Search for sterile neutrino oscillations with the SoLid experiment at BR2 reactor : Energy calibration of the detector and antineutrino signal extraction

Noë Roy

université
PARIS-SACLAY

Search for sterile neutrino oscillations with the SoLid experiment at BR2 reactor: Energy calibration of the detector and antineutrino signal extraction.

Recherche d'oscillation vers un neutrino stérile auprès de l'expérience SoLid située au réacteur BR2 : calibration en énergie du détecteur et extraction du signal antineutrino.

Thèse de doctorat de l'Université Paris-Saclay

École doctorale n° 576, PHENIICS
(Particules, Hadrons, Énergie, Noyau, Instrumentation, Imagerie, Cosmos et Simulation)
Spécialité de doctorat: Physique des particules
Unité de recherche: Université Paris-Saclay, CNRS, IJCLab, 91405, Orsay, France.
Référent: : Faculté des sciences d'Orsay

Thèse présentée et soutenue à Orsay, le 16 décembre 2021, par

Noë ROY

Composition du jury :

Marie-Hélène SCHUNE Directrice de recherche, IJCLab	Présidente
Inés GIL BOTELLA Directrice de recherche, CIEMAT	Rapportrice & Examinatrice
David LHUILLIER Directeur de recherche, CEA Saclay	Rapporteur & Examinateur
Christine MARQUET Directrice de recherche, CENBG	Rapportrice & Examinatrice
Dirk RYCKBOSCH Professeur, Ghent University	Examinateur

Direction de la thèse :

Mathieu BONGRAND Chargé de recherche, SUBATECH - IMT Atlantique	Directeur de thèse
Laurent SIMARD Enseignant chercheur, Université Paris-Saclay - IJCLab	Co-encadrant de thèse

Invité :

Sara BOLOGNESI Chargée de recherche, CEA Saclay	Invitée

Thèse de doctorat

NNT: 2021UPASP142

Acknowledgment

Je tiens tout d'abord à remercier mes rapportrices et rapporteur, Inés Gil Botella, David Lhuillier et Christine Marquet, qui ont accepté, dans un temps court, et malgré les tribulations administratives, de prendre le temps de lire en détail mon manuscrit de thèse. Merci pour vos retours, commentaires et questions. Merci à Marie-Hélène Schune pour avoir accepté de présider mon jury. Merci également aux autres membres du jury, Dirk Ryckbosch et Sara Bolognesi pour le temps que vous avez consacré à ma thèse.

Bien entendu, cette thèse n'aurait pas été possible sans mes deux encadrants, Laurent et Mathieu. Merci à toi Laurent pour ta vision du petit détail (pour certain, mais pour moi ça veut dire beaucoup...) qui coince et que personne n'avait vu et bien entendu les gâteaux, tartes et autres cookies du lundi. Mathieu, merci pour tout. Ce fût trois années plutôt mouvementées, mais tu ne m'as jamais laissé tombé, bien au contraire. Dans les moments de doutes où les résultats ne suivaient pas ou quand la rédaction n'avançait pas, tu as pu trouver les mots pour me remettre d'aplomb. Et la chose la plus importante, c'est qu'avec toi, je savais que je serais toujours soutenu quoi qu'il arrive et c'était probablement la chose la plus importante pour moi en tant qu'apprenti chercheur.

Merci aussi à toute l'équipe ~~gâteaux~~ neutrinos (et plus si affinité) du 200. Merci Luis, sans toi et tes fameuses "traditions", nous n'aurions pas eu de gâteau du lundi. Delphine, pour m'avoir passé le flambeau et m'avoir conseillé de foncer sur une thèse avec Mathieu. Catherine, pour avoir été notre deuxième maman pendant ton séjour au LAL. Bien sûr, les amis de Delight, Pia, Malak (I know that you can speak french now, so I won't write it in english!) ces pauses café dans le bureau des thésards me manqueront ! Monsieur Garrido, tu as été présent sur toute la ligne. De mon entretien pour l'entrée au magistère, en passant par mes premiers pas dans la physique du neutrino en stage de M1 pour en arriver à ma présentation à Mathieu… Et bien entendu aux blagues, pas toujours racontables, dans le bureau entre deux drifts à vélo. Et puis, Cloé, ma partenaire de crime. Toutes ces musiques mises dans nos têtes respectives, toutes ces vidéos les plus débiles les unes des autres, ces plots ROOT qui n'auraient probablement pas été aussi beau sans toi... Parfois un peu de discussion de physique, mais surtout beaucoup de thés, de gâteaux (oui, encore les gâteaux), d'éclat de rire à 19 h le vendredi soir avant d'aller au rock... Bref, grâce à toi cette thèse était plus fun que j'aurais pu l'espérer et merci énormément pour ça.

Plus généralement, merci à toutes les personnes du labo avec qui j'ai eu la chance de travailler, discuter, boire un café. Dominique pour ton aide précieuse sur l'électronique LiquidO. Les anciens et moins anciens thésards des soirées films et occasionnellement jeux. Les amis du bureau des étudiants. Dédicace spéciale à Eli notre marraine, Louis future star montante du showbiz et bien sûr Guillem, partenaire de TP, expert en rap français et pipelette à ses heures perdues.

Merci à tous les collègues de SoLid avec qui j'ai adoré travailler pendant ces trois ans. Tout spécialement Benoit, merci énormément pour ton temps, ton aide pour mes travaux de thèse et pour la relecture de ma thèse. Simon with whom I really enjoyed talking about everything but SoLid on Mattermost. Mike, I hope we will figure out one day why the smiley are that different between France and Poland). I have to admit that I am still amazed by your capacity to drink that much beer, in a short amount of time and to still be able to do a great job at the meeting the day after. Savitri, cette quantité d'images qui collent à toutes les situations imaginables me hantera probablement toute ma vie, mais clairement, chapeau l'artiste ! Rijeesh, you took over the analysis work, and it was really a pleasure to work with you, during the few months we worked together. David, tu as été à l'origine de beaucoup de travaux présentés dans cette thèse et j'ai beaucoup appris en bossant à tes côtés sur cette satanée calibration. Je ne sais pas encore où je serai à l'avenir, mais j'espère vraiment qu'on aura l'occasion de retravailler ensemble. Et merci à Alexandra Elbakyan, pas tout à fait une collègue, mais qui a œuvré grandement pour que la recherche soit plus accessible et par tous, et sans qui mes trois années de travail au sein de SoLid auraient été bien plus compliqué.

Merci aux potos d'Orsay, rencontrés en cours, en danse, à l'ASESCO... Quentin pour nos TP de M1, on n'a pas toujours tout réussi il faut l'admettre, mais on a toujours été créatifs dans l'explication nos échecs. Benjamin, partenaire de judo, de volley, et de discussions enflammées sur le monde. Mathilde et Blandine, mes deux voisines, camarades de rock, puis d'escalade, grâce à qui ces années à la Pacat' ont été bien plus agréables. Valentin, Eugénie, Théophile, on en a vécu des galères avec l'asso, mais ça a valu tout le temps qu'on a passé ensemble, autours d'une pizza, d'un chocolat chaud ou d'un jeu de société, parfois les trois en même temps. Lydia, dire que ces 6 années à Orsay aurait été fade sans ta présence serait un euphémisme. Ton dogmatisme à toute épreuve, nos débats interminables sur des sujets tous aussi inutiles les uns que les autres, ton ton souvent râleur... Et puis les balades, les films, les restaurants, la danse, notre longue discussion à propos de tout et rien qui dure depuis quasiment 5 ans. Une grande personne a dit que la vie, c'était un cosinus, et bien merci d'avoir était là pour toutes les valeurs de la sinusoïde.

Merci aux amis rencontrés au cours de mes pérégrinations. Les amis de prépa. Avec une mention spéciale pour Cécichou, tu m'as introduit aux incertitudes en TP de prépa et comme tu le vois, ça m'a bien été utile ! Les amis du Japon, Emma, Quentin, Elouan, promis je vous rejoins le plus vite possible. Les camarades de jeu, à des horaires parfois peu raisonnables grâce à qui les derniers mois n'ont pas été trop durs ! Solène, merci d'avoir été là ces dernières années, pour toutes les découvertes culinaires (oui, toujours la nourriture) et ta présence à mes côtés. On en a fait du chemin depuis la fameuse lettre du cours de latin. Les amis de la famille, qui sont comme une famille aujourd'hui, tout spécialement les amis des vacances. Merci à vous Monsieur Quéau, qui m'avez donné envie d'étudier la physique et monsieur Seigne qui avez confirmé cette envie.

Enfin, merci à la famille. Les tontons et les tantes de sang ou de cœur ! Kossi, Françoise, malgré les kilomètres, je ressens les ondes positives. Et bien sûr, Papa, Maman, vous êtes les meilleurs, je vous aime. Merci pour tout, d'avoir fait de moi la personne que je suis aujourd'hui et de m'avoir constamment soutenu jusqu'ici ! Et enfin merci à mes deux étoiles porte bonheur.

Contents

List of Figures

List of Tables

Introduction

Neutrinos are one of the most abundant particles in the Universe. Every second, trillions of neutrinos are passing through our bodies without leaving any trace. Those ghostlike particles are created through nuclear processes inside stars, particle accelerators, nuclear reactors and even... Bananas! Despite their abundance, they are still very mysterious particles. Among the mysteries of neutrinos, even their mass is still unknown. Further understanding of the neutrino nature could help to understand the matter-antimatter asymmetry, and thus why there is something instead of nothing. Moreover, if photons created inside the Sun's core take thousands of years to diffuse to the surface while they lose all the information from their creation process, it takes only a few seconds for neutrinos. The study of neutrinos can provide the necessary information on the Sun core for a deeper understanding the nuclear interactions inside stars. The biggest source of neutrinos was the Big Bang. After the Big Bang, the Universe was so dense and hot that even light could not escape. As it cooled down, the Universe became transparent to light and the photons created at that time escaped bearing the information of the Universe state roughly 380 000 years after the Big Bang. This light is the oldest light of the Universe. Neutrinos escaped at around a second after the Big Bang, and the measurement of those relic neutrinos would give a window to the earliest time of the Universe never observed.

The chapter 1 of this thesis, presents the history the of neutrino discovery from the first hypothesis of W. Pauli in 1930 to the latest experiments of the past decade. The neutrino oscillation, one of the most intriguing properties of neutrinos, will then be introduced with its discovery and its theory. This theory of oscillations is successful in explaining almost all the experimental results among a wild variety of sources, energies and baselines. Nevertheless three types of oscillation anomalies were observed by past experiments. Those three types of anomalies, so called *gallium anomaly*, *accelerator anomaly* and *reactor antineutrino anomaly*, could be a hint for the existence of a neutrino state not described by the Standard Model of particle physics: the sterile neutrino. Several experiments that were built during the past decades could probe the oscillation of neutrinos toward an eV-scale sterile neutrino state. Among those experiments, very short baseline antineutrino experiments are among the best candidates to probe the existence of a sterile neutrino state. Those experiments consist in a neutrino detector placed at a few meters from a nuclear reactor core in order to measure the variations of the reactor antineutrino flux at different distances from the core.

The chapter 2 presents the SoLid (**S**hort baseline **o**scillation search using a **Li**thium-6 **d**etector) experiment. It is a very short baseline reactor antineutrino experiment and the detector on which I have worked during my PhD studies. The experiment operates a highly segmented detector located at the BR2

research reactor in Mol (Belgium). The detector uses a novel technology based on a hybrid scintillation with plastic scintillator cubes (PVT) combined with ^{6}LiF:ZnS(Ag) layers. It is composed of 12 800 detection cells connected to 3 200 MPPCs by wavelength shifting (WLS) fibres. This high segmentation allows a signal selection background rejection through the use of the events topologies. The detector is placed at a baseline from 6.3 to 8.9 m from the reactor core and is operational since April 2018. The chapter will detail the detector features, the main backgrounds faced and the reconstruction of the events.

The energy calibration of the detector is presented in detail in the chapter 3. It is the first task presented in this thesis which I conducted. This work consisted first in the measurement of the gain and the pedestal of each MPPC in the detector, and to correct for possible drifts. Then a measurement of each attenuation length of the fibres and the quantification of the optical coupling between the fibres and the MPPCs was performed. The last step of the calibration was the measurement of the light yield of the detection cells to convert the number of visible photons read out by the MPPCs into an energy deposit inside the detector. All this calibration work was then used to tune the simulation of the detector response, and some Data to Monte-Carlo comparisons were performed to show the goodness of this tuning. The calibration of the detector is an essential step for any experiment, as all measurements will depend on the detector calibration. Furthermore, calibrations allow computing the necessary inputs to build a trustworthy simulation of the detector response that can be used later on in an analysis.

The second main task of this thesis is the analysis on the extraction of the antineutrino signal. This analysis is detailed in chapter 4. With the level of background faced on SoLid, an analysis that uses all the information of the signal in the detector is necessary to reject as much background as possible while keeping a signal efficiency sufficient for the search for an oscillation signal toward a sterile neutrino. An analysis based on a boosted decision tree will be presented in this chapter, along with the antineutrino extraction from reactor data through background subtraction.

The last chapter of this thesis, the chapter 5, describes the current situation of the SoLid sensitivity to sterile neutrino oscillation with the known uncertainties and signal. Then, leads to improve the performances of the experiment will be presented with an alternative analysis and the SoLid Phase II data.

Chapter 1

Neutrino physics

1.1 A brief history of the neutrino

In the beginning of the XX$^{\text{th}}$ century, the energy spectrum of the product of the β decay was one of the most troubling issues in the particle physics community. As for the α and γ radioactivities, the electron emitted from a β decay was expected to have a well-defined energy, as it was expected to carry all the energy of the interaction. However, contrary to what was expected, the energy spectrum of the electrons appeared to be continuous. Various attempts to explain this phenomenon were made like secondary effects happening to the electrons, properties of the atoms to have different decay energies, the violation of energy momentum conservation was even proposed. None of those explanations were totally satisfactory. In 1930, in a famous letter to the scientific community, W. Pauli proposed an explanation for the continuous energy spectrum of the electrons produced in β decays: he introduced a new particle that he called the *neutron*. This *neutron* should be neutral, to respect the charge conservation, of spin $1/2$ to respect the spin conservation, and light particle to explain the shape of the electrons energy spectrum. After the discovery of what we call now a neutron, E. Fermi proposed a new name for this particle after a proposition of E. Amaldi: the *neutrino* (little neutron). In 1933, Fermi published his theory of beta decay [1] that finished to set Pauli's particle as the primary suspect for the missing energy in the electron energy spectrum with the now well known equation of the β decay of the neutron:

$$n \rightarrow p + e^{-} + \overline{\nu}_e \tag{1.1}$$

If the neutrino hypothesis seemed to be the perfect answer to the β decay energy spectrum question, it still needed a direct observation with an experimental proof of its existence. The last step before the detection of neutrinos came with the proposition of H. Bethe and R. Peierls [2]. Because neutrinos enter a creation process with the β decay, they proposed an *annihilation* (understood as a destruction mechanism) process with the so-called inverse beta decay (IBD). It consists in a (anti)neutrino, interacting with a proton to produce a neutron and a positron:

$$\overline{\nu}_e + p \rightarrow n + e^{+} \tag{1.2}$$

However, the estimation of the cross-section of the interaction was at the level of 10^{-44} cm^2 for 2 MeV

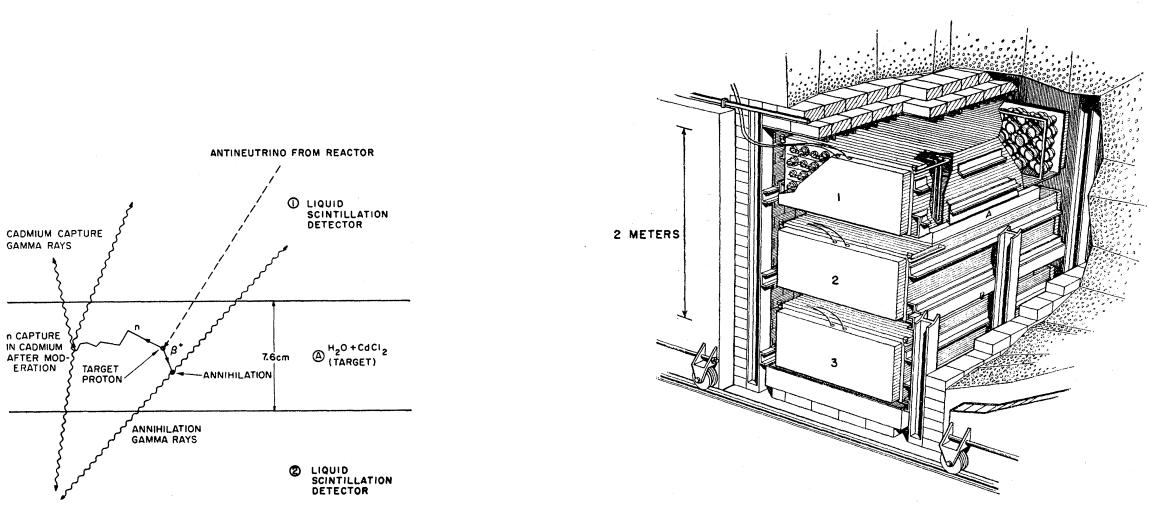


Figure 1.1: Left: Schematic diagram of the antineutrino detection. Right: Sketch of the Reines and Cowan experiment. The scintillating detectors 1, 2 and 3 are placed around the two water tanks A and B [3].

neutrinos. This corresponds to a mean free path of about 1600 light years in water. Given the low cross-section of the interaction, the neutrino detection, required a very intense neutrino source and a really low level of background and was thought hardly impossible to measure by many. The first experiment to perform the detection of electron antineutrinos was the F. Reines and C. Cowan experiment in 1956 [3]. The experiment consisted of two tanks filled with water mixed with $CdCl_2$ surrounded by three liquid scintillator detectors. Antineutrinos were produced by a nuclear reactor at Savannah River. They interacted with the free protons of water via IBD. The resulting positrons annihilated quasi instantly in two annihilation gammas that were detected in coincidence by the scintillation detectors on the two sides of the water tanks. A second signal was then generated from the neutron capture by the cadmium, with the resulting gamma rays measured also in coincidence by two liquid scintillation detectors. A scheme of the detection procedure can be seen on the figure 1.1. The time coincidence between the two signals was the final piece that was needed to have a sufficient background reduction and finally observe (anti)neutrinos. The coincidence method developed by Reines and Cowan for the IBD detection has been the major detection method of reactor antineutrinos for decades and is still used nowadays.

The first detection of another flavoured neutrino was performed in 1962 with the discovery of the muon neutrino at the Brookhaven National Laboratory by L. Lederman, M. Schwartz and J. Steinberger [4]. While the previous neutrino experiments used nuclear reactors as neutrino sources, the neutrino beam was created with accelerated protons at 15 GeV colliding a beryllium target to generate the pions. The charged pion decayed in a muon and a neutrino:

$$\pi^{\pm} \rightarrow \mu^{\pm} + \overset{(-)}{\nu} \tag{1.3}$$

The neutrinos were then detected in an aluminium spark chamber via one of the following interac-

tions:

$$\begin{aligned} \nu_e + n &\rightarrow p + e^- \\ \bar{\nu_e} + p &\rightarrow n + e^+ \\ \nu_\mu + n &\rightarrow p + \mu^- \\ \bar{\nu_\mu} + p &\rightarrow n + \mu^+ \end{aligned} \tag{1.4}$$

The measurement of (anti)electrons or (anti)muons in the direction of the beam could then give the nature of the neutrinos generated in the pion decay. The experiment showed a clear excess of muon events compared to electron events, and was thus the first discovery of muon neutrinos with the proof that they are different from electron neutrinos.

After the tau lepton discovery in 1975 at the SLAC [5], the existence of a third neutrino, the tau neutrino, was postulated. The LEP experiment, that measured the Z^0 decay width, then constrained the number of light active neutrinos to three [6]. The last predicted neutrino was experimentally confirmed in 2000 by the DONUT experiment at Fermilab [7]. The tau neutrinos were generated through collisions of 800 GeV protons accelerated in the Fermilab Tevatron with a tungsten target. Among the particles produced by the collision, the charged meson D_s then decayed into a tau particle and a tau antineutrino:

$$\begin{aligned} D_s^- &\rightarrow \tau + \overline{\nu}_\tau \\ D_s^+ &\rightarrow \tau^+ + \nu_\tau \end{aligned} \tag{1.5}$$

The decays of the $\tau^{(+)}$ also contributed to the neutrino beam:

$$\begin{aligned} \tau &\rightarrow \nu_\tau + l + \overline{\nu}_l \\ \tau^+ &\rightarrow \nu_\tau + l^+ + \nu_l \end{aligned} \tag{1.6}$$

where l represents either an electron or a muon and $\overset{(-)}{\nu_l}$ the associated (anti)neutrino. The neutrinos were then detected with a nuclear emulsion target with charged current interaction as in equation 1.4:

$$\begin{aligned} \nu_\tau + n &\rightarrow p + \tau^- \\ \overline{\nu_\tau} + p &\rightarrow n + \tau^+ \end{aligned} \tag{1.7}$$

In their final result, the DONUT collaboration measured 9 $\overset{(-)}{\nu_\tau}$ events for 1.5 background events, achieving the detection of the last fermion of the Standard Model of the particle physics.

1.2 The Standard Model of particle physics

1.2.1 Global presentation of the Standard Model

During the second half of the XX$^{\text{th}}$ century, the particle physics discoveries did not happen solely in neutrino physics. The building of bigger and bigger particle accelerators led to the discovery of a broad quantity of new particles and new interactions between them. Theories and observations of specific interactions were blooming, but there was a lack of a global framework to join all those measurements. This arrived in the beginning of the 1970s, when the Standard Model of particle physics was built. This model unifies three of the four fundamental interactions of the Universe: the weak interaction, the strong

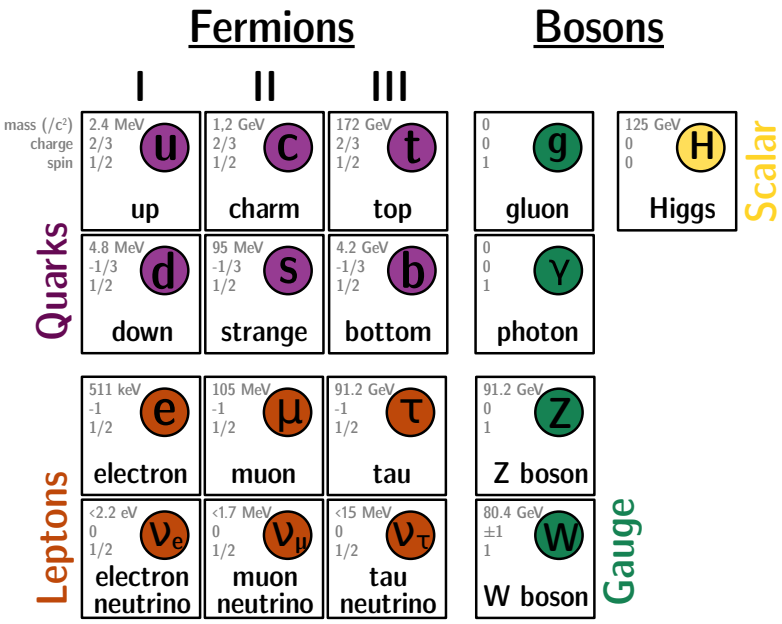


Figure 1.2: The elementary particles of the Standard Model. Figure taken from [8].

interaction and the electromagnetic interaction. As represented on the figure 1.2, the model is composed of elementary particles that can be decomposed into two types:

• **The fermions:** The fermions are the 12 elementary components of matter. They are half spin particles and are composed of two families, the quarks and the leptons. The quarks are sensitive to all three interactions, while the leptons are not sensitive to the strong interaction. The neutrinos, as non-charged leptons, are only sensitive to the weak interaction. This explains why they are elusive particles that hardly interact with matter. Each fermion family is composed of three generations. Each generation has a higher mass than the previous one and a lower lifetime. All generations are supposed to feel the same interactions.

• **The bosons:** The bosons are integer spin particles. Four of them mediate the fundamental interactions and are spin 1 particles. The electromagnetic interaction is carried by photons, the weak interaction is carried by W and Z bosons and the strong interaction is carried by gluons. The last boson is a spin 0 particle: the Higgs boson. It is responsible for the mass creation process of all fermions and massive bosons.

1.2.2 Neutrinos in the Standard Model

In the section 1.1, we have seen that the three generations of neutrinos (ν_e, ν_μ, ν_τ) of the Standard Model have been detected, with the detection of the ν_τ as one of the validation of the model. Each of the generation neutrino is called a flavour. In the Standard Model, a neutrino of a given flavour ν_l is always produced with the counterpart charged lepton l. The model describes very well the neutrino interactions

via the exchange of Z or W bosons, however some features of the neutrinos are still not well understood. First, their true nature is still unknown. As neutral particles, neutrinos could be Majorana particle (equal to the antineutrinos) or Dirac particles (distinct from the antineutrinos). Furthermore, because the weak interaction only interacts with particles having a left-handed chirality, only left-handed neutrinos are seen in nature. In its current formulation, the mass creation in the Standard Model requires both the left-handed and the right-handed forms of a particle. Thus, the Standard Model does not provide any mass creation process for the neutrinos. However, neutrino oscillations confirmed the existence of at least two massive neutrinos. This effect will be described in the section 1.3.2. Finally, various neutrino oscillation anomalies were observed by various experiments with a wrong prediction of the neutrino rates. This could be the hint of sterile neutrinos not predicted by the Standard Model. This last hypothesis is discussed in the section 1.5.

1.3 The neutrino oscillations

1.3.1 First experimental observation: the solar neutrino problem

Because of the low cross-section of neutrino interactions with matter, neutrinos are the best candidates to act as probes from nuclear reactions inside the Sun. J. Bahcall and collaborators developed a Standard Solar Model (SSM) during the end of the XXth century [9] that gave a prediction of the solar neutrino flux and energies. In the late 1960s, the Homestake experiment was designed to measure the solar neutrino flux [10]. The experiment consisted of a tank full of C_2Cl_4 and used the reaction of neutrinos with chlorine to measure the neutrino rate:

$$\nu_e + {}^{37}\mathrm{Cl} \rightarrow e^- + {}^{37}\mathrm{Ar} \tag{1.8}$$

The neutrino flux could then be measured by extracting the ^{37}Ar and counting the decay in a low background environment. This led to the solar neutrino problem: a significant deficit of neutrinos in the Homestake measurement compared to the Standard Solar Model prediction was seen. Both results were tested and verified, but nothing could explain this discrepancy. Later on, SAGE [11] and GALLEX [12], two neutrino experiments using gallium as a target, showed a deficit of solar neutrinos compared to the rate predicted by the Solar Model. In the year 1996, the Super-Kamiokande experiment, successor of the Kamiokande experiment, started its operations in the Kamioka mine, in Japan. The experiment is made of a Cherenkov detector with 50 kt of pure water. When neutrinos interact in the water via elastic scattering on the electrons, the accelerated electrons emit a Cherenkov light read out by photomultipliers. The experiment measured a $45\ \%$ deficit of solar neutrinos compared to the SSM prediction [13]. Furthermore, the Super-Kamiokande detector was sensitive to interactions of higher energy neutrinos created in the atmosphere. $\overset{(-)}{\nu}_e$ and $\overset{(-)}{\nu}_\mu$ could be detected via charged current interaction with the creation of respectively electrons or muons. The two particles were identified with their unique Cherenkov light emission pattern. With a reconstruction of the incident angles of the neutrinos, the Super-Kamiokande experiment showed a lower rate of ν_μ that travelled through the earth compared to the ν_μ that came from above the detector, but not such effect in ν_e [14]. This was another strong hint on the disappearance of neutrinos correlated to their travelled distance.

The final solution to the solar neutrino problem was provided by the SNO experiment [15]. The detector was composed of $1\ \mathrm{kt}$ of heavy water. Like the Super-Kamiokande experiment, the solar neutrinos

could be detected via electron scattering:

$$\nu + e^- \rightarrow \nu + e^- \tag{1.9}$$

This process is accessible for all neutrino flavours via neutral current interaction, with the exchange of a Z boson. It can also be achieved with a charged current interaction, via the exchange of a W boson, for electron neutrino only with a cross-section six time larger than the neutral current interactions. However, the presence of deuteron allowed new detection channels for neutrinos. The first one is a charged current reaction:

$$\nu_l + d \rightarrow l + p + p \tag{1.10}$$

with ν_l a type of neutrino and l the charged lepton associated. Due to the energies of the solar neutrinos, this channel was only sensitive to electron neutrinos. The last detection channel is the neutral current interaction, via the exchange of a Z boson:

$$\nu_l + d \rightarrow \nu_l + p + n \tag{1.11}$$

This last interaction channel is sensitive to all active neutrino flavours. This channel allowed a measurement of the total solar neutrino rate consistent with the SSM prediction, while the charged current channel confirmed the measurement of previous experiments. On the figure 1.3 can be seen the final combined result from Super-Kamiokande and SNO for the solar neutrino flux measurement. As one can see on the figure, both experiments are in agreement in the ν_μ and ν_τ neutrino flux versus ν_e flux measured. The results are also in agreement with the flux predicted by the Solar Model. The solar neutrino problem had found an explanation: the oscillation from electron neutrino to other flavours.

1.3.2 Neutrino oscillation: the theory

The first hypothesis of neutrino oscillations was proposed by B. Pontecorvo in 1957 [17]. He proposed the hypothesis of neutrino oscillations to antineutrinos. The flavour oscillation theory was introduced by Z. Maki, M. Nakagawa, and S. Sakata in 1962 [18]. They proposed an oscillation theory with two neutrinos, muon and electron, as the tau neutrino had not yet been discovered. The oscillation was then extended in the three-flavour model. The theory relies on neutrino being massive particles. If the flavour eigenstates of the neutrinos (ν_e, ν_μ, ν_τ), eigenstates of weak interactions, are different from their mass eigenstates (ν_1, ν_2, ν_3), eigenstates of the propagation in space, a flavour mixing can occur. In the case of three neutrinos, the mixing can be written as

$$\begin{bmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{bmatrix} = U_{PMNS} \begin{bmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{bmatrix} \tag{1.12}$$

with U_{PMNS} the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix that describes the neutrino flavour mixing. This matrix can be described with three mixing angles (θ_{12}, θ_{13}, θ_{23}) and a CP violation phase δ_{CP}. To describe the flavour oscillation, the PMNS matrix can be written as the product of (almost) three rotation matrices, with a complex phase in one of them:

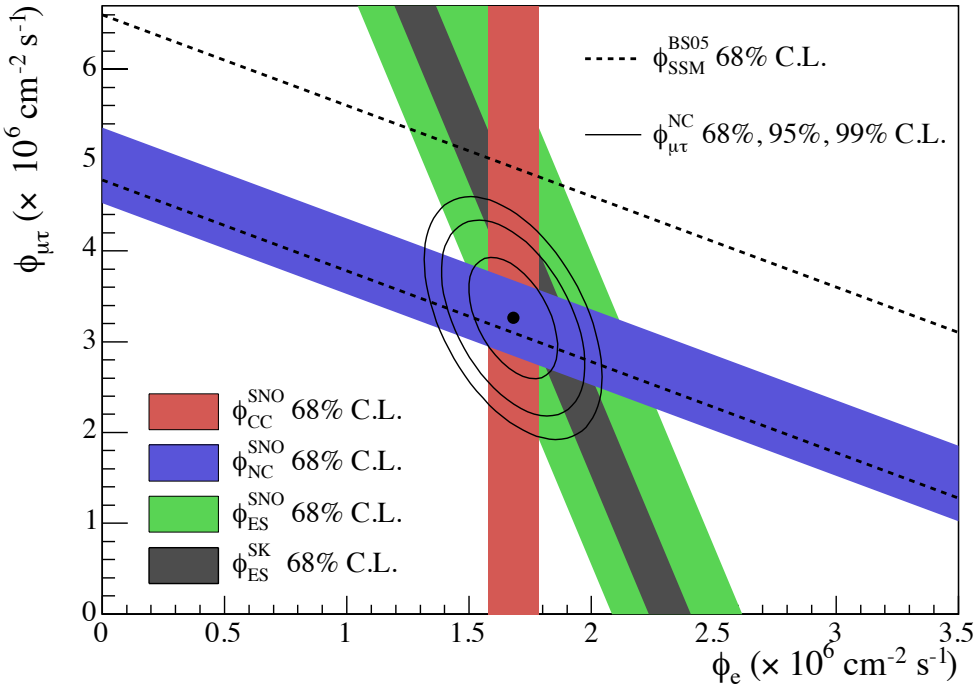


Figure 1.3: ν_μ and μ_τ neutrino flux versus ν_e flux for solar neutrinos. Charged current, neutral current and electron scattering measurements by SNO experiment are represented by the colour bands. The Black band represents the electron scattering in Super-Kamiokande results. The dotted lines represent the prediction from Solar Model calculation and the plain lines represent the 68%, 95% and 99% joint probability of the two flux [16].

$$U_{PMNS} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{bmatrix} \begin{bmatrix} c_{13} & 0 & s_{13}e^{-i\delta_{CP}} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta_{CP}} & 0 & c_{13} \end{bmatrix} \begin{bmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{1.13}$$

where $c_{ij} = \cos(\theta_{ij})$ and $s_{ij} = \sin(\theta_{ij})$. Weak eigenstates, ν_α, which are the eigenstates detected in the neutrino experiments, can be written as:

$$|\nu_\alpha\rangle = \sum_j U^*_{\alpha j} |\nu_j\rangle \tag{1.14}$$

with ν_j the mass eigenstate j and $U^*_{\alpha j}$ the component α, j of Hermitian transpose of the PMNS matrix. To understand the oscillation process, one can compute the wave function of the neutrino at a time t by applying the Schrödinger equation. A mass eigenstate j that travelled a distance L at a time t becomes in natural units ($\hbar = 1$, $c = 1$):

$$|\nu_j(t)\rangle = e^{-i(E_j t - p_j L)} |\nu_j(0)\rangle \tag{1.15}$$

with E_j the neutrino's energy and p_j its impulsion. Because of their very small masses, the neutrinos travel almost at the speed of light, hence one can write $t \simeq L$ and their energy becomes:

$$E_j = \sqrt{p_j^2 + m_j^2} \simeq p_j + \frac{m_j^2}{2p_j} \tag{1.16}$$

Equation 1.15 becomes then:

$$|\nu_j(t)\rangle = e^{-iLm_j^2/(2p_j)} |\nu_j(0)\rangle \tag{1.17}$$

The state $|\nu_\alpha\rangle$ can then be computed at a time t:

$$|\nu_\alpha(t)\rangle = \sum_j U^*_{\alpha j} e^{-iLm_j^2/(2p_j)} |\nu_j(0)\rangle \tag{1.18}$$

Because of the very small mass of the neutrinos, one can replace p_j by E the average energy of the mass eigenstates. By writing the state $|\nu_\alpha(t)\rangle$ in the flavour basis, it becomes:

$$|\nu_\alpha(L, E)\rangle = \sum_j \sum_\beta U^*_{\alpha j} e^{-iLm_j^2/(2E)} U_{\beta j} |\nu_\beta\rangle \tag{1.19}$$

For a neutrino emitted in the state ν_α with an energy E, the probability to be detected with a flavour ν_β after travelling a distance L is then:

$$P_{\alpha\to\beta} = |\langle\nu_\beta|\nu_\alpha\rangle|^2 = \sum_{j,k} U_{\alpha k} U^*_{\beta k} U^*_{\alpha j} U_{\beta j} e^{-iL\Delta m^2_{kj}/(2E)} \tag{1.20}$$

where Δm^2_{kj} is the difference between the squared mass m^2_k and m^2_j. In practice, three mass differences are defined: Δm^2_{21}, Δm^2_{31} and Δm^2_{32}. If β and α represent the different flavours, $P_{\alpha\to\alpha}$ is usually called a disappearance probability, and $P_{\alpha\to\beta}$ is usually called an appearance probability. From the equation 1.20, three oscillation lengths can be derived:

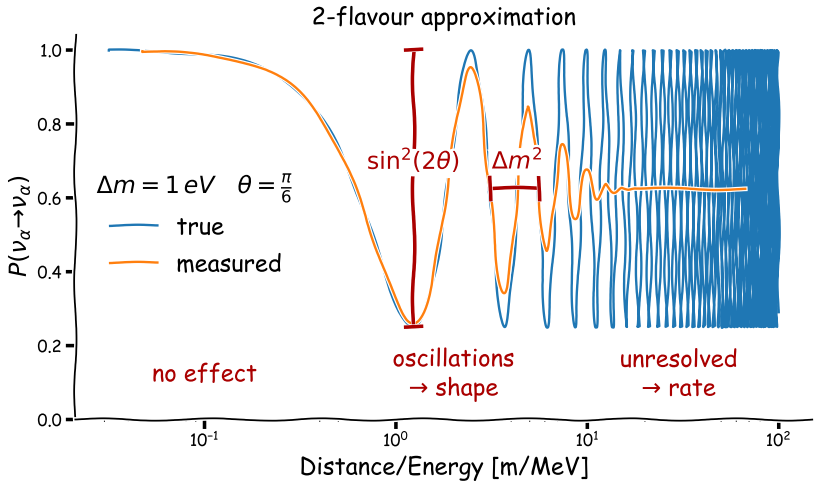


Figure 1.4: Disappearance probability for neutrino ν_α as a function of L/E in the two flavour approximation for a Δm^2 at 1 eV [19].

$$\lambda_{kj} = \left| \frac{4\pi E}{\Delta m^2_{kj}} \right| \tag{1.21}$$

For an oscillation experiment, one wants to maximize the oscillation probability. To do so, the baseline L, defined as the distance between the neutrino source and the detector, has to be as close as possible to one of the oscillation lengths λ_{kj}. Each oscillation length is constrained by the energy E of the neutrino source and the mass difference Δm^2_{kj} one wants to measure. Furthermore, the mass difference Δm^2_{21} is one order of magnitude smaller than Δm^2_{31} and Δm^2_{32} in absolute value, so Δm^2_{31} is assimilated to Δm^2_{32}. Those large mass differences, associated to a large mixing of the neutrinos, allow a two-flavour approximation where a given flavour appearance or disappearance probability is driven by one mass difference Δm^2_{kj} and one mixing angle θ_{kj}. On the figure 1.4 is represented the shape of the disappearance probability as a function of L/E in the two-flavour approximation. The oscillation is maximal for a ratio L/E at the same order of magnitude of Δm^2. For a lower L/E, the oscillation effect is not happening. For higher L/E, the oscillation is averaged by resolution effect and not visible: as one can see on the figure 1.4, for a ratio L/E too high, the oscillation effect is averaged into an average probability of $50\,\%$. The oscillation amplitude, defined as the maximum depth of the oscillation probability, is driven by the mixing angle.

From the two-flavour approximation three sectors were defined to measure the three mixing angles and the two mass differences: the solar sector, the atmospheric sector and the reactor sector.

1.3.2.1 The solar sector

It is a measurement of neutrinos at the MeV scale and at long baselines. The experiments that probe the solar sector are sensitive to θ_{12} and Δm^2_{21}. The solar neutrino experiments presented in the section 1.3.1 are sensitive to this type of oscillation. For those experiments, the oscillation is enhanced by matter effect due to the matter density inside the Sun, described by the Mikheyev-Smirnov-Wolfenstein (MSW) effect [20]. Long baseline reactor neutrinos like the Kamioka Liquid-scintillator Antineutrino Detector

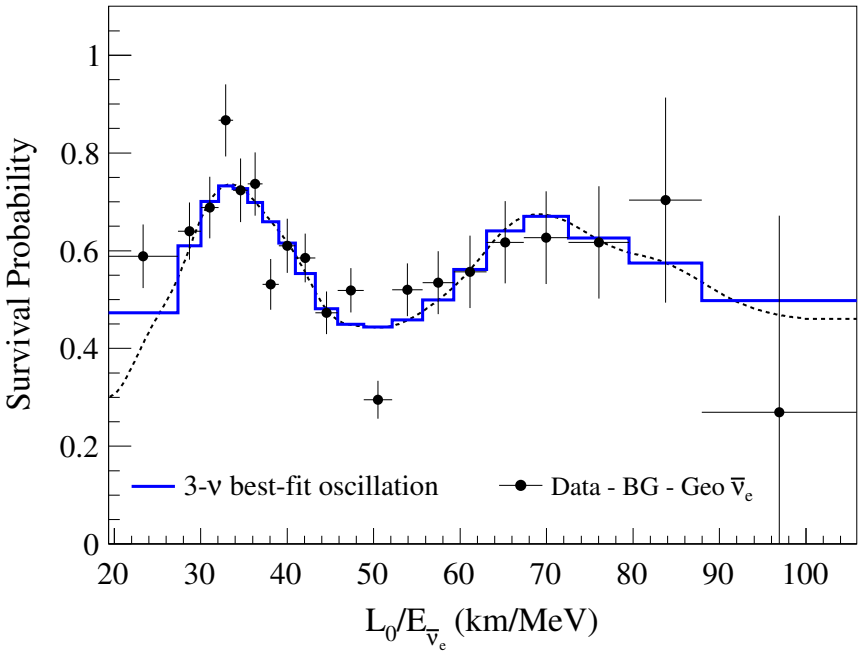


Figure 1.5: Ratio of the observed electron antineutrinos in the KamLAND experiment to the no-oscillation expectation vesus L_0/E with L_0 the average baseline of 180 km [21].

(KamLAND) experiment are also sensitive to θ_{12} and Δm^2_{21}. The KamLAND detector is located in Japan and surrounded by nuclear reactors at an average distance of 180 km. On the figure 1.5, one can see a clear oscillation signal from the experiment as a function of L/E [21].

1.3.2.2 The atmospheric sector

The neutrinos measured in the atmospheric sector are high-energy neutrinos, at the GeV scale. The experiments probing the atmospheric sector are sensitive to θ_{23} and Δm^2_{32}. The neutrinos can be generated by interaction of high energy cosmic rays in the atmosphere that produce $\overset{(-)}{\nu}_\mu$ and $\overset{(-)}{\nu}_e$. The first measurement of the oscillation of atmospheric neutrinos was performed by Super-Kamiokande as said earlier. Atmospheric neutrinos can also be measured by neutrino telescopes like ANTARES [22] or IceCube [23]. Neutrinos ν_μ produced by accelerators are also a source of neutrinos used to probe the atmospheric sector, with experiments like the NuMI Off-Axis ν_e Appearance (NOνA) [24] or the Tokai to kamioka (T2K) [25]. Those experiments use a detector at a short distance from the accelerator to measure the rate of neutrinos without oscillation and a far detector to measure the oscillated neutrino rate. The NOνA experiment uses a neutrino beam created at Fermilab and detects the neutrinos 810 km further in the Minnesota. The T2K experiment uses a neutrino beam created at the J-PARC laboratory and uses the Super-Kamiokande detector as far detector placed at 295km from the neutrino source. Those experiments are also sensitive to the θ_{13} and δ_{CP} when taking into account the exact oscillation probability. The T2K provided the first hint of the non zero value of the angle θ_{13}.

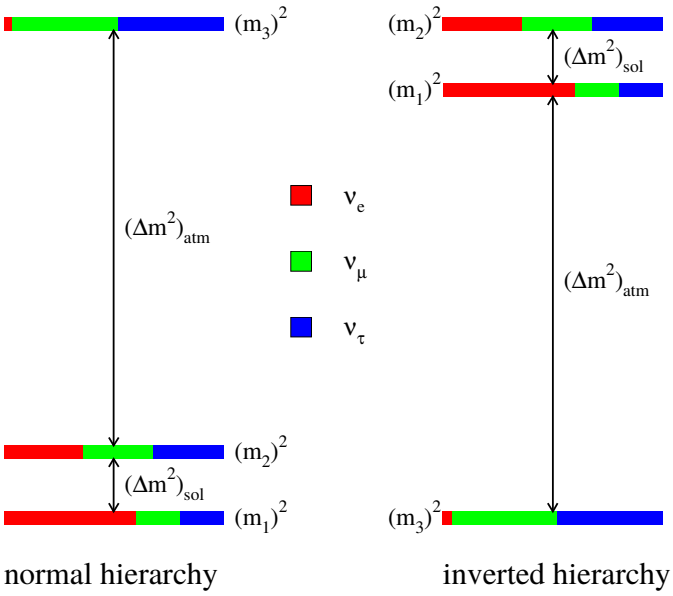


Figure 1.6: Scheme of the two different mass hierarchies. Each colour indicates the probability of detecting a given flavour for a given neutrino mass eigenstate. Δm^2_{sol} represents Δm^2_{21} and Δm^2_{atm} represents either Δm^2_{32} for the normal ordering or Δm^2_{31} for the inverted ordering. The scheme is taken from [29].

1.3.2.3 The reactor sector

The experiments in the reactor sector are sensitive to θ_{13} and Δm^2_{31}. This last angle is the smallest of the three and the most difficult to measure. The first proof of a non-zero value of θ_{13} was performed by the T2K experiment [25]. However, better precision measurements were performed later with short baseline reactor neutrinos like the Double Chooz [26], Daya Bay [27] or Reno experiments [28] with $\bar{\nu_e}$ disappearance measurement. Those three experiments are detecting antineutrinos from nuclear reactor via inverse beta decay in gadolinium-doped liquid scintillator detectors. As for long baseline accelerator experiments, the neutrino measurements of the far detector, at about one kilometre from the reactor, are compared to neutrino models. Those models are constrained by the measurements of antineutrinos at few hundred of meters from the reactor with a near detector.

1.3.2.4 Mass hierarchy

Only squared mass differences can be measured with neutrino oscillation experiments. The direct masses of the neutrinos are thus still unknown. By convention with the matter effect in the Sun, Δm^2_{21} is positive, hence, $m_1 < m_2$. Furthermore, in absolute value, Δm^2_{21} has been measured as one order of magnitude lower than the two other mass differences $\left|\Delta m^2_{32}\right|$ and $\left|\Delta m^2_{31}\right|$. However, the hierarchy between the mass is still unknown, so, as represented on the figure 1.6, there are two possibilities in the neutrino mass ordering:

- Normal ordering (NO): $m_1 < m_2 \ll m_3$
- Inverted ordering (IO): $m_3 \ll m_1 < m_2$

Parameter	Normal ordering	Inverted ordering
θ_{12} [°]	$33.44^{+0.77}_{-0.74}$	$33.45^{+0.78}_{-0.75}$
θ_{23} [°]	$49.2^{+0.9}_{-1.2}$	$49.3^{+0.9}_{-1.1}$
θ_{13} [°]	$8.57^{+0.12}_{-0.12}$	$8.60^{+0.12}_{-0.12}$
Δm^2_{21} [10^{-5}eV^2]	$7.42^{+0.21}_{-0.20}$	$7.42^{+0.21}_{-0.20}$
Δm^2_{3l} [10^{-3}eV^2]	$+2.517^{+0.026}_{-0.028}$	$-2.497^{+0.028}_{-0.028}$
δ_{CP} [°]	217^{+40}_{-28}	280^{+25}_{-28}

Table 1.1: Three-flavour oscillation parameters from NuFIT global fit [30]. The mass difference Δm^2_{3l} represents Δm^2_{32} for the normal ordering or Δm^2_{31} for the inverted ordering.

.

1.3.3 CP violation

From the PMNS matrix presented in 1.13, the last oscillation parameter that was not presented in this section is the CP violation phase δ_{CP}. A value of δ_{CP} different from $0°$ or $180°$ would make the PMNS matrix a non Hermitian matrix ($U_{PMNS} \neq U^*_{PMNS}$), leading to a different oscillation probability for neutrino and antineutrino appearance:

$$P(\nu_\alpha \to \nu_\beta) \neq P(\overline{\nu_\alpha} \to \overline{\nu_\beta}) \tag{1.22}$$

CP violation processes could help to understand the matter-antimatter asymmetry observed in the Universe. This violation has already been observed with baryons, with in particular the K meson decay. However, the level of CP violation in the baryon sector is too small to explain the level of asymmetry that is seen today. A CP violation by leptons could solve the issue: it could allow potential heavy neutral leptons to undergo CP-violating decays that resulted in a larger number of particles than antiparticles in the early ages of the Universe. The long baseline accelerator experiments Minos, NoνA and T2K are the most sensitive to δ_{CP} and are giving the most precise measurement of the phase.

1.3.3.1 Global neutrino oscillation parameter measurement

From the most recent neutrino data of the latest neutrino experiments, the NuFit group [30] and the particle data group [31] conducted a global analysis and aggregated the best fit values for the neutrino mixing parameters. The global $\Delta\chi^2$ analysis can be seen on the figure 1.7 with a separation between both normal ordering and inverted ordering. The final flavour mixing parameters are presented in the table 1.1.

As one can see on the figure 1.7, there are still three measurements that are missing in the three-flavour oscillation parameters: the precise measurements of θ_{23} and δ_{CP} and the neutrino mass ordering. The upgrade of present long baseline experiments and future detectors like DUNE, Hyper-Kamiokande, KM3NeT, JUNO... should provide the precise measurement of those last parameters.

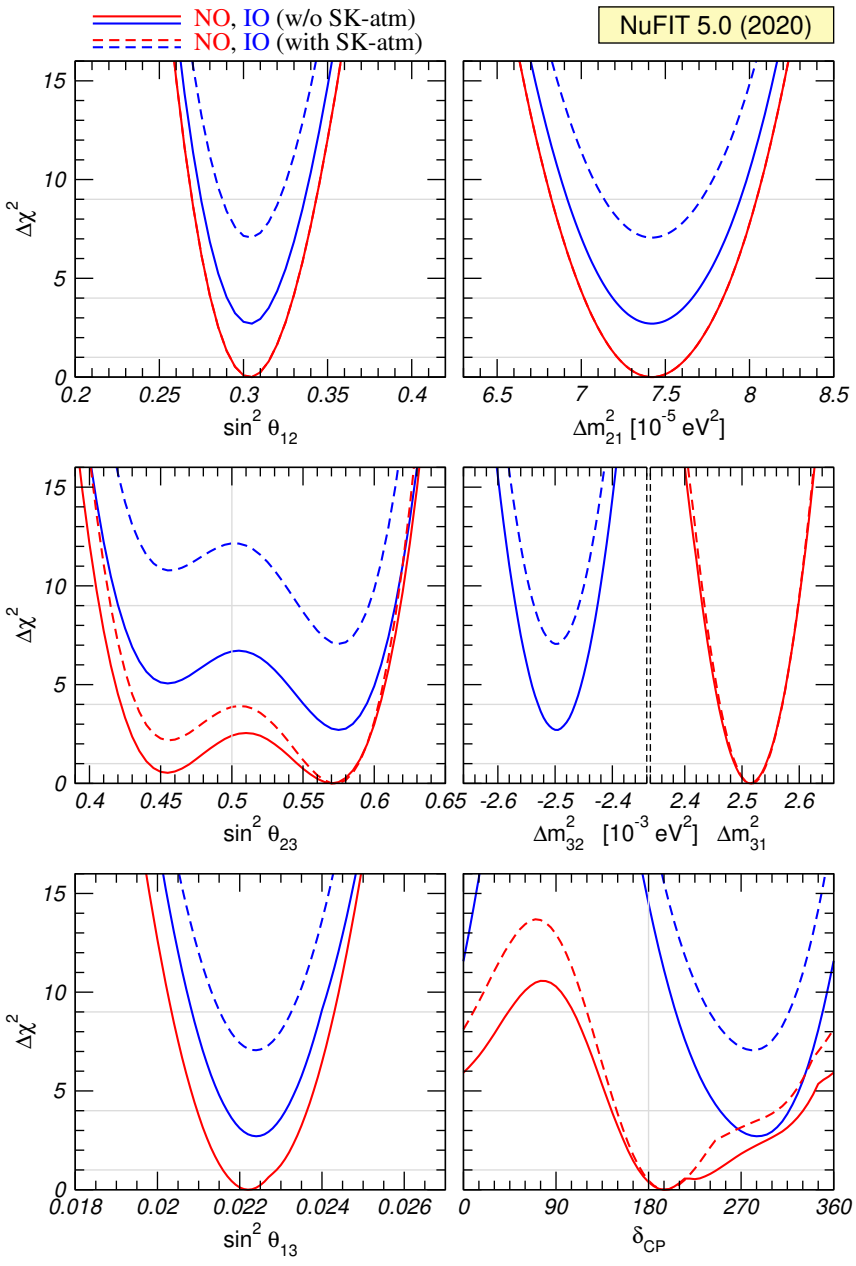


Figure 1.7: Global 3ν analysis by the NuFIT group [30].

1.4 Reactor neutrino experimental anomalies

Despite the high-precision measurements of the mixing parameters achieved by neutrino experiments, some anomalies are still observed in the detection of antineutrinos from nuclear reactors. Those anomalies could come from errors in the nuclear models of reactors that lead to a prediction error. It could also be induced by physics outside the Standard Model, with the oscillation of neutrinos into a sterile neutrino state. In this section, the calculation of the reactor antineutrino energy spectrum will be presented first, then the different experimental anomalies in reactor experiments will be discussed, to finally conclude with the description of other neutrino oscillation anomalies from neutrino beam experiments and gallium experiments.

1.4.1 Reactor antineutrino energy spectrum calculation

In nuclear reactor, the collision between slow neutrons and fissile isotopes induce the fission of the isotopes due to their large neutron-proton number asymmetry. The reaction is highly exothermic, and the energy released is usually converted into electricity. The atoms produced by the fission are highly unstable, with a number of neutrons that is still too large compared to their proton number. A cascade of beta decays starts then for all the fission products until they reach stability:

$$ {}^{\mathrm{A}}_{\mathrm{Z}}\mathrm{X} \rightarrow {}^{\mathrm{A}}_{\mathrm{Z+1}}\mathrm{X}' + e^- + \overline{\nu}_e \tag{1.23} $$

From those cascades rise many different β emitters with various energies and half lives. Furthermore, due to the fuel burn-up, the composition of isotopes involved in the cascades varies over time. A precise calculation of the antineutrino yield and energy is thus essential for antineutrino reactor experiments.

The antineutrino yield $N_{\overline{\nu}_e}(E)$ computation is based on the following equation:

$$ N_{\overline{\nu}_e}(E) = \int_0^{t_f} \frac{P_{\mathrm{th}}(t)}{\sum_k \alpha_k(t)\langle E_k\rangle} \sum_k \alpha_k(t) S_k(E) dt \tag{1.24} $$

where the integral varies between the initial time and t_f, the exposure time. The index k represents the fissile isotopes, mainly ^{235}U, ^{239}Pu, ^{241}Pu and ^{238}U. ^{235}U is the major isotope that contributes the most to the β cascades. The first fraction of the equation 1.24 is the ratio between the thermal power of the reactor P_{th} and the mean energy per fission $\langle E_k\rangle$ weighted with the isotope fraction in the fuel composition $\alpha_k(t)$. This fraction represents the number of fission per second in the reactor core. The second term is the sum of every antineutrino energy spectrum $S_k(E)$ normalized per fission weighted with the isotope fractions α_k. The isotopes fractions are obtained with simulations and the thermal power can be measured at the reactor site. The mean energy per fission has been computed for the four isotopes in [32] and are available in nuclear databases. The last parameters that need to be measured are the antineutrino energy spectra of each isotope $S_k(E)$. Two historical methods were designed to obtain this measurement, the conversion method and the summation method.

1.4.1.1 The conversion method

One of the first measurements of the reactor antineutrino reactor energy spectrum between 2 and 7.5 MeV was performed on ^{235}U at the ILL reactor in Grenoble (France) in 1980 [33]. A ^{235}U target was exposed

to a thermal neutron flux. The thermal neutrons were captured by the ^{235}U and the ^{236}U fissions into two unstable fission fragments. Those fragment then undergo a succession of beta decays as in equation 1.23 to reach stability. The energy of the electrons emitted in the beta decays were measured with the "BILL" spectrometer [34]. The electron energy spectrum was then converted into an antineutrino energy spectrum. To do so, the electron energy spectrum was reconstructed with the weighted sum of several individual beta energy spectra with a precision better than 1 %. Those spectra were computed from virtual branches via tabulated data of a beta spectrum for a given Z (proton number) from the Fermi beta decay theory. The spectra were added one by one to fit the end tail of the measured spectrum. To do so, the electron energy spectrum was split in multiple slices and an individual spectrum was fitted to the data in each slice, starting from the last slice. The antineutrino spectrum was then computed by summing the expected neutrino energy from each energy spectrum:

$$S_{^{235}U}(E) = \sum_i S_i^{\beta}(E_i^0 - E) \tag{1.25}$$

where $S_{^{235}U}(E)$ is the antineutrino spectrum for ^{235}U, i represent the different Z used in the electron spectrum decomposition, S_i^{β} is the weighted electron energy spectrum of the isotope i and E_i^0 is the end point energy of the partial β spectrum i. In 1982 the same experimental setup led to the measurement of the ^{239}Pu and a new measurement of the ^{235}U antineutrino energy spectra [35]. The conversion method was the same as in [33], with a difference on effective Z_i used in the decomposition for more precision. This last correction is due to the precision of the electron energy measurement: with a given binning, several Z could give a spectrum that fits the electron data, so the average Z of the different possibilities is taken as the effective Z_i. The modification did not change the final electron energy spectrum but improved the converted antineutrino spectrum. The measurement of the antineutrino energy spectrum was performed between 1.5 and 8 MeV. With the use of a higher target mass of ^{235}U and the reduction of the reactor power to improve the signal to background ratio, the measurement of the ^{235}U antineutrino energy spectrum has been performed up to 9.5 MeV [36]. In 1988, the measurement of the ^{241}Pu antineutrino energy spectrum between 1.5 and 9 MeV was performed at the ILL reactor [37]. The addition of new radiative and higher order corrections also led to a new evaluation of the ^{239}Pu energy spectrum. In the end, the antineutrino energy spectrum has been measured for ^{235}U, ^{239}Pu and ^{241}Pu with energy bins of 250 keV between 1.5 and 9 MeV.

In 2011, Mueller et al. [38] included sub-leading corrections to the beta spectrum that improved the precision of the antineutrino spectra calculation of the three isotopes ^{235}U, ^{239}Pu and ^{241}Pu. This method used a mixed approach by combining the conversion method to the summation method (presented in section 1.4.1.2) to reduce the systematic uncertainties of the calculation and take into account the effect of long-lived isotopes (off equilibrium effects). This mixed approach allowed the first calculation of the ^{238}U energy spectrum. In 2012, Huber [39] proposed another calculation of the ^{235}U, ^{239}Pu and ^{241}Pu antineutrino spectrum, by adding even further sub-leading effects in the beta decay and using the conversion method with only virtual branches as it was done in [33] [35] [36][37]. A comparison between the Mueller et al. and the Huber ^{235}U antineutrino spectra can be seen on the figure 1.8. The figure shows the relative difference between the calculated spectra and the original spectrum from [36]. Both spectra are consistent between 2 and 5.5 MeV but the Mueller et al. spectrum shows some oscillating behaviour at higher energies. This oscillation may be explained by a larger binning used in calculations [38] and a low statistics at high energies.

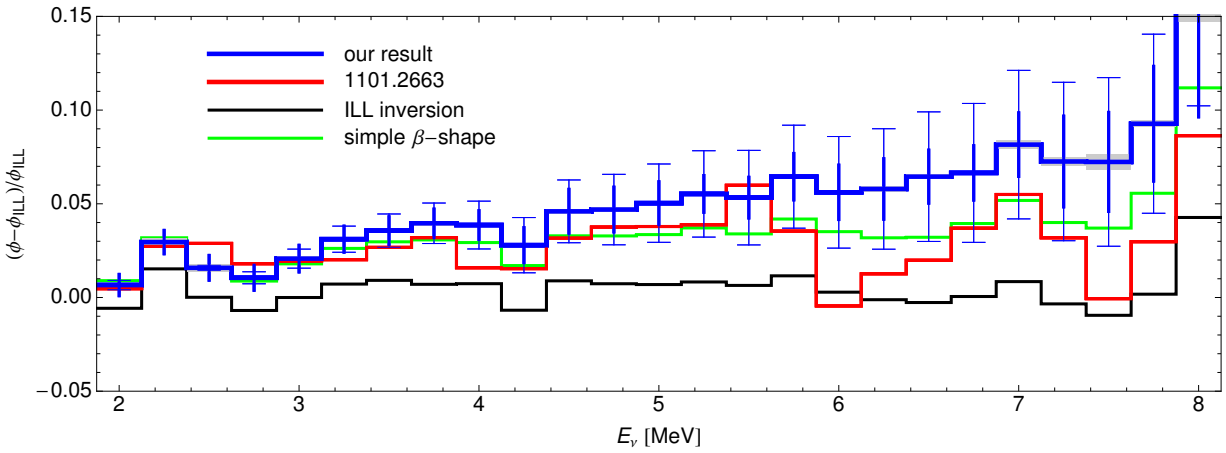


Figure 1.8: Relative difference between different ^{235}U antineutrino energy spectra and the spectrum computed in [36]. The blue line represents the calculation from Huber. The red line represents the calculation from Mueller et al. The green line represent Huber calculation using the beta decay model in [38]. The black line is a calculation from Huber using the method described in [36]. The figure is extracted from [39].

The fission of ^{238}U can only be achieved with fast neutrons, so the ILL facility could not measure the antineutrino spectrum from this isotope. In 2013, the ^{238}U antineutrino energy spectrum was finally measured with the FRM II neutron source in Germany [40]. The same method as the BILL measurement was performed: the measurement of an electron energy spectrum from beta decays induced by the fission of the isotope and the conversion of the spectrum in an antineutrino spectrum. Because of low statistics in the end tail of the energy spectrum, the previous conversion method could not be used. Instead, an empirical method developed in [37] was used. In relativistic cases ($E_e \gg m_e c^2$), both energy spectra of the neutrino and the electron tend to be similar for a given kinetic energy of the particles. The neutrino energy spectrum thus needs to be shifted by 511 keV, the electron mass, and by an additional 50 keV for an average Coulomb interaction between the electron and the nucleus:

$$S_{^{238}U}(E) = S_i^{\beta}(E - 511 \text{ keV} - 50 \text{ keV}) \times k(E) \tag{1.26}$$

with $k(E)$ a correction parameter that was computed with the BILL measurements and the neutrino rate predictions with the summation method. The antineutrino energy spectrum was also measured between 3 and 7.5 MeV with 250 keV binning.

In current experiments, the Huber model is usually used for the ^{235}U, ^{239}Pu and ^{241}Pu isotopes. The ^{238}U spectrum is then computed with the Mueller et al. method in the Huber-Mueller (H-M) model or with the Haag et al. measurement [40] in the Huber-Haag (H-H) model. The improvement of the antineutrino spectrum model through the years led to the observation of a rate anomaly in the antineutrino reactor experiments that will be discussed in the section 1.4.4. However, those models were built only recently and are highly non-trivial and current work is still trying to improve the model taking into account further sub-leading effects in the beta decays.

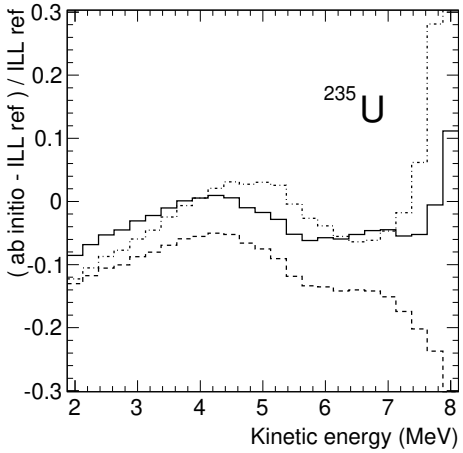


Figure 1.9: ^{235}U electron energy spectrum computed with the summation method minus to the ILL reference measurement divided by the reference measurement. The dotted dashed line represents the predicted spectrum using the ENSDF nuclear database. The dashed line was obtained by correcting the pandemonium effect in some nuclear input data. The solid line was obtained by adding missing beta contributions from the JENDL nuclear database to the previous line. The figure was extracted from [38].

1.4.1.2 The summation method

The summation method, or ab. initio method, is an alternative model to measure the antineutrino spectrum from the fission of an isotope. It was first proposed in 1958 by R. W. King and J. F. Perkins [41] and developed by others later. It relies on the available nuclear data on the beta decay for the fission product. The method is based on the weighted sum of all known beta branches involved in the beta cascades. Each branch is weighted by its activity:

$$S_k(E) = \sum_{fp} A_{fp} S_{fp}(E) \tag{1.27}$$

A_{fp} is the activity of the fission product fp and S_{fp} is the spectrum weighted by the branching ratio of the fission product. As this method does not rely on existing β^- spectrum measurement, it can be used for any reactor fuel or reactor design. However, it requires data from a vast number of nuclear isotopes and branching ratios. Furthermore, nuclear data are affected by the so-called pandemonium effect that creates systematic uncertainties in beta decay studies using Germanium detectors. For low-energy beta decays, the nuclei are in highly excited states. The transition from those states to ground state may lead to a gamma cascade, and some gamma can be missed by the detector due to inefficiencies. This pandemonium effect leads to an underestimation of the low-energy beta transitions. Due to those issues, for a long time the summation method was not precise enough to be used for precise reactor antineutrino measurement, as can be seen on the figure 1.9 from [38]. On the figure, the ^{235}U electron spectrum reproduces the ILL measured spectrum at a 10 % level.

Recent calculations have improved greatly the summation results with the use of new nuclear data [42]. The figure 1.10 shows the ratio of antineutrino energy spectra between the recent summation calculations and the Huber-Model calculations. This shows a good agreement between the two methods for the three isotopes ^{239}Pu, ^{241}Pu and ^{238}U and shows a flat ratio but with a lower rate predicted by

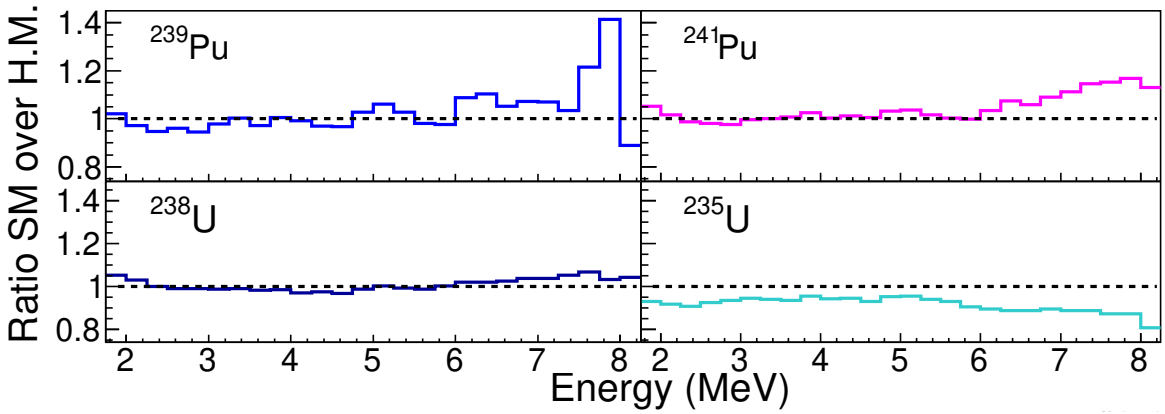


Figure 1.10: Ratio of the antineutrino energy spectra for the four main fissile isotopes between the new summation calculation and the H-M calculation. The figure was extracted from [42].

the summation model. The new calculation show a flat deficit of ^{235}U at about 5% on the whole energy range.

1.4.2 The ^{235}U rate deficit

In the evolution of the reactor antineutrino flux in Daya Bay experiment [43], a 7.8% deficit has been observed between the predicted ^{235}U induced rate compared to the best prediction from the Huber-Mueller model. This measurement has been confirmed by the Double Chooz experiment [44] and the first results from the STEREO experiment tend to also show this deficit [45]. The Daya Bay experiment measurement tends to favour the ^{235}U as the isotope primarily responsible for the reactor antineutrino anomaly. However, this measurement questions the sterile neutrino hypothesis that would affect all isotopes. An analysis of the Daya Bay results in [46], that fitted the ILL data with different forbidden transition proportions, was able to account for the rate deficit.

1.4.3 The 5 MeV bump

The so called 5 MeV bump effect is a spectral distortion in the reactor antineutrino antineutrino spectrum at 5 MeV. It has been presented by the Double Chooz experiment at the Neutrino 2014 conference and was then confirmed by the Daya Bay and RENO experiments [44] [47] [48]. An analysis which compared the reactor antineutrino experiments Double Chooz, RENO, Daya Bay and Bugey-3 was performed in [49]. The figure 1.12 shows the comparison between the measured spectra and the Huber-Muller model prediction. The distortion is not observed on the Bugey-3 data despite the similar fuel composition of the reactor for all the experiments. One explanation of the bump could thus be a calibration effect in the three other experiments. Furthermore all those experiments use liquid scintillator to measure the antineutrino spectrum and liquid scintillators are known to give a non linear energy response. These are the hypothesis investigated in this article [49] and the question is still opened.

In 2017, the Daya Bay experiment published the first extraction of the individual antineutrino spectra from the two main reactor isotopes contributing to the antineutrino emission: ^{235}U and ^{239}Pu [50]. Both isotopes energy spectra were extracted with the study of fission fraction time evolution. On figure 1.11 can be seen the two energy spectra with the Huber-Mueller model predictions. Both predictions are

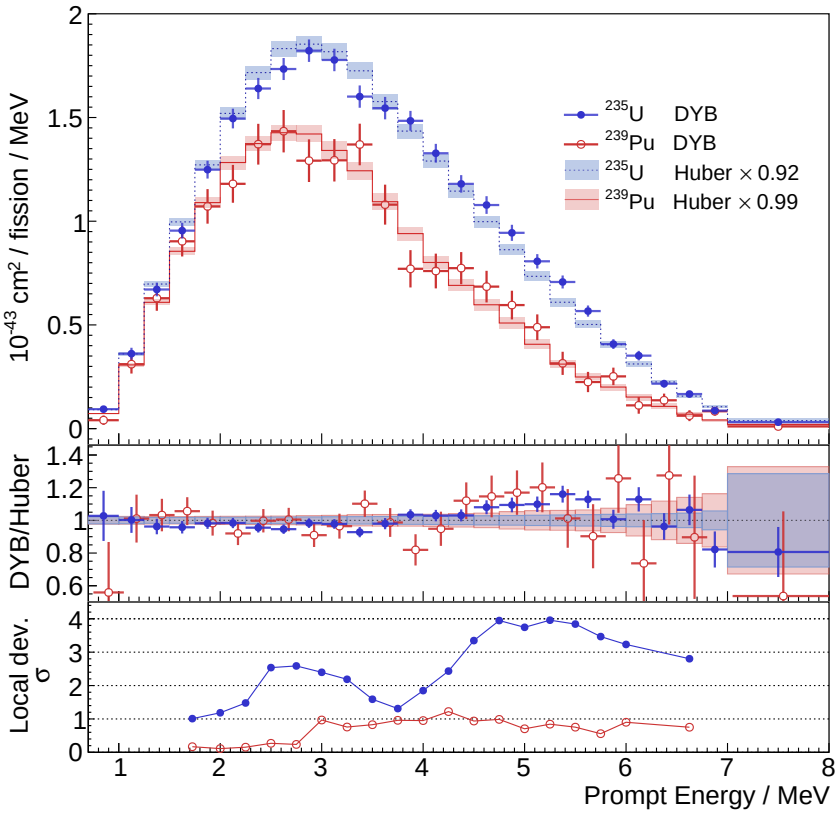


Figure 1.11: Top: Daya Bay extracted antineutrino spectrum from ^{235}U and ^{239}Pu with the normalized Huber-Mueller prediction. Middle: Ratio between the data and the predicted spectra. Bottom: Local significance of both data spectrum compared to the model prediction [50].

normalized to the best-fit number of events for each isotope. A significant spectral distortion at 5 MeV compared to the prediction can be seen for the ^{235}U.

Recently, the PROSPECT and STEREO experiments both published the antineutrino spectrum from highly enriched ^{235}U reactor [51] (PROSPECT), [52] (STEREO). Both experiments find a spectral distortion similar to the previous reactor experiments. Work on the combination of both reactor data to reduce the systematic uncertainties is ongoing. Furthermore, the addition of the forthcoming STEREO phase-III data should increase the experiment sensitivity to the spectral distortion

1.4.4 The reactor antineutrino anomaly

The new evaluation of the ^{235}U, ^{239}Pu and ^{241}Pu antineutrino energy spectrum from [38] induced a shift of about 3 % of the energy spectra with higher rates predicted compared to precedent measurements. In addition, the antineutrino rate of previous reactor experiments have been re-analysed, with a correction of the IBD cross-section with the newly computed spectra and taking into account long-lived isotopes from fissions. This led to significant deviation of the measured antineutrino rate compared to the prediction with a deficit R = 0.943 ± 0.023 meaning a 3σ significance [53]. This discrepancy is called the *reactor antineutrino anomaly* (RAA). A comparison between the expected neutrino rate and the corrected evaluation from reactor experiment can be seen on the figure 1.13. Theoretical calculation pointed out that the

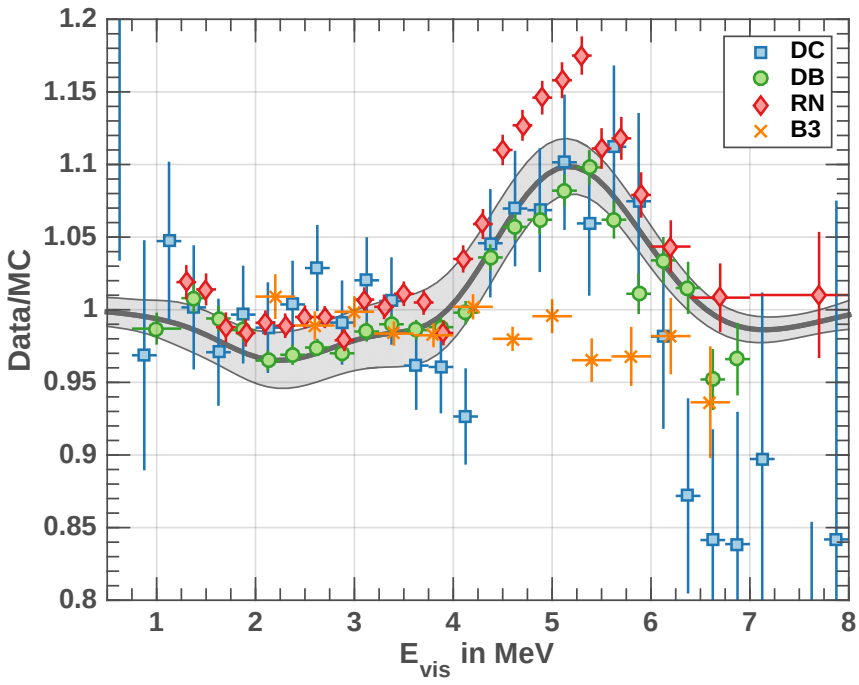


Figure 1.12: Ratio of observed reactor antineutrino spectra to Huber-Muller model prediction. The coloured points represents data from Double Chooz (DC), Daya Bay (DB), RENO (RN) and Bugey 3 (B3). The grey line represents the best fit between the data and the band represents the uncertainty on the fit. The energy represented for all experiments is the positron kinetic energy from the IBD.

current lack of knowledge of forbidden transition in beta decay could account for the antineutrino deficit with higher systematic uncertainties in the measured spectra [54]. More recent calculations showed that a better computation of the first forbidden beta decays could account for the missing neutrinos in the reactor antineutrino anomaly [55].

This anomaly could also be explained by an antineutrino oscillation into a light sterile state at very short distance from the reactor. On the figure 1.13, one can see that in the case of an oscillation to a sterile neutrino state, the oscillation is predicted to be maximal at a few meters from the reactor core. As one can see, most of the antineutrino experiments has been performed at tens or hundreds of meters from a nuclear reactor thus, the oscillation effect are averaged out. This advocates to build very short baseline reactor antineutrino experiments to probe the sterile hypothesis. With the possibility of measuring the antineutrino yield at several distances, those experiments could directly test the oscillation hypothesis without relying on an absolute antineutrino yield prediction.

1.4.5 Other neutrino oscillation anomalies

1.4.5.1 The accelerator anomaly

The Liquid Scintillator Neutrino Detector (LSND) was located at the Los Alamos LAMPF accelerator National laboratory. It studied $\overline{\nu}_\mu \rightarrow \overline{\nu}_e$ oscillations. The experiment used a LINAC to produce pions. The pions were stopped in a target and while mostly π^- interacted with the nuclei inside the target

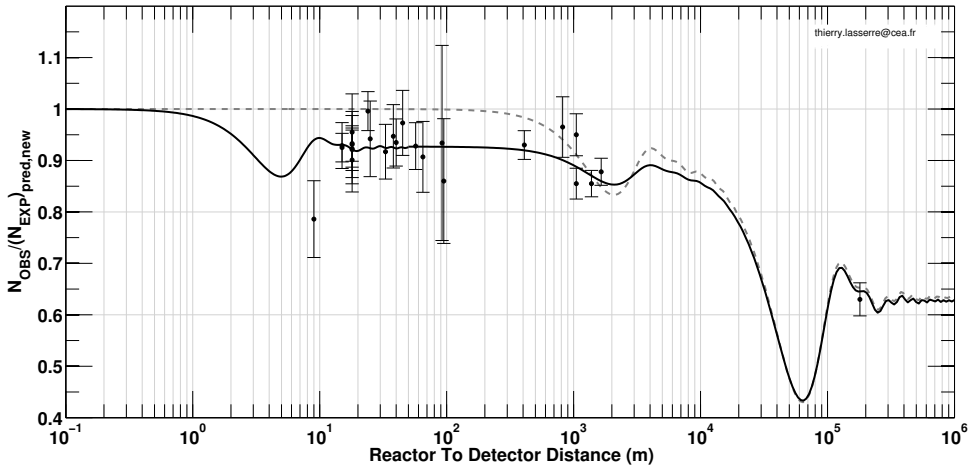


Figure 1.13: Re-analysis of the reactor antineutrino experiments. The dashed line represent the predicted oscillation with only three active neutrinos. The solid line corresponds to a three active neutrinos plus one sterile neutrino model. The figure is extracted from [56].

without neutrino emission, π^+ decays created a neutrino beam:

$$\begin{aligned} \pi^+ &\rightarrow \mu^+ + \nu_\mu \\ \mu^+ &\rightarrow \overline{\nu}_\mu + \nu_e + e^+ \end{aligned} \tag{1.28}$$

The neutrinos were then detected by a 167 t liquid scintillator. After background subtraction, the experiment observed an excess of $87.9 \pm 22.4 \pm 6.0\ \overline{\nu}_e$ events that corresponded to a 3.8 σ significance [57] which could be explained by an oscillation with a Δm^2 of around 1 eV.

The MiniBooNE experiment at Fermilab aimed to verify the oscillation observed by LSND with a similar L/E ratio. It used a neutrino beam composed of both ν_μ and $\overline{\nu}_\mu$. The experiment reported an excess in both ν_e and $\overline{\nu}_e$ with a total of 460.5 ± 99.0 events that corresponded to a 4.7 σ significance [58]. This result is in agreement with LSND first measurement, and by combining the two analysis, the best oscillation fit occurs at $\Delta m^2 = 0.041\text{eV}^2$ and $\sin^2 2\theta = 0.96$ with a 6 σ significance. However, an excess at low energy in the MiniBooNE results is not compatible with LSND measurements. The MicroBooNE experiment will aim to study this discrepancy.

1.4.5.2 The gallium anomaly

The last neutrino flux anomaly comes from the SAGE and GALLEX experiments. Both experiments were experiments studying the solar neutrinos. They both used liquid metallic gallium as a target:

$$^{71}Ga + \nu_e \rightarrow {}^{71}Ge + e^- \tag{1.29}$$

The germanium was then extracted and counted via its radioactive decay. In order to calibrate those experiments, ^{51}Cr and ^{37}Ar sources were placed inside the detectors. Those sources are both mono-

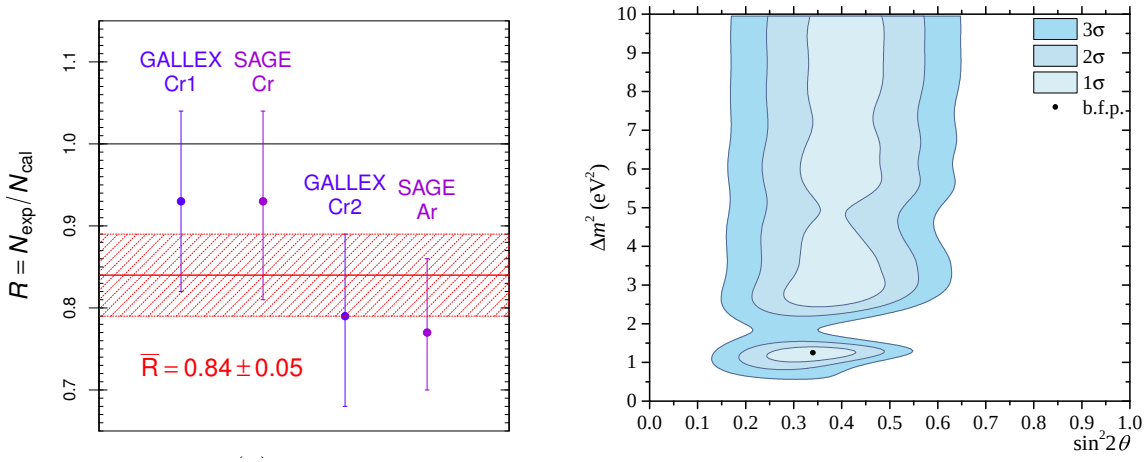


Figure 1.14: Left: Ratio R of the measured (N_{exp}) and calculated (N_{cal}) numbers of electron neutrino events in the GALEX and SAGE radioactive source experiments. The horizontal line and the band show the average ratio and its uncertainty [64]. Right: Exclusion plot for the BEST results combined with the SAGE and GALLEX results. The best fit is found for $\Delta m^2 = 1.25\ \mathrm{eV}^2$ and $\sin^2 2\theta = 0.34$. The figure is extracted from [63].

energetic ν_e emitters via electron capture:

$$\begin{aligned} {}^{51}Cr + e^- &\rightarrow {}^{51}V + \nu_e \\ {}^{37}Ar + e^- &\rightarrow {}^{37}Cl + \nu_e \end{aligned} \tag{1.30}$$

The comparison between the expected neutrino rate from the two ^{51}Cr calibration sources of the GALLEX detector [59] and the two ^{51}Cr, ^{37}Ar from the SAGE experiment [60] showed a deficit of $R = 0.84 \pm 0.05$ observed neutrinos as shown on figure 1.14 (left). This anomaly could indicate the oscillation of neutrinos toward a sterile state at short baseline. This oscillation would be driven by a squared mass $\Delta m^2 \gtrsim 0.35\ \mathrm{eV}^2$ and an amplitude $\sin^2 2\theta \gtrsim 0.07$ [61]. This nearly $3\,\sigma$ observation was reduced to $2.3\,\sigma$ by recent re-estimation if the cross-section of antineutrino interaction on gallium [62]. Recently, in 2021, the BEST experiment aimed to remeasure the gallium anomaly, with a similar process as the SAGE and GALLEX experiments. In [63], the experiment confirmed the gallium anomaly with a measurement 20-24% lower than the prediction. The exclusion contour for the joint analysis of the three gallium experiments can be seen on figure 1.14 (right).

1.5 The sterile neutrino hypothesis

The various neutrino oscillation anomalies (*gallium anomaly*, *accelerator anomaly* and *reactor antineutrino anomaly*), that have been observed by several experiments, could all be explained by an oscillation toward another light neutrino state. Because of the measurement of the width Z^0 boson decay, the number of light active (that can interact through weak interaction) neutrino states is constrained to three. This new neutrino would thus be sterile in the sense that it could not interact with matter via weak interaction and thus could not be detected via a direct interaction. This neutrino would however intervene in the oscillation process with the addition of a new mass eigenstate ν_4 to the formalism presented in section

1.3.2. This new eigenstate leads to three new mixing angles θ_{14}, θ_{24}, θ_{34}, and a new mass m_4. This four neutrinos formalism is usually referred as the 3+1 model. The mixing matrix would thus become :

$$\begin{bmatrix} U_{e1} & U_{e2} & U_{e3} & U_{e4} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} & U_{\mu 4} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} & U_{\tau 4} \\ U_{s1} & U_{s2} & U_{s3} & U_{s4} \end{bmatrix} \tag{1.31}$$

Given the current observations, this mass is expected to be at the eV scale to deal with the oscillation anomalies. The original best fit point of the RAA combining both reactor experiments and the gallium experiments was found at (Δm^2, $\sin^2 2\theta$) = (0.17,2.3 eV^2) [53]. Several types of experiments can study the sterile neutrino hypothesis:

Beta decay experiments. Experiments like KATRIN that aim to measure the neutrino masses via beta decays can test the existence of a sterile neutrino. KATRIN would be sensitive to the existence of a fourth neutrino with the study of spectral distortion in the beta decay spectrum that would be induced by the addition of a new mass term in the 3-flavour neutrino paradigm [65]. The experiment is more sensitive to the existence of a massive neutrino with a mass of a few keV. However, thanks to its large statistics, it could still be sensitive to ~ 1 % effects induced by a neutrino of a mass at the eV scale.

Appearance experiments. At a ratio $L/E \sim 1\ m/\mathrm{MeV}$ the appearance probability can be approximated by:

$$P(\overset{(-)}{\nu}_\alpha \to \overset{(-)}{\nu}_\beta) \simeq \sin^2(2\theta_{\alpha\beta}) \sin^2(\frac{\Delta m^2_{41} L}{4E}) \tag{1.32}$$

With $\sin^2(2\theta_{\alpha\beta}) = 4|U_{\alpha 4}|^2|U_{\beta 4}|^2$. The accelerator neutrino experiments such as MiniBooNE or LSND that study the oscillation $\overset{(-)}{\nu}_\mu \to \overset{(-)}{\nu}_e$ could thus be sensible to oscillations toward a sterile neutrino state.

Disappearance experiments. At a ratio $L/E \sim 1\ m/\mathrm{MeV}$ the disappearance probability can be approximated by:

$$P(\overset{(-)}{\nu}_\alpha \to \overset{(-)}{\nu}_\alpha) \simeq 1 - \sin^2(2\theta_{\alpha\alpha}) \sin^2(\frac{\Delta m^2_{41} L}{4E}) \tag{1.33}$$

With $\sin^2(2\theta_{\alpha\alpha}) = 4|U_{\alpha 4}|^2(1-|U_{\alpha 4}|^2)$. This type of measurement can be performed with muon neutrinos, with neutrino telescopes like IceCube or with accelerator experiments like Noνa or T2K. This measurement can also be performed with electron antineutrino, like the very short baseline nuclear antineutrino experiments.

1.5.1 Very short baseline reactor antineutrino experiments

During the past decade, various very short baseline ($\lesssim 10$ m) experiments were designed to probe the light sterile neutrino hypothesis. Those experiments aim to measure an antineutrino oscillation signal driven by a $\Delta m^2 \sim 1$ eV^2 and by an amplitude $\sin^2 2\theta \sim 10^{-1}$. Different ranges of baselines are studied and different types of reactors with various shapes, power and fuel are used as antineutrino sources. The figure 1.15 shows the variation of the parameter space covered by an experiment according to the

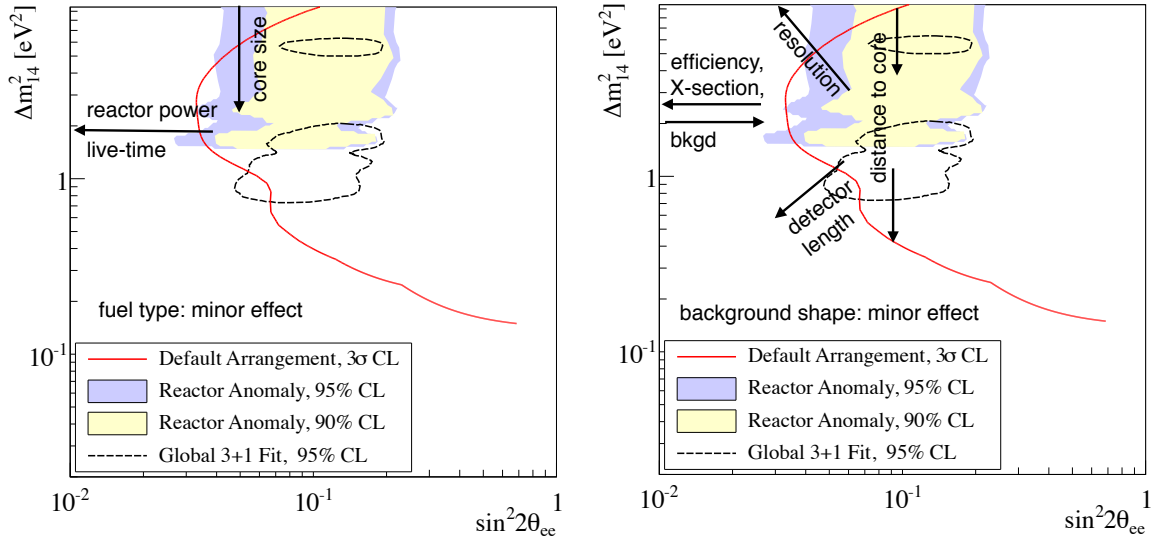


Figure 1.15: Reactor and detector parameters relevant for covering the parameter space [66].

variation of the reactor or the detector properties. On the figure, one can see that the range of amplitudes $\sin^2 2\theta$ that can be scanned depends on the signal to background ratio and the total statistics. A larger baseline will allow a wider range of Δm^2 to be tested. The closer the detector is from the reactor core, the higher Δm^2 values can be tested. A good energy resolution allows scanning larger values of Δm^2 to resolve the high-frequency oscillations.

The current very short baseline reactor antineutrino experiments will now be presented.

1.5.2 STEREO

The STEREO experiment [67] is based at the ILL reactor. The detector is composed of six target cells filled with a gadolinium-loaded liquid scintillator. It allows the measurement of antineutrinos at 6 different distances from the reactor core, from 9.4 to 11 m. The ILL research reactor has a compact core highly enriched in ^{235}U (HEU), with an operating power of 58 MW_{th}. A first analysis of the STEREO Phase-I+II data with 179 days of reactor ON and 235 days of reactor OFF data excludes at more than 99.9 % confidence level (C.L) the best fit point of the RAA [68]. This analysis provided an exclusion contour at 95 % C.L as can be seen on the figure 1.16.

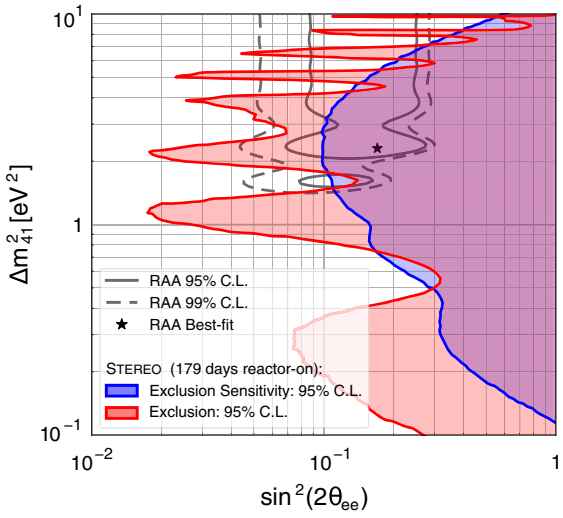


Figure 1.16: Exclusion contour (red) and exclusion sensitivity contour (blue) at 95% C.L. of Stereo phase-I+II. Overlaid are the allowed regions of the RAA (grey) and its best-fit point (star). The figure is extracted from [68].

1.5.3 PROSPECT

The PROSPECT experiment [69] operates a segmented detector based at the HFIR reactor in the USA. The detector contains $\sim$ 4 tons of liquid scintillator loaded with ^{6}Li segmented in 11×14 detection cells on a baseline from 7 to 9 m from the reactor core. The HFIR research reactor uses a HEU compact core, with an operating power of 85 MW_{th}. A first analysis of 96 days of reactor ON data disfavours the RAA best fit point at 99.4 % C.L, and shows a preference for an oscillation at $\Delta m^2 = 1.78\ \mathrm{eV}^2$ and $\sin^2 2\theta = 0.11$. However, this oscillation point is only slightly preferred with respect to the no oscillation hypothesis [51].

1.5.4 Neutrino-4

The Neutrino-4 experiment [70] is based at the SM-3 reactor in Russia. The detector is a $1.8\ \mathrm{m}^3$ gadolinium-loaded liquid scintillator detector, segmented in 10×5 sections. It operates on a movable platform and can detect antineutrinos at a distance from 6 to 12 m from the reactor core. The SM-3 reactor uses a HEU compact core, with an operating power of 90 MW_{th} thermal power. A first analysis of 480 days of reactor ON and 278 of reactor OFF showed an oscillation effect at 3.5 σ (99.98 %) C.L in the vicinity of $\Delta m^2 \simeq 7.26\ \mathrm{eV}^2$ and with $\sin^2 2\theta \simeq 0.38$. With the addition of the gallium anomaly effect, the obtained oscillation parameters become $\Delta m^2 = 7.3 \pm 1.17\ \mathrm{eV}^2$ and $\sin^2 2\theta = 0.36 \pm 0.12$ with a 2.7 σ confidence level. Using a Monte-Carlo Neutrino-4 like dataset, Giunti et al showed however that the resolution effects of the detector might not be taken into account to arrive at such results [71]. Due to the higher value of the best fit Δm^2, this parameter space will be difficult to study by other very short baseline experiments as it requires both an excellent energy resolution and measurements at a distance closer to the core than the current experiments. However, this effect could be verified by a measurement via beta decay as the KATRIN experiment.

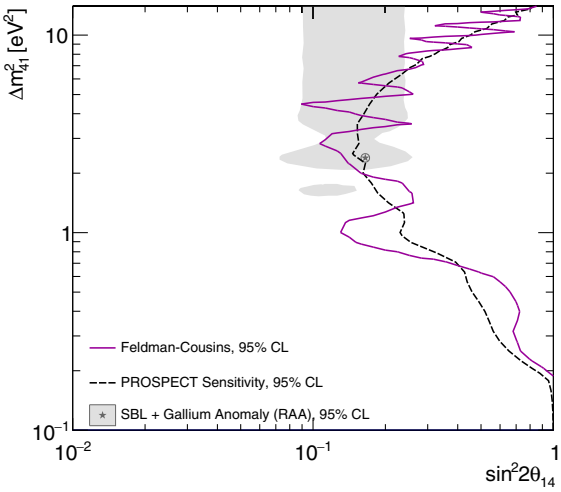


Figure 1.17: PROSPECT sensitivity and 95% confidence level sterile neutrino oscillation exclusion contour from 33 days of reactor ON data. The RAA best fit is excluded at a 2.2σ confidence level. The figure is extracted from [51].

1.5.5 DANSS

The DANSS experiment [72] is located at the Kalinin nuclear plant in Russia. The detector is highly segmented with 2500 plastic scintillator strips of $1\times4\times100$ cm^3 with gadolinium-loaded reflective covers. The detector is placed on a platform under an industrial reactor with a thermal power of 3.1 GW$_{th}$. The detector distance to the reactor core can be varied from 10.7 to 12.7 m with a lifting system that can modify the distance between the center of the detector and the reactor core. Current results from DANSS exclude the best fit point of the RAA with more than a 5σ significance. The best oscillation point from this analysis is $\Delta m^2 = 1.3$ eV^2 and $\sin^2 2\theta = 0.02$, with a significance of 1.5 σ [73].

1.5.6 NEOS

The NEOS experiment [74] operated at the Hanbit nuclear power complex in South Korea. The detector used approximatively 1 ton of gadolinium-loaded liquid scintillator, located at 23.7 m from the reactor core. The Hanbit reactor is an industrial reactor with a thremal power of 2.8 GW$_{th}$. As it was not a segmented detector, the NEOS detector could only measure the antineutrinos at one precise place. The published analysis from NEOS uses the Daya-Bay spectrum as a reference spectrum for the sterile neutrino research. The experiment results from 2017 rejects the RAA best fit point at a 90 % C.L. with no strong evidence for a 3+1 neutrino oscillation [74]. A recent combined analysis with RENO was performed using the 2509 days of RENO near detector data, placed at 294 m from the same nuclear reactor. It concludes with an exclusion at 95 % confidence level for the region of $0.1 < \Delta m^2 < 7$ eV^2 and an allowed region with a best fit at $\Delta m^2 = 2.41 \pm 0.03$ eV^2 and $\sin^2 2\theta = 0.08 \pm 0.03$ in a 68 % C.L allowed region [75].

1.6 Conclusion

In this chapter, the history of the neutrino discovery with the different discoveries through the XXth century were presented. Those discoveries lead to the observation of neutrino oscillation by the Super-Kamiokande and SNO experiments. These oscillations are not described by the Standard Model and could lead to the discovery of physics outside the Standard Model, with various oscillation anomalies. One explanation to the oscillation anomalies could be the neutrino oscillations toward a light sterile state. The sterile neutrino hypothesis can be studied by various type of experiments, however very short baseline reactor antineutrino experiment would be the best candidates to probe the hypothesis. Among those experiments, the SoLid experiment, with its hybrid scintillation technology, aims to provide additional inputs to the sterile neutrino oscillations and to the ^{235}U reactor antineutrino measurement.

Chapter 2

The SoLid experiment

The different oscillation anomalies have been presented in the chapter 1. The development of very short baseline reactor antineutrino experiments could provide the necessary results to study the sterile neutrino hypothesis. During the past decades, detectors were built all around the world at a few meters from nuclear reactors. Among those experiment, the SoLid experiment operates at the BR2 research reactor in Belgium with as main goal the search of the oscillation of electron antineutrino to a sterile state. Thanks to the highly enriched in ^{235}U composition of the reactor core, the experiment also aims to produce a precise measurement of the ^{235}U induced antineutrino spectrum. This chapter will present in detail the SoLid experiment with the BR2 reactor, the SoLid detector and the expected signals and backgrounds.

2.1 The BR2 nuclear reactor

The Belgian Reactor 2 (BR2) is a nuclear reactor located at the SCK CEN research center in Mol. Its main purpose is the testing of materials for nuclear reactors, and it is also involved in medical isotope production and research on doped silicon. The reactor is an open pool type reactor, immersed into water that acts as neutron moderator.

Thanks to its unique twisted design with inclined channels, the reactor core can be very compact ($\sim$0.5 - 1 m diameter) as it can be seen on figure 2.1. It uses highly enriched $^{235}\mathrm{U}$ fuel (93.5%) with a thermal power between 40 and 100 $\mathrm{MW_{th}}$ and a high neutron flux, up to 10^{15} $\mathrm{n/cm^2/s}$ for a rate of emitted electron antineutrinos up to about $2 \cdot 10^{19}$ $\overline{\nu}_\mathrm{e}/\mathrm{s}$. The ^{235}U is responsible for 99% of the fissions.

The reactor undergoes reactor ON and OFF cycles and is ON approximatively 150 days per year with cycles of about one month. The reactor OFF periods are used in the SoLid experiment for background measurements.

2.1.1 Antineutrino flux prediction

The different methods available to compute the antineutrino flux prediction were presented in the section 1.4.1. In SoLid, the prediction of the antineutrino flux is based on a detailed 3D model of the BR2 core, coupled to a Monte-Carlo based method, MCNPX/CINDER90, that produces the fission rates [76]. The MURE code then used to track the burn-up of the fissile products in the reactor core [77]. The energy spectrum is then computed with the Huber-Mueller model with off-equilibrium correction provided by

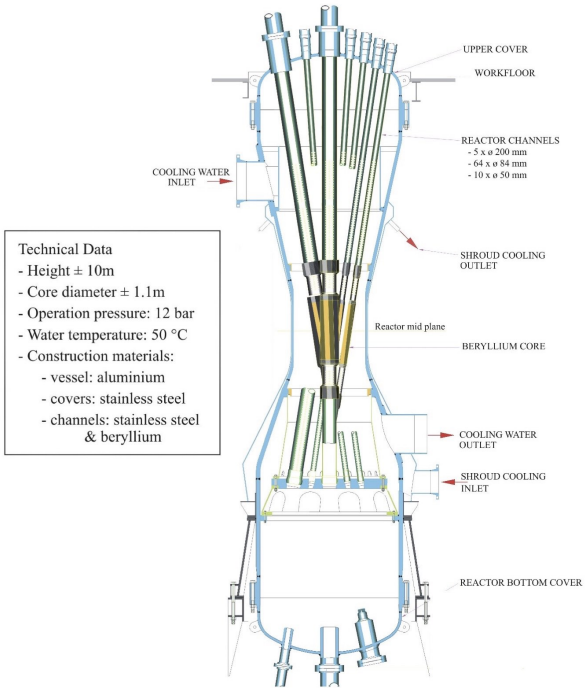


Figure 2.1: Design and technical data of the BR2 reactor core. The special twisted design with the inclined channel can be seen. The core is composed of hexagonal channels containing the nuclear fuel, the control rods and experimental channels all inside a beryllium matrix. In blue is represented the aluminium vessel that is completely under water.

the MURE code. On the figure 2.2 (left) is represented the evolution of fission rates for the four main fissile isotopes in the BR2 core (^{235}U, ^{238}U, ^{239}Pu and ^{241}Pu) for a given cycle. One can see that the ^{235}U contributes for more than 99% of the fissions in the reactor core. On the figure 2.2 (right) is represented the variations of the emitted antineutrino spectrum computed for 28 days of fuel burn-up. The time steps were chosen to taken into account the power variation of the reactor and the needs of the reactor calculation. One can see a clear shape variation as a function of the fuel burn-up.

2.1.2 SoLid detector location

The sensitive volume of the SoLid detector covers a baseline from 6.3 m to 8.9 m from the core centre, see figure 2.3. The detector baseline is aligned with the mean antineutrino flux direction. The detector is located on the 3rd floor of the containment building. There is no nearby experiment, and all the beam ports are shielded with 20 cm of lead. It insures that the detector faces only a low level of reactor induced backgrounds. However, due to its location in a reactor building, it has very low protection against atmospheric background with 8 meters-water-equivalent above it. A passive shielding of 50 cm of polyethylene is placed on the top of the detector to mitigate fast neutrons coming from the atmospheric showers. The detector is then surrounded by a 50 cm thick water wall. Then 2 mm thick cadmium sheets are placed between the shielding and the detector to capture the slow neutrons and 10 cm of polyethylene are placed under the detector.

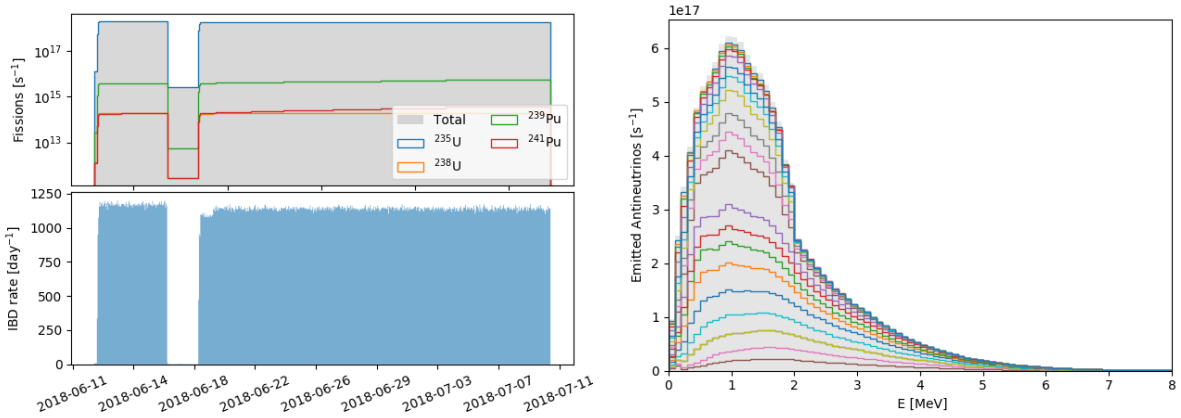


Figure 2.2: Left: Evolution of fission rates for the different fissile isotopes present in the reactor core for a given cycle. The ^{235}U contributes for more than 99% of the fissions. Right: Emitted antineutrino spectrum using the summation method for 28 days of irradiation. The time steps for the reactor evolution were chosen to take into account the power variation, but also the needs of the reactor calculation. Each of them is represented by a colour.

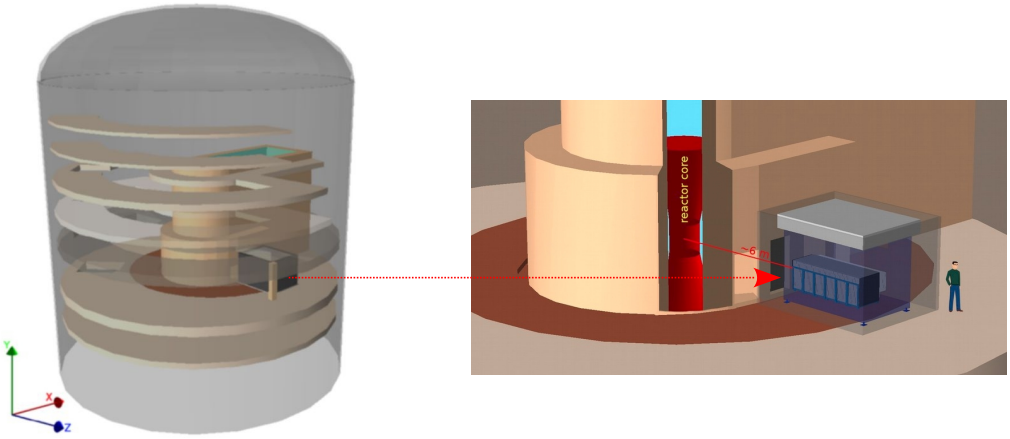


Figure 2.3: Left: GEANT4 Model of the BR2 reactor building with SoLid detector inside. Right: SoLid position in front of the reactor core. The detector is inside a container from 6.3 to 8.9 m of the reactor core.

2.2 Detection principle

2.2.1 Events of interest

SoLid experiment was designed for two main goals: the test of the oscillation toward a light sterile neutrino state and a precise measurement of ^{235}U induced antineutrino energy spectrum. The key interaction

used in reactor neutrino oscillation is the inverse beta decay:

$$\overline{\nu}_e + p \rightarrow e^+ + n \tag{2.1}$$

The energy measured is the positron kinetic energy. To recover the antineutrino energy, one can do a simple mass conservation equation, neglecting the neutron (and proton) kinetic energy with regards to the positron one:

$$E_{\overline{\nu}_e} = E_{e^+} + m_n - m_p = T_{e^+} + m_{e^+} + m_n - m_p \tag{2.2}$$

The target material is a plastic scintillator polyvinyl toluene (PVT) in the form of cubes. After depositing all its energy T_{e^+} in a cube, the positron will annihilate with an electron of the medium and two 511 keV annihilation gammas will be emitted. Those gammas won't necessarily be detected in the same cube. In that case, zero, one or two gammas can be detected depending on how many gammas escape the detector with too low-energy deposits.

The detection method selected in the SoLid experiment is based on a double scintillation strategy, as represented on figure 2.4 (left). The organic PVT is rich in protons and is thus a good antineutrino target. It has a good light yield with in average 10 000 photons per MeV deposit and with a fast scintillating decay time of a few ns. The positron will deposit its energy in the interaction cube. The positron energy is then derived from the PVT scintillation. The emitted neutron will thermalize via scattering with the protons of the PVT cubes and be captured by inorganic ^{6}LiF:ZnS(Ag) screens on the side of the cubes via the break-up reaction:

$$n + {}^6\mathrm{Li} \rightarrow \alpha + {}^3\mathrm{H} \tag{2.3}$$

The α and ^{3}H induce then the ZnS scintillation with a slow decay time of a few microseconds. Due to the neutron thermalization time, a few tens of microseconds, the signal of interest is a delayed time coincidence between the prompt PVT and ZnS scintillation. The two signals are identified with their different decay time and pulse shape feature: one high amplitude peak for the PVT scintillation and several low amplitude peaks for the ZnS, as represented on figure 2.4 (right).

2.2.2 Detection Cells

The detection cell in SoLid is composed of a $5 \times 5 \times 5$ cm^3 PVT cube with a ^{6}LiF:ZnS(Ag) screen on two sides. Each cube is optically isolated with Tyvek wrapping, see figure 2.5 (left). To retrieve the scintillation light, four wavelength shifting fibres are connected to the cube through groves of 5×5 cm^2 on four different faces of the cube. Each fibre guides the light produced by the scintillation of a cube or a neutron screens to a reflective aluminized mylar mirror at one end and to a Multi-Pixel Photon Counter (MPPC)[1] at the other end, see figure 2.5 (right). The use of mirrors allows the optimization of the light read out by each MPPC [78]. The connection between a fibre and the MPPC/mirror is done with optical grease. A cube is thus connected to two vertical and two horizontal fibres on a plane orthogonal to the detector baseline. To cope with fibre attenuation effects, the MPPCs position is alternated from one side to the other side of the detector.

[1] Also called silicon photomultipliers (SiPM)

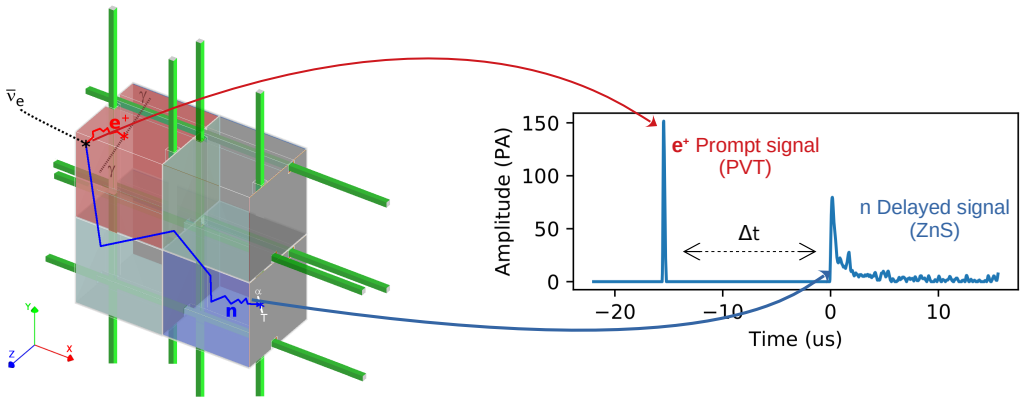


Figure 2.4: Left: Representation of the $\overline{\nu}_e$ interaction inside SoLid detector. The positron is detected with PVT scintillation. The neutron is captured in a ^{6}LiF:ZnS(Ag) screen that induce the breakup of the ^{6}Li and the ZnS scintillation. Right: scintillation pulses in SoLid with a prompt positron narrow pulse and a delayed neutron signal with a slowly decaying pulse.

2.2.2.1 MPPCs

The MPPCs used in the SoLid experiment are Hamamatsu type S12572-050P [79]. They are composed of a matrix of 50×50 μm^2 semiconductors (pixels) on 3×3 mm^2. Each semiconductor uses a P-N junction. The MPPCs operate in Geiger mode, hence the electrons created by the interaction of a scintillation photon in the junction will induce an electronic avalanche, or pixel avalanche (PA). The signal obtained will then be the sum of all pixels that had an avalanche in a MPPC.

To obtain an avalanche, a sufficient voltage needs to be applied to the P-N junction. This is called the *breakdown* voltage V_{br}. For the type of MPPC used in SoLid, the value given by the manufacturer is 65 ± 10 V, but precise measurement have been performed for every MPPC of the detector [80]. The *operating* voltage V_{op} is the actual voltage used. It is higher than the *breakdown* voltage in order to increase the detection efficiency and the signal gain. The difference between V_{op} and V_{br} is called the overvoltage (V_{ov}). The larger is the overvoltage, the larger is the gain and the photon detection efficiency, as shown on figure 2.6.

The value set as overvoltage is limited by two effects. The first one is the dark count rate. This comes from self triggering pixels induced by thermal noise. The larger is the overvoltage, the larger will be the dark noise. The second one comes from the possible emission of ultraviolet photons by a pixel during an avalanche. The larger is the overvoltage, the larger will be the probability of those photons to trigger another pixel of the MPPC. This effect is called the optical cross-talk. Those effects will impact the energy resolution of the detector and can induce fake signals and bias the detection efficiency. The evolution of the dark count rate and the cross-talk probability can be seen on figure 2.7. To diminish the dark count rate, the MPPCs are kept at 11 °C via the chilling of the detector, as presented in section 2.3.2. The working point chosen in SoLid is an overvoltage of 1.8 V with an expected photon detection efficiency

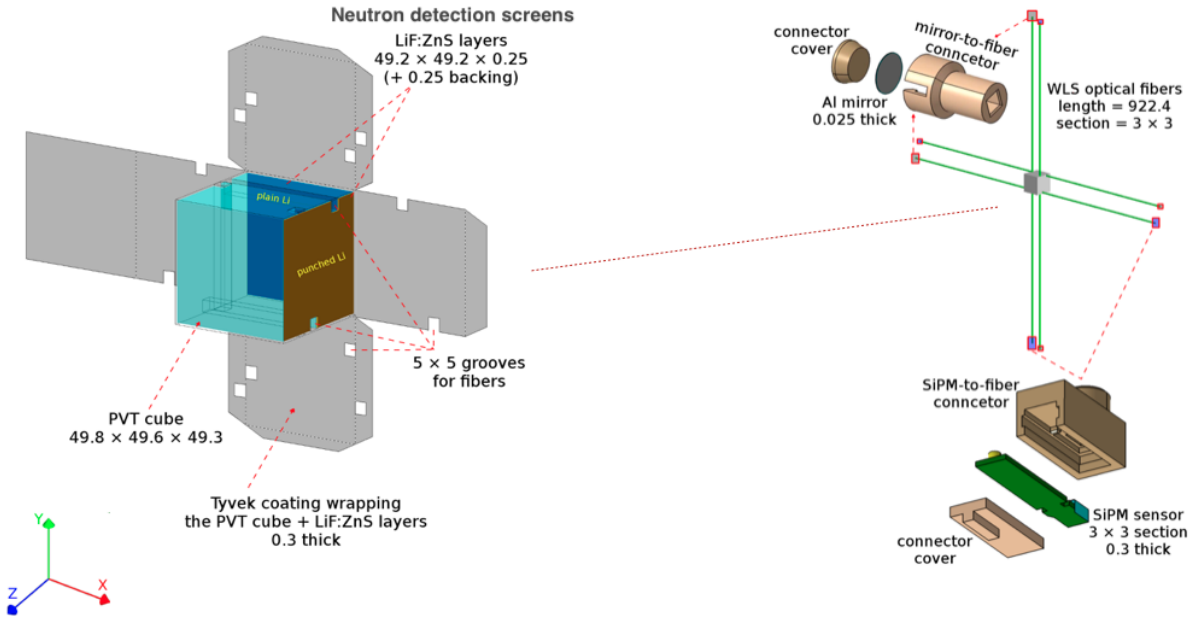


Figure 2.5: Left: A schematic view of the PVT detection cell, with the neutron detection screens and its Tyvek wrapping. Right: Four wavelength shifting fibres cross each detection cell, with alternating positions of MPPCs and mirrors at the fibre ends. They are contained in plastic printed connectors. Dimensions are in mm.

of 32%, a gain of 31.5 ADC counts per PA, a cross talk probability of 20% and a mean dark count rate of 110 kHz per channel. The gain that is set here at a value of 31.5 ADC counts per PA is not directly the gain measured on figure 2.6 but a combination of several amplifiers in the readout chain. However, by convention, it will be called the MPPC gain as this value will always depend on the electronics used to readout the signals.

2.2.2.2 Wavelength shifting fibres

The fibres used in SoLid are Saint-Gobain's BCF-91A double cladded wavelength shifting fibres [81]. They have a squared section of 3×3 mm^2 and are 92 cm long. The fibres are used to guide the light from the scintillators to the MPPCs with exactly the same contact surface. Those specific fibres were chosen to match the emission spectrum of the scintillators and maximize the MPPCs detection efficiency. The fibres have a maximum absorption peak in the blue around 420 nm that matches the 425 nm emission of the PVT cubes [82], another absorption peak at 450 nm that matches the ZnS(Ag) emission peak [83] and a emission peak around 494 nm for a maximum detection efficiency of the MPPCs at 500 nm[79]. The fibre attenuation lengths are measured individually in the calibration procedures in section 3.3.2.

2.2.3 Detector design

The SoLid detector has been designed as a highly segmented detector. It is organized in 50 planes of 16x16 detection cells each. Planes are then regrouped by ten to form a module, for a total of 5 modules. Each plane is surrounded with high-density polyethylene (HDPE) to reflect neutrons that would escape

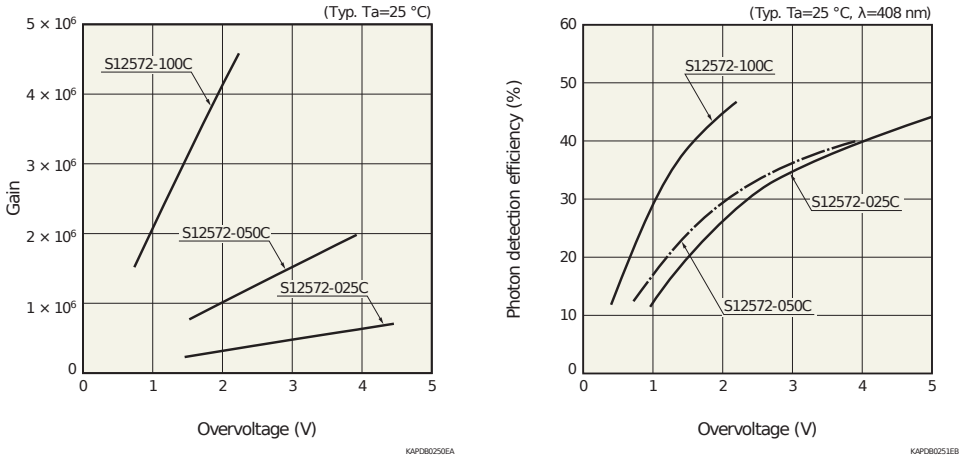


Figure 2.6: Gain and photon detection efficiency as a function of the overvoltage for several types of MPPC from Hamamatsu data sheets [79]. The type used for the SoLid detector have the same characteristics as the S12572-050C. The larger is the overvoltage, the larger are the two parameters. These measurements were made at 25 °C with photons with a wavelength of 408 nm.

the detector when produced at the border. The whole plane is supported by a hollow aluminium frame that contains the MPPCs and their cabling and that acts as a Faraday cage for the sensors. The front-end electronics of the plane, detailed in section 2.4.3.1, are kept in aluminium enclosures next to the frame. Finally, two Tyvek sheets are placed between each plane to ensure a better optical isolation between the planes, see figure 2.8.

2.3 Detector construction and integration

2.3.1 Detector construction and quality assurance

SoLid construction started in 2016. Each of the 12800 PVT cubes has been individually machined, visually inspected and precisely weighted with a 1 mg precision. All the neutron screens have also being weighted. After being wrapped in Tyvek and equipped with the ^{6}LiF:ZnS(Ag) screens, all the cubes were tagged with a QR code and their specificities were entered in a database. The weights before and after the wrapping were included in the database. It is a very important step to know precisely the number of proton targets in each cube for antineutrino interactions. In the detector the mean weight of the cubes is equal to 119.7 g with a RMS of 0.1 g ensuring a per mille control of the number of targets. The cubes were then integrated inside the planes for a first qualification, with their position tracked with the QR code.

The quality assurance was performed with the so-called Calipso test bench [80]. This consisted in a robot that could place radioactive sources in front of each cube of a plane. This quality assurance phase allowed the detection of construction failure such as missing screens, bad fibre connection, MPPCs malfunctioning or wrong cabling. Those issues could be corrected before the integration of a plane inside the detector. A ^{22}Na source was used for a preliminary energy calibration of the detector. AmBe and ^{252}Cf were used for first neutron calibration to determine a preliminary neutron reconstruction efficiency throughout the detector.

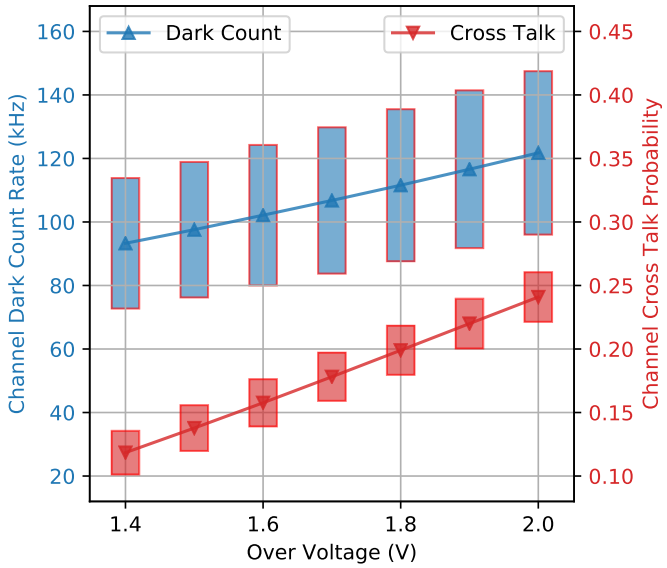


Figure 2.7: Cross-talk probability and dark count rate as a function of the overvoltage at 11°C. The points show the channel mean, and the boxes show the standard deviation across all channels. For the detector operations at 1.8 V overvoltage, the mean dark count rate is at 110kHz and the cross talk probability is at 20%.

2.3.2 Detector container

The detector is placed into a cooled container at BR2, as shown in figure 2.9. It allows having a control on the operation temperature with a chiller and on the humidity inside the detector with an air flushing system. It replaces the humid air of the container with dry air produced by the factory in the reactor building. This also helps to remove radon inside the container that can be a source of background, as presented in section 2.5.2. The rails of the calibration robot (detailed in section 2.4.2) are also fixed on the top of the container and each module is fixed on rails on the ground, this allows the creation of small gaps between the modules for calibration purpose with an automated procedure. Finally, the container acts as a black box for the MPPCs with a total light isolation.

To monitor the environmental conditions such as the pressure, the humidity inside the container, the temperature at different points inside the detector, several sensors have been placed inside the detector and are readout with a Raspberry-Pi device. A PMT coupled to a NaI scintillator, placed inside the container, is also used to monitor the gamma background and a radon detector is used to measure the airborne radon concentration.

2.4 Detector operations

2.4.1 Gain equalization

In practice, the gain of each MPPC is measured with the energy distribution of dark counts with a periodic trigger that allows the recording of unbiased low amplitude signals (see section 2.4.3.2). The energy spectrum is obtained after digitization of the PA distributions by an analog-to-digital converter

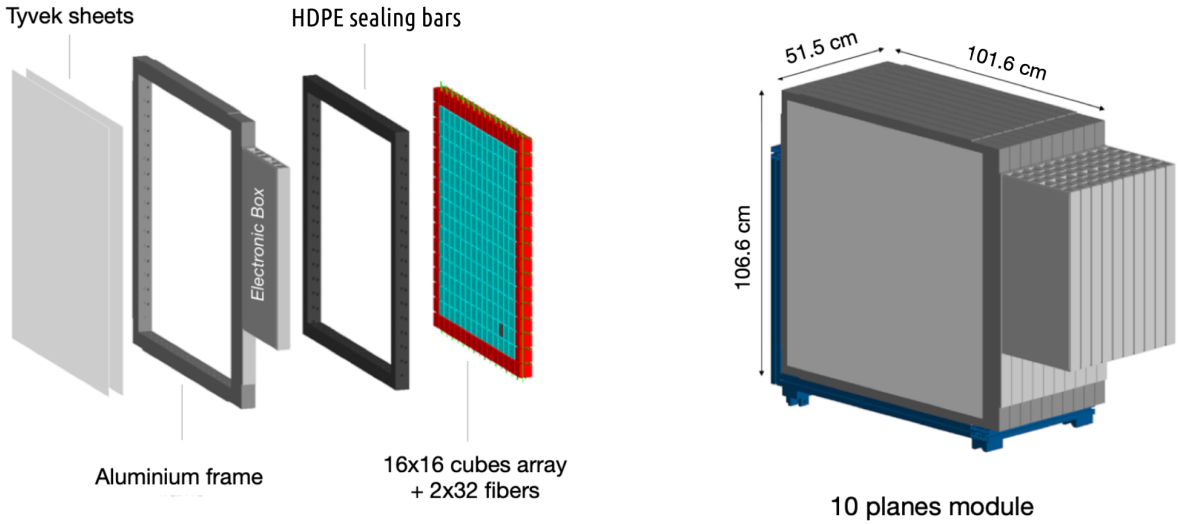


Figure 2.8: Left: Exploded view of a frame. The 16x16 detection cells are represented in blue and red. The HDPE bars for neutron reflection are in black. The aluminium frame holding the plane and the electronic box and the two Tyvek sheets are represented in grey. Right: The association of 10 planes to make a module on its trolley.

(ADC), see section 2.4.3.1, and is expressed in ADC counts. As it is represented on figure 2.10 (left), the ADC spectrum is composed of peaks, particularly when a local maximum filter is applied. Each peak represents a given number of PA. The first one at zero is the pedestal peak. This peak corresponds to the electronics measurement when no input signal is read out. It can be different for each MPPC and thus needs to be measured and subtracted from any measurement. In this case, it is null because of the pedestal subtraction for this measurement. Then the gain can be seen as the gap between two peaks or, as it is showed here, the first non-zero peak position when the pedestal is subtracted. This gain is measured for every channel and an overvoltage scan is performed to set all channels at the same gain by setting individual V_{op} to each of the 3200 MPPCs. The spread observed after this gain equalization procedure is around 1.4%, see figure 2.10 (right).

2.4.2 CROSS in situ calibration system

A calibration robot has been designed in order to perform in-situ calibration of the detector. It consists in a mechanical arm on which a source holder is attached where radioactive sources can be put. The arm is attached to rails that are above the detector. Those rails allow the arm to go in front of the first plane, behind the last one and between each module in gaps momentarily created for the calibration. In each gap, the arm can put the source in 9 different positions to ensure a good coverage of the detector, see figure 2.11.

For the energy calibration, as the calibration sources can only be placed between each module, penetrating sources are needed due to the depth of five planes that must be calibrated with a given source position. For that reason gamma sources were chosen. Due to the size of the PVT cubes and the energies of the available gammas of a few hundreds of keV to a few MeV, only the Compton edges can be used for energy calibration. The Compton edges fitting methods developed by the collaboration and the de-

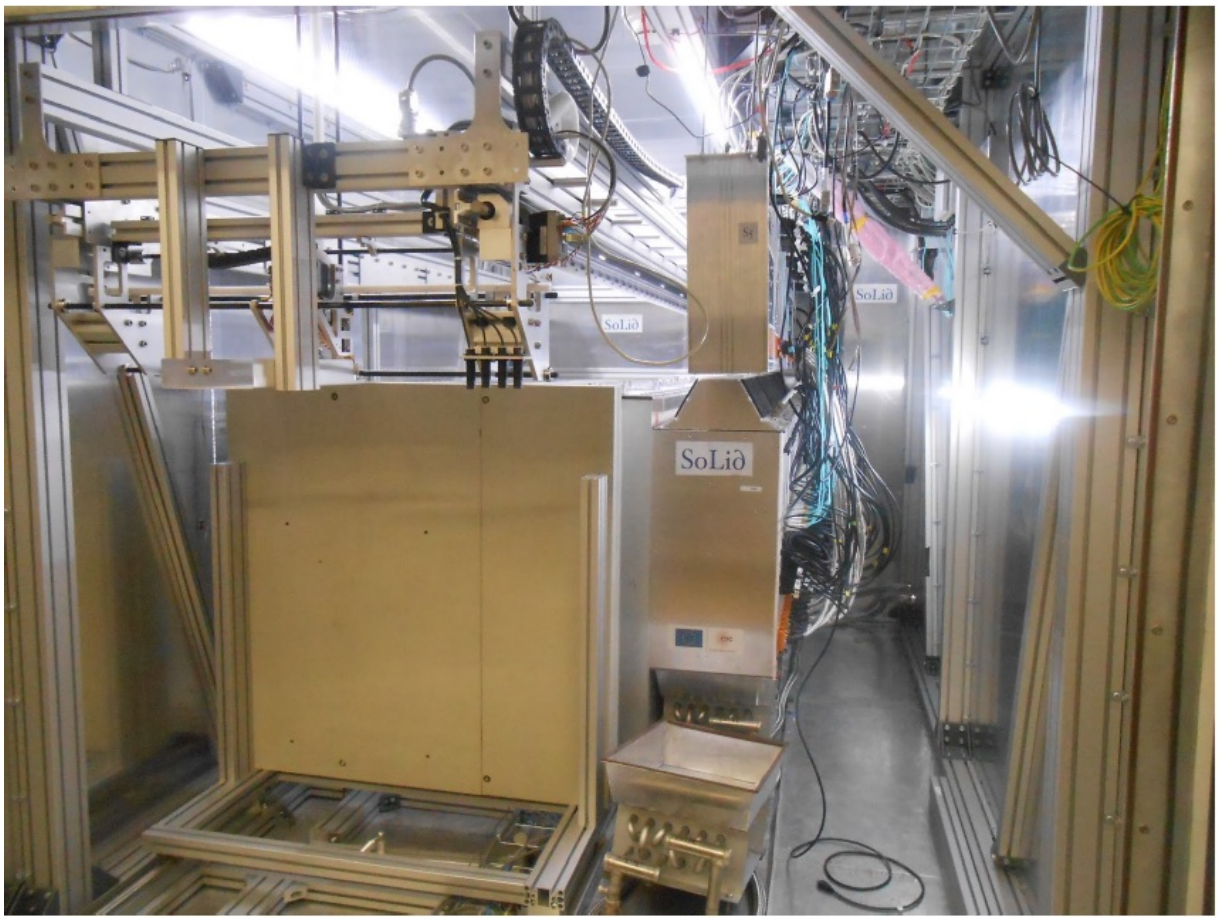


Figure 2.9: Photograph of the detector in its container. On the top of the detector can be seen the rails of the calibration robot. On the bottom the rails on the grounds for calibration purpose. On the right of the detector can be seen all the electronic boxes.

tails on the energy calibration will be presented in the chapter 3. A detail of the available sources, their activities and Compton edge energies can be seen in table 2.1.

For an IBD analysis, the antineutrino detection efficiency depends directly on the neutron detection efficiency of the detector. Neutron calibrations have been performed with AmBe and ^{252}Cf neutron sources. The aim of the neutron calibration is to measure the neutron reconstruction efficiency, to be multiplied by the neutron capture efficiency computed with Geant4 simulations to obtain the neutron detection efficiency. More detail on those calibrations can be seen in Valentin Pestel's thesis [84]. Both efficiencies have been evaluated for each detection cells for the two different calibration sources. The neutron capture efficiency is represented in the figure 2.12. Its average value on the detector is 71%. The reconstruction efficiency is represented on figure 2.13. The difference between the two calibration sources are due to uncertainties in the AmBe source activity. The average detection reconstruction efficiency is 74%. The averaged neutron detection efficiency in the detector is then 52%.

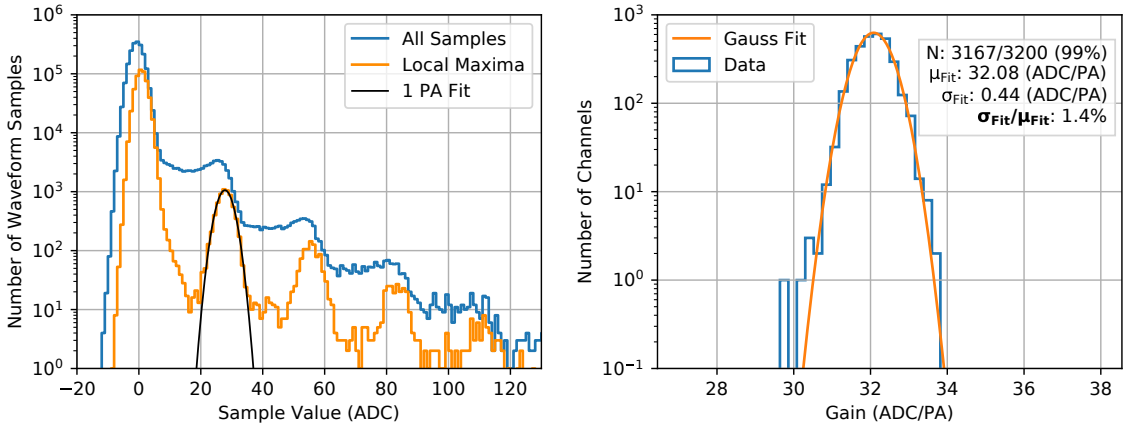


Figure 2.10: Left: Spectrum of ADC samples of one channel of the detector with pedestal subtracted. The gain is then defined as the mean of the Gaussian fit of the first PA peak (around 32 ADC) represented in black. Right: Spread of the gain values for all operational MPPCs after a voltage scan equalization.

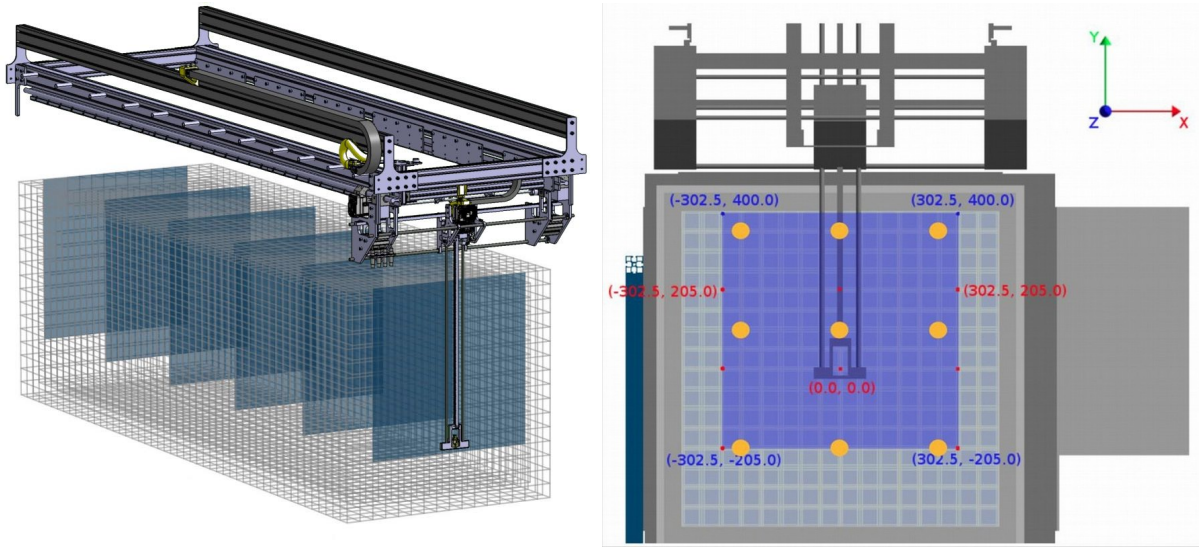


Figure 2.11: Left: CROSS system with a rail on the top of the detector. The source holder can go in the six gaps represented by blue areas. Right: the 9 sources positions available inside a Gap.

2.4.3 Data acquisition

2.4.3.1 Readout system

Each detection plane has its own readout system inside the electronic box of the aluminium frame. It contains the front end electronics, composed of two 32-channel analog boards that provide the power supply of the MPPCs and set the overvoltages. Those boards also shape and amplify the MPPC pulses to prepare the digitization. This operation is performed by a digital board connected to the two analog boards. The signal is digitized with a sampling frequency of 40 MHz. An FPGA based device is then used for data reduction with a Zero-Suppression (ZS) algorithm and plane triggering. Once a trigger signal is

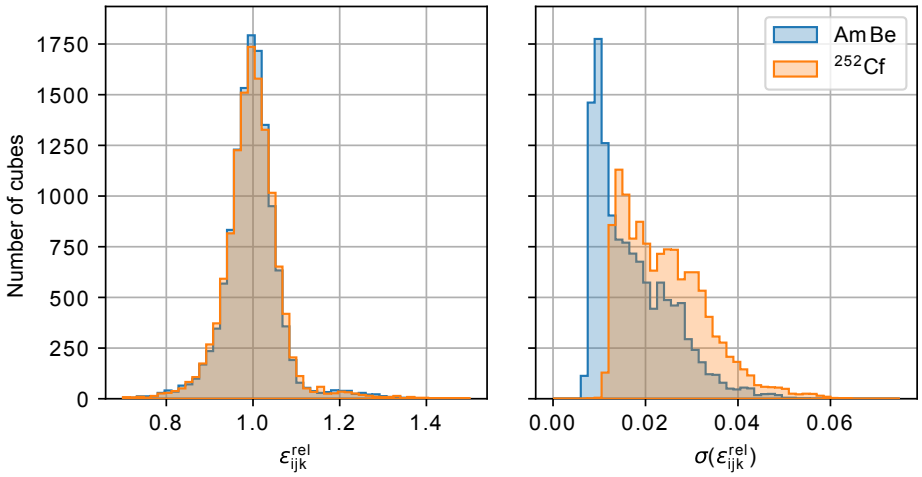


Figure 2.12: (Left) Neutron capture efficiency for the 12800 cells obtained with the AmBe (blue) and ^{252}Cf source (orange). (Right) The total uncertainty on the neutron capture efficiency for all detection cells obtained with sources [84].

emitted by a plane, it is propagated to all other detector planes for a simultaneous readout, depending on the triggering strategy. Those strategies will be described in the next section. All the planes of a module (group of ten planes) are synchronized with a shared clock, and a master clock is used to synchronize all the modules.

Data recorded in each plane are stored into a buffer and gathered into blocks of 6.4 µs. Depending on the triggering decision, a block will be transmitted further or discarded. Keeping the blocks in memory even without a trigger allows detecting a coincidence between triggering neutron signals and electromagnetic signals that happened previously in the detector. The data transmitted by the buffer are then zero-suppressed for data reduction.

2.4.3.2 Trigger strategies

In regular data taking, the following triggers are used simultaneously:

The *neutron trigger*. Due to the high rate of electromagnetic signals (ES) in the detector for antineutrino physics data taking, it is not possible to trigger on all signals. However, removing low-energy deposit would prevent the detection of low energy annihilation gammas. The decision was made to trigger on neutron signals (NS). Once a NS triggers a plane, a space and time region is readout around the signal in order to recover all the events linked to the possible IBD interaction. To do so, a NS detection algorithm has been developed. This algorithm is based on a peak counting algorithm: the plane will trigger if a given number of peaks are above a given threshold in a given time window. In regular data taking, the trigger is set to 17 peaks over 0.6 PA in a 6.4 µs time block [86]. An illustration of the neutron trigger on a waveform can be seen on figure 2.14. Once a plane triggers, ± 3 planes around are readout to collect the signals linked to the positron and the annihilation gammas.

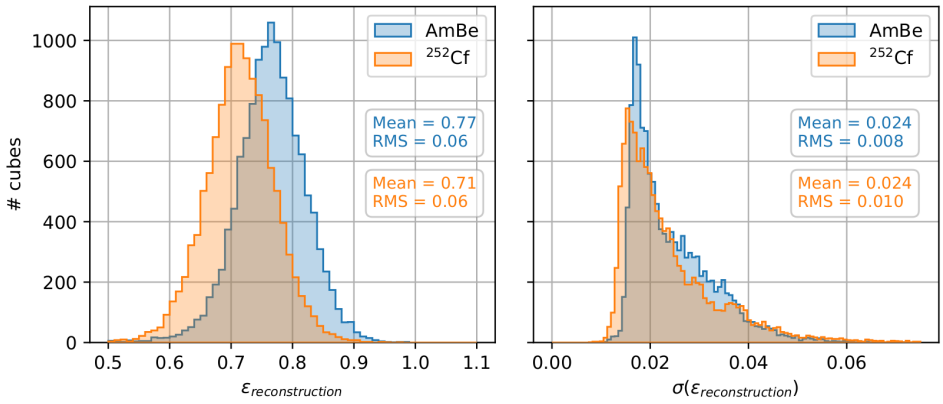


Figure 2.13: (Left) Neutron reconstruction efficiency for the 12800 cells obtained with the AmBe (blue) and ^{252}Cf source (orange). (Right) The total uncertainty on the neutron reconstruction efficiency for all detection cells obtained with the two sources [85].

Source	Activity [kBq]	γ energy [keV]	Compton edge energy [keV]
^{207}Bi	37	569	393
		1063	857
		1770	1547
^{22}Na	37	511	341
		1275	1062
AmBe	1	4438	4196

Table 2.1: List of the different gamma sources used for energy calibration with their activities, their gamma energies and the associated Compton edge energy.

The *threshold trigger*. When two orthogonal channels of a cube are reading a signal above a given threshold, the plane triggers and the whole plane is readout for two time blocks (the triggering block and the next one). This trigger focuses on high-energy ES events such as muon tracks, and is also used for regular calibration campaigns.

The *periodic trigger*. The whole detector is readout in a two time blocks window at a given frequency. This trigger is mostly used for MPPC monitoring and calibrations at low-energy to avoid any bias from an energy threshold.

2.5 Expected signal and backgrounds

2.5.1 Expected antineutrino signal

With a 3D model of the reactor core, the average number of antineutrino interactions in a given cube is computed with the formula:

$$N = \int_R \frac{1}{4\pi L_i^2} n_f N_p \sigma_f dx^3 \tag{2.4}$$

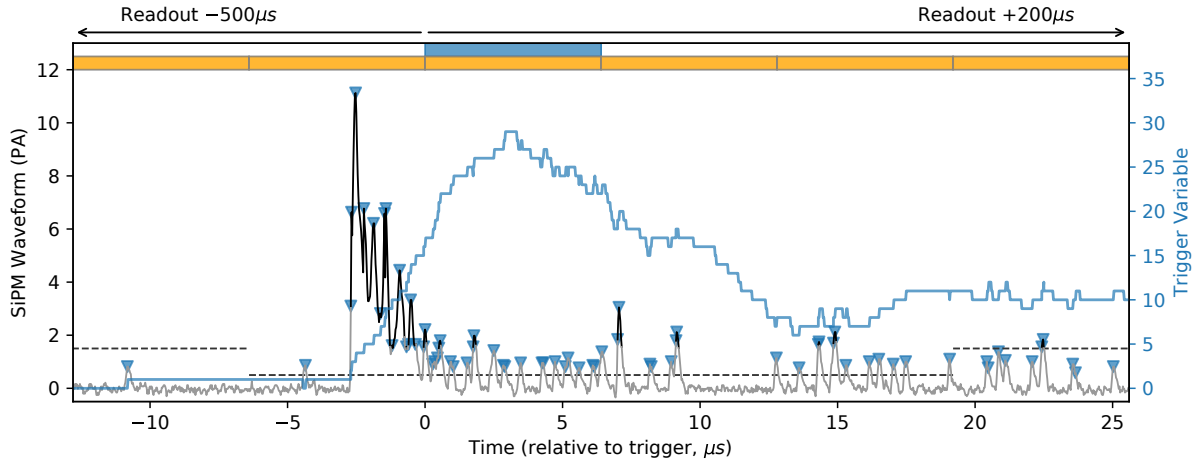


Figure 2.14: Illustration of the neutron triggering algorithm. The neutron waveform is in grey. The zero suppression threshold is at 1.5 PA for regular data taking and lowered to 0.5 PA for a neutron signal. The blue histogram represents the number of peak over threshold in the rolling time window. The blue triangles represent the peaks over threshold. The samples above 1.5 PA are represented in black. The rectangles on the top represent the data blocks. The blue one is the block that triggers the plane, and then the blocks in $500\mu s$ before and $200\mu s$ after the signal will be stored.

with an integral on the whole reactor core volume R, L_i the distance between the integration point (x, y, z) in the core and the centre of the cube, n_f the number of fissions in the point (x, y, z), N_p the number of proton target in the cube and σ_f the cross-section per fission for all the fissile isotopes. This last parameter can be derived from the antineutrino spectrum S_ν and the IBD cross-section σ_{IBD}:

$$\sigma_f = \int S_\nu(E_\nu)\sigma_{IBD}(E_\nu)dE_\nu \tag{2.5}$$

The antineutrino spectrum can be computed as presented in section 2.1.1. Following Vogel and Beacom [87] form, the IBD cross-section can be written as:

$$\sigma_{IBD}(E_\nu) = \frac{2\pi^2}{m_e^5 f^R \tau_n} E_{e^+}\sqrt{E_{e^+}^2 - m_e^2} = K E_{e^+}\sqrt{E_{e^+}^2 - m_e^2} \tag{2.6}$$

with f^R a phase space factor and τ_n the neutron life-time. With recent measurements, the first fraction can be measured [88]: $K = 0.961 \times 10^{-43}\text{cm}^2\text{MeV}^{-2}$. With a geometrical acceptance of the detector (number of neutrinos that crosses the detector over the number of neutrino emitted) of about 0.11 %, the expected yield of detectable neutrino per day is around 1200 [85]. By combining the antineutrino energy spectrum and the IBD cross-section, the expected energy spectrum in the detector can be created as on figure 2.15.

2.5.2 Backgrounds

Among the backgrounds faced in SoLid, two main categories can be defined: the accidental background and the correlated background. The first one is composed of random coincidences between a prompt

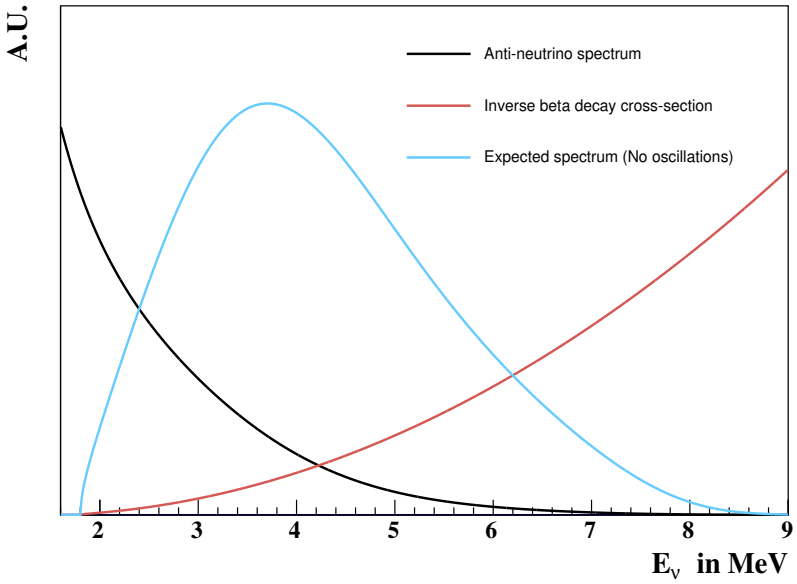


Figure 2.15: Anti neutrino spectrum (black), IBD cross-section (red) and expected antineutrino spectrum (blue) normalized in arbitrary unit.

ES and a delayed NS, produced by two independent processes. The correlated background comes from physical processes that create both an ES and a NS signal in a delayed coincidence spatially and in time. The two principal sources are the atmospheric background and the internal radioactivity. Those backgrounds will be identified for a part in the reconstruction procedure for the muons tracks, see section 2.7, and with offline selection on reconstructed variables that will be detailed in chapter 4.

2.5.2.1 Atmospheric background

The first correlated background faced by SoLid is induced by cosmic rays. High energy primary particles coming from space collide with atoms in the atmosphere and generate high-energy nucleons and elementary particles. After interactions in the atmosphere and decay of the secondary particles, neutrons will be created. Due to the low overburden of the detector, those neutrons can penetrate the detector without being stopped. If some of them are mitigated by the water wall, a part of fast neutrons can still enter the detector. Even if neutrons are neutral particles, they can still create electromagnetic signals with elastic scattering on protons with high-energy recoils. The recoiled protons will ionize the PVT, and induce its scintillation. Due to a scintillation quenching effect, those recoils will be measured in the energy range of the IBD. After thermalization in the detector and capture by a neutron screen, the neutrons will create a NS-ES coincidence mimicking an IBD event as represented on figure 2.16 (left).

Muons can also be created by resulting interactions of cosmic rays in the atmosphere. Those muons can create spallation neutron in the vicinity of the detector and follow the process described above. The muons can also interact directly in the detector. If they leave a track through the planes, they can easily be reconstructed and discarded when looking for IBDs. However, some of them can interact only in the corners, leaving energy in a few number of cubes in coincidence with a neutron previously created by

spallation as represented on figure 2.16 (right). This can be another source of background.

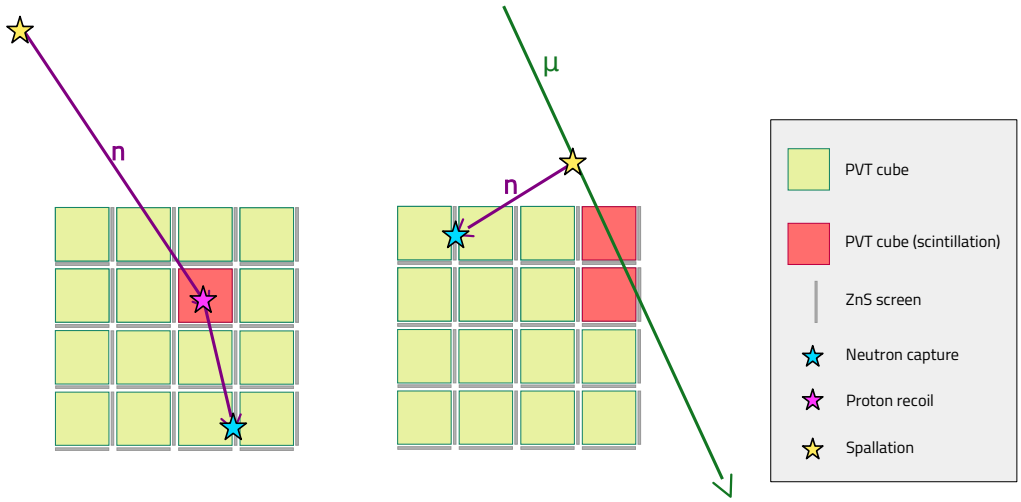


Figure 2.16: Scheme of the atmospheric backgrounds. On the left is represented an atmospheric neutron making a proton recoil and being captured afterwards. On the right is represented a muon creating a spallation neutron outside the detector sensitive volume and being detected in delayed coincidence with the neutron.

2.5.2.2 Internal radioactivity

The second source of correlated background comes from the natural radioactivity of the detector materials. ^{238}U, ^{232}Th and ^{40}K are the three main isotopes found in all materials. This is due to their half-lives of respectively 4.5 billion years, 14 billion years and 1.2 billion years. The ^{40}K decays into ^{40}Ar or ^{40}Ca via electron conversion or β emission. As both daughter nuclei are stable, the decay chain naturally stops and the ^{40}K does not represent a source of correlated background in this experiment. For the other two isotopes, radioactive elements are created in their decay chains and can be found in materials and be a source of background, as seen on figure 2.17. Events that would have an IBD like signature would be a prompt emission of a β with a delayed α with a time constant of tens of micro seconds with a β. This specific signal can be found in the ^{238}U chain with the so called BiPo decay, which consists in the decay of ^{214}Bi to ^{214}Po with the reaction:

$$^{214}_{83}\mathrm{Bi} \xrightarrow[19.9\ \mathrm{min}]{\beta^-} {}^{214}_{84}\mathrm{Po} \xrightarrow[164\ \mu\mathrm{s}]{\alpha} {}^{210}_{82}\mathrm{Pb} \tag{2.7}$$

The β^- emitted by the ^{214}Bi will be detected as a prompt ES and the delayed α from the ^{214}Po decay will produce a NS, exciting the ZnS(Ag) as represented on figure 2.18. The half-life of the ^{214}Po is of 164 µs, which corresponds to a time constant of 237 µs. A good amount of the BiPo events will thus be in the time-space coincidence window that will be defined for the IBD analysis. The maximum energy of the emitted β^- is 3.27 MeV thus in the energy region of interest of the IBDs.

Two sources of BiPo background can be identified. The first is ^{222}Rn emission from the concrete around the detector. ^{222}Rn is a noble gas with a half-life of several days. It can diffuse inside detection

planes, and it can be a homogeneous source of background. This background contribution will depend on the atmospheric pressure and the air flushing that flushes out the contaminated air. The other source is a contamination of the ZnS screens during their fabrication. This source is independent of external parameters and expected to be constant over time.

2.5.2.3 Accidental background

The accidental backgrounds in SoLid are due to an ES and a NS from different origins. The NS can be an α from natural radioactivity or a cosmic induced neutron. The ES can come from muons misidentified, or capture of the neutrons from the reactor by ^{40}Ar, carbon or hydrogen. The accidental background yield and nature thus depends on the reactor activity.

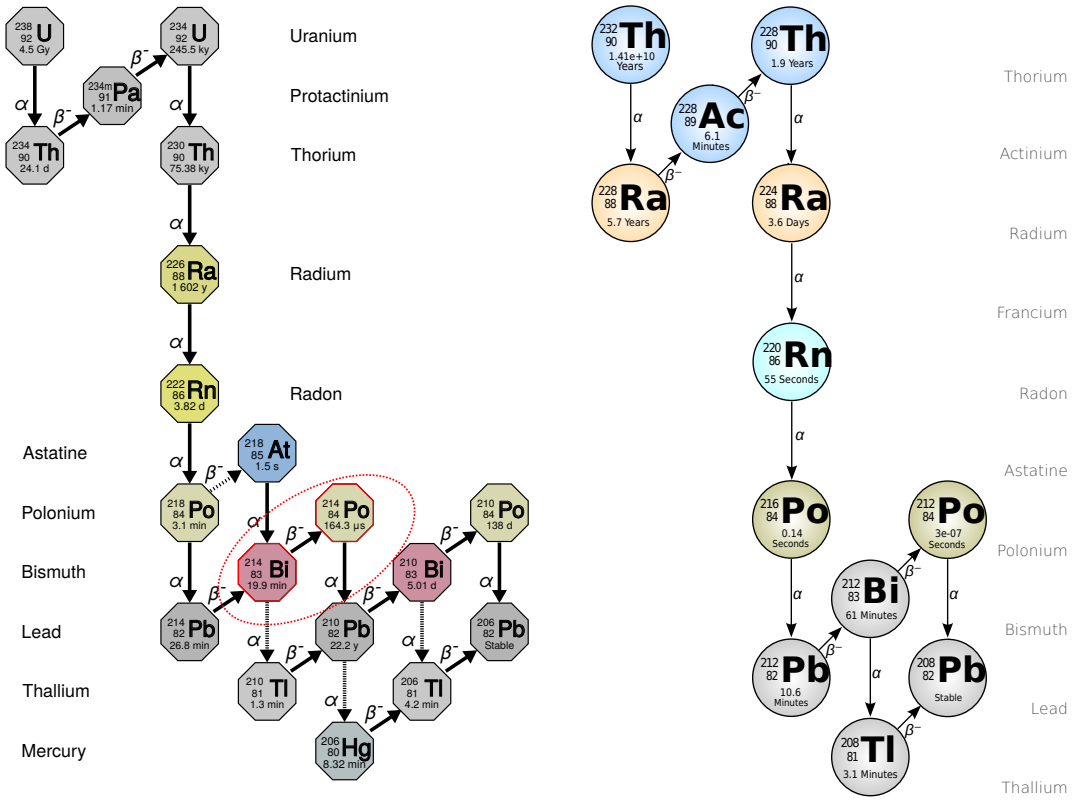


Figure 2.17: Left: Decay chain of the ^{238}U. The decays of the BiPo background are circled in red. The long-lived radioactive elements can contaminate the materials used in the detector and be a source of internal radioactivity background. Right: Decay chain of the ^{232}Th. The alpha emission in the BiPo decay of this chain is too fast to be treated as a delayed coincidence and will not be a part of the correlated background.

2.6 Simulations

The simulations in SoLid are divided into two programs: *SoLidSim* based on GEANT4 [89] to generate the detector geometry and the particle interactions and energy deposits and *ROsim* to reproduce all the detector effects. Both software will be presented in the next sections.

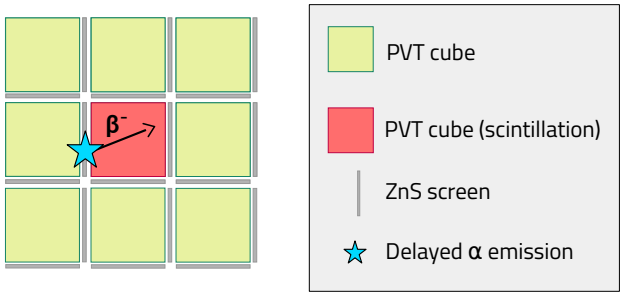


Figure 2.18: Scheme of BiPo interaction in the detector with a prompt β as ES and delayed α emission as NS.

2.6.1 Geant4 simulations

2.6.1.1 Geometry

With the IBD detection principle presented in section 2.2.2 and with a precise description of the geometry of the detector, a model of the detector has been implemented inside the GEANT4 simulation [90]. A precise description of the PVT cubes, the fibres, the Tyvek layers, the ZnS screens, the MPPCs, the aluminium frames, the HDPE neutron reflection screens has been used as described on figure 2.8. The surroundings of the detector have also been modelled with a special care to the CROSS system that enters inside the detector during calibrations. To generate the fast neutron and muon backgrounds, the reactor containment building has also been implemented, based on blueprints and measurements of the reactor building, with the different floors and the reactor pool.

2.6.1.2 Antineutrino generator

The antineutrino events are generated as presented in section 2.5.1. This generation is done with the so-called *SoLo* software that takes as an input the antineutrino spectrum model. This program computes the expected antineutrino interaction points inside the detector with their energies and momentum. The positron and the neutron created by the IBD interaction are also provided by the software, and SoLidSim will generate everything that happens after the IBD interaction. SoLo can also add a fake oscillation pattern to the antineutrino generated for further studies, as shown on figure 2.19 with an expected antineutrino signal without an oscillation patten on the left and with an oscillation generated by SoLo on the right.

2.6.1.3 Background generator

SoLidSim is also used to generate the correlated backgrounds.

- The incoming muons are generated with the Guang parametrization [92] that provides the energy, momentum, and angular distribution of muons at sea level.
- The incoming neutrons resulting from the cosmic rays interactions in the atmosphere are generated using Gordon parametrization [93] based on measurements done in several locations in the USA.

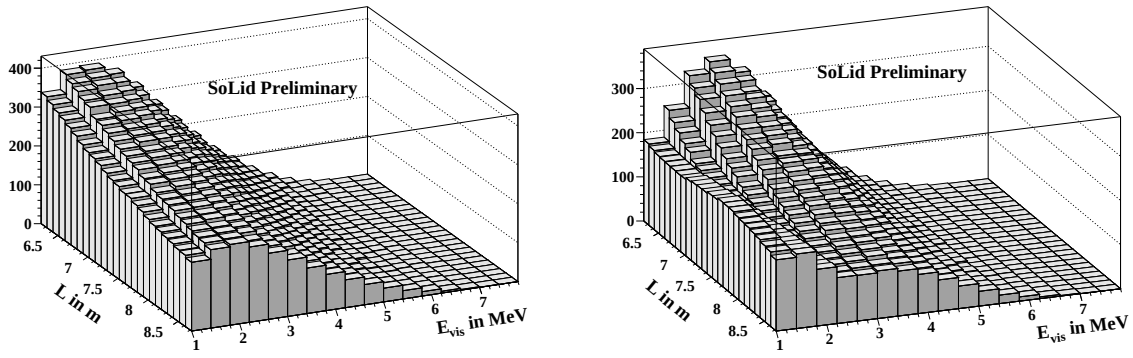


Figure 2.19: Antineutrino yield as a function of the positron energy and the distance travelled in case of no oscillation (left) and with a fake oscillation with $\sin^2(2\theta_{14}) = 0.5$ and $\Delta m^2_{41} = 1.78\ \mathrm{eV}^2$ (right). Here the $\sin^2(2\theta_{14})$ is artificially increased to make the oscillation more visible compared to the best fit point of the RAA at 0.09 [91].

– The BiPo background is simulated by ^{214}Bi uniformly generated in the ^{6}LiF:ZnS(Ag) layers.

2.6.1.4 Calibration sources generator

The gamma sources (^{22}Na, ^{207}Bi, ^{137}Cs) are generated with their radioactive decay and the neutron sources (^{252}Cf and AmBe) are generated with energy spectrum from the ISO norm. The simulation uses the same source position, as presented in section 2.4.2. With the precise CROSS geometry it reproduces the real conditions of calibration.

2.6.2 Detector response

The detector response is simulated by the *ROsim* C++ code that takes as an input a SoLidSim output and transforms it in the same format as the real raw data for analysis. It consists in two main steps, the simulation of the scintillation response and the simulation of the electronics response. Unlike the SoLidSim part that tracked every particle created, only the number of scintillation photons are transmitted from one step to the other and not each individual generated photons.

2.6.2.1 Scintillation response

The detector response simulation starts with the cube and fibre response simulation. For a given energy deposit in a given PVT cube, called the centre cube in the following, a number of scintillation photons is generated depending on the individual cube light yield and quenching effects described by the Birks' law [94]. A part of those scintillation photons leaks with a given probability in the neighbouring cubes, and a part of the neighbouring cube light leaks inside the centre cube. For an energy deposit inside a ^{6}LiF:ZnS(Ag) layer, a ZnS light yield is applied. The cube light yield set inside the simulation is not directly the PVT light yield of 10 000 photons per MeV but an effective light yield. It takes into account inefficiencies such as the light collection inefficiencies of the fibres, MPPCs inefficiencies, or any other light losses that would not have been taken into account by another parameter. The light yield value per cube is tuned using the energy calibration results, as will be discussed in chapter 3.

The remaining number of photons is equally distributed in the four fibres attached to the cube. For a given fibre, half of the photon is directly transmitted to the corresponding MPPC and the other half is transmitted first to the mirror before going to the MPPC. At this stage, three parameters can modify the number of photons detected by the MPPC: the fibre attenuation length, the mirror reflectivity and the coupling parameter that takes into account the goodness of the optical coupling between the fibre and the MPPC. Finally, the number of photons reaching the MPPC is randomized with a Poisson distribution. The time distribution of the incoming photons is generated with the travel time of the photons in the fibres and the time constants of both scintillators. The different steps are illustrated on figure 2.20.

In this process, the different parameters are tuned the following way:

- **The PVT light yield** is tuned with Data to Monte-Carlo comparisons on calibration source energy spectrum. It is then corrected with a cube to cube light yield variation measured with an energy calibration.
- **The ZnS light yield** is computed with data to Monte-Carlo comparisons on alpha from ^{214}Po energy spectrum.
- **The cube light leakages** are measured with horizontal muons. Those are expected to cross only one cube per plane, and the fraction of leakage is obtained measuring the amplitude in the neighbouring cubes. The fraction of leakage is set to 10% for all cubes.
- **The quenching of the light**, described by the Birks' law, is determined with simulations to match the manufacturer measurement for different particles. The value of the Birks' coefficient is $0.14\mathrm{mm/MeV}$.
- **The mirror reflectivity** is an empirical constant set to 0.8. It does not necessarily match the real mirror reflectivity because quantifying it would be too difficult to do it individually. This specific value does not matter as long as the global model describes the data.
- **The attenuation lengths** are measured in the calibration procedure described in section 3.3.2.
- **The fibre to MPPC coupling constants** are measured individually in the calibration procedure described in the section 3.3.2.

2.6.2.2 Electronic response

When a photon time distribution is transmitted to a MPPC, dark noise, cross talk and detection efficiency effects are applied to the photon distribution as described in section 2.2.2. The recovery effect of the different pixels after an avalanche is also treated in the software. A list of pixel avalanches is then created, and each avalanche generates a waveform with an empirical model combining two Gaussian models for a fast rise and a slow fall of the waveform, as shown on figure 2.21. A gain amplification is applied to the waveform to match to data run conditions. The waveforms are sampled to match the SoLid ADCs with a sampling frequency of $40\ \mathrm{MHz}$.

The sampled waveforms are then in a data-like format, and the trigger response model is thus a direct reproduction of the trigger logic presented in section 2.4.3.2. The obtained simulated data are then treated as real data in the event reconstruction software Saffron presented in the section 2.7.

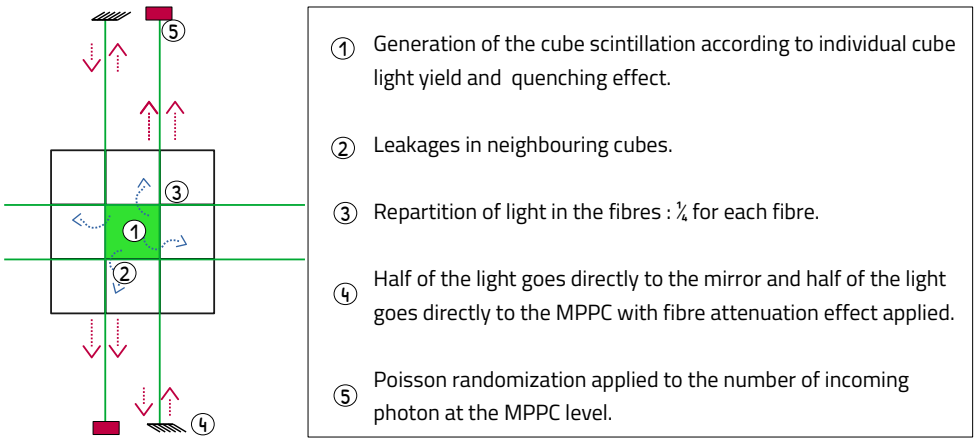


Figure 2.20: Scheme of the different steps happening in the scintillation simulation from the generation of scintillation photons to a measured number of pixel avalanches.

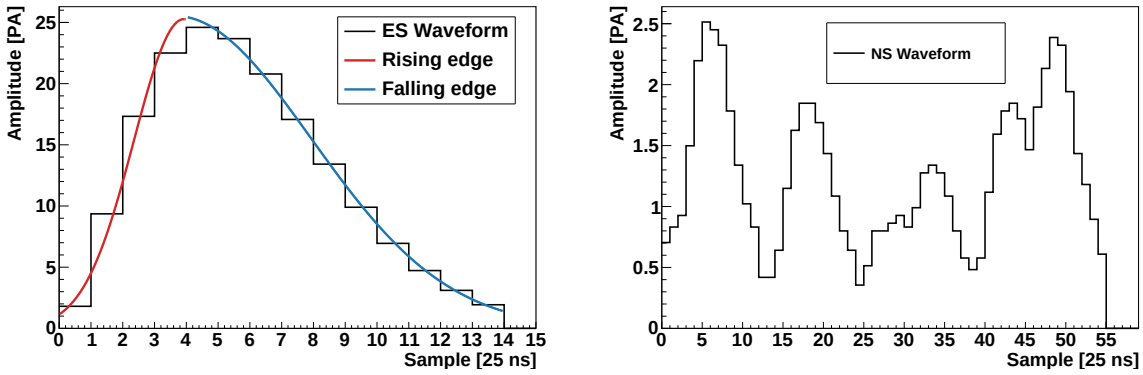


Figure 2.21: Example of data waveforms. Left: example of ES waveform. In red and blue are fitted the fast rise and slow fall of the waveform with the same parameters as the ROsim. Right: example of NS waveform.

2.7 Event reconstruction

The raw or simulated data are treated by the Saffron software. It is a C++ software used for event reconstruction and characterization. The general idea of the code is to aggregate the input waveforms in time and space clusters, and then classify the clusters according to their characteristics. For memory consumption optimization, the software works by cycle. A cycle represents a time window where the reconstruction is applied at once to all the data inside. It starts when a waveform is recorded and stops when a time of at least 2.240 ms (350 time blocks) is reached between two waveforms. The cycle data are stored on disk before freeing the memory for the next cycle to be analysed.

2.7.1 Waveforms time clustering

Inside a cycle, a set of time ordered waveforms, on several channels from several planes, is available. A time clustering plane by plane is performed. A scheme of the time cluster creation algorithm is presented on figure 2.22. Waveforms that are within 7 samples (175 ns) of the start of a given waveform are associated. An initial cluster is made if at least one horizontal and one vertical channel can be associated, otherwise the signal could not be localized in the plane. For data reduction, signals with an amplitude below 2.5 PA are removed from the clustering. This avoids the creation of a lot of clusters made of noise only, and corresponds to the level of data to Monte-Carlo agreement in the experiment.

Then, if several planes were readout by the trigger, clusters can be located in multiple planes and can be added together. The clusters that start within 7 samples are aggregated. To take into account the long queue of the NS, a larger time window is then created. Inside this window, waveforms that are recorded on the channels of the current clusters are added to the cluster. This window is a function of the number of channels in the cluster and the time length of the cluster. The time length is defined as the time difference between the start of the first sample of the first waveform and the time of the last sample of the last waveform. Its precise tuning is essential to avoid merging independents ES events while merging all the waveforms induced by a NS in the same cluster.

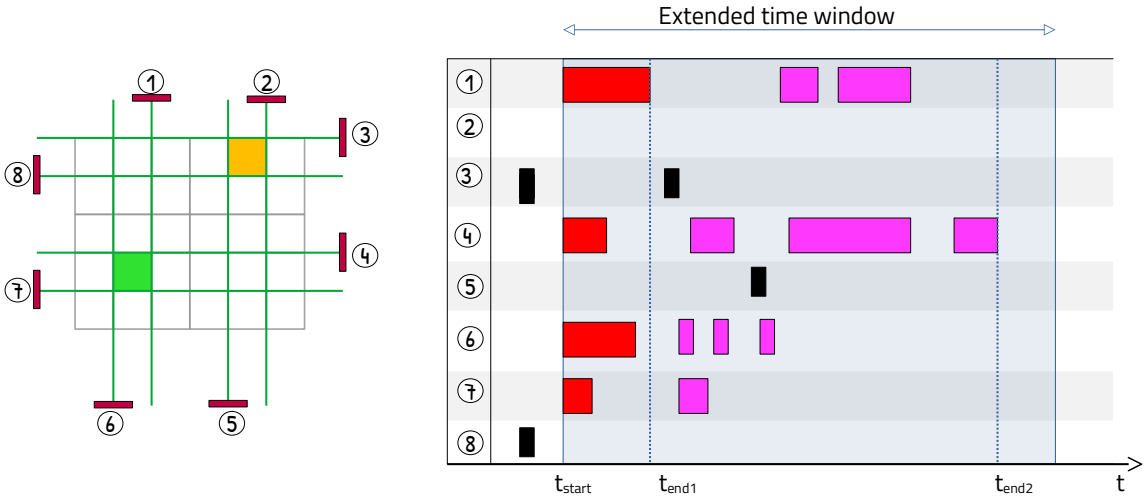


Figure 2.22: Illustration of the time clustering algorithm. Left: Schematic representation of a plane with two cubes that can be reconstructed. The green one has a real signal and the orange one just has channels seeing dark noise. Right: representation of the waveforms seen by each channel. The black rectangles are waveforms with too low amplitude treated as dark noise and are discarded. The red one represents a first cluster that is created between t_{start} and t_{end1}. The extended time window is applied to gather the waveforms in the same channels as the cluster, in pink. The real end of the cluster is then t_{end2} and it is made of red and pink waveforms.

2.7.2 Cluster categories

When a cluster is created, a set of cubes are reconstructed by matching all possible horizontal and vertical channels combinations. The cluster characteristics can be used to determine the physical process at the

origin. Three signals categories are then created: ES, NS and muons. The ES clusters are expected to be short clusters in time with high amplitude, the NS clusters are expected to be longer clusters with lower amplitude signals, and the muons clusters are expected to be tracks of PVT scintillation throughout the detector. The difference between NS waveforms and ES waveforms can be seen on figure 2.21. Three cluster categories can thus be reconstructed in SoLid and the identification methods will be presented in the next sections.

2.7.2.1 NS clusters

The ZnS scintillation characteristic will be used for NS cluster determination. The ^{6}LiF:ZnS(Ag) scintillation is induced by alphas and tritium from ^{6}Li breakup or alphas from radioactive decay. A NS cluster is thus expected to be most likely a cluster with two to four channels on only one cube. If multiple cubes can be reconstructed, the cube that has the largest number of peaks over a threshold of 0.6 PA is treated first. If possible, another cluster is created without the channels of the first cube and is added to the pile of clusters to treat.

The first selection cut applied to the cluster is a minimum length: the cluster must have a temporal length larger than 25 µs. This length is defined as the time difference between the first sample of the first waveform and the last sample of the last waveform. This cut takes into account the long tail of the ZnS scintillation and removes the majority of ES candidates. Then for each channel of the candidate cube, the amplitude of the waveform of maximum amplitude is computed and all the maximum amplitudes are averaged by the number of channels. This variable is noted $\overline{A}$.

$$\overline{A} = \frac{1}{n_c} \sum_{channels} MAX(A_{channel}) \tag{2.8}$$

with n_c the number of channels, $MAX(A_{channel})$ the amplitude of the waveform that has the maximum amplitude. The Integral of the waveforms within 0.75 µs is also computed for each channel and averaged. This will be noted $\overline{Q}$. Then, a selection is applied combining the two variables:

$$\overline{Q} > 10.5 \times \overline{A} - 800 \tag{2.9}$$

This cut helps to clean the contamination of ES clusters with large $\overline{A}$ and low $\overline{Q}$. This cuts also prevents the creation of multiple NS cubes that would come from one NS in one cube sharing channels with an ES or a muon cube in the same row or column.

A last selection is applied on the asymmetry on the cube channels. To avoid miss-reconstructed NS cubes, if the channels in one direction (vertical or horizontal) are receiving more light than the two other channels (horizontal or vertical), it is asked that charge $\bar{Q}$ is high enough so that it is certain that all channels receive enough light. The selection applied is the following:

$$\bar{Q} > 4250 \times |\delta^3_{X-Y}| + 650 \tag{2.10}$$

with δ_{X-Y} an asymmetry term defined as following:

$$\delta_{X-Y} = \frac{\overline{Q_l(X)} - \overline{Q_l(Y)}}{\overline{Q_l(X)} + \overline{Q_l(Y)}} \tag{2.11}$$

with $\overline{Q_l(X)}$ ($\overline{Q_l(Y)}$) the averaged amplitude integral on 25 µs on the channels on the horizontal (vertical) direction. Those selection cuts have been studied extensively by our colleagues in LPC Caen and more details on the selection can be found in [84]. The example of the $\overline{Q}$ and $\overline{A}$ distributions can be seen on figure 2.23

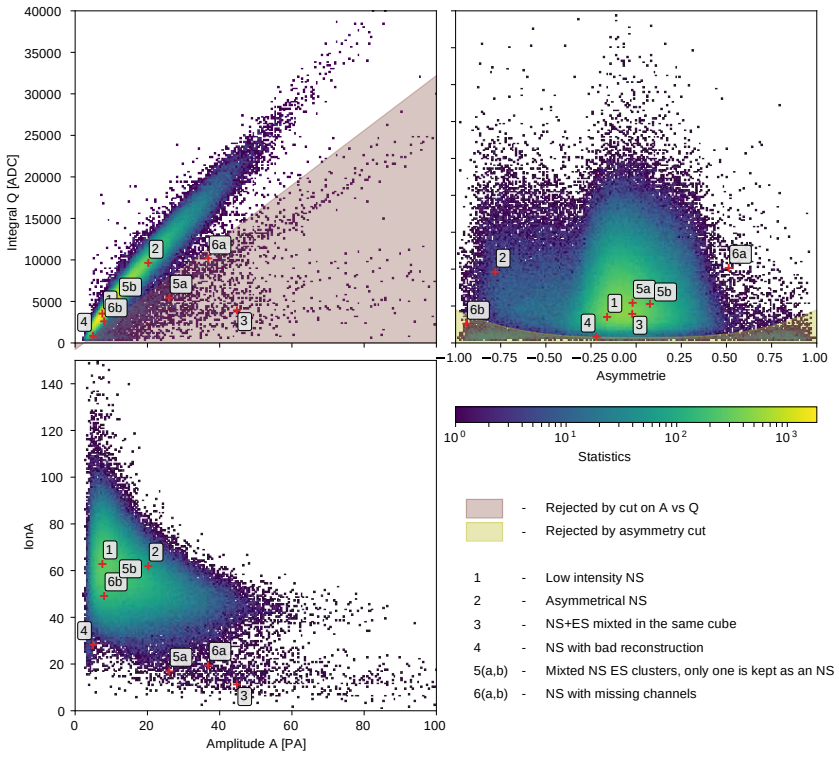


Figure 2.23: Example of NS candidates. Top left: $\overline{Q}$ versus $\overline{A}$ distribution. The red area is the clusters rejected by the selection cut $\overline{Q}$ versus $\overline{A}$. Top right : $\overline{Q}$ versus δ_{X-Y}. The yellow area represents the candidates rejected by the asymmetry cut. Bottom left: integral over amplitude versus the amplitude. The detail of each cluster can be found on the bottom right panel [84].

2.7.2.2 Muon cluster

Unlike NS clusters, the muon clusters are induced by PVT scintillation. Here, the main discrimination point that is used is the number of channels of the cluster. A muon is expected to cross the detector, depositing a large amount of energy in a high number of cubes. A cluster is tagged as a muon if it has at least 11 horizontal or vertical channels with a maximum amplitude above 200 ADC counts ($\sim$ 6 PA).

A fit is then performed on the track both vertically and/or horizontally if the number of channels passes the previous cut. The fit as represented on figure 2.24 is a straight line fitted in the (Z, X) and (Z, Y) projections of the cubes.

From the fit convergence, three types of muons can be determined:

– **Type 0 muons** are muons clipping in the detector. They leave energy in only few cubes on the edge

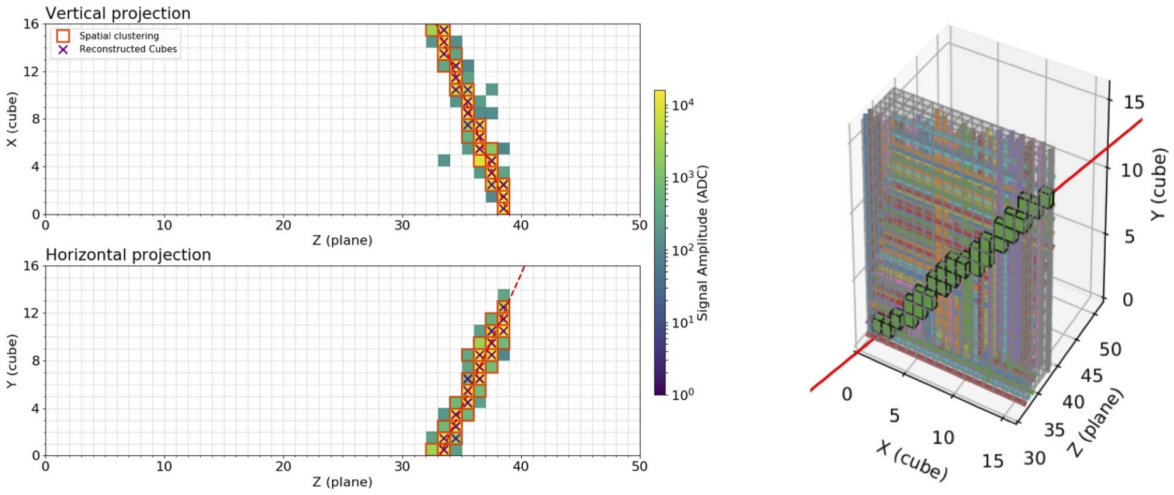


Figure 2.24: Example of a muon cluster reconstructed as type 2. Left: projections on the planes (X,Z) and (Y,Z) of the signal amplitudes. All the cubes represented are part of the muon cluster. A red line is fitted in both projection that allows knowing the penetration angle of the muon. Right: 3D representation of the detector with the muon track fitted represented in red and the associated cubes are in green.

of the detector. For those muons the fit does not converge. Those muons do not leave a track to be fitted and are harder to tag. In the current state of the SoLid reconstruction, those muons are 10% of muons interacting in the detector, but they are not tagged.

– **Type 1 muons** are muons that are reconstructed in only one plane. Those muons leave a track in the detector and are tagged more easily, but there is still an ambiguity on the cubes crossed by the muons due to the fibre sharing.

– **Type 2 muons** are muons that cross several planes. For those muons the penetration angle inside the detector is well known and those are perfectly reconstructed as illustrated on figure 2.24.

2.7.2.3 ES cluster

The remaining clusters not tagged as NS or muons are then tagged as ES. However, not all clusters are reconstructed: to reduce the dark noise, the channels with an amplitude below 2.5 PA are discarded. A simple reconstruction algorithm was performed at start: a cube was reconstructed if at least one horizontal and one vertical channel attached to this cube passed the previous threshold. The cube energy was then defined as the sum of the fibre amplitudes. This reconstruction method can be accurate if a single cube is reconstructed in a plane but can quickly lead to imprecisions in the cube energy and position. As shown in figure 2.25 ambiguities can happen with fake cubes being reconstructed or cubes can be reconstructed with a wrong energy due to pile up in the channels when two cubes are reconstructed sharing some fibres. To cope with those reconstruction issues, a Maximum-Likelihood Expectation-Maximization (ML-EM) [95] based reconstruction have been developed by the SoLid collaboration. This type of algorithm has also been used by the NEXT experiment [96].

The reconstruction, called CCube reconstruction, method is based on an iterative process: from the actual measurements of the channel energies, predictions are made on the repartition of the energy de-

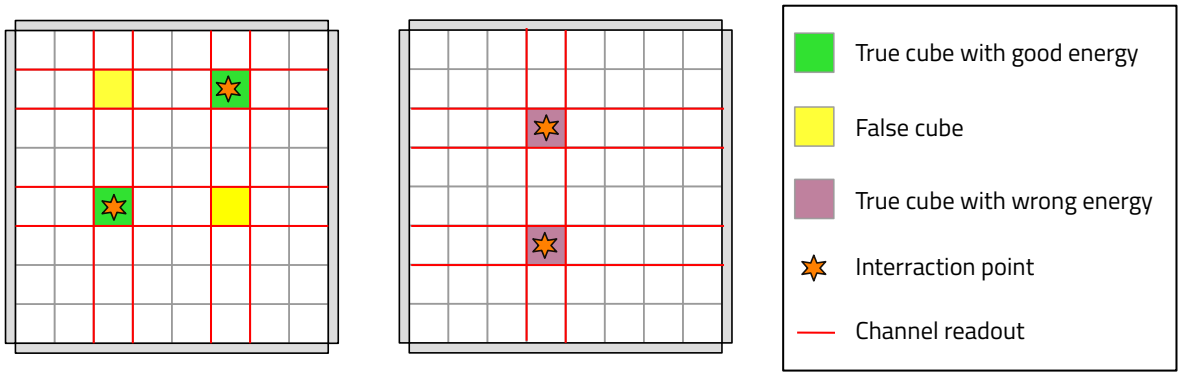


Figure 2.25: Example of cubes with reconstruction issue. Left: the cluster contains two interactions in two cubes. From the two real interactions inside two PVT cubes, two other fake cubes can be created from coincidences with vertical and horizontal channels. Right: The cluster contains two interactions in the same column on two cubes sharing two channels. The energy of both cube will thus be wrongly estimated, with the two vertical channels reading out the sum of both signals.

posits of the cluster in each cube to minimize a likelihood function. The fraction of the cube energies expected to be read out by each MPPC is modelled in a system matrix. The matrix encodes known detector effects such as light yield variations, fibre attenuation, light leakages and fibre to MPPC optical coupling inhomogeneities. A comparison is then performed between the modelled repartition of the energy in the MPPCs and the measurement and a new estimation is created until convergence of the projections, as shown on the figure 2.26. The master equation of the algorithm can be derived as:

$$E_j^{n+1} = \frac{E_j^n}{\sum_i A_{ij}} \sum_i A_{ij} \frac{p_i}{\sum_{\hat{j}} A_{i\hat{j}} E_{\hat{j}}^n} \tag{2.12}$$

with E_j^n the energy of the cube j at the iteration n, i the index of the four MPPCs linked to the cube, p_i the energy measured by the MPPCs, A_{ij} the components of the system matrix and $\hat{j}$ the index of the cubes in the same row or column as the cube j.

In the results presented in this document, a correction is applied to the channel measurements p_i before the algorithm to correct for the coupling to MPPCs inhomogeneities and gain variation from one channel to another. The system matrix is assumed to be flat: A_{ij} is a matrix composed of only 0.25 everywhere. This considers that the four MPPCs of a cube receive the same amount of light. More realistic system matrices are under development to take into account the attenuation lengths of the channels and the light leakages of the cubes.

2.7.2.4 NS-ES coincidence

Once all the clusters have been reconstructed and classified, ES clusters are associated with NS clusters in delayed coincidences. Each NS cluster is associated to all ES clusters that starts within a time window

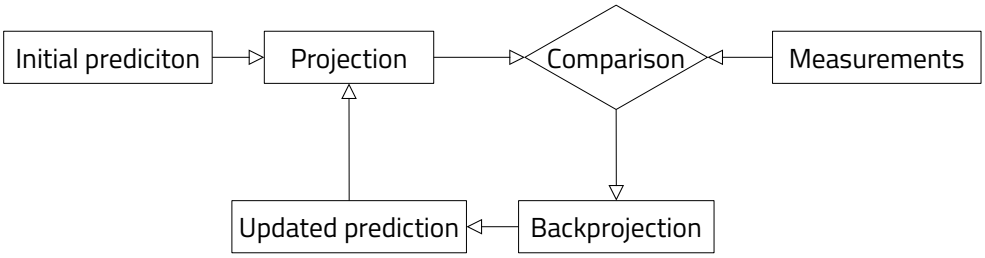


Figure 2.26: Principle of the reconstruction algorithm based on ML-EM.

of $\Delta T_{NS-ES} \in [-200, 500]$ µs and with an ES global energy, defined as the sum of all channel energies, larger than 1 MeV (not to be confused with the previous cube energy defined with the ML-EM algorithm). The positive part of ΔT_{NS-ES} is chosen to study the IBD-like NS-ES coincidences while the negative part of the time difference is chosen to study the accidental background. The reference cube to define the distance between the ES and NS for topological studies is the most energetic cube in which the IBD interaction is expected to happen.

Chapter 3

Energy calibration of the SoLid detector

3.1 Introduction

For any event in SoLid, a set of amplitudes are measured by the detector channels. The amplitude readout by a channel i can be written as:

$$A_i = (\sum_{c \in cubes} N_{c,i}^{PA}) \times G_i + P_i \tag{3.1}$$

where G_i is the gain of the channel, P_i is the pedestal of the channel, $cubes$ is the set of PVT cubes sharing the fibre i where a particle has deposited some energy, $N_{c,i}^{PA}$ is the number of scintillating photons readout by the MPPC of the channel i induced by an energy deposit in the cube c. For an energy deposit E_c, assuming that the light is equally distributed in the four channels of a cube, one can write:

$$N_{c,i}^{PA} = \frac{1}{4} \times E_c \times LY_c \times \epsilon_i^{coupling} \times \epsilon_{i,c}^{atten} \tag{3.2}$$

where LY_c is the yield of the visible scintillation light that reaches the fibres for a given energy deposit in the cube c, $\epsilon_{i,c}^{atten}$ is the fibre attenuation effect along the fibre i for a light collection in the cube c and $\epsilon_i^{coupling}$ is a quantification of the light losses in the fibre not due to attenuation. Those different parameters will be presented in detail in the next sections.

From the 3200 fibres to the 12800 cubes there are 25600 parameters to measure, understand and correct in order to control the energy response of the detector. This chapter presents the methods developed by the collaboration and the results obtained on the energy calibration on which I was involved. The energy calibration is divided in three steps represented on figure 3.1. The first one is the equalization of the electronics response with the gain and pedestal inhomogeneities. The second is the correction of light losses in the fibres and the third is the correction of the individual cube light yields.

3.2 Calibration experimental conditions

The calibration system used for the SoLid collaboration has been presented in section 2.4.2. It consists in an automated robot that can place radioactive sources in different positions inside the detector. Neutron sources (AmBe and ^{252}Cf) were used for the neutron detection efficiency measurements and gamma

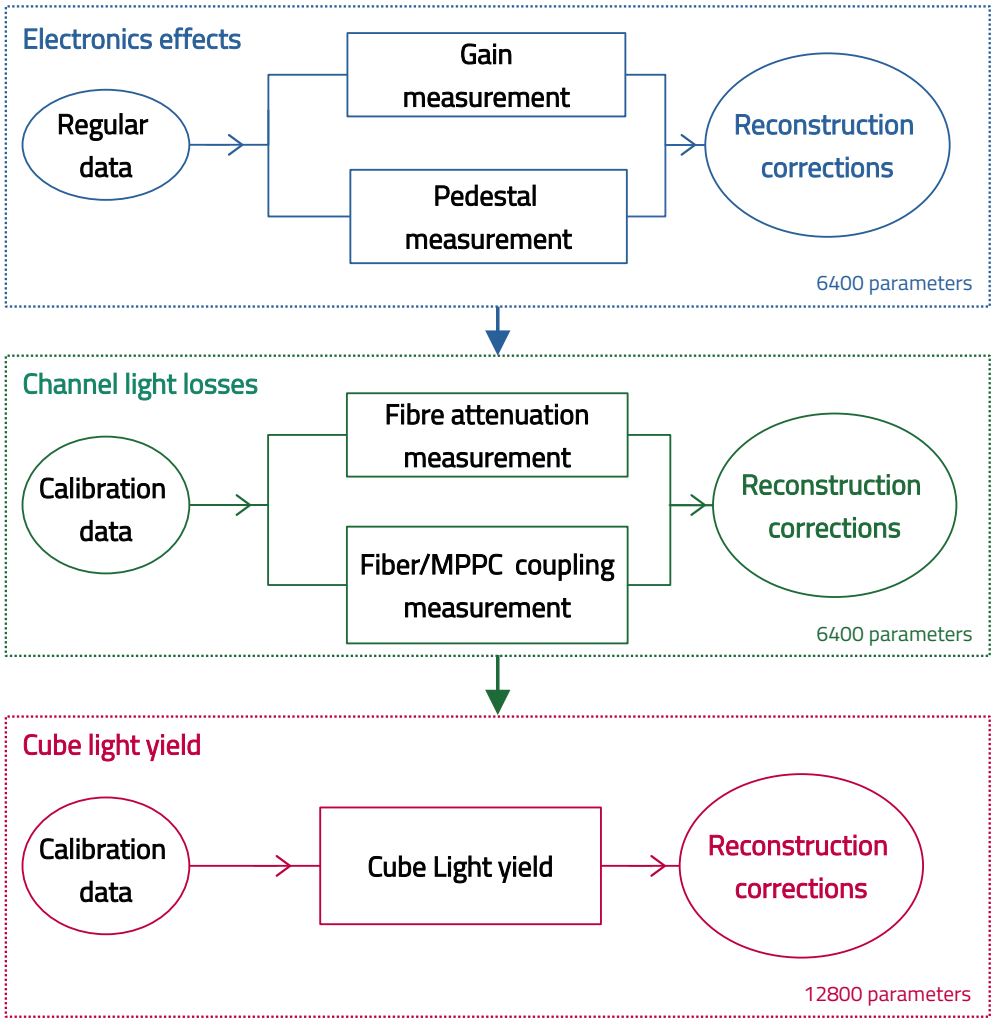


Figure 3.1: Scheme of the energy calibration procedure. The procedure can be decomposed in three main steps: an equalization of the electronics effect of the MPPCs, a correction of the light losses in the fibres and a measurement of the cube light yields.

sources (^{22}Na, ^{207}Bi and AmBe) for the energy calibrations. The gamma emission of the three sources can be described as:

- The ^{22}Na source decays into an excited state of ^{22}Ne via β^+ emission in $\sim 90\,\%$ of the cases [97]. The excited ^{22}Ne emits then a 1275 keV gamma. With the annihilation of the β^+ particle, the available gammas are then 511 keV and 1275 keV gammas. The decay chain of the ^{22}Na is represented on figure 3.2 (right).

- The ^{207}Bi decays into ^{207}Pb via electron capture. The decay can lead to three excited states of the ^{207}Pb and the available gammas are 1770 keV, 1063 keV and 569 keV gammas. The decay chain of the ^{207}Bi is represented on figure 3.2 (left).

- The AmBe source is a neutron source with the following process:

$$\begin{aligned} {}^{241}\mathrm{Am} &\rightarrow \alpha + {}^{237}\mathrm{Np} \\ \alpha + {}^{9}\mathrm{Be} &\rightarrow n + {}^{12}\mathrm{C}^{*} \end{aligned} \tag{3.3}$$

  With the emission of a 4438 keV gamma by the excited carbon atom with a branching ratio around $60\,\%$ [98].

A first calibration campaign was performed in March 2018 with the ^{22}Na calibration source. For this campaign a simple trigger was implemented: the readout was triggered if one channel had an amplitude above 9 PA in all the detector. In that case, the plane containing the channel and $\pm$ two planes around it were readout on two time blocks (the triggering block and the next one, which corresponds to 12.8 µs of data being readout). Due to the high activity of the source (37 kBq) and the simple trigger that led to a high trigger rate, the detector was in dead-time during 90 % of the run. In order to have a good amount of statistics to be able to calibrate all the cubes in the detector, the calibration campaign took seven days. After this campaign, the trigger settings were adapted with the aim to have faster and more efficient calibration campaigns. For the regular ^{22}Na calibration campaigns:

- The readout is triggered if two orthogonal channels in a plane are reading a signal above 7.5 PAs.

- Only the plane that triggers is readout.

- Only the ± 5 planes around the calibration source can be triggered.

All those settings were tuned by the calibration team in order to reduce the data rate drastically and mainly record data useful for the calibration. With those new settings, the calibration time for all the detector was reduced to one day. Calibration campaigns with those settings and with the ^{22}Na source are performed regularly to measure the detector energy response time evolution.

Special calibration campaigns were also performed in September and May-June 2020 with more sources. Those campaigns were dedicated for energy linearity study and data to Monte-Carlo comparisons. Due to the bias of the threshold trigger presented before at low energy, with a trigger designed to discard low energy deposit, another trigger strategy was used for those data. A periodic trigger was used where ± 5 planes around the calibration source were triggered. A detail of the calibration campaigns performed for SoLid Phase I (between September 2018 and June 2020) can be seen on table 3.1 and a detail on the calibration sources used can be seen on table 2.1.

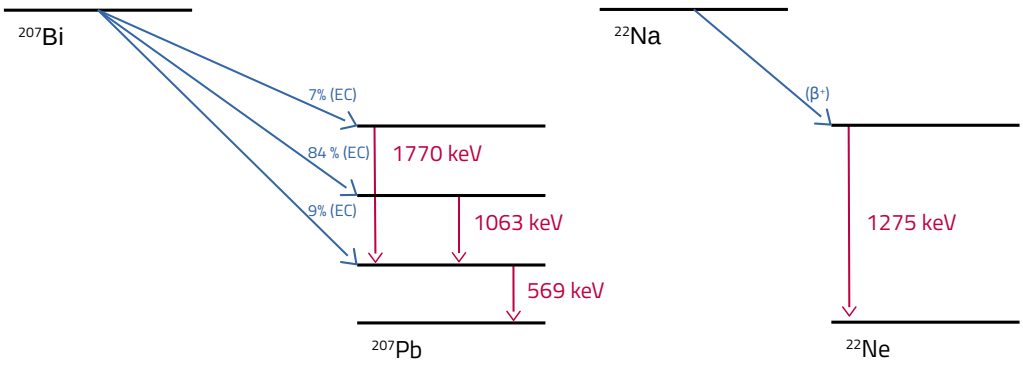


Figure 3.2: Left: main energy levels of the ^{207}Bi decay to ^{207}Po. Right: energy level of the ^{22}Na decay into ^{22}Ne [97].

Date	Source	Trigger type	Position	Calibration type	duration [day]
September 2018	^{22}Na	Threshold	Whole detector	Energy	1
	^{22}Na	Periodic	Gap 2 and 3, 9 positions	Energy	<1
	^{207}Bi	Periodic	Gap 2 and 3, 9 positions	Energy	<1
	AmBe	Threshold	Gap 2 and 3, 9 positions	Energy	<1
	AmBe	Neutron	Whole detector	Neutron	1
	^{252}Cf	Neutron	Whole detector	Neutron	1
October 2018	^{22}Na	Threshold	Whole detector	Energy	1
December 2018	^{22}Na	Threshold	Gaps 1 to 5	Energy	1
January 2019	^{22}Na	Threshold	Whole detector	Energy	1
May 2019	^{22}Na	Threshold	Whole detector	Energy	1
May 2019 bis	^{22}Na	Threshold	Whole detector	Energy	1
August 2019	^{22}Na	Threshold	Whole detector	Energy	1
February 2020	^{22}Na	Threshold	Whole detector	Energy	1
May 2020	^{22}Na	Threshold	Gap 5	Energy	< 1
June 2020	^{22}Na	Periodic	Gap 5, 2 positions	Energy	3
	^{207}Bi	Periodic	Gap 5, 2 positions	Energy	2
	AmBe	Threshold & Neutron	Gap 5, 2 positions	Energy & neutron	3

Table 3.1: Detail of the calibration campaigns performed between September 2018 and June 2020. Two similar calibration campaigns can be seen in May 2019 due to an intervention that took place on the detector between the two calibrations.

3.3 Channel equalization

The first step of the energy calibration is the channel equalization. It consists in three main tasks:

- Gain and pedestal equalization.
- Fibre attenuation length measurement.
- Fibre to fibre inhomogeneity correction, mostly identified as optical coupling differences between the fibres and the MPPCs.

The evaluation of each parameter will be described in the coming sections.

3.3.1 Gain and pedestal measurement

A gain equalization was performed at the beginning of the SoLid Phase I data taking. All the channel gains were set to 31.5 [ADC/PA]. Due to external condition variations (mainly temperature) of a few percent are expected during the two years of data taking.

The pedestal of each MPPC is computed several times a year with a voltage scan (see section 2.4.1) after intervention on the detector or in the containment building that required to shut down the electronics. This value is not updated frequently and do not take into account a possible drift in time.

To correct those effects of gain and pedestal evolution, the two parameters are computed in a 6 hours frequency to correct a possible day/night evolution on processed data. A measurement is performed using a small sample of the data that are processed on every physics run (every ~ 8 minutes). In those data, the amplitude distributions per channels are gathered in 6 hours time range. The amplitude peak positions are then fitted with Gaussians to get the amplitude as a function of the number of pixel avalanches, each amplitude peak corresponding to 1 PA, as represented on figure 3.3. The gain and pedestal drift can then be derived using equation 3.4:

$$\text{Amplitude [PA]} = (\text{Amplitude [ADC]} - \text{Drift [ADC]})/\text{Gain [ADC/PA]} \tag{3.4}$$

With the amplitude measured being expressed in ADC units and the equalized amplitude in pixel avalanches (PA). The amplitudes in PA will be used next. The pedestal drift is then added to the reference pedestal to obtain the real pedestal value per channel in a 6 hours time range and corrected in the data processing. The gain and pedestal time evolution of ten channels is represented on figure 3.4. Periods without any gain or pedestal values are periods where the data taking with the usual running condition was stopped, due to technical matters (chiller failure, pressure test in the containment building, maintenance of the detector) or a calibration campaign. Some drops or peaks of the gain value can be observed after a period without any gain measurement, like the three drops between August 2019 and October 2019 for instance. This can be explained by a temperature not stable in the detector just after a restart of all the electronics and the chiller. After the restart of the detector, it takes a small time for the temperature to be equalized at the same nominal value in all the detector. Those periods of unstable electronics will be removed from the antineutrino analysis after a set of data quality tests.

The variations of the gains are within 2 % percent in average during the whole Phase I period and the pedestal variations are below one percent, assuring a stable response of the channels over time. On the figure 3.5 (3.6) can be seen the value of the gain (pedestal) averaged on the whole Phase I period on

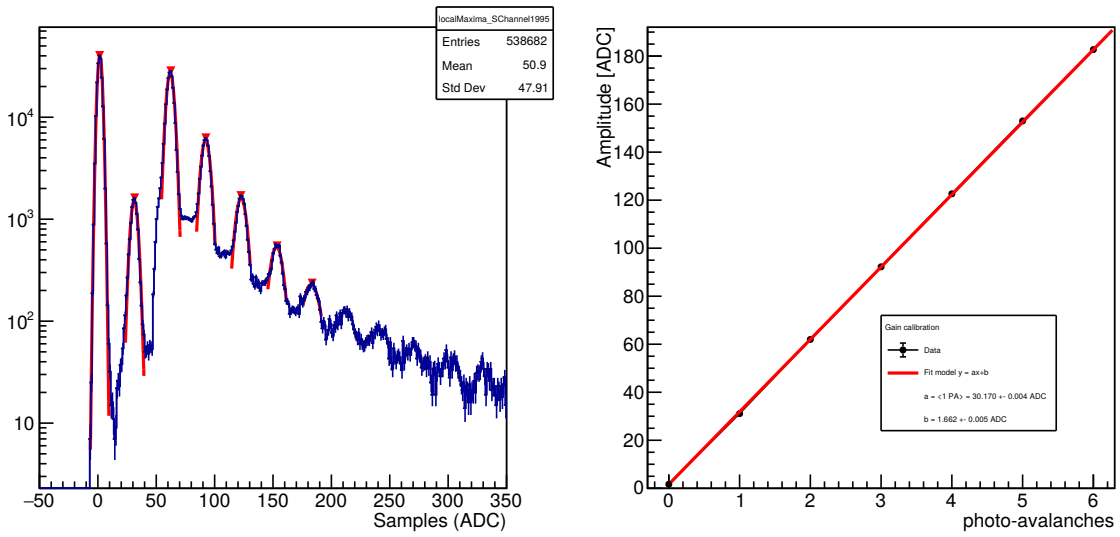


Figure 3.3: Amplitude spectrum for a given channel in ADC units (left). Fit of the Amplitude as a function of the pixel avalanche peak number (right). The slope is the gain and the intercept is the pedestal drift.

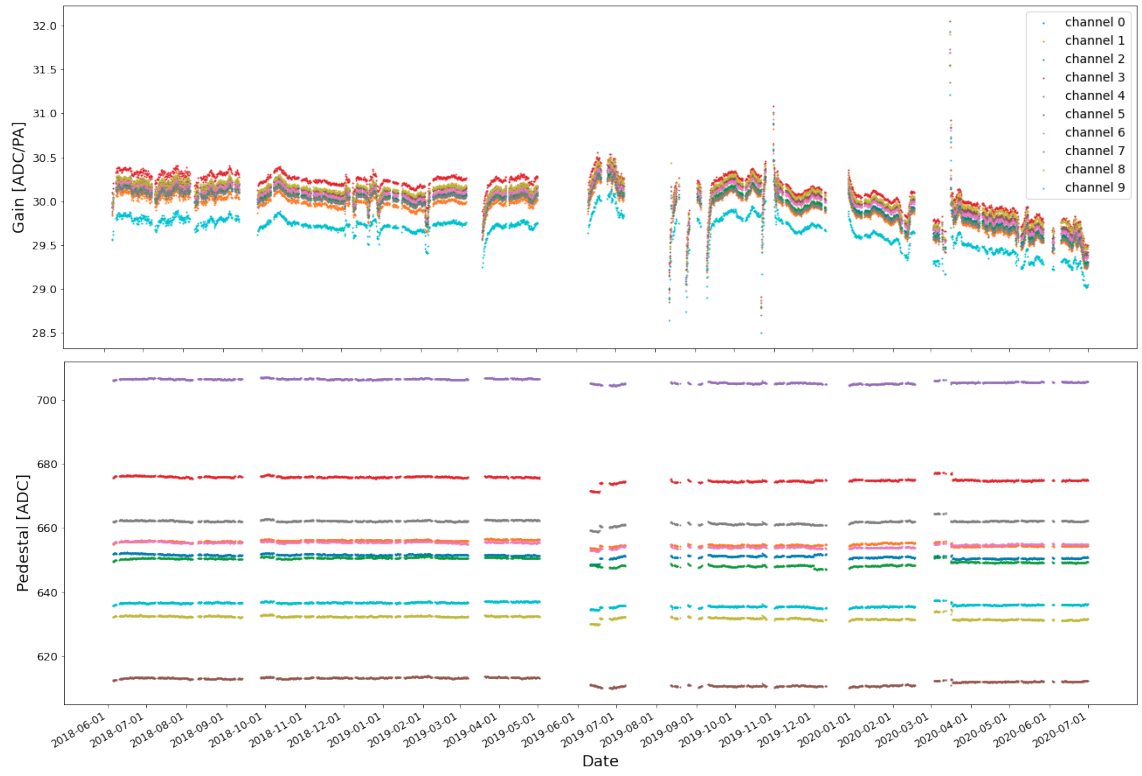


Figure 3.4: channel gain (top) and pedestal (bottom) variation during Phase I for 10 different channels. The periods without any values correspond to periods where no ordinary physics data are taken such as calibration campaigns or stopped data taking.

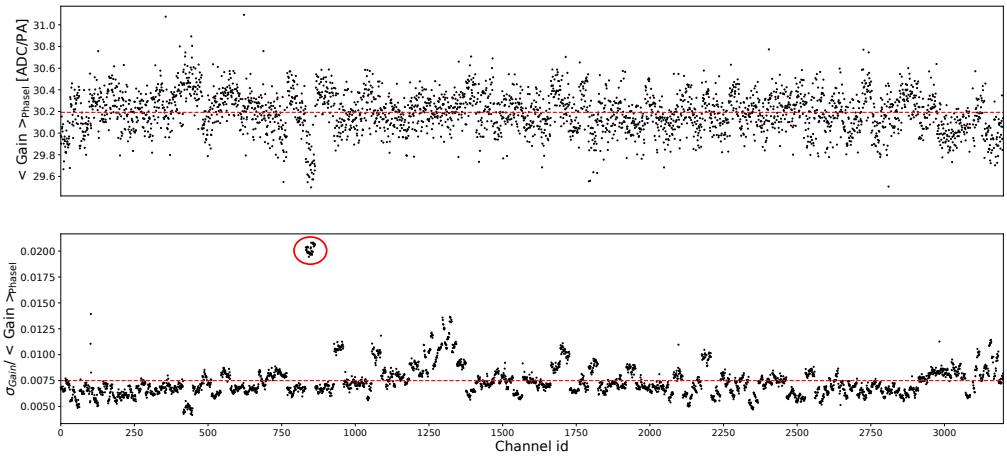


Figure 3.5: Top: gain of the active channels averaged on the two years of data taking. Bottom: standard deviation of the gain divided by the averaged gain on the two years of data taking. The red lines represent the average values. On the bottom plot, the circled points represent channel on the same analogue board that have a larger gain variation through the two years of data taking.

the top plot. On the bottom plot is represented the standard deviation of the gain (pedestal) for each channel computed on the whole Phase I period. On the bottom part of the figure 3.5, one can see a region circled in red. This group of channels that show a larger variation of the gain through the time are all the 32 channels of one analogue board. On figure 3.7 is represented the variation of the gain for ten of those channels. In the end of the year 2019, a gain diminution can be seen on those channels. However, those variations are still within 5 % between the highest and the lowest gain values. For the antineutrino data, the trigger threshold is set at 1.5 PA and the reconstruction threshold, after gain correction, is set at 2.5 PA so the 5 % variation would not induce any rate variation in the reconstructed data. This shows that the gain measurement and equalization is precise enough to have a stable response of the electronics through time.

3.3.2 Fibre light losses

From the light emission in a scintillator after an interaction, to the electronic signal measured in SoLid, the light is transported from the PVT cubes and ZnS layers to the MPPCs by wavelength shifting optical fibres. This process induces light losses that can systematically differ from one channel to another. Those differences can be induced by several effects. The first one is the inhomogeneities of the attenuation length of the fibres. The attenuation effect is due to scattering or absorption processes in the fibres. The second type of light losses regroups the losses due to the interface between a fibre and the cubes, the mirror and the MPPC. This last type of effect will be all treated as a single coupling effect per channel. To measure those parameters, regular ^{22}Na calibration data with the threshold trigger presented in section 3.2 are used.

A first step of correction of the gain and pedestal differences between the channels is performed. Then, a first light yield measurement is made on the calibration data with the method that will be pre-

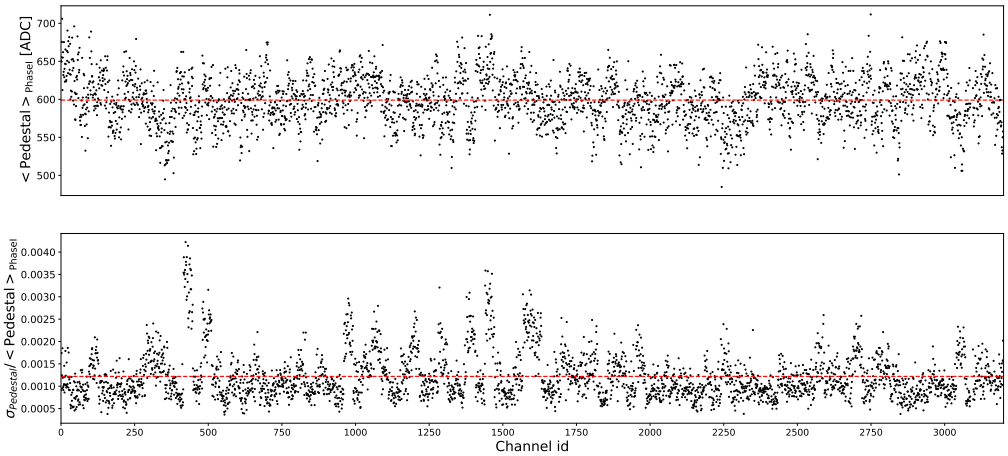


Figure 3.6: Top: pedestal of the active channels averaged on the two years of data taking. Bottom: standard deviation of the pedestal divided by the averaged pedestal on the two years of data taking. The red lines represent the average values.

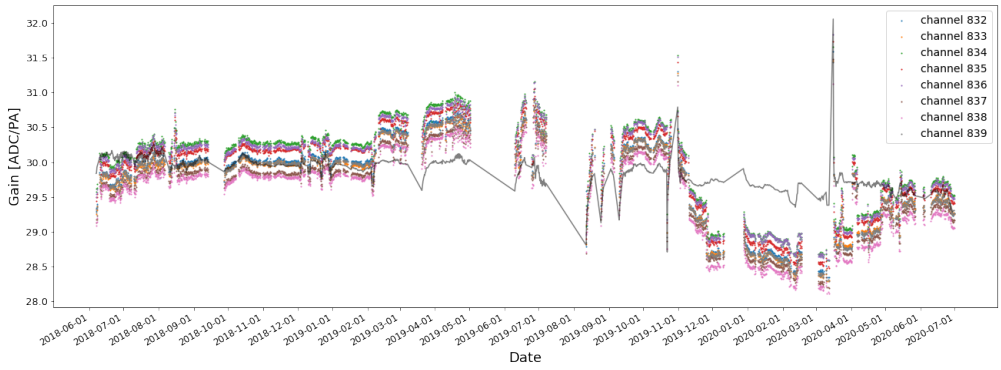


Figure 3.7: Channel gain variations during Phase I for 10 different channels. Those channels have the largest gain variation through time. All ten of them are part of the same analogue board. In black is represented the mean gain value of the 3200 channels.

sented in the section 3.4.2. This first measurement is done in order to select energy deposits that were induced by the $1.27\ \mathrm{MeV}$ gamma from ^{22}Na to avoid any measurement on cubes with a bad reconstruction due to ambiguities, or fake cube creation for the attenuation measurement. A selection of cubes alone in their plane is also used to avoid the pile-up of several energy deposits along the same fibre. This pile-up can be induced by two different gammas interacting in the same plane in the same row or column, or one gamma interacting twice in the same plane, biasing the amplitude readout by two of the four cube channels.

Two methods were developed for the attenuation length and coupling measurements. The first one, is based on a sequential measurement. The attenuation length is measured for each fibre as a first step and then after a correction of the attenuation effect, the residual inhomogeneities between fibres are used to measure each coupling parameter. The first method considers that the attenuation length and coupling can be measured independently one channel at a time and one parameter at a time. However, there is an interdependence of the variables used in the measurement of the channel parameters measurement within a plane. This is why a simultaneous fit per plane was developed by the colleagues in Subatech Nantes, deriving the 128 parameters per plane at once. The two methods will be presented in the following sections.

In this section and the following ones, the MPPCs placed at different sides of the detector might be treated differently. For clarity, the four sides will be designed as top, bottom, left and right. The top (bottom) side represents the MPPCs placed above the cubes in the row y=15 (below the cubes in the row y=0). The left side (right) are the cubes next to the column x=0 (x=15).

3.3.2.1 Sequential method

Attenuation length. To measure the attenuation effect along a fibre, one can compute the deposited energy along the fibre with a known initial energy deposit and derive an attenuation length from measurement of the attenuation effect:

$$A_{i,j} = f_i^{attenuation}(j) \times a_{i,j} \tag{3.5}$$

with $A_{i,j}$ the amplitude measured by the channel i in the cube j, $f_i^{attenuation}(j)$ the attenuation function of the channel i applied to the cube j and $a_{i,j}$ the amplitude without any attenuation effect. However, along a fibre, the amplitudes are measured on cubes with different light yields as described in equation 3.2: $A_{i,j} \propto LY_j$. As both the cube light yields and the attenuation effect vary along the fibre, it is not possible to do this direct measurement.

Instead, one can divide the amplitude of the fibre in the cube by the total amplitude of the cube, defined as the sum of the amplitude of the four channels of the cube. The obtained light fraction becomes independent of any cube light yield:

$$f_{i,j} = \frac{A_{i,j}}{\sum_{k=1}^{4} A_{k,j}} \tag{3.6}$$

With a selection of the energy of the cube between 0.99 and 1.01 MeV and computing the attenuation on 16 different points, one could expect the correlation between the different parameters to average out and the attenuation length of the channel to be obtained. For a given fibre, the fractions of light recorded by the MPPC as a function of the cube distance to the MPPC are thus measured along the fibre. On figure 3.8 are represented the light fractions for the channels in the four sides of a given plane. The light fraction

along the fibre should follow a double exponential shape as described in the equation:

$$f_{i,j} = C \times \left(\frac{1}{2} \exp\left(-\frac{D_{i,j}}{\lambda_i} \right) + \frac{1}{2} R_{mirror} \times \exp\left(-\frac{2L_{fibre} - D_{i,j}}{\lambda_i} \right) \right) \tag{3.7}$$

where $f_{i,j}$ represents the light fraction of the channel i in the cube j, $D_{i,j}$ is the cube j distance to the MPPC i, λ_i attenuation length of the fibre i, L_{fibre} the length of the fibres (92 cm), R_{mirror} the mirror reflectivity and C an arbitrary constant. The first exponential represents the attenuation of half of the light that goes directly to the MPPC, as represented on the figure 2.20. The second exponential is the attenuation of the other half of the light that travels to the reflector at the other end of the fibre before going to the MPPC. The coefficient R_{mirror} was fixed arbitrarily at 0.8 due to difficulties in fitting the attenuation lengths with too many free parameters. The value of 0.8 was chosen after measurements on test bench on several types of mirrors [78]. The same value was used in the simulation of the attenuation effect in order to have an equivalent model in simulation and data reconstructions.

The attenuation pattern was then fitted for every channel of the planes such as presented on figure 3.9 with the data points for one channel and the exponential fits for all channels of a given plane represented. For each channel i, the relative residuals are computed:

$$R_{i,j} = \frac{f_{i,j} - f_i^{attenuation}(j)}{f_{i,j}} \tag{3.8}$$

with $f_i^{attenuation}(j)$ the attenuation model from equation 3.7 fitted on the data for the channel i evaluated on the cube j. Those relative residuals can be seen on figure 3.10 for the four sides of the detector where one can see residuals under control with a mean value below 1% for the four types of channels and a spread of around 5%.

The average value measured for the attenuation length is 97 cm with a spread of around 17%. The attenuation length distribution can be seen on figure 3.11. On this figure, several effects might be the cause of the shape of the distribution. At long attenuation lengths, over 130 cm, the tail of the distribution might be due to batches of optical fibres with a better attenuation length than the others. At lower attenuation length, below 70 cm, there could be a quality selection effect in the manufacturing of the fibres. Only 3128 channels out of 3200 were measured due to 36 dead channels. Since only cubes with 4 active channels were considered, 72 channels were left aside.

Coupling parameter. Once the individual attenuation lengths are measured, the last parameters to be evaluated are the fibre to fibre inhomogeneities inside a plane. This coupling parameter is measured after attenuation correction of the energy deposits inside the cubes. On figure 3.12 can be seen the amplitudes read out by the channels in the four sides of a plane after attenuation correction. The difference between the row/columns amplitudes is defining the coupling parameter. In a given plane, for a given side of the plane (top, bottom, left or right), the average amplitude measured in a channel is divided by the average amplitudes of all the channels on the same side of the plane. The obtained quantity is the coupling parameter:

$$\epsilon_i^{coupling} = \frac{\overline{A}_i}{1/16 \times \sum_{k=1}^{16} \overline{A}_k} \tag{3.9}$$

with $\epsilon_i^{coupling}$ the coupling factor of the channel i, $\overline{A}_i$ the average amplitude of the channel i and $\overline{A}_k$ the average amplitude of the channels in the same side of the plane as the channel i. The distribution of

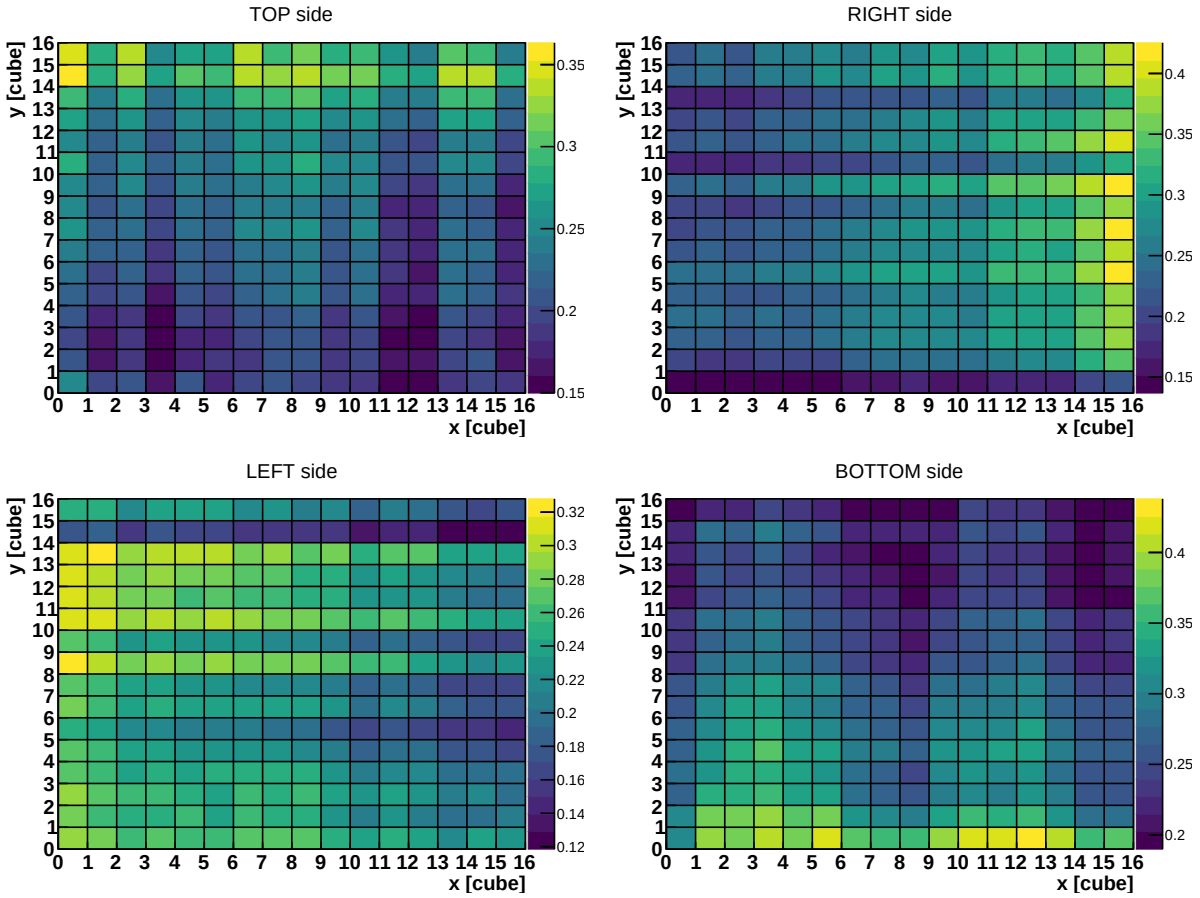


Figure 3.8: Light fraction readout by the channels in the 4 sides of a given plane for a full dataset of ^{22}Na calibration.

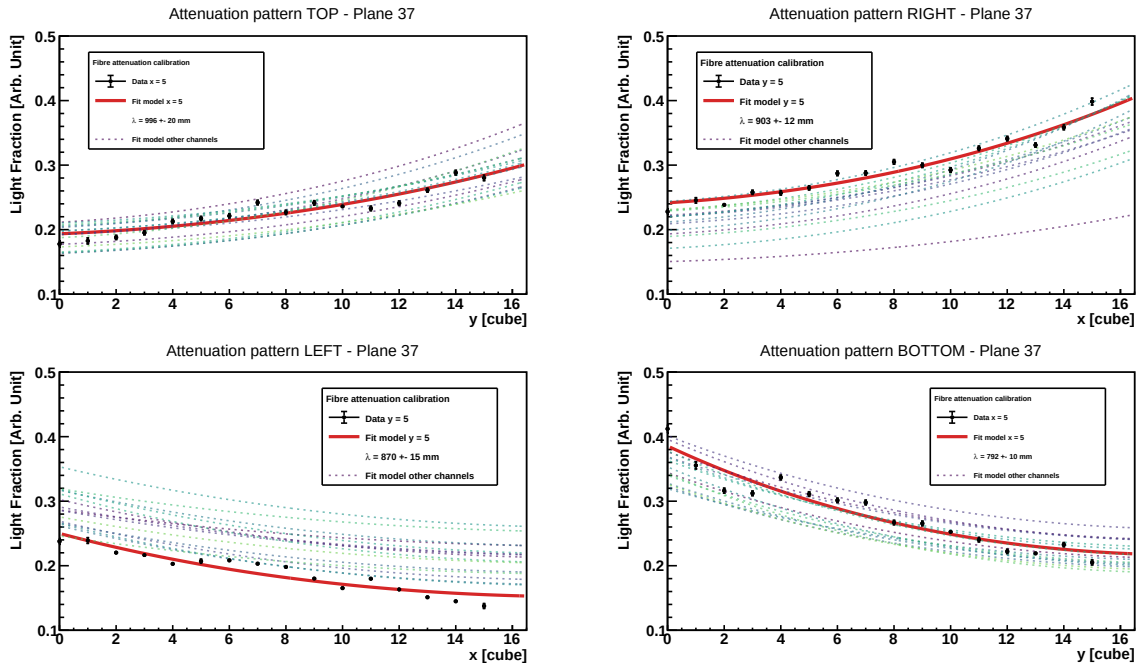


Figure 3.9: Attenuation patterns with the fits for the plane 37 using the equation 3.7. The black points are the data points for a specific channel in the planes in each side. The red line is the attenuation fit for this channel. The dotted lines are the other fits of the plane.

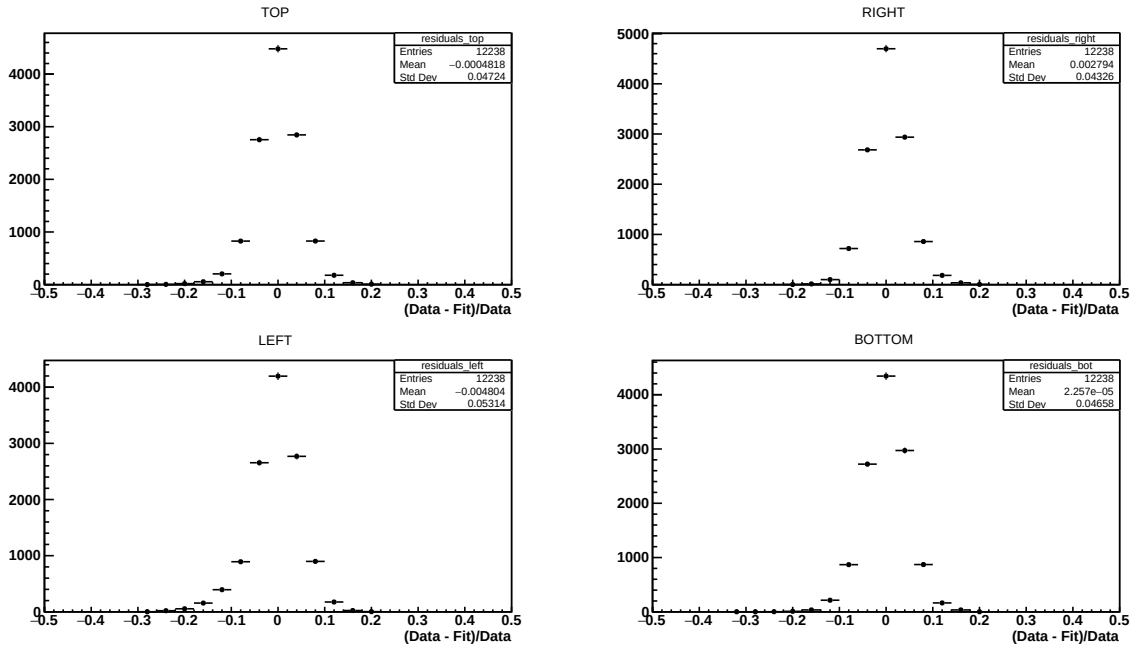


Figure 3.10: Residuals of the attenuation fit for all the active channels. The low value of the residuals ensures a good shape of the exponential fits.

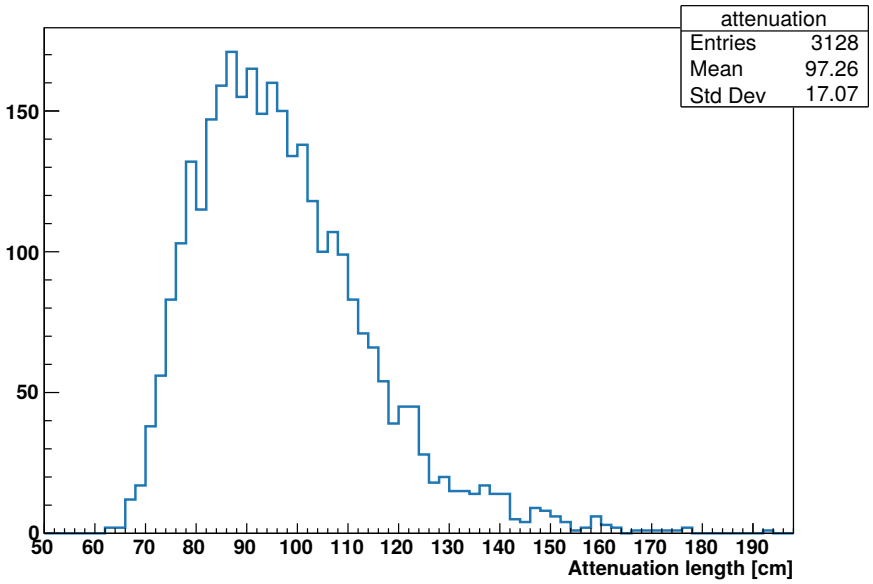


Figure 3.11: Optical fibre Attenuation length distribution for the active channels of the detector.

those coupling parameters is shown in figure 3.13 has a Gaussian shape with a spread of around 15% between the channels.

To assess the goodness of the coupling and attenuation length measurements, ^{22}Na simulations have been produced, and both parameters have been measured for all the channels of one module. The relative difference between the input parameters is represented on the figure 3.14, with a bias of 2.4 % for the attenuation length measurement and 0.1 % on the coupling measurement. The coupling parameter errors follow an expected Gaussian behaviour with a bias under 1 % while a tail can be seen on the attenuation error measurement with a larger bias and spread of the values. To achieve a more precise measurement of the attenuation lengths, another method using a global fit on each plane have been developed by the team at Subatech Nantes.

3.3.2.2 Asymmetry fit

When the attenuation length and coupling are computed for a given fibre, the observables used depend on the other fibres of the plane. One can write the light fraction of a fibre for a given cube as:

$$f_{i,j} = \frac{A_i \times \epsilon_i^{coupling} \times (\epsilon_{i,j}^{att,dir} + \epsilon_{i,j}^{att,ref})}{\sum_{k=1}^{4} A_k \times \epsilon_k^{coupling} \times (\epsilon_{k,j}^{att,dir} + \epsilon_{k,j}^{att,ref})} \tag{3.10}$$

with $f_{i,j}$ the light fraction of the channel i in the cube j, A_i the amplitude of the channel i, $\epsilon_i^{coupling}$ the coupling parameter of the channel i, $\epsilon_{i,j}^{att,dir}$($\epsilon_{i,j}^{att,ref}$) the direct (reflected) attenuation factor of the channel

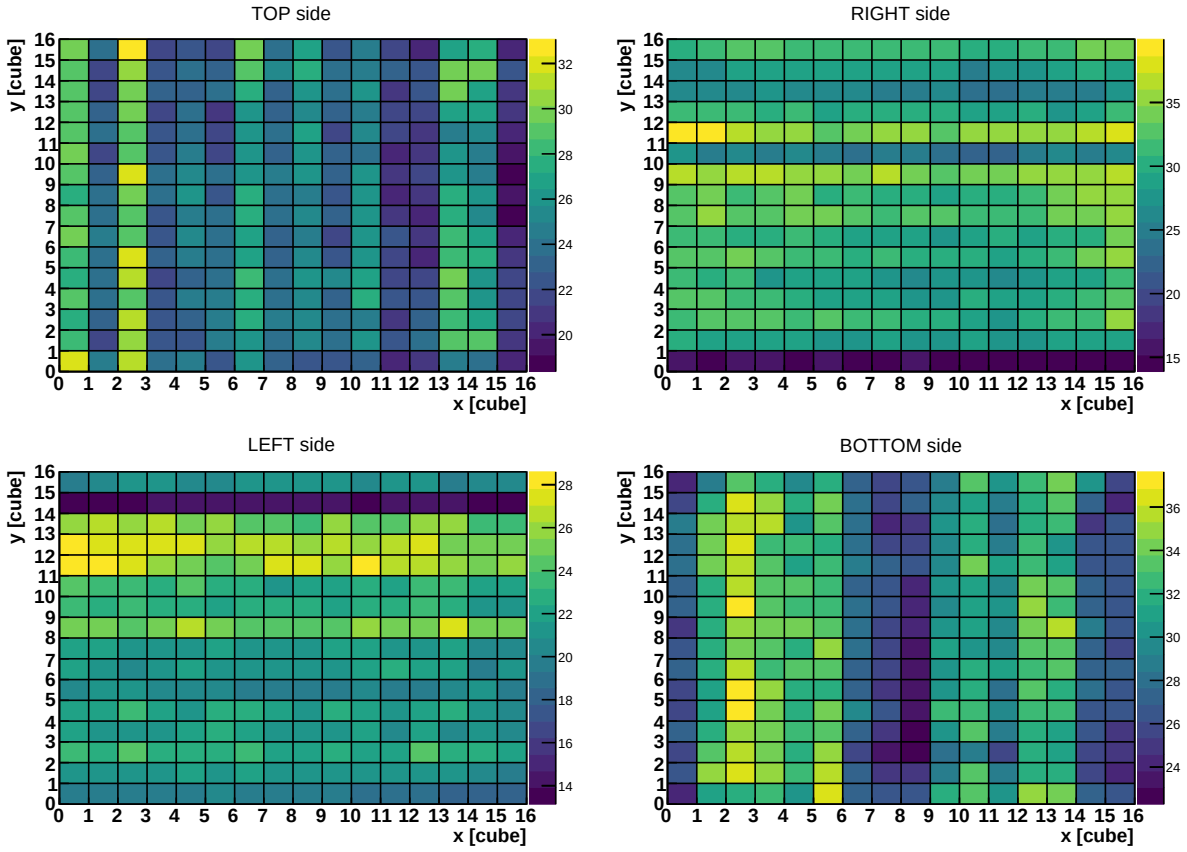


Figure 3.12: Amplitudes measured by the channels in the 4 sides of a plane after attenuation correction.

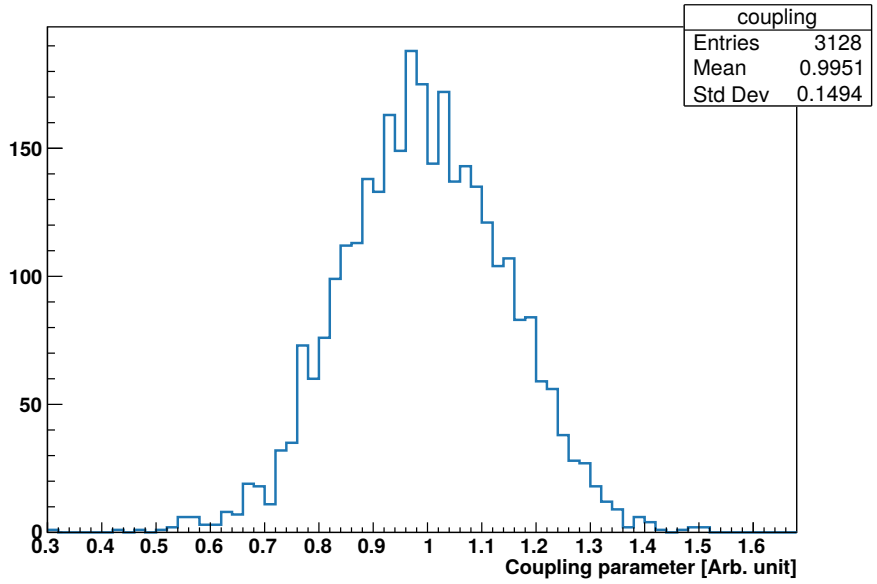


Figure 3.13: Coupling parameter distribution for all the measured channels.

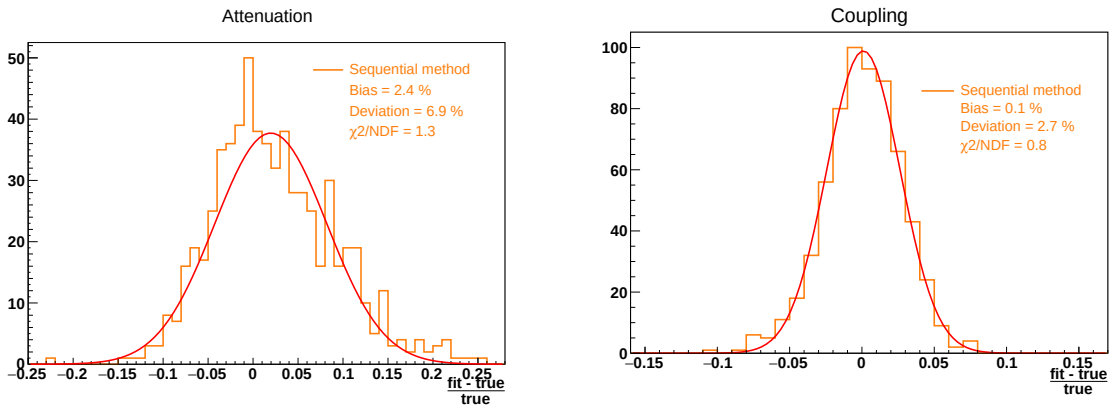


Figure 3.14: Relative difference between the input parameters and the measurements with the sequential for attenuation length (left) and coupling (right).

i in the cube j. Both attenuation factors can be written as:

$$\begin{aligned} \epsilon_{i,j}^{att,dir} &= \frac{1}{2}\exp\left(-\frac{D_{i,j}}{\lambda_i}\right) \\ \epsilon_{i,j}^{att,ref} &= 0.8 \times \frac{1}{2}\exp\left(-\frac{2L_{fibre} - D_{i,j}}{\lambda_i}\right) \end{aligned} \tag{3.11}$$

As defined here, the variable used to measure the attenuation of a channel depends on the coupling and the attenuation of the other channels of the plane and not only on the channel attenuation length as supposed on equation 3.7. Really low couplings on some channels can impact the light fraction of the cubes of other channels and bias the attenuation measurement. To treat those ambiguities, another method was developed to measure the attenuation length and coupling parameter of the channels of a plane simultaneously in a combined fit. This method is based on the asymmetries between the 4 fibres in a cube. Defining for a given cube 3 types of asymmetry (horizontal - vertical, left - right, top - bottom) as described in equation 3.12:

$$\begin{aligned} \mathscr{A}_{H-V} &= \frac{A_L + A_R - (A_T + A_B)}{A_L + A_R + A_T + A_B} \\ \mathscr{A}_{L-R} &= \frac{A_L - A_R}{A_L + A_R} \\ \mathscr{A}_{T-B} &= \frac{A_T - A_B}{A_T + A_B} \end{aligned} \tag{3.12}$$

where $\mathscr{A}_{H-V}$ is the *Horizontal-Vertical* asymmetry, $\mathscr{A}_{L-R}$ is the *Left-Right* asymmetry, $\mathscr{A}_{T-B}$ is the *Top-Bottom* asymmetry, $A_{T,B,L,R}$ is the amplitude of the channel with the MPPC in side top, bottom, left or right. This amplitude can be derived in the same way as the equation 3.2 as:

$$A_{T,B,L,R} = \frac{1}{4} \times E_{cube} \times LY_{cube} \times (\epsilon_{T,B,L,R}^{att,dir} + \epsilon_{T,B,L,R}^{att,ref}) \times \epsilon_{T,B,L,R}^{coup} \tag{3.13}$$

where $\epsilon_{T,B,L,R}^{att,dir(ref)}$ is the attenuation factor for the direct (reflected) light for the fibre in side top, bottom, left or right for the given cube and $\epsilon_{T,B,L,R}^{coup}$ is the coupling factor for the fibre in side top, bottom, left or right, LY_{cube} the light yield of the cube considered and E_{cube} is the energy deposited by the gamma in the cube. Here also, the light yields are cancelled in the definition of the asymmetries. Each attenuation

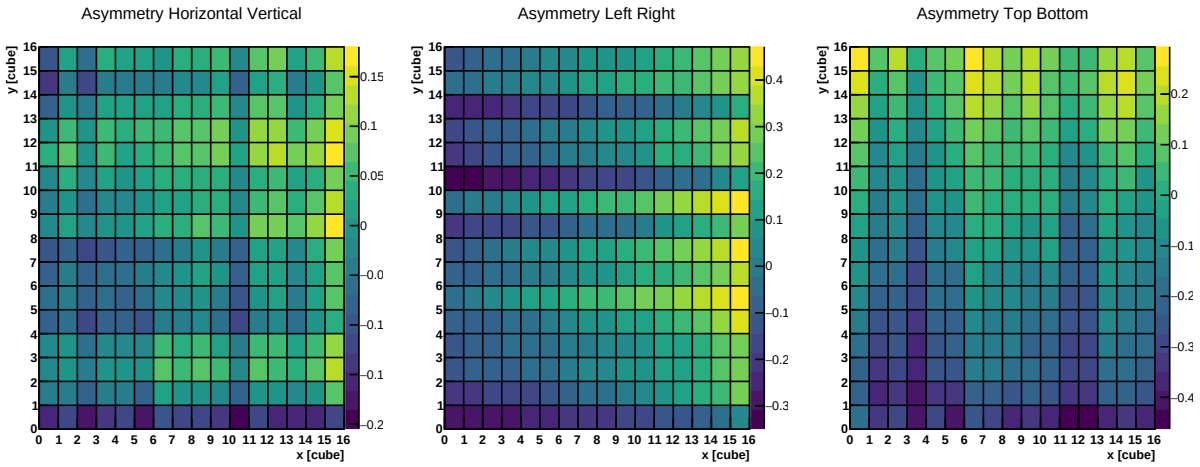


Figure 3.15: Map of the three asymmetries described in 3.12 on a given plane.

factor is defined as in the equation 3.11. An example of the three asymmetries of the cubes of one plane is represented on figure 3.15. On the figure, one can clearly see, especially for the *Left-Right* and *Top-Bottom* asymmetries, the attenuation effect at play.

As each asymmetry is a function of both the fibre attenuation and coupling, the 512 parameters per plane can directly be measured in a global fit defining a global chi square with the three asymmetries:

$$\chi^2_{tot} = \chi^2_{H-V} + \chi^2_{L-R} + \chi^2_{T-B} \tag{3.14}$$

with

$$\chi^2_{type} = \sum_{i=1}^{256} \left(\frac{\mathscr{A}^{fit}_{type}(\vec{\lambda}_{att}, \vec{\epsilon}_{coupling}, i) - \mathscr{A}^{measured}_{type}(i)}{\sigma_i} \right)^2 \tag{3.15}$$

where $type$ defines the type of asymmetry, $\vec{\lambda}_{att}$ contains the attenuation length of the four channels of the cube i, $\vec{\epsilon}_{coupling}$ contains the coupling parameter of the four channels of the cube i, $\mathscr{A}^{fit}_{type}$ is the asymmetry obtained with the fitted model, $\mathscr{A}^{measured}_{type}$ is the measured asymmetry and σ_i is the error on the measured asymmetry. The measurements with the sequential method are used to initialize the fitter due to the large number of free parameters to be fitted. The residuals of those fits in the whole detector are plotted on figure 3.16. The averages of the three residuals are below 3% with a spread of about 25%. This level of agreement is quite good if one considers the large number of parameters and cubes fitted simultaneously.

To choose the most accurate method, a comparison between the fitted and the expected values has been performed on ^{22}Na simulations as shown on figure 3.17. The expected values are defined as the input parameters of the simulations. With the asymmetry method the bias in the measurement of attenuation lengths is better than the sequential method (1.7% compared to 2.4%) with values in a smaller range ($\sigma = 4.2\%$ compared to 6.9%). The measurement on the coupling gives roughly the same values, whatever method considered. Due to the improvement on the parameter measurement, the asymmetry method was chosen as the default fibre calibration method for SoLid calibration. The bias of 0.1 % of the coupling parameter of the channels will have only a low impact on the energy measurement as it induces at most a 0.4 % bias combining the biases on the four channels. Considering the energy resolution of the

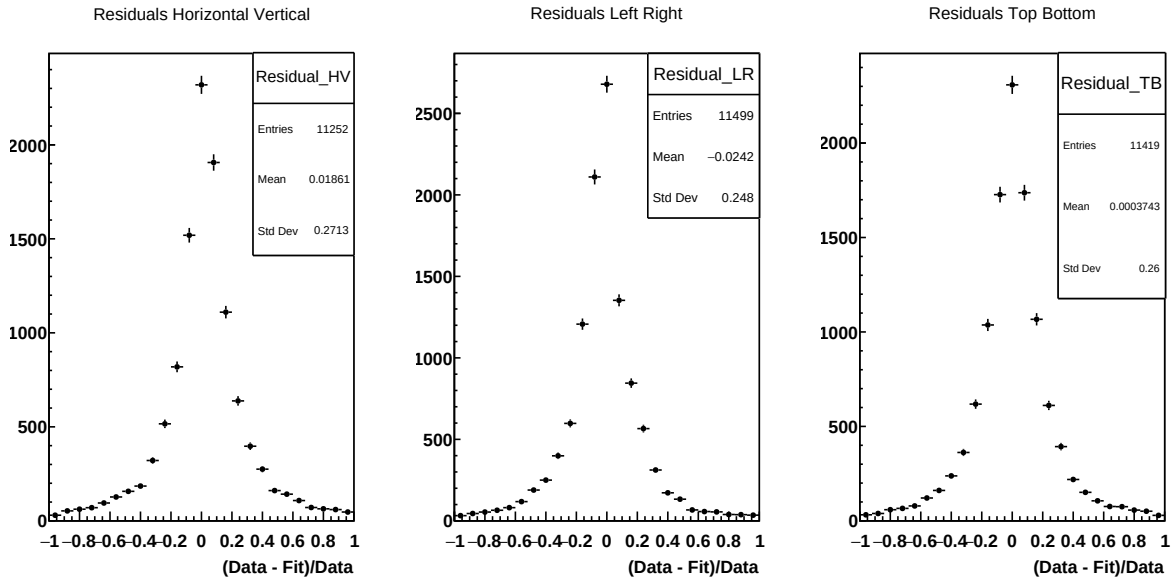


Figure 3.16: Relative residuals of the asymmetry fit for all the active cubes of the detector. The averages of the residuals is below 3% with a spread around 25 %.

cube energies of about 16 % at 1 MeV (see section 3.4.7) and the precision in the energy reconstruction that would be required to see the effect a sub 1 % bias, the precision of the coupling parameter measurement is good enough for the SoLid experiment. Furthermore, in the same way as the gain and pedestal inhomogeneities, the coupling variations are corrected prior to the cube reconstruction. That way, a difference of coupling from one channel to another does not induce a reconstruction of more or less cubes due to a channel reconstruction energy threshold. The effect of the attenuation length measurement bias is more difficult to quantify due to the effect of the attenuation that both depends on the attenuation length of each fibre and the position of the cube being reconstructed. As it becomes a cube effect that will be mixed with the light yield measurements, the data to Monte-Carlo comparisons on the cube energies presented on 3.5 will be used to quantify the precision of the energy calibration.

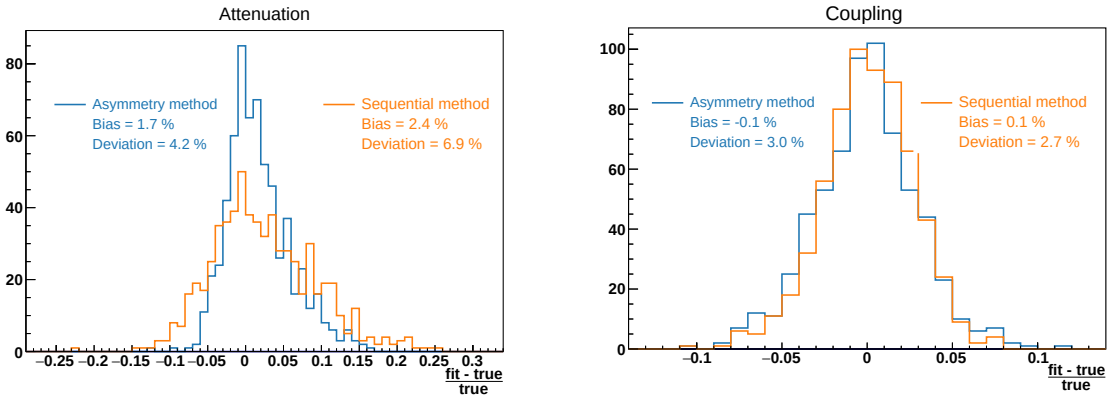


Figure 3.17: Comparison input vs measurement for attenuation length (left) and coupling (right).

3.4 Cube energy calibration

Due to the small size of the detection cells (5x5x5 cm^3) and the use of light elements (PVT) the probability of a photon of around $1\ \mathrm{MeV}$ to deposit all its energy in only one cube is small. After a Compton scattering, the gamma will usually leave the cube in which it interacted. Due to this specificity of the detector, Compton edges are used as energy calibration reference point. Two methods have been developed for this purpose: a first one based on Klein Nishina cross-section for Compton interaction [99]. The second method is based on data/Monte-Carlo comparison with a numerical convolution of the true Monte-Carlo energy deposits. Both methods will be presented in the next sections. In this section and the following ones, for reasons detailed in section 3.4.8.2, the channels are corrected of the gain, pedestal and coupling variation. If not indicated otherwise, the channel attenuation is not corrected before computing the light yields. In this section and the following ones, except indicated otherwise, the energy reconstruction used is the CCube reconstruction presented in section 2.7.2.3.

3.4.1 Klein Nishina analytical fit

This method was developed by the colleagues of Subatech Nantes and more details on the fit can be seen in David Henaff's thesis [100]. The method was then used for the results showed in this thesis without further developments.

In the cases where the gamma will leave a fraction of its energy in a cube, leaving the cube after the interaction without any more energy deposit, the cross-section of the interaction can be derived by the Klein Nishina formula:

$$\frac{d\sigma}{dT} = \frac{\pi r_e^2}{m_e c^2 \alpha^2}\left(2 + \left(\frac{T}{E_0 - T}\right)^2\left(\frac{1}{\alpha^2} + \frac{E_0 - T}{E_0} - \frac{2}{\alpha}\left(\frac{E_0 - T}{T}\right)\right)\right) \tag{3.16}$$

where r_e is the classical radius of the electron, m_e is the electron's mass, α is the fine structure constant, E_0 is the initial energy of the gamma and T is the electron energy after scattering. The cross-section being non-null for $T \in [0, E_c]$ with E_c the Compton edge energy:

$$E_c = E_0\left(1 - \frac{1}{1 + \frac{2E_0}{m_e c^2}}\right) \tag{3.17}$$

For the $1275\ \mathrm{keV}$ of the ^{22}Na, it corresponds to a calibration energy of $1062\ \mathrm{keV}$. A numerical convolution is then applied to the model to introduce the detector energy resolution. The p.d.f[1] is obtained the following way:

$$f_{conv}(x) = \sum_{i=0}^{i_{E_c}} \frac{d\sigma}{dT}(T_i)\frac{1}{\sqrt{2\pi}\sigma_0\sqrt{T_i}}\exp\left(-\frac{(\frac{x}{LY} - T_i)^2}{2\sigma_0^2 T_i}\right) \tag{3.18}$$

where x is the amplitude in PA, and where the two parameters to fit are the light yield, LY and the energy resolution, σ_0. This p.d.f has then to be normalized:

[1] Probability density function

$$f_{conv}(x) = \frac{\sum_{i=0}^{i_{E_c}} \frac{d\sigma}{dT}(T_i) \frac{1}{\sqrt{2\pi}\sigma_0\sqrt{T_i}} \exp\left(-\frac{(\frac{x}{LY}-T_i)^2}{2\sigma_0^2 T_i}\right)}{\sum_{i=0}^{i_{E_c}} \frac{d\sigma}{dT}(T_i)} \tag{3.19}$$

For this method, it is essential to have a precise reconstruction of the energy without pile-up effects. Even if the CCube algorithm is designed to reduce at most the effect of pile-up in channels, an isolation selection is applied to the data: only the events where only the cube being calibrated has an energy deposit in its plane are selected. To have this isolation selection is the best way to be sure that the data will be described by the density function. Furthermore, another selection is made on the four channels of the cube, they should all have an energy above 2.5 PA, in order to avoid energy losses due to one channel being inactive or low-energy inefficiencies.

Some corrections due to reconstruction inefficiencies have then to be taken into account. A GEANT4 simulation is used to compute the efficiency correction. ^{22}Na events are generated with the *SoLidSim* software that simulates all the energy deposits inside the detector, reproducing the calibration conditions. The detector effects are then applied to the events with the readout simulation, and the events are then reconstructed the same way as real data. For each cube, the ratio between the true GEANT4 energy spectrum of the events being reconstructed and selected and the true GEANT4 energy spectrum of all the events is computed to correct the p.d.f as shown on figure 3.18 and treated as the correction factor, ϵ_{reco}.

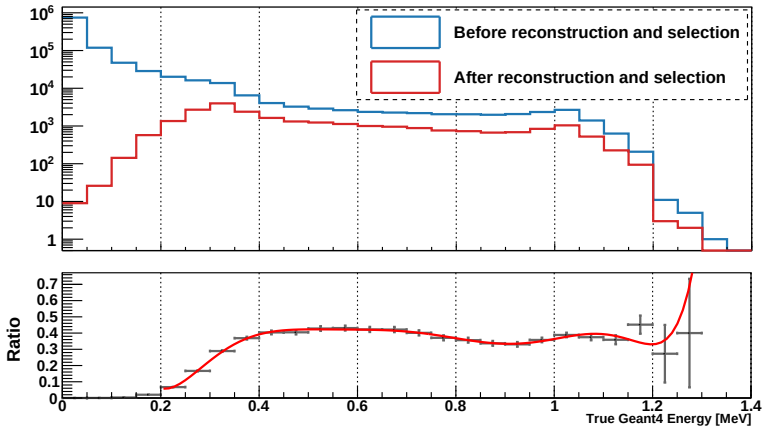


Figure 3.18: Top: GEANT4 energy distribution of the gamma energy deposits in the ^{22}Na events. In blue is represented the distribution of all the events and in red the energy distribution of only the events that passed where reconstructed and passed the selection. Bottom: Ratio between the two energy distributions. The ratio is fitted and stored to create calibration templates.

Before interacting in the cube of interest, the gamma emitted by the source can deposit some energy in the PVT, therefore, another correction needs to be computed, to do so, in GEANT4 simulations, a ratio between the energy spectrum of the cube in front of the source and the energy spectrum of the cube of interest is computed. The cube in front of the source is supposed to receive gammas without energy losses and is thus used as reference. The ratio between the two energy spectra is used as a second correction factor ϵ_{loss}. An example of this ratio can be seen on figure 3.19 for the cube (8 8 34). The effect of the gamma attenuation can be seen in the slope of the ratio between 0.6 and 1.2 MeV: gammas are losing some energies before interacting in the further cubes and the Compton edge is thus less populated. The same effect can be seen around 0.3 MeV, with the attenuation of the annihilation gammas resulting of

the annihilation of the positron emitted in the ^{22}Na decay. The final p.d.f is then :

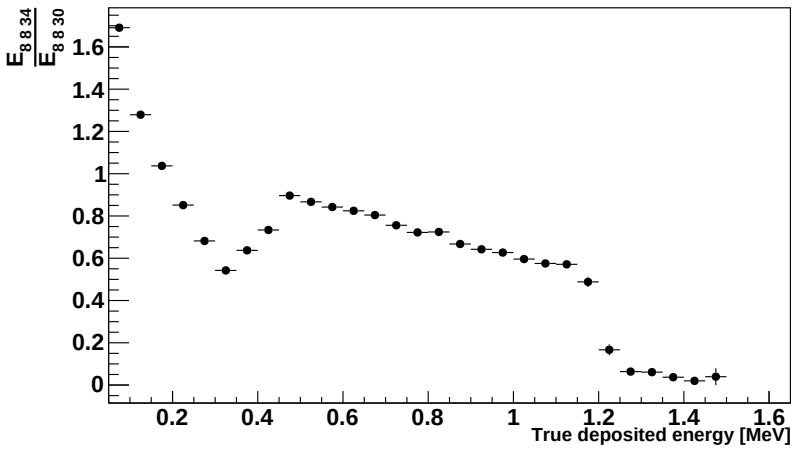


Figure 3.19: Ratio of the true energy deposit in the cube in front of the source (8 8 30) and the cube of interest. This ratio is used to compute the factor ϵ_{loss} per energy bins.

$$f_{conv}(x) = \frac{\sum_{i=0}^{i_{E_c}} \epsilon_{reco}(T)\epsilon_{loss}(T)\frac{d\sigma}{dT}(T_i)\frac{1}{\sqrt{2\pi}\sigma_0\sqrt{T_i}}\exp\left(-\frac{(\frac{x}{LY}-T_i)^2}{2\sigma_0^2 T_i}\right)}{\sum_{i=0}^{i_{E_c}} \epsilon_{reco}(T)\epsilon_{loss}(T)\frac{d\sigma}{dT}(T_i)} \tag{3.20}$$

The example of the fit of one cube can be seen on the figure 3.20. The results on all the cubes will be presented with the comparison between the two methods in section 3.4.3.

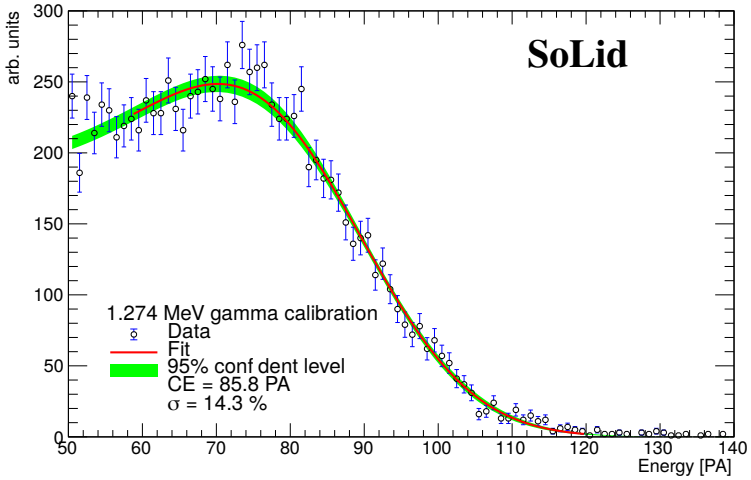


Figure 3.20: Analytical fit for a given cube. The data points are calibration data and the red curves represents the fit of the data with the p.d.f described in the equation 3.20

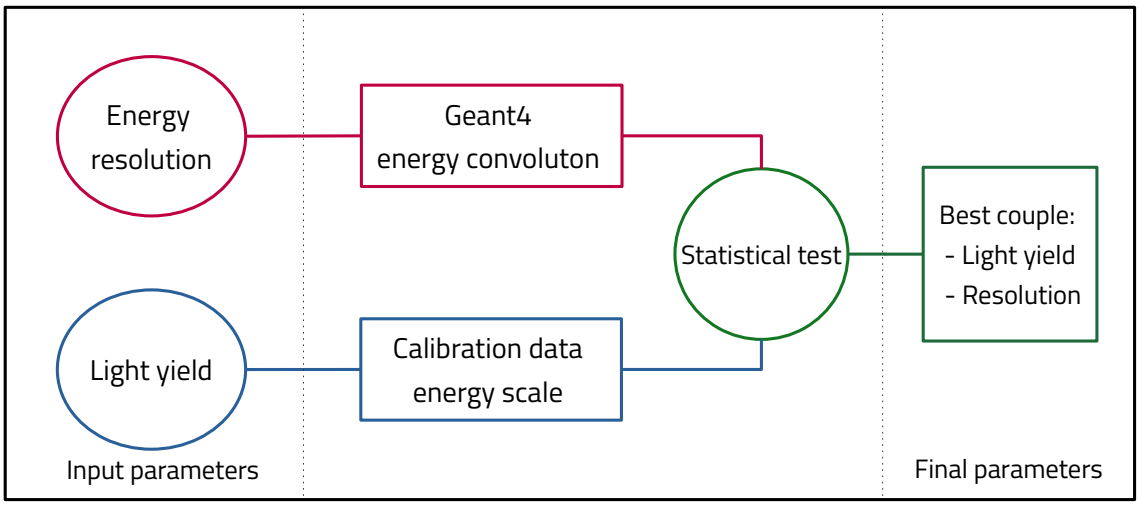


Figure 3.21: Scheme of the numerical method: a loop on the light yield and energy resolution candidates is performed with a statistical test to select the best couple.

3.4.2 Numerical method

A second method of cube energy calibration has been designed based only on Monte-Carlo simulations. This method will be called the numerical method. For a given cube, the true energy spectrum from GEANT4-based Monte-Carlo simulation is numerically convoluted with a given energy resolution to mimic the detector resolution. A global scheme of the method can be seen on figure 3.21

A reconstruction efficiency is applied then, the same way as computed in section 3.4.1.

$$S_{conv}(E,\sigma) = \sum_{i=0}^{i_{E_c}} S_{true}(E_i)\epsilon_{reco}(E_i)\frac{1}{\sqrt{2\pi}\sigma} \times \exp\left(-\frac{E-E_i}{2\sigma^2}\right) \tag{3.21}$$

with S_{conv} the convoluted energy spectrum, S_{true} the GEANT4 energy spectrum, ϵ_{reco} the reconstruction efficiency, σ the considered energy resolution and E the energy in MeV. The energy resolution is assumed to be purely statistical:

$$\sigma = \sigma_0\sqrt{E} \tag{3.22}$$

The effect of the convolution can be seen on figure 3.22. A simulation template is created for each cube and a range of energy resolution between 10 and 25 $\%$ at 1 MeV. In the case of the numerical method, the first event selection chosen is to only consider cubes with four active fibres in order to avoid energy losses that would not be taken into account, with ϵ_{reco} computed with this selection. In the next section, an isolation selection cut will be added to the event selection and ϵ_{reco} will be computed again according to the new selection. Contrary to the analytical fit method, there is no need to correct the gamma energy attenuation before interacting in the cube of interest, as it is a physical effect already taken into account in the GEANT4 simulation and in $S_{true}(E_i)$.

The energy deposits of the calibration data are then scaled with a given scaling factor, light yield in PA/MeV. The best couple (light yield, energy resolution) is chosen with a test of agreement between the

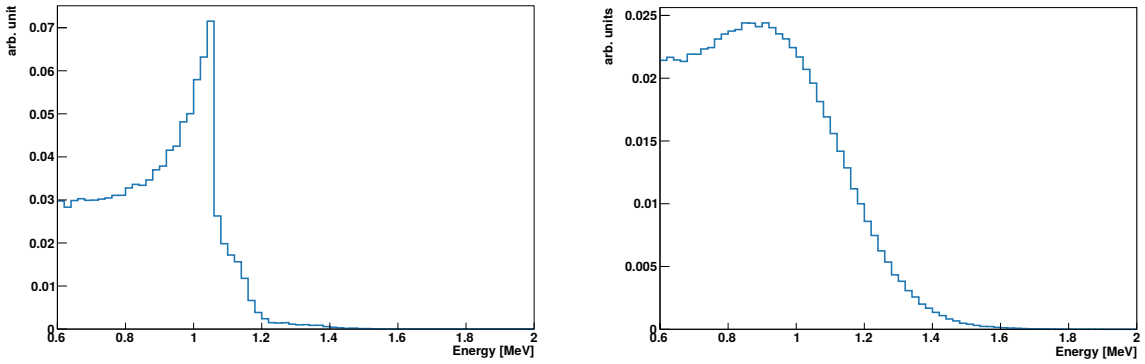


Figure 3.22: Left: true GEANT4 energy distribution for a given cube. Right: convoluted spectrum with a given energy resolution. For each cube the true GEANT4 energy distribution is numerically convoluted with a given Gaussian energy resolution. A distribution template is created for each resolution tested.

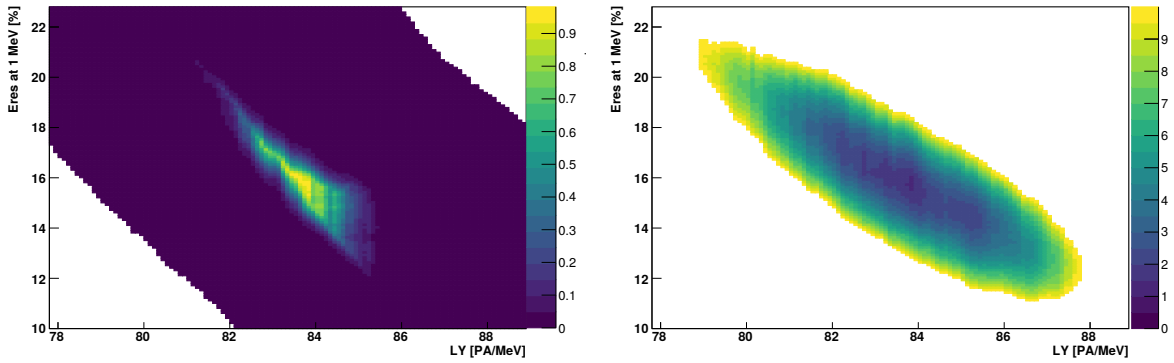


Figure 3.23: Left: Kolmogorov Smirnov scores as a function of the resolution and light yield tested for a given cube. Right: $\Delta\chi^2$ value as a function of the resolution and light yield tested for a given cube. The best couple will be the one that maximises the Kolmogorov-Smirnov test or minimizes the χ^2 test.

scaled data and the convoluted simulation, here a Kolmogorov-Smirnov (K-S) test or a χ^2 test. On figure 3.23 are represented the scans (light yield, energy resolution) performed for one cube with on the left the K-S test and on the right a χ^2 test score. The best couple is the one that maximizes the Kolmogorov-Smirnov test or minimizes the χ^2 test, depending on the test that has been chosen. An example of the fit result for a cube can be seen on figure 3.24. More details on the fits on all the cubes will be presented in the next sections.

To compute the error on the parameters, for the K-S test, the results of the statistical test are projected on each parameter space (light yield or resolution), selecting for the parameter the best K-S score on figure 3.25 left and selecting the light yield variations obtained at half height of the parabola. A more rigorous way to compute the errors is to do a $\Delta\chi^2$ analysis. On figure 3.25 right is represented the $\Delta\chi^2$ distribution of the light yield tests of a given cube. For this cube, the error could be estimated at 2% at one σ with a $\Delta\chi^2$ method.

Historically, the method was derived with the Kolmogorov-Smirnov test only because it is expected to perform well even at low statistics. With the improvement of the calibration data taking with more

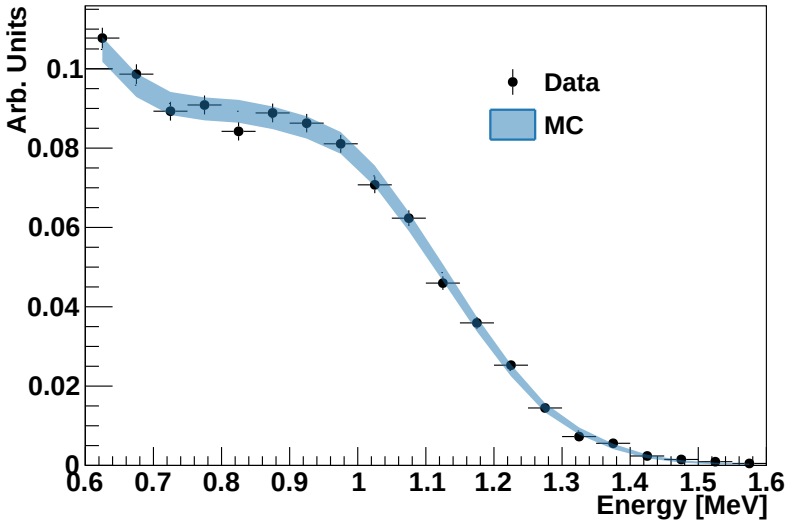


Figure 3.24: Fit for a given cube for ^{22}Na. The blue envelope represents the best convoluted GEANT4 with a 1 σ statistical uncertainty on the convoluted Monte-Carlo and the black points represent the data point from a calibration campaign.

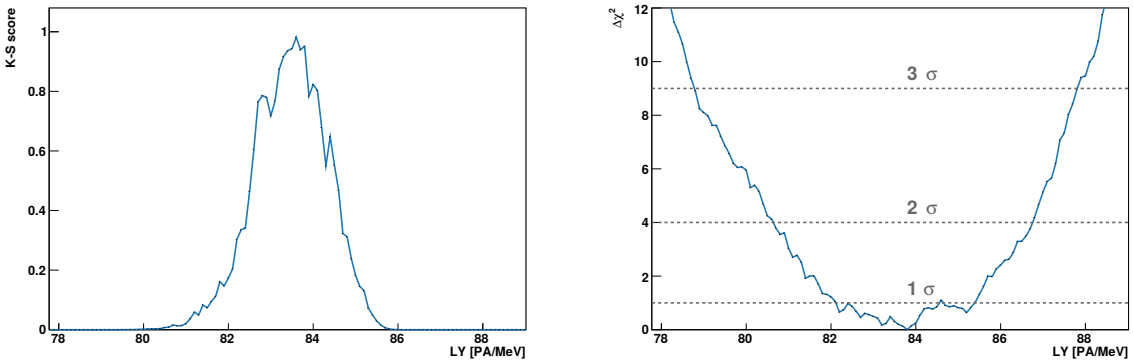


Figure 3.25: Left: Kolmogorov Smirnov test score as a function of the light yield tested for a given cube. Right $\Delta\chi^2$ score as a function of the light yield tested for a given cube. Each point corresponds to the light yield that maximises the likelihood for a given energy resolution.

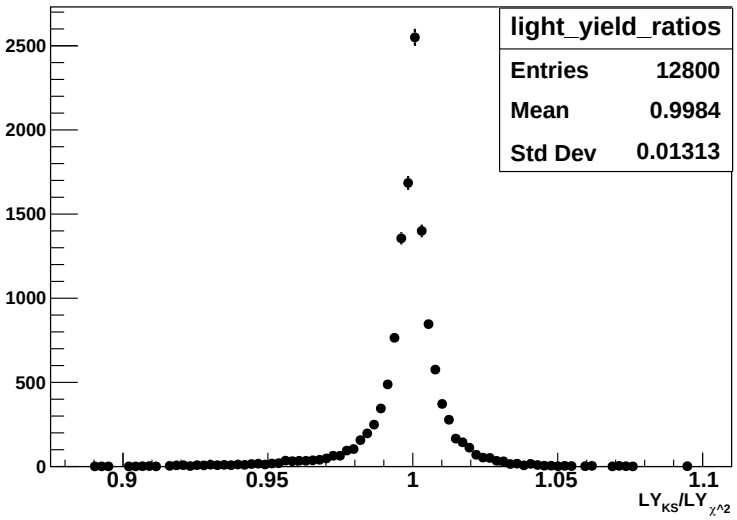


Figure 3.26: Ratio between the light yields computed with the Kolmogorov-Smirnov test and the χ^2 test.

data in less time, a χ^2 test was then added to the procedure and used as the main test in order to have a better handle on the measurement uncertainties. On figure 3.26 is represented the light yield ratios between the two methods for the whole detector. The difference between the two methods being lower than 1 % the use of only the χ^2 test in the calibration was decided.

3.4.3 Comparison between the analytical and numerical methods

3.4.3.1 On calibration data

To compare the two calibration methods, the first step is to use calibration data. To do so, the September 2018 ^{22}Na calibration campaign has been used. All cubes with four active channels have been fitted with both methods, and the relative difference between the fitted light yields can be seen on the figure 3.28. A first comparison was made using no other selection for the numerical method and an isolation in the plane for the cube of interest for the analytical method. An effect can be seen on the planes in front of the sources, with a 4.5 % bias between both methods, and a 2 % bias on the other planes as can be seen on figure 3.27. This effect could be due to some pile up on the fibres occurring in the reconstruction and light leakages with neighbouring cubes depositing some energy in the cube of interest.

To try to have a better agreement between both methods, the isolation selection used previously only on the analytical fit was also applied on the numerical method. The comparison after this new selection can be seen on figure 3.28. A bias lower than 1 % with a spread of 2 % can be observed on both categories of planes (close or further from the source). This result shows that the isolation selection shifted the fitted light yield of the numerical method closer to the analytical one. If the effect was especially expected in the planes in front of the sources, the comparison of the light yields on the other planes is also improved. This result that shows a good agreement between both methods is essential to prove the control of the energy response of the detector, and having two different methods agreeing at that level shows the strength of the calibration work. In the following sections, the isolation selection will be applied on ^{22}Na for the numerical method.

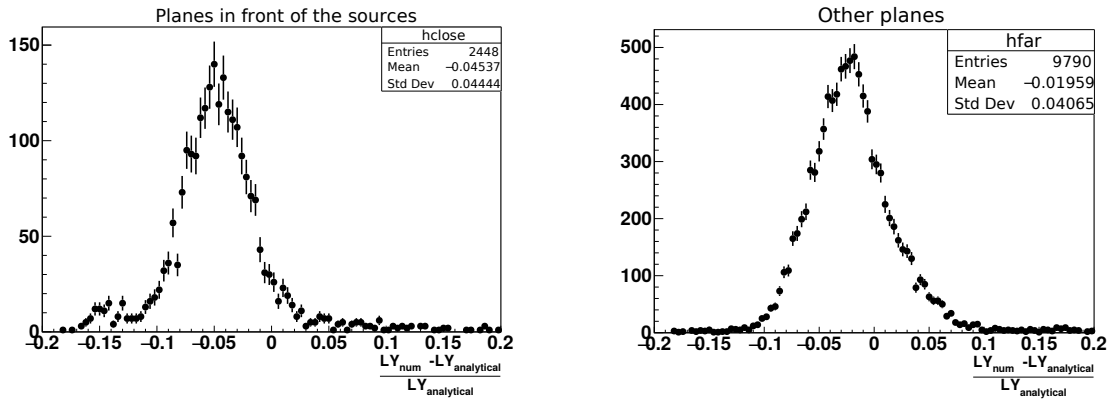


Figure 3.27: Relative difference between the numerical method and the analytical fit with an isolation selection applied only on the analytical fit. A bias of 4.5 % is seen on the planes close from the source and a 2 % bias on the other planes.

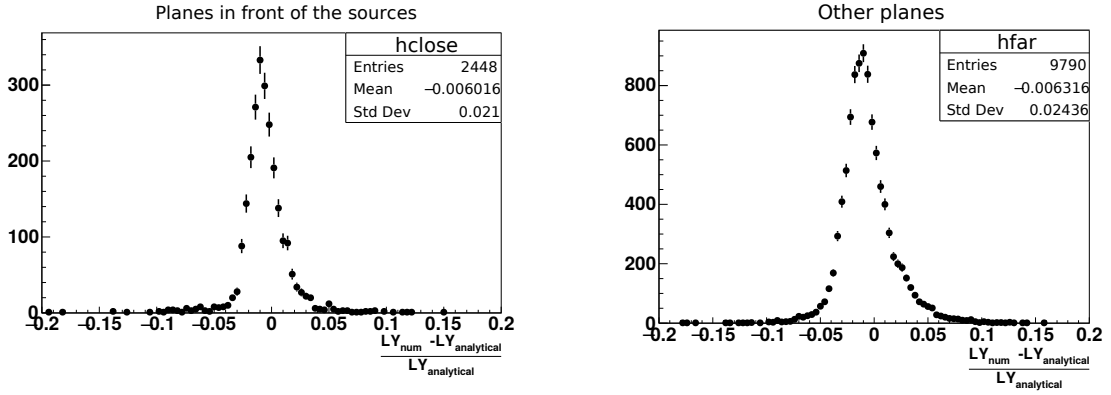


Figure 3.28: Relative difference between both fit methods, with an isolation selection applied in both methods. The bias between both methods is lower than 1 % for the planes close and further from the source.

3.4.3.2 On Monte-Carlo simulation

To study possible biases of both methods, ^{22}Na calibration data have been simulated in five different planes and reconstructed with the official reconstruction software in the same condition as real data. At the same time, 1 MeV electrons have been simulated and reconstructed in the same planes, with the same conditions. The aim is to use the electron energy deposits to evaluate an expected *true* light yield, to compare with the *fitted* one of both methods. The distribution of the relative difference between both types of light yield is presented on the figure 3.29. A bias under 1 % is obtained for both methods, with a standard deviation of around 1 %. This last result confirms that both methods, give a precise and robust measurement of the light yield of each cube in the detector.

3.4.4 Energy response linearity

One of the key characteristics of the PVT scintillation is its good energy response linearity (by opposition to liquid scintillators). This feature can be measured with a multi-sources calibration campaign with

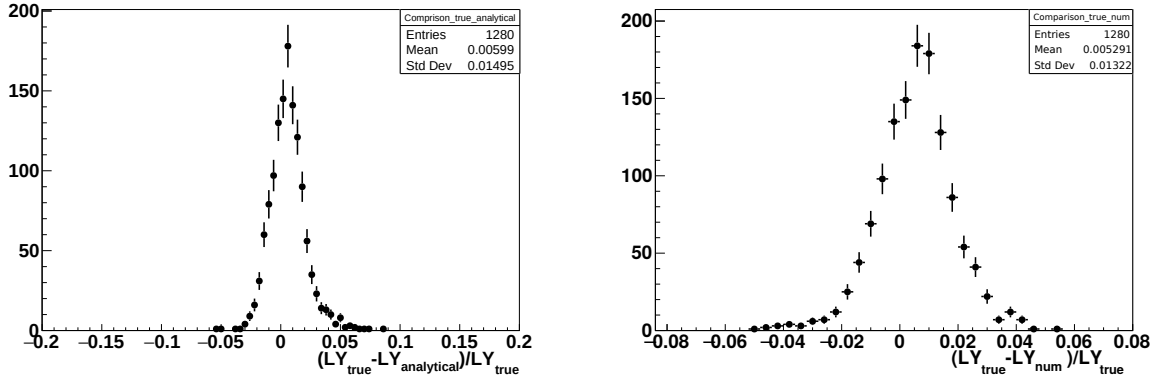


Figure 3.29: Relative difference between expected light yield and fitted light yield with the analytical (numerical) method on the left (right). In both cases, a deviation under 1% is found.

^{22}Na, ^{207}Bi and AmBe. The three gamma sources have been presented in the section 3.2 and the gammas used in this study are the 569 keV, 1063 keV and 1770 keV gammas from the ^{207}Bi source, the 1275 keV gamma from the ^{22}Na source and the 4438 keV gamma from the AmBe source. As the Compton edge of each gamma is used as the calibration point, 5 calibration energies can be used for a linearity study between 0.5 and 4.2 MeV.

To avoid a bias on the energy measurement of low-energy deposits, a periodic trigger has been used for ^{207}Bi and ^{22}Na gamma energy calibrations and a threshold trigger was used for the higher energies of the AmBe to gather more statistics. Due to time limitation during the linearity calibration campaign and to optimize the statistics for each calibration point, the sources were only put at 2 positions: at the middle and at the top corner, electronic box side, of Gap 5. A total of 239 cubes, ~ 2 % of the detector, was tested. For the different calibration sources, different selection criteria were applied to measure the Compton edges:

- For the ^{22}Na and ^{207}Bi, an isolation criterium was applied. Only events where no other cubes in the plane of interest were selected in order to remove the pile up effects along the fibres and the contamination from light leakages of neighbouring cubes. An example of the isolation selection is shown in figure 3.30 for the Compton edge of the 569 keV gamma from ^{207}Bi.

- For the AmBe, as the emission of a neutron is expected at the same time as the gamma, a time correlation selection is performed to consider only events with a neutron signal in the next 500 µs. In order to avoid the contamination of proton recoil induced by the neutron emission, only events where the neutron signal and the electromagnetic signals are in different sides of the source are selected. Therefore, if the source is between the planes 39 and 40 and a neutron is detected before plane 39, the gamma must be tagged in the planes after plane 40. In the case of ~ 4 MeV energy deposits, there is systematically light leakages from the cube of interest to neighbouring cubes, the isolation cut is thus removed for the event selection. With this level of energy, with only one gamma emitted, and a neutron tagged in the other module, one can expect a lower effect of the pile-up on the cube energy.

Each cube light yield was measured for the different gammas with the numerical method presented in 3.4.2 with an example of the fits for the ^{207}Bi and AmBe represented on figure 3.31. Those values were

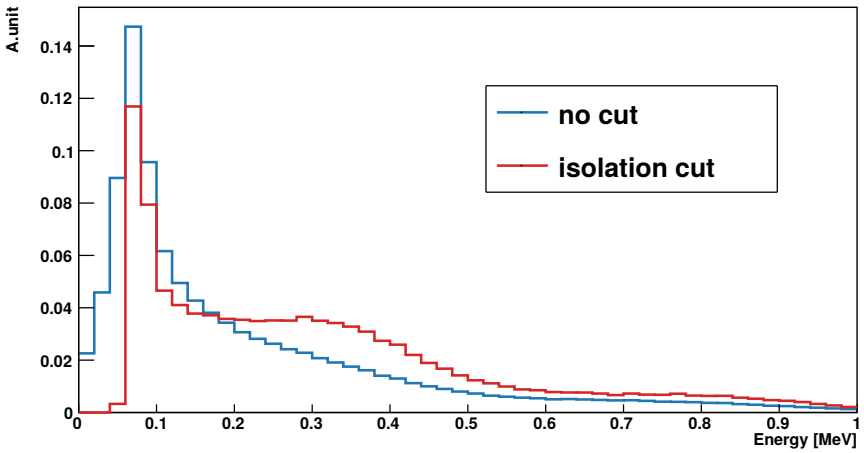


Figure 3.30: Energy distribution for simulated ^{207}Bi around the 569 keV gamma with and without the isolation selection. When an isolation selection is applied (red curve), the cube of interest is not contaminated by pile up, light leakage or Compton scattering in the same plane due to other gammas and the Compton edge becomes more visible.

then averaged over the cubes per calibration gamma to get an averaged linearity measurement on the tested cubes as shown on figure 3.32. The fitted light yields are in agreement within 2 % which confirms the good linearity of the energy response of the PVT. The light yield obtained was also fitted for all the cubes and the slope distribution is represented on the same figure where a spread of about 5% of the light yield is obtained which is coherent with the previous light yield measurements.

3.4.5 Light yield homogeneity

After a measurement of the individual visible cube light yields, the homogeneity of the detector has been measured with a spread of 7% of light yields, without fibre attenuation correction. On figure 3.33 (left) can be seen a map of the light yield in the planes, averaged on all planes of the detector. As the attenuation effect is not corrected in the official reconstruction of SoLid and used only in the simulation, the effect of the attenuation is taken into account in the light yield constants, hence the attenuation pattern with lower light yield in the centre of the detector. On the figure 3.33 (right) is plotted the distribution of the cube light yields in the detector, the average light yield is 96.3 PA/MeV. For all the light yields measurements, the 20 % cross talk of the MPPCs described in the section 2.2.2.1 are not subtracted. The cross talk is directly added inside the simulations to reproduce the data.

Variation of light yield can be observed from one plane to another one with most of them within 5 % as represented on figure 3.34 (top) is represented the light yield variation per plane. Those variations are corrected in the reconstruction and reproduced in the simulations. For the current antineutrino analyses of SoLid, planes are gathered in modules for antineutrino distance travelled measurement. The average light yield difference is at most 5% between the modules, as can be seen on figure 3.34 (bottom), which shows a good homogeneity of the detector on all its baseline. As the zero suppression threshold for the calibration data selects signals around 100 keV and the current analysis threshold for the antineutrino analysis selects events above 1.5 MeV, this module to module light yield variation bellow 5 % will not

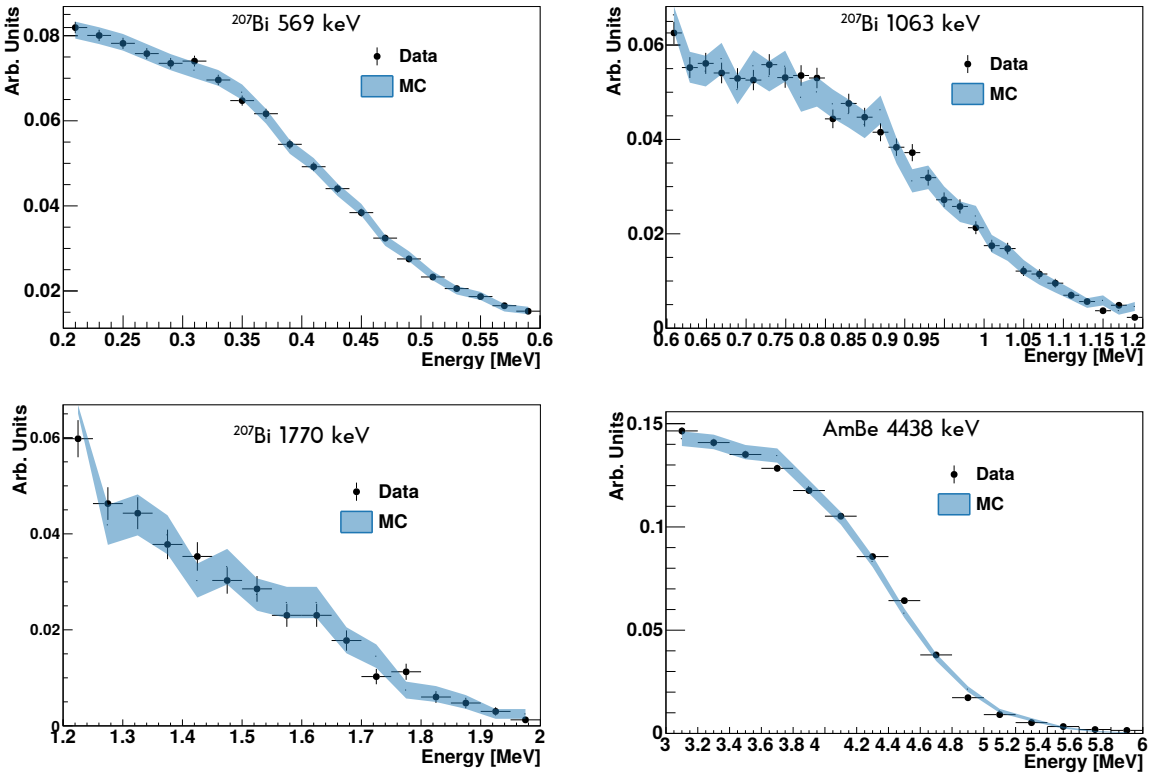


Figure 3.31: Fit for a given cube for ^{207}Bi (top, and bottom left), and AmBe (bottom right). The blue envelope represents the best convoluted GEANT4 with a 1 σ statistical uncertainty on the Monte-Carlo.

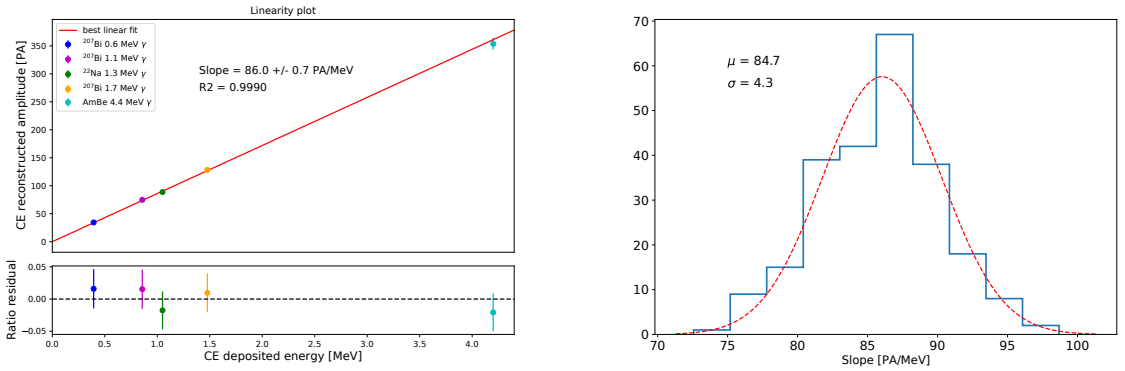


Figure 3.32: Left: average linearity measured on Gap 5 with 2 source positions. Right: slope of the individual cube linearity distribution.

induce difference in data rate from one module to another after analysis selection.

To ensure that all possible calibration effects in each plane are taken into account in the calibration procedure, one can measure the intrinsic light yield variations within each plane of the detector, an attenuation and coupling correction is applied to each cube light yield by averaging all light loss corrections of the four channels composing the cube:

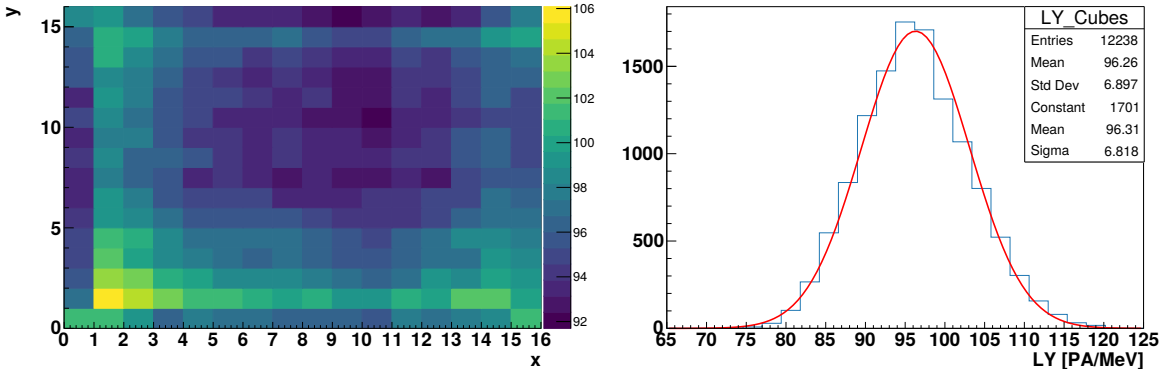


Figure 3.33: Left: light yield by cube positions, averaged in z position. Right: Light yield distribution on the cubes with 4 active channels.

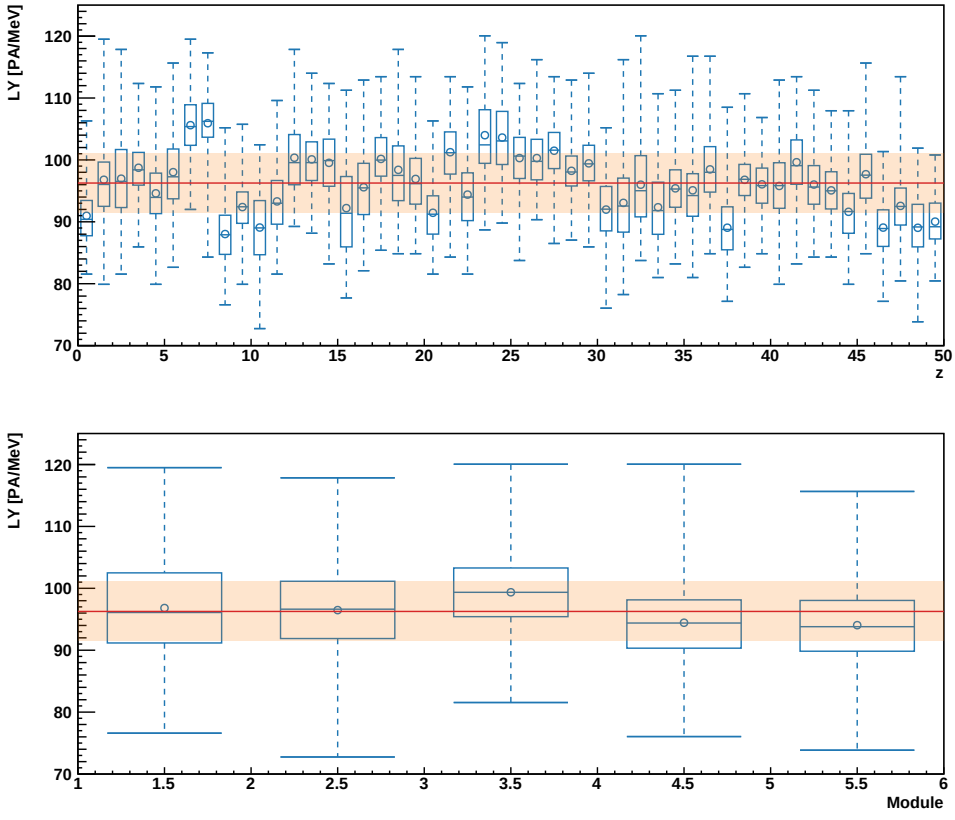


Figure 3.34: Top: light yield dispersion per plane. Bottom light yield dispersion per module. The red lines represent the average light yield of the detector and the orange bands a 5 % variation.

$$LY_j^{intrisic} = \frac{1}{4} LY_j^{measured} \sum_{i \in \text{channels}} 1/(\epsilon_i^{coupling} \times (\epsilon_{i,j}^{att,dir} + \epsilon_{i,j}^{att,ref})) \tag{3.23}$$

with $LY^{measured}$ the light yield measured in the cube j with the method presented on section 3.4.2 without other correction on the data than the gain and pedestal variations, the index i representing the four channels of a cube, $\epsilon_i^{coupling}$ the coupling parameter of the channel i, and $\epsilon_{i,j}^{att,dir}$ ($\epsilon_{i,j}^{att,ref}$) the direct (reflected) attenuation factor of the channel i in the cube j. The intrinsic light yield dispersion within each plane can then be computed as the ratio between the corrected light yield of each cube and the average light yield of its plane. The dispersion obtained in all the detector is 3% which shows a very good homogeneity of the energy response within a plane. This 3 % effect will be treated in a systematic uncertainty study for the antineutrino analysis. A distribution of these variations can be seen on figure 3.35.

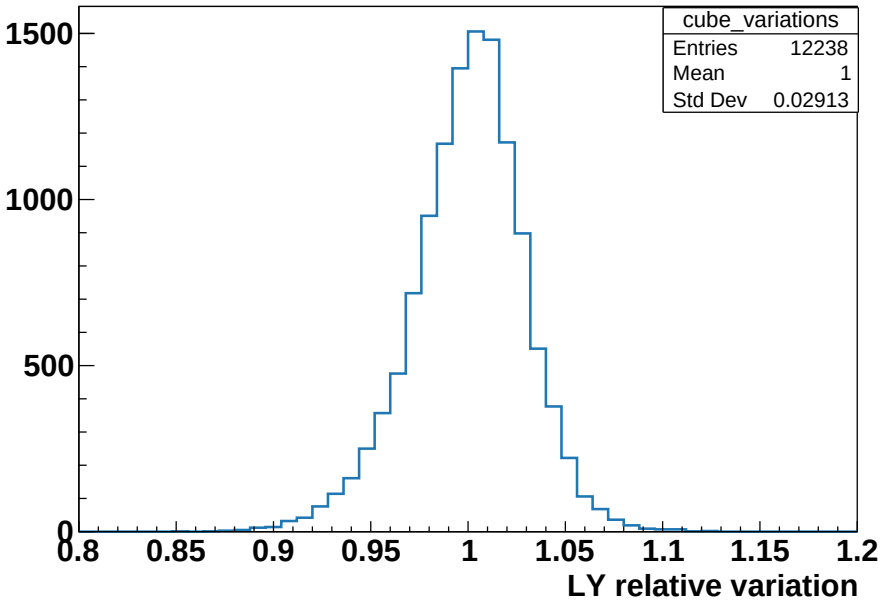


Figure 3.35: Relative light yield variation with respect to the average light yield of each plane. The light yields here are corrected of the attenuation and coupling effects.

3.4.6 Time evolution

To ensure a good knowledge of the energy response of the detector, several calibration campaigns have been performed over time, as presented on table 3.1. For all those periods, the couplings, attenuation lengths and light yields have been measured with the methods presented in the previous sections. The light yield evolution through the two years of data of the so called SoLid Phase I can be seen on figure 3.36. The effect of the PVT ageing can be seen with a global decrease of the average light yield through time. Two periods can be identified: before and after May 2019, where the average light yield dropped of a few percent. This drop happened just after a chiller failure in the detector container and an issue in the monitoring procedure during a maintenance of the detector, leading to an increase of the temperature to

$50\,^{\circ}\mathrm{C}$ at the detector level instead of the usual $12\,^{\circ}\mathrm{C}$. It is believed that this heating caused a degradation of the PVT light yield. The average decreasing of light yield through time seems also to have raised from -1.2% to -3.3% per year.

Furthermore, the mechanical constraints on the fibre and the environmental conditions could have an effect on the coupling and attenuation length, this is why the evolution of those parameters is also monitored as presented on figure 3.37. On this plot are represented the ratio of both attenuation length and coupling between a given calibration campaign and the previous one. From a calibration to another, the average of the attenuation varies less than 1 % with a standard deviation lower than 5% for most of the campaigns, while the coupling varies in average less than 1%. The largest spread can be seen on the red points that represents the evolution before and after the heating incident. The hypothesis behind the variations of the fibre parameters is a modification of the connexion surfaces between the fibres and the mirrors or MPPCs made with optical grease, due to mechanical constraints induced by the heating, deterioration of the optical grease and small modifications of the fibre responses with the temperature. Those last results show the very good stability of the fibre response and the robustness of the calibration procedure over time.

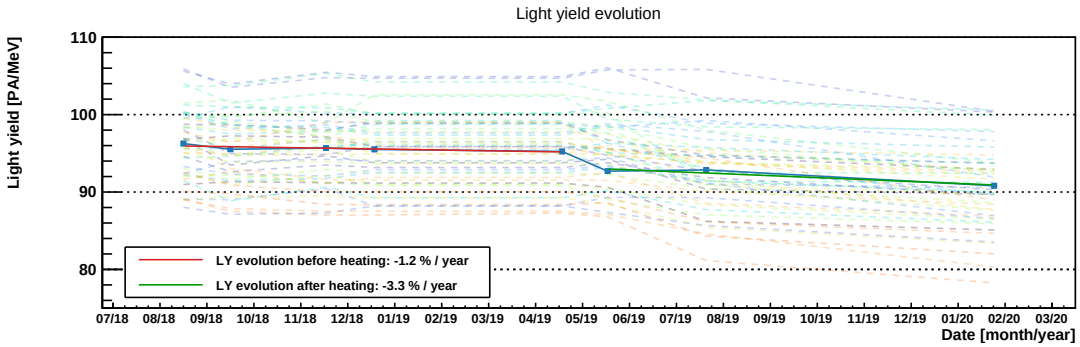


Figure 3.36: Evolution of the average light yield in the detector. Each plan is represented by a doted line and the average variation os represented by the blue dots. The light yield loss between the first and last calibration is -5.6 % in average. An effect of the heating of the detector can be seen around May 2019 with a sudden drop of the light yield assumed to be due to an accelerated ageing of the PVT. In doted are represented the average light yields of each plane.

3.4.7 Energy resolution

With the methods presented above, one can also measure the energy resolution of the detector. The usual description of the energy resolution of a calorimeter can be written as:

$$\frac{\sigma}{E} = \frac{a}{E} \oplus \frac{b}{\sqrt{E}} \oplus c \tag{3.24}$$

where :

– a is a noise contribution induced by electronic noise in the readout.

– b is a stochastic contribution due to statistical fluctuations on the number of avalanches induced by a given energy deposit.

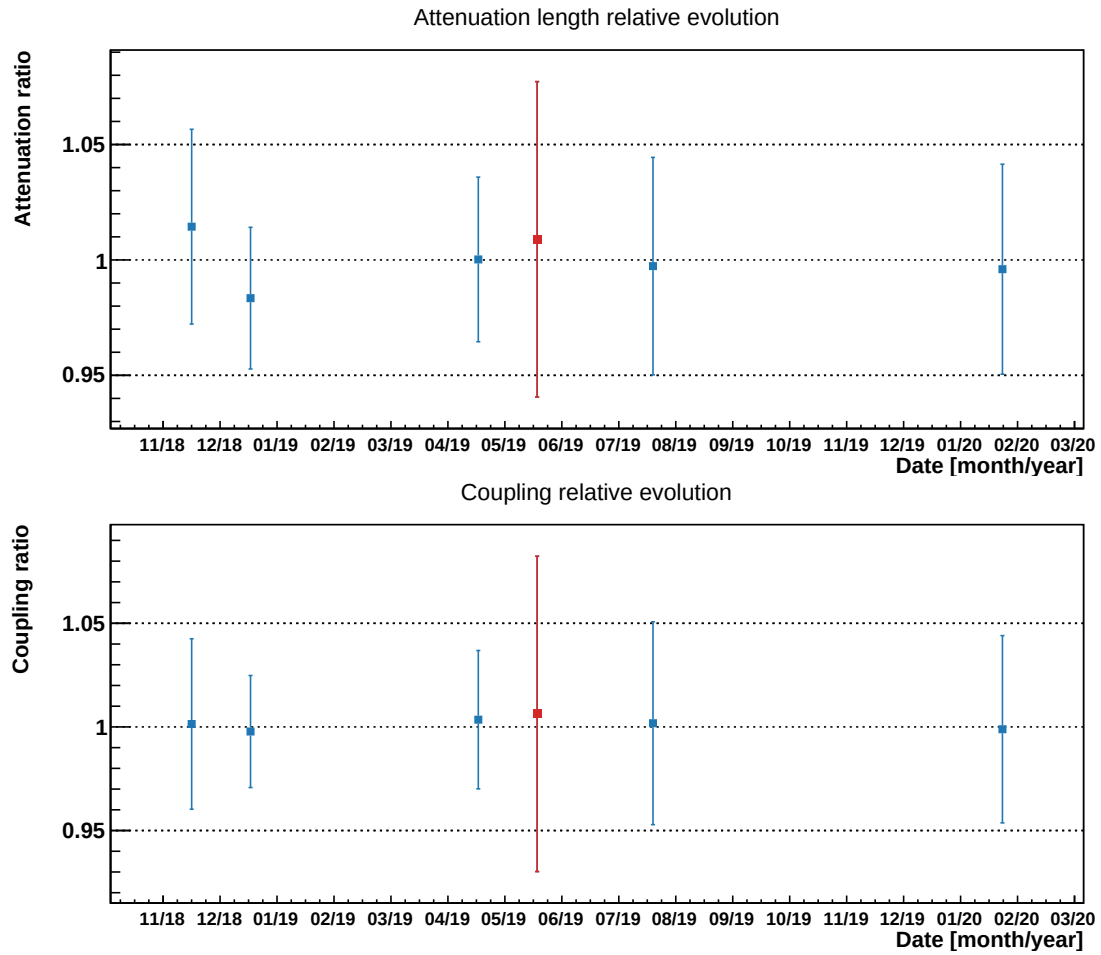


Figure 3.37: Top: Evolution of the attenuation length in time. Bottom: Evolution of the coupling in time. Each point represents the averaged ratio of the parameter between the calibration campaign and the previous one. The error bars represent the standard deviation of the ratio distribution. The red point represents the variations before and after the heating of May 2019.

– c is a constant contribution independent of the energy deposit that comes from light leakages and dead materials in the detector.

In the same way as the linearity, a measurement of the cube energy resolution was performed via the method described in 3.4.2. Those parameters can then be fitted with the equation 3.24 as shown in 3.38. The energy resolution obtained is then:

$$\begin{aligned} a &= 12 \pm 1\% \\ b &= 11 \pm 1\% \\ c &= 3 \pm 1\% \end{aligned} \tag{3.25}$$

An important point to note on both methods presented to measure the energy resolution of the cubes is that they are both under the assumption that there is only a stochastic term in the energy resolution. Both resolution effect applied is computed with only the $b \times \sqrt{E}$ term. On the energy range fit around the Compton edge, for a given a_1, b_1, c_1, set of resolution parameters from equation 3.24, the resolution can be approximated by $b_2 \times \sqrt{E}$ with b_2 an arbitrary parameter. The convolution of a ^{22}Na energy

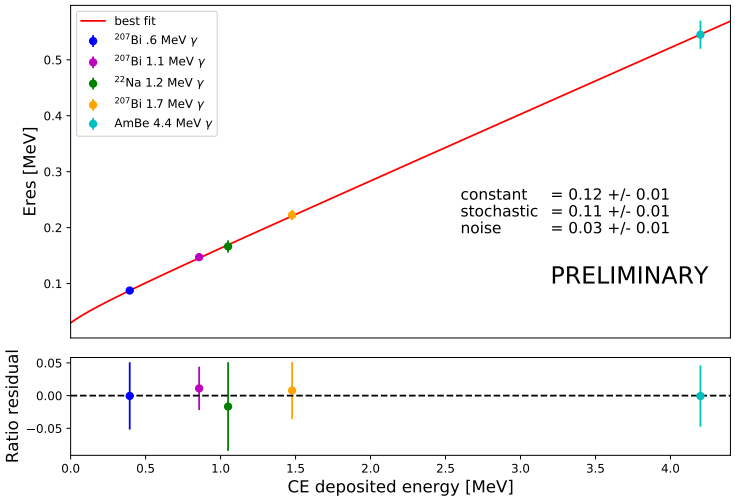


Figure 3.38: Energy resolution measurement for the 5 different calibration energies.

spectrum with both resolutions is represented on figure 3.39 around the Compton edge on the energy range used for the calibration fit. Both convoluted energy spectra are within 5 % on the Compton edge region because both resolution formula are close in that area. For the AmBe ~ 4.2 MeV Compton edge, the resolution is dominated by the stochastic resolution. This can be seen on figure 3.40 with on the left panel both resolutions curves within 1 % and the convoluted spectra on the right panel within 2 % on all the fitted range.

This study could be used to optimize the calibration parameters measurement like the energy ranges for the fits and the convolution of the true GEANT4 energy spectra but the current performances of the calibration and the data to Monte-Carlo comparisons presented on section 3.5 were good enough to leave the method as it is. Furthermore, the resolution measured here is the energy resolution of a cube, whereas in the analyses the energy estimators are clustering cube energies. This energy estimator, as a combination of several cube energies, will have a bigger energy resolution than the individual cubes. Furthermore, the resolution is not directly tuned inside the simulation but obtained after the simulation of all the detector effects. It is thus less important to know directly each cube energy resolution.

3.4.8 Energy calibration implementation

Once all the calibration procedure has been performed, the calibration is used in two different parts of the collaboration software: in the simulation and in the reconstruction.

3.4.8.1 Energy calibration inside the simulation

The use of the energy calibration in the ROsim is partially presented in section 2.6.2.1: the attenuation lengths and couplings are set per channels and used in the simulation of the simulation response. The relative light yield is set per cube using the intrinsic light yield variations for each plane. A plane to plane relative variation is then applied to the simulated light yields. That way, each of the attenuation,

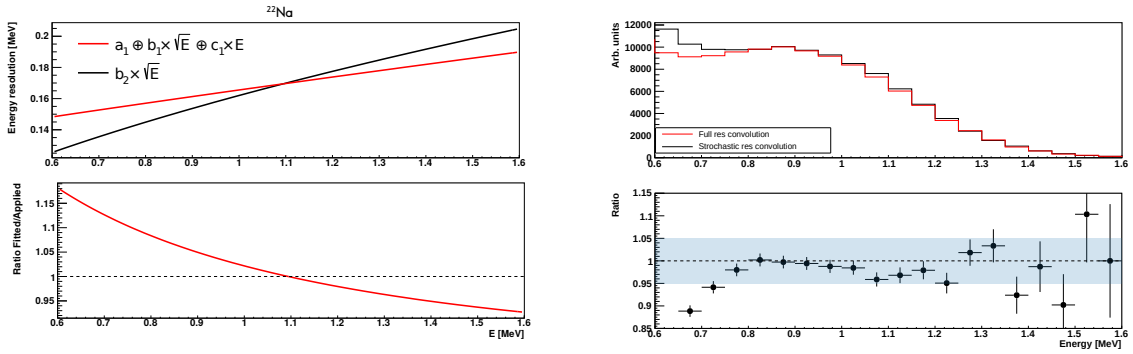


Figure 3.39: Left: energy resolution with only a stochastic term (black) and with the equation 3.24 on the ^{22}Na fit range. Right: convoluted ^{22}Na energy spectrum with the a stochastic resolution (black) and the resolution described in the equation 3.24 (red). On the ratio plot, the ± 5 % band is represented in blue.

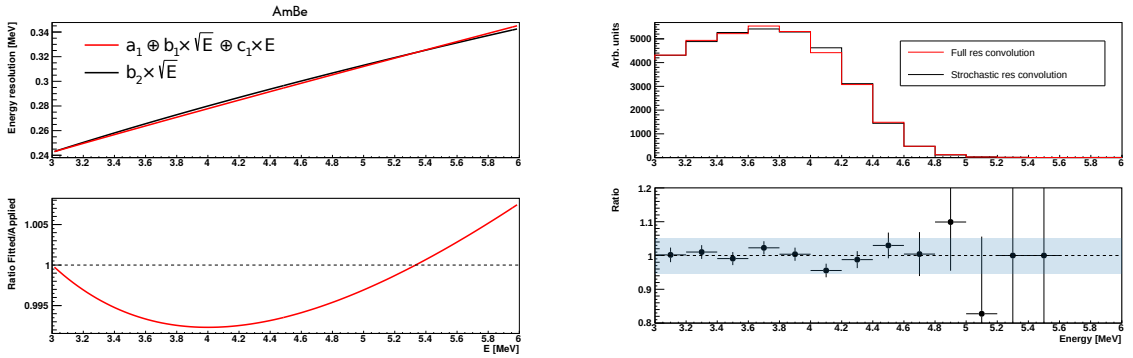


Figure 3.40: Left: Energy resolution with only a stochastic term (black) and with the equation 3.24 on the AmBe fit range. Right: convoluted AmBe energy spectrum with the a stochastic resolution (black) and the resolution described in the equation 3.24 (red). On the ratio plot, the ± 5 % band is represented in blue.

coupling, light yield variation effects are simulated independently and can be simulated or not if needed. The last parameter derived from the calibration used in the simulation is the global light yield: a global constant applied to all cubes to determine, from a given energy deposit, how many photons will be generated. This last parameter tuning is developed in the section 3.5.1. A scheme of the implementation of the calibration is presented on figure 3.41.

The calibration parameters usage can be summarized in the following equation:

$$A_{i,j} \propto \frac{1}{4} \times a_j^{G4} \times LY^{global} \times LY_j^{relative} \times \epsilon_i^{coupling} \times (\epsilon_{i,j}^{att,dir} + \epsilon_{i,j}^{att,ref}) \tag{3.26}$$

with a_j^{G4}, the true energy deposit in the cube j, $A_{i,j}$ the amplitude of the channel i of the cube j, LY^{global} the global light yield input of the whole simulation, $LY_j^{relative}$ the relative light yield of the cube j and the parameters ϵ defined as in the previous sections. Here the light leakages and other electronic effects are not represented in the equation because they are not produced via this calibration work.

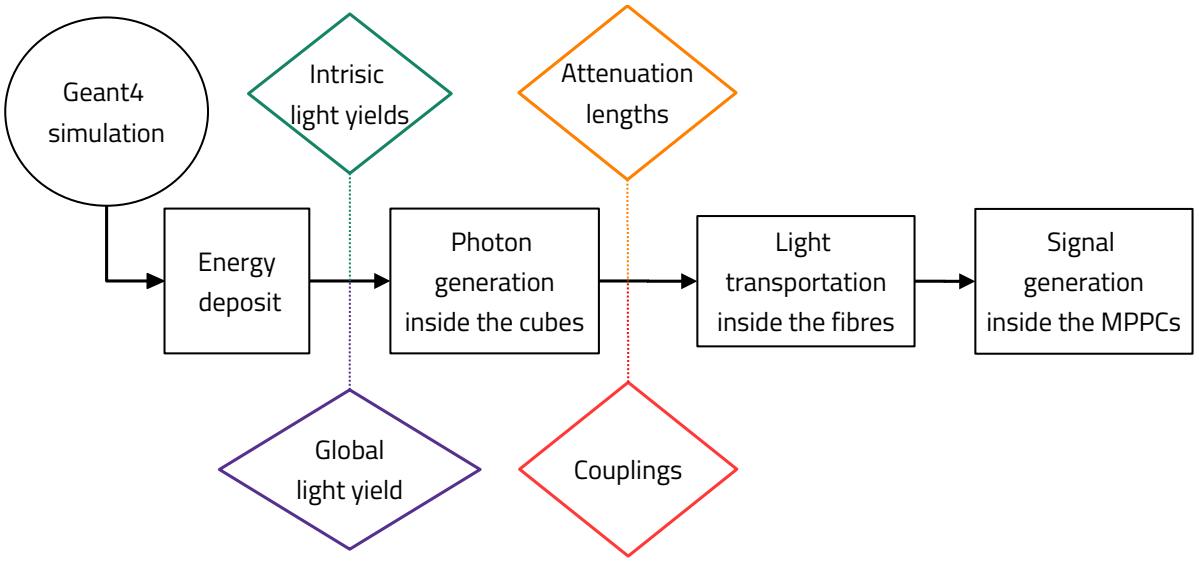


Figure 3.41: Scheme of the usage of the calibration variables inside the Readout simulation.

3.4.8.2 Energy calibration inside the reconstruction

The other software which uses the results from the energy calibration is the reconstruction software Saffron. The reconstruction of the electromagnetic clusters has been presented in the section 2.7.2.3. A scheme of the implementation of the calibration inside the reconstruction is presented on figure 3.42. In this reconstruction, the cubes are created via an iterative process, with the projection of all channel energies on the cubes with the help of a system matrix that encodes the detector effects. Ideally, the system matrix would gather all the calibration parameters to take into account all the detector inhomogeneities in order to distribute the energy inside the cubes. In the current version of the reconstruction software, the system matrix is a totally flat matrix: all the components of the matrix are the same. That means that the matrix does not encode any detector effect for now. The calibration correction must thus be applied outside the iterative algorithm presented on figure 2.26.

A first correction of the channel parameters is applied directly on the channel energies. The gain, pedestal and coupling are corrected at this level. One can write the following equation:

$$A_i[\text{PA}] = \frac{a_i[\text{ADC}] - \text{Pedestal}}{\text{Gain} \times \epsilon_i^{coupling}} \tag{3.27}$$

with $a_i[\text{ADC}]$ the input amplitude measurement before any correction of the channel i and $A_i[\text{PA}]$ the corrected amplitude of the channel i. The fibre attenuation effects are not corrected here because it requires a localization of the cube reconstructed and at this stage, only the fibre energies are reconstructed. The ML-EM algorithm is then applied on the amplitudes A_i to reconstruct in energy and position the electromagnetic clusters. Then a last correction is applied on the cubes. This correction takes into account both the cube light yield and the four channels of the cubes. This visible light yield is the one presented on figure 3.33. It is computed without any attenuation correction, thus it corrects for the light

yield variations and the fibre attenuation effects. The cube energy can then be derived as:

$$E_j[\text{MeV}] = f_{ML-EM}(\vec{A}_i[\text{PA}], j)/LY_j \tag{3.28}$$

with E_j[MeV] the energy of the cube j, $\vec{A}_i$ the amplitudes of all the channels of the cluster, LY_j the visible light yield and $f_{ML-EM}(\vec{A}_i[\text{PA}], j)$ the output of the algorithm for the cube j. In the future, the calibration will be used to build directly the system matrix and everything will be treated in the ML-EM algorithm.

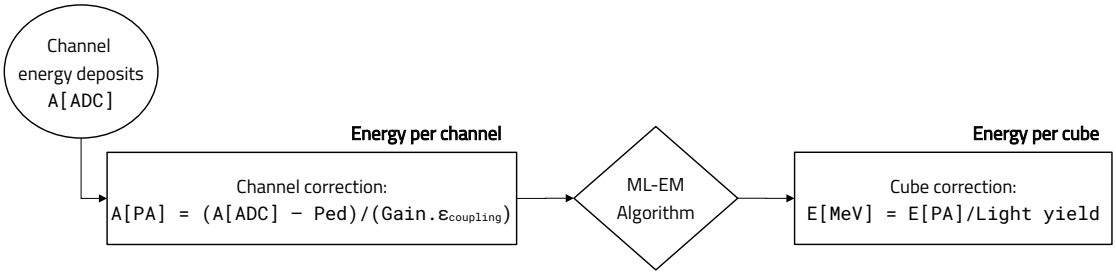


Figure 3.42: Scheme of the usage of the calibration variables inside the reconstruction software. The gain, pedestal and coupling effect are corrected directly at the channel level on the measured amplitudes. A visible cube light yield that encodes both fibre attenuation and light yield variations is applied at the end of the reconstruction chain, once all the cubes have been created.

3.5 Data - Monte Carlo comparison

The last step needed to test the understanding of the detector energy response is Data to Monte Carlo comparisons. Those are performed with *threshold trigger* on regular ^{22}Na calibration campaigns and with *periodic trigger* on special calibration campaigns. In this section, due to the need of selecting cubes with specific channel amplitudes, the energy of a cube will be defined as the sum of the energies of the four channels of the cube. A study made on BiPo background will be presented on section 4.4.2 with data to Monte-Carlo comparison on energy estimators using the CCube algorithm.

3.5.1 Threshold trigger comparisons

The first step is to tune the simulations with the couplings, attenuation lengths and cube to cube relative light yield variations from the calibration measurement. Those measurements were presented in the previous sections. The last parameter to tune is the global cube light yield as an input of the simulations. It is a constant which is the same for all the cubes that is tuned with data to Monte-Carlo comparisons with ^{22}Na. To do so, for a given input light yield, a ratio of the energy distributions data over Monte-Carlo is computed as presented on figure 3.43. For this comparison, only cubes with at least one vertical and one horizontal channel above 7.5 PA and with 4 channels above 2.5 PA are considered. The isolation selection is also applied to the cubes. A good match is expected on the Compton edge of the distributions. Some difference of reconstruction efficiency can be seen at low-energy, at around 35 PA (∼300 keV).

The discrepancy might be explained by a trigger effect not reproduced in the simulations. In the data with the threshold trigger, data in ± 6.4 µs around the triggering event are readout in the triggered plane. During this time, due to the source activity, several ^{22}Na decays can pile-up in the data while the decays are treated one by one in the Monte-Carlo. All the energy deposits larger than the ~ 30 PA for a cube will be readout in both cases, while deposits lower than the threshold are readout only if a plane is triggered by another deposit. If two decays happen in the same 12 µs in the data, there is thus a higher probability that a plane is triggered with a lower energy deposit from one of both decays. This difference could be the reason of excess in the low-energy region in data compared to the Monte-Carlo. Due to this discrepancy, all spectra are normalized at the Compton edge, between 60 PA and 140 PA.

This comparison was done for a selection of cubes around the source for a given source position. On figure 3.44 can be seen the data/Monte-Carlo ratios for multiple cubes. The red profile represents the average of the energy ratios per energy bin. The agreement is within a 5 % in average on the [30-140] PA energy range (∼[300-1400] keV). This shows a good agreement between the data and the Monte-Carlo at the Compton edge region, and validates the simulation of the energy processes.

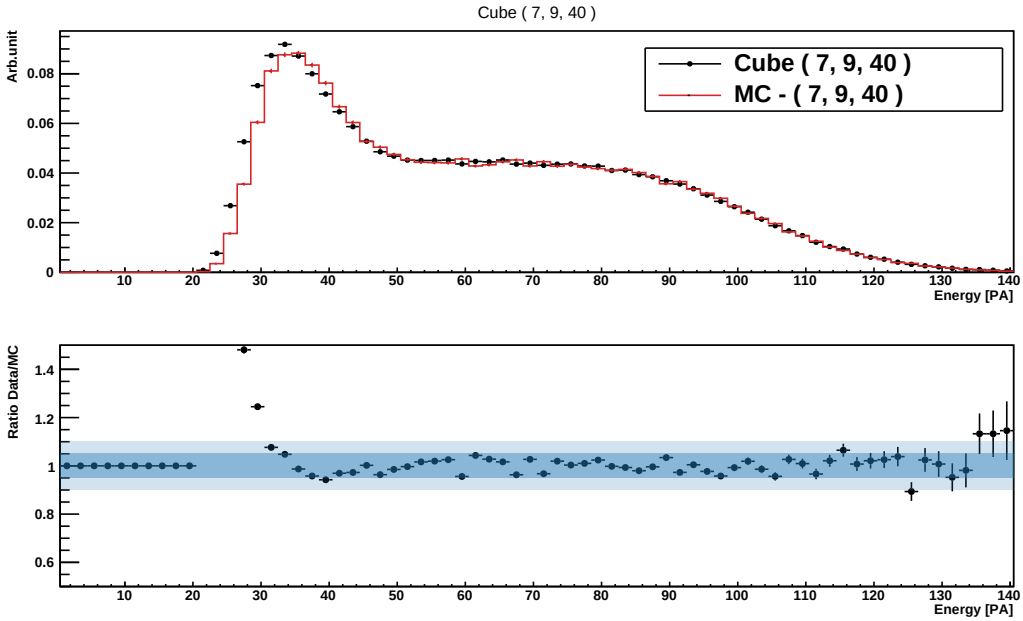


Figure 3.43: Comparison between data and simulation of a given cube with the threshold trigger. The low-energy region is identified as a bias induced by the threshold trigger not totally well reproduced in the simulation.

This type of comparison is used to determine the best input light yield in the simulation: A set of simulations are produced varying the input light yield and the one with the best data - Monte-Carlo agreement is chosen. On figure 3.45 can be seen the tuning of the February 2020 simulation light yield. The optimal light yield profile is plotted in black and in red (green) are showed the profile for +2% (-2%). In practice, this has been done for September 2020 and the optimal light yield is then computed with the ratio of the average measured light yield of the detector after calibration. For instance, for the February calibration, the computed simulation light yield is:

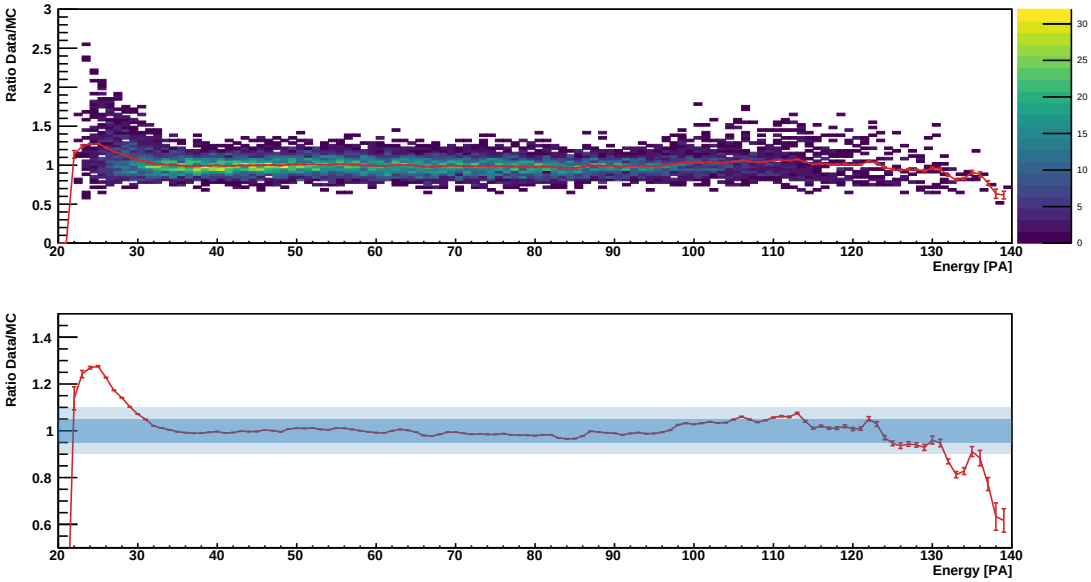


Figure 3.44: Comparison between data and simulation for several cubes. Each bin in the top histogram represents a data/MC energy ratio bin for a given cube. The average per energy bin is plotted in red.

$$LY^{sim}_{\text{February 2020}} = LY^{sim}_{\text{September 2018}} \times \frac{LY^{meas}_{\text{February 2020}}}{LY^{meas}_{\text{September 2018}}} \tag{3.29}$$

the results are represented in table 3.2.

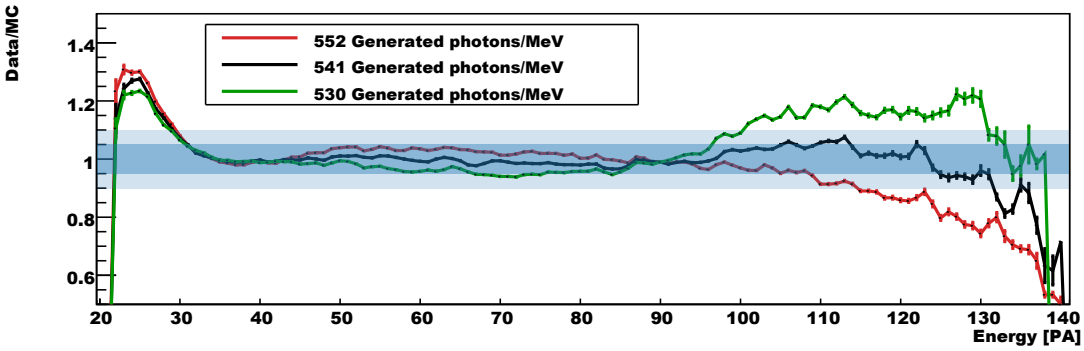


Figure 3.45: Profiles of the data/Monte-Carlo tuning. Each colour represent a value of input light yield in the simulation. The optimal value is plotted in black, and in red (green) are showed the profile for +2% (-2%).

3.5.2 Periodic trigger comparisons

To study the lower energy range without any threshold or trigger bias, a comparison with periodic trigger data and Monte-Carlo is necessary. Only cubes with 4 channels above 2.5 PA are considered. An agreement within 5 % is found in the energy range [15-130] PA with a good reconstruction of the low-energy part as shown on figure 3.46. The tuning of the simulation can also be studied with those

Date	Measured averaged light yield [PA/MeV]	Simulation input generated photons per MeV
September 2018	96.3	574
October 2018	95.5	570
December 2018	95.7	570
January 2019	95.5	570
May 2019 (before intervention)	95.2	568
May 2019 (After intervention)	92.7	553
August 2019	92.9	554
February 2020	90.9	541

Table 3.2: Recap of all the input global light yields in the simulations. Each light yield is obtained by using September 2018 input light yield scaled with the average light yield from the calibration campaigns.

data as represented on figure 3.48 with an optimal value and ± 1 % light yield variations. This shows that the light yield tuning with the measured ratio between a given campaign and September 2018 gives an agreement at the 1 % level.

3.6 Conclusion

The main strength of the SoLid detector, its large segmentation, comes with a challenge to control the detector response on a large number of detection cells. For an antineutrino oscillation analysis, a thorough understanding of the detector is essential.

The first task of the calibration was to equalize the channel electronic responses with small differences evolving through time according to environmental conditions. This has been done with a correction every 6 h to take into account day/night temperature variations. This equalization shows a sub 3% variation of the gains and sub 1% variation of the pedestals during the nearly two years of data taking showing a stable detector without triggering differences induced by the electronics.

Then a fibre light losses quantification has been performed, with fibre attenuation lengths and couplings measurements. Two methods were designed to do so with a first sequential measurement that was then coupled with a global fit proposed by Subatech Nantes team to achieve a measurement with an averaged bias lower than 2 % and a 4 % deviation on all fibres on calibration simulations.

Contrary to experiments with bigger detection cells, the energy calibration necessitated the development of methods using Compton edges. Two methods were developed and presented in this document. A good agreement was found between the analytical fit developed by the Subatech team and the numerical fit, with discrepancies bellow 1 % level between both methods.

With this calibration, the homogeneity of the light yield of 3% within each planes have been measured, with a linearity within a few percent in the [0.5 - 4.2] MeV region.

A measurement of the calibration parameters have been performed regularly during the 2 years of data, showing a stable response of the detector. The calibration allowed a precise tuning of the simulations with a match below 5% on the Compton edge region using the high energy threshold trigger and on a bigger range, down to 10 PA (~ 20 keV) on unbiased data. A data to Monte-Carlo comparison study

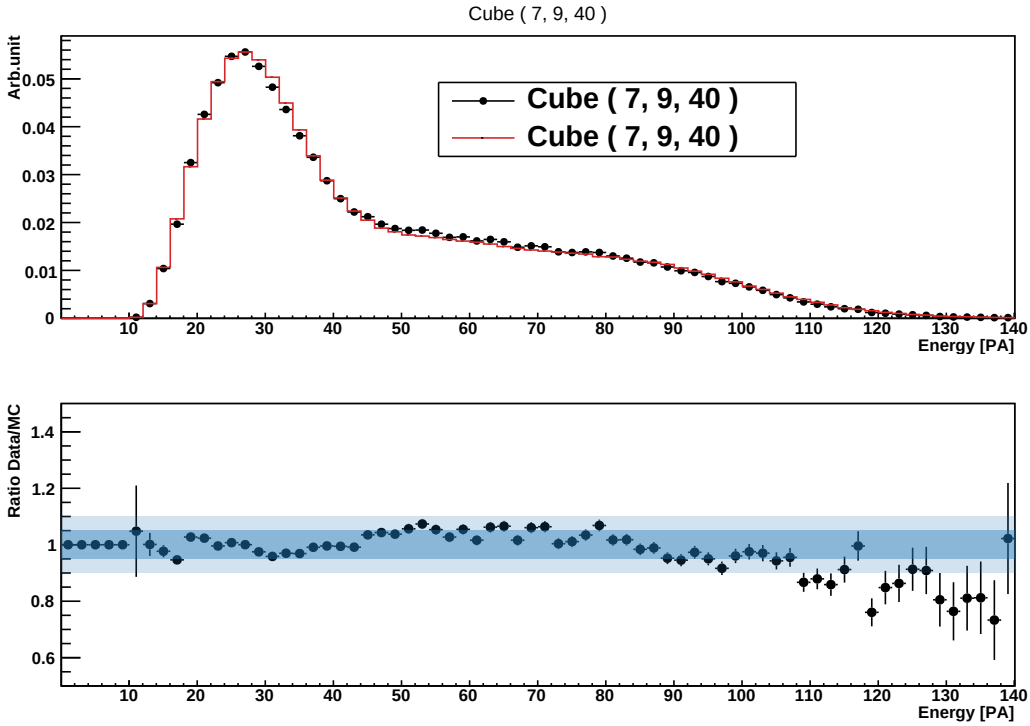


Figure 3.46: Comparison between data and simulation of a given cube with the periodic trigger.

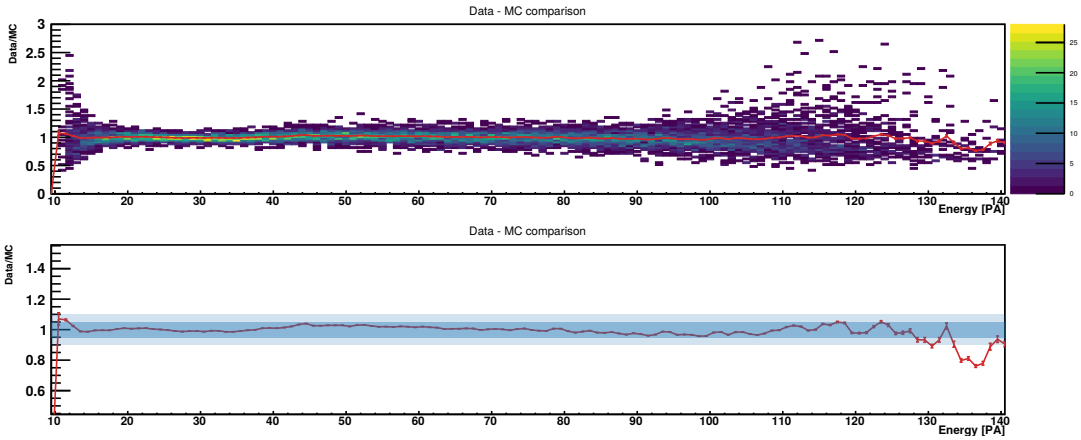


Figure 3.47: Comparison between data and simulation for several cubes with the periodic trigger. Each bin in the top histogram represents a bin on the left histogram for a given cube. The average per energy bin is plotted in red.

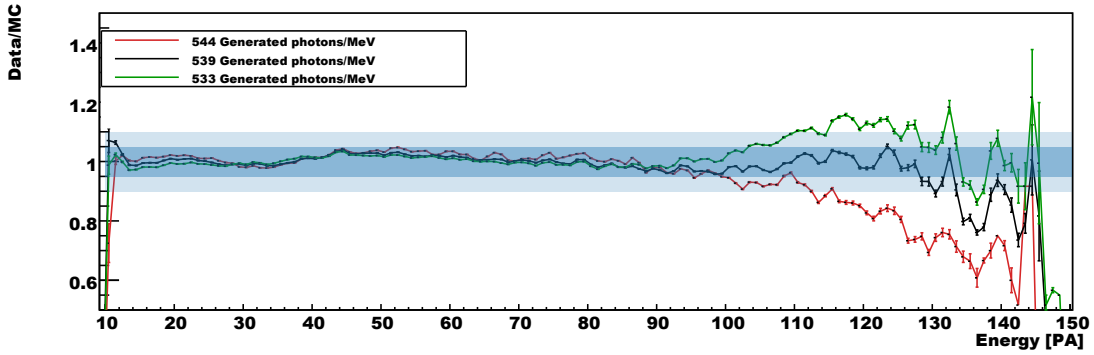


Figure 3.48: Profiles of the data/Monte-Carlo tuning. Each colour represent a value of input light yield in the simulation. The optimal value is plotted in black, and in red (green) are showed the profile for +1% (-1%).

has been performed on ^{22}Na annihilation gammas Compton edge in [100]. This study shows a good reproduction of the annihilation gamma energies with a Compton edge at ~ 340 keV as represented on the figure 3.49.

The next step of the calibration would be to go at the fibre level where small discrepancies are still seen at low energies as shown in figure 3.50. A fibre calibration, with 4 calibration parameter per cube for each channel, that would take into account all the calibration parameters removing the need of any model or any correlation between two parameters could be the final step to obtain a sub 5% agreement at a few pixel avalanches in all the detector. All the calibration parameters would then be encoded inside the system matrix used in the ML-EM reconstruction algorithm. I started to work on this new calibration but did not push it through while the colleagues at LPC Clermont that developed the ML-EM algorithm took over the work.

The level of agreement already obtained in the calibration, with the robust measurement of the 12800 cube light yields, 3200 attenuation lengths and 3200 coupling parameters already shows a good understanding of the energy response at the cube level of a complex detector with a trustworthy simulation of the SoLid detector. The effects of the calibration uncertainties level will be evaluated in dedicated studies in section 5.1.

A data to Monte-Carlo comparison on higher level reconstruction variables will be presented in the section 4.4.2 confirming the good reproduction of variables of interest in the final antineutrino analysis. This is a good step to achieve a rigorous antineutrino analysis. The decision was then made to stop the calibration work and start an antineutrino analysis.

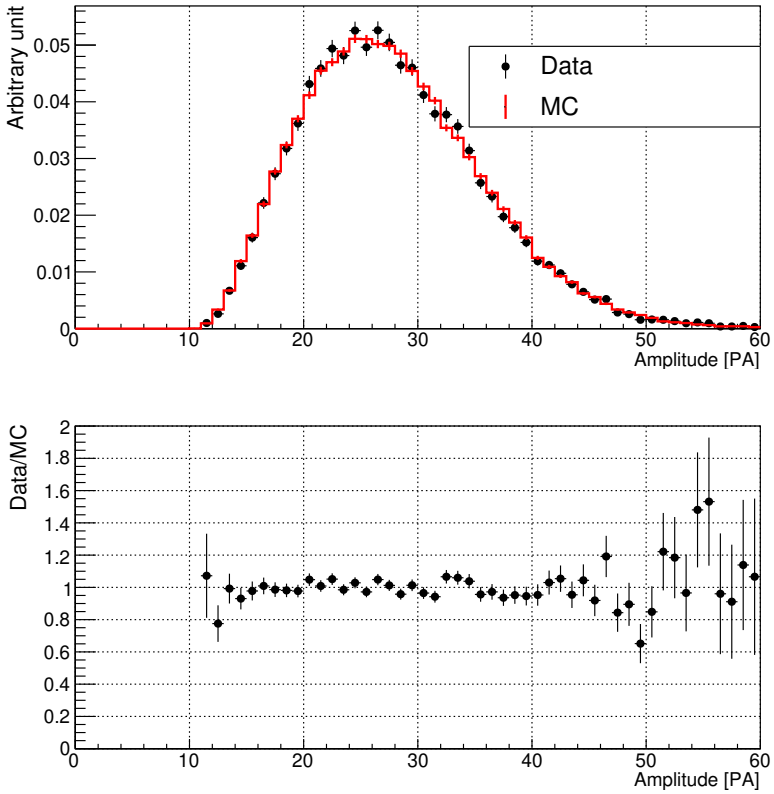


Figure 3.49: Data to Monte Carlo comparison with ^{22}Na and with an annihilation gamma selection. Both spectra are normalized to one [100].

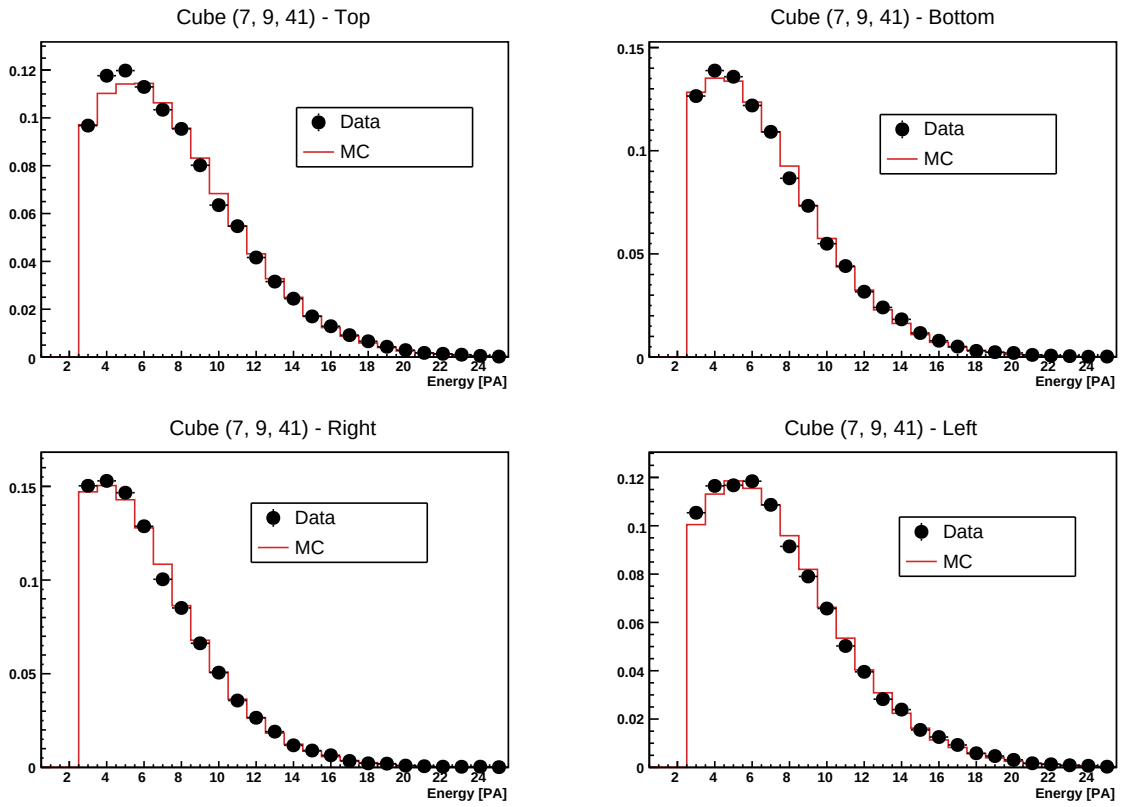


Figure 3.50: Data to Monte Carlo comparison with ^{22}Na with the random trigger for the four channel of a given cube.

Chapter 4

Antineutrino analysis

4.1 Introduction

After an extensive calibration work to understand the detector energy response, the final step before the oscillation to a sterile neutrino state measurement is the extraction of the antineutrino signal. The goal is to measure, for each event, both the energy and the travelled distance between the reactor and the interaction point in the detector. This antineutrino analysis is composed of two main parts. The first one consists in building a data selection that maximizes the antineutrino detection rate and the purity of the sample. This part is done with Monte-Carlo IBD simulation and reactor OFF data. Once a selection has been made, the second part of the analysis is to subtract the remaining backgrounds from the reactor ON data to extract the antineutrinos. Each background component (accidental coincidences, BiPo background and atmospheric background) is subtracted in both reactor ON and reactor OFF datasets. The reactor OFF is used as a control sample to check that the background has been properly removed.

For the antineutrino analysis, a blind analysis strategy has been chosen by the SoLid collaboration. From the nearly two years of data, the analysis will first be performed and optimized on a small sample called the *open dataset* composed of a few days of reactor ON and OFF (RON and ROFF). This dataset will be used to choose the selection cuts, build the analysis algorithms and verify the quality and robustness of the analysis. Once the analysis is defined and the main systematic effects understood, the analysis will be frozen and ready to be applied to a larger sample of the data. This strategy is common in analyses of a large data sample and is good to avoid unintended biases in the final results introduced by the analysis.

In this chapter, some parts of the data have been totally unblinded in order to study the detector stability over time. To do so, reactor OFF data with a selection outside the analysis selection have been used. For instance, data with energy larger than 7 MeV, outside the neutrino energy region of interest, have been used to study the atmospheric background in the section 4.3.3. Enhanced BiPo background data have also been used to study the Data to Monte-Carlo agreement over time in 4.4.2. As the selection on those data does not share any events with the final antineutrino selection, this unblinding will not introduce any bias in the analysis.

The total reactor ON and OFF periods are presented in the table 4.1 and 4.2. In total, approximately 326 days of reactor ON and 187 of reactor OFF are available. The number of reactor OFF days is lower than the reactor ON because all the maintenance, calibrations and BR2 operation happened during reactor OFF periods. Furthermore, the first and last periods are reactor ON and nearly no data was taken be-

Cycle number	Start data	End date	Total run duration (days)
2	24-04-2018	22-05-2018	20.9
3	12-06-2018	10-07-2018	24.3
4	21-08-2018	11-09-2018	19.7
5	02-10-2018	23-10-2018	19.2
6	13-11-2018	11-12-2018	25.8
7	05-02-2019	05-03-2019	24.8
8	26-03-2019	30-04-2019	32.2
9	02-07-2019	06-08-2019	4.17
10	17-09-2019	22-10-2019	32.2
11	05-11-2019	03-12-2019	25.8
12	09-01-2020	07-02-2020	27.0
13	03-03-2020	31-03-2020	22.2
14	17-04-2020	18-05-2020	28.7
15	04-06-2020	02-07-2020	18.7

Table 4.1: Overview of the Phase I reactor ON periods, corresponding to approximately 326 live data-taking days at BR2. The first cycle available is labelled as cycle 2 since the cycle 1 happened before the data taking of SoLid. The cycle 9 has a low number of days available because the detector was in maintenance during most of the cycle.

tween April 2019 and September 2019 due to maintenance and issues in the detector. On those datasets, 21 days of reactor ON (between the 19/06/2018 and the 09/07/2018, in the cycle 3) and 19 days of reactor OFF (between the 12/07/2018 and the 31/07/2018, in the period $3-4$) are used for the open dataset. Later, a larger proportion of reactor OFF data will be used to check the stability of the background subtraction method and the different background stability over time.

In this chapter, the analysis will be presented in different steps, with a simple pre-selection with rectangular selection cuts on low-level reconstructed variables. Then, an annihilation gamma reconstruction and selection will be presented to exploit the event topologies. This gamma reconstruction work was widely designed by David Henaff and also presented in detail in [100]. Lastly, a multivariate analysis will be presented to enhance the detector background rejection power.

4.2 Variables of interest

4.2.1 Selection variables

The main variables used in the analysis are presented in this section. For antineutrino detection, IBD event candidates formed of an ES cluster and an NS cluster in time and space coincidence are built. When an NS cluster is made, it is associated with all possible ES clusters in a time window of $\Delta T \in [-200, 500]$ µs. On those events a first set of low-level variables are built:

Label	Start date	End date	Total run duration (days)
2-3	22-05-2018	12-06-2018	11.2
3-4	10-07-2018	21-08-2018	30.2
4-5	11-09-2018	02-10-2018	4.14
5-6	23-10-2018	13-11-2018	15.3
6-7	11-12-2018	05-02-2019	43.5
7-8	05-03-2019	26-03-2019	6.34
8-9	30-04-2019	02-07-2019	15.4
9-10	06-08-2019	17-09-2019	7.32
10-11	22-10-2019	05-11-2019	5.92
11-12	03-12-2019	09-01-2020	20.6
12-13	07-02-2020	03-03-2020	6.24
13-14	31-03-2020	17-04-2020	13.1
14-15	18-05-2020	04-06-2020	7.18

Table 4.2: Overview of the Phase I reactor OFF periods, corresponding to approximately 187 live data-taking days. Label $i-j$ indicates that the associated reactor OFF period is between reactor ON cycles i and j.

- ΔT is defined as the time difference between the delayed NS and the prompt ES.
- $\Delta X, Y, Z, R$ are defined as the distance (in number of cubes) between the delayed NS and the prompt ES in the X,Y,Z direction or in the volume R.
- $promptNCCube$ is defined as the number of cubes composing the ES cluster.
- $promptVolume$ is used to describe the volume of the ES cluster. It is defined as:

$$\prod_{i \in X,Y,Z} (\max(i) - \min(i))$$

  where $\min(i)$ ($\max(i)$) is the cube with the lower (higher) coordinate in the i direction.

The distribution in these variables for four days reactor OFF data in the open dataset and the IBD Monte-Carlo simulation are represented on the figure 4.1. The details of the reactor OFF distributions will be discussed in the next sections, with a detail of each background component. One can already see a first discrimination possible at high ΔT due to the lower characteristic time of the neutron capture compared to the BiPo decay time constants and compared to the flat ΔT component of the accidental background. As the NS and ES are correlated in space for IBDs compared to the accidental backgrounds, low values of $|\Delta X, Y, Z, R|$ will be selected to maximize the IBD signals. The asymmetries in the ΔX and ΔZ background and IBD distributions are due the position of the ZnS layers inside the cubes that enhances the negative ΔX and positive ΔZ. The effect on Z is even stronger for the IBDs due to the kinematic of the IBD interaction that will favour neutron emission in the positive Z direction.

4.2.1.1 Energy Estimators

The next parameters to present are the different estimators of the energy of the positron resulting from the IBD, as describe in the section 2.2.1. When an ES cluster is reconstructed, a list of cubes is built with, for each of them, an energy and a position. The cube with the highest energy reconstructed is called the annihilation cube (AC). It is expected that the positron resulting of the IBD will deposit the majority of its energy and annihilate in the same cube. The first requirement for an oscillation analysis is to reconstruct the interaction point of the antineutrino, the AC position is considered as the interaction point of the antineutrino. The AC positions will thus be used as the position of the ES clusters. The other requirement for the oscillation analysis is to reconstruct precisely the energy of the antineutrinos. This energy can be either retrieved from the positron deposited energy, or from the sum of the energy deposited by the positron and by the interactions of annihilation gammas. From the set of cubes created in an ES cluster, several energy estimators can be built to tackle the different situations. Those estimators are represented for IBD Monte-Carlo on figure 4.2 and can be detailed as:

- AC_energy is the simplest energy estimator: It is the cube with the highest energy deposit reconstructed with the CCube reconstruction algorithm. It corresponds to the expected positron reconstructed energy, without taking into account any light leakages or the annihilation gammas for an IBD interaction. The main advantage from using this energy estimator is that it does not mix several types of energy deposits, even with a broad selection. However, the positron in not always confined in the annihilation cube and can leave its energy in the neighbouring cubes. Furthermore, due to light leakages of the cube, the energy can be underestimated, and another resolution effect can be added to the event energy. The use of the AC_energy thus requires a precise knowledge of the bias introduced by the positron leaving the annihilation cube and the light leakages in all the cubes of the detector.

- $Prompt\_energy$ is the sum of the energies of all the channels in the spatial buffer of the trigger. Contrary to other estimators this estimator directly sums the channel amplitudes and does not directly rely on a cube energy reconstruction. This avoids any reconstruction issue in the repartition of the energy in an event. For an IBD interaction, this variable should gather the positron deposited energy and the annihilation gamma ones if they interact within the ± 3 planes around the NS. As it is expected to gather the energy deposits from all the products of an IBD, it is easier to recover the antineutrino energy with fewer ambiguities due to the contribution of the annihilation gammas or the positron that would have leave its energy in several cubes. As all the channel amplitudes all over the detector are summed, one cannot calibrate properly the energy with a given cube light yield. For this estimator, the light yield of the annihilation cube is used as a global light yield. This energy estimator thus does not use the calibration of the detector at its full potential.

- $Crown\_energy$ is the sum of the AC_energy and the energies of all the cubes in the 3x3x3 volume around it. It should contain in most of the case the full energy of the positron. If the issue of the escaping positron is treated here, a more precise analysis needs to be performed in order to differentiate the cases where 0, 1 or 2 annihilation gammas deposit their energies in the 3x3x3 volume.

In the next sections, if not specified otherwise, the $Crown\_energy$ will be used as the energy variable. It is a compromise between the large light collection of the $Prompt\_energy$ and the precise energy mea-

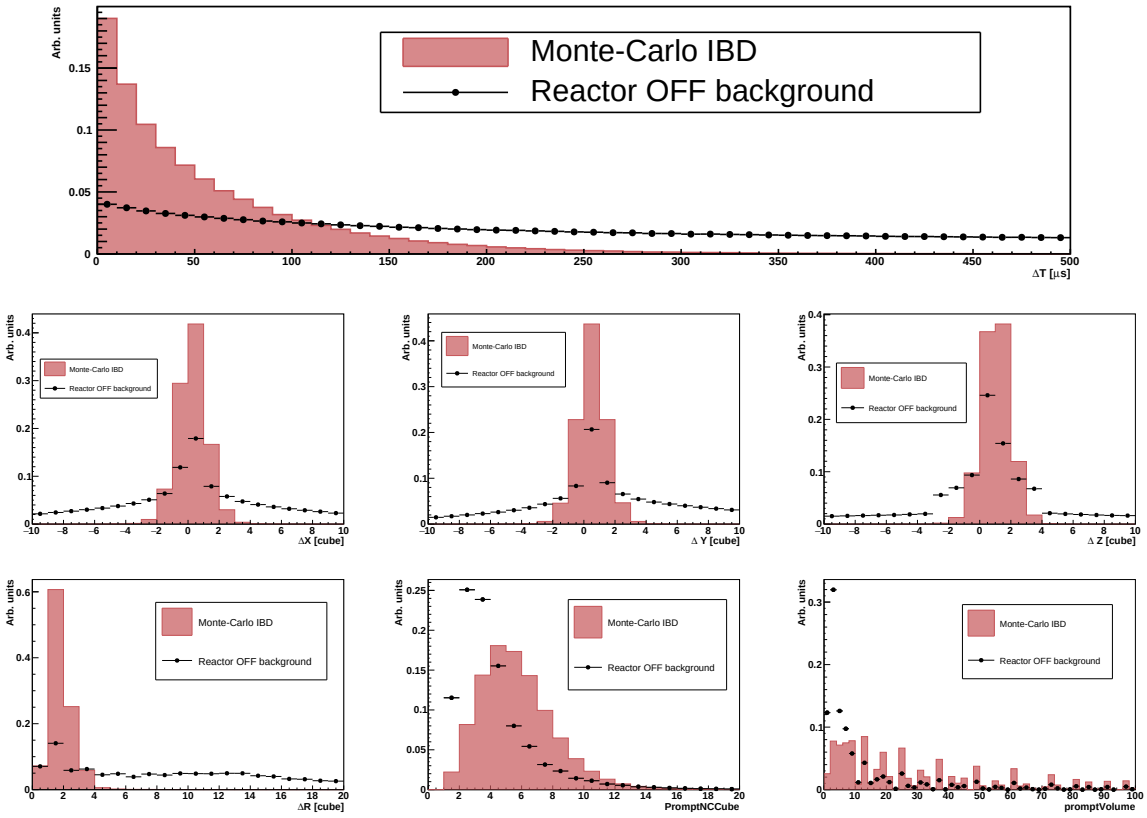


Figure 4.1: Distribution of the low-level discriminative variables with areas normalized to one for four days of reactor OFF background (back points) and IBD Monte-Carlo (red). All background sources are comprised inside the reactor off background. The error bars on the data points are drawn on the plots but masked by the markers.

surement of the AC_energy. Further studies are being developed to choose the best estimator for the final analysis. This can be done by running a full analysis with the three estimators to select the one that gives the best sensitivity to an antineutrino oscillation.

4.3 Background measurement

As presented in section 2.5.2.1, mainly three types of backgrounds are faced for an antineutrino analysis in the SoLid experiment: accidental background, BiPo background and atmospheric background. Contrary to the section 2.5.2.1 that presented the general sources of the different backgrounds, this section will present their specificities in the context of the SoLid analysis and the tools developed to quantify and reduce them. The study of the different background properties is made from data-driven measurement to avoid any possible mismatch with background simulation. To do so, a few days of reactor OFF data from the open dataset will be used in this section.

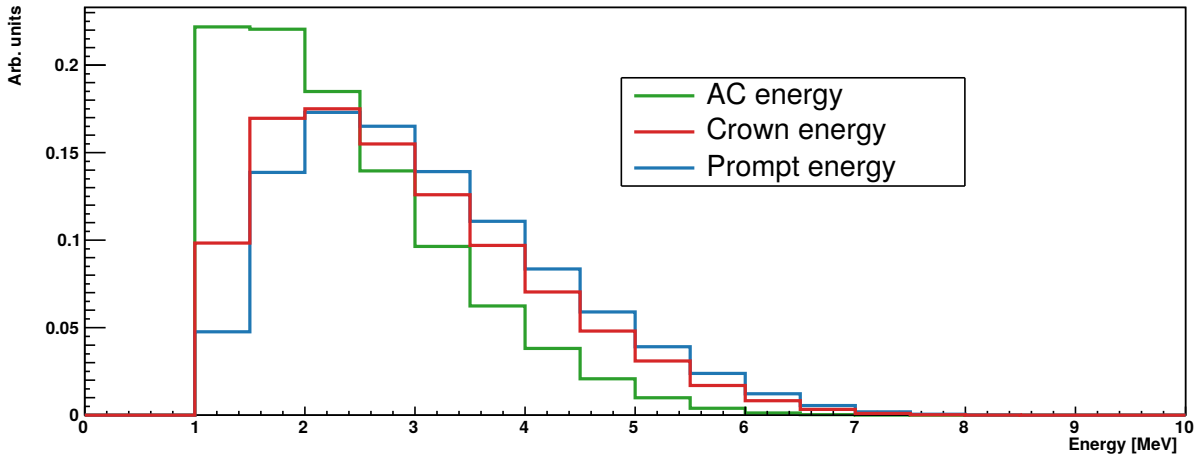


Figure 4.2: Different energy estimators distribution from IBD Monte-Carlo. AC_energy is the energy of the cube, with the higher energy deposit in the event. $Crown\_energy$ is the sum of the energies of the cubes in the 3x3x3 volume around the highest energy cube. $Prompt\_energy$ is the sum of the energies of all the channels of the events.

4.3.1 Accidental background

The accidental background is generated via accidental time coincidence between the NS cluster and ES cluster. As both clusters have no correlation in time or space, the accidental background will show a flat ΔT distribution. A first method was used to compute the accidental rate of a signal sample, selecting the negative part of the ΔT distribution and rescaling it to the signal region. The time window to evaluate the accidental background is:

$$\Delta T \in [-200, -100]\ \mu\text{s} \tag{4.1}$$

This window ends at 100 µs due to correlated signal close to $\Delta T = 0$. Those correlations can be induced by cosmogenic events, where muons create a spallation of a neutron in nuclei and the nuclei decay after the neutron capture. On the figure 4.3 (top) is represented the method to extract the accidental component in the positive ΔT time window (correlated time window). On the figure 4.3 (bottom), one can see that the Accidental time window is not perfectly flat. The effect is about 4% between the extremities of the time window.

In order to avoid this issue, another method has been developed using *false positive neutron triggers* (FPNTs). The FPNTs are events where the neutron trigger condition was fulfilled but no NS signal were reconstructed. This can happen with a muon waveform, where the long tail of the waveform can match the number of peak over threshold condition of the neutron trigger, as presented in section 2.4.3.2. A FPNT is selected only if a muon is identified in the same plane and at the same time where the FPNT has been detected. This condition is used to select muon induced FPNTs (and not neutron induced triggers that would not have been reconstructed for instance). The NS-ES coincidences are then built in the same way for FPNT clusters and NS clusters, as detailed in the section 2.7.2.4: in the buffer window opened by a FPNT, a coincidence is build for all ES that are in the time window $\Delta T_{FPNT-ES} \in [-200, 500]$ µs. To quantify the level of accidental coincidences induced by the NS clusters, one FPNT cluster is selected per real NS cluster that was reconstructed in the same plane during the reconstruction cycle (see section

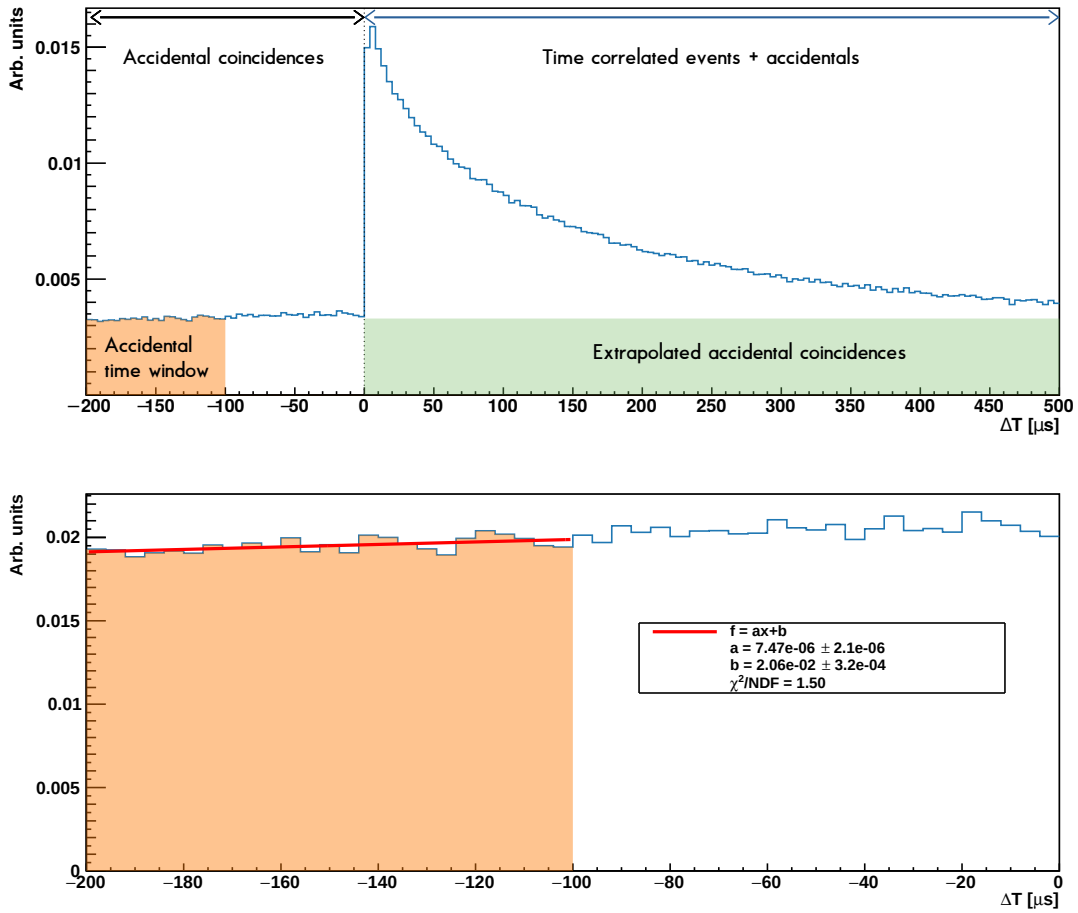


Figure 4.3: Top: ΔT distribution in a Reactor OFF sample. The orange band is the measurement of the accidental component in the $[-200, -100]$ μs time window, and the green band is the extrapolation of the accidental component in the correlated time window. Bottom: distribution zoomed around the accidental time window with a degree one polynomial fitted.

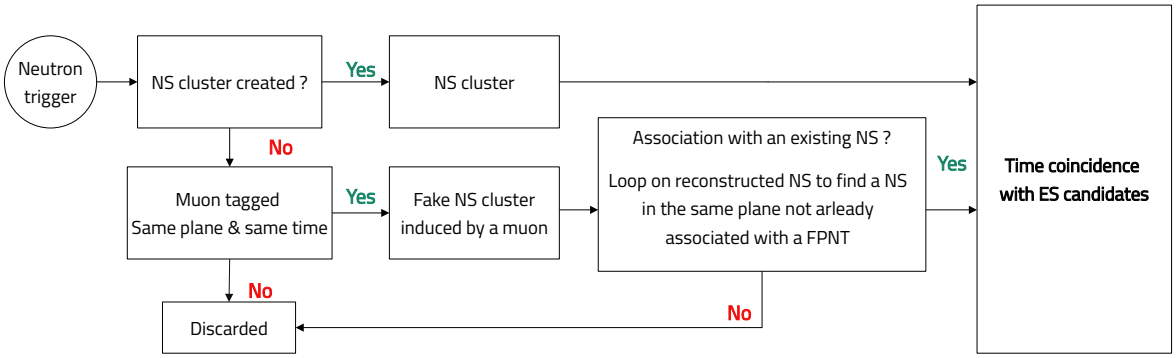


Figure 4.4: Scheme of the FPNTs-ES coincidence creation during a reconstruction cycle.

2.7). Due to the rate of FPNTs in the detector, in the end or an eight minutes run, less than 2% of the NS are not matched with a fake NS induced by a FPNT. A scheme of the FPNT clusters creation process can be found on figure 4.4.

As the FPNT coincidences are mainly coincidences with a muon crossing the detector, one can expect three main components that can be seen on figure 4.5:

- at $|\Delta T_{FPNT-ES}|$ of a few micro seconds, there is the correlation of the FPNT with itself. It consists in an ES created by a mistrack of the same muon or a Michel electron induced by the muon that created the FPNT. As it happens below 5 µs this region will be excluded for the estimation of the accidental contamination of the correlated events.

- in the negative ΔT region, there is a small correlation of after muon events that can be cosmogenics events or proton recoils from spallation neutrons. This small correlation adds up to a flat un correlated distribution and explains why the FPNT-ES coincidence rate here is slightly higher than the NS-ES coincidences rate.

- in the positive ΔT region, only a flat distribution induced by accidental coincidences is seen.

The accidental rates will thus directly be measured, applying the final selection on the FPNT-ES coincidence events.

Contrary to the correlated events, the NS and the ES of the accidental background events are not correlated in space. Thus, the distance between the two signals, encoded in the ΔR variable, is larger than the correlated signal, as can be seen on figure 4.6. On this figure, one can indeed see the expected IBD signal ΔR is much lower than the average ΔR of the accidental background. This will be the main variable to reduce the accidental background.

4.3.2 BiPo Background

The BiPo background is induced by the decay of the ^{214}Bi to ^{214}Po and to ^{210}Pb (see section 2.5.2.2). A large part of the BiPo background is generated by a contamination of the ZnS screens, inside which the BiPo decay happens. The electron resulting of the decay is detected in a neighbouring PVT cube, and an alpha particle is emitted and detected inside the screen. As the ZnS screens are located on two faces of each cube, one on the X direction and one on the Z direction, three main topologies can be seen on BiPo

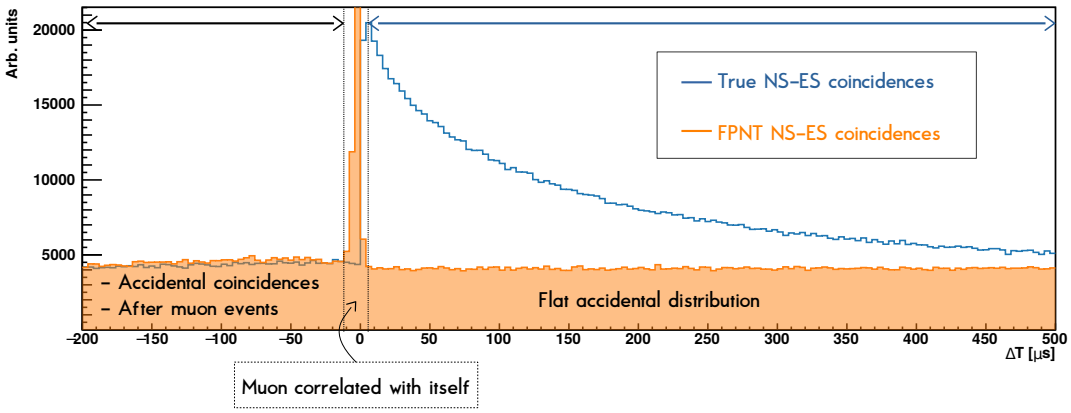


Figure 4.5: ΔT distribution in a four days Reactor OFF data sample in blue. ΔT distribution of the FPNT coincidences in the same sample in orange.

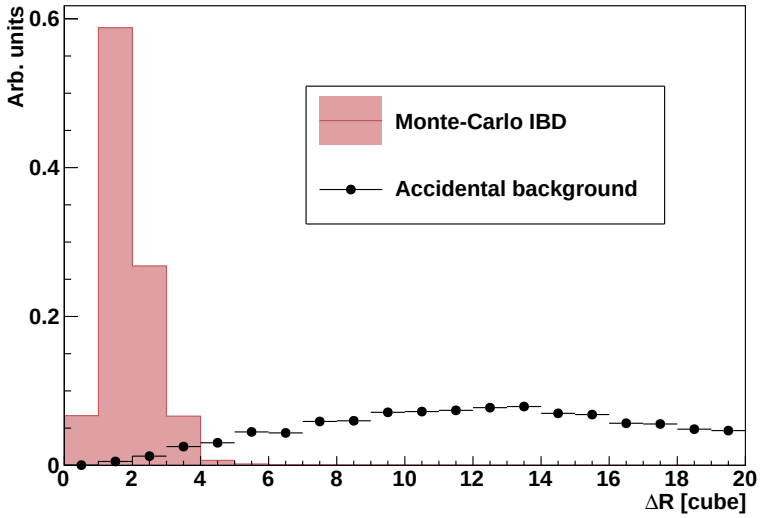


Figure 4.6: ΔR for an accidental sample of 20 days of reactor OFF data in black points and for IBD Monte-Carlo in red.

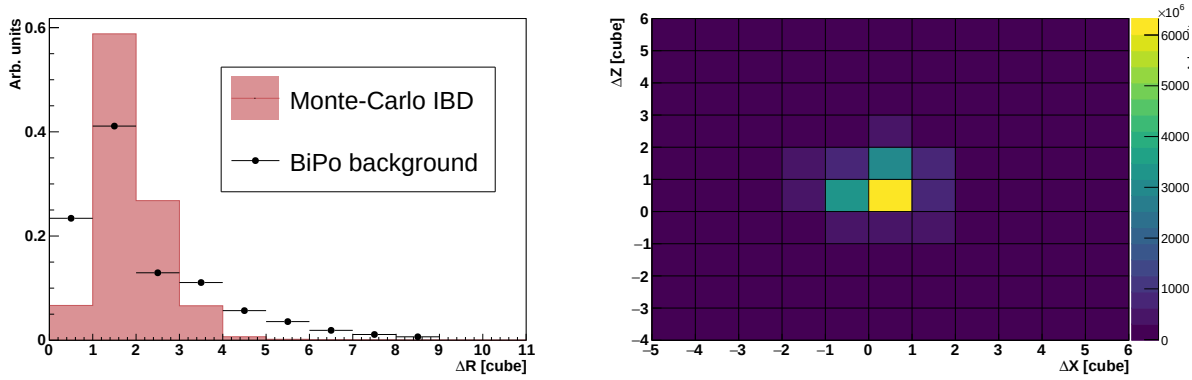


Figure 4.7: Left: ΔR for a BiPo enhanced reactor OFF sample in black points and for Monte-Carlo IBD in red. Right: ΔZ vs ΔX distribution for BiPo background

backgrounds:

$$\begin{aligned} \Delta(X,Y,Z) &= (0,0,0) \\ \Delta(X,Y,Z) &= (-1,0,0) \\ \Delta(X,Y,Z) &= (0,0,1) \end{aligned} \tag{4.2}$$

This selection is used to extract a BiPo enhanced sample from the reactor OFF data. Each topology corresponding to the electron being detected in the same cube as the alpha or crossing the tyvek sheet around the cube in the X or Z direction and being detected in a neighbouring cube. On figure 4.7 are represented the ΔR (left) and the ΔZ vs ΔX (right) distributions for a BiPo enhanced background sample. The three BiPo cases can be observed on the right figure. A selection on $\Delta R > 0$ can remove some BiPo background, as can be seen on the left figure. This removes the first main topology of BiPo events presented in equation 4.2. The two other topologies are also populated by the IBD events, and removing them would impact too much the antineutrino selection efficiency. The enhanced BiPo sample will thus be used to select a pure sample of BiPo background to develop discrimination tools that will use both the ΔT and the NS shape discriminator for the IBD analysis.

The NS signal is induced by the energy deposits of the alpha resulting of the decay of the ^{214}Po. The energy deposited in the ZnS scintillator by the alpha is 7.7 MeV. For a neutron capture by the ^{6}LiF:ZnS(Ag) layer, the alpha and tritium, generated by the ^{6}Li breakup, deposit an energy of 4.8 MeV. Due to attenuation in the screens, light absorption in the PVT, light losses in the fibres, the amplitudes of the ZnS signals fluctuate too much to use it as a direct energy selection [101]. However, the stopping power of the ZnS layers is higher for the products of the ^{6}Li breakup due to their energies than for the BiPo alpha. This leads to nuclei excited in higher states in the ZnS by the alpha and tritium from ^{6}Li than for the BiPo alpha. The scintillation signal resulting of a neutron capture has thus a larger decay time compared to the scintillation induced by the BiPo alpha as shown on figure 4.8 (left). The neutron induced distribution is built with AmBe calibration data (presented in the section 3.2) to have a pure neutron induced sample and the BiPo enhanced sample is obtained using the enhanced BiPo topologies presented in the equation 4.2. A pulse shape discrimination variable is then created by computing the ratio of a waveform integral in the $[0, 87.5]$ µs window and the $[0, 7.5]$ µs window. The variable is called BiPonisher and is shown on figure 4.8 (middle). The BiPonisher was the original BiPo discriminator used

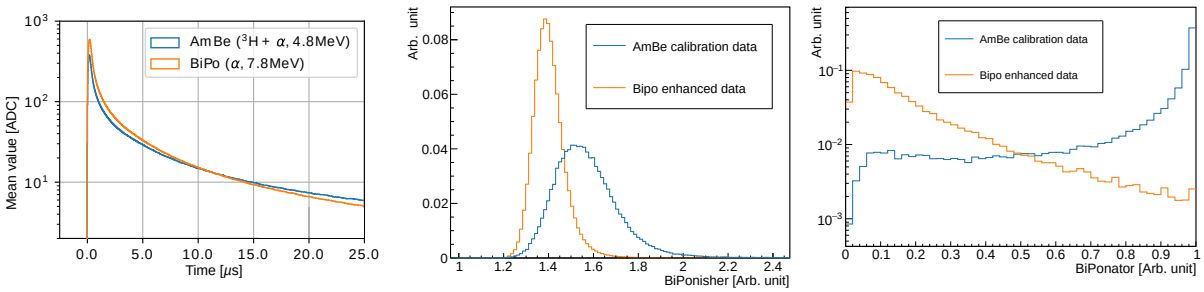


Figure 4.8: Left: Average neutron waveform for AmBe calibration data (blue) and BiPo enhanced sample (orange) [84]. Middle: BiPonisher discrimination variable distribution for AmBe calibration data and BiPo. Right: BiPonator discrimination variable for AmBe calibration data and BiPo.

in the SoLid experiment, however because the BiPonisher shapes for alphas and neutrons are still close, the discrimination between both types of scintillation could be improved. To improve the discrimination power between the neutron induced waveforms and the alpha induced waveforms, the colleagues of LPC Caen used a 1D convolutional neural network to have a more performant discrimination. The distribution of the newly created variable, called BiPonator, for the different types of NS can be seen on figure 4.8 (right). This new variable is expected to be less sensitive to baseline variation over time and improved the alpha rejection efficiency of a factor 3 for similar neutron efficiency.

The other main specificity of the BiPo used in the analysis is the time constant of the ^{214}Po decay. With an half life of the ^{214}Po of 164 µs, the ΔT distribution of the BiPo will follow a decreasing exponential with a time constant of $\tau = 237$ µs. The neutron capture time constant is composed of a fast time constant $\tau_{fast} \simeq 11$ µs for the epithermal neutron capture and a slow thermalization time of $\tau_{slow} \simeq 64$ µs. The ΔT variable is thus a powerful discriminative tool. On figure 4.9 is represented the different ΔT shapes of the different backgrounds for reactor OFF background where each component is obtained by a ΔT fit. Both neutron induced and BiPo backgrounds ΔT distributions follow an exponential shape but with a time constant larger for BiPo. A large ΔT selection will thus tend to select more BiPo while a short ΔT selection will select more neutron signal from atmospheric background and IBDs.

The variables ΔT and BiPonator will be used to build a sample rich in neutron induced events for the antineutrino analysis. To evaluate the level of BiPo contamination inside this sample, a sideband will be created in the 2D space ΔT vs BiPonator. The distribution of the ΔT vs BiPonator is presented on figure 4.10, with low values of ΔT and high values of BiPonator defining the neutron signal region and high values of ΔT with low values of BiPonisher defining the BiPo region. The selection used for both regions are given in the table 4.3.

Variable	Neutron	BiPo
ΔT [µs]	$5 < \Delta T < 141$	$250 < \Delta T < 500$
BiPonator	0.7 < BiPonator < 1	0 < BiPonator < 0.2

Table 4.3: Signal and BiPo selections in the ΔT vs BiPonator space.

The contamination of BiPo in the neutron region is then given by the number of BiPo events in the

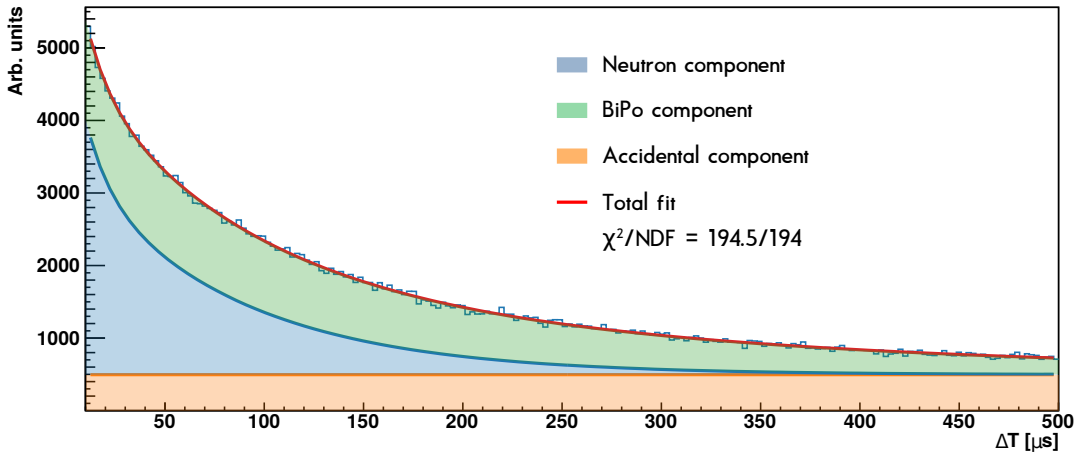


Figure 4.9: ΔT distribution for four days of reactor OFF data. In orange is represented the flat accidental contribution. Stacked over it, in blue, is the atmospheric background component composed of two exponentials for the epithermal neutron capture and the neutron capture after thermalization. The final stack, in green, represents the BiPo components of the background composed of one exponential due to the ^{214}Po decay time.

BiPo sample, scaled with the following scaling factor:

$$S_{BiPo\to neutron} = \frac{\int_{0.7}^{1} R_{BiPo}(\text{BiPonator})\text{dBiPonator}}{\int_{0}^{0.2} R_{BiPo}(\text{BiPonator})\text{dBiPonator}} \cdot \frac{\int_{5}^{141} \exp\left(\frac{-t}{237}\right)dt}{\int_{250}^{500} \exp\left(\frac{-t}{237}\right)dt} \tag{4.3}$$

With $R_{BiPo}(\text{BiPonator})$ the distribution of the BiPonator variable for a pure BiPo sample.

4.3.3 Atmospheric background

The last type of background to treat is the atmospheric background. It is present in all the energy range considered for the IBD analysis, and is a dominant background at energies higher than 3 MeV. Contrary to the accidental background and the BiPo background, there is no way to measure directly the contamination of the atmospheric background inside reactor ON data via a direct event selection or a sideband. As for the IBDs, the atmospheric background consists in a correlated coincidence between an electromagnetic and a neutron signal. The characteristic time of the process will thus be the same for antineutrino detection and atmospheric induced coincidences. For the antineutrino analysis, the atmospheric background is obtained after subtraction of both accidental background and BiPo background. In the remaining sample, the reactor OFF is composed of the atmospheric background and the reactor ON is composed of both the atmospheric background and the IBDs. Because the rate of atmospheric background measured is correlated with the molecule density in the atmosphere, hence the atmospheric pressure, it is a background that will vary through time. It is thus not possible to easily extrapolate the level of atmospheric background of a reactor ON data sample, given the atmospheric background yield in a reactor OFF data sample. To extrapolate the level of contamination inside a sample, one can however build a pressure model:

$$R_{ATM} = a \times P_{atm} + b \tag{4.4}$$

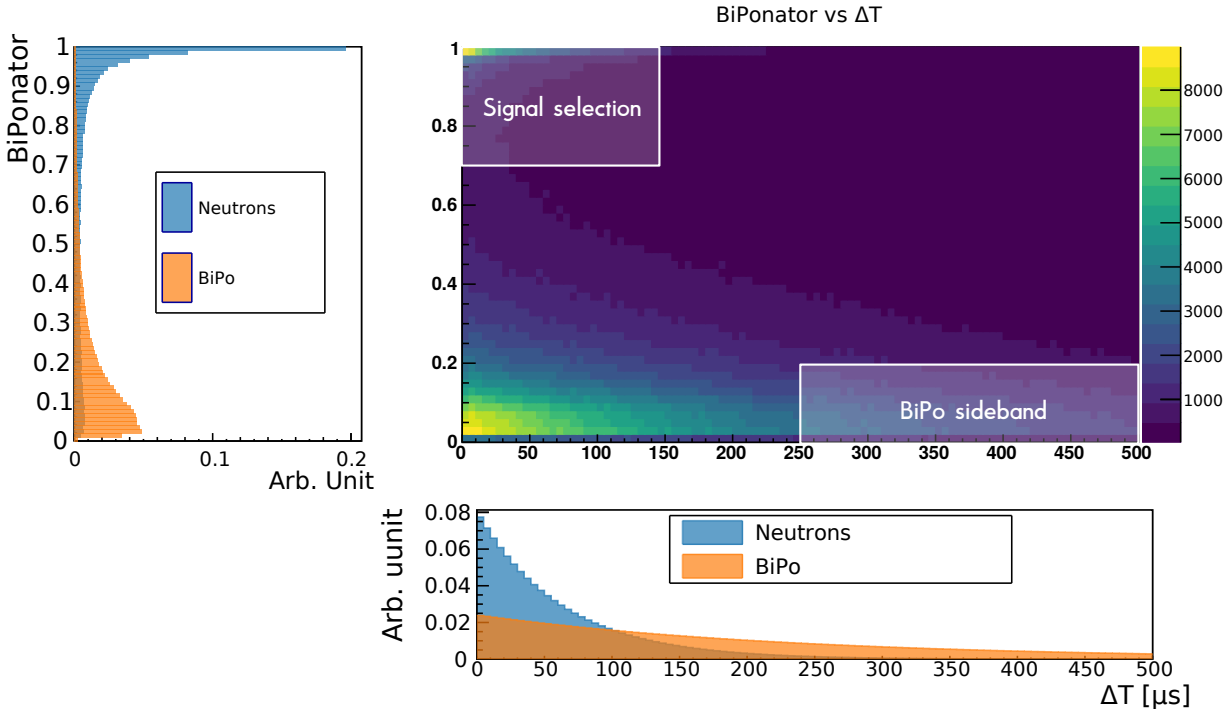


Figure 4.10: 2D histogram: ΔT vs BiPonator distribution for ROFF data. The 1D distribution of both ΔT and BiPonator are represented normalized to one for both neutron induced (blue) and BiPo induced (orange) events. The two white rectangles on the 2D histogram represent the ΔT and BiPonator selection used for the antineutrino analysis and for the BiPo sideband.

with R_{ATM} the rate of atmospheric events per day, P_{atm} the atmospheric pressure averaged per day and a, b two constants to fit in the model. In order to build a pressure model with a larger statistics and atmospheric background rates for given pressure values, one can want to add reactor ON data to build the model. Furthermore, some contamination by other sources of backgrounds, independent of the pressure, could still happen in the atmospheric background samples. For instance a BiPo contamination constant in time and not fully subtracted from the data. For this reason a model with the average rate and pressure removed from the measurement has been built:

$$R_{ATM} - \overline{R_{ATM}^{RON/ROFF}} = a \times (P_{atm} - \overline{P_{atm}}) = f_{model}(P_{atm} - \overline{P_{atm}}) \tag{4.5}$$

where $\overline{R_{ATM}^{RON/ROFF}}$ represents the average of the atmospheric background on the Reactor ON or OFF datasets considered and $\overline{P_{atm}}$ is the pressure average over the whole dataset. Both reactor ON and OFF rates are treated independently because of the IBD signal that is a component of $\overline{R_{ATM}^{RON}}$ but not of $\overline{R_{ATM}^{ROFF}}$. Both averages should thus contain the mean atmospheric rate in both periods, a possible background constant through time and IBD events in $\overline{R_{ATM}^{RON}}$. In order to rescale the pressure model to the actual atmospheric background rate in both samples, the average atmospheric rate from reactor ON and OFF are supposed to be similar. The rate of atmospheric background is thus computed by rescaling the pressure function with the average over the reactor OFF dataset that is expected to be close to the average atmospherics rate in the reactor ON sample:

$$R_{ATM}^{model} = f_{model}(P_{atm} - \overline{P_{atm}}) + \overline{R_{ATM}^{ROFF}} \tag{4.6}$$

Where R_{ATM}^{model} is the predicted number of atmospheric events. f_{model} encodes only the effect of the pressure variations in the sample, while $\overline{R_{ATM}^{ROFF}}$ is used as a global reference rate.

4.3.4 Atmospheric background in the open dataset

In order to check the validity of a pressure model on atmospheric background, a selection has been made on the *open dataset* to remove the BiPo contamination:

$$\begin{aligned} Energy &\in [7, 20]\ \text{MeV} \\ \Delta T &\in [10, 400]\ \mu\text{s} \\ \Delta R &\in [0, 5[\end{aligned} \tag{4.7}$$

By selecting an energy larger than 7 MeV it allows the study of data free of BiPo contamination and without IBDs. The ΔT and ΔR selection are then used to remove a large part of the accidental background remaining in the sample, with a selection on correlated events. The sample should then be only composed of atmospheric events and a small contamination of accidental coincidences. The level of accidentals is nevertheless evaluated with the FPNT method and the accidental rate is removed to the data rate as described in the section 4.3.1. On the figure 4.11 can be seen the ΔT distribution before accidental subtraction, where one can see that there is no more BiPo background. In the sample, one can also note that less than 2% accidental contamination. In the end, with this selection and after a subtraction of the accidental background, one can consider that a pure sample of atmospheric background is available to test the pressure model. Some days can have lower exposure time than the others due to technical issues

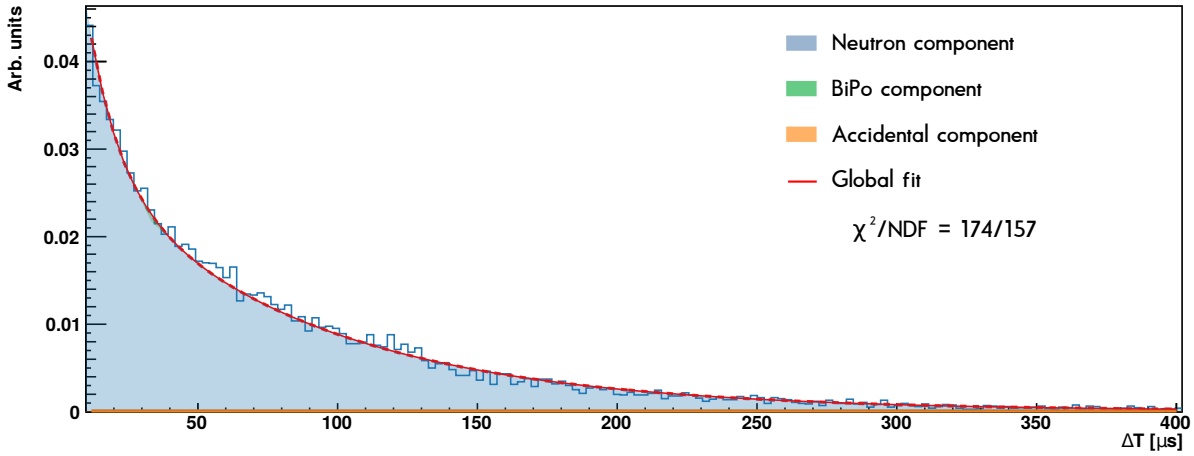


Figure 4.11: ΔT distribution on rector OFF data with the three background components fitted with the selection for optimized atmospheric background described in equation 4.7. Due to the energy selection, no BiPo contamination can be found in the sample and less than 2% of accidental background is found in the sample.

in the data taking or runs discarded after a data quality test. This effect can artificially lower the atmospheric background rate for some days, so the exposure time is corrected on all the data. This selection is a good way to study the pressure model in a region that will be excluded by the analysis. The models derived in this section will not directly be used in the analysis later, as the model depends directly on the event selection, and will have to be rebuilt.

The time evolution atmospheric background rate and the pressure can be seen on the figure 4.12 (top). The transition RON-ROFF at the 10^{th} of July cannot be seen due to the selection performed. A clear anti-correlation can be seen between the atmospheric background and the pressure on the bottom plot of figure 4.12. On this sample, no clear reactor ON/OFF difference can be seen due to the energy selection cut that removes the reactor induced events. The pressure model for the open dataset is represented on the bottom graph, a global fit is performed on the combined reactor OFF and ON data using the model of equation 4.5. From the fit result, with a good χ^2, the pressure model describes well the variation of the atmospheric background rate in the *open dataset*.

4.3.5 Stability of the pressure model on all reactor OFF

In order to check the stability of the model on a larger dataset, all the reactor OFF dataset has been used to generate the pressure model. The pressure model computed over all the data can be seen on the figure 4.13. On the left plot, the model has been fitted to the data, one can see that the χ^2 of the global fit is worse than when it was computed only with the open dataset. On the figure 4.13 (right) can be seen the relative difference between the atmospheric events rate given by the model and measured with the selection of equation 4.7, with the following equation:

$$\frac{R_{ATM}^{model} - R_{ATM}^{Data}}{R_{ATM}^{Data}} \tag{4.8}$$

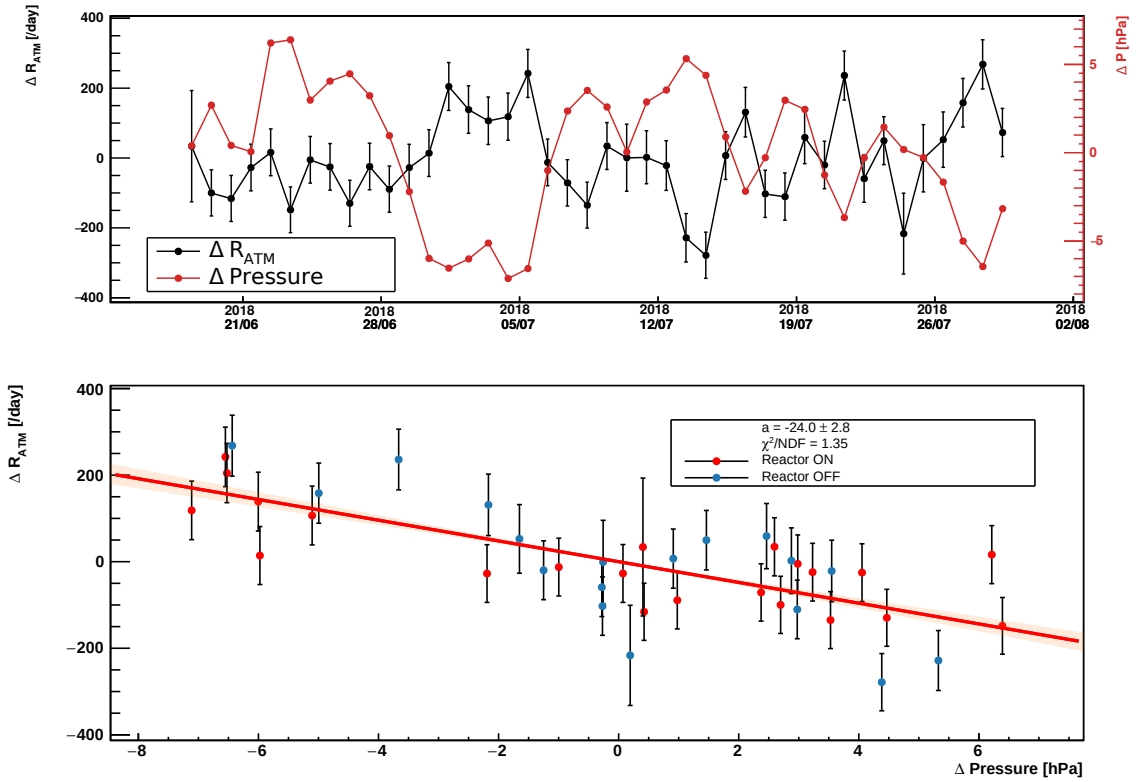


Figure 4.12: Top: time evolution of $R_{ATM} - \overline{R_{ATM}^{RON/ROFF}}$ and $P_{atm} - \overline{P_{atm}}$ through time in the open dataset given the selection of the equation 4.7. Bottom: pressure model for the open dataset. Each point represents a day of data corrected by the exposure time. The red (blue) points represents reactor ON (OFF) days. The red line is the pressure model and the orange band is the error at 1 σ on the slope of the model.

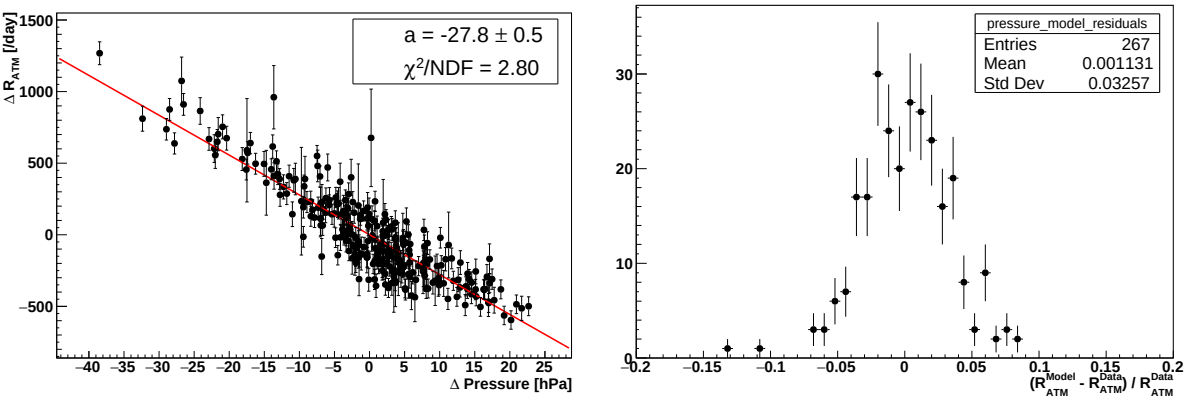


Figure 4.13: Left: atmospheric background rate (per day) as a function of the atmospheric pressure, with the average value of both variables subtracted. Right: residuals of the pressure model computed as in the equation 4.8.

The difference between the model and the measured data is less than 1% in average, for a spread of 3.3%. An uncertainty could be derived from the fitter uncertainty. One can compute the maximum and minimal slope from the model. The uncertainty on the slope gives a maximal uncertainty of 20 events per day. This has to be compared to the average atmospheric rate of 4475 events per day, giving an uncertainty below 1% from the fit itself. The uncertainty on the measurement is thus driven by the statistical uncertainty on the average atmospherics rate, which is 1.5% on this sample. However, it seems necessary to improve the understanding of the larger χ^2 of the model with all the reactor OFF data. With the aim to improve the pressure model, it can be studied period by period as represented on the figure 4.14. A reactor OFF period is defined as in the table 4.2 and consists in the time period between two reactor ON cycles. On the figure, most of the periods are following the average pressure model, which is a good hint to use a global model to evaluate the atmospheric background contamination on reactor ON data. However, a first effect can be seen on the $2-3$ period. It is known that the detector was not totally stable during this period, with large humidity variations and unstable detector response. That could explain why the data on this period does not seem to follow the average model. This period will be removed from the dataset used for the antineutrino analysis. Then, for some periods ($3-4$ or $13-14$) the atmospheric rate is systematically lower or higher than the averaged model. This could be a hint that the atmospheric background rate also depends on environmental parameters other than the pressure, such as the temperature or long term seasonal effects [102]. Furthermore, small variations in the detection efficiency of the atmospheric rates, due to small energy scale variations between two calibration campaigns, for instance.

4.3.5.1 Pressure and temperature model

A model with the ratio of pressure over temperature has been tested on figure 4.15 following the ideal gas relation:

$$n \propto P/T \tag{4.9}$$

with n the gas density, P the gas pressure and T the atmospheric temperature (in K). Compared to the pressure only model, the pressure and temperature model gives a worse description of the atmospheric background: the linear correlation is much less clear with a reduced χ^2 (chi^2 divided by the number of degrees of freedom) of nearly 13 and the spread of the residuals twice higher than before. The model might be too simplistic to describe the atmospheric background with the temperature added. This could also be related to the available pressure and temperature measurements, indeed the sensors are located at the ground level and not directly in the atmosphere.

4.3.5.2 Independent models

If the temperature is not a straightforward variable to use in the model, another method can be used to reduce the effect of long term seasonal effects. Each period can be treated independently by subtracting the average pressure and average rate over the period instead of the average over the whole reactor OFF data. For a given reactor OFF or ON period i the subtracted rates and pressure per day become:

$$\begin{aligned} \Delta R^{ind}_{ATM} &= R_{ATM} - \overline{R^i_{ATM}} \\ \Delta P^{ind} &= P - \overline{P^i} \end{aligned} \tag{4.10}$$

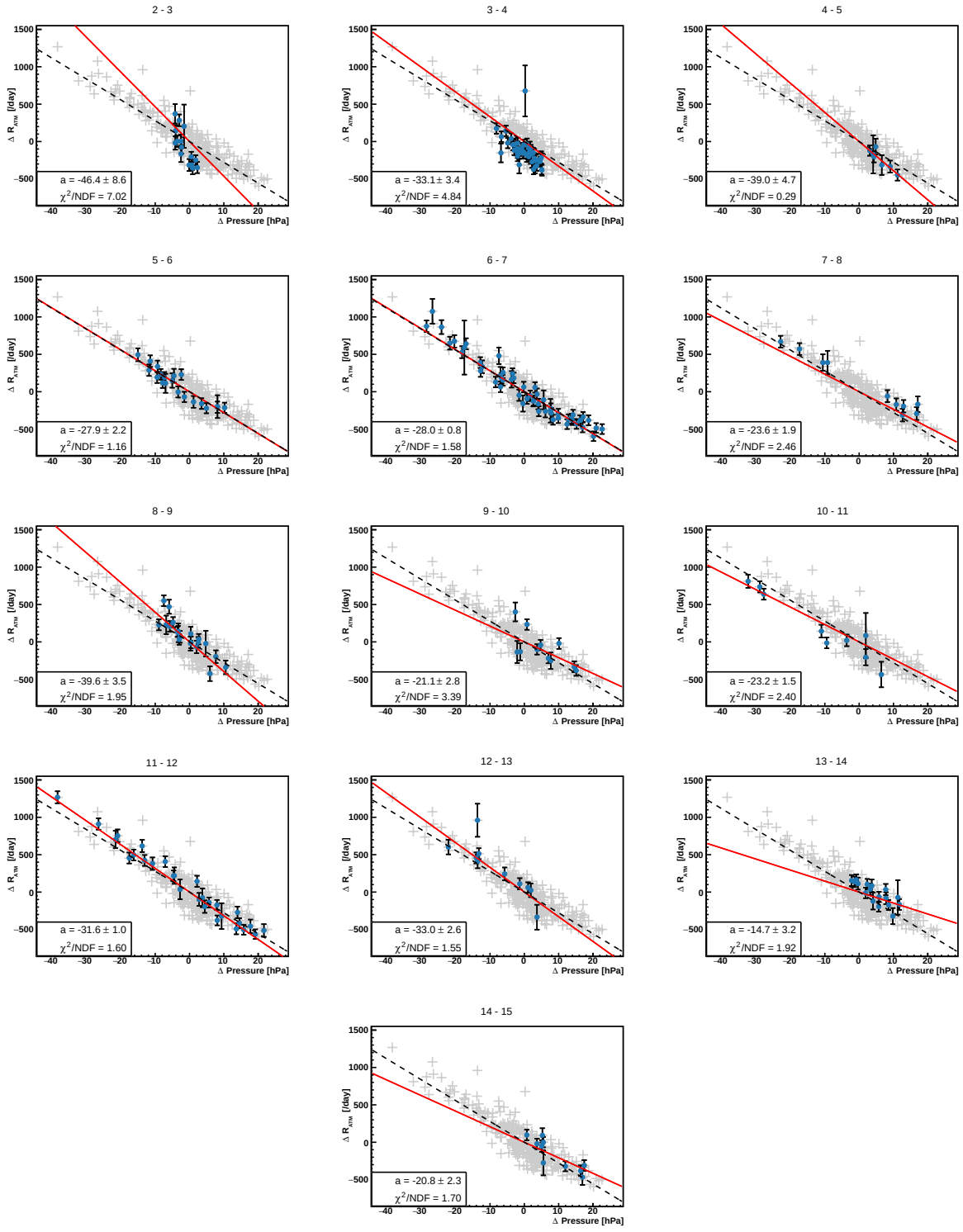


Figure 4.14: Atmospheric background rate as a function of the atmospheric pressure for each reactor OFF period. The grey points represent the rate for the whole Phase I reactor OFF dataset and the red line represent the pressure model fitted on only the data sample, the black doted line represents the global fit on all the data.

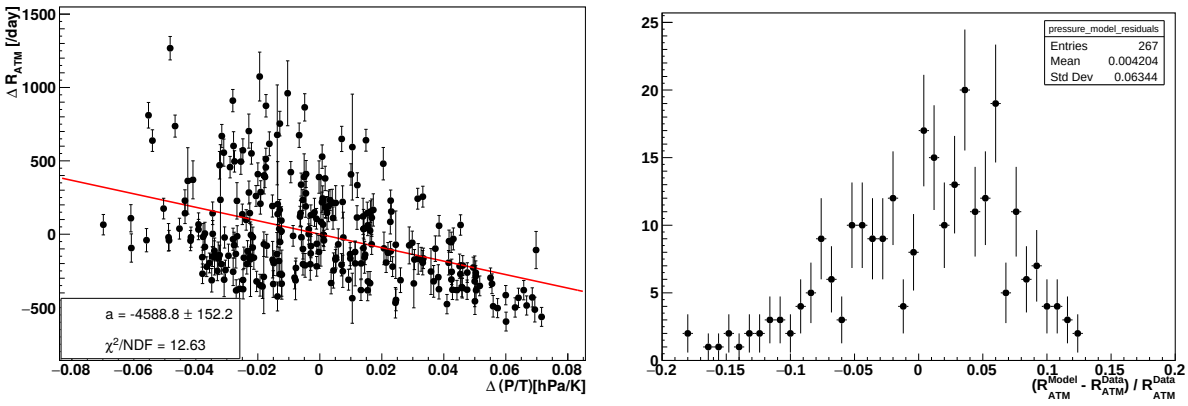


Figure 4.15: Left: atmospheric background rate per day as a function of the ratio between the outside ground temperature at BR2 and the pressure. The average of the variables over the whole dataset are subtracted. Right: residuals of the pressure and temperature model computed as in the equation 4.8.

with $\overline{R^i_{ATM}}$ and $\overline{P^i}$ the atmospheric events rate and the pressure averaged over the time period i equals to a reactor cycle for reactor ON or the time between two cycles for reactor OFF. Once the daily atmospheric rates have been subtracted by the average rate on the period, all the periods can be gathered together to build a new model f^{ind}_{model} in the same way as before:

$$f^{ind}_{model} = a \times \Delta P^{ind} \tag{4.11}$$

In order to retrieve the predicted atmospheric background rate, one still needs to use the average atmospheric rate computed on the whole reactor OFF period as a reference rate. The evaluation of the atmospheric background rate with the pressure model will then be:

$$R^{model,ind}_{ATM} = f^{ind}_{model}(\Delta P^{ind}) + \overline{R^{ROFF}_{ATM}} \tag{4.12}$$

With $\overline{R^{ROFF}_{ATM}}$ the average reactor OFF rate on the whole dataset used in the analysis.

Treating each period independently would reduce the spread of the data but comes with a risk of losing the prediction power of the new pressure model. In order to test the prediction of the model, it is built on the data from the reactor OFF periods $3-4$, $4-5$, $5-6$, $6-7$, $7-8$ and tested on the other reactor OFF periods. The residuals are computed following the equation 4.8. On the figure 4.16 (left) are represented both the new model on the days used to build the model (top) and the days rates not used to build the model (bottom). For the dataset used to build the model, the model gives a better description of the data than the first model presented in the figure 4.13 with a reduced χ^2 at 1.2. On the right panel of the figure are represented the residuals for both periods. For both periods, the average of the residuals is lower than 1% and the spreads of the residuals are at similar level between the dataset used to build the model and the other one, with a spread of 2.4% for the first one and 2.6% for the second one. This shows that the prediction power of the model f^{ind}_{model} is still good. The residuals on all the reactor OFF data is represented in the figure 4.17, with the period $2-3$ included for comparison purpose. It shows a spread of 2.7% of the residuals, it is bigger than the spread on the test period because of the period $2-3$ added in the sample. Compared to the residual of the first model (3.3%) this shows a better characterization of the atmospheric background.

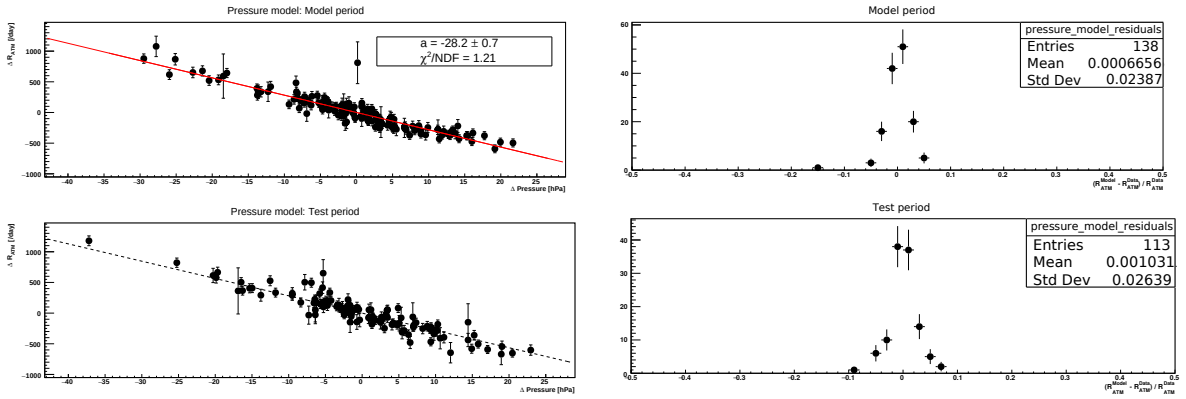


Figure 4.16: Left: atmospheric background rate variations as a function of the pressure variation. The variations are computed according to equation 4.10. On the top panel is represented the dataset used to build the model, and on the bottom panel is represented the dataset used to test the model. The rate and pressure of each day point is subtracted by the averaged rate and pressure on the period containing the day. Right: residuals for the dataset used to build the model and for the other dataset.

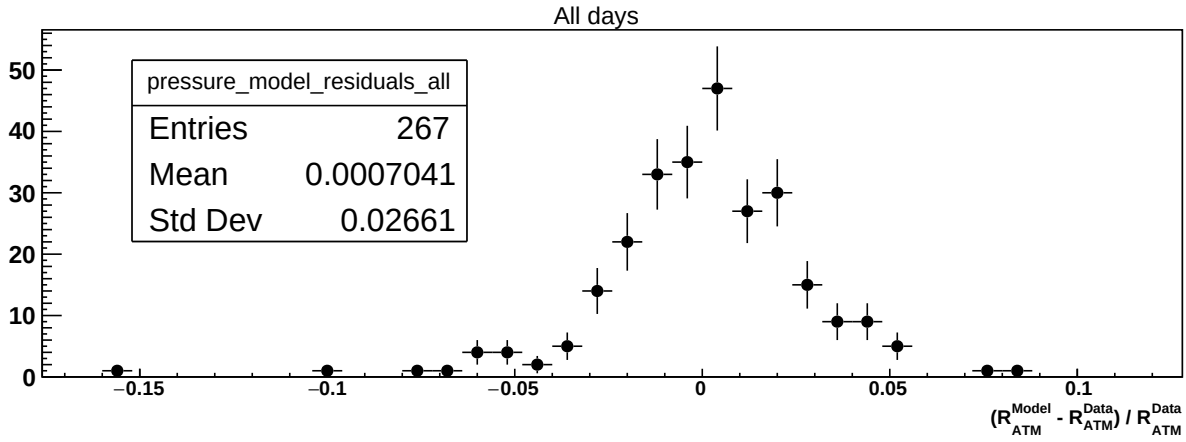


Figure 4.17: Residuals for all the reactor OFF data, except the period $2-3$, of the pressure model using localized average.

The breakdown of the pressure model for each reactor OFF period is represented on figure 4.18. The identified issues with the periods $3-4$ and $13-14$ are then fixed by treating the cycles independently with rates and individual models closer to the global pressure model: the average model gives a better description of each sub dataset. For the antineutrino signal extraction in the section 4.6.1, the atmospheric model presented in equation 4.12 is used to predict the reactor ON atmospheric background contamination.

4.4 Data quality

During the two years of data taking, the environmental parameters have been fluctuating and the stability of the detector response has fluctuated through time. To ensure that all the data used in the final analysis was taken with a stable detector in comparable conditions, a data quality selection is necessary to remove bad data. In this section we will present the criteria chosen by the collaboration to select or

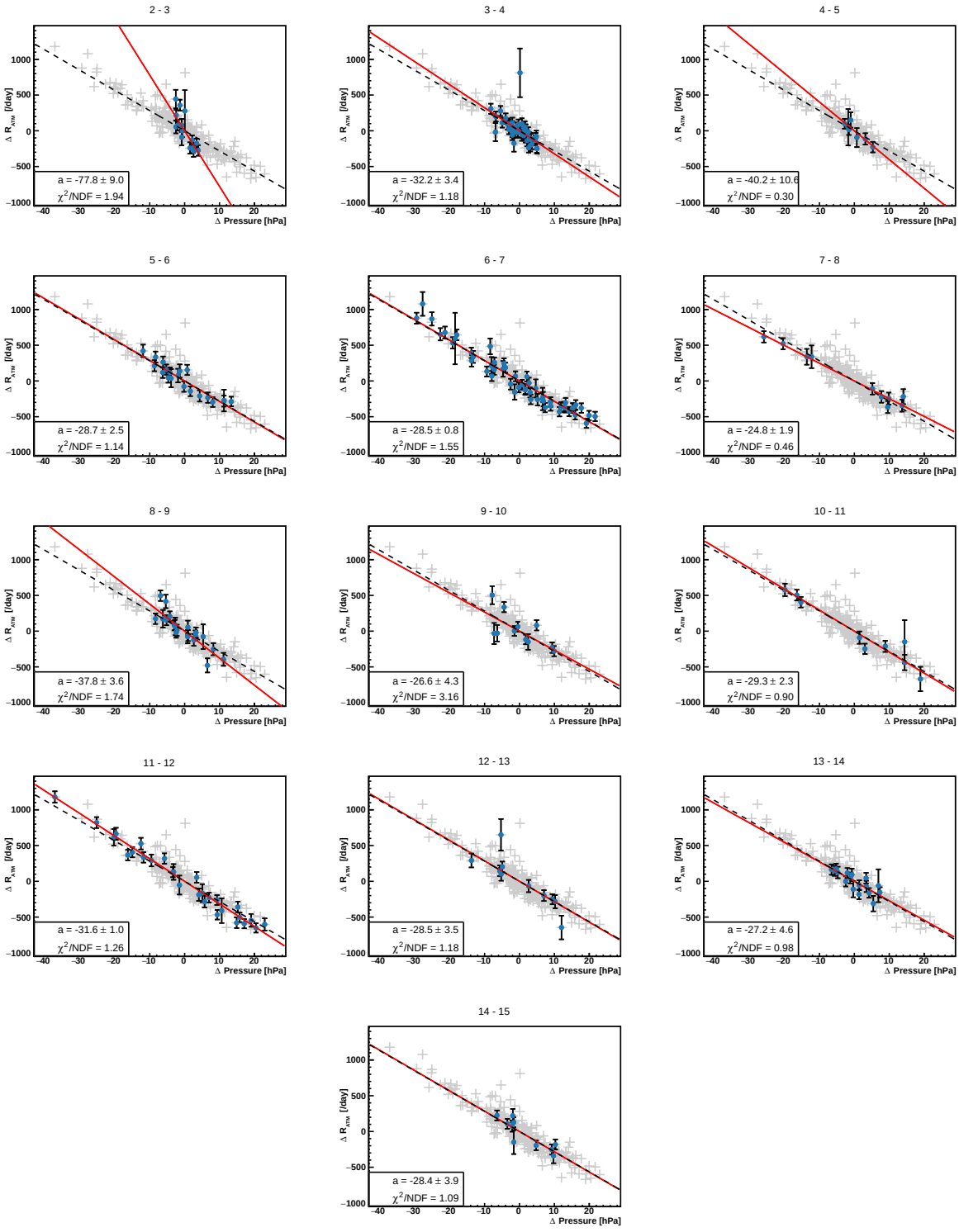


Figure 4.18: Atmospheric background rate as a function of the atmospheric pressure for each reactor OFF period when all periods are treated independently. The grey points represent the rate for the whole Phase I reactor OFF dataset and the red lines represent the pressure model fitted on only the data sample, the black doted line represents the global fit on all the data.

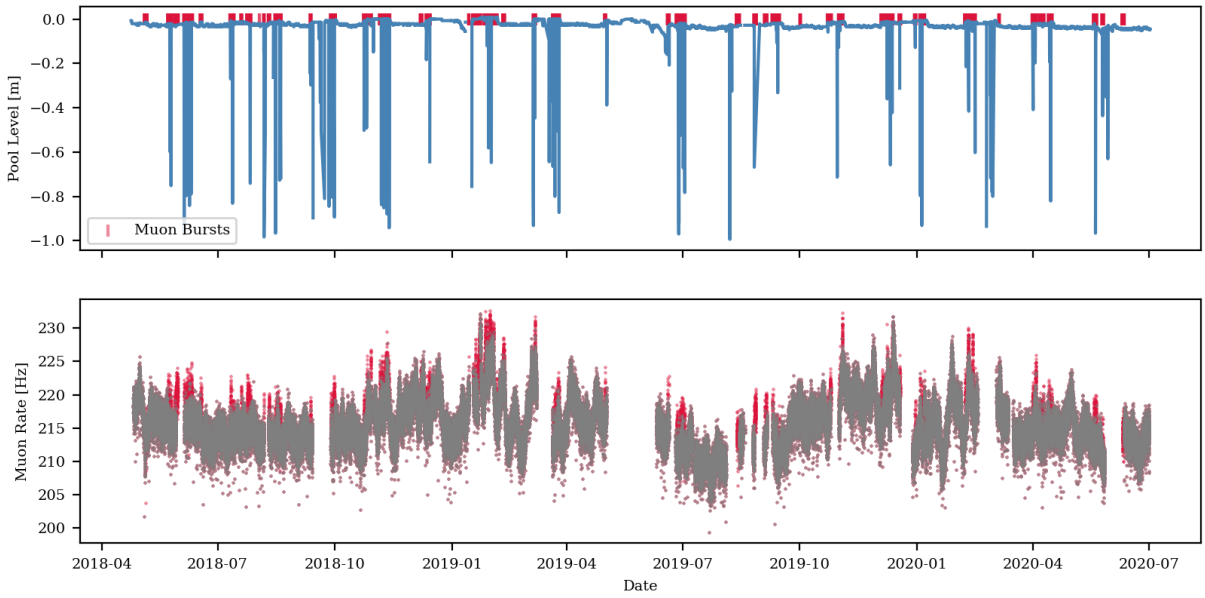


Figure 4.19: Top: relative reactor pool level as a function of time (blue). In red are represented the muon bursts. Bottom: muon rate as a function of time (red) and the muon rate as a function of time without the bursts in grey.

not the data and present the final dataset used in the IBD analysis [103].

4.4.1 Quality selection

The first rough selection that has been made on the data is to ensure the good quality of the external conditions of data taking before any data reconstruction. This selection is based on empirical criteria to reject runs when the detector was not stable. Runs during large variations of the temperature or pressure were removed from the selection. Those variations usually happened during pressure tests in the containment building or when the cooling of the detector was turned off. Calibration runs were also not selected for obvious reasons, and a few runs during some maintenance work were discarded. Apart from those obvious runs, a whole reactor ON cycle has been removed from the selection, between 8$^{\text{th}}$ of July 2019 to the 9$^{\text{th}}$ of September 2019 due to issues affecting a plane in the detector, and because a part of the data being taken without the back wall of the detector, inducing a bigger background rate. If a part of this data sample could be used in the future, it has been decided to leave it aside for the moment to avoid having to treat this specific cycle differently.

Due to a regular partial draining of the reactor pool during the year, the environment around the detector can variate, reducing the passive shielding of the detector. The consequence of this is a quick increase of the muon rate on a small time scale: a muon burst. In order to select the muon bursts to remove them, the muon rates of each eight minute run is compared to the average muon rate on one hour. The runs that show a muon rate greater than 2 Hz compared to the average are then discarded. On figure 4.19 is represented the pool water level variation through time (top) and the muon rate through time (bottom). One can see that the muon burst apparitions correlate with the draining of the pool.

Finally, runs that are not clearly identified as reactor ON or OFF are discarded. The reactor state is defined with the thermal power P_{th} so that reactor ON (OFF) corresponds to $P_{th} > 40\text{MW}$ ($P_{th} < 1\,\text{MW}$).

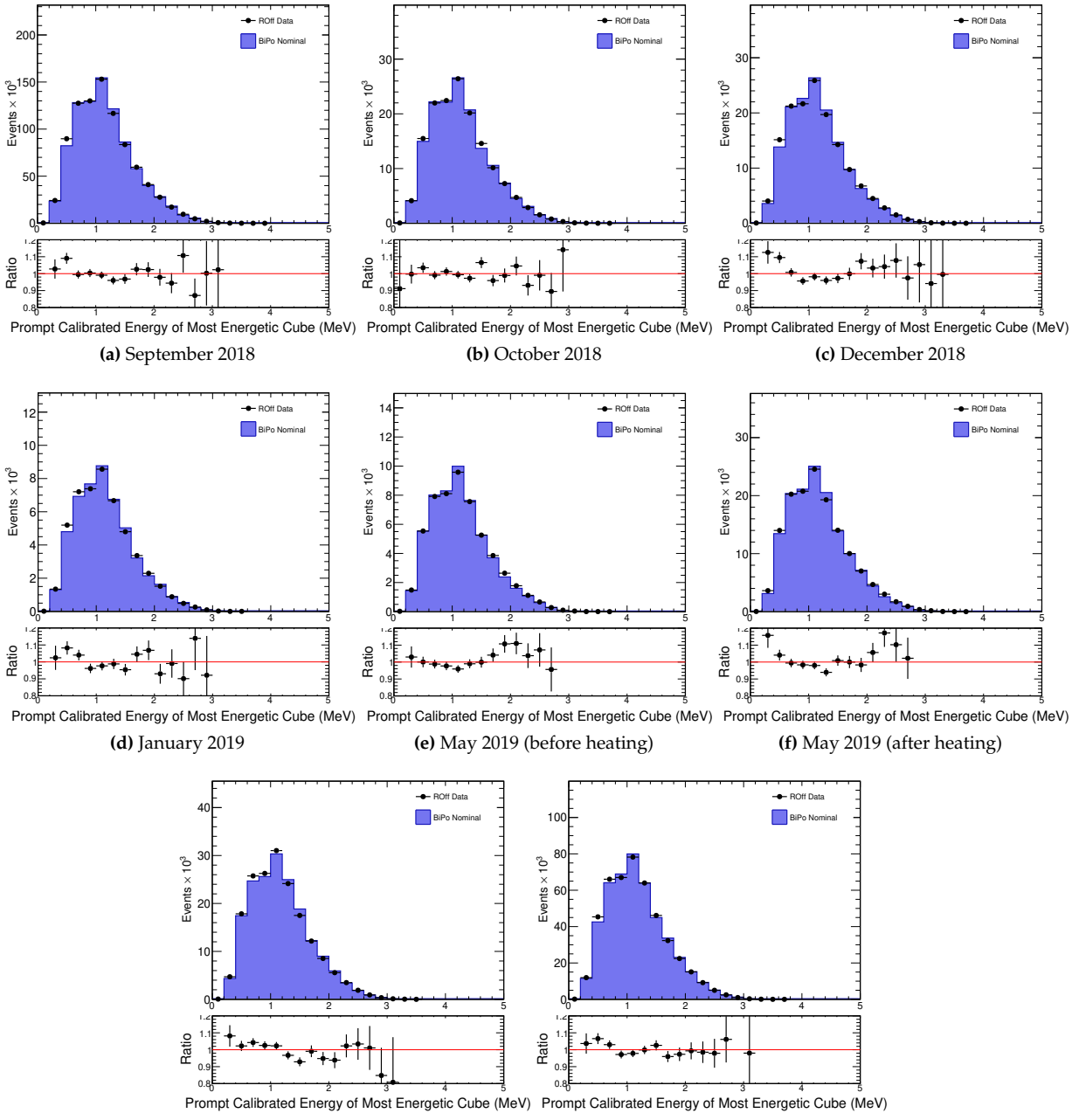
Then, after a selection on first removal of bad runs induced by the environmental conditions, another set of selection cuts are applied on reconstructed variables. Among the reconstructed variables used for the pre-selection, runs with too much deadtime or too low run time are removed from dataset. The trigger rates are supposed to be fairly stable if the detector response is stable, so the periodic trigger, threshold trigger and NS trigger rates are bounded to reference values, as the muon rates or the BiPonisher peak for a BiPo selection. Those cuts are empirical cuts based on the variable distribution on all the Phase I dataset to remove the runs that are too different from the average while keeping the reactor ON/reactor OFF specificities. This set of data quality criteria are described in [103].

4.4.2 Data to Monte-Carlo comparisons on control samples

After a data quality selection, the stability of the detector has been tested using the BiPo background. This background is expected to be constant in time and is thus a good tool to add to the calibration to control the response of the detector. A selection on BiPo background has been performed on non-signal region, avoiding the use of data used later in the oscillation analysis. To do so, the following selection cuts have been applied to the data:

- $\Delta T \in [200, 500]$ µs
- $\Delta X \in [-1, 0]$
- $\Delta Y = 0$
- $\Delta R \leq 1$
- $Prompt\_energy \in [1, 4]$ MeV
- $BiPonator \in [0.22, 0.65]$

From the reactor OFF dataset selected presented in table 4.2 using only the runs validated by the previous data quality work. The BiPo sample can then be used as a data sample to further validate the Monte-Carlo simulations. As presented on the table 3.2, eight different tuning of the simulation by the calibration have been implemented in order to reproduce as precisely as possible the detector effects. The Data to Monte-Carlo comparison for the three different estimators are represented on the figures 4.20,4.21, 4.22. On those figures, one can see that the most stable is the $Crown\_energy$ estimator, with nearly no distortion on the energy spectrum of the BiPo through the time. On figure 4.23 can be seen the sum of all Monte-Carlo and data energy distribution on all the dataset for the three estimators. One can see that the $Prompt\_energy$ variable shows some distortions at 2.5 MeV on the data compared to the Monte-Carlo simulation. This can be explained by the energy scale that is less precise for the $Prompt\_energy$ variable. The two other estimators show a really good data to Monte Carlo agreement above 1 MeV with an agreement below the 5% level on the whole [1-3] MeV region for the $Crown\_energy$ estimator. This confirms the choice of this estimator that is stable over time and is expected a precise reconstruction of the IBD positron energies. This comparison on physics data confirms the good stability of the detector through the phase I dataset and the precise tuning of the energy response of the detector in the simulations.

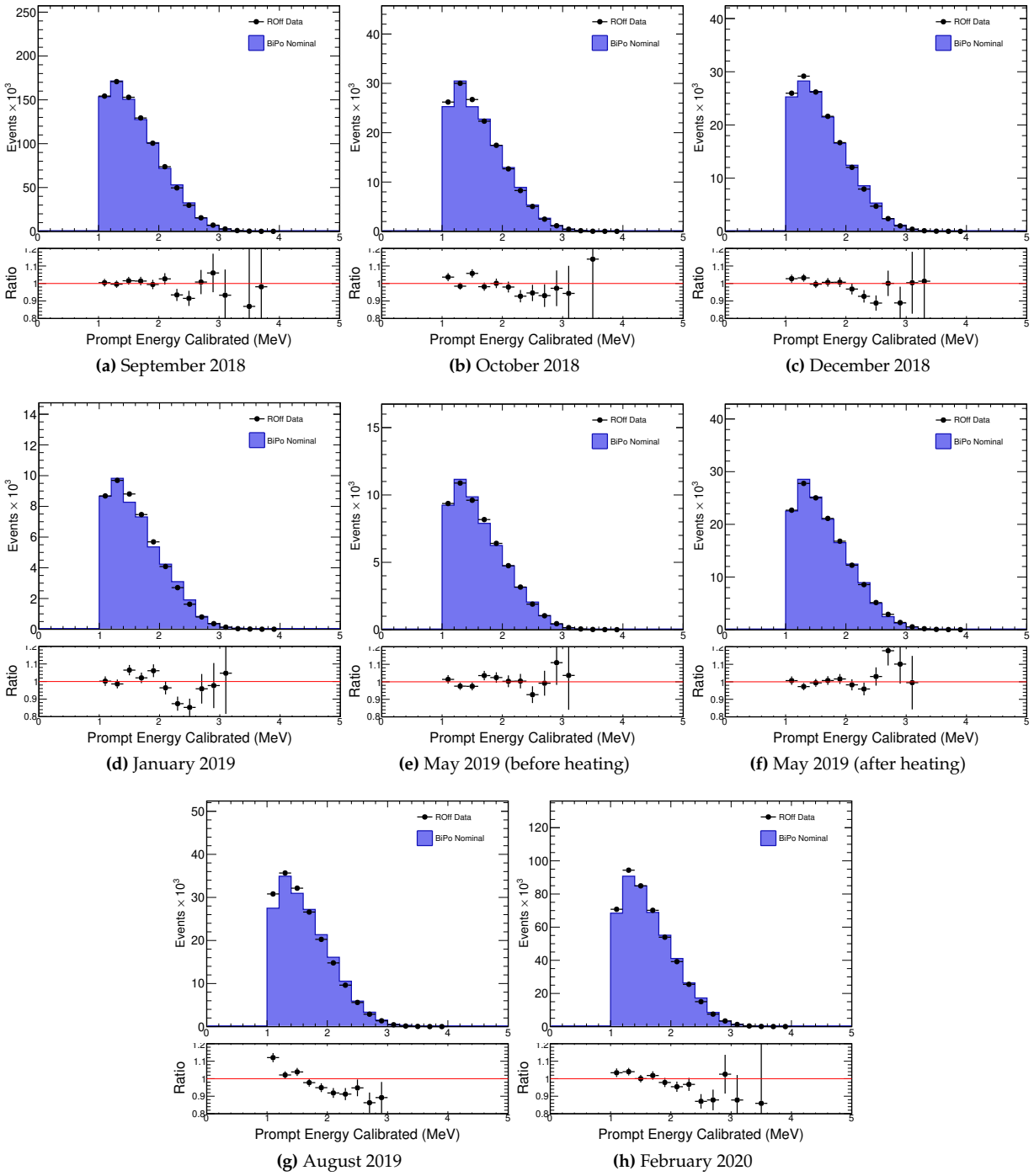


(a) September 2018 **(b)** October 2018 **(c)** December 2018

(d) January 2019 **(e)** May 2019 (before heating) **(f)** May 2019 (after heating)

(g) August 2019 **(h)** February 2020

Figure 4.20: Breakdown of the data to Monte-Carlo comparison for the AC_energy energy estimator on BiPo background. Each plot corresponds to a set of calibration constants applied to the data and the Monte-Carlo tuning associated.

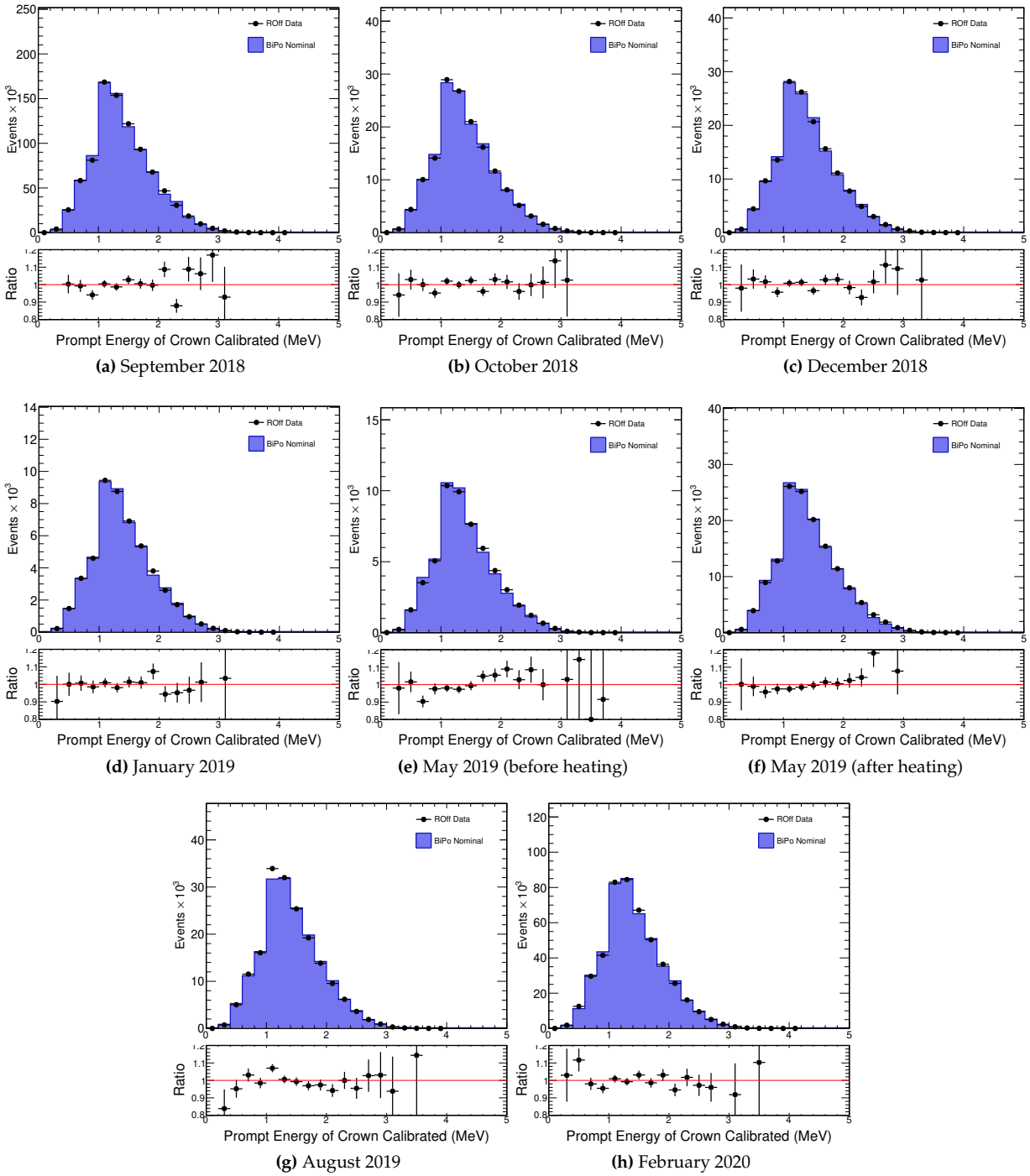


(a) September 2018 **(b)** October 2018 **(c)** December 2018

(d) January 2019 **(e)** May 2019 (before heating) **(f)** May 2019 (after heating)

(g) August 2019 **(h)** February 2020

Figure 4.21: Breakdown of the data to Monte-Carlo comparison for the $Prompt\_energy$ energy estimator on BiPo background. Each plot corresponds to a set of calibration constants applied to the data and the Monte-Carlo tuning associated.

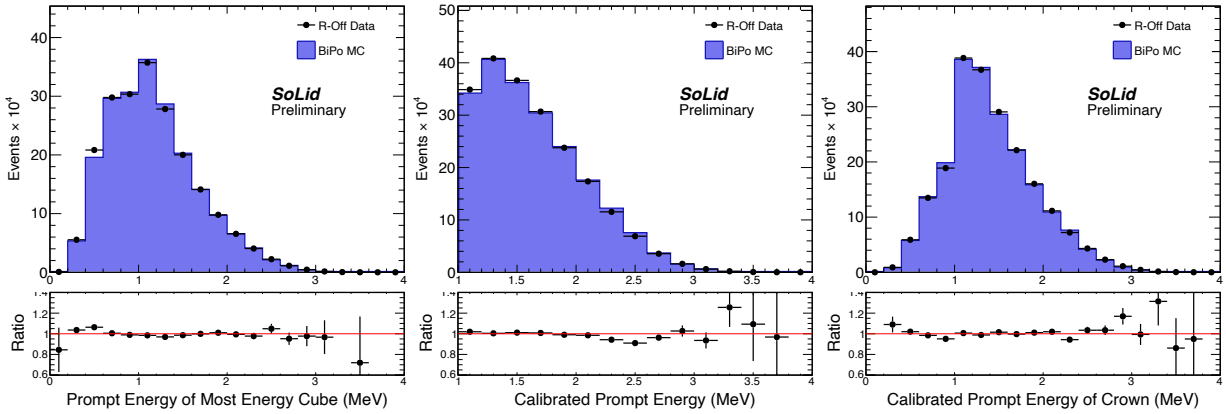


(a) September 2018 **(b)** October 2018 **(c)** December 2018

(d) January 2019 **(e)** May 2019 (before heating) **(f)** May 2019 (after heating)

(g) August 2019 **(h)** February 2020

Figure 4.22: Breakdown of the data to Monte-Carlo comparison for the $Crown\_energy$ energy estimator on BiPo background. Each plot corresponds to a set of calibration constants applied to the data and the Monte-Carlo tuning associated.



(a) AC Energy **(b)** Prompt Energy **(c)** Crown energy

Figure 4.23: Data to Monte-Carlo comparison for the whole phase I reactor OFF dataset. Each energy distribution is the sum of the distributions of each estimator is obtained with the sum of all the distributions presented on the figure 4.20, 4.21, 4.22.

4.5 IBD analysis

In this section, several analyses will be presented, from the most simple to the more complex. This will start with basic cuts using the variables presented in section 4.2. Information on the topologies of the prompt signal will then be added to improve further the discrimination power of the analysis. Finally, multivariate tools used to optimize the background rejection will be presented. To evaluate each analysis, an open dataset of 20 days of reactor OFF data will be used as background sample and IBD Monte-Carlo simulation reconstructed with *ROsim* and *Saffron* will be used as a signal sample. The reactor OFF sample will be taken between July 11$^{\text{th}}$ 2018 and July 31$^{\text{st}}$ 2018. From the two samples, a signal rate (S) and a background rate (B) will be derived from the number of events passing the selection cuts. The expected signal rate can be expressed the following way:

$$S = N_{IBD}^{int} \cdot \frac{N_{IBD}^{sel}}{N_{IBD}^{G4}} \cdot \epsilon_{DT} \tag{4.13}$$

where N_{IBD}^{int} is the expected number of IBD interaction in the detector, ϵ_{DT} is the dead-time, N_{IBD}^{G4} is the number of IBDs generated for the simulation and N_{IBD}^{sel} is the number of IBDs from the GEANT4 sample that are reconstructed and selected after the antineutrino analysis. For the *open dataset* in the reactor cycle 3, 1088 IBD per day are expected to interact in the detector, computed with the method presented in section 2.5.1. ϵ_{DT} is dominated by the time between two runs, of about 33s (about 7% of a 8-minute run) run and the muon veto dead time and has been evaluated at 6% of a run for this given dataset. The neutron capture efficiency has been measured in previous neutron calibration work and its value is sampled from the calibration in the simulation, its value is $\epsilon_n^{IBD} = 52\%$ [84]. The ratio $N_{IBD}^{sel}\ /N_{IBD}^{G4}$ thus encodes the global detection efficiency, including neutron capture and reconstruction efficiencies; as well as all detector effects included in the simulation.

4.5.1 Basic selection cuts

The simplest analysis developed in SoLid is a cut based selection. The values of the selection cuts are the following:

- $Crown\_Energy \in [1.5 - 7]$ MeV
- $\Delta X \in [-3, 3]$
- $\Delta Y \in [-3, 3]$
- $\Delta Z \in [-2, 3]$
- $\Delta R \in]0, 3]$
- $\Delta T \in [5, 141]$ µs
- $BiPonator > 0.7$

The low-energy region is highly contaminated with BiPo background, so the lowest energies are discarded by the energy selection. The $\Delta X, Y, Z, R$ selection are used to remove the accidental background where the NS and the ES are far from each other. The $\Delta R > 0$ selection also removes a large part of the BiPo contamination. The ΔT selection removes the large ΔT BiPo background component. At last, the high BiPonator selection removes the low $BiPonator$ BiPo background contamination. The specific values of the selection cuts were optimized in previous work as in [100] and [84] and if the values could be optimized again with improvement of the reconstruction software and new variables as the BiPonator. As the main analysis of the thesis is the multivariate analysis, the optimization of the variable selection was not performed here and the results give an estimation of the discrimination power in terms of signal and background rates that can be achieved with a simple cut based analysis.

With the basic pre-cut selection, one can achieve a performance described in table 4.4 with the cumulative efficiency of the selection cuts represented in figure 4.24. The cut total efficiency, without taking into account the reconstruction and dead time efficiencies, obtained with the total selection is 42 % for the IBD and 0.6% for the reactor OFF data. With 1088 IBD interactions expected per day in the dataset, by applying the totality of the pre-cut described before, one can achieve a measurement of around 160 antineutrino events for around 5200 background events for a S/B value of 0.03. Those performances are way too low to be able to do a full antineutrino analysis with an oscillation study, where one would like to reach $S/B \sim 1$ for at least a hundred antineutrinos per day. The selection presented here will be used as a preselection in the two other analyses.

The large proportion of remaining background can partially be explained by a BiPo background pollution ten times higher than what we observed with previous detector prototype using the same technology [104]. This larger BiPo contamination could be explained by pollution of the ZnS sheets due to change in the manufacturing process of the sheets. One other cause of the large background rate is the absence of an external muon veto around the detector that would help the rejection of after-muon events, even when the muon does not cross the detector. The last limitation of the detector is the difficulty to do a pulse shape discrimination between electrons and proton recoils that induce the ES compared to liquid scintillator experiments. This level of background contamination in the signal selection is a call for a more thorough use of the detector segmentation through the use of the prompt signal topologies.

Selection	S [$/day$]	B [$/day$]	S/B
$\Delta X \in [-3, 3]$	342	218699	0.002
$\Delta Y \in [-3, 3]$	342	218709	0.002
$\Delta Z \in [-2, 3]$	342	216997	0.002
$\Delta R \in]0, 3]$	304	129269	0.002
$\Delta T \in [5., 141]$ µs	274	104239	0.003
Crown_Energy $\in [1.5 - 7]$ MeV	275	81059	0.003
$BiPonator > 0.7$	259	21922	0.01
Total	149	4892	0.03

Table 4.4: Signal rate, background rate and signal over background ratio for the selection using each selection cut individually.

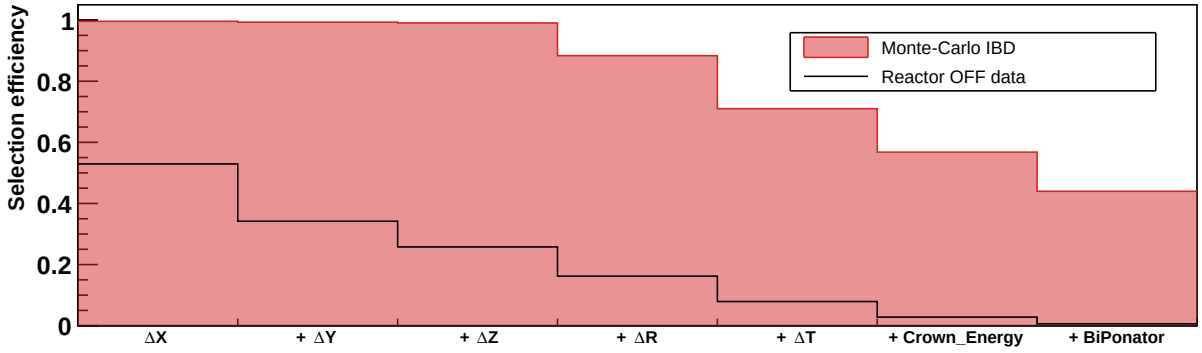


Figure 4.24: Efficiency of the cumulative selection cuts on IBD Monte-Carlo and on reactor OFF data. The reconstruction efficiency is not represented in the plot.

4.5.2 Topology reconstruction

4.5.2.1 Gamma track reconstruction

In order to use the SoLid detector at its full potential, the colleagues of Subatech Nantes developed a tracking algorithm to reconstruct the annihilation gammas. The methods with its main results and performances will be presented in this document and more details can be found in David Henaff's thesis [100].

The main goal of the algorithm is to identify the full topology IBD events composed of one high-energy cube where the antineutrino interacted and two tracks of cubes back to back induced by the annihilation gammas. The gammas mainly interact randomly via Compton scattering in the detector, thus reproducing a gamma track is not as simple as a straight muon one. Furthermore, due to the gamma travel time and the time resolution of about 10 ns for an electromagnetic signal, it is not possible to perform a time tracking of the reconstructed cubes in order to track the annihilation gammas. The last main difficulty with the gamma tracking is the low-energy of the annihilation gammas. Each gamma is emitted at 0.511 MeV and can generate only a few pixel avalanches in the channels, with reconstruction

inefficiencies at that level of energy and a Poissonian energy resolution as discussed in section 3.4.1. To try to tackle those issues, the Compton scattering cross-section will be used in association to the different cube energies and position to build a global likelihood function.

The first information used to build the likelihood function is the mean free path of a gamma with a given energy E inside PVT. In order to compute it, one needs to know the Compton scattering cross-section for a gamma of energy E integrated on the solid angle. This was derived by Klein and Nishina [105]:

$$\sigma_{KN}(E) = 2\pi r_e^2 \left(\frac{1+\alpha'}{\alpha'^2} \left(\frac{2(1+\alpha')}{1+2\alpha'} - \frac{\log(1+2\alpha')}{\alpha'} \right) + \frac{\log(1+2\alpha')}{2\alpha'} - \frac{1+3\alpha'}{(1+2\alpha')^2} \right) \tag{4.14}$$

with α' the ratio $E/m_e c^2$ and r_e the classical radius of the electron ($r_e = 2.8 \cdot 10^{-13}$ cm). The mean free path of a gamma of energy E in the PVT can then be written as :

$$\lambda(E) = \frac{1}{\sigma_{KN}(E) \cdot n_e} \tag{4.15}$$

with n_e the number of electron per cubic centimetre. This can be obtained from the PVT composition given by the manufacturer with for EJ-200 PVT [106]:

$$n_H = 5.17 \cdot 10^{22} \text{ cm}^{-3}$$
$$n_C = 4.69 \cdot 10^{22} \text{ cm}^{-3}$$

and with finally

$$n_e = Z_H \cdot n_H + Z_C \cdot n_C \tag{4.16}$$

with Z_H and Z_C the respective atomic numbers of the Hydrogen and Carbon atoms. The probability for a gamma of energy E to interact after travelling a distance d follows an exponential law and can thus be written as:

$$P_{dist}(d, E) = \frac{1}{\lambda(E)} \exp(-d/\lambda(E)) \tag{4.17}$$

The second information on the reconstructed cubes of a cluster that is used is the cube energy. The same way it was done for the calibration with the analytical p.d.f for the gamma Compton scattering in section 3.4.1, one can derive the probability for a gamma of energy E to transmit an energy T to an electron in a Compton interaction:

$$\frac{d\sigma}{dT}(T, E) = \frac{\pi r_e^2}{m_e c^2 \alpha^2} \left(2 + \left(\frac{T}{E-T} \right)^2 \left(\frac{1}{\alpha^2} + \frac{E-T}{E} - \frac{2}{\alpha} \left(\frac{E-T}{T} \right) \right) \right) \tag{4.18}$$

with α the fine structure constant. In the same way as it was done to compute the calibration p.d.f an energy resolution convolution is applied and the obtained p.d.f is then:

$$P_{energy}(T, E) = \frac{\sum_{i=0}^{i_{E_c}} \frac{d\sigma}{dT}(T_i, E) \frac{1}{\sqrt{2\pi}\sigma_0\sqrt{T_i}} \exp\left(-\frac{(T-T_i)^2}{2\sigma_0^2 T_i} \right)}{\sum_{i=0}^{i_{E_c}} \frac{d\sigma}{dT}(T_i)} \tag{4.19}$$

with E_c the Compton edge energy of a gamma with an energy of E, σ_0 an energy resolution factor evaluated at $0.14 \cdot \sqrt{E}$ that was used as an empirical value when the algorithm was developed.

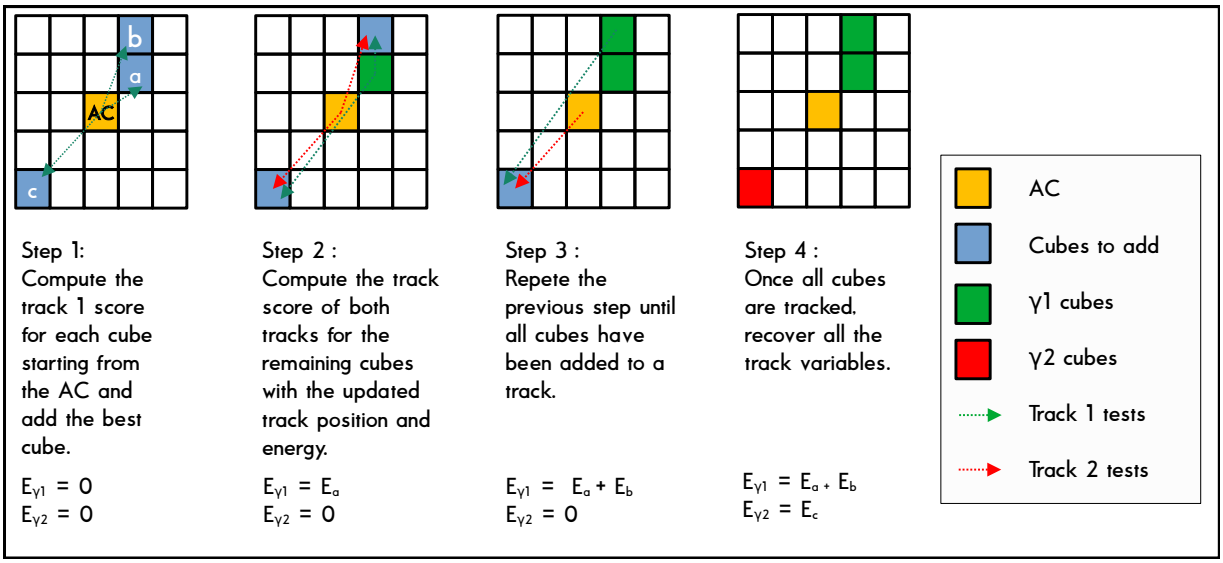


Figure 4.25: Scheme of the gamma tracking algorithm. In the scheme, the three cubes a, b and c are used to create two tracks. The first track is composed of the cubes a and b and the second is composed of the cube c.

With the two probabilities P_{dist} and P_{energy}, one can build a global likelihood score for a gamma of initial energy E to travel a distance d and deposit an energy T:

$$\mathscr{L}_{\gamma}(d, T, E) = -\log(P_{dist}(d, E)) - \log(P_{energy}(T, E)) \tag{4.20}$$

The aim is then to reconstruct at most two tracks t_1 and t_2 with a likelihood minimization algorithm. Both tracks are starting inside the annihilation cube, with an initial energy of 511 keV. The tracks are reconstructed in an iterative way with a loop over all the cubes. For each cube, the likelihoods of being part of the track t_1 or t_2 are computed. Thus, $2 \cdot n_{cubes}$ scores are computed, and the minimal score is selected. The cube with the best score is then added to the corresponding track t_i. In the next iteration, the possible distances travelled by the gamma for the track t_i to compute P_{dist} are computed between the newly added cube and the tested cubes. The likelihoods are then computed with the remaining energy in the track once the 511 keV gamma has deposited its energy in the newly added cube. Once a cube is added to a track, a loop is performed on the remaining cubes with the updated gamma position and energy until no more cube can be added. A scheme of the algorithm can be seen on figure 4.25 with the different steps of the algorithm to build the tracks. For the track creation, the four cubes around the AC ($\Delta X_{AC-cube} = \pm 1$ or $\Delta Y_{AC-cube} = \pm 1$) are not considered, the hypothesis being that they are too often induced by light leakage of the AC in a neighbouring cube or by the positron depositing its energy in two different cubes. Three categories can then be created according to the number of reconstructed tracks:

- Zero gamma is reconstructed. It can happen if the only reconstructed cubes are the AC and the four cubes in the cross around the AC.

- One gamma is reconstructed. In the case where all the cube scores are minimized if they belong to the first track. It can happen if the second gamma escaped the detector as can be seen on figure

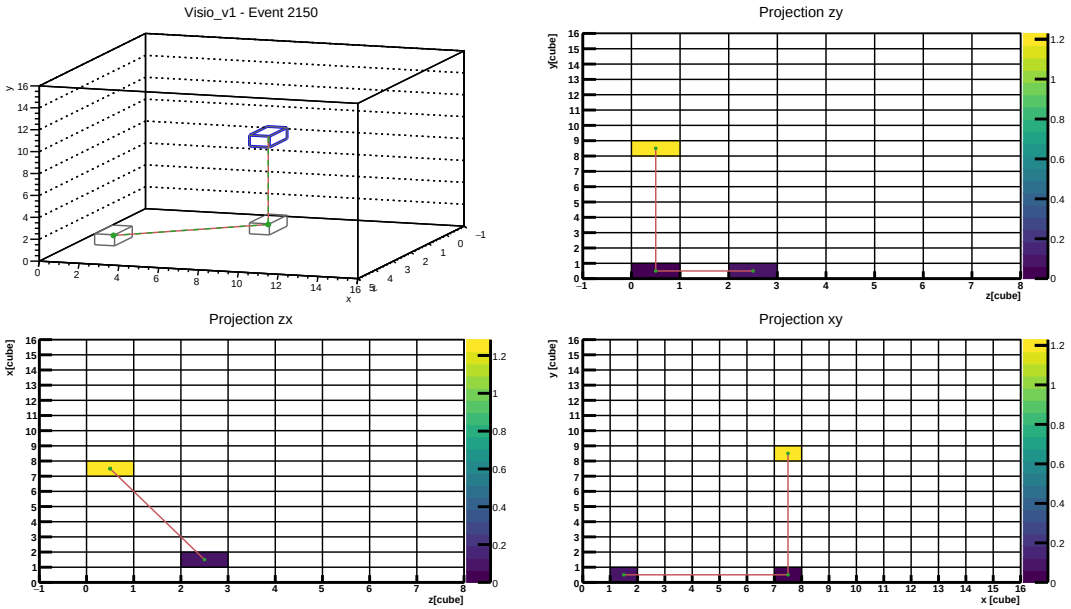


Figure 4.26: Representation of a Monte-Carlo IBD one-gamma event with a custom visualization with a 3D view (top left) and the three projections. On the projections, the yellow rectangle represent the annihilation cube, while the blue rectangles are the other reconstructed cubes of the events. In the event, one track is reconstructed (red line). Each green dot represent a true energy deposit of the gamma inside the detector.

4.26, was not properly reconstructed, as represented on figure 4.27 or if the algorithm failed to reconstruct the second track.

– Two gammas are reconstructed. This is the situation with the most discrimination power because it is specific only to IBDs: one annihilation cube with a large energy deposit and two gamma tracks. An example of a two-gamma category event can be seen on figure 4.28 where both back to back gamma tracks can clearly be seen.

With this new track reconstruction algorithm, one can build new variables:

– Eg_1, Eg_2, the energies of the tracks 1 and 2, defined as the sum of the cube energies composing each track.

– Sc_1, Sc_2, the scores of the tracks 1 and 2, defined as the sum of the likelihood score of the cubes composing each track.

– Nc_1, Nc_2, the number of cubes in each track.

– dot_{12}, the dot product between the two tracks computed the following way:

$$dot_{12} = \frac{(\vec{r}_1 - \vec{r}_{AC}) \cdot (\vec{r}_2 - \vec{r}_{AC})}{|\vec{r}_1 - \vec{r}_{AC}||\vec{r}_2 - \vec{r}_{AC}|}$$

with $\vec{r}_{1,(2)}$ the coordinate of the first cube of the first (second) track and $\vec{r}_{AC}$ the coordinate of the AC in SoLid coordinate system.

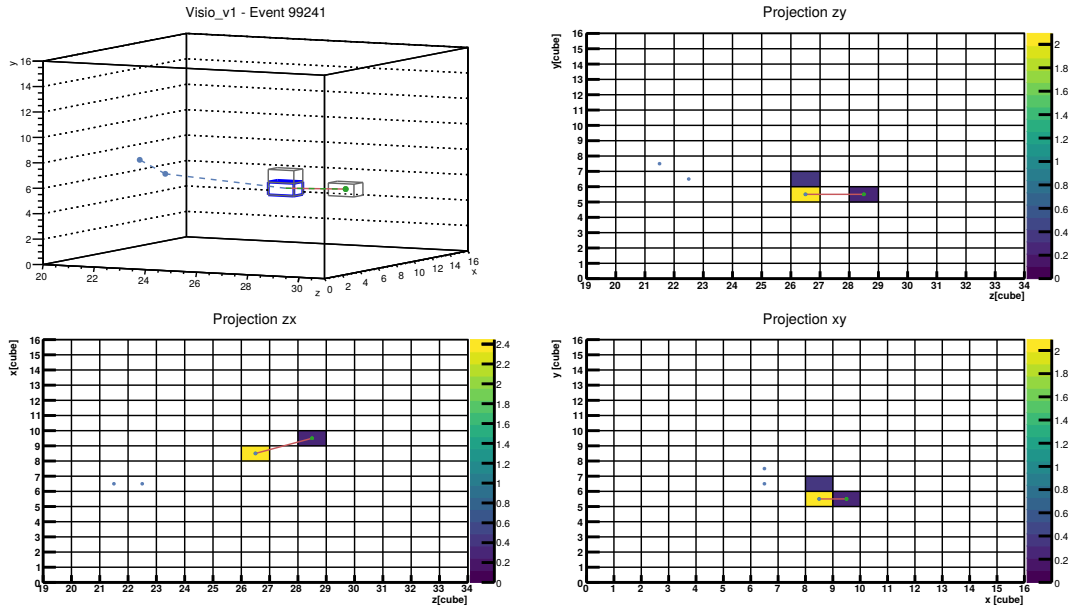


Figure 4.27: Representation of a Monte-Carlo IBD fake one-gamma event with a custom visualization with a 3D view (top left) and the three projections. On the projections, the annihilation cube is represented in yellow while the other reconstructed cubes are represented in blue rectangles. In this event, one track is reconstructed (red line). At the GEANT4 level, two gammas deposited energy in the detector and their true energy deposits are represented by the blue and green dots, but only the deposits of the green gamma were reconstructed.

The distributions of the new variables can be seen on figure 4.29 for the events where one gamma is reconstructed. Here one can see that the most discriminative variable will be the track energy. This discrimination is induced by the difference in the origins of the tracks for IBDs (annihilation gammas) and for the backgrounds (proton recoils for the fast neutrons, gammas that can be emitted in the decay chain of the ^{214}Bi during in the nuclei de-excitation).

On figure 4.30 the distribution of variables for the events with two reconstructed tracks can be seen. As for the one-gamma case, the most discriminative variable is still the gamma energies. The dot product between the tracks is represented on figure 4.31. On the figure, one can see that the IBD will tend to have more event with a negative dot product between the two gammas compared to the background. This difference can be explained by the expected back to back emission of the annihilation gammas. The excess of background events at dot_{12} around one can be explained by proton recoils from fast neutrons that can mostly only travel forward, leading to two gamma tracks in the same direction compared to the cube of maximum energy. The assumption that fast neutrons can only be scattered forward after an elastic scattering on a neutron is derived from the conservation of the energy and momentum in the interaction. As both particles have an almost equal mass, they can only be scattered forward.

4.5.2.2 Tracking performances

A short study has been conducted on the gamma tracking performances. This is performed on IBD Monte-Carlo simulation. For each event, if a gamma leaves at least an energy deposit larger than $50\ \mathrm{keV}$, it is treated as a reconstructible track. The number of reconstructible tracks versus the number of recon-

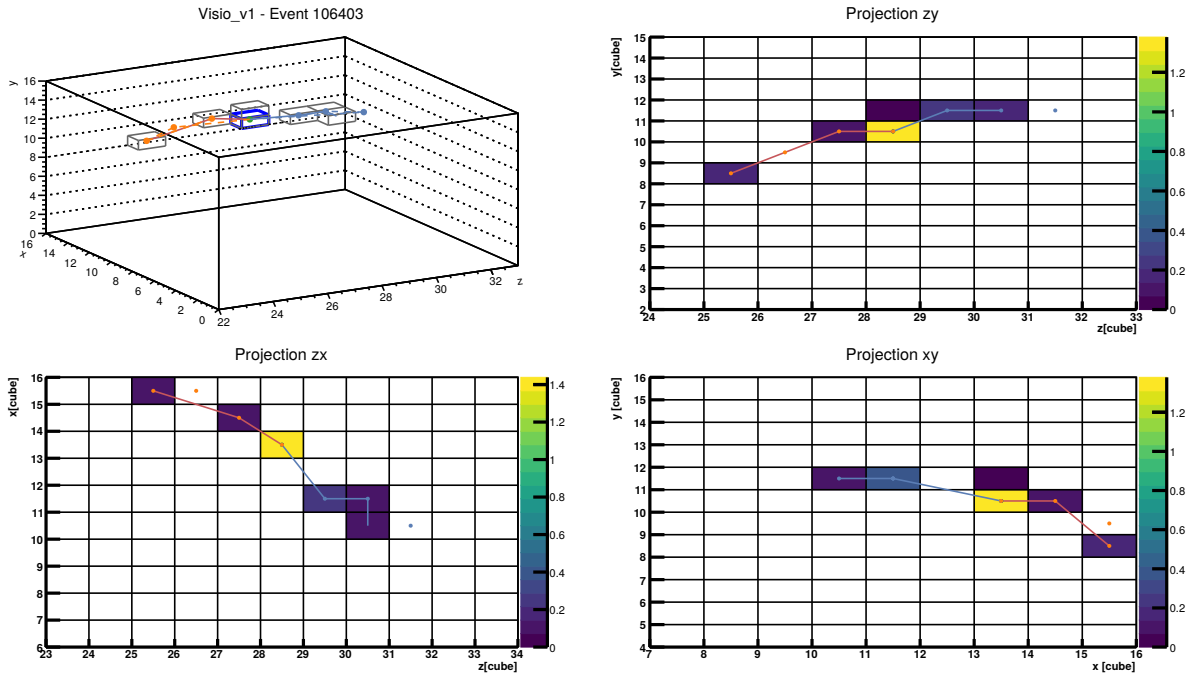


Figure 4.28: Representation of a Monte-Carlo IBD two-gamma event with a custom visualization with a 3D view (top left) and the three projections. The two plain lines (orange and blue) represent the tracks reconstructed with Subatech algorithm, the cubes represent the reconstructed cubes and the dots the true interaction points before reconstruction. The two colours of the dots represent energy deposits from different gammas. On the projections, the annihilation cube is represented in yellow.

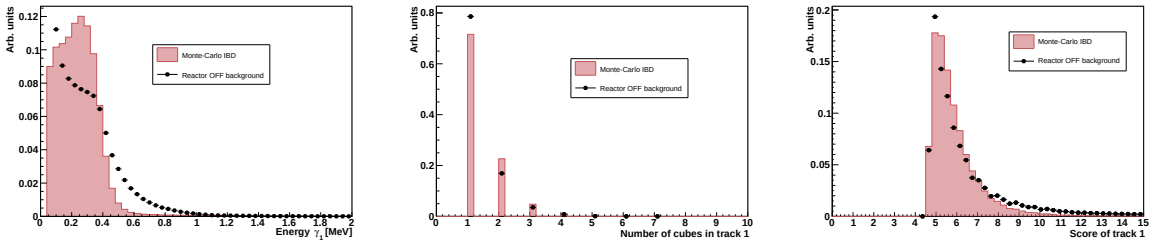


Figure 4.29: Distributions of topology variables for the events in the 1 gamma category for IBD Monte-Carlo (red) and reactor OFF (points). Left: Energy of gamma 1 track. Middle: Number of cubes in the track. Right: total score of the track.

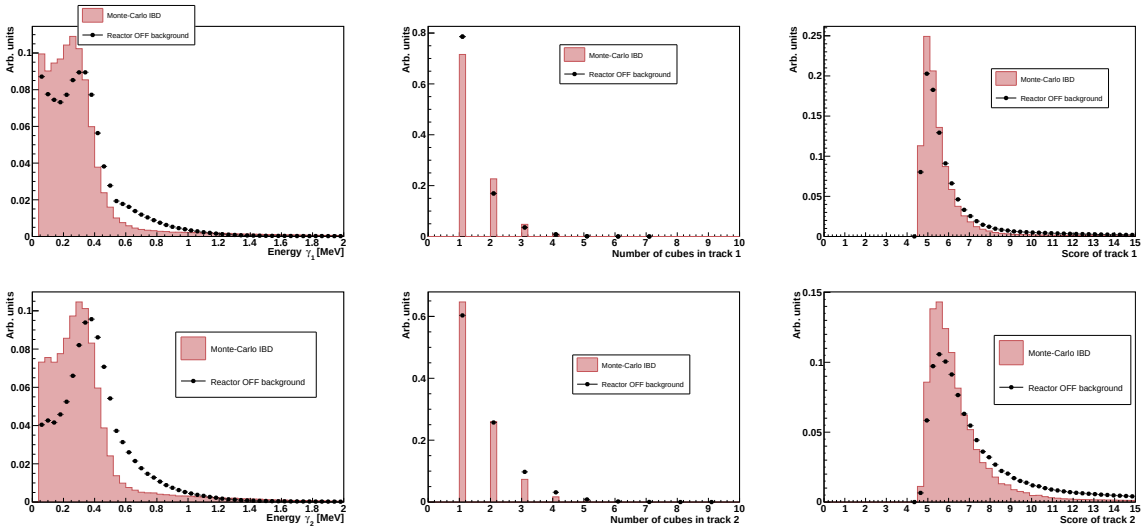


Figure 4.30: Distributions of topology variables for the events in the two-gamma category for IBD Monte-Carlo (red) and reactor OFF (points). Left: Energy of gamma tracks. Center: Number of cubes in the tracks. Right: total score of the tracks. Top: for the first track. Bottom: For the second track.

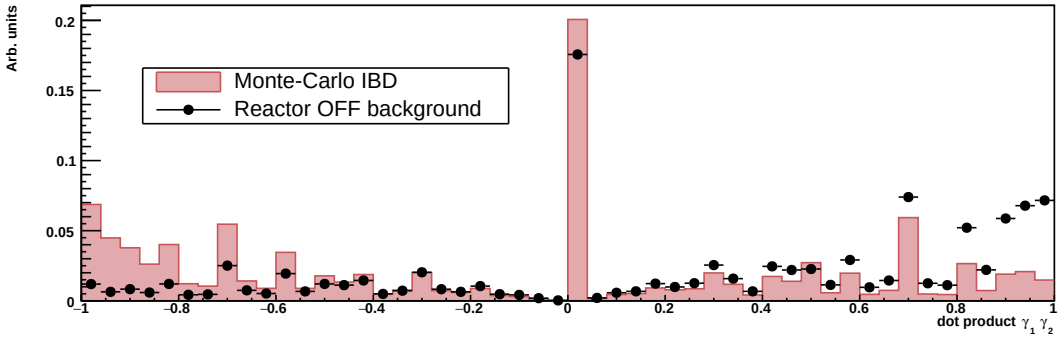


Figure 4.31: Distribution of the dot product between the AC and the first cube of track 1 and track 2 for IBD Monte-Carlo and Reactor OFF data.

structed tracks per event is then compared in the figure 4.32. From all the events where two gammas were reconstructible, 33% of them are reconstructed as one-gamma events. Due to that effect, the ratio $60\%/35\%$ of events in the two-gamma category/one-gamma category at the GEANT4 level drops to $43\%/39\%$ after reconstruction. The number of zero-gamma event can help to understand the difference in the category populations between the reconstructible tracks and the reconstructed ones.

The migration of an event from the one (or two) gamma category at the GEANT4 level into the zero-gamma category after reconstruction can only be explained by the bad reconstruction of the cubes inside the tracks. Indeed, zero-gamma events are only composed of the annihilation cube and the 4 cubes next to it, so if there was another energy deposit in the detector in another cube, it was not properly reconstructed. For the events with one gamma that deposited energy in the detector, 25% of the events were reconstructed as zero-gamma events. This proportion can be treated as the probability for a track not to be reconstructed. When it is applied to the two-gamma events, one can conclude that from the 33% true two-gamma events that are reconstructed into one-gamma events, 15% of the tracks are not reconstructed at all and 18% of the tracks are badly tracked. The performances of the algorithm could be improved with a larger number of cubes being reconstructed with a lower reconstruction threshold or a higher light yield. This is the aim of the detector upgrade started in summer 2020. Ongoing upgrade of the cube reconstruction, algorithm with a better treatment of the light leakages between the cubes and the fibre to fibre inhomogeneities are believed to be also a good help for the gamma tracking with a better localization of the cube in the detector.

4.5.2.3 Neutron track reconstruction

In the same spirit as the gamma tracking presented in the previous section, a neutron tracking algorithm has also been developed by the Subatech team. The tracking is made under the assumption that the neutrons can only travel forward. For an event with a number of cubes n_{cubes} being reconstructed, all possible combinations of the cubes are created to make tracks of n_{cubes} length. From this $n_{cubes}!$ tracks, only the ones that show no backward scattering of neutrons are kept. The neutron tracking algorithm relies on the neutron scattering cross-section on Hydrogen and Carbon:

$$\sigma_n = \sigma_n^H \cdot n_H + \sigma_n^C \cdot n_C \tag{4.21}$$

with $\sigma_n^{H,C}$ the cross-section of neutron scattering on hydrogen (carbon), and $n_{H,C}$ the density of hydrogen (carbon) atoms in the PVT. The cross-sections can be retrieved from ENDF/B-VIII.0 nuclear database [107]. Then, in the same way as equation 4.17 one can build a probability to interact:

$$P_{dist}(d, E) = \frac{1}{\lambda_n(E)} \exp(-d/\lambda_n(E)) \tag{4.22}$$

Because of a quenching effect in the neutron induced scintillation, one cannot directly reconstruct the true energy of the neutron using the energy calibration described in the chapter 3. The energy of the neutron can be described using Birks' law in the following way:

$$E_{true} = E_{vis} \cdot (1 + k_B \cdot \frac{E_{true}}{L_{true}}) \tag{4.23}$$

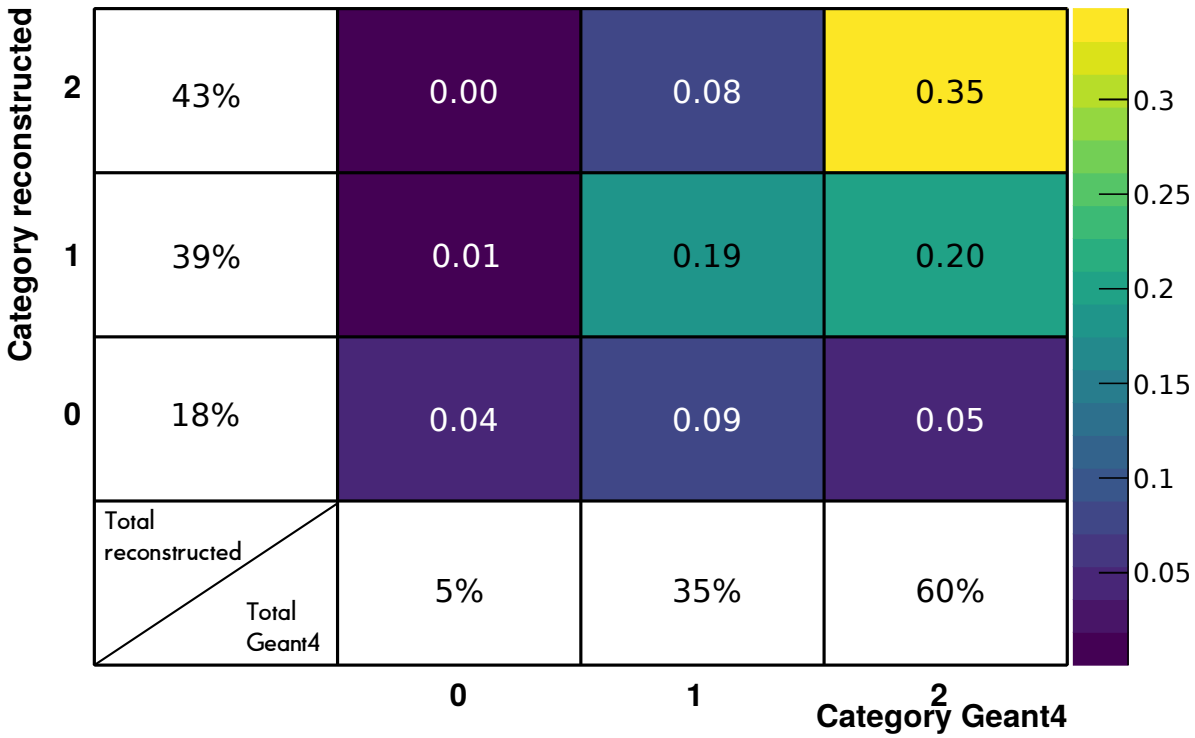


Figure 4.32: Number of reconstructed gammas vs number of real gammas that deposit energy in the detector from GEANT4 before any event selection. The histogram is normalized to 1. The first row in the bottom represents the repartition of events with 0, 1 or 2 gammas in GEANT4 simulation. The first column represents the repartition of the events with 0, 1 or 2 gammas after reconstruction of the events.

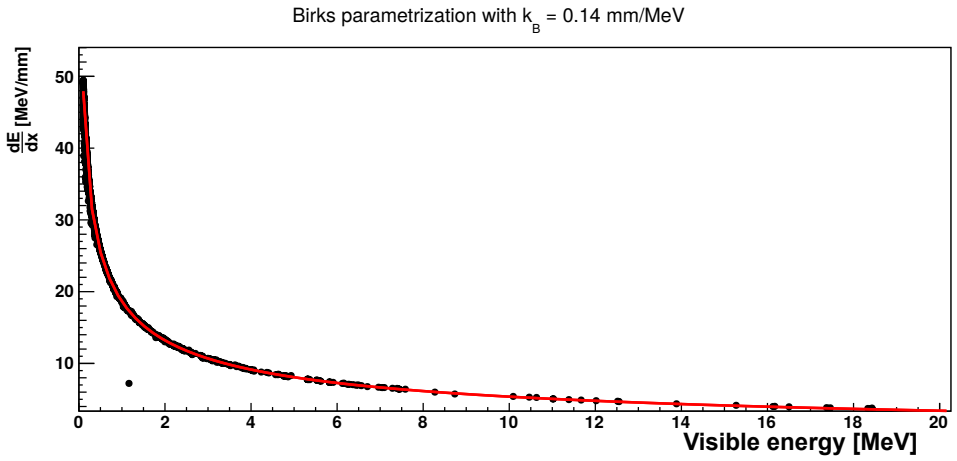


Figure 4.33: $\frac{E_{true}}{L_{true}}$ as a function of E_{vis} fitted with the parametrization from the equation 4.24. E_{true} and L_{true} are obtained from GEANT4 simulation and E_{vis} computed with the equation 4.25 with $k_B = 0.14$ mm/MeV.

Where E_{true} is the true energy of the neutron, L_{true} is the distance of the recoil of the proton induced by the neutron scattering, E_{vis} is the measured energy of the proton recoil and k_B is Birks' constant. For the PVT used in the SoLid detector, k_B was measured to 0.14 mm/MeV. The E_{true}/L_{true} dependence as a function of the energy attenuated by the quenching effect can be parametrized from GEANT4 simulation with neutron simulation using the parametrization represented on figure 4.33:

$$f_{kB}(E_{vis}) = \frac{p_0}{E_{vis}^3} + \frac{p_1}{E_{vis}^2} + \frac{p_2}{E_{vis}^1} + p_3 + \frac{p_4}{\sqrt{E_{vis}}} \tag{4.24}$$

The energy deposited by the neutron in a cube is thus :

$$E_{true} = E_{vis} \cdot (1 + k_B \cdot f_{kB}(E_{vis})) \tag{4.25}$$

And for each track, a global neutron score is computed:

$$Sc_n = \frac{1}{n_{cubes} - 1} \sum_{i_{cube}=2}^{ncubes} -\log\left(P_{dist}^{i_{cube}}(E_{vis}, dist)\right) \tag{4.26}$$

From the previously created variables, several variables can be computed depending on the different gamma categories.

Zero gamma. As no other cubes than the annihilation cube and the 4 cubes around it are available for the zero gamma category, no neutrons are tracked on those events.

One gamma. For the one-gamma category, the following variables are computed:

- Sc_n the neutron score.
- $Sc_{1,n}$, the ratio between the gamma score and the neutron score for the one-gamma category.

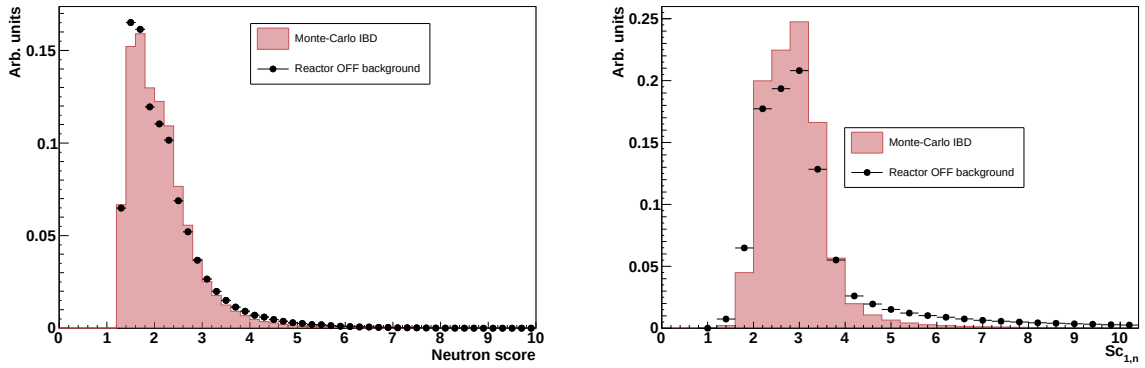


Figure 4.34: Left: distributions of the neutron score for the one-gamma category for reactor OFF data and IBD Monte-Carlo. Right: distributions of the $Sc_{1,n}$ variable.

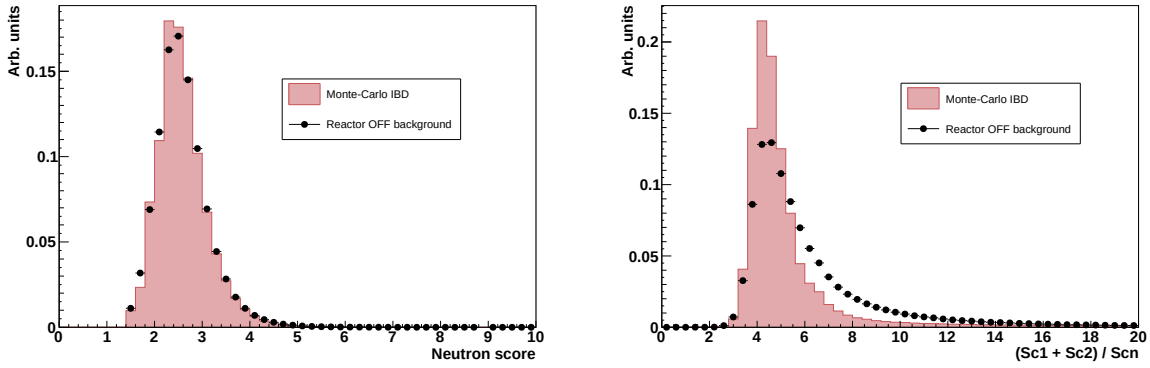


Figure 4.35: Left: distributions of the neutron score for the two-gamma category for reactor OFF data and IBD Monte-Carlo. Right: distributions of the $Sc_{12,n}$ variable.

A distribution of those variables can be seen on figure 4.34. If the neutron score only discrimination power seems quite low, the combination between both neutron and gamma scores with their ratio tends to show a discrimination power in the $Sc_{1,n} \in [2, 4]$ region.

Two gammas. For the two-gamma category, the following variables are computed:

- Sc_n the neutron score.
- $Sc_{12,n}$, the ratio between the sum of the two gamma scores and the neutron score for the two-gamma category.

A distribution of those variables can be seen on figure 4.35. As for the one-gamma category, if the neutron score only does not seem to be a good discriminator by itself, its combination with the other scores gives a discrimination handle between the IBD Monte-Carlo and the reactor OFF data, with a tail in the large scores for the reactor OFF data that is not present in the IBD Monte-Carlo data. This tail is explained by the fact that fake gamma tracks will tend to have a high gamma score and a low neutron score.

4.5.2.4 Selection on topology variables

In this section, the selection used is a simple selection that gives an improvement expected on the antineutrino extraction with a set of rectangular selection cuts on the gamma topology variables. Those cuts are mostly empirical and could be optimized with a proper study that was not performed here. From these newly created event topologies, new handles are available to discriminate the background. The most discriminative variables can thus be used in a simple cut based selection:

- $Eg_1 \in [0, 0.4]$ MeV
- $Eg_2 \in [0, 0.4]$ MeV
- $Sc_{1,n} \in [2, 4]$ (one-gamma category only)
- $dot_{12} < 0.7$ (two-gamma category only)
- $Sc_{12,n} < 5$ (two-gamma category only)

With this selection added to the pre-cuts, one can derive the new selection performances for the one-gamma and two-gamma categories. Those performances can be seen in the table 4.5. The repartition in the different categories after selection is represented on the figure 4.36. The different category populations and their selection power can be described as following:

Zero-gamma category. On the figure 4.36, the zero-gamma category contains less than 10% of the IBDs while containing 35% of the background rate. This is explained by the fact that a large part of the background can be reconstructed in only one cube, with muon clipping in the corner of the detector or BiPo where only the electron from the ^{214}Bi decay is reconstructed in a cube. If that category by itself lacks of discrimination power with no new topology variable available, by removing it from the selection, one can already remove a large part of the background.

One-gamma category. The one-gamma category variable selection efficiency is represented on the figure 4.37 (left). The final selection efficiency (after reconstruction) for this category is 33% for the IBDs and 20% for the reactor OFF data, as presented on the figure 4.36. As it has been presented in section 4.5.2.2, a good part of the IBDs are reconstructed as one-gamma events due to some reconstruction issues. The low population of one-gamma events compared to the two-gamma events in the reactor OFF data can be explained by the philosophy of the tracking algorithm by itself. It is designed to track 511 keV annihilation gammas. With reactor OFF data those type of gammas are not expected in the background sources, thus when multiple cubes are reconstructed in the detector, because the algorithm is forced to gather all the cubes into one or two tracks, it will tend to create the maximum number of tracks possible to avoid accumulating bad scores. It is thus more probable to have zero gamma (only one cube is reconstructed) or two gammas (maximum number of tracks than can be reconstructed) than exactly one gamma tracked. After reconstruction and selection, 49 IBDs per day are left in the one-gamma category for 986 background events per day, for a signal over background of 0.05.

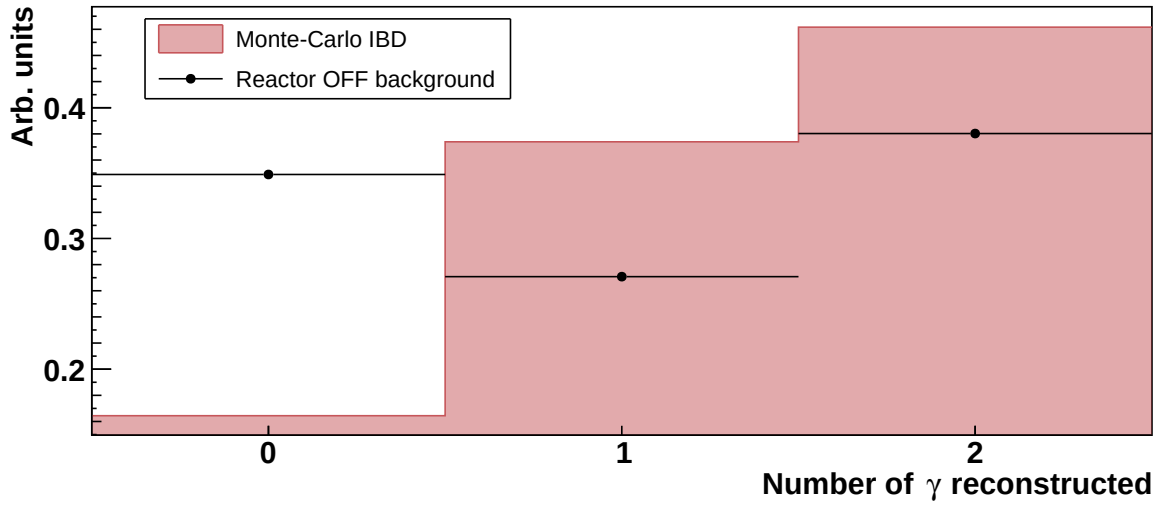


Figure 4.36: Repartition in category for the IBD Monte-Carlo after cut selection.

Two-gamma category. The two-gamma selection efficiency is represented on the figure 4.37 (right). As it is the main IBD category with two tracks well reconstructed, it is the one that gives the best discrimination. With about 5% of the reconstructed reactor OFF data left in the sample after selection, and 25% of the reconstructed IBDs. After reconstruction and selection, 42 IBDs per day are left in the two-gamma category for 317 background events per day, for a signal over background of 0.13.

With the use of gamma tracking reconstruction added with the pre-cuts, the background level was greatly reduced with a total signal over background of 0.07 for 91 antineutrinos per day compared to a signal over background of 0.03 before the use of the topologies. However, the background rejection power of the analysis is still too low and a simple selection with rectangular selection on discriminative variable is thus not enough to obtain a good enough discrimination for an analysis due to the amount of background faced. In order to use the correlation between the variables and improve the analysis discrimination power, a Multivariate analysis was used and will be presented in the next section.

Category	S [$/day$]	B [$/day$]	S/B
1 gamma	49	986	0.05
2 gammas	42	317	0.13
total (1+2)	91	1303	0.07

Table 4.5: Signal rates, background rates and signal over background ratio predictions for one-gamma and two-gamma categories cut selection. The cuts used are the combination of the basic pre-cuts and the selection cuts on the topology variables.

4.5.3 Multivariate analysis

4.5.3.1 Boosted decision trees

Multivariate analyses are widely used in classification problems in high-energy physics. They allow the optimization of a selection with a set of variables and the use of the correlations between them. In the

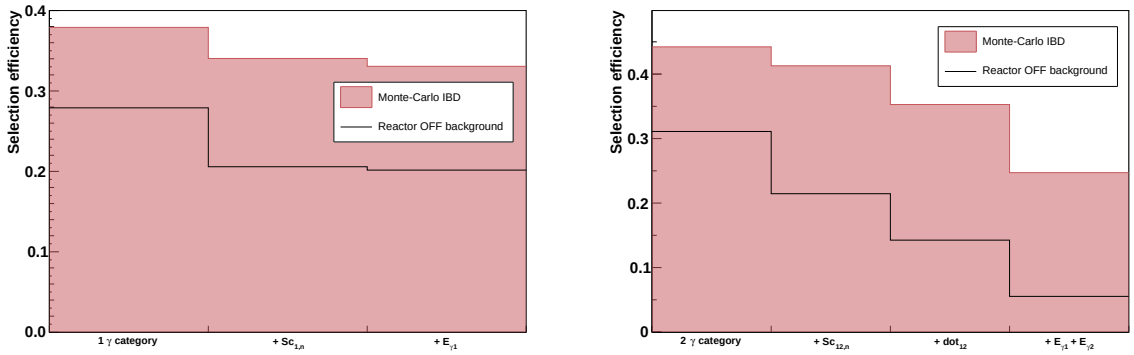


Figure 4.37: Left: efficiency of the gamma topology cuts for the one-gamma category. Right: efficiency of the topology cuts for the two-gamma category. In both cases the efficiencies are computed for a Monte-Carlo simulation of neutrino events interacting through IBD and reactor-off data, representing the background. Each bin represents the efficiency of the selection cut associated to the written variable in addition with the of the bins on the left.

analysis presented in this thesis, a boosted decision tree was used as a robust multivariate tool [108].

A decision tree algorithm is an algorithm based on a set of "weak learners", here the selection variables. The algorithm takes as an input a pure signal sample and a pure background sample, and uses the learners to discriminate as much as possible the two types of data. Each learner will divide the sample in two groups, both with the best purity possible: either mostly background, or mostly signal. In order to increase the discrimination power of a decision tree, a boosting algorithm can then be applied.

Because one tree by itself with only a few number of cuts does not give a high signal-background separation, the principle of the boosted decision trees (BDT) is to compute a larger number of trees and for each tree, build a loss function that quantify the goodness of the algorithm classification. Several loss functions can be used, but the general principle is that the better is the prediction of the BDT compared to the known categories of the event, the lower is the function value. One type of boosting algorithm is based on the gradient descent algorithm in order to minimize the loss function for each new tree created [109]. This algorithm was chosen for the analysis presented in this thesis, as a robust method, already tested in precedent experiments as MiniBooNE [108]. The use of the gradient boosted decision trees (BDTG) is done via the TMVA ROOT package [110] where all the algorithms to train and apply the BDTG are implemented. Once the algorithm is trained, it will classify an event, reading its input variables, from -1 (more likely to be a background event) to 1 (more likely to be a signal event).

4.5.3.2 Algorithm training

The algorithm is trained on two datasets, a background dataset composed of 20 days of reactor OFF data in the *open dataset* and a signal dataset made of IBD Monte-Carlo with $\sim 3.5 \times 10^6$ antineutrinos generated. Half of the dataset is used to train the algorithm and half is used to test the algorithm. This split is done in order to verify that the response of the BDT is the same on the training sample and the test sample. Before the training of the algorithm, a set of pre-cuts presented in section 4.5.1 is applied to the dataset in order to optimize the training of the BDTG. On the pre-cuts, the $BiPonator > 0.7$ and $\Delta T < 141$ µs cuts are not applied to be able to use the BiPo sideband presented in section 4.3.2. Three

Category	type	Number of events
0 gamma	IBD	57191
	Background	29267
1 gamma	IBD	130506
	Background	26410
2 gammas	IBD	161774
	Background	39227

Table 4.6: Number of events in each training sample.

BDTs are trained for the three gamma categories with three different sets of input variables. The number of events used in the training sample is resumed in table 4.6 for each category. The variables used in the algorithm training are the following for the different categories:

0 gamma:		1 gamma:		2 gammas:
X_{NS}		E_{g1}		E_{g2}
Y_{NS}		N_{c1}		N_{c2}
$\Delta X, Y, Z, R$	+	S_{c1}	+	S_{c2}
$promptNCCube$		S_{cn}		dot_{12}
$promptVolume$		$Sc_{1(2),n}$		

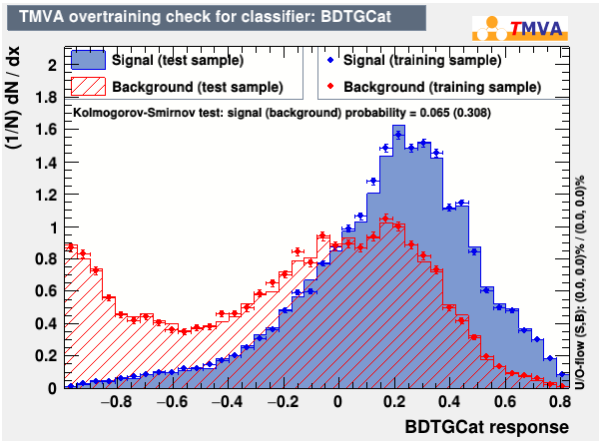
With X_{NS} and Y_{NS} the positions in the planes of the neutron signal. On the figure 4.38 is represented the superposition of the BDTG score for the training sample and the testing sample. A statistical comparison is performed between the two distributions for both backgrounds and signals. The statistical test between the samples shows little to no sign of over-training of the model that could lead to non-optimal performance of the BDTG classification.

The discrimination power of each variable is represented on the figure 4.39. One effect of the BDT can be seen on the less discriminative variable of both category one and two: $Sc_{1,n}$ and Scn are both variables that can be rebuilt from a simple combination of other variables and thus are less used in the end. It is hard to derive precise information from the variable rankings as a small change on the training dataset, in the variables used, the pre-cut, the size of the training sample, can give a totally different ranking for equivalent performances.

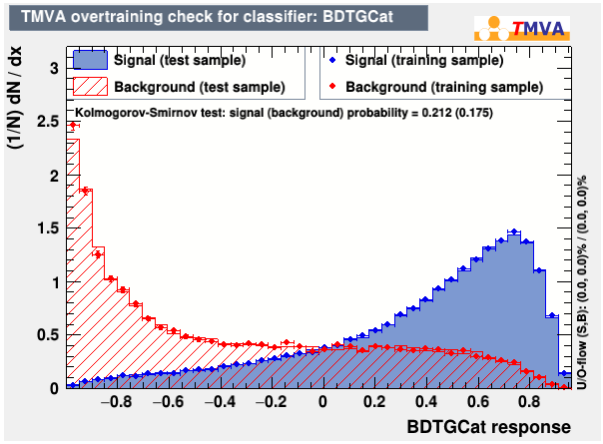
4.5.4 Performance prediction

From the trained algorithm presented in the precedent section, one can use the test sample to make a prediction of the performances on reactor ON data. To do so, one can compute S and B varying the cut on the BDT value. One way to treat the optimization of the BDT scores is to optimize the three BDTs at the same time, with three loops on the scores that gives three signal components S_0, S_1, S_2, and three background components B_0, B_1, B_2. The choice of the selection cut on the BDTG response variables can then be done on a figure of merit with the sum of all background components and signal components.

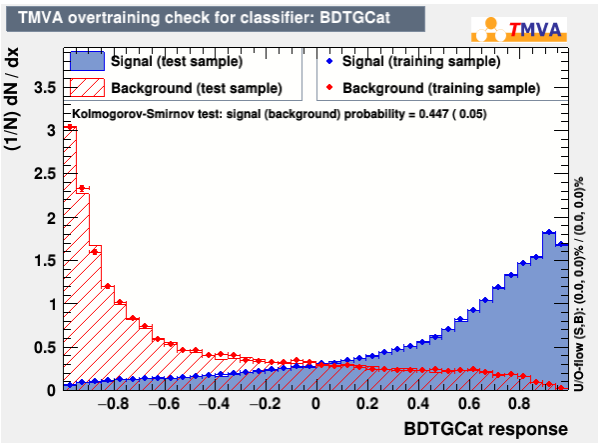
To optimize the selection cut, the usual "statistical significance" $S/\sqrt{S+B}$ figure of merits is used to optimize the BDT cut. However, the analysis is not tied to one figure of merit, and the optimal choice of



(a) 0 gamma



(b) 1 gamma



(c) 2 gammas

Figure 4.38: BDT response function for the signal (blue) and background (red) for each gamma category. The points represent the distribution on the training sample and the filled area represent the test samples.

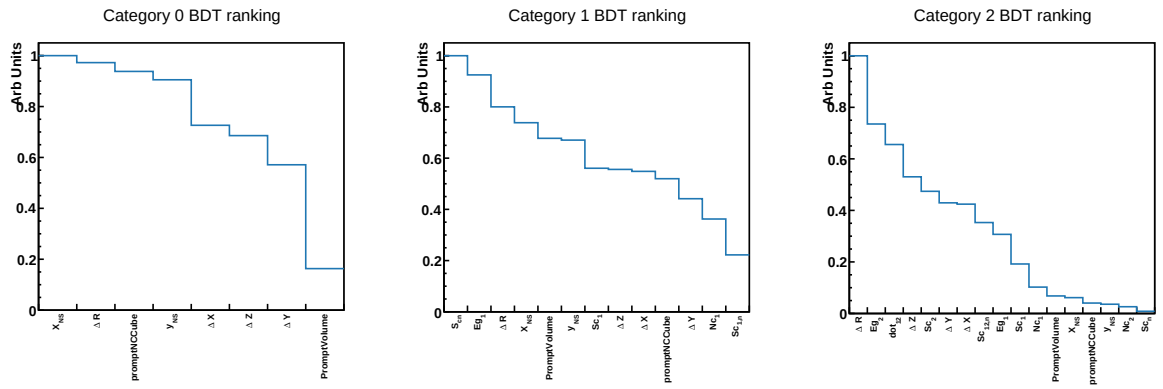


Figure 4.39: Discrimination power ranking for the input variables of the BDT for the three categories of topology.

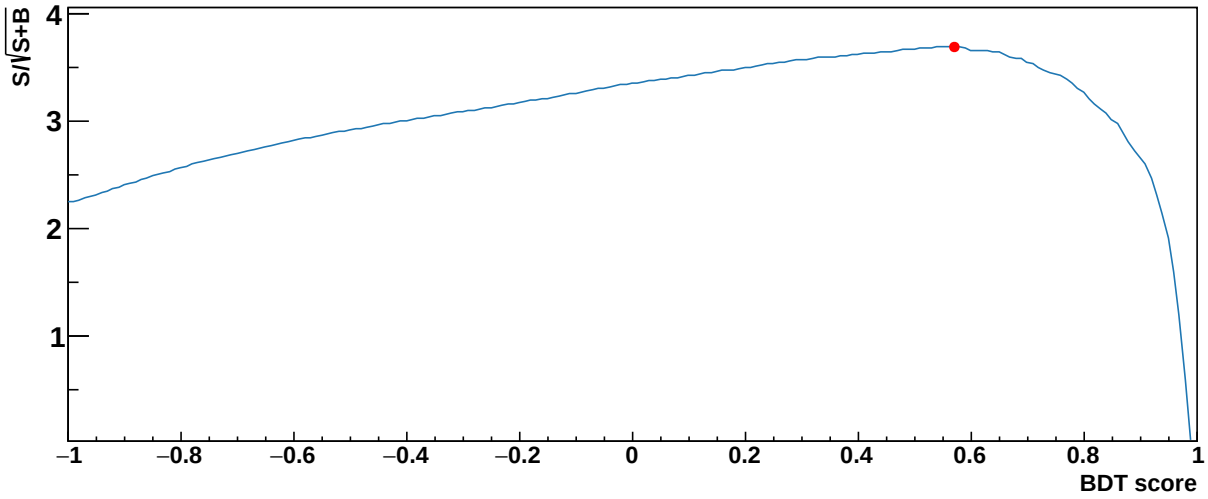


Figure 4.40: Evolution of the predicted significance with the BDT score selection for the one-gamma and two-gamma categories combined. The best significance is obtained for a BDT value of 0.6 and is represented by the red point. On the curve, the score of both categories is set to the same value for each point.

figure of merits (fom) will be derived thanks to the oscillation final analysis and the study of systematic uncertainties. For the $S/\sqrt{S+B}$, the signal and background rates for the optimized BDT selection are represented in table 4.7. The zero-gamma category is totally rejected by the optimization due to its high background contamination for a low number of IBDs and the lack of gamma variables to reject the background. The optimal point is obtained for the two other BDT scores equal to 0.6. This value will be used as the optimal one in the analysis. The variation of the $S/\sqrt{S+B}$ as a function of both BDT scores can be seen on the figure 4.40, on the curve, the score of both categories is set to the same value. The total signal obtained after the sum of the two remaining categories is 65 IBD events expected per day, for a background rate of 242 events per day. It represents a diminution of around 80% of the background rate compared to the results obtained with rectangular cuts in the table 4.5 for a reduction of 30% of signal. The final efficiency in the IBD extraction is 6%. The final S/B predicted with the multivariate analysis is 0.27. For this antineutrino selection analysis, with this level of background, this is the best performance one can get. However, some upgrade of the reconstruction software, are ongoing and might help to slightly improve the IBD selection performance. Those upgrade will be discussed in the section 5.2. The S/B and efficiency variation can be seen on the figure 4.41 Those results will be used in order to compare the predictions and the antineutrino yield that will be extracted from reactor ON data. On those curves, as on the figure 4.40, the same score is used for both categories.

category	S [$/day$]	B [$/day$]	S/B
0 gamma	0	0	-
1 gamma	24	104	0.23
2 gammas	41	138	0.29
total	65	243	0.27

Table 4.7: Signal rates, background rates and signal over background ratio predictions for one-gamma and two-gamma categories after selection on the BDT score optimized with $S/\sqrt{S+B}$.

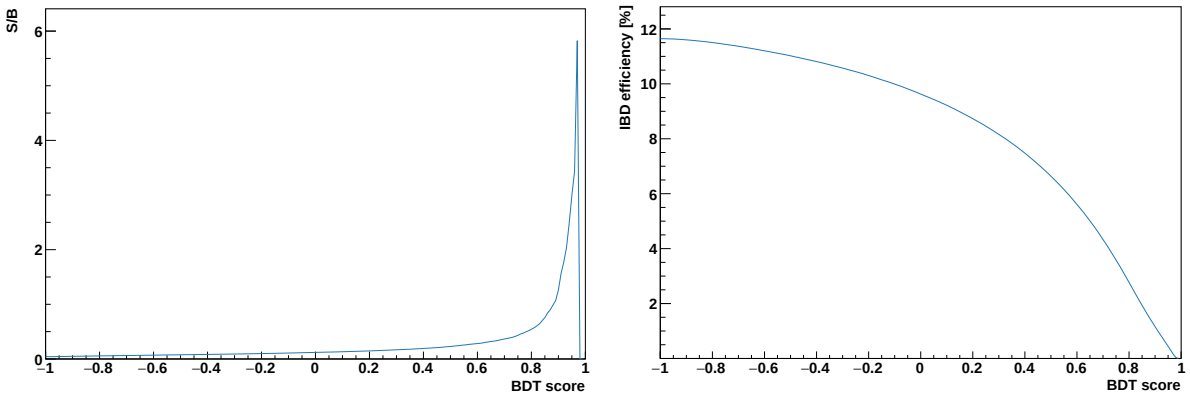


Figure 4.41: Left: signal over background ratio variation with the BDT score. Right: IBD efficiency variation with the BDT score.

4.6 Reactor data subtraction

4.6.1 Subtraction method

Once a selection is fixed for the antineutrino analysis, the final aim is to extract the antineutrinos from the reactor ON data. The main strategy is to remove the background step by step in the data. The first step is to remove the accidentals with the FPNTs sample (described in section 4.3.1):

$$\begin{aligned} S_{TOT}^{ROFF} - S_{FPNTs}^{ROFF} &= S_{BiPo}^{ROFF} + S_{ATM}^{ROFF} \\ S_{TOT}^{RON} - S_{FPNTs}^{RON} &= S_{BiPo}^{RON} + S_{ATM}^{RON} + S_{IBD}^{RON} \end{aligned} \tag{4.27}$$

with S_{FPNTs}, S_{BiPo}, S_{ATM} the FPNTs, BiPo and atmospheric background rates in the reactor OFF or ON samples. The BiPo rate is then evaluated and subtracted using the BiPo sideband defined in table 4.3 rescaled to match the analysis selection with the scaling factor presented in equation 4.3:

$$S_{BiPo} = S_{BiPo \to neutron}(S_{Bipo+Acc}^{sb} - S_{FPNTs}^{sb}) \tag{4.28}$$

where $S_{BiPo \to neutron}$ is the scaling factor for the BiPo sideband to the data selection window, $S_{Bipo+Acc}^{sb}$ is the total number of events in the BiPo sideband and S_{FPNTs}^{sb} is the accidental background yield in the BiPo sideband. $S_{BiPo \to neutron}$ is computed once in the open dataset for all the data. After subtraction of the BiPo components, the obtained rate is:

$$\begin{aligned} S_{ATM}^{ROFF} &= S_{TOT}^{ROFF} - S_{FPNTs}^{ROFF} - S_{BiPo}^{ROFF} \\ S_{ATM}^{RON} + S_{IBD}^{RON} &= S_{TOT}^{RON} - S_{FPNTs}^{RON} - S_{BiPo}^{RON} \end{aligned} \tag{4.29}$$

In the reactor OFF background, the only component left is the atmospheric background component while on reactor ON, there are both atmospheric background and IBDs in the data. As those data are both generated by neutron capture on ^{6}Li events, the global rate of S_{ATM}^{ROFF} will be named $S_{Neutrons}^{ROFF}$ for reactor OFF and $S_{ATM}^{RON} + S_{IBD}^{RON}$ will be called $S_{Neutrons}^{RON}$ for reactor ON. In order to evaluate the level of atmospheric background in the reactor ON sample, an atmospheric background model is built as a function of the pressure as presented in section 4.3.3:

$$\begin{aligned} S^{ROFF}_{Residuals} &= S^{ROFF}_{TOT} - S^{ROFF}_{FPNTs} - S^{ROFF}_{BiPo} - f_{pressure} \\ S^{RON}_{IBD} &= S^{RON}_{TOT} - S^{RON}_{FPNTs} - S^{RON}_{BiPo} - f_{pressure} \end{aligned} \tag{4.30}$$

From that equation, one can also derive the statistical uncertainties. For the direct measurements, the statistical uncertainty is determined using Poisson statistics: the uncertainty of a given yield is the square root of the number of events. As it was discussed in section 4.3.3, the statistical uncertainty of the pressure model is directly the statistical uncertainty of S^{ROFF}_{ATM}:

$$\sigma^2_{f_{pressure}} = S^{ROFF}_{TOT} + S^{ROFF}_{FPNTs} + S^2_{BiPo\to neutron}(S^{ROFF,sb}_{Bipo+Acc} S^{ROFF,sb}_{FPNTs}) \tag{4.31}$$

The total squared statistical uncertainty for the reactor OFF residuals is then:

$$\sigma^2_{residuals} = 2 \times \left(S^{ROFF}_{TOT} + S^{ROFF}_{FPNTs} + (S^{ROFF,sb}_{Bipo+Acc} + S^{ROFF,sb}_{FPNTs}) \right) \tag{4.32}$$

And for the IBD signal:

$$\begin{aligned} \sigma^2_{IBD} = S^{RON}_{TOT} + S^{RON}_{FPNTs} + S^2_{BiPo\to neutron}(S^{RON,sb}_{Bipo+Acc} + S^{RON,sb}_{FPNTs}) \\ + S^{ROFF}_{TOT} + S^{ROFF}_{FPNTs} + S^2_{BiPo\to neutron}(S^{ROFF,sb}_{Bipo+Acc} S^{ROFF,sb}_{FPNTs}) \end{aligned} \tag{4.33}$$

After subtractions of all identified background components, the reactor OFF data should be compatible with zero, with no other background left in the dataset while only IBDs should remain in the reactor ON dataset. From reactor ON subtracted dataset, one can then derive the energy and travelled distance of the neutrino necessary for an oscillation analysis.

4.6.2 Subtraction on the Open dataset

In order to validate the subtraction method, the procedure has been performed on a small dataset before a full-scale study. On figure 4.42 is represented the variation of the rates of the different types of events through time on the open dataset. The top part of the figure represents the rates without any exposure correction. Each rate is computed as described in the section 4.6.1. The bottom part of the figure 4.42 shows the rates corrected by the exposure time, as it was described in section 4.3.4, in order to build the pressure model later. After correction of the exposure time, one can see the good stability of both accidental and BiPo rates. On the variation of the corrected rates, the transition reactor ON-OFF is visible at the tenth of July with a drop of the total rate. The remaining variations of the $S_{Neutrons}$ rates are then expected to be induced by pressure variation. The atmospheric background is thus evaluated with a pressure model as described in section 4.3.3. This model is represented on figure 4.43. As one can see, the variations of pressure are so small on the dataset that the model is nearly flat with even a positive correlation fitted, and the errors on the data points leads to larger error in the pressure model. However, because the variations of pressure are small, the variations of atmospheric background are also small and with this pressure model one can still do a background subtraction test.

Reactor OFF: After subtraction of all backgrounds, the residuals in the reactor OFF data is -2 ± 4 (stat) events per day. It is compatible to a 0 event as expected in the reactor OFF. The absence of excess in the reactor OFF sample is also stable in time, as can be seen on the figure 4.44. The reactor OFF data are then

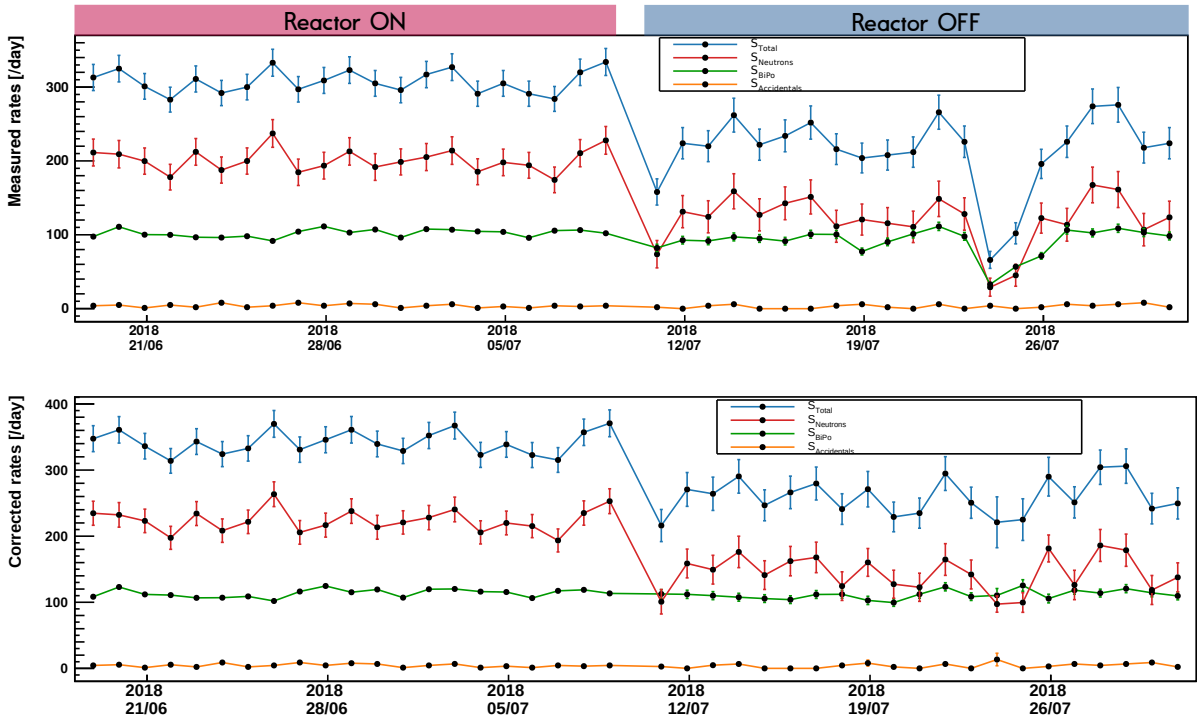


Figure 4.42: Rates for each source of events in the dataset. Top: Rates without any exposure time correction. Bottom: Rates corrected by the exposure time. The $S_{Neutrons}$ rate corresponds to atmospheric background for reactor OFF data and atmospheric background plus IBDs for reactor ON.

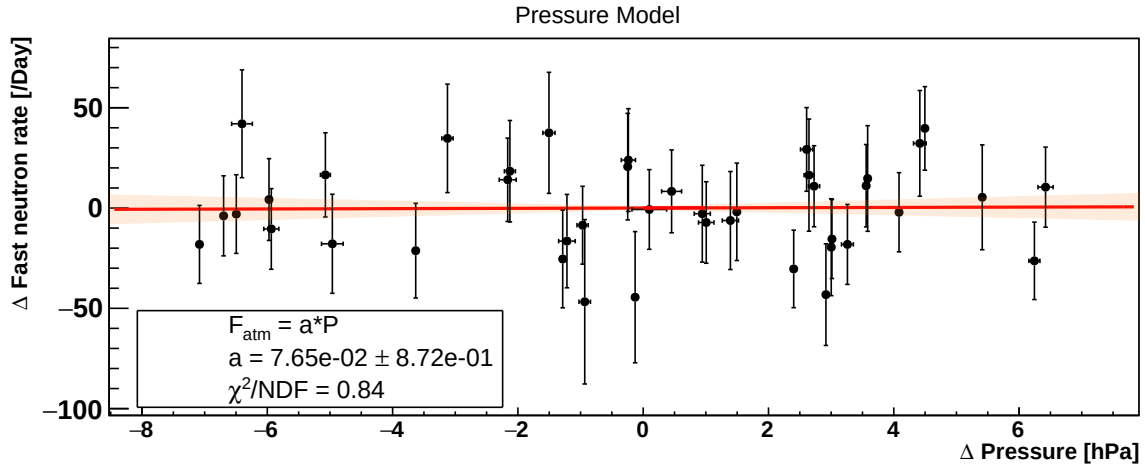


Figure 4.43: Pressure model for the open dataset.

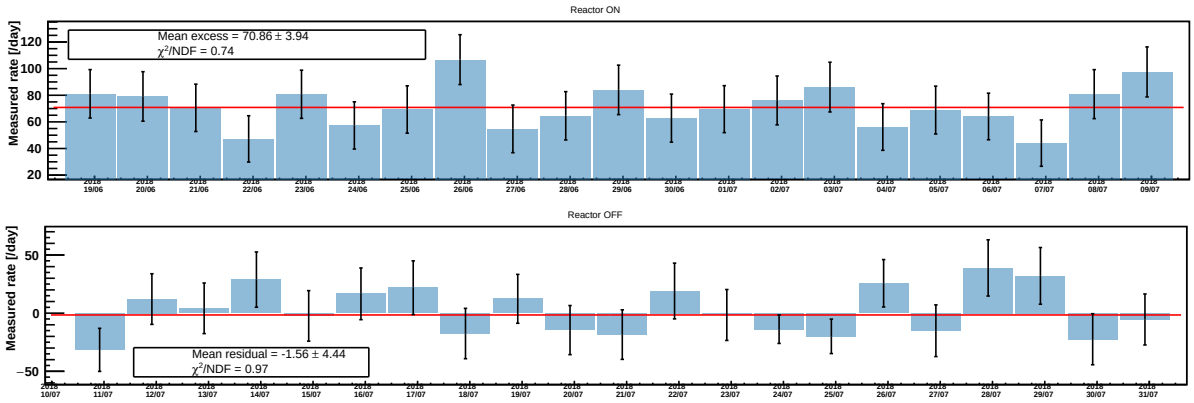
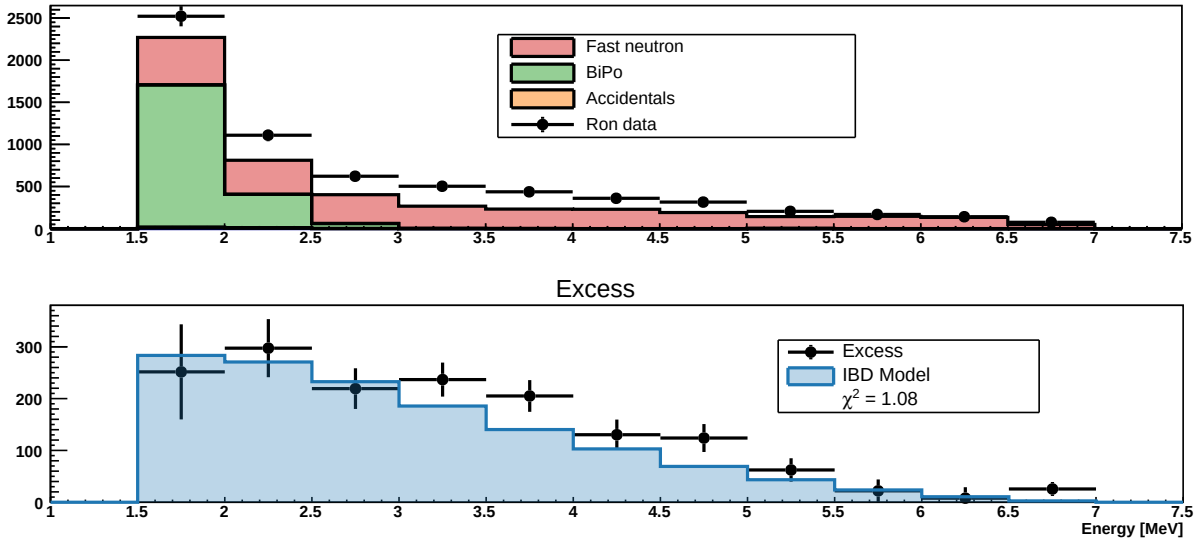


Figure 4.44: Measured event rate after subtraction of all background sources for reactor ON (top) and reactor OFF (bottom). The red line represents a zero degree polynomial fitted to the data.

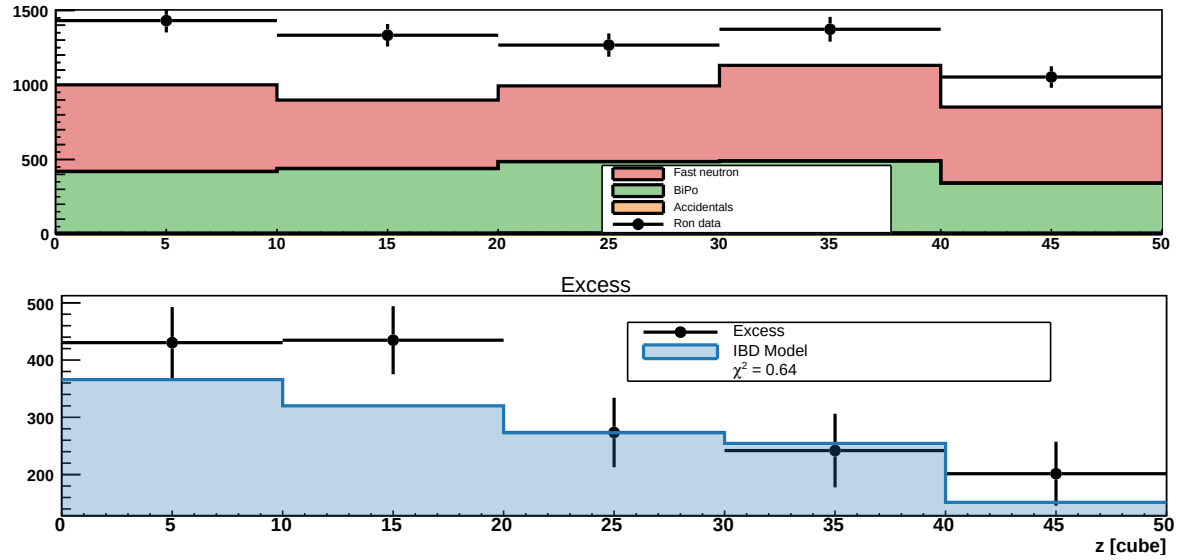
used to extract each background shape for energy and position. Those distributions will then be rescaled with the yield of each component measured on the reactor ON. The BiPo shapes are taken directly in from the BiPo sideband (accidental subtracted). The accidental shapes are taken from the FPNT data. The atmospheric shapes are extracted from the subtraction of the two other components from the reactor OFF data.

Reactor ON: The result of the subtraction for a selection cut on the BDT function larger than 0.57 is represented on figure 4.44. The average excess rate measured is 71 ± 4 (stat) events per day for the reactor ON. For this selection, 65 events were expected per day. Considering only the statistical error on the reactor OFF subtraction, the result is at less than $1.5\ \sigma$ than the prediction, which is expected considering the systematic uncertainties that were not computed here. Furthermore, the residuals are stable over all the reactor OFF period and compatible with 0. On the figure 4.45 can be seen the distribution of the different backgrounds and for the subtracted excess for the energy and Z position inside the SoLid detector. Both distributions are in agreement with the predicted IBD Monte-Carlo predictions, however the statistical uncertainty on the data points is high due to the level of background and the low number of days, 21, used for the subtraction. On the z distribution, one can see the effect of the geometry of the detector: the further is the interaction point from the reactor, the fewer antineutrinos interact in the detector. The last module low detection efficiency is also due to many dead channels in the latest plans of the detector. The evolution of the signal and background rates as a function of the BDT cut is represented on figure 4.46 and this evolution shows a good statistical agreement between the signal prediction and measurement.

This result on the open dataset is a convincing step in order to move forward and open more reactor ON data. It gives the confidence in a stable and robust method to extract antineutrinos from a reactor ON sample. Before opening more reactor cycles, the stability over time of the subtraction procedure must be tested.



(a) Energy



(b) Z position

Figure 4.45: Top: background breakdown of the Energy (a) and the z position of the interaction (b) for the 21 days of Reactor ON data. Each background component rate is computed via the subtraction method, and then all background energy distributions are computed on the Reactor OFF data and scaled with the measured rate. Bottom: Energy and z position distributions of the background subtracted reactor ON excess and of IBD Monte-Carlo. The IBD MC is normalized at the prediction IBD rate.

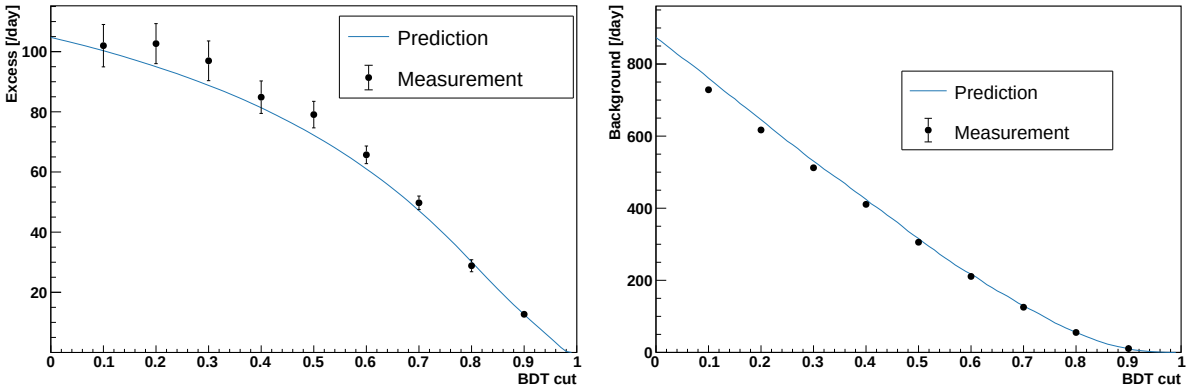


Figure 4.46: Left: excess rate measured and predicted for different values of the BDT cut. Right: background rate measured and predicted for different values of the BDT cut. The BDT values of both one-gamma and two-gamma categories are set to the same values around the curve.

4.6.3 Reactor OFF only subtraction

In order to test the stability of the subtraction over time and derive a systematic error from the procedure itself, a subtraction of two reactor OFF periods can be performed. This is essential to test the subtraction method on reactor OFF only data before unblinding the reactor ON data to ensure all the final results will be understood.

4.6.3.1 Stability of the energy measurement

The first step of the reactor OFF - reactor OFF subtraction is to ensure a good stability of the individual background measurements through time. To do so, three reactor OFF periods have been used: the *open dataset* (07/2018), the period $6-7$ (12/2018, 01/2019 and 02/2019) and the period $11-12$ (12/2019, 01/2020). Those periods were chosen due to their length, with at least 20 days in each period, and the time of several months between each of them. For each period the different background components of each dataset were computed with the FPNTs and the BiPo sideband methods. The atmospheric background component is defined as the residual data after measurement and subtraction of both accidental and BiPo backgrounds, as in equation 4.29. Because the accidental background represents less than 1% of the background contamination after signal selection, only the atmospheric and BiPo backgrounds will be presented here. On the figure 4.47 can be seen the superposition of the energy distributions of both BiPo and atmospheric backgrounds for the *open dataset* and the period $6-7$. For the two types of background sources, the energy distributions are statistically compatible between the two periods. On the BiPo energy spectrum, the energy bins over 3 MeV are irrelevant for the analysis and the large ratios are due to a few accidental coincidences left in the sample. The energy bin at 2.5 MeV shows the largest deviation of the Bipo. It corresponds to 2% of the BiPo distribution and can be explained by a small drift in the energy scale between two calibration campaigns. The data to Monte-Carlo comparisons showed on section 4.4.2 showed however that energy scale was controlled at a few percent. The same comparison is done between the *open dataset* and the cycle $11-12$ on figure 4.48. For those periods, the energy distributions of the atmospheric background are not in agreement any more. The energy distribution of the atmospheric background for the period $11-12$ has a component at low-energy below 3 MeV that is

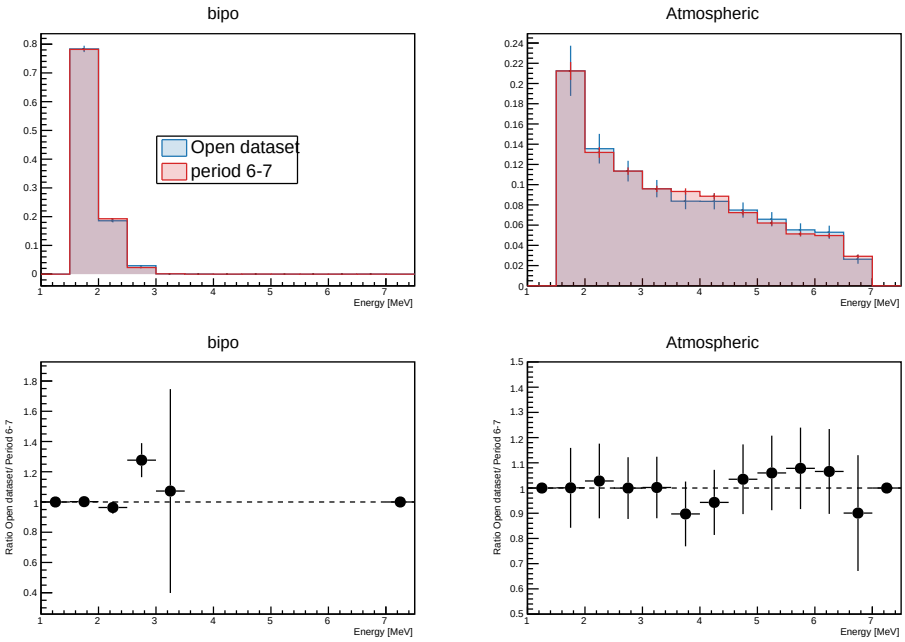


Figure 4.47: Energy distribution of the BiPo and the atmospheric backgrounds from reactor OFF data. The blue distributions are obtained on the open dataset and the red distribution from the $6-7$ period. The ratios between the two distributions can be seen on the bottom plots.

not present in the *open dataset* one. This can be explained by a BiPo contamination that was not measured properly for this period.

The main suspect of this error in the BiPo rate measurement is the scaling factor of the BiPo sideband used to measure the BiPo rate. The scaling factor is computed once with a reference distribution, and the same is used for every period. A small variation of the BiPonator response function could cause a variation of the sideband scaling factor and lead to a wrong measurement of the BiPo rate in a data sample. The BiPonator for the two periods *open dataset* and $11-12$ is represented on the figure 4.49, and shows a variation of the BiPonator between the two periods. The BiPonator distribution for the *open dataset* is higher in the BiPo sideband than for the $11-12$ period. However, the scaling factor in the *open dataset* is lower than the $11-12$ period. It explains why the BiPo contamination of the $11-12$ period is underestimated in the figure 4.48 when the $11-12$ period sideband is rescaled with the *open dataset* scaling factor.

The instability of the BiPonator response can be resolved by two actions: the computation of the scaling ratio with a BiPo enriched data for each period and if a variation of the BiPonator response is seen around a reactor ON period, the IBD rate prediction must be corrected according to the variation. For the reactor OFF - reactor OFF subtraction, once the scaling ratio is computed for each period, the different energy distributions are now in agreement for the *open dataset* and the period $11-12$ as can be seen on the figure 4.50.

4.6.3.2 Stability of the energy subtraction

With the BiPo scaling factor re-computed for every reactor OFF period, one can perform a subtraction using the same method as the reactor ON - reactor OFF subtraction. In this study, the open dataset

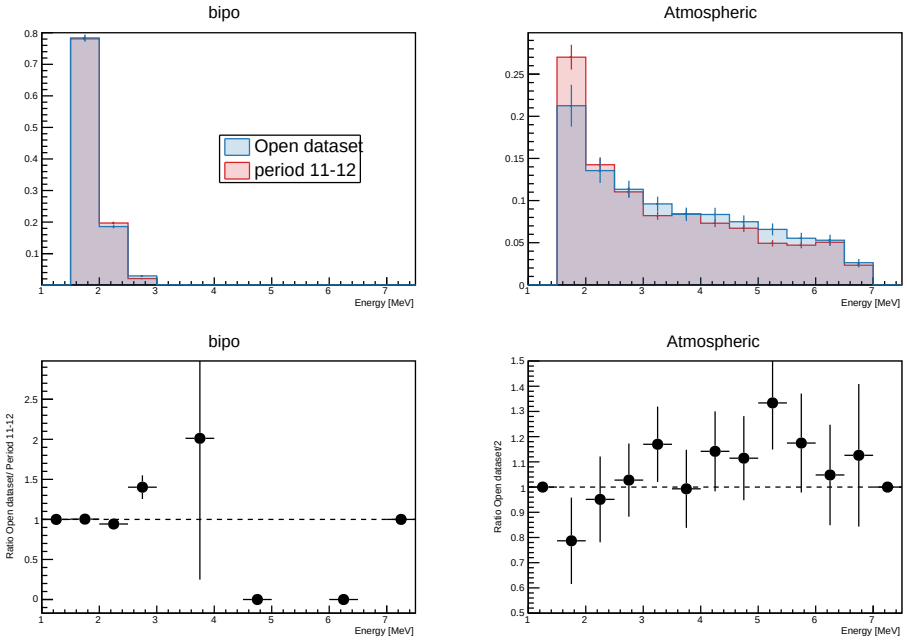


Figure 4.48: Energy distribution of the BiPo and the atmospheric backgrounds from reactor OFF data. The blue distributions are obtained on the open dataset and the red distribution from the $11-12$ period.The scaling factor of the BiPo sideband is computed for each period. The ratios between the two distributions can be seen on the bottom plots.

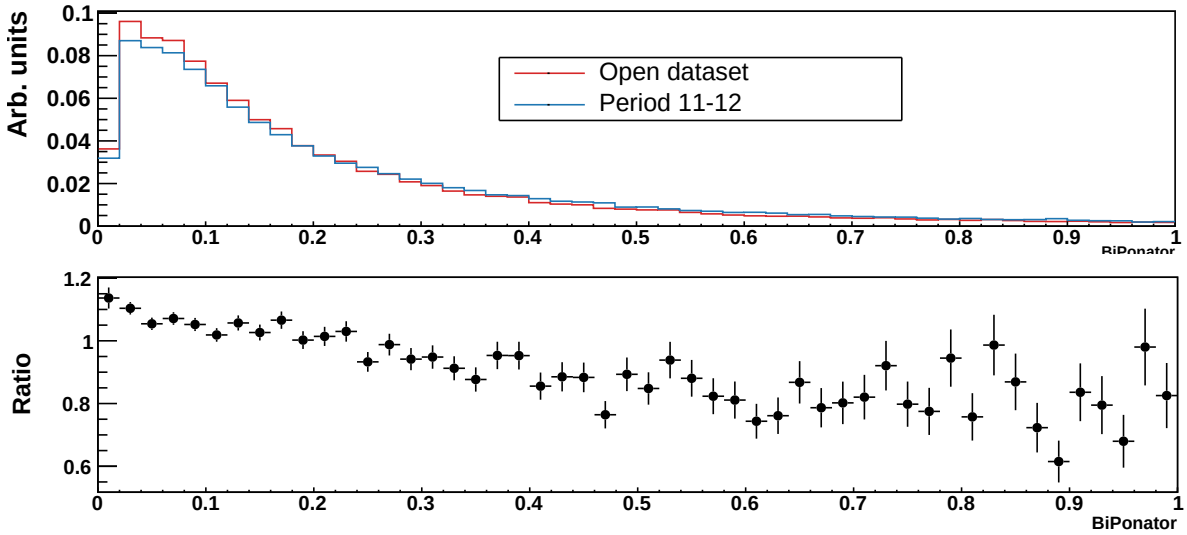


Figure 4.49: BiPonator distribution for BiPo enhanced data for the two *open dataset* and $11-12$ periods.

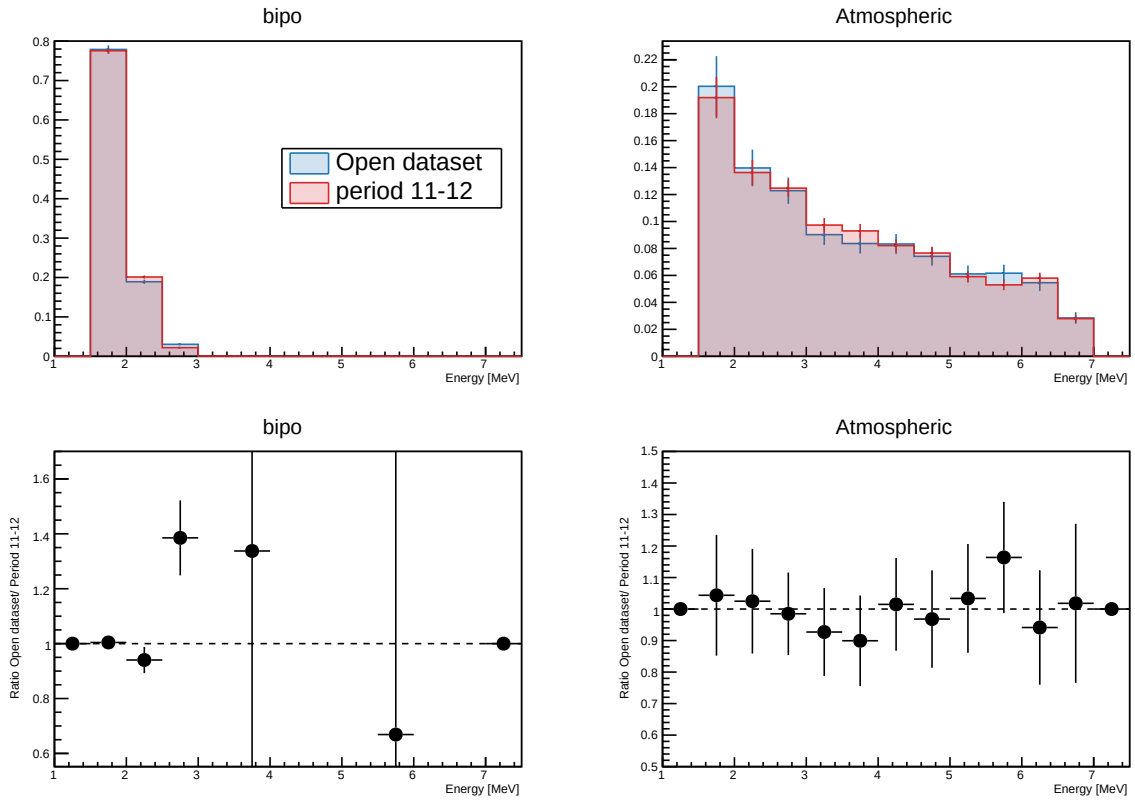


Figure 4.50: Energy distribution of the BiPo and the atmospheric backgrounds for from reactor OFF data. The blue distributions are obtained on the open dataset and the red distribution from the $11-12$ period. The scaling factor of the BiPo sideband is computed for each period. The ratios between the two distributions can be seen on the bottom plots.

has been subtracted to all the periods from the $4-5$ to the $14-15$ except two periods, the $8-9$ and $9-10$: because of the heating of the detector in May 2019, the BiPonator had large day-to-day variations. The energy distribution subtraction for each period individually can be seen on the figure 4.51. On each figure, the *open dataset* is used to generate the background shapes, and each background yield is computed on the period considered. The pressure model is rescaled with the average atmospheric rate from the *Open dataset*.

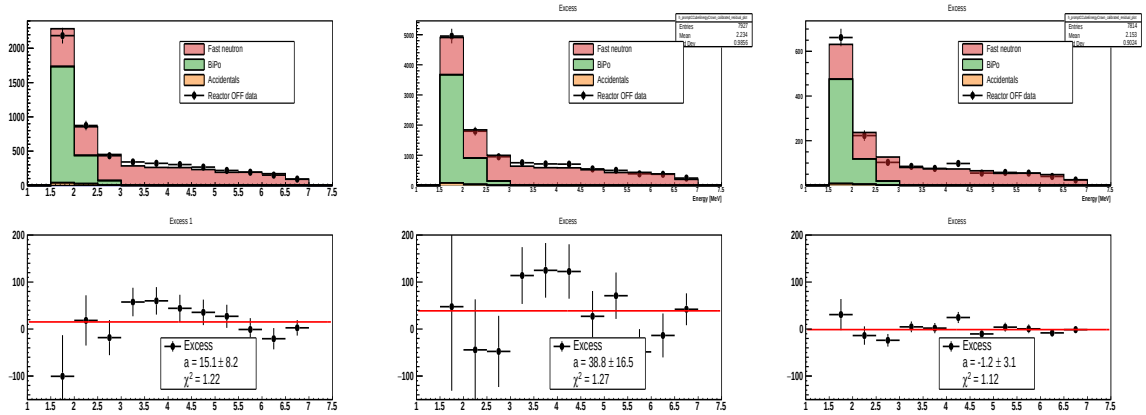
One can see that for some periods as the $7-8$ or the $13-14$ that the first energy bin between 1.5 and 2 MeV that presents a large deficit between the total reactor OFF data compared to the sum of all backgrounds energies. The subtracted distribution for the energies larger than 3 MeV is consistent with a 0 excess, so one can assume that the atmospheric rate is correctly measured while the BiPo component is overestimated. Furthermore, on other periods as the $4-6$ ot the $5-6$, the subtraction gives an excess of events between 3 and 4.5 not compatible with 0. On the figure 4.52, one can see the BiPo rate per day in the BiPo sideband and rescaled to match the IBD selection. The BiPo rate seems constant, around 860 events per day, between the periods $4-5$ and $6-7$. Then, a clear decrease of the BiPo rate can be seen between the periods $6-7$ and $7-8$ with afterwards a stable BiPo at 730 events per day. This underlines the previous conclusion on the BiPo sideband that can vary over time. From that figure, it seems that the BiPo sideband changed only once between the periods $6-7$ and $7-8$. This might be a hint for later studies to understand the stability of the BiPonator. Furthermore, on the bottom plot that represents the BiPo rate rescaled to the selection, regardless of the BiPo rate variations in the sideband, the BiPo rate seems to be relatively stable around 115 events per day. The only period that does not follow the constant trend of the BiPo rate is the period $7-8$ that shows a significantly higher rate of about 150 events per day. This confirms the observation on the figure 4.51 (d) where the BiPo was overestimated in the period $7-8$. The reason of the overestimation of the BiPo in this region is still unknown, but those days might be the first one to study for a deeper BiPonator stability evaluation. The differences seen on the individual subtractions thus underline an instability of the BiPo measurement over time and could advocate for an IBD extraction using only the reactor OFF periods close from each reactor ON cycle.

4.6.3.3 Systematic uncertainty on a global subtraction

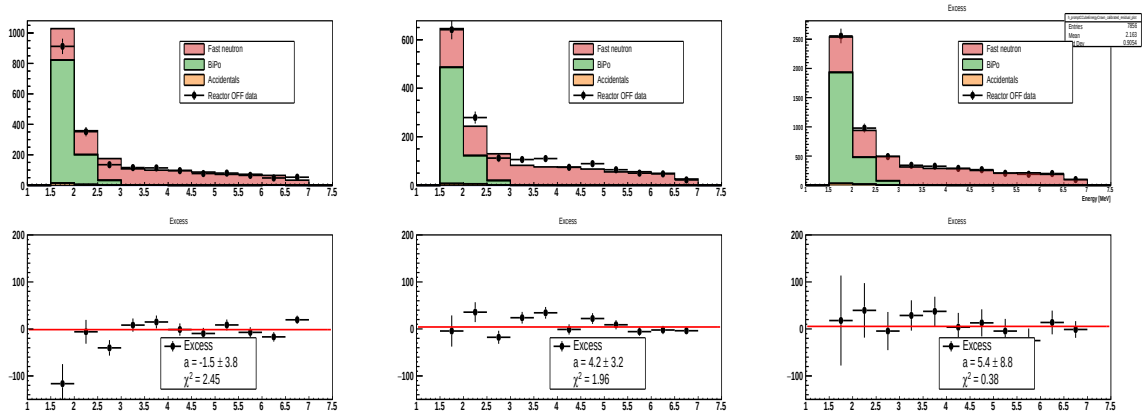
To derive a systematic uncertainty on the antineutrino rate per energy bin per day, one can average all the excess distributions weighted on the number of days used on each period:

$$S_{syst}(E) = \frac{1}{ntot} \sum_{i=4-5}^{14-15} S_i \times ndays_i \tag{4.34}$$

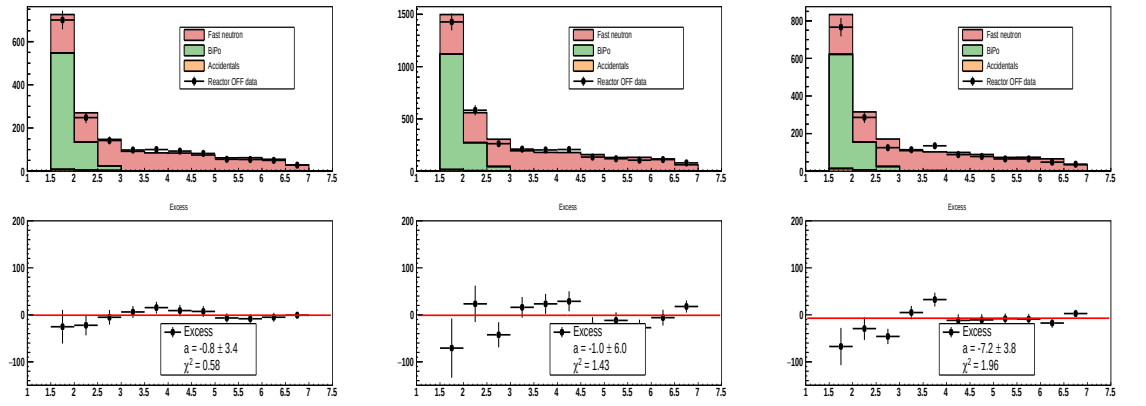
With $S_{syst}(E)$ the systematic uncertainty on the number of measured antineutrinos per day and per energy bin, S_i the excess per energy bin of the reactor OFF period i, $ndays_i$ the number of days in the period i and $ntot$ the total number of reactor OFF days used. The obtained systematic uncertainty on on the antineutrino rate per energy bin can be seen on the figure 4.53. On average there is an excess of 2 events per day and most of the energy bins are compatible with 0 at $1\ \sigma$. However, the energy bins between 3 and 4.5 MeV show a systematic excess of a total of 15 events per day with the subtraction. From the figure 4.51, it is clear that the excess on those energy bins is driven by the subtraction of the period $5-6$. Indeed, by removing this period, the final averaged subtracted spectrum can be seen on the figure 4.54. For illustration, without the $5-6$ period, the error on the bins between 3 and 4.5 MeV is



(a) period 4- 5 **(b)** period 5- 6 **(c)** period 6- 7



(d) period 7- 8 **(e)** period 10- 11 **(f)** period 11- 12



(g) period 12- 13 **(h)** period 13- 14 **(i)** period 14- 15

Figure 4.51: Reactor OFF - Reactor OFF energy subtraction with the *open dataset* subtracted to the periods $4-5$ to $14-15$ without the periods $8-9$ and $9-10$. For each period and each energy distribution, each background component rate is computed via the subtraction method and then all background energy distribution are scaled with the measured rate and stacked to recreate the reactor off energy distribution of the period. The differences between the measured background distribution and the initial reactor off energy distribution is represented in the bottom plots.

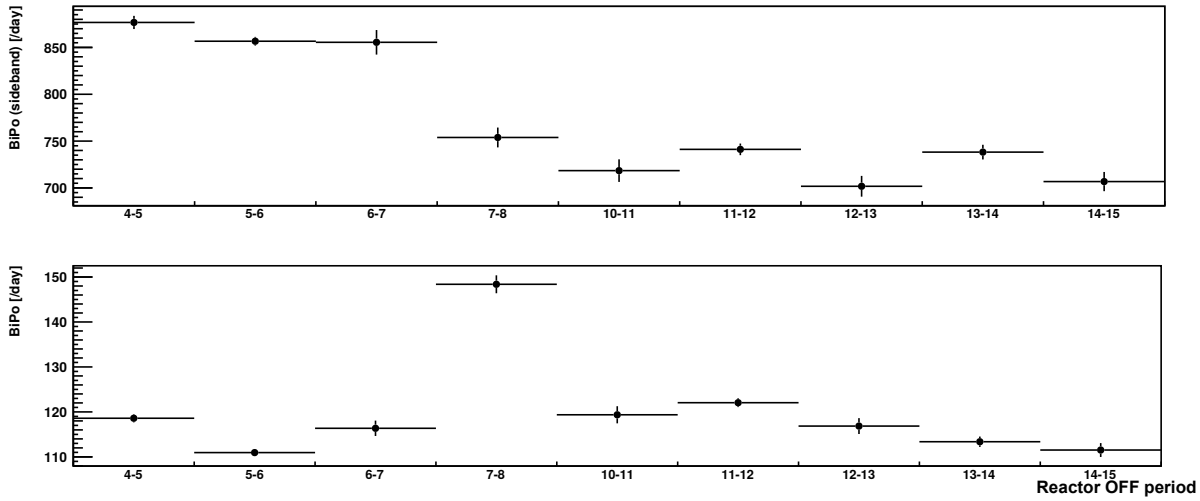


Figure 4.52: BiPo rate per day in the BiPo sideband (top) and in the IBD selection (bottom) for different reactor OFF periods.

lowered to 6 events per day. This last effect is another argument in favour of a reactor ON background subtraction, using only the reactor OFF periods close in time to avoid a too large variations of the energy spectrum decomposition. The same systematic study can be done with the position of the interaction in the detector for all the periods, as can be seen on the figure 4.55. Here again the quality of the subtraction is dependent on the period considered. The period $5-6$ for instance shows an excess not compatible with 0. The averaged residual can be seen on figure 4.56. Again, the errors are quite large compared to the expected signal. However, one can see that the statistical errors on the averaged residuals are also dominant (at the level of the residuals themselves). A conclusion from that work would be that trying to derive an uncertainty on the subtraction method in that way would overestimate the real systematic uncertainty considering the level of background faced. Small statistical fluctuations on the background yields could indeed lead to a large excess on the reactor OFF - reactor OFF subtraction. Because of lack of more precise way to estimate a systematic error induced by the subtraction of the *open dataset* reactor OFF to the reactor ON, this method could be used as a very conservative way to measure the systematic uncertainty related to the subtraction, however no competitive exclusion analysis could be done with that level of errors.

4.7 Conclusion

In a conclusion, a way of extracting the antineutrino yield from the reactor ON data has been presented, with results consistent with the predictions from the Monte-Carlo simulations. New topologies of the prompt event have been derived in order to achieve the best performances possible. However, the performances of the analysis are limited due to the high contamination of background, with a final number of antineutrino expected to be detected of 65 per day for 243 background events. The study of the stability of the subtraction method showed variations of the BiPonator variable over time. This led to errors in the measurement of the BiPo yield in the data that can induce large systematic uncertainties due to the level of background faced in SoLid. In the next chapter, the sensitivity that can be achieved with those

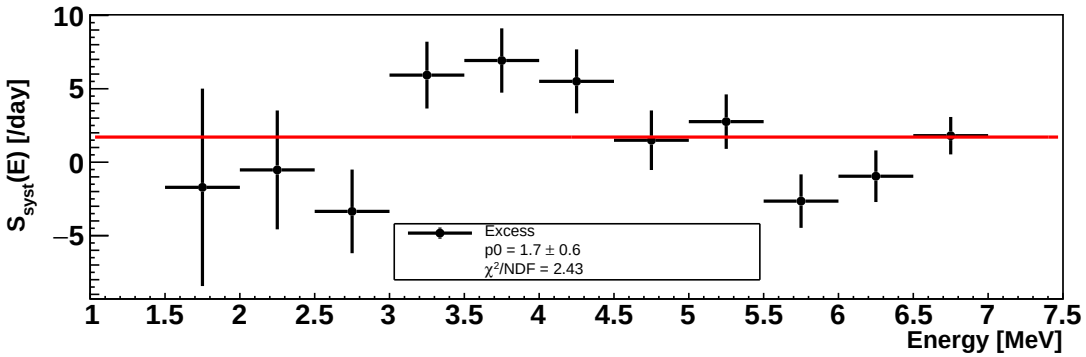


Figure 4.53: Systematic error per day on the energy spectrum subtraction.

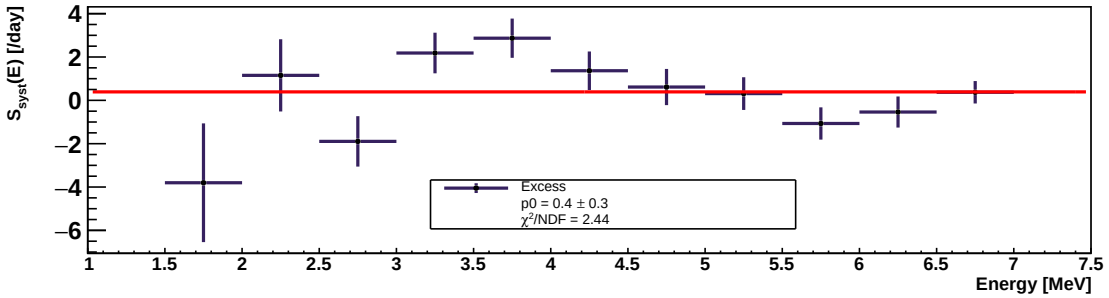
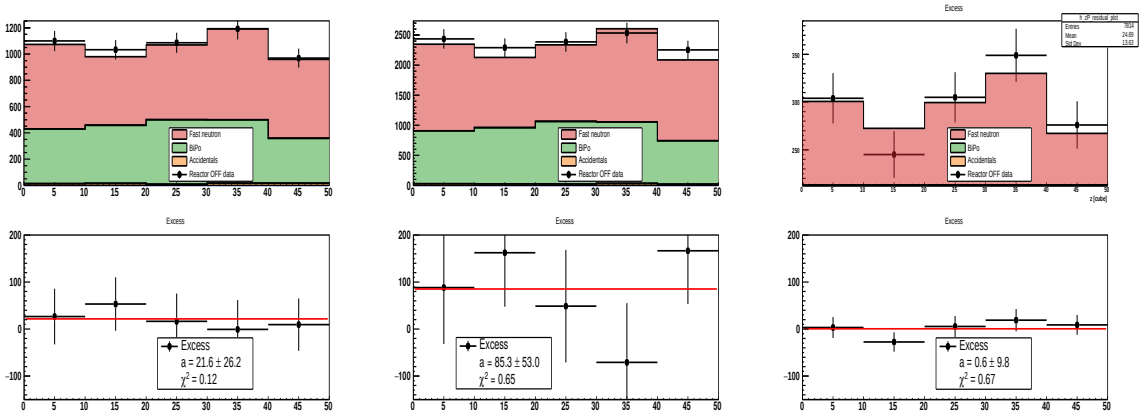
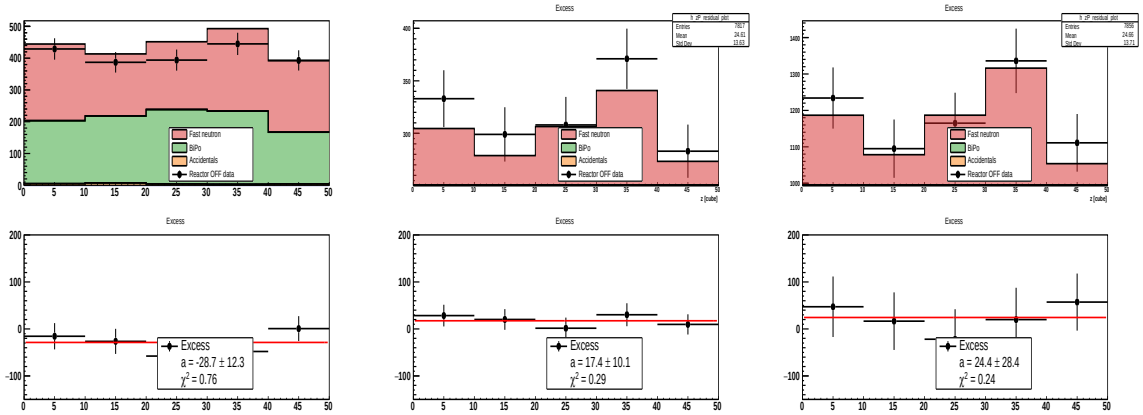


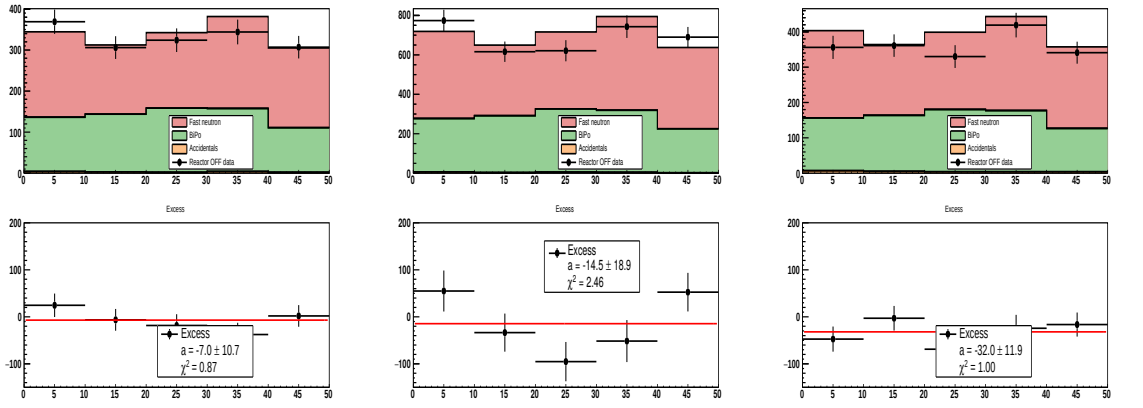
Figure 4.54: Systematic error per day on the energy spectrum subtraction. The period $5-6$ has been removed from the measurement.



(a) period 4- 5 **(b)** period 5- 6 **(c)** period 6- 7



(d) period 7- 8 **(e)** period 10- 11 **(f)** period 11- 12



(g) period 12- 13 **(h)** period 13- 14 **(i)** period 14- 15

Figure 4.55: Reactor OFF - Reactor OFF prompt z position subtraction with the *open dataset* subtracted to the periods $4-5$ to $14-15$ without the periods $8-9$ and $9-10$. For each period and each energy distribution, each background component rate is computed via the subtraction method and then all background energy distributions are scaled with the measured rate and stacked to recreate the reactor off energy distribution of the period. The differences between the measured background distribution and the initial reactor off energy distribution are represented in the bottom plots.

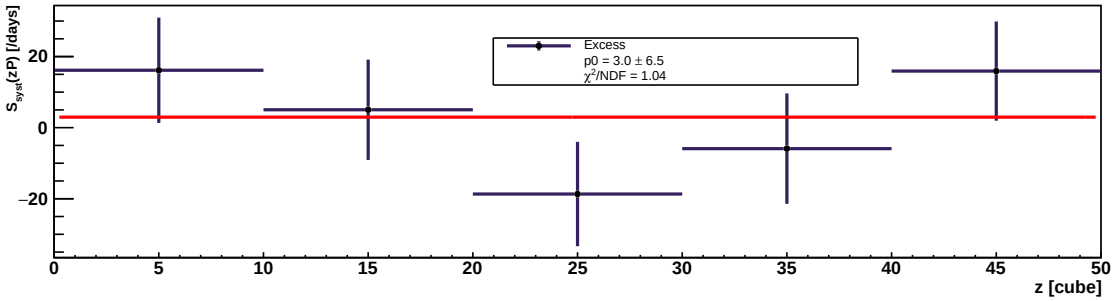


Figure 4.56: Systematic error per day on the Z position subtraction for the Reactor OFF open dataset.

performances will be presented, along with possible ways to improve the antineutrino analysis.

Chapter 5

Outlook and prospects

In this section, the expected sensitivity to neutrino oscillation from the analysis presented in the chapter 4 will be discussed. An alternative analysis that aims to achieve a good purity in the signal selection will then be presented. Lastly, the SoLid Phase II, with the first calibration work will be presented.

5.1 Expected sensitivity

The final step of the SoLid analysis is the search for an oscillation to a sterile neutrino signal, or the rejection of the sterile neutrino hypothesis. In this section, the oscillation fit procedure will be briefly presented, as it was not a part of the analysis I was deeply involved in. More details can be found in Ianthe Michiels' thesis [111]. In order to probe the oscillation toward a sterile neutrino state, one can derive the appearance probability in a two-neutrino approximation:

$$P(\bar{\nu_e}) \rightarrow P(\bar{\nu_e}) = 1 - \sin^2(2\theta_{41}) \sin^2 \left(1.27 \frac{\Delta m^2_{14}[\mathrm{eV}^2] L[\mathrm{m}]}{E[\mathrm{MeV}]} \right) \tag{5.1}$$

where L is the distance travelled by the neutrinos between the reactor core and the interaction point in the detector and E is the antineutrino energy. This is the same equation as the equation 1.33 presented in the section 1.5 but expressed in SI units. The two variables are the direct output of the analysis work presented in chapter 4. For a given couple (Δm^2_{41}, $\sin^2(2\theta_{14})$), one can build a predicted result P, that can be compared with the extracted antineutrino shape D.

Those two datasets are gathered in the form of a ($n_L \times n_E$) matrix, with n_L the number of bins for the neutrino distance and n_E the number of energy bins. In the case of SoLid, the current choice is ($n_L \times n_E = 5 \times 11$) with a binning per detector module for L and a 0.5 MeV bin width for the energy. The choice of the binning is essential for the oscillation analysis, indeed, with a finer binning, an experiment would be more sensitive to higher oscillation frequencies. However, a finer binning would also lead to higher statistical fluctuations in the results. Furthermore, the energy binning is limited by the energy resolution of the detector. From the prediction P and the data D, several types of oscillation analyses can be performed. To avoid having to rely on a theoretical rate of IBD events, a shape only analysis has been chosen in the case of SoLid.

The fit strategy is then the test of a given hypothesis with χ^2 score between the prediction P and the

dataset D:

$$\chi^2 = (D-P)^\intercal C^{-1}(D-P) \tag{5.2}$$

With C the covariance matrix encoding the uncertainties on the antineutrino energy and travelled distance extraction and predictions. The details of the covariance matrix will be discussed in the next section. The χ^2 gives the statistical agreement between the data and the prediction data. A χ^2 test provides a confidence level for the rejection of a given hypothesis, either the no oscillation hypothesis or an oscillation driven by a given couple $(\Delta m^2_{41}, \sin^2(2\theta_{14}))$. The detail of the χ^2 analysis will be presented in the section 5.1.2.

5.1.1 Building a covariance matrix

From the knowledge of the detector, one can build a covariance matrix. This matrix encodes all the known statistical and systematic uncertainties in the antineutrino yield per energy and length bins. Building a covariance matrix is a difficult work as it requires a precise knowledge of all sources of uncertainties of the experiment. Those can be derived from the flux prediction, the detector acceptance, the IBD efficiency, the detector response, the knowledge of the background shape... Current work is ongoing to derive all the systematic uncertainty effects from the detector. One method to generate the covariance matrix that is used in SoLid is the use of IBD Monte-Carlo simulations. Given one parameter u, its nominal value $\bar{u}$ and the uncertainty associated σ_u one can generate a set of simulations by varying the input parameter u according to its uncertainty and nominal value. The covariance matrix can then be built as:

$$C_{\alpha\beta} = \frac{1}{N_{sim}} \sum_{i=1}^{N_{sim}} (P^i_\alpha - \bar{P}_\alpha)(P^i_\beta - \bar{P}_\beta) \tag{5.3}$$

where C is the covariance matrix, N_{sim} is the number of simulations, P^i is a prediction with a varied input parameter, and $\bar{P}$ is the prediction with the nominal input parameter. This method has been used to assess three of the major detector uncertainties:

Light yield uncertainty. From the light yield measurements presented in the chapter 3, a conservative 3 % uncertainty of the light yield measurement has been selected to reflect the cube to cube light yield variation within a plane. A set of simulations was generated with more than 14 million IBD events in total. For each simulation, each of the 3200 cube light yields is randomly generated from a Gaussian distribution centred around the nominal light yield value of the cube with a 3 % width.

Global energy scale. This uncertainty is derived from the ageing of the PVT. Because several months can happen between two calibrations, one needs to take into account the uncertainty related to the PVT light yield drift. This effect is corrected at each calibration campaign, however there are some periods of several months without any calibration campaigns. Simulations with a global light yield shift of -3 % and $+2$ % have been generated to quantify this effect. As the expected shift per year is -3 %, those light yield variations are very conservative.

Neutron detection efficiency. From the results of the neutron calibration detailed in [84], a random variation of the neutron detection efficiency has been applied per module. For the modules from 1 to 4, the neutron efficiency has been varied with a 1 % uncertainty and for the module 5, a 2 % uncertainty has

been used to vary the efficiency. The module 5 has a larger variation of efficiency due to inefficiencies in the last planes due to numerous of dead channels.

A final conservative $1\ \%$ uncertainty has been set globally on the distance travelled by the neutrinos to take into account acceptance effects due to the fuel loading configuration that can change from one cycle to another.

For now, all the uncertainties taken into account are quite conservative in order to have a first estimation of the systematic effects. The major uncertainties that still needs to be computed are the uncertainty on the subtraction method and the light leakage uncertainties. The covariance matrix also encodes the statistical uncertainties per energy and length bins that are derived from the equation 4.33

5.1.2 Frequentist method

Once the covariance matrix C has been computed, for a given couple (Δm^2_{14}, $\sin^2(2\theta_{14})$), with the prediction P one can build a χ^2 score:

$$\chi^2 = (D-P)^\intercal C^{-1}(D-P) \tag{5.4}$$

The χ^2 test allows to exclude the hypothesis of an oscillation with a given confidence level given by the χ^2 distribution for $(n_L \times n_E)$ degrees of freedom. In the case where oscillation hypotheses are tested, the χ^2 test also provides the best fit point. This point is the hypothesis that minimizes the χ^2. The parameter space is then divided into 200×150 bins and 30 000 χ^2 are computed for a given dataset and covariance matrix. The minimal $\chi^2 = \chi^2_{BF}$ gives the so called *best fit* point. In order to derive the allowed and rejected zones in the parameter space, one can follow the Feldman Cousins prescription [112]. For a given point p in the parameter space, $\mathcal{O}(1000)$ fake experiments are generated. Each toy is generated according to the prediction P_p and with a randomization at the level of the covariance matrix uncertainties. For each fake experiment, a $\Delta\chi^2 = \chi^2 - \chi^2_{BF}$ is built. The distribution of all the $\Delta\chi^2$ for the prediction p is then used to generate a rejection confidence level: a $90\ \%$ confidence level is defined by $\Delta\chi_{c,p}$ such that $90\ \%$ of the pseudo experiments have a higher $\Delta\chi^2$. The map of the critical $\Delta\chi^2_{c,p}$ is then generated from the 30 000 points in the oscillation parameter space.

For each of the $\mathcal{O}(1000 \times 30000)$ pseudo experiments, a best fit value has to be computed by scanning all the parameter space. This requires a lot of time and computing power. In order to reduce the time consumption, one can perform a *Raster scan* [113]. For a given hypothesis p ($\Delta m^2_{14,p}$, $\sin^2(2\theta_{14,p})$), the best fit point is searched on a scan of a slice of $\sin^2(2\theta_{14})$ at $\Delta m^2_{14,p}$ fixed. This allows a reduction of the computing time of a factor about 100. However, one needs to note that the best fit χ_{BF} is slightly overestimated. The distribution $\Delta\chi_{c,p}$ measured with the *Raster scan* will thus be a slightly underestimated. This leads to a larger number of points in the parameter excluded when using a *raster scan* compared to a global scan and an overestimation of the discrimination power in $\sin^2(2\theta_{14})$. A global scan would thus be needed for a precise parameter determination.

Because only a few days of reactor ON data have been unblinded for the antineutrino analysis, one cannot derive an exclusion contour yet. However, a preliminary sensitivity contour can be produced. This gives the average expected exclusion an experiment can provide. One way to compute it is to generate an exclusion contour with many IBD Monte-Carlo events, generated with the no oscillation hypothesis as the dataset D. The IBD Monte-Carlo is then scaled to the expected number of events in the whole two years of data taking for the Phase I data. On figure 5.1 can be seen a sensitivity contour

expected for the whole phase I dataset ($\sim$ 300 days of reactor ON and $\sim$ 180 days of reactor OFF). This contour was generated with an expected S/B ratio of 0.2 and an excess of 90 events per days. Those performances are slightly better than what has been presented in the chapter 4 and obtained with an alternative analysis from colleagues at Imperial College of London. From this sensitivity contour, one might expect the final exclusion contour to be at the level of the RAA best fit point, however it will depend on the data themselves and can only be verified with the real exclusion contour. Furthermore, one can see that the current performances of the analysis is not competitive compared to other experiments and will only be able to exclude oscillation parameters already excluded at > 99 % C.L. by several experiments like STEREO or PROSPECT.

5.2 Analysis improvement

The work presented in this section is currently under development by the LPC Clermont-Ferrand team. It relies on the selection of event topologies the less contaminated by backgrounds in order to have the purest signal where the two gammas produced by an IBD interaction are detected. This analysis is briefly presented here with its current results.

5.2.1 Reconstruction software improvement

One lead to further improve the discrimination power of the detector lies in the upgrade of the reconstruction software. This work is mainly developed by the LPC Clermont-Ferrand team. In its current implementation, the CCube algorithm described in section 2.7.2.3 only takes a flat input as the system matrix. This matrix should encode all the repartition of the light in the fibres of a plan for an energy deposit in a given cube. This flat matrix approximation implies that for a given energy deposit in a cube, one can expect a 1/4 distribution of the scintillating photons over the four MPPCs connected to the cube. However, with the attenuation and coupling effects described in the section 3.3.2 one knows that the number of PA read out by one MPPC will not necessarily be the same as the three other ones. Furthermore, due to light leakages from a cube to its neighbouring cubes, for an energy deposit in a cube, more than only four MPPCs are expected to be triggered. Thus, if an annihilation gamma deposit some energy in a cube, more than one cube can be reconstructed and that would induce a more difficult gamma tracking.

A first implementation of the new system matrix has been proposed by using the attenuation lengths computed in the calibration as a first system matrix, combined with an assumption of uniform light leakages at 6 % on all cubes, from first estimations with muon data crossing the detector. This matrix is only a rough estimation of the final system matrix but the improvement allowed to perform an analysis on very low background samples that will be presented in section 5.2.2. The computation of the system matrix can also be performed with horizontal muon. Those muons cross the detector in the z direction, they enter the detector in the first plane, and they exit in the last one (or the other way around). The long track of the muons with many cubes allows a good tracking of the cubes crossed by the muon. On figure 5.2 can be seen the track of a simulated horizontal muon projected on both (x, z) and (y, z) directions. With the identified true cube from the muon energy deposit, one can derive the given light leakage to the neighbouring cubes. Work on that last method would allow a measurement of the leakages expected in individual cubes and would give a finer understanding of the energy reconstruction on the energy

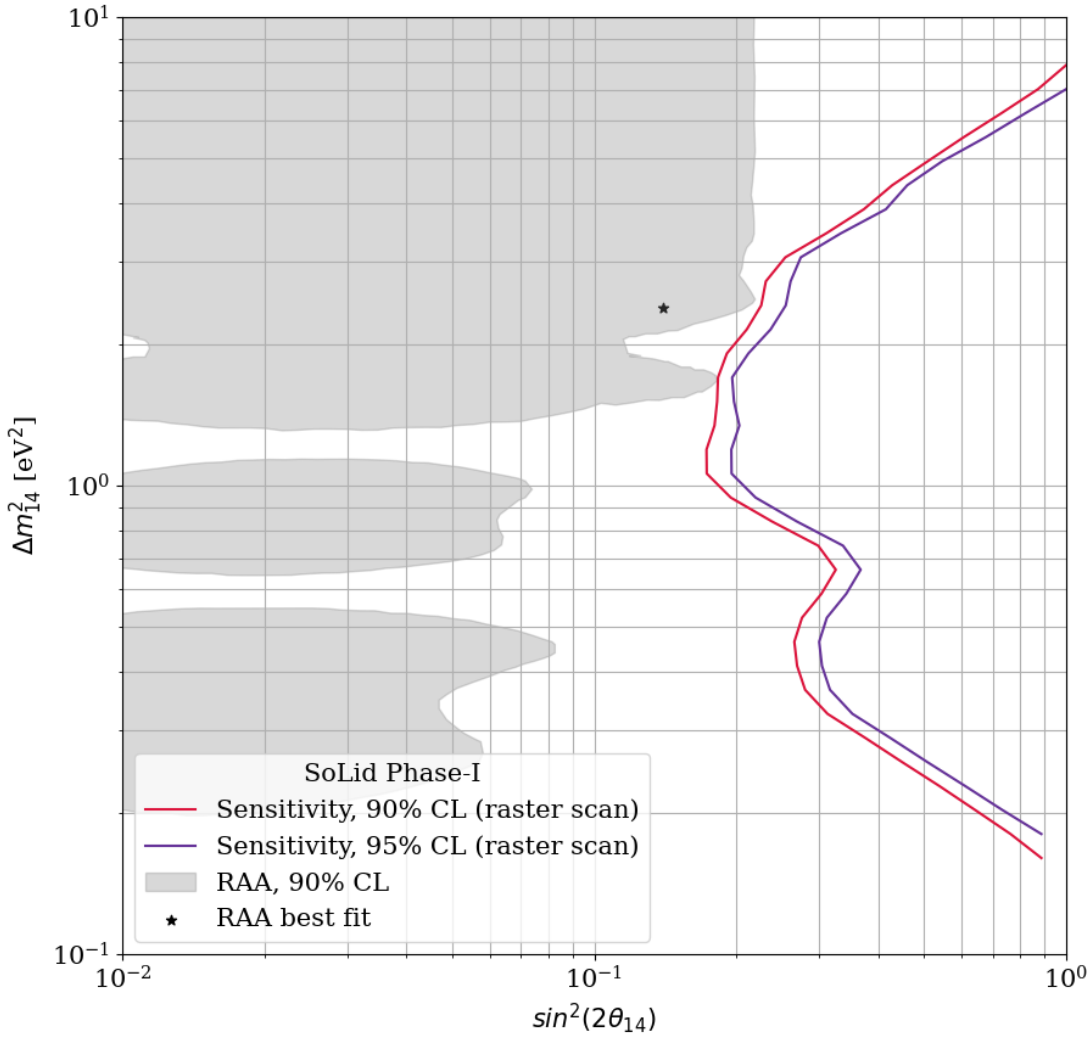


Figure 5.1: Preliminary sensitivity contour with the current antineutrino analysis for $\sim$ 300 days of reactor ON and $\sim$ 180 days of reactor OFF and the first systematic uncertainties. It was computed for a S/B ratio of 0.2 and an excess of 90 events per day.

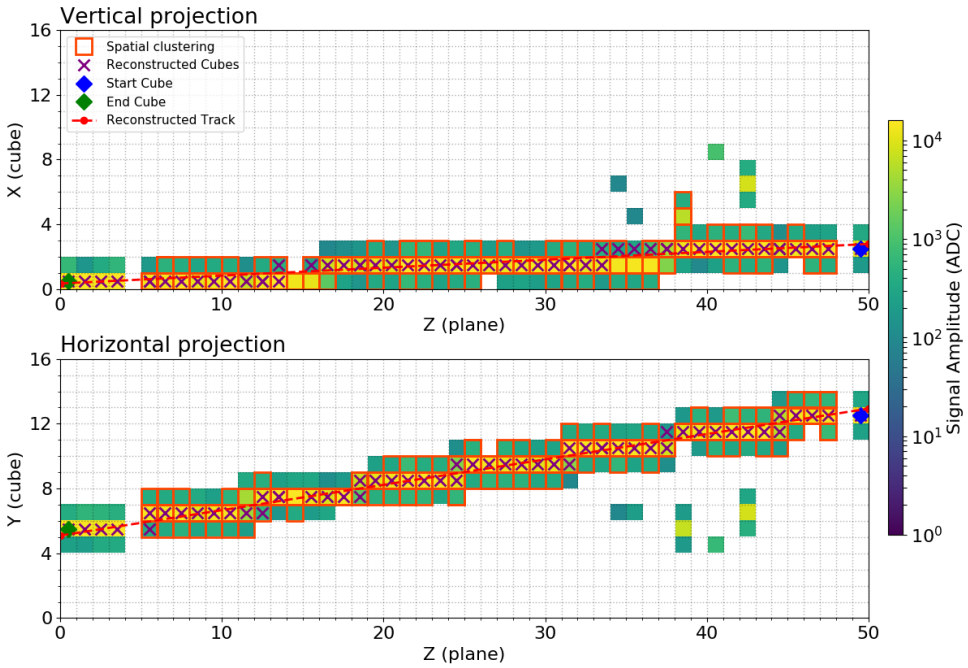


Figure 5.2: Tracking of a horizontal muon simulated in the detector. Top: projection on the plane (x, z). Bottom: projection on the plane (y, z). The cubes with a x inside are the cubes that are expected to be in the muon trajectory.

response of the detector.

5.2.2 Low background analysis

With the integration in the reconstruction software of the rough system matrix that combines calibration measurement and a global light leakage, the LPC-Clermont team derived a new analysis based on the selection of event topologies in order to increase the signal purity at its best. The analysis is based on a cluster research to classify the events based on a 0, 1 or 2 gamma category. An example of the cluster creation process for a two-gamma event can be seen on figure 5.3. A first cluster is built, composed of the annihilation cube and the $3 \times 3 \times 3$ cube envelope around it. This cluster is expected to contain the energy of the positron produced by the IBD. For an IBD event, it is expected that the two annihilation gammas are emitted back to back and create two clusters in two half space of the detector. Thus, if there is another cube in the detector outside the AC envelope, the detector is divided into two half spaces based on the annihilation cube and the second most energetic cube $C1$. This cube is expected to be induced by an energy deposit of a first annihilation gamma. The first half space contains all the cubes C_i that create an angle $(C_i, \widehat{\mathrm{AC}, C1}) < 90°$, and the second half space contains all the other cubes. Three possibilities can then happen:

– No other cube can be found outside the AC envelope. The event is tagged as a 0 gamma event.

– Cubes can be found in only one half space outside the AC envelope. The event is tagged as a 1 gamma event. Those cubes are gathered into a cluster. The barycentre of the half space cluster is

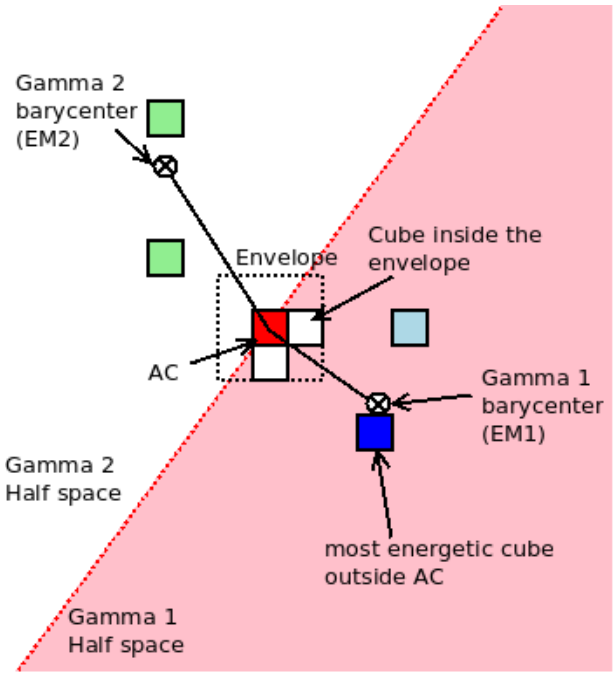


Figure 5.3: Example an event classification of a two gamma event. The AC cluster is represented in the middle, with the red cube and the dashed envelope. The first gamma cluster is represented by the blue cubes and the second gamma cluster is represented by the green cubes.

used as the position of the cluster, and the energy of the most energetic cube of the gamma cluster is used as the cluster energy.

– Cubes can be found in both half spaces outside the AC envelope. The event is tagged as a 2 gammas event. A cube cluster is created in each half space. The cluster positions and energies are computed as in the 1 gamma category.

Each of the gamma categories are then divided into three sub categories if the envelope contains zero, one or more than one cube apart from the annihilation cube. This is done to separate the cases where the positron left all its energy inside the annihilation cube, where the positron left some energy in another cube in the envelope or where one of the two gammas deposited energy inside the envelope.

Compared to the gamma tracking algorithm presented in the section 4.5.2 the cluster classification allows a different understanding of the events. It does not allow a clear tracking of the annihilation gammas, but it does not rely on any model to be built. Thanks to that, the gamma cluster position and energy distributions can be understood in a better way for the backgrounds that does not follow the expected Compton interaction model. Both classification methods require however a good reconstruction of the events and can be impacted by missing or fake cubes. Furthermore, due to the classification into 9 different categories, according to the number of cubes reconstructed around the annihilation cube, this new method necessitates a really good treatment of the light leakages to understand precisely each topology. Both analyses have thus a lot to gain from the upgrade of the reconstruction software.

In order to achieve a good signal purity, the analysis is based on the two-gamma events. A multivariate analysis is performed independently on each of the three topologies. A BDT is trained on both BiPo and atmospheric background sources in order to maximize each background rejection power. A

BiPonator and ΔT selection is used on reactor OFF data to retrieve the background enhanced samples used for the training of each BDT. Until now all the steps presented are somewhat similar to what is done in the analysis presented in the chapter 4, however the major difference between the two methods comes from the extraction of the IBD signal.

In order to extract the yield of each background source, the analysis relies on a multidimensional fit on both ΔT and ΔR simultaneously. As a reminder, ΔT is the time difference between the NS signal and the ES signal and ΔR is the distance (in cube) between the two NS and ES clusters. As it has been discussed in the chapter 4, the ΔT fit is a good tool to extract the different background yields from reactor OFF data. However, due to the similar neutron capture time for the IBD and the atmospheric background, the ΔT information alone cannot be used to determine the yield of IBD in a reactor ON sample. This is what is done on the figure 5.4 where both ΔT and ΔR are fitted simultaneously on a dataset composed of reactor OFF data, and IBD Monte-Carlo. To do so, one needs to create a ΔR p.d.f for all the type of events. The background p.d.fs are taken from reactor OFF data and the IBD p.d.f is built from IBD Monte-Carlo. The accidental p.d.f is taken from the FPNTs sample. The BiPo p.d.f is taken from the BiPo (BiPonator, ΔT) sideband on which accidental is removed. The atmospheric p.d.f is taken from the signal (BiPonator, ΔT) window on which the accidental contamination is removed.

This new antineutrino extraction analysis has been performed on the reactor ON open dataset and preliminary results show that one can measure a rate of 21.8 ± 2.1(stat) IBD events per day for a signal to background ratio slightly larger than 1 (as a comparison, the analysis presented in the precedent chapter as a signal to background ratio of 0.85 in that region). A comparison between the measurement and the prediction shows a very good agreement between the predicted signal and the measurement on figure 5.5.

This new analysis with a high purity selection and a very low efficiency raises the question of the best working point for the analysis. The oscillation analysis performed on the result of both analyses at different (S, S:B) configuration will tell which configuration gives the most exclusion power between the low efficiency analysis with a high purity of this section or an analysis with more efficiency but lower purity. During my Ph.D, I was briefly involved in the use of this method to one gamma topologies. Those have a larger background contamination but could help to improve the final statistics. However, this work was left aside due to time constraint at the end of the thesis but will be continued for the final results of this new analysis.

5.3 SoLid Phase II

Another way to improve the exclusion power of the SoLid experiment is through the use of its phase II data. During the Summer 2020, the SoLid detector was shut down for an upgrade. This upgrade consisted in the replacement of all the MPPCs (type S12572-050C) by a new generation of MPPCs (type S14160-3050HS). Those MPPCs have a higher photon detection efficiency for a lower operation voltage and lower internal cross talk. At similar working point, one can thus expect an increase of the light yield of the detection cells.

A first test bench has been developed at Imperial College in order to measure the expected performances of the new sensors. The test bench was composed of a PVT cube coupled with two WLS fibres. A MPPC was then coupled to one end of each fibre and connected to an analogue electronic board. A ^{90}Sr calibration source was then placed at the top of the cube. The setup was put inside a small dark

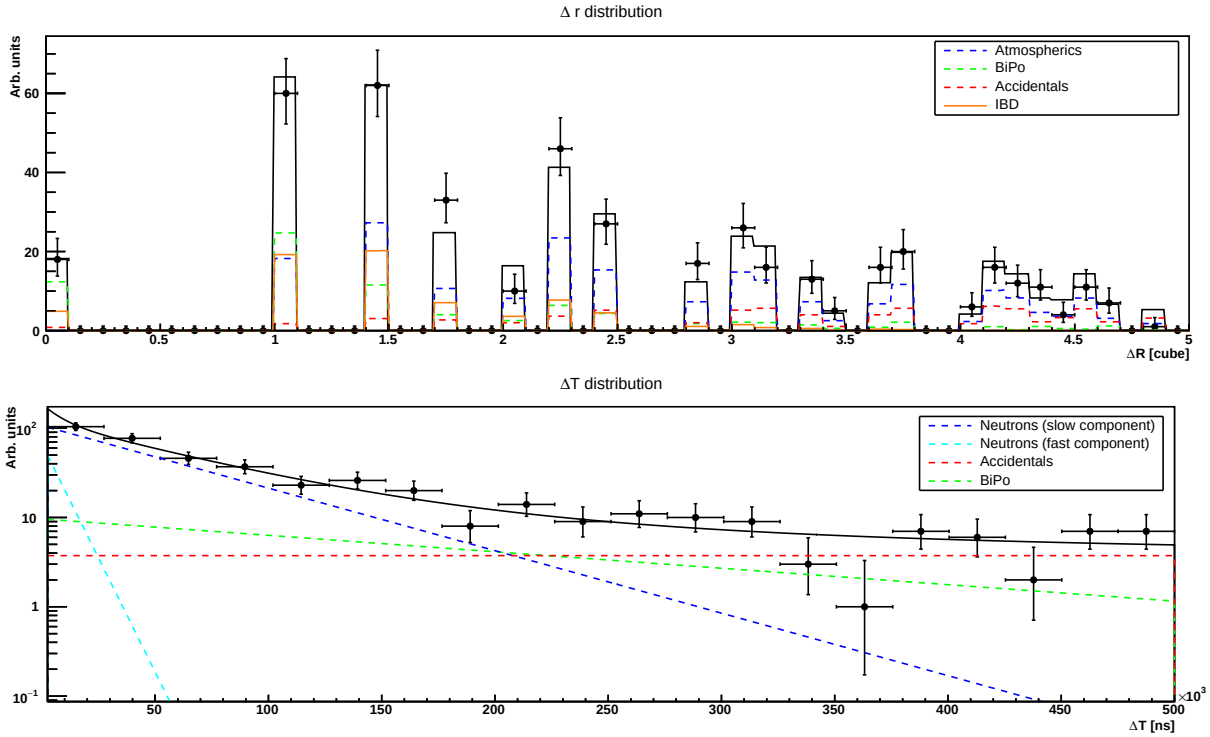


Figure 5.4: Example of a combined ΔT, ΔR fit on reactor OFF data with IBD Monte-Carlo neutrinos added to the sample.

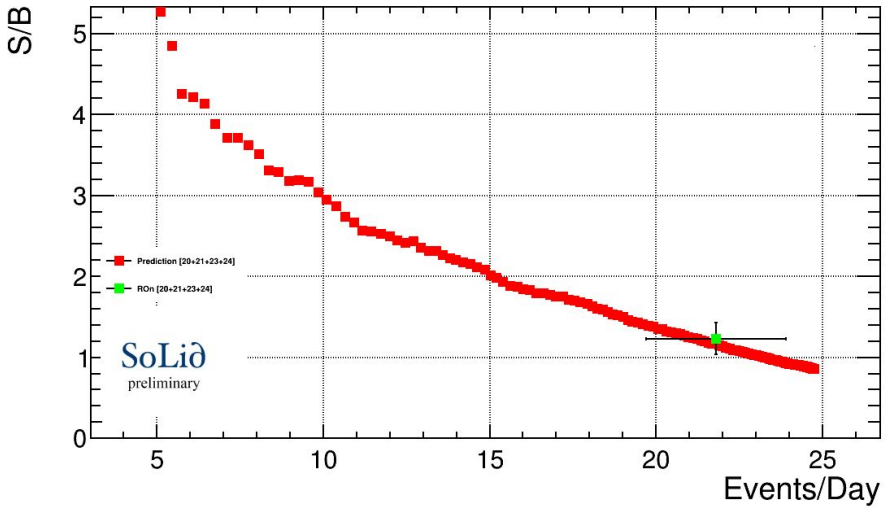


Figure 5.5: S/B ratio versus the antineutrino rate per day. In red are represented the predictions while in green is represented the measurement on the open dataset.

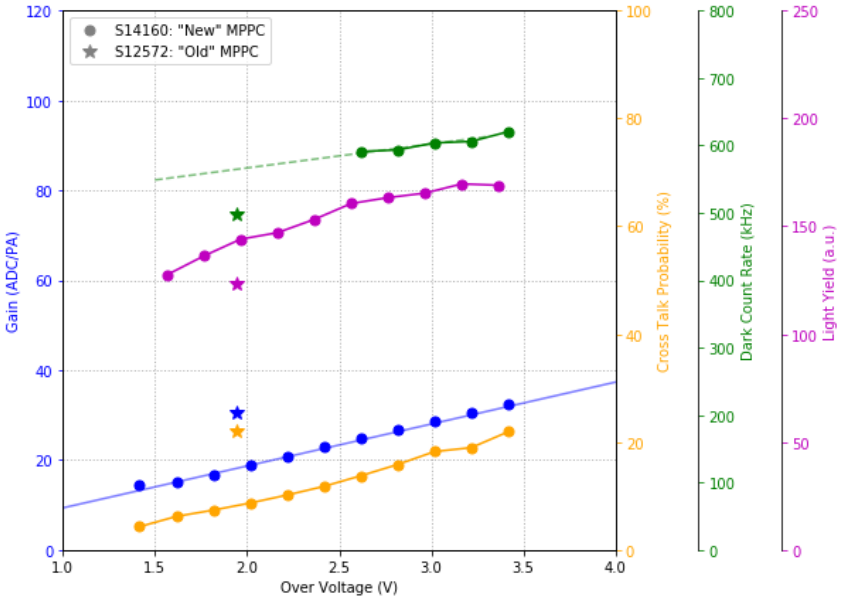


Figure 5.6: Cross talk probability, dark count rate, light yield and gain versus the over voltage for the Phase I ("Old") and Phase II ("New") MPPCs)

room that also acted as a Faraday Cage. The result of this first MPPC characterisation with a comparison with the Phase I MPPC can be seen on figure 5.6. One can see that for the same over voltage, the cross talk of the detector drops significantly with a small increase of both the light yield and the dark count rate. However, the aim of the upgrade is to use higher over voltage values in order to get a higher light yield. This is allowed by the lowered cross talk achieved with the new generation of MPPC. With an over voltage of about 3.2 V, in order to obtain the same gain as the Phase I, the light yield increases of a factor about 40 % with an increased dark count rate around of around 25 % higher. This increase in the dark count rate is not an issue as it is then greatly reduced with the X-Y coincidences asked for the trigger, the cooling of the detector and the trigger threshold. This higher light yield value would lead to a better energy resolution, and a detection of lower energy deposits in the detector. The addition of the lower energy deposit could be an important improvement for the gamma tracking algorithm, with a larger number of deposits than can be taken into account in the tracking algorithm.

The first energy calibration has been performed in November 2020 with the ^{22}Na calibration source and with an over voltage at 3.2 V. It consisted in the different steps described in chapter 3: gain and pedestal equalisation, and the measurement of the channel attenuation lengths, the channel couplings and the cube light yields. The distribution of the measured cube light yields is represented the figure 5.7. The average light yield of the detector is 129 PA/MeV without cross talk subtraction, compared to the value of 90.9 measured in February 2020, this verifies the increase of light yield of about 40 %. This result confirms that the Phase II data will be of great help to improve the antineutrino oscillation analysis. Data in the Phase II configuration are being taken since October 2020 for an undefined number of days.

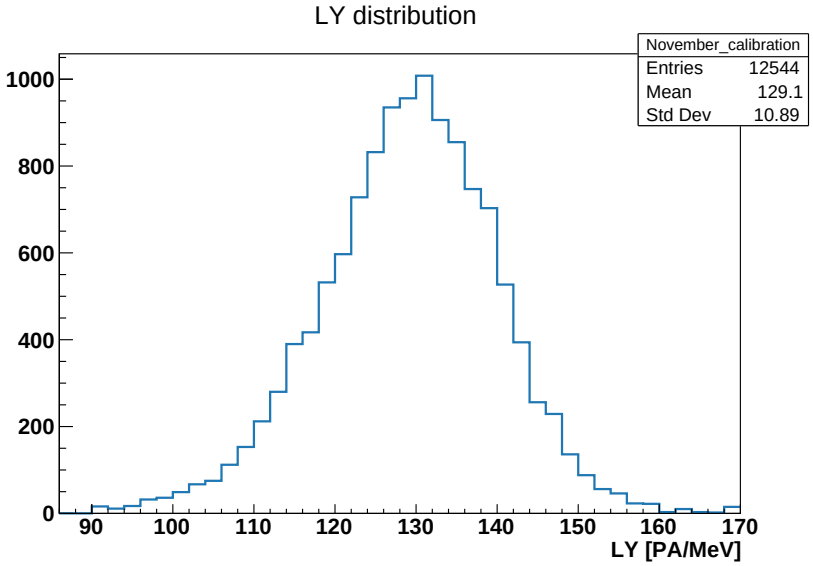


Figure 5.7: Distribution of the measured cube light yields for the Phase II data without cross talk subtraction.

Conclusion

Neutrino physics is a field of research that is still opened for major measurements and discoveries to unveil the true nature of neutrinos. The neutrino discoveries could be seen as a chain of anomalies that lead to the search for sterile neutrinos. It started with the missing energy of the beta decay that led to the discovery of neutrinos. It then continued with the missing neutrinos in experiments that looked for neutrinos from the Sun that led to the discovery of neutrino oscillation. Finally, anomalies were found in oscillation experiments with the *gallium anomaly*, *accelerator anomaly* and *reactor antineutrino anomaly*. If this time the solution has not yet been found, the oscillation toward a light sterile neutrino state could account for those observations.

The SoLid experiment aims to detect antineutrinos at a few meters from the BR2 reactor core. The experiment uses a hybrid scintillating technology with a high segmentation. It is composed of 12 800 detection cells read out by 3 200 MPPCs. The challenging calibration of the detector has been my first work during the thesis. I started on the equalization of the MPPC response with a measurement of the gain and the pedestal variations over time. Those variations were then corrected in the reconstruction software in order to have a uniform response of the photo sensors. The measurement of the gains show a variation of less than 3 % of the gains during the two years of data taking of the SoLid Phase I. For a given channel, less than 1 % pedestal variation was measured throughout the years. The next step of the calibration was the measurement of the fibre attenuation lengths and a quantification of the goodness of their optical coupling with the MPPCs. This was performed with a ^{22}Na gamma source and with a combined method between a sequential method that I developed, and a global fit developed by the colleagues of Subatech Nantes. With this combined method, the attenuation length measurement on calibration data showed a 1.7 % bias and the coupling showed nearly no bias when compared to simulations. The method showed a good stability overtime from the various calibration campaigns, with variations of the average attenuation length and couplings bellow 1 % during the 2 years of data taking. The final step of the calibration was the measurement of the cube light yields. Due to the size of the detection cells in the detector, only Compton edges were available to calibrate the detector. A method has been developed at IJCLab that uses a numerically convoluted GEANT4 spectrum in order to derive the light yield and the energy resolution of each cube. This method has been compared with an analytical fit developed at Subatech Nantes with a difference bellow 1 % between the two methods, ensuring a robust measurement of the cube light yields. This light yield was measured through the 2 years of SoLid Phase I to ensure a good correction of the PVT ageing effects. The use of several calibration sources allowed to test the linearity of the detector response in the $[0.5 - 4.2]$ MeV region and to measure the full energy resolution for a subset of cubes. Those calibration measurements allowed to tune precisely the detector

simulation. Data to Monte-Carlo comparison on ^{22}Na calibration data showed an agreement below 5 % in the Compton edge.

The second work presented in this thesis is the antineutrino analysis. In this work, I used the gamma tracking algorithm developed at Subatech Nantes to derive IBD categories. Those categories coupled with a boosted decision tree based analysis allowed reduction of the background of 95 % compared to simple rectangular selection cuts on low-level variables for a total number of 243 background events per day. However, this came with a great reduction of the IBD efficiency that was reduced of 55 % compared to simple cuts. The final efficiency of antineutrino detection compared to the number of antineutrino interactions in the detector is 6 % for a total number of 65 IBD events per day. The work conducted in this thesis also consisted in a review of the background subtraction procedure from the reactor ON data. The first subtraction was performed on an *open dataset* of 20 days of reactor ON and showed results compatible with the predictions for the antineutrino yield and with the shape of the energy and distance distributions. I have then performed a study to test the stability of the subtraction method. This study showed variations over time of the PSD variable used to discriminate the BiPo background. These variations induce errors in the evaluation of the BiPo background, which at the level of background faced in SoLid can induce large systematic uncertainties in the antineutrino yield and energy and travelled distance distributions.

The first preliminary sensitivity contour that was obtained with the analysis performances showed that the SoLid experiment could not give a competitive exclusion contour with its current results. The analysis of the whole reactor ON data should be done in the near future in order to provide a full antineutrino spectrum and a first exclusion contour. Some improvement could be achieved however with the upgrade of the reconstruction software combined with a very pure analysis. An upgrade of the detector has been performed during the summer 2020 that consisted in the change of all the 3 200 MPPCs of the detector for a newer generation. Those new MPPCs have a lower cross-talk and can be operated at a higher voltage, inducing a better light collection. The SoLid Phase II data with MPPCs that have a light collection 40 % higher than the phase I MPPCs could also help the analysis with a better detection of the IBD annihilation gammas that could help a better rejection of the background.

Bibliography

[1] E FERMI. "Tentativo di una teoria dell'emissione dei raggi «Beta». ricerca scientifica, v. 4". In: (1933) (cit. on p. 3).

[2] Hans Bethe and Rudolph Peierls. "The neutrino". In: *Nature* 133.3366 (1934), pp. 689–690 (cit. on p. 3).

[3] Frederick Reines et al. "Detection of the free antineutrino". In: *Physical Review* 117.1 (1960), p. 159 (cit. on p. 4).

[4] Gaillard Danby et al. "Observation of high-energy neutrino reactions and the existence of two kinds of neutrinos". In: *Physical Review Letters* 9.1 (1962), p. 36 (cit. on p. 4).

[5] Martin L Perl et al. "Evidence for anomalous lepton production in e+- e- annihilation". In: *Physical Review Letters* 35.22 (1975), p. 1489 (cit. on p. 5).

[6] The SLD Electroweak et al. "Precision electroweak measurements on the Z resonance". In: *Physics Reports* 427.5-6 (2006), pp. 257–454 (cit. on p. 5).

[7] K Kodama et al. "Observation of tau neutrino interactions". In: *Physics Letters B* 504.3 (2001), pp. 218–224 (cit. on p. 5).

[8] Cloé Girard-Carillo. "Study of ^{208}Tl background rejection influence on the $0\nu\beta\beta$ decay sensitivity, characterisation of SuperNEMO demonstrator calorimeter timing performance". PhD thesis. Université Paris-Saclay, 2020 (cit. on p. 6).

[9] John N Bahcall et al. "Standard solar models and the uncertainties in predicted capture rates of solar neutrinos". In: *Reviews of Modern Physics* 54.3 (1982), p. 767 (cit. on p. 7).

[10] Raymond Davis Jr, Don S Harmer, and Kenneth C Hoffman. "Search for neutrinos from the sun". In: *Physical Review Letters* 20.21 (1968), p. 1205 (cit. on p. 7).

[11] Johnrid N Abdurashitov et al. "Solar neutrino flux measurements by the Soviet-American gallium experiment (SAGE) for half the 22-year solar cycle". In: *Journal of Experimental and Theoretical Physics* 95.2 (2002), pp. 181–193 (cit. on p. 7).

[12] Florian Kaether et al. "Reanalysis of the GALLEX solar neutrino flux and source experiments". In: *Physics Letters B* 685.1 (2010), pp. 47–54 (cit. on p. 7).

[13] S Fukuda et al. "Solar B 8 and hep Neutrino Measurements from 1258 Days of Super-Kamiokande Data". In: *Physical Review Letters* 86.25 (2001), p. 5651 (cit. on p. 7).

[14] Y Fukuda et al. "Study of the atmospheric neutrino flux in the multi-GeV energy range". In: *Physics Letters B* 436.1-2 (1998), pp. 33–41 (cit. on p. 7).

[15] J Boger et al. "The Sudbury neutrino observatory". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 449.1-2 (2000), pp. 172–207 (cit. on p. 7).

[16] B Aharmim et al. "Electron energy spectra, fluxes, and day-night asymmetries of 8 B solar neutrinos from measurements with NaCl dissolved in the heavy-water detector at the Sudbury Neutrino Observatory". In: *Physical Review C* 72.5 (2005), p. 055502 (cit. on p. 9).

[17] Bruno Pontecorvo. "Mesonium and antimesonium". In: *Zhur. Eksptl'. i Teoret. Fiz.* 33 (1957) (cit. on p. 8).

[18] Ziro Maki, Masami Nakagawa, and Shoichi Sakata. "Remarks on the unified model of elementary particles". In: *Progress of Theoretical Physics* 28.5 (1962), pp. 870–880 (cit. on p. 8).

[19] Sebastian Böser et al. "Status of light sterile neutrino searches". In: *Progress in particle and nuclear physics* 111 (2020), p. 103736 (cit. on p. 11).

[20] SP Mikheyev and A Yu Smirnov. "Resonant amplification of ν oscillations in matter and solar-neutrino spectroscopy". In: *Il Nuovo Cimento C* 9.1 (1986), pp. 17–26 (cit. on p. 11).

[21] Azusa Gando et al. "Reactor on-off antineutrino measurement with KamLAND". In: *Physical Review D* 88.3 (2013), p. 033001 (cit. on p. 12).

[22] Michel Ageron et al. "ANTARES: the first undersea neutrino telescope". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 656.1 (2011), pp. 11–38 (cit. on p. 12).

[23] Mark G Aartsen et al. "The IceCube Neutrino Observatory: instrumentation and online systems". In: *Journal of Instrumentation* 12.03 (2017), P03012 (cit. on p. 12).

[24] Holger Meyer, Mathew Muether, and Nickolas Solomey. "Measurement of the neutrino mixing angle θ_{23} in NOvA". In: (2017) (cit. on p. 12).

[25] K Abe et al. "Indication of electron neutrino appearance from an accelerator-produced off-axis muon neutrino beam". In: *Physical Review Letters* 107.4 (2011), p. 041801 (cit. on pp. 12, 13).

[26] Y Abe et al. "Indication of reactor electron antineutrinos disappearance in the Double Chooz experiment". In: *Physical Review Letters* 108 (2012), p. 131801 (cit. on p. 13).

[27] FP An et al. "Observation of electron-antineutrino disappearance at Daya Bay". In: *Physical Review Letters* 108.17 (2012), p. 171803 (cit. on p. 13).

[28] Jung Keun Ahn et al. "Observation of reactor electron antineutrinos disappearance in the RENO experiment". In: *Physical Review Letters* 108.19 (2012), p. 191802 (cit. on p. 13).

[29] André de Gouvêa et al. "Neutrinos". In: *arXiv preprint arXiv:1310.4340* (2013) (cit. on p. 13).

[30] Ivan Esteban et al. "The fate of hints: updated global analysis of three-flavor neutrino oscillations". In: *Journal of High Energy Physics* 2020.9 (2020), pp. 1–22 (cit. on pp. 14, 15).

[31] Particle Data Group et al. "Review of particle physics". In: *Progress of Theoretical and Experimental Physics* 2020.8 (2020), p. 083C01 (cit. on p. 14).

[32] XB Ma et al. "Improved calculation of the energy release in neutron-induced fission". In: *Physical Review C* 88.1 (2013), p. 014605 (cit. on p. 16).

[33] K Schreckenbach et al. “Absolute measurement of the beta spectrum from 235U fission as a basis for reactor antineutrino experiments”. In: *Physics Letters B* 99.3 (1981), pp. 251–256 (cit. on pp. 16, 17).

[34] W Mampe et al. “The double focusing iron-core electron-spectrometer “BILL” for high resolution (n, e-) measurements at the high flux reactor in Grenoble”. In: *Nuclear Instruments and Methods* 154.1 (1978), pp. 127–149 (cit. on p. 17).

[35] F Von Feilitzsch, AA Hahn, and K Schreckenbach. “Experimental beta-spectra from 239Pu and 235U thermal neutron fission products and their correlated antineutrino spectra”. In: *Physics Letters B* 118.1-3 (1982), pp. 162–166 (cit. on p. 17).

[36] K Schreckenbach et al. “Determination of the antineutrino spectrum from 235U thermal neutron fission products up to 9.5 MeV”. In: *Physics Letters B* 160.4-5 (1985), pp. 325–330 (cit. on pp. 17, 18).

[37] AA Hahn et al. “Antineutrino spectra from 241Pu and 239Pu thermal neutron fission products”. In: *Physics Letters B* 218.3 (1989), pp. 365–368 (cit. on pp. 17, 18).

[38] Th A Mueller et al. “Improved predictions of reactor antineutrino spectra”. In: *Physical Review C* 83.5 (2011), p. 054615 (cit. on pp. 17–19, 21).

[39] Patrick Huber. “Determination of antineutrino spectra from nuclear reactors”. In: *Physical Review C* 84.2 (2011), p. 024617 (cit. on pp. 17, 18).

[40] Nils Haag et al. “Experimental determination of the antineutrino spectrum of the fission Products of U 238”. In: *Physical review letters* 112.12 (2014), p. 122501 (cit. on p. 18).

[41] RW King and JF Perkins. “Inverse beta decay and the two-component neutrino”. In: *Physical Review* 112.3 (1958), p. 963 (cit. on p. 19).

[42] Magali Estienne et al. “Updated summation model: an improved agreement with the daya bay antineutrino fluxes”. In: *Physical review letters* 123.2 (2019), p. 022502 (cit. on pp. 19, 20).

[43] FP An et al. “Evolution of the reactor antineutrino flux and spectrum at Daya Bay”. In: *Physical review letters* 118.25 (2017), p. 251801 (cit. on p. 20).

[44] Hervé de Kerret et al. “Double Chooz θ13 measurement via total neutron capture detection”. In: *Nature Physics* 16.5 (2020) (cit. on p. 20).

[45] H Almazán et al. “Accurate Measurement of the Electron Antineutrino Yield of U 235 Fissions from the STEREO Experiment with 119 Days of Reactor-On Data”. In: *Physical Review Letters* 125.20 (2020), p. 201801 (cit. on p. 20).

[46] AC Hayes et al. “Analysis of the Daya Bay reactor antineutrino flux changes with fuel burnup”. In: *Physical review letters* 120.2 (2018), p. 022503 (cit. on p. 20).

[47] Feng Peng An et al. “Measurement of the reactor antineutrino flux and spectrum at Daya Bay”. In: *Physical review letters* 116.6 (2016), p. 061801 (cit. on p. 20).

[48] Z Atif et al. “Measurement of Reactor Antineutrino Flux and Spectrum at RENO”. In: *arXiv preprint arXiv:2010.14989* (2020) (cit. on p. 20).

[49] G Mention et al. “Reactor antineutrino shoulder explained by energy scale nonlinearities?” In: *Physics Letters B* 773 (2017), pp. 307–312 (cit. on p. 20).

[50] D Adey et al. “Extraction of the U 235 and Pu 239 antineutrino spectra at Daya Bay”. In: *Physical review letters* 123.11 (2019), p. 111801 (cit. on pp. 20, 21).

[51] Manoa Andriamirado et al. “Improved short-baseline neutrino oscillation search and energy spectrum measurement with the PROSPECT experiment at HFIR”. In: *Physical Review D* 103.3 (2021), p. 032001 (cit. on pp. 21, 27, 28).

[52] Helena Almazán Molina et al. “First antineutrino energy spectrum from 235U fissions with the STEREO detector at ILL”. In: (2020) (cit. on p. 21).

[53] G Mention et al. “Reactor antineutrino anomaly”. In: *Physical Review D* 83.7 (2011), p. 073006 (cit. on pp. 21, 25).

[54] AC Hayes et al. “Systematic uncertainties in the analysis of the reactor neutrino anomaly”. In: *Physical Review Letters* 112.20 (2014), p. 202501 (cit. on p. 22).

[55] Dong-Liang Fang and B Alex Brown. “Effect of first-forbidden decays on the shape of neutrino spectra”. In: *Physical Review C* 91.2 (2015), p. 025503 (cit. on p. 22).

[56] A Gando et al. “CeLAND: search for a 4th light neutrino state with a 3 PBq 144Ce-144Pr electron antineutrino generator in KamLAND”. In: *arXiv preprint arXiv:1312.0896* (2013) (cit. on p. 23).

[57] A LSND. “Evidence for neutrino oscillations from the observation of anti-neutrino (electron) appearance in a anti-neutrino (muon) beam”. In: *Phys. Rev. D* 64 (2001), p. 112007 (cit. on p. 23).

[58] AA Aguilar-Arevalo et al. “Significant excess of electronlike events in the MiniBooNE short-baseline neutrino experiment”. In: *Physical review letters* 121.22 (2018), p. 221801 (cit. on p. 23).

[59] Florian Kaether et al. “Reanalysis of the GALLEX solar neutrino flux and source experiments”. In: *Physics Letters B* 685.1 (2010), pp. 47–54 (cit. on p. 24).

[60] JN Abdurashitov et al. “Measurement of the solar neutrino capture rate with gallium metal. III. Results for the 2002–2007 data-taking period”. In: *Physical Review C* 80.1 (2009), p. 015807 (cit. on p. 24).

[61] Carlo Giunti and Marco Laveder. “Statistical significance of the gallium anomaly”. In: *Physical Review C* 83.6 (2011), p. 065504 (cit. on p. 24).

[62] Joel Kostensalo et al. “The gallium anomaly revisited”. In: *Physics Letters B* 795 (2019), pp. 542–547 (cit. on p. 24).

[63] VV Barinov et al. “Results from the Baksan Experiment on Sterile Transitions (BEST)”. In: *arXiv preprint arXiv:2109.11482* (2021) (cit. on p. 24).

[64] Stefano Gariazzo et al. “Light sterile neutrinos”. In: *Journal of Physics G: Nuclear and Particle Physics* 43.3 (2016), p. 033001 (cit. on p. 24).

[65] M Aker et al. “Bound on 3+ 1 active-sterile neutrino mixing from the first four-week science run of KATRIN”. In: *Physical review letters* 126.9 (2021), p. 091803 (cit. on p. 25).

[66] KM Heeger et al. “Experimental parameters for a reactor antineutrino experiment at very short baselines”. In: *Physical Review D* 87.7 (2013), p. 073008 (cit. on p. 26).

[67] N Allemandou et al. “The STEREO experiment”. In: *Journal of Instrumentation* 13.07 (2018), P07009 (cit. on p. 26).

[68] Helena Almazán et al. "Improved sterile neutrino constraints from the STEREO experiment with 179 days of reactor-on data". In: *Physical Review D* 102.5 (2020), p. 052002 (cit. on pp. 26, 27).

[69] PROSPECT Collaboration et al. "First search for short-baseline neutrino oscillations at HFIR with PROSPECT". In: *Physical review letters* 121.25 (2018), p. 251802 (cit. on p. 27).

[70] AP Serebrov et al. "Search for sterile neutrinos with the Neutrino-4 experiment and measurement results". In: *Physical Review D* 104.3 (2021), p. 032003 (cit. on p. 27).

[71] C Giunti et al. "Neutrino-4 anomaly: Oscillations or fluctuations?" In: *Physics Letters B* 816 (2021), p. 136214 (cit. on p. 27).

[72] I Alekseev et al. "DANSS: Detector of the reactor AntiNeutrino based on Solid Scintillator". In: *Journal of Instrumentation* 11.11 (2016), P11011 (cit. on p. 28).

[73] Mikhail Danilov. "New results from the DANSS experiment". In: *arXiv preprint arXiv:2012.10255* (2020) (cit. on p. 28).

[74] YJ Ko et al. "Sterile neutrino search at the NEOS experiment". In: *Physical review letters* 118.12 (2017), p. 121802 (cit. on p. 28).

[75] Z Atif et al. "Search for sterile neutrino oscillation using RENO and NEOS data". In: *arXiv preprint arXiv:2011.00896* (2020) (cit. on p. 28).

[76] Silva Kalcheva et al. "Reactor Core Simulations for Determination of the Antineutrino Spectrum for the SoLid Experiment at BR2 Reactor". In: *M&C 2017*. Jeju, South Korea. URL: https://hal-imt-atlantique.archives-ouvertes.fr/hal-02411158 (cit. on p. 31).

[77] Olivier Méplan et al. "MURE: MCNP Utility for Reactor Evolution-Description of the methods, first applications and results". In: *ENC 2005-European Nuclear Conference. Nuclear Power for the XXIst Century: From basic research to high-tech industry*. European Nuclear Society. 2005, pp. 1–7 (cit. on p. 31).

[78] Y Abreu et al. "Optimisation of the scintillation light collection and uniformity for the SoLid experiment". In: *Journal of Instrumentation* 13.09 (2018), P09005 (cit. on pp. 34, 68).

[79] *Hamamatsu MPPC: online datasheet*. https://seltokphotonics.com/upload/iblock/118/1186141f87bc6da48efe2c465195e38d.pdf (cit. on pp. 35–37).

[80] Y Abreu et al. "Development of a quality assurance process for the SoLid experiment". In: *Journal of Instrumentation* 14.02 (2019), P02014 (cit. on pp. 35, 37).

[81] *Saint Gobain Fibres: online datasheet*. https://www.crystals.saint-gobain.com/sites/imdf.crystals.com/files/documents/fiber-product-sheet.pdf (cit. on p. 36).

[82] https://eljentechnology.com/images/products/data_sheets/EJ-200_EJ-204_EJ-208_EJ-212.pdf (cit. on p. 36).

[83] *SCINTACOR neutron screens: online datasheet*. https://scintacor.com/wp-content/uploads/2015/09/Datasheet-Neutron-Screens-High-Res.pdf (cit. on p. 36).

[84] Valentin Pestel. "Détection de neutrinos auprès du réacteur BR2: analyse des premières données de l'expérience SoLid". PhD thesis. Normandie, 2019 (cit. on pp. 40, 42, 54, 113, 129, 130, 164).

[85] Y Abreu et al. "SoLid: A short baseline reactor neutrino experiment". In: *Journal of Instrumentation* 16.02 (2021), P02025 (cit. on pp. 43, 44).

[86] Y Abreu et al. "Commissioning and operation of the readout system for the SoLid neutrino detector". In: *Journal of Instrumentation* 14.11 (2019), P11003 (cit. on p. 42).

[87] P Vogel and John F Beacom. "Angular distribution of neutron inverse beta decay, $\overline{\nu}_e + p \rightarrow e^+ + n$". In: *Physical Review D* 60.5 (1999), p. 053003 (cit. on p. 44).

[88] Y Abe et al. "Reactor ν e disappearance in the Double Chooz experiment". In: *Physical Review D* 86.5 (2012), p. 052008 (cit. on p. 44).

[89] Sea Agostinelli et al. "GEANT4—a simulation toolkit". In: *Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303 (cit. on p. 47).

[90] I. Piñera-Hernández. "BR2 building geometry & model for SoLid Geant4 simulations". In: *SoLid technical note* () (cit. on p. 48).

[91] Joachim Kopp et al. "Sterile neutrino oscillations: the global picture". In: *Journal of High Energy Physics* 2013.5 (2013), pp. 1–52 (cit. on p. 49).

[92] Mengyun Guan et al. "A parametrization of the cosmic-ray muon flux at sea-level". In: *arXiv preprint arXiv:1509.06176* (2015) (cit. on p. 48).

[93] MS Gordon et al. "Measurement of the flux and energy spectrum of cosmic-ray induced neutrons on the ground". In: *IEEE Transactions on Nuclear Science* 51.6 (2004), pp. 3427–3434 (cit. on p. 48).

[94] JB Birks and FA Black. "Deterioration of anthracene under a particle bombardment". In: *Proc. of the Phys. Soc.(Lond.)* Vol. 64. 874. 1951 (cit. on p. 49).

[95] Arthur P Dempster, Nan M Laird, and Donald B Rubin. "Maximum likelihood from incomplete data via the EM algorithm". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 39.1 (1977), pp. 1–22 (cit. on p. 55).

[96] A Simón et al. "Event reconstruction in NEXT using the ML-EM algorithm". In: *Nuclear and particle physics proceedings* 273 (2016), pp. 2624–2626 (cit. on p. 55).

[97] *LNHB atomic data*. URL: http://www.lnhb.fr/donnees-nucleaires/donnees-nucleaires-tableau/ (cit. on pp. 61, 62).

[98] Ali Asghar Mowlavi and Rahim Koohi-Fayegh. "Determination of 4.438 MeV γ-ray to neutron emission ratio from a 241Am–9Be neutron source". In: *Applied Radiation and Isotopes* 60.6 (2004), pp. 959–962 (cit. on p. 61).

[99] Oskar Klein and Yoshio Nishina. "The scattering of light by free electrons according to Dirac's new relativistic dynamics". In: *Nature* 122.3072 (1928), pp. 398–399 (cit. on p. 76).

[100] David Henaff. "Recherche d'oscillations à courte distance auprès du réacteur BR2 et mesure du spectre d'antineutrinos issus de l'Uranium 235 avec l'expérience SoLid". Theses. Ecole nationale supérieure Mines-Télécom Atlantique Bretagne Pays de la Loire, 2021 (cit. on pp. 76, 99, 100, 104, 130, 131).

[101] Yamiel Abreu et al. "A novel segmented-scintillator antineutrino detector". In: *Journal of Instrumentation* 12.04 (2017), P04024 (cit. on p. 112).

[102] Mario A Acero et al. "Seasonal variation of multiple-muon cosmic ray air showers observed in the NOvA detector on the surface". In: *Physical Review D* 104.1 (2021), p. 012014 (cit. on p. 119).

[103] D Galbinski et al. "A Note on Quality Assurance of Phase-I Data and Stability Studies Conducted using BiPo Background". In: *SoLid technical note* () (cit. on pp. 124, 125).

[104] Y Abreu et al. "Performance of a full scale prototype detector at the BR2 reactor for the SoLid experiment". In: *Journal of Instrumentation* 13.05 (2018), P05005 (cit. on p. 130).

[105] Oskar Klein and Yoshio Nishina. "Über die Streuung von Strahlung durch freie Elektronen nach der neuen relativistischen Quantendynamik von Dirac". In: *Zeitschrift für Physik* 52.11 (1929), pp. 853–868 (cit. on p. 132).

[106] *ELJEN general purpose plastic scintillator: online datasheet.* URL: https://eljentechnology.com/images/products/data_sheets/EJ-200_EJ-204_EJ-208_EJ-212.pdf (visited on 03/17/2021) (cit. on p. 132).

[107] David A Brown et al. "ENDF/B-VIII. 0: the 8th major release of the nuclear reaction data library with CIELO-project cross sections, new standards and thermal scattering data". In: *Nuclear Data Sheets* 148 (2018), pp. 1–142 (cit. on p. 138).

[108] Byron P Roe et al. "Boosted decision trees as an alternative to artificial neural networks for particle identification". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 543.2-3 (2005), pp. 577–584 (cit. on p. 144).

[109] Cheng Li. "A gentle introduction to gradient boosting". In: (2016) (cit. on p. 144).

[110] Andreas Hocker et al. *TMVA-toolkit for multivariate data analysis with ROOT: users guide*. Tech. rep. 2007 (cit. on p. 144).

[111] Ianthe Michiels. "Development of the oscillation analysis framework for the SoLid experiment". PhD thesis. Gent U., 2020 (cit. on p. 163).

[112] Gary J Feldman and Robert D Cousins. "Unified approach to the classical statistical analysis of small signals". In: *Physical Review D* 57.7 (1998), p. 3873 (cit. on p. 165).

[113] L. Lyons. "Raster scan or 2-D approach?" In: *arXiv: High Energy Physics - Experiment* (2014) (cit. on p. 165).

Résumé

Au début du XX^{eme} siècle, le spectre en énergie des électrons émis par désintégration β était l'un des problèmes les plus troublants dans la communauté de la physique des particules. De la même façon que pour la désintégration α ou γ, l'électron émis lors d'une désintégration β était attendu comme monoénergétique avec une énergie bien définie, transportant toute l'énergie de la désintégration dans son énergie cinétique. Néanmoins, contrairement aux prédictions, les spectres en énergie des électrons apparaissaient comme continus. C'est en 1930 que W. Pauli proposa une explication, en introduisant une nouvelle particule dans la désintégration β qui partagerait avec l'électron une partie de l'énergie de la désintégration : le neutrino. En 1956 F. Reines et C. Cowan firent la première détection antineutrinos électroniques au réacteur nucléaire de Savanah River. En 1962 puis 2000, furent découvert les deux autres saveurs de neutrinos : les neutrinos muon et tau, pour un paradigme de neutrino à trois saveurs leptoniques. Après la découverte de différentes saveurs de neutrinos, le phénomène d'oscillation de neutrino a été mis en évidence par les expériences SNO et SuperKamiokande dans la fin du XX^{eme} siècle, ce qui leur a valu le prix Nobel de physique en 2015. Un neutrino émis dans un certain état de saveur, après propagation sur une certaine distance, a une probabilité non nulle d'être détecté dans un autre état de saveur. Si les paramètres d'oscillation sont aujourd'hui connus avec précisions pour une majorité d'entre eux, il reste néanmoins des paramètres qui nécessitent d'être déterminés avec plus de précisions comme la hiérarchie de masse des neutrinos, la phase de violation CP ou encore certaines anomalies expérimentales qui ne trouvent aujourd'hui aucune réponse satisfaisante.

Anomalies expérimentales

Suite à la réévaluation en 2011, par Muller et al., des spectres en énergie des neutrinos de réacteur, un déficit de 6 % de neutrinos par rapport aux prédictions est observé dans les expériences de détection de neutrino à courte distance des réacteurs. Cette anomalie est appelée *l'anomalie d'antineutrino de réacteur* (RAA). De plus, les expériences d'observation de neutrino solaire SAGE et GALLEX ont mesuré un déficit de 16 % de neutrino dans leurs mesures de calibration en utilisant des sources de neutrino monoénergétiques. Cette anomalie est communément appelée *l'anomalie gallium*. Enfin les expériences LSND et MiniBooNE ont toutes deux observé un excès d'antineutrinos électroniques dans leur expérience d'oscillation de $\overline{\nu}_\mu$ vers $\overline{\nu}_e$. Ces anomalies peuvent toutes être expliquées par une oscillation vers un nouvel état de neutrino léger. Du fait des mesures de la largeur de désintégration du boson Z0 effectuées au LEP, on sait aujourd'hui qu'il existe uniquement trois neutrinos actifs et légers. Ce quatrième neutrino serait donc stérile, dans le sens où il ne pourrait pas interagir par interaction faible. Les premières analyses combinant *l'anomalie gallium* et *l'anomalie d'antineutrino de réacteur* obtiennent un

best-fit pour une différence de masses au carré $\Delta m^2 = 1.25\ \mathrm{eV}^2$ et un angle de mélange $\sin^2 2\theta = 0.34$. Parmi les dispositifs expérimentaux pouvant étudier l'hypothèse du neutrino stérile, les expériences de détection d'antineutrinos à courte distance de réacteurs nucléaires sont parmi les plus efficaces. Avec une mesure du taux d'antineutrinos électroniques à plusieurs distances, ces expériences sont capables de s'affranchir des modèles nucléaires sur lesquels sont calculés les prédictions de flux d'antineutrinos. Depuis quelques années, des détecteurs de neutrinos ont été installés à courte distance de réacteurs nucléaires dans différents pays, en utilisant différentes technologies de détection (scintillateur liquide ou solide), avec des détecteurs placés à des distances différentes par rapport au centre des réacteurs, et en utilisant des réacteurs de nature différente (réacteurs de recherche avec un cœur hautement enrichis en ^{235}U ou réacteurs commerciaux avec une grande puissance thermique). Parmi ces expériences, cette thèse a été réalisée sur le détecteur SoLid.

L'expérience SoLid

L'expérience SoLid est basée à Mol, en Belgique, au centre de recherche SCK CEN. Le réacteur est constitué d'un cœur compact hautement enrichi en ^{235}U ce qui permet en supplément d'effectuer une mesure précise du spectre en énergie des antineutrinos émis par le $^{235\mathrm{U}}$U. Le détecteur est placé à très courte distance du réacteur, entre 6.3 et $8.9\ \mathrm{m}$ du cœur du réacteur. Cette proximité est essentielle pour la recherche d'oscillation vers un neutrino stérile, mais implique néanmoins une très faible protection contre les rayonnements cosmiques. Enfin, le réacteur subit des périodes de marche et d'arrêt d'un mois et demi, ce qui permet une analyse complète des bruits de fond attendus dans le détecteur.

L'interaction d'intérêt dans SoLid est la désintégration bêta inverse (IBD):

$$\overline{\nu}_e + p \rightarrow e^+ + n$$

Afin de détecter les antineutrinos, une technologie de scintillation hybride est utilisée, représentée sur la figure 5.8 (gauche). Des cubes de plastique scintillant (PVT) sont utilisés comme cibles des neutrinos et pour la mesure de l'énergie des positrons émis par IBD. Des feuilles de ^{6}LiF:ZnS(Ag) sont ensuite utilisées pour la capture des neutrons par le ^{6}Li après thermalisation ainsi que leur détection par la scintillation du ZnS(Ag) induits par les produits de la rupture des atomes de ^{6}Li:

$$n + {}^6\mathrm{Li} \rightarrow \alpha + {}^3\mathrm{H}$$

À cause du temps de thermalisation des neutrons, de l'ordre de quelques dizaines de microsecondes, le signal d'intérêt est une coïncidence temporelle retardée entre un signal rapide, venant de la scintillation du PVT, et un signal retardé, induit par la scintillation du ZnS après capture du neutron. Ce signal spécifique est représenté en figure 5.8 (droite).

L'utilisation de plastique scintillant dans SoLid permet d'avoir un détecteur hautement segmenté, avec 12800 cellules de détections, réparties en cinq modules, chacun d'entre eux est composé de dix plans, chaque plan possédant 16x16 cellules. Une cellule de détection est composée d'un cube de 3x3x3 cm^3 de PVT, avec une feuille de ^{6}LiF:ZnS(Ag) sur deux faces du cube. Chaque cube est isolé optiquement des autres grâce à une couche de Tyvek. Chaque cube est connecté à quatre fibres optiques qui transmettent la lumière de scintillation à des MPPCs (compteurs de photons). Chaque fibre est connectée à

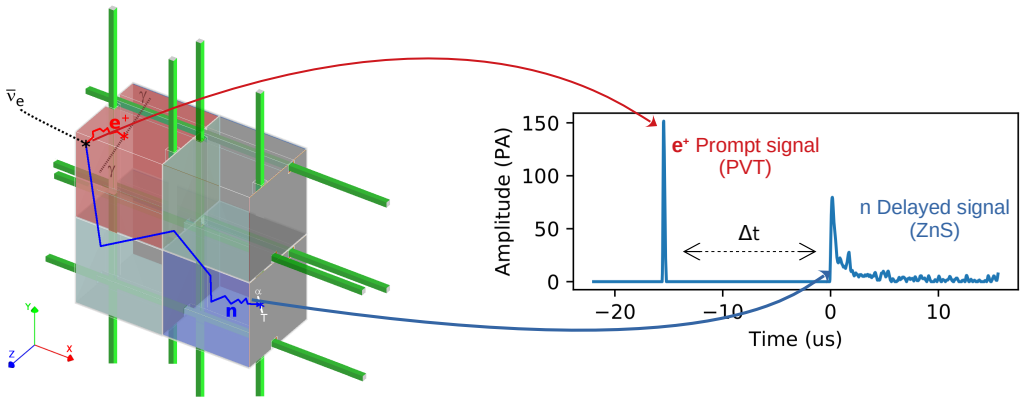


Figure 5.8: Gauche : représentation de l'interaction d'un $\overline{\nu}_e$ dans le détecteur SoLid. Le positron est détecté avec la scintillation du PVT. Le neutron est capturé par un écran de ^{6}LiF:ZnS(Ag) en induisant la rupture d'un atome de ^{6}Li et la scintillation du ZnS. Droite : pulses de scintillations avec un pulse rapide et fin, induit par la scintillation du PVT et un signal retardé, avec une constante de décroissance plus lente, lié à la capture du neutron et la scintillation du ZnS.

un MPPC à une extrémité et un miroir à l'autre pour optimiser la collection de lumière. Un tel nombre de cellules de détection requiert une calibration en énergie intensive du détecteur afin d'atteindre une bonne homogénéité de la réponse de tous les plans. De plus la segmentation du détecteur SoLid a pour but de détecter les gammas de 511 keV émis par l'annihilation des positrons après dépôt de leur énergie cinétique dans le PVT. Il est donc nécessaire de maitriser la réponse à basse énergie du détecteur avec grande précision.

Calibration en énergie

Le premier travail décrit dans cette thèse consiste en la calibration en énergie du détecteur SoLid. Afin de calibrer le détecteur, un système de calibration in situ est utilisé. Le système permet de placer des sources radioactives dans neuf positions entre chaque module, devant le détecteur et derrière le détecteur. Les sources radioactives utilisées doivent donc être des sources pénétrantes afin de calibrer deux demi-modules (devant la source et derrière la source). Pour cette raison, les sources utilisées sont des sources de gammas, dans notre cas du ^{22}Na, du ^{207}Bi et de l'AmBe. L'émissions de gammas par les trois sources peut être décrite de la manière suivante :

– Le ^{22}Na se désintègre en un état excité du ^{22}Ne via désintégration β^+ dans ~ 90 % des cas. Le ^{22}Ne émet ensuite un gamma de 1275 keV. Avec l'annihilation du positron, les gammas à disposition sont des gammas de 511 keV (x2) et de 1275 keV.

– Le ^{207}Bi se désintègre en ^{207}Pb via capture d'électron. La désintégration peut mener à trois états

d'excitation du ^{207}Pb et les gammas à disposition ont pour énergie $1770\ \mathrm{keV}$, $1063\ \mathrm{keV}$ et $569\ \mathrm{keV}$.

– L'AmBe est une source de neutron avec le processus d'interaction suivant :

$$^{241}\mathrm{Am} \rightarrow \alpha + {}^{237}\mathrm{Np}$$
$$\alpha + {}^{9}\mathrm{Be} \rightarrow n + {}^{12}\mathrm{C}^{*}$$

Avec l'émission d'un gamma de $4438\ \mathrm{keV}$ par le carbone excité.

La calibration en énergie du détecteur commence par la correction du gain et du piédestal de chaque MPPC. Si ces paramètres ont été mesurés et corrigés lors de la mise en route du détecteur, une dérive dans le temps peut apparaitre. Une mesure du gain et du piédestal de chacun des 3200 MPPC est effectuée toutes les six heures en analysant les données de physique (sans source de calibration). Les paramètres sont ensuite stockés dans une base de donnée pour corriger au plus proche les dérives qui peuvent être du à des variations des paramètres environnementaux.

Une fois les effets de perte de lumière corrigés l'étape suivante de la calibration en énergie consiste en la mesure du rendement lumineux visible des 12800 cubes de PVT. Dans un premier lieu, le ^{22}Na est utilisé. Dû à la taille des cellules de détections, les gammas ne vont pas déposer toute leur énergie dans un cube et seule l'information des fronts Compton est disponible. Deux méthodes ont été développées pour réaliser la mesure du front Compton. La première se base sur la section efficace de Klein-Nishina pour la diffusion Compton des gammas. La section efficace est ensuite convoluée par une fonction gaussienne dépendant de la résolution en énergie à $1\ \mathrm{MeV}$ σ_0. Une fonction de probabilité de densité est ensuite créée :

$$f_{conv}(x) = \frac{\sum_{i=0}^{i_{E_c}} \epsilon_{reco}(T_i)\epsilon_{loss}(T_i)\frac{d\sigma}{dT}(T_i)\frac{1}{\sqrt{2\pi}\sigma_0\sqrt{T_i}}\exp\left(-\frac{(\frac{x}{LY}-T_i)^2}{2\sigma_0^2 T_i}\right)}{\sum_{i=0}^{i_{E_c}} \epsilon_{reco}(T_i)\epsilon_{loss}(T_i)\frac{d\sigma}{dT}(T_i)}$$

Avec x, le nombre de photons mesurés par les quatre MPPCs d'un cube (en PA), E_c l'energie du front Compton, ϵ_{reco} et $\epsilon_{loss}(T_i)$ des facteurs d'inefficacité, $\frac{d\sigma}{dT}$ la section efficace d'interaction. Les paramètres ajustés avec la fonction de probabilité sont la résolution en énergie ()σ_0) et le rendement lumineux (LY). La deuxième méthode développée est une méthode d'ajustement numérique. En partant de simulations GEANT4 de ^{22}Na, une convolution numérique gaussienne est appliquée avec une résolution en énergie donnée σ_0 et un facteur d'échelle LY est appliqué aux données de mesure ^{22}Na. Pour chacun des couples (σ_0, LY) un test de χ^2 est effectué entre les données mises à l'échelle et les simulations convoluées pour sélectionner le couple avec le score le plus bas.

Une comparaison entre les deux méthodes a ensuite été opérée, celle-ci montre qu'il y a un biais inférieur à $1\ \%$ entre les deux mesures. Enfin, afin de tester la précision des méthodes, une comparaison de mesures a été faite avec des simulations, entre les méthodes utilisant le ^{22}Na et une mesure du rendement lumineux avec des électrons monoénergétiques. Ces derniers ont l'avantage de donner un spectre en énergie très simple à analyser comme il n'est composé que d'une gaussienne, centrée autour de la valeur d'intérêt. Les comparaisons entre les deux mesures donnent une nouvelle fois un biais inférieur à 1 %, ce qui confirme l'excellente précision de la mesure du rendement lumineux en utilisant les fronts Compton.

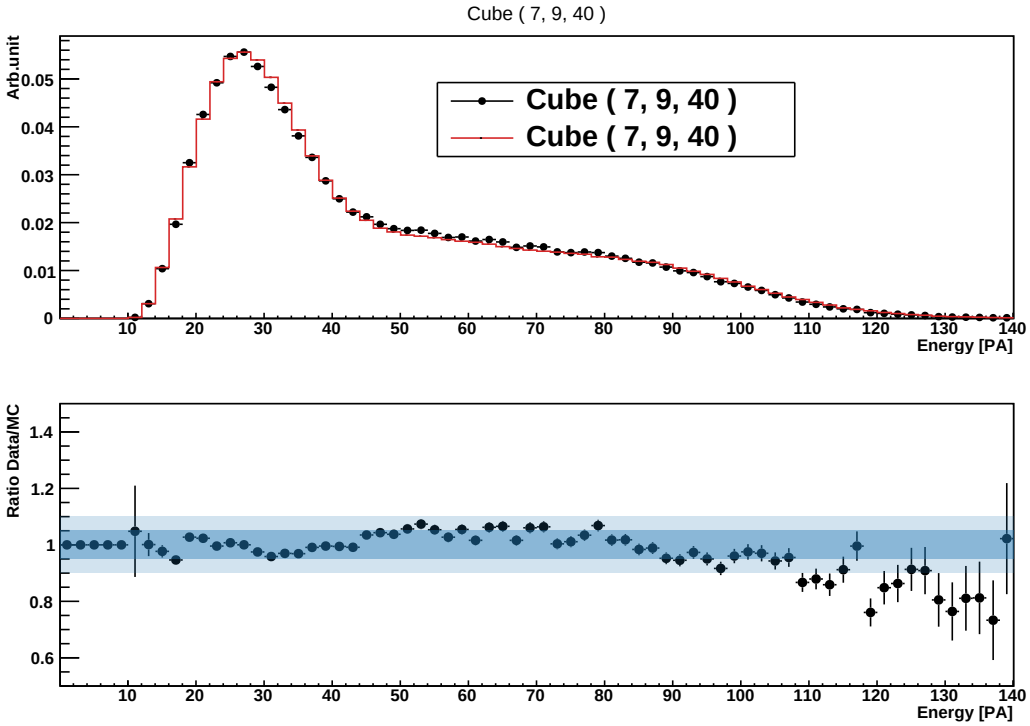


Figure 5.9: Comparaison entre les données et les simulations pour des spectres en énergie de ^{22}Na dans le cube (7,9,40).

Cette calibration en énergie peut ensuite être utilisée afin de mesurer l'homogénéité de la réponse du détecteur. Ainsi, le rendement lumineux moyen du détecteur est de 96 PA par MeV, les variations de rendement sont de 5 % au maximum entre deux modules avec une variation de l'ordre de 3 % dans un plan. De plus, avec l'utilisation des autres sources radioactives, la linéarité de la réponse en énergie du détecteur a pu être mesurée à quelques pourcents.

Enfin, la précision de la calibration en énergie a permis un excellent accord Data/Monte-Carlo en comparant les simulations en ^{22}Na, jusqu'à la centaine de kilo electronvolts. Un exemple de comparaison des spectres en energie entre données et simulation Monte-Carlo peut être vu sur la figure 5.9 pour un cube donné. Cette maitrise de la réponse à basse énergie du détecteur est essentielle pour la recherche de gamma d'annihilations qui ne déposeront que quelques centaines de keV dans le détecteur.

Analyse d'antineutrinos

La deuxième grande partie décrite dans la thèse est l'analyse d'antineutrinos. Cette analyse consiste en plusieurs étapes. Tout d'abord la quantification des différents types de bruit de fonds. Ensuite l'optimisation du signal contre le bruit de fond. Enfin, l'extraction d'antineutrinos. Ces travaux se placent dans le cadre d'une analyse en aveugle : l'analyse est d'abord testée sur un petit jeu de donnée, ici 21 jours de réacteur ON, et un nombre équivalent de réacteur OFF. Une fois l'analyse validée, elle sera appliquée à plus grande échelle. Dans cette thèse, seulement la validation de l'analyse a été effectuée.

Les principaux bruit de fonds dans SoLid sont le bruit de fond accidentel, le bruit de fond BiPo et enfin le bruit de fond atmosphérique. L'analyse de bruit de fond se fait à l'aide de données réacteur OFF dans lesquelles sont présents les trois types de signaux. Le bruit de fond accidentel est composé de coïncidences temporelles fortuites entre un signal retardé neutron (NS) et un signal rapide électromagnétique (ES). Il peut typiquement être induit par la capture de neutron du réacteur par du ^{40}Ar, du carbone ou de l'hydrogène ou bien des muons mal identifiés. Ce bruit de fond est quantifié à l'aide de faux déclencheur neutrons : des signaux qui déclenchent la prise de mesure comme un neutron pourrait le faire, mais qui ne sont pas identifiés comme un neutron par la suite. Ces signaux sont en réalité des muons qui traversent le détecteur et aucune coïncidence temporelle est attendu entre eux et un autre signal ES. Ainsi les coïncidences entre les signaux ES et les faux déclencheurs neutron permettent de quantifier le taux de bruit de fond accidentel dans le détecteur. Cette méthode est appelée, méthode FPNT ou *False Positive Neutron Triggers*.

Le bruit de fond BiPo est un bruit de fond induit par la radioactivité naturelle des feuilles de ^{6}LiF:ZnS(Ag), dû à une contamination inattendue, ainsi que par la désintégration de radon gazeux, présent dans l'air et émis continuellement par les murs de l'enceinte de confinement. La chaine de désintégration responsable de ce bruit de fond est la suivante :

$$^{214}_{83}\mathrm{Bi} \xrightarrow[19.9\mathrm{min}]{\beta^-} {}^{214}_{84}\mathrm{Po} \xrightarrow[164\ \mu\mathrm{s}]{\alpha} {}^{210}_{82}\mathrm{Pb}$$

Le ^{214}Bi se désintègre en ^{214}Po par désintégration β. La particule émise induit la scintillation du PVT et constitue le signal rapide. Le ^{214}Po se désintègre avec l'émission d'un α qui induit la scintillation du ZnS. Cette scintillation constitue le signal retardé. Le temps caractéristique du signal BiPo est d'environ $250\ \mu$s, soit légèrement plus long que celui de l'IBD ($\sim 64\ \mu$s). De plus, la scintillation du ZnS est induite par un α de 7.8 MeV contrairement à une scintillation suite à la capture de neutron qui est induite par un α et un ^{3}H avec une énergie totale de 4.8 MeV. Cette différence d'énergie induit une différence dans les signaux de scintillation qui peut être utilisée pour faire de la discrimination de formes d'onde. À l'aide d'un réseau de neurones convolutionnel, entrainé à différentier les deux types de formes d'onde, une nouvelle variable est créée : le BiPonator. En combinaison avec la différence temporelle entre les signaux NS et ES (ΔT), une fenêtre enrichie en BiPo est créée (bas BiPonator, haut ΔT) ainsi qu'une fenêtre enrichie en signaux induit par des captures de neutrons (haut BiPonator, bas ΔT), comme présenté sur la figure 5.10. La quantité de BiPo dans la fenêtre de signal est ensuite extrapolée de la fenêtre BiPo avec un facteur d'échelle.

Le dernier bruit de fond majoritaire dans SoLid est le bruit de fond atmosphérique. Il est induit par l'interaction de rayons cosmiques de haute énergie dans l'atmosphère. Il se retrouve sous deux formes principales. La première consiste en des neutrons rapides qui effectuent des reculs de protons dans le détecteur, puis sont capturés après thermalisation. Les reculs de protons constituent le signal ES et la capture des neutrons le signal retardé NS. La deuxième forme consiste en un muon traversant le détecteur et produisant un neutron de spallation. Le muon, s'il n'est pas bien identifié comme muon créé le signal ES et le neutron de spallation s'il est capturé après thermalisation dans le détecteur créé le signal retardé NS. Dans ces deux cas, le temps caractéristique de l'interaction est le même que pour les IBDs, $\sim 64\ \mu$s. Dans ce cas-là, aucune variable ne peux être utilisée pour opérer une quantification du bruit de fond atmosphérique dans les données. Néanmoins, le bruit de fond atmosphérique est dépendant de la densité de l'atmosphère et donc de la pression atmosphérique. Ainsi, un modèle de

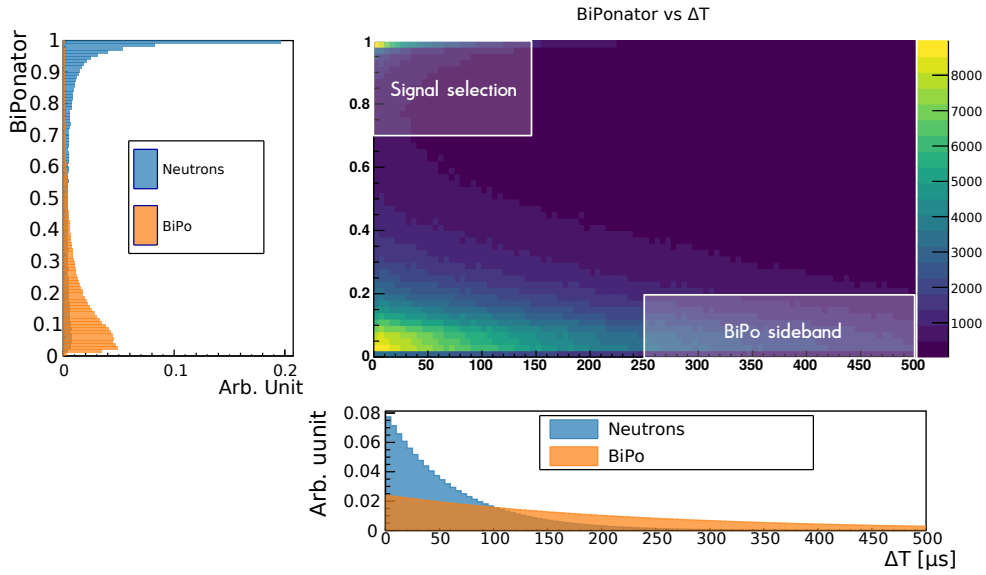


Figure 5.10: Histogramme 2D : distribution ΔT vs BiPonator pour les données réacteur OFF. Les distributions 1D du ΔT et du BiPonator sont représentées normalisées à l'air pour des événements induits par capture de neutron après thermalisation (bleu) et des événements BiPo (orange). Les deux rectangles blancs sur l'histogramme 2D représentent la sélection utilisée pour la fenêtre BiPo et la fenêtre enrichie en signal neutron, utilisée pour l'analyse d'antineutrino par la suite.

pression qui corrèle la pression atmosphérique et le taux d'atmosphérique est créé. Dans les données réacteur OFF et ON, le modèle est testé sur des données au-dessus de 7 MeV. En soustrayant le bruit de fond accidentel via l'utilisation des FPNTs, seul le bruit de fond atmosphérique reste dans ces données, le BiPo étant prédominant sous 3 MeV et l'énergie des antineutrinos étant inférieur à 7 MeV. Avec ce bruit de fond atmosphérique, une très claire anti-corrélation est observée entre les variations de bruit de fond atmosphérique et les variations de pressions comme on peut le voir sur la figure 5.11. Cette anti-corrélation sera par la suite utilisée pour extrapoler le niveau de bruit de fond atmosphérique dans les données réacteur ON a parti des données réacteur OFF.

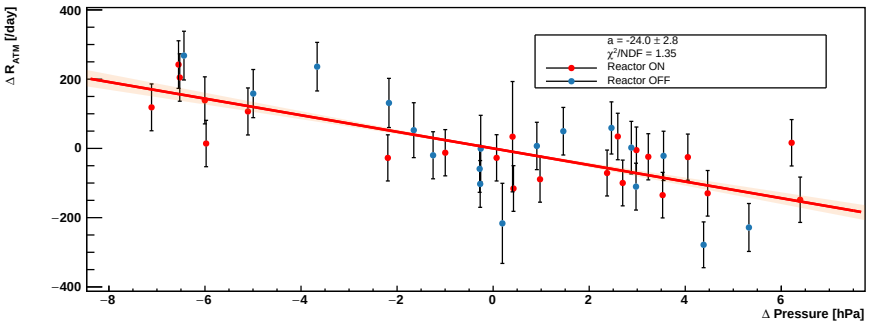


Figure 5.11: Variation du taux de bruit de fond atmosphérique dans l'échantillon de donnée en fonction de la variation de pression. La ligne rouge représente le modèle de pression et la bande orange représente l'erreur à 1 σ sur la pente du modèle.

Une fois les bruit de fonds quantifié, l'étape suivante et d'optimiser la sélection de signal vs bruit de fond. Afin d'évaluer les performances d'une sélection donnée, on étudie à la fois le nombre d'IBD attendu par jours à l'aide de simulation Monte-Carlo, et le nombre d'événements bruit de fond mesurés par jours dans les données réacteur OFF. Les premières modélisations prévoient un nombre d'IBD dans le détecteur de 1088 par jours pour le jeu de donnée utilisé dans notre analyse.

Pour cela, une première sélection est effectuée sur des variables de bas niveau, comme des différences spatiales et temporelles entre les signaux NS et ES, l'énergie et le BiPonator. Après cette première sélection, il reste 149 événements IBD par jours pour 4892 événements bruit de fond, soit un rapport signal/bruit de 0.03. Afin d'améliorer ces performances un algorithme de reconstruction des traces gamma est ensuite utilisé. Cet algorithme, permet la reconstruction des traces d'énergies déposées par les gammas d'annihilation dans le détecteur à l'aide d'un score de vraisemblance basé sur la section efficace de Klein-Nishina. Ainsi, zéro, une ou deux traces peuvent être reconstruites. Des variables sont ensuite associées à ses traces, telles que l'énergie, le score, le nombre de cubes dans la trace, le produit scalaire entre les deux traces... et une deuxième sélection peut être effectuée. Suite à cette deuxième sélection, il reste 91 IBD par jour pour 1303 événements bruit de fond. Pour optimiser une seconde fois la sélection de donnée, un arbre de décision boosté (BDT) est utilisé. Cet algorithme prend en entré un jeu de donnée identifié comme signal (les simulations IBD), un jeu de donnée identifié comme bruit de fond (une partie des données réacteur OFF), et un jeu de variables. Un score est ensuite attribué à chaque événement suivant qu'il ressemble plus à du bruit de fond ou du signal. En utilisant le BDT pour optimiser la sélection, la performance finale de l'analyse est 65 événements IBD par jour pour 243 événements bruit de fond, soit un ratio signal sur bruit de 0.27.

Une fois la sélection des données fixée, la dernière étape est l'extraction d'antineutrino des données réacteur ON. Pour ce faire, la première étape consiste à sélectionner la fenêtre signal dans l'espace BiPonator vs ΔT. Dans cette sélection, le bruit de fond accidentel est quantifié et soustrait à l'aide des FPNTs. Le bruit de fond BiPo est quantifié et soustrait à l'aide de la fenêtre BiPo. Après avoir retiré les deux premières sources de bruit de fond, il ne reste dans notre jeu de donnée réacteur OFF que du bruit de fond atmosphérique et dans les données réacteur ON, le bruit de fond atmosphérique ainsi que les IBDs. Ce dernier bruit de fond est quantifié et soustrait à l'aide du modèle de pression qui permet d'extrapoler à partir du taux d'atmosphériques dans les données réacteur OFF, la contamination des données réacteur ON. Après soustraction de tous les bruits de fond, on obtient un excès de 71 ± 4 événements par jours dans le réacteur ON, et un résidu de -1 ± 4 événements par jours dans le réacteur OFF. Une comparaison des distributions en énergie et de position d'interaction des événements dans le détecteur de cet excès avec des simulations Monte-Carlo d'IBDs montrent un très bon accord statistique, ce qui tend à valider le fait que l'excès observé est bien un excès dû à l'interaction d'antineutrinos dans le détecteur.

Une fois l'extraction d'antineutrinos terminée, la dernière étape consiste à effectuer une première analyse de sensibilité de l'expérience à l'hypothèse du neutrino stérile. Cette étude a été réalisée par des collègues de l'Imperial College of London. À l'aide de distributions en distance traversée et énergie des antineutrinos fournis par l'analyse d'extraction sur les données de 21 jours de réacteur ON, extrapolées à deux an de prises de mesures, le contour de sensibilité est représenté en figure 5.12. Celui-ci a été obtenu en utilisant une méthode fréquentiste prescription de Feldman-Cousins. Ce contour prend en compte les premières incertitudes systématiques que sont l'acceptance du détecteur, l'efficacité de capture des

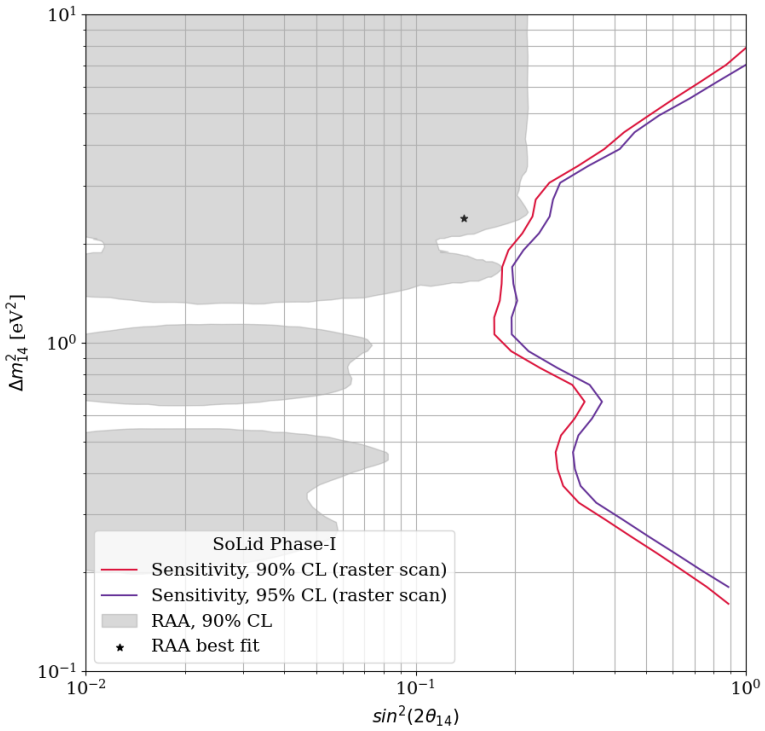


Figure 5.12: Contour de sensitivité préliminaire avec l'analyse d'antineutrino actuelle, pour ~ 300 jours de données réacteur ON, et ~ 180 jours de données réacteur OFF. Le contour prend en compte les incertitudes systématiques suivantes : l'acceptance du détecteur, l'efficacité de capture des neutrons ainsi que l'échelle en énergie.

neutrons ainsi que l'échelle en énergie. Des études sont en cours pour mesurer les incertitudes restantes. Avec la validation du travail d'analyse sur cette thèse, la phase d'ouverture des données pourra avoir lieu prochainement avec la production du premier contour d'exclusion de l'expérience SoLid. Enfin, malgré la faible compétitivité des résultats de SoLid par rapport aux résultats d'expériences similaires comme STEREO, des améliorations de l'analyse sont en cours d'implémentation, avec une meilleure reconstruction des événements et une sélection plus pure d'événements IBDs. Enfin, l'amélioration du détecteur qui a eu lieu pendant l'été 2020 amène une augmentation du rendement lumineux du détecteur de 40 %, ce qui sera une très grande aide pour améliorer le traçage des gammas d'annihilation, ce qui permettra d'améliorer les performances de l'analyse présentée dans cette thèse.

université PARIS-SACLAY

ÉCOLE DOCTORALE
Particules, hadrons, énergie et noyau : instrumentation, imagerie, cosmos et simulation (PHENIICS)

Titre: Recherche d'oscillation vers un neutrino stérile auprès de l'expérience SoLid située au réacteur BR2 : calibration en énergie du détecteur et extraction du signal antineutrino.

Mots clés: Neutrino, Oscillation, Stérile, SoLid, Calibration en énergie

Résumé: SoLid (Search for oscillations with a lithium-6 detector) est une expérience de recherche d'antineutrino à très courte distance située au réacteur BR2 en Belgique. C'est un détecteur de 1.6 tonnes qui couvre une distance entre 6.3 et 8.9 m du cœur du réacteur. L'objectif de l'expérience est la recherche d'oscillation d'antineutrinos électroniques vers un état de neutrino stérile et léger pour sonder "l'anomalie d'antineutrino de réacteur". De plus, le cœur du réacteur BR2, hautement enrichi en uranium-235, permet une mesure précise du spectre en énergie des antineutrinos émis par l'uranium-235. Cette mesure pourra aider à la compréhension de l'anomalie spectrale autours de 5 MeV observée par des expériences précédentes.

Le détecteur SoLid utilise une technologie hybride de scintillation. Cette technologie est basée sur la combinaison de scintillateurs plastiques et de feuilles de ^{6}LiF:ZnS(Ag). L'utilisation de scintillateurs plastiques permet une grande segmentation du détecteur avec 12800 cellules de détection formées de cubes en PVT, mesurant $5{\times}5{\times}5$ cm^3, associés à des feuilles de ^{6}LiF:ZnS(Ag) collées sur deux des côtés des cubes. Les antineutrinos interagissent dans les cubes de PVT via désintégration bêta inverse, avec pour produit un positron et un neutron. L'énergie du positron est mesurée à l'aide de la scintillation du PVT. Le neutron, lui, est capturé par le lithium-6, ce qui induit une scintillation du ZnS. Les photons de scintillation sont ensuite transmis à des MPPCs à l'aide de fibres optiques. Les MPPCs sont utilisés pour lire les signaux lumineux.

Cette thèse comprend une revue des oscillations de neutrinos avec une présentation des derniers résultats expérimentaux, ainsi qu'une description complète de l'expérience SoLid. Les deux principaux travaux de la thèse sont ensuite présentés. Le premier étant la calibration en énergie du détecteur. Cette calibration représente une grande partie du travail décrit dans ce document. Celle-ci comprend la mesure de la quantité de lumière de chacune des cellules en plus de la linéarité et de l'homogénéité de la réponse en énergie du détecteur. Cette calibration a permis une connaissance précise de la réponse en énergie du détecteur ainsi que son évolution dans le temps à l'aide de multiples campagnes de calibration.

La seconde partie du travail présenté dans cette thèse concerne l'extraction du signal antineutrino. Le détecteur est situé à quelque mètres d'un réacteur nucléaire avec peu de protection contre le rayonnement cosmique. Il doit donc faire face à d'importants taux de bruit de fond, ainsi, l'extraction du signal antineutrino est un vrai challenge. Les différentes sources de bruit de fond sont décrites et analysées dans ce document, de même que les méthodes développées pour les rejeter. L'analyse finale à l'aide d'arbres de décisions boostés est ensuite présentée, suivie d'une discussion sur les incertitudes systématiques. La stabilité de l'analyse a été testée et fourni une extraction robuste des antineutrinos. Ce travail est enfin conclu par une analyse de 21 jours de données de réacteur, avec la mesure d'un spectre en énergie d'antineutrinos ainsi que de la distance qu'ils ont parcourue avant d'interagir dans le détecteur. Ces dernières mesures permettent d'avoir les données nécessaires pour déterminer la sensibilité de l'expérience SoLid à la recherche d'une oscillation vers un neutrino stérile.

Title: Search for sterile neutrino oscillations with the SoLid experiment at BR2 reactor: Energy calibration of the detector and antineutrino signal extraction.

Keywords: Neutrino, Oscillations, Sterile, SoLid, Energy calibration

Abstract: SoLid (Search for oscillations with a Lithium-6 detector) is a very short baseline reactor antineutrino experiment based at the Belgian BR2 reactor. It is a 1.6 ton detector covering a distance from 6.3 to 8.9 m from the reactor core. Its main purpose is the search for electron antineutrino oscillation to a light sterile state in order to probe the "reactor antineutrino anomaly". Thanks to the highly enriched in ^{235}U BR2 reactor core, the detector also aims to provide a precise measurement of the antineutrino spectrum from ^{235}U. The measurement of this energy spectrum could help to understand the spectral anomaly at 5 MeV in the reactor antineutrino energy spectrum observed by precedent experiments.

The SoLid detector uses a novel hybrid scintillation technology. It is based on the combination of plastic scintillator and ^{6}LiF:ZnS(Ag) screens. The use of plastic scintillators allows a fine segmentation of the detector with 12800 detection cells in the form of $5 \times 5 \times 5$ cm^3 PVT cubes with ^{6}LiF:ZnS(Ag) screens on two sides. The PVT cubes are used as targets for the antineutrinos via inverse beta decay (IBD) interactions. The energy of the resulting positron is measured with the PVT scintillation, while the resulting neutron is captured by the ^{6}Li, inducing the scintillation of the ZnS. The scintillation photons are then transmitted via wavelength shifting fibres to MPPCs that read out the light signals.

After a review of the neutrino oscillation with the latest experimental results and a complete description of the SoLid experiment, the two main tasks of this thesis are presented. The first main contribution of this thesis to the SoLid experiment is the energy calibration of the detector. A great focus has been put on the calibration of each detection cell with light yields measurement, linearity assessment and the test of the homogeneity of the detector response. The contribution of this work to the energy calibration of the detector allowed a precise knowledge of the energy response of the detector. The evolution through time of the detector energy response has also been studied with an analysis of multiple calibration campaigns.

The second task presented in this thesis is the extraction of the antineutrino signal. As it is a detector located at a few meters from a nuclear reactor, with close to no overburden, SoLid faces a large proportion of various backgrounds. The extraction of the antineutrino signal is thus real a challenge. The backgrounds faced by the experiment are described and analysed in this work, as well as the methods developed to reject them. The final analysis using boosted decision trees is then presented with a discussion on the systematic uncertainties related to it. The stability of the analysis has been tested and gives a robust measurement of the antineutrino rates. The conclusion of this work is an analysis of 21 days of reactor data, with the extraction of an antineutrino energy spectrum and distance travelled before interacting in the detector. This last measurements give the necessary inputs to determine the sensitivity to sterile neutrino of the experiment and to perform the search for a sterile neutrino signal.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France