Toward the search of a sterile neutrino with the SoLid experiment at the SCK-CEN BR2 reactor : detector calibration and design of a multivariate topological selection of antineutrino events

David Henaff

UNIVERSITE MATIERE BRETAGNE MOLECULES LOIRE ET MATERIAUX

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

THÈSE DE DOCTORAT DE

L'ÉCOLE NATIONALE SUPÉRIEURE MINES-TÉLÉCOM ATLANTIQUE
BRETAGNE PAYS DE LA LOIRE – IMT ATLANTIQUE

ÉCOLE DOCTORALE 596
Matière, Molécules, Matériaux
Spécialité : Physique Subatomique et Instrumentation Nucléaire

Par
David HENAFF

Toward the search of a sterile neutrino with the SoLid experiment at the SCK-CEN BR2 reactor : detector calibration and design of a multivariate topological selection of antineutrino events.

Thèse présentée et soutenue à Nantes, le 04 Juin 2021
Unité de recherche : SUBATECH, UMR 6457
Thèse N°2021IMTA0246

Rapporteurs avant soutenance

Christine MARQUET	Directrice de recherche, CNRS, CENBG (France)
Jacob LAMBLIN	Maître de conférences, Université Grenoble Alpes, LPSC (France)

Composition du Jury

Président	**Gines MARTINEZ**	Directeur de Recherche, CNRS, SUBATECH (France)
Rapporteurs	**Christine MARQUET**	Directrice de Recherche, CNRS, CENBG (France)
	Jacob LAMBLIN	Maître de Conférence, Université Grenoble Alpes, LPSC (France)
Examinateurs	**Jorgen D'HONDT**	Professeur, Vrije Universiteit Brussel (Belgique)
	Sara BOLOGNESI	Chercheuse, CEA (France)
Directeur de thèse	**Benoît REVENU**	Directeur de Recherche, CNRS, SUBATECH (France)
Encadrants scientifiques	**Benoît VIAUD**	Chargé de Recherche, CNRS, SUBATECH (France)
	Frédéric YERMIA	Maître de Conférence, Université de Nantes, SUBATECH (France)

À mon père,

Acknowledgements

Ce manuscrit ne présente pas le travail et l'investissement d'une seule personne, mais bien un travail collectif. C'est donc avec beaucoup d'émotion et de souvenir que je souhaite remercier l'ensemble des personnes qui m'ont aidé tout au long de ces trois ans et huit mois.

Je souhaite tout d'abord remercier l'ensemble de mon jury : Gines Martinez pour avoir bien voulu présider mon jury de thèse. Aussi, je tiens à remercier énormément mes deux rapporteur Christine Marquet et Jacob Lamblin pour leur lecture, questions et corrections. J'ai eu de très intéressantes discussions avec vous deux pour préparer la soutenance. Enfin merci à Sara Bolognesi et Jorgen D'Hondt pour leur participation en tant qu'examinateur.

Cette thèse est le résultat d'un travail collaboratif avec mes deux encadrant de thèse : Benoit Viaud et Frédéric Yermia. Merci à vous deux pour m'avoir fait confiance au début de la thèse pour mener à bien ces travaux. C'était loin d'être une thèse simple au vu du contexte, mais vous m'avez apporté de nombreuses discussions pour m'encadrer et m'apprendre de nouvelles choses ainsi que me laisser de la liberté lorsque je voulais développer des idées. Merci à toi Benoit pour m'avoir transmis ta rigueur que ce soit dans l'analyse que dans l'écriture, même si à la fin je n'en pouvais plus j'ai beaucoup appris avec toi. Merci à Fred qui m'a offert mon premier stage en neutrino et ma donnée le goût de la recherche en physique avec Thiago malgré la période compliqué avec la maladie de mon père. Au-delà du travail, on a eu aussi beaucoup de bons moments que ce soit lors de voyage pour les meetings de collaboration : la course à travers l'aéroport Charles de Gaulle en revenant de Bristol ou bien la Jeep à Genève avec du Cardi B à fond.

Un grand merci à l'ensemble de ma famille qui m'a aidé tout au long de cette thèse. Vous m'avez permis de me sortir la tête de l'eau quand j'en avais besoin et été de bon conseil. En particulier, merci à mes parents qui m'ont donné un cadre plus que favorable pour étudier et ma mère pour avoir organisé l'après-soutenance malgré les problèmes de restriction sanitaires. Vous m'avez soutenu tout au long de cette thèse même si mon travail vous paraissait incompréhensible. Vous avez toujours été à mon écoute.

Bien sur une telle aventure s'accompagne toujours des amis thésards, j'ai passé beaucoup trop de bons moments avec vous tous : Manu (Mich'), Antony, David (Escroc), Erwann, Loic, Chloé, Florian, Gabriel, Ophélie, Quentin et le nouvel arrivant dans le groupe neutrino : Victor. Je me souviens surtout des pauses quotidiennes avec Mich', toujours en train de fomenter de sales coups, écrire des paroles et approfondir notre capacité d'agacement. Je crois qu'on a pris beaucoup d'expérience sur

ce dernier point ! Je souhaite aussi remercier les personnes avec qui j'ai pu échanger au sein de la collaboration SoLid : Maja, Yamiel, Dan, Valentin, Luis, Mathieu, Noé (bon courage pour l'écriture et la soutenance!).

Parmi mes amis proches je souhaite remercier en particulier Hugo et Julia. On a toujours réussi à se voir quand j'étais de passage sur Paris ou toi sur Nantes. Ça me faisait un bien fou de venir te voir et passer de longues soirées à discuter (Up5!). Tu m'as fait découvrir les folles soirées parisiennes, toi-même tu sais..! Merci à toi Julia aussi, on a énormément discuté et j'ai appris plein de choses avec toi, je me suis rarement aussi bien entendu en si peu de temps avec quelqu'un. On a passé de sacrées soirées ensemble, et quelles soirées, les rendez-vous de l'erdre, transfert, les fanfaronnades (dédicace à Mich'). Et même durant le confinement, vous étiez les personnes avec qui je parlais le plus grâce aux soirées Rocket League du mercredi, vive la team licorne. Vous êtes les meilleurs. Aussi un grand merci au Planck et son équipe pour avoir soutenu la cause tout au long de la thèse ! Toujours présent, fidèle au poste pour parler de science en buvant une bonne pinte de bière et finalement rentrer à pied.. ! Il est aussi important de citer Laurène qui malgré la distance m'a épaulée pendant une partie de cette thèse. On a partagé beaucoup ensemble que ce soit à l'étranger où en France avec les folles journées par example. Merci pour ces moments.

Contents

List of figures

List of tables

Introduction

The Standard Model of particle physics encompasses our present understanding of this domain. Its success lies in the fact that it can explain all experimental observations carried out in the laboratory, and even predicted many of them, sometimes decades in advance. A growing list of intriguing anomalies appeared in the last decade or two, but still have to be firmly confirmed. On the other hand, there are strong reasons to believe the SM is only the manisfestation at the presently reachable energy scales of a more fundamental physics. One of them resides in the difficulty for the SM to deal with mass, flavors and CP violation. For instance, the level of CP violation possible in this model fails to explain the baryon asymmetry of the Universe. More generally, the SM does not explain the deep origin of mass, flavors and CP violation.

Neutrino physics is a powerful probe of physics beyond the SM (BSM). Neutrino oscillations provide a very rich phenomenology: neutrinos have mass, their flavors mix, and could reveal in the coming decade a new source of CP violation. The pattern associated to this involves a lot of observables. Measuring them precisely is a new way to explore the issues introduced in the previous paragraph. In particular, the extreme smallness of neutrino masses, at least five 5 orders of magnitude below that of the lightest charged fermion, should carry important information. Indeed, in the case of a neutral fermion, Dirac terms are no longer the privileged way to generate mass. Neutrinos could actually be the first observed Majorana fermions. This has implications beyond the SM, in particular the possibility to violate the conservation of the lepton number.

Over the last 20 years, the intensive study of neutrino oscillations allowed the precise description of this phenomenon within the standard 3 flavour model. The three mixing parameters that can be used to express the PMNS matrix in the Unitarity assumption, θ_{12}, θ_{23} and θ_{13}, as well as the squared mass differences Δm^2_{21} and Δm^2_{31} are now known at the few percents level. Despite this progress, neither the neutrino mixing pattern nor the smallness of neutrino masses are explained. A new generation of experiments is beeing built to constrain the PMNS matrix better and explore beyond the 3 flavor model. Long baseline, accelerator based experiments (DUNE, T2(H)K) hope to measure CP violation in the neutrino sector in the coming decade and establish the neutrino mass ordering. The latter is also the goal of JUNO, a medium baseline experiment measuring reactor antineutrinos. Precision oscillation studies possible with these experiments will also challenge the Unitarity of the PMNS matrix, which violation would sign BSM physics. Beyond the study of oscillations, a wealth of experiments search for neutrinoless double bêta decays, which observation would sign the existence of Majorana neutrinos. Another way to explore BSM physics is to search for additional neutrinos, not yet known. This is the main goal of the SoLid experiment.

The **first chapter** of this thesis is an introduction to the SM, and to the role of neutrinos. An experimental state of the art will also be presented, with a focus on measurements of neutrino oscillation parameters. It will also present a handful of anomalies observed at the turn of the century, among them the reactor antineutrino anomaly. Anomalies could be explained by the existence of a new neutrino state, a *sterile neutrino*, at the eV-scale. At such a mass, it could be responsible for an oscillation at very short range : only few meters from the reactor. A worldwide effort has been initiated in order to measure the flux and spectrum of antineutrinos at this distance from the core of experimental reactors.

The SoLid experiment entirely falls under this context. The **second chapter** describes the experimental apparatus, the standard methods and the status of the sterile analysis in Solid, as of late 2019. The experiment is based on a novel technology. It responds to the high background levels that experiments have to face when close to a reactor and therefore counting only on a low overburden. The detector is a highly segmented target with 12,800 cubes, readout by 3,200 WLS fibres each connected to a MPPC. This segmentation is possible thanks to the usage of a plastic scintillator (PVT). Antineutrinos are detected via Inverse Beta Decays (IBD): the positron and two annihilation gammas are detected directly in the PVT, while an inorganic scintillator (^{6}LiF :ZnS(Ag)) is used for neutron detection. This segmentation allows to have a fine reconstruction of the topology and energy of interactions happening in the detector, which strengthens the discrimination against backgrounds.

The **third chapter** is dedicated to the energy calibration of the detector. This is an essential work for several reasons. The way to detect a sterile oscillation is to measure the distortion it causes in the $(E_\nu; L)$ distribution, where L is the distance from the production to the detection point and E_ν the antineutrino's energy. Reconstructing the energy is therefore essential, and the fundamental bricks of this measurement is the calibration of the response of the cubes and fibres. It is also a way to correct for inhomogeneities that could washout the expected oscillation. More generally, the measurement relies on the control of detector effects and the understanding of the energy scale. The tool of choice for that is a simulation able to reproduce the detector response accurately. Among the parameters to tune for this in the readout simulation, the calibration constants, or some parameters closely related to them. Calibration in SoLid is challenging, in particular because of the high number of parameters to determine. Other constraints caused by the high segmentation will also be presented, as will be the original methods that we have developed to carry out the calibration. The work devoted to the tuning of the readout simulation is presented in **chapter 4**. We also demonstrated in this chapter that the level at which we understand the detector allows the exploitation of low energy deposits, one of the conditions to the success of the IBD selection we describe in **chapter 5**.

Finally, **chapter 5** is indeed dedicated to the selection and extraction of the IBD signal. The status of the analysis two years ago was critical : with the selection which was by then the standard, background levels were still too high to hope for a competitive sterile neutrino search. At this moment the signal selection did not use low energy deposits of the annihilation gammas. Their distinctive, back-to-back topology was therefore not exploited. We developed a reconstruction method using these low energy

deposits and derived new discriminative variables. We also worked on multivariate analyses to fully exploit the potential of that. The performance of this selection will finally be presented.

Chapter 1.

Neutrino physics, state of the art

Mounika. *How Are You*, Maju Records, 2017

1.1. Standard Model of particle physics

The theory describing the known elementary of particles composing the Universe and the way how they interact each others is called the *Standard Model of particle physics*. It has been built during the 20th century thanks to both theoretical and experimental developments. This section aims to give an introduction of this theory.

1.1.1. Generalities

1.1.1.1. Particles

All the elementary particles involved in the Standard Model are categorised into two groups: the fermions which are the particles composing the matter with a half-integer spin values and the boson which are the gauge bosons of interactions with an integer spin values. There are three interactions defined in the SM with their corresponding gauge bosons:

- **Electromagnetic:** mediates via the photon γ.
- **Weak:** mediates via W^+, W^- and the Z^0.
- **Strong:** mediates via 8 gluons g.

There are twelve fermions and their corresponding antiparticles in the SM. They are divided into two types: six quarks and six leptons. Depending on the types particles are sensible to different interactions:

- **Quark:** They are sensible to the three interactions.
- **Leptons:** They can only interact via the electromagnetic and the weak interactions.

Fermion Family	Name	Mass (MeV)	Left-handed			Right-handed		
			Electric charge	Weak isospin	Weak hypercharge	Electric charge	Weak isospin	Weak hypercharge
Leptons	ν_e	0	0	$+\frac{1}{2}$	-1	Not included		
	ν_μ	0						
	ν_τ	0						
	e	0.511	-1	$-\frac{1}{2}$	-1	-1	0	-2
	μ	105						
	τ	1777						
Quarks	u	2.2	$+\frac{2}{3}$	$+\frac{1}{2}$	$+\frac{1}{3}$	$+\frac{2}{3}$	0	$+\frac{4}{3}$
	c	1275						
	t	173.10^3						
	d	4.7	$-\frac{1}{3}$	$-\frac{1}{2}$	$+\frac{1}{3}$	$-\frac{1}{3}$	0	$-\frac{2}{3}$
	s	95						
	b	4180						

Table 1.1.: Table giving the fermions masses in the SM and their quantum number with regards to the electromagnetic and electroweak interactions. Masses are taken from [10].

Fermions are also usually categorised into three generations according to their mass as shown by the table 1.1. We can see that neutrinos have a null electrical charge and so they only interact via the weak interaction.

The six quarks: up, down, charm, strange, top and bottom are combined in hybrid particles, called hadrons, and form the observable matter. The combination of quarks to form hadron is explained thanks to the strong interaction. They are chargeless regarding to this interaction. Two types of hadrons have been observed:

- **Baryons:** There are composed of three quarks such as the proton (uud) or neutron (udd).
- **Mesons:** There are composed of only two quarks such as pions $\pi^\pm$ ($u\overline{d}$,$\overline{u}d$) and π^0 ($u\overline{u} - d\overline{d}/\sqrt{2}$).

1.1.1.2. Historical introduction of the theory

The actual version of the Standard Model inherits from several developments made during the 20th century. At the beginning of this century, the Quantum Mechanic as well as the Special Relativity have been developed in order to understand recent observation of the radioactivity phenomena and the dynamic theory of the electromagnetic field proposed by Maxwell. In 1927, Dirac attempted a quantization of the electromagnetic field described classically by the Maxwell's equation. From his work, he obtained the first theory that is consistent with the Quantum Mechanic and Special Relativity principles [11]. In 1934, Fermi proposed his theory of the beta decay, which was the

precursor of the theory describing the weak interaction [12]. In the following years, physicists studied in more details the Dirac's theory and found issues in calculation with infinite values or disagreement with experimental measurement of, for example, the magnetic moment of the electron. In the 40s, Feynman [13], Schwinge [14] and Tomonaga [15] proposed a new version of the Dirac's theory. They have developed a technique called *renormalization*, among other useful tools, in order to suppress infinities in the calculation. Moreover, the comparison with experiment measurement had shown a spectacular agreement. In the coming years, the condition for a quantum field theory to be renormalisable as been determined. Especially, the point-like version of the weak theory introduced by Fermi few year ago has been demonstrated to not be renormalisable. The issue was coming from the point-like description of the interaction, as comparison with QED the interaction is carried by the electromagnetic gauge boson: the photon. Such theories are called *local gauge theory*, the fundamental equation describing the theory *Lagrangian* and particles are treated as fields. This Lagrangian should be invariant under local transformation, described by Lie groups. In particular, the QED corresponds to a gauge theory with the gauge group $U(1)$.

Nowadays the actual form of the Standard Model is described by the direct product of three gauge symmetry groups: $SU(3) \times SU(2) \times U(1)$. The first one described the strong interaction and the two others represent the unification of the weak and the electromagnetic interaction. Then the Standard Model's Lagrangian could be written as a sum of two Lagrangian:

$$\mathcal{L}_{SM} = \mathcal{L}_{QCD} + \mathcal{L}_{EW} \tag{1.1}$$

1.1.1.3. Quantum Chromodynamics sector

The strong interaction is described by the Quantum Chromodynamics (QCD) based on the gauge group $SU(3)$. It acts on quarks which have a *charge color*: red, blue or green. There are eight gauge bosons of this interaction called gluons which also have a color charge. It means then that gluons can interact between them, this imply an interesting property of the interaction called: *asymptotic freedom*. The strength of the interaction, represented by the coupling constant g_s, increases when the system's energy decreases. For this reason, quarks forming hadrons are intensively confined in them.

1.1.1.4. Electroweak sector

As said in the little introduction, the Fermi's theory describing the weak interaction as point-like was not renormalisable, a contrario to the the electromagnetic interaction described by the QED with the gauge group $U(1)$. In the following years, experimental investigation about the weak interaction in β decays has determined the properties of this interaction:

- **Charged bosons:** The β decay ($n \longrightarrow p + e^- + \overline{\nu_e}$) indicates that the gauge boson should be electrically charged.

- **Massive bosons:** The weak interaction is observed at very short range, it implies that this interaction should be mediated by massive bosons.
- **Parity violation:** In 1957, in turns out that the β decay of the ^{60}Co violates the parity symmetry. Suggesting an interaction in the form V-A.

From those observations, Glashow proposed in 1961 a first unification of the electromagnetic and weak interaction based on a larger symmetry group $SU(2)\times U(1)$ allowing to have one boson (γ) responsible of an parity-conserved interaction and three others: W^+, W^- and one additional with no electrical charge Z^0 which are violating parity [16]. Unfortunately, there were no explanation on the origin of elementary particle masses. It is in 1964 that Higgs [17], Englert and Brout [18], proposed the spontaneous breaking symmetry mechanism allowing to give masses to bosons in a gauge theory by the addition of one scalar boson in theory. Besides the ability to have massive bosons, it allows also to give mass to the fermion via the coupling of fermion's field with the Higg's field described by the Yukawa's term. Finally Glashow [19], Salam [20] and Weinberg [21] synthesis all theoretical developments and proposed the actual form of the electroweak Lagrangian:

$$\mathcal{L}_{EW} = \mathcal{L}_{bosons} + \mathcal{L}_{fermions} + \mathcal{L}_{Higgs} + \mathcal{L}_{Yukawa}. \tag{1.2}$$

Finally, the proof of the renormalisability of this theory has been determined by t' Hooft in 1971 [22].

Boson's term $\mathcal{L}_{boson}$

This term treats the interaction of boson's fields and is given by:

$$\mathcal{L}_{boson} = -\frac{1}{4}W_a^{\mu\nu}W^a_{\mu\nu} - \frac{1}{4}B^{\mu\nu}B_{\mu\nu}. \tag{1.3}$$

Where $W_a^{\mu\nu}$ corresponds to a three dimensional ($a = 1,2,3$) vector boson field associated to $SU(2)$ gauge group and $B^{\mu\nu}$ corresponds to $U(1)$ gauge group. Those fields do not represent the gauge boson that we have observed experimentally.

Fermion's term $\mathcal{L}_{fermions}$

The expression of the fermion's term is given by the following equation:

$$\mathcal{L}_{fermions} = \sum_{\text{Generation}} \left(i\left(\overline{\nu_e}\bar{e}\right)_L \gamma_\mu D_L^\mu \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L + i\bar{e}_R\gamma_\mu D_R^\mu e_R + i(\bar{u}\bar{d})_L\gamma_\mu D_L^\mu \begin{pmatrix} u \\ d \end{pmatrix}_L + i\sum_{u,d}\bar{q}_R\gamma_\mu D_R^\mu q_R \right). \tag{1.4}$$

We have represented the Lagrangian only for the first Generation of particles, it has to be summed over the three families composing the model. Where:

- **1st term:** Interaction of left-handed fermions.
- **2nd term:** Interaction of right-handed fermions.
- **3rd term:** Interaction of left-handed quarks.
- **4th term:** Interaction of right-handed quarks.

We can see a difference between the fermion's singlet and quark's singlet. Indeed as right-handed neutrino has never been observed, there were not included in the model. And the two corresponding covariant derivatives:

$$\begin{aligned} D_R^\mu &= \partial^\mu + i g_1 \tfrac{Y}{2} B^\mu \\ D_L^\mu &= \partial^\mu + i g_2 \tfrac{\sigma_a}{2} W^{a\mu} + i g_1 \tfrac{Y}{2} B^\mu. \end{aligned} \tag{1.5}$$

Where, g_1 and g_2 are the coupling constants to B^μ and $W^{a\mu}$ boson fields respectively. Y is the weak hypercharge and $\frac{\sigma_a}{2}$ is the three components of the weak isospin, described by the Pauli's matrices σ_a. We can already see that the electroweak interaction treats differently particles according to their helicity: left or right-handed. Indeed right-handed field are only coupled to the B^μ field.

Higgs' term $\mathcal{L}_{Higgs}$

This term is responsible of the spontaneous breaking of the symmetry mechanism allowing to explain the mass of particles. It introduces two complex scalar fields:

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}. \tag{1.6}$$

The Higgs' term should be invariant under $SU(2)$ local transformation as well as keep the theory renormalisable, for those two reasons it has the following expression:

$$\mathcal{L}_{Higgs} = (D_L^\mu \Phi)^\dagger (D_{\mu L} \Phi) - V(\Phi). \tag{1.7}$$

Where the Higgs' potential $V(\phi)$ is:

$$V(\Phi) = -\mu^2 \Phi^\dagger \Phi + \frac{\lambda}{4} (\Phi^\dagger \Phi)^2. \tag{1.8}$$

The potential is shown on the figure 1.1. We will see in the following sections how this breaks the $SU(2) \times U(1)$ symmetry in order to obtain the three massive gauge bosons of the weak interaction and the photon of the electromagnetic interaction. This spontaneous symmetry breaking happens when the Higgs' potential on figure 1.1 goes to one minima.

Yukawa's term $\mathcal{L}_{Yukawa}$

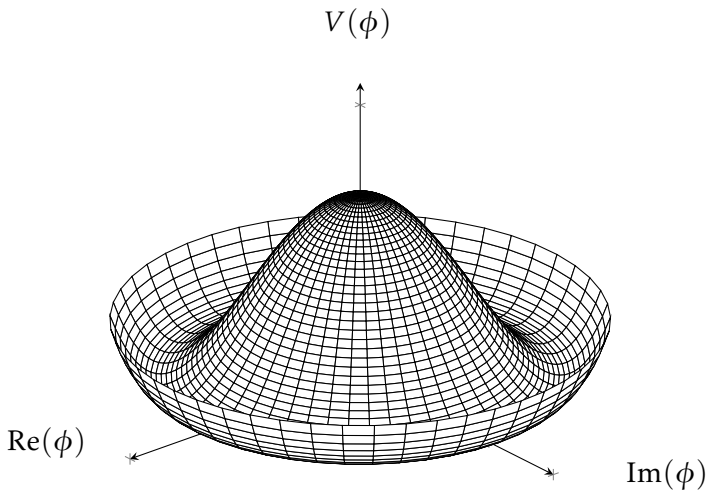


Figure 1.1.: The Higgs' potential with a Mexican's hat form.

This term couples the fermion's field to the Higgs' one in order to give them masses. This term is given by:

$$\begin{aligned}\mathcal{L}_{Yukawa} = &- \sum_{\text{familly}} \left(g_u \left((\bar{u}\bar{d})_L \tilde{\Phi} u_R + \overline{u_R} \tilde{\Phi}^\dagger \begin{pmatrix} u \\ d \end{pmatrix}_L \right) + g_d \left((\bar{u}\bar{d})_L \Phi d_R + \overline{d_R} \Phi^\dagger \begin{pmatrix} u \\ d \end{pmatrix}_L \right) \right) \\ &- \sum_{\text{familly}} g_e \left((\overline{\nu_e}\bar{e})_L \Phi e_R + \bar{e_R} \Phi^\dagger \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L \right).\end{aligned} \tag{1.9}$$

Where we introduce: g_u, g_d and g_e the coupling constant between fermion's and the Higgs' field. We also have $\tilde{\Phi} = i\sigma^2\Phi^*$.

1.1.1.5. Number of parameters

Several parameters have been introduced in the previous section that are not predicted by theory. In fact, the Standard Model has 18 free parameters that should be determined experimentally. Those parameters are:

- The three coupling constants g_1, g_2 and g_s.
- The nine fermion's coupling constant to the Higgs' field, for the first Generation: g_u, g_d and g_e.
- The two parameters introduce to describe the Higgs' potential λ and μ.

- The four parameters describing the Cabibbo-Kobayashi-Maskawa (CKM) matrix allowing quarks of various flavor to mix. This will be introduced in the following section.

1.1.2. Flavour physics

1.1.2.1. Mass generation

In section 1.1.1.4, we have introduced the Lagrangian describing the Electroweak interaction before spontaneous symmetry breaking. As said, this happens when the Higgs' potential value is minimal. There are an infinite possible solutions with no consequence on Physics due to symmetry. For this reason, we can chose one which simplifies equations:

$$\langle \Phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}, \tag{1.10}$$

where $v = \frac{2\mu}{\sqrt{\lambda}}$ and is called the vacuum expectation value. Thus the Higgs' term involving boson fields, $(D_L^{\mu}\Phi)^{\dagger}(D_{\mu L}\Phi)$ becomes:

$$\begin{aligned} \left(D_L^{\mu}\Phi\right)^{\dagger}\left(D_{\mu L}\Phi\right) &= \left| \left(\partial_{\mu} + i g_2 \frac{\sigma_a}{2} W^{a\mu} + i g_1 \frac{Y}{2} B^{\mu}\right) \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \right|^2 \\ &= \frac{v^2}{8}\left[g_2 \left(\left(W^{1\mu}\right)^2 + \left(W^{2\mu}\right)^2 \right) + \left(g_2 W^{3\mu} - g_1 B^{\mu} \right)^2 \right]. \end{aligned} \tag{1.11}$$

Here we can defined new boson fields as a linear combination of the three W^{μ} components and B^{μ}:

- Two charged vector bosons:

$$W^{\mu\pm} = \frac{1}{\sqrt{2}}(W^{1\mu} \mp i W^{2\mu}) \tag{1.12}$$

- Two neutral gauge bosons:

$$\begin{aligned} Z^{\mu} &= \frac{1}{\sqrt{g_2^2 + g_1^2}} \left(g_2 W^{3\mu} - g_1 B^{\mu} \right) \\ A^{\mu} &= \frac{1}{\sqrt{g_2^2 + g_1^2}} \left(g_2 W^{3\mu} + g_1 B^{\mu} \right) \end{aligned} \tag{1.13}$$

The relationship between the Z^{μ}, A^{μ} and $W^{3\mu}$, B^{μ} could be also parametrised by a rotation matrix with the mixing angle θ_W called the Weinberg angle.

Using those definitions, one can write the equation 1.11 such as:

$$\left(D_L^{\mu}\Phi\right)^{\dagger}\left(D_{\mu L}\Phi\right) = \frac{v^2}{8}\left(2g_2^2 W_{\mu}^{+}W^{-\mu} + \left(g_1^2+g_2^2\right)Z_{\mu}Z^{\mu} + \frac{g_2g_1 - g_1g_2}{\sqrt{g_1^2+g_2^2}}A_{\mu}A^{\mu}\right), \tag{1.14}$$

from this equation we can identify mass terms of new boson fields, where:

- $m_W = \frac{g_2 v}{2}$
- $m_Z = \frac{v}{2}\sqrt{g_1^2+g_2^2}$
- $m_{\gamma} = 0$

Thus thanks to the Higgs' mechanism that spontaneously breaks the symmetry $SU(2)\times U(1)$ into $U(1)$ it results three massive bosons, known as $W^{\pm}$ and Z^0 corresponding to the weak interaction and one massless boson γ carrying the electromagnetic interaction.

The same approach could be applied to the Yukawa's term to obtain fermion's mass term, if we evaluate the equation 1.9 with:

$$\Phi = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v \end{pmatrix}, \text{ and } \tilde{\Phi} = \frac{1}{\sqrt{2}}\begin{pmatrix} v \\ 0 \end{pmatrix} \tag{1.15}$$

we obtain:

$$\mathcal{L}_{Yukawa} = -\sum_{\text{familly}} \frac{g_e v}{\sqrt{2}}\left(\bar{e}_L e_R + \bar{e}_R e_L\right) + \frac{g_u v}{\sqrt{2}}\left(\bar{u}_L u_R + \bar{u}_R u_L\right) + \frac{g_d v}{\sqrt{2}}\left(\bar{d}_L d_R + \bar{d}_R d_L\right). \tag{1.16}$$

We can see mass terms appearing for each fermions in the theory, except for neutrino. The mass of a fermion i is then given by: $\frac{g_i v}{\sqrt{2}}$.

1.1.2.2. Charged and neutral interactions

In the previous subsection we have derived the expression of physical bosons mediating the weak interaction as a function of the gauge boson before the spontaneous symmetry breaking. Using those expressions we can also rewrite the $\mathcal{L}_{fermions}$ term as a function of those physical bosons. The weak interaction is expressed as two different currents depending on the boson involved in the interaction:

- **Charged current interaction:** Those interactions are the ones involving the $W^{\pm}$ charged bosons, the equation 1.4 could then be expressed as:

$$\sum_{\text{Generation}} -\frac{g_2}{\sqrt{2}}\left(\bar{\nu}_e\gamma^{\mu}\frac{1-\gamma^5}{2}eW^{\mu+} + \bar{e}\gamma^{\mu}\frac{1-\gamma^5}{2}\nu_e W^{\mu-} + \bar{u}\gamma^{\mu}\frac{1-\gamma^5}{2}dW^{\mu+} + \bar{d}\gamma^{\mu}\frac{1-\gamma^5}{2}uW^{\mu-}\right). \tag{1.17}$$

- **Neutral current interaction:** Those ones correspond to interaction involving the Z^0 neutral boson, we obtain:

$$\sum_f -\frac{1}{2\sqrt{g_1^2+g_2^2}}\bar{f}\gamma_\mu\frac{g_{Vf}-g_{Af}\gamma^5}{2}fZ^\mu \tag{1.18}$$

We can see the projector $\frac{1-\gamma^5}{2}$ which allows to project a Dirac filed onto its left-handed components leading to a parity maximally violated in charged current interaction but not for neutral ones. Indeed depending on the values of g_{Vf} (vectorial coupling) and g_{Af} (axial coupling) the neutral current treats differently right and left-handed fermions' components but does not only consider the left-handed one. Those two couplings could be expressed as a function of the two coupling constants g_1 and g_2.

1.1.2.3. CKM matrix and CP violation

At this point we can see that the Standard Model and especially the weak interaction allows only interaction between fermions of the same generation. For example the π^+ $(u\bar{d})\longrightarrow \nu_\mu+\mu^+$ decay is allowed via the W^+ but not the K^+ $(u\bar{s})\longrightarrow \nu_\mu+\mu^+$ decay because it would require to have a term in the Lagrangian involving quarks of the first and second families. However this decay has been observed and the theory can not handle this observation. The last piece of the actual Standard Model puzzle is the following. The idea is to consider that, for quarks (d,s,b), their weak eigenstates involved in the weak interaction differ from their mass eigenstates. The transformation allowing to compute a vector expressed in a given basis into another one is obtained thanks to a unitary matrix. In this case we need a 3-dimensional matrix, called the CKM (Cabibbo [23], Kobayashi, Maskawa [24]) matrix:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}\begin{pmatrix} d \\ s \\ b \end{pmatrix}, \tag{1.19}$$

where: (d',s',b') are the weak eigenstates and (d,s,b) the mass eigenstates. This matrix represents the transition probability from a quark i into a quark j, and this probability is proportional to $|V_{ij}|^2$. Another representation of such matrices is by a product of three 3x3 rotation matrices describes by: three mixing angle and one complex phase. If the complex phase is different from zero, it introduces interaction violating the CP symmetry. All elements of this matrix has been measured and especially the CP-violation terms which is non-null but not large enough to explain the matter-antimatter asymmetry observed in the Universe.

Finally, the charged current weak interaction involving quarks must take this effect into account. For example, in the third term of equation 1.17 the d field become a superposition of d,s,b fields:

$$d' = V_{ud}d + V_{us}s + V_{ub}b, \tag{1.20}$$

allowing now to have interactions involving quarks of different families.

1.1.3. Limitations

The development of the Standard Model has been a successful story during the whole 20th century. Both theoretical and experimental progresses have allowed to built an almost consistent theory to describe the fundamental structure of the matter. The highlight of this story is probably the prediction of the Higgs boson existence and its discovery in 2012 at the CERN by the CMS [25] and ATLAS [26] collaborations, predicted few years ago. Nevertheless, it remains questions that the theory is not answering, we will quickly discuss the ones that are directly or indirectly related to neutrino physics. Thus we will not talk about: the hierarchy problem, the strong CP problem, the unification of all interactions, etc...

Mass hierarchy As we have seen the spontaneous symmetry breaking mechanism allows to give mass to fermions and bosons. Contrarily to bosons for which their masses are derived via the two coupling constants g_1, g_2 and the vacuum expectation value v, fermion mass' values are not constraint by the theory. Moreover, we will see that the neutrino is actually a massive particle with a mass of the order of few eV. The difference between neutrino mass and the top quark, about nine order of magnitudes, is intriguing.

Lepton and quarks flavour mixing Physicists have proven that neutrinos are massive particles because they observed the ability of them to oscillate from one flavour to another one. This is called the neutrino oscillation phenomena and it is described in the same way as the flavour mixing in the quark sector. But the two matrices, in the quark and the neutrino sector, are very different in terms of values. As opposite to quark flavour mixing which is happening at very short distance (within the atom), neutrino oscillation occurs sometimes at few hundred kilometres.

CP violation The Standard Model could not explain actually the fact that we observe today a Universe composed of matter. Indeed most of cosmological theories assume that at the beginning of the Universe there were the same amount of matter and antimatter. The *Baryogenesis*, is a hypothetical process that should have happened at the early stage of the Universe in order to explain this asymmetry. Conditions to observe such asymmetry are called the Sakharov conditions [27]. Among the three, one of the condition is: the C and CP violation. As we know, the weak interaction maximally violates the C symmetry and the complex phase of the CKM matrix introduces CP violations. However, the CP violation observed in the quark sector is too low. A new way of obtaining matter-antimatter asymmetry is via the *Leptogenesis* process. It requires an extension of the Standard Model by the introduction of one or several heavy right-handed neutrinos to the model. Via the so-called see-saw mechanism (see section 1.2.4.3), it allows to explain the tiny mass of active neutrinos as well as introducing new source of CP violation.

Thus several limitations of the Standard Model are pointing out that new physics, beyond the Standard Model, may be found by studying rare decays at colliders or neutrino physics.

1.2. Neutrino Physics

1.2.1. Discovery and properties

The history about the neutrino started at beginning of the 20th century with the observation of the β decay and the measurement of the electron's energy spectrum by Chadwick [28]. At this time, it was supposed that this interaction involved only two particles in the final state: an electron and the remaining atom. However, instead of γ and α decays the energy spectrum of the emitted particle was not a narrow peak but a continuous spectrum. In 1930, Pauli proposes, in his famous letter [29], an explanation of the continuous spectrum by introducing a new light (almost massless) chargeless particle: the neutrino. In such case, the β decay becomes a three-body decay and the electron's continuous energy spectrum is explained.

It is only in 1956 that Reines and Cowan [30] discovered the electronic antineutrino produced by a nuclear power plant. They have used the same way of detection as we are using with the SoLid experiment: the Inverse Beta Decay which is:

$$\overline{\nu_e} + p \longrightarrow e^+ + n \tag{1.21}$$

a precise description of this interaction is given in section 2.2.1. In 1958 helicity of neutrino was measured by Goldhaber, Grodzins and Sunyar [31]. They found that neutrino are only left-handed particles, since the fermions' mass is obtained via the combination of both left-handed and right-handed fields there is no way to build neutrino mass term. Leading to introduce the neutrino in the Standard Model as a massless particle.

As we have seen there are three fermion families, the two others flavor of neutrino, coupling with the muon and the tau leptons have been discovered in 1962 by Lederman, Schwartz, Steinberger [32] and in 2000 by the DONUT collaboration [33], respectively. The analysis of the Z^0 at the electron-positron LEP collider allows to measure the number of light active neutrinos (with a mass less than the half of Z^0 one and interacting with it). They reported $N_\nu = 2.9841 \pm 0.0083$ [1].

1.2.2. Neutrino oscillation

1.2.2.1. The anomalies

At the beginning of the sixties, the Homestake experiment aimed to measure the flux of neutrino coming from the sun. Star produces neutrinos via fusion reactions which are described by the Standard

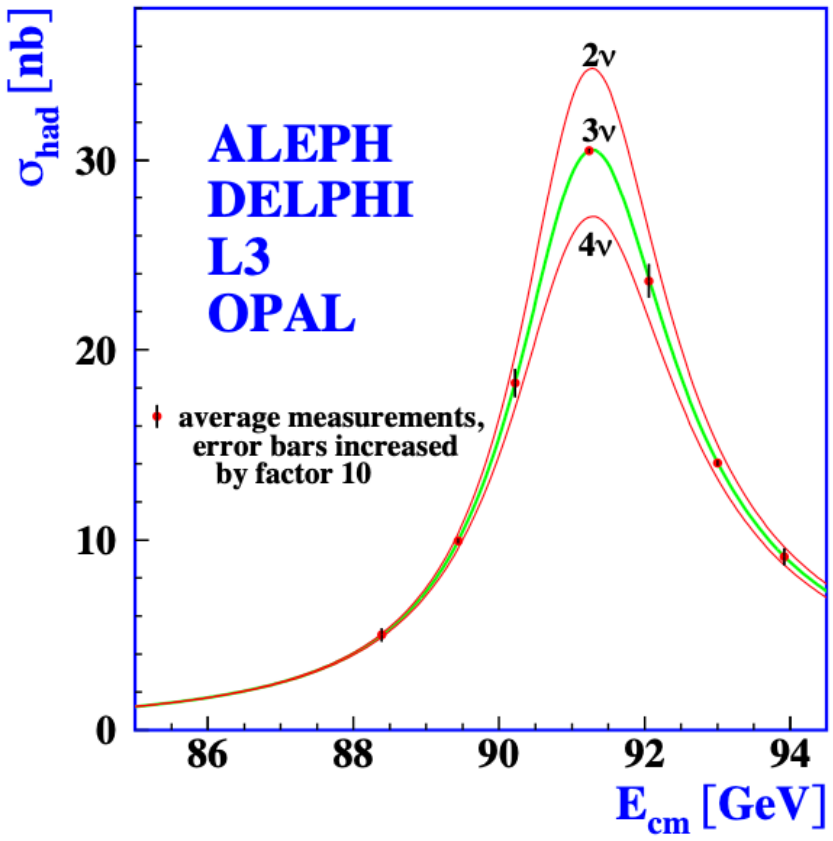


Figure 1.2.: Measurement of the Z^0 width resonance by the four LEP experiments compared with Standard Model prediction in the case of 2, 3 or 4 active neutrinos. [1].

Solar Model [34]. The dominant process generating electronic neutrino is called the pp chain and starts with the following reaction:

$$p + p \longrightarrow d + e^{+} + \nu_e. \tag{1.22}$$

It produces around 91% of solar neutrino. The detection of those neutrinos was performed via the reaction

$$\nu_e + {}^{37}\mathrm{Cl} \longrightarrow {}^{37}\mathrm{Ar} + e^{-}. \tag{1.23}$$

They found in 1968 a 66% deficit between the measured electronic neutrino flux and the one predicted from the Standard Solar Model [35]. This deficit was later confirmed by other experiments and was called the solar anomaly.

In 1988 the Kamiokande experiment observed also a deficit of muonic neutrino produced by cosmic rays by interaction with the atmosphere [36]. This was confirmed later by the Super-Kamiokande experiment. The advantage of those experiments was the ability to reconstruct the direction of the incoming ν_μ, allowing to know if the incoming muonic neutrino have crossed the Earth (downside) before interaction or not (upside). The comparison between the downside and upside fluxes showed a discrepancy with the unity.

The conclusion of the solar anomaly came from the SNO experiment [37]. This experiment, using heavy water as target, has the ability to be sensible of the three different flavours of neutrinos thanks to the following neutral-current reaction

$$\nu + d \longrightarrow n + p + \nu. \tag{1.24}$$

In parallel, the experiment used a charged-current reaction in order to measure only electronic neutrino via the reaction:

$$\nu_e + d \longrightarrow p + p + e^{-}. \tag{1.25}$$

They first confirmed the solar anomaly using the charged-current reaction and found that the flux derived from neutral-current reaction was consistent with the predicted flux from Standard Solar Model. This was a confirmation that electronic neutrinos emitted by the sun have been detected after propagation in the two other flavours: muonic and tauic. The Super-Kamiokande and SNO experiments obtained a Nobel price in 2015 for the discovery of neutrino oscillation.

1.2.2.2. Mathematical framework of neutrino oscillation

The mechanism of neutrino oscillation had been described thanks to the contribution of Pontecorvo, Maki, Nagawa and Sakata [38–40]. The same mathematical framework as the CKM matrix has been used for the description of neutrino oscillation. The idea is to consider that the neutrino

weak eigenstates (involved in the weak interaction) are different than the neutrino mass eigenstates (involved during the particle propagation) are different. The link between the vector of weak and mass eigenstates is given by an unitary matrix. In the three neutrinos framework of the Standard Model this matrix is noted U_{PMNS} and assumed to be unitary:

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = U_{PMNS} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix}. \tag{1.26}$$

As we have seen such matrix could be parametrised using four parameters: three mixing angles and one complex CP violating phase. The term angle for the three first parameters comes from the possibility to decompose the matrix into the product of three rotation matrices as follow:

$$U_{PMNS} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{1.27}$$

where: $c_{ij} = \cos(\theta_{ij})$ and $s_{ij} = \sin(\theta_{ij})$, with θ_{ij} the three mixing angles and δ the CP-violating phase.

1.2.2.3. Oscillations probability in the two flavor case

In order to simplify equation we can derive the oscillation probability in the two flavor case. We assume then two flavor eigenstates $|\nu_\alpha\rangle$, $|\nu_\beta\rangle$ and two mass eigenstates $|\nu_1\rangle$, $|\nu_2\rangle$. Then the relationship between the two eigenbases is given by:

$$\begin{pmatrix} \nu_\alpha \\ \nu_\beta \end{pmatrix} = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \end{pmatrix}, \tag{1.28}$$

where θ is the mixing angle parametrising the oscillation. We want to derive the appearance probability of $|\nu_\alpha\rangle$ as a function of its propagation which is given by:

$$P_{\nu_\alpha \to \nu_\beta}(t) = \left| \left\langle \nu_\beta \mid \nu_\alpha(t) \right\rangle \right|^2 = \left| \left\langle \nu_\beta \mid \hat{H} \mid \nu_\alpha \right\rangle \right|^2. \tag{1.29}$$

where $\hat{H}$ is the Hamiltonian operator. In this simple case we consider the Hamiltonian of a free particle: $\hat{H} = \frac{\hbar^2}{2m}\nabla^2$. We can then express $|\nu_\alpha\rangle$ as a function of the mass eigenstates:

$$|\nu_\alpha\rangle = \cos(\theta)|\nu_1\rangle + \sin(\theta)|\nu_2\rangle, \tag{1.30}$$

and replace it in the probability equation which gives:

$$\begin{aligned} P_{\nu_\alpha \to \nu_\beta}(t) &= \left| \left\langle \nu_\beta \mid \hat{H} \mid \nu_\alpha \right\rangle \right|^2 \\ &= \left| \langle \nu_\beta | \hat{H} \cos(\theta) | \nu_1 \rangle + \langle \nu_\beta | \hat{H} \sin(\theta) | \nu_2 \rangle \right|^2 \\ &= \left| \langle \nu_\beta | e^{-iE_1 t} \cos(\theta) | \nu_1 \rangle + \langle \nu_\beta | e^{-iE_2 t} \sin(\theta) | \nu_2 \rangle \right|^2 \\ &= \left| \cos(\theta)\sin(\theta) \left[e^{-iE_2 t} - e^{-iE_1 t} \right] \right|^2 \\ &= \sin^2(2\theta) \sin^2\left(\frac{\Delta E t}{2} \right). \end{aligned} \tag{1.31}$$

Or in the case of the neutrino is a massive particle, it should be tiny and then could be considered as ultrarelativistic. This implies that, in natural unit, the time is proportional to the travelled distance and we have:

$$E_i = \sqrt{p_i^2 + m_i^2} \approx p_i + \frac{m_i^2}{2p_i}. \tag{1.32}$$

and using $p^2 \approx E$, where E is the average energy of mass eigenstates, we obtain the well-known appearance neutrino oscillation probability:

$$P_{\nu_\alpha \to \nu_\beta}(L, E) = \sin^2(2\theta) \sin^2\left(\frac{\Delta m_{12}^2 L}{4E} \right). \tag{1.33}$$

For simplicity we can express this formula in physics units:

$$P_{\nu_\alpha \to \nu_\beta}(L, E) = \sin^2(2\theta) \sin^2\left(1.27 \frac{\Delta m_{12}^2 (\mathrm{eV}^2) L(\mathrm{m})}{E(\mathrm{MeV})} \right), \tag{1.34}$$

and the equation giving the disappearance probability is:

$$P_{\nu_\alpha \to \nu_\alpha}(L, E) = 1 - P_{\nu_\alpha \to \nu_\beta}(L, E) = 1 - \sin^2(2\theta) \sin^2\left(1.27 \frac{\Delta m_{12}^2 (\mathrm{eV}^2) L(\mathrm{m})}{E(\mathrm{MeV})} \right), \tag{1.35}$$

We can see that this probability depends on the travelled distance of the neutrino L, its energy E and the mass-squared difference. The angle mixing θ drives the oscillation amplitude and the mass-squared difference Δm_{12}^2 drives the oscillation frequency. An example of the oscillation probability evolution as a function of $\frac{L}{E}$ is given on figure 1.3. As the oscillation depends on the mass-squared difference of neutrinos, it implies that at least two among three neutrinos are massive particles which is in contradiction with the actual Standard Model.

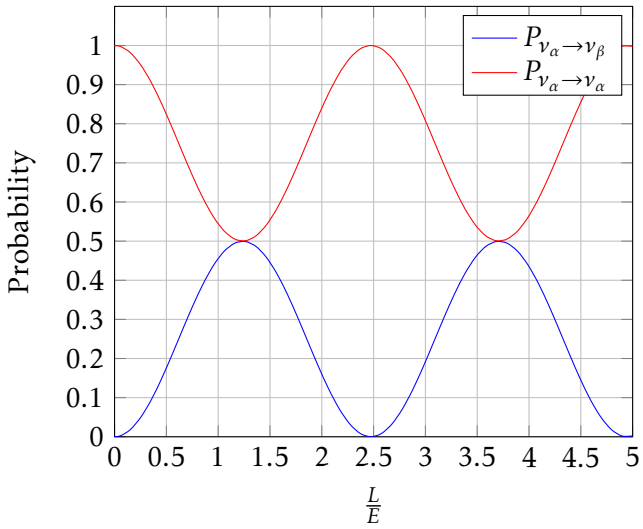


Figure 1.3.: Appearance and disappearance probability in the 2-flavours case as a function of $\frac{L}{E}$ assuming $\theta = \frac{\pi}{8}$ and $\Delta m^2_{12} = 1eV^2$. The oscillation frequency depends on Δm^2_{12} which drives also the position of the maximum amplitude $\left(\frac{L}{E}\right)^{max} = \frac{\pi}{2\times 1,27\times \Delta m^2}$. Whereas the mixing angle θ drives the amplitude.

1.2.2.4. Generalisation: three flavor case

This relation could be generalised to the three flavor case, for this we use the following parametrisation of the PMNS matrix:

$$U_{PMNS} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{pmatrix}. \tag{1.36}$$

The appearance probability given in equation 1.33 for two flavors, becomes:

$$\begin{aligned} P_{\nu_\alpha \to \nu_\beta}(L,E) = \delta_{\alpha\beta} - 4\sum_{k=2}^{3}\sum_{j=1}^{k-1} \mathrm{Re}\left[U^*_{\alpha k}U_{\beta k}U_{\alpha j}U^*_{\beta j}\right]\sin^2\left(\frac{\Delta m^2_{kj}L}{4E}\right) \\ + 2\sum_{k=2}^{3}\sum_{j=1}^{k-1} \mathrm{Im}\left[U^*_{\alpha k}U_{\beta k}U_{\alpha j}U^*_{\beta j}\right]\sin^2\left(\frac{\Delta m^2_{kj}L}{4E}\right) \end{aligned} \tag{1.37}$$

and as in the two flavors case we can derive the disappearance probability given by:

$$P_{\nu_\alpha \to \nu_\alpha}(L,E) = 1 - 4\sum_{k=2}^{3}\sum_{j=1}^{k-1} |U_{\alpha k}|^2 \left|U_{\alpha j}\right|^2 \sin^2\left(\frac{\Delta m^2_{kj}L}{4E}\right) \tag{1.38}$$

It has to be noted that those equations are correct only in the case of neutrino is propagating in vacuum. If not then the Hamiltonian describing the neutrino propagation has an additional term taking into

account the charged and neutral current interaction of it on electron or atom in the middle: for taking into account medium effects.

From those equations we can also see that in order to be sensitive to a particular oscillation parameter, the experimental setup is constraint by:

- Neutrino flavor and energy emitted by the source.
- Neutrino flavor detected by the experiment.
- Distance between the source and the experiment.

1.2.2.5. Neutrino mass hierarchy

We have seen that the measurement of neutrino oscillation is sensible to only the squared-mass difference of neutrinos. This implies that the absolute determination of neutrino masses is not possible via oscillation experiments. Rigorously, we should consider three squared-mass differences: Δm^2_{12}, Δm^2_{23} and Δm^2_{13}. Actually measurements, presented in the following subsection, have shown that the parameter Δm^2_{12} is lower than the two others by two order of magnitudes. So we can consider that:

$$\Delta m^2_{23} \approx \Delta m^2_{13}. \tag{1.39}$$

Using this approximation, it remains one open question about the mass hierarchy. Indeed, we can consider two mass ordering:

- **Normal Ordering (NO):** $m_1 < m_2 << m_3$
- **Inverted Ordering (IO):** $m_3 << m_1 < m_2$

those two different definitions have a direct impact on the determination of the phase CP violation of the PMNS matrix.

1.2.3. State of the art

1.2.3.1. Solar sector: Measurement of θ_{12} and Δm^2_{12}

The so-called solar sector defines the mixing angle and mass-squared difference that dominates at very long baseline. The first measurement of those parameters have been performed by experiments looking at neutrino produced by the Sun such as: Homestake, SAGE, GALLEX, Kamiokande. Long distance reactor experiments such as Kamland are also sensitive to those parameters. Their measurement is shown on the figure 1.4, we can see the clear oscillation pattern appearing.

The future JUNO (Jiangmen Underground Neutrino Observatory) is expecting to perform the measurement of θ_{12} and Δm^2_{12} with a precision below one percent.

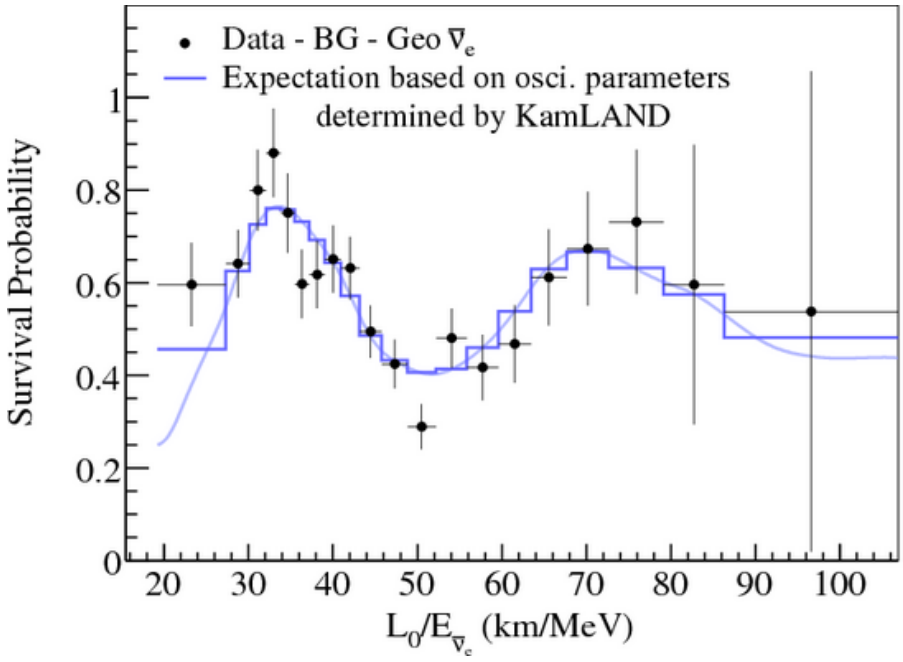


Figure 1.4.: Ratio between the measured $\overline{\nu_e}$ and the prediction assuming the no oscillation hypothesis as a function of L/E from the Kamland experiment. From [2].

1.2.3.2. Atmospheric sector: Measurement of θ_{23} and Δm^2_{23}

The cosmic rays interacting in the atmosphere produce pions and kaons whose the decays produce neutrino and antineutrino into two flavors: electronic and muonic. Atmospheric neutrino has a broad range in energy, from GeV to several TeV, and a high travelled distance, from few km to 10^4km. Again several experiments are using atmospheric neutrino such as: Super-Kamiokande, IceCube or MINOS.

However, they are not the only experiments sensitive to the atmospheric sector oscillation parameters. Neutrinos could be also produced using particle accelerators, the idea is to use accelerator facility to produce pions or kaons by collision on a target. The undesired particles are stopped thanks to a beam dump and soil. Such experimental setup allows to have a better control of the source than for atmospheric neutrino. We can cite those experiments: MINOS, OPERA, T2K or NOvA. They are also sensible to the θ_{13} mixing angle and to the CP violation phase.

1.2.3.3. Reactor sector: Measurement of θ_{13}

The θ_{13} mixing angle was the last one to be determined. The interest on its measurement was enhanced by its impact on possible CP violation in the leptonic sector. Indeed a zero value of this parameter would imply no possible CP violation. The first hint of a non-zero value of θ_{13} was coming from the T2K experiment. This was confirmed in 2012 by the Double Chooz, Daya Bay and RENO experiments. Those experiments are using the same detection technology with a target loaded with liquid scintillator doped with Gadolinium in order to increase neutron detection produced by the inverse beta decay reaction.They all showed that θ_{13} has a small and non-zero value.

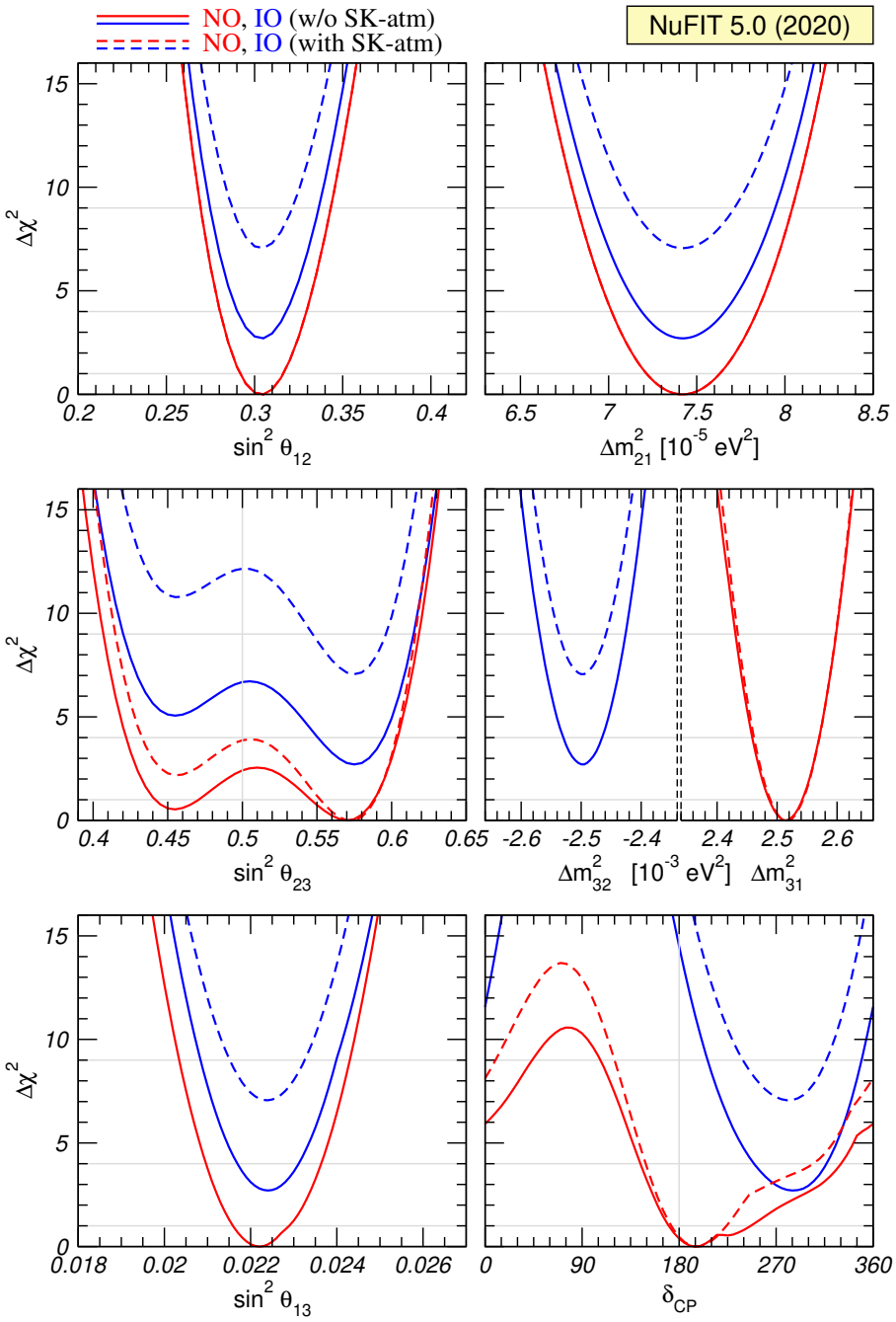


Figure 1.5.: Global fit considering the 3ν oscillation framework. This figure shows the $\Delta\chi^2$ profiles in the case of normal or inverted mass ordering. From [3].

1.2.3.4. Global fits

The actual estimation of neutrino oscillation parameters is given by the NuFit global fit [3]. They are combining measurement of several neutrino oscillation experiments in order to obtain the best measurement of all oscillation parameters. The projection of $\Delta\chi^2$ of the six relevant neutrino oscillation parameters is given on the figure 1.5. The values and the 1σ errors associated are given on the table 1.2. In this global fit the Normal Ordering is favoured at 2.7σ.

	Normal Ordering	Inverted Ordering
$\sin^2(\theta_{12})$	$0.304^{+0.012}_{-0.012}$	$0.304^{+0.013}_{-0.012}$
θ_{12} (°)	$33.44^{+0.77}_{-0.74}$	$33.45^{+0.78}_{-0.75}$
$\sin^2(\theta_{23})$	$0.573^{+0.016}_{-0.020}$	$0.575^{+0.016}_{-0.019}$
θ_{23} (°)	$49.2^{+0.9}_{-1.2}$	$49.3^{+0.9}_{-1.1}$
$\sin^2(\theta_{13})$	$0.02219^{+0.00062}_{-0.00063}$	$0.02238^{+0.00063}_{-0.00062}$
θ_{13} (°)	$8.57^{+0.12}_{-0.12}$	$8.60^{+0.12}_{-0.12}$
δ_{CP} (°)	197^{+27}_{-24}	282^{+26}_{-30}
$\frac{\Delta m^2_{12}}{10^{-5}\text{eV}^2}$	$7.42^{+0.21}_{-0.20}$	$7.42^{+0.21}_{-0.20}$
$\frac{\Delta m^2_{3l}}{10^{-3}\text{eV}^2}$	$+2.517^{+0.026}_{-0.028}$	$-2.498^{+0.028}_{-0.028}$

Table 1.2.: Best fit values of oscillation parameters using the 3ν framework. The error corresponds to 1σ. The convention adopted for Δm^2_{3l} is the following: $\Delta m^2_{3l} \equiv \Delta m^2_{31} > 0$ for Normal Ordering and $\Delta m^2_{3l} \equiv \Delta m^2_{32} < 0$ for Inverted Ordering. [3]

1.2.4. Nature of the neutrino

Neutrinos are massive particles but even if we extend the Standard Model with the PMNS matrix it is still unknown how to give them mass within the Standard Model framework. Indeed, we have never observed right-handed neutrinos which implies that neutrino never couples with the Higgs as other fermions. This question the nature of the neutrino particle: is it a Dirac or a Majorana particle?

1.2.4.1. Dirac mass

This is the most straightforward way to give a mass to neutrinos. The idea is to use the same mechanism as other fermions: via the Yukawa coupling. This extension is called the "minimal extended Standard Model" in which right-handed neutrinos are added to the model. Right-handed neutrinos have unique properties in the theory: they do not interact at all. Indeed the three known neutrinos (called active), only interact via weak interaction but the important property of this interaction is to maximally violates the parity. For this reason we called right-handed neutrino *sterile neutrino*. With this new field one can build neutrino mass term such as:

$$\mathcal{L}^D_{mass} = -m^D_\nu \left(\bar{\nu}_R \nu_L + \bar{\nu}_L \nu_R\right). \tag{1.40}$$

It requires then to have new free parameters of theory corresponding to the Yukawa coupling constant of the three active neutrinos.

1.2.4.2. Majorana mass

Contrarily to other fermions, the neutrinos are the only one with a null electric charge. In this case, we may consider that the neutrino and antineutrino are the same particles. In 1937, Majorana tried to express a right handed field from a left-handed one in order to build massive term [41]. He found that it is possible if the so-called Majorana condition is respected, which is:

$$\psi_R = \xi C \overline{\psi_L}^T, \tag{1.41}$$

where ξ is an arbitrary phase factor, C is the charge conjugate operator. This condition implies that we have

$$\psi^C = (\psi_L + \psi_R)^C = (\psi_L + C\overline{\psi_L}^T)^C = \psi, \tag{1.42}$$

and so that a Majorana particle is its own antiparticle. Under this condition, one can build the following Majorana mass term

$$\mathcal{L}^M_{mass} = -m^M_\nu \left(\overline{\nu^C_L} \nu_L + \overline{\nu_L} \nu^C_L \right). \tag{1.43}$$

As neutrino has a leptonic number of +1 and antineutrino -1, such reaction if it exists violates the conservation of the leptonic number.

1.2.4.3. Seesaw mechanism

The seesaw mechanism relies on the existence of right-handed neutrino field as well as the Majorana nature of neutrino. Under those assumptions one can write new Dirac term:

$$\mathcal{L}^{D+M}_{mass} = -\frac{1}{2} (\bar{\nu}^c_L, \bar{\nu}_R) \begin{pmatrix} m^M_L & m^D \\ m^D & m^M_R \end{pmatrix} \begin{pmatrix} \nu_L \\ \nu^c_R \end{pmatrix}, \tag{1.44}$$

developing this equation one can have both Dirac and Majorana mass terms in the Lagrangian. The mass matrix in the previous equation could be diagonalised and the two eigenvalues are:

$$m_{2,1} = \frac{m^M_L + m^M_R}{2} \pm \sqrt{\left(\frac{m^M_L - m^M_R}{2} \right)^2 + (m^D)^2}. \tag{1.45}$$

If we assume that $m^M_L = 0$ and $m^M_R \gg m^D$, we obtain:

$$\begin{aligned} m_1 &\simeq m^M_R \\ m_2 &\simeq \frac{\left(m^D\right)^2}{m^M_R}. \end{aligned} \tag{1.46}$$

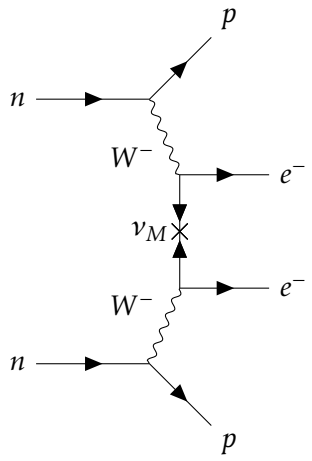


Figure 1.6.: Feynman diagram of the neutrinoless double beta decay.

Then if the right-handed has a huge mass compared to active neutrino we can generate small Dirac mass term for the active neutrino. This mechanism is a crucial component of leptongenesis class model which is a way to explain the matter-antimatter observed nowadays.

1.2.4.4. Neutrino mass and nature experiments

Double beta decay experiments An important experimental effort has been made in order to measure the nature of the neutrino. Such experiments are looking for a specific reaction called the neutrinoless double beta decay ($0\nu\beta\beta$), the Feynman diagram is shown on figure 1.6. The observation of this reaction would imply that the neutrino is a Majorana particle or not.

Experimentally this reaction is highlighted by looking at the sum of energy of the two electrons is equal to the Q-value of the reaction. Otherwise, there are in the final state two electronic antineutrinos carrying a part of the total available energy and thus the sum of the two electrons energy is lower than the Q-value. There is currently no evidence of this reaction. Experiments are putting limits on its half-life.

We can quote dedicated double beta decay experiments: GERDA, CUORE, CUPID, NEMO-3 and its upgrade superNEMO with the superNEMO Demonstrator Module under construction at the Modane Underground Laboratory. Some oscillation neutrino experiments are also currently upgrading their detector in order to perform double beta decay research such as: KamLAND or SNO.

Absolute mass experiment One important contribution in the recent years on the neutrino physics was the first measurement by the KATRIN experiment of effective neutrino mass. They are using the tritium decay and performing a precise measurement of electron's energy in the end-point region.

This measurement is sensible to the following effective neutrino mass:

$$m_\nu^2 = \sum_i |U_{ei}|^2 m_i^2, \tag{1.47}$$

a non-zero value of m_ν^2 implies a shift of the tritium endpoint and a distortion of electron energy spectrum. They have released the best limit on this parameter in 2019 with $m_\nu < 1.1eV$(90 % CL) [42].

1.2.5. Keeping testing the Standard Model

In 2017, the COHERENT has performed the first measurement of the Coherent Elastic Neutrino-Nucleus Scattering (CEνNS) [43] corresponding to the reaction:

$$\nu + A \longrightarrow \nu + A. \tag{1.48}$$

This opened a new way to perform precise test on Standard Model interaction by measuring the cross-section of this interaction. Deviations from theory could imply non standard interaction involving new mediators. Moreover it could bring informations on neutrino properties such as its magnetic dipole moment or its charge radii. CEνNS experiments are also sensitive to light sterile neutrino [44]. Several experiments are performing such measurement, the CONUS, TEXONO and CONNIE are actually taking data and the Nu-GEN, RED-100, MINER, RICOCHET and NUCLEUS which are in commissioning or construction.

1.3. Sterile neutrino

1.3.1. Introduction

The right-handed neutrinos that have been introduced in the previous section to explain neutrino mass would have a particular place in the Standard Model. As we have seen, due to their helicity, they will not interact via the weak interaction. There are also no constraint on their mass. There is also no constrain on the number of right-handed neutrino. For those two reasons, they are usually used in order to explain anomalies between the actual Standard Model and the experimental results.

- **eV-scale:** Sterile neutrino with a mass around the eV is a possible explanation of anomalies observed by reactor neutrino experiments. This is detailed in more details in the section 1.3.2.
- **keV-scale:** One of the crucial question of particle physics is the nature of the Dark Matter. One possibility is to consider a sterile neutrino with a mass of the order of the keV.
- **GeV and higher scale:** As we have seen, a class of model called leptogenesis have been developed to explain the baryonic asymmetry in the Universe. Depending on the model, they require one or several sterile neutrino with a mass from the GeV to 10^{15} GeV.

The only possibility to highlight the presence of sterile neutrino is via the oscillation phenomena. Indeed, adding a sterile neutrino would change the dimension of the PMNS matrix and new oscillation parameter would be associated to it. In the following section, we discuss about the anomalies highlighting the possible presence of a sterile neutrino at the eV-scale.

1.3.2. Sterile neutrino at the eV mass scale

Even if a lot of anomalies have been explained via the oscillation mechanism, it remains non-understood experimental results. They all concerned short baseline measurement.

1.3.2.1. The gallium anomaly

We have already cited two experiments involved on the measurement of solar neutrinos: the GALLEX [45] and SAGE [46] experiments. They used the following reaction in order to detect electronic neutrino:

$$\nu_e + {}^{71}\mathrm{Ga} \longrightarrow {}^{71}\mathrm{Ge} + e^{-}. \tag{1.49}$$

Both have decided to use for calibration purpose radioactive sources producing mono-energetic electronic neutrino. They observed a deficit between the measured electronic neutrino flux and the predicted one: $R = 0.84 \pm 0.05$ (see figure 1.7). A new estimation on the ^{71}Ga cross-section with neutrinos has shown that the significance of this deficit goes from 3σ to 2.3σ.

1.3.2.2. The accelerator anomaly

The LSND experiment was developed for studying the oscillation of muonic antineutrino into electronic antineutrino. The detector was positioned at 30m from the target on which accelerated protons interacted to produce neutrinos and antineutrinos. The detector was a cylinder loaded with 167 tons of liquid scintillator and they detected electronic antineutrinos via the inverse beta decay reaction. They have reported an excess of electronic antineutrino with a significance of 3.8σ [47].

Following this observation the MiniBooNE experiment has been built in order to investigate this anomaly. They used a detector of 800 tons of mineral oil located 541m downstream of the target. Even if the baseline is different with regards to LSND, the antineutrino energy is higher thus they have both investigated the same L/E region. MiniBooNE has reported an excess in the neutrino and antineutrino channel, with a significance of 3.4σ and 2.8σ respectively [48].

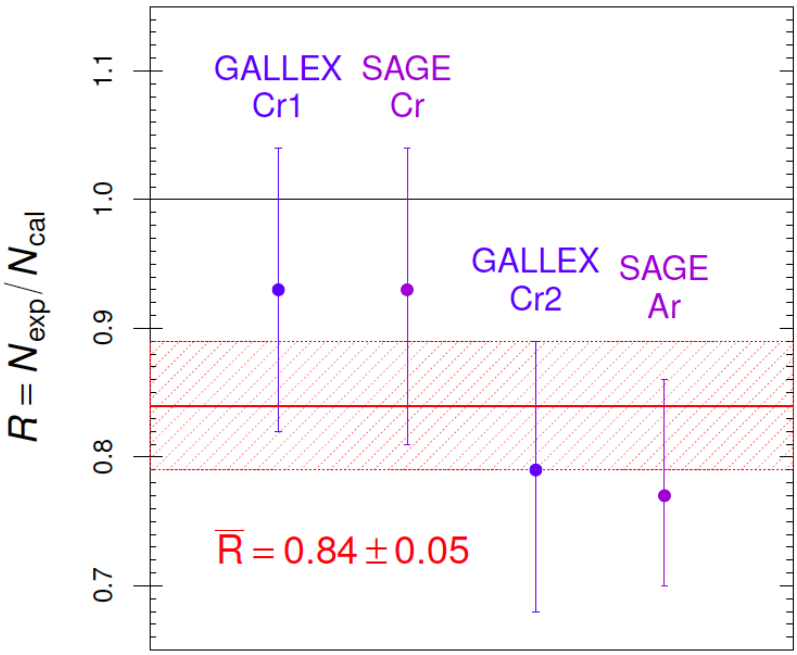


Figure 1.7.: Ratio between the measured and expected number of electronic neutrino for the GALLEX and SAGE experiments. From [4].

1.3.2.3. The reactor anomaly and the distortion at 5 MeV

Several experiments have looked at electronic antineutrino produced by nuclear reactors. After a re-evaluation of the predicted electronic antineutrino flux emitted by nuclear reactors, the comparison with the measured one has shown a deficit of $R = 0.943 \pm 0.023$ with a 3σ significance (see figure 1.8) [49]. This anomaly is known as the Reactor Antineutrino Anomaly (RAA). We can see that this deficit could be interpreted as a fourth neutrino oscillation implying a maximum disappearance at very short baseline: few meters from the reactor.

This is not the only anomaly about the reactor antineutrino. The antineutrino energy spectrum has been firstly computed by reactor experiments. The Double Chooz, RENO and Daya Bay collaborations have together observed a distortion of the measured antineutrino energy spectrum with regards to the prediction. Several explanations are still on discussion, the most considered by the community are:

- An issue on the theoretical estimation of antineutrino flux emitted by nuclear reactors.
- An issue on the experimental reconstruction of antineutrino energy. All those experiments are using a target loaded with liquid scintillator doped with Gadolinium. This kind of target is known to have a non-linear energy response.

This anomaly may have also an impact on the RAA because it relies also on the predictions.

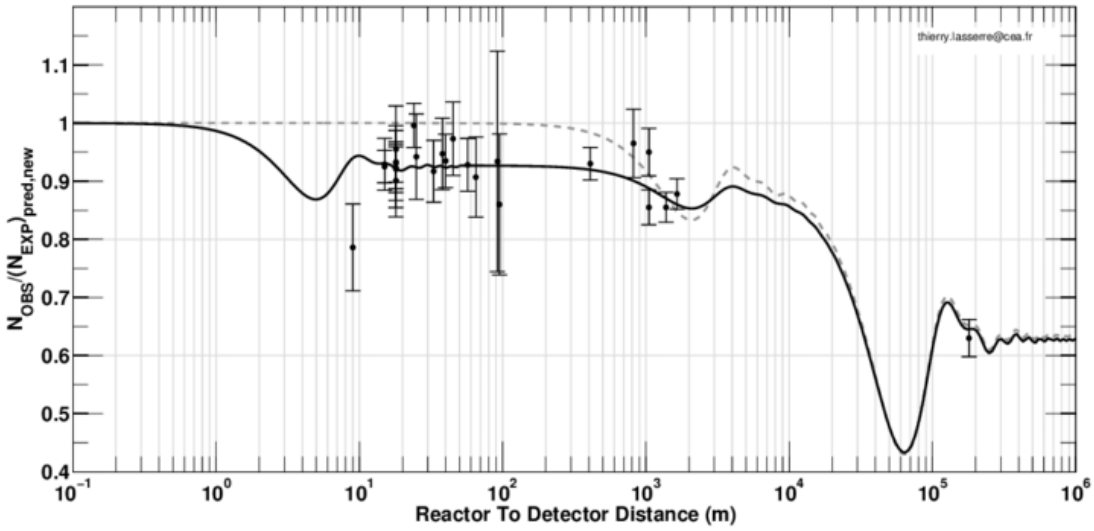


Figure 1.8.: Ratio between the measured and predicted electronic antineutrino flux as a function of baseline experiments after re-estimation of the theoretical predictions. The solid line corresponds to a fit with a fourth neutrinos (three actives plus one sterile) and the dashed line to the actual three neutrino framework. The minimum around 1km corresponds to the atmospheric neutrino experiments and the one around 100km to solar neutrino experiments. From [5].

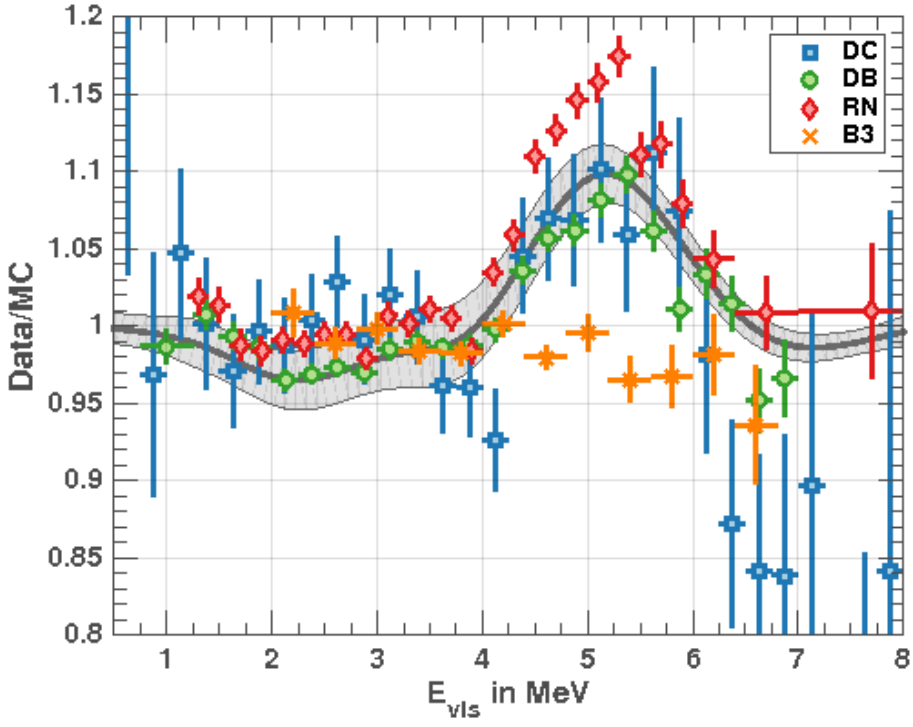


Figure 1.9.: Ratio between the measured and predicted electronic antineutrino energy spectrum. DC: Double Chooz, DB: Daya Bay, RN: RENO, B3: Bugey-3. From [6]

1.3.2.4. A global explanation: Sterile neutrino

All the rates associated anomalies previously presented: Gallium, Accelerator and Reactor could be explained by the addition of (at least) one sterile neutrino. This sterile neutrino would never been observed as it does not interact via weak interaction. But it could actually interfere with known neutrinos via the oscillation phenomena. This requests to extend the PMNS matrix into a 4x4 matrix and adding two other oscillation parameters: θ_{14} and Δm^2_{14}. Global fit has been performed and have shown tensions between Accelerator and the two other datasets.

The previous anomalies constrained the parameter Δm^2_{14} to be around at the eV-scale.This implies that, for reactor antineutrino, the maximal oscillation would take place at very short baseline, few meters. This has motivated a worldwide effort to search for oscillation at very short distance. This is discussed in the following section.

1.3.3. Reactor experiment at very short baseline

Those anomalies conducted to the development of several experiences with different sources, technologies and sensitivity to hypothetical sterile neutrino oscillation. We can distinguish the experiments near a commercial or experimental nuclear reactor which have different properties.

- **Commercial reactor:** Nuclear power plants are built in order to produce electricity. They have a Low Enriched Uranium fuel (LEU) with around 5% ^{235}U. They have a high reactor power, usually the order of few GW_{th}.
- **Experimental reactor:** Those reactors are built for research and fabrication of medical radioisotope purpose. They have a High Enriched Uranium fuel (HEU), > 90% of ^{235}U. The reactor power is lower than nuclear power plant, below 100 MW_{th}.

All experiments are using the same reaction as Reines and Cowan to discover the electronic antineutrino: the Inverse Beta Decay (see equation 1.21). The detection relies on the detection in coincidence of the positron and the neutron. Very short baseline experiments have chosen different technology to detect those two particles. For the positron detection, two kinds of scintillators have been chosen:

- **Liquid scintillator:** Liquid scintillator is a well known technology in reactor neutrino physics. The θ_{13} experiments - Double Chooz, RENO and Daya Bay - are using a detector loaded with a transparent liquid. The constraints of using this technology in a sterile oscillation research is to segment a liquid target as well as deals with non-linearity energy response of the liquid associated to PMTs. However, one strong advantage is the possibility to perform Pulse Shape Discrimination (PSD) allowing to significantly reduce the background.
- **Plastic scintillator:** Plastic scintillator was not usually used by reactor neutrino experiments. The main difference with liquid is the impossibility to perform PSD. However, it has the advantage

Name	Detector				Reactor				Result RAA
	Technology	Segmented	Mass (ton)	Energy Resolution (% at 1 MeV)	Overburden (m.w.e.)	Baseline (m)	Power (MW_{th})	Fuel	
NEOS	LS(Gd)	none	1	5	≈20	23.7	$2.8.10^3$	LEU	Exclude 1.5σ [50]
DANSS	PS(Gd)	2D	1	17	≈50	11 - 13	$3.1.10^3$	LEU	Exclude $> 5\sigma$ [51]
Neutrino-4	LS(Gd)	2D	1.4	-	≈10	6 - 9	100	HEU	Favour 3σ [52]
STEREO	LS(Gd)	1D	1.8	8	≈15	9 - 11	58	HEU	Exclude $> 3\sigma$ [7]
PROSPECT	LS(Li)	2D	3	4.5	≈5	6 - 12	85	HEU	Exclude 2.5σ [8]
SoLid	PS(Li)	3D	1.6	14	≈1	7 - 12	70	HEU	-

Table 1.3.: Table of experiments looking for sterile neutrino oscillation near nuclear reactor and their main properties.

to have a linear energy response reducing uncertainty on the energy reconstruction. It is also easier to machine and so to have segmented target.

For the neutron detection we can distinguish:

- **Lithium:** They are using the capture of neutron on ^{6}Li giving an alpha and a tritium sharing a total energy of 4.78 MeV.
- **Gadolinium:** Liquid scintillator are usually doped with Gadolinium which have a high neutron capture cross-section and result in the emission of several gammas sharing 8 MeV.

The current experiments involved in the very short baseline oscillation research are listed in the table 1.3. The recent results shown by the STEREO, PROSPECT and Neutrino-4 collaborations are in tension. Indeed, the Neutrino-4 has presented a sign at more than 3σ of short baseline oscillation while the two others are excluding a large region of allowed parameters. The current strategy is to perform a joint analysis in order to take advantage of the different energy resolution, baseline and technology to constraint a larger region of the two oscillation parameters with a higher statistical significance. It remains a huge work to combine the data of different experiments.

1.3.3.1. NEOS experiments

The NEOS experiment is based in South Korea near the Hanbit Nuclear Power Complex. This complex is composed of six nuclear power plants with thermal power of $2.6GW_{th}$ each. The RENO experiment is also based near this complex. The detector is positioned at 24m from one of the nuclear reactor.

The detector is a cylinder of $1m^3$ loaded with liquid scintillator doped with Gadolinium. They have a very good energy resolution and S/B. However, the detector is not segmented and thus need electronic antineutrino flux prediction to perform an oscillation search. For this purpose they use the Daya Bay spectra measurement for normalisation.

1.3.3.2. DANSS experiment

The DANSS experiment is located in Russia near the Kalinine power plants. This power plant is composed of four nuclear reactors with a thermal power of $3.1 GW_{\text{th}}$ each. The detector is composed of 2500 one meter long plastic scintillator (PVT) strips with a surface coating loaded with Gadolinium. The detector is on a mobile platform allowing to perform measurement at different baseline: 10.9m, 11.9m and 12.9m. They have a very good S/B but a lower energy resolution degrading their sensibility on sterile neutrino oscillation.

1.3.3.3. Neutrino-4 experiment

The Neutrino-4 experiment takes place in Russia near the SM-3 experimental reactor at the SSC-RIAR nuclear center, Dimitrovgrad. The reactor is using a highly enriched fuel with 95% ^{235}U. It has a thermal power of $100 MW_{\text{th}}$. The detector is composed of 50 cells of $0.225 \times 0.225 \times 0.85 m^3$ loaded with liquid scintillator doped with Gadolinium. The experimental apparatus can be moved allowing them to cover a baseline from 6 to 12 meters.

1.3.3.4. STEREO experiment

The STEREO experiment is located in France near the ILL research reactor, Grenoble. The reactor is also using a highly enriched fuel and is operated at the thermal power of $58 MW_{\text{th}}$. The target of the detector is composed of six cells of 37cm width loaded with liquid scintillator and doped with Gadolinium. They have an energy resolution of 9% at 0.835 MeV and a baseline from 9.4 to 11.1m. They have reached a S/B of 0.9. They released an exclusion contour in [7] disfavouring the RAA best fit with a significance better than 3 σ (see figure 1.10)

1.3.3.5. PROSPECT experiment

The PROSPECT detector is positioned near the High Flux Isotope Reactor (HFIR) in USA. It is operated at a thermal power of $85 MW_{\text{th}}$ with highly enriched fuel. The detector is composed of 154 optically isolated segments of $14.5 \times 14.5 \times 117.6 cm^3$. They are loaded with four tons of liquid Scintillator with 0.08% of 6[Li]. The light collection is ensured by two PMTs per segment positioned at each end. Difference of signal time arrival allows to perform 3D reconstruction of interaction. They have a good

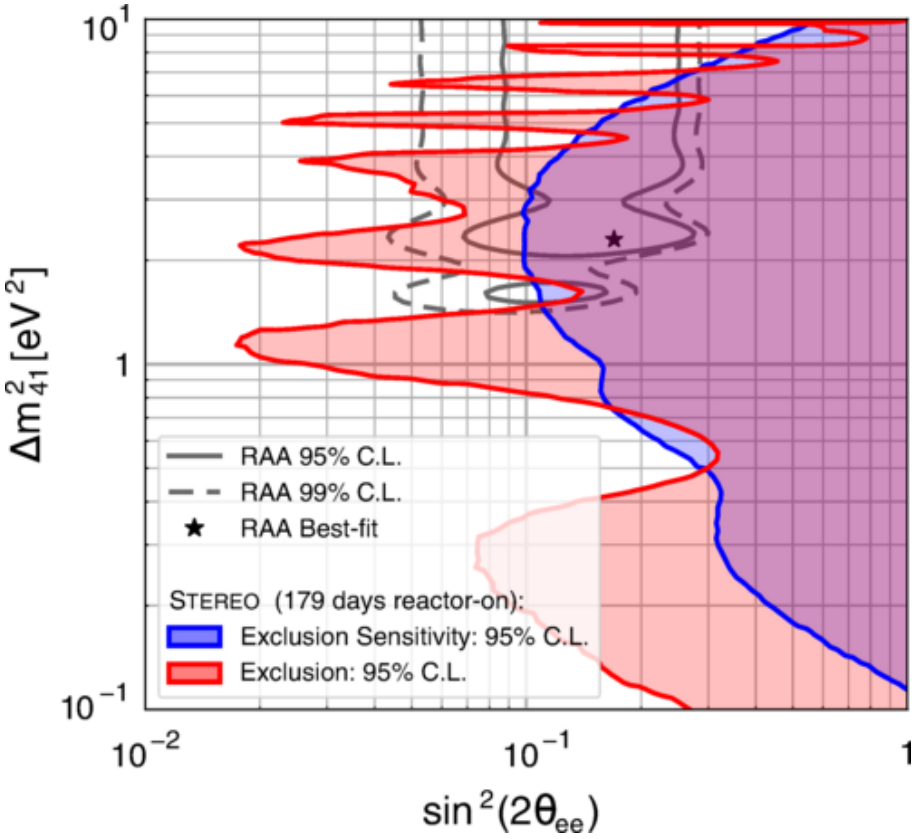


Figure 1.10.: Exclusion contour in red and exclusion sensitivity in blue obtained by the STEREO experiment using 179 days of reactor-on data. The RAA is excluded with a significance better than 3 σ. From [7]

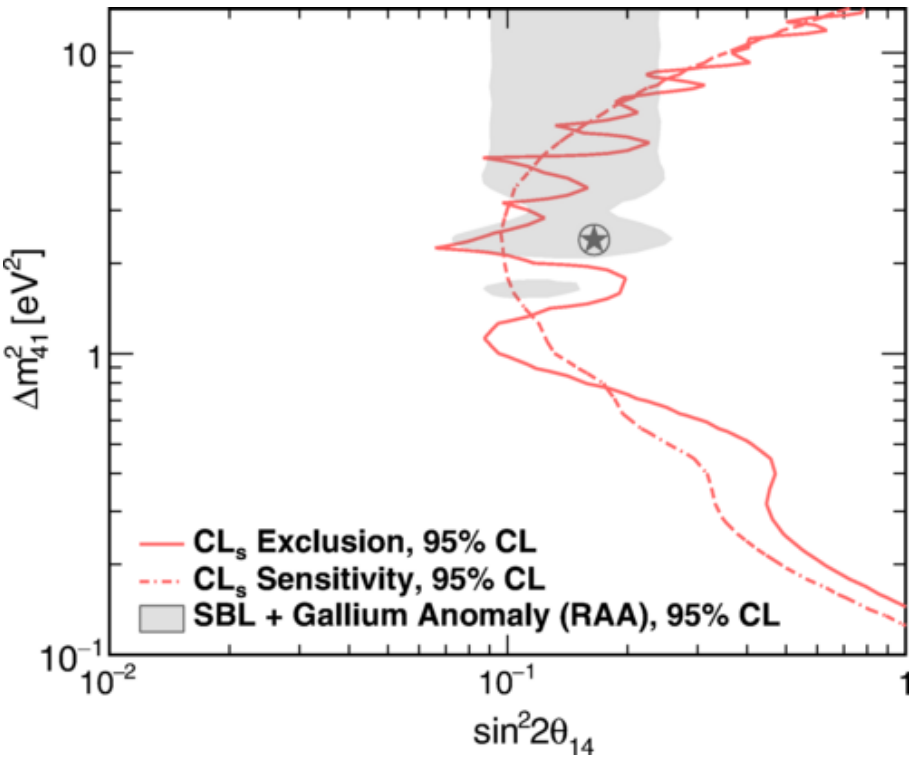


Figure 1.11.: Exclusion contour (full line) and the exclusion sensitivity contour (dashed line) of the PROSPECT experiment using 96 days of reactor-on data. They exclude the RAA best fit with a significance of 2.5 σ. From [8]

energy resolution and regarding to the low overburden reach a good S/B of 1.7. They obtained the exclusion contour shown on figure 1.11 which also exclude the RAA best fit with a significance of 2.5σ.

1.3.4. Further constrains

Another way to constrain sterile neutrino is to study the PMNS matrix under the 3ν framework. The unitary property of this matrix implies mathematical relationship between the $U_{\alpha j}$ parameters. The improvement on oscillation parameters uncertainties would allow to constrain the unitary property of the PMNS matrix. In particular, deviations from this unitarity could question the three flavour framework. Future experiments like JUNO, T2HK or DUNE should be able to perform the measurement on the mass hierarchy, the CP violation phase and reduce the uncertainty on mixing angles. The SuperChooz project (based on LiquidO R&D), located at Chooz, would aim the unitarity test measuring both reactor antineutrinos and solar neutrinos with an exceptional S/B [53].

1.4. Conclusion

The state of research on neutrinos has been presented in this chapter. Over the years, various experiments have confirmed the status of the neutrino as a massive particle with a mixing between its

mass and flavor eigenstates. In recent years, the value of the last unknown mixing angle, θ_{13}, has been measured by reactor experiments (Daya Bay, RENO, and Double Chooz) and accelerator experiments (T2K).

There are still open questions about the value of the δ_{CP} phase, the mass hierarchy, the absolute neutrino mass and the existence of one or more sterile flavors. Regarding the δ_{CP} phase, the large value of θ_{13} will allow future experiments like DUNE to measure it. In the next decade, several experiments (ORCA, JUNO, PINGU...), still under construction, aim to give a measurement of the mass hierarchy at more than 3 sigmas. Concerning the PMNS matrix, we are entering in the precision era of the PMNS parameter measurements with experiments like JUNO, Dune, HK... The sterile eV-scale neutrino quest at short baselines is reaching its end with several experiments like STEREO, Prospect or the SoLid experiment.

Chapter 2.

The SoLid experiment

Hidden Orchestra. *Archipelago*, Tru Thoughts, 2012

Chapter 1 has shown that the energy spectrum and the flux of emitted reactor antineutrinos are not well understood. In particular the flux anomaly leads to the possibility of an additional, neutrino beyond the Standard Model. The community requested new measurements at short baseline to better understand this issue. Some research nuclear reactors use highly enriched fuel in ^{235}U. It gives the opportunity to measure electronic antineutrino spectra coming from a single isotope among the four contributors to the fuel of a commercial reactor. The BR2 reactor at SCK·CEN in Mol is perfectly suited for this kind of measurements and is described in section 2.1.

This measurement represents a challenge due to the high background rate expected at ground level. In order to overcome this difficulty the SoLid collaboration has developed an innovative and highly segmented detector. Section 2.2 is dedicated to the description of the detection principles and of the detector's design with its readout system.

This background has several origins and dominates the event rate recorded by the detector. Its understanding is crucial in order to optimise the selection of the signal and background subtraction. All of this is treated in section 2.3.

The SoLid experiment reconstructs reactor antineutrinos via their interaction in the detector. The interaction mode used is the Inverse bêta decay (IBD), which emits a positron and a neutron. The data recorded by the detector is analysed offline in order to perform low and high level reconstructions. First, reconstruction algorithms group the waveforms arriving in time coincidence into clusters of waveforms corresponding to one physical event. Then each cluster is identified as a delayed-like (case of the neutron), muons or prompt-like event (like the positron). Higher level reconstruction determines the energy of prompt clusters, and the localisation of all clusters with a good precision thanks to the segmentation of the detector. This allows in particular to exploit the topology of positron events by distinguishing deposits due to the positron and those due to annihilation gammas. This is detailed in section 2.4.

Finally, a status of the Collaboration physics analysis activities will be presented in section 2.5. This will explain in particular the performance reached in this domain before we delivered the results of our own analysis work, described in chapter 5.

2.1. The BR2 reactor at SCK·CEN

2.1.1. BR2 reactor

The BR2 reactor (Belgian Reactor 2) is located at the nuclear research centre SCK·CEN in Mol, Belgium. It has a tank-in-pool design with a highly enriched uranium fuel(93.5% ^{235}U). The compactness of the core, of about 20cm diameter, is possible thanks to a unique twisted design (see figure 2.1). It is operated with a thermal power ranging from 40 to 100 MW_{th}. It produces a very high neutron flux, up to 10^{15} n/cm^2/s, and is an intense source of electronic antineutrinos, around $2{\cdot}10^{19}$ $\overline{\nu_e}$/s.

The reactor operates at a nominal power of 65 MW_{th} with a duty cycle between 160 and 210 days per year. On average there are 6 cycles (reactor-on period) of three to four weeks per year. The rest of the time, there is interim maintenance of roughly the same duration (reactor-off period) during which the collaboration performs calibration campaigns, background measurements or detector maintenance.

2.1.2. Reactor simulation

One important ingredient in the sterile oscillation search is the knowledge of the reference antineutrino flux and energy spectrum, evaluated in the absence of oscillation. BR2 is a very versatile reactor with a variable core configuration. Thus it is required to have a dedicated simulation of each cycle. The collaboration calculates the emitted electronic antineutrino spectrum as a function of the fuel loading map and operation history. The spatial fission distribution is also computed to obtain the fraction of emitted antineutrinos going through the detector. The detector geometric acceptance is about 0.11%. The detailed description of the core composition is performed with the code MCNPX [54] or MCNP6 [55]. Then it is coupled with the evolution code CINDER90 [56] in order to have the spatial and time fission rate distributions of all isotopes of interest, ^{235}U, ^{238}U, ^{239}Pu and ^{241}Pu.

The collaboration employs two approaches [57] in order to convert those fission rate distributions into the emitted antineutrino spectrum. The conversion method [58, 59] uses the fission rate as input and the integral β^- spectra measured in the 80's at the ILL reactor in Grenoble, France [60–62]. It was considered for a long time as the reference. The second approach is called the summation method. Instead of using the integral β^- spectra, it takes as input the fission products provided by the code MURE [63]. Then for each individual fission products a β^- spectrum is computed based on measurements or models. Finally, a weighted sum over all contributors is performed to obtain the antineutrino spectra.

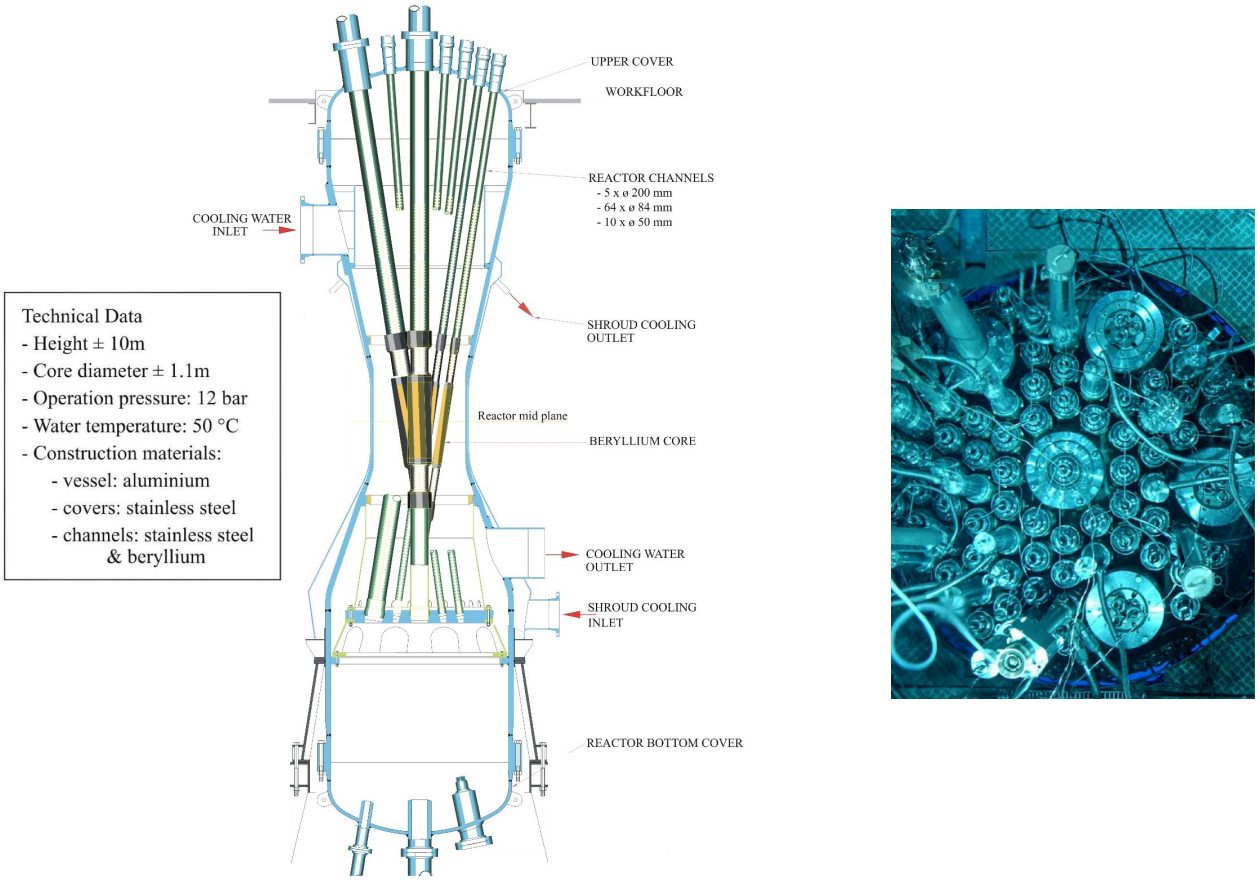


Figure 2.1.: *Left:* Design and technical data of the BR2 reactor core. It is a beryllium matrix composed of 79 hexagonal channels. *Right:* Picture of the top of the reactor.

In the typical conditions of operation - power of 65 MW_{th} - the expected number of IBDs in the detector is between 1000 and 1250 events per day. The difference depends on the core composition, which can differ from a cycle to another one.

2.2. The SoLid detector

2.2.1. Detection principle

Antineutrinos only interact via electroweak processes as seen in chapter 1. There are only two ways for neutrinos to interact: through the exchange of a charged gauge boson ($W^{\pm}$) called the charged current, or through the exchange of the neutral gauge boson (Z^0) called neutral current. In the energy range of reactor antineutrinos, the best interaction to use is the IBD. It has a specific signature in the detector as well as a high cross-section compared to other processes. The equation of this reaction is the following:

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

This reaction has several experimental and theoretical advantages compared to others. The first advantage is that the cross-section is well known, leading to reliable predictions. The second one concerns the signature of the interaction, with the emission of two different particles: a positron and a neutron. Indeed the positron loses its energy quickly via ionisation within the detector. This gives a prompt, high amplitude signal. Instead the neutron, given its neutral charge, loses its energy via multiple scattering on nuclei in the medium before being captured on a nucleus, giving a delayed signal, distinguishable from the positron one. This is called the *thermalisation* process where the neutron's energy will decrease until reaching the thermal equilibrium around 0.025 eV. It is at those energies that the neutron-capture cross-section presents resonances and so increases the probability for the neutron to be captured. The capture leads to the emission of secondary particles, mainly gamma or alpha, allowing to distinguish this event from others. This signature depends on the nuclei used for the neutron capture, such as Hydrogen, Gadolinium or Lithium. For example, the Double Chooz experiment is using Gadolinium and Hydrogen for the neutron capture. It results in the emission of several gammas with a total energy of 8 MeV for the Gadolinium. The capture on Hydrogen is followed by the emission of a single, 2.223 MeV gamma. This difference in energy plays also a role on the analysis, there is no natural radioactivity processes emitting gamma rays of 8 MeV, reducing the amount of background mimicking a neutron capture on Gadolinium.

The time difference between the positron interaction and the neutron capture allows for experiment using the IBD reaction to perform a time coincidence analysis where *prompt* signal, the positron, and *delayed* signal, the neutron capture, are associated. The time between the emission and the capture of the neutron is an important key of antineutrino detection. It mainly depends on the detector geometry and the amount of target available to capture it. For example the Double Chooz detector is composed of a target surrounded by a γ-catcher which differs in the composition of the liquid scintillator. The

target has been doped with Gadolinium nucleus, which has a high neutron capture cross-section. The presence of Gadolinium has decreased the mean neutron capture from 200 μs in the γ-catcher to 30 μs in the target [64]. This coincidence allows to drastically reduce the background as it gives a specific signature of the interaction of antineutrinos inside the detector.

From a kinematic point of view, the IBD reaction has interesting properties. First, it has a threshold depending on the mass of the nucleus the target proton is bound to. The lower threshold is obtained for a free proton, or Hydrogen atom, which therefore constitutes the target of most short baseline experiments. In this case the energy threshold of the IBD reaction is given by

$$E_{\overline{\nu}_e}^{threshold} = \frac{(m_{e^+} + m_n)^2 - m_p^2}{2m_p} = 1.8 \text{ MeV} \tag{2.2}$$

where m_{e^+}, m_n and m_p are the positron, neutron and proton masses, respectively. Above this energy, there is a relationship between the incoming antineutrino energy and the positron one:

$$E_{\overline{\nu}_e} = E_{e^+} + (m_n - m_p) \tag{2.3}$$

So the measurement of the positron's energy gives a direct estimator of the antineutrino one.

The detection of prompt and delayed events in SoLid is performed by combining two different scintillators. The PolyVinyl-Toluene (PVT) acts as hydrogenous target for antineutrino detection. It is a plastic scintillator with a linear response as a function of the deposited energy. Chapter 3 is dedicated to the energy calibration of the SoLid detector and the linearity response has been tested. Furthermore, a plastic allows to have the highly segmented target necessary to reject backgrounds based on topology discrimination with regards to the signal. Thus the fundamental element of the detector is a PVT cube. The neutron is captured via the breakup reaction on ^{6}Li:

$$n + {}^6Li \longrightarrow {}^3H + \alpha + 4.78 \text{ MeV} \tag{2.4}$$

Two sheets per cube of ZnS(Ag) inorganic scintillator doped with ^{6}Li allows the detection of this reaction. The discrimination between waveforms resulting from deposits in PVT or ZnS is based on differences in the scintillation decay times between the two scintillators. It equals 2.1 ns for the PVT and ranges from 10 to 100μs for the ZnS. This variation in ZnS scintillation time depends on the type of particles and their energies. Indeed, for electrons or positrons in the MeV range, the energy deposit per unit length is rather low and populates only states with a lower excitation energy, leading to a short decay time, almost comparable to that in the PVT. Contrarily, neutron capture on ^{6}Li results in highly ionisation paricles (α and ^3H) and thus a large population of higher excited states, leading to a longer decay time [65]. Therefore, as can be seen on figure 2.2, searching for an IBD is searching for a pair of two coincident signals : a prompt signal in the form a sharp and high waveform due to PVT scintillation, found a few dozens microseconds before a delayed signal coming in the form of a lower and longer waveform due to the ZnS scintillation. These two kind of signals are also often referred to ES (electron-like scintillation) and NS (nuclear-like scintillation) signals.

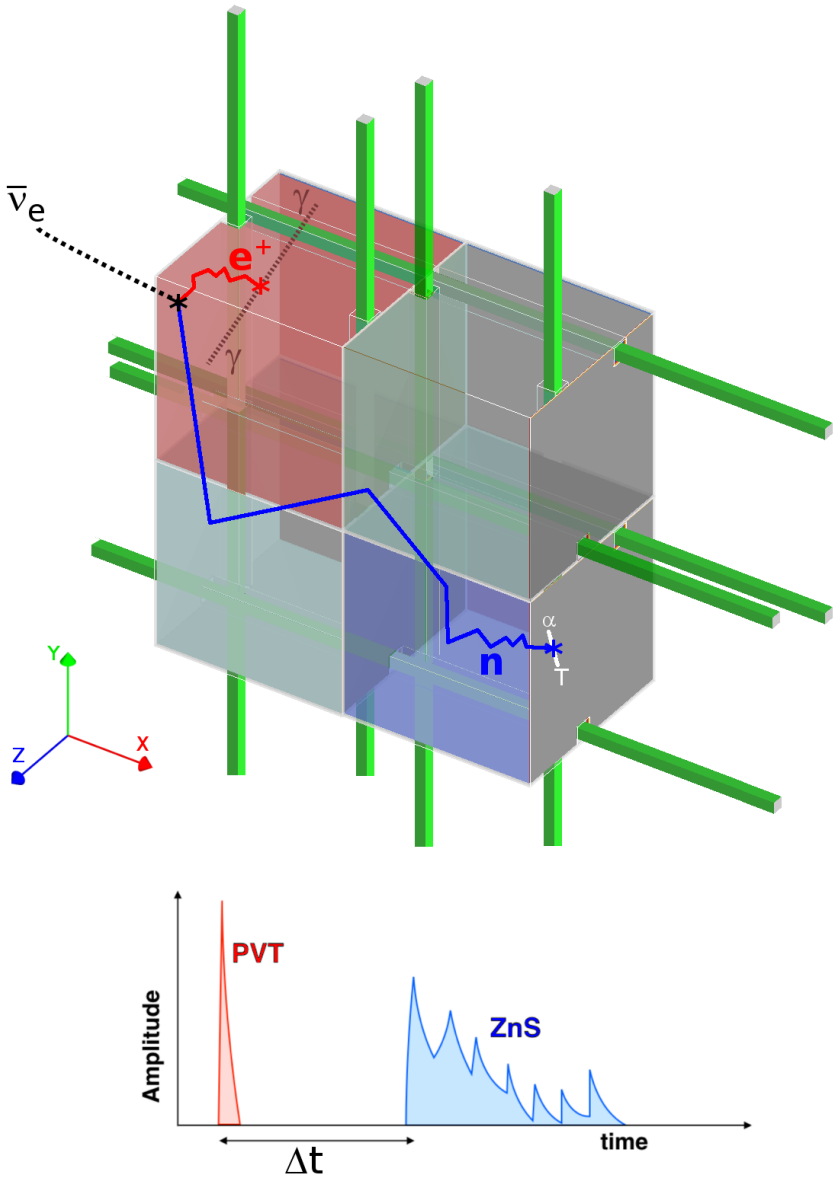


Figure 2.2.: Illustration of an IBD events occurring in the SoLid detector. The red parts show the prompt events which is given by the interaction of the positron plus the two annihilation gammas. The blue parts concern the delayed events which is the thermalisation of the neutron followed by its capture on a ^{6}Li atom. The prompt and delayed events are identified thanks to different waveform shape.

2.2.2. Mechanical design

2.2.2.1. Detection cell

The cube is the fundamental brick of the detector design. It is made of PVT from ELJEN technology [66] with a dimension of 5x5x5 cm^3. In PVT a positron of 10 MeV energy travels less than 48mm thus for a large majority of events, the positron's energy is fully deposited in the cube of interaction. Scintillation photons produced by the PVT are extracted with four optical fibres with a square cross section of 3x3mm^2. Four grooves of 5x5 mm^2 are machined on four faces of each cube in order to accommodate them. On two faces, there is a sheet of ^{6}LiF :ZnS(Ag) from SCINTACOR [67] of 250 μm thickness. One is positioned perpendicularly to the detector-reactor axis (Z-axis) and the other one is perpendicular to X-axis. To finish, cubes are almost optically isolated with a Tyvek wrapping [68]. However the holes to position fibres may lead to the possibility for scintillation to escape and be detected by the first neighbouring channels. Figure 2.3 shows a schematic of the detection cell.

The light produced in the detector is extracted thanks to double-clad wavelength shifting fibres of type BCF-91A from St.Gobain [69]. They are 92cm long with one end placed against a mirror made of a Mylar foil with a reflective aluminium coating to increase the light yield. The optical contact is ensured by a drop of optical gel. On the other end, they are coupled to multi-pixel photons counter (MPPC) of type S12572-050P from Hamamatsu [70]. Each MPPC contains 3,600 pixels arranged on a 3x3mm^2 matrix.

2.2.2.2. Plane and module

The cubes are arranged into detection planes of 16x16 cubes, readout by 64 optical fibres and 64 MPPCs as shown on figure 2.4. They are surrounded by a lining of white high-density polyethylene, 46 mm thick. Its role is to reduce the amount of neutrons escaping the detector. Each detection plane is coupled to its individual electronic box, plugged on one side of the plane, which contain the front-end electronics boards. Finally, each plane is covered by two square Tyvek sheets of that covers on each side the full surface of the plane to ensure the optical isolation.

A module is composed of ten planes and the detector includes a total of 5 modules, meaning 12,800 detection cells and 3,200 MPPCs and optical fibres. In average, a detection plane is composed of 30.634 ± 0.018 kg of PVT and 0.956 ± 0.015 kg of ^{6}LiF :ZnS(Ag). So in total the detector is composed of 1531.721 kg of PVT and 47.766 kg of ^{6}LiF :ZnS(Ag), leading to a fiducial volume of 1.6 tons. These figures come from a very precise measurement of each detector element made by the collaboration during the construction phase.

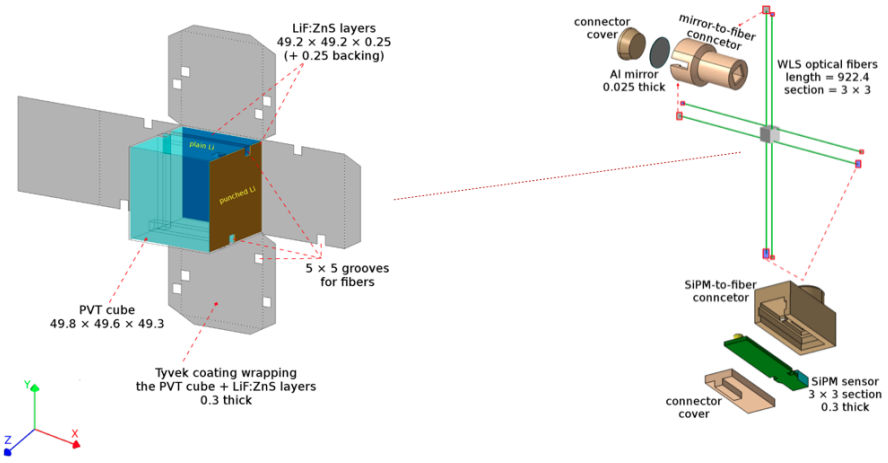


Figure 2.3.: *Left:* Scheme of the SoLid detection cell. The two ^{6}LiF :ZnS(Ag) screens are placed on two faces of the PVT cube. The whole is optically isolated with a Tyvek wrapping. There are grooves in order to position optical fibre and readout the scintillation light. *Right:* Scheme of the light collection with the four optical fibres crossing a cube. At one end of each fibre, there is a mirror and on the other end the SiPM.

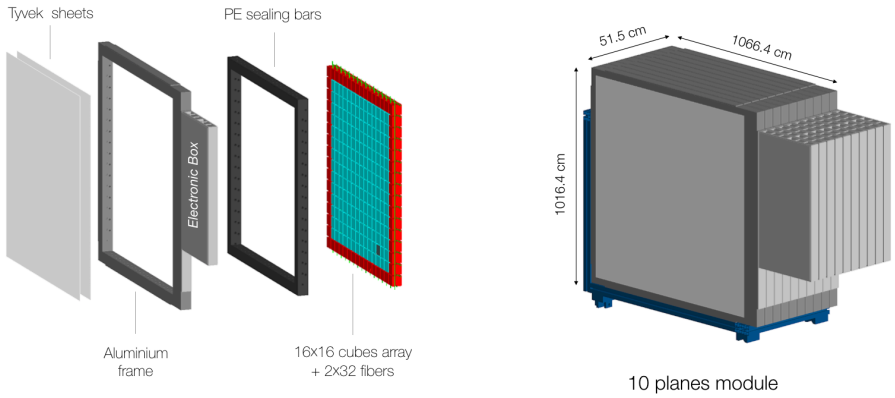


Figure 2.4.: *Left:* Exploded schematic of a full detection plane. The detection volume is surrounded by the HDPE neutron reflector, an aluminium frame also holding the front-end electronics board reading out the plane and two Tyvek sheets for light isolation. *Right:* Schematic of a complete module with 10 detection planes.

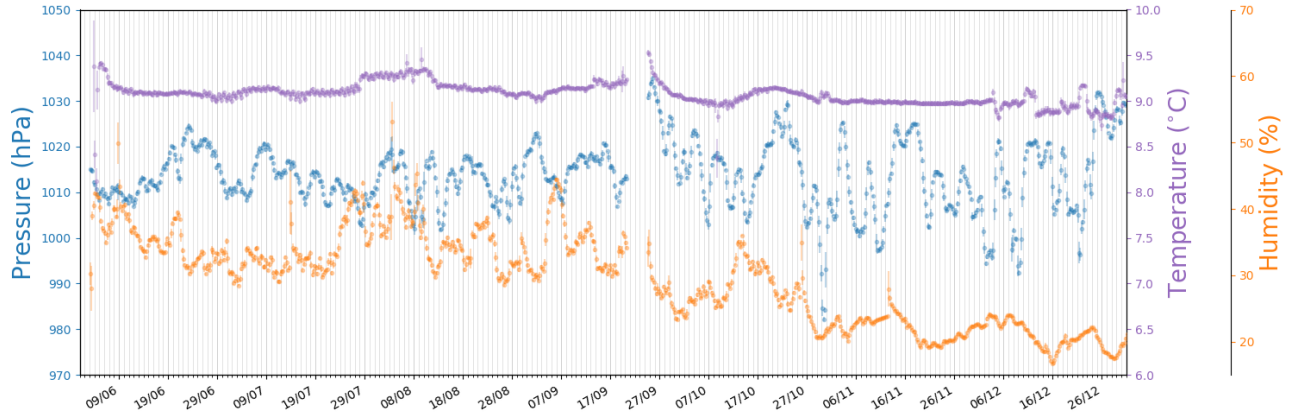


Figure 2.5.: Time evolution of environmental parameters: temperature, humidity and pressure, measured by sensors inside the container, between June and December 2018.

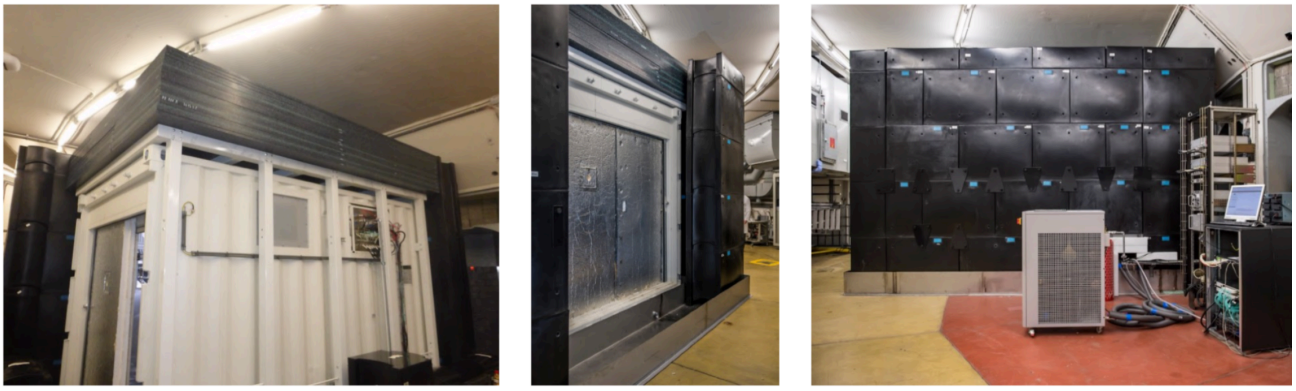


Figure 2.6.: Photographs of the container inside the BR2 containment building. The passive shielding composed of PE layers on the top and a water brick wall surrounding the container.

2.2.2.3. Container

The detector is installed in a cargo container of 2.4x2.6x3.8m^3 as shown in the figure 2.6 . The container is cooled down with a chiller system. During physics runs the temperature is maintained at 11 degrees Celsius so as to reduce the electronic noise from MPPC. The relative humidity in the container is controlled thanks to a permanent flushing with dry air. It ensures a safe and stable environment to prevent PVT ageing and it helps to remove possible Rn emanations from concrete. Environmental parameters such as temperature, pressure and humidity are monitored with a network of a custom sensor controlled by a Raspberry-Pi device. An example of time evolution of those parameters is shown in figure 2.5.

The radioactive background is monitored thanks to two devices located inside the container. A standard PMT coupled to a NaI scintillator allowing a periodic measurement of the gamma background. And a radon detector allows the monitoring of the airborne concentration.

A passive shielding surrounds the container to protect the detector from cosmic induced and reactor backgrounds. On the top of the container, there is a 50 cm thick PE layer, made of 2.5 cm thick slabs that are staggered to avoid gaps. On each side of the detector, there is a water wall made of 50 cm thick bricks (see figure 2.6). It reduces by a factor 10 the cosmic neutron flux in the 1 to 20 MeV energy range, by reducing the energy of incoming neutron via moderation process. The container is positioned at level 3 of the BR2 containment building. This gives an overburden above the detector which is composed of three concrete floors and by the steel roof of the containment building, corresponding to 8 metres-water equivalent.

2.2.2.4. Quality assurance and Calibration system

During the construction of the detector, a quality assurance has been performed to check the response of each detection cells in terms of light collection and neutron efficiency. For that purpose, a specific *ex-situ* system has been designed, called Calispo. This system allows to position automatically a calibration source in front of each cube of a plane. Each plane has been scanned using a ^{22}Na gamma source and ^{252}Cf or AmBe neutron sources. From those measurements, bad fibre-to-SiPM optical couplings have been identified before the integration inside the container and fixed. For more details see reference [71]. It was also a way to start the development of in-situ calibration methods and tools, and to test them.

The calibration of the detector is a crucial task of the analysis. Indeed it is required to control and understand the energy response of the detector in order to provide an accurate measurement of neutrino's energy and tune the simulation. Also the neutron capture efficiency has to be understood. For that purpose a calibration robot, called CROSS, is positioned on the top of the detector. A precise description of the system is given in chapter 4 since the integration of its geometry into the simulation was one of my tasks. It allows to position the calibration sources inside the detector to perform *in-situ* calibration. More details are given in the chapter 3 dedicated to the energy calibration of the detector.

2.2.3. Channel characterisation

The collaboration has chosen to use MPPC (or SiPM) in order to detect the scintillation photons produced inside the detector. A MPPC is composed of an array of semiconductors with a p-n junction. When a scintillation photon creates an electron-hole pair in one of the p-n junctions, an avalanche process is provoked, multiplying the number of carriers. This avalanche is obtained by applying a sufficiently high electric field: the primary charge carrier is accelerated until it could produce secondary charge carrier via impact ionization process. In such case, the MPPC is operated in Geiger mode. The needed voltage to operate a MPPC in this mode is called the *breakdown* voltage V_{br}. Generally, they are operated with a voltage V_{op} higher than V_{br} in order to increase the detection efficiency but it also increases the electronic noise. The difference $\Delta V = V_{op} - V_{br}$ is called the overvoltage. During this process no other incoming scintillation photon can induce another avalanche until the cell recover.

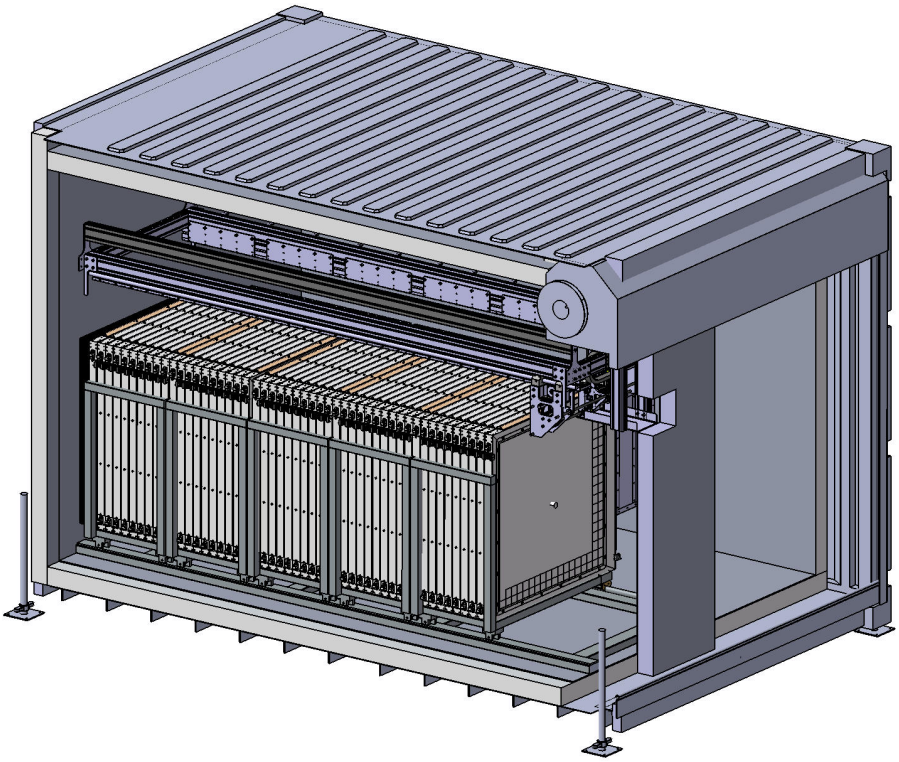


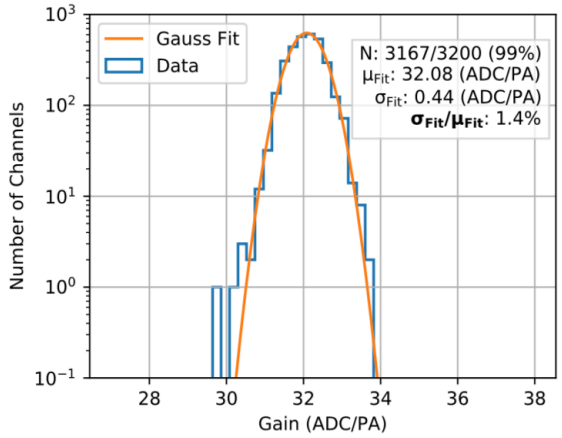
Figure 2.7.: Sketch of the detector inside the container. The CROSS calibration robot is positioned at the top of the detector.

It creates a measurable analogue signal which is then read and digitalised by an ADC converter (see section 2.2.4). The unit usually used to express the digitalised signal is the ADC. We convert the amplitude of the current pulse measured in ADC into a number of Photo-Avalanches (PA). One PA corresponds to the detection of a single scintillation photon. The relationship between ADC and PA is given by the gain due to the applied voltage and will be characterised in the following subsection.

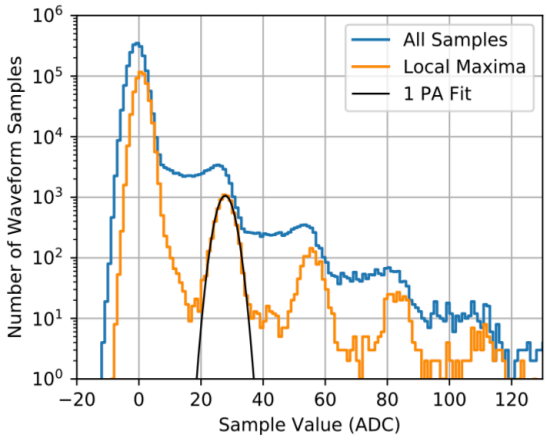
The channel characterisation is important in order to have a homogeneous response of the 3,200 channels. Several parameters have to be determined such as the baseline and the gain in order to correctly convert the ADC into PAs. Moreover MPPCs are subject to electronic backgrounds such as dark counts and crosstalks. Those parameters are determined using unbiased data taken with a specific trigger during physics runs. The different triggers algorithms are described in the section 2.2.4.2.

2.2.3.1. Gain equalisation

The gain is determined by looking at the amplitude spectrum produced by the thermal noise. An example is shown in the figure 2.8b. On this figure we can see peaks corresponding to an integer number of PAs. The first one, centred around 0, gives a measurement of the baseline (already subtracted on this figure). The gain is defined as the difference between the mean value of two adjacent peaks. For each channel of the detector the gain is measured every 6 hours. The distribution of the



(a) Amplitude spectrum of one channel obtained with the periodic trigger. The first peak represents the baseline and others the first, second and third PAs. A Gaussian fit allows access the gain value of the channel.



(b) Gain distribution of all channels, a good homogeneity is observed, with a 1.4% of dispersion.

Figure 2.8.: Amplitude spectrum of one channel obtained with the periodic trigger and gain distribution of all channels of the experiments.

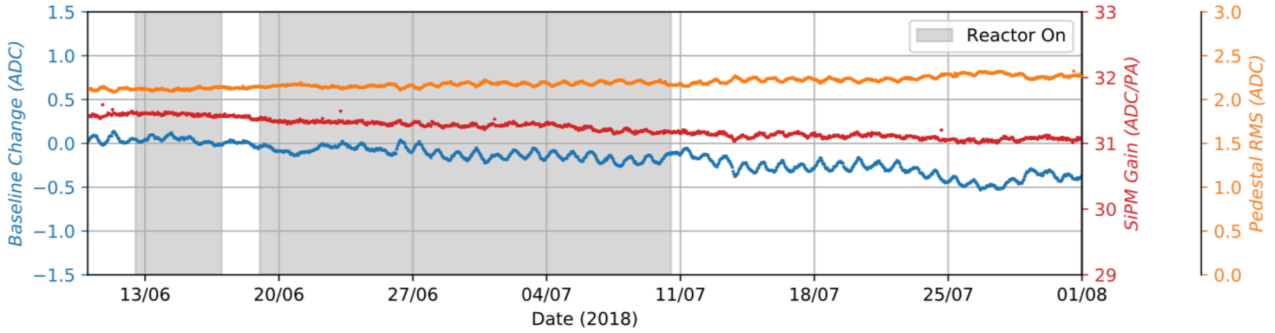


Figure 2.9.: Time evolution of the baseline average and dispersion and gains between June 2018 and August 2018.

gains is shown on figure 2.8a. A dispersion of 1.4% is observed, demonstrating the good homogeneity of response of the detector.

Those values are determined and stored in a database every six hours in order to be used by the reconstruction software during data processing. Figure 2.9 shows the time evolution of the average gain and baseline measurements over the whole detector. A day-night variation is observed due to temperature variations of around 0.5 degrees Celsius. A small increase of the temperature over the period increases the dark court rate resulting in a shift of the baseline of -0.5 ADC.

2.2.3.2. Dark count rate and crosstalk probability

MPPCs are subject to electronic noise, two are relevant for SoLid: Dark counts and crosstalks. Those two noises depends on the temperature and overvoltage.

A dark count is a Photo-Avalanche triggered by the thermal noise. Like the gain, it is monitored with the non-zero-suppressed data acquired with the periodic trigger (see section 2.2.4.2). The dark count rate is defined as:

$$R_{DC} = \frac{1}{t} \int_{0.5PA}^{\infty} f(A)dA \tag{2.5}$$

Where t is the acquisition time, f is the amplitude distribution of the channel as shown in figure 2.8b and A is the amplitude.

A MPPC is composed of an array of pixels. When a PA occurs in a cell, it can trigger another avalanche in a neighbouring cell. This is called the crosstalk effect and the probability to induce another PA is noted μ_{CT}. As the dark count rate is around 110 kHz in SoLid, the probability is small (though not negligible) to have a pile-up of dark count within the sub-microsecond length typical of waveforms due to one or a few PAs. Thus it can be assumed that the observation of a signal with an amplitude above 1.5 PA is due to a dark count that have induced another PA via cross-talk. Thus the crosstalk probability can be defined as:

$$\mu_{CT} = \frac{\int_{1.5PA}^{\infty} f(A)dA}{\int_{0.5PA}^{1.5PA} f(A)dA} \tag{2.6}$$

The evolution of the dark count rate and the crosstalk probability as a function of the overvoltage is given on figure 2.10. The nominal overvoltage applied on MPPC is 1.8 V, giving a dark count rate of 110 kHz and a crosstalk probability of 20%. This dark count rate is high but its impact on the analysis is reduced by selecting signals with an amplitude above 2.5 PAs. With this threshold the dark count rate per channel is reduced to 2-3 kHz. An excessive crosstalk probability could impact the resolution of the energy reconstruction enough to reduce our ability to resolve oscillation pattern. However, we checked in calibration runs that the nominal cube resolution of 14% at 1 MeV is reached. This is shown in chapter 3.

2.2.4. Readout system

2.2.4.1. Description

The signal readout and trigger are performed at three levels: plane, module and full detector. Each plane has its own system installed in its individual electronic box. All planes of a module are sharing

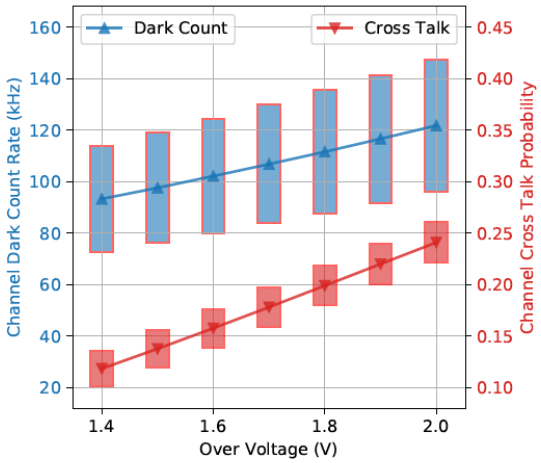


Figure 2.10.: Evolution of the dark count rate and the crosstalk probability of all channels as a function of the overvoltage. The detector is operated with an over-voltage of 1.8 V so the mean dark count rate is 110 kHz and the crosstalk probability is 20%.

a service box containing a DC-DC voltage converter, a clock and distribution board, a network patch panel and a Minnow JTAG programming system. The ten planes are then synchronised thanks to the module clock board. A master clock allows to synchronise the five modules.

The MPPCs of a plane are connected to two 32-channel analogue boards responsible for their alimentation and the signal's pre-amplification. Those two boards are connected to a 64-channel digital board. The signal is sampled with a 40 MHz frequency, corresponding to a sample every 25 ns. Trigger conditions are implemented in a FPGA. When a trigger happens, the information is propagated to neighbouring planes in order to record corresponding data.

2.2.4.2. Triggers and data reduction

There are several triggers and data reduction techniques implemented in each FPGA. The collected data from the Analogue to Digital Converter (ADC) are treated by block of 6.4μs, representing 256 samples of 25 ns.

A first data reduction is performed by applying a Zero-Suppression (ZS) algorithm. It consists to remove samples with an amplitude below a given threshold. The ZS value evolves depending on the type of trigger as well as on the time from it. There are three triggers running in parallel during nominal operation.

Periodic trigger

A periodic trigger records the entire detector during two-time blocks (512 sample or 12.8 μs) without ZS. It gives unbiased data to monitor MPPCs stability. From those data the main MPPC

parameters are derived: the gain, crosstalk and the dark count rate as a function of time. The periodic trigger is operated with a frequency of 1.2 Hz.

Threshold trigger

The second trigger is a threshold trigger aiming to record high amplitude events. The trigger algorithm requires to have one vertical and one horizontal channel of the sample plane with an amplitude simultaneously above 25 PAs (1 MeV). When it happens, all the channels in the plane are readout with a ZS at 1.5 PA. Two time blocks are then read-out (12.8 μs). It allows to reconstruct muons, as muons give an energy deposit of 2 MeV per centimeter in the PVT. Reconstructing muons allows to use them for calibration purposes and to tag and monitor cosmic induced backgrounds.

Neutron trigger

Last but not least is the neutron trigger developped to record IBD events in SoLid. As said in section 2.2.1, signals coming from ZnS are characterised by sporadic pulses emitted over several microseconds. Figure 2.11 shows several examples of waveforms induced by a neutron capture on ^{6}Li. We can see this specific waveform shape signing the capture of a neutron. In particular, on figure 2.11c one can see the difference with an electromagnetic signal happening on one of the four channel (-X) seeing a neutron signal. The interaction in the PVT gives a short and high amplitude signals.

The trigger algorithm takes advantage of this specific neutron capture signal by counting the number of peaks on the waveform in a given time window. The condition requires to have at least 17 peaks in a rolling time window of 6.4 μs. This condition allows a trigger efficiency of 75% with a purity of 20%. This low purity is explained by the contamination of muons events interacting in the ZnS. The high energy density released by the muons in the ZnS scintillator will mimic waveform resulting of the interaction of an alpha and a tritium and then trigger the neutron condition. Those events are identified as muon and rejected offline. The neutron trigger performance could be improved by an upgrade of the firmware and so reduce the amount of data acquired. When this trigger happens, the plane where it happened plus three other planes on either sides of it are readout for a long time window, 500 μs before and 200 μs after the trigger. This large time window allows to record also the prompt candidate associated with the neutron that caused the trigger, to form the IBD candidate from their coincidence. This time window does not affect the IBD selection efficiency as in average the delayed signal happens around 62 μs after the prompt. However, the 7-plane spatial buffer limits the ability to exploit the topology of positron events. Also, these time and spatial buffers are too short to constitute control samples easily. The 200 us time window after the neutron trigger allows to monitor and evaluate the un-correlated background made of random pairs of ES and NS signals (see section 2.3 for the definition of the various backgrounds). This window, however, is close to the neutron signal, and is suspected to contain information correlated to it. As already mentioned, the spatial buffer (+/- 3 planes) is a limitation for the IBD analysis. An upgrade of the firmware would help to record more useful data and possibly increase the recorded spatial buffer. As it will be shown in chapter 5,

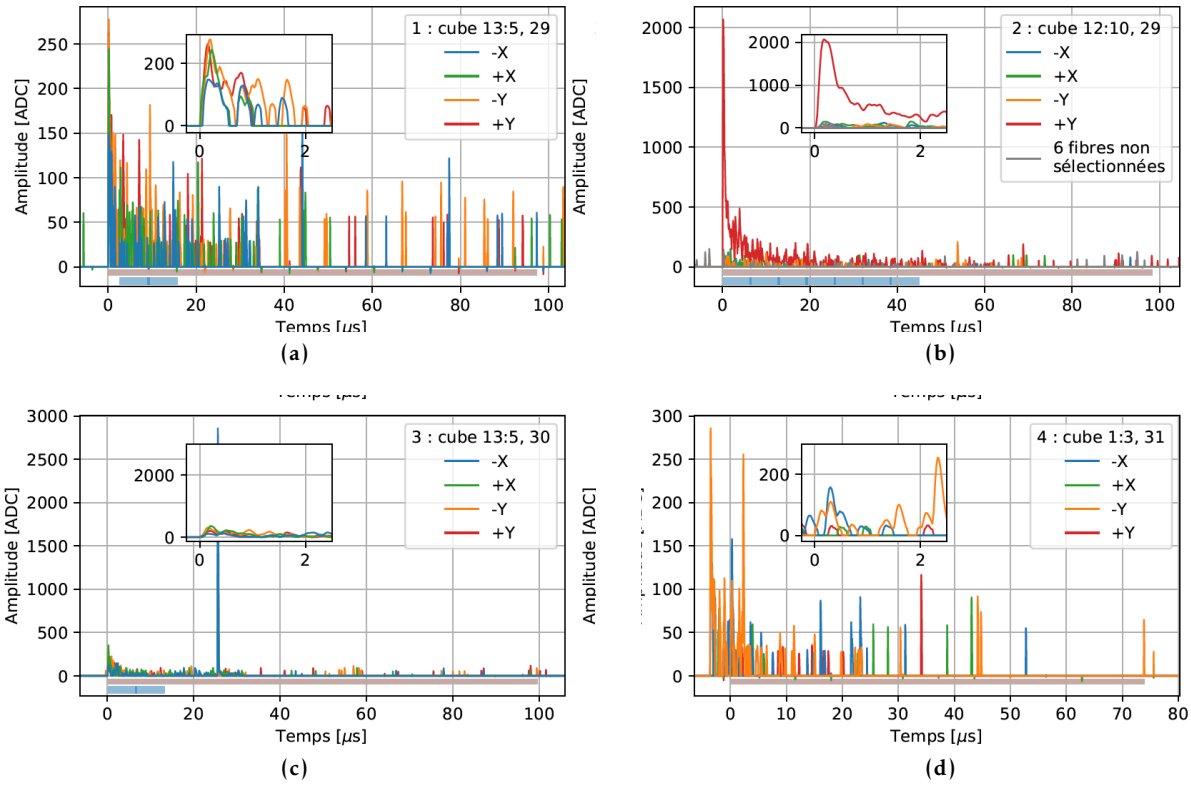


Figure 2.11.: Examples of waveforms following the capture of a neutron on a ^{6}Li nucleus. The -X, +X, -Y and +Y corresponds to channels with a MPPC at the left, right, top and bottom of a detection plane.

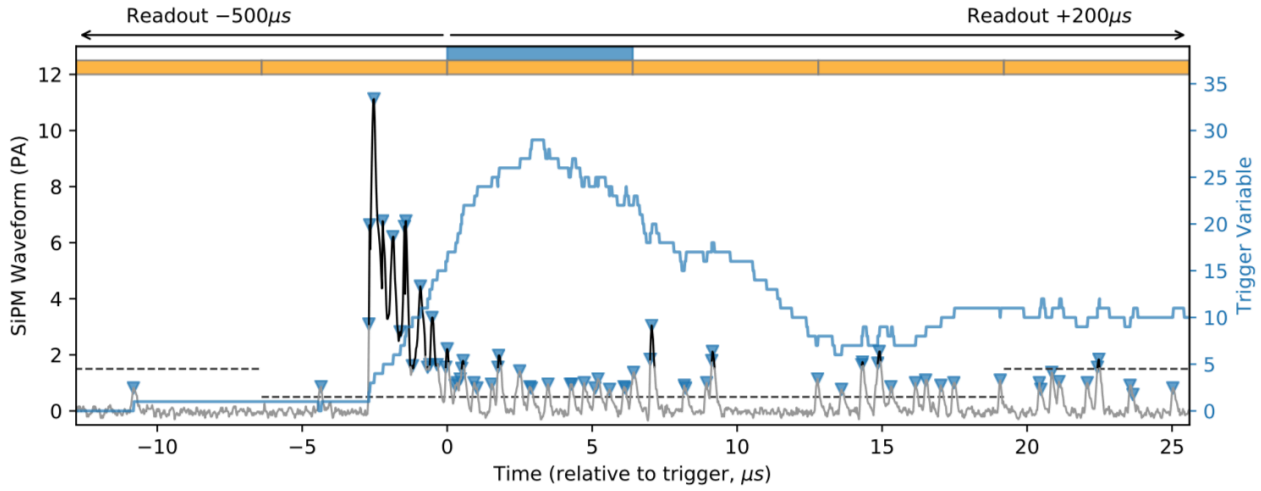


Figure 2.12.: Waveform that has triggered the neutron trigger. The trigger variable is the Peak over Threshold in a rolling time window of 6.4μs. All the sample within a [-500,+200] μs time window are readout. To increase the neutron identification, the ZS is lowered from 1.5 to 0.5 PA in [-6.4,+19.2] μs.

the IBD selection is improved based on the reconstruction and the identification of the annihilation gammas. In the PVT, the mean free path of them is about 10 cm. So if we consider an annihilation gamma emitted perpendicularly to the plane, only in 78% of cases will the first interaction occurs in the recorded part of the detector. Finally, the neutron trigger has a different treatment of the ZS, which is lowered from 1.5 to 0.5 PA in the previous time block as well as the three time blocks around the trigger. This has been done in order to increase the ability to see low amplitude peaks and have a better efficiency on neutrons. An example of a waveform passing the neutron trigger is shown on figure 2.12.

During physics data taking the three triggers are running in parallel. A summary of the trigger settings and the data rates obtained during reactor-on runs is given in table 2.1. All the recorded data are stored in binary file organised by time block. Also the information about the trigger plane position and the type of trigger is available.

Trigger Type	ZS Threshold	Condition	Readout Region		Trigger rate (Hz)	Data rate (MB/s)
			Space	Time (μs)		
Periodic	Disabled	Random 1Hz	Whole detector	12.8	1.2	3.9 (19%)
Threshold	1.5 PA	Waveform sample > 25 PA (horizontal and vertical coincidence)	Triggered plane	6.4	2100	2 (10%)
Neutron	0.5 PA or 1.5 PA	$N_{peaks} \geq 17$ peaks (Width = 6.4 μs , T_{peak} = 0.5 PA)	Triggered plane ± 3 planes	[-500,+200]	40	15 (71%)

Table 2.1.: Summary of trigger settings and associated data rates during reactor-on periods.

2.2.4.3. Dead time

The zero suppressed data is temporarily stored in buffer memory before their storage on the disk. There are two buffers, at the channel and plane levels. In the case of high data rate, those buffers could be saturated.

- **channel dead time:** The channel buffer is able to store a total of 1536 samples. If the limit is reached, the channel stops and is masked from the trigger and readout until the end of a run. During normal data taking about 1% of channels overflow by the end of the run.
- **plane dead time:** The plane buffer keeps all data coming from a plane. It can overflow in case of high trigger rate or if the readout process is not fast enough. In normal data-taking the plane deadtime represents a 1% of data loss.

Even if in physics runs the dead time has a limited impact on the data quality, this is not the case for calibration runs. The DAQ is not able to handle the activity of the source used to calibrate the detector. Consequently, this is one of the difficulties we met during the comparison between data and MC. This is shown in detail in chapter 4.

2.2.4.4. Time shifted planes

During physics analysis, an unexpected behaviour has been highlighted. Few planes are shifted of one time block (6.4 μs) with regards to the rest of the detector. This issue has no explanation at the moment of the writing of this thesis. Nevertheless an offline correction is performed by using horizontal muons. Indeed, in such events, muon cross several planes. Regarding the time sampling of 25 ns, those planes are hit at the same time. Thus the comparison of the reconstructed time for high amplitude event allows to spot shifted planes. In such cases, the processing is stopped and restarted with a time correction applied to the considered plane. But this algorithm is not working properly on calibration runs due to specific DAQ configurations, making the comparison between data and MC (chapter 4) more complicated.

2.3. Backgrounds in the SoLid experiment

The measurement of reactor antineutrinos implies to operate with important background rates. This section details the different backgrounds considered by the collaboration according to their origins.

2.3.1. Reactor induced backgrounds

The position of the detector has been chosen according to the hypothetical oscillation maximum determined with the RAA best fit. It implies to have a distance between the detector and the reactor

core of only few meters. The gamma environment activity evolves as a function of reactor status. Indeed during reactor-off periods, the gamma background is dominated by the natural radioactivity. During reactor-on periods, a few isotopes are activated by the reactor neutrons. The main contributor is the $^{41}Ar^{*}$. It has a decay time of 2.6 hours and emits a γ of 1.3 MeV. As those events are not followed by the emission of a neutron, they only contribute to the accidental rates. As said, nuclear reactors and especially experimental ones are an intense source of neutrons. Those neutrons coming from the reactor are sources of backgrounds, this impact is largely reduced by the shielding surrounding the reactor and the detector.

2.3.2. Cosmic induced background

Particles coming from the cosmos are bombarding the atmosphere regularly, creating cosmic ray air showers. The interaction of the primary particles on a nucleus in the atmosphere creates many energetic hadrons. Most of them are unstable and decay into other particles. In the context of SoLid, two kinds of primary particles are studied: neutrons and muons. Indeed they can produce directly or indirectly IBD-like candidates in the detector. This is one of the two most important correlated backgrounds. We have decided to split the cosmic induced background into two categories depending on this type of primary particles.

2.3.2.1. Muons

Muons reaching the detector have a wide range in energy. They can directly interact inside the detector releasing a lot of energy in a lot of cubes thus those events are easily tagged and removed from the analysis. But they are useful to perform a monitoring of the detector response in energy and also to spot shifted time planes. Most dangerous muons are the one interacting in the surrounding materials of the detector creating fast neutrons via spallation reactions. Muons interacting directly in the detector are excluded thanks to an offline selection. Remaining fast neutrons contribute to the correlated background as will be explained in the following subsection.

Also interaction of cosmic muons on ^{12}C, element of the PVT, via spallation could lead to the production of the two radioisotopes ^{9}Li and ^{8}He. Those isotopes are unstable and are β-n emitters and thus contribute to the correlated background. For neutrino reactor experiments with a significant overburden, this background is higher and not negligible than the one induced by production of fast neutrons. This is not the case in SoLid as the overburden is low and thus the fast neutron related background is largely dominant.

2.3.2.2. Fast neutrons

Fast neutrons are induced by muon interaction in the detector's vicinity as previously described or directly from cosmic air shower. Several processes are possible depending on their energy. Figure 2.13

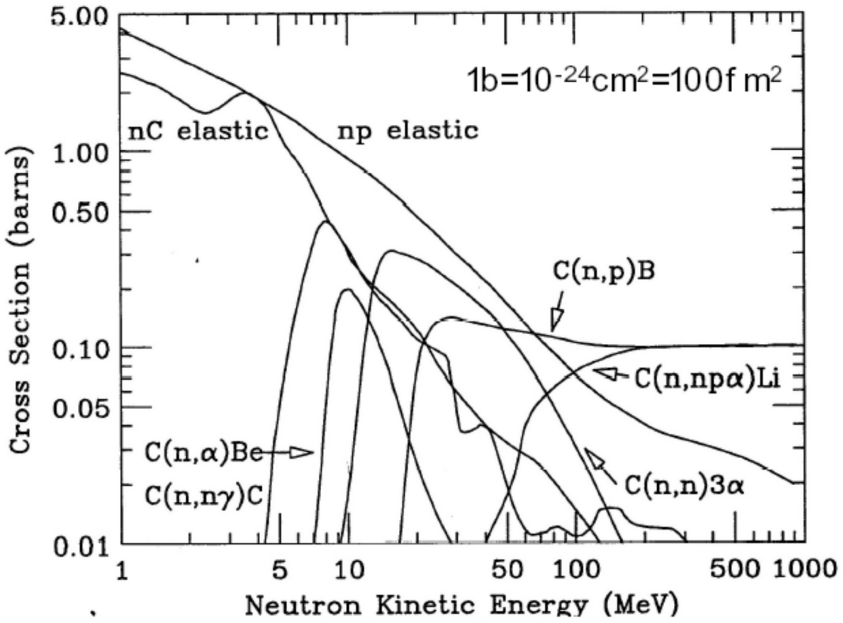


Figure 2.13.: Neutron cross-section involving Hydrogen and Carbon atoms. Several have a neutron among the outgoing particles giving the possibility to have a prompt-like and a delayed-like events.

shows the cross-section of several processes implying a neutron and Hydrogen or a Carbon atom. The elastic scattering process is the dominant one from 1 to around 50 MeV, in this case the neutron is slowed down via multiple elastic scatterings. Those proton recoils mimic the positron interactions. Indeed, the proton is an ionizing particle that causes the typical narrow and high scintillation peak in PVT. The apparent deposited energy is often in the IBD positron energy range. Neutrons often have an energy of a few tens of MeV. They tend to deposit half their energy at each scattering on a proton, but the reconstructed deposit is far smaller due to the quenching effect described in section 3.2.1. Then the neutron is captured by a ^{6}Li, implying a time between the prompt and delayed signals very close to that in the case of IBD, in average 62 μs.

Inelastic interactions may also occur involving carbon atom and with a neutron among the outgoing particles, for example:

- $n + {}^{12}\mathrm{C} \longrightarrow {}^{12}\mathrm{C}^* \longrightarrow {}^{12}\mathrm{C} + \gamma(4.4\ \mathrm{MeV}) + n$
- $n + {}^{12}\mathrm{C} \longrightarrow 3\alpha + n$

In the first reaction the prompt signal is given by the interaction of the 4.4 MeV γ in the PVT, in the second one it is the interaction of the three α particles each one having an energy of few MeV. For both, the delayed signal is due the outgoing neutron.

All those processes mimick the IBD reaction. Their rate is correlated with the atmosphere pressure variation and thus complicates its subtraction in reactor-on period. Unlike some liquid scintillators based experiments like the STEREO experiment, the SoLid detector is not able to discriminate nuclear recoils against electronic recoils using a Pulse Shape Discrimination technique [7]. The ability to reject

those backgrounds relies on the prompt topology interaction: cubes spatial and energetic distribution. This has driven the segmented design of the detector.

2.3.3. BiPo

The other dominant background is one of the backgrounds we expected to play a role, but at a far lower level. Its rate is finally nearly two orders of magnitude above that of IBDs before any selection. The BiPo background comes from the decay chain of the long-lived isotope ^{238}U. Along its decay chain, there are two decays that mimic the IBD:

$$^{214}\mathrm{Bi} \underset{19.9min}{\xrightarrow{\beta^-}} {}^{214}\mathrm{Po} \underset{164.37\mu s}{\xrightarrow{\alpha}} {}^{210}\mathrm{Pb} \tag{2.7}$$

The β^- decay emits an electron with an energy up to 3.2 MeV and is reconstructed as the prompt. A few hundred μs later, an alpha is emitted and mimics the neutron capture on Li atom. Moreover the discrimination against IBD events is difficult because the $\beta-$ decay is often followed shortly by radiative emissions. Those gamma rays mimic to some extent the annihilation gammas following the interaction of the positron from IBDs.

The collaboration has identified two origins for this background:

- **Internal pollution:** During the manufacturing of the ^{6}LiF :ZnS(Ag) sheets by SCINTACOR, a Radon pollution occured. The pollution is 10 times higher than expected, and could be explained by a change in the manufacturing process.
- **External pollution:** The BiPo rate evolves as a function of time. This is explained by an airborne pollution from the ^{222}Rn. The concentration of Rn depends on the pressure variation inside the containment building.

As a summary, there are two main backgrounds to deal with: the cosmic induced and the BiPo one. They are compromising the analyses foreseen by the collaboration as their rates dominate the signal one. We will see in section 2.5 that the first analysis attempted by the collaboration fail to overcome this issue and cannot be used for physics measurements. For this reason, we have developed a more discriminant approach taking advantage of annihilation gammas present in IBDs (see chapter 5).

2.4. Data reconstruction

Data reconstruction is performed offline by the *Saffron* software. It is a c++ program in which the full data reconstruction is implemented.

The algorithms analyse all signals recorded by the different triggers (see section 2.2.4.2) contained in a time window, called a cycle. A cycle starts at the beginning of the data taking and is closed when a gap of 350 time blocks (2.2 ms) long without any signal above ZS is found. All the analysis chain is

applied to this cycle and the reconstruction output is stored on disk before the allocated memory is released. So it allows to preserve the requested memory needed to analyse a full run.

The first step of the reconstruction is to constitute a set of *SWaveform* objects, called waveforms in the rest of the manuscript. The software loads the binary files in order to parse and merge waveforms. A waveform is defined as a continuous list of samples from a channel. The set of waveforms is time ordered to facilitate the analysis. From this point the goal of the reconstruction is to associate waveform corresponding to the same physical event and identify it. In this approach the waveforms are associated in time coincidence. They are combined in an object called a *SCluster* and for simplicity called cluster. This section presents a short summary of the cluster creation. It is a complex procedure. More details are given in [9].

2.4.1. Cluster creation

The goal of the clusterisation is to group waveforms which results from the same interaction of a particle in the detector. As we have seen, the distinction between ES and NS waveforms relies on the time extension of the signal. This complicates the procedure. For example, a NS waveform could be interpreted as several electromagnetic signals. The algorithm described in the following aims to correctly perform this association and identification.

First, a quality selection is applied to remove waveforms not suited for analysis: waveforms with fewer than 2 samples or with a maximum amplitude below 80 ADC (2.5 PA) are removed. Remaining waveforms of the same plane are combined together if they are in a 175 ns coincidence time and a cluster is created. The time reference is the time of the first sample of the waveform. Clusters without at least one vertical and one horizontal channel in the same plane are removed. This allows to reduce the impact of dark counts. Then all clusters coming from different planes are compared and if two have a starting time within 175 ns they are merged together. Also if two clusters share channels and one starts before the other one ends, they are merged. Those two conditions allow to have clusters with waveforms from several planes and to avoid creating new clusters on long signals, especially the one resulting from a neutron capture on ^{6}Li.

Then one has to correctly associate clusters sharing channels, but with a gap on these channels between the end of the wafeform of the first cluster and the start of the first wafeform of another cluster. If the gap is short, the two clusters could still be due to the same interaction. An additional time window is opened after the end waveforms preceding such gaps. If the first waveform of the other cluster on this channel starts within this window, then both clusters are merged. Tuning the size of this window was essential to not merge clusters actually due to independent processes. The channel multiplicity of clusters was also used.

At the end of this algorithm, a set of *SCluster* objects is reconstructed. In the rest of the reconstruction, each cluster is considered as one event occurring in the detector.

2.4.2. Cluster identification

The identification procedure aims to tag each cluster as muons, NS or ES. The identification is performed in a consecutive manner, starting with the muon tagging then NS tagging and the remaining clusters are categorised as ES.

2.4.2.1. Muon tagging

There are three types of muons defined in the reconstruction:

- **Type 0:** It corresponds to clipping muons. They interact in fewer than 3 cubes at the edge of the detector. They are the most difficult muons to identify due to low channel multiplicity.
- **Type 1:** Those muons have a higher multiplicity than type 0 but the tracking algorithm is not able to reconstruct the track. Those events mainly correspond to muons interacting in only one plane.
- **Type 2:** In this case the muons have crossed several planes and the cluster has a high channel multiplicity. The combination of spatial informations allows to reconstruct a track.

First of all, a selection is applied on the channel multiplicity. It requests to have at least 11 horizontal or vertical channels in the cluster with an amplitude above 200 ADC. If this criterion is met by the cluster, it is identified as a muon. It remains to determine the type of this muon, according to the previous definitions. The rest of the identification relies on the ability to reconstruct a track. The tracking algorithm tries to fit a straight line separately using horizontal (Y-Z projection) and vertical (X-Z projection) channels (see figure 2.14). If one of these fits does not converge, the cluster is identified as type 1. Otherwise it is categorised as type 2 and the combination of the informations from the two fits allows to fully describe the muon track.

2.4.2.2. NS tagging

The neutron identification relies first on the length of clusters. The first criteria is to be a cluster of at least 25μs. Also it requests that all the waveforms forming the cluster have an amplitude below 10,000 ADC (i.e. 10 MeV) to exclude misidentified muons.

For each cubes (a cube here is formed by at least one horizontal and one vertical fibres with a signal) in clusters passing the previous selection, three variables are computed:

- **Amplitude:** The amplitude is defined as the average over the 4 channels serving the cube of the highest amplitude found on each channel. We mean here the highest amplitude among the WF merged on this channel by the clustering algorithm.
- **Integral:** The long integral is defined as the mean of the integrals over cube channels. The integral is performed in the [0,25]μs time window.

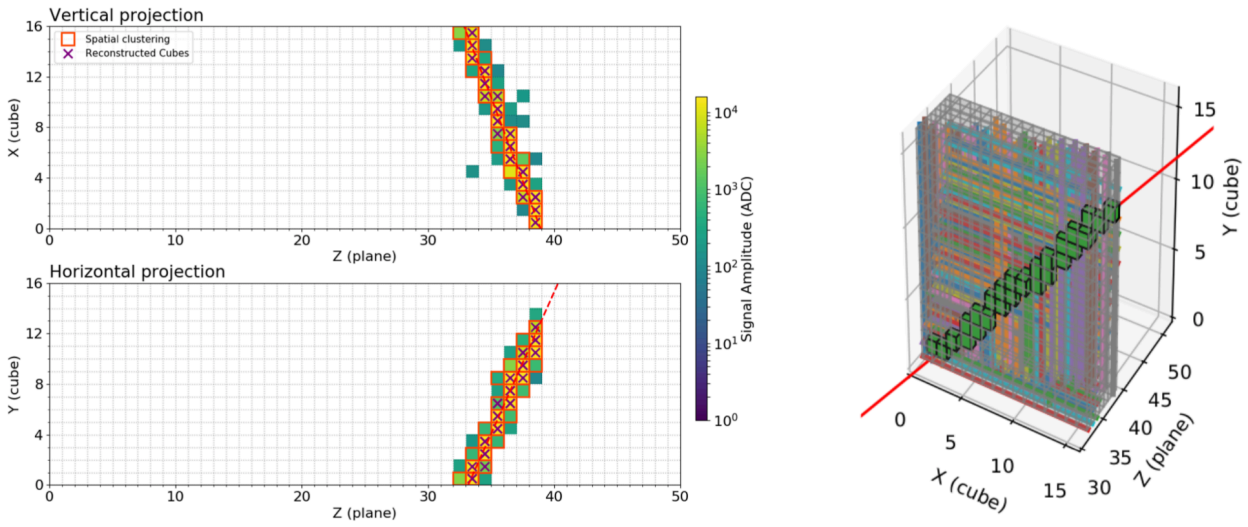


Figure 2.14.: Example of a reconstructed muon track (type 2). *Left:* Vertical and horizontal projections with the signal amplitude. The two projections are fitted with a straight line. *Right:* 3D view of the muon tracks. The reconstructed cubes are shown in green. The combination of the two projections fits allow to obtain the muon's trajectory through the detector.

- **IonA:** This is the ratio between the integral and the amplitude of the signal.

The variables are computed by taking the average over all the channels serving a cube, allowing to apply a selection without taking care on the number of fibres actually seeing a signal. The distribution of the IonA variable is shown on figure 2.15a. We can see two contributions. At low values, that of ES signals. At higher values, the contribution of interest. Thus we keep as candidate only cubes satisfying the condition:

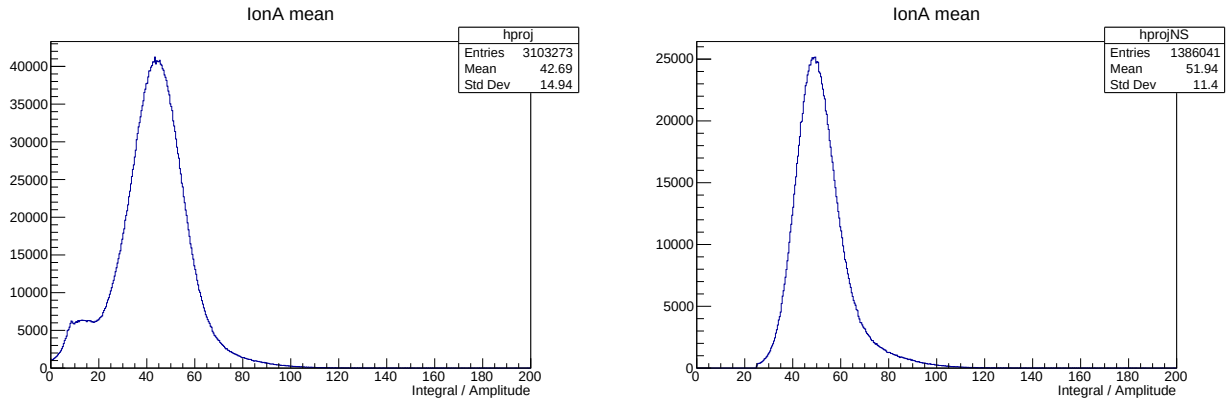
$$\text{IonA} > 25. \tag{2.8}$$

If several cubes from the same cluster satisfy this condition, the number of Peak Over Threshold (PoT) is computed in a 25μs time window. The cluster is redefined based on this variable. Only one cube is kept : to the one in which PoT is maximal. The other cubes (and corresponding channels) are discarded. The IonA distribution of NS cubes is given in 2.15b, we can see that the low IonA contribution has been correctly removed.

2.4.2.3. ES tagging and reconstruction

Cube reconstruction

Remaining clusters, not been identified as NS or muons, are tagged by default as ES clusters. A specific reconstruction of interaction is performed on those events in order to take advantage of the topology.



(a) Cube candidate.

(b) Cube identified as NS.

Figure 2.15.: IonA distribution of NS cube candidates 2.15a and of identified NS cube 2.15b after application of selection.

One of the key parameters of the reconstruction is the fibre analysis threshold. It is a selection applied at the beginning of the algorithm to remove signals with a low amplitude. At the beginning of the thesis the nominal value was 4.5 PA. This value allowed to remove dark counts and also low amplitude cubes were not used by the first signal selections. But this threshold has been lowered, thanks to preliminary studies I carried out showing the need to do that. It is now 2.5 PAs. This increases the efficiency to see low energy deposits coming from annihilation gamma. This is discussed in the chapter 5.

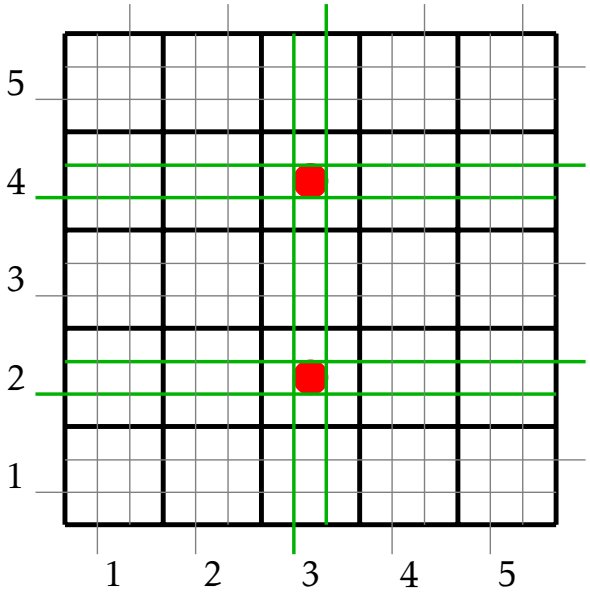
Before introducing the algorithm used by the collaboration to compute the deposited energy in cubes, it is interesting to quickly show the first version of the energy reconstruction, in order to understand why a more sophisticated algorithm had to be developed. Also, the simplicity of this first approach still proved very useful when it came to understand the detector's response (see e.g. chapter 4). Initially the cube's energy was simply computed by doing the sum of the amplitudes of the waveforms found on the 4 fibres serving it:

$$A^{SCube} = A^{Hori,1}_{Fibre} + A^{Hori,2}_{Fibre} + A^{Vert,1}_{Fibre} + A^{Vert,2}_{Fibre} \tag{2.9}$$

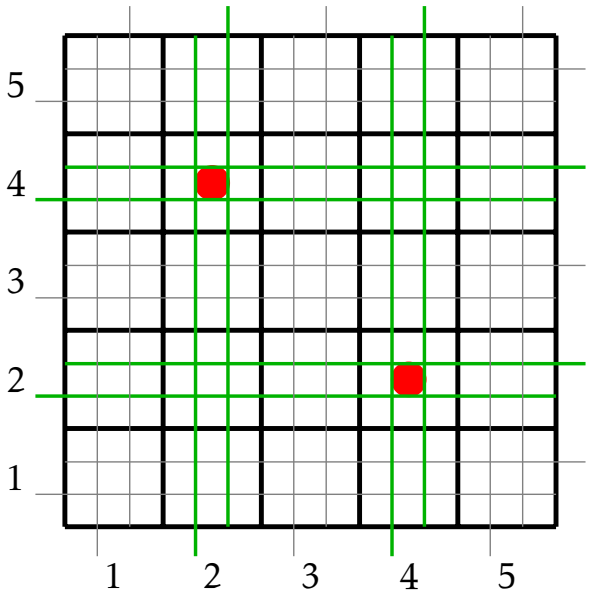
Where: A^{SCube} is the cube's amplitude and $A^{Hori,1}_{Fibre}$ is the amplitude on the first horizontal fibre. This approach to reconstruct the cube amplitude is called *SCube*.

The issue with this kind of reconstruction happens when there are several energy deposits in the same plane creating some degeneracies of the energy information. Indeed two problematic cases appeared:

- **Pile-up:** At least two cubes on the same row or column have been hit.
- **Ambiguities:** At least two cubes not on the same row or column have been hit.



(a) Schematic showing an example of **pile-up**. The two cubes (3,2) and (3,4) are reconstructed in the correct position but with a biased energy using the SCube approach.



(b) Schematic showing an example of **ambiguity**. Two cubes (2,4) and (4,2) are correctly reconstructed with the correct energy but two others (2,2) and (4,4) are false cubes with the SCube approach.

Figure 2.16.: The red circles represent a energy deposit in a cube. The grey lines show the fibres without signal and the green ones are showing fibres seeing a signal.

Using the *SCube* reconstruction, the cube's amplitude will be biased in the first case. In the case of ambiguities, two false cubes will be reconstructed as shown by the figure 2.16.

To overcome this difficulty a new approach was introduced. This reconstruction is performed plane by plane and is based on a Maximum-Likelihood-Expectation-Maximisation (ML-EM) algorithm [72]. This kind of iterative algorithm is used in medical imagery [73] and by the NEXT collaboration [74].

In the context of SoLid, the goal is to reconstruct the deposited energy in each cube of a plane based on the fibres measurement. The amplitude seen on one given fibre gives an estimation of the integrated energy in the corresponding row or column. It tells nothing of the individual energy deposited in cubes. The role of this method is to estimate them. It computes the likelihood of a list of cube positions and true energies to correspond to the ensemble of fibre signals in this plane. This likelihood is maximize iteratively, via this formula is given by:

$$A_j^{n+1} = \frac{A_j^n}{\sum_i a_{ij}} \sum_i a_{ij} \frac{p_i}{\sum_j a_{ij} A_j^n} \tag{2.10}$$

Where A_j^n is the estimated amplitude in thej^{th} cube at the step n, a_{ij} is the projector from the cube j to the i^{th} SiPM and p_i is the amplitude observed on the i^{th} SiPM. The initialisation of the algorithm

consists to set the values of A_j^0 equal to the sum of the four fibres amplitude serving it. Typically, 50 iterations are necessary to converge. This approach to reconstruct the cube's amplitude is called *CCube*.

As the SiPM see at the first order 25% of the light emitted by a cube, all the elements of the matrix a_{ij} are equal to 0.25. Also to be consistent with this value, the fibre's amplitude p_i is corrected by the fibre-to-SiPM optical coupling measurement detailed in section 3.6. The conversion from electronic unit to natural unit is performed at the end of iterations:

$$E_j^{CCube} = \frac{A_j^{CCube}}{C_j} \tag{2.11}$$

Where C_j is the calibration constant in PA/MeV of the jth cube. The determination of coupling and cube calibration constants is detailed in the chapter 3.

An improved version of the algorithm is currently in development. In particular, this new method will take into account the attenuation of light when it is transported by the fibre from the cube to the SiPM. In the present method, the effect of the attenuation on the energy reconstructed in a given cube is treated by an 4-fibre average correction naturally included in C_j. The new method will therefore improve the energy reconstruction by correcting more accurately for detector inhomogeneities.

Time reconstruction

Once the cube's position and energy have been reconstructed, a finest estimation of the time is computed. At first, timing is limited by the sampling of 25 ns applied during the waveform digitalisation. In order to increase the accuracy on the reconstructed time a weighted summation method is applied following the equation:

$$T_{Peak} = \frac{\sum_{i=i_{max}-3}^{i_{max}+3} A(i) \times T(i)}{\sum_{i=i_{max}-3}^{i_{max}+3} A(i)} \tag{2.12}$$

where: i_{max} is the sample with the maximum amplitude, $A(i)$ is the amplitude in ADC of ith sample and $T(i)$ is the time in nanosecond of the ith sample. The sum is performed on the seven samples around the waveform's maximum amplitude. This method has a temporal resolution of about 10 ns.

2.4.3. ES-NS Coincidences

The last step of the reconstruction consists of the creation of coincidences between ES and NS clusters. For each pair a preselection is applied to the cluster's energy and the time difference between the two clusters:

- $E_{cluster}$: It is the sum of all cube's energy reconstructed in the ES cluster.

- ΔT_{ES-NS}: It is the time difference between the ES and the NS clusters.

The preselection requires that:

$$E_{cluster} > 1 \text{ MeV} \text{ and } \Delta T_{ES-NS} \in [-600, 600]\mu s \tag{2.13}$$

The cut on the energy allows to reduce the output file size due to the high background level below 1 MeV.

The ES cluster may have several reconstructed cubes thus it is important to define a reference cube in order to compute, for example, the relative positions between the ES and NS cluster. As said, a prompt event from IBD is composed of a positron carrying the neutrino energy and two annihilations emitted back-to-back with an energy of 0.511 MeV. Thus it is reasonable to use the cube with the highest energy.

For the selected coincidences, all the relevant information to perform the signal selection is kept and written on the output ROOT file.

2.5. SoLid analysis status

There are two physics studies foreseen by the SoLid collaboration: the search for a sterile neutrino and the measurement of the reactor antineutrino energy spectrum from ^{235}U. In this thesis, we will focus on some IBD analyses related to the sterile neutrino search. The energy spectrum has not been studied by the collaboration yet. As introduced in chapter 1, the presence of a sterile neutrino would manifest via an oscillation pattern impacting the antineutrinos baseline and their energy distributions. The amplitude and frequency of those oscillations are driven by two parameters: the mixing angle θ_{ee} and the square of the mass difference Δm^2_{14}. The statistical significance with which the analysis will exclude or not a region of the (Δm^2_{14},$\sin^2 2\theta_{ee}$) phase-space is determined by fitting fake data distributions to the energy vs baseline distribution measured in real data. These fake data distributions assume various values of (Δm^2_{14},$\sin^2 2\theta_{ee}$) and incorporate all detection effects using the full MC simulation. So simulation plays a major role on the results of the analysis and the chapter 4 is dedicated to the tuning of the energy response of the simulation.

2.5.1. Fitting strategy

As already explained in chapter 1, the oscillation is observed by measuring the distortions it causes in the neutrino energy spectrum, and the way they vary with the baseline L to the source. To detect a distortion, one needs a reference. Ideally, one would use predicted energy spectra, assuming no oscillation, provided by reactor physicists. As discussed in 1.3.2.3, these predictions are imperfect. However, the unoscillated spectrum does not vary as a function of L. Therefore, by measuring energy spectra at various distances, the importance of this issue is minimized. Indeed, even if the distortions

measured by comparing them with the predicted spectrum are biased due to the imperfectness of the prediction, their variation with L should still be accurate enough. This principle is followed by SoLid and by most of other very short baseline experiments (see for instance [7]): their detectors are segmented longitudinally in order to measure spectra as a function of L, covering a region of about 2 m, centered a few meters away from the source. These two elements are dictated by the most probable values of (Δm^2_{14},$\sin^2 2\theta_{ee}$) obtained by studying the RAA, which predict a short oscillation, starting a few meters from the source. The longitudinal homogeneity of these detectors also allows to minimize the impact of the uncertainties with which we understand the reconstruction and selection effects (i.e. the detector response) since we expect them to not vary, or slowly, as the function of L. Consequently, in SoLid, fake spectra are generated at various values of L (one per module in the present version of the analysis), starting from predicted spectra, distorted according to (Δm^2_{14},$\sin^2 2\theta_{ee}$) and L, convoluted with the detector's response and fitted to the spectra measured at these distances. In principle, it is also possible to directly compare measured spectra in various modules, taking one module as a reference, in order to be completely independent from predictions. However, this is not optimal a priori: the information from the accuracy of the prediction (e.g. normalisation) is lost and the statistical uncertainties are larger this way (limited statistics in the reference module). Also, acceptance effects as a function of L need a prediction to be provided.

In SoLid, the fit is based on the comparison of the module to module spectrum ratios measured real data with those observed in generated fake data, taking the closest module as a common denominator: then ratios of the detected number of IBD are computed:

$$R_{ij} = \frac{N_{ij}}{N_{0j}} \tag{2.14}$$

where, i id the i^{th} bin of the baseline, j is the j^{th} energy bin and N is the number of detected signal event. Then a chi-square is computed in order to compare the ratios from data and ratios from the predictions evaluated for different values of $\sin^2 2\theta_{ee}$ and Δm^2_{14}. This chi-square is given by:

$$\chi^2 = \sum_{i>0}^{n_L-1} \sum_{j}^{n_E} \sum_{k>0}^{n_L-1} \sum_{l}^{n_E} \left(D_{ij} - R_{ij}\right)^T V^{-1}_{\alpha\beta} \left(D_{kl} - R_{kl}\right) \tag{2.15}$$

where, n_L and n_E are the number of bin on the baseline and energy. D_{ij} is the equivalent of R_{ij} introduced in equation 2.14 but coming from predictions and $V^{-1}_{\alpha\beta}$ is the covariance matrix taking into account the correlated systematic uncertainties as well as the data statistical uncertainties. For a given set of values of $\sin^2 2\theta_{ee}$ and Δm^2_{14}, fake data are generated using the whole simulation chain described in chapter 4. The statistical agreement between those fake data (predictions) and the real one is performed via the chi-square defined in equation 2.15. Using it, one could determine:

- **Sensitivity:** The sensitivity of the experiment by comparing fake data between them, replacing the real data distribution by another fake data sample.

- **Exclusion contour:** An exclusion contour by comparing real data with fake data under the hypothesis of no oscillation.
- **Best parameters:** The best oscillation parameters in the hypothesis of the presence of a sterile neutrino oscillation.

In order to have more details about the development of the fitting procedure, see [75].

2.5.2. Figures of merit

The collaboration has decided to focus on the sterile neutrino analysis for the moment. In [76], the authors show the list of parameters that are playing a role on the sensitivity of an experiment to the sterile neutrino oscillation:

- **Reactor parameters:** The reactor power, the fuel type, the duty cycle and the core dimensions are the parameters influencing such measurement. Indeed, they mostly drive the number of antineutrinos emitted by the reactor and thus the signal rate expected after analysis.
- **Facility parameters:** The detector dimension and position with regards to the reactor core are constraints by the reactor facility. For example, depending on the building structure, the closest distance between the detector and the reactor will differ from a site to another one. As the neutrino oscillation phenomenon depends on the travelled distance, it has to be taken into account.
- **Detector parameters:** The authors list in this category the parameters related to the detector: the fiducial volume and the target mass, the detection efficiency and the position and energy resolutions. For example, in chapter 3 dedicated to the energy calibration of the detector, we show that the energy resolution is within the acceptable range: 14% at one MeV. This does not correspond to the energy resolution on the positron's energy. Indeed the positron energy is not a single energy deposit in a single cube. Its energy is often shared among several cubes, some of them below threshold. Also, the two annihilation gammas can deteriorate the total energy reconstruction if one cannot tell the positron's deposits from the deposits due to the gammas. These issues depend on the ability to reconstruct low energy deposits. Preliminary studies show that the energy resolution on the positron's energy reconstruction is 20% at one MeV. Concerning the position resolution, we have chosen to use modules as baseline definition, meaning 50 cm.
- **Backgrouds:** Reactor experiments operated at very short baseline, such as SoLid, are subject to a challenging background contribution, described at length earlier in this manuscript. The ability to reject and subtract those backgrounds is critical to lower as much as possible the error associated to the measurement.

We can see that most of the parameters are determined by the experimental set-up: the choice of the reactor, the detector technology and its position from the reactor core. It remains for analysers to design the best selection and signal extraction method. The best figure of merit is ideally the area of

the (Δm^2_{14},$\sin^2 2\theta_{ee}$) phase-space excluded by a given analysis. To be produced, it requires to run the whole fitting procedure. To optimise the signal selection in a less CPU demanding way, we use two figures of merit far simpler to obtain and of direct impact on the contours: the signal efficiency (or signal rate) and the signal over background ratio.

2.5.2.1. Signal efficiency

The signal efficiency is defined as the ratio between the number of signal events selected after the reconstruction and selection processes over the number of electronic antineutrino emitted by the reactor. This efficiency can be expressed as the product of several efficiencies:

$$\epsilon_{signal} = \epsilon_{acceptance} \times \epsilon_{IBD} \times \epsilon_{n,capture} \times \epsilon_{n,reco} \times \epsilon_{selection} \tag{2.16}$$

where:

- $\epsilon_{acceptance}$: represents the detector geometric acceptance, about 0.11%.
- ϵ_{IBD}: takes into account the IBD cross-section and the number of protons in the target.
- $\epsilon_{n,capture}$: is the neutron capture efficiency. It mainly depends on the number of ^{6}Li atom available and the detector geometry. It was estimated in [9] to be around 71%.
- $\epsilon_{n,reco}$: is the neutron reconstruction efficiency. It depends on the detector properties: light yield, trigger efficiency and the event identification. It was estimated in [9] to be around 73.9%.
- $\epsilon_{selection}$: is the signal selection efficiency. It corresponds to the percentage of signal event passing the signal selection.

The last efficiency $\epsilon_{selection}$ is the only one on which analysers have some flexibility. It is not possible to obtain a purely data-driven determination of this efficiency. It is determined either fully from simulation, or by determining the signal rate in a part of the reactor-on data and by dividing it by the rate predicted in the detector before the selection. In either case, the numerator of the efficiency is a rate of selected IBDs. It is defined as follows when it is determined from a reactor-on data sample :

$$R^{Data}_{Signal,sel} = \frac{N^{Data}_{Signal,sel}}{t_{livetime}}, \tag{2.17}$$

where: $N^{Data}_{Signal,sel}$ is the number of events identified as signal in the studied dataset and $t_{livetime}$ is the time during which the detector has recorded data. When it is determined in a simulated sample, it is computed this way :

$$R^{MC}_{Signal,sel} = \frac{N^{MC}_{Signal,sel}}{N^{MC}_{Signal,gen}} \times R^{MC}_{Signal,gen}, \tag{2.18}$$

where: $N^{MC}_{Signal,sel}$ is the number of selected events as signal in the MC dataset and $N^{MC}_{Signal,gen}$ is the number of events generated at the Geant4 level to obtain this dataset. $R^{MC}_{Signal,gen}$ is the expected rate of IBD interactions in the detector derived from the reactor predictions. For the cycle studied in the chapter 5 we expect to have in average 1,200 IBD per day.

2.5.2.2. Signal over background

The Signal over Background ratio, noted SoB, expressed the ability of the experiment to reject the background while still maintaining the efficiency high. As for the signal rate, the background rate can be derived from data or simulation. It is practical to derive it from data thanks to reactor-off periods. Indeed those data provides a background only measurement more precise and reliable than a simulation. Moreover even if a cross-check via simulation would give confidence in the results, this was not possible because of a lack of precision to simulate cosmic induced events. Thus those numbers are defined as:

$$\left(\frac{S}{B}\right)^{Data,MC} = \frac{R^{Data,MC}_{Signal,sel}}{R^{Data}_{Background,sel}} \tag{2.19}$$

The ability of the experiment to search a sterile oscillation pattern depends a lot on S/B, in two ways. First, the statistical power of the search depends obviously on S (thus on the efficiency), but also on the statistical fluctuations of the number of background events still present in the selected sample. Therefore, B must be as small as possible compared to S. Moreover, as will be seen later, one needs to subtract the background component to search for the oscillation. It is therefore necessary to estimate it. The method used for that (see section 5.4) suffer systematic uncertainties. The higher B, the higher the systematic uncertainty propagated to the signal spectrum used in the oscillation search.

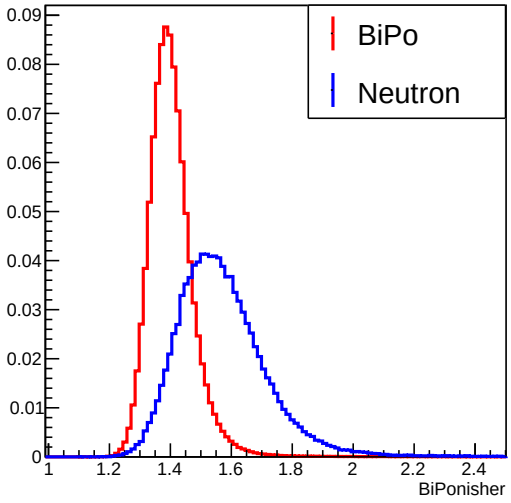
2.5.3. Standard signal selection

This section aims to present the status of the signal selection in late 2019, before I started to develop the analysis presented in chapter 5. It will prove the necessity of this improved analysis. Also the goal is to have a global overview of basic variables involved in the signal selection. This selection has been derived from the work in [9].

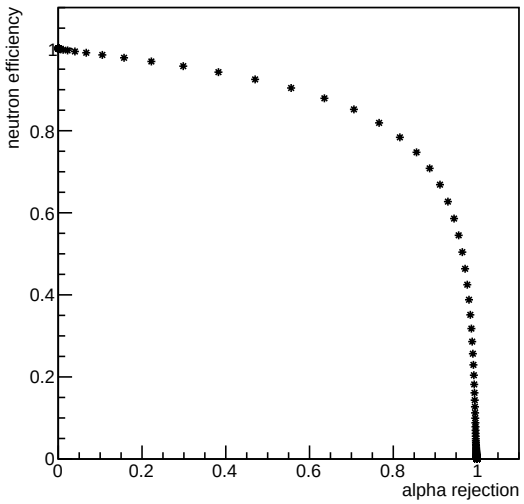
2.5.3.1. Discriminative variables

As we have seen in the section 2.4.3, at the end of the data processing, ES and NS events are associated to form coincidence events. The signal selection is then a set of cut based on NS and ES related variables:

- E_{prompt}: Sum of the energy deposited in the ES cluster.



(a) BiPonisher distribution of enriched selection of BiPo background using reactor-off days (red) and pure neutron capture event (blue).



(b) ROC curve.

Figure 2.17.: Discrimination power between alpha (BiPo) interactions and alpha plus tritium (neutron capture on ^{6}Li) interaction in the ZnS scintillator.

- ΔT_{ES-NS}: Time difference between ES and NS clusters.
- $\Delta X, Y, Z, R_{ES-NS}$: Relative position between ES and NS clusters.

One of the dominant background types is due to the contamination of ^{6}LiF :ZnS(Ag) layer (see section 2.3.3). One of the differences between IBD and a BiPo event concerns the NS event. In the case of a neutron capture, there is 4.8 MeV deposited by tritium and an alpha whereas for the BiPo it is an alpha with an energy of 7.8 MeV. Thus waveforms from BiPo are shorter, with a higher amplitude than those from neutron captures. So it is possible to discriminate the two events by using Pulse Shape Discrimination techniques on NS waveforms. This is defined by a ratio between two amplitude integrals with a different time window: a short one between [0;7.5]us and a long one [0;87.5]us. The variable derived from this calculation is called the BiPonisher. The discrimination power obtained with this variable is shown by figure 2.17.

2.5.3.2. Background distributions

First, we have studied few days of reactor-off data as it gives a pure background measurement. This background could be split in three components depending on the origin: accidental, BiPo or cosmic induced (see section 2.3). At the moment of the writing, the simulation of the cosmic induced background was not reliable enough to use it as prediction. Thus it was required to use a data driven method to better understand the particularities of each background.

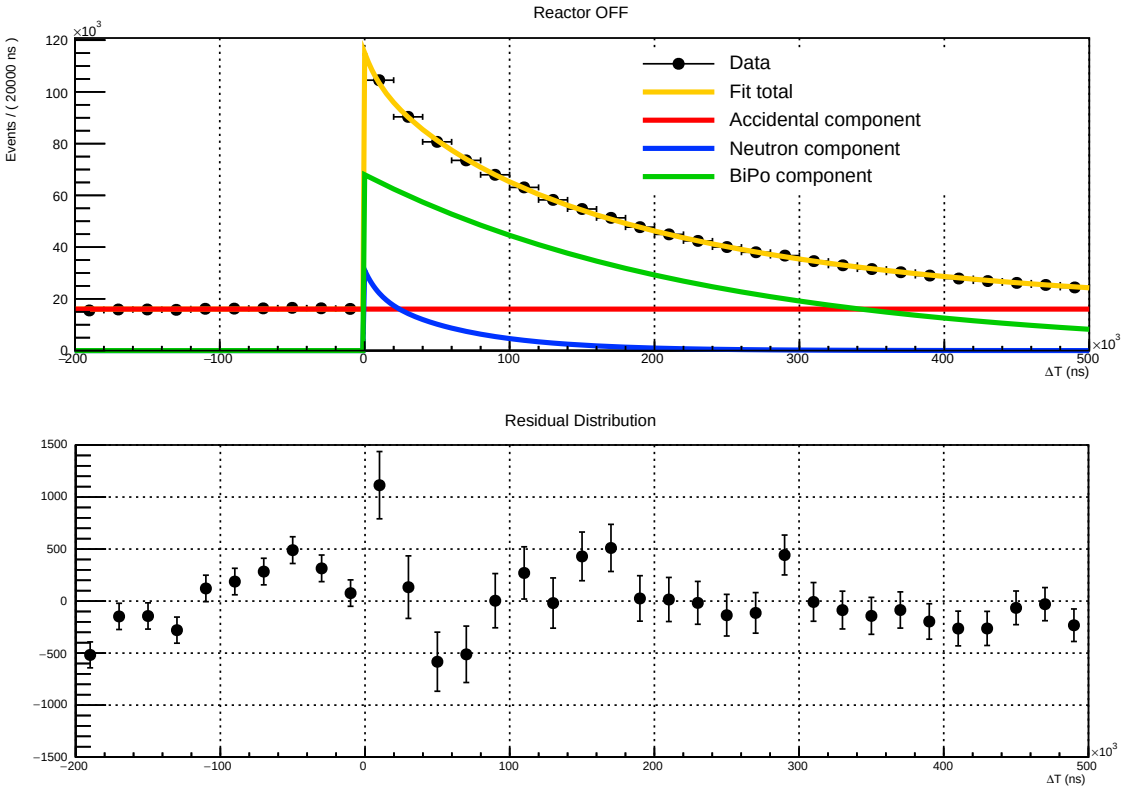


Figure 2.18.: Fit of the reactor-off ΔT_{ES-NS} distribution.

As it will be detailed in the section 5.4.2, each background have a different distribution with regards to the ΔT_{ES-NS} variable. So we can use it in order to extract the distributions of each background with the sPlot method [77]. This method has been applied on processed data, figure 2.18 shows the fitted ΔT_{ES-NS} distribution and the residual. The negative time window corresponds to an almost pure measurement of the accidental background. Whereas the positive time window shows the contribution of all backgrounds: correlated and accidental ones. We can see that the fit describes well the data and that the residuals are contained within 3%. This level of agreement is enough as the idea is to highlight the background particularities and not to perform a background subtraction as it will be done in the chapter 5.

From this, we can obtain the distribution of all components with regards to other discriminative variables introduced in the previous section. The figure 2.20 is showing the distributions of those components. In the rest of the section, we will discuss the shape of each background in order to find the best selection.

Accidental background

We can see on the figure 2.20a, that the accidental background dominates the low energy region of the E_{prompt} variable. This is explained by the activation of the $^{41}Ar^{*}$ isotope by neutron coming from the reactor. The shoulder around 2 MeV could be explained by neutron capture on Hydrogen

giving a gamma of 2.2 MeV. Also, as the accidental background is the result of the association of an ES and NS clusters of different origins, we expect to have a wider spatial distribution between the ES and NS compared to correlated events. This is shown by figures 2.20b to 2.20e. Figure 2.20d shows the ΔZ_{ES-NS} distribution. We can see that there is less accidental coincidences for which the ES and the NS are reconstructed in the same plane ($\Delta Z_{ES-NS} = 0$). This is an artefact of the reconstruction, indeed the accidental component is partly estimated in the negative region of ΔT_{ES-NS}, as shown on figure 2.18. In this region, the ES cluster is reconstructed after the NS cluster. Due to the large scintillation time constant of the ZnS scintillator, waveforms coming from it could hide ES signals. This is not an issue once the signal selection is applied as the accidental background is drastically suppressed.

Based on the observations above, we fight against this background by applying a cut on E_{prompt} to remove the low energy region and in addition require the ES and NS clusters to be close in space. Figure 2.20f shows that accidental and BiPo components have a quite similar BiPonisher distribution. This indicates that most of accidental events are built with a NS cluster due to an α decay.

BiPo background

The prompt energy distribution of BiPo events is, as expected, dominating the low energy region up to 3 MeV. The distribution of BiPo events in $\Delta X, Y, Z, R_{ES-NS}$ is also interesting. Indeed, those distribution can be understood based on the detector geometry, with the position of the ZnS sheets. Indeed the two ZnS sheets are positioned in the X and Z direction. Therefore, we can highlight specific $\Delta X, Y, Z_{ES-NS}$ topologies, for example if we select events for which $\Delta Y_{ES-NS} = 0$, we can obtain figure 2.19. Three topologies are dominating:

- $\Delta X_{ES-NS} = 0$ and $\Delta Z_{ES-NS} = 0$: corresponding to event in which the ES and the NS clusters have been reconstructed in the same cube.
- $\Delta X_{ES-NS} = 0$ and $\Delta Z_{ES-NS} = 1$: corresponding to events in which the electron has been detected in the cube adjacent to the ZnS sheet positioned in the Z direction.
- $\Delta X_{ES-NS} = -1$ and $\Delta Z_{ES-NS} = 0$: corresponding to events in which the electron has been detected in the cube adjacent to the ZnS sheet positioned in the X direction.

Unfortunately the $\Delta X, Y, Z, R_{ES-NS}$ variables do not help a lot to reject this background. But the BiPonisher one is the best discrimination variable as we can see on figure 2.20f. Be reminded that IBDs have the same distribution as cosmic, fast neutrons backgrounds. We can also see that a combination of cuts on the BiPonisher and on ΔT_{ES-NS} isolated an almost pure sample of BiPo background.

Cosmic induced background

The cosmic induced background impacts the whole signal energy range as we can see on figure 2.20a. But compared to other backgrounds, it dominates only at high energy. As for the BiPo

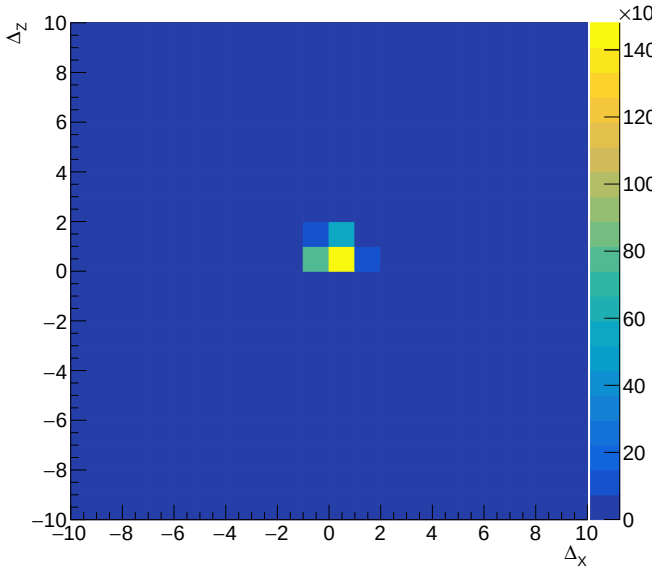


Figure 2.19.: 2-dimensional (ΔZ_{ES-NS},ΔX_{ES-NS}) of BiPo events while keeping only events such as $\Delta Y_{ES-NS} = 0$.

background, the relative position variables between ES and NS clusters are not so useful. Thus, to fight this background, a selection removing the high energy region is required and new discriminative variables are needed in order to reach a decent signal over background ratio. A new approach has been developed during the thesis and this is shown in the chapter 5.

2.5.3.3. Selection

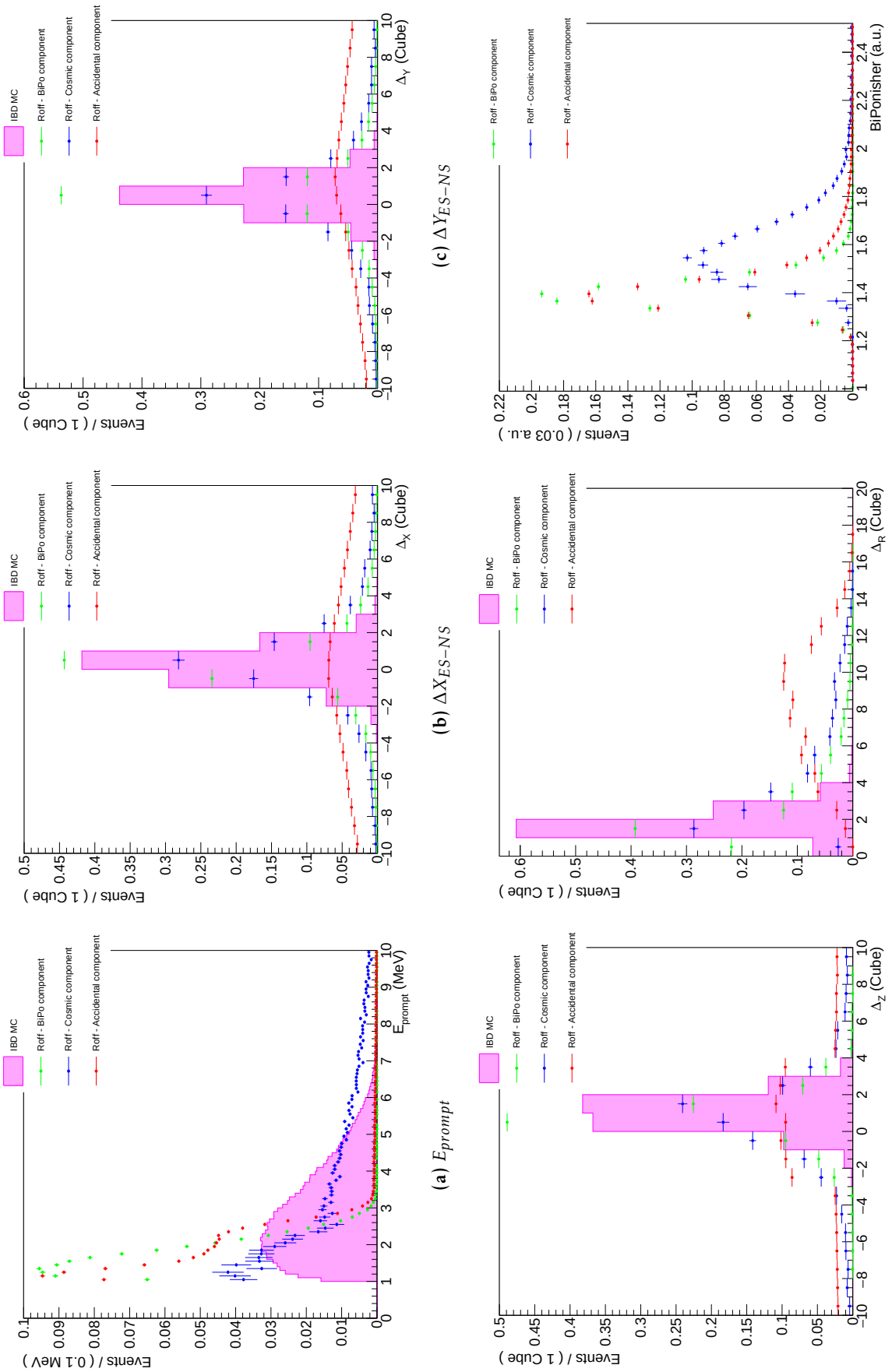
Three cuts are used by all IBD analyses described in the manuscript. They are considered as a preselection :

- $E_{cluster} \in [2,7]$MeV : This range of cut is justified by the presence of a high BiPo background rate below 2 MeV and above 7 MeV there is no sizeable IBD contribution.
- $\Delta T_{ES-NS} \in [1,141]\mu s$: The upper limit allows to remove a large number of BiPo events.
- BiPonisher > 1.44.

After this preselection is applied, the other cuts are optimised by comparing IBD simulation and a few days reactor-off data as shown by the figure 2.20. A first discriminative variable against BiPo events is:

- $\frac{E_2}{E_1}$: Ratio between the energy of the most energetic cube and the energy of the second most energetic. If there is only one cube reconstructed in the event, the E_2 is equal to -1.

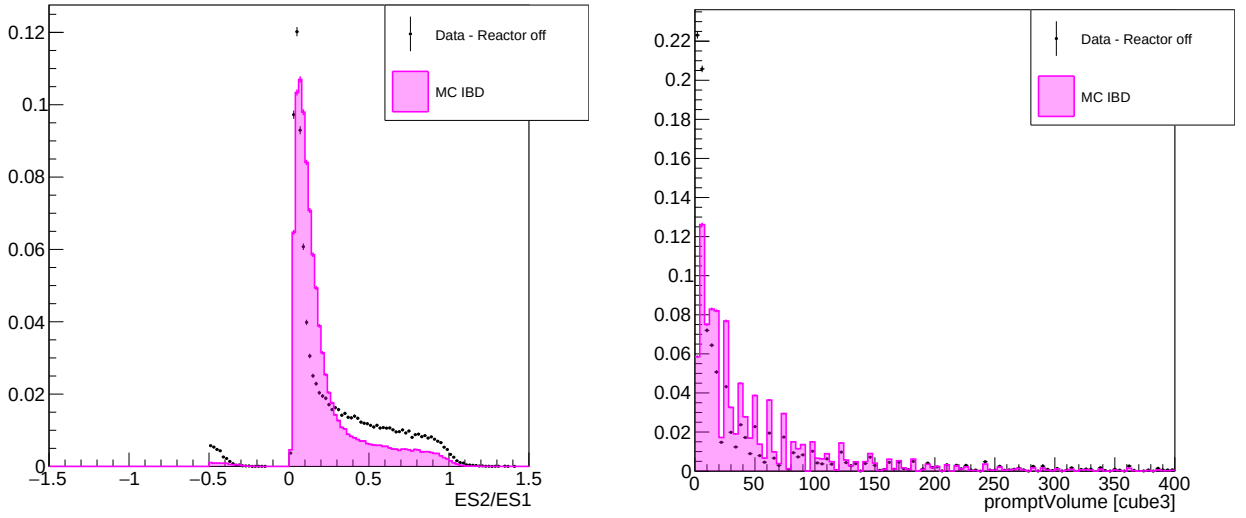
Indeed this ratio differs between the signal and the BiPo. For the signal, the most energetic cube is given by the positron with, in average, an energy of 3 MeV. The second most energetic cube is given by



(a) E_{prompt} **(b)** ΔX_{ES-NS} **(c)** ΔY_{ES-NS}

(d) ΔZ_{ES-NS} **(e)** ΔR_{ES-NS} **(f)** BiPonisher

Figure 2.20: Distributions of each background component: accidental, BiPo and cosmic compared with the simulation of the signal. Each histogram is normalised to the unity.



(a) $\frac{E_2}{E_1}$

(b) Prompt Volume.

Figure 2.21.: Discrimination power between alpha (BiPo) interaction and alpha plus tritium (neutron capture on ^{6}Li) interaction in the ZnS scintillator.

the interaction of an annihilation gamma with an energy between 0.1 to 0.5 MeV. So we can expect a ratio around 0.1. Concerning BiPo events, the second most energetic cube is usually given by a radiative decay with an energy around 1 MeV. Then this ratio will be higher than for the signal.

Then, the prompt volume variable is computed. It gives an idea of the spatial extension of reconstructed CCubes in the ES cluster. For each direction X, Y and Z, we compute:

$$\delta_X = \max X_i - \min X_i, \tag{2.20}$$

where i represents all reconstructed cubes in the ES cluster. Then the prompt volume is :

$$\text{Prompt Volume} = \delta_X \times \delta_Y \times \delta_Z \tag{2.21}$$

Figure 2.21 shows the distribution of the reactor off after the pre-selection applied on those two variables.

Combining all those informations, a signal selection has been built. This selection is given by the equation 2.22.

$$S^{ref}_{Signal} = \begin{cases} E_{cluster} \in [2,7]\text{MeV} \\ \Delta T_{ES-NS} \in [1,141]\mu s \\ \Delta X_{ES-NS} \in [-2,2]\text{cubes} \\ \Delta Y_{ES-NS} \in [-1,2]\text{cubes} \\ \Delta Z_{ES-NS} \in [-1,3]\text{cubes} \\ \Delta R_{ES-NS} \in [0,3.4]\text{cubes} \\ \text{BiPonisher} > 1.44 \\ \text{Prompt volume} \in [0,297] \\ \frac{E_2}{E_1} \in [0,0.37] \end{cases} \tag{2.22}$$

Usually, reactor experiments also use a muon veto. It consists on removing IBD candidates for which a reconstructed muon has been found in coincidence with it. This helps to reduce the background induced by the passage of a muon in the direct detector's environment. The efficiency of such criteria has to be studied since it also introduces an additional source of dead time. Such analysis has been performed by the collaboration and details are shown in [9]. Figure 2.22 shows the distribution of the time difference between the closest reconstructed muon and an ES-NS coincidence, $\Delta T_{\mu-\text{NS}}$. The NS time is taken as time reference. This is obtained from a specific selection of atmospheric events applied on reactor-off data. We can see that by applying an isolation veto of 200 µs, around 10% of coincidences are removed and we introduce a dead time of 6%. We thus conclude that applying such veto is not really useful. This low efficiency is explained by the way events are reconstructed. Indeed the proton recoils due to fast neutrons produced by muons happen only a few microseconds after the passage of the muon. Thus, due to the cluster creation, the signals coming from those energy deposits are mainly merged into the muon cluster. Since we form coincidences only from ES and NS clusters (and not from muon clusters), the cluster identification described earlier actually acts as a veto.

2.5.3.4. Performance

With the selection described on this section, the collaboration reached a performance of 110 antineutrinos selected per day (corresponding to an IBD efficiency of 9.2%) with a signal over background of 0.06. The comparison with other experiments shows that the SoLid experiment has an important background issue. Most of Short Baseline experiments reach a SoB around 1. The factor 17 in SoLid disfavor is understood as the unexpected contamination of ZnS layer. The BiPonisher variable helps, but not enough. In addition, the detector is not able to discriminate a nuclear recoil from an electronic recoil, unlike some of the liquid scintillator based experiments. Therefore, cosmic backgrounds are

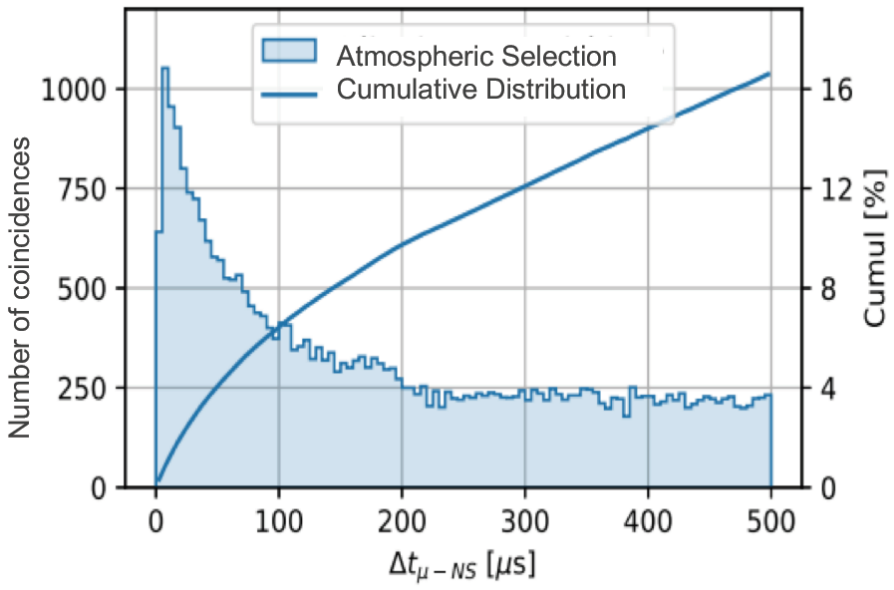
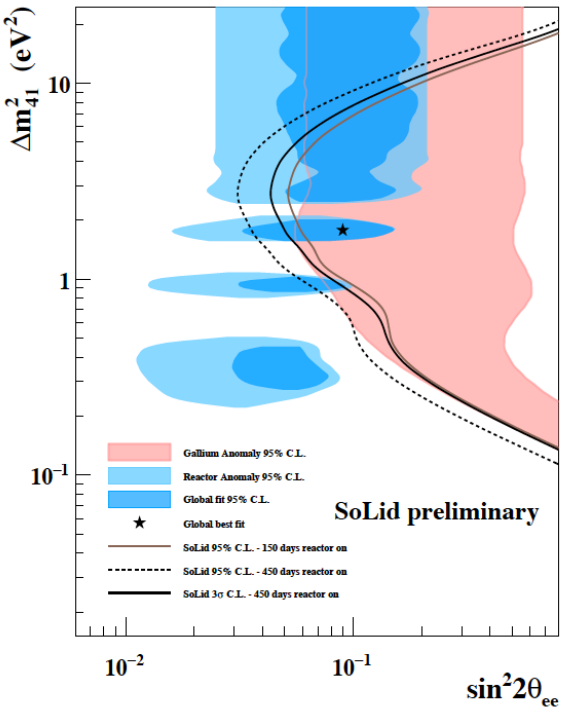


Figure 2.22.: Distribution of $\Delta T_{\mu\text{–NS}}$ between the closest muon cluster and the NS cluster from a coincidence. This is obtained from reactor-off data by applying a specific selection of atmospheric background. From [9]

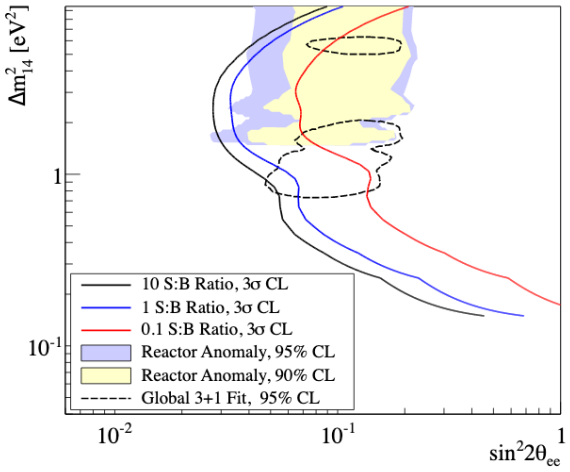
also very high. Chapter 5 is dedicated to the improvement of the signal selection by taking advantage of annihilation gammas topology.

2.5.4. Sensitivity to sterile neutrino oscillation

The sensibility contour estimated during the preparation of the experiment to a sterile neutrino oscillation presented was assuming a signal over background of 3:1. The factor between this expectation and the more realistic evaluation described in this chapter proves the absolute necessity to improve the selection if one wants SoLid to exclude the RAA best fit. Indeed as shown in [76], the impact of the signal over background value is important on the sensitivity contour. Figure 2.23b shows this impact. We have to interpret this figure qualitatively since it does not correspond to the SoLid detector. For this reason a focus has been made in this thesis to develop new discriminative variables in order to improve background rejection and signal efficiency.



(a) The expected SoLid sensitivity assuming a signal over background ratio of 3:1



(b) Evolution of the sensitivity contour as a function of the signal over background ratio. From [76].

Figure 2.23.: Comparison between the expected sensitivity of SoLid and the evolution of sensibility as a function of the signal over background ratio.

Chapter 3.

Energy calibration of the detector

FINAL FANTASY VII REMAKE Original Soundtrack. *Various Artist*, Square Enix Music, 2020

3.1. Introduction

The manifestation of a sterile neutrino is observed by the variation of the number of electronic antineutrino interacting in the detector as a function of the neutrino energy, E_ν, and the baseline L_ν. However, a measurement is always the convolution of the true distribution with the detector response. The signal reconstruction and selection also impact the shape of the (E_ν ; L_ν) distribution and depend on the detector's response as well. Therefore, in order to allow a sensitive and reliable search for a sterile oscillation, this response must be known precisely. In particular, effects depending on the energy or the position of the interaction in the detector - due to inhomogeneities in the detector response - must be understood and corrected as far as possible.

The SoLid collaboration uses a fit to determine if the realisation observed in the data is in favour of the null hypothesis (absence of oscillation) or not. This is done by fitting to the observed (E_{vis}) distribution a set of simulated distributions, that vary in the values assumed for the sterile parameters (no oscillation, or various values of θ_{ee} and Δm_{14}). This procedure relies on the ability of the simulation - in particular the simulated detector's response - to reproduce the data. In particular, if the reconstructed energy is biased, the effect on the oscillation measurement can be cancelled if the same bias is present in the simulation. The energy response of the simulation is tuned from the values extracted during the calibration procedure. So it requires to be confident with the calibration work. Therefore, the role of SoLid's calibration work is not only to provide constants to the reconstruction algorithms to translate reconstructed signals into MeV and correct for inhomogeneities but also to calibrate the simulation of the detector's response. The calibration procedure has been developed to fulfil both requirements.

The amplitude measured by the detector is expressed in electronic unit, the ADC (Analogue to Digital Conversion). Calibration work converts this amplitude in natural units, the MeV.

The function describing the conversion from electronic unit to energy is called the energy scale. The quality with which the energy scale is determined in every cube is therefore important. It will be explained in section 3.2.2, that it is a challenge in such a segmented detector. Also, calibration measurements at several energies are necessary, since the energy scale in a scintillator is in general affected by a quenching effect [78]. A strong feature of the PVT chosen by SoLid is to display a linear response as a function of the energy deposit, that contrast with other experiments. It must be verified.

In this chapter the methods developed or improved along the thesis in order to extract all the parameters needed to model the energy response of the detector are described. The outline is the following:

- An overview of the calibration work performed along the thesis, with an emphasis on the challenges to be faced in the case of a detector like SoLid.
- The calibration tools.
- A description of the DAQ and event reconstruction algorithms employed to acquire calibration data.
- A description of the methods used to evaluate the yield of scintillation light collected per unit of energy deposited in a cube (the visible Light Yield, LY_{vis}).
- A description of the methods used to measure the fibre attenuation length and the fibre to SiPM "coupling" that accounts for the loss of light at the optical coupling between each SiPM and the corresponding fibre.
- An overview of the calibration results after a typical calibration campaign.
- A proposal to determine the LY in the case of low energy deposit, where a dedicated complex treatment is necessary.

The main contributions brought by this thesis work are listed below:

- Development of a new method to extract the LY with the ^{22}Na calibration source to complement the original method.
- Study of the effect of the selection and reconstruction criteria on the energy spectrum of the gamma calibration, and effect of the energy loss due to their interaction upstream the cube to calibrate; incorporation in the above mentioned LY methods.
- Development of a new method to extract the attenuation lengths and the coupling efficiencies, compared study of its performance and of that of original approach.
- Development of an analytical selection efficiency model to extract the light yield of low energy sources.

3.2. Overview of the calibration work

3.2.1. Goals of the calibration

The two goals of the calibration are to convert the amplitudes expressed in electronic unit into natural units, to evaluate the detector's response, and to tune the simulation response. It is achieved assuming a model of the light propagation, from the interaction of particles in the scintillator to the number of PAs detected by the sensors. Figure 3.1 shows a schematic summarising the successive stages between the deposit of energy in the scintillator by an ionising particle and the production of a Photon-Avalanche (PA) in the SiPM.

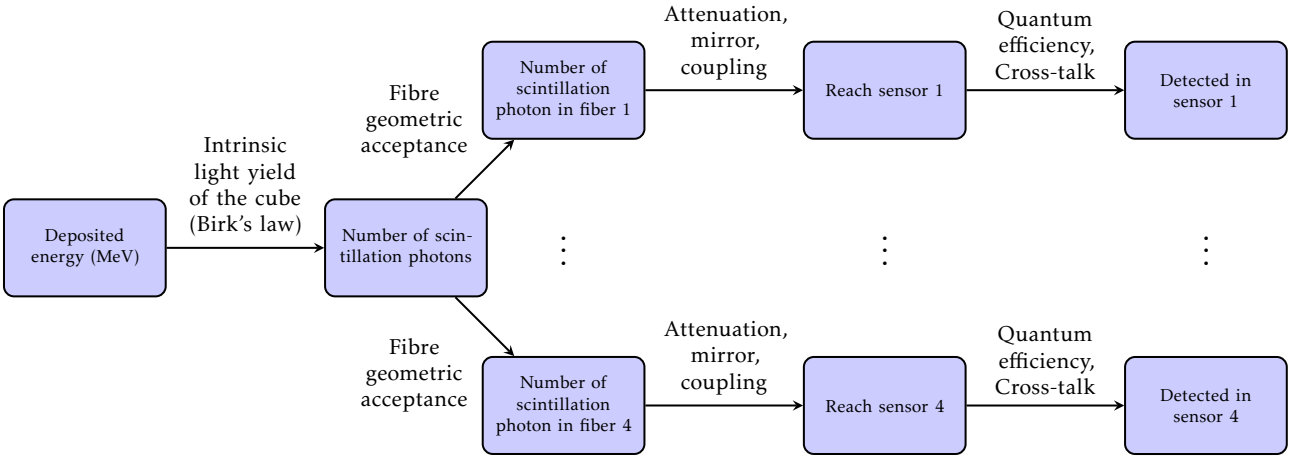


Figure 3.1.: Scheme summarising the successive stages between the deposit of energy in the scintillator by a ionising particle and the production of a Photon-Avalanche (PA) in the SiPM. Once scintillation photons have been trapped in a fibre, the same pattern is seen on each of the four fibres reading out a cube.

This path of light is explained in the context of calibration runs. The calibration particles are gammas which interaction in materials like PVT (low Z) is dominated by Compton scattering. The scattered electron excites the molecules composing the plastic scintillator. The number of scintillation photons produced is proportional to the intrinsic light yield of the scintillator which is about 10,000 photons per MeV. It actually depends on the quenching effect [78], described by the empirical Birks law:

$$\frac{dL}{dx} = L_0 \frac{\frac{dE}{dx}}{1 + k_B \frac{dE}{dx}} \tag{3.1}$$

where $\frac{dL}{dx}$ is the light scintillation yield per length unit, L_0 is the light scintillation yield in the linear regime, $\frac{dE}{dx}$ is the stopping power of the ionising particle and k_B is the Birks constant which depends on the type of scintillator. Scintillation photons propagate within the cube until they are absorbed by the scintillator, reflected by or absorbed in the Tyvek sheet, or captured by a fibre. The probability of a photon to reach a fibre is called the fibre geometric acceptance, noted ϵ_{acc}.

Once the photons are trapped by a fibre, they propagate through it. Along the propagation there is a probability for them to be absorbed. This attenuation effect is quantified by the attenuation length constant of the fibre, noted λ_{att}. From the manufacturer, the fibres are expected to have an attenuation length of 101 cm. Then there are two cases according to the direction in which photons are going. First, photons can go directly to the sensor at one end of the fibre. In this case, one must take into account the coupling efficiency, noted $\epsilon_{coupling}$ between the fibre and the sensor. Alternatively, photons can propagate to the other end of the fibre and have a probability to be reflected by the mirror. This probability is noted ϵ_m and is assumed to be equal to 80%. Then the photons propagate back in the other direction until the SiPM and the coupling efficiency is applied.

Once a photon reaches the SiPM, a Photo-Avalanche (PA) may occur, as described in section. Several electronic effects have to be taken into account to obtain the detected number of PAs as a function of the number of photons reaching sensors. The first effect is the Photon Detection Efficiency (P.D.E. or quantum efficiency) of the sensors, it governs the probability of a scintillation photon reaching the sensor to induce a PA. Considering the generation of MPPC used in the detector and the operation settings, the value is around 32%. Then due to the MPPC geometry there is a probability when a PA is induced to trigger a neighbouring cell. This is called the crosstalk effect and the probability to induce another PA is noted μ_{CT}. It is evaluted around 17%. (see section 2.2.3.2)

Based on this description, the goals of the calibration are therefore to determine 19,200 parameters from data in order to correctly characterise the detector response in energy and tune the simulation. Indeed there are:

- 12,800 cube light yields (LY_{vis}) .
- 3,200 attenuation lengths (λ_{att}).
- 3,200 coupling efficiencies (ϵ_c).

It is a huge challenge to determine all those parameters with enough accuracy.

SoLid's analyses are presently based on the reconstruction of cubes. Thus it requires to identify the list of cubes in which interactions happen. Then, to determine the energy, the straightforward variable is the sum over the fibre amplitudes. Taking into account this reconstruction and the model previously described, one can express with equation 3.2 the cube amplitude as a function of the parameters described above:

$$
\begin{aligned}
A_{cube}(X,Y) &= A^{Hori,1}_{Fibre}(Y) + A^{Hori,2}_{Fibre}(Y) + A^{Vert,1}_{Fibre}(X) + A^{Vert,2}_{Fibre}(X) \\
&= E_{dep} \times LY^{intstrinsic}_{cube}(X,Y) \times \epsilon_{acc} \times \left(\epsilon^{Hori,1}_{att,dir}(Y) + \epsilon^{Hori,1}_{att,indir}(Y) \times \epsilon^{Hori,1}_{m}\right) \times \epsilon^{Hori,1}_{coupling} \\
&+ E_{dep} \times LY^{intstrinsic}_{cube}(X,Y) \times \epsilon_{acc} \times \left(\epsilon^{Hori,2}_{att,dir}(Y) + \epsilon^{Hori,2}_{att,indir}(Y) \times \epsilon^{Hori,2}_{m}\right) \times \epsilon^{Hori,2}_{coupling} \\
&+ E_{dep} \times LY^{intstrinsic}_{cube}(X,Y) \times \epsilon_{acc} \times \left(\epsilon^{Vert,1}_{att,dir}(X) + \epsilon^{Vert,1}_{att,indir}(X) \times \epsilon^{Vert,1}_{m}\right) \times \epsilon^{Vert,1}_{coupling} \\
&+ E_{dep} \times LY^{intstrinsic}_{cube}(X,Y) \times \epsilon_{acc} \times \left(\epsilon^{Vert,2}_{att,dir}(X) + \epsilon^{Vert,2}_{att,indir}(X) \times \epsilon^{Vert,2}_{m}\right) \times \epsilon^{Vert,2}_{coupling}
\end{aligned}
\tag{3.2}
$$

where:

- E_{dep} is the deposited energy in the cube, corrected from the Birks law.
- $\epsilon_{att,dir}$ and $\epsilon_{att,indir}$ are the attenuation functions for photons going respectively directly to MPPC or indirectly.

3.2.2. Challenges

The particular features of SoLid impose the calibration to overcome a number of challenges, often related to its high segmentation.

Number of calibration constants

The first challenge has already been cited. It is the number of parameters that should be determined in order to correctly model the detector response. In a first version, only LY constants were determined (LY constants for the Readout simulation) and generic values were used for the attenuation lengths and couplings, identical for all fibres. This was enough for the first phases of SoLid's analysis, since the goal by then was to prove the ability to see a neutrino excess (enough signal and sufficient S/B) and not yet to perform a sterile analysis proper, which requires the best possible calibration. As will be explained in section 3.6, the latter takes to measure also the attenuation lengths and couplings. Note that the LY determination methods and results presented in section 3.5 are valid whatever the treatment of attenuation lengths and couplings.

Pile up and ambiguities

The planes are designed such that one fibre serves sixteen cubes. It means that the amplitude measured at the end of a fibre depends on the light emitted by all those cubes. This causes *pile-up* when simultaneous energy deposits occur in several cubes sharing fibres, or *ambiguities*, when energy deposits occur in at least two cubes from different rows or column. The latter causes the reconstruction of additional "ghost" cubes, where fibres cross. In case of pile-up or ambiguities, the reconstruction of cubes position and energy is not straightforward (see 2.4.2.3). This reconstruction needs to separate

the contributions of the various cubes to the signal seen on a fibre, and to better identify ghost cubes. For this, it is necessary to compare the signals seen on orthogonal fibres. A good description of this orthogonal information is one of the motivations to measure fibre parameters. Indeed, useful signals in SoLid cannot be restricted to cube signals, made of the combination of signals seen on the 4 fibres and calibrated as an autonomous energy estimator, by a single constant. Individual fibre signals matter since cubes cannot be seen as isolated entities.

Selections are used in section 3.5.1 to discard pile-up or ambiguous cubes, since they could lead to biases in the determination of LY's. For instance, one of them keeps only events in which the cube to calibrate is the only one in its plane with a signal detected above threshold. The efficiency of these selections must be determined. It depends on the efficiency of other fibres in the plane to individually see a signal, thus on their individual attenuation length and coupling.

Usage of gamma sources at restricted positions

The calibration of the detector requires to deploy a radioactive source able to reach enough statistic in all cubes. But the specific detector geometry implies some mechanical constraints. During the data acquisition, the planes should be in contact to keep the target homogeneous. The strong point of the detector is its segmentation and it requires to have a good alignment of all planes. Thus moving one by one each plane during a calibration campaign to insert a radioactive source between each cube of planes is not possible. The calibration facilities are discussed in section 3.3.1 and the solution chosen by the collaboration is to move modules (10 planes). Then moving modules means that the calibration particles should be able to travel through five planes. Thus, gamma sources are well suited. Those particles are indeed penetrative. It exists different isotopes producing mono-energetic gammas with a high activity. The need of high activity sources to perform a fast calibration requires a DAQ able to manage the data rate. Section 3.3.2 details the sources used for the energy calibration and the specific DAQ configurations. The usage of sources causes several difficulties. The size of the cube implies that only a part of the incoming energy is deposited in a cube, where, typically, a calibration photon undergoes only one Compton scattering. Thus the true energy spectrum is not a simple peak but a Compton spectrum. This makes the determination of the visible light yield and the resolution more complicated. Moreover two effects may bias the measurement. First, gammas interacting in cubes away from the source have to travel through the detector. They sometimes interact before the cube of interest. Thus, the true deposited energy spectrum is distorted from the theoretical one. Second, the energy reconstruction and the event selection have an impact on spectrum as well. The section 3.5 explains in detail the energy reconstruction and the two approaches used to determine the visible light yield and the resolution and their validations.

Interdependence between cubes and channels

As will be described in section 3.6, the attenuation length of a fibre is measured by comparing cube LYs measured in cubes located at various positions along this fibre. These LYs also depend on the

signals measured on transverse fibres. Therefore, the value of the attenuation length found for a fibre depends on the quality of the reconstruction on other fibres (ex : well calibrated or not). This kind of situations create correlations between channels, between cubes, between cubes and channels when measuring their attenuation lengths, couplings and light yields. Moreover the determination of fibre parameters is a degenerate problem since there are more parameters than measurements, therefore some arbitrary choices had to be made. The methods used to determine the fibre parameters have then been tested extensively to detect potential biases. This analysis is discussed in section 3.6.

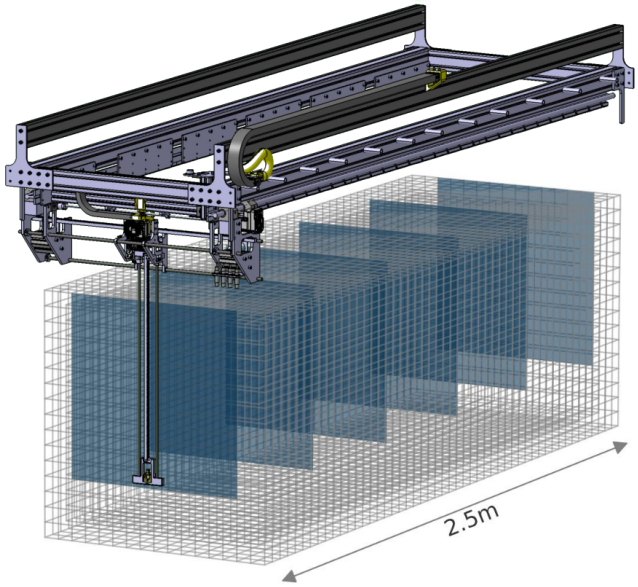
Light readout with optical fibres

The last constraint is the choice to use plastic scintillators read-out by fibres. It allows to have a highly segmented target, but the visible light yield is lower than in detectors based on liquid scintillator, due to the fibre acceptance. This complicates the analysis of low energy deposits observed in the experiment such as annihilation gamma interactions. As an example, considering a gamma with an energy of 511 keV, the average of the deposited energy spectrum by Compton scattering is 174 keV. Assuming a visible cube light yield of 96 PA/MeV it means that on average fibre will see 4.2 PAs. In this regime, the random variable describing the fibre amplitude is not following a Gauss law but is Poissonian. This also complicates the light yield measurement for low energy sources, crucial to prove the linearity response of the detector. Thus a preliminary method has been developed and is described in the annex A.

3.3. The calibration system

3.3.1. Calibration robot: CROSS

A robot has been designed to transport calibration sources inside the detector, called CROSS. At the start of a calibration run, modules move in order to leave a 3 cm "gap" between the modules to calibrate. In addition to 4 intern modules, two gaps at the front and rear of the first and fifth modules do not necessitate such movement. The CROSS automat then places itself above the desired gap. CROSS is made of 4 elements (see figure 3.2). The first one is a vehicle fixed to a pair of rails, situated above SoLid's modules, along which it can move. Between the two parts attached to the rails is the second element. It is held by horizontal rails, which allows to move left or right. It contains a retracted arm which, once the desired horizontal position is reached unfolds down in the gap to reach the desired vertical position. The fourth element is the source holder positioned at the end of the arm. Due to mechanical constraints the source cannot be placed in front of each cube of the detector, the nine positions chosen to perform the calibration are described in section 3.4.1.



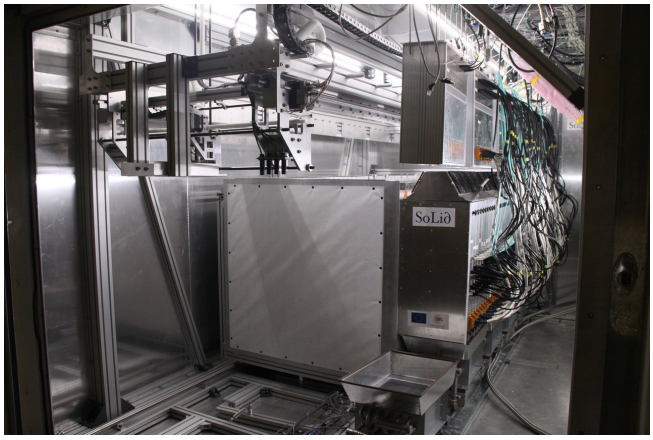


Figure 3.2.: *Left:*Schematic of the calibration robot CROSS above the detector. The blue shadow represents the reachable area of the source. Six gaps allow to calibrate all the cubes of the detector. *Right:* Photograph of the interior of the container. On the top left there is the CROSS robot.

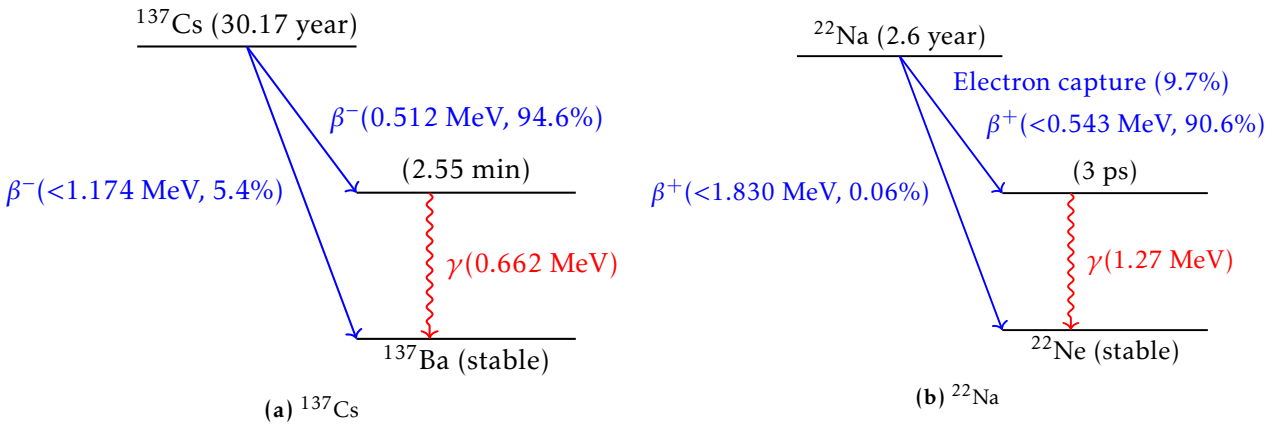
Sources	Activity [kBq]	Gamma energy [MeV]	Compton edge energy [MeV]
^{137}Cs	3	0.662	0.477
^{207}Bi	37	(0.57 - 1.06 - 1.8)	(0.39 - 0.85 - 1.6)
^{22}Na	37	(0.511 - 1.274)	(0.341 - 1.054)
AmBe (n-H)	-	2.2	1.994
AmBe(γ)	1	4.4	4.198

Table 3.1.: List of available gamma calibration sources and the associated Compton edge energy and the trigger used during data taking.

3.3.2. Sources

Gamma sources have advantages:

- **Penetrating power:** The mean free path of γ depends on its energy, in PVT at 1 MeV it is around 14 centimetres. Then even in the farthest cube a sufficient statistics of Compotn scatterings can be gathered, with a manageable time exposure.
- **Various energies:** Many sources exist and with several energies up to 4.4 MeV for AmBe.
- **Mono-energetic:** Contrarily to β sources, one can easily find γ sources emitting mono-energetic particles. Thus it is easier to find and fit a reference point in the energy spectrum.



(a) ^{137}Cs **(b)** ^{22}Na

Figure 3.3.: Decay schematic of the ^{137}Cs and ^{22}Na radioactive sources.

The list of available gamma sources with their activities, energies of emitted particles and Compton edge energies is given in the table 3.1. The collaboration has chosen the ^{22}Na as the principal source. This choice has been motivated by the Compton edge energy around 1 MeV and the emission of annihilation gammas which could be used to study the detector response at low energy. The decay schematic of the ^{22}Na source is given in 3.3. It also shows that of the ^{137}Cs, as this source has been used to test a method to extract the visible light yield at low energy in the annex A. The other sources (^{137}Cs,^{207}Bi and AmBe) are used to prove the linearity of the PVT energy response in most of the range of the IBD energy spectrum. Above 4 MeV there is no source available, this energy region is controlled using muon crossing the detector as explained in the section 2.4.2.1.

3.4. Description of the data acquisition

3.4.1. Data acquisition

3.4.1.1. Source positions

Once the sources have been chosen, the calibration process has to be designed. The source positions, the trigger and the run time duration have been decided by considering those constraints:

- **Time duration:** When a source is inside the detector there is no possibility to take physics runs. Both reactor-on and reactor-off days are needed to perform physics analyses. Thus a calibration campaign should be fast.
- **Mechanical constraints:** The calibration robot CROSS positions the calibration sources inside the calibration gaps. Ideally, it would place the source, on demand, in front of any cubes of the two planes that then surround the source. However, due to mechanical constraints all cubes are

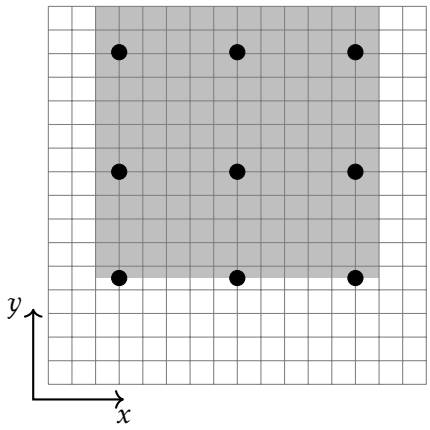


Figure 3.4.: Schematic of source calibration positions chosen (black points). The gray filled area represents the area reachable by the calibration robot.

not reachable. Indeed the container height limits the CROSS's arm length. Also it was possible to set only nine source calibration positions in the software operating the robot.

- **DAQ limitations:** The available sources have a high activity (see table 3.1) which the DAQ is not able to hold. Tests have been performed in order to obtain a compromise between the data rate and the threshold applied to trigger. This was an important source of difficulty during the studies performed to tune the readout simulation (see chapter 4).

With all those considerations, the nine positions have been determined and are shown in the figure 3.4.

3.4.1.2. Calibration triggers

Different settings were used for the trigger, depending on the source or the types of data we wanted to acquire. All the triggers described in the following are based on the three triggers used for physics data and described in 2.2.4.2. The first calibration of the full detector happened in March 2018 using the ^{22}Na source. The goal was to acquire enough data in all the cubes within a limited time duration. At this time, the trigger cuts could be applied only per channel, avoiding the possibility to perform a horizontal-vertical coincidence for example. The threshold trigger was defined as: one channel with an amplitude above 9 PAs found anywhere in the detector. In this case the triggering plane plus or minus two planes around it and during two time blocks (12.8 μs) were readout. Due to the source activity and the quantity of data to readout, the detector was most of the time in deadtime (90%). This high deadtime had a big impact on the duration needed to reach enough statistics in all cubes as it took seven days to complete this calibration campaign.

After this first calibration campaign, the DAQ settings were updated in order to increase the neutron trigger efficiency. The SiPM's over-voltage were increased from 1.5 to 1.8 V, with the consequence to also increase the visible cube light yield. In parallel, a firmware update allowed to base the trigger

decision on the presence in any plane of one coincidence between an horizontal and a vertical channel:

$$A_{\text{Horizontal}} > 7.5 \text{ PAs and } A_{\text{Vertical}} > 7.5 \text{ PAs} \tag{3.3}$$

We participated in the tuning of this cut, and to the studies that led to an adapted DAQ for the calibration. The threshold trigger during ^{22}Na calibration is then defined as:

- Only ± 5 planes around the source are able to trigger and record data.
- Only the triggered plane is recorded.

With these new settings, the data rate has been reduced. This allowed to reduce the dead time enough for the regular calibration campaigns. The request to have a plane *pre-scale* has been made just after the first calibration campaign. The goal is to weight the probability of a plane to trigger and thus to reduce dead time without having to use the high thresholds mentioned above, which prevent from studying the response to low energy (± 300 keV) deposits with regular calibration campaign data. Further reducing the dead time would also have been precious since it was still high enough to complicate the studies related to the readout simulation tuning (see chapter 4). Unfortunately, this solution would have requested an intervention in the firmware and no competent person was available for that.

Those new trigger settings were tuned in September 2018, after the first reactor-on/reactor-off transition had been recorded with the new DAQ settings. It is this period which is analysed in chapter 5. This time, it took only 1 day to obtain enough statistics, which was a big improvement. In order to demonstrate the linear response of the detector, few runs were taken using the other sources and the random trigger. Between September 2018 and the end of Phase-1 in June 2020, the detector was calibrated seven times using the ^{22}Na source with the threshold trigger described above. During four campaigns we experienced issues with CROSS, imposing us to delay the calibration of one or several gaps of a few days. The last calibration campaign, held in June 2020, was foreseen to be the longest, with, in particular, a large amount of runs using the periodic trigger for data/MC studies. Unfortunately, CROSS was able to reach only one gap.

The calibration using other sources such as ^{137}Cs, ^{207}Bi and AmBe used the two other types of trigger implemented in DAQ: the neutron trigger and the random trigger. The random trigger has been used for the visible light yield measurement at low energy, in particular with the ^{137}Cs source which emits a gamma of 0.662 MeV. There, the threshold trigger would totally distort the Compton edge spectrum. It has also the advantage of providing unbiased data very useful in order to carry out data/MC comparisons as shown in chapter 4. The association of the neutron and the threshold triggers have been used to perform the visible light yield measurement at higher energy with the AmBe source, which emits a gamma of 4.4 MeV in coincidence with neutrons.

The list of calibration campaigns performed during the whole operation of Phase-1 is shown in table 3.2.

Date	Gaps	Source	Trigger condition	Duration [days]	Notes
March 2018	1-6	^{22}Na	Threshold	7	
Modification of DAQ settings					
August 2018	1	^{22}Na	Threshold	1	Issue with CROSS to reach other gaps
September 2018	1-6	^{22}Na	Threshold	1	
	3	^{22}Na	Random	<1	
	2	^{137}Cs	Random	<1	
	2	^{207}Bi	Random	<1	
	2	AmBe	Threshold	1	
October 2018	1-6	^{22}Na	Threshold	1	
December 2018	1-5	^{22}Na	Threshold	1	Issue with CROSS to reach gap 6
January 2019	1-6	^{22}Na	Threshold	1	
Beginning May 2019	1-6	^{22}Na	Threshold	1	Issue with CROSS to reach gaps 5 and 6
End May 2019	1-6	^{22}Na	Threshold	1	
February 2020	1-6	^{22}Na	Threshold	1	Issue with CROSS to reach gap 6
June 2020	5	^{22}Na	Threshold	1	Issue with CROSS to reach other gaps.
	5	^{22}Na	Random	3	
	5	AmBe	Physics	3	
	5	^{207}Bi	Random	2	

Table 3.2.: Table summarising all the calibration campaigns performed during Phase 1 operations.

3.4.2. Event reconstruction

During this thesis, two different cube reconstructions have been used by the SoLid collaboration. The first one, called SCube reconstruction, is the historical one. It consists first on looking for fibre signals coincident within 75 ns, to build a "SEvent". The cubes where some of these fibres cross are SCubes. The SCube amplitude is simply the sum of the signal amplitudes measured on all such fibers. In the case where the signals on the 4 fibres serving the cube are above the reconstruction threshold, this writes:

$$A_{cube}(X,Y) = A_{Fibre}^{Hori,1}(Y) + A_{Fibre}^{Hori,2}(Y) + A_{Fibre}^{Vert,1}(X) + A_{Fibre}^{Vert,2}(X). \tag{3.4}$$

This approach is limited by the pile-up and ambiguities (ghost cubes) issues described in 3.2.2. This is why a second approach was developped by the SoLid group at Laboratoire de Physique de Clermont (LPC, in2p3). It is the CCube approach described in section 2.4.2.3. It is the canonical one in SoLid's analysis. Therefore, the constants are obtained by analysing calibration energy distributions reconstructed with this method.

In both methods, the input to the energy reconstruction is the amplitudes of the waveforms measured on the fibres that readout the cube. It is expressed as a number of PAs after the values in ADC counts have been converted thanks to the gain measurement (see section 2.2.3) and, ideally, corrected for the losses due to attenuation along the fibre and to the fibre to SiPM coupling. In the first version of the CCube algorithm, used in this thesis, the attenuation effect was not accounted

for. It is not straightforward since the input to this algorithm is a set of fibre signals, not a list of cubes. Therefore the distance between the energy deposits and the SiPMs is not known. Instead, the attenuation effects are absorbed in the cube LY. In the next version, attenuation will be included from the start, and in every cube will be re-evaluated. The methods described in section 3.5 are transparent to this and can simply be applied again. The same comment applies each time refined values of fibres' attenuation lengths and couplings are obtained.

In the case of the determination of the light yield implemented in the readout simulation (chapter 4), both methods can be used. Data/MC comparisons are involved there. Due to a certain number of difficulties described in chapter 4, we often prefered to use the simplest reconstruction, therefore SCubes. It makes the interpretation of data/MC discrepancies easier, while an accurate simulation of basic signals like (sums of) fibre amplitudes should also result in an accurate simulation of the CCube energy.

3.5. Methods to determine the cube light yield

This section presents the two methods developed in order to extract the visible cube light yield, using the ^{22}Na source.

3.5.1. Events Selection

Ideally, to determine LY in a cube, we need to rely on energy distributions that depend only on energy deposits that occured in this cube. They must not be affected by the pile-up and ambiguity effects described in 3.2.2, which depend on the feature of other cubes and fibres. The CCube algorithms are helpful there. But they cannot be assumed to be perfect at this stage. Therefore, two selections have been used.

The first one is used by the so-called Kolmogorov approach described in section 3.5.2. It relies on the CCube reconstruction as far as pile-up effects are concerned, but removes ghost cubes from energy distributions by applying a selection on a variable called fibre asymmetry. It is based on the fact that for a real deposit in a cube, the four fibres reading out this cube all see in average the same amount of light. Thus the idea is to compute the asymmetry between the amplitudes observed on the two horizontal fibres and the two vertical one as:

$$\mathscr{A}_{HV} = \frac{A_{Fiber}^{Hori} - A_{Fiber}^{Vert}}{A_{Fiber}^{Hori} + A_{Fiber}^{Vert}} \tag{3.5}$$

where $A_{Fiber}^{Hori} = A_{Fiber}^{Hori,1} + A_{Fiber}^{Hori,2}$ and $A_{Fiber}^{Vert} = A_{Fiber}^{Vert,1} + A_{Fiber}^{Vert,2}$. Under the previous assumption this variable is close to zero for a real deposit and different from zero for a ghost cube. It was decided to

apply the following cut:

$$|\mathscr{A}_{HV}| < 0.6. \tag{3.6}$$

This cut also reduces, to some extent, the probability that a second energy deposit occurred on the same cube row or column, therefore the pile-up effect.

Concerning the other approach, called analytical approach, described in section 3.5.3, the selection is more drastic: the cube must be the only one in its plane with 4 fibres seeing a signal above 4.5 PAs. Modulo the efficiency of this threshold, we therefore isolate here cases with no pile-up effects.

Note that the choice to use a certain selection with a certain approach is not strongly imposed by any specific features of these approaches. It could have been the other way around. However, the first selection is more efficient, and therefore better suited to the Kolmogorov approach since it based on G4 samples which generation is time consuming. Having one selection per method was also a way to control the impact of remaining pile-up and ambiguity effects, in case they are large.

3.5.2. The Kolmogorov approach

The first method is called the *Kolmogorov* approach. It has been developed by the SoLid team at IJClab. The idea is to produce a realist prediction of the distribution of the true energy deposits, using the Geant4 simulation. This distribution is subsequently re-scaled by the LY and convoluted with the resolution function given by:

$$f(x;LY,\sigma_0) = \frac{1}{\sqrt{2\pi}\sigma_0\sqrt{T}} e^{-0.5\frac{(\frac{T}{LY}-x)^2}{\sigma_0^2 T}} \tag{3.7}$$

where T is the true deposited energy in MeV, x is the cube amplitude in PA and σ_0 is the assumed resolution at 1 MeV. In all the manuscript we assume a simple resolution model:

$$\frac{\sigma_E}{E} = \frac{\sigma_0}{\sqrt{E}} \tag{3.8}$$

This treatment provides a proxy that can be fitted to the distribution measured in real data - the consistency between the real distribution and that proxy is evaluated via a Kolmogorov-Smirnov test - In each cube, many such proxy distributions are generated, with varying LYs and resolutions values. The values that maximise the probability of the K-S test are returned as the measured LY and resolution at 1 MeV.

3.5.3. The analytical approach

This method assumes that the calibration γ interacts by Compton scattering only once in the cube to calibrate. Under this assumption the deposited energy spectrum is well known and has been derived

by Klein and Nishina [79]:

$$\left(\frac{d\sigma_{KN}}{dT}\right) = \frac{\pi r_e^2}{m_e c^2 \alpha^2}\left(2 + \left(\frac{T}{E_0 - T}\right)^2\left(\frac{1}{\alpha^2} + \frac{E_0 - T}{E_0} - \frac{2}{\alpha}\left(\frac{E_0 - T}{T}\right)\right)\right) \tag{3.9}$$

where E_0 is the energy of the incident photon, T is the energy of the scattered electron, r_e is the classical radius of the electron, α is the fine structure constant and $m_e c^2$ is the electron's mass.

Beyond the Compton Edge (CE), given by $E_{\text{CE}} = E_T(\max) = E\left(1 - \frac{1}{1+\frac{2E}{m_e c^2}}\right)$, the value of this cross-section is null.

Here too a convolution is applied to obtain a proxy to the reconstructed energy spectrum. This convolution is done numerically following the equation:

$$f(x) = \int_0^{E_{CE}} \frac{d\sigma_c}{dT}(T) \times \frac{1}{\sqrt{2\pi}\sigma_0\sqrt{T}} e^{-0.5\frac{(\frac{x}{LY_{vis}} - T)^2}{\sigma_0^2 T}} dT, \tag{3.10}$$

where x is the amplitude in PA. The visible light yield LY_{vis} and the resolution at 1 MeV σ_0 are the two free parameters to determine. This is not actually a *probability density function* since the integral over x is not equal to the unit. One can see that the Gaussian part is already normalised, it is only the Klein-Nishina formula which has to be normalised. So the *p.d.f.* is given by the following equation:

$$f(x) = \frac{\int_0^{E_{CE}} \frac{d\sigma_c}{dT}(T) \times \frac{1}{\sqrt{2\pi}\sigma_0\sqrt{T}} e^{-0.5\frac{(\frac{x}{LY_{vis}} - T)^2}{\sigma_0^2 T}} dT}{\int_0^{E_{CE}} \frac{d\sigma_c}{dT}(T) dT}. \tag{3.11}$$

3.5.4. Motivation for using two fit methods

One of the aims defined by the SoLid collaboration concerning calibration was to understand the cubes energy scale at 2% or better. Therefore, a precise and reliable fit of the visible light yield is crucial. A first difficulty in this task has already been mentioned: the necessity to fit the complex distribution described in 3.5.2 and 3.5.3 rather than a narrow peak. It could not be considered as obvious a priori that it would lead to an unbiased estimate of the visible light yield, even assuming the p.d.f. is correct. Moreover, the fit models described in 3.5.2 and 3.5.3 are approximations. They should therefore lead to a bias. Using two independent fit approaches is a way to control this, especially since they are based on different assumptions. The main assumption of the Kolmogorov approach is that the geometry of the detector and of the CROSS system is perfectly described in the Geant4 simulation, and that, consequently, the way to share the energy of a ^{22}Na 1.274 MeV calibration gamma among cubes is correct. The analytical approach assumes nothing on the geometry. Thus, the

difference between the visible light yield's measured by either methods is most probably a conservative estimation of the associated bias. Also, with the analytical method assuming a single interaction per cube for each photon, the comparison with the Kolmogorov approach also covers imperfection in interactions description by Geant4 (although we think them small).

The issues due to pile up events and ambigious cubes have already been described in section 3.5.1. Having one specific selection to treat such events for each fit approach is a way to control the remaining effect. This is also a way to constrain another effect: the efficiency of these selections is energy-dependent and must be modelled in equation 3.11, so that the pdf of the energy distribution is still correct in the presence of this distortion of the spectrum. Selection efficiencies are modelled using the simulation. It could be imperfect, leading to a bias in the measured visible light yield. An example of such an energy dependence is the following. In a pile-up event, the multiple energy deposits are often due to the same gamma, which was scattered with an angle around 90 degrees when it interacted in the first hit cube. The scattering angle is correlated with the energy transferred to the electron. At this angle, the energy transfer is about 800 keV. Therefore, when requiring that no other cube be reconstructed in the same plane, one depletes the middle region of the deposited energy distribution in the cube to calibrate.

Not all sources of bias are independent between the two methods. The main example here is the role of energy losses between the source and the cube to calibrate. The Geant4 distribution used in the Kolmogorov approach includes automatically this effect. In the case of the analytical approach, energy distributions are compared between the closest cube to the source and the cube to calibrate to give a correction function to be included in equation 3.11. Therefore, both methods depend on the Geant4 simulation: a similar bias should affect them and should not be detected by the comparison. This is explained in chapter 4.

Finally, two groups comparing and discussing their results was a source of improvements, understanding, bug correction, etc...

3.5.5. Validation of the analytical method with toys

The second method has been studied in more details at Subatech to understand the performance using toys and simulations. The performances of the fit are defined based on the ability to extract the visible light yield value assumed when generating such samples.

First, 1000 samples of 10000 pseudo calibration events have been generated, where 10000 is the typical statistics expected in a cube during a calibration run. For each pseudo event, a true energy was sampled from the Klein-Nishina distribution, and modified by a shift drawn from the Gaussian in equation 3.10, where typical values for the visible light yield and the resolution where assumed. The p.d.f. in equation 3.11 was then fit to each sample producing 1000 pseudo-measurements of LY and of the resolution.

The results are given on figure 3.5. The extraction of the assumed light yield is good with almost no bias and an error below the precent. The resolution is determined with a small bias of $(2 \pm 1)\%$ and a 3% precision. This test proves that with the typical statistics expected in a calibration, enough information can be extracted from this distribution to measure the parameters of interest accurately, if the p.d.f. is the correct one.

A more realistic study has also been performed, in which the true energy spectrum is given by the Geant4 simulation. In this case the gamma may interact before, and several times in the cube of interest (see section 3.5.6).

Results are shown on figure 3.6. From the previous study the average of the reduced chi-square distribution is higher, this is expected since the *p.d.f.* is not correct. On both LY and the resolution a bias appeared, 2.1% and 14%, respectively. This bias comes from the double interaction of gamma in the cube of interest. The figure 3.7 shows the comparison of those two distributions. One can see that in the Geant4 distribution, a population above the Compton edge appears. This leads to the observed biases. Since they are understood, they can be corrected.

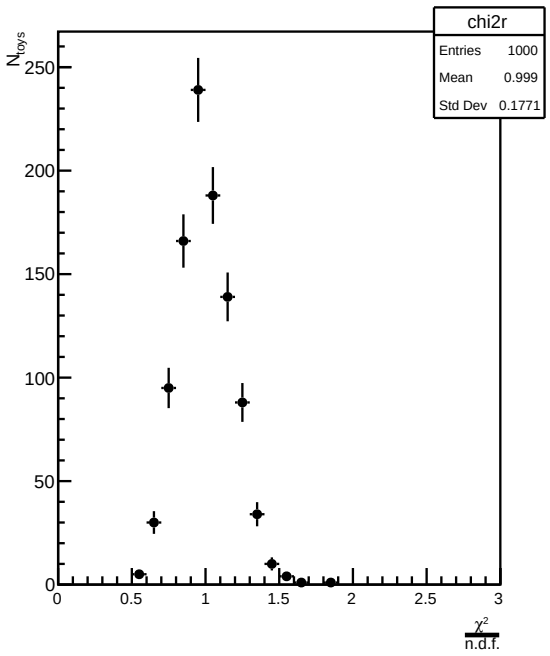
To conclude: the analytical approach can extract the visible light yield at the sub percent level and with no bias if a correction of 2.1% is applied to take into account multi-scattering events which overestimate the measured visible light yield. This will be confirmed in a study reproducing the calibration of one entire gap of the detector and using the complete simulation (Geant4 and the readout simulation) in section 3.5.6.

3.5.6. Corrections for energy losses before the cube of interest and for the selection efficiency

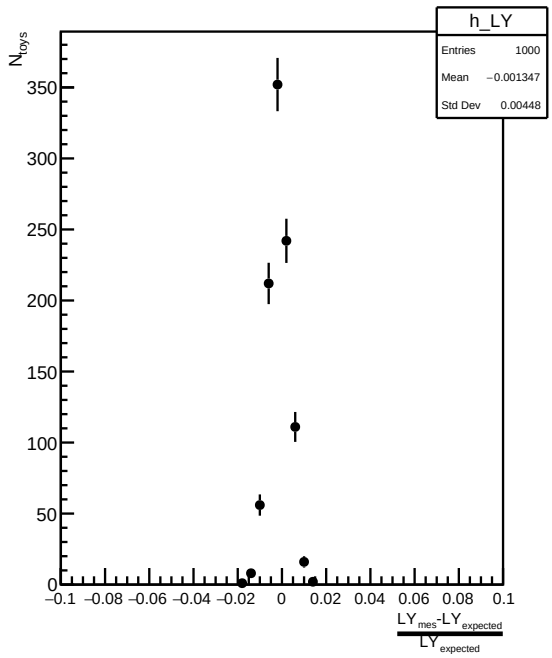
As mentioned before the majority of cubes are not directly in front of the source and the calibration gamma has to travel through a part of the detector until reach those cubes. The probability to have interacted there and lost a part of its energy is not negligible and the shape of the deposited energy spectrum is therefore distorted.

This effect has been studied using the simulation of the detector. For each cube, the deposited energy spectrum is compared to the cube in front of the source used as reference. Thus the ratio of the two spectra gives how it is distorted due to gamma propagation through matter. An example of this ratio is given on figure 3.8. We can see that the ratio decreases as a function of the true deposited energy. This indicates that the reference cube has more event at high energy than the cube of interest. One can see that the CE region is depleted in the farthest cube with regard to the reference one. It means that fewer gammas are interacting at the CE in this cube than in the reference. We also observe a drop in the region of annihilation gamma CE (around 0.340 keV) for the same reason there are less annihilation gammas reaching the farthest cube than the closer one.

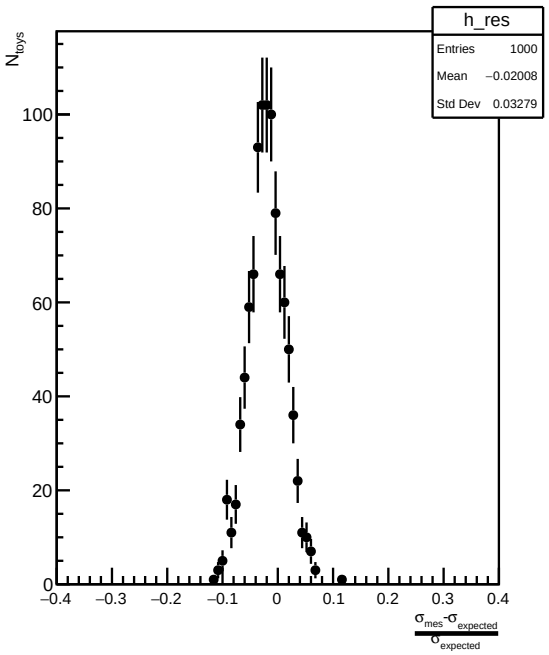
One also needs to account for the efficiency of the reconstruction and the selection described in section 3.5.1. Using the same simulation this efficiency has been studied too. The selection is applied



(a) Reduced chi-square distribution.

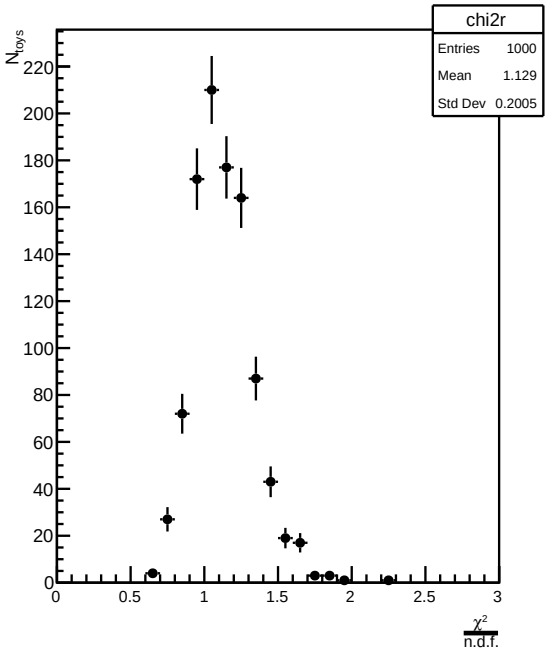


(b) Pull distribution of the fitted visible light yield.

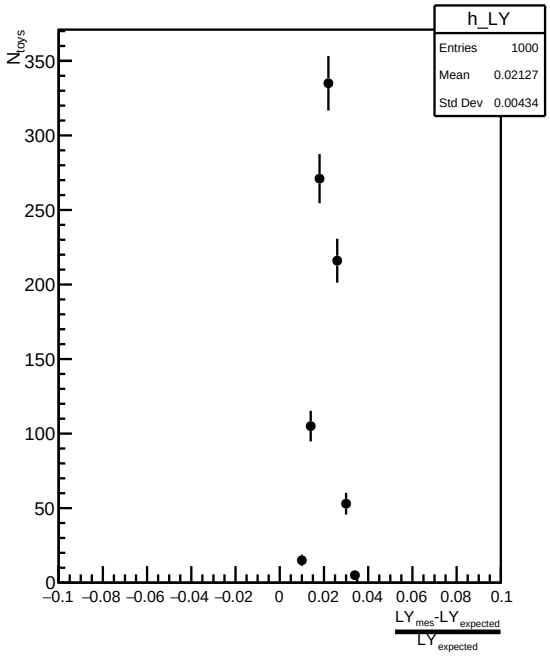


(c) Pull distribution of the fitted resolution.

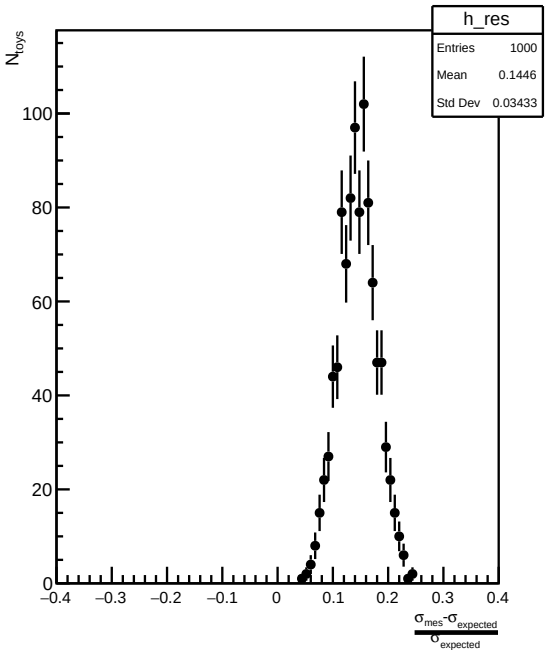
Figure 3.5.: Analytical fit performance for toys generated with the Klein-Nishina cross-section.



(a) Reduced chi-square distribution.



(b) Pull distribution of the fitted visible light yield.



(c) Pull distribution of the fitted resolution.

Figure 3.6.: Analytical fit performance for toys generated with the Geant4 deposited energy distribution in a cube in front of the source.

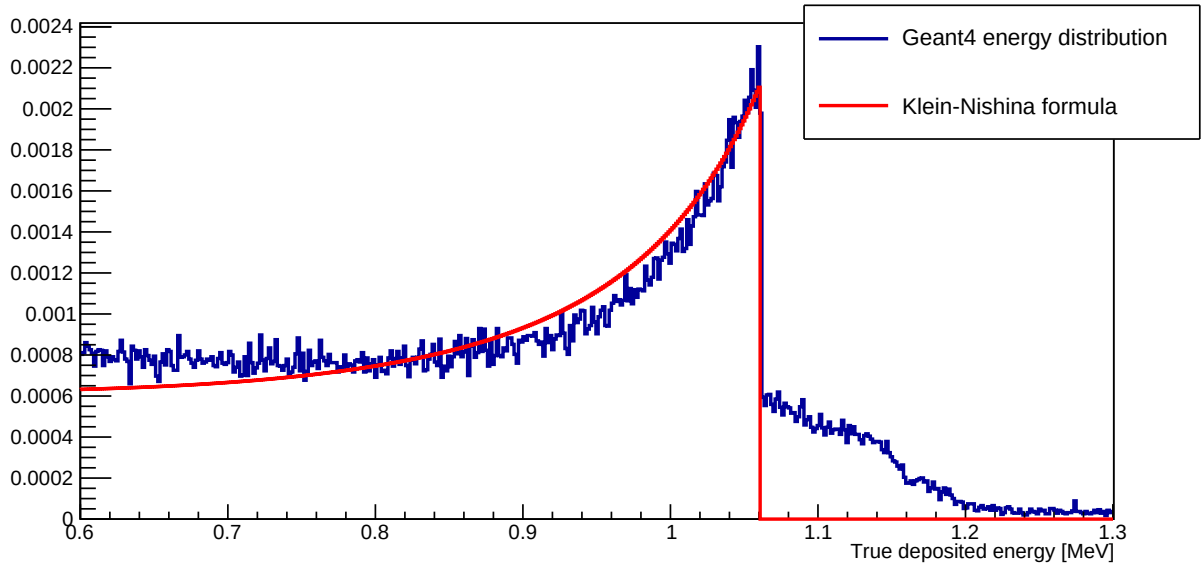


Figure 3.7.: Comparison between the true deposited energy distribution from Geant4 simulation in a cube in front the source and the Compton scattering differential cross-section described by the Klein-Nishina formula considering the 1.274 MeV gamma from ^{22}Na. The population above the CE, at 1.054 MeV, is due to the multiple interaction of the incident gamma in the cube of interest.

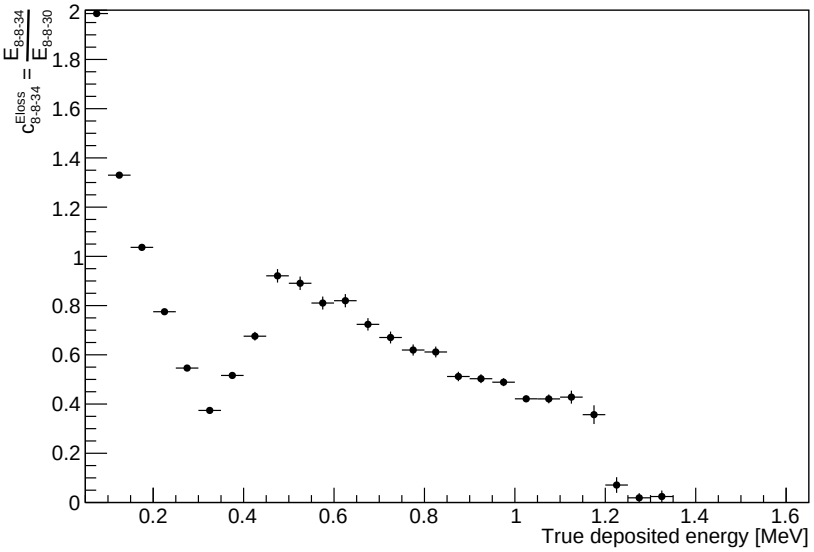
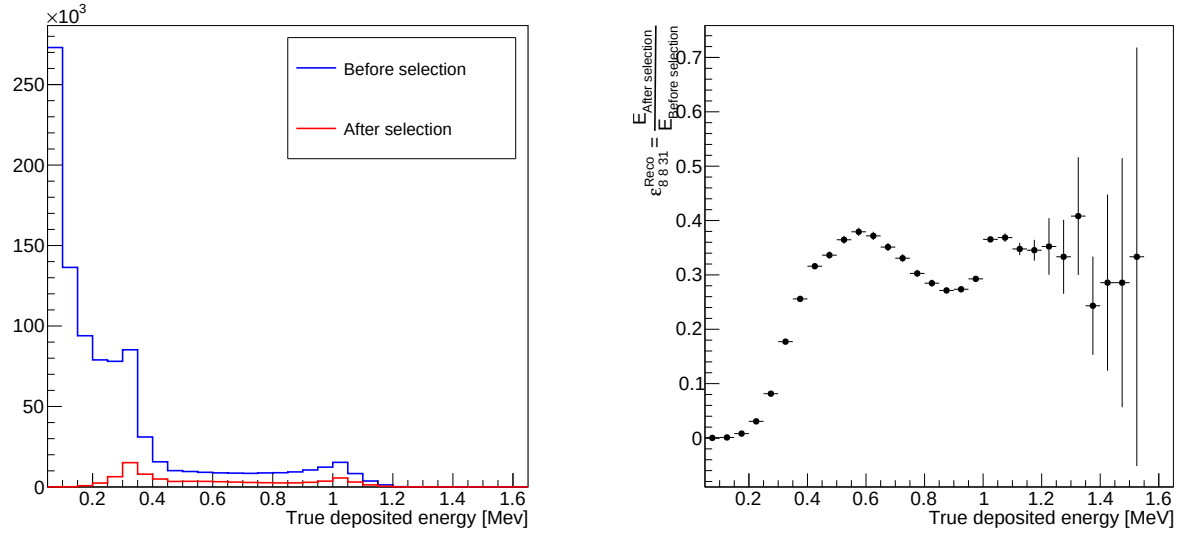


Figure 3.8.: Ratio of true deposited energy spectrum between the cube (8-8-34) and its reference cube (8-8-30) giving the energy loss correction factor c^{Eloss}_{8-8-34}



(a) True deposited energy distribution with and without the selection applied: the cube should be alone in its plane.

(b) Selection efficiency as a function of the true energy deposited.

Figure 3.9.: Reconstruction efficiency of a cube.

after the read-out simulation and a correspondence with Geant4 files allows to know what has really happened in the cube of interest. Thus a comparison between the true deposited energy spectrum before and after the selection (described in 3.5.1) on reconstructed quantities gives the reconstruction and selection efficiency. This is shown on the figure 3.9. The ratio between the true deposited energy distribution after and before the selection in shown on the figure 3.9b. At low energy, the shape is driven by the 4-fibres requirements (see the annex A where the selection efficiency is studied in more details). Above 600 keV a maximum is reached, from this energy the probability to have a cube with 4 fibres is close to the unity and only the isolation criteria is playing a role. An interesting structure appears at intermediate energies. It is due to the relation between the energy deposited by a Compton scattering and the scattering angle (see figure 3.10). This region corresponds about to $\frac{\pi}{2}$, the angle for which the scattered gamma has a chance to deposit energy in the same plane. A larger impact of the isolation cut is therefore expected in this region.

Corrections have been derived for all cubes of a gap. So during the convolution the Klein-Nishina cross-section is weighted with those efficiencies and the equation 3.11 becomes:

$$f(x) = \frac{\int_0^{E_{CE}} c^{Eloss}(T) \times \varepsilon^{Reco}(T) \times \frac{d\sigma_c}{dT}(T) \times \frac{1}{\sqrt{2\pi}\sigma_0\sqrt{T}} e^{-0.5\frac{(\frac{x}{LY}-T)^2}{\sigma_0^2 T}} dT}{\int_0^{E_{CE}} c^{Eloss}(T) \times \varepsilon^{Reco}(T) \times \frac{d\sigma_c}{dT}(T) dT} \tag{3.12}$$

where:

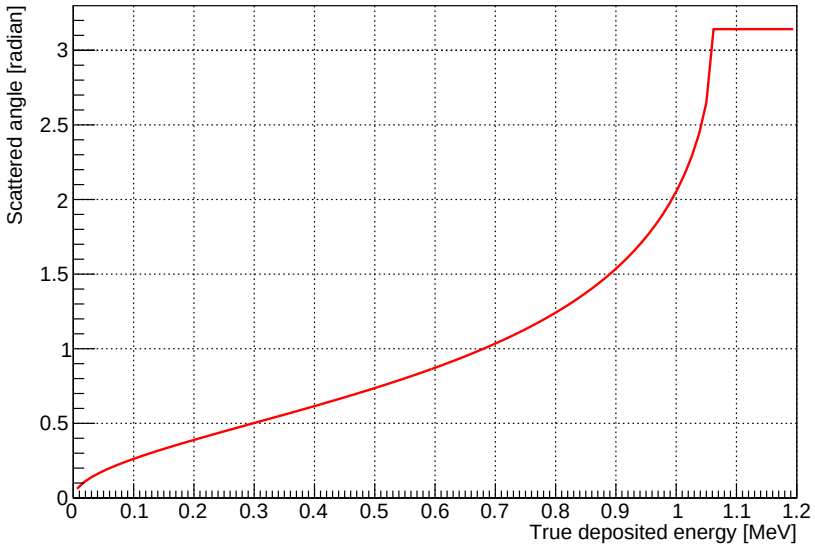


Figure 3.10.: Scattered angle as a function of deposited energy for a gamma of 1.274 MeV interacting by Compton scattering.

- $c_{Eloss}(T)$: is the correction term due to interactions upstream the cube of interest, derived from the figure 3.8
- $\varepsilon_{Reco}(T)$: is the reconstruction efficiency derived from figure 3.9.

A check of the impact of efficiencies was performed using the simulation. Electrons of 1 MeV have been generated in all cubes (with Geant4's particle gun), to measure the expected LY in the simulation. Then the two fitting processes have been applied on ^{22}Na simulation and the comparison between the measured LY and the expected one has been computed.

Figures 3.11 and 3.12 are showing the results of this study without and with the corrections in the equation 3.12. We show these results in 2560 cubes from the five planes on either sides of the gap where the source is placed. Without corrections, the fitted LY with the analytical approach is dropping with the distance of the cube to the source position. This effect is solved by the energy loss correction.

On figure 3.13, we show the precision and accuracy with which the expected LY_{rec}'s are reconstructed once the corrections have been applied. In the case of the analytical approach, no bias remains (we also subtract the bias described at the end of section 3.5.5). In the case of the Kolmogorov approach, a 5% bias is visible for the cubes in the closest plane to the source. It was also visible on figure 3.11 (cubes 1050 to 1550). This effect is presently being studied by our IJCLab colleagues. Preliminary results indicate the bias could be due to pile-up events, more frequent in the first plane due to annihilation gammas. The looser cut used in this approach to mitigate such effects (see section 3.5.1) might not suffice. If a strict isolation cut was used, one could expect the bias to be reduced at the same level as in other planes, namely 2%. Note that the differences between the biases affecting

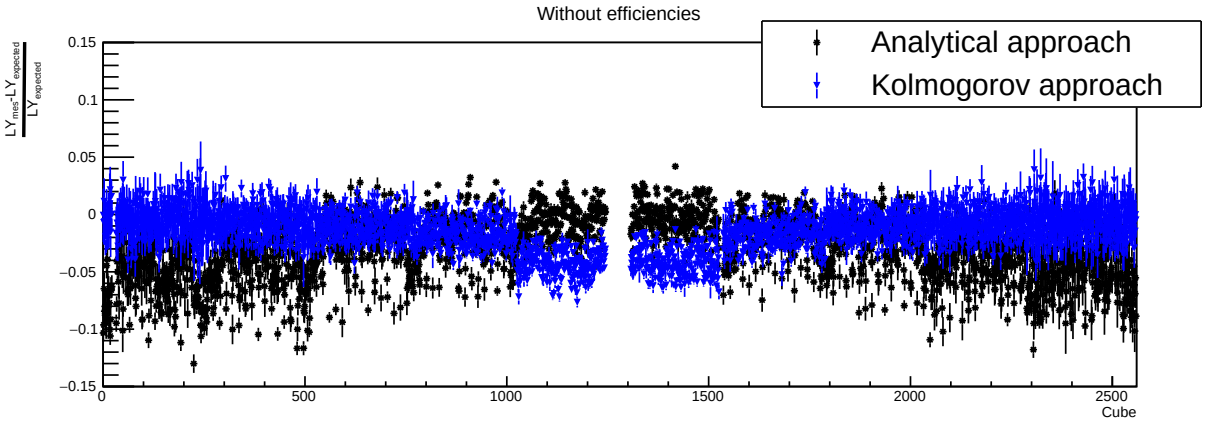


Figure 3.11.: Comparison between the measured visible light yield and the expected one using the full simulation for both the Kolmogorov and analytical approach without efficiencies.

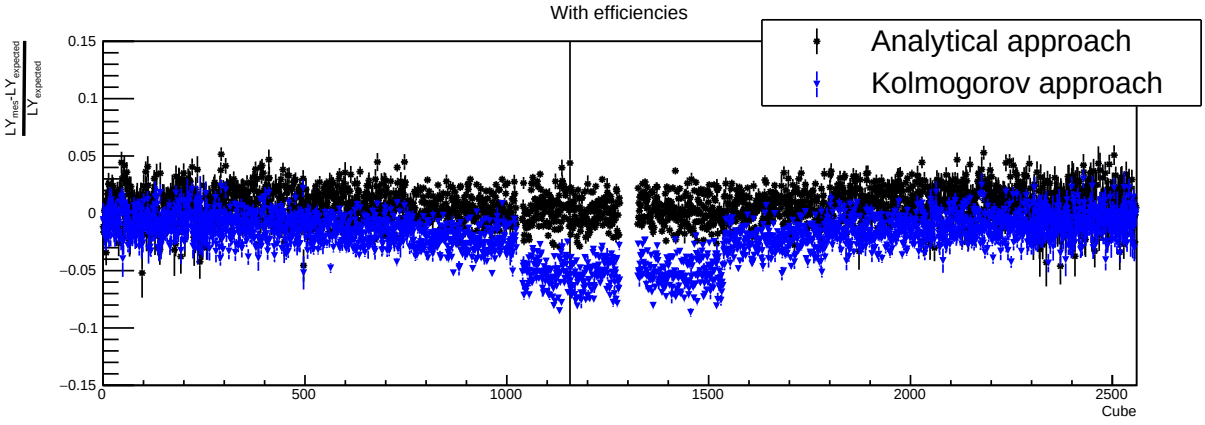


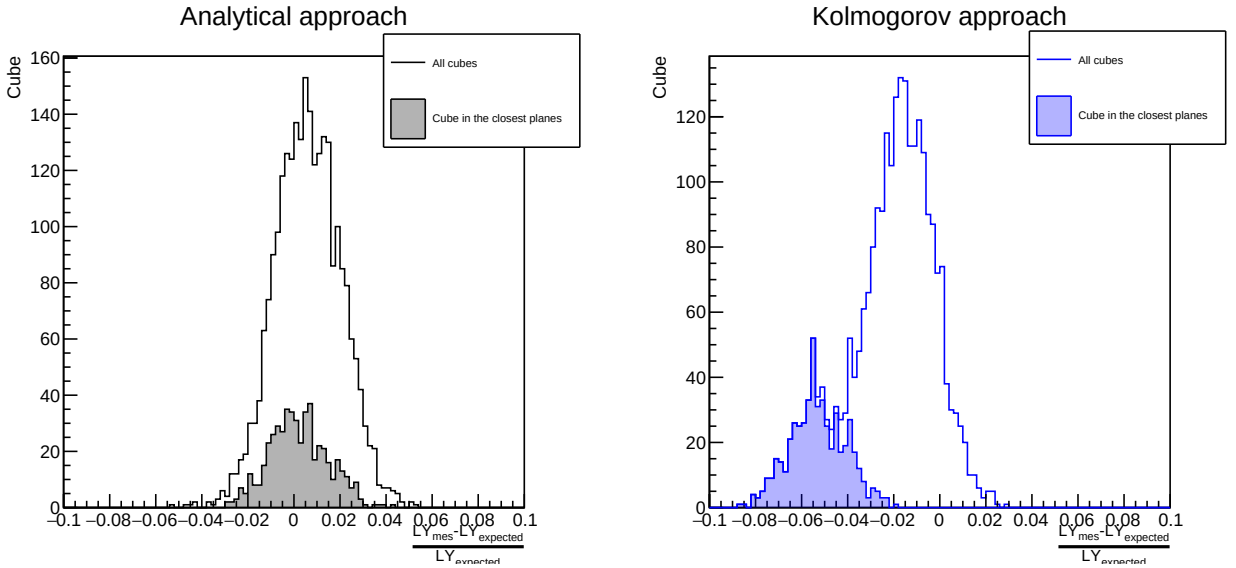
Figure 3.12.: Comparison between the measured visible light yield and the expected one using the full simulation for both the Kolmogorov and analytical approach with efficiencies.

the analytical approach and the Kolmogorov approach are observed as well in real data. Figure 3.17 compares directly the LY_{rec}'s obtained then by both methods.

In conclusion, the typical 2% difference between the two methods, confirmed in real data, sets the scale of the systematic uncertainty due to the fit method assumptions. The issue of this systematic uncertainties on LY are discussed more extensively in chapter 4.

3.5.7. Application on real data

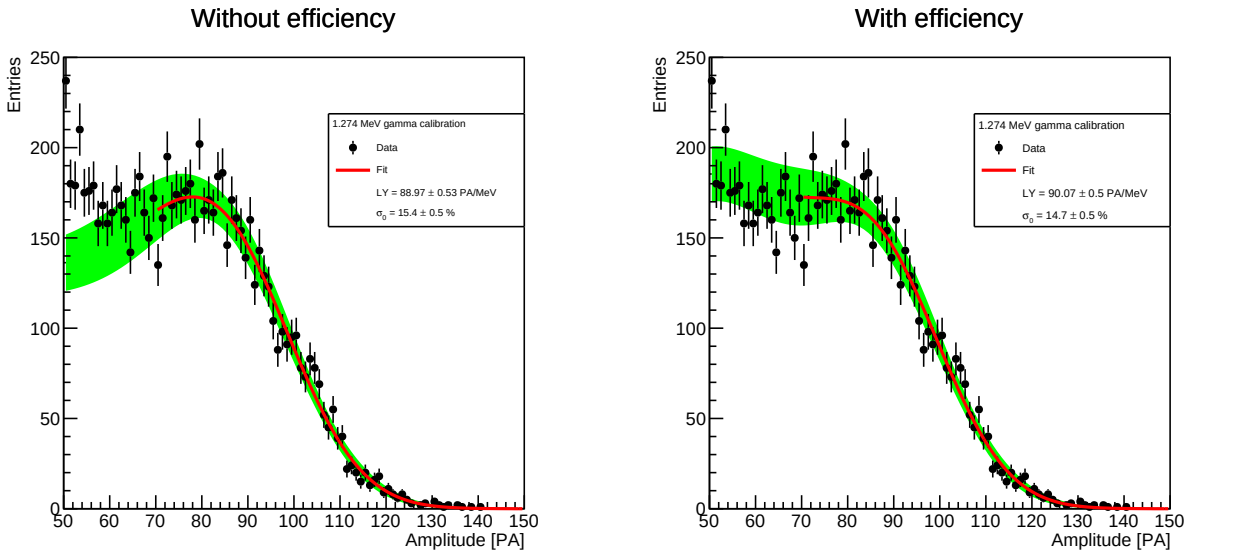
The Kolmogorov and Analytical approaches have been applied on the first calibration campaign with the updated DAQ settings held in September 2018. The figure 3.14 shows an example of cube amplitude distribution fitted with the analytical approach with and without the two corrections discussed in 3.5.6. The cube is positioned in X=7 Y=7 Z=32, two planes away from the source, so the two corrections are playing a role.



(a) Results with the Analytical approach.

(b) Results with the Kolmogorov approach.

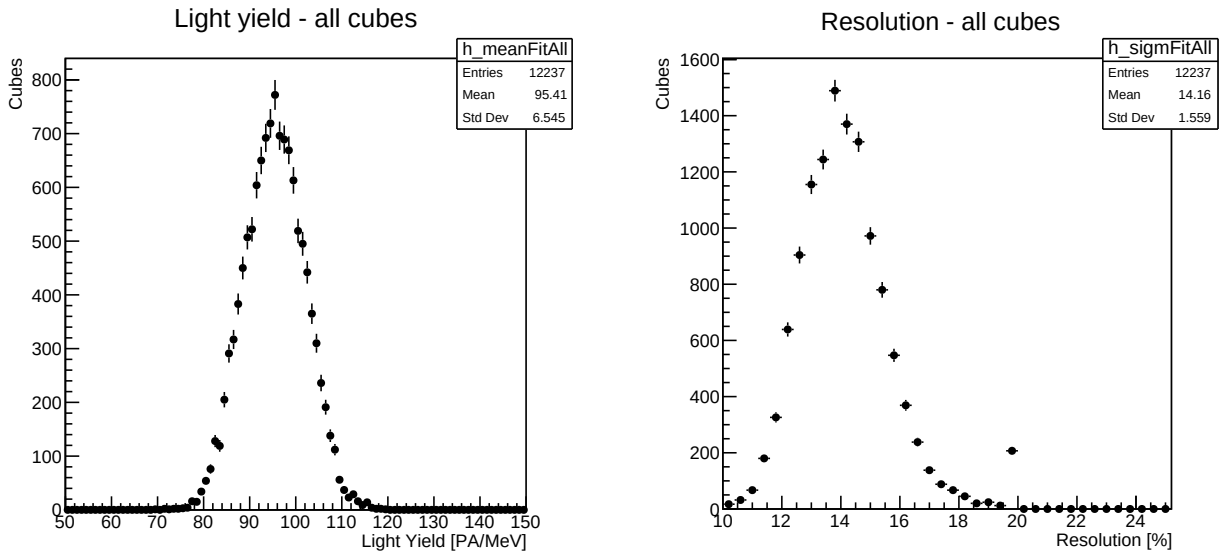
Figure 3.13.: Pull distribution of the measured visible light yield with efficiencies. Filled histograms are the distributions of cubes in the closest planes to the source.



(a) Without efficiencies.

(b) With efficiency.

Figure 3.14.: Analytical fit without and with efficiencies taken into account. We can see the reconstruction efficiency which corrects the *p.d.f.* at low amplitude and the matter effect which increases the measured visible light yield of around 1%. The green shape corresponds to ± 1 σ.



(a) Distribution of the fitted visible light yield.

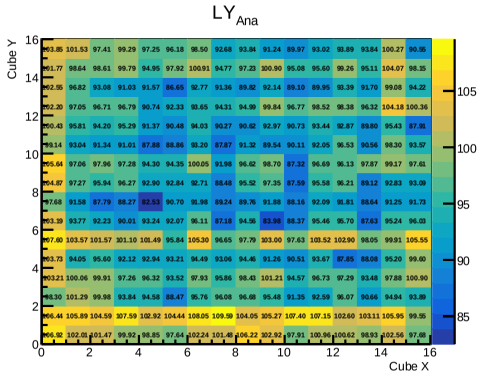
(b) Distribution of the fitted resolution at 1 MeV.

Figure 3.15.: Results of all fitted cube of the detector with the analytical approach.

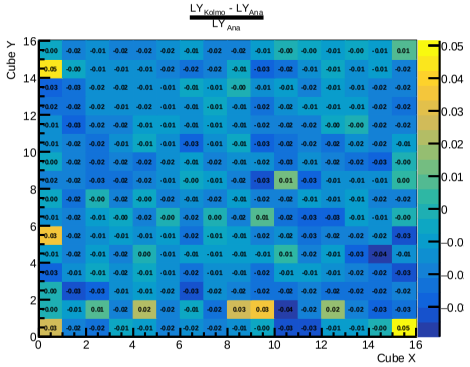
The distributions of the fitted visible light yield and resolution at 1 MeV using the Analytical approach are given on the figure 3.15. In average we found a visible light yield of 95.4 PA/MeV with a resolution at 1 MeV of 14.16%.

The results of a calibration campaign are summarised plane by plane in a set of plots shown on the figure 3.16 for the plane 22. The distribution of the visible light yield and the resolution at 1 MeV fitted with the Analytical approach as a function of the cube position in the plane are given in the figures 3.16a and 3.16d. On the two central plots 3.16b and 3.16e the good agreement between the Kolmogorov and the Analytical approaches is shown. For a few cubes, a discrepancy is observed and it corresponds to cube for which the reduced chi square is higher than expectation. In those cases the fit has not well converged, usually it mostly concerns cubes at the edge of a plane. In those regions, due to the limitation of the calibration source positions the statistic is limited. But the impact is limited, indeed the number of cubes affected is low, the discrepancy is rarely more than 5%, and for physics analyses cubes at plane edges are excluded because of the highest background rates observed.

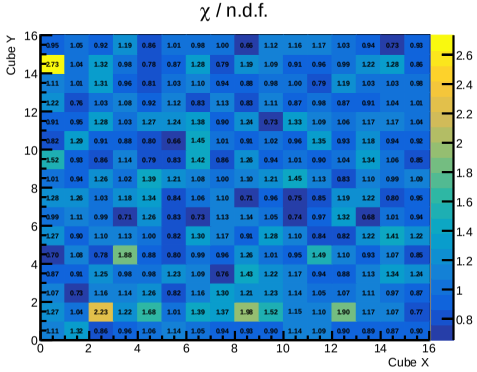
For each plane the difference in LY between the Kolmogorov and the Analytical methods is fitted with a Gaussian. The mean of this fit has been studied as a function of plane distance from the source with and without the corrections discussed in section 3.5.6. This is shown on figure 3.17. If neither the selection and reconstruction efficiencies nor the effects due to energy losses are accounted for, there is a linear evolution of the difference as a function of the plane depth. When they are corrected, this linear behaviour is corrected. The larger difference seen in section 3.5.6 in the first plane is seen here too, confirming the agreement between data and MC on this issue. In other planes, the simulation predicted a shift of around -2% and this is, again, also observed in real data. Those results are a



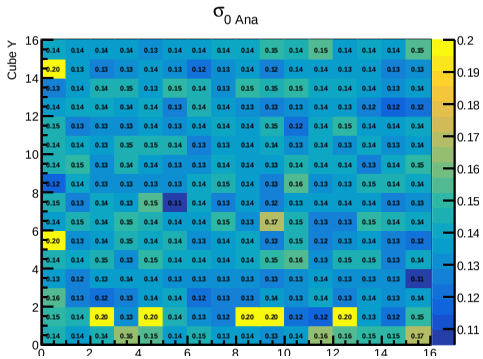
(a) Visible light yield with the Analytical approach.



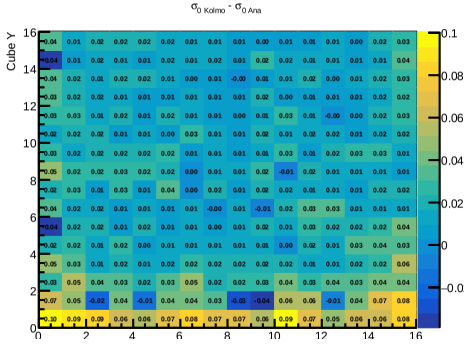
(b) Visible light yield comparison between Analytical and Kolmogorov.



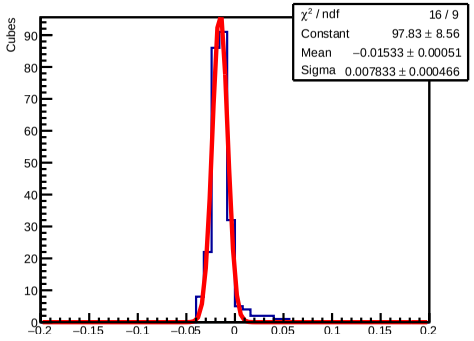
(c) Reduced chi-square of the Anaytical fit.



(d) Resolution at 1 MeV with the analytical approach.



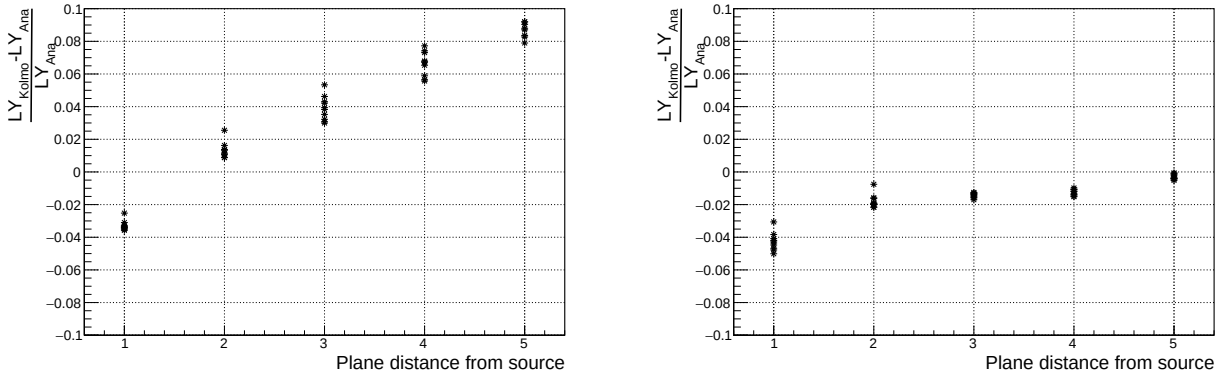
(e) Comparison of the Resolution at 1 MeV between Analytical and Kolmogorov.



(f) Visible light yield comparison between Anaytical and Kolmogorov approaches.

Figure 3.16.: Summary plots of plane 22 during the September 2018 calibration campaign using the ^{22}Na source.

confirmation of the ability of the simulation to reproduce calibration data. Also it gives confidence on the validity of the corrections derived from the simulation.



(a) Without the efficiencies.

(b) With the efficiencies.

Figure 3.17.: Comparison between the visible light yield extracted with the Kolmogorov and the Anaytical approach as a function of the plane distance from the source. Each point is the mean value given by a fit of the plane distribution.

3.6. Method to extract the fibre to SiPM optical coupling ϵ_c and attenuation lengths λ_{att}

In section 3.2.1 the various steps undergone by scintillation photons between their production and their detection via PAs in MPPC was described, and the necessity to measure the fibre attenuation lengths and the fibre to SiPM optical coupling was explained.

The attenuation efficiency affecting the scintillation light emitted in a given cube ϵ_{att} can be expressed as the sum of two terms: one for photons going directly to the MPPC and another one for photons going to the mirror. The number of photons absorbed in the fibre because of the attenuation is modelled by an exponential. Thus the total attenuation efficiency for a given fibre is expressed by Equation 3.13:

$$\begin{aligned}\epsilon^f_{att}(r) &= \epsilon^f_{att,dir}(r) + \epsilon^f_{att,indir}(r) \times \epsilon_m \\ &= \frac{1}{2}\exp\left(-\frac{d_{cube,sensor}(r)}{\lambda_{att}}\right) + \frac{1}{2}\exp\left(-\frac{d_{cube,mirror}(r)}{\lambda^f_{att}}\right)\epsilon_m \exp\left(-\frac{d_{mirror,sensor}}{\lambda^f_{att}}\right)\end{aligned} \tag{3.13}$$

where: r is the position of the cube in the row or column with the convention that $r = 0$ is the closest cube to the SiPM, f is the index identifying the fibre. When we consider explicitly the fibres reading out the same cube, the adopted convention is that $f \in [0,1]$ defined the horizontal fibres and $f \in [2,3]$

defined the vertical fibres. ϵ_m is the mirror reflection efficiency that is considered as constant and equal to 80% as studied in [80], $d_{cube,sensor}$ is the distance between the cube of interest to the sensor, $d_{cube,mirror}$ is the distance between the cube of interest to the mirror and $d_{mirror,sensor}$ is the distance between the mirror and the sensor: the fibre length.

The fibre to SiPM optical coupling does not depend on the cube position contrarily to the attenuation. Thus it is expressed by a constant attributed to each fibre. Putting all together, one can express the amplitude of the waveforms seen in MPPCs as function of cube intrinsic light yield and fibre efficiencies as shown by equation 3.14.

$$A^f(r) = E_{dep} \times LY_{cube}^{intstrinsic}(r) \times \epsilon_{acc} \times \left(\epsilon^f_{att,dir}(r) + \epsilon^f_{att,indir}(r) \times \epsilon^f_m\right) \times \epsilon^f_c \tag{3.14}$$

In the following sections, the methods used to measure the attenuation length and the fibre to SiPM optical coupling are described.

3.6.1. Sequential method

This method has been developed by people from IJClab and Antwerp University. This is the first method used for the determination of the fibre attenuation lengths and the fibre to SiPM couplings. I was in charge of its validation.

3.6.1.1. Attenuation length measurement

In the equation 3.14 the attenuation efficiency appears to only depend on the cube position in the row or column. The event selection is then requiring to have only one reconstructed cube in the plane, to prevent potential issue in the reconstruction via ambiguities or pile-ups, and constraint the deposited energy via the following condition:

$$LY_{vis}(X,Y) \times 0.99 < \sum_{f=0}^{3} A^f(X,Y) < LY_{vis}(X,Y) \times 1.01, \tag{3.15}$$

in order to have roughly the same true deposited energy in all cube events. Then using the fraction of light seen by each fibre divided by this fraction taken in the closest cube of the MPPC, one can

factorise the intrinsic light yield of cubes as well as the coupling as in the following equation:

$$\begin{aligned} F^f(r) &= \frac{A^f(r)}{\sum_{k=0}^{3} A^k(r)} \\ &= \frac{E_{dep} \times LY_{cube}^{intstrinsic}(r) \times \epsilon_{acc} \times \epsilon_{att}^f(r) \times \epsilon_c^f}{\sum_{k=0}^{3} E_{dep} \times LY_{cube}^{intstrinsic}(r) \times \epsilon_{acc} \times \epsilon_{att}^k(r) \times \epsilon_c^k} \\ &= \frac{\epsilon_{att}^f(r) \times \epsilon_c^f}{\sum_{k=0}^{3} \epsilon_{att}^k(r) \times \epsilon_c^k} \end{aligned} \tag{3.16}$$

We can see that the fraction $F^f(r)$ only depends on the attenuation efficiency and the fibre to SiPM optical coupling of fibre. It is computed event per event: for each cube events passing the selection, this ratio is computed and the mean value of the distribution is calculated. This value is noted: $\overline{F^f(r)}$.

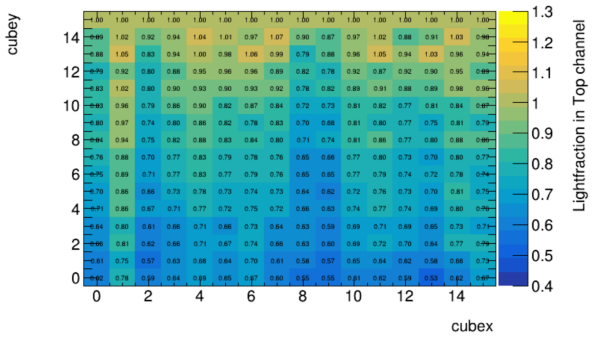
As we have seen the value $\overline{F^f(r)}$ still depends on the fibre to SiPM optical coupling of the fibre of interest: ϵ_c^f. In order to suppress this dependence, the following ratio is used:

$$\begin{aligned} \frac{\overline{F^f(r)}}{\overline{F^f(0)}} &= \frac{\epsilon_{att}^f(r) \times \epsilon_c^f}{\sum_{k=0}^{3} \epsilon_{att}^k(r) \times \epsilon_c^k} \times \frac{\sum_{k=0}^{3} \epsilon_{att}^k(0) \times \epsilon_c^k}{\epsilon_{att}^f(0) \times \epsilon_c^f} \\ &= \frac{\epsilon_{att}^f(r)}{\epsilon_{att}^f(0)} \times \frac{\sum_{k=0}^{3} \epsilon_{att}^k(0) \times \epsilon_c^k}{\sum_{k=0}^{3} \epsilon_{att}^k(r) \times \epsilon_c^k} \end{aligned} \tag{3.17}$$

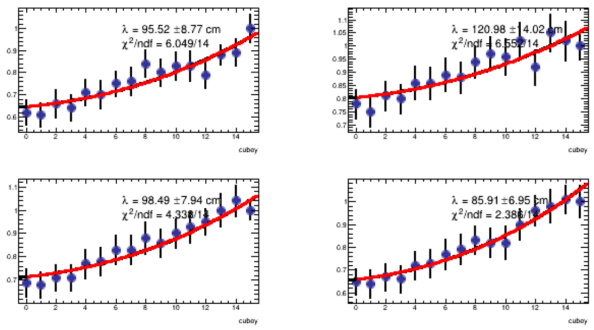
This observable still depends on the attenuation efficiency and the fibre to SiPM optical coupling of other cubes with the second term of the equation 3.17. This explains why we can see in the figure 3.18b some points that are shifted from the expected double exponential law introduced in the equation 3.13. An example is given in the figure 3.18a and examples of fit which gives the attenuation length of each channel are shown in the figure 3.18b. The performance in terms of bias and resolution on the measured attenuation length are shown in 3.6.3.

3.6.1.2. Coupling determination

For each fibre, the amplitude is computed and corrected from the attenuation effect thanks to the measurement in the previous subsection. As can be seen on figure 3.19, showing the corrected amplitude measured for fibres with the SiPM at the top of the planes, attenuation patterns are no longer visible. However, one can clearly see column to column differences. They are due to differences



(a) An example of the light fraction seen on top channels.



(b) An examples of attenuation length fit.

Figure 3.18.: Example of the light fraction measured in one plane and on the top channels and the fits to determine the attenuation length of fibres.

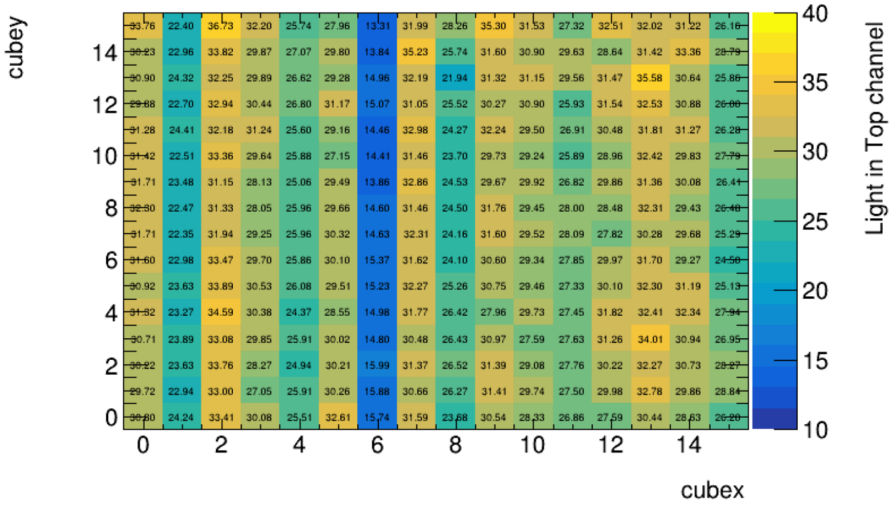


Figure 3.19.: Example of light measured from each cube by the top sensor of a plane. The sensor in X=6 is seeing less light than others, demonstrating a bad coupling between the fibre and the sensor.

in fibre to SiPM optical coupling between fibres. The same is observed in all plane and for all kind of fibres (horizontal, vertical).

From figure 3.19, we obtain 16 measurements of the amplitude per fibre of interest. So we can compute the average of this amplitude over all cubes crossed by this fibre as:

$$\overline{A^f} = \frac{1}{16} \times \sum_r A^f(r) \tag{3.18}$$

So for a given plane, 64 values are computed (one per fibre). Those averaged amplitude are proportional of the fibre to SiPM optical coupling values that we want to determine. Comparing the $\overline{A^f}$'s of the various fibres gives access to relative fibre to SiPM optical couplings. Since no absolute determination is possible, we neglect that fact and express ϵ_c^f's as in equation 3.19. Note that the difference between the values obtained this way and the truth can be absorbed when determining LY_{rec}'s, as can be seen

in equation 3.14.

$$\epsilon^{f}_{coupling} = \frac{\overline{A^f}}{\frac{1}{64}\sum_f \overline{A^f}} \tag{3.19}$$

Then this effect could also be corrected for each event. From the equation 3.14 it remains only the cube intrinsic light yield and the fibre acceptance.

The sequential method's performance are seen on figure 3.21, which shows the precision and accuracy with which it measures the attenuation lengths and fibre to SiPM optical couplings, in two kinds of simulated data (toy data and full simulation) described in the section 3.6.3. We had no clear a priori concerning the resolution of these measurements. However, the fact that these relative residuals distribution are not symmetric is a sign that the procedure is sometimes ill-behaved. This is confirmed in real data. There, we observe that the attenuation lengths of fibres readout at the bottom of plane or the left of the plane are in average 10% higher than the attenuation length of the other kinds of fibres (see figure 3.24). This bias could be due to differences between the simple fitted model (equations 3.13 and 3.14) and the actual behavior. This difference is likely in particular because of the absence of a proper optical model (see section 3.6.5).

Tracking down the precise origin of these issues was out of the scope of this thesis since this method is developped at IJCLab and Antwerp. However, two limitations in the conception of the method can be considered:

- The measurement of the attenuation length of a fibre proceeds via the fit of 16 fractions. As can be seen in equation 3.17, these fractions depend on the cube intrinsic LY, which differs between cubes.
- The denominator of the fraction depends on the characteristics of the three other fibres reading out the cube of interest. For instance, if one of them is affected by a bad coupling, the corresponding A^f is smaller and the fraction higher.

These two effects, which are not related to the attenuation lengths to measure, artificially increase this dispersion of points on figure 3.23, and therefore the uncertainty on the measured attenuation lengths. It can also cause biases, and correlations between attenuation lengths measured on some fibres and the couplings of others. Also, the sequential nature of the method (first attenuation length measurements, then measurement of the couplings) causes a correlation between the attenuation length on a fibre and the fibre to SiPM couplings.

To try to avoid the various issues described above, we have designed a method that use the information available in calibration runs differently. It is described in section 3.6.2.

3.6.2. Simultaneous global method

3.6.2.1. Introduction

In this approach we extract the attenuation lengths and the fibre to SiPM coupling values by doing a single fit using simultaneously the three asymmetries that can be defined and measured in every cube:

- **Horizontal-Vertical (HV):**

$$\mathscr{A}_{HV}(X,Y) = \frac{A^{Hori}(X,Y) - A^{Vert}(X,Y)}{A^{Hori}(X,Y) + A^{Vert}(X,Y)} \tag{3.20}$$

  with $A^{Hori}_{Fiber}(X,Y) = A^0(X,Y) + A^1(X,Y)$ and $A^{Vert}_{Fiber}(X,Y) = A^2(X,Y) + A^3(X,Y)$.

- **Left-Right (LR):**

$$\mathscr{A}_{LR}(X,Y) = \frac{A^0(X,Y) - A^1(X,Y)}{A^0(X,Y) + A^1(X,Y)} \tag{3.21}$$

- **Top-Bottom (TB):**

$$\mathscr{A}_{TB}(X,Y) = \frac{A^2(X,Y) - A^3(X,Y)}{A^2(X,Y) + A^3(X,Y)} \tag{3.22}$$

$\mathscr{A}_{HV}$, $\mathscr{A}_{LR}$ and $\mathscr{A}_{TB}$ have several advantages, the first one being that the asymmetry no longer depends on the intrinsic cube light yield :

$$\mathscr{A}_{HV}(X,Y) = \frac{\sum\limits_{f=0,1} \epsilon^f_{att}(r)\epsilon^f_{coupling} - \sum\limits_{f=2,3} \epsilon^f_{att}(r)\epsilon^f_{coupling}}{\sum\limits_{f=0,1} \epsilon^f_{att}(r)\epsilon^f_{coupling} + \sum\limits_{f=2,3} \epsilon^f_{att}(r)\epsilon^f_{coupling}} \tag{3.23}$$

$$\mathscr{A}_{LR}(X,Y) = \frac{\epsilon^0_{att}(r)\epsilon^0_{coupling} - \epsilon^1_{att}(r)\epsilon^1_{coupling}}{\epsilon^0_{att}(r)\epsilon^0_{coupling} + \epsilon^1_{att}(r)\epsilon^1_{coupling}} \tag{3.24}$$

$$\mathscr{A}_{TB}(X,Y) = \frac{\epsilon^2_{att}(r)\epsilon^2_{coupling} - \epsilon^3_{att}(r)\epsilon^3_{coupling}}{\epsilon^2_{att}(r)\epsilon^2_{coupling} + \epsilon^3_{att}(r)\epsilon^3_{coupling}} \tag{3.25}$$

Also, the attenuation lengths measured thanks to 16 such asymmetries no longer depend on the 32 transverse fibres. These asymmetries also benefit of strong features that anchor the measurement of the attenuation length and fibre to SiPM optical coupling. Indeed, assuming that the couplings and attenuations of the two vertical (horizontal) fibres are the same, the $\mathscr{A}_{LR}$ or ($\mathscr{A}_{TB}$) asymmetry

does cancel between the 8th and 9th cubes of a column (row). Also, the maximum value of these asymmetries, observed on one side of the plane, is the opposite of the minimum value observed on the opposite side. On figure 3.20, one can see that this pattern is essentially respected. The value of these extrema should determine the attenuation length and coupling to first order, while the departure from the ideal behavior provides information on the difference between their actual values for the Left and Right (Top and Bottom) fibres. The information on these differences is, however, probably dominated by $\mathscr{A}_{HV}$. Indeed as can be seen in the equation 3.23, if again the attenuation lengths and fibre couplings were the same for the 4 fibres crossing a given cube, this asymmetry would be close to 0 everywhere in the plane. $\mathscr{A}_{HV}$ also helps to break degeneracies between the optical couplings of the various fibres: they can vary without changing much $\mathscr{A}_{LR}$ and $\mathscr{A}_{TB}$, but this changes $\mathscr{A}_{HV}$, where all the couplings (not a pair) are involved.

3.6.2.2. Fitting procedure

For each calibration event reconstructed in a cube, we compute $\mathscr{A}_{HV}$, $\mathscr{A}_{LR}$ and $\mathscr{A}_{TB}$. The averages of the distribution of these asymmetries are the base of the measurement: in each plane, we fit the 256x3 theoretical asymmetries (3.23 to 3.25) to these measured average, to determine 128 parameters:

- 64 attenuation lengths.
- 64 couplings assuming that the mirror reflection efficiency is constant and equal to 80%.

The best parameters are found by minimising a global chi-square including all asymmetries as shown by the equation 3.26.

$$\chi^2_{global} = \chi^2_{HV} + \chi^2_{TB} + \chi^2_{LR} \tag{3.26}$$

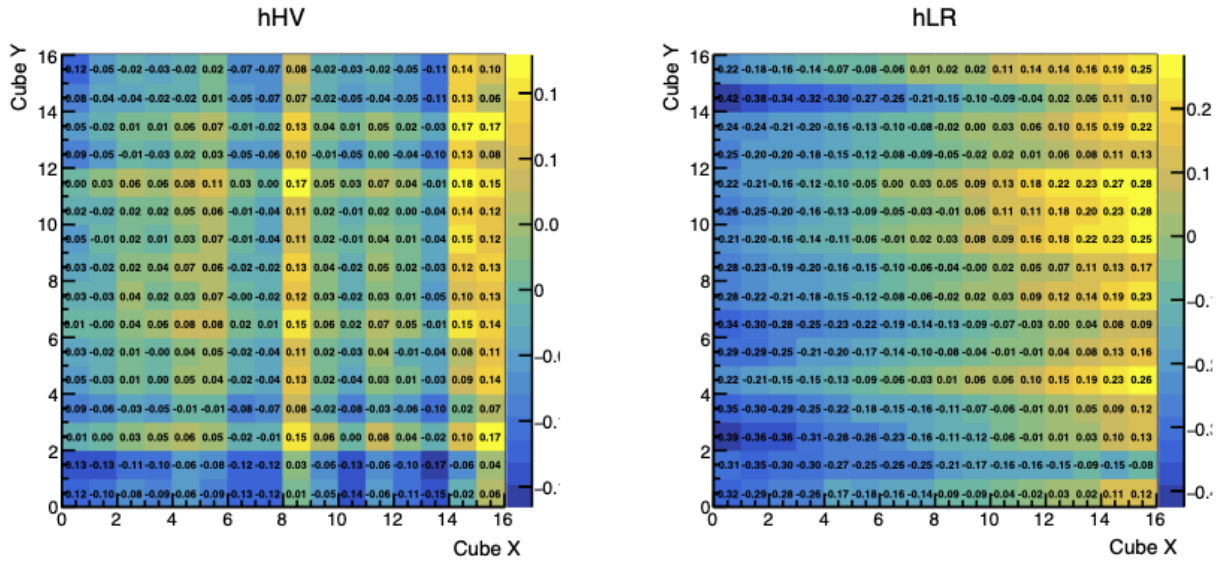
where chi-square is defined such as:

$$\chi^2_t = \sum_{i=0}^{256} \left(\frac{\mathscr{A}_t(\vec{\lambda}_{att}, \vec{\epsilon}_{coupling}, i) - \mathscr{A}_t^{mes}(i)}{\sigma_i} \right)^2 \tag{3.27}$$

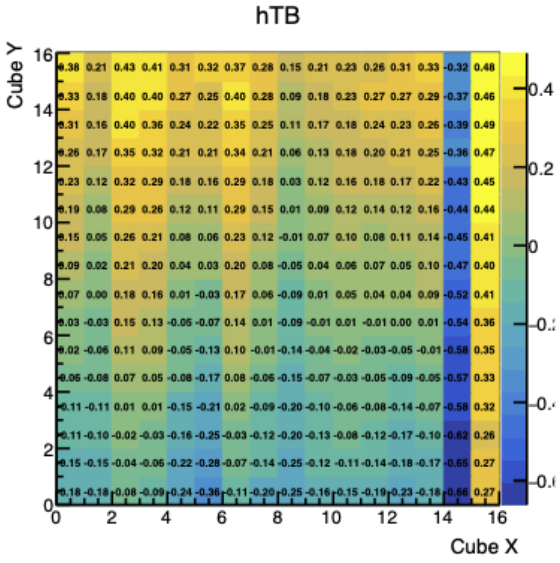
where: t is defining the three asymmetries: Horizontal-Vertical, Top-Bottom and Left-Right. The $\vec{\lambda}_{att}$ and $\vec{\epsilon}_{coupling}$ are vectors containing the 64 attenuation lengths and fibre to SiPM optical couplings to determine. And σ_i^2 is the error associated to the measurement $\mathscr{A}_t^{mes}$.

3.6.3. Toy generation and results

Given the number of parameters to determine this is not an easy fit. To make sure that enough information was available in principle for the fit to perform properly, a first study has been made using



(a) $\mathscr{A}_{HV}$ (b) $\mathscr{A}_{LR}$



(c) $\mathscr{A}_{TB}$

Figure 3.20.: An example of the three asymmetries (Horizontal-Vertical, Left-Right and Top-Bottom) coming from the simulation.

Parameters	low variation toys law	high variation toys law
Visible cube light yield [PA/MeV]	Gaussian(96,10%)	Gaussian(96,30%)
Attenuation length [cm]	Gaussian(101,10%)	Gaussian(101,30%)
Coupling	1-Exp(0.1)	Uniform(0.3,1)

Table 3.3.: Law used to randomly generate the input parameters used for the toys.

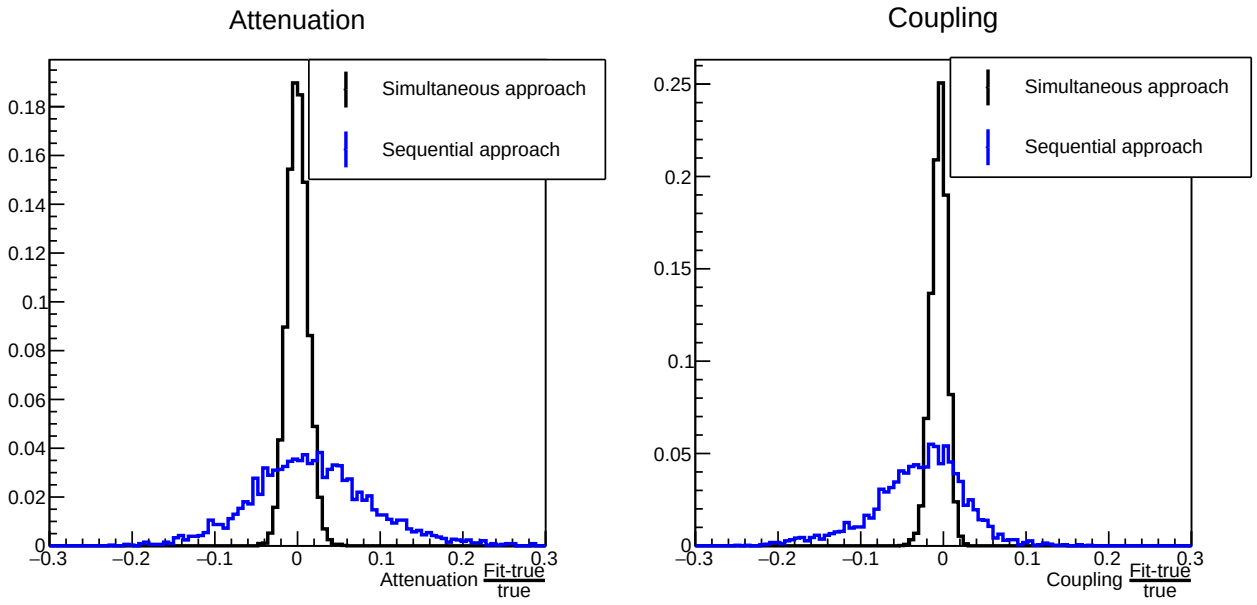
Toys	Parameters	Simultaneous approach		Sequential approach	
		Bias [%]	Resolution [%]	Bias [%]	Resolution [%]
Low variation	$\epsilon_{coupling}$	-0.5	1.0	-3.0	5.5
	λ_{att}	0.0	1.2	1.6	7.2
High variation	$\epsilon_{coupling}$	1.0	2.4	-2.8	13.1
	λ_{att}	0.0	2.1	-0.2	14

Table 3.4.: Bias and resolution on the fibre to SiPM optical couplings and attenuation lengths obtained using the Simultaneous and Sequential approaches from the two toy studies.

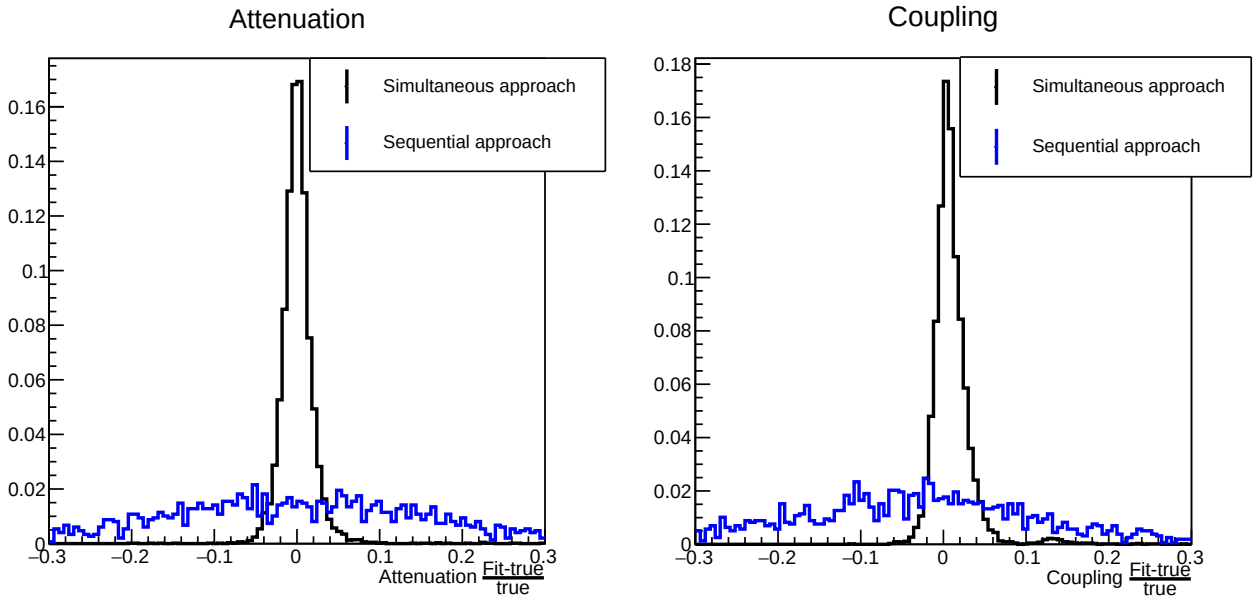
toys. For each toy, random values of attenuation lengths, fibre to SiPM optical couplings and visible cube light yields are generated. Two kinds of toys have been generated: an "easy" case in which these random values are sampled according to a gaussian centered on the typical values observed in previous studies and which σ also reflects the observed dispersion among cubes and fibres. A second study is made where this dispersion is tripled, checking the ability of the fit to determine simultaneously all parameters even in a very inhomogeneous detector. The conditions for the generation are summarised in the table 3.3.

For each configuration random values are chosen at the beginning and then 10,000 events of 1 MeV are simulated in the 256 cubes in a plane. Then, a simulation of the scintillation light between emission and detection is performed assuming a 1 MeV energy deposit. The light propagation from the cube to the MPPC is exactly the same as the full simulation of the experiment, then the number of detected PAs are recorded at the end of the four fibres crossing a cube is recorded for each event. In other words, the toy must reproduce the simulation if we are able to select isolated interaction of 1 MeV, a regime that should be reached by applying the selection previously discussed. At the end one hundred samples were generated and the two analyses (the sequential and simultaneous approaches) are performed on those data.

Results of this study are presented in the figure 3.21a and 3.21b. The blue histogram represents results of the traditional method and the black one is showing the global simultaneous fit. First this confirms the ability of the two methods to recover the true values fixed in input of toys without important bias as shown by the table 3.4 which shows the bias and the resolution of histograms on the figure 3.21. Then we see that the performance of the global simultaneous fit is better: the resolution is



(a) With low variation toys. *Left:* Comparison between fitted attenuation length and the true one. *Right:* Comparison between fitted coupling and the true one.



(b) With high variation toys. *Left:* Comparison between fitted attenuation length and the true one. *Right:* Comparison between fitted coupling and the true one.

Figure 3.21.: Results of toys studies using the sequential method in blue and the simultaneous approach in black.

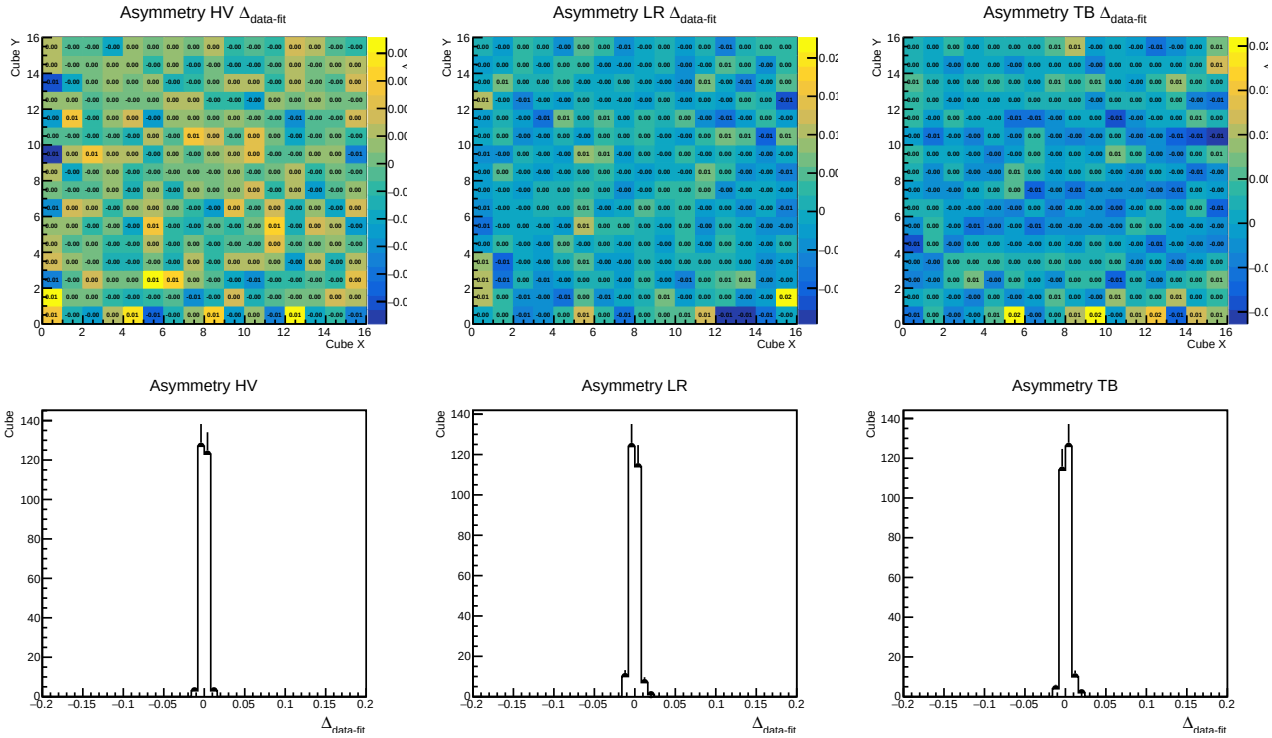


Figure 3.22.: Residual between the fit and the simulation in one plane for the three asymmetries. *Left:* Horizontal-Vertical asymmetry. *Middle:* Left-Right asymmetry. *Right:* Top-Bottom asymmetry.

lower, and the distribution of the relative residuals is more symmetric. Also, these performances are more stable against inhomogeneities in the detector.

3.6.4. Full simulation study

Once the two methods have been validated with a simple toy, we decided to perform the same study in a more realistic case. For this purpose the full simulation of the detector is used, the input values of the attenuation lengths and fibre to SiPM couplings are the one measured by the sequential method from data.

Thus a partial calibration campaign has been simulated, only one gap, using the ^{22}Na source. After the event selection and the computation of asymmetries in each cube of the plane, the fit procedure is launched. The residual between data and the fit are shown in the figure 3.22. The fit has well converged and nicely describes all histogram bins.

The results of this study are shown in the figure 3.23. The global simultaneous approach measures the attenuation lengths and couplings with an acceptable bias and twice smaller a resolution (around 4%) than the sequential approach. This resolution is worse than in the case of the TOY study but still acceptable in this realistic test. These results validate the two methods as far as they can be on simulated data. In particular, this confirms that the asymmetries used by the simultaneous method are

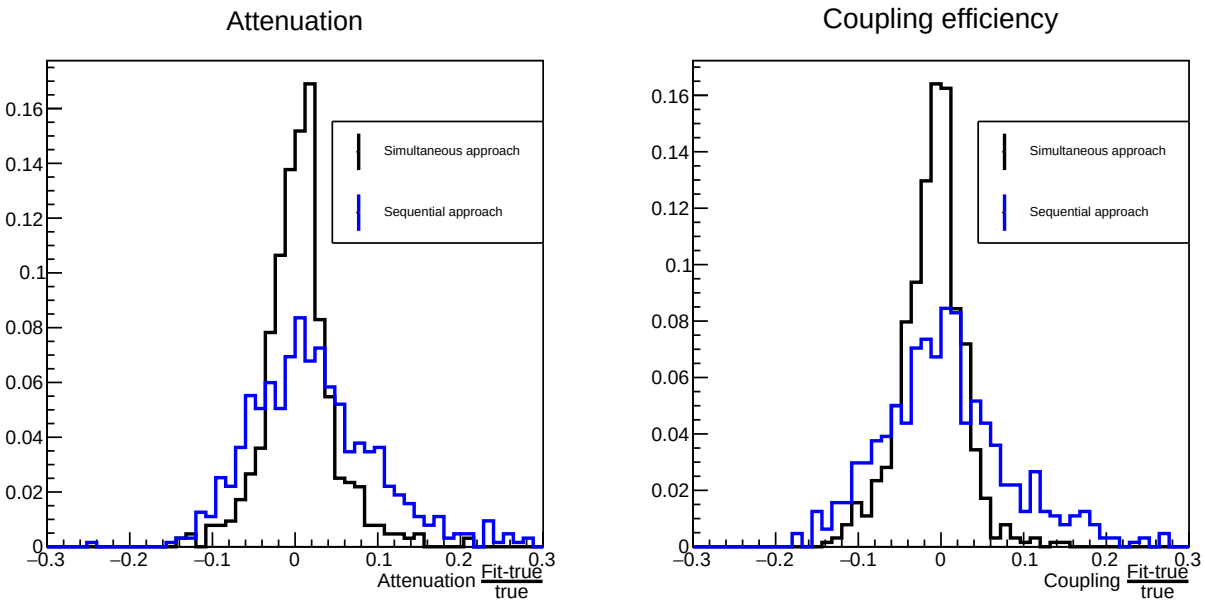


Figure 3.23.: Comparison between measured and true values of attenuation and coupling using the complete simulation of the detector. The error given by the simultaneous approach is almost divided by a factor 2.

Parameters	Simultaneous approach		Sequential approach	
	Bias [%]	Resolution [%]	Bias [%]	Resolution [%]
$\epsilon_{coupling}$	-1.1	3.7	0.5	7.6
λ_{att}	0.5	4.1	2.2	7.5

Table 3.5.: Bias and resolution on the fibre to SiPM optical couplings and attenuation lengths obtained using the Simultaneous and Sequential approaches using complete simulation of the detector.

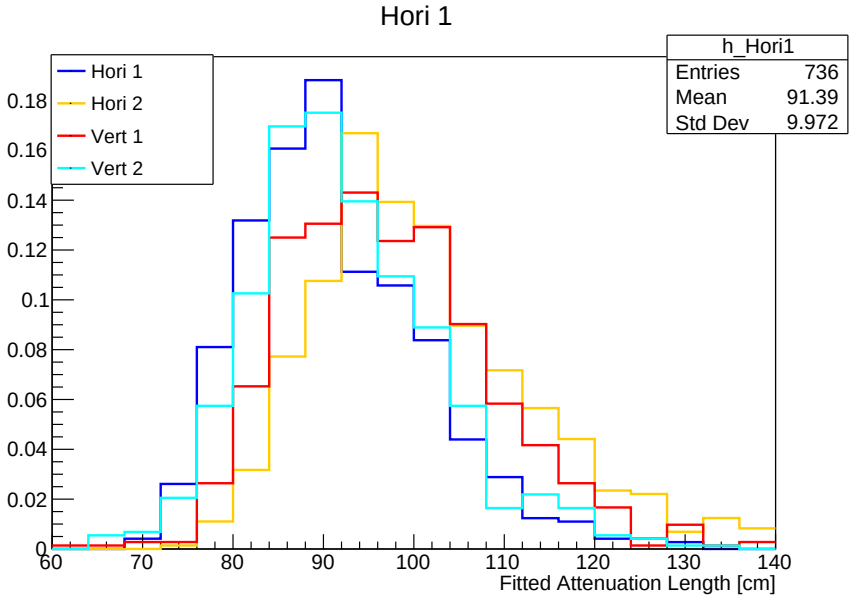


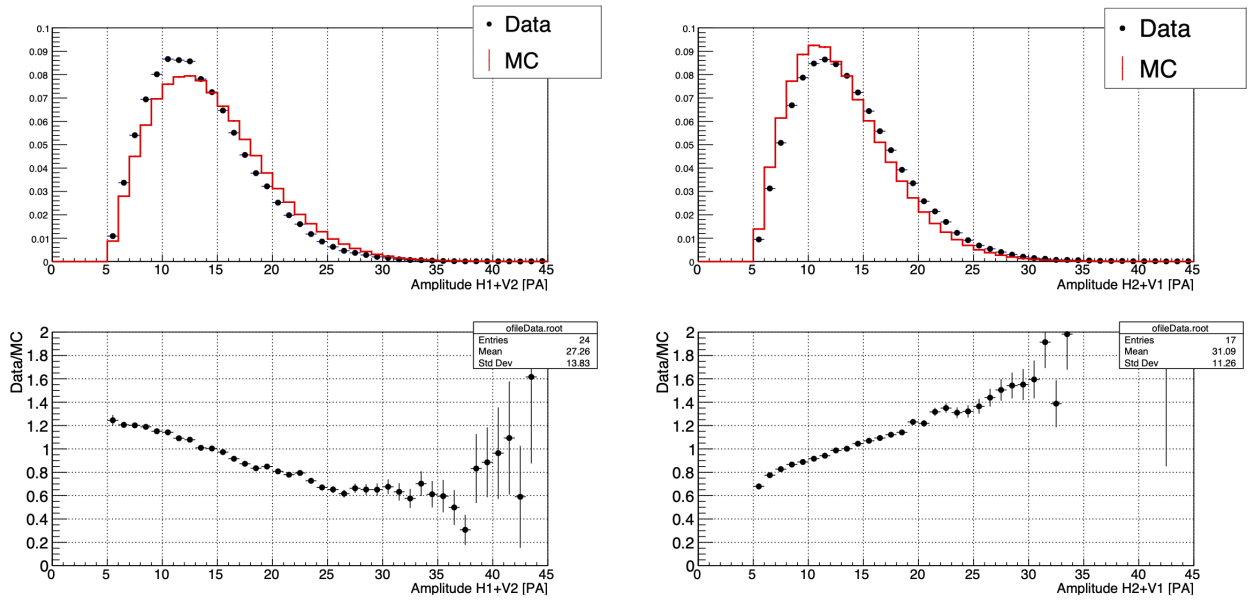
Figure 3.24.: Fitted attenuation lengths distribution with the sequential approach of fibres depending on the fibre orientation.

pertinent variables to access the needed information to measure these parameters. This all assumes that the model assumed in equation 3.14 as well as the simulation are good representations of real data. This will be discussed in section 3.6.5.

3.6.5. Application on data

Sequential approach

After the final validation with the simulation, those two approaches have been applied on real data, from the calibration campaign of September 2018. The figure 3.24 is showing the distribution of attenuation lengths depending on the fibre orientation. We can see two populations. This shift has been studied in more details, in particular using the simulation. The attenuation lengths and couplings measured here have been implemented in the readout simulation (see chapter 4) and a new simulated sample was produced. The distributions of amplitudes measured on individual fibres in events passing the calibration selections in real data have been compared with the same distributions in this new simulated sample. This comparison of fibre's amplitude has been performed in order to check that the extracted parameters allow to reproduce the data. Contrarily to the cube amplitude, the fibre amplitudes are more sensible to the attenuation lengths and fibre to SiPM optical coupling values. We have discovered that the amplitude of two fibres in the simulation are systematically shifted with regards to the data. The figure 3.25 is showing the data/MC comparison between the summed amplitude $A_{H1} + A_{V2}$ and $A_{H2} + A_{V1}$. A clear discrepancy is observable, the amplitude of fibres $A_{H1} + A_{V2}$ is overestimated in the MC whereas the amplitude of fibres $A_{H2} + A_{V1}$ is underestimated. This was a hint that the sequential approach ill-behaves.



(a) Ampltiude H1+V2. **(b)** Amplitude H2 + V1

Figure 3.25.: Comparison between data and simulation of the fibre's amplitude. The simulation is tune using the attenuation lengths and the fibre to SiPM optical coupling value derived with the sequential approach.

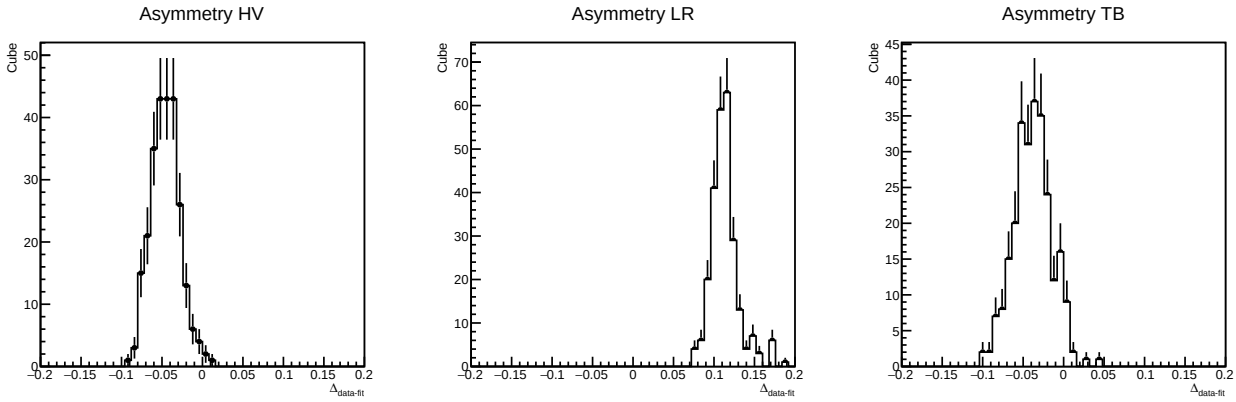


Figure 3.26.: Comparison of the asymmetries measured in each cube of one plane with the model taking the values extracted from the sequential approach.

A last test was to look at the residuals of the best fit in the three asymmetries used by the Simultaneous method. We compared the means of the data distribution of $\mathscr{A}_{HV}$, $\mathscr{A}_{LR}$ and $\mathscr{A}_{TB}$ in every cubes with the model. This is shown on the figure 3.26 for one plane of the detector. We can see that for the three asymmetries the residual distributions between data and the model are biased. The values extracted with the Sequential approach cannot reproduce the asymmetries measured in data even with the best fit. From this we confirm that the Sequential approach must be complemented by an improved method. That is the role of Simultaneous global method.

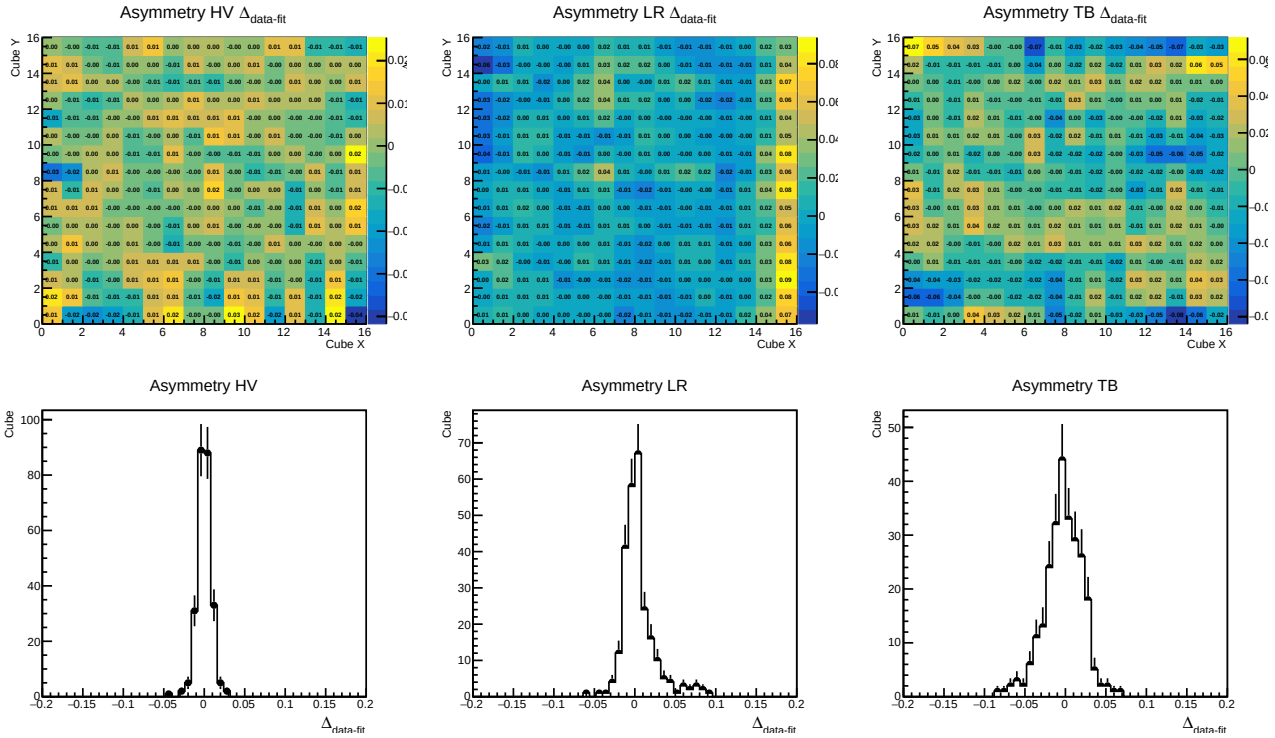


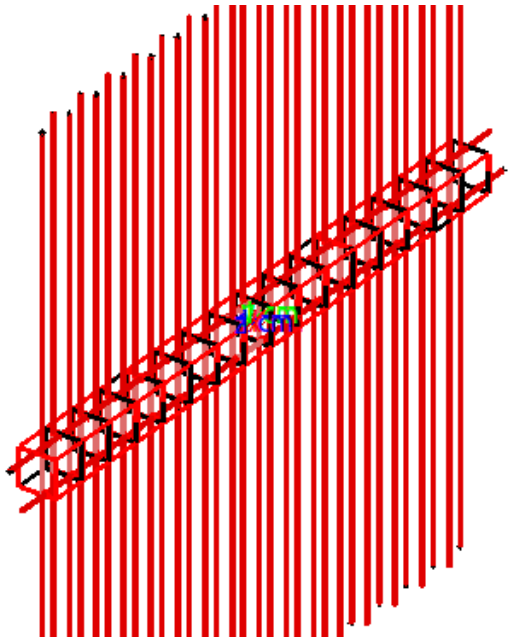
Figure 3.27.: Asymmetries residuals between the simultaneous fit and data. On the top the 2-dimensional distributions and on the bottom the 1-dimensional distributions.

Simultaneous approach

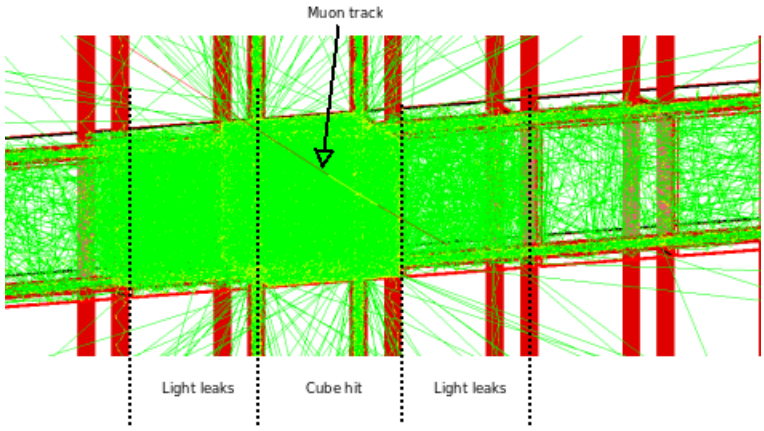
The Simultaneous approach, introduced in section 3.6.2, has been applied on real data. The residuals we obtained are shown on the figure 3.27. A clear improvement is seen with regards to the results obtained with the Sequential approach (see figure 3.26), the residuals are now distributed around zero.

Nevertheless several differences have been spotted with regards to the full simulation study performed in the section 3.6.4. The width of residual distributions is higher than in the simulation study (see figure 3.22) and we can see a tail in the $\mathbb{A}_{LR}$ residual distribution. It corresponds to cubes positioned at the edge of the plane. Also, those differences on the average asymmetries between the model and the data could explain the reason why the fit does not converge most of the time properly. In particular, it returns negative values for the attenuation lengths of fibres.

The team from the LPC-Clermont laboratory carried out a study to understand the light leaks between a cube to a neighbouring one. They have used a Geant4 simulation with one row of cubes and the fibres crossing them, the geometry is shown on the figure 3.28a. In the official simulation of the detector, the scintillation photon are not propagated. Instead the Geant4 simulation provides the amount of deposited energy in each sensitive volume by each particle. The yields of photons reaching fibres is assumed to be the same for the four fibres reading out a cube, and this is assumed in all the cubes of the detector. All effects that might occur during the propagation of the light in a cube, or related to the contact between the fibre and the PVT are neglected in this respect. Then



(a) Schematic of the geometry implemented in the custom Geant4 simulation.



(b) Example of a muon crossing a PVT cube and the scintillation photons produce by the PVT and their propagation.

Figure 3.28.: Schemas of the custom Geant4 simulation of the detector to understand the scintillation light propagation in the detector. The PVT cubes are shown in black and the fibres in red. The green tracks are the scintillation photons.

the average number of scintillation photons reaching a SiPM is determined thanks to a model of the light propagation based on the equation 3.14. In order to know if this model is correct, Clermont has simulated the propagation of scintillation photons and used the GODDeSS (Geant4 Objects for Detailed Detectors with Scintillators and SiPMs) [81] Geant4 extension to describe the geometry of the detector and material properties which is primordial to correctly propagates photons. Figure 3.28b shows the scintillation photons produced by the interaction of a muon in a cube. Knowing the issue we have observed between the fitted model to describe average asymmetries in real data, they have used this simulation to see if the double exponential model describing the attenuation effect is sufficient.

They fired 370 MeV muons crossing each cube horizontally at its centre and studied the propagation of scintillation photons and the $\mathscr{A}_{LR}$ asymmetry variable. We will focus in this manuscript on the comparison between our model (equation 3.14) and the results obtained with this simulation. The figure 3.29 shows the asymmetry between horizontal fibres as a function of the cube position. The red line is a fit using our model (equation 3.14). We can see that the model has difficulties to describe the asymmetry values for cubes positioned at the two extremities as we have seen in real data (see figure 3.27). This was a confirmation that our model which is the one implemented in the official RO simulation, could not fully reproduce the light propagation predicted by the more realistic Clermont simulation. The people in charge of the readout simulation will have to improve it by introducing a realistic treatment of the propagation of scintillation photons in the cube.

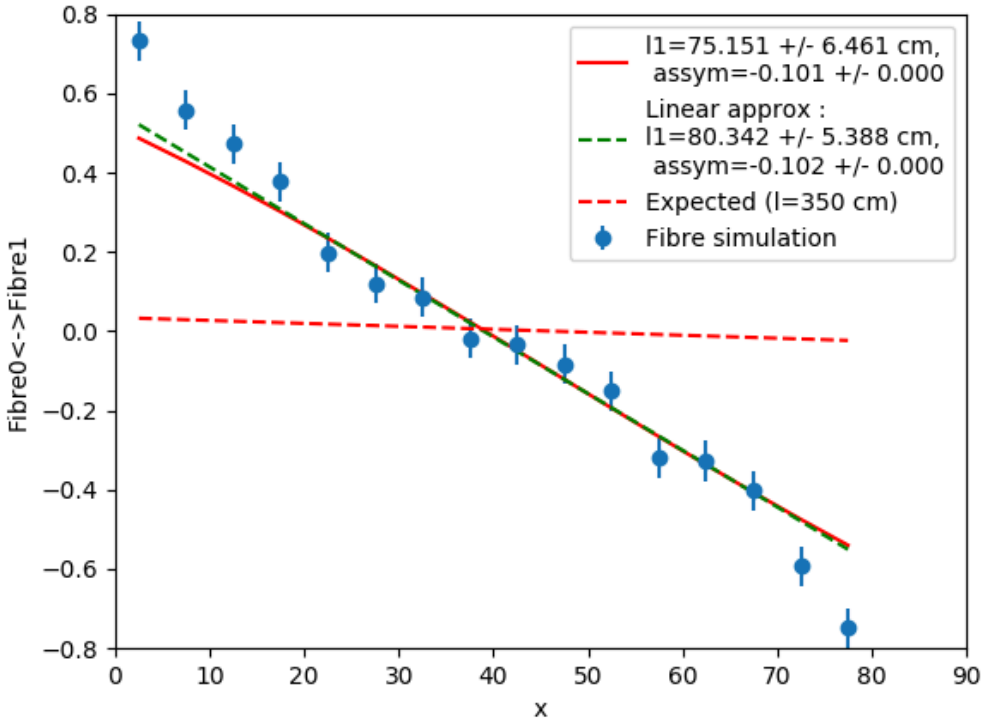


Figure 3.29.: $\mathscr{A}_{LR}$ as a function of the cube position using the LPC Clermont custom simulation. The red line is showing a fit using the equation 3.24.

Combined approach

The discrepancies between the model and measured asymmetries on at the edge of the planes, the width of asymmetry residuals distributions as well as the convergence issues, can also be due partly to a too large distance between the starting values of the fitted parameters and the best fit. The fit could be subject to local minima, for instance. To help in this, we have adopted a combined approach: the values of the attenuation lengths and couplings obtained with the Sequential approach are used as the starting point by the Simultaneous approach. The results on the figure 3.30 show a slight improvement. Moreover, in this combined approach the biases in the $A_{H1} + A_{V2}$ and $A_{H2} + A_{V1}$ distributions is no longer present as we see on figure 3.31.

This method can still be improved with the help of a proper optical model, which was out of our scope. However, as will be seen in chapter 4 the precision achieved on the attenuation lengths and fibre to SiPM couplings values, already ensure a satisfactory data/MC agreement.

The distribution of the attenuation and the coupling efficiencies for all channels are shown in the 3.32. In average the fibres have an attenuation of $\lambda_{att} = 97.33$ cm. By construction, the coupling is distributed around the unity but the dispersion gives an idea of the importance of this correction. In most fibres, the fitted coupling is found within 14% of the average.

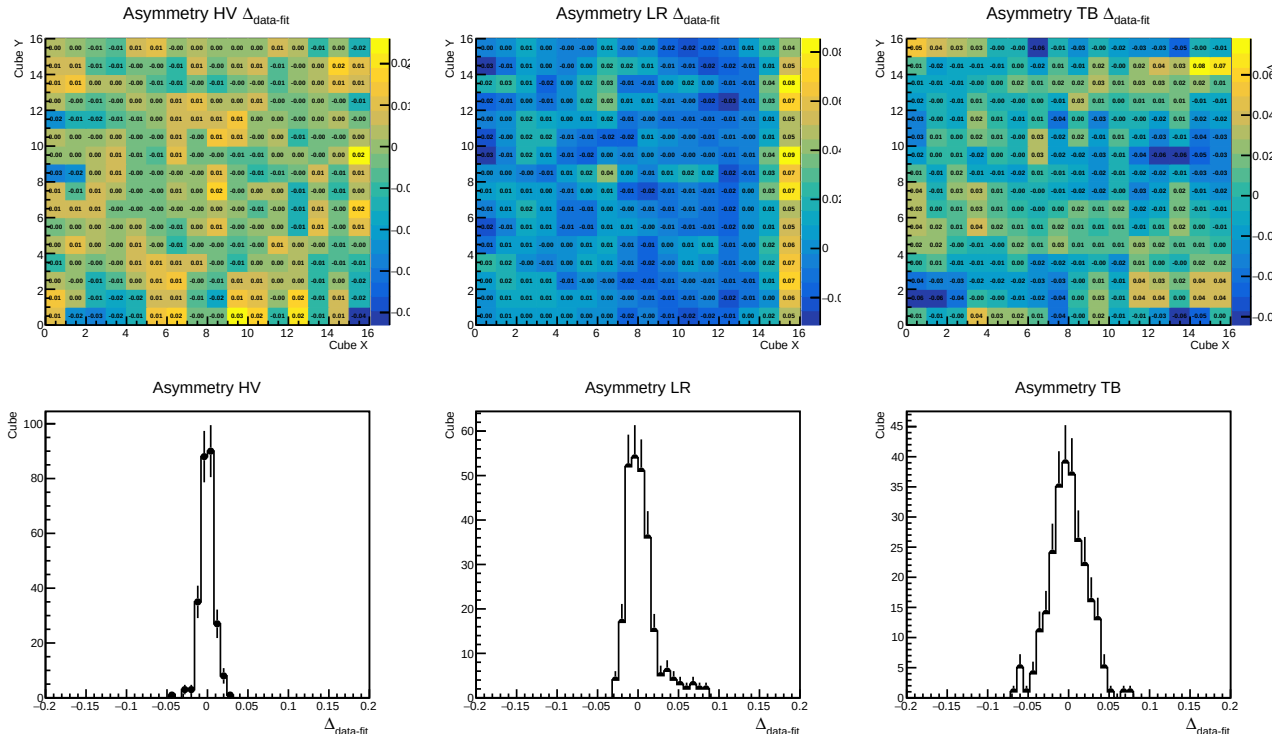
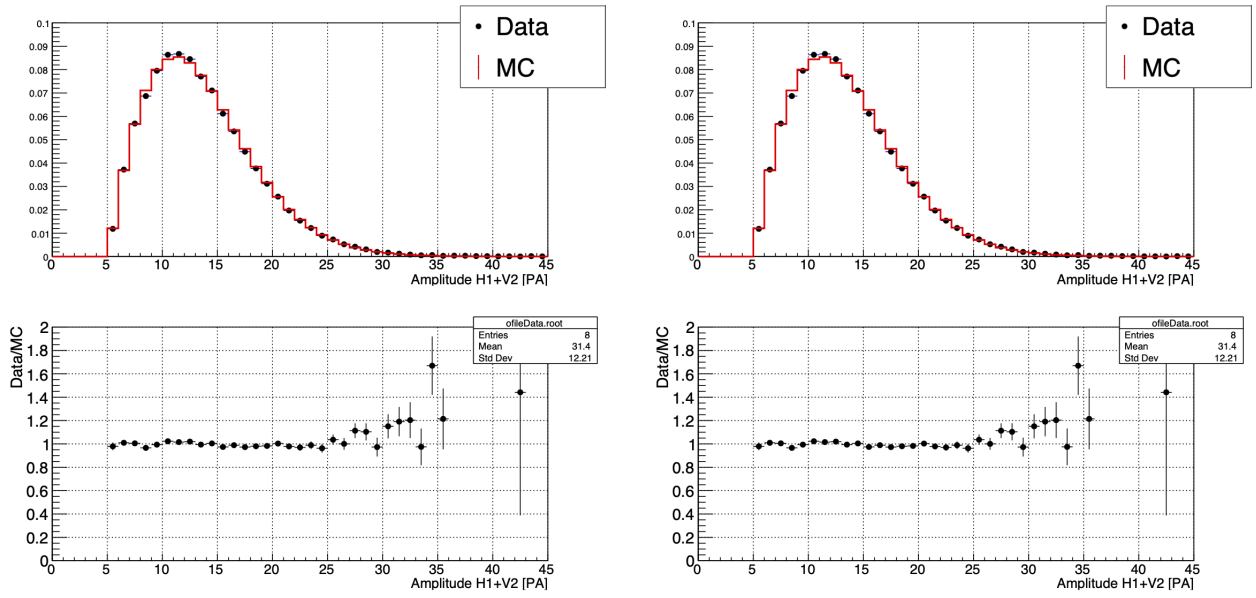
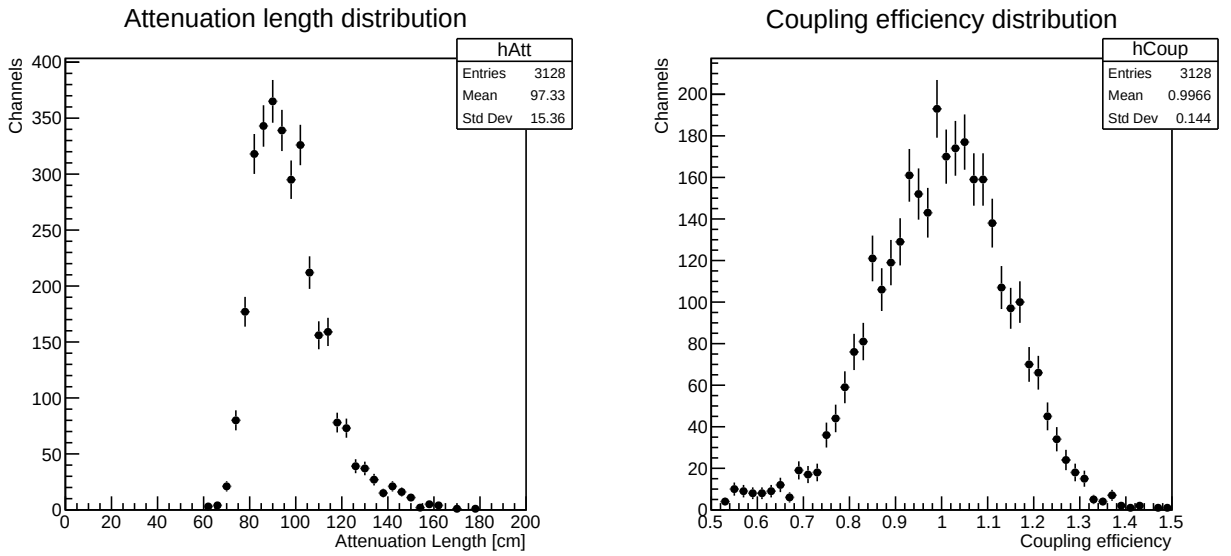


Figure 3.30.: Residuals between the fitted model and data using the simultaneous approach with in input the fitted values from the sequential approach. On the top the 2-dimensional distributions and on the bottom the 1-dimensional distributions.



(a) Ampltiude H1+V2. **(b)** Amplitude H2 + V1

Figure 3.31.: Comparison between data and simulation of the fibre's amplitude. The simulation is tuned using the attenuation lengths and the fibre to SiPM optical coupling value derived with the combined approach.



(a) Attenuation length

(b) Fibre to SiPM optical coupling

Figure 3.32.: Distribution of the attenuation length and coupling efficiency fitted in the data using the combined approach using the September 2018 calibration campaign data with the ^{22}Na source.

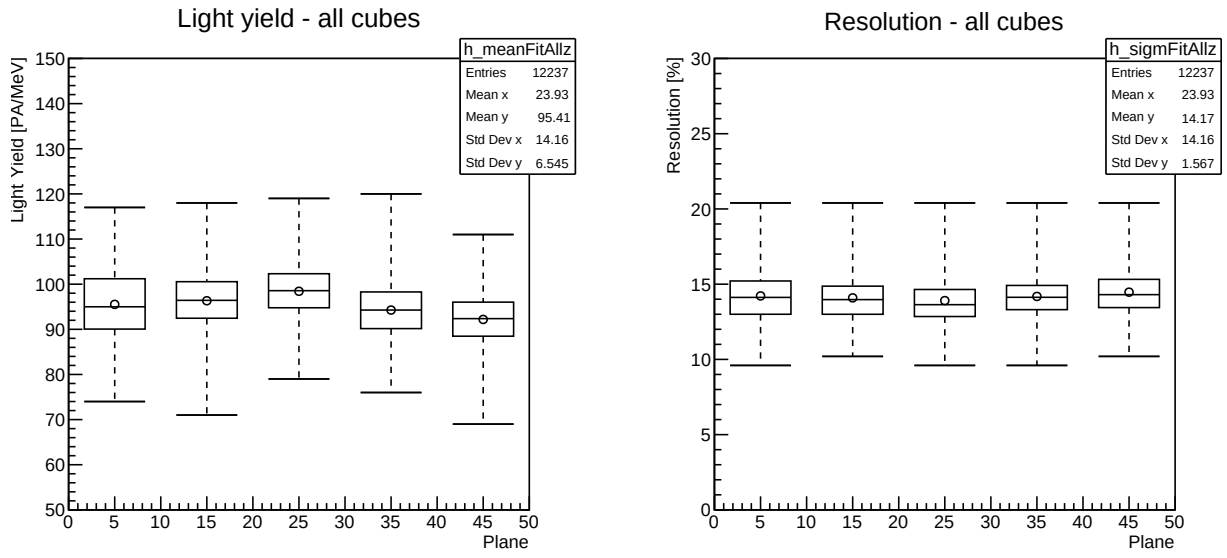
3.7. Overview of the calibration results

Sections 3.3 to 3.6 have described the calibration system and methods. They have been progressively developed since 2017 and applied to a series of calibration campaigns. We present in this section the results obtained: What are the typical values of the visible cube light yield, attenuation lengths, couplings? How homogeneous is the distribution of these values across the detector? A certain number of limitations are also discussed.

3.7.1. Calibration parameters homogeneity

We first have checked the distribution of fitted parameters (cube LY_{rec}, attenuation lengths and fibre to SiPM optical couplings) as a function of the plane position in the detector. It is the most important dependence to study as the oscillation analysis is function of the baseline. Figure 3.33 is showing the distribution of the fitted cube LY_{rec} and resolution (determined in section 3.5) as a function of the baseline. The planes are grouped in module as it will be done in the oscillation analysis. We can see that the detector response is homogeneous for the five modules.

In order to derive the visible cube light yield homogeneity, we have used the cube light yield measured in the section 3.5 and correct them for the attenuation and fibre to SiPM optical coupling effects using the measurement derived in section 3.6. Using the equation 3.14 the fitted cube amplitude distribution in the section 3.5 is:



(a) Light yield

(b) Resolution at 1 MeV

Figure 3.33.: Distribution of the fitted visible cube light yield and resolution at 1 MeV as a function of the modules using the September 2018 calibration campaign data with the ^{22}Na source.

$$\begin{aligned} A_{cube}(X,Y) &= \sum_{f=0}^{3} A^{f}(X,Y) \\ &= E_{dep} \times LY_{cube}^{intstrinsic}(X,Y) \times \left[\sum_{f=0}^{3} \left(\epsilon_{att,dir}^{f}(r) + \epsilon_{att,indir}^{f}(r) \times \epsilon_{m}^{f} \right) \times \epsilon_{coupling}^{f} \right] \end{aligned} \tag{3.28}$$

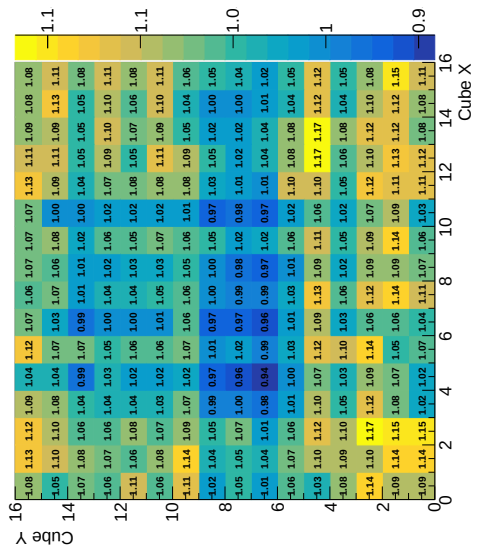
From this equation an estimator of the intrinsic cube light yield could be derived via:

$$LY_{corrected}(X,Y) = LY_{vis}(X,Y) \times \left[\sum_{f=0}^{3} \left(\epsilon_{att,dir}^{f}(r) + \epsilon_{att,indir}^{f}(r) \times \epsilon_{m}^{f} \right) \times \epsilon_{coupling}^{f} \right]^{-1} \tag{3.29}$$

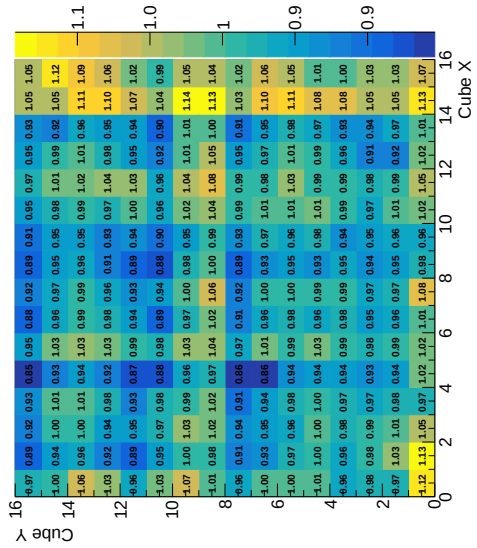
The dispersion of the $LY_{corrected}(X,Y)$ for all cubes of the detector is shown on the figure 3.35. We see a 3% dispersion which is at a satisfactory level since it contains the errors due to the visible cube light yield, attenuation lengths and fibre to SiPM optical coupling measurement as well as the natural variability of the intrinsic cube light yield. We show also on figure 3.34 the homogeneity of LY_{rec} and $LY_{corrected}$ taking into account for the fibre-to-SiPM optical coupling and attenuation length for three planes. We clearly observe the improvement of the homogeneity as a function of X and Y.

3.7.2. Calibration parameters time evolution

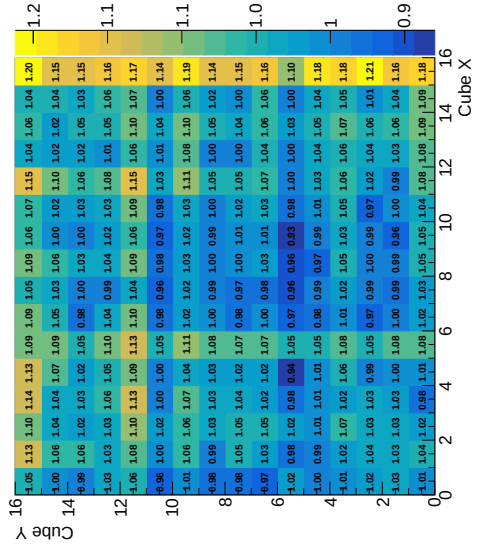
Five calibration campaign have been completed using the ^{22}Na and the threshold trigger. We can see on the table 3.6 that the visible cube light yields seem to slightly decrease with a drop of around 2% over



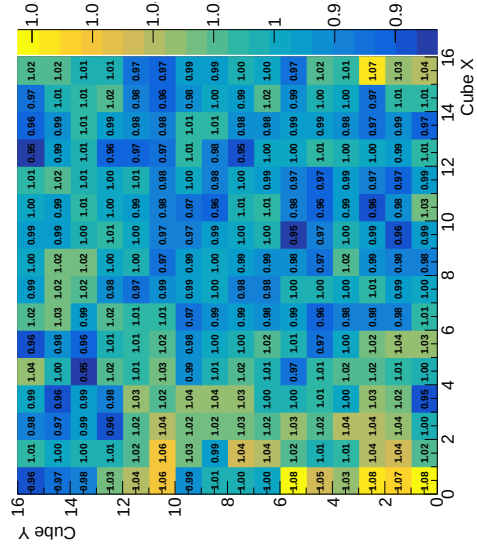
(a) Plane Z=0 LY_{rec}



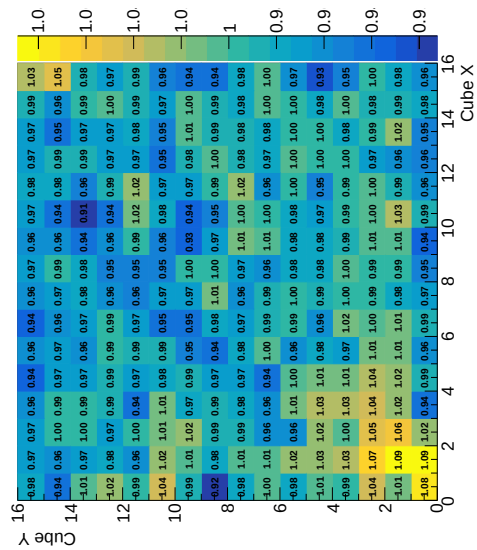
(b) Plane Z=40 LY_{rec}



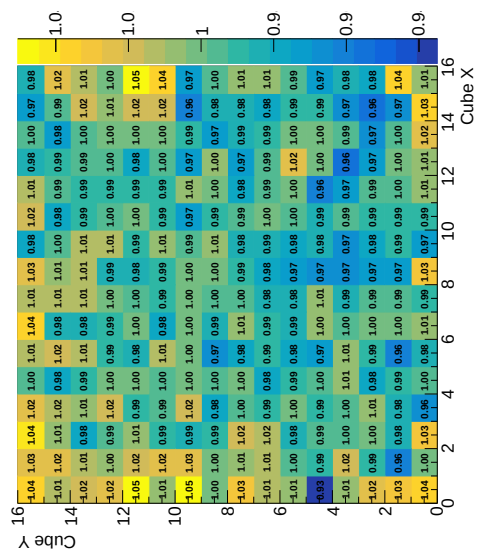
(c) Plane Z=25 LY_{rec}



(d) Plane Z=0 $LY_{corrected}$



(e) Plane Z=40 $LY_{corrected}$



(f) Plane Z=25 $LY_{corrected}$

Figure 3.34.: Distribution as a function of X and Y of LY_{rec} and $LY_{corrected}$ for three planes. For each plane the cube X=8 and Y=8 is taken as reference to highlight the homogeneity. We can see that the homogeneity is improved by taking into account those corrections.

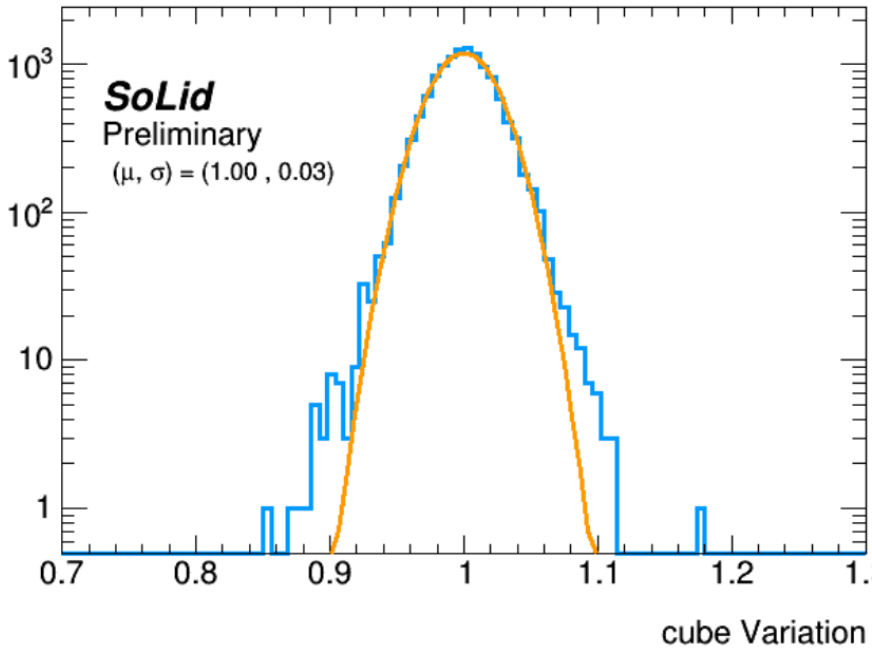


Figure 3.35.: Visible cube light yield corrected from the attenuation and fibre to SiPM coupling effects dispersion. All cubes are distributed with a standard deviation of 3% using the September 2018 calibration campaign data with the ^{22}Na source.

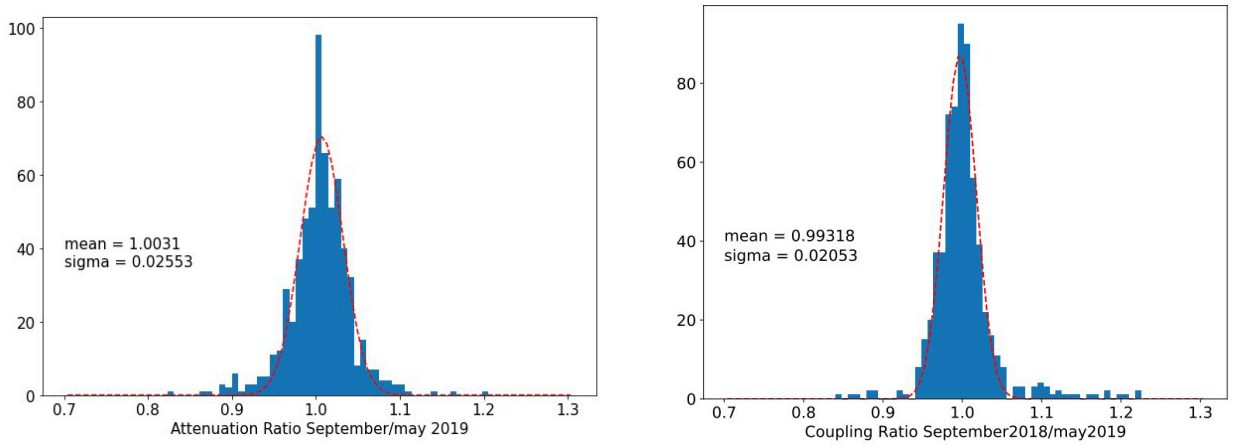
Date	$\overline{\mathrm{LY}}/\overline{\mathrm{LY}_{september}}$
September 2018	1.0
December 2018	1.003
January 2019	0.993
May 2019	0.977

Table 3.6.: Time evolution of the average visible light yield of all cubes of the experiment with regards to September 2018 calibration campaign results.

a period of two years. Plastic scintillators are known to be sensitive to environment conditions [82] which impact their light yield emission properties. In particular PVT-based scintillator are more sensitive to performance degradation over time than Polystyrene-based scintillator [83].

The two others parameters, attenuation length and fibre to SiPM optical coupling, have been also studied as a function of time. This is shown on figure 3.36. We can see no clear deviation between September 2018 and May 2019 measurement. Nevertheless, a tail is seen on the fibre to SiPM optical coupling which is under investigation. A possible explanation would be that the contact between vertical fibres and SiPM deteriorates due to gravity effect.

These ageing effects have to be taken into account while performing the sterile analysis with all the data taken with the detector. When reconstructing an event, the values of a given calibration parameter can be chosen based on the date when this event was taken, at an intermediate value between the



(a) Attenuation length

(b) Fibre to SiPM optical coupling

Figure 3.36.: Comparison between the attenuation lengths and fibre to SiPM optical coupling determined in May 2019 and September 2018.

values measured during the closest calibration campaigns. MC samples will also have to be generated accordingly.

3.7.3. Linearity test

One of the argument in favour of the PVT choice as target is its linear energy response contrarily to liquid scintillator based detector. This linear behaviour between the amount of deposited energy and the produced scintillation light has to be confirmed over the whole IBD energy range, from few hundred keV to 9 MeV. The team from IJClab have performed this study using the calibration runs taken with other sources in September 2018. We describe the result here for information, and because we contributed to the validation of the Kolmogorov approach used here.

The ^{207}Bi emits three gammas at different energies, see the table 3.1. The AmBe source is the most energetic one with a gamma of 4.4 MeV. Above this energy, the muons crossing the detector are used, as described in the section 2.4.2.1. They have used the Kolmogorov approach, described in section 3.5.2, in order to derive the visible light yield from the Compton spectra given by those gammas. Figure 3.37 shows the results of this study. The measurements are in agreement with the expected linear behaviour of the PVT from 600 keV to 4 MeV. It shows that the energy scale in SoLid is consistent not only with a linear behaviour but with a proportional one. Lower energetic sources, such as the ^{137}Cs and annihilation gammas from ^{22}Na, require to correctly model the efficiency to reconstruct such energy deposits under the assumption of 4 fibres above threshold per cube. A proposal is given in the annex A to model this efficiency analytically in order to avoid the running of a full simulation for each tested visible light yield.

Another group (at Gent University) is in charge of completing the inputs to the linearity study by providing LYs measured at high energy, using muons. This is not finalized yet.

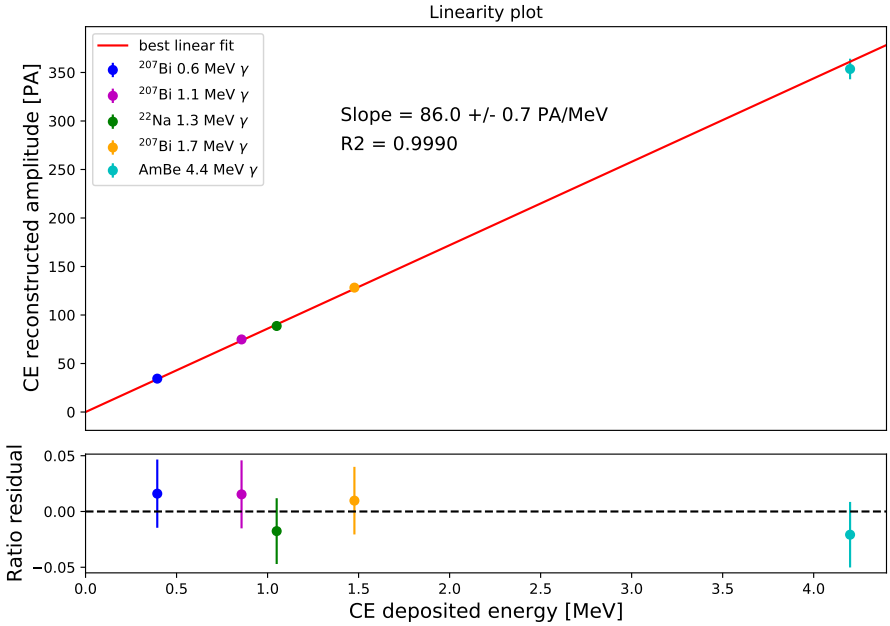


Figure 3.37.: Extracted light yield as a function of the Compton edge energy for several calibration sources. The points are compatible with a linear law as expected. This demonstrates the linear response of the PVT.

3.8. Conclusion

The energy calibration of the detector is a crucial work in order to perform the physic analyses foreseen by the SoLid collaboration. SoLid's detector and DAQ designs make this task a very challenging one. The finite size of the cube, the important number of channels, the fact that different cubes are readout by common fibres, and the limitations of the DAQ are the main issues in this respect.

Firstly, a new method to extract the visible cube light yield has been developed. The usage of this method and of a previous one developped at IJCLab allows to control the associated systematic uncertainties. The major part of the work was performed using the main source of the collaboration, the ^{22}Na, which Compton edge is close to 1 MeV. The two methods have been validated using toys and the detector simulation. The differences in their results have been shown to be similar in MC and real data. It is at the 2% level. A conservative 3% uncertainty on the determination of LYs is confirmed by the dispersion observed after correction of the fibres attenuation lengths and optical couplings to SiPMs. It is consistent with the required precision on the energy scale for a sterile analysis.

The second part of the work has been dedicated to the fibre parameter measurement: the attenuation lengths and the coupling. A review of the traditional method has shown issues in their extraction. Then a new approach has been proposed to perform the extraction and simulation and data/MC

studies have demonstrated a better performance. Using it, the attenuation lengths and the couplings are known up to 4%.

Finally, a method was proposed to determine LYs when energy deposits do not exceed a few hundreds keV. It takes into account from first principles the Poissonian behavior of the number collected PAs and the selection efficiency. This method could be useful to constrain the linearity of the PVT response down to low energies.

Thanks to this work, the calibration constants have been implemented in the reconstruction software and in the simulation. The comparison between data and simulation after tuning of the energy response is the subject of the next chapter.

Chapter 4.

Simulation of the detector: Description and tuning of the energy response

Rodrigo y Gabriela. *11:11*, Rubyworks, 2009

4.1. Introduction

4.1.1. Role of the simulation in SoLid analysis

Simulation is a fundamental tool in particle physics. The ability of the simulation to reproduce data is a strong evidence of understanding and control of the detector response. A reliable simulation is an important tool to design a physics analysis. Among other things, the simulation is useful to :

- Tune the selection of signal events by comparing the distributions of discriminative variables in simulated signal and backgrounds samples.
- Provide distributions to be fitted to the data to determine the signal yield, to measure some physics parameters or establih confidence regions, etc.
- Subtract detector effects from these distributions via reliable ununfolding procedures, trained with simulated dataset.
- Perform systematics uncertainties studies.

As already explained in section 3.2.1, the search of a sterile oscillation with SoLid is performed by comparing the measured $(E_\nu;L_\nu)$ spectrum with spectra predicted by the full simulation, which play the role of the p.d.f. in a fit to real data. The oscillation parameters θ_{ee} and Δm^2_{14} are the floating parameters. Therefore, a strong hypothesis is made on the ability of the simulation to reproduce data. It is required to prove it via comparisons between data and simulation in control samples. It usually starts with the simplest dataset available, such as calibration runs. Then comparisons are performed with more complex events.

The ability to resolve an oscillation is driven, among other things, by the resolution with which the baseline and the neutrino energy can be reconstructed. Those parameters mainly depend on the detector design and technology. This is not the subject of this thesis. Our role is to minimize the additional uncertainties, often stemming from biases due to discrepancies between real data and MC. This is obtained by tuning many parameters in the simulation, based on data measurements. We mainly focussed on the parameters studied in chapter 3. Indeed, the readout simulation has to properly describe the losses of scintillation photons due to the attenuation along the fibre and to the quality of the fibre to SiPM optical couplings: therefore, the attenuation lengths and couplings measured for each fibre have to be implemented in the simulation. Also each true energy deposit in a cube has to trigger in the simulation the creation of a well defined yield of scintillation photons reaching each fibre, derived from the LY_{rec} measured in each cube. Then the quality of the simulated energy response is studied at several levels. First, the comparison of individual cubes and fibres signals compared with those in real data. Once these elementary building blocks are validated, one can study the quality with which *clusters* of cubes are reconstructed. This is necessary in the case of IBD since the energy of the positron and annihilation gammas is often deposited in several cubes. To that end, we studied BiPo events, stemming from the environmental radioactivity, which involve radiative beta decays resembling IBD. This is due to the presence of gammas hitting several cubes in the same event. Then, we studied a part of the systematic uncertainties related to the data/MC discrepancies that remain in ES variables. This is also described in this chapter, after an overview of the main systematic uncertainties that should affect SoLid oscillation search.

4.1.2. Overview of the simulation work in this thesis

At the beginning of this thesis work, it was not possible with the Geant4 simulation of the detector to perform a realistic simulation of detector calibration. Indeed, the calibration robot responsible of calibration sources transportation within the detector was not simulated yet, nor were the calibration sources. My first contribution was to implement the geometry and material of these elements, as well as source generators. This first step was crucial to validate the calibration procedure and this work is presented in the section 4.2.

The first calibration campaign with the ^{22}Na source happened in March 2018 (see table 3.2). The methods to measure the cube visible light yield (described in section 3.5) have been applied to those data. This allowed to perform the first comparison between data and simulation using a control sample. The goal was not to tune all the detector parameters but to determine how well the first version of the simulation - before any extensive tuning- already described real data, and whether tuning only cube scintillation light yields, thus one single parameter per cube, could suffice for the first SoLid's analyses. It was somehow useful to determine the amount of work and allow SoLid to obtain preliminary results. It was performed on only one central cube of a plane in front of the source. In this preliminary work, we attributed to all fibres the same coupling and attenuation length values in the simulation. This simplified approach allowed to track and correct a lot of bugs in the Readout simulation code. This first step on the comparison between data and simulation is presented in section

4.3.1 with the emphasis on the understanding of an important source of discrepancy between data and simulation.

At the end of this preliminary work, the operation of the experiment was improved. The DAQ settings changed to increase the neutron detection efficiency. The SiPMs' over-voltages were increased to increase gains. New calibration data were taken to pursue data/MC studies in these new conditions. Also the reconstruction algorithm changed: the definition of an event, the waveform treatment, the extraction of the amplitude and the computation of the cube amplitude changed. Concerning the calibration work, the attenuation lengths and the couplings of fibres were measured and implemented in the simulation for the first time. This is presented in section 4.3.2.

One of the main backgrounds in SoLid is the BiPo background (see section 2.3.3). It is present in every cube of the detector at high rate. The BiPo event topology is close to that of the signal due to the emission of gammas in addition to the β particle. This causes multi-cube events. This is a good candidate to explore the agreement between data and simulation of variables used in the signal selection. Two ways have been developed to select and extract pure samples of BiPo events and perform the comparison. This is shown in section 4.3.3.

The systematics uncertainties due to the remaining imperfections of the parameter tunings described in this chapter are discussed in the last section 4.4.

4.2. Description of the simulation

4.2.1. Geant4 simulation

The Geant4 framework is a c++ toolkit [84] to simulate the passage of particles through matter. It consists of randomly choose the fate of the primaries and secondaries particles according to the cross-sections derived from measurement or empirical models. Particles are tracked step by step. At each step, the kind of interaction undergone by the particle is chosen this way, is the quantity of energy or momentum exchanged with the medium. In some cases new particles are created, which will be subsequently tracked. When writing a G4-based simulation, a first important stage is to define in detail the geometry of the detector's elements and the materials they are made of.

4.2.1.1. Geometry and material of the detector

The detection volumes are reproduced according to the technical drawing made during the development of the detector (ex: PVT cubes, two ^{6}LiF:ZnS sheets per cubes etc...), as detailed in section 2.2. Also the fibres and MPPC's geometry as well as the Aluminium frames surrounding each detection plane are described. This implementation of the detector geometry also allows to compute the number of Hydrogen atoms available to interact with anti-neutrinos coming from the reactor. Then the envi-

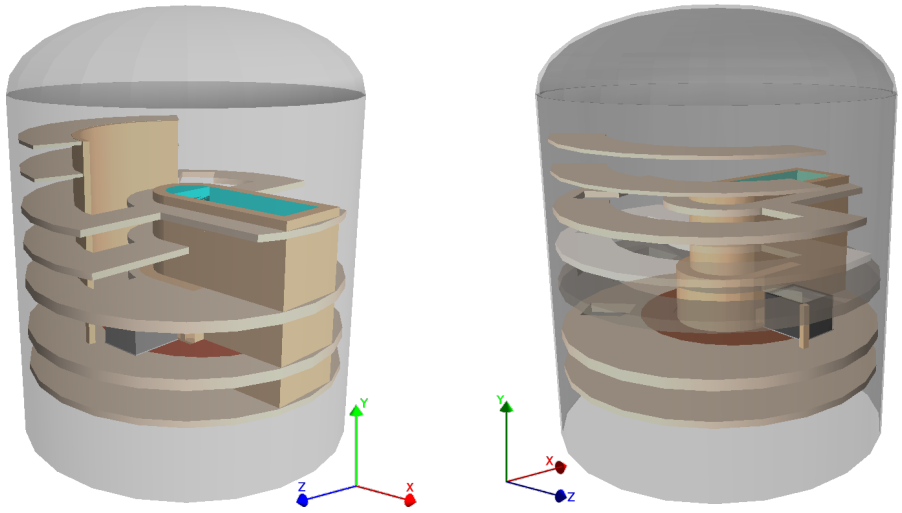
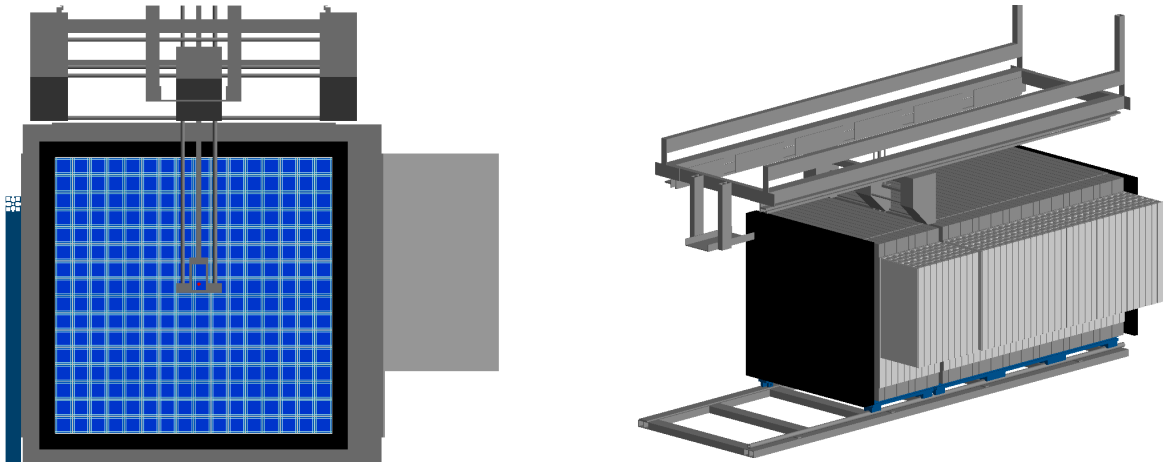


Figure 4.1.: View of the simulated environment around the SoLid detector. The BR2 building made of concrete and the reactor pool are well described in order to correctly reproduce cosmic induced backgrounds.



(a) Front view of CROSS.

(b) CROSS in position in a calibration gap.

Figure 4.2.: Views of the CROSS system in the Geant4 simulation.

ronment of the detector, which is mainly the passive shielding and the container are described, as well as the reactor pool and the BR2 building, as can be seen on figure 4.1.

The first contribution of this thesis to the simulation is the implementation of the calibration robot. This is shown on figure 4.2. In addition to the careful definition of the material and geometry of the various elements constituting this robot. We also contributed to the implementation of calibration source generators.

4.2.1.2. Event generator

Several simulations are produced to help design the oscillation analysis. The production of the signal and background simulated samples is centralised since every one from the collaboration could need

to use it. On the other hand, the simulation of calibration runs is left to people who work on the calibration or on the agreement between data and MC. Events generators have been developed to meet the SoLid's specific need.

Signal generation

When the production of a simulated IBD sample is launched, an input file produced by a software called *SoLo* is read (described in 2.1.2), it contains the positions of IBD interactions in the detector and the incoming neutrino energy and momentum, as well as those of the positron and neutron produced by the IBD. Specific samples are generated for each reactor cycle, with predictions adapted each time, based on the core composition and the power measured at BR2 during this cycle. This is described in 2.1.2. Then *SoLo* reads the MCNP output and returns a list of IBD events which length corresponds to the duration of the run, as well as the spatial and momentum values of the positron and neutron produced for each IBD in the detector. These outcomes are ruled by the number of protons available in the detector, its geometric acceptance and the IBD differential cross-section.

Calibration sources

Two kind of sources have to be distinguished:

- **Neutron sources:** Two neutron sources used to determine the neutron detection efficiency, the 252-Cf and the AmBe. The neutron energy spectrum is taken from the ISO norm.
- **Gamma sources:** Gamma sources are listed in table 3.1. Some of them emit several gammas in coincidence, thus the Geant4 class *G4IonTable* is used to select the isotope of interest in order to fully describe the kinematic of the decays. In those sources, the radioactive material is contained in a polyethylene capsule. We simulated it as well. Indeed, although in many cases (neutron, gamma) it is too thin to significantly affect the calibration particles, it is thin enough for the low energy positron emitted by the ^{22}Na decay to annihilate there, producing annihilation gammas.

Backgrounds

In the oscillation analysis, the correlated background measured during reactor-off period is subtracted from reactor-on data to evaluate the signal. A more detailed description of the background in the detector is given in section 2.3. Thus the understanding of this background is important to control the subtraction. Three kinds of such background samples were produced:

- **Muon simulation:** A part of the correlated background is from muons produced in the atmosphere by cosmic showers. The muon rate depends on the experiment altitude and on the atmospheric pressure. A specific event generator is used to obtain the rate and the angular and energy distributions of muons coming from the atmosphere, called CRY [85]. Those simulations

mainly cover the IBD-like events produced by muon spallation and they are used to derive the muon veto efficiency. Processes that contribute to the background are described in section 2.3.2.

- **Neutron simulation:** Besides atmospheric muons, cosmic showers also produce atmospheric neutrons. Their flux and energy spectrum are simulated using the Gordon model [86]. They are fast neutrons that could reach the detector and imitate the IBD signals, in the same way as the fast neutrons produced by muon spallation, described above.
- **BiPo:** The ^{6}LiF :ZnS(Ag) sheets produced by SCINTACOR company have a severe and unexpected pollution from isotopes of the ^{238}U decay chain. This BiPo background is simulated by firing a ^{214}Bi isotope with the *G4IonTable*, taking into account all possible radiative decays with the corresponding probability. It is randomly positioned in the volume corresponding to the ^{6}LiF :ZnS(Ag) sheets.

4.2.1.3. Information in output of the G4 simulation

During the Geant4 process, the primary particles are transported step by step through the defined volumes (containment building, container, detector, etc..). At each step of the particle, it can interact or not. If an interaction happens in a sensitive volume, i.e. PVT or ^{6}LiF:ZnS sheets, a hit object is created and recorded. It contains the following information about the ionising particles which may create scintillation photons by the scintillator's molecule excitation:

- The amount of deposited energy in MeV transferred to the ionising particle.
- The distance travelled by the ionising particle, called track length.
- The time and position of the hit.

From the point of view of the SoLid's read-out simulation, what matters is what can create scintillation light. Therefore, a ROOT TTree object is created, which contain for each event and each sensitive volume the list of charged particles that deposited energy there. The information on the many steps a particle has undergone in this volume is not fully kept: only the summed deposited energy over all steps (plus the nature of this particle, the position and identifier of the volume, also an index to retrieve the particle in other trees containing more details about creation and destruction processes). This is the output of the Geant4 stage. The corresponding *output file* is generated by the *SoLidsim* software.

4.2.2. Read-out simulation

The read-out simulation is responsible for simulating the detector response and takes as input the file produced by the *SolidSim* software. First, a description of simulated readout chain is given from the reading of this file to the beginning of the reconstruction, where the simulation no longer differs real data. Then a list of the thesis contributions to the full detector simulation is given.

4.2.2.1. Simulated readout chain

The detector response is simulated thanks to a sequential application of a list of classes applied event by event. The simulated readout chain is:

1. **SSimReader:** This part ensures the reading of the Geant4 output produced by the *SoLidSim* software. It takes all energy deposits that have happened in a detector's sensitive volume.

2. **SCubeFibreSim:** In each volume, the total energy deposit (summed over all the particles that interacted in this volume) would be converted into a scintillation yield if a full optical model was used. In the SoLid readout simulation, it is directly converted into a number of photons reaching the fibres, assuming the same number on all the four fibres. The conversion factor is adapted from the measurement of LY_{rec} from calibration data with the methods detailed in section 3.5. Part of the scintillation light produced in a cube may escape and be detected in cubes sharing a face. Thus, a number of photons reaching fibres of these neighbouring cubes is also computed. If real energy deposits also occurred in these cube, they will be added to the numbers due to these deposits.

   The average number of photons reaching the SiPM is computed by applying the attenuation factor due to the fibre attenuation length and to the fibre to SiPM coupling (equation 3.14 in section 3.6). Afterwards this number reaching the SiPM is randomised with a Poisson law and the time of arrival is computed assuming the speed of light in the fibres.

3. **SSensorSim:** This algorithm is responsible for the simulation of the sensors response. For each photon reaching a sensor there is a probability to induce a Photo-Avalanche (PA), this is mostly driven by the Probability Detection Efficiency of the MPPC. There are also effects that trigger other PAs: Dark Counts (DC) and Cross-Talks (CT), as described in section 2.2.3.2. The models describing each effects and giving their rates depend on the voltages applied on the MPPC, which fluctuates when a PA is induced, and on the ability of the MPPC to recover the nominal value. This effect is taken into account in the simulation. In the end a list of PAs per channels is passed on to the next stage.

4. **SElectronicsSim:** For each PA this algorithm creates the corresponding waveform. The waveform model is from an empirical fit of data waveforms. An example of fitted waveform is shown on figure 4.3. Then the baseline fluctuation is added and the list of waveforms is stored.

5. **STriggerSim:** At this stage the simulation is in the same format as the data: a continuous flux of waveforms sampled every 25 ns. The trigger algorithms are described in section 2.2.4.2. If the reconstructed signals in an event meet the requirements of one of these algorithms, one records all the waveforms found in the corresponding space and time trigger buffers. To reduce the amount of data to read, samples with an ADC value below a certain threshold (in recent physics runs, it was 1.5 PAs) are excluded (Zero Suppression). More details about triggers are given in section 2.2.4.2.

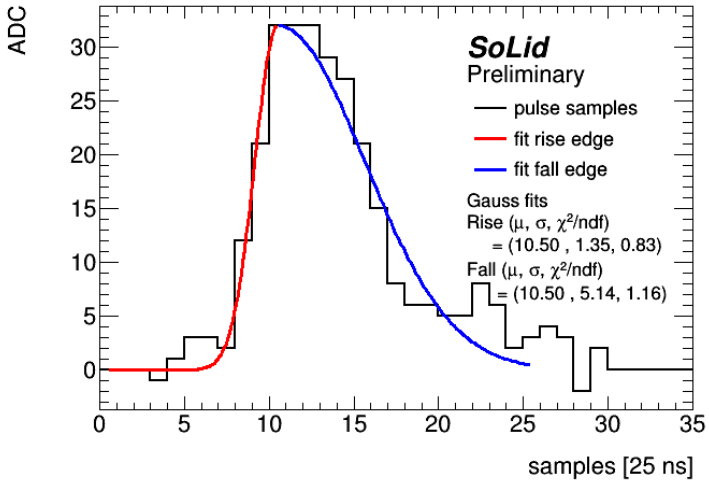


Figure 4.3.: Example of fitted data waveform in which a peak with an amplitude of 1 PA is observed. The fit has two components to correctly reproduced the rising and the falling edge of the peak.

4.2.2.2. List of parameters to tune

The detector's response depends mainly on seven parameters. Therefore, the simulation of this response depends on it too, and has to use accurate values. We present below these parameters, and the way we determined them.

- **Birks constant:** The quenching effect described in 3.2.1 depends on the Birks constant k_B, expressed in mm/MeV. Its value depends on the scintillator chemical composition. The manufacturer of the PVT scintillator [66] quotes $k_B = 0.14$ mm/MeV.

- **Channel gains:** The gain of each channel is determined from un-biased data recorded with the random trigger during physics runs, as said in section 2.2.3.1. Those values are stored in a database every six hours in order to monitor and correct possible time-dependent effects. The gain is obtained by fitting the first peaks of the channel amplitude distribution, each peak corresponding to a number of PAs. The gains are distributed around 32 ADC/PA with a standard deviation of 1.4% (see figure 2.8a). Thus, a good uniformity is obtained across the whole detector. This uniformity allows to consider only one gain for all channels in the simulation as this value is well measured in data.

- **Cube light yield:** This crucial parameter has to be determined in each of the 12,800 cubes the detector is made of. It necessitates an automated - thus systematic - procedure, which I developed. First, Monte Carlo ^{22}Na calibration data was generated using the full simulation, assuming a starting value for the yield of photons reaching the 4 fibres, LY_{RO}. This is taken in a reference cube, located at (X=8, Y=8, Z=30). Then, the corrected Light Yield $LY_{corrected}$ is determined in the

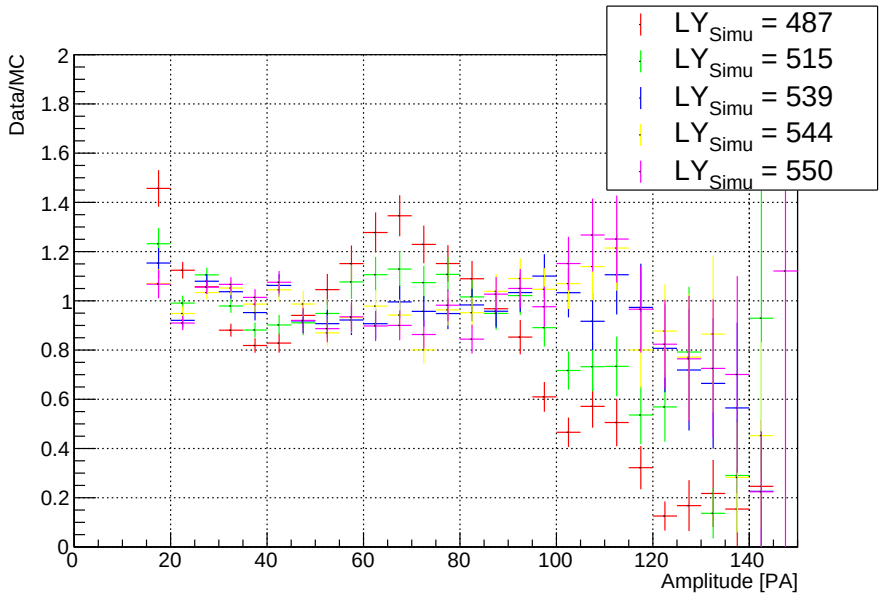


Figure 4.4.: Example of the data/MC agreement found in a central cube (X=8,Y=8,Z=30) for different values of LY_{RO}. Several simulations have been ran and the best value has been found to be 539 photons per MeV per fibre.

PAs distribution generated this way, using the fit procedure described in 3.7 allowing to subtract for fibre attenuations and fibre to SiPM couplings. It is compared with the value measured in real data. Since it is proportional to LY_{RO}, the data/MC ratio can be used to obtain a realistic value of this parameter to implement in the simulation for this reference cube. Then, the values of LY^i_{rec} measured in data in the 12799 other cubes are compared to that in the reference cube : the ratios are used to derive from the reference cube the relevant value of LY^i_{RO} in each cube i.

- **Light leaks:** The light leak is the probability for a scintillation photon produced in a given cube to be measured in a fibre serving a neighbouring cube. This is estimated using neutron capture events and muons. We use horizontal muons that crossed only one cube per detection plane and we measure the amplitude observed by MPPCs serving neighbouring cubes. In the simulation, we assume the quantity of light leaking from a cube to its neighbours to be the same in all of them.

- **Attenuation lengths:** The 3200 attenuation lengths are determined using calibration data (see section 3.6)

- **Couplings:** Same as for the attenuation lengths.

- **Photon Detection Efficiency and Cross-talk:** Those two parameters are measured from the data. In the simulation all channels have the same values. This is justified since a wrong PDE is compensated by the determination of L_{RO} in the end.

4.2.3. Contributions of this thesis work to the simulation

The main contribution of this thesis to SoLid's simulation is to study data/MC discrepancies affecting the measurement of ES quantities. To reach a satisfactory data/MC agreement, we tuned a certain number of parameters and identified and solved a certain number of issues. Most of that is described in detail in section 4.3.

Before this, as already described above, we had taken care of the simulation of the CROSS system. Later, we took part in the implementation in the simulation software of the fibre attenuation lengths, fibre-to-SiPM optical couplings and the LY_{RO}'s. Also, we provided a rigorous and detailed review of the simulation software, which consumed a significant part of our time. It was a necessary step while trying to explain the data/MC discrepancies we observed. We identified issues there. They eventually had limited impact on data/MC comparisons, but we had to correct them to confirm that fact, and help provide the collaboration with a quality tool. Two examples are discussed below.

As previously described the simulation does not propagate the scintillation photons, instead a model based on physics principles is used to determine the average number of scintillation photon reaching a SiPM. This is performed at the beginning of the read-out simulation (see section 4.2.2.1), thus the Geant4 simulation should provide the amount of deposited energy in MeV in each cube. As discussed in 3.2.1 the amount of scintillation photons produced by a scintillator depends on the Birks law (see equation 3.1) on the quenching effect. To take into account this effect, the deposited energy returned by the Geant4 simulation is not actually the exact deposited energy but an effective deposited energy corrected using the equation 3.1. It has been discovered that this correction was not applied correctly. As said the returned information given by Geant4 is the sum of each energy deposited and tracklengths of hits made by a given particle in a given volume:

$$E_{tot} = \sum_{i_{hit}} E_{hit} \text{ and } L_{tot} = \sum_{i_{hit}} L_{hit} \tag{4.1}$$

where, E_{hit} is the deposited energy in one step of the particle and L_{hit} is the corresponding tracklength. In the initial version of the simulation, the effective energy was computed once per particle, via equation 4.1, where dE/dx was approximated by E_{tot}/L_{tot} . We implemented a more correct approach, in which E_{hit} for each hit is replaced by an effective energy, using $dE/dx = E_{hit}/L_{hit}$.

We corrected numerous memory leaks and bugs in the readout simulation code using the *Valgrind* software. Beyond reliability, it allowed to speed up the simulation running time and reduces the quantity of memory needed to run a simulation. This kind of review/bug correction work had to be done numerous times and although it seems worthless to describe it in detail in a thesis, it was rather time consumming.

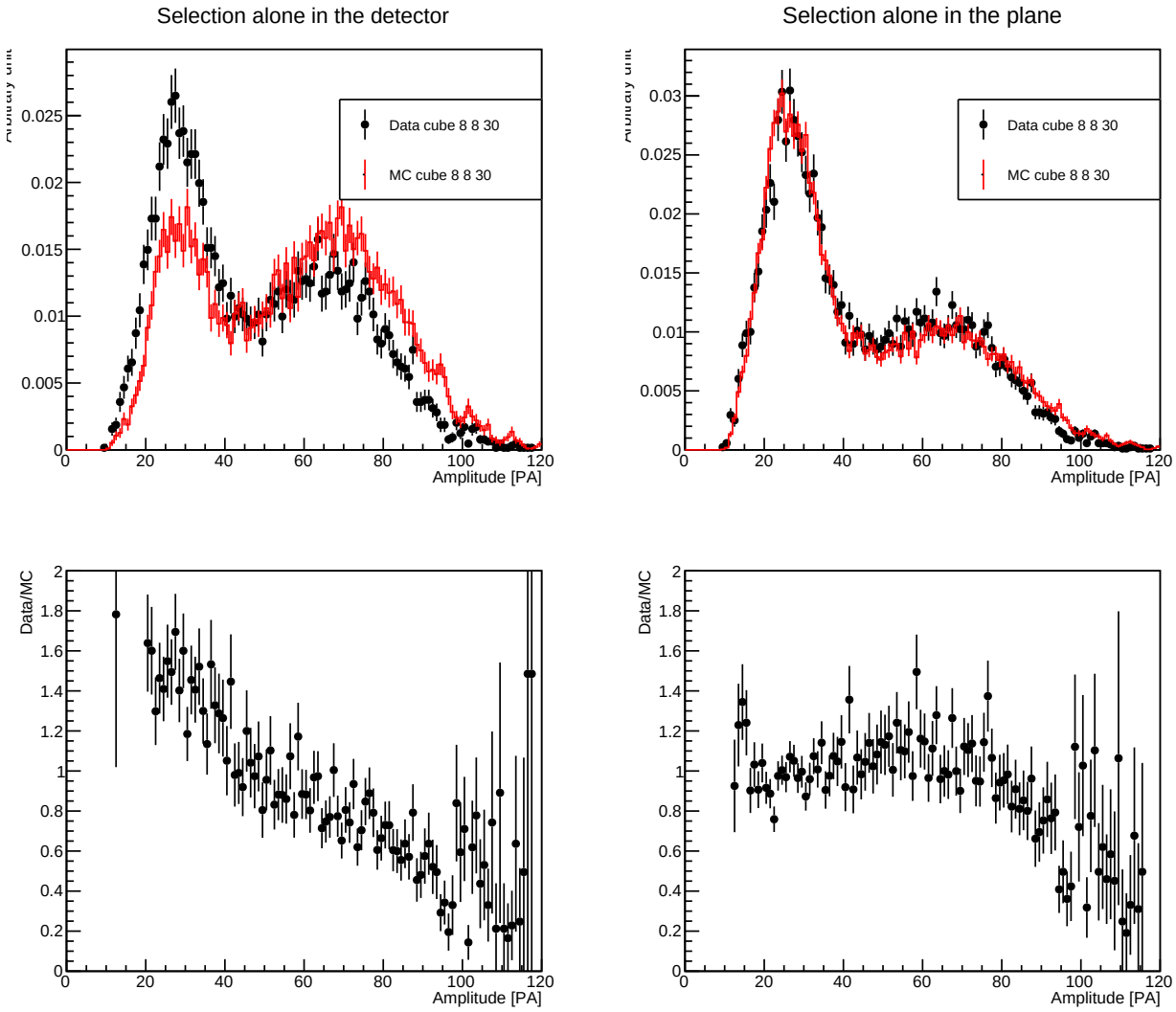
4.3. Energy response tuning

4.3.1. Status early 2018 and First tuning

At the beginning of this thesis the whole structure previously described was already in place. The work to perform consisted to first use 22-Na calibration data with the threshold trigger (see section 3.3.2) to perform a first comparison between data and simulation, to evaluate the amount of work necessary to reach a satisfactory level of data/MC agreement, and to track potential bugs. The detector settings have changed in April 2018 especially the SiPM overvoltage, which was increased and so the visible light yield too (see section 3.4.1). This study was done with a dataset taken before this modification. Also, it was taken with the first version of the calibration trigger: a threshold trigger requesting to have one fibre above a threshold anywhere in the detector. In case of trigger, five planes were read-out, the triggered plane ± two planes around it (see section 3.4.1). This dataset covered the full detector (data had been taken at all the sources positions described in 3.4.1.1).

The idea of this first tuning was to tune only one cube in a central position and determine whether a rough agreement between data and simulation could be found. Several simulations have been running with different values of LY_{RO}, to find the best one. Only a limited number of events were simulated to have the results quickly. The selection required only one cube in the detector. The agreement between data and simulation in the central cube is shown on figure 4.5a. It is very bad. But it was not possible at this stage to state whether it was due only to a bad simulation of the cube or the channels, or to something more global, like the DAQ. Thus a loosened selection was studied by asking to the cube of interest to be only alone in its plane. The new comparison is shown on figure 4.5b. It is indeed significantly better.

The first thing to check was the simulation code especially the part simulating the trigger. Indeed this discrepancy may indicate a problem of trigger efficiency in the simulation. At this moment the developers of SoLid's trigger software had not included a functionality to indicate which channels caused the trigger. Thus a selection was designed to find events in which only one fibre could have triggered and its amplitude was compared between data and simulation in the figure 4.6. The threshold effect is clearly visible in data (the trigger threshold was set at 10 PAs), but not in the simulation. This indicates that the trigger algorithm was not correctly applied in the simulation. A cross-check of the trigger algorithm in the simulation pointed out an issue when a waveform falls between two time blocks and only the second one should be kept. Indeed the algorithm should compute the new waveform beginning timestamp and later the samples before this time should be erased. But this time was not reset properly, implying that wrong samples were erased and even sometimes the one which had triggered. After solving this, the agreement is better between data and simulation on the triggered fibre amplitude distribution. However, the distribution illustrated on the figure 4.5a was only marginally improved.



(a) Cube alone in the detector.

(b) Cube alone in its plane.

Figure 4.5.: Comparison between data and simulation with the old settings using the first ^{22}Na calibration runs. Both simulation and data histograms are normalised to unity and the ratio between data and MC is shown below.

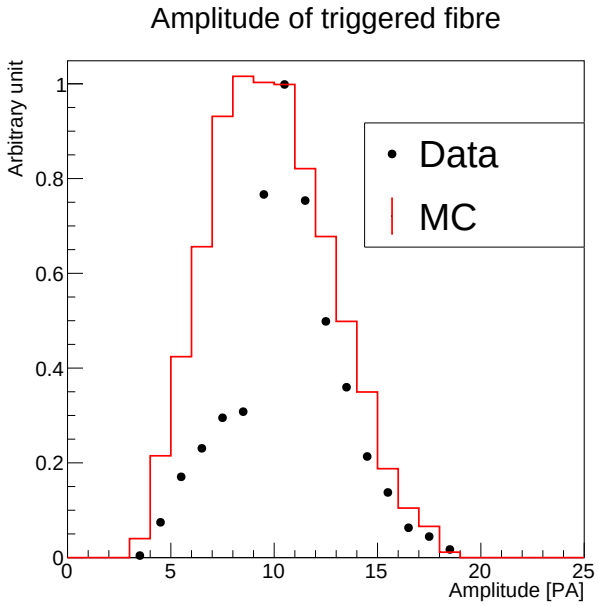


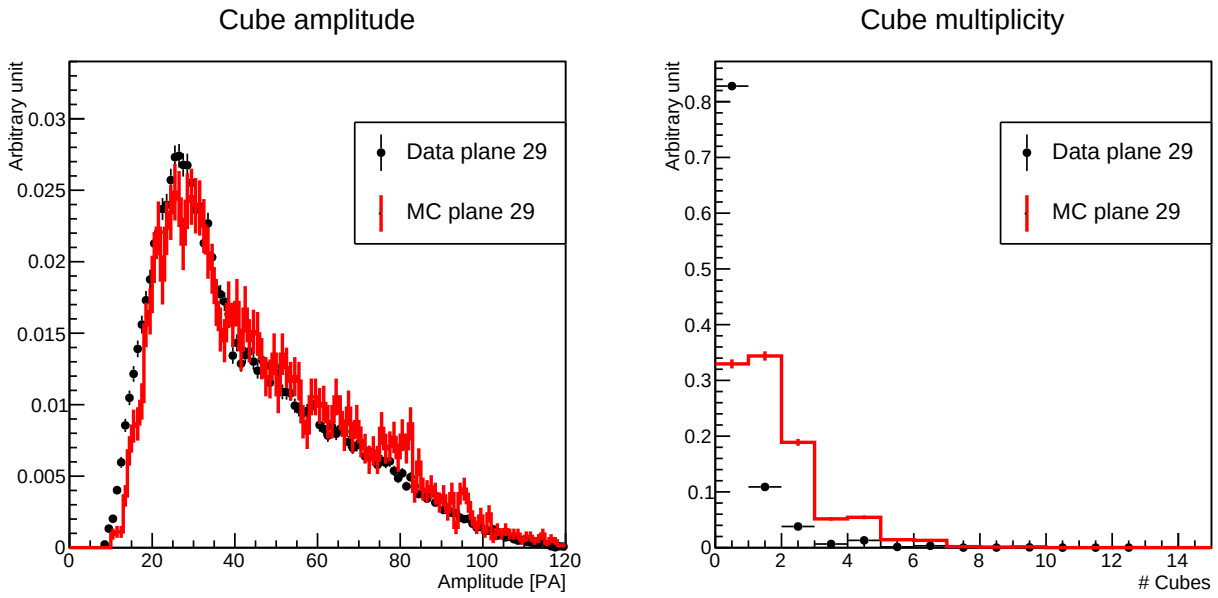
Figure 4.6.: Comparison between data and simulation of the amplitude distribution of the triggered fibre. The imprint of the trigger threshold at 10 PAs is clearly seen in data but not in the simulation indicating an issue in the trigger.

Several other checks were made at low-level objects in the simulation such as the comparison of waveform shape with data. This pointed out another issue on the smearing applied on the sample's amplitude which takes into account the electronic noise. The number of waveforms with two maxima was too high in the simulation but again this could not explain the issue.

DAQ saturation effects

The important data/MC discrepancy at low energy shown on the figure 4.5a could finally be traced down to a DAQ saturation problem, related to the high activity of the source and the absence of a deadtime diagnostic in the SoLid online monitoring.

To determine this, we slightly modified our event selection: several cubes were now allowed in the same event ; one of them had to satisfy $A_{cube} < 30$PAs and be located in one given half module ; and any other cube had to lie in the opposite half module. The idea behind this selection stemmed from the observation that the data/MC disagreement was greatly enhanced when the isolation criterion was to require one cube alone in the closest plane to the source. It pointed out an inability to detect several cubes simultaneously in *different* planes. This new selection assumed that the low energy region in 4.5a is dominated by annihilation gammas, and that additional cubes in the event would be due to the 1.22 gamma emitted at the same time by the ^{22}Na source, that could not be found often enough in the opposite module. If it could be proven that this second gamma is detected more efficiently in MC than in data, then it would explain why events at low energy in the cube of interest are less efficiently selected by the isolation cut in MC.



(a) Cube alone in the detector.

(b) Cube alone in its plane.

Figure 4.7.: Results of the data/MC study while selecting low amplitude cubes in one half-module and looking at the closest plane to the source of the other half-module.

Once this new sample was constituted, we compared, plane by plane, the number of cubes found per event in the opposite module. This can be seen on the figure 4.7b. It is far more frequent in real data to find no cube in the opposite module. This is quantified in the table 4.1, which shows the ratio of the rate of events where at least one cube was found to the rate of empty events in the opposite modules. Both in data and MC it decreases as one gets further from the source, due to the probability for a gamma to cross several planes. However, a large data/MC discrepancy is visible, but decreases in deeper planes. It was subsequently correlated to the amount of deadtime that we measured in these planes. We therefore concluded that the deadtime explained the excessive efficiency in real data of the isolation cut at low energy: deadime of them simply makes impossible to impose the isolation since no additional cube is found in a dead plane. This dead time is absent from the simulation since each simulated event is treated as a whole, with no remaining trace in the DAQ buffers of previous events.

We concluded from this sequence that studying the quality with which MC reproduces the reconstruction of particles requires methods that minimise the influence of the DAQ, which can create discrepancies that have nothing to do with what we need to study. Although most of the discrepancy seen on figure 4.5a is solved by requiring the cube of interest to be isolated only in its plane, see figure 4.5b, it is not clear that all DAQ effects are removed. Deadtime patterns are not precisely understood in SoLid. The work described in this section is partly at the origin of the improvements to the DAQ settings presented in the section 3.4 (Horizontal-Vertical coincidences requirement for the threshold trigger with a careful choice of threshold and random triggers.)

The status of data/MC studies at this point is summarized by the figure 4.5b. The agreement is not yet satisfactory there, which justifies the additional work described in the following sections.

Plane	Ratio (Data)	Ratio (MC)	Data/MC
29	9.1 %	77 %	0.12
28	5.2 %	28 %	0.19
27	2.4 %	11 %	0.22
26	1.8 %	6.3 %	0.29
25	1.5 %	3.4 %	0.44

Table 4.1.: The ratio ($\frac{N_{cube}>=1}{N_{cube=0}}$) of the number of events with at least one cube over the number of events with no cube reconstructed as a function of the plane.

4.3.2. Characterisation of simulated energy response with calibration data

As already explained earlier, the SiPM over-voltages were changed in 2018, and new reconstruction methods were developed (see section 2.4.2.3). Also, the values of attenuation lengths and fibre to SiPM optical couplings measured in section 3.6 were implemented in the readout simulation. It was therefore necessary to pursue the data/MC studies using this new kind of data, although simpler reconstruction objects (fibre level signal, simple sum of such signals for cube reconstruction) still need to be studied too, due to their more direct link to the detector response. Beyond this, it was also time, in late 2018, to investigate the quality of the simulation of the reconstruction of low energy deposits. Indeed it was becoming clear at that time that these deposits, as will be explained in the chapter 5 are crucial to fully exploit the topology of IBD coincidences in the IBD selection. This requires to understand the reconstruction of annihilation gammas (511 keV). In some versions of the analysis, it is made more efficient by considering during event reconstruction all fibre level waveform above 2.5 PAs in amplitude. While other versions kept that threshold at 4.5 PAs, in fear that with lower thresholds, event would be too complex to be dealed with, that background would be higher and that data/MC would be worse. Therefore, in this section, data/MC is not only studied in the low energy region, but with different reconstruction thresholds.

4.3.2.1. First comparison post tuning

Among the various novelties in SoLid's reconstruction: the so-called CCube cube reconstruction, described in section 2.4.2.3. As was explained in this subsection, one of the interests of this method is, in case of pile-up along a fibre, to quantify which part of the fibre signal should be attributed to each cube. This problem was one of the reasons to select isolated cubes in calibration procedures and studying data/MC agreement, as was explained in 3.5.1. It was therefore interesting to study data/MC with this selection criteria off. Moreover, the CCube reconstruction is supposed to be valid even in case of very low energy deposits, when it is likely that only two or three of the four fibres serving a cube will see a signal above the channel reconstruction threshold. We therefore included in the compared samples cubes with less than four fibres above threshold, to study to which extent the simulation can

describe reliably the response of those deposits. The figure 4.8 shows the measured energy spectrum in a typical cube, situated at X=8 Y=8 Z=30, right in front of the source. Three populations are visible both in data and MC, corresponding to cases where two, three or four fibres are above threshold. The relative importance of these populations depend as expected, on the threshold. Above 40 PAs, a region dominated by the four fibres case, the data/MC ratio is imperfect but relatively flat. This worsens at lower energies: a slope is visible and worsens as the reconstruction threshold decreases from 4.5 to 2.5 PAs. Below 10 PAs, the two fibres case dominates, and is more prominent in data than in MC. It is a sign that more fibres are unresponsive in data. To confirm that, we studied a simpler quantity than the CCube energy reconstruction (a likelihood minimisation, see section 2.4.2.3). We studied asymmetries between the waveforms amplitude on individual fibres:

- **Horizontal-Vertical (HV):**

$$\mathscr{A}_{HV}(X,Y) = \frac{A^{Hori}(X,Y) - A^{Vert}(X,Y)}{A^{Hori}(X,Y) + A^{Vert}(X,Y)} \tag{4.2}$$

  with $A^{Hori}_{Fiber}(X,Y) = A^0(X,Y) + A^1(X,Y)$ and $A^{Vert}_{Fiber}(X,Y) = A^2(X,Y) + A^3(X,Y)$.

- **Left-Right (LR):**

$$\mathscr{A}_{LR}(X,Y) = \frac{A^0(X,Y) - A^1(X,Y)}{A^0(X,Y) + A^1(X,Y)} \tag{4.3}$$

- **Top-Bottom (TB):**

$$\mathscr{A}_{TB}(X,Y) = \frac{A^2(X,Y) - A^3(X,Y)}{A^2(X,Y) + A^3(X,Y)} \tag{4.4}$$

We first compared data and MC distributions of those variables in sixteen cubes located in plane 30, surrounding the cube in front of the source ($X \in [6,9]$ and $Y \in [6,9]$). The result is shown on figures 4.9, 4.10 and 4.11. It can be seen on figure 4.9 that values of $\mathscr{A}_{TB}$ equal to -1 or +1, a direct evidence that no signal was found above threshold on either the top or bottom fibre, are far more frequent in data. The same is obtained for $\mathscr{A}_{LR}$ (see figure 4.10). The study of the $\mathscr{A}_{HV}$ provided a complementary information: there, data/MC is poor all across the distribution range, and worse close to -1 or +1. These issues were traced down to bugs in the measurement procedure of attenuation lengths and fibre to SiPM optical couplings and in their transmission to the relevant reconstruction database. This study allowed to correct for this.

Before these corrections, a new comparison was performed, selecting cubes with four fibres above a threshold of 0.5 PA. There, to remain close to basic detector information, we reconstructed the cube energy as the simple sum of the fibre waveforms amplitude. As can be seen on figure 4.12, a better agreement is observed, at the 20% level in most cubes. The fact to tune in the readout simulation the global L_{RO} after the values of the attenuation lengths and fibre to SiPM couplings have been implemented compensates in part the problem mentioned above. However, we considered we had

to try to reach a better agreement (in particular at low energy). This will be described in the next subsections, where a more systematic study (in all the detector's cubes) was also carried out. It can be noted that from now on, we will focus on cubes with four fibres above the threshold 2.5 PAs. As is shown of figure A.3 in section A.2.1, which shows the efficiency of this criteria as a function of the true energy deposited, we see reasonable efficiency is reached down to 300 keV. Moreover data/MC comparison including all types of cubes, whatever the number of fibres above threshold, will be shown in section 4.3.3, using a sample of BiPo events. In order to stay close to basic aspects of the detector's response, we will also study in the following sections energy distributions where the energy is the direct sum of the amplitudes measured on each fibre.

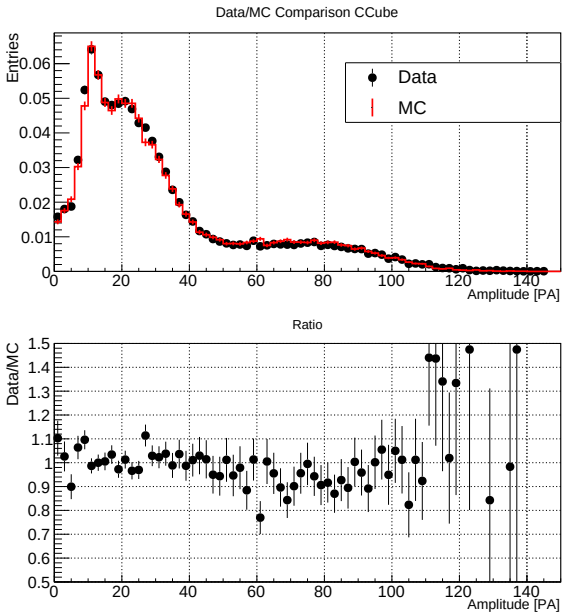
4.3.2.2. First large scale data/MC study

The next phase of this work was carried out between September 2019 and January 2020. The goal was to assess the quality of data/MC agreement once the issues mentioned in the previous section concerning the attenuation lengths and the fibres to SiPM optical couplings had been treated and the simulation global cube light yield retuned accordingly. It was also to provide the first large scale comparison, not focussed on only a few cubes. This was necessary since the collaboration aimed at a first physics results for the 2020 winter conferences.

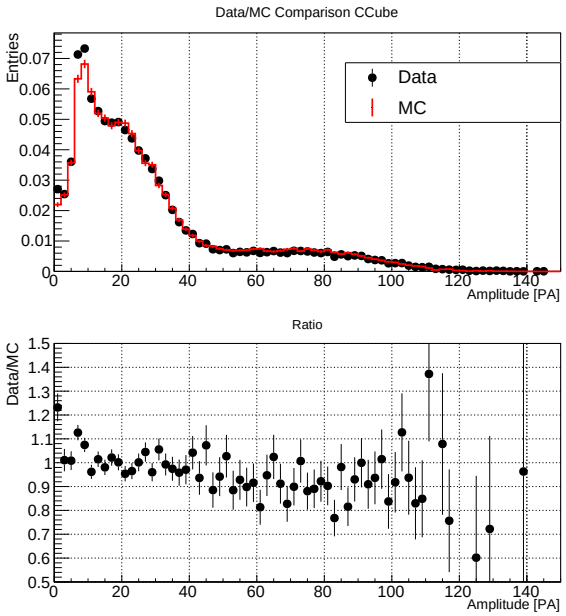
That implied several issues. First, it required to have enough statistics in all the cube in the modules surrounding the source. The calibration ^{22}Na sample taken in September 2018 with the random trigger was the best one for data/MC comparison since it is the least affected by DAQ issues and allows to study low energy region. However, the duration of this run and the random trigger efficiency (see section 3.4) lead to sufficient statistics only in a few cubes in front the source. Therefore, we use the threshold trigger dataset described in the section 3.4, table 3.2.

Using a threshold trigger dataset required to overcome two additional issues. In real data, when a positive trigger is issued, the DAQ system readouts buffered data taken 6.4 μs before and after the signal which caused the trigger. The activity of the source is high enough so that another ^{22}Na can happen during this period of time. In MC, this is not the case: every ^{22}Na decay is an event treated independently from any other decay. Consequently, in MC, every ^{22}Na event has to be the cause of the trigger, thus the energy distribution in any studied cube is distorted in the low energy region by the necessity for two channels to see a signal above 7.5 PAs. This is not the case in real data: a fraction of the entries in a cube energy distribution is not subject to this bias since an earlier or later decay cause the trigger. For this reason, an artificial discrepancy was first observed in data/MC comparison at low energy. To overcome this effect, we computed the reconstructed distributions of energy deposits that could caused the trigger, selecting cube events with two fibres above 7.5 PAs (and the other ones above 5.5 PAs). This configuration forbids to study data/MC below 5.5 PAs (around 250 keV).

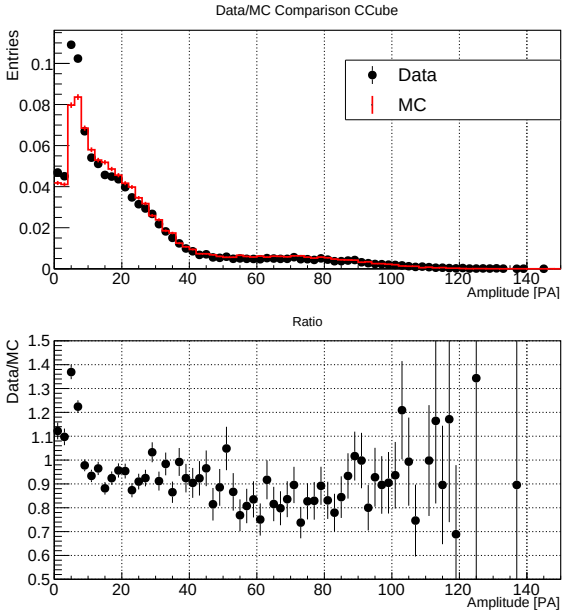
Another issue is yet again due to deadtime. To acquire enough statistics, we first considered in each cube the sum of the nine energy distributions obtained with nine source positions (see 3.4). Their shape differ: the farthest the source from the cube, the more energy lost in the cubes between the source and



(a) Fibre analysis threshold at 4.5 PAs.



(b) Fibre analysis threshold at 3.5 PAs.



(c) Fibre analysis threshold at 2.5 PAs.

Figure 4.8.: Comparison between data and simulation of one cube as a function of the fibres threshold.

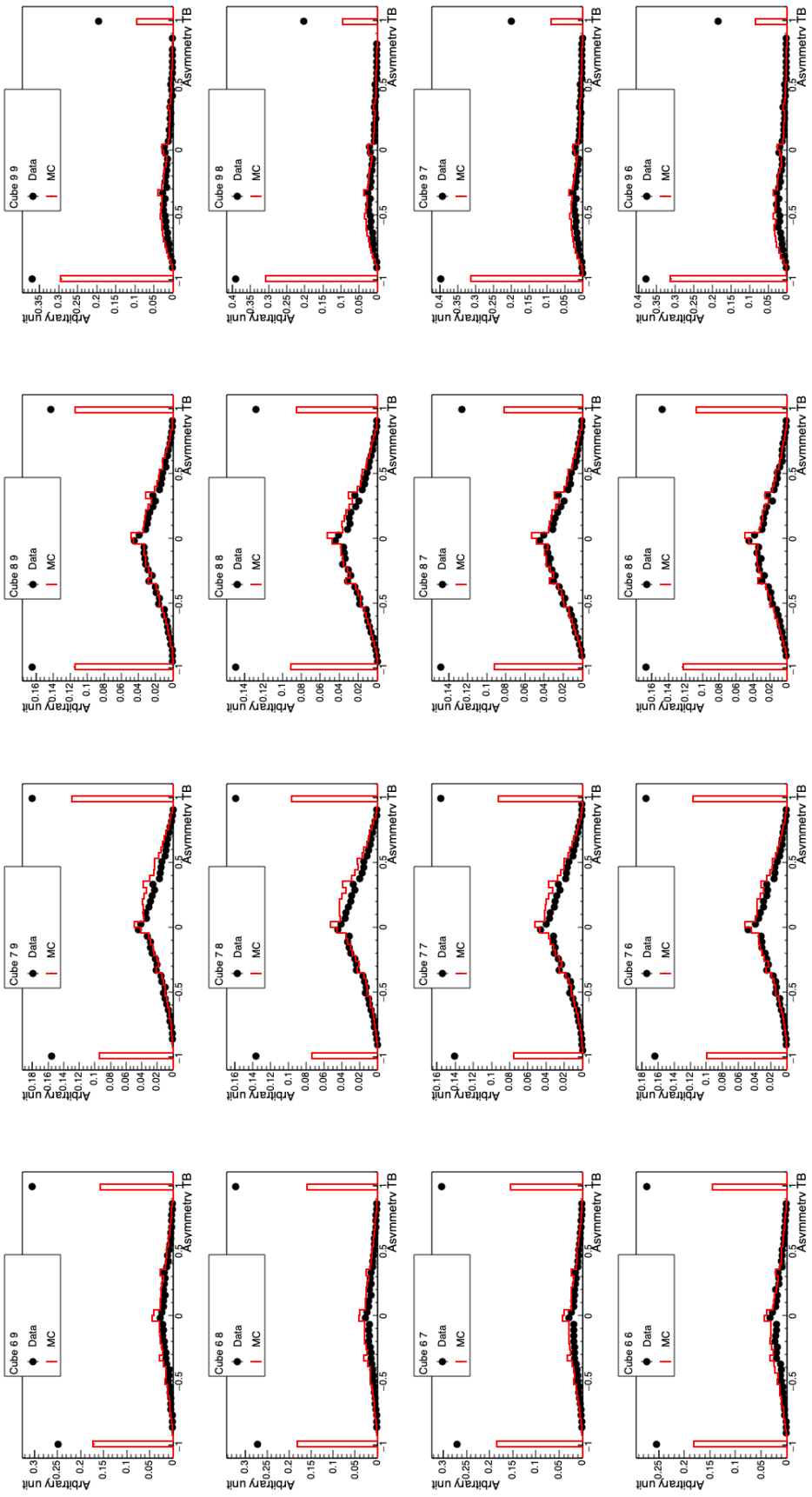


Figure 4.9.: Comparison between data and simulation of the top-bottom asymmetry for sixteen cubes around the source. A clear discrepancy is observed for extreme bins indicating a lack of cubes with less than four fibres in the simulation.

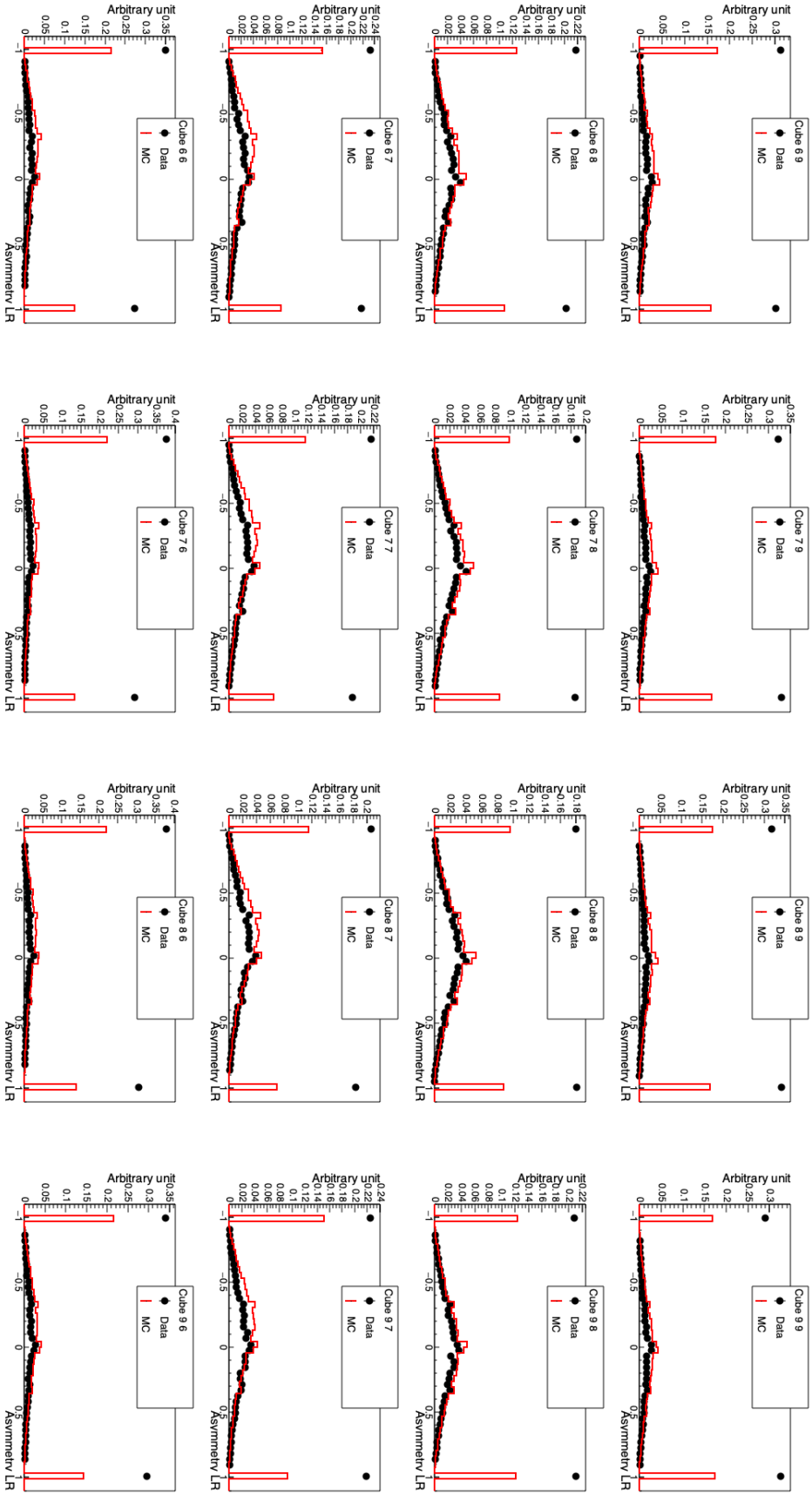


Figure 4.10.: Comparison between data and simulation of the left-right asymmetry for sixteen cubes around the source. A clear discrepancy is observed for extreme bins indicating a lack of cubes with less than four fibres in the simulation.

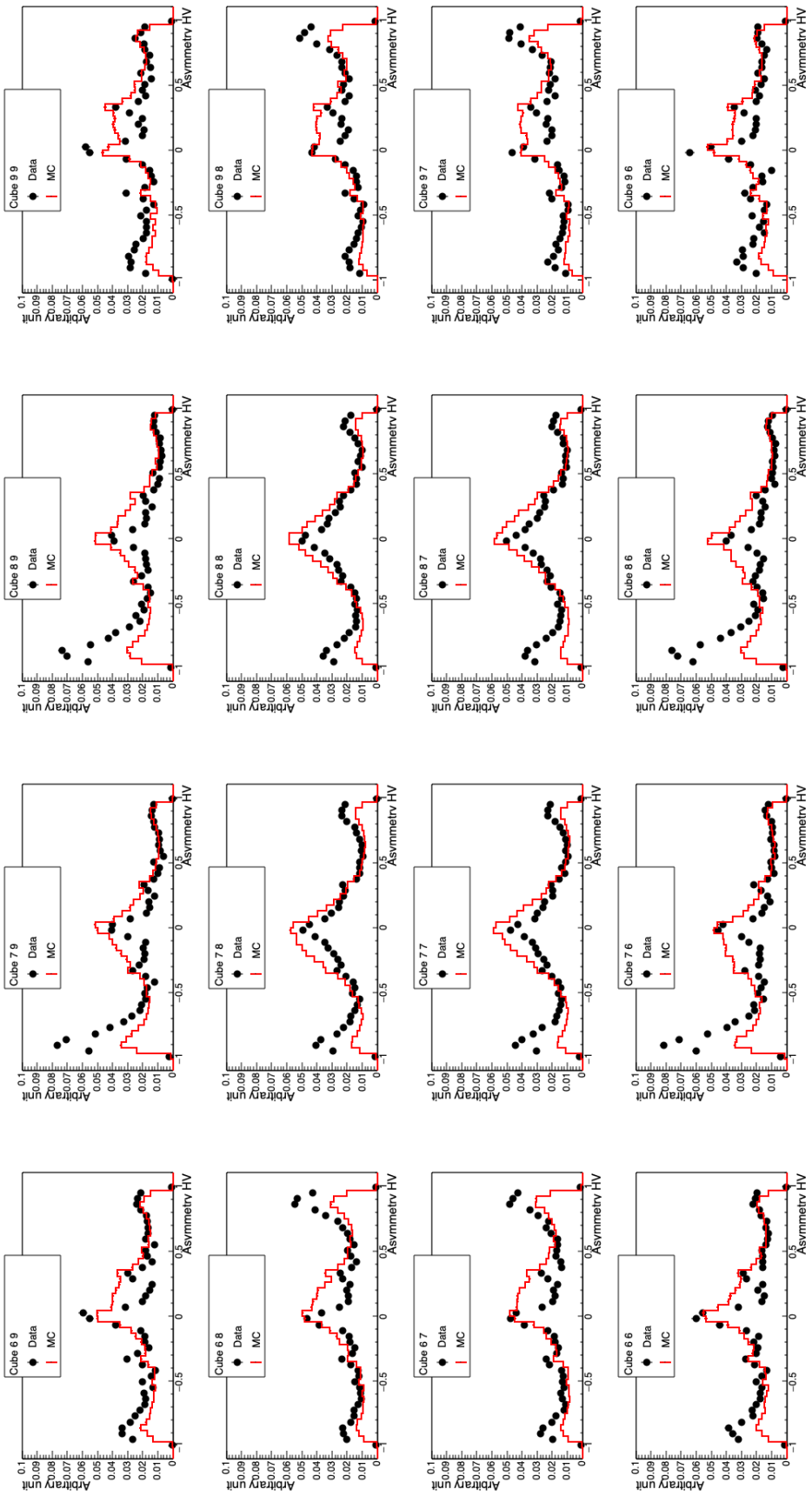


Figure 4.11.: Comparison between data and simulation of the horizontal-vertical asymmetry for sixteen cubes around the source. For most of the cubes a discrepancy is observed indicating a wrong tuning of fibre parameters (coupling and attenuation length).

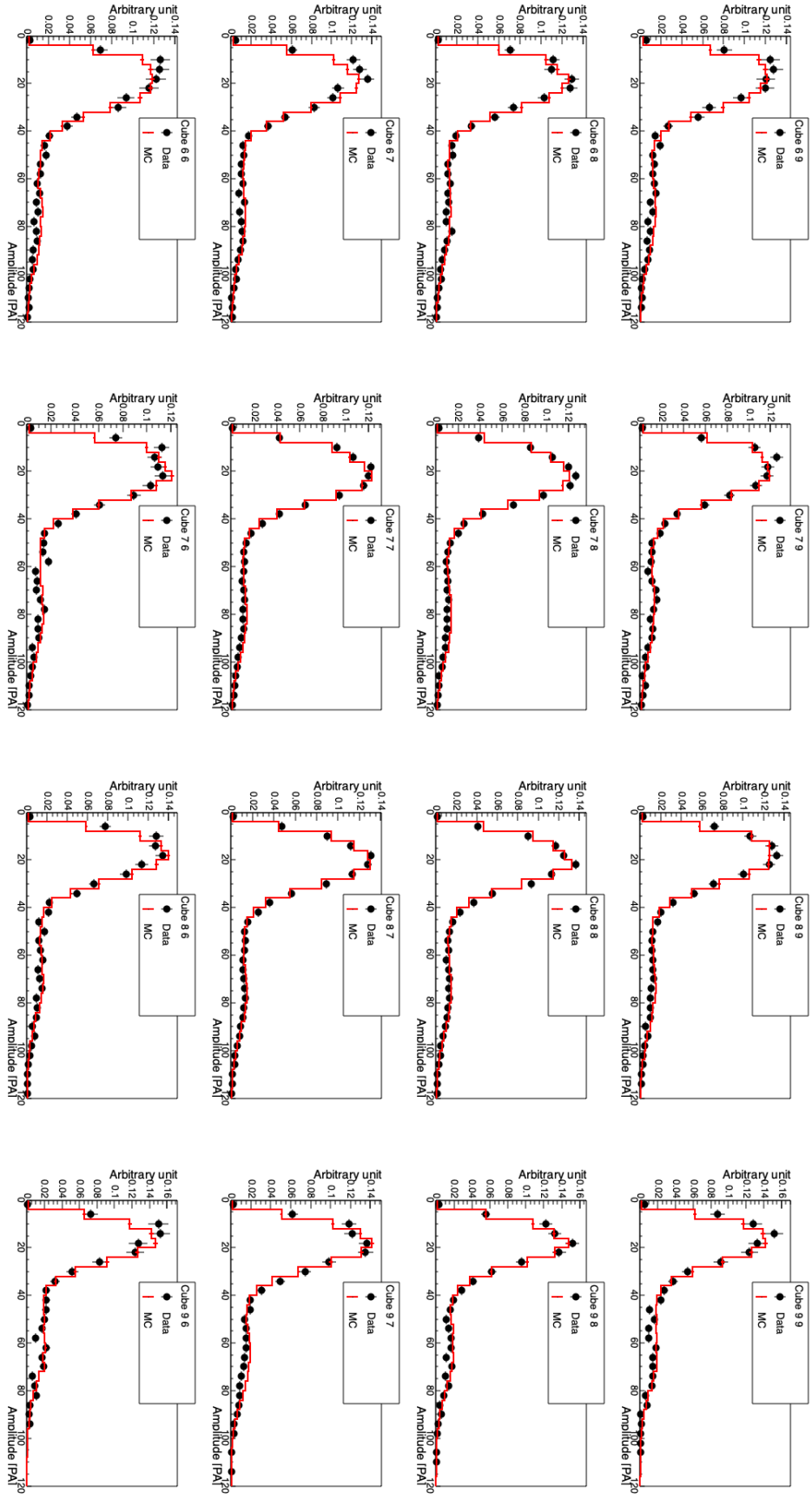


Figure 4.12.: Comparison between data and simulation of the cube amplitude computed with SCube for sixteen cubes around the source. A good agreement is observed.

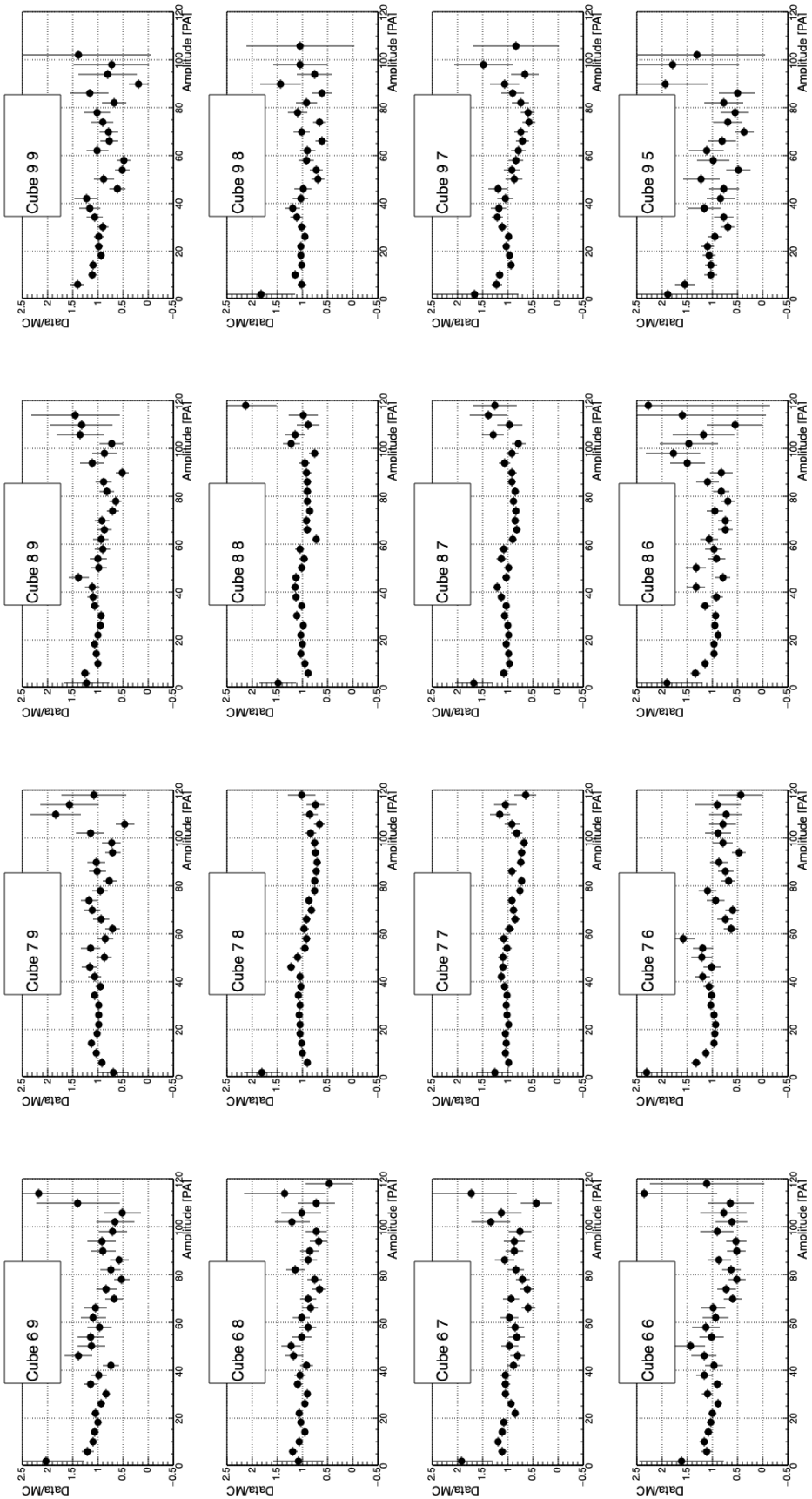


Figure 4.13.: Ratio between data and MC of the figure 4.12.

the cube of interest. To not include an artificial data/MC discrepancy, the exposure time in each source position must be the same in data and MC (actually the relative exposure time). Unfortunately, the time in data was not precisely known due to the unavailability of a reliable deadtime measurement. For this reason, we compared only distributions taken with the same source position.

Ideally, we would study the results in all and every cube individually to prove the details understanding of the detector. This is not easy in practice, since we tuned thousands of cubes. A simple solution would be to compare global distributions, i.e. the sum of the energy distributions measured in all the cubes. However, discrepancies in one region of the detector could be compensated for by an opposite discrepancy in another region, blinding us to a baseline-dependant data/MC issue that would impact the oscillation fit. Presently, the baseline "granularity" of SoLid's oscillation fit is at the level of a module (ten planes). It should therefore be acceptable to regroup all cubes of a module to compare data and MC energy reconstruction. This is what is shown on the figure 4.14. An attempt to keep track of the results in individual cubes will be presented for the next phase of the data/MC work in section 4.3.2.3.

On figure 4.14, one can see that above around 50 PAs the data/MC ratio is flat within 5% or better. Below this value, we observe a clear deficit in data. Interpreting such a discrepancy close to the edge caused by the trigger threshold is delicate. A study in 4.3.2.3 will explain a part of the difference. Another part could be attributed, again, to deadtime effects. Indeed, this low energy region receives three kinds of contribution: the long tail of the 1.274 MeV gamma spectrum, the distribution of annihilation gamma interacting in the same module as the 1.274 MeV gamma and that of the other annihilation gamma, interacting alone in the second module. Although a modification of the threshold trigger - coincident signals above 7.5 PAs on two crossing fibres rather than one are now looked for - allowed to improve the deadtime rate, it is still high (figure 4.15). Importantly, it is higher in the first plane of the module, where annihilation gammas interact more often than 1.274 MeV gamma due to a shorter mean free path. Therefore, more events are lost in the low energy region due to deadtime.

To confirm the quality of the simulation after an additional improvement of the procedure measuring the fibres attenuation lengths and couplings to the SiPMs, and to assess it in a region of low energy not accessible with the threshold trigger, a new phase was undertaken, described in the following section 4.3.2.3.

4.3.2.3. Final data/MC study

In June 2020, a long ES calibration campaign was finally undertaken (see section 3.4.1). In particular, high statistics samples of ^{22}Na random trigger data were taken. They allow to study the low energy region more easily, since no trigger thresholds are applied now, only a reconstruction threshold at 2.5 PAs for all fibres. Besides this, the same selections were applied. The result can be seen on figure 4.16a, that compares the agreement between the data and simulated energy distributions for all the cubes found in planes 25 to 34. We observe here that both distributions agree within a few percents over the whole range.

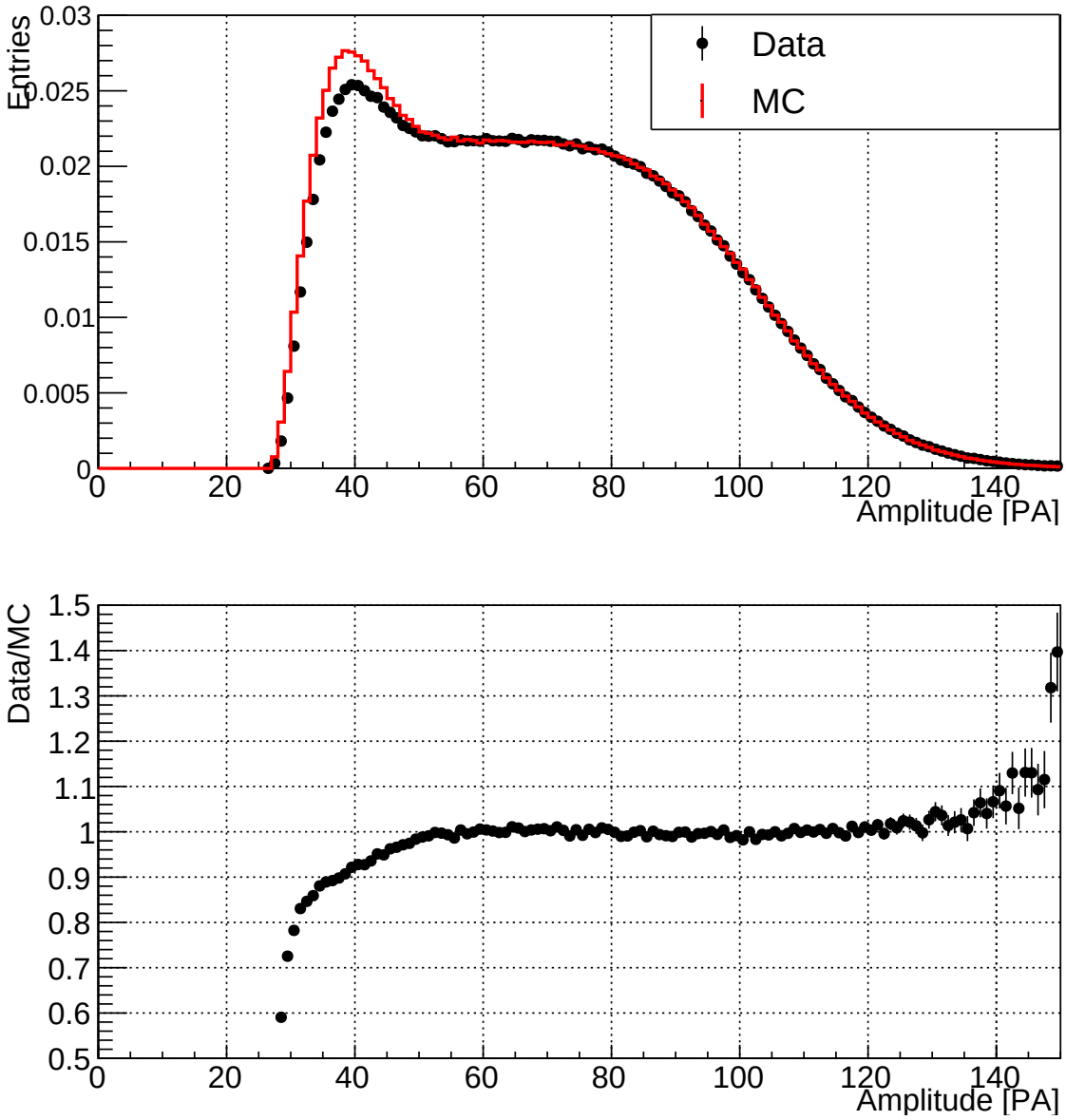


Figure 4.14.: Comparison between data and MC using the threshold trigger runs taken in September 2018. The distributions in the 2560 cubes composing the two half-modules surrounding the source are added together then the data and MC total distributions are normalised to the unity.

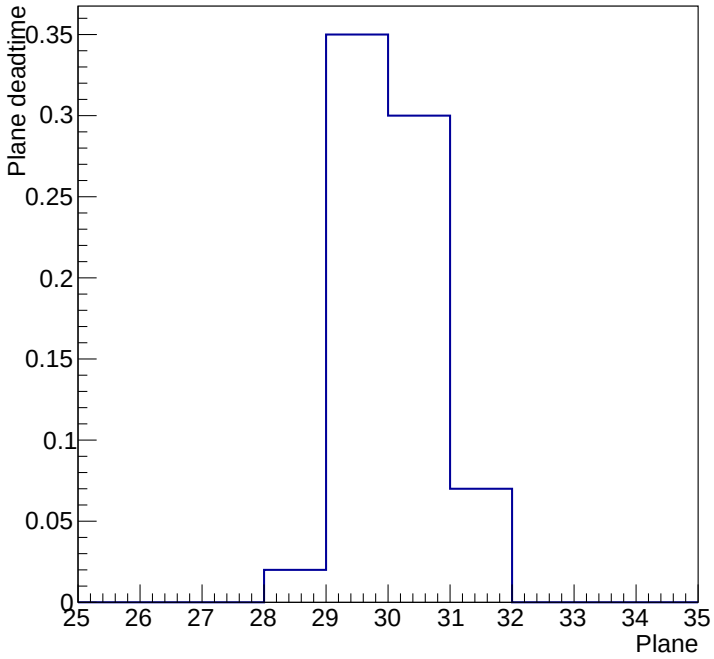
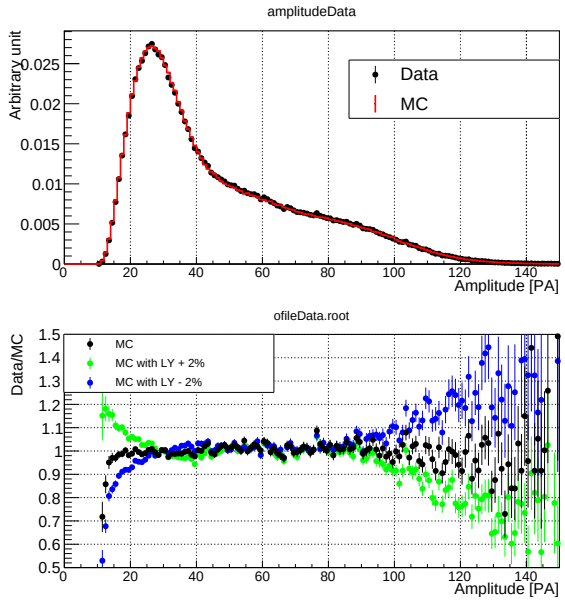
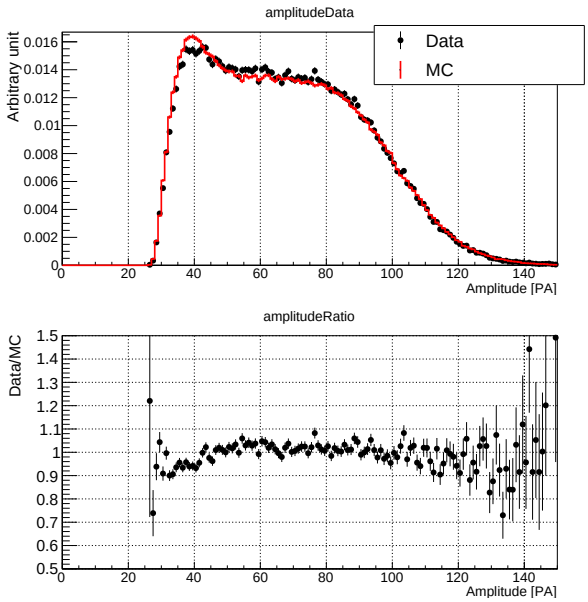


Figure 4.15.: Fraction of time during which the plane is not recording data because of the plane deadtime effect explained in section 2.2.4.3 as a function of the plane. The source in positioned at the center between planes 29 and 30. The data is from one run of the September 2018 calibration campaign with the ^{22}Na.



(a) Cube amplitude asking for each cube to have the four fibres above 2.5 PAs. On the bottom there is the ratio between data and MC. We show two additional ratios which corresponds to an increase (descrease) of the LY_{RO} in the simulation of plus (minus) 2%.

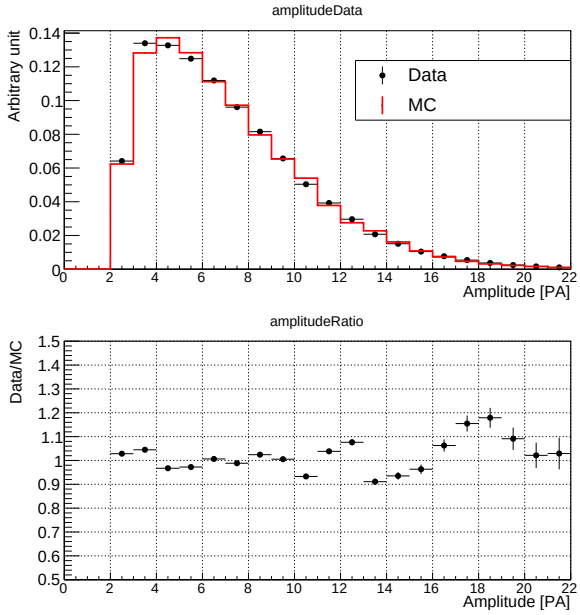


(b) Cube amplitude by asking the cube to be able to trigger the threshold trigger (at least one horizontal and one vertical fibre above 7.5 PAs) and the four fibres above 5.5 PAs, as in the study shown in section 4.3.2.2)

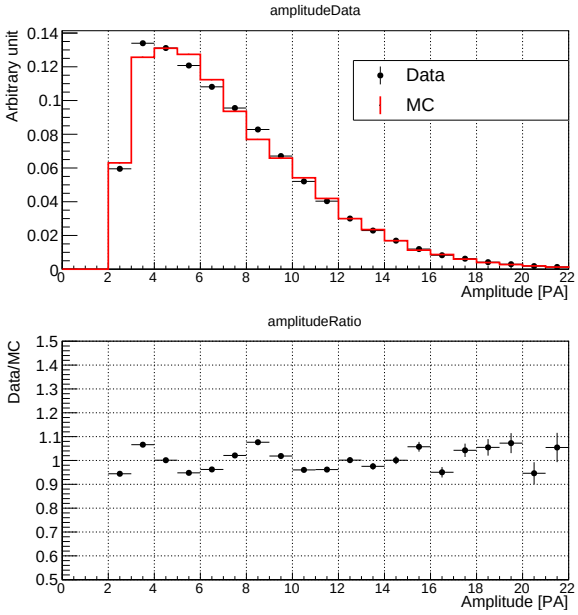
Figure 4.16.: Comparison between data and simulation of the cube amplitude.

We have studied the robustness of this agreement against the higher thresholds that could be necessary in the analysis of physics data (e.g. if for operational reasons or if the idea to use low thresholds explained in chapter 5 eventually proved bad). We selected a subsample of the data shown on figure 4.16a by requiring for each cube two fibres above 7.5 PAs and two above 5.5 PAs (like in 4.3.2.2). On figure 4.16b, one observes a small (around 7%) discrepancy appears at low energy, in the region where the efficiency of these thresholds requirements deviates from 100%, making it possible for data/MC discrepancies to be apparent. To understand the origin of this disagreement, we plotted on figure 4.17 and 4.18 the distributions of signal amplitudes on individual fibres before the threshold described above were applied, and after each individual threshold was applied, to follow sequentially the impact of these cuts. No clear pattern can be observed here: the fibre level distributions show as far as we can tell a satisfactory agreement. Remaining discrepancies are consistent with a parameter tuning at the 1 or 2 % level, i.e. a level of accuracy that we can hardly push further down due to the precision of the measurement of LY_{rec}, attenuations lengths and fibre to SiPM couplings. Adding up these small discrepancies as we cut on the four fibres simultaneously explains the 7% deficit in the low energy region of figure 4.16b. Note that this must also provide a part of the explanation for the discrepancy observed in figure 4.14.

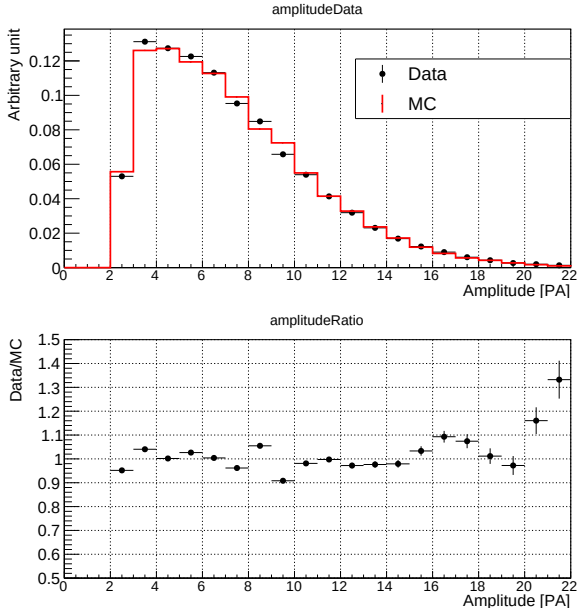
The good agreement observed overall should ideally be confirmed cube by cube. It is hardly possible to study thousands of cubes. To "automate" this study we compute for each cube the χ^2 per degree of freedom between the energy distribution data/MC ratio and the flat hypothesis. This



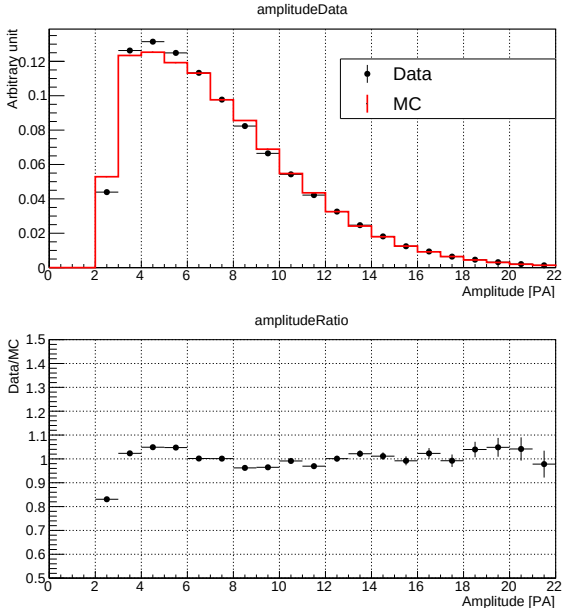
(a) $A_{Fiber}^{Hori,1}$



(b) $A_{Fiber}^{Hori,2}$

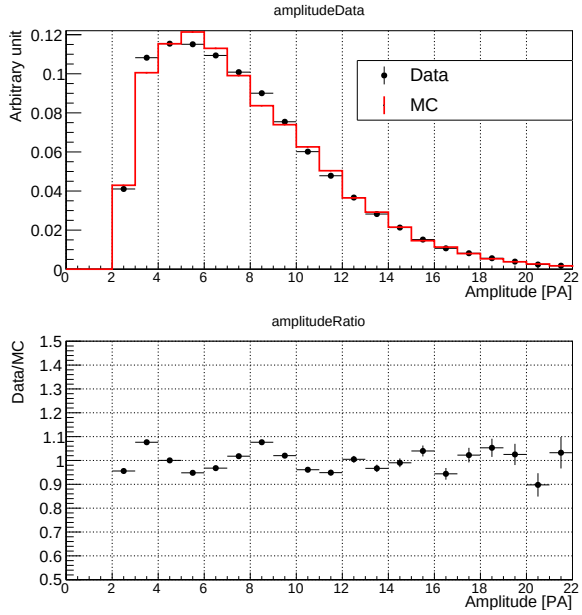


(c) $A_{Fiber}^{Vert,1}$

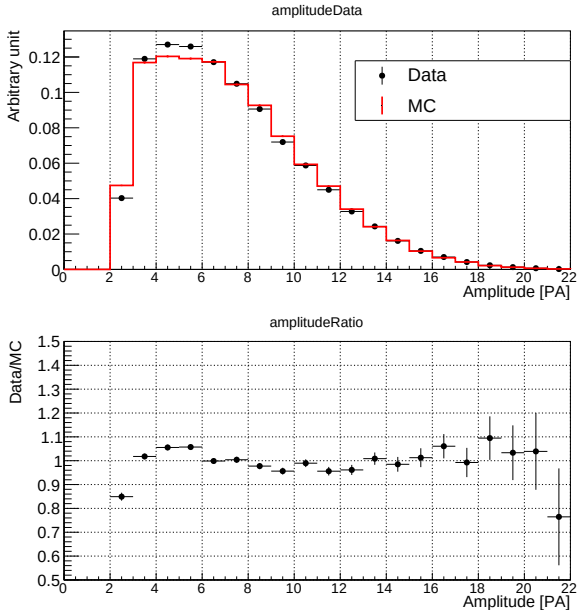


(d) $A_{Fiber}^{Vert,2}$

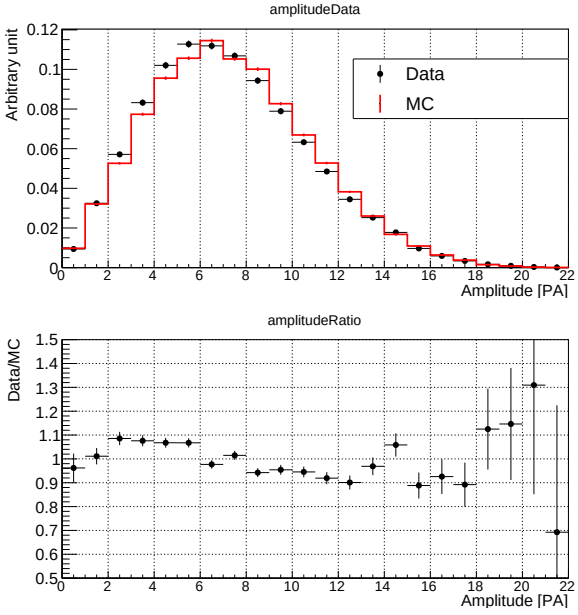
Figure 4.17.: Comparison between data and MC of individual fibres amplitude while selecting cubes such as: the four fibre amplitudes are above 2.5 PAs and the cube amplitude is below 45 PAs in order to highlight a discrepancy at low cube amplitude.



(a) $A_{Fiber}^{Hori,2}$

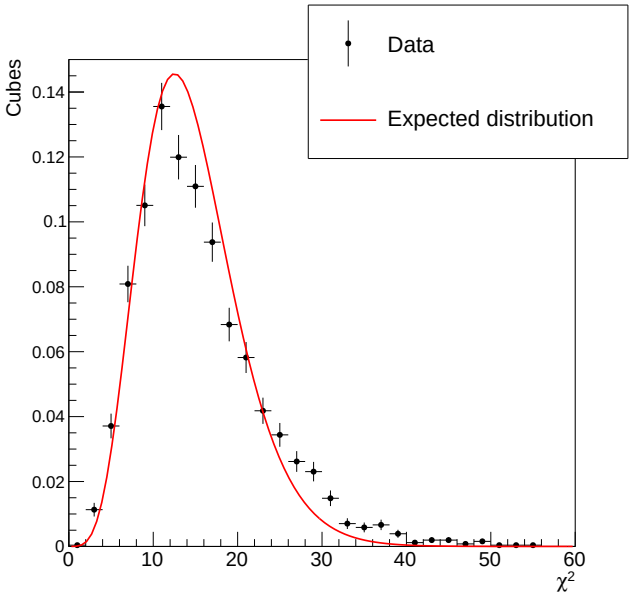


(b) $A_{Fiber}^{Vert,1}$

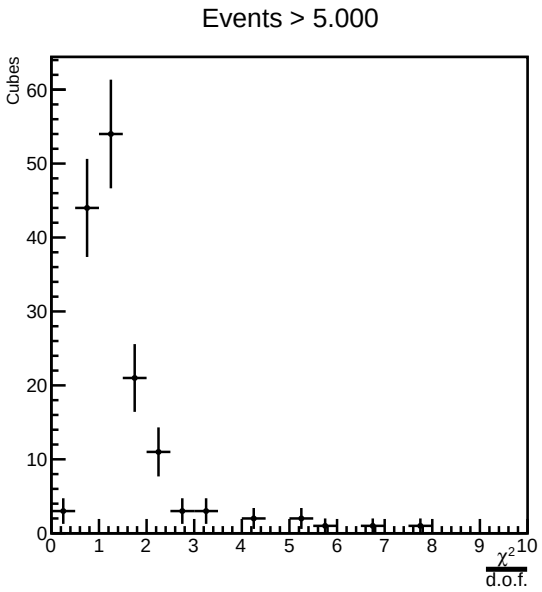


(c) $A_{Fiber}^{Vert,2}$

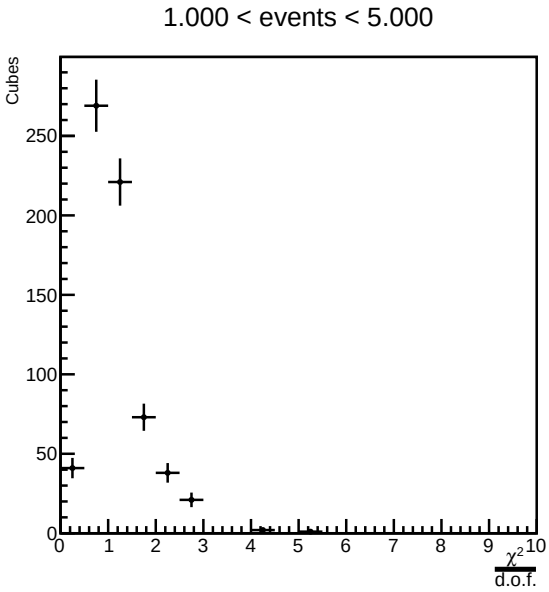
Figure 4.18.: Comparison between data and MC of individual fibres amplitude while applying sequentially the selection: $A_{Fiber}^{Hori,1} > 5.5$ PA and $A_{Fiber}^{Hori,2} > 7.5$ PA and $A_{Fiber}^{Vert,1} > 7.5$ PA on top of the selection shown in 4.17.



(a) χ^2 of all cubes



(b) χ^2_{dof} for cubes with more than 5k entries.



(c) χ^2_{dof} for cubes with a statistic between 1 to 5 k entries.

Figure 4.19.: χ^2 distributions of the data/MC comparison in all the 2,560 cubes of the calibrated module.

chi-square is computed between 60 and 150 PAs in order to quantity the agreement between data and MC in the Compton edge region used for the evaluation of the light yield (see section 3.5). This is shown on figure 4.19 where the χ^2 distribution is plotted and compared with the expected distribution for this number of d.o.f.. Two cases are shown: when the data distribution in a cube has more than 5.000 entries, or when it has between 1.000 and 5.000 entries. This is necessary to not over interpret a bad χ^2_{dof} that can occur even if the relative agreement (residuals) is good, if the statistics is high.

In most cubes, a correct value of the χ^2_{dof} is found. On figures 4.20 (high statistic) and 4.21 (low statistic), a random selection of "good" cubes ($\chi^2_{dof} < 1$) is shown, confirming the good agreement. Figures 4.22 (high statistic) and 4.23 (low statistic) shown a random selection of "bad" χ^2_{dof} ($\chi^2_{dof} > 1$). The shape of the data/MC ratios is typical of shifts on the LY_{RO} (see the effect shown on figure 4.16a, when the light yield in readout simulation is moved by 2%). This is not too surprising, considering the precision with which we can determine this light yield.

A part of the discrepancies in figures 4.20 to 4.23 can also in principles be attributed to the accuracy with which the simulation describes the neighbour fibres and cubes (as already explained in section 3.2.2). The size of this effect is evaluated on figure 4.24, which shows a subsample of the sample on figure 4.16a, when cubes are required to be alone in their plane.

This cut depletes the central part of the energy distribution, due to the kinematics of Compton scattering. The effect is a bit less pronounced in data (by a few percents) which indicates that energy deposits in neighbour cubes are less often reconstructed. The level of agreement here shows the quality we still reach even when the individual remaining data/MC discrepancy from many cubes and fibres is cumulated. It indicates the accuracy with which MC can now describe multi deposits events, an important part of IBD candidates. BiPo studies in subsection 4.3.3 will confirm this.

4.3.2.4. Tagged annihilation gamma

As the selection presented at the ICHEP 2020 conference is based on the reconstruction of annihilation gamma, a specific study has been performed to quantify the discrepancy between data and MC of low energy deposits. The ideal source to use is the ^{22}Na since two annihilation gammas are emitted in coincidence with the 1.274 MeV gamma. The random trigger runs described in 3.4 have been used in order to reduce the probability to have a DAQ issue and to have lower threshold. In order to isolate annihilation gamma the idea is to tag the interaction of 1.274 MeV gamma in one module. This is performed by asking to have above 60 PAs (0.625 MeV) and below 150 PAs (1.56 MeV). The upper selection aims to reject high energetic events such as muons crossing the detector. Once the module in which the 1.274 MeV gamma and one of the annihilation gammas have interacted has been tagged this way, one knows that the second annihilation gamma must have crossed the opposite module in the same time buffer and been the only particle to have done so (modulo rate back-scattering from the initial module). It is therefore possible to study the reconstruction efficiency of pure low energy deposits and the quality of low energy deposits reconstruction.

The distributions of the amplitude (sum of the four fibres amplitudes) measured in all the cubes found in the events of this sample is shown on figure 4.25a which compares the distribution found in real data and in the simulated samples. We can see that the shape of the reconstructed energy distribution is very well predicted by the MC. The ratio of the data distribution over the MC distribution is flat within a few percent across the full spectrum. However, on figure 4.25a histograms have been normalised to the unity. Thus there is no control of the efficiency to reconstruct small energy deposits.

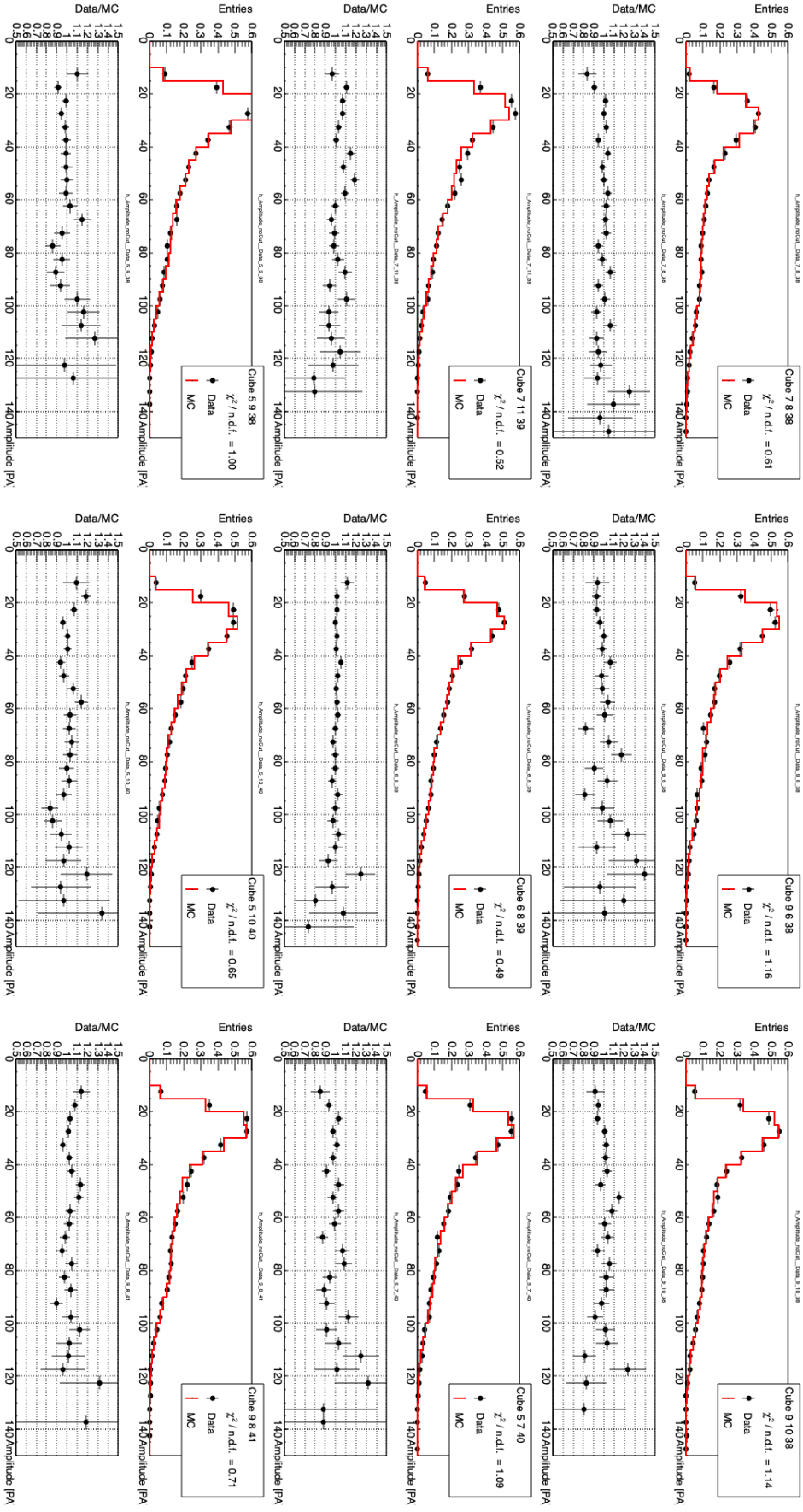


Figure 4.20.: Data and MC comparison for a random cube selection with more than 5.000 events in the data histograms and a good reduced chi-square value.

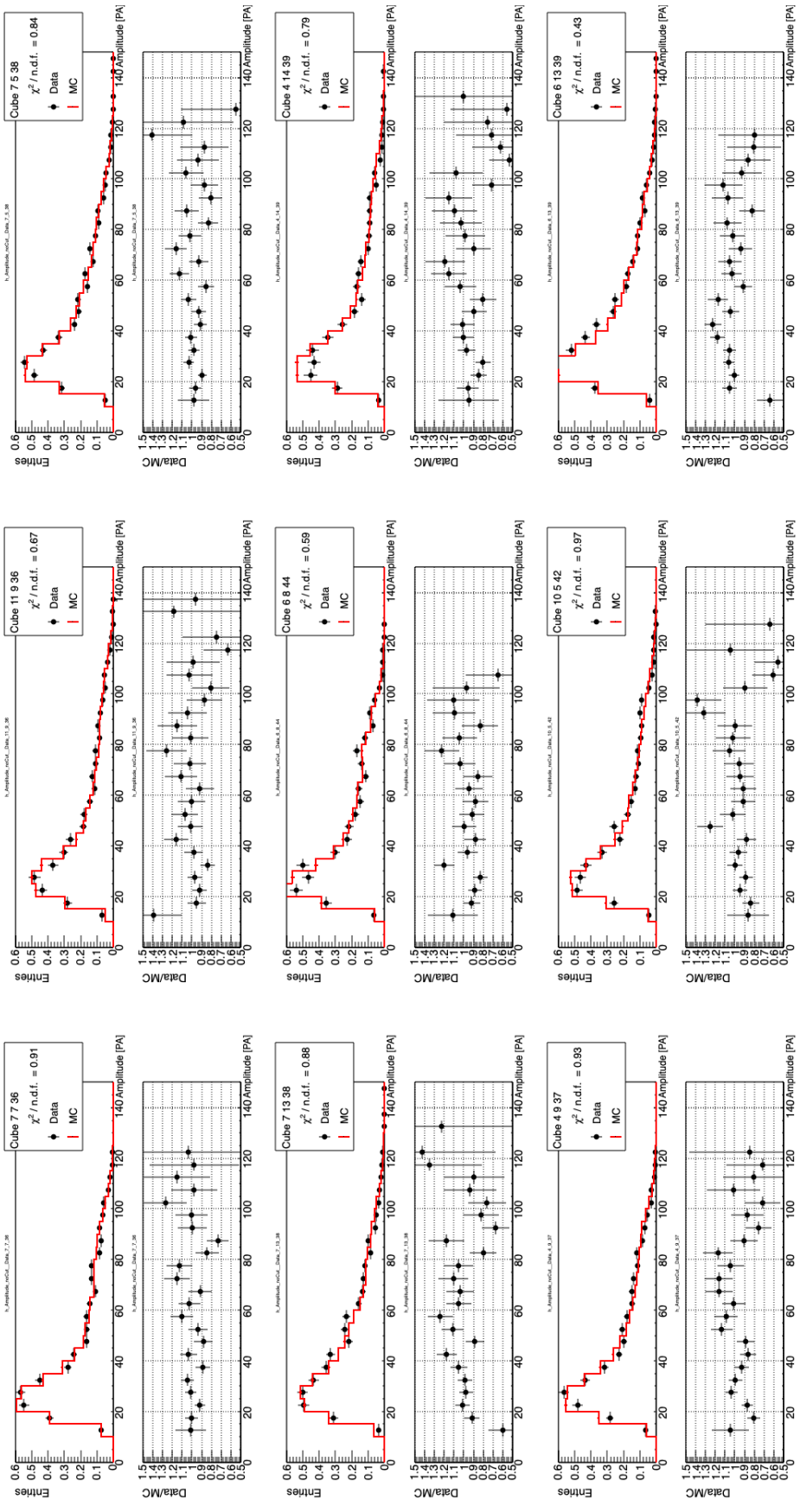


Figure 4.21.: Data and MC comparison for a random cube selection with between 1.000 and 5.000 events in the data histogram and a good reduced chi-square value.

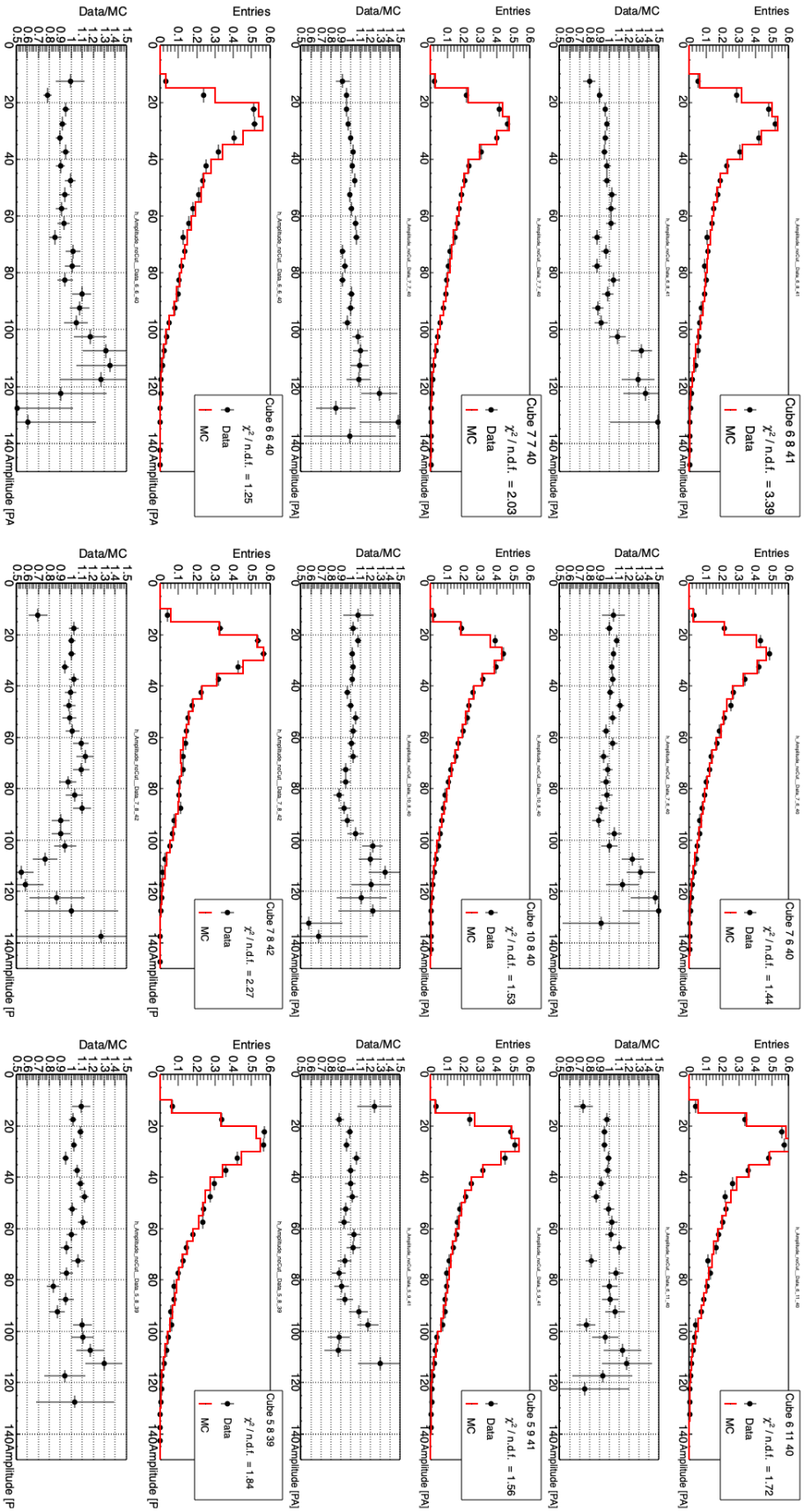


Figure 4.22.: Data and MC comparison for a random cube selection with more than 5.000 events in the data histograms and a bad reduced chi-square value.

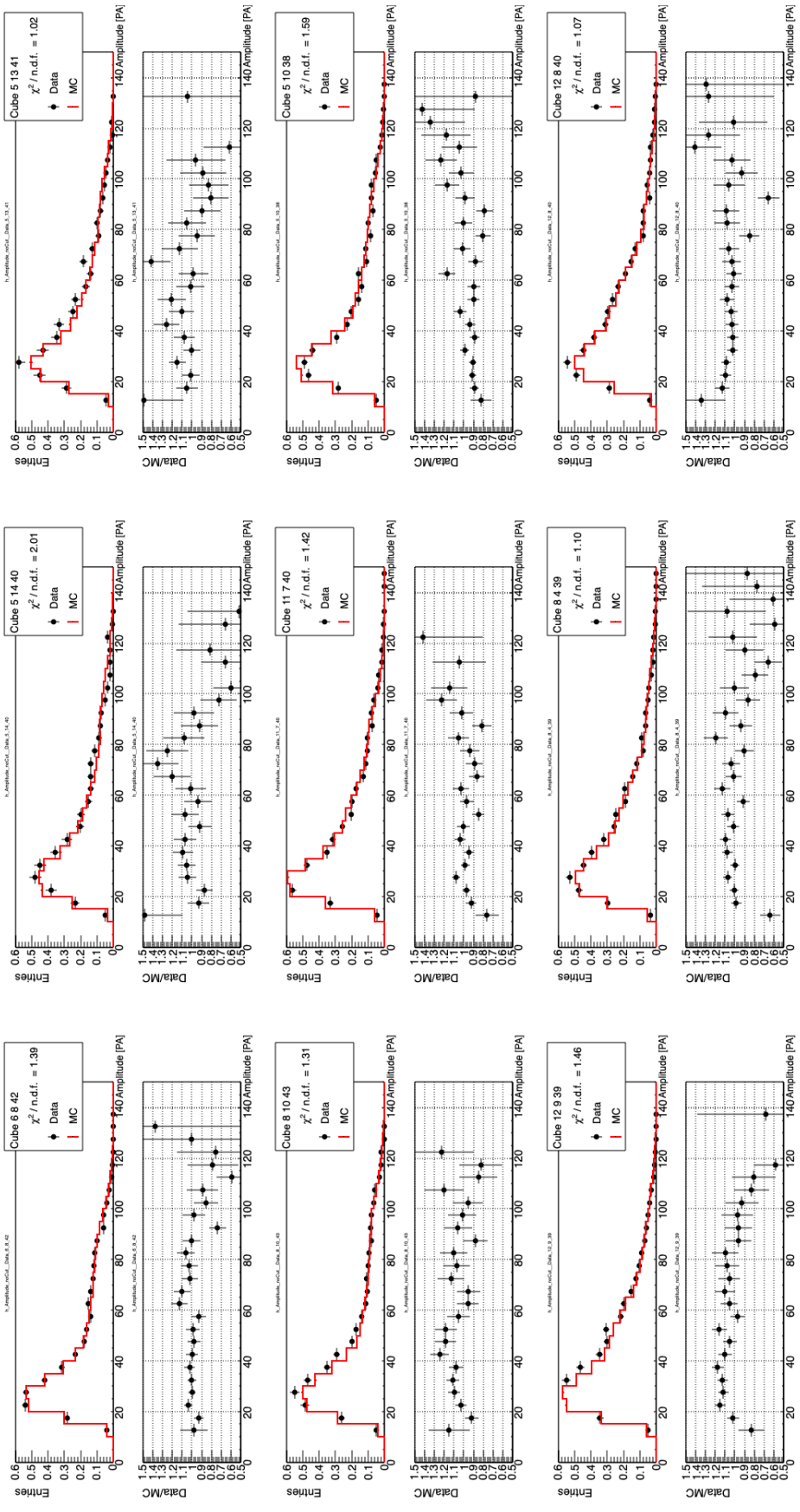


Figure 4.23.: Data and MC comparison for a random cube selection with between 1.000 and 5.000 events in the data histogram and a bad reduced chi-square value.

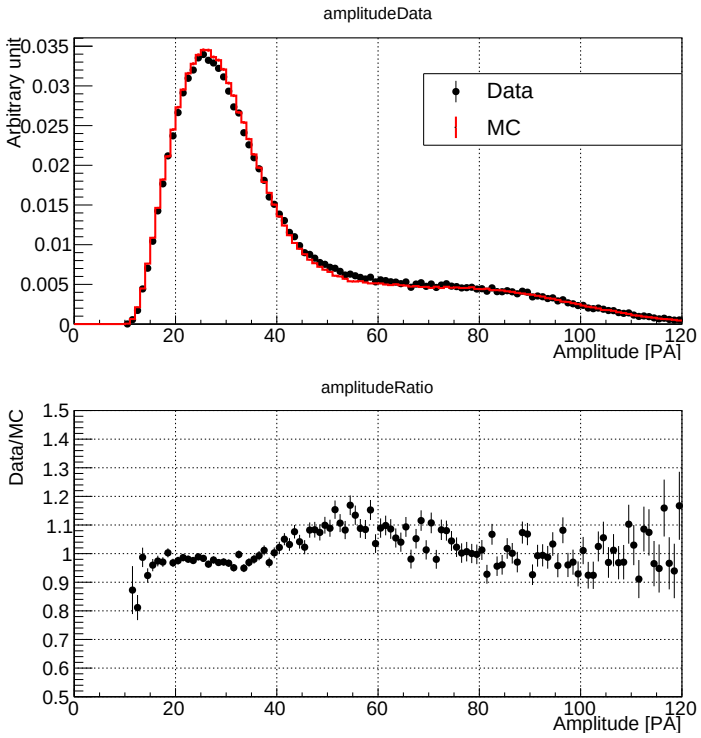
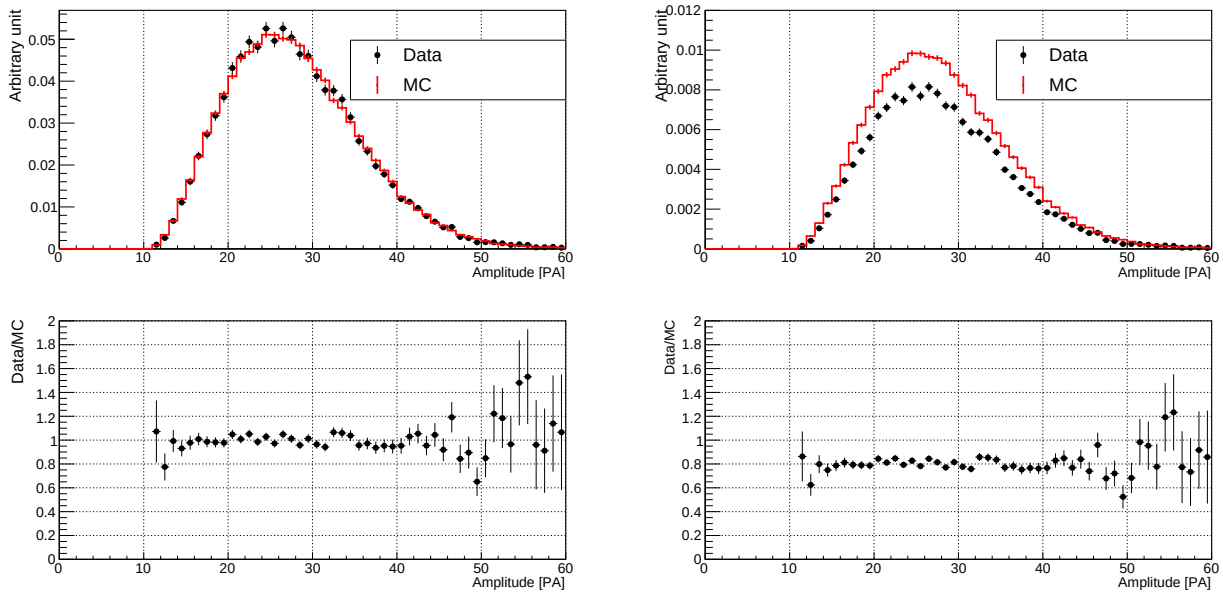


Figure 4.24.: Comparison between data and simulation of the cube amplitude asking to the cube to isolated in its plane. This is a subset of figure 4.16a events.

But histograms can also be normalised to the number of tagged events. The figure 4.25b shows this comparison. A 17% discrepancy on the efficiency is revealed this way.



(a) Normalisation to the unity.

(b) Normalisation to tagged events.

Figure 4.25.: Comparison between data and simulation of the cube amplitude with the annihilation gamma selection.

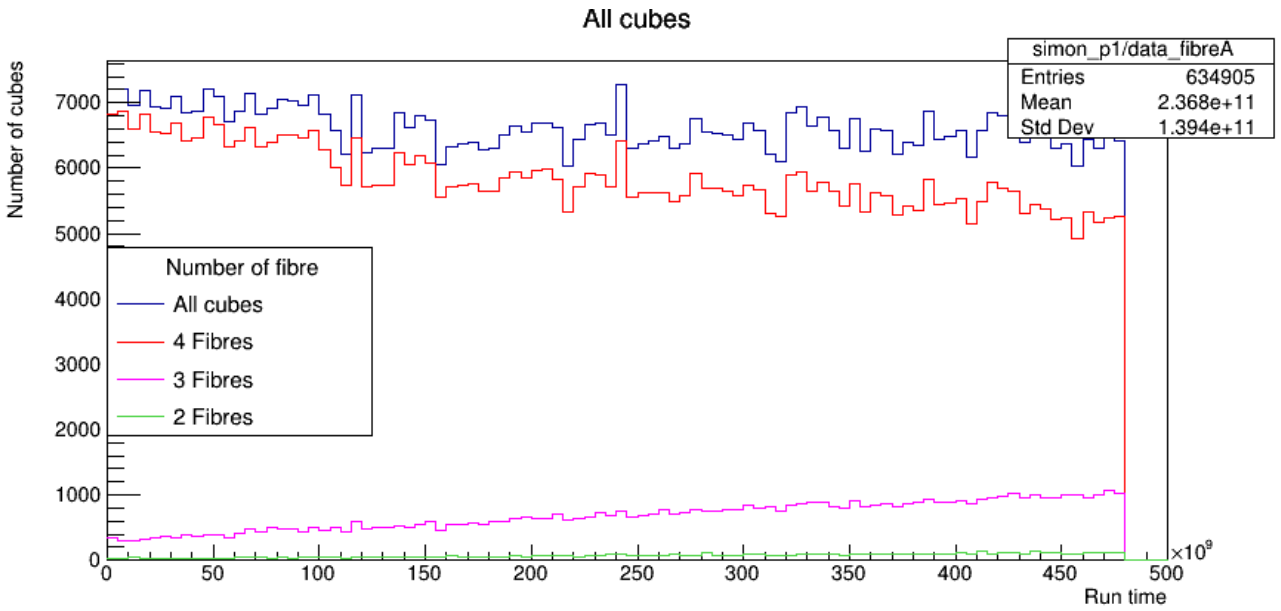


Figure 4.26.: Number of reconstructed cube with 2,3 or 4 fibres as a function of the run time in physics data.

This 17% deficit in the annihilation gamma reconstruction efficiency in data is difficult to explain based on the readout simulation tuning of the fibre and cube parameters (at least those we took care of) since the shape of all the energy distributions shown in this subsection and the previous one agree so well. On the other hand, we encountered along this work many examples of DAQ instabilities: We have shown high deadtimes ; we observed also that a fraction of fibres become irresponsive after a few minutes in physics data (see figure4.26) ; regular cases of desynchronisation between planes are observed by our colleagues. Thus, the problem could be due to yet another problem of this point, for instance issues in the synchronisation between modules.

4.3.2.5. Source based data/MC comparison: Conclusion

Detailed studies concerning the ability of the simulation to reproduce the reconstructed energy spectra of gammas emitted by the ^{22}Na calibration source have been carried out. After having corrected a number of bugs in the readout simulation code, identified and overcome deadtime issues that originally prevented a reliable comparison, and after having implemented in the simulation the best measured values of the fibres attenuation lengths and couplings to SiPM, and tuned in every cube the yield of scintillation photon reaching fibres, we find that data and MC spectra agree within a few percents across their whole energy range. A minority of cubes exhibit larger discrepancies, without changing drastically the quality of the overall detector response. A similar level of agreement is also observed at the fibre level. We measured the efficiency with which small deposits (typical of annihilation gamma's) are reconstructed. A 17% deficit in data is not understood yet but could be attributed to a DAQ issue, since several times since the start of the experiment DAQ instabilities issues were reported, in particular in the demanding calibration data. In section 4.3.3, we will present data/MC comparison based on BiPo background events.

4.3.3. Characterisation of simulation response with BiPo background

To complement the studies presented in section 4.3.2, we carried out other studies, based on BiPo background events collected during regular physics runs and which features are close to that of IBD signal events. The goal is to confirm that the agreement observed with sources is good enough for the first physics results. The decay happens in the ZnS scintillator with the emission of an electron via beta decay followed by an alpha decay (see section 2.3.3 for details). Moreover the beta decay could be followed by radiative emissions giving a signal close to IBDs because the emitted gamma rays following the ^{214}Bi decay could mimic annihilation gammas emitted in IBDs. Also it allows to tune the response of the ZnS scintillator, as the beta decay happens in the ZnS scintillator, in the simulation. A part of the unblinded reactor-on and reactor-off dataset have been used to perform this study:

- Reactor-off period: 11/07/2018 - 25/07/2018
- Reactor-on period: 09/06/2018 - 25/06/2018

The two methods presented in chapter 5 to subtract the background can be used to isolate the BiPo component in this sample. The first one is using a side band selection with a high BiPo purity. Indeed thanks to the BiPonisher variable and the high time constant of the decay, a selection combining the two variables is able to remove a large amount of other correlated components giving a sample of BiPo and accidental events. Using the variables already introduced in section 2.4, the BiPo selection is the following:

- BiPonisher $\in [0, 1.35]$
- $E_{prompt} \in [1, 4]$ MeV
- $\Delta Z_{ES-NS} \in [-3, 3]$ Cubes
- $\Delta T_{ES-NS} \in [300, 500]$ us

Only accidental background remain, which is subtracted by looking at an off-time window. This component is obtained by using the same set of cuts as the previous one expects for the ΔT_{ES-NS}:

- $\Delta T_{ES-NS} \in]-200, -100[$ us

The second approach is using the sPlot techniques [77], an introduction to this method is given in section 5.4.2. Each type of background and the signal have a different time constant than the BiPo. Indeed the signal and the cosmic induced background have a time constant between the prompt and the delayed signal close to 60 μs and the accidental background have, by definition, a flat distribution with regards to the time between the prompt and the delayed signals. Contrarily, for BiPo events the alpha decay happens 250 μ s after the beta decay. Thus a fit of the ΔT_{ES-NS} could be performed and using the sPlot techniques the BiPo distribution of all variables can be extracted, assuming that those variables are not correlated with the ΔT_{ES-NS}. The selection is the same as the previous one without the cut on ΔT_{ES-NS}.

Thus four sets of distribution could be constituted:

- **Reactor-off with sPlot:** It is the set of distributions extracted with the sPlot techniques and using Reactor Off days.
- **Reactor-on with sPlot:** It is the set of distributions extracted with the sPlot techniques and using Reactor On days.
- **Reactor-off side band:** It is the set of distributions extracted by applying the side band selection on Reactor Off days.
- **Reactor-on side band:** It is the set of distributions extracted by applying the side band selection on Reactor On days.

We studied a certain number of variables of interests, like the prompt energy estimators or variables related to the topology of IBD candidates introduced in the section 5.2.3. In the following subsections, the four sets of distributions are compared with the prediction coming from the simulation of BiPo events.

4.3.3.1. Prompt energy estimators:

Firstly, the prompt energy estimators were studied using the CCube reconstruction described in 2.4.2.3. Three different variables are under investigation to know the best one to perform a sterile neutrino analysis.

- $E_{most\ energetic\ cube}$: This is the energy of the most energetic cube in the prompt.
- $E_{prompt\ crown}$: This is the energy reconstructed in the most energetic cube plus its direct neighbours.
- E_{prompt}: This is the energy reconstructed in all the whole detector (at least the part readout following a neutron trigger).

The result is shown on figure 4.27. For all the energy estimators, data and MC agree within a few percents over most of the range. Above 2.5MeV, a discrepancy seems to appear. A part of this is due to the lack of statistics in this region. The sPlot approach, however, allows for more statistics, and seems to confirm this trend. It is a sign that parameters such as the light leakage between neighbouring cubes might be overestimated. In MC high energy cubes would loose too much light (which would not be compensated by light from neighbours since very little energy would be deposited in them since the total BiPo ES energy is around 3 MeV) and migrate to the left of the distribution, while lower energy cubes would receive too much light from neighbours and shift to the right. The fact that discrepancies are lower for estimators that sum more than one cube come in support of that. Other parameters, like the ZnS scintillation light yield caused by the passage of the BiPo's electron through it could also play a role. It could also be a bias in the sPlot procedure, which is known to have a limitation if more than two populations have to be separated, or if the discrimination variable (here ΔT_{ES-NS}) is not

powerful enough. We shall conclude here that the agreement seems to suffice for the first physics results (systematics related to data/MC discrepancies and discussed at the end of this chapter), and invite the colleagues in charge of the BiPo -based MC tuning to repeat this with higher statistics.

One satisfactory observation is the following one: events are selected by requiring $E_{prompt} > 1MeV$, as can be seen on figure 4.27c. The impact of this is clearly visible on figure 4.27a and 4.27b: below $E_{prompt\ crown} = 1$MeV are found events that pass the cuts because of additional energy deposits out of the crown that surround the highest energetic cube. Therefore, the simulation seems able to reproduce correctly events with a pile of multiple energy deposits, some of them low.

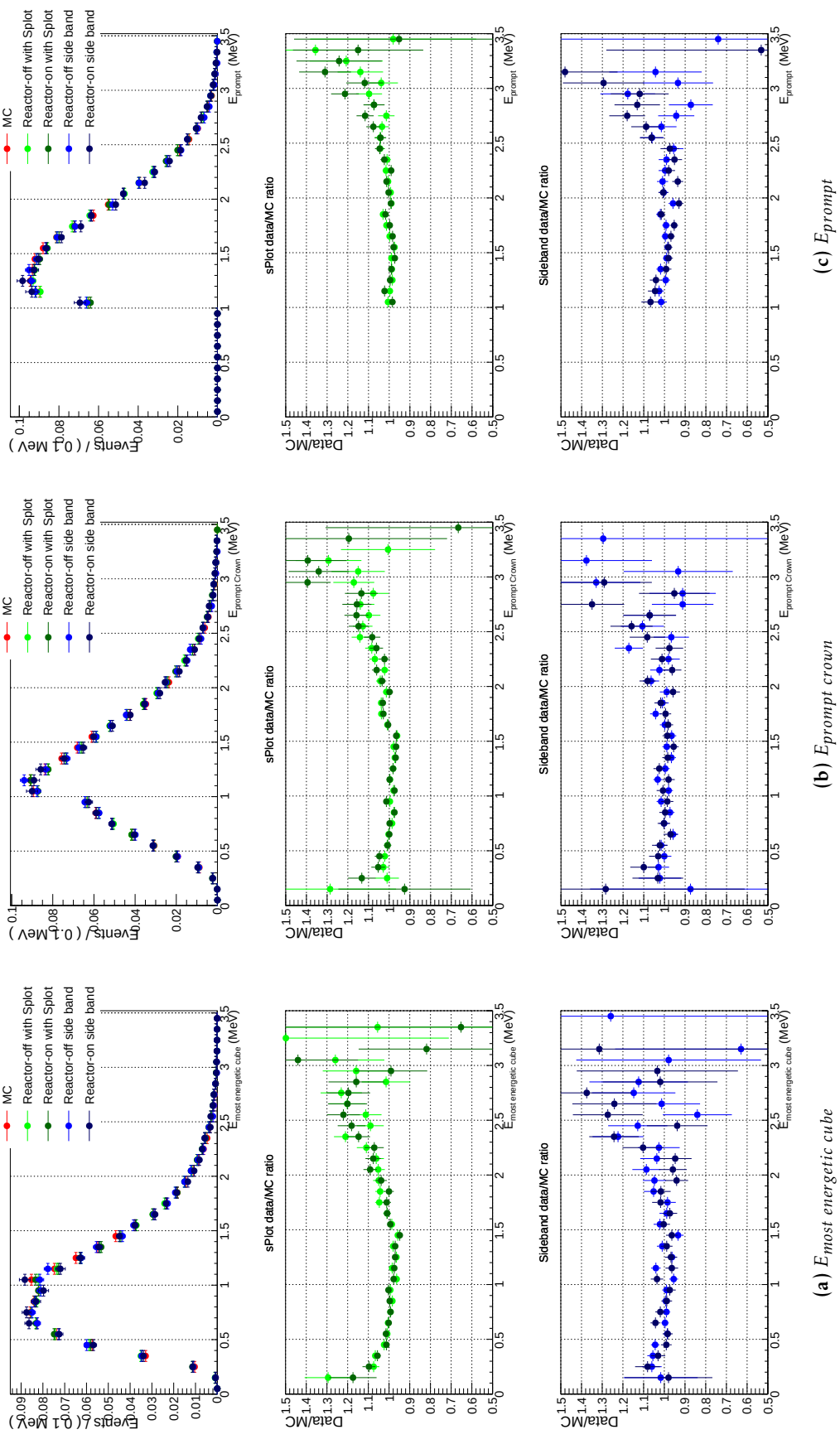
The sterile neutrino analysis is performed as a function of the baseline. For the first planned physics results, the control of the simulation response has to be done per module since this is the granularity with which different baseline will be compared in the oscillation analysis. This comparison is shown in the figure 4.29. Again a good agreement is observed between data and simulation. A remarkable agreement concerns the height and shape of distribution below 1 MeV. It fluctuates from a module to another and this is well described in the simulation. This fluctuation is mostly due to the dispersion of the visible light yield seen by each fibres of the detector. Indeed, in the reconstruction software the selection $E_{prompt} > 1MeV$ is applied before any calibration of fibres amplitudes. Thus the impact of this threshold is expected to fluctuate a bit depending on the position of the triggered channels.

Finally, the last test consists in evaluating the agreement as a function of the cube lateral position. Because the baseline also depends on the lateral position in a plane. Moreover, cubes at high X or Y had to be studied with a limited statistics in source runs, due to their distance to the source positions. Two regions have been defined:

- **Internal:** The cube is reconstructed in the centre of a plane. $X_{prompt}, Y_{prompt} \in [4,11]$
- **External:** The cube is not reconstructed in the centre of a plane. $(X_{prompt} \in [0,3] \text{or} \in [12,15]) \text{and} (Y_{prompt} \in [0,3] \text{or} \in [12,15])$

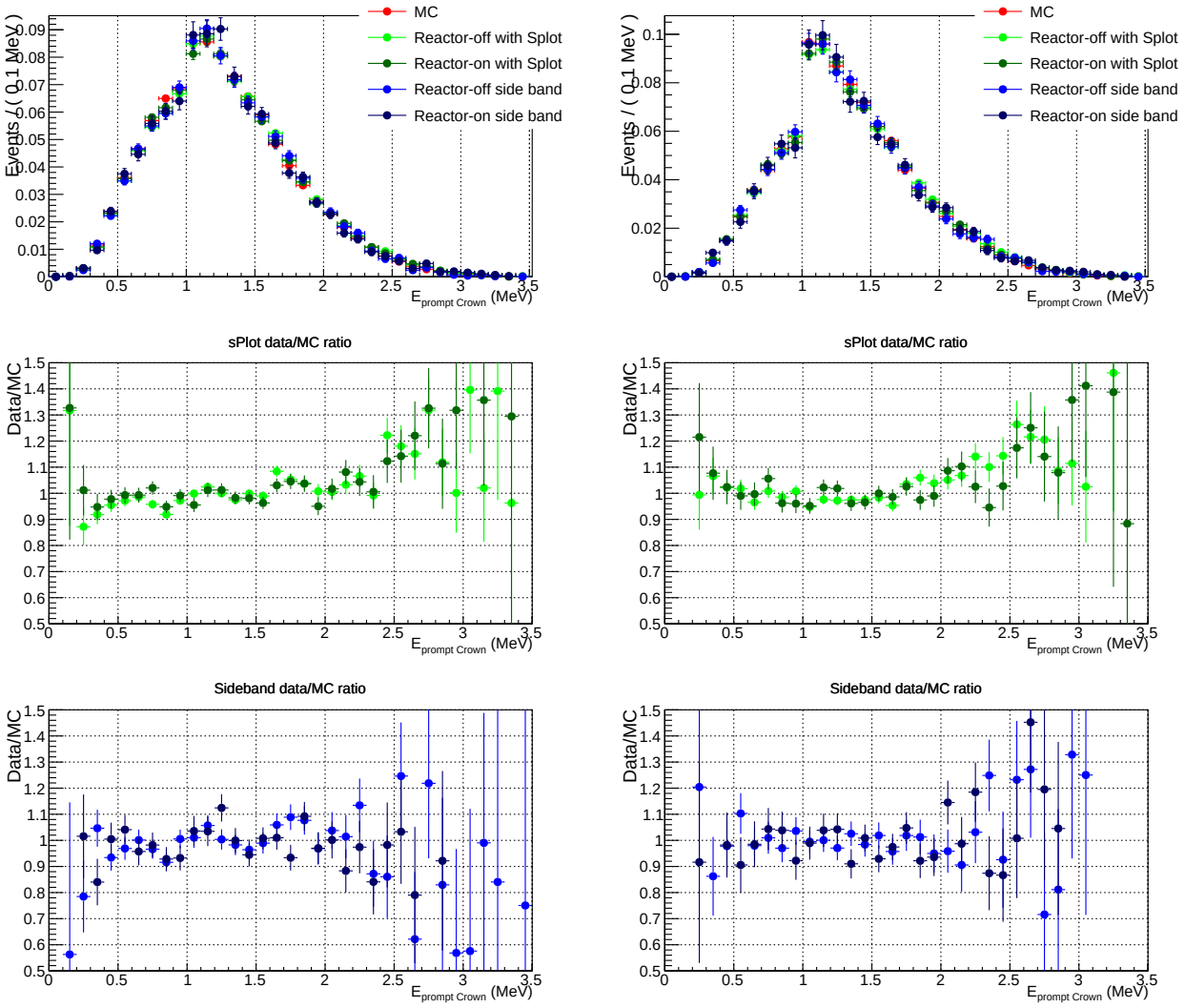
Figure 4.28 shows the result. In both cases, the agreement is at the same level as in the entire detector. Even the size of the threshold at 1 MeV is well reproduced.

This analysis shows the same results as the one from calibration data with an energy scale well reproduced in the simulation even with complex events as the BiPo. This level of agreement remains as function of the baseline. The agreement is also good for cubes close to plane edges, suggesting among other things that the attenuation is well taken into account. The collaboration considers this agreement sufficient for the first physics results. Systematics associated with that readout simulation tuning will be discussed in 4.4.



(a) $E_{most\ energetic\ cube}$ **(b)** $E_{prompt\ crown}$ **(c)** E_{prompt}

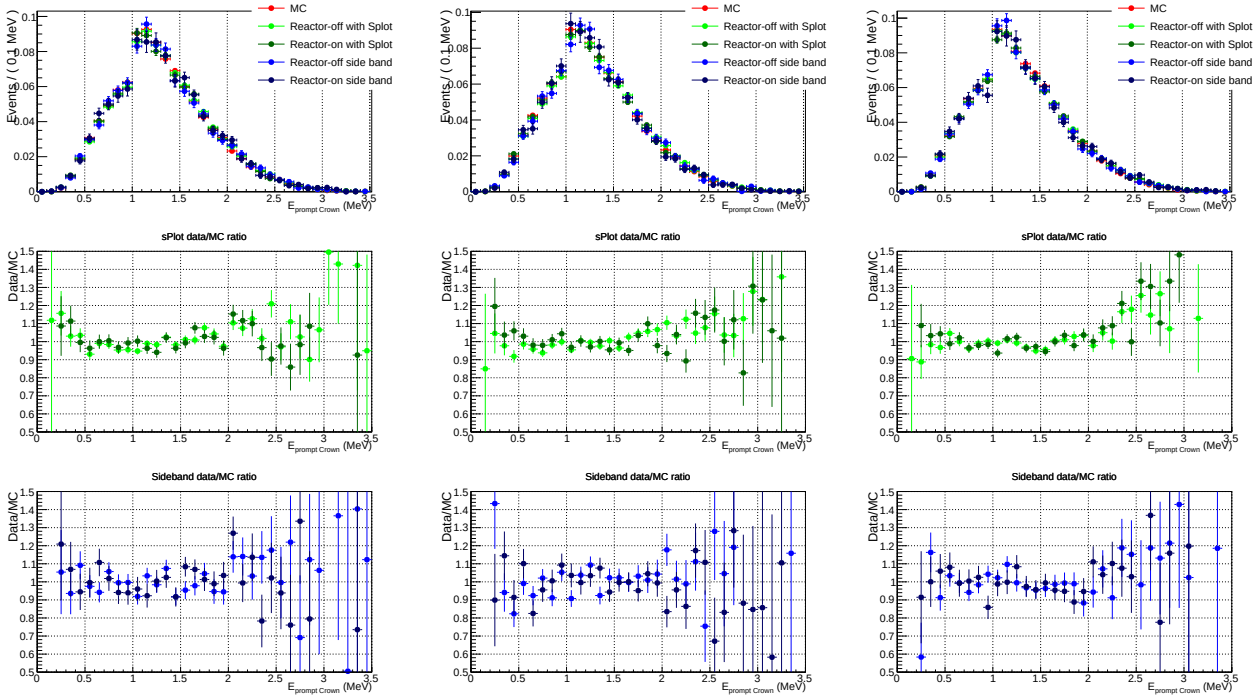
Figure 4.27.: Comparison between data and simulation of the three prompt energy estimators using the two different methods to extract the BiPo background.



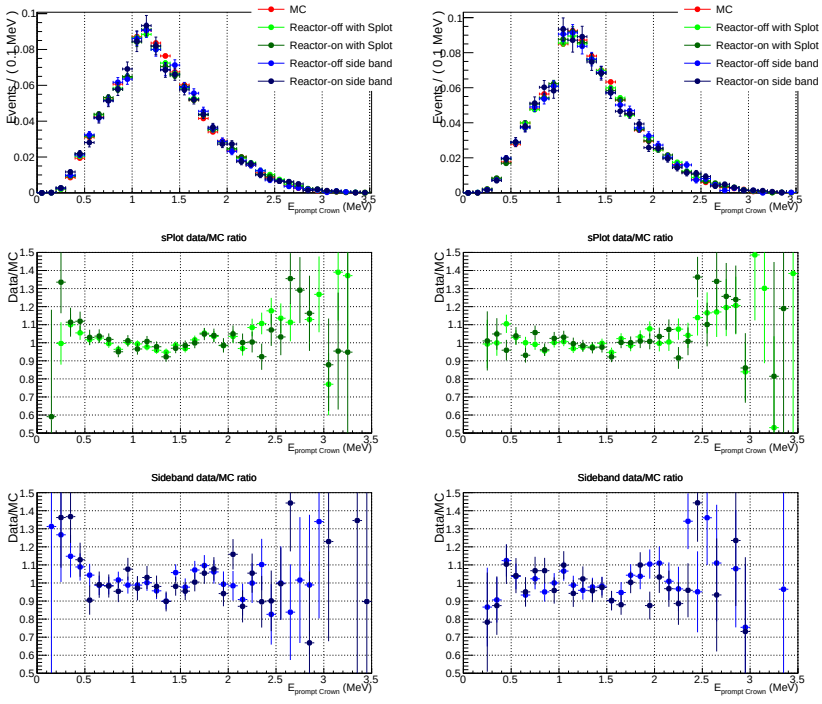
(a) Internal region.

(b) External region.

Figure 4.28.: Comparison between data and simulation of the crown energy as a function of the most energetic cube position in the plane.



(a) Module 1 **(b)** Module 2 **(c)** Module 3



(d) Module 4 **(e)** Module 5

Figure 4.29.: Comparison between data and simulation of the crown energy for the five modules of the detector.

4.3.3.2. Annihilation gamma reconstruction:

The improvements on the signal selection are mainly based on the reconstruction of annihilation gamma. Two reconstructions of annihilation gammas have been developed. The first one is described in 5.2.2. The second one, developed in the framework of this thesis, is described in chapter 5, in section 5.2.3). This reconstruction consists in the assumption that the event is an IBD signal. We then consider low energy cubes as the interaction of annihilation gammas and try to minimise a likelihood in order to correctly associate cubes to the first or the second tracks. To have confidence in this approach it is required to have a good comparison between data and the simulation for the original variables that we developed for this reconstruction. Three categories have been defined based on the number of reconstructed annihilation gamma candidates in the event. This variable is related to the efficiency with which small energy deposits are constructed. A good data/MC agreement in the distribution of this variables would confirm that the 17% deficit observed in section 4.3.2.4 is due more to a DAQ effect than to the tuning of cubes and fibre parameters. This comparison is shown in the figure 4.30, a very good agreement is found between data and simulation.

An even better insight into data/MC agreement for low energy deposits requires to study quantities more directly related to such deposits. First, one can study again the number of reconstructed annihilation gamma candidates, but separately for the internal and external events defined at the end of the previous subsection. In the external category, it is frequent that one or two annihilation gammas exit the detector before interacting. The proportion of such events should be well described by the Geant4 simulation. Therefore, data/MC in this category depends less on the different tuning of the readout simulation. We indeed observe on figure 4.31 a good agreement. The agreement is, logically, worse in the internal category. A 5% effect appears (note that the three categories are highly correlated since the loss of a reconstructed gamma causes migration of events from an upper to a lower category). It again confirms the 17% deficit in section 4.3.2.4 could be due to the DAQ. Other variables relying on low energy deposits are shown on figure 4.32. These variables are involved in the topological analysis presented in chapter 5, section 5.2.3. For all these variables, disagreement within 5% is observed over most of the plotted spectrum. In tails, this goes up to 10%. Considering the cuts applied to these variables, this does not affect the efficiency by more than a few percents. Note that the effect of discrepancies in figure 4.32d are already included in 4.31, since these quantities (ratio of the likelihood scores of the tracking algorithm under the annihilation gamma or neutron assumption) are used to reconstruct annihilation gamma's. This quantity also gives a global vision of the quality of data/MC for low energy deposits and multi-cube events since they combine the information on all small deposits in the event.

The next step of this analysis depends on the impact of such disagreement on a sterile analysis. This the aim of the next section in which systematic uncertainties are discussed.

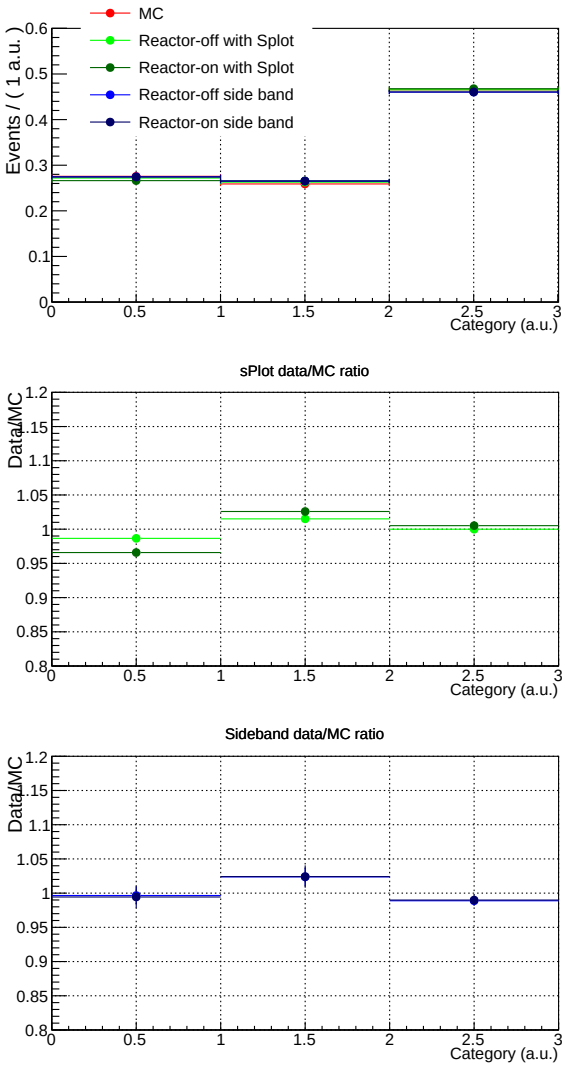
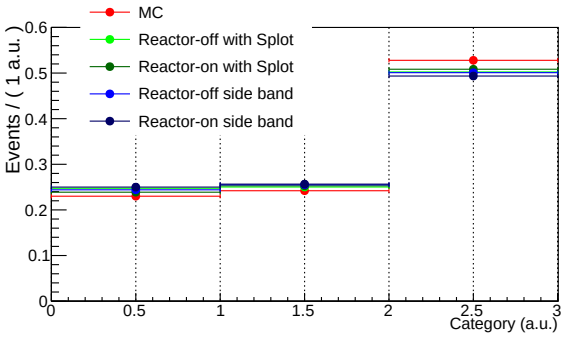
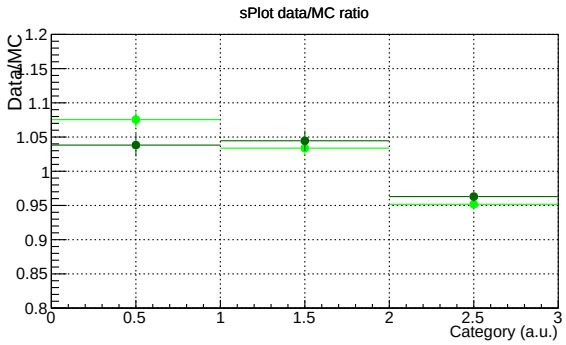
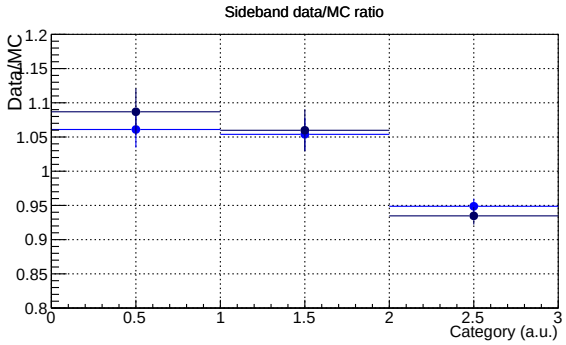


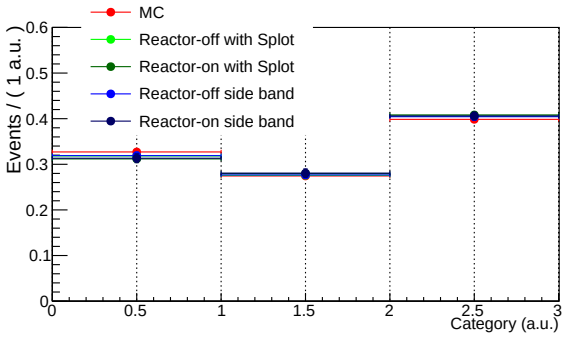
Figure 4.30.: Comparison between data and simulation using BiPo events extracted from reactor-on and reactor-off period with two techniques: number of reconstructed annihilation gammas.

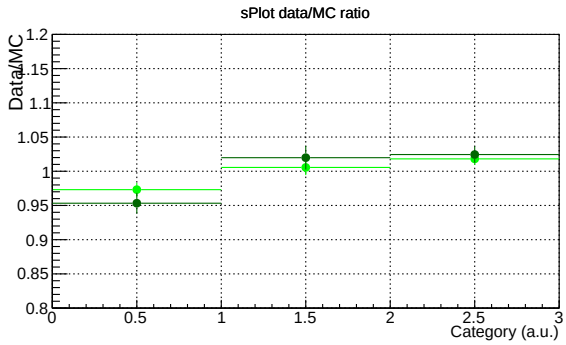


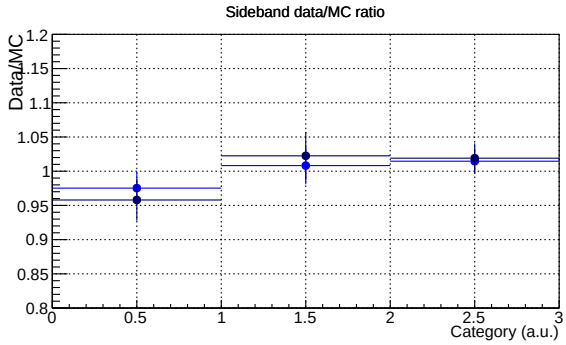




(a) Internal region.

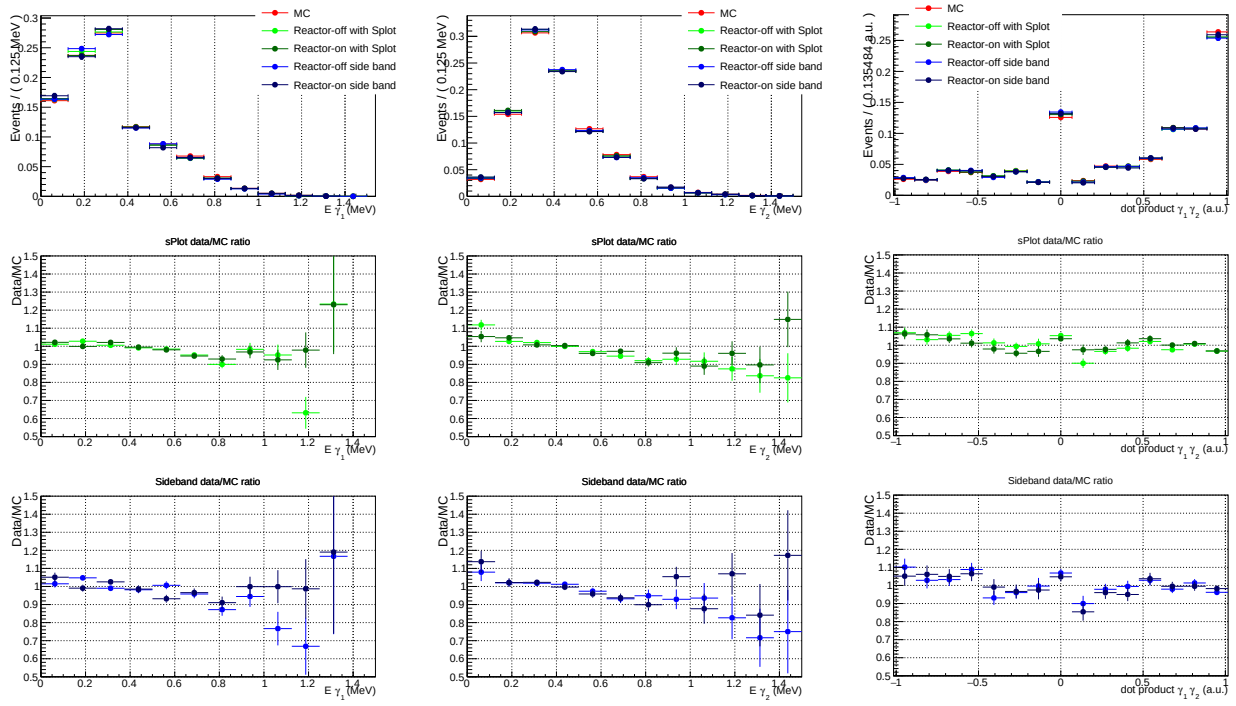




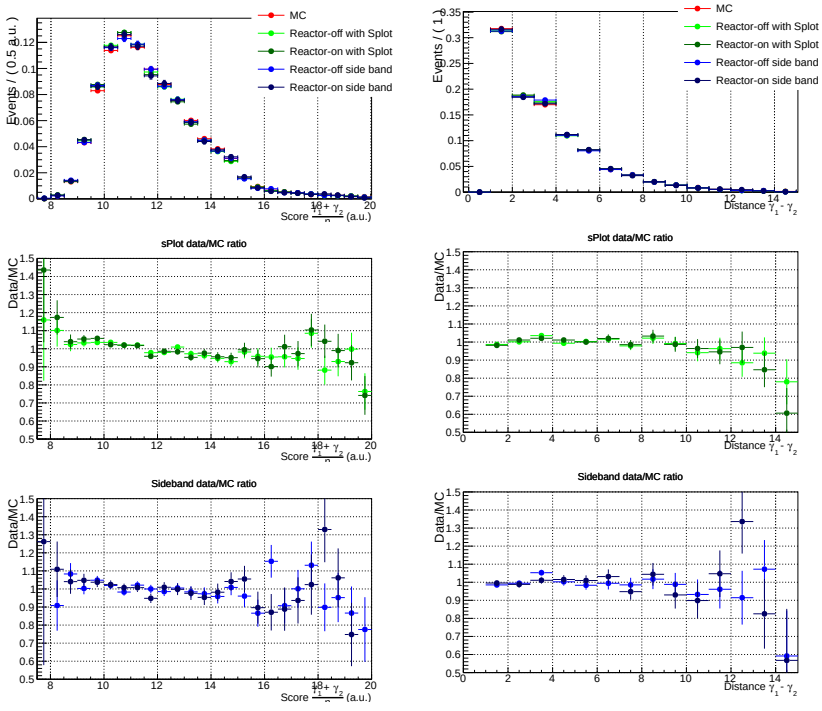


(b) External region

Figure 4.31.: Comparison between data and simulation of the categories as a function of the most energetic cube position in the plane.



(a) $E_{\gamma 1}$ **(b)** $E_{\gamma 2}$ **(c)** Dot product$_{\gamma 1,\gamma 2}$



(d) $\frac{\mathcal{L}_{\gamma 1}+\mathcal{L}_{\gamma 2}}{\mathcal{L}_n}$ **(e)** Distance between the first cube of the two gamma tracks.

Figure 4.32.: Comparison between data and simulation of the annihilation gamma reconstruction variables selecting the 2-gamma category events.

4.4. Systematic uncertainties related with the tuning of the simulation

The search of a sterile oscillation is based on the fit of a pdf to the measured (E_ν:L) distribution. The pdf is provided by the MC simulation. If the distortions of the true spectrum by reconstruction - thus including calibration and detector's response - and selection procedures are the same in MC as in data, then no bias is introduced here. If not, the significance of an oscillation could be wrong, the contours containing the values of the sterile neutrino parameters could be shifted, as could be exclusion contours set through this parameter space. Consequently the oscillation analysis needs to incorporate systematic uncertainties related to data/MC discrepancies. We discuss in this section the part of these uncertainties related to the measurement of positron, electron and gammas, in particular those due to the tuning of the parameters (light yield, attenuation lengths and couplings) we were in charge of. To put things in perspective, we will first describe briefly the main systematic uncertainties in the future first published sterile search, although they do not depend on the work carried out in the framework of this thesis.

The data/MC agreement obtained in the section 4.3.3 is typically at the 5% level or better in energy distributions, and in the distribution of other key analysis variables. We discuss below of ways to translate this a priori promising "by eye" analysis into more quantitative statements on the analysis results.

The impact of the differences between the simulation and reality can be studied in two ways. We have worked first on an "ab initio" study: the difference is partly due to the tuning of parameters in the readout simulation ; thus, we can measure the impact on the oscillation measurements by generating alternative simulated IBD samples, changing between samples the values used in the readout simulation, by amounts decided by the uncertainty with which they are measured. We will focus here on the cube and fibre parameters. At the end of the section, we will discuss briefly a second method, directly based on data/MC ratios.

4.4.1. How to judge of the impact of a systematic uncertainty

A good way to do this is to study how contours in the ($\sin\theta_{ee}$,$\Delta^2 m_{14}$) are modified when systematic effects are accounted for. In addition to the toy experiments generated to produce the contours due to statistical uncertainty, additional sets could be generated, with varying readout simulation parameters. However, the team in charge of the fit framework was not yet ready to provide that tool at the time when this thesis was written. A simpler method was therefore adopted. A team from IJClab produced alternative IBD samples and compared each time the energy distribution with that from the default simulation. An example is shown on figure 4.33 (see subsection 4.4.2). The idea is to check whether the considered systematic effect distorts the distribution more than statistical uncertainties, and is large enough to mimic or hide oscillation. As will be explained in chapter 5, the present state of the analysis, despite important improvements thanks to the topological approach we developed, is still limited in terms of backgrounds and signal efficiency. Therefore, the contours produced by the first

physics results should be dominated by a limited statistical power. In such condition, it might be worthless to evaluate the systematic uncertainties addressed by this section. To anticipate their role in an improved analysis, we assume future improvements will allow to triple the signal efficiency and reduce the background by 25%, to reach a signal over background ratio around the unity. We also assume an analysis based on six reactor cycles, in order to reach a IBD statistics competitive with that in [7], for instance.

4.4.2. Results

Random uncertainties

In principle, we should apply the method described above to the cube light yield tuned in the simulation as well as to the attenuation lengths and fibre to SiPM optical couplings. The uncertainty with which the two latter quantities were determined is presented in section 3.6. However, we think this can be avoided. Indeed, in every cube, the last step of the readout simulation tuning is to tune the cube light yield, as explained in 4.2.2.2. If the fibres serving this cube have wrong values of the attenuation length and optical coupling in the simulation, this last tuning should absorb a large part of the effect. Moreover, as can be seen on figure 3.35, the dispersion of the light yield determined in all the cubes with respect to the average of their plane is 3%. One of the contributions to this is the imperfection of the fibre parameters.

The 3% variation on figure 3.35 also encompasses the random uncertainties of the LY_{rec} fits described in 3.5, but also the natural variability in the actual scintillation yields of the cubes. Therefore, assuming a 3% random uncertainty in each cube for the cube and fibres parameters should be conservative.

IJCLab colleagues have generated alternative IBD simulated sample, drawing for each sample, in each cube independently, a random value of the light yield, chosen on a 3% wide Gaussian around the default value. The effect of this uncertainty can be seen on figure 4.33. We can see that the distortion hardly exceeds that of the statistical uncertainty that will affect each bin in energy in a data sample of this size and S/B. In the present state of SoLid's analysis, we consider it satisfactory.

Biases uncertainties

The light yield fit could also be biased in a systematic way. A first evaluation of this is provided by the comparison between the Kolmogorov and analytical approaches. A systematic 2% shift, that suggest a 1% uncertainty. Is there an additional bias, common to both methods? This is difficult to exclude a priori. For instance, the corrections due to selection efficiency or due to energy between the ^{22}Na source and the cube of interest are in both methods obtained from the simulation. To evaluate the scale of an individual bias, we produced a new version of 4.16a. In addition to the default data/MC comparison, we compared data with a MC distribution obtained by changing all cube light yield's by plus or minus 2% in the readout simulation (same shift in all cubes). One can see that data/MC is

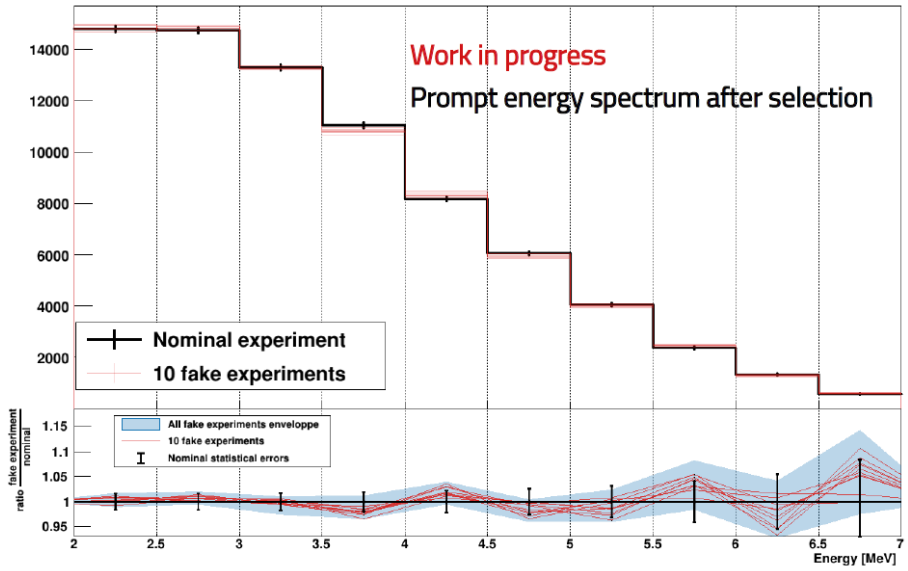


Figure 4.33.: Comparison between a nominal simulation of IBD spectrum with fake experiments for which the cube light yield has been randomly fluctuated in ± 3%.

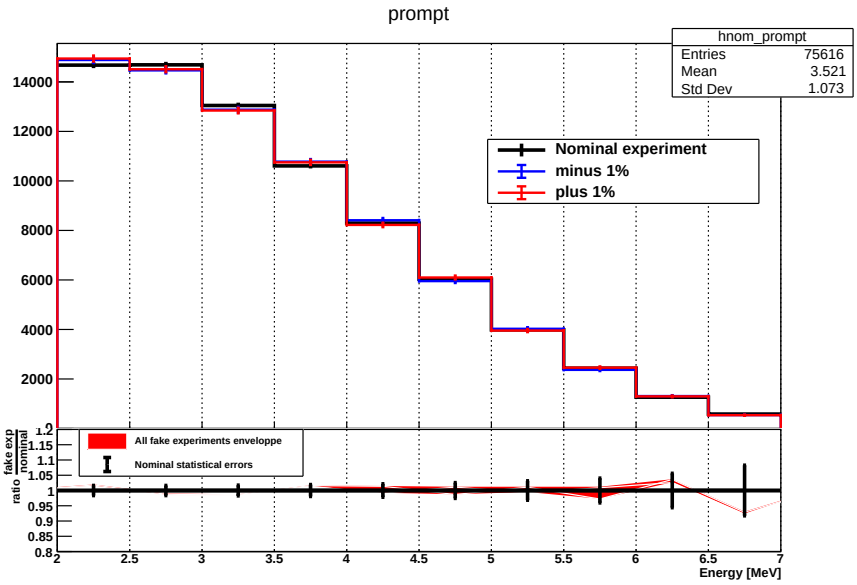


Figure 4.34.: Comparison between the reconstructed energy of IBD of the default simulation and the alternative ones where all cubes light yield have been shifted plus or minus one percent. It shows that a global bias of 1% on the light yield would falls into the statistical uncertainty.

in this case far worse than for the default, suggesting that a systematic bias on the cube light yield should not exceed 1%. To evaluate the role of this on the oscillation analysis, IJCLab colleagues have generated two alternatives samples. In each sample, the cube light yield of all cube of the detector has been shifted by plus or minus 1%. On figure 4.34, the ratio between the energy distribution of simulation with a shifted light yield is compared with the default simulation distribution. Again, no large distortion is found, compared to the expected statistical fluctuations in such a physics sample.

4.4.3. Other possible inputs from data/MC comparison to the systematic uncertainties

Low energy deposits reconstruction efficiency

In section 4.3.2.4, we found a 17% deficit in the data efficiency to reconstruct deposits below 400 keV. This was measured in a ^{22}Na calibration run, and we suspect that DAQ effects might be responsible, rather than reconstruction effects proper. Moreover, we found no sign of such a strong deficit in BiPo data (see 4.3.3). Still, to evaluate the impact of a possible discrepancy there, the people working on systematic uncertainties for the first sterile search will generate alternative IBD simulated sample: in each sample, reconstructed cubes where an energy deposit below 400 keV occurred are randomly chosen to be kept or killed. The killing probability is 9% (a conservative trade off between the deficit in ^{22}Na data and in BiPo data) for cubes that are alone in their row and column. If another cube on the same row or column receives energy and the sum along this row and column is above 800 keV, then no killing is done.

Remaining discrepancies in distribution

The data/MC ratios are flat with 3% (^{22}Na) and 5% (BiPo) in most of the range of distributions shown in sections 4.3.2 and 4.3.3. The precise origin of these remaining discrepancies is difficult to identify, so no ab-initio approach is possible to evaluate a potential effect on the oscillation analysis. Another approach is out of the scope of this thesis, by lack of time. We suggested to use data/MC ratios of the important variables (starting with E_ν, L_ν, selection variables, etc...) to derive weights to reduce or enhance the reconstruction probability of cube or events. This can be done separately in each module or planes. BiPo events are a promising source of information for that. Indeed, most BiPo decays involve an electron in the main cube and a gamma in the neighbouring cubes. It is therefore possible to isolate categories of events comparable to the IBD topology. In addition to weight, the impact of cuts can also be studied by comparing how the (E_ν,L_ν) distribution is modified in data and MC when a looser selection is tightened into the default one.

4.5. Conclusion

The quality of the sterile neutrino search depends heavily on the ability of the simulation to predict reliably the (E_ν,L_ν) reconstructed distribution. At the start of this work, the simulation could reproduce only loosely the energy distributions observed in calibration runs. A large quantity of work was provided in the framework of this thesis to eventually reach a satisfactory agreement: data and MC distributions are consistent within 5% in the greater part of the energy range. This is confirmed by studies using BiPo events, that resemble IBD. Besides energy distributions, variables used in the IBD selection were also studied and showed a satisfactory agreement. The efficiency with which SoLid reconstructs low energy deposits, typical of annihilation gammas, were studied in source samples. This was found in deficit of 17% in real data with respect to MC. A plausible explanation is DAQ related

problems in data. A consistent observation is that low E deposits in BiPo events seem reconstructed with the same performance in simulation and data.

The level at which we determine the energy scale at cube level is estimated to be below 3%. The impact on the E_ν distribution was studied. We conclude it should be acceptable in view of the statistical precision the experiment can presently reach. Ways to address the systematic uncertainties related to the reconstruction of the positron in the first sterile oscillation searh - hopefully ready by summer 2021 - have been proposed.

At the start of this work, the readout simulation was far from the state where it should be to be used for a real physics analysis. We bridged the necessary gap thanks to a combination of improvements in the code, in the calibration trigger, in the understanding of DAQ effects, and in the determination of the intrinsic light yields, attenuation lengths and fibre-to-SiPM optical couplings to be used in the simulation.

Chapter 5.

IBD signal selection and extraction

Joachim Horsley. *Via Havana*, La Café, 2019

5.1. Introduction and overview

When we started this work at the end of 2018, the official selection of the collaboration was close to that presented in the section 2.22. The performance of the selection was not sufficient to perform an oscillation analysis. But this selection does not use the two annihilation gammas emitted in an IBD process. We decided to design a selection that does, by fully exploiting the high segmentation of the detector. We defined new variables, based on the corresponding low energy deposits. Reconstructed this way, the topology of a positron event and the energy of the annihilation gamma candidates differs in the case of an IBD event and a background event.

Such an analysis is a priori optimal if we lower the fibre analysis thresholds, for a better efficiency on these energy deposits. Nevertheless, due to the two-dimensional readout, we were initally afraid to not understand well low amplitude signals, therefore the impact of fake cubes (described in section 2.4.2.3) on the annihilation gamma reconstruction. Indeed, the readout simulation results were not trustable yet. We therefore decided to perform first a smeared Geant4 analysis with a simpler reconstruction. This allowed us, to test several ways to reconstruct annihilation gammas and gave hints on the feasibility and improvements of this topological approach could bring.

We presented in September 2019 a first version of our topological analysis using a lower fibre analysis threshold and showed better performance than the official selection. However, the collaboration decided to not decrease this threshold and to stick to the official selection. The data/MC agreement was not yet considered sufficient even for this threshold, and we focused on its improvement. Despite this decision, in February 2020, colleagues from Imperial College presented a selection based on the reconstruction of annihilation gamma energy deposits drawing the same conclusion as us few months ago: it is more powerful against backgrounds, but to reach a sufficient discrimination, a lower fibre analysis threshold is necessary. The collaboration strategy therefore shifted toward this kind of selection, aiming to show first results at ICHEP-2020.

This new strategy implied to confirm a good data/MC agreement could be reached, in only a few months. However, in parallel, a part of the collaboration was working on the preparation of the detector upgrade planned for the summer. Therefore a huge effort was necessary, relying on few people, to control not only the data/MC agreement with the low fibre analysis threshold but also optimise the selection and the subtraction, derive the systematic uncertainties and perform the oscillation to release an exclusion contour with a reduced manpower.

It was necessary to organise this effort efficiently. For instance, Imperial presented in spring 2020 another version of their analysis, suggesting a 2.76 reduction in background rate when going to low thresholds. Unfortunately, we proved this result was wrong. This does not question the validity of the topological approach with low thresholds, but a precious time was lost to understand this issue. We realised that the priority of the collaboration for summer conferences should be to provide a *robust* proof of the possibility to reach reasonable background levels (the full oscillation analysis was postponed). To ensure this, an important part of the work has therefore consisted on comparing carefully the Imperial selection and ours, in order to have confidence in the results. We lead that comparison, and managed to prove the consistency between the two analyses. In parallel to this, we optimised our selection, finalised the work presented in the chapter 4 to ensure that the data/MC was good enough for low thresholds. Finally, SoLid's new selections were presented at the ICHEP-2020 conference.

In this chapter, we present the two analyses. In the both, the event is considered as an IBD event and the goal is to reconstruct the two hypothetical annihilation gammas and derive from it new discrimination variables. The approach used by Imperial's analysis is based on a work performed by the team from Clermont. It is described in section 5.2.2. They have used a *spatial approach* that define 2 hemispheres, each one attributed the first or the second annihilation gamma. In our approach, a fuller tracking of these gammas is performed, using more information. It is based on the minimisation of a likelihood which takes into account of the probability for an annihilation gamma to travel a given distance and, to release a given energy in certain cubes, and to be deflected there with a certain angle. We called it the *likelihood approach* and it is presented in 5.2.3.

We will first present the definition of the topological variables. Then, we present in the section 5.3 the optimisation of the selection, using a rectangular selection and multivariate tools, that have also been performed differently. Imperial's analysis is based on a Uniform Boosted Decision Tree (uBDT) whereas we decided to use a Neural Network (NN). We also have defined a rectangular selection for comparison purposes with multivariate approaches and with the official selection as it was in the fall 2019. The performance of each selection is derived using IBD simulation and reactor-off dataset.

Finally, section 5.4 presents the approach used by the collaboration to extract the signal in reactor-on data. We also have contributed to the improvement of the main method described in section 5.4.1 and developed another one (see 5.4.2) for cross-check purpose. In the end, a comparison between the extracted IBD excess and prediction for the two analyses is performed. It confirms that the two approaches are close in terms of performance giving us confidence on the reliability of the analyses and their results.

5.2. Topological reconstruction

5.2.1. Introduction

Most of the short baseline detectors are using liquid scintillator as target. This technology constraints the target's design, in particular it is difficult to obtain a segmented one. Thus the interaction of a positron looks like a single energy deposit without the possibility to distinguish the energy from the positron and from the two annihilation gammas. Thanks to the usage of plastic scintillator, the SoLid collaboration manages to obtain a segmented target with a spatial resolution of 5 cm. In the PVT, the mean free path of annihilation gamma is the order of 10 cm. This allows them to escape from the cube in which the positron has been annihilated. In SoLid the signature of a prompt IBD event is constituted of one energetic cube, larger than 1 MeV, and several low energetic cubes. So in this manuscript, the prompt's topology refers to the spatial and energetic distribution of those low energetic cubes with regard to the most energetic cube in the event.

The two reconstructions presented in this section are assuming that the event is an IBD. Then the goal is to correctly label each cube forming the event to the first or the second annihilation gammas. The difficulty of this work relies on the probability to see at reconstruction level the signals from the two particles. In addition to the detector's geometrical acceptance, one has to consider the efficiency to reconstruct energy deposits of few hundred keV (see the annex A). The amplitude observed on a fibre is Poisson distributed and close to the minimum amplitude threshold applied to consider a signal. At the beginning of this work, the threshold was 4.5 PA which corresponds to around 200 keV. Lowering this threshold leads to deal with an increase of un-physical signal coming from, for example, dark counts. This will increase the number of reconstructed false cubes. A qualitative study has been carried out in order to compare the discrimination power of new variables between low and high threshold dataset. This is shown in the annex B. Moreover, it requests also to have a simulation able to reproduce correctly the detector response.

5.2.2. Spatial approach

5.2.2.1. Approach

This approach has been developed by the team from the LPC Clermont. In this method, the goal is to define three clusters of cubes: the positron's cluster and one cluster for the first annihilation gamma and another for the second one.

Firstly, the most energetic cube of the event is assumed to be the position in which the positron has annihilated, this cube is called Annihilation Cube (AC). The positron may lose its energy in more than one cube depending on its direction, energy and proximity to the cube's border. Thus the cubes sharing a face or an apex with the AC cube, called the *envelop*, are integrated to the positron's cluster and are not used for annihilation gamma reconstruction.

The second step is to look for the second most energetic cube in the event, noted C_1 for clarity. In the hypothesis of an IBD event, the two annihilation gammas are emitted back-to-back. Then a plane is defined as the one passing through AC and perpendicular to the straight line defined by the vector $\overrightarrow{\text{AC } C_1}$. This plane allows to split the detector into two hemispheres and the cubes are integrated to the first or the second annihilation gamma cluster according to their positions from this plane. Thus this method labels the cubes based on a purely spatial argument, for this reason it is called the "spatial approach" in the manuscript.

5.2.2.2. Topological variables

Events are categorised according to the number of reconstructed annihilation gammas. This is very useful for the selection optimisation as the number of discriminative variables depends on this number.

0 Gamma topology

If there is no gamma reconstructed, meaning that no cube has been reconstructed outside of the AC's envelope, there is no information to exploit. This category is expected to have less discrimination power. Indeed it is equivalent to the reference selection presented in 2.22.

1 Gamma topology

If there is at least one cube outside the AC envelope, but all reconstructed in the same hemisphere then there is only one gamma annihilation reconstructed. In this case there are informations not previously used and new variables are defined:

- $E_{\gamma 1}$: Sum of cube's energy in the cluster.
- $N_{\gamma 1}$: Cube's multiplicity in the cluster.
- $d_{\text{AC}-\gamma 1}$: Distance between the AC and the barycentre of the cluster.

2 Gamma topology

In this case there are reconstructed cubes on each side of the plane and two annihilation gamma clusters can be defined. The same variables as previously can be defined for the second annihilation gamma and one additional variable is defined as the dot product between the two barycentres.

- $E_{\gamma 2}$: Sum of cube's energy in the cluster.
- $N_{\gamma 2}$: Cube's multiplicity in the cluster.
- $d_{\text{AC}-\gamma 2}$: Distance between the AC cube and the barycentre of the cluster.

- dot product$_{\gamma 1-\gamma 2}$: Dot product between the two cluster's barycentres taking the AC cube as reference.

5.2.3. Likelihood approach

In the previous approach the back-to-back topology of annihilation gammas is used as an assumption to split the detector into two parts. The approach developed during this thesis and described in the following sections is trying to solve a tracking problem. It is a well-known problem in particle physics, the goal is to associate hits in tracks and if possible time ordered. A hit is the interaction of a particle in the detector giving a signal; generally useful information is reconstructed for each hit such as: position, time and energy deposits. Then a track is a time-ordered set of hits which described the particle trajectory. For example, figure 2.14 shows the interaction of muons crossing the detector. Thanks to the high segmentation and the straight line assumption, it is possible to perform a fit in order to reconstruct the muon's track.

The tracking of annihilation gammas is a challenge for several reasons related to physics or the detector. The Compton scattering allows the outgoing gamma to be scattered at every angle. Thus the trajectory to reconstruct is not a simple one. This is one of the inconveniences against the previous approach which assumes that annihilation gammas stayed in their hemispheres, neglecting the backscattering probability. There are also two annihilation gammas to reconstruct which complicates the identification. Moreover there are reconstruction limitations. First, the reconstruction of the hit position is difficult due to possible pile-ups or ambiguities as discussed in section 2.4.2.3. Second, the hit's energy is the order of few hundred keV, at those energies, the amplitude observed on a sensor does not exceed a few PAs. In this Poisson regime, the energy resolution is particularly large. This is problem for the tracking when it is too far from reality (the tracking model depends on energy deposits). Finally, the time resolution in SoLid is around 10 ns, far from sufficient to distinguish the times of the various annihilation gamma interactions.

In order to overcome those difficulties, the theoretical Compton scattering cross-section is scrutinised in the following section. Those relations will help to design a tracking strategy.

5.2.3.1. Compton scattering process

The process of interest is the Compton scattering. It has already been introduced in chapter 3. In this section, more details are given about this process, in particular in the case of the SoLid experiment. The goal is to have a theoretical description of annihilation gamma trajectory.

Travelled distance

n_H [cm^{-3}]	$5{,}15 \cdot 10^{22}$
n_C [cm^{-3}]	$4{,}69 \cdot 10^{22}$

Table 5.1.: Number of hydrogene and carbon atoms per cubic centimetre in the PVT scintillator, from the manufacturer.

The total cross-section associated to the Compton scattering process is given by equation 5.1, with r_e the classical electron radius equal to $2{,}8 \cdot 10^{-13}$cm, $\alpha = \frac{E_0}{m_e c^2}$ and E_0 the energy of the incoming gamma [87].

$$\sigma_{KN}(E_0) = 2\pi r_0^2 \left(\frac{1+\alpha}{\alpha^3} \left(\frac{2\alpha(1+\alpha)}{1+2\alpha} - \ln(1+2\alpha) \right) + \frac{\ln(1+2\alpha)}{2\alpha} - \frac{1+3\alpha}{(1+2\alpha)^2} \right) \tag{5.1}$$

Then using the composition of the PVT, described in table 5.1, one can compute the mean free path of the gamma as a function of the incident energy given by equation 5.2.

$$\lambda(E_0) = \frac{1}{\sigma_{KN}(E_0) \times (Z_H n_H + Z_C n_C)} \tag{5.2}$$

The atomic number appears in the equation since the process only involves electron. The probability of a gamma to interact is given by an exponential law with the mean free path as the constant. So the *p.d.f.* of the distance travelled by a gamma in finally given by equation 5.3.

$$P_{dist}(X = d; E_0) = \frac{1}{\lambda(E_0)} e^{-\frac{d}{\lambda(E_0)}} \tag{5.3}$$

To obtain the correct *p.d.f.* one should apply the function representing the detector response. Thus knowing the energy and the production or last interaction position of the incoming gamma, it is possible to have the probability for it to interact at a given distance.

Deposited energy spectrum

The deposited energy spectrum is given by the famous Klein-Nishina formula given by equation 5.4. In which T is the scattered electron energy, in other words the deposited energy. E_0 is the incoming gamma energy and α is the ratio of E_0 to the electron mass.

$$\frac{d\sigma_c}{dT}(T,E_0) = \frac{\pi r_e^2}{m_e c^2 \alpha^2}\left(2 + \left(\frac{T}{E_0 - T}\right)^2\left(\frac{1}{\alpha^2} + \frac{E_0 - T}{E_0} - \frac{2}{\alpha}\left(\frac{E_0 - T}{T}\right)\right)\right) \tag{5.4}$$

This equation gives the *p.d.f.* of the true energy distribution expected for a mono-energetic gamma undergoing one interaction in the detection volume. SoLid is made of 5 cm cubes and most of the time scattered gammas escape the cube where the interaction happens.

This distribution has to be smeared by a resolution function taking into account all detector effects. This function is usually a Gaussian with an energy-dependent resolution. The resolution law of plastic scintillator is given by: $\sigma(E) = \sigma_0 \cdot \sqrt{E}$. With σ_0 the resolution at 1 MeV. For SoLid this parameter is estimated to equal 14%.

So, in the end, the *p.d.f.* for the deposited energy in a cube is given by equation 5.4:

$$P_{edep}(X = E_{cube}; E_0) = \int_0^{E_{CE}} \frac{d\sigma_c}{dT}(T,E_0) \times \frac{1}{\sqrt{2\pi}\sigma_0\sqrt{T}} e^{-0.5\frac{(E_{cube}-T)^2}{\sigma_0^2 T}} dT \tag{5.5}$$

5.2.3.2. Gamma likelihood for IBD identification

From the two *p.d.f.*s derived in the previous section, it is possible to build a likelihood-like function defining how likely the tested cube is the most probable next cube of a track. It depends on its energy E_{test} and position $\overrightarrow{x_{test}}$. The track should provide the expected energy of the incoming gamma, noted E_{in}, and its emission position $\overrightarrow{x_{in}}$. Then from that information, the travelled distance is computed as follows:

$$d_{travel}(\overrightarrow{x_{test}}, \overrightarrow{x_{in}}) = \sqrt{\sum_{i=0}^{i=2}(x^i_{test} - x^i_{in})^2} \tag{5.6}$$

and a likelihood *score* is attributed to the tested cube by combining the distance and the energy information:

$$\mathcal{L}(C_{test}) = -\log(P_{dist}(X = d_{travel})) - \log(P_{edep}(X = E_{test}; E_{in})) \tag{5.7}$$

Assuming an incoming gamma with a given energy, the likelihood score is then a function depending on two parameters, the distance and the measured energy. An example of this function is given in

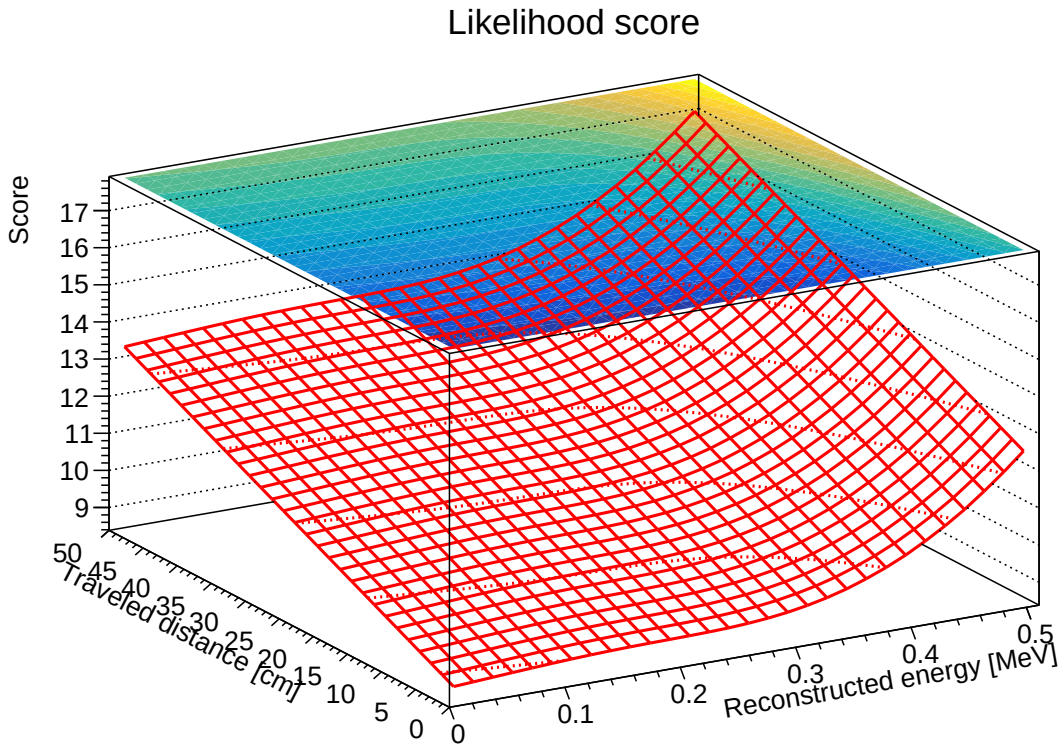


Figure 5.1.: Score of the likelihood as a function of the travelled distance and the measured energy for a 511 keV gamma.

figure 5.1. As the likelihood takes the opposite logarithm of the probability, the cube minimising this function is considered as the most probable next cube of the track.

The algorithm starts with a loop over all reconstructed cubes in the event in order to find the first cube of the first track. It assumes that the incoming gamma is an annihilation gamma with an energy of 511 keV and has been emitted in the AC cube. The cube minimising the function 5.7 is associated to the first track.

Then the algorithm loops again over remaining reconstructed cubes. For each one, the likelihood score in computed assuming if the cube is the next interaction of the first track or the first cube of a second track. Of course when a track has already cubes associated, the incoming gamma energy is recomputed such as: $E_{in} = 0.511 - \sum_{i_{cube}=0}^{i_{cube}=n} E_{i_{cube}}$, where n in the number of cubes already associated to the track and E_i the reconstructed energy of those cubes. Due to the energy resolution, this value could become negative. In order to overcome this difficulty, in this case an arbitrary value of the likelihood score is associated to this cube. The algorithm ends when all cubes have been associated to one of the two tracks.

In the end, global variables are defined for each track:

- **Cube multiplicity:** Number of cubes associated to the track.
- **Track energy:** Sum of the cube's energy in the track.

- **Track score:** Average of the cube's score forming the track. Corresponding to equation 5.7.

As in the spatial reconstruction approach, the events are categorised with respects to the number of annihilation gamma tracks that have been reconstructed.

5.2.3.3. neutron-proton elastic scattering

In parallel of the annihilation gamma likelihood, we developed a neutron likelihood. We saw in section 2.3.2 that cosmic induced backgrounds are mainly composed of fast neutrons. Furthermore the elastic scattering on hydrogen (or protons) and carbon atoms are dominating all processes. Thus the idea is to find a likelihood giving how likely the event is to be a series of neutron-proton elastic scatterings.

Deposited energy spectrum

The kinematic of the elastic neutron scattering on a proton is fully solved and has the particularities that the two particles have roughly the same mass. The energy of scattered neutron E_n^{out} is given by:

$$E_n^{out} = E_n^{in} \frac{1 + \cos\theta_{cm}}{2} \tag{5.8}$$

where E_n^{in} is the energy of the incoming neutron and θ_{cm} the diffusion angle in the centre of mass. This equation shows that the outgoing neutron energy spectrum is flat between zero and E_n^{in}. Thus defining a term of the likelihood based on the measured energy is not discriminant. Nevertheless, one interesting property is that the outgoing neutron is always scattered forward in the laboratory. This property will be used in the algorithm to remove fake track candidates.

Travelled distance

As in the case of annihilation gammas, it is possible to derive the mean free path of a neutron in the detector. But the difference is that the energy of the incoming neutron is not known and the mean free path depends on it. Thus it is required to have an estimation of the neutron energy. This is possible under the assumption that all signals observed in the detector are coming from several elastic scattering of the same neutron. The number of scintillation photons produced in the plastic scintillator by a proton recoil is not equivalent to that by an electron recoil. This quenching effect is described by the Birks law already discussed in section 4.2.1.3. We therefore use a correction based on this relation:

$$E_{true} = E_{vis}(1 + k_B \frac{E_{true}}{L_{true}}) \tag{5.9}$$

However, this requires to have an estimation of the fraction $\frac{E_{true}}{L_{true}}$ which is not accessible experimentally. Fortunately, this ratio can be derived from cosmic simulations by selecting only proton-recoil processes and expressed as a function of the visible energy as shown in figure 5.2. The distribution has been

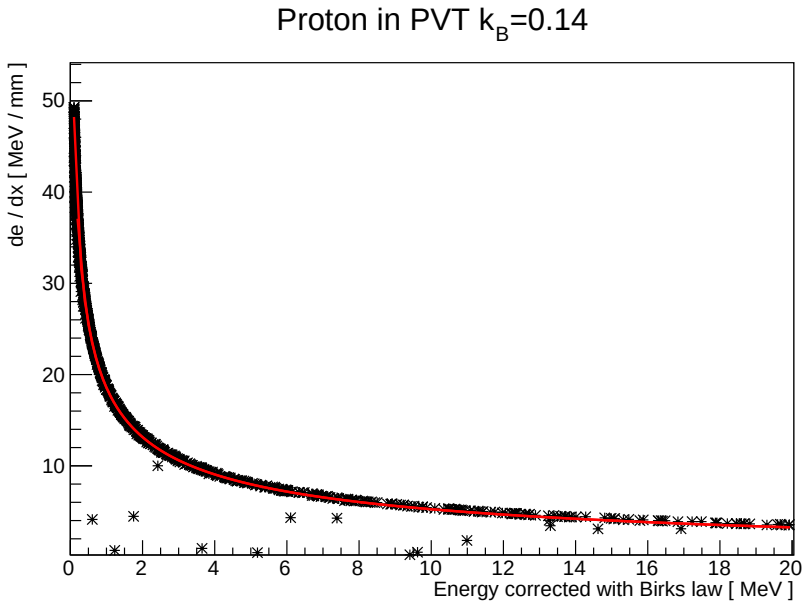


Figure 5.2.: $\frac{E_{true}}{L_{true}}$ as a function of the energy corrected by the Birks law giving the visible energy. This comes from the Geant4 simulation, by selecting only proton-recoil processes. The red line is a fit with the function: $f(x) = \frac{p_0}{x^3} + \frac{p_1}{x^2} + \frac{p_2}{x} + \frac{p_3}{\sqrt{x}} + p_4$

fitted in order to obtain a proxy of the dependence of this ratio with regards to the visible energy with the function:

$$f(x) = \frac{p_0}{x^3} + \frac{p_1}{x^2} + \frac{p_2}{x} + \frac{p_3}{\sqrt{x}} + p_4. \tag{5.10}$$

As shown on the figure, the fit describes well the simulation data. Thus an estimation of the incoming neutron energy in the detector is:

$$E_n = \sum_{i_{cube}=0}^{n_{cube}} E^i_{vis}(1 + k_B \cdot f(E^i_{vis})) \tag{5.11}$$

where n_{cube} is the number of reconstructed cubes in the event including the most energetic one. As for gammas, we now need the expression of the cross-section. We obtained it from the ENDF database [88]. The cross-section of the process of interest is given in figure 5.3.

Then the same *p.d.f.* as in equation 5.3 could be used to derive a likelihood score by taking the negative logarithm of this expression.

5.2.3.4. Neutron likelihood for cosmic induced background rejection

One difficulty of the neutron hypothesis is that its arrival direction is not known. Thus the tracking algorithm starts by creating all possible tracks with all cubes in the events. It means that there are

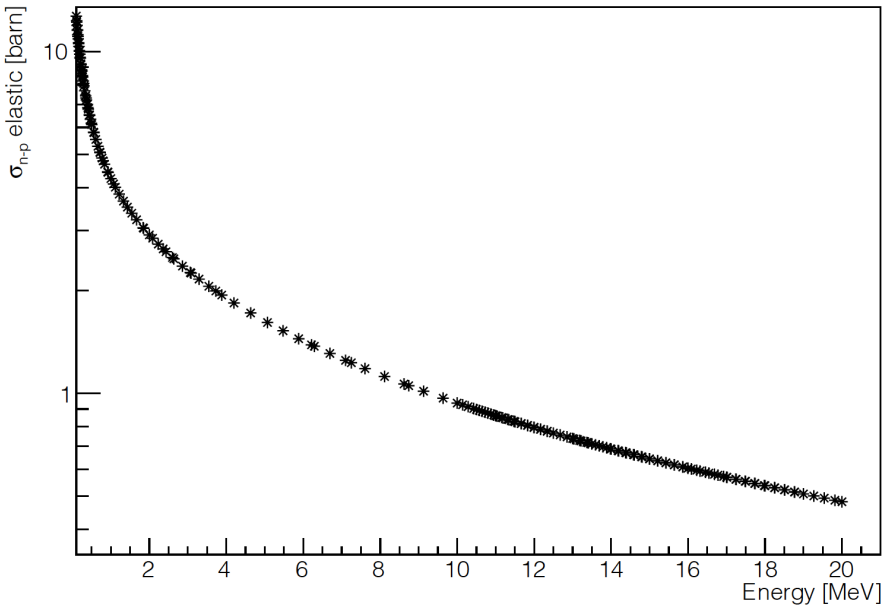


Figure 5.3.: Cross-section in barn of the neutron-proton elastic scattering as a function of the neutron energy.

n_{cube}! tracks. In order to gain in efficiency, the algorithm removed each track in which the next cube is found backward to the two previous ones. Indeed, as reminded in the previous section, the neutron is always emitted forward in the laboratory. Once all possible tracks have been realised, the algorithm computes the total likelihood score of them using the equation:

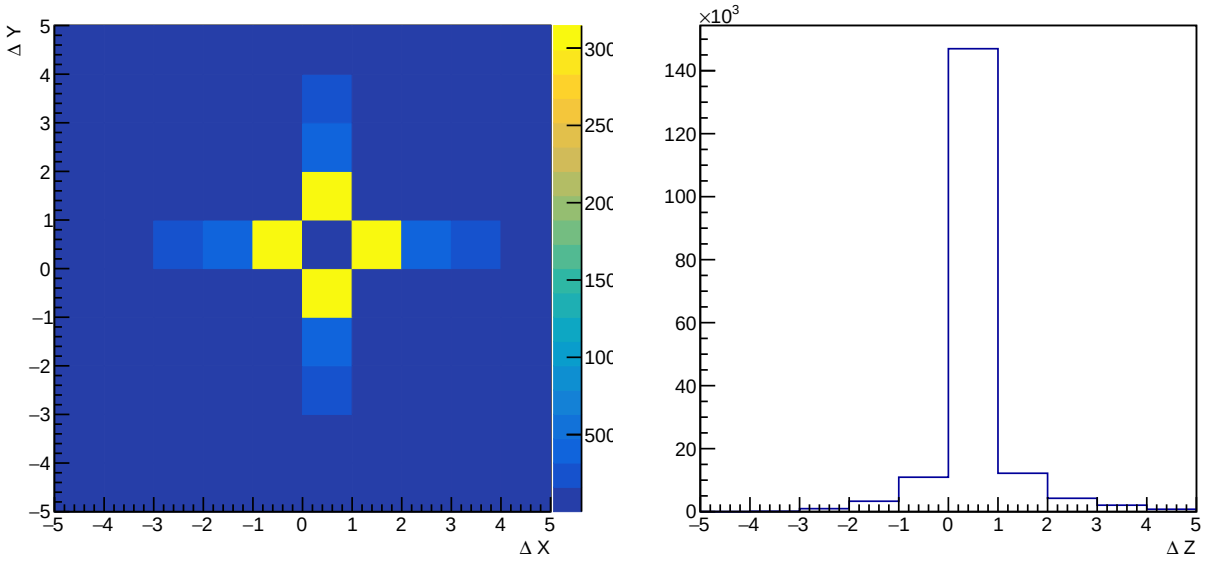
$$\mathcal{L}_n = \frac{1}{n_{cube}-1} \sum_{i_{cube}=1}^{i_{cube}=n_{cube}} -\log(P_{dist}(X = d_{travel})) \tag{5.12}$$

The sum starts with the second cube as the distance travels until the first interaction in the detector can't be known.

If there is no possible permutation of all cubes in the event respecting the forward assumption, the algorithm chooses the track with the minimum likelihood score with $n_{cube} - 1$ cubes.

5.2.3.5. Topological variables

As for the spatial reconstruction, the events are categorised according to the number of annihilation gammas reconstructed. The same variables as the previous approach are defined. In addition, the neutron and gamma likelihood scores are computed as well as the ratio $\frac{\mathcal{L}_{\gamma 1}+\mathcal{L}_{\gamma 2}}{\mathcal{L}_n}$. This ratio should be low for an event looking like an IBD event and so could be used to discriminate signal and background. Also, there is a difference in the dot product between the first and the second gamma tracks. In the spatial approach, it takes the barycentre of the two clusters to compute it. The issue is that the gamma could be scattered at all angles thus after the first interaction, the information about the emission



(a) 2D distribution in X and Y.

(b) Distribution in Z.

Figure 5.4.: Distribution of false cubes relative position with regards to the most energetic cube using IBD simulation.

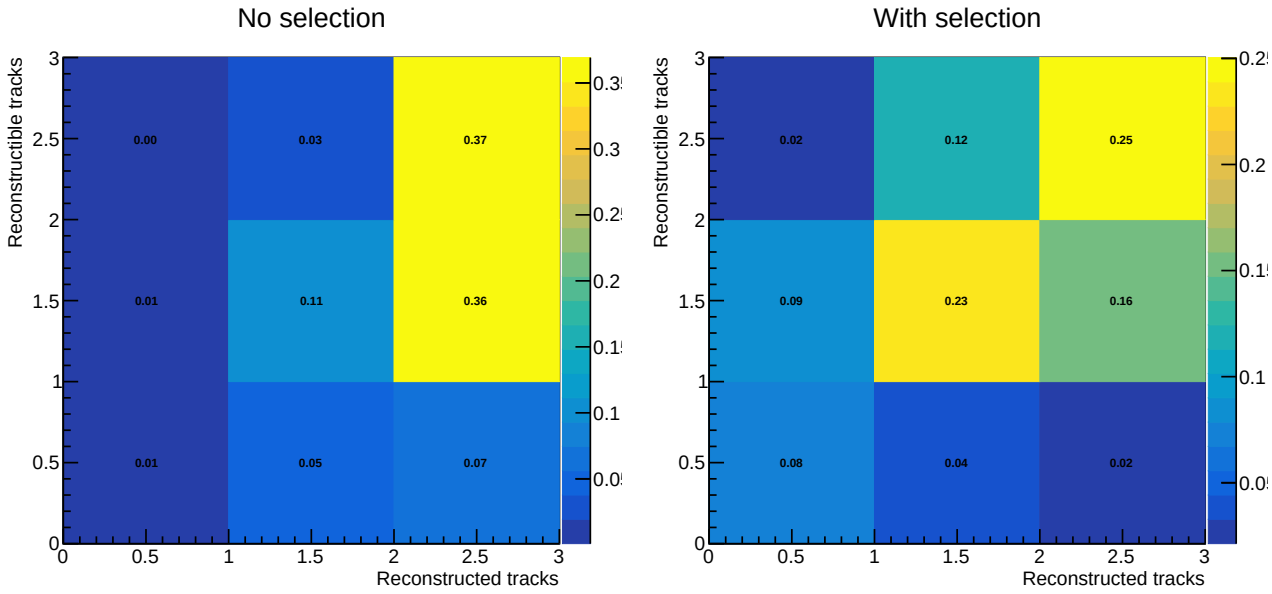
direction of this gamma is partially lost. For this reason the algorithm computes the dot product based on the first cubes added to the tracks.

5.2.3.6. Cube selection

The cube reconstruction discussed in section 2.4.2.3 is not perfect and a selection is required to improve the annihilation gamma reconstruction.

The reconstruction attributes an energy to every cube in which at least one horizontal and one vertical fibres have seen a signal. In the case of a false cube, its energy returned by the algorithm is very low. Applying a threshold on this value has to be compromise between removing false cubes and losing real ones. In particular, the Compton edge of annihilation gamma is at 341 keV. The threshold has been set at 50 keV.

Also, a study has been carried out in order to determine the selection using the simulation of IBD. The goal is to determine a selection removing false cubes. Those false cubes could be identified using the simulation by checking for each reconstructed cubes if there is a real energy deposit at the Geant4 level. The positions with regards to the most energetic cube of the event are shown in the figure 5.4. In majority, those false cubes are positioned in the same plane as the most energetic and share a face with it. Discarding these cubes removes 68% of false cubes. This selection does not impact the estimation of the positron's energy since we only forbid the algorithm to use those cubes in the tracking.



(a) Without cube's selection.

(b) With the cube selection.

Figure 5.5.: 2D distribution of the number of reconstructible tracks as a function of the number of reconstructed tracks with the likelihood approach.

5.2.3.7. Tracking performance

A short study has been carried out in order to see how well the likelihood approach performs. Using the simulation, a definition of a *reconstructible* track has been established: at least one reconstructed cube in the ES cluster corresponds to the interaction of an annihilation gamma. Thus with this definition we take into account the efficiency to reconstruct low energy deposit.

The comparison between the number of reconstructible and the number reconstructed tracks is shown in the figure 5.5. With no cube selection, the algorithm tends to overestimate the number of tracks. In particular, when there is only one reconstructible track, the algorithm is correct for only 23% of events. The cube's selection helps on this point as the algorithm now finds the correct number of tracks in 48%. The algorithm overestimates the number of track in only 33% of cases. In the case of two reconstructible gammas, the algorithm finds the correct number of track in 65% of cases. Remaining mistakes are due to false cubes not removed and events in which the positron loses energy in several cubes.

5.3. Official selections and performance

As explained in the introduction to this chapter, the collaboration developed several selections in order to understand their performance and prove their results reliable. This was particularly justified since the two main ones use multivariate techniques to obtain better performance than a rectangular

selection. Indeed, these approaches are useful to fully exploit the discriminative power of all variables, but they exploit more information, which can cause issues if this information is not understood.

The performance of a selection is given by the signal selection efficiency and the signal over background ratio. The performance is driven by the discrimination power of selection variables between signal and background events.

5.3.1. Dataset

The collaboration decided to perform a blind analysis. However, to help the design of the analysis in the absence of a fully reliable simulation - in particular concerning backgrounds - we unblinded a small part of the dataset. The detector took physics data from May 2018 to June 2020. During this time span, several couples of reactor-on to reactor-off periods were recorded. We unblinded only the first couple. This first cycle is composed of 66 days between the 9th June 2018 to the 16th August 2018. The daily evolution of the reactor power is given in the figure 5.6. The extraction is performed on a daily basis and a day is attributed to a reactor-on or a reactor-off period based on the mean power value of the considered day :

$$\begin{cases} P_{th} < 30\text{MW}_{\text{th}} = \text{Reactor-off} \\ P_{th} > 30\text{MW}_{\text{th}} = \text{Reactor-on} \end{cases} \tag{5.13}$$

An issue related to the reactor operation happened at the beginning of the reactor-on period. It made it necessary to shut down the reactor for two days. Thus the reactor-on period is split into two parts. It is also necessary to compute the detector live-time. This live-time is corrected from the dead time between runs, period during which the DAQ is initialised between two runs, but not from the channel or plane dead time (see section 2.2.4.3). Figure 5.7 shows this value for the whole analysed period. We can see that on average the detector is not online 7.5% of the time. For the analysis we need two independent reactor-off datasets: one for the background subtraction and one to optimise the selection. The statistical independence of the two datasets is primordial in order to prevent the introduction of selection bias. We use the last period of the open dataset for the optimisation. Table 5.2 summarizes which days are considered as reactor-on or reactor-off and how we will use them in the rest of the study.

5.3.2. IBD Selections

A first standard selection was introduced in chapter 2 using only the spatial and time coincidences between the prompt and delayed events. This selection gives a reference to compare the performances of the selections developed here. As shown, the number of variables available depends on the number of reconstructed annihilation gammas. Thus, events are divided in three categories and for each one of them, a specific selection is developed. As a first approach, a *rectangular based selection* (see section

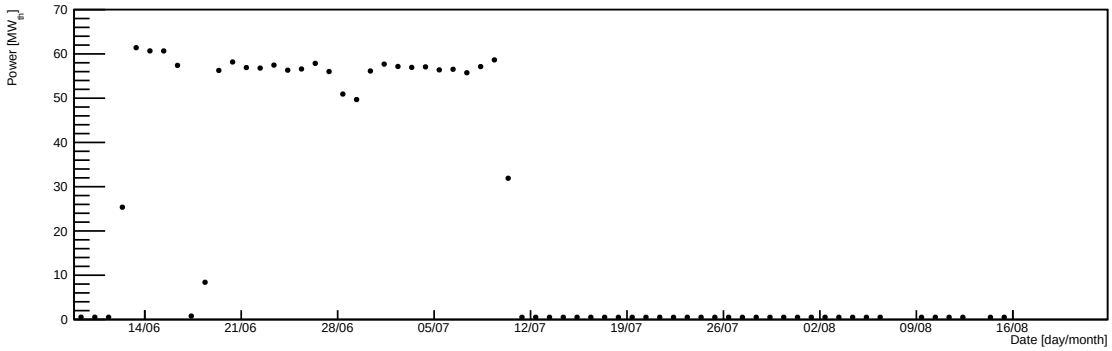


Figure 5.6.: Daily evolution of the power for the first cycle.

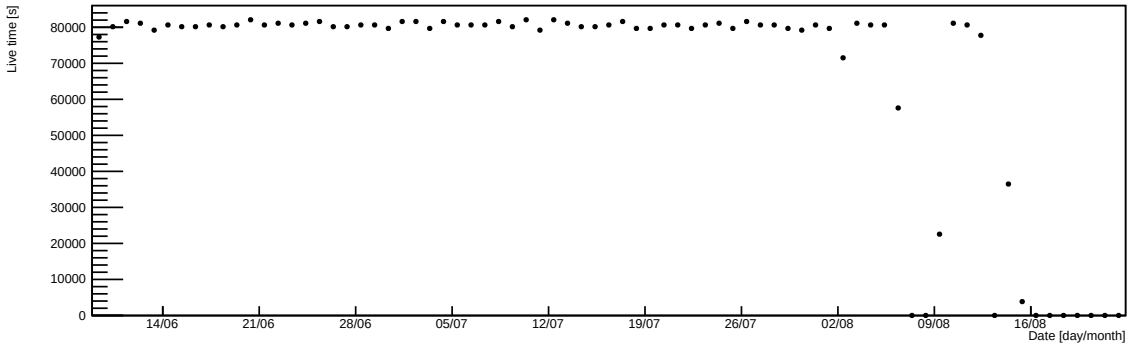


Figure 5.7.: Daily evolution of the detector live time for the first cycle.

Name	First day	Last day	Usage	Duration [days]
Reactor-off 1	9th June 2018	12th June 2018	Subtraction	4
Reactor-on 1	13th June 2018	16th June 2018	Subtraction	4
Reactor-off 2	17th June 2018	18th June 2018	Subtraction	2
Reactor-on 2	19th June 2018	10th July 2018	Subtraction	22
Reactor-off 3	11th July 2018	6th August 2018	Subtraction	27
Reactor-off 4	9th August 2018	16th August 2018	Selection	5

Table 5.2.: Summary table of the open dataset with: the name, the first and last days, the usage and the duration of each sub-dataset

5.3.2.1) has been designed, in order to perform a fair comparison with the standard selection. Then, multivariate selections are developed in order to fully exploit the correlation between all variables to increase the discrimination power. The first one, developed by Imperial, is called uBDT (for uniform boosted decision tree, see section 5.3.2.2.2). The second one, that we developed, is called ANN (for artificial neural network, see section 5.3.2.2.1). This kind of approach reaches the best performances but is very sensitive to the ability of the simulation to correctly reproduce the signal.

We use the same preselection for all selection, based on the standard selection presented in section 2.5.3. A preselection is useful when using multivariate approach because the algorithm then focus on events that are the most complicated to label as signal or background. This preselection uses non-topological variables: the reconstructed energy in the cluster, the time difference between the ES and the NS clusters, the BiPonisher and the relative positions between the ES and NS clusters. These cuts are presented by equation 5.14. We can see that some cuts have been loosened in order to let the multivariate analyses fully exploit the correlation between variables.

$$S_{Signal}^{pre-selection} = \begin{cases} E_{cluster} \in [2,7]\text{MeV} \\ \Delta T_{ES-NS} \in [1,141]\mu s \\ \text{BiPonisher} > 1.44 \\ \Delta X_{ES-NS} \in [-3,3]\text{cubes} \\ \Delta Y_{ES-NS} \in [-3,3]\text{cubes} \\ \Delta Z_{ES-NS} \in [-2,3]\text{cubes} \\ \Delta R_{ES-NS} \in]0,4]\text{cubes} \end{cases} \tag{5.14}$$

5.3.2.1. Topological selection using rectangular cuts

This first selection is performed per category, according to the number of reconstructed annihilation gammas in the event. It will allow to have an idea of the improvement possible thanks to topological variables with respect to the standard selection. It will also give a reference to know the gain of using a multivariate analysis.

The background and signal are not distributed the same with respect to the number of reconstructed gammas, as shown on figure 5.8. The category with no annihilation gamma is the one with the poorest signal over background. No improvement is possible using topology. We simply cut out events from this category. This selects 84% of the signal and 74% of the background.

In the other two categories, we start by tightening the preselection. The initial preselection does not fully exploit the relative positions of the prompt and delayed events, since those cuts have been loosened in order to let multivariate analyses to use at best their potential. Figure 5.9 shows the remaining discrimination between signal and backgrounds. We can see that the most discriminative

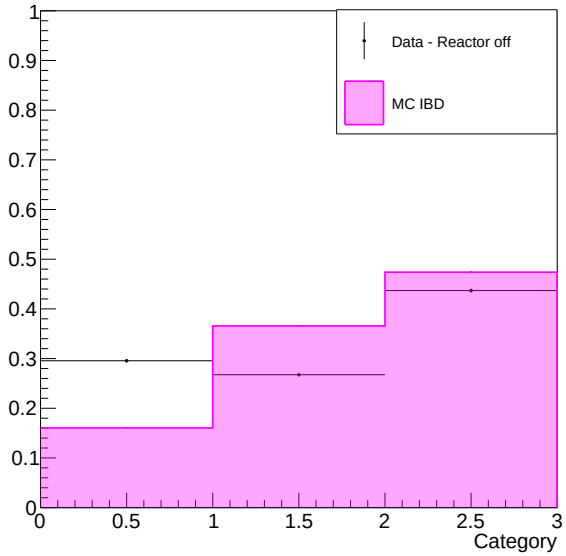


Figure 5.8.: Distribution of signal and background events as a function of the event category.

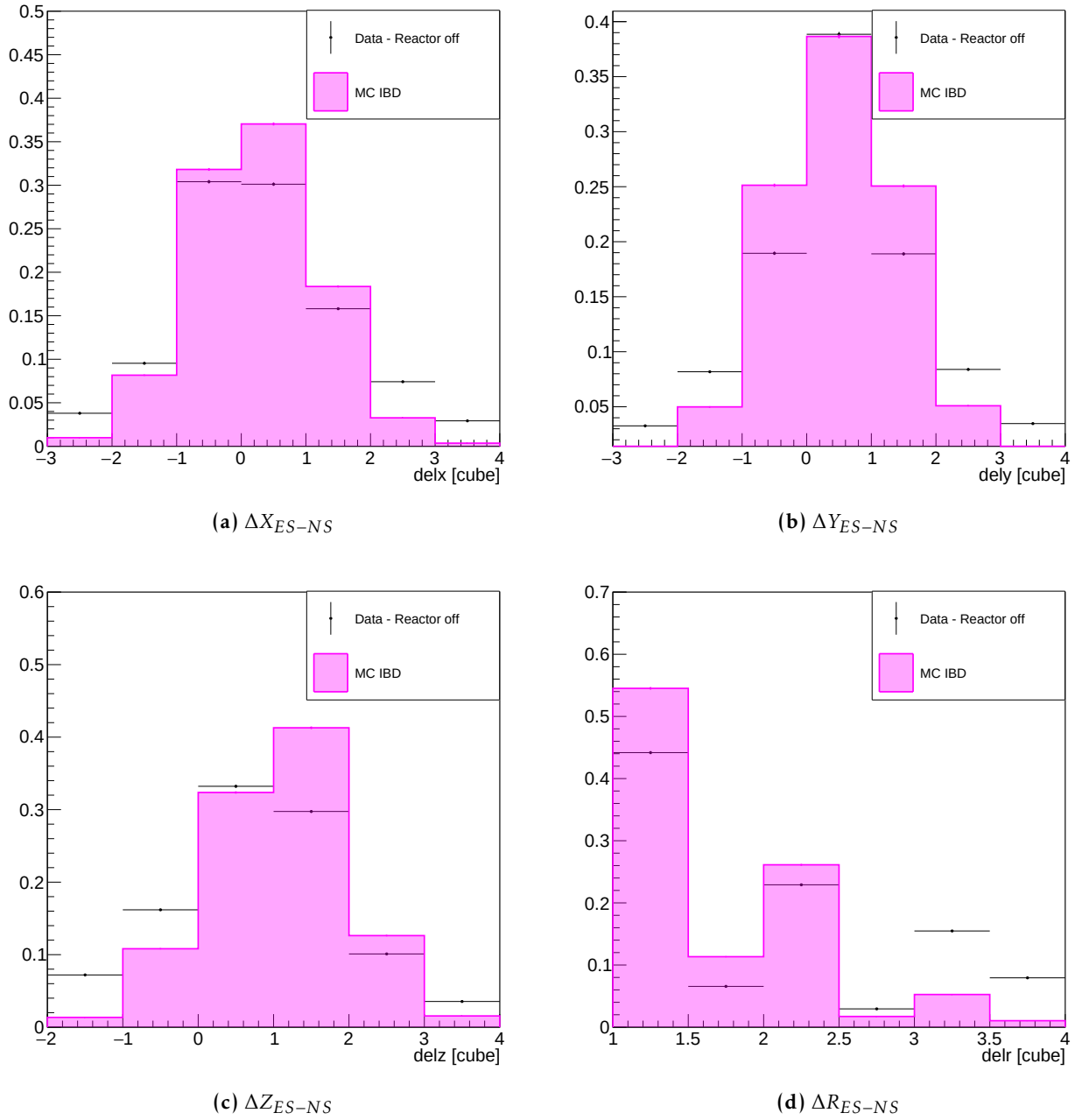
variable is ΔR_{ES-NS} (see figure 5.9d). We have then applied the following cut: $\Delta R_{ES-NS} \in]0,3]$. After this cut, no obvious discrimination power remains in the relative position between ES and NS clusters, since they are correlated with ΔR_{ES-NS}. The potential remaining discrimination power will be exploited later by the multivariate algorithms.

1-gamma category

The comparison between signal and background for the two most discriminative new variables is shown on figure 5.10. For the reconstructed energy of the gamma (figure 5.10a), the signal is peaked at the expected Compton Edge value (0.341 MeV), whereas the background has a long tail at higher energy. Thus we can exclude events for which $E_{\gamma 1}$ is larger than 0.5 MeV. We see also that the background dominates below 0.1 MeV and also exclude those events. Concerning the likelihood score ratio (figure 5.10b), we expecte it to be lower for an IBDs than for backgrounds (see section 5.2.3.2). We can see that it is indeed the case. In particular, we see that the background dominates above 8.5. We lowered a bit this cut from 8.5 to 7.5 in order to tune the IBD rate around 100 events per day, to ease comparisons with other selections. Finally, the selection is :

$$S_{Signal}^{rectangular,1} = \begin{cases} \Delta R_{ES-NS} \in]0,3] \\ E_{\gamma 1} \in [0.1, 0.5] \text{MeV} \\ \frac{\mathcal{L}_{\gamma 1}+\mathcal{L}_{\gamma 2}}{\mathcal{L}_n} \in [0.0, 7.5] \end{cases} \tag{5.15}$$

The impact of each cut is shown on figure 5.11. This selection reduces the background by a factor 6.25 and the signal by a factor 2.5 in the 1-gamma category.



(a) ΔX_{ES-NS}

(b) ΔY_{ES-NS}

(c) ΔZ_{ES-NS}

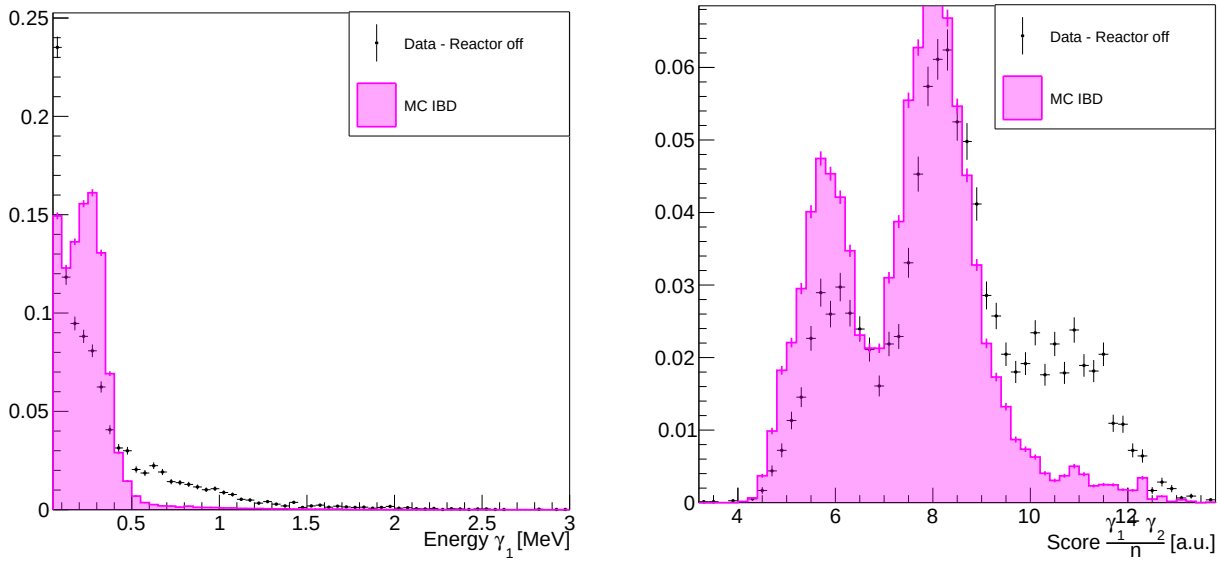
(d) ΔR_{ES-NS}

Figure 5.9.: Discrimination power of the relative position between ES and NS clusters for events of all categories passing the preselection. The signal (pink filled histogram) has an ES and NS clusters closer than the background (black point).

2-gamma category

The second category is more interesting as there are more variables available to perform the selection. Those variables are shown on figure 5.12.

As for the first category the energy of the two reconstructed annihilation gammas are very discriminative: the signal peaks around the Compton Edge and the background extends toward higher



(a) $E_{\gamma 1}$ **(b)** $\frac{\mathcal{L}_{\gamma 1}+\mathcal{L}_{\gamma 2}}{\mathcal{L}_n}$

Figure 5.10.: Discrimination power of the variables used by the IBD selection in the 1-gamma category.

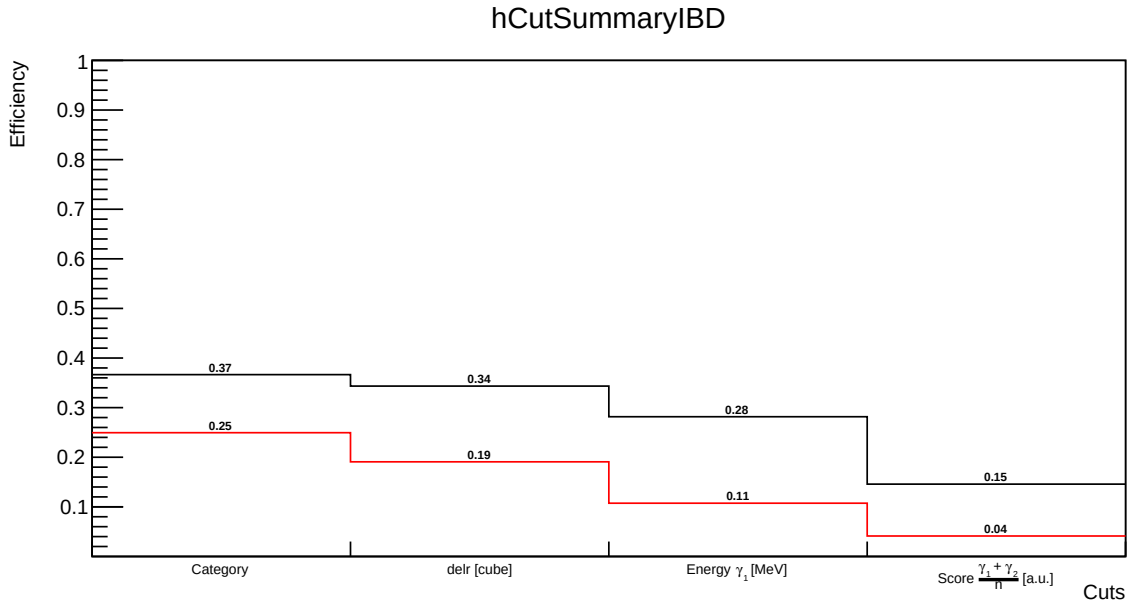


Figure 5.11.: Impact of each cut of the 1-gamma category on the signal (black) and the background (red). The reference is the number of events passing the preselection.

values(figures 5.12a and 5.12b). The dot product between the two tracks is shown on figure 5.12d. Since annihilation gammas are emitted back-to-back we expect to have more signal events in the (-1) region and it is indeed the case. Nevertheless, we observe for both signal and background peaks in the distribution at $|\text{dot product}_{\gamma 1 \gamma 2}| \simeq (0, 0.3, 0.6, 0.75)$. It is due to the tracking algorithm and to the geometry of the detector. The tracking is based on a probability to observe a cube at a given distance

Categories	0 gamma	1 gamma	2 gamma	Total
Signal [event/day]	0	33	56	89
Background [event/day]	0	368	450	818
Background / Signal	-	11.15	8.03	9.19
Signal / $\sqrt{\text{Signal} + \text{Background}}$	-	1.65	2.49	2.96

Table 5.3.: Signal and background rates obtained by applying the sequential cuts analysis. The signal over background ratio is given for each category as well as the daily significance.

and with a given reconstructed energy (see equation 5.7). As the *p.d.f.* describing the travelled distance of a gamma is an exponential decay, closest cubes to the most energetic have a lower likelihood score and thus have more probability to be the first cube of the track. When it happens, only a few discrete values are possible for the angle between the lines joining the central cube to the first cube of the 2 tracks.

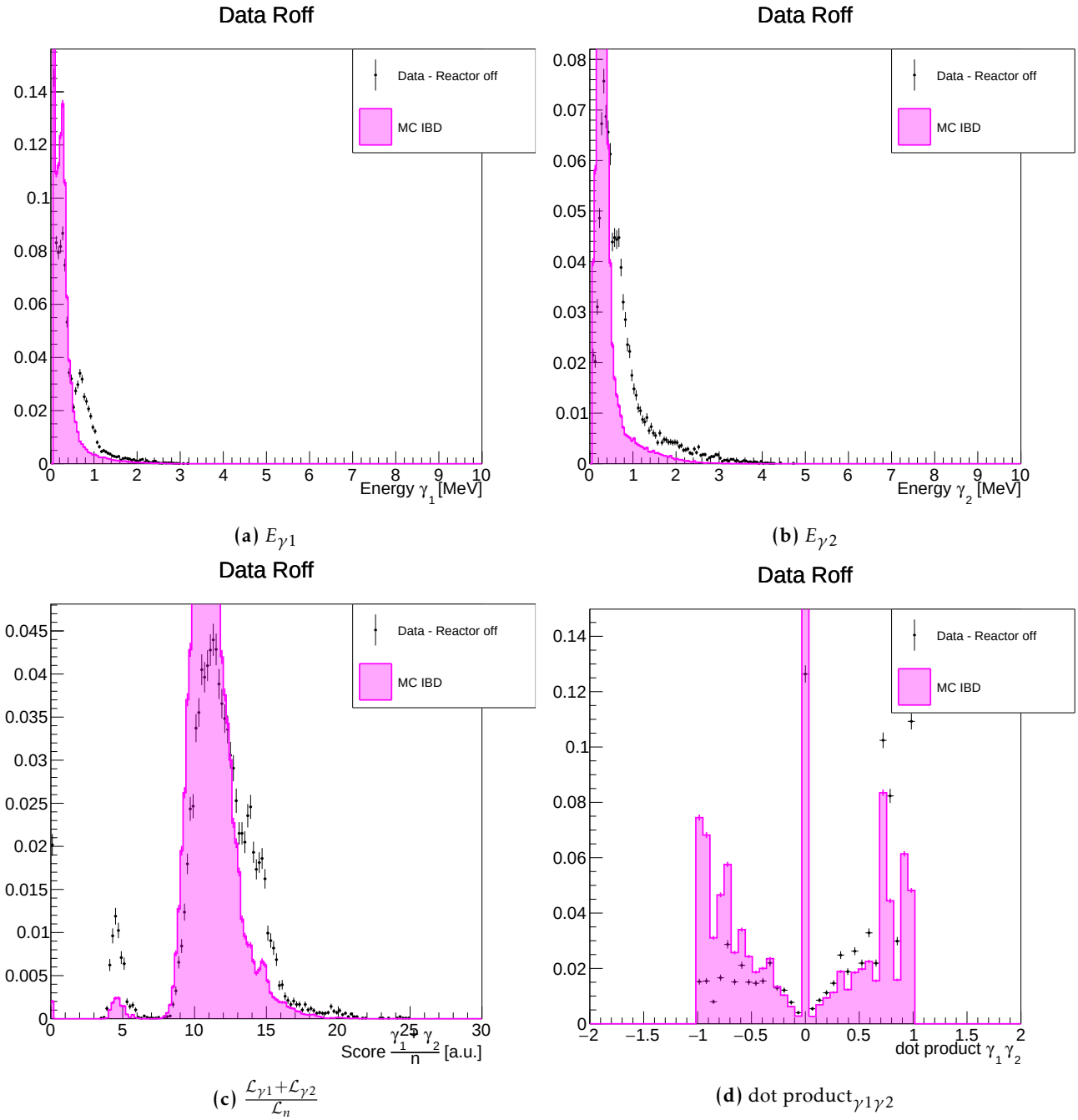
The last variable is the likelihood score. Like in the first category the background has a higher value than the signal. In the end, the selection applied to the second category is:

$$S_{Signal}^{rectangular,2} = \begin{cases} \Delta R_{ES-NS} \in]0,3] \\ E_{\gamma 1} \in [0.0, 0.5]\text{MeV} \\ E_{\gamma 2} \in [0.0, 0.5]\text{MeV} \\ \text{dot product}_{\gamma 1 \gamma 2} \in [-1, 0.7] \\ \frac{\mathcal{L}_{\gamma 1} + \mathcal{L}_{\gamma 2}}{\mathcal{L}_n} \in [0.0, 12.5] \end{cases} \tag{5.16}$$

A summary of the efficiency of each cut is given in the figure 5.13.

Performance

Finally, a summary of the expected signal and background rates is given in the table 5.3. As expected, the 2-gamma category performs better : 89 signal events per day (corresponding to an IBD efficiency of 7.4%) with a signal over background ratio of 0.11. This has to be compared with the performance obtain without topological variables : the background has been divided by a factor 1.7 while the signal goes from 110 to 89 events per day (divided by a factor 1.2). Thus topological variables have improved the background rejection but this remains insufficient to perform a sterile oscillation analysis.



(a) $E_{\gamma 1}$ **(b)** $E_{\gamma 2}$

(c) $\frac{\mathcal{L}_{\gamma 1}+\mathcal{L}_{\gamma 2}}{\mathcal{L}_n}$ **(d)** dot product$_{\gamma 1 \gamma 2}$

Figure 5.12.: Discrimination power of the variables used to perform the IBD selection in the 2-gamma category.

5.3.2.2. Multivariate based selection with topology

Multivariate analyses are commonly used in particle physics to optimise signal selection. Such analysis is based on a trained algorithm that classifies events. Usually there are only two types of events: background and signal. This algorithm is called a classifier. They require to have pure datasets of signal and background events in order to "learn" the differences. The algorithm will iteratively classify events from the training samples and will minimise a loss function which depends on the

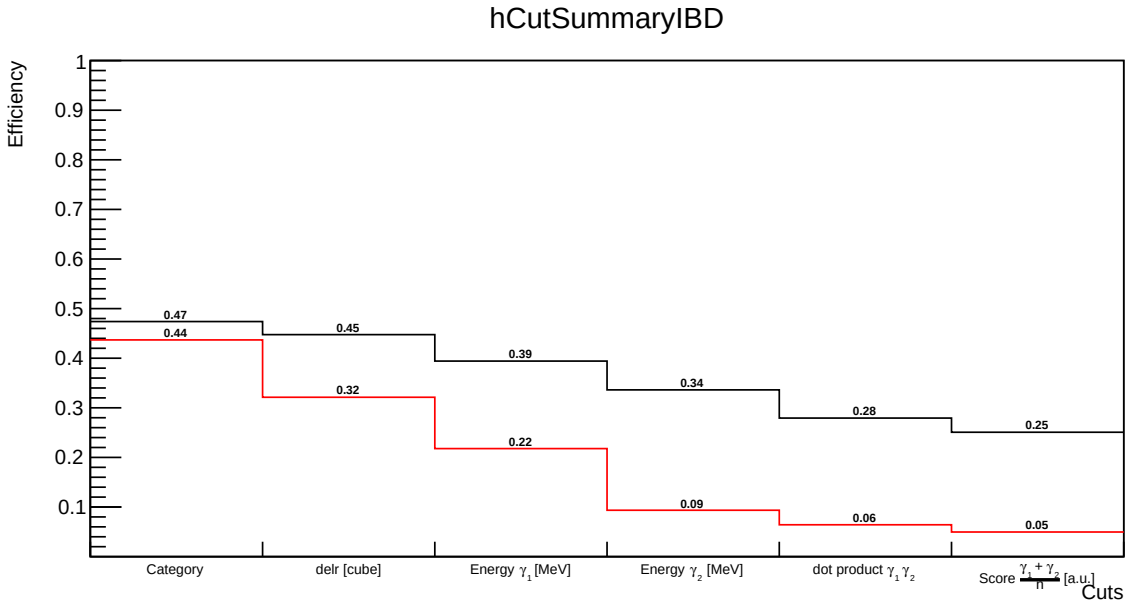


Figure 5.13.: Impact of each cut of the 2-gamma category on the signal (black) and the background (red). The reference is the number of events passing the preselection.

amount of incorrectly classified events. One has to care about the possibility of the algorithm to have learned on statistical fluctuations of the dataset. This effect is called overtraining, to control it another independent dataset is required, named *test dataset*. After training, the response of the classifier is computed for test and training datasets and are compared. If they are statistically in agreement then the classifier is not overtrained. Two very common types of classifier are Boosted Decision Trees (BDTs) and Artificial Neural Networks (ANNs). These two solutions have been adopted for our selection and for Imperial's. A quick introduction to these techniques is given in the following sections.

5.3.2.2.1. Artificial Neural Network (ANN)

During the thesis an ANN solution has been chosen to provide an alternative to the uBDT based technique developed at Imperial. Obtaining with this tool a similar result is a way to check that no information that we do not understand is used by either of these two ML approaches. The choice of the ANN was dictated by the fact that it is a long used technique in HEP, that does not possess the power of more recent NN techniques, but proved its robustness, and is moreover easy to use via the root-interfaced TMVA package. As the goal was to probe the gain to use the topology and multivariate tools, we did not use a technique that explicitly requires the uniformity of the selection efficiency as a function of the energy and the baseline.

All the analyses performed during the thesis have used the ROOT framework thus the simplest way to perform multivariate analysis is to use the TMVA toolkit [89] already included in it. There are four different ANNs available in TMVA, the user guide recommends to use the one denoted MLP, for MultiLayer Perceptron, because it is faster and more flexible than others. Figure 5.14 shows a schematic of such ANN with four input variables and one hidden layer. The output node in red gives

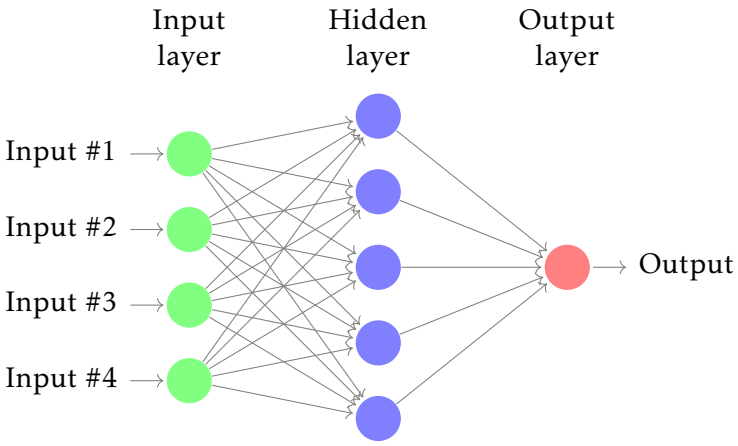


Figure 5.14.: Schematic of Artificial Neural Network with four input variables (Input layer in green) and one hidden layer with five nodes (in blue) and one output (in red). For each line between nodes, there is a weight which is tuned during the training. For each node an activation function could be defined.

the ANN score between 0 (background) and 1 (signal). During the training, the weights between nodes are tuned in order to obtain the best discrimination between background and signal.

The training is performed using the back-propagation algorithm. It is the most common recursive algorithm to adjust the weights between nodes and optimise the classification performance. This performance is characterised by an error function giving a measurement of the neural network misclassification. Thus algorithm tries to minimise this function by moving the weight values.

The ANN is used by the likelihood approach and exploits the variables provided by the reconstruction of annihilation gammas. The analysis has been performed per category according to the number of annihilation gamma reconstructed because the amount of input variables differs from a category to another. The input variables are:

0-gamma category:

- X, Y_{NS}
- $\Delta X, Y, Z, R_{ES-NS}$
- N_{cube}
- A_{dense}

1-gamma category:

- $E_{\gamma 1}$
- $N_{\gamma 1}$
- $\mathcal{L}_{\gamma 1}$
- $\frac{\mathcal{L}_{\gamma 1}+\mathcal{L}_{\gamma 2}}{\mathcal{L}_n}$
- $\frac{E_2}{E_1}$

2-gamma category:

- $E_{\gamma 2}$
- $N_{\gamma 2}$
- $\mathcal{L}_{\gamma 2}$
- Dot product$_{\gamma 1,\gamma 2}$

Where each category uses also the variables of the previous category. We can see that we are using in input the same variables as used for the selection presented in section 2.5. The variables related to the topology are described in section 5.2.3. The A_{dense} variable was introduced by people working on the muon event rejection. Especially for the type-0 muon (see section 2.4.2.1) events which are the most complicated muon events to remove since they have a low channel multiplicity. This variable is

	0-gamma	1-gamma	2-gamma
Signal	21667	49416	64029
Background	8591	7776	12705

Table 5.4.: Number of events in the training sample used for the ANN training and testing as a function of the event category.

the average of all fibre amplitudes in the ES cluster as shown by the following equation:

$$A_{dense} = \frac{1}{n_{fibre}} \sum_{i=0}^{i=n_{fibre}} A_i - C. \tag{5.17}$$

Where, n_{fibre} is the number of fibres with an amplitude above the fibre analysis threshold and A_i is the reconstructed amplitude of the fibre. C is a constant introduced by other analysers in order to have the A_{dense} distribution of type 0 muons from cosmic simulation distributed around 0. It has appeared that this variable is useful to reject muon events. Thus they have recommanded to use it in the selection of IBD events.

A simple ANN architecture has been chosen without any optimisation. There is only one hidden layer with $n_{var} + 5$ nodes, where n_{var} is the amount of input variables.

A few days of reactor-off data, between 10th August 2018 and the 14th August 2018, are used to perform the training and the testing of the ANN. Table 5.4 gives the number of events in this dataset. The distributions of signal and background for each input variable and for each category are given on figures 5.15, 5.16 and 5.17. We can see that the most discriminant variables in the 1-gamma and 2-gamma categories are indeed related to the topological variables such as: the reconstructed energy of annihilation gammas ($E_{\gamma 1}$,$E_{\gamma 2}$), the likelihood score ratio ($\frac{\mathcal{L}_{\gamma 1}+\mathcal{L}_{\gamma 2}}{\mathcal{L}_n}$) and the dot product (Dot product$_{\gamma 1,\gamma 2}$). We observe that the distribution of X_{NS} and Y_{NS} evolves as a function of the category. There are less 2-gamma events at the edge of the detector due to the possibility for one annihilation gamma to escape the detector.

After training, the comparison of the ANN response for the signal and the background between the training and the testing dataset is shown on figure 5.18. ANN response distributions with the training and the testing dataset are in a statistical agreement. This confirms the ANN is not overtrained.

Figure 5.19 shows the background and signal distribution as a function of the ANN response for the three categories. We can see that the discrimination power increases as a function of the gamma category thanks to additional discriminative variables. In particular, we can see the ability of topological variables to discriminate signal and background by comparing the 0-gamma category, figure 5.19a and the 1-2-gamma categories, figures 5.19b and 5.19c.

The evolution of signal distributions with regards to input variables and as a function of the ANN response is given on figures 5.20, 5.21 and 5.22 corresponding to events of 0-gamma, 1-gamma and

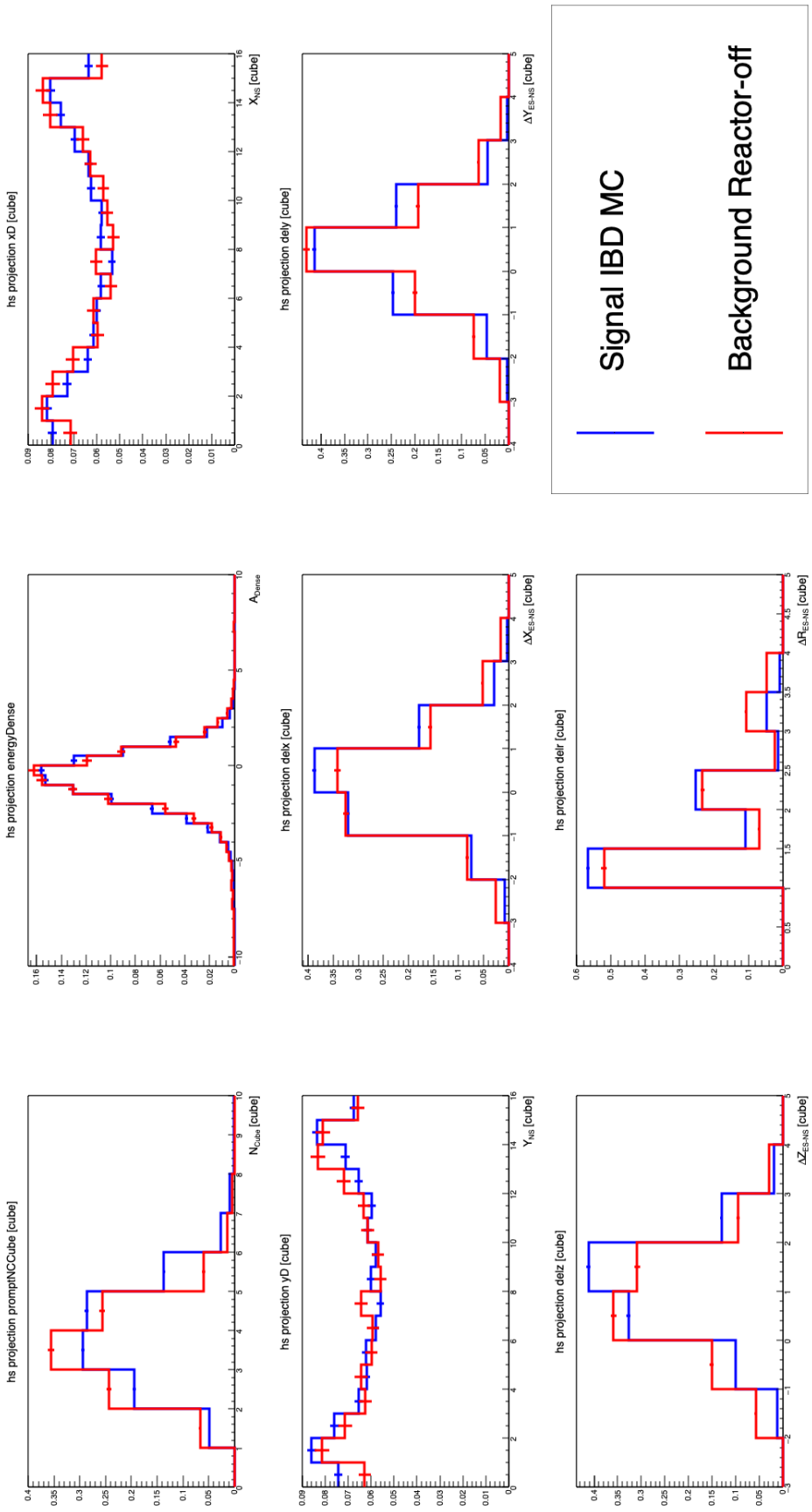


Figure 5.15.: Signal and background distributions with regards to input variables of the ANN for the 0-gamma category.

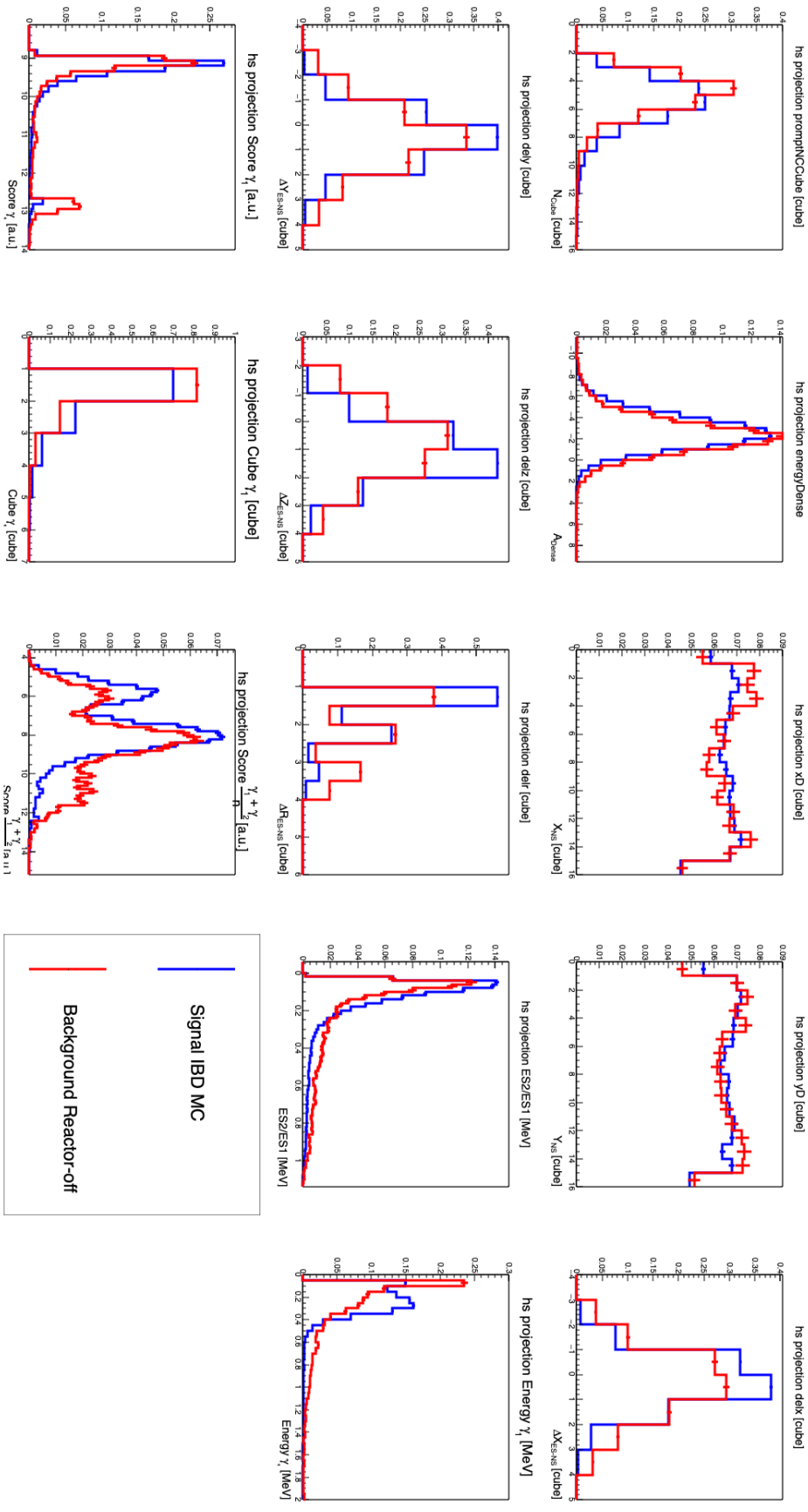


Figure 5.16.: Signal and background distributions with regards to input variables of the ANN for the 1-gamma category.

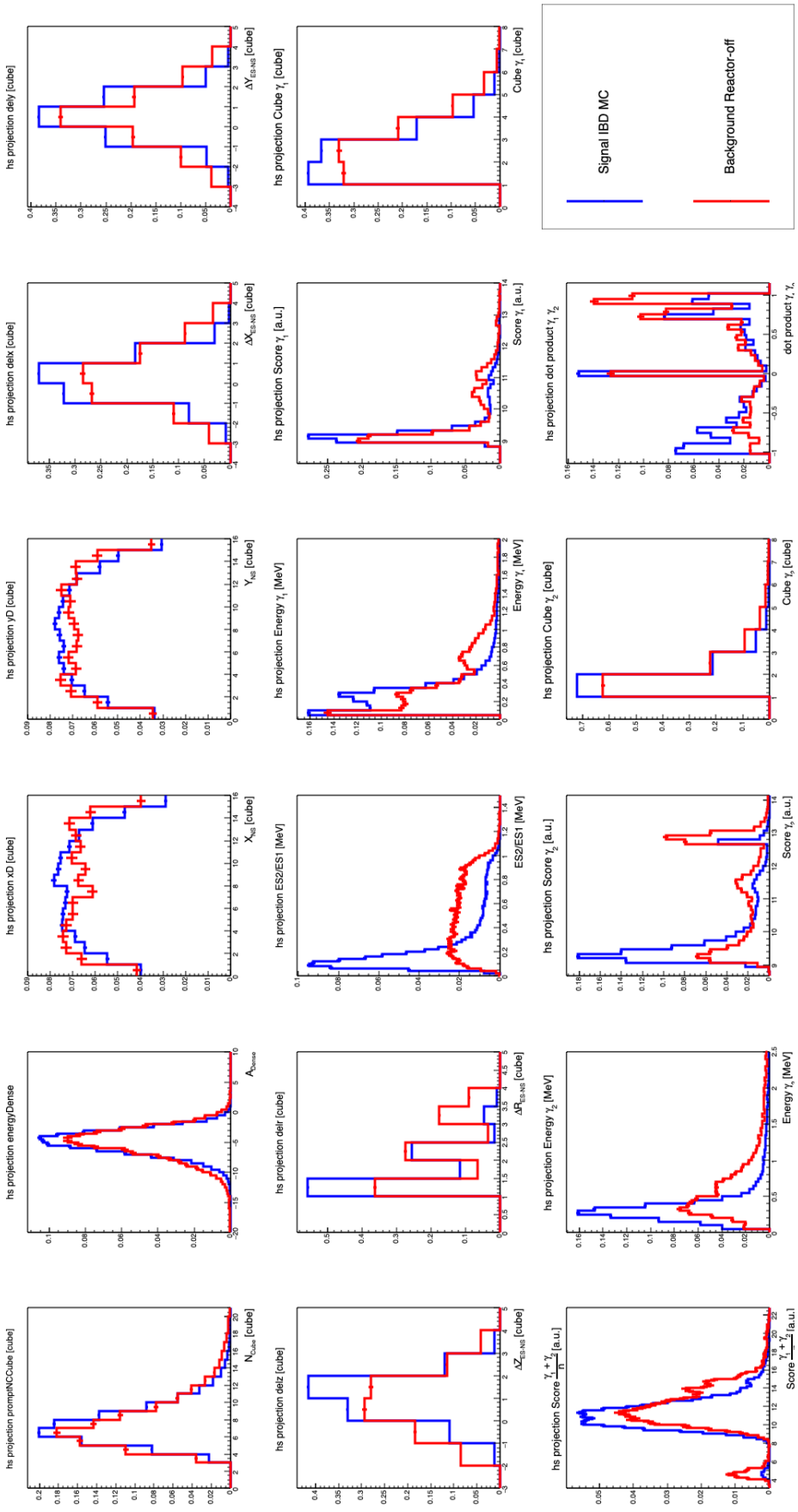


Figure 5.17.: Signal and background distributions with regards to input variables of the ANN for the 2-gamma category.

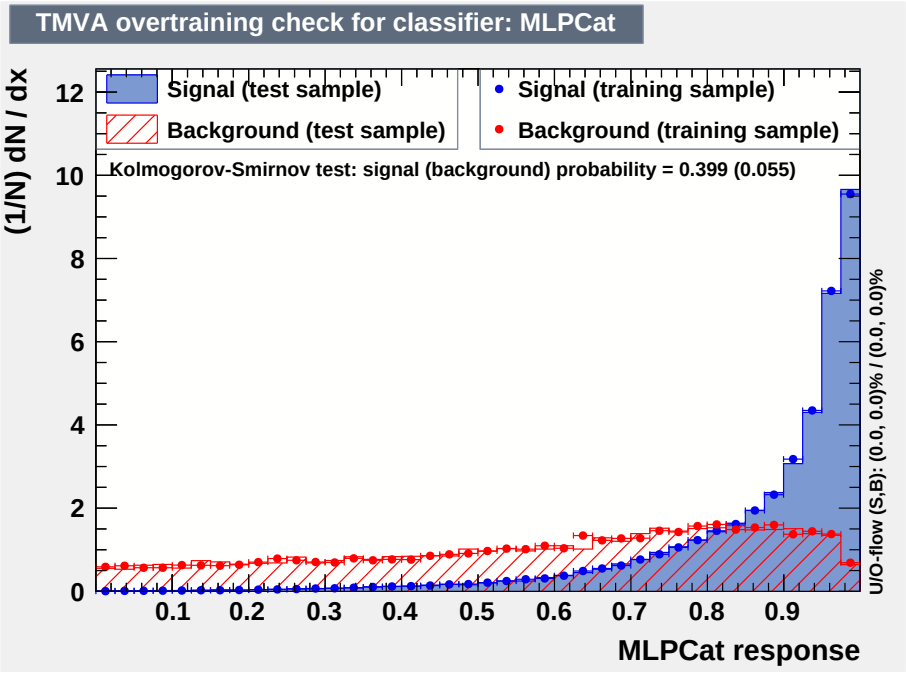


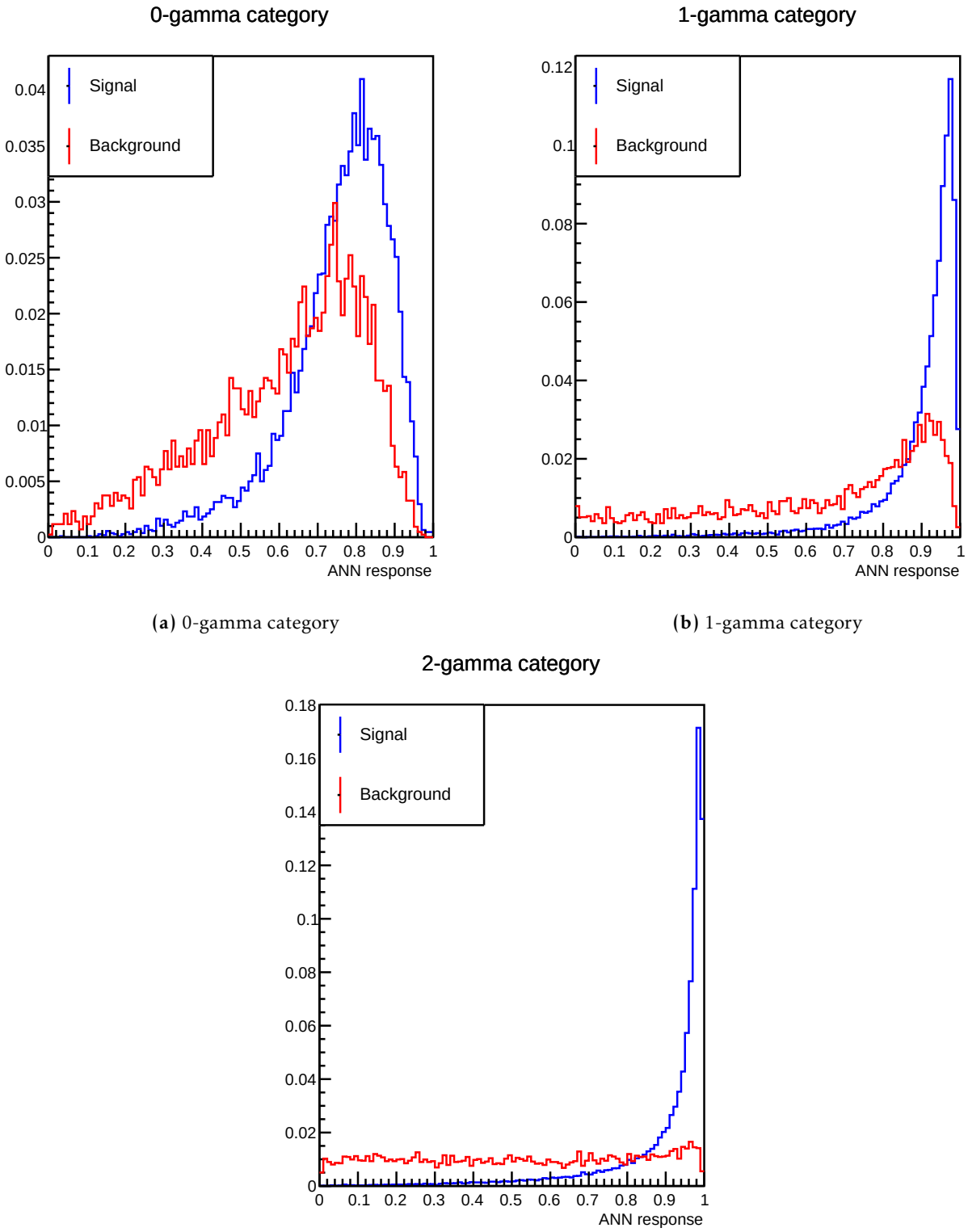
Figure 5.18.: Comparison of the ANN response for the signal (blue) and the background (red) between the training and testing datasets. We can see that for both signal and background the test and the training distributions are in a statistical agreement.

2-gamma category, respectively. The same figures with the background distributions evolution are shown on figures 5.23, 5.24 and 5.25. The histogram in dark blue corresponds to events passing the preselection, no cut on the ANN response, shown in equation 5.14 and are normalised to the unity. Other histograms are also scaled to the unity. We can see that the ANN response tends to select events with a high cube multiplicity and with close ES and NS events. For the 0-gamma category, the ANN is almost removing all signal and background events of this category as there are no clear discrimination between them and it is more efficient to remove them. Concerning events of 1-gamma and 2-gamma category, the ANN removes events for which the energy of the reconstructed annihilation gammas is higher than the Compton edge value (0.341 MeV). The likelihood score ratio is also playing a role by removing events with a value above 10 (14) for the 1-gamma (2-gamma) category. The ANN is also removing a lot of background with a dot product value higher than 0.7. We can conclude that most of the discrimination is coming from topological variables.

The working point of the ANN approach has been determined by maximising the signal significance:

$$\mathrm{FoM}_{ANN} = \frac{S}{\sqrt{S+B}} \tag{5.18}$$

where S is the total predicted signal rates and B is the total background rates. In this approach there is no need to take into account the signal and background event distribution as a function of the category as it is already taken into account during the training. Figure 5.26 shows the evolution of the signal rates 5.26a, the background rates 5.26b and the significance 5.26c as a function of the ANN response.



(a) 0-gamma category

(b) 1-gamma category

(c) 2-gamma category

Figure 5.19.: Distribution of the signal (blue) and background (red) as a function of the ANN response for the three categories.

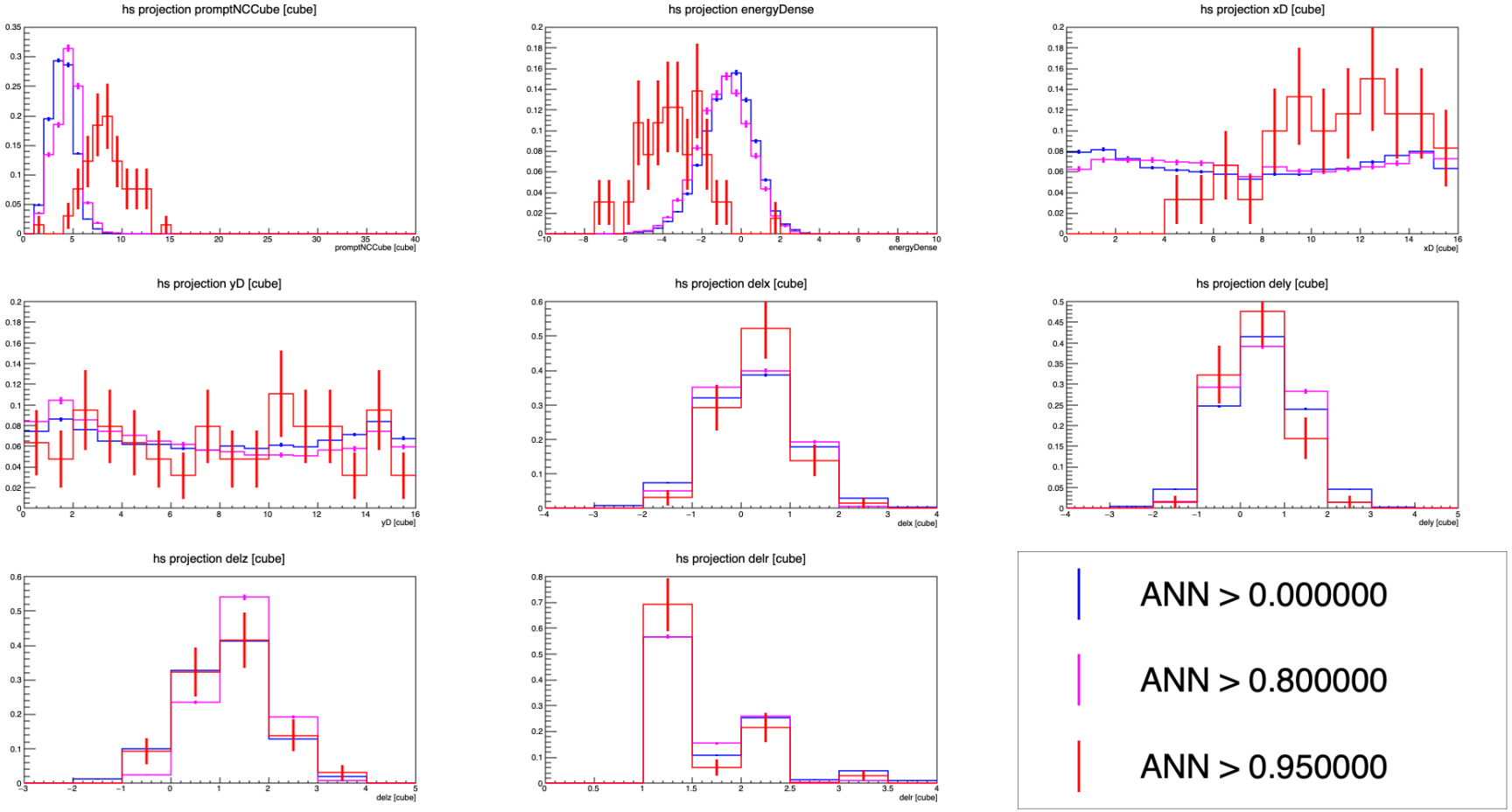


Figure 5.20.: Evolution as a function of the cut on the ANN response of 0-gamma signal event distributions with regards to input variables. The distributions are normalised to the unity.

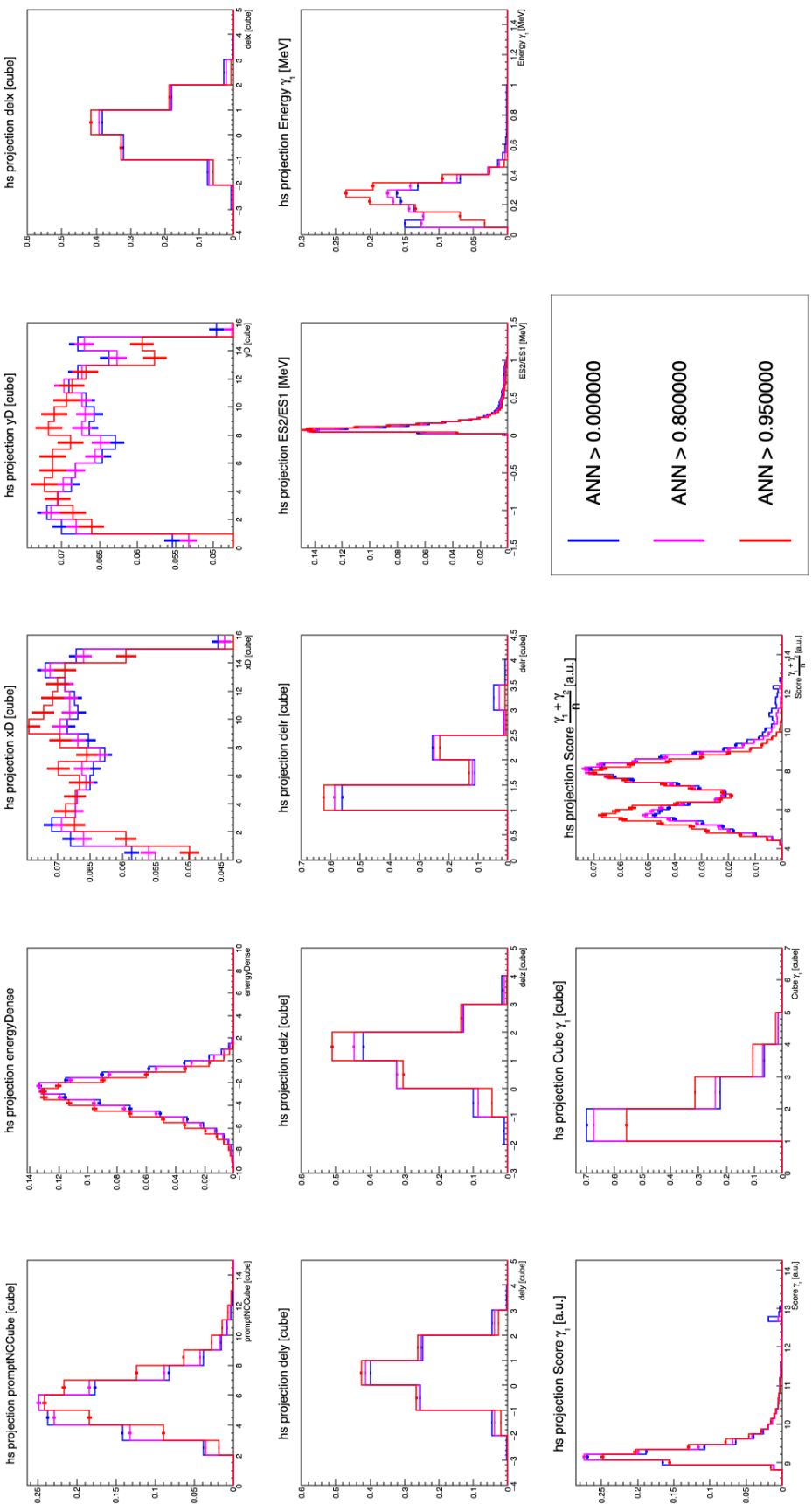


Figure 5.21.: Evolution as a function of the cut on the ANN response of 1-gamma signal event distributions with regards to input variables. The distributions are normalised to the unity.

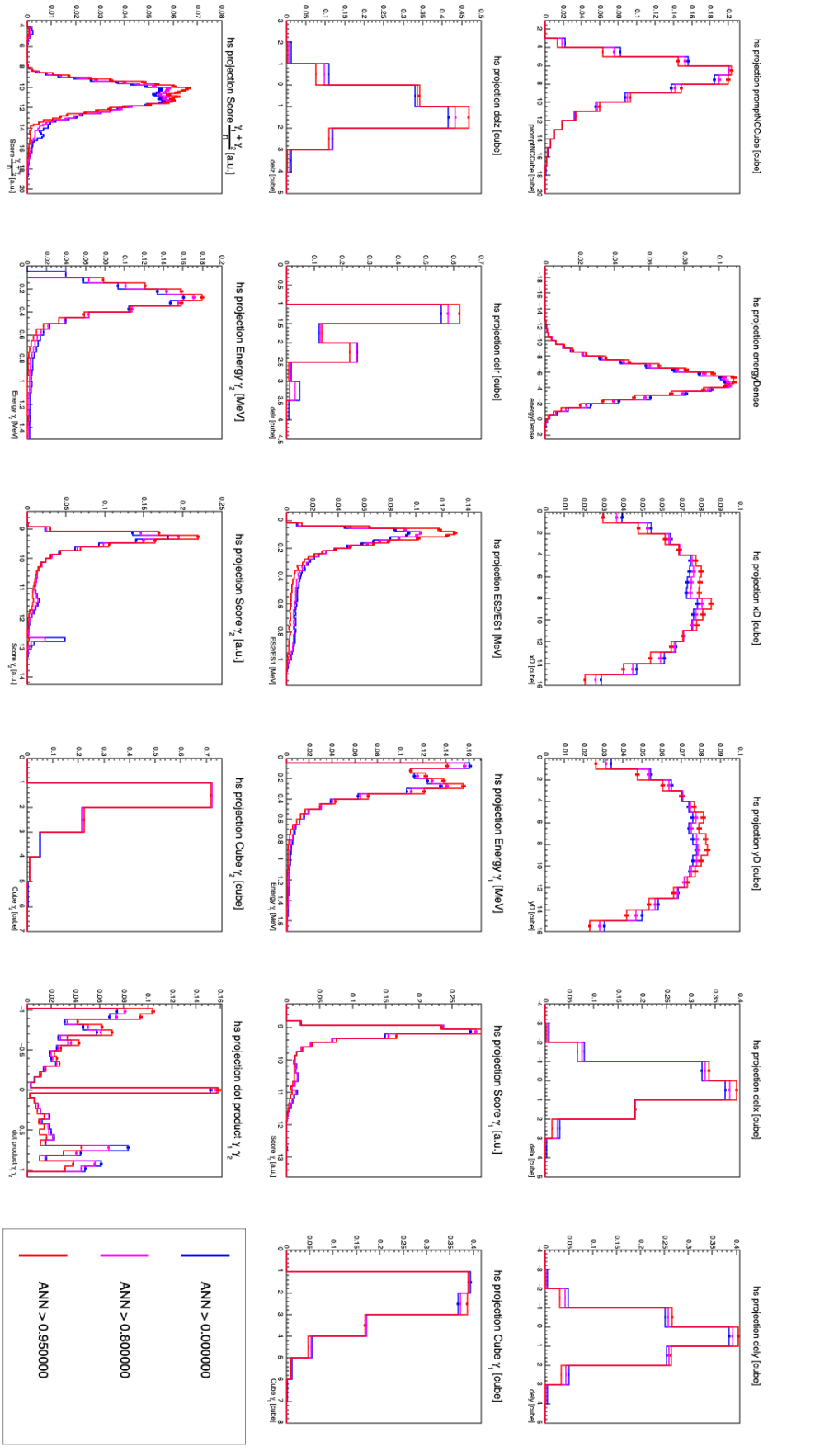


Figure 5.22.: Evolution as a function of the cut on the ANN response of 2-gamma signal event distributions with regards to input variables. The distributions are normalised to the unity.

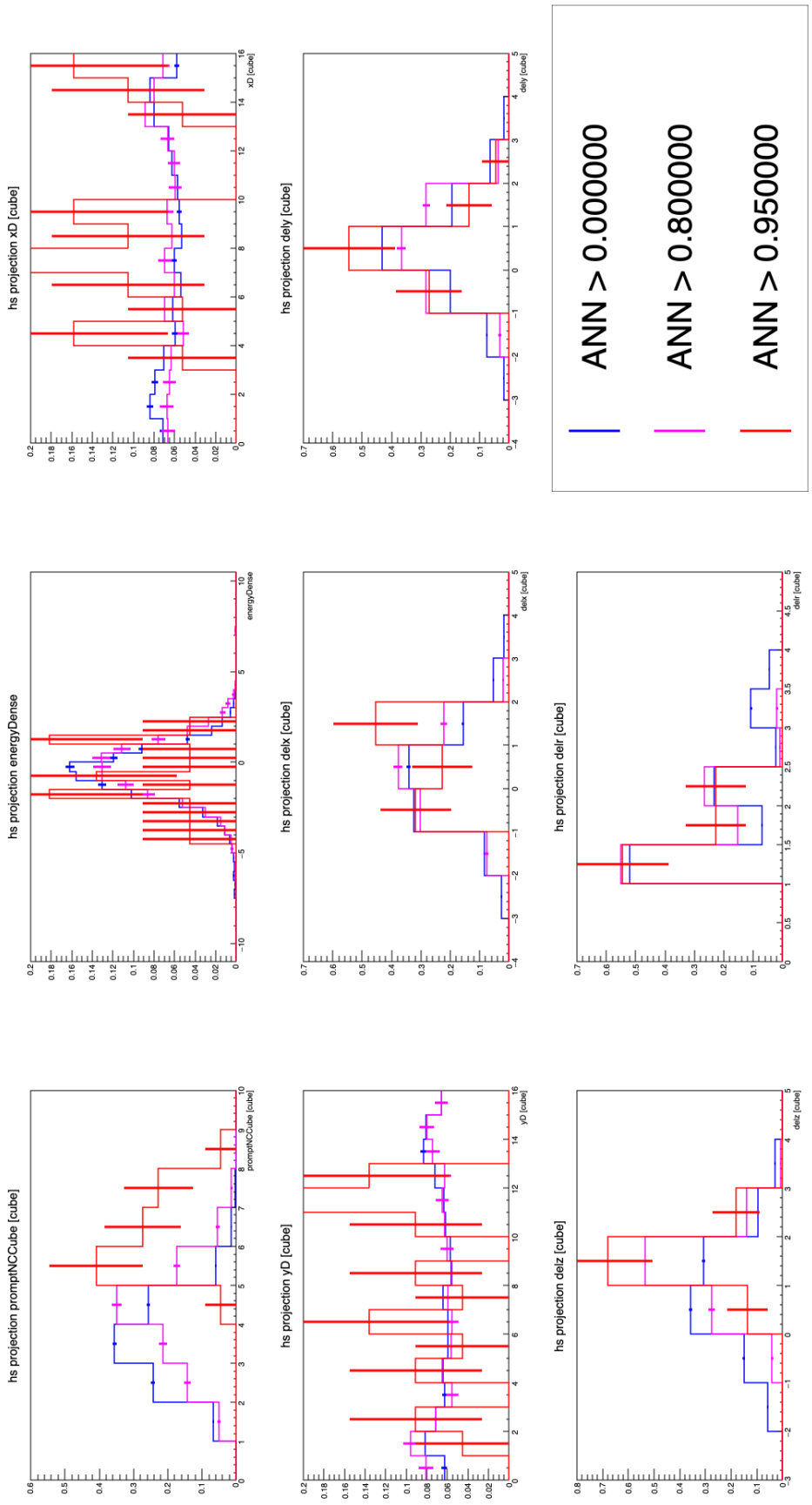


Figure 5.23.: Evolution as a function of the cut on the ANN response of 0-gamma background event distributions with regards to input variables. The distributions are normalised to the unity.

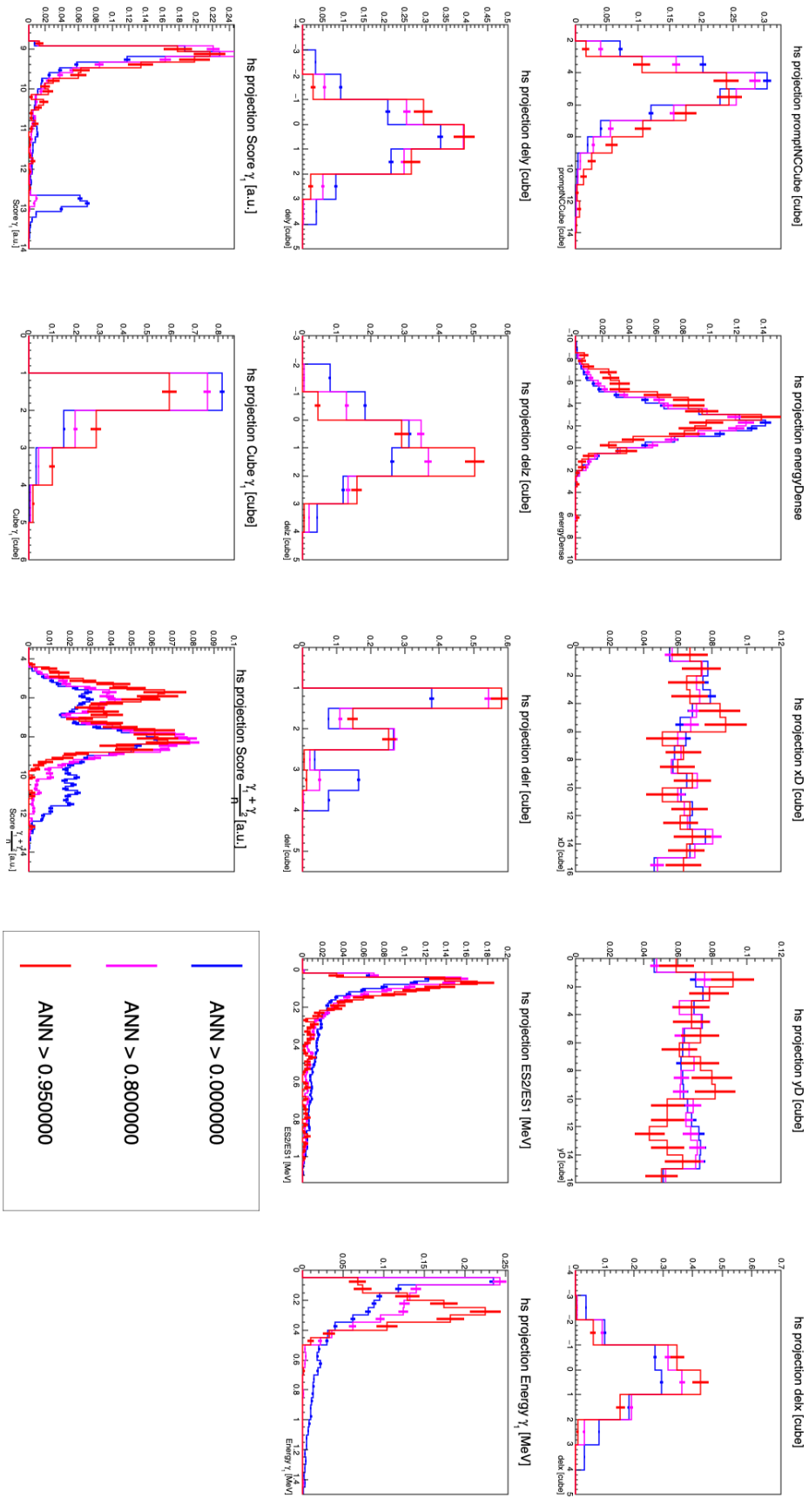


Figure 5.24.: Evolution as a function of the cut on the ANN response of 1-gamma background event distributions with regards to input variables. The distributions are normalised to the unity.

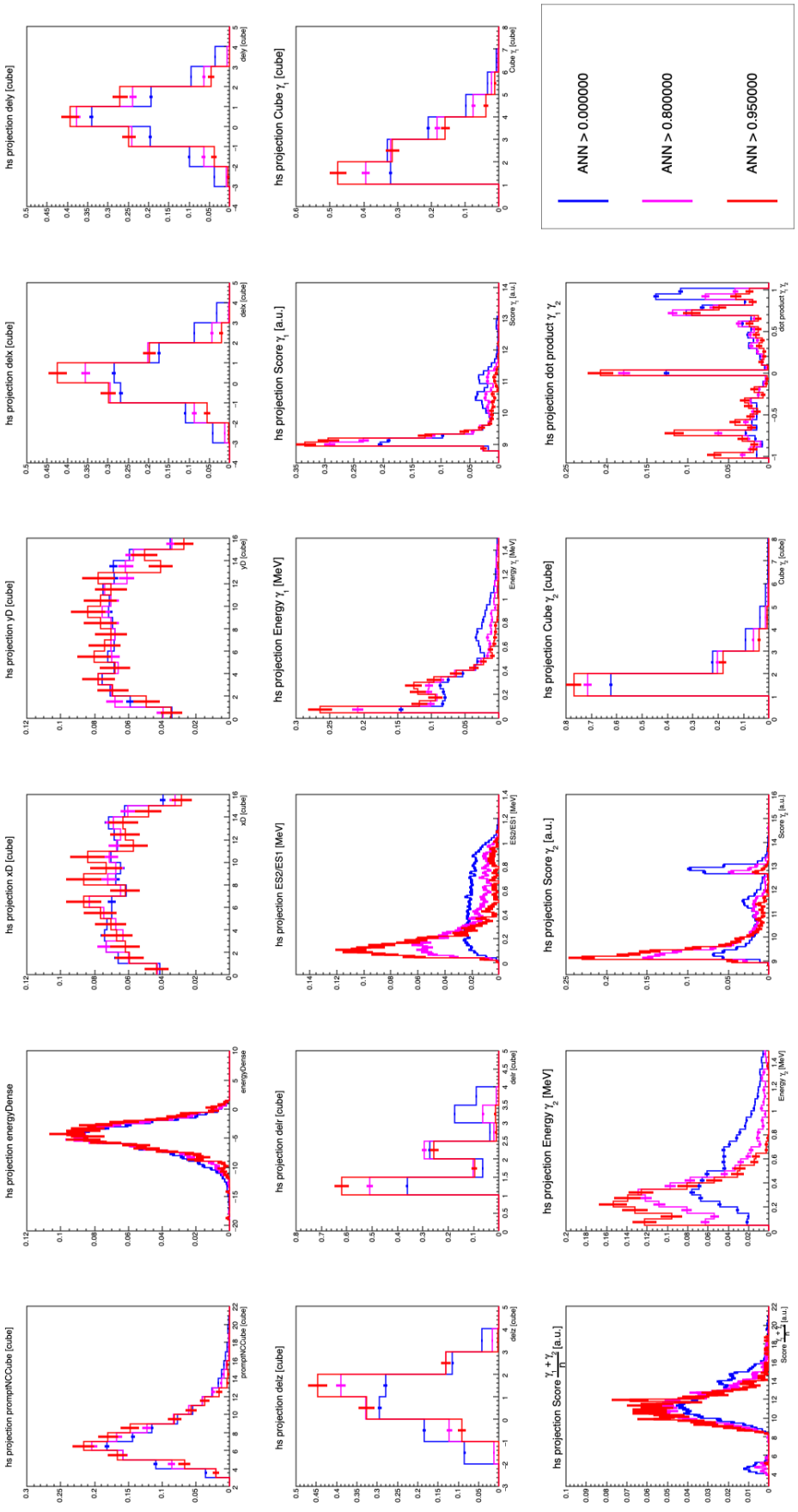
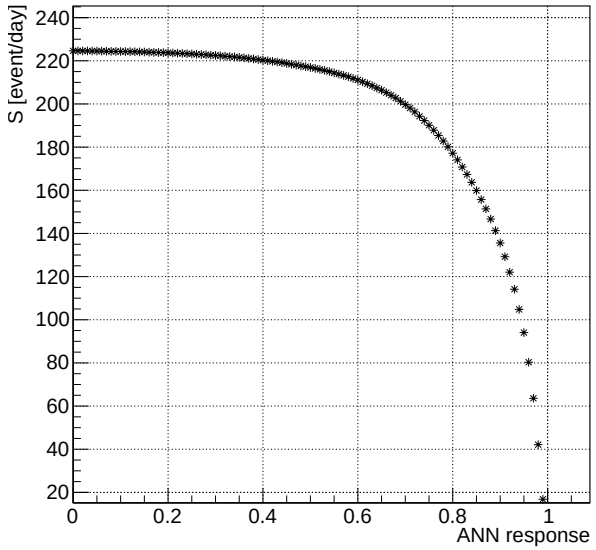
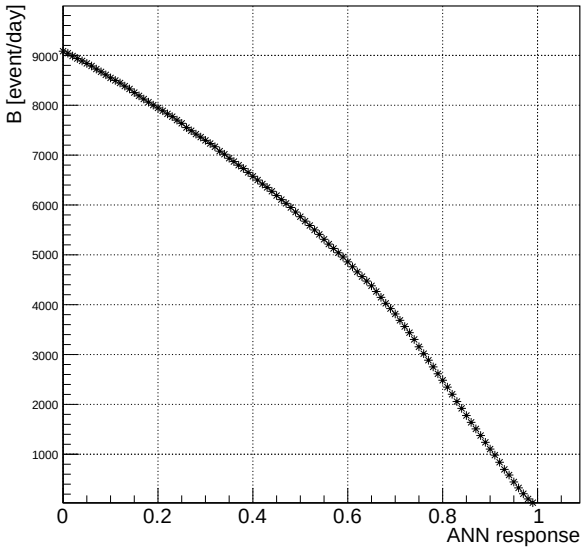


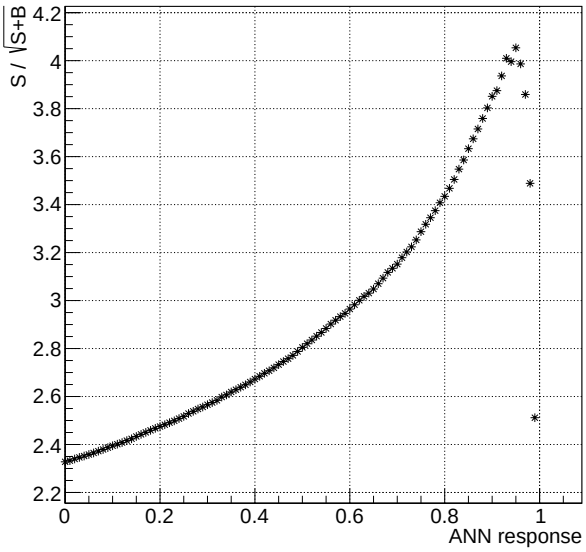
Figure 5.25.: Evolution as a function of the cut on the ANN response of 2-gamma background event distributions with regards to input variables. The distributions are normalised to the unity.



(a) Signal rate per day (IBD simulation).



(b) Background rate per day (Reactor-off test dataset).

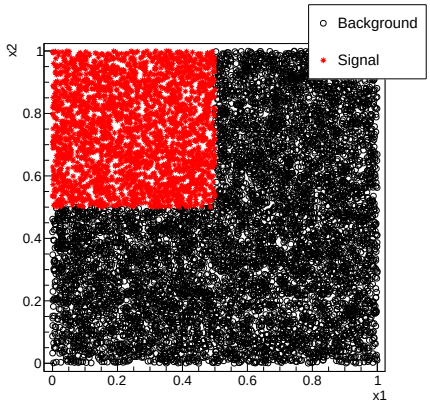


(c) Significance.

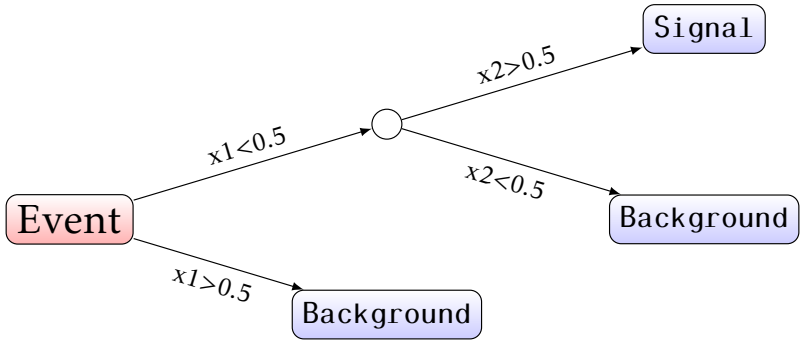
Figure 5.26.: Predicted performances obtained with the testing dataset as a function of the ANN response.

Signal [event/day]	Background [event/day]	B/S	FoM
94.06	444.35	4.72	4.05

Table 5.5.: Summary of the performance obtained using the working point obtained with equation 5.18.



(a) 2D distribution of the signal (red stars) and the background (black circle) as a function of the two input variables x_1 and x_2.



(b) Schematic of the decision tree allowing to fully classify an event according to input variables x_1 and x_2.

Figure 5.27.: Simple example of classification problem with two input variables and two event classes.

We can see on figure 5.26c that the maximum significance is obtained by cutting on the ANN response variable at 0.95. This is the working point of the ANN analysis. At this value the predicted performance are given in table 5.5. We reach an IBD efficiency of 7.8% (94 neutrinos per day) and a signal over background ratio of 0.21. This is a factor 2 better than the standard selection with topology presented in section 5.3.2.1.

5.3.2.2.2. Uniform Boosted Decision Tree (uBDT)

A decision tree is one of the oldest methods for supervised learning. The structure of a decision tree is composed of nodes in which the input data is divided in two sets according to one of input variables. In order to fix ideas, let's assumes a simple example where there are two classes of events *Signal* and *Background* and events are composed of only two variables x_1 and x_2. The figure 5.27a is showing the signal and background distributions in the two-dimensional space (x_1, x_2).

The corresponding decision tree that classifies an event into background or signal would have the structure shown in figure 5.27b. This simple example performs a classification without error, meaning that the purity of events attributed to signal or background is equal to one. But in the case of a classification problem with a lot of variables that cannot fully discriminate the two types of events the decision tree would have a more complex structure and would not give a perfect classification.

In order to fully handle the complexity of the problem, new classes of classifiers have been introduced. They are called Boosted Decision Tree (BDT). The idea is to perform several training on the

same dataset. At the end of each training an weight is assigned to incorrectly classified events in order to give them more importance for the training of the next decision tree. Like this, at each iteration the new decision tree learned from mistake of others. At the end, the decision is taken by a weighted majority vote of all decision trees trained. This technique enhances the classification performance and is called *boosting*.

This analysis has been performed by the Imperial team of SoLid. A sterile oscillation analysis takes as input the signal distribution as a function of the prompt energy and the baseline. Thus it is required to avoid any distortion of those variables while performing the signal selection. For this reason they have decided to use a specific BDT approach called Uniform Boosted Decision Tree (uBDT) [90]. It was originally developed in order to perform Dalitz analyses at LHCb and Flavor Factories. The goal is to add weight to events that are in low efficiency phase-space area (in the same way as the boosting). So after multiple iterations, the global response of all trained classifiers tend to have a uniform selection efficiency with regards to the variables of interest. In the case of this analysis, the two variables that are required to be uniform are E_{prompt} and Z_{ES}.

They used the spatial reconstruction of annihilation gamma discussed in section 5.2.2 as input of the classifier. Three uBDTs are trained independently according to the number of reconstructed annihilation gammas as the number of input variables differs. The input variables as a function of the category are the following:

0-gamma category:

- X, Y_{NS}
- $\Delta X, Y, Z, R_{ES-NS}$
- N_{cube}
- A_{dense}

1-gamma category:

- $E_{\gamma 1}$
- $N_{\gamma 1}$
- $d_{AC-\gamma 1}$
- $\frac{E_2}{E_1}$

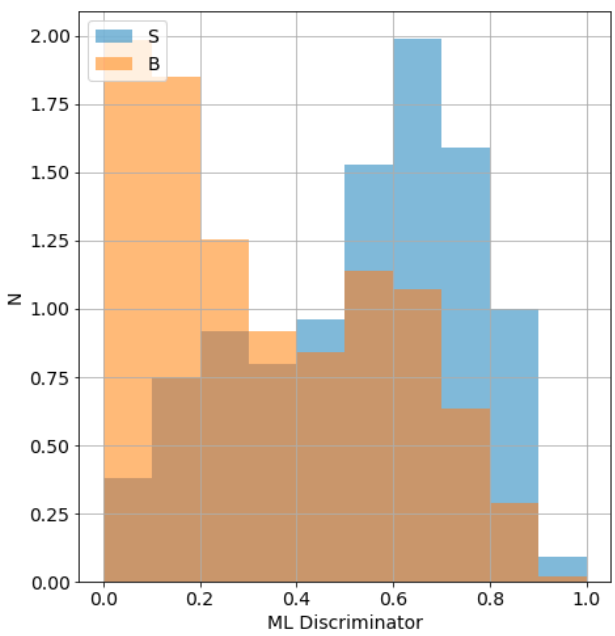
2-gamma category:

- $E_{\gamma 2}$
- $N_{\gamma 2}$
- $d_{AC-\gamma 2}$
- Dot product$_{\gamma 1,\gamma 2}$

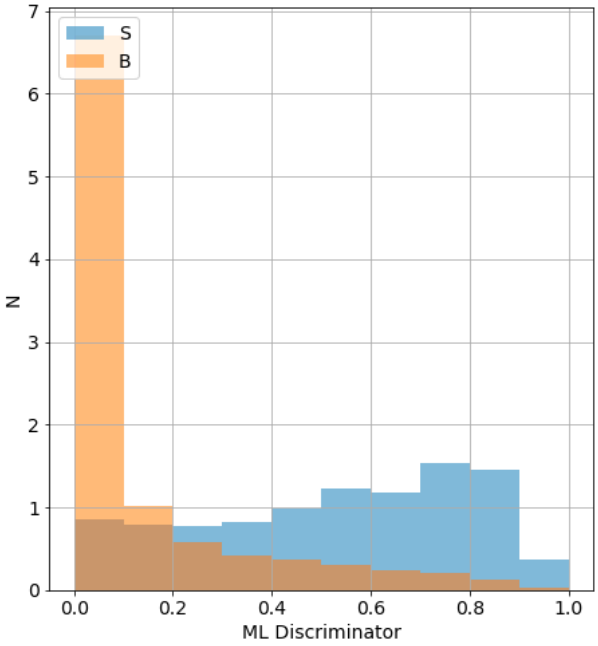
We can see that the same variables are used as input as the ANN approach (see section 5.3.2.2.1), the only difference concerns the topological variables where the Spatial reconstruction of annihilation gamma is used (see section 5.2.2). After training, the discrimination power achieved with this approach between signal and background for the three categories is given on figure 5.28.

As there are three uBDTs, they have to find the best working point taking into account the efficiency and the background rejection in each category. The best set of cuts is obtained by optimising the equation 5.18. Once the cuts have been optimised the uniformity criteria was cross-checked (see figure 5.29). The selection is uniform as a function of the baseline 5.29b within a few percent. The uniformity criteria is less good as function of the prompt energy especially for the 0-gamma category. Nevertheless those corrections could be taken into account during the fitting procedure via the migration matrix (see section 2.5.1).

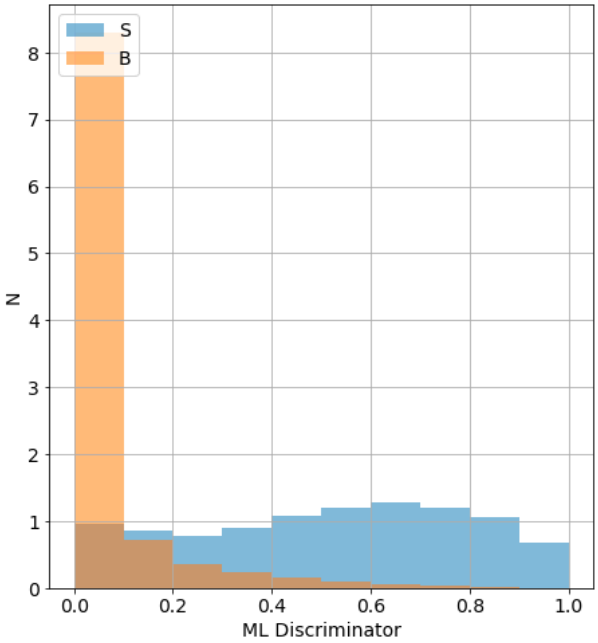
The performance of this selection are shown in the section 5.3.3 where all selections are compared.



(a) 0-gamma category.

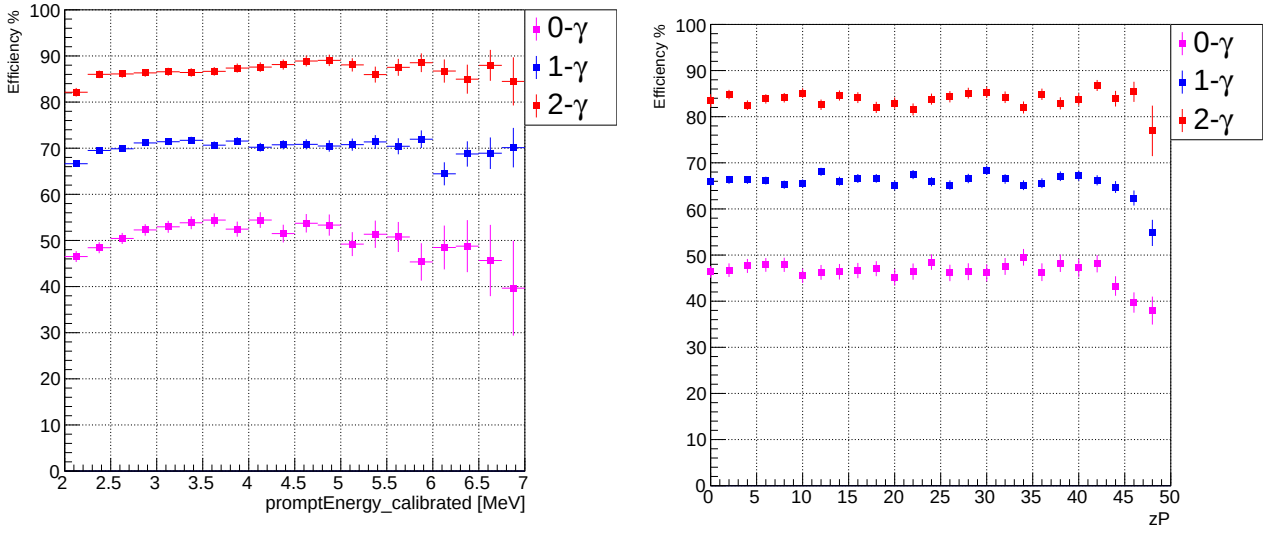


(b) 1-gamma category.



(c) 2-gamma category.

Figure 5.28.: Discrimination power obtained with the uBDT approach for the three categories.



(a) Signal selection efficiency as a function of the prompt energy.

(b) Signal selection efficiency as a function of the baseline.

Figure 5.29.: Signal selection efficiency as a function of the two variables required to be uniform, using the best working point in the uBDT approach. The three categories are shown on the figure.

5.3.3. Performance comparison

The predicted performances of the uBDT and ANN approaches have been compared. The performance are derived using IBD simulation for the signal and few days of reactor-off for the background. The figure 5.30 is showing the background over signal ratio as a function of the predicted excess for the four selections presented in this manuscript:

- The standard approach without topology (see section2.5)
- The standard approach with topology (see section 5.3.2.1)
- The uBDT approach (see section 5.3.2.2.2)
- The ANN approach (see section 5.3.2.2.1)

A clear improvement is seen by using the annihilation gamma reconstruction and the multivariate tools. The first one gives an improvement in the background rejection by a factor 1.7 and second one by a factor 2. It is interesting to see that the uBDT and the ANN approach gives the same level of performance with a predicted excess of 100 neutrinos per day and a background over signal ratio of 5. Nevertheless those results still have to be confirmed using real data. Up to this point, the signal rate was derived from the full simulation of the detector.

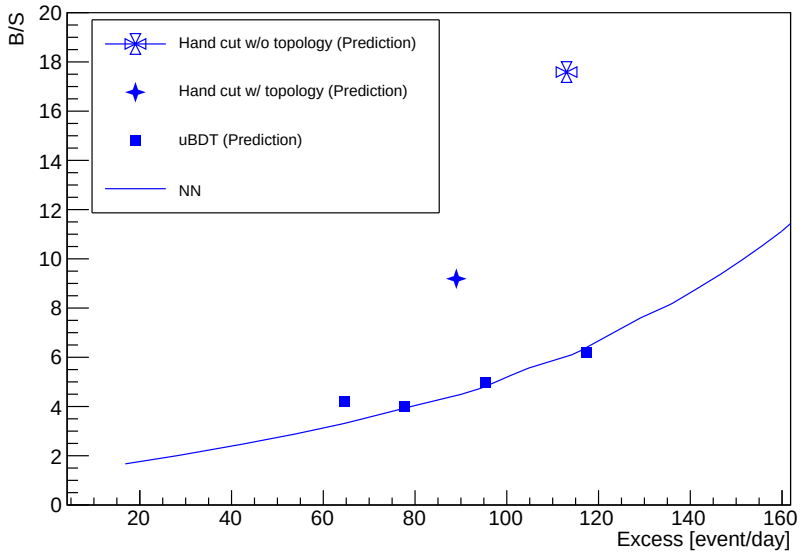


Figure 5.30.: Background over signal as a function of the daily number of IBDs predicted by the simulation, for the four analyses presented in the manuscript. There is a clear improvement by using the annihilation gamma topology and multivariate analysis. Also the two multivariate analyses are in agreement.

5.4. Signal extraction

The SoLid experiment takes advantage of pure background sample provided by reactor-off period. Thus a data-driven background model could be built. Given the rather low signal over background ratio predicted in the previous section, the uncertainty on the background subtraction is expected to be one of the large sources of systematic uncertainties in the oscillation search. Procedures to subtract the background are described below. In this thesis, we use them to evaluate in real data the daily rate of IBDs selected by the various analyses described in the previous sections. This will be compared with the results on figure 5.30 in order to assess the robustness of these evaluations. Indeed, in the previous section, the signal rate was determined with the simulation (see in chapter 2 the section on the figures of merit). In spite of the results obtained in chapter 4, the simulation might still misestimate the signal efficiency.

The reactor-off and reactor-on coincidence rates are the sum of several components given by the following equation:

$$\begin{aligned} R_{\text{Reactor-on}}(t) &= R_{IBD}(t) + R_{Cosmic}(t) + R_{BiPo}(t) + R_{Acc}(t) \\ R_{\text{Reactor-off}}(t) &= R_{Cosmic}(t) + R_{BiPo}(t) + R_{Acc}(t) \end{aligned} \tag{5.19}$$

The difficulty of this signal estimation resides on the control of two distinct correlated backgrounds: the background rates in equation 5.19 are not expected to be strictly the same in reactor-on and

Components	BiPo	Cosmic induced fast neutrons	IBD
τ_{ES-NS} [μs]	256	62	62
BiPonisher	Low	High	High

Table 5.6.: Table summarising the properties of each correlated components with regards to the ΔT_{ES-NS} and BiPonsiher variables.

reactor-off samples. Indeed, the BiPo which evolves because of radon release and the cosmic induced one because of it correlations with the atmospheric pressure.

This section presents the methodology employed to solve that problem. Two methods are actually presented to perform this signal extraction and finally the results are shown and compared to those based on simulation.

The motivation to compare two subtraction methods relates issues that we detected while using the official subtraction method (described in section 5.4.1). We decided to develop another approach (described in section 5.4.2) to cross-check the first one.

For each selection S_{sel}, the corresponding event rate per day is noted R_{sel}. It is obtained by counting the number of events passing the selection S_{sel} in a time bin divided by the live-time in this time bin. All the figures shown in this section have been derived using the ANN selection and cutting at the working point determined in the section 5.3.2.2.1.

5.4.1. Sideband extraction and remaining background estimation

5.4.1.1. Description

The traditional approach developed by the collaboration consists to define sideband selections to have pure samples of accidental and BiPo backgrounds, and to derive the expected amount of these backgrounds in the signal region. This is performed by using the differences in the $\Delta T_{ES-NS} = t_{NS} - t_{ES}$ and BiPonisher distributions between the various event types, a summary of those differences is given in the table 5.6.

5.4.1.2. Accidental component

The uncorrelated background is the association of a prompt and a delayed signal without any correlations, meaning that they have different physics origins, for example a gamma from a ^{41}Ar decay with an atmospheric neutron. This background is flat in ΔT_{ES-NS}.

The measurement of this background is performed in SoLid by using the neutron trigger properties described in the section2.2.4.2. As said when a waveform is passing the neutron trigger condition, the

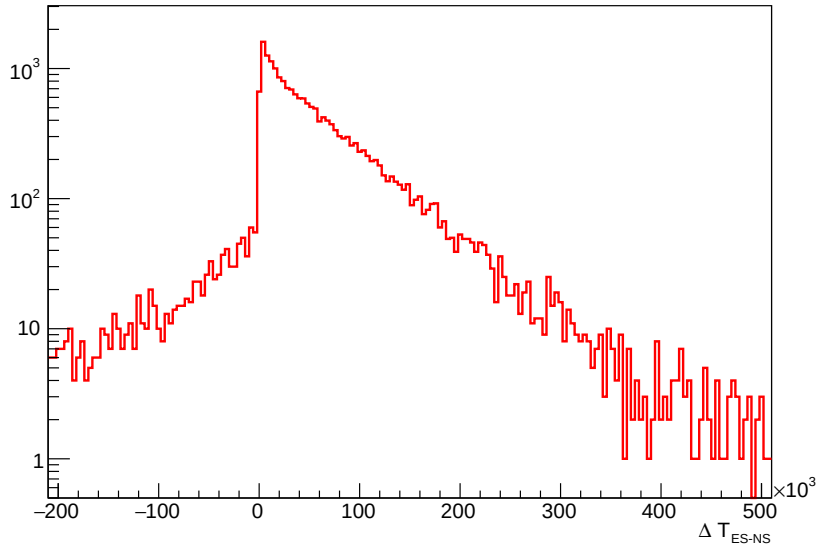


Figure 5.31.: ΔT_{ES-NS} distribution of the simulated cosmic induced background. The events are exponentially distributed in the negative time window showing a pollution in the uncorrelated time window.

data -500 μs before and 200 μs after the first peak of the waveform is recorded. Thus the coincidences with a negative ΔT_{ES-NS} value are the ones with a delayed signal found before the prompt signal. There should be no correlation between the two signals. Actually, it is wrong. Simulations of cosmic induced backgrounds have shown that there is a correlated component in this negative window, shown on figure 5.31. To prevent against this pollution the uncorrelated background is estimated by selecting coincidences such as:

$$\Delta T_{ES-NS} \in [-200, -100]\mu s \tag{5.20}$$

Thus assuming a selection S_{Signal}, the accidental sideband of this selection noted S^{Acc}_{Signal} is obtained by changing the cut on the ΔT_{ES-NS} variable by the condition 5.20. In order to have a consistent rates between the two, a scaling factor is applied according to the time window length requested in S_{Signal}. Then if this cut requires the following condition: $\Delta T^{Signal}_{ES-NS} \in [\Delta T_{min}, \Delta T_{max}]$, the scaling factor $c_{Acc \to Signal}$ is:

$$c_{Acc \to Signal} = \frac{\Delta T_{max} - \Delta T_{min}}{100} \tag{5.21}$$

where the denominator comes from the size of the window selected by the condition 5.20. Then the signal rate without the accidental component is obtained with:

$$R_{Signa-Acc} = R_{Signal} - c_{Acc \to Signal} \times R^{Acc}_{Signal}. \tag{5.22}$$

This is performed for both reactor-on and reactor-off days.

5.4.1.3. BiPo component

The correlated background is less simple to subtract than the uncorrelated one. There are two components in the correlated background: the BiPo and the cosmic induced. As for the accidental background, it is possible to define a sideband for the BiPo background. Those side bands are defined again based on ΔT_{ES-NS} and the BiPonisher variable (see figure 2.17).

As seen the BiPo background has a higher value of τ_{ES-NS} and a lower value of BiPonisher variable compared to other background components. Therefore, the BiPo sideband selection, noted S_{BiPo}, is obtained by changing the cuts on the variables ΔT_{ES-NS} and BiPonisher such as:

$$S_{BiPo} = \begin{cases} S_{Signal} \\ \Delta T_{ES-NS} \in [300, 500]\mu s \\ \text{BiPonisher} \in [0.0, 1.44] \end{cases} \tag{5.23}$$

Of course we also have to subtract from this selection the accidental component. For this reason, we define the BiPo accidental selection, noted S^{Acc}_{BiPo}, as:

$$S^{Acc}_{BiPo} = \begin{cases} S_{Signal} \\ \Delta T_{ES-NS} \in [-200, -100]\mu s \\ \text{BiPonisher} \in [0.0, 1.44] \end{cases} \tag{5.24}$$

For each selection corresponds a rate which is given by the number of selected events divided by the live-time. The accidental free BiPo background rate is obtained via:

$$R_{BiPo-Acc} = R_{BiPo} - c_{Acc \to Signal} \times R^{Acc}_{BiPo} \tag{5.25}$$

It remains to find the scaling factor needed to determine the rate of BiPo in the signal region from $R_{BiPo-Acc}$. First, the BiPonisher cut has been inverted. Indeed in the signal selection we require BiPonisher $\in [1.44, 3]$ and the BiPo sideband requests BiPonisher $\in [0.0, 1.44]$. From figure 2.17, we can derive the scaling factor by computing the integral of the enriched BiPo distribution. This scaling

factor is given by the following equation:

$$c^{BiPonisher}_{BiPo\rightarrow Signal} = \frac{\int_{1.44}^{3} f_{BiPo}(x)dx}{\int_{0}^{1.44} f_{BiPo}(x)dx} = 0.31, \tag{5.26}$$

where f_{BiPo} is the enriched BiPo distribution shown in figure 2.17.

There is also a shift of the cut on ΔT_{ES-NS}. In S_{BiPo} it becomes $\Delta T_{ES-NS} \in [300, 500]\mu s$. The ΔT_{ES-NS} distribution for BiPo events is a exponential with a time constant of 252 μs. The scaling factor due to this cutm noted $c^{\Delta T}_{BiPo\rightarrow Signal}$, is :

$$c^{\Delta T}_{BiPo\rightarrow Signal} = \frac{\int_{1\mu s}^{141\mu s} f(t)dt}{\int_{300\mu s}^{500\mu s} f(t)dt} = 2.77 \tag{5.27}$$

where $f(t)$ is an exponential decay function with the adequate decay time. The total scaling factor is:

$$c_{BiPo\rightarrow Signal} = c^{\Delta T}_{BiPo\rightarrow Signal} \times c^{BiPonisher}_{BiPo\rightarrow Signal} = 0.86 \tag{5.28}$$

and the rate of BiPo events in the signal region is given by:

$$R^{BiPo}_{Signal} = R_{BiPo-Acc} \times c_{BiPo\rightarrow Signal} \tag{5.29}$$

Finally, from the equation 5.19 two components have been estimated: R_{BiPo} and R_{Acc}. Those two contributions are subtracted from the signal sideband as:

$$R_{Signal-BiPo-Acc} = R_{Signal-Acc} - R^{BiPo}_{Signal} \tag{5.30}$$

Figure 5.32 shows the daily evolution of R_{Signal} and R_{BiPo} and their accidental sidebands. We can see that the level of accidental is higher in the BiPo sideband than in the signal region. This is explained by the origin of delayed events in accidental backgrounds (see section 2.5.3.2): they are mostly from alpha decays (see figure 2.20f). The accidental level in the signal selection is also very low, around 5%, this is explained by the topological selection we applied which mostly select events with two reconstructed annihilation gammas whereas prompt event composing accidental are mostly due to the interaction of a single event. Even if it is not visible on this figure, the accidental rate is higher (33%) in reactor-on period than reactor-off (see table 5.8). This is due to the gamma background induced by the reactor. The comparison of R_{Signal}, R_{BiPo} and $R_{Signal-BiPo-Acc}$ shows that the background, neglecting the accidental one, is composed of around 40% by BiPo background. The rest is the cosmic induced background. The comparison between reactor-on and reactor-off periods of $R_{Signal-BiPo-Acc}$ shows a clear excess of around 90 events per day. This is a qualitative agreement with the performance shown in table 5.5.

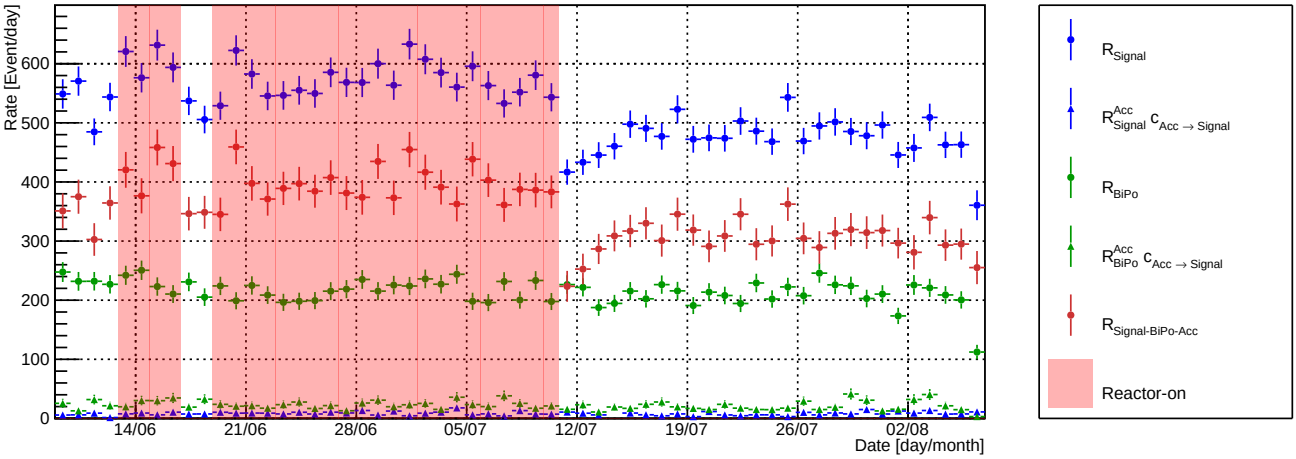


Figure 5.32.: Daily rates of the signal, BiPo and accidental sidebands obtained with the ANN approach and cutting at the working point. The reactor-on days are shown in red.

5.4.1.4. Cosmic induced component

The cosmic induced background is more difficult than the BiPo one to subtract as it has the same time decay constant as the IBD components and it is correlated with the pressure variation. We use sensors at BR2 to monitor environmental variables such as the atmospheric pressure. They are measured at the beginning of every runs. Therefore, it is possible to use the reactor-off and reactor-on datasets to set a predictive model of the cosmic background rate as a function of the atmospheric pressure:

$$f(P_{Atm}) = a \times P_{Atm} + b \tag{5.31}$$

Both reactor-on and reactor-off are exploited to determine this predictive model. The difference between the two datasets is the presence of the IBD component in the reactor-on. It means that in average the reactor-on rate is higher than in reactor-off thus a combined fit using the absolute rate is not possible because the b parameter is different. Instead, we compute the relative rate and pressure in parallel for the two datasets:

$$\begin{aligned} \Delta P_{Atm,i} &= P_{Atm,i} - \overline{P_{Atm}} \\ \Delta R_{Signal-BiPo-Acc,i} &= R_{Signal-BiPo-Acc,i} - \overline{R_{Signal-BiPo-Acc}} \end{aligned} \tag{5.32}$$

where, $\overline{P_{Atm}}$ and $\overline{R_{Signal}}$ are the mean pressure and rate over the full dataset (reactor-on and reactor-off) and i is a 24 hours bin time. Then the scatter plot (ΔR_{Signal},ΔP_{Atm}) is fitted using a linear function to determine the a parameter of the equation 5.31. It returns a measurement of the a parameter equals to -1.86 ± 0.92. Then the share of cosmic background in $R_{Signal-BiPo-Acc}$ can be predicted for the full dataset, knowing the atmospheric pressure in any time bin. It is noted R^{model}_{Cosmic}.

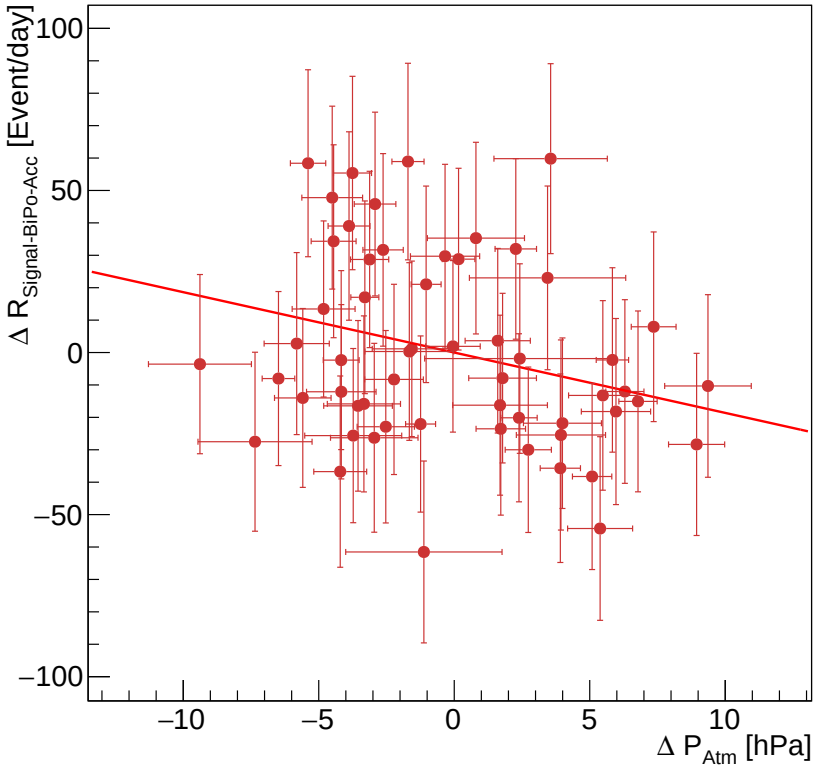


Figure 5.33.: $\Delta R_{Signal-BiPo-Acc}$ as a function of the ΔP_{Atm} for the whole dataset. A linear fit is performed to derive a model of cosmic induced background rates as a function of the atmospheric pressure.

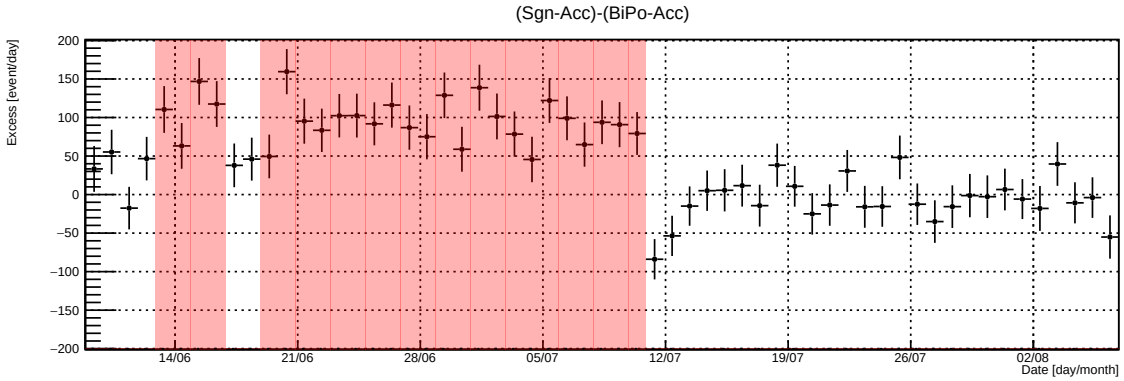


Figure 5.34.: Daily R_{Excess}. A clear excess is seen during the reactor-on period whereas reactor-off days are distributed around 0.

Dataset	Excess [Event/day]
Reactor-on	91.9 ± 5.7
Reactor-off	-4.3 ± 4.6

Table 5.7.: Excesses computed by summing all days available in the dataset for reactor-on and reactor-off periods.

5.4.1.5. IBD Excess in reactor-on days

Finally the excess per bin is computed following the equation:

$$R_{Excess} = R_{Signal-BiPo-Acc} - R^{model}_{Cosmic} \tag{5.33}$$

and the statistical error is computed such as:

$$\Delta R_{Excess} = \sqrt{(\Delta R_{Signal-BiPo-Acc})^2 + (\Delta R^{model}_{Cosmic})^2} \tag{5.34}$$

where $\Delta R^{model}_{Cosmic}$ is equal to the error returned by the fitter on the a parameter. Figure 5.34 shows the daily evolution of R_{Excess}. A clear excess and a reactor-off compatible with 0 are seen while comparing reactor-on and reactor-off days. Summing all day for both periods, the total excess is shown on the table 5.7. The reactor-on excess is in agreement with the prediction shown in table 5.5.

This method was applied for several values of the cut on the ANN response between 0.9 to 0.99. The excesses obtained are shown on figure 5.35. It shows that a very good agreement is observed between the rates obtained from data and the predictions.

We have also derived the distributions of the variables of interest for each background components and the IBD excess, in both reactor-off and reactor-on datasets. For example, we can see on figure 5.36 the distribution in the E_{prompt} variable. As we have already observed in the section 2.5.3.2 the background distribution is dominated by BiPo events below 3 MeV and above by the cosmic induced

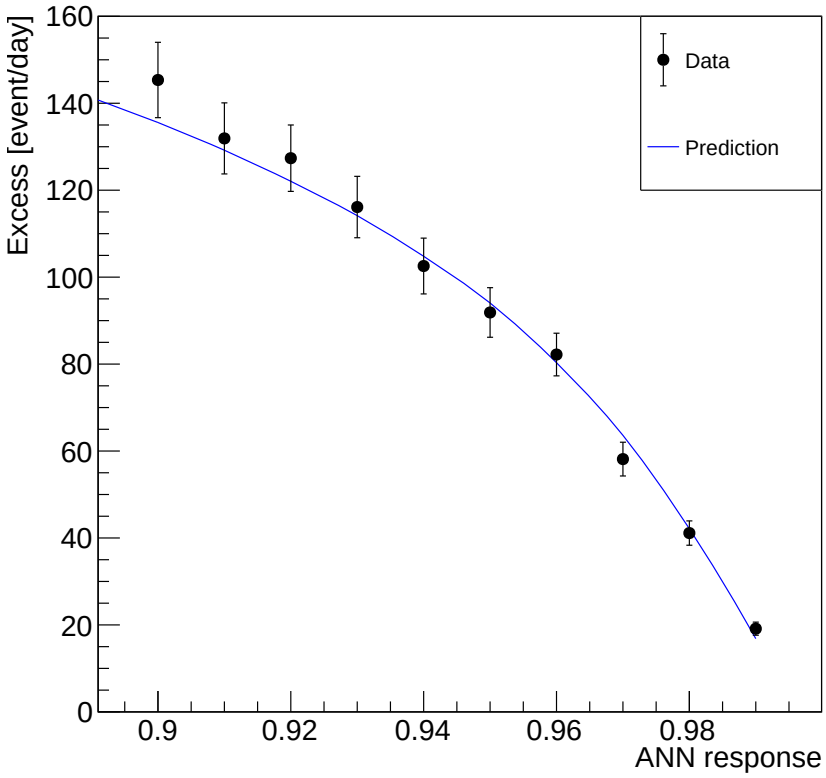
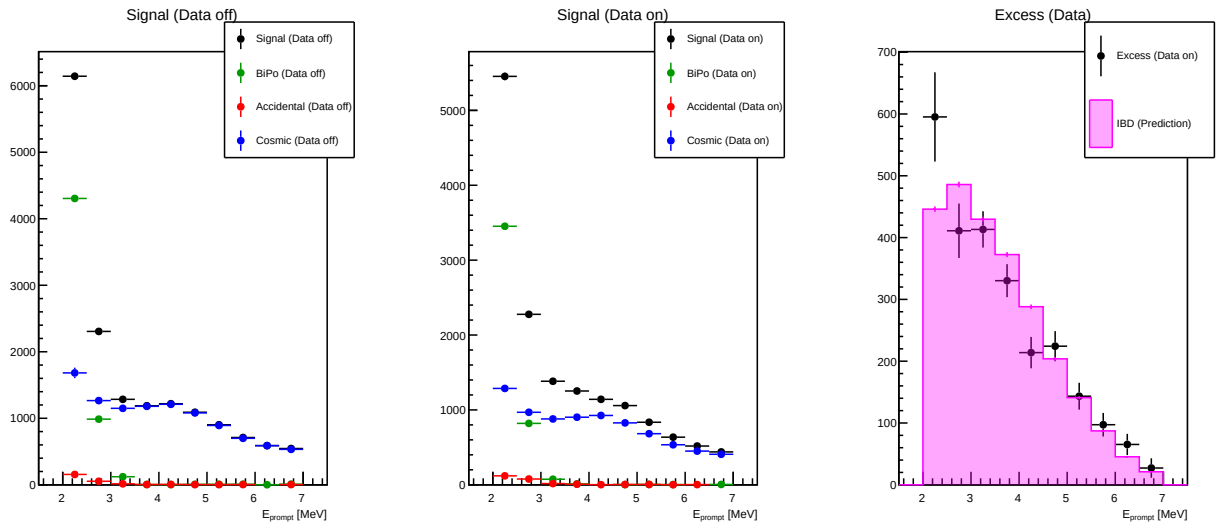


Figure 5.35.: Excesses computed as a function of the cut on the ANN response. The data points are in very good agreement with the prediction (blue line).



(a) Reactor-off. **(b)** Reactor-on. **(c)** Excess.

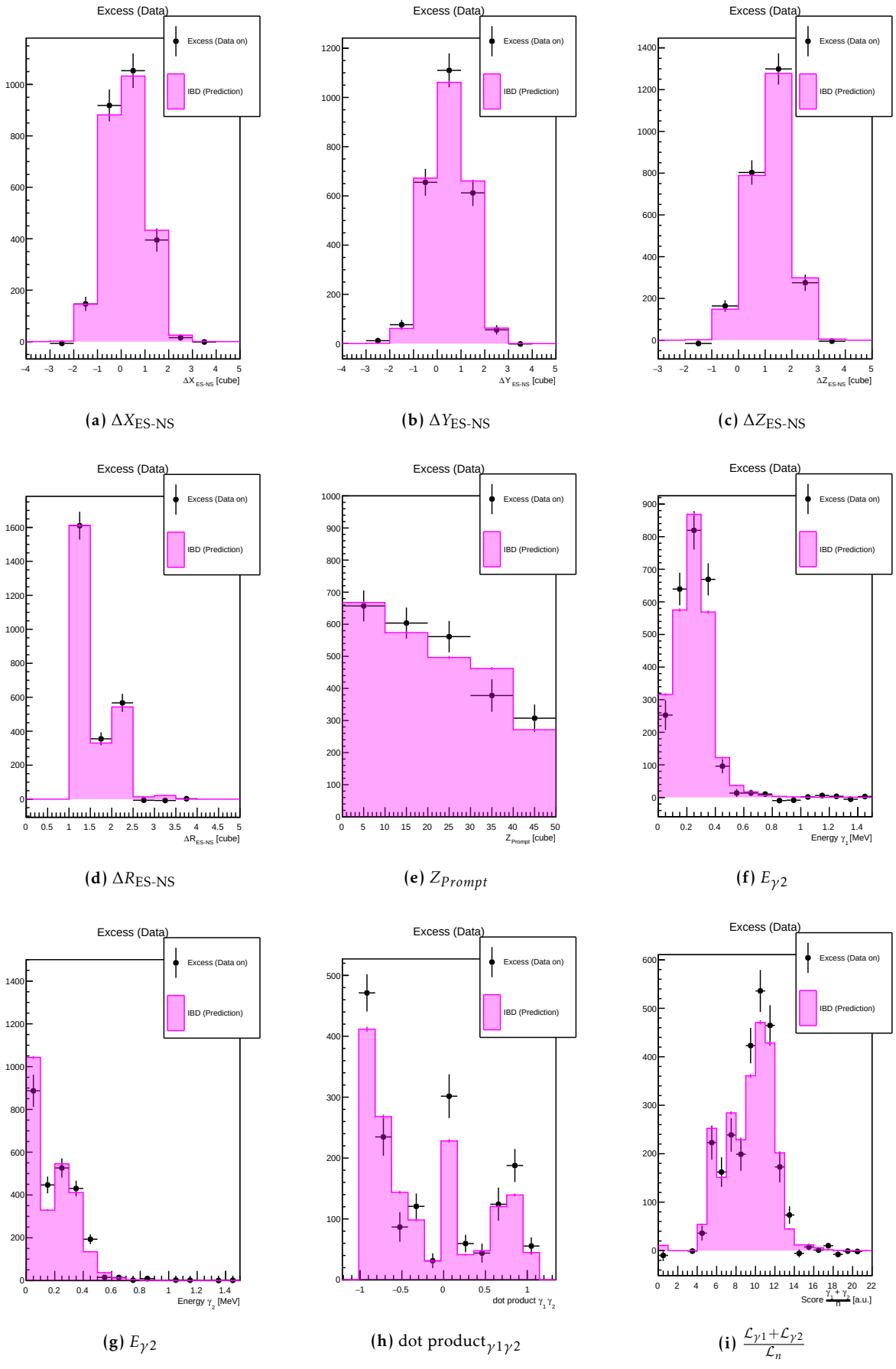
Figure 5.36.: E_{prompt} distribution of the three components (accidental, BiPo, cosmic) using the reactor-off and reactor-on datasets and comparison between the IBD excess in reactor-on with IBD simulation.

one. Figure 5.36c shows the distribution of events in reactor-on data (see figure 5.36b) after subtraction of each background distributions. We can see that the comparison between excess distribution and IBD simulation shows a good agreement. Figure 5.37 shows the distribution of the excess for nine variables of the analysis. We can see that a good agreement is observed with IBD simulation distribution even for topological variables such as the energy of the first and second gamma, the likelihood score ratio and the dot product. In particular, we see on the tails of reconstructed annihilation gamma energies the excess distribution is compatible with zero meaning that the subtraction was well performed. This is a confirmation of the both the ability of the simulation to reproduce the signal as well as the robustness of the subtraction method.

5.4.2. sPlot subtraction method

5.4.2.1. Description

As table 5.6 shows, all components in the reactor-on and reactor-off dataset have a known ΔT_{ES-NS} distribution. From those values, we can build a model describing the ΔT_{ES-NS} distribution in order to perform a fit where only the yield of each components float. From this fit, the sPlot method derives the distributions of the other variables for each components. In particular, we want to determine the dot product distributions since we decided to focus this method on the 2-gamma category. The sPlot technique [77] is a rather common approach to this problem in the particle physics community. It requires to know the number of components in the dataset and to have a discriminative variable. For each species, the p.d.f. describing the distribution of the discriminative variable is required. A fit determines the yield of each species in the dataset. A weight, called sWeight, is computed for each



(a) $\Delta X_{\text{ES-NS}}$ **(b)** $\Delta Y_{\text{ES-NS}}$ **(c)** $\Delta Z_{\text{ES-NS}}$

(d) $\Delta R_{\text{ES-NS}}$ **(e)** Z_{Prompt} **(f)** $E_{\gamma 2}$

(g) $E_{\gamma 2}$ **(h)** dot product$_{\gamma 1 \gamma 2}$ **(i)** $\frac{\mathcal{L}_{\gamma 1}+\mathcal{L}_{\gamma 2}}{\mathcal{L}_n}$

Figure 5.37.: Distribution of IBD excess compared with the IBD simulation of several variables of interest. We observe a good agreement between the excess distributions and the prediction in terms of rate and shape.

event according to the probability of this event to be of a given species. The distributions that we want to determine for a given species are derived weighting the distributions of the full sample with the adequate sWeights. The variables these distributions correspond to are called control variables and should not be correlated to the discriminative variable. In order to illustrate the approach in the following sections, the ANN selection operating at the working point (see section 5.3.2.2.1) has been used.

5.4.2.2. Fit of the ΔT_{ES-NS} distribution

We used the RooFit toolkit [91] included in the ROOT software. It is a c++ library allowing to perform data modelling and statistical study in the ROOT environment. A class is available to use the sPlot technique. In the case of this analysis, the discriminative variable is ΔT_{ES-NS} and all other variables available in the analysis are assumed independent. The p.d.f. describing the distribution of this variable is made of three components:

- **Neutron-capture:** Those events concern the IBD and the cosmic induced background. It is described by a double exponential decay with a fast component ($\tau_{fast} \simeq 10\mu$s) and a slow component ($\tau_{slow} \simeq 64\mu$s). The p.d.f. is then:

$$f_{neutron}(t) = N_{neutron} \times \left(R \times \exp(-\frac{t}{\tau_{slow}}) + (1-R) \times \exp(-\frac{t}{\tau_{fast}}) \right) \tag{5.35}$$

  where $N_{neutron}$ is the neutron yield, R is the ratio between the slow and fast component.

- **BiPo:** Those events are described by a simple exponential decay with a decay time constant τ_{BiPo} equals to 236.9 μ s. Thus the p.d.f. is given by:

$$f_{BiPo}(t) = N_{BiPo} \times \exp(-\frac{t}{\tau_{BiPo}}) \tag{5.36}$$

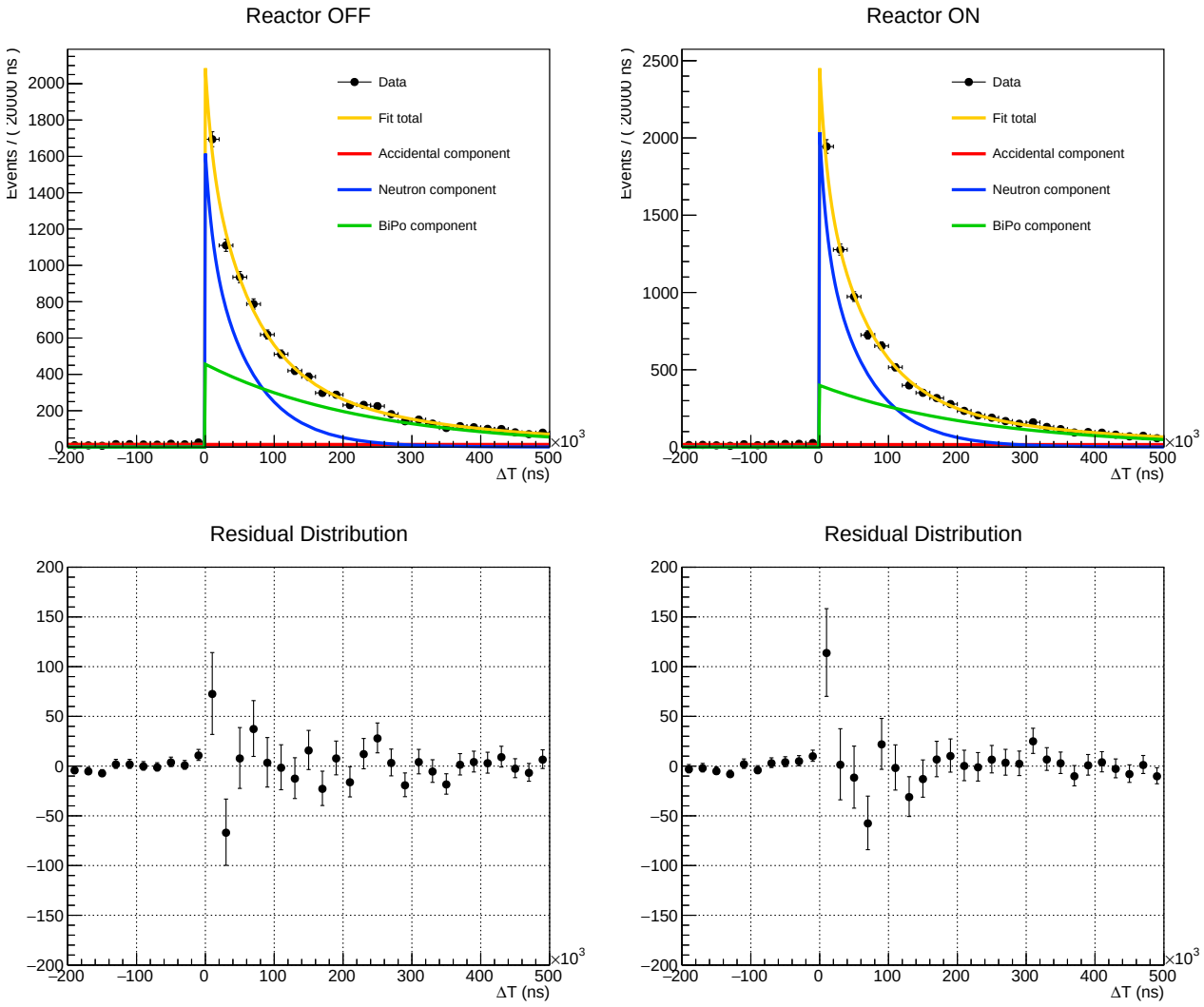
  where N_{BiPo} is the BiPo yield.

- **Accidental:** The last component is the accidental background. It is described by a simple flat p.d.f.:

$$f_{Acc}(t) = N_{Acc} \tag{5.37}$$

  where N_{Acc} is the yield of accidental events.

Thus a global p.d.f. is built by summing those three components and is fit to the reactor-off and reactor-on ΔT_{ES-NS} distributions. The fitted parameters are the yields of each component and the ratio R between the slow and the fast neutron component. An example of the fitted distribution is given in the figure 5.38. There is a small deviation in the negative part of the distribution. The model does not take into account correlated events, mostly coming from the cosmic induced background, in this region. But the residuas show the contamination is negligible (bottom figures of figure 5.38).



(a) Reactor-off distribution.

(b) Reactor-on distribution.

Figure 5.38.: Distribution of reactor-off and reactor-on coincidences using the ANN approach cutting at the working point. The fit determining the yield of the three components are shown and its residuals for the two datasets.

Fitted parameter	Reactor-off	Reactor-on
N_{Acc} [event/day]	21.7 ± 1.8	28.8 ± 2.2
N_{BiPo} [event/day]	205.7 ± 5.8	211.5 ± 6.7
$N_{neutron}$ [event/day]	173.3 ± 5.2	244.6 ± 6.2
R	0.94 ± 0.01	0.93 ± 0.01

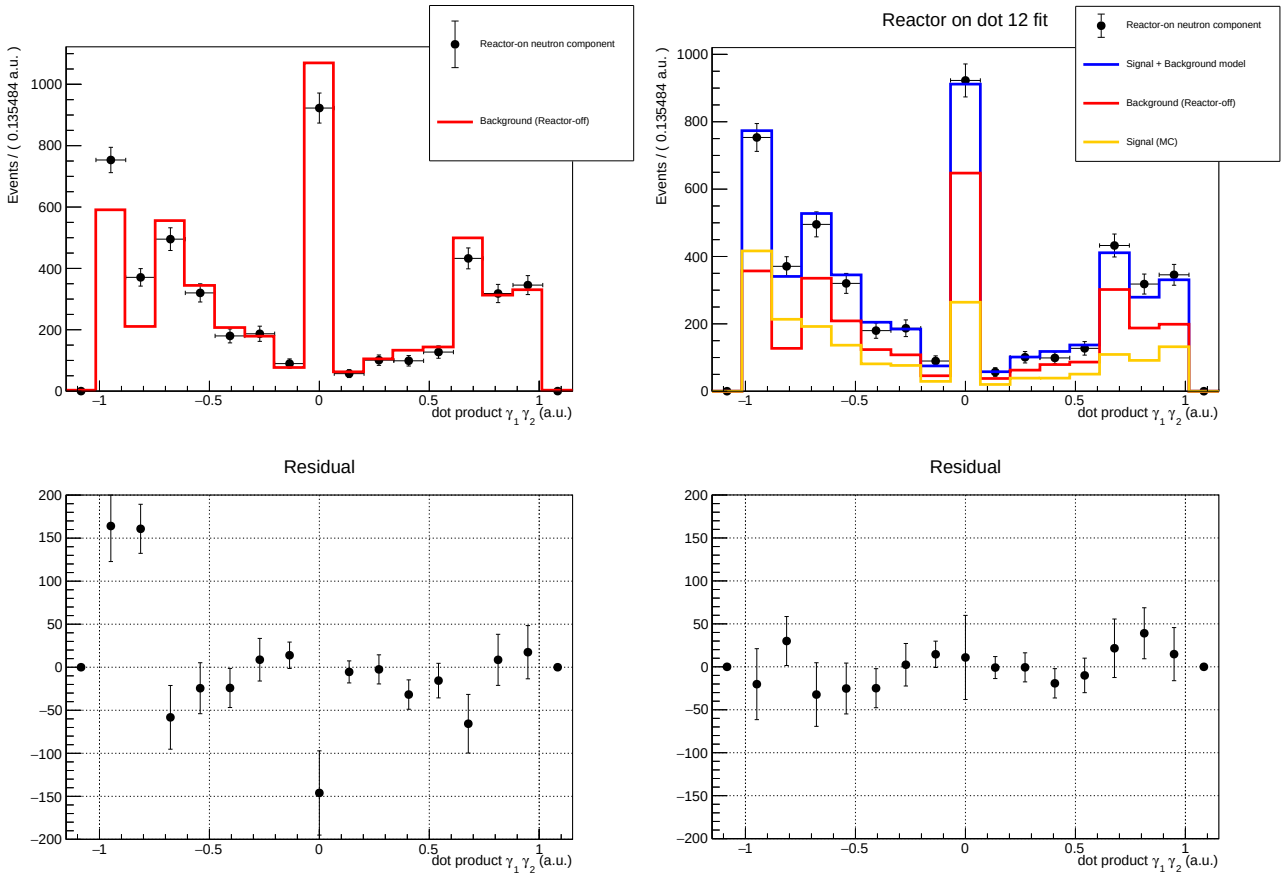
Table 5.8.: Table showing the value and error obtained by fitting the reactor-off and reactor-on ΔT_{ES-NS} distribution.

The fitted values obtain from those two fits are shown in the table 5.8. The errors are directly obtained from the fit method. It is interesting to see that the R and N_{BiPo} fitted values are in agreement in reactor-off and reactor-on dataset. This is expected as the reactor status does not impact the BiPo yield nor the neutron model. We see that the accidental and neutron yields increased in the reactor-on, again this effect is expected. Indeed the accidental background should increase because of the reactor gamma background drastically increases and in the neutron component there is the IBD component. An overall difference in atmospheric pressure could also explain such difference. For this reason we will perform a comparison of the dot product$_{\gamma 1 \gamma 2}$ reactor-off and reactor-on distributions to highlight the presence of the IBD component without taking into account of pressure variation.

5.4.2.3. Excess extraction

Once the ΔT_{ES-NS} distribution of the two datasets has been fitted, the sWeight variables are computed by the algorithm. In particular, it gives the possibility to perform a fine comparison between data and simulation using the BiPo component. This comparison is shown in section 4.3.3. But in the case of the signal extraction, the interest is to obtain the neutron component distribution with regards to other variables of the analysis. Indeed in the reactor-off neutron component there is only cosmic induced background contribution whereas in reaction-on one there are IBD and cosmic induced events.

As seen in section 5.3.2.1 the dot product variable is a discriminative variable of the analysis and is sensible to the back-to-back behaviour of annihilation gammas from IBD events. Thus the goal is to fit the distribution of this variable in the reactor-on sample in order to separate the cosmic induced and signal yields. The reactor-off dataset gives the best measurement of the cosmic induced background distribution as there is no trustable simulation available. Concerning the signal, the IBD simulation is used since the comparison between data and MC for topological variables seems good, as shown in section 4.3.3. Figure 5.39 shows the two fits performed on the reactor-on neutron component. On the left (5.39a) the fitted model is built using only the background derived from the reactor-off dataset. The residual shows a deviation of several standard deviation in the back-to-back region (dot product$_{\gamma 1 \gamma 2} = -1$). Thus the dot product shape of the neutron reactor-on component could not be explained by the neutron reactor-off component. The right part of the figure shows the fit when the



(a) Background only model.

(b) Signal and background model.

Figure 5.39.: Distribution of the reactor-on neutron component extracted with the sPlot technique and fitted with the background model only (left) and the background plus signal model (right). The background distribution is the neutron component extracted with sPlot technique using reactor-off dataset. The signal distribution is from MC.

Parameters	Fitted values
IBD component [event/day]	84.8 ± 11
Cosmic induced component [event/day]	130.4 ± 12

Table 5.9.: Fitted parameters values from the fit shown by the figure 5.39b. The associated errors are directly given by the fitter but do not take into account of model statistical fluctuation.

dot product distribution derived from IBD simulation is added to the fitted model. Now the model fully describes the reactor-on distribution.

Finally, the fitted parameter values returned by the fitter are shown in table 5.9. The errors estimated by the fitter are not correct as the uncertainties affecting the histograms that play the role of the *p.d.f.*'s in the fit are not accounted for. We discuss next section about them.

Parameters	Values
Predicted excess [event/day]	58.6
Mean toy excess [event/day]	74.2
Std Deviation toy excess [event/day]	17.1

Table 5.10.: Results obtained from the toy study of the sPlot approach using the ANN selection and cutting at the working point.

5.4.2.4. Bias and errors

The fit gives an estimation of the signal and cosmic induced background yields in the selected dataset. The issue is to have a correct estimation of the error associated to those numbers. Indeed, uncertainties naturally affect the background histogram extracted with the sPlot method, since this method can't be perfect. RooFit provides a method allowing to perform binned fit with binned model taking into account the statistical uncertainties of each bin of distributions. This method is called the Barlow-Beeston approach [92]. But there is an issue in the implementation, the associated error to each bin is computed assuming the usual Poissonnian error and so overwriting the error computed from sPlot. Also it was interesting to determine if the signal yield derived with this approach is biased or not. Therefore, instead of propagating the error by hand it has been decided to perform a toy study of this method.

The idea of the toy study is to reproduce the signal extraction many times in order to evaluate the precision and bias which the assumptions used to generate the pseudo experiments (like the assumed signal yield) are retrieved by the selection procedure. Here, pseudo reactor-on and reactor-off samples are generated. Each time, ΔT_{ES-NS} distributions are produced, as the sum of the distributions of the various components of these samples (signal, BiPo, etc.). The same is done with dot product$_{\gamma 1 \gamma 2}$ distributions. The models to use for that is obvious in the case of ΔT_{ES-NS} (the same as the one used on figure 5.38). It is not in the case of dot product$_{\gamma 1 \gamma 2}$. For the BiPo and the IBD components, we use the simulation. Concerning the accidental and the cosmic induced background the distribution extracted from reactor-off dataset with the sPlot technique is used.

Once 50 pseudo samples have been generated, the sPlot technique is applied on each of them to derive sPlot-ed distributions of dot product$_{\gamma 1 \gamma 2}$, and the fit to these distributions to determine the IBD yield is performed, just like in the previous section.

Results are shown in the table 5.10 and figure 5.40 shows the fitted excesses for 50 toys. So for this selection, the excess is biased of 27%. Also the error is clearly underestimated by the fitter as expected.

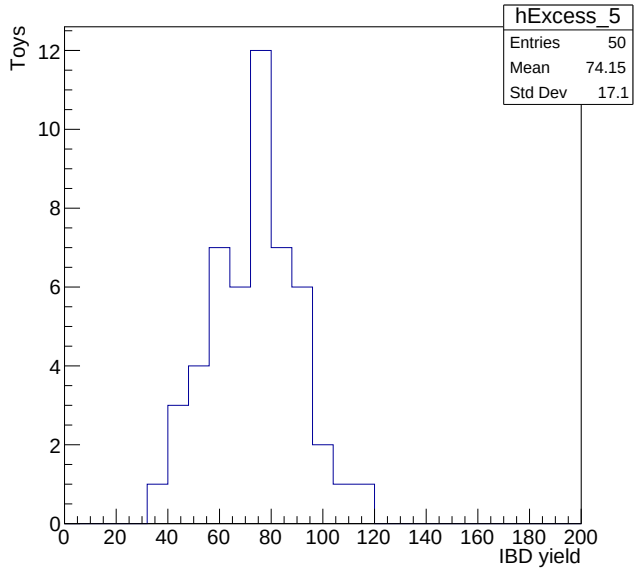


Figure 5.40.: Excesses obtained from the 50 toys reproducing the signal extraction with the sPlot method and cutting on the ANN response at the working point.

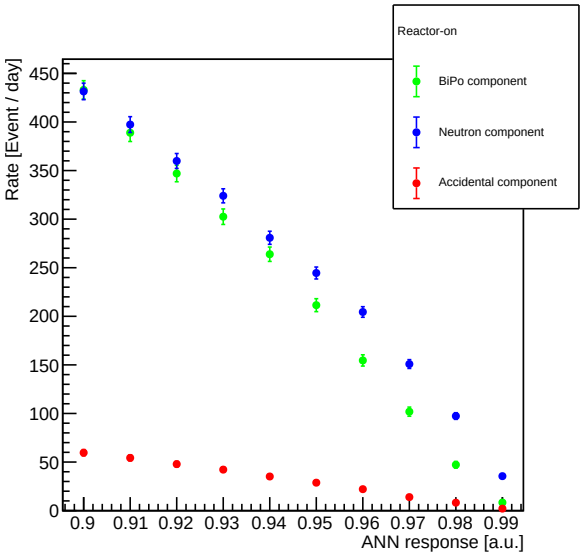
5.4.2.5. Excess

The method detailed in the previous section has been performed for several cut values of the ANN analysis, from 0.9 to 0.99. This section aims to show the stability of the method with regards to the ANN response and thus is a good check of the subtraction method and simulation ability to predict the signal yield and distributions.

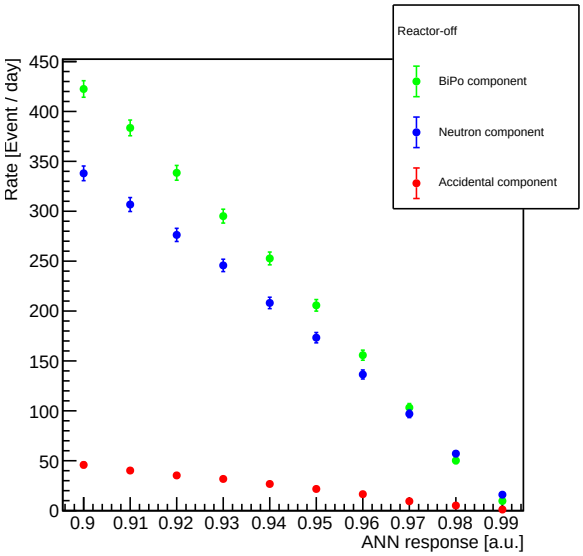
Firstly, the rate of the three fitted components for reactor-off and reactor-on have been computed and compared. This is shown by the figure 5.41. Figures 5.41a and 5.41b show that the accidental background is not dominant as expected. The two correlated backgrounds, BiPo and cosmic induced, have roughly the same rates after applying the signal selection. It is interesting to notice that the neutron components rate ratio shown by figure 5.41c is the only component statistically increasing with regards to the cut on the ANN response. This is expected as this component is the sum of IBD plus the cosmic induced background. Thus the fraction of the signal should increase and the fraction of background should decrease as far as the cut on the ANN response increases.

The dot product$_{\gamma 1 \gamma 2}$ fit stability has also been checked. As said, two models are tested: with only the background and with the background plus the signal. Figure 5.42 shows the reduced chi square of those two fits. As shown, the fit using the model where the IBD component is added to the cosmic background always gives a better reduced chi square and is distributed around the unity.

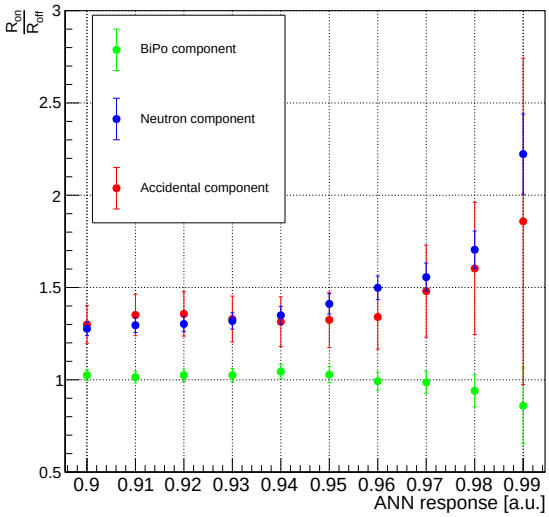
Taking the bias and errors discussed in section 5.4.2.4 into account one can correct the results of data obtained in table 5.9. Finally, table 5.11 gives a global summary of the excess measured with the sPLot technique when the nominal cut on the ANN response is appplied. After corrections for the bias



(a) Absolute rate of each component in reactor-on



(b) Absolute rate of each component in reactor-off



(c) Ratio between reactor-on and reactor-off rate.

Figure 5.41.: Extracted rate of each component (BiPo, Neutron and Accidental) using the fit of the ΔT_{ES-NS} distribution for reactor-on (a) and reactor-off (b) data as a function of the ANN response. Ratio R_{on}/R_{off} (c).

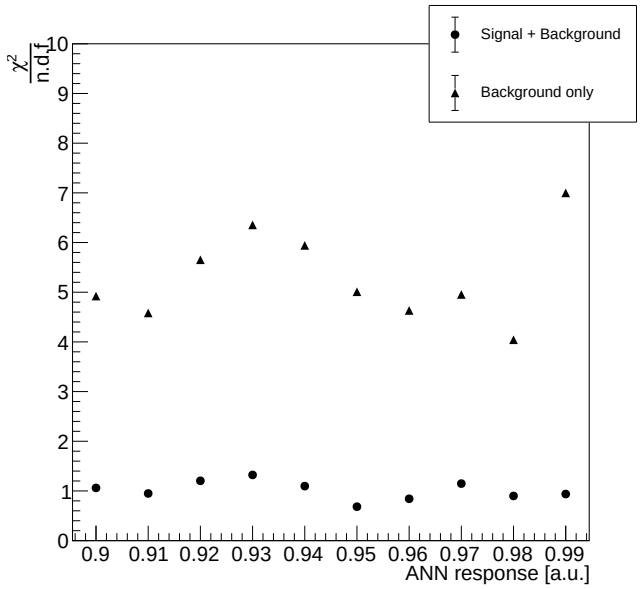


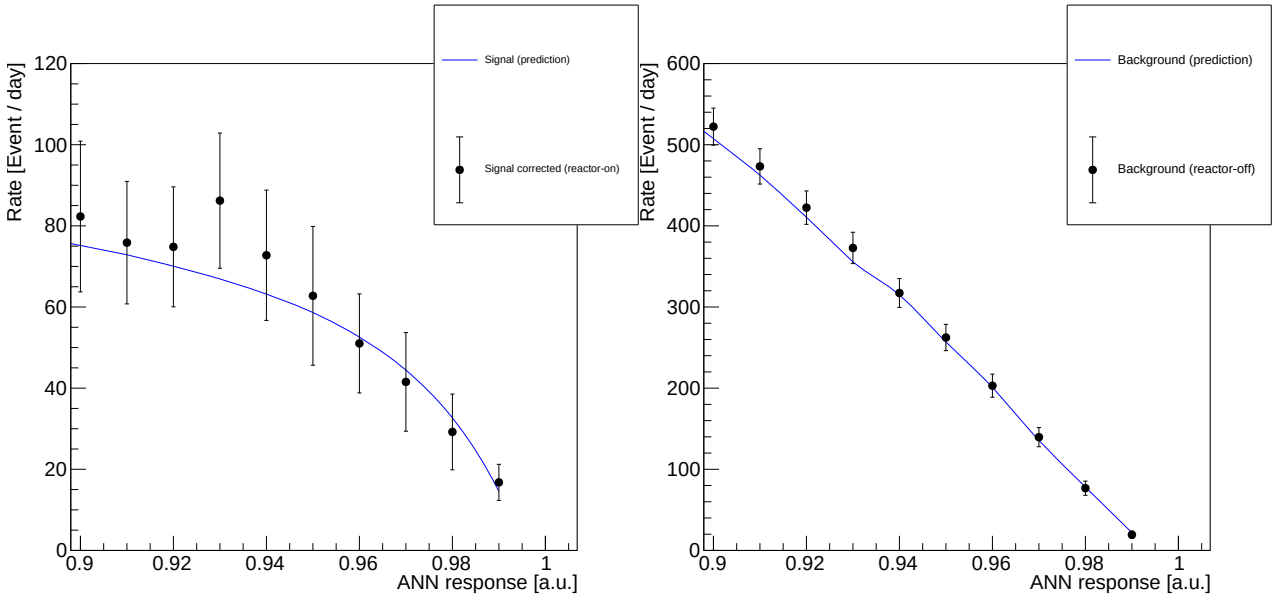
Figure 5.42.: Reduced chi square of the dot product$_{\gamma 1 \gamma 2}$ variables with the background only model (triangle) and the background plus signal model (circle).

Parameters	Values
Fitted Excess (Data) [event/day]	84.8
Toy bias [%]	26.5
Toy error [event/day]	17.1
Corrected Excess (Data) [event/day]	62.2
Predicted Excess (MC) [event/day]	58.6

Table 5.11.: Summary of the excess measurement in the reactor-on sample, and bias corrections. The nominal cit on the ANN score was used here.

determined with the toy study in the previous section, the measured excess is in agreement with the prediction.

Finally, we studied the variation of the measured background and signal rates as a function of the cut on the ANN response. We compared it with the predicted ones. For each cut value, the toy method discussed in section 5.4.2.4 has been performed in order to determine the error and the bias. Figure 5.43 shows this comparison. For the background the reactor-off rates have been used. Both IBD and background rates are in good agreement with the predictions. It gives confidence in the various approaches used in this chapter to derive the IBD yield.



(a) Excess rate.

(b) background rate.

Figure 5.43.: Excess and background rates compared with the simulation prediction. The signal rate has been corrected from the bias determined by the toy analysis and the associated errors.

5.4.3. Results and comparison with predictions

We have seen in this section that two methods have been developed in order to determine the rate and distributions of IBD events in the reactor-on dataset. Extracted rates have been compared with the IBD simulation and we have seen that there are in agreement giving us confidence on the validity of the two subtraction methods. However, the sPlot method is less precise than the first method because it requires to consider only 2-gamma events. For this reason, we have used the method presented in section 5.4.1 to derive result for the ICHEP-2020 conference. Finally, we have added on the figure 5.30 the results obtained from data. This is shown on figure 5.44. We can see that for all the selections presented in this manuscript, performance from data and simulation are in agreement. This is the case, in particular, of the two multivariate analyses. This figure also confirms that they also multivariate reach similar performance. It also shows the improvement brought by the usage of topological variables, and the improvement brought by the multivariate approach. For the same daily IBD rate, the S/B ratio is improved by about almost a factor 4. This chapter is also a great confirmation of the quality of the work performed in chapter 3 and 4, since the ability of the simulation is crucial to optimise the selection and control the subtraction. In particular, we have seen with the first subtraction method in section 5.4.1.5 that the IBD excess distributions from data are in agreement with the prediction, even for complex topological variables. This is also confirmed by the study using the second subtraction method, which is based on the ability of the IBD simulation to predict the dot product$_{\gamma 1 \gamma 2}$ distribution.

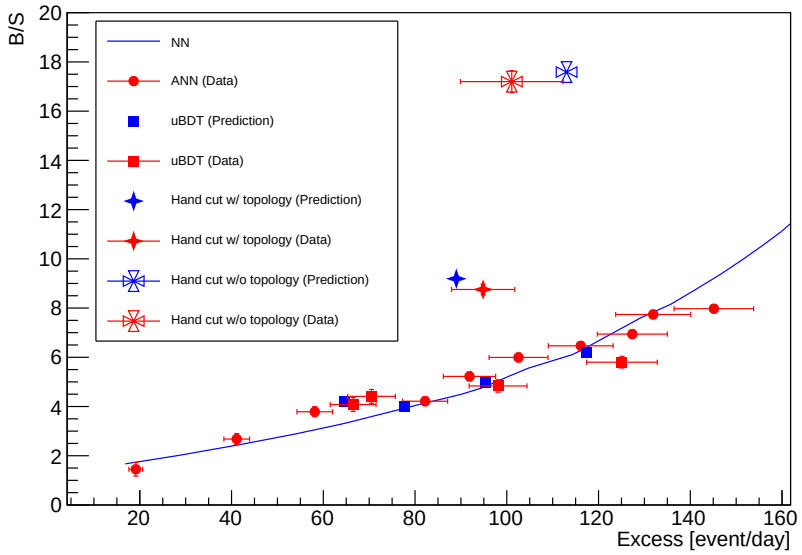


Figure 5.44.: Background over signal ratio as a function of the excess of all selections: Without topology, with topology and with multivariate tools. The prediction (blue) and data (red) are in good agreement for all approaches.

5.5. Conclusion

The SoLid experiment has to fight against important and partially unexpected background rates. The development of the BiPonisher variable allowed to decrease significantly the BiPo background. The standard selection presented in section 2.5.3 shows the status of the analysis at the beginning of the work presented in this chapter. The performance reached in terms of signal rate and signal over background ratio was too low to perform an oscillation analysis.

The solution we proposed is to take advantage of the annihilation gammas resulting from the positron. This is a way to fully exploit the high segmentation of the detector. To do so, we developed a new way to reconstruct of a prompt ES event, that associates together the desposits due to annihilation gammas. It also includes lower energy deposits. The potential gain therefore relies on the ability of the simulation to correctly reproduce the detector response not only to high but also to low energy deposits. This reconstruction is based on the theoretical description of gamma interaction in matter. The variables derived from this reconstruction improve a lot the discrimination power, hence the performance of the selection.

To highlight these gains, we have first designed a selection applying rectangular cuts and compared its performance with that of the standard selection. Then, to fully exploit the new variables, we developed another selection based on a Artificial Neural Network. When they are tuned to maintain the same daily signal rate (~100 IBDs), these two selections reduce the background rate by about a factor 2 or 4, respectively.

The best result above (with the ANN) is confirmed by another selection, based on a uniform BDT and developed by the SoLid Imperial College group. We led the validation tests that proved this. The agreement between the two methods is an important proof of robustness of the result. Moreover, for even more robustness, the performance of both selections was determined in two ways: when the signal rate is computed based on the simulation and on the expected number of IBD interactions in the detector before selection, and when this rate is measured directly in data.

To measure the signal rate in data, we first reproduced the official subtraction method. Then we developed an approach based on the sPlot technique in order, again, to have two different methods cross-checking each other. Finally, the official subtraction method provides less statistical errors and thus has been used to estimate the excess and the signal over background ratio for all selections. Figure 5.44 summarizes our results, id est shows the daily IBD rate and B/S ratio for all the selections and all the evaluation methods.

This chapter was also the occasion to compare the distribution of analysis variables in data and simulation. It had been done in chapter 4 with BiPo event. Thanks to the measurement of the IBD excess performed in this chapter, we could this time do for IBD events. There are no clear deviation between the extracted excess and the IBD simulation, which also gives us more confidence on the reliability of the official subtraction method.

We conclude from this work that:

- The development of annihilation reconstruction gives more discrimination. With it, the high segmentation of the target is fully exploited.
- Multivariate tools also improve crucially the discrimination.
- The simulation which has been tuned and controlled in chapter 4 reproduces sufficienlty well the detector response to perform this study, even using multivariate approaches.

New improvements are foreseen by the collaboration and such studies will benefit of them. The first one is the development of a 1-dimensional Convolutional Neural Network (1D-CNN) in order to improve the BiPonisher discrimination. Instead of using a simple integral ratio of waveforms, it fully exploits the differences in the shape of waveforms. Preliminary tests suggest a reduction of the BiPo background by a factor 3. The second one concerns the upgrade of the detector. New MPPCs were installed during the summer 2020. This new generation has better performance, especially in terms of PDE, which directly impacts the detector light yield. Indeed tests indicate an increase of 40%, it means a better efficiency to see low energy deposits and so more events in the 2-gamma category, and better reconstructed, which gives the hope of a stronger discrimination.

Conclusion

This manuscript presented my contributions to the SoLid experiment. I presented in chapter 1 the context of this experiment: a world wide effort to perform very short baseline measurements of electronic antineutrinos produced by nuclear reactors, in the hope to discover a new state of matter called sterile neutrino. I also explained that such experiments would help improve spectral predictions of reactor neutrinos, that suffer various issues.

The SoLid detector is the result of a innovative detection approach. The choice of a novel technology implies an important work to understand the detector response. This thesis has contributed to the development of several new tools and methods, necessary to reach in this respect the necessary level for a neutrino oscillation measurement.

In the third chapter, I presented the goals and challenges of the energy calibration of this detector. I want to highlight the high number of calibration constants, 19,200, that we had to determine and understand. I worked on a new approach to measure cube light yields using the energy spectrum of Compton interactions. While I was in parallel working on data/MC studies to tune the readout simulation, I showed that calibrating only at cube level is not enough. Because any SoLid cube shares half its WLS fibres with 15 other cubes, it cannot be understood if the other cubes and fibres in the same plane are not well described too by the simulation. For that reason, and to ensure a proper treatment of inhomogeneities in the detector, I also measured for all fibres the attenuation length and the optical efficiency of the fibre-to-SiPM coupling. I contributed to a first method for this measurement, and developed a new one. With the first method, the measurement of the features of a fibre depended on the features of the other fibres of the plane. By using asymmetries between the signals seen by the 4 fibres that readout a cube, my method reduced this dependence and stabilised measurements. This allowed to eliminate biases on some of the attenuation lengths when measured by the first method, and to reduce the uncertainty on these parameters from about 8 to 4%. Once fibre attenuations and couplings have been corrected for, the measured cube light yields show over a full module a dispersion of 3%, that constitutes a conservative upper limit on the precision with which we perform this measurement. Tests on simulated data confirm that.

I implemented in the model used by the readout simulation the attenuation lengths and coupling efficiencies measured above. Also, from the cube light yields measured in calibration runs, I derived the yield of scintillation photons to be used in this model. This, plus an extensive work to check and correct the simulation code, led to a satisfactory data/MC agreement. It was studied first with the simplest sample we had, namely the ^{22}Na calibration dataset. A lot of steps proved necessary until we

reached a satisfactory level of agreement between data and simulation. It took to understand various effects, including unexpected effects in the DAQ. Strategies to improve source data taking were also developed. In the end, based on these data/MC comparisons, we conclude that we understand at better than 3% the energy scale at cube level. Studies by the SoLid oscillation group conclude that this is enough to not deteriorate much the exclusion contours SoLid can establish, given the present performance of IBD selection. The quality of the simulation was also carefully verified using BiPo control samples. Not only the energy reconstruction but other variables used in the IBD selection algorithm are well reproduced. I used two methods, including an original one, to isolate these control distributions. In most of the range of these distributions, data and MC differ by less than 5%, and 10% in some tails. The efficiency with which low energy deposits are reconstructed was also studied there. Again a satisfactory data/MC level was observed, giving the collaboration confidence on the ability of the simulation to reproduce the detector response in this regime. This was determinant since the new signal selection I developed is based on the reconstruction of such low energy deposits.

I finally contributed to the analysis by designing a reconstruction and identification method for annihilation gammas and neutrons. It implies to use small energy deposits. I proved that topological approaches based on the energy and spatial distributions of low energy deposits improve the selection performance. Indeed for a same signal efficiency we reject about 4 times more background events. To reach this, we also trained a Neural Network based on these new variables. Moreover, I led the careful comparison with a similar method developed Imperial's college, which used a different topology reconstruction, and a different multivariate tool. These results were computed first based on the simulation, then based on the measurement of the IBD rate in real data. I worked on two methods to evaluate this rate and to subtract background contributions. With all these methods, we find similar results. This confirms the robustness of these results. I presented them at the ICHEP-2020 conference: we obtain a signal over background ratio of 0.2 for a signal efficiency of 8%, corresponding to a daily rate of the order of 100 neutrinos.

The next step of the collaboration is to analyse the full dataset already taken with SoLid Phase-1 in order to derive a first exclusion contour. This requires to evaluate and propagate all the systematic uncertainties on the measurement, especially the ones coming from the simulation, which is the continuity of my work. This important work is ongoing in the collaboration and first results are expected soon. Thanks to the improvement brought by the topological selection, we might be for the first time in position to exclude the most probable values of Δm^2_{14} and θ_{ee} according to the data of the experiments that found the reactor anomaly. To be more competitive with the best very short baseline experiments, an upgrade of the detector has been performed during the summer 2020. It consists in replacing the MPPCs by a new generation which improves the light collection by 40%. With this upgrade, the collaboration expects to have a higher reconstruction efficiency of low energy deposits, hence to improve the selection performance based on the ability to reconstruct annihilation gammas. Along with the calibration methods, it is one of the legacies of my thesis work.

Appendix A.

Proposal to extract the cube light yield (< 600 keV)

The ability of the SoLid detector to reconstruct small energy deposits is an important ingredient in the search for a sterile oscillation. For instance, the necessary discrimination between IBD events and various backgrounds might be achieved only by using the event features due to the presence of 2 back-to-back annihilation gammas (see chapter 5). Therefore, the ability to reconstruct such gammas and to determine accurately their energy is important. They rarely deposit more than 340 keV (corresponding Compton edge) in a cube. Also, the positron created by the IBD interaction often deposits its energy (before it annihilates) in 2 adjacent cubes. The energy deposited in one of them is often small (a few hundred keV) but has to be measured to reconstruct accuratly the neutrino energy spectrum the oscillation fit relies on. It is therefore important to understand the detector's response to low energy deposits, and, in particular, to verify the linearity even in this region, which has not been achieved yet (see section 3.7.3). It is particularly difficult to determine cubes light yields at low energy. Naively, one could employ the same method as in section 3.5: the convolution of the true Compton scattering energy spectrum with a gaussian. However, with a typical $LY_{rec} = 96$ PA/MeV, a 174 keV energy deposit (average deposit of an annihilation gamma) produces typically 4 PAs per fibre. Fluctuations of the PA yield per fibre - at the origin of the resolution that we account for with a convolution product - must therefore be treated with Poisson statistics when calibrating low energy deposits. As in section 3.5, the efficiency of the calibration event selection must be accounted for, in addition to the convolution, as it varies as a function of the energy deposit. This treatment is particularly important in the case of low deposits, since the efficiency is particulary low, distorting a lot the original spectrum. This can be understood from figure A.1: cutting out events below the 2.5 PAs threshold used in recent SoLid analyses has an efficiency of 80%, meaning that only 40% of cubes that received 174 keV are selected when 4 fibres have to satisfy this criterion.

In this section, we propose a method based on the Poisson distribution that treats these issues simultaneously, and from first principles, in order to obtain a relevent PDF to fit to the measured energy spectrum of low energy photons (annihilation gammas from ^{22}Na and ^{137}Cs), to determine LY_{rec}'s at this scale. Note that simulations could have been used to evaluate the efficiency part. But in this case, the high dependence of the efficiency on LY_{rec} would oblige to generate a large number of simulations, changing each time the readout simulation (for each new evaluation of the light yield).

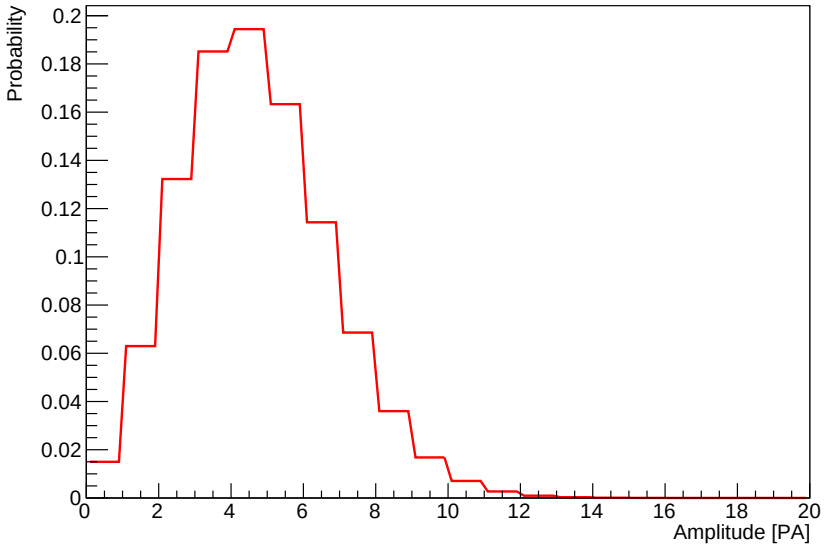


Figure A.1.: Poisson distribution for a mean value of 4.2 PAs

A.1. Fibre amplitude

The first thing to do is to build the probability to detect k PAs on a MPPC assuming a deposited energy E_{dep}, one can write the equation A.1.

$$P_{Fibre}(X=k) = \sum_{i \leq k} P_{Poisson}(X=i, E_{dep} \times LY_{fibre}) \times P_{Binomial}(i, k-i, \mu_{CT}) \tag{A.1}$$

where:

- LY_{Fibre} is the number of captured photons in fibre per MeV. It depends on the attenuation length, the coupling efficiency and the cube light yield.
- E_{dep} is the deposited energy in MeV.
- μ_{CT} is the crosstalk probability that a PA induces another one in a neighbouring cell of the same SiPM.

Thus the probability is not a simple Poisson distribution due to the crosstalk effect. When a scintillation photon reaches a cell of the MPPC a photon avalanche is triggered and there is a probability to trigger a neighbouring cell as well. Meaning that two PAs are detected instead of one. This is the role of the Binomial law which returns the probability to success $k-i$ times among i trials with a probability of success given by μ_{CT}.

Number of fibres with at least 1 PA	Combinations
1	{5,0,0,0}
2	{4,1,0,0} ; {3,2,0,0}
3	{3,1,1,0} ; {2,2,1,0}
4	{2,1,1,1}

Table A.1.: All the partitions of 5 with 4 integers ordered by the number of integers not equal to zero.

Now it is needed to express the first term as a function of the cube visible light yield which is not corrected from the crosstalk. Assuming that the four fibres see on average the same amount of energy, the previous equation becomes A.2.

$$P_{Fibre}(X=k,E_{dep}) = \sum_{i\leq k} P_{Poisson}(X=i, \frac{\frac{LY_{rec}}{4}}{1+\mu_{CT}} \times E_{dep}) \times P_{Binomial}(i,k-i,\mu_{CT}) \tag{A.2}$$

A.2. Cube amplitude

Now that the probability law of a single fibre has been determined, it has to be generalised to the cube amplitude. A cube is read out by four fibres thus all fibre combinations that lead to the given total numbers of PAs have to be taken into account. For example, the table A.1 is given the list of combinations that gives a cube amplitude equal to 5 PAs.

The decomposition of an integer in a sum of integers is a known problem and several algorithms have been developed to return all possible cases, called set of partitions. Ideally, one has to pay attention to the different permutations for each partition since not all channels respond the same way (different couplings, attenuation lengths, etc...). But due to the time needed to evaluate each partition this would drastically increase the CPU time needed to perform one fit. The number of partitions with four integers as a function of the cube amplitude is given in the figure A.2.

Let's defined $\mathbb{A}_{4-Fibres,Selected}$ the set of partitions with four integers that pass a selection, in this case the four fibres should be above 2.5 PAs. The probability to see k PAs in the cube for a deposited energy E_{dep} is given by the equation A.3.

$$P_{Cube}(X=k,E_{dep}) = \sum_{\forall p \in \mathbb{A}_{4-Fibres,Selected}(k)} \left(\sigma(p) \times \left(\prod_{i=1}^{4} P_{Fibre,i}(X_i=k_i,E_{dep}) \right) \right) \tag{A.3}$$

where:

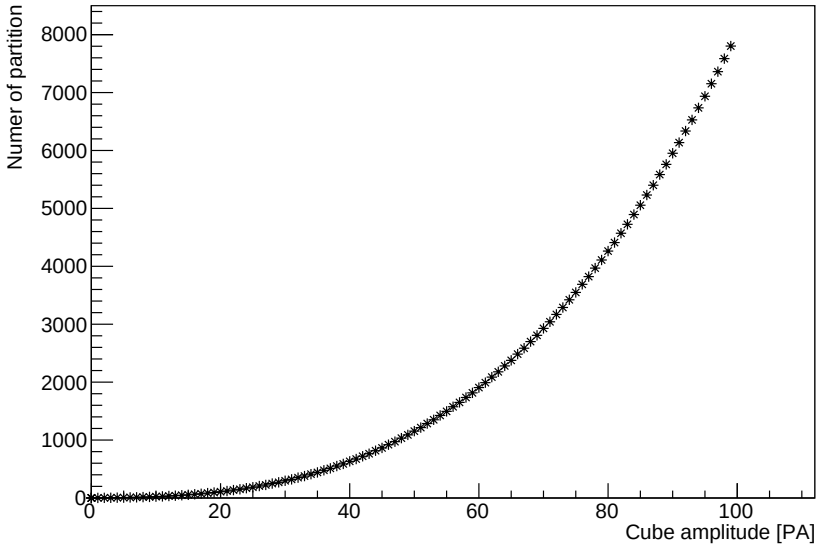


Figure A.2.: Number of partitions with four integers as function of the cube amplitude

- p is a partition from the set $\mathbb{A}_{4-Fibres,Selected}(k)$ for which the sum of integers in equal to k, the cube amplitude.
- $\sigma(p)$ is the number of permutations of the given partition p.
- $P_{Fibre,i}(X_i = k_i, E_{dep})$ is the fibre probability given by the equation A.2.
- E_{dep} is the deposited energy distribution, the fit is using the Klein-Nishina formula as in the Anatyical approach described in the section 3.5.3.

A.2.1. Sanity check with simulation

A first cross-check was to see if the proposed model can reproduce simulation efficiency. For that purposes, dedicated Geant4 samples were generated. It consists of 10,000 events of electrons interacting in a single cube at a central position with the same energy. This has been done to scan an energy window from 100 keV to 1 MeV. Then read-out simulation was run over those files and the number of events passing the selection is counted to derive the efficiency for each tested energy.

Since all the parameters used in the readout simulation are known, the model represented by equation A.3 can use the same parameters, and is therefore expected to reproduce the efficiency determined above. From the equation A.3 one can integrate over k to have the total probability to see an event for a given deposited energy as shown by the equation A.4.

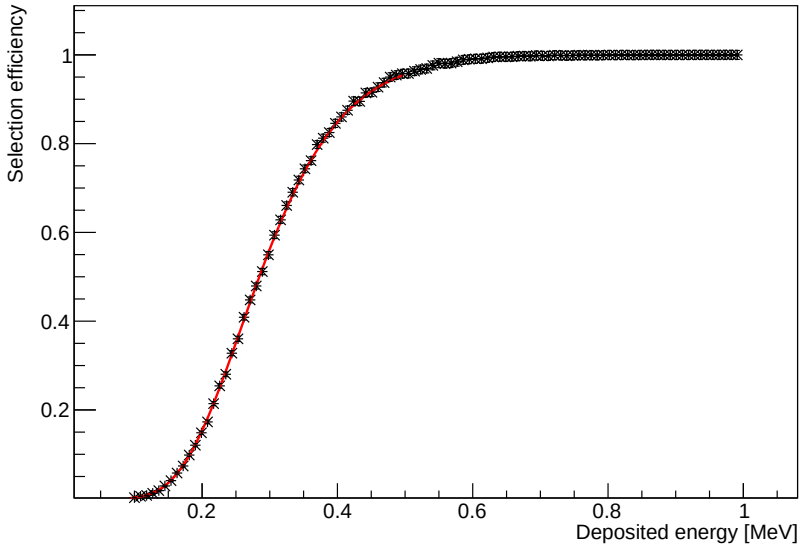


Figure A.3.: Comparison between simulation efficiency to have four fibres above 2.5 PAs in black stars with statistical errors and the model in red as a function of the deposited energy.

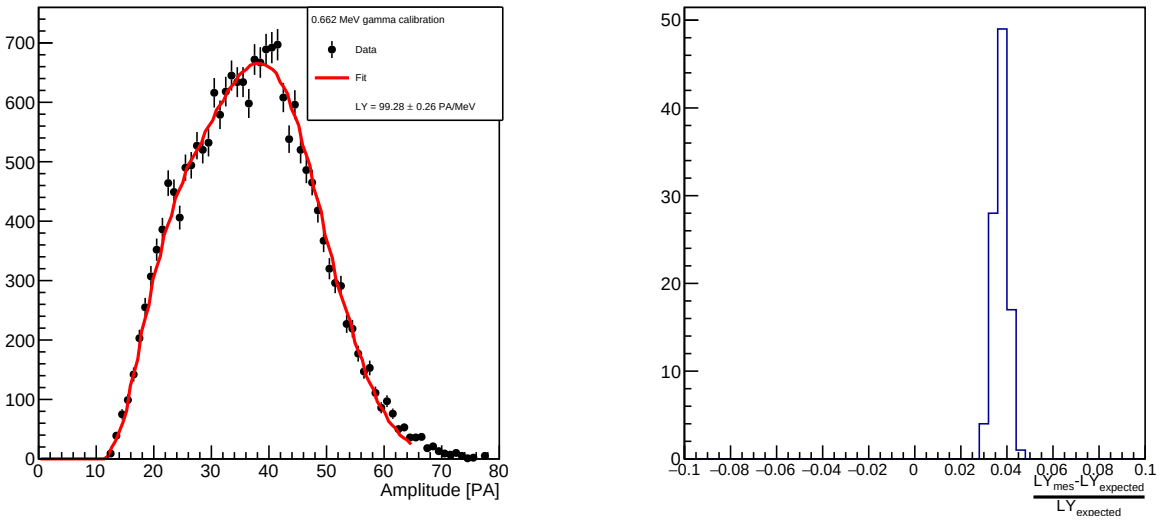
$$\epsilon_{selection}(E_{dep}) = \sum_{k=0}^{\infty} P_{Cube}(X = k, E_{dep}) \tag{A.4}$$

The comparison between the simulation and the model is given in the figure A.3. The two are in good agreement meaning that this model can reproduce the full simulation response at low energy. This is a validation to use the model to fit low energetic source in order to extract the light yield value.

A.3. Toy studies

Another validation work has been made to evaluate possible biases. The same procedure has been followed as in section 3.5.5. First we generated toys samples based on the Klein-Nishina formula. Then, we generated samples where the distributions were drawn from G4 distributions. Only the second analysis is shown since it is the most realistic one.

Given the complexity of the fit, using Minuit algorithm to find the best parameter would be too long. Indeed the algorithm is evaluating the function for light yield to compute the chi-square and its derivative in order to found the minimum and the associated error, this require to evaluate the equation A.3 for a lot of different values of LY_{rec} and all loop over all the partition from the set $\mathbb{A}_{4-Fibres,Selected}(k)$.



(a) An example of one fitted sample

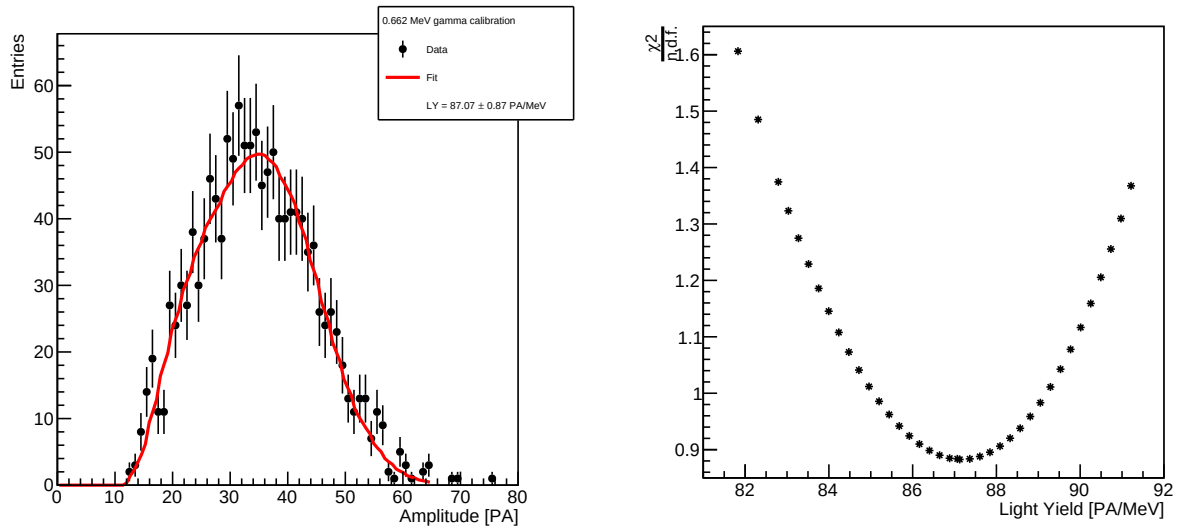
(b) Bias on the light yield observed with 100 toys.

Figure A.4.: Results from the toys study generated with simulation deposited energy. *Left:* An example of one fitted toy.

Figure A.4 is showing the results of this toy study. One hundred toys have been generated using the true deposited energy distribution from a cube in front of the ^{137}Cs source from the Geant4 simulation. The manual fit performs well in the most of the fitted range. In the tail of the distribution, the fit starts to deviate from the simulated data. This is the region where the distribution is the most impacted by multi-scattering interactions. The fitted light yield is more biased than for the ^{22}Na study: 3.7% instead of 2%.

A.4. Preliminary tests on data

This study requires to have un-biased data and thus we have to use random trigger runs. Indeed, the runs acquired with the threshold trigger would distorted the cube amplitude shape. The most simple source available is the ^{137}Cs as it emits only one gamma of 662 keV as shown in the figure 3.3. Also even if few runs with the ^{22}Na source and with the random trigger have been taken in September 2018, it would require to apply a selection to tag and remove the 1.274 MeV gamma from the analysis, as described in the section 4.3.2.4. This selection would lead to a lower efficiency than using the ^{137}Cs data which do not require a selection. Few runs have been taken in September 2018 with this source and the random trigger for one hour. It represents a very low number of exploitable events, only 8 cubes closer to the source are usable. The fit well converges for all cubes, an example of the fit and the evolution of the reduced chi-square as a function of the light yield is shown in the figure A.5. The bias of 3.7% is already corrected.



(a) Fitted cube amplitude distribution.

(b) Scan of the $\frac{\chi^2}{n.d.f.}$ as a function of the light yield.

Figure A.5.: An example of the fit performs on real data with the ^{137}Cs source, the bias observed with the toy study is already subtracted.

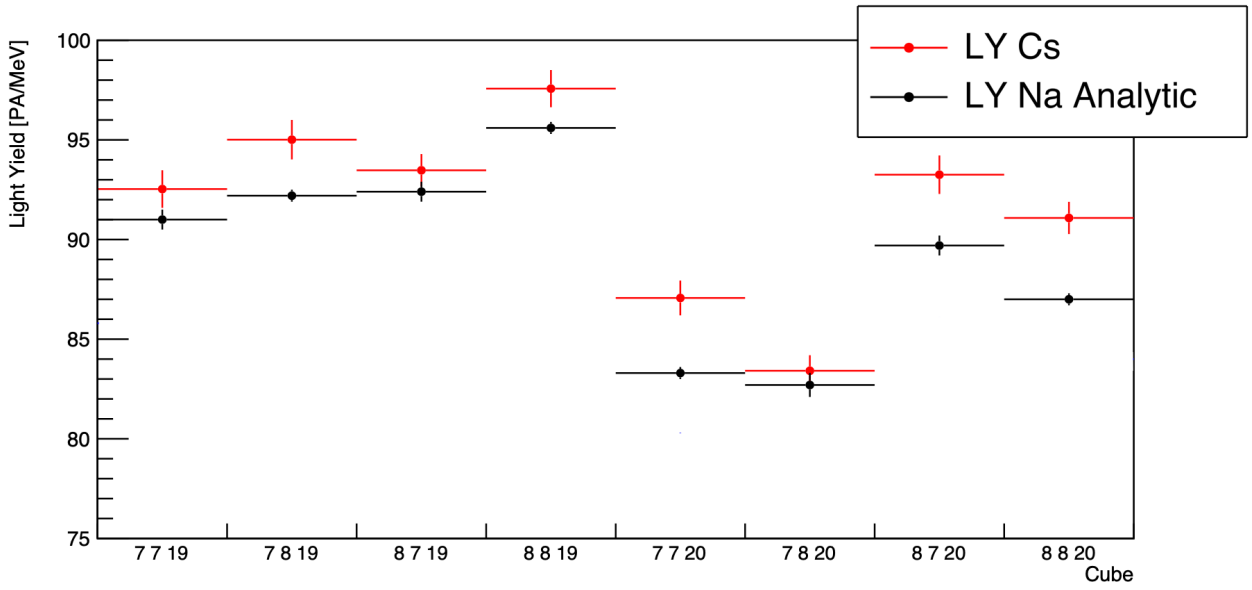


Figure A.6.: Comparison of the light yield extracted from ^{137}Cs and ^{22}Na data.

The model correctly reproduces the data distribution. The comparison between the light yield found with the ^{137}Cs and the LY_{rec} found at "high" energy using the 1.27 MeV gamma from the ^{22}Na sources is shown in the figure A.6.

The ^{137}Cs results correctly follow the cube to cube evolution, which proves the sensibility of this method to the light yield. A discrepancy of few percents is observed between the Analytical approach with the ^{22}Na source at 1.2 MeV and the ^{137}Cs but it remains very promising.

To conclude there are several improvements that could be made:

- Those results are derived using the Klein-Nishina formula as deposited energy spectrum, using the ones extracted from the simulation will improve the performance.
- A more precise study should be done using the full simulation to understand the discrepancy with the light yield extracted from 22 Na data.
- Actually, the light yield determination takes around 15 minutes which is long, thus a modification in the code to scan in parallel several values of the light yield would reduce the time needed.
- During the last calibration campaign, in June 2020, a lot of runs with random triggers and the ^{22}Na source have been taken. So a lot of statistics are available by selecting annihilation gamma and fit them.

Appendix B.

Fibre analysis threshold

The choice of lowering the fibre analysis threshold from 4.5 (200 keV) to 2.5 PA (100 keV) has been motivated by the increase in the efficiency to reconstruct low energy deposits. Naively, this would induce a shift of events in the 1-gamma and 2-gamma categories for which the discrimination power is higher. We wanted to perform a quantitative comparison between the two thresholds before the ICHEP-2020 conference but because of a lack of time we decided to perform a qualitative study. Lowering the threshold does not influence the number of reconstructed events selected by the preselection described in equation 5.14. Indeed the variable used in this preselection did not change while lowering the threshold. Even the E_{prompt} variable does not change a lot since lowering the threshold will add new reconstructed cubes with a low energy. So we expect an impact only on topological variable introduced in section 5.2. Since the discrimination power increases as a function of the number of reconstructed gammas, we expect to reach better performance with a lower fibre analysis threshold. But this gain could not be understood only by the migration of events toward higher categories. We have performed in this section an event by event comparison between low and high threshold of both signal and background. The goal is to compare the discrimination power of events that have changed category by lowering the threshold with regards to events that remain in the same category.

B.1. Low/High threshold association

This study necessitates to perform an event per event comparison. The set of HT events is a subset of the LT events, then the algorithm is looping over LT events and its HT version is found requiring the following variables to equals: ΔT_{ES-NS}, ΔX_{ES-NS}, ΔY_{ES-NS}, ΔZ_{ES-NS}, ΔR_{ES-NS}, $X_{Delayed}$, $Y_{Delayed}$ and $Z_{Delayed}$. A quality criteria of this selection is to look at the number of events that have been found only in the LT set or in the two datasets. Those fractions are given in table B.1 for both the reactor-off and IBD simulation datasets.

The fraction of events found only in the LT dataset is negligible, it corresponds to events with a total cluster energy just below the 1 MeV threshold applied in the reconstruction software. By

	LT	LT and HT
Reactor-off	0.4%	99.6%
IBD simulation	0.06%	99.994%

Table B.1.: Fraction of events found only in the LT dataset or found in the LT and HT datasets. We can see that decreasing the fibre analysis threshold from 4.5 to 2.5 PAs does not bring new events thus we can perform the event per event comparison.

lowering the threshold, new fibres with a low amplitude are added to the cluster and thus its energy increases a bit.

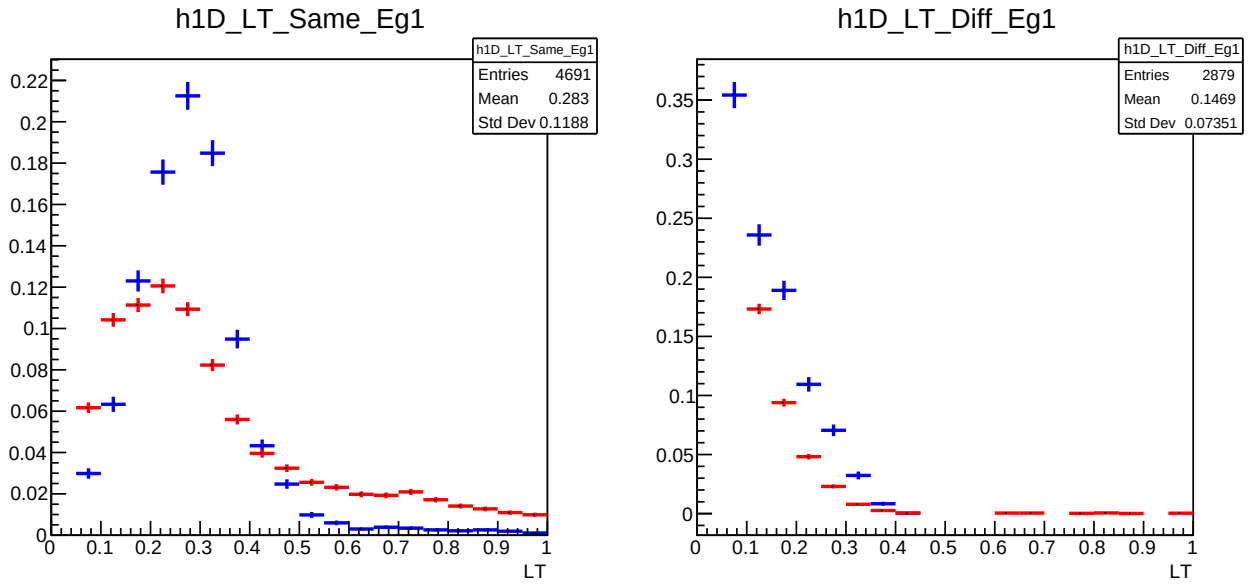
B.2. Comparison between LT and HT

The interest of such analysis is to compare the discrimination power for variables for events that have shifted of gamma category. Thus the events are classified into two groups:

- The number of reconstructed annihilation gamma is conserved.
- The number of reconstructed annihilation gamma increases.

Then for two groups, we compared the LT distribution of all topological variables, derived from the likelihood approach, used by the analysis in order to see if it brings new discrimination. This is shown in the figure B.1 for the $E_{\gamma 1}$ variable. For events that changed category, the discrimination power is less important thus it implies that even if the number of events in the most discriminant categories, 1 or 2 reconstructed annihilation gammas, the gain is less important because the background has a closer distribution from the signal one. Another example is the dot product variable between the two tracks shown on figure B.2. We can draw from this variable the same conclusion, the discrimination power is less important for events that have changed category.

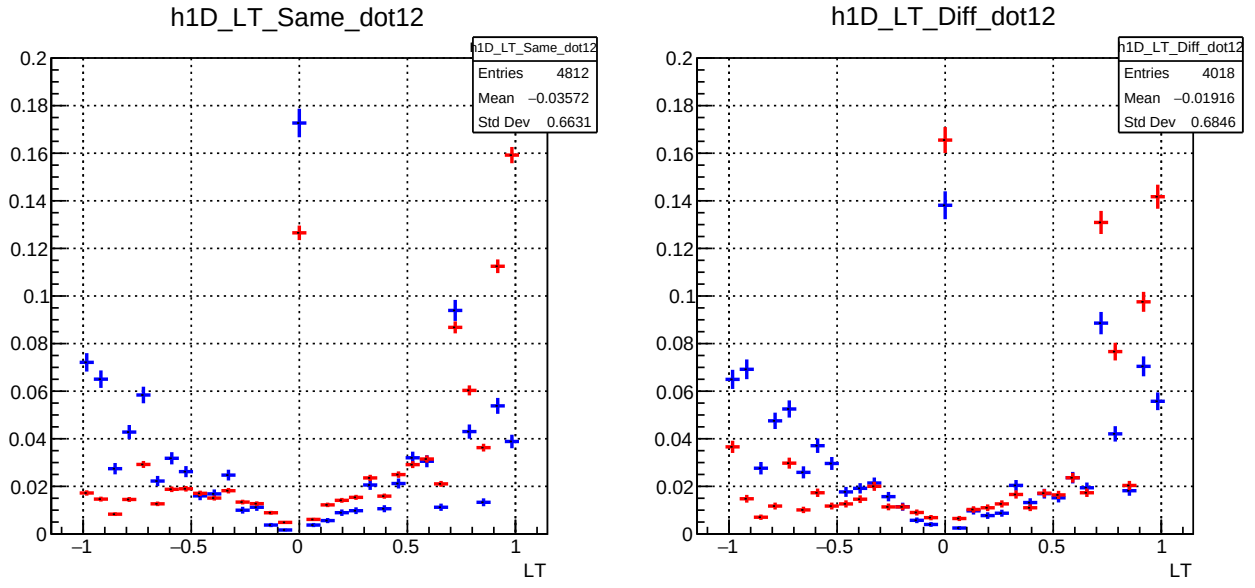
As a conclusion, this study shows that by lowering the analysis thresholds there is a gain in discrimination power, because a larger fraction of events are found in the 1 and 2-gamma categories. But the performance does not scale as we could naively think. Indeed, within each category, the discriminative variables have less power for events that changed category. It's easy to explain in the case of the annihilation gamma E distributions : they tend to be closer between signal and backgrounds since in both cases, new gammas are based on new, low E cubes. Note that with the upgraded detector, the increased light yield, all topological variables should be better reconstructed, including the energy, on which the resolution should be better. This will recover a part of the lost discrimination.



(a) Events not changing of category.

(b) Event changing of category

Figure B.1.: Distribution of signal (blue) and background (red) with regards to $E_{\gamma 1}$ variable using the LT dataset. The events are split into two subsets: events that have not changed of category and events that have changed of category while decreasing the fibre analysis threshold. There is no preselection and the signal distribution is provided by IBD simulation and background distribution comes from the reactor-off.



(a) Events not changing of category.

(b) Event changing of category

Figure B.2.: Distribution of signal (blue) and background (red) with regards to Dot product$_{\gamma 1,\gamma 2}$ variable using the LT dataset. The events are split into two subsets: events that have not changed of category and events that have changed of category while decreasing the fibre analysis threshold. There is no preselection and the signal distribution is provided by IBD simulation and background distribution comes from the reactor-off.

Bibliography

[1] The SLD Electroweak, Heavy Flavour Groups, ALEPH Collaboration, DELPHI Collaboration, L3 Collaboration, OPAL Collaboration, SLD Collaboration, LEP Electroweak Working Group, et al. Precision electroweak measurements on the z resonance. *Physics Reports*, 427(5-6):257–454, 2006.

[2] A. Gando et al. Reactor On-Off Antineutrino Measurement with KamLAND. *Phys. Rev. D*, 88(3):033001, 2013.

[3] Ivan Esteban, Maria Conceptión González-García, Michele Maltoni, Thomas Schwetz, and Albert Zhou. The fate of hints: updated global analysis of three-flavor neutrino oscillations. *Journal of High Energy Physics*, 2020(9):1–22, 2020.

[4] Stefano Gariazzo, C Giunti, M Laveder, YF Li, and EM Zavanin. Light sterile neutrinos. *Journal of Physics G: Nuclear and Particle Physics*, 43(3):033001, 2016.

[5] A. Gando et al. Celand: search for a 4th light neutrino state with a 3 pbq 144ce-144pr electron antineutrino generator in kamland, 2014.

[6] G. Mention, M. Vivier, J. Gaffiot, T. Lasserre, A. Letourneau, and T. Materna. Reactor antineutrino shoulder explained by energy scale nonlinearities? *Phys. Lett. B*, 773:307–312, 2017.

[7] H. et al. Almazán. Improved sterile neutrino constraints from the stereo experiment with 179 days of reactor-on data. *Phys. Rev. D*, 102:052002, Sep 2020.

[8] M. et al. Andriamirado. Improved short-baseline neutrino oscillation search and energy spectrum measurement with the prospect experiment at hfir. *Phys. Rev. D*, 103:032001, Feb 2021.

[9] Valentin Pestel. *Détection de neutrinos auprès du réacteur BR2 : analyse des premières données de l'expérience SoLid*. PhD thesis, Caen U., 2019.

[10] P.A. Zyla et al. Review of Particle Physics. *PTEP*, 2020(8):083C01, 2020.

[11] PAM Dirac. The quantum theory of the emission and absorption of radiation. In *Special Relativity and Quantum Theory*, pages 157–179. Springer, 1988.

[12] E. Fermi. An attempt of a theory of beta radiation. 1. *Z. Phys.*, 88:161–177, 1934.

[13] R. P. Feynman. Relativistic cutoff for quantum electrodynamics. *Phys. Rev.*, 74:1430–1438, 1948.

[14] Julian S. Schwinger. Quantum electrodynamics. I A covariant formulation. *Phys. Rev.*, 74:1439, 1948.

[15] Tomonaga Tati and Sin-itirô Tomonaga. A self-consistent subtraction method in the quantum field theory, i. *Progress of Theoretical Physics*, 3(4):391–406, 1948.

[16] S. L. Glashow. Partial Symmetries of Weak Interactions. *Nucl. Phys.*, 22:579–588, 1961.

[17] Peter W. Higgs. Broken symmetries, massless particles and gauge fields. *Phys. Lett.*, 12:132–133, 1964.

[18] F. Englert and R. Brout. Broken Symmetry and the Mass of Gauge Vector Mesons. *Phys. Rev. Lett.*, 13:321–323, 1964.

[19] S. L. Glashow, J. Iliopoulos, and L. Maiani. Weak interactions with lepton-hadron symmetry. *Phys. Rev. D*, 2:1285–1292, Oct 1970.

[20] A Salam. Elementary particle theory, almquist and wiksells, n. *Svartholm, Stockholm*, 1968.

[21] Steven Weinberg. A Model of Leptons. *Phys. Rev. Lett.*, 19:1264–1266, 1967.

[22] Gerardus't Hooft. Renormalization of massless yang-mills fields. *Nuclear physics: B*, 33(1):173–199, 1971.

[23] Nicola Cabibbo. Unitary Symmetry and Leptonic Decays. *Phys. Rev. Lett.*, 10:531–533, 1963.

[24] Makoto Kobayashi and Toshihide Maskawa. Cp-violation in the renormalizable theory of weak interaction. *Progress of theoretical physics*, 49(2):652–657, 1973.

[25] Serguei Chatrchyan, Vardan Khachatryan, Albert M Sirunyan, Armen Tumasyan, Wolfgang Adam, Ernest Aguilo, Thomas Bergauer, M Dragicevic, J Erö, C Fabjan, et al. Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc. *Physics Letters B*, 716(1):30–61, 2012.

[26] Georges Aad, Tatevik Abajyan, B Abbott, J Abdallah, S Abdel Khalek, Ahmed Ali Abdelalim, R Aben, B Abi, M Abolins, OS AbouZeid, et al. Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc. *Physics Letters B*, 716(1):1–29, 2012.

[27] A. D. Sakharov. Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe. *Sov. Phys. Usp.*, 34(5):392–393, 1991.

[28] J Chadwick. Intensitätsverteilung im magnetischen Spectrum der β-Strahlen von radium B + C. *Verhandl. Dtsc. Phys. Ges.*, 16:383, 1914.

[29] Wolfgang Pauli. Pauli letter collection: letter to Lise Meitner. Typed copy.

[30] Clyde L Cowan, Frederick Reines, FB Harrison, HW Kruse, and AD McGuire. Detection of the free neutrino: a confirmation. *Science*, 124(3212):103–104, 1956.

[31] Maurice Goldhaber, Lee Grodzins, and Andrew W Sunyar. Helicity of neutrinos. *Physical review*, 109(3):1015, 1958.

[32] Gaillard Danby, Jean Maurice Gaillard, Konstantin Goulianos, Leon M Lederman, N Mistry, M Schwartz, and J Steinberger. Observation of high-energy neutrino reactions and the existence of two kinds of neutrinos. *Physical Review Letters*, 9(1):36, 1962.

[33] K Kodama, N Ushida, C Andreopoulos, N Saoulidou, G Tzanakos, P Yager, B Baller, D Boehnlein, Walter Freeman, B Lundberg, et al. Observation of tau neutrino interactions. *Physics Letters B*, 504(3):218–224, 2001.

[34] John N Bahcall. Solar neutrinos. i. theoretical. *Physical Review Letters*, 12(11):300, 1964.

[35] Raymond Davis Jr, Don S Harmer, and Kenneth C Hoffman. Search for neutrinos from the sun. *Physical Review Letters*, 20(21):1205, 1968.

[36] KS Hirata, T Kajita, M Koshiba, M Nakahata, S Ohara, Y Oyama, N Sato, A Suzuki, M Takita, Y Totsuka, et al. Experimental study of the atmospheric neutrino flux. *Physics Letters B*, 205(2-3):416–420, 1988.

[37] Q Retal Ahmad, RC Allen, TC Andersen, JD Anglin, JC Barton, EW Beier, M Bercovitch, J Bigu, SD Biller, RA Black, et al. Direct evidence for neutrino flavor transformation from neutral-current interactions in the sudbury neutrino observatory. *Physical review letters*, 89(1):011301, 2002.

[38] Bruno Pontecorvo. Mesonium and antimesonium. *Zhur. Eksptl'. i Teoret. Fiz.*, 33, 1957.

[39] Ziro Maki, Masami Nakagawa, and Shoichi Sakata. Remarks on the unified model of elementary particles. *Progress of Theoretical Physics*, 28(5):870–880, 1962.

[40] Bruno Pontecorvo. Neutrino experiments and the problem of conservation of leptonic charge. *Sov. Phys. JETP*, 26(984-988):165, 1968.

[41] Ettore Majorana. Theory of the symmetry of electrons and positrons. *Nuovo Cim*, 14(171):50, 1937.

[42] M. et al. Aker. Improved upper limit on the neutrino mass from a direct kinematic method by katrin. *Phys. Rev. Lett.*, 123:221802, Nov 2019.

[43] D. et al. Akimov. Observation of coherent elastic neutrino-nucleus scattering. *Science*, 357(6356):1123–1126, 2017.

[44] O. G. Miranda, D. K. Papoulias, O. Sanders, M. Tórtola, and J. W. F. Valle. Future cevns experiments as probes of lepton unitarity and light sterile neutrinos. *Phys. Rev. D*, 102:113014, Dec 2020.

[45] W Hampel, G Heusser, J Kiko, T Kirsten, M Laubenstein, E Pernicka, W Rau, U Rönn, C Schlosser, M Wójcik, et al. Final results of the 51cr neutrino source experiments in gallex. *Physics Letters B*,

420(1-2):114–126, 1998.

[46] JN Abdurashitov. Measurement of the response of the russian-american gallium experiment to neutrinos from a 51 cr source. *Phys. Rev. C*, 59(hep-ph/9803418):2246, 1999.

[47] A. et al. Aguilar. Evidence for neutrino oscillations from the observation of $\overline{\nu}_e$ appearance in a $\overline{\nu}_\mu$ beam. *Phys. Rev. D*, 64:112007, Nov 2001.

[48] A. A. et al. Aguilar-Arevalo. Improved search for $\overline{\nu}_\mu \to \overline{\nu}_e$ oscillations in the miniboone experiment. *Phys. Rev. Lett.*, 110:161801, Apr 2013.

[49] G. Mention, M. Fechner, Th. Lasserre, Th. A. Mueller, D. Lhuillier, M. Cribier, and A. Letourneau. Reactor antineutrino anomaly. *Phys. Rev. D*, 83:073006, Apr 2011.

[50] Z. Atif and et al. Search for sterile neutrino oscillation using reno and neos data, 2020.

[51] I. Alekseev et al. Search for sterile neutrinos at the danss experiment. *Physics Letters B*, 787:56–63, 2018.

[52] A. P. Serebrov et al. Experiment neutrino-4 search for sterile neutrino and results of measurements, 2021.

[53] Anatael Cabrera. Possible precise neutrino unitarity?, 2019.

[54] Denise Pelowitz, Joe Durkee, Jay Elson, Michael Fensin, John Hendricks, Michael James, Russell Johns, Gregg Mckinney, S. Mashnik, Jerome Verbeke, Laurie Waters, and Taylor Wilcox. MCNPX 2.7.0 extensions, 04 2011.

[55] M. James, D. Pelowitz, A. Fallgren, G. McMath, Thomas Booth, F. Brown, J. Bull, Lawrence Cox, J. Elson, J. Durkee, Michael Fensin, R. Forster, J. Goorley, J. Hendricks, H. Hughes, Russell Johns, B. Kiedrowski, Gregg Mckinney, Roger Martz, and Anthony Zukaitis. MCNP6 User's Manual. Code Version 6.1.1 Beta, 06 2014.

[56] Franz Gallmeier, Phillip Ferguson, Wei Lu, Erik Iverson, Guenter Muhrer, Shannon Holloway, Charles Kelsey, Eric Pitcher, Michael Wohlmuther, and Bradley Micklich. The cinder 90 transmutation code package for use in accelerator applications in combination with mcnpx. *ICANS XIX*, 03 2010.

[57] S. Kalcheva. Reactor core simulations for determination of the antineutrino spectrum for the solid experiment at br 2 reactor. 2017.

[58] Patrick Huber. Determination of antineutrino spectra from nuclear reactors. *Phys. Rev. C*, 84:024617, Aug 2011.

[59] Th. A. Mueller, D. Lhuillier, M. Fallot, A. Letourneau, S. Cormon, M. Fechner, L. Giot, T. Lasserre, J. Martino, G. Mention, A. Porta, and F. Yermia. Improved predictions of reactor antineutrino spectra. *Phys. Rev. C*, 83:054615, May 2011.

[60] K. Schreckenbach, H.R. Faust, F. von Feilitzsch, A.A. Hahn, K. Hawerkamp, and J.L. Vuilleumier. Absolute measurement of the beta spectrum from 235 U fission as a basis for reactor antineutrino experiments. *Phys. Lett. B*, 99:251–256, 1981.

[61] K. Schreckenbach, G. Colvin, W. Gelletly, and F. Von Feilitzsch. DETERMINATION OF THE ANTI-NEUTRINO SPECTRUM FROM U-235 THERMAL NEUTRON FISSION PRODUCTS UP TO 9.5-MEV. *Phys. Lett. B*, 160:325–330, 1985.

[62] F. Von Feilitzsch, A.A. Hahn, and K. Schreckenbach. EXPERIMENTAL BETA SPECTRA FROM PU-239 AND U-235 THERMAL NEUTRON FISSION PRODUCTS AND THEIR CORRELATED ANTI-NEUTRINOS SPECTRA. *Phys. Lett. B*, 118:162–166, 1982.

[63] MCNP MURE. utility for reactor evolution: couples monte-carlo transport with fuel burnup calculations, 2009.

[64] Guillaume Pronost. *Studies for the measurement of the mixing angle θ_{13} with the Double Chooz experiment*. PhD thesis, SUBATECH, Nantes, 2015.

[65] Yamiel Abreu, Y Amhis, L Arnold, G Ban, W Beaumont, M Bongrand, D Boursette, JM Buhour, BC Castle, K Clark, et al. A novel segmented-scintillator antineutrino detector. *Journal of Instrumentation*, 12(04):P04024, 2017.

[66] Eljen general purpose plastic scintillator: online datasheet. Available at:https://eljentechnology.com/images/products/data_sheets/EJ-200_EJ-204_EJ-208_EJ-212.pdf.

[67] Scintacor neutron screens: online datasheet. Available at:https://scintacor.com/wp-content/uploads/2015/09/Datasheet-Neutron-Screens-High-Res.pdf.

[68] Dupont tyvek 1082d: online datasheet. Available at:https://www.dupont.cn/content/dam/dupont/amer/us/en/microsites/tyvek-design/images/documents/2019-C&I_Tyvek_1082D_Datasheet.pdf.

[69] Saint-gobain plastic scintillating fibers: online datasheet. Available at:https://www.crystals.saint-gobain.com/sites/imdf.crystals.com/files/documents/fiber-product-sheet.pdf.

[70] Hamamatsu multi-pixel photon counter: online datasheet. Available at:https://ibook.antpedia.com/attachments/product/212/1426215004-7605.pdf.

[71] Y. Abreu et al. Development of a Quality Assurance Process for the SoLid Experiment. *JINST*, 14(02):P02014, 2019.

[72] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22, 1977.

[73] Gengsheng Lawrence Zeng. *Medical image reconstruction: a conceptual tutorial.* Springer, 2010.

[74] A. Simón et al. Application and performance of an ML-EM algorithm in NEXT. *JINST*, 12(08):P08009, 2017.

[75] Ianthe Michiels. *Development of the oscillation analysis framework for the SoLid experiment*. PhD thesis, Gent U., Proeftuinstraat 86, 9000, Gent, Belgium, 2020.

[76] K. M. Heeger, M. N. Tobin, B. R. Littlejohn, and H. P. Mumm. Experimental parameters for a reactor antineutrino experiment at very short baselines. *Phys. Rev. D*, 87:073008, Apr 2013.

[77] M. Pivk and F.R. Le Diberder. : A statistical tool to unfold data distributions. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 555(1-2):356–369, Dec 2005.

[78] John Betteley Birks. Scintillations from organic crystals: specific fluorescence and relative response to different radiations. *Proceedings of the Physical Society. Section A*, 64(10):874, 1951.

[79] Oskar Klein and Yoshio Nishina. The scattering of light by free electrons according to dirac's new relativistic dynamics. *Nature*, 122(3072):398–399, 1928.

[80] Delphine Boursette. *Neutrino physics with SoLid and SuperNEMO experiments*. PhD thesis, Paris Saclay, 2018.

[81] E Dietz-Laursonn, T Hebbeker, A Künsken, M Merschmeyer, S Nieswand, and T Niggemann. Goddess: a geant4 extension for easy modelling of optical detector components. *Journal of Instrumentation*, 12(04):P04026, 2017.

[82] V Senchishin, V Koba, and O Korneeva. New radiation stable and long-lived plastic scintillator for the ssc.

[83] A Artikov, J Budagov, I Chirikov-Zorin, D Chokheli, M Lyablin, G Bellettini, A Menzione, S Tokar, N Giokaris, and A Manousakis-Katsikakis. Properties of the ukraine polystyrene-based plastic scintillator ups 923a. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 555(1-2):125–131, 2005.

[84] GEANT Collaboration, S Agostinelli, et al. Geant4–a simulation toolkit. *Nucl. Instrum. Meth. A*, 506(25):0, 2003.

[85] Chris Hagmann, David Lange, Jerome Verbeke, and Doug Wright. Cosmic-ray shower library (cry). *Lawrence Livermore National Laboratory document UCRL-TM-229453*, 2012.

[86] MS Gordon, P Goldhagen, KP Rodbell, TH Zabel, HHK Tang, JM Clem, and P Bailey. Measurement of the flux and energy spectrum of cosmic-ray induced neutrons on the ground. *IEEE Transactions on Nuclear Science*, 51(6):3427–3434, 2004.

[87] O. Klein and T. Nishina. Über die Streuung von Strahlung durch freie Elektronen nach der neuen

relativistischen Quantendynamik von Dirac. *Zeitschrift fur Physik*, 52(11-12):853–868, November 1929.

[88] David A Brown, MB Chadwick, R Capote, AC Kahler, A Trkov, MW Herman, AA Sonzogni, Y Danon, AD Carlson, M Dunn, 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. *Nuclear Data Sheets*, 148:1–142, 2018.

[89] Andreas Hocker et al. TMVA - Toolkit for Multivariate Data Analysis. 3 2007.

[90] Justin Stevens and Mike Williams. uboost: A boosting method for producing uniform selection efficiencies from multivariate classifiers. *Journal of Instrumentation*, 8(12):P12013, 2013.

[91] Wouter Verkerke and David P. Kirkby. The RooFit toolkit for data modeling. *eConf*, C0303241:MOLT007, 2003.

[92] Roger J. Barlow and Christine Beeston. Fitting using finite Monte Carlo samples. *Comput. Phys. Commun.*, 77:219–228, 1993.

Résumé en français

Le Modèle Standard de physique des particules englobe notre compréhension actuelle de ce domaine. Son succès provient de sa capacité à expliquer toutes les observations expérimentales menées en laboratoire mais aussi à prédire certaines de ces observations. Cependant, une série intrigante d'anomalies est apparue au cours des dernières années mais nécessite toujours une confirmations. Sur d'autres points, il y a d'importantes raisons de penser que le Modèle Standard est seulement une manifestation d'une théorie plus fondamentale de la physique. Parmi ces raisons, on peut citer en particulier la manière dont la masse et la saveur des particules sont intégrées au sein du Modèle Standard ainsi que la violation de la symétrie Charge-Parité. En effet, l'amplitude de violation de la symétrie Charge-Parité ne parvient pas à expliquer l'asymétrie baryonique observée dans l'Univers.

La Physique des neutrinos est une intéressante voie afin d'étudier ces effets et de trouver de la physique au-delà du Modèle Standard. L'étude du phénomène d'oscillation des neutrinos permet d'effecteur plusieurs séries de mesures afin de contraindre ces nouveaux modèles. La mesure précise des paramètres d'oscillation est une nouvelle façon d'explorer les problèmes évoqués précédemment. En particulier, la masse très faible des neutrinos par rapport aux autres particules, au moins cinq ordres de grandeur plus faibles, devrait apporter d'importantes informations. En effet, dans le cas d'un fermion neutre, les termes de Dirac décrivant la masse des particules dans le Lagrangien, ne sont plus la meilleure façon de générer une masse. Ainsi, les neutrinos pourraient être les premiers fermions dits de Majorana découverts. Cette potentielle découverte aura un impact sur les théories au-delà du Modèle Standard, avec la possibilité d'obtenir des processus violant la conservation du nombre leptonique.

Au cours des vingt dernières années, l'intense étude des oscillations de neutrino a permis de faire une description précise du phénomène dans le cadre standard du modèle à trois saveurs. Les trois angles de mélanges utilisés pour décrire la matrice PMNS dans l'hypothèse d'unitarité ainsi que les carrés des différences de masses sont désormais connus à quelques pourcents. Malgré ce progrès, ni la structure de la matrice PMNS ni la petitesse de leurs masses n'est expliqué. Une nouvelle génération d'expériences est en cours de construction pour contraindre plus précisément les paramètres de cette matrice, mais aussi explorer les possibilités d'avoir plus de trois neutrinos. Les deux expériences, DUNE et T2(H)K, vont permettre de mesurer la violation de symétrie Charge-Parité dans le secteur leptonique ainsi que statuer sur l'ordre des masses des neutrinos. Cette nouvelle ère de précision va aussi permettre de tester l'unitarité de la matrice PMNS, qui si elle est violée mettra en lumière de la nouvelle physique au-delà du Modèle Standard.

Les mesures entreprisent pour contraindre les paramètres de la matrice PMNS décrivant le phénomène d'oscillations des neutrinos ont aussi révélé plusieurs anomalies. En particulier les expériences ayant étudié les antineutrinos électronique provenant des réacteurs nucléaires ont observé un déficit par rapport au flux attendu théoriquement. Cette anomalie est appelée l'anomalie réacteur. Ce déficit peut être interprété par la contribution d'un quatrième neutrino avec une masse de l'ordre de l'Électron-volt. À cette énergie, l'oscillation devient maximale à très courte distance : seulement quelques mètres du cœur du réacteur nucléaire. Un effort mondial a été initié afin d'obtenir de nouvelles mesures proches de réacteurs nucléaires.

L'expérience SoLid s'inscrit totalement dans ce contexte. Le détecteur est positionné entre six et neuf mètres du coeur du réacteur BR2 SCK-CEN en Belgique. Ce réacteur expériemental dispose d'un cœur compact et hautement enrichi en Uranium 235 offrant une meilleure précision sur la distance parcourue par les antineutrinos électronique. Ces particules sont détectées via la décroissance bêta inverse correspondant à l'interaction d'un antineutrino électronique sur un proton résultat dans l'émission d'un positron et d'un neutron. La cible du détecteur est constituée de 12,800 cubes de plastique scintillant en Polyvinyltoluene de taille 5x5x5 cm3 permettant une mesure précise de l'énergie du positron. La détection du neutron est effectuée via sa capture sur des noyaux de Lithium dont résulte l'émission d'un tritium et d'un alpha partageant environ 4 MeV d'énergie. Le Lithium est contenu dans des feuilles de scintillateur inorganique LiF:ZnS(Ag) au nombre de deux par cubes. La combinaison de ces scintillateurs permet de différencier les signaux provenant du Polyvinyltoluene et du LiF:ZnS(Ag). En effet, le temps de scintillation entre les deux scintillateurs diffère d'environ trois ordres de grandeurs. Les cubes sont agencés en cinquante plans de 256 cubes chacun et chaque cube est traversé par quatre fibres optiques afin de récolter les photons de scintillations. Au bout de chaque fibre est placé d'un côté un miroir et de l'autre un MPPC permettant de convertir les photons de scintillation en signal digital exploitable. La haute segmentation de la cible permet une reconstruction fine des événements et en particulier d'exploiter la différence en topologie, spatial et énergétique, entre le signal et le bruit de fond important. En effet, l'expérience souffre de deux bruits de fond principaux : le BiPo qui résulte d'une contamination des feuilles de LiF:ZnS(Ag) lors de leur production et le bruit de fond d'origine cosmique dû au faible enfouissement de l'expérience.

Les premiers travaux présentés dans ce manuscrit concernent la calibration en énergie du détecteur. C'est un travail très important dans l'optique d'effecteur une recherche d'oscillation de neutrino. En effet, la probabilité d'oscillation évolue en fonction de la distance parcourue, mais aussi de l'énergie du neutrino. Il est donc primordial de correctement caractériser la réponse du détecteur en énergie. En particulier, ces mesures permettront d'ajuster la réponse de la simulation du détecteur, outil essentiel afin de savoir si la mesure est statistiquement significative. Les travaux de calibrations concernent principalement la détermination de trois paramètres : le rendement de lumière produit par chaque cube, la constante d'atténuation de chaque fibre optique et une constante de couplage entre la fibre optique et les MPPC. La difficulté de ces mesures réside dans le nombre de paramètres à déterminer, 19 200 ainsi que la difficulté de décoléré chaque effet les uns des autres. Plusieurs méthodes ont été développées pour déterminer ces paramètres et leurs performances ont été déterminées à l'aide de la simulation. Tout d'abord pour la détermination du rendement de lumière des cubes la méthode tire

avantage de la forme du front Compton présent dans la section efficace d'interaction des particules gammas émis par les sources de calibration. Dans un second temps, une nouvelle méthode a été développée pour déterminer en même temps les couplages fibres-MPPC et les longueurs d'atténuations d'un plan. Elle utilise les trois asymétries d'amplitude - haut-bas, gauche-droite et horizontal-vertical - entre fibres, calculées dans chaque cube d'un plan. Un ajustement global des distributions des trois asymétries permet d'obtenir une mesure des deux paramètres pour chaque canaux. Pour finir, la linéarité de la réponse en énergie a aussi été prouvée à quelques pourcents près et une méthode d'extraction du rendement de lumière à basse énergie a été développée.

La détermination des constantes de calibration a permis d'entreprendre l'amélioration de la simulation du détecteur. La première étape a été d'effectuer une comparaison entre les données et la simulation dans un seul cube du détecteur en utilisant les données de calibration. Cette phase a permis de mettre en évidence et de corriger une série d'erreurs dans le code de simulation, mais aussi divers temps mort très important durant la prise de donnée. En effet, plusieurs types de temps mort ont été mis en évidence :

- Au niveau des canaux impliquant leur perte jusqu'à la réinitialisation de la prise de donnée.
- Au niveau du plan qui reprendra la prise de données uniquement lorsque sa mémoire sera vidée.

Pour pallier à ces temps morts, un autre jeu de donnée de calibration a été utilisé afin de finaliser l'ajustement de la simulation en énergie. Finalement, un accord à quelques pour-cent est obtenu sur l'énergie globale du détecteur ainsi que l'énergie reconstruction dans les cubes. L'étape a été d'utiliser le bruit de fond BiPo pour effectuer une comparaison entre les données et la simulation des variables nouvellement développées pour améliorer la sélection du signal.

Finalement, les derniers travaux de thèse ont contribué à l'amélioration de la sélection du signal. Les expériences essayant de mettre en évidence des oscillations de neutrino à très courte distance de réacteur nucléaire souffrent d'important bruit de fond. La sélection du signal alors développée au sein de la collaboration prenait en compte uniquement les variables dépendant de la différence en espace et en temps entre l'événement positron et neutron. Cependant, de l'annihilation du positron résulte l'émission de deux gammas d'annihilation. En plus de la particularité de l'énergie, ils sont émis dos à dos donnant au signal une signature spécifique. Une reconstruction des petits dépôts d'énergie a donc été développée dans le but d'améliorer le rapport signal sur bruit de fond. De ce travail, plusieurs nouvelles variables sont déterminées en fonction du nombre de gammas d'annihilation reconstruits. Leur pouvoir de discrimination a été mis en évidence à l'aide d'une sélection rectangulaire offrant de meilleur performance que la précédente sélection. Pour obtenir la meilleure sélection possible, une analyse multivariée a été développée à l'aide d'un réseau de neurones pour tenir pleinement compte des corrélations entre les différentes variables. Il en résulte une amélioration du rapport signal sur bruit de fond d'un facteur XX tout en gardant le flux de signal constant. Finalement, cette sélection a été appliquée sur les données avec la méthode traditionnelle d'extraction du signal. Le flux ainsi que les distributions de l'excès ont été comparés avec la simulation du signal et sont en accord. La dernière contribution consiste au développement d'une seconde méthode pour l'extraction du signal.

UNIVERSITE MATIERE BRETAGNE MOLECULES LOIRE ET MATERIAUX

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

Titre : Vers la recherche d'un neutrino stérile avec l'expérience SoLid au réacteur BR2 de SCK-CEN : calibration du détecteur et conception d'une sélection topologique et multivariée d'événements antineutrinos

Mots clés : SoLid ; antineutrino ; neutrino stérile ; anomalie réacteur ; oscillation de neutrinos ; calibration en énergie ; simulation ; selection.

Résumé : L'expérience SoLid est basée auprès du réacteur de recherche BR2 à SCK-CEN, Belgique. Son but est de rechercher une oscillation de neutrino à très courte distance, qui expliquerait l'anomalie dite du flux d'antineutrinos de réacteur, par la présence d'une nouvelle particule, le neutrino stérile. En mesurant le spectre en énergie des antineutrinos émis par ce réacteur très enrichi en Uranium 235, SoLid aiderait aussi à comprendre la différence de forme entre spectres prédit et mesuré, constatée par des expériences antérieures ("bump à 5 MeV"). Afin de lutter contre les importants bruits de fond d'origine cosmique inhérents aux expériences situées en surface, SoLid a opté pour une segmentation très élevée : le détecteur se compose de 12,800 cellules de détection lues par 3,200 MPPC. Les antineutrinos électroniques sont détectés via des Décroissances Beta Inverses (IBD), avec une nouvelle technologie de détection s'appuyant sur l'association d'un scintillateur plastique (PVT) et minéral (ZnS). Les contributions de cette thèse peuvent se diviser en trois principales parties. Un premier volet a consisté dans la calibration en énergie du détecteur. Ceci est un défi au vu de la haute segmentation du détecteur. La seconde contribution a porté sur l'ajustement de la réponse en énergie dans la simulation et a amélioré son code. Enfin, la dernière contribution a porté sur la sélection et l'extraction du signal. Tout d'abord avec le développement de nouvelles variables discriminantes exploitant les gamma d'annihilations issus de l'IBD et en utilisant des techniques d'analyse multivariable. Puis via l'amélioration de la méthode officielle d'extraction du signal et le développement d'une nouvelle méthode à des fins de comparaison. Mes travaux permettent entre autres de maîtriser l'échelle en énergie des cubes de SoLid à environ 2% et d'améliorer d'un facteur 4 le rapport signal sur bruit de la sélection d'événements IBD, en maintenant constante son efficacité sur le signal.

Title : Toward the search of a sterile neutrino with the SoLid experiment at the SCK-CEN BR2 reactor : detector calibration and design of a multivariate topological selection of antineutrino events

Keywords : SoLid ; Antineutrino ; sterile neutrino ; reactor anomaly ; neutrino oscillation ; energy calibration ; simulation ; selection.

Abstract : The SoLid experience is based near the BR2 research reactor at SCK-CEN, Belgium. Its aim is to search for neutrino oscillation at very short baseline, which would explain the so-called Reactor Antineutrino Anomaly via the presence of a new particle, a sterile neutrino. The measurement of antineutrino energy spectrum from the reactor highly enriched in Uranium 235, SoLid would help to understand the disagreement between predicted and measured observed by previous experiment (so-called 5 MeV bump). In order to fight against the important cosmic induced background inherent in ground-level experiment, the SoLid detector is high segmented : with 12,800 detection cells read-out by 3,200 MPPCs. Electronic antineutrinos are detected via Inverse Beta Decay (IBD), thanks to a new detection technology involving the combination of plastic scintillator, the PVT, and mineral scintillator, the ZNS. Thesis contributions could divided in three main parts. First consisted in the energy calibration of the detector. This is a challenge due to the high segmentation of the target. The second one extended the calibration by adjusting the energy response in the simulation improved its code. Finally, the last one conceived the selection and extraction of the signal. With the development of new discriminating variables which exploits the annihilation gammas from IBD and using multivariate analysis tools. Then via the improvement of the official signal extraction method and the development of a new one for comparison purpose. Among other things, my work allows to control the energy scale of SoLid cubes at around a 2% level and improve by a factor 4 the signal over background ratio for the same signal efficiency.