UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia "Galileo Galilei"

Master's Degree in Physics

Final Dissertation

First analysis of LEGEND-200 commissioning data: searching for neutrinoless double beta decay in Germanium-76

Thesis supervisor

Prof. Alberto Garfagnini

Thesis co-supervisor

Dr. Sofia Calgaro

Candidate

Giovanna Saleh

Academic Year 2022/2023

Abstract

Neutrinoless double beta decay ($0\nu\beta\beta$) is a hypothetical lepton-number-violating rare process which could take place only if neutrinos were Majorana fermions, namely if neutrinos were their own antiparticles. If observed, this decay would shed light on neutrinos' nature and would be an experimental evidence for lepton number violation, which is a Beyond Standard Model phenomenon. Moreover, from the study of this decay it would be possible to extract information concerning neutrino masses.

The LEGEND experiment (Large Enriched Germanium Experiment for Neutrinoless double beta Decay) searches for the $0\nu\beta\beta$ decay of ^{76}Ge employing active ^{76}Ge-enriched detectors. In its first phase, operated at Gran Sasso National Laboratories (LNGS) in Italy, 200 kg of Germanium are deployed, aiming at a half-life sensitivity of over 10^{27} years; the objective of the second phase is to operate 1000 kg of Germanium, to achieve a sensitivity beyond 10^{28} years.

This thesis work is devoted to the study of the first LEGEND commissioning data, collected with 60 kg of Germanium, for an exposure of approximately 2.248 kg yr. The analysis focuses on the energy spectrum acquired by Germanium detectors: starting from raw data a series of analysis cuts is implemented, aimed at removing noise and background signals, so as to retain only the good $\beta\beta$ candidates. Finally, a Bayesian analysis of the distribution of the signals in the region of interest around $Q_{\beta\beta} = 2039$ keV is performed, in order to give an estimate of the background index $B.I.$ which quantifies the presence of residual background events, and eventually to put an upper bound on the strength of the $0\nu\beta\beta$ signal. The resulting parameters are $B.I. = 2.03^{+6.67}_{-1.38} \times 10^{-3}$ (keV kg yr)$^{-1}$ considering a 90% two-tailed C.I. and $T^{0\nu}_{1/2} > 3.56 \times 10^{24}$ yr at 90% C.I.

Contents

Introduction

Many reasons can be adduced to justify a study addressing neutrinoless double beta decay ($0\nu\beta\beta$). First of all, an eventual observation of $0\nu\beta\beta$ would unambiguously solve the open issue of neutrinos' nature, since it would be the decisive proof that neutrinos are Majorana fermions, namely that neutrinos are their own antiparticle. This aspect already locates $0\nu\beta\beta$ beyond the Standard Model of particle physics (BSM). Another feature that makes it a fully BSM process is that it violates the lepton number conservation by two units, meaning that it would not be a fundamental symmetry of Nature. The occurrence of $0\nu\beta\beta$ would have also remarkable cosmological implications: it would pave the way to the explanation of the matter-antimatter asymmetry encountered in the Universe, via leptogenesis mechanism. Finally, from the measurement of the half life of $0\nu\beta\beta$, some information about the neutrino masses and their mass hierarchy can be gained.

Among the experiments searching for $0\nu\beta\beta$ there is LEGEND (Large Enriched Germanium Experiment for Neutrinoless double beta Decay), built combining the best technologies from GERDA and MAJORANA Demonstrator experiments. LEGEND started acquiring its first commissioning data in Fall 2022: it is therefore utmost important to perform immediately a preliminary analysis of these commissioning data. The rationale is two-fold: on the one hand it allows to check the performances of the experiment and of the acquisition system, and to verify the data quality; on the other, it is allows to identify and validate a suitable analysis routine for the search of $0\nu\beta\beta$ signals, replicable on the final dataset at the end of the full experimental campaign.

The goal of this thesis work is to perform this first analysis of the physics data collected during the LEGEND commissioning phase and eventually to give a preliminary estimate of the achieved background index ($B.I.$).

This thesis work is structured as follows:

- **Chapter 1** provides a theoretical overview of the physical case: some elements of neutrino physics are introduced and in particular the phenomenon of interest, neutrinoless double beta decay ($0\nu\beta\beta$), is discussed.
- **Chapter 2** is dedicated to the description of the LEGEND experiment.
- **Chapter 3** explains the structure of the data taking and introduces the analyzed dataset, presenting its main features.
- **Chapter 4** provides an explanation of the data production flow, namely of how the raw acquired data are processed in order to produce the samples employed in the analysis. Crucial steps in this signal processing are the reconstruction of the energy of the acquired signals and its calibration.
- **Chapter 5** contains the main analysis, aimed at removing from the energy spectrum collected by Germanium detectors the noise and the non-physical signals, but also the signals coming from physical processes which constitute background with respect to the searched $0\nu\beta\beta$ decay. This purpose is achieved implementing a series of analysis cuts on the dataset:

– **Pulser signal removal** (Section 5.1)

The pulser injects in the system artificial signals with a precise shape and frequency: having a known signal periodically appearing among the collected data allows to check the stability of the acquisition and of the detector's performances during the physics data taking. Pulser signals are removed from the energy spectrum since they are non-physical.

– **Quality Cuts** (Section 5.2)

The Quality Cuts aim to discriminate the physical signals form the noise on the basis of the shape of the waveforms. In fact, the waveforms of physical signals are expected to have a well defined shape, comprehending a flat baseline, a fast positive-sloped rising edge up to a maximum amplitude, and then a slow exponential fall; the waveforms of noise signals, instead, can have heterogeneous shapes due to oscillations, spikes, inverted polarity and saturation. In practice, the shape of the waveforms of the good signals is synthesized in a set of analytical conditions on the main parameters of the waveforms themselves. Each signal is examined individually: if it respects all the conditions, it is classified as a good, physical signal; otherwise it is classified as noise, and removed from the spectrum.

– **Muon veto** (Section 5.3)

LEGEND setup includes an active veto devoted to identify muons reaching the experimental site. Events tagged as *muon* are removed from the energy spectrum.

– **Multiplicity cut** (Section 5.4)

We define *multiplicity* ($\mathcal{M}$) of an event the number of Ge detectors acquiring a non-zero signal in coincidence. Therefore, studying the multiplicity of an event means studying the coincidences between multiple Germanium detectors. It is known from theory that the energy deposit in $\beta\beta$ is highly localized: all the available energy is deposited in a small volume, largely contained within the volume of a single detector. Consequently, good $\beta\beta$ candidates are only the events in which one and only one Ge detector acquires a non-zero signal ($\mathcal{M}1$); instead, all the other events can be classified as background for the sake of this analysis and removed from the spectrum.

– **Liquid Argon veto** (Section 5.5)

Germanium detectors need to be operated at cryogenic temperatures: for this reason they are immersed in a cryostat filled with Liquid Argon. This liquid Argon is instrumented to detect eventual energy releases, so as to act as an active veto, exploiting again the topology of $\beta\beta$ events. Events vetoed by the Liquid Argon are identified as background and removed from the energy spectrum.

– **Pulse Shape Discrimination** (Section 5.6)

This cut focuses on the shape of the waveforms to discriminate among good $\beta\beta$ candidates and background signals. The key idea is that the topology of the event influences the shape of the waveform of the acquired signal: this is exploited to isolate Single Site Events in the active volume of the detectors, which are the only good $\beta\beta$ candidates; all the other types of events are removed from the energy spectrum. The chosen PSD parameter is $A/E := \frac{\text{Current amplitude}}{\text{Energy}}$, which by construction is sensible to the shape of the waveform, since the current signal is the derivative of the charge signal (waveform).

In Section 5.7 the joint results of all the performed analysis cuts are presented.

- **Chapter 6** is dedicated to a Bayesian analysis of the background and of the eventual signal appearing in the $0\nu\beta\beta$ Region Of Interest (ROI) of the energy spectrum after all the analysis cuts. The goal of this analysis is to extract the background index ($B.I.$) in the ROI and eventually to put an upper limit on the strength of the $0\nu\beta\beta$ signal, which translates to a lower limit on the half life for this decay, $T^{0\nu}_{1/2}$.

Chapter 1

The physical case

1.1 Neutrinos in the Standard Model of Particle Physics

The existence of neutrinos was postulated by Pauli in 1930 to account for two features of the β decay: the continuous energy spectrum of the emitted electrons and the angular momentum violation in the process.

Pauli first presented his hypothesis in a letter [1] he sent from Zurich to the participants of a conference on radioactivity in Tubingen. These are Pauli's words, from the English translation published in [2]:

> *Dear Radioactive Ladies and Gentlemen,*
>
> *[...] I have hit upon a desperate remedy to save the [...] energy [conservation] theorem. Namely [there is] the possibility that there could exist in the nuclei electrically neutral particles that I wish to call neutrons, which have spin 1/2 and obey the exclusion principle, and additionally differ from light quanta in that they do not travel with the velocity of light: the mass of the neutron must be of the same order of magnitude as the electron mass and, in any case, not larger than 0.01 proton mass. The continuous β-spectrum would then become understandable by the assumption that in β decay a neutron is emitted together with the electron, in such a way that the sum of the energies of neutron and electron is constant.*

The actual *neutron* was discovered two years later by Chadwick [3, 4] and Pauli's particle was renamed *neutrino*. Of course now we know that its mass is much smaller than the electron mass, but apart from that the intuition was right: β decay is a three-body process and the energy available in the final state is shared among the electron and (anti)neutrino. The β^- decay then reads as follow:

$$n \to p + e^- + \bar{\nu}_e \tag{1.1}$$

More in general, when a nucleus having atomic number Z and mass number A undergoes a beta decay, a neutron from its nucleus is converted to a proton, and an electron and an electronic antineutrino are emitted. The process than reads as:

$$(A, Z) \to (A, Z+1) + e^- + \bar{\nu}_e \tag{1.2}$$

In 1934 the symmetric process, called β^+ decay, or positron emission, was observed [5]:

$$(A, Z) \to (A, Z-1) + e^+ + \nu_e \tag{1.3}$$

In the following we will refer to β decay implying that what we are writing is valid both for β^- and β^+ decay.

The first experimental observation of neutrinos took place more than twenty-five years after Pauli's statement: neutrinos were discovered in 1956 by Cowan and Reines in the Savannah River experiment [6] exploiting the $p + \bar{\nu}_e \to n + e^+$ reaction in Cadmium chloride enriched water.

The idea of the experiment is the following: assuming that the (anti)neutrino exists, it interacts with the proton via inverse beta decay and a neutron and a positron are produced. The positron annihilates with an electron of the medium and two γ rays are emitted. The neutron, instead, is captured by an absorber nucleus of the medium; the newly-formed nucleus will be in an excited state and another γ ray is emitted when it de-excites to ground state. The signature of this process is then the observation of delayed coincidences between the two γs from the electron-positron annihilation and the *gamma* from the de-excitation of the nucleus that captured the neutron. Of course, the fact that the inverse beta decay takes place implies the existence of the (anti)neutrino.

For the realization of this experiment Cowan and Reines employed reactor (anti)neutrinos impinging on a tank of water enriched in Cadmium chloride: water itself provides the target proton for the inverse beta decay to occur; the produced positron annihilates with an electron and the produced neutron is efficiently absorbed by cadmium nuclei. The γ ray emitted by cadmium is expected to be revealed $\approx 5\,\mu s$ after the the two γs emitted in the electron-positron annihilation.

In fact, Cowan and Reines observed these delayed coincidences: they were able to conclude that the inverse beta decay took place and therefore that the (anti)neutrinos exist.

Let us now come to the theoretical framework developed to take into account the existence and the interactions of this new particle (weak processes, beyond QED).

The first model developed to account for Pauli's three-body beta decay and for similar weak processes was proposed by Fermi in 1934 [7, 8]. Fermi's model consisted in a four fermion contact interaction, described by the lagrangian

$$\mathcal{L}_{\text{Fermi}}^{\text{WI}} = -\frac{G_F}{\sqrt{2}}(\overline{\psi}_p\,\gamma^\mu\,\psi_n)\,(\bar{\psi}_e\,\gamma_\mu\,\psi_{\nu_e}) \tag{1.4}$$

This model, though, will turn out to be unsuitable to account for two experimental evidences collected in the following years: in 1954 the experiment carried out by Wu showed that parity is not conserved in weak interactions [9]; in 1958 the helicity of neutrinos was measured [10] and the result was that only left-handed neutrinos ν_L couple to the weak interaction. In order to account for these phenomena, the chiral V-A theory was introduced. Also this model, though, soon proved to be insufficient to explain the observed phenomena, and it was overcome with the introduction of the Interacting Vector Boson theory [11].

The IVB theory is based on the assumption that massive vector bosons ($m > 0$, spin $= 1$) exist and are the mediator of weak interactions. The IVB lagrangian, then, has the following structure:

$$\mathcal{L}_{IVB} = \mathcal{L}_F + \mathcal{L}_B + \mathcal{L}_{INT} \tag{1.5}$$

in which $\mathcal{L}_F$ is the fermion lagrangian, accounting for the kinetic and mass term of the fermions, $\mathcal{L}_B$ is the boson lagrangian, accounting for the kinetic and mass term of the bosons and $\mathcal{L}_{INT}$ is the interaction lagrangian, accounting for the fermion-boson interaction. The IVB is still a chiral theory, therefore the left and right components of the particles appear in the lagrangian according to the experimental observations.

Coming to the Standard Model of particle physics (SM) notation and results, neutrinos are neutral leptons which can be classified in three flavor families, together with the charged lepton they are associated with:

$$\begin{pmatrix} e \\ \nu_e \end{pmatrix} \quad \begin{pmatrix} \mu \\ \nu_\mu \end{pmatrix} \quad \begin{pmatrix} \tau \\ \nu_\tau \end{pmatrix} \tag{1.6}$$

In particular, for all the flavor families (generic lepton l), having $SU(2) \times U(1)$ as group of symmetry for the SM, the left-handed chirality particles are doublets in $SU(2)$, while the right-handed chirality particles are singlets in $SU(2)$:

$$\begin{pmatrix} l \\ \nu_l \end{pmatrix}_L \quad ; \quad l_R, \ \cancel{\nu_R} \tag{1.7}$$

Since in the Standard Model lagrangian the mass term of each fermion (described by the spinor Ψ) reads

$$\mathcal{L}^{SM}_{\text{mass}} = m\,\overline{\Psi}\,\Psi = m(\overline{\Psi}_L\,\Psi_R + \overline{\Psi}_R\,\Psi_L) \tag{1.8}$$

namely it contains both the chirality states of the particle, and since no ν_R (or $\bar{\nu}_L$) is observed, the neutrino is forced and necessarily assumed to be massless in the Standard Model:

$$m_\nu^{SM} = 0 \tag{1.9}$$

1.2 Flavor oscillations and neutrino masses

Again, an experimental evidence was about to falsify the just stated assumption ($m_\nu = 0$) and to push towards a refinement of the theory: few years apart, around the beginning of the new millennium, two collaborations found an evidence for neutrino flavor oscillation, a phenomenon which can occur only if neutrino masses are not all zero.

1.2.1 Experimental observations

In 1998 the Super-Kamiokande collaboration, which studied atmospheric ν_μ, namely neutrinos produced as a consequence of the interaction of cosmic rays with the atmosphere and in particular in the kaons and pions decay and in the following muons decay, presented its results [12]: the flux of atmospheric muonic neutrinos detected at ground level was lower than expected, and in particular it exhibited a zenith angle dependent deficit. This observation was compatible with the hypothesis that neutrinos changed their flavor between their production and their detection. This phenomenon is referred to as *flavor oscillation*: neutrinos *oscillate* between different flavor states while propagating.

Another evidence of this phenomenon came few years later from the SNO (Sudbury Neutrino Observatory) collaboration [13, 14]. In particular, SNO's research managed to solve the so called solar neutrino problem: it was known that the Sun produces electronic neutrinos ν_e in its thermonuclear reactions; these neutrinos are expected to reach the Earth after propagating in space, and in fact they have been detected, but their flux detected on Earth was much lower than expected from the predictions of the Solar Standard Model. This result was obtained by several experiments, from the Sixties to the early 2000s: first by the Homestake experiment, and then also by Kamiokande, SAGE, GALLEX and Super-Kamiokande itself [15]. The SNO experiment managed to solve this problem, but only under the assumption that neutrinos flavor oscillation takes place. In particular, the SNO experiment was sensitive to two reactions, one involving only ν_e, the other involving all the three flavor families of neutrinos (ν_e, ν_μ and ν_τ). Being this the case, they were able to measure both the electronic neutrinos flux (Φ_{ν_e}) and the total neutrinos flux ($\Phi_\nu^{tot} = \Phi_{\nu_e} + \Phi_{\nu_\mu} + \Phi_{\nu_\tau}$). What they found out was that Φ_{ν_e} was, as in the previous experiments, much lower than the SSM prediction; instead, Φ_ν^{tot} was fully compatible with the theoretical predictions for Φ_{ν_e}. The interpretation they gave to this evidence was that part of the ν_e emitted by the Sun oscillate to a different flavor state (becoming

ν_μ, ν_τ) while propagating from the Sun to the Earth; therefore, the total neutrino flux is conserved, but the amount of ν_e arriving to Earth is only a fraction of the original flux. This interpretation is of course consistent with the result from Super-Kamiokande for atmospheric neutrinos.

Later, experiments on flavor oscillations were performed also employing neutrino beams produced in accelerator. An example is the T2K experiment [16], in which a pure beam of ν_μ was produced and sent towards two detectors: the initial properties (flux and composition) of the beam were measured in the *near detector*, located close to the production site; its final properties were instead measured in the *far detector*, located hundreds of kilometers away. The result was that a fraction of ν_e, not observed in the near detector, appeared instead in the far detector. Again, then, the conclusion was that neutrinos oscillate between flavor families while propagating from the production point to the detection point: flavor is not conserved. The $\nu_\mu \to \nu_\tau$ oscillation was finally observed by OPERA [17].

The theoretical implication of the observation of flavor oscillations is serious. In fact, as we will see the oscillation probability is non-zero only if the neutrino masses are non-zero, therefore the experimental observation of the oscillation is then a demonstration of the fact that neutrino masses are non-zero, contrarily to the SM assumption.

1.2.2 Theoretical interpretation

Neutrinos oscillations can be explained assuming that flavor states (ν_e, ν_μ, ν_τ) do not coincide with their mass eigenstates (ν_1, ν_2, ν_3). The flavor basis and the mass basis, though, are connected, and a generic flavor state ν_α, with $\alpha = e, \mu, \tau$ can be expressed as a linear combination of mass eigenstates ν_i, with $i = 1, 2, 3$:

$$|\nu_\alpha\rangle = \sum_i U_{\alpha i} |\nu_i\rangle \tag{1.10}$$

Assuming the existence of three flavor families and three mass eigenstates, the matrix U responsible for the basis transformation is the so called U_{PMNS} matrix, a 3x3 unitary matrix which took its name from the physicists who proposed this model (Pontecorvo-Maki-Nakagawa-Sakata) [18].

If neutrinos are Dirac particles, namely if they *do not* coincide with their own antiparticle, the U_{PMNS} matrix can be parametrized as follow, in terms of three mixing angles θ_{12}, θ_{23}, $\theta_{13} \in [0, \pi/2]$ and a Dirac phase $\delta \in [0, 2\pi]$, in complete analogy with the V_{CKM} mixing matrix of the quark sector:

$$U^D_{PMNS} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{1.11}$$

In which the following notation is used: $s_{ij} = \sin(\theta_{ij})$ and $c_{ij} = \cos(\theta_{ij})$. If neutrinos are Majorana particles, namely if they coincide with their own antiparticle (see Section 1.3), an additional term must be introduced in the U_{PMNS} matrix, containing two additional Majorana complex phases κ, λ:

$$U^M_{PMNS} = U^D_{PMNS} \begin{pmatrix} e^{i\kappa} & 0 & 0 \\ 0 & e^{i\lambda} & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{1.12}$$

In the following we will simply write U, implying that the Dirac or Majorana parametrization of the $PMNS$ matrix can be employed according to what is needed.

The time evolution of a flavor state can be obtained through the time evolution of the mass eigenstates, with $\hbar = 1$ and $c = 1$:

$$|\nu_\alpha(t)\rangle = \sum_i e^{-iE_i t} U_{\alpha i} |\nu_i(t=0)\rangle \tag{1.13}$$

Defining $W^{ij}_{\alpha\beta} = U^*_{\beta i} U_{\alpha i} U_{\beta j} U^*_{\alpha j}$, it can be shown that the oscillation probability from a flavor family α to a flavor family β is then given by:

$$P_{\alpha\beta} := P(\nu_\alpha \to \nu_\beta) = |\langle \nu_\beta(0) | \nu_\alpha(t) \rangle|^2 = \delta_{\alpha\beta} - 4 \sum_{i<j} \mathrm{Re}(W^{ij}_{\alpha\beta}) \sin^2\left(\frac{\Delta m^2_{ij} L}{4E}\right) \pm 2 \sum_{i<j} \mathrm{Im}(W^{ij}_{\alpha\beta}) \sin\left(\frac{\Delta m^2_{ij} L}{4E}\right) \quad (1.14)$$

Equation 1.14 is originally a function of time, since it contains a time evolution, but employing $c = 1$ and assuming $E \gg m_\nu$ (and then $v \approx c$), one obtains $t = L$, namely the dependence on the time of propagation t can be replaced with a dependence on the distance L covered in the propagation.

What is clear from the equation is that the neutrino oscillation probability is non-trivial only if $\Delta m_{ij} \neq 0$. In fact:

- If $\Delta m_{ij} = 0$, the two terms containing its sine are identically zero, and therefore the equations degenerate to the trivial case $P_{\alpha\beta} = \delta_{\alpha\beta}$, which is of course equal to 1 if $\alpha = \beta$ and equal to 0 if $\alpha \neq \beta$. In this case, then, no flavor oscillation are expected: the flavor is conserved during the propagation.
- If $\Delta m_{ij} \neq 0$, the two terms containing its sine are *not* identically zero. In this case, then, there is a non-zero probability of an oscillation between two different flavor states α and β.

Since as said there are several experimental evidences of neutrino flavor oscillations, we can conclude that neutrino $\Delta m_{ij} \neq 0$, namely that neutrino masses are not identically zero.

The observation of neutrino oscillations, then, gives direct evidence for the existence of new, Beyond Standard Model (BSM) physics, since this phenomenon cannot be explained within the SM. Despite this remarkable result, though, many features of neutrinos' nature and behavior are still unknown.

First, from the observation of oscillations we conclude that neutrino masses cannot be identically zero, but we have no clue about the absolute mass scale, or at least the mass ordering. From solar neutrinos studies it was possible to conclude that $m_2 > m_1$ [19], but up to now there is no information concerning m_3. Two scenarios are therefore still possible: the normal mass hierarchy, in which $m_3 > m_2 > m_1$, and the inverted hierarchy, in which $m_2 > m_1 > m_3$ (Fig. 1.1).

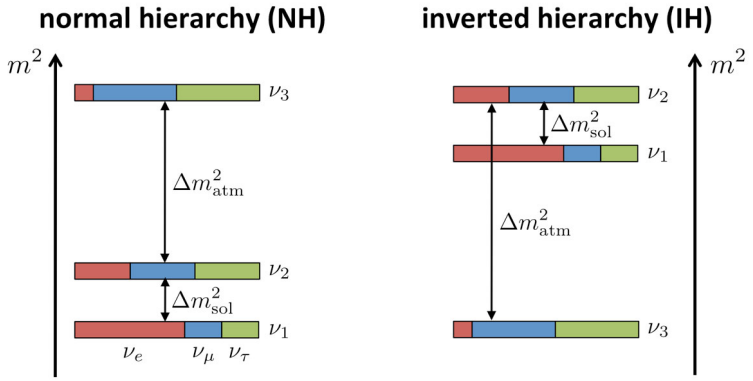


FIGURE 1.1: Schematic representation of the two possible mass orderings. In the so-called normal hierarchy $m_3 > m_2 > m_1$, while in the inverted hierarchy $m_2 > m_1 > m_3$. Which of the two is the actual neutrino mass hierarchy is still an open question. Picture from [20].

Several experiments aiming to determine the actual neutrino mass ordering are currently running or are under development, among these NOvA, JUNO, DUNE, Hyper-Kamiokande and KM3NeT [21].

Another important neutrino feature that is still unknown is whether neutrinos are Dirac or Majorana particles, namely whether they have distinct antiparticles or they are their own antiparticle. In order

to gain this information, a process that can occur only in one case is studied: neutrinoless double beta decay.

1.3 Double beta decay

As anticipated in Section 1.2, there are two competing hypothesis concerning the nature of neutrinos. They could be:

- **Dirac fermions** → neutrinos and antineutrinos are different particles;
- **Majorana fermions** → neutrinos are their own antiparticles.

The only known way to verify which of these hypothesis is correct is to study a process which could take place only in one of the two cases: a process of this type is neutrinoless double beta decay ($0\nu\beta\beta$), a hypothetical lepton number violating rare process which could take place only if neutrinos were Majorana particles. If observed, then, this phenomenon would be the clear proof that neutrinos are Majorana fermions.

Let us now briefly outline the theoretical background and the phenomenological features of (neutrinoless) double beta decay [22, 23, 24].

The possibility for a nucleus to undergo a double beta decay, namely to face the simultaneous emission of two electrons and two neutrinos, was first considered by Goeppert-Mayer in 1935 [25]. The considered process is

$$2\nu\beta\beta: \quad (A, Z) \rightarrow (A, Z+2) + 2e^- + 2\bar{\nu}_e \tag{1.15}$$

Again, we explicitly present the double β^- decay, but the same considerations can be done for the corresponding double β^+ decay, $(A, Z) \rightarrow (A, Z-2) + 2e^+ + 2\nu_e$. In the following we will talk about double β decay, referring to both of them, and we will call for brevity *electrons* and *neutrinos* the emitted leptons: what we imply is again *electrons and electronic antineutrinos* for $2\beta^-$ decay and *positrons and electronic neutrinos* for $2\beta^+$ decay.

The $2\nu\beta\beta$ is a second order weak process, therefore the probability of its occurrence is in principle *low* (half lives in Tab. 1.1). In general, this channel is in fact strongly suppressed, but there are some nuclei for which it is allowed: even-even nuclei in which the single β decay itself is strongly suppressed or energetically forbidden [26]. This situation occurs because even-even nuclei have lower mass than their odd-odd isobars [27], as shown in Fig. 1.2: the even-even parent nucleus is forced to decay double beta (two nucleons transformed) to an even-even daughter nucleus since the single beta decay (only one nucleon transformed) towards the closest odd-odd nucleus is not energetically possible.

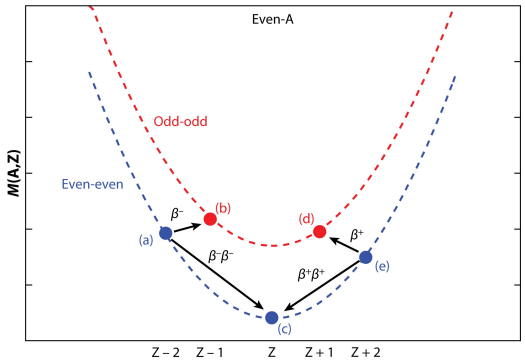


FIGURE 1.2: Sketch of the mass parabolas for even-A isobars: single β decay from a (e) to b (d) is not possible, so double β decay towards c becomes favorite. Picture from [26].

The Feynman diagram of $2\nu\beta\beta$ is shown in Fig. 1.3: two neutrons transform simultaneously in two protons, and a couple electron-neutrino is produced in each decay.

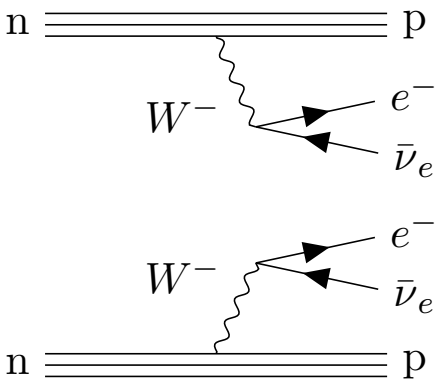


FIGURE 1.3: Feynman diagram of $2\nu\beta\beta$.

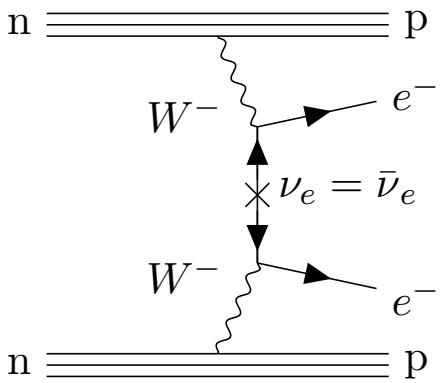


FIGURE 1.4: Feynman diagram of $0\nu\beta\beta$.

From the Feynman diagram of $2\nu\beta\beta$ (Fig. 1.3) it is clear that *if* the neutrino is its own antiparticle, then another process becomes possible: the neutrino*less* double beta decay $0\nu\beta\beta$ (Fig. 1.4). In fact, if the neutrino is its own antiparticle, the two electronic antineutrinos emitted in the standard (2ν) double beta decay may be regarded as a couple neutrino-antineutrino, and therefore annihilate. If this happens, no neutrino in the final state is found. The corresponding process would then be:

$$0\nu\beta\beta: \quad (A, Z) \rightarrow (A, Z + 2) + 2e^- \tag{1.16}$$

The $0\nu\beta\beta$ decay has first been theorized by Furry [28]: he was trying to interpret and explain the double beta decay in Majorana's theory framework, and the result was that the $\beta\beta$ decay was still possible, but with a different final state, in which only the electrons appeared, and no neutrinos.

A fundamental difference lies between $2\nu\beta\beta$ and $0\nu\beta\beta$, though. In $2\nu\beta\beta$ the lepton number is conserved, since in the initial state there are no leptons ($L = 0$) and in the final state there are two leptons and two antileptons ($L = +2 - 2 = 0$): this process is therefore allowed by the Standard Model. In $0\nu\beta\beta$, instead, the lepton number is *not* conserved, and in particular, it is violated by two units, since in the initial state there are no leptons ($L = 0$), while in the final state there are two ($L = 2$): this process is therefore *not* allowed by the Standard Model.

Up to now, the $2\nu\beta\beta$ has been observed in various isotopes: a list of the measured half lives can be found in [29]. In particular, the result for ^{76}Ge, obtained by the GERDA Collaboration, is $T^{2\nu}_{1/2} = (1.926 \pm 0.094) \cdot 10^{21}$ yr [30].

The $0\nu\beta\beta$, instead, has never been observed yet. Only a lower bound to its half life can then be put for now: the best result was obtained by the KamLAND-Zen Collaboration employing ^{136}Xe, $T^{0\nu}_{1/2} > 2.3 \times 10^{26}$ at 90% C.L. [31]; the best result obtained employing ^{76}Ge is owed instead to the GERDA Collaboration, $T^{0\nu}_{1/2} > 1.8 \cdot 10^{26}$ yr [32]. A strong experimental effort is currently ongoing to find an evidence for this decay, as further discussed in Section 1.4.

$2\nu\beta\beta$	$0\nu\beta\beta$
$(A, Z) \rightarrow (A, Z + 2) + 2e^- + 2\bar{\nu}_e$	$(A, Z) \rightarrow (A, Z + 2) + 2e^-$
Allowed by SM, already observed	Beyond SM, not observed yet
$\Delta L = 0$	$\Delta L = 2$
$T^{2\nu}_{1/2} \approx 10^{18} - 10^{22}$ yr [29]	$T^{0\nu}_{1/2} > 2.3 \times 10^{26}$ yr [31]

TABLE 1.1: Summary of the main features of two neutrinos double beta decay ($2\nu\beta\beta$) and neutrinoless double beta decay ($0\nu\beta\beta$). The half life depends on the decaying isotope: for the $2\nu\beta\beta$ it has been measured for several nuclei and the presented range contains all the results; for the $0\nu\beta\beta$, instead, it has never been measured and the presented result is the current best experimental limit.

Due to the different kinematical features of $2\nu\beta\beta$ and the hypothetical $0\nu\beta\beta$, it would be possible to distinguish the two contributions in the E_{2e} energy spectrum, which is the distribution of the sum of the energies of the two emitted electrons (Fig. 1.5). Let us underline that in both cases the detected particles are only the emitted electrons, since the neutrinos escape from the detector after being produced.

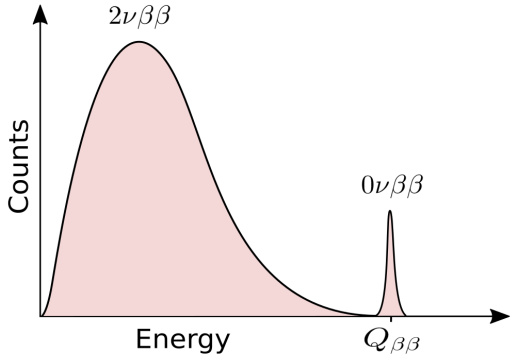


FIGURE 1.5: Schematic representation of the expected spectrum of E_{2e}, the sum of the energy of two electrons emitted by a double β decaying nucleus. The two contributions are well separated: a continuous distribution with $E_{2e} \in [0, Q_{\beta\beta}]$ for $2\nu\beta\beta$, and a sharp peak at $E_{2e} = Q_{\beta\beta}$ for $0\nu\beta\beta$.

If we assume the kinetic energy of the recoiling nucleus and the neutrino masses to be negligible (as $Q_{\beta\beta}$ is of $\mathcal{O}$(MeV)), the sum of the kinetic energies of all the emitted leptons is equal to the available energy, which is the Q-value of the transition $Q_{\beta\beta} = m_i - m_f - 2m_e$,

The standard $2\nu\beta\beta$ is a four-bodies decay ($2\nu + 2e$ in the final state), so the available energy $Q_{\beta\beta}$ is shared among the four emitted leptons. The sum of the energies of the two emitted electrons will therefore assume values in a continuous distribution spanning from $E_{2e} = 0$ to $E_{2e} = Q_{\beta\beta}$, as represented in Fig. 1.5. In particular, it is extremely improbable for the two neutrinos or for the two electrons to take all the energy available in the process, therefore the probability goes to zero for $E_{2e} \to 0$ and $E_{2e} \to Q_{\beta\beta}$.

The $0\nu\beta\beta$ decay, instead, would be a two-bodies decay (only $2e$ in the final state), so the sum of the energies of the two electrons E_{2e} would be identically equal to $Q_{\beta\beta}$.

The signature of the occurrence of the $0\nu\beta\beta$ would therefore be the presence of a sharp and isolated peak at $E_{2e} = Q_{\beta\beta}$ in the distribution of the sum of the energies of the two electrons. Note that in practice the width of the $0\nu\beta\beta$ peak depends also on the energy resolution of the employed detectors

This feature of the $2e$ energy spectrum is a powerful experimental tool in the search for $0\nu\beta\beta$: in fact the key idea of all the experiments searching this decay is to acquire the $2e$ energy spectrum and to look for a peak at $Q_{\beta\beta}$.

If the peak at $E_{2e} = Q_{\beta\beta}$ is observed, it implies that the $0\nu\beta\beta$ decay took place, and this implies that neutrinos are Majorana particles.

The eventual observation of the peak at $Q_{\beta\beta}$ would allow to extract an important additional information on neutrinos' effective Majorana mass $m_{\beta\beta}$. In fact, the following relation holds [24]:

$$\left[T^{0\nu}_{1/2}\right]^{-1} = G^{0\nu}(Q_{\beta\beta}, Z)\ |M^{0\nu}|^2\ \frac{|m_{\beta\beta}|^2}{m_e^2} \tag{1.17}$$

In which $G^{0\nu}(Q_{\beta\beta}, Z)$ is an integral over the phase space whose values are known and tabulated, Z is the atomic number of the decaying isotope, $M^{0\nu}$ is the nuclear matrix element of the transition and m_e the electron mass.

Therefore, if $T^{0\nu}_{1/2}$ is measured experimentally, neutrinos' effective Majorana mass $m_{\beta\beta}$ can be extracted inverting Eq. 1.17.

Moreover we know from theory [24], that in the standard interpretation of $0\nu\beta\beta$, which explains the decay with the exchange of a light Majorana neutrino, $m_{\beta\beta}$ can be written as a function of neutrino mass eigenvalues m_i (with $i = 1, 2, 3$), mixing angles θ_{ij} and Majorana phases κ, λ:

$$m_{\beta\beta} = \sum_{i=1}^{3} U_{1i}^2 m_i = (c_{12}^2 c_{13}^2 e^{2i\kappa})\, m_1 + (s_{12}^2 c_{13}^2 e^{2i\lambda})\, m_2 + (s_{13}^2 e^{-2i\delta})\, m_3 \tag{1.18}$$

This means that the effective Majorana mass is sensible to the neutrino mass hierarchy and to the Majorana phases: an experimental measurement of $m_{\beta\beta}$ (extracted from $T^{0\nu}_{1/2}$) would therefore give some information also about these topics.

1.4 Experimental search of $0\nu\beta\beta$

The hypothetical $0\nu\beta\beta$ is an extremely rare process: for this reason a crucial and critical point common to all the experiments aimed at studying this decay is the background, which requires to be suppressed and modeled so as to have few and known background contributions in the acquired energy spectrum.

A generic experiment having a mass M of the $\beta\beta$ decaying isotope, taking data for a time T, with an energy resolution σ and an eventual background B, will have a half life sensitivity [33]

$$T^{0\nu}_{1/2} \propto \begin{cases} M \cdot t & \text{Background-free} \\ \sqrt{\frac{M \cdot t}{B \cdot \sigma}} & \text{With background} \end{cases} \tag{1.19}$$

In which the background-free regime is defined as a condition in which the expectation is to observe less than one background count in the energy region of interest. If we define *exposure* the product of the mass of the detector multiplied by the time in which it acquires data (livetime):

$$\text{Exposure [kg yr]} = \text{Detector mass [kg]} \cdot \text{Livetime [yr]} \tag{1.20}$$

it is clear that the background-free regime is particularly favorable since then half life the sensitivity scales linearly with exposure, instead of with its square root.

Several factors can contribute to keep the background as low as possible. The experiment must be shielded from external radiation: this is why most of the experiments, LEGEND included, are built in underground facilities. The materials employed in the construction of the experiment itself must be also radiopure, in order to limit the internal background sources. This translates in a huge experimental effort to determine the most suitable materials in order to conciliate the background needs and the overall performances. Finally, active background suppression techniques and pulse shape discrimination analysis techniques must be employed to tag and remove the residual background.

Even after choosing the best location and the best materials, the background will not be identically zero in all the energy spectrum. For this reason it must be modeled and compared to the acquired data. The key idea of the background modeling is that knowing the geometry and the material composition of all the components of the experiment, the expected background can be extracted with a simulation.

In nature there are several double beta decaying isotopes (a list can be found in [34]), but not all of them are equivalently suitable candidates to perform an experiment. In fact, when planning an experiment, several factors must be taken into account to establish which is the best isotope to employ.

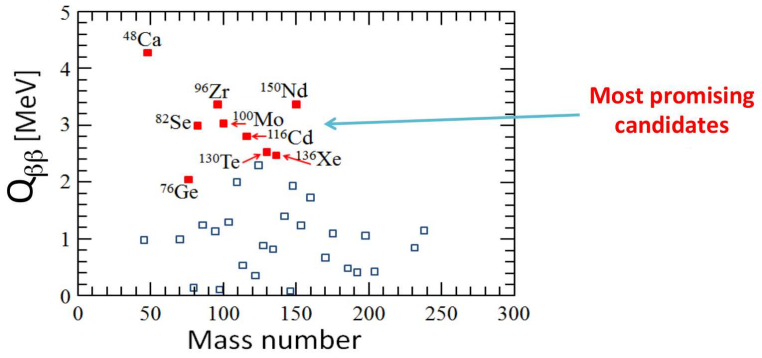


FIGURE 1.6: Good candidates for $\beta\beta$ searches are required to have te following features: high $Q_{\beta\beta}$, high (natural or achievable) isotopic abundance, and compatibility with viable detection techniques. Picture adapted from [35, 36].

The main factors guiding the isotope selection for an experiment are the following [23]. A good candidate isotope should have a high $Q_{\beta\beta}$: this leads to a larger phase space, since $G^{0\nu} \propto {Q_{\beta\beta}}^5$ in the standard interpretation and to higher powers of $Q_{\beta\beta}$ if other assumptions are made; also, a high $Q_{\beta\beta}$ grants a stronger background control, since at higher energies it is less probable to have a background processes which mimics the $0\nu\beta\beta$ signal of interest (γ background highly suppressed at high energy). Another valuable feature for the isotope is to have high natural isotopic abundance and/or the possibility of being enriched. Finally, the isotope has to be compatible with some viable detection techniques, otherwise the information about the energy of the emitted electrons is lost.

Depending on the chosen isotope and on the desired experimental design, several detection techniques can be employed to search the $0\nu\beta\beta$ decay [37]. A common feature to most of the experiments (LEGEND included) is to employ active detectors, namely to design a setup in which the detector itself is made up of a double beta decaying isotope: the same component acts then both as source and as detector medium.

- **Semiconductor detectors**

  When a charged particle (in the case of $0\nu\beta\beta$ the two emitted electrons) crosses a semiconductor detector, it ionizes the material; the ionization charges (electrons and holes clusters) drift towards the electrodes due to the electric field present in the active volume of the detector. Thus a charge signal is generated and read out, whose amplitude is proportional to the energy of the impinging particle. In case of active detectors, the source that decays $\beta\beta$ emitting the electrons that we want to detect is the semiconductor itself.

  An example of this class of experiments is LEGEND itself, as thoroughly discussed in Chapter 2, like its predecessors GERDA [38] and MAJORANA Demonstrator [39], in which ^{76}Ge-enriched High Purity Germanium Detectors are used.

  The main advantage of these detectors is the excellent energy resolution they can achieve, $\mathcal{O}$(keV) at $Q_{\beta\beta}$, namely $\mathcal{O}(0.1\%)$. Besides this, favorable is also the maturity of this detection technique and technology, extensively employed since the Sixties. The main disadvantages are instead the relatively low $Q_{\beta\beta}$ (2039.061(7) keV for ^{76}Ge [40]) and the high production cost, also driven by the need of enriching the Germanium before operating it (the natural abundance of ^{76}Ge is around 8% while the aimed fraction for experimental purposes is usually above 80%).

- **Bolometers**

  In a bolometer the energy release due to the impinging radiation is extracted from the measurement of temperature rising, with thermistors directly attached to the absorber medium. Since the temperature variation due to the absorption of the two electrons emitted in a double beta decay is extremely low, bolometric experiments must be performed at cryogenic temperatures ($\approx$ 10 mK).

Among the experiments employing the this detection technique there are CUORE [41] and CUPID [42] (^{130}Te), and AMoRE [43] (^{100}Mo).

Bolometers provide good energy resolution $\mathcal{O}(0.2\%)$ and can be operated using several isotopes, including some having high natural abundance, such as ^{130}Te. The critical point, instead, is of course to operate a large-scale detector at the required ultra-low temperatures.

- **Scintillators**

  Scintillating materials emit light when absorbing a charged particle: the number of produced photons is proportional to the energy of the impinging particle.

  Examples of experiment searching for $0\nu\beta\beta$ decay with a liquid scintillator are KamLAND-Zen [44], which employs enriched Xenon (^{136}Xe), and SNO+ [45], which employs Tellurium (^{130}Te).

  The main advantage of this type of detectors is the easy mass scalability, namely the possibility of deploying extremely large masses of the chosen material. The disadvantage instead, is the relatively poor energy resolution $\mathcal{O}$(7-10%).

- **Time Projection Chambers (TPCs)**

  When a charged particle crosses a TPC, the detection material produces both ionization charge and scintillation light. The combination of these two signals allows the reconstruction of event topology, position, energy and particle type.

  A common isotope choice as source and detection medium in $0\nu\beta\beta$ experiments employing TPCs is ^{136}Xe. An example of such experiment is EXO [46] and the upcoming NEXT [47].

  TPCs show good mass scalability and provide several background discrimination variables. The energy resolution, of $\mathcal{O}(1\%)$ at $Q_{\beta\beta}$, despite being better than in scintillators, cannot compete with that of the semiconductor detectors.

Fig. 1.7 shows the discovery sensitivity of the most promising ongoing and planned $0\nu\beta\beta$ experiments. The LEGEND experiment, both in the short term (LEGEND-200) and even more in the long term (LEGEND-1000) has a competitive and promising discovery sensitivity.

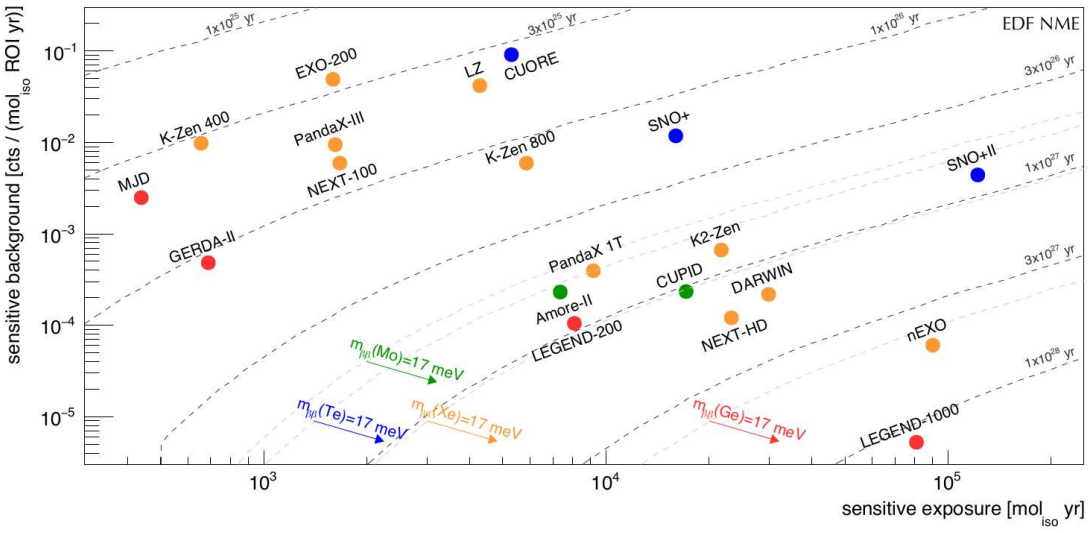


Figure 1.7: Comparison of the discovery sensitivity of completed (GERDA, MJD, K-Zen-400, EXO-200, CUPID-0), current and planned $0\nu\beta\beta$ experiments, color-coded by employed isotope: ^{76}Ge in red, ^{136}Xe in orange, ^{130}Te in green and ^{100}Mo in blue. The dashed lines represent the $0\nu\beta\beta$ half life for which the experiments would have discovery potential; the result is given assuming 10 yr of data taking for the current and planned experiments. The arrows indicate the half-life sensitivities required to test the whole inverted hierarchy scenario for the corresponding isotope. Picture from [48].

Chapter 2

The LEGEND Experiment

LEGEND (Large Enriched Germanium Experiment for Neutrinoless Double-beta decay) is an experiment devoted to the search of $0\nu\beta\beta$ in ^{76}Ge: the searched decay is then $^{76}\text{Ge} \rightarrow {}^{76}\text{Se} + 2e^-$. Its working principle is based on the usage of active High Purity Germanium detectors (HPGe), serving both as source of the decaying isotope and as detectors for the emitted electrons. The natural abundance of ^{76}Ge is about 8%: in order to improve the efficiency of the experiment, the Ge diodes are enriched in ^{76}Ge to a fraction above 90%. A feature of the $0\nu\beta\beta$ which is crucial for the design of the experiment and for the subsequent analysis is that this decay is expected to produce a highly localized signal in Germanium, with all the energy released in about 1 mm^3 [32]: therefore the experiment needs to operate multiple detectors to be sensible to coincidences, and all the events in which a signal is detected in coincidence in more than one detector can be identified as background with respect to $0\nu\beta\beta$.

Unless noted otherwise, the information presented in this Chapter come from the *LEGEND-1000 Preconceptual Design Report* [49].

2.1 Location and background mitigation

Since $0\nu\beta\beta$ is a rare process, eventually providing a low-statistics signal, a major issue to deal with when performing an experiment searching for this decay is to find a balance between the desired increase of the exposure and the needed reduction of the background. To minimize the presence of background coming from cosmic radiation, neutrino experiments are often build in underground facilities in order to exploit the fact that the above layer of rock absorbs most of this radiation, providing a clean environment. LEGEND belongs to this category of experiments: it is currently built in Gran Sasso National Laboratories (LNGS), a deep underground experimental area located under the Gran Sasso mountains, in Italy. The experimental site benefits from about 1400 m rock overburden, translating to a 3500 m water equivalent depth, which removes the hadronic components of cosmic ray showers and reduces the muon flux at the experiment by six orders of magnitude with respect to above-ground values, to a rate of about 1.2 muons/(m^2·h) [50].

Another background source which must be minimized is the radioactivity of the materials and components appearing in the experimental area and employed for the construction of the experiment itself (Section 2.4). This background is mitigated choosing the most radiopure materials and wrapping the core Germanium detectors in different layers of shielding. Besides these passive mitigation strategies, also active reduction techniques are deployed in order to identify and possibly remove the residual background signals: the liquid Argon (LAr) veto and the muon veto are precisely intended to detect and tag as background the signals which cannot come from a double beta decay.

2.2 LEGEND operation phases

The LEGEND experiment, built combining the best technologies from GERDA and MAJORANA Demonstrator, is planned to have two operation phases, primarily differentiated by the amount of deployed Germanium and consequently by the sensitivity they can achieve (Tab. 2.1).

The first phase, referred to as LEGEND-200 since 200 kg of Germanium are deployed, is operated in the pre-existing GERDA's infrastructure in LNGS. LEGEND-200 just started acquiring data in 2023 and is planned to achieve a background index $B.I. \leq 2 \cdot 10^{-4}$ counts/(keV kg yr), leading to a 3σ half-life discovery sensitivity for the searched decay of $1.5 \cdot 10^{27}$ years with five years of data taking, namely with an exposure of about 1 ton·yr. Using a nuclear matrix element $M^{0\nu}$ in the range of 2.66-6.34 for ^{76}Ge, a phase space factor and a value of g_A (contributing to $G^{0\nu}$) of $2.363 \cdot 10^{-15}$ yr and 1.27 respectively, the LEGEND-200 half life discovery sensitivity corresponds to an $m_{\beta\beta}$ upper limit in the range of 27-63 meV via Eq. (1.17) [23].

The objective of the second phase, LEGEND-1000, is to operate 1000 kg of Germanium (four modules of 250 kg each) and to achieve a background index $B.I. \leq 1 \cdot 10^{-5}$ counts/(keV kg yr), leading to a half life discovery sensitivity beyond $1.3 \cdot 10^{28}$ years with ten years of data taking, namely with an exposure of about 10 ton·yr. This half life corresponds to a $m_{\beta\beta}$ upper limit in the range of 9-21 meV via Eq. (1.17) [23].

In order to achieve this goal in LEGEND-1000, many modifications and upgrades must be deployed with respect to the LEGEND-200 setup. First of all, the GERDA infrastructure is too small to host the LEGEND-1000 setup: a bigger experimental hall is needed and two candidate sites are being considered, again LNGS or SNOLAB (Canada). Then, in order to achieve the desired sensitivity, the background mitigation must be pushed even further than in LEGEND-200: a neutron moderator will be added, to prevent external neutrons from reaching the detectors; each Germanium detectors array will be deployed in a reentrant tube filled with Underground sourced Liquid Argon (UGLAr), which has a lower concentration of ^{42}Ar with respect to the atmospheric LAr, in order to avoid this contamination at least in the region closer to the detectors; the reentrant tubes themselves are immersed in a cryostat filled with LAr.

Thanks to these technical upgrades and to the advanced analysis techniques deployed to identify (and eventually remove) background signals appearing in the energy spectrum, LEGEND-1000 is expected to achieve a quasi-background free regime.

Parameter	LEGEND-200	LEGEND-1000
Germanium mass	200 kg	1000 kg
Background Index ($B.I.$)	$\leq 2 \cdot 10^{-4}$ counts/(keV kg yr)	$\leq 1 \cdot 10^{-5}$ counts/(keV kg yr)
Half life sensitivity ($T^{0\nu}_{1/2}$)	$1.5 \cdot 10^{27}$ yr	$1.3 \cdot 10^{28}$ yr
Effective Majorana mass sensitivity ($m_{\beta\beta}$)	27-63 meV	9-21 meV

Table 2.1: Summary of the key parameters expected for the two phases of the LEGEND experiment. Ten years of data taking are assumed in the calculation of the half life and the effective Majorana mass. Data from [23].

The analysis presented in this work is performed on data collected in September 2022 during the commissioning phase of LEGEND-200, with 60 kg of Germanium already deployed.

2.3 Experimental setup

The basic design of the LEGEND experiment is shown in Fig. 2.1: bare Germanium detectors, organized in strings (1), are deployed in a cryostat (2) filled with about 64 m^3 of liquid Argon, serving both as refrigerant for the Germanium detectors and as veto for background signals; this whole setup is immersed in a 590 m^3 volume of ultrapure water (3) serving as veto for the muons reaching the experimental site.

Figure 2.1: Sketch of the whole LEGEND experimental setup. From the core: the Germanium detectors strings (1), the cryostat filled with liquid Argon (2) and the water tank (3). Picture credits: Patrick Krause [51].

2.3.1 Germanium detectors

FIGURE 2.2: Sketch of the Germanium detectors array. The nylon mini-shroud wrapping each string is visible in transparency. Credits: Patrick Krause [51].

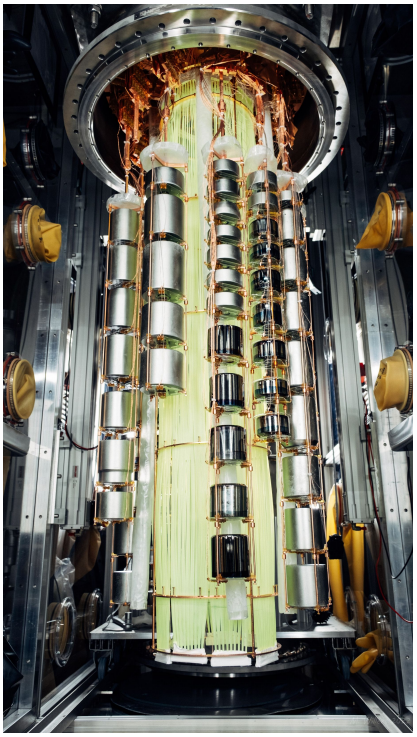


FIGURE 2.3: Photo of the Gemanium strings deployed around the inner fiber barrel (see Section 2.3.2). Credits: Mike Willers [51].

The core of the LEGEND experiment is the array of HPGe detectors aimed at detecting the $0\nu\beta\beta$ signal. These detectors are arranged in strings, as shown in Fig. 2.2 and 2.3, thanks to a mounting system including Electro-Foamed Copper (EFCu) vertical supports and Polyethylene Naphthalate (PEN) baseplates for each detector. Each string is wrapped by a thin ($\approx$ 125 µm) nylon layer, referred to as mini-shroud (MS), introduced to mitigate the flux of ^{42}K ions reaching the detectors.

In the LEGEND-200 phase several types of Germanium detectors are employed, produced by both ORTEC and Mirion companies: Broad Energy Germanium detectors (BEGe), P-type Point-Contact Germanium detectors (PPC), Inverted-Coaxial [Point-Contact] Germanium detectors (IC[PC]), and semi-Coaxial Germanium detectors (Coax). These classes of detectors differ in shape, mass, position of the electrodes and consequently shape and strength of the inner electric field.

In the commissioning phase, only ICPC detectors are deployed, therefore we focus on their features. Their shape, sketched in Fig. 2.5, is cylindrical with a large cavity in the upper part. The typical dimensions of an ICPC detector are about 8 cm diameter and 10 cm height. The p^+ electrode (boron implantation), located at the bottom of the detector, is small, and this grants a high electric field gradient in that area, while the whole top face acts as n^+ electrode (lithium layer). The advantages of ICPC detectors are that they have rather large masses (average mass of a diode about 2 kg) and that they have excellent Pulse Shape Discrimination (PSD) properties. LEGEND ICPC detectors reach an average energy resolution at $Q_{\beta\beta}$ around ≈3 keV FWHM, as discussed in Chapter 4. The electric field and the weighting potential inside an ICPC detector are represented in Fig. 2.5 and 2.6, respectively. In LEGEND, the ICPC Germanium detectors are operated at a High Voltage (HV) in the range 2-5 kV [52].

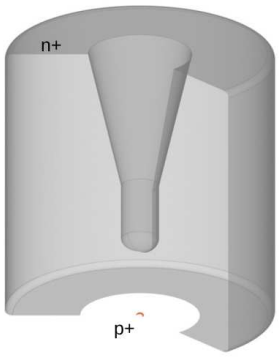


FIGURE 2.4: ICPC detector geometry. Credits: David Hervas Aguilar [51].

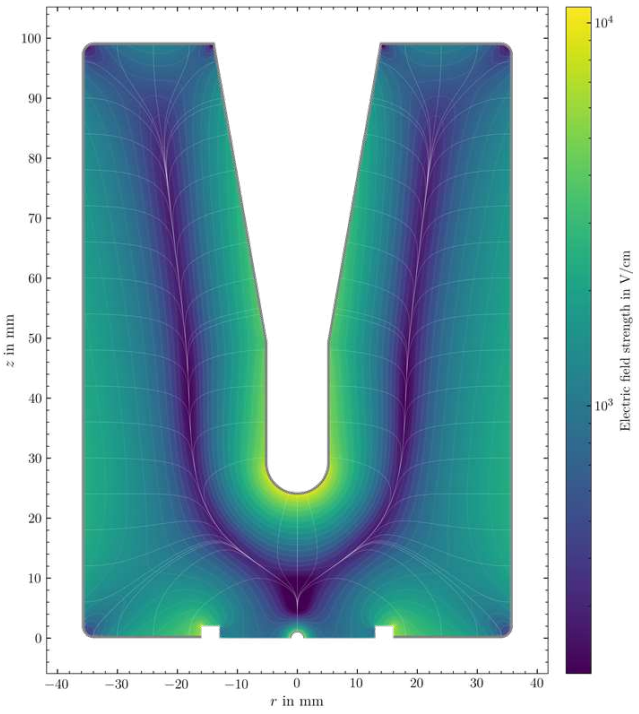


FIGURE 2.5: Model of the electric field strength in the active volume of an ICPC detector. The thin lines in the foreground represent the electric field lines, while the thicker lines in the background represent lines of equal field strength. Contacts are shown in gray at the borders of the detector: the p+ as a thin line and the n+ as a thick line. Credits: David Hervas Aguilar [51].

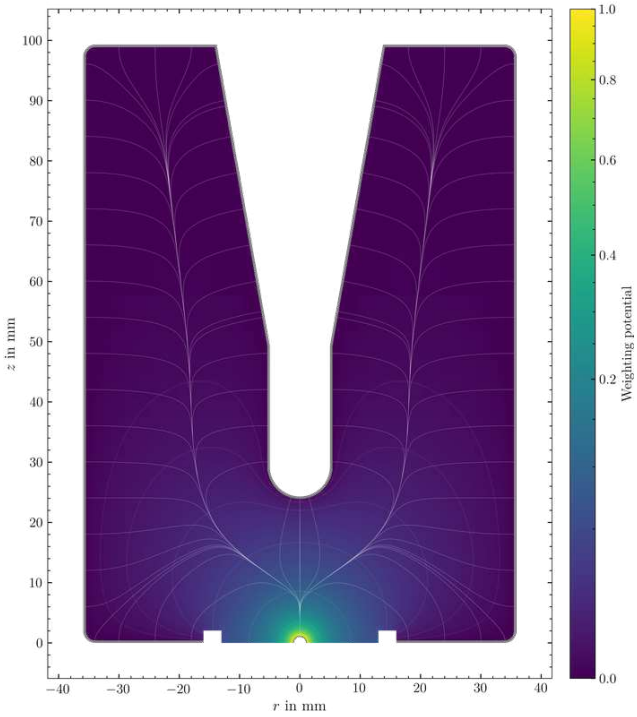


FIGURE 2.6: Model of the weighting potential in the activevolume of an ICPC detector. The thin lines in the foreground represent the electric field lines, while the thicker lines in the background represent lines of equal weighting potential. Contacts are shown in gray at the borders of the detector: the p+ as a thin line and the n+ as a thick line. Credits: David Hervas Aguilar [51].

When a charged particle crosses a Germanium detector, it loses energy producing electron-hole couples: due to the electric field, these ionization charges drift towards the corresponding electrode, where they are collected and read out. In order to maximize the active volume, namely the volume in which the produced ionization charge can be properly collected, the detectors must be fully depleted: in this regime the active volume nearly coincides with the whole volume of the detector, with the exception of a thin layer, of $\mathcal{O}$(1mm) [53], close to the surface, referred to as dead layer.

In the LEGEND setup the charge collected at the electrodes of each Germanium detector reaches a cryogenic charge-sensitive preamplifier (CC4) and then the signal is digitized by a FC250b FlashCam module at 62.5MHz/16bit [51]. Since each Germanium detector is read out individually, it can be univocally identified by the FlashCam (FC) channel in which its signals are digitized.

In order to check the stability of the acquisition and of the performances of the detectors, a pulser is introduced in the acquisition chain, providing a known signal periodically appearing among the collected data. In fact, the pulser is a pulse generator which injects in the system signals with a precise shape and frequency. Further details on the pulser signals are discussed in Chapter 5.1.

In order to extract meaningful information from the collected signals, the Germanium detectors must be calibrated, namely a relation must be found between the amplitude of such signals and the amount of energy that was released in the interaction. As thoroughly discussed in Chapter 4, the calibration is performed exposing the detectors to a known source and evaluating the position of γ lines in the collected energy spectra. The calibration source employed in LEGEND-200 commissioning phase is ^{228}Th: 13 sources, having a total activity of about 24 kBq [51], are lowered into the cryostat and brought close to the detectors by a mechanical system referred to as Source Insertion System (SIS). In order to achieve a uniform exposure of all the detectors of the array, calibration data are acquired with the sources located in three different vertical positions: P_1 = 8200 mm, P_2 = 8500 mm and P_3 = 8799 mm, defined as vertical distances from the SIS, located on top of the cryostat.

2.3.2 Cryostat and LAr veto

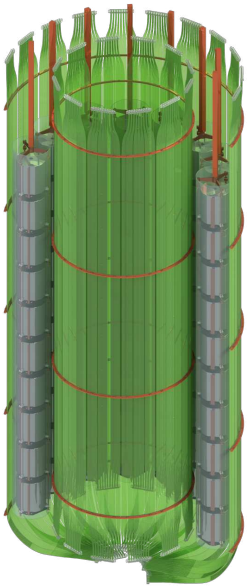


Figure 2.7: Sketch of the inner and outer fiber barrel around the Germanium Detectors array. Picture from [54].

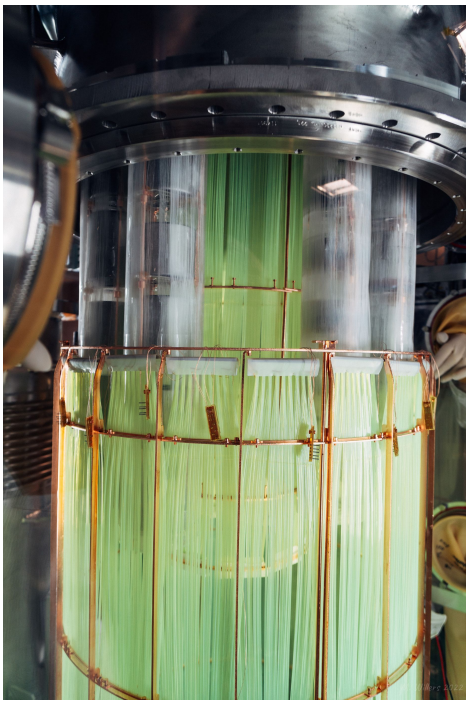


Figure 2.8: Photo of the Germanium detectors array, surrounding the inner fiber barrel, being lowered into the outer fiber barrel. Credits: Mike Willers [51].

Germanium detectors must be operated at cryogenic temperatures: in fact at higher temperatures the electrons in the valence band of the semiconductor would spontaneously pass to the conduction band by thermal energy, drift towards the anode due to the electric field and thus mimic the effect of a signal. For this reason LEGEND's Germanium detectors array is deployed in a cryostat filled of liquid Argon at a temperature around 88.8 K. The dimensions of the cryostat are about 4 m of diameter and 5 m of height, for a total volume of 64 m^3 of LAr [55]. Besides serving as a cooler and,

automatically, as a passive background shield for the germanium detectors, the liquid Argon has some precious scintillation properties which make it possible to employ it as an active veto for background signal.

To achieve this goal, the liquid Argon is properly instrumented: two concentric curtains (Inner barrel and Outer barrel) of TetraPhenyl Butadiene (TPB) coated, double-cladded WaveLength Shifting (WLS) fibers are deployed in the inner and outer side of the Germanium detectors strings (Fig. 2.7 and 2.8). The fibers are read out, at the top and at the bottom, by a SiPM and the signals from the SiPMs are then digitized.

The working pronciple of this instrumentation is sketched in Fig. 2.9, with time flowing from top to bottom, from the interaction in the LAr to the readout.
When some ionizing radiation interacts with liquid Argon, the latter emits 128 nm Vacuum Ultra-Violet (VUV) light. This light hits then the TPB layer wrapping the fibers which shifts the VUV light to visible blue light. Finally the light hits the WLS fiber which shift the blue light to green light. This green light, after traveling in the fiber, is read out by a Silicon PhotoMultiplier (SiPM).
Therefore, when a SiPM detects a signal, it means that some ionizing radiation has interacted with the active volume of liquid Argon. In particular, the total number of SiPM channels in the commissioning phase of LEGEND-200 is 58: 29 at the top and 29 at the bottom and, at each end, 9 in the inner barrel and 20 in the outer. Each SiPM channel consists in an array of 9 distinct SiPMs, read out in parallel. Each SiPM is a 3x3 mm^2 PM33100T cell by Ketek made up of pixels in which the fibers end and are read out (Fig. 2.10)

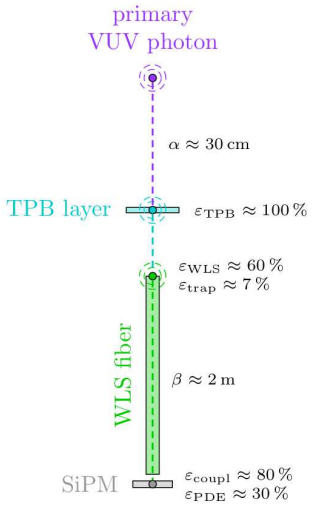


FIGURE 2.9: Simplified sketch of the LAr light collection chain, with efficiencies in ideal conditions. Picture from [56].

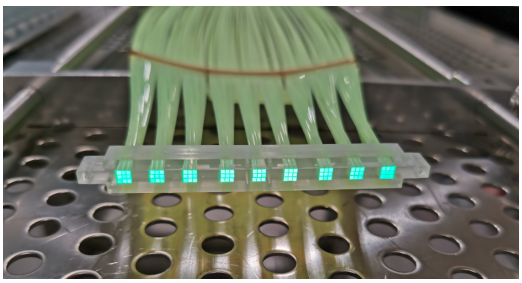


FIGURE 2.10: One SiPM channel: parallel connection of the fibers in 9 cells, each then read out by a 3x3 mm^2 SiPM. Credits: Patrick Krause [51].

The SiPMs channels are connected to the Germanium global trigger: when a Germanium detector detects a signal, also all the SiPM acquire data simultaneously.

Since the energy released in $0\nu\beta\beta$ events is expected to be highly localized, when a signal is detected in coincidence by a Germanium detector and a SiPM channel, the corresponding event is not $0\nu\beta\beta$, but rather can be identified as background with respect to this decay. This background tagging procedure is referred to as LAr veto. The action of the LAr veto on the acquired Germanium energy spectrum is discussed in Section 5.5.

2.3.3 Water tank and muon veto

The whole cryostat, containing the LAr and the Germanium detectors array, is immersed in 590 m^3 of ultra-pure water (Fig. 2.1).

Again, this has a double function: passive shield and active veto. In particular, water acts as a shield mostly against radiation coming from the experimental hall (gammas and neutrons) and as active veto against muons. In fact, when an energetic charged particle crosses water, Cherenkov light is emitted. This light is collected by 66 dedicated PhotoMultiplier Tubes (PMTs) located on the surface of the water tank and their signal is then read out.

In order to maximize the light collection efficiency, a 206 µm thick reflective foil covers the inner walls of the water tank, its floor and the outer walls of the cryostat. Finally, considering that the muons could pass undetected through the neck of the cryostat, an array of plastic scintillators is mounted on the roof of the clean room. Considering the results obtained by GERDA with the same setup, the expected muon rejection efficiency is $> 99\%$ [57].

Let us underline that in principle this setup would detect the signal of all the charged cosmic radiations passing through the water volume; in practice, though, only muons reach the experimental site and can be detected. For this reason, this part of the setup is referred to as muon veto.

Also in this case, exploiting the expected high localization of the energy release from $0\nu\beta\beta$, when a signal is detected in coincidence by a Germanium detector and a PMT, the corresponding event is not $0\nu\beta\beta$, but rather can be identified as a muon.

2.4 Residual background contributions

All the materials employed in the construction of the experiment are chosen for their radiopurity, to minimize the presence of background source close to the detectors. Despite this effort, some contaminants are still present.

In particular, besides the external muons reaching the experimental site, the most relevant background sources for LEGEND are the following:

- **^{238}U and ^{232}Th decay chains and ^{40}K internal contamination**

  Due to their long half lives and high natural abundance, ^{238}U, ^{232}Th, and ^{40}K appear in trace quantities in the materials employed in the fabrication of all the experimental components. ^{238}U and ^{232}Th decay to ^{206}Pb and ^{208}Pb respectively, through a series of short-lived isotopes. Among these, ^{208}Tl and ^{214}Bi are of primary concern for ^{76}Ge $0\nu\beta\beta$ searches ($Q_{\beta\beta} = 2039\,\text{keV}$), since they can emit gammas of energy 2104 keV and 2119 keV respectively: as further discussed in Chapter 5.7, these two energy regions will be excluded from the Region Of Interest (ROI) for the $0\nu\beta\beta$ analysis.

  The most relevant contribution from ^{40}K is the 1461 keV γ line, which is expected to be clearly visible in the energy spectrum.

- **Cosmogenic isotopes in Ge: ^{60}Co and ^{68}Ge**

  When enriched Ge material is exposed to cosmic rays, ^{60}Co and ^{68}Ge are produced: ^{60}Co decays β with a Q-value of 2.8 MeV to ^{60}Ni, which then de-excites emitting two γ rays in coincidence, having energy 1173 keV and 1333 keV respectively; ^{68}Ge decays β to ^{68}Ga, which then decays by electron capture or β^+ emission with a Q-value of 2.9 MeV.

  In the case of LEGEND, the *enriched Ge material* is represented by Germanium detectors themselves: when operated underground, they are not exposed to cosmic rays, but they are indeed exposed during the production and transportation phases, which are carried out above ground. To estimate the expected contamination, the exposure above ground is carefully recorded.

- **Surface α decays**

  During detector components fabrication and handling, contaminants like ^{210}Po and ^{210}Pb can deposit on their surface. These contaminants can emit high-energy α particles (up to 5.3 MeV in the case of ^{210}Po) which can penetrate the surface and deposit part of their energy in the active volume of the detector.

- **^{42}K from liquid Argon**

  The atmospheric Ar that fills the cryostat contains traces of the long-lived ^{42}Ar, which β decays to ^{42}K with a Q-value of 599 keV: it does not produce background in the ROI for ^{76}Ge $0\nu\beta\beta$. ^{42}K, though, decays again β to ^{42}Ca, with a Q-value of 3525 keV: this decay can then produce background in the ROI.

  Liquid Argon contains also traces of ^{39}Ar, which decays β to ^{39}K with a Q-value of 565 keV. This contribution, despite being expected to be dominant in the low energy region of the acquired spectrum, appearing as a continuous distribution from 0 to 565 keV, is not considered harmful for the $0\nu\beta\beta$ search since its endpoint is much below ^{76}Ge $Q_{\beta\beta}$ and its progeny is stable.

- **External γ rays**

  Also the materials appearing in the cryostat, in the water tank and in the outer experimental area can contain radioactive isotopes contributing to the detected background. In particular, the most relevant expected γ line with this origin is 2615 keV from ^{208}Tl (from ^{232}Th chain) from the stainless steal cryostat and water tank.

- **External Neutrons**

  The cryostat material can produce also neutrons, which also contribute to the background: α decays from the ^{238}U and ^{232}Th chains within the apparatus can generate neutrons via (α, n) reactions. Neutrons produced out of the water tank, instead, are effectively absorbed by the water itself.

Another crucial contribution which is expected to be clearly visible in the energy spectrum is the continuous distribution from 0 keV to $Q_{\beta\beta} = 2039\,\text{keV}$ from the $2\nu\beta\beta$ decay of ^{76}Ge itself.

Chapter 3

Data taking

3.1 Data taking structure

The data taking is organized in runs, each characterized by a fixed configuration of the setup and of the acquisition settings and parameters. Data from a single run are therefore consistent and can be analyzed together.

In order to evaluate the stability of the acquisition during a run, some parameters are monitored, such as waveforms' baseline value, leakage current, pulser gain and number of discharges. All these parameters should in principle remain constant during one run: if their actual distribution results not to be constant over time, it means that the detector is not stable. Also, some of the monitored parameters, such as the leakage current and the number of discharges are required to be sufficiently low: if they are not, the dataset is not suitable for the analysis.

For the LEGEND-200 commissioning phase, the runs in which the acquired data are most stable over time are runs 025, 026 and 027. The following analysis is therefore performed on these three runs.

Considering the structure of the detectors, it is crucial to distinguish these two concepts, defined here and employed in throughout this work:

- **Signal**: charge pulse detected by a single detector;
- **Event**: set of the signals detected by all the detectors of the array at a certain instant, due to the global trigger.

Therefore, while a signal refers to the single detector, an event refers to the full array: an event is made up of all the signals detected simultaneously by the detectors of the array.

3.2 Run structure

Each run is composed of two phases:

- **Calibration**

  This phase takes place at the beginning of each run, on a weekly basis. It consists in positioning a well known radioactive source (^{228}Th) close to the detectors and acquiring their signals. The purpose of this operation is to get a relation among the observed signals (ADC counts) and the energy deposited in the detectors (keV).

- **Physics**

  Once the system is calibrated, the radioactive sources are removed and data are acquired: this phase provides the physical data to be analyzed looking for the $0\nu\beta\beta$ peak.

During the physics data taking, the Germanium detectors array is operated in global trigger mode: when one Germanium detector detects a signal whose energy is above acquisition-threshold, all the detectors of the array, all the SiPMs of the LAr veto and all the PMTs of the muon veto acquire data simultaneously. This triggering criterion is chosen in order to allow a proper coincidence study both among the Germanium detectors, between Germanium and SiPMs and between Germanium and PMTs. This will be crucial in the analysis for the search for $0\nu\beta\beta$, since all the events in which multiple signals are detected in coincidence (either Ge-Ge or Ge-SiPM or Ge-PMT) cannot be identified as $0\nu\beta\beta$.

Note that even if there is no need to modify the configuration of the setup and of the acquisition settings, after about one week of data taking the acquisition is interrupted in order to perform a new calibration: in this way a new run begins.

The rationale for this structure for the data taking is to ensure a periodic energy calibration of the collected data and to highlight immediately potential instabilities of the system or of the acquisition. In facts, if the system is stable, calibration parameters will be constant as long as the acquisition parameters remain unchanged. Instead, if the optimal calibration parameters vary in time, it means that the system is not stable. Only after a proper energy calibration data from different runs become comparable and can be analyzed together.

The collected data are stored consistently with this acquisition structure: for each run calibration data (*cal*) and physics data (*phy*) are saved and can be accessed separately.

3.3 Dataset employed in this analysis

Details on the duration of the calibration for the three considered runs can be found in Table 3.1. Data from these calibration acquisitions are not explicitly analyzed in this work, though: the energy calibration procedure developed and performed by the Analysis Group of the Collaboration is briefly presented in Chapter 4, but our study is focused directly on the calibrated energy of the signals acquired during the physics data taking.

Details on the duration of the physics data taking can be found instead in Table 3.2: this will be the dataset for the analysis presented in this work.

Run	Source position	Start	Stop	Duration
025	P_1	09-09-2022 19:02:52 UTC	09-09-2022 21:04:09 UTC	2h 01m 17s
	P_2	09-09-2022 21:09:02 UTC	09-09-2022 23:34:26 UTC	2h 25m 24s
	P_3	09-09-2022 23:37:31 UTC	10-09-2022 01:32:39 UTC	1h 55m 08s
026	P_1	16-09-2022 12:02:15 UTC	16-09-2022 13:56:00 UTC	1h 53m 45s
	P_2	16-09-2022 14:05:25 UTC	16-09-2022 15:56:44 UTC	1h 51m 19s
	P_3	16-09-2022 16:01:31 UTC	16-09-2022 18:35:21 UTC	2h 33m 50s
027	P_1	23-09-2022 16:54:18 UTC	23-09-2022 18:51:58 UTC	1h 57m 40s
	P_2	23-09-2022 18:57:23 UTC	23-09-2022 20:55:05 UTC	1h 57m 42s
	P_3	23-09-2022 21:01:27 UTC	23-09-2022 22:57:28 UTC	1h 56m 01s

TABLE 3.1: Calibration phase for the runs of interest. In each calibration run the source is lowered in the cryostat and data are taken with the sources in three different positions: $P_1 = 8200$ mm, $P_2 = 8500$ mm and $P_3 = 8799$ mm, defined as vertical distances from the SIS, located on top of the cryostat. About two hours of data are collected for each position.

Run	Start	Stop	Duration	Total net duration available for analysis
025	10-09-2022 02:12:26 UTC 12-09-2022 15:30:12 UTC	11-09-2022 02:19:27 UTC 16-09-2022 08:54:37 UTC	1 day, 0h 07m 01s 3 days, 17h 24m 25s	4 days, 16h 34m 00s
026	16-09-2022 20:07:35 UTC 18-09-2022 16:29:13 UTC	18-09-2022 15:51:10 UTC 23-09-2022 09:51:29 UTC	1 day, 19h 43m 35s 4 days, 17h 22m 16s	6 days, 11h 56m 20s
027	23-09-2022 23:52:51 UTC	02-10-2022 20:25:58 UTC	8 days, 19h 11m 20s	8 days, 20h 33m 07s

TABLE 3.2: Physics data taking phase for the runs of interest.

Our dataset	Total net duration available for analysis
Runs 025+026+027	19 days, 23h 41m 40s

TABLE 3.3: Total net duration of the data acquisition available for the analysis presented in this work.

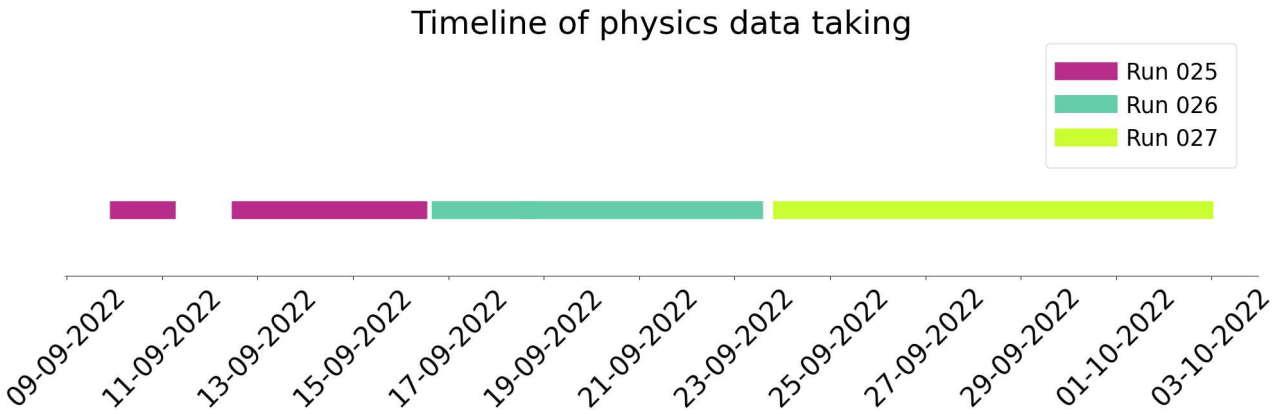


FIGURE 3.1: Timeline showing when the physics data employed in this analysis were collected.

- **Run 025**

  After slightly more than one day of data acquisition, the data acquisition system (DAQ) experienced a crash [52] that caused the interruption of the data taking for more than 37 hours before being able to start collecting data again. This interruption is clearly visible in the timeline (Fig. 3.1). Since data before and after the crash are consistent and appear not to be affected by the interruption, they are analyzed together.

- **Run 026**

  Also run 026 experienced a short interruption, due to an update of the DAQ codes [52], which again did not impact the data acquired before and after it. Being very short (less than 10 minutes), this interruption is not visible in the timeline (Fig. 3.1), but of course is taken into account when evaluating the duration of the run.

- **Run 027**

  During run 027, instead, data are acquired without any interruption for its full duration.

During the three runs the average trigger rate has been rather stable: about 0.83 events/s in run 025, 1.06 events/s in run 026 and 0.88 events/s in run 027.

Data acquired in one run are not stored in one single file, but rather segmented in multiple files, of the duration of about one hour of acquisition each, which are saved one after the other at run time. The rationale for this procedure is that in case of upcoming instabilities or in case of a crash of the DAQ system, only the interested data are lost, and not the entire run. In this storage process, few seconds of acquisition are lost between the end of one file and the start of the following one. When evaluating the net duration of each run then, we consider only the time of actually available acquired data, so we exclude the time span between different data files. In particular, the net time duration of each run is computed from the number of acquired pulser signals, exploiting the fact that they have a fixed and known frequency.

The dataset employed for our analysis is thus given by the union of the data properly acquired and saved in the three runs: its total net duration, crucial for the calculation of the exposure, is provided in Tab. 3.3.

3.4 Germanium detectors in runs 025, 026 and 027

During the considered runs, 28 ICPC Germanium detectors are deployed in the LAr bath. These detectors have different shapes and masses and they are organized in four strings (string 1, 2, 7 and 8), as shown in Fig. 3.2. Their technical specifications can be found in Table 3.4.

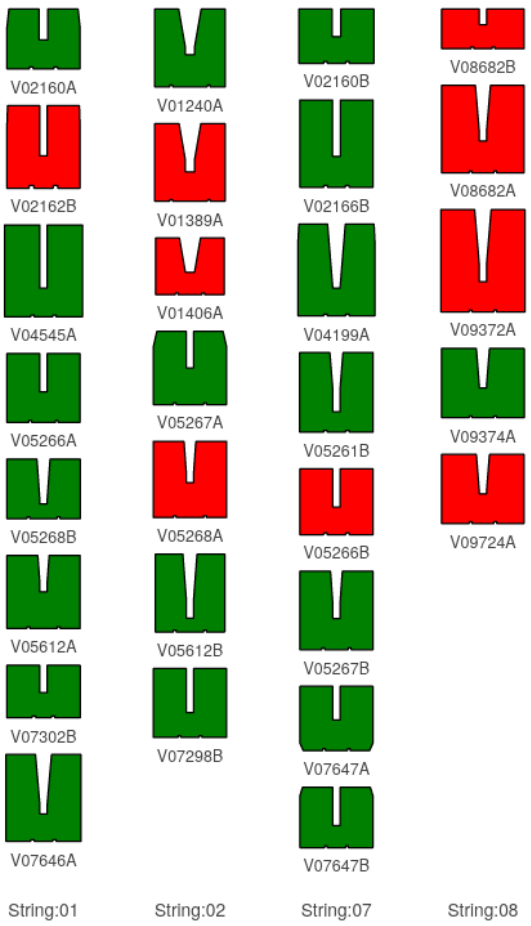


Figure 3.2: Germanium detectors string configuration in LEGEND-200 commissioning phase. The detectors are color-coded by status: in green the detectors whose data (at least part of them) are employed in this analysis, in red the discarded detectors. Picture from [58].

Not all the mounted detectors, though, provide reliable data, suitable for the analysis: during the physics data taking, some of them presented instabilities in the monitored parameters, especially in the pulser gain and in noise. An example of this behavior is shown in Fig. 3.3 for string 8: only one out of the five installed detectors appears to be stable in time (ch005). Similar instabilities appear also in a small fraction of the detectors from the other strings.

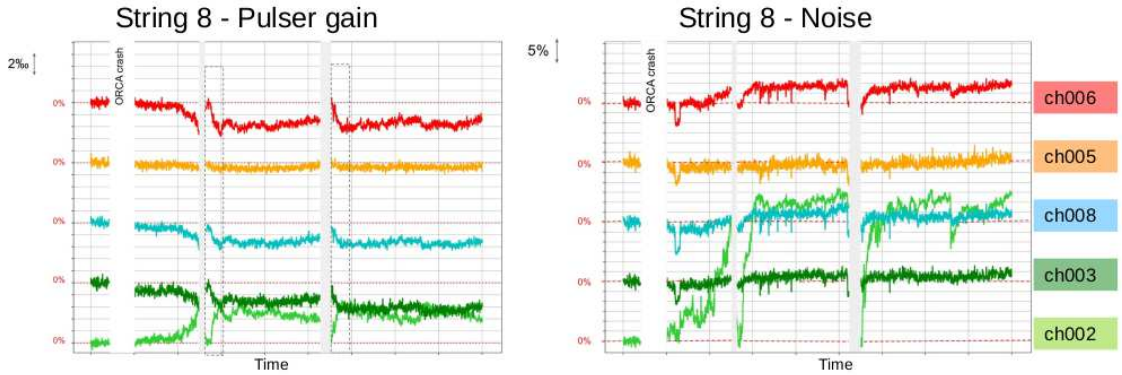


FIGURE 3.3: Examples of monitoring plots for detectors of string 8. The considered parameters are the pulser gain and the noise. The plots show that the only stable channel is ch005 (orange), while all the others suffer significative instabilities over time. Picture adapted from [59].

Data from all the detectors presenting those instabilities, color-coded in red in Fig. 3.2 and shaded in Tab. 3.4, are not considered in the analysis [60].

The stable detectors are instead color-coded in green in Fig. 3.2: their data are those actually employed in the analysis. For these detectors we calculate and present in Tab. 3.4 also the exposure, defined as the product of the mass of the detector multiplied by its livetime (Eq. (1.20)).

All but one of the stable detectors are considered for the full duration of runs 025, 026 and 027, reported in Tab. 3.3: this is therefore the time duration employed in the calculation of their exposure. The only exception is ch028 (string 1), color-coded in blue in Tab. 3.4: it is considered for runs 025 and 026, but discarded for run 027 since it started presenting some instabilities. In particular, its pulser gain suddenly increases at the beginning of run 027, making uninformative the reconstructed energy of the signal. in fact, if the gain changes during the acquisition, the same signal would be associated to different energies depending on when it was acquired The effective acquisition time for ch028, employed in the calculation of its exposure, is then given by the sum of the net durations of run 025 and 026 only: 11 days, 4h 30m 20s.

String	Position	FlashCam channel	Detector name	Mass [kg]	Exposure [kg yr]
1	1	ch023	V02160A	1.75	0.096
	2	ch024	V02162B	2.48	—
	3	ch025	V04545A	3.138	0.172
	4	ch026	V05266A	2.073	0.113
	5	ch027	V05268B	1.791	0.098
	6	ch028	V05612A	2.201	0.067
	7	ch029	V07302B	1.592	0.087
	8	ch016	V07646A	2.63	0.144
2	1	ch009	V01240A	2.1	0.115
	2	ch010	V01389A	2.093	—
	3	ch011	V01406A	1.383	—
	4	ch012	V05267A	2.183	0.119
	5	ch013	V05268A	2.298	—
	6	ch014	V05612B	2.092	0.114
	7	ch015	V07298B	2.085	0.114
7	1	ch037	V02160B	1.719	0.094
	2	ch038	V02166B	2.634	0.144
	3	ch039	V04199A	2.987	0.163
	4	ch040	V05261B	2.393	0.131
	5	ch041	V05266B	1.988	—
	6	ch042	V05267B	2.362	0.129
	7	ch007	V07647A	1.893	0.104
	8	ch043	V07647B	1.779	0.097
8	1	ch002	V08682B	1.517	—
	2	ch003	V08682A	3.34	—
	3	ch008	V09372A	4.046	—
	4	ch005	V09374A	2.648	0.145
	5	ch006	V09724A	2.643	—

TABLE 3.4: Active Germanium detectors in LEGEND-200 commissioning phase, color-coded by status: in black the stable detectors whose data are actually employed in this analysis, for the full duration of the data taking, in gray the discarded detectors, in blue the detector exploited only for a subset of the data taking period. In particular, ch028 is included in the analysis of runs 025 and 026, but discarded in run 027. For the selected detectors, the exposure is calculated according to Eq. (1.20). All the detectors deployed in this phase are ICPC.

Finally, Tab. 3.5 summarizes the features of the detectors actually employed in our analysis: their total mass and their total exposure. Please note that the available mass is comparable with the total Germanium mass available during the GERDA Phase II experiment, 44.2 kg.[1] Of course, since our acquisition time is much shorter than GERDA's, our exposure (and consequently our statistics) is much lower: 2.248 kg yr against the 103.7 kg yr collected by GERDA [32]. This translates in the possibility of putting just looser bounds to the half life of the searched $0\nu\beta\beta$ decay (Eq. (1.19)).

Total mass [kg]	Total exposure [kg yr]
42.050	2.248

TABLE 3.5: Total mass and total net exposure of the actually employed detectors.

[1] In its final phase, GERDA operated 6 coaxial detectors (14.6 kg), 30 BEGes (20 kg) and 5 IC detectors (9.6 kg).

Chapter 4

Data production

The raw acquired signals needs to be processed in order to extract the parameters then employed in the $0\nu\beta\beta$-targeted analysis. The data production flow is processed in multiple steps, referred to as *tiers*, each providing higher level information about the data. A crucial step in this process is the evaluation of the energy of the signals, which can be performed with different reconstruction methods and which needs the inputs coming from the calibration in order to be finalized.

4.1 Data production flow and tiers structure

The data production flow, schematized in Fig. 4.1, is organized in multiple subsequent tiers, starting from the digitized signals produced by the DAQ and leading to a set of high level parameter characterizing these signals.

After the acquisition, the three main levels are the following: the *raw* tier, collecting raw signals (waveforms) and the basic parameters provided by the DAQ; the *dsp* tier, standing for digital signal processing (of the single waveform), providing the most relevant features of the signals, except for those requiring the energy calibration; the *hit* tier, which benefits from the *dsp* parameters and from the results of the energy calibration, and collects the final set of parameters needed to characterize a physical signal.

Data are saved according to this structure: for each sample of collected data there are three different `.lh5` files containing, for each signal, its *raw*, *dsp* and *hit* parameters respectively.

Data processing is performed employing `pygama` [61], a Python package developed precisely for physics data processing.

Tab. 4.1 provides a summary of the most relevant observables employed in this analysis, grouped by the tier they belong to. Further details about each parameter will be provided when explicitly employing it.

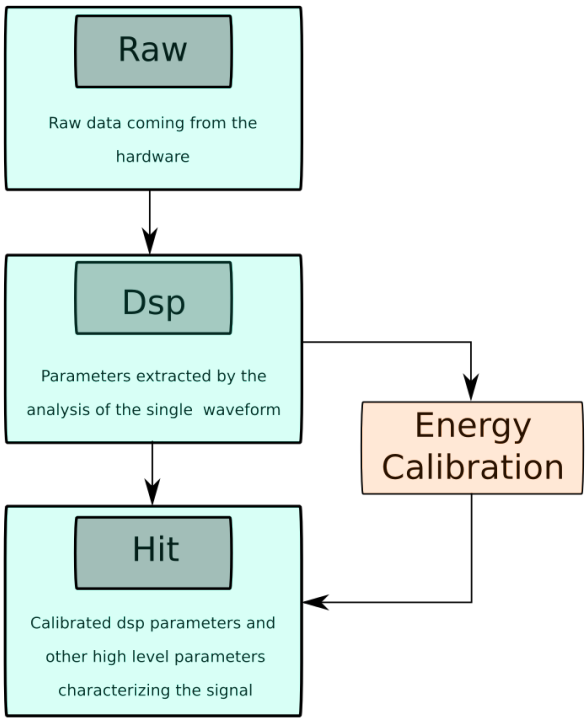


FIGURE 4.1: Extremely simplified representation of the data production flow leading from DAQ raw data to high level parameters characterizing the collected signals.

Tier	Parameter	Unit of measure	Description
Raw	`waveform`	ADC	Array of points sampled from the waveform (rate 62.5 MHz, one sample every 16 ns)
Dsp	`timestamp`	s	Timestamp of the event (UTC+00 format)
	t_0	ns	Start of the risetime
	t_{10}, t_{50}, t_{90}	ns	Time in which the amplitude is 10, 50, 90% of the maximum
	`t_max`	ns	Time in which the amplitude of the signal is maximum
	`baseline`	ADC	Baseline of the raw waveform,
	`bl_mean`	ADC	Actual mean value of the baseline, after baseline subtraction
	`wf_min`	ADC	Minimum amplitude of the waveform
	`wf_max`	ADC	Maximum amplitude of the waveform
	`dt_eff`	ns	Effective drift time
	T_{rec} { `trapEmax/` `cuspEmax/` `zacEmax`	ADC	Uncalibrated energy estimators
Hit	`AoE_Corrected`	—	A/E value after corrections
	`AoE_Classifier`	—	A/E classifier
	`AoE_Double_Sided_Cut`	—	Boolean parameter summarizing PSD cut results
	E_{rec} { `trapEmax_ctc_cal/` `cuspEmax_ctc_cal/` `zacEmax_ctc_cal`	keV	Calibrated and charge-trapping-corrected energy estimators

TABLE 4.1: List of the observables employed in this analysis, grouped by tier.

4.1.1 Energy reconstructions and corrections

Among the parameters provided by the data production, the energy estimators deserve a special comment. From Tab. 4.1 it can be seen that multiple parameters can be employed as energy estimators, both uncalibrated and calibrated. This is because various methods can be employed to reconstruct the energy of a signal, given its waveform.

In principle, if the waveform were a perfect step function, the energy of the signal would be proportional to the amplitude of the waveform. In practice, since the waveform is *not* a perfect step function, the value extracted with this method is not precise, since it is extremely sensible to any fluctuation at the start and at the end of the rising edge, and to systematic discrepancies of the real waveform from the ideal one, such as pole-zero effects. To overcome these issues and provide a precise and accurate value for the energy of the signals, more complex filters have been thus developed, all based on the idea of modifying in a known, controlled and functional way the raw waveform, in order to obtain a shape still containing all the information of the original waveform, but from which the energy can be extracted with higher precision. The suitability of each filter depends on the type and level of noise affecting the waveforms.

Let us underline that the energy reconstruction is performed before calibration: the filters are applied to the raw waveform in order to extract $\texttt{T}_{\texttt{rec}}$, which is then calibrated and charge-trapping corrected to obtain $\texttt{E}_{\texttt{rec}}$. The three calibrated energy estimators presented in Tab. 4.1 are therefore the result of performing the calibration of each uncalibrated energy estimator.

In LEGEND commissioning phase, the considered energy reconstruction methods are the following [62, 63]:

- **TRAP filter**

  The shaping is performed with a trapezoidal filter considering two moving windows with a delay between the two, and the energy is extracted by integration of the trapezoid.

- **CUSP filter**

  The shaping is performed employing an exponentially rising section and a symmetrical falling section separated by the flat section, convolved with the exponential decay added by the electronics to the waveform.

- **ZAC filter**

  The ZAC filter, standing for Zero Area Cusp filter, is an optimization of the CUSP: the waveform shaped with the CUSP method is convolved with two parabolas to obtain a total area equal to zero.

The criterion we employ to determine which energy reconstruction method is more suitable for the acquired data is to select the one that optimize the energy resolution at $Q_{\beta\beta}$ (Section 4.2.1).

An additional factor that affects the energy resolution is the occurrence of charge trapping, namely the fact that some of the charge produced in the interaction of the ionizing radiation with the detector may not be collected, at least within the acquisition time window. This effect is particularly relevant when the energy is deposited at the top of the detector: long drift times make charge trapping more likely. To face this problem, the charge trapping correction (labeled as `ctc` in the name of the parameters) is introduced in the energy reconstruction procedure, for all the employed filter.

4.2 Energy calibration

In order to obtain a relation between the uncalibrated energy estimator ($\texttt{T}_{\texttt{rec}}$, in ADC) and the corresponding physical energy of a signal ($\texttt{E}_{\texttt{rec}}$, in keV), the energy calibration is needed. Let us underline that calibration is thus also a necessary step to pass from the *dsp* to the *hit* tier.

The methodology of the calibration of LEGEND follows the steps developed in the context of the GERDA experiment, widely presented and discussed in [64]. Note that the calibration is performed individually for each detector and for each possible energy reconstruction algorithm.

To perform the calibration, the source is located in front of the detectors and data are acquired: of course, the energy of the signals coming from the source is expressed in ADC, since it is an output of the digitizer. The uncalibrated energy spectrum is thus built, its peaks are identified and put in correspondence with their expected energy value, which is known from literature, and their distribution is fitted with the calibration function, which provides the relation of $\texttt{T}_{\texttt{rec}}$ with $\texttt{E}_{\texttt{rec}}$.

For the calibration of LEGEND Germanium detectors in the commissioning phase ^{228}Th is employed as source. The decay chain of ^{228}Th is shown in Fig. 4.2. The peaks typically employed for the construction of the calibration curve are reported in Tab. 4.2.

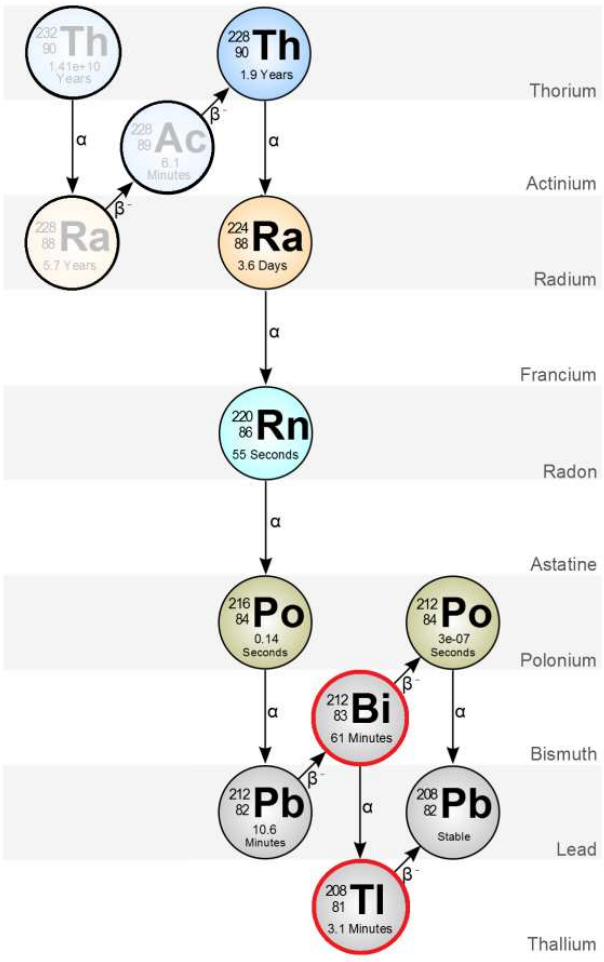


FIGURE 4.2: ^{228}Th decay chain. In red the main sources of gammas employed in the calibration. Picture adapted from [65].

Isotope	Energy [keV]
^{208}Tl	583
^{212}Bi	727
^{208}Tl	860
^{208}Tl	1592 (DEP)
^{212}Bi	1620
^{208}Tl	2103 (SEP)
^{208}Tl	2614 (FEP)

TABLE 4.2: Energy values of the peaks employed in the calibration of LEGEND ICPC detectors in the commissioning phase. In particular, the three ^{208}Tl lines tagged as FEP, SEP and DEP are the Full energy peak, Single Escape Peak and Double escape peak, respectively. The FEP comes from the events in which all the energy is absorbed in the same detector; being this energy above 1021 keV = $2m_e$, pair production can occur: the SEP from the events in which one annihilation photon leaves the detector and the DEP from the events in which both the annihilation photons leave the detector: the distance from the FEP is then exactly 511 keV for the SEP and 1021 keV for the DEP. As discussed in [64], not all these peaks appearing in the calibration spectrum are actually employed for the calibration of each detector: the selection is based on analytical conditions such parameters as the signal-to-noise ratio and the energy resolution (FWHM) of the peak.

The identified peaks in the uncalibrated spectrum are fitted with a suitable function, which takes into account the gaussian shape of the peak, the presence of background and eventual tails of other overlapping distributions.

In particular [64], a gaussian (Eq. (4.1)) is employed to model the signal peak,

$$g(E) = \frac{n}{\sqrt{2\pi}\sigma}\big[-\frac{(E-\mu)^2}{2\sigma^2}\big] \tag{4.1}$$

in which n is the amplitude of the peak, μ is its mean and σ its width; a linear function (Eq. (4.2)) to model the background,

$$f_{lin} = a + bE \tag{4.2}$$

in which a is the intercept and b the slope; an additional step function (Eq. (4.3)) to model the flat background appearing at low energy due to multiple Compton scattering in case of high statistics peaks,

$$f_{step}(E) = \frac{d}{2}\mathrm{erfc}\big(\frac{E-\mu}{\sqrt{2}\sigma}\big) \tag{4.3}$$

in which d is the height of the step function and erfc the complementary error function; finally the function in (4.4) to model the eventual low-energy tail due to incomplete charge collection and residual pile-up events,

$$h(E) = \frac{c}{2\beta}\exp\big(\frac{E-\mu}{\beta} + \frac{\sigma^2}{2\beta^2}\big)\mathrm{erfc}\big(\frac{E-\mu}{\sqrt{2}\sigma} + \frac{\sigma}{\sqrt{2}\beta}\big) \tag{4.4}$$

in which β is the height of the tail, c its slope and μ, σ the parameters from the gaussian.

Each of the peak positions determined by the fit can be associated uniquely to the corresponding expected energy in keV (Tab. 4.2). The positions of the identified peaks in the uncalibrated energy estimator spectrum (in ADC) are plotted against their physical energies (in keV) and the distribution is fitted with a function that will have the role of calibration curve, namely will provide the desired analytical relation between $\texttt{T}_{\texttt{rec}}$ and $\texttt{E}_{\texttt{rec}}$. A linear function is employed for this purpose.

The result of the calibration on the spectra acquired with the source for runs 025, 026 and 027 is shown in Fig. 4.3. Since the peaks are perfectly overlapped, it is clear that, after the calibration, data from the three runs are fully compatible and can be treated as a unique sample. Fig. 4.4 is a zoom of the spectrum appearing in Fig. 4.3 and shows in detail the ^{208}Tl SEP (Single Escape Peak): the

peaks in the three calibration runs are perfectly aligned with each other and with the expected value of the γ line. Also, the statistics in the three dataset is comparable, and this is consistent with the fact that the calibration phase in the three runs had a comparable duration (Tab. 3.1).

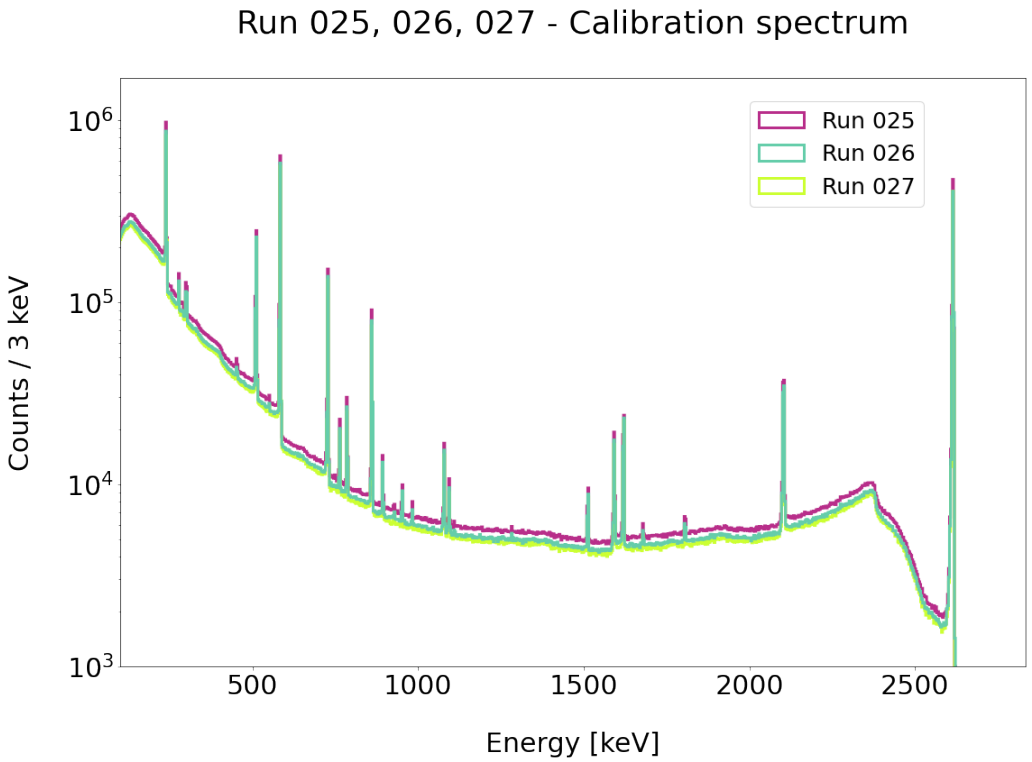


FIGURE 4.3: Calibrated energy spectrum of the signals acquired with the ^{228}Th source (*cal* phase) for the three runs of interest. The considered energy estimator is `trapEmax_ctc_cal`. Data from all the detectors are plotted together, despite the calibration being performed for each detector individually, precisely to check the compatibility of data from different detectors after the calibration. The γ lines from the three distribution are mostly overlapped due to the effectiveness of the calibration procedure and to the similar statistics of the datasets.

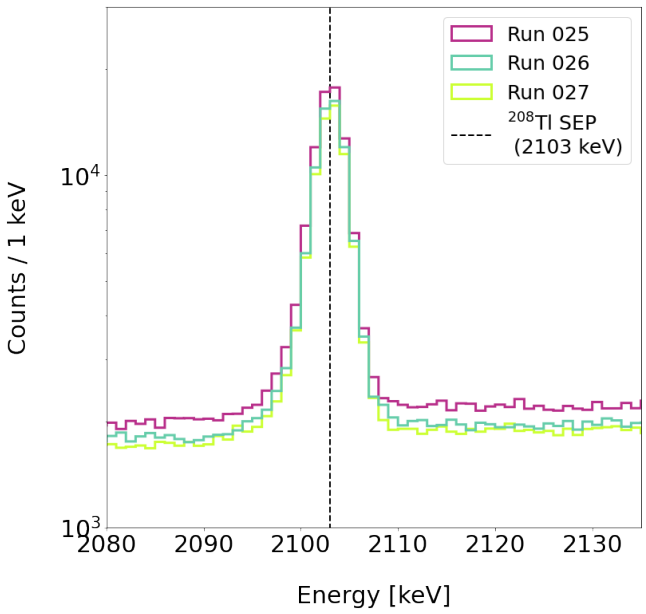


FIGURE 4.4: Zoom on the ^{208}Tl SEP from the calibrated energy spectrum. The three distributions are nicely overlapped and centered on the expected value of SEP = FEP - m_e = 2614 keV - 511 keV = 2103 keV [66].

4.2.1 Energy resolution and energy reconstruction choice

The energy calibration is performed individually for each detector and for each possible energy reconstruction algorithm. This means that a set of single-detector, single-energy-reconstruction spectra is available and can be employed for the evaluation of the energy resolution achieved by each detector with the different energy reconstruction methods.

In particular, from the fit (gaussian + background corrections) of the peaks appearing in the calibrated energy spectrum collected with the ^{228}Th source by one detector, it is possible to extract the energy resolution of that detector at the peaks' energies.

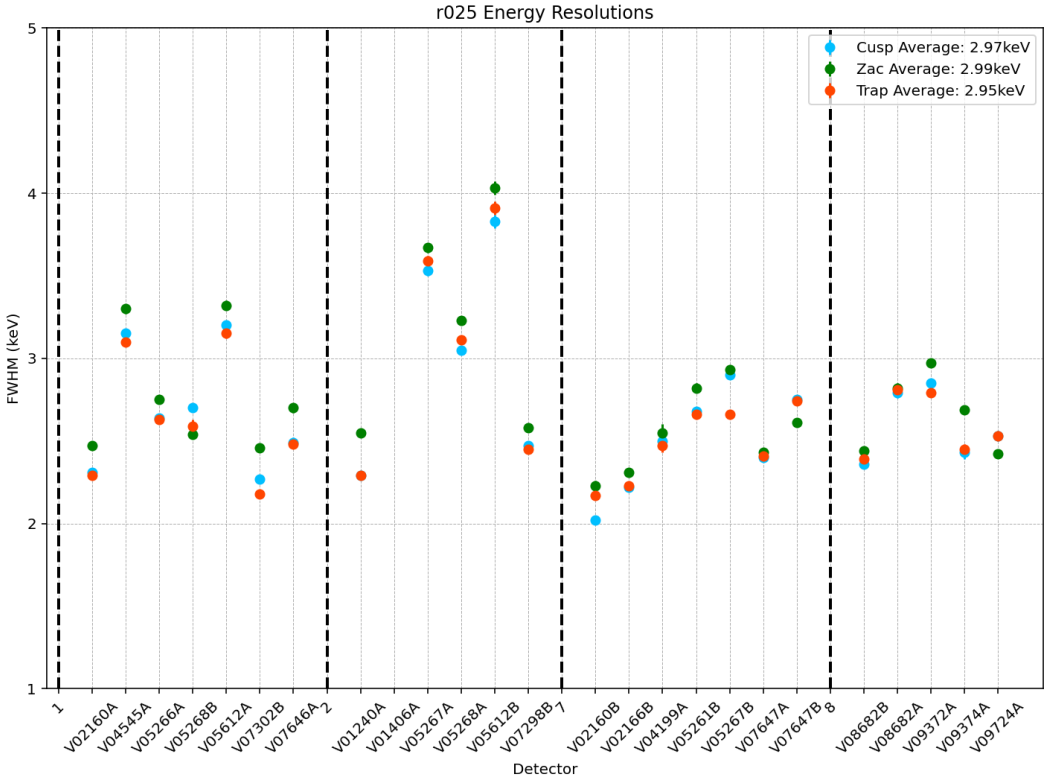
These energy resolutions are then plotted against the energies at which they were calculated and the distribution is fitted with the function

$$\sigma(E) = \sqrt{a + bE} \tag{4.5}$$

with a, b fit parameters and E the energy. In particular, a accounts for the contributions from electronic noise, while b accounts for statistical fluctuations in the number of charge carriers [64], which are the main parameters on which the energy resolution of a Germanium detector depends [67]. The energy resolution at $Q_{\beta\beta}$ can be extrapolated simply evaluating the function at $Q_{\beta\beta} = 2039\,\text{keV}$. The results of this procedure for the runs of interest, performed by the Analysis Group of the Collaboration for all the detectors deployed in LEGEND commissioning phase, for the three considered energy reconstruction methods, are shown in Fig. 4.5.

As anticipated, the mean FWHM resolution at $Q_{\beta\beta}$ of the ICPC Germanium detectors is around 3 keV, with individual values ranging between 2 keV and 4 keV. The values in the three runs appear highly stable in time, with some detectors having a systematically higher (or lower) resolution.

The three energy reconstruction methods appear consistent among each other, with discrepancies in the results of $\mathcal{O}(0.1\,\text{keV})$. The TRAP filter, though, appear to systematically provide a slightly better overall resolution, as confirmed by the values averaged on all the detectors, reported in the plots. For this reason, the analysis presented in this work is performed using as energy estimator the value obtained employing the TRAP filter, namely `trapEmax_ctc_cal`.



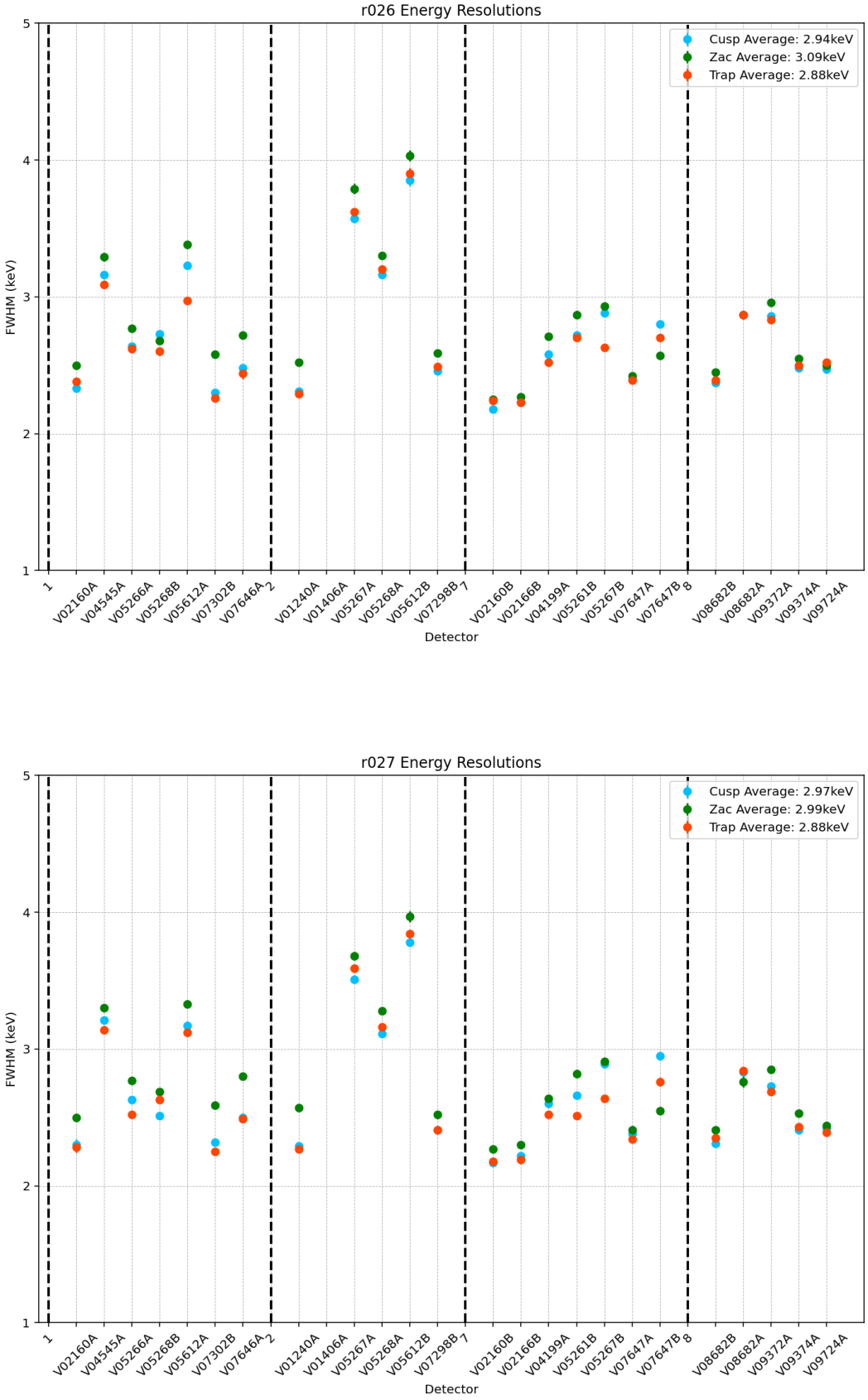


FIGURE 4.5: Energy resolution, in terms of FWHM, of the active Germanium detectors at $Q_{\beta\beta}$ during runs 025 (top), 026 (center) and 027 (bottom). The value of the FWHM at $Q_{\beta\beta} = 2039\,\text{keV}$ is extrapolated from the energy resolution curve (Eq. (4.5)), obtained fitting the width of the calibration peaks, as discussed in [64]. Plots from [58].

4.3 The raw acquired spectrum

After the energy calibration, data from the different runs and from different detectors can be analyzed together. Fig. 4.6 shows the raw energy spectrum built gathering all the signals collected by the Germanium detectors during runs 025, 026, 027.

Tab. 4.3 collects the total number of signals in the raw energy spectrum. Two energy regions are considered: the full energy range of the acquired signals, spanning from 0 keV to 10000 keV, and the Region Of Interest (ROI) for $0\nu\beta\beta$, around $Q_{\beta\beta} = 2039\,\text{keV}$. In particular, the energy range selected as ROI for this thesis work is the same considered in GERDA's analysis [32], namely

$$\text{ROI} = [1930, 2190]\ \text{keV} \tag{4.6}$$

with the exclusion of the intervals (2104 ± 5) keV and (2119 ± 5) keV, containing two known background peaks, from ^{208}Tl and ^{214}Bi, respectively [68].

The purpose of the following analysis is to isolate, among these signals, the good $0\nu\beta\beta$ candidates by removing from the spectrum noise and signals representing background with respect to the physical process of interest.

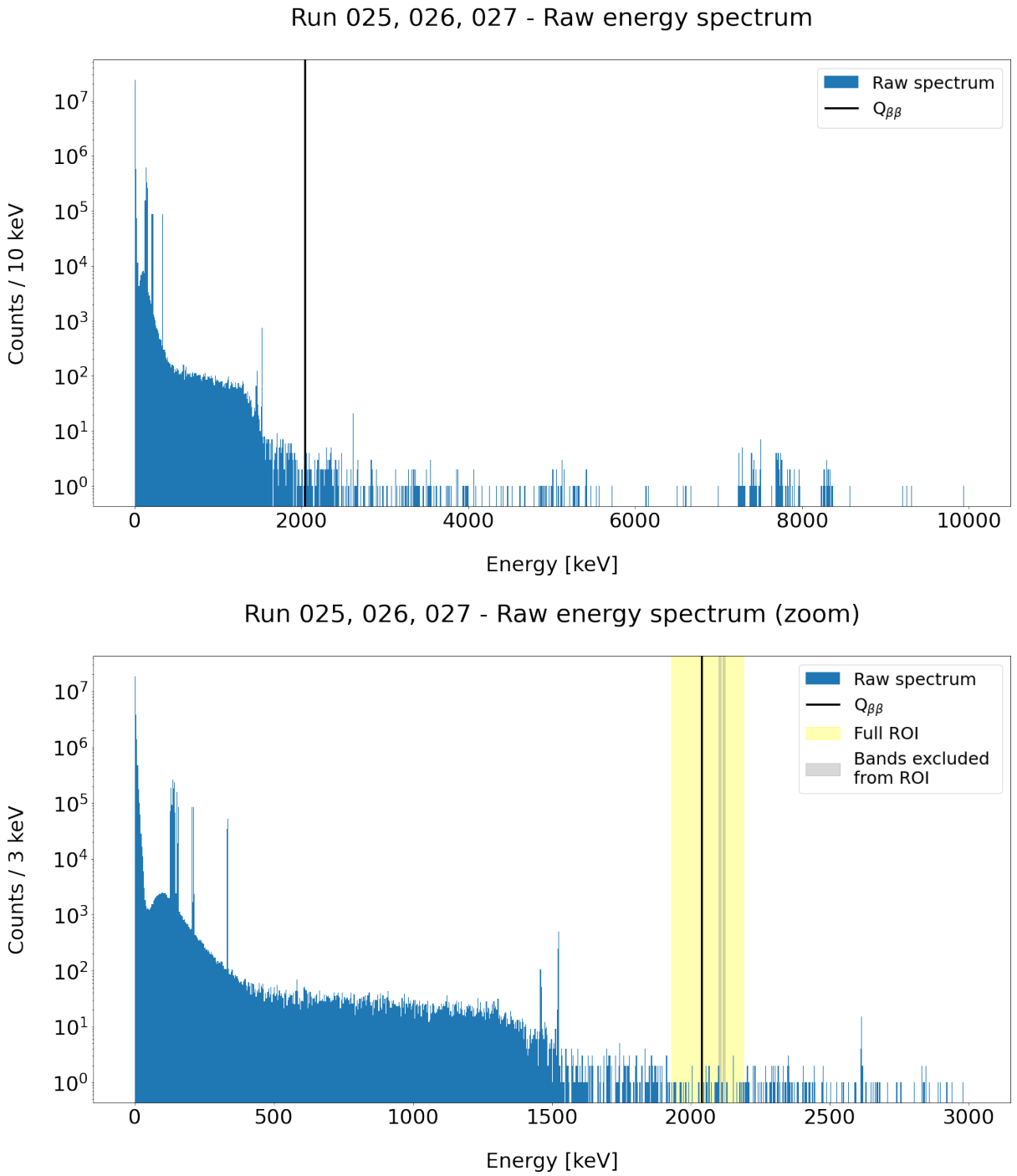


FIGURE 4.6: Raw energy spectrum of signals from run 025, 026 and 027. TOP: full energy range; BOTTOM: zoom in the energy range where most of the structures of interest are concentrated.

	Total number of raw signals
Full energy range	29 738 052
ROI	38

TABLE 4.3: Total number of raw signals acquired by Germanium detectors during runs 025, 026, 027. Two energy regions are considered: the full energy range of the spectrum and the ROI around $Q_{\beta\beta}$ for the $0\nu\beta\beta$ analysis, spanning from 1930 keV to 2190 keV, with the exclusion of the 10 keV wide regions (2104 ± 5) keV and (2119 ± 5) keV from ^{208}Tl and ^{214}Bi peak, respectively.

Chapter 5

Analysis cuts

The main goal of the analysis presented in this work is to remove noise and background signals from the energy spectrum collected by Germanium detectors, in order to retain only good $0\nu\beta\beta$ candidates.

This goal is achieved by implementing a series of analysis cuts, discussed in the next sections: Pulser Signal Removal (Section 5.1), Quality Cuts (Section 5.2), Muon veto (Section 5.3), multiplicity cut (Section 5.4), LAr veto (Section 5.5) and Pulse Shape Discrimination (Section 5.6).

The cuts are developed to maximize the acceptance of $\beta\beta$ signals while minimizing the noise an background acceptance. All the analysis is performed in Python, with a fully original code developed for this work. For the parameters estimate we employ the results of `pygama` data processing.

At the end of each section the results of the corresponding cut are presented, comparing the energy spectrum obtained after the considered cut (orange) with the energy spectrum before that cut (blue). Note that at each step we consider as starting point the sample (and therefore the spectrum) obtained as output of the previous cut, and *not* the original raw spectrum.

With each cut, a sample of data is removed by the main spectrum: the spectra of these discarded events are collected in Appendix A and referred to as *residual spectra*.

5.1 Pulser Signals Removal(PSR)

In this Section, we will start the analysis of the physics data collected during LEGEND-200 commissioning phase, and in particular during run 025, 026 and 027.

As anticipated in Chapter 2, a pulse generator is introduced in the acquisition chain in order to have a known signal periodically appearing among the collected data, which allows to check the stability of the acquisition and of the performances of the detectors.

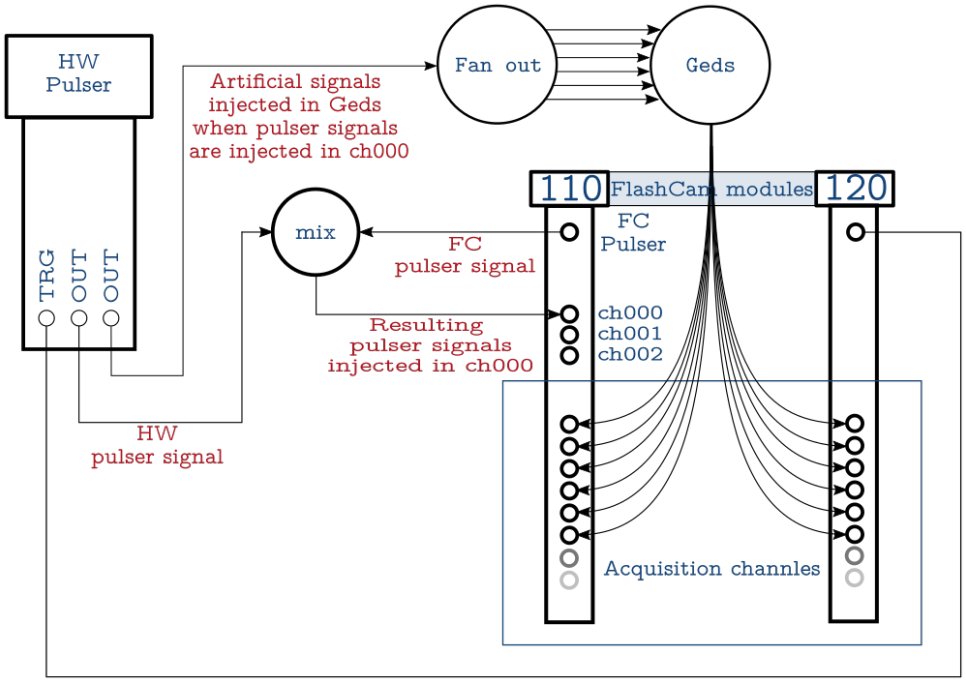


FIGURE 5.1: Sketch of the electronics connections of the external and internal pulse generators to the FlashCam modules and to the Germanium detectors array, in the LEGEND-200 commissioning phase setup. In blue the main elements of the chain, in red the signals. The number of channels represented in the FlashCam is merely indicative, to visualize the connections.

5.1.1 Identification of pulser signals

During the commissioning phase, the ch000 of the digitizer (FlashCam module 110) is connected both to a PB-5 external pulser (*HW pulser*) from BNC (Berkeley Nucleonics Corporation) and to the internal pulser channel on the FlashCam module itself (*FC pulser*), as shown in Fig. 5.1 [69, 70].

- **FlashCam pulser signals (FC)**

  These signals are produced by the pulser channel internal to the FlashCam, and injected to the ch000 of the FlashCam itself.

  The signals are squared waves injected with the frequency of $f_{FC} = 0.1\,\text{Hz}$. They have a height of $\approx 5000\,\text{ADC}$ and a width of 4000 ns with $\approx 10\,\text{ns}$ of rise time and fall time (Fig. 5.2 and 5.3).

- **HardWare pulser signals (HW)**

  These signals are produced by an external PB-5 BNC pulser module and injected in the ch000 of the FlashCam.

  The signals are square waves injected with a frequency $f_{HW} = 0.05\,\text{Hz}$. They are much narrower and higher with respect to the FC signals, having a height of $\approx 13\,500\,\text{ADC}$ and a width of 150 ns, with $\approx 10\,\text{ns}$ of rise time and fall time (Fig. 5.2 and 5.3).

With this input structure, we expect ch000 to record and acquire three types of signals, which will then be found in the analyzed database:

1. **Non-pulser events**

   The ch000 is not connected to a Germanium detector, so it cannot record and acquire physical signals on its own. Despite this, ch000 is connected to the global trigger of the array: therefore every time a Germanium detector reveals a signal and the global trigger is fired, also ch000 acquires data, simultaneously to all the other channels. These signals, collected by ch000 just because a Germanium channel triggered the acquisition, will mostly be flat, laying at the value of the baseline (namely at zero after baseline subtraction), as shown in Fig. 5.3 (top left). These flat signals are not related to or influenced by the pulser inputs.

2. **Pure FC pulser signals**

   When the FC pulser signal *is not* in coincidence with a HW pulser signal, once every 20 s seconds, the ch000 records only the FC flag, as shown in Fig. 5.2 (blue) and 5.3 (top right).

3. **FC + HW pulser signals**

   When the FC pulser signal *is* in coincidence with a HW pulser signal, once every 20 s, the ch000 records a signal which is the sum of the two injected waveforms, as shown in Fig. 5.2 (red) and 5.3 (center). The amplitude of the waveforms of these signals will therefore correspond to the sum of the amplitudes of FC and HW pulses: $\texttt{wf\_max} = \texttt{wf\_max}_{\mathrm{FC}} + \texttt{wf\_max}_{\mathrm{HW}}$.

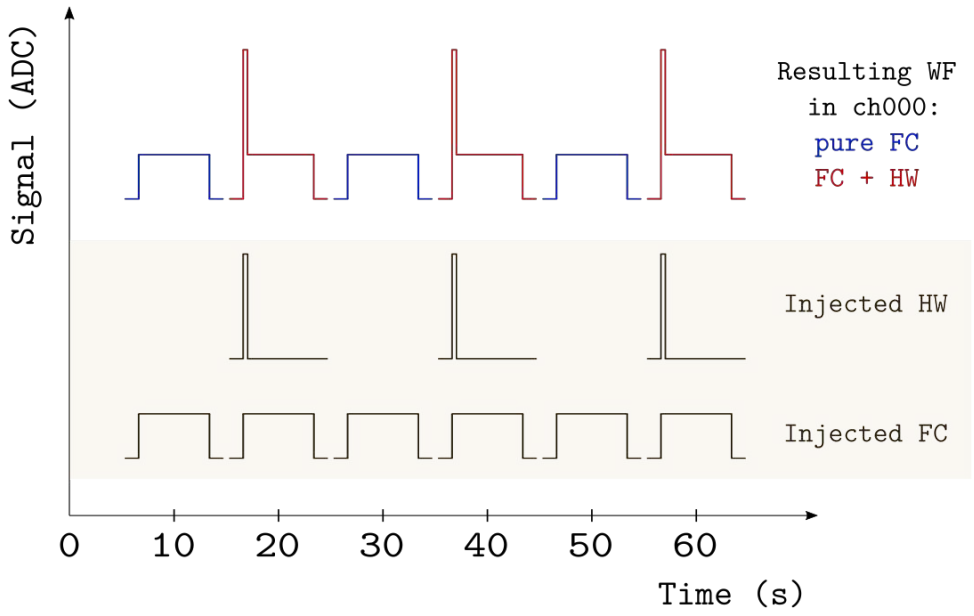


FIGURE 5.2: A schematic representation of the pulses produced by the pulse generators and injected in ch000, and of their interaction. In the two lower bands are represented the injected pulses (*FC pulser signal* and *HW pulser signal*, see Fig. 5.1); in the upper band the waveforms actually appearing in ch000, as a result of the interaction of the injected pulses. Pure FC flags (in blue) trigger the acquisition of the full array; no artificial signals, though, are injected in the Germanium detectors, therefore the acquired data are not influenced by the presence of the pulser. When FC+HW signals (in red) are acquired in ch000, instead, an artificial signal is injected in all the Germanium channels, at the jFET level, in order to monitor the performances of the DAQ; the signals acquired by the Germanium detectors simultaneously to a FC+HW pulse in ch000 are therefore non-physical and must be removed before proceeding with the analysis of the energy spectrum. Note: signal duration not to scale.

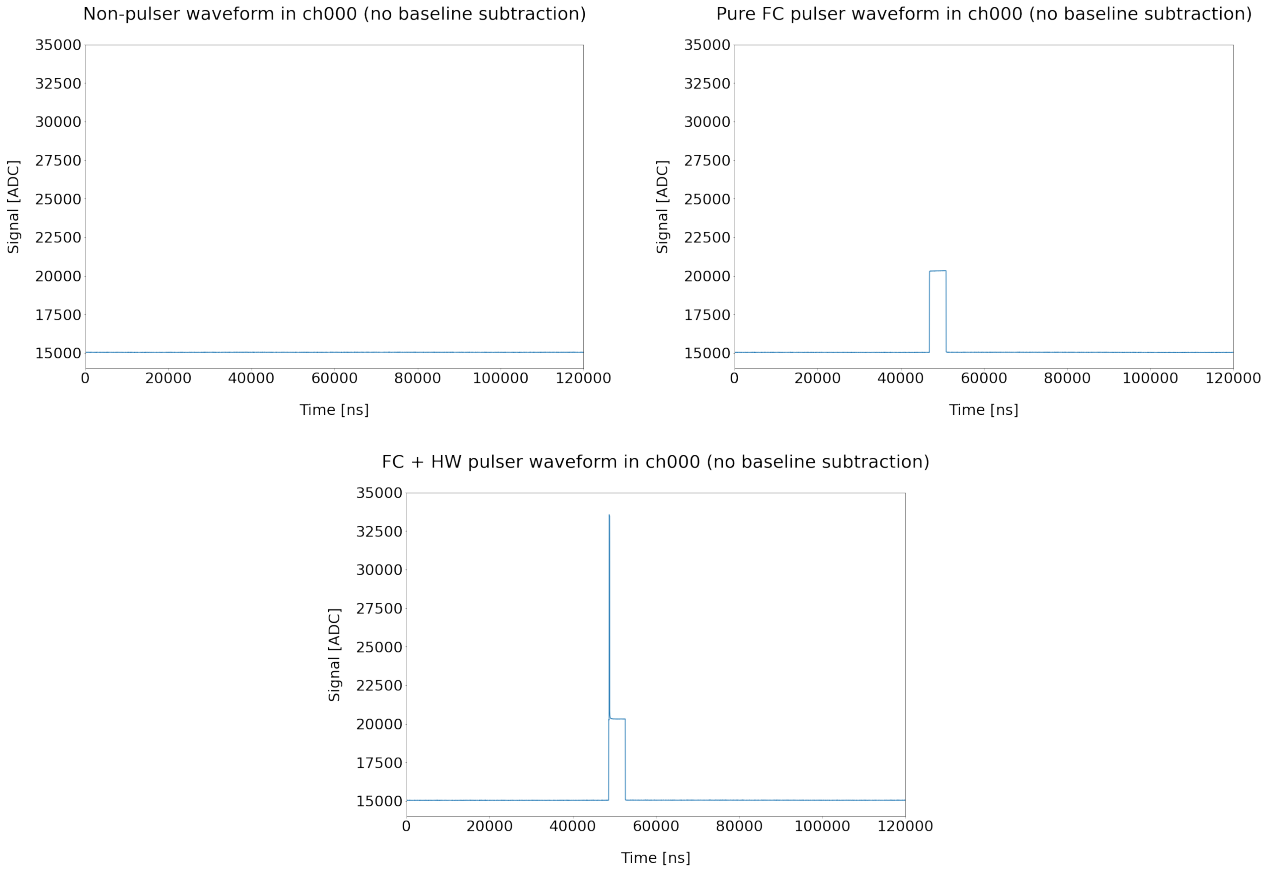


FIGURE 5.3: Experimental waveforms collected by ch000. TOP LEFT: example of non-pulser waveform; TOP RIGHT: example of pure FC pulser waveform; BOTTOM: example of FC+HW pulser waveform.

The presence of these three types of signals leads to a three-bands structure in the energy spectrum from ch000 and in the distribution of the maximum of the waveforms acquired in that channel. In practice, we study the maximum of the waveforms (`wf_max`) acquired over time, instead of the energy, so as to avoid any possible artifact related to the chosen energy reconstruction method.

Moreover, this analysis is performed without subtracting the baseline from the amplitude of the signals (namely we considered the raw `wf_max`), in order to avoid the introduction of an additional parameter, since not strictly necessary in this phase. In facts, what we are interested in is the band structure, and in particular the possibility to isolate each band from the others: the rigid translation brought by the baseline subtraction does not modify the results in this phase.[1]

We expect therefore three well defined and isolated bands in the distribution of the maximum of the waveform acquired over time in ch000, corresponding to the three types of resulting pulser signals stated above. The experimental data from runs 025, 026, 027 are plotted in Fig. 5.4: their distribution is consistent with the expectations.

[1]Despite this, we tried to subtract the baseline from all the data, to check that there are no undesired offsets: as expected the lower band is rigidly translated towards zero, confirming the acquisition of zero-energy flat signals. The two other bands are correspondingly lowered.

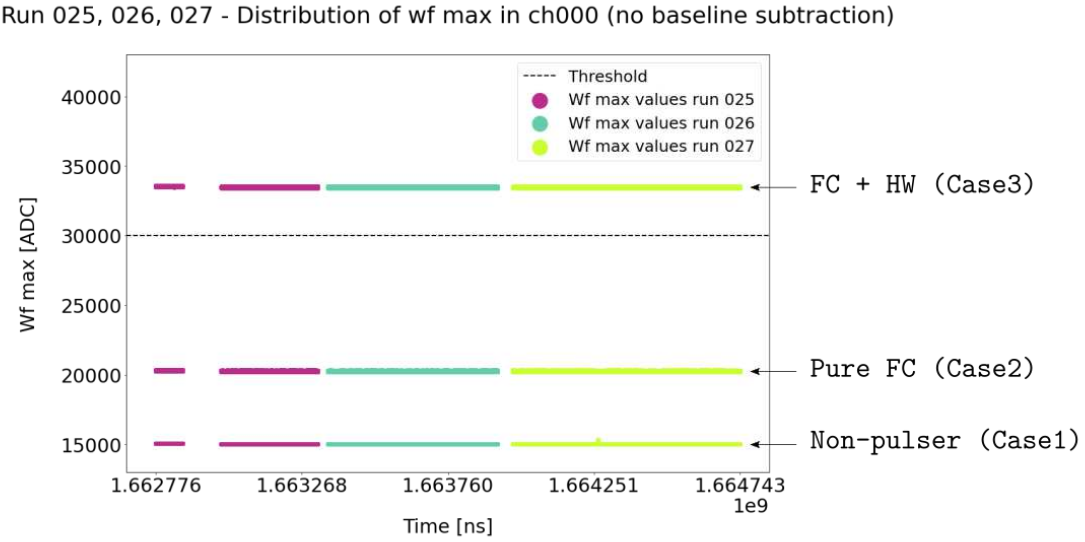


FIGURE 5.4: Distribution of the parameter `wf_max` (maximum of the waveform) over time, in data acquired by ch000 during runs 025, 026, 027. The three-band structure is the outcome of the three different types of signals that can be acquired in ch000 with the given input and trigger configuration.

5.1.2 Consequences of pulser signals on data acquired by Germanium detectors

The question now is: what do all the other Germanium channels see in these three cases? Do they see physical signals, or just artificial signals produced as a consequence of the pulser injection?

In Case 1 and 2 (non-pulser and pure FC), the signals in all the other channels of the array are not influenced by what happens in ch000; in Case 3 (FC+HW), instead, an artificial signal is injected in all the Germanium channels of the array:

1. In case of non-pulser signals in ch000, it is ch000 that acquires data as a consequence of the global trigger. The signals acquired by the Germanium detectors are not affected by the ch000 acquisition.

2. When a pure FC signal is acquired in ch000, it fires the global trigger: all the Germanium detectors acquire data simultaneously and since the acquisition is not due to the presence of a physical signal in one of the Ge channels, but rather to the ch000 signals, the signals acquired by the Germanium detectors will mostly be flat.

   In this case, then, ch000 triggers the acquisition but does not influence what the Germanium detectors see at the moment of the acquisition.

   The flat signals acquired by the Ge detectors are employed to give an estimate of the baseline of the acquisition system.

3. When a FC+HW pulser signal appears in the ch000 of the FlashCam, an artificial signal is produced and injected in all the Germanium channels of the array.

   As shown in Fig. 5.1, this artificial signal is generated by the external pulser (OUT connection), duplicated via a fan-out and directed towards all the Germanium detectors. The detectors are then read out by the FlashCam itself, in channels [002-043] of module 110 and module 120 [51].

   The artificial signal is injected in the Germanium channels in order to have a known signal appearing with a known frequency in the Germanium detectors acquired data: if everything in the acquisition chain works properly, this signal is also read out and saved in the database with the same rate. The injection of mock-physical signals in the Germanium channels serves therefore in the monitoring of the DAQ performances and status.

The signals injected in the Germanium channels simultaneously to a FC+HW pulser signal in ch000 will be simply referred to as *pulser signals* and will be further analyzed in the next sections.

5.1.3 Pulser signals in Germanium channels

When a FC+HW signal is recorded in ch000, another signal, of fixed amplitude, is sent to the Germanium detectors. As anticipated, this secondary signal, generated by the external pulser, passes through an analogue fan out that duplicates it: each copy of the original pulser signal is then sent to a specific Germanium detector (Fig. 5.1).

The pulser signal is actually injected at the level of the jFET, which is the component of the front-end electronics where the charge (coming from the Germanium when dealing with a physical signal or, in this case, constituting the artificial signal) is collected before being sent to the readout electronics. With this configuration, the pulser signal injection allows to check the working status of all the acquisition chain, from the jFET to the data storing.

Fig. 5.5 shows a sample of pulser signals acquired during run 025, color-coded by the channel in which they are acquired. Data from run 026 and 027 show exactly the same behavior.

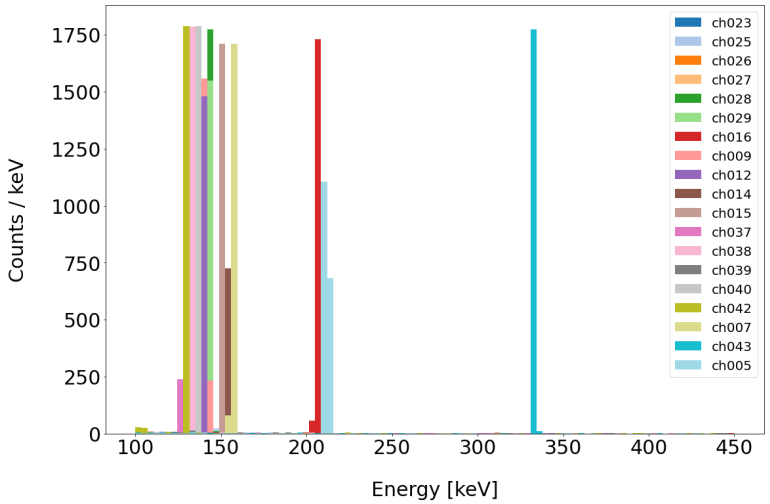


FIGURE 5.5: Sample of data from the pulser peaks from run 025 color coded by acquisition channel. It is evident from the plot that most of the pulser peaks lay in a rather narrow energy range, as expected considering that the original pulser signal (before fan out) is only one, with a fixed amplitude. In three of the detectors, instead, ch016, ch005 and ch043 the pulser peak has a completely different energy, highlighting some unexpected process at some point of their readout chain (further discussion in main text).

For most of the channels, the pulser peaks are confined in a quite narrow energy range, despite not being all identical. This result is consistent with the expectations. The signal generated in response to the FC+HW pulser is only one and has a well defined energy, justifying thus the concentration of the peaks in a small portion of the possible energies. The spread of the peak positions, instead, finds an explanation in the structure of the electronic chain that the signal, once generated, must pass through before being acquired.

In particular [71]:

- Each channel has a different amplifier, and they can have slightly different gain;
- The pulser is distributed by an analog fanout which may also have slightly different gain;
- The pulser signal is coupled directly at the front end board (jFET) with a 0.25 pF capacitor which can have up to 20-30% different size for intrinsic difficulties in building such small components.

In three of the channels, instead, the acquired pulser signal has a completely different energy (Fig. 5.5. In particular, there is a cluster at ≈210 keV (ch005, ch016), and a single peak at ≈340 keV (ch043).

The deviance of the energy of these peaks from the main cluster cannot be explained by the causes stated above. In facts, no conclusive explanation for this behavior has been found yet.

We can still proceed, though, since the absolute value of the pulser peak position, as long as it stays constant over time, does not affect the information that can be extracted from the data. Note that if, instead, the position of the pulser peaks drifted over time, it would be the symptom of a severe instability of the acquisition chain, chich would compromise the usability of the acquired data. In fact, if the energy of the pulser peak changes in time in a certain channel, it means that the gain of the amplifier of that channel is changing: also the energy of the acquired physical signals would be meaningless since drifting in time.

After exploiting the pulser signals to monitor the performances of the DAQ, we go on with the analysis of the Germanium energy spectrum.

5.1.4 PSR implementation and results

The pulser signals, artificially injected in the Germanium channels when a FC+HW pulser signal appears in ch000, must be removed from the energy spectrum before performing any other analysis, since they do not have any physical origin.

This type of event has a precise signature: the signal acquired by a Germanium detector, whose waveform resembles a perfectly shaped physical signal, is simultaneous to a signal having a fixed and known amplitude (`wf_max`) in ch000, corresponding to the amplitude of the FC+HW signal, namely belonging to the upper band of Fig. 5.4.

It is therefore possible to remove these artificial signals from the Germanium energy spectrum by cutting away all the signals collected simultaneously to a signal having `wf_max`$>$ *threshold* in ch000.

From Fig. 5.4 it is clear that the pulser response has been stable during the three runs: we can therefore establish a common threshold. The threshold in ch000 chosen to safely remove all the signals from the upper band is $threshold = 30\,000$ ADC. All the signals collected by a Germanium detector simultaneously to a signal having `wf_max` $> 30\,000$ ADC in ch000 are identified as *pulser signals* and removed from the spectrum.

In Fig. 5.6 the resulting spectrum is shown, superimposed to the original raw spectrum. The three pulser peaks are nicely removed by the cut. Few other signals, at different energies, are also removed: they could be signals coming from physical events occurring simultaneously to the pulser signal injection, or noise signals (such as discharges) simultaneous to the pulser signals, for which the reconstructed energy value is meaningless.

The rate of collected pulser signals gives information about the acquisition status and stability of the system. We know the rate of FC+HW pulser signals injection, and we know that everytime a FC+HW pulser signal appears in ch000, a pulser signal is injected simultaneously in all the Germanium channels. Therefore, in optimal acquisition conditions, all the channels acquire exactly the same rate of pulser signals. This actually happens in the dataset under analysis: all the channels acquire exactly the same number of pulser signals ($N_{pulser} = 86345$) during the total acquisition time of runs 025, 026 and 027 (see Table 5.3, 5.4 and 5.5 for single run results presented channel by channel).

Tab. 5.1 shows the number of events surviving to pulser signal removal. Note that in ROI the surviving fraction is 100%: this is consistent with the expectations, since the energy of the artificial signal injected by the pulser is chosen to be well below the ROI, in order to avoid any impact in that region.

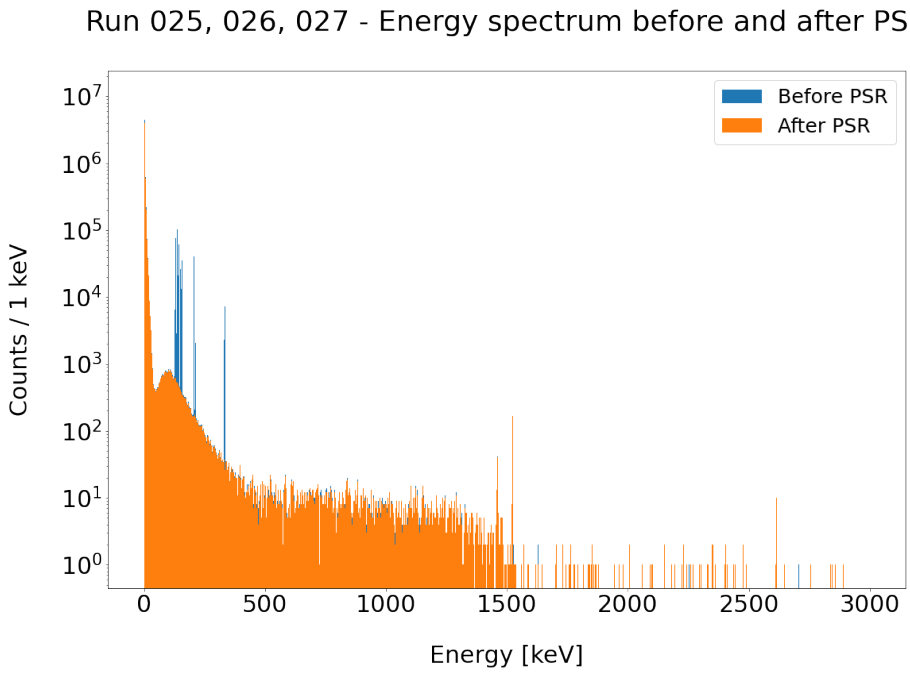


FIGURE 5.6: Energy spectrum before and after pulser signal removal: signals from all Germanium detectors, from run 025, 026 and 027 (zoom).

	Number of signals before PSR	Number of signals after PSR	Surviving fraction (%)
Full energy range	29 738 052	27 253 386	91.64 %
ROI	38	38	100%

TABLE 5.1: Number of signals before and after pulser signals removal. The surviving fraction is calculated as ratio between the two numbers, so it is referred only to the present cut. Two energy regions are considered: the full energy range of the spectrum and the ROI around $Q_{\beta\beta}$ for the $0\nu\beta\beta$ analysis.

5.2 Quality cuts (QC)

Now that the artificial pulser signals have been removed from the energy spectrum, we can proceed with the analysis of the physical signals collected by the Germanium detectors.

The aim of this Section is to discriminate the physical events from noise and to remove from the energy spectrum the events identified as noise. The discrimination is performed on the basis of the shape of the waveforms of the collected signals [72].

The waveforms of physical signals have a well defined shape (Fig. 5.7), including a flat baseline, a fast positive-sloped rising edge ($\mathcal{O}(100$ ns)) up to a maximum amplitude, and then a slow exponential fall (τ of $\mathcal{O}(100\,\mu$s)). This enables us to identify waveforms representing physical events based on their shape, and to discriminate them from all the other waveforms collected by the detectors, representing noise, which must be removed before proceeding with the rest of the analysis.

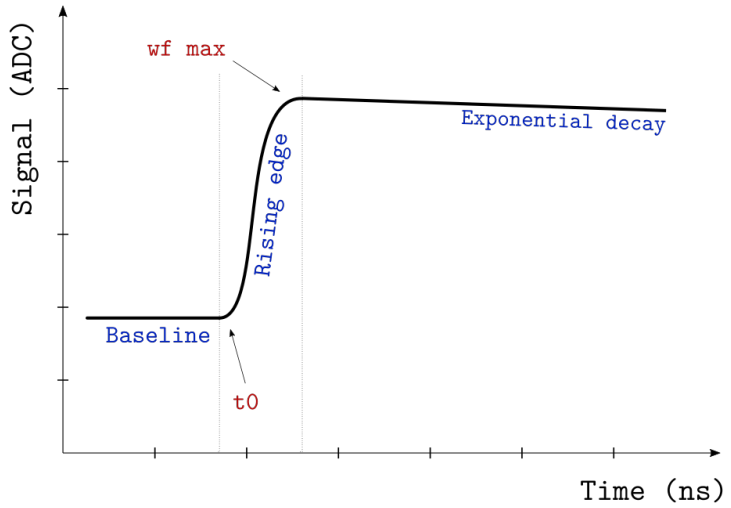


FIGURE 5.7: Schematic representation of the expected shape of a waveform collected from a physical event. Its main features, highlighted in the sketch, will be employed in the following analysis to discriminate physical signals from noise. For each signal, a time window of 130 000 ns is acquired and saved on disk (e.g. Fig. 5.8, 5.9 and 5.10): enough to include, for well-shaped physical signals, some samples from the baseline, all the rising edge and the first part of the exponential decay.

In practice, the shape of the waveforms of the good signals is synthesized in a set of features and these features are expressed as conditions on the main parameters of the waveforms themselves. The discrimination is then performed by checking whether or not the collected waveforms respect these conditions (quality cuts, QC). Each signal is examined individually: if it respects all the conditions, it is classified as a good, physical signal; otherwise, if it does not respect one or more of the conditions, it is classified as noise, and removed from the dataset.

Before presenting the implementation of the QC, just a comment about **noise**. Why don't we have only physical events? Where does noise come from? The noise sources are multiple, but most of them can be traced back to the electronics involved in the system [72]: discharges can lead to spikes and oscillating signals; when the amplitude of the signal is larger than the available range, uninformative saturated traces appear. The purpose of the QC, then, is to remove this noise from the dataset.

Moreover, when the digitization and acquisition is not fast enough with respect to the events' counting rate, **pile-up events** can appear (interfering effects between a pulse and the previous one): also these signals are removed by quality cuts.

Finally, another class of signals that is worth mentioning before performing the cuts is **flat signals**, namely signals which are acquired despite being completely flat ($Amplitude = 0$), apart from statistical fluctuations. There are two reasons why these signals appear among the saved data. For some of

them, the statistical fluctuations might be above the acquisition threshold, so the signal is actually interpreted as an event and acquired. Most of them, though, are acquired because of the trigger structure of the acquisition system. As said, the system works in global trigger mode, meaning that when a trigger is generated from one subsystem (namely one channel), all channels are read out. In practice, this means that when one channel records an event above acquisition threshold, all the channels acquire simultaneously: this mechanism leads to a huge component of flat signals in the database. As a gross estimate (not taking into account the possibility of physical events in which more than one detector is hit) during the LEGEND-200 commissioning phase, with approximately 20 active Germanium detectors, there will be a nearly 20:1 ratio of flat signals to non-flat signals.

In Fig. 5.8, 5.9 and 5.10 (top) some examples of raw waveforms before any cut, from the dataset under analysis are shown: in most of them, the fraction of noise signal is manifestly dominant.

5.2.1 Implemented QC

The parameters and observables considered in the evaluation of the goodness of the signals are presented below. The conditions have been developed starting from [73], and updating them to the current cuts employed by the Analysis Group of the Collaboration for the LEGEND-200 analysis [74, 75].

In particular, the presented conditions identify good, physical signals, and therefore the signals that *do not* respect all these conditions must be rejected. Note that the conditions are to be applied on raw waveforms, before baseline subtraction.

- **Saturation:**
  - $\texttt{wf\_min} > 0\,\text{ADC}$

    The minimum of the waveform must be strictly bigger than zero. Imposing this condition removes signals which are saturated from below, namely which hit the lower bound of possible values that the digitizer can read.
  - $\texttt{wf\_max} < 65\,520\,\text{ADC}$

    The maximum of the waveform must be strictly smaller than 65 520 ADC. This is a conservative condition imposed to remove signals which are saturated from above, namely which hit the upper bound of possible values that the 16-bit digitizer can read (2^{16}).
- **Start of the rising edge ($\mathtt{t_0}$):** $45\,000\,\text{ns} \leq \mathtt{t}_0 \leq 55\,000\,\text{ns}$

  The start of the rising edge must be within 45 000 and 55 000 ns.

  This condition comes from the trigger settings of the acquisition. The global trigger is fired when a channel records a signal above threshold. When this happen, all the channels save the data acquired from 50 000 ns *before* the trigger to 80 000 ns *after* the trigger (thus covering a time window of 130 000 ns). Therefore, in the saved data, the start of the rising edge must be at around 50 000 ns. All the signals not respecting this condition must be removed.
- **Risetime**
  - $\mathtt{t}_{10-90\%} \geq 96\,\text{ns}$

    With $\mathtt{t}_{10-90\%} := \mathtt{t}_{90} - \mathtt{t}_{10}$, the time it takes to the signal to go from 10% to 90% of its amplitude; it must be bigger than 96 ns.
  - $\mathtt{t}_{10-50\%} \geq 32\,\text{ns}$

    With $\mathtt{t}_{10-50\%} := \mathtt{t}_{50} - \mathtt{t}_{10}$, the time it takes to the signal to go from 10% to 50% of its amplitude; it must be bigger than 32 ns.

These two conditions allow to control the duration of the risetime and therefore the shape of the rising edge. In particular, putting a lower bound to the risetime allows to remove spikes, in which the signal rises abruptly and more quickly than the physical signals.

- **Instant of maximum amplitude (`t_max`):** $45\,000\,\text{ns} \leq \texttt{t\_max} \leq 120\,000\,\text{ns}$

  The instant in which the signal reaches its maximum amplitude must be after the start of the trigger time window ($\geq 45\,000\,\text{ns}$), and before the very end of the acquisition time window ($\leq 120\,000\,\text{ns}$).

  The lower bound allows to remove inverted signals, namely signals which have the right shape, but developed in the negative semi-plane with respect to the baseline. These signals, which appear as good signals reflected with respect to the x axis, are originated by cross talk between different channels [76] and must be removed.

  The upper bound allows to remove signals which are not fully contained in the acquired time window, namely signals which reach their maximum amplitude at the very end of the acquired time window and which would probably keep growing if further acquired. These signals must be removed since their energy reconstruction is meaningless.

- **Effective drift time:** $\texttt{dt\_eff} > 0\,\text{ns}$

  The effective drift time must be bigger than 0 ns, so it must be a positive, non-zero, time duration.

  The drift time is defined as $\text{dt} = \texttt{t}_{100} - \texttt{t}_0$, in which t_0 is the time in which the charge is produced in the interaction, and t_{100} the time in which it is collected. The drift time gives an estimate of the time it takes to the produced charge to be collected and is therefore a positive defined quantity.

  The effective drift time gives the same information, and is calculated as $\texttt{dt\_eff} = \frac{\text{Q}_{\text{drift}}}{\text{Energy}}$, in which Q_{drift} is the charge produced in the interaction and drifting toward the anode electrodes of the detector [77].

  The effective drift time must retain the original property of being positive.

- **Baseline mean:** $\texttt{bl\_mean} > -400\,\text{ADC}$

  The parameter `bl_mean` is defined as the mean of the actual baseline of the signal *after* baseline subtraction; it must be bigger than $-400\,\text{ADC}$.

  Since `bl_mean` is calculated after baseline subtraction it should be zero for all signals. In facts, it is often not exactly zero, giving thus an estimate of the correction that must be made offline to the baseline estimation performed online and stored in the database.

Tab. 5.2 synthesized the implemented cuts. Note that a signal is kept in the energy spectrum if it respect *all* the conditions.

Observable	Condition to be a **good** signal
Saturation	$\texttt{wf\_min} > 0\,\text{ADC}$ $\texttt{wf\_max} < 65\,520\,\text{ADC}$
Start of the signal	$45\,000\,\text{ns} \leq \texttt{t}_0 \leq 55\,000\,\text{ns}$
Risetime	$\texttt{t}_{10-90\%} \geq 96\,\text{ns}$ $\texttt{t}_{10-50\%} \geq 32\,\text{ns}$
Time maximum amplitude	$45\,000\,\text{ns} \leq \texttt{t\_max} \leq 120\,000\,\text{ns}$
Effective drift time	$\texttt{dt\_eff} > 0\,\text{ns}$
Baseline mean	$\texttt{bl\_mean} > -400\,\text{ADC}$

Table 5.2: Summary of the implemented quality cuts (QC)

5.2.2 Effects of QC

In Fig. 5.8, 5.9 and 5.10 we show some examples of the effects of the quality cuts on the raw waveforms samples acquired by the Germanium detectors.

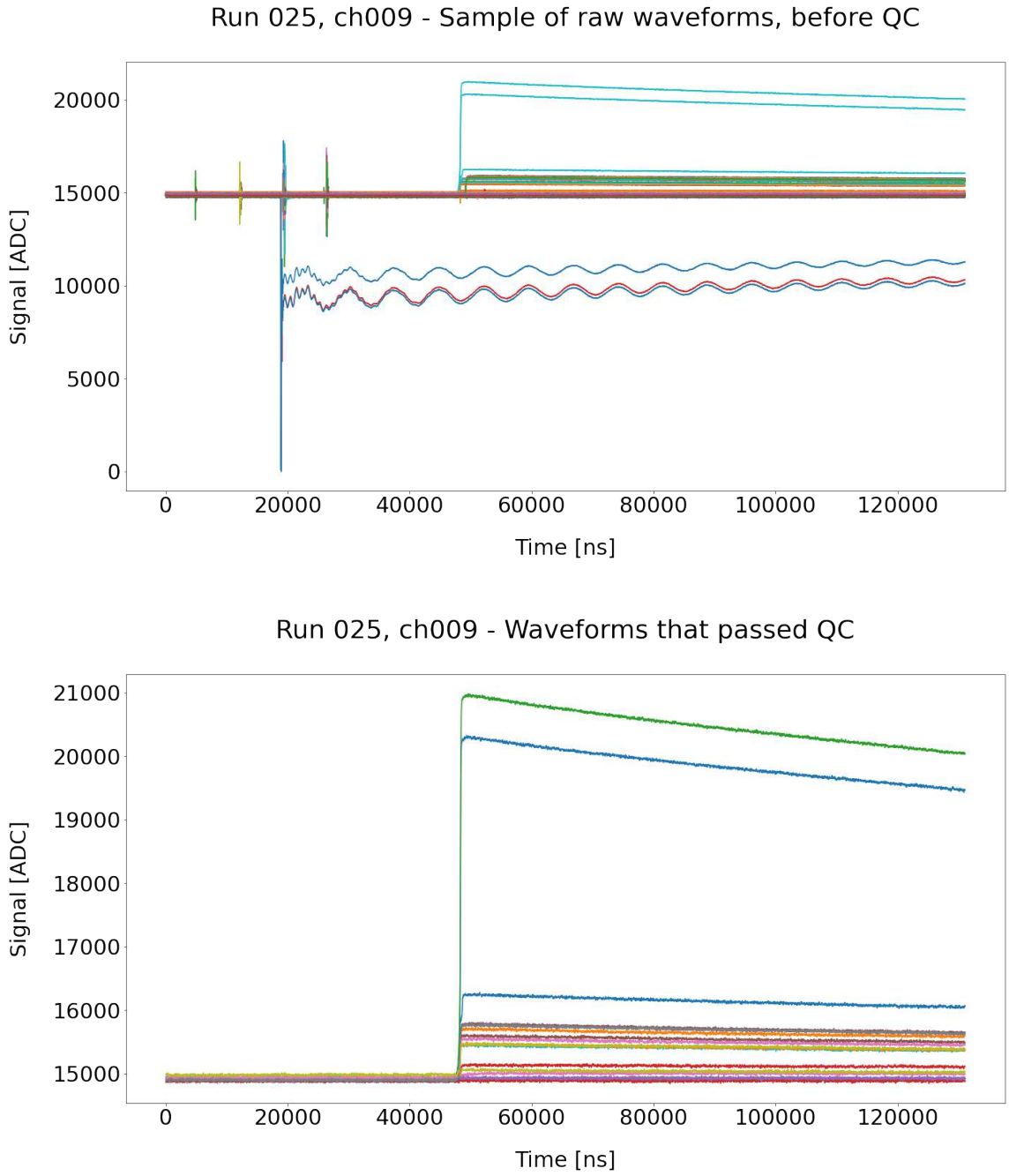


Figure 5.8: TOP: Sample of waveform acquired by the Germanium detector in ch009 during a short time slot from run 025. BOTTOM: Waveforms that passed the quality cuts, among those showed above

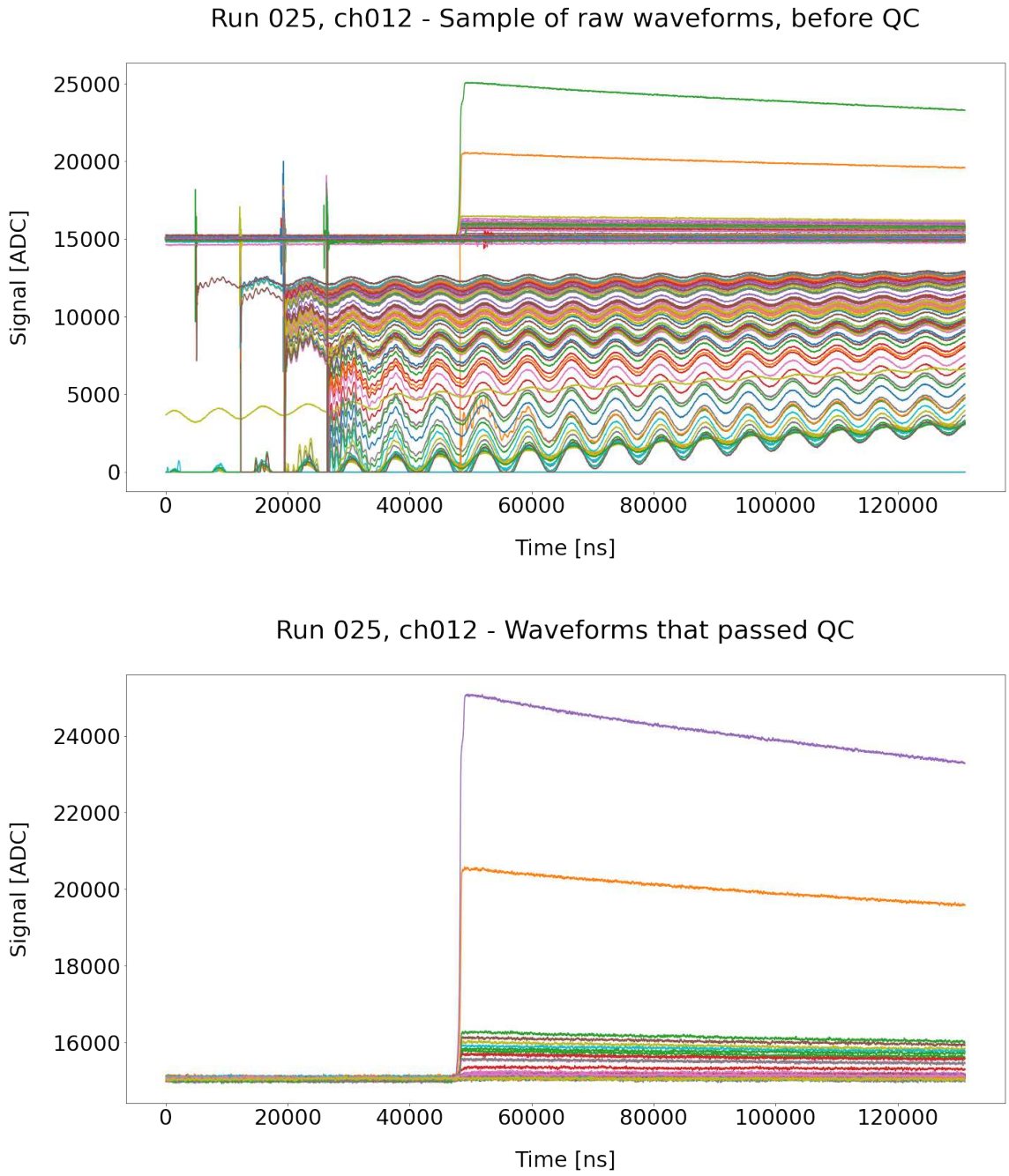


FIGURE 5.9: TOP: Sample of waveform acquired by the Germanium detector in ch012 during a short time slot from run 025. BOTTOM: Waveforms that passed the quality cuts, among those showed above

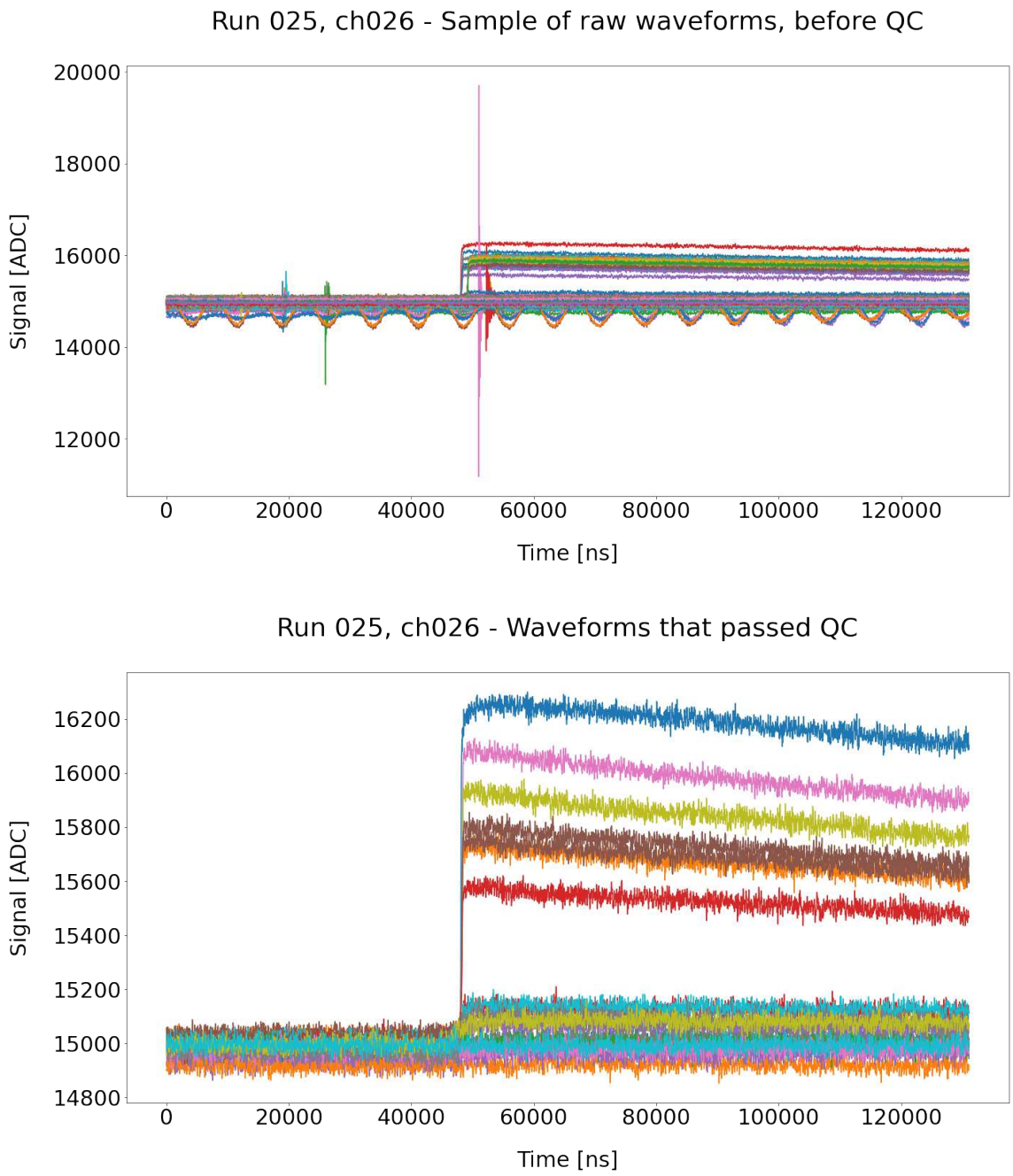


FIGURE 5.10: TOP: Sample of waveform acquired by the Germanium detector in ch026 during a short time slot from run 025. BOTTOM: Waveforms that passed the quality cuts, among those showed above. Notice the different y-axis scale.

Similar results are obtained in all the data collected in the three runs under analysis.

It is clear from the plots that the quality cuts successfully remove most of the undesired noise.

Tab. 5.3, 5.4 and 5.5 summarize the results of the QC in terms of fraction of events that manage to pass all the cuts. The analysis is performed channel by channel, in order to monitor the performances of the single detectors. The results are then presented grouping together the detectors physically located in the same string, in order to check whether there is any correlation between the noise level and the string.

Moreover, this analysis is performed considering each run individually in order to verify the stability of the acquisition and of the noise level over time, from one run to the other.

If the results from the three runs are compatible, indicating that the noise level has been stable in time, data will be again put together as a unique sample for the rest of the analysis.

Run 025					
String	Channel	Number raw signals	Number signals after PSR	Number signals after QC	Surviving fraction (%)
1	ch023	335562	315300	1670	0.5
	ch025	335562	315300	2318	0.69
	ch026	335562	315300	2389	0.71
	ch027	335562	315300	2342	0.7
	ch028	335562	315300	2337	0.7
	ch029	335562	315300	1578	0.47
	ch016	335562	315300	2421	0.72
2	ch009	335562	315300	2408	0.72
	ch012	335562	315300	3205	0.96
	ch014	335562	315300	2457	0.73
	ch015	335562	315300	2334	0.7
7	ch037	335562	315300	1798	0.54
	ch038	335562	315300	2139	0.64
	ch039	335562	315300	2153	0.64
	ch040	335562	315300	1755	0.52
	ch042	335562	315300	2169	0.65
	ch007	335562	315300	1256	0.37
	ch043	335562	315300	973	0.29
8	ch005	335562	315300	1975	0.59
Total		6375678	5990700	39677	0.62

TABLE 5.3: Summary of the action of PSR and quality cuts on data from run 025. The surviving fraction is calculated from the number of events passing QC, with respect to the original number of raw signals. Results are shown detector by detector; the last row contains the total number of events at each step and the average surviving fraction for this run.

Run 026					
String	Channel	Number raw signals	Number signals after PSR	Number signals after QC	Surviving fraction (%)
1	ch023	459180	438949	1878	0.41
	ch025	459180	438949	2284	0.5
	ch026	459180	438949	2659	0.58
	ch027	459180	438949	2736	0.6
	ch028	459180	438949	2671	0.58
	ch029	459180	438949	1833	0.4
	ch016	459180	438949	2701	0.59
2	ch009	459180	438949	2282	0.5
	ch012	459180	438949	2988	0.65
	ch014	459180	438949	2222	0.48
	ch015	459180	438949	2192	0.48
7	ch037	459180	438949	3434	0.75
	ch038	459180	438949	8742	1.9
	ch039	459180	438949	2092	0.46
	ch040	459180	438949	1690	0.37
	ch042	459180	438949	2014	0.44
	ch007	459180	438949	1187	0.26
	ch043	459180	438949	826	0.18
8	ch005	459180	438949	1990	0.43
Total		8724420	8340031	48421	0.56

Table 5.4: Summary of the action of PSR and quality cuts on data from run 026. The surviving fraction is calculated from the number of events passing QC, with respect to the original number of raw signals.Results are shown detector by detector; the last row contains the total number of events at each step and the average surviving fraction for this run.

Run 027					
String	Channel	Number raw signals	Number signals after PSR	Number signals after QC	Surviving fraction (%)
1	ch023	669966	631952	3006	0.45
	ch025	669966	631952	3837	0.57
	ch026	669966	631952	4158	0.62
	ch027	669966	631952	4347	0.65
	ch029	669966	631952	2798	0.42
	ch016	669966	631952	4693	0.7
2	ch009	669966	631952	3693	0.55
	ch012	669966	631952	4964	0.74
	ch014	669966	631952	3755	0.56
	ch015	669966	631952	3789	0.57
7	ch037	669966	631952	4472	0.67
	ch038	669966	631952	9743	1.45
	ch039	669966	631952	3794	0.57
	ch040	669966	631952	2969	0.44
	ch042	669966	631952	3645	0.54
	ch007	669966	631952	2225	0.33
	ch043	669966	631952	1494	0.22
8	ch005	669966	631952	3487	0.52
Total		12059388	11375136	70869	0.59

TABLE 5.5: Summary of the action of PSR and quality cuts on data from run 027. The surviving fraction is calculated from the number of events passing QC, with respect to the original number of raw signals.Results are shown detector per detector; the last row contains the total number of events at each step and the average surviving fraction for this run.

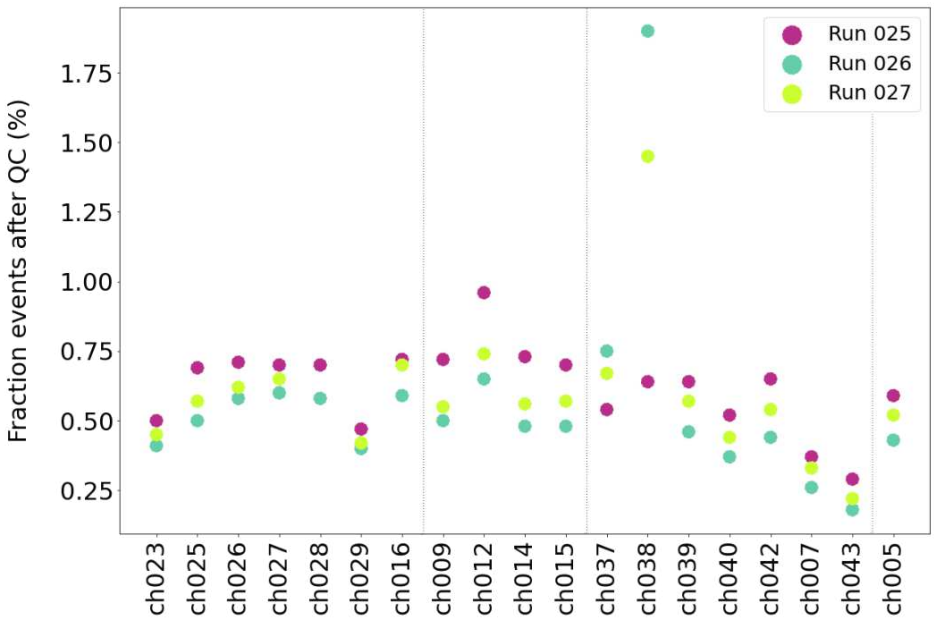


FIGURE 5.11: Fraction of signals that passed PSR and QC in runs 025, 026, 027. These are the signals identified as physical.

Fig. 5.11 shows the results obtained in the three runs: the fraction of events surviving the quality cuts is rather uniform in the available channels. Some exceptions are present, though. The fraction of good signals in ch043 is sensibly lower with respect to the other channels, meaning that ch043 is more noisy than the others. We recall the fact that ch043 presented an anomalous behavior also when performing the pulser signals analysis (Fig. 5.5): its pulser peak was shifted in energy with respect to the other channels. We may then infer that something in the electronics of ch043 does not work fully properly, leading to an anomalous abundance of noise and to a distortion of the pulser signals energy. Further analysis and monitoring should be done in order to check the performances of this channel. In particular, the behavior of ch043 data will be checked in the future runs. Nevertheless, we assume that the signals identified as physical in this channel, despite being few, are still informative, and so they are retained for the rest of the analysis.

Another channel showing an anomalous behavior is ch038: here, during run 026 and 027, the fraction of signals identified as physical is much higher than in the other channels. The position of the detector read out in ch038 is not anomalous: it is the second detector in string 7, which is a quite stable string (all the detectors but one are stable enough to be kept in the analysis); its mass, also, is fully compatible with many others of the deployed detectors. Also in this case, then, the behavior should be investigated more with some dedicated analysis. A possibility could be to check if the detector is located close to any specific background source, having high activity: this would lead to a high rate of physical signals, having thus the correct shape to pass the quality cuts. Another possibility could be to check whether this channel acquires anomalously shaped signals, that are not removed with the implemented quality cuts, and eventually to develop some new dedicated cut. For this work, anyway, we will rely on the performed cuts and retain all the signals thus identified as physical.

Finally, a systematic trend in the data is observed: the fraction of physical signals tends to be systematically higher during run 025 and lower during run 026, with run 027 results in between. This suggests that the electronics may have been slightly more noisy during the runs in which the fraction of good events is lower, possibly due to the parameters employed for the operation of the system (HV alimentation or threshold for the acquisition). Also in this case, further analysis should be performed to understand up to which point these results are significant, and which are the optimal conditions to acquire a clean dataset.

Apart from this features, the distribution is rather uniform and stable over time.

5.2.3 Monitoring of QC effects

We want now to check the consistency among the different quality cuts. In other words we want to check whether or not the action of one cut is enhanced and supported by the action of the other cuts.

A possible way of performing this check is to consider one of the observables on which the cuts are made and to compare its distribution before performing any of the cuts with its distribution after performing all the cuts except the one on the observable itself. The point of this comparison is to see whether the action of all the other cuts manage to identify the region of interest of the studied observable.

This procedure is based on the assumption that *good signals* owe simultaneously all the stated features, so constitute a well defined volume in the parameters space, while features of *noise signals* are assumed to be uniformly distributed. Performing all but one QC should therefore remove signals from the noise volume while keeping untouched the *good signals* volume.

An example of this procedure is shown in Fig. 5.12. The chosen observable is $\mathtt{t}_0$, the start of the risetime: its distribution in raw data from runs 025, 026 and 027 (blue) is compared with its distribution in the dataset obtained performing all the quality cuts presented in Tab. 5.2 but that on $\mathtt{t}_0$ (orange). If one then implements the cut on $\mathtt{t}_0$ itself on the latter (orange) distribution, the lateral shaded regions are removed, and the the resulting dataset turns out to be the final, after-QC dataset.

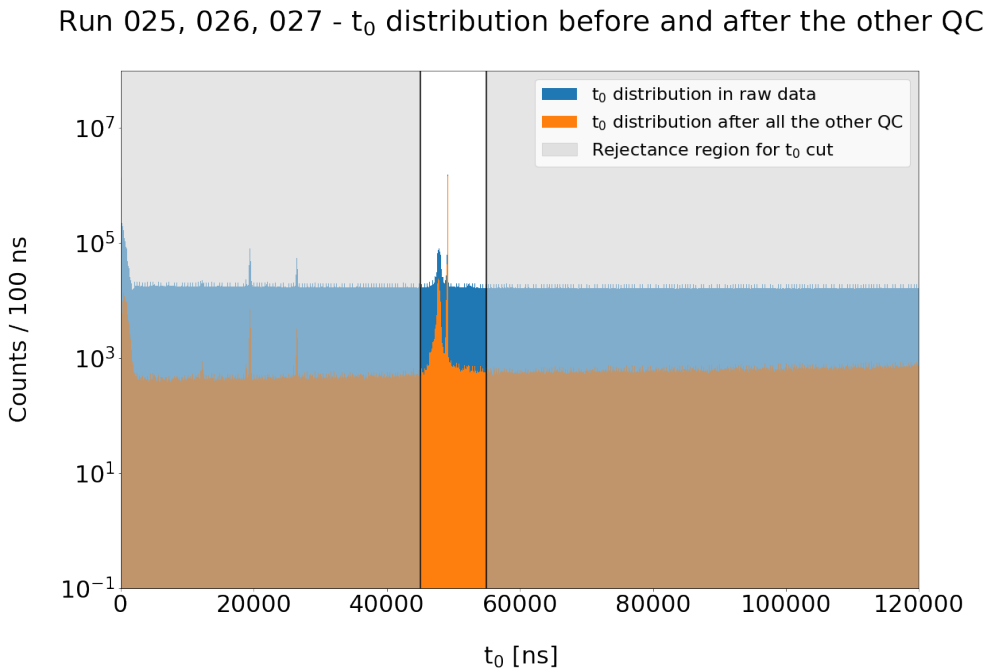


FIGURE 5.12: The behavior of $\mathtt{t}_0$ distribution under quality cuts is analyzed: its distribution in the sample obtained from raw data performing all QC except the one on $\mathtt{t}_0$ itself (orange) is compared with its distribution in the original raw dataset (blue). The trigger peak, which is the expected $\mathtt{t}_0$ for *good signals* is enhanced in the cut dataset. The shaded regions represent the action of the cut on $\mathtt{t}_0$. The final dataset, constitued by signals passing all QC, is therefore represented by the central orange volume.

From Fig. 5.12 it is clear that the quality cuts on the other observables successfully clean the peak area of $\mathtt{t}_0$ distribution, enhancing the trigger peak, which represent the expected start of risetime for correctly shaped physical signal. In facts, the cut on $\mathtt{t}_0$ selects precisely the region of the peak enhanced by the other quality cuts.

5.2.4 Results of QC

In Fig. 5.13 the spectrum made up of the events that passed all quality cuts is compared to the original spectrum. Please note that as *original spectrum* we consider the spectrum obtained as output of the preceding cut, so the spectrum *after* pulser signals removal.

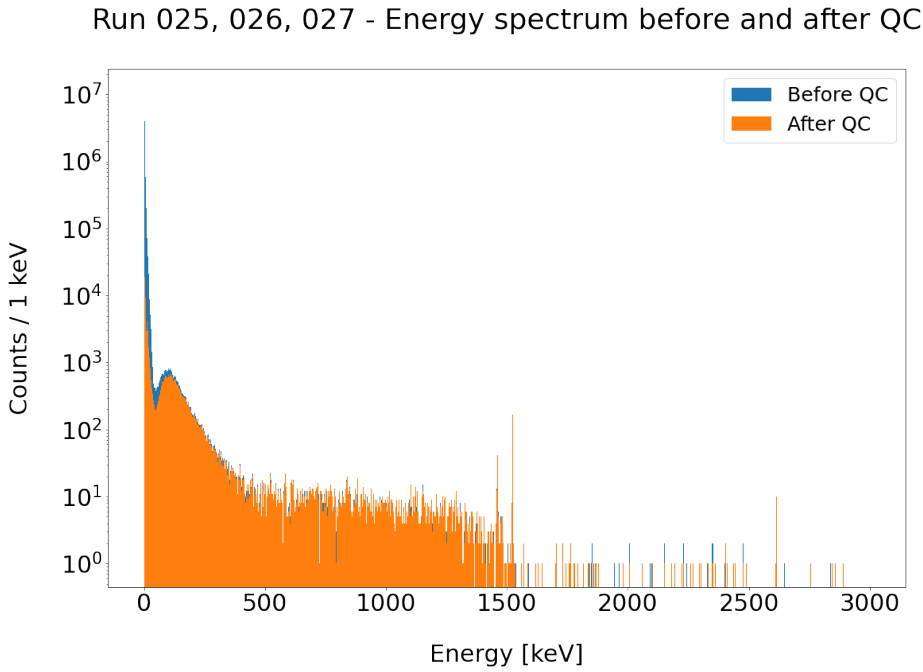


FIGURE 5.13: Full energy spectrum from run 025, 026, 027, before and after quality cuts (zoom).

From Fig. 5.13 and 5.14 it is clear that the effects of quality cuts are mostly concentrated in the low energy region, and in particular against the peak close to $Energy = 0$, whose height decreases by nearly four orders of magnitude.
This is consistent with the expectations, since to this peak mostly contribute noise signals and signals which do not bring any physical information: all the flat signals, collected only because of the global trigger, actually represent zero energy events; for extremely irregular signals, with interplay of positive and negative spikes, and for sinusoidal signals, the analytical result of the reconstructed energy is often close to zero because positive and negative regions compensate, but in facts the resulting value is not meaningful and does not have any physical interpretation.

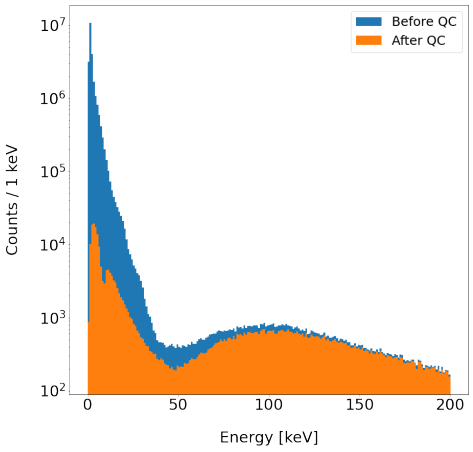


FIGURE 5.14: Zoom of the low energy region of Fig. 5.13, where the effects of quality cuts are concentrated.

Table 5.6 summarizes the impact of quality cuts on the number of events: it is clear that physical signals, on which the following analysis will be performed, constitute just a small fraction of the original dataset. Fig. 5.15 shows the action of quality cuts in the ROI.

	Number of signals before QC	Number of signals after QC	Surviving fraction (%)
Full energy range	27 253 386	233 351	0.86%
ROI	38	22	57.89 %

TABLE 5.6: Number of signals before and after quality cuts. The surviving fraction is calculated as ratio between the two numbers, so it is referred only to the present cut. Two energy regions are considered: the full energy range of the spectrum and the ROI around $Q_{\beta\beta}$ for the $0\nu\beta\beta$ analysis.

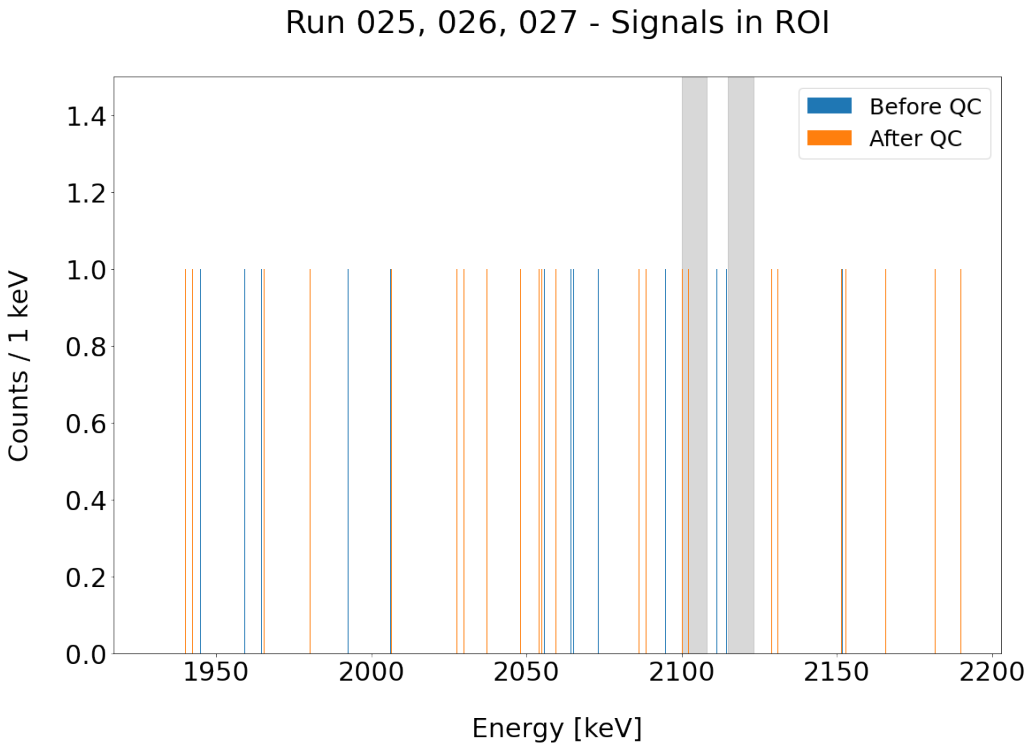


FIGURE 5.15: Signals appearing in the ROI before and after Quality Cuts. The considered ROI spans form 1930 keV to 2190 keV; the energy regions in the two shaded bands are excluded from ROI due to the possible presence of signals from the ^{208}Tl (2104 keV) and ^{214}Bi (2119 keV) γ lines [32, 68].

With the presented analysis, artificial and noise signals have been removed from the spectrum. The events remaining in the spectrum are physical and their signals have the expected shape.

Some of these physical events, though, still represent background contributions: external muons and radioactivity coming both from the surrounding materials and from the materials of which the experiment is made of. These background contributions are retained as low as possible by the location of the experiment, by its shielding, and by the choice of materials for its construction and operation; despite this, though, they are still present and must be considered, identified and removed in order to perform meaningful analysis on $2\nu\beta\beta$ and eventually $0\nu\beta\beta$ events.

The discrimination will be performed exploiting the intrinsic characteristics of double beta decay, which allow to distinguish it from many other processes: double beta decays are single site events (SSE), in which all the energy is released within an extremely small volume. This, in practice, means events in which only one detector acquire a non-zero energy signal.

The following analysis, then, will focus on identifying, studying and removing from the double beta

spectrum all the events in which a non-zero energy signal is detected by the PMTs (Muon Veto), by another Germanium detector in coincidence (Multiplicity cut) or by the SiPMs (LAr Veto) in coincidence with a signal in a Germanium detector.

5.3 Muon veto

Some of the signals acquired by Germanium detectors are expected to come from muons reaching the experimental site. These signals constitute background with respect to the searched $0\nu\beta\beta$ signal, therefore they must be identified and removed from the energy spectrum. To achieve this goal, the muon veto is introduced: a series of auxiliary detectors (PMTs) read out the water tank surrounding the cryostat; when these PMTs detect a signal in coincidence with a signal in the Germanium detectors, the event is tagged as muon and can be removed in the analysis.

Considering the deep underground location of the experimental site, muons are the only component of cosmic radiation which manage to reach it and which, therefore, must be considered in the background studies. In fact, the muon flux, despite being reduced by a factor 10^6 with respect to the above-ground intensity to an average value of about 1.2 muons/(m^2·h) [50], is not negligible within the desired sensitivity of the experiment.

When an energetic muon reaches the experimental site, it is expected to cross all the layers of the experiment, to interact with the encountered materials, and to produce a wide range of signals in the deployed detectors. Also its reaction products may undergo several interaction, producing additional signals. The signature of a muon event is therefore the acquisition of multiple signals in coincidence, thorough all the detection system.

In particular, Cherenkov light is emitted when an energetic charged particle travels at high speed across the ultrapure water contained in the water tank, with:

$$v_p > \frac{c}{n_w} \quad \rightarrow \quad v_p > 0.75 \cdot c \tag{5.1}$$

In which v_p is the speed of the particle, c the speed of light in vacuum and $n_w \approx 1.33$ the refractive index of the medium, in this case water.

The water tank is properly instrumented with 66 PMTs which read out this Cherenkov light. These PMTs are connected to the global trigger: when a Germanium detector detects a signal above acquisition-threshold, also the PMTs acquire and save data. Then, if in a certain event both the Germanium detectors and the PMTs acquire a non-zero signal, the event is identified as an impinging muon.

5.3.1 Definition of the veto condition

In the LEGEND-200 commissioning phase setup, when the PMTs detect a non-zero signal in coincidence with a signal in the Germanium detectors, a fast feedback signal is injected in the auxiliary ch001 of the FlashCam (Fig. 5.16, RIGHT). Instead, when no signal is detected in the PMTs, the ch001 of the FlashCam acquires just a flat waveform because of the global trigger (Fig. 5.16, LEFT).

Thanks to this configuration, in order to understand whether or not an event was originated by a muon, it is not necessary to study explicitly the signals collected by the PMTs, but rather it is sufficient to check what type of signal appears in the ch001 of the FlashCam:

- Flat signal in FC ch001 (Fig. 5.16, LEFT) → Non-muon event
- Fast signal appearing in FC ch001 (Fig. 5.16, RIGHT) → Muon event

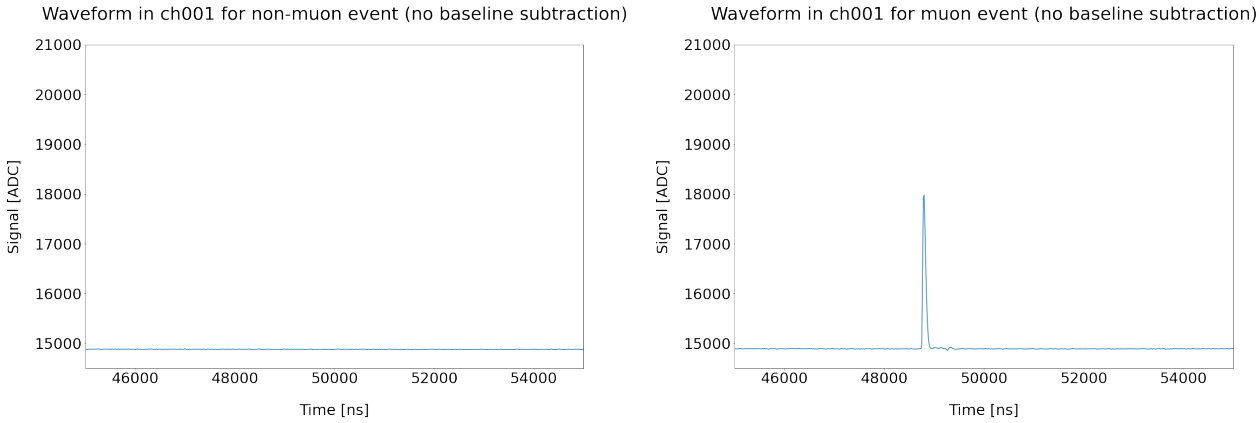


FIGURE 5.16: Examples of the two types of waveforms that can appear in ch001. LEFT: waveform in ch001 in case of an event that is *not* tagged as muon by the PMT system. In this case ch001 still acquires a waveform because of the global trigger, but the signal is flat since nothing is injected in ch001; RIGHT: waveform in ch001 in case of an event that is tagged as muon by the PMT system. In this case a fast signal is injected in ch001. It is this fast signal, in practice, which tags the events as *muon event*, without needing to check manually the data acquired by the PMTs.

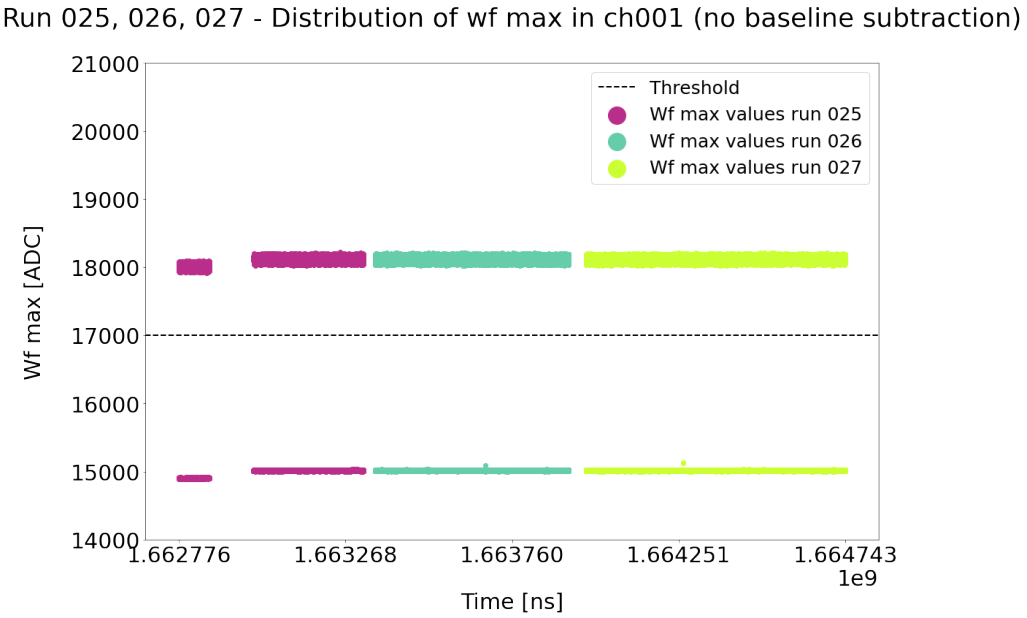


FIGURE 5.17: Distribution of the parameter `wf_max` (maximum of the waveform) over time, in data acquired by ch001 during runs 025, 026, 027. The two-band structure is the outcome of the two different types of signals that can be acquired in ch001: flat signals (lower band) in case of non-muon events and fast, non-zero signals (upper band) in case of muon events.

In practice, to identify and remove muon events from the energy spectrum, we can proceed analogously to what we did to identify and remove pulser events.

With the two stated types of signals acquired in ch001 (Fig. 5.16) we get a two-bands structure in its `wf_max` distribution (Fig. 5.17). We tag as *muon events* the events collected by Germanium detectors in coincidence with a non-zero signal in ch001, namely in coincidence with a signal belonging to the upper band of the `wf_max` distribution. Since, apart from a slight increase after the DAQ crash in run 025, the position of the bands appear stable during the three runs, a common threshold can be established at 17 000 ADC.

Therefore, we tag as muon events all the events having `wf_max` $>$ 17 000 ADC in the ch001, and we remove all the corresponding signals from the Germanium detectors' energy spectrum.

5.3.2 Results of muon veto

The results of the muon veto are summarized in Tab. 5.7, showing the number of events identified as muons and non-muons, respectively.

Total number of events	Number of *muon* events	Number of *non-muon* events
217387	50239	167148

TABLE 5.7: Number of events identified as *muons* and *non-muons* by the muon veto. Note that the total number of events (first column) refers to the sample of physical signal only, provided by the quality cuts: in other words, the first column contains the number of events in which at least one signal survives to the quality cuts.

In particular, from the number of identified muon events, knowing the duration of the data taking (t) and the surface (Σ) crossed by the muons, we can give a rough estimate of the muon flux reaching the experiment:

$$\begin{aligned} \Phi_\mu &= \frac{\text{Number muon events}}{\Sigma \cdot t} \\ &\approx \frac{\text{50239 muon events}}{\text{86.15 m}^2 \cdot \text{479.69 h}} \\ &\approx 1.22 \text{ muons}/(\text{m}^2 \cdot \text{h}) \end{aligned} \tag{5.2}$$

The time t is the total duration of the data taking (Tab. 3.3), expressed in hours.

For the estimate of the surface Σ we argue that if a particle crosses both the water tank (in which the PMTs are located) and the Germanium detectors array, it must cross the surface of the cryostat at some point. Also, if we assume the muons contributing to the flux to come only from above-ground, we can refine the estimate and employ as effective surface Σ crossed by the muons only the upper and lateral surface of the cryostat. The cryostat employed in LEGEND-200 is the same employed in GERDA, having a total surface of about 100 m^2 and a radius of about 2.1 m [78]. If we model the cryostat as a cylinder for simplicity, we obtain $\Sigma = \Sigma_{tot} - \Sigma_{bottom} \approx 100\text{m}^2 - \pi 2.1^2 \approx 86.15\text{m}^2$.

The result is fully compatible with the expected value of the muon flux at LNGS [50], confirming thus the efficacy of the muon veto system in identifying impinging muons.

The effects of the cut on the spectrum under analysis in the full energy range, in the low energy region and in the ROI are shown in Fig. 5.18, 5.19 and 5.20 respectively; the numerical results are presented in Tab. 5.8. The energy spectrum of the identified muons, instead, is shown in Fig. A.1.

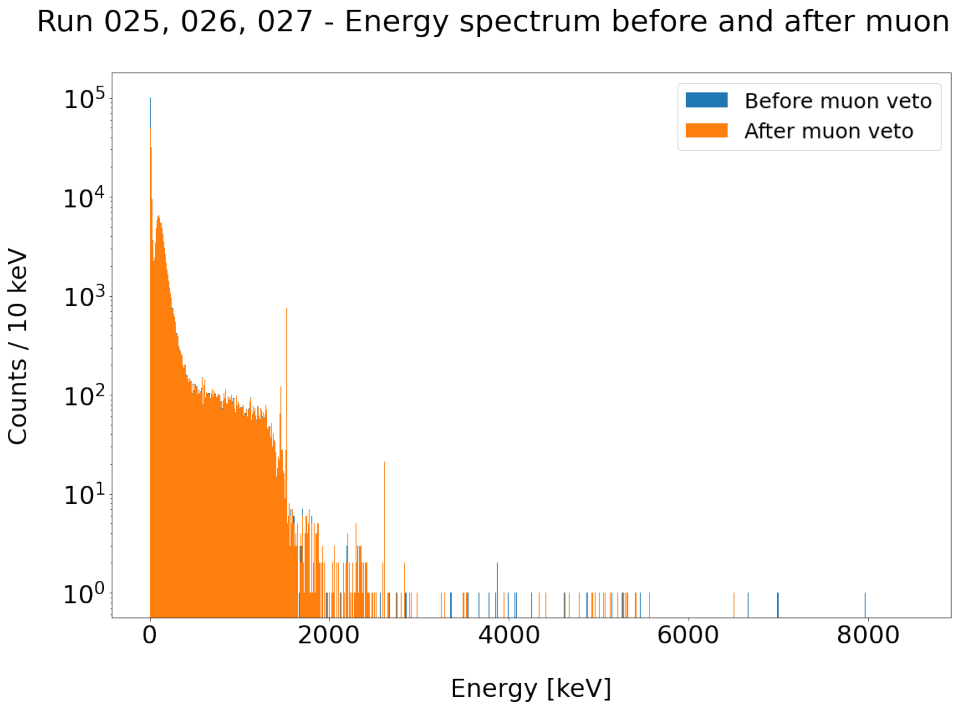


FIGURE 5.18: Full energy spectrum from run 025, 026, 027, before and after muon veto.

From Fig. 5.18 it is clear that the fraction of muon signals in the energy spectrum acquired by Germanium detectors is low: in fact, the spectrum before and after muon veto is almost indistinguishable by eye.
With a careful observation, though, it is possible to notice that some signals are removed from the high energy region ($\gtrsim 4000$ keV) and from the zero-energy noise peak.
In principle we expect muons to produce high energy signal, since they must have extremely high energy to penetrate that deep into rocks, and on average they are not completely absorbed in the experiment, but rather pass through it before continuing their propagation.

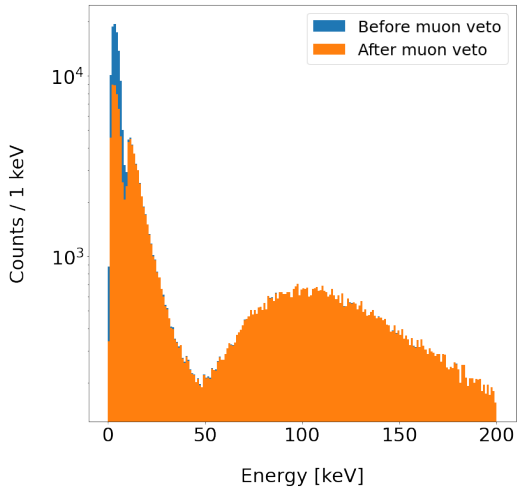


FIGURE 5.19: Zoom of the low energy region of Fig. 5.18.

The reason why a dominating fraction of the signals coming from events identified as muons belongs to the zero-energy peak (Fig. 5.19) can be found in the trigger configuration of the system: in fact, when the muon or its reaction products cross the Germanium detectors array, not all the detectors will be hit (in general, just few of them will be hit, instead), but all the channels will acquire the corresponding signal. Thus, most of these signals will be flat waveforms, populating the zero-energy peak.

Note that the spectra in Fig. 5.18 and 5.19 are built considering single signals, and not complete events.

In order to reconstruct the total energy released by a muon in the Germanium detectors, we must sum the energies of all its signals, namely we must calculate the total energy of the event. The muon spectrum shown in Fig. A.1 is built in this way, with one entry *per event*, rather than one entry *per signal*. Note that even the total energy of the event does not represent the total energy of the muon itself: to obtain it we should consider also all the energy released in the surrounding components, other than Germanium detectors. The total energy of the event gives just an idea of the energy released by the muon in the Germanium detectors, as discussed in Appendix A.1.

	Number of signals before muon veto	Number of signals after muon veto	Surviving fraction (%)
Full energy range	233 351	180 667	77.42%
ROI	22	20	90.90 %

TABLE 5.8: Number of signals before and after muon veto. The surviving fraction is calculated as ratio between the two numbers, so it is referred only to the present cut. Two energy regions are considered: the full energy range of the spectrum and the ROI around $Q_{\beta\beta}$ for the $0\nu\beta\beta$ analysis.

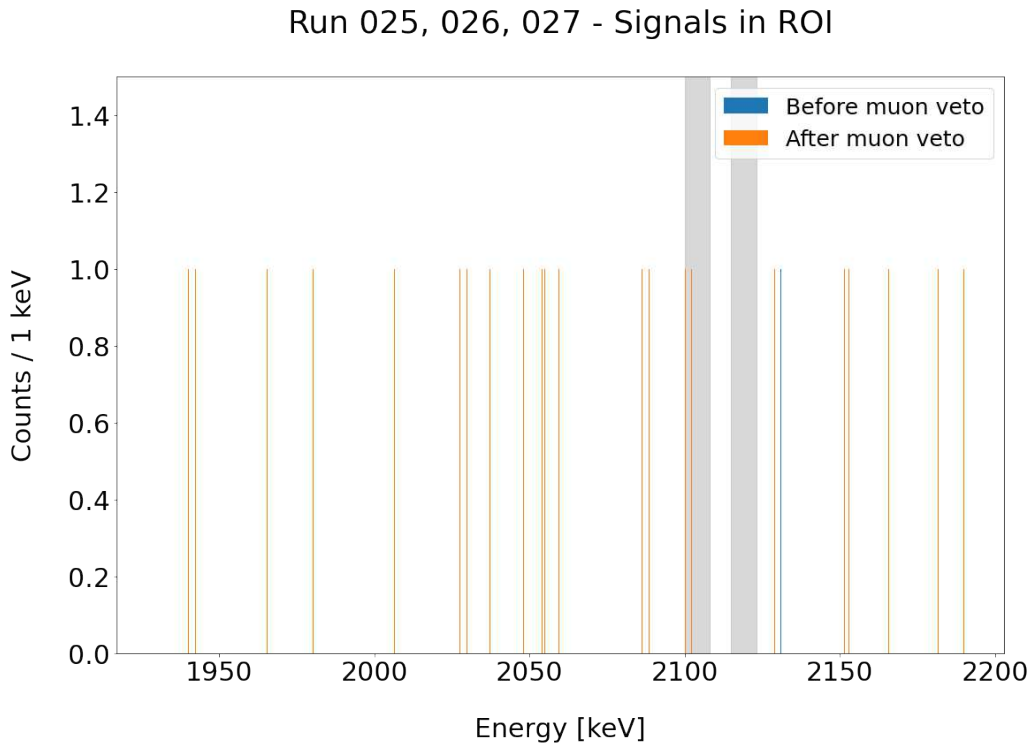


FIGURE 5.20: Signals appearing in the ROI before and after muon veto. The considered ROI spans form 1930 keV to 2190 keV; the energy regions in the two shaded bands are excluded from ROI due to the possible presence of signals from the ^{208}Tl (2104 keV) and ^{214}Bi (2119 keV) γ lines [32, 68].

5.4 Multiplicity cuts

This step of background removal exploits the topology of double beta events. we know from theory that the two electrons emitted in a $\beta\beta$ decay can travel in Germanium only for a range of order $1\,\mathrm{mm}^3$ [32]. This means that all the available energy is deposited in a small volume surrounding the interaction point: this volume is well within the volume of a single Germanium detector. Consequently, all the events in which some energy is released in more than one Germanium detector cannot be traced back to double beta decays, and therefore can be classified as background for the sake of this analysis.

In this Section, we study the multiplicity ($\mathcal{M}$) of the acquired events, defined as the number of Germanium detectors in which a non-zero signal is acquired in coincidence: studying the multiplicity of the events, then, means studying the coincidences among Germanium detectors. The aim of this study is to identify and isolate the events in which one and only one Germanium detector acquires a non-zero signal: only these events are good candidates for being the outcome of a double beta decay.

5.4.1 Dataset and possible cases

The total number of signals after muon veto is 180 667 (Tab. 5.8), coming from 167 148 distinct events (Tab. 5.7). Note that for the analysis of the multiplicity, signals must be clustered back in events, so as to be able to check how many detectors acquired a non-zero signal at a fixed time.

For these events, several scenarios are possible, depending on how many signals per event passed the QC: some events now are made of a single signal, and this can be zero or non-zero; others are still made of more than one signal, and these signals can be all zero, all non-zero or some zero and others non-zero. In which by *zero signals* we mean zero-energy, flat signals, while by *non-zero signals* we mean non-zero energy, non-flat signals. These scenarios can be summarized in terms of multiplicity ($\mathcal{M}$) of each event, namely in terms of number of detectors that acquired a non-zero signal which survived up to the muon veto. The considered cases are the following:

- **$\mathcal{M}$0**

  An event has multiplicity $\mathcal{M} = 0$ if all the detectors acquire only zero signals.

  These events are interpreted as pure background events, namely not containing any physical signal of interest for the $\beta\beta$ analysis. The $\mathcal{M}0$ spectrum (Fig. A.2) will therefore be made up of the the flat signals populating the zero-energy peak in the spectrum.

- **$\mathcal{M}$1**

  An event has multiplicity $\mathcal{M} = 1$ if precisely one detector acquires a non-zero signal, while all the others (eventual) signals are compatible with zero.

  These events are the good candidates as $\beta\beta$: one detector acquires a non-zero, physical signal while all the other acquire only flat signals. The $\mathcal{M}1$ spectrum, containing only the non-zero signal from each $\mathcal{M}1$ event, will be the main spectrum, on which all the following analysis, aimed at $\beta\beta$ study, will be performed. Note that being $\mathcal{M}1$ is a necessary but not sufficient condition for an event to be $0\nu\beta\beta$: also γ rays from the environmental radioactivity can release all their energy in a single detector. Therefore, further analysis will be performed to remove these spurious signals.

- **$\mathcal{M}\geq$ 2**

  An event has multiplicity $\mathcal{M} \geq 2$ if at least two detectors acquire a non-zero signal in coincidence.

  These events are mostly originated by photons: γ rays coming from environmental radioactivity, traveling through the array and releasing part of their energy in all the detectors they cross before stopping. For these events, the energy of the decay is given by the sum of the energies released in the hit detectors. Therefore, in the $\mathcal{M}\geq2$ spectrum (Fig. A.3) one entry *per event*, rather than one entry *per signal*, will appear, given by the sum of the energies of all the non-zero signals.

The aim of the analysis presented in this Section is to separate these three different contributions in the energy and in particular to extract the $\mathcal{M}1$ spectrum, on which the search for double beta decays will be pursued. What we need to do, then, is to establish a practical condition to discriminate *non-zero* signals from *zero* signals.

5.4.2 Working definition of multiplicity

We define *zero* signals the zero-amplitude and zero-energy signals acquired just because of the global trigger, and *non-zero* signals the non-zero-amplitude and non-zero-energy signals. According to this definition, we would tag as non-zero all the signals having amplitude[2] $\texttt{A} > 0$ or equivalently energy $\texttt{E}_{\texttt{rec}} > 0$. Of course these criteria are not suitable in practice: each waveform is made up of 8192 sampled points and in the case of a flat signal, they will not all lay precisely at amplitude $\texttt{A} = 0$, but rather they will be normally distributed around $\texttt{A} = 0$ with non-zero gaussian noise. Consequently, neither the amplitude nor, consequently, the reconstructed energy of these signals will be exactly zero, but rather it will assume random values around zero.

What we need to do is then to establish a consistent threshold either on $\texttt{E}_{\texttt{rec}}$ or on $\texttt{A}$, which manages to perform the desired discrimination among *zero* and *non-zero* signals, taking into account also the statistical fluctuations of the parameter itself.

In particular, we put a threshold on the energy $\texttt{E}_{\texttt{rec}}$, since this parameter is less sensible than the amplitude $\texttt{A}$ to the statistical fluctuations of the single points of the waveform. In fact, as discussed in Chapter 4, the energy reconstruction filters are introduced precisely to mitigate the impact of small amplitude fluctuations due to single points positions.

Then for the following analysis we define:

$$\textit{Non-zero} \text{ signals} \rightarrow \texttt{E}_{\texttt{rec}} > \texttt{E}_{\texttt{mt}}$$

$$\textit{Zero} \text{ signals} \rightarrow \texttt{E}_{\texttt{rec}} < \texttt{E}_{\texttt{mt}}$$

with $\texttt{mt}$ standing for *multiplicity threshold*.

The value of the threshold is chosen in order to isolate the zero-energy peak from the rest of the spectrum. The rational of this criterion lays of course in the physical interpretation of the data distribution, widely discussed in [68]: signals in the zero-energy peak are interpreted as pure noise, while signals in the rest of the spectrum are interpreted as physical ($\beta\beta$ signals embedded in other background signals coming from the radioactivity of the surrounding materials).

Considering the low energy region of the spectrum, the most relevant contribution in the neighborhood of the zero-energy peak is made up of the signals coming from the β decay of ^{39}Ar, present in LAr [68],[50].

In fact, the LAr employed as refrigerant and active veto is made of atmospheric Argon[3]. Its most abundant constituents are stable isotopes: ^{40}Ar (99.6%), ^{36}Ar (0.334%) and ^{38}Ar (0.0630%); besides these, though, it contains also traces of radioactive isotopes, and in particular of ^{39}Ar and ^{42}Ar [66]:

- ^{39}Ar decays 100% β^- towards ^{39}K: $^{39}\text{Ar} \rightarrow {}^{39}\text{K} + \text{e}^- + \bar{\nu}_e$, with $\text{Q} = 565\,\text{keV}$.
- ^{42}Ar decays 100% β^- towards ^{42}K: $^{42}\text{Ar} \rightarrow {}^{42}\text{K} + \text{e}^- + \bar{\nu}_e$, with $\text{Q} = 599\,\text{keV}$.

The spectrum of these decays is therefore a continuous distribution ranging from zero to the corresponding Q.

[2] The employed operational definition of the amplitude is $\texttt{A} = \texttt{wf\_max} - \texttt{corrected\_baseline}$, in which $\texttt{wf\_max}$ is the *dsp* parameter providing the maximum value of the waveform and $\texttt{corrected\_baseline}$ is a user-defined parameter calculated as $\texttt{corrected\_baseline} = \texttt{baseline} + \texttt{bl\_mean}$, with $\texttt{baseline}$ and $\texttt{bl\_mean}$ the *dsp* parameters representing respectively the baseline value calculated at run time by the FC and the mean value of the baseline calculated offline after baseline subtraction, representing a correction to the raw $\texttt{baseline}$ value.

[3] Atmospheric Argon is employed in GERDA Phase II, LEGEND-200 commissioning and LEGEND-200. For LEGEND-1000 it is planned to partially substitute it with radiopure UnderGround sourced Liquid Argon (UGLAr). This should reduce the background coming from Argon decays, expecially the contribution of ^{39}Ar and ^{42}Ar [49].

In particular, the low energy distribution appearing in our spectrum is dominated by the ^{39}Ar contribution: the fraction of events coming from ^{42}Ar is negligible in that region (to the point that in [32] and [68] ^{42}Ar is not even mentioned as background source in that region).[4]

In the following we will therefore refer to the Argon distribution in the low energy region (≈[0, 565] keV) as from ^{39}Ar, accounting for its main constituent. What we implicitly mean is *distribution that we expect to be dominated by ^{39}Ar β decay, with a small contribution from ^{42}Ar βdecay.* The analysis procedure, despite being referred to ^{39}Ar, does not assume a strictly pure distribution.

In principle, we want the energy threshold to separate the zero-energy noise peak from the Argon decay distribution. In practice, we choose as threshold the energy at which we expect the contribution from ^{39}Ar to drop to less than 50% of the total amount of data acquired at that energy. The procedure and the results are presented in Fig. 5.21.

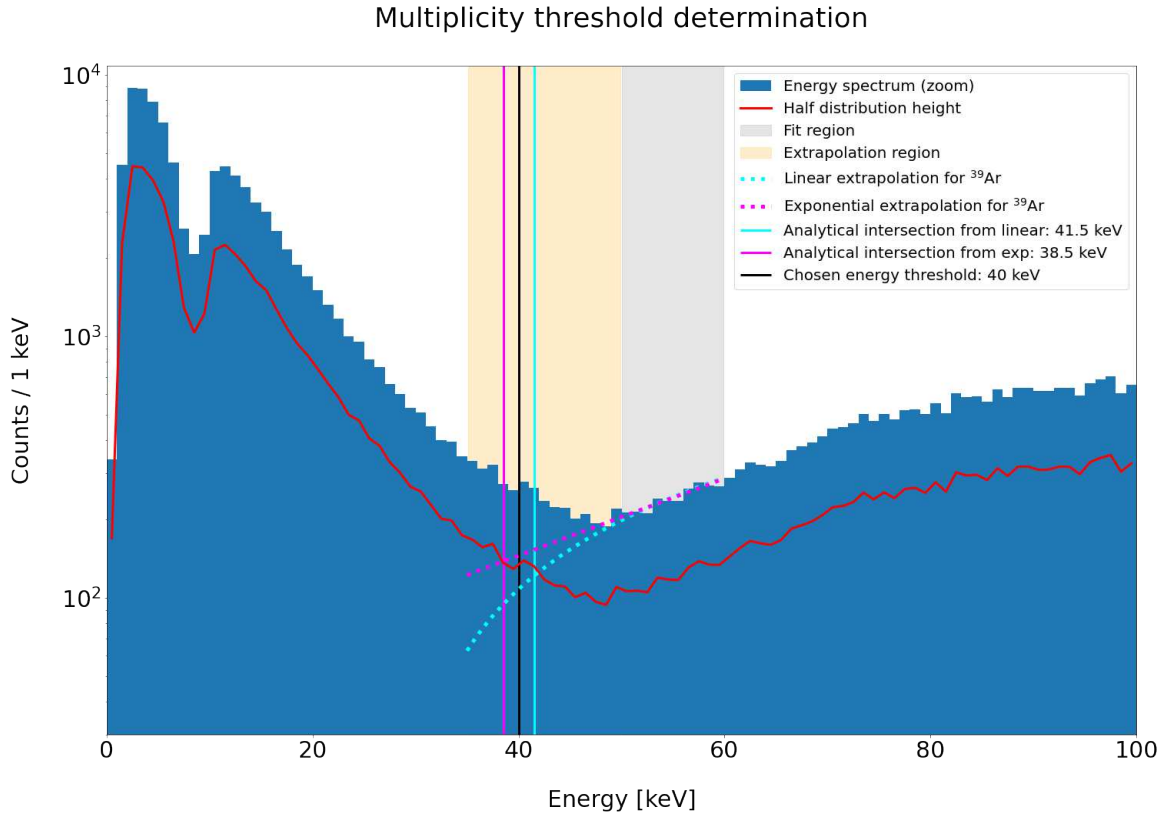


FIGURE 5.21: Zoom on the low energy region of the spectrum of the physical signals that passed the muon veto. The histogram is plotted with 1 keV per bin. In principle, we could have built it with a thinner binning, in order to determine with higher precision the analytical threshold as intersection point (in pink/blue) between the half height distribution (in red) and the extrapolated argon trend (in dashed pink/blue). In practice, though, with a thinner binning the (half) height of the data distribution is dominated by fluctuations, making it impossible to determine consistently the intersection point.

Given the energy spectrum, we extrapolate the trend of the ^{39}Ar peak towards the low energy region. We find analytically the energy at which the height of this extrapolated distribution coincides with half of the height of the distribution of the actually acquired data (in red in Fig. 5.21). At this energy, we expect the 50% of the events to come from the zero-energy peak, and the 50% to come from the ^{39}Ar peak.

Two possible models are considered for the extrapolation of the ^{39}Ar distribution towards low energy: a linear model, as first approximation of the distribution, and an exponential model, as approximation

[4] Note, though, that the daughter nucleus of the ^{42}Ar's beta decay, ^{42}K, gives instead a significant and sensible contribution to the background around $Q_{\beta\beta}$: ^{42}K too decays β^- towards ^{42}Ac with a Q value of 3525 keV; the spectrum of this decay is therefore a continuous distribution from 0 to 3525 keV, including this time the $Q_{\beta\beta}$ region. The impact of the background coming from ^{42}K on the B.I. in the region around $Q_{\beta\beta}$ is not negligible.

of the left side of the continuous β distribution. Please note that Fig. 5.21 is a log-linear plot, therefore the linear fit (dashed blue) appears logarithmic curve and the exponential fit (dashed pink) as a straight line.

For both of the models the ^{39}Ar distribution is fitted in the range [50,70] keV and extrapolated in the range [35,50] keV. The results of the fits are presented in Tab. 5.9.

	Linear $y = Mx + Q$	Exponential $y = Ae^{Bx}$
Fit parameters	$M = (9.0 \pm 0.6)$ keV^{-1} $Q = -250 \pm 40$	$A = 37 \pm 5$ $B = (0.034 \pm 0.002)$ keV^{-1}
χ^2	≈ 5.59	≈ 5.36
N_{dof}	13	13
P-value	≈ 0.02	≈ 0.02

TABLE 5.9: Linear and exponential fit results

The performed hypothesis tests allow to accept at 95% CL the null hypothesis that data come from both a linear and an exponential distribution. Note that this result must be interpreted carefully: the employed models makes sense only if interpreted as a local approximation of the actual ^{39}Ar distribution; in order to obtain a proper description of the total ^{39}Ar distribution, the full background model must be employed [68]. In our case, it can be acceptable to employ such models just because the energy value at which we need to extrapolate the height of the ^{39}Ar distribution is near to the region in which we performed the fit. Only under this condition the obtained results for the energy at which the contribution of ^{39}Ar drops below 50% are meaningful.

The identified energy region falls within $\approx$[38.5, 41.5] keV, considering as extremes the values from the two fit models. This result is fully consistent with the energy threshold of 40 keV employed for the multiplicity analysis of the GERDA Phase II data, as presented and discussed in [68].

We have now all the elements to choose a suitable threshold $\mathtt{E_{mt}}$. As said, both of the employed models provide a rough approximation of the actual ^{39}Ar distribution; therefore, none of the two analytical results can be considered fully representative of the energy at which the 50% condition is fulfilled. The above mentioned 40 keV threshold, instead, has two remarkable features. On the one hand, it comes from an accurate background modeling of GERDA Phase II data, and we do not expect any substantial variation of the ^{39}Ar distribution when coming to LEGEND-200 commissioning phase setup. On the other hand it falls in the middle of the analytical (though approximate) energy range we found.

Based on these observations, we choose to employ as energy threshold for the multiplicity analysis $\mathtt{E_{mt}}$ = 40 keV.

The definition of *zero* and *non-zero* signals becomes therefore:

$$\textit{Non-zero} \text{ signals} \rightarrow \mathtt{E_{rec}} > 40\,\text{keV}$$

$$\textit{Zero} \text{ signals} \rightarrow \mathtt{E_{rec}} < 40\,\text{keV}$$

In the following analysis, then, we define the multiplicity of an event as the number of Germanium detectors in which an energy of at least 40 keV is deposited.

The three considered scenario are therefore:

- $\mathcal{M}0$: events in which no signals above 40 keV are recorded;
- $\mathcal{M}1$: events in which only one signal above 40 keV is acquired;
- $\mathcal{M}\geq2$: events in which two ore more signals above 40 keV are acquired.

5.4.3 Results of multiplicity cut

The results of multiplicity analysis on our dataset are presented in Tab. 5.10. As expected, a considerable fraction of events is tagged as $\mathcal{M}0$: in fact, with the 40 keV threshold, all the events fully contained in the zero-energy peak are interpreted as $\mathcal{M}0$. Instead, few $\mathcal{M}\geq 2$ events are identified, since the statistics at high energy is much lower, and the distribution is dominated by $2\nu\beta\beta$ events, which are $\mathcal{M}1$.

Total number of events	Number of $\mathcal{M}0$ events	Number of $\mathcal{M}1$ events	Number of $\mathcal{M} \geq 2$ events
167148	80110	86289	749

TABLE 5.10: Summary of the results of multiplicity cut. Note that the total number of events refers only to the events that passed all the cuts performed up to now. Events in the full energy range are considered.

The spectrum of the events that passed the multiplicity cut (events identified as $\mathcal{M}1$) is shown and compared to the spectrum before multiplicity cut in Fig. 5.22. The numerical results of the multiplicity cut on our dataset are presented in Tab. 5.11, referring both to the full energy range and to the ROI. A zoom of the spectrum in the ROI is presented in Fig. 5.23. Finally, the residual $\mathcal{M}0$ and $\mathcal{M}\geq 2$ spectra are discussed in Appendix A (Fig. A.2, A.3).

Note that the $\mathcal{M}1$ spectrum contains, for each event identified as $\mathcal{M}1$, only the above-threshold signal; all the *zero* (below-threshold) signals eventually acquired within the same event are discarded from the $\mathcal{M}1$ spectrum and from the dataset on which the next analysis will be performed, since considered non-physical or at least non-informative. After the multiplicity cut then, the number of signals comes to coincide with the number of events in the dataset of interest: only one signal per event is kept, so from now on each considered signal comes from a unique event.

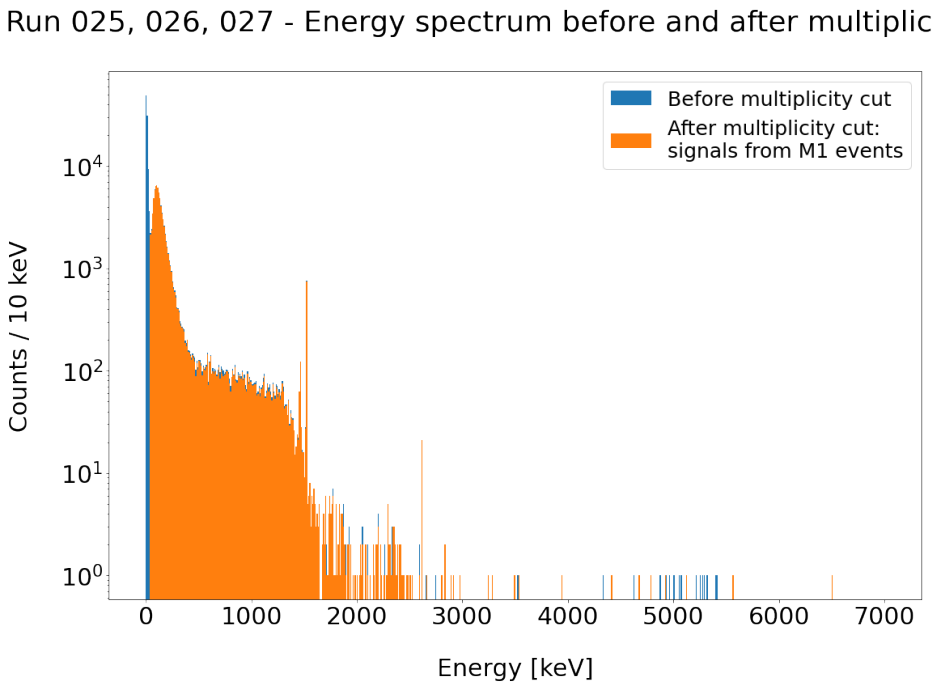


FIGURE 5.22: The energy spectrum of the events that passed the multiplicity cut, namely the spectrum made up of the above-threshold signals from events identified as $\mathcal{M}1$ by the multiplicity analysis (in orange), is superimposed to the spectrum before multiplicity cut (in blue).

	Number of signals before multiplicity cut	Number of signals after multiplicity cut	Surviving fraction (%)
Full energy range	180 667	86 289	47.76%
ROI	20	19	95%

TABLE 5.11: Number of signals before and after multiplicity cut. The surviving fraction is calculated as ratio between the two numbers, so it is referred only to the present cut. Two energy regions are considered: the full energy range of the spectrum and the ROI around $Q_{\beta\beta}$ for the $0\nu\beta\beta$ analysis.

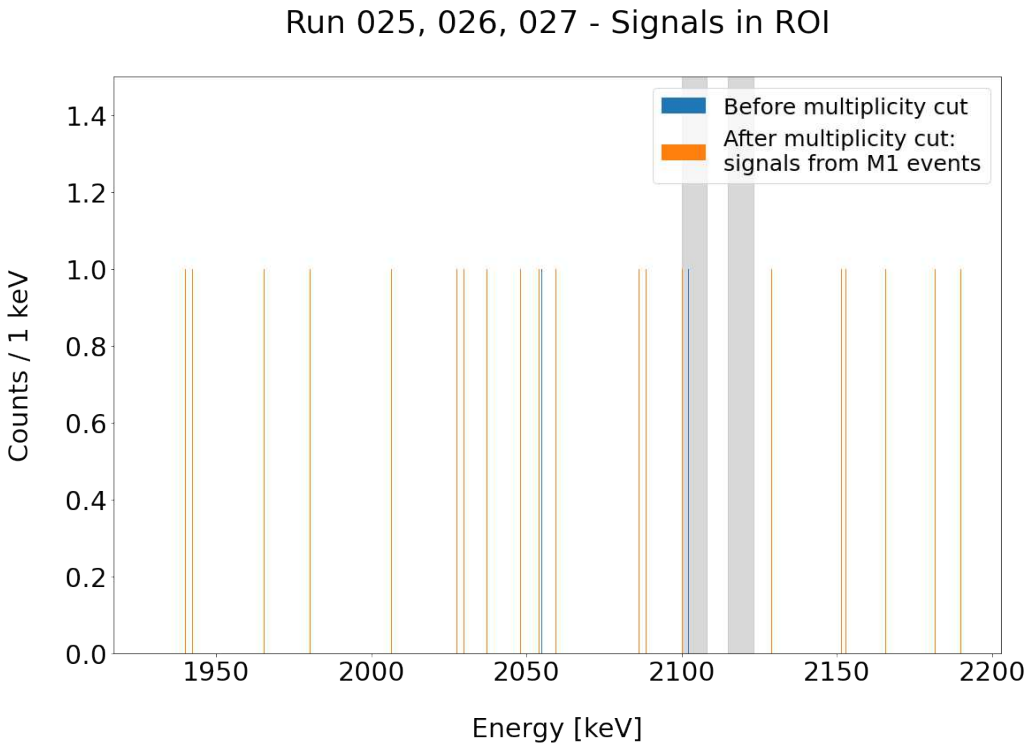


FIGURE 5.23: Signals appearing in the ROI before and after multiplicity cut. In particular, one of the signals originally appearing in the ROI is discarded by the multiplicity cut: in fact, it belonged to a $\mathcal{M}\geq 2$ event, namely it was acquired simultaneously to (at least) another above-threshold signal. As usual, the considered ROI spans form 1930 keV to 2190 keV; the energy regions in the two shaded bands are excluded from ROI due to the possible presence of signals from the ^{208}Tl (2104 keV) and ^{214}Bi (2119 keV) γ lines [32, 68].

5.5 Liquid Argon (LAr) veto

Germanium detectors need to be operated at cryogenic temperatures: for this reason they are immersed in a cryostat filled with liquid Argon. Besides its cooling function, and besides being an additional passive shield against external environmental radioactivity, the liquid Argon LAr is properly instrumented to act as an active veto for background signals, as described in Chapter 2.

Similarly to the muon veto, the LAr veto exploits the expected topology of $0\nu\beta\beta$ events, and in particular their high localization. The key idea of the LAr veto is that the energy release in a $0\nu\beta\beta$ event will be confined in the Germanium detector in which the decay takes place: then, if some energy is released both in a Germanium detector and in the liquid Argon, the process that caused that energy release is not a $0\nu\beta\beta$ decay. In practice, this means that if a signal is detected in coincidence in a Germanium detector and in the SiPMs reading out the LAr, the corresponding event is not a $0\nu\beta\beta$ decay, and therefore can be removed from the energy spectrum.

5.5.1 Definition of the veto condition

As in the previous steps of the analysis, we must determine an analytical condition to tag the events we want to remove from the energy spectrum, namely, in this case, the events *releasing energy both in the Germanium and in the LAr*. To do this effectively, we must take into account both the geometry of the apparatus and the possibility of having false coincidences among signals in the Germanium and in the LAr.

In fact, in principle we would remove from the energy spectrum all the signals collected by the Germanium detectors in coincidence with any signal detected by the SiPMs. In practice, though, this is not possible, due to the high rate of signals in the SiPMs, not always really associated to an event in Germanium.

An effective way to find a condition to identify physical Ge-SiPMs coincidences is discussed in [79]. The idea is to consider a class of signals acquired by Germanium detectors which should not produce any coincident signal in SiPMs. For example, the pulser signals can be considered: they are artificially injected in the Germanium detectors' DAQ system, and are not correlated to any signal in the SiPMs. The veto condition is then optimized maximizing the survival probability of pulser events and minimizing the acceptance of physical coincidences.

Several conditions have been tested by the Analysis Group of the Collaboration, considering both the geometrical distribution of the SiPMs and the amount of energy released in them.

The optimal veto condition, identified in [79], is that at least four different SiPMs channels acquire a non-zero energy, and the total energy acquired by the SiPMs in the considered event is at least 4 photoelectrons (p.e.):

$$\mathcal{N}_{\text{SiPMs}} \geq 4 \;\; \texttt{AND} \;\; E_{\text{tot}} \geq 4 \text{ p.e.} \tag{5.3}$$

All the events that *verify* this condition are removed from the energy spectrum.

The condition on the number of SiPM channels accounts for the fact that if a physical event happens and reaches both the Germanium and the SiPMs, it is probable for multiple SiPMs to acquire some energy; if instead only one SiPM acquires a non-zero signal, it is reasonable to assume that it comes from an internal event or discharge, not involving the Germanium detectors. The condition on the minimum energy is empirical and accounts for the fact that a zero-energy threshold in practice is not informative, due to statistical fluctuations of the signals.

The veto condition in Eq. (5.3) is chosen to grant a 95.2% pulser signals acceptance [79], while nicely suppressing the undesired peaks in the experimental energy spectrum, as discussed in next Section (Fig. 5.24, 5.25 and 5.26).

5.5.2 Results of LAr veto

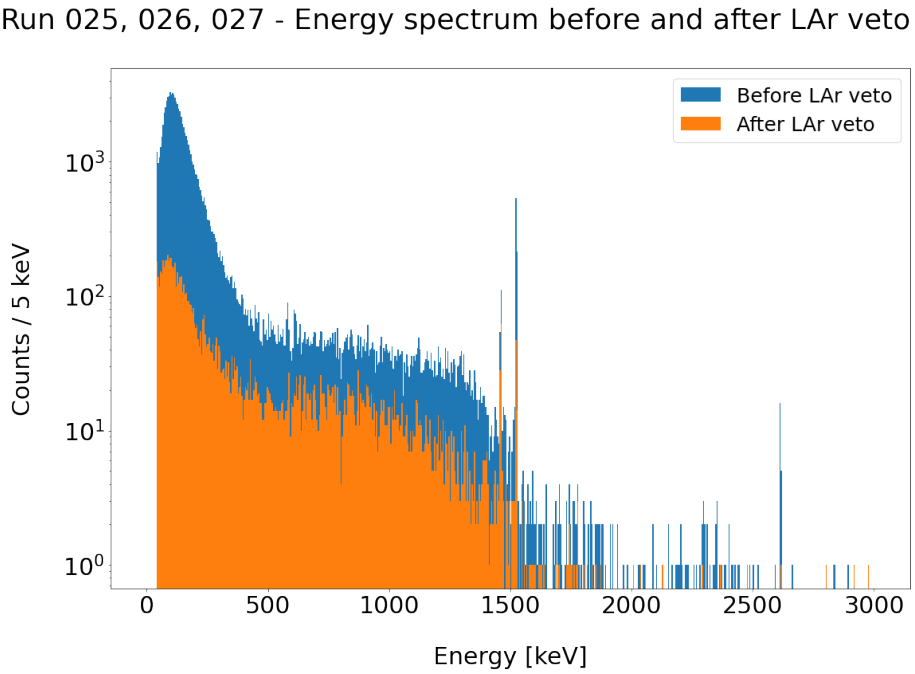


FIGURE 5.24: Energy spectrum from run 025, 026, 027, before and after LAr veto (zoom).

The impact of LAr veto on the acquired energy spectrum is clear from Fig. 5.24, showing the part of the spectrum in which almost all the signals are concentrated.

The ^{39}Ar distribution results primarily suppressed (note that the plot is in logarithmic scale on the y axis). This is consistent with the expectations: the source of ^{39}Ar is liquid Argon itself, therefore we expect these signals to be effectively detected by the LAr instrumentation.

The distribution in the range 500 - 1400 keV, which is the $2\nu\beta\beta$ region, appears strongly suppressed as well. It is crucial to underline that the removed signals *do not* come from the $2\nu\beta\beta$ distribution: analogously to the $0\nu\beta\beta$, also the $2\nu\beta\beta$ is a highly localized interaction, whose energy is fully absorbed within the volume of a single Germanium detector. Therefore, $2\nu\beta\beta$ events are absolutely not expected to release energy in the liquid Argon, and then are not expected to be vetoed in this step. The removed signals come instead from the Compton edge and continuum of the Potassium peaks: the γs emitted by ^{40}K and ^{42}K at 1461 keV and 1525 keV respectively can scatter on the electrons of the surrounding materials, lose some energy, and then be detected by the LAr instrumentation. In fact, note that the shape of the distribution before and after the LAr veto is slightly different: after the LAr it finally assumes the rounded shape typical of the $2\nu\beta\beta$ spectrum (see, for example, [32]). The distribution of these removed signals is instead shown in the residual spectrum (Fig. A.5).

The effects of the LAr veto on the three main γ lines still appearing in the spectrum, 1461 keV from ^{40}K, 1525 keV from ^{42}K and 2614 keV from ^{208}Tl FEP is shown more in detail in Fig. 5.25 and 5.26.

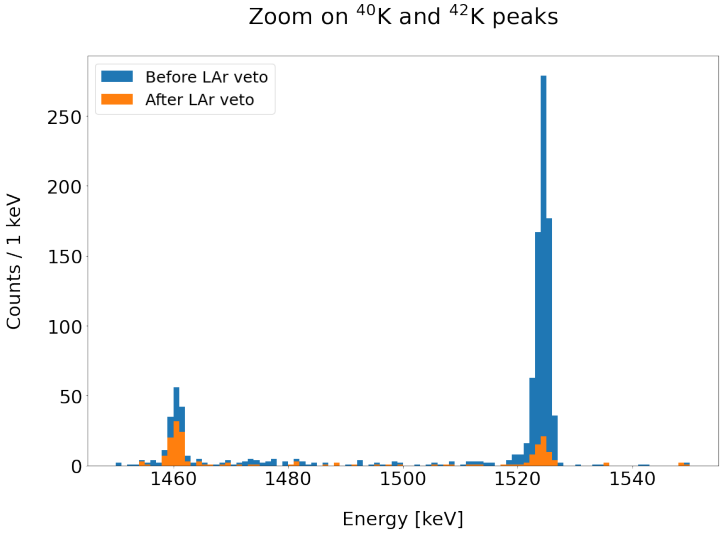


FIGURE 5.25: Zoom of the Potassium peaks energy region of Fig. 5.24 (1461 keV from ^{40}K and 1525 keV from ^{42}K).

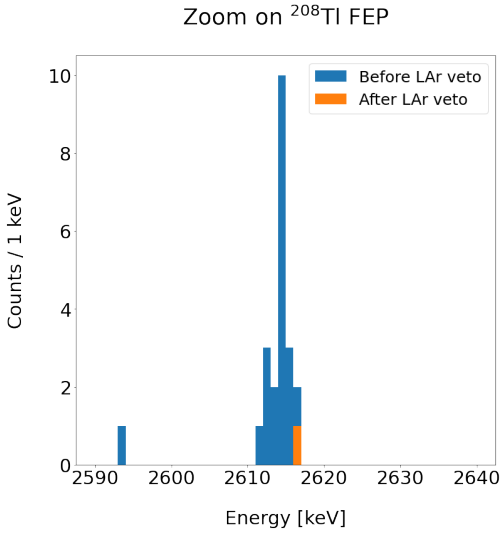


FIGURE 5.26: Zoom of the Thallium FEP energy region of Fig. 5.24 (2614 keV).

	^{40}K peak (1461 keV)	^{42}K peak (1525 keV)	^{208}Tl peak (2614 keV)
Number of signals before LAr veto	166	748	21
Number of signals after LAr veto	90	61	1
Survival probability	$\approx 54\%$	$\approx 8\%$	$\approx 4\%$

TABLE 5.12: Number of signals before and after LAr veto in the peaks plotted in Fig. 5.25 (^{40}K, ^{42}K) and 5.26 (^{208}Tl). The survival probability is calculated as ratio between the two numbers.

From the number of events in the peaks before and after LAr veto, we can estimate their survival probability (Tab. 5.12). In particular, for each peak we counted the events in an energy region of 10 keV centered in the nominal value of the peak, corresponding to an interval of about $\pm 2\cdot$ FWHM resolution of the detectors at the energy of the peaks [5].

In particular, we expected the LAr veto to be more effective in suppressing the ^{42}K peak than the ^{40}K peak, due to the different location of the two sources and to the different decay chain they face. In fact, ^{42}K is the progeny of ^{42}Ar's β decay, therefore it is produced directly in the liquid Argon and it is likely for the γs produced in its decay chain to be detected by the LAr instrumentation. Also, ^{42}K decays β to ^{42}Ca, which then can de-excite via multiple γ emissions (one of which is indeed the 1525 keV γ): it is then probable to have multiple energy depositions in coincidence in LAr (Fig 5.28).

^{40}K is instead present in most of the screening and construction materials (such as cables, front-end electronics, fiber shroud, mini-shroud, and detectors holders) [68], so the γ rays it emits do not always trigger the LAr instrumentation. Also, its decay chain (Fig. 5.27) does not produce many γ rays in coincidence, so it is harder for these signals to fulfill the veto condition.

Finally, ^{208}Tl's peak is effectively suppressed as expected, due again to the multiple γs emitted in coincidence in its decay chain (Fig. 5.29).

[5] The FWHM resolution of LEGEND's ICPC detectors spans from about 2 keV to 3.5 keV in the energy range containing the three considered peaks[58]

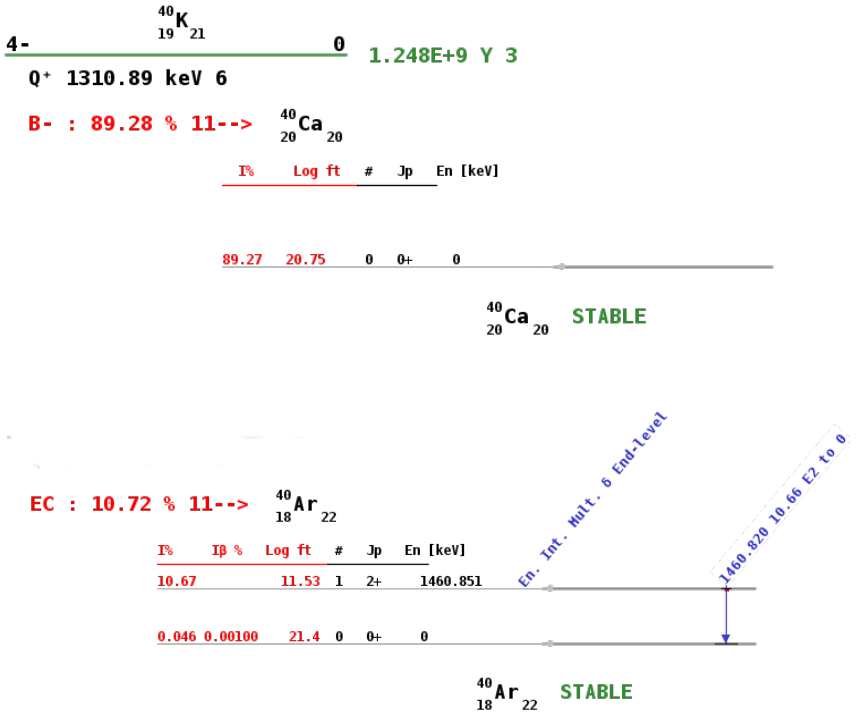


FIGURE 5.27: ^{40}K decay chain. Picture adapted from [80]

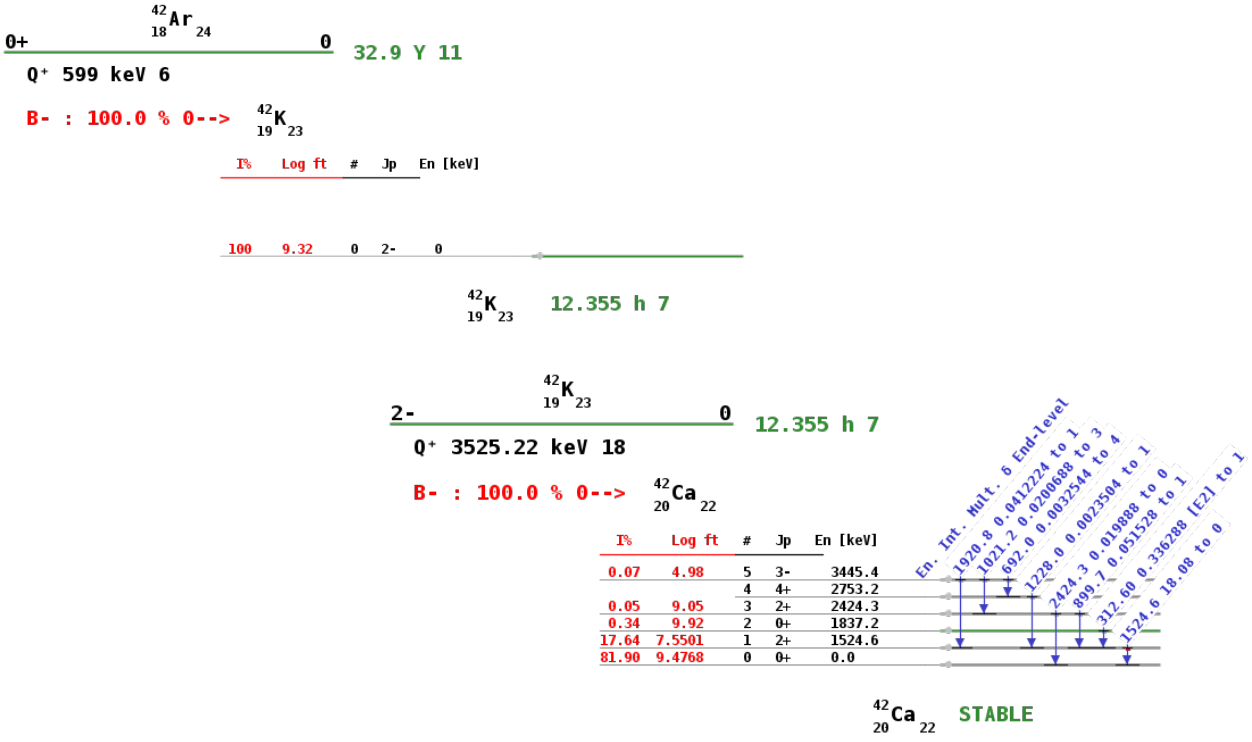


FIGURE 5.28: ^{42}K decay chain, including also its parent nucleus ^{42}Ar. Picture adapted from [80]

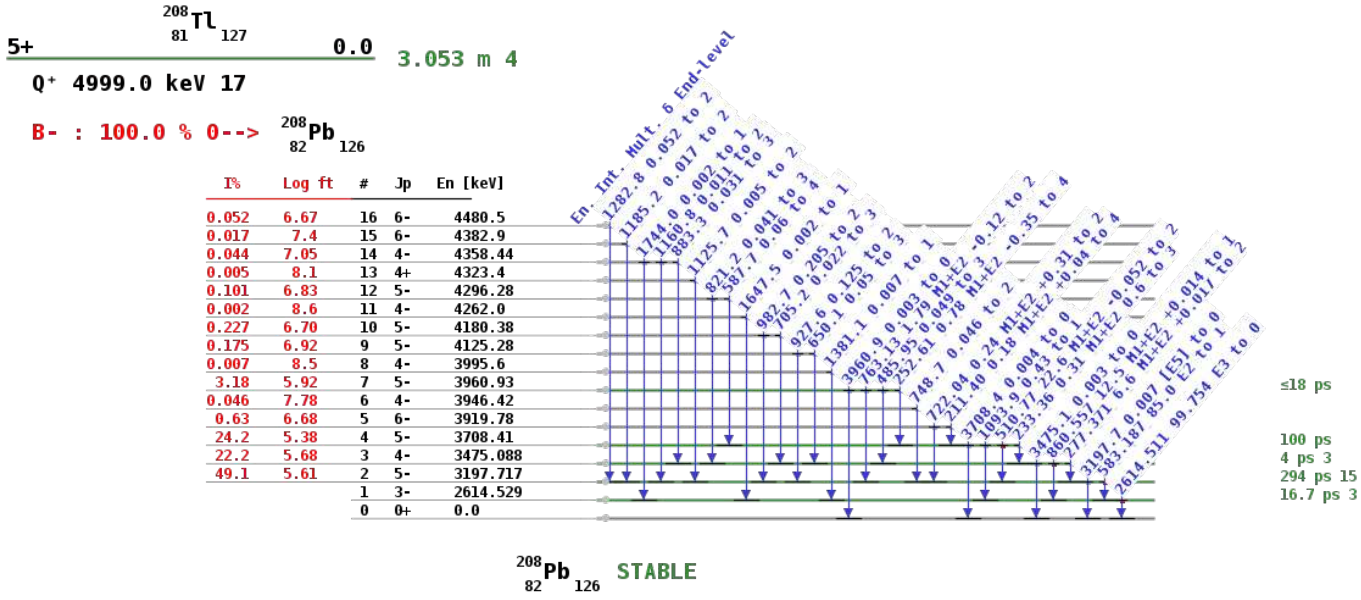


FIGURE 5.29: ^{208}Tl decay chain. Picture adapted from [80]

The overall effects of LAr veto on the acquired energy spectrum are summarized in Tab. 5.13, showing the total number of events that passed the cut. Fig. 5.30 shows the effects of the veto in the ROI: most of the signals are vetoed since coincident to a signal in LAr; only 4 signals remain after this cut.

	Number of signals before LAr veto	Number of signals after LAr veto	Surviving fraction (%)
Full energy range	86 289	8828	10.23%
ROI	19	4	21.05 %

TABLE 5.13: Number of signals before and after LAr veto. The surviving fraction is calculated as ratio between the two numbers, so it is referred only to the present cut. Two energy regions are considered: the full energy range of the spectrum and the ROI around $Q_{\beta\beta}$ for the $0\nu\beta\beta$ analysis.

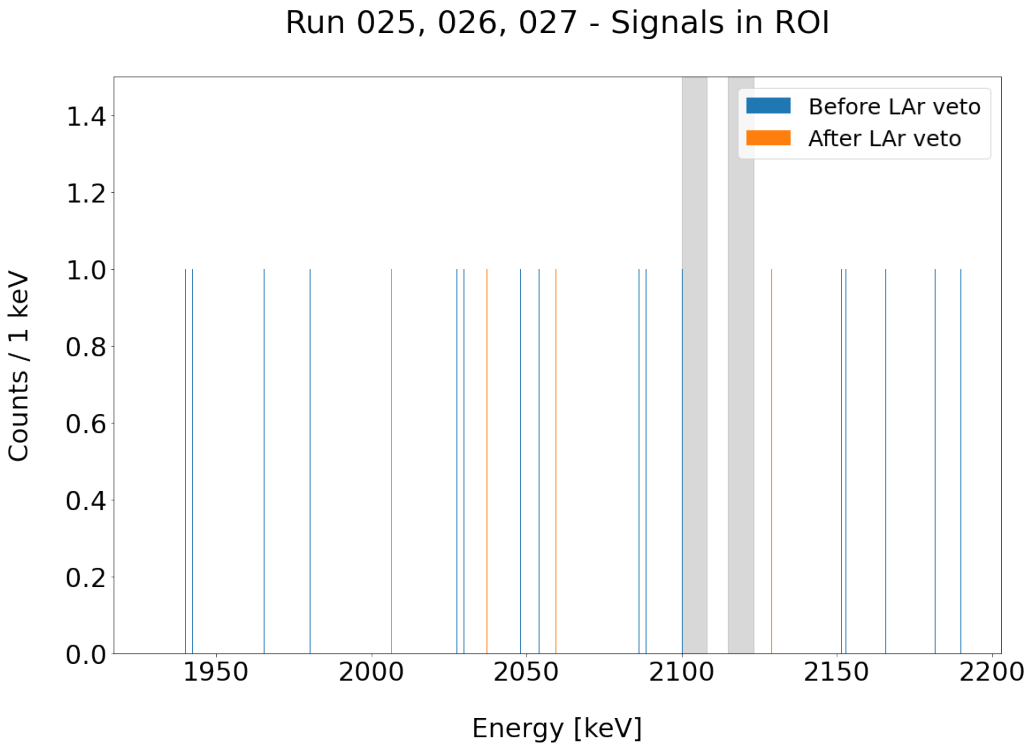


FIGURE 5.30: Signals appearing in the ROI before and after LAr veto. This cut is specifically designed to be effective in the Region Of Interest, and in fact many residual background signals are removed. As usual, the considered ROI spans form 1930 keV to 2190 keV; the energy regions in the two shaded bands are excluded from ROI due to the possible presence of signals from the ^{208}Tl (2104 keV) and ^{214}Bi (2119 keV) γ lines [32, 68].

5.6 Pulse Shape Discrimination (PSD)

The importance of studying the shape of the collected waveforms has been first highlighted when performing the quality cuts (Chapter 5.2). At that stage of the analysis we exploited the basic features of the time profile of physical signals to distinguish them from noise.

Now that noise and background contributions have been removed from our sample, the study of the time profile of the collected waveforms becomes again crucial: the last step of the background suppression strategy, referred to as Pulse Shape Discrimination (PSD), is based on the fact that the time profile of individual pulses provides precious information on the topology of the corresponding events, which allow to perform an accurate discrimination of signal-like events from background-like events.

The main references for this part of the analysis are the reviews of the pulse shape discrimination techniques employed in GERDA phase I [81] and II [82].

5.6.1 Definition of the PSD parameter

The key idea of the PSD is to discriminate among different classes of physical events based on the shape of the signals they produce in Germanium detectors. In particular, we exploit the fact that the high localization of the energy release in $0\nu\beta\beta$ events translates in a unique shape of the collected waveforms, different from the shape of the waveforms collected in case of events with different topology.

In practice, to perform this discrimination, we need to identify a parameter which manages to condense all the needed information about the shape of the waveform. A suitable parameter to achieve this result is A/E, defined as the ratio between the amplitude of the current pulse (A) and the energy deposited in the interaction (E):

$$\mathrm{A/E} = \frac{\text{Amplitude current pulse}}{\text{Energy}} \tag{5.4}$$

This parameter, by construction, contains detailed information on the shape of the waveform. In fact, the current pulse is the derivative of the charge pulse (which indeed is the waveform itself), therefore it is highly sensible to small variations of the time profile of the waveform, which as we pointed out is related to the topology of the underlying physical event. In particular, if we normalize the amplitude of the current pulse to the total energy of the signal, we remove the dependence on the absolute amplitude of the waveform, and retain only the information on the shape of the rising edge.

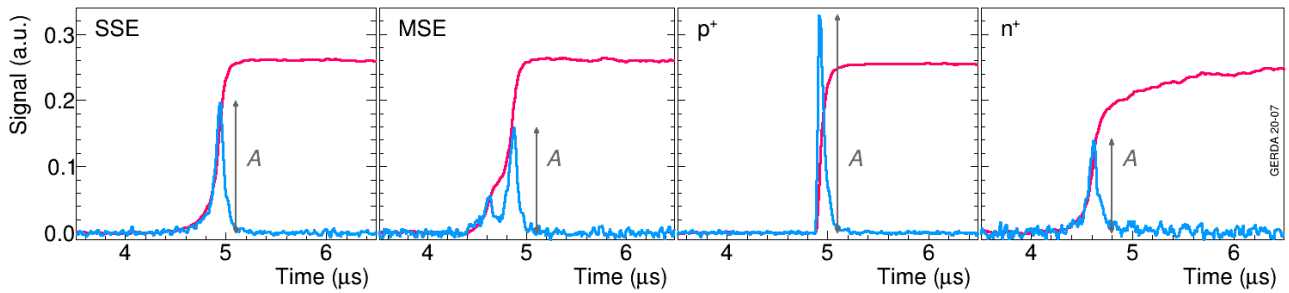


FIGURE 5.31: From left to right: Single Site Event (SSE), Multi Site Event (MSE), p^+ surface event and n^+ surface event. Picture from [82].

Fig. 5.31 shows the main types of waveforms that can be acquired, and their time derivative. Each of them come from a different class of physical events:

- **Single Site Events (SSE)**

  SSE are the good $0\nu\beta\beta$ candidates since the energy deposition is localized in time and space: the energy is deposited all at the same time, and in a specific point of the detector's volume. The produced ionization charge, then, drifts towards the electrodes as a compact cluster and when it is read out it produces a signal with a single sloped rising edge. Consequently, the current pulse has a single peak.

- **Multi Site Events (MSE)**

  MSE are events in which the energy is deposited in multiple parts of the detector's volume, or in (slightly) different moments ($\Delta t <$ acquired waveform length): different clusters of ionization charge then drift to the electrodes, producing a signal with different rising edges overlapping. The result is a rising edge with variable slope, translating to multiple peaks in the current pulse. Then, when normalized to the energy of the signals, the amplitude of the current pulse in case of a MSE is smaller than for a SSE (since the area of the distribution must be conserved): $(\mathrm{A/E})_{\mathrm{MSE}} < (\mathrm{A/E})_{\mathrm{SSE}}$.

- **p^+ surface events**

  In case of ICPC detectors, the p^+ surface is just the point-like electrode at the bottom of the detector. When some energy is released close to that electrode, the ionization charge is immediately read out, producing a fast signal with a steep rising edge, and consequently a high and sharp current peak. For this reason, $(\mathrm{A/E})_{\mathrm{p}^+} > (\mathrm{A/E})_{\mathrm{SSE}}$.

- **n^+ surface events**

  Events in which the energy is released close to the n^+ surface are usually α particles produced outside the detectors and immediately absorbed when they reach Germanium. In case of ICPC detectors, the n^+ surface is the whole lateral and upper surface of the detector. The ionization charge produced in n^+ surface events faces then a long drift time before being collected and it is possible for it not to be collected completely. For this reason the waveform can have a slow rise, translating to a low current peak. In this case, $(\mathrm{A/E})_{\mathrm{n}^+} < (\mathrm{A/E})_{\mathrm{SSE}}$.

$0\nu\beta\beta$ candidates are SSE, for which the following relation holds:

$$[(\mathrm{A/E})_{\mathrm{MSE}},\ (\mathrm{A/E})_{\mathrm{n}^+}] < (\mathrm{A/E})_{\mathrm{SSE}} < (\mathrm{A/E})_{\mathrm{p}^+} \tag{5.5}$$

Therefore, the defined A/E PSD parameter can be employed to identify SSE by performing a double-sided cut, removing both the events in which the A/E is outside an allowed region. The only good $0\nu\beta\beta$ candidates, retainde in the energy spectrum, are therefore events having

$$\texttt{AoE\_low\_cut} < \mathrm{A/E} < \texttt{AoE\_high\_cut} \tag{5.6}$$

What we need to do now is of course to identify the best thresholds to perform this cut on our data.

5.6.2 A/E cut optimization

The optimal A/E cut is the one that manages to isolate SSE, so as to remove from the energy spectrum the other types of events, while keeping the SSE sample untouched. The calibration spectrum can be employed to find a suitable upper-side and lower-side threshold to achieve this result, as discussed in [81, 82].

In particular, some of the peaks appearing in the calibration spectrum provide nearly pure samples of SSE and MSE, respectively (Fig. 5.32):

- 1593 keV ^{208}Tl DEP $\rightarrow$ SSE
- 1621 keV ^{212}Bi FEP, 2103 keV ^{208}Tl SEP and 2614 keV ^{208}Tl FEP $\rightarrow$ MSE

The goal is therefore to find a cut (namely a value of `AoE_low_cut` and `AoE_high_cut`) that maximizes the survival probability for the events in the SSE-like peak while minimizing it for the events in the MSE-like peaks. This optimization is done by the Analysis Group of the Collaboration and the results are sown in Fig. 5.32: the survival probability for SSE-like events is around 90% while for the MSE-like less than 20%.

Another energy region that deserves to be considered is the $Q_{\beta\beta}$ region. In the calibration spectrum, this region is dominated by the Compton continuum from the Thallium γ lines (Fig. 5.32): the survival probability of these events is around 40-50% (Fig. 5.32).

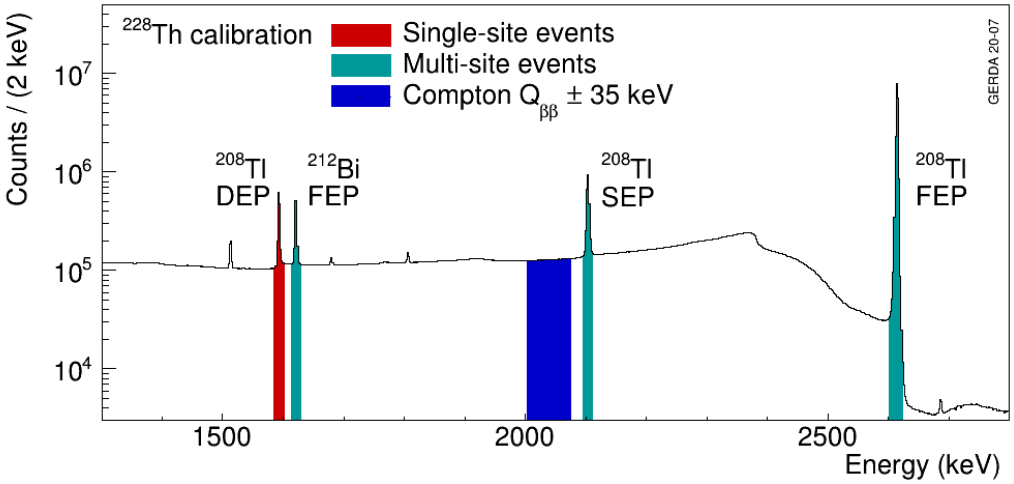


FIGURE 5.32: Example of calibration spectrum with main peaks color-coded by topology: SSE (red), MSE (teal) and events from the Compton continuum in the energy range arounf $Q_{\beta\beta}$ (blue). The effect of PSD on these signals is employed to optimize the cut, namely to maximize the SSE acceptance and minimize the MSE acceptance, while checking the impact on the events from the Compton continuum. Picture from [82].

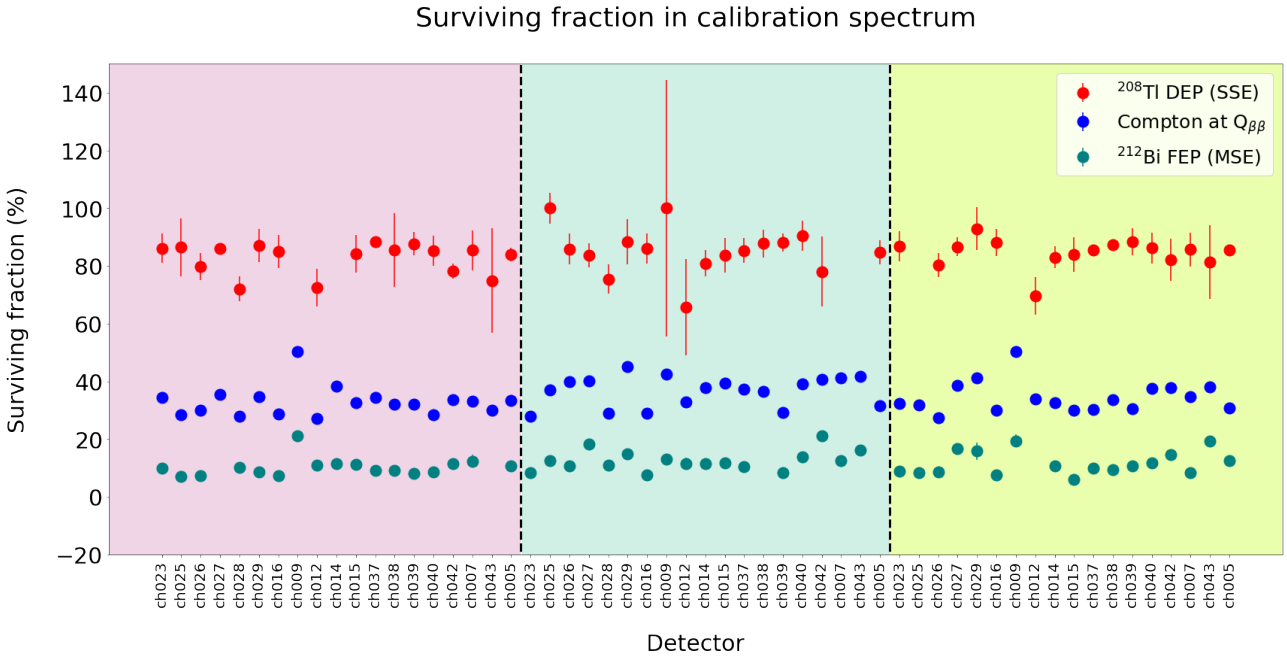


FIGURE 5.33: Surviving fraction of the SSE-like events from 1593 keV ^{208}Tl DEP (red), Compton continuum at $Q_{\beta\beta}$ (blue) and 1621 keV ^{212}Bi FEP (teal), obtained with the optimized PSD parameters. Since the PSD cut thresholds are optimized for each run and for each detector, the surviving fractions are plotted correspondingly. Results are grouped by run: left run 025, center run 026 and right run 027.

The parameter employed for the optimization is actually not the raw A/E, but the A/E Classifier ζ defined as:

$$\zeta = \frac{\frac{\mathrm{A/E}}{\mu_{\mathrm{A/E}}(E)} - 1}{\sigma_{\mathrm{A/E}}(E)} \tag{5.7}$$

In which $\mu_{\mathrm{A/E}}$ is the mean value of the A/E distribution and $\sigma_{\mathrm{A/E}}$ its standard deviation. By construction, the distribution of the A/E Classifier is centered in zero and has a standard deviation equal to one for SSE.

Therefore, in practice the employed PSD cut is:

$$\texttt{AoE\_Classifier\_low\_cut} < \zeta < \texttt{AoE\_Classifier\_high\_cut} \tag{5.8}$$

Note that the optimization is performed run per run, and detector per detector, meaning that for each run and for each detector the values of `AoE_Classifier_low_cut` and `AoE_Classifier_high_cut` that grants the desired survival fractions are different.

The distribution of the ζ parameter before and after PSD cut is shown in Fig. 5.34.

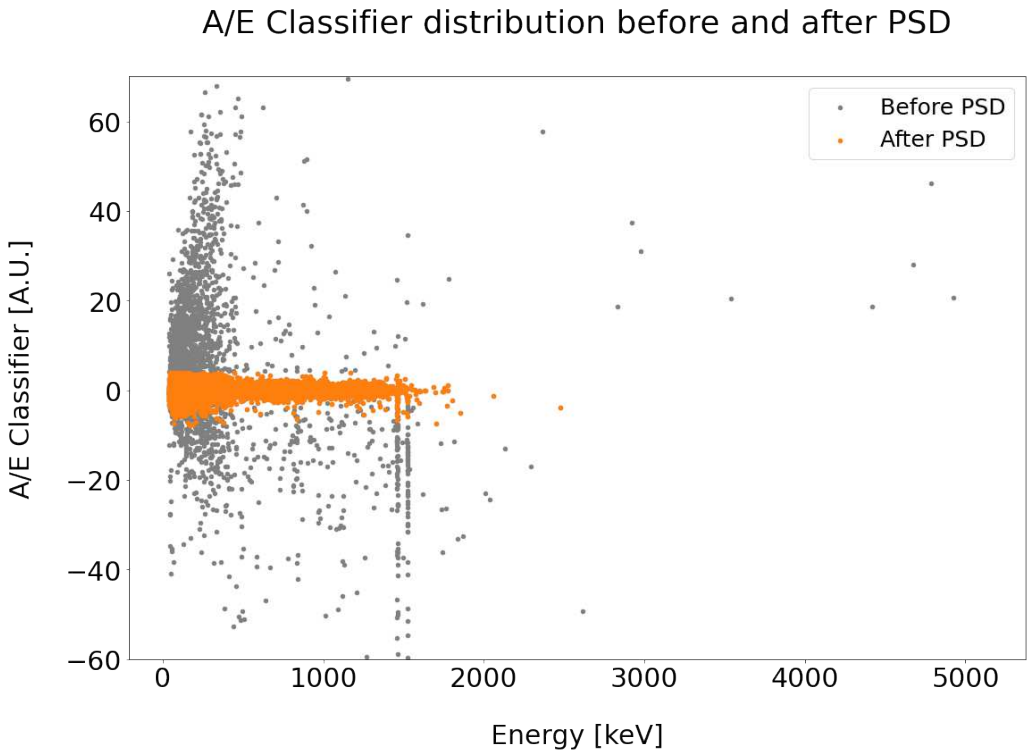


Figure 5.34: Distribution of the PSD parameter ζ before and after PSD double-sided cut. As expected, the distribution is centered at $\zeta = 0$ and the parameter acquires values close to zero for SSE. The two lines at an energy of about 1500 keV are the events from the ^{40}K and ^{42}K lines.

5.6.3 Results of PSD

As shown in Fig. 5.35, the PSD cut manages to further suppress the ^{39}Ar distribution, the ^{40}K and ^{42}K peaks and most of the residual background in the $Q_{\beta\beta}$ region. As expected, the $2\nu\beta\beta$ region is instead not sensibly affected by the PSD cut: in fact, also in $2\nu\beta\beta$ events the energy release is highly localized, so their waveforms are not expected to have different features from those of $0\nu\beta\beta$ events, therefore the performed PSD analysis is not expected to distinguish them.

The number of events before and after PSD cut is presented in Tab. 5.14. Note that after PSD cut only one events survives in the Region Of Interest (Fig. 5.36).

Finally, the spectrum of the events removed by PSD cut is shown in Fig. A.6.

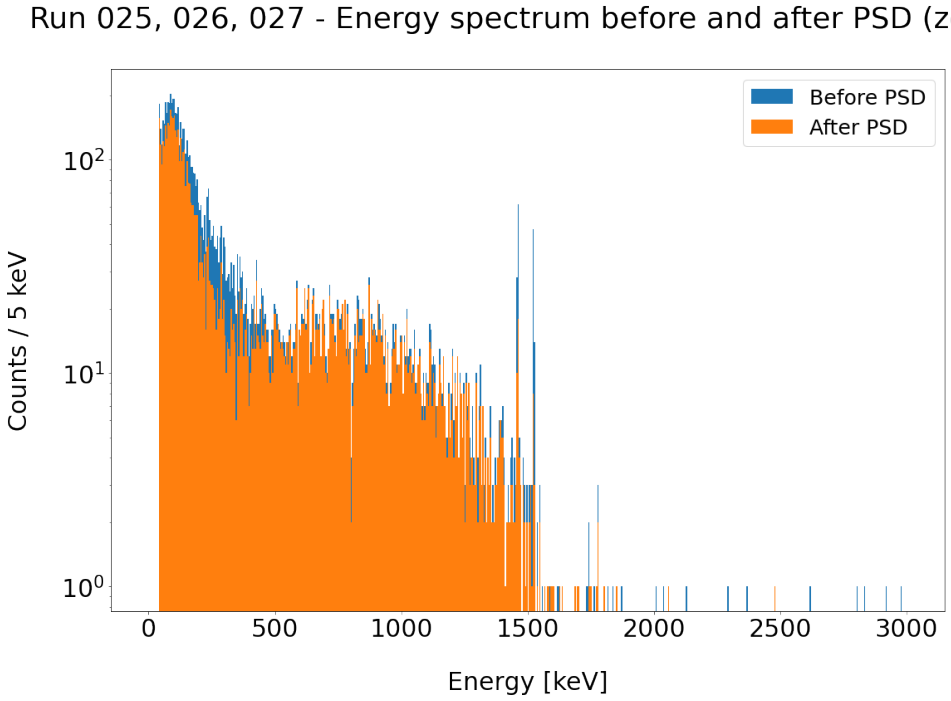


FIGURE 5.35: Energy spectrum from run 025, 026, 027, before and after PSD (zoom).

	Number of signals before PSD	Number of signals after PSD	Surviving fraction (%)
Full energy range	8828	6807	77.10%
ROI	4	1	25.00 %

TABLE 5.14: Number of signals before and after PSD cut. The surviving fraction is calculated as ratio between the two numbers, so it is referred only to the present cut. Two energy regions are considered: the full energy range of the spectrum and the ROI around $Q_{\beta\beta}$ for the $0\nu\beta\beta$ analysis.

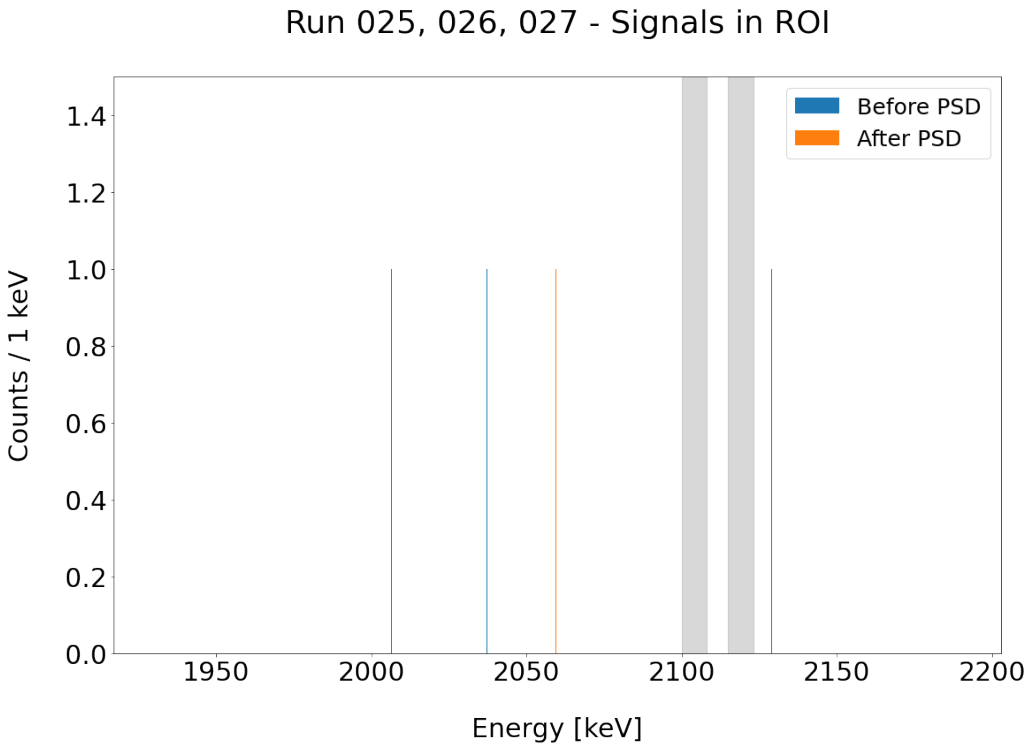


FIGURE 5.36: Signals appearing in the ROI before and after PSD cut. This cut is specifically designed to be effective in the Region Of Interest, and in fact just one signal survives in the ROI after this last cut. As usual, the considered ROI spans form 1930 keV to 2190 keV; the energy regions in the two shaded bands are excluded from ROI due to the possible presence of signals from the ^{208}Tl (2104 keV) and ^{214}Bi (2119 keV) γ lines [32, 68].

5.7 Results of the analysis cuts

Let us now summarize the steps of the performed analysis. We started from the raw energy spectrum acquired by Germanium detectors during runs 025, 026 and 027, shown in Fig. 4.6. Then we implemented a series of analysis cuts aimed at removing:

- Artificial signals injected to monitor the performances of the system → Pulser Signal Removal (Section 5.1);
- Noise signals → Quality Cuts (Section 5.2);
- Background signals, coming from physical processes different from $\beta\beta$ decay → Muon veto (Section 5.3), multiplicity cut (Section 5.4), LAr veto (Section 5.5) and Pulse Shape Discrimination (Section 5.6)

The joint effect of all the performed cuts on the collected energy spectrum is shown in Fig. 5.37: the top panel shows the full range energy spectrum (0 - 10000 keV energy range), before and after the analysis cuts; the bottom panel shows a zoom of that spectrum (40 - 3000 keV energy range), in which most of the structures of interest are concentrated and including of course the ROI.

The zero-energy noise peak and the pulser peaks are completely removed. The background suppression is also remarkable: at low energy, below 500 keV, the ^{39}Ar distribution is strongly suppressed; in the 500 - 1400 keV energy range, the Compton continuum from the Potassium peaks is suppressed, revealing the shape of the underlying $2\nu\beta\beta$ distribution; the 1461 keV peak from ^{40}K and 1525 keV peak from ^{42}K are strongly suppressed too, and so is the 2614 keV peak from ^{208}Tl; finally, the high energy background and the signals form α radioactivity are completely suppressed.

In the region of interest, background signals are successfully suppressed, too: out of the 38 signals originally appearing in the spectrum, only one, having energy $E = 2059.66$ keV, survives all the analysis cuts (Fig. 5.38).

Tab. 5.15 collects the number of signals surviving in the energy spectrum after each cut. Two energy ranges are considered: the full energy range and the ROI.

Finally, Fig. 5.39 shows the energy spectrum after all the analysis cuts, only. Note that in this case the range 0 - 3000 keV is actually the full energy range, since all the signals at higher energy have been removed. The main contributions are labeled in the spectrum: starting from the low energy side, we observe the continuous distributions of ^{39}Ar and of $2\nu\beta\beta$, and then the γ lines, and in particular the 1461 keV line from ^{40}K, ant the 1525 keV line from ^{42}K; the 1764 keV line from ^{214}Bi was embedded in other background signals and becomes visible only after background suppression. Instead, the 2314 keV line from ^{208}Tl , clearly visible in the original spectrum, is nearly completely suppressed by the analysis cuts. The peak identification is based on [32].

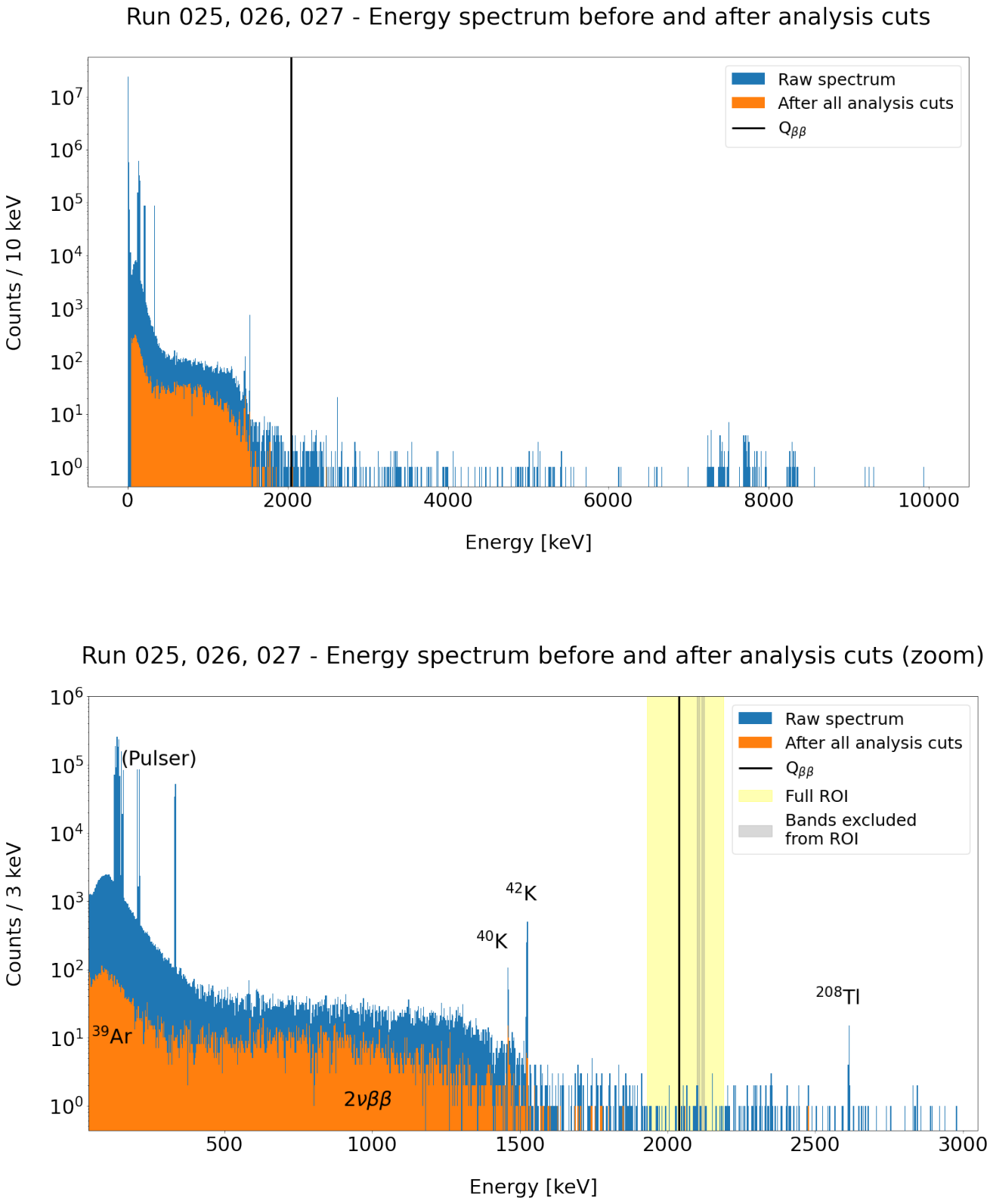


FIGURE 5.37: Energy spectrum before and after all the analysis cuts: TOP: full energy range; BOTTOM: zoom in the energy range where most of the structures of interest are concentrated, including ROI and the main gamma lines, identified according to [32].

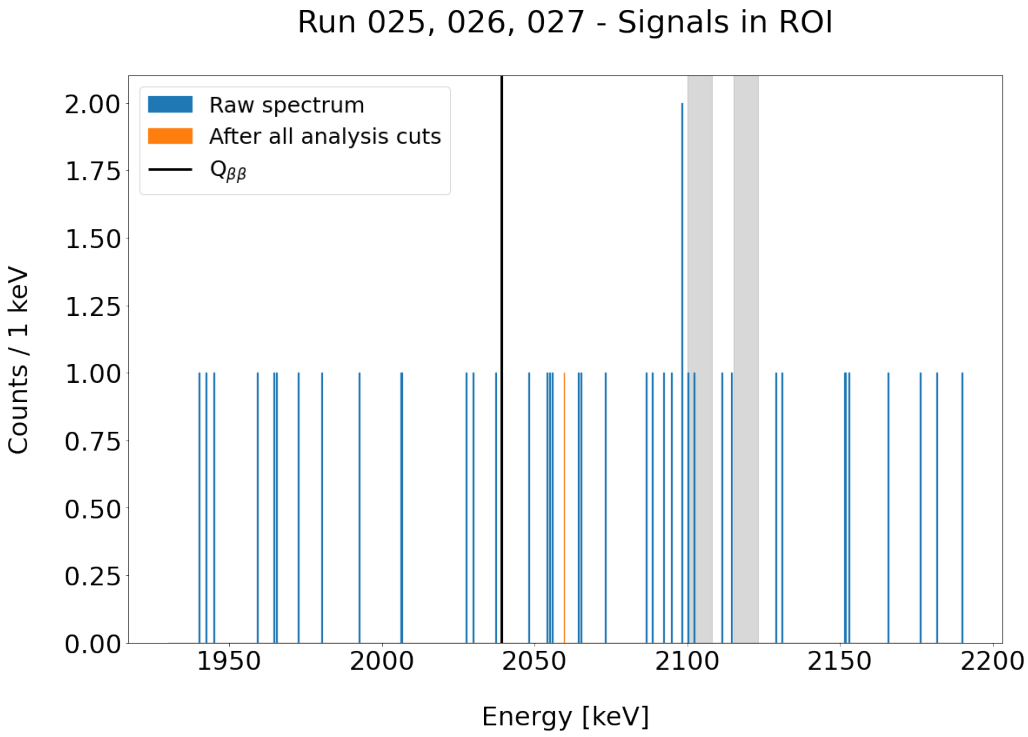


FIGURE 5.38: Signals appearing in the ROI before and after all the analysis cuts. As usual, the considered ROI spans form 1930 keV to 2190 keV; the energy regions in the two shaded bands are excluded from ROI due to the possible presence of signals from the ^{208}Tl (2104 keV) and ^{214}Bi (2119 keV) γ lines [32, 68].

	Total number of raw signals	PSR	QC	Muon veto	Multiplicity cut	LAr veto	PSD	Surviving fraction (%)
Full energy range	29 738 052	27 253 386	233 351	180 667	86 289	8828	6807	0.029%
ROI	38	38	22	20	19	4	1	2.63 %

TABLE 5.15: Summary of the impact of each performed analysis cut in terms of number of signals survived at each step. The surviving fraction is calculated at the end of the analysis cuts procedure (after PSD), with respect to the original total number of row signals.

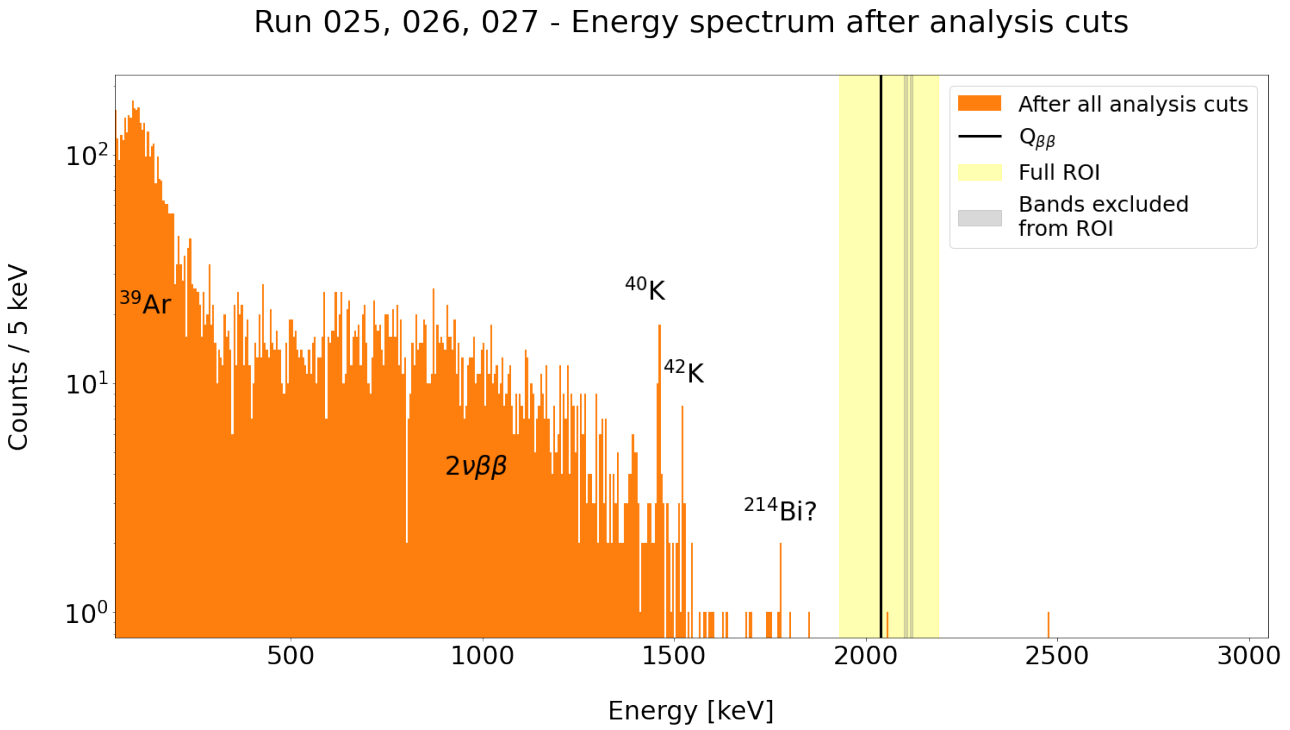


FIGURE 5.39: Full energy spectrum after analysis cuts from run 025, 026 and 027. The γ lines are identified according to [32].

This energy spectrum is the starting point for any quantitative analysis of $2\nu\beta\beta$ and eventually $0\nu\beta\beta$ decay properties. A proper analysis of the $2\nu\beta\beta$ distribution aimed at the extraction of $T^{2\nu}_{1/2}$ is out of the scope of this work. Instead, we focus on the distribution of the signals appearing in the $0\nu\beta\beta$ ROI: the goal is to give an estimate of the signal and background contribution in that energy region. In particular, with the available exposure, we do not expect to observe any $0\nu\beta\beta$ signal. Therefore, our goal is to give an estimate of the background index $B.I.$ in the ROI and eventually to put an upper limit to the strength of the $0\nu\beta\beta$ signal $\mathcal{S}$ based on the available data.

Chapter 6

Bayesian analysis of signal and background in $0\nu\beta\beta$ ROI

In order to extract information on the achieved background index $B.I.$ and eventually on the strength of the $0\nu\beta\beta$ signal $\mathcal{S}$, a Bayesian analysis of the distribution of the signals appearing in ROI is performed. Unless noted otherwise, the reference for the analysis presented in this Chapter is [50].

6.1 The model

The parameters considered in the analysis are the following:

- **Strength of a possible $0\nu\beta\beta$ signal ($\mathcal{S}$)**

  $\mathcal{S}$ is defined as $\mathcal{S} = 1/T^{0\nu}_{1/2}$.

  The expected number of $0\nu\beta\beta$ events as a function of $\mathcal{S}$ is modeled as:

  $$\mu_S = \log(2)\,\frac{N_A}{m_{Ge}}\,\epsilon\,\mathcal{E}\,\mathcal{S} \tag{6.1}$$

  In which:

  - $N_A = 6.022\,140\,76 \times 10^{23}$ mol^{-1} is the Avogadro number.
  - $m_{Ge} = 75.65 \times 10^{-3}$ kg/mol is the molar mass [83].
  - ϵ is the signal detection efficiency, taking into account the enrichment fraction, the detector active volume and the probability that an eventual $0\nu\beta\beta$ is detected, is successfully analyzed and appears in ROI. Note that this value is not directly known for LEGEND detectors. This point will be discussed when performing the calculations.
  - $\mathcal{E} = 2.248$ kg yr is the total exposure.

- **Background Index ($B.I.$)**

  The expected number of background events as a function of $B.I.$ is modeled as:

  $$\mu_B = \mathcal{E}\,\Delta E\,B.I. \tag{6.2}$$

  In which:

  - $\mathcal{E} = 2.248$ kg yr is again the total exposure.
  - $\Delta E = 240$ keV is the net amplitude of the ROI, which spans from 1930 keV to 2190 keV with the exclusion of the two 10 keV bands in correspondence of the ^{208}Tl and ^{214}Bi peaks.

Let us call $\mathcal{D}$ the dataset made up of the experimental signals appearing in the ROI. In general, the likelihood function describing the probability of obtaining such dataset given a certain set of parameters $\mathcal{S}$ and $B.I.$ can be parametrized as in Eq. (6.3): the model is built assuming a flat distribution for the background and a Gaussian distribution for the signal, both weighted with the Poisson term appearing as prefactor.

$$\mathcal{L}(\mathcal{D}|\mathcal{S}, B.I., \theta) = \frac{e^{-(\mu_s+\mu_B)}(\mu_S+\mu_B)^{N_{\text{obs}}-1}}{N_{\text{obs}}!} \cdot \prod_{j=1}^{N_{\text{obs}}} \mu_B \frac{1}{\Delta E} + \mu_S \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{(E_j - Q_{\beta\beta})^2}{2\sigma^2}} \tag{6.3}$$

In which:

- μ_S and μ_B are the expected number of signal and background events, respectively, as defined in Eq. (6.1) and (6.2).
- N_{obs} is the number of signals appearing in $\mathcal{D}$.
- $\sigma = \frac{FWHM}{2\sqrt{2\log(2)}}$ is the energy resolution.
- $\theta = \{\epsilon, \sigma\}$ are the nuisance parameters assumed to be known in the model.

The likelihood function in Eq. (6.3) is built starting from the model presented in [50], considering just one dataset $\mathcal{D}$ and assuming no systematic energy offsets in the acquired data.

In our case, the dataset $\mathcal{D}$ is made up of the signals appearing in the ROI after all the analysis cuts: therefore, in practice, it contains only one signal ($N_{obs} = 1$), having energy $E = 2059.66$ keV. In this case, the analytical expression of the likelihood is simplified:

$$\mathcal{L}(\mathcal{D}|\mathcal{S}, B.I., \theta) = e^{-(\mu_s+\mu_B)} \cdot \left(\mu_B \frac{1}{\Delta E} + \mu_S \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{(E - Q_{\beta\beta})^2}{2\sigma^2}}\right) \tag{6.4}$$

Within a Bayesian analysis framework, we can compute $\mathcal{P}(\mathcal{S}, B.I.|\mathcal{D}, \theta)$ (Eq. (6.5)), the posterior PDF (probability density function) of the two unknown parameters $\mathcal{S}$ and $B.I.$, as the product of the likelihood $\mathcal{L}(\mathcal{D}|\mathcal{S}, B.I., \theta)$ (Eq. (6.4)) and the prior distributions $\mathcal{P}(\mathcal{S})$ and $\mathcal{P}(B.I.)$ of the unknown parameters:

$$\mathcal{P}(\mathcal{S}, B.I.|\mathcal{D}, \theta) \propto \mathcal{L}(\mathcal{D}|\mathcal{S}, B.I., \theta) \cdot \mathcal{P}(\mathcal{S})\mathcal{P}(B.I.) \tag{6.5}$$

In particular, for the signal strength $\mathcal{S}$ we consider a uniform prior between 0 and 10^{-15} yr^{-1} (Eq. (6.6)). The choice of the range is highly conservative. The current best lower bound on $T^{0\nu}_{1/2}$ for ^{76}Ge is of $\mathcal{O}(10^{26})$ yr [32], which translates to a signal upper bound $\mathcal{S} = 1/T^{0\nu}_{1/2}$ of $\mathcal{O}(10^{-26})$ yr^{-1}; though, since we do not expect to have the same sensitivity as GERDA, due to the lower exposure, we impose a much looser bound, namely we allow the parameter $\mathcal{S}$ to span in a wider range of values. Note that the case $\mathcal{S} = 0$ yr^{-1}, namely $T^{0\nu}_{1/2} = +\infty$, accounts for the case of $0\nu\beta\beta$ not occurring.

$$\mathcal{P}(\mathcal{S}) \propto \begin{cases} 1 & \text{for } \mathcal{S} \in [0, 10^{-15}] \text{ yr}^{-1} \\ 0 & \text{otherwise} \end{cases} \tag{6.6}$$

For the background index $B.I.$ we consider a uniform prior between 0 and 0.1 (keV kg yr)$^{-1}$. Again, the choice of the range is highly conservative. A rough reference value for the upper limit of the $B.I.$ can be extracted from its definition, $B.I. = \frac{\mu_B}{\mathcal{E}\Delta E}$, assuming that all the acquired signals contribute to the background, namely in our case, $\mu_B = 1$: in this way we get $B.I. = \frac{1}{240 \text{ keV} \cdot 2.248 \text{ kg yr}} = \mathcal{O}(10^{-3})$.

$$\mathcal{P}(B.I.) \propto \begin{cases} 1 & \text{for } B.I. \in [0, 0.1] \text{ (keV kg yr)}^{-1} \\ 0 & \text{otherwise} \end{cases} \tag{6.7}$$

The one-dimensional posteriors $\mathcal{P}(\mathcal{S}|\mathcal{D},\theta)$ and $\mathcal{P}(B.I.|\mathcal{D},\theta)$ can be extracted via marginalization of the total posterior $\mathcal{P}(\mathcal{S},B.I.|\mathcal{D},\theta)$ (Eq. (6.5)).

Note that all the parameters appearing in the posterior are known, except for the signal efficiency of the detectors ϵ. In order to properly extract this value, a simulation of the performances of the full apparatus should be performed. This simulation, though, has not been performed for LEGEND yet. Instead, the full simulation had been performed for the GERDA setup: in particular, the estimated efficiency for ICPC Germanium detectors was $\epsilon = (66.0 \pm 1.8)\%$ [32]. Since LEGEND's ICPCs are built with the same criterion of GERDA's ICPC, we consider reasonable to assume that their efficiency should be comparable. Of course, when the full LEGEND simulation will be performed, this analysis must be updated to the proper LEGEND efficiency values.

In practice, we consider the efficiency ϵ as an additional unknown nuisance parameter in the likelihood, and consequently in the posterior, having a Gaussian prior, centered in GERDA's result:

$$\mathcal{P}(\epsilon) \propto \begin{cases} \frac{1}{\sqrt{2\pi}\sigma_G} e^{\frac{(\epsilon-\epsilon_G)^2}{2\sigma_G^2}} & \text{for } \epsilon \in [0,1] \\ 0 & \text{otherwise} \end{cases} \tag{6.8}$$

In which $\epsilon_G = 0.66$ and $\sigma_G = 0.018$ are the results form GERDA. Note that, being an efficiency, it is defined only in the [0,1] interval, namely the probability is zero out of that range.

The full posterior becomes therefore:

$$\mathcal{P}(\mathcal{S},B.I.|\mathcal{D},\theta) \propto \mathcal{L}(\mathcal{D}|\mathcal{S},B.I.,\theta) \cdot \mathcal{P}(\mathcal{S})\mathcal{P}(B.I.)\mathcal{P}(\epsilon) \tag{6.9}$$

6.2 Running the MCMC

The computation and marginalization of the posterior is performed via Markov Chain Monte Carlo (MCMC). The key idea of the MCMC algorithm is to sample from the posterior distribution so as to get an approximated result for the posterior itself. The theory supporting MCMC methods is discussed for example in [84].

The numerical implementation of the MCMC is performed employing the `emcee` Python library [85, 86].

The considered posterior $\mathcal{P}(\mathcal{S},B.I.|\mathcal{D},\theta)$ is the function presented in Eq. (6.9), built as a product of $\mathcal{L}(\mathcal{D}|\mathcal{S},B.I.,\theta)$ (Eq. (6.4)), $\mathcal{P}(\mathcal{S})$ (Eq. (6.6)), $\mathcal{P}(B.I.)$ (Eq. (6.7)) and $\mathcal{P}(\epsilon)$ (Eq. (6.8)). Note that `emcee` requires the log-likelihood and the log-priors: these functions are computed analytically as $\log(\mathcal{L}(\mathcal{D}|\mathcal{S},B.I.,\theta))$ and $\log(\mathcal{P}(\mathcal{S}))$, $\log(\mathcal{P}(B.I.))$ and $\log(\mathcal{P}(\epsilon))$ respectively.

The parameters' space is of course tri-dimensional, and the sampled parameters are sets $p_i = \{\mathcal{S}_i, B.I._i, \epsilon_i\}$.

We run 16 chains, with a random starting point in the allowed parameters' space, sampling 100 000 sets of parameters each, for a total of 1 600 000 samples. The obtained chains are shown in Fig. 6.1. A burn-in period of 250 samples is removed; a thinning factor of 30 is introduced to mitigate the autocorrelation effects of the sampled parameters. The generated chains appear to converge nicely to the same unimodal distribution. The resulting distribution is showed in Fig. 6.2: in the bottom-left corners, the $\{\mathcal{S}_i, B.I._i\}$, $\{\mathcal{S}_i, \epsilon_i\}$ and $\{B.I._i, \epsilon_i\}$ couples sampled from the bi-dimensional parameters' space, and on the diagonal the the projected one-dimensional marginalized posteriors for $\mathcal{S}$, $B.I.$ and ϵ.

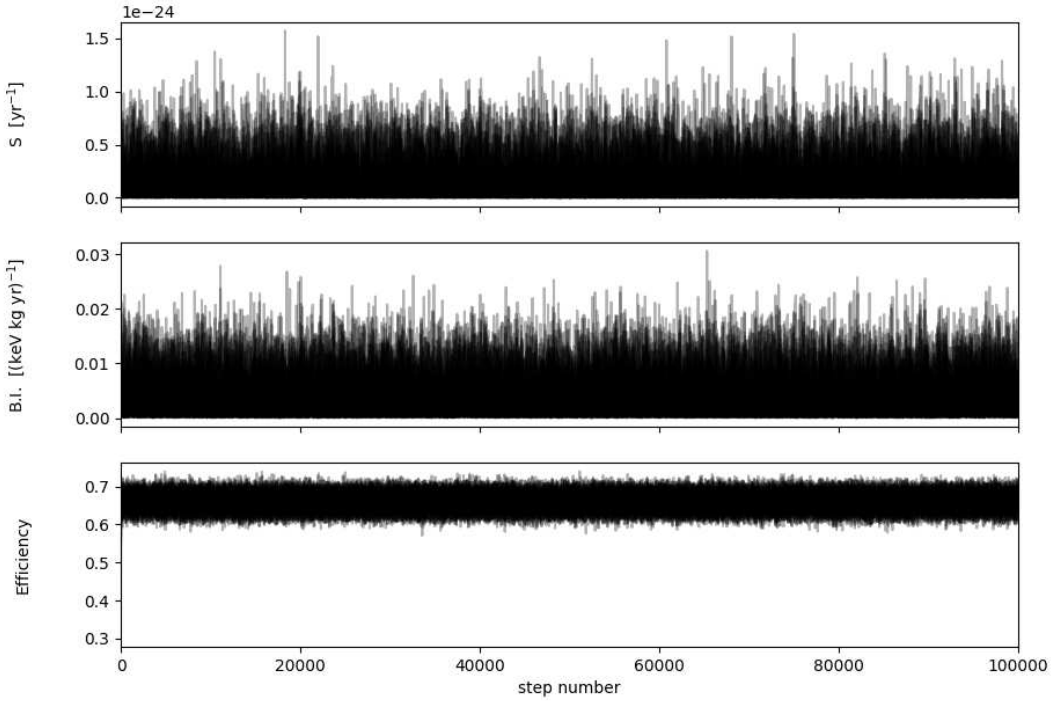


FIGURE 6.1: MCMC chains of sampled parameters.

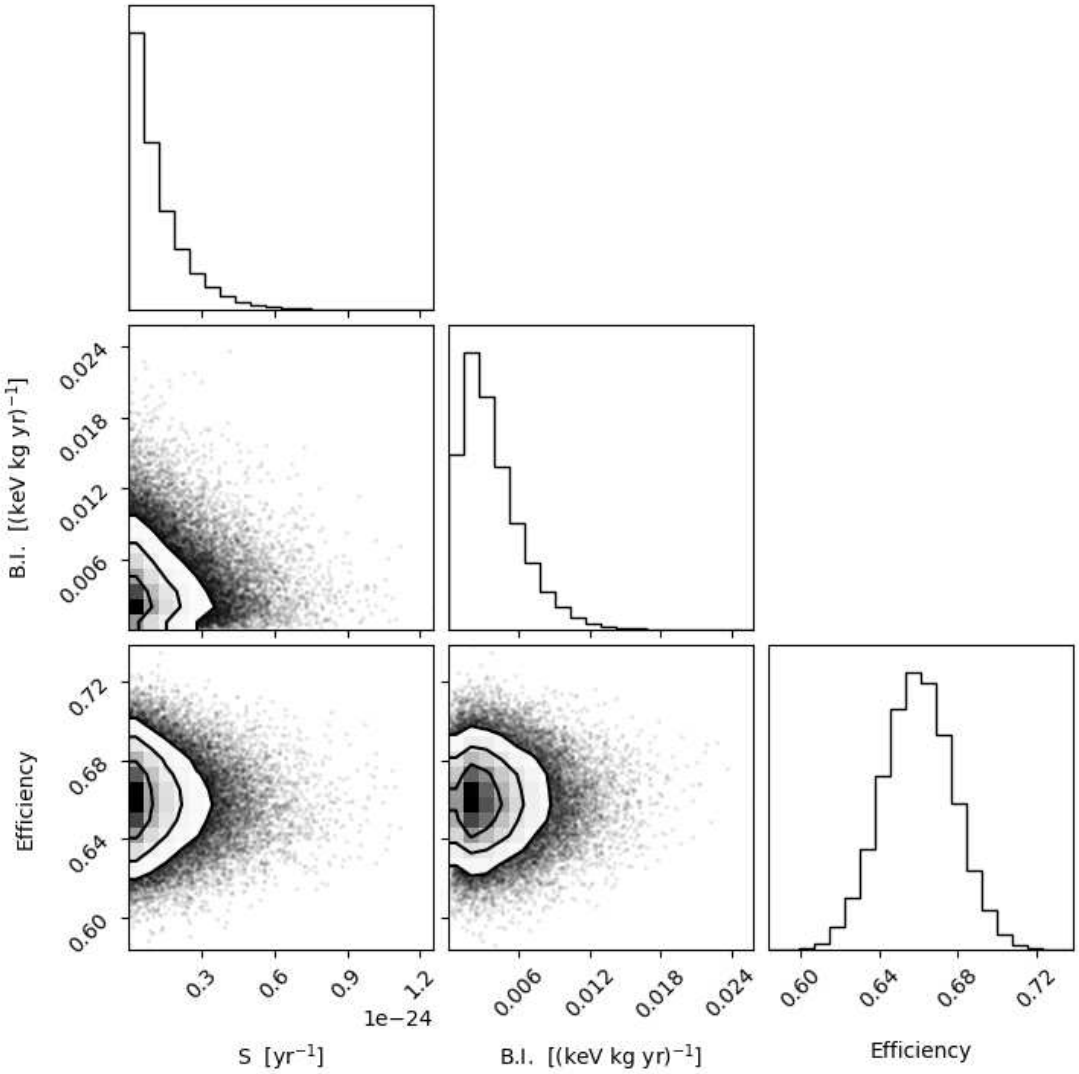


FIGURE 6.2: Corner-plot of the sampled parameters. In the bottom-left panels there are the sampled couples in the bi-dimensional parameters' space; on the diagonal the projected marginalized distribution, each in its one-dimensional parameters' space.

6.3 Results of the Bayesian analysis

Let us now focus on the marginalized posterior for $\mathcal{S}$ and $B.I.$, in order to give an estimate of these parameters.

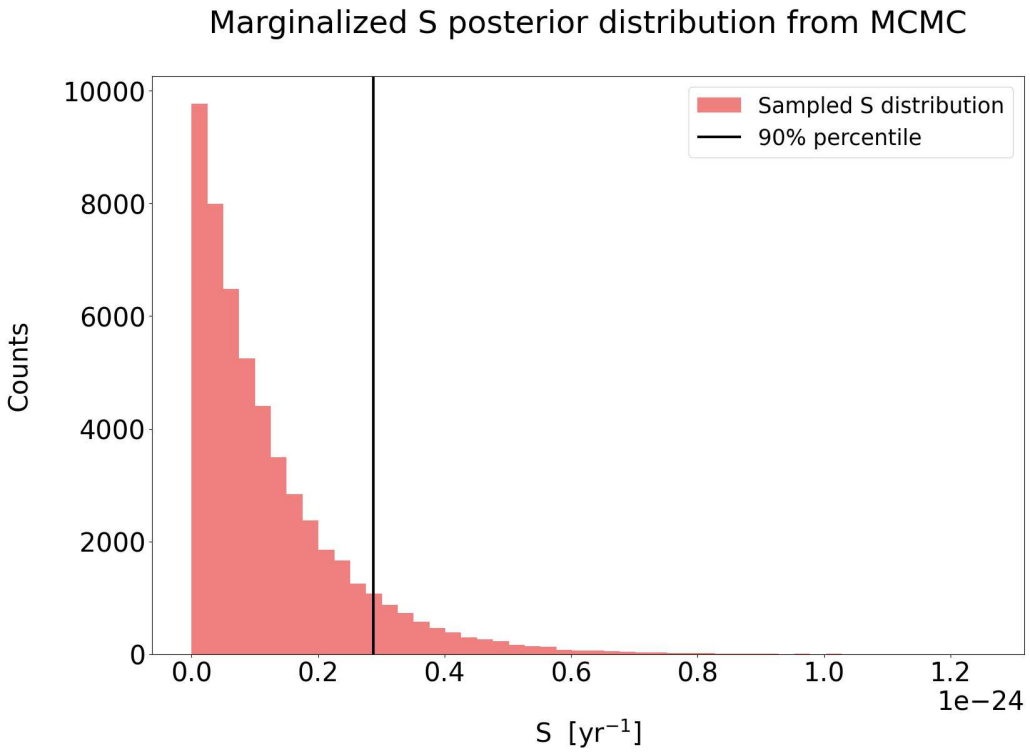


FIGURE 6.3: Marginalized posterior distribution for the strength of the signal $\mathcal{S}$, $\mathcal{P}(\mathcal{S}|\mathcal{D},\theta)$. Since the mode of the distribution is compatible with the absence of signal ($\mathcal{S} = 0$ yr^{-1}), only an upper limit on the strength of the signal can be determined: $\mathcal{S} < 2.81 \times 10^{-25}$ yr^{-1} at 90% C.I.

Fig. 6.3 shows $\mathcal{P}(\mathcal{S}|\mathcal{D},\theta)$, the one-dimensional marginalized posterior distribution of the strength of the signal $\mathcal{S}$. The mode of the distribution is fully compatible with the case $\mathcal{S} = 0$ yr^{-1}, namely with the absence of $0\nu\beta\beta$ signal.

In this situation, we cannot provide an estimate for the strength of the signal, but only an upper limit to its value:

$$\mathcal{S} < 2.81 \times 10^{-25} \text{ yr}^{-1} \qquad (90\% \text{ C.I.}) \tag{6.10}$$

This upper limit on the strength of the signal translates, by definition ($\mathcal{S} = 1/T^{0\nu}_{1/2}$), to a lower limit on the half life of the searched $0\nu\beta\beta$ decay:

$$T^{0\nu}_{1/2} > 3.56 \times 10^{24} \text{ yr} \qquad (90\% \text{ C.I.}) \tag{6.11}$$

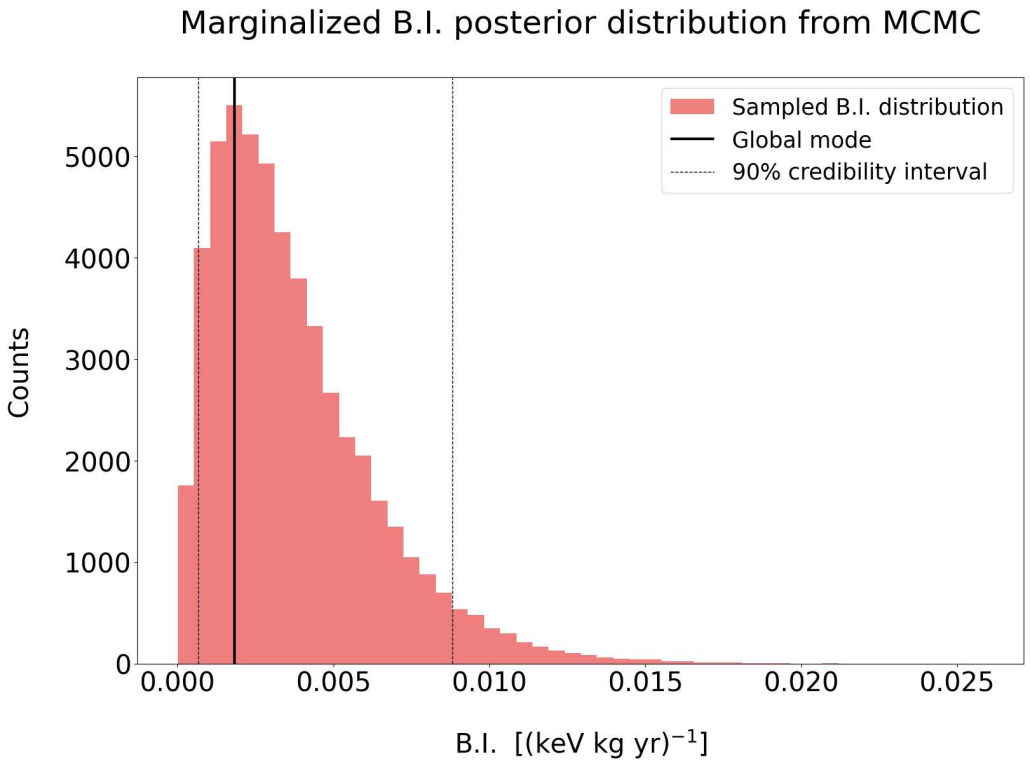


FIGURE 6.4: Marginalized posterior distribution for the background index $B.I.$, $\mathcal{P}(B.I.|\mathcal{D},\theta)$. In this case an estimate of the $B.I.$ is given as global mode of the distribution, since this is different from zero. The dashed lines show the two-tailed 90% C.I. .

Fig. 6.4 shows $\mathcal{P}(B.I.|\mathcal{D},\theta)$, the one-dimensional marginalized posterior distribution of the background index $B.I.$. For this distribution, the global mode is shifted with respect to zero: for this reason we can give it as estimate of the most probable value of the parameter of interest.

Since the distribution is not symmetric around the mode, the two-tailed Credibility Interval providing the 90% probability of containing the most probable $B.I.$ value corresponds to an asymmetric energy range. The interval is built considering as limits the 5% and 95% quantiles.

The resulting value of the background index is:

$$B.I. = 2.03^{+6.76}_{-1.38} \times 10^{-3} \ \text{(keV kg yr)}^{-1} \qquad \text{(90\% C.I.)} \tag{6.12}$$

The two-tailed [5-95]% Credibility Interval is therefore $B.I. \in [0.65\times10^{-3}; 8.79\times10^{-3}]$ (keV kg yr)$^{-1}$. Another relevant interval that can be provided is the 68% Credibility Interval, again two-tailed and symmetric in probability, [16-84]%, while asymmetric in energy, which results to be $[1.30\times10^{-3}; 6.06\times10^{-3}]$ (keV kg yr)$^{-1}$.

The obtained result is consistent with the expectations and with the rough estimate of $B.I. = \frac{1}{240\ \text{keV}\cdot 2.248\ \text{kg yr}} = \mathcal{O}(10^{-3})$.

We do not perform, in this work, a dedicated analysis of $\mathcal{P}(\epsilon|\mathcal{D},\theta)$, the one-dimensional marginalized posterior distribution of the efficiency.

This because the likelihood (Eq. (6.4)) does not contain enough information on the efficiency, so it does not manage to modify enough the shape of the considered prior. To compensate for this and to avoid spoiling the precision of the other estimates, we employ an informative Gaussian prior $\mathcal{P}(\epsilon)$ (Eq. (6.8)) based on the information available from GERDA. As it can be seen in the corner plot (Fig. 6.2, bottom right panel) the shape of the obtained posterior is too bound to the shape of the prior to extract meaningful information on the efficiency itself. A proper simulation of the full experimental setup performances must be implemented to obtain this information.

What we can do to give an idea of the possible results for $\mathcal{S}$ and $B.I.$ for different values of the efficiency (Fig. 6.5), is to repeat the MCMC simulation assuming different (fixed) ϵ values.

We consider a set of possible efficiency values in the range [0.4, 0.9], and we scan it by steps of 0.05. Note that at each step the value of the efficiency is fixed to a precise numerical value, so the MCMC is run considering the posterior in Eq. (6.5), with only two unknown parameters, $\mathcal{S}$ and $B.I.$: the sampling is thus performed in a bi-dimensional parameters space.

For each value of the efficiency, we determine the marginalized posterior distribution for $\mathcal{S}$ and $B.I.$ and we give an estimate for the two parameters as discussed above:

- Lower limit on the strength of the signal at 90% C.I. (in red in Fig. 6.5);
- Best estimate of the background index as global mode of its marginalized posterior distribution, with its [5, 95]% C.I (in blue in Fig. 6.5).

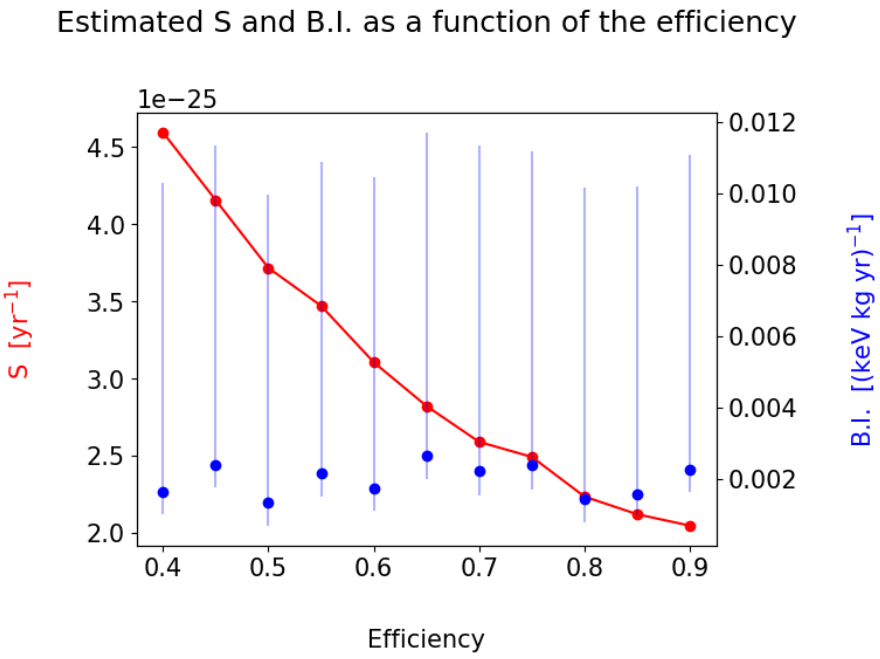


FIGURE 6.5: Lower limit to the strength of the signal $\mathcal{S}$ at 90% C.I. (red) and best estimate of the background index $B.I.$ as global mode of the distribution, with the [5, 95]% C.I. (blue). Results presented as a function of the efficiency, scanned by 0.05 step in the range [0.4, 0.9]. For each ϵ value the MCMC is run sampling $\{\mathcal{S}_i, B.I._I\}$ from a bi-dimensional parameters space.

The results are consistent with the expectations. The $B.I.$ appears independent from the signal detection efficiency, as it should be by definition (Eq. (6.2)). The limit we can put on the strength of the signal, instead, given the fixed dataset $\mathcal{D}$, containing only one event at energy $E = 2059.66$ keV $\neq Q_{\beta\beta}$, decreases with the increase of the signal detection efficiency, as expected from Eq. (6.1).

Note that by definition the $T^{0\nu}_{1/2}$ is the inverse of $\mathcal{S}$: this means that with the increase of the detection efficiency, the lower limit we can put to the half life grows.

6.4 Comments to results and future perspectives

Overall, the obtained results are consistent with the expectations: this is promising, since it shows that the LEGEND experiment provides clean and reliable data and that the analysis methodology is solid.

The achieved background index of 2.03×10^{-3} (keV kg yr)$^{-1}$ shows that, even during the commissioning phase, LEGEND's background suppression strategy is successful, both at the experiment level (radiopure materials, passive shieldings and active vetoes) and at the analysis level (analysis cuts)[1].

The result obtained for the $0\nu\beta\beta$ half life lower bound, $T^{0\nu}_{1/2} > 3.56 \times 10^{24}$ yr at 90% C.I., as expected suffers from the low exposure available for this analysis, of about 2.248 kg yr[2]. In fact, in the pursued quasi background-free regime, the half life sensitivity scales linearly with the exposure: this is of course the reason why the full experimental campaign will involve a higher Germanium mass (200 kg) and will last for at least five years, so as to grant an exposure of about 1000 kg yr.

Fig. 6.6 shows the perspective of the $0\nu\beta\beta$ half life sensitivity achievable with the next phases of the LEGEND experiment. It is clear from the plot that only in a background-free regime the sensitivity on $T^{0\nu}_{1/2}$ is maximized.

Achieving this background-free regime is therefore utmost important, and with this analysis we have shown that LEGEND successfully reaches it. The developed analysis methodology will be capable of producing extremely informative results as soon as the collected exposure become higher.

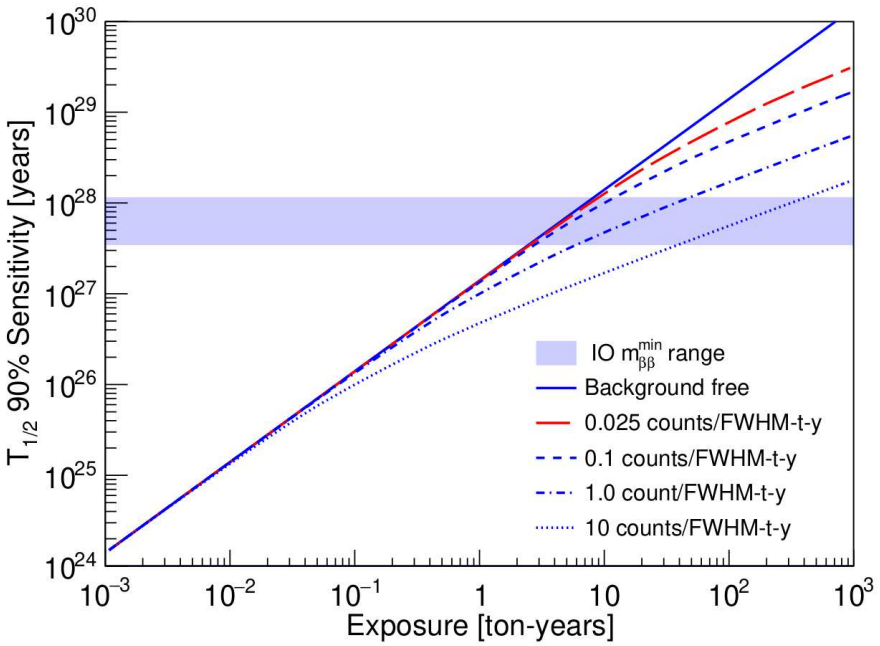


FIGURE 6.6: Projected $0\nu\beta\beta$ half life sensitivity achievable as a function of the exposure at different background index regimes. Only in a background-free regime the sensitivity scales linearly with the exposure: only in this regime, then, we make the most of the deployed Germanium mass. This is the reason why achieving a background free regime is so important for LEGEND. Picture from [49].

[1] *B.I.* achieved by GERDA: $B.I. = 5.2 \times 10^{-4}$ (keV kg yr)$^{-1}$ [32]. This is the lowest background index ever achieved by a $0\nu\beta\beta$ experiment.

[2] $T^{0\nu}_{1/2}$ lower bound achieved by GERDA: $T^{0\nu}_{1/2} > 1.8 \times 10^{26}$ at 90% C.L. [32]. This result was obtained with an exposure of 103.7 kg yr.

Conclusions

This thesis work is dedicated to the analysis of the data acquired by LEGEND-200 during its commissioning phase, and in particular to the study of the energy spectrum collected by Germanium detectors. The motivation to undertake this work is two-fold:

- Checking the performances of the experiment and the quality of the collected data, before starting the long-term data taking;
- Developing and validating a suitable analysis routine for the search or $0\nu\beta\beta$ signals, replicable on the full dataset at the end of the experimental campaign.

The core of the performed $0\nu\beta\beta$-oriented analysis is to remove noise and background signals from the energy spectrum, so as to be able to check whether or not the characteristic peak at $E = Q_{\beta\beta}$ appears. To reach this goal, multiple analysis cuts are implemented (Chapter 5): Pulser Signal Removal (Section 5.1), Quality Cuts (Section 5.2), Muon veto (Section 5.3), multiplicity cut (Section 5.4), LAr veto (Section 5.5) and Pulse Shape Discrimination (Section 5.6). The selected cuts exploit the unique features of $\beta\beta$ signals, coming first and foremost from the topology of the decay, in which the energy release is highly localized, within a volume of about 1 mm^3. As discussed in Section 5.7 the cuts successfully suppress most of the appearing background: starting with a raw sample of 29 738 052 signal in the full energy range, only 6807 pass all the analysis cuts, which means a surviving fraction around 0.029%. In the considered 240 keV ROI around $Q_{\beta\beta} = 2039\,\text{keV}$ the original number of signals is 38, dropping to just 1 after the cuts, for a surviving fraction of about 2.63%.

After the cuts, once retained only the good $\beta\beta$ candidates, a Bayesian analysis of the distribution of the residual signals in the ROI is performed, aimed at giving an estimate of the achieved background index (Chapter 6). The result is $B.I. = 2.03^{+6.67}_{-1.38} \times 10^{-3}$ (keV kg yr)$^{-1}$ considering a 90% two-tailed symmetric C.I.. Eventually, since at 90% C.I. there is no evidence for a $0\nu\beta\beta$ signal, it is possible to put a limit on the $0\nu\beta\beta$ half life based on the available data of $T^{0\nu}_{1/2} > 3.56 \times 10^{24}$ yr.

Considering the low exposure available for this analysis, of about 2.248 kg yr, the result for the $B.I.$ is extremely promising: it proofs that a successful background suppression can be achieved with the employed experimental setup and analysis routine. The effectiveness of the background suppression strategy is crucial for the future of the LEGEND experiment, considering that only in a background-free regime the half life sensitivity scales linearly with the exposure.
Projecting our half life result to the targeted 1000 kg yr exposure for LEGEND-200, we expect the design $T^{0\nu}_{1/2}$ sensitivity of $\mathcal{O}(10^{27}$ yr) to be successfully reached.

Appendix A

Residual spectra of interest

At each cut step of the analysis, part of the signals were considered good double beta candidates, and then kept in the primary spectrum, while others were discarded since identified as background.

In the main text we showed the primary spectrum after each cut, namely the spectrum made up of the signals that passed that cut, e.g. Fig. 5.18 (muon veto), 5.22 (multiplicity cut), 5.24 (LAr veto) and 5.35 (PSD).

Here, instead, we present the residual spectra, made up of the events discarded by the cuts. Note that here we show only the residual spectra made up of events identified as physical, but considered background for the search and study of $\beta\beta$ decays. The residual spectra made up of unphysical events, instead, are not shown, since the reconstructed energy is not considered meaningful: this is the case, for example, of the spectrum of the events that did not pass quality cuts.

A.1 Muon spectrum

When, in coincidence with a signal in a Germanium detector, a signal is acquired also by the PMTs reading out the water tank, the event is identified as a muon event (see Chapter 5.3).

The total number of events thus identified as *muons* by the muon veto system is $N_\mu^{\text{ev}} = 50239$. In some of these events, more than one Germanium detector acquires a non-zero signal: the total number of signals belonging to muon events is $N_\mu^{\text{sig}} = 52684$. These signals are of course removed from the main spectrum for the double beta search, since they constitute background in that context.

The total energy of an incoming muon would be equal to the sum of the energies it released in all its interactions. What we can reconstruct, though, is only the energy it released in the Germanium detectors: this is equal to the sum of the energies of all the signals acquired within the same event.

$$E_\mu^{\text{Ge}} = \sum_i E_\mu^{\text{Ge,i}} \tag{A.1}$$

Again, the calculated energy E_μ^{Ge} does not give an estimate of the total energy of the muon, but only of the energy it released within our detection system. The energy spectrum of these muon events is shown in Fig. A.1.

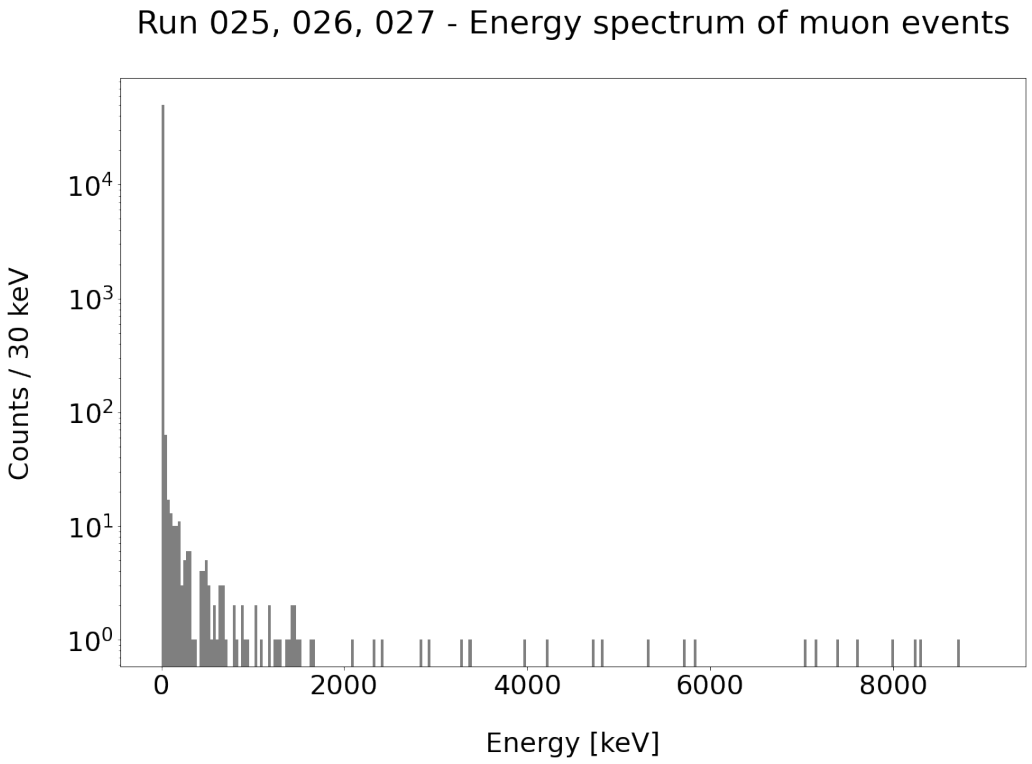


Figure A.1: Full energy spectrum of muon events.

We notice that muon events are often high energy events: they constitute therefore a background source in the ROI, and also at higher energies.

Besides that, the muon spectrum does not show any peculiar structure, at least with the acquired statistics. In fact, the low statistics of the muon spectrum is a confirmation of the efficacy of LEGEND's shielding system. As anticipated in Chapter 2, the location of the experiment is chosen specifically to achieve a strong attenuation of the cosmic rays flux: in particular, the 1400 m of rock overburden (acting as about 3500 m water equivalent shielding) is expected to reduce the muon flux at the experiment by six order of magnitudes with respect to above-ground rates, to a value of $\Phi_\mu \approx 1.2$ muons/(m^2·h) [50]. As discussed in Chapter 5.3, the muon flux reaching the detection system is $\Phi_\mu^{\text{calc}} \approx 1.22$ muons/(m^2·h), fully compatible with the expected value.

A.2 $\mathcal{M}0$ and $\mathcal{M}\geq 2$ spectrum

The multiplicity cut selects as principal spectrum the $\mathcal{M}1$ events: events in which only one signal has $\texttt{E}_{\texttt{rec}} \geq 40$ keV (see Chapter 5.4).

The residual spectrum, then, is made up of two contributions, $\mathcal{M}0$ events, in which no signals above 40 keV are acquired (Fig. A.2), and $\mathcal{M}\geq 2$, in which two or more signals above 40 keV are acquired (Fig. A.3).

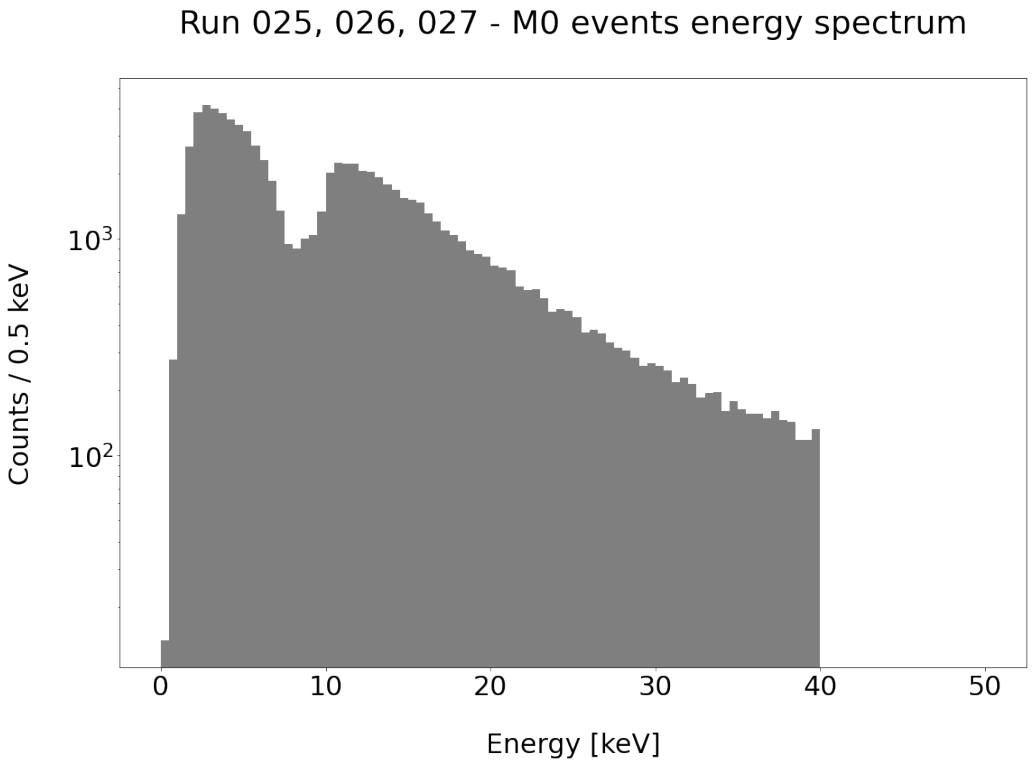


Figure A.2: $\mathcal{M}0$ energy spectrum.

The $\mathcal{M}0$ spectrum is of course fully contained within the [0, 40] keV range. In this spectrum, the energies of the single signals from each event appear separately: in fact, with the performed analysis there is no way to know whether or not multiple below-threshold signals acquired simultaneously (appearing formally in the same event) are correlated (actually come from the same physical event), and consequently whether or not it is meaningful to sum their energies.

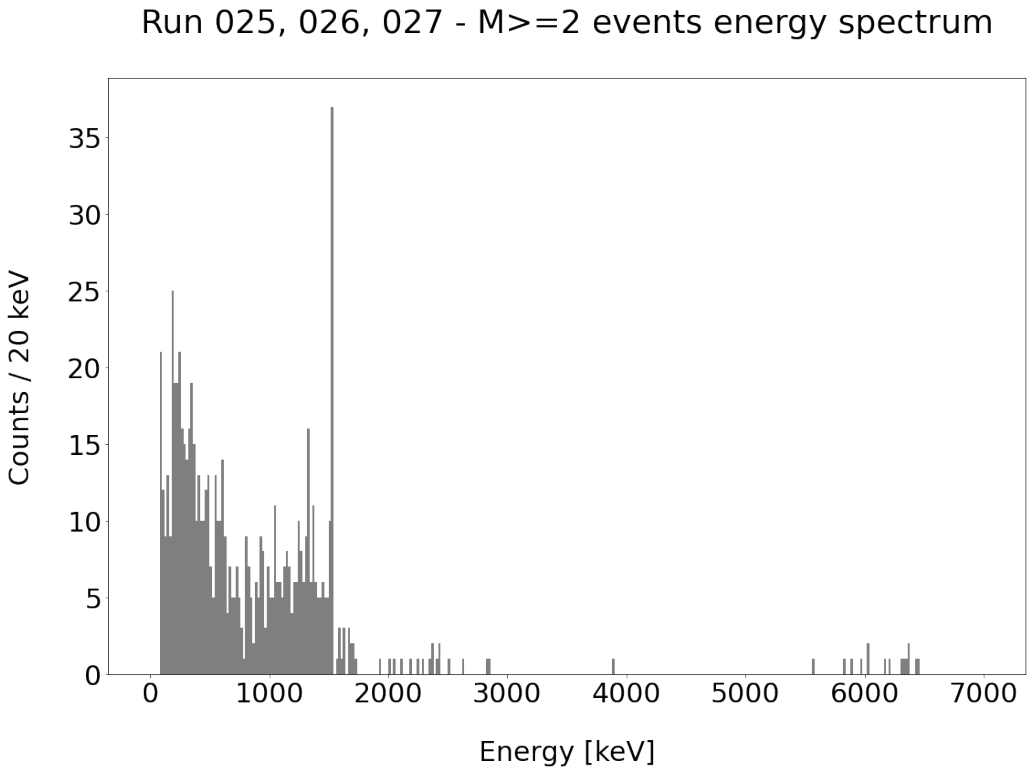


Figure A.3: Full $\mathcal{M}\geq 2$ energy spectrum

The $\mathcal{M}\geq 2$ spectrum, instead, is built summing the energies of all the above-threshold signals within

each event. Then, in this spectrum there is not one entry *per signal*, but one entry *per event*, having energy equal to the sum of all the above-threshold energies acquired in that event. This procedure is based on the assumption that most of the times multiple above-threshold signals within the same event are originated by a single physical process: for example, a radioactive decay producing γ-rays that cross several detectors before stopping, and release in each of them part of their energy. The total energy of the event could then be reconstructed as the sum of all the released energies.

Note that this conservative assumption is at the basis of the multiplicity cut itself: $\mathcal{M}\geq 2$ events are excluded from the main dataset precisely because we consider likely for them to come from the same physical event, and then we don't consider them good $\beta\beta$ candidates, which are known to be highly localized.

The validity of this assumption is confirmed by the presence, in $\mathcal{M}\geq 2$ spectrum, of the peaks at 1461 keV and 1525 keV, coming from the decay chain of ^{40}K and ^{42}K respectively (Fig. 5.27 and 5.28) [68].

We observe that the energy of the peaks extracted from their gaussian fit (see legend of Fig. A.4) are slightly lower than the expected values. This result is consistent with the physical interpretation we gave to the collected signals and with the procedure employed to evaluate the full energy of the event.

The photon travels across several detectors before stopping: in principle its total energy can be reconstructed as the sum of all the released energies. In practice, though, we consider in the calculation only the above-threshold energies. This means that if in the last crossed detector an energy < 40 keV is deposited, this energy is not taken into account in the calculation, since it is interpreted as *zero*. The analytical sum of the energies of the above-threshold signals will therefore be slightly lower than the real total energy of the event. The same interpretation can be given to the signal excess at ≈ 1520 keV: the energy deposited in the last crossed detector, despite being non-zero, was below-threshold and therefore not taken into account in the calculation.

This spectrum is therefore not optimal for estimating the energy of the gammas emitted by Potassium, but is useful in order to validate the analysis procedures and the interpretation of its results.

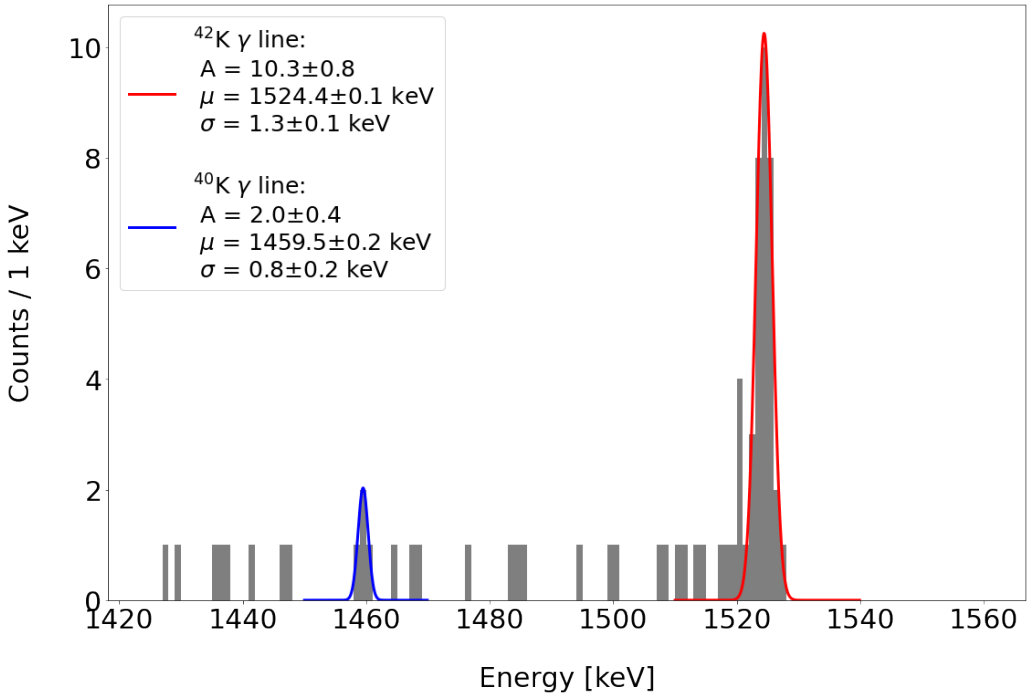


Figure A.4: Gaussian fit of Potassium peaks from $\mathcal{M}\geq 2$ energy spectrum.

A.3 LAr vetoed events spectrum

Fig. A.5 shows the energy spectrum of the events discarded by LAr veto (Section 5.5). It is particularly interesting since it is a pure physical background spectrum, showing the contribution of the radioactive materials present within the experimental site.

Starting from the low energy side, the ^{39}Ar distribution is clearly visible (meaning that it is well suppressed in the main spectrum) in the region below 500 keV. In the 500-1400 keV region the distribution is dominated Compton edge and continuum from the Potassium peaks. Note that the vetoed signals in this region *do not* come from $2\nu\beta\beta$, since these signals are highly localized (as much as the $0\nu\beta\beta$) and do not release energy in the LAr. Then the three main γ line are visible: ^{40}K (1461 keV), ^{42}K (1525 keV) and ^{208}Tl FEP (2614 keV). Some residual background signals are vetoed also in the $Q_{\beta\beta}$ region.

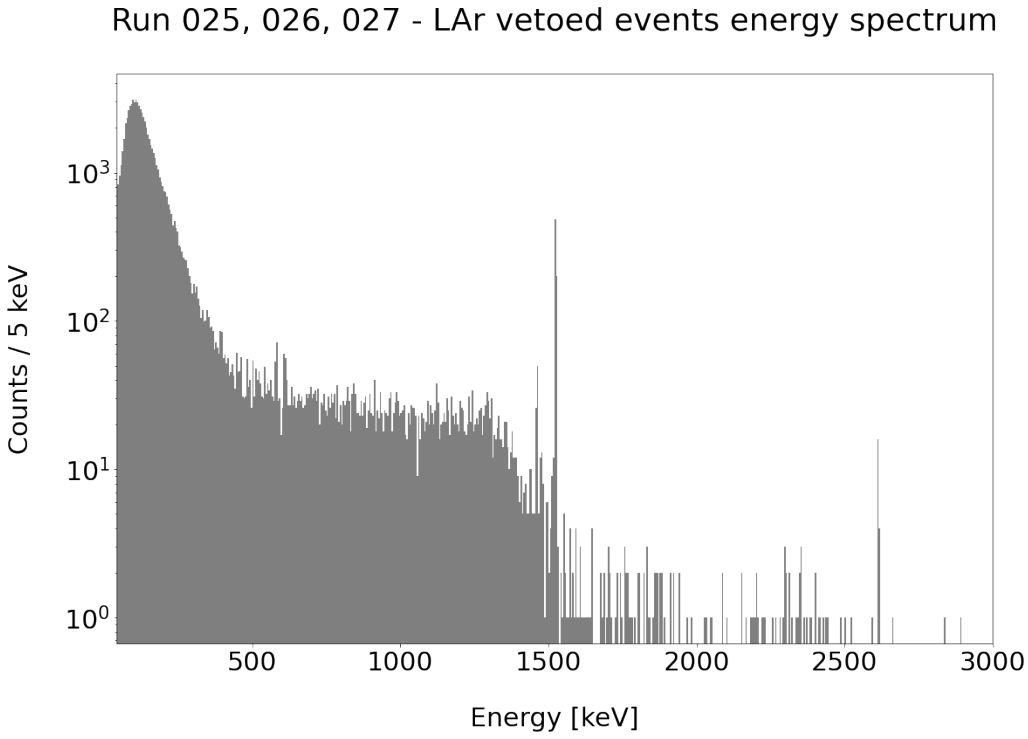


Figure A.5: Energy spectrum of the events vetoed by LAr veto during run 025, 026, 027 (zoom).

A.4 Spectrum of events removed by PSD

The PSD cut (Section 5.6) manages to remove a sensible fraction of the remaining signals from the ^{39}Ar distribution, from the ^{40}K and ^{42}K peaks and from the $Q_{\beta\beta}$ region. Few signals are removed also from the $2\nu\beta\beta$ region, but again the removed signals *do not* come from the $2\nu\beta\beta$ distribution, since they are SSE as the $0\nu\beta\beta$.

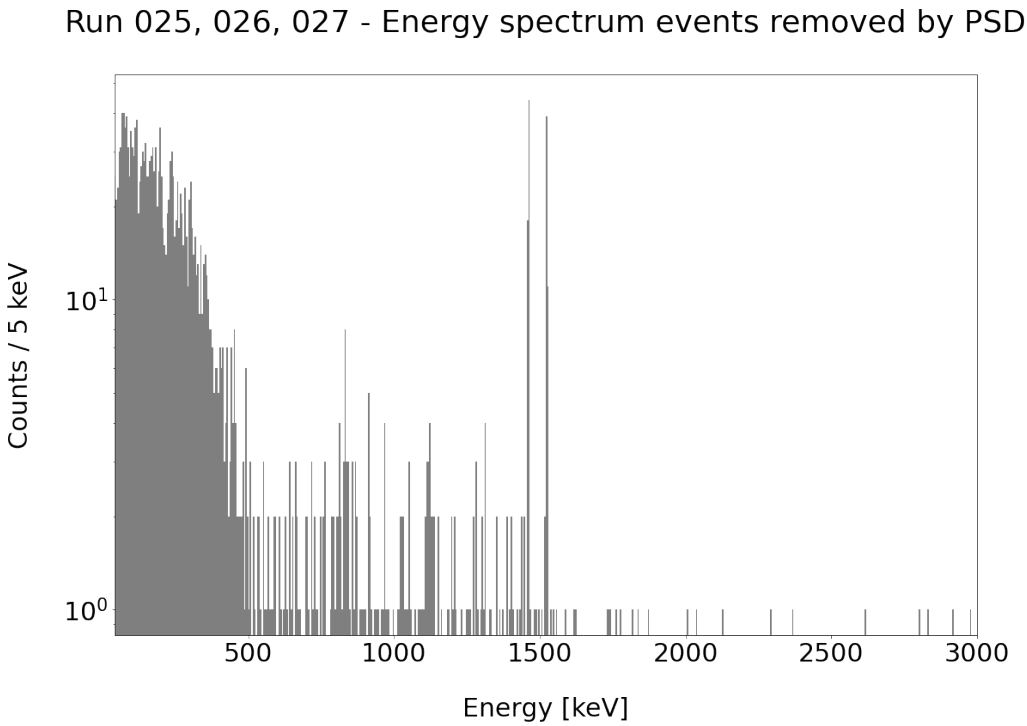


Figure A.6: Energy spectrum of the events vetoed by PSD during run 025, 026, 027 (zoom).

Appendix B

Glossary of acronyms and abbreviations

$0\nu\beta\beta$ = Neutrinoless double beta decay

$2\nu\beta\beta$ = Two neutrinos double beta decay

$\beta\beta$ = Double beta decay (with or without neutrinos)

BEGe = Broad Energy Germanium detector

$B.I.$ = Background Index

BSM = Beyond Standard Model

chXXX = FlashCam Channel XXX

C.I. = Credibility interval

Coax = semi-Coaxial Germanium detectors

DAQ = Data Acquisition

DEP = Double Escape Peak

DSP = Digital Signal Processing

EFCu = Electro-Foamed Copper

FC = FlashCam

FEP = Full Energy Peak

FWHM = Full Width at Half Maximum

HPGe = High Purity Germanium detectors

HV = High Voltage

HW = Hardware

IC[PC] = Inverted-Coaxial [Point-Contact] Germanium detectors

IO = Inverted Ordering (of neutrino mass hierarchy)

LAr = Liquid Argon

LEGEND = Large Enriched Germanium Experiment for Neutrinoless $\beta\beta$ decay

LNGS = Laboratori Nazionali del Gran Sasso

$\mathcal{M}$n = Multiplicity n

$m_{\beta\beta}$ = Effective Majorana neutrino mass in $0\nu\beta\beta$ decay

MCMC = Markov Chain Monte Carlo

MS = Mini-Shroud

MSE = Multi Site Events

NO = Normal Ordering (of neutrino mass hierarchy)

PDF = Probability Density Function

PEN = PolyEthylene Naphthalate

PMT = PhotoMultipliers

PPC = P-type Point Contact Germanium detectors

PSD = Pulse Shape Discrimination

PSR = Pulser Signals Removal

$Q_{\beta\beta}$ = Q-value of the $\beta\beta$ decay

QED = Quantum ElectroDynamics

QC = Quality Cuts

ROI = Region Of Interest

$\mathcal{S}$ = Strength of the (eventual) $0\nu\beta\beta$ signal

SEP = Single Escape Peak

SiPM = Silicon PhotoMultiplier

SIS = Source Insertion System

SM = Standard Model

SSE = Single Site Events

$T^{0\nu}_{1/2}$ = Half life for $0\nu\beta\beta$

$T^{2\nu}_{1/2}$ = Half life for $2\nu\beta\beta$

TPB = TetraPhenyl Butadiene

VUV = Vacuum Ultra-violet

WLS = WaveLength Shifter

Bibliography

[1] W. Pauli, *Pauli letter collection: letter to Lise Meitner*, Typed copy, URL: http://cds.cern.ch/record/83282.

[2] L. M. Brown, *The idea of the neutrino*, in: Physics Today **31**.9 (September 1978), pp. 23–28, DOI: 10.1063/1.2995181.

[3] J. Chadwick, *Possible existence of a neutron*, in: Nature **129**.3252 (1932), p. 312, URL: https://doi.org/10.1038/129312a0.

[4] J. Chadwick, *The existence of a neutron*, in: Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character **136**.830 (1932), pp. 692–708, URL: https://doi.org/10.1098/rspa.1932.0112.

[5] C. D. Ellis and W. Henderson, *Artificial radioactivity*, in: Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character **146**.856 (1934), pp. 206–216.

[6] F. Reines and C. L. Cowan, *The neutrino*, in: Nature **178**.4531 (1956), pp. 446–449, URL: https://doi.org/10.1038/178446a0.

[7] E. Fermi, *Versuch einer Theorie der β-Strahlen. I.* In: Z. Phys. **88** (1934), pp. 161–177, DOI: 10.1007/BF01351864.

[8] F. L. Wilson, *Fermi's theory of beta decay*, in: American Journal of Physics **36**.12 (1968), pp. 1150–1160, URL: https://doi.org/10.1119/1.1974382.

[9] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson, *Experimental Test of Parity Conservation in Beta Decay*, in: Phys. Rev. **105** (1957), pp. 1413–1415, DOI: 10.1103/PhysRev.105.1413.

[10] M. Goldhaber, L. Grodzins, and A. W. Sunyar, *Helicity of Neutrinos*, in: Phys. Rev. **109** (1958), pp. 1015–1017, DOI: 10.1103/PhysRev.109.1015.

[11] S. L. Glashow, *Partial-symmetries of weak interactions*, in: Nuclear Physics **22**.4 (1961), pp. 579–588, ISSN: 0029-5582, DOI: https://doi.org/10.1016/0029-5582(61)90469-2.

[12] Super-Kamiokande Collaboration, *Evidence for Oscillation of Atmospheric Neutrinos*, in: Physical Review Letters **81**.8 (1998), pp. 1562–1567, DOI: 10.1103/physrevlett.81.1562.

[13] SNO Collaboration, *Measurement of the Rate of $\nu_e + d \rightarrow p + p + e^-$ Interactions Produced by ^8B Solar Neutrinos at the Sudbury Neutrino Observatory*, in: Phys. Rev. Lett. **87** (2001), p. 071301, DOI: 10.1103/PhysRevLett.87.071301.

[14] SNO Collaboration, *Direct Evidence for Neutrino Flavor Transformation from Neutral-Current Interactions in the Sudbury Neutrino Observatory*, in: Phys. Rev. Lett. **89** (2002), p. 011301, DOI: 10.1103/PhysRevLett.89.011301.

[15] M. Nakahata, *History of Solar Neutrino Observations*, 2022, DOI: 10.1093/ptep/ptac039.

[16] T2K Collaboration, *Indication of Electron Neutrino Appearance from an Accelerator-Produced Off-Axis Muon Neutrino Beam*, in: Phys. Rev. Lett. **107** (July 2011), 041801, DOI: 10.1103/PhysRevLett.107.041801.

[17] OPERA Collaboration, *Final results on neutrino oscillation parameters from the OPERA experiment in the CNGS beam*, in: Phys. Rev. D **100**.5 (2019), DOI: 10.1103/physrevd.100.051301.

[18] Z. Maki, M. Nakagawa, and S. Sakata, *Remarks on the Unified Model of Elementary Particles*, in: Progress of Theoretical Physics **28**.5 (1962), pp. 870–880, DOI: 10.1143/PTP.28.870.

[19] M. Nakahata, *History of solar neutrino observations*, in: Progress of Theoretical and Experimental Physics **2022**.12 (2022), ISSN: 2050-3911, DOI: 10.1093/ptep/ptac039.

[20] Fermilab, *All Things Neutrino*, URL: https://neutrinos.fnal.gov/.
[21] A. Cabrera et al., *Synergies and prospects for early resolution of the neutrino mass ordering*, in: Scientific Reports **12**.1 (2022), 5393, URL: https://www.nature.com/articles/s41598-022-09111-1.
[22] F. Mandl and G. Shaw, *Quantum field theory*, John Wiley & Sons, 2010, pp. 381–385.
[23] M. Agostini, G. Benato, J. A. Detwiler, J. Menéndez, and F. Vissani, *Toward the discovery of matter creation with neutrinoless $\beta\beta$ decay*, in: Rev. Mod. Phys. **95**.2 (2023), 025002, DOI: 10.1103/RevModPhys.95.025002.
[24] S. Bilenky and C. Giunti, *Neutrinoless double-beta decay: A brief review*, in: Modern Physics Letters A **27**.13 (2012), p. 1230015, URL: https://arxiv.org/pdf/1203.5250.pdf.
[25] M. Goeppert-Mayer, *Double Beta-Disintegration*, in: Phys. Rev. **48** (1935), pp. 512–516, DOI: 10.1103/PhysRev.48.512.
[26] R. Saakyan, *Two-Neutrino Double-Beta Decay*, in: Annual Review of Nuclear and Particle Science **63**.1 (2013), pp. 503–529, DOI: 10.1146/annurev-nucl-102711-094904.
[27] C. v. Weizsäcker, *Zur theorie der kernmassen*, in: Zeitschrift für Physik **96**.7-8 (1935), pp. 431–458, URL: https://link.springer.com/article/10.1007/BF01337700.
[28] W. H. Furry, *On Transition Probabilities in Double Beta-Disintegration*, in: Phys. Rev. **56** (1939), pp. 1184–1193, DOI: 10.1103/PhysRev.56.1184.
[29] A. Barabash, *Precise Half-Life Values for Two-Neutrino Double-β Decay: 2020 review*, 2020, eprint: 2009.14451, URL: https://arxiv.org/abs/2009.14451.
[30] GERDA Collaboration, *Results on $\beta\beta$ decay with emission of two neutrinos or Majorons in ^{76}Ge from GERDA Phase I*, in: The European Physical Journal C **75**.9 (September 2015), DOI: 10.1140/epjc/s10052-015-3627-y.
[31] KamLAND-Zen Collaboration, *Search for the Majorana Nature of Neutrinos in the Inverted Mass Ordering Region with KamLAND-Zen*, in: Physical Review Letters **130**.5 (January 2023), DOI: 10.1103/physrevlett.130.051801.
[32] GERDA Collaboration, *Final Results of GERDA on the Search for Neutrinoless Double-β Decay*, in: Physical Review Letters **125**.25 (2020), DOI: 10.1103/physrevlett.125.252502.
[33] C. Wiesinger, *No neutrinos not found*, PhD Thesis, Munich, Tech. U., PhD thesis, 2020, URL: https://www.mpi-hd.mpg.de/gerda/public/2020/phd2020_wiesinger_christoph.pdf.
[34] V. I. Tretyak and Y. G. Zdesenko, *Tables of Double Beta Decay Data — An Update*, in: Atomic Data and Nuclear Data Tables **80**.1 (2002), pp. 83–116, ISSN: 0092-640X, DOI: https://doi.org/10.1006/adnd.2001.0873.
[35] F. Šimkovic, *Double beta decay: A problem of particle, nuclear and atomic physics*, in: Progress in Particle and Nuclear Physics **64**.2 (2010), Neutrinos in Cosmology, in Astro, Particle and Nuclear Physics, pp. 219–227, ISSN: 0146-6410, DOI: https://doi.org/10.1016/j.ppnp.2009.12.015.
[36] C. Nones, *Neutrinoless Double Beta Decay: present and future*, Presentation at "32nd Rencontres de Blois on Particle Physics and Cosmology", October 2021, URL: https://indico.cern.ch/event/997281/contributions/4574209/attachments/2332505/.
[37] L. Pertoldi, *Search for new physics with two-neutrino double-beta decay in Gerda data*, PhD Thesis, 2020.
[38] D. Budjá š et al., in: Journal of Instrumentation **8**.04 (April 2013), pp. P04018–P04018, DOI: 10.1088/1748-0221/8/04/p04018.
[39] MAJORANA Collaboration, *The* MAJORANA DEMONSTRATOR *Neutrinoless Double-Beta Decay Experiment*, 2013, eprint: 1308.1633, URL: https://arxiv.org/abs/1303.6768.
[40] B. J. Mount, M. Redshaw, and E. G. Myers, *Double-β-decay Q values of 74Se and 76Ge*, in: Phys. Rev. C **81** (2010), 032501, DOI: 10.1103/PhysRevC.81.032501.
[41] CUORE Collaboration, *CUORE: a cryogenic underground observatory for rare events*, in: Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **518**.3 (February 2004), pp. 775–798, DOI: 10.1016/j.nima.2003.07.067.

[42] The CUPID Interest Group, *CUPID: CUORE (Cryogenic Underground Observatory for Rare Events) Upgrade with Particle IDentification*, 2015, eprint: 1504.03599, URL: https://arxiv.org/abs/1504.03599.

[43] H. Bhang et al., *AMoRE experiment: a search for neutrinoless double beta decay of 100Mo isotope with 40Ca100MoO4 cryogenic scintillation detector*, in: Journal of Physics: Conference Series **375**.4 (July 2012), p. 042023, DOI: 10.1088/1742-6596/375/1/042023.

[44] A. Gando et al., *Measurement of the double-β decay half-life of 136 Xe with the KamLAND-Zen experiment*, in: Physical Review C **85**.4 (2012), p. 045504, URL: https://journals.aps.org/prc/abstract/10.1103/PhysRevC.85.045504.

[45] SNO+ Collaboration, *Current Status and Future Prospects of the SNO+ Experiment*, in: Advances in High Energy Physics **2016** (2016), pp. 1–21, DOI: 10.1155/2016/6194250.

[46] M. Auger et al., *The EXO-200 detector, part I: detector design and construction*, in: Journal of Instrumentation **7**.05 (May 2012), pp. P05010–P05010, DOI: 10.1088/1748-0221/7/05/p05010.

[47] J. J. Gómez Cadenas et al., *Present status and future perspectives of the NEXT experiment*, in: Advances in High Energy Physics (2014), URL: https://arxiv.org/abs/1307.3914.

[48] J. Detwiler, *Future Neutrinoless Double Beta Decay Experiments*, Presentation at "Neutrino 2020", July 2020, DOI: 10.5281/zenodo.3959552.

[49] LEGEND Collaboration, *LEGEND-1000 Preconceptual Design Report*, 2021, DOI: 10.48550/ARXIV.2107.11462.

[50] GERDA Collaboration, *Background-free search for neutrinoless double-β decay of 76Ge with GERDA*, in: Nature **544**.7648 (2017), 47–52, DOI: 10.1038/nature21717.

[51] LEGEND Collaboration, *LEGEND Confluence*, Internal Documentation, URL: https://elog.legend-exp.org/.

[52] LEGEND Collaboration, *LEGEND Elog*, Internal Documentation, URL: https://elog.legend-exp.org/.

[53] V. Biancacci, *Detailed active volume determination of HPGe detectors and its impact on the measurement of the half-life of $2\nu\beta\beta$ decay of 76Ge in the Gerda and LEGEND-200 experiments*, PhD Thesis, University of Padova, 2022, URL: https://www.research.unipd.it/handle/11577/3471174.

[54] M. Schwarz et al., *Liquid Argon Instrumentation and Monitoring in LEGEND-200*, in: EPJ Web Conf. **253** (2021), 11014, DOI: 10.1051/epjconf/202125311014.

[55] K. Knöpfle and B. Schwingenheuer, *Design and performance of the GERDA low-background cryostat for operation in water*, in: Journal of Instrumentation **17**.02 (February 2022), P02038, DOI: 10.1088/1748-0221/17/02/P02038.

[56] GERDA collaboration, *Liquid argon light collection and veto modeling in GERDA Phase II*, 2022, eprint: 2212.02856, URL: https://arxiv.org/abs/2212.02856.

[57] K. Freund et al., *The performance of the Muon Veto of the Gerda experiment*, in: The European Physical Journal C **76**.5 (2016), DOI: 10.1140/epjc/s10052-016-4140-7.

[58] F. Henkes, *LEGEND dashboard*, Internal Documentation, URL: https://legend.edm.nat.tum.de/.

[59] S. Calgaro and M. Morella, *Online Data Monitoring*, LEGEND Collaboration Meeting, 10-14 October 2022, GSSI, L'Aquila (Italy) - Internal Documentation, October 11, 2022.

[60] M. Agostini, *L60 r25/26/27 Analysis*, LEGEND Collaboration Meeting, 10-14 October 2022, GSSI, L'Aquila (Italy) - Internal Documentation, October 11, 2022.

[61] LEGEND Collaboration, *Pygama github repository*, URL: https://github.com/legend-exp/pygama.

[62] M. Salathe and T. Kihm, *Optimized digital filtering techniques for radiation detection with HPGe detectors*, in: Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **808** (2016), pp. 150–155, URL: https://arxiv.org/abs/1504.02039.

[63] GERDA Collaboration, *Improvement of the Energy Resolution via an Optimized Digital Signal Processing in GERDA Phase I*, 2015, eprint: 1502.04392, URL: https://arxiv.org/abs/1502.04392.

[64] GERDA Collaboration, *Calibration of the Gerda experiment*, in: The European Physical Journal C **81**.8 (2021), DOI: 10.1140/epjc/s10052-021-09403-2.
[65] PhysicsOpenLab, *Thorium Gamma Spectrometry*, URL: https://physicsopenlab.org/2016/01/31/thorium-gamma-spectrometry/.
[66] *National Nuclear Data Center - Nuclide Chart*, URL: https://www.nndc.bnl.gov/ (Last accessed: February 2023).
[67] GERDA Collaboration, *Improvement of the energy resolution via an optimized digital signal processing in GERDA Phase I*, in: Eur. Phys. J. C **75**.6 (2015), 255, DOI: 10.1140/epjc/s10052-015-3409-6.
[68] GERDA Collaboration, *Modeling of GERDA Phase II data*, in: Journal of High Energy Physics **2020**.3 (2020), DOI: 10.1007/jhep03(2020)139.
[69] T. Kihm, *Ge-Pulser*, LEGEND DAQ Call - Internal Documentation, August 31, 2022, (Last accessed: February 2023).
[70] S. Sailer, *FlashCam notes*, Internal Documentation.
[71] T. Kihm, Personal communication, January 2023.
[72] G. F. Knoll, *Radiation detection and measurement*, John Wiley & Sons, 2010.
[73] G. Marshall, *L60 Germanium Analysis*, LEGEND Collaboration Meeting, 10-14 October 2022, GSSI, L'Aquila (Italy) - Internal Documentation, October 11, 2022.
[74] G. Marshall, *L200 Data Structure, Content, Flow*, LEGEND Collaboration Meeting, 8-11 May 2023, Roma Tre University, Rome (Italy) - Internal Documentation, May 9, 2023.
[75] LEGEND Collaboration, *LEGEND github repository*, Internal Documentation, URL: https://github.com/legend-exp/.
[76] A. Lazzaro, *Signal processing and event classification for a background free neutrinoless double beta decay search with the GERDA experiment*, PhD Thesis, Munich, Tech. U., PhD thesis, 2019, URL: https://www.mpi-hd.mpg.de/gerda/public/2019/phd2019_AndreaLazzaro.pdf.
[77] G. Marshall, S. Sullivan, A. Alexander, and M. Agostini, *Characterization of Analysis Routines for the Digital Signal Processing of Ge-detector data*, Internal Documentation, August 24, 2021.
[78] K. Knöpfle and B. Schwingenheuer, *Design and performance of the GERDA low-background cryostat for operation in water*, in: Journal of Instrumentation **17**.02 (2022), P02038, DOI: 10.1088/1748-0221/17/02/p02038.
[79] P. Krause, *Liquid Argon Stream: a first view on the L60 LAr data*, LEGEND Collaboration Meeting, 10-14 October 2022, GSSI, L'Aquila (Italy) - Internal Documentation, October 11, 2022.
[80] IAEA Nuclear Data Services, *National Nuclear Data Center - Nuclide Chart*, URL: https://www-nds.iaea.org/relnsd/vcharthtml/VChartHTML.html.
[81] M. Agostini et al., *Pulse shape discrimination for Gerda Phase I data*, in: The European Physical Journal C **73**.10 (2013), DOI: 10.1140/epjc/s10052-013-2583-7.
[82] GERDA Collaboration, *Pulse shape analysis in Gerda Phase II*, in: The European Physical Journal C **82**.4 (2022), DOI: 10.1140/epjc/s10052-022-10163-w.
[83] GERDA Collaboration, *Limit on Neutrinoless Double Beta Decay of 76Ge by GERDA*, in: Physics Procedia **61** (2015), 13th International Conference on Topics in Astroparticle and Underground Physics, TAUP 2013, pp. 828–837, ISSN: 1875-3892, DOI: https://doi.org/10.1016/j.phpro.2015.06.002.
[84] D. J. MacKay, *Information theory, inference and learning algorithms*, Cambridge University Press, 2003.
[85] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, *emcee: The MCMC Hammer*, in: Publications of the Astronomical Society of the Pacific **125**.925 (2013), pp. 306–312, DOI: 10.1086/670067.
[86] *emcee*, Package Documentation, URL: https://emcee.readthedocs.io/en/stable/.