UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia "Galileo Galilei"

Master Degree in Physics

Final Dissertation

Characterization and tests of 39 channels of the JUNO large PMT electronics

Thesis supervisor

Prof. Alberto Garfagnini

Thesis co-supervisor

Dr. Filippo Marini

Candidate

Riccardo Callegari

Academic Year 2019/2020

Abstract

The Jiangmen Underground Neutrino Observatory (JUNO) is a multi-purpose neutrino experiment under construction in South China. It will make use of 20 kton of highly transparent liquid scintillator (Linear Alkyl-Benzene) contained in an acrylic sphere surrounded by 18000 20" PMTs and 25000 3" PMTs, respectively, aiming at an energy resolution better than 3% at 1 MeV. JUNO is expected to resolve the neutrino mass hierarchy, significantly improve accuracy of the solar oscillation parameters and make a significant impact to several other Neutrino Physics topics.

The JUNO Padova research unit is responsible for the design and development of the large PMT readout electronics; in order to verify its performances, a small JUNO mock-up system, equipped with 48 small size PMTs reading out the light generated by a 20 liter liquid scintillator detector, has been assembled and setup at the Legnaro INFN National Laboratories (LNL).

Within this context, my thesis work concerns the development of a software for the processing and the analysis of the readout data of the test system at LNL and the evaluation of the performances of the full JUNO electronics chain.

In particular, the first part of my thesis focuses on the development of a ROOT based program whose main task is reading the binary files produced by the DAQ system and extracting all the information they provide, such as, for example, the event timestamps or the acquired signal waveforms; this information, in addition to some other important characteristics of the acquired signals, such as the baseline contribution or the integrated charge, is then stored in ROOT files for further elaboration.

Another important activity of my thesis work concerns the development of several methods for data quality monitoring of the experimental setup; these methods include in particular a study on the stability, over a long period of time, of the trigger rate and of the baseline contribution to the signals. The data quality monitoring is performed by means of some ROOT based programs, which are developed for this purpose and properly optimised using data acquired in several long term acquisition runs.

Furthermore, studies on time correlation and synchronization of signals coming from different PMTs are carried out; these signals are acquired in runs with an external trigger on cosmic muons, generated by three external plastic scintillators.

These studies are useful in defining a specific procedure for reconstructing the total charge released by these particles in the liquid scintillator; a detailed analysis of the resulting spectrum allows to obtain, after a careful comparison with the expectation based on theoretical calculations and Monte Carlo simulations, some important results on the evaluation of the setup response.

Finally, some additional analyses concerning a more detailed calibration of the apparatus are presented and discussed.

Contents

Chapter 1

Introduction

Precise measurements of the θ_{13} neutrino oscillation parameter by the Daya Bay [1], Reno [2], and Double Chooz [3] experiments, have opened the path to the determination of the neutrino mass hierarchy. Indeed whether the ν_3 neutrino mass eigenstate is heavier or lighter than the ν_1 and ν_2 mass eigenstates is one of the remaining undetermined fundamental aspects of the Standard Model (SM) in the lepton sector [4]. Its determination would represent an important step in the formulation of a generalized model, and it would have a profound impact on the quest of the nature of neutrinos (Dirac or Majorana) and the search for a theory of flavor.

In fact, several experiments with solar, atmospheric, reactor, and accelerator neutrinos have determined, with remarkable accuracy, most of the necessary parameters that describe the three-generation oscillations. However, several crucial pieces are still missing: the neutrino mass hierarchy (or the ordering of the neutrino masses), the magnitude of the CP (charge conjugation and parity) phase δ_{CP}, the absolute scale of the neutrino mass, and whether neutrinos are Dirac or Majorana fermions. In particular, the study of the energy spectra of neutrinos coming from nuclear reactors at a medium distance is a very clean experimental method to determine the mass hierarchy without exploring neutrinos matter effects [5].

In this context, the Jiangmen Underground Neutrino Observatory (JUNO) is a reactor neutrino experiment under construction at Kaiping, Jiangmen in Southern China composed by a large liquid scintillator detector surrounded by thousands of large and small PhotoMultiplier Tubes (PMTs), a water cherenkov detector and a top tracker detector. Thanks to the large 20 kton active mass and unprecedented energy resolution (3% at 1 MeV), it will allow to determine the neutrino mass hierarchy with good sensitivity and to precisely measure the neutrino mixing parameters, θ_{12}, Δm^2_{21} and Δm^2_{ee}, below the 1% level [6].

Most of the information presented in this chapter has been retrieved from [6–10].

1.1 Neutrino oscillations

The standard electroweak model is a successful theory which not only unifies the electromagnetic and weak interactions but also explains almost all the phenomena of this nature observed at or below the electroweak scale. When this theory was first formulated by Weinberg in 1967 [11], neutrinos were assumed to be massless and hence there was no lepton flavor mixing. But just one year later the solar neutrinos were observed [12], and a deficit of their flux as compared with the prediction from the Standard Solar Model (SSM) was also established by Bahcall et al. [13, 14]. Such an anomaly turned out to be solid evidence for new physics beyond the SM, because it was found to be attributed to the neutrino oscillation, a spontaneous and periodic change from one neutrino flavor to another, which does not take place unless neutrinos have finite masses and lepton flavors are mixed.

In the SM the fact that the quark fields interact with both scalar and gauge fields leads to a non-trivial mismatch between their mass and flavor eigenstates, which is just the dynamical reason for quark flavor mixing and CP violation; therefore, one may expect a straightforward extension of the SM in which the phenomena of lepton flavor mixing and CP violation emerge for a similar reason. We

can in fact assume that the neutrinos ν_e, ν_μ and ν_τ (so-called flavor eigenstates) that couple with the gauge bosons $W^\pm$ through weak interactions are coherent super-positions of three mass eigenstates ν_i, i = 1, 2, 3, by the action of a unitary 3x3 matrix U:

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{pmatrix} \cdot \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix}$$

The lepton mixing matrix U (usually denoted as the PMNS matrix for Pontecorvo, Maki, Nakagawa and Sakata who were the pioneers of the field) can be parametrized in terms of the three angles θ_{13}, θ_{23}, and θ_{12} and the CP phase δ_{CP}:

$$U = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \cdot \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta_{CP}} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta_{CP}} & 0 & c_{13} \end{pmatrix} \cdot \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} e^{i\alpha} & 0 & 0 \\ 0 & e^{i\beta} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

where e.g. $c_{ij} = cos\theta_{ij}$, $s_{ij} = sin\theta_{ij}$ and α and β are the so-called Majorana phases that are decoupled from the phenomenon of neutrino oscillation.

Since the flavor eigenstates are a super-position of the three mass eigenstates, it happens that a flux of neutrinos of a well-determined flavor measured at a specific position has a non-null probability of exposing a different flavor if measured after a sufficiently large distance; this probability depends on the parameters of the PMNS matrix, accordingly to the flavors of the considered neutrinos and the distance between the positions of the first and the second measurements, but also on the difference between the squared mass eigenvalues Δm^2_{ij}. This description clearly suggests that neutrinos have a non-null mass and that the three mass eigenvalues must be different, otherwise the oscillation phenomenon should not occur.

parameter	best fit value $\pm$ 1σ	3σ range
$\sin^2\theta_{12}$	$0.304^{+0.012}_{-0.012}$	(0.270, 0.344)
θ_{12} (degrees)	$33.48^{+0.77}_{-0.74}$	(31.30, 35.90)
$\sin^2\theta_{23}$	$[0.451^{+0.001}_{-0.001}]$ or $0.577^{+0.027}_{-0.035}$	(0.385, 0.644)
θ_{23} (degrees)	$[42.2^{+0.1}_{-0.1}]$ or $49.4^{+1.6}_{-2.0}$	(38.4, 53.3)
$\sin^2\theta_{13}$	$0.0219^{+0.0010}_{-0.0011}$	(0.0188, 0.0251)
θ_{13} (degrees)	$8.52^{+0.20}_{-0.21}$	(7.87, 9.11)
δ_{CP} (degrees)	251^{+67}_{-59}	(0, 360)
Δm^2_{21} $\times 10^{-5}$ eV^2	$7.50^{+0.19}_{-0.17}$	(7.03, 8.09)
(normal) Δm^2_{31} $\times 10^{-3}$ eV^2	$+2.458^{+0.046}_{-0.047}$	(+2.325, +2.599)
(inverted) Δm^2_{32} $\times 10^{-3}$ eV^2	$-2.448^{+0.047}_{-0.047}$	(-2.590, -2.307)

Table 1.1: Neutrino oscillation parameters taken from [15]. For the atmospheric mass-squared splitting Δm^2_{3x}, the best fit results for both the normal and the inverted mass hierarchy are shown.

Table 1.1 summarizes the present knowledge of neutrino masses and mixings including neutrino mixing angles θ_{ij}, the CP phase δ_{CP}, and neutrino mass-squared differences $\Delta m^2_{ij} = m^2_i - m^2_j$, based on recent fits.

As we can notice from the pattern of neutrino masses and mixings schematically shown in Figure 1.1, at the present time we cannot decide whether the ν_3 neutrino mass eigenstate is heavier or lighter than the ν_1 and ν_2 neutrino mass eigenstates in nature: the scenario in which the ν_3 is heavier is referred to as the normal mass hierarchy (NH), while the other scenario, in which the ν_3 is lighter, is referred to as the inverted mass hierarchy (IH).

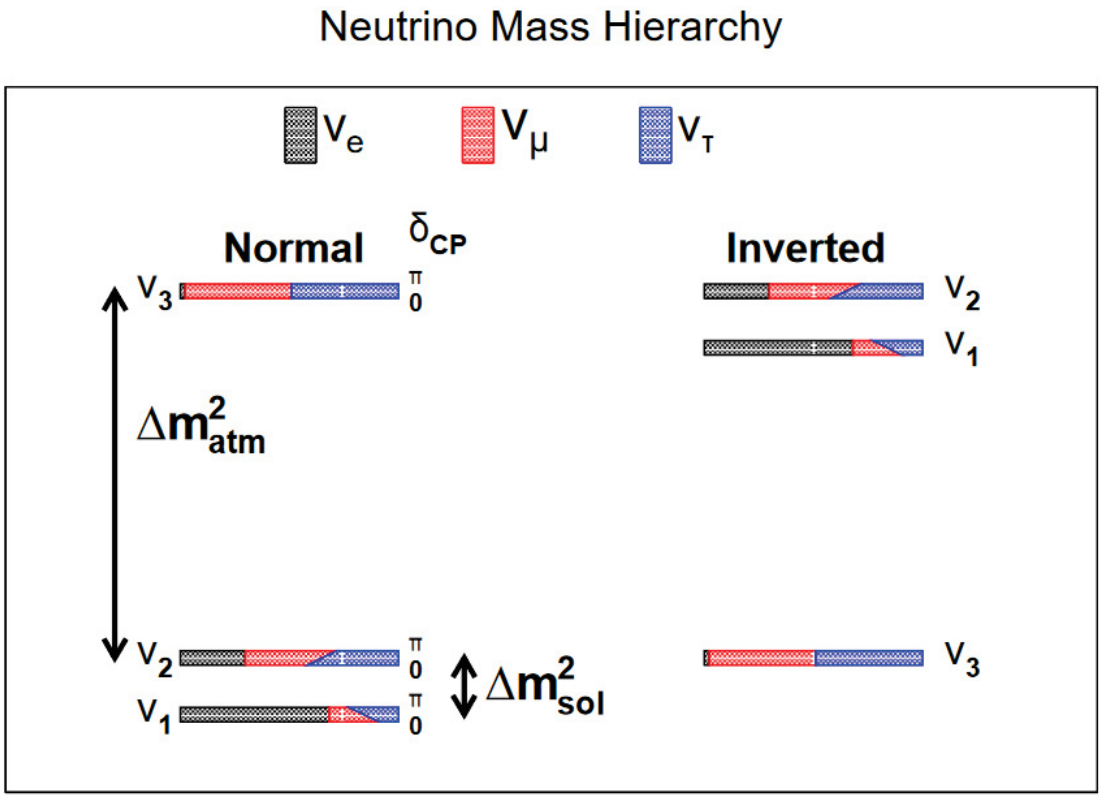

Figure 1.1: Pattern of neutrino masses for the normal and inverted hierarchies is shown as mass squared. Flavor composition of the mass eigenstates as the function of the unknown CP phase δ_{CP} is indicated. $\Delta m^2_{atm} \sim |\Delta m^2_{31}| \sim |\Delta m^2_{32}|$ and $\Delta m^2_{sol} \sim \Delta m^2_{21}$ stands for the atmospheric and the solar mass-squared splitting, respectively. Picture taken from [7].

Even if the difference between the two mass orderings depicted in Figure 1.1 seems to be striking, the determination of the true mass hierarchy (MH) is still very challenging. However, despite the difficulties, it is vitally important to solve this problem, since MH determination would strengthen, or eliminate, roughly half of the proposed particle physics models, and correspondingly a great effort is devoted to its solution.

Similarly to most of the parameters describing neutrino mass and mixing, the neutrino MH can be accessed through the neutrino flavor oscillation. As shown in Table 1.1, there are two small parameters in the neutrino oscillation description: the mixing angle θ_{13} ($sin^2\theta_{13} \sim 0.022$) and the ratio $\Delta m^2_{21}/\Delta m^2_{31} (\sim 3\%)$. Due to this feature, most oscillation results are reasonably well described in the framework of only two neutrinos mixing; therefore, in this case, the probability of flavor change along a distance L in vacuum is given by:

$$P(\nu_l \to \nu_{l'}) = sin^2 2\theta \cdot sin^2\Big(1.27 \cdot \frac{\Delta m^2(eV^2) \cdot L(m)}{E(MeV)}\Big)$$

and, obviously, the sign of Δm^2 (the mass hierarchy) cannot be determined in such case.

Therefore, in order to find effects that are sensitive to the sign of Δm^2_{31} or Δm^2_{32}, one has to either go beyond the vacuum oscillation or go beyond the simple framework of two-neutrino mixing. Correspondingly, there are two direct ways to determine MH: in the first method, one uses the propagation in matter where an additional phase from the interaction between neutrinos and the matter constituents makes the hierarchy determination possible (thanks to the Mikheyev-Smirnov-Wolfenstein (MSW) effect or the matter effect [16–18]), while in the second method one explores the small difference Δm^2_{21} between Δm^2_{31} and Δm^2_{32}.

The approach adopted by JUNO for determining the neutrino mass hierarchy consists in detecting neutrino oscillations at a medium baseline ($\sim$ 50 km) from the source; in particular, the experiment will detect electron antineutrino interactions in the main detector thanks to the inverse beta decay reaction $\overline{\nu}_e + p \rightarrow e^+ + n$. The energy deposited by the positron annihilating in the liquid scintillator is strictly proportional to the energy of the incoming antineutrinos. The time coincidence between the positron interaction and the subsequent neutron capture on protons allows to identify efficiently the neutrino interaction, even in the presence of uncorrelated background. The electron antineutrino survival probability can be written as:

$$P_{ee} = 1 - sin^2 2\theta_{13} \cdot (cos^2\theta_{12} sin^2\Delta_{31} + sin^2\theta_{12} sin^2\Delta_{32}) - sin^2 2\theta_{12} \cdot cos^4\theta_{13} sin^2\Delta_{12}$$

where $\Delta_{ij} = \dfrac{\Delta m^2_{ij} L}{4E_\nu}$; its trend as a function of the distance from the nuclear power plant is represented in Figure 1.2.

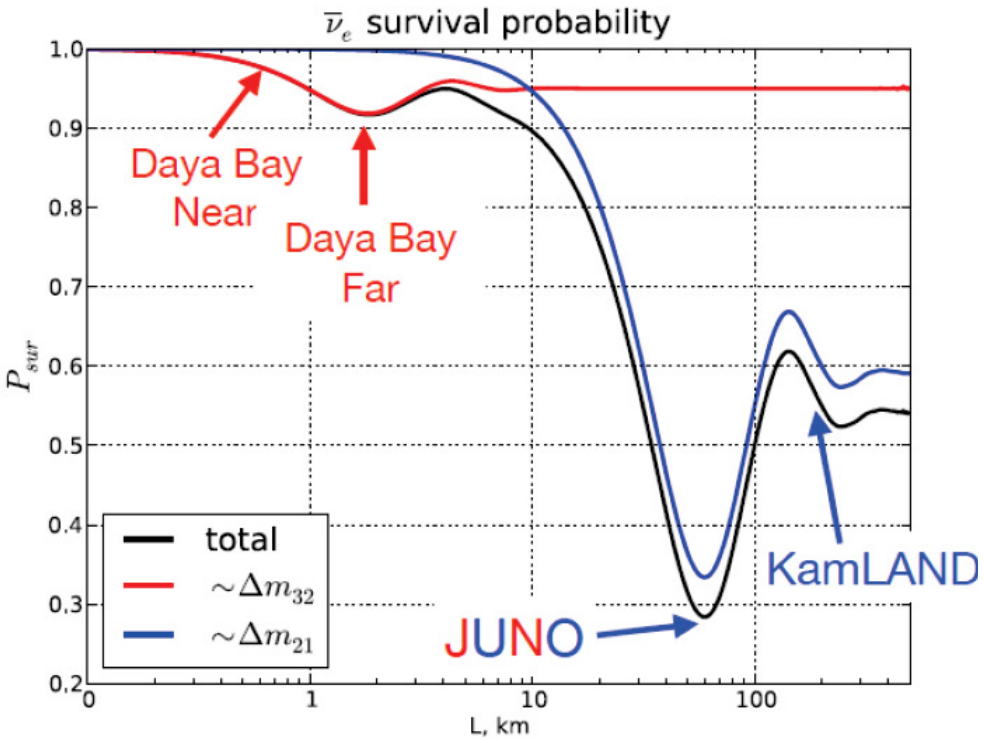

Figure 1.2: The $\overline{\nu}_e$ survival probability as a function of the distance from the nuclear power plant (NPP). The distance of other neutrino reactor experiments (KamLAND and Daya Bay) from the NPP are indicated and compared with the JUNO distance at 53 km. Picture taken from [9].

Considering the approximation $\Delta m^2_{32} \approx \Delta m^2_{31}$, we can rewrite this formula in order to make explicit the mass hierarchy dependence:

$$\begin{aligned} P_{ee} = 1 &- cos^4\theta_{13} sin^2 2\theta_{12} sin^2\Delta_{21} - sin^2\theta_{13} sin^2|\Delta_{31}| \\ &- sin^2\theta_{12} sin^2 2\theta_{13} sin^2\Delta_{21} cos 2|\Delta_{31}| \\ &\pm (sin^2\theta_{12}/2) \cdot sin^2 2\theta_{13} sin^2 2\Delta_{21} sin^2 2|\Delta_{31}| \end{aligned}$$

The sign flip in the last term is due to the neutrino mass hierarchy possibility: the direct hierarchy gives a positive contribution, while it is negative for the inverse mass ordering. The effect is small, but fortunately not negligible due to the relatively large value for θ_{13}.

The effect of the neutrino mass hierarchy on the electron antineutrino energy spectrum, measured at a medium-baseline of $\sim$ 53 km, is shown in Figure 1.3, where the normal ordering is represented with a blue line and the inverted one with a red line.

The finite resolution of detectors tends to wash away the characteristic features of the ripples that can be distinguished in the oscillated spectrum in Figure 1.3; if the resolution deteriorates further the peaks and throats of the spectrum would start to disappear, making it very hard or almost impossible to unfold the neutrino mass hierarchy from the collected spectra. Detailed studies have shown that assuming a large scintillator detector with a 20 kton mass, a 3% energy resolution and an exposure time of 6 years, can provide a 4σ discrimination power; systematic effects, such as the shape uncertainty of the reactor spectrum and the background discrimination, would affect the result and worsen the discrimination power, therefore they must be studied and evaluated in detail [6].

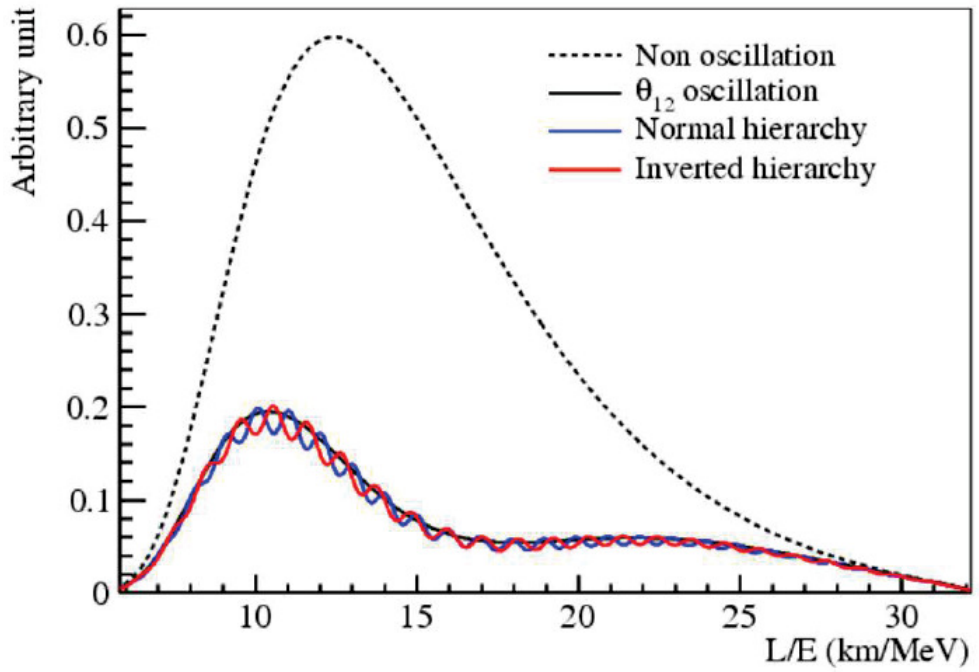

Figure 1.3: The black line shows the un-oscillated spectra of reactor anti-neutrino at the distance of 53 km (distance betweeen JUNO site and nuclear power plants), while the red and the blue line show the oscillated spectra for the two mass orderings (blue is the normal order and red is the inverted order) [6].

Beyond mass hierarchy and precision determination of the neutrino oscillation parameters, a large liquid scintillator detector can provide fundamental results on many topics in astroparticle physics, like supernova burst and supernova diffuse neutrinos, solar neutrinos, atmospheric neutrinos, geo-neutrinos, nucleon decay, indirect dark matter searches and a number of additional exotic searches. A reference to the rich physics program of JUNO can be found in [6].

1.2 The Jiangmen Underground Neutrino Observatory experiment

JUNO [6] will detect reactor antineutrinos from nuclear power plants (NPP). The mass hierarchy determination requires equal baselines from the detector to all the reactor cores to avoid cancellation of the oscillation dephasing effect. The site location is optimized to have the best sensitivity for the mass hierarchy determination, which is at 53 km from both the Yangjiang and Taishan NPPs, where six and four nuclear cores are planned, respectively. The neutrino detector is a liquid scintillator detector with a 20 kton fiducial mass, deployed in an underground laboratory about 700m deep. The experimental site and the detector will be described in the following.

1.2.1 JUNO experimental site

The JUNO experiment locates in Jinji town, Kaiping city, Jiangmen county, Guangdong province. The geographic location is East longitude 112°31'05" and North latitude 22°07'05". 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: Guangzhou, Hong Kong, Macau, Shenzhen, and Zhuhai, all in about 200 km drive distance, as shown in Figure 1.4.

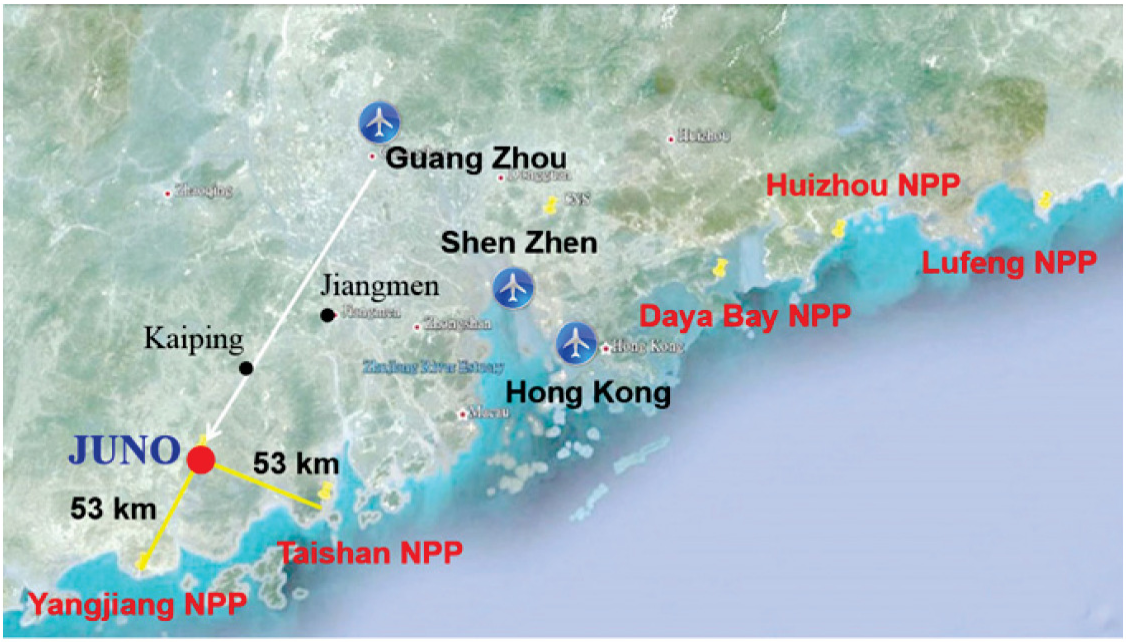

Figure 1.4: Location of the JUNO site. The distances to the nearby Yangjiang NPP and Taishan NPP are both 53 km. Daya Bay NPP is 215 km away. Three metropolises, Hong Kong, Shenzhen, and Guangzhou, are also shown.

The experimental site is at 53 km from the Yangjiang NPP and Taishan NPP. Yangjiang NPP has six reactor cores of 2.9 GW_{th} each (thermal power). All cores are the 2nd generation pressurized water reactors CPR1000 [19]. The distances between any two cores of Yangjiang NPP are between 88 m and 736 m. Taishan NPP has planned four cores of 4.59 GW_{th} each. All cores are the 3rd generation pressurized water reactors EPR. The distances between any two cores are between 252 m and 1110 m. The total thermal power of the Yangjiang and Taishan NPPs will be 35.73 GW_{th}. Daya Bay complex 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 GW_{th}. The Daya Bay complex is 215 km away from the JUNO detector, and will contribute about 2.8% of the reactor antineutrino events. There is no other NPP or planned NPP in 500 km around the JUNO experimental site. The thermal power of all cores and the baselines are listed in Table 1.2. The distances from the detector site to the Yangjiang and Taishan cores are surveyed with a Global Positioning System (GPS) to a precision of 1 meter. All these NPPs are constructed and operated by the China General Nuclear Power Group (CGNPG).

Cores	YJ-C1	YJ-C2	YJ-C3	YJ-C4	YJ-C5	YJ-C6
Power(GW)	2.9	2.9	2.9	2.9	2.9	2.9
Baseline(km)	52.75	52.84	52.42	52.51	52.12	52.21
Cores	TS-C1	TS-C2	TS-C3	TS-C4	DYB	HZ
Power(GW)	4.6	4.6	4.6	4.6	17.4	17.4
Baseline(km)	52.76	52.63	52.32	52.20	215	265

Table 1.2: Summary of the thermal power and baseline to the JUNO detector for the Yangjiang (YJ) and Taishan (TS) reactor cores, as well as the remote reactors of Daya Bay (DYB) and Huizhou (HZ) [6].

In absence of high mountains in the allowed area where the sensitivity to the mass hierarchy is optimized, the detector will be deployed in an underground laboratory under the Dashi hill. The activities of the ^{238}U, ^{232}Th, and ^{40}K in the rock around the experimental hall are 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, estimated by simulation with the surveyed mountain profile taken into account [6].

1.2.2 JUNO experimental apparatus

The JUNO detector, shown in Figure 1.5, consists of a central detector (CD) filled with liquid scintillator (LS) readout by a PhotoMultiplier Tube (PMT) system. VETO detector and a calibration system complete the detector structure. In order to reach $3\%/\sqrt{\mathrm{E(MeV)}}$ energy resolution, the central detector will build a super acrylic sphere and a stainless-steel truss, which will hold 20-kton liquid scintillator, 18000 20-inch PMTs and 25000 3-inch PMTs. The VETO detector will be divided into a top tracker and a water Cherenkov detector. The calibration system will provide different methods for JUNO calibration, with its operation being of a fundamental importance to obtain the required design resolution and to have a checkup of the energy scale of the detector.

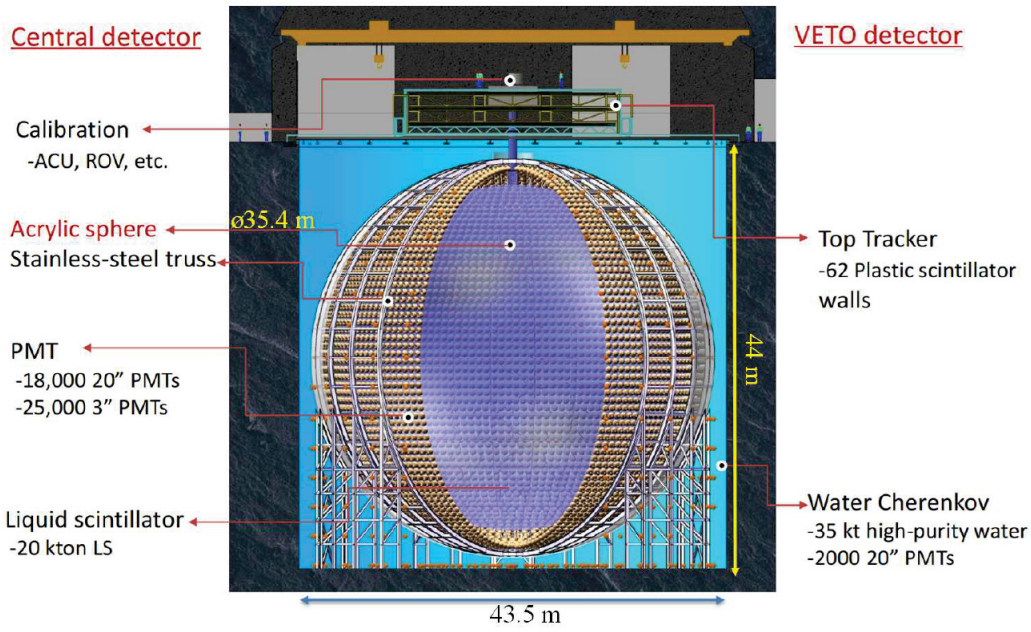

Figure 1.5: A schematic view of the JUNO detector. The main parts are indicated: the central detector composed by the liquid scintillator, the acrylic sphere, the PMTs and the calibration system; the veto detector composed by the water Cherenkov pool and the top tracker.

In order to reach the required energy resolution, the following requirements have been set:

- PMT photocathode covergage $\geq 75\%$;
- PMT photocathode quantum efficiency $\geq 35\%$;
- LS attenuation length ≥ 20 m at 430 nm, which corresponds to an absorption length of 60 m with a Rayleigh scattering length of 30 m [20].

JUNO liquid scintillator

The LS has similar recipe as the Daya Bay LS without gadolinium loading. Linear Alkyl-Benzene (LAB), a straight alkyl chain of 10-13 carbons attached to a benzene ring [21], is used as the detection medium due to its excellent transparency, high flash point, low chemical reactivity, and good light yield. The liquid scintillator also consists of 3 g/L 2,5-diphenyloxazole (PPO) as the fluor and 15 mg/L p-bis-(o-methylstyryl)-benzene (bis-MSB) as the wavelength shifter; its total density is 0.859 g/ml. 20 kton LS is contained in a spherical container of radius of 17.7 m. The light emitted by the LS is watched by about 18000 large 20-inch PMTs and 25000 small 3-inch PMTs, which are necessary to achieve the resolution requirements of the experiment. In fact, to quantify in a basic and clear approach the resolution characteristics and requirements of JUNO, we can resort to the following simplified relation:

$$\frac{\sigma_{\mathrm{E}}}{\mathrm{E}} = \sqrt{\left(\frac{a}{\sqrt{E}}\right)^2 + b^2 + \left(\frac{c}{\sqrt{E}}\right)^2}$$

where a represents the stochastic term governed via the overall maximization of light (i.e. through the photo-coverage, PMT Quantum Efficiency (QE) and features of the LS), and b and c are non-stochastic terms controlled by the minimization of systematic effects. Such a control in the detector will be achieved in two ways, by an accurate calibration strategy and by the cross check measurements performed via the auxiliary system of the small 3-inch PMTs. They will allow a full complementarity of the event identification, with particular emphasis on time resolution, spatial reconstruction, dynamic range and triggering strategy. The PMTs are installed on a spherical structure of a radius of 19.5 m, and submerged in a buffer liquid to protect the LS from the radioactivity of the PMT glass.

JUNO PMT system

Photon detectors measure the scintillation light created by interactions of the neutrinos with LS and they are key components for accomplishing the physics goals of JUNO. Important requirements for photon detectors used in JUNO include high detection efficiency for scintillation photons, large area, low cost, low noise, high gain, stable and reliable, long lifetime.

Two types of large PMTs will be used in the JUNO experiment:

- the Hamamatsu R12860-HQE is a 20-inch large photomultiplier presenting a 460 mm diameter hemispherical bi-alkaline photocathode light-sensitive for wavelength that ranges between 300 nm and 650 nm, with a maximum sensitivity around 400 nm and the relative quantum efficiency around 30% [22]. Its chain of amplification has a linear focusing geometry structure, which at 2 kV has a $1 \cdot 10^7$ nominal gain;
- the NNVT MCP-PMT comes with a 20-inch glass window and incorporate a Micro-Channel Plate (MCP) in place of the conventional discrete dynodes. The MCP consists in a two-dimensional array of channels with a 6-20 μm diameter bundled in parallel. Each channel acts as an independent electron multiplier. The photoelectrons emitted from the photocathode enter the channels of the MCP and impinge on the inner wall where they are multiplied by means of secondary emission. This process is repeated along the channels, and finally a large number of electrons are collected by the anode as an output signal. The quantum efficiency of this photomultiplier is lower than the R12860, 22.86%, while the gain is the same, $1 \cdot 10^7$.

JUNO readout electronics

The average number of photoelectrons (PE) generated by a single large PMT ranges from one PE for low-energy events up to thousands PE. In both extreme cases, the photoelectrons time profiles have to be determined and the energy of the event has to be measured. The energy represent a fundamental parameter for the neutrino mass hierarchy determination. A basic limitation on the energy resolution arises from the statistics of detected PE and it must not be worsened significantly by the effect of electronics. For an energy release of 1 MeV in the central detector, which corresponds to an average of 1100 PE, a 3% design energy resolution is expected. The layout of the fully-submerged readout scheme is sketched in Figure 1.6.

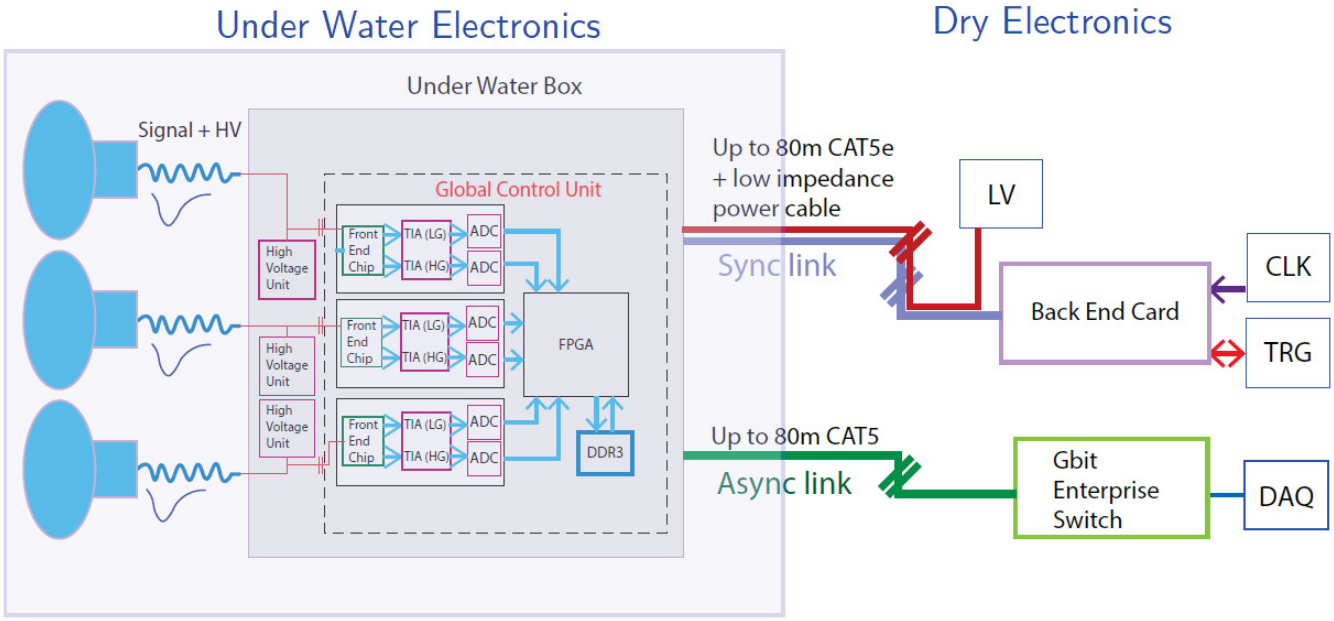

Figure 1.6: JUNO readout electronics architecture.

Considering the underwater architecture sketched in Figure 1.6, we can notice that groups of three PMTs are connected to water-tight housings which contain the essential front-end and readout electronics of the system:

- three modules that generate the High Voltage (HV) from a low-voltage input for operating the 3 PMTs independently;
- the Global Control Unit (GCU), integrating in particular Analogue to Digital Units (ADUs) and an FPGA; the GCU performs the data digitization, buffering and processing, and, monitors and controls all the relevant parameters.

The readout electronics main task is the reconstruction of the event. Because of the distributed nature of the complete event information, each PMT can only collect one fragment of it. The acquired data are analyzed online by the GCUs that, employing a specific threshold level, determine if there are any possible fragments of the event and consequently send a trigger request to the Central Trigger Unit (CTU). Subsequently, the CTU makes a global correlation in order to generate a trigger validation which is then sent back to the GCUs, which finally transfer, through Ethernet, all the fragments of the event to the Data Acquisition system (DAQ).

The underwater scheme implements an intelligent PMT concept minimizing the deteriorating effects of analog signal transmission over long cables.

1.3 The integration test facility for the JUNO large PMT electronics

Due to the fact that the JUNO experiment Front-End electronics will be inaccessible after its underwater installation, reliability is a fundamental aspect to consider during development; in particular, in order to achieve the sensibility required for the mass hierarchy determination, less than 0.5% defective channels over 6 years are tolerated. Besides the reliability of the employed electronic components, another important aspect is the optimization of the whole electronics chain, including both hardware and firmware; therefore, it is necessary to have a full characterization of the whole system response, which can be obtained by means of an integration test facility that allows to test the JUNO electronics performances.

For this purpose, a small JUNO mock-up system, equipped with 48 PMTs, has been built and set in operation at the *Laboratori Nazionali di Legnaro* (LNL), in Italy; this system plays a crucial role in testing the JUNO experiment electronics.

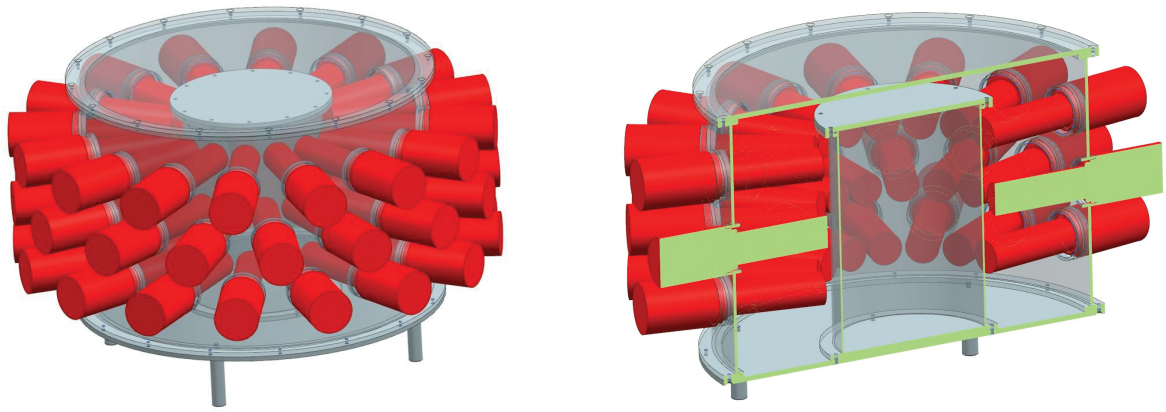

Figure 1.7: Mechanical design of the structure of the test system; the internal cylindrical vessel (grey) is surrounded by the 48 PMTs (red); the PMTs are inside the plastic structure (grey), while their bases are outside (red).

This test facility, shown in Figure 1.7 and Figure 1.8, is composed by:

- an internal cylindrical vessel, made of transparent Plexiglas, containing $\sim$ 17 l of liquid scintillator; the liquid scintillator is composed by a solvent (LAB) doped with Poly-Phenylene Oxide (PPO) and p-bis-(o-MethylStyryl)-Benzene (bisMSB), used as wavelength shifter to match the PMT response; the same components are used by the Daya Bay neutrino experiment [1], with concentrations of 3 g/l of PPO and 15 mg/l of bisMSB;
- the internal cylinder is surrounded by 48 Philips XP2020 PMTs (diameter of $\sim$ 5 cm) equipped with their proprietary bases; as it can be seen in Figure 1.7, the PMTs are placed around the lateral surface of the cylindrical vessel, divided in three rings, each with 16 PMTs; the employed PMTs are suitable for high and medium energy physics where the number of photons to be detected is very low and they feature a good linearity, a very low background noise (typical anode dark noise is up to $\sim$ 900 counts/s) and extremely good time characteristics and good single electron spectrum. The 48 PMTs are finally connected, in groups of 3, to 16 Global Control Units (GCUs);
- a black plastic structure of cylindrical shape that surrounds the inner vessel and the PMTs, whose main task is to support the PMTs; in addition, it also prevents external light from penetrating and induce misleading results;
- 3 plastic scintillators, one placed on the top of the apparatus and the other two on the bottom of the black plastic structure, which are used to provide a trigger mechanism based on cosmic rays by means of the AND coincidence of these three scintillators; the distance between the scintillating plate on the top of the setup and the first scintillator placed below the vessel is 52 cm, while the distance between the two scintillators on the bottom is 3 cm;

- a LED light has been introduced on the top of the setup using an optical fiber; it provides light at 405 nm and it can be used to calibrate the detector.

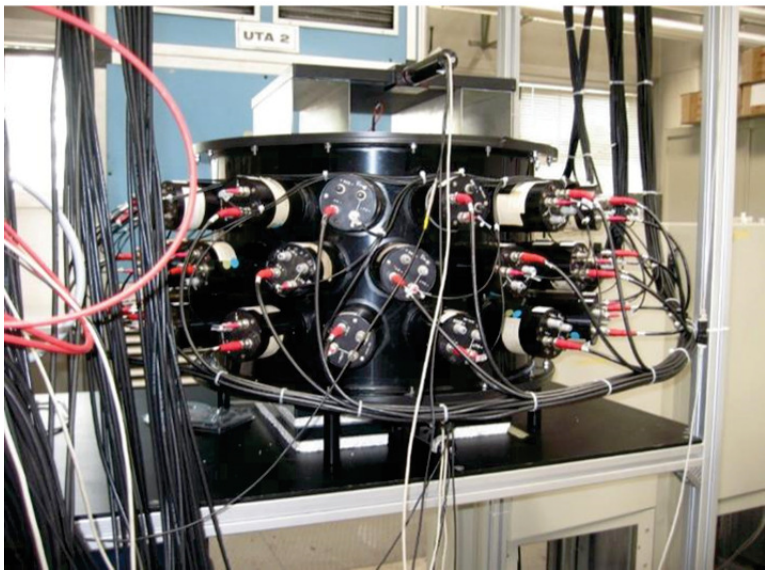

Figure 1.8: Picture of the integration test facility operating at the LNL; only the black plastic structure and the PMT bases are visible.

1.3.1 The electronics chain of the Legnaro test facility

The photons produced in the liquid scintillator are collected and converted by each PMT into an electric signal, which has to pass through several steps of the system electronics chain, before being stored in the raw data files produced by the DAQ program.

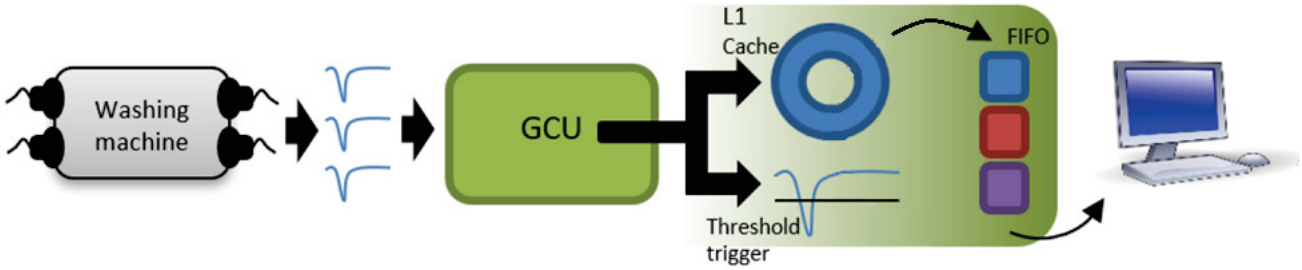

Figure 1.9: Scheme of the electronics chain of the Legnaro test system.

The electronics chain, whose schematic description is reported in Figure 1.9, is composed by the following main steps:

1. all the 48 PMTs are connected, in groups of 3, to 16 GCUs; in particular, during all my thesis work, only 13 GCUs were actually active, which correspond to 39 available acquisition channels (see Figure 1.10 and Figure 1.11); all the channels acquire their signals asynchronously;
2. for each active channel, the analog signal produced by the PMT is digitized by an Analog-to-Digital Unit (ADU) and subsequently doubled, so that one of the two signal copies is registered with its GCU timestamp into a L1 cache, while the other one is analyzed with a specific threshold trigger algorithm;

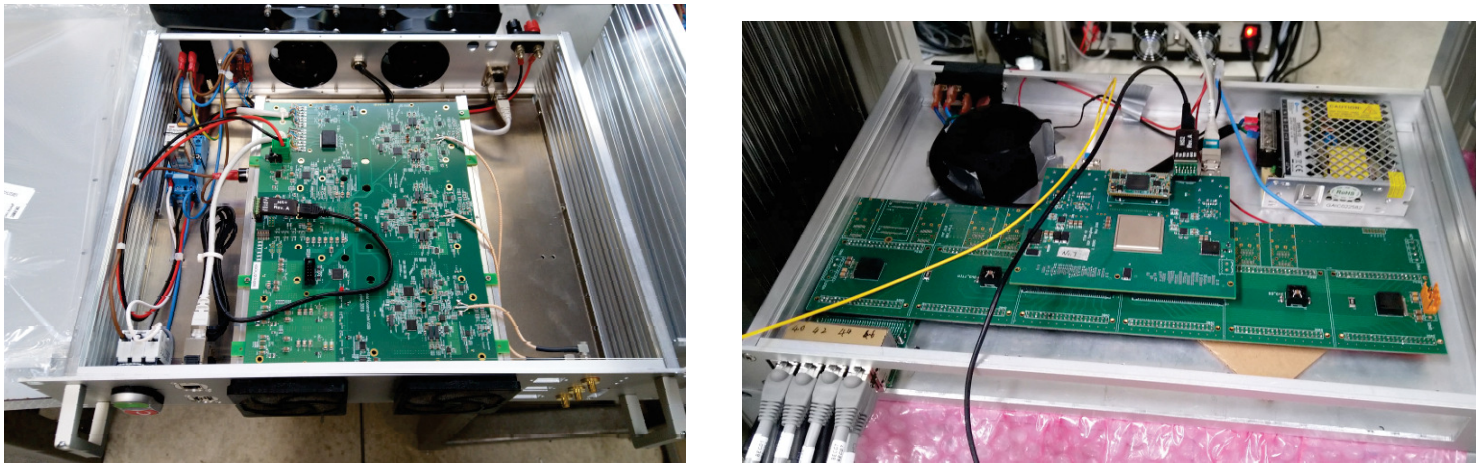

Figure 1.10: Left panel: picture of a single Global Control Unit. Right panel: picture of a Back End Card.

3. if the signal exceeds the threshold, the GCU sends a trigger request to a Back End Card (BEC, see Figure 1.10) which then takes a "global trigger" decision, based on the specific trigger logic that has been chosen, and it sends back to all the GCUs the same timestamp, chosen by the trigger validation; clearly, for this step, GCUs and BEC must be properly synchronized in time. In particular the global trigger logic could be based either on a logic OR of all the 3 channels of a single GCU, or on the multiplicity of the acquired event, i.e. a logic AND between two or more channels;
4. after the trigger validation by the BEC, the firmware extracts the signals with the selected timestamp from the L1 cache and it transfers them into a First-In-First-Out (FIFO) unit; the content of the FIFO is then transferred to the PC through an ethernet switch, where the DAQ program stores the raw data accordingly to a specific structure, which is discussed in detail in the following chapter.

Figure 1.11: Picture of the 13 GCUs set in operation.

In addition to this trigger validation procedure, the system implements an additional external trigger which can be used, for instance, to trigger on cosmic muon events thanks to the three plastic scintillators positioned above and below the experimental setup.

Chapter 2

Raw data processing

In this chapter a detailed description of the structure of the acquired raw data, and of the information they provide, is reported; in addition, we will present step by step the processing software that has been developed, in the first part of this thesis work, in order to read, extract and store all the information carried by the raw data, underlying the main features and implemented methods.

2.1 Raw data file format

The data acquired by the JUNO DAQ software from each one of the 16 GCUs of the experimental setup at the LNL have a fixed structure: they are grouped in *data packets* and each packet is composed of a variable number N_w of sequences of 16 bits, which are referred to as *words*, and stored in binary files. Every data packet is wrapped by two sequences of 8 words, the *header* and the *trailer*, which allow to recognise and separate each packet from the others, providing also important information which uniquely characterises that specific data packet. The header is composed of:

- the header starting word ($0x805a$, in hexadecimal numeral system), which is fixed for all the data packets;
- the GCU channel number, which can be 0, 1 or 2;
- the data packet size, including header and trailer, given in units of 8 words;
- the cyclic trigger count;
- the timestamp, a 64 bits sequence which represents the time reference given in units of 8 ns.

the trailer is composed of:

- the trailer starting sequence, a 6-words pattern fixed for all the data packets;
- the GCU ID number;
- the trailer ending word ($0x0869$), which is fixed for all the data packets.

The header and trailer structure is reported in Figure 2.1.

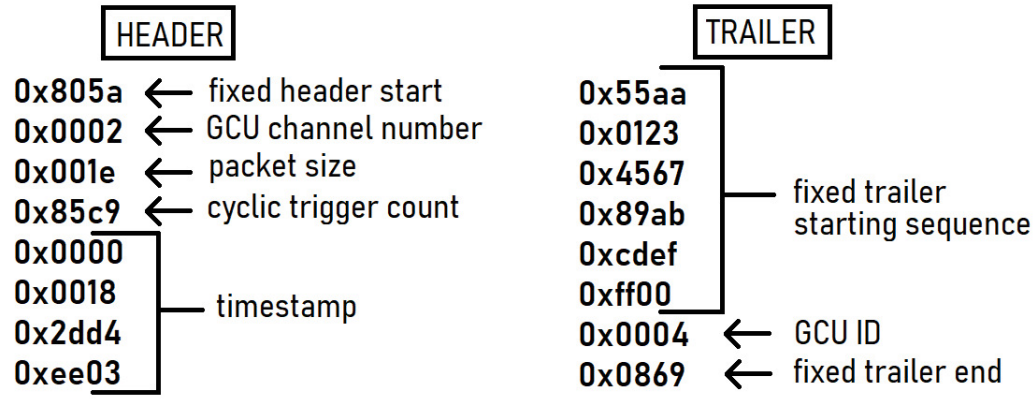

Figure 2.1: Data packet header and trailer structure (each word is reported as an hexadecimal number).

Between header and trailer we find the actual data, stored as a sequence of (N_w-16) words, which represents the signal waveform collected in a time window of (N_w-16) ns, starting from the reference timestamp reported in the header; a typical waveform is plotted in Figure 2.2.

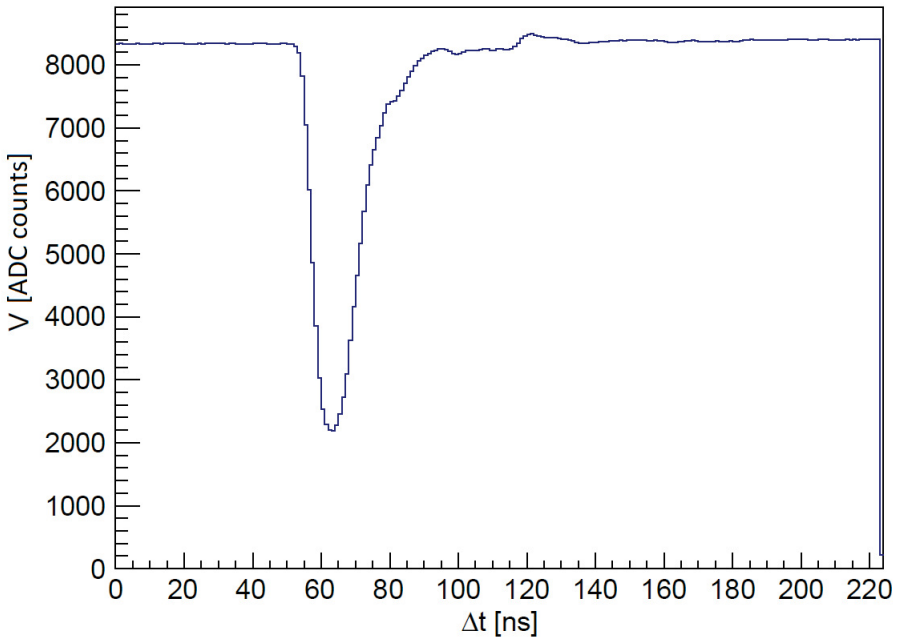

Figure 2.2: Example of waveform plot from a single data packet; the voltage V is given in units of *ADC counts* and a difference $\Delta V = 1$ *ADC count* corresponds to $\Delta V = 75\mu V$.

2.2 Raw data file reading

The reading of the binary files is performed by means of a ROOT [24] based program, which has been developed to identify and separate each single data packet from the others, extracting all the carried information; the information is then stored in a ROOT `TTree` object, which will be then used for further analysis. During the reading and storing phase, the program performs integrity checks on the raw data packets, to identify possible missing words, bit sequence mismatches or data corruption. A schematic description of the main tasks performed by the program is reported in Figure 2.3.

2.2.1 The Boyer-Moore algorithm

An important characteristic of the long bit sequence contained in the binary files is that it is unlikely to start exactly with the fixed header starting word (i.e. $0x805a$, in hexadecimal numeral system) of the first data packet, but it contains a limited sequence of bits which belong to a truncated data packet produced in previous data acquisitions; in order to correctly start the reading of the whole file, this first sequence has to be eliminated, searching for the actual useful data packet. This research is performed collecting the first kB of information, divided into 16-bits words, contained in the binary file in a C++ `std::vector<uint16_t>`, and searching for the first fixed trailer starting sequence, composed of 6 words as described in Section 2.1. The program looks for the whole trailer starting sequence, instead of searching directly for the header starting pattern, because it could be also contained in the waveform payload; searching for the whole trailer starting sequence allows to drastically reduce the probability of a wrong tagging of the start of the first useful data packet.

In order to perform this task the program employs the so called *Boyer-Moore* algorithm [25], which is an efficient string-searching algorithm developed by Robert S. Boyer and J Strother Moore in 1977; it searches for a short string, called the *pattern*, contained in a longer string, called the *text*. The Boyer–Moore algorithm uses information gathered during some preprocessing steps to skip sections of the text, resulting in a lower constant time factor than many other string search algorithms; its key feature is the research method based on the matching on the tail of the pattern rather than on the head, allowing to skip along the text in jumps of multiple characters, instead of searching every single character in the whole text.

This algorithm is opportunely applied to our case, where the 16-bit words are effectively considered as strings, so that the searched pattern corresponds to the trailer starting sequence and the

investigated text corresponds to the first kB of information of the binary file. The implementation of this searching method is done by means of the `std::search` template [26], and the related `std::boyer_moore_searcher` [27], defined in the `algorithm` standard library of C++17 and directly applied to the `std::vector<uint16_t>` filled with the first kB of information of the binary file. Once the position of the first trailer starting sequence, and, as a consequence, the following header start, is found, the program reads all the data packets in cascade until the end of the file is reached, employing a user-developed C++ class.

During this first pattern searching phase, and for the subsequent whole reading phase, the program also properly swaps the bit sequence of each word, in order to obtain for the data the correct endianness accordingly to the description presented in Section 2.1, which is exactly the opposite of the one provided by the employed JUNO DAQ software (for example the expected header starting word $0x805a$, is provided as $0x5a80$ by the JUNO DAQ software).

By the way, in several data acquisitions we performed, as it is discussed in the following chapters, we employed also an additional Python-based DAQ software, which provides the raw data in the endianness we would expect accordingly to Section 2.1.

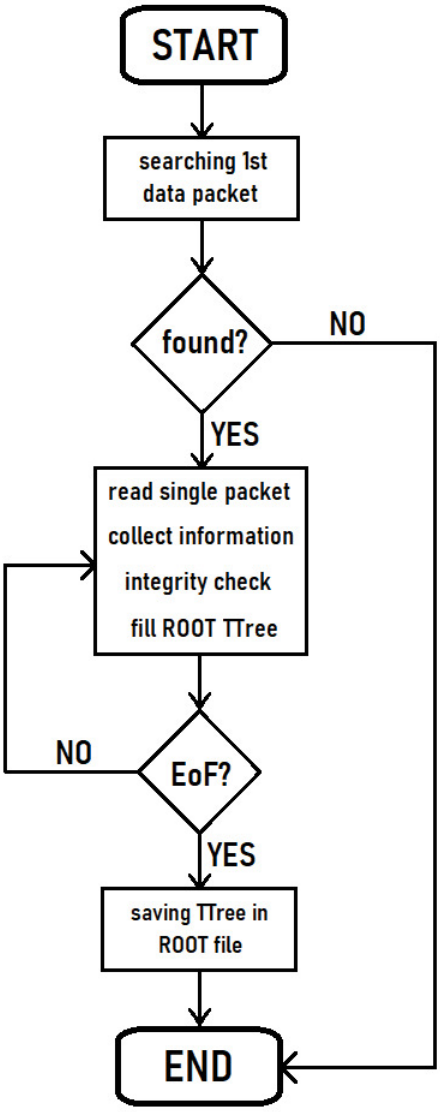

Figure 2.3: Schematic description of the main tasks of the raw data processing code.

2.2.2 Extracting and collecting the data

As anticipated in Section 2.2.1, the program that has been developed reads each data packet one after the other up to the end of the file and it collects all the interesting information that they carry, checking also their integrity.

This task is performed with support from a user-developed C++ class, called `gcu_reader` (a schematic Unified Modeling Language (UML) model is reported in Figure 2.4), which in turn employs a user-developed C++ structure, called `gcu_raw`, used as a container for temporarily storing all the information provided by a single data packet; therefore, some of the most important variables implemented in `gcu_raw` are used for storing information on the GCU and channel number of the acquired waveform, its trigger timestamp and the actual data packet size, but there are also some Boolean flags, used for spotting undesired data corruption problems.
A more detailed, technical description of the implemented methods is reported in Appendix A.

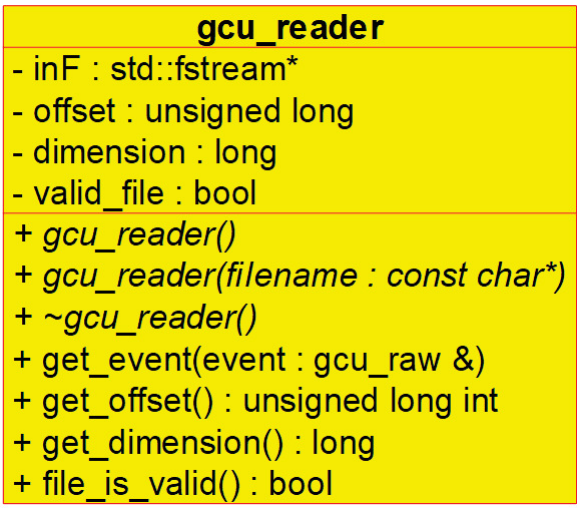

Figure 2.4: Unified Modeling Language (UML) model of the `gcu_reader` class.

2.2.3 Saving in a ROOT `TTree` object

As we have already said in the previous sections, the processing software stores the data in a ROOT `TTree` object, that is finally written in a ROOT file. The ROOT `TTree` class is designed for storing a large number of same-class objects and it is optimized to reduce disk space and enhance access speed; it consists of a hierarchy of branches and leaves which allows a flexible organization of data, that is fundamental for the analysis that will be considered in the following sections.

The `TTree` object written in the output ROOT file is composed of three different branches:

gcu_raw which contains all the `gcu_raw` structures that have been filled by the `get_event` method of the `gcu_reader` class, as discussed in detail in Appendix A;
baseline which contains important information on the baseline contribution to every signal waveform;
charge which stores the value of the integrated charge extracted from each waveform;
threshold flag which describes if the waveform amplitude exceeds a specific threshold;
timing info which contains some timing information, such as the waveform risetime and falltime.

The details on how the baseline, the charge and the timing information is extracted and employed will be detailed in the following chapters.

2.3 Interesting results on trailer consistency

The implementation of the `valid_data` flag and of the `actual_pkt_size` attribute in the `gcu_raw` structure allow to detect undesired bugs in the binary files produced by the GCUs, which are mainly due to a lack of consistency in the header and trailer of the data packets.

For example, thanks to these variables, we have been able to spot some important bugs affecting specifically the trailer consistency:

- we noticed that a small, but non negligible, fraction of data packets had an actual size of 8 words, instead of the expected N_w words, and we easily discovered that this was due to the presence of packets containing a double trailer (see Figure 2.5), which should not be present in the binary files;
- a smaller percentage of data packets contained an anomalous trailer sequence where some of the expected words were missing, because they were substituted by several repetitions of part of the trailer starting sequence (see Figure 2.6); this led to an anomalous actual packet size of $2 \cdot N_w$ words, which was easily spotted thanks to the `actual_pkt_size` variable.

Performing some specific acquisitions setting all the channels in auto-trigger, i.e. each channel acquired on its own threshold trigger validation, and with very different trigger rates from each other, highlighted that these anomalies were more likely to occur with a high rate of acquisition, while their presence was negligible in the other cases.

```
7220 7720 7120 7320 aa55 2301 6745 ab89
efcd 00ff 0100 6908 aa55 2301 6745 ab89
efcd 00ff 0100 6908 5a80 0000 1e00 8d41
0000 0ea0 d336 9ddf 5420 5620 5320 5c20
5820 5220 5420 5120 5620 5120 5220 5520
5620 5720 5620 5420 5620 5920 5720 5620
5720 5520 5520 5620 5220 5420 5620 5920
5520 5220 5720 5720 5420 5620 5520 5620
5620 5720 5720 5320 5420 5320 5720 5420
```

Figure 2.5: Example of anomalous data packet with double trailer.

```
6620 6b20 6920 6c20 aa55 2301 6745 ab89
efcd 00ff efcd 00ff efcd 00ff efcd 00ff
0000 ad4e 2a0d 138d 5c20 5820 5a20 5820
5220 5420 5720 5720 5b20 5720 5920 5920
5b20 5a20 5520 5720 5a20 5b20 5920 5820
5720 5c20 5520 5920 5820 5720 5620 5620
5220 5520 5720 5620 5920 5c20 5720 5a20
5620 5720 5520 5d20 5a20 5820 5b20 5620
```

Figure 2.6: Example of anomalous data packet with trailer containing pattern repetitions.

Chapter 3

Raw data analysis

Once the processing software has been developed and optimized, the focus has shifted to a first analysis of the raw data, involving in particular the development of several ROOT based methods, which have been employed to perform some data quality monitoring (DQM) of the GCU raw data and that have been useful for the implementation of run time quality checks in further acquisitions.

This chapter will present in detail, after a brief summary of all the data acquisitions performed during the whole thesis work, some of the DQM studies that have been carried out and some of the main results that have been obtained on important features of the acquired signals, such as the estimate of the baseline contribution or the reconstruction of the trigger rate trend in time; finally, we will present some of the first studies on time correlation of signals acquired by different PMTs of the experimental setup, which has been fundamental in the physics data analysis presented in the following chapters.

3.1 Summary of all the performed data acquisitions

RUN	Duration/Date	Trigger Logic	DAQ software	Trigger Rate
1	20 days (from 12.03 to 01.04)	GT (OR of GCU0 channels) - intermediate threshold	JUNO	$\sim$ 11 Hz
2	10 days (from 08.05 to 18.05)	ET	Python-based	$\sim$ 4.5 Hz
3	1 day - (22.05)	AT - two different threshold values	Python-based	different for each PMT
4	4 days (from 29.05 to 01.06)	ET	Python-based	$\sim$ 4.5 Hz
5	few hours - (04.06)	GT (GCU0 CH2) - low threshold	Python-based	$\sim$ 15 Hz
6	few hours - (05.06)	GT (AND of CH1 of GCU0, 3, 4) - intermediate threshold for each	Python-based	$\sim$ 3.5 Hz
7	15 minutes - (28.08)	GT (AND of 3 GCUs) - threshold of $6\sigma_B$	JUNO	$\sim$ 1.97 kHz
8	15 minutes x 4acq. - (28.08)	GT (AND of 3 GCUs) - threshold of $6\sigma_B$	JUNO	$\sim$ 3 kHz

Table 3.1: Brief summary of all the main data acquisitions performed during the thesis work, with details on duration, trigger logic and employed DAQ software.

During the whole thesis work several data acquisitions have been performed for different purposes and, consequently, in different trigger conditions; in particular, we used mostly a global trigger (GT) logic based on BEC validation, accordingly to the description reported in Section 1.3.1, and an external trigger on cosmic muons (ET), based on the use of the three external plastic scintillators; but also some acquisitions with all the channels set in auto-trigger (AT), i.e. each channel acquires on its own threshold trigger validation, have been carried out.

A brief summary of the main data acquisitions is reported in Table 3.1, while a more detailed description of characteristics and purposes of each run will be presented in the following sections.

3.2 Data Quality Monitoring

In order to perform the first studies related to the data quality monitoring, we acquired data for a long period of time (RUN1 in Table 3.1, 20 days, starting from 12.03.2020, stopping in 01.04.2020), dividing them in 20 one-day long acquisition runs. Since in this first phase we were not interested in specific physics signals, but only in the response of the acquisition system, we set a high threshold level, in order to suppress undesired dark noise contributions, and we used an OR of all the 3 channels of the first GCU as trigger logic. For each of the 38 out of the 39 GCU channels (GCU1, CH2 was not active during this acquisition), the JUNO DAQ software produced 20 files, each of which contained the data of a single day acquisition, opportunely extracted by the processing software, described in Chapter 2.

3.2.1 Trigger rate stability study

We performed a first study on the event trigger rate stability counting the number of collected events every 10 minutes and analysing its trend over a single day of acquisition. As can be seen in the example reported in Figure 3.1, the trigger rate seems to be quite stable during the whole day, except for the first point in each run; thanks to this study we were able to spot this anomaly which affects in the same way all the 38 channels in every run and which is connected to a huge time difference between the timestamps of two consecutive signals among the first collected in each run.

We discovered that this anomaly is simply due to a feature of the GCU firmware, which stores in the internal FIFO a fixed and limited number of waveforms belonging to previous acquisitions, which are not eliminated from the FIFO until a new acquisition starts; as a consequence, the first data packets of each file acquired before the two affected by this anomalous time difference will not be considered in the analysis from now on.

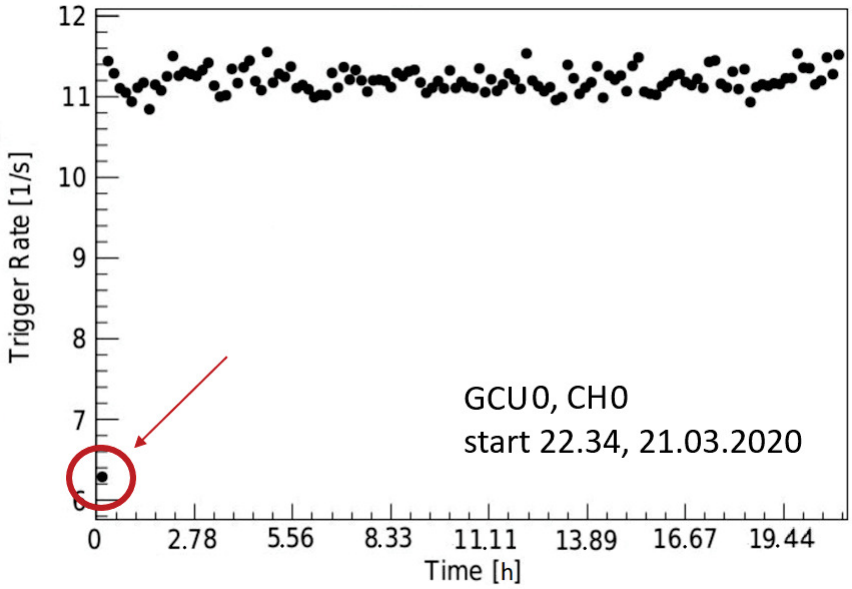

Figure 3.1: Preliminary study on the stability of the trigger rate.

3.2.2 Further studies on the trigger rate stability

Another study of the trigger rate stability has been performed, following a little more complex procedure; first of all, for each of the 38 channels, we calculated the time difference Δt between two consecutive timestamps over 1 hour of acquisition. By plotting the results in a histogram, an exponential distribution can be extracted, similar to the one that is reported in Figure 3.2.

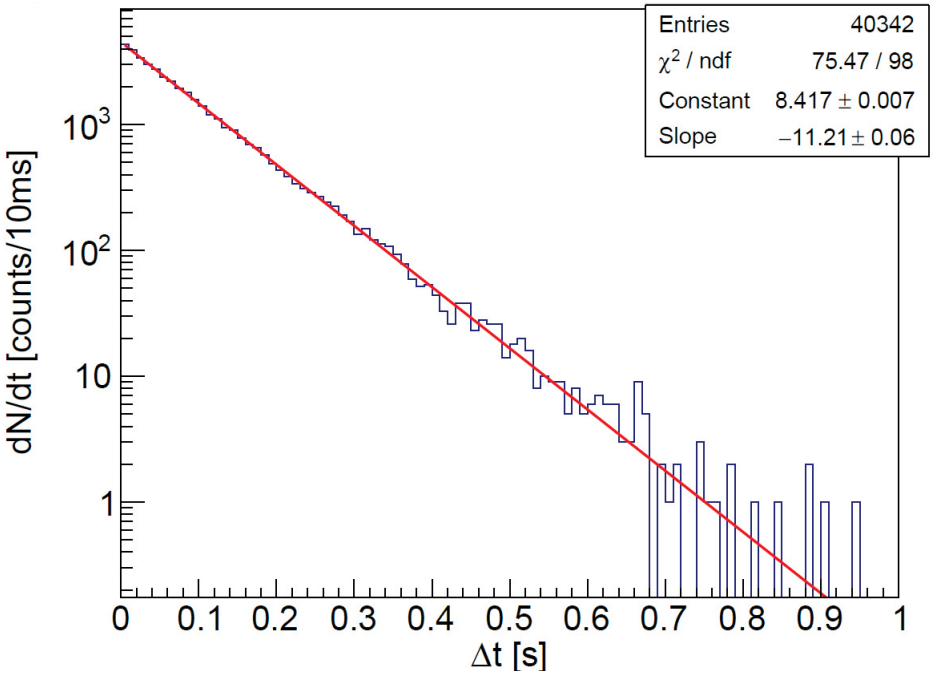

Figure 3.2: Distribution of the consecutive timestamp differences Δt within a time interval of 1 hour, with exponential fit (GCU0, CH0).

In order to obtain an estimate of the trigger rate for every hour of the 20 days of acquisition, we performed a fit on these distributions with the function $Ae^{-\lambda\Delta t}$, where:

- λ represents the trigger rate;
- A is a constant.

Plotting the results for each channel allows the analysis of the trigger rate behaviour over the whole period of acquisition. Of course, since the BEC triggers all the channels equally, the trend does not differ between different GCU channels.

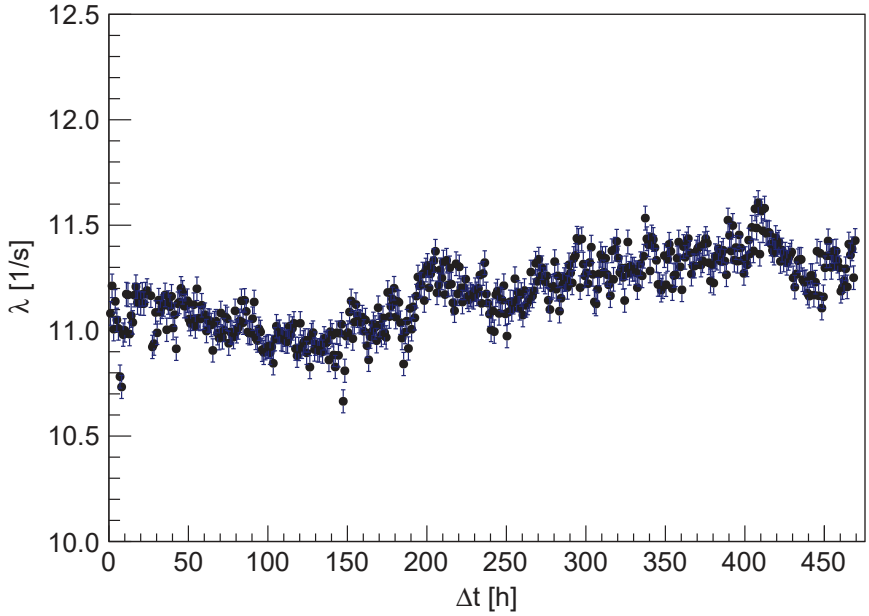

Figure 3.3: Time distribution of the trigger rate λ over 20 days of acquisition (GCU0, CH0).

As we can notice in Figure 3.3, even if there are not significant variations of λ throughout a single day, we can notice a non-negligible shift in its value during the whole acquisition. We are not able to exactly determine the reasons of this shift, which is likely to be caused by a simple variation of temperature during the considered 20 days.

3.2.3 Study on the baseline mean value stability

An estimate of the baseline contribution to the signal waveform is extracted, during the processing of the raw data; the baseline is calculated as the mean value of the contents of the first N_{pre} bins (highlighted in Figure 3.4), where N_{pre} is a variable number which corresponds to the time width, in ns, of the pre-trigger region, i.e. the time interval which precedes the exceeding of the threshold level, which can be dynamically adjusted during acquisition.

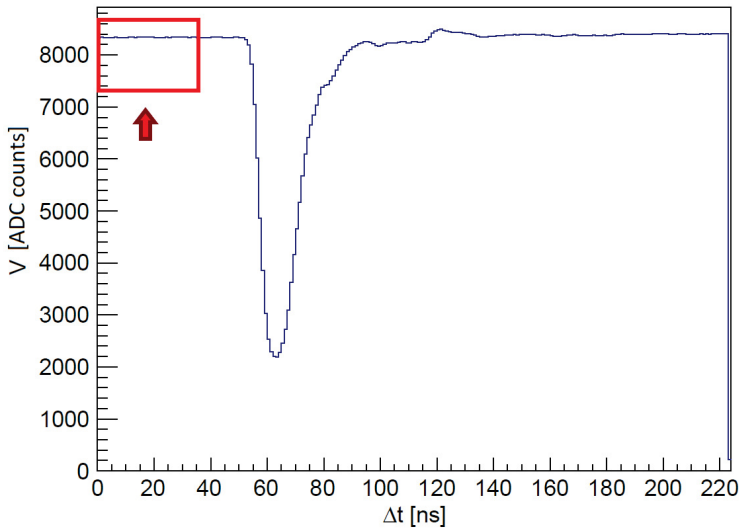

Figure 3.4: Example of acquired waveform with the highlighted pre-trigger region used in the baseline calculation.

The stored baseline estimates are then grouped within time intervals of 30 minutes and for each of these groups a distribution, as the one reported in Figure 3.5, is produced; in order to evaluate the behaviour of the baseline contribution during the whole data acquisition, the obtained histograms are fitted with a Gaussian function, extracting their centroids μ_B and sigmas σ_B.

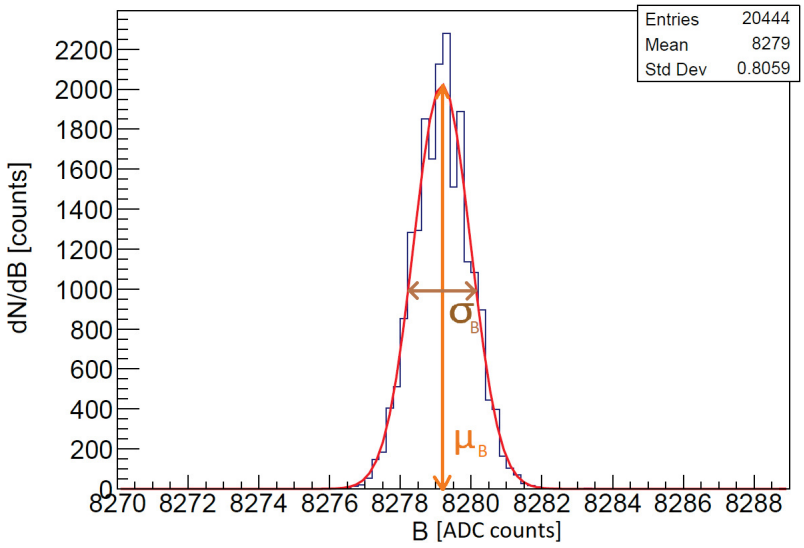

Figure 3.5: Distribution of the baseline estimates within a time interval of 30 minutes, with Gaussian fit (GCU0, CH0).

The values of μ_B and σ_B, for each channel, are then both plotted as functions of the acquisition time, as it is shown, as an example, in Figure 3.6; all the GCU channels analysed so far have a similar trend. Looking to the reported results and considering that 1 ADC count corresponds to a voltage difference of 75 μV, we can notice that both μ_B and σ_B are quite stable during the whole period of acquisition.

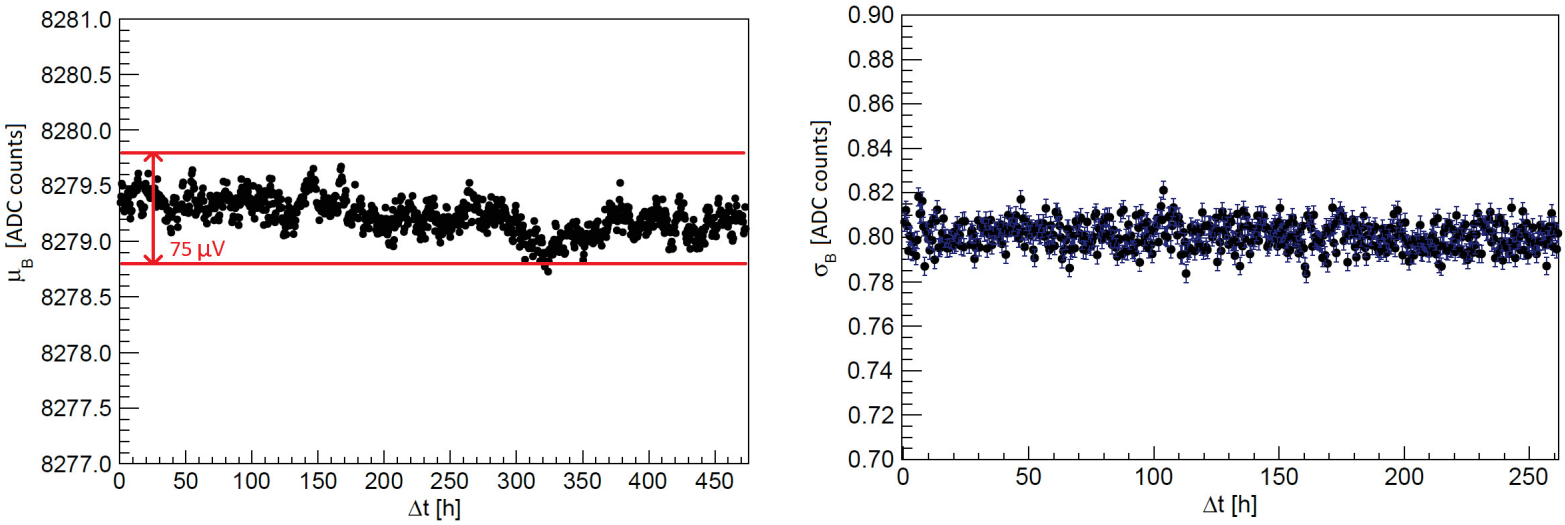

Figure 3.6: Time distribution of B mean value μ_B and B sigma σ_B (for GCU0, CH0).

3.2.4 Study on the baseline spread

To further improve the data taking quality monitoring software, different approaches have been pursued:

- study on the stability of the baseline standard deviation S, which is evaluated for each waveform during the raw data processing phase, as well as the mean baseline value, considering the first N_{pre} bins of the pre-trigger region;
- study on the stability of the baseline dispersion D, calculated for each waveform as the difference between the maximum and the minimum bin content among all of the N_{pre} bins of the pre-trigger region.

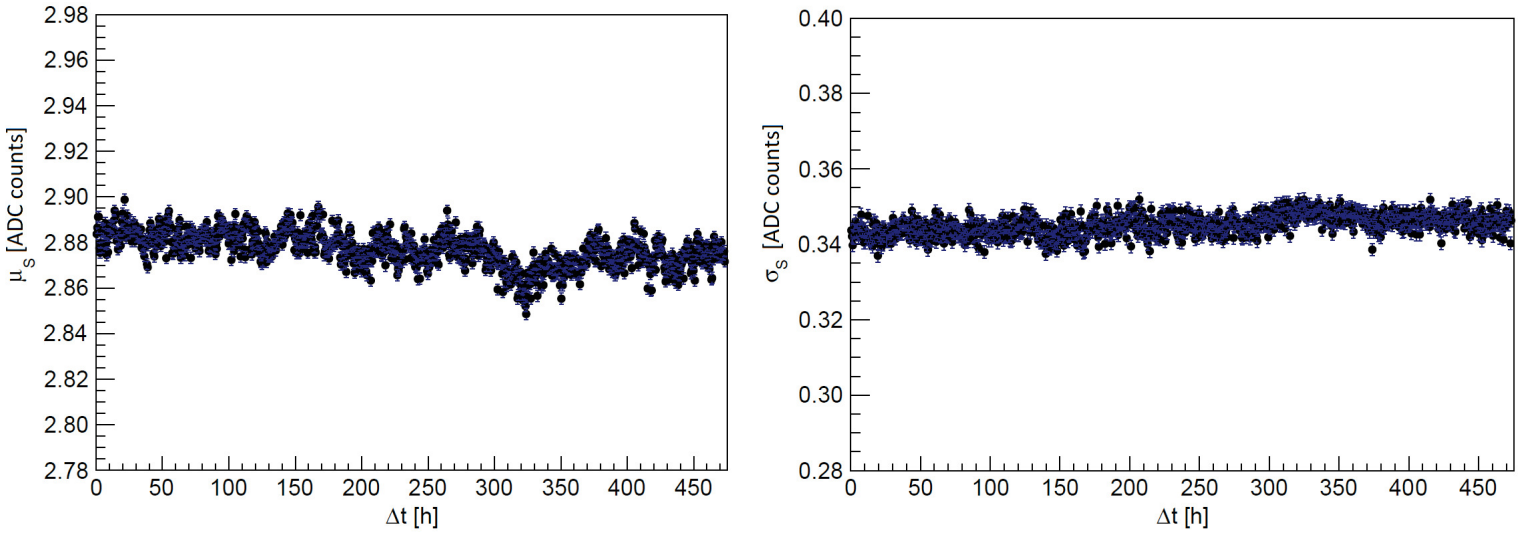

Figure 3.7: Time distribution of S mean value μ_S and S sigma σ_S (for GCU0, CH0).

Following the same procedure described in the previous section, we divided, for each channel, the values of S and D within time intervals of 30 minutes and we extracted the centroids μ_S, μ_D and the sigmas σ_S, σ_D by means of a simple Gaussian fit on their distributions; we finally represented them as functions of time, obtaining several plots all similar to those reported as an example in Figure 3.7 and Figure 3.8. As we can notice, also in these cases the evaluated quantities happen to be stable during all the acquisition period.

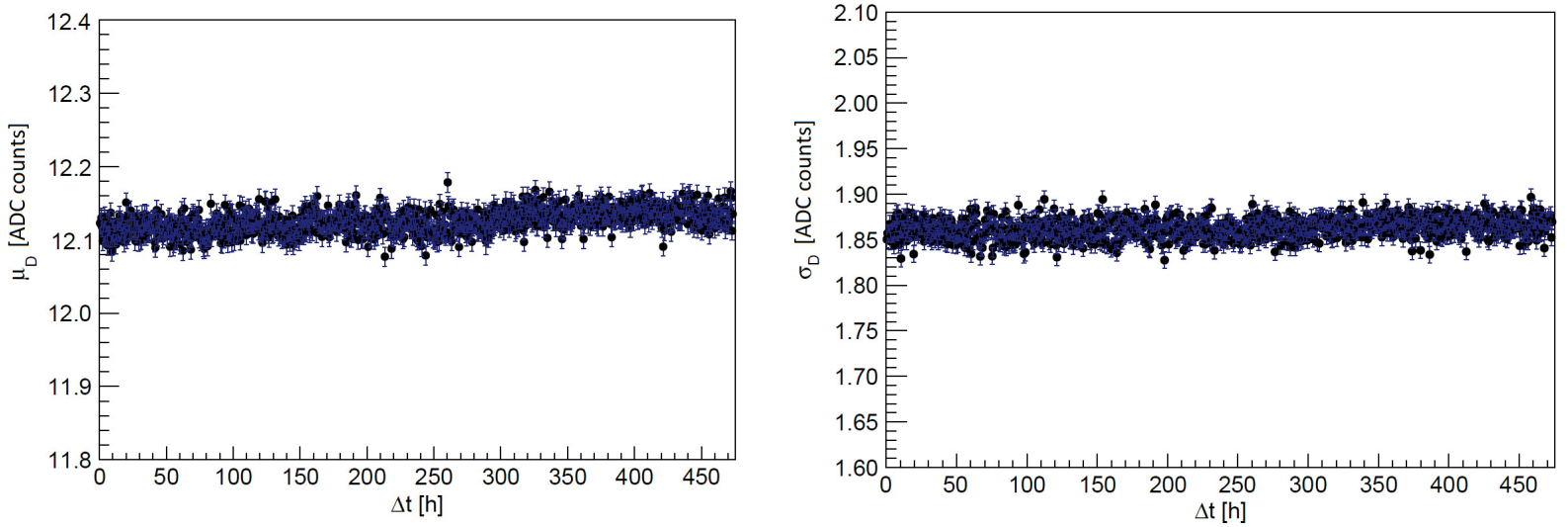

Figure 3.8: Time distribution of D mean value μ_D and D sigma σ_D (for GCU0, CH0).

3.3 Time correlation studies

Using the same data acquired for the analysis discussed in the previous section, we performed some studies involving the time correlation of signals acquired by different PMTs. The main goal of this part of the thesis work was to develop some methods in order to:

- check the signal and pulse integrity for all the PMTs of the setup;
- evaluate the timing correlation properties between different channels;
- reconstruct the total amount of charge released in the liquid scintillator.

3.3.1 Detecting and plotting time correlated waveforms

The first step consisted in the development of a ROOT based macro whose main task is to identify, given a specific timestamp, all the time correlated signals from the files produced during a single day of acquisition; in addition, a PDF file, containing the waveform plots of all the active PMTs, is produced in output (an example is reported in Figure 3.9). This is useful in terms of data quality monitoring since when possible anomalies in the data taking are spotted, it allows to check any affected raw data waveform. In fact, we noticed that in all the events that missed one or more of the 38 out of the 39 expected waveforms (we remind that GCU1, CH2 was not available during the whole acquisition in RUN1), the problem was due to bugs affecting the data packets produced by the DAQ system, which had to be solved.

In addition, after a careful study of the waveform plots for several timestamps in the same acquisition day, we noticed that the GCU3, CH0 produced always a very small signal, with respect to all the other channels and GCUs. This feature was due to the strong and undesired presence of signal ringing and reflection, as we can clearly notice in Figure 3.10. Due to this problem, the information provided by this PMT will not be considered in the following analyses.

Finally, we noticed that not all the channels had the same baseline value B and this could be a problem for those which are characterised by B $\sim$ 8000 ADC counts, which do not exploit the whole ADC dynamic range, as it can be noticed for example for GCU8, CH1 in Figure 3.11; in order to avoid

this problem, the baseline value will be set at B $\sim$ 12000 ADC counts in further acquisitions.

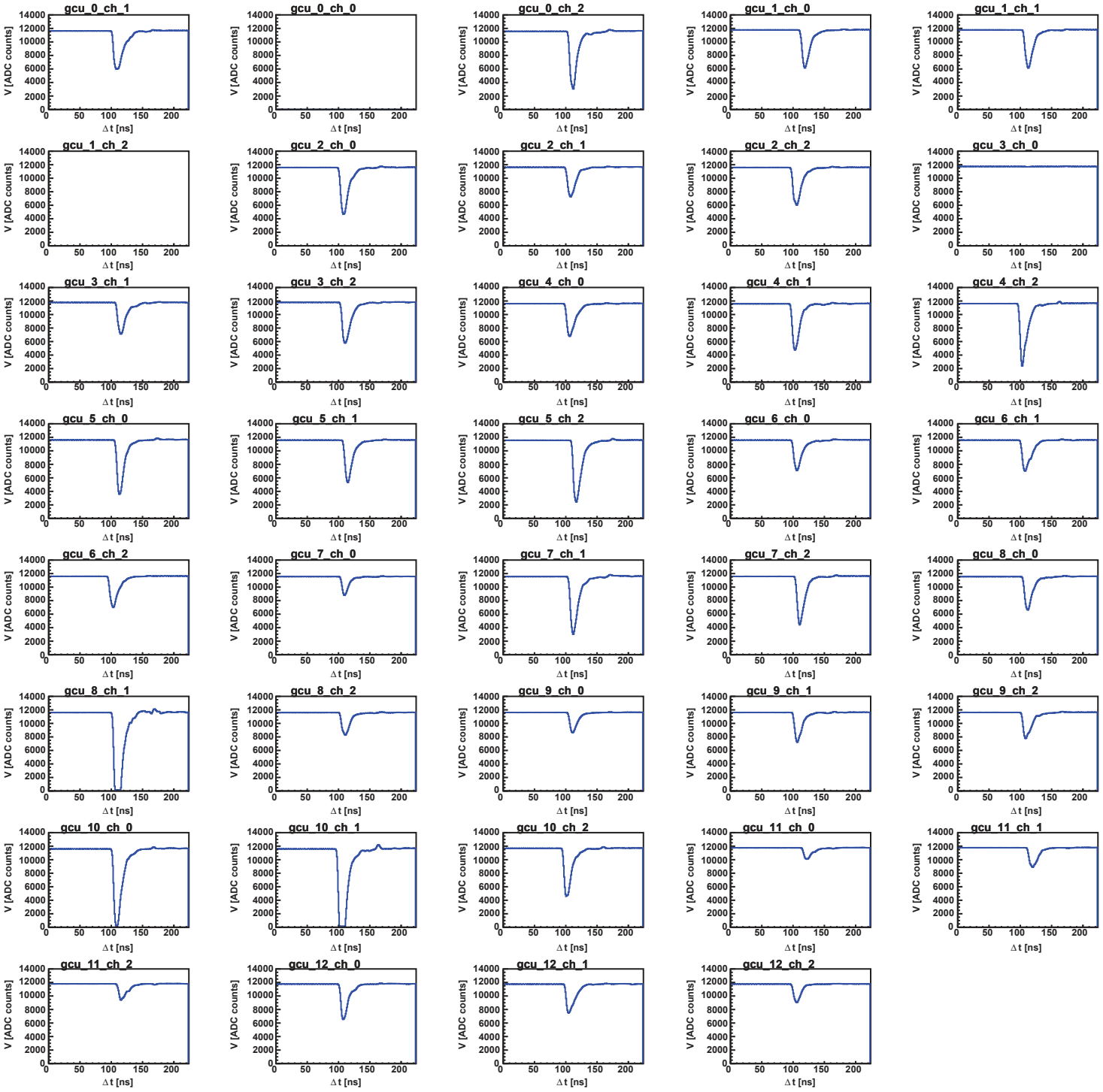

Figure 3.9: Example of time correlated waveform plots, after the setting of the same baseline value for all the channels. GCU0 CH0 and GCU1 CH2 are missing, because the corresponding PMTs did not acquire data.

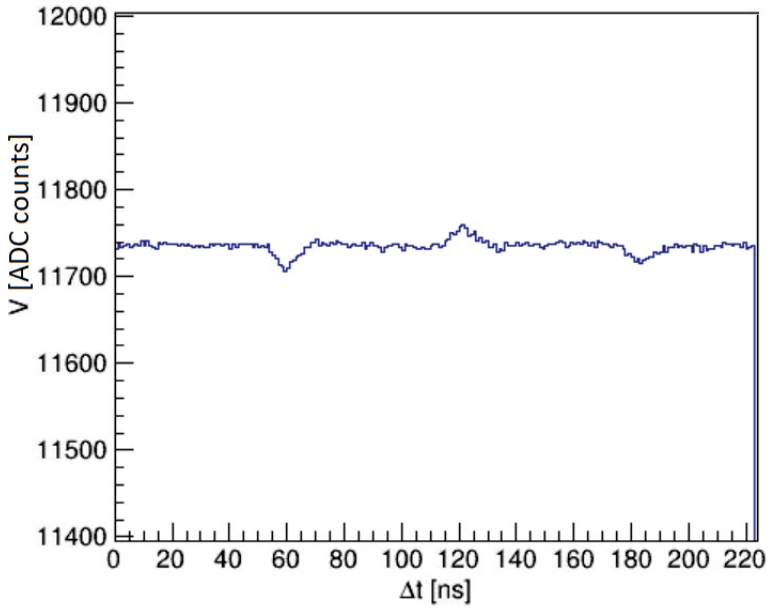

Figure 3.10: Zoom on the signal acquired by GCU3, CH0 with ringing effects.

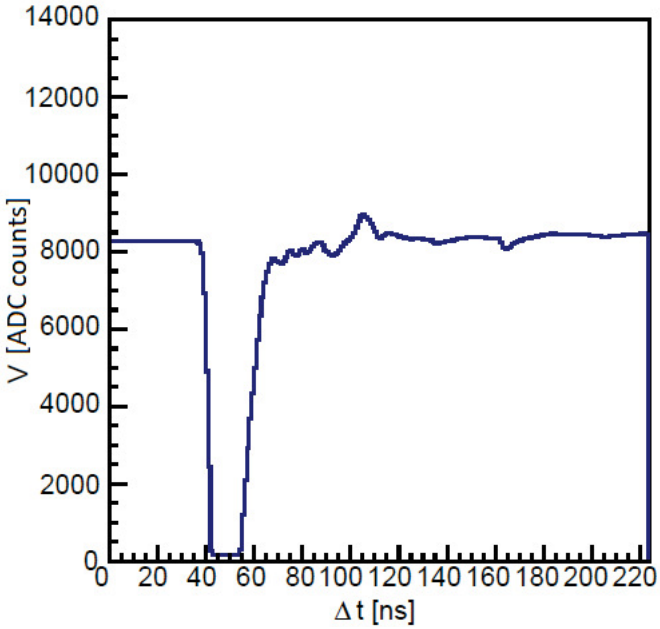

Figure 3.11: Saturation of the signal acquired by GCU8, CH1.

3.3.2 Study on the trigger jitters of time correlated signals

An important study (using data acquired in a single day of acquisition in RUN1), which will be useful in particular for defining the procedure of total charge reconstruction, discussed in detail in the next section, is the study of possible trigger jitters between signals related to the same event, i.e. characterised by equal timestamp. In fact, we expected the time correlated signals to be synchronised within 8 ns top, as a consequence of the specific operating conditions of our system [10].

In order to analyse the time synchronization of the signals, we needed to define a reference time t_H for all the waveforms at a certain timestamp, so that we could evaluate possible delays. We decided to define t_H as the time (in units of ns, and evaluated with respect to the trigger characteristic

timestamp) in which the signal reaches half of its maximum amplitude with respect to its baseline; t_H is calculated during the raw data processing phase and stored in the *timing info* branch of the ROOT `TTree` produced in output. In order to correctly implement this calculation procedure, we needed to be sure that a signal was present; therefore we searched for a pulse in a limited and fixed time interval of the trigger window, where we expected to find it, and we calculated t_H only if the signal exceeded a threshold of 6σ above its baseline.

Once the t_H for all the channels had been calculated, we evaluated the time differences $\Delta t_H = t_H^0 - t_H^i$, where t_H^i is the reference time of the i-th channel, while t_H^0 is the reference time of the first channel of GCU0, which is the GCU considered for activating the trigger system; the plot is presented in Figure 3.12. As we can notice, the majority of the Δt_H are within a time interval centered in zero and characterised by a width of $\sim$ 10-15 ns; all the cases which strongly deviate from this behaviour are characterised by very small signals, for which the adopted procedure for estimating t_H is not very suitable.

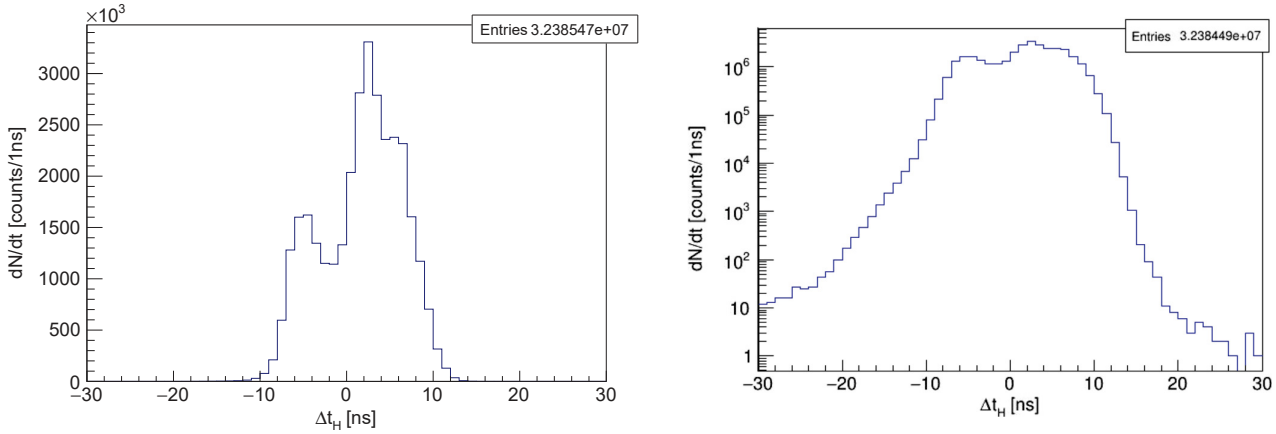

Figure 3.12: Δt_H distribution considering all the active channels of the DAQ system (on the right side the y-axis in log scale). Data acquired in a single day of acquisition in RUN1.

In order to obtain a better interpretation of the result and a more detailed discrimination of the behaviour of the different channels, we decided to restrict the same study considering first only the time differences of the GCU0 channels with respect to t_H^0, and then only the time differences of CH0 and CH1 of GCU5 always with respect to t_H^0; we considered the channels of GCU0 and GCU5 in order to have results on time synchronization of both signals acquired by the same GCU, which, in addition, is the one that activated the trigger system, and signals of a different GCU with respect to the reference one.

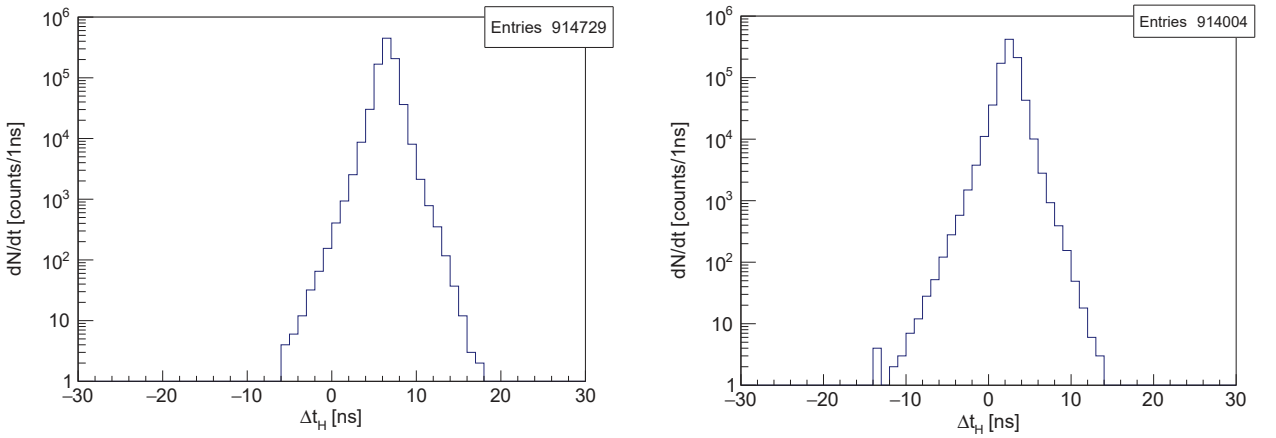

Figure 3.13: Δt_H distribution considering only the channels of GCU0 (GCU0 CH0 − GCU0 CH1 on the left side, GCU0 CH0 − GCU0 CH2 on the right side).

Looking to the results, reported in Figure 3.13 and Figure 3.14, it is interesting to notice that every channel has its own characteristic time difference with respect to the reference one; in particular, considering the cases of GCU5, the signals happen to be always after the signal of the reference channel. In addition, we can again notice that the majority of the signals deviate in time from t_H^0 by $\sim$ 10-15 ns top; this result is not in agreement with the time difference of 8 ns that was observed in previous analyses on time synchronization in the response of the DAQ system electronics [10]; the difference that we notice is probably due to the fact that we should consider also possible delays introduced during the PMT photon collection, which is not taken into account in the analyses of the response of the DAQ system electronics, in addition to a procedure which can be more suitable for signals which have very different amplitudes from each other.

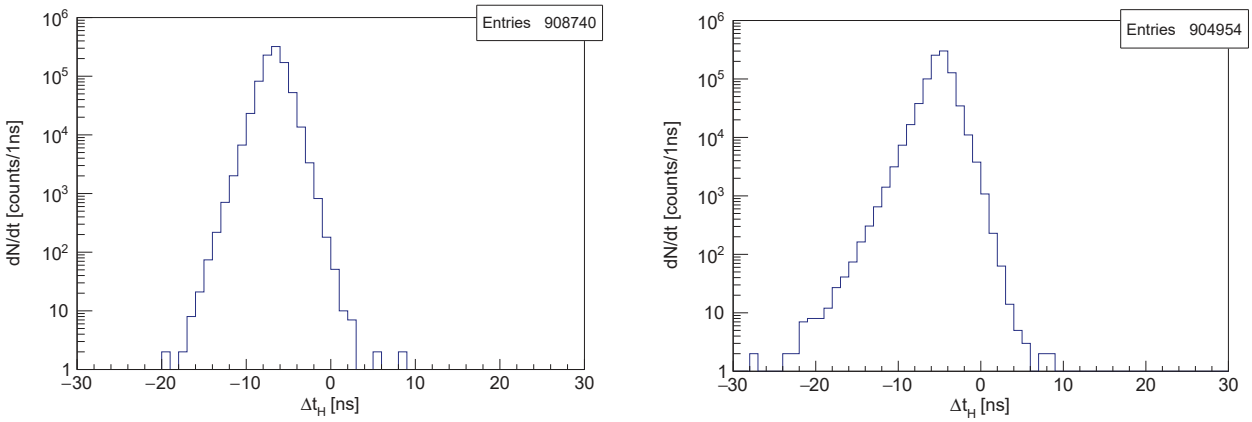

Figure 3.14: Δt_H distribution considering only CH0 (GCU0 CH0 − GCU5 CH0, on the left side) and CH1 (GCU0 CH0 − GCU5 CH1, on the right side) of GCU5.

3.3.3 Reconstruction of the total charge released in the liquid scintillator

The results of the time synchronization study, discussed in the previous section, are very useful to define a proper procedure to extract from the acquired signals the information on the charge released by particles in the liquid scintillator; this information is very important because it is related to the energy of the particles detected by the experimental setup, as it will be discussed in detail in the following chapters. The procedure we adopt for reconstructing the total integrated charge is the following:

1. during the raw data processing phase, we calculate the integral of the signal waveform within a fixed time interval Δt_S, where we expect to find the signal, corresponding to N_B bins, as it is shown in Figure 3.15; Δt_S is properly chosen taking into account possible jitters between time correlated signals, whose entity has been studied in detail in the previous section, and it can be varied accordingly to the DAQ parameters. In order to obtain an estimate of the integrated charge for the single waveform (expressed in μC), we use the following formula:

$$Q = \Delta t_S \cdot \frac{\sum_{i=1}^{N_B} |N_i - B| \cdot 75\mu V}{R}$$

   where N_i is the content (in *ADC counts*) of the i-th bin, B is the baseline mean value (in *ADC counts*) of the considered waveform, R is assumed to be equal to 50Ω, i.e. the expected impedance seen by the PMTs, and, finally, 75 μV is the voltage corresponding to 1 ADC count; obviously, the charge value is extracted only if a signal is actually present, i.e. if it exceeds a threshold of 6σ above its baseline;
2. the estimated charge is then stored in its specific branch of the ROOT `TTree` produced as output of the processing phase;
3. for each acquired event, using a procedure similar to the one previously described in Section 3.3, we identify all the waveforms characterised by its corresponding timestamp and we sum over

all the active channels, checking also that none of the signals is missing as a consequence of consistency problems of the raw data packets;

4. finally, considering only the events for which the signals of all the active channels have been found, the estimate of the total integrated charge Q_T is plotted, producing a distribution as the one reported in Figure 3.16 as an example.

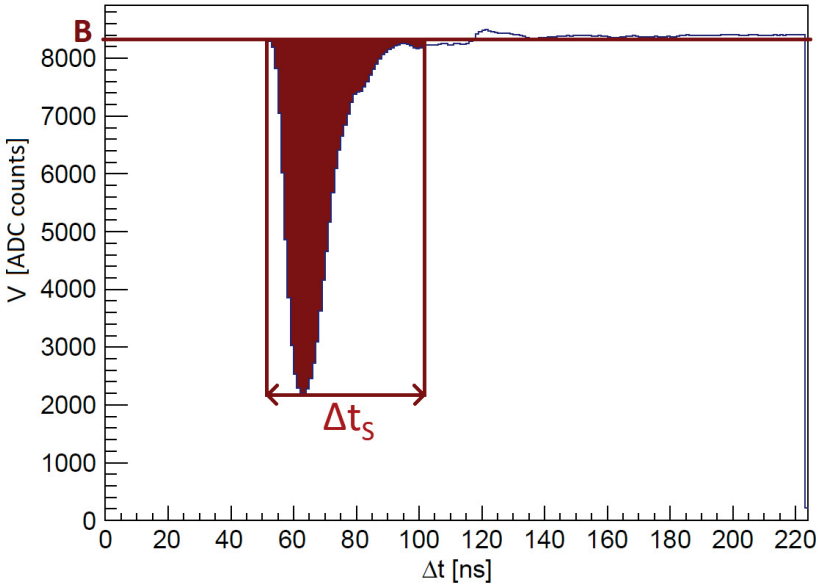

Figure 3.15: Description of the procedure for the integrated charge calculation.

We initially considered a slightly different procedure for charge calculation, which integrated the voltage signal within an interval of fixed width but whose starting time was not always the same, depending on the position of the signal minimum; comparing the resulting integrated charge of this alternative method (Q_{notfixed}) with the one obtained with the adopted procedure (Q_{fixed}), we noticed the two algorithms were not exactly equivalent, since Q_{notfixed} often happened to be greater than Q_{fixed} (see Figure 3.17).

This result suggests that the alternative procedure is not always reliable, because it could provide an integrated charge value considerably different from zero even if there is not any pulse within the acquired trigger window. This is due to the fact that the alternative procedure simply search for a minimum without considering if the signal is actually present in the time range where we expect to find it.

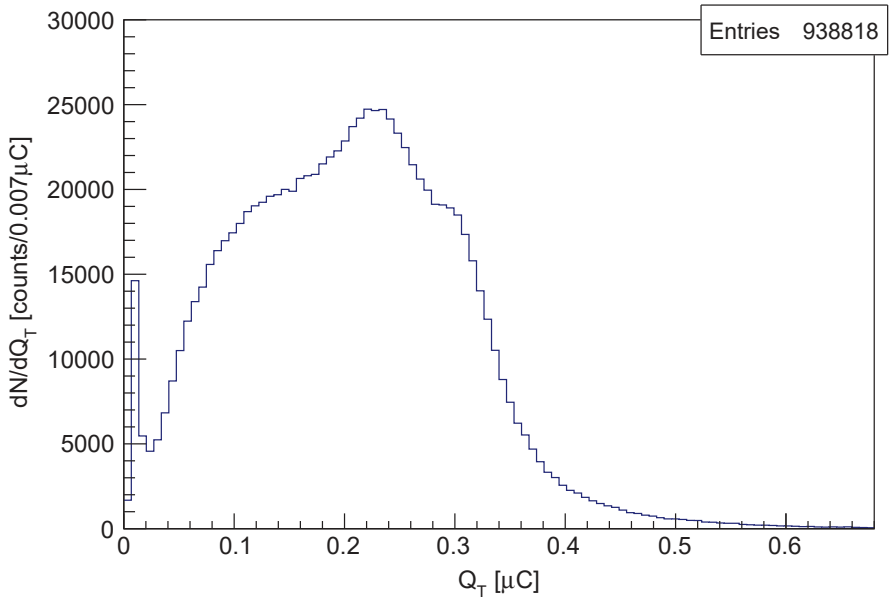

Figure 3.16: Example of total integrated charge distribution (data acquired in 21.03.2020, during RUN1).

Figure 3.16 contains the total integrated charge collected during a single day of acquisition in RUN1, and it is in agreement with the DAQ condition of high threshold set for this run. In order to go deeply in the study of our experimental setup response, we need to acquire new data using a different trigger mechanism, which should select events corresponding to well-known physics phenomena; this allows us to compare the experimental results with some Monte Carlo simulations, as it will be discussed in detail in the following chapters.

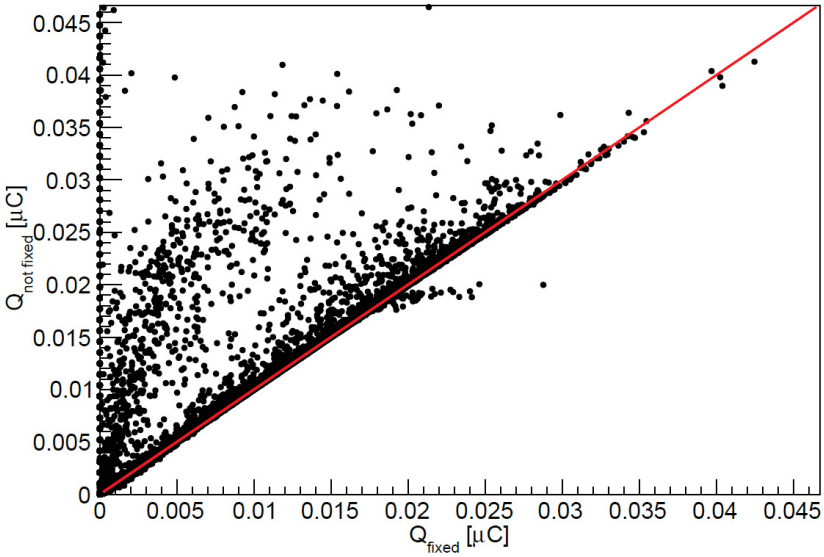

Figure 3.17: Comparison of the results of the two possible algorithms for total charge calculation (fixed/not-fixed start of the integration range).

3.4 Characterization of the PMT electronic noise and single-photon measurements

An interesting study that we performed, applying the methods discussed in the previous sections, concerns the characterization of the electronic noise affecting the PMT response and the detection of single-photon events.

These measurements have been performed using the available LED light, at 405 nm, connected to the top of the LS vessel by means of an optic fiber (a photo of the LED circuit and the optic fiber is reported in Figure 3.18).

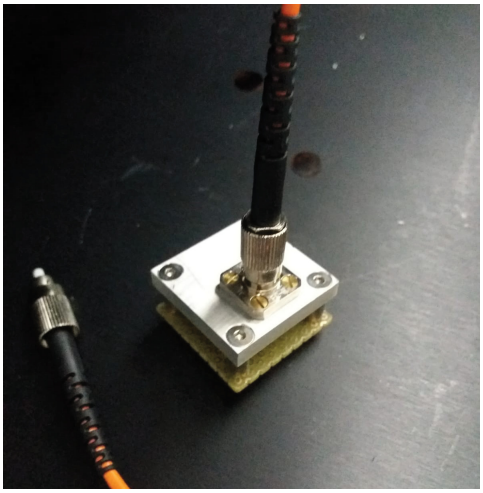

Figure 3.18: Photo of the employed LED circuit and the optic fiber used to connect it to the top of the setup.

Connecting the LED to a pulse generator, we produced square pulses characterised by an amplitude A = 6 V and a width W = 10 ns, which have been opportunely set in order to be able to detect single-photon signals with our setup, employing an external trigger provided by the pulse generator, with a resulting rate of 500 Hz. It is important to underline that, due to the position of the LED light, not all the PMTs were illuminated during these measurements.

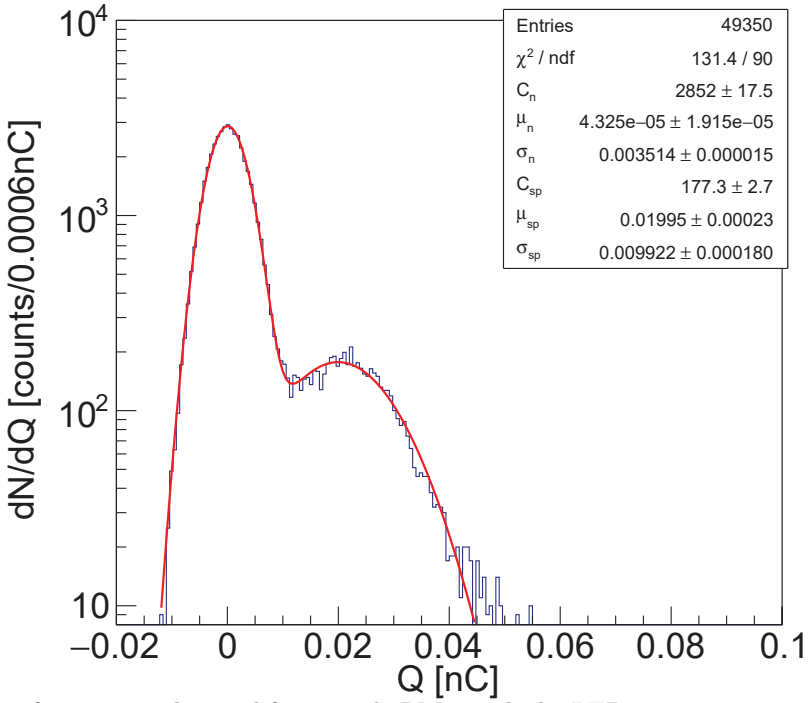

Figure 3.19: Example of spectrum obtained for a single PMT with the LED measurements; the double Gaussian function used for the fit is reported in red. The parameters C_n, C_{sp} are reported in units of counts/$6 \cdot 10^{-4}$ nC, while μ_n, σ_n, μ_{sp} and σ_{sp} in units of nC.

A typical example of detected spectrum for a single PMT is reported in Figure 3.19, where we can notice the presence of two different peaks. The first peak, centered in zero, has been obtained even in measurements performed without the LED light and in the same trigger validation conditions; an example of resulting spectrum for these measurements is reported in Figure 3.20. This result suggests that this peak is due to the electronic noise affecting the PMT, while the other one is the actual single-photon peak.

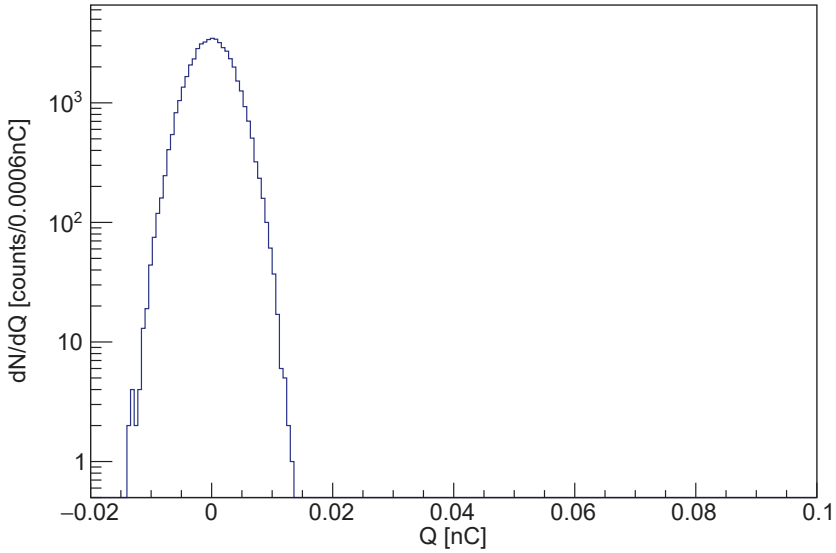

Figure 3.20: Example of spectrum obtained without the LED light, using the pulse generator as external trigger.

In order to extract some interesting information from the spectrum reported in Figure 3.19, we decided to perform a fit with the following function G_2, that is basically a double Gaussian:

$$G_2 = C_n \cdot e^{-\frac{(x-\mu_n)^2}{2\sigma_n^2}} + C_{sp} \cdot e^{-\frac{(x-\mu_{sp})^2}{2\sigma_{sp}^2}}$$

Using the estimated sigma of the noise peak, $\sigma_n = (351 \pm 2) \cdot 10^{-5}$ nC, we can evaluate its FWHM $= (827 \pm 5) \cdot 10^{-5}$ nC, which can be considered as a good estimate of the electronic noise affecting the PMT. In addition, using also the estimates of the single-photon peak sigma, $\sigma_{sp} = (0.0010 \pm 0.0002)$ nC, and of the centroids of the two peaks, $\mu_n = (4 \pm 2) \cdot 10^{-5}$ nC and $\mu_{sp} = (0.0200 \pm 0.0002)$ nC, we can obtain an estimate of the resolution for the single-photon peak, calculated as $R = \dfrac{\sigma_{sp}}{\mu_{sp} - \mu_n}$ $= (0.50 \pm 0.01)$. Finally, we can also calculate the peak-to-valley ratio R_{pv}, calculated as the ratio between the maximum of the single-photon peak and the local minimum of the double Gaussian, positioned between the two peaks. The estimated $R_{pv} = 1.29$ tells how well we can discriminate the single-photon signal from the noise, and it is an important parameter for the characterization of the setup. The spectra and the results of the fitting procedure for all the active PMTs are reported in Appendix B.

3.5 Efficiency tests of the setup response

During our studies we noticed that in acquisitions characterised by higher trigger rates the system sometimes failed in providing a specific timestamp from all the active channels. Therefore, we decided to perform some specific tests on the experimental setup in order to characterize its efficiency ϵ in storing the whole event information, provided by all the active channels, with respect to different acquisition rates.

Employing the Python-based DAQ software, we performed some tests to evaluate the behaviour of ϵ firstly for a single channel, estimated as the the ratio between the number of stored waveforms and the total number of trigger validations, as function of the trigger rate; then, we performed a similar study for the efficiency of the single GCU and the whole setup, where ϵ has been calculated in these cases as the ratio between the number of retrieved channels N_f for a fixed timestamp and the total number of active channels N_t ($N_t = 3$ for the single GCU, $N_t = 37$ for the whole setup). The acquisition rate was controlled by an external pulse generator which provided square waves at the desired frequency as input to the external trigger system. The acquired waveform window was kept at 144 ns for all the channels throughout the whole test.

The results of these studies are reported in Figure 3.21, Figure 3.22 and Figure 3.23. Considering the efficiency of the setup for the single channel, reported in Figure 3.21, we can notice that ϵ starts to decrease at $\sim$ 5 kHz with a low slope; this decrease becomes more drastic when we move to the studies regarding the single GCU and, even more, the whole setup, where we found a 100% efficiency up to respectively $\sim$ 6 kHz and $\sim$ 2.4 kHz.

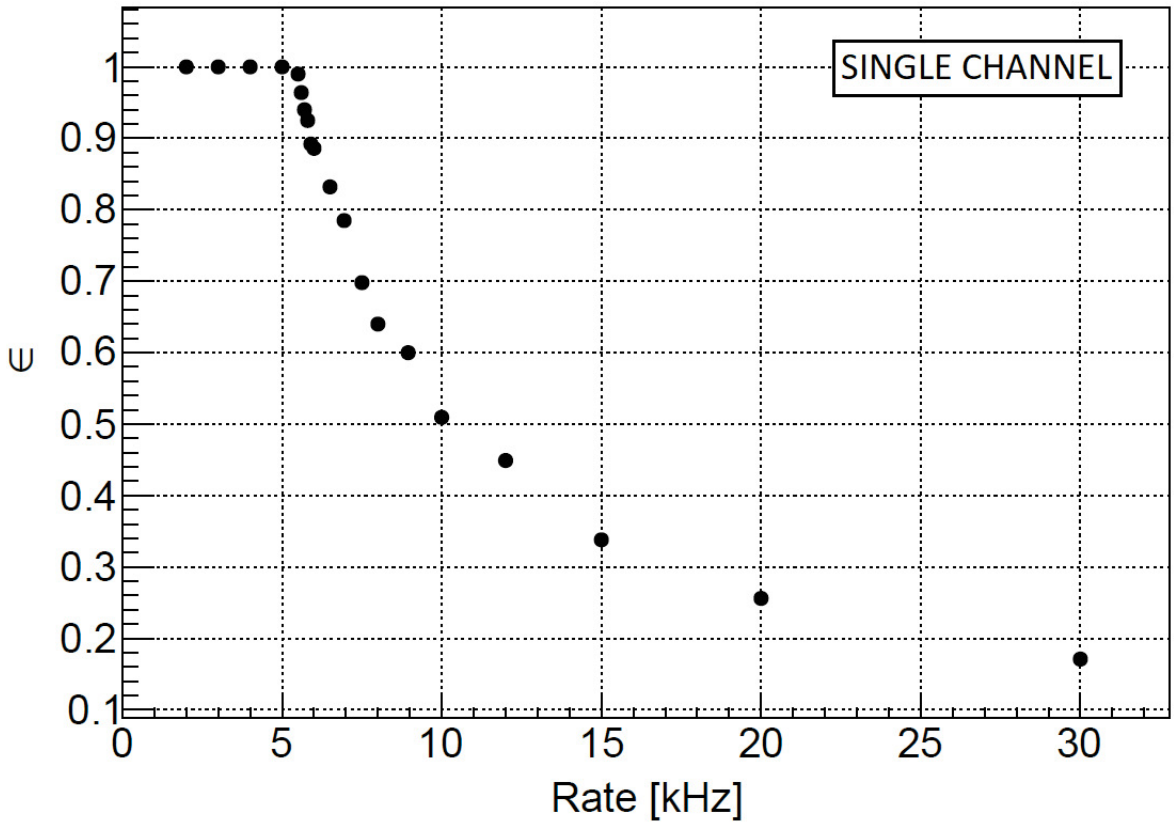

Figure 3.21: Efficiency ϵ trend as function of the acquisition rate for the single channel, using the Python-based acquisition software.

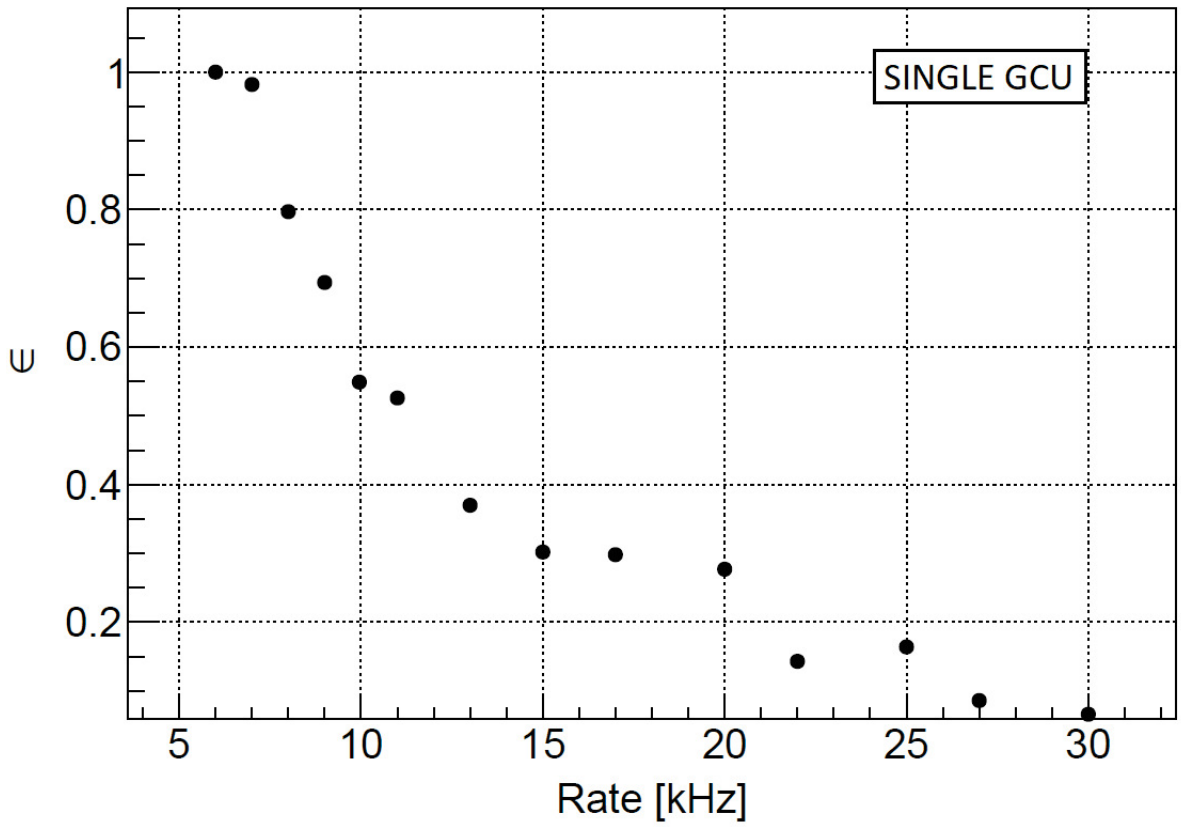

Figure 3.22: Efficiency ϵ trend as function of the acquisition rate for the single GCU, using the Python-based acquisition software.

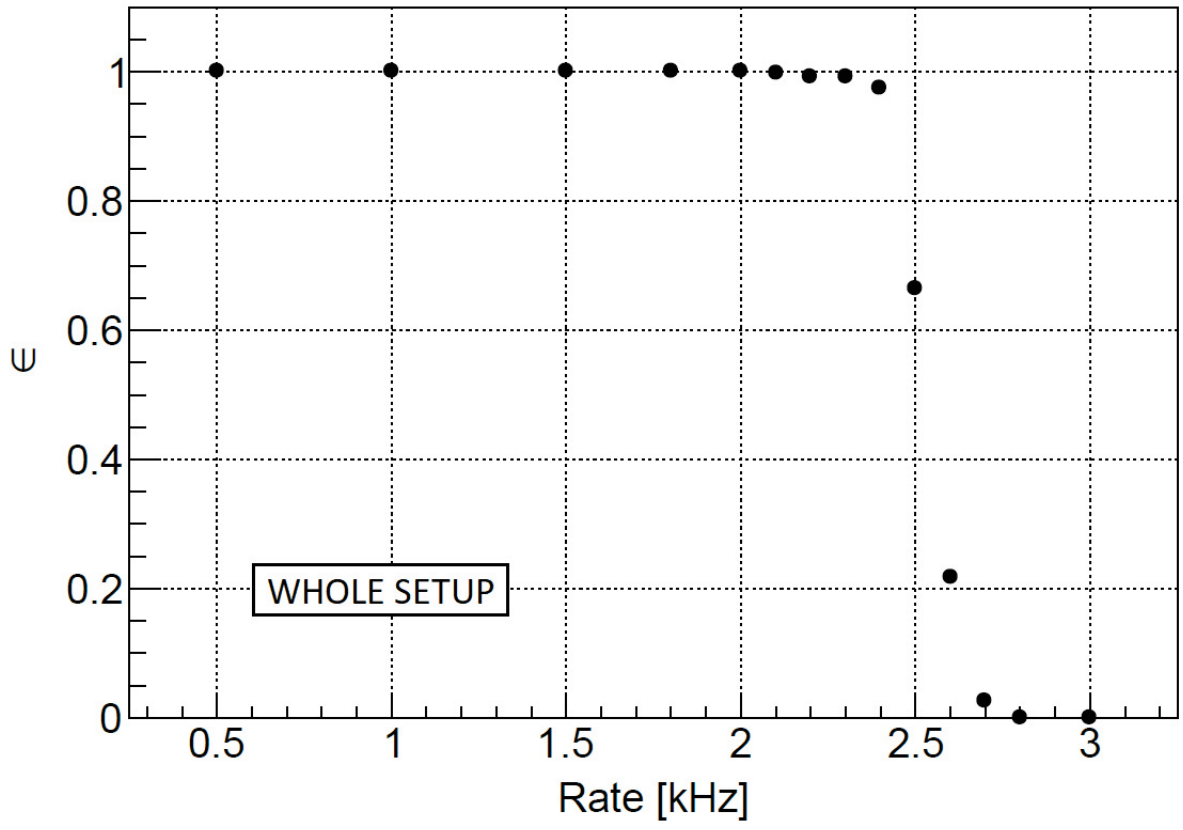

Figure 3.23: Efficiency ϵ trend as function of the acquisition rate for the whole setup, using the Python-based acquisition software.

We finally performed a similar test, just for the whole setup (37 active channels again), using the JUNO DAQ software and setting a fixed waveform window of 304 ns for all the channels; the result is reported in Figure 3.24. Making a comparison between Figure 3.23 and Figure 3.24, we can notice that using the JUNO DAQ software the efficiency starts to decrease at higher rates, even if the acquired waveform windows was set to a larger value than the one set in the acquisition with the Python-based software. In particular Figure 3.24 shows a first, slight decrease of ϵ at ~ 5 kHz and a subsequent drastic decrease at ~ 15 kHz.

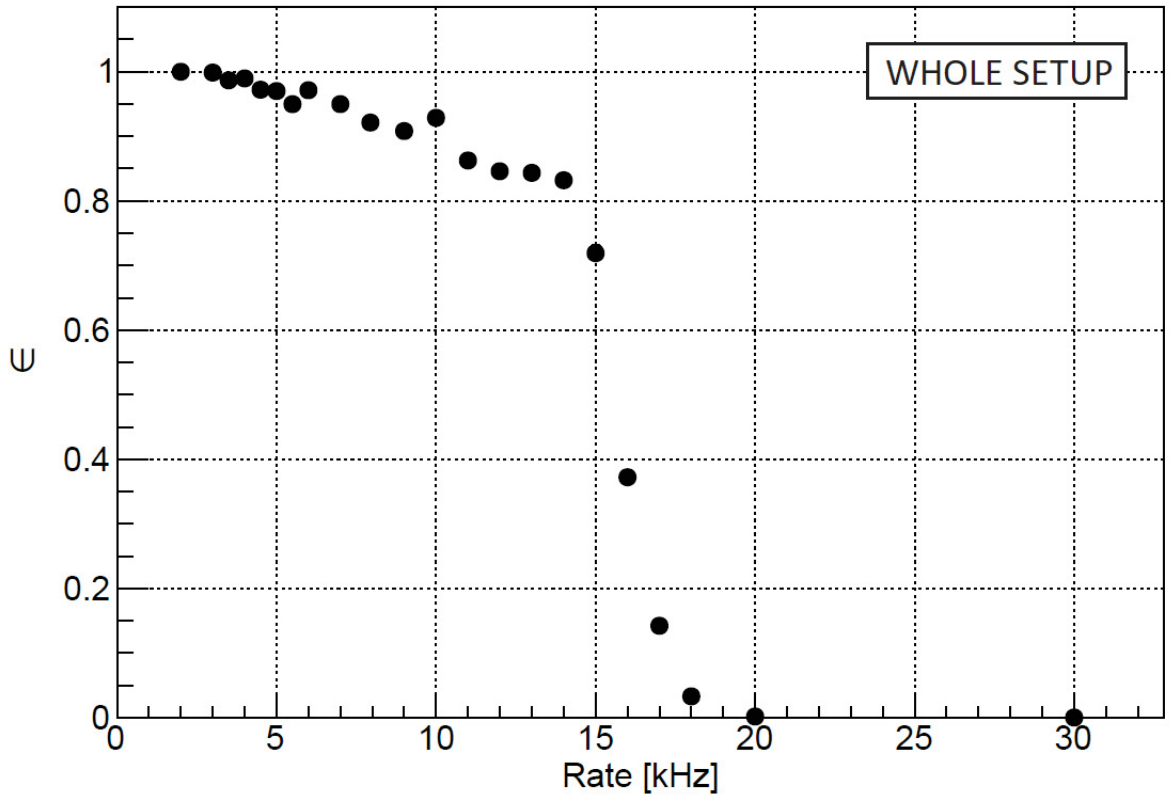

Figure 3.24: Efficiency ϵ trend as function of the acquisition rate for the whole setup, using the JUNO acquisition software.

The results that we have obtained shows that the system is able to acquire data without losing information even at rates of the order of kHz; this is important, for example, in acquisitions with radioactive sources, which are used to provide an energy calibration of the experimental setup, as it is discussed in Chapter 5, and which are characterized by typical rates of the order of kHz.

Another important information provided by these studies is the difference between the efficiency obtained with the JUNO DAQ software, developed in C++, and the one provided by the Python-based script. This difference suggests that a possible bottleneck consists in the different use of the CPU in the acquisition server; therefore, a software based on C++ drastically reduces the use of the CPU, improving the performances.

However, determining the bottlenecks and optimizing the system settings, such as, for example, the dimension of the acquired waveform window, it is possible to considerably increase the limits on the acquisition rates.

The plot reported in Figure 3.25 [28] shows the expected performances, for a single channel, of the IPbus protocol for the data transmission through Ethernet, developed at CERN and employed also by our system. In the acquisitions that we performed the dimension of the IPbus data packet was $\simeq 10$ kB, corresponding to a throughput of $\simeq 200$ Mbps, for the single channel.

As we can notice, the expected maximum throughput is around 500 Mbps, therefore we can evaluate the maximum sustainable trigger rate ν_{max} expected for a single channel, acquiring data in the same conditions we set for the efficiency tests with the JUNO DAQ software. With a data packet size of 320x16 bits, which corresponds to 5120 bits per single data packet, we estimate $\nu_{max} \simeq 97.7$

kHz; this limit clearly suggests that there is vast room for improvement for our system, which can be achieved properly optimizing the acquisition settings.

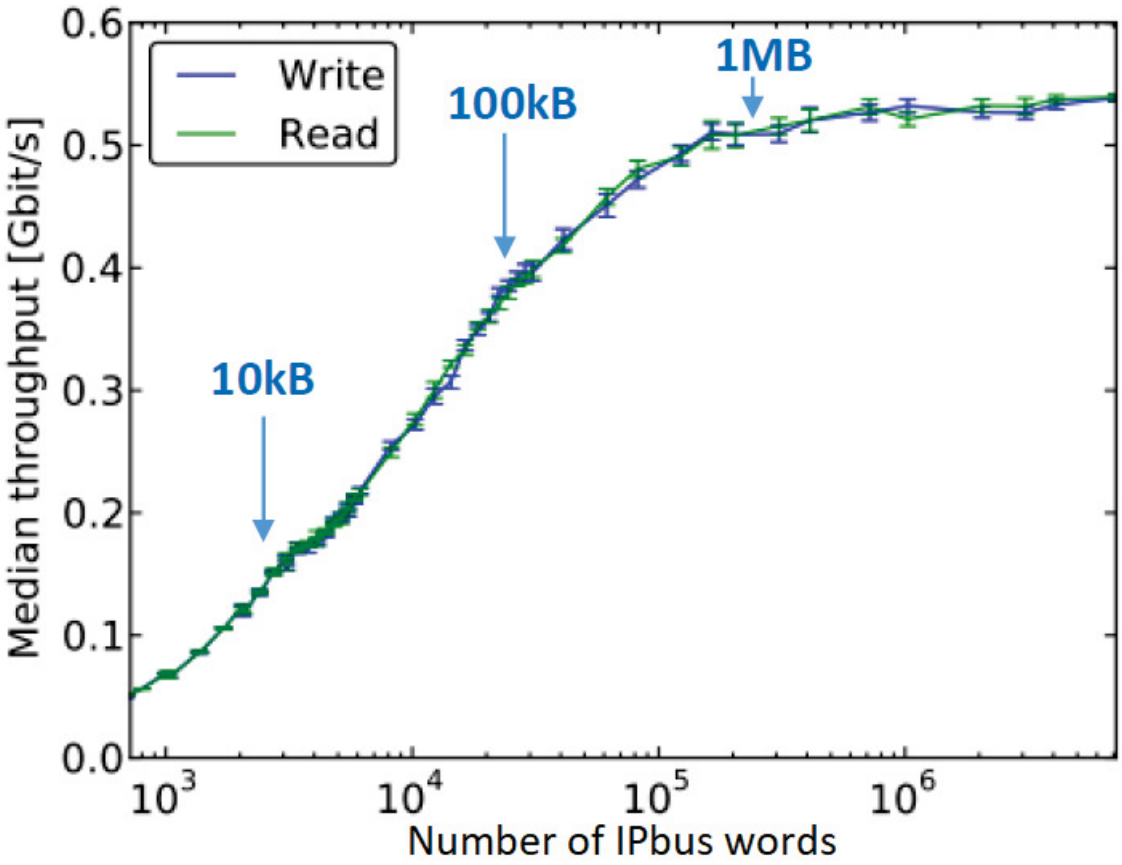

Figure 3.25: Trend of the median throughput as function of the number of IPbus words expected for the IPbus protocol employed by the acquisition system (single-channel measurements) [28].

Chapter 4

Data analysis on cosmic muons

A more detailed characterization of the experimental setup has been pursued considering data which could be compared to robust theoretical expectations, for example based on the results of Monte Carlo simulations of the setup response to well-known physics phenomena. Therefore, we decided to perform a 10-days long acquisition (referred to as RUN2 in Table 3.1, starting from 08.05.2020, stopping in 18.05.2020) setting an external trigger on cosmic muons, provided by the use of the three external plastic scintillators; in particular, the trigger is activated only if a particle passes through all the three scintillating slabs. During this run, in addition to GCU1 CH2, also GCU0 CH0 did not acquire signals. This chapter will present a detailed analysis of the acquired data, considering, in addition to some preliminary checks on the setup response in this different trigger configuration, a deep study concerning the reconstruction and interpretation of the energy spectrum of the detected events.

4.1 Data Quality Monitoring Studies

First of all, using the developed methods presented and discussed in Chapter 3, we performed some data quality monitoring over the whole 10-days long acquisition of RUN2. Some examples of stability studies are reported in Figure 4.1, Figure 4.2 and Figure 4.3, concerning the trigger rate, the mean value and the spread of the baseline contribution to the signals; even in this case we do not report the plots for all the channels because they all show a similar trend.

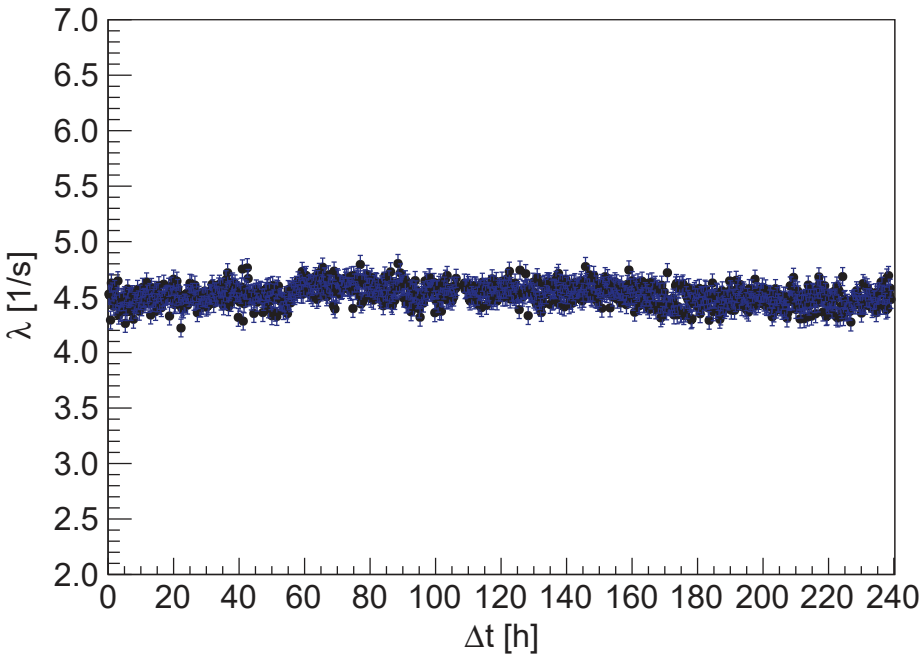

Figure 4.1: Trigger rate stability study during the whole acquisition of RUN2.

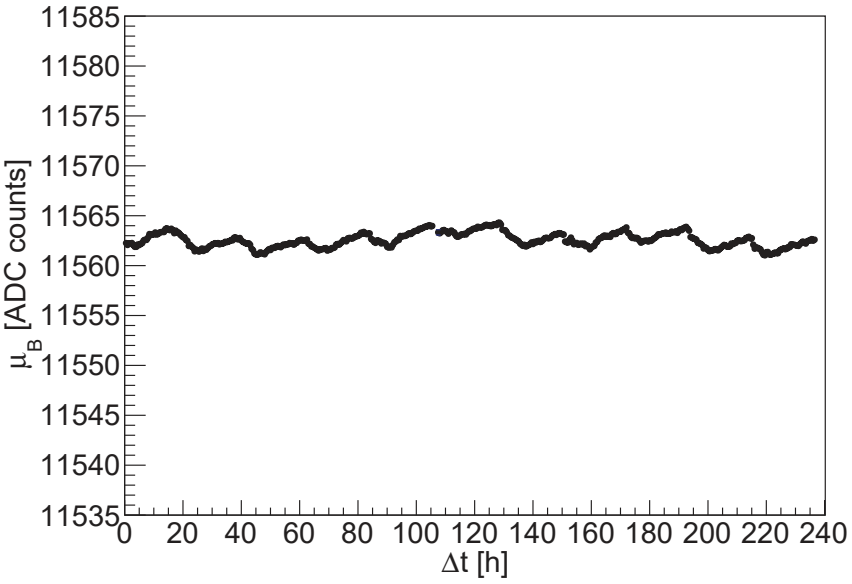

Figure 4.2: Example of stability study for the baseline mean value μ_{B} during RUN2 (GCU0, CH2).

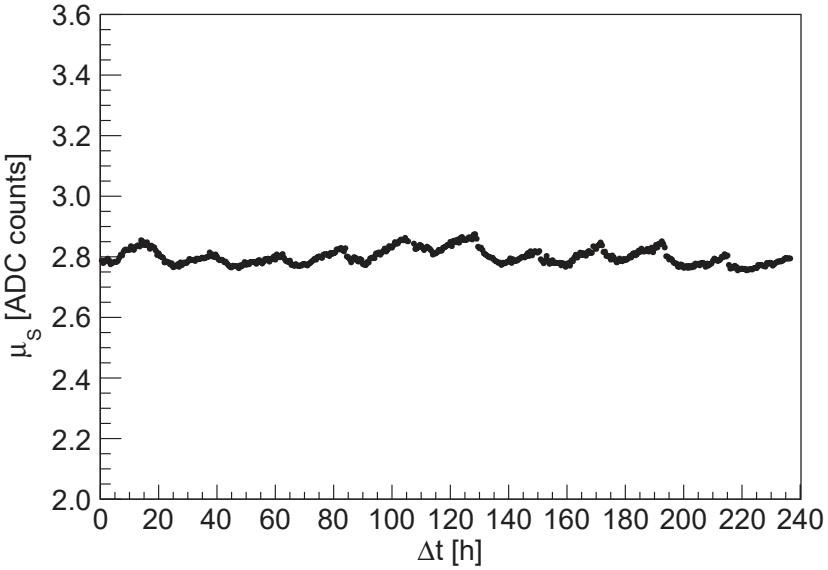

Figure 4.3: Example of stability study for the baseline standard deviation μ_{S} during RUN2 (GCU0, CH2).

In Figure 4.1 we can notice a very good stability of the trigger rate, while looking to Figure 4.2 and Figure 4.3, we can notice a slightly variation of the baseline characteristics, which is however negligible considering that 1 ADC count is equal to a voltage difference of 75 μV. The measured trigger rate λ_{exp} is slightly smaller than the expected one, $\lambda_{th} \simeq 7.8$ Hz, estimated considering the integral intensity of vertical muons at sea level $I \sim \dfrac{1}{\mathrm{cm}^2 \cdot min}$ [29] and the horizontal section of the liquid scintillator vessel $A \simeq 0.047\mathrm{m}^2$; nonetheless, they are of the same order of magnitude.

4.1.1 Characterization of the signal rise time and fall time

We decided to perform an additional study for characterizing the acquired waveforms, which consisted in evaluating their rise time t_R and fall time t_F distributions; the adopted algorithm for t_R and t_F calculation is composed of the following steps:

- in a fixed time interval, where we expected to find the pulse, the algorithm checks for its presence with the usual threshold level of 6σ above the baseline;

- then the signal maximum amplitude is evaluated, searching for the bin with the minimum content within the selected time window;
- using the information on the amplitude, we retrieve the time $t_{10\%}$ and $t_{90\%}$ in which the signal respectively reaches 10% and 90% of the total amplitude, both on the rising and on the falling edge of the signal;
- finally, we calculate the rise time and fall time of the signal as $t_R = t_{R,90\%} - t_{R,10\%}$ and $t_F = t_{F,10\%} - t_{F,90\%}$.

The resulting distributions are reported in Figure 4.4 and Figure 4.5; we can notice that both the plots show a prominent peak around $t_R \sim 5$ ns and $t_F \sim 18$ ns, respectively for the rise time and for the fall time, with FWHM ~ 3 ns for the rise time distribution and FWHM ~ 7 ns for the fall time. It is interesting to report that we found a limited amount of signals which happened to have a very large fall time (over 40 ns); we tried to plot the corresponding waveform (an example is reported in Figure 4.5), discovering that this anomaly was due to the presence of a double pulse within the usual, fixed time window considered for t_R and t_F calculation and where we expected to find just one signal; this clearly compromised the adopted procedure. However, the number of similar anomalous cases is negligible with respect to the total amount of considered signals.

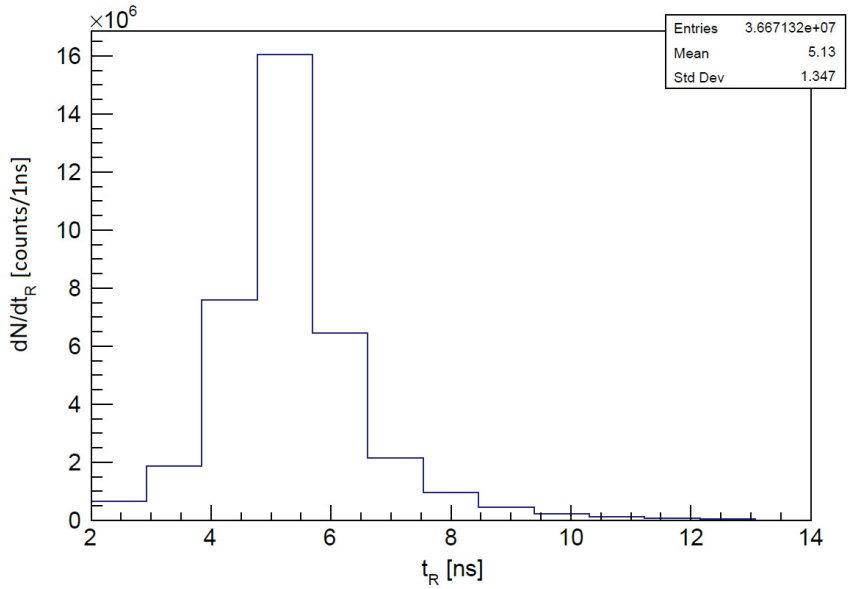

Figure 4.4: Signal rise time distribution.

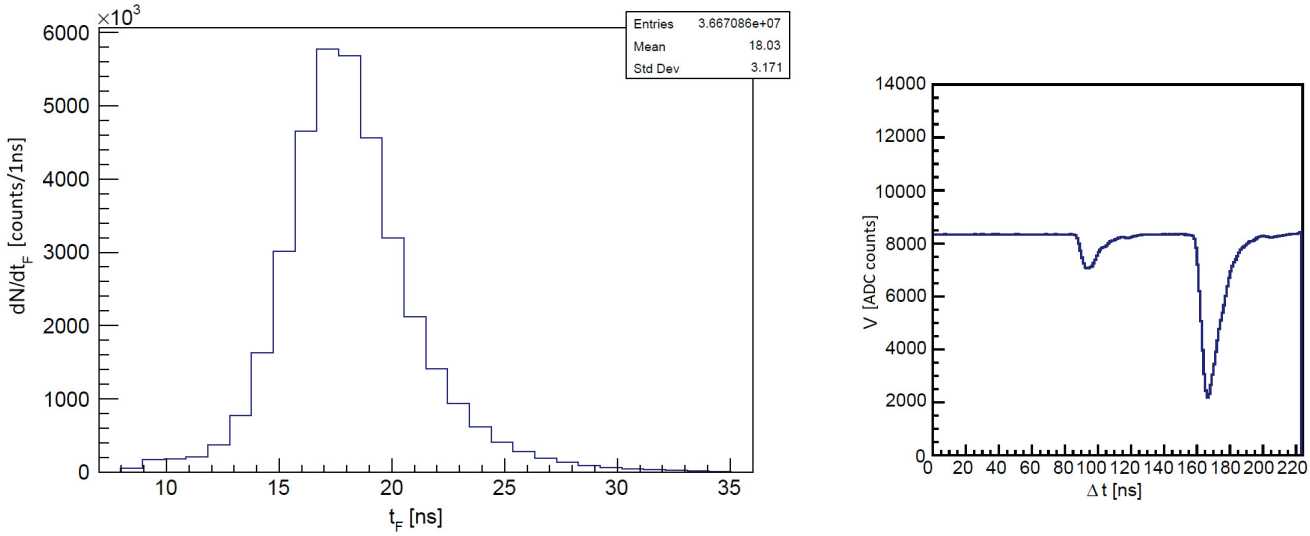

Figure 4.5: Left panel: signal fall time distribution. Right panel: example of signal with double pulse.

4.2 Total Charge Reconstruction

Another fundamental step in the analysis of the data acquired in RUN2 consisted in the reconstruction of the total charge Q_T released by the muons in the liquid scintillator and, consequently, in the interpretation of the corresponding energy spectrum. Therefore, following the same procedure previously developed, and discussed in detail in Section 3.3.3, we obtained the distribution reported in Figure 4.6, which contains the total charge produced as a consequence of the muon energy loss in the liquid scintillator.

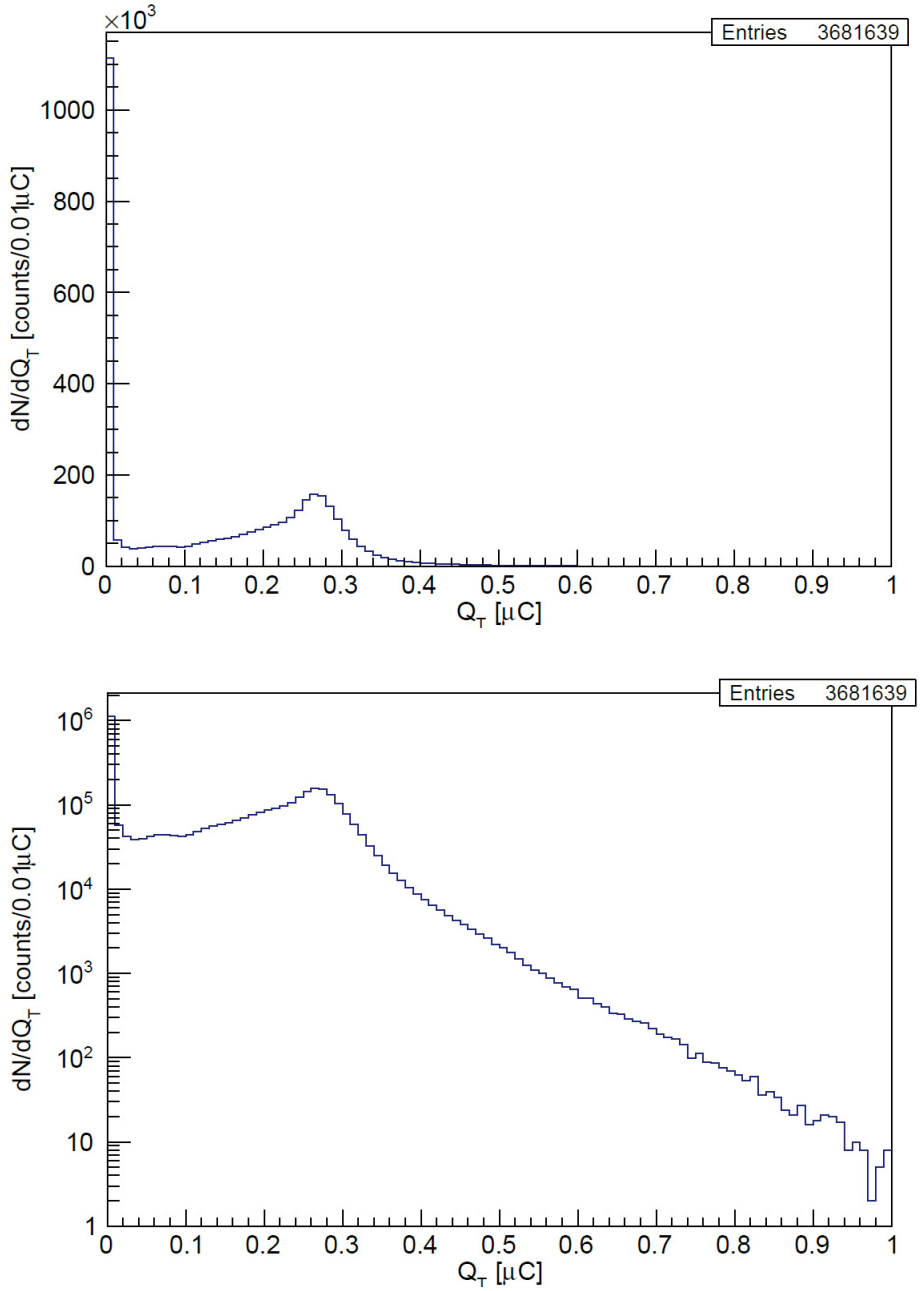

Figure 4.6: Total charge Q_T distribution for data acquired in RUN2 (in the bottom panel the y-axis in log scale).

Our result reported in Figure 4.6 shows the typical trend of an energy loss distribution for a minimum ionizing particle (MIP), which is in good agreement with our expectation on muons detected by the experimental setup. Nonetheless, we needed to perform some additional studies, based also on the comparison with the results of Monte Carlo simulations on the response of the apparatus to cosmic

muons, in order to obtain a better interpretation of the experimental result.

4.2.1 Study on the event multiplicity

In the first study we focused on the analysis of the total charge distribution taking into account the multiplicity ω of the acquired events, i.e. considering for each event how many active channels detected a signal exceeding a threshold of 6σ above the baseline.

This study allowed to separate the events which mainly contributed to the low-charge region of the spectrum in Figure 4.6 from those responsible for the characteristic peak at $Q_T \sim 0.3$ μC; in fact we expected that events associated to a low integrated charge were characterised by a low multiplicity, since the latter condition is more likely related to particles which traverse such a short distance in the liquid scintillator that only a very limited number of PMTs is able to detect it, while we expected that particles with a higher energy loss had triggered a greater number of channels.

We report in Figure 4.7 and Figure 4.8 only few examples of charge distributions respectively characterised by $\omega \leq 9$ and $\omega \geq 28$, which are those in which we are more interested for our analysis, since they are related respectively to particles that travelled for a very short distance in the detector and, on the other hand, to muons that released the largest amount of energy, crossing larger distances in the liquid scintillator.

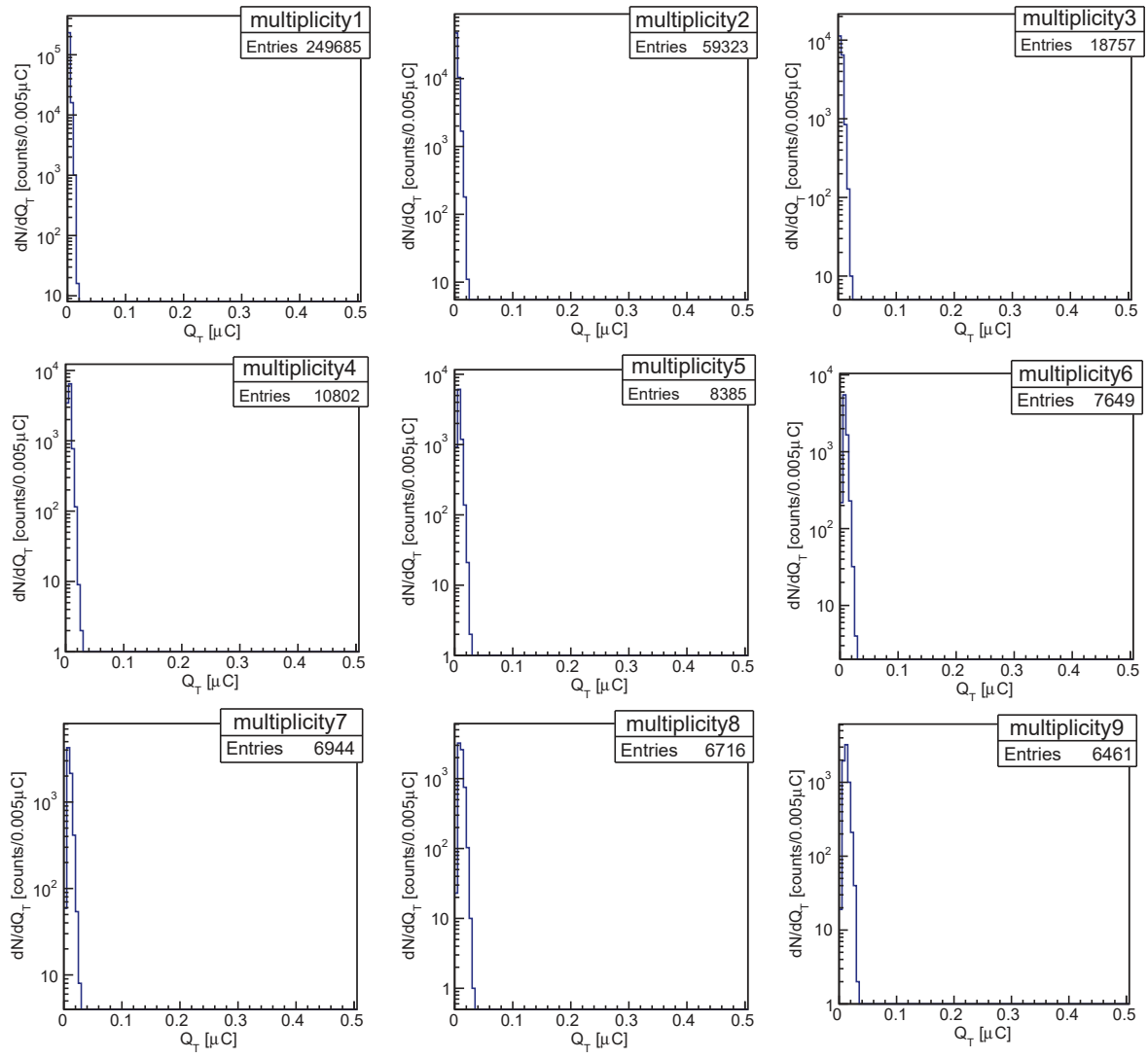

Figure 4.7: Total charge Q_T distribution for data acquired in RUN2 with $\omega \leq 9$.

Considering Figure 4.7 and Figure 4.8, we can notice that our expectation seems to be confirmed, since the events with low ω essentially contribute only to the lowest region of the spectrum, while the peak at $Q_T \sim 0.3$ μC is mainly due to events with high ω.

Focusing in particular on Figure 4.8, we can notice that the number of entries for each distribution increases with the considered multiplicity up to $\omega = 32$, then it starts to decrease; considering that GCU0 CH0 and GCU1 CH2 did not acquired data during RUN2 and that GCU3 CH0 did not provide any valid signal due to ringing effect (as discussed in Section 3.3), we would have expected an increase of the number of entries up to $\omega = 36$. Since the result we obtained seemed to be in contrast with our expectation, we decided to perform some additional analyses regarding the response, in terms of charge, of each, single active PMT, as it will be discussed in the following section.

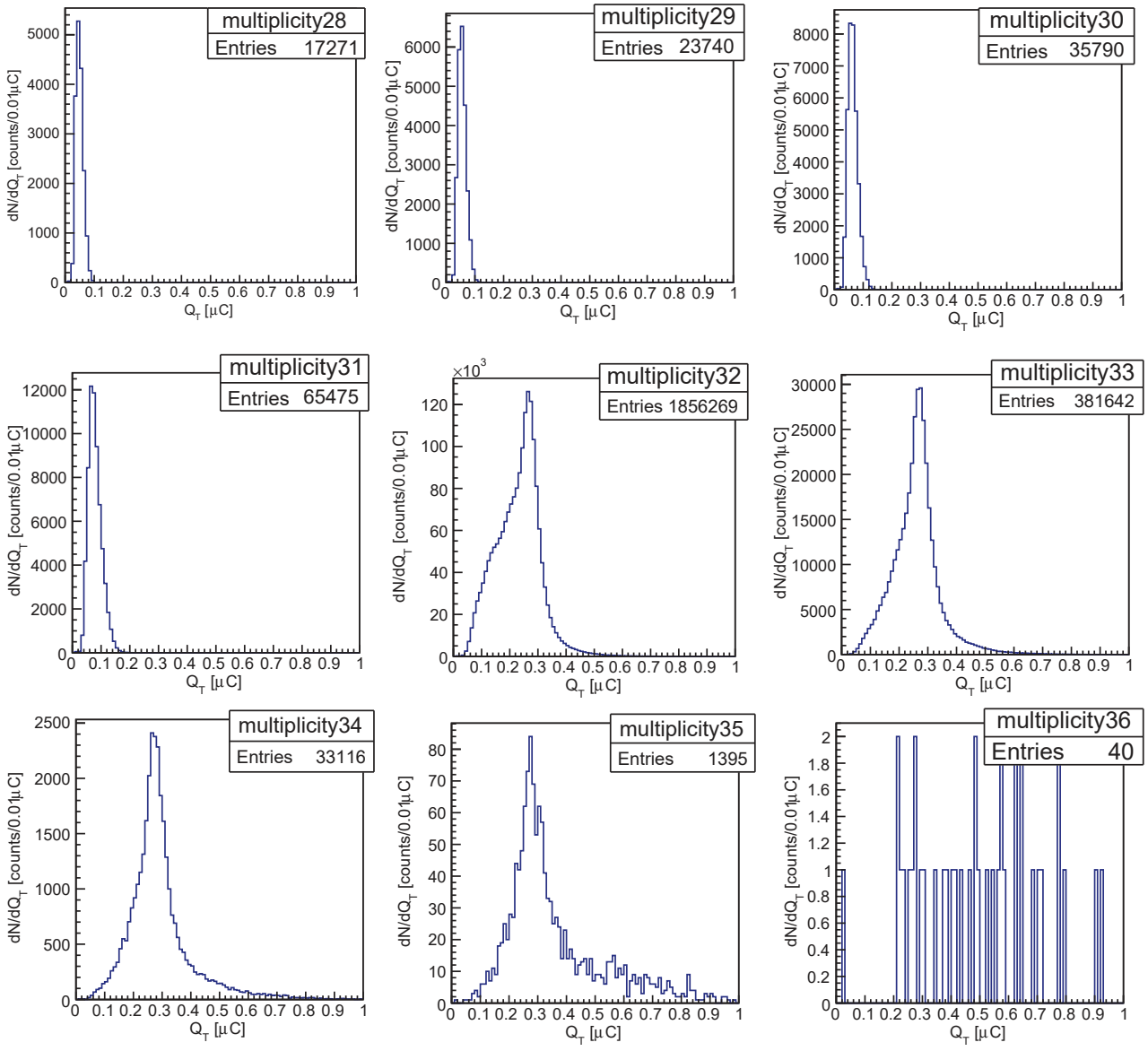

Figure 4.8: Total charge Q_T distribution for data acquired in RUN2 with $\omega \geq 28$.

To be thorough, we finally report in Figure 4.9 a plot of the total charge as function of the multiplicity, which provides an additional information on the contribution to the total charge spectrum of the events characterised by intermediate values of ω.

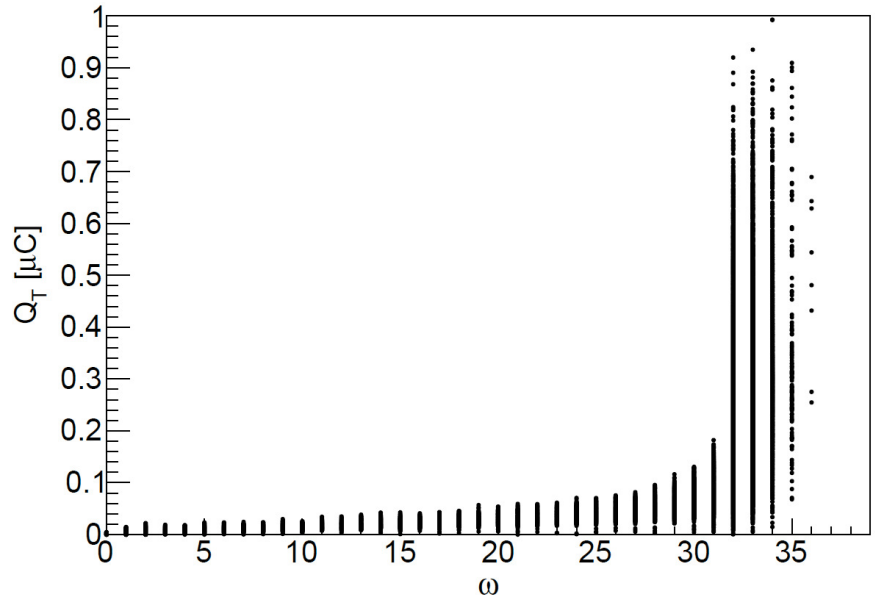

Figure 4.9: Total charge Q_T as function of multiplicity ω.

4.2.2 Charge distribution for single PMT

As anticipated, we decided to perform some additional studies focusing on the response, in particular in terms of charge, of each active channel, in order to spot and evaluate those which could be the causes of the discrepancy we noticed in the previous section; however, these studies were useful also to have a further characterization of the setup response.

First of all, we evaluated, for each channel, the mean value of the integrated charge μ_Q, calculated considering all the acquired signals; if everything was working properly, we would have expected μ_Q to assume similar values for all the channels. In Figure 4.10 we report the plot we obtained, where we can notice that, in addition to the usual problematic channels (GCU0 CH0, GCU1 CH2 and GCU3 CH0) which did not acquire any valid signal, other 4 channels (GCU2 CH2, GCU4 CH2, GCU5 CH1 and CH2) seem to provide null signals. In addition, we can notice that GCU8 CH1 (PMT n. 25) is characterised by a mean charge which is larger than the one related to the other active channels; this could be due to a greater light exposure of this PMT with respect to the others.

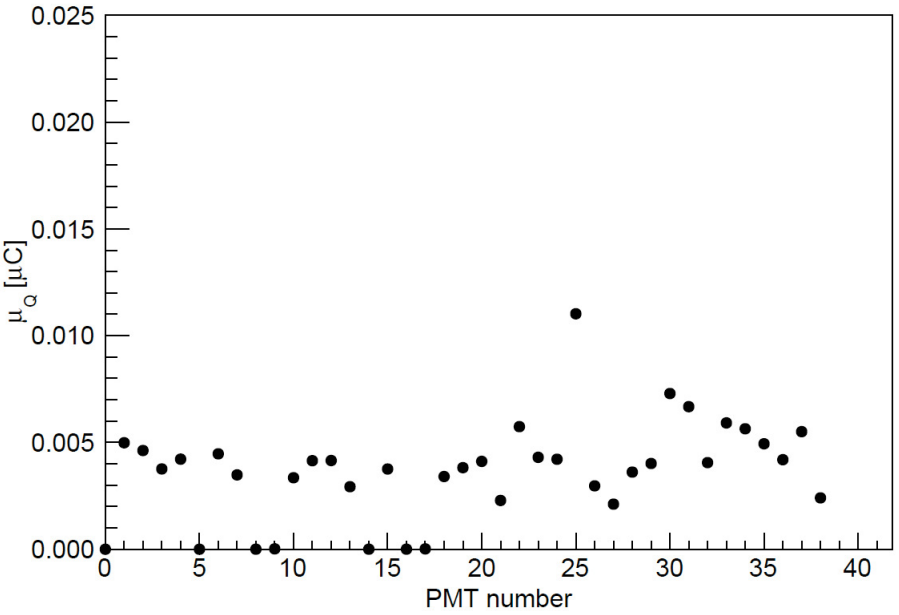

Figure 4.10: Mean value of the integrated charge μ_Q evaluated for all the 39 channels (enumerating them from 0 up to 38, where 0 is assigned to GCU0 CH0, 1 is assigned to GCU0 CH1, etc.).

In order to check if the signal of these channels was actually absent, we produced:

- some waveform plots associated to different events acquired during RUN2 (an example is reported in Figure 4.11);

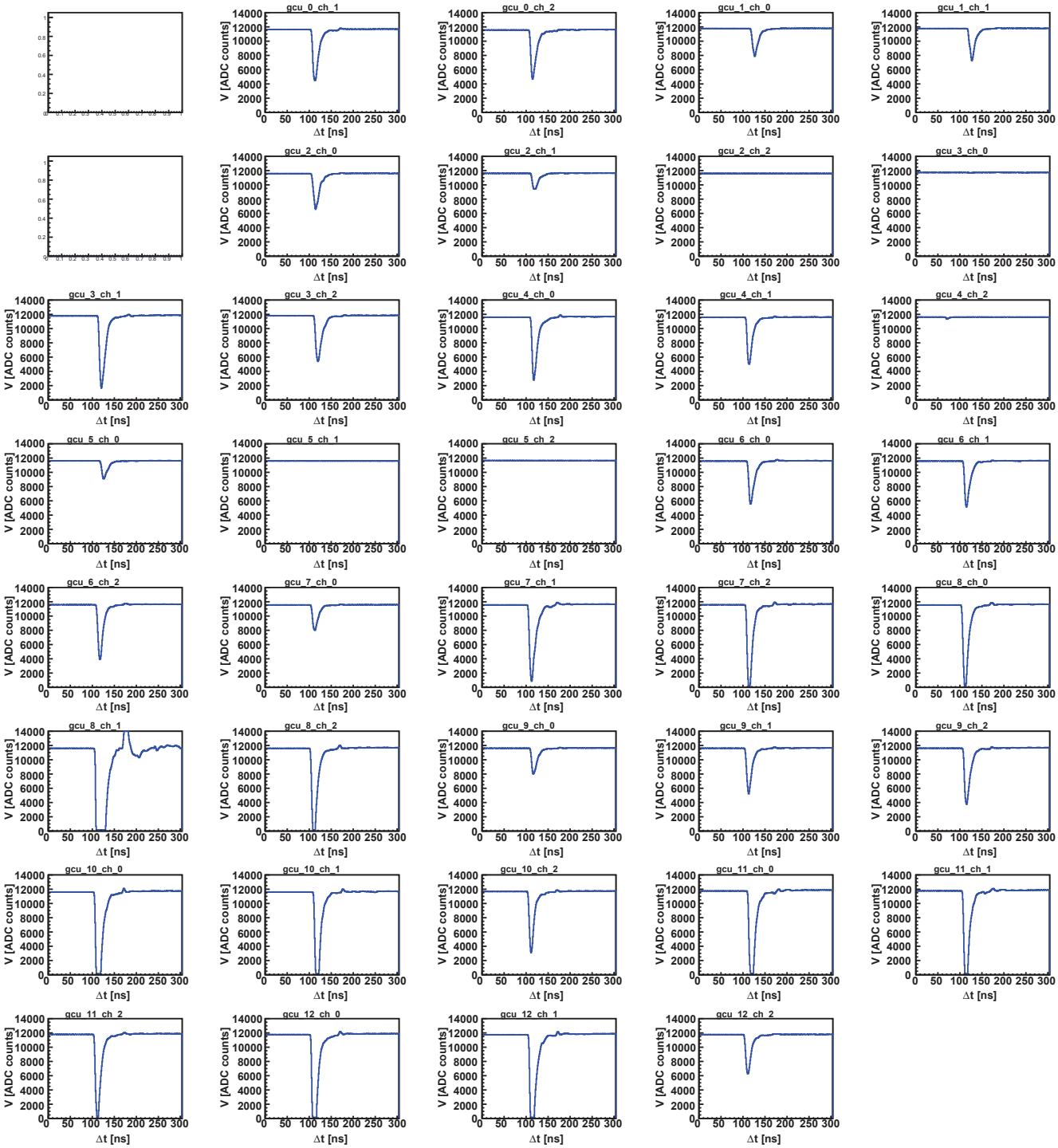

Figure 4.11: Example of waveform plot for an event acquired during RUN2.
GCU0 CH0 and GCU1 CH2 did not acquire data during the whole acquisition.

- the distributions of integrated charge evaluated for each active channel, of which we report some examples related to channels providing good signals in Figure 4.12, to be compared to those obtained for the problematic channels, reported in Figure 4.13.

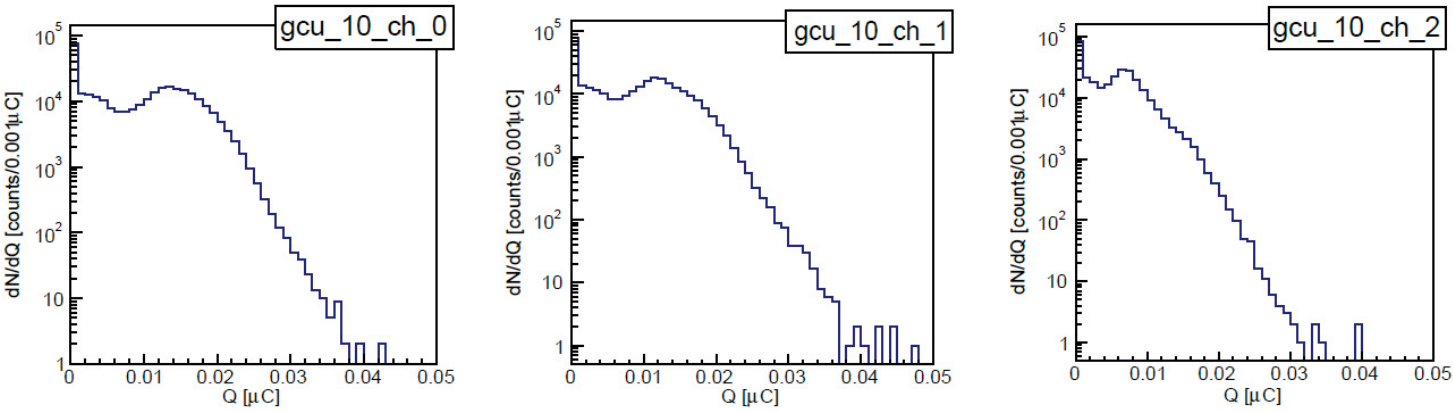

Figure 4.12: Example of integrated charge distributions for channels with good signals.

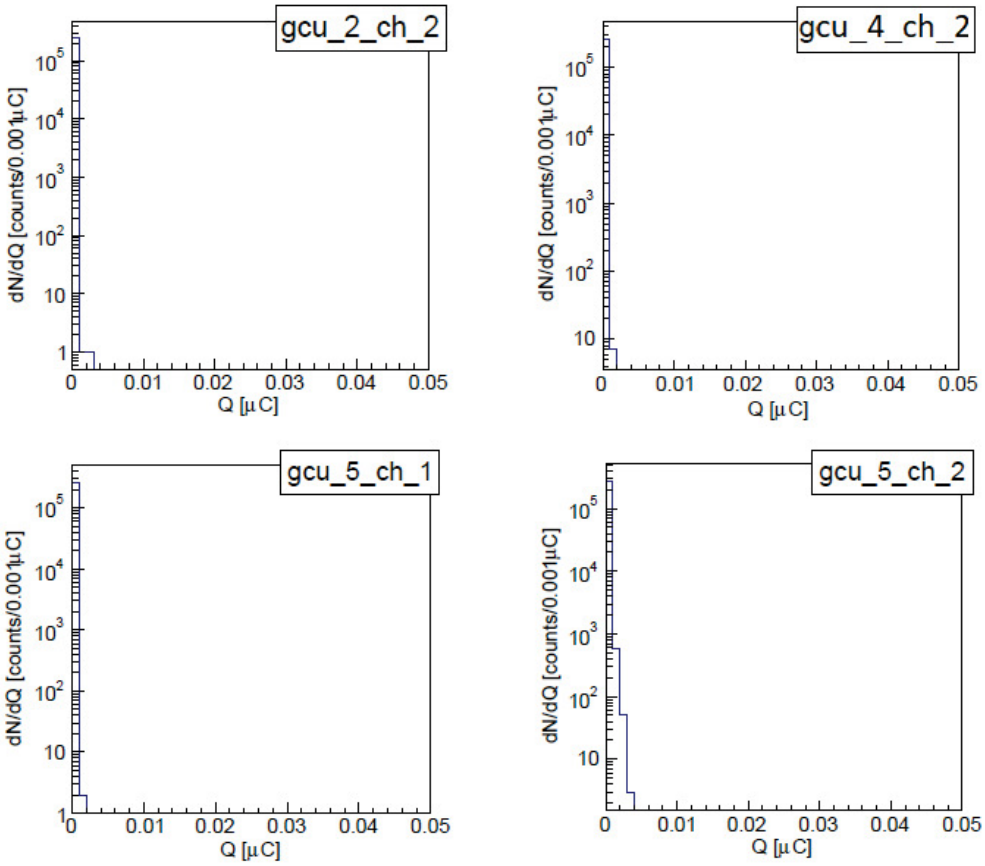

Figure 4.13: Integrated charge distributions for the 4 problematic channels.

For all the events considered in our analysis we obtained waveform plots very similar to those shown in Figure 4.11, where all the active channels show a non-negligible pulse, except for the 4 problematic channels which only show the baseline contribution to the signal.

One more confirmation that there was a problem with the signals acquired by these 4 channels is provided by the comparison between their charge distributions, reported in Figure 4.13, and those obtained for some correctly operating PMTs, shown in Figure 4.12: it is clear that the signals are actually missing.

In order to determine whether the problem was directly related to the PMT response or to GCU firmware bugs, we decided to perform some additional brief acquisitions (referred to as RUN3, RUN4 and RUN5 in Table 3.1), using three different trigger validation logics:

- RUN3 was performed in auto-trigger, therefore all the channels acquired on their own trigger validation;
- in RUN4 we employed the external trigger provided by the three external plastic scintillators;
- data in RUN5 have been acquired with a global trigger based on GCU0 CH2, setting a low threshold.

First of all we checked if there were any anomalies in the acquisition rate λ of the active channels, producing the plot reported in Figure 4.14, where we can notice that λ was very stable for all the PMTs (except for the usual GCU0 CH0 and GCU1 CH2, which did not work during all the acquisitions).

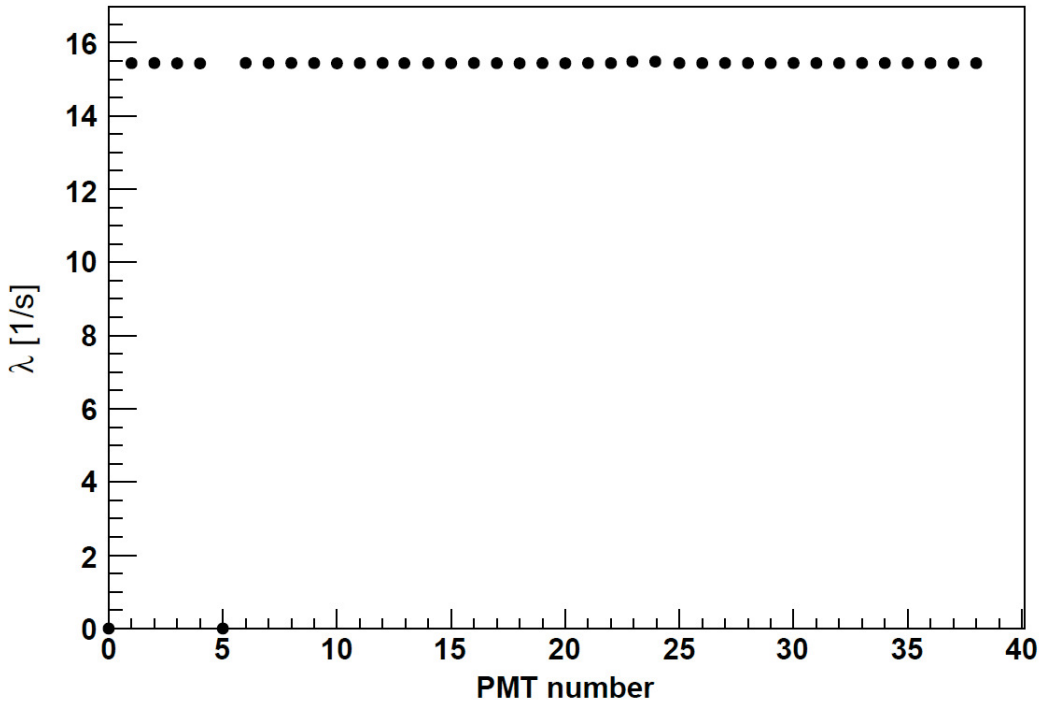

Figure 4.14: Acquisition rate of the different active channels in RUN5, performed with global trigger based on GCU0 CH2 (enumerating them from 0 up to 38, where 0 is assigned to GCU0 CH0, 1 is assigned to GCU0 CH1, etc.).

Considering the data of RUN3 acquired in auto-trigger, so that all the channels acquired on their own trigger validation, and setting a very high threshold level, we understood that the problem was actually due to bugs of the GCU firmware; in fact, in the conditions set for RUN3, we expected that all the channels stored a sufficiently high pulse at every trigger validation, but performing the analyses on the waveform plots and on the integrated charge distributions, we obtained results similar to those reported in Figure 4.11 and Figure 4.13. Therefore, even in this case, the 4 problematic channels did not show any pulse above their baseline contribution to the signal.

The problem was due to timing synchronization loss at the L1 cache level; therefore, after a correction of the firmware, we were able to retrieve the correct signals from all the active channels.

4.3 Comparison with the theoretical expectation

Another important step in the interpretation of the experimental result consisted in the comparison with the theoretical expectation of the amount of charge released in the liquid scintillator; in fact, as anticipated, the plot we reported in Figure 4.6 shows the typical trend of a MIP energy loss distribution. Therefore, we firstly expected to be able to properly fit the peak at $Q_T \sim 0.3$ μC with a Landau function L, in order to characterize the result extracting the most probable value of Q_T from the distribution [30].

The analytic expression of the Landau function used for fitting the peak is the following:

$$L(x;\mu,\sigma) = \frac{1}{\pi\sigma}\int_0^\infty e^{-t}\, cos\Big(t\,\Big(\frac{x-\mu}{\sigma}\Big) + \frac{2t}{\pi}\, log\Big(\frac{t}{\sigma}\Big)\Big)\, dt$$

where x represents the value of the total integrated charge Q_T, μ is a location parameter and it corresponds approximately to the most probable value (MPV), while σ is a scale parameter (so it does not correspond to the sigma of the full distribution, which is not defined due to the particular shape of the Landau function) [31, 32]. The result of the fitting procedure, comprehensive of the estimates of the parameters μ and σ, is reported in Figure 4.15.

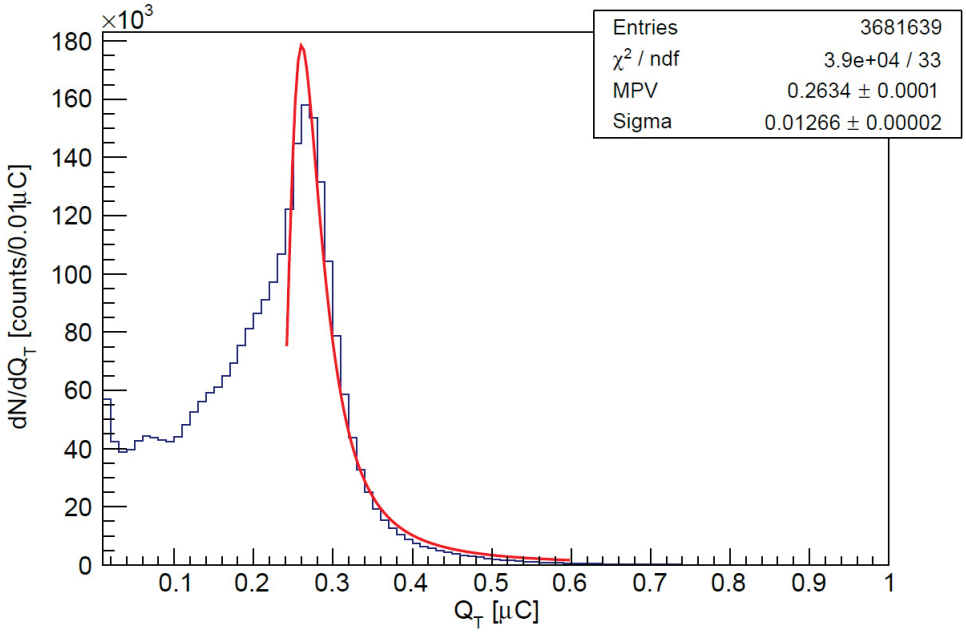

Figure 4.15: Total charge distribution fitted with Landau function [31, 32]. The estimated parameters MPV and Sigma are reported in units of μC.

As we can notice in Figure 4.15, the adopted function fits well the right tail of the distribution, which corresponds to events with a higher charge deposition, but it seems to fail in fitting the left region, where the amount of events characterized by a lower multiplicity is not negligible. A better result seemed to be obtained fitting the distribution with the so-called Crystal Ball function f_{CB} [33–35], named after the Crystal Ball Collaboration, which is a probability density function often used to model various loss processes in high-energy physics, as a further development of Landau's model; it consists of a Gaussian core portion and a power-law low-end tail, below a certain threshold, with two free parameters that are used to properly parametrize the shape of the left tail of the distribution.

The function can be expressed as:

$$f_{CB}(x;\alpha,n,\mu,\sigma) = N\cdot\begin{cases} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, & \text{for } \frac{x-\mu}{\sigma} > -\alpha \\ A\cdot\Big(\frac{n}{|\alpha|} - |\alpha| - \frac{x-\mu}{\sigma}\Big)^{-n}, & \text{for } \frac{x-\mu}{\sigma} \le -\alpha \end{cases} \qquad \text{with} \quad A = \Big(\frac{n}{|\alpha|}\Big)^n \cdot e^{-\frac{|\alpha|^2}{2}}$$

where μ and σ are the usual mean and standard deviation of the Gaussian peak, n is the exponent of the tail function, α gives the connecting point of the Gaussian and tail function, and N is a normalization factor; the result of the fitting procedure is reported in Figure 4.16.

However, both the adopted fitting procedures provides an estimate of the most probable value of deposited charge $Q_T \sim 0.263$ μC, that is basically what we needed to make a comparison with our expectation.

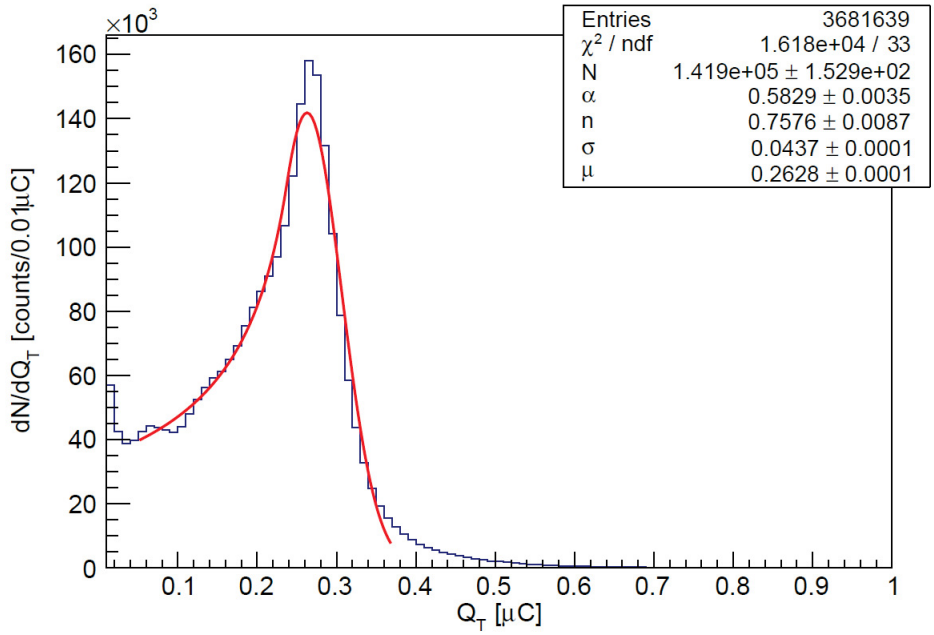

Figure 4.16: Total charge distribution fitted with Crystal Ball function [35]. The estimated parameters μ and σ are reported in units of μC, while N is reported in counts/0.01μC.

In order to obtain a theoretical estimate of the expected deposited charge, we needed to consider a chronological overview of all the processes involved in the realization of the plot shown in Figure 4.6, starting from the moment in which a muon crosses the detector, up to the moment in which we evaluate the value of the total integrated charge used to fill the histogram.

A schematic description of the main processes is presented in Figure 4.17, while a more detailed discussion on how all these processes have been considered and treated is presented in the following sections.

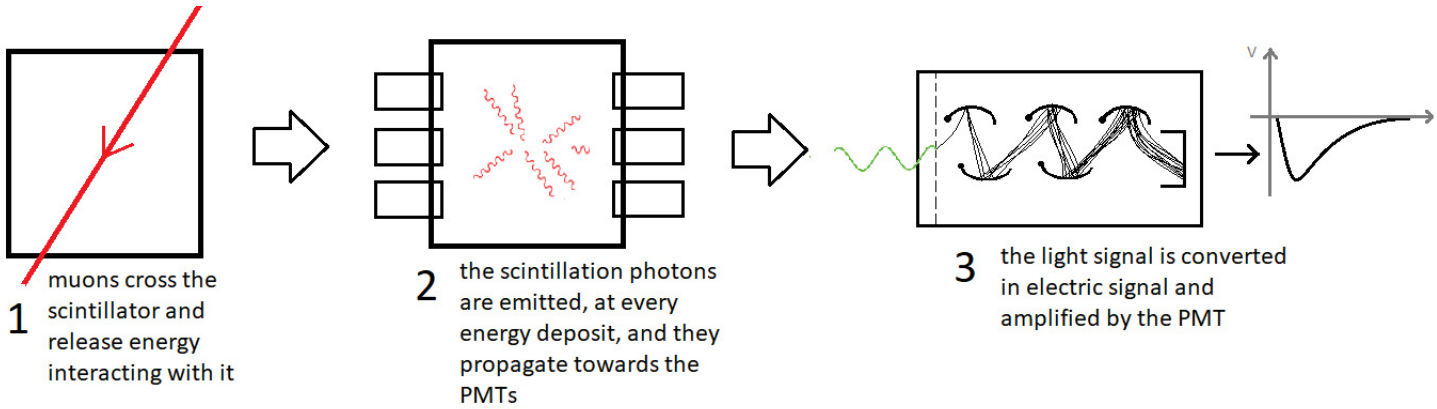

Figure 4.17: Schematic description of the main processes involved in the detector response.

4.3.1 Muon energy loss in LAB scintillator

The first process that takes place in the long chain presented in Figure 4.17 is the deposition of energy by the muons when they cross the internal vessel of the experimental setup and interact with the LAB scintillator; as a consequence, a certain quantity of photons, characterised by a specific wavelength, is emitted depending on the energy released in the scintillator.

The amount of deposited energy per unit length, also known as *stopping power* $\rho \left\langle \frac{dE}{dx} \right\rangle$, usually depends on the characteristics of the scintillating material, in particular its mass density and composition, and on the energy of the interacting particles and it can be precisely calculated with the Bethe-Bloch formula [36]. Nonetheless, considering some of the main studies and experimental results concerning the flux of cosmic rays at sea level (such as those reported in [37–39]), it is well-known and demonstrated that muons can be considered, to good approximation, as minimum ionizing particles (MIPs), so the released energy does not directly depend on their own energy.

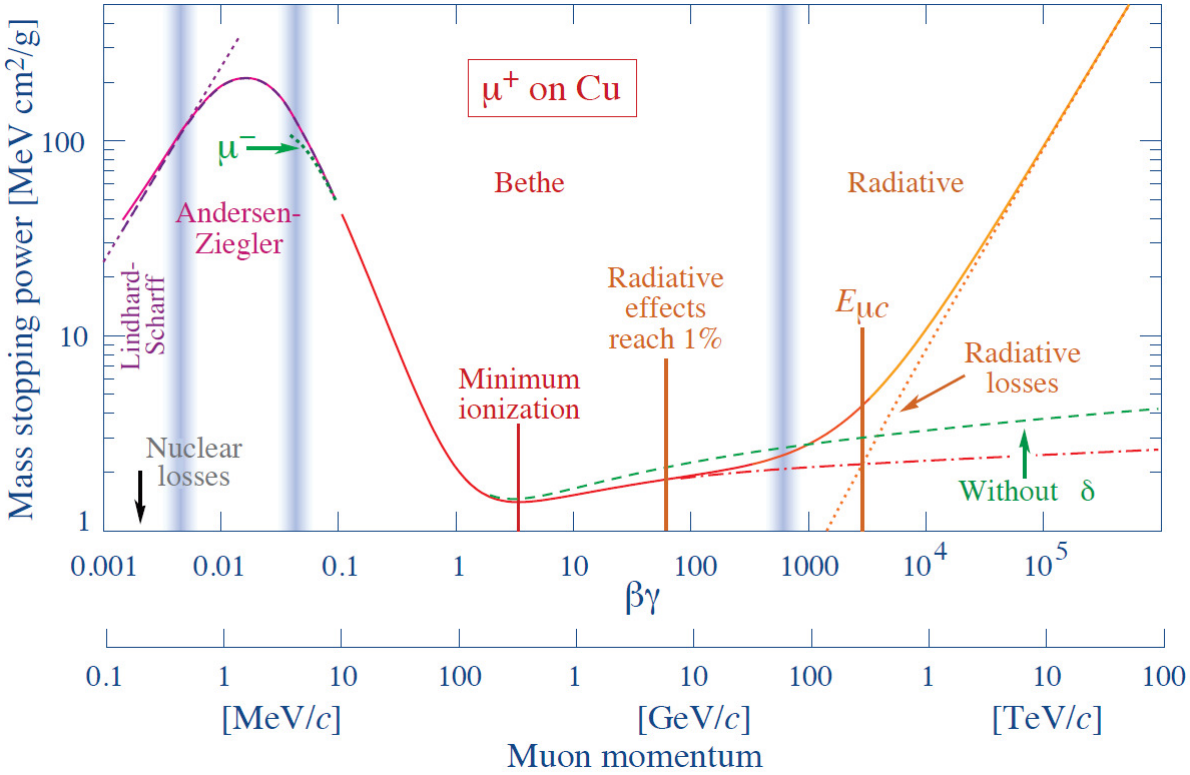

Figure 4.18: Mass stopping power for positive muons in copper as a function of $\beta\gamma = p/Mc$ over nine orders of magnitude in momentum (12 orders of magnitude in kinetic energy); solid curves indicate the total stopping power. The plot is taken from [41].

Therefore, we evaluated the *mass stopping power* $\left\langle \frac{dE}{dx} \right\rangle$, which does not depend on the scintillating material, from the plot reported in Figure 4.18, considering the muons as MIPs: $\left\langle \frac{dE}{dx} \right\rangle \sim 1.5 \, \frac{\text{MeV cm}^2}{\text{g}}$. We then calculated $\rho \left\langle \frac{dE}{dx} \right\rangle$ knowing that the mass density of the LAB scintillator [1] is $\rho_{\text{LAB}} = 0.9 \, \frac{\text{g}}{\text{cm}^3}$; the resulting deposited energy per unit length for a muon is $\rho \left\langle \frac{dE}{dx} \right\rangle \sim 1.35 \, \frac{\text{MeV}}{\text{cm}}$.

In order to evaluate the total amount of deposited energy, the length of the muon track in the liquid scintillator has to be considered; since we wanted to retrieve the value of the charge corresponding to the peak of the plot in Figure 4.6, that we demonstrated was produced by the events with the highest multiplicities (see Section 4.2.1), we considered only the muon tracks which could have most likely triggered the majority of the PMTs. Therefore, we considered a length value l between the height of the internal vessel h and the longest track l_{max}, corresponding to the diagonal length of

the rectangular section reported in Figure 4.19; the resulting range is 35 cm $< l <$ 42.67 cm and, subsequently, the range of the total deposited energy E is 47.25 MeV $< E <$ 57.6 MeV.

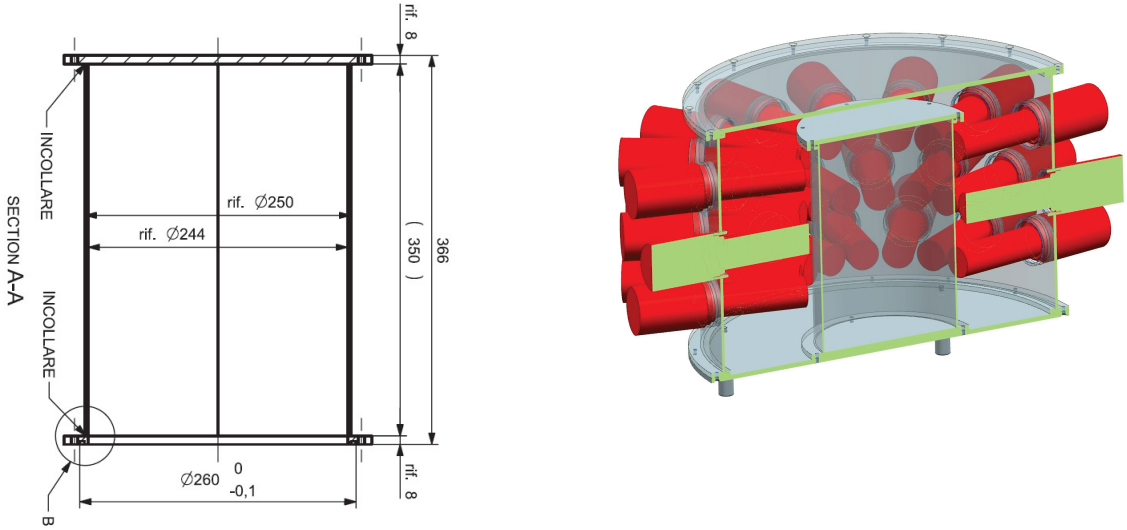

Figure 4.19: Left panel: section of the internal vessel of the experimental setup at the LNL. Right panel: section and mechanical design of the whole setup.

4.3.2 Scintillation light emission and collection

As a consequence of the interaction between the particles and the liquid scintillator, a specific amount of photons are emitted isotropically, depending on the composition of the scintillating material and on the total released energy; for the employed LAB scintillator the typical light output is $\sim 10^4$ ph/MeV [1], therefore the amount of emitted photons n_{ph} is $4.725 \cdot 10^5 < n_{\text{ph}} < 5.76 \cdot 10^5$. As we can notice in Figure 4.20, the emitted light has a wavelength mainly ranging from 400 nm up to 440 nm; in particular, the red line represented in Figure 4.20 corresponds to the maximum sensitivity of the PMTs of the experimental setup. Once the light is emitted, it propagates in the scintillator and it

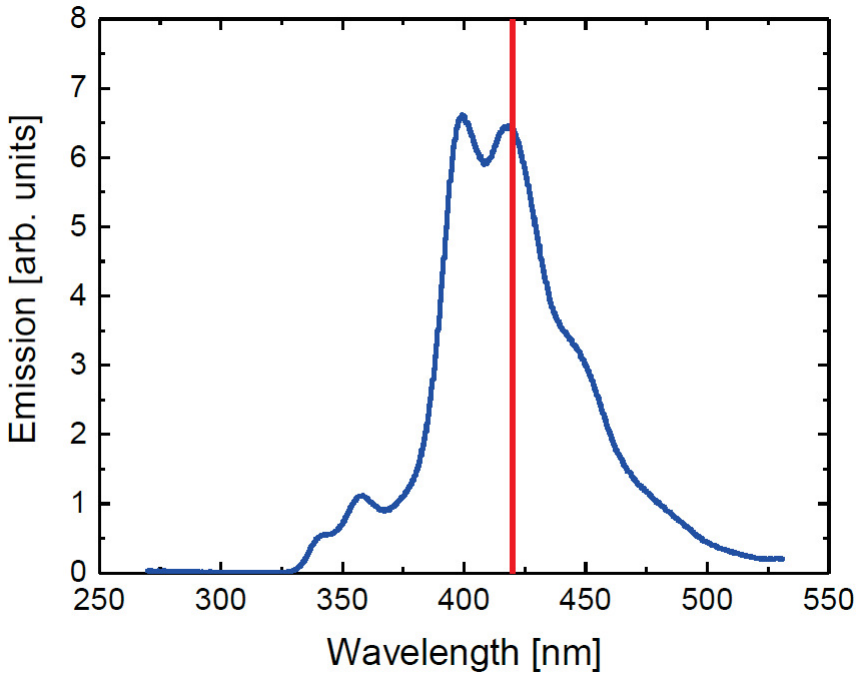

Figure 4.20: Emission spectrum of the LAB scintillator (measurement performed by dr. Fausto Ortica, INFN, Perugia division). The red line at 420 nm corresponds to the maximum sensitivity of the employed PMTs (taken from the datasheet).

finally reaches the PMTs where the light signal is converted into an electric signal; usually during the propagation there is an attenuation of the light signal, which is proportional to the distance D between the emission point and the position of the PMTs and it is described by a decreasing exponential law $n_{ph}(D) = n_{ph,0} \cdot e^{-\frac{D}{\lambda}}$, where λ is the so-called *attenuation length* and $n_{ph,0}$ is the initial amount of photons emitted by the scintillator. In JUNO experiment this light attenuation is surely important, but it is negligible in our test system due to the small vessel size; therefore, we assumed that no photon was lost along the distance D.

In addition to this, it is important to underline that for the non-vertical tracks the amount of light collected by each PMT could be very different from one PMT to another, but due to the fact that we focused on events characterised by high multiplicities, we could assume that all of them collected the same quantity of light. Finally, we had to consider that not all the lateral surface of the internal vessel is sensitive to photons, since the PMTs cover only a limited fraction, called *photo-coverage*, of this surface; therefore, considering that photons are emitted isotropically by the scintillator and that the photo-coverage of the setup is $\sim 50\%$, we evaluated that the number of detected photons ranges from $\sim 2.4 \cdot 10^5$ to $\sim 2.9 \cdot 10^5$.

4.3.3 Light signal conversion and amplification

Part of the photons that reach the PMTs are then converted into photo-electrons as a consequence of the photoelectric effect, which takes place, as a probabilistic process, in the PMT photo-cathodes; clearly, not all the N_{ph} incident photons are absorbed by the photo-cathode, but the number of emitted photo-electrons N_{e^-} can be evaluated considering the characteristic *quantum efficiency* $QE = \frac{N_{e^-}}{N_{ph}}$ of the PMT, which depends on the specific wavelength of the incident light.

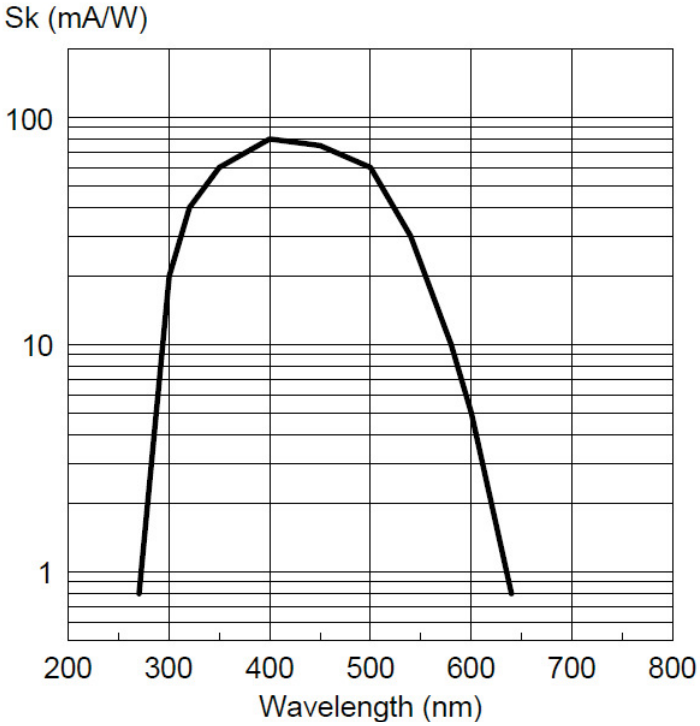

Figure 4.21: The typical spectral characteristics of the XP2020 PMTs taken from the datasheet, on the y axis the radiant sensitivity of the PMT indicated with S_k.

The plot reported in Figure 4.21 shows the relation between the *radiant responsivity* S_K of the employed PMTs as a function of the photons' wavelength λ; S_K is defined as the photoelectric current generated by the photo-cathode divided by the incident radiant flux at a given wavelength, expressed in units of milliamperes per watts (mA/W), which, from a practical point of view, is easier to be

estimated than QE. However, the following relation between the two quantities holds [42]:

$$QE = \frac{S_K[A/W]}{\lambda} \cdot \frac{h\,c}{e} \sim \frac{S_K[A/W]}{\lambda} \cdot 1240\,\frac{W\,nm}{A}$$

where λ is the wavelength of the incident light, h is the Planck constant, c is the speed of light in vacuum and e is the elementary charge.

Comparing the plot in Figure 4.21 and the emission spectrum reported in Figure 4.20, we could assume a wavelength of $\lambda = 420$ nm for the incident light, considering it approximately monochromatic, which led to $QE \simeq 0.236$; therefore, the number of emitted photo-electrons is $5.6{\cdot}10^4 < N_{e^-} < 6.8{\cdot}10^4$.

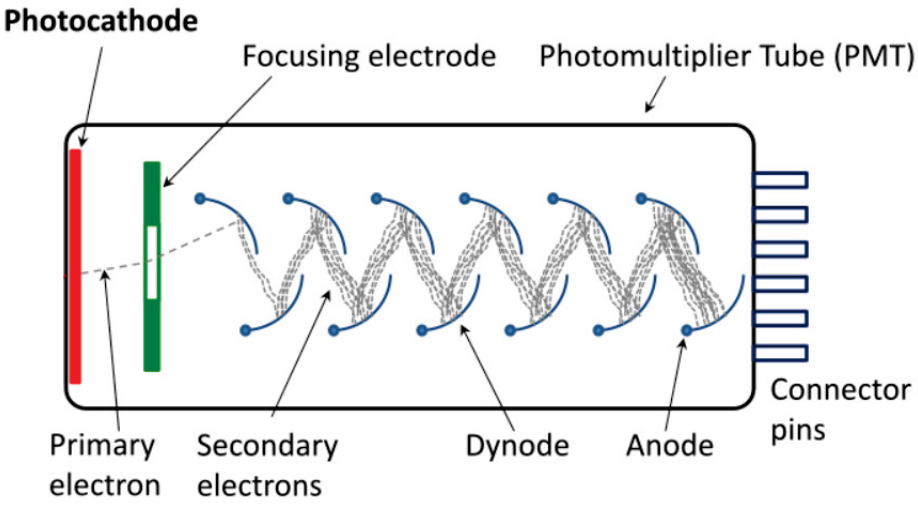

Figure 4.22: Schematic description of the PMT electron multiplier.

Since the number of emitted photo-electrons is typically too small to produce a non-negligible voltage signal, they are directed by the focusing electrode towards an electron multiplier, consisting of a number of electrodes, the *dynodes*, with an increasing potential, where the electrons are multiplied by the process of secondary emission (see Figure 4.22). Therefore, the charge we reconstruct with the procedure described in Section 3.3.3 is not exactly the one released by the particle in the scintillator, but it is related to it by a multiplying factor, which is the *gain* g of the PMT, depending on the supply voltage with which it has been set in operation. In particular, for our setup the supply voltages have been properly set in order to obtain $g \sim 10^7$ for each PMT; thus, we estimated that the final value of the charge, $\mathrm{Q}_T^{th} = N_{e^-} \cdot g \cdot e$, should range in:

$$0.09\,\mu\mathrm{C} < Q_T < 0.1\,\mu\mathrm{C}$$

Comparing this theoretical result with the value of the peak position we retrieved experimentally, $\mathrm{Q}_T^{exp} = (0.2628 \pm 0.0001)\,\mu$C, we can notice that they have the same order of magnitude, even if the expected value is slightly lower than the measured one; the small discrepancy that is present is certainly due to the several approximations we made in the different steps of our theoretical calculations, in addition to possible processes we did not take into account, regarding in particular the production of the PMT output voltage signal, its conversion into a digital signal by the ADU and the final signal integration procedure. All these steps could affect the final estimate of Q_T^{exp}, but we were limited in correctly estimating their actual contribution to the final result.

Chapter 5

Energy calibration of the experimental setup

An additional and important step in the interpretation of the spectrum obtained experimentally is energy calibration.

A simple procedure for calibrating in energy the response of an experimental setup is based on the use of a radioactive source characterised by a simple and well-known decay scheme; in fact, the identification of the peaks of its spectral emission allows a proper reconstruction and calibration of the acquired spectrum, which can then be applied also to other acquisitions.

We decided to employ four different radioactive sources:

- ^{241}Am, that decays 100% by alpha transitions to ^{237}Np. Most of the decays (84.6%) populate the excited level of ^{237}Np with energy of 59.54 keV, so decaying to the ground level it emits gamma radiation with this characteristic energy, which we expected to measure [44]; our source was produced in 04.01.2014 with an activity A = 371 kBq;
- ^{137}Cs, which disintegrates by β^- emission to the ground state of ^{137}Ba (5,6%) and via the 661 keV isomeric level of ^{137}Ba (94,4%), which has a half-life of 2.55 min [44]; the latter decay is the one we expected to detect with our setup; our ^{137}Cs source was produced in 31.12.2014 with A = 411 kBq;
- ^{22}Na, that disintegrates by β^+ emission predominantly to the 1275 keV level of ^{22}Ne, while a very small fraction (0,056%) disintegrates to the ground state of ^{22}Ne [44]; our ^{22}Na source was produced in 03.12.2014 with A = 388 kBq;
- ^{60}Co, which disintegrates by β^- emission to excited levels of Ni-60, that decay to the ground state by emitting gamma radiation mainly at $E_{\gamma,1}$ = 1173 keV and $E_{\gamma,2}$ = 1332 keV [44]; our ^{60}Co source was produced in 03.12.2014 with A = 376 kBq.

5.1 Preliminary Monte Carlo simulations of the setup response

Monte Carlo simulations [45] of the setup response have been performed in order to set the proper acquisition conditions based on a simulated expectation. The radioactive sources have been assumed as point-like and positioned on the top of the liquid scintillator vessel, simulated both as collimated, in order to select only the photons with vertical and downward direction, and as uncollimated, in order to simulate photons emitted isotropically in all the possible directions.

The plot reported in Figure 5.1 shows the expected event multiplicities for all the sources and for the muon events; as we can notice, we should expect a very low multiplicity for the events related to the source activities, which clearly implies, in terms of acquisition, higher trigger rates and, subsequently, a large amount of stored information even in a very limited acquisition time. For this reason, the processing code used so far would have required increasingly longer times for elaborating and analysing data from all the active channels, therefore some modifications have been implemented, in order to speed up the reconstruction of the acquired data (see Appendix C for details).

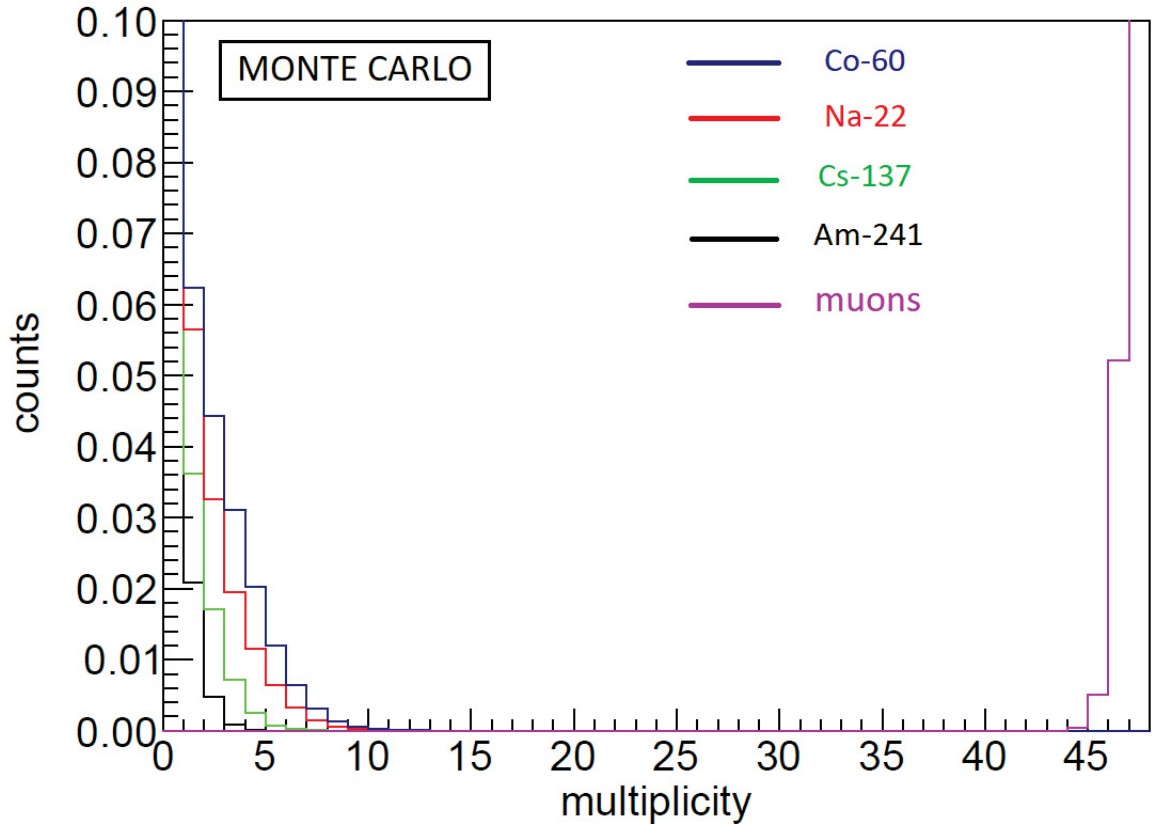

Figure 5.1: Simulated multiplicity of the events related to the activity of the collimated radioactive sources and to the cosmic muons.

The plots reported in Figure 5.2 and Figure 5.3 shows the simulated distributions of the number of photons N_γ absorbed by the PMT photocathodes, both with collimated and uncollimated sources; the number of absorbed photons is clearly proportional to the deposited charge and, consequently, to the energy released in the liquid scintillator. Both the plots provide a very important information regarding the setup response, since we can notice that the system is not expected to detect the full-energy peaks of the radioactive sources, but only their Compton continuum; this is probably due to the characteristic density of the employed scintillator in combination with the small size of its vessel, which allows the gammas to escape from the detector without releasing all their energy. Therefore, the energy calibration with the radioactive sources should be based on the determination of the Compton edges, whose energy $E_{e,max}$ is related to the expected full gamma energy E_γ through the relation $E_{e,max} = \frac{2E_\gamma^2}{2E_\gamma + m_e}$:

- ^{241}Am: $E_{e,max} = 11$ keV;
- ^{137}Cs: $E_{e,max} = 477$ keV;
- ^{22}Na: $E^1_{e,max} = 341$ keV and $E^2_{e,max} = 1062$ keV;
- ^{60}Co: $E^1_{e,max} = 963$ keV and $E^2_{e,max} = 1118$ keV.

Comparing the two plots, we can notice that the two distributions show a very similar trend; however, we decided to proceed acquiring data with collimated sources, using Pb shields.

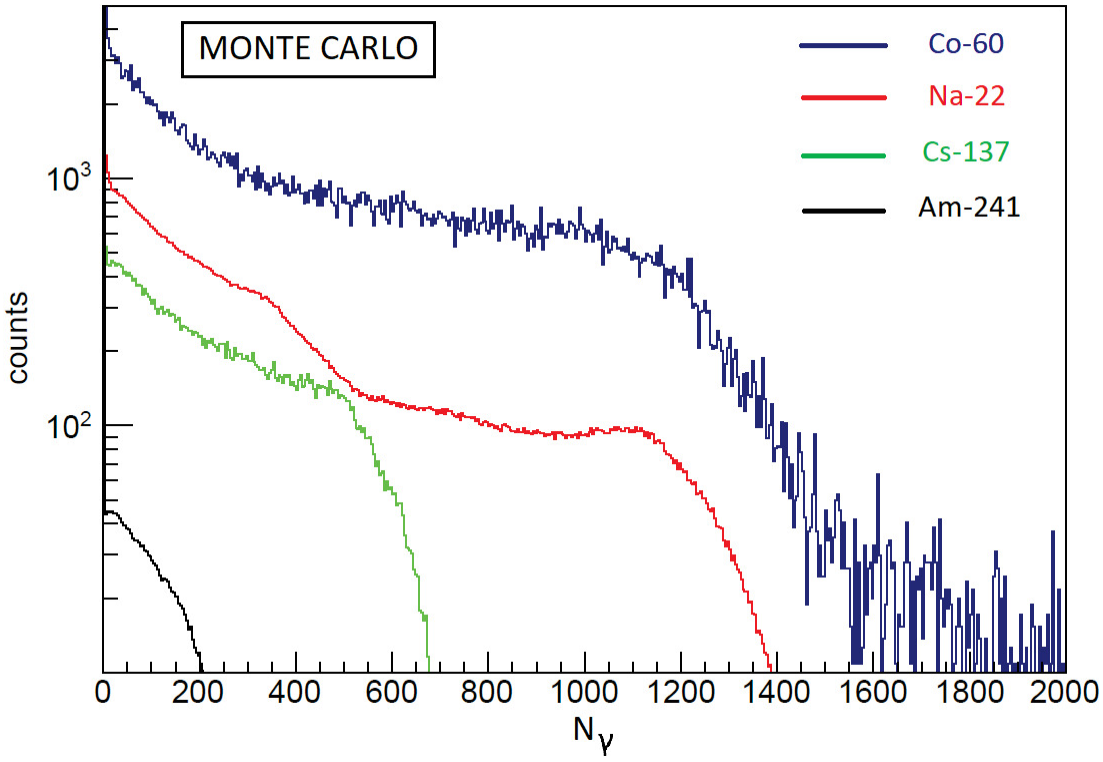

Figure 5.2: Simulated distributions of the number of photons absorbed by the PMT photocathodes with collimated sources.

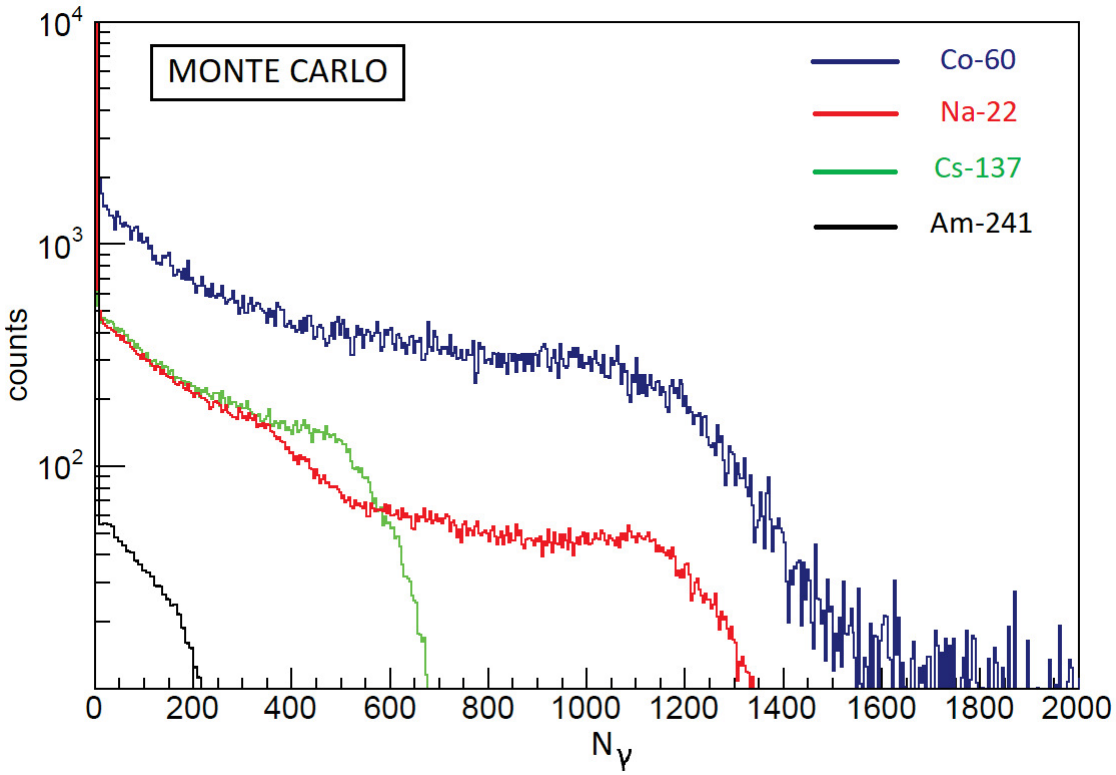

Figure 5.3: Simulated distributions of the number of photons absorbed by the PMT photocathodes with uncollimated sources.

5.2 Results of the acquisitions with the radioactive sources

Using the results of the Monte Carlo simulations, we decided to perform 5 different acquisitions with the following settings:

- duration of 15 minutes each;
- using a global trigger validation based on the logic AND of 3 PMTs;
- the threshold level set for each channel was equal to 6 baseline standard deviations σ_B from the baseline contribution to the signal.

The first acquisition was performed without any radioactive source in order to correctly determine the background contribution to the acquired data, mainly due to environmental radioactivity; then, four different acquisitions have been performed for each radioactive source, opportunely positioned on the top of the setup and collimated using a Pb shield, with the following, resulting trigger rates:

- background only: 1.97 kHz;
- with ^{241}Am: 2.00 kHz;
- with ^{137}Cs: 2.45 kHz;
- with ^{22}Na: 2.39 kHz;
- with ^{60}Co: 3.30 kHz.

A comparison between the spectrum acquired without any radioactive source and the one acquired with the ^{22}Na source, opportunely normalized in time, is reported in Figure 5.4, where we can appreciate the difference between the two acquisitions.

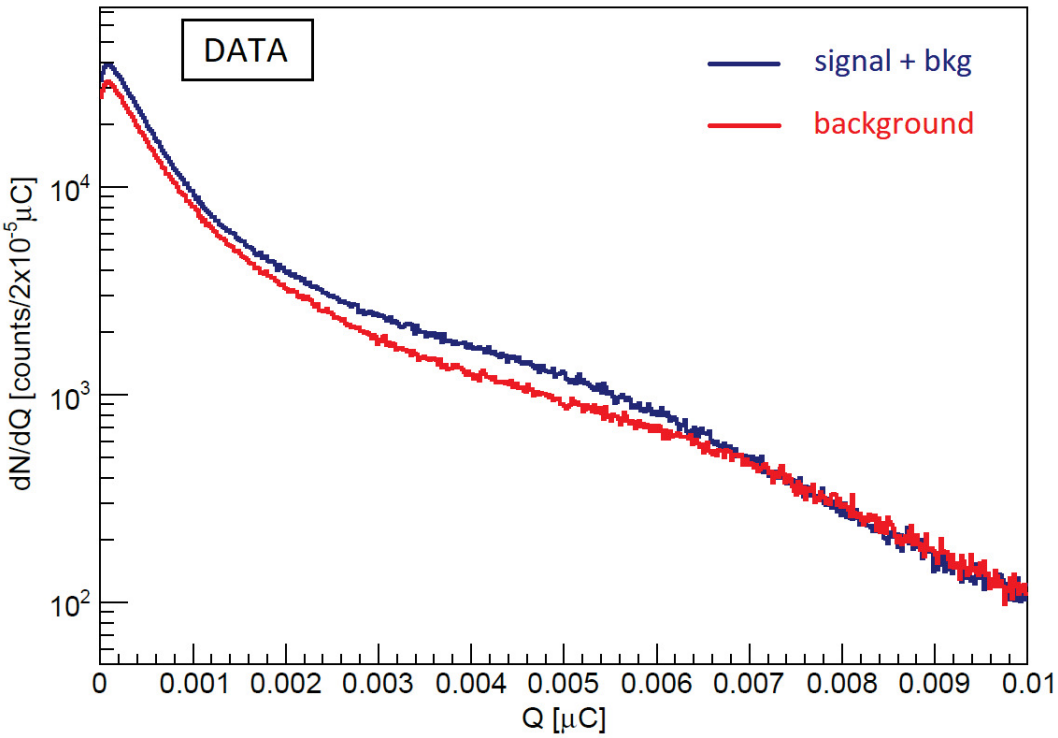

Figure 5.4: Comparison between acquired background spectrum (in red) and spectrum acquired with the ^{22}Na source (in blue), opportunely normalized in time.

Then, the background spectrum is subtracted from each spectrum acquired with the radioactive sources, producing the resulting plot reported in Figure 5.5. Comparing the experimental result reported in Figure 5.5 and the results of the Monte Carlo simulations reported in Figure 5.2 and Figure 5.3, we can notice that there is a good agreement with the expectation.

The main difference with the results of the Monte Carlo simulations consists in the fact that the

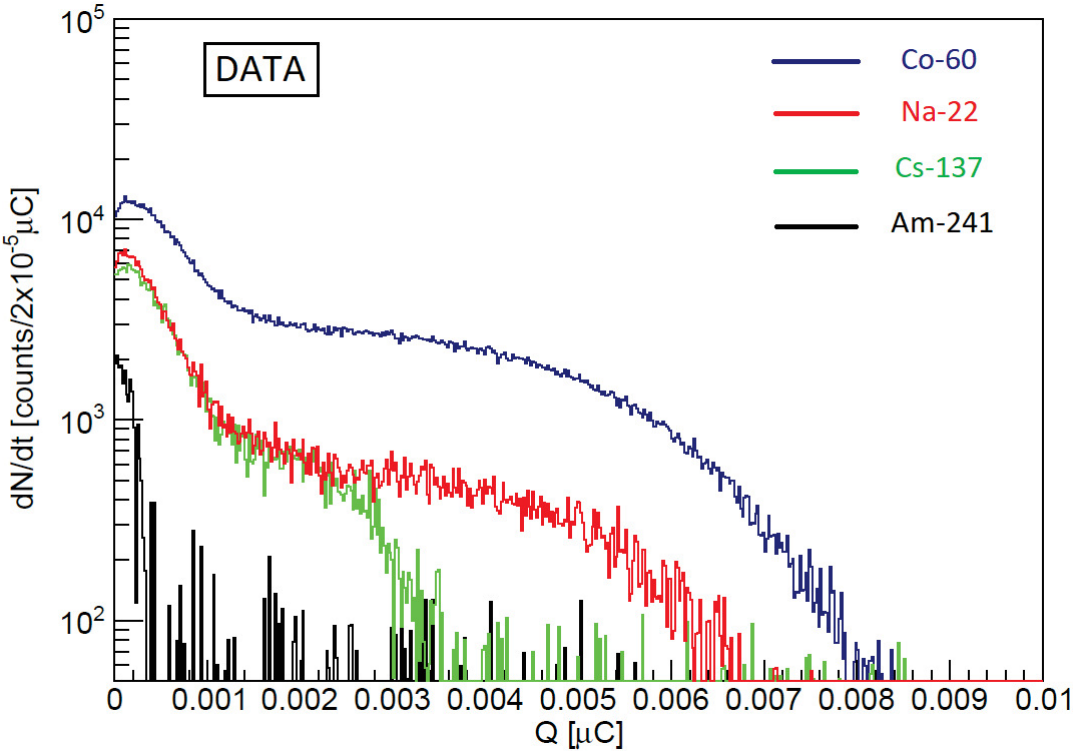

Figure 5.5: Net radioactive sources spectra obtained with the experimental measurements.

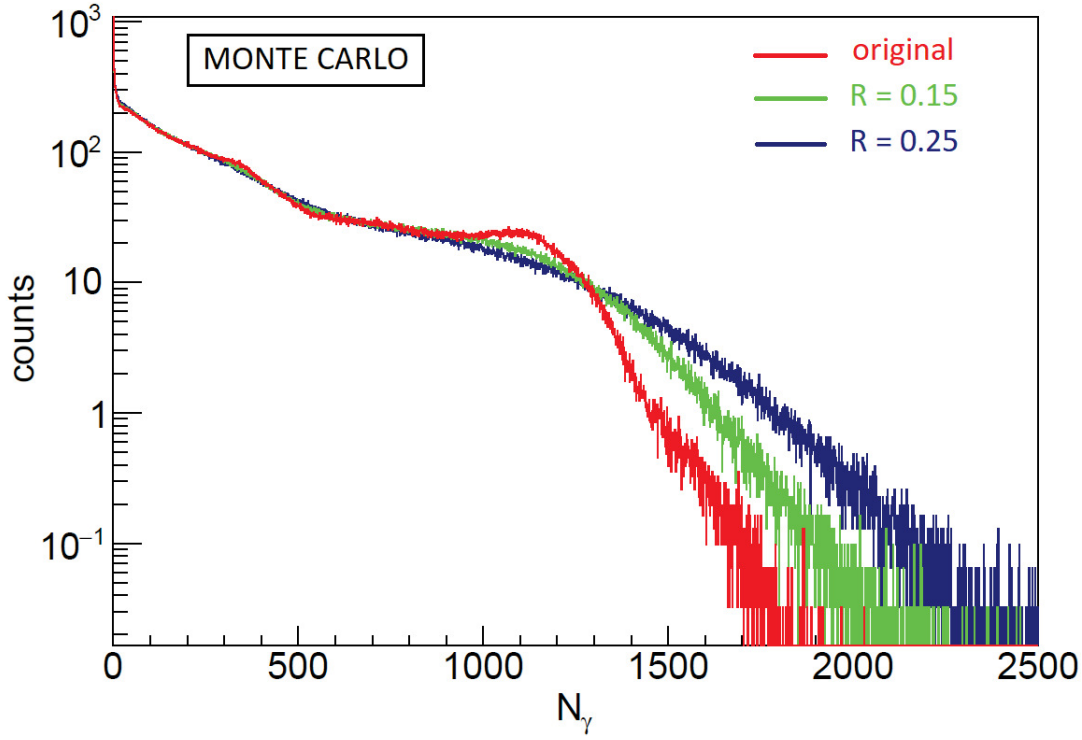

Figure 5.6: Original simulated spectrum and spectra resulting from the smearing procedure with different resolution R, for the ^{22}Na collimated source.

experimental Compton edges appear much more smoothed; this happens because the result of the ideal simulation does not take into account the finite resolution of the experimental setup.

A simple demonstration which could support this consideration is provided by the comparison of one of the spectra reported in Figure 5.2 with the result of a smearing procedure, which can simulate the effects of a finite resolution of the detector, applied to the considered spectrum.

The smearing procedure consists in producing a new spectrum from the original one, where each bin is substituted by a Gaussian distribution characterised by $\mu = x_i$ and $\sigma = x_i \cdot R$, where x_i corresponds to the i-th bin center and R is the simulated resolution. Figure 5.6 shows a comparison between the original ^{22}Na spectrum, obtained in the previously discussed simulation of the collimated sources, and two new spectra obtained applying the smearing procedure, assuming R = 0.15 and R = 0.25. As we can notice, the worse is the resolution, the smoother are the resulting spectra.

5.3 Determination of the energy calibration parameters

In order to extract the positions of the Compton edge for each source, which were necessary for providing an energy calibration of the setup, we decided to perform a fit on the experimental spectra using the complementary error function *Erfc* [46]:

$$Erfc(x) = \frac{2A}{\pi} \int_{\frac{x-\mu}{\sigma}}^{\infty} e^{-t^2} dt + \mathrm{C}$$

where μ is the position of the curve half height, corresponding to the estimated position of the Compton edge, σ is a parameter associated to the smearing of the curve and A, C are constants.

An example of fit on the experimental spectrum is reported in Figure 5.7, while the estimated parameters for all the considered sources are reported in Table 5.1.

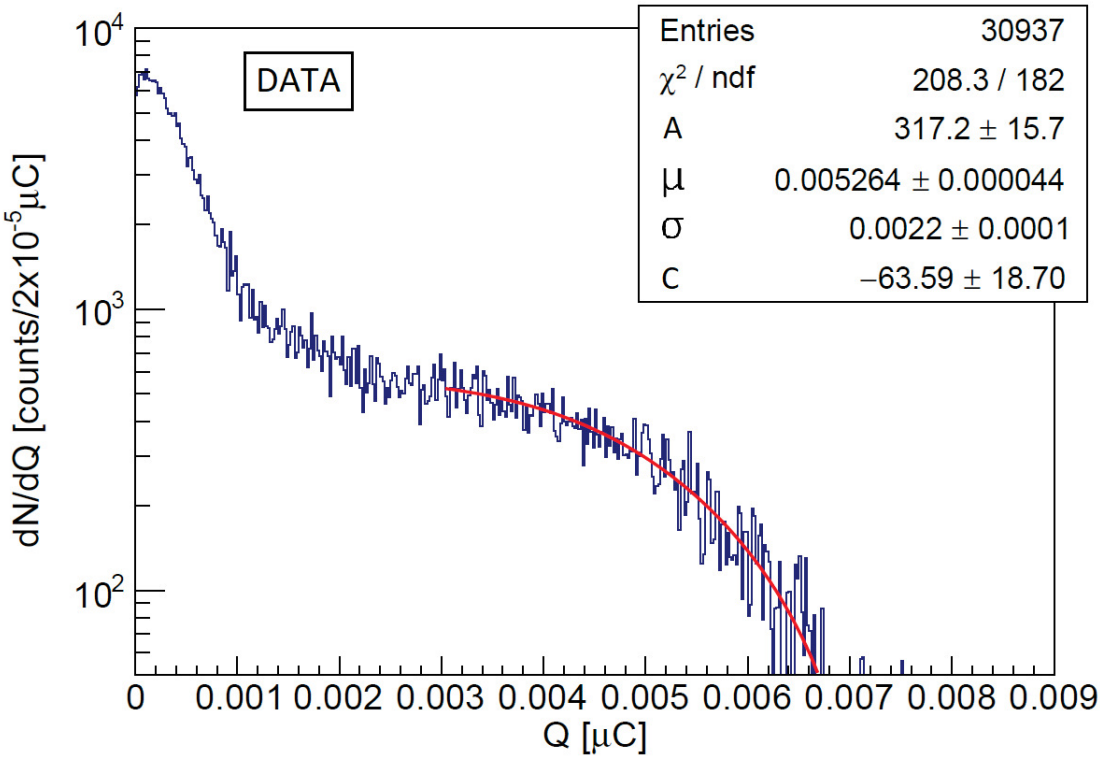

Figure 5.7: Example of fit with the complementary error function on the ^{22}Na experimental spectrum. The parameters μ and σ are reported in units of μC, while A and C are reported in units of counts/$2 \cdot 10^{-5}\mu$C.

source	A [counts/$2 \cdot 10^{-5}\mu$C]	μ [$10^{-5}\mu$C]	σ [$10^{-5}\mu$C]	C [counts/$2 \cdot 10^{-5}\mu$C]	$E_{e,max}$ [keV]
^{241}Am	369 ± 11	20.6 ± 0.5	13.7 ± 0.5	30.2 ± 0.7	11
^{137}Cs	359 ± 9	267 ± 2	82 ± 3	9 ± 4	477
^{22}Na	317 ± 16	526 ± 4	220 ± 10	-64 ± 19	1062
^{60}Co	1225 ± 8	548 ± 3	180 ± 4	12 ± 8	1118

Table 5.1: Results of the fitting procedure with the complementary error function, associated to the expected values for the Compton edge, for all the employed sources.

The estimated Compton edge positions μ and the associated expected values in keV $E_{e,max}$ are reported in Table 5.1; since we noticed that we had not sufficient data to obtain an accurate estimate of the ^{241}Am Compton edge, we decided to exclude it from the fitting procedure performed to determine the calibration parameters. The associated quantities μ and $E_{e,max}$, for ^{137}Cs, ^{22}Na and ^{60}Co, have been used to perform a linear fit, whose result is shown in Figure 5.8, in order to obtain the parameters for the energy calibration; the estimated values for the angular coefficient m and for the intercept q are $m = (0.228 \pm 0.003)$ MeV/nC and $q = (-0.13 \pm 0.01)$ MeV.

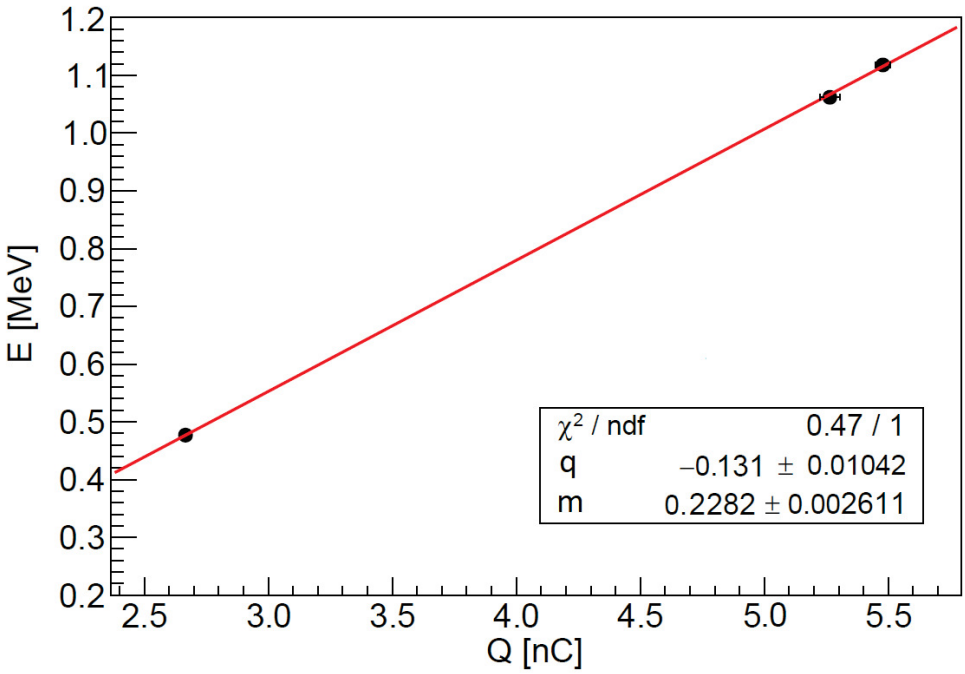

Figure 5.8: Result of the linear fit for the determination of the calibration parameters, excluding ^{241}Am; m is reported in MeV/nC and q is reported in MeV.

The estimated parameters m and q have been used to calibrate the muon spectrum presented in Section 4.2, obtaining the result reported in Figure 5.9. It is interesting to compare this plot with the result of a Monte Carlo simulation that was performed considering cosmic muons at sea level, only with vertical direction; the result of this simulation is reported in Figure 5.10, fitted with a Landau function to extract its centroid value. As already discussed in previous sections, we expected muons with vertical direction to cross the largest distance in the liquid scintillator, releasing, as a consequence, the largest amount of energy and producing the characteristic peak that can be observed in the spectrum shown in Figure 5.9. Therefore, the peak obtained in the Monte Carlo simulation should correspond to the peak present in the experimental spectrum; as we can notice there is a very good agreement between the two results, since they are both centered in ~ 59 MeV.

Finally, a comparison between the extrapolated value for the cosmic muon peak and the Compton edge energies of the radioactive sources is reported in Figure 5.11, where we can have a clear view of the dynamic range of the experimental setup.

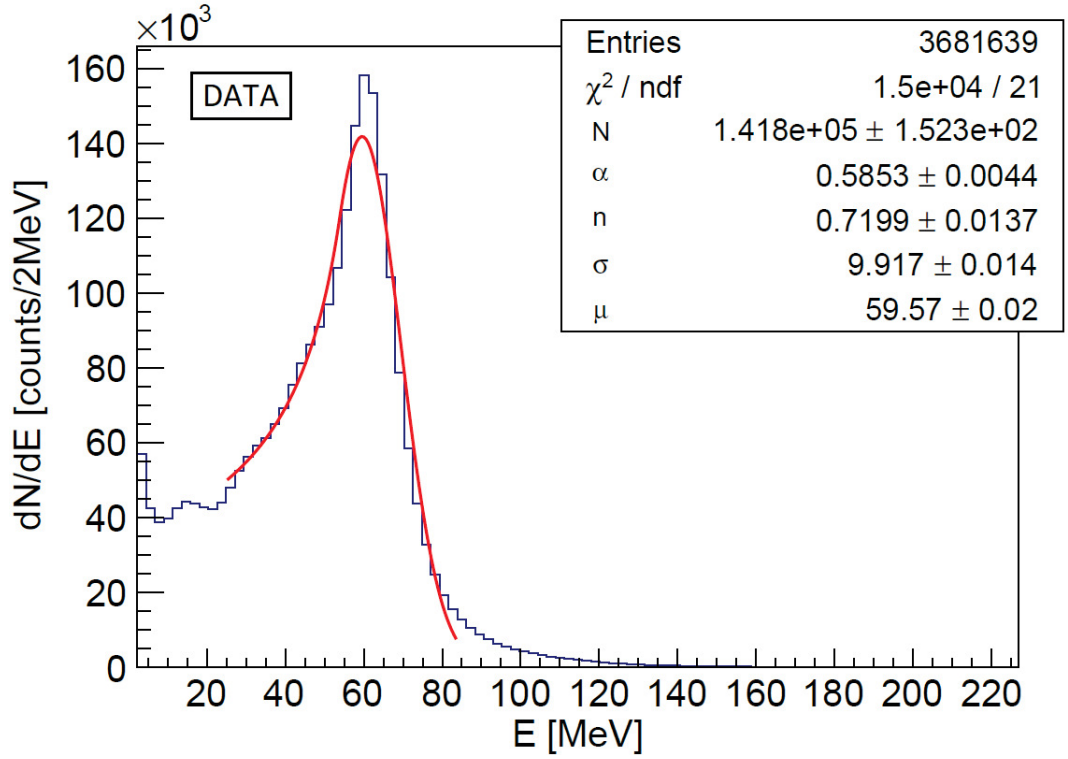

Figure 5.9: Experimental muon energy spectrum, fitted with Crystal Ball function [35]. The estimated parameters μ and σ are reported in units of MeV, while N is reported in counts/2MeV.

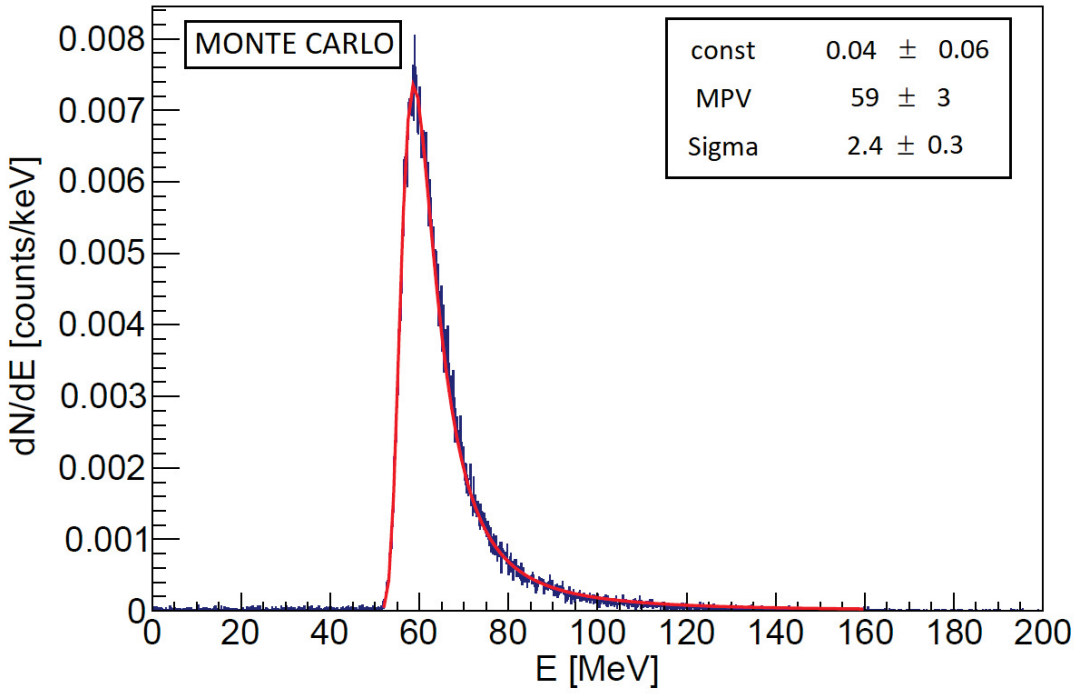

Figure 5.10: Simulated energy spectrum of cosmic muons at sea level, with vertical direction; the peak is fitted with a Landau function. The estimated parameters MPV and Sigma are reported in units of MeV, while const is reported in counts/keV.

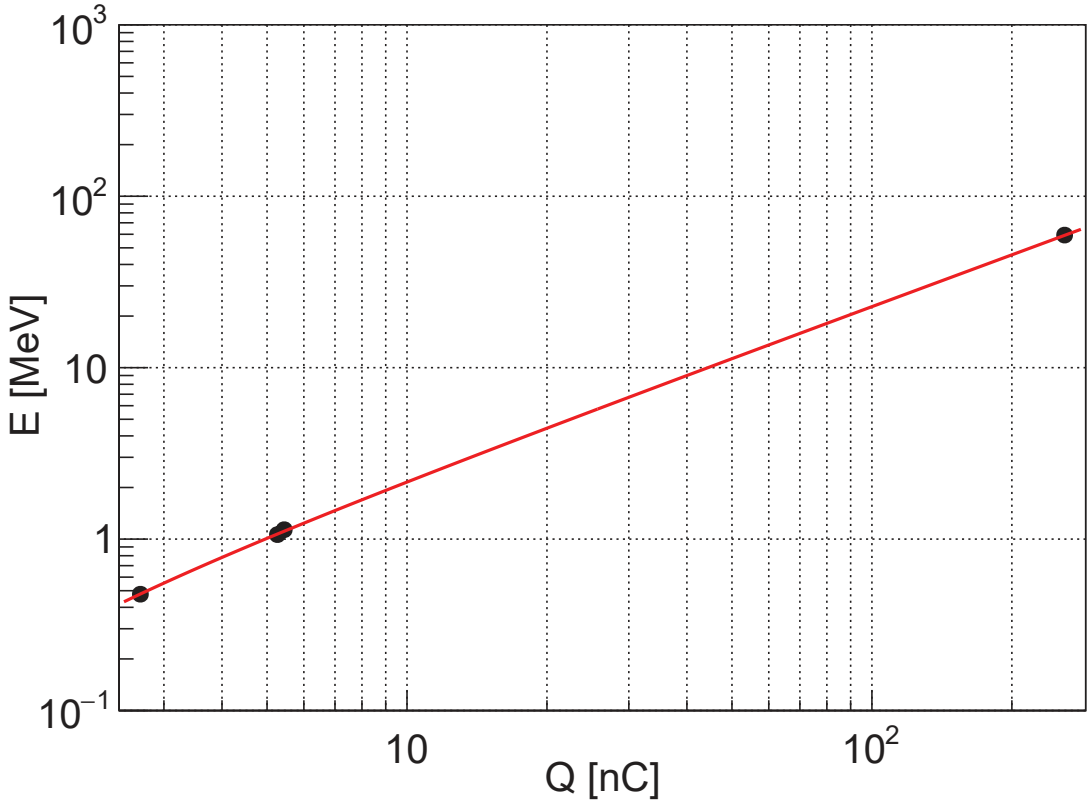

Figure 5.11: Relation between released energy in MeV and deposited charge in nC for the radioactive sources and the cosmic muons.

Chapter 6

Conclusions and future perspectives

Several studies have been performed in order to test and optimize the software developed in the first part of this thesis work for processing and efficiently elaborating the raw data acquired by the JUNO setup at LNL; we developed additional methods which have been successfully applied, and are still available for future acquisitions, for monitoring the quality of the acquired data over both short and long periods of time, taking into account some of their most important characteristics, such as the stability of the baseline contribution to the signals and the stability of the trigger rate; by the way, the stability of these quantities has been observed and proved more than once. In addition, these methods have been fundamental in spotting several bugs affecting the DAQ system and they have helped in ameliorating the features of the GCU firmware; moreover, they can always be used for online checks in further measurements.

The acquisitions we performed have allowed to test and characterize the response of the experimental apparatus using many different trigger validation conditions; the acquired data have been used in particular to define and apply a proper procedure for estimating the total charge released by particles in the liquid scintillator. As a consequence, this procedure has been successfully applied in the analysis of the spectrum of cosmic muons, detected using an external trigger provided by the use of three plastic scintillators. An important part of this thesis work was devoted in formulating a detailed interpretation of the experimental results on cosmic muons, by means of dedicated studies and comparisons with the theoretical expectation, which has been, by the way, well fulfilled.

Furthermore, we focused on some specific studies for providing a precise energy calibration of the setup, employing some radioactive sources characterised by simple and well-known decay schemes. Thanks to the support of the results of some preliminary Monte Carlo simulations of the setup response, we could set the proper trigger validation conditions for the acquisitions and, in particular, we discovered that, due to the small size of the LS vessel, we could not detect the full-energy peaks of the sources. However, we have been able to measure their Compton continuum, which happened to be in agreement with the expectation; extracting the Compton edge positions for the considered radioactive sources, and associating them to their expected values in units of keV, we could obtain a proper calibration of the experimental setup. Finally, calibrating in energy the experimental cosmic muon spectrum, we could compare it to the result of a Monte Carlo simulation, obtaining a very good agreement with the expectation.

For further improvements, a more precise calibration procedure could be taken into account, for example by employing a laser source characterised by a known wavelength, which would allow a precise characterization of the experimental response at different and well-known energy inputs. Moreover, some additional improvements of the raw data processing code could be thought and implemented, in order to optimize and reduce its computational times even with larger amount of data. Finally, we should expect in the near future the installation of the 9 missing channels, since during the thesis work only 13 out of the planned 16 GCUs were installed, which will clearly imply further tests and analyses of the whole setup response.

Appendix A

The gcu_raw structure and the gcu_reader class

As anticipated in Section 2.2.2, the program that has been developed for the processing of the raw data reads the data packets from the binary files and it stores all the interesting information, employing a user-developed C++ structure, called `gcu_raw` and reported in Figure A.1, opportunely written for containing the details extracted from each data packet.

```
struct gcu_raw {

 int gcu_id = 0;
 int ch_n = 0; //GCU channel number
 int seq_n = 0; //packet sequential number in the binary file
 int trg_n = 0; //cyclic trigger count of the GCU channel
 unsigned long time_cnt = 0; //timestamp
 int pkt_size = 0; //expected packet size (including header and trailer)
 int data_size = 0; //expected data size
 int actual_pkt_size = 0; //actual packet size (including header and trailer)
 bool valid_data = false;
 int data_vraw[data_size] = {0};
};
```

Figure A.1: The implemented, user-developed, C++ structure `gcu_raw`.

Some considerations on the attributes of the `gcu_raw` structure:

- the attribute `pkt_size` is the value of the packet size that is reported in the header, as described in Section 2.1, while `actual_pkt_size` is used to save the actual dimension of the data packet, calculated during the reading process; it is important to underline that they are not necessarily equal, there could be in fact undesired mismatches or bugs which could affect the actual size of a packet, causing consistency problems which can be spotted thanks to these variables;
- the boolean variable `valid_data` is used as a flag to identify the presence of corrupted data packets;
- the array `data_vraw` is used to collect the information contained in the waveform payload.

While the `gcu_raw` structure can be considered as a container for the information carried by a single data packet, the `gcu_reader` class actually provides the methods used for reading the binary file and extracting the corresponding `gcu_raw` structure. A simple Unified Modeling Language (UML) model of the `gcu_reader` class, comprehensive of all its private and public members, is reported in Figure A.2.

After the user sets the filename, the class calculates the total file dimension, which is then stored in the `dimension` variable, and, employing the Boyer-Moore algorithm, looks for the first header start and, if it is found, its position is saved in the `offset` variable; then, moving along the whole bit se-

quence of the binary file, by means of the `inF` pointer, the `get_event` method reads each data packet one after the other and it fills the corresponding `gcu_raw` structure, given in input to it. Finally, the two additional methods, `get_offset()` and `get_dimension()`, are used to get access respectively to the values of the `offset` and the `dimension` variables.

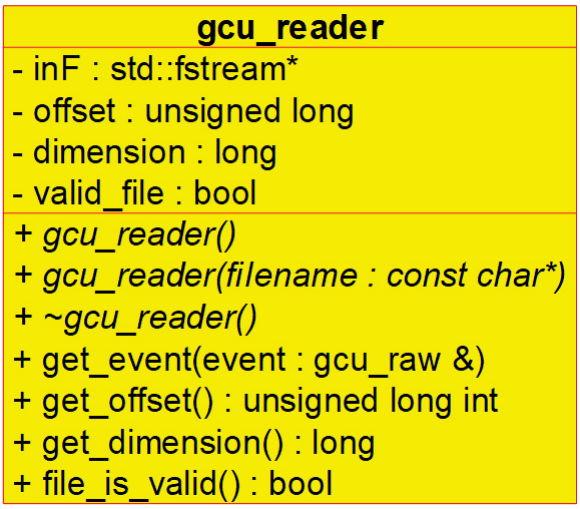

Figure A.2: UML model of the `gcu_reader` class.

Appendix B

Results of the single-photon measurements

This chapter reports the results that we obtained for the single-photon measurements for all the PMTs of the setup, as anticipated in Section 3.4; in particular Figure B.2 - B.6 contain the spectra with the characteristic noise and single-photon peaks, properly fitted with a double Gaussian function, as described in Section 3.4. The results of the fitting procedure, and the related derived quantities, such as the estimate of the resolution R, or the peak-to-valley ratio R_{pv}, are reported in Table B.1, Table B.2 and Table B.3. As it can be seen in the plots of the reconstructed charge spectra, not all the channels show a well-defined single-photon peak; this is probably due to the fact that not all the PMTs were correctly illuminated by the LED during the measurements.

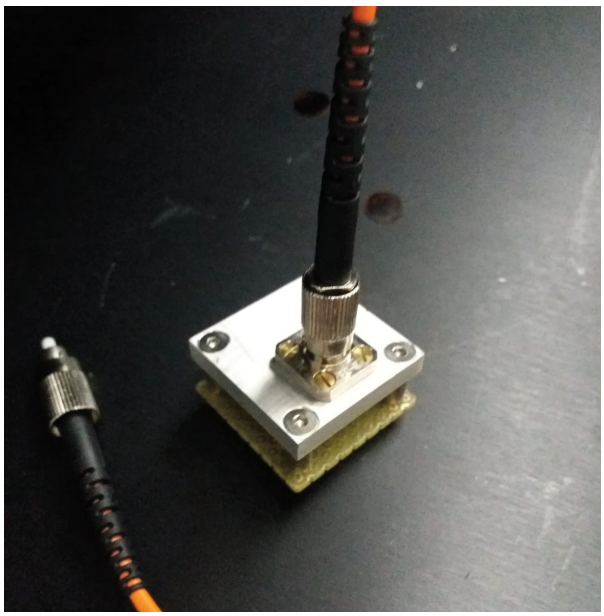

Figure B.1: Photo of the employed LED circuit for the single-photon measurements and the optic fiber used to connect it to the top of the setup.

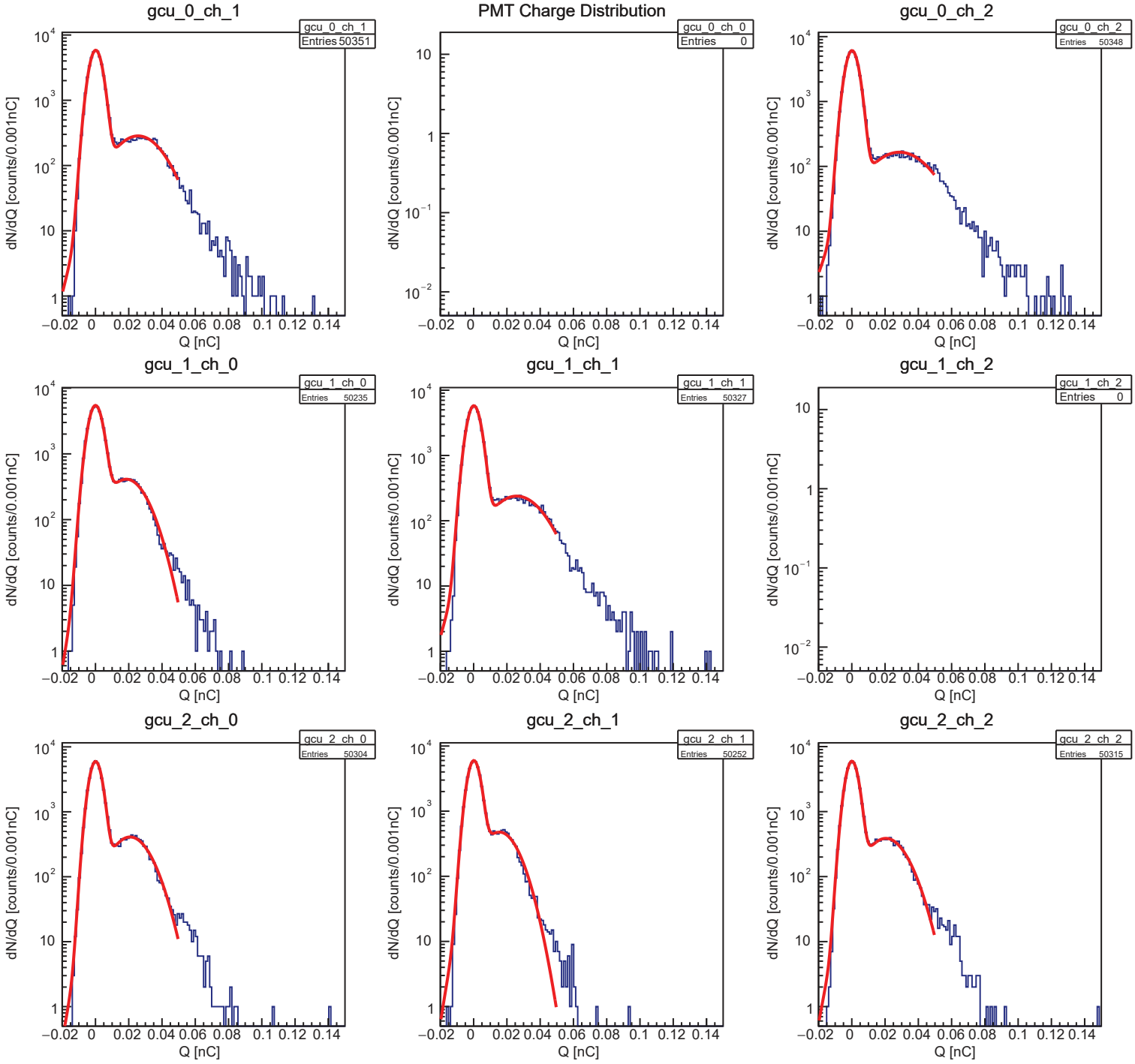

Figure B.2: Measured single-photon spectra - PAGE 1.
GCU0 CH0 and GCU1 CH2 did not acquired data.

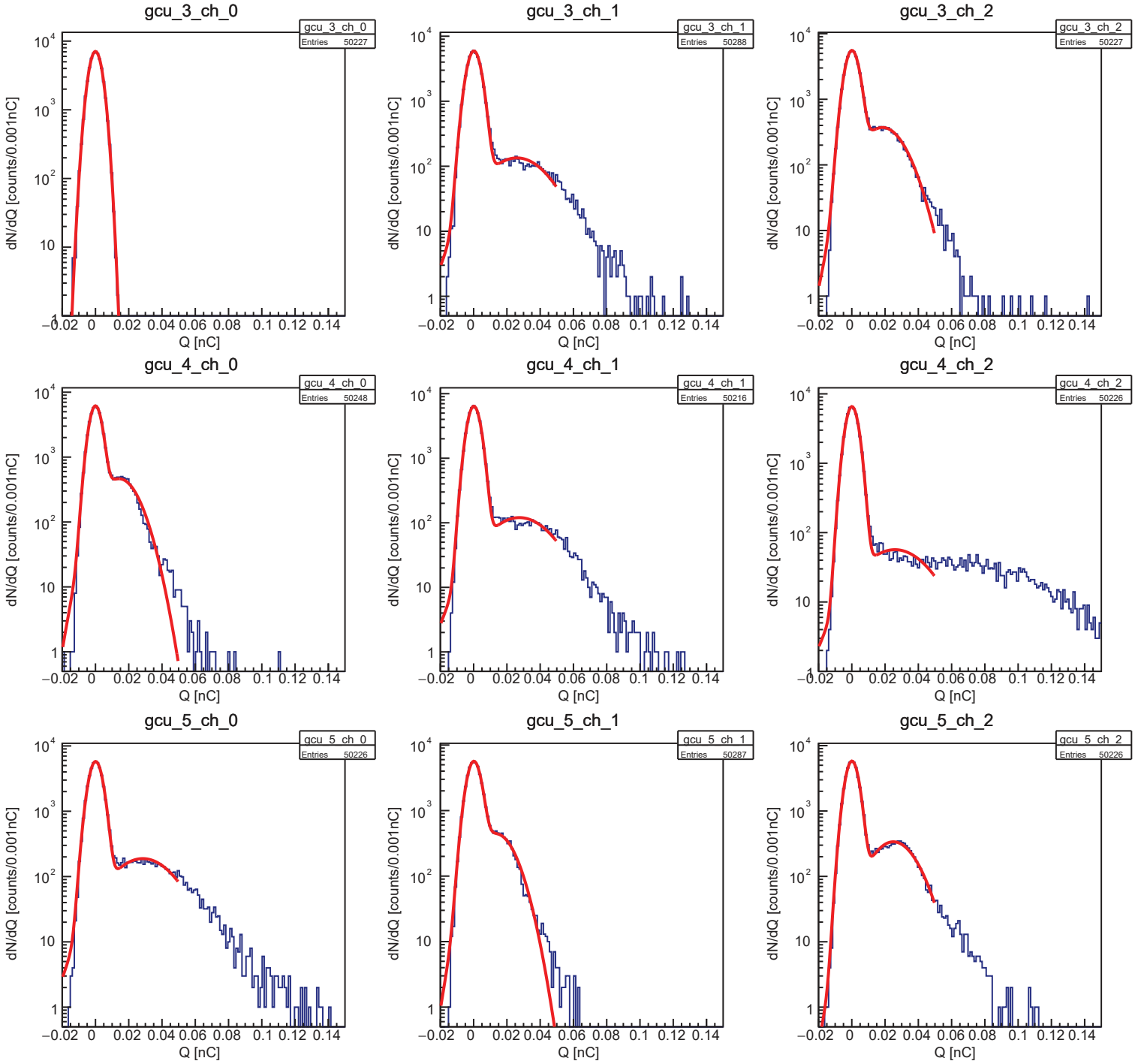

Figure B.3: Measured single-photon spectra - PAGE 2.

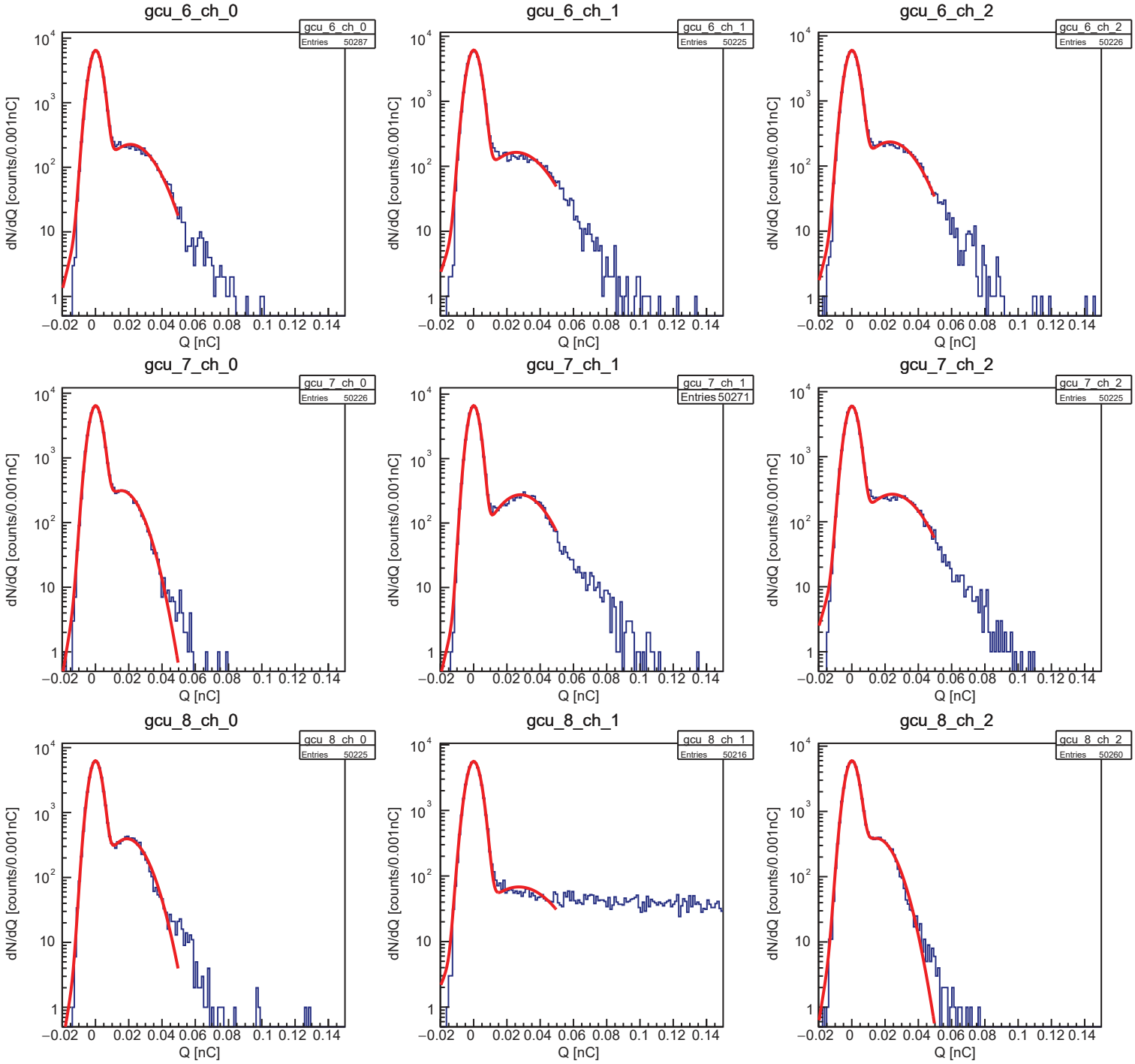

Figure B.4: Measured single-photon spectra - PAGE 3.

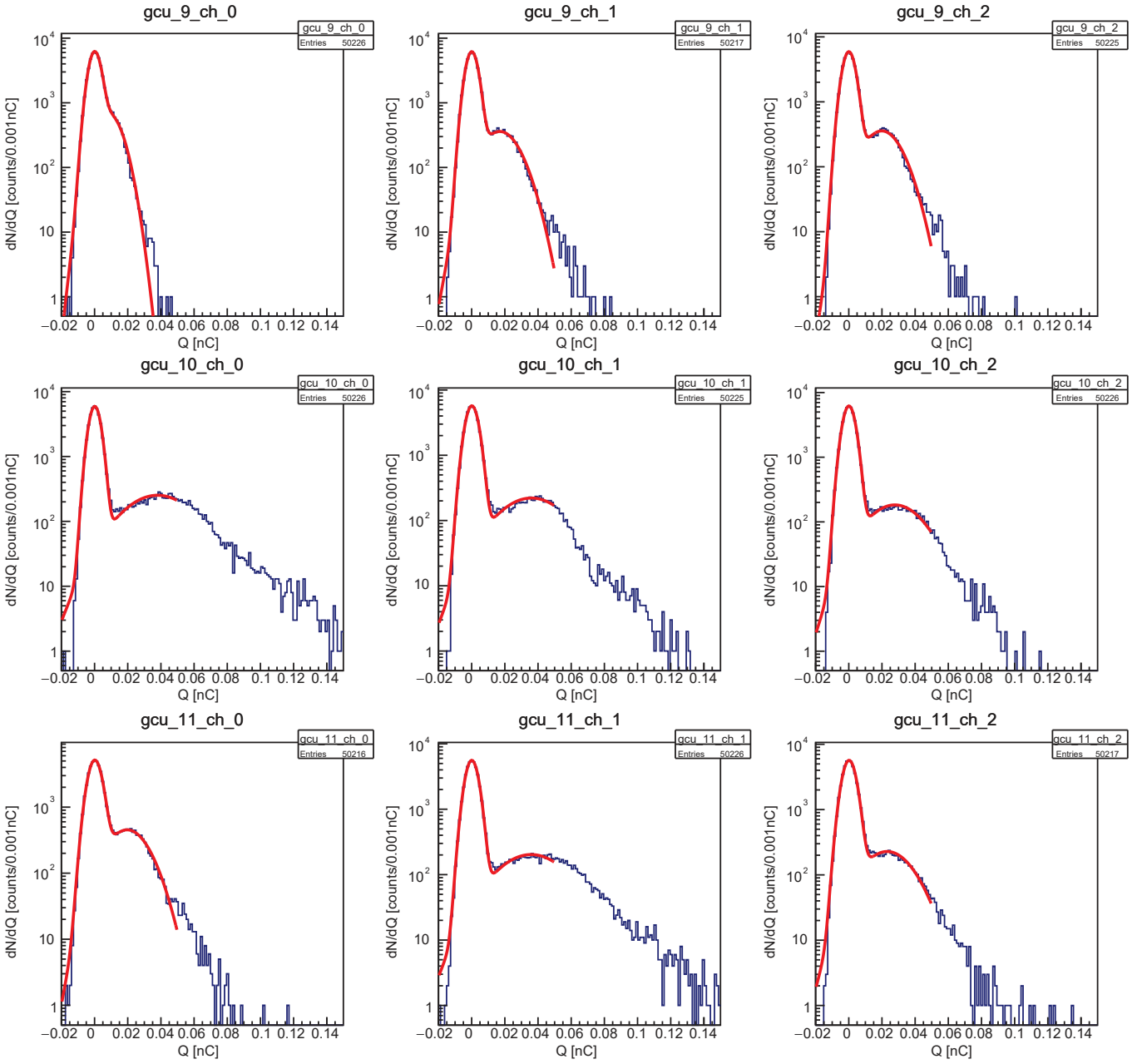

Figure B.5: Measured single-photon spectra - PAGE 4.

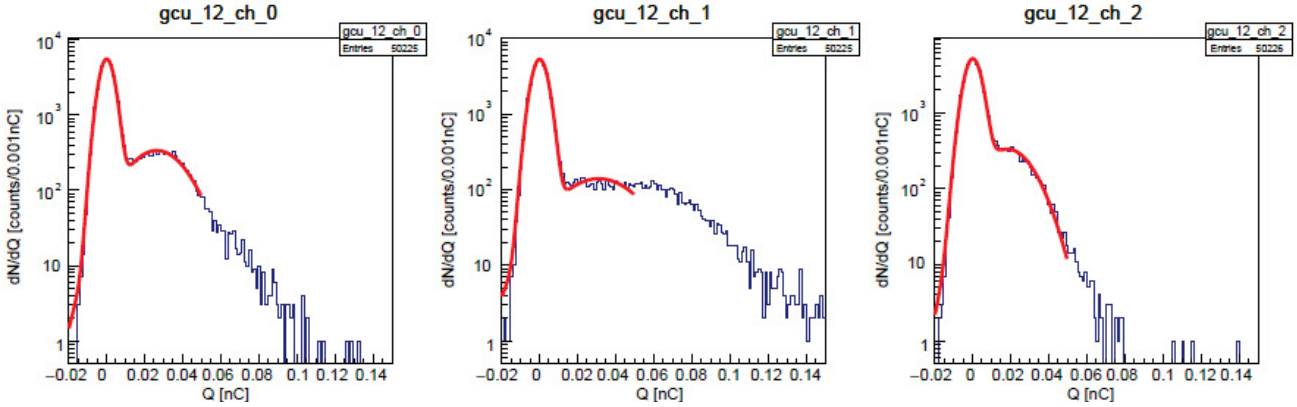

Figure B.6: Measured single-photon spectra - PAGE 5.

GCU	CH	C_n [counts]	C_n err [counts]	μ_n [nC]	μ_n err [nC]	σ_n [nC]	σ_n err [nC]
0	0	/	/	/	/	/	/
0	1	5799	35	0.00014	1e-05	0.00352	1e-05
0	2	6056	35	0.00012	1e-05	0.00355	1e-05
1	0	5372	33	-9e-06	2e-05	0.00371	1e-05
1	1	5737	34	0.00012	1e-05	0.00362	1e-05
1	2	/	/	/	/	/	/
2	0	5905	36	5e-05	1e-05	0.00336	1e-05
2	1	5857	36	9e-05	1e-05	0.00336	1e-05
2	2	5889	36	5e-05	1e-05	0.00338	1e-05
3	0	6207	35	-0.0005	0.0006	0.0032	0.0001
3	1	5894	34	0.00012	1e-05	0.00373	1e-05
3	2	5459	34	6e-05	2e-05	0.00366	1e-05
4	0	6008	38	-3e-05	1e-05	0.00328	1e-05
4	1	6365	36	0.00010	1e-05	0.00347	1e-05
4	2	6540	37	7e-05	1e-05	0.00343	1e-05
5	0	5743	33	7e-05	1e-05	0.00363	1e-05
5	1	5506	34	8e-05	2e-05	0.00365	1e-05
5	2	5773	35	7e-05	1e-05	0.00349	1e-05
6	0	6414	38	0.00014	1e-05	0.00334	1e-05
6	1	6101	35	9e-05	1e-05	0.00355	1e-05
6	2	5937	35	9e-05	1e-05	0.00354	1e-05
7	0	6342	38	7e-05	1e-05	0.00335	1e-05
7	1	6612	40	6e-05	1e-05	0.00311	1e-05
7	2	5933	35	8e-05	1e-05	0.00344	1e-05
8	0	6220	38	2e-06	1e-05	0.00326	1e-05
8	1	5580	33	3e-05	1e-05	0.00374	1e-05
8	2	5867	36	4e-05	1e-05	0.00351	1e-05
9	0	5892	40	-4e-06	2e-05	0.00332	1e-05
9	1	6077	37	1e-05	1e-05	0.00337	1e-05
9	2	5860	35	3e-05	1e-05	0.00351	1e-05
10	0	5894	36	0.00010	1e-05	0.00320	1e-05
10	1	5692	34	9e-05	1e-05	0.00357	1e-05
10	2	6168	36	7e-05	1e-05	0.00346	1e-05
11	0	5064	32	0.00011	2e-05	0.00376	1e-05
11	1	5559	34	4e-05	1e-05	0.00355	1e-05
11	2	5626	33	0.00016	1e-05	0.00374	1e-05
12	0	5467	34	0.00012	1e-05	0.00356	1e-05
12	1	5263	31	0.00012	1e-05	0.00389	1e-05
12	2	4986	31	0.00015	2e-05	0.00407	1e-05

Table B.1: Electronic noise peak parameters extracted with the double Gaussian fit on the single-photon spectra, for all the setup channels.

GCU	CH	C_{sp} [counts]	C_{sp} err [counts]	μ_{sp} [nC]	μ_{sp} err [nC]	σ_{sp} [nC]	σ_{sp} err [nC]
0	0	/	/	/	/	/	/
0	1	2847	4	0.0256	0.0002	0.0136	0.0002
0	2	165	2	0.0286	0.0003	0.0165	0.0003
1	0	411	5	0.0185	0.0002	0.0105	0.0001
1	1	238	3	0.0258	0.0002	0.0145	0.0002
1	2	/	/	/	/	/	/
2	0	409	5	0.0208	0.0002	0.0107	0.0001
2	1	476	6	0.0155	0.0001	0.0096	0.0001
2	2	387	5	0.0204	0.0002	0.0111	0.0001
3	0	1379	1847	0.003	0.001	0.0029	0.0002
3	1	134	2	0.0259	0.0004	0.0165	0.0003
3	2	371	5	0.0184	0.0002	0.0114	0.0001
4	0	463	6	0.0143	0.0001	0.0098	0.0001
4	1	120	2	0.0274	0.0004	0.0171	0.0003
4	2	56	1	0.0259	0.0007	0.0179	0.0005
5	0	188	3	0.0284	0.0003	0.0166	0.0003
5	1	440	7	0.0134	0.0002	0.0095	0.0001
5	2	337	4	0.0249	0.0002	0.0119	0.0001
6	0	225	3	0.0209	0.0002	0.0127	0.0001
6	1	163	3	0.0255	0.0003	0.0155	0.0002
6	2	234	3	0.0229	0.0002	0.0136	0.0002
7	0	311	5	0.0155	0.0002	0.0097	0.0001
7	1	274	4	0.0281	0.0002	0.0133	0.0002
7	2	264	4	0.0243	0.0002	0.0144	0.0002
8	0	394	5	0.0189	0.0002	0.0101	0.0001
8	1	68	1	0.0271	0.0006	0.0178	0.0005
8	2	385	6	0.0148	0.0002	0.0096	0.0001
9	0	668	16	0.0084	0.0001	0.00708	7e-05
9	1	360	5	0.0170	0.0002	0.0104	0.0001
9	2	356	5	0.0198	0.0002	0.0104	0.0001
10	0	252	3	0.0379	0.0005	0.0193	0.0004
10	1	220	3	0.0360	0.0005	0.0187	0.0004
10	2	182	3	0.0277	0.0003	0.0157	0.0003
11	0	453	5	0.0196	0.0002	0.0113	0.0001
11	1	202	3	0.0358	0.0005	0.0190	0.0004
11	2	227	3	0.0231	0.0002	0.0138	0.0002
12	0	336	4	0.0264	0.0002	0.0139	0.0002
12	1	139	2	0.0311	0.0005	0.0189	0.0004
12	2	328	5	0.0185	0.0002	0.0120	0.0001

Table B.2: Single-photon peak parameters extracted with the double Gaussian fit on the single-photon spectra, for all the setup channels.

GCU	CH	FWHM$_n$ [nC]	FWHM$_n$ err [nC]	R	R err	R$_{pv}$
0	0	/	/	/	/	/
0	1	0.00829	3e-05	0.536	0.009	1.48282
0	2	0.00836	3e-05	0.58	0.01	1.45294
1	0	0.00873	3e-05	0.570	0.009	1.1095
1	1	0.00854	3e-05	0.56	0.01	1.38678
1	2	/	/	/	/	/
2	0	0.00793	3e-05	0.515	0.008	1.37296
2	1	0.00792	3e-05	0.625	0.008	1
2	2	0.00797	3e-05	0.547	0.008	1.28954
3	0	0.0077	0.0003	0.9	0.4	1
3	1	0.00879	3e-05	0.64	0.01	1.23082
3	2	0.00863	3e-05	0.620	0.009	1.07949
4	0	0.00774	3e-05	0.685	0.008	1
4	1	0.00818	3e-05	0.62	0.01	1.33439
4	2	0.00808	2e-05	0.69	0.02	1.20051
5	0	0.00856	3e-05	0.58	0.01	1.42407
5	1	0.00861	4e-05	0.71	0.01	1
5	2	0.00823	3e-05	0.478	0.008	1.65734
6	0	0.00787	3e-05	0.60	0.01	1.20582
6	1	0.00837	3e-05	0.61	0.01	1.29688
6	2	0.00834	3e-05	0.59	0.01	1.25027
7	0	0.00790	3e-05	0.63	0.01	1
7	1	0.00733	2e-05	0.475	0.009	2.05774
7	2	0.00811	3e-05	0.594	0.009	1.33711
8	0	0.00767	3e-05	0.535	0.008	1.25871
8	1	0.00882	3e-05	0.65	0.02	1.21901
8	2	0.00827	3e-05	0.65	0.01	1
9	0	0.00783	4e-05	0.84	0.01	1
9	1	0.00795	3e-05	0.615	0.009	1.04182
9	2	0.00826	3e-05	0.524	0.009	1.25067
10	0	0.00754	3e-05	0.51	0.01	2.32364
10	1	0.00841	3e-05	0.52	0.01	1.96809
10	2	0.00816	3e-05	0.56	0.01	1.4792
11	0	0.00886	4e-05	0.580	0.008	1.14995
11	1	0.00836	3e-05	0.53	0.01	1.91698
11	2	0.00882	3e-05	0.60	0.01	1.20705
12	0	0.00838	3e-05	0.531	0.009	1.54439
12	1	0.00916	3e-05	0.61	0.01	1.38506
12	2	0.00959	4e-05	0.65	0.01	1.01802

Table B.3: Further parameters for the single-photon and electronic noise measurements, for all the setup channels.

Appendix C

Main updates of the raw data processing code

As anticipated in Chapter 5, an important upgrade of the raw data processing code has been implemented, in order to ameliorate and speed up the elaboration of the acquired data; this upgrade was necessary due to the higher trigger rates characterising the new acquisitions, which led to a very large amount of stored information even in a very limited acquisition time. For this reason, the original processing code required increasingly longer times for elaborating and analysing data from all the active channels.

Despite many analogies with the original software, the main characteristic of the new code consisted in the fact that it produced in output a single ROOT file, where all the waveforms describing the same event were already time correlated during the processing phase. With the original script, as many ROOT files as the number of active channels were produced and a subsequent elaboration was needed for correlating in time all the acquired waveform; clearly, in this condition, when the number of data to be processed was excessively large, the required elaboration time became unsustainable.

An additional feature of the new code was the possibility of retrieving the multiplicity of a specific event directly during the processing phase, allowing a faster discrimination of undesired events detected during the acquisition.

In order to implement these new features, the main modifications were applied to the user-developed `gcu_reader` class, whose new UML model is reported in Figure C.1, and to the related `gcu_raw` structure.

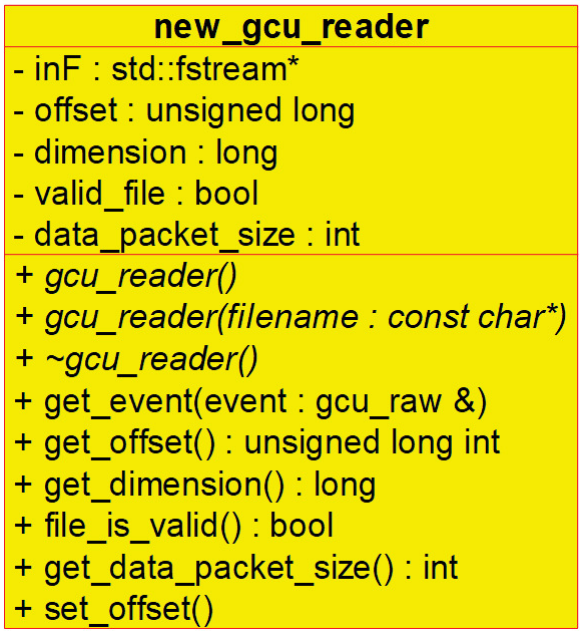

Figure C.1: UML model of the new `gcu_reader` class.

In the UML model reported in Figure C.1, we can notice the presence of a few new attributes and methods with respect to the original `gcu_reader` class:

- the attribute `data_packet_size`, and the related `get_data_packet_size()`, which are used respectively for storing and accessing the dimension of the raw data packets;
- the `set_offset()` method, which allows to vary the reading position in the binary files.

All these new attributes are fundamental for retrieving the time correlated waveforms acquired by the different active channels, since they allow to move back and forth in the long bit sequences contained in the binary files; in fact, as we anticipated, the main feature of the new code consists in producing in output a single ROOT file where all the waveforms describing the same event are already time correlated during the processing phase, allowing a faster, further elaboration.

However, the implementation of this new feature was possible thanks to the important changes introduced in the `gcu_raw` structure, substituted with the user-developed C++ structures `event_structure` and `event_structure_wf`, that are reported in Figure C.2.

```
struct event_structure{

    double baseline[number_of_channels] = {0.};
    double baseline_sigma[number_of_channels] = {0.};
    double charge[number_of_channels] = {0.};
    double total_charge = 0.;
    int active_channels_n = 0;
    int multiplicity = 0;
    int gcu_id[number_of_channels] = {0};
    int ch_n[number_of_channels] = {0};
    int trg_n[number_of_channels] = {0};
    int actual_pkt_size[number_of_channels] = {0};
    int valid_data[number_of_channels] = {0};
    int threshold_flag[number_of_channels] = {0};
    unsigned long time_cnt[number_of_channels] = {0};
};

struct event_structure_wf{

    int gcu_id_wf[number_of_channels] = {0};
    int ch_n_wf[number_of_channels] = {0};
    int trg_n_wf[number_of_channels] = {0};
    int valid_data_wf[number_of_channels] = {0};
    int waveforms[number_of_channels][wf_window_stop-wf_window_start];
    unsigned long time_cnt_wf[number_of_channels] = {0};
};
```

Figure C.2: The user-developed C++ structures `event_structure` and `event_structure_wf`.

As we can notice in Figure C.2, in order to be able to store all the information related to a single event, the simple variables previously used in the `gcu_raw` structure for saving the information regarding the single acquired waveform are now substituted by arrays, whose dimension depends on the number of active channels. Some additional considerations on the attributes of `event_structure`:

- the `total_charge` attribute is used to store the value of the total integrated charge, calculated summing the charge values extracted from each active channel, which are however saved in the `charge` array;

- the `active_channels` attribute contains the information on how many channels have been retrieved at a fixed timestamp and it is used to check if there is lack of information at high acquisition rates;
- the `multiplicity` attribute stores the number of channels which have a signal exceeding a threshold of $6\sigma_{\mathrm{B}}$ above the baseline and it is calculated thanks to the boolean variables stored in the `threshold_flag` array.

Finally, an additional structure, called `event_structure_wf`, has been implemented mainly for saving the waveforms themselves, which can be used in particular for debugging purpose; the stored waveform window is not necessarily as wide as the one retrieved from the binary files, but it can be properly adjusted using the variables `wf_window_start` and `wf_window_stop` in order to select only the most interesting range and reducing the amount of stored information. The use of a separated structure only for the waveforms is due to the fact that this allows an easier handling of data which can produce an important reduction in the time needed for their elaboration.

Bibliography

[1] F.P. An et al., *Measurement of electron antineutrino oscillation based on 1230 days of operation of the Daya Bay experiment*, Phys. Rev. D, **95**, 072006 (2017).

[2] S. H. Seo et al., *Spectral Measurement of the Electron Antineutrino Oscillation Amplitude and Frequency using 500 Live Days of RENO Data*, Phys. Rev. D **98**, 012002 (2018).

[3] Y. Abe at al., *Measurement of θ_{13} in Double Chooz using neutron captures on hydrogen with novel background rejection techniques*, JHEP01 **163** (2016).

[4] F. Capozzi et al., *Global constraints on absolute neutrino masses and their ordering*, Phys. Rev. D **95** (2017).

[5] P. Ghoshal and S.T. Petcov, *Neutrino mass hierarchy determination using reactor antineutrinos*, JHEP 1103:058, 2011, arXiv:1011.1646.

[6] F.P. An et al., *Neutrino Physics with JUNO*, J. Phys. G43 (2016) 030401, arXiv:1507.05613.

[7] Qian, X., and P. Vogel, *Neutrino Mass Hierarchy*, Progress in Particle and Nuclear Physics **83** (2015): 1–30.

[8] A. Garfagnini, *Precision Oscillation Measurements with Next Generation Reactor Neutrinos*, Proceedings of Science **023** (2017). DOI: 10.22323/1.283.0023.

[9] A. Giaz, *Status and perspectives of the JUNO experiment*, arXiv:1804.03575 [physics.ins-det] (2018).

[10] F. Marini, *Design and tests of the FPGA embedded trigger algorithms for the large PMT JUNO Electronics*, Master Degree Thesis, Dipartimento di Fisica e Astronomia, Università degli Studi di Padova (2018).

[11] S. Weinberg., Phys. Rev. Lett. **19**, 1264 (1967).

[12] R. Jr Davis, D.S. Harmer and K.C. Hoffman, Phys. Rev. Lett. **20**, 1205 (1968).

[13] J.N. Bahcall, N.A. Bahcall and G. Shaviv, Phys. Rev. Lett. **20**, 1209 (1968).

[14] J.N. Bahcall, W.F. Huebner, S.H. Lubow, P.D. Parker and R.K. Ulrich, Rev. Mod. Phys. **54**, 767 (1982).

[15] M.C. Gonzalez-Garcia, Michele Maltoni, Jordi Salvado, and Thomas Schwetz, *Global fit to three neutrino mixing: critical look at present precision*, JHEP, 1212:123, 2012.

[16] L. Wolfenstein, *Neutrino oscillations in matter*, Phys. Rev., D17:2369, 1978.

[17] S.P. Mikheev and A.Yu. Smirnov, *Resonant amplification of neutrino oscillations in matter and solar neutrino spectroscopy*, NuovoCim., C9:17–26, 1986.

[18] S.P. Mikheev and A.Yu. Smirnov, *Resonance enhancement of oscillations in matter and solar neutrino spectroscopy*, Sov. J. Nucl. Phys., 42:913–917, 1985.

[19] Steven Lau, *CPR1000 Design, Safety Performance and Operability*, Daya Bay Nuclear Power Operations and Management Company, 2011.

[20] Q. Liu et al., Nucl. Instrum. Meth. A 795, 2015, 284, arXiv:1504.01001.

[21] W. Konetschny and W. Kummer, Phys. Lett. B 70, 1977, 433.

[22] Xiang-Cui Lei et al., *Evaluation of new large area PMT with high quantum efficiency*, Chinese Physics C 40 (2016), 26002.

[23] Pedretti D. et al. (2018) *The Global Control Unit for the JUNO Front-End Electronics.* In: Liu ZA. (eds) Proceedings of International Conference on Technology and Instrumentation in Particle Physics 2017. TIPP 2017. Springer Proceedings in Physics, vol 212. Springer, Singapore. https://doi.org/10.1007/978-981-13-1313-4_37

[24] R. Brun, F. Rademakers, *ROOT: an object oriented data analysis framework*, Nucl. Instr. Meth. A, **389**, 81-86 (1997).

[25] R.S. Boyer, J S. Moore, *A fast string searching algorithm*, Communication of the ACM, **20**, (1977).

[26] C++ `std::search` reference: https://en.cppreference.com/w/cpp/algorithm/search

[27] C++ `std::boyer_moore_searcher` reference: https://en.cppreference.com/w/cpp/utility/functional/
/boyer_moore_searcher

[28] https://indico.cern.ch/event/299180/contributions/1659676/attachments/563129/775787
/IPbus_TWEPP_20140924_v1.pdf

[29] PDG reference for integral intensity of vertical muons I at sea level: https://pdg.lbl.gov/2011
/reviews/rpp2011-rev-cosmic-rays.pdf

[30] L. Landau, *On the energy loss of fast particles by ionization*, J. Phys. (USSR), 8: 201, (1944).

[31] ROOT reference for Landau function: https://root.cern.ch/doc/master/
/namespaceTMath.html#a656690875991a17d35e8a514f37f35d9

[32] K.S. Kölbig and B. Schorr, *A program package for the Landau distribution*, Computer Phys. Comm. 31, 97-111 (1984).

[33] J.E. Gaiser, *Charmonium Spectroscopy from Radiative Decays of the J/Psi and Psi-Prime*, Appendix-F, Ph.D. Thesis, SLAC-R-255 (1982).

[34] J. Cheng, Z. Wang, L. Lebanowski et al., *Determination of the total absorption peak in an electromagnetic calorimeter*, Nucl. Instrum. Methods Phys. Res. Sect. A 827, 165-170 (2016).

[35] ROOT reference for Crystal Ball function: https://root.cern/doc/master/CrystalBall_8C_source.html

[36] H. Bethe, J. Ashkin, *Experimental Nuclear Physics*, ed. E.Segré, J.Wiley, New York, p.253 (1953).

[37] M. Tanabashi et al. (Particle Data Group), Phys. Rev. D **98**, 030001 (2018), p. 424-432. DOI: 10.1103/PhysRevD.98.030001

[38] T.K. Gaisser, *Spectrum of cosmic-ray nucleons, kaon production, and the atmospheric muon charge ratio*, Astroparticle Physics **35**, 801 (2012). DOI: 10.1016/j.astropartphys.2012.02.010

[39] M. Boezio, *Measurement of the flux of atmospheric muons with the CAPRICE94 apparatus*, Physical Review D: Particles and fields **62**(3) (2000). DOI:10.1103/PhysRevD.62.032007

[40] A. Bettini, *Introduction to Elementary Particle Physics*, Cambridge University Press, 2008.

[41] PDG reference for mass stopping power: http://pdg.lbl.gov/2020/reviews/rpp2020-rev-passage-particles-matter.pdf

[42] A. Rogalski, K. Adamiec and J. Rutkowski, *Narrow-Gap Semiconductor Photodiodes*, SPIE Press, 2000.

[43] Knoll, G., *Radiation Detection and Measurement*, John Wiley and Sons, Inc., 2010.

[44] Recommended data of the Laboratoire National Henri Becquerel: http://www.nucleide.org/DDEP_WG/DDEPdata.htm.

[45] Yu.M. Malyshkin, A.N. Fazliakhmetov, A.M. Gangapshev et al., *Modeling of a MeV-scale particle detector based on organic liquid scintillator*, Nuclear Inst. and Methods in Physics Research, A **951** (2020), 162920. DOI: https://doi.org/10.1016/j.nima.2019.162920.

[46] Test bench for measurements of the NOνA scintillator properties: https://astronu.jinr.ru/wiki/index.php/Test_bench_for_measurements_of_the_NOvA_scintillator_properties