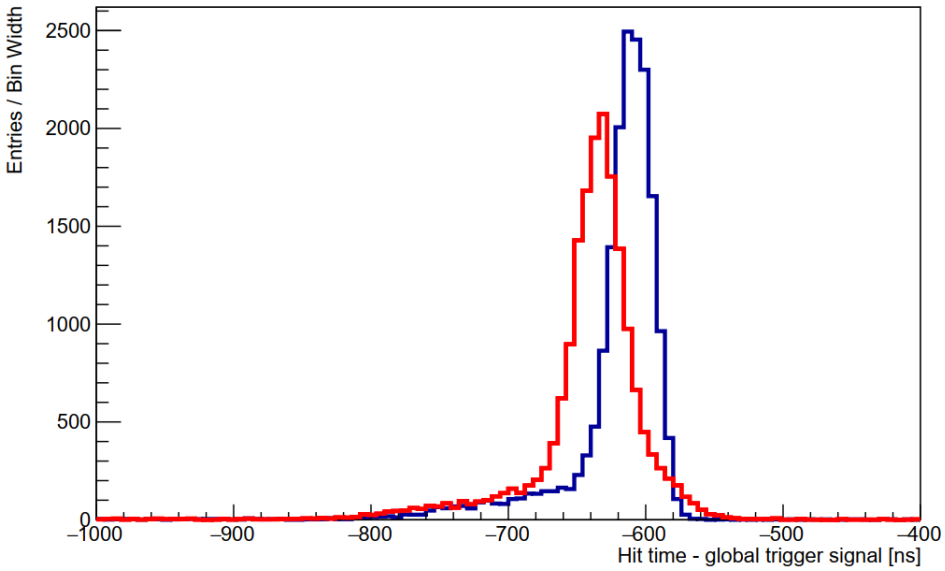

Figure 4-2: Time distribution of the CRT hits (red) and the optical flashes (blue) with respect to the digitized trigger signal for a subsample of the Run1 [44].

As the PMT system is triggered the information of the optical hits is stored in an optical flash object that collects all of the hits within one microsecond. The CRT-PMT matching consists of coupling one or more CRT hits with an optical flash using only their time information, pairing CRT hits and optical flashes that occur within a frame of ±100 ns from each other. Depending on the topology of the cosmic track, more than one CRT hit can be matched with an optical flash. An example is a crossing particle entering through the top CRT and exiting through the side CRT.

Each optical flash gets assigned a position from its light barycenter, weighted by the amplitude of each optical hit, and the direction of the detected cosmic particle is determined by the sign of the difference between the the flash and the CRT hit times, defined as:

$$\Delta t_{CRT-Flash} = t_{CRT} - t_{Flash} \tag{4-1}$$

where the direction is defined as entering for $\Delta t_{CRT-Flash} < 0$ ns and exiting for $\Delta t_{CRT-Flash} > 0$ ns. It is worth noting that the difference in times is

not a direct measure of the time of flight, as it also includes the time of propagation of the scintillation light from the ionization point to the PMTs.

Information about the distribution of the CRT times matched with an optical flash is avaliable at both data and monte carlo level. The figure **4-4** shows the distribution of the top CRT hits for data and monte carlo, mostly populated in the region $<$ 0ns. The negative value is expected as cosmic particles cross the CRT before entering the TPC volume. The Gaussian-like distribution is due to the cosmic incidence angle, which could determine a longer or smaller time of flight[59].

Figure **4-5** shows the distribution of $\Delta t_{CRT-Flash}$ for the side CRT hits matched with an optical flash, in this case the distribution is different from the one obtained for the top CRT, showing two peaks, one on the negative and one on the positive region. Similarly to the top CRT the negative peak is due to cosmic particles that enter from the side modules and generate a flash in the liquid argon; while the second peak is generated by the cosmic particles that exit the detector. In general negative peaks are enhanced in the regions located in the upper half of the detector volume, as for the lower half the entering cosmic rays are disfavored by the zenith angle dependence. The positive peak, on the other hand, favors the lower regions of the side CRT as cosmic rays entering from the Top CRT and exiting from the upper regions of the side one would not cross the liquid argon volume. The distribution of the side CRT hits along the height of the detector as a function of $\Delta t_{CRT-Flash}$ is shown in figure **4-3**.

.

Further analysis on the CRT planes allows to locate the regions for which the distribution of the time differences changes more drastically between data and monte carlo.

A study on the data and monte carlo distributions of the time differences was done on each CRT region, being the distributions for all the planes of the Top CRT very similar. The distributions of the side CRT are of more interest as there are regions for which the inefficiencies on the data acquisition are clearly noticeable as they are the East-south (region 45) shown in figure **4-37**,

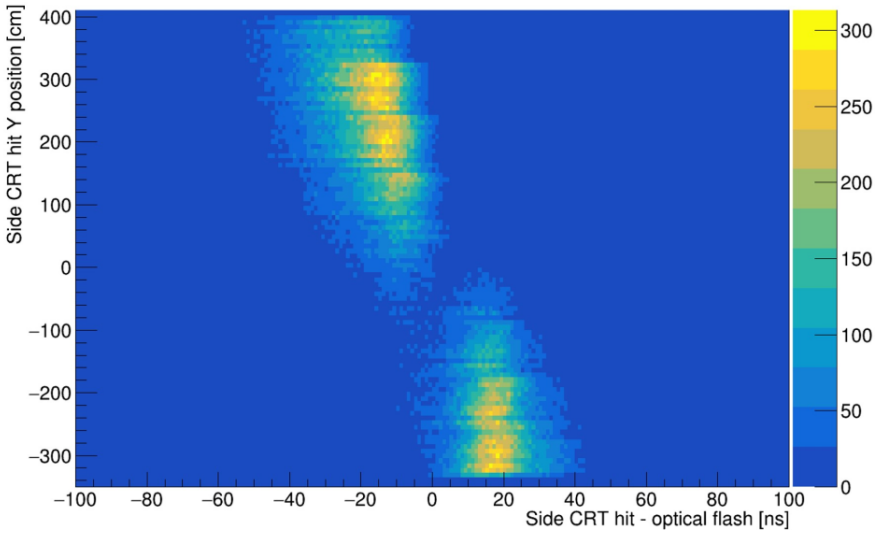

Figure 4-3: Side CRT position along the vertical direction with respect to the time differences [44].

while others as the West-South (region 40), shown on figure **4-32**, perform better. All the $\Delta t_{CRT-Flash}$ distributions are shown in the appendix A. .

.

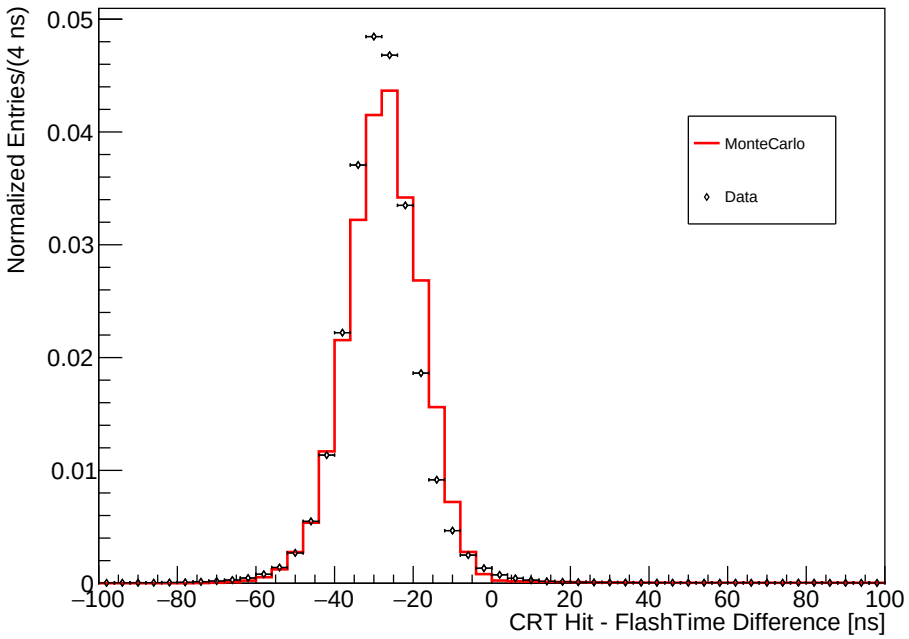

Figure 4-4: Area-normalized distribution for the Top CRT and optical flashes time difference between data (black) and Monte Carlo (red)

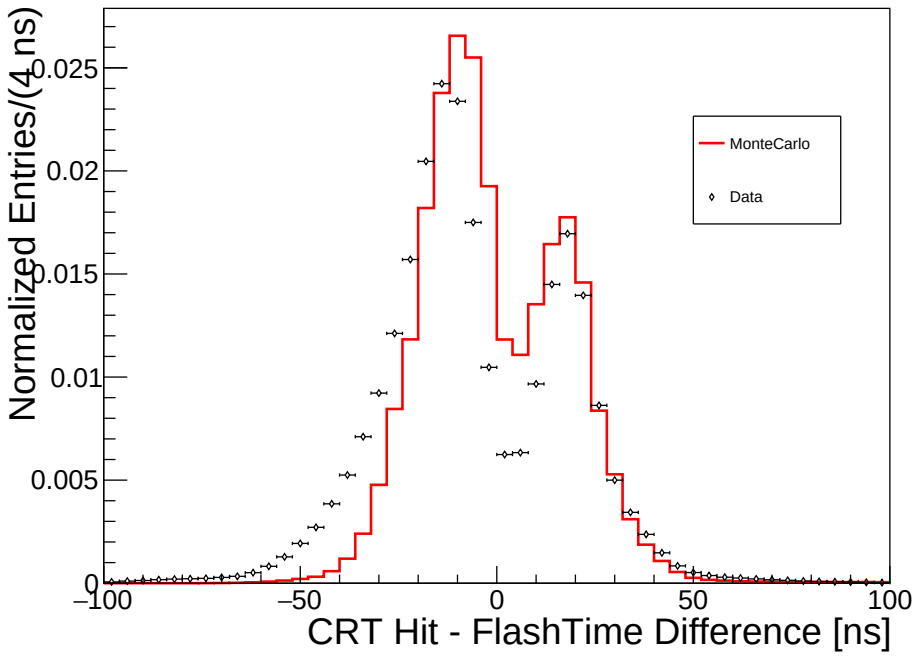

Figure 4-5: Area-normalized distribution for the Side CRT and optical flashes time difference between data (black) and Monte Carlo (red)

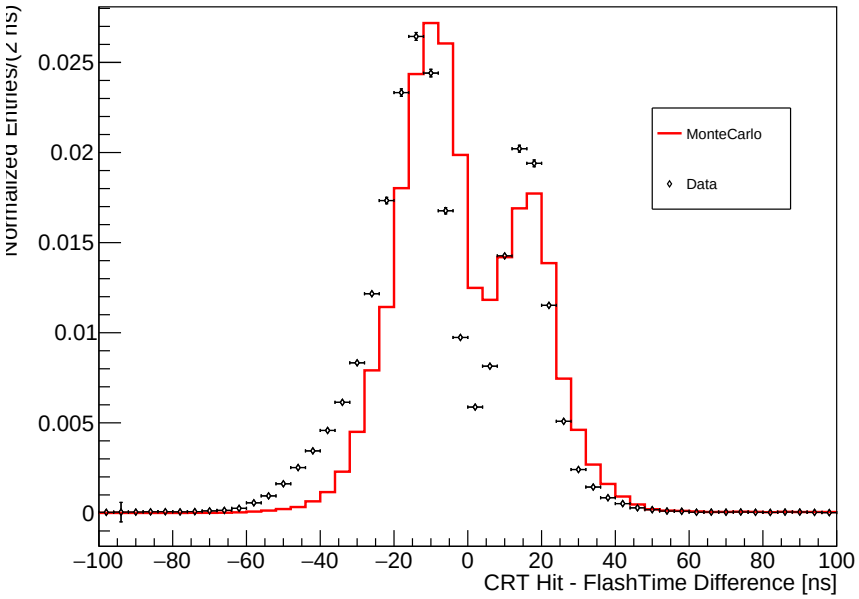

Figure 4-6: Comparison between Side CRT hit from region 40 matched with optical flashes for Run 1 data (black) and Monte Carlo out-of-time cosmics (red).

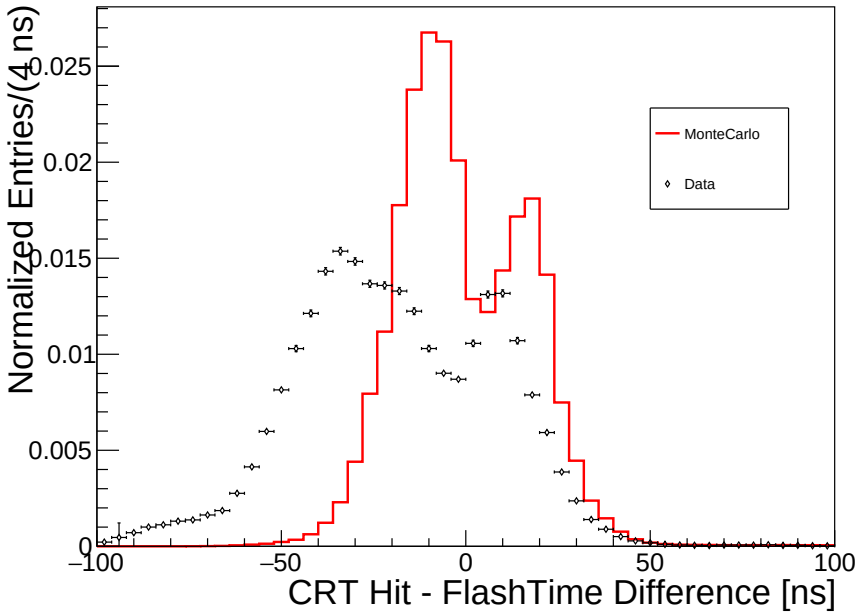

Figure 4-7: Comparison between Side CRT hits from region 45 matched with optical flashes for Run 1 out-of-time data (black) and Monte Carlo out-of-time cosmics (red).

4.1.2 In time event selection and background rejection

A preliminary classification of events can be done by means of the time of the CRT hits matched with an optical flash. The categories are:

- optical flashes with no matching CRT hits,
- optical flashes preceded by a top CRT hit,
- optical flashes preceded by a side CRT hit,
- optical flashes preceded by a top CRT hit and followed by a side CRT hit,
- optical flashes followed by a top CRT hit,
- optical flashes followed by a side CRT hit,
- all other possible cases.

These categories are mutually exclusive and form a classification system from which it is possible to distinguish cosmic from possible neutrino events. A CRT-PMT classification was performed on a BNB Majority data sample. Figure **4-9** displays the percentages of the out-of-time optical flashes that fall into each category by counting the number of CRT hits associated with the flash time.

The distribution shows that the majority of optical flashes are preceded by one top CRT hit without any CRT signal (61.5%) while a smaller fraction (10.3 %) is only preceded side CRT hit. Flashes that are not matched with a CRT hit can be explained by effects related to: geometric acceptance, efficiency, threshold, dead time of the CRT boards, cosmic rays that back scatter and enter from the bottom of the detector, and eventually, neutral particles. In addition to these effects, optical flashes originated from late scintillation light from the liquid argon or incorrectly reconstructed optical hits cannot be matched with CRT activity. The labeling of a particle as *entering* or *exiting* particle relies on the time resolution of the $\Delta t_{CRT-Flash}$ measurement. For the case of the Top CRT the resolution of ≤ 4 ns is not limiting since for a mip this could be interpreted as a position uncertainty

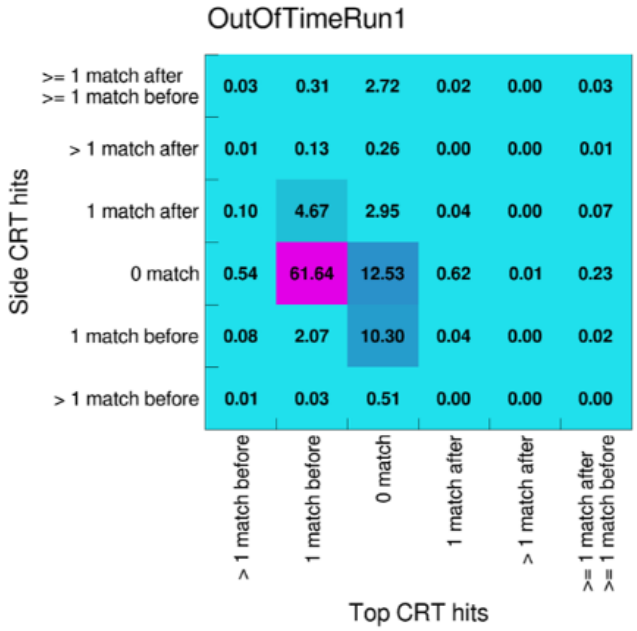

Figure 4-8: Relative percentage of optical flashes matching with Top and side CRT hits for an out-of-time BNB data sample.

of less than 1.5 m, well within the 4 m distance between the Top CRT and the TPC. For the side modules, on the other hand, the scintillator panels are a few cm away from the cryostat walls, therefore they could not be able to resolve the direction of a track.

Nevertheless the aforementioned inefficiencies, the CRT-PMT matching is still a good preliminary event selection algorithm, filtering interesting events (flashes not matched with any CRT hits) without going through the entire processing of all TPC waveforms. This is shown in figure **4-8** where the proportion of flashes that are not matched with any CRT hit on the BNB in-time window increases up to (33 %), successfully labeling possible neutrino events just by means of the time differences in the CRT and PMT detectors and providing a filter for events for the in-time window.

A further improvement of the classification algorithm would be the inclusion

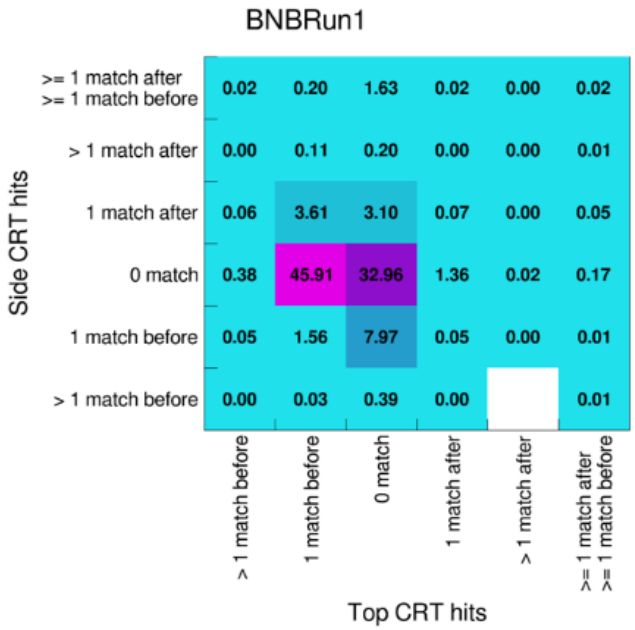

Figure 4-9: Relative percentage of optical flashes matching with Top and side CRT hits for an on-time BNB data sample.

of the TPC track information. The inclusion of the TPC subsystem information and the development of a CRT-TPC connection, integrated with the aforementioned CRT-*PMT* matching, allows for a more robust selection and classification of the events as cosmic or neutrino-like.

4.2 Out-of-time cosmic background

4.2.1 TPC-PMT-CRT combined analysis

The TPC-PMT-CRT combined analysis, referred as triple-matching, associates a reconstructed TPC track with an optical flash, already matched with one or more CRT hits by means of their timing information. The three detector association is performed by first combining independently the CRT hits and the optical flashes as mentioned in the previous section while in par-

allel combining the optical flashes with the reconstructed TPC tracks. The TPC-PMT match is done by pairing all of the flashes with each one of the tracks contained in an event. After these two processes are performed and a group of preliminary cuts is applied, a final set of pairs is obtained and from this a resulting set of triples is obtained matching the common flash time in both pairs. The construction of the triples is explained in detail on the following section.

4.2.2 Dataset preliminaries

The formation of the TPC-PMT pairs and CRT-PMT pairs is done using two parameters that serve as selection cuts on the likeliness that the same particle produces all the signals. The cut on the CRT-PMT pairs corresponds to pairing only signals that occur within a $\pm$ 100 ns interval, as explained in the previous sections, while the cut applied on the TPC-PMT pairs is the proximity of the light and charge signals in the LAr volume.

An interaction that produces an optical signal in the PMT will have a scintillation light distribution correlated with the reconstructed charge in the TPC. However, as scintillation light is propagated on a spherical surface while charge is only propagated along the drift direction, the easiest way to compare light and charge is by means of their corresponding barycenters, which should be close to each other. The track charge and flash barycenter matching is currently being used in neutrino event selection by looking for interactions that occur within 1 m along the longitudinal coordinate of the TPC (z) from the triggering flash. The z axis was chosen as it is directed along the neutrino beam and it presents higher PMT segmentation. For the triple-matching analysis the same flash and track pairing criteria was used, but the window was enlarged to $\pm$ 3 m to maximize efficiency and introduce a background component for optimization studies described in the following sections.

As the reconstruction in the drift coordinate of the track depends on the drift time inside the TPC, a quality cut was applied to the pairing of the TPC-PMT signals to check whether the flash time is compatible with the track or not. The drift coordinate of the track can be reconstructed using

the following equation:

$$\Delta x_{reco} = (t_{TPC} - 850 - \frac{t_{flash}}{N_{ticks}}) * N_{ticks} * 0.157 \frac{cm}{\mu s} \tag{4-2}$$

Where $_{reco}$ in equation 4-2 corresponds to the change in the drift coordinate due to the flash time, t_{TPC} to the tics stored in the digitizer from which the signal arrives in the TPC, t_{flash} is the optical flash time, 0.517 $\frac{cm}{\mu s}$ is the drift velocity in the TPC. N_{ticks} and 850 correspond to the number of tics measured by the TPC sampling and a tic offset of the pre-trigger digitizer. One tic corresponds to 0.4 μs.

If, when reconstructing the track dritf coordinates at the time of the paired flash, part of the tracks are outside the TPC volume the track is discarded.

As each one of the TPC-PMT pairs has a time that is common to at least one PMT-CRT pair, a triple of TPC-PMT-CRT signals is formed using the PMT flash time as the mediator between both sets, this association relates the TPC tracks with the CRT hits and allows the evaluation of new quantities from TPC and CRT values. The total number of triples of an event is given by all possible combinations of TPC tracks, optical flashes and CRT hits associated with that flash in the event

$$N_{Triples} = N_{Tracks} \times N_{Flashes} \times N_{CRTPMThits} \tag{4-3}$$

Each triple contains:

- the spatial coordinates and charge barycenter of the drifted TPC track assuming the time of the flash,
- the flash time and barycenter of the flash,
- the spatial coordinates, the time and the plane of the CRT hit.

The direction of the drifted tracks is determined by fitting the three dimensional hit points by means of a Principal Component Analysis (PCA), assuming that the tracks can be approximated as straight lines, their direction is determined by the eigenvector of the PCA principal component. The fitted line is then extrapolated to the CRT hit plane to determine the interception between the track and the plane $\mathrm{x_{ext}}, \mathrm{y_{ext}}, \mathrm{z_{ext}}$.

To increase the performance of the PCA analysis a filtering is done on the tracks that form the candidate triples. To assure that the number of points collected is robust enough to obtain a reliable value for the direction, tracks for which the number of good reconstructed points is below 5 are discarded, as well as tracks for which the length is below 30 centimeters as the extrapolation of tracks of this kind is very prone generate errors when extrapolated on the CRT plane. A third parameter, called track point density; That quantifies the distance between two reconstructed points of the track per centimeter for each track length is applied as a quality cut. A density of at least 1.6 $\frac{points}{cm}$ is selected as it discards shower-like or splited tracks without losing volume on the reconstructed tracks,as most of them have a density of 5 points per centimeter.

The distance between the extrapolated point and the hit in the CRT plane is evaluated as:

- $\sqrt{(\mathrm{x_{ext}} - \mathrm{x_{CRT}})^2 + (\mathrm{z_{Proj}} - \mathrm{z_{CRT}})^2}$ for the planes of the top CRT in which the vertical coordinate (y) is constant,
- $\sqrt{(\mathrm{x_{ext}} - \mathrm{x_{CRT}})^2 + (\mathrm{y_{Proj}} - \mathrm{y_{CRT}})^2}$ for the planes of the CRT in which the longitudinal coordinate (z) is constant, namely north and south ,
- $\sqrt{(\mathrm{y_{ext}} - \mathrm{y_{CRT}})^2 + (\mathrm{z_{Proj}} - \mathrm{z_{CRT}})^2}$ for the planes of the side CRT in which the drift coordinate (x) is constant, namely west and east.

A sample of $\sim$ 24 million triples was constructed from the Run2 of ICARUS events with BNB majority and a sample $\sim$ 17 million of triples was constructed by monte carlo simulations of the Run2. The construction of the triples and implementation of the triple-matching algorithm for both samples was done on a LArSoft analysis module. LArsoft is a novel software tool designed specifically to work on the reconstructed signals of LArTPC experiments.[60].

A distribution of the distances for the extrapolated track interceptions to the CRT in all planes done in the data sample is shown in **4-10**. The distribution peaks around 20 cm from the hit on each module, which is expected as 20 cm is the length of a bar in the module.

The track extrapolaiton and CRT distance is referred to as Extrapolated Distance to the plane (EDP) and not only connects the signals of the detectors inside the LAr with the CRT, but also allows to perform an analysis on the quality of the matched triples and how likely they are to be a set of signals that correspond to the same particle crossing the LAr volume in the out-of-time window.

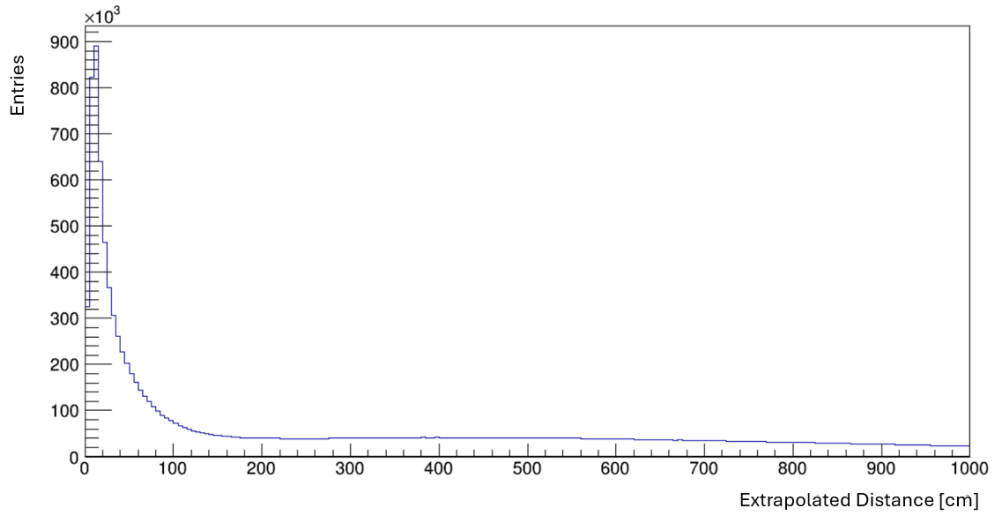

Figure 4-10: Distributions of the distances extrapolared to the CRT plane for all the candidate triples in the Run2 BNB Majority data sample. The preliminary selection includes the signal region (low EDP) and a background component (almost flat distriubtion for EDP¿200 cm

Additionally, a distribution of the projected points for each plane was done to evaluate the precision with which the EDP populates each CRT module. Both figures **4-11** and **4-12** show a distribution of the extrapolated coordinates on the Top CRT in an interval that corresponds to the width of a strip (a projected distance of less than 30 centimeters). The data distribution shows features such as the malfunctioning of one bar and the dead space between the module, while the monte carlo distribution shows the lack of simulation of single bar inefficiencies and dead space. On both distributions

points are enclosed in what corresponds to the TPC fiducial volume below the top CRT. [61].

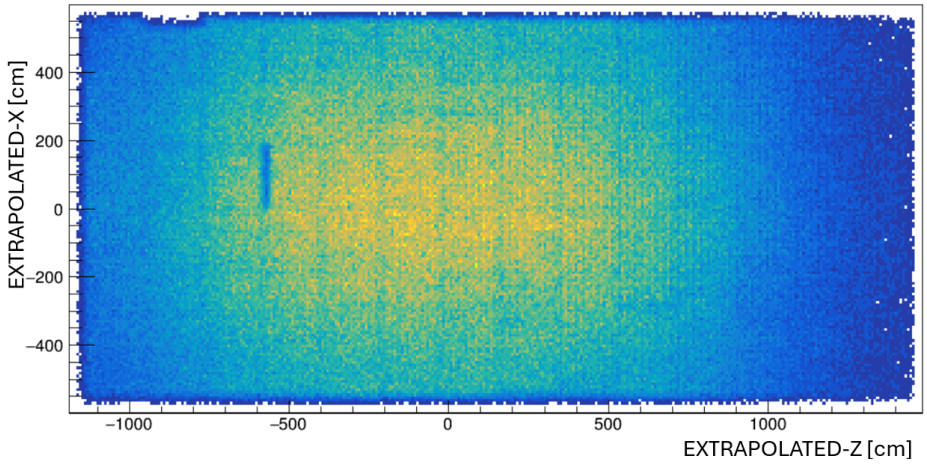

Figure 4-11: Distribution of the extrapolated crossing points for a data sample of BNB RUN2

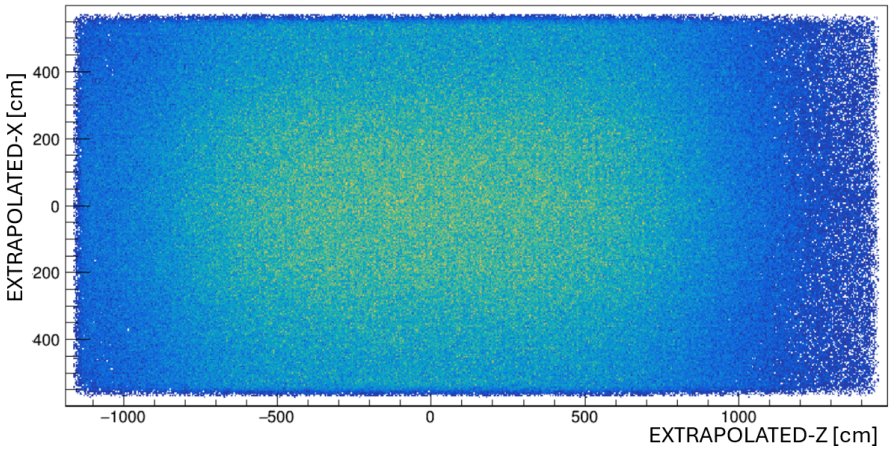

Figure 4-12: Distribution of the extrapolated crossing points for a RUN2 monte-carlo production

4.2.3 The triple matching algorithm

A diagram that summarizes the spatial and temporal coordinates on each detector saved on each triple along with the quantity that serves as the connection for the triple is shown in **4-13**. The steps of the triple-matching, including the selection and quality cuts, can be recalled as:

1. tracks that fulfill the resolution constraint are paired with each one of the flashes of the same event and only the pairs for which the barycenter distance in z is less than 3 meters are considered,
2. the tracks are drifted to the time of the flash to which they are paired. Pairs for which the drifted track is not fully contained in the TPC are discarded,
3. To each pair the CRT hit that corresponds to the flash time is added and the triples are formed,
4. In each triple the TPC track is projected to the CRT plane and the EDP is saved as an additional quantity.

As the EDP performs a good approximation of the eventual CRT coordinates of a cosmic track reconstructed in the TPC, it is functional as a criteria for the goodness of the triples formed, meaning by goodness how likely is that one triple corresponds to the reconstructed signals of the same particle or not. However, this criteria cannot be used as the only selection cut that determines if the matched values correspond to the same particle, since there can be multiple cosmic tracks for which the extrapolated distance falls satisfactorily within the area of the CRT module and multiple flash times for which the EDP of the drifted track gives a close distance of the intercept to the CRT module.

The inclusion of the distance between the light and charge barycenters as a measure of the goodness of the triple also contributes to the selection of *good* triples over the ones that are wrongly assigned. As it can be seen in figures **4-14** and **4-15** there is a strong correspondence between a short light-charge barycenter distance and a short distance between the extrapolated and CRT hit coordinates, so a parameter space in which both quantities are below a

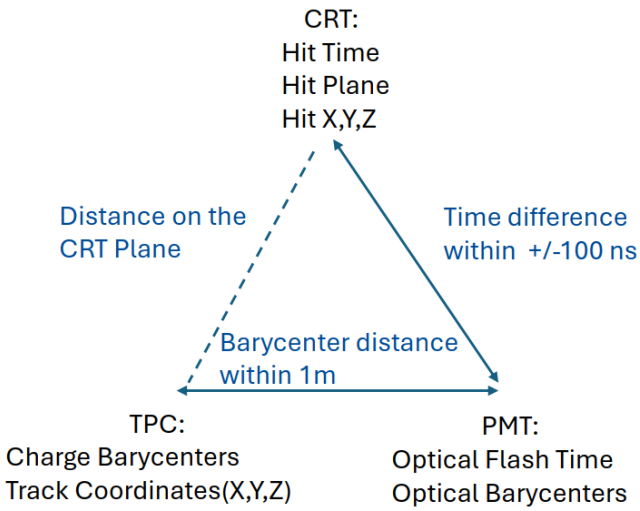

Figure 4-13: Schema of the coordinates that form each triple (vertex) and the quantities that relate each detector on the formation of a triple (lines). The dotted line on the TPC-CRT match indicates that no cut is performed on the EDP.

given value can be a strong indicator of the goodness of the triple.

However, an analytical relation for the two variables is not easy to determine,as there is no reason to assume that a shorter light-charge barycenter distance will result in a good match rather than a close extrapolated distance or viceversa.

For these reasons a more robust statistical analysis was performed on the triples on both the data and monte carlo samples; characterizing not only the two quantities mentioned but other ones that determine the goodness of the triple as the track length,the time of flight of the particle, the directions of the track and the region of the CRT in which it is projected.

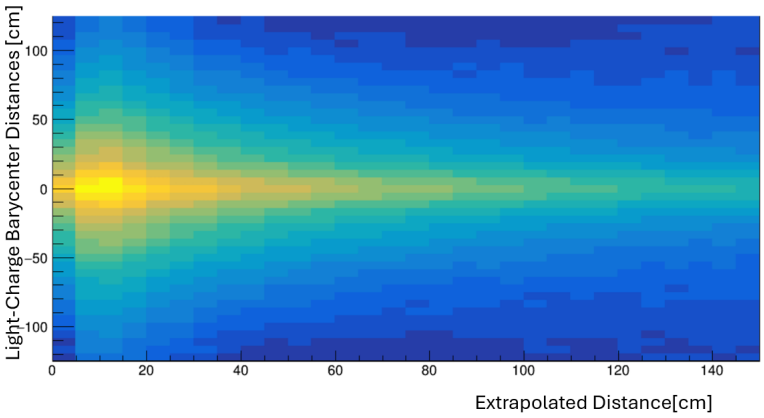

Figure 4-14: Distribution of the EPD with respect of the barycenter distance for the Run 2 monte carlo production.

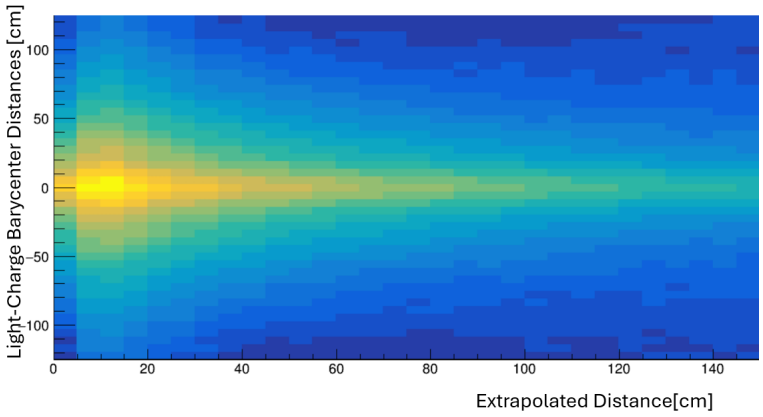

Figure 4-15: Distribution of the EPD with respect of the barycenter distance for the Run 2 BNB data.

4.3 Multivariate classification of the matches

Due to the difficulties in finding an analytical relation to determine the goodness of each triple, a Boosted Decision Tree was trained on the parameters that conform the triples to provide a score that defines the likeliness of each triple to be a *good* match. A Boosted Decision Tree(BDT) is a multivariate machine learning technique that takes a set of inputs and gets trained on

examples from which it weights each of the inputs and outputs a classification of the data as eventual signal or background, based on the values of the inputs for each entry [62]. This label was implemented at analysis level by keeping the information of the Geant4 ID of the particle that generates the signal on each detector. If a triple has the information on each detector that corresponds to the same particle ID it is defined as a *True-Match*. For the purposes of the training of the BDT the *True-Match* triples were defined as signal and the ones for which the values on the detector correspond to a different particle as background.

The sample of monte carlo on which the BDT was trained contained $\sim$ 15 million of triples of which $\sim$ 9.3 million corresponded to false matches and $\sim$ 5.8 million to true-matches, which corresponds to a ratio of $1 : 1.6$ for signal and background triples. This ratio however is not assured to be the same on the data since in the monte carlo used for the training, the proportion between in-time cosmic rays and neutrinos was not realistic and there are detector features that are not simulated. Therefore a further refinement on the BDT results was performed after applying the BDT output on the data sample.

The input variables of the BDT were:

- the absolute value of the director cosines of the TPC track,
- the length of the TPC track,
- the PMT flash time,
- the distance between the Barycenters in all coordinates,
- the total Extrapolate Distance to the Plane (EDP) and the extrapolated distance for each coordinate on the plane,
- the back-tracked distance from outermost point of the TPC track to the position of the projected point in the CRT,
- The time of flight of the particles from the CRT to the LAr volume, calculated as the difference between the CRT and PMT times over the back-tracked distance,

- the initial and end points of the track on the drift coordinate.

The BDT was implemented using the Toolkit for Multivariate Analysis with ROOT(TMVA) on the samples of data an monte carlo, both stored as root files [63].

The distribution of the events classified as signal and background for the monte carlo sample is shown in **4-16**, the output of the classifier corresponded to $\sim$ 6.1 million of signal events and $\sim$ 8.5 million of background events, a ratio of 1:1.39, which is a good approximation to the true triples present in the sample. Both the training and test samples were set to half the size of the triples that formed the monte carlo, the size of the training set was good since it is usually recommended a training set that is at least one order of magnitude bigger than the number of variables to train, and indeed no overtraining was seen on neither of the samples.

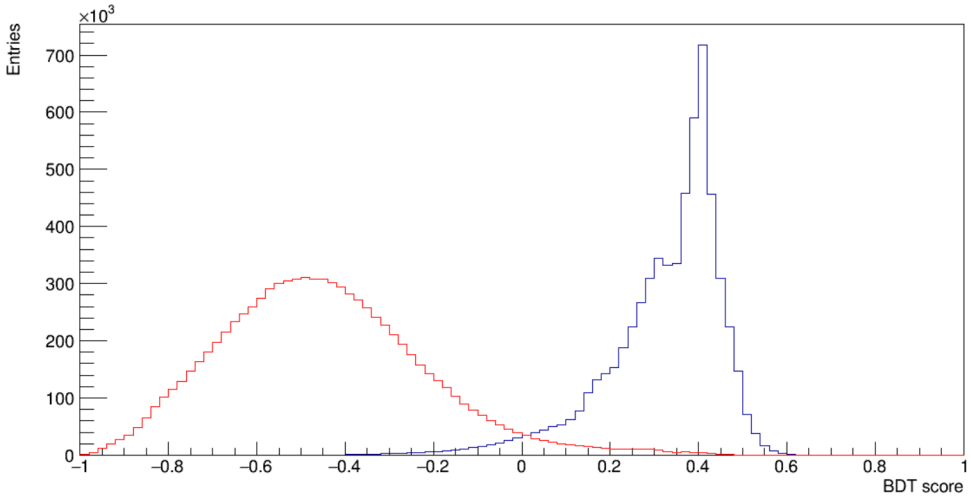

Figure 4-16: Distribution of events classified as background (red) and signal (blue) in terms of the BDT score for Run2 BNB monte carlo sample.

For each one of the input variables signal and background monte carlo truth distributions is done. Of particular interest are the distribution of the EDP for both shown in **4-17** and the light and charge barycenter distances along the z coordinate (shown in **4-18**), in which a clear distinction between the signal and background can be noticed in the distributions.

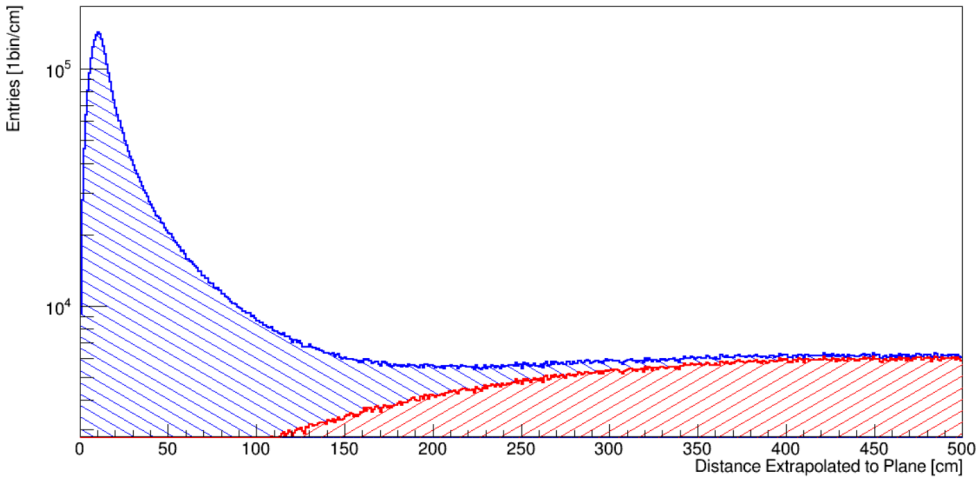

Figure 4-17: Log distribution of the Extrapolate distance to the plane for events classified as background(red) and signal(blue) with the BDT, the distribution of the signal shows a clear peak around 10 centimeters, expected for cosmic tracks goodly matched with signals inside of the detector. The background distribution shows a more uniform distribution at higher distances, in agreement with a random association of distances.

After the BDT was trained on the monte carlo the weights of the outputs defined by the classifier were used on the data sample of $\sim$ 24 million of triples to classify the events as signal or background, and the value of the parameter that serve as the score for the goodness of the triple was set. Nevertheless,as mentioned before, is not accurate to assume that the relation of signal and background on the data is the same that of the monte carlo sample since for the first one there are inefficiencies in the PMT and and CRT systems that are not simulated.

The assessment on how the combination of not simulated events alters the ratio of signal and background for the triples would require a intensive study of the reconstruction performed by the monte carlo simulation, the ratio for the data was determined fitting the BDT distribution obtained for the data sample to the monte carlo one. The background fraction was used as the parameter to fit the distribution and the whole process was performed on

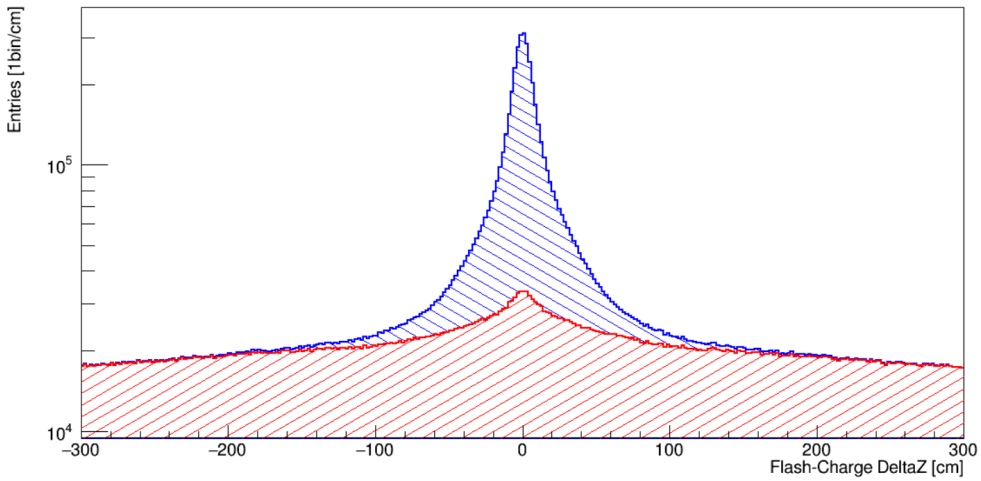

Figure 4-18: distribution of the distance between light and charge barycenters for events classified as background(red) and signal(blue) with the BDT, the distribution of the signal shows a distribution mostly populated within one meter, in agreement with the selection cuts applied on neutrino event selection analysis. The background distribution shows a more uniform distribution around the whole interval of distances.

RooFit [64].

The fit was a composite model formed by the signal and background of both the samples,expressed by the equation 4-4:

$$M(x) = (1 - f) * S_{Data} + f * B_{Data} \tag{4-4}$$

where m defines the composite model that describes the signal and background of the monte carlo after applying the BDT classifier and S_{Data}, B_{Data} to the number of signal and background triples in the data sample. After the fit was applied the background fraction f was found to be of 0.6, which corresponds to a ratio of signal to background of $1 : 1.66$. The distribution of the signal and background for the monte carlo sample and the data is shown in figure **4-19**,where is possible to see that although well fitted, there are are still some inconsistencies between the data and the monte carlo distributions; which lead to think that a detailed analysis on the detector inefficiencies and

their simulation needs to be performed either way.

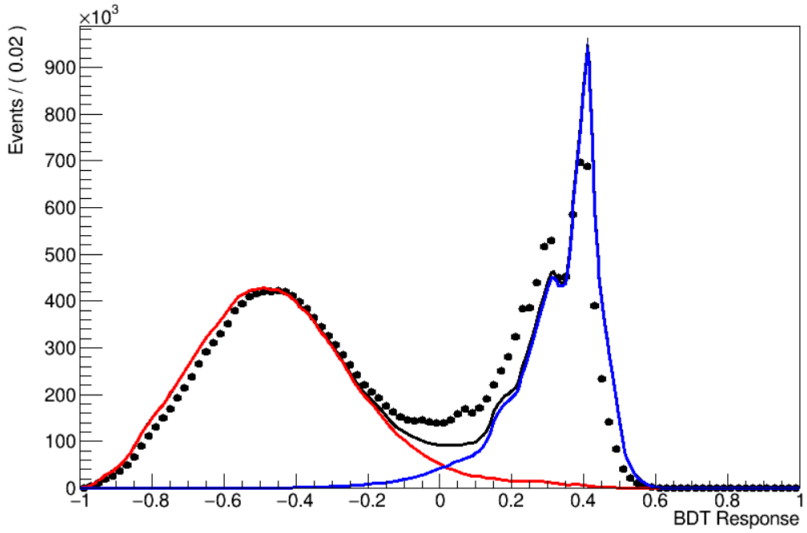

Figure 4-19: Distribution of the signal (correctly matched triples) and background(wrongly matched triples) for the data (dotted black) and the monte carlo signal (blue) and background (red) after the implementation of the BDT algorithm. It is possible to see some shifts on the distribution, particularly on the side on the positive score

With the new fraction of signal and background obtained for the data, the efficiency and purity of the BDT classifier was calculated using the TMVA framework [63], obtaining a signal efficiency of 96.7 ± 0.1 % and a signal purity of the 95.1 ± 0.1 % after a setting a cut on the BDT score at **0 ((.0081))**. The distributions of the signal efficiency, signal purity and background efficiency are shown in figure **4-25** This results show that with a signal efficiency and purity of over the 90% without taking into account some of the detector inefficiences and features on the monte carlo simulation, the triple matching is a reliable pairing method to group the signals from the three subdetectors of ICARUS. However further studies on the detector systematics and response need to be performed to improve the selection power of the algorithm and discriminate satsifactoraly neutrino interactions from cosmic ones in the out-of-time window. As a first step further calibrations on the top

CRT system are expected to be done by the summer, aligning the simulation geometry with the real one, determining a more correct data-monte carlo comparison. This corrections can increase the performance of the developed score.

After the data fit was done a plot on the distributions of the BDT inputs in the data and monte carlo signal and background was done normalized to the number of triples on each sample, showing that the shape of the monte carlo distributions corresponds to the one of the data for most of the variables. As with the first BDT output for the monte carlo, the distributions of the EDP and the light-charge distance of the barycenters is shown in figures **4-20** and **4-22**, the rest of the figures are present in the appendix A.

For the top CRT figure **4-20** represents the distribution of data,monte carlo signal, and background, while figure **4-22** corresponds to the extrapolated distances i z of the top CRT plane, for which and offset is present, while on the x extrapolated distances all the coordinates are centered to 0 as is shown in figure **4-21**.

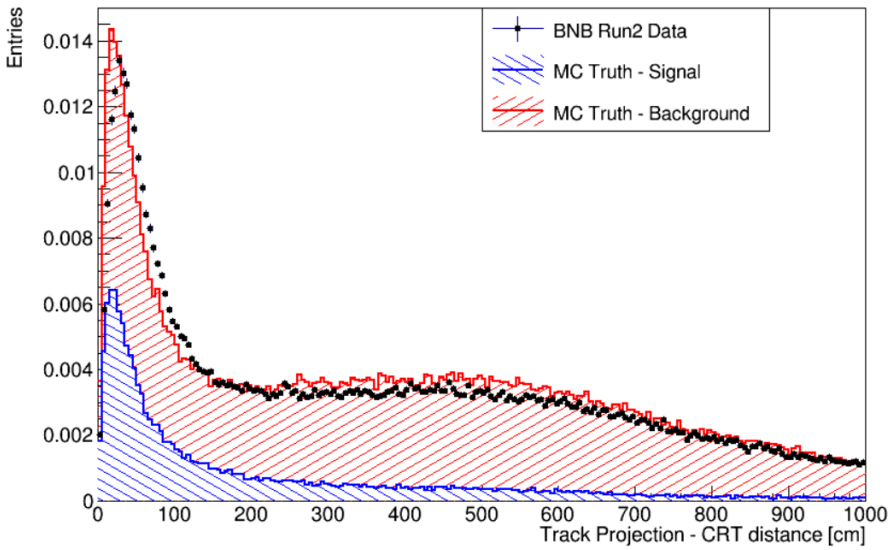

Figure 4-20: Distribution of the EDP for the data superimposed to the distributions of signal and background. The Data distribution is normalized to the number of triples in the data while the monte carlo one to the total number of simulated triples

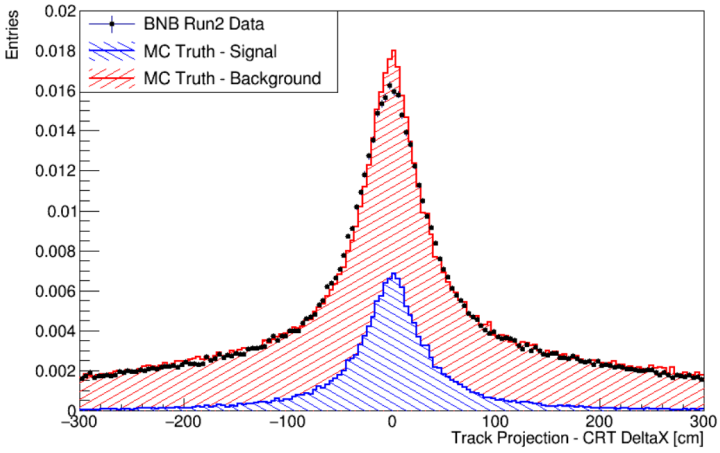

Figure 4-21: Distribution of the x coordinate distance on the top CRT for matched TPC tracks. Data and monte carlo distributions are area normalized

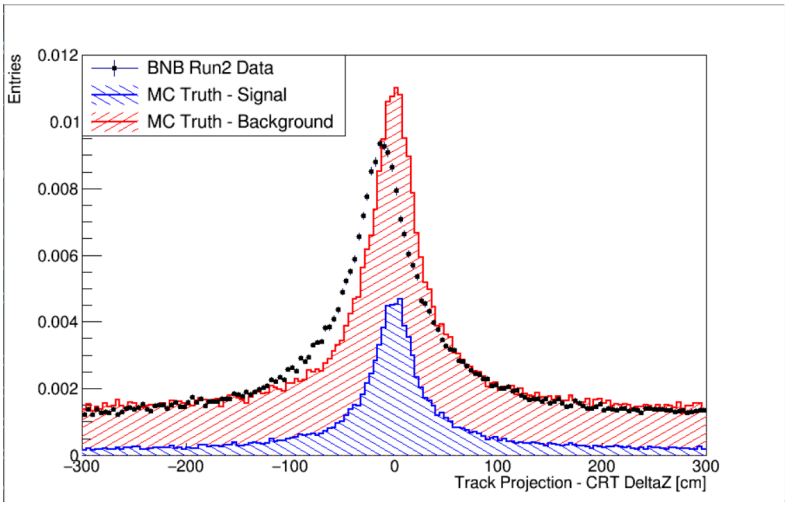

Figure 4-22: Distribution of the z coordinate distance on the top CRT for matched TPC tracks where is possible to notice the offset of the CRT modules for the data. Data and monte carlo distributions are area normalized

The distribution of the barycenter distances along the longitudinal axis presented in **4-23** shows that signal is mostly found within ± 1 m.Background on the other hand, although peaked around the same interval present additional values that make the distribution more populated on the sides, qualitatively resembling a gaussian distribution convoluted with the geometric properties of the detector. The distribution of the barycenter distances along the drift

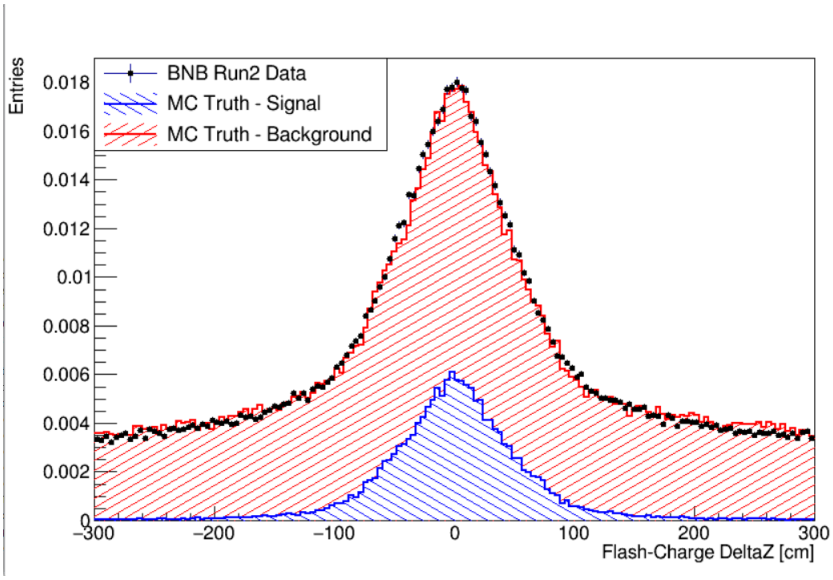

Figure 4-23: Distribution of the flash and charge barycenter distances in z for the matched triples. It is possible to see that the signal is peaked at 0 for smaller distances and that the closer the distance the more likely is that the triple corresponds to signal. Data and monte carlo distributions are area normalized

coordinate shows also a distribution that agrees with the results of the longitudinal coordinate, being one meter the interval for which the signal is localized, while the background is distributed around the whole interval considered. This behavior is shown in figure **4-24**. For this case the influence of the geometric properties of the detector are noticeable.

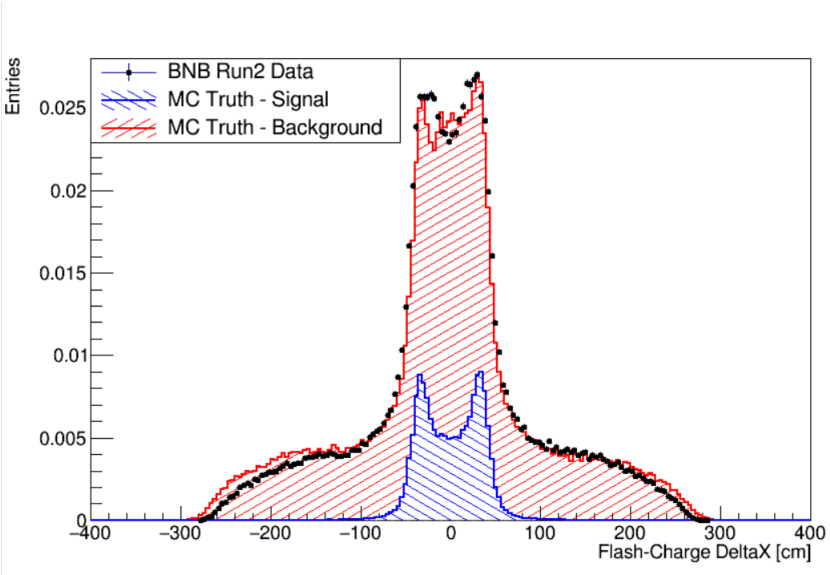

Figure 4-24: Distribution of the flash and charge barycenter distances in x for the matched triples. Data and monte carlo distributions are area normalized

4.3.1 Results and future prospects

The triple matching algorithm is then an efficient method of classification for out-of-time interactions reconstructed in ICARUS, and the BDT score can be used in future experimental analysis to assign a time to more than 90 % of the tracks reconstructed in the out-of-time window for a given set of events.

Further improvements in both the simulation of the events and the calibration of the detector will lead to a better data an monte carlo agreement, improving the BDT efficiency and purity, opening the door to a more comprehensive analysis on neutrino and cosmic events in LAr detectors.

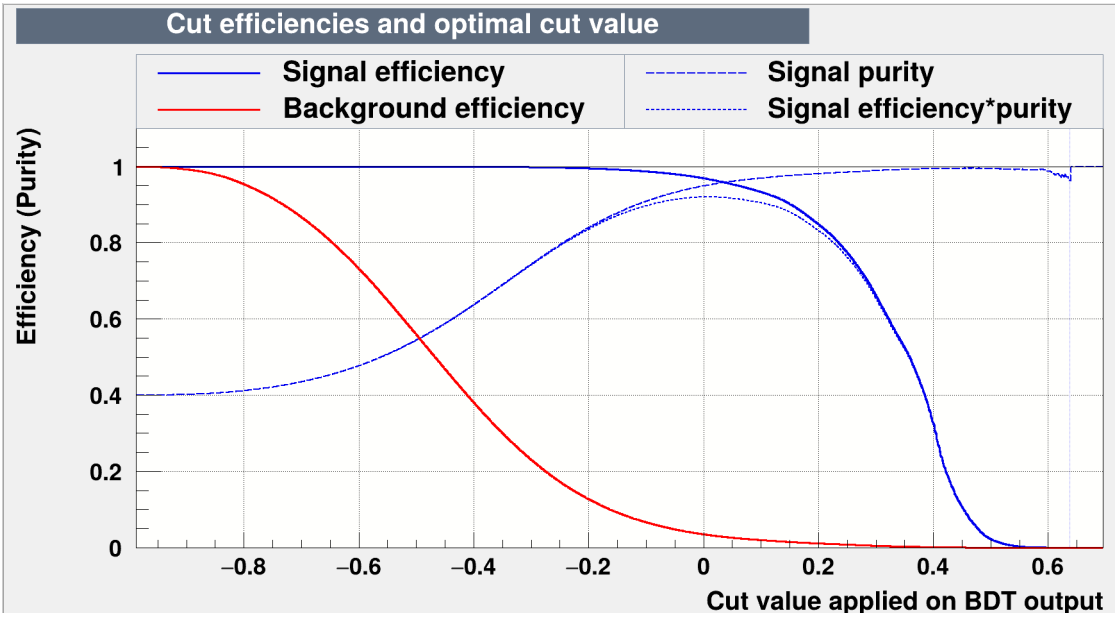

Figure 4-25: signal efficiency and signal purity for the BDT with respect of the score for a sample of 40000 signal triples events and 60000 background triple [63].

Conclusions

The ICARUS detector has been actively taking data on both the BNB and NuMI neutrino beams since June 2022.

This thesis work presented a multivariate analysis tool that for the first time combines the three different ICARUS subdetectors, exploiting each of their features in order to assign to each of the reconstructed track a precise timing information. The method chosen for the multivariate classification is a Boosted Decision Tree that determined the likeliness of a TPC track, an optical flash and a CRT hit being produced by the same particle.

The results of the algorithm show that on MC simulation an efficiency of 96.7$\pm$ 0.1 % and a purity of 95.1 $\pm$ 0.1% can be achieved. The algorithm was also applied to a large real data sample and, for the vast majority, data and MC distributions are compatible, suggesting the eventual integration of the algorithm into the ICARUS background rejection analysis.

The multivariate analysis can be improved by refinements on the Monte Carlo simulations or the calibration of the detectors. The first improvement is the inclusion of a more realistic rate of in-time cosmic rays and of the trigger efficiency of the PMT system. The second can be the re-alignment of the CRT modules.

Appendix A

CRT PMT Matches

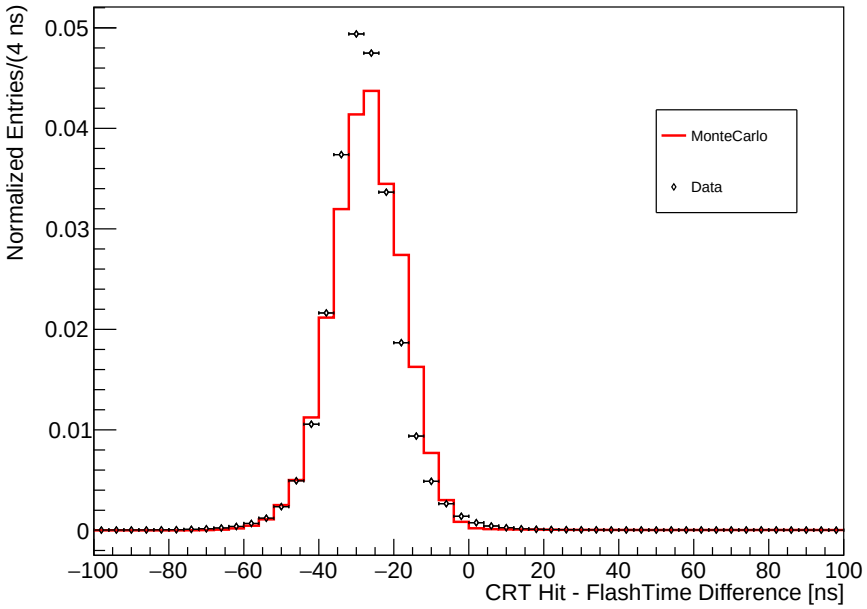

Figure 4-26: Comparison between top CRT hits from region 30 matched with optical flashes for Run 1 out-of-time data (black) and Monte Carlo out-of-time cosmics (red).

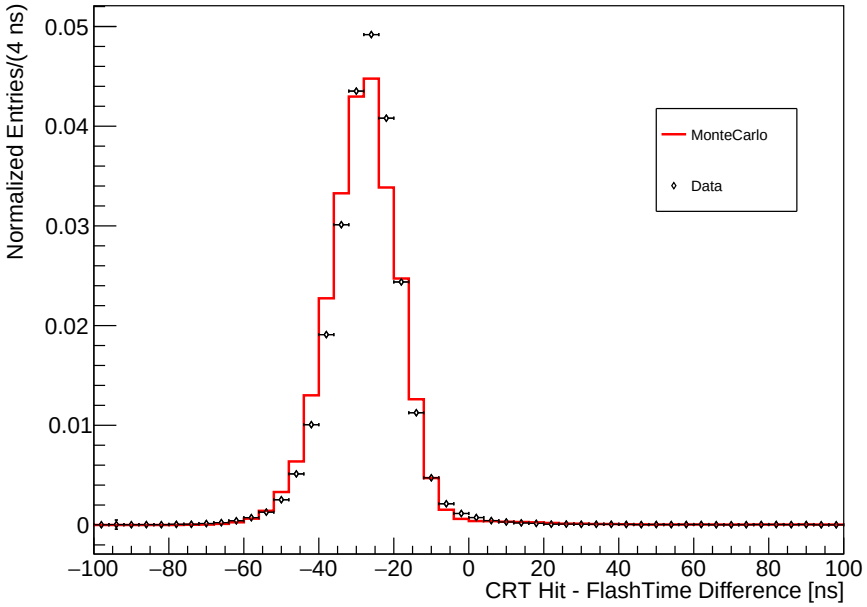

Figure 4-27: Comparison between top CRT hits from region 31 matched with optical flashes for Run 1 out-of-time data (black) and Monte Carlo out-of-time cosmics (red).

BDT outputs

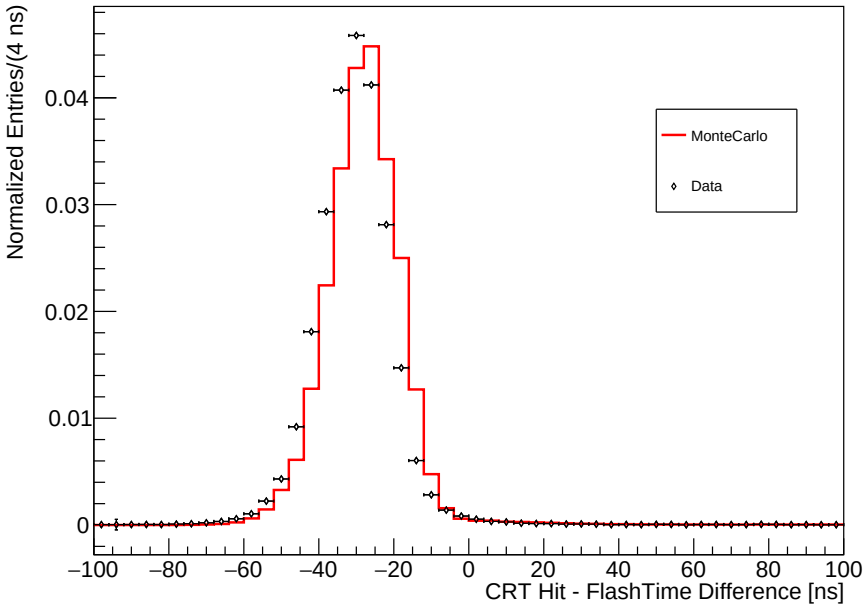

Figure 4-28: Comparison between top CRT hits from region 32 matched with optical flashes for Run 1 out-of-time data (black) and Monte Carlo out-of-time cosmics (red).

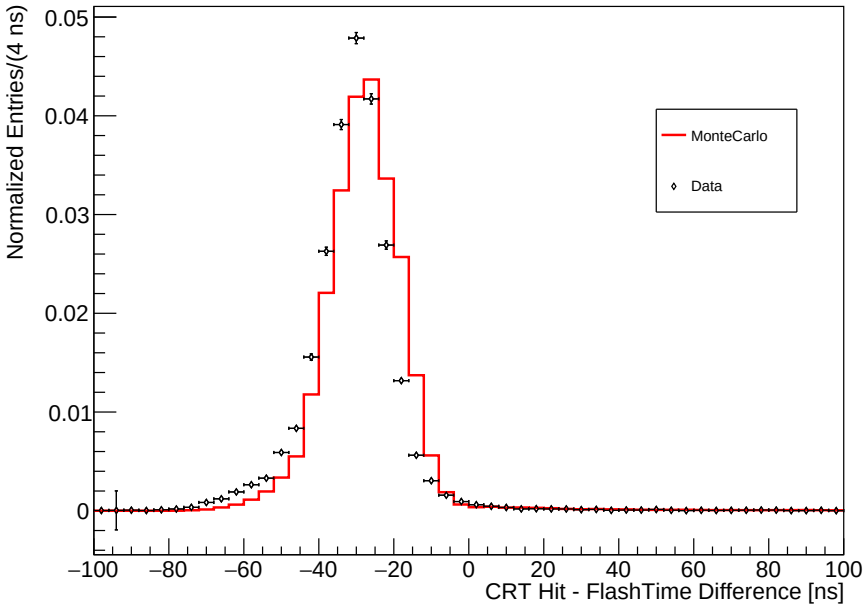

Figure 4-29: Comparison between top CRT hits from region 33 matched with optical flashes for Run 1 out-of-time data (black) and Monte Carlo out-of-time cosmics (red).

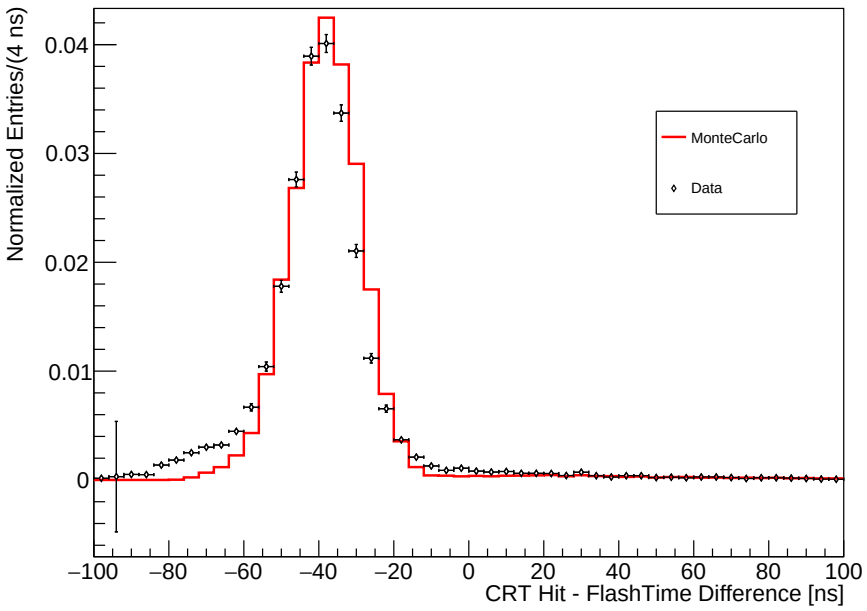

Figure 4-30: Comparison between top CRT hits from region 34 matched with optical flashes for Run 1 out-of-time data (black) and Monte Carlo out-of-time cosmics (red).

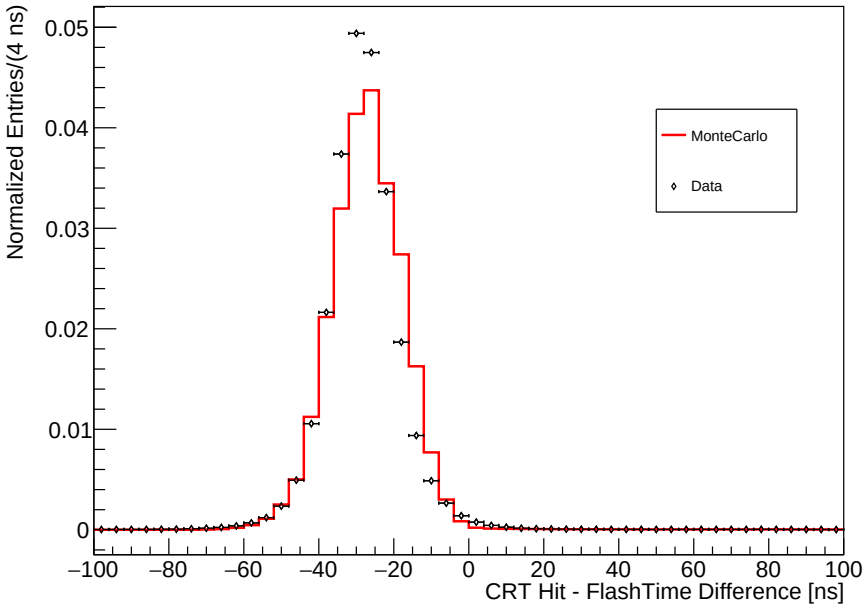

Figure 4-31: Comparison between top CRT hits from region 30 matched with optical flashes for Run 1 out-of-time data (black) and Monte Carlo out-of-time cosmics (red).

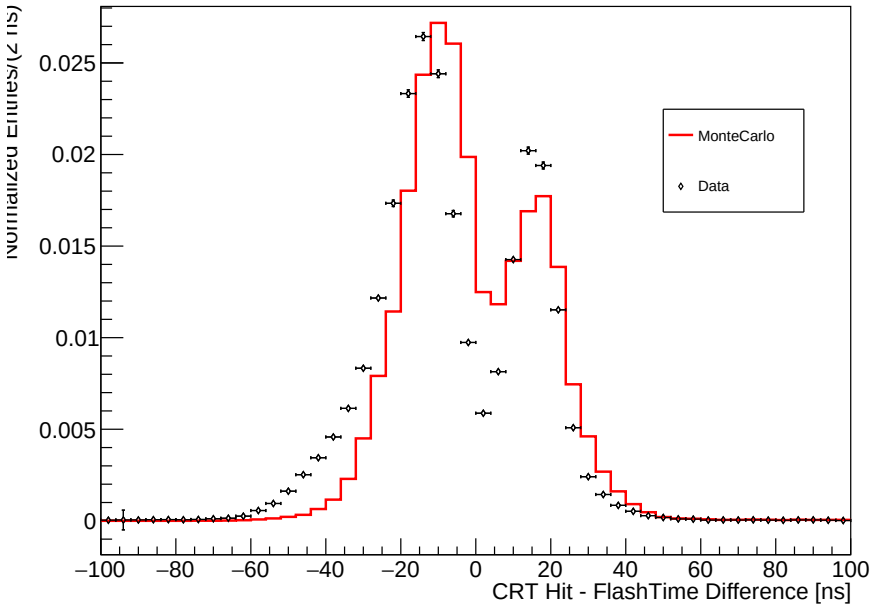

Figure 4-32: Comparison between Side CRT hit from region 40 matched with optical flashes for Run 1 data (black) and Monte Carlo out-of-time cosmics (red).

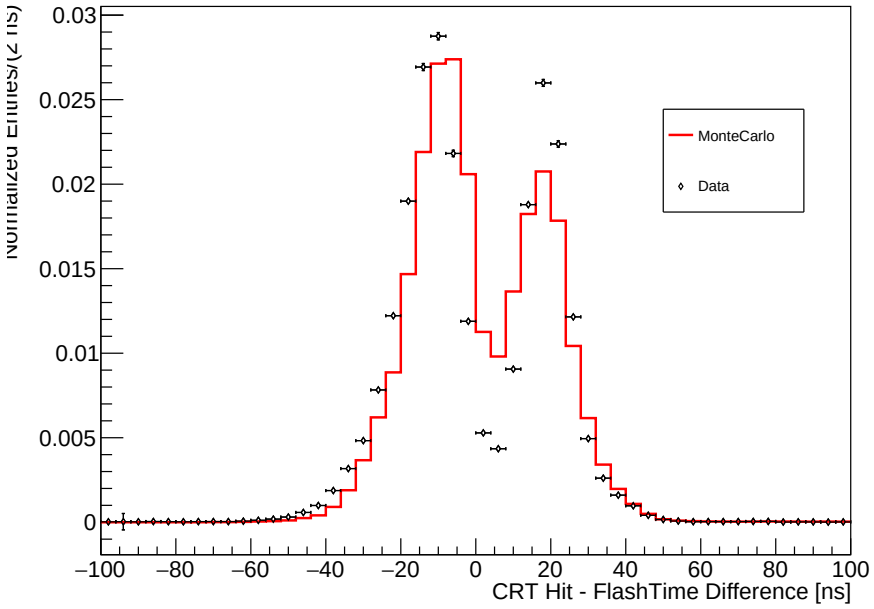

Figure 4-33: Comparison between Side CRT hit from region 42 matched with optical flashes for Run 1 data (black) and Monte Carlo out-of-time cosmics (red).

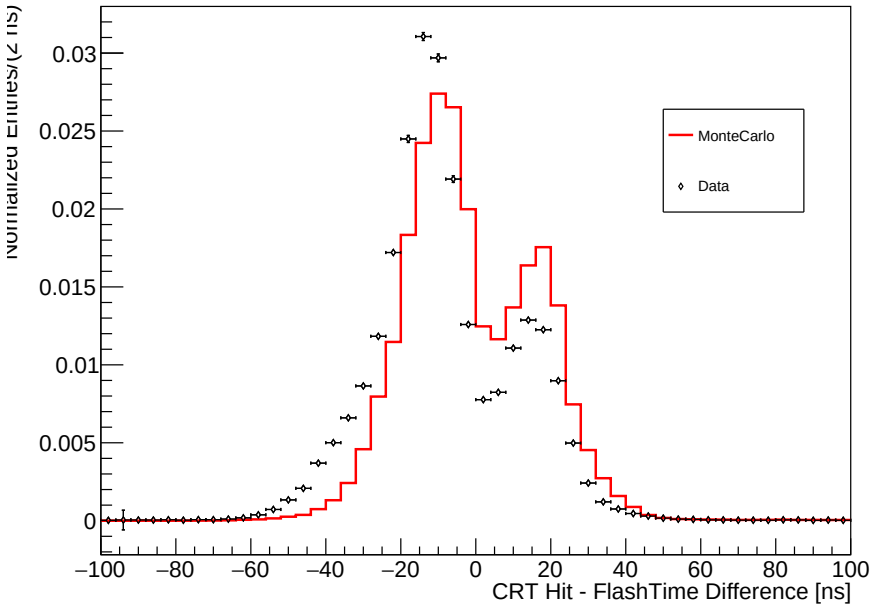

Figure 4-34: Comparison between Side CRT hit from region 42 matched with optical flashes for Run 1 data (black) and Monte Carlo out-of-time cosmics (red).

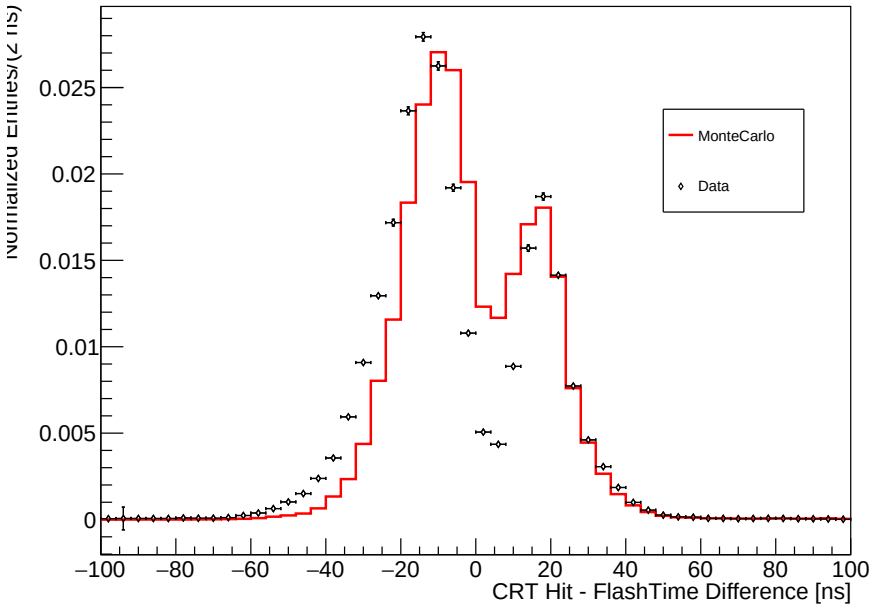

Figure 4-35: Comparison between Side CRT hit from region 43 matched with optical flashes for Run 1 data (black) and Monte Carlo out-of-time cosmics (red).

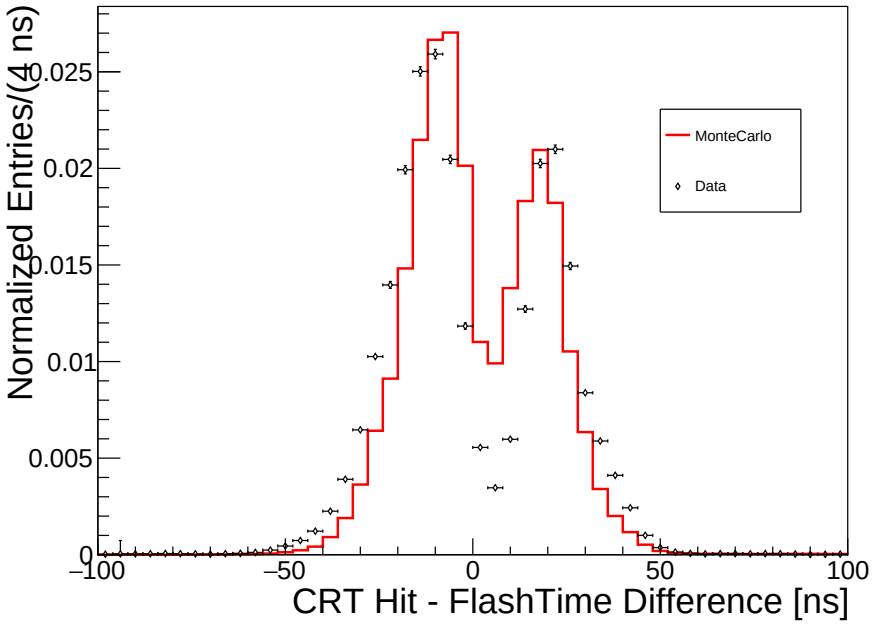

Figure 4-36: Comparison between Side CRT hit from region 44 matched with optical flashes for Run 1 data (black) and Monte Carlo out-of-time cosmics (red).

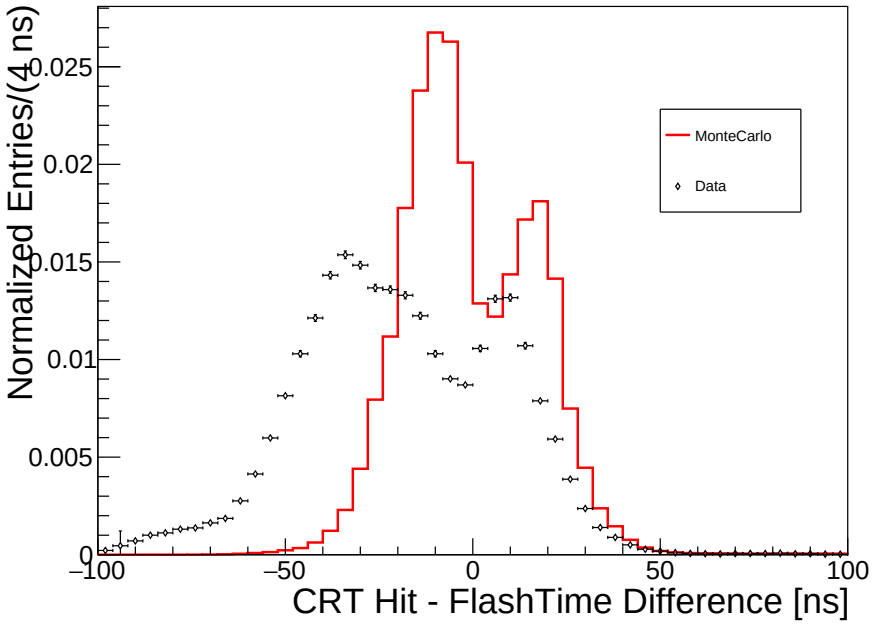

Figure 4-37: Comparison between Side CRT hit from region 45 matched with optical flashes for Run 1 data (black) and Monte Carlo out-of-time cosmics (red).

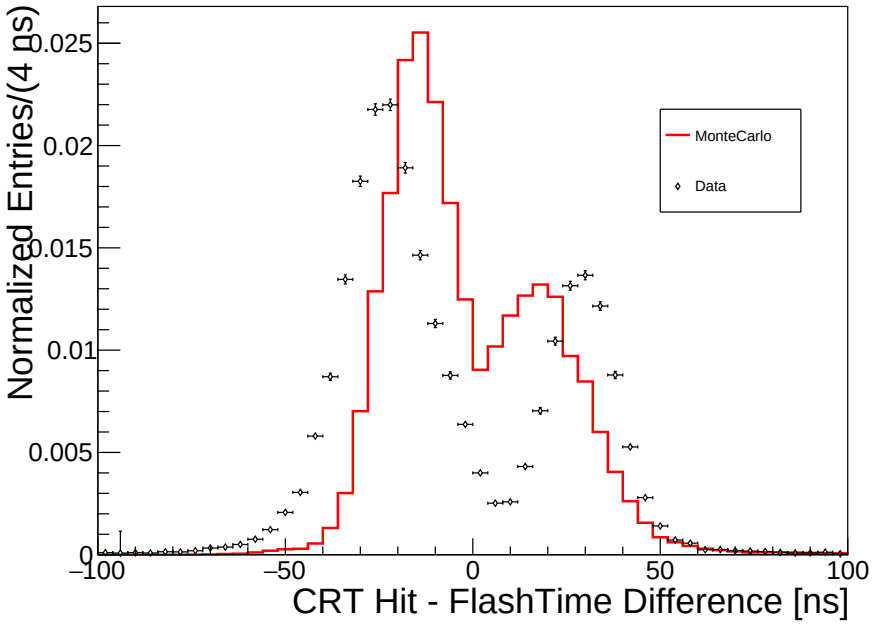

Figure 4-38: Comparison between Side CRT hit from region 46 matched with optical flashes for Run 1 data (black) and Monte Carlo out-of-time cosmics (red).

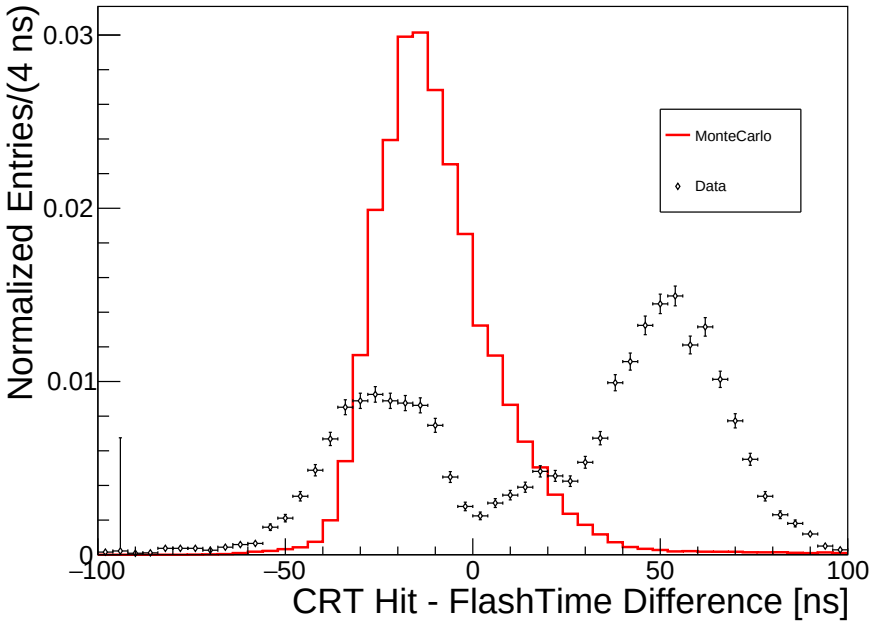

Figure 4-39: Comparison between Side CRT hit from region 47 matched with optical flashes for Run 1 data (black) and Monte Carlo out-of-time cosmics (red).

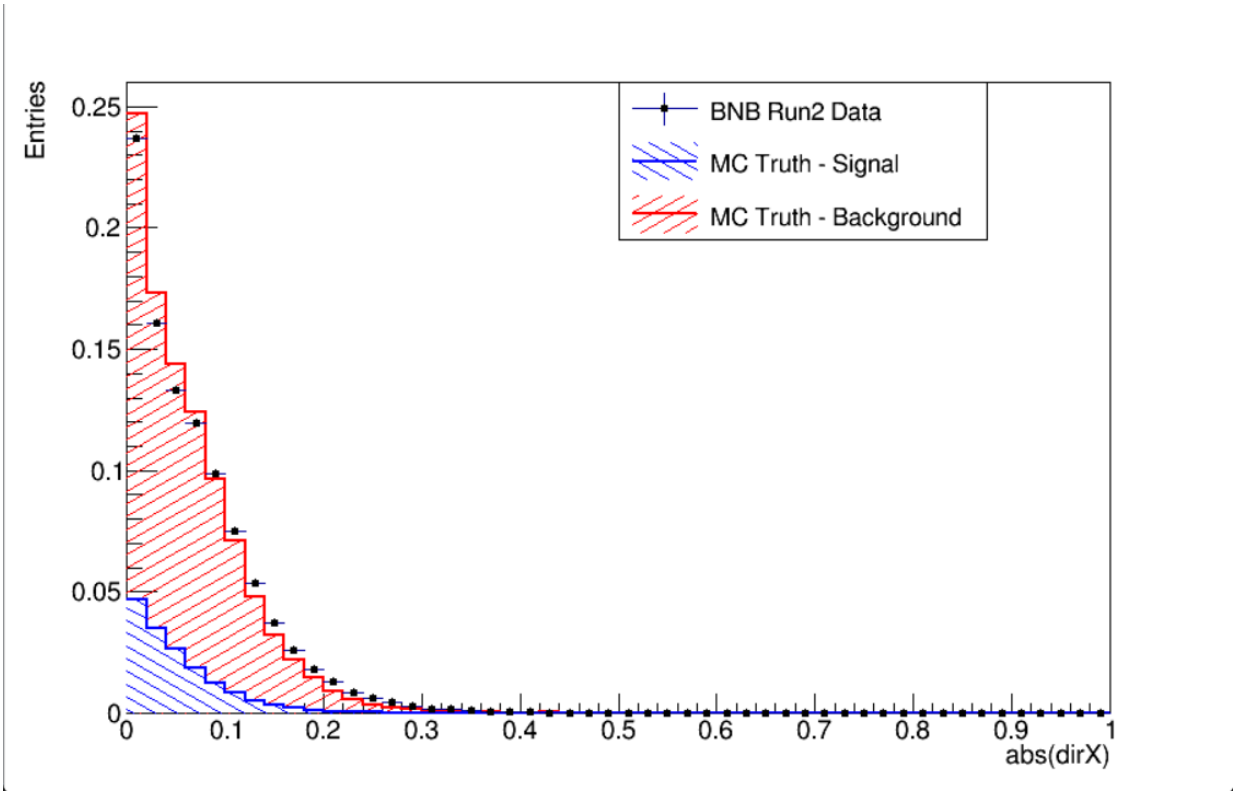
Entries
0.25
0.2
0.15
0.1
0.05
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
abs(dirX)
BNB Run2 Data
MC Truth - Signal
MC Truth - Background

Figure 4-40

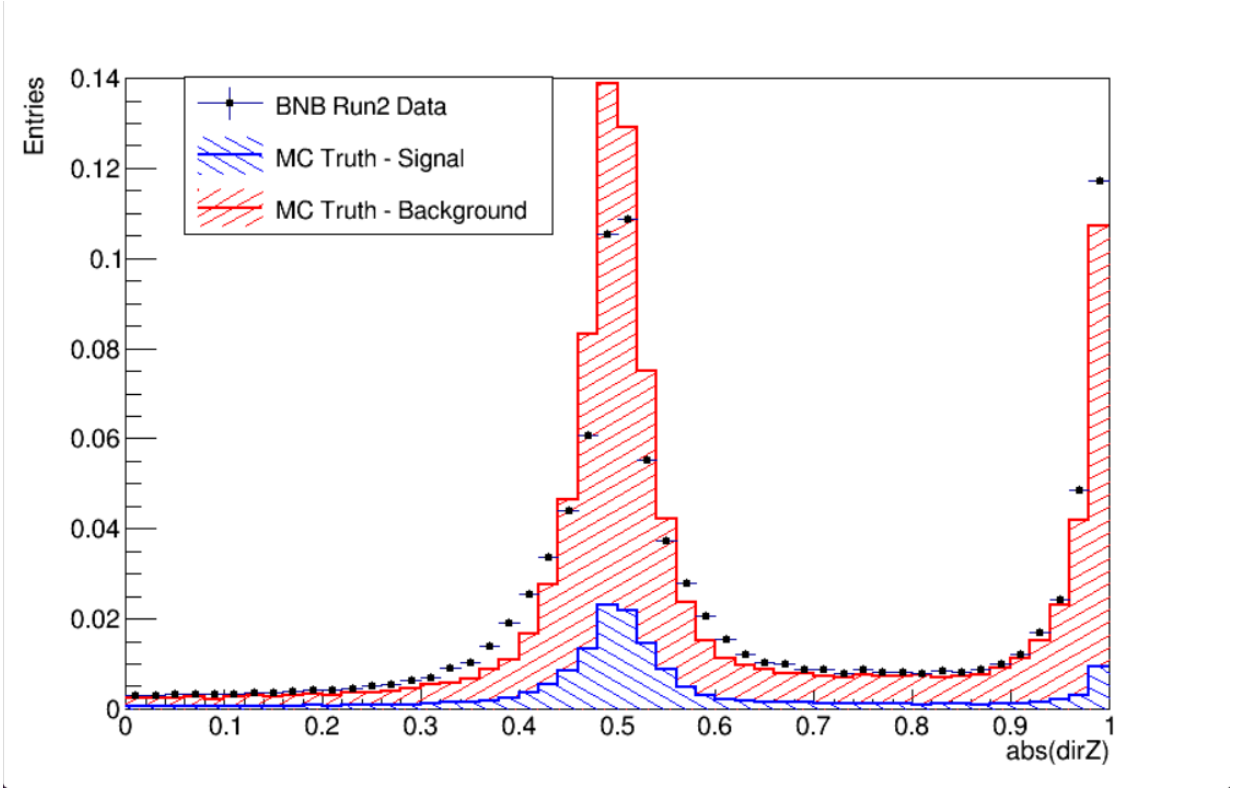
Entries
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
BNB Run2 Data
MC Truth - Signal
MC Truth - Background
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
abs(dirZ)

Figure 4-41

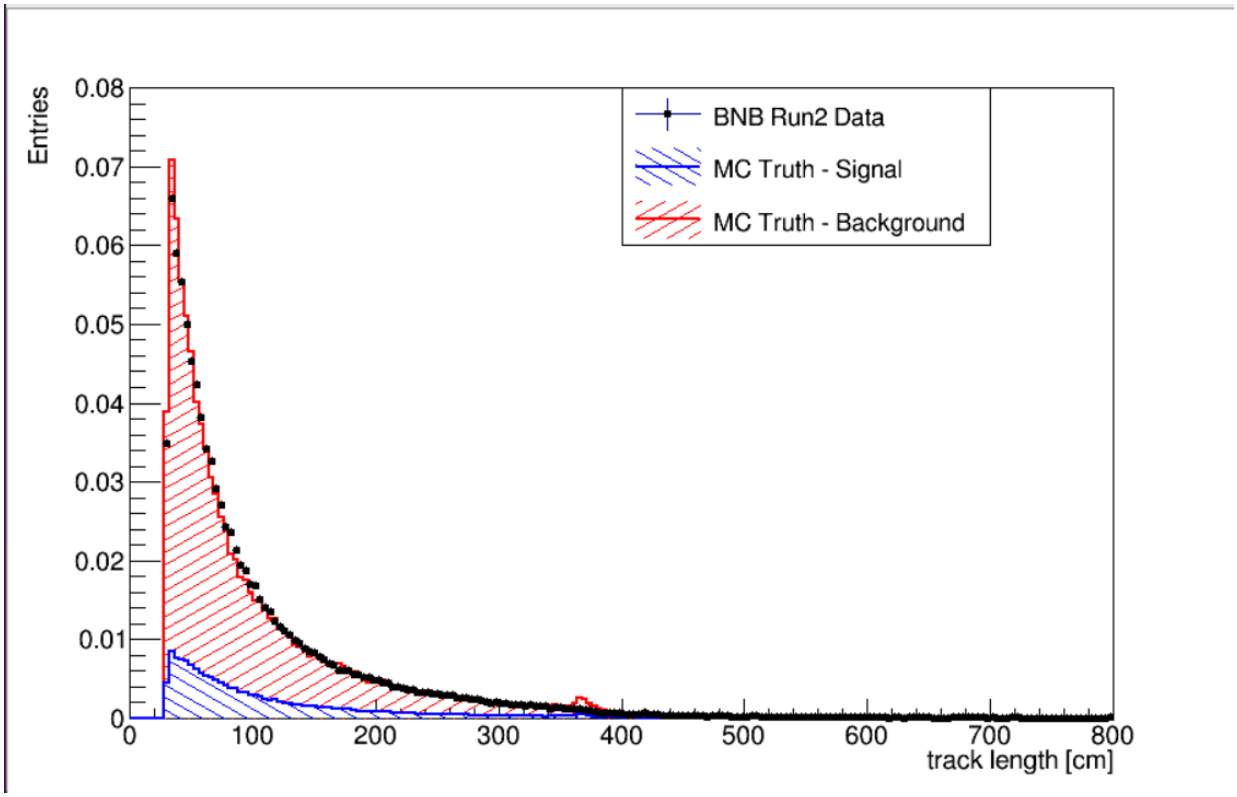
Entries
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
0
100
200
300
400
500
600
700
800
track length [cm]
BNB Run2 Data
MC Truth - Signal
MC Truth - Background

Figure 4-42

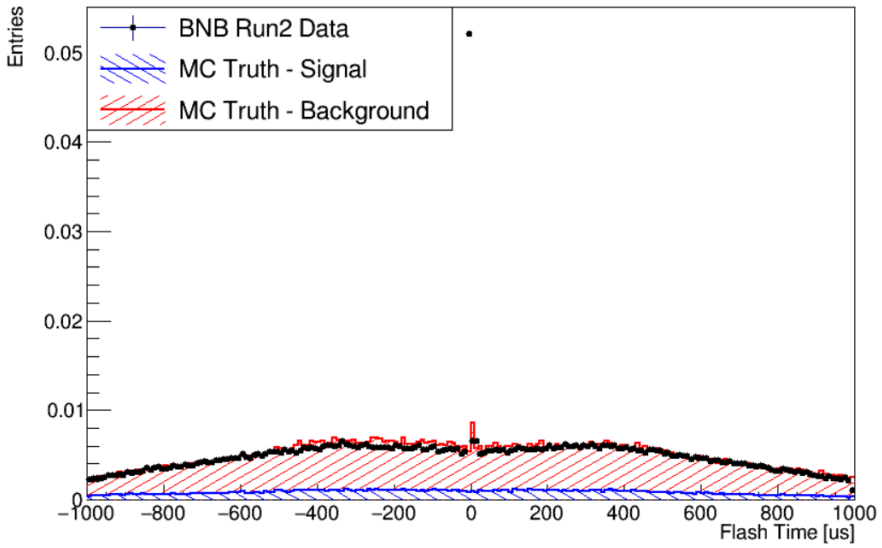

Figure 4-43

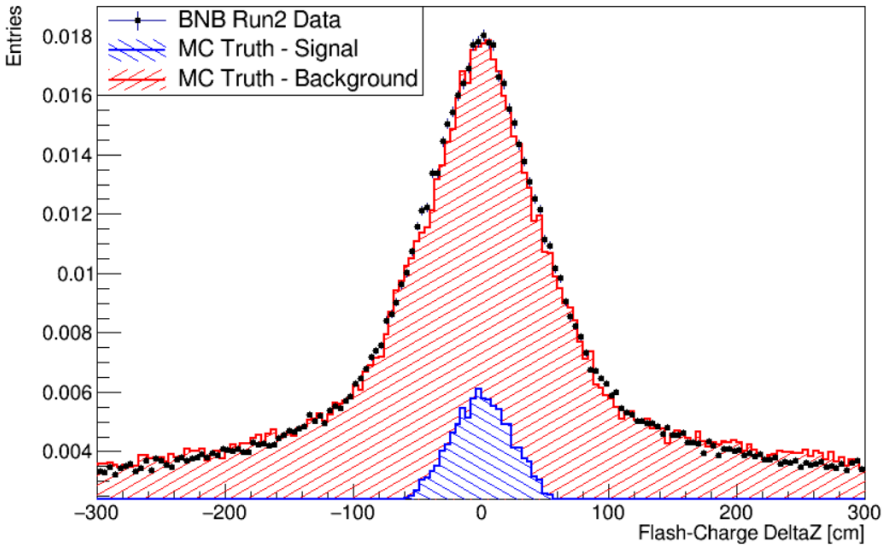

Figure 4-44

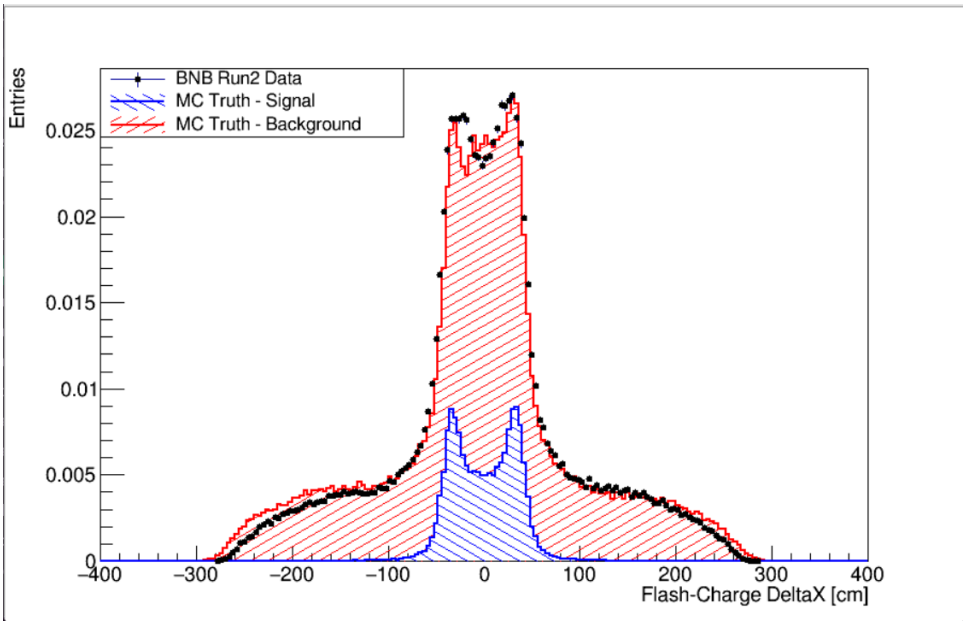

Figure 4-45

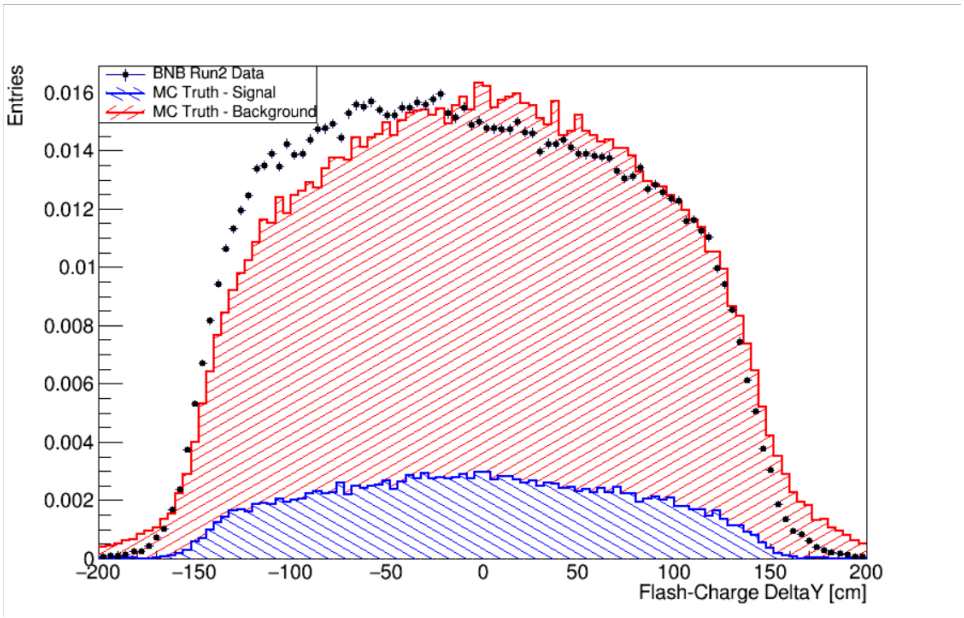

Figure 4-46

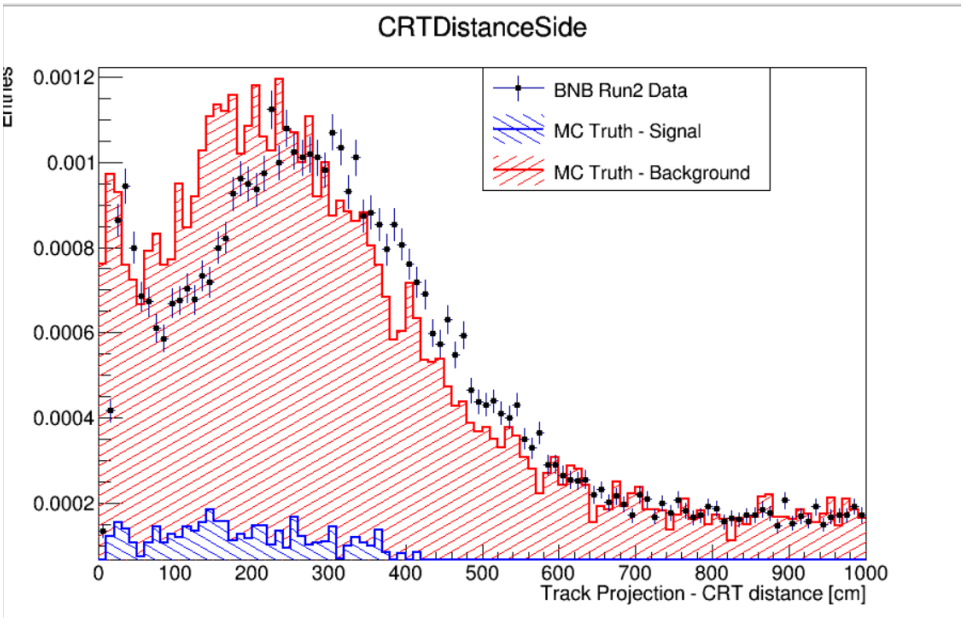

Figure 4-47

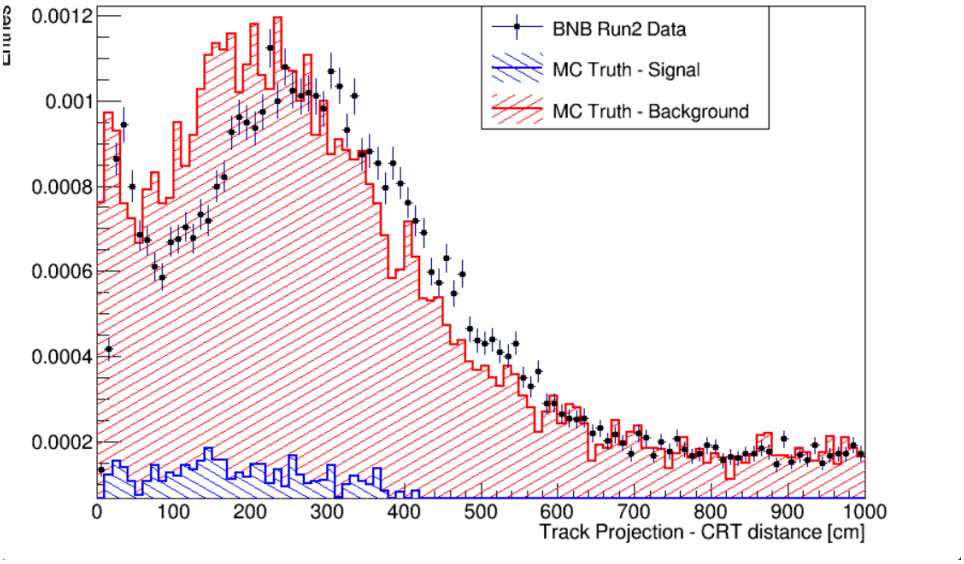

Figure 4-48

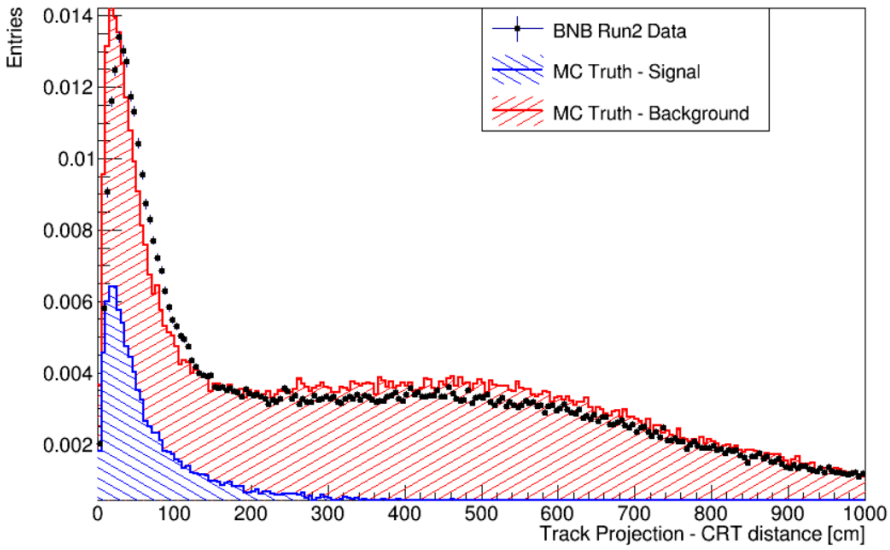

Figure 4-49

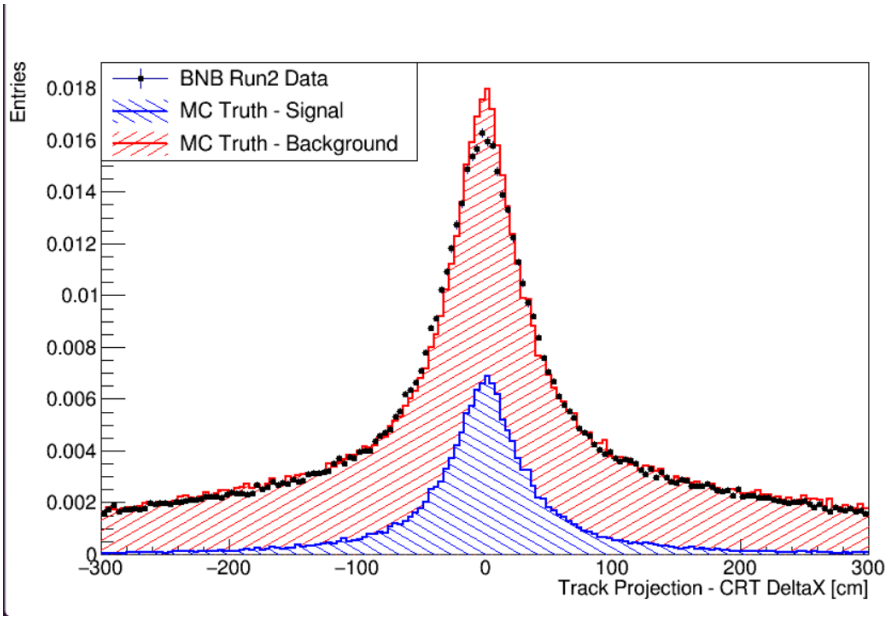

Figure 4-50

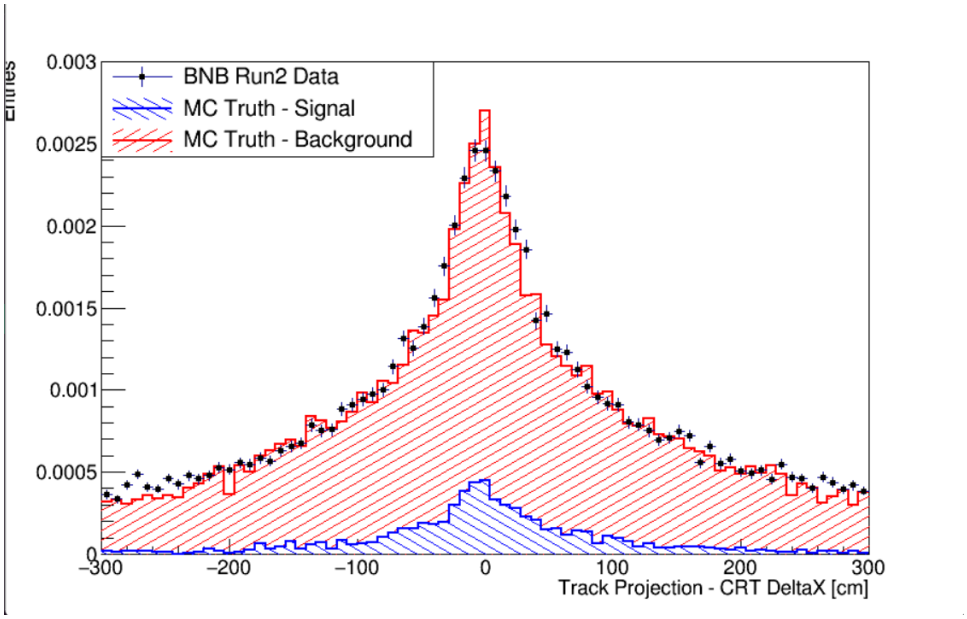

Figure 4-51

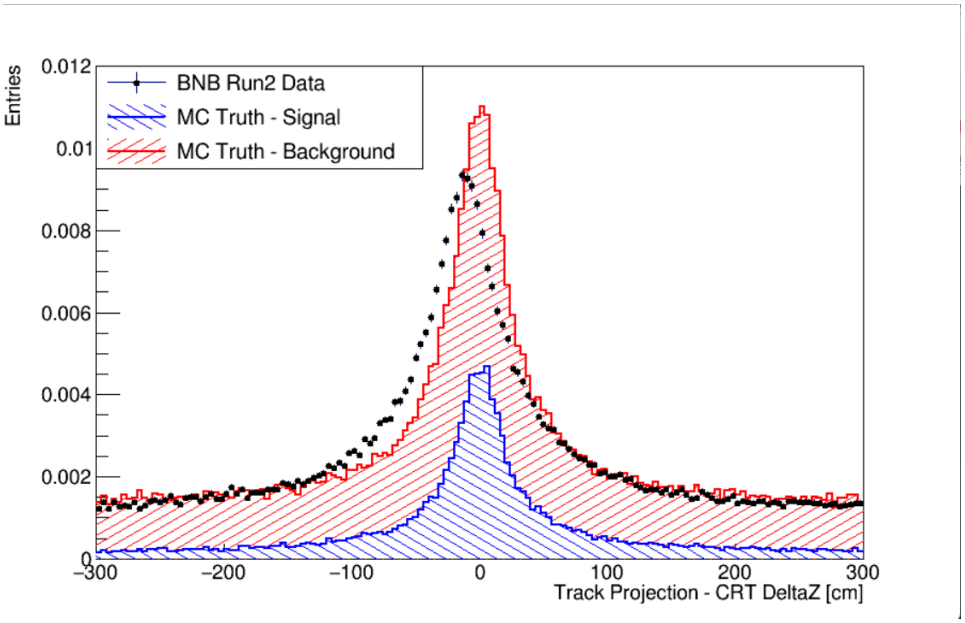

Figure 4-52

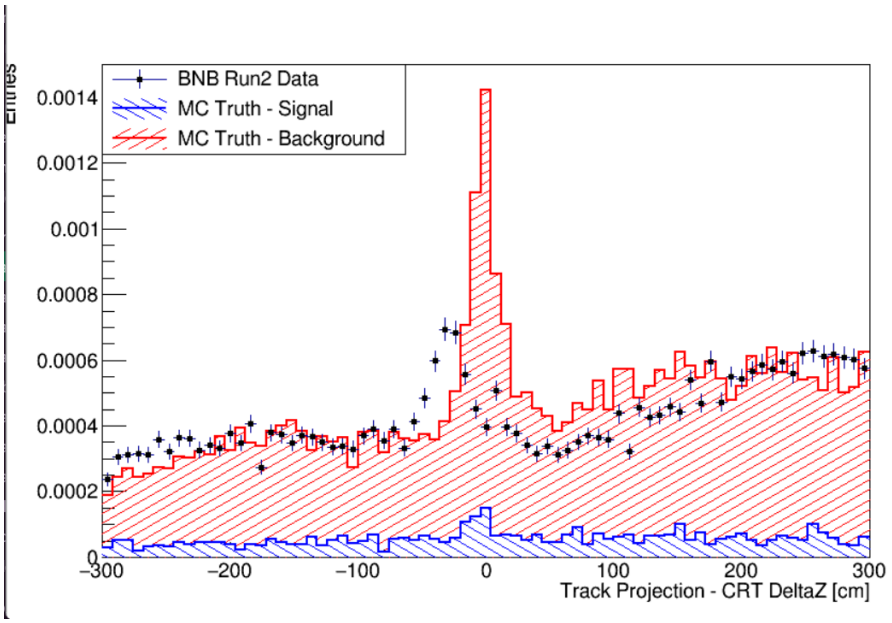

Figure 4-53

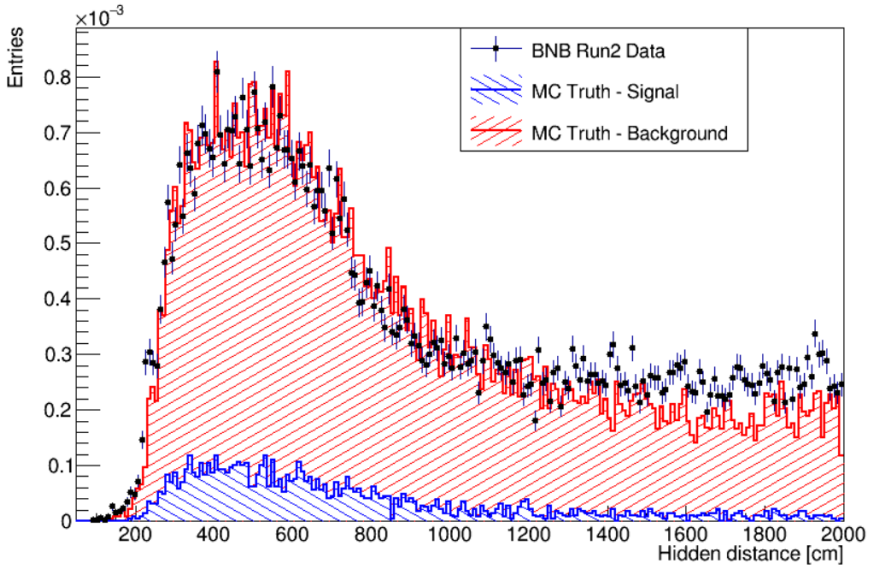

Figure 4-54

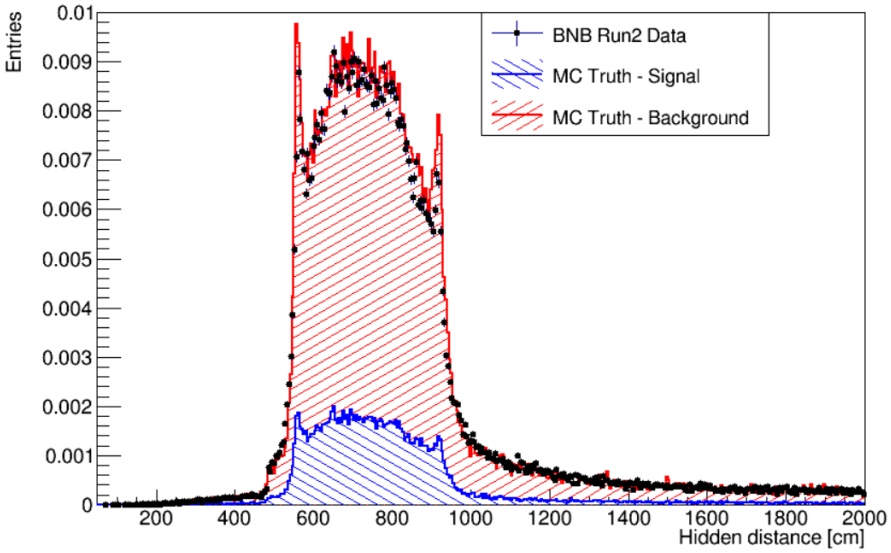

Figure 4-55

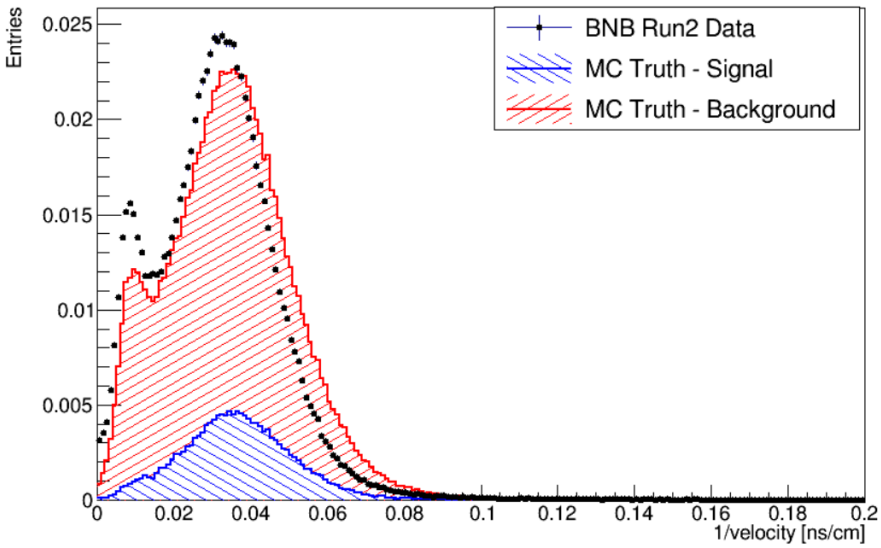

Figure 4-56

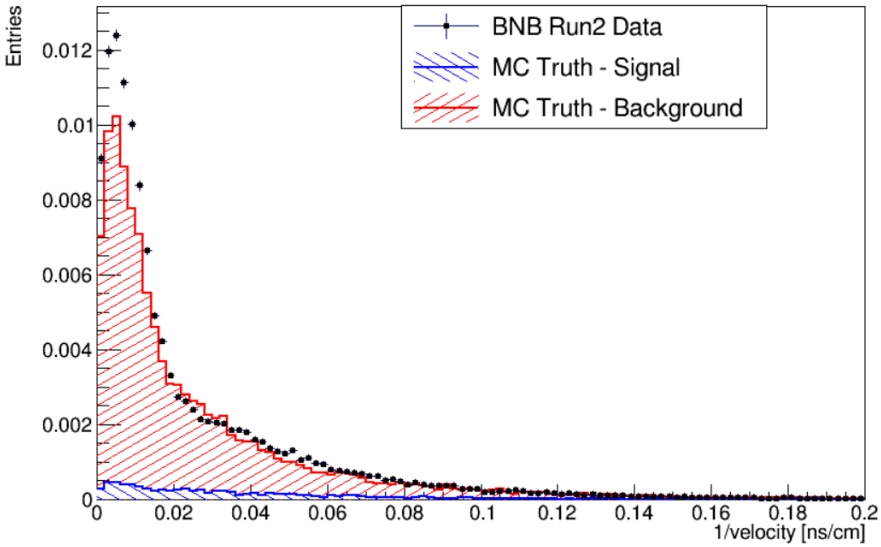

Figure 4-57

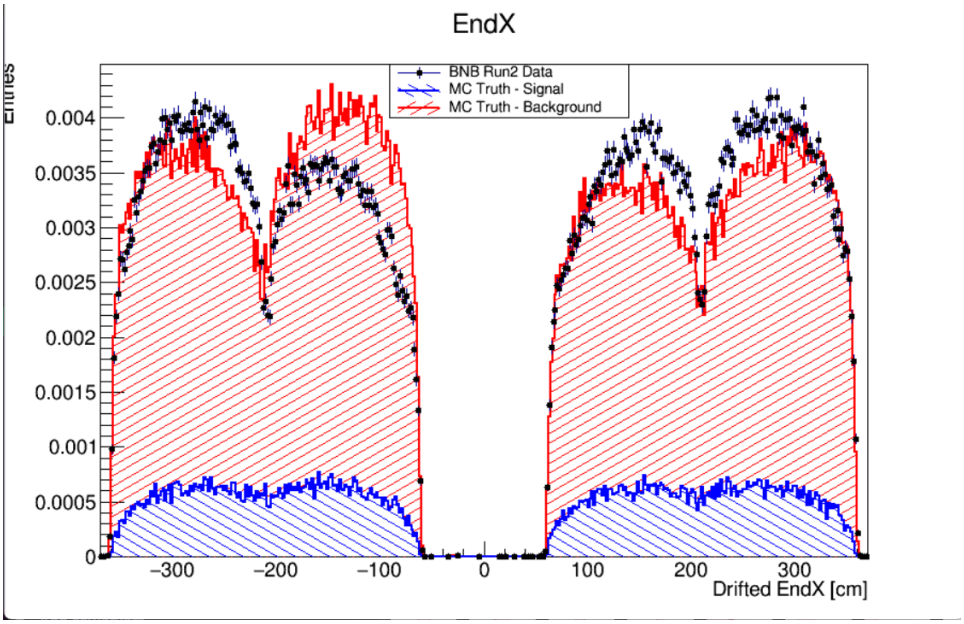

Figure 4-58

Bibliography

1. Thurner, F. & Mann, U. Measurement of the Rate of $\nu_e + d \rightarrow p + p + e$ interactions produced by solar neutrinos at the Sudbury Neutrino Observatory. *Phys.Rev.Lett* **87,** 1–6 (2001).
2. Coadou, Y. Number of neutrino families from LEP1 measurements. *Artificial Intelligence for High Energy Physics* (2022).
3. Giunti, C. & Kim, C. W. *Fundamentals of Neutrino Physics and Astrophysics* (Oxford university press, London, 2007).
4. D.Waller & the SNO collaboration. Results from the Sudbury Neutrino Observatory. *SLAC Summer Institute on Particle Physics* (2004).
5. Suzuki, Y. The super-kamiokande experiment. *European Physical Journal C* **79** (2019).
6. Alimonti, G. & et al. The borexino detector at the laboratori nazionali del gran sasso. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment,* **600,** 568–593 (2009).
7. Allison, W. W. M. & et al. Measurement of the atmospheric neutrino flavor composition in Soudan-2. *Phys. Lett* **B391,** 491–500 (1997).
8. Giacomelli, G. Neutrino physics and astrophysics with the MACRO experiment at the Gran Sasso lab. *Braz. J. Phys.* **33,** 211–217 (2003).
9. F.An & et al. Measurement of electron antineutrino oscillation based on 1230 days of operation of the daya bay experiment. *Phys. Rev. D* **95** (2017).
10. Kim, S. New results from reno and prospects with reno-50. *Nuclear and Particle Physics Proceedings* **265,** 93–98 (2015).

11. Crespo-Anadón, J. Double Chooz: Latest results. *Nucl. Part. Phys. Proc* **265,** 99–104 (2015).
12. Decowsk, M. & the KAMLAND collaboration. KamLAND's precision neutrino oscillation measurements. *Nuc. Phys B* **908,** 52–61 (2016).
13. hernandez, P. *Neutrino Physics,Proceedings of the 2015 CERN–Latin-American School of High-Energy Physics* 2016.
14. Observatory, J. U. N. *Introduction to JUNO* 2023.
15. Sioli, M. *Flavour physics in the lepton sector* 2022.
16. Boyd, S. "Recent results from the k2k (kek-to-kamioka) neutrino oscillation experiment. *Nuclear Physics B - Proceedings Supplements* **98,** 175–181 (2001).
17. Evans, J. The MINOS Experiment: Results and Prospects. *Adv. High Energy Phys,* 182–537 (2013).
18. Agafonova, N. & et al. Final Results of the OPERA Experiment on Appearance in the CNGS Neutrino Beam. *Physical Review Letters* **120** (2018).
19. Vannucc, F. The nomad experiment at cern. *Advances in High Energy Physics,* 1–20 (2014).
20. Cocco, A. G. Results from CHORUS experiment at CERN. *Phys. Rept* **307,** 319–324 (1998).
21. Letizia, P. Search for heavy neutrinos at the na48/2 and na62 experiments at cern. *EPJ Web Conf* **182** (2018).
22. Abazajian, K. N. & et al. Light sterile neutrinos: A white paper. *https://doi.org/10.48550/ARXIV.1* (2012).
23. Giunti, C. & et al. Reactor antineutrino anomaly in light of recent flux model refinements. *Physics Letters B* **829** (2022).
24. Hampel, W. & et al. GALLEX solar neutrino observations: Results for GALLEX IV. *Physics Letters B* **447,** 127–133 (1999).
25. Abdurashitov, J. N. & et al. Measurement of the response of a gallium metal solar neutrino experiment to neutrinos from a 51Cr source. *Phys rev C* **59,** 2246–2263 (1999).
26. M.Acero. & et al. White paper on light sterile neutrino searches and related phenomenology. *ArXiv* (2022).

27. Serebrov, A. & et al. Search for sterile neutrinos with the neutrino-4 experiment and measurement results. *Physical Review D* **104** (2021).
28. Aguilar, A. & et al. Evidence for neutrino oscillations from the observation of $\bar{\nu}_e$ appearance in a $\bar{\nu}_\mu$beam. *Physical Review D* **64** (2001).
29. Aguilar-Arevalo, A. A. & et al. Neutrino flux prediction at MiniBooNE. *Physical Review D* **79** (2009).
30. Dentler, M. & et al. Updated global analysis of neutrino oscillations in the presence of eV-scale sterile neutrinos. *Journal of High Energy Physics* **8** (2018).
31. Rubbia, C. The liquid-argon time projection chamber: a new concept for neutrino detectors. *CERN, Geneva, Tech. Rep* (1977).
32. Antonello, M. & et al. A Proposal for a Three Detector Short-Baseline Neutrino Oscillation Program in the Fermilab Booster Neutrino Beam. *arXiv:1503.01520 [physics.ins-det]* (2015).
33. Shiltsev, V. Fermilab Proton Accelerator Complex Status and Improvement Plans. *Mod. Phys. Lett. A* **32** (2017).
34. Adamson, P. & et al. The NuMI Neutrino Beam. *Nucl. Instrum. Meth.A* **806,** 279–306 (2016).
35. Machado, A. & Segreto, E. Arapuca a new device for liquid argon scintillation light detection. *Journal of Instrumentation* **11** (2016).
36. Machado, A. A. & et al. The X-ARAPUCA: An improvement of the ARAPUCA device. *arXiv: 1804.01407 [physics.ins-det],* 279–306 (2018).
37. Acciarri, R. & et al. Design and construction of the MicroBooNE detector. *Journal of Instrumentation* **12** (2017).
38. P.Abratenko & et al. Search for an anomalous excess of charged-current quasielastic ν_e interactions with the microboone experiment using deeplearning-based reconstruction. *Phys. Rev D* **105** (2022).
39. P.Abratenko & et al. Search for neutrino-induced neutral-current Δ radiative decay in microboone and a first test of the miniboone low energy excess under a single-photon hypothesis. *Phys. Rev. Lett* **128** (2022).
40. Abratenko, P. & et al. First Constraints on Light Sterile Neutrino Oscillations from Combined Appearance and Disappearance Searches with the MicroBooNE Detector. *Phys. Rev. Lett* **130** (2023).

41. Arneodo, F. & et al. Performance Of A Liquid Argon Time Projection Chamber Exposed To The WANF Neutrino Beam. *Phys. Rev.D* **74** (2006).
42. Antonello, M. & et al. Operation and performance of the icarus t600 cryogenic plant at gran sasso underground laboratory. *Journal of Instrumentation* **10** (2015).
43. The ICARUS collaboration. Cosmogenic background suppression at the icarus using a concrete overburden. *Journal of Physics: Conference Series* **2156** (2021).
44. Poppi, F. *THE COSMIC RAY TAGGER OF THE ICARUS DETECTOR: CONSTRUCTION, COMMISSIONING AND PERFORMANCES IN THE FIRST PHYSICS RUN* Phd Thesis (Univesità di Bologna, Bologna, 2023).
45. Donatella Torretta, J. K. *ICARUS Operation report – Dec 2022* 2022.
46. C.Rubbia. The Liquid Argon Time Projection Chamber: A new concept for neutrino detectors. *EP international report* **08,** 1–14 (1977).
47. Collaboration, T. I. Overhaul and Installation of the ICARUS-T600 Liquid Argon TPC Electronics for the FNAL Short Baseline Neutrino Program. *Journal of instrumentation* **16,** 1–23 (2021).
48. Collaboration, T. I. Design, construction and tests of the ICARUS T600 detector. *Nuclear Instruments and Methods in Physics Research A* **527,** 329–410 (2004).
49. Collaboration, T. I. New read-out electronics for ICARUS-T600 liquid Argon TPC. Description, simulation and tests of the new front-end and ADC system. *Journal of Instrumentation* (2018).
50. Collaboration, T. I. Precise 3D Track Reconstruction Algorithm for the ICARUS T600 Liquid Argon Time Projection Chamber Detector. *Advances in High Energy Physics* **2013,** 1–16 (2013).
51. Kisiel, J. Present status of the ICARUS experiment. *Acta Physica Polonica B* **36,** 3227–3223 (2005).
52. Marshall, J. & Thomson, M. The pandora software development kit for pattern recognition. *The European Physical Journal C* **75,** 1–16 (2015).
53. Acciarri, R. & et al. The Pandora multi-algorithm approach to automated pattern recognition of cosmic-ray muon and neutrino events in

the MicroBooNE detector. *The European Physical Journal C* **78,** 1–25 (2018).

54. Abud, A. A. & et al. Reconstruction of interactions in the protodune-sp detector with pandora. *The European Physical Journal C* **83** (2023).
55. Artero Pons, M. *Study of the reconstruction of ν_μ CC QE events from the booster neutrino beam with the ICARUS detector* Phd Thesis (Univesità degli studi di Padova, Padova, 2024).
56. Ali-Mohammadzadeh, B. & et al. Design and implementation of the new scintillation light detection system of ICARUS T600. *Journal of Instrumentation* **15** (2020).
57. De Kerret, H. & et al. The double chooz antineutrino detectors. *The European Physical Journal C* **82** (2022).
58. Michael, D. & et al. The magnetized steel and scintillator calorimeters of the minos experiment. *Nuclear Instruments and Methods in Physics Research Section A* **596,** 190–228 (2008).
59. Heggestuen, A. Light detection and Cosmic Rejection in the ICARUS LArTPC at Fermilab. *Journal of Instrumentation* **19** (2024).
60. The LArsoft collaboration, E. *The Liquid Argon Software (LArSoft)* `Larsoft.org` [Accessed: 17/06/2024]. 2024.
61. The ICARUS collaboration, E. ICARUS at the Fermilab Short-Baseline Neutrino Program - Initial Operation. *https://arxiv.org/pdf/2301.08634* **1,** 1–33 (2023).
62. Chemarin, M. Boosted decision trees. *Nuc. Phys. B,* 67–71 (2000).
63. The TMVA collaboration, C. *TMVA - Toolkit for Multivariate Data Analysis* `https://tmva.sourceforge.net` [Accessed: 21/06/2024]. 2007.
64. Wouter Verkerke, D. P. K. *The RooFit toolkit for data modeling* `https://root.cern/manual/roofit/` [Accessed: 23/06/2024]. 2003.