Correlated and Integrated Directionality in Borexino

Using Cherenkov Directionality for the Measurement of Solar Neutrinos
in High Light-Yield Liquid Scintillator Detectors

Johann Martyn
Born in Rudnyj, Kazakhstan

Dissertation Submitted for the Award of the Title
Doctor of Natural Sciences
to the Faculty of Physics, Mathematics and Computer Science

13. September 2023

Johannes Gutenberg University Mainz

Abstract

Observations of low energy neutrinos from astrophysical sources are performed currently by either liquid scintillator detectors or water Cherenkov detectors. To date, liquid scintillator detection in particular is the only technique that allows precision spectroscopy of sub-MeV solar neutrinos, as demonstrated by the Borexino detector at the Gran Sasso National Laboratory in Italy. This is made possible by the high light-yield of the scintillation, which is also responsible for a low energy threshold of ~ 0.2 MeV, together with the unprecedented radio-purity of the detector materials. In contrast, water Cherenkov detectors are only able to measure solar ^{8}B neutrinos above a few MeV with high precision. They reconstruct the event direction via the corresponding Cherenkov photon hits, to differentiate between the solar neutrino signal and the background events. A combination of the directional Cherenkov information and the high light-yield and low energy threshold of the scintillation spectroscopy can enable an unprecedented precision in the measurement of solar neutrinos and other physics goals. Exploring this novel hybrid detection approach, this thesis describes the first measurement of sub-MeV solar neutrinos using their associated directional Cherenkov photons, as well as a directional measurement of the CNO-neutrino rate, using the Borexino detector. These analyses are based on the specially developed *Correlated and Integrated Directionality* method, where the individual photon hits of the events are correlated to the known position of the Sun. The integration of these angular hit values over a large number of events allows the statistical inference of the number of solar neutrino events. The directional measurement of the 0.862 MeV line of the ^{7}Be-neutrinos is performed in an energy window between 0.56 MeV to 0.76 MeV, which is selected through the dominant scintillation light. The measured rate is $R(^7\mathrm{Be})_{\mathrm{CID}} = 51.6^{+13.9}_{-12.5}\,(\mathrm{stat.} + \mathrm{syst.})\,\mathrm{cpd}/100\,\mathrm{t}$, with an exclusion of the zero neutrino hypothesis of $> 6\sigma$. The directional measurement of the CNO-neutrino rate is performed in an energy region between 0.85 MeV to 1.3 MeV. The measured rate is $R_{\mathrm{CNO}} = 7.2^{+2.8}_{-2.7}\,(\mathrm{stat.} + \mathrm{syst.})\,\mathrm{cpd}/100\,\mathrm{t}$, with an exclusion of the zero CNO-neutrino hypothesis of $> 5\sigma$. These results correspond to the first detection of low energy solar neutrinos using only their directional Cherenkov information in a large-scale, high light-yield scintillator detector. Additionally, this is also a very direct proof of the solar origin of the ^{7}Be- and CNO-neutrino signal events. This demonstration of a directional sensitivity in a monolithic liquid scintillator target provides an experimental proof-of-principle for the concept of hybrid event detection, which could be implemented in future, purpose-built neutrino detectors.

Contents

List of Figures

List of Tables

List of Abbreviations

ASU	**A**cceptance **S**ampling **U**nfolding
CC	**C**harged **C**urrent
CCD	**C**harged and **C**oupled **D**evice
CI	**C**onfidence **I**nterval
CID	**C**orrelated and **I**ntegrated **D**irectionality
CL	**C**onfidence **L**evel
CNO	**C**arbon and **N**itrogen **O**xygen cycle
cpd	**c**ounts **p**er **d**ay
ES	**E**lastic **S**cattering
ET	**E**qual-**T**ail confidence interval definition
FoM	**F**igure **o**f **M**erit
FV	**F**iducial **V**olume
GEANT4	**GE**ometry **AN**d **T**racking
G4Bx2	The **G**eant**4** based simulation framework of **B**ore**x**ino
HD	**H**ighest-**D**ensity confidence interval definition
HZ	**H**igh Metallicity (**Z**)
LS	**L**iquid **S**cintillator
LZ	**L**ow Metallicity (**Z**)
LMA	**L**arge **M**ixing **A**ngle
MC	**M**onte **C**arlo simulation
MLP	**M**ulti **L**ayer **P**erceptron
MSW	**M**ikheyev-**S**mirnov-**W**olfenstein
NC	**N**eutral **C**urrent
ndf	**n**umber of **d**egrees of **f**reedom
PC	**P**seudo**C**umene. The solvent 1,2,4-trimethylbenzene
PDF	**P**robability **D**ensity **F**unction
PMT	**P**hoto**M**ultiplier **T**ube
PPO	The wavelength shifter 2,5-diphenyloxazole
QE	**Q**uantum **E**fficiency
RoI	**R**egion **o**f **I**nterest
TFC	**T**hree **F**old **C**oincidence
ToF	**T**ime **o**f **F**light
SM	**S**tandard **M**odel of particle physics
SSM	**S**tandard **S**olar **M**odel
SSS	**S**tainless **S**teel **S**phere
stat.	**stat**istical uncertainty
syst.	**syst**ematic uncertainty
WbLS	**W**ater **b**ased **L**iquid **S**cintillator
WCh	**W**ater **Ch**erenkov

Introduction

This thesis presents the first directional detection and measurement of low energy neutrinos produced in the Sun. This is done with the Borexino detector, that uses about 280 t of liquid scintillator as a target material. Here, solar neutrinos scatter elastically off the target electrons, which then excite the scintillator, that subsequently emits scintillation light. These recoil-electrons also emit Cherenkov photons, which are strongly sub-dominant to the scintillation. The measurements presented in this thesis are based on these Cherenkov photon hits, in contrast to the typical scintillation spectroscopy of the Borexino experiment.

For this task a novel analysis technique has been developed, validated and optimised. This has been done through the development of new software tools and necessary changes in the existing Monte Carlo model framework of the experiment. The systematic uncertainties have been investigated and quantified through the variation of the different data selection and fit parameters, through the use of calibration sources and background events outside the region of interest, and through toy Monte Carlo studies, as well as more complex pseudo data analyses that include the full detector response.

The thesis is divided into eight chapters. The first Chapter 1 provides a rough introduction on solar neutrinos, with a focus on the neutrino flavour oscillations, as well as an overview of neutrino production processes inside the Sun and the solar neutrino flux predicted by the Standard Solar Model. It also gives an overview of solar neutrino detectors, to provide a context for the novelty of the analyses in this thesis. The second Chapter 2 describes the Borexino detector, gives an overview of the signal and background components and explains the reconstruction and particle discrimination algorithms used in this thesis. The third Chapter 3 then explains the *Correlated and Integrated Directionality*, as well as the changes introduced in the existing Borexino Monte Carlo simulation.

The next two Chapters 4 and 5 present the measurement of the effective scintillation group velocity and the effective Cherenkov group velocity correction, respectively. These parameters are important for the time-of-flight correction of the detected photon hits, which allows for a time sorting of the Cherenkov and scintillation hits and therefore improve the sensitivity of the analysis. The Chapter 6 provides a first experimental proof-of-principle for the feasibility of a directionality based analysis in a liquid scintillator detector, using the sub-MeV ^{7}Be-neutrinos. This analysis provides important lessons and avenues for improvement, which are implemented in the Chapter 7, where the CNO-neutrino rate is measured. Here, the potential systematic uncertainties are investigated in a more sophisticated and in-depth fashion. The last Chapter 8 summarises the results of this thesis and gives an outlook for the potential future use of the *Correlated and Integrated Directionality* in upcoming or proposed neutrino detectors.

This work has been performed in the context of the *analysis group* of the Borexino collaboration and it therefore includes work by people other than myself. These individuals and their direct contributions are explicitly mentioned at the beginning of each chapter to which they have contributed.

Chapter 1

Solar Neutrinos

The direct study of solar neutrinos has been of interest since at least 1968, where radiochemical experiments [1, 2, 3] proved that fusion provided the energy of the Sun. At the same time, the measured neutrino flux was significantly lower than the Standard Solar Model predictions. This led to the so called *solar neutrino problem* and its subsequent solution through the experimental discovery of neutrino flavour oscillations and a non-zero neutrino mass [4, 5, 6]. Solar neutrinos continue to be of great interest as they stand at the crossroad between astro-, nuclear- and particle physics. On one hand, they can be used as astrophysical messengers, probing the nuclear processes in the deep interior of our Sun in real time. At the same time, the large solar neutrino flux can be used to measure the properties of the neutrinos themselves and to search for physics beyond the Standard Model of particle physics.

This chapter introduces the basic concepts of neutrino physics and the production processes of solar neutrinos, to provide a context for the measurement of solar neutrinos performed in this thesis. First, Section 1.1 gives a short description of neutrinos in the framework of the Standard Model of particle physics, including the neutrino flavour oscillation in vacuum and in matter. The second Section 1.2 explains the principles behind the production of solar neutrinos, and it provides a description of the Standard Solar Model with its predicted neutrino flux. The last Section 1.3 gives an overview of different types of solar neutrino detectors, as the topic of this thesis is the development of a new, hybrid detection method for solar neutrinos.

Although the histories of the Standard Model of particle physics and the Standard Solar Model are both a rich and interesting topic, they will not be discussed here in any detail. A good overview can be found for example in [7], or other textbooks. Likewise, there exist a wide range of different neutrino sources, such as the atmosphere, accelerators, reactors, Supernovae, the earth (geo-neutrinos) and so on. While they all provide a rich field of scientific discovery, they are beyond the scope of this thesis and will therefore not be discussed further.

1.1 Properties of Neutrinos

The existence of neutrinos has been first proposed by Pauli in 1930 [8] as a potential explanation for the continuous energy spectrum of electrons emitted in the β-decay. This was unexpected at the time, as the β-decay was assumed to be a two-body process, which should provide a mono-energetic electron. By adding a third, neutral decay particle with spin $1/2$, the experiment could be made compatible again with the conservation of energy.

In the following years Fermi [9] proposed a first version of the weak interaction framework, including the neutrino. The theory was now in agreement with the experimental data of the time. Fermi already came to the conclusion that the rest mass of the neutrino is either zero or at least very small with respect to the mass of the electron, based on the shape of the electron energy endpoint region[1]. Then, in 1956 Cowan and Reines provided the first observation of

[1] A method that is still used in the current measurements of the neutrino mass in the KATRIN experiment [10].

the (electron anti-)neutrino [11], using a nuclear reactor as a neutrino source, resulting in the award of the Nobel Prize [12]

Since then neutrinos have been observed from a wide range of sources, both natural and man-made, from the Earth itself, the atmosphere, the Sun, from a supernova, nuclear reactors and from particle accelerators. This section is intended to provide a short description of the neutrinos, with a focus on their flavour oscillations and their relationship to solar neutrinos. A deeper introduction to the fundamentals of neutrino physics can be found in textbooks, such as [13], while the particle Data Group [14] provides a comprehensive review of the entire field of particle physics.

1.1.1 Neutrinos in the Standard Model of particle physics

The Standard Model of particle physics [14] (SM) is one of the most successful theories of physics, which elegantly describes three of the four known fundamental forces (the electromagnetic, the weak and the strong interactions) in the framework of quantum field theory. The success of the SM comes from its ability to provide predictions for a variety of particle properties that have been confirmed with remarkable precision. Probably most famously in recent times was its prediction of the Higgs particle, detected in 2012 [15, 16, 17].

The SM is based on the gauge group $SU(3)_{\mathrm{C}} \times SU(2)_{\mathrm{L}} \times U(1)_{\mathrm{Y}}$, where the C, L and Y denote color charge, left-handed chirality and weak hypercharge, respectively. There are the eight massless gluons that mediate strong interactions, corresponding to the eight generators of $SU(3)_C$, while the electroweak interactions are governed by three massive gauge bosons ($W\pm$ and Z^0) and one massless boson (the photon γ).

The interaction of neutrinos is only governed by the electroweak part of the SM, which is mediated either through the weak Charged Current (CC) via the exchange of the $W^{\pm}$ ($m_{\mathrm{W}} \simeq 80\,\mathrm{GeV}$ [14]) bosons or through the weak Neutral Current (NC), via the exchange of the Z^0 boson ($m_{\mathrm{Z}} \simeq 91\,\mathrm{GeV}$ [14]). They gain their mass through their interaction with the Higgs field through the electroweak symmetry breaking. The large masses of the exchange bosons are the reason for the relatively small cross-sections seen in the weak interactions of neutrinos, which are on the order of $10^{-44}\,\mathrm{cm}^{-2}$ at MeV scales (see Figure 2.3 in the next chapter). This necessitates large target masses, to measure even large fluxes, such as is the case for solar neutrinos. Figure 1.1 shows examples of the NC, where no charge is exchanged, and the CC interactions.

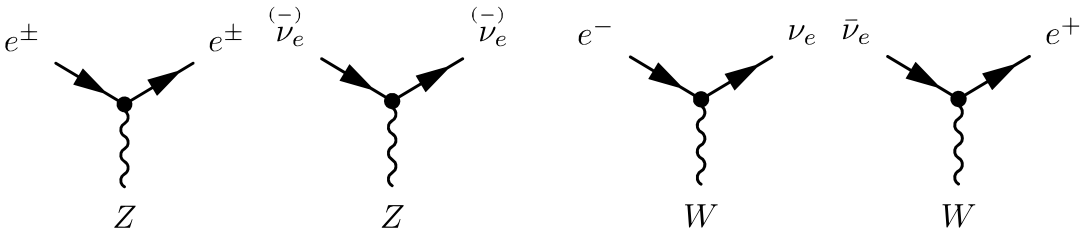


FIGURE 1.1: Examplary trilinear couplings of the weak neutral current (left) and the weak charged current (right), represented by Feynman diagrams.

The decay width of Z^0 depends on the number of light, active (left handed) neutrinos states. Therefore, the measurement of this Z^0 decay width it is possible to infer the number these active neutrino states. At present the measurement provided $N_\nu = 2.984 \pm 0.008$ [14]. Thus, nature provides a number of three generations of fermions, divided in two categories of quarks and leptons, where the neutrinos are named after their corresponding charged lepton:

The gauge bosons Z^0 and $W^{\pm}$ couple only to left-handed particles or right-handed anti-particles, due to the maximal parity violating nature of weak interactions. Therefore, the

Generation	1	2	3
Leptons	e (electron) ν_e (electron-neutrino)	μ (muon) ν_μ (muon-neutrino)	τ (tau) ν_τ (tau-neutrino)

left-handed charged leptons and their corresponding neutrinos form a doublet state of the weak interaction, while the right-handed charged leptons are described by singlet states:

$$L_{L,l} = \begin{pmatrix} \nu_l \\ l \end{pmatrix}_L \qquad E_{R,l} = (l)_R \qquad l = \{e, \mu, \tau\} \tag{1.1}$$

In the SM, the masses of the charged fermions are generated via the Yukawa coupling of the scalar Higgs field to the right-handed charged lepton fields E_R and left-handed lepton doublets L_L of the fermion. However, since the SM does not contain right-handed neutrinos, no Yukawa coupling can be built for the neutrinos. Consequently, neutrinos are truly massless fermions in the SM, where there is neither mixing nor CP violation.

Currently, the absolute neutrino mass is being measured by such experiments as KATRIN [10] or Project8 [18], both of which investigate the β^--decay spectrum of tritium. However, the best current constrain of the neutrino mass comes from the Planck measurements of the cosmic microwave background [19] as $\sum m_\nu < 0.12\,\text{eV}$ (95%). It has to be noted here, that the observation of neutrino flavour oscillations exclude a neutrino mass of exactly zero. The small, but non-zero neutrino mass is in need of an explanation, for which a number of mass models exit [20]. Some mass models make the realisation of a Dirac mass in principle possible, but this would require the existence of right handed neutrino states. These states would be inactive and are called sterile neutrinos. While they would not participate in the weak interaction they would still be accessible through flavour oscillations. Here, the existence of light (eV-scale) sterile neutrinos has been significantly constrained in the recent past [21, 22]. Some other mass models require the neutrino to be a Majorana particle, i.e. its own anti-particle. The experimental search for the Majorana nature of the neutrino is typically performed through the search for the neutrinoless double beta decay, in a number of different experiments and with different candidate isotopes [23].

It is quite clear that neutrinos in general continue to be of great interest as there are still many remaining questions about their nature left open. The experimental evidence for neutrino masses provided an unambiguous signal of new physics. This evidence came through the observation of neutrino flavour oscillations in vacuum and matter, which are described in the following sections.

1.1.2 Neutrino oscillations in vacuum

As mentioned at the beginning of this chapter, the earliest measurements of solar neutrinos were in disagreement with the expectations, only providing about 30%–50% of the predicted electron-neutrino flux. The experimental results of atmospheric neutrinos [4], solar neutrinos [5, 24] and reactor neutrinos [25], and others, have established the neutrino flavour oscillations [26] as the reason for the missing solar neutrinos. While all solar neutrinos are produced as electron-neutrinos, some of them oscillate into muon- or tau-neutrinos, giving rise to the apparent loss of solar neutrinos in the measurement. This phenomenon is energy dependent and requires a non-zero neutrino mass, contrary to the Standard Model.

The oscillations emerge because the neutrinos are produced via the weak-interaction in their flavour eigenstates $|\nu_e\rangle$, $|\nu_\mu\rangle$, $|\nu_\tau\rangle$, which are linear combinations of the Hamiltonian (mass) eigenstates $|\nu_1\rangle$, $|\nu_2\rangle$, $|\nu_3\rangle$. Thus, a flavour eigenstate $|\nu_\alpha\rangle$ is linked to the mass

eigenstates $|\nu_k\rangle$ through the *Pontecorvo-Maki-Nakagawa-Sakata* [27, 28] (PMNS) mixing matrix U, which is a unitary 3×3 matrix:

$$\begin{aligned} |\nu_\alpha\rangle &= \sum_{k=1}^{3} U^*_{\alpha k} |\nu_k\rangle \\ |\nu_k\rangle &= \sum_{\alpha}^{e,\mu,\tau} U_{\alpha k} |\nu_\alpha\rangle \end{aligned} \tag{1.2}$$

The time evolution of the flavour state in vacuum $|\nu_\alpha\rangle$ can be described by the solution of the Schrödinger equation in the plane wave approximation:

$$|\nu_\alpha(t)\rangle = \sum_{k=1}^{3} U^*_{\alpha k} \mathrm{e}^{-iE_k t} |\nu_k\rangle = \sum_{\beta}^{e,\mu,\tau} \left(\sum_{k=1}^{3} U^*_{\alpha k} \mathrm{e}^{-iE_k t} U_{\beta k} \right) |\nu_\beta\rangle \tag{1.3}$$

Here, the unitarity of U is used to express of the pure flavour state $|\nu_\alpha\rangle\,(t=0)$ as a superposition of different flavour states evolving in time. In other words, the different mass eigenstates have a different velocity, resulting in a variation in time of the relative mass phases. Thus, depending on the time of observation, the flavour eigenstate has a certain probability $P(\alpha \to \beta)(t)$ of being detected either as its original flavour at α time of production, or as different flavour β:

$$P(\alpha \to \beta)(t) = \left| \langle \nu_\beta | \nu_\alpha(t) \rangle \right|^2 = \sum_{k,j} U^*_{\alpha k} U_{\beta k} U_{\alpha j} U^*_{\beta j} \mathrm{e}^{-i\left(E_k - E_j\right)t} \tag{1.4}$$

The oscillation probability $P(\alpha \to \beta)(t)$ is only influenced by the relative phase $\Delta E_{kj} t = (E_k - E_j)\,t$ between the different mass eigenstates. In the highly relativistic case the energy E_i can be approximated as:

$$E_i = |\vec{p}_i| \Longrightarrow E_i = \sqrt{p_i^2 + m_i^2} \approx E + \frac{m_i^2}{2E} \tag{1.5}$$

Additionally, the travel time t can be well approximated by the distance traveled L[2]. The oscillation probability can then be written as:

$$P(\alpha \to \beta)(E,L) = \delta_{\alpha\beta} - 4 \sum_{k>j} \Re \left[U^*_{\alpha k} U_{\beta k} U_{\alpha j} U^*_{\beta j} \right] \sin^2 \left(\frac{\Delta m^2_{kj} L}{4E} \right) \tag{1.6}$$

$$+ 2 \sum_{k>j} \Im \left[U^*_{\alpha k} U_{\beta k} U_{\alpha j} U^*_{\beta j} \right] \sin \left(\frac{\Delta m^2_{kj} L}{2E} \right) \tag{1.7}$$

It can be seen that the mass-splitting Δm^2_{kj} must be non-zero for at least one combination of mass eigenstates to allow for oscillations and that L/E is the physical value over which the oscillation can be experimentally accessed. The values of Δm^2_{kj} are realised in nature in such a way, that different experiments are sensitive to different oscillation parameters. A global analysis of neutrino experiments [26] results in mass splittings of $\Delta m^2_{21} = (7.35 \pm 0.17) \cdot 10^{-5}\,\mathrm{eV}^2$ and $\Delta m^2_{31} \approx \Delta m^2_{32}\,(2.455 \pm 0.035) \cdot 10^{-3}\,\mathrm{eV}^2$. The first squared–mass difference Δm^2_{21} is responsible for the oscillation of solar neutrinos, while the oscillation of atmospheric neutrinos is influenced more by Δm^2_{31}. For example, if an experiment is sensitive to the small solar mass splitting Δm^2_{21}, such that $\Delta m^2_{21} L/4E \approx 1$, then the corresponding oscillation phase for Δm^2_{31} is $\Delta m^2_{31} L/4E \gg 1$ and therefore averaged to $\langle \sin^2 x \rangle = 1/2$, independent of L/E.

[2] In natural units

For the opposite case, where $\Delta m^2_{31} L/4E \approx 1$ it also follows that $\Delta m^2_{21} L/4E \ll 1$, meaning that these *solar* terms do not not contribute to the oscillation of atmospheric neutrinos.

For this reason the PMNS-matrix is typically represented as a multiplication of three matrices, that depend on the mixing angles $\theta_{12}, \theta_{13}, \theta_{23}$. Given the assumption that neutrinos are Dirac particles a phase factor δ can introduce a violation of the charge-parity (CP) symmetry, if it is different from zero. Should neutrinos be Majorana particles, i.e. their own antiparticle, two additional phase factors can exist.

$$U = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13} e^{-i\delta_{\mathrm{CP}}} \\ 0 & 1 & 0 \\ -s_{13} e^{i\delta_{\mathrm{CP}}} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{1.8}$$

Here, the mixing angles $\sin\theta_{kj}$, $\cos\theta_{kj}$ are abbreviated as s_{kj} and c_{kj}, respectively. Typically, solar neutrino experiments and short-baseline reactor experiments are sensitive to the oscillation parameters $(\Delta m^2_{21}, \theta_{12})$, while atmospheric neutrino experiments and long-baseline accelerator experiments are sensitive to the oscillation parameters $(|\Delta m^2_{23}|, \theta_{23})$. It is worth to note that matter effects in the Sun make it possible to constrain the sign of the solar mass splitting Δm^2_{21} to be positive, while the sign of the atmospheric mass-splitting is still unknown. Therefore, two possible arrangements of the mass hierarchy are possible: The normal ordering with $m_1 < m_2 < m_3$ or the inverted ordering, with $m_3 < m_1 < m_2$.

It is possible to simplify the three neutrino scenario to describe the neutrino oscillation with an effective model considering only two neutrinos, due to the decoupling effect of the large difference between the mass-splittings Δm^2_{21} and Δm^2_{23}. In this case the mixing matrix U is given by a rotation matrix, with a single angle $\theta = \theta_{12}$. For solar neutrinos this corresponds to the rightmost matrix of Equation 1.8. The probability that a neutrino is detected in the same flavour eigenstate $|\nu_\alpha\rangle$ in which it has been produced is called survival probability. In the simplified two neutrino case it can be written as:

$$P(\alpha \to \alpha)(E,L) = 1 - \sin^2(2\theta) \sin^2\left(\frac{\Delta m^2 L}{4E}\right) \tag{1.9}$$

For solar neutrinos the survival probability due to oscillations in vacuum is effectively averaged out. The reason for this is the large size of L in combination with the detector energy resolution and the relative large size of the fusion zone, where solar neutrinos are produced:

$$\langle P_{\text{solar, vacuum}}(e \to e)\rangle = 1 - \frac{1}{2}\sin^2(2\theta_{12}) \tag{1.10}$$

In Borexino, and in this thesis, the expected rate of solar neutrino events in the detector is calculated using the oscillation parameters in the normal ordering, from the global analysis presented in [26]. The global best fit values are summarised in Table 1.1.

1.1.3 Neutrino oscillation matter effect

Due to their weakly interacting nature, the neutrinos that are produced in the centre of the Sun are well able to escape the dense stellar core and can therefore be used as a direct probe of the nuclear processes in the Sun. At the same time, the solar neutrinos experience an influence of the matter potential on their way out of the Sun, which causes a significant, energy dependent change in the electron-neutrino survival probability.

The idea of neutrino oscillations in matter was first proposed in [29], with additional contributions in [30], resulting in the name of *Mikheyev-Smirnov-Wolfenstein* (MSW) effect, after the authors. Ultimately, the solar neutrino problem has been solved by the combination

Parameter	Ordering	Best fit	1σ range
$\Delta m^2_{21}/10^{-5}\text{eV}^2$	Both	7.34	7.20 - 7.51
$\sin^2\theta_{12}/10^{-1}$	Normal	3.04	2.91 – 3.18
	Inverted	3.03	2.90 – 3.17
$\sin^2\theta_{13}/10^{-2}$	Normal	2.14	2.07 – 2.23
	Inverted	2.18	2.11 – 2.26
$\lvert\Delta m_{13}\rvert/10^{-3}\text{eV}^2$	Normal	2.455	2.423 – 2.490
	Inverted	2.441	2.406 – 2.474
$\sin^2\theta_{23}/10^{-1}$	Normal	5.51	4.81 – 5.70
	Inverted	5.57	5.33 – 5.74
δ_{CP}/π	Normal	1.32	1.14 – 1.55
	Inverted	1.52	1.37 – 1.66

TABLE 1.1: Best fit values and allowed 1σ region of the neutrino oscillation parameters, for the normal and inverted mass ordering. From the global fit performed in [26].

of vacuum oscillations and the MSW effect. This is the so called LMA-MSW (large mixing angle) solution, due to the (unexpectedly) large size of the solar mixing angle θ_{12}.

The MSW effect relies on the fact that ordinary matter is made up of electrons. Therefore, the ν_e can interact with electrons with both the CC and NC, while $\nu_{\mu,\tau}$ can only interact via the NC. This introduces an additional potential for ν_e, that is dependent to the Fermi constant G_F and on the electron density N_e of the matter:

$$V_{\text{CC}} = \sqrt{2} G_F N_e \tag{1.11}$$

In the following, the MSW effect is illustrated in the two neutrino (ν_e, ν_μ) approximation. The evolution of a neutrino state with initial flavour α is described by the Schrödinger equation:

$$i\frac{d}{dt}\begin{pmatrix}\nu_1\\ \nu_2\end{pmatrix} = \frac{1}{2E}\begin{pmatrix}m_1^2 & 0\\ 0 & m_2^2\end{pmatrix}\begin{pmatrix}\nu_1\\ \nu_2\end{pmatrix} = H^k\begin{pmatrix}\nu_1\\ \nu_2\end{pmatrix} \tag{1.12}$$

Here H^k is the diagonal Hamiltonian matrix (mass matrix), meaning the mass eigenstates in vacuum are eigenstates of H^k. Using the mixing matrix U, the Hamiltonian matrix can be written in the representation of flavour eigenstates with the mixing angle $\theta_{12} = \theta$:

$$H^\alpha = UH^kU^\dagger = \frac{1}{2E}\begin{pmatrix}m_1^2\cos^2\theta + m_2^2\sin^2\theta & \left(m_2^2 - m_1^2\right)\sin\theta\cos\theta\\ \left(m_2^2 - m_1^2\right)\sin\theta\cos\theta & m_1^2\sin^2\theta + m_2^2\cos^2\theta\end{pmatrix} \tag{1.13}$$

In the presence of matter the mass of ν_e is modified trough an additional potential with a value of V_{CC}:

$$H_m^\alpha = H^\alpha + \begin{pmatrix}V_{\text{CC}} & 0\\ 0 & 0\end{pmatrix} \tag{1.14}$$

Transforming this matrix back into the mass representation (ν_1, ν_2) results in:

$$H_m^k = U^\dagger H_m^\alpha U = \frac{1}{2E}\begin{pmatrix}m_1^2 + 2\cos^2\theta EV_{\text{CC}} & 2\cos\theta\sin\theta EV_{\text{CC}}\\ 2\cos\theta\sin\theta EV_{\text{CC}} & m_2^2 + 2\sin^2\theta EV_{\text{CC}}\end{pmatrix} \tag{1.15}$$

The neutrino mass eigenstates in matter (ν_{1m}, ν_{2m}) are now different from the mass eigenstates in the vacuum, due to the coherent forward-scattering off electrons. The additional on-diagonal terms will alter the effective mass squared difference and thus change the oscillation lengths, while the off-diagonal allows the transition between mass eigenstates. However, if the electron density n_e is constant or varies only slowly (adiabaticity) then the so called crossing probability is negligible. Therefore, the neutrino oscillations can now be described in terms of the effective matter eigenvalues (m^2_{1m}, m^2_{2m}), their mass splitting Δm^2_m and the effective mixing angle θ_m:

$$m^2_{1m,2m} = \frac{1}{2}\left[m_1^2 + m_2^2 + EV_{\mathrm{CC}} \mp \sqrt{\left(EV_{\mathrm{CC}} - \Delta m^2_{21}\cos 2\theta\right)^2 + \left(\Delta m^2_{21}\sin 2\theta\right)^2}\right] \tag{1.16}$$

$$\Delta m^2_m = m^2_{2m} - m^2_{1m} = \sqrt{\left(EV_{\mathrm{CC}} - \Delta m^2_{21}\cos 2\theta\right)^2 + \left(\Delta m^2_{21}\sin 2\theta\right)^2} \tag{1.17}$$

$$\sin 2\theta_m = \frac{\Delta m^2_{21}\sin 2\theta}{\Delta m^2_m} \tag{1.18}$$

It can be seen from the above equations, that for very small neutrino energies the matter effects can become negligible. In the Sun this is the case for neutrino energies below 1 MeV, where the survival probability is given by the vacuum oscillation. Furthermore, the condition $EV_{\mathrm{CC}} = \Delta m^2_{21}\cos 2\theta$ results in a matter enhanced resonance, maximising the mixing angle $\sin 2\theta_m = 1$. What is most interesting in the context of solar neutrinos is that for large values of EV_{CC} the mixing angle can become very small $\sin 2\theta_m \to 0$ as the effective mass splitting in matter is $\Delta m^2_m \propto EV_{\mathrm{CC}}$. This is the case in the centre of the Sun for neutrino energies above 10 MeV. These neutrinos are now produced in the electron flavour, but since $\theta_{12} \approx 33^\circ$ they are associated entirely to the mass eigenstate ν_{2m}. The neutrino then leaves the Sun in the corresponding vacuum mass eigenstate ν_2, as the density profile of the Sun changes adiabatically. Because ν_μ has a larger admixture of ν_2 than ν_e (or vice versa ν_2 has a larger admixture of ν_μ than of ν_e) this has the consequence that the electron neutrino survival probability P_{ee} is now decreased due to the matter effect. It follows for sufficiently large neutrino energies (in the two neutrino approximation):

$$P_{ee} = \left|U^2_{12}\right| = \sin^2\theta_{12} \tag{1.19}$$

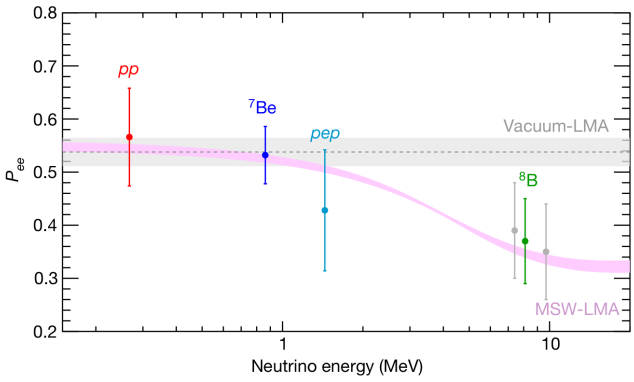


FIGURE 1.2: The electron-neutrino survival probability P_{ee} as a function of the neutrino energy. The vacuum oscillation is shown as a grey band, while the MSW-LMA prediction is shown as a pink band. The data points show the results of the Borexino measurement of solar neutrinos at different energies, named after their production process. Taken from [31].

This is illustrated in Figure 1.2, where P_{ee} is plotted against the solar neutrino energy. The vacuum oscillation of the LMA parameters is shown as a grey band around $P_{ee} \sim 54\%$, while the MSW effect introduces a transition region between $\sim$ 1 MeV-10 MeV, after which the matter effect defines the value of $P_{ee} \sim 33\%$. The data points show the result of the Borexino neutrino spectroscopy [31], that is well in agreement with the MSW-LMA solution.

Additionally, the MSW effect also plays a role for the neutrino propagation through the matter of the Earth [32]. The electron neutrino flux is expected to be partially regenerated, leading to an increase of the measured neutrino interaction rates during night time. The Earth matter effect is small and proportional to the neutrino energy; for 10 MeV neutrinos it regenerates on average about $\sim 3\%$ of the ν_e flux in the detector at night.

1.2 Standard Solar Model

Standard Solar Models (SSMs) constitute a well-defined framework that provides a description of the Sun at the present time, from which predictions can be made. They provide a well defined reference for different fields of research, ranging from solar and stellar modeling to solar neutrinos and particle physics. This section provides a short introduction to the SSM, as it is used for the calculation of the expected fluxes of the solar neutrinos. More details can be found in [7, 33, 34, 35].

The actual predictions of the SSM require detailed model calculations, using underlying assumption of the stellar evolution, summarised as follows:

- The Sun is in a hydrostatic equilibrium, meaning that the radiative and particle pressure of the model is balanced by the gravitational attraction.
- The energy transport is dominated by radiative processes for the interior of the Sun and by convection for its outer layer. Additional transport due to acoustic or gravity waves are considered negligible in the SSM.
- The primary energy source of the Sun is the nuclear fusion of four protons into a Helium nucleus, which produces the observable solar luminosity and neutrinos. Typically, small effects of contraction or expansion are also included in the SSM.
- The primordial solar interior is assumed to be chemically homogeneous and changes in the local abundances of individual isotopes occur only by nuclear reactions. This implies that mass loss is negligible during the life time of the Sun.

Given these assumptions, the SSM is then the result of the evolution of a $1\,M_\odot$ stellar model up to the present solar age $\tau_\odot = 4.57\,\text{Gyr}$. The SSM is required to match the solar luminosity $L_\odot$, the solar radius $R_\odot$, and the surface (photospheric) metal[3]-to-hydrogen ratio $(Z/X)_\odot$, at the present time $\tau_\odot$. The calibration of the SSM is done by adjusting the mixing length parameter (α_{MLT}) and the initial helium (Y_{ini}) and metal (Z_{ini}) mass fractions. These three parameters are correlated with each other, as they all depend on the three observational constraints. The SSM is of importance not only for the understanding of solar physics but for astrophysics in general, as it constitutes a benchmark for stellar evolution models. Additionally, the SSMs depend on other physical inputs, such as the radiative opacities, cross sections of nuclear reactions and others.

1.2.1 Energy and neutrino production in the Sun

The Sun is powered by two sets of nuclear reactions, where four protons are fused into one Helium nucleus, which releases an energy of 26.7 MeV. These sets are the pp-chain [36] and the CNO-cycle[4] [37, 38, 39]. The solar neutrino fluxes can be predicted from the SSM by estimating these nuclear reaction rates. They depend on the fusion cross sections, temperature,

[3] Here every element heavier than helium is called a metal, as is usual in Astronomy.
[4] Also known as Bethe–Weizsäcker cycle.

nuclei density, and chemical composition inside the Sun. In general, the interaction rate R between two nuclei can be written as [7]:

$$R = \frac{n_a n_b}{1+\delta_{ab}} \langle \sigma v \rangle \tag{1.20}$$

Here, n_a, n_b are the particle densities, σ the energy dependent cross-section, v their relative velocity and δ_{ab} is used to avoid double counting of identical nuclei. The temperature averaged product $\langle \sigma v \rangle$ assumes a Maxwell-Boltzmann distribution. This does not provide the particles with enough energy to overcome the Coulomb barrier, given thermal energies of several keV inside the Sun. Thus, the fusion reactions are possible thanks to quantum tunneling, which provides an energy dependent probability to cross the Coulomb potential of the respective fusion reaction. The cross sections σ are of fundamental importance as input parameters for the SSM, but they are hard to measure at the solar energies due to their small size. Therefore, they are typically extrapolated through Taylor expansion from higher energy measurements, or through *ab initio* calculations, when possible. See [40] for more details on the solar fusion cross sections. Due to this dependence on the particle densities and their temperature, the different species of solar neutrinos show different radial distributions in their production rate. All solar neutrinos are produced well within a radius of $0.3 R_\odot$, while neutrinos produced in fusion processes involving heavier nuclei tend to have a radial production rate distribution that is shifted closer to the centre, i.e. to higher temperatures and densities.

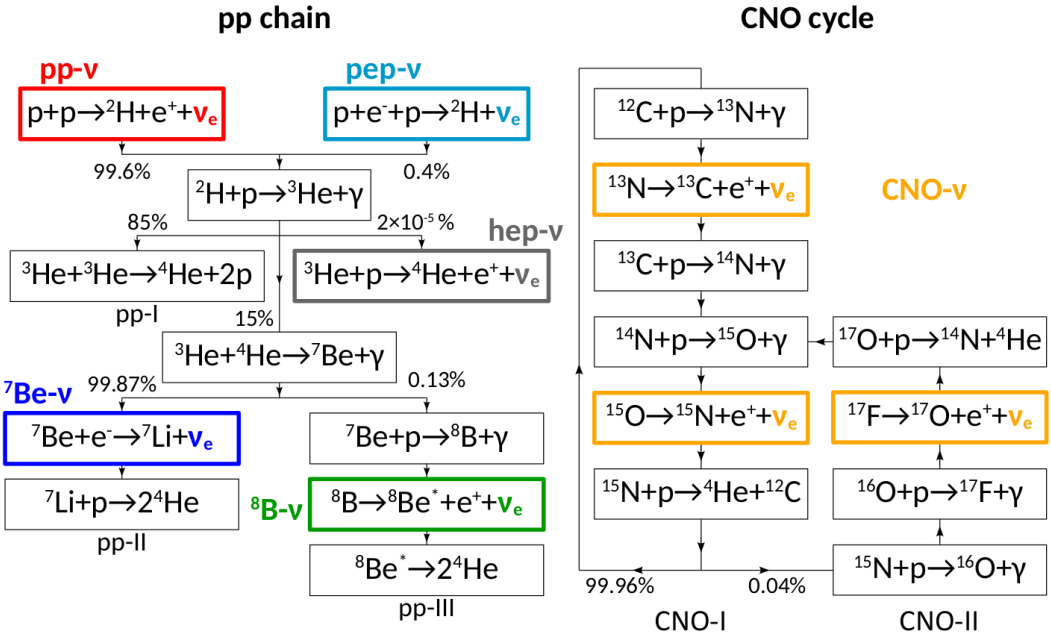


FIGURE 1.3: The schematic overview of the Hydrogen fusion from the pp-chain (left) and the CNO-cycle (right). Taken from [31].

The pp-chain

The pp-chain is responsible for about $\sim 99\%$ of the total energy production in the Sun, while the CNO-cycle accounts for the remaining $\sim 1\%$. It is shown schematically in 1.3 on the left, where the neutrinos are named after the nuclei involved their production. The first reaction step, and basis of the entire pp-chain, is the fusion of protons into a deuteron, either through pp (99.6%, $E_\nu \leq 0.42\,\text{MeV}$) or pep (0.4%, $E_\nu = 1.44\,\text{MeV}$) fusion:

$$\textbf{pp-}\nu: \qquad p + p \rightarrow {}^2\text{H} + e^+ + \nu_e \tag{1.21}$$

$$\textbf{pep-}\nu: \quad p + e^- + p \rightarrow {}^2\mathrm{H} + \nu_e \tag{1.22}$$

The ^{2}H then absorbs a proton and afterwards the reaction chain divides into branches that are called pp-I, pp-II, pp-III and hep, where only the latter three produce neutrinos. In the pp-I branch two ^{3}H fuse directly into ^{4}H and two protons are emitted. For the pp-II branch a ^{3}H and ^{4}H nucleus produce ^{7}Be, which subsequently decays through electron capture into ^{7}Li and emits a neutrino:

$$\textbf{}^{\mathbf{7}}\textbf{Be-}\nu: \quad {}^7\mathrm{Be} + e^- \rightarrow {}^7\mathrm{Li} + \nu_e \tag{1.23}$$

The ^{7}Be decays with a probability of 89.48% into the ground state of ^{7}Li, emitting a mono-energetic neutrino with $E_\nu = 0.862\,\mathrm{MeV}$. For the remaining 10.52% it decays to the first excited state ^{7}Li* with a subsequent emission of a neutrino with $E_\nu = 0.384\,\mathrm{MeV}$ and a 0.478 MeV γ-ray. The pp-II branch then terminates trough the production of two ^{4}H via $^7\mathrm{Li} + p \rightarrow 2\,{}^4\mathrm{H}$.

The pp-III branch starts similar to the pp-II branch, but here the ^{7}Be captures a proton, producing a ^{8}B. This ^{8}B is unstable and its β^+–decay provides another neutrino with $E_\nu \leq 15\,\mathrm{MeV}$:

$$\textbf{}^{\mathbf{8}}\textbf{B-}\nu: \quad {}^8\mathrm{B} \rightarrow {}^8\mathrm{Be}^* + e^+ + \nu_e \tag{1.24}$$

The excited ^{8}Be* then dissociates into two ^{4}H, closing the pp-III chain. At last, ^{4}H is also produced with a very low fraction of $2 \cdot 10^{-5}$ by a proton capture on ^{3}H, which produces a neutrino with $E_\nu \leq 18.77\,\mathrm{MeV}$:

$$\textbf{hep-}\nu: \quad {}^3\mathrm{H} + p \rightarrow {}^4\mathrm{H} + e^+ + \nu_e \tag{1.25}$$

The CNO-cycle

The CNO-cycle is a catalytic process, which relies on the presence of C, N and O in the Sun. This is schematically shown in Figure 1.3 on the right. Almost all the energy produced by the CNO catalysed fusion is due to the CNO-I cycle (CN-cycle) with a small contribution of the CNO-II cycle (NO-cycle). The CNO-I cycle consists of a loop of reactions that starts with the proton capture on ^{12}C and ends in the proton capture on ^{15}N, again producing a ^{12}C and a ^{4}H. The sub-dominant CNO-II cycle starts with the proton capture on ^{15}N and ends with the proton capture on ^{17}O and the production ^{14}F and ^{4}H.

$$\textbf{CNO-I}: \quad {}^{12}\mathrm{C} + p \rightarrow {}^{13}\mathrm{N} + \gamma \tag{1.26}$$

$$ {}^{13}\mathrm{N} \rightarrow {}^{13}\mathrm{C} + e^+ + \nu_e \quad (E_\nu \leq 1.20\,\mathrm{MeV}) \tag{1.27}$$

$$ {}^{13}\mathrm{C} + p \rightarrow {}^{14}\mathrm{N} + \gamma \tag{1.28}$$

$$ {}^{14}\mathrm{N} + p \rightarrow {}^{15}\mathrm{O} + \gamma \tag{1.29}$$

$$ {}^{15}\mathrm{O} \rightarrow {}^{15}\mathrm{N} + e^+ + \nu_e \quad (E_\nu \leq 1.73\,\mathrm{MeV}) \tag{1.30}$$

$$ {}^{15}\mathrm{N} + p \rightarrow {}^{12}\mathrm{C} + {}^4\mathrm{He} \tag{1.31}$$

$$\textbf{CNO-II}: \quad {}^{15}\mathrm{N} + p \rightarrow {}^{16}\mathrm{O} + \gamma \tag{1.32}$$

$$ {}^{16}\mathrm{O} + p \rightarrow {}^{17}\mathrm{F} + \gamma \tag{1.33}$$

$$ {}^{17}\mathrm{F} \rightarrow {}^{17}\mathrm{O} + e^+ + \nu_e \quad (E_\nu \leq 1.74\,\mathrm{MeV}) \tag{1.34}$$

$$ {}^{17}\mathrm{O} + p \rightarrow {}^{14}\mathrm{F} + {}^4\mathrm{He} \tag{1.35}$$

$$ \tag{1.36}$$

The CNO-I cycle (CN-cycle) conserves the number abundance of the solar metals, but changes their distribution as it eventually achieves an equilibrium. In the Sun this leads

to the conversion of most of the primordial ^{12}C in the core into ^{14}N. This change in the chemical composition over time has an impact on the SSM as it changes, for example, the core opacity and the heavy element mass fraction Z. The CNO-cycle requires preexisting metals (C, N and O) to fuse the protons into ^{4}He, where the reactions are characterised by larger Coulomb barriers than for the pp-chain. As such, the CNO-cycle energy production has a much steeper temperature dependence compared to the pp-chain. For the CNO-cycle the energy dependence of the ^{4}He production rate is proportional $\propto T^{17}$, while for the pp-chain it is $\propto T^{4}$ [40]. Consequently, while the CNO-cycle is sub-dominant in the Sun, it dominates the energy production in heavier main-sequence stars.

1.2.2 Solar neutrino flux predictions

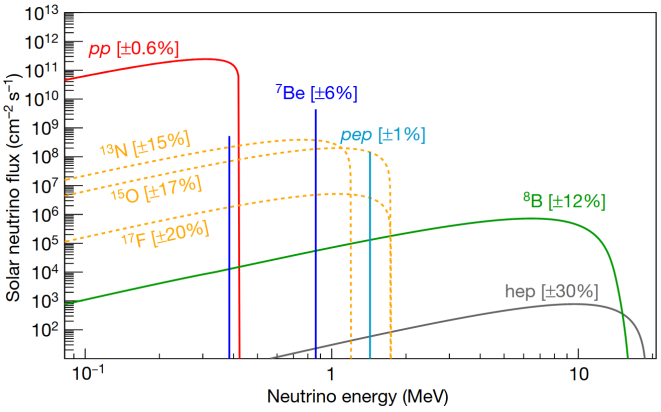


FIGURE 1.4: The energy spectra of the solar neutrinos [41] with the updated fluxes from [35]. CNO is the sum of the ^{13}N, ^{15}O and ^{17}F neutrinos. Taken from [31].

The SSM provides a complete description of the Sun, after its parameters have been correctly calibrated. One of its most important predictions is the production rate of solar neutrinos from the nuclear reactions, described in the previous section. Figure 1.4 shows the (un-oscillated) energy spectrum of the expected solar neutrino flux at the earth according to the SSMs of [35], with the uncertainty of the SSM prediction shown in the brackets. The constituents of the CNO-cycle are shown as dotted, yellow lines, while the neutrinos from the pp-chain are shown as solid, coloured lines. The pep- and ^{7}Be-neutrinos can be seen as mono-energetic lines.

As stated in the previous section, the SSM depends on the abundance of chemical elements in the Sun. These abundances are determined primarily through the spectroscopy of the solar photo-sphere. For this type of measurement it is necessary to model the solar atmosphere, to determine its radial temperature and density distribution, as well as performing detailed radiation transfer calculations. It is then then possible to infer the elemental abundances from the spectral line distributions. New three-dimensional, time-dependent hydro-dynamical modeling of the solar atmosphere [42] lead to lower estimations of the solar metallicity by about $\sim 20\%$, compared to previous estimations [34, 43].

Among other things, the solar metallicity also impacts the opacity and the sound speed profile of the Sun. Here, helioseismology [44] provides a tool for the investigation of the interior properties of the Sun, where the measurement of the global acoustic eigenmodes has allowed the reconstruction of the interior structure of the Sun with great precision. While the older, high-metallicity SSM predictions are in agreement with the sound speed profiles inferred from helioseismic data, the newer, low-metallicity SSM predictions show some tension. This is sometimes called the *solar metallicity puzzle*. A higher metal abundance increases the opacity and therefore the temperature of the Sun's interior, which in turn is correlated to the production rate of the solar neutrinos. Therefore, a direct measurement of the CNO-ν flux is

interesting because it is correlated directly to the solar metallicity and can help to provide a solution to the metallicity puzzle [45].

The Borexino measurements are typically compared to, or constrained by, the two different SSM predictions presented in [35], which is also the case for this thesis. The high-metallicity model B16-GS98 (HZ-SSM) is calculated with a metal abundance of $(Z/X)_{\odot} = 0.02292$, while the low-metallicity model B16-AGSS09met (LZ-SSM) has $(Z/X)_{\odot} = 0.01780$. Here, the metal abundance in the Sun has been calculated with information from the composition of meteorites in addition to the photospheric measurements. The predictions for the solar neutrino fluxes (without oscillations) of these two models is shown in Table 1.2:

Flux [$cm^{-2}s^{-1}$]	HZ-SSM (B16-GS98) [cpd / 100 t]	LZ-SSM (B16-AGSS09met) [cpd / 100 t]	(HZ-LZ) / HZ [%]
pp	$5.98\ (1\pm0.006)\cdot10^{10}$	$6.03\ (1\pm0.005)\cdot10^{10}$	-0.84
pep	$1.44\ (1\pm0.01)\cdot10^{8}$	$1.46\ (1\pm0.009)\cdot10^{8}$	-1.39
hep	$7.98\ (1\pm0.30)\cdot10^{3}$	$8.25\ (1\pm0.30)\cdot10^{3}$	-3.38
^{7}Be	$4.93\ (1\pm0.06)\cdot10^{9}$	$4.50\ (1\pm0.06)\cdot10^{9}$	$+8.72$
^{8}B	$5.46\ (1\pm0.12)\cdot10^{6}$	$4.50\ (1\pm0.12)\cdot10^{6}$	$+17.6$
^{13}N	$2.78\ (1\pm0.15)\cdot10^{8}$	$2.04\ (1\pm0.14)\cdot10^{8}$	$+26.6$
^{15}O	$2.05\ (1\pm0.17)\cdot10^{8}$	$1.44\ (1\pm0.16)\cdot10^{8}$	$+29.8$
^{17}F	$5.29\ (1\pm0.20)\cdot10^{6}$	$3.26\ (1\pm0.18)\cdot10^{6}$	$+38.4$
CNO	$4.88\ (1\pm0.16)\cdot10^{8}$	$3.51\ (1\pm0.19)\cdot10^{8}$	$+28.1$

TABLE 1.2: The predicted neutrino fluxes of the HZ-SSM and LZ-SSM according to [35]. CNO is the sum of the ^{13}N, ^{15}O and ^{17}F fluxes.

1.3 Solar Neutrino Detectors

While the Sun provides the highest natural neutrino flux on Earth, the detection of neutrinos is extremely challenging because of the very low reaction cross sections provided by the weak interaction. Thus, the measurements of solar neutrinos requires large detectors to gather enough statistics, as well as sufficiently low background levels. For this reason the typical neutrino detector is situated deep underground, to shield from cosmic rays and cosmic muon induced background. This section is intended to provide an overview of such solar neutrino detectors, starting from the historical radio-chemical experiments, to the current water Cherenkov and liquid scintillator detectors , to proposed future hybrid detectors, using both scintillation and Cherenkov radiation. The intention here is to provide a context for the directional and spectral measurement of solar neutrino detection and to highlight future experiments that may benefit from the new analysis method developed in this thesis. An overview of the potential, future use of noble liquid (Xenon, Argon) experiments for the detection of solar neutrinos is left out here, but can be found for example in [46, 47]. A good, recent review on the future of solar neutrino detection in general can be found in [48].

1.3.1 Radio-chemical detectors

As mentioned previously, the radio-chemical experiments constituted the first generation of solar neutrino detectors and were critical in the understanding of the solar neutrino problem. Here, the neutrinos are detected through the CC reaction on a specifically selected target

isotope, which produces a radioactive daughter nucleus. This daughter nucleus is then chemically extracted from the target and can then be detected through its own decay signature. The Homestake experiment [2], named after its location in the Homestake gold mine in the USA, used the inverse beta decay of ^{37}Cl:

$$^{37}\mathrm{Cl} + \nu_e \rightarrow {}^{37}\mathrm{Ar} + e^- \tag{1.37}$$

This reaction has an energy threshold of 814 keV and the ^{37}Ar has a lifetime of $\tau_{^{37}\mathrm{Ar}} \approx 35$ d. It was chemically extracted every two months and detected through the ^{37}Ar electron decay in small proportional counters. The Homestake experiment in particular gave rise to the solar neutrino problem and operated between 1967 and 1994. It measured an average neutrino capture rate of $2.56 \pm 0.16(\mathrm{stat.}) \pm 0.16(\mathrm{syst.})$ SNU, compared to the 9.3 ± 1.3 SNU predicted by the SSM of that time [2]. The SNU stands for Solar Neutrino Unit, which is a convenient unit of measurement corresponding to 10^{-36} neutrino captures per target nucleus and second. The principle detection of solar neutrinos was awarded with a part of the Nobel prize in 2002 to R. Davis [3].

The Homestake experiment was then followed by two experiments using the inverse beta decay of ^{71}Ga: The first was SAGE (Soviet American Gallium Experiment) [49], which began its activity in 1989 at the Baksan Neutrino Observatory in Russia. While it is still running, it is renamed and is focused on the so-called gallium anomaly nowadays [50]. The second was GALLEX (GALLium EXperiment) [51], which ran at the INFN Gran Sasso National Laboratory in Italy from 1991 to 1997.

$$^{71}\mathrm{Ga} + \nu_e \rightarrow {}^{71}\mathrm{Ge} + e^- \tag{1.38}$$

The detection threshold for ^{71}Ge of 233 keV is significantly lower than for ^{37}Cl and well below the 422 keV end point of the pp-neutrino spectrum. For these experiments the measured neutrino signal was once again smaller than the SSM prediction at about $\sim 60\%$.

The radio-chemical detection method can only provide an integrated measurement of the solar neutrino flux above a certain energy threshold, but it is not able to separate the contributions of the different solar neutrino spectra. Today the radio-chemical detection of solar neutrinos is still an area of ongoing research; an example being the lorandite experiment (LOREX) [52]. Here, the proposed use of the isotope ^{205}Tl provides a threshold value of only 52 keV. The daughter isotope ^{205}Pb has a half-life of several million years and could thus provide the possibility to measure the average solar neutrino flux over the last million years.

1.3.2 Water Cherenkov detectors

A breakthrough in solar neutrino physics came from the advent of purpose-built water Cherenkov detectors, in particular the Sudbury Neutrino Observatory (SNO) [53] situated in a nickel mine in Canada, which was initiated in 1984 explicitly to solve the solar neutrino problem. The central element of the detector were 1000 t of heavy water ($^2\mathrm{H_2O}$, $\mathrm{D_2O}$) in a 12 m diameter acrylic vessel. The use of heavy water allowed for three different interaction channels for the solar neutrinos:

$$\text{Neutral Current (NC):} \quad \nu_X + d \rightarrow \nu_X + p + n \tag{1.39}$$

$$\text{Elastic Scattering (ES):} \quad \nu_X + e^- \rightarrow \nu_X + e^- \tag{1.40}$$

$$\text{Charged Current (CC):} \quad \nu_X + d \rightarrow e^- + p + p \tag{1.41}$$

While the NC and ES interactions are sensitive to all neutrino flavours, the CC interaction is sensitive only to electron neutrinos. This way it was possible to directly prove that the total measured neutrino rate was in good agreement with the predictions of the SSM, but only a

fraction of the solar neutrinos conserved their electron flavour. This experimental evidence led to the Nobel Prize in 2015 to A.B. McDonald [6], together with T. Kajita and the associated results of the Super-Kamiokande experiment.

The fast moving electrons produced in the CC and ES reactions provide Cherenkov radiation which was detected with $\sim$ 9500 20 cm photo-multiplier tubes (PMTs). The NC reaction was detected in three different ways for separate phases of the experiment. The neutrons of the NC reaction were captured on deuterium in Phase 1 and on chlorine from dissolved NaCl in Phase 2. The subsequent emission of γ-rays provided Cherenkov light that could be detected, due to Compton scattered electrons. In Phase 3 the NC produced neutrons were detected by an array of ^{3}He-filled neutron counters. The principal results of SNO clearly showed that the solar ^{8}B-neutrinos change their flavour in transit to Earth.

The infrastructure of SNO has been reused for the SNO+ experiment [54], which started data taking in 2017. In the first operating phase of the experiment the detector was filled with light water (H_2O) and a second phase has started, which uses 780 t of liquid scintillator. For the third phase it is planned to load the scintillator with 3.9 t of tellurium. The main goal of SNO+ is the search for neutrinoless double beta decay, the detection of which would prove the Majorana nature of neutrinos, meaning that the neutrino is its own antiparticle. In addition, the physics program also includes the studies of geo- and reactor antineutrinos, supernova and solar neutrinos, and exotic physics.

The second important water Cherenkov detector in the context of solar neutrinos is Super-Kamiokande [55], the successor of Kamiokande (Kamioka Nucleon Decay Experiment). Kamiokande was initially built to search for the proton decay, but it was able to provide a direct proof of the solar origin of the detected neutrinos [56] as well as a detection of neutrinos from the supernova 1987A [57]. For this achievement a part of the Nobel prize was awarded in 2002 to M. Koshiba [3].

Super-Kamiokande is a large, cylindrical, water Cherenkov detector, containing 50 kt of ultra-pure water, instrumented with $\sim$ 12 000 PMTs, that started data taking in 1997. Here, the detection channel for solar neutrinos is only given by the ES reaction (Equation 1.40). The direction of each event can be reconstructed, based on the PMT hit pattern of the recoil-electron Cherenkov cone. As the direction of the recoil-electron is correlated to the direction of the solar neutrinos it is possible to efficiently differentiate between the background events and the solar neutrino signal events.

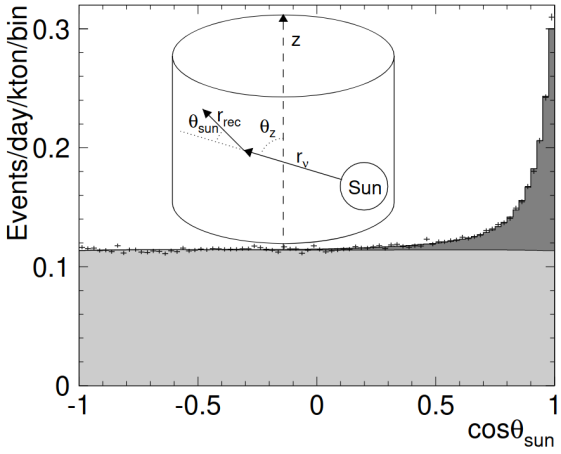


FIGURE 1.5: Exemplary angular event distribution in Super-Kamiokande for 3.5 MeV to 19.5 MeV of kinetic energy of the recoil-electron. The distribution of the angle $\cos\theta_{\text{sun}}$ between the reconstructed event direction $\vec{r}_{\text{rec}}$ and the incoming neutrino direction $\vec{r}_{\nu}$. The data is shown as black points, while the best fit is shown as a histogram. The neutrino and background contributions are shown as dark and light shaded areas, respectively. Taken from [58]

This is illustrated in Figure 1.5 from [58], which shows the angular distribution $\cos\theta_{\text{sun}}$ between the reconstructed event direction $\vec{r}_{\text{rec}}$ and the solar neutrino direction $\vec{r}_{\nu}$. It can be clearly seen that the background (light shaded area) follows a uniform distribution, while the solar neutrino signal (dark shaded area) shows a peak at $\cos\theta_{\text{sun}} = 1$, smeared by the

resolution of the event direction reconstruction. A typical analysis in water Cherenkov detectors is performed on this angular distribution, binned for a number of energy regions to make use of both the directional and spectral information of the solar neutrinos.

Both SNO and Super-Kamiokande provide a low energy threshold for their solar neutrino analysis at 3.5 MeV [58, 59], for the kinetic energy of the recoil-electrons. This is due to the combination of the relatively low light-yield of the Cherenkov radiation at these energies, with only about 30 detected photo-electrons per event, and the presence of a residual radioactive background in the water. Therefore, water Cherenkov detectors are only sensitive to the high energy part of the solar neutrino spectrum and have only measured ^{8}B neutrinos, for which Super-Kamiokande provides the most precise flux measurement [58] to date.

In addition to the solar neutrino program Super-Kamiokande famously provided evidence for the neutrino oscillation through the measurement of atmospheric neutrinos [4], for which a Noble prize was awarded in 2015 to T. Kajita [6]. Here, the data exhibited a zenith angle dependent deficit of muon neutrinos that was inconsistent with the expectations while it was consistent with the two-flavour $\nu_\mu \leftrightarrow \nu_\tau$ oscillations. Furthermore, Super-Kamiokande was also used as the far detector for the investigation of accelerator neutrinos from the long-baseline K2K [60] and T2K [61] experiments, the latter of which is still ongoing.

The next generation of the water Cherenkov concept will be the Hyper-Kamiokande detector [62], south of the Super-Kamiokande site. It is expected to start data taking in 2027 [63]. It will be filled with 260 kt of ultra-pure water, providing a ~ 8 times larger fiducial volume than Super-Kamiokande, and it will be instrumented with 40 000 PMTs. The enormous scale of Hyper-Kamiokande will allow it for a rich physics program, ranging from oscillation studies from accelerator, atmospheric and solar neutrinos, sensitivity studies on the proton decay, the search for astrophysical neutrinos, and so on. The search for a CP-violating phase, visible through asymmetries in accelerator neutrino and antineutrino oscillations, is quoted as one of the major goals of Hyper-Kamiokande.

1.3.3 Liquid scintillator detectors

Liquid scintillator (LS) detectors are the second workhorse of the current generation of solar neutrino experiments. Charged particles deposit their kinetic energy in the LS, mostly through ionisation, which excites the LS molecules, resulting in the emission of isotropic scintillation light. Typically the LS consists of a solvent, for example pseudocumene or linear alkylbenzene, and a secondary fluorophore [5] at a concentration at about 1.5 g/l to 2.5 g/l. This secondary fluorophore is called a wavelength shifter, as it absorbs the energy from the solvent and then emits the scintillation light at a wavelength outside of the absorption region of the solvent. This increases the attenuation length of the scintillation light drastically and makes it possible to construct large scale LS detectors. Thus, the LS provides a relatively high light-yield per deposited energy, compared to the light-yield of the Cherenkov process. The main detection channel of solar neutrinos is again their ES off the target electrons, but now the information on the neutrino energy is provided by the scintillation. The high light-yield of the LS allows for a relatively low energy threshold, as well as a high precision position and energy reconstruction, compared to the water Cherenkov detectors. The solar neutrinos are then typically measured through a fit of the detected energy spectrum, but without the directional information from the Cherenkov radiation.

Furthermore, the high light-yield also allows for the efficient detection of electron anti-neutrino events, which, for example, are produced by nuclear reactors or by the decay of ^{238}U and ^{232}Th within the Earth, providing the so called geo-neutrinos. The detection method for electron anti-neutrinos is based on the inverse beta decay and the coincidence of the prompt

[5] Often called a *fluor* in the context of scintillation experiments.

positron signal and the delayed signal ($\sim 260\,\mu$s) of a neutron capture hydrogen, which then emits a 2.2 MeV γ-ray:

$$\bar{\nu}_e + p \rightarrow e^+ + n \tag{1.42}$$

Since the produced neutron is heavier than the target proton, the inverse beta decay has a kinematic threshold of 1.8 MeV.

The KamLAND (Kamioka Liquid Scintillator Anti-Neutrino Detector) experiment started operation in 2002, and is situated in the Kamioka Observatory; same as Super-Kamiokande. The detector consists of 1000 t of LS, contained in a transparent nylon-vessel. The scintillation light is detected by an array of about 1800 17" and 20" PMTs, mounted on the inner surface of an 18 m stainless steel sphere. It is also surrounded by a 3.2 kt water Cherenkov detector, which tags cosmic muons and absorbs γ-rays and neutrons from the detector structure and the surrounding rock. One of the main achievements of the experiment was the precise measurement of the Δm^2_{12} mass splitting, as well as improving the precision of θ_{12} in combination with the solar neutrino data. This was done through the detection of the high statistics, distorted reactor (electron) anti-neutrino spectrum at a 180 km baseline, which allowed to establish the LMA-MSW neutrino oscillation mechanism as the solution to the solar neutrino problem [25]. Additionally, the anti-neutrino spectrum allowed KamLAND to perform the first experimental study of geo-neutrinos [64]. In the context of solar neutrinos KamLAND performed a measurement of ^{7}Be-neutrinos [65] and ^{8}B-neutrinos [66]. It suffered from an unfortunately high contamination of radioactive background, specifically ^{210}Bi, which made the ^{7}Be-neutrino measurement less precise by almost an order of magnitude compared to the Borexino result. The current iteration of the experiment is called KamLAND-Zen [67], where the LS now is loaded with ^{136}Xe in the search of neutrinoless double beta decay.

The Borexino detector was situated at the INFN Gran Sasso National Laboratory in Italy and took data from 2007 to 2021. It consisted of 280 t of LS, monitored by $\sim$ 2000 20" PMTs. The main physics goal of the experiment was the measurement of solar neutrinos, which was successfully achieved thanks to its unprecedented radio-purity. Borexino managed to provide a measurement of the full pp-chain [31], with the exception of the hep-neutrinos, as well as the first direct, experimental evidence for the existence of CNO-cycle neutrinos in the Sun [68]. Details about the Borexino detector will be presented in Chapter 2.

The main improvements necessary for the next generation of organic LS detectors can achieved by the significant increases in the target mass, as well as lower background levels or improved capabilities for background discriminating. The upcoming JUNO (Jiangmen Underground Neutrino Observatory) [69] detector takes the first approach, where it will use a 20 kt LS target, monitored by $\sim$ 17 000 20" PMTs. in addition to $\sim$ 25 000 3" PMTs. It is scheduled to start data taking in 2024. The JUNO LS will consist of linear-alkylbenzene as a solvent and two wavelength shifters: PPO at 2.5 g/l and bis-MSB at 1.5 mg/l, which further increases the attenuation length of the scintillation light as this is crucial for such a large detector. Due to the large light-yield, large attenuation length and the high (80%) PMT coverage, JUNO is expected to provide an unprecedented energy resolution of $3\%\sqrt{E/\mathrm{MeV}}$. JUNO is expected to detect around $\sim$ 1350 photo-electrons per MeV of deposited energy, which is drastically higher than the performance of Borexino at $\sim$ 500 photo-electrons per MeV. The main goal of JUNO is the determination of the mass hierarchy, through the measurement of reactor anti-neutrinos at a baseline of 53 km. The physics programs further includes the precision measurement of neutrino oscillation parameters, the search for the diffuse supernova (anti)-neutrino background, the study of geo-neutrinos and solar neutrinos, and others. However, the analysis of solar neutrinos is much more susceptible to the impact of background in the detector, compared to the detection of anti-neutrinos, as the latter provide the an efficient coincidence tagging. The cosmogenic background in JUNO will be relatively high, as the overburden is is only 700 m thick and the LS may not be able to reach a sufficiently

high radio-purity for the solar neutrino analysis. For JUNO the greatest potential lies in the measurement of the ^{8}B-neutrinos, with a possibility to measure ^{7}Be-neutrinos, while a measurement of the pep- and CNO-neutrinos is hindered by the cosmogenic background [69].

The proposed Serappis (SEarch for RAre PP-neutrinos In Scintillator) experiment [70] takes the opposite approach, where the detector is small with a LS volume of only $\sim 20\,\mathrm{m}^3$. The detector would benefit from the existing JUNO infrastructure, as it would be an upgrade of the OSIRIS [71] sub-system, used to monitor the LS radio-purity during the filling of JUNO. The goal of the experiment would be the measurement of the low energy solar pp-neutrino at a few-percent level. This would allow for a sensitive test of the solar luminosity constrain and contribute to the study of solar neutrino oscillations. For this, Serappis aims to provide an excellent energy resolution ($\sim 2.5\%$ at 1 MeV) and a low internal background level of ^{14}C, compared to ^{12}C with a ratio of $\sim 1.9 \cdot 10^{-18}$. The small size of the detector is crucial, as it makes the pile-up of ^{14}C decays with themselves and other spectral components negligible. This pile-up would otherwise smear the lower energy ^{14}C events into the endpoint region of the pp-neutrinos and add a systematic uncertainty.

1.3.4 Hybrid detectors

The idea of so called hybrid detection is to use both the unique morphology of Cherenkov light in combination with the benefits of high light-yield scintillation in the same detector. The Cherenkov light would provide directional information, as well as the potential for particle discrimination given its dependency on the particle momentum, for example between γ-rays, electrons and muons. The scintillation would provide a low energy threshold, as well as drastically improve the performance of the energy and position reconstruction. This approach could potentially achieve unprecedented levels of background rejection that would allow for a rich physics program.

One such proposed hybrid detector is Theia [72] (after the Titan goddess of light), located at the Sanford Underground Research Facility, with a volume of 25 kt–100 kt water-based liquid scintillator [73]. This water-based liquid scintillators (WbLS) is a dispersion of LS micelles in ultra-pure water, stabilised by a surfactant. This LS dilution reduces the absorption of Cherenkov photons in the bulk material, while still providing scintillation. Furthermore, the background of ^{14}C and muon spallation products from ^{12}C is also reduced, relative to pure LS. The main challenges for WbLS to be used as a large scale hybrid detection medium are Rayleigh scattering on the scintillator micelles, as well as the radio-purification of the water phase. The current proposal is for Theia to be realised in phases, where the first phase consists of lightly-doped WbLS and very fast photosensors and has as its main physics goals the measurement of long-baseline accelerator neutrino oscillations, the ^{8}B-neutrino flux, as well as the search for the diffuse supernova neutrino background. The second phase would feature a high light-yield WbLS, or pure but slow LS, which would allow for the additional investigation of the MSW transition region and the measurement of solar CNO- and pep-neutrinos, as well as reactor- and geo-neutrinos. The water phase makes the loading of metal isotopes possible, which could offer an additional CC interaction for the measurement of solar neutrinos or which could be used in a third phase for the search of the neutrinoless double beta decay.

The Jinping Neutrino Experiment (Jinping) [74] is a proposed detector at the Jinping Underground Laboratory in the Sichuan province of China, with a rock shielding of 2400 m. Its main goals would be the measurement of solar neutrinos of both the pp-chain and the CNO-cycle, as well as geo-neutrinos and supernova neutrinos. The proposal is to build two neutrino detectors with a total fiducial target mass of 2000 t for solar neutrino physics, equivalently to 3000 t for geo-neutrino and supernova neutrino physics[6], and a light collection range between 200 to 1000 photo-electrons per MeV. Jinping considers to use a slow LS

[6] As they are anti electron-neutrinos.

as the target material, to allow for the hybrid detection of the events. Here, the Cherenkov light and scintillation could be differentiated through their differences in the PMT hit time distribution. The prompt Cherenkov light can be used for the directional reconstruction of the events and particle discrimination, while the slow scintillation light can be used for the energy and position reconstruction of the events. These different timing structures between the scintillation and the Cherenkov light would require a dedicated electronics system to record the waveform output from the PMTs, which then could also help to discriminate between different particles. A 1 t prototype LS detector is running since 2017 [75], to test the hardware and to measure the radioactive background for Jinping.

1.4 Final remarks

In recent decades, the study of solar neutrinos has yielded a profound and, at the time, surprising experimental discovery: The non-zero mass and flavour oscillation of neutrinos. It seems likely that in the future solar neutrinos will be investigated as a part of a broader physics program by large scale, multipurpose detectors with a focus on different main physics goals, from rare event searches to long-baseline neutrino experiments. Nonetheless, these next-generation neutrino detectors will feature higher event statistics, and novel and improved detection technologies.

One topic of interest for solar neutrinos is the vacuum to matter transition region. While the two extreme flavour conversion regimes have already been measured, where either the vacuum term or the matter term dominates, there exists no direct measurement of the transition from one to the other. A high statistics sample of CNO-, pep- and the lower energy region of ^{8}B-neutrinos could make such a measurement possible. This would also allow for a sensitive search of non-standard interactions affecting the neutrino propagation. In the same context of neutrino matter transitions, the present experiments have so far struggled to provide a significant detection of the day-night asymmetry, i.e. the regeneration of the electron-neutrino flavour, expected from the matter effects of neutrinos traversing the Earth. Another important investigation would be a precise measurement of the pp-neutrino flux, as this is the bottleneck of the pp-chain through which the vast majority of the solar energy is produced. The comparison between the pp-neutrino flux and the well known solar luminosity would allow new opportunities to study the Sun, as well as search for physics beyond the Standard Model. Furthermore, a precision measurement of CNO-cycle neutrinos would allow for a direct measurement of the total C and N abundance in the solar core, almost free from the uncertainties of the Standard Solar Model. This would not only help to improve the SSM, solving the metallicity problem, but it would also provide a means to study the evolution of the proto-planetary disk to the solar system of today [76].

The future study of solar neutrinos will continue to offer a unique insight, not only into our Sun, but also into the most fundamental laws of physics. This may, yet again, lead to surprising and drastic discoveries, but it will certainly help our understanding of the universe in which we live.

Chapter 2

The Borexino Experiment

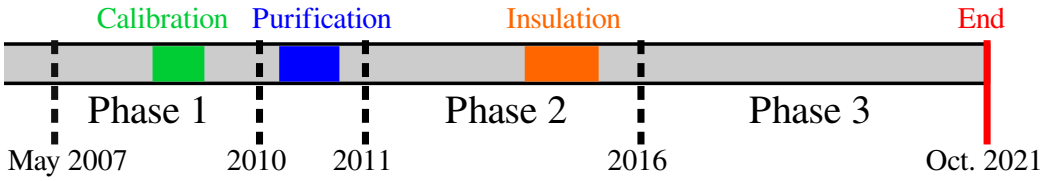


FIGURE 2.1: The Borexino data taking time line.

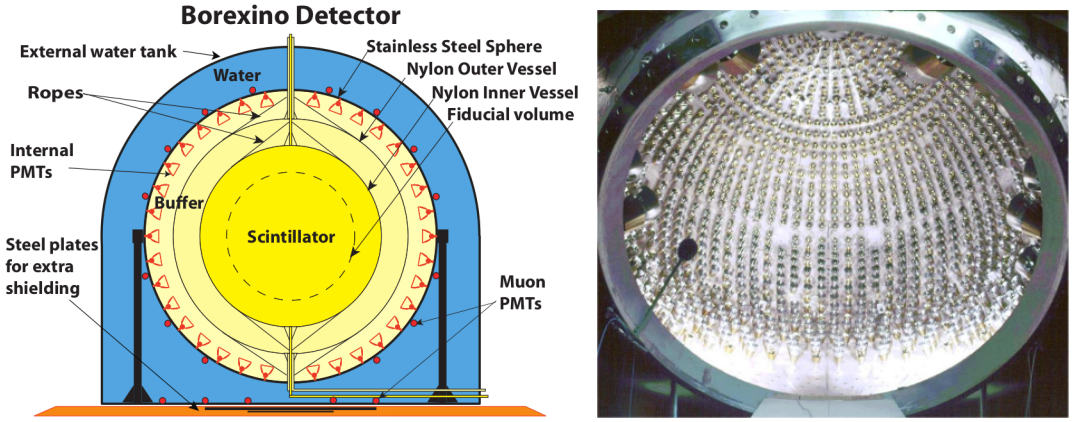
Borexino is a large liquid scintillator (LS) detector, primarily devoted to the spectroscopic measurement of low energy solar neutrinos. It arose from the proposed BOREX detector in 1990 [77] and in the following years new techniques for the purification and handling of the scintillator have been developed [78, 79, 80, 81]. In 1998 the counting test facility [82] has been used to measure the radio-purity of the LS and proved the feasibility of multi-ton, low-background LS detectors. The high radio-purity is one of the fundamental achievements of the Borexino detector, on which all subsequent measurements and analyses are based. For example, the LS had a ^{238}U residual contamination of $(5.3 \pm 0.5)10^{-18}$ g/g in Phase 1 [83] and in Phase 2 this value has been reduced to $< 9.4 \cdot 10^{-20}$ g/g (95% CL) [84], following the additional purification campaign in 2010-2011. This corresponds to a reduction of the natural background by about 9–10 orders of magnitude.

Borexino started data taking in May 2007 and ended data taking in October 2021, as shown in the time line in Figure 2.1. A calibration campaign [85] has been performed during Phase 1, and the LS underwent a further purification between Phase 1 and Phase 2. Borexino provided the measurement of the full pp-chain (pp-ν, ^{7}B-ν, pep-ν, ^{8}B-ν) [83, 31, 86, 87, 88], with the exception of hep-neutrinos, that have a very low flux. Following the thermal insulation of the detector in Phase 2 and the development of new analysis techniques [89], Borexino has also been able to provide the first experimental evidence for the existence of CNO-cycle neutrinos in the Sun [68]. This result has been improved in a subsequent measurement, using the entirety of Phase 3 [45]. In addition to the solar neutrino program, Borexino has also measured geo-neutrinos [90], provided a limit on the neutrino magnetic moment [91] and performed other measurements that are not listed here.

This chapter will first give a brief overview of the detector design in Section 2.1. The signal and background components that are relevant for Borexino are discussed in Section 2.2 and Section 2.3. Then, Section 2.4 gives a rough overview over the data acquisition procedure and explains some event reconstruction algorithms that are of importance for the analyses performed throughout this thesis. The last Section 2.5 describes the Monte Carlo (MC) simulation framework. An example of the standard spectral fit is also shown, to illustrate its contrast to the main analyses of this thesis, which measures the solar neutrinos solely through they directionality, without direct use of the spectral shape.

The electronic trigger scheme and data read-out are not discussed here. A more detailed detector description can be found in [83, 92, 93] and the references therein.

2.1 Detector design



(a) (b)

FIGURE 2.2: (a) Schematic illustration of the Borexino detector. (b) A picture taken of the stainless steel sphere interior from its entrance door. The PMTs can be seen with their light collecting cones. Both pictures are taken from [92].

The Borexino detector is situated underground, in the Hall C of the *Laboratori Nazionali del Gran Sasso* (LNGS), with about 3800 m water equivalent (w.e.) rock shielding above it. This corresponds to a suppression of the muon flux of about 10^{-6}, relative to the surface. Figure 2.2(a) shows the schematic design of the detector, which consists of two sub-detectors: The outer, water Cherenkov detector and the inner liquid scintillator detector, which are separated by the stainless steel sphere (SSS). The external water tank has a diameter of 18 m and a height of about 17 m. Within, the outer detector is filled with 2400 t of ultra-pure water and instrumented with 208 8" photo-multiplier tubes (PMTs). It serves as a passive shielding against the external γ and neutron background radiation, as well as an active muon veto, which allows for the reconstruction of the muon track [94].

The SSS has a radius of 6.85 m and contains the inner detector, which is segmented into three sub-volumes through two transparent nylon vessels [95]. The outer nylon vessel shields the inner volume (IV) from radioactive contamination that can detach from the PMTs and the SSS. The inner vessel has a nominal radius of 4.25 m and contains 280 t of ultra-pure LS and is used as an additional barrier against radioactive contamination from the outside. The LS is composed of pseudocumene (PC, 1,2,4-trimethylbenzene, $C_6H_3(CH_3)_3$) as a main solvent and the additional fluorophore PPO (2,5-diphenyloxazole, $C_{15}H_{11}NO$) with a concentration of 1.5 g/l. The nylon vessels are hold in place with nylon ropes that connect to the vessel end cap structures at the top and bottom of the respective vessel.

The PPO acts as a wavelength shifter, where the PC molecules are first excited by a charged particle and then they transfer this energy non-radiatively to the PPO. The emission of the scintillation light then occurs from the PPO, with a wavelength distribution that peaks at ~ 330 nm, now removed from the absorption region of the PC. At these wavelengths the LS has an attenuation length of ~ 10 m and it has a scintillation light-yield of about 10k photons per MeV of deposited energy. This composite LS has a decay time that can be modeled with a multi-exponential distribution, where the fastest component contributes $\sim 90\%$, with a decay time $\tau_1 = 3.5$ ns. The buffer liquid outside the inner vessel is composed of PC and 5 g/l DMP (dimethyl phytalate, $C_{10}H_{10}O_4$), which quenches the intrinsic scintillation of the PC by a factor of ~ 20. This strongly suppresses the potential signals outside the LS volume.

The SSS is instrumented with a number of 2212, 8" ETL-9351, PMTs, which are used for the detection of optical photons. A number of 1838 PMTs are equipped with light collecting

cones, made out of aluminium. These cones enhance the effective light collection area by a factor ~ 2 and at the same time limit the field of view to the inner vessel volume. The total optical coverage of the inner detectors SSS surface is about 30%. Figure 2.2(b) shows a photo of the SSS with the mounted PMTs, taken trough the entrance door. The PMTs have a typical transit time spread (TTS) of ~ 1 ns and a typical, maximum quantum-efficiency of $\sim 30\%$ at 360 nm. The characteristics of the PMTs can be found in [92, 96]. Different PMTs can have a different single photo-electron response function, i.e. a different charge distribution corresponding to the detection of a single photon. Additionally, the PMTs can have different quantum efficiencies, as well as a small time offset between each other, which includes the impact of the electronic PMT read out system. For this reason a laser calibration system exists, to align the PMT times and their single photo-electron charges [92, 97]. The effective PMT quantum efficiencies are also continuously calibrated, using the intrinsic ^{14}C decay events [98].

2.2 Solar neutrinos in Borexino

Solar neutrinos are detected in Borexino through the elastic scattering off electrons of the LS, where a fraction of the neutrino energy is transferred as kinetic energy to the electron.

$$\nu_{e,\mu,\tau} + e^- \rightarrow \nu_{e,\mu,\tau} + e^- \tag{2.1}$$

The recoil-electron in turn deposits its energy in the LS through ionisation, exciting the LS molecules, which then de-excite through the emission of photons. Thus, the measurable electron energy spectrum is continuous even for mono-energetic solar neutrinos. It has an endpoint of $T_{\max}$, defined by the neutrino energy E_ν and the electron rest mass $m_e c^2$. The recoil-electron spectra of the mono-energetic ^{7}Be-ν (0.862 MeV) and pep-ν (1.44 MeV) exhibit a Compton-like edge, with the endpoints of 0.665 MeV and 1.22 MeV, respectively.

$$T_{\max} = \frac{E_\nu}{1 + \frac{m_e c^2}{2E_\nu}} \tag{2.2}$$

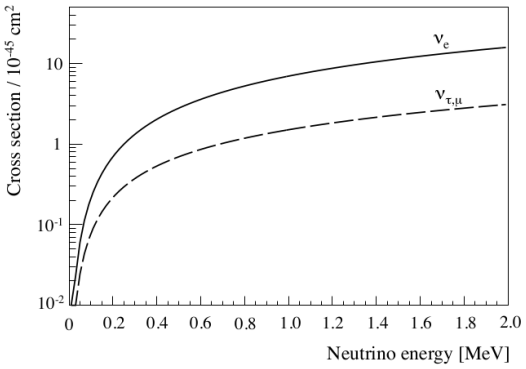


FIGURE 2.3: The energy dependent cross section of the neutrino-electron elastic scattering. The ν_e (solid line) can interact via the CC and NC, while the $\nu_{\mu,\tau}$ (dashed line) can only interact via the NC. Taken from [83].

The following is a summary of the neutrino rate calculation in Borexino, based on [83] and the updated, internal document [99]. The rate of neutrino–electron elastic scattering interactions is given by the product of the incoming neutrino–flux Φ_ν, the elastic scattering cross section $\sigma_{e,\mu,\tau}$ and the number of electrons in the respective target n_e. For Borexino the latter value is $n_e = (3.307 \pm 0.015) \cdot 10^{31}$ per 100 t [68]. Neutrinos of all flavours can be detected in Borexino, where the ν_e interact via both the Charged Current (CC) and Neutral Current (NC), while the ν_μ andν_τ interact only via the NC. Thus, the ν_e have a larger cross section than the ν_μ and ν_τ, which are shown as energy dependent functions in Figure 2.3

as a solid line and a dashed line, respectively. These cross sections $\sigma_{e,\mu,\tau}$ are obtained from the electroweak Standard Model and they include radiative corrections, as described in [100]. The electron neutrino survival probability P_{ee} is calculated according to [32, 101], where the neutrino mixing parameters are taken from the global analysis results of [26]. The total expected neutrino interaction rate R_ν in Borexino is given by the integration over the recoil-electron energy T and the differential neutrino energy spectrum $d\lambda/dE_\nu$ [83]:

$$R_\nu = n_e \phi_\nu \int \frac{d\lambda}{dE_\nu} dE_\nu \int \left(\frac{d\sigma_e(E_\nu,T)}{dT} P_{ee}(E_\nu) + \frac{d\sigma_\mu(E_\nu,T)}{dT} [1 - P_{ee}(E_\nu)] \right) dT \qquad (2.3)$$

Here, the neutrino energy spectra are taken from [41] and the different neutrino fluxes are given by the HZ-SSM (B16-GS98) and LZ-SSM (B16-AGSS09met) [35]. The resulting neutrino rates of the low energy solar neutrinos in Borexino are summarised in Table 2.1:

ν	HZ-SSM (B16-GS98) [cpd / 100 t]	LZ-SSM (B16-AGSS09met) [cpd / 100 t]
pp	131.28 ± 1.74	132.38 ± 1.69
^{7}Be (0.384 MeV)	1.88 ± 0.11	1.72 ± 0.10
^{7}Be (0.862 MeV)	46.01 ± 2.82	42.00 ± 2.57
^{7}Be	47.90 ± 2.82	43.72 ± 2.57
pep	2.74 ± 0.04	2.78 ± 0.04
^{13}N	2.30 ± 0.34	1.68 ± 0.24
^{15}O	2.55 ± 0.43	1.79 ± 0.29
^{17}F	0.066 ± 0.013	0.041 ± 0.007
CNO	4.92 ± 0.78	3.52 ± 0.52

TABLE 2.1: The expected neutrino rates in Borexino. The ^{7}Be branching ratios at 0.384 MeV and 0.862 MeV are 0.1052 and 0.8948, respectively [83]. CNO is the sum of the ^{13}N, ^{15}O and ^{17}F rates.

2.3 Backgrounds in Borexino

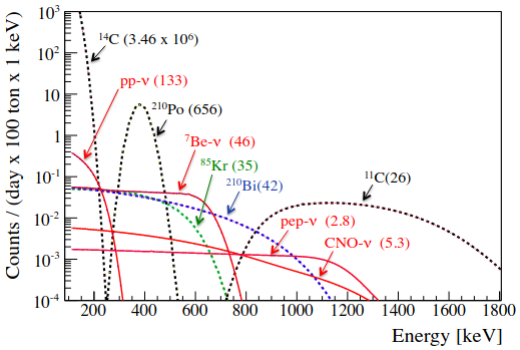


FIGURE 2.4: The expected energy spectra of the solar neutrinos (solid, red lines) and the main background components (black and coloured, dotted lines), for Phase 1. The rates used for the production of this plot are shown in the parenthesis in units of cpd / 100 t. Taken from [83].

While Borexino managed to purify the LS to unprecedented levels of radio-purity there still exist a residual contamination. Additionally, there is also an irreducible background of ^{14}C, intrinsic to the organic scintillator, and the background that is continuously produced by cosmic muons. Figure 2.4 shows the expected energy spectra of the low energy neutrinos in

comparison to the most relevant background contributions, following the assumptions made in Phase 1 [83]. The neutrinos are shown as red solid lines, while the background contributions are shown as dotted lines, for ^{14}C, ^{210}Po and ^{11}C in black, for ^{85}Kr in green and for ^{210}Bi in blue. This section is meant to give a short overview of some background species in Borexino, which can be put in three categories: Internal, external, and cosmogenic background.

2.3.1 Internal background

The internal background stems from the decays of radioactive isotopes contaminating the LS. The most abundant background is ^{14}C (β^-, Q-value = 156 keV, τ = 8270 a), which constitutes a fraction of the organic LS carbon atoms. It is chemically identical to ^{12}C and therefore cannot be reduced through chemical processes. Instead, the Borexino LS is derived from petroleum that is pumped from relatively deep wells, where the levels of ^{14}C are reduced roughly by a factor of a million down to a concentration of $\sim 2 \cdot 10^{-18}$ g/g [83]. The average ^{14}C rate is around $\sim 3.5 \cdot 10^6$ cpd/100 t, which is reduced to $\leq$ 30 Hz by the trigger threshold of 50 keV. This ^{14}C, as well as the random coincidence of two ^{14}C decays, is the most important background for the spectral measurement of the low energy pp-neutrinos.

The isotope ^{85}Kr (β^- (99.57% branching ratio), Q-value = 687 keV, τ = 15.4 a) has a spectral shape that is very similar to that of the ^{7}Be-ν electron recoil, which makes it an important background to reduce for the spectral analysis. It is present in the atmosphere with an average activity of 1 Bq m^{-3}, and can be efficiently removed by stripping the LS with nitrogen gas [81]. For Phase 1 it has been measured around $\sim$ 30 cpd / 100 t [83], while it has been reduced to $<$ 7.5 cpd / 100 t (95% CL) [84] in Phase 2, following the purification campaign.

Another background of great concern for the spectral analysis is ^{210}Bi (β^-, Q-value = 1.16 MeV, τ = 7.23 d), which has a shape that is similar to that of the CNO- and pep-neutrinos, given the expected statistics. The ^{210}Bi is supported by the detector contamination with its parent nucleus ^{210}Pb (β^-, Q-value = 63.5 keV, τ = 32 a). The strong spectral degeneracy between the ^{210}Bi and the CNO-neutrinos makes the measurement of the latter very challenging. It has been first proposed in [102] that it is possible to constrain the ^{210}Bi rate through measuring the rate of its own daughter ^{210}Po (α, Q-value = 5.4 MeV, τ = 200 d):

$$^{210}\text{Pb} \xrightarrow[32\,\text{a}]{\beta^-} {}^{210}\text{Bi} \xrightarrow[7.23\,\text{d}]{\beta^-} {}^{210}\text{Po} \xrightarrow[200\,\text{d}]{\alpha} {}^{206}\text{Pb} \tag{2.4}$$

This idea has been successfully realised in Borexino after the installation of the thermal insulation, using the so called *low polonium field* method [89]. The Phase 3 constraint of ^{210}Bi to $\leq 11.5 \pm 1.3$ cpd / 100 t made it possible to provide the first experimental evidence for the existence of the CNO-cycle neutrinos in the Sun [68] .

The ^{210}Po decay has a large Q-value, but the α-particle it produces has a high ionisation energy loss dE/dx. The amount of scintillation light produced by an α-particle is quenched by a factor of $\sim$ 10, relative to the same energy deposited by an electron [103], which is explained further in Chapter 3. Additionally, the α-particles can be efficiently differentiated from electrons, as α-particles induce a slower scintillation decay time. The discrimination method is outlined in Section 2.4.3.

Typically, the internal background is assumed to be uniformly distributed within the LS volume. This is not the case for ^{210}Po, which has a so called *vessel component* that comes from the convective migration of ^{210}Po from the inner nylon vessel into the LS volume. Here, the ^{210}Po is produced from the ^{210}Pb $\rightarrow^{210}$ Bi $\rightarrow^{210}$ Po decay chain, where the ^{210}Pb and ^{210}Bi stay on the inner vessel. As such, the main purpose of the thermal insulation has been the suppression of this ^{210}Po convective motion, to be able to constrain the ^{210}Bi rate.

2.3.2 External background

The main source of the external background is the radioactivity of the detector materials that contain the LS, but γ radiation from outside the detector may also reach the LS. These are, for example, the SSS, the PMTs and their light collecting cones, the vessel support structures and the other hardware components mounted on the SSS. The external background is strongly suppressed by the shielding provided by the quenched buffer region and the only relevant background that can reach the active, inner scintillation volume are the γ-rays from the ^{40}K (Q-value = 1.46 MeV), ^{214}Bi (Q-value = 2.45 MeV) and ^{208}Tl (Q-value = 2.61 MeV) decays. These γ-rays deposit their energy trough Compton scattering on the LS electrons. This attenuation in the LS volume results in an exponential radial distribution, which allows for the efficient reduction of the γ-background through the selection of a fiducial volume. The total contribution of these external background is ~ 6 cpd / 100 t [84] and the impact of these external γ-rays is small for the measurement of ^{7}Be-neutrinos, but it is relevant for the pep-ν and CNO-ν measurement.

2.3.3 Cosmogenic background

The 3800 m w.e. overburden provides a reduction of the cosmic muon flux by about six orders of magnitude, resulting in a flux of about $\sim 1.2\,\mathrm{m}^{-2}\,\mathrm{h}^{-1}$ [94], corresponding to ~ 4300 muons crossing the inner detector per day. These muons can be tagged with good accuracy by the outer, water Cherenkov detector and also by the inner detector. They can produce a variety of unstable isotopes via spallation processes along their path through the detector. These cosmogenic isotopes have a stable rate in Borexino, given by the equilibrium of their respective production and decay rates. Consequently, they cannot be reduced through the purification of the LS. A dead time of 300 ms is applied after each passing muon, to efficiently suppress the contribution of most spallation products (see Table XI in [83]).

After this cut, the dominant muon–induced cosmogenic background is ^{11}C (β^+, Q-value = 960 keV, $\tau = 29.4$ min), which is accompanied by a neutron:

$$\mu + {}^{12}\mathrm{C} \longrightarrow \mu + {}^{11}\mathrm{C} + \mathrm{n} \tag{2.5}$$

The ^{11}C undergoes β^+ decay with a $Q = 960$ keV, followed by the electron-positron annihilation with 2×511 keV. This provides a total energy deposition between 1.02 MeV and 1.98 MeV.

$$^{11}\mathrm{C} \longrightarrow {}^{11}\mathrm{B} + e^+ + \nu_e \tag{2.6}$$

This energy region is is very important for the measurement of the pep-ν and CNO-ν rates in Borexino. Therefore, it is necessary to reduce the contribution of the ^{11}C, as the expected pep-ν and CNO-ν rates are small compared to the ~ 26 cpd / 100 t rate of the ^{11}C. This is achieved trough a space-time coincidence tagging between the passing muon, the subsequent neutron capture on hydrogen and the delayed signal from the ^{11}C decay, which is explained in Section 2.4.4.

2.4 Event reconstruction

Charged particles deposit energy in the LS, which leads to the emission of scintillation photons and their subsequent detection by the PMTs mounted on the SSS. The number of detected PMT hits is expected at about $N_h = 400$ for an electron-equivalent energy deposition of 1 MeV, or a total charge of all PMTs, in units of photo-electrons, of about $N_{pe} = 500$. This corresponds to an energy resolution of about 5% at 1 MeV. The PMT charge information is

obtained through the integration of the first 80 ns of the PMT pulse, which is followed by a 60 ns dead-time, to write the data to the internal memory buffer.

At the energies of interest around 1 MeV the PMTs operate mainly in the single photo-electron regime, but multi-photon hits become more likely at off-centre position for PMTs close to the energy deposition. Should a PMT detect multiple photon hits in the 80 ns time window, then the total charge of all photon-hits is collected, but only the start of the PMT pulse is recorded as the single time information of these photon-hits. This means that the earliest detected photo-electron defines the PMT hit time, while the timing information of the additional hits is lost within the 80 ns time window.

The Borexino trigger is implemented with a digital logic, due to the large number of PMTs and their dark count rate of 20 kHz [96]. Typically, the trigger threshold is set to a number of 20-25 detected PMT hits within a time window of 100 ns, which corresponds to an energy threshold of around ~ 50 keV. The information on the PMT hits (time and charge) is acquired in a 16.5 µs time window, following the trigger of the inner detector. This normal trigger type for the inner detector requires that the outer detector did not see a signal at the same time. Another trigger type concerns the internal muons, where both the inner and outer detectors are triggered. Here, a second time window of 1.6 ms is opened, following the first acquisition window. This guarantees a high detection efficiency of cosmogenic neutrons that follow the passing of an internal muon. These are used for the rejection of the ^{11}C background through the so called three-fold coincidence method, outlined in Section 2.4.4. Additionally, there exits a range of other trigger types that are used for monitoring and calibration purposes.

The set of PMT hits, each with their corresponding hit time and charge, is called a *trigger event* in Borexino, where the event information also includes the absolute time of the trigger, given in the Global Positioning Systems (GPS) time format. Each trigger event is reconstructed using the ROOT [104] based reconstruction software ECHIDNA, developed by the Borexino collaboration. A first reconstruction algorithm removes non-physical PMT hits and identifies hits that belong to the same scintillation signal of an energy deposition. These so called *clustered events* are then the fundamental objects of the Borexino data, to which the subsequent reconstruction algorithms are applied. As such, the terms *event* and *PMT hit*, used throughout this thesis, correspond to these clustered events and their set of PMT hit information. The raw PMT hits are reconstructed in batches of so called *runs*, which correspond to a typical data taking time of around six hours.

This section provides a short description of the reconstruction algorithms that are relevant for this work. These are the energy estimators, the position reconstruction algorithm as well as the methods for the discrimination of α-particles and ^{11}C events. Details about the electronic signal processing, the trigger system and the data acquisition system can be found in [92, 93].

2.4.1 Energy estimators

ECHIDNA provides a number of energy estimators for the events, based on the corresponding cluster of PMT hits. The true energy distribution of the different events is smeared, due to the finite number of PMT hits. The different analyses performed in this thesis make use of two different estimators, which are the total number of PMT hits N_h and the geometrically normalised number of PMT hits N_h^{geo}. The N_h is simply the total number of detected PMT hits, including multiple hits on the same PMT following the 80 ns integration and 60 ns dead time of the previous hit. The charge information is not used in N_h.

The number of live PMTs $N_h^{live}(t)$ has decreased over time in Borexino, as the PMTs have been dying at a rate of 50-80 PMTs per year. In order to account for this variation of active PMTs the estimator is normalised to a number of 2000 working PMTs through the relation

factor f_{eq}:

$$\mathrm{N_h^{norm}} = f_{\mathrm{eq}}(t) \cdot \mathrm{N_h}(t) = \frac{2000}{\mathrm{N_h^{live}}(t)} \cdot \mathrm{N_h}(t) \tag{2.7}$$

This value $\mathrm{N_h^{norm}}$ shows a non-linearity for high energies, as multiple hits are are not resolved within the 80 ns integration time. A charge based energy estimator $\mathrm{N_{pe}}$ suffers less from such a non-linearity, but at the same time it could be more affected by dark noise. The $\mathrm{N_h}$ estimator seems like the better choice in the context of this thesis, given the relative low energy region of interest.

However, if an energy deposition happens off-centre, at a large radial position, the PMTs close by are able to detect more hits than PMTs further away, simply due to the geometric acceptance given by the solid angle of the respective PMTs. This introduces a radial dependence on the energy estimators. As the main analyses in this thesis select relatively large fiducial volumes it is helpful to use an energy estimator, which corrects for this radial dependence. Such an estimator exists with the geometrically normalised number of PMT hits $\mathrm{N_h^{geo}}$, which makes use of the reconstructed event position $\vec{r}_{\mathrm{ev}}$:

$$\mathrm{N_h^{geo}} = \frac{2000 \cdot \alpha_c}{\sum_i \alpha_i^{\mathrm{live}}(\vec{r}_{\mathrm{ev}})} \cdot \mathrm{N_h}(t) \tag{2.8}$$

Here, $\alpha_c = 0.00298623$ is the value of the solid angle seen by a PMT for an event in the detector centre and the denominator $\sum_i \alpha_i^{\mathrm{live}}(\vec{r}_{\mathrm{ev}})$ is the sum of the solid angles of all live PMTs at the time of the event detection. The solid angle α_i is defined for each PMT individually, depending on the PMT position (x_i, y_i, z_i) relative to the reconstructed event position $(x_{\mathrm{ev}}, y_{\mathrm{ev}}, z_{\mathrm{ev}})$:

$$\alpha_i(x_{\mathrm{ev}}, y_{\mathrm{ev}}, z_{\mathrm{ev}}) = \frac{\pi R^2}{d_i^2} \cdot \frac{x_i(x_i - x_{\mathrm{ev}}) + y_i(y_i - y_{\mathrm{ev}}) + z_i(z_i - z_{\mathrm{ev}})}{d_i \sqrt{x_i^2 + y_i^2 + z_i^2}} \tag{2.9}$$

Where R is the radius of the PMT cathode, and d_i is the distance between the PMT position and the reconstructed event position:

$$d_i = \sqrt{(x_i - x_{\mathrm{ev}})^2 + (y_i - y_{\mathrm{ev}})^2 + (z_i - z_{\mathrm{ev}})^2} \tag{2.10}$$

The $\mathrm{N_h^{geo}}$ is an energy estimator that is normalised to a number of 2000 PMTs, as if they have detected a number of hits for an event at the centre of the detector. Thus, it takes into account the decreasing number of live PMTs over time, as well as the radial dependence of the number of detected hits, which is beneficial for the main analyses of this thesis.

2.4.2 Position reconstruction

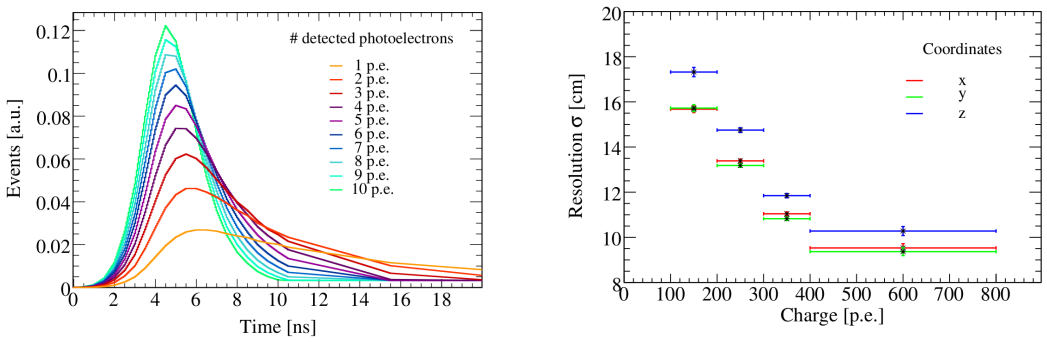
The position reconstruction in Borexino is a log-likelihood fit $\mathscr{L}_{\mathrm{pos}}$, based on the detected PMT hit times t_i and photo-electron charges q_i. The value of $\mathscr{L}_{\mathrm{pos}}$ is calculated from the charge dependent hit time probability density functions (PDFs) $P(t,q)$ and minimised over the proposed event position $\vec{r}_{\mathrm{ev}}$ and the associated event time t_{ev}:

$$\mathscr{L}_{\mathrm{pos}}(\vec{r}_{\mathrm{ev}}, t_{\mathrm{ev}} \mid \{\vec{r}_i, t_i, q_i\}) = -\sum_i^I \log\left[P(t_i,\, q_i \mid \vec{r}_{\mathrm{ev}}, t_{\mathrm{ev}})\right] \tag{2.11}$$

Here, the dependence of $P(t_i,\, q_i \mid \vec{r}_{\mathrm{ev}}, t_{\mathrm{ev}})$ is treated implicitly through the time-of-flight correction of the PMT hit times with a distance d_i to the event position, defined in Equation 2.10. This correction is calculated with the use of the effective refractive index n_{eff}, which

describes the effective group velocity of the photons, given the speed of light c:

$$P(t_i, q_i \mid \vec{r}_{ev}, t_{ev}) = P(t_i - \frac{n_{eff}}{c} d_i + t_{ev},\ q_i) \tag{2.12}$$



(a) (b)

FIGURE 2.5: (a) The photo-electron charge dependent PMT hit time PDFs used for the position reconstruction. (b) Performance of the position reconstruction algorithm. Both plots are taken from [85].

The reference PDFs $P(t,q)$ are shown in Figure 2.5(a), for ten different values of the detected PMT charge, in units of photo-electrons (p.e.). They have been produced for electron events using the Borexino Monte Carlo (MC) simulation [105], briefly described in Section 2.5. The PDF shapes follow that of an exponentially modified Gaussian distribution, which becomes more narrow and shifts to earlier times for larger values of the detected charge. This behaviour stems from the 80 ns integration window of the PMTs, where the first detected hit defines the hit time t_i. This means that for multiple photon hits, which then produce a large detected charge, the hit time peak must shift to earlier times and become more narrow, as the PMT hit time is selected as the earliest time from a set of photon hits. The PDFs for detected charges that are in-between the presented distributions are produced through linear interpolation of the nearest neighbor charge values.

The position reconstruction algorithm has been optimised and tested during the calibration campaign, where the active, inner volume has been mapped with calibration sources of different energies [85]. For example, the PDFs are modified to have a constant, minimal probability value following the peak, which allows relatively late PMT hits to be moved around in time without changing their contribution to $\mathscr{L}_{pos}$. This particulate choice has been used simply because it provides a good position resolution for a broad range of event energies and true calibration source positions. Thus, the n_{eff} have also a certain value, that is different for the data and the MC simulation, just to provide a good position reconstruction in combination with the particular choice of PDFs. The values of $n_{eff}^{data} = 1.68$ and $n_{eff}^{MC} = 1.66$ do not necessarily represent the true group velocity of the photons in the data and in the MC.

The exact positions of the calibration sources have been determined by CCD cameras, with a precision of ~ 1 cm [85]. The result of the calibration campaign for the position reconstruction can be seen in Figure 2.5(b), where the resolution σ is plotted against the energy estimator of the total collected charge of all PMTs. The values for the x, y, and z-axes are shown in red, green and blue, respectively. The position resolution in the energy region of interest for this work, between 300 p.e. to 700 p.e. around 1 MeV, is between 10 cm to 12 cm.

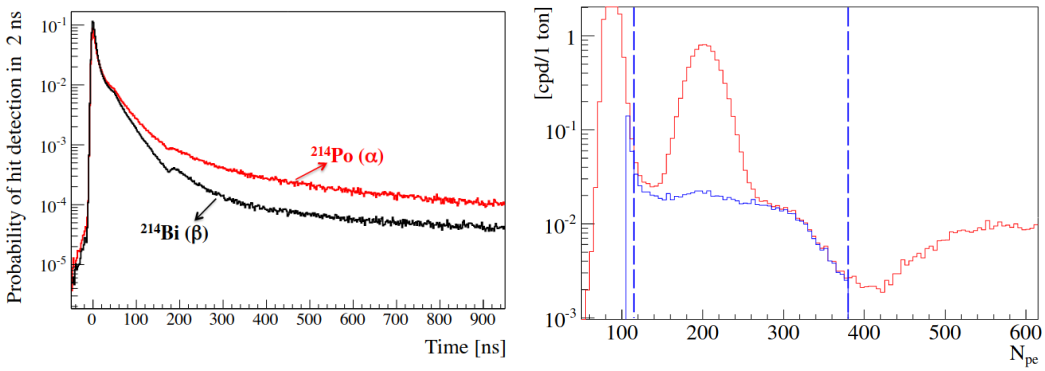
A last note is that the event time t_{ev} corresponds to the time of the event energy deposition, relative to the time of the PMT hit detection, such that the measured PMT hit time distribution aligns well with the PDFs. In Borexino this value is not centered around zero, but it has an

arbitrary constant offset of ∼35 ns added, which is the same for all events and positions. This constant corresponds to the time-of-flight of photons that are emitted in the detector centre, and therefore represents the normalisation of the PMT hit time distributions, as if all photons have been emitted in the detector centre.

2.4.3 Discrimination of α and β events

The time distribution of the scintillation depends on the particle type and its energy loss dE/dx in the LS. As such, it is in principle possible to differentiate particle types based on their detected event hit time distribution. This is most prominently the case for the discrimination between β-like particles and highly ionising α-particles.

Typically, the scintillation molecules are excited to their first singlet state, from which they can de-excite through the emission of a photon. For the α-particles however, due to their large dE/dx, a large fraction of the PPO molecules is excited into the first triplet state, from which a de-excitation into the singlet ground state is spin-forbidden. Thus, the slow component of the scintillation time decay, also called the delayed scintillation, is more pronounced for α-particles than for β-particles. For more details see [106], or other textbooks.



(a) (b)

FIGURE 2.6: (a) PMT hit time distribution for ^{214}Bi β-particles (black) and ^{214}Po α-particles (red). Taken from [83]. (b) Measured energy spectra of the Phase 2 data set, once with (blue) and once without (red) the MLP cut. The ^{210}Po peak is greatly reduced, with a small remaining residue. Taken from [107].

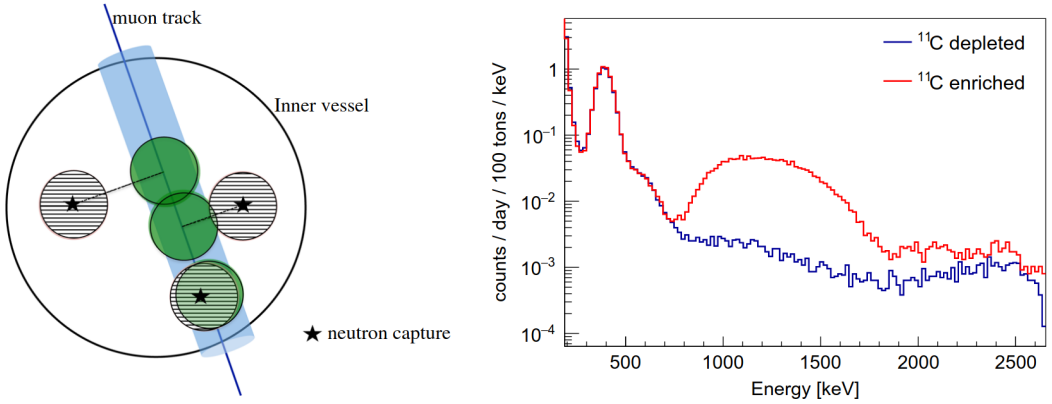
ECHIDNA has different pulse-shape discrimination algorithms available, which aid in the distinction of α-like and β-like energy depositions. The parameters of the algorithms have been tuned using a clean sample of α and β events from the coincidence of the ^{214}Bi(β^-)–^{214}Po(α) decays [83]. Their time-of-flight corrected PMT hit time distributions are shown in Figure 2.6(a), for the sum of a large number of events. Here the pulse-shape of the ^{214}Po(α) is shown in red and the ^{214}Bi(β^-) is shown in black. The highly ionising α-particles result in a pulse-shape that has a larger contribution at late times, compared to the β-particles. The features visible in the distributions around ∼60 ns and ∼180 ns come from the reflections on the SSS and the PMT cathodes, and the PMT dead-time after the charge integration, respectively.

This work makes use of the so called MLP parameter for the α/β pulse-shape discrimination. MLP stands for Multi-Layer Perceptron, which is a neutral network architecture used for supervising binary classifiers. It has been trained with the events selected from the ^{214}Bi(β^-)–^{214}Po(α) coincidence. The MLP uses a number of 13 α/β input variables to discriminate each event, according to the distribution of the time-of-flight correct PMT hit times. These input variables are for example, the tail-to-total ratio, the mean, variance,

skewness of the hit time distribution, and so on [108] . The MLP parameter is used, for example, for the measurement of the geo-neutrinos [90], and it plays an important role in the ^{210}Bi-constraint for the spectroscopic CNO-ν measurement [89, 68]. Figure 2.6(b) , from [107], shows the event energy spectra with and without the MLP cut, as blue and red lines, respectively. The large ^{210}Po peak around $N_{pe} = 200$ is efficiently reduced, with only a small residual contribution, while the β-like events are almost unaffected.

2.4.4 Discrimination of ^{11}C events

The cosmogenic ^{11}C events are a major background component, mostly important for the measurement of the pep-ν and CNO-ν rates. In order to reduce this background, the so-called three-fold coincidence (TFC) algorithm has been implemented in Borexino [83, 109]. The basic working principle of the TFC is based on the space-time correlation between the passing of a muon in the inner volume, the accompanying neutron captures on hydrogen and the delayed ^{11}C decays.



(a) (b)

FIGURE 2.7: (a) Three-fold-coincidence (TFC) scheme. Veto regions are defined, after the passing of a muon through the inner volume. The veto volumes are: A cylinder around the muon track (blue) and spheres around the neutron capture points (shaded) and around their projection upon the track (green). (b) The spectrum of the ^{11}C-depleted (blue) and ^{11}C-enriched (red) data events of Phase 2, using the TFC algorithm. The spectra are normalised to the same exposure. Both figures taken from [109].

The muons cross the detector at a rate of about three per minute, where they interact with the ^{12}C of the LS and produce ^{11}C and a neutron through spallation. This neutron is captured on hydrogen relatively fast, with a lifetime of $\tau_n \sim 260\,\mu$s, which results in the emission of a 2.2 MeV γ-ray. The ^{12}C decays via positron emission with a relatively long lifetime of $\tau_{^{11}C} = 29.4$ min. The TFC now works by defining a list of space-time veto regions, as is depicted schematically in Figure 2.7(a). Following a muon track, a *neutron window* is opened, a second 1.6 ms acquisition window after the first 16.5 µs window. Events that fall in this window are called neutron events. If at least one neutron event is found, a number of veto regions is defined as follows: (1) a cylindrical volume with a radius of 0.7 m of around the reconstructed muon track; (2) a sphere with a radius of 1.2 m around every reconstructed neutron event position; (3) a spherical projection of every neutron event position onto the muon track. These latter projections correspond to the likely regions of the actual ^{11}C events, as they have to be produced by the muon along the track. If a high neutron multiplicity is detected or if the neutron events have a position reconstruction that is considered unreliable, then the cylindrical volume is vetoed for two hours. Otherwise the spherical volumes are vetoed for two hours. A more detailed veto scheme can be found in [83, 108]

The TFC algorithm provides a TFC-tagged (^{11}C-enriched) data set and a TFC-subtracted (^{11}C-depleted) data set, as is shown in Figure 2.7(b) in red and blue, respectively. The TFC-tagged data set includes a large fraction of non-^{11}C events, as the veto time is relatively long. For Phase 2 the TFC cut results in a $> 90\%$ rejection of the ^{11}C events, while at the same time it provides an exposure of all events in the TFC-subtracted spectrum of about $\sim 60\%$. The exact parameters used for the TFC cut are tuned using toy-MC studies [83], such that the TFC algorithm can provide a good compromise between the ^{11}C rejection and the amount of exposure which is left in the TFC-subtracted spectrum used for the analysis.

2.4.5 Standard data selection cuts

The events can be selected according to an energy region of interest, a fiducial volume, and the particle discrimination cuts, after the ECHIDNA reconstruction chain. These selection choices are dependent on the particular analysis goal and are therefore different for different analyses. Before that, however, a battery of standard cuts is applied for the selection of all data events, to remove muons and short-lived cosmogenic background, as well as noise events. Only a rough overview is given here, further details can be found in [83].

- *Muon and muon daughter cut:* If a muon is detected by either the outer detector or by the inner detector a 300 ms veto is applied to the entire detector. This means that events detected in this time window are not selected for further analysis, which de-selects most of the unstable isotopes associated with the muon.
- *Single energy deposit*: Only a single cluster is selected for each event. The event is rejected if the reconstruction algorithm fails to find even a single cluster in the 16.5 µs acquisition window. Should multiple clusters be found, then only the first cluster is used.
- *Fast coincidence events:* Events that occur in a time window of 2 ms and that are reconstructed with a distance smaller than 1.5 m between each other are rejected. This cut is used to remove the ^{214}Bi-^{214}Po coincidence events, possible correlated events of unknown origin and noise events.
- *Charge quality control:* The total collected charge of all PMTs is compared to the number of PMTs that saw a hit. The event is deselected if the ratio between the total charge and the average charge per PMT is too small or too high.
- *Additional noise removal:* The cluster that has been responsible for the event trigger has a well defined position in the the data acquisition window. Thus, an event is only accepted if the cluster start time is within a 1.7 µs window around a fixed position in the data acquisition window. Additionally, noise events are deselected if more than 75% of the detected PMT hits are readout and powered by a single electronics rack.

2.5 Monte Carlo simulation

The Borexino collaboration has developed a GEANT4 [110, 111, 112] (v4.10.5) based Monte Carlo (MC) simulation package called G4BX2, which is used to simulate the full detector response for both the neutrino signal and background events. This section gives a short overview of G4BX2 and more details can be found in [105].

Technically, the software is structured as a chain of three separate parts. First is the simulation of the physics processes and the light tracking in the detector. For this purpose a wide range of event generators exist, from solar neutrino interactions, to radioactive decays, geo-neutrinos, and the calibration source events. The particles are simulated according to their

theoretical energy spectra, which for solar neutrinos is taken from [41]. The energy loss of each specific particle is simulated for every material present in the detector. Then a number of scintillation or Cherenkov photons are generated, considering the particle energy loss in the media and the properties of the scintillator or the buffer volumes. The path of each single optical photon is tracked until a PMT is hit or the photon is absorbed somewhere else. The interactions of the photons with the medium are also simulated, such as reflection, scattering, and absorption and re-emission by the LS.

The second part is the simulation of the electronics response, after a photon hits a PMT. Here, the quantum efficiency of each individual PMT is considered, and the PMT signal pulse takes into account the specific design of the electronics chain in Borexino. Then the trigger generation is simulated and a raw file is saved. This filet has the same structure as the one produced for real data.

The last part is then the event reconstruction of the MC simulation with ECHIDNA, including the energy estimators, position reconstruction and pulse-shape parameters. The ECHIDNA code is the same for both the real data and the MC simulation.

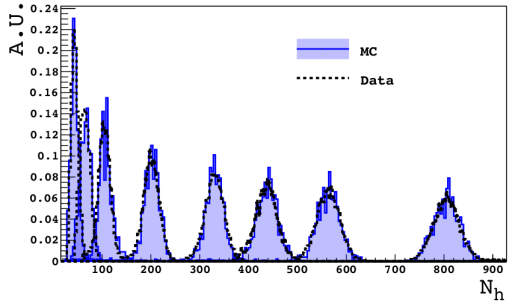


FIGURE 2.8: The N_h distribution of the γ-ray calibration sources for the data (black) and the MC (blue) at the centre of the detector. From [105].

The primary use of the G4BX2 MC is to produce the fit models used for the spectral analyses of the solar neutrinos, including the full detector response. The full simulation of the light production and propagation requires a large number of parameters, that describe the various features of the LS and other materials. Examples for these parameters are the attenuation lengths, the reflectivity, the re-emission probabilities and the refractive indices. Here, the calibration campaign has provided a number of different radioactive sources, which have been used to optimise these parameter values, to validate the accuracy of the MC simulation and to quantify the systematic differences between the source data and the MC. The results of the calibration campaign can be found in [85] and an exemplary plot for the MC validation is shown in Figure 2.8. The energy distribution N_h is plotted for a number of different γ-ray calibration sources, at the centre of the detector, together with the 2.2 MeV γ-ray from the neutron capture on hydrogen, provided by the ^{241}Am–^{9}Be source. The source data is shown in black, while the MC simulations are shown in blue. It can be seen that both the mean values and the resolutions of these distributions are well reproduced by the MC. In total, the systematic difference between the data and the MC energy response is estimated to be $< 0.5\%$ [68], including inaccuracies in the energy scale and in the description of the non-uniformity and the non-linearity of the response.

2.6 Exemplary spectral fit

Almost all analyses performed with Borexino have been performed using the energy spectrum. This is explicitly different for the solar neutrino measurement of this thesis, which makes use of the solar neutrino directionality. Therefore, this section presents an overview of the typical spectral fit used in Borexino, to make the differences between the standard energy and

directional analysis more understandable. For this, the analysis of Phase 2 is shown as an example of the spectral fit. The details can be found in [84, 31, 113].

The spectral analysis uses a fiducial volume cut on the reconstructed event radius R and z-position with $R < 2.8\,m$, $-1.8\,m < z < 2.2\,m$, which selects the innermost 71.3 t of the LS and excludes the residual radioactive background from the nylon vessel and the vessel end cap structures at $z = \pm 4.25\,m$. The analysis is performed with a multivariate fit, that uses the spectrum the N_h energy estimator, the reconstructed radial distribution of the events and a e^+/e^- pulse shape discriminator, based on the likelihood value of the position reconstruction algorithm. The energy range for the fit corresponds to 0.19 MeV–2.93 MeV.

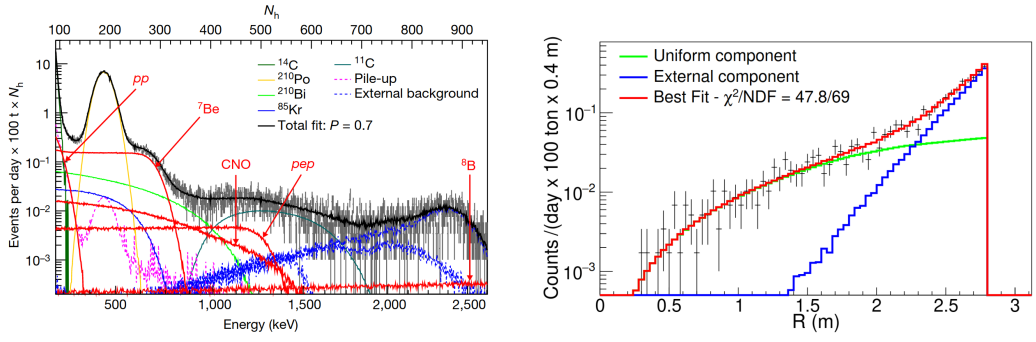
First, the data is split into two data sets consisting of the TFC-tagged and TFC-subracted events, which are then fitted simultaneously through the maximisation of the binned likelihood function $\mathscr{L}_{3D}$:

$$\mathscr{L}_{3D}(\mathbf{k} \mid \theta) = \mathscr{L}_{\text{TFC-sub}}(\mathbf{k} \mid \theta) \cdot \mathscr{L}_{\text{TFC-tag}}(\mathbf{k} \mid \theta) \tag{2.13}$$

Here, θ symbolises the set of arguments over which the function is maximised, i.e. the rates of the solar neutrino signal events and the background events. The measured data is symbolised by $\mathbf{k}$. The likelihoods $\mathscr{L}_{\text{TFC-sub}}(\mathbf{k} \mid \theta)$ and $\mathscr{L}_{\text{TFC-tag}}(\mathbf{k} \mid \theta)$ are given by the standard Poissonian likelihood functions for three- and two-dimensional histograms, respectively:

$$\begin{aligned} \mathscr{L}_{\text{TFC-sub}}(\mathbf{k} \mid \theta) &= \prod_j^{N_E} \prod_l^{N_R} \prod_m^{N_P} \frac{\lambda_{jlm}(\theta)^{k_{jlm}}}{k_{jlm}!} e^{\lambda_{jlm}(\theta)} \\ \mathscr{L}_{\text{TFC-tag}}(\mathbf{k} \mid \theta) &= \prod_j^{N_E} \prod_l^{N_R} \frac{\lambda_{jl}(\theta)^{k_{jl}}}{k_{jl}!} e^{\lambda_{jl}(\theta)} \end{aligned} \tag{2.14}$$

Here the number of data entries k_{jlm} in the energy bin j, the radial bin l, and pulse shape parameter bin m are compared to the expected number of bin entries from the G4BX2 MC model $\lambda_{jlm}(\theta)$. The value of $\lambda_{jlm}(\theta)$ depends on the rates and the MC spectra of the relevant event species. These rates are kept in common between the TFC-tagged and TFC-subtracted spectra, except for the cosmogenic ^{11}C itself and the ^{210}Po, as it has a non-uniform position distribution in the fiducial volume.



(a) (b)

FIGURE 2.9: (a) Phase 2 result of the multivariate fit for the TFC-subtracted energy spectrum. The total number of PMT hits N_h is shown on the top. From [31]. (b) An example of the radial distribution of the TFC-subtracted events, with a selection of $N_h > 290$. From [84].

Figure 2.9 shows the results of the multivariate fit for Phase 2, for the TFC-subtracted energy spectra on the left and an example of the radial distribution on the right. For the energy spectra, estimated with the normalised number of PMT hits N_h^{norm}, the data is shown in black, the MC solar neutrino signal events are shown in red and the different MC background spectra

are shown in different colours. It can be seen that the spectral shape of data is well reproduced by the MC model. The presented radial distribution is an example, where the events are selected according to $N_h > 290$ from the TFC-subtracted data. The data is shown in black, the uniform components are shown in green and the external γ-ray background is shown in blue. Again, the MC model is well able to reproduce the measured data.

Ultimately, the fundamental achievement of an unprecedented low background level has allowed Borexino to measure all neutrino components of the pp-chain except hep-ν, as well as to provide the first direct detection and measurement of the CNO-cycle neutrinos in the Sun. As will be seen in the following chapters, it is also these same conditions that allow for a measurement of solar neutrinos through a fundamentally different approach: Using the directional information of the solar neutrinos provided by the Cherenkov radiation of their recoil-electrons. The current solar neutrino results of Borexino are summarised in Table 2.2:

Solar-ν	Rate [cpd / 100 t]	Flux [$cm^{-2}\,s^{-1}$]	HZ-SSM flux [$cm^{-2}\,s^{-1}$]
pp	$134 \pm 10^{+6}_{-10}$	$\left(6.1 \pm 0.5^{+0.3}_{-0.5}\right) \cdot 10^{10}$	$5.98\,(1.0 \pm 0.006) \cdot 10^{10}$
^{7}Be	$48.3 \pm 1.1^{+0.4}_{-0.7}$	$\left(4.99 \pm 0.11^{+0.08}_{-0.06}\right) \cdot 10^{9}$	$4.93\,(1.0 \pm 0.06) \cdot 10^{9}$
pep (HZ)	$2.43 \pm 0.36^{+0.15}_{-0.22}$	$\left(1.27 \pm 0.19^{+0.08}_{-0.12}\right) \cdot 10^{8}$	$1.44\,(1.0 \pm 0.01) \cdot 10^{8}$
pep (LZ)	$2.65 \pm 0.36^{+0.15}_{-0.24}$	$\left(1.39 \pm 0.19^{+0.08}_{-0.13}\right) \cdot 10^{8}$	$1.44\,(1.0 \pm 0.01) \cdot 10^{8}$
$^{8}B_{HER}$	$0.223^{+0.015\,+0.006}_{-0.016\,-0.006}$	$\left(5.68^{+0.39\,+0.03}_{-0.41\,-0.03}\right) \cdot 10^{6}$	$5.46\,(1.0 \pm 0.12) \cdot 10^{6}$
CNO	$6.7^{+2.0}_{-0.8}$	$\left(6.6^{+2.0}_{-0.9}\right) \cdot 10^{8}$	$4.88\,(1.0 \pm 0.11) \cdot 10^{8}$
hep	< 0.002 (90% CL)	$< 2.2 \cdot 10^{5}$ (90% CL)	$7.98\,(1.0 \pm 0.30) \cdot 10^{3}$

TABLE 2.2: The current solar neutrino results of Borexino, as reported in [31, 88, 45]. The results are presented with their separated statistical and systematic uncertainties, except for the CNO-ν rate, where the total uncertainty is quoted. For hep-ν only a limit can be given.

Chapter 3

Correlated and Integrated Directionality

At present, the measurement of solar neutrinos is performed with two different detector designs: Water Cherenkov (WCh) detectors like Super-Kamiokande [55], SNO [53] and liquid scintillator (LS) detectors like Borexino, KamLAND [65] or JUNO [69]. Both types detect solar neutrinos indirectly via elastic scattering off electrons and the light produced by these recoil electrons is typically detected with photo-multiplier tubes (PMTs). The fundamental difference between these detector types is the process of light production, as well as the subsequent analysis procedure. Using either Cherenkov light or scintillation as the main signal component leads to different sets of advantages and disadvantages between these detectors.

For WCh detectors, the use of Cherenkov light from the recoil electrons (or other leptons) allows for the reconstruction of the event direction. This is a powerful tool for background suppression, essential for the measurement of solar neutrinos [58] in these detectors. Furthermore the Cherenkov ring morphology allows for particle identification, such as for electron and muon events for the measurement of atmospheric neutrinos [114]. The primary disadvantage of WCh detectors is the relatively low light-yield at MeV neutrino energies. Compared to LS detectors, this leads to a higher low-energy threshold, due to the presence of background. Likewise, WCh detectors have a worse reconstructed energy and position resolution than LS detectors. Another disadvantage is that charged particles can only be detected if their velocity in the medium is greater than the speed of light. For water, with a refractive index of $n \approx 1.33$, this results in a kinetic energy threshold of about 0.26 MeV for electrons and about 54 MeV for muons. The effective low-energy threshold is higher in practice because of the presence of background, as the number of Cherenkov photons emitted is low close to the threshold. For example, taking into account the detector coverage, photon detection efficiency the presence of radioactive background results in an effective, (kinetic) low-energy threshold of ~ 3.5 MeV for Super-Kamiokande [58] and SNO [59]. At this energy, only about 30 photo-electrons are detected per event, resulting in an energy resolution of $\sim 20\%$.

In comparison, LS detectors offer a relatively high light-yield and consequently a better energy resolution, more precise event position reconstruction and a lower energy threshold, provided they have a sufficiently low radioactive background level. Here, the neutrino recoil electrons excite the LS molecules, which in turn emit scintillation light isotropically. For Borexino this results in 400 PMT hits at 1 MeV deposited energy, assuming the nominal number of 2000 live PMTs. This corresponds to an energy resolution of $\sim 5\%$ and the effective low-energy threshold for the spectral analysis in Borexino is ~ 0.19 MeV [31]. The main disadvantage of LS detectors is that an event direction reconstruction is not possible. For Borexino, with a refractive index of $n \approx 1.5$, the electron kinetic energy threshold for the Cherenkov process is about 0.17 MeV, but for a kinetic energy of 1 MeV the expected number of direct Cherenkov PMT hits is ~ 1.5 out of 400 hits in total. The presence of the Cherenkov photons is actively taken into account in Borexino, as it contributes $\sim 5\%$ to the visible energy, where the vast majority of Cherenkov photons is absorbed and re-emitted

by the LS. Consequently, this impacts the non-linearity of the energy spectrum [105, 115]. Although the directional Cherenkov information is in principle present in LS detectors, it has until now not been possible to make use of it.

In recent years there has been an ongoing effort from a broad community to pursue the idea of hybrid event detection. The goal is to combine the directionality and morphology of Cherenkov light with the excellent energy resolution and low-energy threshold of scintillation, as for example in the proposed THEIA detector [72]. This pursuit is motivated by the potential for a rich physics program, including the measurement of solar neutrinos, especially an improved measurement of solar CNO-neutrinos [116] and the search for neutrino-less double beta decay, where solar neutrinos constitute a background [117]. Another possible application is in future long-baseline neutrino oscillation experiments [72], where the scintillation light provides a means to reconstruct hadronic recoils in the final states of GeV neutrino interactions.

To achieve this goal of hybrid detection there is currently a diverse range of research and development activities ongoing [118], which can be grouped into four categories:

- New target materials, such as water based liquid scintillator [119], slow scintillator [120] or the use of novel wavelength shifting fluorophores such as quantum dots [121]. These new targets potentially allow for a better separation of Cherenkov light and scintillation by tuning the scintillation time profile and wavelength distribution.
- Fast photo-detectors, to explicitly improve the time separation of Cherenkov light and scintillation, such as large area picosecond photo-detectors (LAPPDs) [122, 123].
- Hardware for the spectral sorting and separation of Cherenkov light and scintillation, for example with bandpass or dichroic filters [124, 125].
- New reconstruction algorithms, making explicit use of both scintillation and Cherenkov light on an event-by-event basis [126, 127, 128, 129], to reconstruct the event position, direction and energy.

These promising target materials, the new hardware and the reconstruction techniques have so far only been investigated experimentally on a relatively small scale, such as in CHESS [130], FlatDot [131], or ANNIE [132]. The feasibility and impact of hybrid event detection in large scale detectors has so far only been studied with Monte Carlo simulations [72, 116, 133]. To summarise this introduction: Combining the directional Cherenkov information with the high light-yield of scintillation is of great interest for a diverse number of neutrino related physics goals, but the experimental proof of this hybrid detection in an existing, large scale detector has been outstanding. The provision of that proof is the primary topic of this thesis and it is successfully achieved [134, 135] through a novel analyses method called the *Correlated and Integrated Directionality* (CID), using the Borexino detector. The idea of the CID is not to perform an event-by-event direction reconstruction, but instead correlate the measured photon direction of each hit PMT with the known direction of the solar neutrinos and then integrate over a large number of events. The resulting distribution of angular hit values allows then to perform a statistical inference on the number of neutrino events. The current chapter introduces the general idea of the CID method as well as some details of its implementation in the Borexino detector. The initial idea of this integrated approach came from [redacted], as part of his work on the proposed Theia experiment [72].

The first Section 3.1 explains the general principle of the CID method, which should be applicable for all types of solar neutrino detectors that have a reasonably precise event position reconstruction, such as KamLAND [65], JUNO [69] or the LS phase of SNO+ [136]. Then, Section 3.2 explains the production of the CID Monte Carlo model, and Section 3.3 introduces the Nth-Hit time like variable, which is necessary for the Cherenkov-scintillation time separation. After this, the following Chapters 4 and 5 present the in-situ calibration

of the effective group velocity of scintillation and Cherenkov photons, respectively. The measurement of ^{7}Be-ν events in Chapter 6 is then used as the proof-of-principle for the CID method, making it the first experimental hybrid detection of solar neutrinos in a monolithic, high light-yield LS detector. This ^{7}Be-ν analysis provides important lessons for the application of CID in Borexino. Subsequently the CID analysis is improved in the final Chapter 7, for the measurement of CNO-ν events, providing a detection as well as a direct proof of the solar origin of the CNO-cycle neutrinos.

3.1 Method

The basic principle of the *Correlated and Integrated Directionality* (CID) method is schematically illustrated in Figure 3.1. First, a solar neutrino with an energy E_ν interacts in the LS via elastic scattering off an electron, transferring the kinetic energy T. This electron with mass m_e now has the scatter angle θ_e, relative to the initial direction of the neutrino:

$$\cos\theta_e = \left(1+\frac{m_e}{E_\nu}\right)\sqrt{\frac{T}{T+2m_e}}, \tag{3.1}$$

For relative large values of T the electron is scattered in forward direction. If the corresponding particle velocity v is larger than the speed of light c in the LS, given by the refractive index n, then the electron emits Cherenkov radiation in an angle θ_{ch} [14]:

$$\cos\theta_{\text{ch}} = \frac{c}{nv} = \frac{1}{n\beta} \tag{3.2}$$

The direction of these Cherenkov photons $\vec{d}_{\text{ch}}$ is now correlated to the initial direction of the solar neutrino $\vec{d}_\nu$. Now each measured PMT hit can be correlated to the known position of the Sun, i.e the neutrino direction $\vec{d}_\nu$, with the correlation angle $\cos\alpha$. It is necessary to estimate the photon direction $\vec{d}_{\text{hit}}$, using the known PMT position $\vec{r}_{\text{PMT}}$ and the reconstructed event position $\vec{r}_{\text{ev}}$:

$$\begin{aligned}\vec{d}_{\text{hit}} &= \frac{\vec{r}_{\text{PMT}}-\vec{r}_{\text{ev}}}{|\,\vec{r}_{\text{PMT}}-\vec{r}_{\text{ev}}\,|}\\ \cos\alpha &= \vec{d}_{\text{hit}}\cdot\vec{d}_\nu\end{aligned} \tag{3.3}$$

The integration of a large number of events results in a $\cos\alpha$ distribution. For a high light-yield LS detector the PMT hits are made up mostly of scintillation light, but a fraction of PMT hits, at early hit times, consists of Cherenkov photons. This results in an expected $\cos\alpha$ distribution for neutrino signal events which is made up of two components: First, a relatively large baseline, a flat distribution from scintillation hits, which are isotropic and thus uncorrelated to the direction of the solar neutrinos. Second, on top of the scintillation baseline, a $\cos\alpha$ distribution with a distinct Cherenkov peak corresponding to $\cos\theta_{\text{ch}}$. This peak is smeared by the position resolution and Rayleigh scattering inside the LS, as well as the distribution of $\cos\theta_e$, according to the selected energy region. The $\cos\alpha$ distribution of the intrinsic, radioactive residual background events is expected to be flat, as the direction of background events is not correlated to position of the Sun.

Because the signal and background $\cos\alpha$ distributions look different, it is now possible to statistically infer the number of neutrino events, given a sufficiently high event statistic. The use of the directional Cherenkov information for background suppression is normally performed through event-by-event *direction* reconstruction, but CID effectively transforms this problem to the event-by-event *position* reconstruction, a task at which LS detectors excel.

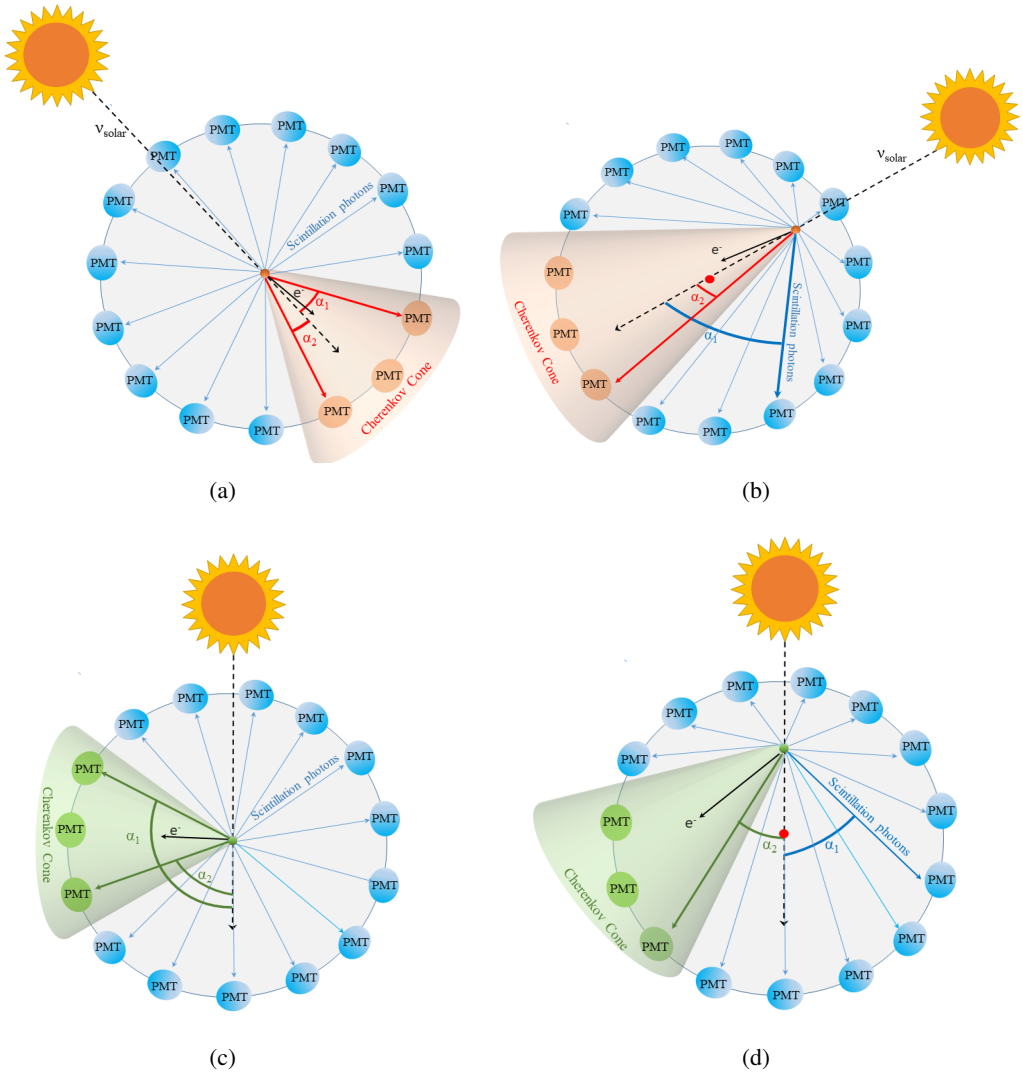


FIGURE 3.1: Schematic representation of the angular correlation α, between the reconstructed direction of the detected photon hits and the position of the Sun, with respect to Borexino. The photon direction is estimated using the reconstructed event position and the hit PMT position. From [135, 134]. (a) A solar neutrino recoil-electron at the centre of the detector produces Cherenkov light in a cone (red arrows), pointing forward in the direction of the Sun. At the same time it produces isotropic scintillation light (blue arrows). The directional angles α_1 and α_2 correspond here to the first and second detected photons of the event, respectively. The Cherenkov photons are correlated to the incoming direction of the solar neutrinos and so are the PMT hits they trigger. (b) The first detected photon here is a scintillation photon, therefore not correlated to the direction of the solar neutrino, and the second detected photon is a Cherenkov photon. Compared to (a), this event would result in a flatter angular distribution. Additionally, this event happens off-centre. (c) A radioactive background, electron event also produces Cherenkov light (green arrows) and isotropic scintillation photons (blue arrows). Again, α_1 and α_2 are the directional angles of the first and second photons of the background event. Here these are Cherenkov photons, but they have no correlation to the Sun's direction. (d) The same as (c), but here the event is off-centre, where the first photon is a scintillation photon and the second detected photon is a Cherenkov photon. These background events result in a flat angular distribution.

Calculation of the solar direction

The Cherenkov information of the CID is given by the $\cos\alpha$ distribution of the integrated event hits. For the calculation of $\cos\alpha$ it is first necessary to know the position of the Sun with sufficient precision. In Borexino this is accomplished by the use of the solar position algorithm from [137], which has also been used previously for the study of the solar neutrino day-night asymmetry [138].

The algorithm uses the event trigger time and the geographic coordinates of the Borexino detector (latitude = 42.421, longitude = 13.515) to calculate the solar position in horizontal coordinates (azimuth, altitude). The direction of solar neutrinos can then be expressed in the Borexino (spherical) coordinate system, where φ, in contrast to the azimuth, is defined in the mathematically positive sense and φ is moreover offset from the geographic north by 52°:

$$\begin{aligned} \varphi &= 308^\circ - \text{azimuth} \\ \cos\theta &= \cos(90^\circ + \text{altitude}) \end{aligned} \tag{3.4}$$

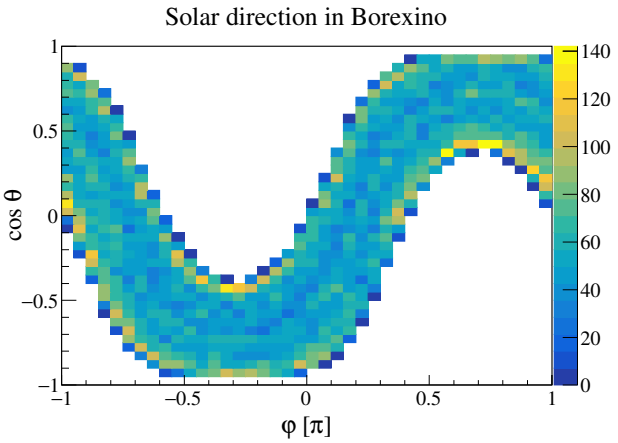


FIGURE 3.2: The distribution of the solar neutrino direction in Borexino detector coordinates. The sinusoidal shape corresponds to the daily movement of the Sun, while the $\cos\theta$ width originates from the seasonal distribution of the solar direction at the geographic location of Borexino.

Figure 3.2 shows the distribution of the solar direction for Phase 1 (with the data events selected in Chapter 6). The daily motion of the Sun is visible as the sinusoidal shape, while the seasonal modulation of the Sun's altitude is given by the $\cos\theta$ width of the distribution. The important thing to note here is that the solar direction is not uniformly distributed. This means that the event morphology, averaged over many events, does not necessary need to produce a flat $\cos\alpha$ distribution for background events. The reason for this is that the background event hits still use the solar direction in the calculation of $\cos\alpha$. Only a certain set of PMT positions is even able to contribute hits with $\cos\alpha = 1$, for example, while only a different set of PMTs is able to contribute to $\cos\alpha = 0$. The background hits may be isotropically distributed on the PMTs, but the PMTs do not contribute uniformly to the $\cos\alpha$ distribution.

3.2 Production of the Monte Carlo model

To be able to perform a statistical inference on the number of neutrino events it is necessary to produce a model of the CID $\cos\alpha$ distribution, that is able to fully reproduce the features of the measured data $\cos\alpha$ distribution. This is done here with the GEANT4 [110, 111] based Borexino G4BX2 Monte Carlo (MC) framework [105]. The parameters of the MC have been tuned with the calibration campaigns in the period between 2008 and 2011 [85] and G4BX2 is well capable of building the MC model required for the spectral fit. As Borexino has never been intended to make direct use of the Cherenkov light it is now necessary to modify the

standard MC for the correct production of the CID $\cos\alpha$ distributions. Furthermore, CID specific calibrations of the MC must also be performed, as explained in the Chapters 4, 5.

3.2.1 Simulated step length

The first important change is the correct selection of the GEANT4 simulation parameters, more specifically the choice of a sufficiently small simulation step length. The following part is intended for readers who are explicitly trying to implement the CID in a simulation framework and to draw attention to the fact that even single parameters for the simulation of the physics model can introduce substantial errors.

In GEANT4, when a particle is shot and transported, all physics processes associated to the particle propose a step length, dependent on the process cross section. The particle is then moved according to the shortest proposal and all continuous processes (e.g. Cherenkov radiation, ionisation and scintillation) are executed *along the step*. Then, the post step phase of the process that limited the step is executed and the procedure begins anew. Here, the continuous energy loss imposes a limit on the step size because of the energy dependence of the cross sections. The MC generally assumes that the particle cross sections are approximately constant along a step. The step size should be small enough that the change in the cross section, from the beginning of the step to the end, is sufficiently small. To achieve a good compromise between the step length and the computation time, GEANT4 is limiting the step length by not allowing the stopping range of the particle to decrease by more than a ratio of $k = \Delta R / R$. This results in a faster computation time, but it effectively averages out the particle path in the presence of multiple scattering. This is a simplified explanation, please see the GEANT4 handbook for more details [110].

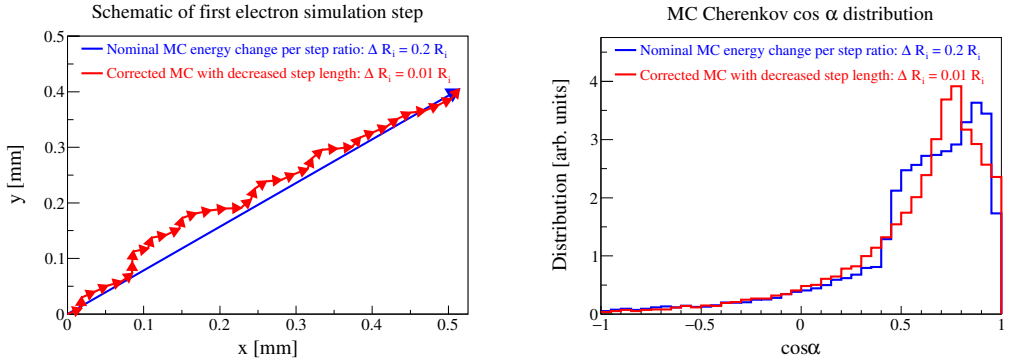


FIGURE 3.3: Dependence of $\cos\alpha$ on the simulated minimal step length of the energy deposition. For GEAN4 physics processes with a simulated continuous energy loss, such as Cherenkov radiation, the maximum step length is limited by the relative change of the ratio $k = \Delta R/R$. (a) Schematic illustration of the electron path for the first energy deposition step. For the normal Borexino energy step ratio $k = 0.2$ (blue) and the corresponding path for the corrected ratio of $k = 0.01$ (red), used in the production of the CID MC. (b) The $\cos\alpha$ distribution between the initial electron direction and the true direction of the Cherenkov photons, for $k = 0.2$ (blue) and $k = 0.01$ (red). The $\cos\alpha$ distribution is incorrect, if the step length is too large. The electrons are simulated with $T = 0.64\,\text{MeV}$, with an expected peak at $\cos\alpha \approx 0.72$.

By default $k = 0.2$ for electrons, which is also the value used for the nominal Borexino MC. For all purposes of the standard Borexino analysis this value is well able to reproduce the data energy spectrum, but for the CID $\cos\alpha$ distribution this value produces incorrect results. This can be seen in Figure 3.3, where the left side is a schematic illustration of the very first simulated step of electrons with a kinetic energy of $T = 0.64\,\text{MeV}$. The nominal value $k = 0.2$

results always in a first step length of 0.7 mm (blue), out of a total path length of ~ 3.6 mm for this electron energy in the LS. Then, the energy loss due to ionisation, subsequent scintillation and Cherenkov radiation is distributed uniformly along this path, as they are simulated as continuous energy loss processes. For the isotropic scintillation this is not of importance, but for the directional Cherenkov radiation this imposes a single Cherenkov angle for a large number of emitted photons. Now a smaller step ratio of $k = 0.01$ is shown in red in Figure 3.3, with a path, comparable to that of $k = 0.2$. The smaller step-size results in a significantly different angular distribution of the emitted Cherenkov photons, as the electron changes its direction after each small step.

This is shown on the right side of Figure 3.3, with the $\cos\alpha$ distribution between the initial electron direction and the true direction of the Cherenkov photons. For $T = 0.64$ MeV the expected value of the Cherenkov angle is $\cos\alpha \approx 0.72$, which is then smeared by the multiple scattering change of the electron direction. For $k = 0.2$ (blue) this peak is split in two by the Cherenkov photons of the first step, with a relative large simulated step size, while for $k = 0.01$ (red) the expected Cherenkov peak is clearly visible in the MC. Smaller values of $k < 0.01$ produce the same results as $k = 0.01$, which is therefore selected as the maximum step length ratio for the CID MC. This step length does not change the total number of scintillation or Cherenkov hits produced.

3.2.2 Cherenkov and scintillation photons

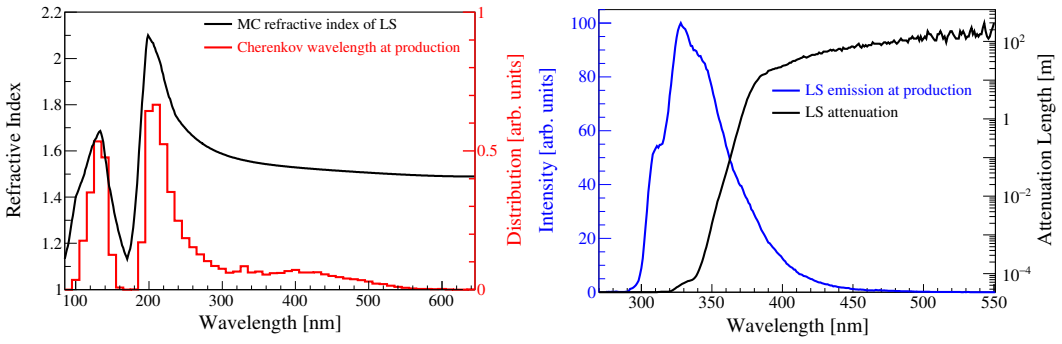
Ultimately the analyses of the CID $\cos\alpha$ distribution provides a number of detected signal Cherenkov hits, correlated to the solar position, for a given number of selected data events. In order to then infer the number of solar neutrino events from the $\cos\alpha$ distribution, it is necessary to have the correct ratio of detectable Cherenkov hits per neutrino event in the MC model In G4Bx2 the average number of Cherenkov photons N for electron events per path-length dx and wavelength $d\lambda$ is calculated according to the Frank-Tamm equation [14, 105]:

$$\frac{d^2N}{dxd\lambda} = \frac{2\pi\alpha}{\lambda^2}\left(1 - \frac{1}{n(\lambda)^2\beta^2}\right) \tag{3.5}$$

Here α is the fine-structure constant, $n(\lambda)$ is the wavelength depended refractive index of the LS and β is the electron velocity over c. From this, the number of simulated photons is calculated from a Poisson distribution with a mean proportional to the simulated step length dx, following the λ^{-2} and $n(\lambda)^{-2}$ dependencies.

Figure 3.4 shows the refractive index implemented in G4Bx2 in black. The corresponding wavelength distribution of the simulated Cherenkov photons is shown in red for electrons with $T = 0.64$ ($\beta = 0.896$). Most Cherenkov photons are below 300 nm, as is expected from $N \sim \lambda^{-2}$. Around ~ 170 nm the simulated energy is too small to produce Cherenkov photons for the given refractive index, as the condition of $\beta > n(\lambda)^{-1}$ is not fulfilled. On the right side of Figure 3.4 the emission profile of the LS is shown in blue and the attenuation length of the LS is shown in black, in a logarithmic scale. In the MC, photons below 310 nm are immediately absorbed, while above 310 nm the physics processes associated to the attenuation (e.g. absorption, Rayleigh scattering) are taken fully into account for the different LS constituents (PC, PPO). If a photon is absorbed in the LS it has a 82% probability to be re-emitted with a wavelength following the PPO scintillation emission profile.

The PMTs are sensitive between about 290 nm to 550 nm, but the LS absorption effectively limits the wavelength of detected photons to $\lambda > 370$ nm. The Cherenkov photons contribute about $\sim 5\%$ to the total light-yield, but most of this contribution is absorbed and re-emitted as isotropic scintillation light, following the emission and time profile of the LS. The direct Cherenkov hits, which are detected without being absorbed and re-emitted, and are thus able to provide the directional information, make up only $< 0.5\%$ of all hits in Borexino.



(a) (b)

FIGURE 3.4: (a) Refractive index (black) used in the Borexino MC and the corresponding wavelength distribution of the simulated Cherenkov photons (red) at time of production, before absorption. (b) Attenuation length used in the MC (black), in a logarithmic scale. In the simulation photons with a wavelength below 310 nm are absorbed immediately. For comparison the scintillation emission profile of the LS (blue) is also shown before absorption.

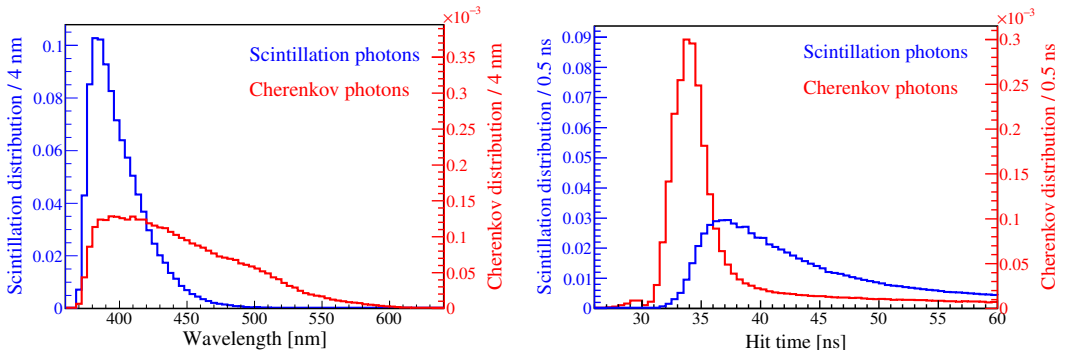
Throughout this thesis the term "Cherenkov hits" is therefore used explicitly for photons that have been produced in the Cherenkov process but have not been absorbed in the LS, while the term "scintillation hits" now includes both the initial scintillation photons, as well as those Cherenkov photons that are absorbed and re-emitted by the LS.

The number of emitted scintillation photons and their time distribution depend on the molecular interactions between solvent (PC) and the wavelength shifter (PPO), as well as the energy loss processes of the charged particles in the LS. In G4BX2 the number of emitted scintillation photons is modeled with the Birks formula [105, 103], an empirical description of the light-yield Y_p^{ph} of a particle p that deposits the energy E in the LS over a path-length dx, with a stopping power dE/dx:

$$\begin{aligned} Y_p^{\mathrm{ph}} &= Y_0^{\mathrm{ph}} Q_p(E) E \\ Q_p(E) &= \frac{1}{E}\int_0^E \frac{dE}{1+k_{\mathrm{B}} dE/dx} \end{aligned} \tag{3.6}$$

Here $Q_p(E)$ is the particle specific quenching factor, which is material dependent through the Birks factor k_{B} and the stopping power dE/dx. This $Q_p(E)$ describes the ionisation quenching, which introduces an intrinsic non-linearity between the deposited energy E and the number of emitted scintillation photons. Both the primary scintillation yield Y_0^{ph} and k_{B} are empirical parameters that have to be determined for every particular scintillator. In G4BX2 these values are set to $Y_0^{\mathrm{ph}} = 11700\,\mathrm{photons/MeV}$ and $k_{\mathrm{B}} = 0.0109\,\mathrm{cm/MeV}$ for electron events.

The left of Figure 3.5 shows the wavelength distribution for Cherenkov light (red) and scintillation (blue), as detected by the PMTs in the MC. This includes the effects of absorption and re-emission, as well as the quantum efficiencies (QE) of the PMTs. For comparison, the Cherenkov distribution is scaled to the scintillation, but it has its own y-axis. The right side of Figure 3.5 shows the time-of-flight corrected PMT hit time distribution for Cherenkov light (red) and scintillation (blue). In the MC Cherenkov photons are emitted instantly, while the scintillation follows a multi-exponential decay, with a fastest decay time of 3.7 ns (see Table VI in [105]). The detected PMT hit times are further delayed and smeared through absorption and re-emission, scattering in the LS and on the detector materials, the PMT transit time spread, electronic signal readout and the event reconstruction. The simulated PMT hits in Figure 3.5 are from ^{7}Be-ν events, where the recoil electrons are selected close to the



(a) (b)

FIGURE 3.5: (a) Borexino MC wavelength spectra for Cherenkov light and scintillation as detected by the PMTs, including the effects of absorption and re-emission. (b) MC distributions of the hit times corrected with their time-of-flight, for solar ^{7}Be-ν recoil electrons from Chapter 6. The left y-axis shows the scintillation (blue), where the area is normalised to 1 and the right y-axis corresponds to Cherenkov light (red), normalised to the number of Cherenkov hits relative to scintillation ($< 0.4\%$). The scintillation includes those Cherenkov photons that have been absorbed and re-emitted by the LS.

Compton-like edge ($T \approx 0.6\,\mathrm{MeV}$), following the data selection from Chapter 6. The $\sim$34 ns offset corresponds to the time-of-flight of photons emitted in the detector centre and the detected Cherenkov hits are mostly distributed in the peak around this hit time. In comparison, the scintillation hits are distributed much more broadly and are detected, on average, much later. However, scintillation hits are ~ 250 times more abundant and as such Cherenkov hits are sub-dominant even at early times.

The following conclusions can be drawn from these plots: It is necessary to perform some sort of time-cut to improve the Cherenkov to scintillation hit ratio, as otherwise the Cherenkov signal would be lost in the statistical noise of the scintillation hits. Since such a time-cut must be applied to both the data and the MC, it can introduce a systematic uncertainty if the hit-time distribution of the data differs from the hit-time distribution of the MC. The refractive index implemented in the MC is taken from a fit on a laboratory measurement with finite statistics and possible systematic uncertainties. This means that the MC refractive index can be systematically different than the real refractive index in data. Likewise, the wavelength distribution of the detected PMT hits can be different in the MC and the data, as it has never been measured in-situ. As such, the effective group velocities for photon propagation can be different in the data and the MC, introducing a difference in the detected PMT hit time distributions. All previous Borexino calibrations are explicitly performed in the context of the spectral analysis, where the focus is to optimise the accuracy of the energy scale between the MC simulation and the data. The total light-yield and the scintillation parameters are provided by laboratory measurements, as well as the tuning of the MC with radioactive calibration sources [105, 85]. Additionally, the intrinsic background events (^{11}C, ^{210}Po, cosmogenic neutrons) are used for a continuous calibration of the PMT quantum efficiencies, and to measure the agreement of the energy scales between the data and the MC over time. This results in a systematic uncertainty on the energy response of 0.34% [139].

In contrast, the calibration of the MC hit time behaviour has been performed with less effort, where the scintillation decay times are tuned to give a general agreement between the data and the MC hit time histograms, without a specific focus on early times. For the spectral analysis the hit time distribution is of secondary importance, as its most important use is the position reconstruction algorithm used for the definition of the fiducial volume (FV) and for the α/β-particle discrimination. Both these use cases are evaluated quantitatively under

the conditions of the tuned MC scintillation parameters and are considered good enough, i.e. with sufficiently small systematic differences between the data and the MC. This effectively bypasses an in-depth analysis of the PMT hit time differences between the data and the MC, because the impact such differences would have on the relevant algorithms is measured explicitly. Moreover, the Cherenkov hit time behaviour has not been investigated at all for the Borexino LS. This is negligible for the spectral analysis, but it now necessitates a CID specific calibration, to investigate and minimise possible systematic differences in the PMT hit time behaviour between the MC and data, especially at early hit times. The first step for this task is explained in Section 3.3, introducing time like variable "Nth-Hit".

3.2.3 Implementation of the solar direction

The last property missing in G4BX2 for the production of the CID model is the implementation of the solar direction. For this it is important to correctly reproduce the CID hit morphology of the data. Borexino started data taking in May 2007 with 2072[1] live PMTs and it ended data taking in October 2021 with 1130 live PMTs [140], where 942 PMTs have died during the Detector live time [2]. The MC simulation is performed with the knowledge of the live PMT distribution on a run-by-run basis. For individual runs some live PMTs can also have been turned off due to some misbehaving characteristics, such as too much dark noise. This means that the underlying, possible PMT hit morphology can look drastically different for different events, given the corresponding position of the Sun at the time of the event trigger. For example, if the Sun is positioned at the north-east, but the south-west region of the detector has no live PMTs, then no neutrino signal Cherenkov hits can be detected and here the possible hit morphology would be such, that signal and background $\cos\alpha$ distributions would look the same. As a data taking run is about six hours long, even events with the same live PMT distribution can have a vastly different underlying, possible $\cos\alpha$ distribution, because the Sun can be situated at significantly different positions for different events.

For this reason the production of the CID MC is performed on a data event-by-event basis. Given a data event and its corresponding solar direction $(\varphi, \cos\theta)$ a number of 200 MC events are simulated using the same live PMT distribution as this data event. For each data event the neutrino signal is simulated with the initial solar direction $(\varphi, \cos\theta)$ and the recoil-electron is scattered with the angle given by Equation 3.1, according to the sampled, transferred kinetic energy T. Then, the background MC is also produced for each selected data event, where now the initial direction of the corresponding background particle is sampled isotropically. Both MC signal and background events are simulated within a sphere of 15 cm around the reconstructed position of the associated data event. The data is selected according to a fiducial volume cut and within an energy region of interest and finally the MC events are then also selected according to the same cuts. Thus, for a given data event the MC production of the $\cos\alpha$ distributions takes fully into account the live PMT distribution.

The question now is if this procedure introduces a systematic uncertainty. Ideally the number of MC events per data event should be infinite to correctly produce the possible, underlying PMT hit morphology, but due to computation time constraints only 200 MC events are produced per data event. There could be a dependence of the MC $\cos\alpha$ distribution on the simulated position, due to the geometric acceptance of the PMTs; closer PMTs have a higher chance to get hit than PMT further away from the event position. Lastly, the neutrino signal is uniformly distributed, while the data background can be non-uniformly distributed within the fiducial volume. The MC production cannot take this into account easily, as each data event is used for the production of both MC signal and background. All these things could

[1] In total 2212 PMTs have been installed, while some PMTs were already inactive at the start of the data taking.

[2] It is interesting to note that PMTs with lower QE die faster. The same properties that give a PMT a high QE also seem to give it a longer live time. Consequently, the average PMT QE increases over time [141].

add systematic differences between the CID data and the MC model, potentially introducing a systematic uncertainty in the MC $\cos\alpha$ distributions. This is explored in Chapter 6, where the proof of the CID principle is provided and it is found that none of the aforementioned problems contribute a systematic uncertainty. The CID MC model, as will be shown later, is well able to reproduce the CID data $\cos\alpha$ distribution.

3.3 The Nth-Hit time like variable

Each event has a reconstructed event position $\vec{r}_{\text{ev}}$, an associated reconstructed event time t_{ev} and a number of PMT hits. These PMT hits have a hit time $t_{\text{hit}}^{\text{raw}}$ each, normalised to 0 ns for the earliest detected PMT hit, and the associated PMT has a position $\vec{r}_{\text{PMT}}$. Then the time-of-flight (ToF) corrected hit time $t_{\text{hit}}^{\text{ToF}}$ is calculated using the effective refractive index $n_{\text{g}}^{\text{eff}}$ and the distance d_{hit} between the event position and the hit PMT:

$$\begin{aligned} d_{\text{hit}} &= \mid \vec{r}_{\text{ev}} - \vec{r}_{\text{PMT}} \mid \\ t_{\text{hit}}^{\text{ToF}} &= t_{\text{hit}}^{\text{raw}} + t_{\text{ev}} - \frac{n_{\text{g}}^{\text{eff}}}{c} \cdot d_{\text{hit}} \end{aligned} \tag{3.7}$$

This refractive index $n_{\text{g}}^{\text{eff}}$ describes the effective group velocity $v_{\text{g}}^{\text{eff}} = c/n_{\text{g}}^{\text{eff}}$ of the detected photon hits, which represents the average velocity of the detected photon hits in the LS. The reconstructed event time t_{ev} is an estimation of the time difference from the first detected PMT hit relative to the start time of the event energy deposition, plus an arbitrary time offset of $\sim$34 ns that corresponds to the ToF from photons emitted in the detector centre. This value is the same constant for all events and positions. Thus, the ToF corrected PMT hit times $t_{\text{hit}}^{\text{ToF}}$ correspond to a reconstruction of the photon emission time for a start time of the event energy deposition at 0 ns, normalised to the detector centre position. The $t_{\text{hit}}^{\text{ToF}}$ is the fundamental parameter on which a time sorting, or a time cut of some sort can be applied. This is necessary to improve the Cherenkov over scintillation hit ratio, as Cherenkov photons on average are emitted at earlier times relative to the scintillation.

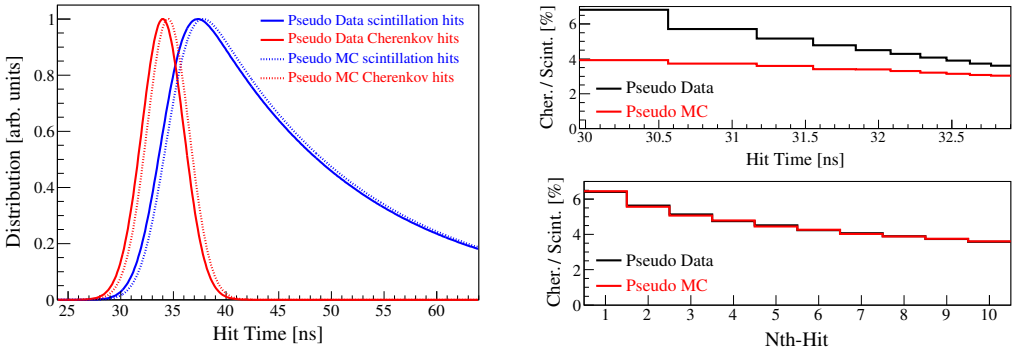
For the CID application in Borexino the time like Nth-Hit variable is used instead of the absolute hit time in units of nanoseconds. The Nth-Hit is defined by the sorted time position of the ToF corrected PMT hits of each individual event. Every event has a number of PMT hits, with an associated $t_{\text{hit}}^{\text{ToF}}$ (Equation 3.7) and $\cos\alpha$ (Equation 3.3). The $t_{\text{hit}}^{\text{ToF}}$ with the smallest value is the first Nth-Hit (or just first hit), the next is the second Nth-Hit (second hit) and so on. Integrating over all selected events results in a $\cos\alpha$ histogram for each Nth-Hit of all events. The first hits of the events should then have the largest Cherenkov / scintillation hit ratio and each subsequent Nth-Hit $\cos\alpha$ histogram should have a smaller Cherenkov / scintillation ratio. The Nth-Hit variable is of fundamental importance for the CID analysis in Borexino, as it allows for the time sorting of Cherenkov and scintillation hits, greatly improving the CID analysis sensitivity. This time sorting could also be performed on the absolute time scale, but the use of the Nth-Hit variable effectively reduces the number of free parameters that are necessary to describe the potential systematic differences in the hit time distributions of the data and the MC. The reasons for this are twofold:

1. Nth-Hit is a *relative* time variable, where the *absolute* differences in the underlying hit time distributions between data and MC are irrelevant.

2. Every event contributes a single hit for each Nth-Hit $\cos\alpha$ histogram, whether it is a neutrino signal event or a background event, independent of the background type.

To illustrate and explain the impact of Nth-Hit, compared to the use of an absolute time scale, it is sensible to look at a simplified model of the hit time distribution. This is done here

with a toy-MC of the hit time distribution, modelled as the convolution of a Gaussian with an exponential decay for the scintillation and as a Gaussian for the Cherenkov hits. The mean μ describes the start time of the event plus the ToF of photons from the detector centre, the standard deviation σ describes the detector response and τ represents the effective exponential decay of the scintillation, including the delay from absorption and re-emission.



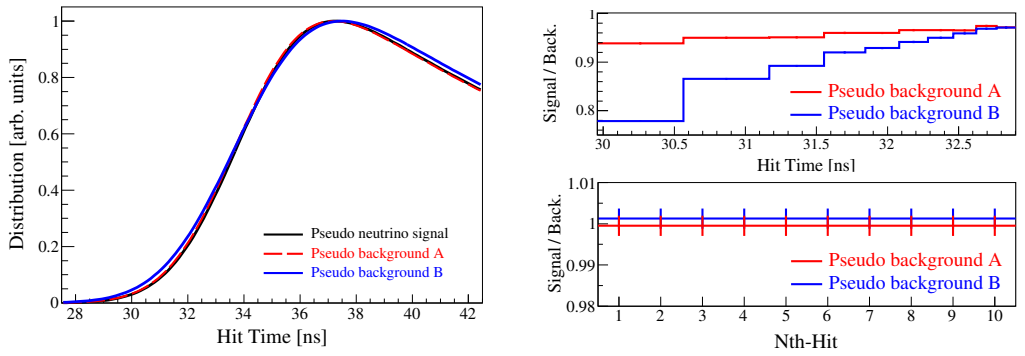
(a) (b)

FIGURE 3.6: The impact of the Nth-Hit variable on the time dependent Cherenkov / scintillation hit ratio, illustrated with a toy-MC model. (a) Hit time distributions of pseudo-data (solid lines) and pseudo-MC (dotted lines) for Cherenkov hits (red) and scintillation hits (blue). Cherenkov hits are modeled as a Gaussian (μ, σ), scintillation hits are modeled as an exponentially (τ) modified Gaussian. Pseudo-data and pseudo-MC are different only in their Gaussian mean value $\mu_{\text{data}} = 34.0\,\text{ns}$, $\mu_{\text{MC}} = 34.5\,\text{ns}$. The injected ratios of Cherenkov / scintillation hits is 0.4% for the full scintillation time profile. (b) Using the hit time distribution from (a) 10k toy-MC events are sampled with 500 hit times each. At the top, the pseudo-data is used to define time bins, such that each bin includes exactly 10k hits and the Cherenkov / scintillation hit ratio is plotted for pseudo-data (black) and pseudo-MC (red). The ratios are significantly different between pseudo-data and pseudo-MC, where pseudo-MC has lower values due to the constant time offset. At the bottom the Cherenkov / scintillation hit ratio is plotted for the time relative Nth-Hit variable, where pseudo-data and pseudo-MC are in agreement

Figure 3.6(a) now shows the model neutrino signal event hit time distribution of pseudo-data as a solid line and pseudo-MC as a dotted line, for Cherenkov hits in red and scintillation hits in blue. The parameters are $\mu_{\text{data}} = 34.0\,\text{ns}$, $\mu_{\text{MC}} = 34.5\,\text{ns}$, $\sigma_{\text{data}} = \sigma_{\text{MC}} = 2\,\text{ns}$ and $\tau_{\text{data}} = \tau_{\text{MC}} = 15\,\text{ns}$. These parameters model the effect of a constant offset between the data and the MC hit time distributions. Such a difference could, for example, come from a small difference in the group velocity (refractive index) between the data and the MC or from a small misalignment in the PMT time calibration. The right Figure 3.6(b) shows the Cherenkov / scintillation ratio of the toy-MC with 10k pseudo events, each with 500 hits sampled from the corresponding hit time distributions. The injected ratio of Cherenkov / scintillation hits is 0.4% for the full scintillation time profile, two Cherenkov hits per pseudo event. The time dependent Cherenkov / scintillation ratio is shown in black for the pseudo-data and in red for the pseudo-MC. At the top, the hit time distribution is plotted as a histogram with equi-statistic bins, where each data bin includes exactly 10k hits and the same bin borders are used for the pseudo-MC. The first bin begins at 0 ns but is shown to start at 29 ns for illustration purposes. It can be seen that such a binning, effectively a time-cut, improves the Cherenkov / scintillation ratio drastically from 0.4% up to 7% for first time bin. Nonetheless, the scintillation hits are still dominant even at the earliest times. The Cherenkov / scintillation ratio for each bin is different between the pseudo-data and the pseudo-MC, because the bin width is defined by the pseudo-data from which the pseudo-MC has a constant offset. Consequently, the application of a time-cut here would result in a different number of Cherenkov hits per neutrino signal event between the data and the MC.

At the bottom of Figure 3.6(b) the same plot is shown for the Nth-Hit variable. Again each Nth-Hit bin has exactly 10k hits, as each of the 10k events contributed exactly one hit for each Nth-Hit bin. Now the pseudo-data and the pseudo-MC have the same distribution of the Cherenkov / scintillation hit ratio, given the simulated statistics. Thus, the use of the time-relative Nth-Hit variable excludes systematic differences that have the form of a constant hit time difference between the data and the MC.

Ultimately the CID method provides only information about the number of neutrino signal Cherenkov hits, from which then the number of actual neutrino events is inferred using the ratio of "Cherenkov hit per neutrino event". This ratio is estimated from the G4BX2 MC model and includes systematic uncertainties. Contrary to the presented toy-MC, in reality there are differences in all parameters μ, σ, τ between data and MC. While the Cherenkov / scintillation hit ratio of the Nth-Hit variable is still dependent on σ, τ, its use effectively decreases the number of systematic uncertainty sources by one. When using the Nth-Hit variable it is sufficient to measure three parameters for the description of the systematic differences in the hit time distribution between data and MC, to construct an adequate CID MC model: The effective refractive indices (group velocities) of scintillation in the data and the MC and additionally the relative group velocity correction of the Cherenkov hits in the MC. The effective refractive indices of the scintillation are explained and measured in-situ in Chapter 4, using α-particle calibration sources. The MC Cherenkov group velocity correction is explained and measured in Chapter 5, using γ-calibration sources, as there are no dedicated Cherenkov sources for Borexino.



(a) (b)

FIGURE 3.7: The impact of the Nth-Hit variable on the time dependent signal / background hit ratio. (a) The toy-MC hit time distributions are modeled as exponentially modified Gaussian distributions (μ, σ, τ). The neutrino signal (black) and the background A (red) have $\mu = 34$ ns, $\sigma = 2.0$ ns, $\tau = 15$ ns, where their difference is that the signal has a Cherenkov / scintillation hit ratio of 0.4% and background A has 0.8%. The background B (blue) has the Cherenkov / scintillation hit ratio of 0.4% and has a different $\sigma = 2.1$ ns. (b) Using the pseudo-data hit time distribution, made up of equal parts from the signal, background A and background B (1:1:1) to simulate 10k events with 500 hits each. At the top, the pseudo-data is used to define time bins, such that each bin includes exactly 10k hits. The resulting signal / background hit ratios for background A (red) and background B (blue) are shown for each time bin. Both background types contribute significantly more hits than the signal and they result in significantly different ratios. At the bottom, the signal / background hit ratio is plotted for the Nth-Hit variable, where each bin again contains 10k hits. Here both background types give ratios in agreement with each other and in agreement with the injected signal / background *event* ratio

Furthermore, the Nth-Hit variable effectively reduces the number of free parameters necessary for the fit of the CID $\cos\alpha$ distribution, which is illustrated through a toy-MC study in Figure 3.7. Figure 3.7(a) shows the model event hit time distributions for the neutrino signal in black, for one background A in red and for a different background B in blue. For the

CID it is necessary to select a large number of data events to gain enough statistical sensitivity for a measurement. This means that data events are selected within an energy window, e.g. at the ^{7}Be-ν edge region or a region with a large number of CNO+pep-ν events. In this selected energy region the average energy of the neutrino signal events can be different than the average energy of the background events. Thus, if background events have a larger average energy than the ν-signal events, they also produce a larger ratio of Cherenkov / scintillation hits. This is the case at the ^{7}Be-ν edge region for the ^{210}Bi background, for example. Here, this is modeled as the difference between the signal and the background A, where both have the same hit time parameter values $\mu = 34\,\text{ns}$, $\sigma = 2.0\,\text{ns}$, $\tau = 15\,\text{ns}$, but now the signal has a Cherenkov / scintillation hit ratio of 0.4%, while the background A has an exaggerated ratio of 0.8%. Additionally, there are different types of background in Borexino, such as the ^{210}Bi electrons, ^{210}Po α-particles, ^{11}C positrons and γ-particles. Every different background particle has a different, underlying hit time distribution and Cherenkov to scintillation hit ratio. Thus, the background B is modeled to represent ^{11}C positrons also with a Cherenkov / scintillation hit ratio of 0.4%, but with an increased $\sigma = 2.1\,\text{ns}$.

The goal of this toy-MC study is not to investigate the potential impact of systematic differences, but to illustrate the effect of the intrinsic hit time differences between the background species on the signal / background ratio. Figure 3.7(b) shows the resulting signal / background ratios for the background A in red and the background B in blue. A number of 10k pseudo-data events are simulated according to an equal ratio between the signal, the background A and the background B, each with a probability of $1/3$. Every pseudo-event is sampled with 500 hit times from the relevant hit time distributions. At the top, the pseudo-data is shown in equi-statistic time bins, where each bin has exactly 10k hit entries. It can be seen that the signal / background hit ratio is time dependent and both the background A and the background B contribute significantly more hits then the signal, for each shown time bin. The different backgrounds result in significantly different signal / background hit ratios, relative to each other. This means that a fit of the $\cos\alpha$ distribution, given a certain, absolute time-cut, would result in a signal to total *hit ratio* that is different then the signal to total *event ratio*. Thus, the calculation of the number of neutrino signal events would require the knowledge of the number of background events for each different background species. In the best case this would require another free parameter in the $\cos\alpha$ fit, for every additional background species. As different backgrounds have indistinguishable $\cos\alpha$ distributions this would introduce an even greater fit uncertainty in the CID measurement.

At the bottom of Figure 3.7(b) the same 10k pseudo-data events are used to produce the signal / background hit ratios using the Nth-Hit variable. Each Nth-Hit bin has 10k hit entries, and now both the background A and the background B are in agreement with each other and both contribute the same amount of hits as the signal events. Because every event contributes exactly one hit per Nth-Hit there is no time dependent over-selection of background hits and each Nth-Hit has the same signal / background *hit ratio* which is also equal to the signal / background *event ratio*. Using the Nth-Hit variable, a fit of the $\cos\alpha$ distribution only needs a single parameter to describe the number of neutrino events: The signal to total event ratio, given the known number of total selected data events. The conclusion is then, that the relative, time-like Nth-Hit is a powerful variable that minimises both the number of free parameters in the CID $\cos\alpha$ fit, as well as the number of possible sources of systematic uncertainty. It provides a significant contribution to the observed robustness of the CID analysis of the ^{7}Be-neutrinos presented in Chapter 6, as well as the measurement of CNO-neutrinos in Chapter 7.

Chapter 4

Measurement of the Effective Scintillation Group Velocity

Abstract

The total Cherenkov to scintillation ratio in Borexino is relatively small, with $< 0.5\%$ for 1 MeV electron events, which is obstructive to the *Correlated and Integrated Directionality* analysis of the solar neutrinos. To improve this ratio it is possible to perform a hit time sorting using the Nth-Hit variable and select only early hits, as Cherenkov light is emitted instantly while the scintillation follows an exponential decay. To estimate this emission time, it is necessary to correct the time-of-flight of the corresponding PMT hits, which requires the measurement of the effective group velocity of the scintillation photons v_g^{eff}. Here, the ^{14}C–^{222}Rn α-event calibration source is used to produce hit time distributions for each individual PMT and for every source position. Fitting these hit time histograms results in a distribution of the expected arrival times μ of the earliest detectable photon hits, as a function of the known distance d between the PMT and the source position. This distribution allows for the in-situ measurement of the effective group velocity, expressed by the effective refractive index $n_g^{\text{eff}} = c/v_g^{\text{eff}}$. For the data the result is $n_g^{\text{eff}}(d) = (1.6867 \pm 0.0026) - d \cdot (0.296 \pm 0.020)\,10^{-2}\,\text{m}^{-1}$, while for MC it is $n_g^{\text{eff}}(d) = (1.7141 \pm 0.0026) - d \cdot (0.575 \pm 0.015)\,10^{-2}\,\text{m}^{-1}$. The linear distance dependence takes into account the dispersion of the detected photons for the full length of the scintillator volume. These results have a precision of the corresponding time-of-flight correction that is negligible compared to the position resolution of >10 cm. The measurements of the effective refractive indices performed in this chapter provide a good foundation for the calibration of the Cherenkov hit time behaviour and the subsequent measurement of solar neutrinos in the following chapters.

Cherenkov and scintillation light have fundamentally different emission time profiles, where the Cherenkov light is emitted instantly and the scintillation follows a multi-exponential decay profile. For this reason it is possible to apply a time sorting of the detected PMT hits to increase the relative contribution of Cherenkov light in the CID $\cos\alpha$ distribution. In the CID method this is done through the Nth-Hit variable, whose calculation depends on the ToF correction using the effective group velocity v_g^{eff}. This parameter is described by the effective refractive index $n_g^{\text{eff}} = c/v_g^{\text{eff}}$, as defined in Equation 3.7. Prior to this work, the effective refractive index in Borexino has been quoted as $n_{\text{eff}}^{\text{data}} = 1.68$ for data and as $n_{\text{eff}}^{\text{MC}} = 1.66$ for MC events. These values have been determined using calibration data to optimise the position reconstruction algorithm [105]. The exact procedure on how these values have been determined is not explained in any Borexino publication and the values are given without

uncertainty. Furthermore, they are explicitly linked to the position reconstruction algorithm and depend on the particular choices made for the hit time PDF used in the likelihood fit. As such, these values are not so much a property of the LS, but more so a property of the position reconstruction algorithm; a different PDF would result in different optimal values of n_{eff}. The conclusion is then that the correct effective refractive index of the LS must be measured for the data and for the MC and its uncertainty must be estimated, as it represents a systematic uncertainty on the production of the Nth-Hit $\cos\alpha$ histograms. For this reason a new method has been developed to perform an in-situ measurement of the effective group velocity, using the ^{14}C–^{222}Rn source, deployed during the 2008-2009 calibration campaign [85]. As the source has been deployed at a large number of different positions in the detector it is possible to produce distance dependent PMT hit time distributions. Fitting these hit time distributions for each individual PMT, allows one to plot the average arrival time of the earliest detected photons, as a function of the distance between the PMT and the source position. The effective refractive index $n_{\text{g}}^{\text{eff}}$ can then be estimated through a polynomial fit.

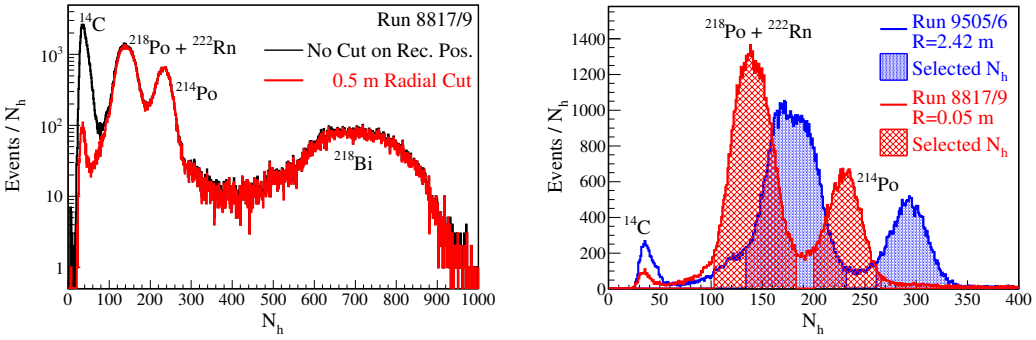
This measurement is explained in some detail, because the methods developed here could be potentially useful for future, large scale LS experiments. First, Section 4.1 explains the event selection from the ^{14}C–^{222}Rn source spectrum. Then, Section 4.2 defines the effective group velocity $v_{\text{g}}^{\text{eff}}$ and the effective refractive index $n_{\text{g}}^{\text{eff}}$ in the context of the CID method and the Borexino detector. Section 4.3 explains in detail the measurement procedure and Section 4.5 presents the results of the measurement and the systematic uncertainties. Finally Section 4.6 concludes this chapter with a summary.

4.1 ^{14}C–^{222}Rn calibration source event selection

During the calibration the vast majority of the calibration points have been obtained with the compound ^{14}C–^{222}Rn source, as it provides α, β, and γ particles in a relatively large energy region. The main purpose of this source has been the measurement of the position reconstruction uncertainty, necessary for the estimation of the FV systematic, as well as the measurement of the position dependent non-uniformity of the energy response [85, 142]. The ^{222}Rn decay chain provides three α emitters: ^{222}Rn, ^{218}Po and ^{214}Po, with energies of 5.5 MeV, 6.0 MeV and 7.4 MeV, respectively. As α particles have a large stopping power dE/dx, they consequently have a large ionisation quenching factor, reducing the β equivalent visible energy. In the Borexino LS this quenching factor is $Q(E) = 20.3 - 1.3E$ [MeV] [143]. For the measurement of the effective scintillation group velocity these α events are of primary interest, as they exclusively provide scintillation without a contamination of Cherenkov light. Additionally, the total number of scintillation photons is relatively low, which is important because it makes sure that the PMT hit time distributions are well within the single photon-electron regime. The impact that multiple photon-electrons could have on the detected hit time does not need to be taken into account here. The β-emitters ^{14}C and ^{214}Bi are not of interest here, but they are also present in the source spectrum.

Figure 4.1(a) shows the energy distribution of the events (black) from Run 8817 and Run 8819, for which the ^{14}C–^{222}Rn source is deployed close to the detector centre. The energy estimator N_{h} used here is the raw number of hit PMTs, without a geometric correction and without a normalisation to 2000 live PMTs. The true position of the source is known from pictures taken by CCD cameras, with a precision of $\lesssim 1$ cm [142]. Applying a cut of 0.5 m between the reconstructed event position and the true source position drastically reduces the contamination from non-source events, mostly from ^{14}C. The peaks of ^{214}Bi, ^{14}C and ^{214}Po are clearly visible, while ^{222}Rn, ^{218}Po are not distinguishable due to the N_{h} statistics.

Two independent energy regions are selected for the measurement of the effective scintillation group velocity: The ^{214}Po peak and the ^{222}Rn+^{218}Po peak. This makes it possible



(a) (b)

FIGURE 4.1: (a) Distribution of the number of hit PMTs N_h for the ^{14}C–^{222}Rn source for Run 8817 and Run 8819, with the source deployed in the centre of the detector. In black: All detected events, in red: Events with a reconstructed position with a 0.5 m distance to the source position. (b) The N_h distribution with the source at different positions. For a radial position of 2.42 m in blue and for 0.05 m in red. Events are selected if the reconstructed position is within 0.5 m of the source position. The measurement of the effective group velocity is performed two times, on independent energy regions. These are shown as the red and blue bands for the respective source positions.

to investigate a potential energy dependent systematic uncertainty. Figure 4.1(b) shows the selected N_h region for two different source positions. In red, the source is close to the detector centre with a position radius of $R = 0.05$ m, while in blue the source is off-centre at $R = 2.42$ m. It can be seen that the ^{214}Po and ^{222}Rn+^{218}Po peaks are shifted to higher N_h values for the larger radial position. This is due to the greater geometric acceptance of the PMTs that are now closer to the source position. The events of interest are selected for each peak within $\pm 2\sigma$ around the mean of a Gaussian fit, shown as the red and blue bands. In total there are 151 source positions used in this analysis. The calculation of the ^{214}Po and ^{222}Rn+^{218}Po selected N_h regions has been automated, but verified by eye. Here the focus of the event selection is to exclude a contamination of the events by ^{14}C on the left and by ^{214}Bi on the right, while providing a clear separation between the ^{214}Po and ^{222}Rn+^{218}Po selected N_h regions. It is not a problem if there is some fraction of ^{214}Po events within the N_h region of ^{222}Rn+^{218}Po, or vice versa, as the they are all α emitters. For the investigation of an energy dependent systematic uncertainty it is enough that the average event energy is different for the two selected N_h regions. The nominal analysis is performed on the ^{214}Po events for the data and for the MC, while the analysis result of ^{222}Rn+^{218}Po is used for the estimation of the energy dependence systematic uncertainty.

4.2 The effective group velocity

The group velocity of a photon wave packet is different than its phase velocity due to the effect of dispersion in the medium. Given the wavelength dependent refractive index $n(\lambda)$ the group velocity $v_g(\lambda)$ can then be expressed with $n_g(\lambda)$, as shown in Equation 4.1. Here $n_g(\lambda)$ corresponds effectively to the function of the refractive index in the absence of dispersion:

$$\begin{aligned} n_g(\lambda) &= n(\lambda) - \lambda \frac{dn}{d\lambda} \\ v_g(\lambda) &= \frac{c}{n_g(\lambda)} \end{aligned} \tag{4.1}$$

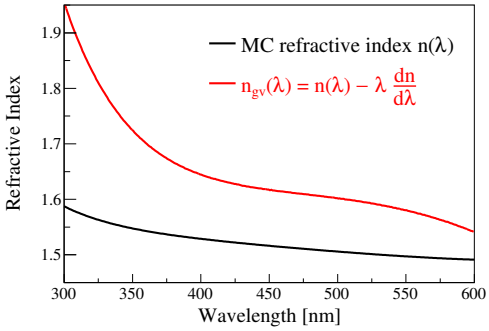
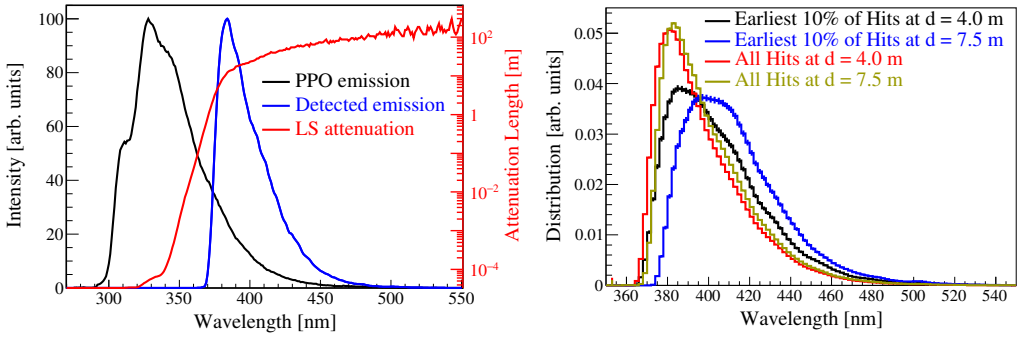


FIGURE 4.2: The refractive index $n(\lambda)$ of the Borexino MC (black) in comparison to $n_g(\lambda)$ (red), which effectively describes the photon group velocity through $v_g(\lambda) = c/n_g(\lambda)$.

Figure 4.2 shows the comparison of the normal MC refractive index $n(\lambda)$ in black with $n_g(\lambda)$ in red. In the wavelength region where photons can be detected, between about 370 nm to 600 nm, the refractive index varies by about $\Delta n = 0.05$. The variation of $n_g(\lambda)$ and the group velocity is much greater at $\Delta n_g = 0.14$, due to the impact of the derivative $dn/d\lambda$. This also means that relatively small differences in the refractive index between the data and the MC can introduce relatively large systematic differences in the resulting group velocity and propagation time of the photons.



(a) (b)

FIGURE 4.3: (a) The MC PPO emission spectrum (black) compared to the detected spectrum at the PMTs for electron events simulated uniformly in a 3 m radius. The change between the emitted and detected scintillation spectrum is due to the absorption by PPO, which dominates the LS attenuation (red) below ~390 nm. (b) The MC wavelength distribution of the detected hits for different distances d between the true event position and the hit PMT. At $d = 4.0$ m (red) there is a small, but significant difference compared to $d = 7.5$ m (yellow) if all PMT hits are considered. If only the earliest 10% of all detected PMT hits are considered, then $d = 4.0$ m (black) and $d = 7.5$ m (blue) have a much larger difference between each other. Longer distances result in more red-shifted wavelength distributions due to PPO absorption and Rayleigh scattering.

The group velocity can be calculated for a single photon wavelength from the refractive index. Because the PMTs detect a broad wavelength distribution of photons the effective group velocity could then be defined as the average value of the product from the group velocity function, weighted by the wavelength probability distribution. In the context of the Cherenkov directionality this is more complicated, because there is a need to focus on the detection time of *early* PMT hits to be able to perform a time separation between Cherenkov light and scintillation. Figure 4.3(a) shows the PPO emission spectrum (black) as it is used by the Borexino MC, the PPO attenuation (red) for the concentration of 1.5 g/L in the LS and the PPO scintillation spectrum as it is detected by the PMTs (blue), after travelling through the detector. The events are simulated in a 3 m radius spherical volume around the detector center. The scattering length in the LS is not shown separately, but it is negligibly large compared

to the absorption length of PPO below $\sim 390\,\text{nm}$. The PPO emission spectrum peaks around 330 nm which is well within the region of its self absorption and consequently the detected scintillation spectrum is shifted to higher wavelength values with a peak around 380 nm. This means that most of the emitted scintillation photons are absorbed and re-emitted and only photons with a wavelength $> 370\,\text{nm}$ are detectable in Borexino, according to the MC.

The effect of the PPO absorption and Rayleigh scattering is illustrated in Figure 4.3(b). Here the MC wavelength of the detected PMT hits is shown for different distances d between the simulated event position and the hit PMT. Looking at all detected hits with $d = 4.0\,\text{m}$ (red) in comparison to $d = 7.5\,\text{m}$ (yellow) shows a small, but significant red-shift for the larger distance between their respective distributions. This is in accordance with the expected behaviour due to PPO absorption. Looking specifically at the earliest 10% of the detected PMT hits for the same distances $d = 4.0\,\text{m}$ (black), $d = 7.5\,\text{m}$ (blue) shows a much larger difference in the their wavelength distribution. Due to the effects of Rayleigh scattering, PPO absorption as well as the wavelength dependence of the group velocity, there is a selection bias for the wavelength distribution of early hits. First, those photons that are produced at larger wavelengths travel faster through the LS because they have a larger group velocity (smaller n_g), relative to photons with a smaller wavelength. Second, those photons that underwent the least amount of scattering and no absorption and re-emission tend to be detected earlier because they have the shortest, most direct path. Therefore, the detected wavelength distribution of the earliest 10% of PMT hits is much broader and with a more pronounced tail to longer wavelengths, than the wavelength distribution that includes all PMT hits. Larger distances select PMT hits at longer wavelengths, which is more pronounced for early PMT hits, where the impact of Rayleigh scattering becomes more important, compared to all (time independent) PMT hits, where only the PPO absorption influences the detected scintillation spectrum.

This means that the effective scintillation group velocity of early PMT hits depends on the refractive index of the LS, the scintillation wavelength distribution as well as the PPO absorption length and Rayleigh scattering length in the LS. All these parameters are implemented in the MC from laboratory measurements and have statistical uncertainties, as well as unknown systematic uncertainties. Thus, the group velocity of early PMT hits is likely different between the data and the MC due to the complex interplay and selection bias of these factors. For the production of the CID Nth-Hit $\cos\alpha$ distributions it is necessary to provide an in-situ measurement of the effective group velocity of early hits for both the data and the MC. This is done in the following sections using the ^{14}C–^{222}Rn calibration sources.

4.3 Fitting of the PMT hit time distributions

The following measurement of the effective group velocity of early PMT hits is defined by a phenomenological formula, through the use of the effective refractive index n_g^eff. Let $\mu(d)$ be the expected, average, true emission time of the earliest detectable photons for a given distance d between the source position and the hit PMT. Consequently, the Tof of the photons is given by the effective group velocity $v_\text{g}^\text{eff} = c/n_\text{g}^\text{eff}$. The ToF distribution is already smeared, for a given distance, due to the broad distribution of the scintillation wavelength. The Gaussian standard deviation σ describes (approximately) this ToF smearing in addition to the detector response from the PMT transit time spread, the read-out electronics, and event reconstruction algorithms. With this the empirical definition of the effective group velocity of early PMT hits follows from a measurement of $\mu(d) = d \cdot v_\text{g}^\text{eff} = d \cdot c/n_\text{g}^\text{eff}$.

The value of $\mu(d)$ can be estimated by fitting the measured PMT hit time t_hit distribution, which is defined similarly to Equation 3.7, but here $t_\text{hit} = t_\text{hit}^\text{raw} + t_\text{ev}$ explicitly excludes the ToF correction. The multi-exponential time decay of the scintillation, as well as the effect of absorption and re-emission is modeled effectively through a bi-exponential decay with the

parameters τ_1, τ_2, and the relative contribution of each exponential decay component R. The fit is then defined as the convolution between the Gaussian and the exponential decay:

$$\begin{aligned} f(t) =& A\, R\, e^{\frac{1}{2}\left(\frac{\sigma}{\tau_1}\right)^2 - \frac{t-\mu}{\tau_1}}\, \mathrm{erfc}\left(\frac{1}{\sqrt{2}}\left(\frac{\sigma}{\tau_1} - \frac{t-\mu}{\sigma}\right)\right) + \\ & A\,(1-R)\, e^{\frac{1}{2}\left(\frac{\sigma}{\tau_2}\right)^2 - \frac{t-\mu}{\tau_2}}\, \mathrm{erfc}\left(\frac{1}{\sqrt{2}}\left(\frac{\sigma}{\tau_2} - \frac{t-\mu}{\sigma}\right)\right) + \\ & A_{\mathrm{ref}}\, R\, e^{\frac{1}{2}\left(\frac{\sigma_{\mathrm{ref}}}{\tau_1}\right)^2 - \frac{t-\mu_{\mathrm{ref}}}{\tau_1}}\, \mathrm{erfc}\left(\frac{1}{\sqrt{2}}\left(\frac{\sigma_{\mathrm{ref}}}{\tau_1} - \frac{t-\mu_{\mathrm{ref}}}{\sigma_{\mathrm{ref}}}\right)\right) + \\ & A_{\mathrm{ref}}\,(1-R)\, e^{\frac{1}{2}\left(\frac{\sigma_{\mathrm{ref}}}{\tau_2}\right)^2 - \frac{t-\mu_{\mathrm{ref}}}{\tau_2}}\, \mathrm{erfc}\left(\frac{1}{\sqrt{2}}\left(\frac{\sigma_{\mathrm{ref}}}{\tau_2} - \frac{t-\mu_{\mathrm{ref}}}{\sigma_{\mathrm{ref}}}\right)\right) \end{aligned} \tag{4.2}$$

Here A is the total amplitude of the function, given the statistics of the data and A_{ref}, μ_{ref}, σ_{ref} is the effective description of a possible reflection peak. The reflections can occur on the stainless steel vessel, on the PMT cathode, and inside the PMT light collection cones. This can introduce a non-negligible secondary peak at later times in the hit time distribution. This relatively complex behaviour is included here, as otherwise the choice of a particular time cut would be necessary to exclude this reflection peak, which could introduce a systematic fit bias. Given the measured PMT hit time distribution $N(t_{\mathrm{hit}})$ in form of a histogram, with a number of I bins, the fitting is then performed with a Poissonian log-likelihood ratio test statistic λ_{LR}:

$$\lambda_{\mathrm{LR}} = -2\sum_{i=0}^{I} \log\left(\frac{f(t_{\mathrm{i}})^{P(t_{\mathrm{i}})}\; e^{-f(t_{\mathrm{i}})}}{N(t_{\mathrm{i}})!}\right) \tag{4.3}$$

The hit time distribution $N(t_{\mathrm{hit}})$ of the data is produced for each PMT and each source position individually. For the 1827 live PMTs during the calibration campaign this could result in $\sim 270\,\mathrm{k}$ hit time histograms. In reality, most source positions do not provide enough hit statistics for most PMTs to be fitted, resulting in a number of $\sim$ 54k usable histograms for the data. Here, the minimum statistic is limited to histograms that have a peak value of at least 30 hit entries. The hit time distribution of the MC is produced in the same way as the data: For each source position a number of ^{214}Po events is simulated with about ~ 1.5 times the statistics of the selected data events. The minimisation of $\lambda_{\mathrm{LR}} = \lambda_{\mathrm{LR}}(A, \mu, \sigma, R, \tau_1, \tau_2, A_{\mathrm{ref}}, \mu_{\mathrm{ref}}, \sigma_{\mathrm{ref}})$ is performed with the MINUIT [144] algorithm of ROOT [104]. The fit is performed once with a binning of 1.0 ns and once with a binning of 0.5 ns for the estimation of the systematic uncertainty from the binning selection.

The parameter of interest here is $\mu(d)$, while the other parameters give supplementary information, that could in principle be used for a relatively precise fine-tuning of the MC. A plot of $\mu(d)$ vs. d allows then the correct calculation of the effective reflective index of early hits for scintillation light in Borexino, including the effects of the photon wavelength distribution, Rayleigh scattering and the full detector response.

The fitting is done by an automated script and a sample of the results is cross-checked by eye. The performance of the fit function from Equation 4.2 is exemplary illustrated in Figure 4.4 for a relatively small PMT-source distance $d = 3.58\,\mathrm{m}$. The measured hit time distribution of PMT No. 1426 (black dots) is plotted on the left for the data and on the right for the MC. The yellow line is the best fit function $f(t)$, while the red bands represent the Poissonian 68.3% and 95.5% confidence intervals. The blue line shows the position of the best fit value of μ, with the uncertainty of the fit $\Delta\mu = \sim 0.1\,\mathrm{ns}$. The influence of the (effective) exponential decay times can be seen by the relative large difference between the peak of the function and μ, which represents the average arrival time of the earliest detectable photon hits.

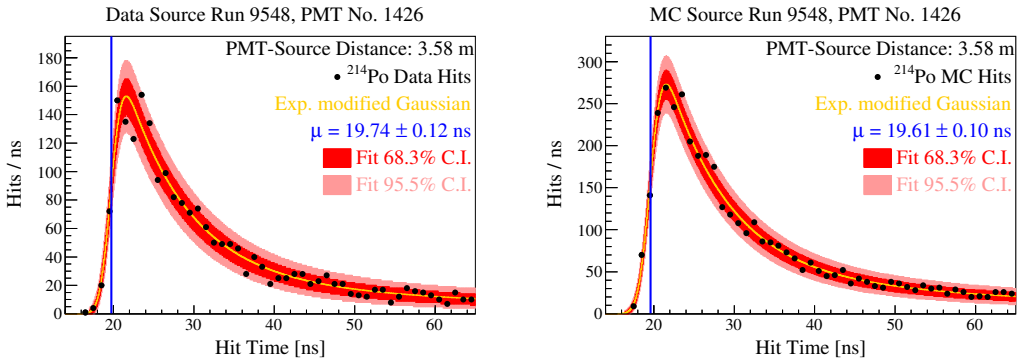


FIGURE 4.4: Exemplary fit for a short PMT-source distance of 3.58 m. The hit time distribution is shown as black dots, the best fit as a yellow line and the Poissonian confidence intervals as red bands. The best fit for the fit parameter μ, describing the average arrival time of the earliest detectable photon hits, is shown as a blue line. (a) ^{214}Po data fit. (b) ^{214}Po MC fit.

The reflection peak (A_{ref}) is not visible here, as it is cut off; it arrives at around $\sim$100 ns and it is negligibly small due to the geometric acceptance of this PMT.

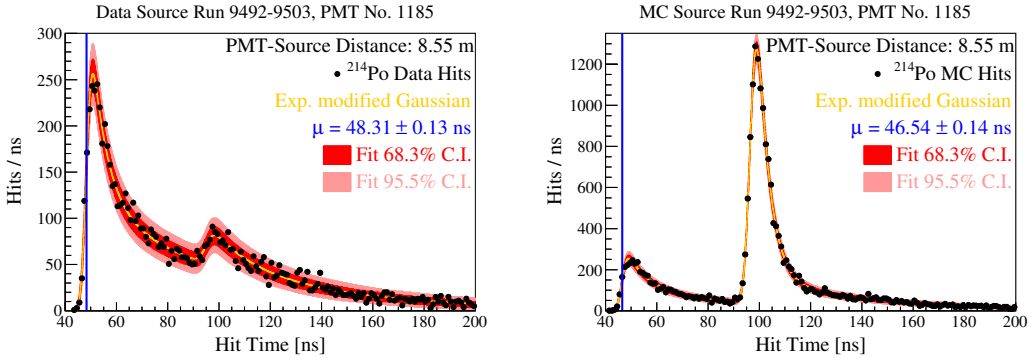


FIGURE 4.5: Exemplary fit for a large PMT-source distance. This is an extreme, but rare example between the possible differences in the qualitative hit time distribution of data and MC. The source is positioned directly on the line between two opposite PMTs with light collection cones, giving rise to a very large reflection peak in MC, likely due to a slightly incorrect geometry of the light collector. The hit time distribution is shown as black dots, the best fit as a yellow line and the Poissonian confidence intervals as red bands. The best fit for the fit parameter μ, describing the average arrival time of the earliest detectable photon hits, is shown as a blue line. (a) ^{214}Po data fit. (b) ^{214}Po MC fit.

This is different in Figure 4.5, where a particular large PMT-source distance of 8.55 m is plotted for PMT No. 1185. Again, the data is plotted on the left and the MC is plotted on the right. For the data the reflection peak is clearly visible and the fit function is well able to reproduce the full hit time distribution. This PMT and source position is selected to showcase the biggest observed difference between the data and the MC in this analysis. The source here is situated in such a way, that it lies directly on the connecting line between two PMTs, where one PMT is close by and PMT No. 1185 is far away, and both PMTs have a light collecting cone. For the MC the reflection peak is about $\sim$ 5 times larger than the primary peak. This explicit investigation of the individual PMTs brought forth a systematic difference between the MC and the data, where the light collecting cones of MC likely have an incorrect geometry. Under rare circumstances the MC hit time distribution is grossly incorrect, but for most PMTs

the hit time distributions are in qualitative agreement and such extreme differences no not occur often. Such a behaviour is not visible in a hit time distribution that is averaged over all PMTs, as has been done in the Borexino calibration [85], and as such its impact is negligible for the spectral analysis. Likewise this is not of concern for the CID method, where the focus is on early PMT hits to improve the Cherenkov to scintillation hit ratio, but it shows the robustness of the fit function 4.2.

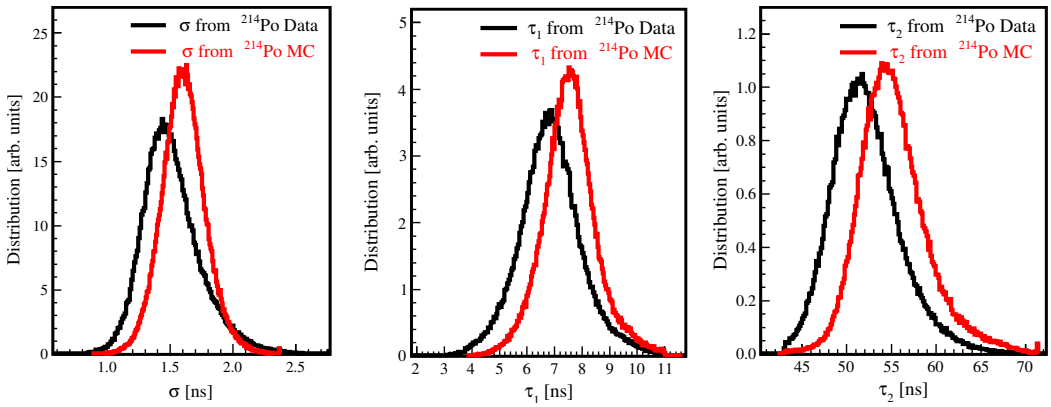


FIGURE 4.6: The distribution of fit parameters for data (black) and MC (red). (left) Detector response smearing $\langle\sigma_{\text{data}}\rangle = 1.47\,\text{ns}$, $\langle\sigma_{\text{MC}}\rangle = 1.60\,\text{ns}$. (middle) Fast component of the effective exponential decay $\langle\tau_{1,\,\text{data}}\rangle = 6.8\,\text{ns}$, $\langle\tau_{1,\,\text{MC}}\rangle = 7.5\,\text{ns}$. (right) Slow component of the effective exponential decay $\langle\tau_{1,\,\text{data}}\rangle = 51.5\,\text{ns}$, $\langle\tau_{1,\,\text{MC}}\rangle = 54.6\,\text{ns}$.

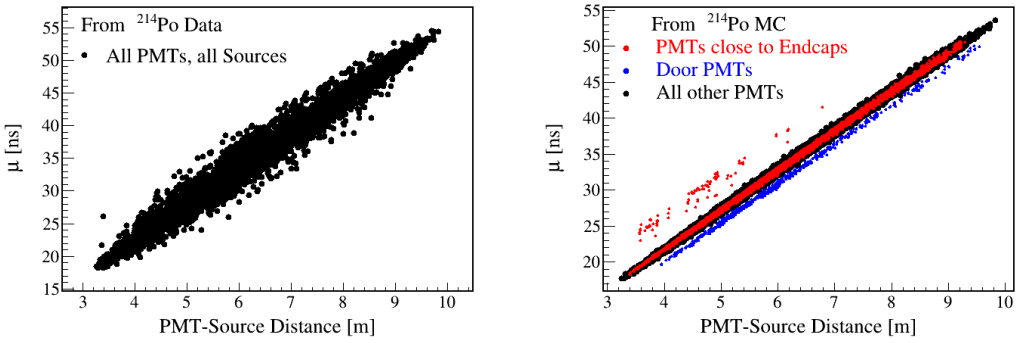
The last Figure 4.6 of this section shows the distribution of the best fit values of the σ, τ_1, τ_2 parameters. It can be seen that all parameter distributions are significantly different between the data and the MC. Here only the difference of $\langle\sigma\rangle_{\mathbf{MC}} - \langle\sigma\rangle_{\mathbf{data}} = 0.13\,\text{ns}$ is of some interest, because σ is an effective description of the detector response smearing for the scintillation. If the underlying time distribution of scintillation hits is different between the data and the MC, the Cherenkov to scintillation hit ratio will also be different and is therefore in need of correction. This is done in Chapter 5. The differences of τ_1, τ_2 are not important for any analysis of this thesis because they describe the time behaviour of α-particle scintillation, which is different from the neutrino recoil electron time behaviour. Nonetheless, these plots show that the fitting of the individual PMT hit time distributions could be potentially useful for future large scale LS detectors, if there is a need for a very precise tuning of the MC PMT hit time behaviour. This could be the case for hybrid detectors, which aim for event-by-event direction reconstruction.

4.4 Polynomial fit of the scintillation effective refractive index

The next step for the measurement of the effective refractive index $n_{\text{g}}^{\text{eff}}$ is a polynomial fit of the distance dependent arrival time of the earliest emitted photons $\mu(d)$. The effective scintillation group velocity $v_{\text{g}}^{\text{eff}}$ is described by $n_{\text{g}}^{\text{eff}}$ through Equation 4.1.

$$\mu(d) = \frac{n_{\text{g}}^{\text{eff}}(d)}{c} \cdot d \left(= \frac{d}{v_{\text{g}}^{\text{eff}}(d)} \right) \tag{4.4}$$

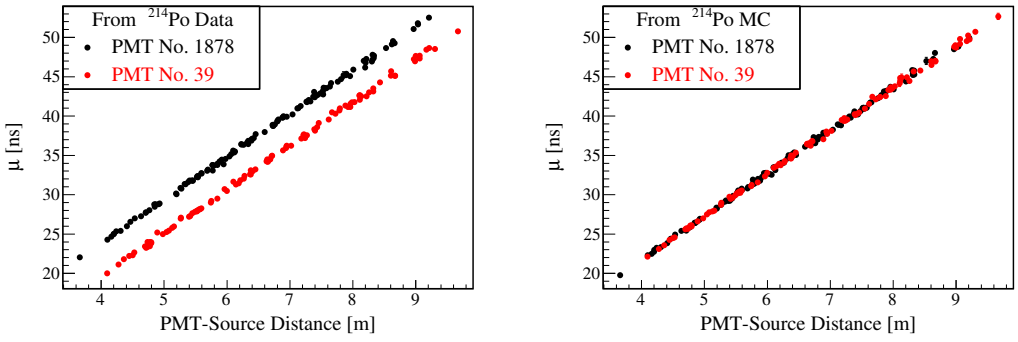
Equation 4.4 describes the measured ToF $\mu(d)$ of the earliest detectable hits as a function of the effective refractive index $n_{\text{g}}^{\text{eff}}$. This dependence is visible in Figure 4.7, where the best fit values of $\mu(d)$ are plotted against the known distance between the hit PMT and the source position d. Here all PMTs are plotted for all source positions, on the left for the data and on the right for the MC. First, it can be seen that both the data and the MC roughly follow a linear



(a) (b)

FIGURE 4.7: Distribution of best fit $\mu(d)$ vs. d for all PMTs for the ^{214}Po event selection. (a) For data. (b) For MC, with the PMTs close to the vessel end-caps ($|z_{\text{PMT}}| > 6\,\text{m}$) shown in red, the PMTs in the main SSS door shown in blue and all other PMTs in black.

trend, as is expected from the Tof of the photons. The data has a much broader distribution of $\mu(d)$ compared to the MC, while the MC has some discernible features. In the MC the bulk of the PMTs is shown in black, while PMTs close to the vessel end-caps are shown in red, where some source positions produce visible outliers that have too large values of $\mu(d)$, compared to most other source positions. The 19 PMTs which are positioned inside the stainless steel sphere (SSS) door of Borexino all have a systematically decreased value of $\mu(d)$ and are shown in blue. The reason for this has been identified as a mistake in G4BX2, where these door PMTs are simulated at a position with a radius 6.5 m, while in reality they have a radial position of 6.9 m. The indentation and geometry of the door in the stainless steel sphere is correctly implemented, but the "door PMTs" are incorrectly simulated with the same radial position as all other PMTs. This mistake has not been identified before this work, but it likely is negligible. Thus, here the blue "door PMTs" in Figure 4.7(b) have a fit value for $\mu(d - 0.4\,\text{m})$ at an assumed distance of only d, which can easily be corrected.

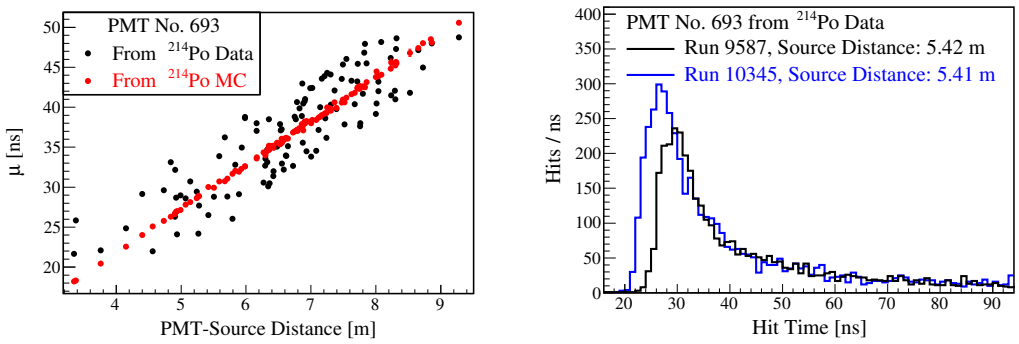


(a) (b)

FIGURE 4.8: The distribution of $\mu(d)$ vs. d for the PMT No. 1878 in black and PMT No. 39 in red. (a) For data these PMTs have a significant constant time offset between each other of ∼4.5 ns. These two PMTs are selected to illustrate this offset behaviour and most other PMTs have a smaller but still significant offset between each other. (b) For MC the same PMTs are well in agreement and no such constant offset is visible.

The next Figure 4.8 illustrates the first reason for the relative broad $\mu(d)$ distribution of the data compared to the MC. As an example, the PMT No. 39 is shown in red and the PMT

No. 1878 is shown in black, on the left for the data and on the right for the MC. It can be seen that the individual PMTs have qualitatively the same linear trend in both the data and the MC. But in the data the two PMTs have a constant ∼4.5 ns offset between each other. For MC this is not the case and both PMT $\mu(d)$ vs. d distributions are well aligned. The origin of this constant offset is not known, but it can be seen here that it is stable in time for hundreds of different source runs. Most PMTs in the data have such a constant offset between each other, which is also often larger than the fit uncertainty $\Delta\mu(d)$. This is the main reason for the relative broadness of $\mu(d)$ in the data, while in the MC this offset does not exit.

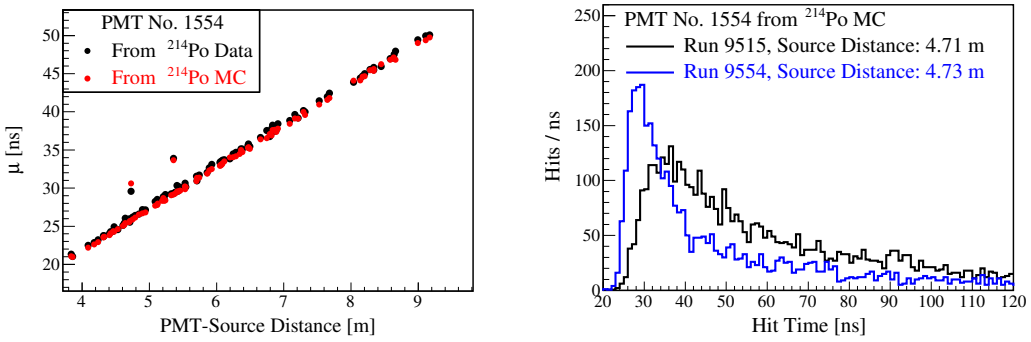


(a) (b)

FIGURE 4.9: The distribution of $\mu(d)$ vs. d for the PMT No. 693. (a) The data (black) shows a large spread of $\mu(d)$ for different calibration runs, even when the source is positioned at similar distances. For the MC this effect does not exits. The PMT is selected to illustrate this behaviour and in most other data PMTs this effect is much smaller, but non-negligible. (b) The measured hit time distributions (without ToF correction) of the same PMT No. 693 for different calibration runs with similar distances. Run 9575 (black) has 5637 hits, while Run 10345 (blue) has 4624 hits. The shapes of the hit time distributions are similar to each other, but they have a clear constant offset between them.

The second reason for the relative broad data $\mu(d)$ distribution is exemplary illustrated in Figure 4.9(a). On the left the best fit $\mu(d)$ values are plotted against d for PMT No. 693, which is a normal PMT with a light collecting cone, that is neither in the door nor close to the end-caps of Borexino. For data (black) the points show a large spread for different source runs, even for comparable distances d. This spread is significantly larger then the corresponding fit uncertainty $\Delta\mu(d)$ and this behaviour is present in most data PMTs, albeit to a much smaller extent. For MC (red) this effect does not exit and the spread of $\mu(d)$ is in agreement with $\Delta\mu(d)$. The origin of this behaviour is unknown and Figure 4.9(b) on the right shows the data hit time distribution for PMT No. 693, for a $d = 5.42\,\text{m}$ in black and for a different source run with $d = 5.41\,\text{m}$ in blue. These distances are the same within the uncertainty of the CCD camera position reconstruction, but the hit time distributions have a clear offset from each other. It is also important to note, that the shape of these distributions, estimated by the best fit parameters, is quantitatively the same except for the amplitude A and an offset in $\mu(d)$. The conclusion is then that the data PMTs have a hit time misbehaviour, where different runs have a different constant time offset between each other, but this time offset is stable within a single run. This effect exits for most PMTs, where it is relatively small, but still larger than the fit uncertainty $\Delta\mu(d)$ for some runs. For some PMTs this effect is very large, with PMT No. 693 being the worst performing PMT. The origin of this misbehaviour could be some unknown error in the PMT time alignment with the laser system [97, 98], but it is unclear exactly what the problem could be.

The last Figure 4.10 concerns the $\mu(d)$ outliers that exist in both the data and the MC, but only for PMTs close to the vessel end-caps. These PMTs typically have an absolute z



(a) (b)

FIGURE 4.10: The distribution of $\mu(d)$ vs. d for PMT No. 1554, which has a position of $x = 0\,\text{m}$, $y = 1.8\,\text{m}$, $z = -6.1\,\text{m}$, close to the vessel end-cap. (a) Both the data (black) and the MC (red) show two $\mu(d)$ outliers for the same source positions. The positions correspond to distances, for which the direct photon path is obstructed by the end-cap structure. (b) The MC PMT hit time distribution of PMT No. 1554 for source runs of similar distance. The calibration Run 9515 (black) with 5697 hits as an obstructed photon path, while for Run 9554 (blue) with 4276 hits it is unobstructed. The distributions have a significantly different shape that cannot be explained by a simple constant hit time offset.

position of $|\,z\,| > 6\,\text{m}$ $(\cos\theta > 0.9)$ and they number about 108 PMTs. Figure 4.10(a) on the left shows data in black and MC in red for the same PMT No. 1554. The distribution of $\mu(d)$ is similar between data and MC, where both exhibit two outliers for the same source positions. The reason for these outlying $\mu(d)$ values is shown in Figure 4.10(b), where the hit time distribution of the MC PMT No. 1554 is plotted for two runs with similar distances of $d \approx 4.7\,\text{m}$. The source position of Run 9515 (black) is such that the direct path between the source and the PMT intersects the vessel end-cap structure, while the direct path for the source position of Run 9554 (blue) is unobstructed. The hit time distribution of the unobstructed path is qualitatively similar to most other hit time distributions, while the hit time distribution of the obstructed path is much broader with a peak position shifted to later hit times.

This likely means that the end-caps have a shadowing effect, which only allows photon paths to be detected when they are scattered around the end-cap structure. The conclusion is then that these outliers can safely be excluded for the calculation of the effective refractive index n_g^{eff}. They are not the sign of an instability in the $\mu(d)$ fit, but they are rather a geometric hit selection bias that is correctly reproduced in the MC.

To summarise the qualitative differences and similarities between the data and the MC best fit $\mu(d)$ distributions:

- $\mu(d)$ has a low order polynomial dependence on d for both the data and the MC.
- For the data the hit time distributions are offset by a constant time between the PMTs.
- For the data the PMTs have a different constant time offset of the hit time distribution for different runs. For some PMTs this results in a large spread of $\mu(d)$.
- The data and the MC have some outliers in $\mu(d)$ for PMTs close to the end-caps due to the shadowing effect of the vessel end-cap structure for some source positions.

Thus, the effective refractive index n_g^{eff} of the earliest detectable scintillation hits can be estimated with a polynomial χ^2 fit of the $\mu(d)$ vs. d distributions:

$$n_g^{\text{eff}}(d) = p_1 + p_2 \cdot d \tag{4.5}$$

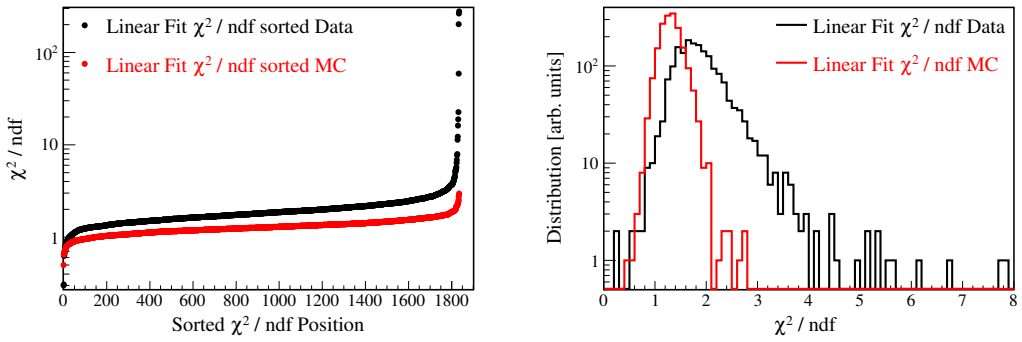
$$\chi_i^2(p_1,p_2,C_i) = \sum_j \left(\frac{\mu(d_j) - C_i - n_g^{\text{eff}} \cdot d_j/c}{S_i \cdot \Delta\mu(d_j)} \right)^2$$

$$\chi^2(p_1,p_2,C_1,...,C_{1827}) = \sum_{\text{PMT } i}^{1827} \chi_i^2(p_1,p_2,C_i) \tag{4.6}$$

$$\chi^2(p_1,p_2) = \underset{C_1,...,C_{1827}}{\arg\min} \left[\chi^2(p_1,p_2,C_1,...,C_{1827})\right]$$

Due to the wavelength selection bias of early PMT hits there could be a complex distance dependence on their effective refractive index. For the purpose of the ToF correction it is enough to perform a phenomenological Taylor approximation that is well able to describe the measured $\mu(d)$ vs. d distribution, given the statistics. Here $n_g^{\text{eff}}(d)$ is estimated as a constant as well as a linear function of d and it is found that the constant approximation is insufficient, while the linear function is fully sufficient to explain the $\mu(d)$ distribution. Please note that the terms *constant* and *linear* are used here for the number of parameters in the $n_g^{\text{eff}}(d)$ function. It is not a description of the $\mu(d)$ vs. d distribution, which additionally includes C_i. These parameters $C_1,...,C_{1827}$ are the constant time offset values for each individual PMT to accommodate the discovered systematic offsets between the data PMTs. The factors $S_1,...,S_{1827}$ are used to increase the fit uncertainty $\Delta\mu(d)$ to include the systematic spread of $\mu(d)$ for different runs, as seen in Figure 4.9. These two consideration of the systematic differences between the individual PMT allows the correct estimation of the effective refractive index for the scintillation. Additionally, the source positions that produce the $\mu(d)$ outliers of the PMTs close to the vessel edge are not included in the fit. The MC is fitted in the same way and with the same PMT and source position selection as the data.

4.5 Fit results

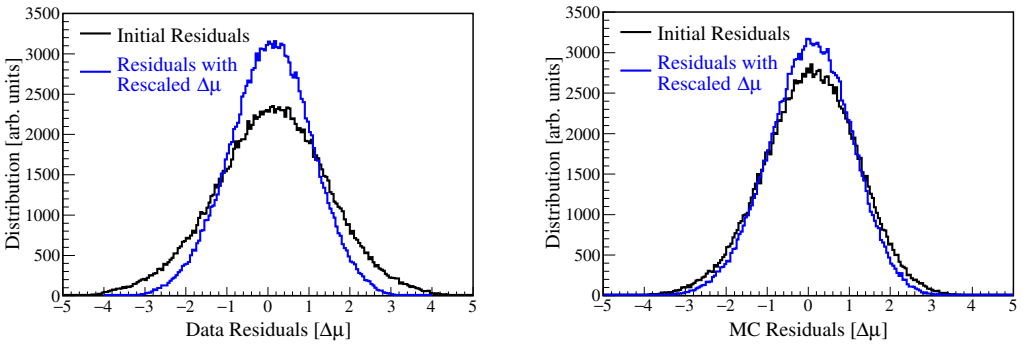


(a) (b)

FIGURE 4.11: (a) Ascending values of the individual χ_i^2/ndf_i values for the 1827 fitted PMT $\mu(d)$ vs. d distributions. Here ndf is the number of source positions used for each individual PMT minus one. The data (black) has consistently larger values than MC (red) and some outliers of up to $\chi_i^2/\text{ndf}_i = 278.2$ (b) The distribution of the same χ_i^2/ndf_i values for the data (black) and the MC (red).

The first step to perform the fit is to deselect those PMTs, which have the worst systematic $\mu(d)$ spread from run to run, such as PMT No. 693 in Figure 4.9. In total there have been 1827 live PMTs available during the calibration campaign, but here 10 PMTs are deselected to avoid a possible systematic shift of the best fit $n_g^{\text{eff}}(d)$ parameters. The deselection process is illustrated in Figure 4.11(a), where the χ_i^2/ndf_i value of each individual PMT is plotted, in

ascending order. For the data (black) the χ_i^2/ndf_i values are always larger than the χ_i^2/ndf_i values of the MC (red) and the largest data χ_i^2/ndf_i values show the impact of the run-by-run spread of $\mu(d)$. The 10 worst performing data PMTs are deselected and their $\mu(d)$ vs. d plots can be found in the Appendix A.1. Figure 4.11(b) shows the χ_i^2/ndf_i distribution of the individual, selected PMTs for the data in black and the MC in red. It can be seen that the data χ_i^2/ndf_i is much broader distributed than the MC and with a skew. This is the effect of the systematically different run-by-run hit time offset of the individual data PMTs. The fit uncertainty $\Delta\mu(d)$ itself is not necessarily perfect, as it results in a mean value of $\langle\chi_i^2/\mathrm{ndf}_i\rangle = 1.3$ for MC, while $\Delta\mu(d)$ is too small for the data with a mean value of $\langle\chi_i^2/\mathrm{ndf}_i\rangle = 1.9$.

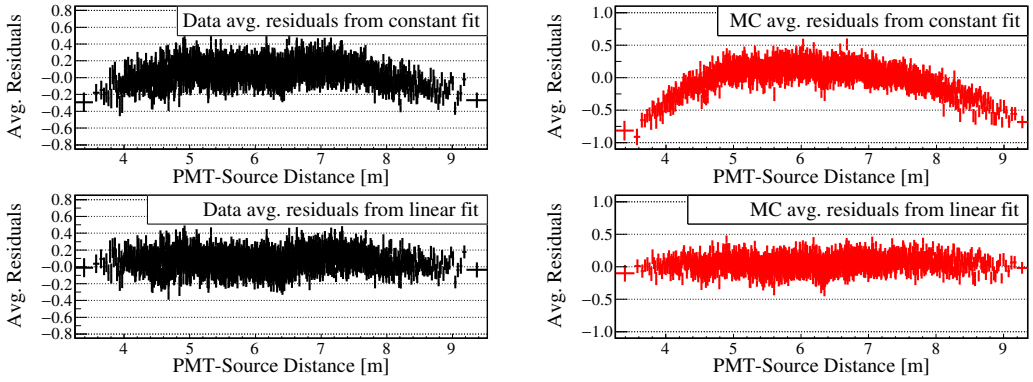


(a) (b)

FIGURE 4.12: Distribution of residuals before (black) and after scaling (blue) of the uncertainties $\Delta\mu(d)$ in the denominator. (a) For the data the standard deviation changes from $1.4\Delta\mu(d)$ to $1.0\Delta\mu(d)$. This value is large without the scaling because of the systematic run-by-run spread of $\mu(d)$, which is needs to be accounted for in the uncertainty. (b) For the MC the change is less severe from $1.1\Delta\mu(d)$ to $1.0\Delta\mu(d)$, as there no such systematic misbehaviour exists.

The run-by-run spread of $\mu(d)$ needs to be taken into account, which is done here through the scaling factor S_i, individually for each PMT. Every PMT has a corresponding number of fitted $\mu(d)$ values and a resulting χ_i^2 value. The value of S_i is selected for each fitting value, during the minimisation process, such that $\chi_i^2/\mathrm{ndf}_i = 1$ for every individual PMT. This rescaling increases the fit uncertainty, to take into account the systematic uncertainty from the change of the hit time offset for each run of a given PMT. The effects of the rescaling are shown in Figure 4.12 for the data on the left and the MC on the right. The distribution of all residuals $(\mu(d_j) - C_i - n_\mathrm{g}^\mathrm{eff}(d_j)\cdot d_j/c)/\Delta\mu(d_j)$ is shown in black, while the distribution of the rescaled residuals $(\mu(d_j) - C_i - n_\mathrm{g}^\mathrm{eff}(d_j)\cdot d_j/c)/(S_i\cdot\Delta\mu(d_j))$ is shown in blue. Without the rescaling the residuals have a standard deviation of 1.4 for the data and 1.1 for the MC and with the rescaling these values are by definition 1.0. Using this method results in a conservative estimation of the uncertainty $\Delta n_\mathrm{g}^\mathrm{eff}$. The χ_i^2/ndf_i values and residual standard deviations of the PMTs can be > 1.0 due to simple statistical fluctuations, but here the rescaling assumes that these values are > 1.0 only due to the unaccounted systematic spread of $\mu(d)$ for different runs. This means that the values of $S_i > 1.0$ are underscaling the summed χ^2 value of all PMTs, as it is forced to take a value that is smaller than expected from the given statistics. This results in an overestimation of the uncertainty $\Delta n_\mathrm{g}^\mathrm{eff}$, as it is given by the $\Delta\chi^2 = \pm 1$, corresponding to the 68.3% (1σ) CI.

The next step is to decide the degree of the polynomial $n_\mathrm{g}^\mathrm{eff}(d)$ function. Due to the introduction of the scaling factor S_i it is not helpful to look at the absolute values of χ^2 for different $n_\mathrm{g}^\mathrm{eff}$ models, as they are set to the same χ^2 value by definition of the S_i factors. A hypothesis test for the number of polynomial parameters is not possible and instead it



(a) (b)

FIGURE 4.13: The distance dependence of the residuals, including the scaling of the uncertainty $S_i \cdot \Delta\mu(d)$. Here the residual values are averaged in bins of ten residuals, to make an incorrect fit model more visible. For data on the left (a) and for MC on the right (b). On the top, the constant fit of $n_g^{\text{eff}} = p_1$ shows clearly a non-flat distance dependence of the residuals. The linear fit $n_g^{\text{eff}}(d) = p_1 + p_2 \cdot d$ at the bottom has a much flatter distribution and no clear distance dependence is visible. Please note that the residual values are relatively small due to the averaging and to a smaller extent due to the application of the uncertainty up-scaling with S_i.

is necessary to look at the actual distance distribution of the residuals. This is shown in Figure 4.13 for the data on the left and the MC on the right. The ten residuals that are closest to each other are averaged to make deviations from zero more visible. On the top, the averaged residual distribution is given for the constant fit of $n_g^{\text{eff}} = p_1$, while the bottom shows the averaged residual distribution for a linear fit of $n_g^{\text{eff}}(d) = p_1 + p_2 \cdot d$. It can be seen for both the data and the MC that a constant n_g^{eff} function is not well able to describe the measured $\mu(d)$ vs. d distribution. The average residual distributions have the shape of a negative parabola, which is much more pronounced in the MC. The residual values are significantly smaller than 1.0 due to the averaging of ten residuals and the error bars correspond to the uncertainty of the average. Using the linear $n_g^{\text{eff}}(d)$ function shows a much flatter residuals distribution. This does not necessarily mean that the true, underlying $n_g^{\text{eff}}(d)$ function is fully described by a linear function. Here $n_g^{\text{eff}}(d)$ only needs to approximate the true behaviour well enough, such that the systematic differences in the ToF correction between the true group velocity and its linear approximation are negligible, compared to the other uncertainties on the PMT hit time distribution of the earliest detectable hits.

Figure 4.14 shows the $\Delta\chi^2(p_1, p_2)$ distribution in units of the Gaussian equivalent σ ($1\sigma = 68.3\%$, $2\sigma = 95.5\%$, $1\sigma = 99.7\%$). The data is shown on the left and the MC is shown on the right. The profile is produced through a scan of the two dimensional parameter space, where for each selection of (p_1, p_2) the PMT constants (C_1, ..., C_{1827}) are treated as nuisance parameters, see Equation 4.6. Thus, the absolute χ^2 value is calculated with C_i values such that χ^2 is minimal for the selected values of (p_1, p_2). The two dimensional confidence intervals are then calculated from the $\Delta\chi^2$ profile with two degrees of freedom. It can be seen that the parameters p_1, p_2 are anti correlated in the fit and the quoted uncertainties Δp_1 and Δp_2 take this into account. The best fit results of the effective refractive index $n_g^{\text{eff}} = p_1 + p_2 \cdot d$ are:

- For the data: $p_1 = 1.6867 \pm 0.0023$, $p_2 = (-0.296 \pm 0.018)\,10^{-2}\text{m}^{-1}$
- For the MC: $p_1 = 1.7141 \pm 0.0015$, $p_2 = (-0.575 \pm 0.012)\,10^{-2}\text{m}^{-1}$

The results of the data and the MC are significantly different between each other, with relatively small statistical uncertainties of $\Delta n_g^{\text{eff}} / n_g^{\text{eff}} < 0.2\%$ for distances up to 9 m. The

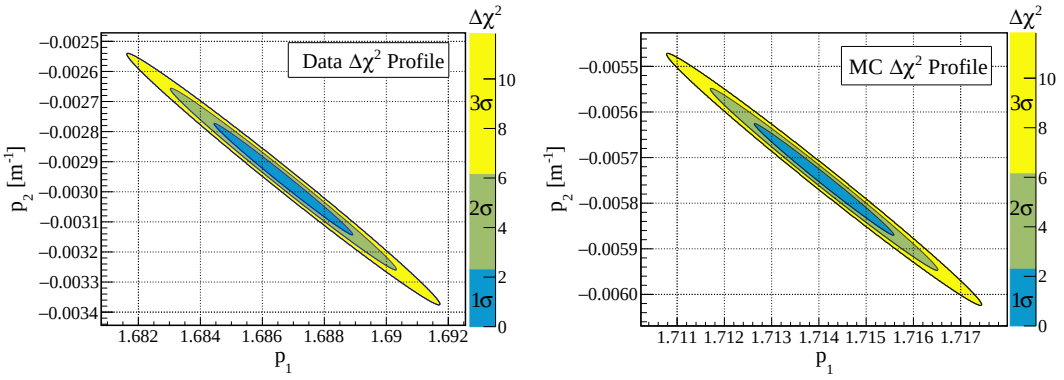


FIGURE 4.14: The best fit results for the linear model of the effective refractive index $n_g^{\text{eff}}(d)$. They are shown as a two-dimensional $\Delta\chi^2$ profile, in units of the two dimensional Gaussian standard deviation σ. The parameters p_1 and p_2 are strongly anti-correlated in the fit. (a) For the data the fit values are $p_1 = 1.6867$, $p_2 = -0.00296\,\text{m}^{-1}$. (b) For the MC the fit values are $p_1 = 1.7141$, $p_2 = -0.00575\,\text{m}^{-1}$.

quoted uncertainty of the data is bigger than that of the MC due to the systematic run-by-run spread of $\mu(d)$, accommodated through S_i. The negative slope p_2 results in a lower effective refractive index and consequently in a faster effective group velocity for larger distances. This behaviour is qualitatively in agreement with the expectations from Figure 4.3, where the wavelength distribution of photons with a larger distance d is selected from longer wavelengths due scattering. Longer wavelengths correspond to a larger group velocity (smaller refractive index). The constant model of $n_g^{\text{eff}} = p_1$ results in values of $n_g^{\text{eff}} = 1.6631 \pm 0.0002$ for data and $n_g^{\text{eff}} = 1.6531 \pm 0.0001$ for MC.

4.5.1 Final results with systematic uncertainties

It follows the estimation of the systematic uncertainties for the $n_g^{\text{eff}}(d)$ analysis. First is the effect of the scaling factors S_i: The PMTs in MC do not (and can not) exhibit the systematic run-by-run spread of $\mu(d)$. Thus, the incorrect application of S_i on MC can be used to estimate the potential systematic uncertainty from the use of the scaling factors. The uncertainty values are estimated from the absolute difference of the MC best fit results with and without S_i, resulting in systematic uncertainties of $\Delta p_1(\text{scaling}) = \pm 0.0008$ and $\Delta p_2(\text{scaling}) = \pm 0.007 \times 10^{-2}\,\text{m}^{-1}$. The application of S_i on the data is a conservative method for the inclusion of the systematic $\mu(d)$ spread in the total uncertainty budget of $\Delta\mu(d)$. The systematic uncertainty from S_i in the data is then estimated to be the same as for the MC.

Next is the selection of the PMT hit time binning. The nominal bin width used for the analyses is 1 ns, where the hit time histograms are fitted for $\mu(d)$, only when the peak of the distribution has more than 30 entries. Applying a bin width of 0.5 ns, with a requirement of the histogram peak to have at least 15 entries. This results in a slightly different selection of PMTs and source positions. Additionally, there could be small scale structures in the PMT hit time distribution that could influence the $\mu(d)$ fitting procedure. Thus, the smaller bin width selection can be used to estimate the potential systematic uncertainty from these two sources. For the data this results in an absolute difference of the best fit values of $\Delta p_1(\text{binning}) = \pm 0.0001$ and $\Delta p_2(\text{binning}) = \pm 0.003 \times 10^{-2}\,\text{m}^{-1}$.

The last point is the systematic uncertainty from the energy selection of the events. For the measurement of n_g^{eff} it is important to be in the single photo-electron regime for all distances, as otherwise there could be a potential distance dependence on the best fit $\mu(d)$, coming from the impact of multiple photon hits on a single PMTs. The best fit value of $\mu(d)$ would

shift to earlier times for multiple photon hits because only the time of the first detected photo-electron is registered. For the selected ^{214}Po events with less than ~ 400 hits over ~ 1800 PMTs this is likely already the case, but the selected ^{222}Rn+^{218}Po events have about 1.7 times less hits on average. At the same time there are more ^{222}Rn+^{218}Po events, such that the total PMT hit number is larger than for the ^{214}Po events. This results in a slightly different selection of PMTs which pass the minimal statistics condition at different source positions. The ^{222}Rn+^{218}Po events are only analysed in the data with the best fit results of $p_1(^{222}\mathrm{Rn}+^{218}\mathrm{Po}) = 1.6859 \pm 0.0014$, $p_2(^{222}\mathrm{Rn}+^{218}\mathrm{Po}) = (-0.292 \pm 0.011)\,10^{-2}\mathrm{m}^{-1}$. Again the absolute differences between the results with the ^{214}Po events and the ^{222}Rn+^{218}Po events are used as an estimation of the potential energy depended systematic uncertainty. They follow as $\Delta p_1(\text{energy}) = \pm 0.0008$ and $\Delta p_2(\text{energy}) = \pm 0.004 \times 10^{-2}\,\mathrm{m}^{-1}$

All these potential systematic uncertainties are estimated maximally conservative, where the absolute difference between the different $n_g^{\text{eff}}(d)$ results is used. It is possible that the analysis results are different from each other, mostly due to simple statistical fluctuations, but at the same time this method of estimation is very simple. The final values of the $n_g^{\text{eff}}(d) = p_1 + p2 \cdot d$ measurement are given in the following, under the assumption that the systematic uncertainties are uncorrelated:

- For the data: $p_1 = 1.6867 \pm 0.0023\,(\text{stat.}) \pm 0.0012\,(\text{syst.})$
  $p_2 = (-0.296 \pm 0.018\,(\text{stat.}) \pm 0.009\,(\text{syst.}))\,10^{-2}\,\mathrm{m}^{-1}$
- For the MC: $p_1 = 1.7141 \pm 0.0015\,(\text{stat.}) \pm 0.0012\,(\text{syst.})$
  $p_2 = (-0.575 \pm 0.012\,(\text{stat.}) \pm 0.009\,(\text{syst.}))\,10^{-2}\,\mathrm{m}^{-1}$

4.5.2 Validation of the MC result

The question now is how well this analysis procedure is able to measure the true, underlying value of interest, which is the effective group velocity of the earliest detectable PMT hits of the scintillation. This velocity is necessary to perform a ToF correction on the PMT hits, such that the Nth-Hit time like variable can be used to improve the Cherenkov to scintillation ratio for the CID analysis.

What are these "earliest detectable PMT hits"? They are those hits that are emitted almost instantly by the LS, do not undergo absorption and re-emission and are only minimally scattered in the detector. Using the G4BX2 simulation it is possible to produce the hit time distribution for exactly these scintillation hits of interest. For this purpose the re-emission probability is set to zero in G4BX2 and likewise the scintillation emission decay times are also set to zero, i.e. an instantaneous photon emission. Additionally, the production of Cherenkov photons is turned off. The absorption and scattering inside the LS is kept on because those are necessary to select the same, distance dependent photon wavelength distribution as for the full MC $n_g^{\text{eff}}(d)$ analysis.

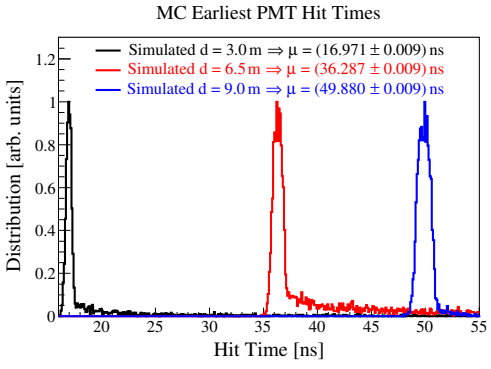


FIGURE 4.15: Validation of the n_g^{eff} measurement for the MC. Here the simulation is modified, such that the scintillation photons are emitted instantaneous, and without re-emission after absorption in the LS. This results in the MC true hit time distribution, without the PMT transit time spread and event reconstruction uncertainties. The peaks of the hit time distributions for different distances correspond to the Tof of the earliest detectable photons.

Figure 4.15 shows the distribution of the MC true PMT hit times of the scintillation photons, for different distances between the event position and the hit PMT. First, the electron events are simulated at a position of $x = 3.5\,\mathrm{m}$ and the four closest PMTs are selected to produce the black distribution, corresponding to $d = 3.0\,\mathrm{m}$. The red distribution is for electron events in the centre and all PMTs with light collecting cones are selected, except the door PMTs, for a distance of $d = 6.5\,\mathrm{m}$. The blue distribution is for electrons simulated at $x = 2.5\,\mathrm{m}$, but now the opposite four PMTs are selected, which are furthest away from the event position with a distance of $d = 9.0\,\mathrm{m}$. All distributions have a distinct peak with a tail at later times. This peak then corresponds to those scintillation hits, that are emitted instantly in the scintillation process and do not undergo absorption and re-emission. The spread of the peaks comes from the wavelength dependence of the group velocity for the different photons, given the scintillation spectrum. The tails of the hit time distributions correspond to photons that are affected by Rayleigh scattering, multiple scattering or by reflection on some surface.

The peaks are fitted with a Gaussian, while safely excluding the tail region. This results in the expected ToF values for the "earliest detectable PMT hits" as $\mu_{\mathrm{true}}(3.0\,\mathrm{m}) = (16.971 \pm 0.009)\,\mathrm{ns}$, $\mu_{\mathrm{true}}(6.5\,\mathrm{m}) = (36.335 \pm 0.009)\,\mathrm{ns}$, $\mu_{\mathrm{true}}(9.0\,\mathrm{m}) = (49.880 \pm 0.009)\,\mathrm{ns}$. The expected ToF values μ_{fit} are calculated according to Equation 4.4 with $n_{\mathrm{g}}^{\mathrm{eff}}$.

	$d = 3.0\,m$	$d = 6.5\,m$	$d = 9.0\,m$
μ_{true}	$(16.971 \pm 0.009)\,\mathrm{ns}$	$(36.287 \pm 0.009)\,\mathrm{ns}$	$(49.880 \pm 0.009)\,\mathrm{ns}$
μ_{fit}	$(16.969 \pm 0.020)\,\mathrm{ns}$	$(36.329 \pm 0.047)\,\mathrm{ns}$	$(49.871 \pm 0.071)\,\mathrm{ns}$
$\frac{(\mu_{\mathrm{true}} - \mu_{\mathrm{fit}})}{\sqrt{\Delta\mu_{\mathrm{true}}^2 + \Delta\mu_{\mathrm{fit}}^2}}$	-0.12	0.88	-0.14

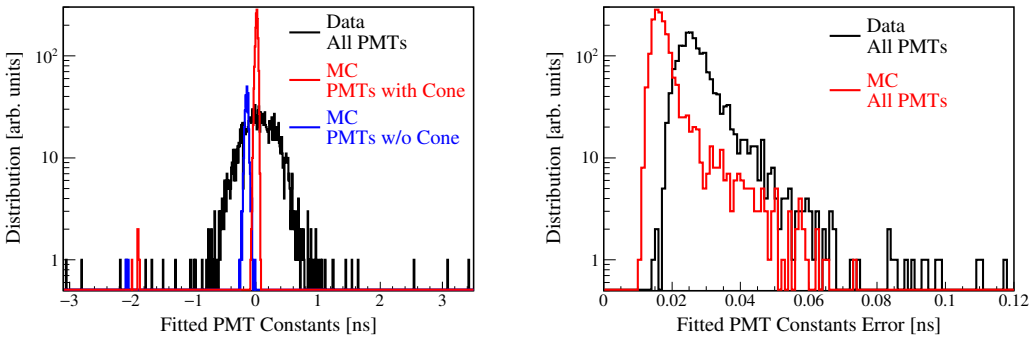
TABLE 4.1: Validation of MC effective refractive index $n_{\mathrm{g}}^{\mathrm{eff}}(d)$.

Table 4.1 compares the ToF from the modified MC simulation μ_{true} with the expected values μ_{fit} according to the $n_{\mathrm{g}}^{\mathrm{eff}}(d)$ measurement. The absolute difference between the ToF MC truth and fit values is smaller than 0.05 ns and the results are well in agreement within their (conservative) uncertainties. The ToF uncertainties $\Delta\mu_{\mathrm{fit}}$ of the fit are negligibly small, compared to the fitted standard deviation of the early hit times with $\sigma = 1.6\,\mathrm{ns}$, which includes the transit time spread of the PMT and the event reconstruction (see Figure 4.6). It can be concluded that the $n_{\mathrm{g}}^{\mathrm{eff}}(d)$ fitting procedure, using the ^{14}C–^{222}Rn calibration source, is well able to describe the effective group velocity of the earliest detectable PMT hits. As this is certainly the case in the MC it is then reasonable to assume that it is also true for the data. The linear $n_{\mathrm{g}}^{\mathrm{eff}}(d)$ model is sufficient for a phenomenological description of the effective group velocity and is thus used in the following CID-analyses of the solar neutrinos.

4.5.3 Effect of the PMT time constants on the position reconstruction.

The analysis of the individual PMT hit times is used mainly for the measurement of the effective refractive index, by fitting the best fit $\mu(d)$ vs. d distributions of the individual PMTs. As previously stated this analysis approach makes it in principle possible to perform a very precise fine tuning of the MC to reproduce the data results. Here an additional use-case of the fitting procedure is explored through the potential improvement of the position reconstruction algorithm. In the Borexino calibration campaign it has been noticed, that the position reconstruction of the data exhibits a position dependent systematic shift in -Z of up to

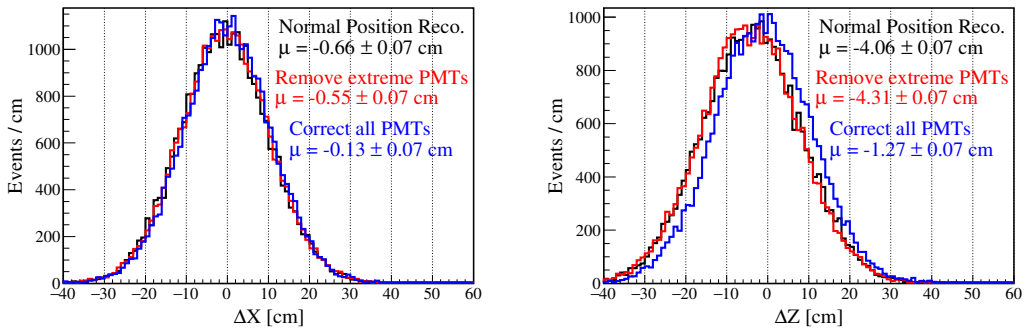
~ 4 cm and this shift is maximal in the centre of the detector, see Figure 22 in [85]. The origin of this systematic shift has never been found out but its impact negligible for the spectral fit.



(a) (b)

FIGURE 4.16: (a) The distribution of the best fit constant hit time offsets C_i of the individual PMTs for the data (black) and the MC (red, blue). The data has a much larger distribution than the MC, while in the MC a substructure is visible for PMTs with (red) and without (blue) the light collecting cones. (b) The corresponding fit uncertainty ΔC_i distribution. The data has larger uncertainties due to the increased uncertainty $\Delta\mu(d)$.

Figure 4.16(a) on the left shows the distribution of the best fit PMT constants C_i, relative to their average value. For the data (black) the bulk of the distribution is concentrated in a Gaussian, with a standard deviation of ~ 0.3 ns, while there are some outliers of up to ± 3 ns. The MC is shown in red for PMTs with a light collecting cone and in blue for PMTs without one. The reflection within the light collecting cones results in a small hit time offset of ~ 0.2 ns relative to the PMTs without cones. Additionally, again the incorrect simulation of the door PMTs is visible around -2 ns. Figure 4.16(b) on the right shows the distribution of the corresponding fit uncertainties of the PMT constants ΔC_i, for the data in black and the MC in red. The data has a higher uncertainty, mostly due to the impact of the systematic run-by-run spread of $\mu(d)$.



(a) (b)

FIGURE 4.17: The effect of the PMT hit time constants on the position reconstruction, for calibration runs at the detector centre. The difference between the true source position and the reconstructed position is shown for the X-axis on the left (a) and for the Z-Axis on the right (b). The normal position reconstruction (black) shows a relatively large systematic offset for the Z-axis. Excluding the most extreme offset PMTs with $|C_i| > 1$ ns (red) does not improve the position reconstruction. Correcting the PMTs with their hit time constant C_i (blue) does improve the position reconstruction significantly, but it is still not in agreement with the true source position.

Because the position reconstruction algorithm is a fit of the detected PMT hit times it is reasonable to assume that the different PMT constants in the data can have a noticeable influence. This is shown in Figure 4.17, where the ^{14}C–^{222}Rn source is deployed in the detector centre. The difference between the source position and the reconstructed event positions is shown on the left for the X-axis and on the right for the Z-axis. The nominal position reconstruction is shown in black, while the position reconstruction that excludes the PMTs with the largest absolute constant values $|\,C_i\,| > 1\,\text{ns}$ is shown in red. Just excluding the most extreme offset PMTs does not improve the systematic mis-reconstruction, the ΔZ offset is even increased in comparison to the nominal position reconstruction. The blue distribution shows the result of the position reconstruction, where all PMTs are corrected with their hit time constants C_i. This results in a significant improvement compared to the nominal position reconstruction. The systematic mis-reconstruction is reduced from $\Delta Z = (-4.31 \pm 0.07)\,\text{cm}$ to $\Delta Z = (-1.27 \pm 0.07)\,\text{cm}$. This is still significantly different from zero but it is nonetheless a relatively large improvement in absolute terms. Thus, the systematic offset in the data position reconstruction is mostly due to the combined effect of all the PMTs that have different constant hit time offsets between them and not only due to a few PMTs with a large offset. Here the effect of the systematic run-by-run spread of the PMT hit time offsets is likely the origin of the remaining systematic position mis-reconstruction, together with the finite precision of the measured hit time constants C_i used for the correction.

These PMT hit time constants can have an impact on the CID analysis and should be taken into account for the calculation of the ToF corrected hit times. Otherwise some PMTs can be over-selected or under-selected for the first Nth-Hits $\cos\alpha$ histograms. This could introduce a bias in the $\cos\alpha$ distribution that is not present in the corresponding $\cos\alpha$ distribution of MC. In contrast, the impact of the PMT constants on the systematic event position mis-reconstruction for the ToF correction $n_\text{g}^\text{eff} \cdot d/c$ is likely negligible, because it is small compared to the position resolution with $\sigma > 10\,\text{cm}$.

4.6 Conclusion and outlook

To perform a CID analysis it is necessary to improve the Cherenkov to scintillation ratio through time sorting, as early PMT hits are more likely to have been produced in the Cherenkov process than hits detected at later times. As such, the hit time sorting must be applied on the reconstructed emission time of the corresponding photon. This reconstruction is performed trough the time-of-flight correction of the distance from the reconstructed event position to the position of the hit PMT. In this chapter the effective group velocity of scintillation light v_g^eff, necessary to calculate the ToF correction, is expressed trough the effective refractive index $v_\text{g}^\text{eff} = c/n_\text{g}^\text{eff}$.

It is possible to construct distance dependent hit time distributions for the individual PMTs, using the ^{14}C–^{222}Rn (α-particle) calibration source at different positions inside the detector. Fitting these distributions with a bi-exponentially modified Gaussian distribution results in a distribution of the earliest detectable hit times $\mu(d)$ vs. d, for each individual PMT. It is found that the $\mu(d)$ vs. d distribution of all PMTs is well described by a linear function of $n_\text{g}^\text{eff}(d) = p_1 + p_2 \cdot d$. The data PMTs exhibit a constant hit time offset between each other, with a standard deviation of $\sim 0.3\,\text{ns}$, where some PMTs have a relative difference of up to $\pm 3.0\,\text{ns}$. Furthermore, the data PMTs show a systematic uncertainty of $\mu(d)$ on a run-by-run basis, which is larger than the uncertainty from the fit $\Delta\mu(d)$, and it is different for different PMTs. Accommodating these systematic differences between the individual PMTs results in best fit results of the effective refractive index for the data as $p_1 = 1.6867 \pm 0.0026$, $p_2 = (-0.296 \pm 0.020)\,10^{-2}\,\text{m}^{-1}$. The MC is analysed in the same manner as the data and gives the best fit results as $p_1 = 1.7141 \pm 0.0026$, $p_2 = (-0.575 \pm 0.015)\,10^{-2}\,\text{m}^{-1}$. The

linear dependence of n_g^{eff} comes from the wavelength distribution of the scintillation light and the dispersion in the LS. This results in an over-selection of photons with a longer average wavelength, and a faster average group velocity for larger distances, which is reflected in the measured $n_g^{\text{eff}}(d)$ parameter values. It follows from the effective refractive index, that the group velocity of the detectable photons is larger for the data than for the MC, given the distances between 3 m to 9 m.

The uncertainties include the parameter anti-correlation of p_1, p_2 in the fit, as well as the systematic uncertainties of the selection of the energy region, the hit time histogram binning, and the run-by-run systematic uncertainty of $\mu(d)$. These systematic uncertainties are estimated maximally conservative as the absolute difference between the fit results of the different selections. These results have a relative small uncertainty, compared to the position resolution of $\sim$10 cm, for events with 1 MeV electron equivalent energy. For example, using Equation 4.4, at $d = 6.5$ m this position resolution provides a ToF uncertainty of up to $\Delta t = 0.55$ ns, compared to the ToF uncertainty from the effective refractive index of data, with only $\Delta t = 0.06$ ns.

The measurement of the effective refractive index for MC is validated trough a modified G4BX2 simulations, where the scintillation is emitted instantly and re-emission is turned off. It is found that the analysis method, using the α-source at many different positions, gives ToF results that are well in agreement with the MC true, underlying group velocity of the early PMT hits. This means that the analysis method used here provides a valid measurement of the effective refractive index and group velocity of the earliest detectable PMT hits. This measurement is taken in-situ, including the effect of the scintillation wavelength distribution and the dispersion in the full LS volume. The significant differences between the data and the MC results can stem from a variety of different sources, including differences in the scintillation spectrum, absorption and attenuation length and the refractive index of the LS. Ultimately, the origin of the differences can be summarised such, that the data and the MC have a different wavelength distribution of the detected scintillation hits and a difference of the refractive index, resulting in a different effective group velocity between data and MC.

Following this it would be possible to fine tune the G4BX2 parameters of the scintillation time profile to improve the agreement between data and MC, but this is not done here for two reasons: First, it is a prohibitively large amount of work, which would make it also necessary to re-perform the entire calibration for the position reconstruction, α/β discrimination, light-yield, and so on. The MC is already well able to model the spectral shape of the data and it is successfully used for the spectral fitting. Secondly, the fine tuning of the scintillation would not improve the knowledge of the Cherenkov time behaviour and there would still be a systematic difference left between Cherenkov light and scintillation in the data and the MC. Instead, only a small correction will be introduced in the G4BX2 event simulation, which is the relative Cherenkov group velocity correction gv_{ch}. This parameter is explained and measured in the following Chapter 5, using γ-calibration sources, as no dedicated Cherenkov sources exits for the Borexino detector.

The detected mis-behaviour, where the individual data PMTs show a constant hit time offset between each other, should be included and corrected for in the ToF calculation of the PMT hits. For MC this mis-behaviour does not exist, which is why it can introduce a systematic uncertainty. From the measurement here it is not clear how stable this behaviour is in time, as the calibration sources have only been deployed in a relatively short time window between 2008–2009. Consequently, the measured PMT constants must not necessarily be applicable for the entire detector live time. Considering the run-by-run hit time offset of individual PMTs it is well possible that the effective hit time offset is different from the results of this chapter compared to the full live time of the different detector phases. Therefore, additional measures are taken to select PMTs with a good time behaviour for the CID analysis. First, for the CID proof-of-principle using the ^{7}Be-ν edge energy region in Chapter 6, the

PMTs are selected according to a statistical argument between the number of the first detected PMT hits in data and MC. For the measurement of the CNO neutrinos in Chapter 7 a more sophisticated method is developed, where the cosmogenic ^{11}C events are used to produce the PMT hit time distributions for the full live-time of each phase. This takes into account the effective change that the PMT hit time behaviour can undergo throughout a single detector phase.

The analysis method presented here could also be potentially useful for future, large scale LS detectors like JUNO [69]. JUNO has a much larger detector volume of 20 kt compared to Borexino with 278 t, and with a radius of 19 m. This will likely result in a greater influence of the dispersion and consequently the effective refractive index function could need a higher order polynomial > 2 to correctly model the scintillation photon ToF behaviour. Likewise, for hybrid detectors like the proposed THEIA [72] the fitting of the individual PMT hit times distributions potentially allows for a very precise fine tuning of the MC hit time behaviour. This could be helpful for the identification of systematic uncertainties, and for the improvement of a time based event-by-event direction reconstruction algorithm.

Chapter 5

Cherenkov Group Velocity Calibration using γ Events

Abstract

The CID measurement of neutrino events necessitates a selection of early PMT hits, to increase the Cherenkov over scintillation hit ratio, as otherwise the high scintillation hit statistics would overwhelm the directional information of the Cherenkov hits. When applying such a selection of early PMT hits it is necessary to have a correct knowledge of the Cherenkov to scintillation hit ratio for each Nth-Hit, such that the number of neutrino events can be inferred from the number of Cherenkov hits in the CID data set. This can be quantified through a calibration of the Cherenkov hit time behaviour, relative to that of the scintillation hits, between the data and the MC. This chapter presents such a calibration, modeled through the so called "effective Cherenkov group velocity correction" $\mathrm{gv_{ch}}$. The $\mathrm{gv_{ch}}$ parameter has the units of $[\mathrm{ns\,m^{-1}}]$ and it is applied on the lowest level of the G4BX2 simulation, shifting the true PMT hit times of Cherenkov photons to earlier times. This correction is effectively an increase of the MC Cherenkov group velocity, relative to that of the MC scintillation photons. The $\mathrm{gv_{ch}}$ parameter is measured using γ calibration sources, for which it is possible to reconstruct the initial event direction, as no dedicated electron calibration source exists for Borexino. The calibration result is $\mathrm{gv_{ch}} = (0.108 \pm 0.006(\mathrm{stat.}) \pm 0.038(\mathrm{syst.}))\ \mathrm{ns\,m^{-1}}$, where the large systematic uncertainty is dominated by the systematic direction mis-reconstruction between the data and the MC γ events. This $\mathrm{gv_{ch}}$ parameter is an effective parameter, changing the MC Cherenkov group velocity in such a way that the angular distribution of PMT hits is in agreement between the data and the MC for the early Nth-Hits of the events.

This Chapter deals with the relative hit time calibration of the Cherenkov hits. The previous Chapter 4 showed that there is a significant difference between the data and the MC in the group velocity of the earliest detectable scintillation photons. For the scintillation hits of the data a larger group velocity (smaller effective refractive index) is measured than for MC. The likely, underlying origin of this difference is the combination of a difference in the scintillation wavelength spectrum as well as a difference in the refractive index of the LS. This means that the Cherenkov hits must also have a difference between the data and the MC, which must be measured and corrected for. Ideally, this should be done with a dedicated electron Cherenkov calibration source, where the direction of the electron events is well known. There is no such source in Borexino, as the Cherenkov directionality has not been of interest at the time of the calibration campaign. Instead, the calibration of the Cherenkov hit time is performed here on ^{54}Mn- and ^{40}K-γ events. Previously, the ^{54}Mn+^{40}K-γ-source has been used for the calibration of the energy scale, in combination with other γ-sources [85, 142].

The first Section 5.1 explains the Cherenkov calibration method using the γ-sources. The reason for the use of the γ-source, instead of an electron source, is that it is reasonably possible to perform a direction reconstruction on the γ-events, while this is not possible for the electron sources available in Borexino. The second Section 5.2 shows the event selection and the angular distribution of the detected hits, i.e. the equivalent of $\cos\alpha$, while the subsequent Section 5.3 explains the choice of the effective model for the Cherenkov calibration. The next Section 5.4 explains the origin and correction of the most relevant systematic uncertainty of this calibration: The direction reconstruction systematic. Then, Section 5.5 defines the test statistic and shows the result of the Cherenkov calibration, including additional, but relatively small systematic uncertainties. The last Section 5.6 presents a summary and conclusion of the effective Cherenkov calibration trough γ events.

The analysis method of this chapter has been jointly developed with [redacted] and results have been cross-checked and independently reproduced by her as part of her dissertation [145]. Parts of the analysis presented here are also published in the associated CID paper [135].

5.1 Method

To be able to perform a Cherenkov calibration it would be ideal to have a calibration source that either produces pure Cherenkov light without scintillation or a source where the Cherenkov light has a well known direction. Ideally, both conditions should be fulfilled. Such a source has not been produced for the Borexino calibration campaign, as the Cherenkov light is not of direct importance for the spectral neutrino analyses. While electron calibration sources exit (^{214}Bi for example), it is not possible to perform a direction reconstruction of these electron events, as the relative number of Cherenkov to scintillation hits is far too small.

Nonetheless, there is a type of calibration source, for which the Cherenkov direction can be estimated through the reconstruction of the event direction: The γ-sources[1]. These sources have been used for the energy calibration in Borexino, which necessitated them to be mono-energetic, and where the scintillation light from the associated α or β emissions had to be suppressed. To achieve this, the radioisotopes of interest have been dissolved in de-ionised water and filled in a 1" diameter quartz sphere [85].

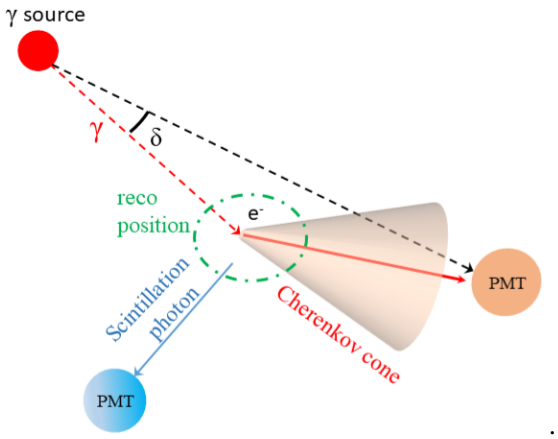


FIGURE 5.1: Schematic illustration of the $\cos\delta$ angle used for the Cherenkov calibration with a γ-source. A γ event (dashed, red line) is emitted from the calibration source position $\vec{r}_{\text{source}}$ (red circle). The initial direction of the Compton-scattered electrons (solid, red line) tends to be correlated to the γ direction. These electrons emit Cherenkov light (orange cone) that is then also correlated to the initial γ direction, while scintillation photons (blue) are uncorrelated. The directional angle $\cos\delta$ is then given by the reconstructed gamma direction and the photon hit direction. Used in [135]

Figure 5.1 shows the schematic idea of the Cherenkov calibration trough the γ-source events. The source position (red sphere) is known with a precision of <1 cm from the CCD cameras [142]. Thus, the origin of the γ-source events is known with a precision of ∼1.5 cm, given the radius of the spherical vial. The γ event deposits its energy through Compton

[1] [redacted] had the original idea of using the γ-sources

scattering in the LS. This results in a total event energy deposition that is not point-like, but rather a cloud made up of a number of individual point-like energy depositions from the Compton scattered electrons. This "deposit cloud" is indicated by the green ellipse in the figure. The energy deposits have a standard deviation of about $\sim$20 cm around the barycentre, according to the MC true deposit positions, for ^{40}K-γ events with 1.46 MeV. This deposit cloud is reconstructed as a single event position $\vec{r}_{\text{source}}$, given the γ event PMT hit times. Using the source position $\vec{r}_{\text{source}}$ and the reconstructed γ event position $\vec{r}_{\gamma}$ it is possible to estimate the γ event direction. The Compton scattered electrons can have enough energy to produce Cherenkov photons, shown in the red cone of Figure 5.1. Using the known position $\vec{r}_{\text{PMT}}$ of the hit PMTs it is then possible to calculate the angle δ, which is an analogue to the definition of $\cos\alpha$ for the solar neutrinos. Here, $\cos\delta$ describes the angle between the estimated γ direction $\vec{d}_{\gamma}$ and the reconstructed photon hit direction $\vec{d}_{\text{hit}}$:

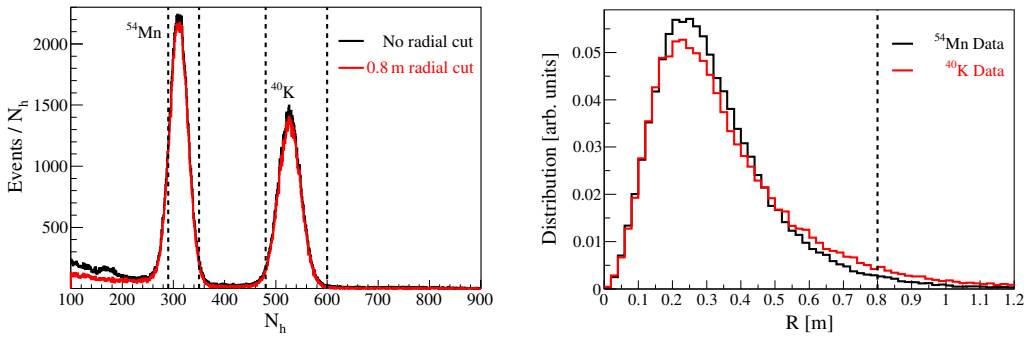
$$\begin{aligned} \vec{d}_{\gamma} &= \frac{\vec{r}_{\text{ev}} - \vec{r}_{\text{source}}}{|\,\vec{r}_{\text{ev}} - \vec{r}_{\text{source}}\,|} \\ \vec{d}_{\text{hit}} &= \frac{\vec{r}_{\text{PMT}} - \vec{r}_{\text{ev}}}{|\,\vec{r}_{\text{PMT}} - \vec{r}_{\text{ev}}\,|} \\ \cos\delta &= \vec{d}_{\text{hit}} \cdot \vec{d}_{\gamma} \end{aligned} \tag{5.1}$$

The kinetic energy of the scattered electrons is largest for a small Compton angle, thus they produce the most number of Cherenkov photons when they travel in the direction of the γ. Due to multiple Compton scattering, the overall distribution of Cherenkov hits is expected to be strongly smeared out, with respect to the reconstructed γ direction. It it reasonable to expect that there is a larger contribution of Cherenkov hits for $\cos\delta > 0$, relative to $\cos\delta < 0$, as the Cherenkov direction is correlated to the direction of the γ events, while the direction of scintillation hits is uncorrelated. Therefore, the $\cos\delta$ distribution for a large number of γ events should have some sensitivity on the time dependent number of Cherenkov hits and the Cherenkov to scintillation hit ratio. The time like variable Nth-Hit is used to perform a time sorting of the PMT hits, as is explained in Chapter 3. The ToF correction (Equation 3.7), necessary for the correct estimation of the emission time of the photons, and for the Nth-Hit sorting, uses the effective refractive indices measured in Chapter 4. Additionally, the relative hit time offsets between the PMTs from Chapter 4 are also included in the ToF correction. The general idea of the Cherenkov calibration is then to produce a model of the difference in the Cherenkov hit time distribution between the data and the MC and then to perform a fit of the model parameters on the Nth-Hit $\cos\delta$ distributions of the data.

5.2 The γ event selection

The Borexino calibration has made use of a number of γ-sources, namely ^{85}Sr (0.514 MeV), ^{54}Mn (0.835 MeV), ^{65}Zn (1.115 MeV), ^{60}Co (1.173 MeV + 1.333 MeV), ^{40}K (1.461 MeV) [85]. Only ^{54}Mn and ^{40}K are able to provide enough event statistic at a high enough energy, such that enough Cherenkov hits are included in the $\cos\delta$ distribution to perform a measurement. The reasons for this are that ^{85}Sr has a too low energy, ^{65}Zn has a too low event statistic. The ^{60}Co is unusable because one event is made up of two consecutive γ emissions with an uncorrelated direction, which makes the direction reconstruction impossible.

The ^{54}Mn- and ^{40}K-γ event selection is illustrated in Figure 5.2. On the right the number of hit PMTs, N_h distribution is plotted for the combined ^{54}Mn+^{40}K-source. The source is positioned close to the detector centre, at $x = 8.6$ cm, $y = -8.8$ cm, $z = 0.2$ cm (Runs 10415+10416). The black histogram shows all events during the run time, while the red



(a) (b)

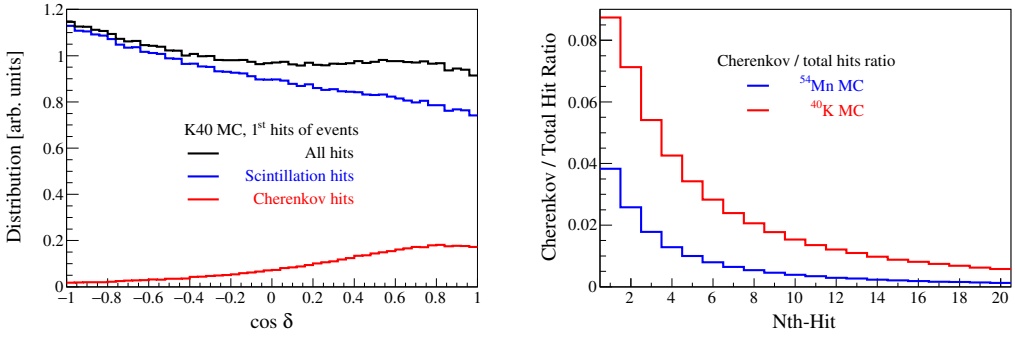
FIGURE 5.2: The selection of the ^{54}Mn- and ^{40}K-γ data events. (a) Distribution of the N_h, number of hit PMTs of the events, with (red) and without (black) a radial cut of 0.8 m around the source position. The dotted lines show the selected energy regions for the respective γ events. (b) Radial distribution of the data events within the selected energy regions of ^{54}Mn (black) and ^{40}K (red). The dotted line shows the radial cut.

histogram shows the event selection after a radial cut of 0.8 m between the source and reconstructed event position. The main peaks of the ^{54}Mn and ^{40}K events are well visible at around 310 N_h and 530 N_h, respectively. Below $\sim$ 260 N_h the influence of the intrinsic LS background is visible (mostly ^{210}Po and ^{14}C), as well as its reduction by a factor ~ 2, due to the radial cut. The energy region of interest is selected as $290 < N_h < 350$ for the ^{54}Mn events and $480 < N_h < 600$ for the ^{40}K events. These selections are shown as the dotted, black lines. The number of intrinsic LS background events within the RoI is considered safely negligible compared to the number of γ events. The number of selected data events are $N(^{54}\text{Mn}) = 86414$ and $N(^{40}\text{K}) = 72842$.

The right side of Figure 5.2 shows the corresponding radial distribution of these energy regions, where the black and red distributions correspond to ^{54}Mn and ^{40}K, respectively. The radial cut is shown as a dotted line. While electron events at these energies provide point-like energy depositions with a path length <5 mm, the γ events have a much larger free path length. The average difference between the source position and the reconstructed event position is $\sim$32 cm and $\sim$36 cm, for ^{54}Mn and ^{40}K, respectively. This is larger than the fit uncertainty of the reconstructed position, which allows for the reconstruction of the initial γ direction $\vec{d}_\gamma$, as defined in Equation 5.1. The necessity of the direction reconstruction, using the position reconstruction, introduces the most relevant systematic uncertainty of this Cherenkov calibration, which is discussed in Section 5.4.

The G4BX2 MC model is produced with a statistic of 20 $\times$ data, with the same energy and radial event selection as for the data. The Nth-Hit $\cos\delta$ histograms are produced according to Equation 5.1, using the measured effective refractive index from Chapter 4 for the ToF correction. Figure 5.3(a) shows the $\cos\delta$ distribution of the ^{40}K MC, for the first hits of the events, Nth-Hit = 1. The black histogram shows all first hits on the PMTs, while the blue and red histograms show only the scintillation or Cherenkov hits, respectively. The contribution of Cherenkov hits follows the qualitative expectations, where the hits are concentrated in forward direction $\cos\delta > 0$, with a broad peak around $\cos\delta \approx 0.8$. The scintillation hits dominate over the Cherenkov hits and they do not have a flat $\cos\delta$ distribution. Instead, their $\cos\delta$ distribution follows a negative slope, with a decreasing contribution for larger $\cos\delta$.

This effect stems from a selection bias of the earliest detected hits, introduced by the ToF correction. First, the position reconstruction has a fit uncertainty for the barycentre of the γ event deposit cloud. Those PMTs, which are closest to the reconstructed position also



(a) (b)

FIGURE 5.3: (a) The $\cos\delta$ distribution of the ^{40}K MC for the first hits of the events (black), with the contribution of Cherenkov hits (red) and scintillation hits (blue). (b) The Nth-Hit dependence of the Cherenkov to total hit ratio. The ^{54}Mn MC (0.835 MeV) (blue) has a maximum at the first hits of the events of 3.8% and the ^{40}K MC (1.461 MeV) (red) has a maximum of 8.7%. The overall MC Cherenkov to total hit ratio is 0.2% and 0.5% for ^{54}Mn and ^{40}K, respectively.

correspond to the positive $\cos\delta > 0$ values, as the reconstructed direction is given simply by the normalised reconstructed position. Likewise, this means that the PMTs which are furthest away from the reconstructed position can only contribute to negative $\cos\delta < 0$ values. Should the reconstructed position now be closer to some PMT, along the reconstructed event direction, compared to the true barycentre of the event, then the hits on those PMTs, which are furthest away ($\cos\delta < 0$) are over-corrected for their ToF. The close by PMTs ($\cos\delta > 0$) are under-corrected for the same reason. The consequence is then that the corresponding, reconstructed emission time of those PMT hits with $\cos\delta < 0$ is estimated at too early times. These hits are then over-selected as the first hit of the events, resulting in a higher probability for the selection of $\cos\delta < 0$ hits, i.e a negative slope in $\cos\delta$. This effect is amplified by the cloud like distribution of energy deposits for the single γ events. Later Nth-Hits > 10 show a flat $\cos\delta$ distribution of the scintillation hits.

Figure 5.3(b) shows the fraction of MC Cherenkov hits as a function of the Nth-Hits. The ^{54}Mn MC (0.835 MeV) is shown in blue, with a maximum of 3.8% and the ^{40}K MC (1.461 MeV) is shown in red, with a maximum of 8.7%. As expected, the earliest hits have the highest Cherenkov hit fraction, much higher than the overall MC Cherenkov to total hit ratio of 0.2% and 0.5% for ^{54}Mn and ^{40}K, respectively. From this it follows that the Nth-Hit dependent $\cos\delta$ distribution can in principle be used for the calibration of the Cherenkov hit time behaviour of the MC, as there is a sufficient Cherenkov hit statistic, as well as a clear difference in the $\cos\delta$ shape of the Cherenkov and scintillation hits.

5.3 Choice of the effective Cherenkov calibration model

The goal of this section is to define a reasonable model for the calibration of the MC Cherenkov hit behaviour. In the last Chapter 4 the effective group velocity of scintillation hits is measured to be faster in the data compared to the MC. It is now reasonable to assume that the same is true for the effective group velocity of the detectable Cherenkov hits, as the underlying reason for this data–MC difference likely comes from a difference in their respective refractive index functions. Therefore, the Cherenkov calibration here consist of an effective change in the group velocity of the Cherenkov hits. This change is performed on the lowest level of the G4BX2 MC: The MC true hit time t^{true} of the Cherenkov photon hit on the PMT cathode, before any electronics simulation or event reconstruction. The so called "effective Cherenkov

group velocity correction" $\mathrm{gv}_{\mathrm{ch}}$ is defined as:

$$\begin{aligned} t^{\mathrm{corr}} &= t^{\mathrm{true}} - \mathrm{gv}_{\mathrm{ch}} \cdot L^{\mathrm{true}} \\ &= t^{\mathrm{true}} - \frac{\Delta \mathrm{n}_{\mathrm{ch}}}{c} \cdot L^{\mathrm{true}} \end{aligned} \tag{5.2}$$

This $\mathrm{gv}_{\mathrm{ch}}$ has the units of $[\mathrm{ns\,m^{-1}}]$ and corresponds to the difference in the effective refractive index $\Delta \mathrm{n}_{\mathrm{ch}}$ of Cherenkov photons between data and MC, which results in a difference of the ToF, given the MC true photon path length L^{true}. The true MC hit time of Cherenkov photons t^{true} is changed to a corrected value t^{corr} trough $\mathrm{gv}_{\mathrm{ch}}$, such that the Nth-Hit $\cos\delta$ distributions are in agreement between the data and the MC. This means that the number of Cherenkov hits is effectively increased for the early Nth-Hits in the MC, such that it corresponds to the Nth-Hit dependent number of Cherenkov hits in the data $\cos\delta$ distributions. It is expected here that $\mathrm{gv}_{\mathrm{ch}}$ is positive, because the effective group velocity of the scintillation is faster in the data than in the MC and this should also be true for the Cherenkov hits, i.e. a decrease of t^{corr} relative to t^{true}.

While the effective refractive index of scintillation is well described by a linear function, the Cherenkov calibration parameter $\mathrm{gv}_{\mathrm{ch}}$ is modeled here as a constant. This does not necessarily mean that the true, underlying difference between the data and the MC is correctly modeled through this single parameter. It is rather expected that a higher order approximation is not sensitive, given the statistical and systematic uncertainties of this γ calibration. Ultimately, the $\mathrm{gv}_{\mathrm{ch}}$ is an effective parameter, that just needs to be good enough to explain the difference between the data and the MC $\cos\delta$ histograms.

The calibration is performed on γ events, while the neutrinos are detected through electron events. This makes it necessary to introduce the effective Cherenkov group velocity correction $\mathrm{gv}_{\mathrm{ch}}$ at the lowest level of the MC, such that the hit time differences of the Cherenkov and the scintillation hits can take into account the photon path lengths L^{true}, given the true position of the γ event energy depositions. At the same time this correction is minimally invasive, as it only changes the time behaviour of the sub-dominant Cherenkov hits, instead of, for example, the shape of the refractive index function. The consequence of this is that no re-calibration of other MC parameters is necessary. The $\mathrm{gv}_{\mathrm{ch}}$ corresponds to an effective calibration of the Cherenkov time behaviour given the full detector response, including the ToF correction using the measured refractive index from the previous chapter. Any systematic differences in the Cherenkov time behaviour, which are introduced by the constant model of $\mathrm{gv}_{\mathrm{ch}}$, will be investigated in the respective CID measurement of the solar neutrinos. Ultimately, it has turned out that this model is sufficient for the measurement of both the ^{7}Be- and CNO-neutrinos, as explained in the respective Chapters 6, 7.

Both the ^{54}Mn and ^{40}K MC is simulated with four different $\mathrm{gv}_{\mathrm{ch}}$ values: ($0.08\,\mathrm{ns\,m^{-1}}$, $0.10\,\mathrm{ns\,m^{-1}}$, $0.16\,\mathrm{ns\,m^{-1}}$, $0.22\,\mathrm{ns\,m^{-1}}$). These full G4Bx2 γ simulations are then used to produce the $\cos\delta$ histograms for all values with $0.08\,\mathrm{ns\,m^{-1}} < \mathrm{gv}_{\mathrm{ch}} < 0.22\,\mathrm{ns\,m^{-1}}$ through the linear interpolation of the nearest neighbours.

5.3.1 Nth-Hit(max) selection

To maximise the sensitivity of the $\mathrm{gv}_{\mathrm{ch}}$ measurement it is necessary to select a number of early PMT hits, to get an increased Cherenkov to total hit ratio. From Figure 5.3(b) one could select the very first hit of each event, as it has the highest Cherenkov to total hit ratio. This is not necessarily the best choice, as later hits, with a lower Cherenkov ratio, can still have enough Cherenkov hits to provide useful information for the measurement, through their respective $\cos\delta$ histograms.

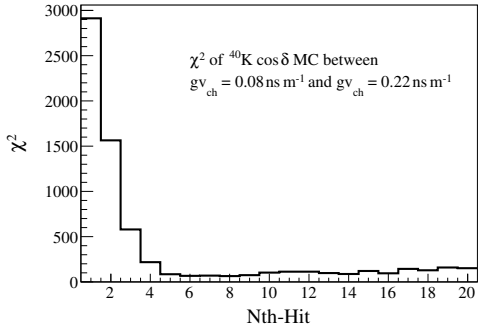
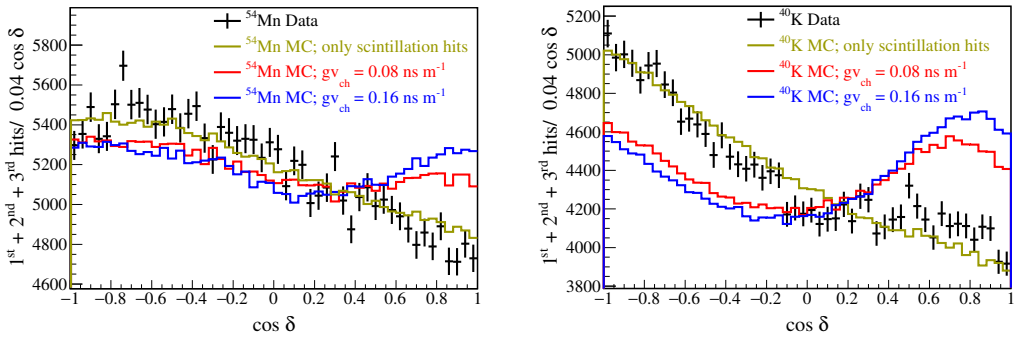


FIGURE 5.4: The χ^2 value between the individual Nth-Hit $\cos\delta$ histograms of the ^{40}K MC, with the two different values of $gv_{ch} = 0.08\,ns\,m^{-1}$ and $gv_{ch} = 0.22\,ns\,m^{-1}$. Larger χ^2 values come from larger differences in the $\cos\delta$ histograms. This corresponds to a larger expected sensitivity for the gv_{ch} measurement.

Thus, the number of early hits to consider in the analysis, Nth-Hit(max), is estimated through Figure 5.4. Here, the absolute χ^2 value of the individual Nth-Hit $\cos\delta$ histograms is calculated between the parameter values of $gv_{ch} = 0.08\,ns\,m^{-1}$ and $gv_{ch} = 0.22\,ns\,m^{-1}$ for the ^{40}K MC. A large χ^2 value means, that this Nth-Hit $\cos\delta$ histograms have large differences for the two gv_{ch} values, which corresponds to a large expected sensitivity for the gv_{ch} measurement. These χ^2 value are calculated using the statistics of the MC and a binning of 50. It can be seen that Nth-Hit $= 1,2,3,4$ provide the largest expected sensitivity, while later Nth-Hits are approximately flat in χ^2. The number of early hits, until which the analysis is then performed is selected as Nth-Hit(max) $= 3$. The values of Nth-Hit(max) $= 2$ and Nth-Hit(max) $= 4$ are used to estimate the systematic uncertainty coming from this particular Nth-Hit(max) $= 3$ choice.

5.4 Systematic direction mis-reconstruction

In the course of this analysis it has been found out that the biggest systematic uncertainty of the gv_{ch} is given by the direction reconstruction. This section describes the origin, impact and solution to the systematic direction mis-reconstruction between the data and the MC γ events.



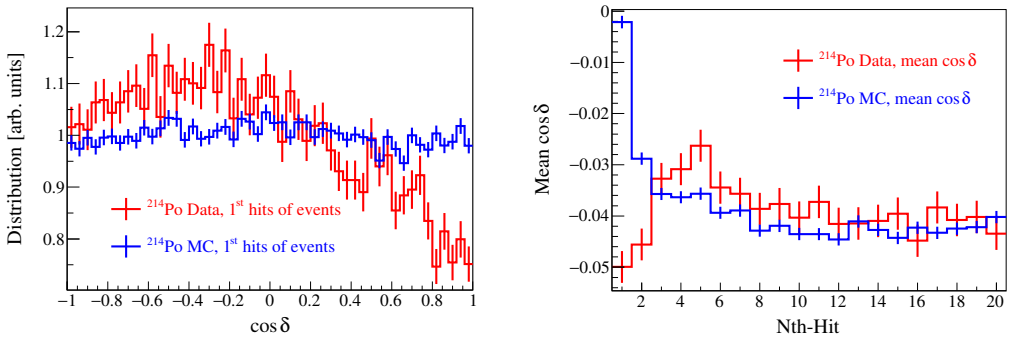
(a) (b)

FIGURE 5.5: Comparison of the data and the MC $\cos\delta$ for the sum of the three hits of the events (Nth-Hit =1, 2, 3). The data (black) shows a negative slope, but no pronounced Cherenkov peak. The MC with $gv_{ch} = 0.08\,ns\,m^{-1}$ (red) and $gv_{ch} = 0.16\,ns\,m^{-1}$ (blue). Both also have a negative slope for $\cos\delta < 0$, but here a Cherenkov peak is clearly visible. This peak is larger for $gv_{ch} = 0.16\,ns\,m^{-1}$, as is expected, due to the larger contribution to early hits from the faster Cherenkov photons. No value of gv_{ch} is able to describe the $\cos\delta$ histogram of the data well, while the pure scintillation hits (yellow) show a much better agreement between the data and the MC. (a) For ^{54}Mn. (b) For ^{40}K.

Figure 5.5 shows the summed first, second and third hits of the events (Nth-Hit = 1, 2, 3) $\cos\delta$ distributions of the data in black in comparison to the gv_{ch} corrected MC. The ^{54}Mn

events are shown on the left and the ^{40}K events on the right. Exemplary MC distributions with the values of $gv_{ch} = 0.08$ ns and $gv_{ch} = 0.16$ ns are shown in red and blue, respectively. The pure MC scintillation hits are shown in yellow and all MC distributions are normalised to the statistics of the data. It can be seen for both ^{54}Mn and ^{40}K events, that no gv_{ch} value is able to describe the $\cos\delta$ distribution of the data. In fact, the $\cos\delta$ distribution of pure MC scintillation hits gives the best agreement with the data. For the ^{54}Mn it would even require a non-physical, negative number of Cherenkov hits to provide a good agreement between the data and the MC. This means not only that the selected model of the Cherenkov hit time gv_{ch} is unable to explain the data, but rather that no Cherenkov hit time model of any sort is able to bring the MC and the data $\cos\delta$ distributions in agreement with each other.

Thus, the underlying difference between the data and the MC $\cos\delta$ distributions must have a systematic uncertainty contribution from some source. This source has been identified as the difference of the direction reconstruction, which provides systematically different results between the data and the MC, for the $\cos\delta$ value of the earliest Nth-Hits.



(a) (b)

FIGURE 5.6: (a) The $\cos\delta$ distribution of ^{214}Po events for a source position in the centre of the detector. The ^{214}Po α events have no directionality, as the energy deposition is point-like and no Cherenkov photons can be produced from α particles at these energies (7.4 MeV). Nonetheless, the $\cos\delta$ distributions of the first hits of the events look different for the data (red) and the ^{214}Po MC (blue). Consequently, this $\cos\delta$ difference must come from a systematical difference of the direction (position) reconstruction. (b) The Nth-Hit dependent mean value of the $\cos\delta$ distributions. For early Nth-hits the data and the MC have the largest difference, while for Nth-Hit > 12 both distributions are in agreement.

The effect can be seen in Figure 5.6, where the Nth-Hit dependent $\cos\delta$ distributions are produced for the ^{214}Po source, for the source position in the centre of the detector, and with the event selection from Chapter 4. The source provides only α-particle events, it has no Cherenkov photons at all, and it has a point-like energy deposition. The ^{214}Po source has no event direction, but the definition of Equation 5.1 can still be applied to investigate the performance of the direction reconstruction, based on the position reconstruction. The left side shows the $\cos\delta$ distribution of the first hits of the ^{214}Po events for the data in red and for the MC in blue. These are purely scintillation hits, that show some structure for the data and a relative flat behaviour for the MC. Unexpectedly, the data and the MC are significantly different from each other. This can also be seen on the right side of Figure 5.6, where the average $\cos\delta$ value is plotted vs. the Nth-Hit. The first Nth-Hit has the biggest difference, while the $\cos\delta$ distributions are in agreement between the data and the MC for Nth-Hit > 12.

Since the ^{214}Po–α events neither produce Cherenkov photons nor have an event direction, the only explanation is a difference in the direction reconstruction between the data and the MC. There exists some form of correlation between the early hits of an event and its reconstructed event direction, which is different between the data and the MC. The origin of

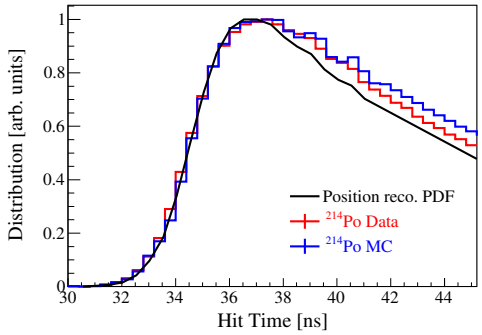
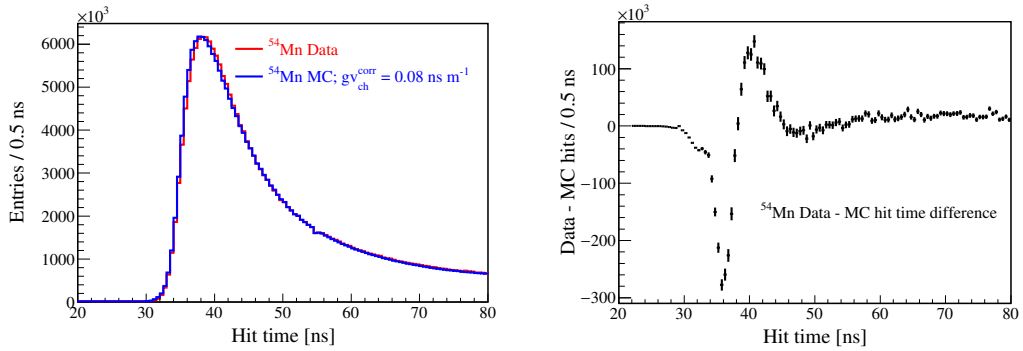


FIGURE 5.7: The ToF corrected PMT hit time distribution of ^{214}Po events for the data (red) and the MC (blue). These distributions are compared to the single photo-electron position reconstruction PDF (black). The PDF has been produced for electron events, which explains the difference to the α events. The data and the MC have a small, but significant difference in their hit time distributions, while both are reconstructed with the same PDF.

this behaviour is shown in Figure 5.7, where the ToF corrected PMT hit time of the ^{214}Po data is shown in red and the MC is shown in blue. The probability density function (PDF) for the single photo-electron, used in the position reconstruction algorithm, is shown for comparison in black. It can be seen that the ^{214}Po distributions are different from the PDF, as it has been produced from electron events and has therefore a faster decay-time.

What is important here, is that the ^{214}Po event hit time distributions are different for the data and the MC, which must result in some difference in the position (direction) reconstruction of the individual events. Even at early times, in the rising hit time slope, the data and the MC hit time distributions are significantly different from each other. These early hit times have a different impact on the position (direction) reconstruction, which then causes the difference of the direction reconstruction, resulting in the $\cos\delta$ distributions from Figure 5.6.



(a) (b)

FIGURE 5.8: (a) The ToF corrected PMT hit time distribution of the ^{54}Mn events for the data (red) and the MC (blue), normalised to the statistics of the data. While both distributions look similar, they are in fact significantly different, given the statistics. This can be seen in (b), where the absolute difference between the two histograms is shown with the corresponding uncertainty.

The conclusion is then that this difference in the direction reconstruction also exists for the γ events and must be corrected first, before a calibration of the Cherenkov hit time behaviour is possible. This can be seen in Figure 5.8, where the right side shows the ToF corrected hit time distributions of the ^{54}Mn data in red and the MC in blue. The left side shows the absolute difference between these histograms, which are significantly different from each other. Different values of $\mathrm{gv_{ch}}$ result in slightly different MC hit time distributions.

How then is the direction reconstruction equalised between the data and the MC? As the MC hit time distribution cannot easily be changed, here instead the position (direction) reconstruction is modified. The $\mathrm{gv_{ch}}$ calibration is then performed in a two step approach. First, the idea is to use different, separate position (direction) reconstruction PDFs for the data and the MC, such that the $\cos\delta$ distributions are in agreement between the data and the MC.

This is done independently for every value of gv_{ch} on the ^{54}Mn source, as the information on gv_{ch} is lost due to the modification of the direction reconstruction. It is a-priori not clear which gv_{ch} value is correct and as such all gv_{ch} values must be treated equally. Correcting the direction reconstruction systematic should then result in a good agreement between the ^{54}Mn $\cos\delta$ distributions of the data and the MC, independently of the gv_{ch} value. After such a good modification of the position (direction) reconstruction is found, the second step is then to use the same PDF for the direction reconstruction of the ^{40}K events. it follows then that there should be a sensitivity for the measurement of the effective Cherenkov group velocity correction, as the higher energy ^{40}K events provide more Cherenkov hits than the ^{54}Mn events.

The position (direction) reconstruction PDF could be modified with any arbitrary function and number of parameters; the parameter space is in principle infinite, as it is an effective correction. Here, the choice of the PDF modification is informed by the shape of the hit time difference in Figure 5.8. The PDF is modified through the addition of this function $f(t)$, a Gaussian derivative described by the amplitude A, the mean value μ and the standard deviation σ:

$$f(t) = A \cdot (\mu - t) \cdot \exp\left(-\frac{1}{2}\frac{(t-\mu)^2}{\sigma^2}\right) \tag{5.3}$$

These parameters (A, μ, σ) are effectively treated similar to nuisance parameters for the measurement of gv_{ch}. Therefore, for each value of the gv_{ch} parameter the position of all events must be reconstructed again, which makes this calibration tedious. The selection of the parameters is not based on a minimisation algorithm, but rather on a number of position re-reconstructions for all events, for different sets of parameters (A, μ, σ), which are selected by hand. Then, the best performing sets of parameters are used for the position (direction) reconstruction of the ^{40}K events. The different (A, μ, σ) parameter values used for the position re-reconstruction are given in Appendix B. The data and each gv_{ch} MC has six sets of parameters.

To summarise: First the (A, μ, σ) parameters are used to calibrate the direction mis-reconstruction between data and MC on and then the corrected ^{40}K events are used to measure the effective Cherenkov group velocity correction gv_{ch}.

Modified MC performance on position reconstruction

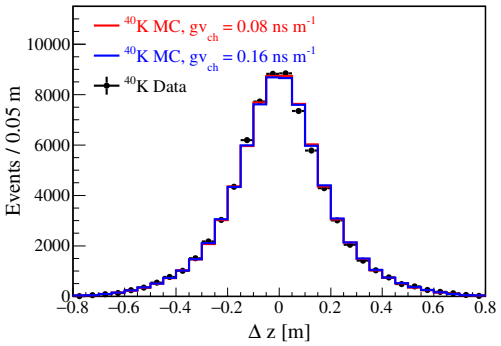


FIGURE 5.9: Exemplary performance of the position reconstruction on the z-position with $\Delta z = z_{source} - z_{rec}$. The ^{40}K data (black), MC with $gv_{ch} = 0.08\,\mathrm{ns\,m^{-1}}$ (red) and $gv_{ch} = 0.16\,\mathrm{ns\,m^{-1}}$ (blue) all show the same general performance of the position reconstruction. Additionally, these distributions are produced with different values of (A, μ, σ) resulting in slightly different position reconstruction PDFs.

Figure 5.9 shows an example of the impact on the position reconstruction, given the ^{40}K data in black and the MC in red, blue for different values of gv_{ch}, which also have different (A, μ, σ) values for the modified position reconstruction PDF. It can be seen that the general performance of the position reconstruction is not impacted by either the gv_{ch} parameter or the small change in the PDF. This means that the general idea of a minimally invasive Cherenkov calibration is satisfied.

5.5 Fit definition and result

In this section the result of the effective Cherenkov group velocity correction $\mathrm{gv_{ch}}$ measurement is presented. As described above this is done through a two step approach. The systematic direction mis-reconstruction between the data and the MC is corrected through the modification of the position reconstruction PDF using the parameters (A,μ,σ) for both the data and the MC, dependent on $\mathrm{gv_{ch}}$. Then a χ^2 scan of the $\mathrm{gv_{ch}}$ parameter is performed on the $\cos\delta$ histograms of both γ-sources, as defined in Equation 5.4:

$$\chi^2\left(\mathrm{gv_{ch}},(A,\mu,\sigma)_{\mathrm{data}},(A,\mu,\sigma)_{\mathrm{MC}}\right)=\sum_{n=1}^{\text{Nth-Hit(max)}}\sum_{i=1}^{I}\left(\frac{(\mathrm{norm}\cdot M_i^n-D_i^n)^2}{\mathrm{norm}\cdot M_i^n+\mathrm{norm}^2\cdot M_i^n}\right) \tag{5.4}$$

The number of histogram entries at the $\cos\delta$ bin number i, of the Nth-Hit $= n$ is given by D_i^n and M_i^n for the data and the MC, respectively. Here, "norm" is the scaling factor between the MC and the data event statistics. The term "$\mathrm{norm}^2\cdot M_i^n$" in the denominator takes into account the finite statistics of the MC. The selected binning for the $\cos\delta$ histograms is $I=50$ and the selected number of Nth-Hits until the analysis is performed is Nth-Hit(max) $= 3$, given the estimation from Figure 5.4. The $\cos\delta$ distribution of the data depends on the position reconstruction PDF $(A,\mu,\sigma)_{\mathrm{data}}$, while the MC depends on $(A,\mu,\sigma)_{\mathrm{MC}}$ as well as on the $\mathrm{gv_{ch}}$ parameter of interest. The $\chi^2(\mathrm{gv_{ch}})$ values are calculated for both the ^{54}Mn and the ^{40}K events, using the same values of $A,\mu,\sigma)_{\mathrm{data}}$, $(A,\mu,\sigma)_{\mathrm{MC}}$.

It is time prohibitive to let a minimisation algorithm run for the selection of $(A,\mu,\sigma)_{\mathrm{data}}$ and $(A,\mu,\sigma)_{\mathrm{MC}}$, as the position reconstruction must be re-performed for every change in (A,μ,σ) and for all the data and the MC events. For this reason a number of six sets of parameters (A,μ,σ) have been produced for the data and each fully simulated MC value of $\mathrm{gv_{ch}}$ (see Appendix B). The analysis is then considered valid only under the condition that both $\chi^2(^{54}\mathrm{Mn})/\mathrm{ndf}$ and $\chi^2(^{40}\mathrm{K})/\mathrm{ndf}$ are smaller than 1.5:

$$\begin{aligned}\chi^2\left(\mathrm{gv_{ch}},(A,\mu,\sigma)_{\mathrm{data}},(A,\mu,\sigma)_{\mathrm{MC}}\mid{}^{54}\mathrm{Mn}\right)/\mathrm{ndf}<1.5\\ \chi^2\left(\mathrm{gv_{ch}},(A,\mu,\sigma)_{\mathrm{data}},(A,\mu,\sigma)_{\mathrm{MC}}\mid{}^{40}\mathrm{K}\right)/\mathrm{ndf}<1.5\end{aligned} \tag{5.5}$$

This somewhat arbitrary choice of the χ^2/ndf value introduces a systematic uncertainty, which needs to be evaluated. This is done using a toy-MC analysis described later.

First, Figure 5.10 illustrates the general improvement of the analysis after the application of the direction (position) reconstruction modification through (A,μ,σ), in comparison to the nominal position reconstruction PDF used in Figure 5.5. The $\cos\delta$ histograms of the first three hits of the events (Nth-Hit = 1, 2, 3) are summed up for illustration purposes. The black crosses are the γ data for the ^{54}Mn source on the left and the ^{40}K source on the right. Both γ-sources use the same $(A,\mu,\sigma)_{\mathrm{data}}$ values for the direction reconstruction. It can be seen that both γ-sources have a negative slope imposed on the $\cos\delta$ distribution, as described before. The ^{40}K events provide a visible Cherenkov peak, while this is not visible for the lower energy ^{54}Mn events. The impact of the modified direction reconstruction is visible for the ^{54}Mn source on the left, where the corresponding MC $\cos\delta$ histograms are shown for a value of $\mathrm{gv_{ch}}=0.08\,\mathrm{ns\,m^{-1}}$ in red and $\mathrm{gv_{ch}}=0.16\,\mathrm{ns\,m^{-1}}$ in blue. Both $\mathrm{gv_{ch}}$ values show a good agreement with the data. The ^{54}Mn source has no sensitivity for the $\mathrm{gv_{ch}}$ measurement, but it can clearly be used to constrain the $(A,\mu,\sigma)_{\mathrm{data}}$, $(A,\mu,\sigma)_{\mathrm{MC}}$ parameters. The ^{40}K source on the other hand does have a sensitivity for the measurement of $\mathrm{gv_{ch}}$, as shown on the right side of Figure 5.10. Using the same values $(A,\mu,\sigma)_{\mathrm{MC}}$ to reconstruct the direction of the ^{40}K events results in a clear difference for the different values of $\mathrm{gv_{ch}}=0.08\,\mathrm{ns\,m^{-1}}$ and $\mathrm{gv_{ch}}=0.16\,\mathrm{ns\,m^{-1}}$. This is in contrast to the previous Figure 5.5 without the direction reconstruction correction. From this plot it is possible the roughly estimate the best fit value

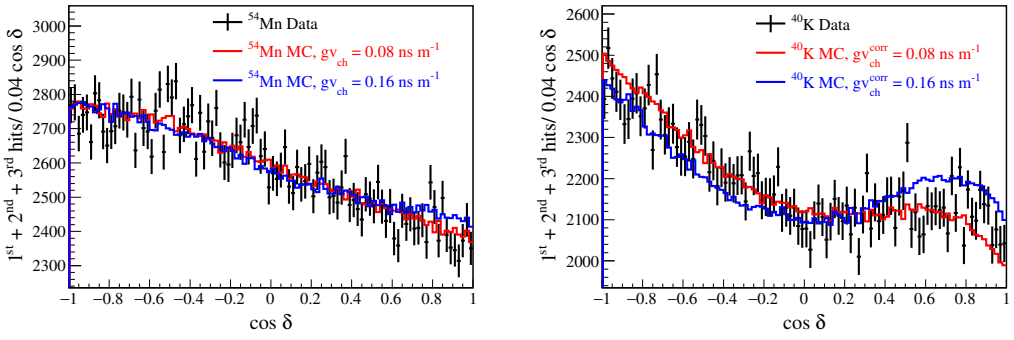


FIGURE 5.10: Comparison of the data and the MC $\cos\delta$ for the sum of the three hits of the events (Nth-Hit =1, 2, 3), after correcting the direction mis-reconstruction between the data and the MC. (a) For ^{54}Mn. (b) For ^{40}K. The data (black) shows a negative slope, and a Cherenkov peak for the ^{40}K source. The MC with $gv_{ch} = 0.08\,\text{ns}\,\text{m}^{-1}$ (red) and $gv_{ch} = 0.16\,\text{ns}\,\text{m}^{-1}$ (blue) both also have a negative $\cos\delta$ slope, while now only the ^{40}K MC has a visible Cherenkov peak. For ^{54}Mn all values of gv_{ch} are able to describe the $\cos\delta$ histogram of the data reasonably well, while the ^{40}K $\cos\delta$ distribution has a clear sensitivity for the gv_{ch} measurement.

of gv_{ch} to be between $0.08\,\text{ns}\,\text{m}^{-1}$ and $0.16\,\text{ns}\,\text{m}^{-1}$.

From the 144 produced combinations between $(A,\mu,\sigma)_{\text{data}}$ and $(A,\mu,\sigma)_{\text{MC}}$ only four best fit gv_{ch} results pass the condition of Equation 5.5:

- $gv_{ch}^{a} = (0.109 \pm 0.006)\,\text{ns}\,\text{m}^{-1}$, $\chi^2/\text{ndf} = 1.32$
- $gv_{ch}^{b} = (0.102 \pm 0.006)\,\text{ns}\,\text{m}^{-1}$, $\chi^2/\text{ndf} = 1.43$
- $gv_{ch}^{c} = (0.110 \pm 0.007)\,\text{ns}\,\text{m}^{-1}$, $\chi^2/\text{ndf} = 1.47$
- $gv_{ch}^{d} = (0.112 \pm 0.007)\,\text{ns}\,\text{m}^{-1}$, $\chi^2/\text{ndf} = 1.48$

The statistical uncertainty is estimated trough $\Delta\chi^2 = \pm 1$. The weighted average gives a value of $gv_{ch} = (0.108 \pm 0.006)\,\text{ns}\,\text{m}^{-1}$ and these results are illustrated in Figure 5.11 in red.

5.5.1 Systematic uncertainties

The biggest systematic uncertainty of this analysis is given by the need for the direction reconstruction modification through (A,μ,σ). This systematic is estimated with a toy-MC study, as follows:

1. The MC is used to sample pseudo-data, using a true, injected value of $gv_{ch} = 0.1\,\text{ns}\,\text{m}^{-1}$, for both ^{54}Mn and ^{40}K.
2. The pseudo-data is produced for all different values of $(A,\mu,\sigma)_{\text{MC}}$
3. Every $(A,\mu,\sigma)_{\text{MC}}$ set of pseudo-data $\cos\delta$ histograms is analysed according to Equation 5.4, again for every set of $(A,\mu,\sigma)_{\text{MC}}$ for the MC $\cos\delta$ histograms. This gives six best fit results for every pseudo-data set.
4. The analysis is performed 10k times for every pseudo-data $(A,\mu,\sigma)_{\text{MC}}$ set. In total there are then 36 best fit gv_{ch} distributions, given the injected value of $0.1\,\text{ns}\,\text{m}^{-1}$.
5. The average value of these best fit gv_{ch} distributions then gives the systematic offset of interest, introduced by performing the analysis with an incorrect set of $(A,\mu,\sigma)_{\text{MC}}$.

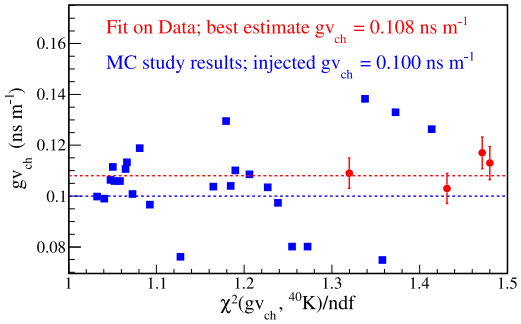


FIGURE 5.11: The results of gv_{ch} measurement, after correction of the direction mis-reconstruction between the data and the MC. The best fit gv_{ch} values are plotted vs. their $\chi^2(^{40}\mathrm{K})/\mathrm{ndf}$. The data results (red) passing the selection of $\chi^2/\mathrm{ndf} < 1.5$ are shown with their statistical uncertainty. The MC values (blue) show the average expected best fit gv_{ch} for a number of 10k toy-analyses, for an injected value of $gv_{ch} = 0.1\,\mathrm{ns\,m^{-1}}$. Used in [135].

The result of these toy-MC analyses can be seen in Figure 5.11 in blue. Here, the best fit gv_{ch} values are plotted vs. the $\chi^2(^{40}\mathrm{K})/\mathrm{ndf}$ value, selected according Equation 5.5. If the pseudo-data and the analysis MC use the same values of $(A,\mu,\sigma)_{\mathrm{MC}}$ then the corresponding $\chi^2(^{40}\mathrm{K})/\mathrm{ndf}$ is relatively small and the systematic offset between the injected gv_{ch} value and the average best fit gv_{ch} value is zero, as is expected.

For large $\chi^2(^{40}\mathrm{K})/\mathrm{ndf}$ values this systematic offset increases. The systematic uncertainty, associated to the direction mis-reconstruction, is then estimated from the largest observed difference between the injected and extracted value, with $\Delta gv_{ch} = \pm 0.038\,\mathrm{ns\,m^{-1}}$. This value is situated at $\chi^2(^{40}\mathrm{K})/\mathrm{ndf} = 1.34$, which is within the region of the data results. This toy-analysis is also performed for an injected value of $gv_{ch} = 0.16\,\mathrm{ns\,m^{-1}}$, which gives a similar but smaller value of $\pm 0.035\,\mathrm{ns\,m^{-1}}$. The larger uncertainty is used for the systematics estimation. The other systematic uncertainties are sub-dominant and are estimated as follows:

- Nth-Hit(max) selection: There could be a Nth-Hit dependence on the best fit result, which is estimated by performing the full analysis again, but with a different choice of Nth-Hit(max) = 2 and Nth-Hit(max) = 4. The biggest difference between the best Nth-Hit(max) dependent gv_{ch} result and the nominal result is used as the systematic uncertainty: $\Delta gv_{ch} = \pm 0.006\,\mathrm{ns\,m^{-1}}$.

- N_h energy cut: The energy cut is applied to select the data ^{54}Mn, ^{40}K events in such a way that the contribution of non-γ events is negligible. This is estimated now by varying the energy cut by $\Delta N_h = \pm 10$ around the nominal low and high N_h cut values. The full analysis is then performed again and the biggest difference of the gv_{ch} results compared to the nominal N_h selection is used as the systematic uncertainty. The value is $\Delta gv_{ch} = \pm 0.004\,\mathrm{ns\,m^{-1}}$.

Thus, the final result on the calibration of the effective Cherenkov group velocity correction is $gv_{ch} = (0.108 \pm 0.006(\mathrm{stat.}) \pm 0.038(\mathrm{syst.}))\,\mathrm{ns\,m^{-1}}$, where the systematic uncertainty is dominated by the direction mis-reconstruction between the data and MC.

5.6 Conclusion and outlook

In this Chapter the group velocity of Cherenkov photons has been calibrated using the parameter gv_{ch}. This effective Cherenkov group velocity correction is applied on the lowest level of the G4BX2 simulation, changing the Cherenkov hit time on the PMTs before performing the electronics simulation and before the event reconstruction. The value of this parameter is calibrated trough the use of the ^{54}Mn, ^{40}K γ calibration sources, with a best fit result of $gv_{ch} = (0.108 \pm 0.039(\mathrm{stat.} + \mathrm{syst.}))\,\mathrm{ns\,m^{-1}}$. This number can be expressed as an effective change of the refractive index. Assuming an effective refractive index value of $n = 1.6$ for the Cherenkov photons results in a change of $\Delta n_{ch} = 0.032 \pm 0.012$. This corresponds to a

relative correction of about $(2.0 \pm 0.8)\%$ for the effective refractive index of the Cherenkov hits. This seems reasonable, considering that the differences between the data and the MC are measured at about 1% for the effective refractive index of scintillation.

The uncertainty is dominated by the systematic mis-reconstruction of the γ event direction, which results in a relatively high uncertainty of 37%. As expected, the use of γ-sources and their Compton scattered electrons is not optimal for the calibration of the Cherenkov PMT hit time behaviour. Future hybrid detectors, using both scintillation and Cherenkov hits will benefit from using a dedicated electron Cherenkov calibration source also at low energies $\sim 1\,\mathrm{MeV}$. An idea for such a source could be the use of ^{144}Ce-^{144}Pr, as illustrated in the internal document [146]. Preliminary toy-MC studies show, that the knowledge of the initial electron direction could provide a relative uncertainty as low as 1% on the $\mathrm{gv}_{\mathrm{ch}}$ measurement, but this will not be discussed any further here.

One important consideration of this calibration has been its dependence on the full detector response. The γ calibration sources have only been deployed in 2009. This means that the events of Phase 2 and Phase 3 are further removed in time from this calibration than Phase 1. At the time of the publication of [134, 135] it has been not clear if the results of this $\mathrm{gv}_{\mathrm{ch}}$ calibration are also applicable to Phase 2 and Phase 3, because there could exists some unknown change in the detector response over time, which would result in a different $\mathrm{gv}_{\mathrm{ch}}$ value. If this γ calibration would have been performed in 2021, would the best fit $\mathrm{gv}_{\mathrm{ch}}$ result be compatible with the result of 2009? As this question could not be answered at the time of the CID measurement of the ^{7}Be-ν events, it has been decided that this following CID measurement is only performed for Phase 1.

Following the publication of [134, 135] the $\mathrm{gv}_{\mathrm{ch}}$ value for Phase 1 and Phase 2+3 is also measured using the ^{7}Be-ν edge region neutrino events. This is done in the context of the CID measurement of CNO neutrinos. In the last Chapter 7 of this thesis it is found that the $\mathrm{gv}_{\mathrm{ch}}$ value measured with the γ-sources is well in agreement with the $\mathrm{gv}_{\mathrm{ch}}$ calibration using the neutrino events at the ^{7}Be-ν edge region. This means that the initial decision of only using Phase 1 for the CID ^{7}Be-ν measurement has been overly conservative.

Chapter 6

Proof of Principle for the Correlated and Integrated Directionality

Abstract

Currently, the most precise measurements of sub-MeV solar neutrinos are only performed with liquid scintillator (LS) detectors, such as Borexino. Water Cherenkov detectors have a relative high energy threshold and are only able to measure neutrinos above $\sim$3.5 MeV, due to the low light yield of the Cherenkov process in the presence of background. The high light-yield of Borexino makes it possible to perform a precise spectroscopic neutrino measurement due to the correspondingly good energy resolution and a sufficiently low background level. Until now the directional Cherenkov information has been inaccessible at the sub-MeV neutrino energy range. In this chapter the so called *Correlated and Integrated Directionality* (CID) is used to provide a proof-of-principle for the possible use the directional Cherenkov hits even for sub-MeV neutrino events, using the Borexino detector. The CID method exploits the correlation between the early detected Cherenkov PMT hits and the known position of the Sun. Integrating a large number of events makes it possible the infer the number of contributing neutrino events, through the statistical inference on the angular hit distribution. As the Cherenkov hits make up only $< 0.4\%$ of all hits it is necessary to increase the sensitivity by selecting only the early PMT hits of the events. The CID analysis is performed in an energy region from 0.56 MeV to 0.76 MeV around the Compton-like edge of the ^{7}Be neutrinos, for Phase 1 of the Borexino detector. This results in a measurement of the ^{7}Be-ν rate with $R(^7\mathrm{Be})_{\mathrm{CID}} = 51.6^{+13.9}_{-12.5}\,(\mathrm{stat.} + \mathrm{syst.})\,\mathrm{cpd}/100\,\mathrm{t}$. The zero-neutrino hypothesis can be excluded with a $\Delta\chi^2(\mathrm{N}_{\mathrm{solar}-\nu} = 0) = 38.7$, corresponding to a $> 6\sigma$ exclusion. These results are the first directional detection of sub-MeV neutrino events, as well as the first experimental evidence that the sub-dominant Cherenkov information is accessible, even in a monolithic liquid scintillator detector. This provides the experimental proof-of-principle for the potential use of hybrid event detection at low neutrino energies.

As described in Chapter 3, the hybrid event detection, i.e. combining the directional Cherenkov information with the low energy threshold of scintillation, is of interest for a diverse number of neutrino related physics goals. The present chapter provides the experimental proof for the feasibility of the hybrid event detection at sub-MeV neutrino energies, using the high light-yield Borexino detector. This is done with the so called *Correlated and Integrated Directionality* (CID), which makes it possible to access the Cherenkov information even for neutrino events with a very low expected number of Cherenkov hits per individual event. The basic idea behind CID is to correlate the photon hit direction with the know direction of the solar neutrinos. The photon direction can be estimated through the reconstructed event position and the hit PMT position. Thus, each hit is assigned an angular correlation angle

$\cos\alpha$, relative to the Sun. For a large number of events it is then possible to make a statistical inference on the total number of neutrino events in the event integrated $\cos\alpha$ distribution. This is in contrast to the typical use of the Cherenkov hits, where the event direction is reconstructed on an event-by-event basis and then afterwards compared to the position of the Sun. See Chapter 3 for more details.

In Borexino it is possible to trigger sub-MeV neutrino events and to reconstruct the event vertex using the scintillation hits and their hit time distribution. Thus, the events can be selected according to their reconstructed position and energy.. The first Section 6.1 explains this event selection for an advantageous energy region of interest and a fiducial volume, providing a large expected ratio of neutrino signal over background events. This region corresponds to the Compton-like edge of the ^{7}Be-ν events. Next, Section 6.2 explains the MC model used in the measurement of the number of solar neutrino events. While the parameter of interest is the number of solar neutrino events, here it is also important to identify and understand the nuisance parameters that govern the possible shape of the $\cos\alpha$ distribution of the neutrino signal and the background. To maximise the sensitivity of the CID measurement it is necessary to select early hits of the events. This is effectively a time sorting of the event hits, where early hits have a higher probability of coming from the Cherenkov process compared to later hits. This selection is performed in Section 6.3.

In the previous Chapter 4 it has been found out that the data PMTs can show a misbehaviour in their hit time distribution between each other and in comparison to the MC PMT hit time behaviour. For this reason it is necessary to make a selection of PMTs that are in agreement between the data and the MC. This PMT selection is explained in Section 6.4. Then Section 6.5 defines the χ^2 test statistic and provides the measurement on the number of solar neutrino events, based solely on the Cherenkov information from the $\cos\alpha$ histograms. Additionally, supplementary analyses are performed to produce a picture of the Sun as seen with the CID. Section 6.6 investigates the possible sources of systematic uncertainty between the data and the MC model, and gives a conservative estimation on the value of the relevant systematic uncertainties. Following that, Section 6.7 gives the final results of the CID measurement, for the number of solar neutrinos at the ^{7}Be-ν edge, as well as the rate of ^{7}Be neutrinos in Borexino. The last Section 6.8 concludes this chapter with a summary and an outlook.

The results of this analysis and the method of the PMT selection have been cross-checked and reproduced by [redacted] as part of her dissertation [145]. Furthermore, the toy-MC studies that are investigating the potential systematic uncertainty of the CID background [147] have been conducted by [redacted] as part of her dissertation, as well as the selection of the number of $\cos\alpha$ histogram bins [148]. The fiducial volume (FV) and energy region of interest (RoI) have been optimised by [redacted] as part of her master thesis [149], including the calculation of the corresponding event selection efficiency. Her work was based on a previous study performed by [redacted]. This CID measurement of the solar neutrinos at the ^{7}Be-ν edge is published in the associated papers [134, 135].

6.1 Selection of the data events

The principal goal of this analysis is to provide the proof-of-principle for the hybrid detection of low energy neutrinos. For this purpose the data is selected in an energy RoI, for which the expected ratio of neutrino signal to background events is large. The energy region with the largest expected ratio of neutrino events is around $\sim 0.66\,\mathrm{MeV}$. This corresponds to the Compton-like edge of the recoil electrons from the $0.862\,\mathrm{MeV}$ line of the ^{7}Be-ν. The FV and RoI are selected according to the figure of merit (FoM), defined as follows:

$$\mathrm{FoM}(\mathrm{R}, \mathrm{N}_\mathrm{h}^\mathrm{geo}(\mathrm{min}), \mathrm{N}_\mathrm{h}^\mathrm{geo}(\mathrm{max})) = \frac{\mathrm{N}_{\mathrm{solar}-\nu}}{N_{\alpha-\mathrm{back.}} + \sqrt{\mathrm{N}_{\mathrm{solar}-\nu} + N_{\beta-\mathrm{back.}}}} \tag{6.1}$$

Here, $N_{solar-\nu}$ is the expected number of solar neutrino events, while $N_{\beta-back.}$, $N_{\alpha-back.}$ are the expected number of α- and β-like background events. These numbers depend on the selected energy RoI N_h^{geo}(min, max) and the spherical FV with a radius of R. The idea behind the FoM is to maximise the signal over the expected statistical event uncertainty. The α events are treated separately, to incur a stronger penalty than the β-like background. The reason for this is that α events in the RoI come from the decay of ^{210}Po, which is distributed highly non-uniformly inside the inner vessel [83, 89, 150]. This non-uniformity could introduce a systematic difference in the $\cos\alpha$ distribution of the ^{210}Po background, compared to other background species with a uniform position distribution. The main de-selection of ^{210}Po is performed through the application of the MLP parameter > 0.7, with a large α rejection efficiency [107]. The exposure of non-α events is estimated as $99.5\% \pm 1.0\%$, measured through the ^{214}Bi-^{214}Po event coincidence, during the LS purification campaign in 2010-2011 [145].

The number of expected events N_{solar}, $N_{\beta-back.}$, $N_{\alpha-back.}$ is calculated according to the spectroscopic Borexino results of the event rates [83, 86, 68]. To calculate the number of events from the event rate it is necessary to know the event trigger efficiency and the fraction of events that fall within the selected energy region N_h^{geo}(min, max). These value is called ε_E here and it is estimated using the full G4BX2 simulation of the different event species.

Phase	Live time	R [m]	Fiducial mass [t]	N_h^{geo}(min, max)	N_{total}
1	740.7 days	3.3	132.2	(225, 305)	19904
2	1291.5 days	3.3	132.2	(210, 310)	29686
3	1596.5 days	3.0	99.3	(200, 320)	29042

TABLE 6.1: Selected RoI for the CID analysis at the ^{7}Be-ν edge and the number of data events N_{total}.

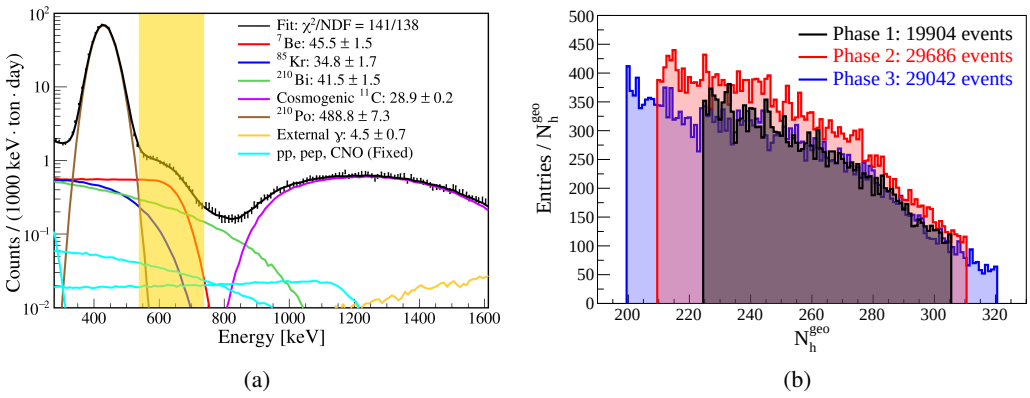
Table 6.1 shows the selected RoI and FV for Phase 1, 2, and 3 separately. The FV radius R decreases for later phases, as the loss of live PMTs over time results in a decrease of the position resolution. This means that for later phases more external γ events are reconstructed inside the FV, which makes it necessary to decrease the value of R. In contrast to this, the energy RoI N_h^{geo}(min, max) increases for later phases, as the internal ^{210}Bi background decreases over time. Please note, that ^{210}Bi has a half life of five days but it is a daughter of, and supplied by ^{210}Pb with a half life of 22.3 years. The fiducial mass is calculated from the spherical FV, using a LS density of $(0.878 \pm 0.004)\,g\,cm^{-3}$ [90].

The next Table 6.2 shows the combined trigger and energy efficiency ε_E for the neutrino signal events of ^{7}Be, pep, and CNO neutrinos. The expected number of neutrino events is given by the product of the neutrino rate, the live time, the fiducial mass, and ε_E. The neutrino rates are calculated with the neutrino fluxes from the low (LZ) and high-metallicity (HZ) B16 standard solar models (SSM) [35, 99]. These values are later used to compare the CID measurement with the independent model predictions. Additionally, N_{SSM}^{pep}+N_{SSM}^{CNO} account for $\sim 12\%$ of the total neutrino signal events, and these values are used to constrain the calculation of the neutrino rate R(^{7}Be-ν) [cpd / 100 t].

The selection of the RoI for Phase 1 is shown in Figure 6.1(a) as a yellow band, in the context of the spectroscopic fit of Phase 1 [83]. The ^{7}Be-ν events comprise the main signal component and the corresponding MC spectrum is shown as a red line. The main background is given by ^{210}Bi, while ^{210}Po is safely excluded. On the right side, Figure 6.1(b) shows the N_h^{geo} spectrum for Phase 1 in black, Phase 2 in red and Phase 3 in blue. The distinct Compton-like shoulder can be seen in all phases.

Phase	$\varepsilon_E^{^7Be}$	ε_E^{pep}	ε_E^{CNO}		$N_{SSM}^{^7Be}$	N_{SSM}^{pep}	N_{SSM}^{CNO}
1	0.193	0.158	0.147	HZ	9059 ± 530	423 ± 6	705 ± 112
				LZ	8264 ± 473	429 ± 6	505 ± 75
2	0.245	0.198	0.189	HZ	20001 ± 1178	929 ± 14	1588 ± 251
				LZ	18256 ± 1073	943 ± 14	1136 ± 168
3	0.278	0.237	0.225	HZ	21145 ± 1178	1028 ± 15	1754 ± 278
				LZ	19299 ± 1135	1043 ± 15	1255 ± 185

TABLE 6.2: Energy efficiency ε_E and SSM expected number of solar-neutrino events in the RoI. The SSM rates are R(^{7}Be) = {47.90 ± 2.82 (HZ), 43.72 ± 2.57 (LZ)} cpd / 100 t, R(CNO) = {4.92 ± 0.78 (HZ), 3.52 ± 0.52 (LZ)} cpd / 100 t, R(pep) = { 2.74 ± 0.04 (HZ), 2.78 ± 0.04 (LZ)} cpd / 100 t.



(a) (b)

FIGURE 6.1: (a) The spectral fit of Phase 1, taken from [83]. The data is shown in black, while the best fit solar-neutrino and background contributions are shown as coloured lines. The RoI selected for the CID analysis is shown as a yellow band around the ^{7}Be-ν shoulder. (b) The N_h^{geo} distribution for the selected data events from Phase 1 (black), Phase 2 (red) and Phase 3 (blue), following the selection criteria of Table 6.1. The ^{7}Be-ν shoulder is visible for all phases.

The primary CID measurement of the number of neutrino events N_{solar} is performed only on Phase 1. The reason for this is that the CID measurement here has been intended as proof-of-principle for the hybrid detection method, and as such a very conservative approach has been adopted for this analysis. The important parameter of the *effective Cherenkov group velocity correction* gv_{ch} is measured in the previous Chapter 5, using the γ-sources from the calibration campaign. The γ-sources have been deployed in 2009, which is closest in time to Phase 1. The measurement of gv_{ch} suffers from a relatively large systematic uncertainty of $\sim 36\%$. At the time of this analysis it has not been clear if the gv_{ch} value could be assumed to be stable over the full detector live time, or what undiscovered systematic differences between the data and the MC could contribute to its measurement. If the gv_{ch} measurement would have been performed again in 2021, would the results be compatible with that of 2009, or would a difference in the detector response introduce a difference in the measured gv_{ch} value? Therefore, only Phase 1 is used in the current chapter for the measurement of N_{solar} and subsequently the corresponding rate R(^{7}Be-ν) [cpd / 100 t]. The combined CID $\cos\alpha$ histograms of Phase 1+2+3 are instead used for the signal model independent rejection of the pure background hypothesis, without performing a fit on N_{solar}.

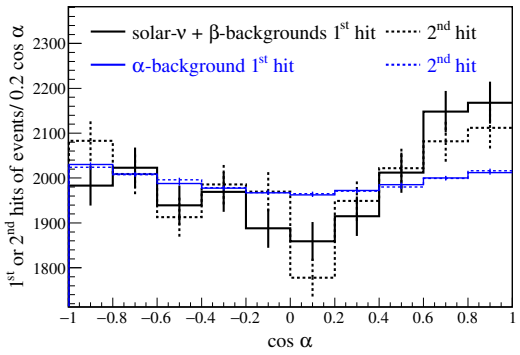


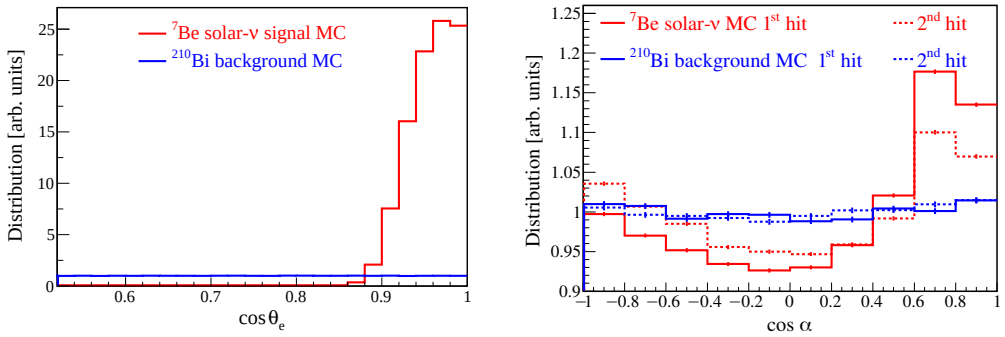
FIGURE 6.2: The CID $\cos\alpha$ distribution of Phase 1, for the first (solid line) and second (dotted line) hits of the events. The 19904 events selected in the RoI (black) are made up out of neutrino signal events and β-background events. For comparison 1.8 million MLP tagged ^{210}Po events (blue) are also shown, normalised to 19904 events. Used in [135].

Figure 6.2 provides the first look at the Phase 1 CID data $\cos\alpha$ histogram. The solid lines show the first hits of the events (Nth-Hit=1), while the dotted lines show the second hits of the events (Nth-Hit = 2). The black histograms show events from the selected energy region of $N_h^{geo}(\text{min, max}) = (225, 305)$, which are composed of neutrino signal events and ^{210}Bi (electron) background. The blue histograms shows MLP selected ^{210}Po-α events around the peak at $N_h^{geo} = 160$ (0.4 MeV), normalised to the statistic of the nominal CID data. Starting from $\cos\alpha > 0.2$ a clear peak can be seen for the neutrino rich RoI, for both early hits. Additionally, there is a negative slope for $-1.0 < \cos\alpha < 0.0$. In contrast, the almost pure ^{210}Po event selection provides a relatively flat $\cos\alpha$ distribution and shows no Cherenkov peak, as expected. This α-background $\cos\alpha$ distribution is not perfectly flat, but rather follows a quadratic function, with a minimum around $\cos\alpha \approx 0.1$. This first $\cos\alpha$ plot already shows that the neutrino events do in fact include Cherenkov information, and they are clearly distinguishable from the background events.

6.2 Neutrino signal MC and background MC

The CID analysis is used to perform a statistical inference on the number of neutrino events contributing to the data $\cos\alpha$ distribution. This is done through the use of a MC model, using the G4BX2 simulation, as explained in Chapter 3.2. It is necessary to estimate the average number of detected Cherenkov hits per neutrino signal event, as the $\cos\alpha$ distribution is produced from the individual hit information, for a given Nth-Hit. This is done here through the use of the *effective Cherenkov group velocity correction* gv_{ch} in the MC, which is measured in Chapter 5. Furthermore, the impact of the live PMT position distribution and solar direction

needs to be taken into account to produce the correct shape of the signal and background MC $\cos\alpha$ distributions. This is accomplished by performing the MC simulation on an event-by-event basis. For each individual data event 200 MC events are simulated for both the ^{7}Be-ν signal and the ^{210}Bi background, within a 15 cm sphere around the reconstructed data event position. The known position of the Sun in the data is used to simulate the correct ^{7}Be-ν MC event direction, while the recoil-electrons in the LS are scattered in an angle $\cos\theta_e$ according to Equation 3.1. The ^{210}Bi background MC event direction is sampled randomly, as it is uncorrelated to the Sun.



(a) (b)

FIGURE 6.3: (a) The angular distribution $\cos\theta_e$ between the solar neutrino directions and the simulated recoil-electrons, from Equation 3.1. The ^{7}Be-ν MC (red) shows that the recoil electrons are scattered in forward direction for the selected Phase 1 RoI of $225 < N_h^{geo} < 305$. The direction of the ^{210}Bi background MC electrons (blue) are not correlated to the solar direction. (b) The CID $\cos\alpha$ distribution for the first (solid line) and second (dotted line) hits of the events. The ^{7}Be-ν signal MC (red) shows a clear Cherenkov peak at $\cos\alpha \approx 0.7$, which is decreased for the second hits. The ^{210}Bi background MC (blue) has a relative flat distribution and no difference between the first and second hits of the events can be seen. Used in [135].

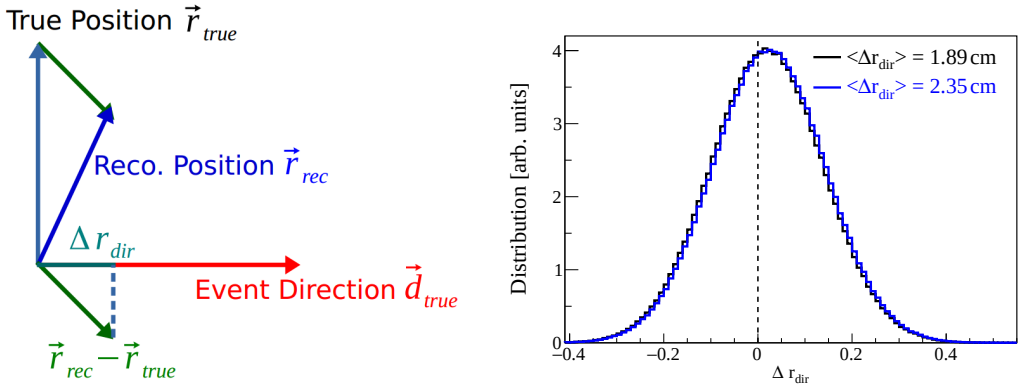
Figure 6.3(a) shows the MC $\cos\theta_e$ angular distribution between the simulated solar direction and recoil-electron direction in red. It can be seen that the electrons are scattered in forward direction for the RoI, as is expected for these events around the Compton-like ^{7}Be-ν shoulder. The background MC is shown in blue and it is uncorrelated to the solar direction. Figure 6.3(b) on the right shows the CID $\cos\alpha$ distribution of the ^{7}Be-ν signal MC in red and for the ^{210}Bi background MC in blue. The first hits of the events (Nth-Hit=1) are shown as a solid line, while the second hits of the events (Nth-Hit=2) are shown as a dotted line. The background MC $\cos\alpha$ distribution is looking flat, and there is no difference between the first and second hits of the events. The signal Cherenkov peak around $\cos\alpha \approx 0.7$ is clearly visible and corresponds to the expectation of the Cherenkov cone opening angle for electrons with $T \approx 0.66$ MeV. Likewise, the Nth-Hit=1 signal $\cos\alpha$ distribution also follows the expectation, as it has a more pronounced Cherenkov peak than the Nth-Hit=2 $\cos\alpha$ distribution. Another notable feature is the negative slope for $\cos\alpha < 0$, which comes from a bias in the position reconstruction, similar to, but less severe than the observation discussed in Chapter 5. This biased position mis-reconstruction is explained in the following section.

6.2.1 MC model nuisance parameters

The primary value of interest is the number of solar neutrino events $N_{\text{solar}-\nu}$. From Figure 6.3 it can be seen that the neutrino signal and the background $\cos\alpha$ distributions are clearly distinguishable. This means that the data $\cos\alpha$ distribution can be fitted with the signal and background MC to estimate $N_{\text{solar}-\nu}$. Furthermore, it has been identified that the signal

$\cos\alpha$ distribution are influenced by two additional parameters. The first parameter is the effective Cherenkov group velocity correction $\mathrm{gv_{ch}}$, defined in Equation 5.2. It governs the Cherenkov to scintillation hit ratio of the early PMT hits. This parameter is measured as $\mathrm{gv_{ch}} = (0.108 \pm 0.039)\,\mathrm{ns\,m^{-1}}$, using the γ calibration sources in Chapter 5. The second parameter is the mis-reconstruction in the initial electron direction Δr_{dir}. The definition of which is given by the (initial) electron direction $\vec{d}_{\mathrm{true}}$, and the true position of the neutrino interaction (energy deposition) $\vec{r}_{\mathrm{true}}$ in respect to the reconstructed event position $\vec{r}_{\mathrm{rec}}$:

$$\Delta r_{\mathrm{dir}} = (\vec{r}_{\mathrm{rec}} - \vec{r}_{\mathrm{true}}) \cdot \vec{d}_{\mathrm{true}} \tag{6.2}$$



(a) (b)

FIGURE 6.4: (a) Schematic definition of the mis-reconstruction in the initial electron direction Δr_{dir}. The reconstructed event position $\vec{r}_{\mathrm{rec}}$ has a small bias to be shifted away from the true event position (position of energy deposit) $\vec{r}_{\mathrm{true}}$, following the electron direction $\vec{d}_{\mathrm{true}}$. (b) The Δr_{dir} distribution for a ^{7}Be-ν MC with a simulated $\mathrm{gv_{ch}} = 0.108\,\mathrm{ns\,m^{-1}}$ value results in an average offset of $\langle \Delta r_{\mathrm{dir}} \rangle = 1.89\,\mathrm{cm}$ (black). A value of $\mathrm{gv_{ch}} = 0.228\,\mathrm{ns\,m^{-1}}$ results in an average offset of $\langle \Delta r_{\mathrm{dir}} \rangle = 2.35\,\mathrm{cm}$ (blue). A larger Cherenkov to scintillation hit ratio at early times increases the Δr_{dir} offset. The standard deviation of both distributions is 11.9 cm.

The definition of Δr_{dir} is schematically illustrated in Figure 6.4(a). Figure 6.4(b) on the right shows the distribution of Δr_{dir} for the ^{7}Be-ν MC. The average offset depends on the implemented value of $\mathrm{gv_{ch}}$ in the MC. The MC with a nominal value of $\mathrm{gv_{ch}} = 0.108\,\mathrm{ns\,m^{-1}}$ is shown in black and has an average offset of $\langle \Delta r_{\mathrm{dir}} \rangle = 1.89\,\mathrm{cm}$, while the ^{7}Be-$\nu$ MC simulated with a value of $\mathrm{gv_{ch}} = 0.228\,\mathrm{ns\,m^{-1}}$ results in a $\langle \Delta r_{\mathrm{dir}} \rangle = 2.35\,\mathrm{cm}$, shown in blue. Likewise a simulation with $\mathrm{gv_{ch}} = -0.012\,\mathrm{ns\,m^{-1}}$ results in $\langle \Delta r_{\mathrm{dir}} \rangle = 1.50\,\mathrm{cm}$, which is not shown.

This means that the early Cherenkov hits tend to pull the reconstructed position towards their own direction. The greater the fraction of early Cherenkov hits, the stronger this effect is. This means that this position reconstruction bias is consistently correlated to the initial solar neutrino direction, because the Cherenkov direction is correlated to the initial direction of the solar neutrinos. The Δr_{dir} offset is small compared to the standard deviation of $\sim 12\,\mathrm{cm}$ and is thus likely negligible for single events, but this effect becomes important after the integration of a large number of events.

The value of $\langle \Delta r_{\mathrm{dir}} \rangle$ can be different between the data and the MC. First, because they could have a different, underlying $\mathrm{gv_{ch}}$ value between them. Second, the $\langle \Delta r_{\mathrm{dir}} \rangle$ can be different because the position reconstruction performs slightly different between the data and the MC, as has been observed in Chapter 5.4. This is also conveyed in Figure 6.5, where the PMT hit time distribution of the data events in the Phase 1 RoI is plotted in black, in comparison to the corresponding MC events in red. The ^{210}Bi MC and ^{7}Be-ν MC exhibit hit time distributions that are fully in agreement, given the statistics of the MC, which is why only

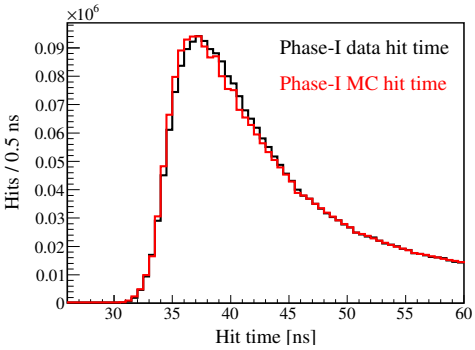
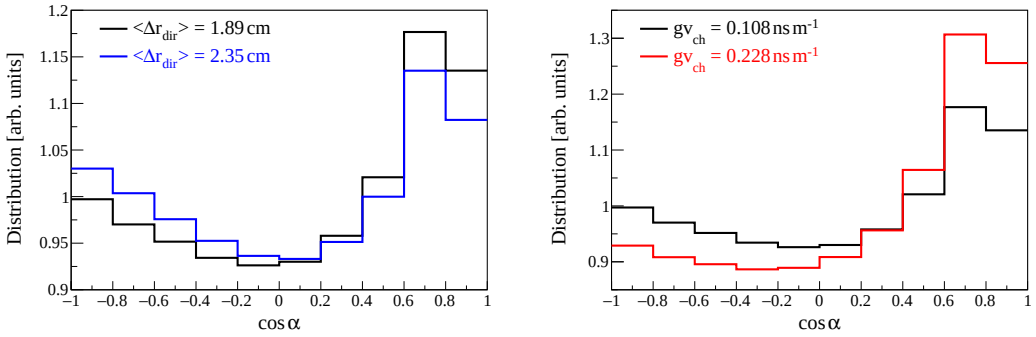


FIGURE 6.5: ToF corrected hit times for the Phase 1 RoI of the data (black) and the ^{7}Be-ν MC (red), normalised to the data statistics. Both ^{7}Be-ν MC and ^{210}Bi MC are comparable with each other, given their statistics and only the ^{7}Be-ν MC is shown. The data and MC distributions are significantly different from each other. Used in [135].

the ^{7}Be-ν MC is shown. What is important to note is that the data and the MC distributions are significantly different from each other, given the statistics of the data. This means that the data and the MC can have a different value of $\langle \Delta r_{\text{dir}} \rangle$, as the position reconstruction is governed by the underlying event hit time distribution.

Unlike the gv_{ch} parameter, the $\langle \Delta r_{\text{dir}} \rangle$ value cannot be measured for the data. It would require a dedicated electron calibration source with a known event direction, but such a source has not been deployed in Borexino. This means that $\langle \Delta r_{\text{dir}} \rangle$ and gv_{ch} are both nuisance parameters in the context of the measurement of the number of solar neutrino events in Borexino. The effective Cherenkov group velocity correction can be constrained by the measurement of Chapter 5 using the γ calibration sources, while the mis-reconstruction in direction $\langle \Delta r_{\text{dir}} \rangle$ must be left free in the fit.



(a) (b)

FIGURE 6.6: The impact of the nuisance parameters on the ^{7}Be-ν signal $\cos\alpha$ distribution for the first hits of the events (Nth-Hit=1) of Phase 1. (a) The nominal MC with $\text{gv}_{\text{ch}} = 0.108\,\text{ns}\,\text{m}^{-1}$ and $\langle \Delta r_{\text{dir}} \rangle = 1.89\,\text{cm}$ (black) is compared to the MC with $\text{gv}_{\text{ch}} = 0.108\,\text{ns}\,\text{m}^{-1}$ where $\langle \Delta r_{\text{dir}} \rangle$ is set to 2.35 cm (blue) by shifting the reconstructed event position in the MC neutrino direction. The number of Cherenkov hits is the same for both $\cos\alpha$ distributions. The Cherenkov peak is decreased for the larger $\langle \Delta r_{\text{dir}} \rangle$ value, due to the increase in the negative $\cos\alpha$ slope. (b) The nominal MC (black) with $\text{gv}_{\text{ch}} = 0.108\,\text{ns}\,\text{m}^{-1}$ is compared to the MC with $\text{gv}_{\text{ch}} = 0.228\,\text{ns}\,\text{m}^{-1}$, where $\langle \Delta r_{\text{dir}} \rangle$ is set to 1.89 cm (red) by shifting the reconstructed event position in opposition of the neutrino direction. The Cherenkov to scintillation hit ratio for the first hits of the events is increased from 8% to 15%.

Figure 6.6 shows the impact of these two nuisance parameters on the ^{7}Be-ν signal MC, on the CID $\cos\alpha$ distribution of the first hits of the events (Nth-Hit=1). The left plot shows the nominal ^{7}Be-ν MC with $\text{gv}_{\text{ch}} = 0.108\,\text{ns}\,\text{m}^{-1}$ and $\langle \Delta r_{\text{dir}} \rangle = 1.89\,\text{cm}$ in black in comparison to the a modified value of $\langle \Delta r_{\text{dir}} \rangle = 2.35\,\text{cm}$ in blue. This new $\langle \Delta r_{\text{dir}} \rangle$ value is produced from the same $\text{gv}_{\text{ch}} = 0.108\,\text{ns}\,\text{m}^{-1}$ simulation, by shifting each single reconstructed event position in the known neutrino direction by 0.46 cm. The total amount of Cherenkov hits in both $\cos\alpha$

distributions is the same, with a Cherenkov to scintillation hit ratio of 8%. The increase in $\langle \Delta r_{\text{dir}} \rangle$ results in an increase in the negative slope at $\cos\alpha < 0$.

This negative slope is superimposed on the Cherenkov peak at $\cos\alpha \approx 0.7$, which has the consequence that an increase in $\langle \Delta r_{\text{dir}} \rangle$ results in a decrease of the Cherenkov peak without changing the number of Cherenkov hits. It is induced by the non-zero value of $\langle \Delta r_{\text{dir}} \rangle$ due to the necessary application of the ToF correction. The events are reconstructed in direction of the neutrino, which corresponds to PMT hits with $\cos\alpha > 0$. These hit PMTs are now, on average, closer to the reconstructed event position than they should be, which in turn means that their hit times are under-corrected by the photon ToF. Likewise, those PMTs which are opposite to the neutrino direction, i.e. $\cos\alpha < 0$, are over-corrected by their hit photon ToF. The combination of these ToF corrections is then such that PMT hits with the lowest $\cos\alpha$ values have a higher chance to be selected as the early hits of the events (Nth-Hit=1, 2, 3, ...), while the PMT hits with the largest $\cos\alpha$ values have the lowest chance to be selected as the early hits.

The right Figure 6.6(b) shows the $\cos\alpha$ distribution of the first hits of the events for the nominal ^{7}Be-ν MC with $\text{gv}_{\text{ch}} = 0.108\,\text{ns}\,\text{m}^{-1}$ and $\langle \Delta r_{\text{dir}} \rangle = 1.89\,\text{cm}$ in black, in comparison to a simulation with $\text{gv}_{\text{ch}} = 0.228\,\text{ns}\,\text{m}^{-1}$ in red. This latter simulation is also set to $\langle \Delta r_{\text{dir}} \rangle = 1.89\,\text{cm}$ by shifting the reconstructed event direction by $-0.46\,\text{cm}$, to look only at the impact of gv_{ch} on the $\cos\alpha$ distribution. The larger value of gv_{ch} results in a larger Cherenkov peak, due to an increase in the number of Cherenkov hits at Nth-Hit=1. This is as expected, as the gv_{ch} parameter effectively increases the Cherenkov photon velocity relative to that of scintillation light. The Cherenkov to scintillation hit ration is 8% and 15% for the values of $\text{gv}_{\text{ch}} = 0.108\,\text{ns}\,\text{m}^{-1}$ and $\text{gv}_{\text{ch}} = 0.228\,\text{ns}\,\text{m}^{-1}$, respectively.

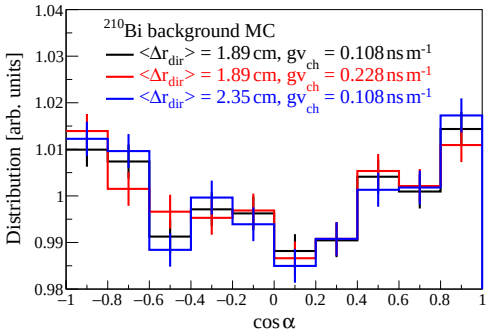


FIGURE 6.7: The ^{210}Bi background $\cos\alpha$ distribution of the first hits of the events for Phase 1 in the selected RoI. The different values of gv_{ch}, $\langle \Delta r_{\text{dir}} \rangle$ result in distributions that are well in agreement with each other, given the statistics of the MC.

Figure 6.7 shows the impact of the nuisance parameters on the ^{210}Bi background $\cos\alpha$ distribution for the first hits of the events. The expectation is that different values of gv_{ch} and $\langle \Delta r_{\text{dir}} \rangle$ should not produce any difference in this $\cos\alpha$ distribution, as the Cherenkov light and initial electron direction of the background are uncorrelated to the position of the Sun. This is the case for the investigated values of $\text{gv}_{\text{ch}} = 0.108\,\text{ns}\,\text{m}^{-1}, 0.228\,\text{ns}\,\text{m}^{-1}$ and $\langle \Delta r_{\text{dir}} \rangle = 2.35\,\text{cm}$ in black, red and blue, respectively. The background MC $\cos\alpha$ distribution is not uniform, but rather follows a parabolic shape, with a minimum around $\cos\alpha \approx 0.1$ and a peak to valley difference of about 2%. This shape is caused by the live PMT position distribution in combination with the non-isotropic distribution of the solar direction. The possible systematic uncertainty from this background shape is studied in Section 6.6 and is found to be negligible.

6.2.2 Linear interpolation of $\cos\alpha$ parameters

For the ^{210}Bi background MC it is enough to simulate a single set of gv_{ch}, $\langle \Delta r_{\text{dir}} \rangle$ parameter values, as the background does not depend on them. In contrast to this, the ^{7}Be-ν signal MC

needs to be produced for different sets of gv_{ch}, $\langle \Delta r_{dir} \rangle$ parameter values. The parameter gv_{ch} is implemented at the lowest level of the G4BX2 simulation, according to Equation 5.2, which makes it necessary to run the full MC simulation multiple times. The value of $\langle \Delta r_{dir} \rangle$ can be changed following the event position reconstruction by shifting every reconstructed ^{7}Be-ν event position by the appropriate value towards the neutrino direction or opposite to it. This means that all possible values of $\langle \Delta r_{dir} \rangle$ can be produced from a single simulation with any gv_{ch} value.

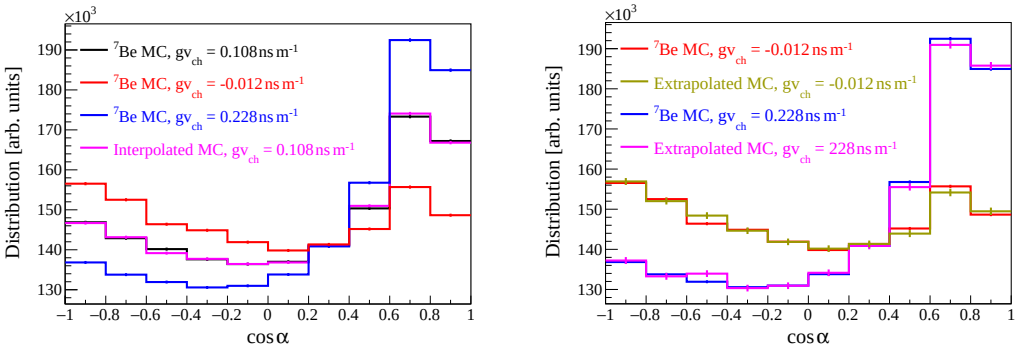


FIGURE 6.8: (a) The linear interpolation of gv_{ch} dependent $\cos\alpha$ histogram of the ^{7}Be-ν signal in the Phase 1 RoI. Using $gv_{ch} = -0.012\,\mathrm{ns\,m^{-1}}$ (red) and $gv_{ch} = 0.228\,\mathrm{ns\,m^{-1}}$ (blue) results in an interpolated $\cos\alpha$ histogram with $gv_{ch} = 0.108\,\mathrm{ns\,m^{-1}}$ (magenta) that is well in agreement with the full MC $gv_{ch} = 0.108\,\mathrm{ns\,m^{-1}}$ $\cos\alpha$ histogram. $\chi^2 = 7.7/9$, p-value $= 0.56$. (b) Linear extrapolation using $gv_{ch} = -0.012\,\mathrm{ns\,m^{-1}}$ (red) and $gv_{ch} = 0.108\,\mathrm{ns\,m^{-1}}$ (see (a)) to produce $gv_{ch} = 0.228\,\mathrm{ns\,m^{-1}}$ (magenta) and using $gv_{ch} = 0.228\,\mathrm{ns\,m^{-1}}$ (blue) and $gv_{ch} = 0.108\,\mathrm{ns\,m^{-1}}$ to produce $gv_{ch} = -0.012\,\mathrm{ns\,m^{-1}}$. The comparison between the extrapolated and full MC $\cos\alpha$ histograms gives a $\chi^2 = 10.3/9$, p-value $= 0.33$.

It is time inefficient to produce the full G4BX2 simulation for a very fine binning of different gv_{ch} values. Thus, the $\cos\alpha$ histograms for different gv_{ch} values are instead interpolated between just two full gv_{ch} simulations. This is shown in Figure 6.8, where the $\cos\alpha$ distribution of the full G4BX2 simulation of the ^{7}Be-ν signal is plotted for three different gv_{ch} values. These are $gv_{ch} = -0.012\,\mathrm{ns\,m^{-1}}$, $gv_{ch} = 0.108\,\mathrm{ns\,m^{-1}}$ and $gv_{ch} = 0.228\,\mathrm{ns\,m^{-1}}$ in red, black and blue, respectively. These values correspond to the $\sim 3\sigma$ level of the measurement with $gv_{ch} = (0.108 \pm 0.039)\,\mathrm{ns\,m^{-1}}$. The magenta line in Figure 6.8(a) shows the interpolation of $\cos\alpha$ with a value of $gv_{ch} = 0.108\,\mathrm{ns\,m^{-1}}$, using the other two histograms. This interpolated histograms is well in agreement with the true $\cos\alpha$ distribution with a $\chi^2 = 7.7/9$, p-value $= 0.56$.

Additionally, the right Figure 6.8(b) shows the extrapolation of the $\cos\alpha$ histogram with $gv_{ch} = -0.012\,\mathrm{ns\,m^{-1}}$ in yellow and $gv_{ch} = 0.228\,\mathrm{ns\,m^{-1}}$ in magenta. The comparison to the full MC $\cos\alpha$ histograms shows again a good agreement with $\chi^2 = 10.3/9$, p-value $= 0.33$, given the statistics of the MC. The conclusion is then that the contribution of Cherenkov hits to the $\cos\alpha$ distribution is well described by a linear interpolation between $gv_{ch} = -0.012\,\mathrm{ns\,m^{-1}}$ and $gv_{ch} = 0.228\,\mathrm{ns\,m^{-1}}$. Therefore, it is sufficient to use these two full gv_{ch} simulation to produce all relevant signal $\cos\alpha$ distributions. This behaviour does not necessarily need to hold true at higher event energies which provide a larger number of Cherenkov hits per event.

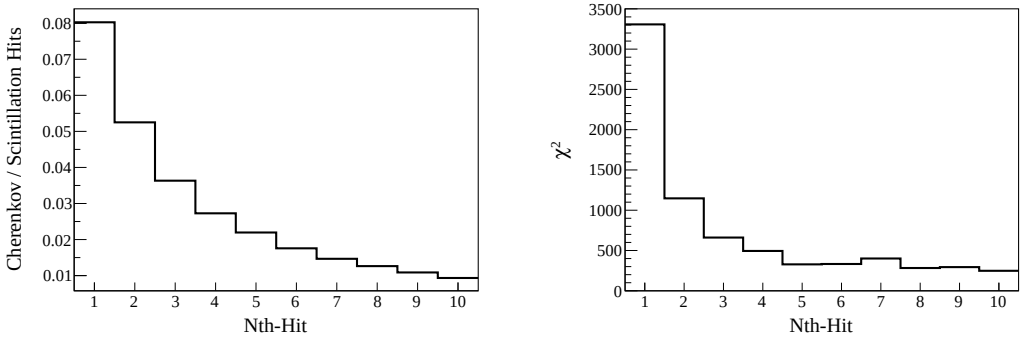


FIGURE 6.9: (a) The Nth-Hit dependence of the Cherenkov to total hit ratio for the ^{7}Be-ν signal MC with a value of $gv_{ch} = 0.108\,ns\,m^{-1}$. The first hits of the events have a maximum of 8%, while the overall Cherenkov to scintillation hit ratio is about 0.34%. (b) The χ^2 test value between the individual Nth-Hit $\cos\alpha$ histograms of the ^{7}Be-ν signal MC and the ^{210}Bi background MC. Larger χ^2 values come from greater differences between the $\cos\alpha$ histograms, which corresponds to a larger expected sensitivity for the CID measurement.

6.3 Selection of the Nth-Hit cut off and number of histogram bins

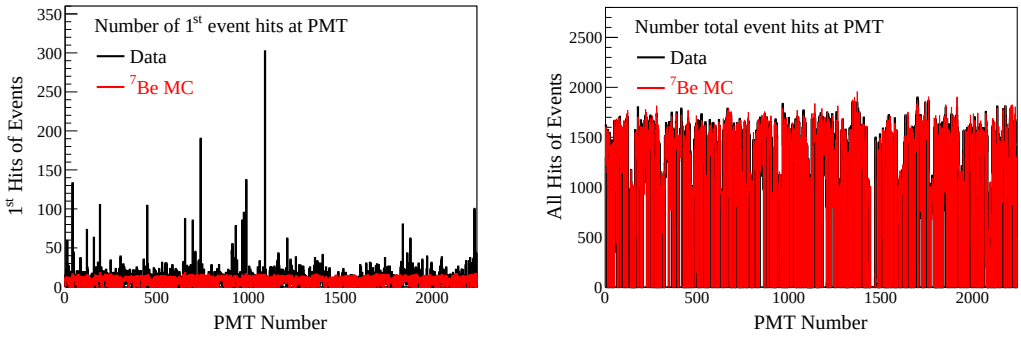
Using the $\cos\alpha$ distribution of the signal and background MC it is now possible to select a value for the Nth-Hit(max) cut off, until which the analysis should be performed. The left Figure 6.9(a) shows the Nth-Hit dependence of the Cherenkov to scintillation hit ratio for a ^{7}Be-ν simulation with a value of $gv_{ch} = 0.108\,ns\,m^{-1}$. The first hits of the events show the largest Cherenkov fraction, which is still sub-dominant to the scintillation, while later hits show a decrease in the ratio.

The right Figure 6.9(b) is a plot of the χ^2 test values between the ^{7}Be-ν MC and ^{210}Bi background MC $\cos\alpha$ distribution for a given Nth-Hit. The $\cos\alpha$ histograms have a binning of 60 and the χ^2 value is calculated using the statistical uncertainty of the MC $\cos\alpha$ bins. It can be seen again, that the first hits of the events have the greatest difference between the $\cos\alpha$ histograms, while later Nth-Hits have a decline that is much steeper, relative to the decline in the Cherenkov to scintillation hit ratio on the left. Starting from Nth-Hit $= 5$ the χ^2 values are flat. A larger χ^2 value means that the signal and background $\cos\alpha$ distributions have a greater discriminatory power between them, which corresponds to a larger expected sensitivity in the CID analysis. Thus, the Nth-Hit until which the CID analysis is performed is chosen as Nth-Hit(max) $= 2$, while the values of Nth-Hit(max) $= 3$, Nth-Hit(max) $= 4$ are used to estimate the systematic uncertainty that could stem from this particular choice in Section 6.6.

The $\cos\alpha$ histogram binning is selected according to a similar argument, based on a toy-MC study [148]. For an injected number of solar neutrino events the estimated uncertainty of the analysis results shows a decrease until a number of 20 bins, followed by a stable value of the expected fit uncertainty. The number of 60 histograms bins is then selected for the CID analysis and this somewhat arbitrary choice is varied in Section 6.6 to estimate a potential systematic uncertainty.

6.4 Selection of good PMTs

The last selection that needs to be performed for the CID analysis is the selection of suitable PMTs, that show the same behaviour between the data and the MC. This is illustrated in Figure 6.10(a), on the left, where each PMT is plotted against the number of times it detected



(a) (b)

FIGURE 6.10: (a) The number of the first hits of the events (Nth-Hit = 1) at the different PMTs. The data (black) shows an over-selection of some PMTs, while the ^{7}Be-ν MC (red), normalised to the data statistics, shows a relatively flat distribution. (b) The number of total hits at the different PMTs. The data (red) and the normalised ^{7}Be-ν MC (red) are in agreement with each other. Different PMTs detect a different number of hits in the distributions due to different quantum efficiencies as well as the presence or absence of the light collecting cone. Both these effects are well reproduced by the MC.

the the first hits of the events. The data of Phase 1 in the RoI is shown in black, while the corresponding ^{7}Be MC is shown in red, normalised to the 19904 data events. It can be seen that some PMTs are grossly over represented in the data, while the MC shows a relative flat distribution. The right Figure 6.10(b) shows the same data and MC, but now for all PMT hits. The data and MC are well in agreement, with a visible structure for the number of detected PMT hits. The difference between different PMTs comes from the difference in the PMT quantum efficiency, as well as the impact of the light collecting cone or its absence. Both effects are well reproduced in the MC compared to the data.

This means that the over- and under-selection of the first hits of the PMTs for certain PMTs is coming from a time mis-alignment of individual PMTs in the data. This is a problem, because it influences the cos α distribution shape of the early hits independent of the Cherenkov to scintillation hit ratio and the bias of the position reconstruction. Some early PMT hits in the data are selected not because they are hit by Cherenkov photons, but rather because the PMTs have an intrinsic time misbehaviour. This has been already observed in Chapter 4, were some PMTs have a run-by-run misbehaviour in their hit time distribution, while all PMTs have a small, constant time offset between their hit time distributions. Therefore, these constant time offsets C_i between the individual PMTs are corrected in the data, using the results of the C_i measurement with the ^{214}Po calibration source from Chapter 4.5.3. These corrections alone are not enough to remedy the PMT misbehaviour of the data, as some PMTs remain over-represented in their number of first hits, even after the time offset correction. This is not unexpected, because the calibration campaign has been performed in a relative short time window between 2008-2009, while Phase 1 ran from 2007-2010. It is possible, that the time behaviour of some PMTs can degrade or change over time and some PMTs that have been alive before 2008 have already been dead by the time of the calibration. Thus, the results of the ^{214}Po calibration are not sufficient to correct the PMT time misbehaviour for the entirety of Phase 1.

To correct the systematic difference between the early PMT hits of the data and the MC only a sub-set of all live PMTs is selected to be used in the CID analysis. These so called "good PMTs" are selected here based on a statistical argument. First, the data PMTs are corrected with their measured constant hit time offsets C_i and the PMTs with the worst run-by-run hit time mis-alignment are deselected in both the data and the MC, following the results of

Chapter 4 (see Figure A.1). Then, the same distribution of the number of first hits at each PMT as in Figure 6.10(a) are used to compare each individual PMT between the data and the MC, normalised to the data statistic. If the number of first hits is in a statistical agreement between the data and the MC, than the PMT is selected to be used in the CID analysis. This "statistical agreement" is defined as the two-sided Poissonian 68.3% confidence interval (CI), where the Poissonian distribution is used because the expected number of first hits per PMT is at about ~ 12. The calculation of the CI follows from the relationship between the Poissonian and χ^2 distribution function, as described for example in [151] (using the "usual" intervals). For a coverage probability of $(1-\alpha)$ $(= 68.3\%)$ and an observed number of first hits x at the PMT the resulting interval is then defined as:

$$\frac{1}{2}\chi^2_{[\mathrm{ndf}=2x],\ (\alpha/2)} \quad \text{to} \quad \frac{1}{2}\chi^2_{[\mathrm{ndf}=2x+2],\ (1-\alpha/2)} \tag{6.3}$$

Equation 6.3 can be solved numerically or by using the inverse of the cumulative function (χ^2 quantiles). The PMTs are used in the CID analysis only if the intervals of the data and the MC are overlapping. PMTs that do not fulfill this condition are deselected in both the data and the signal and background MC for the production of the respective $\cos\alpha$ histograms. The same PMT selection procedure has also been performed in Chapter 5.

6.5 Measurement of the number of solar-neutrino events

The χ^2 test statistics depends on the number of solar neutrinos $\mathrm{N}_{\mathrm{solar}-\nu}$, the mis-reconstruction in direction Δr_{dir}, the Cherenkov group velocity correction $\mathrm{gv}_{\mathrm{ch}}$ and is defined as follows:

$$\chi^2(\mathrm{N}_{\mathrm{solar}-\nu}, \mathrm{gv}_{\mathrm{ch}}, \Delta r_{\mathrm{dir}}) = \sum_{n=1}^{2}\sum_{i=1}^{60}\left(\frac{(\mathrm{norm}\cdot M_i^n - D_i^n)^2}{\mathrm{norm}\cdot M_i^n + \mathrm{norm}^2\cdot M_i^n}\right) + \left(\frac{\mathrm{gv}_{\mathrm{ch}} - 0.108\,\mathrm{ns\,m^{-1}}}{0.039\,\mathrm{ns\,m^{-1}}}\right)^2$$

$$\chi^2(\mathrm{N}_{\mathrm{solar}-\nu}) = \underset{\mathrm{gv}_{\mathrm{ch}},\Delta r_{\mathrm{dir}}}{\arg\min}\left[\chi^2_{\mathrm{gv}_{\mathrm{ch}}}(\mathrm{N}_{\mathrm{solar}-\nu}, \mathrm{gv}_{\mathrm{ch}}, \Delta r_{\mathrm{dir}})\right] \tag{6.4}$$

Here, D_i^n represents the number of entries in the $\cos\alpha$ histogram for Nth-Hit=n and at bin i. The number of entries in the MC M_i^n depends on $\mathrm{gv}_{\mathrm{ch}}$, Δr_{dir} and is normalised to the statistics of the data with "norm". The term "$\mathrm{norm}^2\cdot M_i^n$" in the denominator describes the uncertainty from the finite statistics of the MC simulation. The number of early hits to use for the analysis is selected as Nth-Hit(max) $= 2$, while the number of histogram bins is 60. The right part of the equation is a pull-term constraint on $\mathrm{gv}_{\mathrm{ch}}$, given the results of Chapter 5. The Δr_{dir} cannot be measured in Borexino and is left free from zero to 4 cm, which is large enough such that a further increase has no impact on the χ^2 profile. A negative Δr_{dir} value is excluded, as it would mean that the data event position is reconstructed with a consistent bias in the opposite direction of the Cherenkov hits. This is in opposition to the principle behaviour of the position reconstruction algorithm. Only $\mathrm{N}_{\mathrm{solar}-\nu}$ is of interest for the CID analysis, which is why $\mathrm{gv}_{\mathrm{ch}}$, Δr_{dir} are treated as nuisance parameters. This means that the three dimensional χ^2 profile is minimised over $\mathrm{gv}_{\mathrm{ch}}$, Δr_{dir} for each given value of $\mathrm{N}_{\mathrm{solar}-\nu}$.

The neutrino signal MC $\cos\alpha$ histogram is produced through linear interpolation of the $\mathrm{gv}_{\mathrm{ch}}$, Δr_{dir} parameters, as described in 6.2.2. The full MC $\cos\alpha$ histogram (M) is then the weighted sum of the ^{7}Be-ν signal MC ($M_{\mathrm{solar}-\nu}$) and the ^{210}Bi background MC ($M_{\mathrm{background}}$), constrained by the total number of selected data events $\mathrm{N}_{\mathrm{total}} = 19904$:

$$M = \frac{\mathrm{N}_{\mathrm{solar}-\nu}}{\mathrm{N}_{\mathrm{total}}}\cdot(M_{\mathrm{solar}-\nu}(\mathrm{gv}_{\mathrm{ch}},\ \Delta r_{\mathrm{dir}})) + \frac{\mathrm{N}_{\mathrm{total}} - \mathrm{N}_{\mathrm{solar}-\nu}}{\mathrm{N}_{\mathrm{total}}}\cdot M_{\mathrm{background}} \tag{6.5}$$

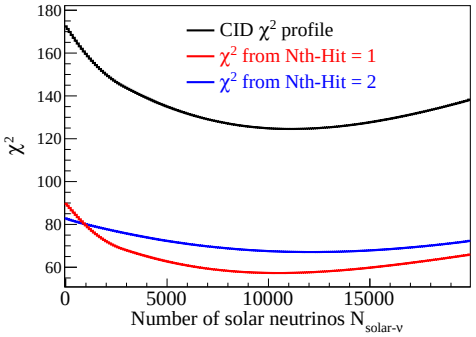
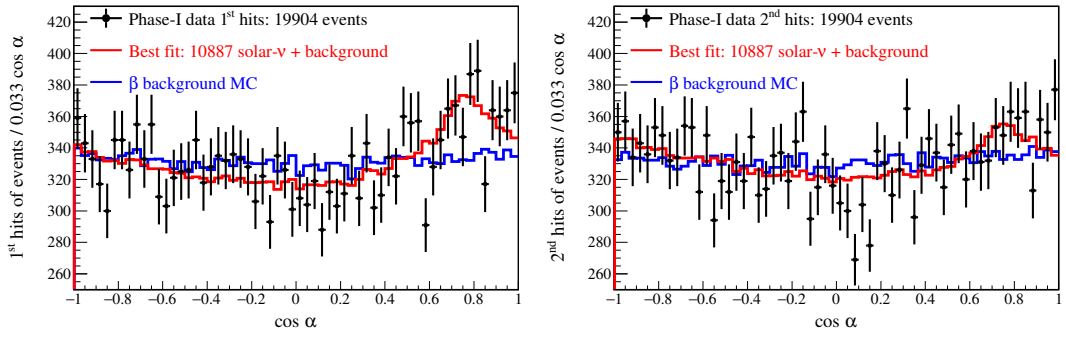


FIGURE 6.11: The CID analysis χ^2 profile (black) in comparison to the individual contribution of Nth-Hit = 1 (red) and Nth-Hit = 2 (blue) $\cos\alpha$ histograms. The number of histograms bins is 60. The performance of the three parameter CID model provides a minimum at $N_{solar-\nu} = 10887$ with $\chi^2/\text{ndf} = 124.6/117$, p-value $= 0.30$, with $\chi^2(\text{Nth-Hit} = 1) = 57.3$, $\chi^2(\text{Nth-Hit} = 2) = 67.3$.

Using this χ^2 test statistics defined in Equation 6.4 results in a best fit value of $N_{solar-\nu} = 10887^{+2386}_{-2103}(\text{stat.})$, with $\chi^2/\text{ndf} = 124.6/117$, p-value $= 0.30$. The statistical uncertainty is estimated from the χ^2 profile trough $\Delta\chi^2 = \pm 1$, corresponding to the Gaussian equivalent 1σ level, i.e. a 68.3% CI. This χ^2 profile can be seen in Figure 6.11 in black, where the individual χ^2 of the $\cos\alpha$ histograms corresponding to Nth-Hit $= 1$, Nth-Hit $= 2$ are shown in red and in blue, respectively. In the absence of additional systematic uncertainties the pure background, zero-neutrino hypothesis can be excluded with a $\Delta\chi^2 = 46.9$. The Nth-Hit $= 2$ χ^2 profile is flatter than the Nth-Hit $= 1$ profile, as is expected from the larger number of Cherenkov hits for Nth-Hit $= 1$. The kink around $N_{solar-\nu} \approx 1800$ comes from the physical boundary of $\Delta r_{dir} \geq 0.0$ cm. For a given, low number of $N_{solar-\nu}$ the fit accommodates the real Cherenkov peak in data by reducing Δr_{dir} and increasing gv_{ch}, i.e. increasing the MC Cherenkov peak by decreasing the negative $\cos\alpha$ slope and increasing the Cherenkov hit per neutrino event ratio. This is only possible until $\Delta r_{dir} = 0.0$ cm reaches zero, and a further decrease in $N_{solar-\nu}$ results in a steeper χ^2 rise.



(a) (b)

FIGURE 6.12: The $\cos\alpha$ distributions of the first (a) and second (b) hits of the data events in Phase 1 (black). This is compared with to the MC model $\cos\alpha$ distributions of the best fit result (red) and the pure background MC (blue). The MC histograms are normalised to the statistics of data. Used in [135].

The best fit results are presented in Figure 6.12, on the left for the $\cos\alpha$ distribution of first hits of the events, and on the right for the second hits of the events. The data is shown in black, with the corresponding statistical uncertainty. The best fit MC $\cos\alpha$ histogram is shown in red, while the pure background MC is shown in blue. The three parameter CID $\cos\alpha$ model seems to able to describe the data well, given the $\chi^2/\text{ndf} = 124.6/117$ value and the appearance of these $\cos\alpha$ distributions. The best fit result of the constrained gv_{ch} parameter is $0.094\,\text{ns}\,\text{m}^{-1}$, which well within the 68.3% confidence interval from the measurement. The relevant systematic uncertainties are estimated in the following Section 6.6 and the calculation of the ^{7}Be-ν rate from the measured $N_{solar-\nu}$ is performed in Section 6.7.

Search for the Sun

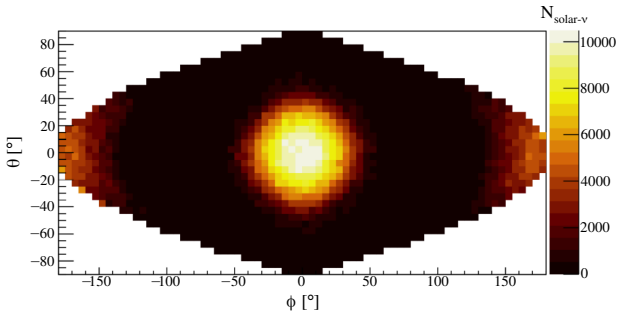


FIGURE 6.13: The CID analysis is performed with different, assumed neutrino directions around the Sun. The plot shows the sinusoidal projection of the best fit results given the assumed neutrino direction, where the true solar direction is defined as $\phi = 0°$, $\theta = 0°$.

The CID analysis works by assuming the known position of the Sun as the origin of the neutrino events. It is now also possible to assume a different origin of the neutrinos and produce equivalent angular distributions $\cos\tilde{\alpha}$. This is presented in Figure 6.13. Here, the true solar direction is defined in spherical coordinates as $(\phi = 0°, \theta = 0°)$ and the sinusoidal projection shows the best fit number of neutrinos $N_{solar-\nu}$ for a given direction (ϕ, θ) around the Sun. The true solar direction shows the largest number of measured neutrino events, while the assumed directions perpendicular to the solar direction have a best fit of $N_{solar-\nu} \approx 0$. The $N_{solar-\nu}$ values at 180° are again non-zero, but significantly smaller than for the true solar direction. This effect comes from the negative slope, induced by the mis-reconstruction in direction Δr_{dir}, which can mimic a small Cherenkov peak. Figure 6.13 is then a picture of the Sun, as seen by the sub-MeV ^{7}Be neutrinos. This is similar to the picture of Super-Kamiokande for ^{8}B neutrinos [152]. While this representation has no additional sensitivity, compared to the nominal CID $\cos\alpha$ histogram analysis, it clearly shows the solar origin of the sub-MeV neutrino events in Borexino.

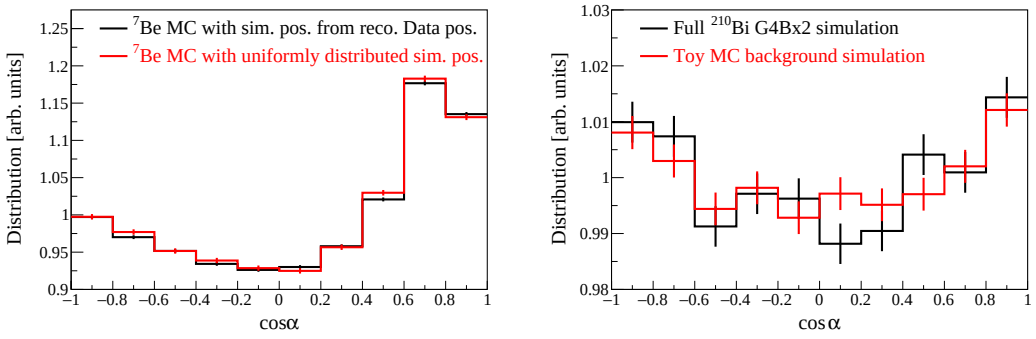
6.6 Investigation of systematic uncertainties

The primary goal of this analysis is the proof-of-principle of the CID method by performing a measurement on sub-MeV ^{7}Be neutrinos. The last step necessary for this measurement is the estimation of the relevant systematic uncertainties, which is performed in this section. As there is not much experience with the CID analysis the quantification of the relevant systematic uncertainties are performed in a simple, but maximally conservative fashion. This can accommodate the potential systematic uncertainties that have not been considered at the time of this analysis, while not endangering the qualitative goal of the analysis. It seems reasonable to err on the side opposing the analysis goal and then still be able to proof the feasibility of the CID method, than to be overly optimistic and to potentially gloss over the result. However, not all of the sources that are examined here have proven to be relevant.

6.6.1 MC production method

The first investigated source of a systematic uncertainty concerns the production method of the MC $\cos\alpha$ histograms for the neutrino signal and the background. The MC is simulated with 200 MC events per data event, using the correct solar direction for the neutrino signal and simulating the MC events in a 15 cm sphere around the reconstructed data event position.

The idea behind the 15 cm sphere is to minimise the potential impact of the event position distribution on the $\cos\alpha$ distribution. The calculation of $\cos\alpha$ depends on the reconstructed event position, as it defines the reconstructed photon direction. This is investigated in Figure 6.14(a), where the nominal ^{7}Be-ν signal MC, in black, is compared to a signal MC simulation, that is simulated with a uniform but random event position, in red. The comparison



(a) (b)

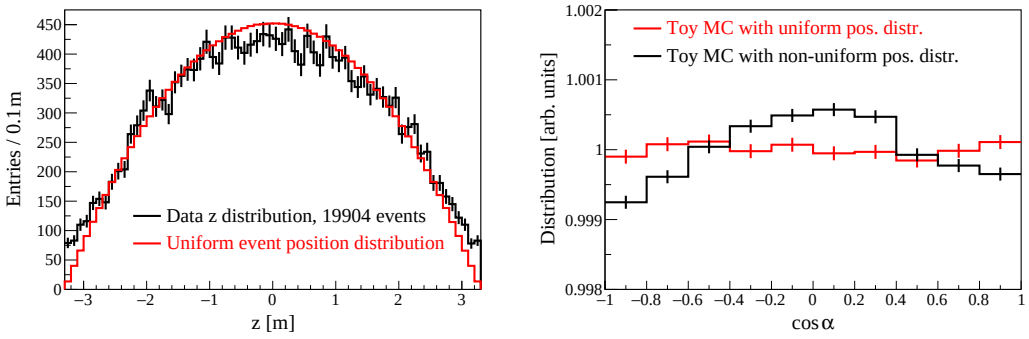
FIGURE 6.14: (a) The ^{7}Be-ν MC $\cos\alpha$ distribution, where the true MC event positions are simulated in a 15 cm sphere around the reconstructed event position of the corresponding data event (black). This is compared to the MC simulation, where the true event position is sampled uniformly in the detector (red). Both distribution are in agreement with $\chi^2/\text{ndf} = 9.8/9$, p-value $= 0.36$. (b) The $\cos\alpha$ distribution of the first hits of the events for the full MC simulation of the ^{210}Bi background (black) in comparison with a toy-MC background simulation (red). The toy-MC takes into account only the live PMT distribution and the solar direction distribution but has no information about the hit timing. Both distributions are well in agreement with $\chi^2/\text{ndf} = 9.4/9$, p-value $= 0.40$.

between the two histograms gives a $\chi^2/\text{ndf} = 9.8/9$, p-value $= 0.36$, given the statistics of the MC. For a given neutrino direction the significantly different event positions still produce an integrated $\cos\alpha$ distribution that is compatible with the more constrained 15 cm sphere method. This means that the CID $\cos\alpha$ distribution is robust, relative to the exact position distribution of individual events.

Another concern is the exact shape of the $\cos\alpha$ distribution, given the distribution of selected PMT positions, and the solar direction distribution. This is investigated in Figure 6.14(b), where the nominal ^{210}Bi background MC, in black, is compared to a toy-MC study of the background, in red. The toy-MC uses the same selected PMT positions as the data and the full MC, and simulates pseudo events uniformly in the detector, with a random photon direction and with a solar direction sampled from Figure 3.2 It has no PMT timing information and it is produced with about 1.5 times the statistic of the full MC. The $\cos\alpha$ distributions of the full MC and the toy-MC are well in agreement with a $\chi^2/\text{ndf} = 9.4/9$, p-value $= 0.40$. This means that the parabolic shape of the background $\cos\alpha$ distribution is produced by the geometric effect of the selected PMT positions, which is well reproduced through the event-by-event simulation approach. The conclusion is then that no systematic uncertainty can be attributed to this MC production procedure.

6.6.2 Non-uniformity of ^{210}Bi background

The next possible source for a systematic uncertainty is the non-uniform distribution of the data background events. This can be seen in Figure 6.15(a), where the reconstructed z-position distribution is shown in black, in comparison to the expected distribution for a uniform event distribution, in red. These distributions are significantly different from each other, giving a $\chi^2/\text{ndf} = 311.6/65$. There is an over-representation of events in the data close to the vessel end-caps at large z-values. The potential impact of this non-uniformity has also been studied with a toy-MC simulation, which is presented in Figure 6.15(b). Here, all simulated photons crossing a radius of 6.5 m are defined as detected, which then corresponds to a full photo coverage of the detector. The solar direction is again sampled according to the true data distribution shown in Figure 3.2 The toy-MC simulation is performed for a uniform position



(a) (b)

FIGURE 6.15: (a) The distribution of the reconstructed event position in z. The Phase 1 data (black) in the RoI is compared to the expected distribution following a uniform event position distribution (red). The distributions are significantly different from each other with $\chi^2/\text{ndf} = 311.6/65$. (b) The $\cos\alpha$ distribution of a toy-MC background simulation with a uniform position distribution (red), compared to the $\cos\alpha$ distribution produced by an event position distribution following the data in (a). A detector coverage of 100% is assumed, i.e. there is no geometric effect from a finite number of PMT positions. Both distributions are simulated with 10^9 photon hits and they are significantly different from each other. The absolute difference is negligible given the statistics of the data.

distribution in red and for the non-uniform distribution seen in the data in Figure 6.15(a), in black. These simulations are performed with a statistic of 10^9 photon hits.

It can be seen that the uniform position distribution now produces a flat $\cos\alpha$ distribution, which agrees with the previous observation, that the finite PMT position distribution is the cause of the parabolic $\cos\alpha$ shape for the background. In contrast, the toy-MC with the non-uniform event position distribution is significantly different from a flat $\cos\alpha$ distribution, but the absolute difference is smaller than 0.1%. This difference is so small, that it is considered safely negligible given the statistics of the MC and the other systematic uncertainties. Please note, that Figure 6.14(b) and Figure 6.15(b) show different background $\cos\alpha$ distributions, as the former uses the live PMT distribution and coverage, while the latter toy-MC assumes a full detector coverage.

6.6.3 Uncertainty of the effective scintillation group velocity

The CID analysis is performed on those PMTs hits, which have the earliest estimated emission time. This photon emission time is estimated trough the ToF correction of the path length between the hit PMT and the reconstructed event position. This ToF correction depends on the effective scintillation group velocity, which is measured as the effective scintillation refractive index $n_g^{\text{eff}}(d)$ in Chapter 4. It is found, that $n_g^{\text{eff}}(d)$ is well described by a linear dependency on the d, due to the attenuation in the LS. The uncertainty $\Delta n_g^{\text{eff}}(d)$ needs to be considered, as it can influences which PMT hits are selected as Nth-Hit $= 1,2$. Varying $n_g^{\text{eff}}(d)$ according to its relatively small uncertainty results in a expected ToF difference of $\Delta t \sim 0.06\,\text{ns}$ at $d = 6.5\,\text{m}$.

At the time of this analysis and its publication in [134, 135] the incorrect, constant n_g^{eff} model has been used for the ToF calculation. This mistake has been found only after the publication of the results in a re-evaluation of the $n_g^{\text{eff}}(d)$ measurement. Using the correct linear model of $n_g^{\text{eff}}(d)$ for the ToF correction results in a best fit value of $N_{\text{solar}-\nu}^{\text{best}} = 11015^{+2412}_{-2127}$. This means that even using the incorrect $n_g^{\text{eff}}(d)$ model introduces only a relatively small systematic uncertainty of the 1.2% compared to the nominal result. The results presented in the next Section 6.7 are the same as the published ones for the sake of compatibility. The best

fit difference is relatively small and the other systematic uncertainties are already estimated maximally conservative to take into account exactly these types of mistakes. As such, an update of the best fit $N_{solar-\nu}$ value compared to the published results [134, 135] would not change the conclusion of this CID analysis.

6.6.4 Selection of Nth-Hit(max)

The selection of Nth-Hit(max) = 2, for the number of early hits to be considered in the CID analysis, could introduce a systematic uncertainty. The Nth-Hit dependence of the Cherenkov to scintillation hit ratio is influenced by the gv_{ch} parameter, which could have unaccounted for systematic differences between the γ-source measurement and the implementation in the G4BX2 neutrino simulation. For this reason the CID analysis is performed two more times, for Nth-Hit(max) = 3,4 and the biggest observed difference is used to estimate the systematic uncertainty. The best fit values are $N^{best}_{solar-\nu}$(Nth-Hit(max) = 2) = 10887, $N^{best}_{solar-\nu}$(Nth-Hit(max) = 3) = 10370, $N^{best}_{solar-\nu}$(Nth-Hit(max) = 4) = 10540. Thus, the systematic uncertainty is conservatively taken as $\pm$517, given the results between Nth-Hit(max) = 2 and Nth-Hit(max) = 3.

6.6.5 Selection of $\cos\alpha$ histogram binning

The analysis is performed on histograms with a number of 60 bins, selected according to the results of a toy-MC study regarding the expected sensitivity. This particular choice is somewhat arbitrary, as all numbers of bins > 20 are expected to perform equally well. This choice could introduce a systematic uncertainty, should there be a unconsidered difference between the data and MC $\cos\alpha$ distributions. For example, there could be some large scale structures in the data or the MC, for which a too fine binning could decrease their potential impact due to low bin statistics. Likewise, should there could be some $\cos\alpha$ outliers due to some undetected effect in the data and those would be smeared out by a too large bin-width. Thus, the CID analysis is performed again with a number of 30, 40 and 120 $\cos\alpha$ histogram bins. Again, the largest difference between these analysis results and the nominal best fit value is used to estimate the systematic uncertainty. The best fit values are $N^{best}_{solar-\nu}$(30 bins) = 10991, $N^{best}_{solar-\nu}$(40 bins) = 11343, $N^{best}_{solar-\nu}$(120 bins) = 10931. Thus, the systematic uncertainty is conservatively taken as $\pm$456, given the results between 40 and 60 bins.

6.6.6 Selection of good PMTs

The data PMTs exhibit a misbehaviour in their hit time distribution, which can over- or under-select early PMT hits, based on the particular hit PMT. This misbehaviour is taken into account for the production of the $\cos\alpha$ histograms, by only selecting PMTs that show a good agreement between the data and the MC. This "good agreement" is defined here as the Poissonian 68.3% (1σ) CI for the number of first hits at the PMTs between the data and the MC. It is clear that this statistical approach is de-selecting a relatively large number of PMTs that do not have any systematic difference between the data and the MC, while there is still a chance that bad PMTs are not de-selected. To estimate the possible systematic uncertainty from the PMT selection method the analysis is performed again, with a less stringent selection of PMTs according to the 95.5% (2σ) CI. The result is $N^{best}_{solar-\nu}$(95.5%CI PMT selection) = 10247, which gives a systematic uncertainty of $\pm$640.

6.6.7 CNO-ν and pep-ν constraint

The next systematic uncertainty concerns the measurement of the ^{7}Be-ν rate in Borexino, in units of [cpd / 100 t]. For this, the measured number of neutrino events $N_{solar-\nu}$ must be

subtracted by the expected number of CNO-ν + pep-ν events. Their B16 SSM [35, 99] expected event numbers are given in Table 6.2. Overall, the CNO-ν + pep-ν events are expected to contribute about $\sim 12\%$ of the total number of neutrino signal events. The CNO-ν + pep-ν events can only be subtracted from $N_{solar-\nu}$ if their MC $\cos\alpha$ distributions have a reasonable agreement to the $\cos\alpha$ distribution of the ^{7}Be-ν MC, which is used as the signal histogram in the CID analysis.

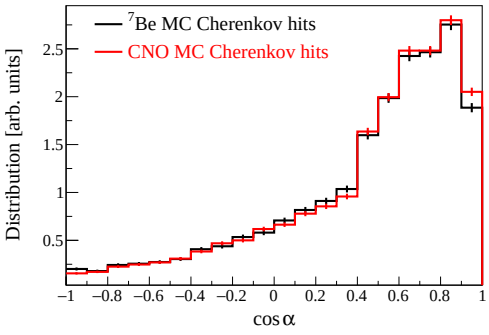


FIGURE 6.16: The $\cos\alpha$ distribution of all Cherenkov hits for the ^{7}Be-ν MC (black) and a CNO-ν MC (red). The latter assumes the most extreme case, for a neutrino energy of $E_\nu = 1.74\,\text{MeV}$ at the endpoint of the CNO spectrum. The histograms are comparable with each other, $\chi^2/\text{ndf} = 24.8/19$, p-value $= 0.16$.

This is evaluated with Figure 6.16, where the nominal ^{7}Be-ν MC ($E_\nu = 0.862\,\text{MeV}$) is compared to a simulated neutrino event corresponding to the endpoint of the CNO-ν with $E_\nu = 1.74\,\text{MeV}$. This corresponds to the most extreme possible difference in the neutrino-electron angle $\cos\theta_e$. Figure 6.16 shows the $\cos\alpha$ distribution of all Cherenkov hits between the ^{7}Be-ν MC in black and the CNO-ν in red. The comparison of these histograms gives a $\chi^2/\text{ndf} = 24.8/19$, p-value $= 0.16$. It is possible that there is a relatively small systematic difference between the neutrino signal $\cos\alpha$ distributions, where the CNO-ν $\cos\alpha$ distribution is shifted to larger $\cos\alpha$ value, compared to the ^{7}Be-ν $\cos\alpha$ distribution. Given the sub-dominant contribution of CNO-ν + pep-ν events and the non-significant $\cos\alpha$ difference between the most extreme neutrino energy differences, it seems reasonable to consider this $\cos\alpha$ shape systematic uncertainty negligible.

Thus, the relevant systematic uncertainty here comes only from the SSM expected number of CNO-ν + pep-ν events. This value is $N^{SSM}_{pep+CNO} = 1128^{+120}_{-230}$, where the HZ model is used as the central value with the corresponding model uncertainties and the difference between the HZ and LZ models is used as an additional systematic uncertainty.

6.6.8 Exposure

The last relevant systematic uncertainty concerns the calculation of the ^{7}Be-ν rate. This calculation is performed by simply multiplying the measured number of $N_{^7Be-\nu}$, after the CNO-ν + pep-ν correction, with the effective exposure f. This effective exposure is the product of the live time, the fiducial mass, the energy efficiency ε_E and the MLP exposure, which has an uncertainty of 1%. The exposure uncertainty is estimated conservatively by using the single largest observed difference in the reconstructed position between the data and MC calibration sources and calculating the fiducial volume accordingly. This difference is 5 cm (see Figure 22 in [85]), which results in a relative uncertainty of 4.6%, given the selected spherical fiducial volume with a radius of $R = 3.3\,\text{m}$.

6.6.9 Summary

In this chapter, the relevant sources of systematic uncertainty have been examined. The contribution of different parameter selections, such as the Nth-Hit(max), the $\cos\alpha$ histogram binning and the PMT selection, has been estimated maximally conservative. It is clear that different parameter choices will result in different best fit outcomes purely due to statistical

Source	Uncertainty [%]
Selection of binning	4.2
Selection of Nth-Hit(max)	4.8
Selection of good PMTs	5.9
Total $\frac{\Delta N_{solar-\nu}}{N_{solar-\nu}}$	8.7

Source	Uncertainty[%]
MLP exposure	1.0
$\frac{\Delta N^{SSM}_{pep+CNO}}{N_{solar-\nu}}$	$^{+1.2}_{-2.3}$
Fiducial mass	4.6
Total $\frac{\Delta R(^7Be)}{R(^7Be)}$	$^{+10.0}_{-10.2}$

TABLE 6.3: Relevant systematic uncertainties for the CID measurement of $N_{solar-\nu}$ and the ^{7}Be-ν rate.

effects, and potentially also due to systematic differences between the data and the MC model. Here, these different best fit outcomes are instead attributed fully to systematic uncertainties of unknown origin. While this method likely overestimates the uncertainty it is at the same time relatively simple to perform. Likewise, the exposure has also been estimated maximally conservative, using the largest possible difference in the FV. This approach does not hinder the principle analysis goal, which is the proof-of-principle of the CID method, and not a maximally sensitive measurement of the ^{7}Be-ν rate. The relative systematic uncertainties are summarised in Table 6.3, relative to the best fit value for the number of neutrino events $N_{solar-\nu}$.

These estimations can be improved by a more thorough study, using toy-MC simulation to disentangle the purely statistical and potential systematic components of the best fit result differences. This is performed and explained in detail for the CID measurement of the CNO-ν rate in Chapter 7.

6.7 Final results

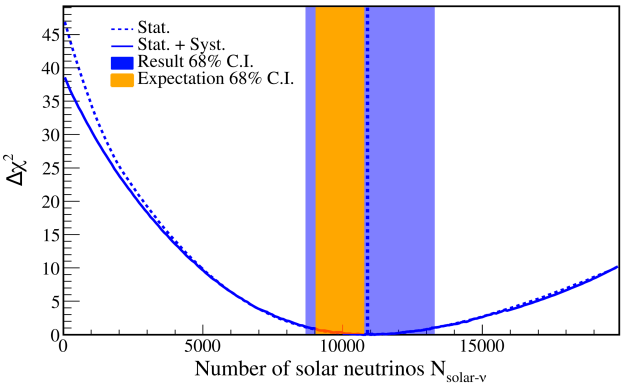


FIGURE 6.17: The $\Delta\chi^2$ profile of the CID analysis, with (solid line) and without (dotted line) the systematic uncertainties. The 68.3% CI (blue band) is given by $N_{solar-\nu} = 10887^{+2386}_{-2103}(stat.) \pm 947(syst.)$. This is in agreement with the SSM expectation of $N^{SSM}_{solar-\nu} = 10187^{+541}_{-1127}$ (orange band). Used in [135, 134].

The final result on the CID measurement of the number of solar neutrino events now includes the systematic uncertainties: $N_{solar-\nu} = 10887^{+2386}_{-2103}(stat.) \pm 947(syst.)$. These systematic uncertainties are included in the measured $\Delta\chi^2(N_{solar-\nu})$ profile through the convolution of the likelihood function $\exp(-\frac{1}{2}\Delta\chi^2)$ with a Gaussian distribution. Its standard deviation is given by the quadratic sum of all relevant systematic uncertainties. The $\Delta\chi^2$ profile, including the systematics, is shown in Figure 6.17 as a solid blue line, in comparison to the $\Delta\chi^2$ profile without systematics as a dotted blue line. The blue band shows the 68.3% CI (1σ) of the CID measurement, while the orange band shows the SSM expected value of $N^{SSM}_{solar-\nu} = 10187^{+541}_{-1127}$. This value uses the HZ model as a central value and the difference of the HZ and LZ models

as an additional model uncertainty. The SSM expectations and the CID measurement are well in agreement with each other, given the relatively large uncertainty of the measurement of -22%, $+24\%$. The exclusion of the zero-neutrino, pure background hypothesis is given by $\Delta\chi^2(\mathrm{N}_{\mathrm{solar}-\nu}=0)=38.7$, p-value $=5\cdot10^{-10}$. This corresponds to an exclusion at the $>6\sigma$ level, i.e. a detection of sub-MeV solar neutrinos with a fit that depends only on the directional information of the Cherenkov hits.

It has to be noted here that the so called statistical uncertainty includes the impact of the nuisance parameters, meaning it is not actually purely defined by the number of data events and PMT hits. Fixing the nuisance parameters to their best fit value results in a $\Delta\chi^2$ profile with a purely statistical uncertainty of $\Delta\mathrm{N}_{\mathrm{solar}-\nu}=\pm1523$. Thus, the impact of the nuisance parameters Δr_{dir}, $\mathrm{gv}_{\mathrm{ch}}$ can be estimated as $(+1837, -1450)$.

Next is the calculation of the ^{7}Be-ν interaction rate in the Borexino detector. This is done by constraining $\mathrm{N}_{\mathrm{solar}-\nu}$ with the number of expected events $\mathrm{N}^{\mathrm{SSM}}_{\mathrm{pep+CNO}}=1128^{+120}_{-230}$ and then scaling the resulting number of $\mathrm{N}_{^7\mathrm{Be}-\nu}$ with the effective exposure $f=189.0\,(100\,\mathrm{t}\times\mathrm{days})$:

$$R(^7\mathrm{Be})=\frac{1}{f}\left(\mathrm{N}_{\mathrm{solar}-\nu}-\mathrm{N}^{\mathrm{SSM}}_{\mathrm{pep+CNO}}\right)=51.6^{+13.9}_{-12.5}\,(\mathrm{stat.}+\mathrm{syst.})\,\frac{\mathrm{cpd}}{100\,\mathrm{t}} \tag{6.6}$$

This rate $R(^7\mathrm{Be})$ corresponds to both the mono-energetic lines of the ^{7}Be neutrinos with 0.862 MeV and 0.384 MeV.

Systematic free exclusion of the zero-neutrino hypothesis

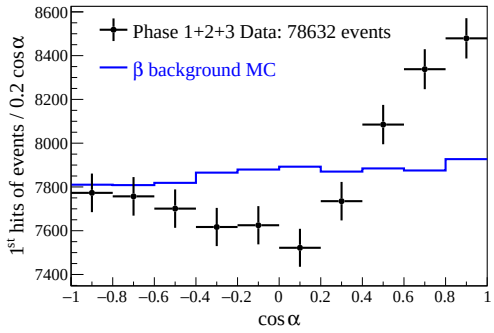


FIGURE 6.18: The $\cos\alpha$ distribution of the full detector live time Phase 1 + Phase 2 + Phase 3 (black), compared with the background MC distribution (blue), normalised to the statistics of data. The distributions are significantly different from each other, with $\chi^2/\mathrm{ndf}=107.0/9$, p-value $=6\cdot10^{-19}$. This plot is updated relative to the published result [135], which used only a part of the Phase 3 data.

The principal CID measurement is performed here only on Phase 1, due to the potential impact of a change in the detector response of the $\mathrm{gv}_{\mathrm{ch}}$ measurement, as well as the measurement of the constant hit time offsets between the different PMTs. These parameters are measured using the calibration sources deployed in relatively short time intervals between 2008-2010. At the time of this analysis it has not been clear how applicable these values could be for Phase 2 and Phase 3. Thus, the $\cos\alpha$ distributions of all phases is used here, not to perform a measurement, but to provide an additional exclusion of the zero-neutrino hypothesis.

This can be seen in Figure 6.18, where the $\cos\alpha$ histogram of the sum of the Phase 1+2+3 events is shown in black, in comparison to the corresponding background MC. The full detector live time of 3628.7 days provides in total 78632 data events. A Cherenkov peak can be clearly seen in the data, albeit at $\cos\alpha\approx0.9$ instead of the expected $\cos\alpha\approx0.7$, likely due to statistics. The comparison between the data and the pure background MC gives a $\chi^2/\mathrm{ndf}=107.0/9$, p-value $=6\cdot10^{-19}$. This means that the background alone is not able to explain the $\cos\alpha$ shape of the data and there must be a contribution of solar-neutrino events. This exclusion of the zero-ν hypothesis is free from the previous systematic uncertainties and does not depend on the three parameter MC model of the CID measurement. It is an additional

reinforcement for the proof of the CID principle, where the directional information of the solar neutrinos can be accessed through the Cherenkov hits, even for sub-MeV neutrinos.

6.8 Conclusion and outlook

The goal of this Chapter is to provide the proof-of-principle for the detection of sub-MeV neutrino events, using their directional Cherenkov hit information in the Borexino LS detector. The analysis is performed on the *Correlated and Integrated Directionality* (CID) $\cos\alpha$ distribution. This distribution is produced by correlating the known position of the Sun with the detected PMT hits for every event. The integration of these angular correlation PMT hit values $\cos\alpha$ over a large number of events then provides a $\cos\alpha$ distribution from which the number of neutrino events can be inferred statistically.

The data events are selected in a beneficial RoI and FV, with a RoI of $225 < N_h^{geo} < 305$, This efficiently selects the Compton-like shoulder of the recoil-electrons from the 0.862 MeV line of ^{7}Be-ν. These borders correspond to a range of 0.56 MeV $< E_e <$ 0.76 MeV, smeared by the finite PMT hit statistic. It has to be noted here that this event selection makes explicit use of the dominant scintillation hits, as the Cherenkov to scintillation hit ratio at these energies is only about 0.34%.

The model of the $\cos\alpha$ distribution is produced with the G4Bx2 MC and uses the two nuisance parameters gv_{ch} and Δr_{dir} in addition to the number of neutrino events. The gv_{ch} can be measured in Borexino through the use of γ calibration sources while Δr_{dir} cannot be measured due to the lack of a dedicated electron Cherenkov calibration source. Consequently, the gv_{ch} can be constrained with a Gaussian pull-term, while Δr_{dir} is left entirely free in the fit of the number of solar neutrino events. These two nuisance parameters are effectively the largest contribution to the systematic uncertainty budget. The other sources of relevant systematic uncertainties are estimated maximally conservative from the biggest difference between the nominal analysis and variations in the Nth-Hit(max), histogram binning and the selection of good PMTs.

The final result on the number of solar neutrinos is then $N_{solar-\nu} = 10887^{+2568}_{-2307}$ (stat.+syst.), which is well in agreement with the SSM expected value of $N^{SSM}_{solar-\nu} = 10187^{+541}_{-1127}$. This results in a rate of $R(^7\text{Be})_{CID} = 51.6^{+13.9}_{-12.5}$ (stat. + syst.) cpd/100 t, after constraining the CNO+pep-ν event number. The CID measured result is well in agreement with the result of the corresponding Phase 1 spectroscopy of $R(^7\text{Be}) = 47.87 \pm 2.28$ cpd/100 t [83][1]. It is clear that the distinct ^{7}Be-ν shoulder can be fitted rather well by the spectroscopic analysis using the scintillation hits, resulting in much smaller uncertainties compared to the CID results. Nonetheless, the CID method also provides a measurement of the ^{7}Be-ν rate, using only the directional Cherenkov information for the fitting. The CID analysis gives a $\Delta\chi^2(N_{solar-\nu} = 0) = 38.7$, corresponding to a $> 6\sigma$ rejection of the zero-ν hypothesis. Thus, the combination of the event selection through the scintillation hits and the subsequent analysis of the directional $\cos\alpha$ Cherenkov information makes this analysis a hybrid detection of the neutrino events. The CID analysis provides the first direct use of the Cherenkov photons from sub-MeV solar neutrinos and this is an additional proof of the solar origin of the signal events in Borexino. The conclusion is then, that it is quite possible to access the directional Cherenkov information, even in a monolithic, high light-yield LS detector, such as Borexino. These CID results are published in [134, 135].

Therefore, the ongoing development of hybrid target materials and detection hardware [119, 120, 124, 122] can be seen in a positive light. All developments that are intended to improve the event-by-event direction reconstruction will likely also improve the CID method. For future experiments it seems reasonable to suggest the use of a dedicated electron

[1] Summing both the 0.862 MeV and 0.384 MeV lines of ^{7}Be-ν

Cherenkov calibration source, in addition to all the other improvements that are necessary for future generation (solar) neutrino detectors. The immediate improvement of the CID analysis in Borexino is performed in the following Chapter 7, where the CNO-ν rate will be measured. In the current chapter only Phase 1 is used for the analysis, which is a deliberate conservatism. The main argument for this has been, that at the time of this analysis it has not been clear how well applicable the measured gv_{ch} pull-term is on Phase 2 and Phase 3, as well as the constant PMT hit time offsets. Both these properties are measured with calibration sources deployed between 2008-2009, i.e. only during Phase 1. The following CID CNO-ν analysis will use the full detector live time and measure these values for each phase separately, using the ^{7}Be-ν edge events, as well as ^{11}C background events.

Chapter 7

Correlated and Integrated Directionality Measurement of CNO-Neutrinos

Abstract

Borexino has provided the first experimental evidence for existence of CNO-cycle neutrinos in the Sun, by performing a multivariate, spectral fit. This has been achieved through the thermal stabilisation of the detector in Phase 3, which allowed for the constraint of the internal ^{210}Bi background. The strong degeneracy between the spectra of the ^{210}Bi background and the CNO-ν signal would otherwise not allow for a sensitive measurement of the latter. The *Correlated and Integrated Directionality* (CID) method now offers a unique crosscheck and complementary measurement to this spectral fit of the CNO-ν rate, where the CID does not depend on a ^{210}Bi constraint and can therefore be performed for the full detector live time. The CID analysis has been further improved in this chapter, compared to the previous proof-of-principle of the hybrid event detection of sub-MeV neutrinos. First, the effective Cherenkov group velocity correction $\mathrm{gv_{ch}}$ is measured trough the CID analysis on the neutrino dominated region around the Compton-like ^{7}Be-ν edge, around $\sim$0.5 MeV–0.8 MeV. For this, the number of solar neutrinos is treated as a nuisance parameter for the measurement of $\mathrm{gv_{ch}}$, where the ^{7}Be-ν and pep-ν neutrinos are constrained according to the Standard Solar Model predictions. The CID measurement of the CNO-neutrinos is then performed in a decoupled energy region around 0.85 MeV–1.3 MeV. Further improvements of the CID sensitivity are provided by an optimised selection of the early PMT hits to use for the analysis, as well as an improved selection of well behaving PMTs. Two separate data sets are selected, Phase 1 and Phase 2+3, to account and investigate for the potential change of the detector response over time. No systematic differences have been found between the CID analyses of Phase 1 and Phase 2+3. The full CID measurement of the CNO-ν rate results in $R_{\mathrm{CNO}} = 7.2^{+2.8}_{-2.7}$ (cpd / 100 t), using the entire detector live time of 3628.7 days, trough the combination of the Phase 1 and Phase 2+3 results. The associated probability for the zero CNO-ν hypothesis is $P(N_{\mathrm{CNO}} = 0) = 7.8 \cdot 10^{-8}$. This corresponds to a detection of the solar CNO-cycle neutrinos with more than 5σ. Therefore, the CID method provides a relatively direct proof of the solar origin of the CNO-ν signal. Future, large scale liquid scintillator or hybrid detectors may be able to further benefit from the CID for the measurement of low energy solar neutrinos by combining the spectral and CID analyses.

Borexino has successfully measured the fluxes of solar neutrinos from the entire pp-chain of fusion reactions [86, 87, 84, 31, 88], with the exception of hep neutrinos. Nonetheless,

the spectral measurement of the solar neutrinos from the CNO-cycle requires an even more attentive treatment of the residual radioactive background. The main hurdle here is the degeneracy between the ^{210}Bi background spectrum and the CNO-ν recoil electron spectrum. Additionally, the endpoint energy region of the CNO-ν spectrum has a large contamination of cosmogenic ^{11}C events. Therefore, the measurement of CNO-neutrinos using the spectral fit is dependent on an external estimation of the ^{210}Bi background [153]. Such an estimation requires a sufficiently high thermal stability of the detector and the understanding of the fluid-dynamical, convective motions inside the detector, as well as certain assumptions about the transport mechanisms of the ^{210}Bi and ^{210}Po daughters of the ^{210}Pb decay chain. For this purpose the detector has been thermally insulated in 2015, to minimise the convective transport of ^{210}Po from the vessel surface to the inside of the FV. The thermal behaviour is continuously monitored and compared to fluid dynamics simulations [154, 155]. With this understanding of the convective transport, the ^{210}Bi rate can be indirectly constrained through its daughter ^{210}Po, using the so called low-polonium field [89].

With these efforts Borexino has managed to provide the first direct experimental evidence for the existence of the CNO fusion cycle inside the Sun, as well as a measurement of the CNO-ν rate [68, 45]. Here, the data taking live time is limited by the thermal stability to 1431.6 days of Phase 3 [45], starting from January 2017. Furthermore, the solar origin of the measured CNO-ν events is proven only indirectly, by the subtraction of the background and pep-ν events from the measured energy spectrum, after which the remainder is in good agreement with the spectrum expected from CNO-ν events.

Naturally, the CID analysis method now offers itself as an interesting cross-check and supplementary measurement of the CNO-ν rate. It can potentially provide a direct proof of the solar origin of the (assumed) CNO-ν events. The CID $\cos\alpha$ shape has a fundamentally smaller degeneracy between the ^{210}Bi background and CNO-ν events. The $\cos\alpha$ shape of all background events originates from the distribution of live PMTs and it is relatively flat, while the signal $\cos\alpha$ shape of all solar neutrino species exhibits a distinct Cherenkov peak around $\cos\alpha \sim 0.7$, given sufficient statistics.

This means that the CID can be used throughout the entire detector data taking time from Phase 1, Phase 2 to Phase 3 and does not depend on an external ^{210}Bi constraint. It needs no thermal insulation or fluid dynamics simulation and the CID does not depend on any assumptions about the convective motions of the different background species. Given enough statistics and a sufficiently high sensitivity, the CID has the potential to provide a measurement of CNO-neutrinos, independent and without the challenges of the multivariate fit.

To achieve this goal, first an advantageous fiducial volume (FV) and energy region of interest (RoI) is selected in Section 7.1, based on the expected number of neutrino events given by the standard spectral fit. Then, the following Section 7.2 explains the production of the MC signal and background $\cos\alpha$ histograms used for the CID analysis, given the selected FV and energy RoI. The previous CID analysis of ^{7}Be-neutrinos provided important lessons, which open up a multitude of avenues for the improvement of the current CID analysis, maximising its sensitivity for the measurement of the CNO-ν rate. The first of which is the optimised selection of the Nth-Hit(max) cutoff, described in Section 7.3. The new selection increases the expected sensitivity by a factor of about ~ 1.5, compared to the selection used for the CID ^{7}Be-ν measurement. Next is the selection of good PMTs to use for the production of the CID $\cos\alpha$ histograms, described in Section 7.4. Here, the PMTs are deselected according to their individual hit time distributions, allowing for a less conservative selection of a greater number of PMTs to use, further improving the sensitivity.

The CID analysis chain is explained in Section 7.5, where Phase 1 and Phase 2+3 are analysed separately to investigate a potential change of the CID detector response over time. First, the effective Cherenkov group velocity correction $\mathrm{gv_{ch}}$ is calibrated on the ^{7}Be-ν edge region, $\sim$0.5 MeV–0.8 MeV. For this, the expected number of neutrino events are now

treated as a nuisance parameter and gv_{ch} is the parameter of interest in the CID analysis. The standard solar models (SSM) B16 [35] are used for the calculation of the expected neutrino rates, with the high metallicity model GS98 (HZ-SSM) and the low metallicity model AGSS09met (LZ-SSM). Then this gv_{ch} measurement is used as a pull-term for the measurement of the CNO+pep-neutrinos in their RoI, decoupled from the ^{7}Be-ν edge region at 0.85 MeV–1.3 MeV. This approach has the advantage that a potential change of the detector response over time is effectively included in the gv_{ch} calibration, while at the same time the gv_{ch} uncertainty is smaller compared to the gv_{ch} calibration using the γ-sources of the calibration campaign, for both Phase 1 and Phase 2+3.

Next is the study of the systematic uncertainties in Section 7.6. Here, toy-MC studies are used in the investigation of the different selection choices made for the analysis and the CID background is examined with events outside of the energy RoI. While the previous CID analysis of ^{7}Be neutrinos is intended as a proof-of-principle and therefore uses maximally conservative estimates of the systematic uncertainties, the emphasis of this chapter is on more reasonable estimates, that require more thorough investigations. Following this train of thought, Section 7.7 examines the fit response itself, where a fitting bias is found and corrected through a Bayesian toy-MC sampling of the posterior. The final results of the CID measurement of the number of solar neutrinos $N_{\mathrm{CNO+pep}}$ and the CNO-ν rate are presented in Section 7.8, including the systematic uncertainties and with the correction of the fit response bias. The last Section 7.9 concludes the CID analysis of CNO-neutrinos with a summary.

The presented CID results, as well as the ROOT scripts used for their production, have been cross-checked and reproduced by [redacted] as part of her dissertation. She also performed the γ background systematic toy-MC study [156] and the calculation of the TFC-cut exposure efficiency. The FV and RoI selection has been primarily performed by [redacted] as part of her master thesis [149] and Section 7.1 is based on her Borexino internal document [157].

7.1 Selection of CNO- and pep-neutrino region of interest

In this section the data event selection is explained, where the first selection is the data taking live time to be used. The CID analysis will be performed independently for Phase 1 and the sum of Phase 2+3. The use of at least two independent CID data sets makes it possible to crosscheck the results, which could potentially help in the investigation of systematic uncertainties stemming from a change of the detector response over time. At the same time, using the sum of Phase 2+3 is expected to maximise the absolute number of CNO+pep-neutrinos, as Phase 2 and Phase 3 have both a longer data taking live time than Phase 1 and a reduced number of ^{210}Bi background events.

The next selection parameter is the cut on ^{11}C events, using the three fold coincidence (TFC) [83, 109] algorithm, where otherwise ^{11}C events could dominate over the ν-signal events. Thus, the main CID analyses in the following sections are performed on ^{11}C TFC-cut data, while the TFC-tagged data is not used directly for the CID analyses. Next, it is necessary to define a RoI which has a favourable number of CNO- and pep-neutrinos $N_{\mathrm{CNO+pep}}$, relative to the total number of events N_{total}. At the same time the CNO+pep-ν region should be decoupled from the previously selected ^{7}Be-ν edge region to avoid potential correlations from the calibration of gv_{ch}, as it is used as a pull-term for the $N_{\mathrm{CNO+pep}}$ measurement. The spherical FV cut on R is limited mostly by the external γ events (^{40}K, ^{214}Bi and ^{208}Ti) as their number increases exponentially for larger R.

A figure of merit (FoM) is defined here as the ratio between the expected number of CNO+pep-neutrinos $N_{\text{CNO+pep}}$ over the statistical uncertainty of N_{total}:

$$\text{FoM}(R,\ \text{N}_\text{h}^{\text{geo}}(\text{min}),\ \text{N}_\text{h}^{\text{geo}}(\text{max})) = \frac{N_{\text{CNO+pep}}}{\sqrt{N_{\text{total}}}} \tag{7.1}$$

The FoM is calculated for each phase independently and maximised to find the best combination of the spherical FV cut R and the upper and lower energy cuts of $\text{N}_\text{h}^{\text{geo}}$. The estimation of $N_{\text{CNO+pep}}$ and N_{total} is based on the results of the standard spectral fit for a given combination of R and $\text{N}_\text{h}^{\text{geo}}(\text{min})$, $\text{N}_\text{h}^{\text{geo}}(\text{max})$. The exception are the external γ background events, which have been fixed by hand in the fit to give reasonable results. The systematic uncertainties of the standard fit are not considered in this estimate. The FV radius R is varied between 2.90 m and 3.10 m in steps of 0.05 m and the spectral fit is performed at each FV to calculate the FoM with the fit results of $N_{\text{CNO+pep}}$ and N_{total}, for a range of $340 < \text{N}_\text{h}^{\text{geo}}(\text{min}) < 400$ and $400 < \text{N}_\text{h}^{\text{geo}}(\text{max}) < 600$. The RoI is then selected according to the highest FoM given the parameters R, $\text{N}_\text{h}^{\text{geo}}(\text{min})$ and $\text{N}_\text{h}^{\text{geo}}(\text{max})$.

	R [m]	$\text{N}_\text{h}^{\text{geo}}(\text{min})$	$\text{N}_\text{h}^{\text{geo}}(\text{max})$	N_{total}
Phase 1	3.05	340	520	2990
Phase 2+3	2.95	340	515	5974

TABLE 7.1: Selected RoI for CNO+pep-ν CID analysis and the number of selected data events N_{total}.

	Live Time	Fiducial Mass	TFC Exposure
Phase 1	740.7 days	104.3 tons	55.77%
Phase 2+3	2888.0 days	94.4 tons	63.97%

TABLE 7.2: Live time, fiducial mass and TFC exposure for the selected RoI.

Table 7.1 summarises the RoI parameters which maximise the FoM for Phase 1 and Phase 2+3. The RoI for the combined analysis of Phase 2+3 is reduced to Phase 3 for the sake of simplicity, as Phase 2 and Phase 3 have very similar RoI values, $R(\text{Phase 2}) = 3.00\,\text{m}$, $R(\text{Phase 3}) = 2.95\,\text{m}$ and no difference in $\text{N}_\text{h}^{\text{geo}}(\text{min})$ and $\text{N}_\text{h}^{\text{geo}}(\text{max})$. The selected R decreases for later phases as more external γ events are reconstructed inwards, due to the decrease in position resolution from the loss of PMTs over time. This also has the consequence of a reduced energy resolution, such that more external γ events are reconstructed inside the energy RoI, which results in a lower $\text{N}_\text{h}^{\text{geo}}(\text{max})$ selection Phase 2+3, compared to Phase 1.

Table 7.2 shows the data taking live time, the fiducial mass and TFC-cut exposure for the selected RoI of each phase. Phase 1 has a data taking live time of 740.7 days [83], Phase 2+3 has 2888.0 days (1291.5 days from Phase 2 [84], 1596.5 days from Phase 3 [158]). Using the LS density of $(0.878 \pm 0.004)\,\text{g}\,\text{cm}^{-3}$ [90] results in a fiducial mass of (104.3 ± 0.5) tons for Phase 1 and (94.4 ± 0.4) tons for Phase 2+3. The TFC-cut exposure has been calculated using the standard toy-MC TFC exposure study [83] for the selected FV. The TFC exposure values are 55.77% for Phase 1 and 63.97% for Phase 2+3, both with a negligible uncertainty of $\pm 0.02\%$. The TFC algorithm has been explicitly optimised for Phase 2 and Phase 3, for the neutrino measurement using the standard spectral fit. For this reason Phase 1 provides a lower exposure compared to Phase 2+3, but the updated TFC algorithm nonetheless provides a better Phase 1 exposure compared to the old result of 48.5% [83].

Figure 7.1 shows the distribution of $\text{N}_\text{h}^{\text{geo}}$ for the selected TFC-cut data events of Phase 1 and Phase 2+3 in black and red, respectively. There is no discernible structure visible for

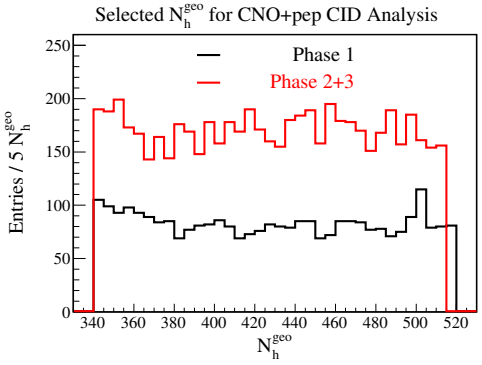


FIGURE 7.1: The N_h^{geo} distribution of data events for Phase 1 and Phase 2+3 in the selected RoI. Phase 1 in black, with $R < 3.05\,\text{m}$ and $340 < N_h^{geo} < 520$. Phase 2+3 in red with $R < 2.95\,\text{m}$ and $340 < N_h^{geo} < 515$.

these N_h^{geo} distributions.

ν	ε_E(Ph. 1)	ε_E(Ph. 2+3)	Rate [cpd / 100 t]	N_ν^{SSM}(Ph. 1)	N_ν^{SSM}(Ph. 2+3)
^{7}Be	$5 \cdot 10^{-5}$	$1.2 \cdot 10^{-4}$	47.90 ± 2.82 (HZ)	1.1 ± 0.1	10.0 ± 0.6
			43.72 ± 2.57 (LZ)	1.0 ± 0.1	9.1 ± 0.6
CNO	0.099	0.091	4.92 ± 0.78 (HZ)	210 ± 33	783 ± 124
			3.52 ± 0.52 (LZ)	150 ± 24	560 ± 83
pep	0.343	0.316	2.74 ± 0.04 (HZ)	405 ± 6	1510 ± 22
			2.78 ± 0.04 (LZ)	411 ± 6	1531 ± 22

TABLE 7.3: Energy efficiency ε_E and SSM expected number of solar neutrinos in the RoI.

Table 7.3 shows the expected number of solar neutrinos according to the HZ-SSM (B16-GS98 [35]) and LZ-SSM (B16-AGSS09met [35]) for the selected phase and RoI, after the TFC-cut. The fraction of events inside the energy RoI, including the trigger efficiency, is called the energy efficiency ε_E here. The expected number of neutrino events is calculated as the product of ε_E, detector live time, fiducial mass, TFC exposure (Table 7.2) and the SSM expected neutrino rates. For the sum of CNO- and pep-neutrinos the expected HZ-SSM signal to total ratio is $N_{CNO+pep}/2990 = 20.6 \pm 0.8\%$ for Phase 1 and $N_{CNO+pep}/5974 = 38.4 \pm 1.5\%$ for Phase 2+3.

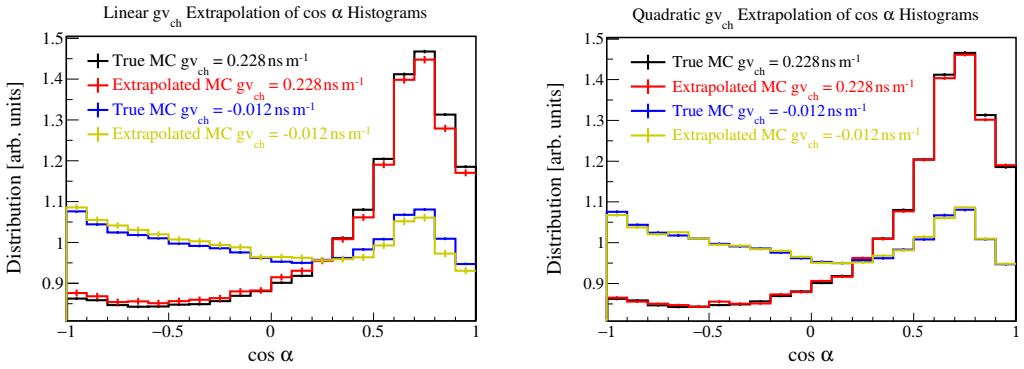
7.2 Production of neutrino signal MC and background MC

The general G4BX2 production of the CID MC $\cos\alpha$ histograms for the neutrino signal and background events is described in Section 3.2. For the selected CNO+pep-ν RoI, again, 200 MC events are simulated for every single data event, each with the corresponding position of the Sun and live PMT distribution.

The main background components are ^{210}Bi, ^{11}C and a smaller number of external γ events. The expected, relative contribution of the background species are $\sim 31\%$ for ^{210}Bi, $\sim 58\%$ for ^{11}C and $\sim 11\%$ for the external γ's, according to the results of the Phase 2 spectroscopy [84], scaled with the corresponding ε_E of the selected RoI. Both main background species ^{210}Bi and ^{11}C are simulated, to study a potential, energy dependent systematic uncertainty for a uniformly distributed CID background in Section 7.6.4. As no relevant systematic differences are found between the ^{210}Bi and ^{11}C events the background MC $\cos\alpha$ histograms used for the CNO+pep-ν region is the sum of ^{210}Bi and ^{11}C. The external γ events are not uniformly distributed in the FV, as they follow an exponential attenuation in the LS. For this reason they could potentially exhibit a non-negligible systematic difference in their $\cos\alpha$ distribution compared to the other background species. This systematic is studied by means of a toy-MC,

as well as through the comparison of the full ^{210}Bi and ^{11}C CID MC with a selection of pure external γ data events in Section 7.6.5. The full G4Bx2 simulation of external γ events is time-prohibitive.

For the signal MC both CNO-ν and pep-ν events are simulated, as they have significantly different neutrino energy distributions. The CNO-ν spectrum is continuous, with an endpoint energy of 1.74 MeV, while pep-neutrinos are mono-energetic at 1.44 MeV [41]. The impact of the differences between the $\cos\alpha$ histograms of CNO-ν and pep-ν is treated as a systematic uncertainty and quantified in Section 7.6.6. For the ^{7}Be-ν edge region of both Phase 1 and Phase 2+3 only the ^{7}Be-neutrinos are simulated as the signal and only ^{210}Bi events are simulated for the background, as both are highly dominant relative to the other signal and background species. The CID signal MC is produced for different values of $\mathrm{gv_{ch}}$, as described in Section 6.2. It is not practical to perform the full G4Bx2 simulation for every possible $\mathrm{gv_{ch}}$ value, as it would take too much time. Instead only a small number of $\mathrm{gv_{ch}}$ values are simulated with the full simulation chain and then used to produce the corresponding signal $\cos\alpha$ histograms. All other $\mathrm{gv_{ch}}$ dependent $\cos\alpha$ histograms are produced through interpolation of these fully simulated $\cos\alpha$ histograms. For the ^{7}Be-ν edge region it is found that a linear interpolation between only two histograms with $\mathrm{gv_{ch}} = -0.012\,\mathrm{ns\,m^{-1}}$ and $\mathrm{gv_{ch}} = 0.228\,\mathrm{ns\,m^{-1}}$ is enough to reproduce the $\cos\alpha$ histograms of the full MC chain, for all $\mathrm{gv_{ch}}$ values in between.



(a) (b)

FIGURE 7.2: (a) Linear extrapolation of histogram with $\mathrm{gv_{ch}} = 0.228\,\mathrm{ns\,m^{-1}}$ from $\mathrm{gv_{ch}} = -0.012\,\mathrm{ns\,m^{-1}}$, $0.108\,\mathrm{ns\,m^{-1}}$ (red) compared to the full MC histogram (black): $\chi^2/\mathrm{ndf} = 44.6/19$. Linear extrapolation of histogram with $\mathrm{gv_{ch}} = -0.012\,\mathrm{ns\,m^{-1}}$ from $\mathrm{gv_{ch}} = 0.108\,\mathrm{ns\,m^{-1}}$, $0.228\,\mathrm{ns\,m^{-1}}$ (yellow) compared to the full MC histogram (blue): $\chi^2/\mathrm{ndf} = 50.6/19$. (b) Quadratic extrapolation of histogram with $\mathrm{gv_{ch}} = 0.228\,\mathrm{ns\,m^{-1}}$ from $\mathrm{gv_{ch}} = -0.012\,\mathrm{ns\,m^{-1}}$, $0.108\,\mathrm{ns\,m^{-1}}$ (red) compared to the full MC histogram (black): $\chi^2/\mathrm{ndf} = 20.7/19$. Quadratic extrapolation of histogram with $\mathrm{gv_{ch}} = -0.012\,\mathrm{ns\,m^{-1}}$ from $\mathrm{gv_{ch}} = 0.108\,\mathrm{ns\,m^{-1}}$, $0.228\,\mathrm{ns\,m^{-1}}$ (yellow) compared to the full MC histogram (blue): $\chi^2/\mathrm{ndf} = 18.4/19$.

For the selected CNO+pep-ν RoI this linear interpolation is not able to adequately reproduce the $\cos\alpha$ histograms of the full MC, as the larger electron energies at this RoI produce a larger number of Cherenkov photons compared to the ^{7}Be edge region. This can be seen in Figure 7.2(a), where the signal $\cos\alpha$ histograms are produced for three $\mathrm{gv_{ch}}$ values with $\mathrm{gv_{ch}} = -0.012\,\mathrm{ns\,m^{-1}}$, $\mathrm{gv_{ch}} = 0.108\,\mathrm{ns\,m^{-1}}$ and $\mathrm{gv_{ch}} = 0.228\,\mathrm{ns\,m^{-1}}$, using the full MC chain. The true $\cos\alpha$ histograms of $\mathrm{gv_{ch}} = 0.228\,\mathrm{ns\,m^{-1}}$ is shown black and for $\mathrm{gv_{ch}} = -0.012\,\mathrm{ns\,m^{-1}}$ in blue. They are compared to the extrapolated histograms in red and yellow, respectively. The extrapolation is performed with a linear fit of the $\cos\alpha$ value for each bin of the two histograms from the other $\mathrm{gv_{ch}}$ values. The true $\cos\alpha$ histograms are significantly different from the extrapolated histograms, with a $\chi^2/\mathrm{ndf} = 44.6/19$, p-value $= 8\cdot10^{-4}$ for $\mathrm{gv_{ch}} = 0.228\,\mathrm{ns\,m^{-1}}$ and $\chi^2/\mathrm{ndf} = 50.6/19$, p-value $= 1\cdot10^{-4}$ for $\mathrm{gv_{ch}} = -0.012\,\mathrm{ns\,m^{-1}}$.

Figure 7.2(b) instead shows that using a quadratic fit between three gv_{ch} values provides an extrapolation that is in agreement with the full MC $\cos\alpha$ histograms. Using a set of four gv_{ch} values $\{-0.012\,\mathrm{ns\,m^{-1}}, 0.068\,\mathrm{ns\,m^{-1}}, 0.148\,\mathrm{ns\,m^{-1}}, 0.228\,\mathrm{ns\,m^{-1}}\}$ to perform a quadratic fit on the first three gv_{ch} values to extrapolate the last $\cos\alpha$ histogram with $gv_{ch} = 0.228\,\mathrm{ns\,m^{-1}}$ gives $\chi^2/\mathrm{ndf} = 20.7/19$, p-value $= 0.35$ and using the last three gv_{ch} values to extrapolate the first $\cos\alpha$ histogram with $gv_{ch} = -0.012\,\mathrm{ns\,m^{-1}}$ gives $\chi^2/\mathrm{ndf} = 18.4/19$, p-value $= 0.49$. This means that the underlying dependency of the neutrino signal $\cos\alpha$ distribution on gv_{ch} is approximated well by a quadratic function for the CNO+pep-ν region.

The signal $\cos\alpha$ histograms for different values of the nuisance parameter Δr_{dir} are produced from the linear interpolation between $\Delta r_{\mathrm{dir}} = 0\,\mathrm{cm}$ and $\Delta r_{\mathrm{dir}} = 6\,\mathrm{cm}$. As Δr_{dir} cannot be calibrated and consequently cannot be constrained by a pull-term in the CID analysis the performance of this method compared to the full MC does not matter. The parameter Δr_{dir} is allowed to take any value, necessary to reproduce the $\cos\alpha$ histograms of data with the corresponding MC histograms, as good as possible.

7.3 Selection of the Nth-Hit cutoff and CID-histogram binning

Previously the Nth-Hit cutoff for the Phase 1 CID analysis at the ^{7}Be-ν edge has been selected conservatively as Nth-Hit(max) $= 2$. The reason for this is twofold:

1. The $gv_{ch} = (0.108 \pm 0.039)\,\mathrm{ns}$ pull-term is measured with the ^{40}K γ-source for a cutoff of Nth-Hit(max) $= 3$ and has been unclear how applicable this gv_{ch} value would be for a very different Nth-Hit cutoff in the neutrino measurement.

2. The χ^2 values between ^{7}Be-ν and ^{210}Bi MC $\cos\alpha$ histograms, as a function of individual Nth-Hits, have shown an unchanging, flat distribution starting from the fifth hit to the tenth. This has been incorrectly interpreted, such that there is only Cherenkov information until the fourth hits of the events. Values for Nth-Hit> 10 have not been investigated and only a single set of nuisance values for gv_{ch} and Δr_{dir} has been used.

For this reason the analysis has been performed on Nth-Hit(max) $= 2$ and the systematic uncertainty from this choice has been estimated maximally conservative, as the absolute difference between Nth-Hit(max) $= 2$ and Nth-Hit(max) $= 3$. The final goal of this CID measurement is to maximise the sensitivity on the CNO-ν rate, which makes it is necessary to revisit the Nth-Hit(max) selection and perform a more sophisticated estimation of the corresponding systematic uncertainties. The method for optimising the Nth-Hit cutoff is a comparison of the χ^2 values between the pure signal and background MC $\cos\alpha$ histograms, as a function of the individual Nth-Hits. For this toy-MC study both signal MC and background MC $\cos\alpha$ histograms are produced with the statistics of data for the relevant phase and RoI. A χ^2-test is then calculated between the two histograms, independently for each selected Nth-Hit and the χ^2 is subtracted by the ndf (#Bins-1), i.e. the expected χ^2 value for a perfect statistical agreement between the signal and background.

$$\Delta\chi^2 = 10^{-4}\sum^{10^4}\left(\sum_i^{\#\mathrm{Bins}} \frac{(N(\cos\alpha_i)^{\mathrm{signal}} - N(\cos\alpha_i)^{\mathrm{back}})^2}{N(\cos\alpha_i)^{\mathrm{signal}} + N(\cos\alpha_i)^{\mathrm{back}}}\right) - (\#\mathrm{Bins} - 1) \tag{7.2}$$

Let $N(\cos\alpha_i)$ be the number of entries for the histogram bin i, out of a total number of bins #Bins. Equation 7.2 then gives the value of the Cherenkov information as $\Delta\chi^2$ for each Nth-Hit between the neutrino signal MC and background MC histograms, averaged over 10k toy simulations. This calculation is now performed up to a Nth-Hit for which $\Delta\chi^2 = 0$ and beyond, as $\Delta\chi^2 = 0$ means that signal and background events are indistinguishable, given the statistics of data. The bigger the difference between signal and background MC is, the larger is $\Delta\chi^2$

value and the expected CID sensitivity. The benefit of this procedure is that the calculation of $\Delta\chi^2$ is simple, as it does not depend on the full CID analysis with nuisance parameters. It gives an estimate on how well neutrino signal and background can be distinguished. This toy-MC study is performed for multiple sets of nuisance parameter values and the same method is also used to optimise the $\cos\alpha$ histogram binning again, as Phase 2+3 has not been investigated previously.

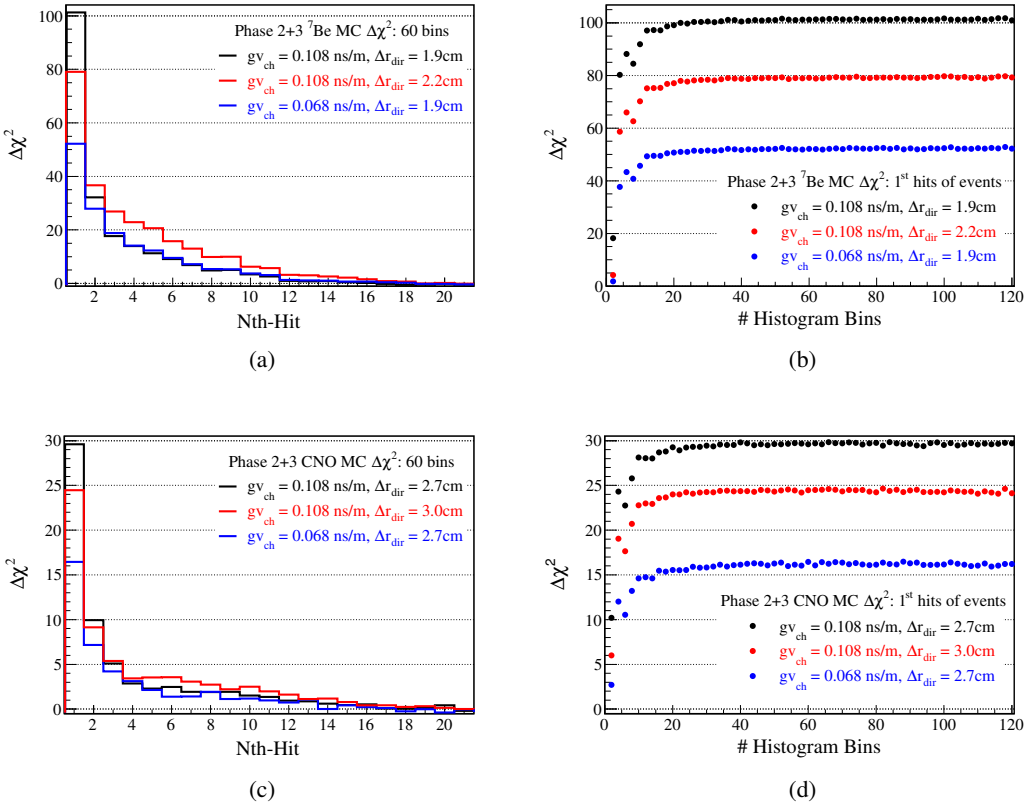


FIGURE 7.3: The toy-MC study of the Cherenkov information $\Delta\chi^2$ as a function of the individual Nth-Hit $\cos\alpha$ histograms and the histogram binning, for Phase 2+3. The nominal MC Δr_{dir} value and $gv_{ch} = 0.108\,ns\,m^{-1}$ is shown in blue, an increased Δr_{dir} value is shown in red and the nominal MC Δr_{dir} with a decreased $gv_{ch} = 0.068\,ns\,m^{-1}$ is shown in blue.

(a) $\Delta\chi^2$ vs. the Nth-Hits for the ^{7}Be-ν region.
(b) $\Delta\chi^2$ vs. $\cos\alpha$ histogram binning for the ^{7}Be-ν region.
(c) $\Delta\chi^2$ vs. the Nth-Hits for the CNO+pep-ν region.
(d) $\Delta\chi^2$ vs. $\cos\alpha$ histogram binning for the CNO+pep-ν region.

Figure 7.3 shows the result of the $\Delta\chi^2$ toy-MC study for Phase 2+3. The corresponding plots for Phase 1 can be found in the Appendix C.1. On the right $\Delta\chi^2$ is plotted as a function of the number of histogram bins for the first hits of the events, while on the left $\Delta\chi^2$ is plotted as a function of the Nth-Hits of the events for histograms with 60 bins. The statistical uncertainty on $\Delta\chi^2$ is about 0.7 and smaller than the plotted points. Three sets of gv_{ch} are investigated: the expected value of $gv_{ch} = 0.108\,ns\,m^{-1}$ from the γ-calibration and the nominal MC value of Δr_{dir} for each RoI in black, an increased Δr_{dir} and the same gv_{ch} value in red and a decreased $gv_{ch} = 0.068\,ns\,m^{-1}$ with the nominal Δr_{dir} value in blue. As expected, the $\Delta\chi^2$ value shows a decrease for later Nth-Hits. The first hits of events is highest for $gv_{ch} = 0.108\,ns\,m^{-1}$ and the nominal MC Δr_{dir} value, while it is lowest for the $gv_{ch} = 0.068\,ns\,m^{-1}$. The lower gv_{ch}

value consistently gives a lower $\Delta\chi^2$ until the fourth (third) hits of events for the CNO+pep-ν region (^{7}Be-ν edge), where both injected gv_{ch} values give the same $\Delta\chi^2$ within the statistical uncertainty. This means that the *direct* information from Cherenkov hits is only relevant up until Nth-Hit = 4. On the other hand the effect of an increased Δr_{dir} is clearly visible for the high statistics ^{7}Be-ν edge region of Phase 2+3. While the first hit gives still a larger $\Delta\chi^2$ for the nominal Δr_{dir} value, already starting from the second hit the $\Delta\chi^2$ is larger for the increased Δr_{dir} value. Due to lower statistics this is less pronounced for the CNO+pep-ν region, but it is still visible.

The reason for this behaviour is that the parameter Δr_{dir} describes the influence of Cherenkov hits on the position reconstruction, where the Cherenkov hits on the PMTs apply a bias on the position reconstruction, towards the initial direction of the electron. All PMT hits are ToF corrected on the reconstructed position of a given event. This means that the $\cos\alpha$ value of all PMT hits is *indirectly* influenced by the Cherenkov hits through the position reconstruction. The lower $\Delta\chi^2$ value of the first hits of the events given by the larger Δr_{dir} is the effective decrease of the Cherenkov peak, due to an increase of the Δr_{dir} induced negative $\cos\alpha$ slope. For later Nth-Hits this indirect Cherenkov information continues to influence the scintillation hits and so there is still a $\Delta\chi^2 > 0$ until Nth-Hit$\sim$ 17. The total information given by this indirect influence of the Cherenkov hits is given by the area under the $\Delta\chi^2$ histogram, as the CID analysis sums up the χ^2 values of all Nth-Hits until Nth-Hit(max). Hence, it can be estimated that the *indirect* Cherenkov information from the scintillation hits is smaller, but of similar size as the Cherenkov information from the early, *direct* Cherenkov hits of the events.

The optimised value for Nth-Hit(max) is selected according to the condition of $\Delta\chi^2 = 0$, where no difference between signal and background MC $\cos\alpha$ histogram can be expected anymore, given the statistics of data. For Phase 2+3 this is Nth-Hit(max) = 17 and for Phase 1 this is Nth-Hit(max) = 15, given that the true Δr_{dir} value of data is unknown. These selections are made by eye and as such are somewhat arbitrary. If the MC would perfectly reproduce data it would in principle be possible to use every hit of the events ($>$ 200 hits), as the CID analysis summation of the χ^2 profiles of the late hits $>$ 17 would only contribute noise, effectively adding only a constant to the summed χ^2 profile. This cannot be guaranteed and even a hypothetical systematic difference between the data background and the MC background, that is negligible for the first $\sim$ 20 hits of the events, could become significant if 200 hits are used for the CID analysis. It is thus beneficial to select Nth-Hit(max) in such a way that all available direct and indirect Cherenkov information is used, given an unknown true value of Δr_{dir}. At the same time it is reasonable to avoid adding too many hits without any Cherenkov information ($\Delta\chi^2 = 0$), as they could potentially contribute a systematic uncertainty. Consequently, the study of such a potential systematic uncertainty between the data background and the MC background is performed in Section 7.6.5 for the selected Nth-Hit region. The systematic impact of the somewhat arbitrary choice the Nth-Hit cutoff is studied in Section 7.6.1 for a range of Nth-Hit(max) values around the nominal Nth-Hit(max) selection.

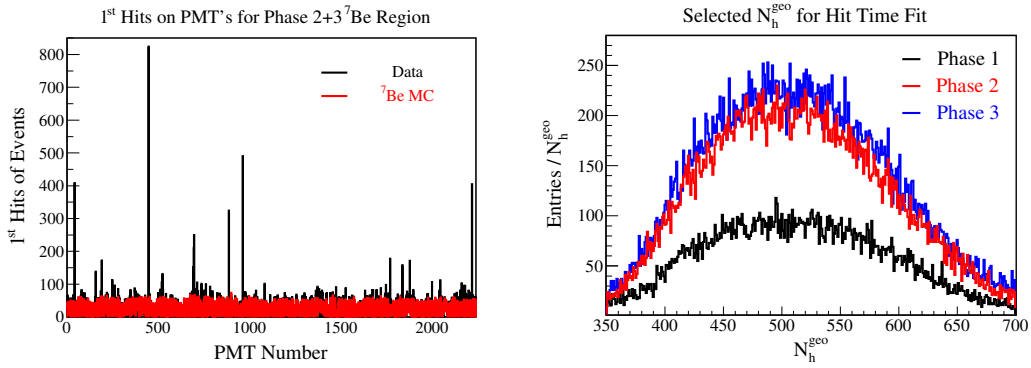
Looking at the binning optimisation, for both the ^{7}Be-ν edge region in Figure 7.3(a) (top) and the CNO+pep-ν region 7.3(c) (bottom), the $\Delta\chi^2$ increases as the number of histograms bins increases and saturates to a constant value at $\sim$ 30 bins. For less than 20 bins there are structures in the graphs that are larger than the uncertainty on the averaged $\Delta\chi^2$, indicating a strong dependence of the signal $\cos\alpha$ shape on the exact number of bins for $<$ 20 bins. For $gv_{ch} = 0.108\,\mathrm{ns\,m^{-1}}$ and the nominal Δr_{dir} MC value the shown $\Delta\chi^2$ values are the highest, while it is lower for the increased Δr_{dir} and lowest for $gv_{ch} = 0.068\,\mathrm{ns\,m^{-1}}$. This is as expected for the first hits of the events, because an increase in Δr_{dir} means a large negative slope, which effectively decreases the Cherenkov peak at $\cos\alpha \sim 0.7$, without changing the number of Cherenkov hits. Here, $gv_{ch} = 0.068\,\mathrm{ns\,m^{-1}}$ gives the least number of Cherenkov hits so the corresponding $\Delta\chi^2$ values are also the smallest. To optimise the selection of the number of $\cos\alpha$ histogram bins the $\Delta\chi^2$ must be maximised. This is the case for any number of bins

> 30, independent of the RoI and the exact value of gv_{ch} and Δr_{dir}. Here, a number of 60 bins is chosen for the main CID analysis, for both Phase 1 and Phase 2+3 and for both the ^{7}Be-ν edge region and CNO+pep-ν region. A possible systematic uncertainty from this choice of binning is investigated in Section 7.6.1. Table 7.4 summarises the selection of Nth-Hit(max) and $\cos\alpha$ histogram binning, which are the same for the ^{7}Be-ν edge region and CNO+pep-ν region for a given phase.

	Nth-Hit(max)	Binning
Phase 1	15	60
Phase 2+3	17	60

TABLE 7.4: Optimised selection of Nth-Hit and $\cos\alpha$ histogram binning.

7.4 PMT selection and correction of the relative PMT time offset



(a) (b)

FIGURE 7.4: (a) The number of the first hits of the events on a given PMT, for the ^{7}Be-ν region of Phase 2+3. data (black) shows over-selected PMT peaks, while the ^{7}Be-ν MC (red), normalised to data statistics, shows a relatively flat distribution. (b) The distribution of N_h^{geo} for the TFC-tagged ^{11}C data events used for the production of the PMT hit time distributions.

A major systematic difference between the CID data and the MC has been identified as the over-selection of some PMTs for early hits in data. This can be seen in Figure 7.4(a) for the ^{7}Be-ν region of Phase 2+3, where the comparison to MC clearly shows some data PMTs show peaks for the number of detected first hits of the events.

Previously, this systematic difference has been corrected by selecting PMTs based on a statistical argument, where the distribution of first hits of the events on the PMTs is compared between data and MC. If the number of detected first hits on a data PMT is outside of the expected Poissonian 68.3% CI of the corresponding MC PMT, then the hits on this PMT are ignored for the production of the $\cos\alpha$ histogram of data and MC; the PMT is deselected. This is explained in Section 6.4 and the method is by design a conservative approach that likely deselects too many non-misbehaving PMTs. The reason for this conservatism is that the initial ^{7}Be-ν CID measurement has been used to study the principal feasibility of CID and all methods and uncertainty estimations have been maximally conservative. Here on the other hand the goal is perform a CID measurement of CNO-neutrinos with a reasonably high sensitivity, i.e. without being maximally conservative at every step of the analysis. For this

reason a better method to deselect the misbehaving PMTs is presented in this section, based on the PMT hit time distributions of the separate PMTs.

The underlying cause for the over-selection (and under-selection) of the first hits of events for some PMTs, as seen in Figure 7.4(a), is the time behaviour of those individual PMTs relative to the other PMTs. This effect has already been noticed in Chapter 4, where some PMTs have shown a run to run time misbehaviour as well as a statistically significant, relative time offset between the PMTs. There the analysis is based on the PMT hit time distributions of the ^{214}Po calibration source events. The calibration campaign has been performed only at the end of Phase 1 and only for specific source positions for a number of calibration runs close in time. Furthermore, these ^{214}Po-source studies have been produced with a known source position and without the ToF correction, precisely because there the effective refractive index, necessary for the ToF correction, has been measured for the purpose of the CID analysis. As the Nth-Hit definition for the production of the CID $\cos\alpha$ histograms relies on the reconstructed position and ToF correction this means that effectively the PMT time misbehaviour and relative time offset between the hit time distributions of the PMTs could also rely on the performance of the position reconstruction algorithm and ToF correction itself. It is therefore not clear if a PMT selection based on the results from Section 4.5 is applicable to Phase 2+3, or even to the entirety Phase 1.

Thus, in this section the ToF corrected hit time distributions of each individual PMT are investigated using ^{11}C-tagged data events [1], separately for Phase 1, Phase 2 and Phase 3. The events are selected within $350 < N_h^{geo} < 700$, ^{11}C-tagged and within the CNO+pep-ν region FV for each phase (Phase 1: $R < 3.05\,$m, Phase 2: $R < 2.95\,$m, Phase 3: $R < 2.95\,$m). The reason for this selection is that ^{11}C is produced uniformly throughout the detector and has a reasonably large number of events and at the same time a large number of PMT hits per event in the selected RoI. Additionally, the ^{11}C-tagged data is independent from the ^{11}C-cut data of the CNO CID analysis, excluding potential systematic correlations between the PMT selection and the CID analysis.

Figure 7.4(b) shows the N_h^{geo} distribution of ^{11}C-tagged data events in the corresponding RoI for each phase. Phase 1 has 20270 ^{11}C-tagged events, Phase 2 has 42769 events and Phase 3 has 47932 events. The average number of PMT hits per event is $\langle N_h^{geo} \rangle \approx 500$. The ToF corrected hit time t-distribution of each individual PMT is produced for each phase separately, according to the definition of Equation 3.7. The reason to look at the t-distribution separately for each phase is shown in Figure 7.5(a), exemplary for PMT No. 1006. For this particular PMT the t-distribution of Phase 3 is significantly different compared to Phase 1 and Phase 2, which is not the case for most other PMTs, which show a time behaviour that is comparable between all phases. The t-distributions are fitted with a bi-exponentially modified Gaussian distribution similar, but less complex than Equation 4.2. The fit function $f(t)$ is defined here as:

$$\begin{aligned} f(t) =& A_1\, e^{\frac{1}{2}\left(\frac{\sigma}{\tau_1}\right)^2 - \frac{t-\mu}{\tau_1}}\, \mathrm{erfc}\left(\frac{1}{\sqrt{2}}\left(\frac{\sigma}{\tau_1} - \frac{t-\mu}{\sigma}\right)\right) + \\ & A_2\, e^{\frac{1}{2}\left(\frac{\sigma}{\tau_2}\right)^2 - \frac{t-\mu}{\tau_2}}\, \mathrm{erfc}\left(\frac{1}{\sqrt{2}}\left(\frac{\sigma}{\tau_2} - \frac{t-\mu}{\sigma}\right)\right) \end{aligned} \tag{7.3}$$

Where t is the ToF corrected hit time, "erfc" is the complementary error function ($\mathrm{erfc}(x) = 1 - \mathrm{erf}(x)$), A_1, A_2 are the amplitudes, τ_1, τ_2 are the exponential time constants and μ and σ are the Gaussian mean and standard deviation, respectively. The test statistic used to fit these

[1] The low energy ^{14}C events have also been investigated, but have been found to be unusable for this analysis. The position reconstruction uncertainty is too large, given the number of PMT hits of $N_{geo}^{h} < 160$.

parameters is the likelihood-ratio λ_{LR}, defined as follows:

$$\lambda_{\mathrm{LR}} = -2\sum_{i}^{I}\left[\log\left(\frac{e^{-f(t_i)}f(t_i)^{d_i}}{d_i!}\right) - \log\left(\frac{e^{-d_i}d_i^{d_i}}{d_i!}\right)\right] \tag{7.4}$$

Where the PMT hit time distribution is given as a histogram with a binning of 1 ns from $i = 0$, to $I = 55$. The number of data entries in each bin i is given by d_i, whereas the number of expected entries is given by the function value $f(t_i)$ of Equation 7.3. The likelihood-ratio λ_{LR} is then the sum of the logarithm of the Poissonian likelihood for each bin. The constant likelihood, where d_i is assumed to be its own expected value is subtracted to make λ_{LR} more comparable to a Gaussian χ^2 value.

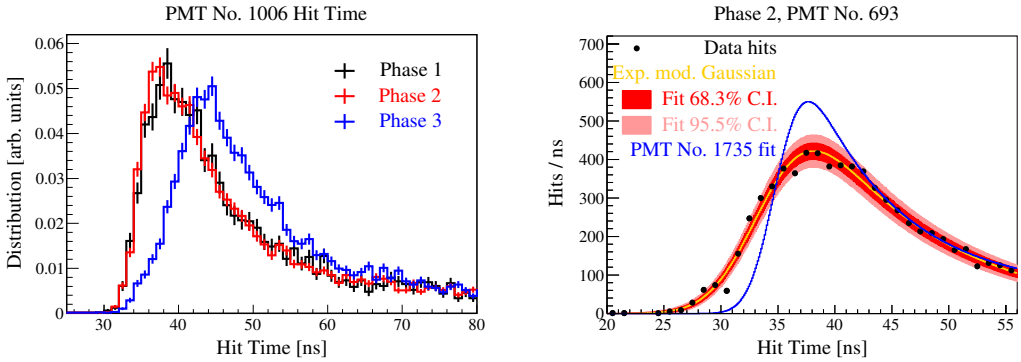


FIGURE 7.5: (a) The normalised hit time distribution of the same PMT No. 1006 for Phase 1 (black), Phase 2 (blue) and Phase 3 (red) as an example that time behaviour of individual PMTs can change drastically over time. (b) Example of the PMT hit time distribution fit for the misbehaving PMT No. 693 in Phase 2. The data hits (black) are fitted with the bi-exponentially modified Gaussian distribution (yellow) and the resulting confidence intervals are shown in shades of red. For comparison the best fit distribution of a normal PMT is also shown (blue).

An example of the fit is given in Figure 7.5(b), where for Phase 2 the hit time distribution of PMT No. 693 is shown in black, with the corresponding best fit in yellow and the 68.3% (1σ), 95.9% (2σ) confidence intervals in shades of red. It can be seen that some data points are outliers, but the overall $\lambda_{\mathrm{LR}}/\mathrm{ndf} = 1.9$ is not intolerably large. For the purpose of illustration the best fit result of the normal PMT No. 1735 is also shown in blue. In this case "normal" means that the best fit parameter values of PMT No. 1735 are distributed similar to the bulk of the best fit values of all PMTs. It can be seen that PMT No. 1735 and 693 have a clearly different hit time distribution and that PMT No. 693 gets over-selected for the first hits of an event, not because it sees more Cherenkov hits, but because the ToF corrected PMT hit time distribution is much broader and contributes more hits at earlier times compared to most other PMTs. Thus, under the assumptions that most PMTs in data behave comparably to the PMTs simulated in MC it is possible to deselect individual PMTs if they show a ToF corrected hit time distribution significantly different from the bulk of the PMTs. Correspondingly, the deselection of misbehaving PMTs is based on the distribution of the best fit parameters from Equation 7.3 and is performed as follows:

1. Produce and fit the ToF corrected hit time t-distributions for each PMT.

2. Plot the distribution of the best fit results of (a) $\lambda_{\mathrm{LR}}/\mathrm{ndf}$, (b) μ, (c) the corresponding fit uncertainty $\Delta\mu$, (d) σ, and (e) the early hits over total ratio, where early hits are defined as $t < \mu - 2\sigma$. This ratio is chosen because it is sensitive to distributions which are to

broad or have a hit time distribution at early times that is not well explained by the fit function.

3. Deselect PMTs which have outlying best fit result compared to the bulk of the PMTs for any of the above mentioned parameters. The μ cut values are to deselect PMTs outside of ± 1 ns around the average of the bulk distribution $\langle\mu\rangle$, i.e. a Gaussian fit of the μ distribution, while explicitly excluding outliers. The values are $\langle\mu\rangle$(Phase 1) = 35.34 ns $\langle\mu\rangle$(Phase 2) = 35.32 ns and $\langle\mu\rangle$(Phase 3) = 35.33 ns each with a statistical uncertainty of 0.01 ns. The other cut values are the same for all phases: deselect PMTs with $\Delta\mu > 0.3$ ns, $\lambda_{\mathrm{LR}}/\mathrm{ndf} > 2$, or early hits over total ratio with values larger than 0.015.

4. Fix σ in the fitting algorithm to the peak value of the distribution of only the selected PMTs. These fixed values are $\langle\sigma\rangle$(Phase 1) = 1.808 ns, $\langle\sigma\rangle$(Phase 2) = 1.840 ns and $\langle\sigma\rangle$(Phase 3) = 1.842 ns, each with a statistical uncertainty of 0.006 ns.

5. Perform the fit and PMT selection again, now with the fixed σ value. This is done to improve the selection power of $\lambda_{\mathrm{LR}}/\mathrm{ndf}$, where misbehaving PMTs give larger, best fit λ_{LR} values as their hit time distribution shape cannot be accommodated by the fit with the fixed σ value. For example the fit of PMT No. 693 in Phase 2 gives $\lambda_{\mathrm{LR}}/\mathrm{ndf} = 1.9$ with a free σ parameter and $\lambda_{\mathrm{LR}}/\mathrm{ndf} = 3.8$ with the fixed value.

Figure 7.6 shows an example of the fit parameter distributions used for the PMT deselection in Phase 2. The black histograms show the distribution of the parameters for all PMTs, before any deselection and the red histograms show the distribution of the parameters after fixing σ in the fit and after the PMT deselection. The distribution of μ in Figure 7.6(a) shows the bulk of the PMTs around 35.3 ns and this value is the same within the statistics for all phases, which is a good sign that the collective hit time behaviour of most PMTs stays stable between the phases. It can be seen that there are some PMTs with outlying μ values, which are deselected. The same is true for the other plots of the $\Delta\mu$, $\lambda_{\mathrm{LR}}/\mathrm{ndf}$ and early hits over total ratio distributions. The $\lambda_{\mathrm{LR}}/\mathrm{ndf}$ values of the PMTs are shifted to slightly higher values after the cut on parameters, i.e. after fixing σ.

The effective result of this hit time distribution based deselection of PMTs is shown in Figure 7.7 for Phase 2. The distribution of PMTs with a certain number of first hits of the events is shown for the data without the deselected PMTs in black and for data with the PMT deselection in red. The data without deselected PMTs has PMTs with up to 286 first hits and the bulk is shifted to smaller numbers compared the distribution of data after the PMT deselection. This is compared the expected distribution according to MC in blue. The expectation is produced by normalising the number of MC events to the statistics of data (as in Figure7.4(a)) and applying a Poissonian smearing around the expected number of first hits.

After the PMT deselection the data is in good agreement with the MC expected distribution. The corresponding plots for Phase 1 and Phase 3 are in the Appendix C.2, where the data is also in good agreement with the MC expectation after the PMT deselection.

Method of PMT selection	Phase 1 #PMTs	Phase 2 #PMTs	Phase 3 #PMTs
First hit statistics	1432	1326	975
^{11}C hit time distribution	1678	1499	1121

TABLE 7.5: Number of selected PMTs for the two different methods of PMT selection.

Table 7.5 compares the numbers of selected PMTs for both methods of PMT deselection, the statistical deselection of Section 6.4 and the method based on the hit time distribution of

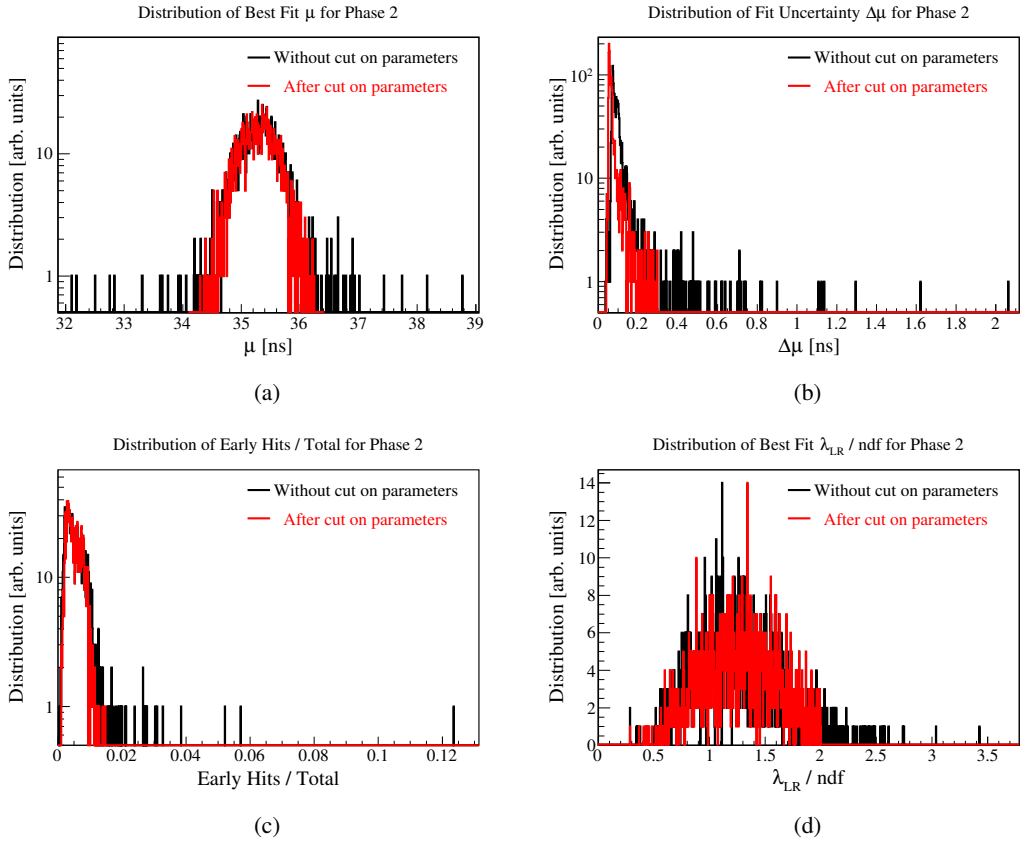


FIGURE 7.6: Example distributions of the four parameters used to select good PMTs in Phase 2. The black histograms show the distributions for all PMTs with a free σ parameter. The red histograms show these parameter distributions for the final PMT selection, after fixing $\sigma = 1.842\,\text{ns}$.

(a) Distribution of the Gaussian mean value μ.

(b) Distribution of the fit uncertainty on the Gaussian mean value $\Delta\mu$.

(c) Distribution of the early hits ($t < \mu - 2\sigma$) over total ratio.

(d) Distribution of the best fit negative log-likelihood λ_{LR} results.

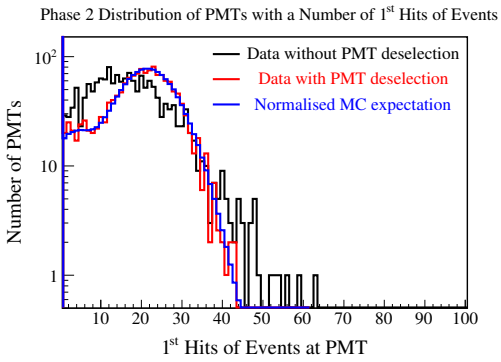


FIGURE 7.7: Distribution of the number of first hits of the events on the PMTs for the ^{7}Be-ν region of Phase 2. The data without a PMT deselection (black) has PMTs with up to 286 first hits. The comparison of the data with the PMT deselection (red) and the expected distribution from MC (blue) show a good agreement.

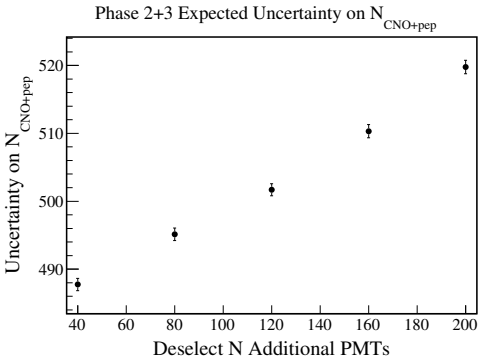


FIGURE 7.8: Toy-MC estimated uncertainty on the $N_{\mathrm{CNO+pep}}$ analysis of Phase 2+3 for a number of additionally deselected PMTs. The nominal selection of PMTs from this section is given as $N = 0$.

this section. The latter keeps more PMTs selected and thus gives a higher expected sensitivity, under the assumption that all misbehaving PMTs are deselected.

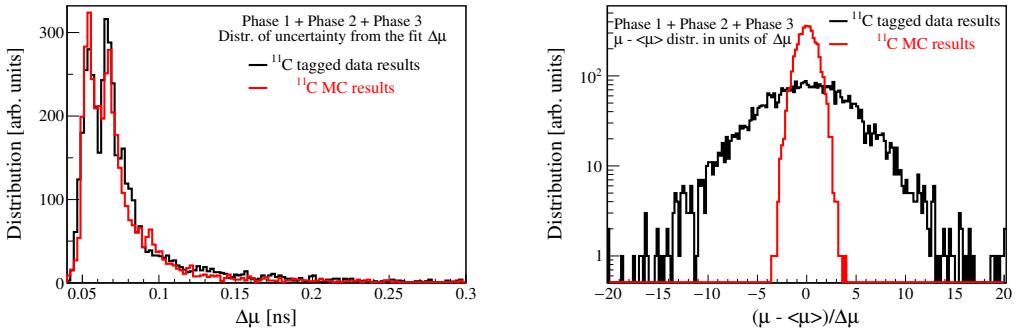
The dependence on the statistical uncertainty of the CID analysis is shown in Figure 7.8 for the fit of $N_{\mathrm{CNO+pep}}$ in Phase2+3. Here, 1000 toy-MC samples have been produced with a simulated value of $N_{\mathrm{CNO+pep}} = 2293$ events out of 5974 according to the HZ expectation. The nuisance parameters are simulated as $\mathrm{gv}_{\mathrm{ch}} = 0.108\,\mathrm{ns}$ and $\Delta r_{\mathrm{dir}} = 2.7\,\mathrm{cm}$. The toy-analyses are performed with the selections for the nominal data analysis of Nth-Hit = 17 and 60 bins for the $\cos\alpha$ histograms. The PMTs used in the $\cos\alpha$ histograms production are selected by the hit time distribution method of this section and additional PMTs have been deselected by further constraining the cut values of the fit parameters. Deselecting additional PMTs increases the statistical uncertainty, where the nominal PMT selection gives an expected uncertainty of ~ 480 and each additional deselected PMT increases the uncertainty by about ~ 0.2 ($\sim 0.09\%$). This is because the signal neutrino MC and background MC $\cos\alpha$ histograms are looking more similar to each other for a lower number of selected PMTs.

Although here the selection of the PMTs is based on the underlying hit time distribution, instead of just statistical arguments, the cut values of the parameters are still selected somewhat arbitrary. There is no a-priori reason to select $\pm 1.0\,\mathrm{ns}$ around $\langle\mu\rangle$ instead of $\pm 1.2\,\mathrm{ns}$ for example. For this reason Section 7.6.2 has a dedicated study of the PMT selection systematic uncertainty, where additional PMTs are deselected in the same way as above, to see whether a significant change of the data CID analysis results occurs.

Relative PMT time offset correction

The fit parameter μ shows a certain distribution around a central value $\langle\mu\rangle$. It is shown in the following that the μ distribution of data cannot be explained from the statistical uncertainty of the PMT hit time distribution fit itself, but rather that the PMTs have effectively a constant offset relative to each other. For this the ^{11}C MC PMT hit time distributions are fitted in the same way as data, where the MC PMT hit time distribution is reduced to the statistics of the corresponding data PMT for each phase. The MC analysis is not used to deselect PMTs but the same PMT selection is applied as on data. Figure 7.9(a) shows the resulting PMT hit time fit parameter distribution of the fit uncertainty $\Delta\mu$ between data in black and MC in red, for the sum of all phases. Both $\Delta\mu$ distributions of data and MC show a similar shape, where a double peak structure with an exponential tail is visible. The peak at smaller $\Delta\mu$ values comes from Phase 2 and Phase 3 which have higher event statistics compared to Phase 1 and therefore produce smaller uncertainties. The exponential tail likely comes from the different hit statistic on the PMTs, stemming from the difference in quantum efficiencies and spatial distribution of the PMTs, as well as the presence of a light collecting cone.

The fit procedure results in a reasonable agreement between the data and the MC for the uncertainty $\Delta\mu$. Figure 7.9(b) on the other hand shows a significant difference between the data and the MC for the μ distribution around the central value $\langle\mu\rangle$, in units of the uncertainty



(a) (b)

FIGURE 7.9: (a) Distribution of the fit uncertainty $\Delta\mu$ for data in black and for MC normalised to data statistics in red. The results of Phase 1, Phase 2 and Phase 3 are summed up. (b) Distribution of the best fit time mean values around the average of all mean values $(\mu - \langle\mu\rangle)$ in units of fit uncertainty $\Delta\mu$. The MC (red) has a standard deviation of 1.0, while the data (black) has a standard deviation of 4.7.

$\Delta\mu$. Here, PMTs with and without light collecting cone are plotted relative to their own central value. If the value of μ around the central value is only smeared by the statistical uncertainty of the fit, then the $(\mu - \langle\mu\rangle)/\Delta\mu$ distribution should have a standard deviation of 1.0. This is the case for the MC, which means that the fit procedure itself behaves correctly, as it is known that MC has no systematic differences of the time behaviour between the PMTs. For the data on the other hand the distribution has a standard deviation of 4.7, which corresponds here to 0.3 ns. It is not possible that the data standard deviation of μ is systematically increased by the fit procedure itself, as the fit is performed the same way for the data and the MC. The only explanation is then, that the data PMTs show an effective constant time offset relative to each other. This has also been the case for the measurement of the effective refractive index in Section 4.5, but in this ^{11}C analysis the ToF correction on the reconstructed position is explicitly included in the same way as for the production of the CID $\cos\alpha$ histograms. This should be a better estimation, as it includes the effective detector response throughout the full data taking time of the respective phase, instead of just a small number of calibration source runs close in time.

The origin of the relative time offset between the PMTs is not known. It could be from a small, systematic misalignment of the PMT laser calibration system [97], differences in the PMT rise time and also an effect of the position reconstruction, which is known to be asymmetric in data (see. Section 4.5.3). It could be a combination of any of those and other effects, but ultimately the origin of this relative time offset between the data PMTs is not important. It is only important that the data and the MC behave systematically different, as this systematic difference translates into a difference between the data and MC $\cos\alpha$ histograms, which depend on the relative hit times between the PMTs. Consequently, the relative, constant time offset $(\mu - \langle\mu\rangle)$ of the data PMTs should be corrected. This correction introduces a separate systematic uncertainty which is studied in Section 7.6.3, as $(\mu - \langle\mu\rangle)$ is only known with an uncertainty $\Delta\mu$.

The last thing that has to be noted here is that $(\mu - \langle\mu\rangle)$ is calculated on ^{11}C tagged events, which are positron events. The neutrino signal used in the CID analysis comes from the neutrino recoil electrons. This difference is not a problem because only the relative, constant time differences between each PMT are relevant for the calculation of the Nth-Hit value. This is the case here as the hit time distributions are produced from the same ^{11}C tagged events for each and every PMT. The absolute time differences between different event species are not important for the Nth-Hit variable, as each event provides a $\cos\alpha$ value for every Nth-Hit,

independently from all other events.

7.5 Measurement of $N_{\mathrm{CNO+pep}}$ without systematic uncertainties

In the following section the full analysis chain for the measurement of the number of CNO+pep-neutrinos $N_{\mathrm{CNO+pep}}$ is performed for a first time. This is done without including the various systematic effects and uncertainties, as to explain the analysis procedure.

First, the calibration of the effective Cherenkov group velocity correction $\mathrm{gv_{ch}}$ is performed for Phase 1 and Phase 2+3 on their respective ^{7}Be-ν edge region. By applying a pull-term on the expected number of solar neutrinos N_ν and treating them as a nuisance parameter it is then possible to produce a χ^2-profile for $\mathrm{gv_{ch}}$. The two χ^2-profiles for Phase 1 and Phase 2+3 represent an effective calibration of $\mathrm{gv_{ch}}$ which could include unknown systematic differences between Phase 1 and Phase 2+3 and as such they are then used independently to constrain $\mathrm{gv_{ch}}$ for the measurement of $N_{\mathrm{CNO+pep}}$ of Phase 1 and Phase 2+3.

All $\cos\alpha$ histograms used in the CID analyses are produced with the PMT selection and time offset correction $(\mu(\mathrm{PMT}) - \langle\mu\rangle)$ described in the previous Section 7.4, but without the systematic uncertainty from $\Delta\mu(\mathrm{PMT})$. In this section only the plots for Phase 2+3 are shown due to their higher statistics, while the corresponding plots of Phase 1 can be found in the Appendix C.3.

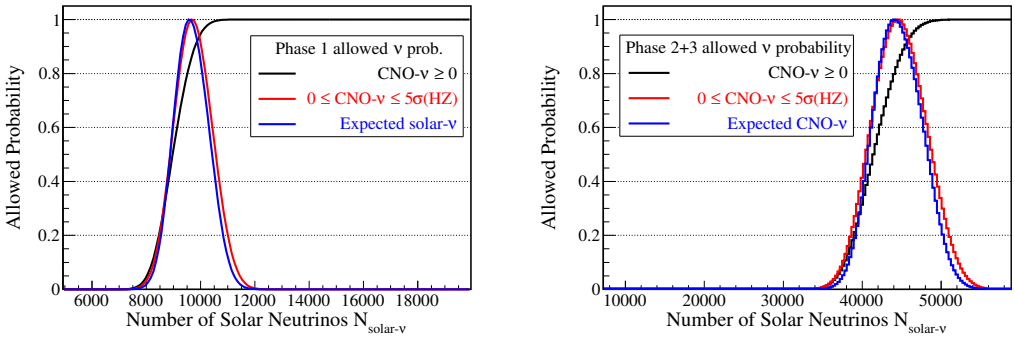
7.5.1 Calibration of $\mathrm{gv_{ch}}$ at the ^{7}Be-ν edge region

The CID analysis of the ^{7}Be-ν edge region is used to calibrate the effective group velocity correction $\mathrm{gv_{ch}}$ in this section. The selection of the data events in the FV and RoI, as well as the production of the CID MC for the ^{7}Be-ν signal and ^{210}Bi background is described in the previous Section 6.1. Phase 1 has $N_{\mathrm{total}} = 19904$ and Phase 2+3 has $N_{\mathrm{total}} = 58728$ selected data events at the ^{7}Be-ν edge region.

$$\begin{aligned}\chi^2_{\mathrm{gv_{ch}}}(N_\nu, \mathrm{gv_{ch}}, \Delta r_{\mathrm{dir}}) &= \sum_{n=1}^{\text{Nth-Hit(max)}} \sum_{i=1}^{I} \left(\frac{(\mathrm{norm}\cdot M_i^n - D_i^n)^2}{\mathrm{norm}\cdot M_i^n + \mathrm{norm}^2 \cdot M_i^n} \right) - 2\ln(P(N_\nu)) \\ \chi^2_{\mathrm{gv_{ch}}}(\mathrm{gv_{ch}}) &= \min_{N_\nu, \Delta r_{\mathrm{dir}}} \left[\chi^2_{\mathrm{gv_{ch}}}(N_\nu, \mathrm{gv_{ch}}, \Delta r_{\mathrm{dir}}) \right]\end{aligned} \tag{7.5}$$

The χ^2-test statistic for the calibration of $\mathrm{gv_{ch}}$ is given by Equation 7.5. Here, $I = 60$ is the selected binning for the $\cos\alpha$ histograms and the selected number of Nth-Hits until the analysis is performed is Nth-Hit(max) = 15 for Phase 1 and Nth-Hit(max) = 17 for Phase 2+3. The number of histogram entries at the $\cos\alpha$ bin i of the Nth-Hit $= n$ is given by D_i^n and M_i^n for data and MC, respectively, where "norm" is the scaling factor between the MC and data event statistics. The term "$\mathrm{norm}^2 \cdot M_i^n$" in the denominator takes into account the finite statistics of MC. The number of solar neutrinos N_ν are constrained according to the distribution $P(N_\nu)$ from the SSM expected values and their corresponding uncertainties. It is then possible to produce the $\chi^2(\mathrm{gv_{ch}})$ profile by treating N_ν and the mis-reconstruction in direction Δr_{dir} as the nuisance parameters, where for each given $\mathrm{gv_{ch}}$ the three dimensional $\chi^2_{\mathrm{gv_{ch}}}(N_\nu, \mathrm{gv_{ch}}, \Delta r_{\mathrm{dir}})$ is minimised over N_ν and Δr_{dir}.

Table 7.6 lists the HZ-SSM and LZ-SSM expected number of solar neutrino events for Phase 1 and Phase 2+3 according to their respective FV and selected energy region at the ^{7}Be -ν edge. Here, the MLP-cut is applied, but no TFC-cut. Using these numbers Figure 7.10 then shows the allowed probability distributions $P(N_\nu)$ for Phase 1 on the left and Phase 2+3 on the right. The SSM uncertainty is assumed to be Gaussian and the HZ and LZ



(a) (b)

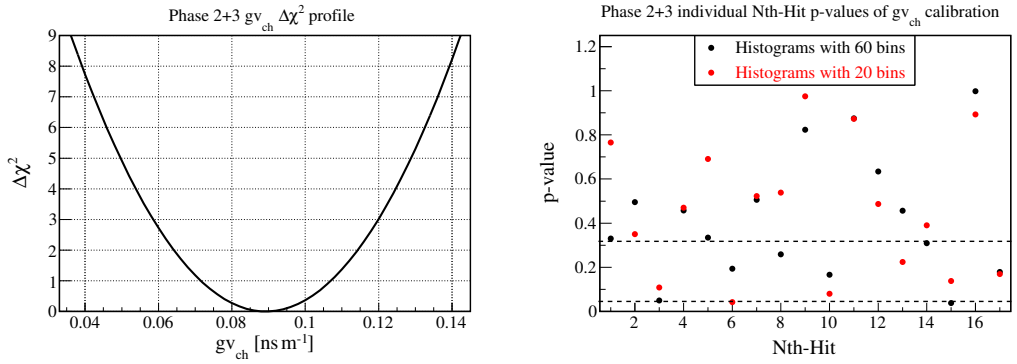
FIGURE 7.10: Different possible distributions for the allowed number of neutrino events, used as a pull-term for the gv_{ch} calibration. Blue: Normalised sum of the LZ-SSM and HZ-SSM expected values of ^{7}Be + pep + CNO-neutrinos. Black: LZ-SSM and HZ-SSM expected values of ^{7}Be + pep, where CNO is left free. Red: The LZ-SSM and HZ-SSM expected values of ^{7}Be + pep, where CNO is uniformly distributed between zero and the upper 5σ confidence level of HZ-SSM. This is the selected distribution for the pull-term on the gv_{ch} calibration. (a) For Phase 1. (b) For Phase 2+3.

	$N_{^7\mathrm{Be}}$ (HZ)	$N_{^7\mathrm{Be}}$ (LZ)	N_{pep} (HZ)	N_{pep} (LZ)	N_{CNO} (HZ)
Phase 1	9059 ± 530	8264 ± 473	423 ± 6	429 ± 6	705 ± 112
Phase 2+3	41146 ± 2405	37555 ± 2208	1958 ± 29	1986 ± 29	3342 ± 530

TABLE 7.6: Expected number of solar neutrinos at ^{7}Be-ν edge region for Phase 1 and Phase 2+3. The total number of data events is 19904 for Phase 1 and 58728 for Phase 2+3.

expectations are treated as equally likely, by summing up their respective distributions and then normalising the summed distribution. This creates the blue distributions for the expected number of solar neutrino events. These distributions are not used as a pull-term for the gv_{ch} calibration, because it makes an explicit assumption about the possible number of CNO-ν events, which later could introduce a bias for the CID measurement of CNO-ν with a pep-ν constraint. Instead two other options are presented in black and red. For the black distribution the number of CNO-ν events is left entirely free from zero to the total number of measured data events and for the red distribution the number of CNO-ν events is constrained from zero to a value corresponding to the upper 5σ confidence level of the HZ-SSM. This means that the distribution of $N_{^7\mathrm{Be+pep}}$ is convoluted with a flat distribution of $0 < N_{\mathrm{CNO}} < N_{\mathrm{total}} - N_{^7\mathrm{Be+pep}}$ for the black distribution and $0 < N_{\mathrm{CNO}}(\mathrm{Phase\ 1}) < 1265$, $0 < N_{\mathrm{CNO}}(\mathrm{Phase\ 2+3}) < 5992$ for the red distribution of Phase 1 and Phase 2+3, respectively. This red distribution is then used as the actual $P(N_\nu)$ constraint on the number of expected neutrino events, because it is a tighter constraint than leaving the CNO-ν rate entirely free, while at the same time it is very unlikely to introduce a bias that could affect the CID measurement of CNO-ν with a pep-ν constraint later on. An important note is that these distributions explicitly include the statistical uncertainty on the expected number of neutrinos and not only the SSM uncertainty.

Performing the gv_{ch} calibration according to Equation 7.5 produces the $\Delta\chi^2(gv_{ch})$ profile shown in Figure 7.11(a) for Phase 2+3. The best fit value is $gv_{ch} = (0.089 \pm 0.018)\,\mathrm{ns\,m^{-1}}$, where the uncertainty is estimated from $\Delta\chi^2(gv_{ch}) = \pm 1$, corresponding to a Gaussian 68.3% (1σ) CI. This value is in agreement with the results of the γ calibration sources of $gv_{ch} = (0.108 \pm 0.039)\,\mathrm{ns\,m^{-1}}$ (see Chapter 5), even though both calibration methods and Nth-Hit(max) selection are very different. Overall, the best fit MC $\cos\alpha$ histogram is able to



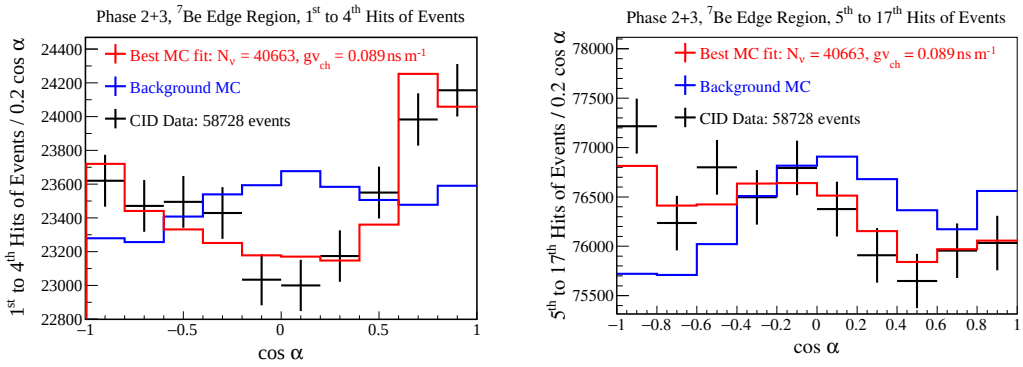
(a) (b)

FIGURE 7.11: (a) Resulting $\Delta\chi^2(\mathrm{gv_{ch}})$ profile of Phase 2+3, with $\chi^2/\mathrm{ndf} = 1036.3/1017$, p-value = 0.33. The best fit value is $\mathrm{gv_{ch}} = (0.089 \pm 0.018)\,\mathrm{ns\,m^{-1}}$. (b) The χ^2 corresponding p-values of each individual Nth-Hit $\cos\alpha$ histogram between data and MC. The MC histograms are not fitted individually on data, but use the best fit results of the full Nth-Hit(max) = 17 analysis. The p-values are calculated for a number of bins of 60 (black) and 20 (red). The dotted lines correspond to p-value = 0.32 andp-value = 0.05.

explain data well, with an absolute $\chi^2/\mathrm{ndf} = 1036.3/1017$, p-value = 0.33. Using the best fit results of the analysis with 60 bins per histogram and reducing the number of bins to 20 gives $\chi^2/\mathrm{ndf} = 336.4/337$, p-value = 0.50, which also shows a good agreement between data and MC. If there would be large scale structures in the $\cos\alpha$ histograms that are systematically different for data and MC, then a binning of 60 could potentially hide those due to lower statistics per bin, but this is not the case here.

For the same reason, the corresponding, combined p-value of 17 $\cos\alpha$ histograms with 60 bins each could mask a potential difference between data and MC for the $\cos\alpha$ histograms at the individual Nth-Hits. Figure 7.11(b) shows the χ^2 corresponding p-values for each individual $\cos\alpha$ histogram at a given Nth-Hit for a binning of 60 in black and 20 in red. This is not an independent analysis of each individual Nth-Hit, but it is using the best fit results of the full Nth-Hit(max) = 17 analysis. The dotted lines represent p-value = 0.32 and p-value = 0.05. The number of points with p-value < 0.32 is six and zero with p-value < 0.003 out of 17, for both 60 bins and 20 bins histograms. This means that no single $\cos\alpha$ histogram is significantly different between data and MC and the overall goodness of fit performance is satisfactory, as statistically about 5.4 values with p-value < 0.32 are expected, while six are realised.

It is not easy possible to give an intuitive, graphical understanding of the fit performance by showing all 17 histograms with 60 bins each. For this reason Figure 7.12 instead shows the sum of the early Nth-Hits and late Nth-Hits $\cos\alpha$ histograms, with 10 bins each. On the left the sum of the first to fourth Nth-Hits is shown and on the right the sum of the fifth to 17$^{\mathrm{th}}$ Nth-Hits. The data is shown in black, while the best fit MC is shown in red. For comparison, the pure background MC is shown in blue. In the early hits the Cherenkov peak is clearly visible for $\cos\alpha \sim 0.7$, as well as the second peak with a negative slope for $-1 < \cos\alpha < 0.0$. This second peak comes from the effect of Δr_{dir}, as explained in Section 6.2.1 and it is also clearly visible for the later Nth-Hits on the right side. This shows that the optimised selection of Nth-Hit(max) is able to use the indirect Cherenkov information of late Nth-Hits through Δr_{dir}, to provide a non-negligible sensitivity for the CID analysis, even though no peak from direct Cherenkov hits is visible. The shape of the background $\cos\alpha$ histograms is defined by the position distribution of the selected, live PMTS. The goodness of fit performance and observations on the $\cos\alpha$ shape are qualitatively also true for Phase 1, whose plots can be



(a) (b)

FIGURE 7.12: Summed $\cos\alpha$ histograms of the Phase 2+3 ^{7}Be region used for the gv_{ch} calibration. data is shown in black, pure ^{210}Bi background MC in blue and in red the best fit MC histogram with $gv_{ch} = 0.089\,\mathrm{ns\,m^{-1}}$, $N_\nu = 40663$ fitted signal events and 18065 background events out of 58728 events in total. (a) The sum of the first to fourth Nth-Hits. The peak of direct Cherenkov hits is visible around $\cos\alpha \sim 0.7$. (b) The sum of the fifth to 17th Nth-Hits. The negative $\cos\alpha$ slope of neutrino events due to $\Delta r_{dir} = 1.66\,\mathrm{cm}$ is visible in data. The non-flat $\cos\alpha$ shape of the MC background comes from the distribution of selected, live PMTs.

found in Appendix C.3.

Table 7.7 summarises the results of the gv_{ch} calibration at the ^{7}Be-ν edge region. Comparing the Phase 1 result of $gv_{ch}(\text{Phase 1}) = (0.138 \pm 0.028)\,\mathrm{ns\,m^{-1}}$ with the result $gv_{ch}(\gamma\text{-calib.}) = (0.108 \pm 0.039)\,\mathrm{ns\,m^{-1}}$ from the γ sources shows that they are in agreement with 0.6σ, while the comparison between the Phase 1 and the Phase 2+3 results gives a difference of 1.5σ.

	χ^2/ndf	p-value	gv_{ch} [ns m^{-}1]	N_ν
Phase 1	868.3/897	0.75	0.138 ± 0.028	9400
Phase 2+3	1036.3/1017	0.33	0.089 ± 0.018	40663

TABLE 7.7: Results of the gv_{ch} CID calibration without systematic uncertainties. The total number of data events is 19904 for Phase 1 and 58728 for Phase 2+3.

The gv_{ch} parameter and its calibration can be considered a robust estimation of the underlying Cherenkov photon properties that are systematically different between data and MC. The calibration at the ^{7}Be-ν edge has good p-values for the full analysis and for each individual Nth-Hit, meaning that the MC model is able to correctly reproduce the $\cos\alpha$ shape of data. In total, the three independent gv_{ch} measurements of Phase 1, Phase 2+3 and on the γ sources give agreeable results, even for the drastically different analysis method between the CID approach and the γ sources, as well as the different selection of Nth-Hit(max) between Phase 1 and Phase 2+3. Due to the 1.5σ difference of the Phase 1 and Phase 2+3 results a conservative approach is taken in the following CID analysis of the CNO+pep-ν energy region, where gv_{ch} is constrained only by the results of the corresponding phase and not by the weighted average of all gv_{ch} measurements. This approach keeps the statistical uncertainty higher than it would be for a combined gv_{ch} pull-term, but it can potentially account for small, systematic differences in gv_{ch} between Phase 1 and Phase 2+3, should they exist.

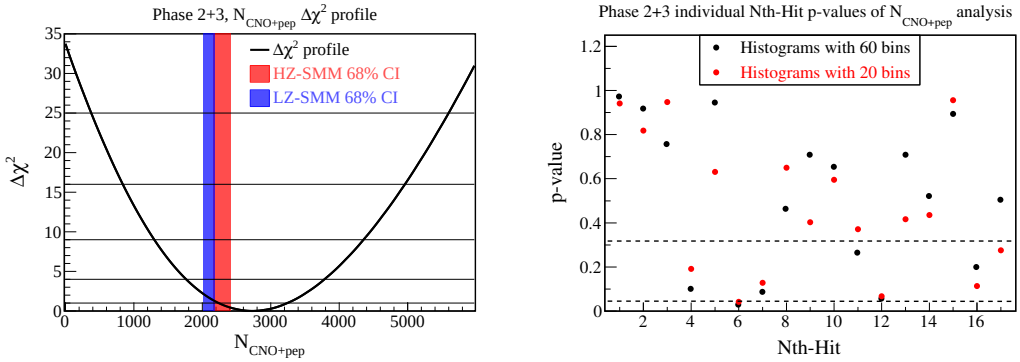
7.5.2 Measurement of $N_{\mathrm{CNO+pep}}$

$$\chi^2_\nu(N_{\mathrm{CNO+pep}}, \mathrm{gv}_{\mathrm{ch}}, \Delta r_{\mathrm{dir}}) = \sum_{n=1}^{\text{Nth-Hit(max)}} \sum_{i=1}^{I} \left(\frac{(\mathrm{norm} \cdot M_i^n - D_i^n)^2}{\mathrm{norm} \cdot M_i^n + \mathrm{norm}^2 \cdot M_i^n} \right) + \Delta\chi^2_{\mathrm{gv}_{\mathrm{ch}}}(\mathrm{gv}_{\mathrm{ch}})$$

$$\chi^2_\nu(N_{\mathrm{CNO+pep}}) = \underset{\mathrm{gv}_{\mathrm{ch}}, \Delta r_{\mathrm{dir}}}{\arg\min} \left[\chi^2_\nu(N_{\mathrm{CNO+pep}}, \mathrm{gv}_{\mathrm{ch}}, \Delta r_{\mathrm{dir}}) \right] \tag{7.6}$$

The CID analysis of $N_{\mathrm{CNO+pep}}$ is performed using the optimised RoI selection from Section 7.1. Phase 1 has $N_{\mathrm{total}} = 2990$ and Phase 2+3 has $N_{\mathrm{total}} = 5974$ selected data events in the CNO+pep-ν RoI and FV. A χ^2-test statistic is used to compare the data and MC $\cos\alpha$ histograms, as shown in Equation 7.6.

The selected binning for the histograms is $I = 60$ and the selected number of Nth-Hits until the analysis is performed is Nth-Hit(max) = 15 for Phase 1 and Nth-Hit(max) = 17 for Phase 2+3. The number of entries at the $\cos\alpha$ bin i of the n^{th} hits of the events is given by D_i^n[2] and M_i^n for data and MC, respectively, where "norm" is the scaling factor between the MC and data event statistic. The term "$\mathrm{norm}^2 \cdot M_i^n$" in the denominator takes into account the finite statistics of MC. The value of $\mathrm{gv}_{\mathrm{ch}}$ is constrained using the $\Delta\chi^2_{\mathrm{gv}_{\mathrm{ch}}}(\mathrm{gv}_{\mathrm{ch}})$ profile, produced from the CID analysis at the ^{7}Be-ν edge region in the previous section. By treating $\mathrm{gv}_{\mathrm{ch}}$ and Δr_{dir} as nuisance parameters, the three dimensional $\chi^2_\nu(N_{\mathrm{CNO+pep}}, \mathrm{gv}_{\mathrm{ch}}, \Delta r_{\mathrm{dir}})$ is minimised over $\mathrm{gv}_{\mathrm{ch}}$, Δr_{dir} to a one dimensional $\chi^2_\nu(N_{\mathrm{CNO+pep}})$ profile.



(a) (b)

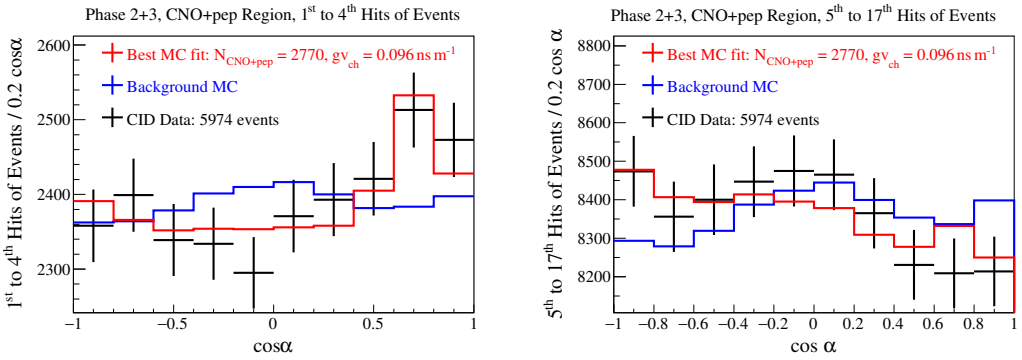
FIGURE 7.13: (a) Resulting $\Delta\chi^2(N_{\mathrm{CNO+pep}})$ profile of Phase 2+3, with $\chi^2/\mathrm{ndf} = 1000.5/1017$, p-value = 0.64. The best fit value is $N_{\mathrm{CNO+pep}} = 2770^{+515}_{-491}$ out of 5974 events. The red and blue bands represents the 68.3% CI of the SSM expectations of HZ and LZ, respectively. (b) The χ^2 corresponding p-values of each individual Nth-Hit $\cos\alpha$ histogram between data and MC. The MC histograms are not fitted individually on data, but use the best fit results of the full Nth-Hit(max) = 17 analysis. The p-values are calculated for a number of bins of 60 (black) and 20 (red). The dotted lines correspond to p-value = 0.32 and p-value = 0.05.

The $\Delta\chi^2(N_{\mathrm{CNO+pep}})$ profile of this CID analysis is shown in Figure 7.13(a) for Phase 2+3. The best fit value is $N_{\mathrm{CNO+pep}} = 2770^{+515}_{-491}$ out of 5974 total events, where the uncertainty is estimated from $\Delta\chi^2(\mathrm{gv}_{\mathrm{ch}}) = \pm 1$, corresponding to a Gaussian equivalent 68.3% (1σ) CI. This best fit $N_{\mathrm{CNO+pep}}$ is well in agreement with the HZ-SSM ($N^{\mathrm{HZ}}_{\mathrm{CNO+pep}} = 2293 \pm 125$) with a 0.9σ difference, while The LZ-SSM ($N^{\mathrm{LZ}}_{\mathrm{CNO+pep}} = 2092 \pm 85$) has a difference of 1.4σ. The no-neutrino hypothesis can be excluded with $\Delta\chi^2 = 34.8$, $> 5\sigma$. The best fit parameters

[2]The continuity correction for the Gaussian approximation of the Poisson distribution ($D_i^n + 0.5$) does not impact the results, as the expected entries per bin are about 50 and 100 for Phase 1 and Phase 2+3, respectively.

provide a $\chi^2/\text{ndf} = 1000.5/1017$, p-value = 0.64 and reducing the number of bins to 20 for the same best fit results gives $\chi^2/\text{ndf} = 331.2/337$, p-value = 0.58. There is a good agreement between data and MC, but these values of $\chi^2/\text{ndf} < 1$ could be an indication of some sort of small, but not critical overfitting. It is possible that here the uncertainty from the bin statistics is relatively large compared to the model differences between MC histograms with different gv_{ch}, Δr_{dir} values.

Figure 7.13(b) shows the χ^2 corresponding p-values for each individual $\cos\alpha$ histogram for a given Nth-Hit with a binning of 60 and 20 in black and red, respectively. These values are calculated with the best fit results of the CID $N_{\text{CNO+pep}}$ analysis using 60 bins and Nth-Hit(max) = 17. The dotted lines represent p-value = 0.32 and p-value = 0.05. The number of p-values < 0.32 is six and there are zero p-values < 0.003, for both histogram binning selections. The early Nth-Hits < 5 in particular have large p-values, independently from the later Nth-Hits. Thus, the observation of $\chi^2/\text{ndf} < 1$ is not driven by noisy, late Nth-Hits, which are expected to have only small Cherenkov hit features. This is good, because it means that the gv_{ch}, Δr_{dir} parameters that govern the MC $\cos\alpha$ shape are likely not strongly influenced by random fluctuations in data of late Nth-Hits. The overall goodness of fit performance is satisfactory, meaning that the CID MC model is well able to describe the data.



(a) (b)

FIGURE 7.14: (a) Summed $\cos\alpha$ histograms of the Phase 2+3 CNO+pep-ν region. data is shown in black, pure MC background in blue and in red the best fit MC histogram with $N_\nu = 2770$ fitted signal events, 3204 fitted background events out of 5974 events in total and $\text{gv}_{\text{ch}} = 0.096\,\text{ns}\,\text{m}^{-1}$. (a) The sum of the first to fourth Nth-Hits. The peak of direct Cherenkov hits is visible around $\cos\alpha = 0.7$. (b) The sum of the fifth to 17$^{\text{th}}$ Nth-Hits. The negative $\cos\alpha$ slope of neutrino events due to $\Delta r_{\text{dir}} = 2.69\,\text{cm}$ is visible in data. The non-flat $\cos\alpha$ shape of MC background comes from distribution of selected, live PMTs.

To illustrate the results of the fit and the agreement between data and MC the $\cos\alpha$ histograms of the sum of early Nth-Hits and late Nth-Hits are shown in Figure 7.14. For the sum of the first to fourth Nth-Hit histograms the peak from direct Cherenkov hits is visible at $\cos\alpha = 0.7$. For the sum of the late hits from the fifth to the 17$^{\text{th}}$ Nth-Hit, the indirect effect of the Cherenkov hits, Δr_{dir} is also visible by eye. The best fit MC histogram explains the negative slope better than the pure background. The non flat $\cos\alpha$ distribution of MC background is governed by the position distribution of the selected live PMTs, which is well described in MC as it is simulated on an data event-by-event basis. The observations made on the $N_{\text{CNO+pep}}$ CID analysis of Phase 2+3 are qualitatively also true for Phase 1, where the CID data is well explained by the best fit MC histogram also for individual Nth-Hits. The relevant plots can be found in the Appendix C.3.

The results of the CID $N_{\text{CNO+pep}}$ measurement are summarised in Table 7.8 and the Phase 1 value $N^{\text{Phase 1}}_{\text{CNO+pep}} = 676^{+235}_{-224}$ is well in agreement with both the HZ-SMM ($N^{\text{HZ}}_{\text{CNO+pep}} = 615 \pm 34$)

and LZ-SSM ($N^{\mathrm{HZ}}_{\mathrm{CNO+pep}} = 561 \pm 25$). For Phase 1 the no-neutrino hypothesis can be excluded with $\Delta\chi^2 = 10.3, > 3\sigma$.

	χ^2/ndf	p-value	$N_{\mathrm{CNO+pep}}$
Phase 1	886.9/897	0.59	676^{+235}_{-224}
Phase 2+3	1000.5/1017	0.64	2770^{+515}_{-491}

TABLE 7.8: Results of the $N_{\mathrm{CNO+pep}}$ CID analysis without systematic uncertainties. The total number of data events is 2990 for Phase 1 and 5974 for Phase 2+3.

A last observation is the difference of the best fit Δr_{dir} values between the phases and energy regions. For Phase 1 these values are $\Delta r_{\mathrm{dir}} = (1.89 \pm 0.19)$ cm and $\Delta r_{\mathrm{dir}} = (3.20 \pm 0.77)$ cm for the ^{7}Be-ν edge and CNO+pep-ν region, respectively. For Phase 2+3 these values are $\Delta r_{\mathrm{dir}} = (1.66 \pm 0.12)$ cm and $\Delta r_{\mathrm{dir}} = (2.69 \pm 0.31)$ cm. The Δr_{dir} values between Phase 1 and Phase 2+3 are comparable with each other due to the relatively large uncertainty, but Phase 1 shows consistently higher values for both energy regions. This is expected, as the loss of PMTs over time increases the position reconstruction uncertainty, which should decrease the effect of Δr_{dir}. The Δr_{dir} values are larger for the CNO+pep-ν region, which is a significant effect for the combined differences of Phase 1 and Phase 2+3. This is also expected, as the larger electron energies at the CNO+pep-ν energy regions result in a higher Cherenkov/scintillation ratio, which gives a bigger Δr_{dir} mis-reconstruction in the initial electron direction.

Here, the detailed study of systematic uncertainties is still outstanding. Neither are these the final results of the $N_{\mathrm{CNO+pep}}$ CID measurement, nor is the number of pep-ν events constrained to perform a measurement of the CNO-ν rate. The different systematic uncertainty studies are explained in the following Section 7.6 and 7.7. The final results of the CID analysis of $N_{\mathrm{CNO+pep}}$ and N_{CNO}, including systematic uncertainties, are given in Section 7.8.

7.6 Estimation of systematic uncertainties

The biggest systematic uncertainty of the CID analysis comes from the uncertainty of the parameters $\mathrm{gv}_{\mathrm{ch}}$ and Δr_{dir}, which are used to model the data $\cos\alpha$ histograms. Their influence is automatically included in the $\Delta\chi^2$ profiles because they are treated as nuisance parameters in the fit. The other possible sources of systematic uncertainties can be categorised into two groups. First, there are those sources which only influence the CID measurement regarding the number of neutrino events in the selected data set:

- The selection of Nth-Hit(max) until the analysis is performed.
- The selection of the number of $\cos\alpha$ histogram bins.
- The selection of good PMTs to be considered for the analysis.
- The relative time alignment between different data PMTs.
- The position distribution of different background species.
- The $\cos\alpha$ shape differences between pep-ν and CNO-ν events.
- The expected number of neutrino events that are not of interest.

The second category of systematic uncertainty sources are those that impact the calculation of the neutrino rate [cpd / 100 t] from the measured number of neutrino events and vice versa:

- The selection of the fiducial volume.
- The selection of the energy region of interest.
- The exposure efficiency of the MLP and TFC algorithms.

For the previous CID measurement of ^{7}Be-neutrinos the uncertainties have been estimated with simple, but maximally conservative methods. The intention of this previous measurement has been to provide a robust proof of principle for the CID method and the overestimation of the systematic uncertainties does not impede this strategic analysis goal. As the primary goal of this analysis is to provide an actual measurement of CNO-neutrinos, now the listed sources are investigated in detail and with a reasonable estimation of their systematic uncertainties values. The methods used for these investigations are toy-MC studies and the comparison of pure data background events outside the selected energy RoI with MC background events.

7.6.1 Selection of Nth-Hits(max) and number of histogram bins

The number of Nth-Hit(max) $\cos\alpha$ histograms used for the CID analysis is optimised to take into account the sensitivity on direct Cherenkov hits, as well as the indirect information of Cherenkov hits, provided through the biased position mis-reconstruction Δr_{dir}. In principle it is possible to use all hits of the events, as each Nth-Hit is independent from the others and starting from some late Nth-Hits, they only provide random noise and no Cherenkov information of any kind. Nonetheless the analysis is constrained to Nth-Hit(max) $= 15$ for Phase 1 and Nth-Hit(max) $= 17$ for Phase 2+3, which is a somewhat arbitrary selection and an equally good Nth-Hit(max) value could have been selected between about Nth-Hit(max) ≈ 10 to 20, or more. The reason that very large Nth-Hit(max) values are not considered, is that there could be systematic differences between the $\cos\alpha$ histograms of data and MC, dependent on the Nth-Hits, without adding more CID sensitivity. Such systematic differences could stem from the MC and data background for example, where even a very small difference, that is negligible for individual Nth-Hits, would dominate the analysis for the use of hundreds of Nth-Hits. Likewise, the selection of the number of bins for the $\cos\alpha$ histogram is also somewhat arbitrary. The existence of a systematic uncertainty through the choice of binning seems unlikely, but it could be possible that some large scale $\cos\alpha$ structures are different between the data and the MC and hidden through the lower bin statistics for a larger binning. Conversely there could exist systematic $\cos\alpha$ outliers in data, which then would be averaged out and effectively hidden by the use of too few bins.

Therefore, the impact of the selection of Nth-Hit(max), as well as the binning on the CID result and systematic uncertainties are studied here, through the use of a toy-MC. The toy-MC used here is the same as for the *Acceptance Sampling Unfolding* used to unfold the fit response, described in the later Section 7.7 and the procedure is as follows:

1. The values of the CID parameters N_ν^{sim}, $\mathrm{gv}_{\mathrm{ch}}^{\mathrm{sim}}$, $\Delta r_{\mathrm{dir}}^{\mathrm{sim}}$ are selected randomly from the entire parameter space. This means that for the ^{7}Be-ν edge region N_ν^{sim} is sampled according to the allowed probability distribution $P(N_\nu)$, while for $\mathrm{gv}_{\mathrm{ch}}^{\mathrm{sim}}$, $\Delta r_{\mathrm{dir}}^{\mathrm{sim}}$ the values are sampled uniformly in a region of interest. For the CNO+pep-ν region $\mathrm{gv}_{\mathrm{ch}}^{\mathrm{sim}}$ is sampled according to the $\Delta\chi^2(\mathrm{gv}_{\mathrm{ch}})$ corresponding probability distribution $P(\mathrm{gv}_{\mathrm{ch}}) = \exp(-\frac{1}{2}\Delta\chi^2(\mathrm{gv}_{\mathrm{ch}}))$[3], while for $N_{\mathrm{CNO+pep}}^{\mathrm{sim}}$, $\Delta r_{\mathrm{dir}}^{\mathrm{sim}}$ all physically allowed values are equally likely to be selected.

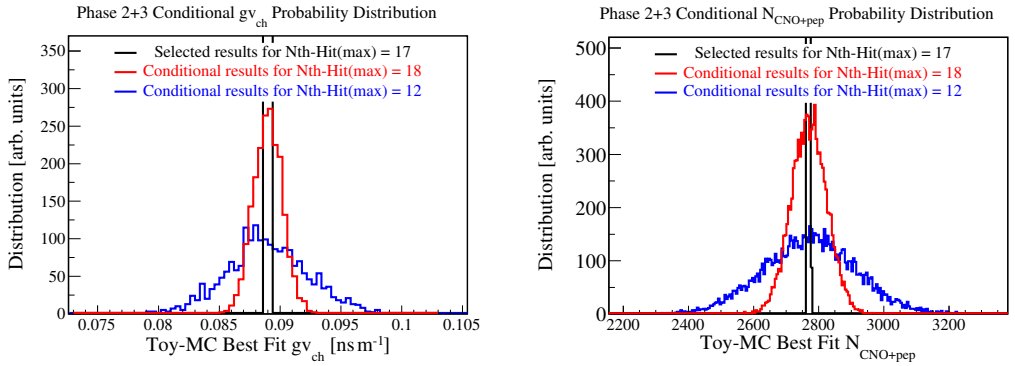
2. The toy-data is produced from MC $\cos\alpha$ histograms according to the sampled parameters and with the statistics of data. Then the toy-data is analysed in exactly the same

[3] This is the same as the Bayesian posterior from the $\mathrm{gv}_{\mathrm{ch}}$ measurement, now used as a prior.

way as the true data, with the nominal selection of Nth-Hits(max) and number of bins as well as including the relevant pull-terms. This results in the best fit values for the toy analysis as N_ν^{fit}, $\text{gv}_{\text{ch}}^{\text{fit}}$, $\Delta r_{\text{dir}}^{\text{fit}}$.

3. If all best fit values of the toy analysis are the same as the results of the true data analysis, within a small window, then the toy-data analysis is performed a second time where the selection of either Nth-Hit(max) or the number of histogram bins is changed relative to the nominal analysis. The best fit parameter distribution of this second analysis is the conditional probability distribution to see a certain result for a changed selection of Nth-Hit(max) or the number of bins, given the results from the nominal data analysis.

4. The standard deviation of the toy-MC fit result distributions then corresponds to the expected statistical uncertainty for the variation of the Nth-Hit(max) and binning selection. This makes it possible to identify a potential systematic uncertainty in fit results of the data , should they be statistically unlikely.

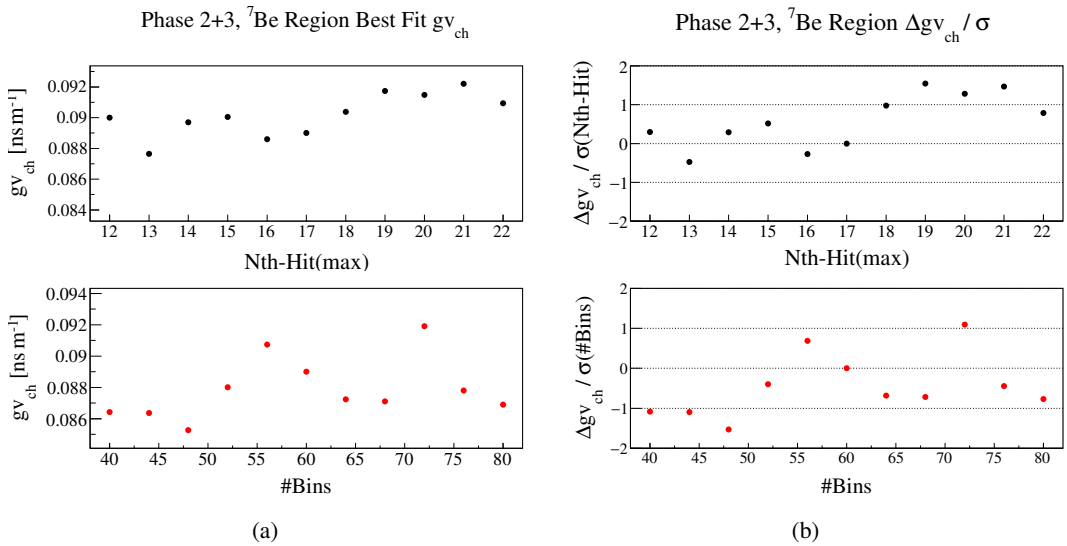


(a) (b)

FIGURE 7.15: Conditional probability distributions produced by toy-MC, given the true results of data analysis with a nominal selection of Nth-Hit(max) = 17 in black. For the toy-MC analyses with Nth-Hit(max) = 18 in red and Nth-Hit(max) = 12 in blue. (a) For the gv_{ch} calibration at the ^{7}Be-ν edge region, with 2000 events. (b) For the $N_{\text{CNO+pep}}$ analysis, with 10k events.

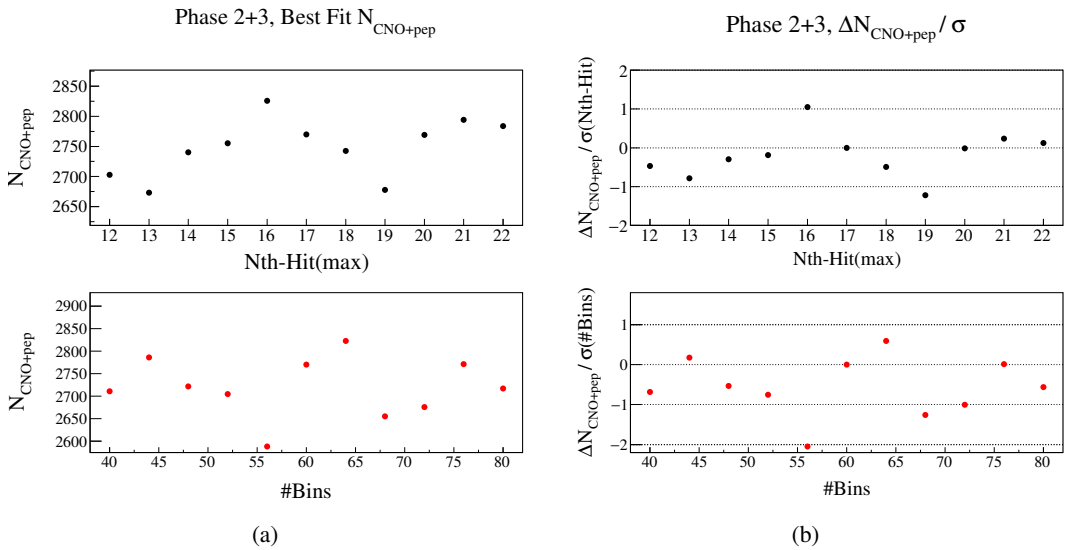
Figure 7.15 shows the toy-MC produced, conditional PDFs for the nominal selection of Nth-Hit(max) = 17 in black, Nth-Hit(max) = 18 in red and Nth-Hit(max) = 12 in blue, for the ^{7}Be region gv_{ch} calibration on the left and for the $N_{\text{CNO+pep}}$ measurement on the right. The black distribution shows the width of the acceptance region, i.e. the toy-MC results must give the same results as the true data analysis within this small window. Given these results of the pseudo-data analysis the conditional PDFs show that the mean value stays the same, independent of Nth-Hit(max), as is expected. The larger the difference between the investigated Nth-Hit(max) and the nominal selection is, the larger the standard deviation becomes, which is also as expected. While it is shown here that Nth-Hit(max) = 12 is much broader than Nth-Hit(max) = 18, this is also true for all other values, Nth-Hit(max) = 22 is about as broad as Nth-Hit(max) = 12 and so on. The PDFs for the binning selection and for Phase 1 are not shown, as they are qualitatively the same.

Given the results of the CID analysis at Nth-Hit(max) = 17 of $N_{\text{CNO+pep}} = 2770$, what is the probability to get the result of $N_{\text{CNO+pep}} = 2742$ at Nth-Hit(max) = 18? With the expected statistical uncertainty provided by the conditional PDFs of the toy-MC it is possible to search for systematic uncertainties from the Nth-Hit(max) and binning selection. This is shown for Phase 2+3 in Figure 7.17, for the ^{7}Be region gc_{ch} calibration at the top and $N_{\text{CNO+pep}}$ measurement at the bottom. The corresponding plots of Phase 1 are in the Appendix C.10. The results



(a) (b)

FIGURE 7.16: (a) Results of the gv_{ch} calibration at the ^{7}Be-ν edge region, for Phase 2+3. On the top (black) for a range of Nth-Hit(max) values around the nominal value of Nth-Hit(max) = 17, all with 60 bins for the $\cos\alpha$ histogram. On the bottom (red) for a range of number of bins around the nominal value of 60 bins, all with Nth-Hit(max) = 17. (b) Difference between the varied gv_{ch} results and the nominal analysis result of $gv_{ch} = 0.089\,\mathrm{ns\,m^{-1}}$ in units of the expected standard deviation.



(a) (b)

FIGURE 7.17: (a) Results of the $N_{CNO+pep}$ measurement, for Phase 2+3. On the top (black) for a range of Nth-Hit(max) values around the nominal value of Nth-Hit(max) = 17, all with 60 bins for the $\cos\alpha$ histogram. On the bottom (red) for a range of number of bins around the nominal value of 60 bins, all with Nth-Hit(max) = 17. (b) Difference between the varied $N_{CNO+pep}$ results and the nominal analysis result of $N_{CNO+pep} = 2770$ in units of the expected standard deviation.

of the true data analysis are shown on left for a range of Nth-Hit(max) = $\{12, 13, ..., 21, 22\}$ and #Bins = $\{40, 44, ..., 76, 80\}$. On the right the difference between the results of the current Nth-Hit(max), binning selection and the result of the nominal selection is shown in units of the expected statistical uncertainty. No significant pattern or structure can be seen for any of the plots. Between Phase 1 and Phase 2+3 some points are outside of 1σ, two points in total are outside of 2σ and none are outside of 3σ.

Results $> \lvert 1\sigma \rvert$	Nth-Hit(max)		#Bins		Total
	gv_{ch}	$N_{CNO+pep}$	gv_{ch}	$N_{CNO+pep}$	
Phase 1	2	4	4	3	25
Phase 2+3	3	4	2	3	

TABLE 7.9: Number of results, where a varied Nth-Hit(max), #Bins fit result differs from the nominal analysis by more than 1σ. The expected number is 3.2 for each table entry and 25.4 in total.

This is summarised in Table 7.9 for Phase 1 and Phase 2+3, where the number of results outside 1σ is given for each corresponding plot, with the expected number of results outside of 1σ is 3.2 as each plot has ten points. In total a number of 25.4 is expected, while a number of 25 is realised. All plots show a behaviour that is fully consistent with pure statistical fluctuations and the conclusion is then that there is no relevant systematic uncertainty for the selection of Nth-Hit(max) or the selection of the number of histogram bins, in the investigated regions.

7.6.2 Selection of good PMTs

The PMTs selection is optimised for the CID analysis through the use of the ToF corrected hit time distribution of the individual PMTs in data. Fitting these hit time distributions with a Gaussian modified exponential PDF results in a distribution of the fit parameters and the PMTs are deselected based on the difference between their fit parameter values relative to the mean values of the full parameter distributions (see Section 7.4). Because the Nth-Hit time like variable used for the CID analysis is sensitive only to the relative hit time differences between the PMTs this approach is the correct way to select PMTs, as the applied cuts means that PMTs are considered good to use if they have a hit time distribution similar to most other PMTs. The stricter the parameter cuts are, the less likely it is that the selected PMTs have some systematic, relative differences in their hit time distributions between each other. The exact cuts applied for the PMT selection are somewhat arbitrary as they have been selected by eye and are not based on a rigorous toy-MC study.

For this reason it is possible that the particular PMT selections could still have some systematic differences between the relative hit time distribution of the data PMTs, which would result in a systematic uncertainty for the CID analyses. This is investigated here for Phase 2+3 by applying stricter parameter cuts on the PMT selection in such a way that an additional number of PMTs N^{-}_{PMTs} is deselected in steps of 40 PMTs, $N^{-}_{\text{PMTs}} = \{40, 80, 120, 160, 200\}$. Should there be a systematic uncertainty left from the selection of good PMTs, then such a stricter PMT selection would potentially show a difference in the CID analyses results that would be noticeably larger than what is expected from pure statistical fluctuations. This is in principle the same type of systematics study as the previous investigation of the Nth-Hit(max) and binning selection, although with some differences in the calculation of the expected standard deviation. The relevant plots of Phase 1 can be found in the Appendix C.4.

Again, a toy-MC is used to estimate the expected, purely statistical standard deviation from the deselection of an increased number of PMTs. This toy-MC can not be performed through random sampling of toy events from the $\cos\alpha$ histogram of the full MC, because

different PMT selection give different $\cos\alpha$ histograms for the same selection of MC events. Thus, the previous method is not able to produce conditional probability distributions, because the random events sampled from the $\cos\alpha$ histogram of the nominal PMT selection are independent from the events sampled from the histograms of the stricter PMT selection. Instead a more laborious toy-MC study must be performed:

1. Because the full CID MC is produced on a data event-by-event basis, here for each data event only a single corresponding MC event is randomly selected, either from the neutrino MC or from the background MC, such that the resulting MC event selection has the correct signal to background ratio. Only MC events which pass the relevant N_h^{geo} and FV cuts are considered, where the MC event selection has the same number of events as the data for each phase and RoI.

2. A number of 1000 random MC event selections are performed, where the true MC parameters N_ν, gv_{ch}, Δr_{dir} are sampled from their respective posterior distribution, given the results of the nominal PMT selection analysis results (see Section 7.7.2). This procedure makes sure that the possible underlying, true CID parameter values are taken into account correctly.

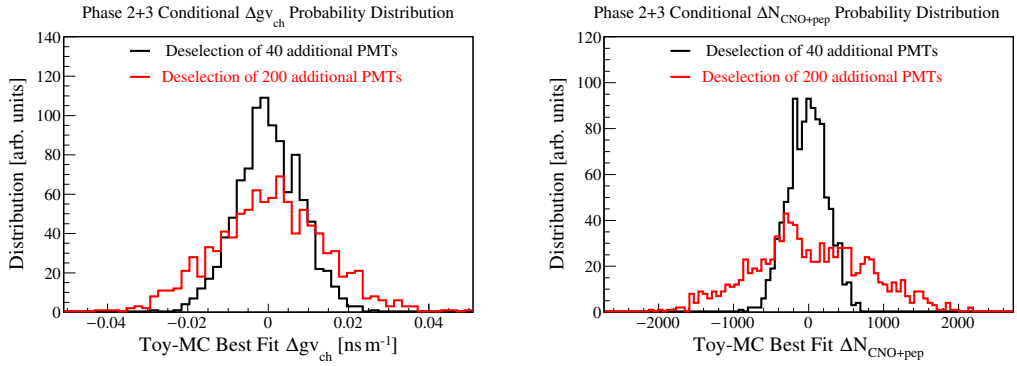
3. The full MC events are used to produce the CID $\cos\alpha$ histograms for the different PMT selections. Thus each set of toy-data events produces six $\cos\alpha$ histograms, one for the nominal and five for the stricter N^-_{PMTs} PMT selections. These toy-data $\cos\alpha$ histograms are now correlated to each other, as they use the exact same set of underlying, full MC events, just with a different selection of PMT hits from the sampled events.

4. The normal CID analysis is performed on the toy-data $\cos\alpha$ histograms with the relevant full MC histograms, which now are produced with the same PMT selection as the toy-data. Each set of toy-data events and the corresponding analysis produces six best fit values for each CID parameter. This results in five difference distributions for each CID parameter, where the toy analysis fit result given by the additionally deselected N^-_{PMTs} PMTs is subtracted with the toy analysis result of the nominal PMT selection. The standard deviation from these distributions is the desired N^-_{PMTs} dependent estimation of the purely statistical uncertainty.

5. Comparing the differences of the true data CID analysis results with this toy-MC standard deviation makes it possible to identify a potential systematic uncertainty from the method of PMT selection.

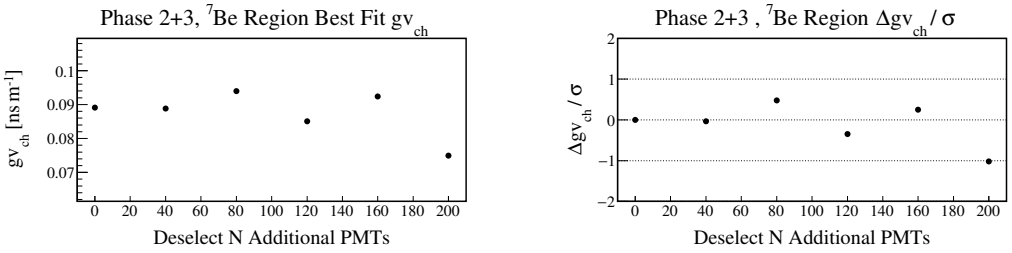
Figure 7.18 shows examples of the N^-_{PMTs} dependent distributions which give the estimation on the expected standard deviation used for the study of the PMT selection systematic uncertainty. On the left Δgv_{ch} from the ^{7}Be-ν edge region is shown and on the right $\Delta N_{CNO+pep}$. The additional deselection of $N^-_{PMTs} = 40$ PMTs is shown in black, while the additional deselection of $N^-_{PMTs} = 200$ PMTs is shown in red. The standard deviation gets larger for a larger number of additionally deselected PMTs, as is expected. It is interesting to note that for $N^-_{PMTs} = 200$ the standard deviation of $\Delta N_{CNO+pep}$ is 780, which is much larger than the estimated uncertainty from the nominal data analysis $\Delta\chi^2(N_{CNO+pep})$ profile of ~ 490, while for Δgv_{ch} the estimated standard deviation is $0.014\,ns\,m^{-1}$, which is smaller than the uncertainty from the data analysis $\Delta\chi^2(gv_{ch})$ profile of $0.018\,ns\,m^{-1}$.

It seems that the low event statistic at the CNO+pep-ν region makes the analysis more dependent on the exact distribution of selected PMTs. This likely comes from the reduced effect of averaging over a large number of different solar directions, compared to the ~ 10 times higher event statistic at the ^{7}Be-ν edge region. For a low number of events the exact distribution of the position of the Sun relative to the selected PMTs becomes more influential on the CID analysis.



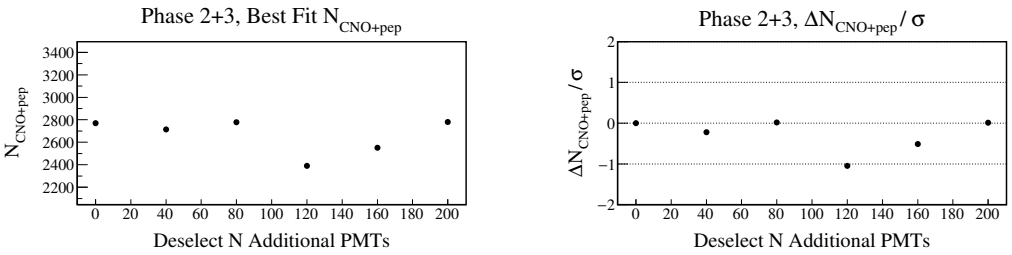
(a) (b)

FIGURE 7.18: The conditional probability distributions of the difference between the CID results obtained with the nominal PMT selection and CID results obtained with the deselection of additional PMTs. The distributions are produced with 1000 toy-MC analyses, with the additional deselection of 40 PMTs in black and the additional deselection of 200 PMTs in black. (a) For the gv_{ch} calibration at the ^{7}Be-ν edge region. (b) For the $N_{CNO+pep}$ analysis.



(a) (b)

FIGURE 7.19: (a) Results of the data gv_{ch} calibration at the ^{7}Be-ν edge region of Phase 2+3 for the nominal and stricter selection of PMTs. (b) Difference between the gv_{ch} results nominal and stricter selection of PMTs in units of the expected statistical uncertainty.



(a) (b)

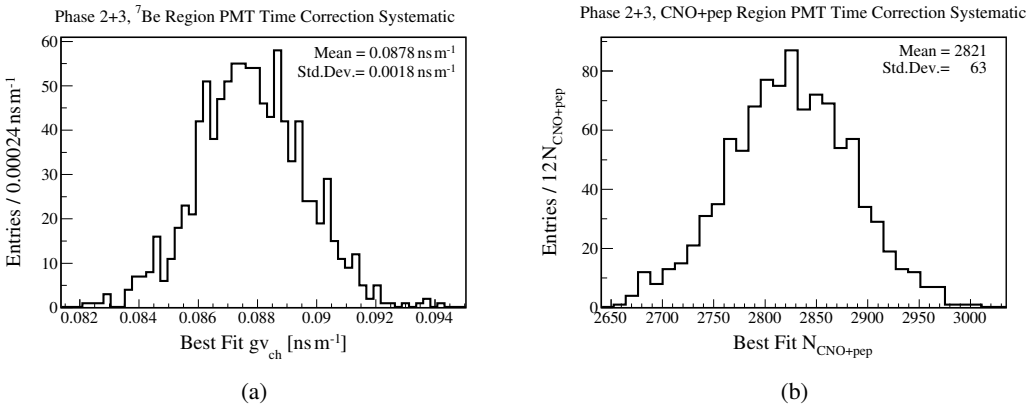
FIGURE 7.20: (a) Results of the data $N_{CNO+pep}$ measurement of Phase 2+3 for the nominal and the stricter selections of PMTs. (b) Difference between the $N_{CNO+pep}$ results of the nominal and the stricter selections of PMTs in units of the expected statistical uncertainty.

Figures 7.19(a) and 7.20(a), on the left, show the absolute fit results of the CID analysis for the gv_{ch} calibration at the ^{7}Be-ν edge region and for the $N_{CNO+pep}$ measurement. On the right 7.19(b) and 7.20(b) show the results for the difference between the different N^-_{PMTs} selections and nominal PMT selection, in units of the expected statistical uncertainty σ. There is no significant systematic uncertainty observed, as no result is outside of 2σ and for Phase 2+3 both N^-_{PMTs} dependent CID analyses show only a single result outside 1σ each, while for Phase 1 both CID analyses show only two results outside 1σ each. Overall there is a good agreement between the expected difference of the N^-_{PMTs} dependent results compared to the estimated statistical uncertainty, as for 20 results in total 6.3 are expected to be outside 1σ, while only six are observed for Phase 1 and Phase 2+3. The conclusion is then that the nominal PMT selection can be safely used for the CID analysis, without contributing a systematic uncertainty.

7.6.3 Systematic uncertainty from PMT time correction

The study of the ToF corrected hit time distributions of the individual PMTs is showing that the selected data PMTs have a constant time difference $\mu(\text{PMT})$ between each other, while still showing a good agreement for the overall shape of their hit time distributions (see Section 7.4). These constant offsets are significantly larger than the fit uncertainty $\Delta\mu(\text{PMT})$ for data, while the corresponding values of the same analysis on MC are well in agreement with the fit uncertainty. It is then reasonable to correct each data PMT with its measured constant offset $(\mu(\text{PMT}) - \langle\mu\rangle)$, relative to all the other PMTs, as otherwise there could still be an over-selection for some PMTs and under-selection for other PMTs for the early Nth-Hits.

Thus, the nominal CID analysis chain is performed with exactly this PMT time offset correction $(\mu(\text{PMT}) - \langle\mu\rangle)$. Because $\mu(\text{PMT})$ is only known with a finite precision it is necessary to propagate the uncertainties $\Delta\mu(\text{PMT})$ of each selected PMT through the entire CID analysis chain to include the systematic uncertainty from the small, but significant time differences of the data PMTs.



(a) (b)

FIGURE 7.21: Distribution of Phase 2+3 best fit results, where the PMT hit time correction is randomly varied a number of 10k times, using to the uncertainty $\Delta\mu(\text{PMT})$ as a Gaussian standard deviation. (a) For the gv_{ch} calibration at the ^{7}Be-ν edge region. (b) For the measurement of $N_{CNO+pep}$.

The error propagation of $\Delta\mu(\text{PMT})$ is done by performing the CID analysis an additional 1000 times, where each time the nominal time corrections $(\mu(\text{PMT}) - \langle\mu\rangle)$ of the PMTs are changed by a Gaussian smearing with a standard deviation of $\Delta\mu(\text{PMT})$. The distribution of the Phase 2+3 best fit results for the relevant CID parameters is shown in Figure 7.21, for the ^{7}Be-ν edge region gv_{ch} calibration on the left and the $N_{CNO+pep}$ analysis on the right.

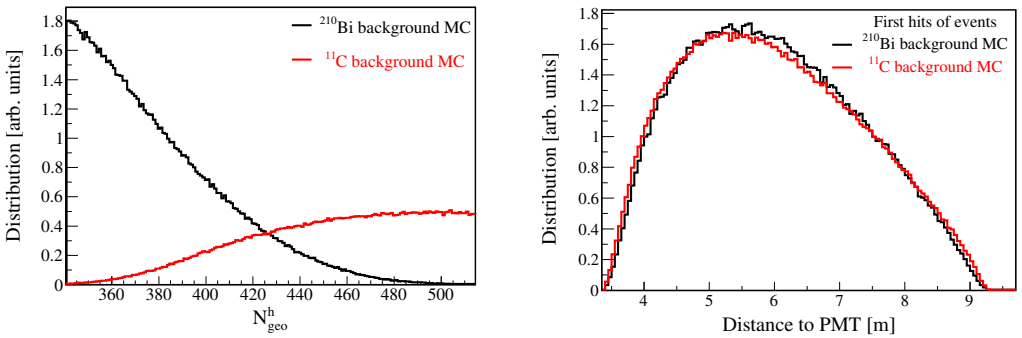
The propagated systematic uncertainty from the finite precision on the PMT time constant correction is estimated with the standard deviation of the best fit result distribution. It has to be noted that the systematic uncertainty is propagated through the full analysis chain, thus the $N_{\mathrm{CNO+pep}}$ analyses are performed after the calculation of the $\mathrm{gv_{ch}}$ systematic uncertainty and the average $\Delta\chi^2(\mathrm{gv_{ch}})$ profile is used as the pull-term for the $N_{\mathrm{CNO+pep}}$ analyses, explicitly including the systematic $\mathrm{gv_{ch}}$ uncertainty. The corresponding plots of Phase 1 can be found in the Appendix C.11.

	$\langle \mathrm{gv_{ch}} \rangle \pm \langle \mathrm{stat.} \rangle \pm \mathrm{syst.}\ [\mathrm{ns\,m^{-1}}]$	$\langle N_{\mathrm{CNO+pep}} \rangle \pm \langle \mathrm{stat.} \rangle \pm \mathrm{syst.}$
Phase 1	$0.1394 \pm 0.0278 \pm 0.0052$	$681^{+236}_{-226} \pm 26$
Phase 2+3	$0.0878 \pm 0.0178 \pm 0.0018$	$2821^{+519}_{-497} \pm 63$

TABLE 7.10: Systematic uncertainties from PMT time constant correction.

Table 7.10 summarises the propagated systematic uncertainty from the PMT time constant correction. Each variation of the PMT time constant correction gives a different best fit result with a corresponding estimated Gaussian uncertainty from the profile $\Delta\chi^2 = 1$. The resulting average value $\langle N_{\mathrm{CNO+pep}} \rangle$ of the distribution is different from the single result given by the nominal CID analysis. Overall, the finite precision of the constant PMT time correction contributes a relatively small systematic uncertainty on $N_{\mathrm{CNO+pep}}$ of $\sim 4\%$ and $\sim 2\%$ for Phase 1 and Phase 2+3, respectively.

7.6.4 Comparison of ^{11}C and ^{210}Bi background MC



(a) (b)

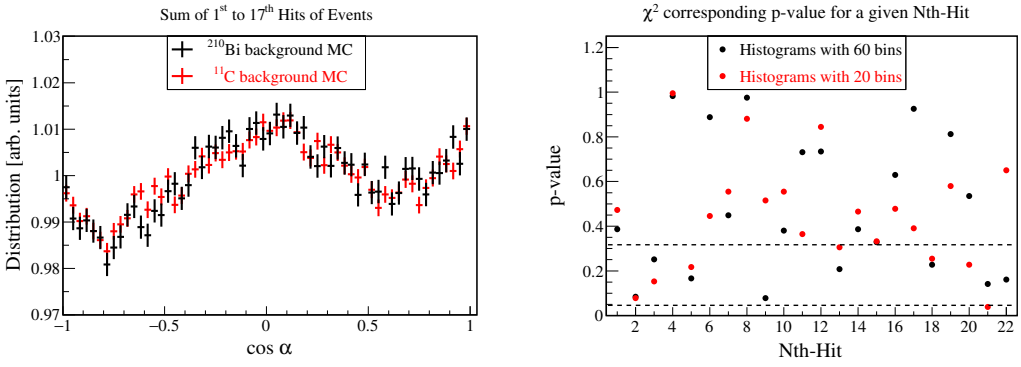
FIGURE 7.22: (a) Distribution of N_h^{geo} for ^{11}C MC (red) and ^{210}Bi MC (black) for Phase 2+3. (b) Distribution of the distance between reconstructed event position and the PMTs which detected the first hits of the events for ^{11}C MC (red) and ^{210}Bi MC (black) for Phase 2+3. The distribution of ^{11}C is shifted to lower distances, as ^{11}C events have a larger average energy which causes an over-selection of close PMTs, relative to ^{210}Bi events.

The two main background components of the CNO+pep-ν region are expected to be ^{11}C and ^{210}Bi. In principle there should be no difference in the $\cos\alpha$ histograms between those two event types, but they have significantly different energy distributions which could potentially introduce a systematic difference between them. For this reason the compatibility of the $\cos\alpha$ histograms between ^{11}C MC and ^{210}Bi MC is investigated in this section.

Figure 7.22(a) shows the N_h^{geo} distribution of ^{11}C MC in red and ^{210}Bi MC in black for the Phase 2+3 RoI. The distributions are significantly different, with an average value of

$\langle N_h^{geo} \rangle \approx 380$ for ^{210}Bi and $\langle N_h^{geo} \rangle \approx 510$ for ^{11}C. This large difference could potentially introduce a systematic difference in the $\cos\alpha$ histograms for early Nth-Hits because the definition of the Nth-Hit depends on the relative PMT hit time for a given event. The PMTs cannot resolve individual photon hits, but integrate all photon hits in a 80 ns window after the first detected photon, with a 60 ns dead time following that. The PMT hit time is then defined as the time of the first detected photon. This means that the underlying PMT hit time distribution shifts to earlier times and becomes more narrow for a larger number of detected photon hits (see also the position reconstruction PDF in Chapter 2).

At the same time PMTs closer to the event position have a higher probability to detect a photon due to their larger geometric acceptance. This means that there could be a selection bias of the first hits of the events for ^{11}C, relative to ^{210}Bi, which would over select close PMTs simply because ^{11}C events have a larger pool of photons to select the earliest hit from. Figure 7.22(b) shows this effect for the first hits of the events, where the distance between the reconstructed event position and the position of the PMT which detected the first hit is plotted. The general shape of the distribution is governed by the geometric acceptance of the PMTs and the number of PMTs at a given distance. The distance distributions of ^{11}C and ^{210}Bi are significantly different, with a $\chi^2/\text{ndf} = 1686/115$, given the statistics of MC. The distribution of ^{11}C is shifted to smaller distances due to the larger ^{11}C event energy compared to ^{210}Bi, which validates the concerns discussed above. At the same time the absolute difference between the distributions is small, with the average values of (5.972 ± 0.001) m for ^{11}C and (5.983 ± 0.001) m for ^{210}Bi.



(a) (b)

FIGURE 7.23: (a) The sum of the 1st to 17th hits of events $\cos\alpha$ histograms of ^{11}C MC (red) and ^{210}Bi MC (black) for Phase 2+3, normalised to the same integral. (b) p-values corresponding to the χ^2 between ^{11}C MC and ^{210}Bi MC $\cos\alpha$ histograms for a given Nth-Hit, with 60 bins (black) and 20 bins (red). The dotted lines show p-value = 0.32 and p-value = 0.05.

Figure 7.23 now shows the comparison of the ^{11}C MC and ^{210}Bi MC $\cos\alpha$ histograms. The sum of the first to 17th hits of the events is shown on the right 7.23(a). The histograms are comparable with a $\chi^2/\text{ndf} = 64.2/59$, p-value = 0.30 for 60 bins and $\chi^2/\text{ndf} = 23.3/19$ (p-value = 0.22) for 20 bins, given the statistics of MC. The additional binning of 20 is selected to investigate potentially larger structures in the $\cos\alpha$ histogram, as those could be obfuscated by the lower bin statistics for a binning of 60. As the sum of these $\cos\alpha$ histograms has a relatively large number of entries, compared to the single Nth-Hit $\cos\alpha$ histograms, a general systematic difference between the in ^{11}C MC and ^{210}Bi MC $\cos\alpha$ histograms seems unlikely.

Following this result, Figure 7.23(b) on the left shows the χ^2 corresponding p-value for 60 bins and 20 bins between the ^{11}C MC and ^{210}Bi MC $\cos\alpha$ histograms, for each individual Nth-Hit. The dotted lines highlight p-value = 0.32 and p-value = 0.05. The first hits of

the events are in agreement between both MC backgrounds and overall there is no Nth-Hit dependency visible, such as a linear trend or a clustering of low p-values at early Nth-Hits. For the 22 histograms investigated the number of p-values below 0.32 is expected to be seven and there are eight for 60 bins and seven for 20 bins. At the same time there is no p-value below 0.003. The conclusion is then, that there is no systematic difference between the MC $\cos\alpha$ histograms for ^{11}C and ^{210}Bi, given the statistics of MC. There is no energy dependence of $\cos\alpha$ for the background events and the concerns discussed above are not relevant in the selected RoI. For this reason the background MC used for the CID analysis of the CNO+pep-ν region is the sum of the ^{11}C MC and the ^{210}Bi MC. The background MC study for the RoI of Phase 1 is not shown as it gives qualitatively the same results, where no systematic differences are found.

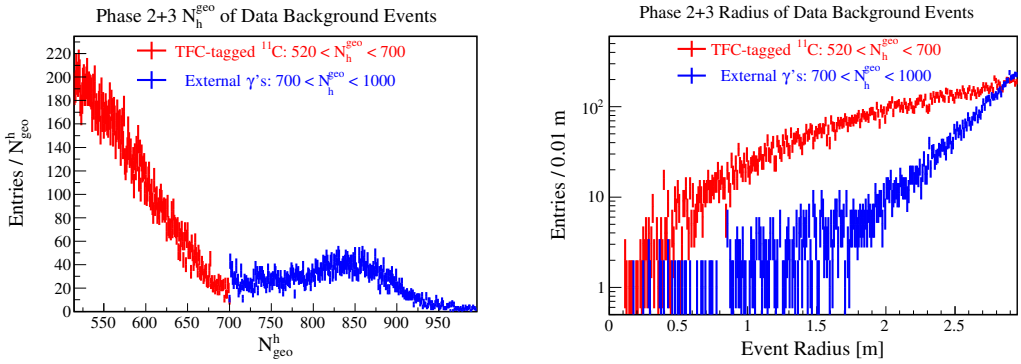
7.6.5 Investigation of systematics between data and MC background

In this section the potential systematic differences between the background MC (^{11}C MC plus ^{210}Bi MC), the TFC-tagged ^{11}C data background and external γ data background is studied. The ^{11}C data events are chosen because it is possible to efficiently tag them with the TFC algorithm, which provides a pure selection of a relatively large number of events, uniformly distributed in the FV. From the investigation of the MC it is expected that ^{210}Bi and ^{11}C background give the same CID $\cos\alpha$ distribution. But, as previously there is the concern about the impact of the ^{11}C event hits on $\cos\alpha$, due to the relatively larger number of PMT hits. This is because data and MC events have a small, but significant difference between their absolute hit time distributions. It could be possible that this small difference introduces a noticeable change in the $\cos\alpha$ histograms between ^{11}C data and MC, as these hit time distribution differences are more pronounced here due to the larger number of PMT hits (event energy) for ^{11}C events, compared to ^{210}Bi events. Another reason is that ^{11}C events are positron events and there is no dedicated positron calibration in Borexino, which could mean there are non-negligible differences between ^{11}C MC and data for the CID analysis [4].

For the external γ events there is no dedicated CID MC production, as it is prohibitively time consuming, but instead the ^{11}C MC plus ^{210}Bi MC background MC is used for all types of CID background. Consequently, there is again the concern regarding the difference in hit time distribution between external γ data and MC background. Of special concern is the unique difference between the external γ events and MC, as their position distribution is highly non-uniform. They are coming from the outside of the detector and their position distribution follows an exponential increase for larger event radii, given by the attenuation length of the LS for these γ energies. For this reason it is important to study their general $\cos\alpha$ shape, as well as the single Nth-Hit $\cos\alpha$ histograms. As for the CID analysis, here the $\cos\alpha$ histograms are produced with the PMT selection, as well as the PMT time correction of data PMTs of Section 7.4. The investigation of pure background events can also be used as a supplementary estimate of this selection. Should the PMT selection be insufficient or the PMT time correction drastically wrong, then this would introduce a Nth-Hit dependence in the statistical agreement between the data and MC $\cos\alpha$ histograms. All these systematic studies are performed in this section on Phase 2+3, while the relevant plots for Phase 1 can be found in the Appendix C.4.

Figure 7.24(a) shows the selected ^{11}C and external γ data events, in red and blue, respectively. The ^{11}C events are selected trough the TFC-tagging with $520 < N_h^{geo} < 700$ and the external gammas with neither TFC-cut not tag with $700 < N_h^{geo} < 1000$. The number of selected data events is $N_{^{11}C} = 19302$ and $N_\gamma = 7026$, both of which are larger than the total

[4] For example, the difference in the hit time distribution of ^{11}C MC and data has lead to the abandonment of the position reconstruction likelihood parameter (Fig. 2 and Fig. 3 in [84]) of the multivariate fit after Phase 2.



(a) (b)

FIGURE 7.24: (a) The selected N_h^{geo} region for the TFC-tagged ^{11}C (red) and external γ (blue) data events of Phase 2+3. The γ events are neither TFC-cut nor TFC-tagged. The events numbers are $N_{^{11}C} = 19302$ and $N_\gamma = 7026$. (b) The corresponding distribution of the reconstructed event radius, in a log-scale. ^{11}C follows a third order polynomial distribution, as expected from a uniform event distribution vs. the radius. The external γ's show the expected exponential increase for larger radii.

number of events in corresponding the CNO+pep-ν RoI with $N_{\text{total}} = 5974$ for Phase 2+3. The same is true for Phase 1 with $N_{^{11}C} = 9318$, $N_\gamma = 3840$ and $N_{\text{total}} = 2990$.

Figure 7.24(b) shows the radial distribution of the reconstructed event positions for these events. The uniformly distributed ^{11}C events correctly show a third order polynomial distribution, where the bin entries are proportional to the volume of the 1 cm wide radial shells given the histogram binning. The external γ's show an exponential increase with larger radii.

Toy-MC study of external γ background non-uniform position distribution

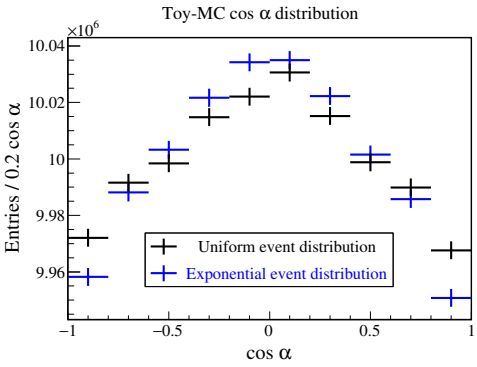


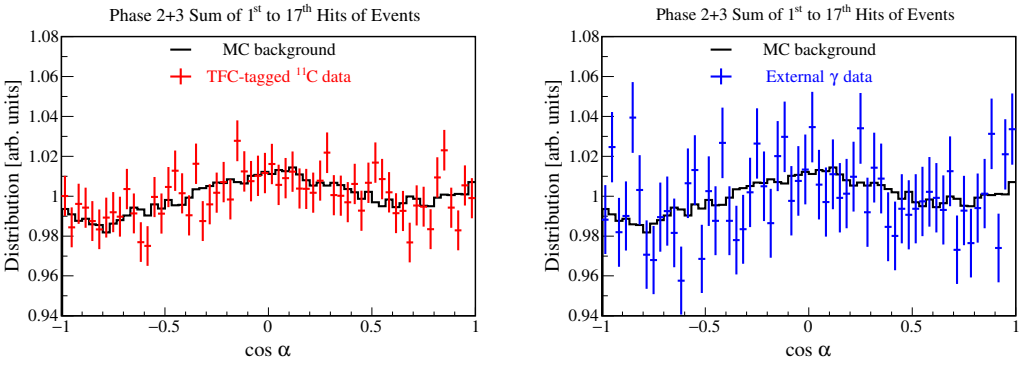
FIGURE 7.25: Toy-MC $\cos\alpha$ histogram for uniform (black) and exponential radial event position (blue) distributions. Using the live PMT distribution of Run 17328 a number of 10^8 toy events is simulated for each position distribution. First, the solar direction is randomly selected from the measured distribution and the event position is sampled according to the relevant distribution. Then a single photon direction for each event is sampled isotropically and the $\cos\alpha$ value is calculated.

The measured exponential radial distribution is used in the following to perform a toy-MC study, where the difference in $\cos\alpha$, introduced by the pure geometric differences between a uniform and exponential radial event distribution, is investigated.

Figure 7.25 shows the $\cos\alpha$ distribution of toy-MC simulations with a uniform event distribution in black and with the exponentially increasing radial distribution of the external γ's in blue. The toy-MC study is performed with the live PMT profile of run 17328, with a statistic of 10^8 PMT hits. First, the direction of the Sun is randomly sampled from the measured distribution of the solar orientation in Borexino coordinates (Figure 3.2). Then the event position is sampled from the relevant radial distribution and a photon direction is sampled isotropically and $\cos\alpha$ is calculated if a PMT is hit. These PMT hits have no time information, as to estimate the purely geometrical effect of the exponential radial distribution on the $\cos\alpha$ shape of the external γ background.

Given the statistics of the toy-MC, the two $\cos\alpha$ histograms are significantly different from each other with a $\chi^2/\text{ndf} = 40.1/9$, p-value $= 6.4 \cdot 10^{-6}$. This means that there is in principal an influence of the position distribution of the events on the $\cos\alpha$ shape, where the exponential radial distribution gives an increase around $\cos\alpha \approx 0.0$ and a correlated decrease around $\cos\alpha \approx \pm 1.0$. At the same time however, the absolute $\cos\alpha$ difference is small with an average, relative difference $\Delta\cos\alpha$ between the uniform and exponential radial distribution of $\langle\Delta\cos\alpha\rangle = 0.08\% \cdot \langle\cos\alpha\rangle$. This value is relatively small, even in comparison to the $\sim 0.15\%$ statistical uncertainty of the full MC $\cos\alpha$ histograms with 60 bins and with the 200 simulated MC events per data event in the CNO+pep-ν RoI. The effect of the drastically non-uniform position distribution of the external γ events is therefore safely negligible given the statistics of data. Even if the data would be made up entirely of γ events, the $\cos\alpha$ shape would be described well enough by the ^{11}C MC plus ^{210}Bi MC background MC.

Summed data $\cos\alpha$ histogram comparison



(a) (b)

FIGURE 7.26: Comparison of the $\cos\alpha$ histograms summed over the 1st to 17th hits of events, for data and MC background of Phase 2+3. The histograms are normalised to the same integral. (a) TFC-tagged ^{11}C data compared to MC background with a $\chi^2/\text{ndf} = 53.5/59$, p-value $= 0.67$. (b) External γ data compared to MC background with a $\chi^2/\text{ndf} = 60.4/59$, p-value $= 0.42$.

The principle, statistical compatibility between the data and MC background events is investigated with the sum of the first to 17th hits of the events $\cos\alpha$ histograms. Should there be a relevant difference in the $\cos\alpha$ histograms between data and MC due to a geometric difference, such as the live PMT distribution or non-uniform event position distribution, it would be visible here, given to the large statistics compared to the $\cos\alpha$ histogram of a single Nth-Hit. Figure 7.27 shows this sum of the $\cos\alpha$ histograms for the TFC-tagged ^{11}C events in red and the external γ events in blue, in comparison with the background MC in black. The histograms are normalised for easier comparison between the ^{11}C and the external γ events. There are no significant differences between the data and MC histograms visible. Both data background species are well in agreement with the background MC, given the results of the χ^2 test. The $\cos\alpha$ histograms are also compared with a binning of 20 to investigate potential structures of a larger $\cos\alpha$ scale. Again, both data background species are in agreement with the MC background.

The χ^2/ndf and p-values of both background species and number of bins for Phase 1 and Phase 2+3 are summarised in Table 7.11. Phase 1 and Phase 2+3 give qualitatively the same results, where the data background and the MC background are statistically comparable to each other, with no p-value below 0.32. There are no relevant systematic difference in the principal $\cos\alpha$ shape between data background and MC background, given the statistics of

	TFC-tagged ^{11}C data vs. MC		External γ data vs. MC	
	χ^2/ndf	p-value	χ^2/ndf	p-value
Phase 1 (20 bins)	15.8 / 19	0.67	16.5 / 19	0.62
Phase 1 (60 bins)	48.6 / 59	0.83	62.8 / 59	0.34
Phase 2+3 (20 bins)	20.7 / 19	0.35	21.0 / 19	0.33
Phase 2+3 (60 bins)	53.5 / 59	0.68	60.4 / 59	0.42

TABLE 7.11: χ^2 test comparison between the data background and the MC background.

the selected data background events. This validates again the result of the previous toy-MC studies for γ events, as well as the selection of good PMT's.

Nth-Hit dependent data $\cos\alpha$ histogram comparison

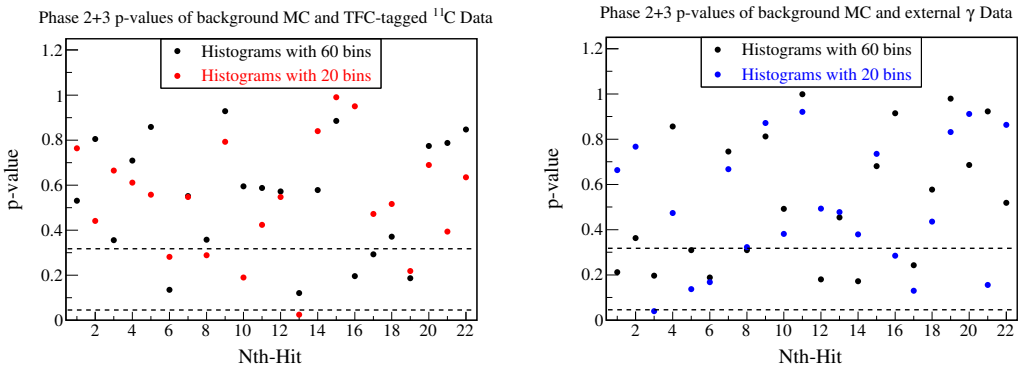


FIGURE 7.27: The χ^2/ndf corresponding p-values between data and MC background of Phase 2+3, as a function of the individual Nth-Hit $\cos\alpha$ histograms. The dotted lines show p-value = 0.32, p-value = 0.05. (a) TFC-tagged ^{11}C data compared to MC background for a binning of 60 (black) and 20 (red). (b) External γ data compared to MC background for a binning of 60 (black) and 20 (blue).

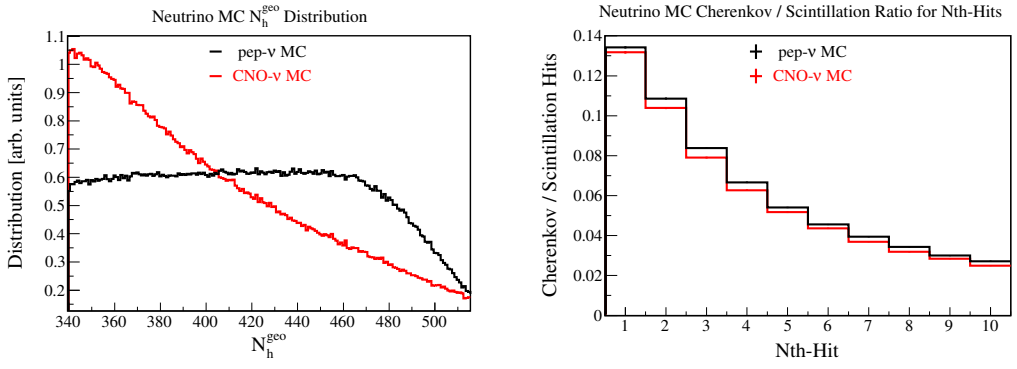
The next investigation concerns the Nth-Hit dependence of the $\cos\alpha$ shape. The absolute time differences of the external γ and TFC-tagged ^{11}C data events, compared to the MC background, could potentially introduce a systematic $\cos\alpha$ difference for early hits between data background and MC background. The applied PMT time correction from Section 7.4 could also introduce such a systematic difference, should it be drastically incorrect.

Figure 7.27 shows the χ^2/ndf corresponding p-value as a function of the individual Nth-Hit $\cos\alpha$ histograms for the TFC-tagged ^{11}C data on the left and the external γ's on the right, both with 20 and 60 histogram bins. The dotted lines highlight p-value = 0.32 and p-value = 0.05. The p-values are calculated until the 22$^{\text{nd}}$ hits of the events. The $\cos\alpha$ histograms of both data species are comparable with the MC background and there is no Nth-Hit dependence visible, such as a linear trend or clustering of p-values. Overall, the data background Nth-Hit dependent $\cos\alpha$ histograms show a good agreement with the MC. For 22 histograms the expected number of p-values below 0.32 is seven, while the graphs show five for ^{11}C for both 20 and 60 bins and six and eight for 20 and 60 bins, respectively, for the external γ events. For Phase 2+3 there are no p-values below 0.003, while for Phase 1 there is a single p-value = 0.0016 for Nth-Hit = 8 of the external γ events. This is the case only for 60 bins, while for 20 bins it has p-value = 0.0031. This implies a significant difference

between the data and MC $\cos\alpha$ histogram for this particular Nth-Hit, but the neighbouring Nth-Hit $\cos\alpha$ histogram are again in agreement. In total 84 histograms (22+22 from Phase 2+3 and 20+20 from Phase 1) are investigated, which gives an expected number of p-values below 0.003 as 0.25. Using a Binomial distribution with the probability $p = 0.003$ and the total number of points $n = 84$ gives a probability of 22% to have at least one point with a p-value below 0.003. Thus, this single point is likely just a statistical outlier and cannot be used to estimate a systematic difference between the γ data and MC background. Continuing with the important $\cos\alpha$ histogram of the first hits of the events shows a p-value ≈ 0.2 for the γ events, but only for a binning of 60 and not for 20, while the p-value of the ^{11}C is well above 0.32. This seems not enough to imply a relevant systematic difference between the data and the MC and the same is true for Phase 1, where, again the γ events show a p-value ≈ 0.2, but now for 20 bins and not for 60 bins.

The statistics of the selected data background events are larger then the total number of events in the corresponding CNO+pep-ν RoI for Phase 1 and Phase 2+3. Therefore it can be concluded from the ^{11}C data events that the systematic uncertainty from the PMT hit time correction can only be small compared to the statistical uncertainty. Likewise, the systematic uncertainty from an absolute difference of the hit time distribution of external γ data events compared to ^{11}C MC plus ^{210}Bi MC background is likely negligible, as is expected from the use of the event-by-event Nth-Hit method. To summarise: There is no relevant systematic uncertainty from the geometric differences or from the differences in the absolute hit time distributions between any species of background in the data and the MC. Furthermore, there are no significant Nth-Hit dependent systematic differences visible, which is in agreement with the explicit systematic studies of the PMT selection, PMT hit time correction and Nth-Hit(max) selection of the previous sections.

7.6.6 CNO and pep neutrino signal MC



(a) (b)

FIGURE 7.28: Comparison between pep-ν (black) and CNO-ν (red) MC. (a) N_h^{geo} distribution. (b) Cherenkov/scintillation ratio of individual Nth-Hits, with a simulated $gv_{ch} = 0.108\,\mathrm{ns\,m^{-1}}$. The total ratio is 0.475% for pep-ν MC and 0.469% for CNO-ν MC.

The main signal components for the CID analysis are the CNO-ν and pep-ν events, where the relative contribution is unknown for the purpose of the $N_{\mathrm{CNO+pep}}$ measurement. The corresponding MC $\cos\alpha$ histograms of these neutrino species are very similar, such that a simultaneous CID analysis of both N_{CNO} and N_{pep} number of neutrino events is not able to discriminate them in any usable way. Thus, the nominal $N_{\mathrm{CNO+pep}}$ analysis is performed with pep-ν MC $\cos\alpha$ histograms only. Nonetheless, their $\cos\alpha$ histograms are in fact different as their energy distributions are different and the corresponding systematic uncertainty is

quantified in this section. The ν-MC N_h^{geo} distributions can be seen in Figure 7.28(a). The mean values are about $\langle N_h^{geo} \rangle \approx 404$ and $\langle N_h^{geo} \rangle \approx 421$ for CNO-ν and pep-ν events, respectively. For the pep-ν events the N_h^{geo} distribution has a Compton-edge like shape, as the pep-neutrinos are mono-energetic with 1.44 MeV. CNO-ν events on the other hand follow a continuous distribution, with an endpoint at 1.74 MeV. This results in a significantly different N_h^{geo} distribution, where the event probability rises for lower event energies.

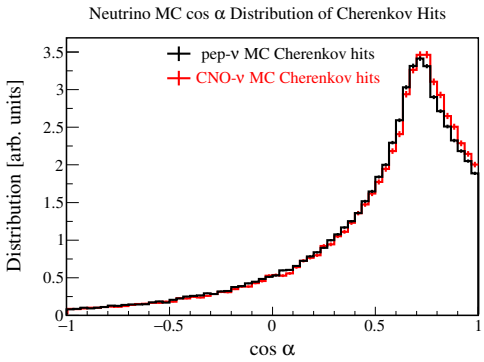


FIGURE 7.29: The MC $\cos\alpha$ histograms of pure Cherenkov hits for pep-ν (black) and CNO-ν (red).

The average energy of the scattered electrons from pep-ν events is higher than the average energy of the electrons corresponding to the CNO-ν events. This results in a larger ratio of Cherenkov over scintillation hits, which is shown in Figure 7.28(b) on the right, where the Cherenkov over scintillation ratio is plotted for the first ten hits of the events, for a simulated value of the Cherenkov group velocity correction of $gv_{ch} = 0.108\,ns\,m^{-1}$. The overall ratio for all PMT hits is 0.475% for pep-ν events and 0.469% for CNO-ν events, according to the MC. The number of Cherenkov hits per event is about 1% lower for the CNO-ν events relative to the pep-ν events. A naive expectation would be that the best fit value of $N_{CNO+pep}$ performed with the CNO-ν MC $\cos\alpha$ histograms should be about 1% higher, compared to the nominal results using the pep-ν MC histograms, as the ratio of Cherenkov hits per data event should be kept equal. Moreover, the shape of the $\cos\alpha$ histograms are also explicitly different between the CNO-ν and pep-νevents for the selected CNO+pep-ν RoI, due to their different energy distributions. It is expected that the lower average electron energies of CNO-ν events should shift the average Cherenkov angle and consequently also the CID $\cos\alpha$ distribution to larger $\cos\alpha$ values, compared to pep-ν events. Exactly this can be seen in Figure 7.29, where the sum of the first ten Cherenkov hits of all MC events (without scintillation hits) is shown in black for pep-ν and in red for CNO-ν. The $\cos\alpha$ difference is relatively small with a shift of the peak value by $\Delta\cos\alpha \approx 0.02$, but this is significant within the ten times increased MC statistics as the χ^2-test between the histogram gives a $\chi^2/ndf = 184.7/59$.

$N_{CNO+pep}$	pep signal MC	CNO signal MC	$\Delta N_{CNO+pep}$
Phase 1	681	695	14 (2.1%)
Phase 2+3	2821	2768	-53 (1.9%)

TABLE 7.12: Systematic uncertainty from CNO and pep MC $\cos\alpha$ histogram differences.

To estimate the systematic uncertainty on the $N_{CNO+pep}$ measurement, introduced by the differences in the $\cos\alpha$ histograms of CNO-ν and pep-ν MC, the CID analysis is performed once nominally with the pep-ν signal MC histograms and a second time with the CNO-ν signal MC histograms, including the systematic uncertainty from the PMT time correction. The difference between the two $\langle N_{CNO+pep} \rangle$ results is used as the systematic uncertainty. This is a maximally conservative approach as it assumes that $N_{CNO+pep}$ is made up entirely either of CNO-ν or pep-ν events, while in reality it is a mixture of both neutrino species. The

values of this systematic uncertainty are summarised in Table 7.12. It is interesting to note that for Phase 1 the $N_{\mathrm{CNO+pep}}$ value increases with the use of the CNO-ν MC as the signal $\cos\alpha$ distribution, while it decreases for Phase 2+3. This likely means that the difference in the $\cos\alpha$ shape dominates the systematic uncertainty, compared to the effect of a decreased Cherenkov to scintillation ratio when using the CNO-ν MC for the CID analysis. Overall, the systematic uncertainty from the CNO-ν and pep-ν MC $\cos\alpha$ histogram difference is relevant, but relatively small with $\sim 2\%$ of the nominal best fit $N_{\mathrm{CNO+pep}}$ value.

7.6.7 Constraint on expected ^{8}B-ν, ^{7}Be-ν and pep-ν events

While it is expected that CNO-ν and pep-ν dominate the signal events in data, there is still a small contribution of the ^{8}B-ν and ^{7}Be-ν events, which is taken into account as a systematic uncertainty on the $N_{\mathrm{CNO+pep}}$ measurement in this section.

Table 7.3 shows the HZ- and LZ-SSM expected number of ^{7}Be-ν events in the RoI and their averaged values are $N_{^7\mathrm{Be}}(\text{Phase 1}) = 1.0 \pm 0.1$ and $N_{^7\mathrm{Be}}(\text{Phase 2+3}) = 9.5 \pm 0.7$. For the expected number of ^{8}B neutrino events the Super-Kamiokande measurement of the flavour-stable[5] ^{8}B neutrino flux is used: $2.345 \pm 0.014\,(\mathrm{stat}) \pm 0.036\,(\mathrm{syst}) \cdot 10^6\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ [58]. For Borexino this corresponds to a ^{8}B-ν rate of 0.420 ± 0.007 cpd / 100 t [88], for the full ^{8}B-ν energy range. The energy efficiency, including the trigger efficiency ε_{E} is given by the full G4BX2 MC simulation as $\varepsilon_{\mathrm{E}}(^8\mathrm{B}, \text{Phase 1}) = 0.088$ and $\varepsilon_{\mathrm{E}}(^8\mathrm{B}, \text{Phase 2+3}) = 0.081$. With the corresponding FV, detector live time and TFC-exposure this leads to the expected number of ^{8}B neutrino events of $N_{^8\mathrm{B}}(\text{Phase 1}) = 15.9 \pm 0.3$ and $N_{^8\mathrm{B}}(\text{Phase 2+3}) = 59.0 \pm 1.0$. Those neutrino species, that are not of interest here, are subtracted from the best fit $N_{\mathrm{CNO+pep}}$ value with their number of expected ^{8}B-ν, ^{7}Be-ν events. Their corresponding uncertainties are treated as additional systematic uncertainties.

The previous investigation of the systematic uncertainty, given by the difference of the CNO-ν and pep-ν MC in Section 7.6.6, shows that a difference of the $\cos\alpha$ shape can either increase as well as decrease the best fit CID result. As the neutrino energies are different for ^{8}B-ν and ^{7}Be-ν compared to pep-ν and CNO-ν the exact $\cos\alpha$ shape is unknown without a full CID MC simulation. In the best case the $\cos\alpha$ histograms of ^{8}B and ^{7}Be would look comparable to the histograms of CNO-ν and pep-ν, which seems reasonable for the mono-energetic ^{7}Be-ν (0.862 MeV). For the ^{8}B neutrino events, with an energy distribution maximum around 6.4 MeV an endpoint near 15 MeV [159], it is not clear how the correct ^{8}B $\cos\alpha$ histograms would influence the $N_{\mathrm{CNO+pep}}$ result. Thus, the assumption on the ^{8}B-ν number can be considered a somewhat conservative correction on the $N_{\mathrm{CNO+pep}}$ measurement. This is because the ^{8}B-ν $\cos\alpha$ distribution should be slightly flatter than for pep-ν and CNO-ν, due to the larger average angle between the neutrino direction and the recoil electron direction.

It is additionally necessary to constrain the number of pep-ν events N_{pep} on the measured $N_{\mathrm{CNO+pep}}$, for the CID measurement of N_{CNO}. This is done in the toy-MC estimation of the posterior distribution, described in Section 7.7.2, where the expected distribution $P(N_{\mathrm{pep}})$ is included as a prior. This distribution is assumed to be the sum of two Gaussian distributions with the mean and standard deviation given by the values of the HZ-SSM and LZ-SSM, respectively. Furthermore, the Poissonian, statistical uncertainty from the expected number of all ν events is also explicitly taken into account, as the expected statistical uncertainty is larger than the model uncertainties given by the SSM. This is done, so as to be able to marginalise over all possible, true values for the number of ^{8}B-ν, ^{7}Be-ν and pep-ν events in the data sample. Thus, the full correction can be estimated as $\Delta N_{^7\mathrm{Be}+^8\mathrm{B}+\mathrm{pep}}(\text{Phase 1}) = -425 \pm 29\,(\text{expected stat. + SSM})$ and $\Delta N_{^7\mathrm{Be}+^8\mathrm{B}+\mathrm{pep}}(\text{Phase 2+3}) = -1588 \pm 48(\text{expected stat. + SSM})$, which adds an additional systematic uncertainty to the CID measurement of N_{CNO}.

[5] Meaning it includes the flavour transformation assuming the MSW-LMA solution, shown in [68] and is therefore smaller than the pure SSM expectation.

7.6.8 Fiducial volume selection

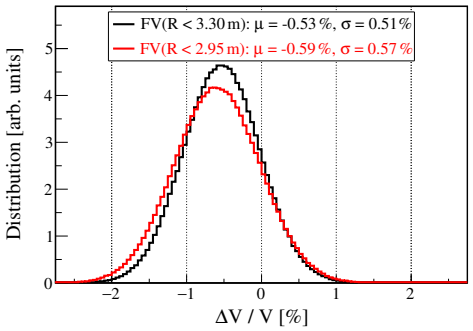


FIGURE 7.30: The toy-MC estimation of the systematic uncertainty on the FV, for R < 3.30 m in black and R < 2.95 m in red. In the toy-MC the radial cut of the FV is randomly sampled according to the systematic uncertainty of the reconstructed data event radius, which is known from the results of the calibration campaign. The plot shows the distribution of the relative difference ΔV / V between the sampled volume and the nominal volume V.

ΔV / V	R < 3.30 m	R < 3.05 m	R < 3.00 m	R < 2.95 m
Mean	-0.53 %	-0.57 %	-0.58 %	-0.59 %
Std. Dev.	0.51 %	0.55 %	0.56 %	0.57 %

TABLE 7.13: Relative systematic uncertainty of the different spherical FV used in the CID analyses.

The data events are selected trough a cut on the reconstructed event radius, which results in a FV with a certain target mass. The FV is used for the calculation of the neutrino rate in [cpd / 100 t] from the measured number of neutrino events and vice versa. The uncertainty on the FV is calculated in this section, using the standard Borexino toy-MC method which has been used in most Borexino publications. For this the $^{222}\mathrm{Rn}+^{14}C$ source positions close to the FV edge at R > 3 m are selected and the reconstructed event position radius is compared to the nominal one obtained from the CCD cameras. For the southern hemisphere the mean offset and standard deviation is $\Delta R = (1.4 \pm 0.8)$ cm, while for the northern hemisphere it is $\Delta R = (-2.6 \pm 0.8)$ cm, where this offset is dominated from the systematic mis-reconstruction of the z-position in data [85].

The toy-MC volumes are produced by randomly sampling the edge of the FV according to these ΔR values and then comparing them to the nominal volume. Figure 7.30 shows the resulting distributions of ΔV / V for the FV of R < 3.30 m in black, which is used for Phase 1 and Phase 2 at the ^{7}Be-ν edge region and R < 2.95 m in red, which is used for Phase 2+3 at the CNO+pep-ν region. The systematic uncertainties are summarised in Table 7.13 for the different FV used throughout the CID analysis chain. Larger FV produce smaller relative uncertainties and all FV have a negative offset and a standard deviation of about 0.6 %. The offset to smaller volumes in the data compared to the MC is due to the systematic position mis-reconstruction of the z-position in the data. The uncertainty of the LS density of (0.878 ± 0.004) g cm^{-3} [90] should also be included here, which results in a total uncertainty of 0.7 % for the fiducial mass.

7.6.9 Energy selection cuts

The data events are selected in an optimised energy region, using the N_h^{geo} estimator. This results in an efficiency ε_E on the energy cut, which is the fraction of events within the selected energy window. The value of ε_E is calculated for each event species from the N_h^{geo} spectrum of the full G4BX2 MC simulations. It is then used for the calculation of the neutrino rate from the measured number of neutrino events within the RoI and vice versa.

There is a systematic difference in the energy scale between the data and the MC, which results in a systematic uncertainty on ε_E. The true N_h^{geo} region of the data can be different

$\Delta\varepsilon_E / \varepsilon_E$	^{7}Be-ν edge region			CNO+pep-ν region	
	^{7}Be-ν	pep-ν	CNO-ν	pep-ν	CNO-ν
Phase 1	1.3 %	1.6 %	1.5 %	0.8 %	1.4 %
Phase 2	1.0 %	1.3 %	1.2 %	0.8 %	1.4 %
Phase 3	0.8 %	1.1 %	1.0 %	0.8 %	1.4 %

TABLE 7.14: Relative systematic uncertainty of ε_E from the N_h^{geo} cuts for the relevant energy regions and neutrino species of the CID analyses.

to the region that is used to calculate ε_E from the MC N_h^{geo} spectrum. Using the calibration sources, cosmogenic neutrons (from the muon-^{12}C spallation) and the internal ^{218}Po and ^{11}C events the non-linearity of the energy scale, light-yield stability over time, and spatial non-uniformity have been studied by the Borexino collaboration [68]. The relative uncertainty on the energy scale is 0.32% and the uncertainty from the non-uniformity and non-linearity of the response is at the level of 0.06% and 0.09%, respectively [139].

The total, relative uncertainty on the energy response of 0.34% is used as a Gaussian standard deviation, to randomly sample the edge of the selected N_h^{geo} cuts. Then the resulting ε_E values are compared to the nominal value for the different neutrino species. Table 7.14 summarises the standard deviation of the toy-MC $\Delta\varepsilon_E / \varepsilon_E$ distributions. The mean values for all energy regions and neutrino species are centred around zero. The sub-percent uncertainty on the energy scale can result in a relative systematic uncertainty above one percent on ε_E. The reason for this is that the N_h^{geo} energy distributions of the different neutrino species are not flat and have different probability values for the low and high N_h^{geo} cuts. For the selected N_h^{geo} cut around the ^{7}Be-ν edge region the relative uncertainties decrease as the regions are broader for later phases, while for the CNO+pep-ν region the selected N_h^{geo} cuts are too similar between Phase 1 and Phase 2+3 to produce a relevant difference for $\Delta\varepsilon_E / \varepsilon_E$.

7.6.10 Summary of systematic uncertainties

The systematic uncertainties, relevant for the CID analysis, have been studied in this section and a number of possible sources of uncertainties can be excluded. The selection of Nth-Hit(max), as well as a general Nth-Hit dependence of the CID analysis has been of great concern, as the MC model of the data neutrino events through the CID parameter gv_{ch} is only a linear approximation of the possible, underlying difference in the Cherenkov time behaviour between data and MC. It turns out that there is no relevant systematic uncertainty from the Nth-Hit(max) selection and neither is there a general Nth-Hit dependence. The best fit result provides a good p-value for each individual Nth-Hit and the pure data background and the MC background are also in agreement with each other for all Nth-Hits, given the statistics of the selected data background. Toy-MC studies confirm that the differences of the CID results for different Nth-Hit(max) are well explained with only statistical fluctuations. Likewise, toy-MC studies show that the selection of the number of bins for the $\cos\alpha$ histograms, as well as the selection of the good PMTs to use for the CID analysis do not contribute a systematic uncertainty.

The $\cos\alpha$ shape of background is only determined by the distribution of the selected, live PMTs relative to the positions of the Sun for the selected events. There is no $\cos\alpha$ dependence on the background energy as the ^{210}Bi MC and ^{11}C MC histograms are in agreement within the large statistics of MC. In principle the external γ events could produce a different $\cos\alpha$ distribution, compared to uniformly distributed events, but a toy-MC study shows that this effect is safely negligible given the statistics of the data. This is also confirmed through the

comparison of the external γ data at high energies with the background MC. Combining the observations of the uncertainty studies and the fit results it can be concluded that the use of the Nth-hit time like variable greatly helps the CID analysis, as absolute time differences between event species become irrelevant. Likewise, using $\mathrm{gv_{ch}}$ and Δr_{dir} works well enough as to not introduce a detectable systematic uncertainty between the MC model and the data.

Following that, a number of systematic uncertainties have been identified that are relevant for the CID analysis. First is the contribution from the uncertainty of the time constant correction $\Delta\mu$(PMT), applied for each PMT. Another source is the difference between CNO-ν and pep-ν MC $\cos\alpha$ histograms. Unlike the background, the neutrino signal $\cos\alpha$ histogram does depend on the energy. The Cherenkov over scintillation ratio increases for larger recoil electron energies and also because the $\cos\alpha$ shape is influenced by the energy dependent angle of the Cherenkov cone. As such, the choice to perform the $N_{\mathrm{CNO+pep}}$ CID analysis with either the CNO-ν or the pep-ν MC histograms adds a systematic uncertainty. Furthermore, there are additional neutrinos species which are not of interest and must be constrained, which also introduces an uncertainty. These are ^{8}B+^{7}Be-ν for the measurement of $\mathrm{N_{CNO+pep}}$ and additionally pep-ν for the measurement of $\mathrm{N_{CNO}}$.

The next category of systematic uncertainties does not directly impact the CID analysis, but it is used for the calculation of the neutrino rate in [cpd / 100 t] from the CID measurement of the number of neutrino events, and vice versa. Here, the selection of the FV as well as the $\mathrm{N_h^{geo}}$ RoI introduces uncertainties, which are evaluated through calibration data and toy-MC studies. The MLP algorithm is known to provide an exposure of $99.5\% \pm 1.0\%$, while the uncertainty of the TFC exposure is negligible with a value of $\pm 0.02\%$. Overall, the total contribution from the relevant systematic uncertainties is small compared to the uncertainty given by the CID fit.

The last observation of these systematics studies is that Phase 1 and Phase 2+3 are principally comparable to each other. Phase 1 and Phase 2+3 have been analysed independently to cross-validate the CID method and to search for potential systematic uncertainties arising from a change of the detector response over time, such as the loss of PMTs. A potential difference between the phases is given by the difference in the effective $\mathrm{gv_{ch}}$ calibration which is not significant and $\mathrm{gv_{ch}}$ is constrained independently for Phase 1 and Phase 2+3. Furthermore, the selection good PMTs to use for the analyses and the relative PMT hit time correction is different between the phases, as well as the uncertainties on the FV and ε_{E}. Apart from this no systematic differences are found between Phase 1 and Phase 2+3, given the investigation of pure background data, the performance of the best neutrino fit on individual Nth-Hit $\cos\alpha$ histograms, as well as the lack of a Nth-Hit(max) dependence of the fit results. This shows that the CID method is also robust over time, where the detector exhibits a decrease in position and energy resolution given the loss of PMTs and other potential changes of the detector response. The conclusion is then that the results of Phase 1 and Phase 2+3 can be combined as they are (conditionally) independent from each other and no prohibitively strange behaviour has been observed between them.

The systematic uncertainties are summarised in Table 7.15 with their absolute and relative values. The relative values are given in comparison to the expected N_ν at the ^{7}Be-ν edge region and to the final result of the CID N_{CNO} measurement. The ΔN_ν values at the ^{7}Be-ν edge region are included in the $-2\ln(P(N_\nu))$ pull-term and affect the $\mathrm{gv_{ch}}$ calibration only indirectly.

7.7 Fit bias and Bayesian interpretation of the CID results

The final goal of this CID analysis is to provide a limit on the zero-neutrino hypothesis for $N_{\mathrm{CNO+pep}}$ and N_{CNO}, as well as the measurement of the CNO-ν rate. For this, the production of

gv_{ch} calibration at ^{7}Be-ν edge region

Source	Phase 1	Phase 2+3
PMT time correction	0.0052 ns m^{-1} (3.7%)	0.0018 ns m^{-1} (2.1%)
ΔN_ν(ΔFV/FV)	67 (0.7%)	320 (0.7%)
ΔN_ν($\Delta\varepsilon_{eff}/\varepsilon_{eff}$)	130 (1.3%)	423 (0.9%)
ΔN_ν(MLP)	97 (1.0%)	470 (1.0%)

CNO-ν measurement

Source	Phase 1	Phase 2+3
PMT time correction	26 (9.6%)	63 (5.5%)
CNO-ν vs. pep-ν MC	14 (5.2%)	53 (4.6%)
pep+^{8}B+^{7}Be-ν constraint	29 (10.8%)	48 (4.2%)
ΔFV/FV	0.7%	0.7%
$\Delta\varepsilon_{eff}/\varepsilon_{eff}$	1.4%	1.4%

TABLE 7.15: Relevant systematic uncertainties for the CID analyses and the neutrino rate conversion.

confidence intervals is of great importance, but at the same time the method of their production, as well as the correct interpretation of the CID $\Delta\chi^2$ profiles, is not trivial.

The CID analysis is performed with a χ^2-test between the data and the MC $\cos\alpha$ histograms, where χ^2 is calculated as a function of three parameters: the number of solar neutrinos N_ν, the effective Cherenkov group velocity correction gv_{ch} and the mis-reconstruction in the initial electron direction Δr_{dir}. This model is able to describe the data well, given the selection of good PMTs to use, as is shown in Section 7.5 and Section 7.6.

One peculiarity of the fit is that the parameters gv_{ch} and Δr_{dir} only influence the shape of the $\cos\alpha$ histogram for the neutrino signal but not of background. The Cherenkov hits of background events are not correlated to the position of the Sun and neither is the initial direction of electrons from the ^{210}Bi background, for example. The analysis is furthermore performed as a chain of two sub-analyses, where first N_ν is treated as a nuisance parameter with a pull-term for the effective calibration of gv_{ch} at the ^{7}Be-ν edge region and this is then in turn used as nuisance parameter with a pull-term for the $N_{CNO+pep}$ fit.

Last is the problem of physical boundaries present in the analysis: there cannot be a negative number $N_{CNO+pep}$ and neither can there be a negative rate of CNO-ν after introducing the pep-ν constraint. The relative large statistical and systematic uncertainties imply that the physical boundaries have a real impact on the interpretation of the CID results, more so for Phase 1 than for Phase 2+3 and most importantly for the CNO-ν rate. The peculiarities of the CID fitting procedure are investigated, exemplary on Phase 2+3, in the first part of this Section 7.7.1. The relevant plots for Phase 1 can be found in the Appendix C.5.

As far as the production of frequentist confidence intervals is concerned it is in principle possible to avoid non-physical regions using the Feldman-Cousins method [160], but then the question opens up on how to treat the nuisance parameters and how to include the systematic uncertainties as well as the pep-ν constraint. For the Feldman-Cousins method it is necessary to simulate toy-experiments, which depend on the values of the nuisance parameters. The choice of which true nuisance values to simulate is somewhat arbitrary and different methods have been proposed, for example using the best fit values or their posterior, likelihood distributions [161]. A seemingly popular method for the inclusion of systematic uncertainties

in the confidence interval production is by integrating over probability density functions which parameterise the uncertainties [162, 163]. This marginalisation of the systematic uncertainties is itself a Bayesian method and it is sometimes called a semi-Bayesian or hybrid method. As such, the production of frequentist confidence intervals [164, 14] under the conditions of the CID analysis seems possible but not trivial.

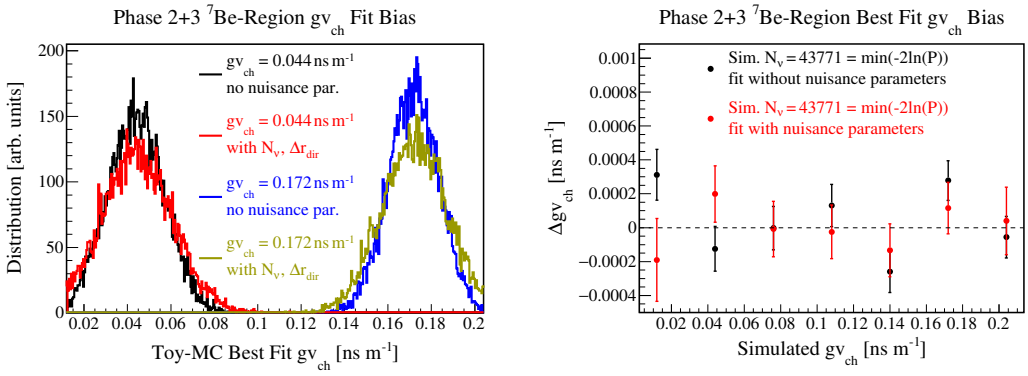
In contrast, the correct production of Bayesian credible intervals [14] seems relatively simple, including the effect of the physical boundaries, the nuisance parameters and the systematic uncertainties. Furthermore, the interpretation of the credibility intervals is relatively intuitive, where the probability is interpreted as a degree of belief: how much do we think that something is true given the evidence of the data. For example a 68.3% interval means that the true, unknown parameter value has 68.3% chance to be inside this interval (see [165] or other textbooks). For this reason I have decided to use Bayesian statistics for the interpretation of the CID analysis, where the complete results are given by the posterior distributions $P(N_{\mathrm{CNO+pep}} \mid \mathrm{data})$ and $P(N_{\mathrm{CNO}} \mid \mathrm{data})$. The production method of the Bayesian posterior distributions is called *Acceptance Sampling Unfolding* in this work and it is explained in the second part of this Section 7.7.2, together with the equations used for the production of the credible intervals.

7.7.1 Qualitative investigation of the fit response bias

For a given set of true parameter values $\mathbf{t} = \left(N_\nu^{\mathrm{true}}, \mathrm{gv}_{\mathrm{ch}}^{\mathrm{true}}, \Delta r_{\mathrm{dir}}^{\mathrm{true}}\right)$, the fitting procedure itself can introduce a systematic shift on the best fit values $\mathbf{b} = \left(N_\nu^{\mathrm{fit}}, \mathrm{gv}_{\mathrm{ch}}^{\mathrm{fit}}, \Delta r_{\mathrm{dir}}^{\mathrm{fit}}\right)$, relative to $\mathbf{t}$. This fit response bias is now studied qualitatively by performing 10k toy-MC analyses, where the simulated (true) parameters are set to certain values $\mathbf{s} = \left(N_\nu^{\mathrm{sim}}, \mathrm{gv}_{\mathrm{ch}}^{\mathrm{sim}}, \Delta r_{\mathrm{dir}}^{\mathrm{sim}}\right)$. Then toy-data is sampled from the MC $\cos\alpha$ histograms according to $\mathbf{s}$ with the statistics of real data and the analysis is performed normally, including the respective pull-terms. The Gaussian mean of the resulting best fit distributions of interest $P(N_\nu^{\mathrm{fit}} \mid \mathbf{s})$, $P(\mathrm{gv}_{\mathrm{ch}}^{\mathrm{fit}} \mid \mathbf{s})$ is then compared to the simulated values $\mathbf{s}$.

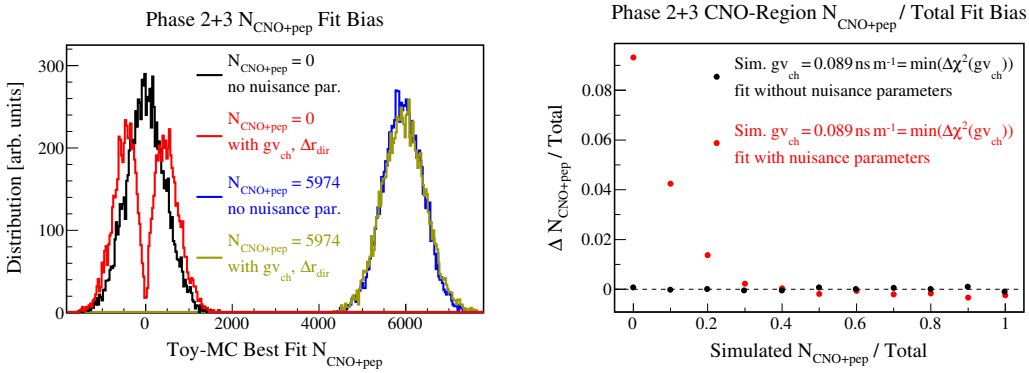
Figure 7.31(a) shows the $\mathrm{gv}_{\mathrm{ch}}^{\mathrm{fit}}$ distribution for $\mathrm{gv}_{\mathrm{ch}}^{\mathrm{sim}} = 0.044\,\mathrm{ns\,m^{-1}}$ (black, red) and $\mathrm{gv}_{\mathrm{ch}}^{\mathrm{sim}} = 0.172\,\mathrm{ns\,m^{-1}}$ (blue, yellow) for the ^{7}Be-ν region, where the simulated number of neutrinos is set to the minimum of the $-2\ln(P(N_\nu))$ pull-term, $N_\nu^{\mathrm{sim}} = 43771$. The nuisance parameter $\Delta r_{\mathrm{dir}}^{\mathrm{sim}}$ is set to 1.8 cm. The $\mathrm{gv}_{\mathrm{ch}}$ fit is performed with both nuisance parameters N_ν, Δr_{dir} (red, yellow) and also only with $\mathrm{gv}_{\mathrm{ch}}$ as a one-dimensional fit (black, yellow) to investigate the effect of the nuisance parameters, i.e. the nuisance parameters are known with infinite precision. As expected, the one dimensional fit without nuisance parameters produces a more narrow distribution and the mean value of the $\mathrm{gv}_{\mathrm{ch}}^{\mathrm{fit}}$ distribution is in agreement with the injected value for both configurations of the fit. This can be seen for a range of injected $\mathrm{gv}_{\mathrm{ch}}^{\mathrm{sim}}$ values in Figure 7.31(b), where $\Delta\mathrm{gv}_{\mathrm{ch}} = \mathrm{gv}_{\mathrm{ch}}^{\mathrm{fit}} - \mathrm{gv}_{\mathrm{ch}}^{\mathrm{sim}}$ is in agreement with zero with (red) and without (black) the nuisance parameters. There is no bias in the fit of $\mathrm{gv}_{\mathrm{ch}}$ at the ^{7}Be-ν region, for an injected N_ν^{sim} at the minimum of the N_ν pull-term.

For the $N_{\mathrm{CNO+pep}}$ fit the general behaviour is different, as can be seen in Figure 7.32(a). The $N_{\mathrm{CNO+pep}}^{\mathrm{fit}}$ distributions are plotted for the injected values of $N_{\mathrm{CNO+pep}}^{\mathrm{sim}} = 0$ (black, red) and $N_{\mathrm{CNO+pep}}^{\mathrm{sim}} = 5974$, which corresponds to a signal to total ratio of 0.0 and 1.0, respectively. Again the $N_{\mathrm{CNO+pep}}$ fit is performed once normally (red, yellow) and once without the nuisance parameters $\mathrm{gv}_{\mathrm{ch}}$, Δr_{dir}. Their injected values are set to $\mathrm{gv}_{\mathrm{ch}}^{\mathrm{sim}} = 0.089\,\mathrm{ns\,m^{-1}}$, which corresponds to the minimum of the $\chi^2(\mathrm{gv}_{\mathrm{ch}})$ pull-term and the value of $\Delta r_{\mathrm{dir}}^{\mathrm{sim}} = 2.7\,\mathrm{cm}$. Now, the inclusion of the nuisance parameters shows a clear effect for a low signal to total ratio, where the best fit value of $N_{\mathrm{CNO+pep}}^{\mathrm{fit}} = 0$ is maximally disfavoured. The fit is allowed beyond the physical limits to illustrate this effect in Figure 7.32(a) and a double peak structure is visible at $N_{\mathrm{CNO+pep}}^{\mathrm{fit}} \approx \pm 500$, while the fit without the nuisance parameters shows a Gaussian distribution around zero.



(a) (b)

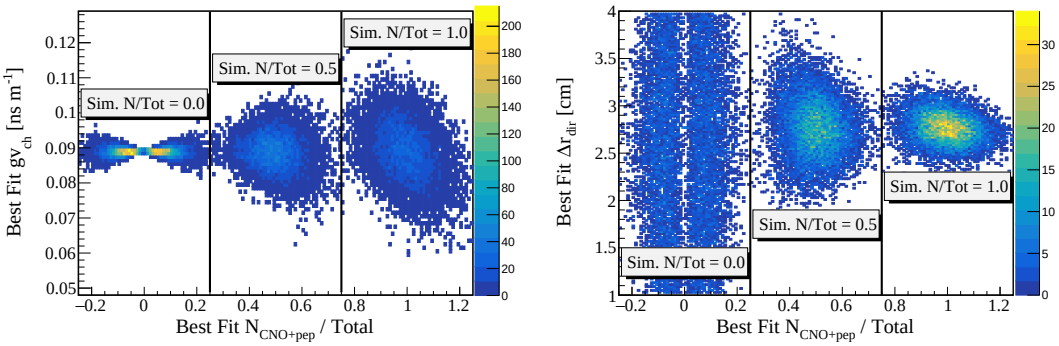
FIGURE 7.31: Fit response on gv_{ch} at the ^{7}Be-ν edge region. The simulated value of N_ν^{sim} is set to the minimum of the pull-term. (a) Distribution of the fit results for 10k toy-MC analyses, where the simulated value is set to $gv_{ch}^{sim} = 0.044\,ns\,m^{-1}$ (black, red) and $gv_{ch}^{sim} = 0.172\,ns\,m^{-1}$ (blue, yellow). The analyses are performed once, including the two nuisance parameters N_ν, Δr_{dir} of the fit (red, yellow) and for comparison with fixed values of the nuisance parameters and only gv_{ch} as a free parameter (black, blue). (b) Difference between the mean value of the gv_{ch}^{fit} distribution and the simulated gv_{ch}^{sim}. The mean value is estimated with a Gaussian fit.



(a) (b)

FIGURE 7.32: Fit response on $N_{CNO+pep}$, where the simulated value of gv_{ch}^{sim} is set to the minimum of the pull-term. (a) Distribution of the fit results for 10k toy-MC analyses, where the simulated value is set to $N_{CNO+pep}^{sim} = 0$ (black, red) and $N_{CNO+pep}^{sim} = 5974$ (blue, yellow). The analyses are performed including the two nuisance parameters of the fit (red, yellow) and for comparison only with $N_{CNO+pep}$ as a free parameter (black, blue). The fit is not constrained to the physical limits ($0 \leq N_{CNO+pep} \leq 5974$) to illustrate the impact of the gv_{ch}, Δr_{dir} nuisance parameters. (b) Offset between the mean value of the $N_{CNO+pep}^{fit}$/total distribution and the simulated $N_{CNO+pep}^{sim}$/total. The mean value is estimated with a Gaussian fit performed within the physical limits of $N_{CNO+pep}$.

Figure 7.32(b) plots the relative difference $\Delta N_{\mathrm{CNO+pep}} / N_{\mathrm{total}}$ against the injected value. Here, $\Delta N_{\mathrm{CNO+pep}} = N^{\mathrm{fit}}_{\mathrm{CNO+pep}} - N^{\mathrm{sim}}_{\mathrm{CNO+pep}}$ is calculated as the difference between the Gaussian mean and the injected value. The Gaussian fit is performed only within the physical limits $0 \leq N^{\mathrm{fit}}_{\mathrm{CNO+pep}} \leq 5974$. While the one dimensional fit (black) of only $N_{\mathrm{CNO+pep}}$ shows a good agreement with zero for all simulated values $N^{\mathrm{sim}}_{\mathrm{CNO+pep}}$, the normal analysis including the nuisance parameters (red) shows a significant bias for low signal to total ratios. The lower the number of $N^{\mathrm{sim}}_{\mathrm{CNO+pep}}$ the bigger is the bias towards a higher fitted number of neutrinos $N^{\mathrm{fit}}_{\mathrm{CNO+pep}}$. For a $\mathrm{gv}^{\mathrm{sim}}_{\mathrm{ch}}$ value set to the minimum of the pull-term the bias stops after a signal to total ratio of ~ 0.4, corresponding to $N_{\mathrm{CNO+pep}} \sim 2400$ in Phase 2+3 which is also true for Phase 1, where a signal to total ratio of 0.4 corresponds to $N_{\mathrm{CNO+pep}} \sim 1200$.



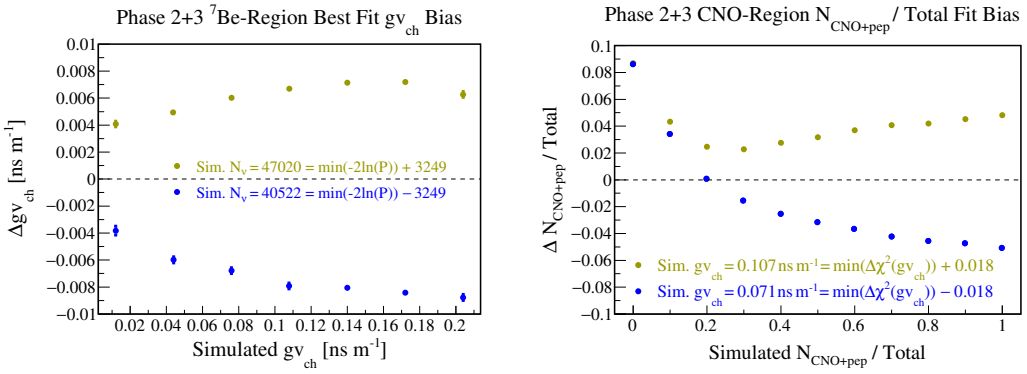
(a) (b)

FIGURE 7.33: Distribution of the fit results for 10k toy-MC analyses of $N_{\mathrm{CNO+pep}}$, where the simulated $\mathrm{gv}^{\mathrm{sim}}_{\mathrm{ch}}$ is set to the minimum of the pull-term. Three simulated $N_{\mathrm{CNO+pep}}$ values are shown, with $N_{\mathrm{CNO+pep}} = 0$ (left), $N_{\mathrm{CNO+pep}} = 2987$ (middle) and $N_{\mathrm{CNO+pep}} = 5974$ (right). (a) $N_{\mathrm{CNO+pep}}$ vs. $\mathrm{gv}_{\mathrm{ch}}$. (b) $N_{\mathrm{CNO+pep}}$ vs. Δr_{dir}.

The bias at low signal to total ratios comes from the asymmetrical influence of the nuisance parameters on the CID neutrino signal compared to the CID background. As a consequence larger neutrino ratios effectively increase the degrees of freedom of the analysis up to a point, because $\mathrm{gv}_{\mathrm{ch}}$, Δr_{dir} only influence the $\cos\alpha$ shape of the neutrino events. Fitting close to pure background means that the analysis effectively only has a single degree of freedom: signal vs. background, while (incorrectly) increasing the absolute value of $N^{\mathrm{fit}}_{\mathrm{CNO+pep}}$ scales up how much the $\cos\alpha$ histogram is influenced by $\mathrm{gv}_{\mathrm{ch}}$, Δr_{dir}. Thus, for a low number of injected neutrinos $N^{\mathrm{sim}}_{\mathrm{CNO+pep}}$ the fit prefers to increase $N^{\mathrm{fit}}_{\mathrm{CNO+pep}}$ as random noise in the toy-data $\cos\alpha$ histogram can then be better fitted through the increased impact of $\mathrm{gv}_{\mathrm{ch}}$, Δr_{dir} on the MC $\cos\alpha$ histogram used for the fit.

This can be seen in Figure 7.33, where the distribution of $N^{\mathrm{fit}}_{\mathrm{CNO+pep}}$ is plotted against $\mathrm{gv}^{\mathrm{fit}}_{\mathrm{ch}}$ (7.33(a)) and $\Delta r^{\mathrm{fit}}_{\mathrm{dir}}$ (7.33(b)), for three different values of $N^{\mathrm{sim}}_{\mathrm{CNO+pep}}$, corresponding to signal to total ratios of 0.0, 0.5 and 1.0. The injected values of the nuisance parameters are again $\mathrm{gv}^{\mathrm{sim}}_{\mathrm{ch}} = 0.089\,\mathrm{ns\,m^{-1}}$, which corresponds to the minimum of the $\mathrm{gv}_{\mathrm{ch}}$ pull-term and $\Delta r^{\mathrm{sim}}_{\mathrm{dir}} = 2.7\,\mathrm{cm}$. For the simulation of pure background $N^{\mathrm{sim}}_{\mathrm{CNO+pep}} = 0$ there is no correlation between $\mathrm{gv}^{\mathrm{fit}}_{\mathrm{ch}}$, $\Delta r^{\mathrm{fit}}_{\mathrm{dir}}$ and $N^{\mathrm{fit}}_{\mathrm{CNO+pep}}$. The distribution of $\mathrm{gv}^{\mathrm{fit}}_{\mathrm{ch}}$ is constrained by the pull-term, but broadens for larger absolute values of $N^{\mathrm{fit}}_{\mathrm{CNO+pep}}$, while the distribution of $\Delta r^{\mathrm{fit}}_{\mathrm{dir}}$ is unconstrained and does not depend on $N^{\mathrm{fit}}_{\mathrm{CNO+pep}}$, given $N^{\mathrm{sim}}_{\mathrm{CNO+pep}} = 0$. Arbitrary values of $\mathrm{gv}^{\mathrm{fit}}_{\mathrm{ch}}$, $\Delta r^{\mathrm{fit}}_{\mathrm{dir}}$ can fit the injected, pure background better by using a larger absolute value of $N^{\mathrm{fit}}_{\mathrm{CNO+pep}}$. For larger values of the simulated number of neutrinos there is a clear anti-correlation visible between the $\mathrm{gv}^{\mathrm{fit}}_{\mathrm{ch}}$, $\Delta r^{\mathrm{fit}}_{\mathrm{dir}}$ and $N^{\mathrm{fit}}_{\mathrm{CNO+pep}}$, where the anti-correlation increases for increasing $N^{\mathrm{sim}}_{\mathrm{CNO+pep}}$.

This is expected, as the analysis is effectively a counting of the early Cherenkov hits and corresponding them to a total number of neutrinos events. Increasing the value of $\mathrm{gv}_{\mathrm{ch}}^{\mathrm{fit}}$ corresponds to a higher ratio of early Cherenkov hits per neutrino event, for a given number of Cherenkov hits in the toy-data, and consequently less neutrino events are needed to explain the fixed number of toy-data Cherenkov hits.



(a) (b)

FIGURE 7.34: Fit response of the respective parameters of interest, for simulated nuisance parameters which are offset from their pull-term minimum. The simulated values of the nuisance parameters are bigger than the minimum of the pull-term for the yellow graph and smaller for the blue graph. (a) Difference between the mean value of the $\mathrm{gv}_{\mathrm{ch}}^{\mathrm{fit}}$ distribution and the simulated $\mathrm{gv}_{\mathrm{ch}}^{\mathrm{sim}}$. (b) Offset between the mean value of the $N_{\mathrm{CNO+pep}}^{\mathrm{fit}}$/total distribution and the simulated $N_{\mathrm{CNO+pep}}^{\mathrm{sim}}$/total.

The anti-correlation of the fit parameters with each other has the effect, that the pull-terms also influence multiple fit parameters and not only the single parameter on which the pull-term is explicitly applied. For the ^{7}Be-ν region analysis of $\mathrm{gv}_{\mathrm{ch}}$ this means that if the best fit value of N_ν^{fit} is offset from the minimum of the N_ν pull-term, then $\mathrm{gv}_{\mathrm{ch}}^{\mathrm{fit}}$ will be also shifted relative to the true value of $\mathrm{gv}_{\mathrm{ch}}^{\mathrm{sim}}$. For the CNO+pep-$\nu$ region analysis of $N_{\mathrm{CNO+pep}}$ the same is true if now $\mathrm{gv}_{\mathrm{ch}}^{\mathrm{fit}}$ is offset from the minimum of its $\mathrm{gv}_{\mathrm{ch}}$ pull-term. This behaviour is shown in Figure 7.34(a) for the ^{7}Be-ν region and in Figure 7.34(b) for the CNO+pep-ν region analysis. An increased N_ν^{sim} is pulled to lower N_ν^{fit} values and thus $\mathrm{gv}_{\mathrm{ch}}^{\mathrm{fit}}$ is also increased. These biases are relatively symmetric around zero for the region around $\mathrm{gv}_{\mathrm{ch}} = 0.089\,\mathrm{ns\,m^{-1}}$, which suggests that they cancels each other for a symmetric $\mathrm{gv}_{\mathrm{ch}}^{\mathrm{sim}}$ region around the best fit. This means that for the calibration of $\mathrm{gv}_{\mathrm{ch}}$ the fit procedure itself is unlikely to introduce a bias. Again, this looks very different for $N_{\mathrm{CNO+pep}}^{\mathrm{fit}}$, where simulated low signal to total ratios are systematically fitted to larger values for both values of $\mathrm{gv}_{\mathrm{ch}}^{\mathrm{sim}}$ and these pull-term induced biases on $N_{\mathrm{CNO+pep}}^{\mathrm{fit}}$ are symmetric around zero only for a signal to total ratio larger ~ 0.4.

To summarise this section: The overall CID fit procedure has three peculiarities that give rise to a complex fit response which introduces a bias for the measured number of CNO+pep-ν events $N_{\mathrm{CNO+pep}}^{\mathrm{fit}}$ relative to the true value:

- The fit parameters $\mathrm{gv}_{\mathrm{ch}}$ and Δr_{dir} only influence the shape of the $\cos\alpha$ histogram of the neutrino signal, but not of background. Low signal to total ratio fit values are disfavoured up to a point, because a higher fitted ratio has effectively a larger degree of freedom to fit noise.

- All three fit parameters have a (anti-)correlation between each other in the fit. The size and sign of the correlation factor depends on the event statistic and on the true, underlying values of the parameters.

- The two pull-terms for the effective gv_{ch} calibration and $N_{CNO+pep}$ measurement fulfill their function by pulling the relevant fit parameters towards the minimum of the pull-term histogram. Because of the (anti-)correlations between the fit parameters this means that all parameters are influenced by the pull-term in some way.

Under optimal circumstances the effect of pull-terms and correlations of the nuisance parameters does not introduce a bias for the credible intervals from the $\Delta\chi^2$ profile, as normally the pull-term bias is symmetric and averages out over the (posterior) likelihood distribution. The pull-term is effectively just the prior of the relevant parameter. This is likely the case for the gv_{ch} calibration at the ^{7}Be-ν edge region. The reason that here no fit bias is expected is that the expected N_ν / total ratio is relatively large, with relatively small uncertainties with $\sim 48.6\% \pm 4\%$ for Phase 1 and $\sim 74.5\% \pm 6\%$ for Phase2+3 and as such the asymmetric influence of the signal to total ratio can be expected to be only small for Phase 1 and certainly negligible for Phase 2+3. This is different for the CNO+pep-ν region, where the previous fit, without systematic uncertainties, gives ratios of $\sim 22.0\% \pm 7.6\%$for Phase 1 and $\sim 47.2\% \pm 8.5\%$ for Phase 2+3. Due to the relatively large statistical uncertainty here, the estimated region of the underlying $N_{CNO+pep}$ posterior distribution goes well below the $\sim 40\%$ ratio, for which the signal to total bias quickly increases and so the fit bias is expected be non-negligible. Thus, it is not possible to use a simple calculation of the likelihood distribution $P = \exp(-\frac{1}{2}\Delta\chi^2)$, normalise it and recalculate the credible intervals within the physical boundaries. It is necessary to first unfold the fit response, quantify and correct the fit bias to produce the correct posterior distribution of $N^{true}_{CNO+pep}$, which is explained in the following section.

7.7.2 MC sampling of the posterior: Acceptance Sampling Unfolding

This section explains the production of the Bayesian posterior distribution through the so called *Acceptance Sampling Unfolding*. Let $\mathbf{t} = \left(N_\nu^{true}, gv_{ch}^{true}, \Delta r_{dir}^{true}\right)$ be the set of the true, underlying parameters that give rise to the CID data. Typically $\mathbf{t}$ is smeared by the finite statistics of data and the analysis then results in a set of best fit values $\mathbf{b} = \left(N_\nu^{fit}, gv_{ch}^{fit}, \Delta r_{dir}^{fit}\right)$ which provide the best description of data using the MC model. The posterior distribution $P(\mathbf{t} \mid \text{data})$ is then given by Bayes' theorem:

$$P(\mathbf{t} \mid \text{data}) = \frac{P(\text{data} \mid \mathbf{t})\pi(\mathbf{t})}{\int P(\text{data} \mid \mathbf{t}')\pi(\mathbf{t}')d\mathbf{t}'} \tag{7.7}$$

Where "data" means the $\cos\alpha$ histograms of data. In the case of the CID test statistic (Equations 7.5, 7.6) the posterior is expressed as a $\Delta\chi^2$ profile and the priors $\pi(\mathbf{t}) = \pi(N_\nu) \cdot \pi(gv_{ch}) \cdot \pi(\Delta r_{dir})$ are implemented as pull-terms on N_ν and gv_{ch} for the respective analyses. The other parameter priors are effectively uniform distributions, bound by the parameter space that is scanned for the analysis.

Because the best fit results $\mathbf{b}$ MC histograms are able to reproduce the data $\cos\alpha$ shape well, $\mathbf{b}$ can be used as an estimate on the most likely parameter values that give rise to the data $\cos\alpha$ histograms. As the fit procedure itself introduces a systematic bias on $\mathbf{b}$, relative to $\mathbf{t}^{best}$, it follows that the simple $\Delta\chi^2$ likelihood is not a correct estimation of the posterior distribution. This bias can be expressed as a response matrix M, which transforms $\mathbf{t}^{best}$ into $\mathbf{b} = M\mathbf{t}^{best}$ To reconstruct or unfold $\mathbf{t}^{best}$ from the measured values of $\mathbf{b}$ we could use toy-MC studies to produce the three dimensional matrix M. This would be done by scanning the true parameter space with a certain binning and then measuring the bias of the fit parameter space, as has been done for example in Figure 7.31(b) for only one dimension. Then M could be inverted to produce an estimate $\tilde{\mathbf{t}} = M^{-1}\mathbf{b}$. Applying M^{-1} on all possible parameter values makes it then possible to construct a bias corrected posterior distribution. This process is potentially

complex, tedious to perform and could itself introduce a systematic bias based on the algorithm used for the matrix inversion, such as the choice of bin width and the interpolation method between the bins.

Instead, here a method of estimating the posterior distribution of $\mathbf{t}$ is applied in which neither the explicit knowledge of M, nor a matrix inversion is necessary. This method is called *Acceptance Sampling Unfolding* (ASU) and an idea very similar to ASU is expressed in [166], where it is called *Unfolding by Folding*. Ultimately, the ASU method is simply the estimation of the posterior distribution through toy-MC sampling:

1. A sample of CID parameters $\mathbf{t}^{\text{sim}} = \left(N_\nu^{\text{sim}}, \text{gv}_{\text{ch}}^{\text{sim}}, \Delta r_{\text{dir}}^{\text{sim}}\right)$ is randomly selected from their independent prior distributions $\pi(N_\nu)$, $\pi(\text{gv}_{\text{ch}})$, $\pi(\Delta r_{\text{dir}})$.

2. The corresponding toy-data $\cos\alpha$ histograms are sampled with the statistics of the true data from the full signal and background MC $\cos\alpha$ histograms, according to the randomly selected CID parameters $\mathbf{t}^{\text{sim}}$.

3. The analysis is performed in the same way as for real data, including the relevant pull-terms. Each toy-analysis provides a set of best fit values $\mathbf{b}^{\text{sim}}$, which are compared to the measured best fit result of real data $\mathbf{b}^{\text{data}}$.

4. The parameter values of $\mathbf{t}^{\text{sim}}$ are accepted and saved in a histogram, if and only if $\mathbf{b}^{\text{sim}} = \mathbf{b}^{\text{data}} + \varepsilon$, where ε is a set of parameter values that defines the acceptance region around $\mathbf{b}^{\text{data}}$. The size of ε is given by the relevant values of the systematic uncertainties for each CID parameter. The acceptance region is fuzzy and modeled as a Gaussian, meaning the $\mathbf{t}^{\text{sim}}$ is only accepted with a certain probability given by the standard deviation ε.

5. The toy-analyses are performed 4000 times for the ^{7}Be-ν edge region and 20k times for the CNO+pep-ν region to produce a distribution $P(\mathbf{t}^{\text{sim}} \mid \mathbf{b}^{\text{sim}} = \mathbf{b}^{\text{data}})$.

The best fit results $\mathbf{b}$ MC histograms are able to reproduce the data $\cos\alpha$ shape well, so the use of $P(\mathbf{t}^{\text{sim}} \mid \mathbf{b}^{\text{sim}} = \mathbf{b}^{\text{data}})$ as an estimate for $P(\mathbf{t}^{\text{sim}} \mid \text{data})$ is a reasonable approximation, as otherwise the exact reproduction of the $17 \cdot 60 = 1020$ $\cos\alpha$ histogram bin entries through random sampling would take too long. The selection of the prior distributions from which to sample the CID parameters is as follows: At the ^{7}Be-ν edge region N_ν^{sim} is sampled according to the allowed neutrino distribution, while gv_{ch} is sampled uniformly within $-0.012\,\text{ns}\,\text{m}^{-1} \leq \text{gv}_{\text{ch}} \leq -0.228\,\text{ns}\,\text{m}^{-1}$ and Δr_{dir} is sampled uniformly from 0.0 cm to 4.0 cm. At the CNO+pep-ν region $\text{gv}_{\text{ch}}^{\text{sim}}$ is then sampled from the marginalised posterior distribution of the gv_{ch} calibration. $N_{\text{CNO+pep}}$ is sampled uniformly within the physical boundaries $0 \leq N_{\text{CNO+pep}} \leq 2990$ for Phase 1 and $0 \leq N_{\text{CNO+pep}} \leq 5974$ for Phase 2+3 and Δr_{dir} is sampled uniformly from 0.0 cm to 6.0 cm. The choice for Δr_{dir} is informed by the physical limit at 0.0 cm, as the position reconstruction algorithm can only introduce a systematic mis-reconstruction in the direction of the early Cherenkov hits, while the upper limit is at most given by total position resolution of ~ 12 cm. A smaller maximum value is chosen to speed up the simulation, as the contribution of toy-analyses with $\Delta r_{\text{dir}}^{\text{sim}} > 6$ cm are safely negligible given the statistics of the ASU simulation. A larger selected $\Delta r_{\text{dir}}^{\text{sim}}$ region does not change the results. The same argument is used for the selection of the uniform $\text{gv}_{\text{ch}}^{\text{sim}}$ prior at the ^{7}Be-ν edge region, for which otherwise no physical limit is assumed.

In this way the ASU method produces a distribution $P(\mathbf{t}^{\text{sim}} \mid \mathbf{b}^{\text{data}})$ of the simulated, underlying CID parameters $\mathbf{t}^{\text{sim}}$ that are able to reproduce exactly the measured best fit CID results $\mathbf{b}^{\text{data}}$. Therefore, the distribution of $\mathbf{t}^{\text{sim}}$ is an estimator on the true distribution of the

CID parameters $\mathbf{t}$ and the distribution of $\mathrm{gv_{ch}}$ or $N_{\mathrm{CNO+pep}}$ is obtained through marginalisation:

$$\begin{aligned} P(\mathbf{t}^{\mathrm{sim}} \mid \mathbf{b}^{\mathrm{sim}} = \mathbf{b}^{\mathrm{data}}) &= P(\mathbf{t} \mid \mathbf{b}^{\mathrm{data}}) \\ P(\mathrm{gv_{ch}} \mid \mathbf{b}^{\mathrm{sim}} = \mathbf{b}^{\mathrm{data}}) &= \iint P(\mathbf{t}^{\mathrm{sim}} \mid \mathbf{b}^{\mathrm{sim}} = \mathbf{b}^{\mathrm{data}})\, dN_\nu\, d\Delta r_{\mathrm{dir}} \\ P(N_{\mathrm{CNO+pep}} \mid \mathbf{b}^{\mathrm{sim}} = \mathbf{b}^{\mathrm{data}}) &= \iint P(\mathbf{t}^{\mathrm{sim}} \mid \mathbf{b}^{\mathrm{sim}} = \mathbf{b}^{\mathrm{data}})\, d\mathrm{gv_{ch}}\, d\Delta r_{\mathrm{dir}} \end{aligned} \tag{7.8}$$

The advantage of ASU is that there is no direct matrix inversion involved and because the method takes into account the full fit response for every possible combination of the CID parameters it cannot introduce any bias to the distribution $P(\mathbf{t}^{\mathrm{sim}} \mid \mathbf{b}^{\mathrm{sim}} = \mathbf{b}^{\mathrm{true}})$. This distribution can then be used to construct the unbiased credible intervals and corresponding Gaussian equivalent $\Delta\chi^2$ profiles for the parameters of interest.

The treatment of the nuisance parameters is a simple marginalisation and the inclusion of the systematic uncertainties also follows from the automatic marginalisation of the Gaussian acceptance region with the standard deviation ε. The implementation of the $N_{^8\mathrm{B}+^7\mathrm{Be}}$ correction, as well as the N_{pep} constraint, can either be done during the ASU simulation, where these parameters are sampled according to their correction values and uncertainties, or it can be done after the ASU simulation through a Gaussian convolution. The ASU method is also used in the previous Section 7.6.1 for the production of conditional probability distributions, used for the study of the systematic uncertainty from the Nth-Hit(max) selection.

7.7.3 Equations used for the final results

The full result of the CID analysis is the posterior probability distribution $P(\mathbf{t}|\mathrm{data})$. The problem is that the ASU distribution can only be produced with a finite number of simulations, which means that it cannot be used for the production of credible intervals or exclusion limits with probabilities that are small compared to the statistics of the ASU. To solve this problem the ASU distribution is fitted with a single parameter β that represents a simple shift of the $\Delta\chi^2$ corresponding likelihood distribution $P(\Delta\chi^2(N_{\mathrm{CNO+pep}})) = \exp\left(-\frac{1}{2}\Delta\chi^2(N_{\mathrm{CNO+pep}})\right)$:

$$\beta^{\mathrm{best}} = \arg\min_{\beta} \left\{ -2 \sum_n \ln\left(\mathrm{Poisson}\left(P(n \mid \mathbf{b}^{\mathrm{data}}) \mid \mathrm{norm} \cdot P(\Delta\chi^2(n-\beta))\right)\right) \right\} \tag{7.9}$$

$$P(N_{\mathrm{CNO+pep}} \mid \mathrm{data}) = \mathrm{norm} \cdot P(\Delta\chi^2(N_{\mathrm{CNO+pep}} - \beta^{\mathrm{best}})) \tag{7.10}$$

Here, β^{best} is the value of the fit response bias and it has an uncertainty that depends on the simulated ASU statistic. The same procedure is also used to fit $P(N_{\mathrm{CNO}} \mid \mathrm{data})$, where β^{best} corresponds to the sum of the bias and the mean value of the N_{pep} constraint.

From this fit of the posterior distribution it is possible to calculate all probabilities p within the physically allowed $0 \leq N_{\mathrm{CNO+pep}} \leq N_{\mathrm{total}}$ region. The probability p can then be expressed as a fraction δ of the standard deviation for the same probability of a Gaussian distribution and the corresponding χ^2 value for a single degree of freedom. These δ, χ^2 values can be calculated numerically from the Gaussian cumulative distribution $\Phi(\delta)$ and the χ^2 cumulative distribution $F(\chi^2)$, respectively:

$$\begin{aligned} p = 2\Phi(\delta) - 1 &= F(\chi^2 \mid \mathrm{ndf} = 1) \\ \delta^2 &= \chi^2 \end{aligned} \tag{7.11}$$

Using the posterior distribution it is possible to perform a hypothesis test between the HZ-SSM and LZ-SSM as $P(\text{HZ-SSM} \mid \mathrm{data})/P(\text{LZ-SSM} \mid \mathrm{data})$, where each probability is

given by the multiplication of the expected SSM distribution and the posterior:

$$P(\text{SSM} \mid \text{data}) = \sum_{n=0}^{N_{\text{total}}} P(n \mid \text{data}) \cdot P(n \mid \text{SSM}) \tag{7.12}$$

While the full result of the CID analysis is the posterior distribution it is necessary to communicate the results as credible intervals to make sense of them. For the exclusion of the zero-neutrino hypothesis this is a simple one-sided credible interval. On the other hand here the production of two-sided credible intervals is ambiguous, as the physical boundaries effectively introduce a skew and the peak (mode), mean and median of the posterior distribution are different from each other. For this reason two different credible interval definitions are used here. First, the equal-tail (ET) credible interval for a probability p is defined such, that the same probability $0.5(1-p)$ is on the left side and on the right side outside the interval. The central value of the ET-CI is the median of the posterior distribution, which is the minimum value of the equivalent δ, χ^2 distributions. For a number of neutrinos N the corresponding credible interval probability $p^{\text{ET}}(N)$ is calculated as:

$$p^{\text{ET}}(N) = 1 - 2\min\left\{\sum_{n=0}^{N} P(n \mid \text{data}), \sum_{n=N}^{N_{\text{total}}} P(n \mid \text{data})\right\} \tag{7.13}$$

The other definition is the highest density (HD) credible interval which is the narrowest interval as it includes the values of highest probability density. This HD-CI always includes the peak of the posterior, the minimum values of the equivalent Δ, χ^2 correspond to the highest probability. For a number of neutrinos N_1, N_2 around the peak, which must have the closest probability to each other $P(N_1 \mid \text{data}) = P(N_2 \mid \text{data})$, the CI probability $p^{\text{HD}}(N)$ is:

$$p^{\text{HD}}(N_1) = p^{\text{HD}}(N_2) = \sum_{n=N_1}^{N_2} P(n \mid \text{data}) \tag{7.14}$$

The last thing to consider is the calculation of the CNO-ν rate R from the number of neutrinos N. This is done by using the effective exposure factor f, which is the product of the detector live time, TFC-exposure, fiducial mass and energy efficiency ε_{E} of each phase. For Phase 1 this factor is $f_1 = 42.5 \cdot 100\,\text{t}$ days and for Phase 2+3 it is $f_{2+3} = 158.2 \cdot 100\,\text{t}$ days. The posterior distribution of the CNO-ν rate $P(R)$ then follows from the law of total probability:

$$P(R) = \sum_{n=0}^{N_{\text{total}}} P(n) \cdot \text{Poisson}(n \mid f\,R) = \sum_{n=0}^{N_{\text{total}}} P(n)\,\frac{(f\,R)^n e^{-(f\,R)}}{n!} \tag{7.15}$$

Effectively, this Equation 7.15 corresponds to the simple scaling $R = n/f$ of the x-axis, for expected values smaller than the total number of measured data events $f\,R < N_{\text{total}}$. This is different for the calculation of the combined CNO-ν rate posterior distribution $P(R_{1+2+3})$, which is correctly given by the law of total probability using the independent posterior probability distributions of Phase 1, $P_1(n_1)$ and Phase 2+3, $P_{2+3}(n_{2+3})$), i.e. a two dimensional marginalisation over n_1 and n_{2+3} [6]:

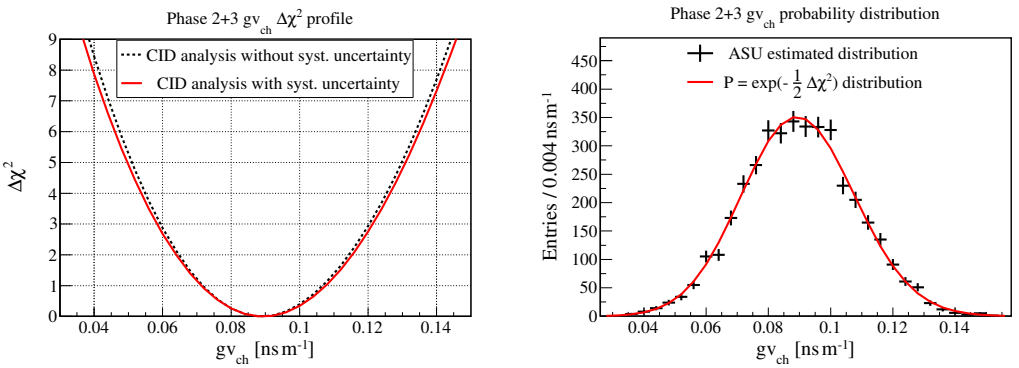
$$P(R_{1+2+3}) = \sum_{n_1=0}^{N^1_{\text{total}}} \sum_{n_{2+3}=0}^{N^{2+3}_{\text{total}}} P_1(n_1) \cdot P_{2+3}(n_{2+3}) \cdot \text{Poisson}(n_1 + n_{2+3} \mid (f_1 + f_{2+3})\,R) \tag{7.16}$$

[6] This can also be written as a one dimensional equation, after the convolution $P_1(n) * P_{2+3}(n) = P_{1+2+3}(n)$.

7.8 Final results

In this section the full CID analysis chain performed for a last time, including all relevant systematic uncertainties and with the correction of the fit response bias. The results are presented as Gaussian equivalent $\Delta\chi^2$ profiles, which take into account the physical limits of the corresponding parameters. Here, the plots are only shown for Phase 2+3 and the relevant plots of Phase 1 can be found in the Appendix C.6.

7.8.1 Calibration of gv_{ch}



(a) (b)

FIGURE 7.35: Final results of the gv_{ch} calibration for Phase 2+3. (a) Resulting $\Delta\chi^2(gv_{ch})$ profile including the systematic uncertainties (red), with $\chi^2/ndf = 1036.2/1017$, $p = 0.33$. The best fit value is $gv_{ch} = 0.089 \pm 0.019\,ns\,m^{-1}$. (b) The ASU distribution of gv_{ch} (black) for 4000 toy-analyses compared to the likelihood distribution (red) of the $\Delta\chi^2$ profiles. There is no fit response bias.

The gv_{ch} calibration of the ^{7}Be-ν edge region is performed as described in Section 7.5.1 and the result is shown in Figure 7.35. The red $\Delta\chi^2$ profile in Figure 7.35(a) is produced with the $-2\ln(P(N_\nu))$ pull-term that now includes the systematic uncertainty on the FV and ε_E, while $\Delta\chi^2$ now explicitly includes the systematic uncertainty from the propagation of the uncertainty on the PMT time constant correction $\Delta\mu$(PMT). The black, dotted line shows the $\Delta\chi^2$ profile without these systematic uncertainties and it can be seen that their contribution is relatively small. The absolute performance of the fit is good, with a $\chi^2/ndf = 1036.2/1017$, p-value $= 0.33$ and the best fit values are $\mathbf{b}^{data} = (N_\nu^{fit} = 40140,\ gv_{ch}^{fit} = 0.0894\,ns\,m^{-1},\ \Delta r_{dir}^{fit} = 1.65\,cm)$. The acceptance region ε is given by the systematic uncertainty propagated from $\Delta\mu$(PMT), $\varepsilon = (N_\nu^{syst} = \pm 247,\ gv_{ch}^{syst} = \pm 0.0018\,ns\,m^{-1},\ \Delta r_{dir}^{syst} = \pm 0.01\,cm)$.

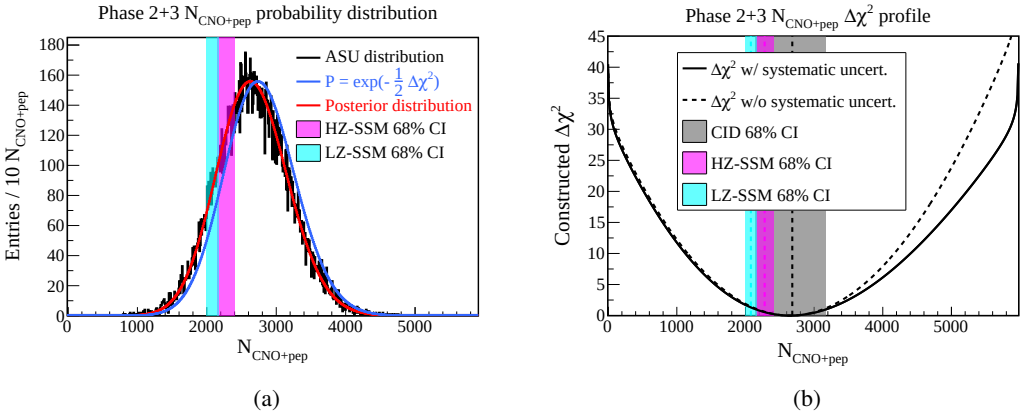
These values are used to produce the black ASU distribution in Figure 7.35(b) with 4000 toy-MC analyses. The ASU distribution is compared to the red probability distribution $P = \exp(-\frac{1}{2}\Delta\chi^2)$, from the corresponding $\Delta\chi^2$ profile of the CID data analysis. As expected, both distributions are well in agreement with each other and the difference between the mean values is $\Delta\langle gv_{ch}\rangle = (0.0001 \pm 0.0003)\,ns\,m^{-1}$. A possible fit response bias can only be negligibly small, which is also the case for Phase 1.

The final result of the gv_{ch} calibration is $gv_{ch} = 0.089 \pm 0.019\,ns\,m^{-1}$ for Phase 2+3 and $gv_{ch} = 0.140 \pm 0.029\,ns\,m^{-1}$ for Phase 1, where the uncertainty is given by the Gaussian 68.3% CI, corresponding to $\Delta\chi^2 = \pm 1$. These results differ by 1.5σ and are used separately for Phase 1 and Phase 2+3 in the $N_{CNO+pep}$ analysis to include possible systematic differences of the detector between Phase 1 and Phase 2+3. The results are summarised in Table 7.16:

	Fit bias $[\mathrm{ns\,m^{-1}}]$	$(\mathrm{gv_{ch}}$ ± stat. ± syst.$)$ $[\mathrm{ns\,m^{-1}}]$
Phase 1	$(1\pm3)10^{-4}$	0.1404 ±0.0278 ±0.0058
Phase 2+3	$(-1\pm4)10^{-4}$	0.0894 ±0.0178 ±0.0043

TABLE 7.16: Results of the $\mathrm{gv_{ch}}$ CID analysis with systematic uncertainties. The fit performance is $\chi^2/\mathrm{ndf} = 874.9/897$, p = 0.70 for Phase 1, $\chi^2/\mathrm{ndf} = 1036.2/1017$, p = 0.33 for Phase 2+3.

7.8.2 Measurement of $N_{\mathrm{CNO+pep}}$



(a) (b)

FIGURE 7.36: Final results of the $N_{\mathrm{CNO+pep}}$ measurement for Phase 2+3. (a) The ASU distribution of $N_{\mathrm{CNO+pep}}$ with 20k toy-analyses (black), compared to the nominal likelihood distribution $\exp(-\frac{1}{2}\Delta\chi^2(N_{\mathrm{CNO+pep}}))$ (blue) and the distribution, corrected for the fit response bias (red). The latter is the posterior distribution $P(N_{\mathrm{CNO+pep}} \mid \text{CID data})$ for the probability of the true, underlying number of neutrino events. The 68.3% CI for the LZ-SSM and HZ-SSM is shown as a cyan and magenta band, respectively. They include the systematic uncertainties of FV, E_{eff}. (b) The equal-tail credible intervals of $N_{\mathrm{CNO+pep}}$, represented in a Gaussian equivalent $\Delta\chi^2$ profile (solid line). The corresponding 68.3% CI ($N^{\mathrm{ET}}_{\mathrm{CNO+pep}} = 2650^{+525}_{-501}$) is shown as a grey band. For comparison the $\Delta\chi^2$ profile without systematics and with fixed nuisance parameters is also shown (dotted line).

The $\Delta\chi^2(N_{\mathrm{CNO+pep}})$ analysis is performed again, as described in Section 7.5.2, including all relevant systematic uncertainties. The absolute performance of the fit is good with a $\chi^2/\mathrm{ndf} = 1000.7/1017$, p = 0.64 and the mean values of the best fit results from the PMT time constant correction systematic are $\mathbf{b}^{\mathrm{data}} = (N^{\mathrm{fit}}_{\mathrm{CNO+pep}} = 2805,\ \mathrm{gv}^{\mathrm{fit}}_{\mathrm{ch}} = 0.0965\,\mathrm{ns\,m^{-1}},\ \Delta r^{\mathrm{fit}}_{\mathrm{dir}} = 2.69\,\mathrm{cm})$. The relevant systematic uncertainties are used to define the ASU acceptance region as $\varepsilon = (N^{\mathrm{syst}}_{\mathrm{CNO+pep}} = \pm88,\ \mathrm{gv}^{\mathrm{syst}}_{\mathrm{ch}} = \pm0.0015\,\mathrm{ns\,m^{-1}},\ \Delta r^{\mathrm{syst}}_{\mathrm{dir}} = \pm0.03\,\mathrm{cm})$. The uncertainties on the FV and ε_{E} are not included here, as they are only relevant for the neutrino rate, but not for the number of neutrinos.

These $\mathbf{b}^{\mathrm{data}}$, ε values are used for the production of the black ASU distribution in Figure 7.36(a), including the ^{8}B+^{7}Be-ν subtraction. The comparison with the $\Delta\chi^2$ corresponding likelihood distribution in blue clearly shows the expected fit response bias β, which is quantified according to Equation 7.10 with a value of $\beta = -109 \pm 4$ for Phase 2+3 and $\beta = -50 \pm 4$ for Phase 1, where the uncertainty comes from the finite statistics of the ASU distribution.

The shifted $\Delta\chi^2$ likelihood distribution describes the simulated ASU distribution well. The χ^2-test between the distributions for a binning of $\Delta N_{\mathrm{CNO+pep}} = 10$ and ≥ 10 entries per bin results in $\chi^2/\mathrm{ndf} = 253.5/243$, p-value = 0.31, while a binning of $\Delta N_{\mathrm{CNO+pep}} = 20$ results in $\chi^2/\mathrm{ndf} = 127.0/134$, p-value = 0.63. As such, the estimation of the posterior distribution $P(N_{\mathrm{CNO+pep}} \mid \text{CID data})$ can be considered valid. This leads to the posterior

probability for zero neutrino events $P(N_{\text{CNO+pep}} = 0 \mid \text{CID}) = 1.5 \cdot 10^{-10}$ for Phase 2+3 and $P(N_{\text{CNO+pep}} = 0 \mid \text{CID}) = 3.5 \cdot 10^{-5}$ for Phase 1, which corresponds to an exclusion of the zero-neutrino hypothesis with 6.4σ and 4.1σ, respectively. The 68.3% CI for the HZ-SMM and LZ-SSM are also shown in Figure 7.36(a) and the probability of these models is $P(\text{HZ} \mid \text{CID})/P(\text{LZ} \mid \text{CID}) = 0.060\%/0.043\%$, which means that the HZ-SSM is 1.4 times more likely to be true than the LZ-SSM, given the CID data of Phase 2+3, while for Phase 1 the value is 1.04.

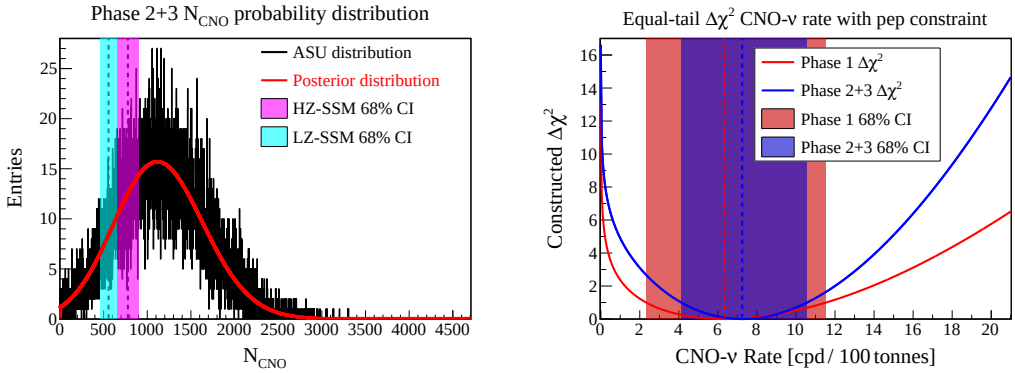
Figure 7.36(b) shows the equal-tail credible intervals, represented as a Gaussian equivalent $\Delta\chi^2$ profile for Phase 2+3. The impact of the physical boundaries can be seen as an increasing slope at the edges. The dotted line shows the $\Delta\chi^2$ profile for a posterior distribution without systematic uncertainties and without the effect of the nuisance parameters, which are fixed to their best fit values and without their marginalisation. This shows the asymmetric influence of the nuisance parameters, as an increased signal to total ratio generates a bigger systematic uncertainty.

The final 68.3% (1σ) credible intervals with all relevant systematic uncertainties are summarised in Table 7.17. The equal-tail and high density CI definitions give only a small difference here.

	Fit bias	P(0|CID)	P(LZ|CID)	P(HZ|CID)	$N^{\text{ET}}_{\text{CNO+pep}}$	$N^{\text{HD}}_{\text{CNO+pep}}$
Phase 1	-50 ± 2	$3.5 \cdot 10^{-5}$	0.163%	0.170%	626^{+239}_{-226}	613^{+237}_{-225}
Phase 2+3	-109 ± 4	$1.5 \cdot 10^{-10}$	0.043%	0.060%	2650^{+525}_{-501}	2626^{+523}_{-499}

TABLE 7.17: Results of the $N_{\text{CNO+pep}}$ CID analysis with systematic uncertainties. The fit performance is $\chi^2/\text{ndf} = 884.8/897$, p = 0.61 for Phase 1, $\chi^2/\text{ndf} = 1000.7/1017$, p = 0.64 for Phase 2+3. The $N_{\text{CNO+pep}}$ uncertainties correspond to the equal-tail (ET) and highest density (HD) 68.3%CI (1σ).

7.8.3 Measurement of the CNO-ν rate



(a) (b)

FIGURE 7.37: Final results of the N_{CNO} measurement for Phase 2+3. (a) The ASU distribution of N_{CNO} with 20k toy-analyses (black) and the corresponding posterior distribution $P(N_{\text{CNO}} \mid \text{CID data})$ (red). The 68.3% CI for the LZ-SSM and HZ-SSM is shown as a cyan and magenta band, respectively. (b) The equal-tail CI of the CNO-ν rate R_{CNO}, represented in a Gaussian equivalent $\Delta\chi^2$ profile. For Phase 1 (red) and Phase 2+3 (blue), with the corresponding 68.3% CI $R_{\text{CNO}}(\text{Phase 1}) = 6.4^{+5.2}_{-4.1}\,(\text{cpd}/100\,\text{t})$, $R_{\text{CNO}}(\text{Phase 2+3}) = 7.3^{+3.4}_{-3.2}\,(\text{cpd}/100\,\text{t})$.

The production of the ASU distribution for the number of CNO-ν events N_{CNO} uses the same $\mathbf{b}^{\mathrm{data}}$, ε values for $N_{\mathrm{CNO+pep}}$, where the number of pep-ν events N_{pep} is constrained. N_{pep} is sampled according to the SSM expectation, including the model uncertainty, as well as the uncertainties from the FV, ε_{E}, and the HZ-SMM and LZ-SMM values are selected with equal probability. Then the number of CNO-ν events N_{CNO} is sampled from $0 \leq N_{\mathrm{CNO}} \leq N_{\mathrm{total}} - N_{\mathrm{pep}}$. Effectively this is the same as the convolution of the $P(N_{\mathrm{CNO+pep}} \mid \mathrm{CID})$ posterior with the distribution of the expected (negative) number of pep-ν events.

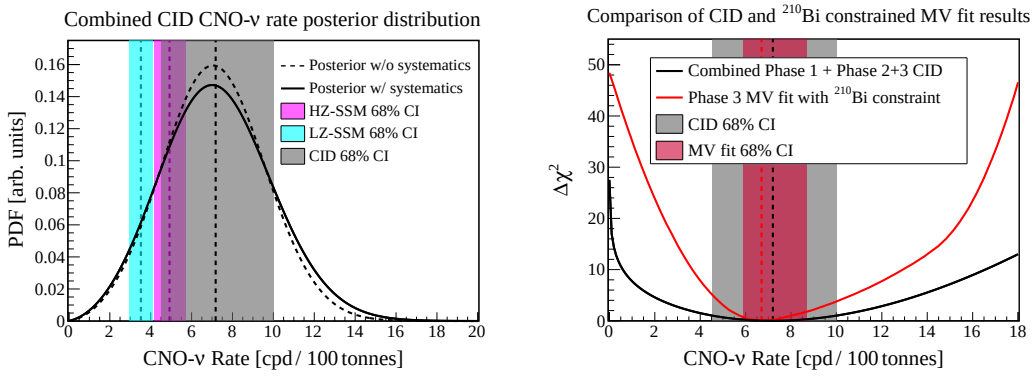
Figure 7.37(a) shows the N_{CNO} ASU distribution in black for Phase 2+3, where the number of entries at $N_{\mathrm{CNO}} = 0$ is $1/20\mathrm{k} = 5.0 \cdot 10^{-5}$ for Phase 2+3 and $25/20\mathrm{k} = 1.25 \cdot 10^{-3}$ for Phase 1 (C.19(a)). The ASU estimated posterior distribution $P(N_{\mathrm{CNO}} \mid \mathrm{CID})$ is shown in red, and given the CID data of Phase 2+3, results in a probability for exactly zero CNO-ν events of $P(N_{\mathrm{CNO}} = 0 \mid \mathrm{CID}) = 5.78 \cdot 10^{-5} (= 1.16/20\mathrm{k})$, which corresponds to a one-sided exclusion of the zero CNO-ν hypothesis with $> 4\sigma$. For Phase 1 this value is $P(N_{\mathrm{CNO}} = 0 \mid \mathrm{CID}) = 1.35 \cdot 10^{-3} (= 27.0/20\mathrm{k})$ corresponding to a one-sided $> 3\sigma$ exclusion. Both posterior probabilities are well in agreement with the estimations from the ASU distributions. The probability that a true, underlying number of exactly zero CNO-ν events is responsible for the measured data $\cos\alpha$ histograms in both Phase 1 and Phase 2+3 is given by the product of Phase 1 and Phase 2+3 as $P(N_{\mathrm{CNO}} = 0 \mid \text{Phase 1 and Phase 2+3}) = 7.80 \cdot 10^{-8}$. This means that the zero CNO-ν hypothesis can be excluded with 5.3σ. This value depends on the assumptions made on the correction and systematic uncertainties from the number of non-CNO neutrinos. It is therefore assumed to be a slightly conservative estimation, as the uncertainty from pep-ν vs. CNO-ν MC is included, as well as the statistical uncertainty on the expected number of pep-ν events.

The N_{CNO} 68.3% CI for the HZ-SMM and LZ-SSM are also shown in Figure 7.37(a) and the probability of these models is $P(\mathrm{HZ} \mid \mathrm{CID})/P(\mathrm{LZ} \mid \mathrm{CID}) = 0.062\%/0.042\%$, which means that the HZ-SSM is 1.5 more likely to be true than the LZ-SSM, given the CID data of Phase 2+3, while for Phase 1 the value is 1.03. The N_{CNO} posterior distributions of Phase 1 and Phase 2+3 result in 68.3% credible intervals of $N^{\mathrm{ET}}_{\mathrm{CNO}}(\text{Phase 1}) = 270^{+221}_{-171}$, $N^{\mathrm{HD}}_{\mathrm{CNO}}(\text{Phase 1}) = 208^{+188}_{-180}$ and $N^{\mathrm{ET}}_{\mathrm{CNO}}(\text{Phase 2+3}) = 1146^{+526}_{-494}$, $N^{\mathrm{HD}}_{\mathrm{CNO}}(\text{Phase 2+3}) = 1115^{+520}_{-496}$. The calculation from the number of CNO-ν events N_{CNO} to a neutrino rate R is given by Equation 7.15.

Figure 7.37(b) shows the ET-CI of the CNO-ν rate R_{CNO} for Phase 1 and Phase 2+3 in red and blue, respectively, represented as a Gaussian equivalent $\Delta\chi^2$ profile for Phase 2+3. The impact of the physical boundary can be seen for the ET-CI, where value of $\Delta\chi^2$ increases for $R_{\mathrm{CNO}} \to 0$. As the posterior distribution is bound by $R_{\mathrm{CNO}} \geq 0$ the probability for the true CNO-ν rate to be outside the ET-CI also goes to zero when R_{CNO} goes to zero. Consequently, the ET-CI representation in units of $\Delta\chi^2$ diverges to infinity for values of R_{CNO} getting closer to zero. The HD-CI in Appendix C.19(b) does not show this effect, as it only makes a statement about the probability of R_{CNO} being inside the CI. It does not make a statement on how probable is is to find R_{CNO} on the left of the HD-CI, relative to the probability to find it on the right side of the HD-CI. Instead the HD-CI corresponding $\Delta\chi^2$ profile exhibits a kink, where the lower boundary of the CI has reached the minimum rate of zero cpd/100 t. The final 68.3% (1σ) credible intervals of R_{CNO} with all relevant systematic uncertainties are summarised in Table 7.18.

Figure 7.38(a) shows the posterior density function for the combined CNO-ν rate of Phase 1 and Phase 2+3, which is calculated according to Equation 7.16. The black line shows the posterior density function[7] including the systematic uncertainties, while the dotted line shows a hypothetical density function without systematic uncertainties and without the influence of

[7] Please note that the number of neutrinos is a non-negative integer providing a probability distribution for each individual number value, while the neutrino rate is a non-negative real number, with a probability density function.



(a) (b)

FIGURE 7.38: The final result of the Phase 1 and Phase 2+3 combined CNO-ν rate R_{CNO}, with $R^{\mathrm{ET}}_{\mathrm{CNO}} = \left(7.2^{+2.8}_{-2.7}\right)$ (cpd / 100 t). (a) The CNO-ν rate posterior density function with all relevant systematics (solid line) is compared to the (expected) PDF without systematic uncertainties and without the effect of the CID nuisance parameters (dotted line). The 68.3% CI for the LZ-SSM and HZ-SSM expectation of the CNO-ν rate is shown as a cyan and magenta band, respectively. (b) The comparison of the CID measurement of the CNO-ν rate (black) to the CNO-ν rate from the multivariate fit with the ^{210}Bi constraint (red) [45]. The 68.3% equal-tail CI of the CID measurement is shown as a grey band, while the 68.3% CI of the MV fit is shown in red with $R^{\mathrm{MV}}_{\mathrm{CNO}} = \left(6.7^{+2.0}_{-0.8}\right)$ (cpd / 100 t).

the nuisance parameters, which are fixed here to their best fit values. The 68.3% CI for the LZ-SSM and HZ-SSM is shown as a cyan and magenta band, respectively. This results in a probability ratio for these models of $P(\mathrm{HZ} \mid \mathrm{CID})/P(\mathrm{LZ} \mid \mathrm{CID}) = 0.107\%/0.062\%$, where the HZ-SSM is 1.7 times more likely to be true than the LZ-SSM, given the combined Phase 1 and Phase 2+3 CNO-neutrino rate posterior distribution.

The equal-tail and highest density 68.3% (1σ) credible intervals are:

- $R^{\mathrm{ET}}_{\mathrm{CNO}} = \left(7.2^{+2.5}_{-2.5}\,(\mathrm{stat.})^{+1.2}_{-0.9}\,(\mathrm{syst.})\right)\frac{\mathrm{cpd}}{100\,\mathrm{t}} = \left(7.2^{+2.8}_{-2.7}\right)\frac{\mathrm{cpd}}{100\,\mathrm{t}}$
- $R^{\mathrm{HD}}_{\mathrm{CNO}} = \left(7.0^{+2.5}_{-2.5}\,(\mathrm{stat.})^{+1.2}_{-0.9}\,(\mathrm{syst.})\right)\frac{\mathrm{cpd}}{100\,\mathrm{t}} = \left(7.0^{+2.8}_{-2.7}\right)\frac{\mathrm{cpd}}{100\,\mathrm{t}}$

These credible intervals are dominated by their statistical uncertainty and the positive systematic uncertainties are larger than the negative systematic uncertainties because of the asymmetric impact of the nuisance parameters on the neutrino signal MC $\cos\alpha$ histograms, while they do not impact the background MC histograms. The two-sided 5σ ET-CI is $0.008\,(\mathrm{cpd}/100\,\mathrm{t}) < R_{\mathrm{CNO}} < 22.795\,(\mathrm{cpd}/100\,\mathrm{t})$.

Figure 7.38(b) shows the Gaussian equivalent $\Delta\chi^2$ profile of the ET-CI of the combined CID analysis in black, in comparison to the improved measurement of CNO-ν rate using the ^{210}Bi constrained multivariate fit [45] in red. Both the CID and MV fit results are fully in agreement with each other. The MV fit has smaller uncertainties with a 68.3% CI of $R^{\mathrm{MV}}_{\mathrm{CNO}} = \left(6.7^{+2.0}_{-0.8}\right)$ (cpd / 100 t) and also a more significant exclusion of the zero CNO-ν hypothesis at about 7σ CL [45]. It is interesting to note that the CID uncertainties are only bigger by a factor of 3.4 for the negative uncertainty and 1.5 for the positive uncertainty. The final CID results on the CNO-neutrino rate R_{CNO} with all relevant systematic uncertainties are summarised in Table 7.18.

7.9 Conclusion and outlook

In this chapter the *Correlated and Integrated Directionality* (CID) has been applied on the energy region for which CNO+pep-ν events are expected to dominate the signal contribution.

Results	P($N_{\mathrm{CNO}} = 0$ |CID)	P(LZ|CID)	P(HZ|CID)	$\mathrm{R}^{\mathrm{ET}}_{\mathrm{CNO}}$ [cpd / 100 t]	$\mathrm{R}^{\mathrm{HD}}_{\mathrm{CNO}}$ [cpd / 100 t]
Phase 1	$1.35 \cdot 10^{-3}$	0.198%	0.204%	$6.4^{+5.2}_{-4.1}$	$4.9^{+4.5}_{-4.3}$
Phase 2+3	$5.78 \cdot 10^{-5}$	0.042%	0.062%	$7.3^{+3.4}_{-3.2}$	$7.0^{+3.3}_{-3.2}$
Combined	$7.80 \cdot 10^{-8}$	0.062%	0.107%	$7.2^{+2.8}_{-2.7}$	$7.0^{+2.8}_{-2.7}$

TABLE 7.18: Results of the CNO-ν rate CID analysis with systematic uncertainties. The uncertainties correspond to the equal-tail (ET) and highest density (HD) 68.3% credible intervals.

The goal has been to provide a measurement on the number of CNO+pep-neutrinos $N_{\mathrm{CNO+pep}}$, as well as a measurement of the CNO-ν rate where pep-ν is constrained to the SSM expected value. For this purpose the CID analysis has been significantly improved, relative to the previous measurement of ^{7}Be-neutrinos.

The two data sets of Phase 1 and Phase 2+3 are investigated and analysed independently, to search for a potential systematic uncertainty related to the change of the detector response over time. The data is selected in a optimised fiducial volume and a $N_{\mathrm{h}}^{\mathrm{geo}}$ region of interest, to maximise the expected neutrino signal over background ratio. This includes the application of the TFC-cut to minimise the contribution of cosmogenic ^{11}C. Likewise, the CID parameters, namely the number Nth-Hit(max) $\cos\alpha$ histograms and the binning of the those histograms, are selected to optimise the sensitivity of the CID analysis. This is done by maximising the χ^2-test difference between the pure background and pure neutrino signal MC $\cos\alpha$ histograms. Here, a value of Nth-Hit(max) is chosen which is larger than what is necessary to fully include the information of the direct Cherenkov hits on the PMTs. The reason for this is that the systematic mis-reconstruction in the initial electron direction Δr_{dir} contains the indirect information of the Cherenkov hits, as they systematically influence the position reconstruction. This effect has a noticeable impact on the $\cos\alpha$ shape of the scintillation hits, which then propagate this indirect Cherenkov hit information to much later Nth-Hits than the information provided by the direct Cherenkov hits

As the CID analysis is based on the time ordered Nth-Hits of the events on the PMTs it is necessary to select PMTs with a good hit time behaviour and deselect PMTs that exhibit systematic differences between the data and the MC model. For this selection the TFC-tagged ^{11}C region is used, where the time-of-flight corrected hit time distribution of each individual PMT is fitted. Systematically incompatible PMTs are easily identifiable by their different fit results. This analysis also provides a correction of the relative hit times between the data PMTs, which is necessary to minimise systematic differences between the data and MC CID $\cos\alpha$ histograms. The PMT selection and hit time corrections are applied individually on Phase 1, Phase 2 and Phase 3, to take into account the changing detector response over time due to the loss of live PMTs.

Another improvement of the CID analysis here is the calibration of the effective Cherenkov group velocity correction $\mathrm{gv}_{\mathrm{ch}}$ on the high statistics ^{7}Be-ν edge region. This means, that the neutrinos themselves are used to calibrate the Cherenkov hit time behaviour between the data and the MC. The effective calibration results of $\mathrm{gv}_{\mathrm{ch}}(\text{Phase 1}) = 0.140 \pm 0.029\,\mathrm{ns\,m^{-1}}$, $\mathrm{gv}_{\mathrm{ch}}(\text{Phase 2+3}) = 0.089 \pm 0.019\,\mathrm{ns\,m^{-1}}$ show a 1.5σ difference between Phase 1 and Phase 2+3 and they have a substantially smaller uncertainty, compared to the calibration using γ-sources $\mathrm{gv}_{\mathrm{ch}}(\gamma\text{-sources}) = 108 \pm 0.039\,\mathrm{ns\,m^{-1}}$. Using the $\mathrm{gv}_{\mathrm{ch}}$ results independently is a conservative approach, as the statistical uncertainty for a combined calibration would be smaller, but it potentially takes into account systematic differences between Phase 1 and Phase 2+3, should they exist.

The systematic uncertainties of the gv_{ch} calibration and $N_{CNO+pep}$ measurement are investigated independently for Phase 1 and Phase 2+3, which allows a thorough examination of the CID systemics on two independent energy regions and two independent time periods. There is an effort to perform the estimation of the systematic uncertainties in a reasonable manner and not maximally conservative, which makes an in-depth examination necessary. For this purpose, toy-MC studies are performed for the production of conditional probability distribution to investigate the Nth-Hit(max) selection, PMT selection and histogram binning selection. The results of these studies is that all these selections do not contribute to the systematic uncertainties. Likewise, the different background species of ^{210}Bi, ^{11}C, and external γ's do not contribute a systematic uncertainty, as they are indistinguishable in their $\cos\alpha$ distributions, given the statistics of the data. In this regard, the CID is more robust than the multivariate (MV) fit, as all background species look the same and are well distinguishable from the $\cos\alpha$ distribution of the neutrino signal.

The relevant systematic uncertainties come from the difference in the energy distribution of CNO-ν and pep-ν events and the uncertainty of the relative data PMT hit time correction. The different neutrino recoil-electron energies result in a slightly different Cherenkov/scintillation ratio of the neutrino events, as well as a slightly different electron scattering angle. This provides a systematic uncertainty of 5.2% for Phase 1 and 4.6% for Phase 2+3, relative to the final result of the CNO-ν measurement. The data PMT hit time correction corresponds to a relative systematic uncertainty of 9.6% for Phase 1 and 5.5% for Phase 2+3. The rest of the systematic uncertainties are not unique to the CID analysis but are also relevant for the MV fit, such as the correction for the number of ^{8}B-ν, ^{7}Be-ν and pep-ν events, as well as the uncertainty on the FV and RoI. They provide a combined, relative systematic uncertainty of 10.9% for Phase 1 and 4.5% for Phase 2+3, for the measurement of the CNO-ν rate.

The CID specific fit response has been investigated and it has been found that the fit itself introduces a bias between the best fit value of $N_{CNO+pep}$ and the true, underlying number of neutrino events. The reason for this is that the nuisance parameters gv_{ch}, Δr_{dir} only affect the signal $\cos\alpha$ shape, which shifts the best fit $N_{CNO+pep}$ to larger values. This bias is estimated with the so called *Acceptance Sampling Unfolding*, which is a toy-MC sampling of the Bayesian posterior distribution for the measurement of $N_{CNO+pep}$, and N_{CNO} with the pep-ν constraint. Here, the results of toy-MC analyses are saved if and only if they reproduce the best fit result of the data analysis, which provides a distribution of the true, underlying values of $N_{CNO+pep}$, i.e. the posterior. The bias introduces an overestimation of the naive $\Delta\chi^2$ best fit $N_{CNO+pep}$ value of $\sim 8\%$ for Phase 1 and $\sim 4\%$ for Phase 2+3, compared to correct calculation of the posterior.

Phase 1 and Phase 2+3 are analysed independently to investigate if the detector response changes over time to such a degree, that the CID method becomes unusable. This is not the case, as the CID measurement results, as well as the systematic behaviour of the background, for Phase 1 and Phase 2+3 show no visible systematic differences between each other. Therefore, it is possible to combine the independent results of Phase 1 and Phase 2+3, which provides an exclusion of the zero CNO-ν hypothesis with a probability of $7.8 \cdot 10^{-8}$, or $> 5\sigma$ CL. Thus, the CID measurement of N_{CNO} corresponds to the detection of solar CNO-neutrinos. The equal-tail 68.3% (1σ) credible interval of the CNO-neutrino rate is $R^{CID}_{CNO} = \left(7.2^{+2.8}_{-2.7}\right)$ (cpd / 100 t). The MV fit with the ^{210}Bi background and pep-ν constraint has a better exclusion of the zero CNO-ν hypothesis at about $\sim 7\sigma$ and a CNO-ν rate measurement with smaller uncertainties $R^{MV}_{CNO} = \left(6.7^{+2.0}_{-0.8}\right)$ (cpd / 100 t) [45].

On the other hand, the MV fit only provides a relatively indirect proof of the solar origin of the measured signal. There, the proof that the signal corresponds to solar CNO-neutrinos comes from the background subtracted energy distribution (Figure 3 in [45]), which is well in agreement with the expected energy distribution of solar CNO-ν events in Borexino. Compared to that, the CID provides a more direct proof for the solar origin of the measured

signal. By construction the measured signal can only originate in the Sun. Thus, the CID method provides an epistemologically more convincing argument for the solar origin of the assumed N_{CNO} signal, compared to the MV fit[8]. Furthermore, the CID does not depend on any background constraint and needs no assumptions about the fluid dynamics of the Borexino detector [154] and the transport mechanisms of the ^{210}Bi and ^{210}Po isotopes in the LS. The last difference between the CID and the MV fit is that the combined CID result uses the entire detector live time, corresponding to 3628.5 days, while the MC fit only uses a part of Phase 3, corresponding to 1431.6 days. The conclusion is then, that the CID measurement can be understood as a supplementary proof of the solar origin of the CNO-ν events, which compliments the more sensitive, ^{210}Bi constrained MV fit. Both measurements taken together provide a more reliable experimental proof for the existence of the CNO-cycle neutrinos than each single measurement on its own.

Therefore, the natural, next step for the CNO-ν measurement in Borexino is the combination of both the MV fit with the CID. Here, a very first estimation can be given by simply combining the posterior, likelihood distributions, which results in a best fit value of $\left(6.8^{+1.6}_{-0.8}\right)$ (cpd / 100 t), with an exclusion of the zero-CNO-ν hypothesis by more than 8σ. A more sophisticated approach would be to implement the CID posterior distribution directly into the MV fit, such that the neutrino signal to background ratio can be constrained trough the directionality of the neutrino events. This would not only provide a direct increase in the sensitivity for the measurement of the CNO-ν rate, but it would simultaneously apply an additional, independent constraint on the problematic ^{210}Bi background. This second effect is then expected to increase the sensitivity on the CNO-ν rate even further. Ultimately, the full implementation of the CID into the MV fit would provide the best currently achievable measurement of the CNO-ν rate [167], all thanks to the unprecedented radio-purity of the Borexino detector. This assumes that the MC fit and CID analyses are independent from each other, which seems reasonable, but should be investigated further. The task of this implementation is currently being worked on by [redacted], as part of his dissertation.

[8] Not that there is any reasonable doubt about that in the MV fit approach. The CID solves more of a philosophical problem, compared to the practical problem of the actual CNO measurement.

Chapter 8

Conclusions and Outlook

The main topic of this thesis concerns the measurement of low energy solar neutrinos, trough their associated directionality in the high light-yield liquid scintillator (LS) Borexino detector. On the one hand, Borexino is the ideal candidate for such an endeavour, as it offers an unprecedented radio-purity. On the other hand, the directionality is not easily accessible in Borexino, exactly because its monolithic scintillator design and high light-yield make the directional Cherenkov hits exceptionally sub-dominant to the scintillation. The task is further complicated by the relatively fast decay time of the scintillator, in combination with the lack of dedicated hardware and electronics systems that could allow for the separation between the scintillation and Cherenkov light.

Borexino was never intended nor expected to be able to make use of the directional Cherenkov information. It was utilised nonetheless, through the *Correlated and Integrated Directionality* (CID) method, specifically developed as part of this thesis. The CID method works by calculating the angle between the reconstructed PMT hit photon direction and the known position of the Sun for the first few hits of every event. The integration over a large number of events then provides an angular distribution, which can be used to make a statistical inference on the number of solar neutrino events. Ultimately, the CID accesses the neutrino directionality not through the event direction reconstruction, but rather through the event position reconstruction. This is a task that is well suited for liquid scintillation detectors. Another important aspect of the CID is the use of the relative time-like variable called *Nth-Hit*, which is the time sorted sequence of PMT hits. The use of only the earliest Nth-Hits provides an improved Cherenkov to scintillation hit ratio, which increases the sensitivity of the analysis. Additionally, the Cherenkov hits imprint their directionality through a bias on the position reconstruction. This in turn influences the angular distribution of the scintillation hits, but only for solar neutrino signal events, which also increases the sensitivity of the CID analysis.

CID measurements

The first use of the CID is the measurement of the 0.862 MeV line of the ^{7}Be-neutrinos, as a proof-of-principle for its feasibility. For this measurement the systematic mis-behaviour of the Cherenkov hit times between the data and the MC model has been measured using the available γ calibration sources. The analysis is performed in a conservative fashion, using only Phase 1 of the detector and estimating the systematic uncertainties maximally conservative. The measurement is performed close to the Compton-like ^{7}Be-ν edge, between 0.56 MeV–0.76 MeV, which selects a total number of 19904 events. The zero-neutrino hypothesis can be excluded with $> 6\sigma$, using the CID analysis. This corresponds to a detection of sub-MeV solar neutrinos with a fit that depends solely on the directionality provided by the Cherenkov hits. The measured number of solar neutrinos is $N_{\text{solar-}\nu} = 10887^{+2386}_{-2103}(\text{stat.}) \pm 947(\text{syst.})$. This is well in agreement with the expected values of the SSM [35] $\text{N}^{\text{SSM}}_{\text{solar}-\nu} = 10187^{+541}_{-1127}$, where the uncertainty includes the difference between low (LZ) and high (HZ) metallicity models. The ^{7}Be-ν rate is then calculated as $R(^7\text{Be})_{\text{CID}} = 51.6^{+13.9}_{-12.5}(\text{stat. + syst.})\text{cpd}/100\text{t}$,

after constraining the number of CNO+pep-neutrino events. This CID measurement is well in agreement with the result of the corresponding Phase 1 spectroscopy of $R(^7\text{Be}) = 47.87 \pm 2.28\,\text{cpd}/100\,\text{t}$ [83]. The spectral fitting provides a significantly greater sensitivity for the measurement of the ^{7}Be-neutrinos. The sensitivity of the CID analysis is limited here due to the low number of Cherenkov hits per neutrino events. Nonetheless, this result is the first directional detection and measurement of sub-MeV solar neutrinos. At the same time, it is also a hybrid detection of sub-MeV neutrinos, as the events are explicitly triggered by the scintillation hits. This proofs the principle feasibility of the hybrid detection concept, even in a large scale, monolithic LS detector.

The second use of the CID, with significant improvements of the analysis, is the measurement of the CNO-ν rate. Here, the effective Cherenkov group velocity correction gv_{ch} is quantified on the ^{7}Be-ν edge region, while the main analysis is performed in a decoupled energy window between 0.85 MeV–1.3 MeV, optimised on the expected ratio between the neutrino signal to the background. This means that the neutrinos themselves are used to constrain the nuisance parameter gv_{ch}, which provides a better sensitivity than the use of γ calibration sources. Another improvement of this CID analysis is the use of both the earliest, direct Cherenkov hits, as well as later PMT hits that are influenced indirectly by the Cherenkov hits. This indirect directionality stems from the influence of the Cherenkov hits on the position reconstruction, which applies a bias in direction of the solar neutrino. This bias then also impacts the later scintillation hits of the solar neutrino signal events to some extent, while it is not present for the angular distribution of the background. The full detector live time of 3628.7 days is analysed in two separate data sets for Phase 1 and Phase 2+3, to search for potential systematic differences stemming from a change of the detector response over time. The final results of this CID analysis is calculated through the toy-MC sampling of the Bayesian posterior distributions. For Phase 1 the number of CNO+pep-ν events is measured as $N_{\text{CNO+pep}} = 626^{+239}_{-226}\,(\text{stat.} + \text{syst.})$, out of 1990 selected events, and for Phase 2+3 the result is $N_{\text{CNO+pep}} = 2650^{+525}_{-501}\,(\text{stat.} + \text{syst.})$. Both results are well in agreement with the SSM [35] expectations of $N^{\text{LZ-SSM}}_{\text{CNO+pep}} = 561 \pm 25$, $N^{\text{HZ-SSM}}_{\text{CNO+pep}} = 615 \pm 34$ for Phase 1, and $N^{\text{LZ-SSM}}_{\text{CNO+pep}} = 2091 \pm 86$, $N^{\text{HZ-SSM}}_{\text{CNO+pep}} = 2293 \pm 126$ for Phase 2+3. The systematic uncertainties are investigated in depth, using toy-MC studies, as well as independent background data sets. It is found, that the change of the detector response over time does not introduce a prohibitive systematic difference between Phase 1 and Phase 2+3. Therefore, the result of the CNO-ν rate measurement is given by the combination of Phase 1 and Phase 2+3, after constraining the number of pep-ν events in each data set with the SSM predictions. This combined CID analysis provides a probability for the zero CNO-ν hypothesis of $P(N_{\text{CNO}} = 0) = 7.8 \cdot 10^{-8}$, which corresponds to a detection of CNO-neutrinos with $> 5\sigma$, using only their directionality. This result is a very direct proof of the solar origin of the assumed CNO-ν events. The measured CNO-ν rate is $R_{\text{CNO}} = \left(7.2^{+2.5}_{-2.5}\,(\text{stat.})^{+1.2}_{-0.9}\,(\text{syst.})\right)\,(\text{cpd}/100\,\text{t})$. The CID posterior distribution implies that the investigated HZ-SSM is 1.7 times more likely to be true than the corresponding LZ-SSM. This CID measurement is in agreement with the result of the spectral, multi-variate (MV) fit of $R^{\text{MV}}_{\text{CNO}} = \left(6.7^{+2.0}_{-0.8}\right)\,(\text{cpd}/100\,\text{t})$ [45]. The MV fit provides smaller uncertainties, but it depends on an external constraint of the ^{210}Bi background. This constraint requires a good thermal stability of the detector, to limit convective motions of the LS, and it is only possible to apply it for Phase 3. In this regard the CID analysis is more robust than the spectral fit, as the CID signal and background distributions have no degeneracy between their angular distributions and an external background constraint is not necessary.

Outlook for the CID

What then could the future of the CID method be? First, the parameters used for the CID analysis seem reasonable well decoupled from the MV fit, where the energy spectrum is fitted

simultaneously with the radial distribution of the selected events. The position of the Sun, i.e the true direction of the solar neutrino events, does not depend on the neutrino energy nor on the reconstructed event position r in the Borexino detector. Similarly, the time-of-flight corrected PMT hit time, by design, does not depend on the reconstructed event radius. Thus, for a given energy selection, the CID measurement of $N_{\mathrm{CNO+pep}}$ is (conditionally) independent from the MV fit result and consequently both analyses can be combined. This can be achieved by implementing the $N_{\mathrm{CNO+pep}}$ to total ratio measurement of the CID as an independent pull-term into the MV fit. This would improve its sensitivity while at the same time taking correctly into account the correlations of the pep-ν, CNO-ν and ^{210}Bi parameters in the MV fit. At the time of writing this idea is actively under investigation by the collaboration. The goal is to produce the best possible measurement of the CNO-ν rate using the Borexino detector.

Second, the uncertainty of the CID measurement is dominated by the statistical uncertainty, which could potentially be drastically improved in future experiments. The size of the statistical uncertainty is not only realised due to the finite number of neutrino events, but more so due to the unfavourable Cherenkov to scintillation hit ratio present in Borexino. For the same number of data events the impact of a better separation of the neutrino signal Cherenkov to scintillation ratio can be estimated through a toy-MC. For example, all other things being equal, improving the Cherenkov to scintillation ratio by a factor of 2, from $\sim 0.47\%$ to $\sim 0.95\%$, would reduce the statistical uncertainty on the CNO-ν rate measurement by a factor of ~ 2 from 2.5 (cpd / 100 t) down to 1.2 (cpd / 100 t). A hypothetical, perfect separation of Cherenkov and scintillation hits would provide a reduction of the statistical uncertainty by a factor of ~ 8, down to 0.3 (cpd / 100 t). Of course, in reality a simple reduction of the number of scintillation hits would increase the position reconstruction uncertainty and consequently the spread from the time-of-flight correction, which would then negatively affect the Cherenkov to scintillation ratio. The true potential of CID under advantageous circumstances must be studied by taking into account the full detector response. Examples for target materials with a better Cherenkov to scintillation ratio would be water based LS [73, 130], or the use of a slow LS with a decay constant larger than PPO, such as linear-alkylbenzene (LAB) with a slow fluorophore [168]. While the exact sensitivity of a CID analysis used in a purpose-built hybrid detector, such as Theia [72] or Jinping [74, 75], needs to be estimated trough a dedicated sensitivity study it seems almost certain that the use of improved target materials, better hardware [123, 125, 124] and improved vertex reconstruction algorithms [126, 129, 128] will likely empower the CID method. Every development that helps an event-by-event direction reconstruction in a LS should in principle also improve the CID analysis.

A very concrete example for the application of CID in a future detector is JUNO [69], a 20 kton LS detector at a depth of 1800 m water equivalent, which is currently under construction and is expected to start data taking in 2024. The main goal of JUNO is the determination of the neutrino mass ordering, but it has also a large number of other physics goals, including the measurement of solar neutrinos. For this task the main challenges are the intrinsic radio-purity of the LS, the electronic PMT pile-up due to a large rate of ^{14}C events from the enormous LS volume, as well as the relatively high muon flux and consequent production of ^{11}C background, compared to Borexino (at 3800 m water equivalent). The measurement of ^{8}B-neutrinos above 2 MeV is a prime candidate due to the large size of JUNO [169], while measuring ^{7}Be-neutrinos is considered challenging [69], given the expected background rate. In this context CID could potentially provide a substantial improvement to the solar neutrino measurement, due to the expected large neutrino signal event statistics, as well as the use of the relatively slow LAB scintillator. The latter should in principle allow for a better separation between Cherenkov and scintillation hits, based on their arrival time. On the other hand, JUNO uses PMTs with a relatively large transit time spread, as well as PPO as a primary and bis-MSB as a secondary wavelength shifter [170]. This could have the consequence that more Cherenkov photons are absorbed and re-emitted as isotropic scintillation light, compared to

Borexino. The exact potential of the CID method in the JUNO experiment must be explored in a dedicated sensitivity study, a task that has been recently started by [redacted] as part of her dissertation.

It seems likely that a dedicated event-by-event direction reconstruction will in general provide a better measurement sensitivity in a purpose-built detector than the CID. At the same time it seems also reasonable that for a sufficiently low trigger threshold, there will be a low energy region, in which CID will outperform the event-by-event direction reconstruction analysis methods. The reason for this is that below a certain signal energy the number of Cherenkov photons per event will become so small, that an event-by-event direction reconstruction becomes useless, as many events will produce an insufficient number of Cherenkov hits to be detected. The CID on the other hand can still perform well for even a very small Cherenkov hits per event ratio, as it can make up the small Cherenkov hit statistic with a sufficiently large number of neutrino events. This is already the case for the Borexino detector and the ^{7}Be-ν and CNO-ν CID measurements presented in this thesis.

Another interesting, potential candidate for the application of CID is the measurement of pp-neutrinos. The maximum energy of pp-neutrinos is 0.42 MeV [41], where the maximum energy transfer for the electron scattering is $T = 0.26\,\text{MeV}$. This is larger than the minimal energy $T > 0.16\,\text{MeV}$ required for Cherenkov radiation in a typical LS with a refractive index of ~ 1.55. Thus, it seems likely that the CID can be applied for the measurement of pp-neutrinos, while they are almost certainly outside of the scope of a directional analysis using an event-by-event direction reconstruction. Such a measurement could be tried in Borexino, but the end of data taking in October 2021 and the following, slow dissolution of the collaboration make this proposal unlikely. Thus, a potential CID measurement of pp-neutrinos is likely reserved for future detectors with advantageous conditions, such as in the proposed Serappis detector [70]. Here again, it is reasonable to suggest a sensitivity study for the possible impact that CID would have on the pp-ν measurement, as well as a simultaneous study of the detector setup, under which the CID sensitivity could be maximised. This task was recently started by [redacted] as the topic of his master thesis in the group of [redacted] in Mainz.

The *Correlated and Integrated Directionality* offers a novel and robust method for solar neutrino detection. It can be used on its own, as a supplementary measurement, or even as an additional constraint in a spectral, multivariate analysis. The Sun is always shining in neutrinos and hopefully it will continue to shine on the CID.

Appendix A

Measurement of the Effective Scintillation Group Velocity

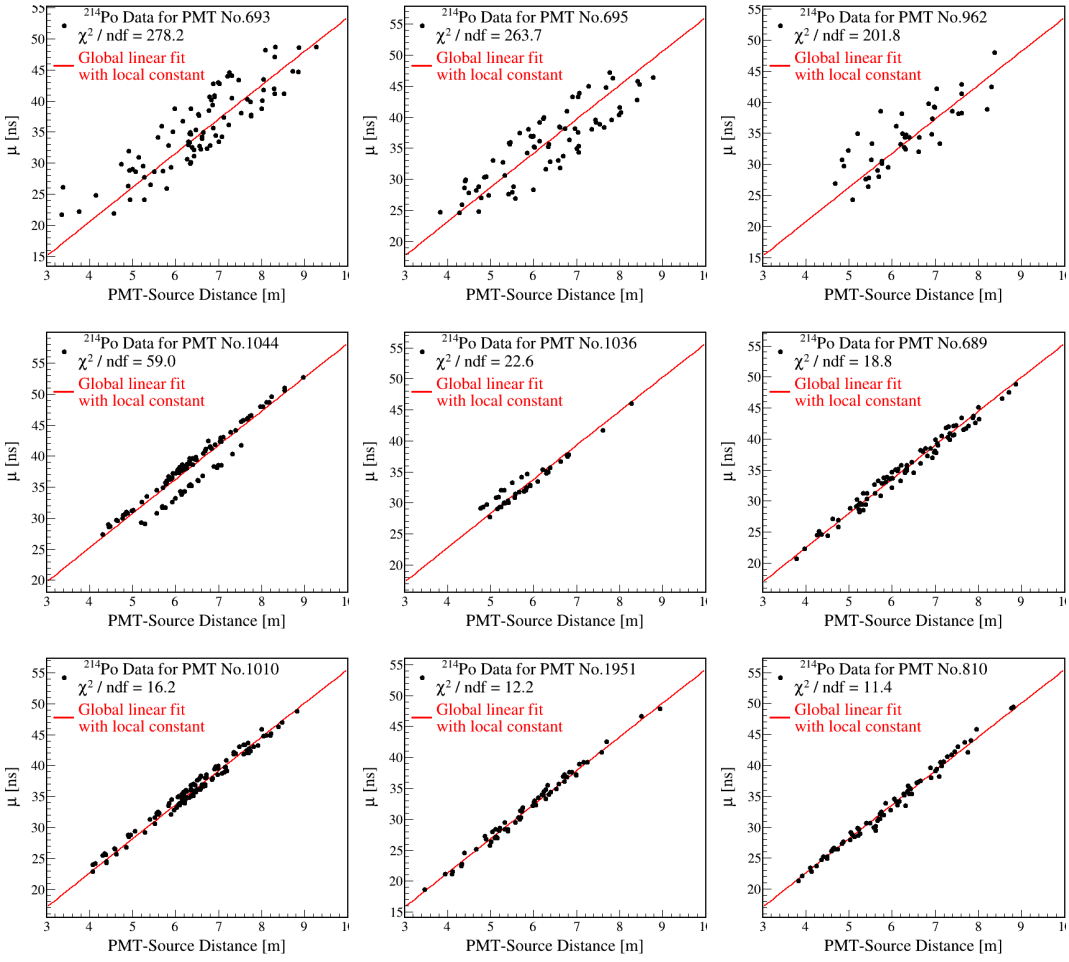


FIGURE A.1: The nine PMTs with the worst χ^2/ndf for a linear fit of the early hit mean times vs. distance, for the ^{214}Po source. The PMTs are sorted from worst to least bad PMT. They have been deselected for the calculation of the effective scintillation group velocity. The tenth deselected PMT is not shown.

Appendix B

Cherenkov Group Velocity Calibration using γ Events

PDF	$A[\mathrm{ns}^{-1}]$	$\mu[\mathrm{ns}]$	$\sigma[\mathrm{ns}]$
Data	-0.100	33.0	0.75
	-0.094	32.3	0.91
	-0.060	33.0	0.75
	-0.300	33.0	1.00
	-0.120	32.5	0.75
	-0.080	33.0	0.75
MC, $\mathrm{gv_{ch}} = 0.08\,\mathrm{ns\,m^{-1}}$	0.078	32.8	1.08
	0.070	32.2	1.30
	0.060	33.5	1.40
	0.070	32.3	1.35
	-0.090	34.7	1.30
	-0.142	33.1	1.16
MC, $\mathrm{gv_{ch}} = 0.10\,\mathrm{ns\,m^{-1}}$	0.065	32.3	1.45
	0.060	33.7	1.70
	0.060	32.2	1.45
	0.092	32.7	1.00
	-0.090	34.7	1.30
	-0.125	33.1	1.15
MC, $\mathrm{gv_{ch}} = 0.16\,\mathrm{ns\,m^{-1}}$	0.060	32.5	1.65
	0.095	32.4	1.35
	0.065	33.6	1.65
	0.094	32.8	1.15
	-0.075	34.3	0.80
	-0.110	32.9	1.15
MC, $\mathrm{gv_{ch}} = 0.22\,\mathrm{ns\,m^{-1}}$	0.080	32.7	1.65
	0.075	32.5	1.65
	0.085	33.1	1.40
	0.065	33.7	1.85
	-0.100	32.7	1.05
	-0.100	34.7	1.05

TABLE B.1: The parameters used for the modification of the position reconstruction PDF, necessary for the $\mathrm{gv_{ch}}$ calibration using the γ sources.

Appendix C

Correlated and Integrated Directionality Measurement of CNO-Neutrinos

C.1 Selection of the Nth-Hit cut-off and CID-histogram binning

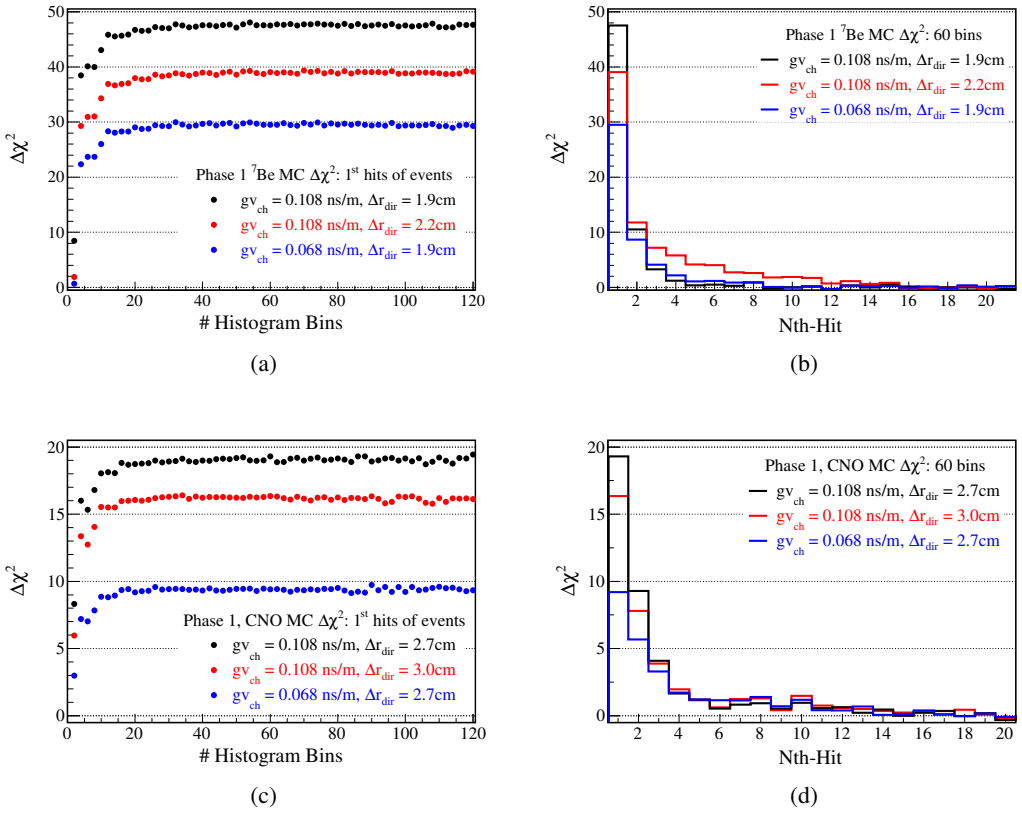
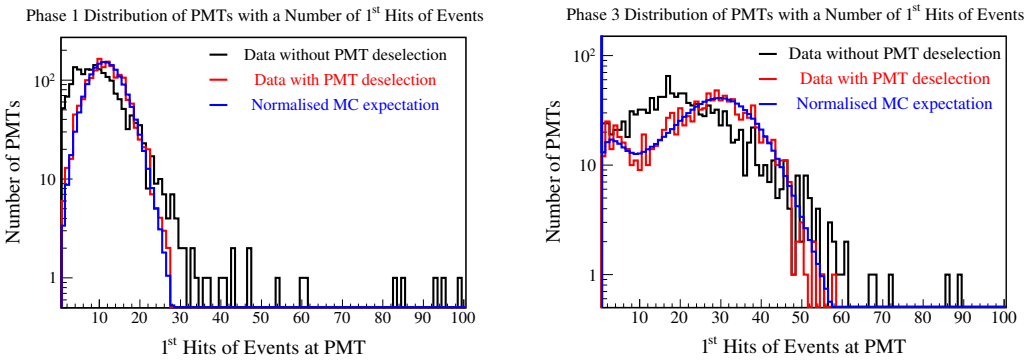


FIGURE C.1: The toy-MC study of the Cherenkov information $\Delta\chi^2$ as a function of the individual Nth-Hit $\cos\alpha$ histograms and the histogram binning, for Phase 1. The nominal MC Δr_{dir} value and $gv_{\text{ch}} = 0.108\,\text{ns}\,\text{m}^{-1}$ is shown in blue, an increased Δr_{dir} value is shown in red and the nominal MC Δr_{dir} with a decreased $gv_{\text{ch}} = 0.068\,\text{ns}\,\text{m}^{-1}$ is shown in blue. (a) $\Delta\chi^2$ vs. the Nth-Hits for the ^{7}Be-ν region. (b) $\Delta\chi^2$ vs. $\cos\alpha$ histogram binning for the ^{7}Be-ν region. (c) $\Delta\chi^2$ vs. the Nth-Hits for the CNO+pep-ν region. (d) $\Delta\chi^2$ vs. $\cos\alpha$ histogram binning for the CNO+pep-ν region.

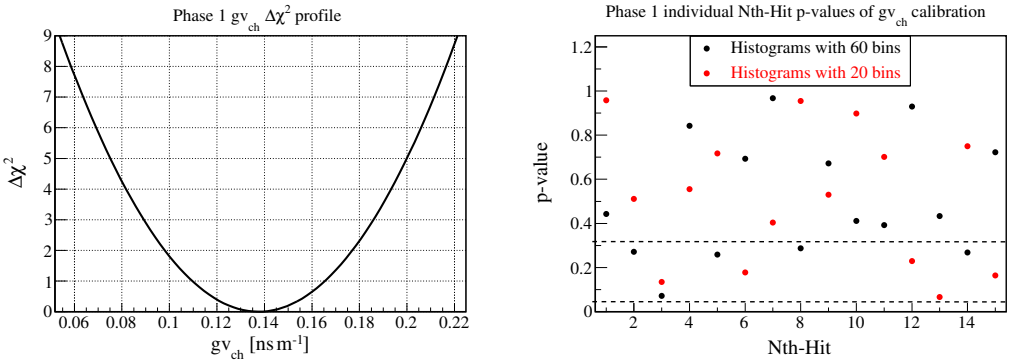
C.2 PMT selection and correction of the relative PMT time offset



(a) (b)

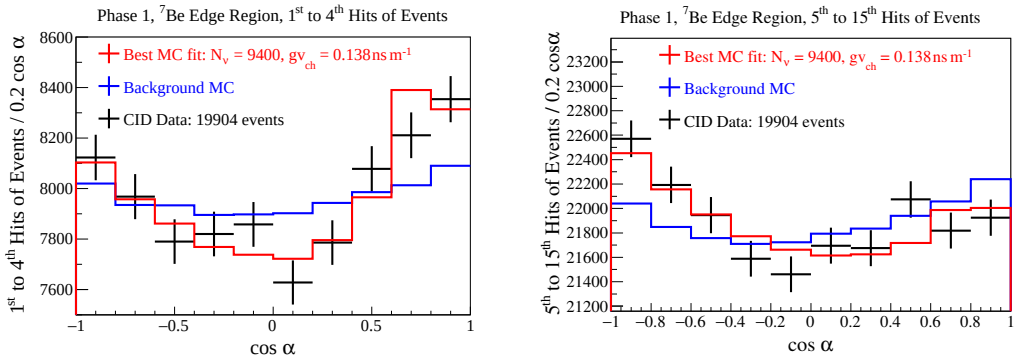
FIGURE C.2: Distribution of the number of first hits of the events on the PMTs for the ^{7}Be-ν region. The data without a PMT deselection (black) is compared with the data after the PMT deselection (red) and the expected distribution from MC (blue). (a) For Phase 1. (b) For Phase 3.

C.3 Measurement of $N_{\mathrm{CNO+pep}}$ without systematic uncertainties



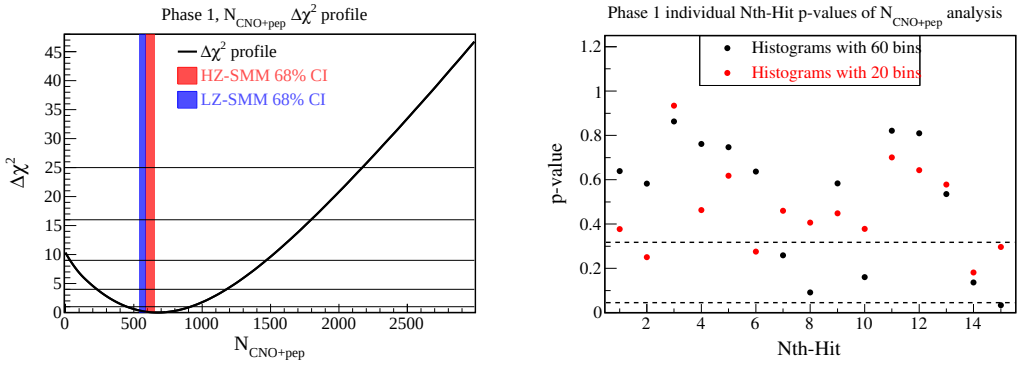
(a) (b)

FIGURE C.3: (a) Resulting $\Delta\chi^2(\mathrm{gv_{ch}})$ profile of Phase 1, with $\chi^2/\mathrm{ndf} = 868.3/897$, p-value = 0.75 for a binning of 60 and $\chi^2/\mathrm{ndf} = 277.2/297$, p-value = 0.79 for a reduced binning of 20 using the same best fit parameters. The best fit value is $\mathrm{gv_{ch}} = (0.138 \pm 0.028)\,\mathrm{ns\,m^{-1}}$. (b) The χ^2 corresponding p-values of each individual Nth-Hit $\cos\alpha$ histogram between data and MC. The MC histograms are not fitted individually on data, but use the best fit results of the full Nth-Hit(max) = 15 analysis. The p-values are calculated for a number of bins of 60 (black) and 20 (red). The dotted lines correspond to p-value < 0.32 andp-value < 0.05.



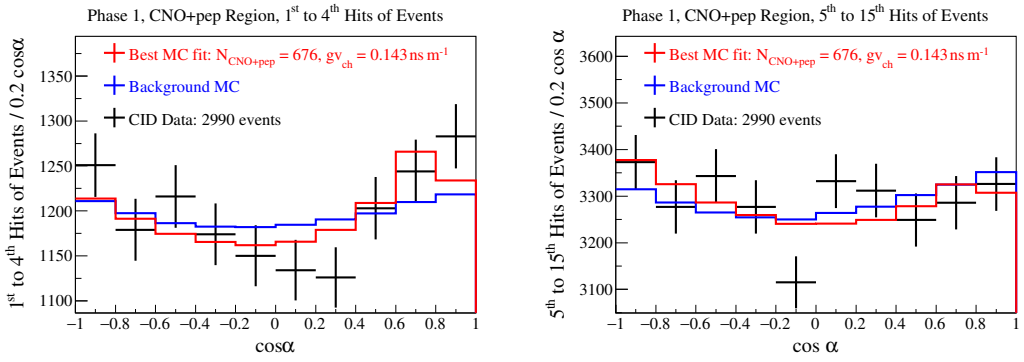
(a) (b)

FIGURE C.4: Summed $\cos\alpha$ histograms of the Phase 1 ^{7}Be region used for the gv_{ch} calibration. data is shown in black, pure ^{210}Bi background MC in blue and in red the best fit MC histogram with $gv_{ch} = 0.138\,\mathrm{ns\,m^{-1}}$ and $N_\nu = 9400$ fitted signal events, 10504 fitted background events out of 19904 events in total. (a) The sum of the first to fourth Nth-Hits. The peak of direct Cherenkov hits is visible around $\cos\alpha \sim 0.7$. (b) The sum of the fifth to 15$^{\text{th}}$ Nth-Hits. The negative $\cos\alpha$ slope of neutrino events due to $\Delta r_{\text{dir}} = 1.90\,\mathrm{cm}$ is visible in data. The $\cos\alpha$ shape of the MC background comes from the run dependent distribution of live PMTs and the PMT selection.



(a) (b)

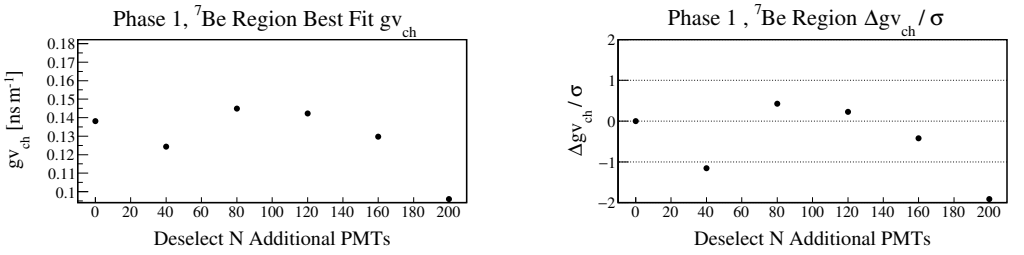
FIGURE C.5: (a) Resulting $\Delta\chi^2(N_{\text{CNO+pep}})$ profile of Phase 1, with $\chi^2/\text{ndf} = 886.9/897$, p-value = 0.59 for a binning of 60 and $\chi^2/\text{ndf} = 284.3/297$, p-value = 0.69 for a reduced binning of 20 using the same best fit parameters. The best fit value is $N_{\text{CNO+pep}} = 676^{+235}_{-224}$ out of 2990 total events. The red and blue bands represents the 68.3% CI of the SSM expectations of HZ and LZ, respectively. (b) The χ^2 corresponding p-values of each individual Nth-Hit $\cos\alpha$ histogram between data and MC. The MC histograms are not fitted individually on data, but use the best fit results of the full Nth-Hit(max) = 15 analysis. The p-values are calculated for a number of bins of 60 (black) and 20 (red). The dotted lines correspond to p-value < 0.32 and p-value < 0.05.



(a) (b)

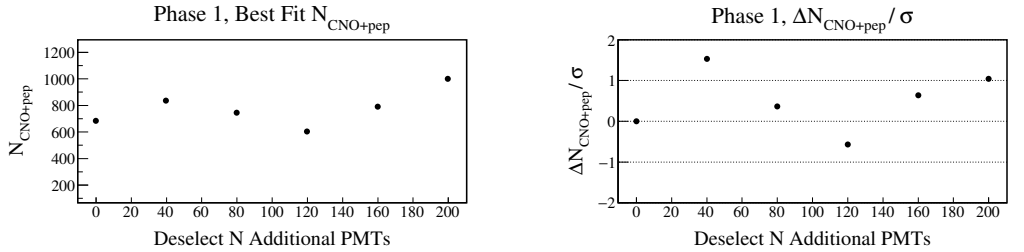
FIGURE C.6: (a) Summed $\cos\alpha$ histograms of the Phase 1 CNO+pep region. data is shown in black, pure MC background in blue and in red the best fit MC histogram with $N_\nu = 676$ fitted signal events, 2309 fitted background events out of 2990 events in total and $\mathrm{gv_{ch}} = 0.143\,\mathrm{ns\,m^{-1}}$. (a) The sum of the first to fourth Nth-Hits. The peak of direct Cherenkov hits is visible around $\cos\alpha \sim 0.7$. (b) The sum of the fifth to 15$^{\text{th}}$ Nth-Hits. The negative $\cos\alpha$ slope of neutrino events due to $\Delta r_{\mathrm{dir}} = 3.20\,\mathrm{cm}$ is visible in data. The $\cos\alpha$ shape of MC background comes from the run dependent distribution of live PMTs and the PMT selection.

C.4 Estimation of systematic uncertainties



(a) (b)

FIGURE C.7: (a) Results of the data $\mathrm{gv_{ch}}$ calibration at the ^{7}Be edge region of Phase 1 for the nominal and stricter selection of PMTs to use in the CID analysis. (b) Difference between the $\mathrm{gv_{ch}}$ results of the nominal PMT selection and the stricter PMT selections in units of the expected standard deviation, which is estimated through a toy-MC study.



(a) (b)

FIGURE C.8:]

(a) Results of the data $N_{\mathrm{CNO+pep}}$ measurement of Phase 1 for the nominal and stricter selections of PMTs to use in the CID analysis. (b) Difference between the $N_{\mathrm{CNO+pep}}$ results of the nominal PMT selection and the stricter PMT selections in units of the expected standard deviation, which is estimated through a toy-MC study.

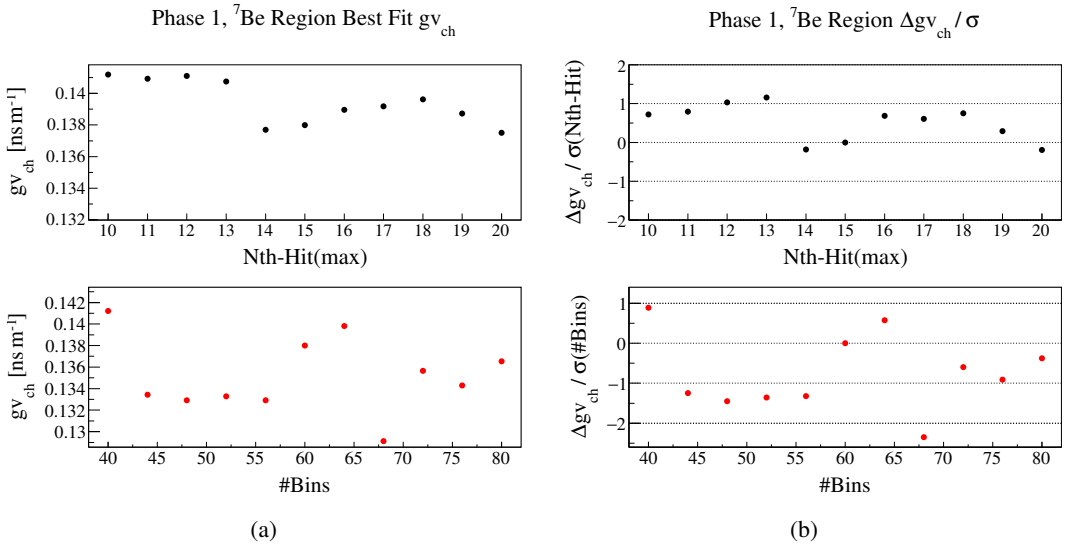


FIGURE C.9: (a) Results of data gv_{ch} calibration at the ^{7}Be-ν edge region, for Phase 1. Top (black): a range of Nth-Hit(max) values around the nominal Nth-Hit(max) = 15, with 60 bins. Bottom (red): a range binning around the nominal 60 bins, with Nth-Hit(max) = 15. (b) Difference between the gv_{ch} results and the nominal result $gv_{ch} = 0.138\,\mathrm{ns\,m^{-1}}$ in units of the expected statistical uncertainty.

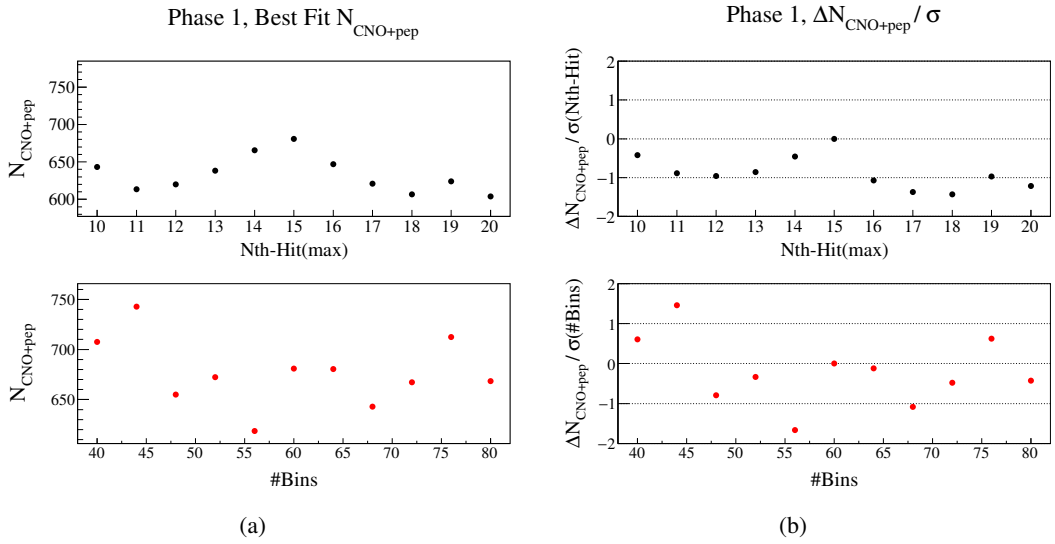
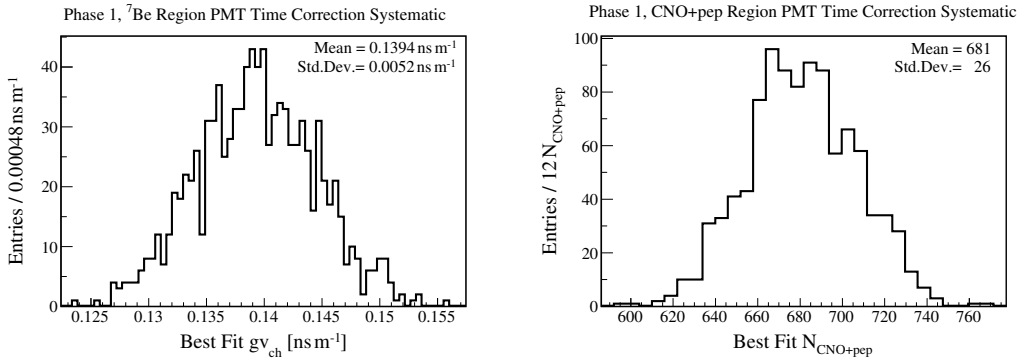
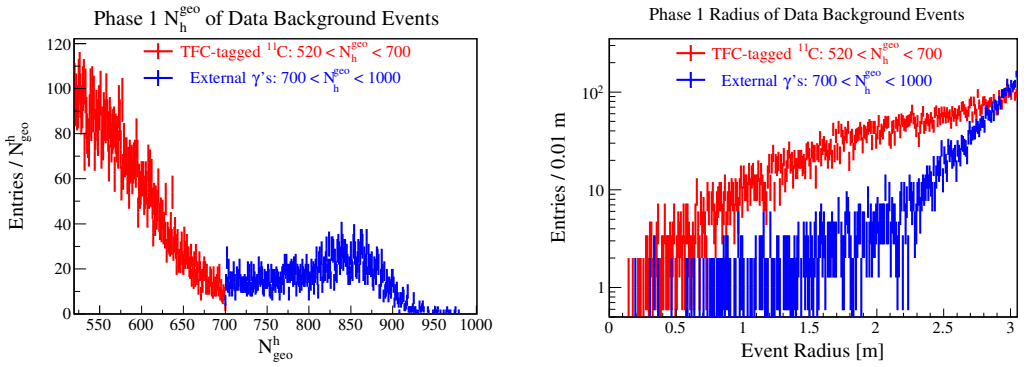


FIGURE C.10: (a) Results of data $N_{CNO+pep}$ measurement, for Phase 1. Top (black): a range of Nth-Hit(max) values around the nominal Nth-Hit(max) = 15, with 60 bins. Bottom (red): a range of number of bins around the nominal 60 bins, with Nth-Hit(max) = 15. (b) Difference between the $N_{CNO+pep}$ results and the nominal $N_{CNO+pep} = 676$ in units of the expected statistical uncertainty.



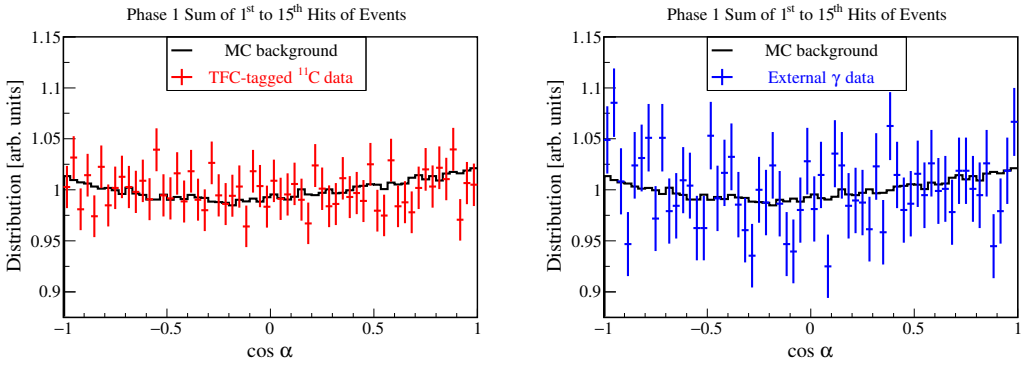
(a) (b)

FIGURE C.11: Distribution of Phase 1 best fit results, where the PMT hit time correction is randomly varied a number of 1000 times, using to the uncertainty $\Delta\mu(\text{PMT})$ as a Gaussian standard deviation. (a) For the gv_{ch} calibration at the ^{7}Be edge region. (b) For the measurement of $N_{\text{CNO+pep}}$.



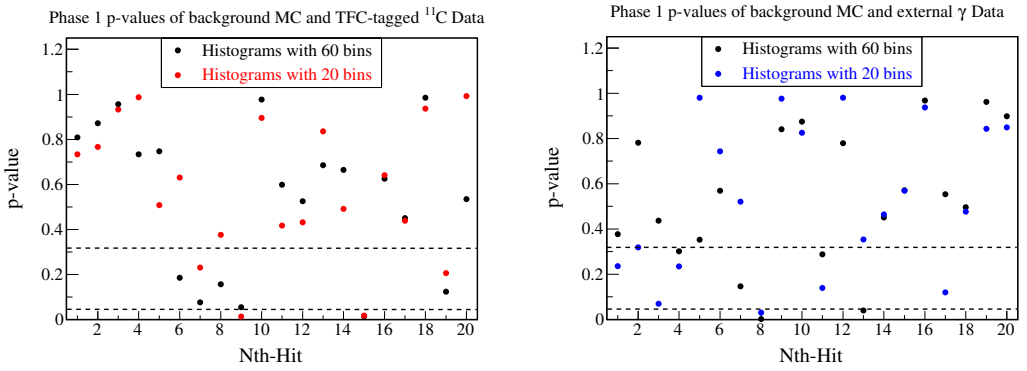
(a) (b)

FIGURE C.12: (a) The selected $\text{N}_{\text{h}}^{\text{geo}}$ region for the TFC-tagged ^{11}C (red) and external γ (blue) data events of Phase 1. The γ events are neither TFC-cut nor TFC-tagged. The events numbers are $N_{^{11}\text{C}} = 9318$ and $N_{\gamma} = 3840$. (b) The corresponding distribution of the reconstructed event radius, in a log-scale. ^{11}C follows a third order polynomial distribution, as expected from a uniform event distribution plotted against the radius. Likewise, the external γ's show the expected exponential increase for larger radii.



(a) (b)

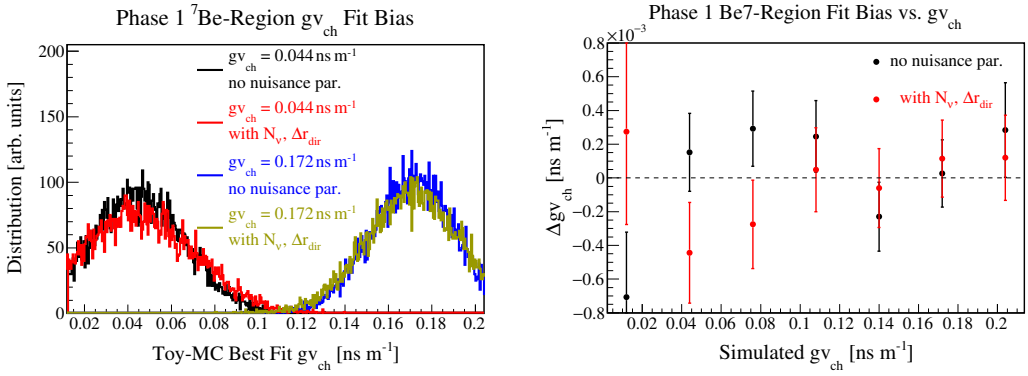
FIGURE C.13: Comparison of the $\cos\alpha$ histograms summed over the 1st to 15th hits of events, for data and MC background of Phase 1. The histograms are normalised to the same integral, as to allow for an easier comparison between the figures. (a) TFC-tagged ^{11}C data compared to MC background with a $\chi^2/\text{ndf} = 48.6/59$, p-value $= 0.83$. (b) External γ data compared to MC background with a $\chi^2/\text{ndf} = 62.8/59$, p-value $= 0.34$.



(a) (b)

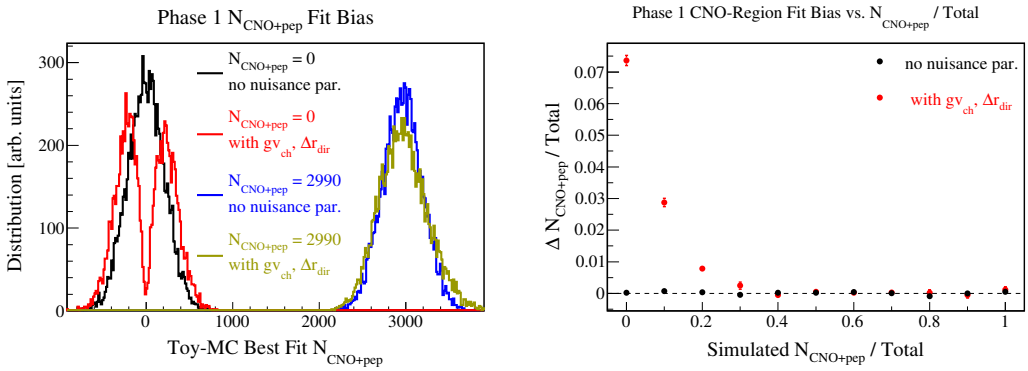
FIGURE C.14: The χ^2/ndf corresponding p-values between data and MC background of Phase 1, as a function of the individual Nth-Hit $\cos\alpha$ histograms. The dotted lines correspond to p-value $= 0.32$ and p-value $= 0.05$. (a) TFC-tagged ^{11}C data compared to MC background for a binning of 60 (black) and 20 (red). (b) External γ data compared to MC background for a binning of 60 (black) and 20 (blue).

C.5 Fit bias and Bayesian interpretation of the CID results



(a) (b)

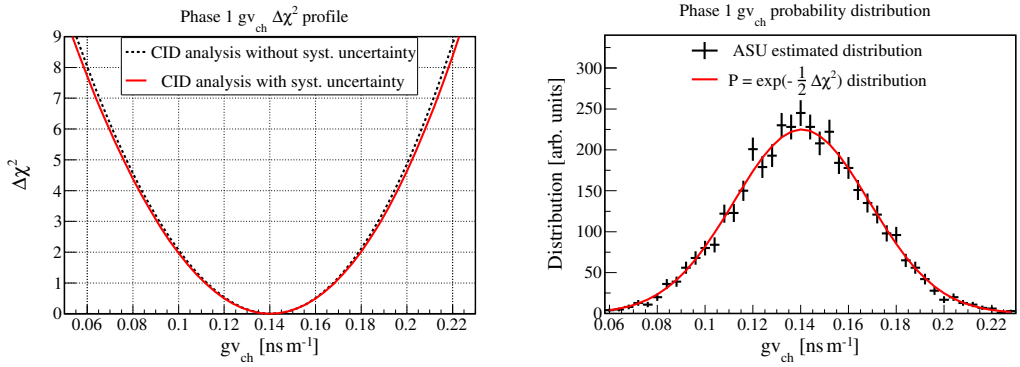
FIGURE C.15: (a) Distribution of the fit results for 10k toy-MC analyses, where the simulated gv_{ch} value is set to $gv_{\mathrm{ch}} = 0.044\,\mathrm{ns\,m^{-1}}$ (black, red) and $gv_{\mathrm{ch}} = 0.172\,\mathrm{ns\,m^{-1}}$ (blue, yellow). The analyses are performed including the two nuisance parameters of the fit (red, yellow) and and for comparison with fixed values of the nuisance parameters and only gv_{ch} as a free parameter (black, blue). (b) Difference between the mean value of the best fit distribution and the simulated gv_{ch}. The mean value is estimated with a Gaussian fit.



(a) (b)

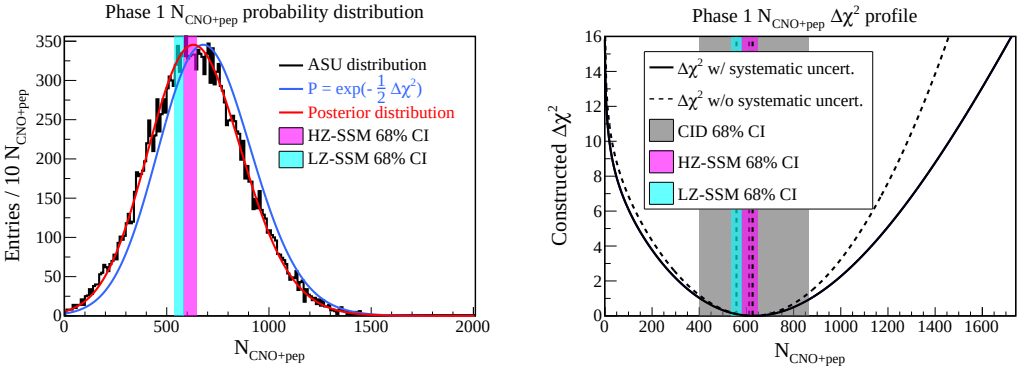
FIGURE C.16: (a) Distribution of the fit results for 10k toy-MC analyses, where the simulated $N_{\mathrm{CNO+pep}}$ value was set to $N_{\mathrm{CNO+pep}} = 0$ (black, red) and $N_{\mathrm{CNO+pep}} = 5974$ (blue, yellow). The analyses are performed including the two nuisance parameters of the fit (red, yellow) and for comparison only with $N_{\mathrm{CNO+pep}}$ as a free parameter (black, blue). The fit is not constrained to the physical limits ($0 \leq N_{\mathrm{CNO+pep}} \leq 5974$) to illustrate the impact of the nuisance parameters better gv_{ch}, Δr_{dir}. (b) Relative offset between the mean value of the best fit distribution and the simulated $N_{\mathrm{CNO+pep}}$. The mean value is estimated with a Gaussian fit performed within the physical limits of $N_{\mathrm{CNO+pep}}$.

C.6 Final results



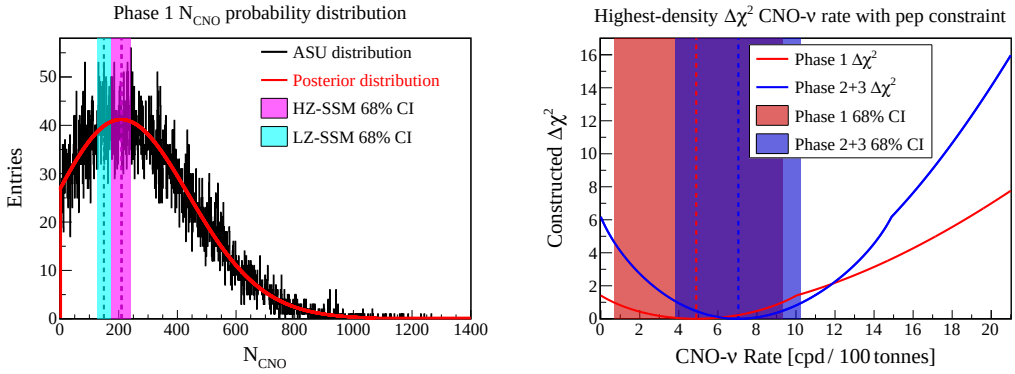
(a) (b)

FIGURE C.17: Final results of the $\mathrm{gv_{ch}}$ calibration for Phase 1. (a) Resulting $\Delta\chi^2(\mathrm{gv_{ch}})$ profile including the systematic uncertainties (red), with $\chi^2/\mathrm{ndf} = 874.9/897$, $\mathrm{p} = 0.70$. The best fit value is $\mathrm{gv_{ch}} = 0.140 \pm 0.029\,\mathrm{ns\,m^{-1}}$. (b) The ASU distribution of $\mathrm{gv_{ch}}$ (black) for 4000 toy-analyses compared to the likelihood distribution (red) of the $\Delta\chi^2$ profiles. The difference between the mean values is in agreement with zero: $\Delta\langle \mathrm{gv_{ch}}\rangle = (-0.0001 \pm 0.0004)\,\mathrm{ns\,m^{-1}}$.



(a) (b)

FIGURE C.18: Final results of the $N_{\mathrm{CNO+pep}}$ measurement for Phase 1. (a) The ASU distribution of $N_{\mathrm{CNO+pep}}$ with 20k toy-analyses in black, compared to the nominal likelihood distribution $\exp(-\frac{1}{2}\Delta\chi^2(N_{\mathrm{CNO+pep}}))$ in blue and the fit response bias corrected distribution in red. The latter is the posterior distribution $P(N_{\mathrm{CNO+pep}} \mid \text{CID data})$ for the probability of the true, underlying number of neutrino events. The 68.3% CI for the LZ-SSM and HZ-SSM is shown as a cyan and magenta band, respectively. They include the systematic uncertainties of FV, E_{eff}. (b) The equal-tail credible intervals of $N_{\mathrm{CNO+pep}}$, represented in a Gaussian equivalent $\Delta\chi^2$ profile (solid line). The corresponding 68.3% CI $N^{\mathrm{ET}}_{\mathrm{CNO+pep}} = 626^{+239}_{-226}$ is shown as a grey band. For comparison the $\Delta\chi^2$ profile without systematics and with fixed nuisance parameters is also shown (dotted line).



(a) (b)

FIGURE C.19: Final results of the N_{CNO} measurement for Phase 1. (a) The ASU distribution of N_{CNO} with 20k toy-analyses in black and the corresponding posterior distribution $P(N_{\mathrm{CNO}} \mid \text{CID data})$. The 68.3% CI for the LZ-SSM and HZ-SSM is shown as a cyan and magenta band, respectively. (b) The highest-density CI of the CNO-ν rate R_{CNO}, represented in a Gaussian equivalent $\Delta\chi^2$ profile. For Phase 1 (red) and Phase 2+3 (blue), with the corresponding 68.3% CI $R_{\mathrm{CNO}}(\text{Phase 1}) = 4.9^{+4.5}_{-4.3}\,(\text{cpd}/100\,\text{t})$, $R_{\mathrm{CNO}}(\text{Phase 2+3}) = 7.0^{+2.8}_{-2.7}\,(\text{cpd}/100\,\text{t})$.

Bibliography

[1] R. Davis, D. S. Harmer, and K. C. Hoffman. "Search for Neutrinos from the Sun". In: *Physical Review Letters* 20 (1968). DOI: 10.1103/PhysRevLett.20.1205. URL: https://link.aps.org/doi/10.1103/PhysRevLett.20.1205.

[2] B. T. Cleveland et al. "Measurement of the Solar Electron Neutrino Flux with the Homestake Chlorine Detector". In: *The Astrophysical Journal* 496.1 (1998). DOI: 10.1086/305343. URL: https://dx.doi.org/10.1086/305343.

[3] Nobel Foundation. https://www.nobelprize.org/prizes/physics/2002/press-release/.

[4] Super-Kamiokande Collaboration. "Evidence for Oscillation of Atmospheric Neutrinos". In: *Physical Review Letters* 81 (1998). DOI: 10.1103/PhysRevLett.81.1562. URL: https://link.aps.org/doi/10.1103/PhysRevLett.81.1562.

[5] SNO Collaboration. "Direct Evidence for Neutrino Flavor Transformation from Neutral-Current Interactions in the Sudbury Neutrino Observatory". In: *Physical Review Letters* 89 (2002). DOI: 10.1103/PhysRevLett.89.011301. URL: https://link.aps.org/doi/10.1103/PhysRevLett.89.011301.

[6] Nobel Foundation. https://www.nobelprize.org/prizes/physics/2015/press-release/.

[7] K. Zuber. *Neutrino Physics*. 3rd Edition. CRC Press, 2020. ISBN: 9781315195612. DOI: https://doi.org/10.1201/9781315195612.

[8] W. Pauli. *Collected Scientific Papers*. Vol. 2. Interscience, 1964.

[9] E. Fermi. "Versuch einer Theorie der β-Strahlen. I". In: *Zeitschrift für Physik* 88.3 (1934). ISSN: 0044-3328. DOI: 10.1007/BF01351864. URL: https://doi.org/10.1007/BF01351864.

[10] KATRIN Collaboration. "Direct neutrino-mass measurement with sub-electronvolt sensitivity". In: *Nature Physics* 18.2 (2022). ISSN: 1745-2481. DOI: 10.1038/s41567-021-01463-1. URL: https://doi.org/10.1038/s41567-021-01463-1.

[11] C. L. Cowan et al. "Detection of the free neutrino: A Confirmation". In: *Science* 124 (1956). DOI: 10.1126/science.124.3212.103.

[12] Nobel Foundation. https://www.nobelprize.org/prizes/physics/1995/press-release/.

[13] C. Giunti and C. W. Kim. *Fundamentals of Neutrino Physics and Astrophysics*. Oxford University Press, 2007. ISBN: 978-0-19-850871-7.

[14] R. L. Workman et al. "Review of Particle Physics". In: *Progress of Theoretical and Experimental Physics* 2022 (2022). DOI: 10.1093/ptep/ptac097.

[15] ATLAS Collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Physics Letters B* 716.1 (2012). ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2012.08.020. URL: https://www.sciencedirect.com/science/article/pii/S037026931200857X.

[16] CMS Collaboration. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Physics Letters B* 716.1 (2012). ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2012.08.021. URL: https://www.sciencedirect.com/science/article/pii/S0370269312008581.

[17] Nobel Foundation. https://www.nobelprize.org/prizes/physics/2013/press-release/.

[18] Project 8 Collaboration. *The Project 8 Neutrino Mass Experiment*. 2022. DOI: 10.48550/ARXIV.2203.07349. URL: https://arxiv.org/abs/2203.07349.

[19] Planck Collaboration. "Planck 2018 results - VI. Cosmological parameters". In: *Astronomy & Astrophysics* 641 (2020). DOI: 10.1051/0004-6361/201833910. URL: https://doi.org/10.1051/0004-6361/201833910.

[20] A. de Gouvêa. "Neutrino Mass Models". In: *Annual Review of Nuclear and Particle Science* 66.1 (2016). DOI: 10.1146/annurev-nucl-102115-044600. URL: https://doi.org/10.1146/annurev-nucl-102115-044600.

[21] STEREO Collaboration. "STEREO neutrino spectrum of ^{235}U fission rejects sterile neutrino hypothesis". In: *Nature* 613.7943 (2023). ISSN: 1476-4687. DOI: 10.1038/s41586-022-05568-2. URL: https://doi.org/10.1038/s41586-022-05568-2.

[22] MicroBooNE Collaboration. "First Constraints on Light Sterile Neutrino Oscillations from Combined Appearance and Disappearance Searches with the MicroBooNE Detector". In: *Physical Review Letters* 130 (2023). DOI: 10.1103/PhysRevLett.130.011801. URL: https://link.aps.org/doi/10.1103/PhysRevLett.130.011801.

[23] M. J. Dolinski, A. W.P. Poon, and W. Rodejohann. "Neutrinoless Double-Beta Decay: Status and Prospects". In: *Annual Review of Nuclear and Particle Science* 69.1 (2019). DOI: 10.1146/annurev-nucl-101918-023407. URL: https://doi.org/10.1146/annurev-nucl-101918-023407.

[24] Super-Kamiokande Collaboration. "Solar ^{8}B and hep Neutrino Measurements from 1258 Days of Super-Kamiokande Data". In: *Physical Review Letters* 86 (2001). DOI: 10.1103/PhysRevLett.86.5651. URL: https://link.aps.org/doi/10.1103/PhysRevLett.86.5651.

[25] M.P. Decowski. "KamLAND's precision neutrino oscillation measurements". In: *Nuclear Physics B* 908 (2016). Neutrino Oscillations: Celebrating the Nobel Prize in Physics 2015. ISSN: 0550-3213. DOI: https://doi.org/10.1016/j.nuclphysb.2016.04.014. URL: https://www.sciencedirect.com/science/article/pii/S0550321316300529.

[26] F. Capozzi et al. "Current unknowns in the three-neutrino framework". In: *Progress in Particle and Nuclear Physics* 102 (2018). ISSN: 0146-6410. DOI: https://doi.org/10.1016/j.ppnp.2018.05.005. URL: https://www.sciencedirect.com/science/article/pii/S0146641018300383.

[27] B. Pontecorvo. "Inverse Beta Processes and Nonconservation of Lepton Charge". In: *Journal of Experimental and Theoretical Physics (U.S.S.R.)* 34 (1958). eprint: http://jetp.ras.ru/files/pontecorvo1958_en.pdf.

[28] Z. Maki, M. Nakagawa, and S. Sakata. "Remarks on the Unified Model of Elementary Particles". In: *Progress of Theoretical Physics* 28.5 (1962). ISSN: 0033-068X. DOI: 10.1143/PTP.28.870. URL: https://doi.org/10.1143/PTP.28.870.

[29] L. Wolfenstein. "Neutrino oscillations in matter". In: *Physical Review D* 17 (1978). DOI: 10.1103/PhysRevD.17.2369. URL: https://link.aps.org/doi/10.1103/PhysRevD.17.2369.

[30] S. P. Mikheyev and A. Yu. Smirnov. "Resonance Amplification of Oscillations in Matter and Spectroscopy of Solar Neutrinos". In: *Soviet Journal of Nuclear Physics* 42 (1985).

[31] Borexino Collaboration. "Comprehensive measurement of pp-chain solar neutrinos". In: *Nature* 562.7728 (2018). ISSN: 1476-4687. DOI: 10.1038/s41586-018-0624-y. URL: https://doi.org/10.1038/s41586-018-0624-y.

[32] P.C. de Holanda, W. Liao, and A.Yu. Smirnov. "Toward precision measurements in solar neutrinos". In: *Nuclear Physics B* 702.1 (2004). ISSN: 0550-3213. DOI: https://doi.org/10.1016/j.nuclphysb.2004.09.027. URL: https://www.sciencedirect.com/science/article/pii/S0550321304007333.

[33] J. N. Bahcall. *Neutrino Astrophysics*. Cambridge University Press, 1989. ISBN: 0-521-35113-8.

[34] A. Serenelli. "Alive and well: A short review about standard solar models". In: *The European Physical Journal A* 52.4 (2016). ISSN: 1434-601X. DOI: 10.1140/epja/i2016-16078-1. URL: https://doi.org/10.1140/epja/i2016-16078-1.

[35] N. Vinyoles et al. "A New Generation of Standard Solar Models". In: *The Astrophysical Journal* 835.2 (2017). DOI: 10.3847/1538-4357/835/2/202. URL: https://doi.org/10.3847/1538-4357/835/2/202.

[36] H. A. Bethe and C. L. Critchfield. "The Formation of Deuterons by Proton Combination". In: *Physical Review* 54 (1938). DOI: 10.1103/PhysRev.54.248. URL: https://link.aps.org/doi/10.1103/PhysRev.54.248.

[37] C.F. von Weizsäcker. "Über Elementumwandlungen in Innern der Sterne I". In: *Physikalische Zeitschrift* 38 (1937).

[38] C.F. von Weizsäcker. "Über Elementumwandlungen in Innern der Sterne II". In: *Physikalische Zeitschrift* 39 (1939).

[39] H. A. Bethe. "Energy Production in Stars". In: *Physical Review* 55 (1939). DOI: 10.1103/PhysRev.55.434. URL: https://link.aps.org/doi/10.1103/PhysRev.55.434.

[40] E. G. Adelberger et al. "Solar fusion cross sections. II. The *pp* chain and CNO cycles". In: *Reviews of Modern Physics* 83 (2011). DOI: 10.1103/RevModPhys.83.195. URL: https://link.aps.org/doi/10.1103/RevModPhys.83.195.

[41] *J.N. Bahcall's webpage with the energy spectra of solar neutrinos*. https://www.sns.ias.edu/~jnb/SNdata/Export/. Accessed: 2022-10-04.

[42] M. Asplund et al. "The Chemical Composition of the Sun". In: *Annual Review of Astronomy and Astrophysics* 47.1 (2009). DOI: 10.1146/annurev.astro.46.060407.145222. URL: https://doi.org/10.1146/annurev.astro.46.060407.145222.

[43] N. Grevesse and A. J. Sauval. "Standard Solar Composition". In: *Space Science Reviews* 85.1 (1998). ISSN: 1572-9672. DOI: 10.1023/A:1005161325181. URL: https://doi.org/10.1023/A:1005161325181.

[44] J. Christensen-Dalsgaard. "Helioseismology". In: *Reviews of Modern Physics* 74 (2002). DOI: 10.1103/RevModPhys.74.1073. URL: https://link.aps.org/doi/10.1103/RevModPhys.74.1073.

[45] Borexino Collaboration. "Improved Measurement of Solar Neutrinos from the Carbon-Nitrogen-Oxygen Cycle by Borexino and Its Implications for the Standard Solar Model". In: *Physical Review Letters* 129 (2022). DOI: 10.1103/PhysRevLett.129.252701. URL: https://link.aps.org/doi/10.1103/PhysRevLett.129.252701.

[46] L. Baudis et al. "Neutrino physics with multi-ton scale liquid xenon detectors". In: *Journal of Cosmology and Astroparticle Physics* 2014.01 (2014). DOI: 10.1088/1475-7516/2014/01/044. URL: https://dx.doi.org/10.1088/1475-7516/2014/01/044.

[47] D. Franco et al. "Solar neutrino detection in a large volume double-phase liquid argon experiment". In: *Journal of Cosmology and Astroparticle Physics* 2016.08 (2016). DOI: 10.1088/1475-7516/2016/08/017. URL: https://dx.doi.org/10.1088/1475-7516/2016/08/017.

[48] G. D. Orebi Gann et al. "The Future of Solar Neutrinos". In: *Annual Review of Nuclear and Particle Science* 71.1 (2021). DOI: 10.1146/annurev-nucl-011921-061243. URL: https://doi.org/10.1146/annurev-nucl-011921-061243.

[49] SAGE Collaboration. "Measurement of the solar neutrino capture rate with gallium metal. III. Results for the 2002–2007 data-taking period". In: *Physical Review C* 80 (2009). DOI: 10.1103/PhysRevC.80.015807. URL: https://link.aps.org/doi/10.1103/PhysRevC.80.015807.

[50] V. V. Barinov et al. "Results from the Baksan Experiment on Sterile Transitions (BEST)". In: *Physical Review Letters* 128 (2022). DOI: 10.1103/PhysRevLett.128.232501. URL: https://link.aps.org/doi/10.1103/PhysRevLett.128.232501.

[51] GALLEX Collaboration. "GALLEX solar neutrino observations: results for GALLEX IV". In: *Physics Letters B* 447.1 (1999). ISSN: 0370-2693. DOI: https://doi.org/10.1016/S0370-2693(98)01579-2. URL: https://www.sciencedirect.com/science/article/pii/S0370269398015792.

[52] M. K. Pavićević et al. "Lorandite from Allchar as geochemical detector for pp-solar neutrinos". In: *Nuclear Instruments and Methods in Physics Research Section A* 895 (2018). ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2018.03.039. URL: https://www.sciencedirect.com/science/article/pii/S0168900218303899.

[53] A. Bellerive et al. "The Sudbury Neutrino Observatory". In: *Nuclear Physics B* 908 (2016). Neutrino Oscillations: Celebrating the Nobel Prize in Physics 2015. ISSN: 0550-3213. DOI: https://doi.org/10.1016/j.nuclphysb.2016.04.035. URL: https://www.sciencedirect.com/science/article/pii/S0550321316300736.

[54] SNO+ collaboration. "The SNO+ experiment". In: *Journal of Instrumentation* 16.08 (2021). DOI: 10.1088/1748-0221/16/08/P08059. URL: https://dx.doi.org/10.1088/1748-0221/16/08/P08059.

[55] Super-Kamiokande Collaboration. "The Super-Kamiokande detector". In: *Nuclear Instruments and Methods in Physics Research Section A* 501.2 (2003). ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(03)00425-X. URL: https://www.sciencedirect.com/science/article/pii/S016890020300425X.

[56] K. S. Hirata et al. "Real-time, directional measurement of ^{8}B solar neutrinos in the Kamiokande II detector". In: *Physical Review D* 44 (1991). DOI: 10.1103/PhysRevD.44.2241. URL: https://link.aps.org/doi/10.1103/PhysRevD.44.2241.

[57] K. S. Hirata et al. "Observation of a neutrino burst from the supernova SN1987A". In: *Physical Review Letters* 58 (1987). DOI: 10.1103/PhysRevLett.58.1490. URL: https://link.aps.org/doi/10.1103/PhysRevLett.58.1490.

[58] Super-Kamiokande Collaboration. "Solar neutrino measurements in Super-Kamiokande IV". In: *Physical Review D* 94 (2016). DOI: 10.1103/PhysRevD.94.052010. URL: https://link.aps.org/doi/10.1103/PhysRevD.94.052010.

[59] SNO Collaboration. "Low-energy-threshold analysis of the Phase I and Phase II data sets of the Sudbury Neutrino Observatory". In: *Physical Review C* 81 (2010). DOI: 10.1103/PhysRevC.81.055504. URL: https://link.aps.org/doi/10.1103/PhysRevC.81.055504.

[60] K2K Collaboration. "Measurement of neutrino oscillation by the K2K experiment". In: *Physical Review D* 74 (2006). DOI: 10.1103/PhysRevD.74.072003. URL: https://link.aps.org/doi/10.1103/PhysRevD.74.072003.

[61] T2K Collaboration. "Constraint on the matter–antimatter symmetry-violating phase in neutrino oscillations". In: *Nature* 580.7803 (2020). ISSN: 1476-4687. DOI: 10.1038/s41586-020-2177-0. URL: https://doi.org/10.1038/s41586-020-2177-0.

[62] Hyper-Kamiokande Proto-Collaboration. *Hyper-Kamiokande Design Report*. 2018. DOI: 10.48550/ARXIV.1805.04163. URL: https://arxiv.org/abs/1805.04163.

[63] T. Tashiro. "Hyper-Kamiokande experiment". In: *The European Physical Society Conference on High Energy Physics* 398 (2022). DOI: 10.22323/1.398.0234. URL: https://pos.sissa.it/398/234.

[64] KamLAND Collaboration. "Experimental investigation of geologically produced antineutrinos with KamLAND". In: *Nature* 436.7050 (2005). ISSN: 1476-4687. DOI: 10.1038/nature03980. URL: https://doi.org/10.1038/nature03980.

[65] KamLAND Collaboration. "^{7}Be solar neutrino measurement with KamLAND". In: *Physical Review C* 92 (2015). DOI: 10.1103/PhysRevC.92.055808. URL: https://link.aps.org/doi/10.1103/PhysRevC.92.055808.

[66] KamLAND Collaboration. "Measurement of the ^{8}B solar neutrino flux with the KamLAND liquid scintillator detector". In: *Physical Review C* 84 (2011). DOI: 10.1103/PhysRevC.84.035804. URL: https://link.aps.org/doi/10.1103/PhysRevC.84.035804.

[67] KamLAND-Zen Collaboration. *First Search for the Majorana Nature of Neutrinos in the Inverted Mass Ordering Region with KamLAND-Zen*. 2022. DOI: 10.48550/ARXIV.2203.02139. URL: https://arxiv.org/abs/2203.02139.

[68] Borexino Collaboration. "Experimental evidence of neutrinos produced in the CNO fusion cycle in the Sun". In: *Nature* 587 (2020). DOI: 10.1038/s41586-020-2934-0. URL: https://doi.org/10.1038/s41586-020-2934-0.

[69] JUNO Collaboration. "JUNO physics and detector". In: *Progress in Particle and Nuclear Physics* 123 (2022). ISSN: 0146-6410. DOI: https://doi.org/10.1016/j.ppnp.2021.103927. URL: https://www.sciencedirect.com/science/article/pii/S0146641021000880.

[70] L. Bieger et al. "Potential for a precision measurement of solar *pp* neutrinos in the Serappis Experiment". In: *The European Physical Journal C* 82 (2022). DOI: 10.1140/epjc/s10052-022-10725-y. URL: https://doi.org/10.1140/epjc/s10052-022-10725-y.

[71] JUNO Collaboration. "The design and sensitivity of JUNO's scintillator radiopurity pre-detector OSIRIS". In: *The European Physical Journal C* 81.11 (2021). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-021-09544-4. URL: https://doi.org/10.1140/epjc/s10052-021-09544-4.

[72] M. Askins et al. "Theia: an advanced optical neutrino detector". In: *The European Physical Journal C* 80.5 (2020). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-020-7977-8. URL: https://doi.org/10.1140/epjc/s10052-020-7977-8.

[73] M. Yeh et al. "A new water-based liquid scintillator and potential applications". In: *Nuclear Instruments and Methods A* 660.1 (2011). ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2011.08.040. URL: https://www.sciencedirect.com/science/article/pii/S0168900211016615.

[74] J. F. Beacom et al. "Physics prospects of the Jinping neutrino experiment". In: *Chinese Physics C* 41.2 (2017). DOI: 10.1088/1674-1137/41/2/023002. URL: https://dx.doi.org/10.1088/1674-1137/41/2/023002.

[75] Y. Wu et al. *Performance of the 1-ton Prototype Neutrino Detector at CJPL-I*. 2022. DOI: 10.48550/ARXIV.2212.13158. URL: https://arxiv.org/abs/2212.13158.

[76] M. Kunitomo and T. Guillot. "Imprint of planet formation in the deep interior of the Sun". In: *Astronomy & Astrophysics* 655 (2021). DOI: 10.1051/0004-6361/202141256. URL: https://doi.org/10.1051/0004-6361/202141256.

[77] T. Kovacs et al. "Borex: Solar neutrino experiment via weak neutral and charged currents in boron-11". In: *Solar Physics* 128.1 (1990). ISSN: 1573-093X. DOI: 10.1007/BF00154147. URL: https://doi.org/10.1007/BF00154147.

[78] J. Benziger et al. "A scintillator purification system for the Borexino solar neutrino detector". In: *Nuclear Instruments and Methods in Physics Research Section A* 587.2 (2008). ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2007.12.043. URL: https://www.sciencedirect.com/science/article/pii/S016890020800003X.

[79] J. Benziger et al. "The fluid-filling system for the Borexino solar neutrino detector". In: *Nuclear Instruments and Methods in Physics Research Section A* 608.3 (2009). ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2009.07.035. URL: https://www.sciencedirect.com/science/article/pii/S0168900209014636.

[80] Borexino Collaboration. "The liquid handling systems for the Borexino solar neutrino detector". In: *Nuclear Instruments and Methods in Physics Research Section A* 609.1 (2009). ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2009.07.028. URL: https://www.sciencedirect.com/science/article/pii/S0168900209014247.

[81] J. B. Benziger and F. P. Calaprice. "Large-scale liquid scintillation detectors for solar neutrinos". In: *The European Physical Journal A* 52.4 (2016). ISSN: 1434-601X. DOI: 10.1140/epja/i2016-16081-6. URL: https://doi.org/10.1140/epja/i2016-16081-6.

[82] G. Alimonti et al. "A large-scale low-background liquid scintillation detector: the counting test facility at Gran Sasso". In: *Nuclear Instruments and Methods in Physics Research Section A* 406.3 (1998). ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(98)00018-7. URL: https://www.sciencedirect.com/science/article/pii/S0168900298000187.

[83] Borexino Collaboration. "Final results of Borexino Phase-I on low-energy solar neutrino spectroscopy". In: *Physical Review D* 89 (2014). DOI: 10.1103/PhysRevD.89.112007. URL: https://link.aps.org/doi/10.1103/PhysRevD.89.112007.

[84] Borexino Collaboration. "Simultaneous precision spectroscopy of *pp*, ^{7}Be, and *pep* solar neutrinos with Borexino Phase-II". In: *Physical Review D* 100 (2019). DOI: 10.1103/PhysRevD.100.082004. URL: https://link.aps.org/doi/10.1103/PhysRevD.100.082004.

[85] Borexino Collaboration. "Borexino calibrations: hardware, methods, and results". In: *Journal of Instrumentation* 7.10 (2012). DOI: 10.1088/1748-0221/7/10/p10018. URL: https://doi.org/10.1088/1748-0221/7/10/p10018.

[86] Borexino Collaboration. "Neutrinos from the primary proton–proton fusion process in the Sun". In: *Nature* 512.7515 (2014). ISSN: 1476-4687. DOI: 10.1038/nature13702. URL: https://doi.org/10.1038/nature13702.

[87] Borexino Collaboration. "First Evidence of *pep* Solar Neutrinos by Direct Detection in Borexino". In: *Physical Review Letters* 108 (2012). DOI: 10.1103/PhysRevLett.108.051302. URL: https://link.aps.org/doi/10.1103/PhysRevLett.108.051302.

[88] Borexino Collaboration. "Improved measurement of ^{8}B solar neutrinos with $1.5 \cdot \mathrm{kt} \cdot \mathrm{y}$ of Borexino exposure". In: *Physical Review D* 101 (2020). DOI: 10.1103/PhysRevD.101.062001. URL: https://link.aps.org/doi/10.1103/PhysRevD.101.062001.

[89] S. Kumaran et al. "The Low Polonium Field of Borexino and its significance for the CNO neutrino detection". In: *arXiv* (2021). DOI: 10.48550/ARXIV.2105.13209. URL: https://arxiv.org/abs/2105.13209.

[90] Borexino Collaboration. "Comprehensive geoneutrino analysis with Borexino". In: *Physical Review D* 101 (2020). DOI: 10.1103/PhysRevD.101.012009. URL: https://link.aps.org/doi/10.1103/PhysRevD.101.012009.

[91] Borexino Collaboration. "Limiting neutrino magnetic moments with Borexino Phase-II solar neutrino data". In: *Physical Review D* 96 (2017). DOI: 10.1103/PhysRevD.96.091103. URL: https://link.aps.org/doi/10.1103/PhysRevD.96.091103.

[92] Borexino Collaboration. "The Borexino detector at the Laboratori Nazionali del Gran Sasso". In: *Nuclear Instruments and Methods in Physics Research Section A* 600.3 (2009). ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2008.11.076. URL: https://www.sciencedirect.com/science/article/pii/S016890020801601X.

[93] D. D'Angelo. "Towards the detection of low energy solar neutrinos in BOREXino: data readout, data reconstruction and background identification". PhD thesis. Technische Universität München, 2006.

[94] Borexino Collaboration. "Muon and cosmogenic neutron detection in Borexino". In: *Journal of Instrumentation* 6.05 (2011). DOI: 10.1088/1748-0221/6/05/P05005. URL: https://dx.doi.org/10.1088/1748-0221/6/05/P05005.

[95] J. Benziger et al. "The nylon scintillator containment vessels for the Borexino solar neutrino experiment". In: *Nuclear Instruments and Methods in Physics Research Section A* 582.2 (2007). ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2007.08.176. URL: https://www.sciencedirect.com/science/article/pii/S0168900207018712.

[96] A. Ianni et al. "The measurements of 2200 ETL9351 type photomultipliers for the Borexino experiment with the photomultiplier testing facility at LNGS". In: *Nuclear Instruments and Methods in Physics Research Section A* 537.3 (2005). ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2004.07.249. URL: https://www.sciencedirect.com/science/article/pii/S0168900204018078.

[97] B. Caccianiga et al. "A multiplexed optical-fiber system for the PMT calibration of the Borexino experiment". In: *Nuclear Instruments and Methods in Physics Research Section A* 496.2 (2003). ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(02)01762-X. URL: https://www.sciencedirect.com/science/article/pii/S016890020201762X.

[98] M. Redchuk. "Looking inside the Sun with the Borexino experiment: detection of solar neutrinos from the proton-proton chain and the CNO cycle". PhD thesis. RWTH Aachen, 2020.

[99] A. C. Re. "Low energy solar-ν: expected values for flux, survival probability, and interaction rate in Borexino". In: *Internal Borexino Document Database* 944 (2021). URL: https://bxddb.lngs.infn.it/ddbcgi/ShowDocument?docid=944.

[100] J. N. Bahcall, M. Kamionkowski, and A. Sirlin. "Solar neutrinos: Radiative corrections in neutrino-electron scattering experiments". In: *Physical Review D* 51 (1995). DOI: 10.1103/PhysRevD.51.6146. URL: https://link.aps.org/doi/10.1103/PhysRevD.51.6146.

[101] V. Antonelli et al. "Solar Neutrinos". In: *Advances in High Energy Physics* 2013 (2013). ISSN: 1687-7357. DOI: 10.1155/2013/351926. URL: https://doi.org/10.1155/2013/351926.

[102] F.L. Villante et al. "A step toward CNO solar neutrino detection in liquid scintillators". In: *Physics Letters B* 701.3 (2011). ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2011.05.068. URL: https://www.sciencedirect.com/science/article/pii/S0370269311006125.

[103] J. B. Birks. "Scintillations from Organic Crystals: Specific Fluorescence and Relative Response to Different Radiations". In: *Proceedings of the Physical Society. Section A* 64.10 (1951). DOI: 10.1088/0370-1298/64/10/303. URL: https://dx.doi.org/10.1088/0370-1298/64/10/303.

[104] R. Brun and F. Rademakers. "ROOT — An object oriented data analysis framework". In: *Nuclear Instruments and Methods in Physics Research Section A* 389.1 (1997). New Computing Techniques in Physics Research V, pp. 81–86. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(97)00048-X. URL: https://www.sciencedirect.com/science/article/pii/S016890029700048X.

[105] Borexino Collaboration. "The Monte Carlo simulation of the Borexino detector". In: *Astroparticle Physics* 97 (2018). ISSN: 0927-6505. DOI: https://doi.org/10.1016/j.astropartphys.2017.10.003. URL: https://www.sciencedirect.com/science/article/pii/S0927650517301330.

[106] G. F. Knoll. *Radiation Detection and Measurement*. 605 Third Avenue, New York, USA: John Wiley & Sons, 2010. ISBN: 978-0-470-13148-0.

[107] Borexino Collaboration. "Seasonal modulation of the ^{7}Be solar neutrino rate in Borexino". In: *Astroparticle Physics* 92 (2017). ISSN: 0927-6505. DOI: https://doi.org/10.1016/j.astropartphys.2017.04.004. URL: https://www.sciencedirect.com/science/article/pii/S0927650517300245.

[108] S. Marcocci. "Precision Measurement of Solar ν Fluxes with Borexino and Prospects for $0\nu\beta\beta$ Search with ^{136}Xe-loaded Liquid Scintillators". PhD thesis. Gran Sasso Science Institute, 2016.

[109] Borexino Collaboration. "Identification of the cosmogenic ^{11}C background in large volumes of liquid scintillators with Borexino". In: *European Physical Journal C* 81 (2021). DOI: 10.1140/epjc/s10052-021-09799-x. URL: https://link.springer.com/article/10.1140/epjc/s10052-021-09799-x.

[110] Geant4 Collaboration. "GEANT4 — a simulation toolkit". In: *Nuclear Instruments and Methods in Physics Research Section A* 506.3 (2003). ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(03)01368-8. URL: https://geant4.web.cern.ch/.

[111] J. Allison et al. "Recent developments in GEANT4". In: *Nuclear Instruments and Methods in Physics Research Section A* 835 (2016). ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2016.06.125. URL: https://www.sciencedirect.com/science/article/pii/S0168900216306957.

[112] Geant4 Collaboration. *Physics Reference Manual*. Geant4 Collaboration, 2017. URL: https://indico.cern.ch/event/679723/contributions/2792554/attachments/1559217/2454299/PhysicsReferenceManual.pdf.

[113] D. Guffanti. "First simultaneous measurement of low-energy pp-chain solar neutrinos and prospects for CNO neutrino detection in Borexino". PhD thesis. Gran Sasso Science Institute, 2019.

[114] Super-Kamiokande Collaboration. "Atmospheric neutrino oscillation analysis with improved event reconstruction in Super-Kamiokande IV". In: *Progress of Theoretical and Experimental Physics* 2019.5 (2019). 053F01. ISSN: 2050-3911. DOI: 10.1093/ptep/ptz015. URL: https://doi.org/10.1093/ptep/ptz015.

[115] Z. Bagdasarian, X. Ding, and A. Vishneva. "Analytical response function for the Borexino solar neutrino analysis". In: *Journal of Physics: Conference Series* 1342.1 (2020). DOI: 10.1088/1742-6596/1342/1/012105. URL: https://doi.org/10.1088/1742-6596/1342/1/012105.

[116] R. Bonventre and G. D. Orebi Gann. "Sensitivity of a low threshold directional detector to CNO-cycle solar neutrinos". In: *The European Physical Journal C* 78.6 (2018). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-018-5925-7. URL: https://doi.org/10.1140/epjc/s10052-018-5925-7.

[117] C. Aberle et al. "Measuring directionality in double-beta decay and neutrino interactions with kiloton-scale scintillation detectors". In: *Journal of Instrumentation* 9.06 (2014). DOI: 10.1088/1748-0221/9/06/p06012. URL: https://doi.org/10.1088/1748-0221/9/06/p06012.

[118] J. R. Klein et al. "Future Advances in Photon-Based Neutrino Detectors: A SNOWMASS White Paper". In: *arXiv* (2022). DOI: 10.48550/ARXIV.2203.07479. URL: https://arxiv.org/abs/2203.07479.

[119] M. Yeh et al. "A new water-based liquid scintillator and potential applications". In: *Nuclear Instruments and Methods in Physics Research Section A* 660.1 (2011). ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2011.08.040. URL: https://www.sciencedirect.com/science/article/pii/S0168900211016615.

[120] Z. Guo et al. "Slow liquid scintillator candidates for MeV-scale neutrino experiments". In: *Astroparticle Physics* 109 (2019). ISSN: 0927-6505. DOI: https://doi.org/10.1016/j.astropartphys.2019.02.001. URL: https://www.sciencedirect.com/science/article/pii/S0927650518302718.

[121] E. Graham et al. “Light yield of Perovskite nanocrystal-doped liquid scintillator”. In: *Journal of Instrumentation* 14.11 (2019). DOI: 10.1088/1748-0221/14/11/p11024. URL: https://doi.org/10.1088/1748-0221/14/11/p11024.

[122] A.V. Lyashenko et al. “Performance of Large Area Picosecond Photo-Detectors (LAPPDTM)”. In: *Nuclear Instruments and Methods in Physics Research Section A* 958 (2020). Proceedings of the Vienna Conference on Instrumentation 2019. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2019.162834. URL: https://www.sciencedirect.com/science/article/pii/S0168900219312690.

[123] T. Kaptanoglu et al. “Cherenkov and scintillation separation in water-based liquid scintillator using an LAPPD”. In: *The European Physical Journal C* 82 (2022). DOI: 10.1140/epjc/s10052-022-10087-5. URL: https://doi.org/10.1140/epjc/s10052-022-10087-5.

[124] T. Kaptanoglu et al. “Spectral photon sorting for large-scale Cherenkov and scintillation detectors”. In: *Physical Review D* 101 (2020). DOI: 10.1103/PhysRevD.101.072002. URL: https://link.aps.org/doi/10.1103/PhysRevD.101.072002.

[125] T. Kaptanoglu, M. Luo, and J. Klein. “Cherenkov and scintillation light separation using wavelength in LAB based liquid scintillator”. In: *Journal of Instrumentation* 14.05 (2019). DOI: 10.1088/1748-0221/14/05/t05001. URL: https://doi.org/10.1088/1748-0221/14/05/t05001.

[126] B.S. Wonsak et al. “Topological track reconstruction in unsegmented, large-volume liquid scintillator detectors”. In: *Journal of Instrumentation* 13.07 (2018). DOI: 10.1088/1748-0221/13/07/p07005. URL: https://doi.org/10.1088/1748-0221/13/07/p07005.

[127] A. Elagin et al. “Separating double-beta decay events from solar neutrino interactions in a kiloton-scale liquid scintillator detector by fast timing”. In: *Nuclear Instruments and Methods A* 849 (2017). ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2016.12.033. URL: https://www.sciencedirect.com/science/article/pii/S0168900216312979.

[128] R. Jiang and A. Elagin. “Space-Time Discriminant to Separate Double-Beta Decay from ^{8}B Solar Neutrinos in Liquid Scintillator”. In: *arXiv* (2019). DOI: 10.48550/ARXIV.1902.06912. URL: https://arxiv.org/abs/1902.06912.

[129] W. Luo et al. “Reconstruction Algorithm for a Novel Cherenkov Scintillation Detector”. In: *arXiv* (2022). DOI: 10.48550/ARXIV.2209.13772. URL: https://arxiv.org/abs/2209.13772.

[130] J. Caravaca et al. “Characterization of water-based liquid scintillator for Cherenkov and scintillation separation”. In: *The European Physical Journal C* 80 (2020). ISSN: 9. DOI: 10.1140/epjc/s10052-020-8418-4. URL: https://doi.org/10.1140/epjc/s10052-020-8418-4.

[131] J. Gruszko et al. “Detecting Cherenkov light from 1–2 MeV electrons in linear alkylbenzene”. In: *Journal of Instrumentation* 14.02 (2019). DOI: 10.1088/1748-0221/14/02/p02005. URL: https://doi.org/10.1088/1748-0221/14/02/p02005.

[132] E. Tiras. “Detector R & D for ANNIE and Future Neutrino Experiments”. In: *arXiv* (2019). DOI: 10.48550/ARXIV.1910.08715. URL: https://arxiv.org/abs/1910.08715.

[133] B. J. Land et al. “MeV-scale performance of water-based and pure liquid scintillator detectors”. In: *Physical Review D* 103 (2021). DOI: 10.1103/PhysRevD.103.052004. URL: https://link.aps.org/doi/10.1103/PhysRevD.103.052004.

[134] Borexino Collaboration. "First Directional Measurement of Sub-MeV Solar Neutrinos with Borexino". In: *Physical Review Letters* 128 (2022). DOI: 10.1103/PhysRevLett.128.091803. URL: https://link.aps.org/doi/10.1103/PhysRevLett.128.091803.

[135] Borexino Collaboration. "Correlated and integrated directionality for sub-MeV solar neutrinos in Borexino". In: *Physical Review D* 105 (2022). DOI: 10.1103/PhysRevD.105.052002. URL: https://link.aps.org/doi/10.1103/PhysRevD.105.052002.

[136] SNO+ collaboration. "Development, characterisation, and deployment of the SNO+ liquid scintillator". In: *Journal of Instrumentation* 16.05 (2021). DOI: 10.1088/1748-0221/16/05/p05009. URL: https://doi.org/10.1088/1748-0221/16/05/p05009.

[137] I. Reda and A. Andreas. "Solar position algorithm for solar radiation applications". In: *Solar Energy* 76.5 (2004). ISSN: 0038-092X. DOI: https://doi.org/10.1016/j.solener.2003.12.003. URL: https://www.sciencedirect.com/science/article/pii/S0038092X0300450X.

[138] Borexino Collaboration. "Absence of a day–night asymmetry in the ^{7}Be solar neutrino rate in Borexino". In: *Physics Letters B* 707.1 (2012). ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2011.11.025. URL: https://www.sciencedirect.com/science/article/pii/S037026931101389X.

[139] L. Pelicci and X. Ding. "Systematics studies summary". In: *Internal Borexino Document Database* 1044 (2022). URL: https://bxddb.lngs.infn.it/ddbcgi/ShowDocument?docid=1044.

[140] C. Ghiano. "Final Status report on PMTs, electronics and data taking". In: *Internal Borexino Document Database* 993 (2021). URL: https://bxddb.lngs.infn.it/ddbcgi/ShowDocument?docid=993.

[141] M. Redchuk et al. "Effective Quantum Efficiency - New Improved Approach". In: *Internal Borexino Document Database* 446 (2019). URL: https://bxddb.lngs.infn.it/ddbcgi/ShowDocument?docid=446.

[142] S. Hardy. "Measuring the ^{7}Be Neutrino Flux From the Sun:Calibration of the Borexino Solar Neutrino Detector". PhD thesis. Virginia Polytechnic Institute and State University, 2010.

[143] M. Leung. "The Borexino Solar Neutrino Experiment:Scintillator Purification and Surface Contamination". PhD thesis. Princton University, 2006.

[144] F. James. "MINUIT Function Minimization and Error Analysis: Reference Manual Version 94.1". In: 1994. URL: https://root.cern.ch/download/minuit.pdf.

[145] S. Kumaran. "Geoneutrinos and solar neutrinos with the Borexino experiment". PhD thesis. RWTH Aachen, 2022.

[146] J. Martyn. "First Look at possible Cherenkov Calibration Source". In: *Internal Borexino Document Database* 778 (2020). URL: https://bxddb.lngs.infn.it/ddbcgi/ShowDocument?docid=778.

[147] A. Singhal. "Analysis strategies used in directional analysis of sub-MeV solar neutrinos in liquid scintillator detector (DPG 2022 talk)". In: *Internal Borexino Document Database* 998 (2022). URL: https://bxddb.lngs.infn.it/ddbcgi/ShowDocument?docid=998.

[148] A. Singhal. "Optimisation of number of bins for cos α histograms and background toy study". In: *Internal Borexino Document Database* 1010 (2022). URL: https://bxddb.lngs.infn.it/ddbcgi/ShowDocument?docid=1010.

[149] A. Weßel. "Directionality measurement of solar neutrinos CNO solar neutrino analysis with Borexino". MA thesis. RWTH Aachen, 2022.

[150] V. Di Marcello et al. "Fluid-dynamics and transport of 210Po in the scintillator Borexino detector: A numerical analysis". In: *Nuclear Instruments and Methods in Physics Research Section A* 964 (2020). ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2020.163801. URL: https://www.sciencedirect.com/science/article/pii/S0168900220303284.

[151] E. A. Maxwell. "Chi-square Intervals for a Poisson Parameter - Bayes, Classical and Structural". In: (2011). DOI: 10.48550/ARXIV.1102.0822. URL: https://arxiv.org/abs/1102.0822.

[152] Super-Kamiokande Collaboration. *Solar neutrinos in Super-Kamiokande*. https://www-sk.icrr.u-tokyo.ac.jp/en/sk/about/research/. Accessed: 2022-12-16.

[153] Borexino Collaboration. "Sensitivity to neutrinos from the solar CNO cycle in Borexino". In: *The European Physical Journal C* 80.11 (2020). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-020-08534-2. URL: https://doi.org/10.1140/epjc/s10052-020-08534-2.

[154] D. Bravo-Berguño et al. "The Borexino Thermal Monitoring and Management System and simulations of the fluid-dynamics of the Borexino detector under asymmetrical, changing boundary conditions". In: *Nuclear Instruments and Methods in Physics Research Section A* 885 (2018). ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2017.12.047. URL: https://www.sciencedirect.com/science/article/pii/S0168900217314535.

[155] V. Di Marcello et al. "Natural Convection and Transport of Background Contamination in the Borexino Neutrino Detector". In: *Journal of Fluids Engineering* 144.8 (2022). 081210. ISSN: 0098-2202. DOI: 10.1115/1.4053895. URL: https://doi.org/10.1115/1.4053895.

[156] A. Singhal. "Optimisation of number of bins for cos α histograms and background toy study". In: *Internal Borexino Document Database* 1075 (2022). URL: https://bxddb.lngs.infn.it/ddbcgi/ShowDocument?docid=1075.

[157] A. Weßel. "CID for CNO - Data Selection Cuts". In: *Internal Borexino Document Database* 1048 (2022). URL: https://bxddb.lngs.infn.it/ddbcgi/ShowDocument?docid=1048.

[158] D. Basilico et al. "CNO results with Phase-III final dataset". In: *Internal Borexino Document Database* 1040 (2022). URL: https://bxddb.lngs.infn.it/ddbcgi/ShowDocument?docid=1040.

[159] J. N. Bahcall et al. "Standard neutrino spectrum from ^{8}B decay". In: *Physical Review C* 54 (1996). DOI: 10.1103/PhysRevC.54.411. URL: https://link.aps.org/doi/10.1103/PhysRevC.54.411.

[160] G. J. Feldman and R. D. Cousins. "Unified approach to the classical statistical analysis of small signals". In: *Physical Review D* 57 (1998). DOI: 10.1103/PhysRevD.57.3873. URL: https://link.aps.org/doi/10.1103/PhysRevD.57.3873.

[161] The NOvA Collaboration. "The Profiled Feldman-Cousins technique for confidence interval construction in the presence of nuisance parameters". In: *arXiv* (2022). DOI: 10.48550/ARXIV.2207.14353. URL: https://arxiv.org/abs/2207.14353.

[162] R. D. Cousins and V. L. Highland. "Incorporating systematic uncertainties into an upper limit". In: *Nuclear Instruments and Methods in Physics Research Section A* 320.1 (1992). ISSN: 0168-9002. DOI: https://doi.org/10.1016/0168-9002(92)90794-5. URL: https://www.sciencedirect.com/science/article/pii/0168900292907945.

[163] J. Conrad et al. "Including systematic uncertainties in confidence interval construction for Poisson statistics". In: *Physical Review D* 67 (2003). DOI: 10.1103/PhysRevD.67.012002. URL: https://link.aps.org/doi/10.1103/PhysRevD.67.012002.

[164] S. Greenland et al. "Statistical tests, P values, confidence intervals, and power: A guide to misinterpretations". In: *European Journal of Epidemiology* 31 (2016). DOI: 10.1007/s10654-016-0149-3.

[165] D. S. Sivia and J. Skilling. *Data Analysis - A Bayesian Tutorial*. 2nd. Oxford Science Publications. Oxford University Press, 2006.

[166] P. Vischia. "Unfolding by Folding: a resampling approach to the problem of matrix inversion without actually inverting any matrix". In: *arXiv* (2020). DOI: 10.48550/ARXIV.2009.02913. URL: https://arxiv.org/abs/2009.02913.

[167] L. Pelicci. "New CNO results with multivariate fit with CID constraint". In: *Internal Borexino Document Database* 1115 (2022). URL: https://bxddb.lngs.infn.it/ddbcgi/ShowDocument?docid=1048.

[168] J. Dunger, E. J. Leming, and S. D. Biller. "Slow-fluor scintillator for low energy solar neutrinos and neutrinoless double beta decay". In: *Physical Review D* 105 (2022). DOI: 10.1103/PhysRevD.105.092006. URL: https://link.aps.org/doi/10.1103/PhysRevD.105.092006.

[169] JUNO Collaboration. "Feasibility and physics potential of detecting ^{8}B solar neutrinos at JUNO". In: *arXiv* (2020). DOI: 10.48550/ARXIV.2006.11760. URL: https://arxiv.org/abs/2006.11760.

[170] A. Abusleme et al. "Optimization of the JUNO liquid scintillator composition using a Daya Bay antineutrino detector". In: *Nuclear Instruments and Methods in Physics Research Section A* 988 (2021). ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2020.164823. URL: https://www.sciencedirect.com/science/article/pii/S0168900220312201.