University of Padova

Department of Physics and Astronomy "Galileo Galilei"

Bachelor Thesis in Physics

Collaudo dell'elettronica di readout del rivelatore OSIRIS dell'esperimento JUNO

Commissioning of the JUNO sub-detector OSIRIS readout electronics

Supervisor
Prof. Alberto Garfagnini
University of Padova

Co-supervisor
Dr. Andrea Serafini
University of Padova

Bachelor Candidate
Lorenzo Vincenzo D'Auria

Student ID
2000365

Academic Year
2022-2023

Abstract

The Jiangmen Underground Neutrino Observatory (JUNO) will be the largest liquid scintillator (LS) based neutrino detector in the World, for the next decade. Thanks to its very large active mass (20 kton) and state of the art performances (3% effective energy resolution at 1 MeV), it will be able to perform important measurements in neutrino physics. The radio-purity of the liquid scintillator is a crucial ingredient for most of the measurements that JUNO will perform. The Online Scintillator Internal Radioactivity Investigation System (OSIRIS) is a stand-alone detector designed to monitor the radiopurity of the LS while the JUNO Central Detector (CD) is filled and to confirm the proper operation of the purification plants. The aim is to guarantee that the concentrations of U-238 and Th-232 in the LS do not exceed the given limits of 10^{-15} g/g or 10^{-16} g/g for the reactor or solar neutrino measurement, respectively. The OSIRIS detector, under construction at the JUNO experimental site, will use the same readout electronics that has been developed and produced for the JUNO Central Detector. The present thesis work will analyze the data from the first dry run (which took place in June 2023) and is structured into the following parts: introduction (the objective and structure of the JUNO experiment will be presented, including the electronics and purification chain, with a focus on OSIRIS), raw data analysis (after a brief presentation of the data under analysis, a sample of the data will be selected from which fundamental quantities will be extracted for the subsequent discussion, and their correct behavior will be verified), event vertex reconstruction (following the introduction of the Charge Center Method, it will be applied to propose the reconstruction of event vertices in OSIRIS), conclusion (the primary findings of the study will be summarized, and subsequent key developments will be discussed)

Il Jiangmen Underground Neutrino Observatory (JUNO) sarà, per la prossima decade, il rivelatore di neutrini basato su Liquido Scintillatore (LS) più grande al mondo. Grazie alla sua massa attiva molto ampia (20 kton) e alle prestazioni all'avanguardia (risoluzione energetica efficace del 3% a 1 MeV), sarà in grado di effettuare importanti misurazioni nella fisica dei neutrini. La radio-purezza del LS è un elemento cruciale per la maggior parte delle misure che JUNO effettuerà. L'Online Scintillator Internal Radioactivity Investigation System (OSIRIS) è un rivelatore autonomo progettato per monitorare la radio-purezza del LS durante il riempimento del rivelatore centrale di JUNO (CD) e per confermare il corretto funzionamento degli impianti di purificazione. L'obiettivo è garantire che le concentrazioni di U-238 e Th-232 nel LS non superino i limiti stabiliti di 10^{-15} g/g o 10^{-16} g/g rispettivamente per le misurazioni dei neutrini da reattore o solari. Il rivelatore OSIRIS, attualmente in costruzione presso il sito sperimentale di JUNO, utilizzerà la stessa elettronica di lettura sviluppata e prodotta per il CD di JUNO. Il presente lavoro di tesi analizzerà i dati della prima dry run (svoltasi nel giugno 2023) e sarà strutturato nei seguenti capitoli: introduzione (verranno presentati l'obiettivo e la struttura dell'esperimento JUNO, compresi l'elettronica e la catena di purificazione, con un focus su OSIRIS), analisi dei dati grezzi (dopo una breve presentazione dei dati in analisi, verrà selezionato un campione di dati da cui saranno estratte grandezze fondamentali per la successiva discussione e verrà verificato il loro corretto andamento), ricostruzione del vertice degli eventi (dopo aver introdotto il Charge Center Method, questo verrà applicato per proporre la ricostruzione dei vertici degli eventi in OSIRIS), conclusione (saranno riassunti i risultati principali dello studio e saranno discussi i principali possibili sviluppi).

1 Introduction

1.1 JUNO

The Jiangmen Underground Neutrino Observatory (JUNO) [1] is a next-generation neutrino experiment under construction in South China, whose aim is to tackle unresolved issues in neutrino physics and astrophysics. The experiment has been proposed with the main goal of determining the neutrino mass ordering (NMO) at 3σ significance within six years of operation [2], and providing a measurement of the neutrino oscillation parameters with sub-percent precision [3]. JUNO will measure the energy spectra of neutrinos coming from Nuclear Power Plants (NPP) at a medium distance, O(50km) (which is a very clean experimental method to determine the mass hierarchy without exploring neutrinos matter effects [4]).

The JUNO experiment is located in Jinji town, Kaiping city, Jiangmen county, Guangdong province. The experimental site is 43 km to the southwest of the Kaiping city, a county-level city in the prefecture-level city Jiangmen in Guangdong province. There are five big cities in the neighborhood: Guangzhou, Hong Kong, Macau, Shenzhen, and Zhuhai, all in about 200 km drive distance, as shown in *figure* 1.1.

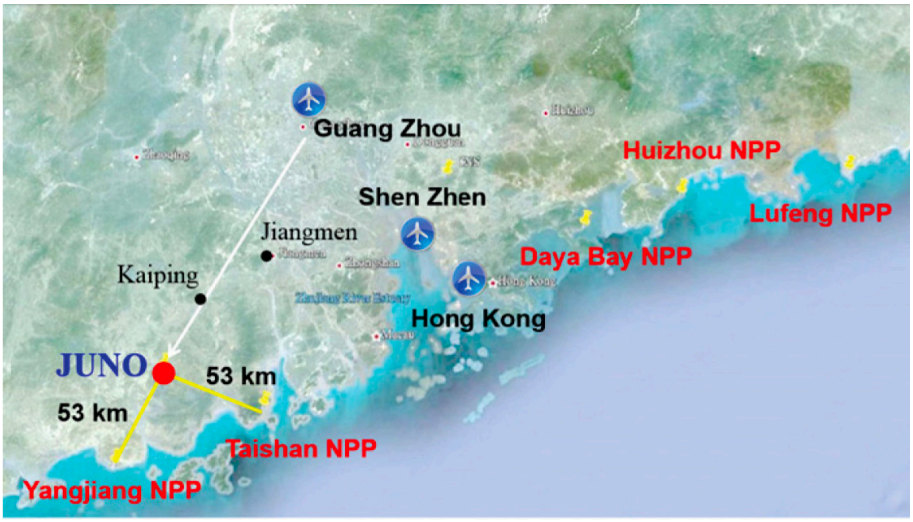

Figure 1.1: JUNO location in South China.

The experimental site is at 53 km from the Yangjiang and Taishan NPPs: Yangjiang has six reactor cores of 2.9 GW_{th} each (the distances between any two cores of Yangjiang NPP are between 88 m and 736 m) and all cores are 2nd generation pressurized water reactors CPR1000 [5]. Taishan NPP has planned four cores (3rd generation pressurized water reactors EPR) of 4.59 GW_{th} each. Other NPPs are in the Daya Bay complex (215 km away from the JUNO detector) that includes Daya Bay NPP, Ling Ao NPP, and Ling Ao-II NPP in a spread of 1.1 km, each with 2 cores of 2.9 GWth. There is no other NPP or planned NPP in 500 km around the JUNO experimental site. In absence of high mountains in the allowed area where the sensitivity to the mass hierarchy is optimized, the detector is deployed in an underground laboratory under the Dashi hill (650m overburden): the activities of the ^{238}U, ^{232}Th, and ^{40}K in the rock around the experimental hall have been measured to be 130, 113, and 1062 Bq/kg, respectively.

The muon rate and average energy in the JUNO detector are expected to be 0.0030 Hz/m^2 and 215 GeV respectively, estimated by simulation with the surveyed mountain profile taken into account [2].

The JUNO central detector (CD) contains 20 kton of higly transparent liquid scintillator (LS) unclosed in an acrylic sphere: particle interactions in the LS generate scintillation and Cherenkov photons, which are then converted into photo-electrons (PEs) by 17612 20-inch Photomultiplier Tubes (PMTs) (Large-PMTs), and 25600 3-inch PMTs (Small-PMTs). In addition, 2400 Large-PMTs are installed in the instrumented Water Pool detector, working as water Cherenkov muon veto system, in which the CD is immersed [6]. A schematic view of JUNO detector is shown in *figure* 1.2.

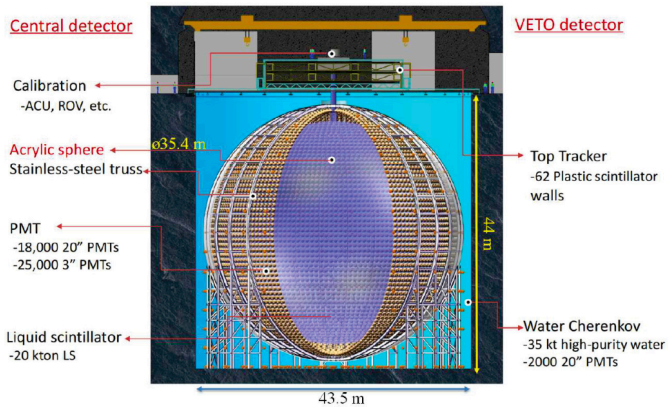

Figure 1.2: JUNO experimental apparatus.

The readout electronics modules of the PMTs will be placed in sealed boxes at a maximum water depth of 43 m, making it impossible to repair or access them after installation. As JUNO is expected to collect data for over a decade, the reliability of the readout electronics was one of the main concerns during the design phase, which translated into a careful selection of the electronics components in order to reach a maximum failure rate of 0.5% over 6 years of operation. The full electronics chain of JUNO Large-PMT electronics (*figure* 1.3) is composed of two parts [6]: the front-end (FE) or wet electronics (very close to the PMT output, inside the JUNO Water Pool) and the dry electronics (installed in the electronics room of the JUNO underground laboratories), consisting of the back-end (BE) electronics and data acquisition (DAQ) system. Dry and wet electronics are collected by two independent ethernet cables: a so-called "synchronous link" (S-link), which provides the clock and synchronization to the boards and handles the trigger primitives, and an "asynchronous link" (A-link) which is fully dedicated to the DAQ and to the Detector Control System (DCS).

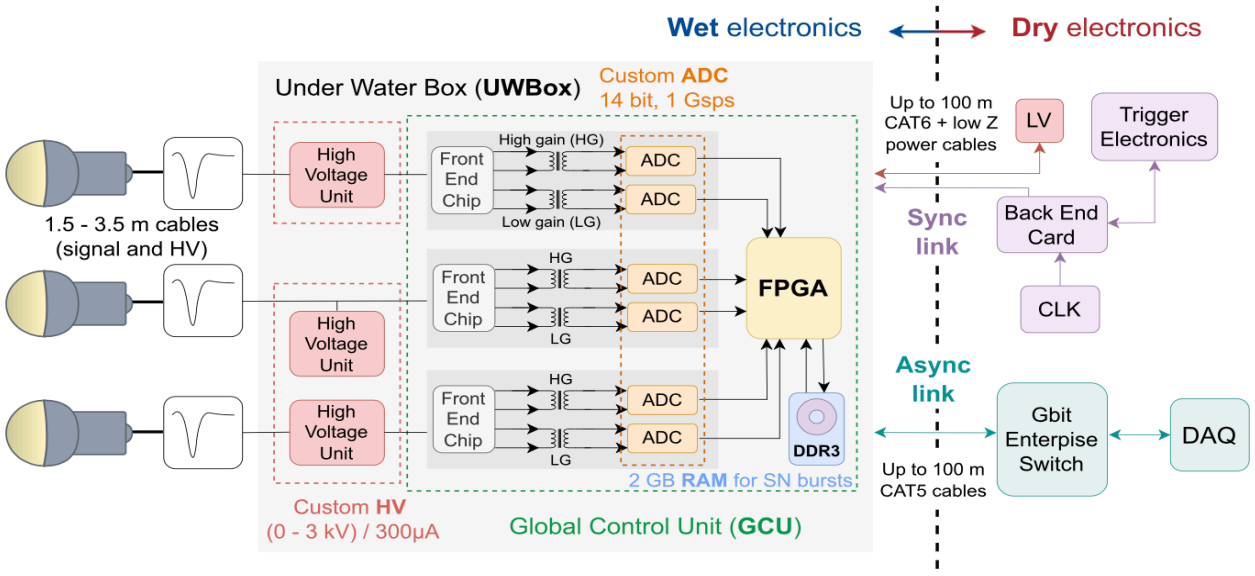

Figure 1.3: JUNO large PMT electronics Read-Out electronics scheme. From [6].

Front-End Electronis: The FE electronis is installed underwater under a water-tight box (the so-called Under Water Box, UWbox). Three PMT output signals are fed to one UWbox (connected with a 50Ω, coaxial cable, with a fixed length of 2m for the CD PMTs and 4m for the VETO PMTs [7]), witch cointains three High Voltage Units (HVU) and a Global Control Unit (GCU), a motherboard incorporating the FE and Readout electronics components. The PMT analog signal reaching the GCU is processed by a custom FE Chip, witch splits the input signal and amplifies it with two different gains; the two signals are further converted to a digital waveform by a 14 bit custom Flash Analog-to-Digital Converter (FADC). The usage of two FADCs per readout channel is driven by the stringent requirements on the charge resolution to be achieved over a wide dynamic range: 0.1PE resolution between 1PE and 100PEs (high-gain stream), and 1PE resolution between 100PEs and 1000 PEs (low-gain stream) [2].

Back-End Electronics: The BE electronics is composed of a Back End Card (BEC) with the Trigger and Time Interface Mezzanine (TTIM) and of the Trigger Electronics: the Reorganize and Multiplex Units (RMU) and the Central Trigger Unit (CTU).

1.2 OSIRIS

Shielding against environmental radiation and a careful selection of radiopure detector materials play a crucial role in JUNO: radioactivity in LS may result in events mimicking the inverse beta decay (IBD) coincidence signals of electron antineutrinos or cause pile-up to the single events distorting the energy scale, both having direct impact on the sensitivity of the neutrino mass hierarchy measurement. This is why the JUNO collaboration has set stringent upper limits on the contamination level of the LS with uranium and thorium chain isotopes: $\leq 10^{15} g/g$ for the IBD-based physics program (IBD-level) and $\leq 10^{16} - 10^{17} g/g$ for solar neutrino detection (solar-level). To achieve these radiopurity requirements, an extensive purification program needs to be performed on the LS before filling the JUNO detector. The primary fluor of JUNO LS is 2,5-diphenyloxazole (PPO) added at a concentration of 2.5 g/l, with p-bis-(o-methylstyryl)-benzene (Bis-MSB) as wavelength shifter at 3 mg/l [8]. These components are mixed and purified on-site in a chain of purification plants (*figure* 1.4): column chromatography in alumina, distillation, water extraction and steam stripping. As the last stage in this purification chain, the Online Scintillator Internal Radioactivity Investigation System (OSIRIS) will serve as a stand-alone detector to verify the efficiency of the upstream purification plants and to monitor the LS radiopurity during the filling of the JUNO CD [9].

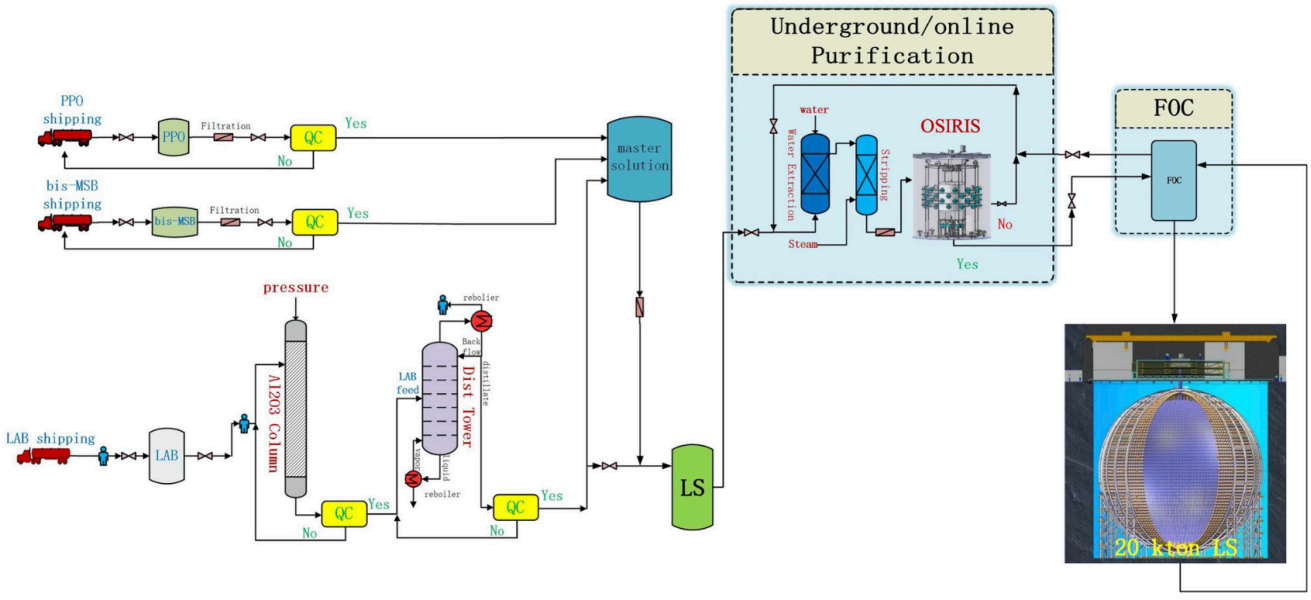

Figure 1.4: JUNO LS purification chain.

The OSIRIS setup has been optimized for the purpose of detecting the residuals of natural U/Th contamination of the LS: it will hold a 18-ton sample of LS enclosed in an Acrylic Vessel (AV) for screening, surrounded by an extensive water shield. The search is based on the fast coincidence decays of ^{241}Bi-^{214}Po and ^{212}Bi-^{212}Po present in the decay chains of U and Th, respectively. The detector dimensions are outlined by the outer Water Tank that features about 9.4 m diameter and height (given the underground location, the footprint and the height are limited). The setup is logically divided into an acryling Inner Detector (ID) containing the LS volume and surrounding PMTs and an Outer Detector (OD) equipped with few PMTs and utilizing the water shielding as a Cherenkov muon veto. A simplified scheme of the detector is shown in *figure* 1.5.

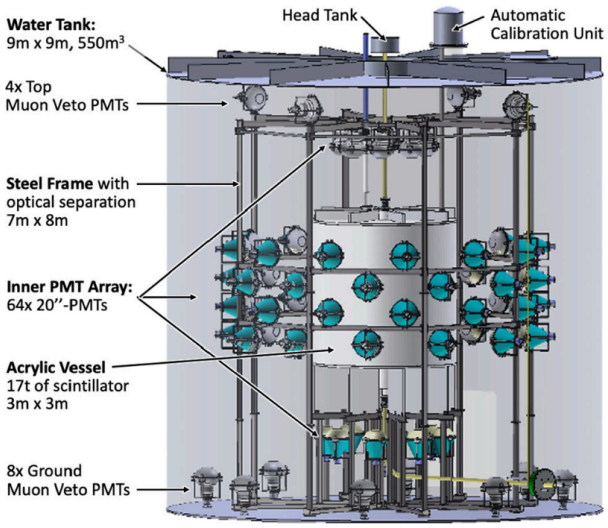

Figure 1.5: Layout of the OSIRIS detector.

Some parts of OSIRIS are [9]:

- **Scintillator target:** the LS batch to be tested is contained in a transparent vessel with an inner volume of 21 m^3 (18 tons of LS), located in the center of the setup, surrounded by a water buffer for shielding, separated from the LS by the Acrylic Vessel;
- **Acrylic vessel (AV):** is a cylindrical vessel of 3 m height and diameter which holds the LS at the center of the detector; it is held in place by eight vertical acrylic plates of 1m height, which are in turn mounted on the inner Steel Frame of 2 m height;
- **Steel Frame (SF):** it reaches a height of 8m and a diameter of 7 m; while its inner section supports the AV, the outer section holds the photomultipliers, calibration and sensor systems as well as the black-and-white PET sheets constituting the Optical Separation between ID and OD;
- **Inner PMT Array:** the scintillation light produced by events inside the LS volume is recorded by 64 20"-MCP PMTs mounted on the SF and facing inwards. For shielding the γ-rays emitted from PMT glass, the tubes are placed at a distance of 1.3 m from the AV surface (2.8 m from the detector center);
- **Water Tank (WT):** made of carbon steel plates, the cylindrical WT holds a 9m height and diameter volume, offering a 4π shielding of more than 3m of water from external gamma rays emitted by the cavern rock of the Scintillator Hall. The residual γ-flux has been determined to be sufficiently low for an effective Bi–Po coincidence search;
- **Muon Veto:** cosmic muons, secondary neutrons and radioactive isotopes created in spallation processes feature a finite probability to mimic the fast coincidence signals of the Bi–Po decays. Therefore, a secondary array of 12 20"-PMTs watches the volume between SF and tank walls, using the ultrapure water as a Cherenkov radiator for crossing muons;
- **Calibration systems:** in order to calibrate the detector, a picosecond laser pulses will be inserted into the detector by a system of optical fibers with the emission points mounted to the SF. Moreover, an Automatic Calibration Unit (ACU) [10] refurbished from the Daya Bay experiment is connected to the AV via a steel pipe in order to lower radioactive sources and an LED directly into the LS volume to calibrate the detector response. Source positions can be cross-checked using a CCD system mounted to the SF.

2

Raw Data analysis

In this chapter, we'll first introduce the datasets that will be analysed and propose an analysis code (Python-based) for the unprocessed waveforms. The aim of this code is to extract specific observables, assess their stability, and compare them with their pre-processed counterparts. The purpose of this section is to provide a general overview, without aiming for exhaustiveness, of the selected quantities. This will establish a solid foundation for the subsequent chapter, which focuses on the reconstruction of the event vertex. The chapter is organized into two sections: datasets presentation and raw data analysis.

2.1 Datasets presentation

The data analyzed in this work pertain to two distinct test runs conducted in the OSIRIS setup on two separate days: June 9, 2023 (1st cycle test of the dry detector, day 1) and June 11, 2023 (1st cycle test of the dry detector, day 3). The significant difference between the two days lies in the employed experimental setup configuration: the first difference can be traced to the fact that on day 1, OSIRIS was empty and not filled with LS. In contrast, on day 2, an LS ball (with a diameter of 7.17 cm and a volume of 193 ml) was inserted close to the AV lateral surface, and its position (i.e., the event vertex) is the subject of analysis. The first condition will be referred to as "background," and the second as "signal". The total number of readout channels (number of PMTs) in the detector is 64 (CD) + 12 (Veto), for a total of 76 active channels. The second difference resides in the distinct trigger choice during data acquisition: data of the "Background" run are acquired in self-trigger mode (i.e., every signal detected by at least one PMT is considered valid), while the "Signal" run is taken by requiring a PMT signal coincidence, nPMT=13. Essential information regarding the data acquisition conditions is summarized in *Table* 2.1.

	Name:	**nPMT**	$\Delta t_{acquisition}$ **[s]**
Background	TCP_27GCU_06-09-0925_1008_1000	Self Trigger	300
Signal	TCP_27GCU_06-11-1048_1008_1000	10	2000

Table 2.1: Data acquisition conditions.

Here $\Delta t_{acquisition}$ denotes the data acquisition time, and nPMT stands for the chosen trigger configuration during the acquisition phase. It will become evident in the subsequent sections how this value plays a crucial role in measurement quality. To highlight this fact, an appropriate variable will be implemented in the code to increase its value for a more accurate event selection (see *Chapter* 3 for a more detailed explanation). For each of the presented datasets, two distinct files exist:

- *.root $\longrightarrow$ these represent the raw data with full waveform information;
- *_ana.root $\longrightarrow$ reconstructed data with fundamental quantities (hereinafter referred to as "pre-processed").

The aforementioned files were obtained through a sequence of processes that transformed the PMTs' output into binary files and subsequently organized them into ROOT [11] TTRee objects, stored in ROOT files. The ROOT TTree class is designed for efficiently storing a large number of objects of the same class. In this chapter, only some of the variables present in the TTree, from a subsample of Signal dataset, will be considered. Specifically, the variable "chData" (representing the set of signal values in ADC counts recorded by various channels during events) will be extracted from the *.root file. First, the trend of timestamps was examined to identify any issues in the event selection for analysis. Second, the study of the reconstructed quantities 'baseline,' 'hit time,' and 'charge' was performed, extracting valuable information for the subsequent analysis. The total duration of the waveform acquisition window is 1 μs. In *Figure* 2.1, a zoomed-in view of a waveform is shown, highlighting the mentioned variables.

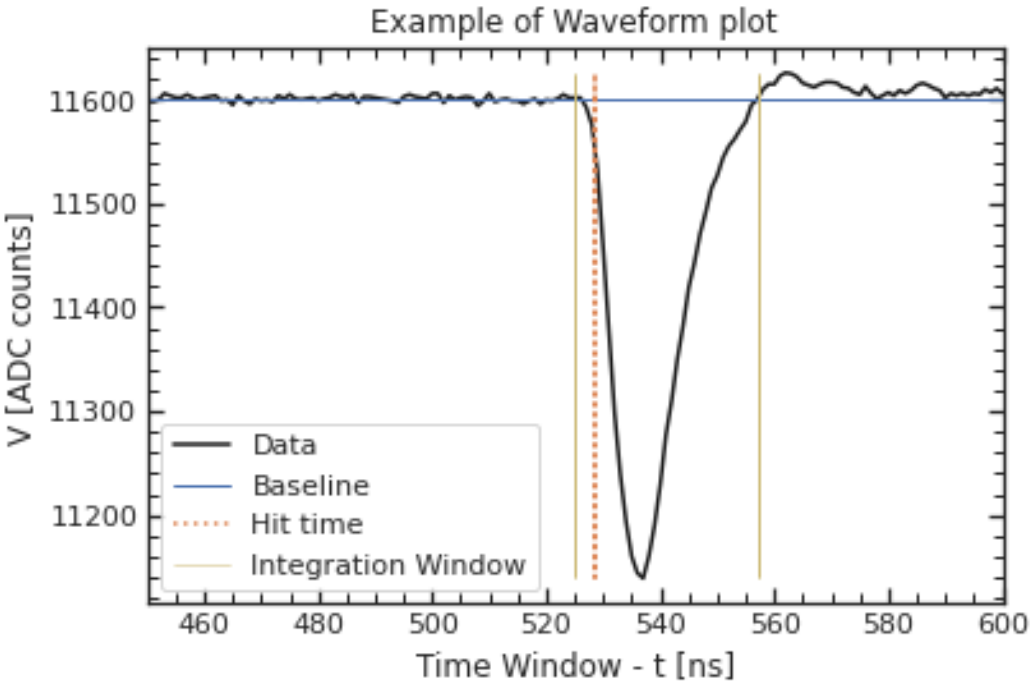

Figure 2.1: Zoom into the waveform of event 2683 with the quantities of interest highlighted: the baseline in blue, the hit time in orange, and the integration window in yellow.

The blue horizontal line represents the baseline value (i.e., the voltage level recorded under no-event conditions), the orange vertical line represents the hit time t_{hit} (i.e., the instant at which the waveform crosses the trigger threshold), and finally, the yellow region depicts the integration window (i.e., the time interval over which the waveform is integrated after baseline subtraction for the determination of the accumulated charge). Further details about the calculation methods for these quantities will be presented in the following section.

2.2 Raw Data Analysis

As first step, an analysis of the "timestamps" variable has been performed. The timestamp is the absolute time of the event, in unit of 8ns. This variable proves particularly useful in understanding the presence of any gaps in the data collection. The trend for a portion of the "Signal" dataset is depicted in *Figure* 2.2.

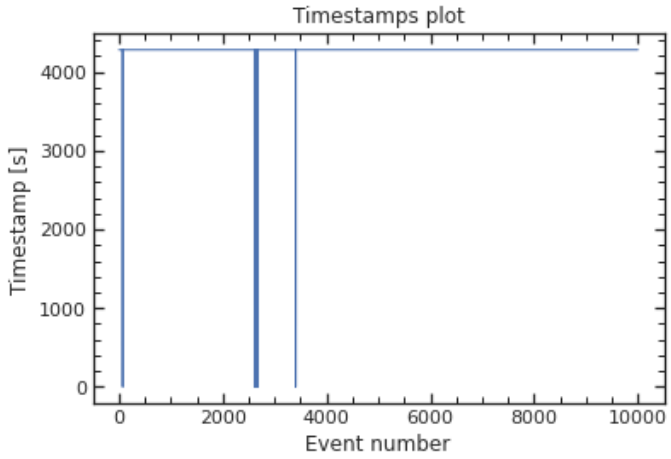

Figure 2.2: Timestamps plot - Subsample of Signal run.

It is clearly noticeable that there are few gaps in the data. To identify anomalous events, a loop has been implemented in the program to retrieve the event number corresponding to null timestamps. Once these events are identified, the timestamps dataset is purified by removing them, and a new plot is generated (shown in *Figure* 2.3).

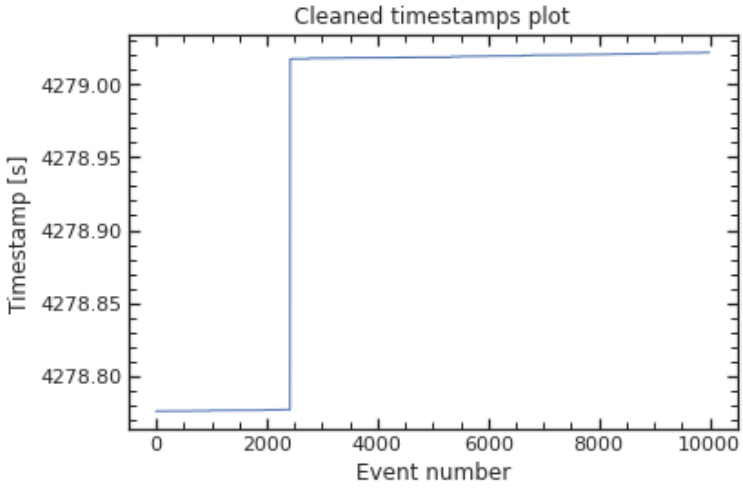

Figure 2.3: Timestamps plot - cleaned.

A clear and constant trend is evident in two regions of the plot, connected by a coincident jump at event 2420. Since the "Signal" run, from which the analyzed subsample is extracted, is acquired with a coincidence trigger of nPMT=10, it is expected that, for each event, all the waveforms from the 76 active PMTs should have the same timestamps. As the trend becomes constant starting from event 2420, in addition to events with null timestamps, the first 2420 events have been excluded from further analysis (each GCU has its own initialization time from which it begins data collection).

From the study of the timestamps dataset constructed, it is possible to derive the event rate for the "Signal" run. First, the difference Δt between consecutive timestamps was calculated, and it exhibited a decreasing exponential behavior. By performing an exponential fit of the form $A \cdot e^{-\lambda t}$ to this profile, it was possible to obtain the event rate as $R = \lambda^{-1}$, resulting in $R = 12.61\,Hz$. In *Figure* 2.4, the Δt profile is shown, highlighting the exponential fit.

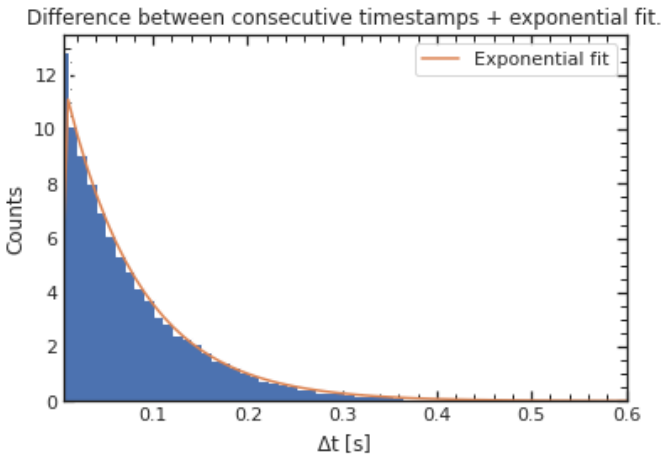

Figure 2.4: Histogram of differences between consecutive timestamps + exponential fit.

Firstly, due to the multitude of recorded events, it was necessary to perform a selection based on fundamental event characteristics. This selection ensured that the waveform associated with an event had a non-zero total charge (i.e., the integral of the waveform within a integration window matching the event duration was non-zero) and exhibited only a single hit time point (i.e., the time at which the signal crossed the trigger threshold, set at the value baseline-5σ). As an illustrative example, *Figure* 2.5 presents six waveforms that prominently display the aforementioned characteristics.

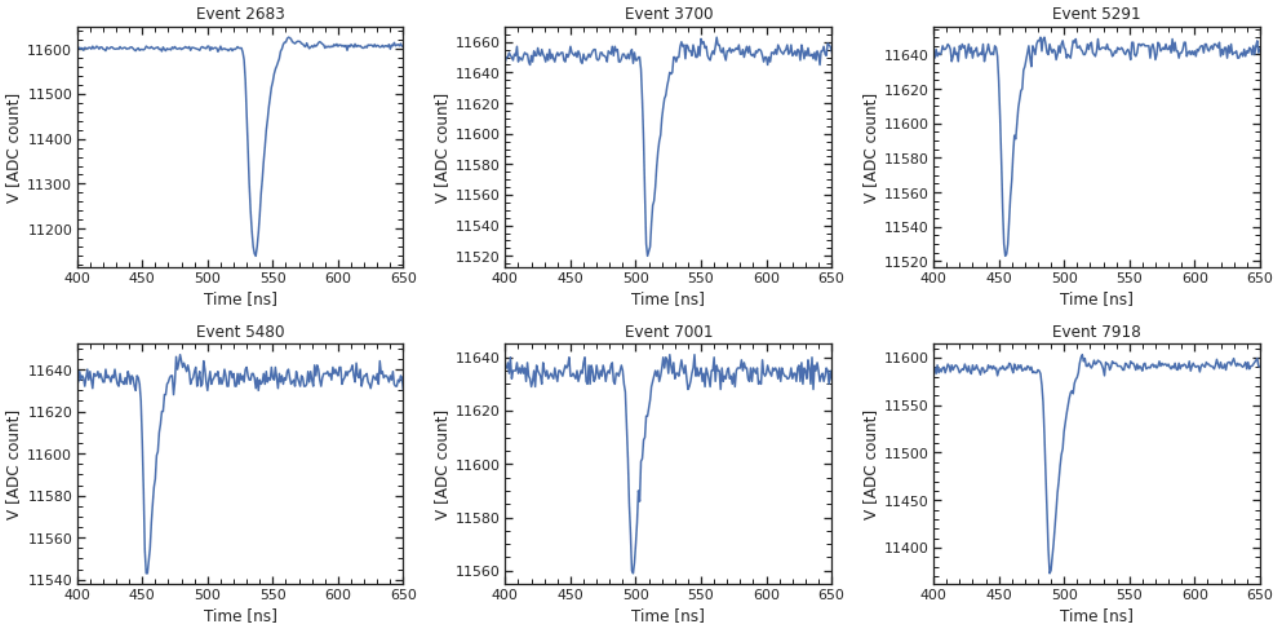

Figure 2.5: Waveform Examples.

The focus of the subsequent paragraphs is the analysis of the stability of quantities (particularly baseline, hit time, and charge) and their correspondence with pre-processed values. The analysis will run in parallel for the six previously selected events, aiming to highlight any differences or anomalies within the dataset and to confirm the correct behavior of the estimated quantities with greater certainty.

2.2.1 Baseline

The calculation of the baseline value is performed over the first N_{pre}=460 bins, where N_{pre} represents a variable number that corresponds to the time width, in nanoseconds, of the pre-trigger region (i.e., $N_{pre} = 460$ bins corresponds to 460ns). This region precedes the threshold crossing and can be dynamically adjusted during acquisition. The baseline behavior for the considered events, overlaid with a Gaussian fit, is shown in *Figure* 2.6.

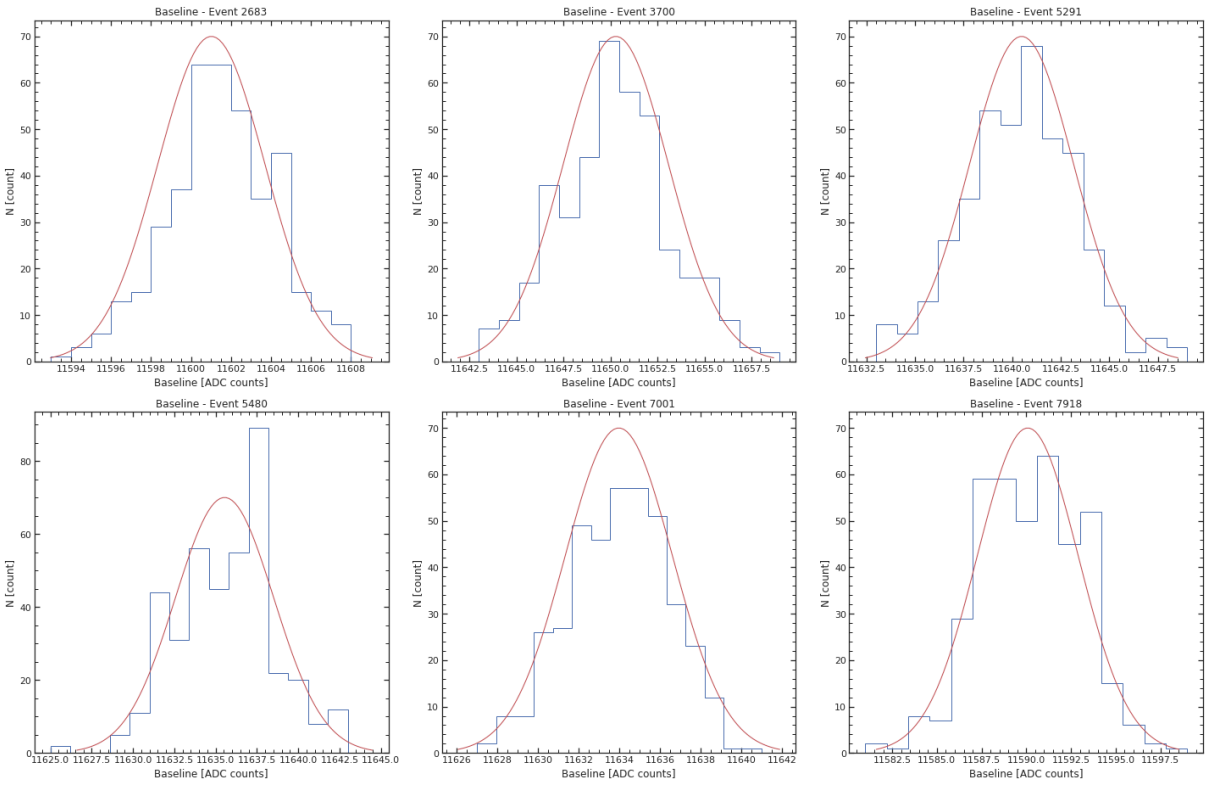

Figure 2.6: Histogram of Baseline.

From the Gaussian fit, certain information about the distribution's centroid (average baseline value) μ_B and standard deviation σ_B can be extracted. The trend of the obtained μ_B and σ_B values is shown in *figure* 2.7 and 2.8.

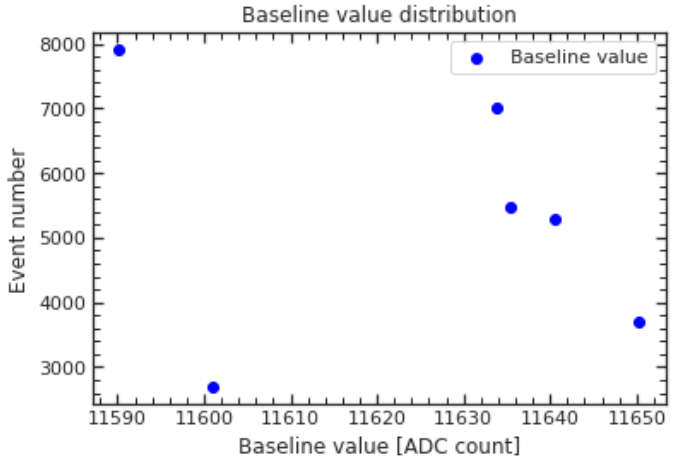

Figure 2.7: μ_B distribution.

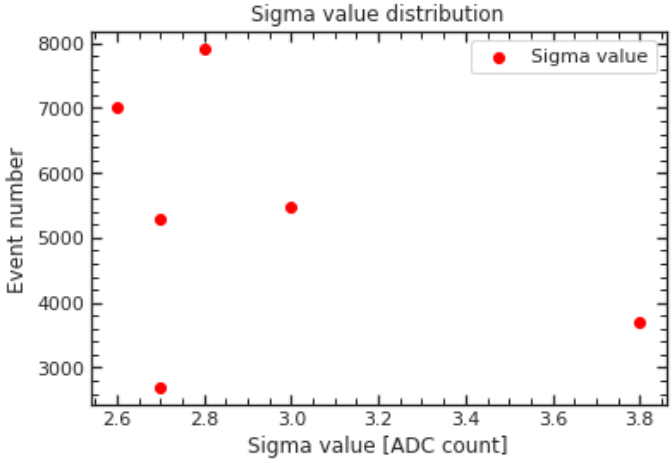

Figure 2.8: σ_B distribution.

In order to verify the correct trend of the estimated baseline values compared to the pre-processed ones, a linear fit of pre-processed versus experimental values is shown in *figure* 2.9.

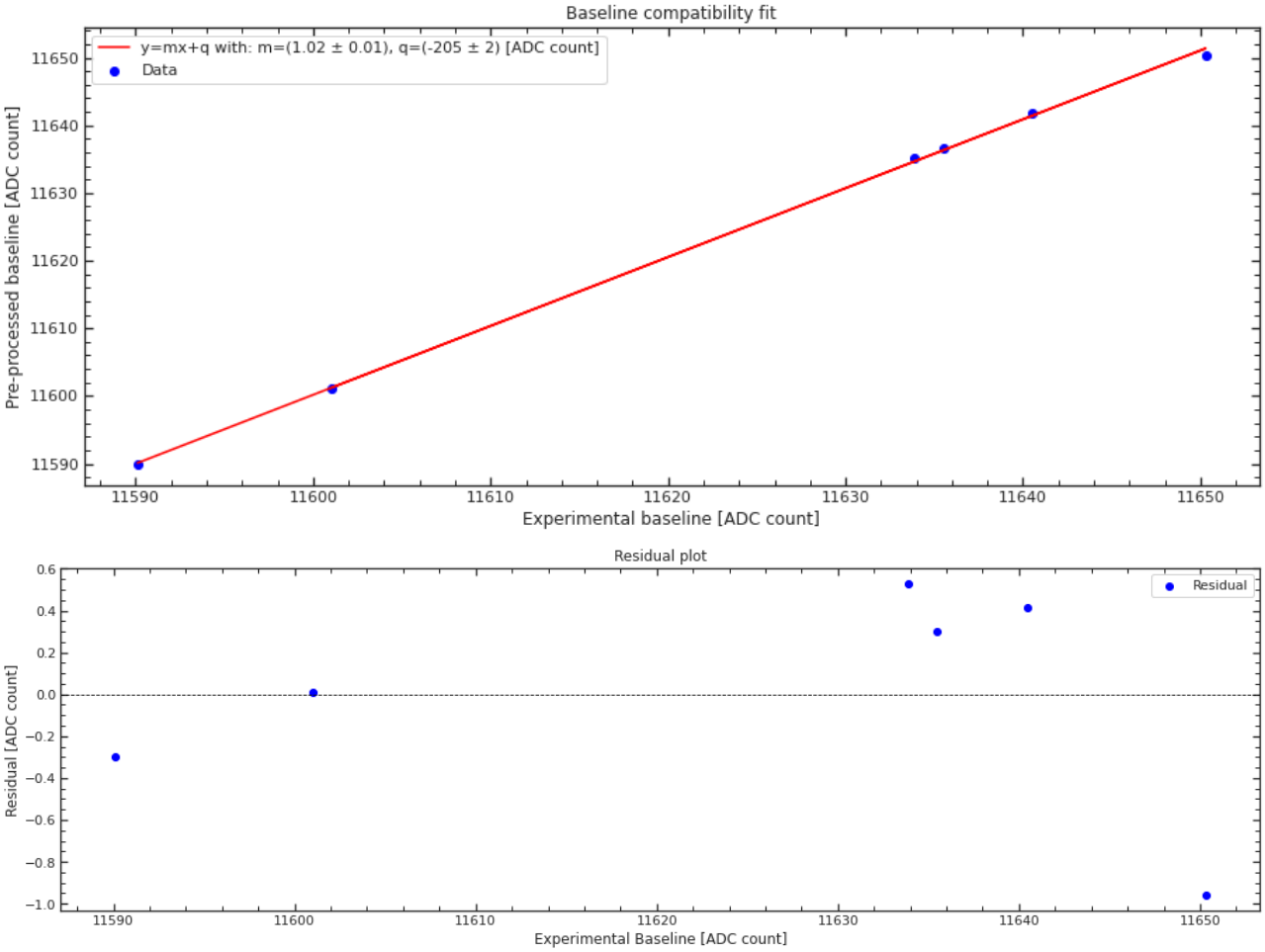

Figure 2.9: Baseline compatibility fit + residual.

A strong compatibility of the slope coefficient m from the linear fit with unity is evident, indicating a good agreement between experimentally obtained values and pre-processed ones.
The baseline trend for all events was then analyzed as it varies across the PMTs. In *Figure* 2.10, the trend of baseline values for three PMTs is shown (the trend is similar for all PMTs). In *Figure* 2.11, the temporal trend of baseline values for events from a single PMT is displayed.

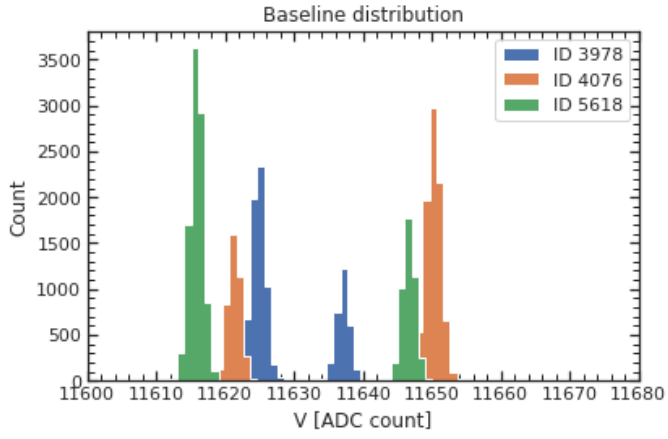

Figure 2.10: Baseline value distribution for IDs 3978, 4076 and 5618.

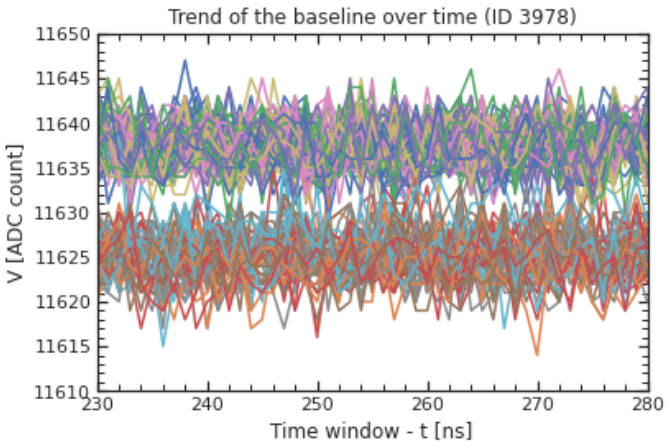

Figure 2.11: Baseline over time (ID 3978).

As can be observed, each PMT records baseline values that are distributed along two Gaussian profiles shifted by approximately $10 - 40$ ADC counts from each other. This confirms the expected Gaussian behavior of the baseline; however, the stability of this behavior diminishes, as it oscillates between two values during the run, even though the discrepancy between these values may be considered negligible.

2.2.2 Hit time t_{hit}

The determination of the hit time value is the time when the waveform signal crosses the trigger threshold (which is set at t_{hit}=Baseline-5σ). Referring to the data presented earlier, the baseline value (μ_B) and standard deviation ($\sigma = \sigma_B$) are used, providing an estimation of t_{hit} for each event. The trend of Hit time values is shown in *figure* 2.12.

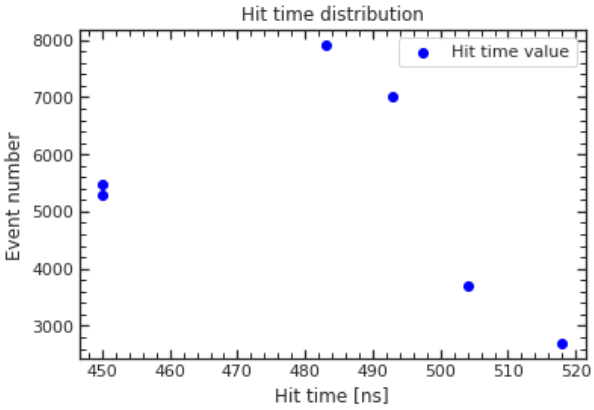

Figure 2.12: Hit time distribution.

In *figure* 2.13, a linear fit of the trend between pre-processed t_{hit} values and reconstructed is shown.

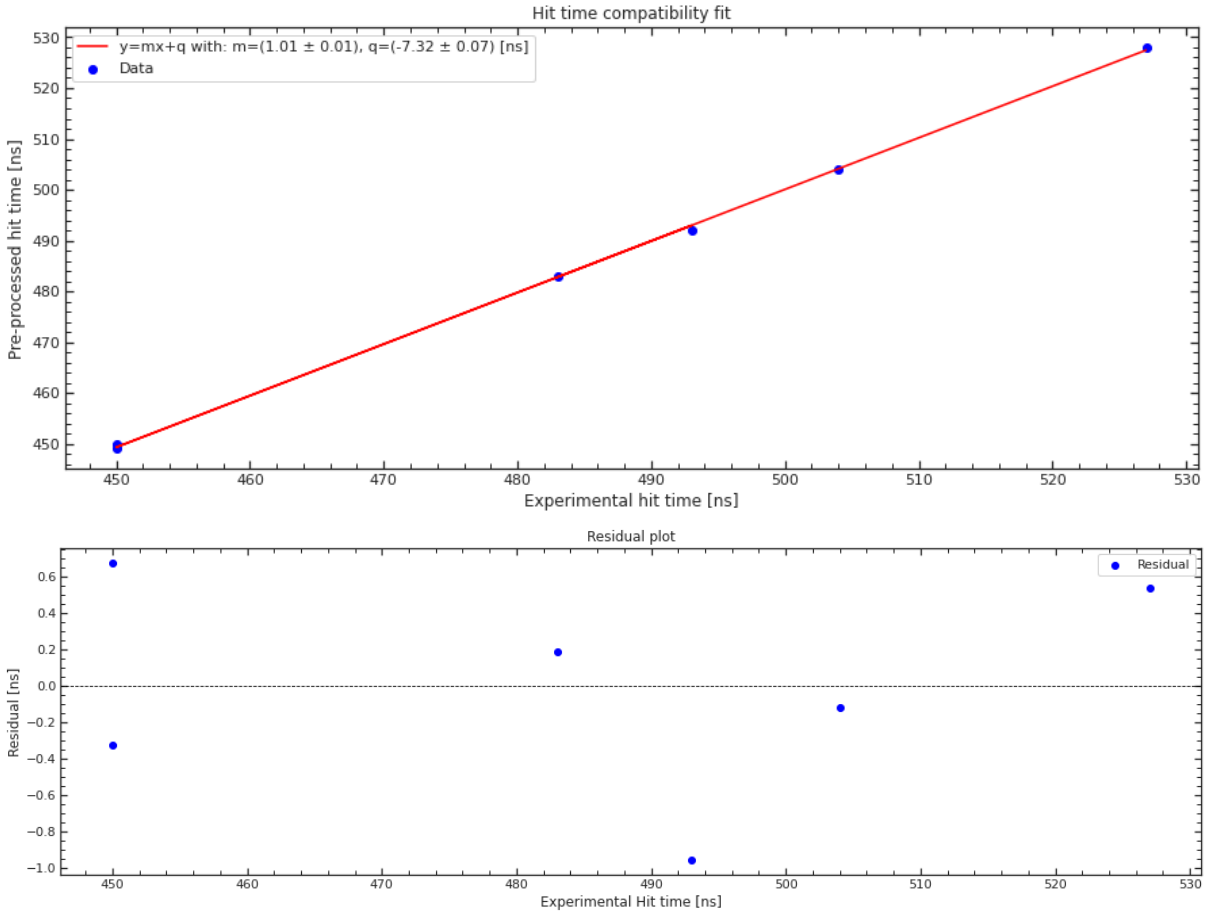

Figure 2.13: Baseline compatibility fit + residual.

2.2.3 Integrated charge

In this section, the aim is to identify a suitable procedure for extracting information regarding the charge released by particles in the Liquid Scintillator (LS) from the acquired signals. The total charge can be estimated as the area enclosed between the waveform and the baseline within a fixed integration range Δt_S. To compute an estimation of the integrated charge for a single waveform (expressed in μC), the following formula is employed [12]:

$$Q = \Delta t_S \frac{\sum_i^{N_B} |N_i - B| \cdot 75\mu V}{R}$$

In this equation, N_i represents the content (in ADC counts) of the i-th bin, while B indicates the mean baseline value (in ADC counts) of the considered waveform. R is 50Ω (i.e., the expected PMT imput impedance), and $75\mu V$ corresponds to the voltage equivalent of 1 ADC count. For estimating the released charge, the only relevant variable is the choice of the integration window. Therefore, two approaches have been adopted to evaluate any differences:

1. Set an integration window $\Delta t_S^{NotFixed}$ that varies for each event, adapting to the waveform profile of the considered event (see *Figure* 2.1 for reference): this integration window is defined by taking as reference the two points where the waveform intersects with the baseline immediately before and after the event;

2. Set a constant integration window Δt_S^{Fixed} for all events, encompassing all waveform shapes of the considered events (refer to *Figure* 2.14).

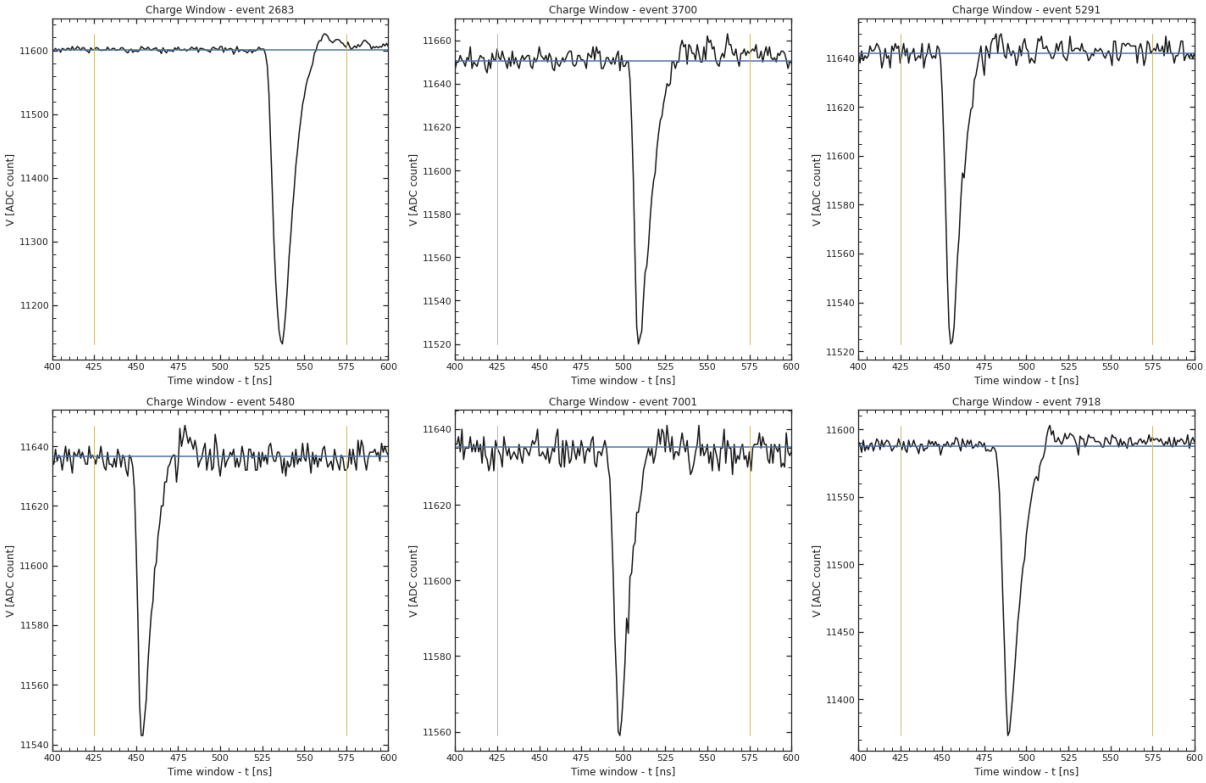

Figure 2.14: Examples of fixed integration windows.

The linear fits (Q, Q_1), (Q, Q_2), and (Q_1, Q_2) are shown in *Figure* 2.15, where Q is the pre-processed charge value, Q_1 and Q_2 are the charge values obtained using the two methods. The results of the fits are reported in *table* 2.2.

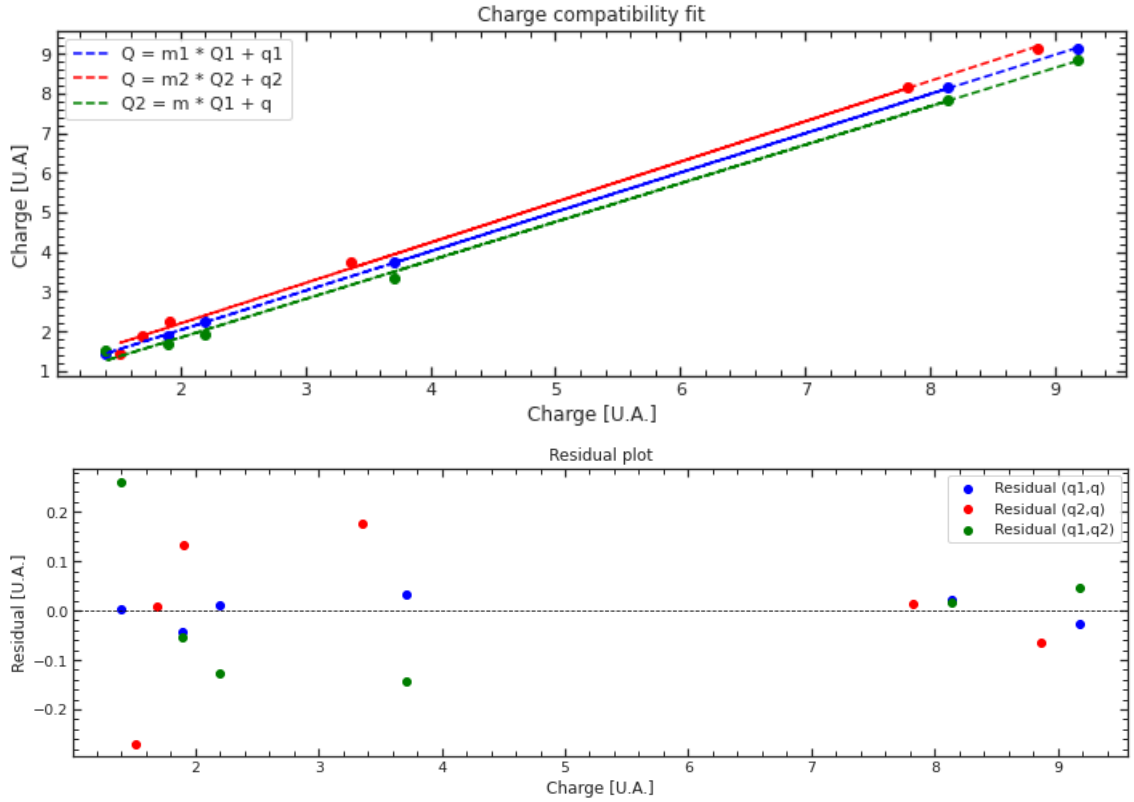

Figure 2.15: Charge compatibility fit + residual.

Fit	$m \pm \sigma_m$	$q \pm \sigma_q\,[U.A.]$
Q **vs** Q_1	0.991 ± 0.004	0.049 ± 0.003
Q **vs** Q_2	1.010 ± 0.020	0.160 ± 0.020
Q_2 **vs** Q_1	0.970 ± 0.020	-0.10 ± 0.02

Table 2.2: Results of Charge compatibility fit.

It is evident the strong compatibility of the slope values m with 1, indicating the correctness of both methodologies (although the $\Delta t^{NotFixed}$ methodology appears slightly more precise). Furthermore, from the (Q_1, Q_2) fit, it is possible to confirm that the methods can be considered roughly equivalent, a sign that there are no evident systematic effects.

3

Event Vertex Reconstruction

The capability to reconstruct the event vertex with good precision is another important requirement for JUNO, because it allows to [6]:

1. Tag signal events through spatial coincidences;
2. Accurately define a fiducial region within the detector volume;
3. Precisely map the spatial response of the detector.

On the electronics side, this requirement brings the need to precisely reconstruct the arrival time of the scintillation photons onto the PMTs.

The chapter will be organized into two main sections:

1. The first section will compare the accumulated total charge in various runs. This analysis aims to identify potential anomalies and verify proper trends;

2. The second section will introduce the Charge Center Method (CCM), which is used for vertex reconstruction. This section will also include a discussion of the employed datasets. Additionally, after qualitatively presenting plot trends with varying trigger thresholds, the chapter will propose and discuss the results obtained from event vertex reconstruction.

3.1 OSIRIS 1st DRY RUN: PHYSICS MEASUREMENTS OF TOTAL CHARGE

Firstly, it was necessary to extract the total charge collected by the PMTs from the *_ana.root file. To achieve this, a function was constructed that takes input variables event_IDs, chCounts (identifying the i-th PMT collecting the charge), hit times and charges (as detailed in the previous chapter) and calculates the accumulated total charge (considering only events with at least one valid hit time and non-zero charge). This function was further refined to enable the calculation of the interaction vertex position. The charge distribution will be reconstructed and discussed separately for the background and signal runs.

Background run

The spectrum of the total accumulated charge in the background run is shown in *Figure* 3.1.

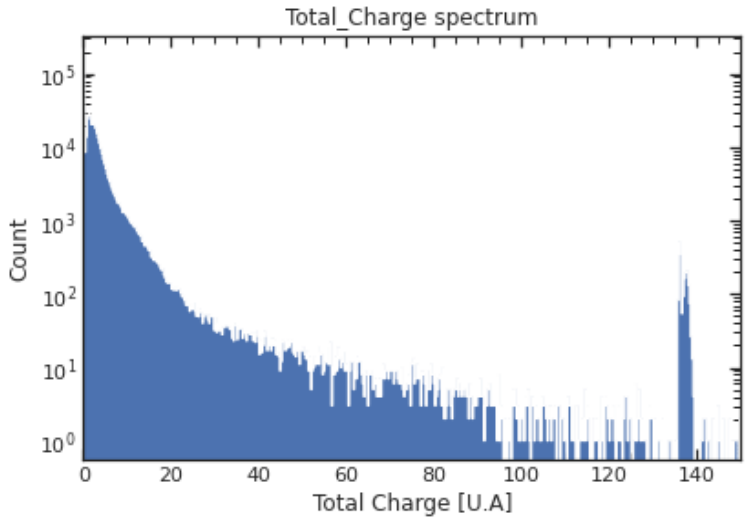

Figure 3.1: Total Charge - Background run.

An anomalous peak is identified for charge values $Q_{tot} > 130$ U.A. Consequently, a data mask was applied to define the events within this peak in order to study the cause of this excess charge. First, the trend of timestamps was analyzed to identify any gaps in the data and potentially reject a portion of them. In *Figure* 3.2 the trend of this variable is shown for all events, while *Figure* 3.3 shows the timestamp for $Q_{tot} > 130$ U.A. (corresponding to 1938 events).

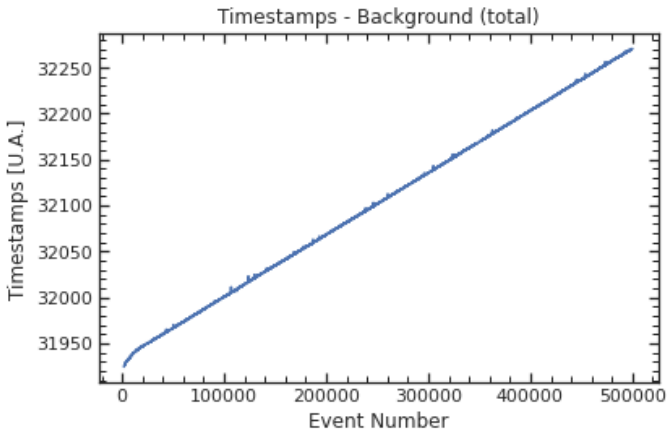

Figure 3.2: Total timestamps - Background run.

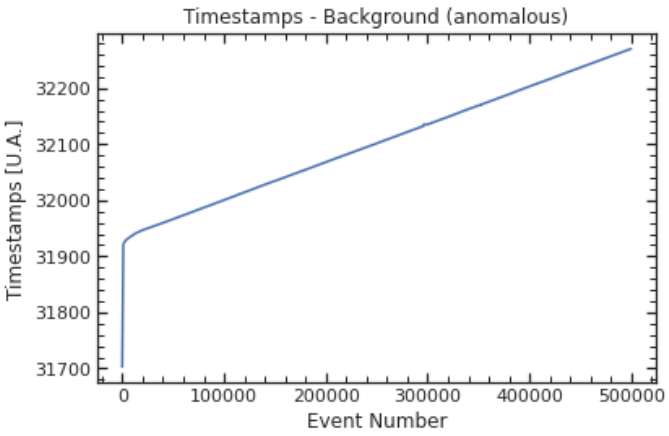

Figure 3.3: Anomalous peak timestamps - Background run.

A completely random trend is evident in the distribution of anomalous events, to the extent that the two plots (total and anomalous) perfectly overlap, except for the initial data. Given the lack of any specific trend for most of the timestamps of the anomalous events, it was decided to discard only the first 10,000 events, for which the timestamps are notably anomalous and deviate from the overall trend.

Since the analysis of timestamps did not provide an explanation for the anomalous peak of excess charge, attention was turned to the waveforms of the events producing such values. In *Figure* 3.4, for clarity, the behavior of one of

the waveforms from the affected events is shown (the others are similar).

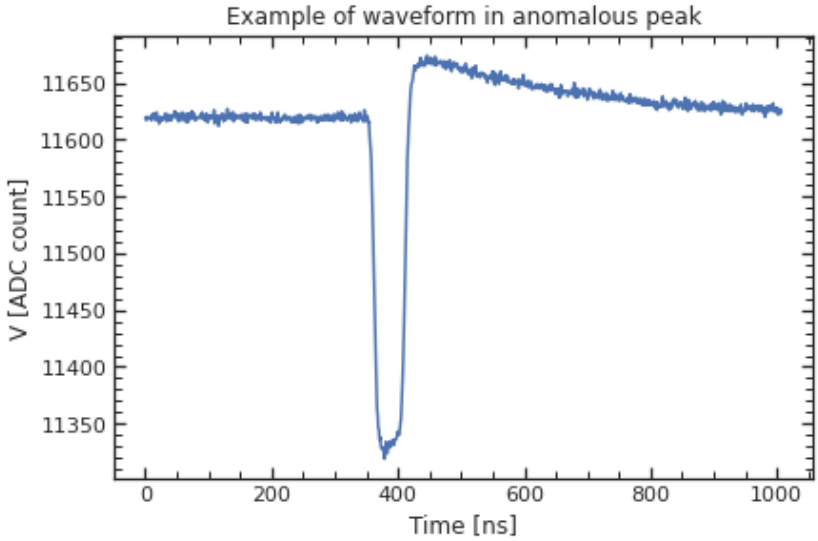

Figure 3.4: Anomalous waveform.

The unusual behavior of the waveform compared to those presented in the previous chapter is particularly evident. In particular, it can be observed that after the event, the signal does not return to the baseline but exhibits a strong overshoot. This unusual profile consequently leads to an incorrect estimate of Q_{tot}. However, such waveforms cannot be considered well-recorded, and in order not to skew the results of the subsequent analysis, it was decided to exclude these events.

Signal run

The trend of the total accumulated charge in the Signal run is shown in *Figure* 3.5.

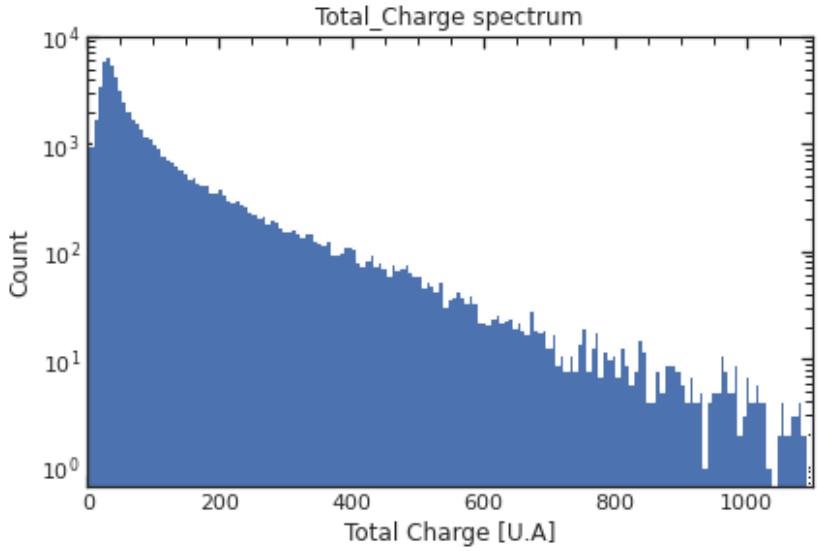

Figure 3.5: Total Charge - Signal run.

The charge trend shows a smooth behavior without any notable defects. The behavior and stability of the main variables in this run have been thoroughly discussed in the previous chapter. Consequently, this dataset is deemed reliable, and the next section proceeds to estimate the interaction vertex position.

3.2 Vertex Reconstruction

For estimating the interaction vertex position of an event in the LS ball, the "charge center method" (CCM) is employed. This method is based on the analysis of charges collected by the PMTs. When a charged particle interacts with a scintillating material, it ionizes the material and creates atomic excitations. These excitations release energy in the form of scintillating light (optical photons) as they return to their ground state. The generated photons are detected by PMTs or similar detectors. In the CCM, it is assumed that the generated photons are detected at positions proportional to the energy released along the particle's path. Consequently, the average position of the detected photons is used as an estimate for the interaction vertex position of the event. This method is particularly useful when scintillating light is collected from multiple points in the detector (as in the case of OSIRIS). Schematically, the method is implemented as follows:

1. Light pulses generated by the event are detected by multiple PMTs or similar sensors positioned at different locations.

2. For each PMT, a weighted position is calculated based on the released energy and the distance from the PMT. This weighted position represents the "charge center" for that sensor.

3. The average charge center is calculated by considering the weighted positions of all PMTs involved in the event.

4. The average charge center is then used as an estimate for the interaction vertex position of the event in the LS detector, according to the relationship [13]:

$$\vec{r}_0 = \frac{\sum_i q_i \vec{r}_i}{\sum_i q_i}$$

   where $\vec{r}_0$ is the interaction vertex position, q_i is the detected charge of the i-th PMT, and $\vec{r}_i$ is the position of the i-th PMT.

This method will be implemented in the following subsections. First, an analysis of the background will be conducted to detect measurements of background events. Subsequently, the analysis of two signals will be presented.

Preliminarily, let us recall two essential pieces of information necessary for the understanding of the following plots: we will refer to nPMTs (or triggering), as already presented in the previous chapter, as the minimum number of PMTs that must detect an event for it to be considered valid. In particular, this threshold is characteristic of the data-taking conditions (see *table* 2.1), but it is possible during analysis (and has been done and presented in the following sections) to "mask" events with higher triggering thresholds. It will be shown how such masking will significantly affect the analysis results. A second important element to recall is multiplicity: with multiplicity, we refer to the number of events "seen" by a certain number of PMTs. The multiplicity parameter is independent of the mask imposed during data analysis and is therefore dependent on the value of nPMTs set during data acquisition. In the following paragraphs, the trend of multiplicity will first be shown (confirming the prediction associated with the data-taking conditions), and then the vertex reconstruction plots will be presented, highlighting characteristic elements of each run.

Background

The run in question was taken with the PMTs in self-trigger mode which ensures the acquisition of all occurring events. In this condition, we therefore expect that the vast majority of events are associated with low values of multiplicity (i.e., an event is always detected by a single PMT). This trend is confirmed by the plot in *figure* 3.6.

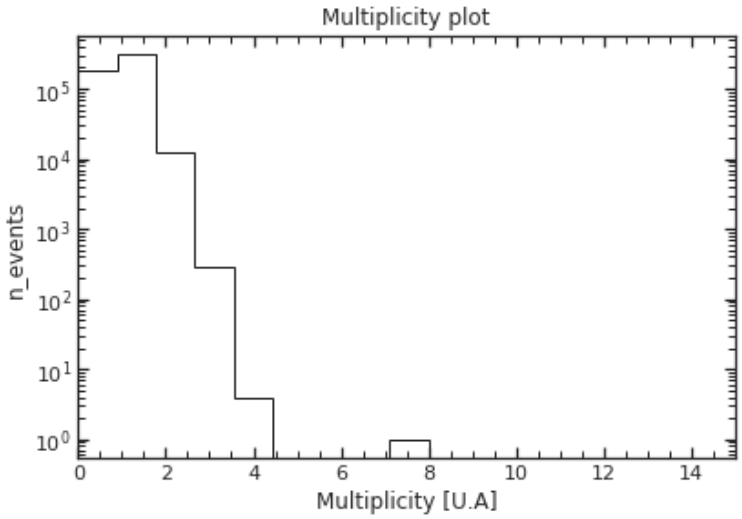

Figure 3.6: Multiplicity plot - Background run.

The CCM has been implemented, and the trend of events in the XY and XZ planes has been plotted. The trend is shown in *figures* 3.7 and 3.8. For completeness and reference, the positions of the PMTs in that plane are also shown in black.

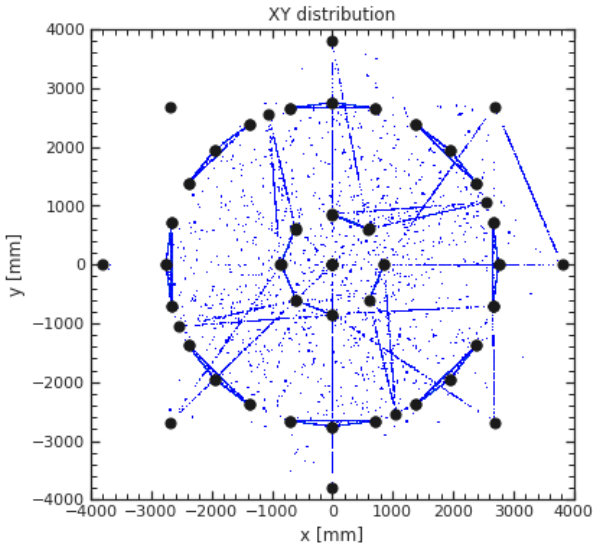

Figure 3.7: XY distribution - Background run.

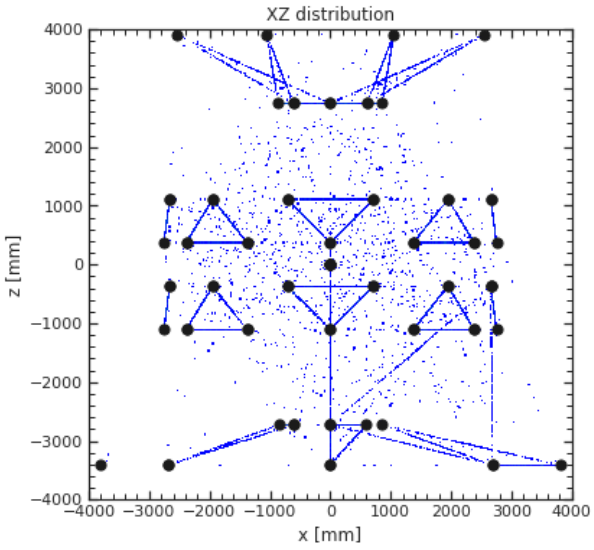

Figure 3.8: XZ distribution - Background run.

A pattern is highlighted in the background: upon closer examination of the connections between the PMTs and the highlighted "triangles" in the plots, a direct correspondence between correlated events and PMTs belonging to the same GCU can be observed. Therefore, it can be hypothesized that under self-trigger conditions, once an event is detected by one PMT, the other two PMTs in the same GCU also detect the same event. What results is a simultaneous "activation" of the three PMTs, forming a distinct pattern of "triangles" linking the PMTs.

Signal run

The run in question was taken with a much higher multiplicity (nPMT=10) compared to the previous one (self-trigger). In *figure* 3.9, a significantly different trend is observed compared to the previous section, but it is consistent with the predictions for this triggering value.

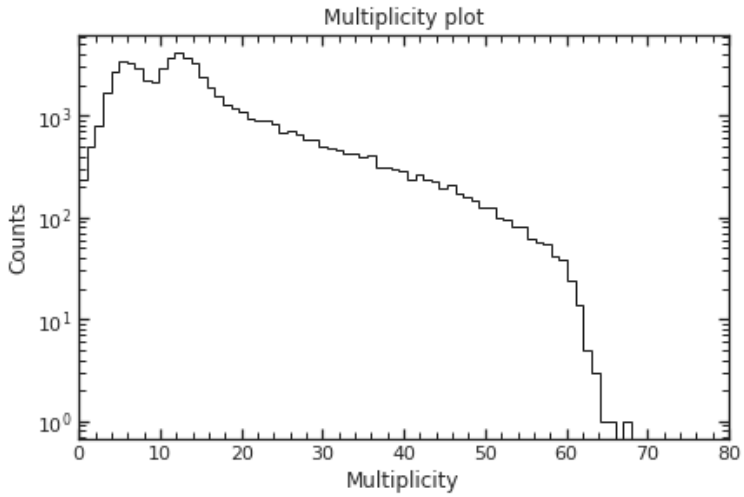

Figure 3.9: Multiplicity plot - Signal run.

The presence of multiplicity values less than 10 can be justified for two main reasons: first, because the threshold value used in the analysis phase may be different from that used in the data acquisition phase, and second, due to event cuts applied in the definition of the analysis function for this section. In particular, the multiplicity value was obtained by counting the detected events, defining a "valid event" as one that had at least one waveform (i.e., at least one hit time t_{hit}) and had a non-zero charge.

In this run, during data analysis, a triggering mask (mul) was applied incrementally, starting from a low value (mul=10) and gradually increasing it to a higher value (mul=50) to achieve better event localization. The total number of events analyzed without any mask imposed is 4,905,616. When a mul=10 mask is applied, the number of "valid" events decreases dramatically to 37,957, and further increasing the mask to mul=50 results in a very limited number of events, 786. In *Figure* 3.10 and subsequent figures, the trend of event position reconstruction is shown as the mul variable increases. Additionally, *Figure* 3.14 presents a three-dimensional reconstruction of the spatial position of events relative to the PMTs.

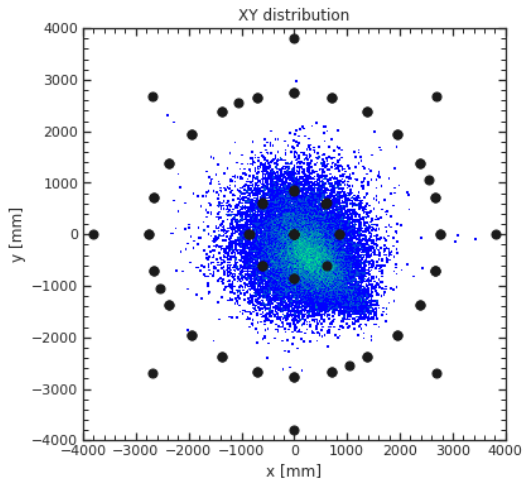

Figure 3.10: XY distribution (mul=10) - Signal run.

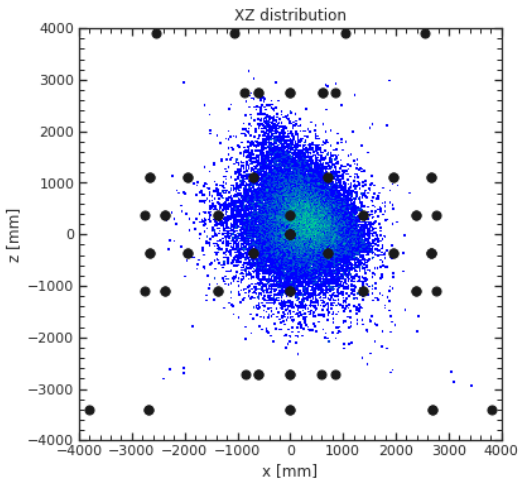

Figure 3.11: XZ distribution (mul=10) - Signal run.

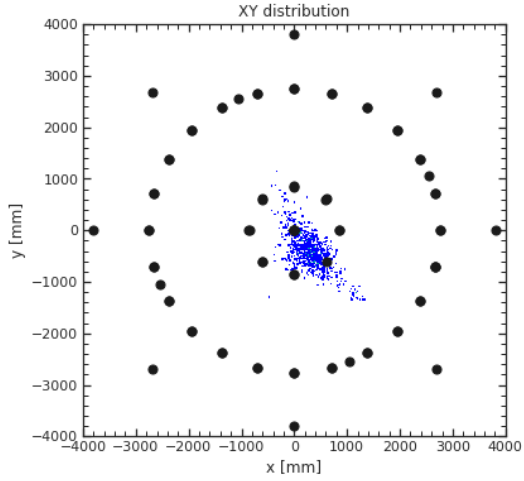

Figure 3.12: XY distribution (mul=50) - Signal run.

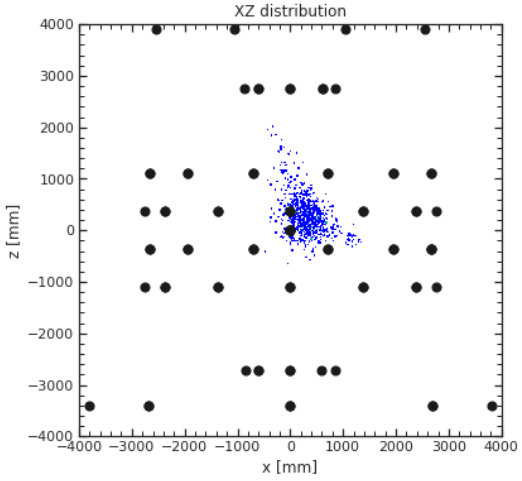

Figure 3.13: XZ distribution (mul=50) - Signal run.

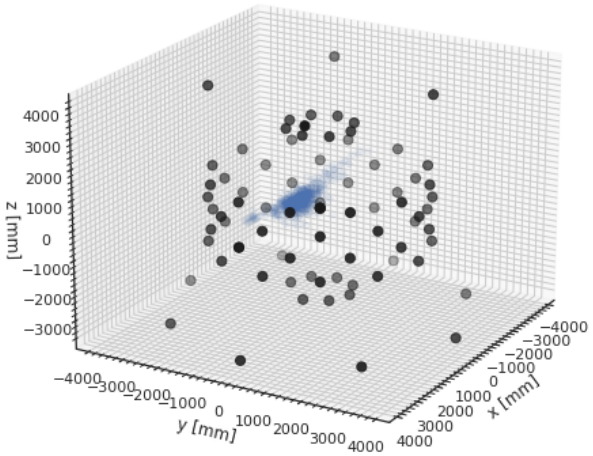

Figure 3.14: Event Vertex Reconstruction 3D (mul=50) - Signal run.

It is clearly evident that increasing the mask value results in a better localization of the event center position, in line with a decrease in the number of valid events. The brighter region in *Figure* 3.10 and 3.11 is well localized, and as the mask value increases, it becomes more apparent in the plots in *Figure* 3.12 and 3.13. In conclusion, by imposing a higher triggering threshold, it is possible to better localize the region of space where the interaction event occurs (i.e., the experimental condition where the LS ball is positioned). With a value of mul=50, the position of the LS ball is estimated to be

$$\begin{cases} r_x = (33 \pm 26)\, cm \\ r_y = (-44 \pm 28)\, cm \\ r_z = (24 \pm 26)\, cm \end{cases}$$

4 Conclusion

The present thesis work has undertaken the analysis of two datasets from the 1st OSIRIS dry run that occurred on June 9, 2023, and June 11, 2023: one background run and one signal run. The primary distinction between these runs lies in the experimental setup, particularly in the presence or absence of a scintillator liquid ball (LS ball) serving as the event vertex. Another difference pertained to the distinct trigger settings during data acquisition. Hence, the central goals of this thesis were to analyze the waveforms from the signal run and to characterize the interaction vertex, with a specific emphasis on the importance of the triggering system for accurate estimation.

Signal stability, conformity with pre-processed values, and the overall temporal behavior of the waveforms were examined, searching for data gaps and reconstructing physically meaningful quantities (such as the total accumulated charge in each event). Following a brief analysis of the total accumulated charge in the two runs (based on a data analysis selection of noise events), the Charge Center Method was introduced. An analysis code was implemented, and the vertex study for the two runs was presented. Particularly, responses were provided for the suspicious pattern in the background run (by revealing an interesting connection among PMTs belonging to the same GCU in a "Self Trigger" data taking) and the characterization of the interaction vertex (i.e., the position of the LS ball) in the signal run. This demonstrated the significance of the triggering system in accurately defining the position.

Therefore, this thesis has successfully achieved its goal of characterizing the 1st dry run, providing an intriguing foundation for the implementation of more sophisticated data analysis methods. Specifically, this work can be extended by employing various vertex position estimation techniques, utilizing different analysis methodologies compared to the CCM. As examples, the following methods are mentioned:

- Time to flight (TOF) method: this approach leverages the flight time of particles within the detector to estimate their point of origin. By utilizing the arrival times of signals from scintillation detectors, precise particle flight times can be calculated, thereby determining their distance from the event source. While TOF may be sensitive to fluctuations in flight times due to variations in materials traversed by particles, it could serve as a valid alternative to the charge center method in certain scenarios.

- Clustering Algorithms (CA): these algorithms seek to identify clusters of signals from correlated particles within the detector. Once clusters are identified, their centroids can be computed, potentially serving as estimates for event vertices. This approach could prove particularly useful in complex events involving multiple interacting particles within the detector.

- Convolutional neural networks (CNNs): CNNs are extensively used in image analysis within the field of machine learning and could be adapted to process detector data. By training a CNN on a large dataset of simulated events, precise vertex estimations could potentially be derived from detected information.

References

[1] **A. Abusleme et al., JUNO Collaboration**, "JUNO physics and detector," *Progr. Part. Nucl. Phys. 123 (2022) 103927*.

[2] **F. An et al.**, "Neutrino Physics with JUNO," *J. Phys. G43 (2016) 030401*.

[3] **A. Abusleme et al., JUNO collaboration**, "Sub-percent precision measurement of neutrino oscillation parameters with JUNO," *Chin. Phys. C46 (2022) 123001*.

[4] **P. Ghoshal and S.T. Petcov**, "Neutrino mass hierarchy determination using reactor antineutrinos," *JHEP 03 (2011) 058 + Addendum*, JHEP 09 (2012) 115.

[5] **S. Lau**, "CPR100 Design, Safety Performance and Operability," *Data Bay Nuclear Power Operations and Management Company*, 2011.

[6] **V. Cerrone et al.**, "Validation and integration tests of the JUNO 20-inch PMT readout electronics," *Nucl. Inst. Meth. A 1053 (2023) 168322*.

[7] **A. Abusleme et al., JUNO Collaboration**, "Mass testing and characterization of 20-inch PMTs for JUNO," *Eur. Phys. J. C82 (2022) 1168*.

[8] **A. Abusleme et al., JUNO and DAYA-BAY Collaborations**, "Optimization of the JUNO liquid scintillator composition using a Daya Bay antineutrino detector," *Nucl. Instrum. Meth. A 988 (2021) 164823*.

[9] **A. Abusleme et al., JUNO Collaboration**, "The design and sensitivity of JUNO's scintillator radiopurity pre-detector OSIRIS," *Eur. phys. J. C 81 (2021) 973*.

[10] **J. Liu et al.**, "Automated calibration system for a high-precision measurement of neutrino mixing angle θ_{13} with the Daya Bay antineutrino detectors," *Nucl. Instr. Meth. A 750 (2014) 19*.

[11] **R. Brun and F. Rademakers**, "ROOT-An object-Oriented Data Analysis Framework," *Nucl. Instr. Meth. A 389 (1997) 81*.

[12] **R. Callegari**, "Charaterization and tests of 39 channels of the JUNO large PMT electronics," *Master Thesis in Physics, Padova University, A.A. 2019/2020*.

[13] **W. Zhimin and N. Rodphai**, "Pre-OSIRIS Vertex Reconstruction," *Private communication*.