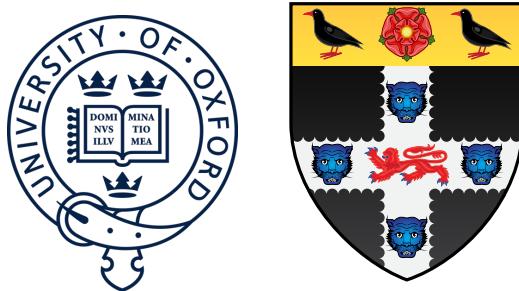


First Measurement of the Solar Neutrino Oscillation Parameters via Boron-8 Solar Neutrinos in SNO+



Daniel Cookman

Christ Church College
University of Oxford

A thesis submitted in fulfilment of the requirements for the degree
of
Doctor of Philosophy
Trinity Term 2022

To my parents

and

To Oscar Jacobsson:

The best of us

Acknowledgements

And I would like to acknowledge ...

Abstract

Formal 1-page summary of the work completed in the thesis.

Statement of Authorship

The nature of contemporary experimental particle physics is that the research work is highly collaborative. The SNO+ experiment, which this thesis is about, is an 80+ member Collaboration and as such is no exception. As a result, the work discussed in this thesis is in many places built upon the tireless work of many others within the Collaboration. This Statement, as well as throughout the thesis itself, endeavours to note where the author himself did the research, and where work was taken from others.

Chapters 1 and 2 cover the background knowledge of neutrino physics as well as the SNO+ detector needed to best understand the work performed in the rest of this thesis. This information has been compiled from a variety of books, journal articles, theses, and SNO+ internal technical reports. The author was trained as both a detector ‘operator’ and ‘expert’, taking dozens of shifts to monitor the detector whilst it was taking data, as well as being on-call to help if any issues arose with the detector electronics. The author also made numerous contributions to the Collaboration’s central software package, RAT.

Chapters 3–5 cover the work done by the author on the SMELLIE optical calibration system for SNO+. This calibration system historically been built and analysed by multiple people, including: Krishanu Majumdar, Esther Turner, Stephanie Langrock, and Jeff Lidgard. The author installed multiple pieces of new hardware for SMELLIE on-site, with help from Armin Reichold and Jeff Lidgard. With the same team, the author also helped to integrate the new hardware with

the existing SMELLIE server software so that the newly-installed hardware could actually be effectively used. The author has taken dozens of hours of calibration data using SMELLIE throughout the course of the DPhil, both by himself and alongside Armin Reichold, Jeff Lidgard, and Ana Sofia Inácio.

Chapter 4 considers improvements to the simulation of SMELLIE events; the work done by Esther Turner on this subject is considered as a starting point. The rest of the chapter covers a new simulation approach created and developed by the author. Similarly, the scattering analysis performed within Chapter 5 was inspired from the analyses done by Krishanu Majumdar, Esther Turner, and Stephanie Langrock. However, the work in this thesis uses a new method created by the author, and uses new scintillator phase data taken by the author. Also in this chapter is a separate analysis of this same data to measure the scintillator’s extinction length. This was a novel analysis for SMELLIE, designed and implemented entirely by the author.

Finally, Chapter 6 describes an analysis of scintillator phase data performed by the author to measure the solar neutrino oscillation parameters. An initial background-free sensitivity study of this topic was originally performed by Javier Caravaca. However, the work done by the author in this thesis analyses actual data, includes all relevant backgrounds appropriately, and considers systematics. Projections of expected sensitivity with greater livetime were also performed by the author. The analysis made by the author also uses a Bayesian MCMC approach, using the `OXO` signal extraction software framework initially built by Jack Dunger. A number of further people have since improved `OXO`, including the author: substantial improvements to the handling of systematics floated within the MCMC fit, as well as allowing for the floating of oscillation parameters within the fit.

Table of contents

| | |
|---|------------|
| List of Figures | xv |
| List of Tables | xix |
| List of Acronyms | xxi |
| Introduction | 1 |
| 1 The Theory of Neutrino Physics | 3 |
| 1.1 The Standard Model and Neutrinos | 3 |
| 1.1.1 A Brief Introduction to the Standard Model | 3 |
| 1.1.2 Neutrinos within the Standard Model | 4 |
| 1.2 Neutrino Oscillations and Neutrino Masses | 5 |
| 1.2.1 The Evidence for Neutrino Oscillations | 5 |
| 1.2.2 The Phenomenology of Neutrino Oscillations | 6 |
| 1.2.3 The Origins of Neutrino Mass | 6 |
| 2 The SNO+ Detector | 9 |
| 2.1 Detector Geometry | 9 |
| 2.2 Experimental Phases | 10 |
| 2.3 Detecting and Recording an Event in SNO+: A Journey | 12 |
| 2.3.1 Particle Interactions with Matter | 12 |
| 2.3.2 Optical Processes | 18 |

| | | |
|----------|---|-----------|
| 2.3.3 | Detection by PMTs | 25 |
| 2.3.4 | Data Acquisition and Triggering | 28 |
| 2.3.5 | Operation of the Detector | 30 |
| 2.4 | Calibrations and Detector Modelling | 31 |
| 2.4.1 | Detector Monitoring | 32 |
| 2.4.2 | Electronic and PMT Calibrations | 32 |
| 2.4.3 | Energy and Optical Calibrations | 34 |
| 2.4.4 | Event Reconstruction | 36 |
| 2.4.5 | Event Simulation | 39 |
| 3 | The SMELLIE Calibration System | 41 |
| 3.1 | The SMELLIE Hardware | 44 |
| 3.1.1 | Lasers | 44 |
| 3.1.2 | Controlling Laser Intensities | 46 |
| 3.1.3 | Propagation of Light into the Detector | 50 |
| 3.1.4 | The Monitoring PMT Unit | 53 |
| 3.1.5 | Event Triggering and Data Acquisition | 55 |
| 3.2 | Software for SMELLIE Data-taking | 58 |
| 4 | Simulating SMELLIE Events | 61 |
| 4.1 | Improving the SMELLIE Generator Algorithm | 62 |
| 4.1.1 | Previous Attempts at SMELLIE Event Simulation | 62 |
| 4.1.2 | The new generator | 65 |
| 4.2 | Improving the beam profiles | 69 |
| 4.2.1 | Combining beam profile datasets | 72 |
| 4.2.2 | Results & Discussion | 78 |
| 4.3 | Comparisons between Data and Simulation | 82 |
| 4.3.1 | Forward Hemisphere Discrepancies | 83 |
| 4.3.2 | Emission Time Discrepancies | 83 |

| | | |
|----------|---|-----------|
| 4.3.3 | Backward Hemisphere Discrepancies | 83 |
| 5 | Analysis of SMELLIE Data in the Scintillator Phase | 85 |
| 5.1 | Extinction Length Analysis | 85 |
| 5.1.1 | Motivation | 85 |
| 5.1.2 | Analysis Approach | 86 |
| 5.1.3 | Validation of the Analysis in Simulation | 87 |
| 5.1.4 | Results in Data | 87 |
| 5.2 | Scattering Analysis | 88 |
| 5.2.1 | Historical Approaches and the Problem of Systematics . . | 88 |
| 5.2.2 | New Methodology | 89 |
| 5.2.3 | Results | 90 |
| 6 | Solar Oscillation Analysis | 91 |
| 6.1 | Analysis Methodology | 92 |
| 6.1.1 | Observational Principle | 92 |
| 6.1.2 | Background Processes | 95 |
| 6.1.3 | The Log-likelihood Test Statistic | 103 |
| 6.1.4 | The Bayesian Statistical Approach | 104 |
| 6.1.5 | Markov Chain Monte Carlo | 105 |
| 6.1.6 | Choosing Priors | 107 |
| 6.1.7 | Including Systematics in the Fit | 108 |
| 6.1.8 | Including Oscillations in the Fit | 110 |
| 6.2 | Analysis on Scintillator-Phase data | 112 |
| 6.2.1 | Dataset and Livetime | 112 |
| 6.2.2 | Event Selection | 114 |
| 6.2.3 | Expected Rates and their Constraints | 118 |
| 6.2.4 | Systematics | 125 |
| 6.2.5 | Results | 128 |

| | |
|--|------------|
| 6.3 Sensitivity Projections | 139 |
| 6.4 Summary and Suggestions for Further Work | 144 |
| 7 Conclusions | 149 |
| References | 151 |

List of Figures

| | | |
|------|--|----|
| 2.1 | 3D model of the SNO+ detector | 10 |
| 2.2 | Comparison of the SNO+ detector media's emission properties, versus optical phase | 14 |
| 2.4 | Scattering lengths in simulation for UPW, LAB, and 2.2 g/L LABPPO | 22 |
| 2.5 | Measured properties of the UPW and acrylic in the water phase. . | 23 |
| 2.6 | Refractive indices of acrylic, UPW, and LABPPO as a function of wavelength; also reflectance as a function of incidence angle . . . | 24 |
| 2.8 | Efficiencies of the R1408-type PMTs used as standard within SNO+ | 27 |
| 2.9 | Example charge spectra for a PMT as a function of the true npe generated | 27 |
| 2.10 | Schematic of the front-end electronics used for data acquisition and triggering in SNO+ | 28 |
| 3.1 | Diagram of the SMELLIE calibration system within SNO+ | 42 |
| 3.3 | Diagram of the connections between the various bits of SMELLIE hardware and the rest of the SNO+ detector | 45 |
| 3.4 | Emission wavelength and timing spectra of the lasers used in SMELLIE | 47 |
| 3.5 | Typical impact of driving voltage intensity for PQ446 laser on observed maximum occupancy in the detector | 48 |

| | | |
|------|--|----|
| 3.6 | Distribution of the number of PMT hits observed per event for laser PQ446 | 49 |
| 3.7 | Picture of the VFA during installation into the ELLIE hardware rack | 50 |
| 3.8 | Maximum occupancy in the detector as a function of the attenuation parameter value used | 51 |
| 3.9 | Fibre emission points and nominal pointing directions | 53 |
| 3.10 | Comparison between typical MPU CAEN traces generated by the PQ495 before and after the Summer 2022 hardware upgrades . . . | 55 |
| 4.1 | Comparison of SMELLIE between data and MC using a 1D generator | 63 |
| 4.2 | Number of attempts taken for existing 2D SMELLIE generator to accept direction | 66 |
| 4.3 | Intensity map of existing beam profile, and interpolated sampling map | 67 |
| 4.4 | Histogram of interpolated intensities within the 2D direction-space | 68 |
| 4.5 | Comparison of water-phase data to MC generated using both the old and new 2D beam profile generator approaches, with the updated beam profiles | 70 |
| 4.6 | Plot of $\Delta\chi^2$ for both single subruns of a typical PMT, and when all relevant subruns are combined | 77 |
| 4.7 | Comparison between old and updated beam profiles for fibre FS055, after combining multiple data sets | 79 |
| 4.8 | Interpolated intensity map for the new updated beam profile of fibre FS055 | 80 |
| 4.9 | Residuals from subruns at two different wavelengths, both compared to the combined beam profile model for fibre FS055 | 82 |
| 6.1 | True ${}^8\text{B}$ E_ν versus true E_e | 93 |

| | | |
|------|--|-----|
| 6.2 | The evolution of energy distributions related to ${}^8\text{B}$ solar neutrino detection. | 94 |
| 6.3 | P_{ee} as a function of true neutrino energy, scanned over a variety of oscillation parameter values | 95 |
| 6.4 | The ${}^{238}\text{U}$ and ${}^{232}\text{Th}$ decay chains. | 96 |
| 6.5 | ${}^{214}\text{Bi-Po}$, ${}^{212}\text{Bi-Po}$, and ${}^8\text{B}$ observed energy spectra before and after out-of-window and in-window cuts | 98 |
| 6.6 | Schematic of (α, n) interactions. | 99 |
| 6.7 | (α, n) reconstructed energy spectrum in MC within the analysis ROI, before and after out-of-window cuts have been applied | 100 |
| 6.8 | Radial dependence of external backgrounds considered in this analysis, as compared to the ${}^8\text{B}$ signal | 101 |
| 6.10 | Demonstration of how to handle an energy scaling systematic correctly | 109 |
| 6.11 | Distribution of $\cos \theta_z$, for the events within this dataset | 114 |
| 6.12 | Comparison of tagged ${}^{214}\text{Bi}$ reconstructed energy distributions between data and MC | 127 |
| 6.13 | Examples of the sampled values for parameters within a given MCMC chain as a function of the step number, after tuning the proposal distribution width parameters. | 130 |
| 6.14 | Plot showing the auto-correlation as a function of step delay for three variables in a typical chain | 131 |
| 6.15 | Marginalised 1D posterior density distributions for all non-oscillation parameters. | 132 |
| 6.16 | Correlation matrix between all parameters in the MCMC fit | 134 |
| 6.17 | 2D posterior density distribution marginalised onto Δm_{21}^2 and θ_{12} | 134 |
| 6.18 | Comparison of 1D posterior density distribution marginalised onto θ_{12} for different flux constraints | 135 |
| 6.19 | Comparison of data to MC for each radial slice | 138 |

| | |
|---|-----|
| 6.20 Posterior density posterior distributions marginalised onto the two solar oscillation parameters, for each livetime scenario | 141 |
| 6.21 Posterior density posterior distributions marginalised onto each of θ_{12} and Δm_{21}^2 , for each livetime scenario | 142 |
| 6.22 Comparison of projections over 1 year of livetime, with different background expectations. | 145 |

List of Tables

| | | |
|-----|---|-----|
| 2.1 | Current values used to model scintillator emission from electrons in 2.2 g/L LABPPO | 15 |
| 3.1 | SMELLIE fibre names, their associated mounting nodes on the PSUP, and their pointing direction. | 52 |
| 4.1 | Water-phase runs used for new beam profiling. | 73 |
| 6.1 | Cuts used in this analysis. | 115 |
| 6.2 | Impact of each cut on the quantity of events in data. | 117 |
| 6.3 | Combined impact of cuts on each MC process. | 117 |
| 6.4 | Measured rates of the external backgrounds during the water phase of SNO+, by Tony Zummo | 123 |
| 6.5 | Number of events expected both before and after cuts have been applied, along with any constraints. | 124 |
| 6.6 | Comparison of the fit parameter values obtained when getting the HPD values after marginalising onto each parameter, versus looking at the sampled point with the maximal likelihood found. | 137 |

List of Acronyms

ADC Analogue-to-Digital Converter

AmBe Americium-Beryllium radioactive source

AMELLIE Attenuation Module for ELLIE

AV Acrylic vessel

BD 1,2-Butanediol

BHT Butylated hydroxytoluene

BisMSB 1,4-Bis(2-methylstyryl)benzene

CI Credible Interval

CTC Crate Trigger Card

DAQ Data acquisition (system)

DB Daughter Board

DDA N,N-Dimethyldodecylamine

ECA Electronic Calibration

ELLIE Embedded LED/Laser Light Injection Entity

EXTA External Asynchronous (Trigger)

FEC Front-End Card

FV Fiducial Volume

GTID Global Trigger Identification number

GT Global Trigger

HPD (The point of) Highest Posterior Density

IW In-window events

LAB Linear alkylbenzene

MPU Monitoring PMT Unit (for SMELLIE)

MTC/A+ Analogue Master Trigger Card

MTC/D Digital Master Trigger Card

NI Unit National Instruments DAQ Unit (for SMELLIE)

npe Number of photoelectrons

OOW Out-of-window events

OWLs Outward-looking PMTs

PCA PMT Calibration

PMTIC PMT Interface Card

PMT Photomultiplier Tube

PPO 2,5-Diphenyloxazole

PQ PicoQuant

PSUP PMT support structure

QE Quantum efficiency (of a PMT)

QHL Charge with high gain over a ‘long’ integration time (390 ns)

QHS Charge with high gain over a ‘short’ integration time (60 ns)

QLX Charge with low gain over a ‘long’ integration time (390 ns)

ROI Region of Interest (of a physics analysis)

SMELLIE Scattering Module for ELLIE

SNO Sudbury Neutrino Observatory

TAC Time-to-amplitude Converter

TeA Telluric acid, Te(OH)₆

TELLIE Timing subsystem for ELLIE

TeLS Tellurium-loaded liquid scintillator

TIR Total Internal Reflection

TOF Time-of-flight (of a photon)

TTS Transit time spread (of a PMT)

TUBii Trigger Utility Board Mark ii

UPW Ultra-pure water

VFA Remotely-controllable Variable Fibre Attenuator

ZDAB Zebra Database (file format)

nhit Number of PMT hits in an event

RATDB RAT Database

RAT Reactor Analysis Tool

Introduction

1

Couple of pages outlining document's structure and contents (this is what each of
the chapters is here for). Less formal than the abstract, also explaining what the
expected audience of this thesis is: who will find this document useful!

2

3

4

Chapter 1

1

The Theory of Neutrino Physics

2

*Light**Light**The visible reminder of Invisible Light*

3

The Rock

T. S. ELIOT

1.1 The Standard Model and Neutrinos

4

1.1.1 A Brief Introduction to the Standard Model

5

Covering how the SM works at the highest level, including:

6

- Quantum Field Theory and the Lagrangian dynamical framework
- The connection between symmetries of a QFT model and its gauge fields that describe the model's forces
- The SM's fundamental symmetries, and associated forces, but —

7

8

9

10

- 1 • Not (exactly) what we see “normally”! The electromagnetic and weak forces
 2 appear distinct, and the weak gauge bosons have mass. To explain this, we
 3 need a further component, the Brout-Englert-Higgs (BEH) Mechanism.
- 4 [2 pages total]

5 **1.1.2 Neutrinos within the Standard Model**

- 6 • Basic description of where neutrinos fit into SM: 3 kinds of neutral fermion,
 7 the counterparts to the charged fermions. Interacts with the weak force
 8 only.
- 9 • Summary of the experimental evidence for this picture: mainly, the dis-
 10 covery of electron anti-neutrinos by Cowan and Reines, the muon neutrino
 11 by Lederman, Schwartz, and Steinberger, and the tau neutrino by the
 12 DONUT Collaboration. Further critical experiments include the first mea-
 13 surement of a neutrino’s helicity by Goldhaber et al. as well as Danby et
 14 al.’s demonstration that ν_μ are distinct from ν_e .
- 15 • More detailed description, via Feynman diagrams, of the two fundamental
 16 modes of interaction by neutrinos with the weak force: charged- and neutral-
 17 current interactions. A brief mention of the quantitative theory that underlies
 18 description: Gashow, Salam, and Weinberg’s Electroweak Theory. This
 19 explains not only the V–A structure of charged-current interactions, but
 20 also predicted accurately the nature of neutral-current interactions. (Given
 21 space constraints, I see no reason to go into much of the details of the theory,
 22 or the many experimental tests of its structure.)

23
$$\frac{d\sigma_{\nu_i}}{dT} = BLAH \quad (1.1)$$

24 [4 pages]

1.2 Neutrino Oscillations and Neutrino Masses

1

1.2.1 The Evidence for Neutrino Oscillations

2

- Describe status quo ante of massless nature of neutrinos: BEH mechanism as exists cannot allow for neutrinos to have mass as only left-handed neutrinos have been observed.
3
4
5
- Furthermore, strong experimental limits on neutrino masses, from e.g. tritium-decay endpoint measurements by the KATRIN experiment and cosmological inferences from the CMB by the Planck satellite.
6
7
8
- But — then neutrino oscillations are observed over a variety of experiments and contexts. Summarise critical bits of evidence:
9
10
- Electron neutrino disappearance in solar neutrino experiments, including Ray Davis' Homestake experiment, the SAGE/GALLEX experiments, and SNO. For the latter, the comparison of charged-current and neutral-current modes of interaction was clear evidence of neutrino oscillations over other types of process (e.g. neutrino decay).
11
12
13
14
15
- Include in the above a brief description of Bahcall's Standard Solar Model.
16
- Muon neutrino disappearance in atmospheric and long-baseline accelerator neutrino experiments, such as Super-Kamiokande, T2K, and Nova.
17
18
- A few further observations to note are: reactor electron anti-neutrino disappearance from both KamLAND and Daya Bay; tau neutrino appearance at the OPERA experiment; short-baseline neutrino anomaly within LSND and MiniBooNE (with recent contrary evidence from MicroBooNE).
19
20
21
22

1.2.2 The Phenomenology of Neutrino Oscillations

- 2 • Describe the current phenomenological model of 3-flavour neutrino oscillations that can explain all of this evidence: the PMNS mixing matrix.
- 4 • Describe also the MSW effect, which is critical for explaining solar neutrino oscillations.
- 6 • Show the formula for solar neutrino oscillations, given this MSW effect in both the Sun and Earth. Note the dependence of solar neutrino oscillations on only the “solar” oscillation parameters. This is all particularly useful for the solar analysis chapter.

$$10 \quad P_{ee} (\tan 2\theta_{12}^M, \sin \theta_{13}^M, \Delta m_{21,M}^2) = BLAH \quad (1.2)$$

11 [3 pages]

1.2.3 The Origins of Neutrino Mass

- 13 • Observed neutrino oscillations require at least two neutrino mass states to be non-zero. Given constraints of the current SM, two main ways of adding neutrino masses: a Dirac mass term (i.e. allowing for sterile neutrinos), and a Majorana mass term.
- 17 • For latter, briefly describe what a Majorana particle is, and how with the Seesaw Mechanism (just the simple Type 1 described in-text) one can not only get neutrino masses but also explain their lightness relative to the other massive SM particles. Note that there exist more elaborate versions of this theory.
- 22 • Furthermore, with reference to the Sakharov conditions, describe qualitatively how the Seesaw Mechanism also allows for possible leptogene-

sis/baryogenesis in the early Universe, and hence could explain its matter-¹
antimatter asymmetry.²

- Describe briefly the nuclear physics behind double-beta decay (i.e. why it can happen at all over just normal beta decay), and then how Majorana neutrinos allow for neutrinoless double beta decay, $0\nu\beta\beta$.³
- Describe the experimental signature of $0\nu\beta\beta$: a spike of events of observed energy equal to the Q-value of the decay.⁴
- Note Schechter-Valle Theorem ensures that any observation of $0\nu\beta\beta$ must be the result of neutrinos being Majorana. I.e. the Universe cannot conspire against us and have $0\nu\beta\beta$ without Majorana neutrinos.⁵
- Very briefly note the current status of the search for $0\nu\beta\beta$, describing the main varieties of experimental setup seen, along with a nice canonical example of such an experiment and their best limit. In particular, the Germanium-crystal detectors such as GERDA, Xenon-TPC detectors like EXO-200, and large-scale liquid scintillators such as KamLAND-Zen.¹¹¹²¹³¹⁴¹⁵

[3 pages]¹⁶

[CHAPTER TOTAL: 17 pages]¹⁷

Chapter 2

1

The SNO+ Detector

2

The light-soaked days are coming.

3

JOHN GREEN

2.1 Detector Geometry

4

The SNO+ detector is a large, multi-purpose neutrino detector built in the SNOLAB underground laboratory near Sudbury, Canada. Its main detector structure is taken from the Sudbury Neutrino Observatory (SNO) [1], which can be seen in Fig. 2.1. At the heart of the detector lies the main detector medium, which changes depending on the phase of the experiment — more on the specifics of this in Section 2.2. This medium is held within a 12 m diameter sphere known as the Acrylic Vessel (AV). The AV floats within a body of ultra-pure water (UPW), beyond which is a stainless steel support structure (PSUP) that holds 9362 inward-facing Photomultiplier Tubes (PMTs). It is these PMTs that detect the light generated from physics events that occur within the detector medium. The AV is kept in place relative to the PSUP through a series of ‘hold-up’ and ‘hold-down’ tensylon ropes. All of these components are suspended within a large cylindrical cavity also filled with UPW. 91 outward-looking PMTs (OWLs) are

5

6

7

8

9

10

11

12

13

14

15

16

17

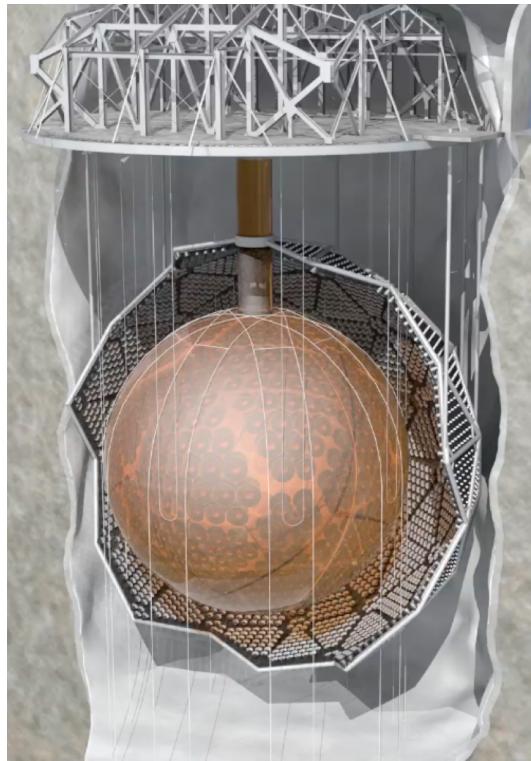


Fig. 2.1: 3D model of the SNO+ detector [2].

¹ also affixed to the outside of the PSUP, allowing for the effective vetoing of cosmic
² ray muons.

³ Directly above the detector is the Deck, within which all the detector electronics
⁴ are kept. Access within the AV for calibration tools and filling is possible only
⁵ through the acrylic ‘neck’ on top of the AV. Full details of the design of the current
⁶ detector can be found in [2].

⁷ 2.2 Experimental Phases

⁸ As mentioned earlier, SNO+ was designed to fulfil a number of physics goals over
⁹ multiple ‘phases’ of the detector’s lifetime. The phases are distinguished by the
¹⁰ medium that fills the AV. The first main phase (after a brief **Air Fill Phase** used
¹¹ only for detector commissioning) was that of the **Water Fill Phase**, with data
¹² taken between May 2017 and July 2019. This was used to perform fundamental

optical calibrations of the detector [3], measurements of the solar neutrino flux [4],
1
observation of neutrino oscillations in reactor anti-neutrinos [5], and searches for
2
nucleon decay [6, 7].
3

After this, the detector was filled with 780 tonnes of liquid scintillator known as
4
linear alkylbenzene (LAB), mixed with the fluor 2,5-diphenyloxazole (PPO). More
5
information on the physics of scintillators can be found in Section 2.3.1. Filling
6
of the LABPPO cocktail had to be paused in March 2020 due to the COVID-19
7
pandemic, leading to the detector having its bottom half still filled with UPW, and
8
the top half filled with LAB and PPO at 0.5 g/L. This impromptu phase became
9
known as the **Partial Fill**, and allowed for some creative analyses to be performed:
10
an initial neutrino oscillation analysis from reactor anti-neutrinos [8], as well as
11
the first ever observation of directionality in a high light yield scintillator [9, 10].
12
Eventually, filling of the detector with liquid scintillator completed in May 2021.
13
At that point, the concentration of PPO in the detector was at 0.6 g/L, markedly
14
below the target level of 2.0 g/L. A further ‘PPO top-up’ campaign then proceeded,
15
finishing in April 2022 with a final concentration of 2.2 g/L PPO. Thus began
16
the **Scintillator Phase** of the experiment, which continues on during the time
17
of writing. The main goals for this phase include a number of solar neutrino
18
analyses (including the one described in Chapter 6), a precision measurement
19
of the neutrino oscillation parameter Δm_{21}^2 using reactor anti-neutrinos [8], and
20
further calibrations of the detector and its backgrounds.
21

Finally, in the near future the detector will be loaded with Tellurium for the
22
Tellurium Phase, allowing for the flagship analysis of the experiment to begin:
23
neutrinoless double beta decay. In order to load Te within the liquid scintillator
24
in a stable manner, a chemical loading process has been developed, as described
25
in [11]. The Te starts within Te(OH)_6 (telluric acid, otherwise known as TeA),
26
which after purification will be reacted with 1,2-butanediol (BD) via heating and
27

¹ addition of N,N-Dimethyldodecylamine (DDA), which acts as a stabiliser. What
² results is tellurium-loaded scintillator, TeLS.

³ Two further chemicals are planned to be added to the scintillator cocktail.
⁴ The antioxidant butylated hydroxytoluene (BHT) will be added to capture any
⁵ free-radicals within the liquid scintillator, hopefully preventing any oxidation
⁶ reactions that could lead to the ‘yellowing’ of the scintillator, a degradation of its
⁷ optical properties. The addition of BHT is not expected to impact the detector’s
⁸ optics in any substantial way. However, the other substance to also be added,
⁹ 1,4-Bis(2-methylstyryl)benzene (bisMSB), will impact the optics. BisMSB acts as
¹⁰ a ‘wavelength-shifter’ which enables the scintillator cocktail to transmit light with
¹¹ a greater overall detection efficiency — more on the details of this in Section 2.3.1.

¹² **2.3 Detecting and Recording an Event in SNO+:**

¹³ **A Journey**

¹⁴ To understand the SNO+ detector well, it is worth thinking about how the
¹⁵ information contained in a physics event, e.g. a solar neutrino interaction, gets
¹⁶ observed. This section follows the journey of such an event.

¹⁷ **2.3.1 Particle Interactions with Matter**

¹⁸ All observable physics events within the detector begin by the generation of some
¹⁹ form of ionising radiation: α , β^\pm , γ , p or n . These can be created via numerous
²⁰ processes, both exciting (e.g. $0\nu\beta\beta$ or interactions of neutrinos) and annoying
²¹ (e.g. decay of background radioisotopes): see Section 6.1.2 for some of them.
²² Regardless of their origin, these particles begin propagating through the detector,
²³ and interacting with the detector medium. A number of mechanisms then allow
²⁴ for the generation of optical-wavelength light as a result of these interactions.

Cherenkov Light Emission

Whenever a charged particle travels through a dielectric medium at speeds faster than the speed of light in that medium, light is generated from the ‘wake’ of induced dipoles. This is known as **Cherenkov light**, a process much akin to the ‘sonic boom’ that occurs when an object travels at supersonic speeds through a medium. This light emanates outwards in a cone along the direction of the charge’s travel; the angle of the cone θ_γ being purely a function of the speed of the charged particle relative to the speed of light in vacuum, β , and the refractive index of the medium $n(\omega)$ at a given frequency ω : $\cos \theta_\gamma(\omega) = \frac{1}{n(\omega)\beta}$. There is then a minimum speed necessary for Cherenkov light to be generated: $\beta_{\min}(\omega) = 1/n(\omega)$.

In addition to the characteristic cone shape of the light, the spectrum of the light generated is also distinctive. Igor Tamm and Ilya Frank determined the expected energy dE emitted per unit length travelled by the charged particle, dx , as [12]:

$$\frac{dE}{dx} = \frac{q^2}{c^2} \int_{\beta n(\omega) > 1} \omega \left(1 - \frac{1}{\beta^2 n^2(\omega)} \right) d\omega. \quad (2.1)$$

Here, q is the charge of the moving particle. The Cherenkov emission spectrum during the water phase is shown in the black dotted line of Fig. 2.2.

All SNO+ detection media allow Cherenkov light to be generated, as long as sufficiently high energy particles traverse it. In the water fill phase of the detector, Cherenkov light was the only means by which light could be generated. Light from Cherenkov emission can still be created in liquid scintillator, but it tends to be swamped by another form of light generation: scintillation.

Scintillation

For certain special classes of material, the excitation and ionisation of atomic electrons nearby a moving charged particle can lead to the generation of optical-wavelength light, in a process known as **scintillation** (often generally referred to

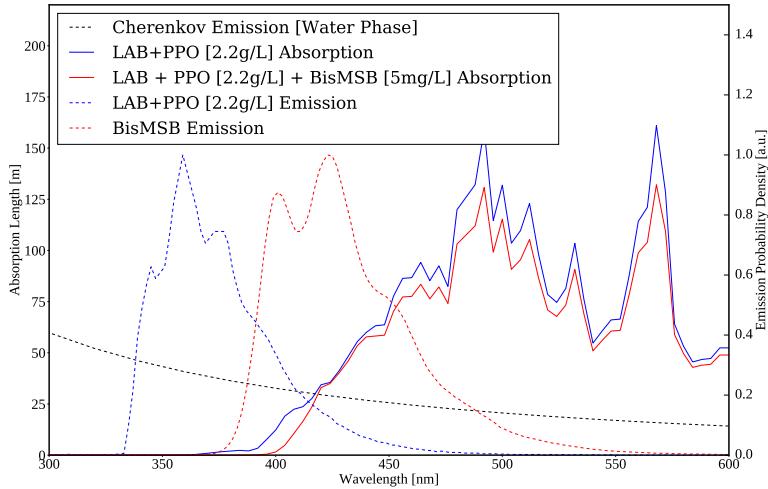


Fig. 2.2: Comparison of the SNO+ detector media’s emission and absorption properties, versus optical phase [13, 14].

as ‘luminescence’ or ‘fluorescence’). Although multiple varieties of scintillator exist, the one used in SNO+ is that of an organic liquid scintillator. For such liquids, scintillation light is generated from the de-excitation of delocalised electrons within carbon–carbon ‘ π -bonds’ [15]. A major example of these π -bonds are found in benzene rings, which are present in LAB, PPO, and bisMSB.

Because of this delocalised structure, excited atomic π -electrons can stay in what is typically the first-excited state for somewhat longer than typical excited states: lifetimes of $\mathcal{O}(10^{-9} \text{ s})$ as opposed to $\mathcal{O}(10^{-12} \text{ s})$. This is what gives scintillation light its characteristic ‘slow’ response relative to the instantaneous light generated by the Cherenkov process. Moreover, decays from this state can emit light typically in the optical-wavelength range. π -electrons can end up in the first-excited state either by direct excitation, or by ionisation followed by recombination. Because the ground state of these electrons are spin-singlet states, atomic spin selection rules [15] strongly prefer any direct excitations to stay in a spin-singlet state. As a result, so-called “inter-system crossing” from an excited singlet state to an excited triplet state is strongly suppressed.

| Component | A_i | τ_i [ns] |
|-----------|--------|---------------|
| 1 | 0.665 | 7.35 |
| 2 | 0.218 | 5.45 |
| 3 | 0.083 | 117.5 |
| 4 | 0.0346 | 425 |
| Rise | – | 0.8 |

Table 2.1: Current values used to model scintillator emission from electrons in 2.2 g/L LABPPO [16, 17].

However, ionised electrons that recombine have no such restriction, and so readily form excited triplet states. Once in such a state, the same spin selection rules strongly suppress the decay of these excited triplet electrons back down to the singlet ground state. This leads to scintillation light having, at the very least, a ‘fast’ and ‘slow’ time component. In SNO+, we currently model emission of scintillation light from LAB with 4 time components, following the timing distribution $f(t)$ given by:

$$f(t) = \sum_i A_i \left(\frac{e^{-t/\tau_i} - e^{-t/\tau_{\text{rise}}}}{\tau_i - \tau_{\text{rise}}} \right), \quad t > 0. \quad (2.2)$$

Here, A_i and τ_i correspond to the fraction of light emitted and decay constant for each component respectively, and τ_{rise} is a common rise time. The current values for these parameters used in simulations for the emission from electron tracks can be seen in Table 2.1. These were obtained by Rafael Hunt-Stokes through the fitting of tagged ^{214}Bi β -decay events within the detector with 2.2 g/L LABPPO [16]. A plot from R. Hunt-stokes showing this fit between data and simulation is shown in Fig. 2.3.

When using just a single scintillating compound, the very same energy levels that can generate scintillation light are those that can absorb it. This can be a problem for large-scale detectors like SNO+, which depend on scintillation light being unobstructed in its path to the PMTs. This problem can be addressed with

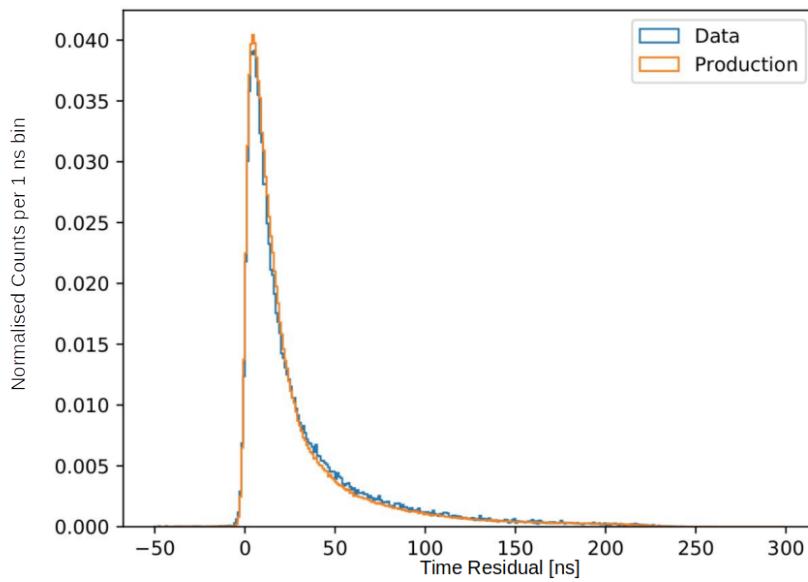


Fig. 2.3: Comparison between the observed emission time distribution of electrons from tagged ^{214}Bi β -decays and a production of matching simulated events, after fitting the timing constants. Taken from [16].

¹ the addition of another scintillating component, known (somewhat confusingly)

² as the primary fluor. In SNO+, this is the PPO added to the LAB.

³ When an LAB molecule is excited, that energy can be transferred to a PPO

⁴ molecule through what is known as a ‘non-radiative transfer’. In short, this transfer

⁵ of energy occurs not through the emission and absorption of optical photons, but

⁶ through the coupling of the molecules’ electric dipoles during a collision. When

⁷ the now-excited PPO molecule de-excites to emit scintillation light, the different

⁸ molecular structure it has generates a different emission spectrum to that of LAB.

⁹ These longer wavelengths of light are no longer able to be absorbed by the LAB,

¹⁰ allowing for a scintillator with less optical absorption.

¹¹ In SNO+ we plan on also adding in the compound BisMSB to the scintillator

¹² cocktail. This is a ‘wavelength-shifter’: scintillation light at short wavelengths is

¹³ absorbed, and then re-emitted at longer wavelengths, where the detection efficiency

¹⁴ of the PMTs is greatest (~ 420 nm). More on the properties of the PMTs in SNO+

¹⁵ can be found in Section 2.3.3. The net effect of the three scintillating components

¹⁶ within SNO+ can be seen in Fig. 2.2. Note how, as energy is transferred from one

scintillation component to another, the wavelength of light emitted gets necessarily longer as energy is lost to heat.

The light yield of a scintillator, i.e. the amount of optical photons generated per unit of energy deposited into the scintillator, is a function not just of the scintillator but also the incident particle’s ionisation strength. In particular, α particles are far more effective at exciting and ionising nearby atoms, and so can deposit far more of its energy into the scintillator per unit volume. However, the strength of this ionisation for α s can actually become at detriment to the generation of scintillation light. Empirically, scintillators follow to first order Birks’ Law for their scintillation light yield [18]:

$$\frac{dL}{dx} = S \frac{\frac{dE}{dx}}{1 + k_{\text{Birks}} \frac{dE}{dx}}, \quad (2.3)$$

where $\frac{dL}{dx}$ is the number of photons emitted per unit track length, $\frac{dE}{dx}$ is the energy loss of the incident particle per unit track length, S is the scintillator’s characteristic light yield constant, and k_{Birks} is the scintillator’s “Birks’ Constant”. For minimum-ionising particles such as a 6 MeV electron $\frac{dE}{dx} \approx 2 \text{ MeV cm}^{-1}$ [], meaning the denominator of this equation is close to 1, and so the amount of scintillation light generated is just $\frac{dL}{dx} \approx S \cdot \frac{dE}{dx}$. However, for α -particles generated in radioactive decays, this denominator can become substantial, and in the limiting case we have merely $\frac{dL}{dx} \approx \frac{S}{k_{\text{Birks}}}$. For example, α -particles are generated at 5.304 MeV from the decays of ^{210}Po nuclei [19]. However, in the 2.2 g/L LABPPO scintillator currently within SNO+, we observe that these events generate light equivalent to a 0.45 MeV event. For this scintillator, S and k_{Birks} are measured to be approximately 14,000 γ/MeV and $0.077 \text{ mm MeV}^{-1}$, respectively [20].

2.3.2 Optical Processes

Once optical-wavelength photons have been created within the detector, various processes can then occur that can hinder its path towards a PMT, and therefore modify the observed signal. This subsection covers the main optical processes, with a focus on Rayleigh scattering, as an understanding of this phenomenon is critical for Chapters 3–5.

7 Rayleigh Scattering

Optical scattering is the general process of how light is scattered by particles within a medium. This is fundamentally an electrodynamical process: an electromagnetic wave is incident on the set of particles within the medium, which induces these particles to oscillate within the field, and therefore generating their own electromagnetic radiation in response. Usually, this ‘scattered’ radiation has the same frequency as that of the incident radiation, and therefore the scattering is said to be *elastic*. It is possible under certain circumstances for this scattered radiation to be of a longer wavelength than the incident radiation: in which case, energy was absorbed by the particles and so the scattering was *inelastic*. However, this latter type of scattering, also known as Raman scattering, is not relevant for SNO+.

The general solution to elastic optical scattering was first described by Gustav Mie [21] and Ludvig Lorenz [22] in what is now known as *Mie Theory*. In this theory, it is assumed that a plane wave of wavelength λ is incident on a dielectric sphere of radius a . While the general solution to the problem of Mie scattering is somewhat complicated (if tractable), in certain regimes one can make further simplifying assumptions that substantially reduce the complexity of the result. In particular, if one assumes that the size of the particle is much smaller than the wavelength of light, and that any induced dipole moment can

actually be established in the time window allowed by the oscillation period of the electromagnetic field [23], then one can obtain *Rayleigh scattering*. This simpler case is so-called because of its initial formulation by Lord Rayleigh [24].

One can show that the differential cross-section associated with Rayleigh scattering of unpolarised light off a single particle, $\frac{d^2\sigma_{\text{Ray}}}{d\theta d\phi}(\theta, \phi)$, is given by [25]:

$$\frac{d^2\sigma_{\text{Ray}}}{d\theta d\phi}(\theta, \phi) = \frac{8\pi a^6}{\lambda^4} \left(\frac{n_{\text{par}}^2 - 1}{n_{\text{par}}^2 + 2} \right)^2 (1 + \cos^2 \theta). \quad (2.4)$$

Here, θ and ϕ correspond respectively to the polar and azimuthal angles of the scattered waves relative to the incoming wave, and n_{par} is the refractive index of the scattering particle. Most important to notice about this equation is that the cross-section follows a strong $1/\lambda^4$ dependence, meaning that short wavelengths of light will be scattered to far greater extents than that of longer wavelengths. Secondly, the light is not scattered isotropically, but according to a $1 + \cos^2 \theta$ dependence. This means that most light is either scattered directly forwards or backwards, and little gets scattered orthogonally to the direction of the incident light. This is useful when it comes to trying to measure scattering in the SNO+ detector, as it provides a handle upon which to distinguish scattered light from isotropically-emitted scintillation light.

Of course, we care about the scattering that occurs within an entire bulk medium, not just the scattering off of a single molecule. From a macroscopic perspective, the key quantity of interest is a material's *Rayleigh scattering length*, l_{Ray} : the mean distance a photon is expected to travel before Rayleigh scattering. One can show that, assuming the above differential scattering cross-section, the Rayleigh scattering length is given by [26]:

$$l_{\text{Ray}} = \left[\frac{16\pi}{3} R \right]^{-1}. \quad (2.5)$$

¹ R is the *Rayleigh ratio*, $R = \frac{1}{V} \frac{d^2\sigma_{\text{Ray}}(90^\circ)}{d\theta d\phi}$, where V is the volume taken up by
² one scattering particle within the medium. R is then equivalent to the power of
³ the scattered light per unit volume of the scattering medium per unit incident
⁴ intensity at $\theta = 90^\circ$.

⁵ This can lead to a few changes to Rayleigh scattering that are worth noting.
⁶ Firstly, unlike for a single particle, the electric polarisability of a material can be
⁷ *anisotropic*. Anisotropic materials have a modified angular dependence on their
⁸ differential cross-section, governed by the *depolarisation ratio*, δ . In particular, the
⁹ $(1 + \cos^2 \theta)$ dependence becomes $\left(1 + \frac{1-\delta}{1+\delta} \cos^2 \theta\right)$. For isotropic materials, $\delta = 0$,
¹⁰ and so the angular dependence reduces to the original form.

¹¹ Secondly, the above model has been shown to be insufficient to describe liquids
¹² or solids [27], because of the non-negligible strength of their inter-molecular forces.
¹³ Fortunately, Einstein [28], Smoluchowski [29], and Cabannes [30] developed a
¹⁴ theory for describing how photons can scatter off of the local charge density
¹⁵ fluctuations that naturally are present in a medium because of the thermal motion
¹⁶ of molecules. The theory shows that the Rayleigh ratio of a medium is related to
¹⁷ the medium's dielectric constant, ε , by:

$$\text{18} \quad R = \frac{\pi^2}{2\lambda^4} \left[\rho \left(\frac{\partial \varepsilon}{\partial \rho} \right)_T \right]^2 k_B T \kappa_T \frac{6 + 6\delta}{6 - 7\delta}, \quad (2.6)$$

¹⁹ where ρ is the density of the medium, $\left(\frac{\partial \varepsilon}{\partial \rho} \right)_T$ is the partial derivative of the dielectric
²⁰ constant with respect to a changing density assuming a constant temperature T ,
²¹ k_B is the Boltzmann Constant, and κ_T is the medium's isothermal compressibility.
²² This latter quantity is given by the rate of change of volume given a changing
²³ pressure of the medium, all at a constant temperature.

²⁴ Furthermore, the Eykman Equation [26, 31] has been shown to be an effective
²⁵ empirical formula relating how ε is impacted by density fluctuations to the

medium's refractive index, n_{med} :

$$\rho \left(\frac{\partial \varepsilon}{\partial \rho} \right)_T = \frac{(n_{\text{med}}^2 - 1)(2n_{\text{med}}^2 + 0.8n_{\text{med}})}{n_{\text{med}}^2 + 0.8n_{\text{med}} + 1}. \quad (2.7)$$

This leads to a final empirical formula for the Rayleigh scattering length:

$$l_{\text{Ray}} = \left[\frac{8\pi^3}{3\lambda^4} \left(\frac{(n_{\text{med}}^2 - 1)(2n_{\text{med}}^2 + 0.8n_{\text{med}})}{n_{\text{med}}^2 + 0.8n_{\text{med}} + 1} \right)^2 k_B T \kappa_T \frac{6 + 3\delta}{6 - 7\delta} \right]^{-1}. \quad (2.8)$$

In-situ measurements of the scattering of the UPW were made indirectly during SNO [32], and then subsequently in the water phase of SNO+ had a divisive scaling factor of $(1.28 \pm 0.05(\text{stat.}) \pm 0.14(\text{sys.}))$ determined by Esther Turner [33]. Ex-situ measurements of the Rayleigh scattering within LAB and LABPPO have also been made by groups in both the SNO+ and JUNO Collaborations [34, 35, 36, 37, 38], but no in-situ measurements have been made prior to this thesis. Fig. 2.4 shows the scattering lengths for UPW, LAB, and 2 g/L LABPPO from these measurements, with the lines showing what is currently being used in simulations for SNO+. Measurements of the scattering lengths in scintillator are a major focus of Chapters 3–5.

Absorption and Re-emission

In addition to scattering, an optical medium is also able to absorb light that propagates through it. For a given medium, the *absorption length* l_{abs} is analogous to l_{Ray} described above, and is typically strongly a function of wavelength. For most materials, absorbed light is forever lost, converted into heat. However, for the special case of scintillators, re-emission of absorbed light is possible: this is because of the physics described in Section 2.3.1.

Because both scattering and absorption impede a photon's ability to propagate through a medium directly, it is often possible to measure their combined impact

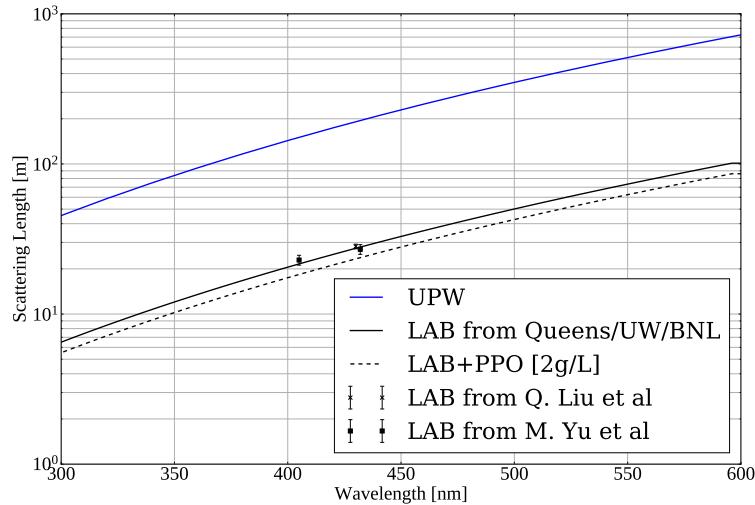


Fig. 2.4: Scattering lengths used in simulation for UPW, LAB, and 2 g/L LABPPO. The UPW shape is taken from indirect in-situ measurements in [32] with an additional divisive scaling factor from [33]. The LAB shape is taken from ex-situ measurements by SNO+ members at Queen’s University, University of Washington, and Brookhaven National Laboratory [34, 35]. An additional divisive scaling factor of 1.176 due to PPO was made by [36]. For comparison, measurements by members of the JUNO Collaboration for LAB are also shown [37, 38].

1 through what is known as the attenuation/extinction length, l_{ext} :

$$2 \quad \frac{1}{l_{\text{ext}}} = \frac{1}{l_{\text{abs}}} + \frac{1}{l_{\text{Ray}}}. \quad (2.9)$$

3 In the water phase, the ‘Laserball’ calibration system was used by Ana Sofia Inácio
4 to measure various optical properties of the detector, including the extinction
5 lengths of the UPW and acrylic as a function of wavelength [3, 39]. Using the
6 water phase scattering measurements made by E. Turner, Eq. 2.9 allowed for the
7 estimation of the absorption lengths of these two materials, shown in Figure 2.5.
8 Measurements of the extinction length in the scintillator phase is discussed in
9 detail in Chapter 5.

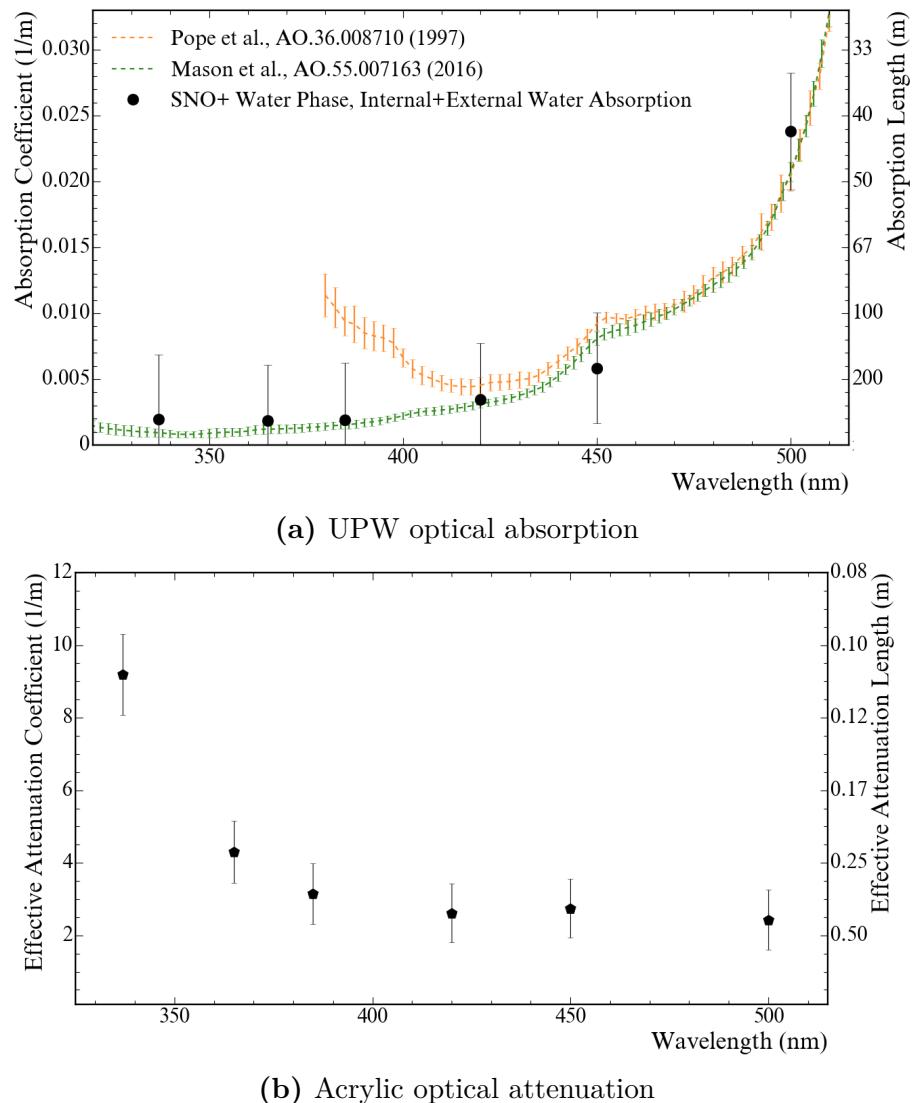


Fig. 2.5: Properties of the UPW and acrylic in the water phase, measured by A. S. Inácio in [3, 39].

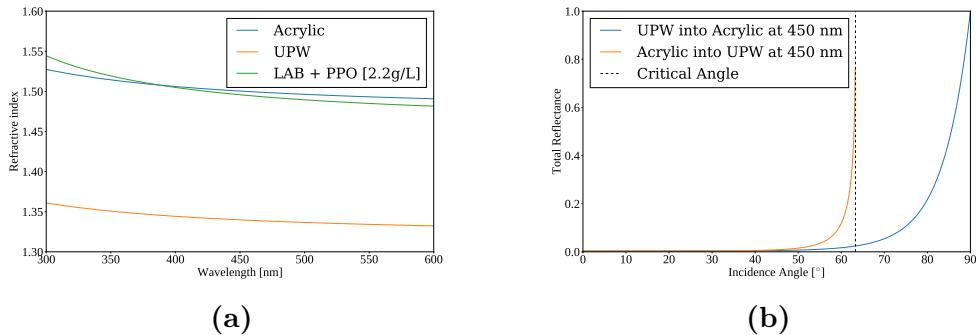


Fig. 2.6: (a): Refractive indices of acrylic, UPW, and LABPPO as a function of wavelength. The values for acrylic and UPW come from model fits to data made in SNO [32, 40], whereas those for LABPPO come from data taken in [41]. (b): Reflectance of an unpolarised beam of light at 450 nm going between UPW and acrylic.

1 Surface reflection and refraction

2 When light travels through the boundary of one medium to another, both reflection
 3 and refraction can be possible, depending on the relative refractive indices of
 4 the two media as well as the angle of incidence. The refractive indices of the
 5 UPW, acrylic, and LABPPO are shown as a function of wavelength in Figure 2.6a.
 6 Note that, for most optical wavelengths, LABPPO has a very close refractive
 7 index to acrylic, whereas UPW is somewhat farther away. By consequence,
 8 negligible refraction is expected in most cases for light travelling between the
 9 liquid scintillator and the acrylic; however, substantial refraction and reflection are
 10 possible for light travelling between acrylic and UPW. Because of this, isotropically-
 11 emitting point-like physics events within the AV that are close enough to the
 12 acrylic will have some of their light undergo Total Internal Reflection (TIR) at
 13 the AV, reflecting back into the AV instead of continuing outward into the outer
 14 water.

15 Even when not undergoing TIR, some light at a boundary will still reflect.
 16 The fraction of light that reflects is known as the *reflectance* R , compared to that
 17 which is able to transmit through the boundary, the *transmittance* $T = 1 - R$.

The *Fresnel Equations* determine the reflectance of an interface [42]:

$$R_s = \left| \frac{n_1 \cos \theta_i - n_2 \cos \theta_t}{n_1 \cos \theta_i + n_2 \cos \theta_t} \right|^2, R_p = \left| \frac{n_1 \cos \theta_t - n_2 \cos \theta_i}{n_1 \cos \theta_t + n_2 \cos \theta_i} \right|^2, \quad (2.10)$$

where R_s and R_p are the reflectances of s - and p -polarised light, n_1 and n_2 are the refractive indices of the first and second optical media, and θ_i and θ_t are the angles of incidence and refraction, respectively. For SNO+, we are only interested in unpolarised light, so the total reflectance $R = (R_s + R_p) / 2$.

The total reflectance going from UPW into acrylic, as well as from acrylic into UPW, for an unpolarised beam of light with wavelength 450 nm is shown in Fig. 2.6b. For the latter case, the critical angle at which TIR occurs is clear.

2.3.3 Detection by PMTs

The final step for photons in our journey is detection by a PMT. Almost all PMTs in SNO+ are of the Hamamatsu R1408 design [1]. These PMTs within SNO+ are housed within an 18-segment reflecting Winston cone known as a ‘concentrator’. The combined PMT–concentrator ‘bucket’, shown in Fig. 2.7a, is designed to maximise the collection efficiency of light emanating from within the AV, whilst minimising the collection efficiency of light outside the AV [43]. The so-called ‘angular response’ of the PMT buckets has been measured in both SNO and SNO+ using the Laserball, which describes the relative collection efficiency as a function of the polar angle of the incident light ray relative to the direction in which the PMT bucket points. The results of this can be seen in Fig. 2.7b.

Once a photon is incident on the PMT’s photocathode, it is possible for that photon to be absorbed and generate a photoelectron. The probability of this happening is governed by the photocathode’s Quantum Efficiency (QE) at the photon’s wavelength, as well as the collection efficiency of a photoelectron onto the first dynode of the PMT. The combined measured efficiency of PMTs tested

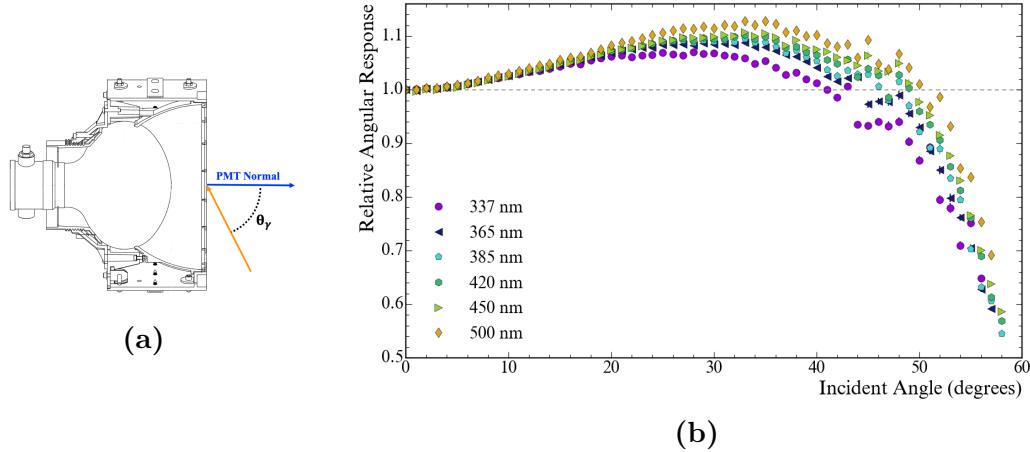


Fig. 2.7: (a): Diagram of the PMT and concentrator ‘bucket’ used within SNO+, showing also the definition of the incidence angle. (b): Plot of the measured relative angular response of the PMTs in SNO+, as a function of both incidence angle and wavelength. Both figures taken from [3].

¹ *ex-situ* for SNO can be seen in Fig 2.8. Once this photoelectron has been created,
² the dynodes within the PMT generate a cascade of electrons that produce an
³ observable voltage signal. The dynamics of this cascade are such that there is
⁴ a natural spread of possible times between the creation of a photoelectron and
⁵ the generation of the voltage pulse in the PMT’s anode. This is known as the
⁶ ‘Transit Time Spread’ (TTS) of the PMTs: for SNO+, the RMS of the TTS for
⁷ the R1408-type PMTs is ~ 1.7 ns [1].

⁸ Finally, if multiple photons generate photoelectrons on the same PMT close
⁹ enough in time, the amount of charge generated increases in proportion to the
¹⁰ number of photoelectrons (npe). Much like with the transit time, the strength
¹¹ of the signal observed by the PMT is governed by a distribution, a function
¹² of the npe generated. Examples of these distributions can be seen in Fig. 2.9.
¹³ The relatively large widths of these charge distributions precludes the ability to
¹⁴ straightforwardly determine the npe purely from charge when the npe is small.
¹⁵ To work around this, various techniques can be employed to try and estimate the
¹⁶ npe in a given PMT — an example of one such method can be seen in Section 4.2.

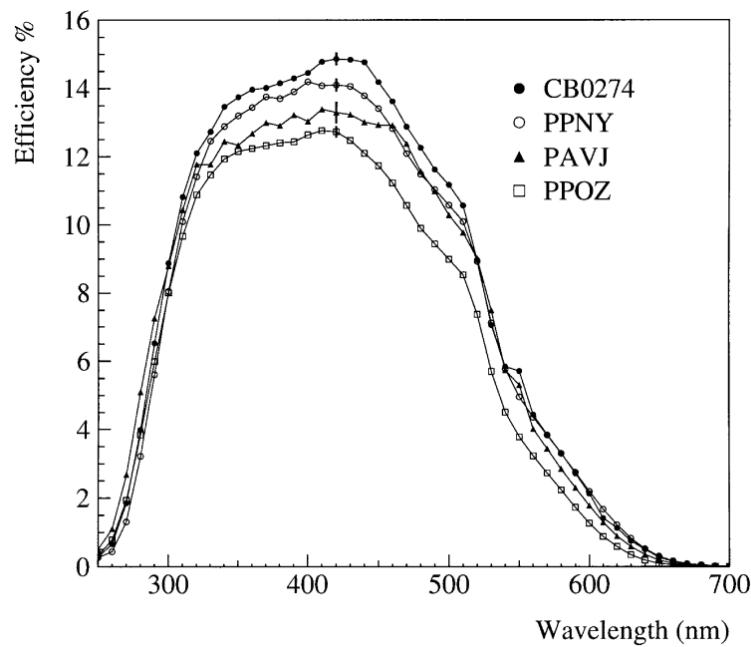


Fig. 2.8: Efficiencies of four R1408-type PMTs tested for calibration by [44].

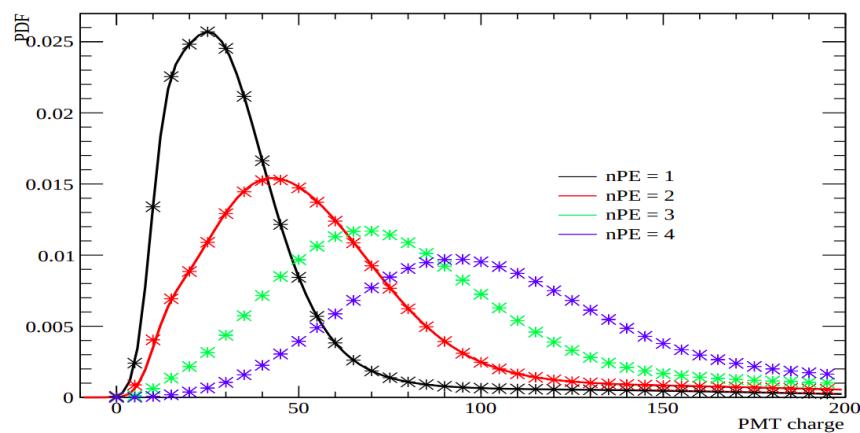


Fig. 2.9: Example charge spectra for a PMT as a function of the true npe generated. Figure taken from [45].

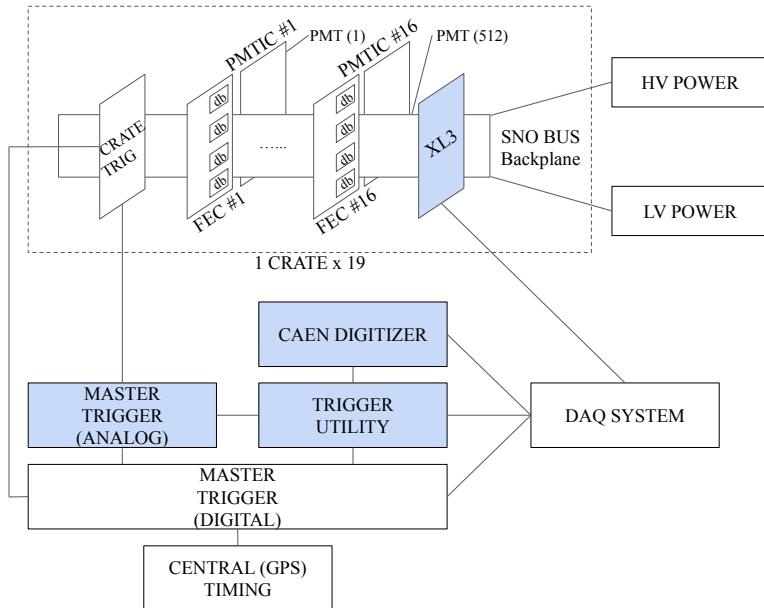


Fig. 2.10: Schematic of the front-end electronics used for data acquisition and triggering in SNO+, taken from and discussed in [2].

2.3.4 Data Acquisition and Triggering

Once a signal reaches the cable attached to a PMT, it travels along up to the front-end electronics on the deck above the detector. The job of these electronics, known as the data acquisition (DAQ) and triggering system, is to convert raw electronic signals from the PMTs into recorded digital ‘events’ that can be used for analysis. A schematic showing the setup of the electronics is shown in Fig 2.10, with full details in [2].

A signal passes first through the PMT Interface Card (PMTIC), which then sends it through to one of the Daughter Boards (DBs) which are stored on Front-End Cards (FECs) within one of 19 electronic crates on deck. The DBs determine if the analogue signal along a given PMT channel has crossed a pre-defined charge threshold, at which point we say that we have detected a ‘hit’ on that PMT’s channel. When this occurs, the DB performs a set of important actions:

1. Begins a timer for that channel, in the form of a Time-to-Amplitude Con-
2 verter (TAC). TAC also corresponds to the resulting analogue voltage on
3 the output of the TAC being measured.

4. Begins integrating the total charge signal for that channel in three ways,
5 known as QHS, QHL, and QLX. These correspond to using different inte-
6 gration times and gain settings.

7. Generates trigger pulses for that channel for each available trigger type.
8 Three main trigger signals are the ‘N20’ (a square pulse for 20 ns), ‘N100’
9 (a square pulse for 100 ns), and ‘ESUMHI’ (a pulse copying the shape of
10 the voltage signal for that channel). The reason for these names will be
11 explained shortly.

Whilst the TAC, QHS, QHL, and QLX are being calculated, the trigger signals
from each channel are sent over to the Crate’s Trigger Card (CTC), where the
signals are then summed for each trigger type. These crate-level trigger signals are
then sent over to the 7 detector-level Analogue Master Trigger Cards (MTC/A+),
which further sum the signals by trigger type from all the crates in the detector.
It is at this point of the process where the names of the trigger types becomes
clear: the combined N20 and N100 signals are proportional to the total number
of hit PMTs within a 20 ns and 100 ns time window, respectively, whilst the total
ESUMHI signal corresponds to the total charge seen over all the PMTs.

If these trigger signals go above certain pre-defined thresholds, then a signal
is sent for that trigger type to the Digital Master Trigger Card (MTC/D). The
MTC/D receives all trigger signals from the detector, and if a given trigger type
has been ‘masked in’ (i.e. activated) the Card will generate a Global Trigger
(GT) for the detector with a time stamp from its 50 MHz clock. Under certain
circumstances, such as calibrations, a trigger signal can be generated externally
and asynchronous to the MTC/D: these are ‘EXTA’ triggers. Any such EXTA

¹ trigger signal is first handled by an electronics box named ‘TUBii’ (Trigger Utility
² Board Mark ii), before then being passed onto the MTC/D. TUBii functionality
³ will become relevant when discussing the calibration electronics described in
⁴ Chapter 3.

⁵ Once a GT signal is generated, it is then sent back to all the CTCs, which
⁶ then orders the integration of time and charge to be stopped on all channels for
⁷ that crate. The time and charge information that has been temporarily stored
⁸ on each channel’s CMOS chip on the FEC is then sent to the crate’s ‘XL3’ card,
⁹ which packages the crate’s raw information via ethernet over to a set of computers.
¹⁰ The trigger signals for a triggered event are also digitised by a CAEN brand
¹¹ Analogue-to-Digital Converter (ADC), and sent to the same DAQ computers. The
¹² total window of time in which data is gathered from one GT signal is 400 ns, with
¹³ data from up to 180 ns before and 220 ns after the GT has arrived. There is then
¹⁴ a necessary ‘dead time’ of 420 ns after a given GT has been given in which no
¹⁵ further GTs can be made.

¹⁶ Finally, the raw data from the crates and trigger system arrives in a set of
¹⁷ computers, which organise all of this into an individually-packaged ‘event’, stored
¹⁸ on disk in the ‘Zebra Database’ (ZDAB) format. A given built event contains the
¹⁹ TAC and charge information from each hit PMT, the CAEN digitised waveforms,
²⁰ a unique identifying number for that triggered event (the GTID), as well as the
²¹ times from both the MTC/D’s 50 MHz clock and a GPS-calibrated 10 MHz clock.
²² The former time is used for measuring relative times between events whilst the
²³ latter is used for knowing the time of day of an event: both are used in Chapter 6.

²⁴ **2.3.5 Operation of the Detector**

²⁵ Control of the detector’s DAQ system is handled through a custom-built GUI
²⁶ known as ORCA [46]. This program allows operators of the detector to modify

settings in the detector electronics at both a low- and high-level. It also allows operators to monitor the current status of the detector, such as voltage and current levels within each crate.

ORCA allows for the detector to have its data split into ‘runs’ of different types. For the majority of the time, the detector is run in the ‘Physics’ mode, with individual runs split into 1 hour periods. It is this data that is used for almost all high-level physics analyses, such as the one described in Chapter 6. To help with the movement and processing of data, runs of raw data are split into ZDAB files of maximum size 1 GB. Because of this, the number of files generated per run is proportional to the trigger rate of the detector. In the current 2.2 g/L LABPPO scintillator phase under nominal conditions, the trigger rate is ~ 2.5 kHz, leading to 15 ZDAB files being generated per run of 1 hour in length.

Other detector run types include ones for detector maintenance, as well as for calibrations of various kinds. During certain calibration runs, data can be further split into ‘subruns’ where necessary. This allows for data taken from a given calibration source to have the different settings used (e.g. different wavelength settings) all kept within one run, but still appropriately separated. Operation of specific calibration sources, including the one described in Chapter 3, can be performed through the ORCA GUI.

2.4 Calibrations and Detector Modelling

Once the raw data from triggered events has been stored in files, certain extra steps must be taken before effective analysis of that data can be achieved. This section covers those steps.

¹ 2.4.1 Detector Monitoring

² No data taken from the detector can reasonably be used for analysis unless its
³ quality has been approved. This is done in a number of ways on SNO+. Firstly,
⁴ a number of automated systems monitor all aspects of the detector, including
⁵ voltage levels in the crates, trigger rates, as well as ‘slower’ quantities such as
⁶ the tensions on the ropes holding the AV in place. Problems in any of these
⁷ measured parameters trigger an automatic alarm system, which notifies a human
⁸ detector operator. A human detector operator monitors the detector 24/7 whilst
⁹ the detector is live.

¹⁰ In addition to systems that monitor whether anything has gone wrong, infor-
¹¹ mation about the state of the detector during each run is stored in a database
¹² known as RATDB. This information includes, amongst other things, a recording of
¹³ which PMT channels have actually been raised to high voltage for that run, as
¹⁴ well as any channels/cards/crates that have been flagged for having a known poor
¹⁵ data quality (e.g. being overly noisy).

¹⁶ 2.4.2 Electronic and PMT Calibrations

¹⁷ The lowest level of calibrations performed in SNO+ are the Electronic and PMT
¹⁸ Calibrations: ECAs and PCAs, respectively. These calibrations convert the raw
¹⁹ time and charge values recorded by the DAQ into quantities that can actually be
²⁰ used in analysis.

²¹ During an ECA, two main quantities are measured. Firstly, because of noise
²² the integrated charge measured on each channel is offset by some amount. This
²³ offset, known as the ‘pedestal’, is recorded for each channel. The other quantity
²⁴ is the ‘time slope’ for each channel, which allows one to convert from the ADC
²⁵ TAC counts into an uncalibrated hit time of that channel’s PMT in ns. Both
²⁶ of these quantities are measured by sending external signals to channels in the

crates, forcing them to start measuring TAC and charge even though no PMTs were actually hit. Running ECAs also enable us to spot any channels with unusual behaviour, so that they are not used during analysis. ECAs are typically done on a fortnightly basis, or after maintenance to the DAQ system has been performed.

Using ECAs alone is not enough to have fully-calibrated time and charge data. The lengths of cables between PMTs and PMTICs are all slightly different, leading to differences in the so-called ‘cable delay’ of each channel. This means that two PMTs that have a photoelectron generated at the same time can generate slightly different TAC values. Furthermore, because the start time of the TAC is determined by when the channel’s signal goes above a constant threshold, if a signal is very large (e.g. when numerous photoelectrons have been generated on one PMT) then the start time of the TAC will be systematically earlier. This is known as the ‘time walk’. Both of these quantities get measured during PCAs.

PCAs can be performed by either the Laserball or by the TELLIE calibration system. The latter is a series of 92 optical fibres attached at various points of the PSUP, through which optical-wavelength light can be fired from LEDs. TELLIE is the Timing subsystem of the ELLIE calibration system: the Embedded LED/Laser Light Injection Entity. The other two fibre-based optical calibration subsystems, AMELLIE and SMELLIE, are introduced in Section 2.4.3. For both the Laserball and TELLIE calibration systems, the cable delay and time walk are measured by firing light from the source at a known time and with a known hit occupancy, and observing when the signal arrives in each PMT channel.

On top of calibrating the PMT hit times, PCAs also further calibrate the charge information. In particular, the charge spectrum generated by a single photoelectron is determined for each channel. This allows us to convert the pedestal-corrected charge ADC counts into an approximate number of photoelectrons.

Using the data gathered from both ECAs and PCAs, the raw data stored in ZDABs is processed into a new file format known as RATDS files. These files

¹ contain all the information of an event, but now the timing and charge information
² have been calibrated. It is this file type used in the optical calibration work of
³ Chapters 3–5.

⁴ 2.4.3 Energy and Optical Calibrations

⁵ The next stage of calibrating the detector is modelling its optical properties. These
⁶ properties include all the processes covered in Section 2.3.2, such as scintillator
⁷ emission, optical absorption, re-emission, and Rayleigh scattering. This is crucial,
⁸ as it allows us to reconstruct information about events within the detector: more
⁹ on event reconstruction shortly.

¹⁰ In addition to deployments of the Laserball (discussed in Section 2.3.2), two
¹¹ further calibration sources are used in SNO+ to measure properties of light
¹² propagation: AMELLIE and SMELLIE. These are the ‘Attenuation Module’ and
¹³ ‘Scattering Module’ for the ELLIE calibration system. Like TELLIE, AMELLIE
¹⁴ and SMELLIE consist of optical light sources that shine through optical fibres
¹⁵ into the detector. The former uses LEDs from TELLIE, whilst the latter uses
¹⁶ optical wavelength lasers. Despite the names both subsystems are similar enough
¹⁷ that they are both capable of measuring attenuation and scattering within the
¹⁸ detector. More details about the SMELLIE hardware can be read in Chapter 3.

¹⁹ Another critical component of the detector to calibrate well is the energy
²⁰ response: given a specific amount of energy deposited in the water/scintillator,
²¹ how many hits are observed? For this, a number of radioactive sources are used at
²² a variety of energies. In the scintillator phase, there are three main sources. The
²³ first is an americium-beryllium (AmBe) source inherited from SNO [47], which
²⁴ contains ^{241}Am that α -decays, which can be captured by the ^9Be within the source.
²⁵ This capture leads to the emission of a neutron as well as production of a ^{12}C
²⁶ nucleus, which 60% of the time is in an excited state. When this excited state

decays, a 4.4 MeV γ is emitted promptly. Eventually the neutron is captured by hydrogen in the detector, leading to a characteristic 2.2 MeV delayed γ being generated [2]. Both the prompt and delayed peak energies from the AmBe source can be used for energy calibration. In addition to this, one can calibrate the neutron detection efficiency with the AmBe source [47], which is important for the analysis of antineutrino IBD events.

Another deployable radioactive source is the ^{16}N source, also originally used for SNO [48]. The ^{16}N isotope β -decays to ^{16}O , with a distinctive 6.1 MeV γ also being generated 66% of the time. It is the γ that can make it out to the detector, whilst the β can be tagged by a block of scintillator and PMT held within the source container.

In theory, sources can be deployed both within the AV ('internally') and in the water shielding ('externally'). Both internal and external deployments were used for the optical calibration of the water phase with the Laserball [3]. Although external deployments during the scintillator phase with the Laserball, N16, and AmBe sources have been made, there have been no such internal source deployments. This is a result of the substantial work currently underway [49] to overhaul the internal deployment hardware, in order to satisfy the much more stringent radiopurity requirements of the scintillator phase over the water phase.

Alongside these deployed sources, a different kind of radioactive source has been used during the scintillator phase for energy calibration: the existing radioactive background spectra within the detector. Backgrounds such as ^{14}C , ^{210}Po , $^{214}\text{BiPo}$, and ^{208}Tl all have distinctive peaks in the energy spectrum of SNO+ during the scintillator phase, and can be used to calibrate the scintillator's energy response. Using $^{214}\text{BiPo}$ events in particular has been used for energy scale calibration with the solar oscillation analysis, as discussed in Section 6.2.3.

¹ 2.4.4 Event Reconstruction

² Once the detector has been calibrated, event ‘reconstruction’ becomes possible.
³ This is the process of deriving high-level physics quantities about a triggered event
⁴ within the detector, based upon the calibrated hit information. In SNO+, our
⁵ base assumption in most cases of event reconstruction is that a triggered event was
⁶ due to a single electron track. Reconstructing an event involves running a number
⁷ of algorithms, which in the scintillator phase are together called the **ScintFitter**.

⁸ The first critical pieces of information that gets determined by **ScintFitter**
⁹ is the event’s position and time. The reconstructed position corresponds to the
¹⁰ point in the detector where the triggered event most likely came from (assuming
¹¹ the event was approximately point-like in extent), whilst the reconstructed time
¹² is the starting emission time of the event, relative to the event’s trigger time. The
¹³ position of an event is critical to know, as far fewer background events occur near
¹⁴ the centre of the detector compared to the edges. It is also important to know the
¹⁵ emission time of an event, as this allows us to build the so-called ‘time residual’
¹⁶ (t_{res}) distribution of an event. For a point-like physics event in the detector, t_{res}
¹⁷ for a given PMT hit is defined as:

$$\text{¹⁸ } t_{res} = t_{\text{hit}} - t_{\text{TOF}} - t_{\text{emm}}, \quad (2.11)$$

¹⁹ where t_{hit} is the calibrated hit time of the PMT, t_{TOF} is the time one expects
²⁰ for light to travel directly from the reconstructed position to that PMT (the
²¹ time-of-flight), and t_{emm} is the reconstructed emission time.

²² Whilst a number of algorithms have been developed for reconstructing position
²³ and time in SNO+, they all work on the same basic principle. Because of the
²⁴ spherical symmetry of the detector, if an event occurs at the centre of the detector

one expects direct light to hit PMTs throughout the detector at the same time¹.
 However, if an event happens some distance away from the detector’s centre then
 direct light will arrive at the PMTs it is closer to sooner. Therefore, by looking
 at the distribution of hit times for PMTs that were hit earliest as a function of
 the PMTs’ positions in the detector (ignoring PMT hits that arrived much later,
 presumably because the photon paths were not direct) one can try and estimate
 where the position of the event was. Reconstructed positions and reconstructed
 times are linked by the time residual equation described above.

Currently on SNO+, a likelihood-based approach is used to reconstruct position
 and time. The algorithm endeavours to maximise the combined likelihood of
 the observed calibrated hit times of the hit PMTs, given proposed points in the
 four-dimensional (position, time) parameter space [50, 51, 52]. However, regardless
 of algorithm there are two factors that limit the position and timing reconstruction
 of an event. The TTS of the PMTs used as well as the speed of the scintillator
 emission timing defines the fundamental timescale — and hence also length scale
 — by which events can be reconstructed. Secondly, if more photons are able to
 generate prompt hits in the detector from a given event, then more information
 can be used to determine the position and time. Under current conditions, a
 2.5 MeV event in the centre of the detector will have a position resolution of
 100 mm [53].

The other critical piece of reconstructed event information is the event’s energy.
 More precisely, this is the kinetic energy of an electron that has been assumed
 to have generated the event. By consequence, events due to α -decay will obtain
 reconstructed energies well below the actual energy of the α particle, because of
 scintillator quenching (see Section 2.3.1 for details).

¹It is possible for the centre of the AV and PSUP to not be completely aligned, i.e. there is an ‘AV offset’. Then, the spherical symmetry can be very slightly broken by refraction through the AV. Fortunately, we account for this when coordinating our position fitters.

At its simplest, assuming that an event is from an electron of moderate energy (e.g. at 2.5 MeV), then we can expect the number of PMT hits observed to be directly proportional to the energy of the event. Given that the number of hits observed in an event (called the `nhit`) is governed by a Poisson distribution, then the uncertainty in energy will just be proportional to the square root of the number of hits. As a result, the reconstructed energy resolution in SNO+ is determined by the scintillator’s light yield and absorption length, as well as the coverage and QE of the PMTs.

There are second-order corrections to the energy reconstruction that need to be considered to minimise bias. At low energies, scintillator quenching becomes non-negligible, so an understanding of the scintillator’s Birks’ constant is needed. At high energies, many PMTs will have had multiple photoelectrons generated, so merely using the `nhit` will give an underestimate of the true energy. Finally, the detection efficiency of photons is non-uniform as a function of position in the detector. The current energy reconstruction algorithm used within SNO+ attempts to deal with all of these effects [54, 55].

After position, time, and energy reconstruction, **ScintFitter** calculates a number of additional quantities from what are known as classifiers. These describe a wide number of properties about an event, often using the derived t_{res} distribution of an event as the basis for classification. Some examples of classifiers used in analysis are discussed in Section 6.1.2.

All Physics data runs, as well as certain calibration data such as AmBe and ^{16}N , has the **ScintFitter** algorithm run over it after having been processed for time and charge calibration. This results in what is known as a fully-processed RATDS file, as well as a new file type known as an `ntuple`. This latter file type has much of the hit-level information removed, and contains only event-level information such as the reconstructed energy and position. Because these files are

much smaller, they are the ones typically used in the high-level physics analyses
on the experiment, such as the one described in Chapter 6.

2.4.5 Event Simulation

Simulations of events in SNO+ are performed using the software **RAT** [2]. Built
on the **GEANT4** particle physics software framework, **RAT** is capable of simulating
all aspects of the physics of an event within the detector via a Monte Carlo (MC)
approach. This includes any particle physics that defines an event’s generation,
propagation and interactions of those particles in the detector media, the generation
of light by both scintillation and Cherenkov processes, the propagation of that light,
as well as the detection of that light by PMTs and simulation of the expected DAQ
response. **RAT** is then used to process both simulation and data in the same way.
In addition to being highly customisable, **RAT** can use the **RATDB** tables generated
from a given data run when simulating to try and match those particular run
conditions as closely as possible. **RAT** also offers a suite of tools to assist with
analysis of data. The software is constantly being updated with new features —
the work done in this thesis uses **RAT** versions between 6.18.8 and 7.0.9, inclusive.

Chapter 3

1

The SMELLIE Calibration

2

System

3

There's a certain Slant of light,

Winter Afternoons —

That oppresses, like the Heft

4

Of Cathedral Tunes —

EMILY DICKINSON

As mentioned in Section 2.4.3, one of the principal systems for calibrating the optics of the SNO+ detector is SMELLIE. This calibration device consists of 5 different optical wavelength lasers able to be fired through 15 optical fibres, whose endpoints are attached to the PSUP. A collimator is attached at the end of each fibre, ensuring the emitted light forms a narrow beam across the detector. A diagram of SMELLIE in the detector is shown in Fig. 3.1.

5

6

7

8

9

10

The primary goal of SMELLIE is the measurement and monitoring of optical scattering within the detector over the lifetime of the experiment. By firing light from SMELLIE into the detector, some fraction of the photons will be scattered by the detector medium, a fraction of those scattering at large angles relative to

11

12

13

14

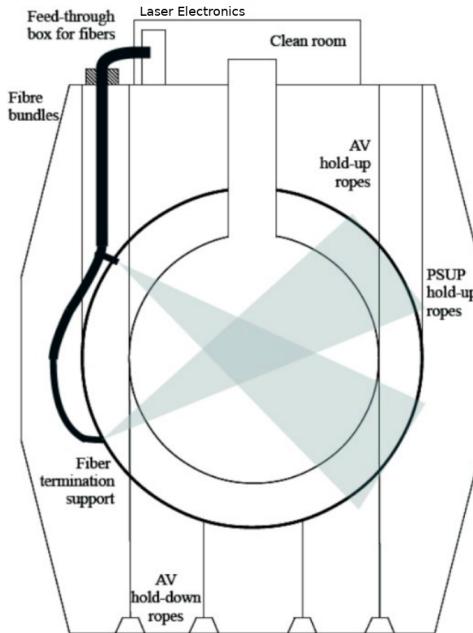


Fig. 3.1: Diagram of the SMELLIE calibration system within SNO+. Modified from [56].

the direction of the SMELLIE beam. This strongly scattered light can be detected by PMTs far from the ‘beamspot’, and will also arrive substantially later than light which travelled directly from the fibre to those PMTs. By isolating this scattered light signal, and comparing the quantity observed in data to equivalent simulations with varying scattering lengths, in principle one can measure the detector medium’s scattering length. If one takes SMELLIE data with various wavelengths of light at various points in time, we can get a dynamical picture of the optical scattering in SNO+. An analysis of optical scattering in the scintillator phase is made in Section 5.2.

Another substantial measurement that can be made with SMELLIE is the extinction length of the detection media as a function of wavelength and time. This can be done by comparing the fraction of light emitted by the fibre that gets observed in the beamspot after passing through the detector. Section 5.1 covers this analysis in the scintillator phase. Once measurements of both the scattering

length and extinction length have been made, it is then possible to derive the
absorption length from Eq. 2.9.

Why is measuring the scattering and absorption lengths of the detector medium
important? Both optical characteristics of the detector impact the propagation of
light from physics events, and hence which PMTs get hit along with the times
of those hits. If these lengths are systematically off within simulation, this can
lead to negative consequences for reconstructing events. In particular, for the
scintillator phase, if there is more optical absorption occurring than expected, then
because not all absorbed light is re-emitted a larger fraction of photons are lost.
Therefore, because energy reconstruction is strongly dependent on the number
of PMT hits observed in an event, the energy of events will be systematically
under-estimated. Alongside this, light that is re-emitted will only do so after
some time delay, and the direction of this re-emission unlikely to be in the same
direction as before. This leads to systematic changes in the observed time residual
distributions, impacting position reconstruction, as well as any classifiers that use
the time residual distribution.

If there is more optical scattering than expected within the scintillator, this
also indirectly leads to a greater loss of light because of the increased path length
that a photon will typically have to travel before being detected. By consequence,
there will be a second-order impact on the energy reconstruction from systematics
in the scattering length. Much like changes in the quantity of re-emitted light,
increasing the amount of scattering will also systematically effect the position
reconstruction and many classifiers.

This chapter covers the hardware used to take SMELLIE data, as well as the
steps taken to ensure high quality data was taken throughout the scintillator
phase of the experiment.

Fig. 3.2

¹ 3.1 The SMELLIE Hardware

² A full description of the initial hardware setup that was used during the air fill
³ and early parts of the water fill phase can be read in [1]. Since then, a series of
⁴ hardware upgrades have been made, with [2] covering the hardware status used in
⁵ data taken throughout the water phase. Fig. 3.2 shows a timeline of the hardware
⁶ upgrades as well as the data taking campaigns performed throughout the current
⁷ lifetime of the experiment. The current layout of the SMELLIE hardware, showing
⁸ the connections between each of the devices within the system, can be seen in
⁹ Fig. 3.3. The rest of this section briefly summarises the current contents of the
¹⁰ calibration system, along with descriptions of the major hardware changes made
¹¹ since the water phase.

¹² 3.1.1 Lasers

¹³ Fundamental to the SMELLIE calibration system are 5 optical-wavelength lasers.
¹⁴ Four of these are fixed-wavelength pulsed-diode laser heads from the company
¹⁵ PicoQuant. These ‘PQ’ laser heads each emit with a different narrow wavelength
¹⁶ spectrum, peaking at 375 nm, 407 nm, 446 nm, and 495 nm. These are referred
¹⁷ to as the PQ375, PQ407, PQ446, and PQ495 lasers, respectively. In addition to
¹⁸ these lasers, a SuperK Compact laser made by NKT Photonics (hereafter referred
¹⁹ to as the SuperK laser) is also used¹. Unlike the PQ laser heads, the SuperK
²⁰ is a super-continuum laser able to produce laser light over the whole optical
²¹ wavelength spectrum. Because we are almost always interested in determining
²² optical properties at specific wavelengths, a variable bandpass filter also built
²³ by NKT Photonics known as the SuperK Varia has been included. This allows

¹ Apologies to those more familiar with ‘SuperK’ referring to the SuperKamiokande experiment based in Japan: this laser bears no relation.

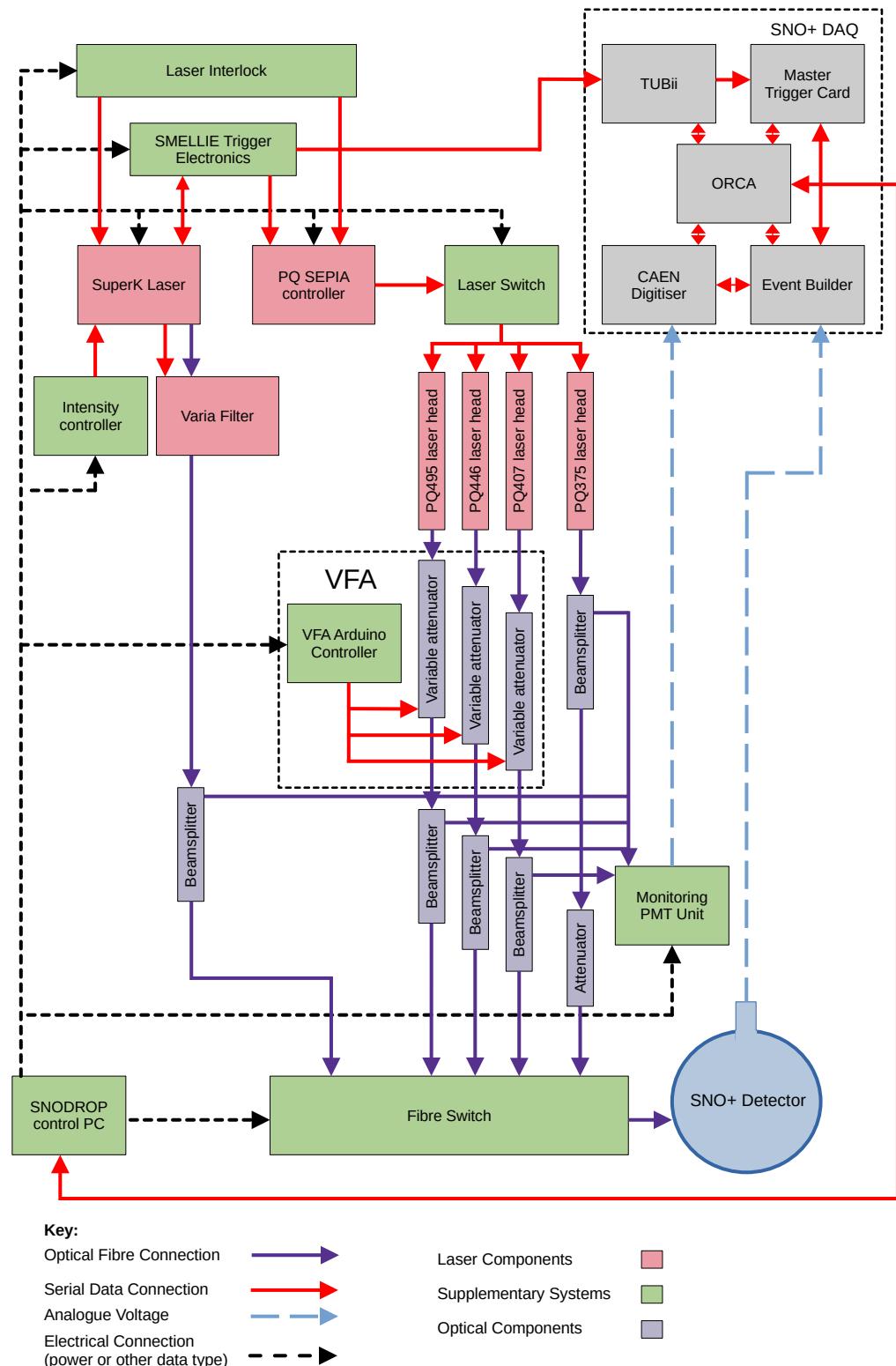


Fig. 3.3: Diagram of the connections between the various bits of SMELLIE hardware and the rest of the SNO+ detector, after the changes made in the Summer of 2022. Adapted from [33].

the user to select any wavelength interval between 400–700 nm, with a minimum bandwidth of 10 nm. The wavelength and emission timing characteristics of all five lasers are shown in Fig. 3.4. The $\mathcal{O}(1\text{ ns})$ time widths for the lasers are essential for scattering analyses as they allow for much greater precision in knowing when photons should arrive at PMTs in the detector than with an LED source. This is why SMELLIE uses lasers for its light generation, unlike the LEDs used for AMELLIE and TELLIE.

3.1.2 Controlling Laser Intensities

It is important to be able to control the quantity of light that enters the detector from a given pulse of one of the lasers. Too much light, and because of the detector’s DAQ analysis of data becomes difficult, and in the extreme case permanent damage could be done. Too little, and insufficient statistics are collected by the detector to actually do any analysis. Controlling the pulse intensity of light into the detector is done in two parts. For the SuperK laser, the raw power of the beam in a pulse can be set as a percentage of the maximum possible power for that wavelength. Once a light pulse has been generated, it is then attenuated by a neutral density filter contained within the SuperK laser hardware.

For the PQ lasers, a PicoQuant-brand SEPIA II laser driver is used to set the raw pulse intensity and firing rate. Unlike the SuperK, the intensity control is given by the laser driver’s driving voltage, as a fraction of the maximum possible voltage. Problematically, the dependence on the raw output intensity of the PQ laser heads are highly nonlinear with respect to the amplitude of the driving voltage. To demonstrate, consider the observed npe per shot within the detector as a function of driving voltage for SMELLIE events using the PQ lasers: see Fig. 3.5. For low driving voltages, the resulting npe is very small, and rises slowly. However, near a certain ‘lasing threshold’ the npe observed increases super-exponentially,

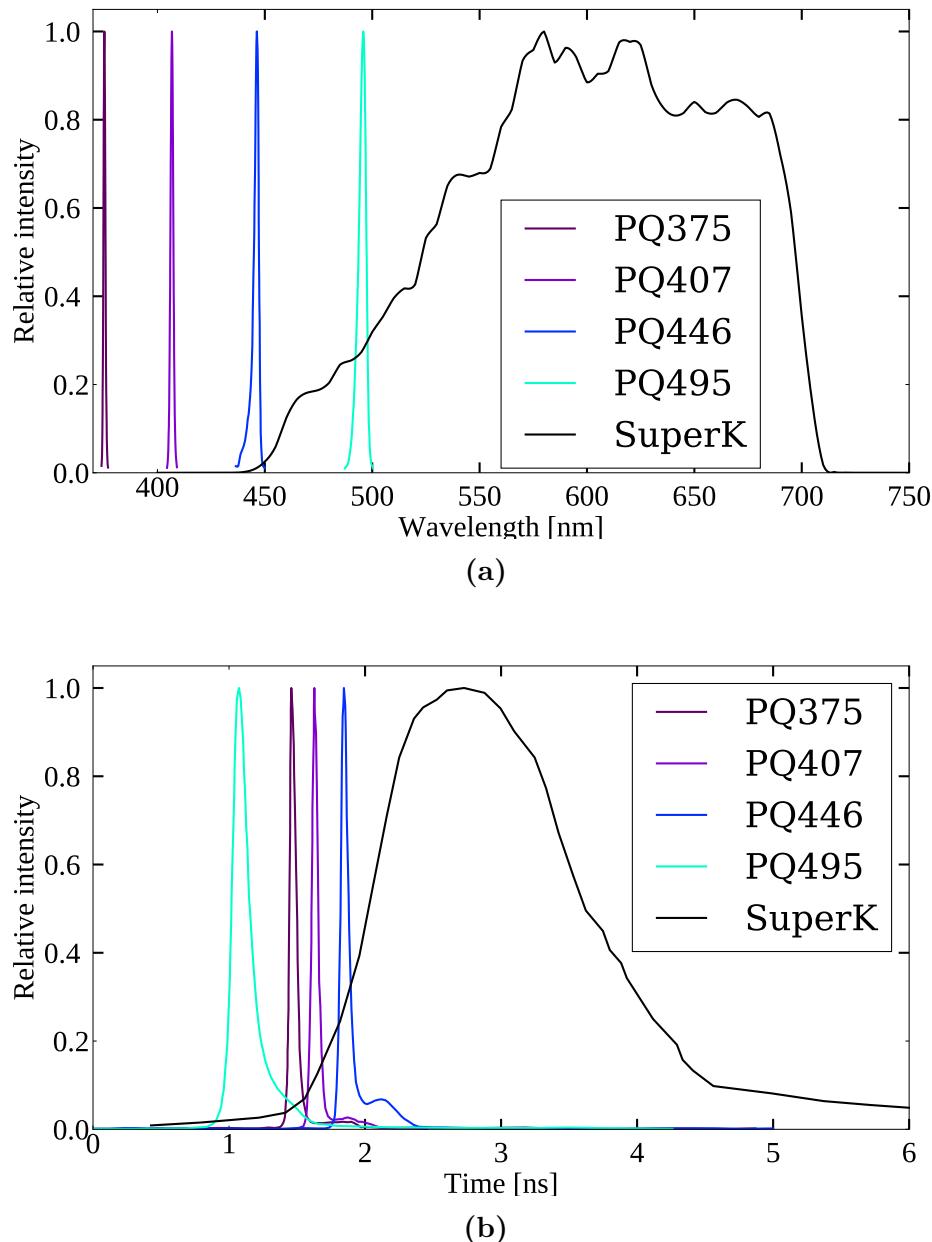


Fig. 3.4: Emission wavelength (a) and timing (b) spectra of the lasers used in SMELLIE. PQ spectral information taken from the manufacturer []; SuperK information measured by J. Lidgard [].

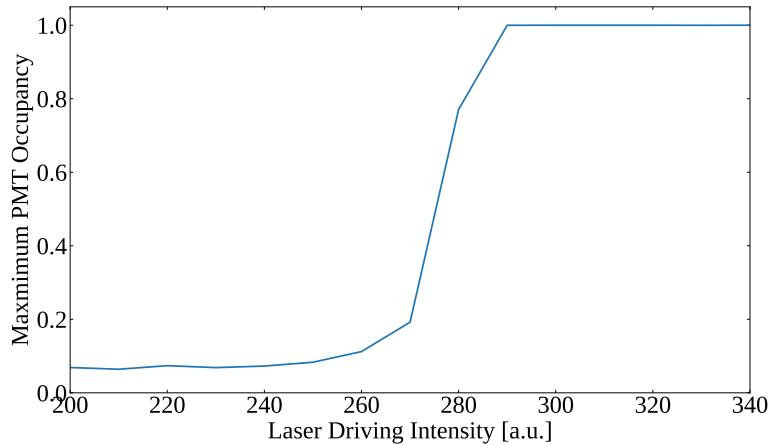


Fig. 3.5: Typical impact of driving voltage intensity for PQ446 laser on observed maximum occupancy in the detector. Data taken on February 22nd, 2021.

1 rapidly climbing many orders of magnitude in intensity. Above this threshold,
 2 increasing the driving voltage once again changes the observed intensity relatively
 3 slowly. A result of this is that if one requires an npe in the detector associated with
 4 a driving voltage that lies near the threshold, then any uncertainty in that driving
 5 voltage value can lead to dramatic changes in the npe observed. Furthermore,
 6 when driving a laser head near its lasing threshold the shot-to-shot variation in
 7 intensity can also become substantial: see Fig. 3.6 for an example of this occurring.
 8

9 During the water phase and for much of the scintillator phase, much of the
 10 data taken, especially using the PQ407 and PQ446 lasers, suffered from large
 11 shot-to-shot intensity variations. Throughout this period, after the light was
 12 generated by a PQ laser head it would then be passed through an attenuator,
 13 fixed to some nominal attenuation setting for each laser. In theory, one could
 14 solve the intensity variation problem by deliberately setting the intensity well
 15 beyond the lasing threshold, and then changing the attenuation of the attenuator
 16 to obtain the npe within the detector one is interested in. However, under the
 original hardware this was untenable as this would require someone to manually

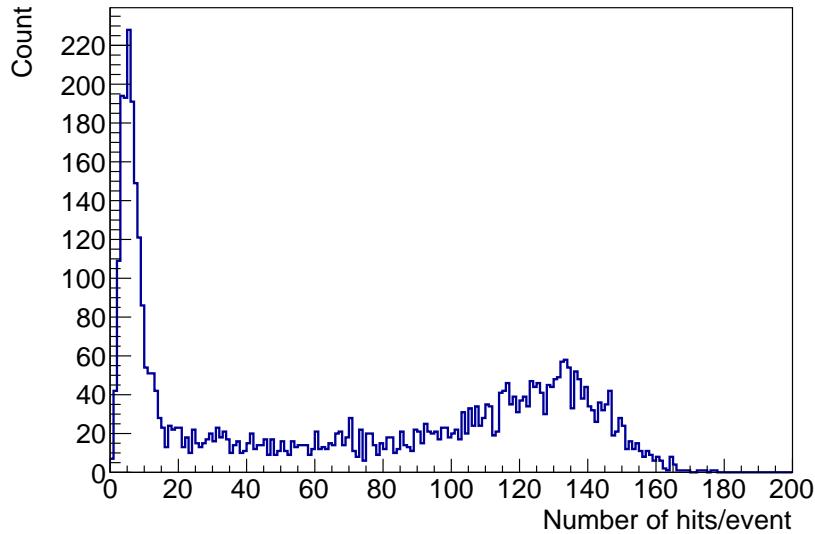


Fig. 3.6: Distribution of the number of PMT hits observed per event for laser PQ446 when the driving intensity is 280. Data taken on February 22nd, 2021.

change the attenuations in-person every time a different set of SMELLIE run
1
conditions were proposed.
2

Instead, Jeff Lidgard built a piece of hardware called the remotely-controllable
3
Variable Fibre Attenuator (VFA), shown in Fig. 3.7. Contained within a metal
4
housing were a ‘precision variable attenuator’ from DiCon Fibreoptics [] for each
5
PQ laser, along with an Arduino running firmware written by Jeff to enable
6
communication with each of the attenuators. Commands could be sent to a given
7
attenuator asking for a specific attenuation between 0–3000 dB. Following ex-situ
8
testing by Jeff Lidgard and Jasmine Simms, the VFA was installed underground
9
by myself and Armin Reichold in July 2022, with some assistance from Jeff in
10
integration of the hardware and SMELLIE server software.
11

During testing of the VFA in-situ, it was discovered that the inherent attenua-
12
tion of the variable attenuator at the minimum setting of 0 dB for the PQ375 laser
13
was so strong that negligible light was ever observed in the detector. Because of
14
this, the PQ375 was not hooked up to the VFA, and kept its original attenuator
15
setup. After fixing the driving voltage settings for PQ407, PQ446, and PQ495
16

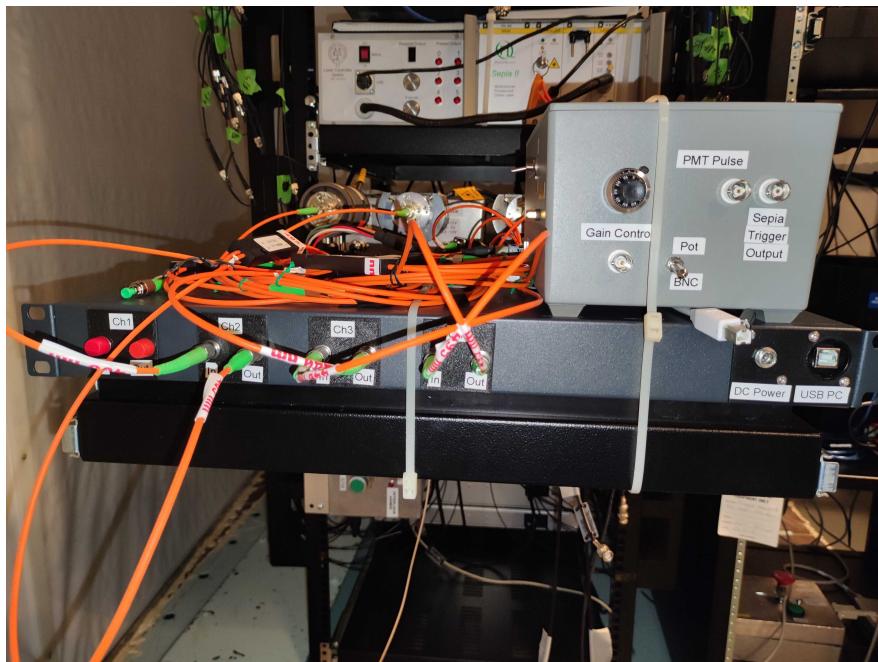


Fig. 3.7: Picture of the VFA during installation into the ELLIE hardware rack. The VFA box rests beneath the fibres connecting to the beamsplitters, as well as the newly-installed Monitoring PMT Unit (see Section 3.1.4 for details).

to be 1000, 750, and 750 respectively, the observed npe in the detector was once again compared to the input intensity setting. For PQ375 this input parameter remained the driving voltage; for the others, the attenuation setting was now used. The results can be seen in Fig. 3.8. As hoped for, the three PQ lasers hooked up to the VFA can now have stable intensities of light observed in the detector over multiple orders of magnitude of observed intensity.

3.1.3 Propagation of Light into the Detector

Once a pulse of optical light has been generated by the lasers and attenuated to the desired intensity, the next step is to navigate that light into the detector. This is achieved through a network of Corning-brand “InfiniCor SXi” multimode optical fibres [1]. These fibres were chosen in part for their low intrinsic radioactivity [1], as well as having a graded index as a function of radius. This latter property enables lower dispersion between different modes of the light, so that the initial sharpness

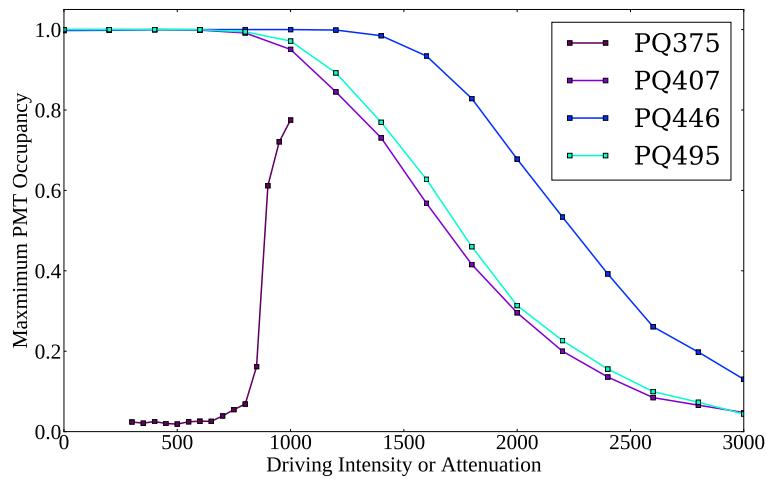


Fig. 3.8: Maximum occupancy in the detector as a function of the attenuation parameter value used, for each of PQ407, PQ446, and PQ495. Because PQ375 is not plugged into the VFA, its dependence on the driving intensity is shown instead. Data taken on the 12th of January, 2023.

of any given light pulse in time is maintained. However, because these fibres were mainly designed for telecommunication purposes, their nominal operating wavelengths are out in the near-infrared, 750–1450 nm. As SMELLIE only fires wavelengths in the range 375–700 nm, there is a small but non-negligible amount of light lost when propagating through the fibres.

After some light is split off by a beamsplitter to allow for ex-situ monitoring of the light pulse (see Section 3.1.4), it is sent to the Fibre Switch, two boxes manufactured by Laser Components UK [] that allows a user to remotely-control which of the fibres to send the light down into the detector.

Finally, the light that passes through the fibre switch propagates along one of the 15 optical fibres that have been submerged in the SNO+ cavity, whose ends are mounted to the PSUP. Specifically, sets of three fibres are mounted to a given node of the PSUP, with associated node numberings: nominally 07, 25, 37, 55, and 21. These provide for a variety of positions within the detector from which light can be emitted. Each mounting which holds three of the optical fibres

| Fibre | Node | Pointing direction |
|-------|------|--------------------|
| FS007 | 07 | 0° |
| FS107 | 07 | 10° |
| FS207 | 07 | 20° |
| FS025 | 25 | 0° |
| FS125 | 25 | 10° |
| FS225 | 25 | 20° |
| FS037 | 37 | 10° |
| FS137 | 37 | 0° |
| FS237 | 37 | 20° |
| FS055 | 55 | 10° |
| FS155 | 55 | 20° |
| FS255 | 55 | 0° |
| FS093 | 21 | 0° |
| FS193 | 21 | 10° |
| FS293 | 21 | 20° |

Table 3.1: SMELLIE fibre names, their associated mounting nodes on the PSUP, and their pointing direction. Taken from [33].

¹ also contains collimators, designed to reduce the possible range of angles with
² which the light can be emitted from. This is particularly important for SMELLIE,
³ because unlike the other ELLIE systems, a thin ‘pencil’ beam of light across
⁴ the detector is ideal for measuring scattering [57]. One last thing the mounting
⁵ achieves is to point each of the three fibres in different directions: 0°, 10°, and 20°
⁶ from the direction radially towards the centre of the detector.

⁷ Each fibre is given a name that nominally refers to both its mounting position
⁸ and its pointing direction. For example, the label ‘FS107’ corresponds to the
⁹ SMELLIE fibre mounted at node 07 with a pointing direction of 10°. Unfortunately,
¹⁰ during installation some fibres were mislabelled, leading to the node mounting
¹¹ points and pointing directions of some fibres being inconsistent with the labelling
¹² convention. The actual pointing directions for each fibre can be seen in Table 3.1.
¹³ The 3D positions and pointing directions of all the fibres are shown in Fig. 3.9.

¹⁴ When performing analysis with SMELLIE, a fibre-centric spherical-polar
¹⁵ coordinate system is used for clarity. This coordinate system was first fully-

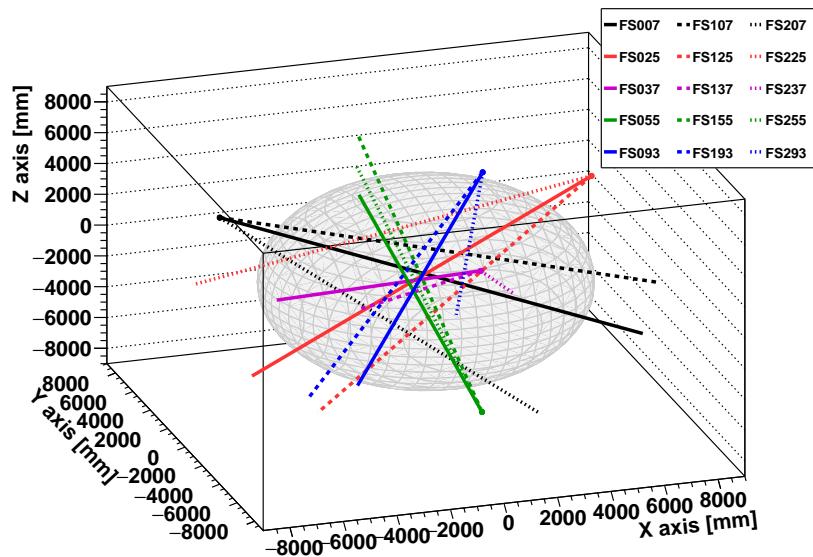


Fig. 3.9: Fibre emission points and nominal pointing directions within the detector.

developed by E. Turner; full details can be read in [33]. A position \mathbf{x} is measured relative to a given fibre's mounting position, \mathbf{f} (hereafter referred to as simply the fibre's position). Polar angles, labelled α , are measured relative to the fibre's nominal pointing direction, $\hat{\mathbf{u}}$. As a result, a line from the fibre position along $\alpha = 0$ across the detector should in theory hit the centre of the fibre's beamspot. Finally, in order to define the azimuthal angle ϕ , the plane orthogonal to $\hat{\mathbf{u}}$ is considered. Both \mathbf{x} and the unit vector in the vertically-upwards direction, $\hat{\mathbf{z}}$, are projected onto this plane, forming the vectors \mathbf{x}_{proj} and $\hat{\mathbf{z}}_{proj}$, respectively. ϕ is then defined as the angle going from $\hat{\mathbf{z}}_{proj}$ to \mathbf{x}_{proj} if one were to look at the plane in the direction of $\hat{\mathbf{u}}$. Mathematically, this can be written as:

$$\tan \phi = \frac{(\mathbf{x}_{proj} \times \hat{\mathbf{z}}_{proj}) \cdot \hat{\mathbf{u}}}{\mathbf{x}_{proj} \cdot \hat{\mathbf{z}}_{proj}}. \quad (3.1)$$

3.1.4 The Monitoring PMT Unit

As mentioned in the previous section, part of the light generated by the lasers gets split off from the main fibre path down into the detector, and instead is used for monitoring purposes. This is achieved with a box known as the Monitoring

¹ PMT Unit (MPU). As the name suggests, the MPU contains a small PMT that
² generates an electronic signal pulse from the laser light. This signal is then shaped
³ by electronics, and passed to the detector's central CAEN digitiser to have that
⁴ pulse digitised.

⁵ One problem with the existing MPU within SMELLIE was that the pulse
⁶ it produced was so broad that 300 ADC samples were needed to capture the
⁷ full shape (the CAEN samples at a rate of 1 every 4 ns). This led to a large
⁸ fraction of data being generated by SMELLIE events coming not from the PMT
⁹ hit information, but simply the MPU's signal digitisation. A natural consequence
¹⁰ of this was the rate at which the lasers could be fired had to be limited to
¹¹ typically 1 kHz, otherwise the detector was not able to handle the rate of data
¹² being generated.

¹³ Because of this, a new MPU was commissioned. Built by Adam Baird and
¹⁴ Johan Fopma from the Oxford Physics Central Electronics Group, this MPU had
¹⁵ updated electronics such that the pulse was shaped shorter. In addition, the rise
¹⁶ time of the pulse was made faster, in the hopes that the emission time of the light
¹⁷ pulse for a given event could be captured more accurately. The new MPU was
¹⁸ installed by myself and Armin Reichold at the same time as the VFA, in Summer
¹⁹ 2022.

²⁰ Alongside the installation of the new hardware, the settings in ORCA for the
²¹ CAEN digitisation of the MPU signal were updated. In particular, the number of
²² samples made by the CAEN was shortened from 300 down to 124. The timing of
²³ the CAEN sampling and delay on the TUBii trigger (more on the trigger shortly)
²⁴ was also modified. As a result of these changes, the shortest possible trace was
²⁵ now being generated, without missing any part of the MPU pulse or moving the
²⁶ observed TAC for hits in the detector outside the trigger window. Fig. 3.10 shows
²⁷ a comparison between typical MPU pulses taken before and after the upgrades.

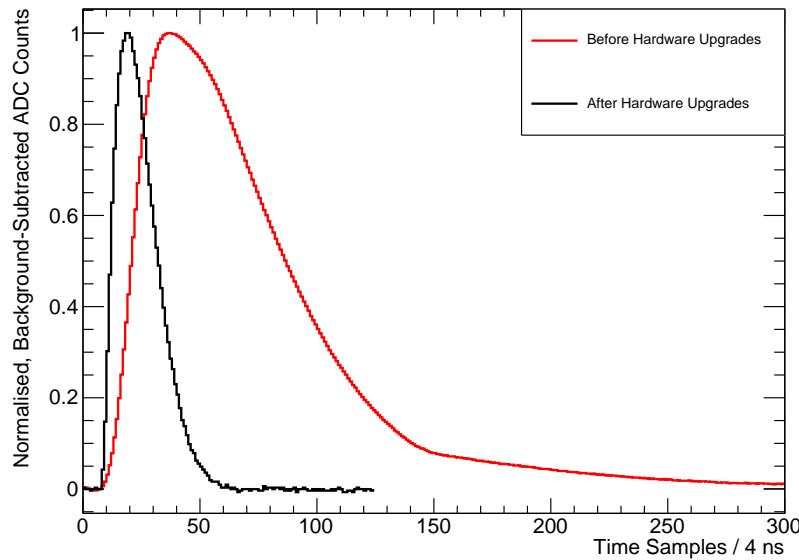


Fig. 3.10: Comparison between typical MPU CAEN traces generated by the PQ495 before and after the Summer 2022 hardware upgrades. The data before was taken on July 24th 2022, whilst the data after was taken on June 17th 2023. The traces have had their baselines subtracted off, and their peak values normalised to 1.

3.1.5 Event Triggering and Data Acquisition

As mentioned in Section 2.3.4, it is possible to trigger the SNO+ detector electronics via an external asynchronous trigger, EXTA. Taking data with SMELLIE takes advantage of this capability: instead of waiting for the normal ‘physics’ triggers such as N100 to pick up the event, because we know when we are firing the laser we can send an EXTA signal to trigger the detector precisely when a light pulse is within the detector.

Trigger signal pulses are created by the National Instruments DAQ Unit (the ‘NI Unit’), in place alongside the rest of the SMELLIE electronics. These trigger pulses are sent to either the PQ’s SEPIA controller or the SuperK laser to induce the firing of the relevant laser. If a PQ laser was fired, a trigger pulse is simultaneously sent to TUBii, which after a delay sends an EXTA trigger signal to the MTC/D, which then issues a GT. In contrast, the SuperK has a substantial variation in the emission time of laser light relative to a given driving

Fig. 3.11

¹ pulse. Instead of sending the trigger signal to TUBii in parallel to the SuperK
² laser, a photodiode contained within the SuperK laser system is able to detect
³ any generated light, and from that detection a new trigger signal is sent to TUBii.

⁴ A major problem with the handling of the trigger signal delay by TUBii was
⁵ present throughout the collection of water phase SMELLIE data. This delay was
⁶ ‘latched’ to the 100 MHz clock present within the electronics of TUBii. As a result,
⁷ the observed hit times of PMTs in the detector (which are relative to the GT time)
⁸ was a convolution of the hit times without the latching and a top hat function of
⁹ width 10 ns.

¹⁰ An example of this effect in action can be seen in Fig. 3.11. Light from the
¹¹ PQ495 laser was fired through fibre FSXXX by E. Turner in June 2018, during
¹² the water phase. Looking only at PMTs that were hit in the beamspot, defined
¹³ here as PMTs with $\alpha < 3^\circ$, the trigger times can be fit to the convolution of a
¹⁴ Gaussian with a 10 ns top hat.

¹⁵ During the partial fill phase, Tony Zummo updated the firmware of TUBii so
¹⁶ that this latching would no longer occur, and the arrival times to the MTC/D
¹⁷ were truly asynchronous to any clocks. The results are also shown in Fig. 3.11:
¹⁸ notice how the width of the peak is now much sharper. This width is determined
¹⁹ by the TTS of the hit PMTs, with the width of the PQ495 timing pulse being a
²⁰ subdominant effect.

²¹ Because this fix to TUBii was not in place until after all water phase data
²² taking had been completed, any analysis of SMELLIE data from the water phase
²³ had to contend with this 10 ns trigger ‘jitter’. Fortunately, this jitter was global
²⁴ to all hit times of a given event, so if measured a correction could be made.

²⁵ In this thesis, two similar approaches are used to measure the event-by-event
²⁶ emission time t_{emm} , the time at which laser light first emanates from the fibre in

Fig. 3.12

a given event. In both, the calibrated hit times of PMTs within the beamspot t_{hit} have the time-of-flight of light travelling from the fibre to the given PMT, t_{TOF} subtracted. t_{TOF} is calculated by using the “Light Path Calculator” algorithm developed within **RAT**. In one method, used both within the analysis work of E. Turner [33] and in Chapter 4 of this thesis, t_{emm} is measured as the second-earliest value of $t_{\text{hit}} - t_{\text{TOF}}$ in the beamspot. This method, called the “ $t_{\text{emm}} = t_2$ ” approach, skips the earliest event in the beamspot to be robust to noise hits.

Alternatively, in the “ $t_{\text{emm}} = t_{\text{med}}$ ” approach, the median value of $t_{\text{hit}} - t_{\text{TOF}}$ in the beamspot is used. We shall see in Chapter 5 that using t_2 has a bias as a function of the number of hits in the beamspot in a given event, whereas t_{med} does not.

This event-by-event approach to reconstructing t_{emm} is only necessary when the trigger system has unresolved problems. Because of the fix to the TUBii firmware, data taken using the PQ lasers during the scintillator phase did not in theory require using either of the t_{emm} reconstruction methods described above. However, for the SuperK laser a new problem was made clear. In Fig. 3.12 one can see the hit times of beamspot PMTs for SuperK light of wavelengths in the interval [490, 500]nm from fibre FSXXX, taken in XXXDATEXXX. A clear double-peaked structure can be seen. Also shown on the plot is the distribution of t_2 values. If the laser light were being emitted in two pulses for a given event, then we would expect to see only one bump in this distribution. However, the shape of the t_2 is clearly bimodal in a manner matching that of the overall timing distribution: this indicates that the actual emission times of light from the fibre can come at two different times relative to the GT time.

The origin of this double-peaked structure for SuperK events remains unresolved. However, the fact that events from the PQ lasers do not see this effect

¹ indicates that the issue likely lies somewhere in the part of the triggering system
² unique to the SuperK laser. For work within this thesis, it was decided to continue
³ to use an event-by-event approach to t_{emm} for data from all phases and lasers, to
⁴ ensure consistency.

⁵ 3.2 Software for SMELLIE Data-taking

⁶ The process of taking data with SMELLIE is largely automated. The user begins
⁷ by deciding on the settings — the number of laser pulses and their frequency, as
⁸ well as which fibre, laser, wavelength, intensity, and attenuation (if relevant) —
⁹ that will be used for each subrun. One run of SMELLIE data taking can have
¹⁰ an arbitrary number of subruns within it, each with their own setting. This is
¹¹ organised into a “run plan” JSON file for that run, and uploaded to a central
¹² CouchDB server for storage.

¹³ To begin data taking, ORCA is used to put the DAQ into a ‘SMELLIE’ run
¹⁴ type, which disables almost all the usual ‘Physics’ triggers such as N100 and N20.

¹⁵ The only triggers still masked into the MTC/D are:

¹⁶ • EXTA: This is obviously needed to capture the trigger signals sent from
¹⁷ SMELLIE.

¹⁸ • ESUMHI: This residual physics trigger is kept in with a high threshold, to
¹⁹ capture any high-nhit events not coming from SMELLIE, e.g. if a supernova
²⁰ were to go off whilst SMELLIE data taking was occurring.

²¹ • PULSEGT: This is a special trigger that fires at a rate of 50 Hz independent
²² of all other systems. Using this trigger the noise rate can be calculated on a
²³ run-by-run basis for each channel.

²⁴ The noise rates derived from the PULSEGT triggers are automatically calculated
²⁵ during the processing of runs, and are stored in RATDB tables. As mentioned

in Section 2.4.5, RAT can use these tables to replicate the noise levels in each
1
PMT when generating simulations to match data. When performing analysis on
2
SMELLIE data, only events from an EXTA trigger are usually considered.
3

With ORCA, one can then load in any run plan already uploaded to the
4
CouchDB server, and then execute it. ORCA then sends commands describing
5
the instructions laid out in the run plan to a server running on a computer called
6
SNODROP, which is housed in the same pair of electronic racks as the rest of the
7
SMELLIE hardware. This server converts these high-level commands into the low
8
level instructions understood by the specific pieces of SMELLIE hardware needed
9
to actually send the correct number of pulses of laser light of the right wavelength
10
and intensity through the correct fibre.
11

Whilst SMELLIE data is being taken, a “run description” JSON file is generated
12
and then uploaded to the SMELLIE CouchDB server, describing the settings of
13
the SMELLIE run actually executed. This differs from the run plan slightly in two
14
main ways: firstly, the same run plan can be run on multiple occasions. However,
15
a run description file is associated to exactly one run of data that was actually
16
taken. Secondly, if a SMELLIE run was terminated early either by the detector
17
operator, or from a hardware problem, the run description file will correctly show
18
only the subruns that were actually performed. The code repository used in all
19
SMELLIE analysis has been updated by myself to use a given run description
20
file in concert with the associated SMELLIE data file, so that the subrun-level
21
metadata can be known during analysis.
22

Chapter 4

1

Simulating SMELLIE Events

2

Max Power : *Kids. From now on there are three ways of doing things: the right way, the wrong way, and the Max Power way.*

Bart Simpson : *Isn't that just the wrong way?*

Max Power : *Yes, but faster!*

THE SIMPSONS

Critical to extraction of scattering information from SMELLIE data is an accurate Monte Carlo (MC) simulation of the SMELLIE system. By modelling the laser light emission into the detector correctly, we can simulate how SMELLIE light will be impacted by changing scattering lengths in the detector. Because of the complexity of the optics of the optical fibres used to direct the laser light into the detector, a given SMELLIE event is simulated as a partially-collimated “flash” of visible photons emanating from the emission point of the fibre into the detector. This flash then requires a number of parameters to be correctly described. In particular:

4

5

6

7

8

9

10

11

12

- 1 • **Fibre emission positions** were recorded during the installation of the
2 fibres.
- 3 • **Wavelength and emission timing distributions** of light pulses were
4 taken from measurements of the laser heads by their manufacturers [], or by
5 colleague Jeff Lidgard in the case of the SuperK wavelength distribution [].
- 6 • **The “pulse magnitude”**, defined as the mean number of photons simu-
7 lated per event, is determined on a subrun-by-subrun basis, and is assumed
8 to fluctuate as a Poisson distribution.
- 9 • **The beam profiles**, which describe the angular emission distributions of
10 each fibre, is the focus of this chapter. These are necessary because unlike
11 scintillation light, the light emitted from SMELLIE fibres is not isotropic.
- 12 • **Nominal fibre emission directions** attempt to define the centre of the
13 beam for a given fibre.

14 This chapter is split into three sections. Improvements to the existing simula-
15 tion algorithm for the beam profiles are first made, and then the beam profiles
16 themselves are updated. Finally, comparisons between data and simulation are
17 made after the upgrades to investigate any remaining discrepancies.

18

4.1 Improving the SMELLIE Generator Algo- 19 rithm

20

4.1.1 Previous Attempts at SMELLIE Event Simulation

21 Before we can determine the beam profiles, we must first decide how to specify them.
22 Previous observations show that different fibres can have notably different beam
23 profiles [57], so we let each fibre’s beam profiles be unique. We assume for now that

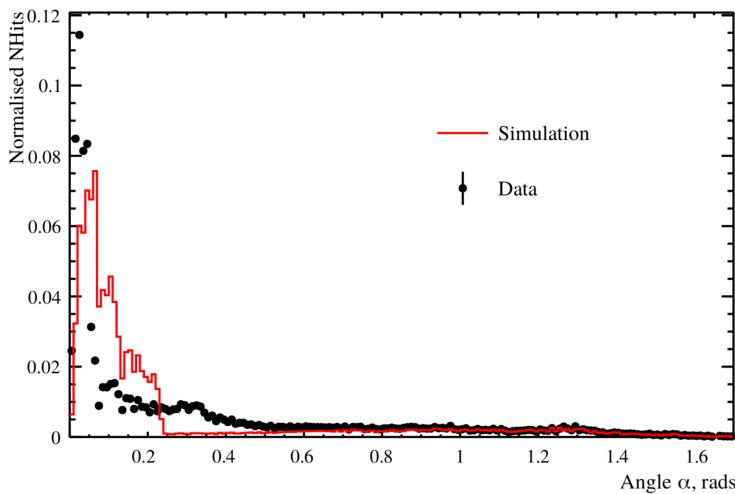


Fig. 4.1: Comparison between a simulation of one of the fibres, made from the 1D beam profile generator (red), with the associated data subrun that was used to create that beam profile (in black). For both MC and data, what is plotted is the PDF of observed PMT hits, as a function of the α angle. Poissonian errors have been added to the data points, but are too small to see. Clearly, this 1D generator does not replicate the observed beam profile correctly. Figure taken from [33].

a given fibre's beam profile is stable over time, and independent of the wavelength of light fired. A straightforward, naïve approach to parameterising a beam profile would be as follows: specify some nominal fibre direction, corresponding to the direction light takes travelling from the fibre to the centre of the “beamspot” observed on the other side of the detector. Then, specify a 1D beam profile, corresponding to the probability density of firing a photon at a given polar angle α relative to the nominal direction. One might even assume this distribution is Gaussian. The distribution in azimuthal direction, ϕ , is assumed to be uniform.

This 1D beam profile approach was used initially for SMELLIE, and remains in use for the other ELLIE sub-systems within SNO+. However, when SMELLIE data was taken in the water-phase of the experiment, simulations using these beam profiles failed to match them well at all - see figure 4.1 for an example. Not only was the distribution in α not Gaussian, a distinct speckle-pattern can be observed within the beamspot that is not uniform in ϕ . This fact led to colleague Esther Turner building a SMELLIE generator that could handle 2D beam profiles:

1 dependent on both α and ϕ . The distribution was stored as a map from each
2 inward-pointing PMT in the detector to a relative intensity value. This was chosen
3 because the beam profile shapes were calibrated from existing SMELLIE data —
4 more on this in section 4.2.

5 This original 2D generator then sampled the beam profile via a rejection
6 sampling approach, outlined as follows:

- 7 1. Propose a test direction (α, ϕ) , by generating ϕ uniformly in the interval
8 $[0, 2\pi]$, and α according to some pre-determined Gaussian distribution,
9 known as the Gaussian envelope.
- 10 2. Given this test direction, calculate where a line following this direction from
11 the fibre of interest will hit the PSUP on the other side of the detector. Find
12 the 3 closest PMTs to that point.
- 13 3. From those PMTs, obtain their relative intensity values from the beam
14 profile mapping, and perform an interpolation based on how close each
15 PMT is to the PSUP intersection point. This gives an interpolated relative
16 intensity value for this test direction.
- 17 4. Because we are sampling using the angular coordinates (α, ϕ) , differential
18 area elements over this space of directions do not have the same size. We
19 can correct for this fact by multiplying our interpolated relative intensity by
20 $\sin \alpha$, which corresponds to the Jacobian of the direction-space.
- 21 5. Calculate the value for the Gaussian envelope along this test direction.
- 22 6. Throw a random number uniformly between 0 and the Gaussian envelope
23 value. If the random number is less than the interpolated intensity, then
24 this test direction is accepted, and a photon is generated with that direction.
25 Otherwise, we reject the direction and try the whole process again.

This generator certainly works, but has a key problem: efficiency. The 1D generator was able to generate a SMELLIE event (that is, to fully specify the starting parameters of all the photons emitted from a fibre) at a speed of ~ 1 ms. However, the 2D generator specified here could take upwards of ~ 50 s *per event* to generate. Because a typical SMELLIE analysis requires simulating many millions of events, the CPU time taken to perform this quickly became unfeasible. Fixing this generator speed problem was a high priority for the SMELLIE analysis.

4.1.2 The new generator

On careful inspection of the existing 2D generator, the main reason for the slowness of the algorithm is the use of a rejection approach. Even with use of the Gaussian envelope, which was included to help with speed, the vast majority of proposed directions are never selected. Figure 4.2 shows a histogram of number of attempts per event it took for a valid direction to be chosen for a representative SMELLIE simulation. Moreover, the calculations needing to be done for every proposed direction are relatively complex, notably trying to find the 3 nearest PMTs to some point on the PSUP.

A new 2D generator was built with these thoughts in mind. Firstly, the rejection method would no longer be used, given its inefficiency. We would also endeavour to try and “pre-calculate” as much as possible before run-time. Starting with the existing PMT relative intensity maps, we plot these in the 2D direction-space $(1 - \cos \alpha, \phi)$: see Figure 4.3a. In a toy-MC simulation, 500,000 directions are then thrown uniformly in this 2D space per fibre. For each direction, the same method of obtaining an interpolated intensity value from the nearest PMTs to the corresponding point on the PSUP as from the original 2D generator was performed, the only difference being that these calculations were done well before

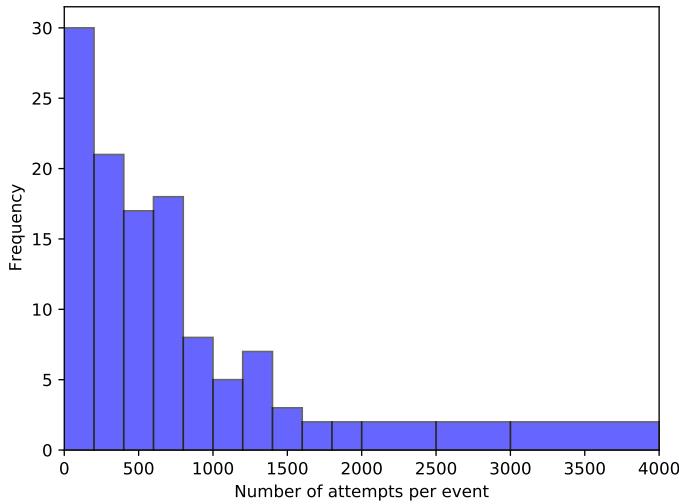
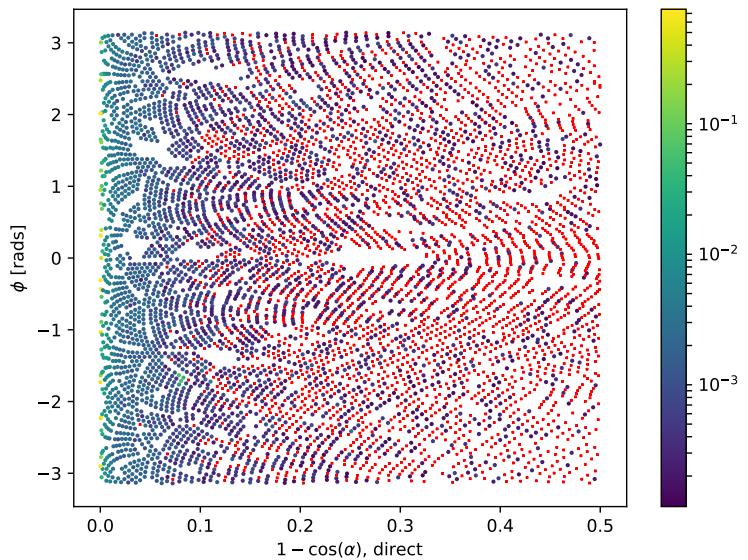


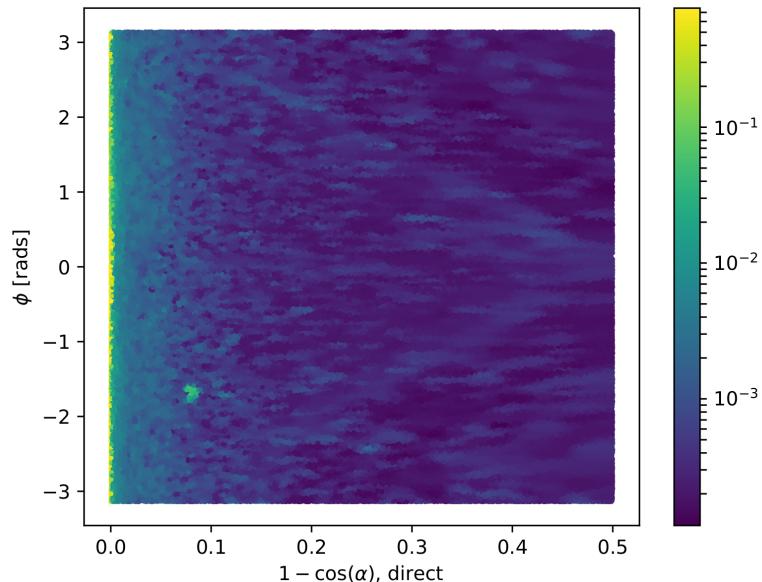
Fig. 4.2: Typical distribution of the number of attempts it takes for the existing 2D generator before the test direction gets accepted, per event.

¹ any actual SMELLIE simulation. Figure 4.3b shows the interpolated intensities
² obtained for one fibre.

³ Following this, the sampled intensities were then binned into a 2D histogram,
⁴ where the bin value corresponds to the sum of all intensities for all directions
⁵ found within this bin. Choosing a sensible binning procedure is important: too
⁶ few bins, and necessary information about the shape of the beam is lost, whilst
⁷ too many bins can oversample the data and capture statistical artefacts in the
⁸ sampling process instead of just the beam profile. As a balance, 15 bins were
⁹ chosen along the ϕ direction, and 60 in $r = 1 - \cos \alpha$. This was chosen to ensure
¹⁰ that a reasonable number of PMTs were located within each bin, lessening the
¹¹ impact of any statistical fluctuations. Although the bins in ϕ were chosen to have
¹² uniform width, this was decided to be not the case for the other axis, as there is
¹³ far more important information near $r = 0$ (the beamspot). Instead, the width
¹⁴ of the bins in r were calculated so that roughly the same total probability was
¹⁵ contained in each r -strip. By consequence, bins near the beamspot typically are
¹⁶ of significantly smaller size than ones much further out. This allows us to both



(a) PMT relative intensity map



(b) Interpolated intensity map

Fig. 4.3: The first step in the new method for preparing the new generator. In (a), the relative intensities used for the existing beam profile of fibre labelled FS055 are shown for each PMT, the position on the plot indicating the location of that PMT in the fibre coordinates. The colour indicates the relative intensity; PMTs marked red have an intensity of zero. Figure (b) shows the result of throwing 500,000 directions uniformly over this 2D space, the intensity of each point given by interpolating the intensities of nearby PMTs.

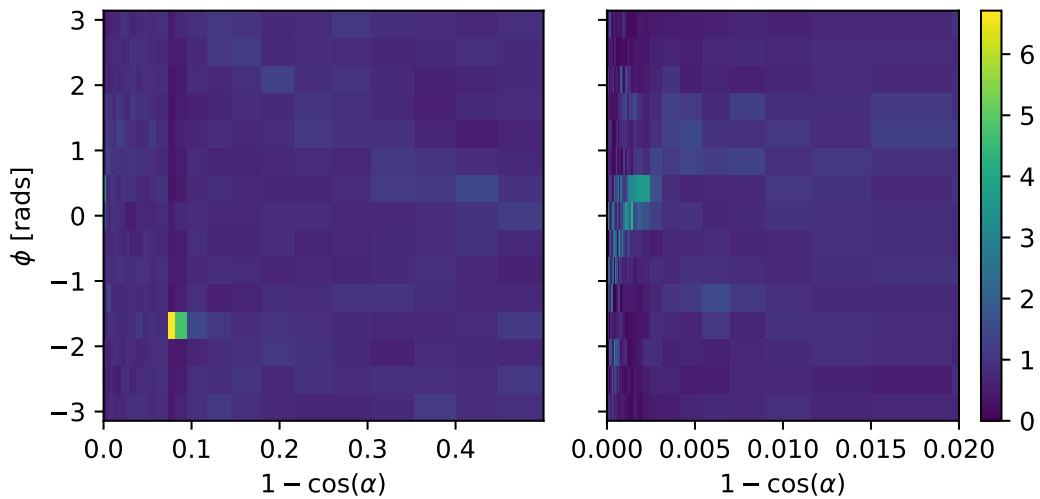


Fig. 4.4: Histogram of interpolated intensities within the 2D direction-space. The left view shows the full histogram; the right is a zoomed-in version near the beamspot. Unlike the binning in ϕ , the bin widths in r are not at all uniform. Instead, they have been determined such that the area summed over a given “strip” of bins of constant r will be the same.

capture any rapid changes in intensity near the beamspot, where this matters greatly, and smooth out the very-low intensities seen at larger polar angles. One of these histograms can be seen in Figure 4.4: the large change in bin widths as a function of r is clear. One can also see that near the beamspot notable dependence on the intensity as a function of ϕ . The mysterious “spot” at $r = 0.08$, well out of the beamspot, is an indication that the underlying beam profile data being used requires improvement: more on this in section 4.2.

The Cumulative Density Function (CDF) of this intensity histogram as a function of bin was then produced, where the bins were ordered through a raster-scan: scanning first over ϕ , and then r . The CDF was then normalised to 1 so that it was well-defined. It is this CDF object that is then loaded in and sampled from during event generation. To do this, an “inverse-CDF” approach was used, which has the major benefit over rejection sampling of always producing a valid direction for every sample made. The algorithm works as follows:

- 15 1. Throw a random number uniformly in $[0, 1]$.

-
2. Perform a binary search to find the bin that has the largest CDF value 1
below this random number. 2
3. Look at the bin edges in ϕ of this selected bin: use linear interpolation of 3
the random number to obtain a ϕ value located between these two ϕ -values. 4
4. Look at the selected bin's r -bin edges, and select a value of r by throwing a 5
second random number uniformly between the two edges. Convert this r 6
into a polar angle α . 7
5. The photon's direction is defined by the (α, ϕ) chosen by this process. 8

Because of the relative simplicity of this algorithm compared to the previous 2D 9
generator, the speed improvement was very large: generation now took ~ 1 ms per 10
SMELLIE event, a speed improvement of nearly 50,000. Event generation became 11
as fast as it was when the 1D generator was being used. Furthermore, because 12
of the approach taken, this major speed improvement comes at no sacrifice in 13
accuracy. Figure 4.5 shows a comparison of the average number of photoelectrons 14
(npe) per event per PMT between water-phase SMELLIE data and simulations 15
with both the old and new 2D generator. One can see clearly that both generators 16
are as accurate as one another. Note that this plot uses the updated beam profiles 17
as explained in the next section. 18

4.2 Improving the beam profiles

Even with the new 2D profile generator, a problem remains: the simulation fails 20
to reasonably recreate data, and much of this appears to be because of the poor 21
beam profile data being used. The curious “spot” for one of the fibres was already 22
noted in the previous section that doesn't seem to be physical, and more broadly 23
at large angles for all the fibres there are large swathes of PMTs with an intensity 24

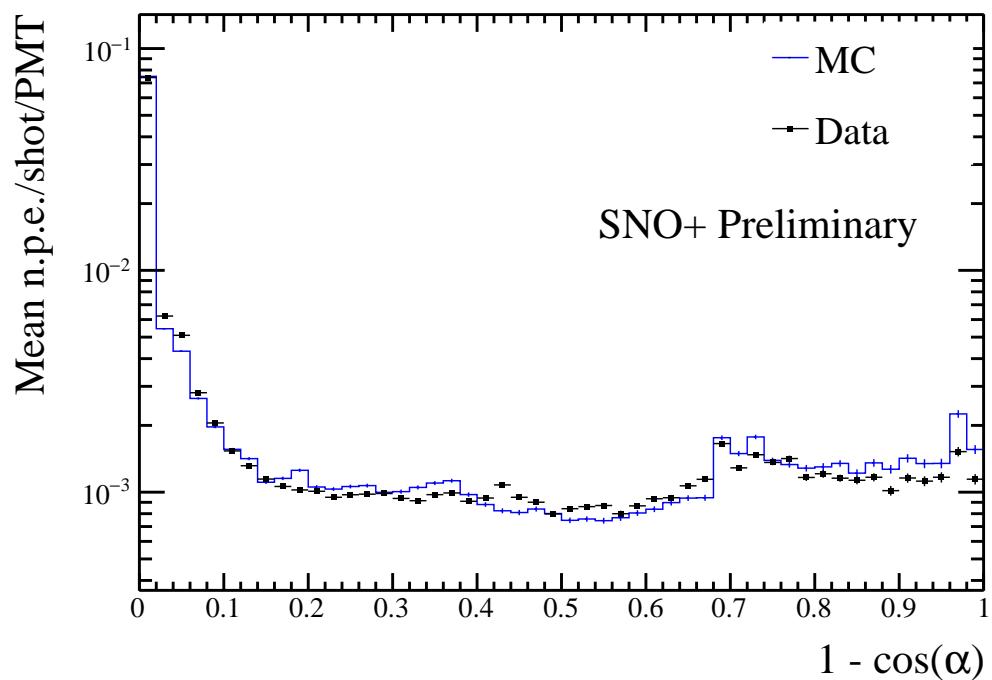


Fig. 4.5: Comparison of water-phase data to MC generated using both the old and new 2D beam profile generator approaches, with the updated beam profiles. Both versions of the generator are consistent with one another, but the new generator is many times faster.

of zero, providing little useful information about the beam shape. It was shown in [33] that with the old 2D generator, the systematic uncertainty on the beam profiles was the dominant source of error in the main SMELLIE analysis. To help improve this situation, it was decided to update the existing beam profiles.

These old beam profiles were originally determined by looking at SMELLIE data taken during the water-phase. Specifically, a “medium”-intensity subrun with one of the lasers firing at a wavelength of 495 nm, was chosen for each fibre. “Medium”-intensity corresponds to firing the relevant laser at a set intensity determined during an earlier commissioning process, for which the maximum occupancy of PMT hits at that intensity, i.e. the proportion of hits per event, corresponded to roughly 80%. This value was chosen as it allowed for high statistics in a relatively short run-time, but not so intense that the occupancy of any given PMT in the beamspot was 100%. Because Rayleigh scattering is strongly-dependent on wavelength, the long wavelength of light was chosen so that impacts from this scattering were small in the data.

SNO+ PMTs are unable to distinguish the exact number of photoelectrons being generated. One is typically only able to know if a PMT has been triggered at all, by any number of photoelectrons. As a result, the occupancy of a PMT over a number of SMELLIE events, o , is a biased estimator of the mean number of photoelectrons generated, μ . Assuming the number of photoelectrons generated in a given event follows Poisson statistics, the probability of generating k photoelectrons is:

$$P(k|\mu) = \frac{\mu^k e^{-\mu}}{k!}. \quad (4.1)$$

The probability of observing a “hit” in a given PMT corresponds to generating at least one photoelectron:

$$P(\text{hit}|\mu) = P(k \geq 1|\mu) = 1 - P(k = 0|\mu) = 1 - e^{-\mu}, \quad (4.2)$$

which implies after rearrangement that one can determine the mean number of photoelectrons per event from the occupancy by:

$$\mu = \ln(1 - o). \quad (4.3)$$

This is the reason why we want to avoid PMTs with occupancies of 100%: they preclude one's ability to convert into a value for μ by looking at occupancy alone. We call this conversion from occupancy into npe the “multi-hit correction”. The impact of this correction is typically small for most PMTs, but can become very significant in a fibre's beamspot.

Once the npe mapping from data was obtained, a correction was then made for the detector's optics: even ignoring a fibre's beam profile, we still expect certain PMTs to be illuminated more than others because of e.g. reflections off the AV, or the solid angle subtended by the PMT bucket opening. For each fibre, a simulation was made where the beam profile was set as isotropic, and the corresponding npe mapping obtained: this map held information about the detector optics only. The beam profile mapping was then derived by simply dividing each fibre's npe mapping from data to its associated isotropic MC npe map. It is these maps that were first used in section 4.1.2.

4.2.1 Combining beam profile datasets

Fortunately, much more SMELLIE data was taken during the water-phase than was used for the original beam-profiling analysis. This additional data can be combined with that which was already used to far better constrain the beam profiles. In particular, given the existing assumption that scattering effects are

| Run Number | Run Type | Comments |
|------------|---|---|
| 114,018 | All PQ lasers; SuperK laser in 400–500 nm range | Only PQ495 laser and SuperK at 495 nm is used |
| 114,023 | SuperK laser in 500–600 nm range | Part 1 of this wavelength range; crash occurred on last subrun, so that subrun is ignored |
| 114,034 | SuperK laser in 500–600 nm range | Part 2 of this wavelength range |

Table 4.1: Water-phase runs used for new beam profiling.

minimal above wavelengths of ~ 490 nm, all data taken with wavelengths above this can also be used. The specific runs (and associated comments about their specifics) are described in Table 4.1. Because high-intensity runs require a different analysis approach (PMTs with high occupancies must use charge, not occupancy, to estimate npe), for this analysis we only considered subruns that used low or medium intensity set-points.

For each subrun j of data per fibre, we look only at PMT hits for each PMT i that has been identified as “good” for that subrun¹, $i \in G_j$. G_j here represents the set of good PMTs in subrun j . In particular, a “good” PMT must have valid electronic and timing calibrations, be at high voltage and masked into the detector’s trigger system for that subrun. In addition, an angular cut of $\alpha < 60^\circ$ was made to remove PMTs that are well outside any reasonable beam direction. To isolate the hits arriving directly from the fibre without reflecting, scattering, or being noise, a time cut was also made. Because what matters is the time relative to emission from the fibre, and the expected time-of-flight from fibre to different PMTs varies, a quantity known as the time residual was used. Starting with the calibrated hit time of a given PMT relative to the event’s trigger time, t_{hit} , the expected time-of-flight t_{TOF} from the fibre to the PMT was subtracted, estimated with the collaboration’s “Light Path Calculator”. Then, the emission time was also subtracted, t_{emm} , estimated by looking at the second-earliest value of $t_{hit} - t_{TOF}$

¹Strictly speaking, a PMT’s “goodness” is only determined on a run-by-run, not a subrun-by-subrun level, but this has no impact on the analysis.

within the fibre's central beamspot, defined as the PMTs for which $\alpha < 3^\circ$. It was found that a “loose” time residual cut of $t_{res} \in [-10, +12]\text{ns}$ was sufficient to remove the vast majority of non-direct light with little signal sacrifice. In the situation where a subrun with intensity was very small, it would not regularly have at least two hits in the beamspot, and so the time residuals calculated would not be valid for many events. To avoid this situation, a cut was made on any subruns with mean intensities below 9 within their beamspot. This value was chosen as it would mean a 2σ fluctuation downwards of $2 \cdot \sqrt{9} = 2 \cdot 3 = 6\text{npe}$ would still have more than the 2 hits necessary for timing reconstruction. One fibre, FS207, has no data subruns that satisfy this condition, and as such will have to be dealt with separately. For the time being, this fibre was ignored.

Extracting the underlying beam profiles from these data required some careful thought, especially because different subruns could have different intensities. Considering a PMT i in subrun j , the mean number of photoelectrons generated per event in that PMT for that subrun, μ_{ij} can be decomposed as follows:

$$\mu_{ij} = I_j k_i = I_j b_i f_i. \quad (4.4)$$

I_j is the intensity of the subrun, i.e. the mean number of photons generated from the fibre in that subrun per event. k_i is the probability that a given photon generated at the fibre source ends up generating a photoelectron in PMT i . This itself can be further split into two components: b_i , the probability that a given photon at the fibre source points in the direction of PMT i ; and f_i , the probability that a given correctly-pointed photon actually makes it to the PMT and successfully generates a photoelectron. It is b_i that is the actual beam profile we would like to measure.

Letting p_{ij} be the probability of observing a hit for a given event on a given PMT, the probability of observing m_{ij} hits out of N_j events in the subrun will be

binomially-distributed:

$$P(m_{ij}|\mu_{ij}) = L(\mu_{ij}|m_{ij}) = \binom{N_j}{m_{ij}} p_{ij}^{m_{ij}} (1-p_{ij})^{N_j-m_{ij}} = \binom{N_j}{m_{ij}} (1-e^{-\mu_{ij}})^{m_{ij}} e^{-\mu_{ij}(N_j-m_{ij})}. \quad (4.5)$$

Here we have used equation 4.2, and noted that this probability distribution in m can be re-framed as a likelihood function for the parameter μ_{ij} . Considering only a single subrun of data, the maximum likelihood estimate of the parameter μ_{ij} can be shown to be:

$$\langle \mu_{ij} \rangle = -\ln \left(1 - \frac{m_{ij}}{N_j} \right) = \ln (1 - o_{ij}) \quad (m_{ij} \neq N_j), \quad (4.6)$$

where o_{ij} is just the occupancy of PMT i in subrun j . This is just the multi-hit correction formula seen in equation 4.3, which makes sense.

When looking at multiple subruns for the same fibre, the total likelihood function for a given PMT when considering all the data for a given fibre will be the product of the likelihoods from each dataset,

$$L(\{I_j\}, k_i | \{m_{ij}\}) = \prod_j L(I_j, k_i | m_{ij}) = \prod_j \binom{N_j}{m_{ij}} (1-e^{-I_j k_i})^{m_{ij}} e^{-I_j k_i (N_j-m_{ij})}. \quad (4.7)$$

This leads to a log-likelihood distribution of

$$\mathcal{L}(\{I_j\}, k_i | \{m_{ij}\}) = \sum_j \left[\ln \left(\binom{N_j}{m_{ij}} \right) + m_{ij} \ln \left(1 - e^{-I_j k_i} \right) - I_j k_i (N_j - m_{ij}) \right]. \quad (4.8)$$

Formally, one could combine the likelihoods of all the PMTs together, and by looking at the maximum likelihood estimates for each of the parameters measure the parameter values this way. However, the set of equations one obtains through this approach quickly become analytically intractable, because the PMTs are coupled by the intensity values I_j . Even a direct numerical approach would be

¹ liable to fail: for a given fibre there can be dozens of subruns, and many thousands
² of PMTs of relevance, so the dimensionality of the system of equations would be
³ far too large.

⁴ Because of this, a different approach was taken. It is expected that in a subrun
⁵ the total npe, summed over all good PMTs, should be proportional to the intensity
⁶ value I_j . One must be careful about this construction — different subruns can
⁷ have different sets of good PMTs, so two subruns with identical I_j values could
⁸ have a larger summed npe merely because more PMTs were good in that subrun.
⁹ To counter-act this effect, only PMTs that were classified as good in *all* subruns
¹⁰ being analysed for that fibre would be used for the npe summation. In other
¹¹ words, we use data from PMT i for summing only if:

$$\text{12} \quad i \in \mathcal{I} = \bigcap_j G_j. \quad (4.9)$$

¹³ We can then define the summed npe for a given subrun as $S_j = \sum_{i \in \mathcal{I}} \text{npe}_{ij}$, and
¹⁴ assert that $I_j = cS_j$. By finding a value proportional to I_j , there is now enough
¹⁵ information to maximise the log-likelihood $\mathcal{L}(k_i | \{m_{ij}\}, \{I_j\})$ with respect to k_i
¹⁶ for each PMT independently, and hence obtain estimates for these k_i parameters.

¹⁷ Of course, what is actually wanted are the underlying b_i values, not k_i . This is
¹⁸ where isotropic simulations come in. For each run of data used, a matching isotropic
¹⁹ MC was produced. As an example, a simulation for run 114,023 contained 200,000
²⁰ events for each fibre using an isotropic beam profile, over the full wavelength
²¹ range considered in this run, 500–600 nm, using the same run conditions as in
²² data (which PMTs were at high voltage, etc.).

²³ For each isotropic MC run, both I_j^{MC} and k_i^{MC} were calculated via the method
²⁴ described above. Because the simulations were isotropic, the underlying value
²⁵ for b_i was constant across all the PMTs, and so $ak_i^{MC} = f_i$. By doing some

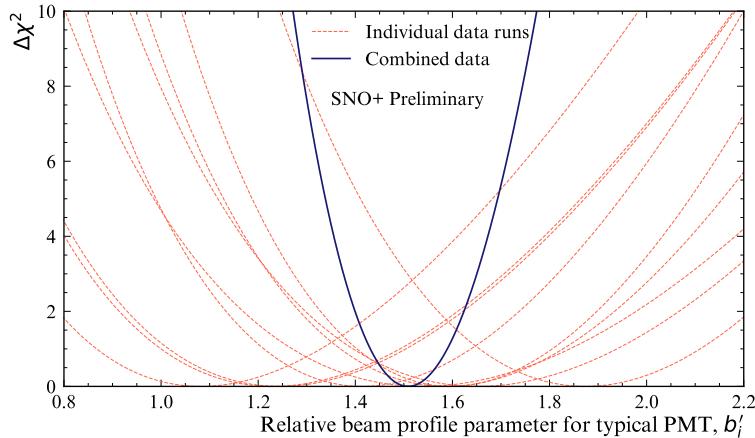


Fig. 4.6: Plot of $\Delta\chi^2 \simeq X_i$, twice the negative log-likelihood ratio, for both single subruns of a typical PMT, and when all relevant subruns are combined.

rearranging of equation 4.4, we find that:

$$\mu_{ij} = I_j b_i f_i = c S_j b_i a k_i^{MC} = (acb_i) S_j k_i^{MC}. \quad (4.10)$$

As a result of this, given the set $\{S_j\}$ and k_i^{MC} , one can maximise the log-likelihood \mathcal{L} with respect to $b'_i = acb_i$ numerically, to obtain the maximum likelihood estimate of b'_i . Because a and c were global constants of proportionality, they would become irrelevant as soon as the beam profile was normalised in the CDF-creation process outlined in 4.1.2.

Figure 4.6 shows the shape of this log-likelihood distribution for a particular PMT when considering fibre FS007's beam profile. One can see how individual subruns provide much more information when combined than if one looked at a single subrun alone.

Another benefit of using this log-likelihood approach is that the resulting distribution's shape can be used for uncertainty estimation. In almost all cases, Wilks Theorem [58] allows us to produce 1σ confidence intervals about the maximum

likelihood estimate for b'_i , $\langle b'_i \rangle$, because

$$X(b'_i) = -2 [\mathcal{L}(b'_i) - \mathcal{L}(\langle b'_i \rangle)]$$

¹ approximates a χ^2 -distribution. As a result, the error bounds on our parameter
² estimate are given by when $X = 1$. The fact that the shape of X can be well-
³ approximated by a quadratic in the region near $X = 0$ indicates the validity of
⁴ Wilks' Theorem being used here.

⁵ Only a couple of exceptions to this approach of parameter estimation are
⁶ possible. In the case where $m_{ij} = N_j$, i.e. a PMT has 100% occupancy, no
⁷ maximum likelihood estimate exists: we need not worry about this, as subruns
⁸ where this occurs have not been used. On the other end, however, there are some
⁹ PMTs for certain fibres where after all subruns of data have been included, there
¹⁰ remains no hits. In this scenario, one can show that the log-likelihood becomes
¹¹ linear in the beam profile parameter:

$$\mathcal{L}(b'_i | \{m_{ij} = 0\}) = b'_i k_i^{MC} \cdot \sum_j [I_j N_j]. \quad (4.11)$$

¹² This scenario is very much reminiscent of rare-decay searches, and a similar
¹³ approach can be used. A 1σ upper limit on the possible value for b'_i can be
¹⁴ analytically-calculated to be:

$$b'_{i,ulim} = -\frac{k_i^{MC} \sum_j [I_j N_j]}{\ln [1 - \text{erf} (1/\sqrt{2})]}, \quad (4.12)$$

¹⁵ where $\text{erf}(x)$ is the error function. [\[18 pages for above two sections\]](#)

¹⁶ 4.2.2 Results & Discussion

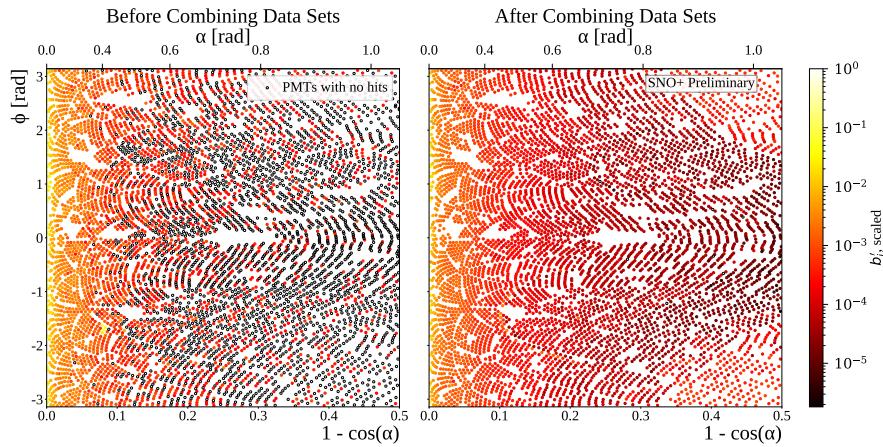


Fig. 4.7: Comparison between old and updated beam profiles for fibre FS055, after combining multiple data sets. Once again, the relative intensities (b'_i) for each PMT are given by the colour of each point, the position of each plotted in the 2D (r, ϕ) -space. The relative intensities have been both scaled here so that the largest value equals 1. Hollowed-out points are PMTs that, even after all relevant subruns have been combined, have no PMT hits.

WARNING: contents of this subsection will be gutted, focusing merely on impact
of combining data sets. Details about discrepancies will be covered in the next
section. [Will just be 1 page] Figure 4.7 shows the impact of using additional
subruns of data on a typical beam profile. One can clearly see the great reduction
in the number of PMTs with no hits in data. That many more data sets were
included allowed for the major increase in dynamic range available for measuring
these b'_i values. One can also note that by including additional data the curious
spot that was seen in the old beam profile our at $r \approx 0.08$ has gone, further
indicating that it was an artefact of that single data set.

Further details can be gathered from the interpolated intensity maps, one of
which can be seen in figure 4.8. There are two curious stand-out features that can
be seen here: firstly, there are multiple distinct parabolic arcs. These correspond
to the shadows of the ropes that hold up/down the AV. More precisely, they are
the mismodelling of those shadows — if the shadows were in the right place in
the isotropic MC, then they would correctly cancel out any decreased intensity
seen in the data of shadowed PMTs. These shadows could be mismodelled either

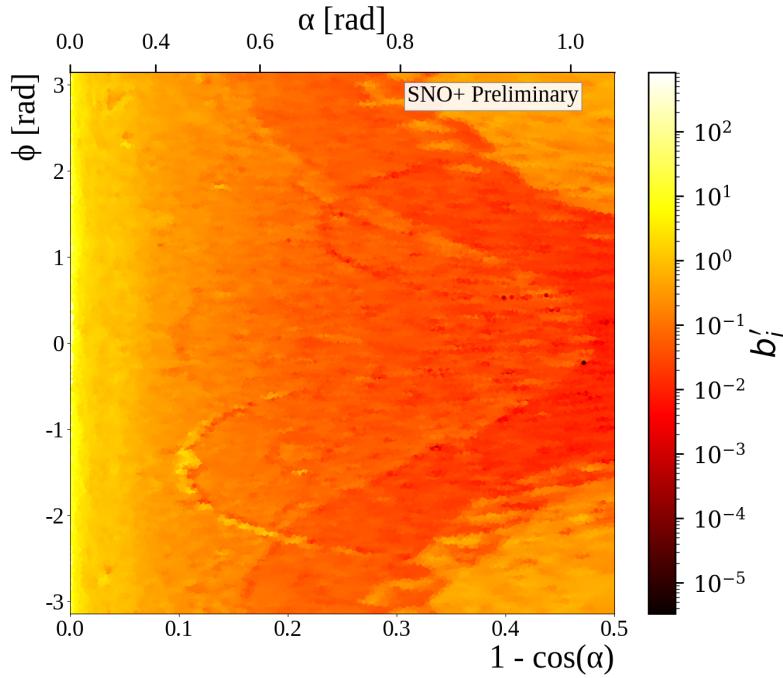


Fig. 4.8: Interpolated intensity map for the new updated beam profile of fibre FS055. The misalignment of rope shadows and AV effects, can both be seen.

because the positions of the ropes in the MC are in the wrong place, or the fibre's emission position is wrong. Note that any mismodelling of the fibre's nominal emission direction has no impact on this shadowing problem, as changing that direction merely causes a change of basis in the (r, ϕ) -space. The latter possibility of incorrect fibre positions are more likely, and in fact these arcs in the beam profiles could be used as an effective way to correct for this problem.

The second distinctive feature of this intensity map is the large band of lower intensity varying between $r \approx 0.2 – 0.5$, followed by larger intensity out at large r values. This feature comes from light reflecting off the AV surface, or internally-reflecting. The reason for this band's functional dependence on ϕ is that this particular fibre, FS055, has a nominal fibre direction $\sim 10^\circ$ from pointing radially-towards the detector's centre. This feature appears in the updated beam profiles of all fibres, but its shape depends on the particular fibre's direction — for fibres pointing directly towards the detector's centre, there is little ϕ -dependence

observed. Like the ropes, this feature must come from some form of mismodelling of the optics of the AV. A de-facto shadowing of PMTs in line with tangents from the AV surface which intersect the fibre position is to be expected. One also expects PMTs at polar angles larger than this to have their observed intensities boosted from reflected light off the AV. However, the discontinuities seen in the beam profiles indicate that for whatever reason this effect has been over-emphasised in the simulation.

There is a further phenomenon that can be seen, by comparing beam profile values obtained from a single subrun to the updated combined beam profile. This can be done by calculating the residuals corresponding to the single subrun, relative to the combined data set. The residual is negative if the combined data sets have a b'_i below the equivalent for a given single subrun; that is, the combined model underestimates this subrun for that PMT.

This information was plotted for two different subruns from the same fibre, seen in figure 4.9. One subrun was the same one used by Esther Turner for the original 2D beam profiling, with a wavelength of 495 nm; the latter was at the longer wavelength of 595 nm. For both subruns, most PMTs are seen to have intensities well-modelled by the combined model. However, there appears to be a significant amount of mismodelling within the beamspot. There also appears to be some systematic shift between data and model at somewhat larger polar angles. Moreover, this mismodelling seems not to be merely random, but a function of wavelength: at shorter wavelengths the beamspot tends towards being overestimated and then underestimated at larger values of α . At longer wavelengths, the beamspot becomes underestimated, with larger angles getting overestimated. This indicates that there appears to be a wavelength-dependence on the beam profiles, contradicting one of the main assumptions which we used to combine the water-phase data in the first place! All three of these features — rope shadows, AV reflections, and wavelength dependence — add systematic

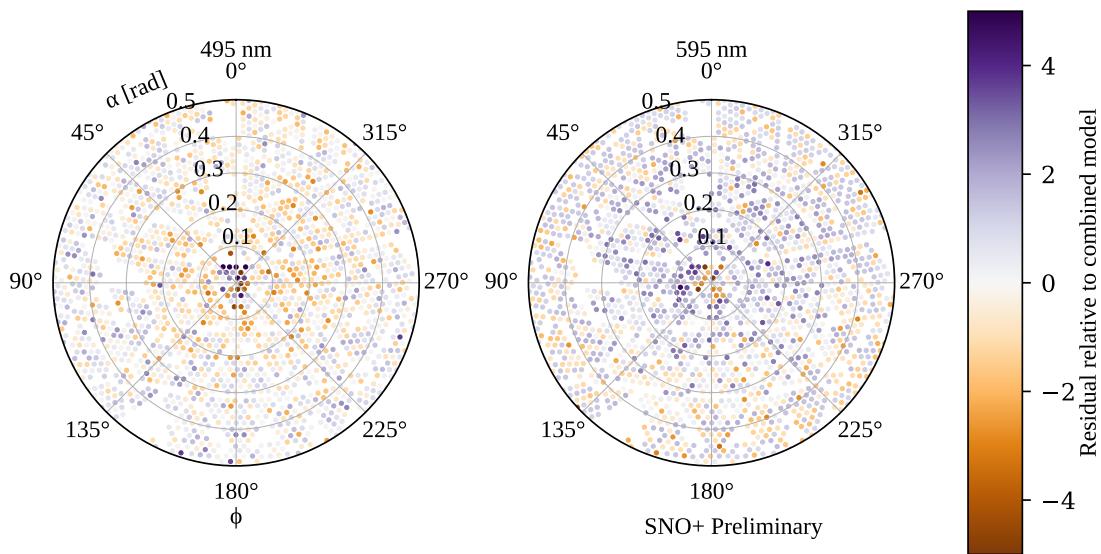


Fig. 4.9: Residuals from subruns at two different wavelengths, both compared to the combined beam profile model for fibre FS055. A negative sign, and hence bluer colours, indicate that the combined model underestimates the observed intensity for that particular subrun. Values with a magnitude beyond 5 are shown capped at this maximal value for the purposes of this plot. These PMTs are plotted in the polar fibre coordinates (α, ϕ).

uncertainty to the beam profiles, beyond the statistical uncertainty as measured by the width of the likelihood distribution. Certainly if one wanted to further improve the uncertainties in the beam profiles, tackling these challenges would be key.

4.3 Comparisons between Data and Simulation

- This focuses on disagreements noticed between data and MC even after the new generator and beam profiles have been used.
- Important to mention that none of these are necessarily game-ending, they are just systematics that may or may not be substantial in a given analysis with SMELLIE..

4.3.1 Forward Hemisphere Discrepancies

- The continued disagreement between data and MC when it comes to measured npe in various parts of the “forward” hemisphere. This includes:
 - The central beamspot,
 - The TIR region,
 - Rope shadows,
 - A noticeable wavelength-dependence.
- This is pretty much most of the contents of Section 4.2.2.

[4 pages]

4.3.2 Emission Time Discrepancies

- For certain lasers, a strong mismatch in the observed hit time residuals for prompt light.
- A mysterious +18 ns bump seen for the PQ495 laser.
- A trigger jitter in the SuperK laser.

[3 pages]

4.3.3 Backward Hemisphere Discrepancies

- The observed distribution of hits vs time and angle in MC does not match data in a number of ways for PMTs near the fibre emission point.
- Includes the outer-water scattering length, rope reflections, and investigations into whether certain modifications to the optics could plausibly fix things (so far, no).

¹ [6 pages]

² [32 PAGES TOTAL]

Chapter 5

Analysis of SMELLIE Data in the Scintillator Phase

This chapter contains two sets of analyses: measurements of the extinction lengths of the scintillator as a function of wavelength and time, as well as monitoring the Rayleigh scattering length over time.

5.1 Extinction Length Analysis

5.1.1 Motivation

- Explain motivating observations for this analysis: a substantial discrepancy between MC and data seen in the radial profiles of nhitsCleaned for ^{210}Po after the PPO top-up campaign. Performed by Serena.
- Hypothesis of a shortening of the absorption/scattering length proposed, further strengthened by Ben Tam's ex-situ absorption measurements with scintillator taken from the detector, as well as knowledge about a likely "cooking" of PPO during the PPO-fill.

- 1 • Describe the provisional optics model decided on based on these measure-
- 2 ments, which includes an additional non-re-emitting component of the
- 3 scintillator.
- 4 • As a further cross-check, SMELLIE should be sensitive to changes in the
- 5 overall extinction length of the scintillator, especially for short extinction
- 6 lengths relative to the size of the detector.
- 7 • More straightforward in measuring extinction length compared to scatter-
- 8 ing length — no need to distinguish between scintillator re-emission and
- 9 scattering.
- 10 • Further uses: can be used to monitor the extinction length over time!

11 [4 pages]

12 5.1.2 Analysis Approach

13 The 2-PMT Case

- 14 • Outline theoretical approach for how one could measure the extinction
- 15 length of scintillator through a comparison of SMELLIE data between the
- 16 scintillator and water phases, in the simplified 1 dimensional case with only
- 17 2 PMTs.

18 [3 pages]

19 Combining Results Between PMTs

- 20 • Not doing analysis with just 2 PMTs, of course! Can combine results from
- 21 multiple PMTs within a beamspot: I explain how here.

22 [2 pages]

Corrections Between the Water and Scintillator Phases

1

- Note the complications that we have to deal with. Namely, the differing refractive indices of the media bending the beamspot differently in the phases, as well as the method used to estimate t_{emm} .
- Explain how we deal with these, the former through MC simulation.

4

5

[2 pages]

6

5.1.3 Validation of the Analysis in Simulation

7

- Show results of this approach being used to measure the extinction length in simulation. How well does it do?

8

9

[3 pages]

10

5.1.4 Results in Data

11

- Describe the data used in this analysis, both water and scintillator, which can be shown in a table.
- Show examples of analysis of data in action for 375 nm data: typical t_{res} distributions of backscattered and beamspot PMTs; calculation of that particular extinction length measurement, followed by the graph for extinction length in 375 nm over all fibres and time periods.
- Discuss what results can be seen in this plot: consistency between fibres, the expected change as a function of PPO concentration, and stability of the extinction length during the main 2.2 g/L scintillator phase.
- Compare results to those made by Ben ex-situ: are they in agreement? If not, what possible systematics could there be? The main one for my

12

13

14

15

16

17

18

19

20

21

22

1 analysis is likely to be uncertainties in the simulated beam profile that leak
2 through into the refractive index correction of the beamspot. For the ex-situ
3 analysis, the value of the extinction length obtained is achieved through
4 background subtraction at some long wavelength, and the particular choice
5 of this wavelength can lead to systematic changes in the obtained extinction
6 length.

- 7 • Look at results at longer wavelengths: can anything reasonably be said at
8 these longer wavelengths? Why/why not?
- 9 • Finally: describe any conclusions that can be reached, in particular whether
10 we can affirm the optics model we use in RAT.

11 [8 pages]

12 5.2 Scattering Analysis

13 5.2.1 Historical Approaches and the Problem of System- 14 atics

- 15 • Comparison to MC is necessary in scattering analysis, compared to merely
16 being needed as a correction factor. This is because of the angular de-
17 pendence of scattering. As a result, we can be far more susceptible to
18 systematics from poor modelling!
- 19 • As a warning, show how Krish's/Esther's approach to the SMELLIE scatter-
20 ing analysis suffers majorly from these systematic effects. Requires describing
21 their analysis approach briefly, and then explaining how the systematics
22 described in Section 4.3 lead to major problems with this approach.

- Motivates the need for either reduced systematics, or an alternative analysis approach that is more robust to them!

[2 pages]

3

5.2.2 New Methodology

4

Signal Region Selection

5

- Propose the new analysis approach: looking at light in the “bad light-path” PMT region. Define what this region is.
- Give qualitative argument for why we expect this region to be robust to the beam profile systematics: dominated by the scattered signal as no direct light can make it here, and changes to beam profile should get “smeared out” after scattering.
- Show how simulations indicate this should be a region with a very high purity of scattered light, and (assuming all else being equal) robust to beam profile uncertainties.
- Confirm robustness of selected PMT region to uncertainties in AV offset and fibre position.

[5 pages]

17

Measuring the Emission Intensity

18

- Remaining systematics is now in the calculation of an average absolute emission intensity.
- Show how various methods don’t work particularly well: whole detector npe, beamspot npe, backscattered light npe, “bad light-path” PMTs but at later times. Explain why it goes wrong for each method.

23

- 1 • Look at "beamspot but excepting the central bit": if that works well, then
2 we can continue!
- 3 • Otherwise, we'll have to live with measuring relative scattering lengths
4 instead of absolute amounts, using the outer water back-scattering as a
5 measure of the relative emission intensity.

6 [4 pages]

7 **5.2.3 Results**

- 8 • Actually do the proposed analysis on data, versus time and wavelength. Do
9 the results seem consistent between fibres? Are they sensible values?

10 [5 pages] [33 PAGES TOTAL]

Chapter 6

1

Solar Oscillation Analysis

2

Driving out into the Sun

Let the ultraviolet cover me up

Looking for a Creation Myth

Ended up with a pair of black lips

3

This is the End

PHOEBE BRIDGERS

Measuring the “solar” neutrino oscillation parameters Δm_{21}^2 and θ_{12} is one of the principal aims of the SNO+ detector during the scintillator phase. There are, in fact, two complementary methods of measuring these parameters: the oscillations of anti-neutrinos from terrestrial nuclear reactors, and the oscillations of neutrinos from the Sun.

4

5

6

7

8

9

10

11

12

13

14

This chapter focuses on the latter approach, using ${}^8\text{B}$ neutrinos coming from the Sun to measure the solar oscillation parameters. An initial background-free study was performed by Javi Caravaca [59], which demonstrated that it was indeed possible to make such a measurement in the detector. The work in this chapter builds substantially from that analysis. This chapter also draws on the associated reactor anti-neutrino analysis built by Iwan Morton-Blake [8], and more broadly

¹ from the general techniques used in the $0\nu\beta\beta$ analysis of Tereza Kroupova [55]
² and Jack Dunger [60].

³ This chapter begins by explaining how it is possible to measure the solar
⁴ oscillation parameters via ^8B events. Then, the framework used to perform the
⁵ analysis is explained: that of a *Bayesian Analysis using Markov Chain Monte*
⁶ *Carlo techniques*. After the method has been described, the dataset upon which
⁷ the analysis is performed is then introduced. The results and associated validation
⁸ are then given. Given these results, a projection is then made for the expected
⁹ sensitivity to θ_{12} as a function of livetime.

¹⁰ 6.1 Analysis Methodology

¹¹ 6.1.1 Observational Principle

¹² How can we measure neutrino oscillation parameters via solar neutrinos in the
¹³ SNO+ detector? As discussed in Chapter 1, it is possible to detect all flavours
¹⁴ of neutrino through elastic scattering with electrons in the detector. If this
¹⁵ interaction was purely neutral-current, then there would be no way of telling
¹⁶ the flavour-state of an interacting neutrino. However, electron neutrinos are
¹⁷ able to interact through an additional charged-current mode. This modifies the
¹⁸ cross-section for electron neutrinos, and means that as the survival probability
¹⁹ for electron neutrinos generated from the Sun, P_{ee} , is modified, the interaction
²⁰ probability of neutrinos with the detector will also. The resulting differential
²¹ interaction rate R for solar neutrinos as a function of the neutrino’s energy, E_ν , is
²² given by:

$$\frac{dR}{dE_\nu} = \Phi_{^8\text{B}} S_\nu(E_\nu) n_e \left[P_{ee}(E_\nu) \sigma_{\nu_e}(E_\nu) + (1 - P_{ee})(E_\nu) \sigma_{\nu_{\mu,\tau}}(E_\nu) \right], \quad (6.1)$$

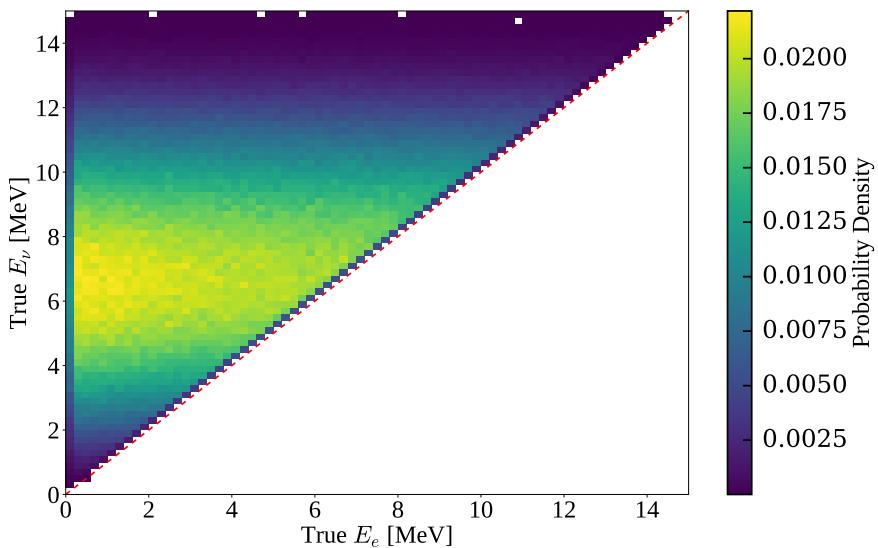


Fig. 6.1: 2D probability distribution comparing the true neutrino energy from a ${}^8\text{B}$ ν_e to the true energy of the scattered electron. Also shown in red is the line $E_\nu = E_e$.

where $\Phi_{{}^8\text{B}}$ is the total flux of ${}^8\text{B}$ solar neutrinos, S_ν is the normalised incident energy spectrum of the solar neutrinos, n_e is the number of electron targets in the detector detection medium, and σ_{ν_i} is the neutrino-electron elastic scattering cross-section for flavour i .

Of course, we do not directly measure neutrino energies in the detector — only the associated scattered electron. If there were no correlation between the observed electron energy and its associated neutrino, then the only effect of neutrino oscillations would be to change the overall observed rate of events due to this process. There would be no change in the shape of the event's energy spectrum, even though neutrino oscillations are a function of neutrino energy. Fortunately, there is some dependence of E_ν on that of the scattered electron, E_e . This dependence can be seen in Fig. 6.1 for ${}^8\text{B}$ electron neutrinos interacting in SNO+. As can be seen, the dependence is weak, and comes mostly from basic energy conservation: If one observes a 10 MeV electron event in the detector, it cannot have come from a 5 MeV neutrino.

In Fig. 6.2 we can see the impact each physical process has on the energy spectrum that we eventually observe. We start with a broad energy distribution

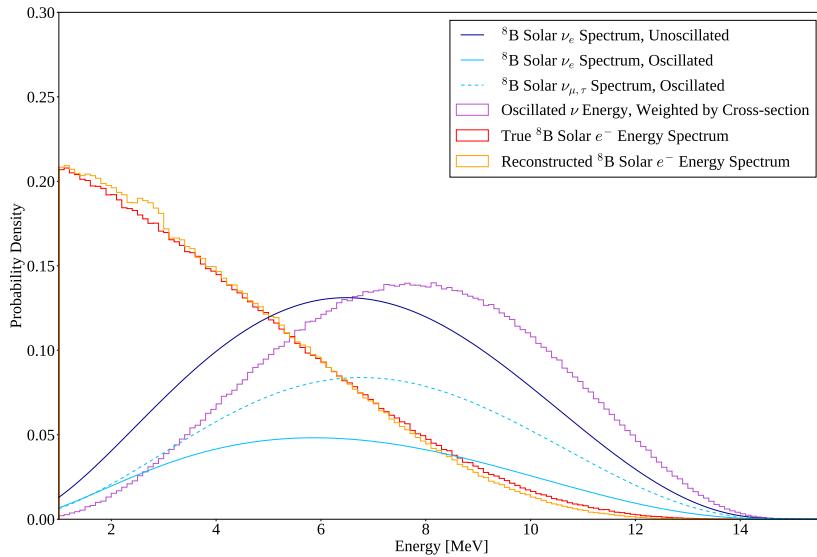


Fig. 6.2: The evolution of energy distributions related to ${}^8\text{B}$ solar neutrino detection. The unoscillated neutrino spectrum is taken from [61]; neutrino oscillations assume oscillation parameters from the current global fit results [62] and ν_e survival probabilities calculated via the method described in Section 6.1.8. The latter three distributions were obtained from MC production as described in Section 6.2.1, with the cross-section formula coming from [63].

1 of ${}^8\text{B}$ electron neutrinos generated in the Sun. These neutrinos then oscillate
 2 their flavour state as they propagate to the detector, in an energy-dependent
 3 manner. When a (tiny) fraction of these neutrinos interact with the electrons in
 4 our detector, there is both an energy- and flavour-dependence on the cross-section.
 5 The scattered electrons gain a kinetic energy with some mild dependence on the
 6 inciting neutrino's energy, which is then measured by the detector to within some
 7 energy resolution.

8 Let us now consider the dependence of P_{ee} on the individual neutrino oscil-
 9 lation parameters. Recall from Eq. 1.2 that, after considering matter-induced
 10 oscillations due to neutrinos passing through the Sun and possibly the Earth,
 11 $P_{ee} = P_{ee}(\tan 2\theta_{12}^M, \sin \theta_{13}^M, \Delta m_{21,M}^2) = P_{ee}(\theta_{12}, \theta_{13}, \Delta m_{12}^2, \Delta m_{13}^2)$. Fig. 6.3 shows
 12 the dependence of each of these four oscillation parameters on $P_{ee}(E)$. We can see
 13 that in reality only the two parameters Δm_{21}^2 and θ_{12} have a substantial impact
 14 on $P_{ee}(E)$ and hence the observed electron energy spectrum. Because of this, for

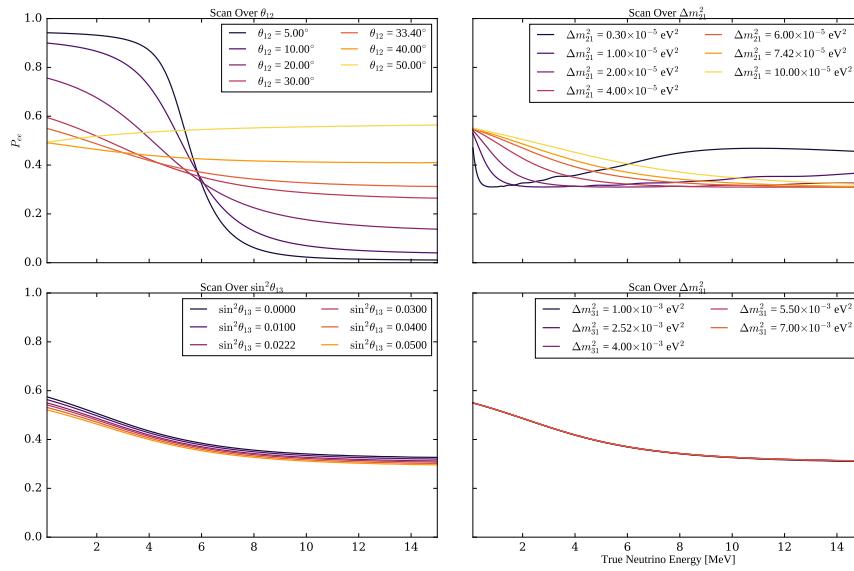


Fig. 6.3: P_{ee} as a function of true neutrino energy, scanned over a variety of oscillation parameter values. For a given oscillation parameter being scanned over, all other oscillation parameters are set at the NuFit 5.1 global fit values. P_{ee} values are calculated using the `PSelmaa` algorithm, as described in Section 6.1.8.

this analysis we will only ever vary these two oscillation parameters, and keep θ_{13} and Δm_{13}^2 at their current “NuFit 5.1” global fit values¹ of $\sin^2 \theta_{13} = 0.0222$ and $\Delta m_{13}^2 = +2.515 \times 10^{-3} \text{ eV}^2$ [62].

6.1.2 Background Processes

Sadly, elastically-scattered electrons from ${}^8\text{B}$ neutrinos are not the only events we see in the SNO+ detector during the scintillator phase. There are a number of background processes that our signal must compete against. Below a reconstructed energy of ~ 2.5 MeV, it is known that various backgrounds completely swamp any possible ${}^8\text{B}$ signal, and so for this analysis we only consider processes that can generate reconstructed energies of at least $E_{\min} = 2.5$ MeV. The following subsections explain each of these backgrounds, as well as methods that have been used to mitigate them as much as possible.

¹We use the global fit results excluding Super-Kamiokande’s atmospheric data, and assuming normal ordering of the neutrino mass hierarchy. This choice has a tiny impact on the magnitudes of these two fixed parameters, the main impact being the sign of Δm_{13}^2 .

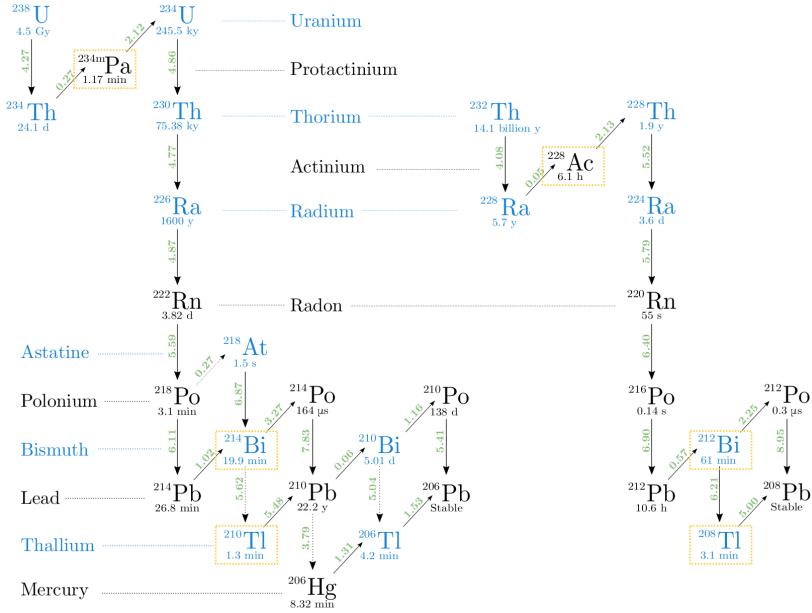


Fig. 6.4: The ^{238}U and ^{232}Th decay chains, taken from [55]. Isotopic half-lives are given below their symbol; the Q-values for each decay, in MeV, is given in green. Downward arrows indicate an α -decay; diagonal arrows indicate β -decay. Isotopes highlighted in gold are potential backgrounds for this solar analysis.

¹ Internal Uranium- and Thorium-Chain Backgrounds

² Although every effort has been made to make the scintillator cocktail that fills
³ SNO+ to be as radio-pure as possible, there inevitably remain trace amounts of
⁴ the radioactive isotopes that derive from the decay chains of the ^{238}U and ^{232}Th
⁵ isotopes. Fig. 6.4 shows these two decay chains. Fortunately, only a fraction of
⁶ the radioactive isotopes in these chains actually are capable of generating events
⁷ in the detector with energies above E_{\min} : these have been highlighted in Fig. 6.4
⁸ in gold.

⁹ Of particular note are the decays of ^{212}Bi and ^{214}Bi . Both are capable of
¹⁰ either α - or β -decays to Tl or Po isotopes, respectively. For the former, it is the
¹¹ subsequent β -decay of the Tl that can have a reconstructed energy above E_{\min} .
¹² For the latter, the Bi decay is the part of the pair of decays that can lie above
¹³ E_{\min} . Although the α -decays here certainly have Q-values well above 2.5 MeV,
¹⁴ the liquid scintillator quenches the observed energy well below E_{\min} . The so-called

“Bi–Po” decays are particularly special because the lifetimes of ^{212}Po and ^{214}Po are 1
300 ns and 164 μs , respectively, which are short enough to allow for highly-effective 2
coincidence tagging. 3

There are two main classes of Bi–Po event in the detector: “out-of-window” 4
(OOW) events for which the Bi and Po occur in separate triggered event windows, 5
and “in-window” (IW) events whereby the Bi and Po occur within the same event 6
window. These lead to two distinct strategies for tagging these kinds of events. For 7
out-of-window Bi–Pos, we look for a delayed coincidence of two events. Using the 8
tagging algorithm suggested in [55, 64] as a starting point, the chosen procedure 9
was as follows. There must be two events that trigger the detector within 4 μs of 10
one another, and both have a valid `scintFitter` position reconstruction within 11
2 m of one another. The delayed candidate event must also have at least 100 12
cleaned PMT hits. This very broad coincidence tagging procedure was designed 13
to ensure that the cut was as *efficient* in tagging (and hence, rejecting) Bi–Pos 14
as possible, whilst negligibly impacting the solar signal. This is in contrast to the 15
cuts chosen by Rafael Hunt-Stokes in [65], which try and obtain a highly *pure* 16
sample of Bi–Po tags. 17

Of course, the above delayed coincidence procedure cannot catch any of the 18
in-window Bi–Po events. For these, we use a different approach. Because two 19
decays happened in the same event window, we expect to see two distinct peaks 20
in the event’s time residual spectrum. In order to look for this event topology, a 21
likelihood-ratio classifier was run over events, first developed by Eric Marzec [66] 22
and re-coordinated for the 2.2 g/L LABPPO scintillator optics by Ziping Ye [67]. 23
This classifier calculates the likelihood ratio between the null hypothesis of a $0\nu\beta\beta$ 24
event (a proxy in this analysis for single-site events such as our ^8B signal) and 25
the alternative hypothesis of an in-window Bi–Po event. The more negative the 26
value of the result, `alphabeta212`, the greater the evidence there is for rejecting 27

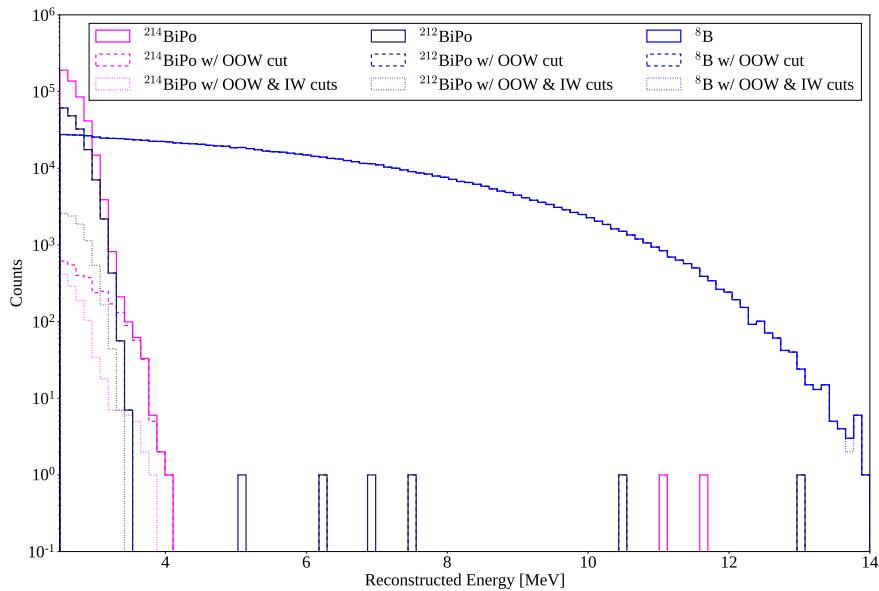


Fig. 6.5: $^{214}\text{Bi-Po}$, $^{212}\text{Bi-Po}$, and ^8B observed energy spectra in MC before and after out-of-window and in-window cuts. The relative normalisations before these cuts have been applied is arbitrary.

the null hypothesis of a single-site event. Events with `alphabeta212 < 0`, or the equivalent for $^{214}\text{Bi-Po}$ events, `alphabeta214`, were then rejected.

Combining both out-of-window and in-window Bi–Po tagging, the impact on $^{212}\text{Bi-Po}$, $^{214}\text{Bi-Po}$, and ^8B ν_e events can be seen in Fig. 6.5. We consider here only events that pass all other cuts used in this analysis (other than the cuts for externals defined shortly): see Section 6.2.2 for the specifics of the cuts used. Because of the different lifetimes of the decays, $^{214}\text{Bi-Po}$ decays predominantly fall out-of-window whilst $^{212}\text{Bi-Po}$ events are typically in-window. This explains why the out-of-window tagging is substantially better at cutting $^{214}\text{Bi-Po}$ decays, whereas the in-window tagging far better tags $^{212}\text{Bi-Po}$ decays. Overall, within the analysis region of interest (ROI), the two combined cuts are able to tag 99.77% of $^{214}\text{Bi-Po}$ triggered events, 94.84% of $^{212}\text{Bi-Po}$ triggered events, whilst retaining 99.85% of ^8B ν_e signal events.

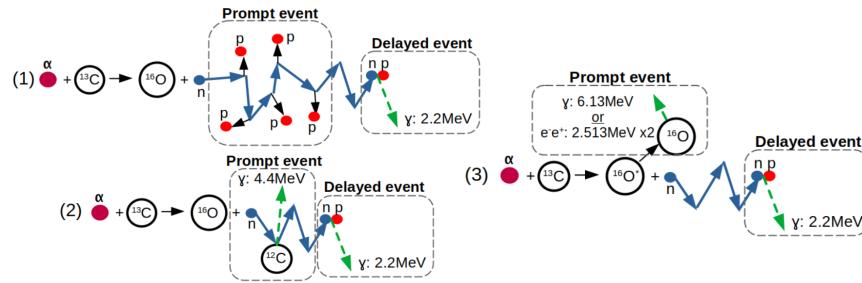


Fig. 6.6: Schematic of the three dominant modes of (α, n) interaction, taken from [8].

(α, n) Reactions

The impact of ^{238}U - and ^{232}Th -chain isotopes does not simply end at their direct decays. It is possible for the α s generated during these decays to undergo their own interactions with nuclei in the detector. Within the organic scintillator of SNO+, the dominant interaction of this type is when an α collides with a ^{13}C nucleus, emitting a neutron: $\alpha + ^{13}\text{C} \longrightarrow ^{16}\text{O} + n$. This is known as an (α, n) reaction.

The topology of this reaction in the detector is a delayed coincidence, as shown in Fig. 6.6. The prompt signal can be generated through a number of processes, including the decay of an excited ^{16}O state. The neutron generated in the interaction then thermalises and gets captured by another nucleus — usually hydrogen in SNO+ — which creates an excited state that then eventually decays, creating a γ that creates the delayed signal in the detector.

As can be seen in Fig. 6.7, (α, n) interactions can lead to events reconstructed at a wide variety of energies, which could be an issue for this analysis. However, because they are delayed coincidence events with a typical decay time of ~ 100 ns, the aforementioned out-of-window and in-window Bi–Po tagging algorithms also efficiently tag (α, n) events. Looking again at Fig. 6.7, simply by using the out-of-window and in-window Bi–Po taggers without any further modifications 99.37% of events in the ROI are cut.

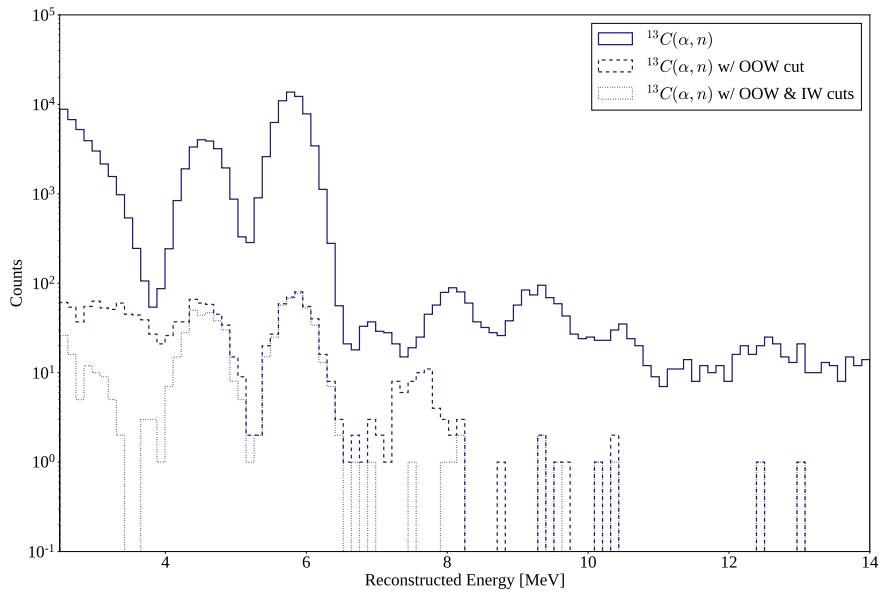


Fig. 6.7: (α, n) reconstructed energy spectrum in MC within the analysis ROI, before and after out-of-window cuts have been applied. The relative normalisation of the distribution before the additional cuts have been applied is arbitrary.

1 External Backgrounds

2 All materials within the SNO+ detector are radioactive, not just the liquid
 3 scintillator cocktail. This includes the acrylic, ropes, external water, and PMTs.
 4 These components have had their radiopurity “assayed” (that is, measured)
 5 throughout the detector’s lifetime, often back to the construction of the original
 6 SNO detector itself. The materials other than the liquid scintillator are known to
 7 have far higher background levels, especially in the important ^{238}U - and ^{232}Th -
 8 chain backgrounds [68]. To distinguish between the inherent backgrounds within
 9 the scintillator, and the backgrounds from materials at larger radii, we use the
 10 terminology “internal” and “external”, respectively.

11 Although there are numerous external backgrounds, with a suitably accurate
 12 and precise position reconstruction algorithm they can be suitably handled. The
 13 simplest approach is with a so-called “fiducial volume” (FV) cut: just throw out
 14 all events that reconstruct beyond some radius. The only external background
 15 events that will reach within the FV are those that have reconstructed very poorly,

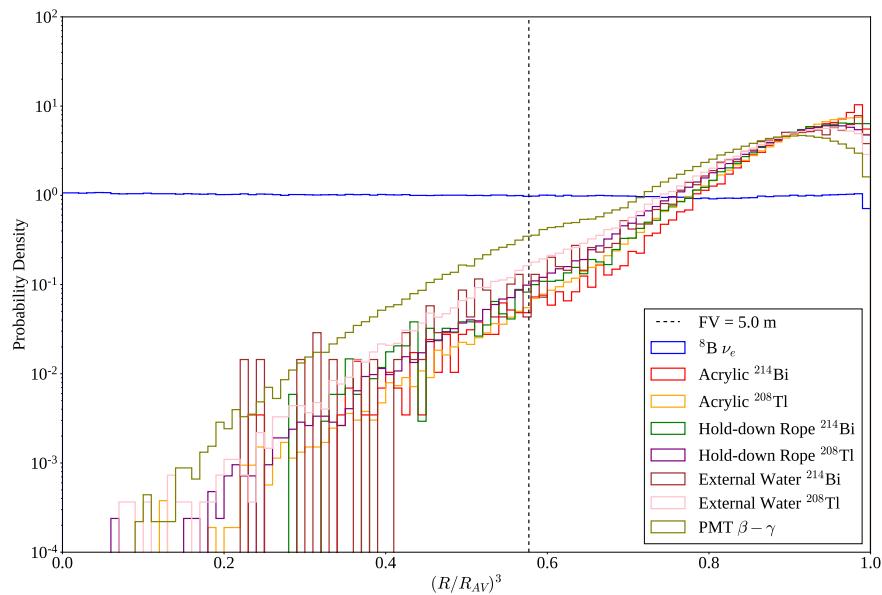


Fig. 6.8: Radial dependence of external backgrounds considered in this analysis, as compared to the ${}^8\text{B}$ signal. Events for all processes shown here had to pass the first 6 cuts of Table 6.1.

or have some long-distance radiation that manages to deposit radiation close to the centre of the AV. Because α and β radiation can only travel short distances through the detector, it is only γ radiation that can realistically travel far enough into the detector to be able to reconstruct anywhere near the centre. Moreover, the intensity of this γ radiation attenuates exponentially towards the centre of the detector, meaning only a tiny fraction of the total number of external events reconstruct within a 3.5 m FV, say. This strong radial-dependence can be seen in Fig. 6.8.

What this figure also demonstrates is that our solar signal has a completely different radial dependence to these backgrounds. As a result, if one considers not just the energy of events but also their reconstructed radius, then it is possible to get an additional handle on the external backgrounds. The FV cut can then be pushed further out to larger radii, allowing one to gain more signal statistics.

Work by Tereza Kroupova [55] allows for additional means of distinguishing external backgrounds from the solar signal. The underlying assumption in the

reconstruction of SNO+ events is that there was an electron at a single point, which is entirely valid for ^8B elastic scattering events. However, external backgrounds can fail this assumption in two ways. Firstly, these radioactive decays often generate γ radiation in addition to the main α/β particle, which creates a multi-site event. Because the `scintFitter` position reconstruction algorithm is not prepared for a distribution of energy depositions in the scintillator, the t_{res} distribution will broaden. This allows an event classifier to be built that distinguishes between the t_{res} distributions of single-site events and externals, known as the “external background timing classifier” (`ext0NuTimeTl208AVNaive`).

Secondly, because external backgrounds that do reconstruct close to the centre of the detector typically have a γ that travelled a long distance towards the centre of the detector, we expect the earliest light that hits the PMTs to arrive most often along the direction of the reconstructed position vector. A distribution of PMT hits for a given event as a function of their angular distribution relative to the direction of position reconstruction can be built, and compared to the expected distributions for single-site and external background events. This is known as the “external background topological classifier” (`ext0NuAngleTl208AV`). Much like the classifier described in Section 6.1.2, the single-site events used for comparison were $0\nu\beta\beta$ events, but these have a similar single-site structure to the solar signal of interest in this analysis.

Fig. 6.9 shows the correlation between the two classifier results for both acrylic ^{208}Tl and $^8\text{B} \nu_e$ events in MC. These events were within the analysis ROI and also below 5 MeV, as external backgrounds never reconstruct above this energy. The other external backgrounds in MC have a similar distribution to that of the acrylic ^{208}Tl .

As can be seen, the external background is far more spread out along both axes compared to the signal. This allows for discrimination of external backgrounds from the signal in the 2.5–5.0 MeV range. For this analysis, a cut on both discriminants

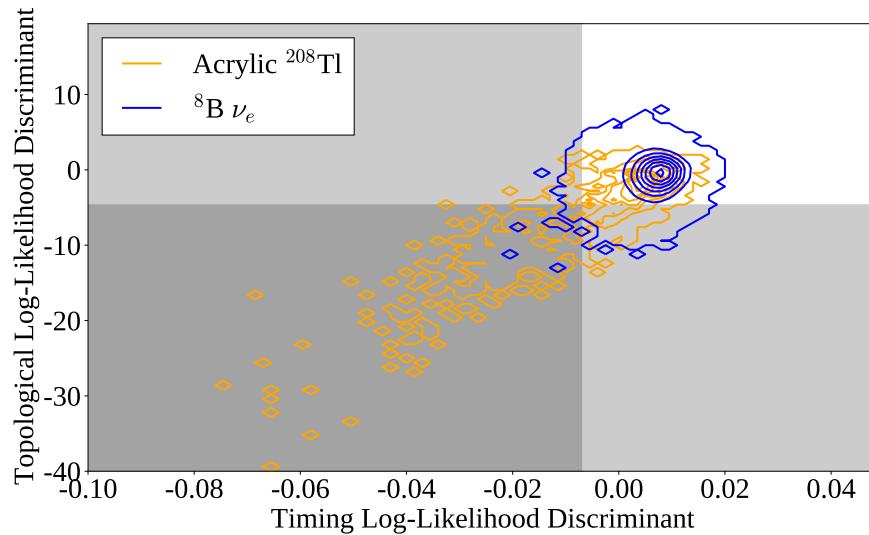


Fig. 6.9: Distributions of the two external classifiers, for both the ${}^8\text{B}$ signal and acrylic ${}^{208}\text{Tl}$ background MC. Contours indicate lines of equal probability density. All events shown were not just in the analysis ROI, but also below 5 MeV. Grey regions indicate the cuts used in this analysis.

were used: to pass, an event required `ext0NuTimeTl208AVNaive > -0.007` and `ext0NuAngleTl208AV > -4.7`. By using these cuts, 37.4% of external acrylic ${}^{208}\text{Tl}$ were cut in the analysis ROI in MC, whilst keeping 99.3% of the ${}^8\text{B} \nu_e$ signal.

6.1.3 The Log-likelihood Test Statistic

At the highest level, this analysis involves taking the data observed in the scintillator-fill after applying a certain set of cuts, along with simulated PDFs for all processes believed to build up the observed data with those same cuts applied, and then attempting to fit the combined energy and radial distributions of the MC to that of the data. Given a set of PDFs, to try and match the distribution of observables in data we can modify a number of parameters. These consist of the normalisations of each PDF (i.e. the total number of events observed due to that process), and any systematic parameters that could modify the shapes of these distributions. For this analysis, the neutrino oscillation parameters act as *de facto* systematic parameters, as they modify the shape of the ${}^8\text{B}$ PDFs. Of course,

¹ unlike usual systematics the oscillation parameters are what we are actively trying
² to measure.

³ The test statistic used for this analysis is the *binned extended log-likelihood*:

$$\mathcal{L}(\mathcal{N}, \boldsymbol{\theta} | \mathbf{n}) = - \sum_{i=1}^{N_{\text{PDFs}}} \mathcal{N}_i + \sum_{j=1}^{N_{\text{bins}}} n_j \ln \left(\sum_{i=1}^{N_{\text{PDFs}}} \mathcal{N}_i P_{ij}(\boldsymbol{\theta}) \right), \quad (6.2)$$

⁶ where \mathcal{N}_i is the normalisation parameter for the i^{th} PDF out of N_{PDFs} in total,
⁷ n_j is the number of observed events in the j^{th} bin of data out of N_{bins} total bins,
⁸ and $P_{ij}(\boldsymbol{\theta})$ is the probability of observing an event of type j in bin i , assuming a
⁹ set of systematic parameters $\boldsymbol{\theta}$. \mathcal{N} and \mathbf{n} are the set of normalisation parameters
¹⁰ and observed bin values, respectively.

¹¹ 6.1.4 The Bayesian Statistical Approach

¹² There are two main schools of statistical inference, “Frequentist” and “Bayesian”.
¹³ In the former, probabilities describe the fraction of times a situation can be found
¹⁴ within the whole ensemble of possible worlds. For the latter, we care not about an
¹⁵ ensemble of worlds but instead our degree of belief about matters in this current
¹⁶ one. We update our beliefs as we acquire knowledge about the world through
¹⁷ Bayes’ Theorem:

$$P(\boldsymbol{\mu} | \mathbf{x}) = \frac{\mathcal{L}(\boldsymbol{\mu} | \mathbf{x}) P(\boldsymbol{\mu})}{P(\mathbf{x})}. \quad (6.3)$$

¹⁹ Here, $\boldsymbol{\mu}$ is the set of parameters that model our system, $P(\boldsymbol{\mu})$ is our *prior* (pre-
²⁰ existing) distribution for those model parameters, and \mathbf{x} is the data taken in
²¹ our experiment. The updated, *posterior* distribution $P(\boldsymbol{\mu} | \mathbf{x})$ is then the prior
²² multiplied by the likelihood of parameters $\boldsymbol{\mu}$ given observations \mathbf{x} , $L(\boldsymbol{\mu} | \mathbf{x})$, and
²³ divided by the total probability $P(\mathbf{x})$ of observing \mathbf{x} under any circumstance.

²⁴ Both approaches to statistics are widely-used in statistical analysis, in both
²⁵ particle physics and beyond. The Bayesian approach was used for this analysis,

as it was believed that this helps keep transparent what assumptions are being
made in the analysis.

Now, if one is able to determine the overall posterior distribution, then it is
possible to derive best-fit values with uncertainties for all parameters in the fit.
This is done by “marginalising” the posterior distribution, i.e. integrating over
all parameters other than the one of interest. A sensible best-fit value is then
the point of Highest Posterior Density (HPD). The uncertainty on this value is
derived from the spread of the marginalised posterior, by the calculation of the
 1σ Credible Interval (CI): this is the set of values for a given parameter which has
a total posterior probability of 68.3%, and contain the best-fit value. There are
in fact an infinite number of CIs that satisfy this property; for this analysis, the
values are chosen in decreasing order of marginalised posterior probability density.

6.1.5 Markov Chain Monte Carlo

Of course, all of this assumes that one can accurately determine the posterior
density distribution. Whilst the likelihood and prior distribution are straight-
forward enough to calculate, often-times $P(\mathbf{x})$ (which acts as a normalisation)
is very challenging to determine. This is because calculating this normalisation
involves integrating the likelihood over all the parameter space, and if there are
a large number of parameters this can become enormously numerically complex.
An alternative approach is needed!

That alternative comes in the form of *Markov Chain Monte Carlo*, MCMC. A
Markov Chain is any mathematical system for which the next state of the system
is dependent only on its current state; the system is in some sense “memoryless”.
For a large class of Markov Chains — those that are “ergodic” and “aperiodic” —
one can prove that regardless of the initial position on the chain, the probability
distribution converges to the same distribution [69]. MCMC uses such a Markov

¹ Chain which attempts to converge towards the posterior density distribution in
² particular. In MCMC, after choosing the initial position in the parameter space,
³ successive states are chosen at random with a probability dependent only on the
⁴ properties of the current position in parameter space and the proposed position.
⁵ The convergence property of Markov Chains means that the set of steps made in
⁶ the parameter space will have a distribution that converges to that of the posterior
⁷ density distribution.

⁸ There are a number of MCMC algorithms, and the particular one used in this
⁹ analysis is that of the *Random-Walk Metropolis Algorithm*. In this algorithm,
¹⁰ given an initial position in the parameter space $\boldsymbol{\mu}$, a new step is proposed, $\boldsymbol{\mu}'$.
¹¹ This step is chosen at random from a multivariate Gaussian distribution centred
¹² on the current position, with widths in each dimension of the parameter space
¹³ chosen beforehand as constants for tuning the MCMC process. This choosing of a
¹⁴ new proposed step at random is what gives the algorithm its Monte Carlo and
¹⁵ Random Walk titles. Once a new step is proposed, it is accepted as the new step
¹⁶ with a probability $S(\boldsymbol{\mu}'|\boldsymbol{\mu})$ according to the condition of *detailed balance*:

$$\begin{aligned} S(\boldsymbol{\mu}'|\boldsymbol{\mu}) &= \min\left(1, \frac{P(\boldsymbol{\mu}'|\mathbf{x})}{P(\boldsymbol{\mu}|\mathbf{x})}\right) = \min\left(1, \frac{L(\boldsymbol{\mu}'|\mathbf{x}) P(\boldsymbol{\mu}')}{L(\boldsymbol{\mu}|\mathbf{x}) P(\boldsymbol{\mu})}\right) \\ &= \min\left(1, \exp\left[\mathcal{L}(\boldsymbol{\mu}'|\mathbf{x}) - \mathcal{L}(\boldsymbol{\mu}|\mathbf{x}) + \ln \frac{P(\boldsymbol{\mu}')}{P(\boldsymbol{\mu})}\right]\right). \end{aligned} \quad (6.4)$$

²⁰ It is the detailed balance condition that ensures convergence of the MCMC
²¹ algorithm to specifically the posterior density distribution. Crucially, because
²² it is only dependent on the ratio of posterior densities, the hard-to-calculate
²³ normalisation $P(\mathbf{x})$ in both posterior density terms cancels out, meaning one only
²⁴ needs to calculate the likelihood and priors for each step.

²⁵ The specific implementation of MCMC used for this analysis is that of **OXO**,
²⁶ a C++ analysis framework first developed by Jack Dunger [60] and built on for

this analysis. `OXO` is able to run the Metropolis algorithm on multidimensional binned data, using the log-likelihood defined in 6.1.3. This framework also allows one to include systematic parameters that can float within the fit, and define non-uniform priors for normalisations and systematics that have constraints.

6.1.6 Choosing Priors

For this analysis, the suggestions made by Biller & Oser in [70] about choosing prior distributions are followed: for parameters that do not have some pre-existing constraint, a flat prior is used. A nice consequence of this choice is that $\ln \frac{P(\mu')}{P(\mu)} = 0$, so the actual value of the prior for these variables never needs to be calculated when running the MCMC algorithm. For this analysis, uniform priors are assumed on the neutrino oscillation parameters Δm_{21}^2 and θ_{12} , as the magnitudes of these parameters are now well-established.

For parameters with existing asymmetric constraints $\beta_{-\sigma_-}^{+\sigma_+}$, this analysis uses an asymmetric Gaussian prior, equivalent to the logarithm of the prior being an asymmetric quadratic:

$$\ln P(\mu) = \mathcal{A} - \begin{cases} \frac{(\mu-\beta)^2}{2\sigma_+^2} & \text{if } \mu \geq \beta, \\ \frac{(\mu-\beta)^2}{2\sigma_-^2} & \text{if } \mu < \beta. \end{cases} \quad (6.5)$$

Here, \mathcal{A} is the logarithm of the prior's normalisation constant, and cancels out in the detailed balance condition. For parameters with symmetric constraints, $\sigma_+ = \sigma_-$, then $\ln P(\mu)$ reduces to a quadratic with maximum at $\mu = \beta$.

Fit parameters often also have basic physical limits on what values they can hold. Normalisation parameters, for example, must be positive. We also restrict the oscillation parameters to be in the very broad range $3.0 \times 10^{-6} \text{ eV}^2 \leq \Delta m_{21}^2 \leq 1.0 \times 10^{-3} \text{ eV}^2$ and $5^\circ \leq \theta_{12} \leq 65^\circ$. The prior density beyond these limits is zero,

¹ meaning that any proposed step outside the allowed region can be immediately
² rejected.

³ 6.1.7 Including Systematics in the Fit

⁴ One important implementation detail is how systematics are applied within the
⁵ MCMC fitting process. Once systematics are added to the fit, at every step the
⁶ binned PDFs for all the processes considered in the fit must get modified appropri-
⁷ ately, which can become extremely computationally-intensive if not approached
⁸ carefully. The strategy used in the `OXO` framework starts by thinking of the
⁹ contents of a binned PDF as a vector of bin probabilities, $\mathbf{p} = (p_1, p_2, \dots, p_{N_{\text{bins}}})^T$.
¹⁰ Then, we can think of a systematic acting on the PDF as a linear transformation,
¹¹ and hence a matrix S acting on this vector: $\mathbf{p}' = S\mathbf{p}$. We only need to calculate
¹² this matrix once for a given set of systematic parameter values, and can then
¹³ use the same matrix on all the PDFs in the fit. Furthermore, when multiple
¹⁴ systematics are applied, the matrix for each systematic can then be combined via
¹⁵ matrix multiplication into one single “detector response” matrix. `OXO` uses the
¹⁶ `Armadillo` [71, 72] linear algebra package for efficient matrix manipulation.

¹⁷ There is a problem that can arise when considering the impact of systematics
¹⁸ near the edge of the analysis ROI. Many systematics such as shifts, scalings,
¹⁹ and convolutions use information about the contents of nearby bins to determine
²⁰ the contents of a particular bin. However, for bins near the edge some of that
²¹ information does not exist — it has been lost to the cuts that define the ROI.
²² This can lead to a bias in the generation of the modified PDFs, and therefore also
²³ the posterior distribution.

²⁴ As an example, consider the impact of an energy scale systematic on the
²⁵ energy distribution of PMT $\beta - \gamma$ events in the detector, shown in Fig. 6.10.
²⁶ Because the events seen for this process in the ROI are merely the high-energy

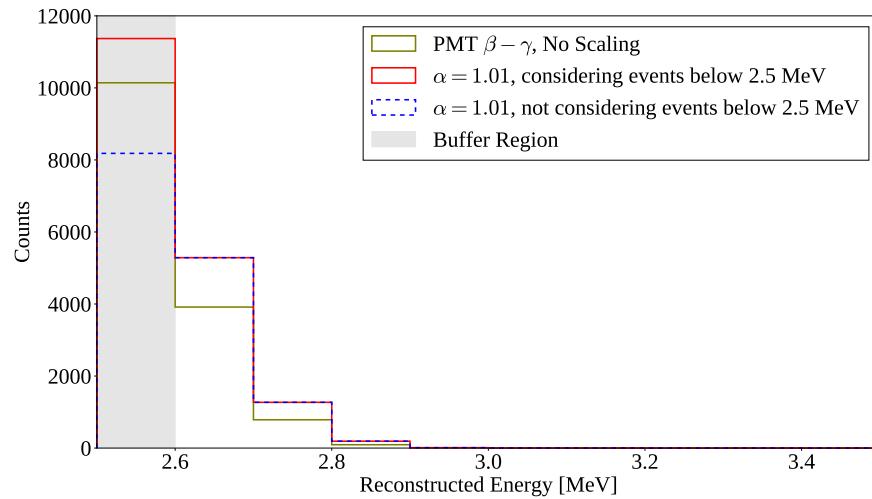


Fig. 6.10: Demonstration of how to handle an energy scaling systematic correctly. The distribution of PMT $\beta - \gamma$ events in the analysis ROI is shown in energy, both before and after an energy scaling of $\alpha = 1.01$ has been applied. If the distribution below 2.5 MeV is not known, then the scaled distribution is biased systematically. This is solved through the use of a buffer region, as indicated.

tail, any systematic energy scaling $E'_{\text{reco}} := \alpha E_{\text{reco}}$ should have a large impact on the number of events observed at the low end of the ROI. However, given that the information about data below E_{\min} is lost to the ROI cuts, any energy scaling of $\alpha > 1$ will not be applied correctly at all.

The solution to this problem is defining a “buffer region” of bins on either side of the ROI, which allow for tracking of events in and out of the ROI due to systematics, but aren’t considered when calculating the likelihood. This is also shown in Fig. 6.10. After the scaling systematic is applied, although incorrect bin values are found in the buffer region, this is fine because we are no longer calculating the likelihood with those bins. Note that because of this modification, the normalisation parameters we put into the model no longer represent the expected number of events in the ROI. Instead, they represent the number of events expected in both the ROI and buffer region, before any systematics have been applied.

6.1.8 Including Oscillations in the Fit

Within the analysis MCMC code, the process of neutrino oscillations are thought of as a *de facto* systematic that acts only on the ${}^8\text{B}$ ν_e and ν_x signal spectra. Three parameters relevant to the signal are floated within the MCMC fit: Δm_{21}^2 , θ_{12} , and $\varphi_{{}^8\text{B}} = \Phi_{{}^8\text{B}}^{\text{meas}} / \Phi_{{}^8\text{B}}$, the unoscillated ${}^8\text{B}$ flux relative to the expected rate. For the two signal PDFs, a third “bookkeeping” dimension is added on top of reconstructed energy and radius: the true neutrino energy, E_ν . This is necessary for correctly applying oscillations, as the oscillation probability is a function of the neutrino’s energy, not the scattered electron’s. Before the fit, these 3D PDFs are given normalisations corresponding to the expectation of the number of events for each type, ν_e and $\nu_{\mu,\tau}$, after cuts but before oscillations have been applied. Strictly speaking there should be zero $\nu_{\mu,\tau}$ events before neutrino oscillations: the pre-oscillation rate used here is the post-cut number of events expected if 100% of the neutrinos oscillated to the $\nu_{\mu,\tau}$ type.

During the MCMC fit, for a given set of parameters $\boldsymbol{\theta} = (\Delta m_{21}^2, \theta_{12}, \varphi_{{}^8\text{B}})$ the following is performed to oscillate the signal PDFs. Firstly, the normalisations are scaled by the factor $\varphi_{{}^8\text{B}}$. Then, for each E_ν bin the survival probability $P_{ee}(E_\nu, \Delta m_{21}^2, \theta_{12})$ is calculated. Each bin then has their probability scaled by either P_{ee} or $1 - P_{ee}$, for ν_e and $\nu_{\mu,\tau}$ respectively. Of course, within the structure of the **OXO** framework these bin-by-bin scaling aren’t immediately applied, but instead a matrix describing the impact of oscillations on each of the PDFs is made. Because the oscillation transformation is purely a bin-by-bin scaling, the resulting matrices are diagonal, with diagonal elements $\varphi_{{}^8\text{B}} \cdot P_{ee}(E_\nu, \Delta m_{21}^2, \theta_{12})$ or $\varphi_{{}^8\text{B}} \cdot (1 - P_{ee}(E_\nu, \Delta m_{21}^2, \theta_{12}))$ for ν_e and $\nu_{\mu,\tau}$ respectively. After the oscillation matrix along with all other systematic matrices are applied to the signal PDFs, the PDFs are then marginalised over the E_ν dimension so that the signal PDFs match the dimensionality of all other PDFs.

Calculations of the survival probability are handled with **PSelmaa**, an algorithm written by Nuno Barros for the SNO 3-phase Analysis [73]. This considers not only the neutrino oscillations through the vacuum of space between the Sun and Earth, but also the impact of matter effects in both the Sun and Earth. This can usually be a very computationally-intensive process, but **PSelmaa** takes advantage of the assumption that the solar oscillation parameters are in the so-called “Large Mixing Angle” regime, making the calculation much faster. As seen in Section 1.2.1, previous solar oscillation experiments demonstrate that this assumption is reasonable. For this analysis, the standard MSW effect is assumed with neutrinos obeying the Normal Hierarchy, with the Sun following the B16_GS98 metallicity model [74] and the PREM model being used for the Earth [75].

One final thing **PSelmaa** needs to know to calculate survival probabilities is the distribution of solar zenith angles during the data-taking. The solar zenith θ_z is the angle between the two following vectors: one going from the centre of the Earth through the centre of the SNO+ detector, and another starting from the detector’s centre and pointing towards the Sun. As an example, if the Sun were ever to be directly above the detector, both vectors would be along direction \hat{z} in detector coordinates, leading to a solar zenith angle of $\theta_z = 0$. The position of the SNO+ detector on Earth, as well as the times at which the detector was live, determine the solar zenith angle distribution. If not accounted for, this can lead to a bias in the result of the analysis, as a preponderance of livetime taken at night (say) would lead to a larger fraction of solar neutrinos having to pass through the bulk of the Earth to get to the detector, and hence the impact of the MSW effect would be greater.

Even after using the Large Mixing Angle approximation, having to call **PSelmaa** numerous times for every step in the MCMC algorithm would lead to exorbitant run times for the fitting. Therefore, a further approximation is made. Before running the MCMC fit, **PSelmaa** is used to calculate P_{ee} over the necessary 3D

¹ space of parameters. To get a fine scan of this space, 101 E_ν values from 1 MeV
² to 20 MeV, 101 Δm_{21}^2 values from 3×10^{-6} eV² to 1×10^{-3} eV², and 151 values
³ for θ_{12} from 5° to 65° were looked over. This 3D grid of $101 \cdot 101 \cdot 151$ P_{ee} values
⁴ is then written to disk, and loaded into memory for use during the fit as a lookup
⁵ table. At run-time, as the Metropolis-Hastings algorithm samples this 3D space
⁶ the survival probability is estimated through a trilinear interpolation of the 3D
⁷ grid loaded in: a version of linear interpolation for three dimensions.

⁸ 6.2 Analysis on Scintillator-Phase data

⁹ 6.2.1 Dataset and Livetime

¹⁰ The data used in this analysis was chosen to be scintillator phase data after the
¹¹ end of the PPO top-up campaign that completed in April 2022. Following an
¹² initial validation of the analysis tools was performed on data between 29th April
¹³ and 10th May 2022 [76], data for the full analysis was chosen from runs taken
¹⁴ after these dates. Not all data taken during this time was considered usable
¹⁵ for this analysis, however: this is the role of *run selection*. The Collaboration’s
¹⁶ ‘Preliminary Scintillator Gold’ run selection list was used as the basis for this
¹⁷ analysis. This run list requires:

- ¹⁸ • The run type must be in ‘Physics’, as opposed to running in calibration or
¹⁹ maintenance;
- ²⁰ • The run must last at least 30 minutes;
- ²¹ • Detector electronics must be working in a stable manner without any alarms,
²² and with all crates online;
- ²³ • There are no abnormal rates of tagged muons, and the OWL PMTs are
²⁴ correctly functioning;

- There are no unusual conditions from e.g. earthquakes, blasting activity in the mine, or loss of power.

In the end, data for this analysis used ‘Gold’ runs selected between 17th May–30th November 2022, run numbers 300733–306498. The raw livetime associated with this dataset was calculated by looking at the start and end times of each run using the detector’s 10 MHz clock: 84.977 days.

Importantly, the livetime of the data actually used in the analysis ends up being somewhat less than this raw value. This is because two of the cuts used in the analysis (described in Section 6.2.2), the muon tagger and the `high-nhit` event tagger, also veto events for a time period following a tagged event. This means that any events in such a time window are automatically thrown out of consideration for analysis. An algorithm was written to determine the loss of livetime from both of these tagging processes for all the runs selected in this analysis, allowing for the net livetime to be calculated. To ensure accuracy in the value of this lost livetime, the algorithm took care to handle any overlaps in the time veto windows. This ends up being a quite common occurrence, as tagged muon events and their followers often have a very high `nhit`. The net livetime was calculated to be 80.615 days. Data processing and simulations used RAT versions 7.0.8 and 7.0.9.

As discussed in Section 6.1.8, in order to account for the impact of the MSW effect through the Earth on P_{ee} , the solar zenith angular distribution of the dataset is needed. This was achieved by taking the recorded trigger time of each event within the dataset as given by the GPS-calibrated 10 MHz clock, determining the position of the Sun at that time, and then deriving the value of $\cos \theta_z$ given that solar position. The resulting distribution of $\cos \theta_z$ is shown in Fig. 6.11. The structure of this distribution are a result of the times of day and year of the runs selected for this analysis.

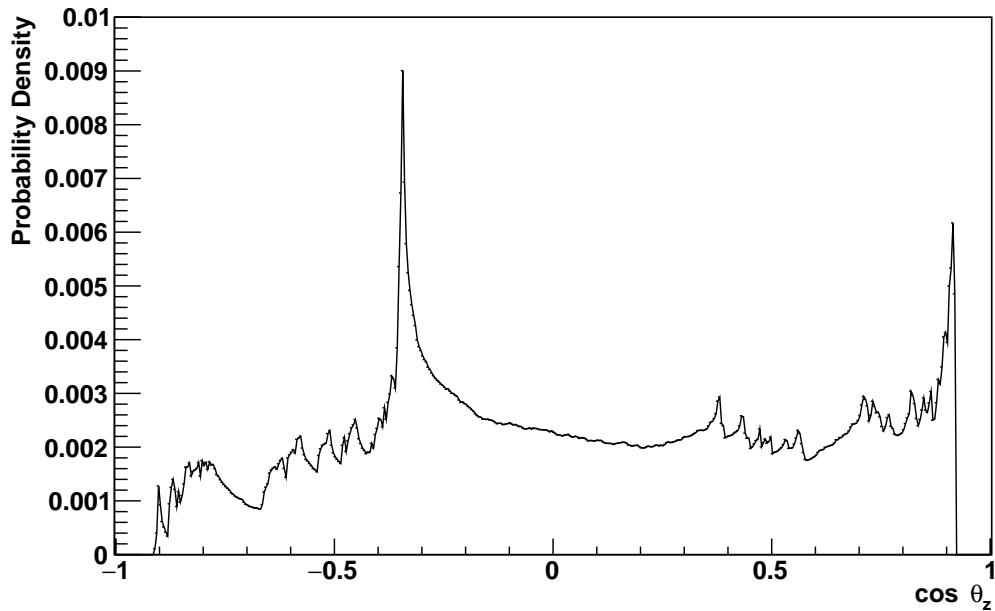


Fig. 6.11: Distribution of the cosine of the solar zenith angle, $\cos \theta_z$, for the events within this dataset. If an event has $\cos \theta_z > 0$, then at the time of that event the position of the Sun is above the detector's equator.

¹ 6.2.2 Event Selection

² Once the runs of processed data and matching MC have been selected, the next
³ step was to perform cuts on both in order to obtain our analysis ROI. The full list
⁴ of cuts used are shown in Table 6.1, most of which have already been explained in
⁵ Section 6.1.2.

⁶ Most of the other cuts used in this analysis that haven't yet been discussed
⁷ come from handling differences between data and MC. In MC, it is possible to
⁸ simulate an event but not have it trigger the detector. In this case, the stored
⁹ index of that event (`evIndex`) will be less than zero. We never see these 'events'
¹⁰ in data of course, but they do need to be removed from simulation. On the flip
¹¹ side, there are types of events that we observe in the detector but do not simulate.
¹² Many of these events are caught during 'data cleaning', which looks for a wide
¹³ variety of problems in the data when the data is being processed. Any subset
¹⁴ of the checks made during data cleaning can be chosen to consider in a given

| Cut Description | Cut |
|---|---|
| Event triggers detector | <code>evIndex ≥ 0</code> |
| Pass data cleaning | <code>ANALYSIS_MASK = 0x2100000042c2</code> |
| High-nhit veto | <code>correctedNhits ≥ 5000; veto for 20 s</code> |
| <code>ScintFitter</code> used | <code>scintFit = true</code> |
| Results of <code>ScintFitter</code> valid | <code>fitValid = true</code> |
| In energy Region of Interest | $2.5 \text{ MeV} < E < 14.0 \text{ MeV}$ |
| Fiducial Volume cut | $R < 5.0 \text{ m}$ |
| Remove BiPo Out-of-window tags | See Section 6.1.2 |
| Remove BiPo In-window tags | Tagged if $\text{alphaBeta212} > 0$ and $\text{alphaBeta214} > 0$ Tagged if $E < 5.0 \text{ MeV}$, |
| Remove External tags | $\text{extONuTimeTl208AVNaive} > -0.007$, and $\text{extONuAngleTl208AV} > -4.7$ |

Table 6.1: Cuts used in this analysis.

analysis by defining a specific ‘analysis mask’; a number that specifies for each of its binary digits whether to consider the data cleaning process associated with that bit. The mask chosen used in this analysis is the Collaboration’s current standard for the scintillator phase, 0x2100000042c2, which corresponds to:

- **00 cut:** Tags every event which has a GTID ending in 00 in hexadecimal, which is associated with a long-standing issue in the detector triggering system.
- **Junk cut:** Tags events that have channels with multiple recorded hits for that event. These are associated with so-called ‘orphaned’ hits that the event building process doesn’t know what to do with.
- **OWL cut:** Tags events that have at least three OWL PMTs hit, a sign that the event came from a cosmic-ray muon.
- **Polling cut:** Tags all events that occur during ‘polling’, a detector monitoring process.
- **CAEN cut:** Tags all events that have missing CAEN trace data.

- ¹ • **Muon cut:** Tags muon-like events.
- ² • **Muon follower cut:** Tags all events coming up to 20 s after a tagged muon event.

⁴ Beyond the formal data cleaning cuts, a high-nhit veto was used to tag and
⁵ remove various instrumental backgrounds which can generate large amounts of
⁶ light in the detector. This method also catches any muons left untagged by the
⁷ data cleaning mask. After any event with a sufficiently large number of hits, both
⁸ that event and any following it for 20 s are removed. This method is likely to be
⁹ fairly conservative, and there exist different methods for cutting flashers within
¹⁰ data cleaning. However, these were originally coordinated for the water phase,
¹¹ and had not been updated for the scintillator phase at the time of writing.

¹² The results of all these cuts can be seen for the data in Table 6.2. 654 out of
¹³ the 18,112,495,770 events within the dataset pass all cuts. These same cuts were
¹⁴ also run over the MC, with results summarised in Table 6.3. Important to note
¹⁵ was the difference between the number of physics events simulated for a given
¹⁶ process and the number of triggered events. The latter ends up being larger than
¹⁷ the former for all processes discussed here, because a given physics event in the
¹⁸ detector will often generate a retrigger event following the primary triggered event.
¹⁹ Also shown in this table are the combined cut efficiencies for each process. This
²⁰ efficiency is defined as the number of triggered events that pass all cuts, divided
²¹ by the number of simulated physics events for that process.

²² Variable bin widths in energy were chosen for this analysis: 0.1 MeV bins
²³ between 2.5 MeV–5.0 MeV, 0.25 MeV bins between 5.0 MeV–13.0 MeV, and a single
²⁴ energy bin between 13.0 MeV–14.0 MeV. Four equally-spaced bins were used in the
²⁵ parameter $r_3 = (r_{\text{reco}}/R_{\text{AV}})^3$ in the range $0 \text{ m} \leq r_{\text{reco}} < 5.0 \text{ m}$. The r_3 parameter
²⁶ was used to allow for equal volume weighting for each radial bin.

| Cut | # Events Remaining | Cut Efficiency (%) |
|---------------------------|--------------------|--------------------|
| Before cuts | 18,112,495,770 | 100.00 |
| Event triggers detector | 18,112,495,770 | 100.00 |
| Data Cleaning | 17,663,711,108 | 97.52 |
| high-nhit veto | 17,039,514,332 | 96.47 |
| ScintFitter used | 10,433,875,757 | 61.23 |
| ScintFitter results valid | 3,431,328,125 | 32.89 |
| Energy cut | 132,673 | 0.00387 |
| FV cut | 2093 | 1.58 |
| BiPo OOW cut | 817 | 39.03 |
| BiPo IW cut | 719 | 88.00 |
| Externals cut | 652 | 90.68 |

Table 6.2: Impact of each cut on the quantity of events in data.

| MC process | # Physics Events Simulated | # Triggered Events Simulated | # Remaining Events after cuts | Overall Cut Efficiency (%) |
|------------------------------------|-------------------------------|---------------------------------|----------------------------------|----------------------------|
| $^8\text{B } \nu_e$ | 2,830,425 | 5,511,568 | 929,551 | 32.8 |
| $^8\text{B } \nu_{\mu,\tau}$ | 1,898,088 | 3,675,261 | 572,949 | 30.2 |
| AV ^{214}Bi | 27,138,894 | 38,267,431 | 127 | 0.000468 |
| Ropes ^{214}Bi | 67,541,698 | 84,734,663 | 221 | 0.000327 |
| AV ^{208}Tl | 4,071,850 | 6,689,331 | 1471 | 0.0361 |
| Ropes ^{208}Tl | 6,105,956 | 8,758,457 | 2407 | 0.0394 |
| External Water ^{214}Bi | 81,300,835 | 85,056,233 | 62 | 0.0000763 |
| External Water ^{208}Tl | 36,602,319 | 38,963,663 | 2786 | 0.00761 |
| PMT $\beta - \gamma$ | 4,069,055 | 6,615,601 | 13,570 | 0.333 |
| Internal $^{212}\text{BiPo}$ | 9,761,554 | 24,331,685 | 8276 | 0.0848 |
| Internal ^{208}Tl | 816,233 | 1,626,536 | 409,629 | 50.2 |
| Internal $^{214}\text{BiPo}$ | 4,883,417 | 19,293,285 | 1030 | 0.0211 |
| Internal ^{210}Tl | 815,256 | 1,625,012 | 342,415 | 42.0 |
| Surface (α, n) (combined) | 22,627,103 | 48,834,413 | 2238 | 0.00458 |
| Internal (α, n) | 1,219,279 | 4,740,313 | 655 | 0.0537 |
| Internal ^{228}Ac | 9,764,480 | 19,294,643 | 0 | 0 |
| Internal ^{234m}Pa | 16,959,995 | 30,806,833 | 1 | 0.0000059 |

Table 6.3: Combined impact of cuts on each MC process. Overall cut efficiency is defined as the number of remaining triggered events after all cuts have been applied, divided by the number of physics events simulated for that process. Processes have been split into three general categories: signal, externals, and internals. A final category corresponds to various processes that have been considered for this analysis, but either all events are cut out, or Section 6.2.3 will show a negligible number of these events are expected in our dataset.

¹ 6.2.3 Expected Rates and their Constraints

² Because this analysis is dependent on deriving the shape and size of the ${}^8\text{B}$
³ spectrum from the data observed, it is important to have expectations about the
⁴ number of events expected to be seen within the dataset for each process, both
⁵ signal and background. This is done in two stages: firstly, the expected rates for
⁶ each process before any cuts applied is determined, along with any constraints.
⁷ Then, by using the cut efficiencies calculated in the previous section, we can derive
⁸ estimates for the expected number of events after cuts. When combined with a
⁹ constraining uncertainty, this can be used as a prior within the MCMC fit. The
¹⁰ resulting expected rates before and after cuts, along with any constraints being
¹¹ used, are shown in Table 6.5.

¹² ${}^8\text{B}$ Signal

¹³ Starting with the ${}^8\text{B}$ signal, numerous theoretical predictions and experimental
¹⁴ measurements have been made of the ${}^8\text{B}$ flux. For this analysis, two different
¹⁵ values have been used. The SSM predicts a relatively loose constraint of $\Phi_{{}^8\text{B}} =$
¹⁶ $(5.46 \pm 12\%) \times 10^6 \text{ cm}^{-2} \text{ s}^{-1}$ [74], whereas a recent global fit of neutrino oscillation
¹⁷ experiments by Bergström *et al* leads to a much stronger constraint of $\Phi_{{}^8\text{B}} =$
¹⁸ $(5.16^{+2.5\%}_{-1.7\%}) \times 10^6 \text{ cm}^{-2} \text{ s}^{-1}$ [77]. We shall mainly use the latter result to both
¹⁹ calculate the expected rate of the signal events in the detector, and the fractional
²⁰ uncertainties used to constrain this rate. The looser constraint coming from the
²¹ SSM will be used for comparison.

²² To calculate the number of electrons within the liquid scintillator, we follow
²³ the method used in [39], which uses the formula:

$$\text{24} \quad n_e = \frac{(f_{\text{LAB}} n_{\text{LAB}} + f_{\text{PPO}} n_{\text{PPO}}) N_A M_{\text{LAB}}}{m}. \quad (6.6)$$

Here, f_{LAB} and f_{PPO} are the fraction by weight of the LAB and PPO within the scintillator cocktail, respectively. Because this analysis is done over data taken during the scintillator phase with a PPO concentration of 2.2 g for every litre of LAB, these take values 99.744% and 0.256%, respectively. n_{LAB} and n_{PPO} are the mean number of electrons per molecule of LAB and PPO, respectively. PPO has the chemical formula C₁₅H₁₁NO, leading to $n_{\text{PPO}} = 116$. The LAB used in SNO+ has varying alkyl chain lengths, leading to a varying number of electrons per molecule. This distribution is known to have changed between batches of LAB made by the manufacturer, and is also impacted by the distillation process used during the purification of the LAB before it was put into the AV. At the time of writing, no final molecular breakdown has been made for the LAB within the detector; for now we shall use the breakdown provided here [78], from a representative tanker truck of LAB: there $n_{\text{LAB}} = 131.68$. Finally, $N_A = 6.02 \times 10^{23} \text{ mol}^{-1}$ is Avogadro's Constant, $M_{\text{LAB}} = 780.2 \text{ t}$ is the total mass of scintillator within a sphere of radius the size of the AV, and $m = 235 \text{ g mol}^{-1}$ is the molecular weight of the scintillator. This leads to a value for the number of electron targets:

$$n_e = 2.63 \times 10^{32} \text{ electrons.}$$

We can modify Eq. 6.1 to get an equation for the total rate of solar neutrino events by flavour, before oscillations or analysis cuts are considered:

$$R_i = \Phi_{^8\text{B}} n_e \int S_\nu(E_\nu) \sigma_{\nu_i}(E_\nu) dE_\nu. \quad (6.7)$$

Note that for $i = \mu, \tau$, this rate corresponds to the interaction rate if 100% of the neutrinos have flavour $\nu_{\mu, \tau}$.

Using the cross-section formula from Eq. 1.1, the ${}^8\text{B}$ spectral shape from [61], and $\Phi_{^8\text{B}} = 5.46 \times 10^6 \text{ cm}^{-2} \text{ s}^{-1}$ ($5.16 \times 10^6 \text{ cm}^{-2} \text{ s}^{-1}$), we expect a rate before

¹ considering oscillations or cuts of:

$$\text{2} \quad R_i = \begin{cases} 2743.2(2592.5) \text{ events/yr} & \text{for } i = e, \\ 489.7(462.8) \text{ events/yr} & \text{for } i = \mu, \tau. \end{cases}$$

³ In the above rates, we have included an additional volume correction factor of
⁴ 1.0139, because in MC events are simulated within the neck of the AV in addition
⁵ to the main spherical bulk [79].

⁶ Internal Uranium- and Thorium-Chain Backgrounds

⁷ As mentioned in Section 6.1.2, $^{214}\text{Bi-Po}$ and $^{212}\text{Bi-Po}$ decays, coming from the
⁸ ^{238}U and ^{232}Th decay chains respectively, are capable of generating distinctive
⁹ delayed coincidence events. Rafael Hunt-Stokes was able to use a series of cuts
¹⁰ to isolate both types of coincidence signals [65], looking at coincidence signals
¹¹ within the same 5.0 m FV as my analysis. After correcting for his cut efficiencies,
¹² he obtained a derived rate in the whole detector of [80]:

$$\text{13} \quad \text{BiPo rate} = \begin{cases} 0.94 \text{ events/hour} & \text{for } ^{212}\text{Bi-Po}, \\ 6.06 \text{ events/hour} & \text{for } ^{214}\text{Bi-Po}. \end{cases}$$

¹⁴ These rates assume a uniform concentration of ^{222}Rn and ^{220}Rn throughout the
¹⁵ detector: this is how internal backgrounds are simulated by default in RAT. However,
¹⁶ it has been shown that this is not at all the case within SNO+; radon ingress
¹⁷ from the AV neck propagates through the bulk of the detector via convection
¹⁸ currents induced by the thermal gradient present throughout the detector [81].
¹⁹ This radon ingress also breaks the secular equilibrium of the two decay chains, so
²⁰ that one can no longer make straightforward predictions about rates for decays in
²¹ the chains above radon. Fortunately for this analysis, Table 6.3 showed that the

cut efficiencies for ^{228}Ac and ^{234m}Pa decays, both in their respective decay chain above radon, are negligible. This means that even if their decay rates were much greater than what would be predicted from the tagged BiPo rates, a negligible number of events are expected for both in the livetime of the dataset considered in this analysis.

The combination of broad coincidence tagging as well as using an in-window BiPo classifier cut ensures that the expected rate of BiPo events that survive all the cuts is 1.5 and 2.5 for $^{212}\text{Bi}-\text{Po}$ and $^{214}\text{Bi}-\text{Po}$, respectively. Because the branching ratio of $^{212}\text{Bi} \rightarrow ^{208}\text{Tl}$ is 36%, and the branching ratio of $^{214}\text{Bi} \rightarrow ^{210}\text{Tl}$ is 0.021% [82, 83], after considering the cut efficiencies in this analysis we expect 512.4 internal ^{208}Tl and 1.1 internal ^{210}Tl events.

In theory, the non-uniformity of all the radon-induced backgrounds could impact the results of the oscillation analysis. However, the rates of each after cuts have been applied mean that only the internal ^{208}Tl could plausibly have any effect. For this particular background, each slice of the binned PDF in r_3 was allowed to float independently, with no constraints. The other three related backgrounds considered in the fit (^{210}Tl , $^{212}\text{Bi}-\text{Po}$, and $^{214}\text{Bi}-\text{Po}$) did not have this splitting applied to them. A loose 25% constraint on each of these three processes' rates was applied, to account for the current uncertainty in the calculation of R. Hunt-Stokes' tagging efficiency.

(α, n) Reactions

As discussed in Section 6.1.2, (α, n) reactions are induced by α -particles generated in the detector. In the SNO+ scintillator phase, the dominant source of α -particles are decays of ^{210}Po located both within the scintillator and on the AV surfaces. Because of substantial quenching by the scintillator, the reconstructed energy of the ^{210}Po α events fall substantially below E_{\min} . The rate of the internal and

¹ surface ^{210}Po events have been tracked throughout the scintillator phase by Serena
² Riccetto and Shengzhao Yu, respectively.

³ Within the 5 m FV, S. Riccetto measured a total of 1.60×10^8 ^{210}Po events
⁴ over the runs considered for this analysis [84, 85]. Assuming the internal rate of
⁵ these events are uniform throughout the scintillator, this leads to a total rate of
⁶ 2.78×10^8 events for all the scintillator. After using a conversion rate of 6×10^{-8}
⁷ neutrons generated per α on ^{13}C in liquid scintillator [8, 86], we can derive an
⁸ expected number of internal (α, n) events before cuts to be 17. However, because
⁹ a coincidence cut is used in this analysis, the vast majority of these should be
¹⁰ removed, leading to merely 0.009 events expected after cuts.

¹¹ Similarly, S. Yu measured the surface ^{210}Po rate to vary between $2\text{--}5 \text{Bq m}^{-2}$
¹² over the time period of this analysis as well as a function of height in the detec-
¹³ tor [87]. Using the midpoint value of 3.5Bq m^{-2} , one can derive an expected rate
¹⁴ of 554.4 surface (α, n) events occurring over the dataset's livetime, before cuts.
¹⁵ Once again, the coincidence cut as well as the FV cut removes the vast majority
¹⁶ of these events, so that only 0.07 events are expected after all cuts have been
¹⁷ applied. Because both classes of (α, n) events have negligible expected rates in
¹⁸ our dataset, they were not included within the MCMC fit.

¹⁹ External Backgrounds

²⁰ During the water phase of SNO+, an analysis of the rates of the external back-
²¹ grounds was performed by Tony Zummo [88]. The results of this analysis are
²² shown in Table 6.4, giving the measured rates as a fraction of the nominal values
²³ given in [68]. Although the statistical uncertainty for these measurements were
²⁴ quite small, the total systematic uncertainty was substantial because T. Zummo's
²⁵ analysis involved looking at events in the tail of energy distributions, so that any
²⁶ uncertainty in the energy scale systematic had an outsized impact on the number
²⁷ of events observed within that energy window. These measured rates and their

| Background Type | Rate (Fraction of Nominal) |
|-----------------|----------------------------------|
| AV & Ropes | $0.21 \pm 0.009^{+0.64}_{-0.21}$ |
| External Water | $0.44 \pm 0.003^{+0.32}_{-0.27}$ |
| PMTs | $1.48 \pm 0.002^{+1.65}_{-0.60}$ |

Table 6.4: Measured rates of the external backgrounds during the water phase of SNO+, by Tony Zummo [88].

systematic uncertainties were used to predict the expected rate and constrain the external backgrounds in this scintillator phase solar analysis. For the AV, ropes, and PMTs, this seems reasonable as we do not expect there to be any substantial change in these backgrounds between the phases. The exception to this is the external water, where various aspects of the water purification process over the years since T. Zummo’s analysis dataset was taken [89]. Because of this, the water phase measured rate was used as a starting point for the fit, but no constraint was applied.

Before the expected number of triggered events can be calculated for each external background process, two subtleties must be dealt with. Firstly, MC production of the externals did not include the hold-up ropes at the time of performing the analysis. This is not a major problem, because the hold-up ropes are of substantially lower mass than the hold-down ropes, and also their average radius from the centre of the detector is much larger. This means that the expected rate of events from the hold-up ropes that manage to reconstruct inside the FV should be sub-dominant to their hold-down counterparts. For the purposes of this thesis, the combined rate for both kinds of ropes were calculated, with the overall cut efficiency and derived PDFs coming from just the hold-down ropes.

Secondly, because the attenuation length of a 2.6 MeV γ particle is 23 cm in water [90, 91], the vast majority of external backgrounds which start from a substantial distance away from the AV do not generate γ s that make it into the scintillator. As a result, an enormous amount of computational resources could be wasted on simulating events that never get into the ROI. To work around this,

| MC process | Expected # Events Pre-Cuts | Overall Cut Efficiency (%) | Expected # Events Post-Cuts | Con- straint (%) |
|----------------------------------|-------------------------------|-------------------------------|--------------------------------|---------------------|
| $^8\text{B } \nu_e$ | 572.58 | 32.8 | 187.94 | +2.5% -1.7% |
| $^8\text{B } \nu_{\mu,\tau}$ | 102.22 | 30.2 | 30.86 | +2.5% -1.7% |
| AV ^{214}Bi | 594,000 | 0.000468 | 2.78 | +304.8% -100% |
| Ropes ^{214}Bi | 140,000 | 0.000327 | 0.46 | +304.8% -100% |
| AV ^{208}Tl | 69,600 | 0.0361 | 25.14 | +304.8% -100% |
| Ropes ^{208}Tl | 79,900 | 0.0394 | 31.48 | +304.8% -100% |
| External Water ^{214}Bi | 4,490,000 | 0.0000763 | 3.42 | — |
| External Water ^{208}Tl | 190,000 | 0.00761 | 14.46 | — |
| PMT $\beta - \gamma$ | 16,800 | 0.333 | 56.12 | +111.5% -40.5% |
| Internal $^{212}\text{Bi-Po}$ | 1814 | 0.0848 | 1.54 | $\pm 25\%$ |
| Internal ^{208}Tl | 1021 | 50.2 | 512.39 | — |
| Internal $^{214}\text{Bi-Po}$ | 11,728 | 0.0211 | 2.47 | $\pm 25\%$ |
| Internal ^{210}Tl | 2.5 | 42.0 | 1.05 | $\pm 25\%$ |
| Surface (α, n) (combined) | 554.4 | 0.00458 | 0.07 | — |
| Internal (α, n) | 17.0 | 0.0537 | 0.009 | — |
| Internal ^{228}Ac | 2835 | 0 | 0 | — |
| Internal ^{234m}Pa | 11,726 | 0.0000059 | 0.0007 | — |

Table 6.5: Number of events expected both before and after cuts have been applied, along with any constraints. As in Table 6.3, processes have been split into broad categories, with the last one being processes which have negligible rates after cuts. These processes are not included within the analysis fit.

¹ PMT $\beta - \gamma$ events are modelled as 2.6 MeV γ s generated at a radius of 6.2 m (just
² outside the AV), pointing radially inwards. Making some basic assumptions about
³ these events, one can derive the expected survival rate of these γ particles as a
⁴ function of radius [91, 92], which leads to a correction factor of 1.17×10^{-6} .

⁵ Similarly, simulations of background events in the external water and ropes are
⁶ restricted from starting beyond certain maximum radii depending on the specific
⁷ process, such that there remains a negligible impact on the fraction of simulated
⁸ events that actually deposit any energy into the scintillator. This leads to rate
⁹ correction factors of 0.35 and 0.50 for external water ^{214}Bi and ^{208}Tl events, as
¹⁰ well as 0.50 and 0.35 for hold-down and hold-up ropes [93].

6.2.4 Systematics

There are a number of systematic effects in the analysis that could possibly have some impact on the resulting posterior densities of θ_{12} and Δm_{21}^2 . Consideration of these effects is important to ensure that the uncertainty in the measurement is not underestimated. However, we shall find in Section 6.3 that for this dataset, the analysis is statistically-limited, and so it is not necessary to perform exhaustive measurements to determine the contribution of all possible systematic contributions. Instead, the focus shall be on the subset of the systematics that can plausibly have the most impact on the final measurement.

In this analysis, the measurement of the oscillation parameters is based on the fitted shape and normalisation of the ${}^8\text{B}$ signal energy spectrum. Therefore, for a systematic effect to have an impact on the final measurement it must first impact the fit of the signal PDF. One straightforward way this could occur is for some background events to be misattributed as signal events within the fit (or vice versa). If the rates of signal and background processes are strongly constrained to incorrect values, then this could easily happen. This could be a result of either getting the rate pre-cuts or the cut efficiencies incorrect. In this analysis, the strongest constraint on the rate of events comes from the global fit constraint of $\Phi_{{}^8\text{B}}$. Because of this, the fit will also be run with the looser SSM constraint on the signal flux as a means of comparison.

Another class of systematic effects that can plausibly have a substantial impact come from the mismodelling of the detector response, particularly on the reconstructed energy and radius, as they are the observables used within the fit. The most important of these for this analysis is a global mismodelling of the energy calibration by some linear factor known as the energy scale, α : $E_{\text{reco}} \rightarrow \alpha E_{\text{reco}}$.

Because of this importance, it is worthwhile constraining this α parameter. This can be done by comparing the reconstructed energy distributions of ${}^{214}\text{BiPo}$

¹ events tagged by R. Hunt-Stokes [80] in the dataset to the equivalent production
² MC, in the same 5 m FV. For a given α , the energies of the tagged ^{214}Bi in MC
³ were scaled by α , and then compared to the equivalent (unscaled) distribution in
⁴ data, via a log-likelihood ratio \mathcal{L}_{Bi} :

$$\mathcal{L}_{\text{Bi}} = \sum_{i=1}^{N_{\text{bins}}} n_i^{\text{data}} \ln \frac{n_i^{\text{MC}}}{N_{\text{MC}}}. \quad (6.8)$$

⁶ Here, n_i^{data} and n_i^{MC} are the number of events for data and (scaled) MC in energy
⁷ bin i , with N_{bins} being the total number of bins and N_{MC} being the total number of
⁸ tagged events in MC. To account for the edge effects discussed in Section 6.1.7, the
⁹ first and last bins in energy were used as buffers and not considered in calculation
¹⁰ of \mathcal{L}_{Bi} .

¹¹ Values of α were scanned over, generating a log-likelihood distribution. In
¹² the usual way, a constraint on α was obtained by obtaining the energy scaling
¹³ factor which maximises the log-likelihood, followed by looking for the α values
¹⁴ with values of \mathcal{L}_{Bi} less than the maximum by 1/2. The constraint was found to
¹⁵ be $\alpha = 0.9969 \pm 0.0022$, with the data and scaled MC at the best fit value of α
¹⁶ shown in Fig. 6.12.

¹⁷ To handle energy scaling as a systematic in this analysis, we allow the energy
¹⁸ scale to be floated within the MCMC fit, with the constraint given above. The
¹⁹ first and last energy bin of the MCMC fit were used as buffer bins to let this
²⁰ floating occur; this allows for the energy scale to float in the range $\frac{13}{14} \leq \alpha \leq \frac{2.6}{2.5}$.

²¹ In addition to a global energy scaling, it is possible that the resolution of the
²² energy reconstruction could be systematically off. Because of the effects discussed
²³ in Section 6.1.1, changes to the oscillation parameters only impact the observed
²⁴ signal distribution over a broad range of energies. This means that mismodelling
²⁵ the energy resolution of the detector is only liable to have an impact on the
²⁶ measurement if that mismodelling is substantial. However, by looking again at

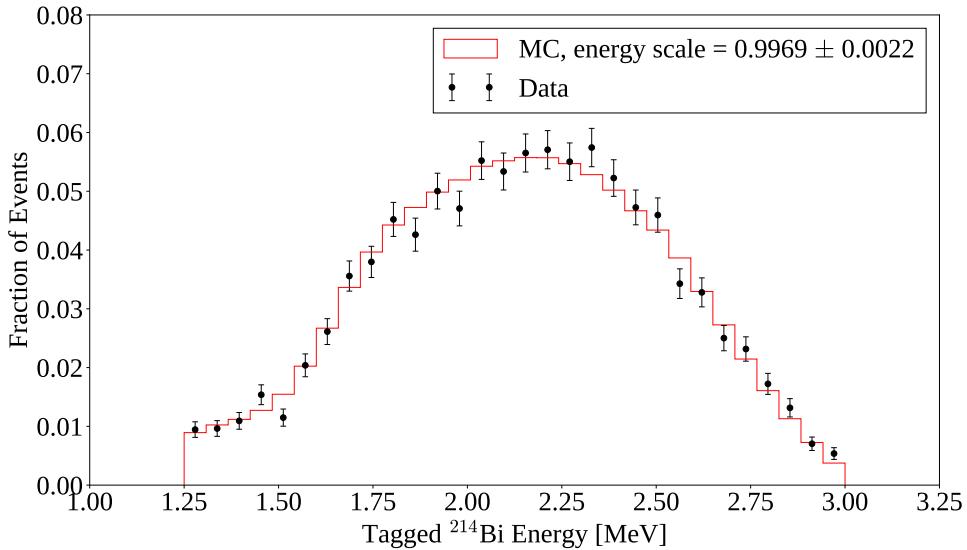


Fig. 6.12: Comparison of tagged ^{214}Bi reconstructed energy distributions between data and MC, after MC has been scaled by the best-fit factor $\alpha = 0.9969$.

Fig. 6.12 we see that there appears to be a good qualitative match of the energy resolution between data and MC for the tagged ^{214}Bi events. As a result, the systematic effect of additional energy smearing will not be considered in this thesis.

Similar to the systematics in reconstructed energy, there are also systematics possible in reconstructed position. These can be decomposed into three kinds of systematic: radial scaling, radial resolution, and position bias along each coordinate axis. The dominant effect that these systematics will have on the PDFs is to change their normalisations. Because we are floating all normalisations within the MCMC fit, and the normalisation constraints of almost all PDFs are relative large, we expect the impact of these systematics to be sub-dominant.

Finally, there can also be mismodelling of the radial distributions for given processes in MC compared to data. For example, it is known that the internal ^{208}Tl events are non-uniform as a function of radius; this is why the normalisation of each radial slice for this background is floated independently. The external backgrounds can also plausibly have PDF shape systematics, especially if the

1 assumptions made to model them (such as the use of simulation shells) are not
2 fully accurate. These shape systematics can also arise from the lack of statistics
3 in the PDFs used to model them: as an example, Table 6.3 shows that only
4 62 external water ^{214}Bi events are used in the creation of the associated PDF.
5 Handling this type of modelling uncertainties for the external backgrounds is not
6 considered in this thesis.

7 6.2.5 Results

8 Fit Validation

9 In theory, the convergence of the sampled distribution within an MCMC fit to
10 the posterior density distribution is guaranteed to eventually occur, regardless of
11 the specific shape of the multivariate Gaussian proposal distribution. However,
12 the speed of this convergence is highly dependent on the widths of the proposal
13 distribution. This is very important as in practice we can only run the MCMC fit
14 for a finite amount of time. For a given parameter within the fit, if the associated
15 width (also known as the step size) in the proposal distribution is too large, then
16 the vast majority of proposals will be to points in the parameter space where
17 the log-likelihood is less than before, and so these proposals will be consistently
18 rejected. Alternatively, if the width is chosen too small, the acceptance rate of
19 the chain will be very high, but too much time will be spent for the chain to go
20 from sampling one point in the space to another.

21 By choosing sensible width parameters for the proposal distribution, as well
22 as running the chain for as long as possible, one can attempt to maximise the
23 “effective sample size” of the MCMC fit. It has been shown that, for a wide variety
24 of situations, an appropriate size for the width of a given parameter is the standard
25 deviation of the true posterior density distribution [94]. Because knowing this
26 posterior density distribution is precisely the aim of the whole MCMC procedure,

this can make choosing these widths challenging in practice. In this analysis, an iterative approach was used: an MCMC fit was first run using width parameters that were guessed. Once this fit completed, the resulting sampled distribution marginalised onto each fit parameter were looked at to find approximately the magnitude of their standard deviations. This then informed the widths chosen for the next MCMC fit. This process was repeated until the autocorrelation of the sampled distribution for each fit parameter appeared to be minimal; this is an indication that the effective sample size has been maximised. A chain length of 1,000,000 steps was found to be sufficient to obtain large enough effective sample sizes for this analysis.

In order to guard against the possibility of a given chain missing some important part of the parameter space because of its starting position, 100 chains were run simultaneously. The initial positions of each in the parameter space were chosen randomly according to the overall prior distribution. Fig. 6.13 shows the sampled values for four different parameters in a given chain, after the tuning of the width parameters had been completed.

Given some initial start point in the parameter space, a given chain will typically take some time in finding where the region of greatest likelihood is. This leads to an initial set of steps where the chain moves a large distance in the parameter space in the same direction, so that the autocorrelation of the chain in this period is substantial. These initial samples are not representative of the actual posterior density distribution: as a result, in this analysis the first 100,000 steps of each chain were declared as a “burn-in” period. This number of steps for the burn-in was motivated in part by looking at auto-correlation plots for the MCMC chains, as seen in Fig. 6.14. By the time the number of steps differed by 100,000, the autocorrelation of parameters and the log-likelihood itself of each chain was at zero, indicating the likelihood evaluations in the chains were independent when they differed by at least 100,000 steps. The overall set

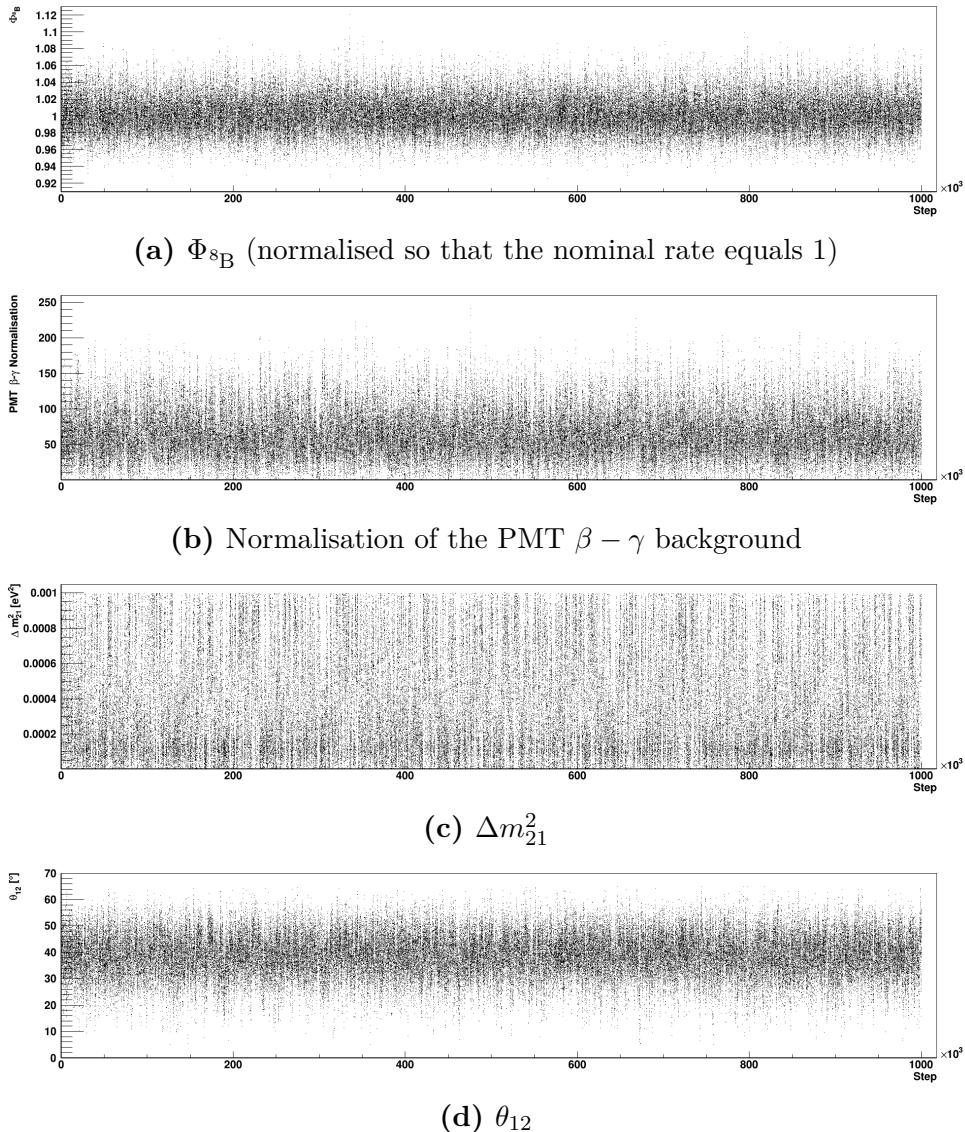


Fig. 6.13: Examples of the sampled values for parameters within a given MCMC chain as a function of the step number, after tuning the proposal distribution width parameters.

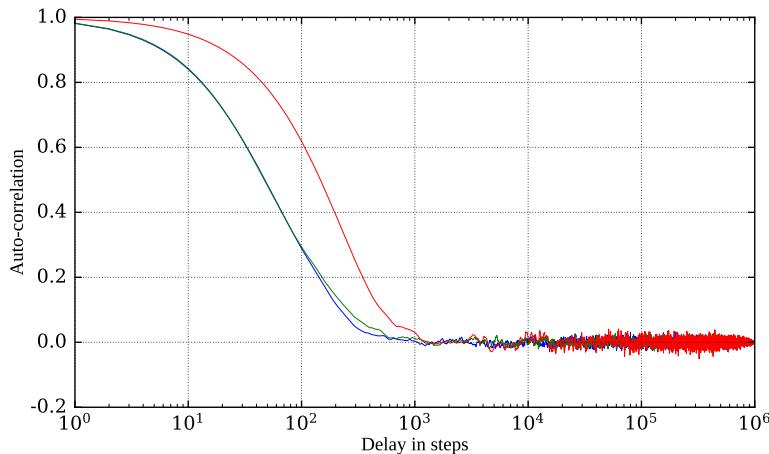


Fig. 6.14: Plot showing the auto-correlation as a function of step delay for three variables in a typical chain: log-likelihood (blue), Δm_{21}^2 (red), and θ_{12} (green).

of samples used to approximate the posterior density distribution corresponds to
1 collection of all sampled points on all the chains, after burn-in.
2

Oscillation Fit Results

Now that there is confidence in the convergence of the MCMC fit to the posterior
4 density distribution, we can look at this distribution to obtain the results of this
5 analysis. Fig. 6.15 shows the posterior density distribution, marginalised onto
6 each parameter of the fit other than the two oscillation parameters. We can see
7 that each marginalised distribution appears smooth, another qualitative indication
8 of a sufficient effective sample size in the MCMC fit. For each 1D distribution,
9 the bin centre associated with the HPD has been indicated, and the vertical band
10 indicates the Bayesian 1σ CI for that parameter. Also shown in these plots is the
11 prior distribution for each of the parameters, for comparison.
12

For a number of the parameters, the posterior distributions are more strongly
13 peaked than their associated priors. This indicates that there was enough infor-
14 mation in the dataset further constrain those parameters. One such parameter
15 is the rate of the external water ^{214}Bi . The resulting measurement of this is
16 $34.1^{+18.5}_{-18.9}$ events in the dataset. This is 10 times larger than the expected rate
17

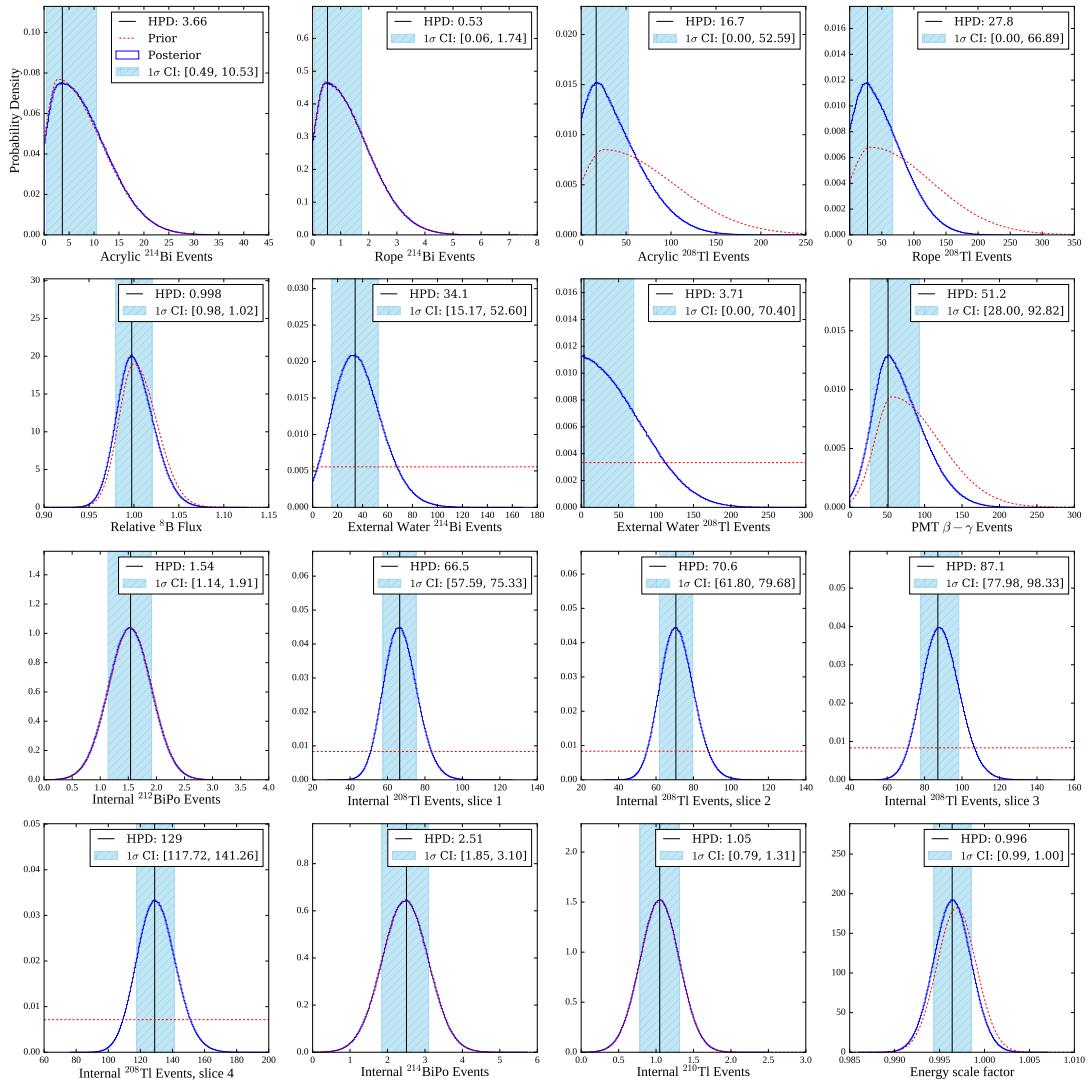


Fig. 6.15: Marginalised 1D posterior density distributions for all non-oscillation parameters, shown in blue. The HPD and 1 σ Bayesian CI are shown for each parameter. Also shown in dashed red are the associated prior distributions.

shown in Table 6.5, providing good evidence for a change in the level of this particular background relative to the water phase. In contrast, all other external backgrounds have HPDs consistent with their prior expected values.

Another set of parameters that have been well-constrained in the fit are the normalisations of each of the internal ^{208}Tl slices. As expected, slices corresponding to larger radii have a greater fitted rate, with the number of events measured in the outermost radial slice to be twice that of the innermost one. However, the posteriors of the other internal ^{238}U - and ^{232}Th -chain are the same as their priors, indicating there was no evidence from data within the fit to constrain those processes any further. Similarly, the posteriors of the energy scale parameter and Φ_{sB} only show very mild differences with their strong priors.

If instead of looking only at the 1D marginalised distributions of each parameter we look at the equivalent 2D distributions for every combination of parameters, we can generate a table of correlation coefficients. This is shown in Fig. 6.16. From this, we can see that there is some clear anti-correlation between a number of the external backgrounds. This seems intuitive: the PDFs for many of these external backgrounds are similar, and so the fit struggles to distinguish between them.

The only other pair of parameters which have a substantial correlation between one another are the two oscillation parameters. Fig. 6.17 shows their 2D marginalised posterior distribution, along with contours for a number of Bayesian CIs. Because the priors on both oscillation parameters are uniform, this can be directly converted into the log-likelihood ratio $\ln L_R$ for a given bin, where $L_R = p/p_{max}$: p is the posterior density of a given bin in the 2D space, and p_{max} is the posterior density of the HPD bin, equivalent to the bin with the maximum likelihood. This allows for Frequentist CI to be derived as well. For this plot, the HPD is at $\theta_{12} = 34.1^\circ$, $\Delta m_{21}^2 = 1.80 \times 10^{-5} \text{ eV}^2$.

As can be seen from this plot, the fit has been able to successfully constrain θ_{12} to some success. However, beyond a value of $5 \times 10^{-4} \text{ eV}^2$, the posterior density

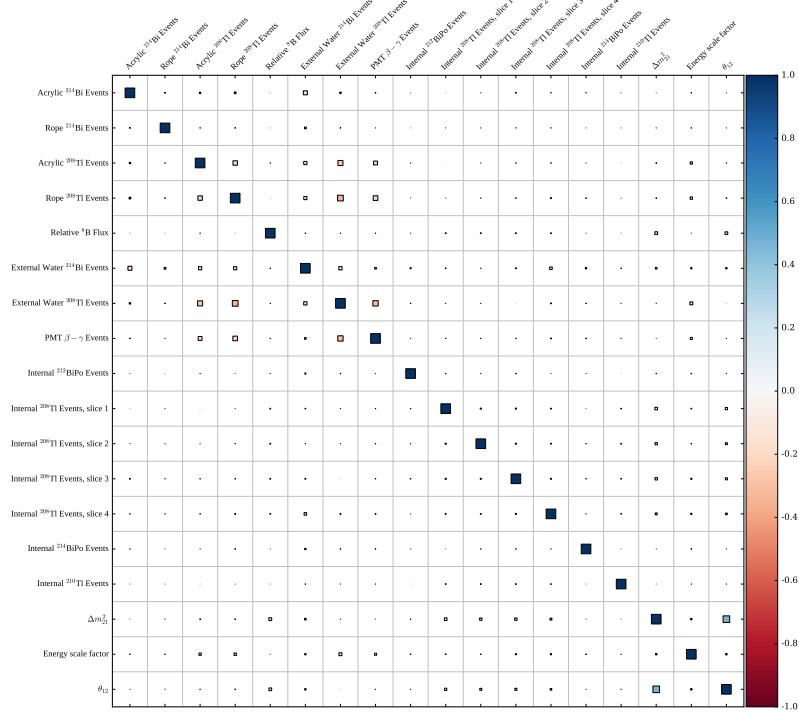


Fig. 6.16: Correlation matrix between all parameters in the MCMC fit.

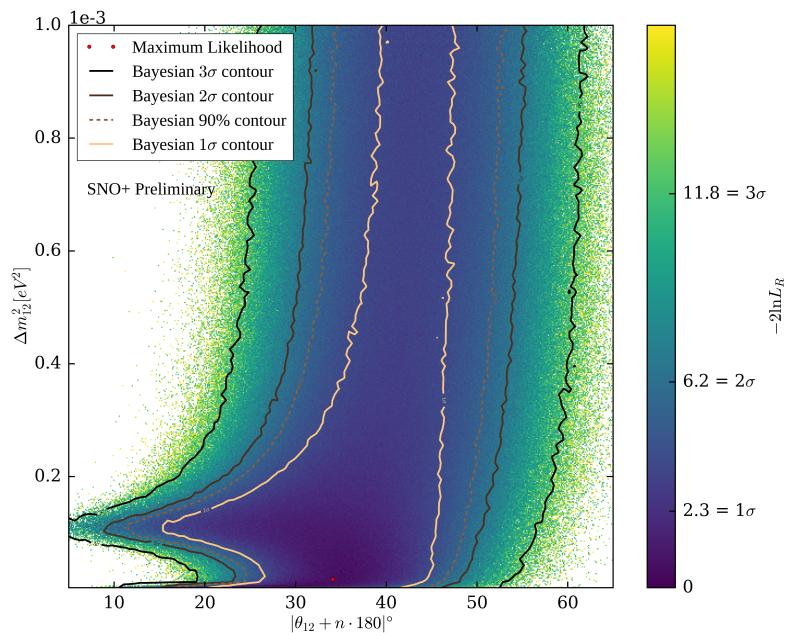


Fig. 6.17: 2D posterior density distribution marginalised onto Δm_{21}^2 and θ_{12} . The colour axis shows $-2 \ln L_R$, proportional to the natural logarithm of the posterior density, with labels corresponding to the values associated with the $N\sigma$ Frequentist CIs. Also shown, in shades of brown, are the Bayesian CIs.

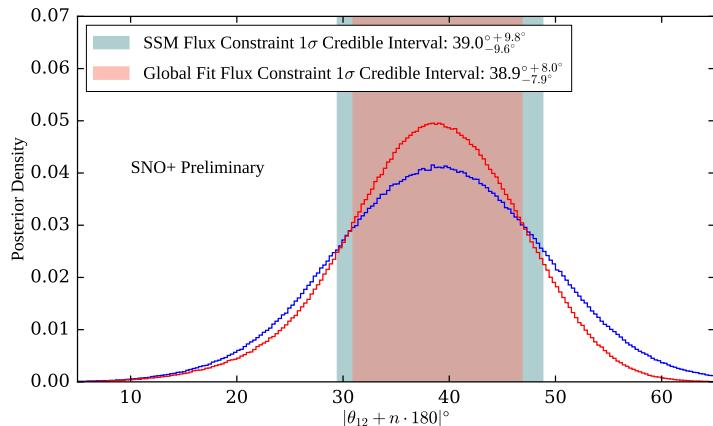


Fig. 6.18: Comparison of 1D posterior density distributions marginalised onto θ_{12} , using both forms of ${}^8\text{B}$ flux constraint. The 1σ Bayesian CI are shown for both.

appears flat as a function of Δm_{21}^2 . This implies that this dataset provides little ability to reject the possibility of large values of Δm_{21}^2 .

Given this lack of constraining power on Δm_{21}^2 , it is reasonable to only obtain a measurement for θ_{12} . The result of marginalising over Δm_{21}^2 to obtain the 1D posterior distribution in θ_{12} is shown in Fig. 6.18. From this, we can obtain a measurement of:

$$|\theta_{12} + n \cdot 180|^\circ = 38.9^{\circ+8.0}_{-7.9}.$$

Here, the HPD has been estimated by fitting a quadratic function to the peak of the posterior distribution, and the uncertainties by obtaining the Bayesian 1σ CI. For comparison, the NuFit 5.1 global fit results have [62]:

$$|\theta_{12} + n \cdot 180|^\circ = 33.44^{\circ+0.77}_{-0.74}.$$

As can be seen, the measurement made in this thesis is consistent with the global fit value, although with a substantially larger uncertainty. This is not surprising, given the limited statistics of the 80.6 days of data used.

The fact that the measured value of θ_{12} is somewhat above the global fit value also has a reasonable explanation. Looking back at Fig. 6.17, we find that the

¹ preferred values of θ_{12} for large Δm_{21}^2 are somewhat greater on average than for
² Δm_{21}^2 values close to the global fit value of 7.4×10^{-5} eV². As a result, the final
³ value quoted for θ_{12} is dependent on our choice of the prior distribution for Δm_{21}^2 :
⁴ this is an inevitable result of using a Bayesian framework for this analysis. We
⁵ can see the impact of choosing a less conservative prior for Δm_{21}^2 by ignoring all
⁶ sampled points with $\Delta m_{21}^2 \geq 5 \times 10^{-4}$ eV², say. The resulting measurement of θ_{12}
⁷ becomes:

$$\theta_{12} + n \cdot 180^\circ = 36.4^\circ_{-7.8^\circ}{}^{+8.0^\circ}.$$

⁹ Although there is negligible change to the uncertainty of the measurement, this
¹⁰ change in prior shifts the HPD value down by 2.5° , in the direction of the global
¹¹ fit value.

¹² Also shown in Fig. 6.17 is the posterior density for θ_{12} if the looser SSM flux
¹³ constraint is used. Under these conditions, the measurement becomes:

$$\theta_{12} + n \cdot 180^\circ = 39.0^\circ_{-9.6^\circ}{}^{+9.8^\circ}.$$

¹⁵ Using this constraint has little impact on the HPD value, but does increase
¹⁶ the uncertainty by $\sim 1.8^\circ$. This gives an indication of the extent to which the
¹⁷ strong global flux constraint allows θ_{12} to be constrained on its own, without any
¹⁸ information from the dataset.

¹⁹ One could naïvely expect that the point in parameter space where each fit
²⁰ parameter corresponds to the HPD value obtained in the above 1D marginalisation
²¹ would correspond to the maximum likelihood. This is not the case, because of
²² the correlations between parameters in the fit. Because an MCMC fit does not
²³ actually attempt to directly find this maximum likelihood point, deciding on the
²⁴ overall ‘best-fit’ point in parameter space is a little ambiguous. For this analysis,
²⁵ the sampled point in all of the chains which had the greatest log-likelihood

| Fit Parameter | 1D HPD value | Maximum likelihood value |
|---|--------------|--------------------------|
| Relative ${}^8\text{B}$ flux | 0.9975 | 1.006 |
| $\Delta m_{21}^2 [\times 10^{-5} \text{ eV}^2]$ | 11.02 | 5.50 |
| $\theta_{12} [^\circ]$ | 39.05 | 33.73 |
| AV ${}^{214}\text{Bi}$ Events | 3.66 | 8.49 |
| Ropes ${}^{214}\text{Bi}$ Events | 0.53 | 0.52 |
| AV ${}^{208}\text{Tl}$ Events | 16.71 | 19.08 |
| Ropes ${}^{208}\text{Tl}$ Events | 27.81 | 31.20 |
| External Water ${}^{214}\text{Bi}$ Events | 34.09 | 45.93 |
| External Water ${}^{208}\text{Tl}$ Events | 3.71 | 88.03 |
| PMT $\beta - \gamma$ Events | 51.22 | 46.88 |
| Internal ${}^{212}\text{BiPo}$ Events | 1.54 | 1.45 |
| Internal ${}^{208}\text{Tl}$ Events, Slice 1 | 66.52 | 68.78 |
| Internal ${}^{208}\text{Tl}$ Events, Slice 2 | 70.60 | 66.62 |
| Internal ${}^{208}\text{Tl}$ Events, Slice 3 | 87.09 | 87.33 |
| Internal ${}^{208}\text{Tl}$ Events, Slice 4 | 128.92 | 125.25 |
| Internal ${}^{214}\text{BiPo}$ Events | 2.51 | 2.56 |
| Internal ${}^{210}\text{Tl}$ Events | 1.05 | 1.00 |
| Energy Scale Factor, α | 0.9964 | 0.9970 |

Table 6.6: Comparison of the fit parameter values obtained when getting the HPD values after marginalising onto each parameter, versus looking at the sampled point with the maximal likelihood found.

was declared as the best-fit point. A comparison between the parameter values obtained at the HPD of the 1D marginalisations to the maximum likelihood point is shown in Table 6.6.

Using the maximal likelihood fit parameters, the comparison of data to MC is shown in Fig. 6.19. We see that the ${}^8\text{B}$ signal distribution only dominates above 5 MeV, as well as at low energies for the first two radial slices. This is because for larger radii, external backgrounds become the dominant processes in the 2.5–2.8 MeV region. Between 3–5 MeV, at all radii the internal ${}^{208}\text{Tl}$ shows a clear peak well above the ${}^8\text{B}$ signal. Qualitatively, the MC appears to give a good fit to the data.

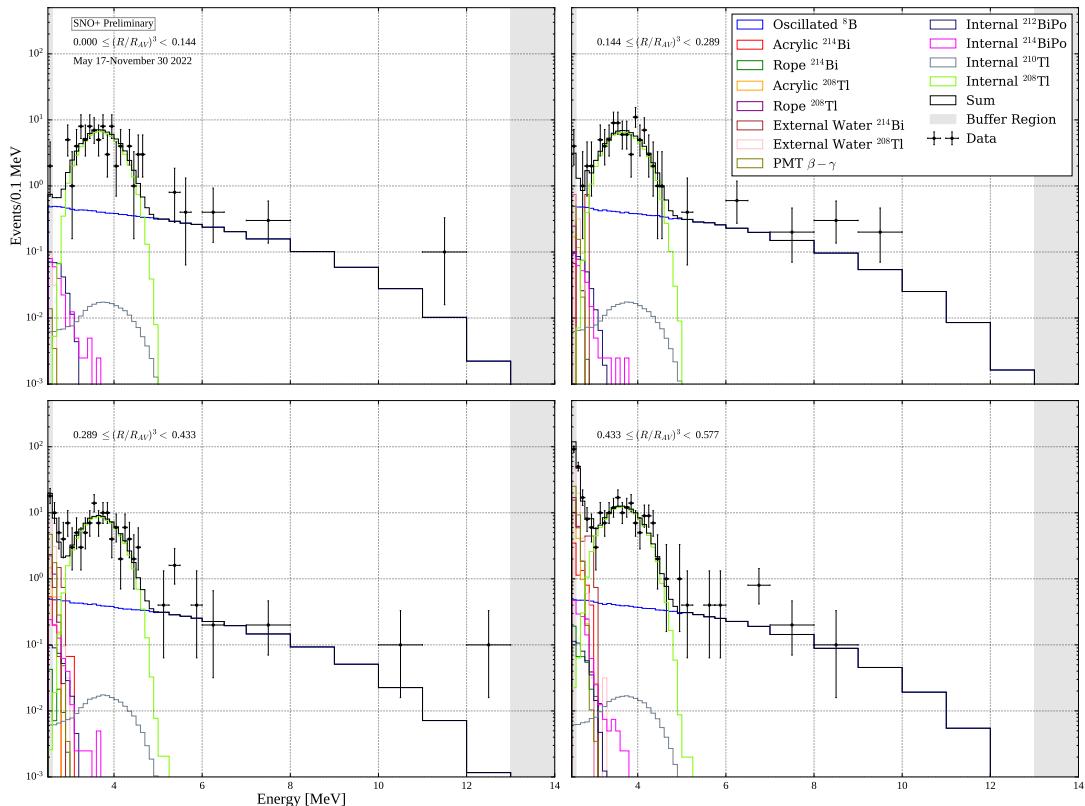


Fig. 6.19: Comparison of data to MC for each radial slice, using the best-fit parameter values derived from the MCMC results. The buffer bins used for varying the energy scale parameter within the fit are shown in grey.

6.3 Sensitivity Projections

Given the limited statistics used in the dataset analysed in Section 6.2, the question of how much better we could measure the solar oscillation parameters with more data naturally arises. In particular, it is worthwhile knowing the expected sensitivity of this analysis as a function of livetime, and hence whether SNO+ could eventually make a world-leading measurement of θ_{12} via ^8B neutrinos. Furthermore, it is useful to find out to what extent improvements such as additional background reduction could help the measurement.

To perform these sensitivity projections, the same analysis method was employed as with the real dataset. However, for each projection scenario a fake “Asimov” dataset was generated. This was a 2D histogram binned in the same way as the MC PDFs, corresponding to the total expected rate of all signal and background processes. Because the same PDFs were used to build both the MC and fake dataset, the maximum likelihood should occur when the fit parameters are identical to those which generated the fake dataset. Although this method of performing sensitivity projections leads to the slightly odd situation of having non-integer numbers of events in the bins of the fake dataset, it has been shown that this approach allows us to estimate the median sensitivity of a scenario [95].

The assumptions used to make the fake datasets are as follows. To begin with, no cuts were changed between the main analysis and this projection, and the same PDFs were used in both. Furthermore, the ^8B rate was determined by the same global fit flux value and neutrino oscillation parameters used in the rest of this analysis. Because some of the dataset used in the main analysis was taken during a time of known elevated ^{238}U - and ^{232}Th -chain backgrounds, R. Hunt-Stokes provided an additional estimate of the $^{214}\text{Bi-Po}$ and $^{212}\text{Bi-Po}$ rates in the 5.0 m FV for only the period of the dataset where the internal background levels had stabilised. Those rates went from 6.06 events per hour before excluding

the higher background period down to 4.87 events per hour for the ^{214}Bi –Po, and from 0.94 events per hour down to 0.89 events per hour. These relative changes for the two background chains were used to scale the rates from their expectation in the original dataset. For the internal ^{208}Tl , the rate of each slice relative to one another was taken from their fitted HPD values in the original analysis, with the absolute value of those rates scaled by 0.89/0.94.

External backgrounds were in general set to the rates expected from T. Zummo, as described in Section 6.2.3. The exceptions were the backgrounds from the external water: for the ^{208}Tl component, the nominal rate without T. Zummo’s correction factor was used. For the ^{214}Bi component, the higher rate as given by the fitted HPD value was used instead. The same constraints on all the rates were used as in the main analysis. For the energy scale parameter, because no additional systematics were applied to the Asimov dataset the true and expected value of α was set to 1, with the same uncertainty of ± 0.0022 as before.

Using this baseline set of assumptions, fake datasets were generated and fit over a series of livetime scenarios between 150 days and 5 years. The results of these MCMC fits, in terms of the 2D and 1D posterior densities for the oscillation parameters are shown in Figs. 6.20 and 6.21. Note that the length of burn-in was increased to 250,000 steps for these projections to ensure the convergence of all chains. As can be seen, by ~ 2 years of livetime, there is sufficient evidence to confidently reject large Δm_{21}^2 values above $2 \times 10^{-4} \text{ eV}^2$. This leads to a fairly substantial decrease in uncertainty as well as less bias in the 1D HPD value for θ_{12} . There does remain some bias, however: this is because of the non-Gaussian “boomerang” structure seen in the 2D posterior density plots.

The general story told by these projections is that, assuming the same conditions and analysis, increasing the livetime of the dataset will lead to a substantial improvement to the precision of the measurement of θ_{12} , as well as the beginnings

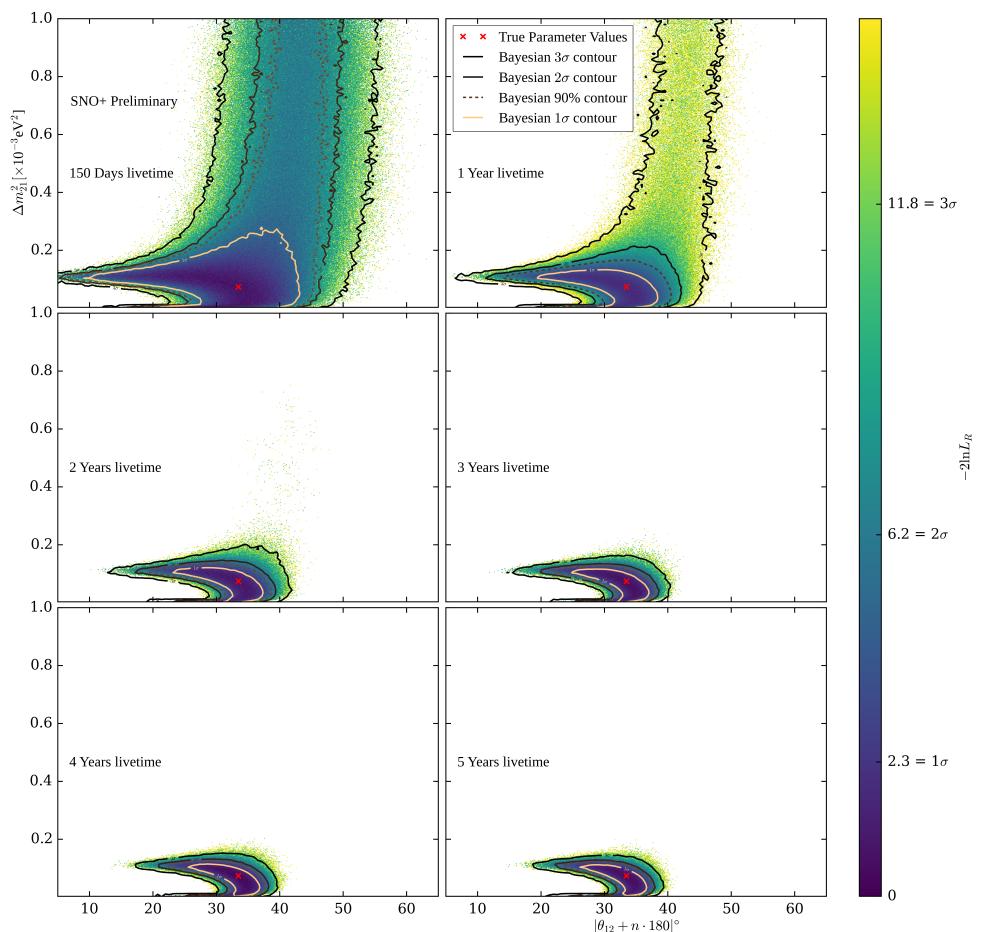


Fig. 6.20: Posterior density distributions marginalised onto the two solar oscillation parameters, for each livetime scenario in the pure scintillator phase under the current backgrounds.

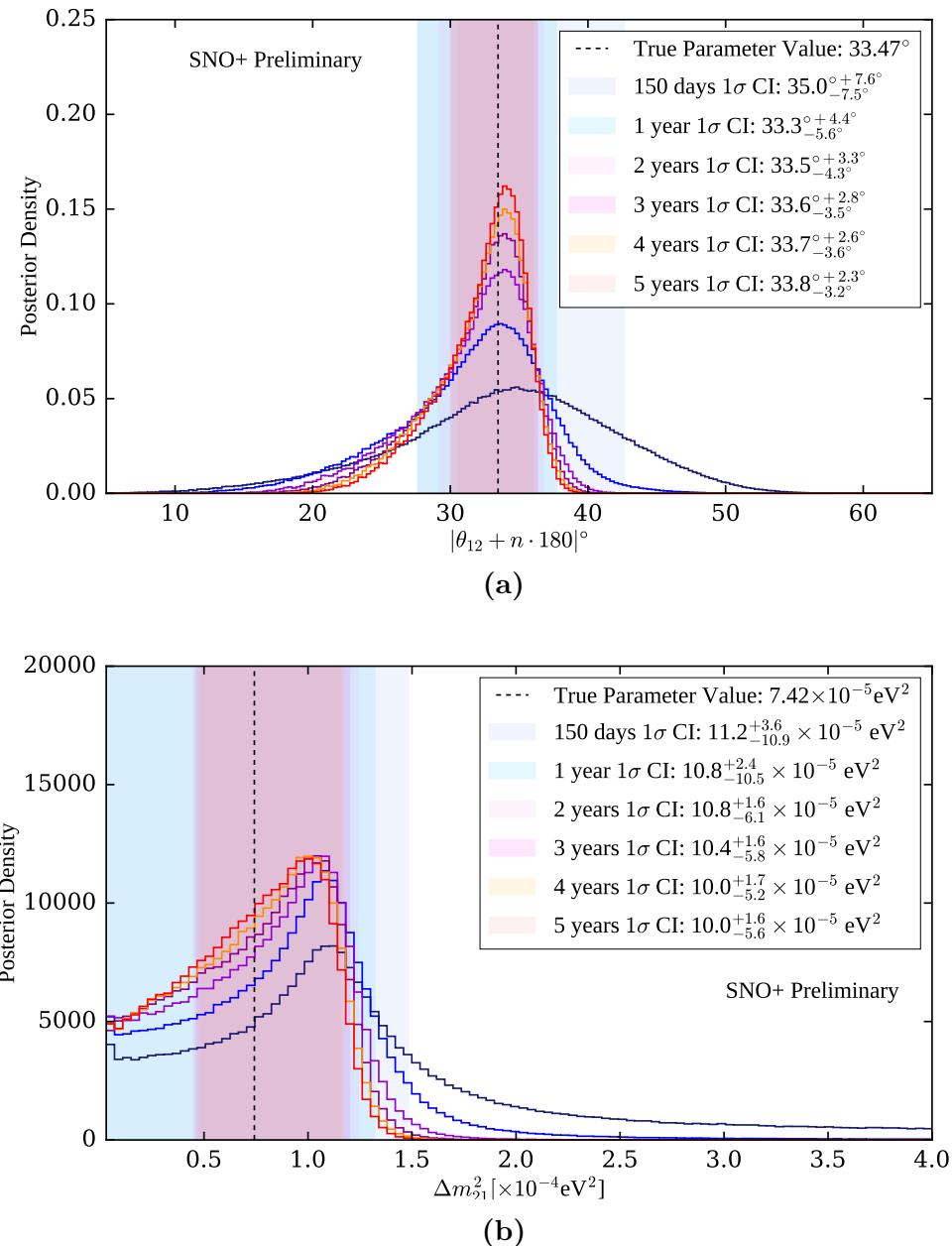


Fig. 6.21: Posterior density distributions marginalised onto each of θ_{12} and Δm^2_{21} , for each livetime scenario in the pure scintillator phase under the current backgrounds.

of a measurement of Δm_{21}^2 . However, after ~ 3 years of livetime the rate of progress slows down somewhat.

The choices of assumptions for this baseline set of projections were deliberately chosen to be fairly conservative. However, they do not account for the loading of BisMSB and TeLS into the detector that is likely to occur in the coming years. For BisMSB, the main impact is expected to be a dramatic increase in the observed light yield in the detector, leading to a smaller energy resolution and hence improved measurement [2]. Once the TeLS is also loaded, the observed light yield will likely decrease back down to levels similar to that of the current LABPPO. One of the biggest expected impact of the Te phase on this analysis will instead be to the background levels within the detector: the target rate of ^{238}U -chain backgrounds in the Te phase is ~ 100 times that of the scintillator phase [68].

Because of this, two further scenarios were considered over a period of 1 year of livetime. In one, the internal ^{238}U - and ^{232}Th -chain backgrounds were raised to the nominal level expected during the Te-loaded phase [55]. Unique to the Te phase will be a number of other background processes. These include $2\nu\beta\beta$ decay as well as cosmogenic isotopes created from the spallation of Te nuclei. Because the Q-value of $2\nu\beta\beta$ decay for ^{130}Te is 2.53 MeV, only a tiny fraction of those events should be expected to make it above this analysis' 2.5 MeV energy threshold. When the cuts described in Table 6.1 were run over the production of $2\nu\beta\beta$ MC made for the dataset used in the above analysis, no events survived. For the cosmogenics, assuming that the purification of the TeLS works as expected, less than one cosmogenic event of any kind is expected in the $0\nu\beta\beta$ ROI in a year [55]. Because of this, both of these additional backgrounds have been ignored for this scenario.

In the other background scenario, the external water backgrounds were set to the level measured by T. Zummo in the water phase, and the internal ^{208}Tl were

¹ reduced by a factor of 10. This latter scenario reflects the possibility of better
² control of the radioactivity in the external water, as well as a way of using new
³ analysis methods to cut out 90% of the internal ^{208}Tl events.

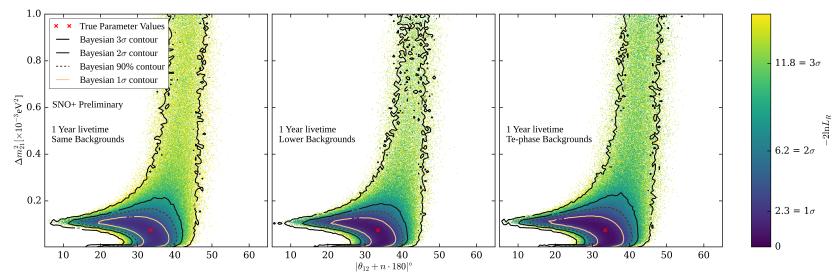
⁴ The results of running these two scenarios can be seen in Fig. 6.22. Unsur-
⁵ prisingly, the 1σ CI for θ_{12} is wider when backgrounds are greater, and thinner in
⁶ the lower background scenario. However, the magnitudes of the change in width
⁷ are somewhat different. For the high backgrounds case, the width only increases
⁸ by 2%. This implies that the higher internal backgrounds expected during the
⁹ Te phase should not, on their own, substantially hamper the sensitivity of the
¹⁰ measurement of θ_{12} . With lower external water and internal ^{208}Tl backgrounds, it
¹¹ appears that sensitivity improvements $\mathcal{O}(10\%)$ are achievable. This is because,
¹² over 1 year of livetime, the statistical uncertainties present in the signal process
¹³ still dominate.

¹⁴ 6.4 Summary and Suggestions for Further Work

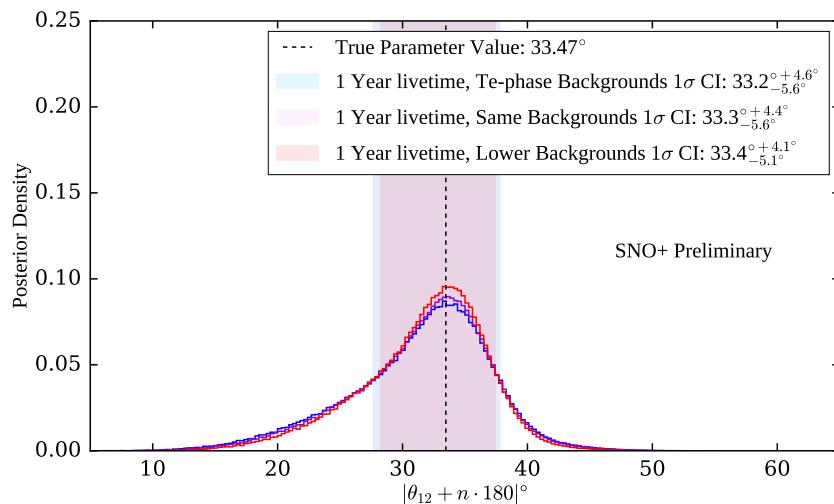
¹⁵ In this chapter, a full analysis for measuring the solar neutrino oscillation param-
¹⁶ eters via ^8B neutrinos was built and demonstrated on SNO+ scintillator-phase
¹⁷ data. Using 80.6 days of livetime, assuming the global fit constraint on the ^8B
¹⁸ flux of $\Phi_{^8\text{B}} = (5.16_{-1.7\%}^{+2.5\%}) \times 10^6 \text{ cm}^{-2} \text{ s}^{-1}$ [77], the oscillation parameter θ_{12} was
¹⁹ measured to be $|\theta_{12} + n \cdot 180|^\circ = 38.9_{-7.9}^{+8.0}{}^\circ$.

²⁰ Furthermore, assuming an identical analysis approach and similar detector
²¹ conditions to those seen in the above dataset, the precision of this measurement is
²² liable to improve by a factor of 2 in under 2 years of total livetime. This is partly
²³ because there become enough statistics to confidently reject large values of Δm_{21}^2
²⁴ above $\sim 1.5 \times 10^{-4} \text{ eV}^2$.

²⁵ Of course, major changes to the detector's scintillator cocktail are expected
²⁶ in that timeline. We have shown that the negative impacts expected from the



(a) Posterior densities marginalised onto the two oscillation parameters.

(b) Posterior densities marginalised onto θ_{12} .**Fig. 6.22:** Comparison of projections over 1 year of livetime, with different background expectations.

1 higher internal backgrounds of the Te-loaded phase should have minimal impact
2 on the sensitivity to the oscillation parameters over a livetime of 1 year. The
3 sensitivity obtained in this scenario is expected to be somewhat conservative, as the
4 scintillation emission time profile is expected to become substantially shorter [11],
5 leading to substantial improvements in the position resolution of the detector as
6 well as the discrimination power of various time-based classifiers.

7 If the external water backgrounds are able to return to the levels observed
8 during the water phase, and the event selection of internal ^{208}Tl events improves
9 substantially, there is some scope for improvement to the measurement of θ_{12}
10 when considering 1 year of livetime. There are three main possibilities for removal
11 of internal ^{208}Tl events. Firstly, these β -decays come as the delayed partner to
12 the α -decay branch of ^{214}Bi nuclei. Therefore, a coincidence tagging approach to
13 remove these events could be possible. Furthermore, alongside the β -decay of ^{208}Tl
14 are typically γ s, which should modify the observed time residual spectrum these
15 events compared to a single-site event like the signal process. As a result, creating
16 a custom multi-site classifier to exclude internal ^{208}Tl events from single-site
17 events could help further. Finally, solar neutrino and ^{208}Tl events can in theory
18 be distinguished by their direction. Via a method such as the one developed in [9],
19 it might be possible to gain some discrimination power of the solar signal over
20 this background.

21 On the subject of classifiers, the choice of cuts used for the externals classifiers
22 were chosen to be deliberately conservative. By tuning those cuts, it is possible
23 that the sensitivity could be improved through the increased rejection of external
24 events. An even more sophisticated approach would be to include these classifiers
25 as dimensions of the PDFs. This method yielded substantial improvements in the
26 $0\nu\beta\beta$ study performed in [55].

27 Of course, what has been seen in this analysis is that the greatest barrier to
28 greater precision is simply the signal statistics. This is a function of exposure,

not just livetime. Therefore, substantial improvements are likely to be made by
1 increasing the FV used for this analysis. For example, if the maximum radius used
2 in this analysis was moved from 5.0 m out to 5.7 m, then the increased volume
3 leads to an increased rate of signal events of 48%. Admittedly, for energies below
4 ~ 3.0 MeV at larger radii we expect external backgrounds to completely dominate
5 over the signal, so the benefits in the signal statistics will mostly be seen above
6 5 MeV.
7

Because the time of day impacts the extent to which solar neutrinos will be
8 travelling through the Earth, there is expected to be some additional power in
9 considering the difference between events observed during the day versus the
10 night. This approach was used to boost the expected sensitivity for the solar
11 oscillation analysis to be used on the JUNO experiment, for example [96, 97].
12 Also considered in that sensitivity study were additional interaction modes of
13 solar neutrinos with a liquid scintillator, such as the CC interaction of ν_e on ^{13}C
14 within the detector. These modes could provide further sensitivity.
15

Finally, analysis of neutrino oscillations from reactor anti-neutrinos in SNO+
16 is expected to lead to a world-leading limit of Δm_{21}^2 in under 4 years of livetime [8].
17 Although both analysis approaches have sensitivity to Δm_{21}^2 and θ_{12} , the reactor
18 anti-neutrino method is more sensitive to Δm_{21}^2 whilst the solar analysis shown
19 above is more sensitive to θ_{12} . Because of this complementarity, it is possible that
20 a combined analysis using both types of oscillating neutrino signal incident on
21 SNO+ will lead to even better measurements of both oscillation parameters.
22

Chapter 7

1

Conclusions

2

Say what has been achieved in this thesis! In particular:

3

- Substantial improvement to the SMELLIE generator in terms of speed and dynamic range 4
- A much stronger understanding of the discrepancies between data and MC in SMELLIE 6
- The creation of two analyses of SMELLIE data, designed explicitly around being robust to these systematics 7
- A measurement of the extinction length of scintillator *in-situ* with SMELLIE at 375 nm, monitored over time 9
- A first measurement of the scattering length of the scintillator *in-situ*, monitored over time 10
- The creation of an analysis of ^8B solar neutrinos in the scintillator phase to measure the solar neutrino oscillation parameters 11
- The first measurement of θ_{12} using ^8B neutrinos in SNO+ 12
- Projections of this solar analysis' precision at longer livetimes 13

14

15

16

17

¹ [3 PAGES TOTAL]

References

- [1] **SNO Collaboration**, J. Boger *et al.*, The Sudbury Neutrino Observatory, Nuclear Instruments and Methods in Physics Research Section A **449**, 172 (2000). 1
2
3
4
- [2] **SNO+ Collaboration**, V. Albanese *et al.*, The SNO+ experiment, Journal of Instrumentation **16**, P08059 (2021). 5
6
- [3] **SNO+ Collaboration**, M. Anderson *et al.*, Optical calibration of the SNO+ detector in the water phase with deployed sources, Journal of Instrumentation **16**, P10021 (2021). 7
8
9
- [4] **SNO+ Collaboration**, M. Anderson *et al.*, Measurement of the B 8 solar neutrino flux in SNO+ with very low backgrounds, Physical Review D **99**, 012012 (2019). 10
11
12
- [5] **SNO+ Collaboration**, A. Allega *et al.*, Evidence of Antineutrinos from Distant Reactors using Pure Water at SNO+, Physical Review Letters **130**, 091801 (2023), arxiv:2210.14154. 13
14
15
- [6] **SNO+ Collaboration**, M. Anderson *et al.*, Search for invisible modes of nucleon decay in water with the SNO+ detector, Physical Review D **99**, 032008 (2019). 16
17
18
- [7] **SNO+ Collaboration**, A. Allega *et al.*, Improved search for invisible modes of nucleon decay in water with the SNO+ detector, Physical Review D **105**, 112012 (2022), arxiv:2205.06400. 19
20
21
- [8] I. Morton-Blake, *First Measurement of Reactor Antineutrinos in Scintillator at SNO+ and Study of Alternative Designs for Large-Scale Liquid Scintillator Detectors*, DPhil Thesis, University of Oxford, 2021. 22
23
24
- [9] J. Paton, *First Observation of Directionality in the SNO+ Scintillator Phase*, DPhil Thesis, in preparation, University of Oxford. 25
26
- [10] **SNO+ Collaboration**, A. Allega *et al.*, First Observation of Directionality in a high Light Yield Scintillator Detector, in preparation. 27
28
- [11] D. J. Auty *et al.*, A Method to Load Tellurium in Liquid Scintillator for the Study of Neutrinoless Double Beta Decay, Nuclear Instruments and Methods in Physics Research Section A **1051**, 168204 (2023), arxiv:2212.12444. 29
30
31
- [12] I. Frank and Ig. Tamm, Coherent visible radiation of fast electrons passing through matter, Compt. Rend. Acad. Sci. URSS **14**, 109 (1937). 32
33

- 1 [13] T. Kaptanoglu, L. Segui, A. Mastbaum, N. Barros, and I. T. Coulter, Optics
2 Overview and Proposed Changes to RAT, 2016, SNO+ Internal Document
3 DocDB #3461.
- 4 [14] T. Kaptanoglu, Documentation for Attenuation Studies, 2022, SNO+ Internal
5 Document DocDB #7608.
- 6 [15] J. B. Birks, Chapter 3: The Scintillation Process in Organic Materials I,
7 in *The Theory and Practice of Scintillation Counting*, pp. 39–67, Pergamon
8 Press, , 2nd ed., 1967.
- 9 [16] R. Hunt-Stokes, Emission Timing Tuning - 2.2 g / L Optical Model, 2022,
10 SNO+ Internal Document DocDB #7621.
- 11 [17] A. LaTorre, New Measurements for the Timing Spectrum of Electrons and
12 Alphas in LAB+PPO, 2016, SNO+ Internal Document DocDB #4081.
- 13 [18] J. B. Birks, Chapter 6: The Scintillation Process in Organic Materials II, in
14 *The Theory and Practice of Scintillation Counting*, pp. 185–234, Pergamon
15 Press, , 2nd ed., 1967.
- 16 [19] F. Kondev, Nuclear Data Sheets for $A = 206$, Nuclear Data Sheets **109**,
17 1527 (2008).
- 18 [20] S. Riccetto, RAT optics 2.2g/L - after CM_Aug2022, 2022, SNO+ Internal
19 Document DocDB #7571.
- 20 [21] G. Mie, Contributions to the Optics of Turbid Media, particularly of Colloidal
21 Metal Solutions, Annalen der Physik **25**, 377 (1908).
- 22 [22] L. Lorenz, Sur la lumière réfléchie et réfractée par une sphère (surface)
23 transparente, Oeuvres scientifiques de L. Lorenz , 403 (1898).
- 24 [23] H. C. van de Hulst, *Light Scattering by Small Particles*Structure of Matter
25 Series (Dover Publications, New York, 1981).
- 26 [24] L. Rayleigh, On the transmission of light through an atmosphere containing
27 small particles in suspension, and on the origin of the blue of the sky, The
28 London, Edinburgh, and Dublin Philosophical Magazine and Journal of
29 Science **47**, 375 (1899).
- 30 [25] J. D. Jackson, Section 10.1: Scattering at Long Wavelengths, in *Classical
31 Electrodynamics*, Wiley, , 3rd ed., 1998.
- 32 [26] X. Zhou *et al.*, Rayleigh scattering of linear alkylbenzene in large liquid
33 scintillator detectors, Review of Scientific Instruments **86**, 073310 (2015),
34 arxiv:1504.00987.
- 35 [27] N. G. Jerlov and E. S. Nielsen, *Optical Aspects of Oceanography : [Collection
36 of the Papers Presented at the Symposium on Optical Aspects of Oceanography
37 ... Held at the Institute of Physical Oceanography in Copenhagen, 19th-23rd
38 June 1972]* (Academic Press / Symposium on Optical Aspects of Oceanogra-
39 phy (1972 : Copenhagen), London ; New York, 1974).

-
- [28] A. Einstein, Theory of Opalescence of Homogeneous Fluids and Liquid Mixtures near the Critical State, *Annalen der Physik* **338**, 1275 (1910). 1
2
- [29] M. S. Smoluchowski, Molecular kinetic theory of gases in critical opalescence state, as well as some related phenomena, *Annalen der Physik* , 205 (1908). 3
4
- [30] J. Cabannes, Relationship between the degree of Polarisation and the intensity of light Scattered by anisotropic molecules. New determination of the Avogadro constant, *Journal de Physique et Le Radium* **1**, 129 (1920). 5
6
7
- [31] J. F. Eykman, Recherches réfractométriques (suite), *Recueil des Travaux Chimiques des Pays-Bas* **14**, 185 (1895). 8
9
- [32] B. Moffat, *The Optical Calibration of the Sudbury Neutrino Observatory*, PhD thesis, Queen's University, Kingston, Canada, 2001. 10
11
- [33] E. Turner, *A Measurement of Scattering Characteristics of the Detection Medium in the SNO+ Detector*, DPhil Thesis, University of Oxford, 2022. 12
13
- [34] M. Chen and I. T. Coulter, Optical properties in RAT, 2012, SNO+ Internal Document DocDB #1664. 14
15
- [35] L. Segui, Scintillator Model: Comparison between new data and old model and its performance in RAT, 2015, SNO+ Internal Document DocDB #2774. 16
17
- [36] Y. Liu, Attenuation and Scattering of TeBD and the Cocktail, 2016, SNO+ Internal Document DocDB #3880. 18
19
- [37] Q. Liu *et al.*, Rayleigh scattering and depolarization ratio in linear alkylbenzene, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **795**, 284 (2015). 20
21
22
23
- [38] M. Yu *et al.*, Measurements of Rayleigh Ratios in Linear Alkylbenzene, arXiv:2203.03126 [hep-ex, physics:physics] (2022), arxiv:2203.03126. 24
25
- [39] A. S. Inácio, *Data Analysis of the Water and Scintillator Phases of SNO+: From Solar Neutrino Measurements to Double Beta Decay Sensitivity Studies*, PhD thesis, Universidade de Lisboa, 2022. 26
27
28
- [40] R. Boardman, *The Detection of Cherenkov Radiation from Neutrino Interactions*, DPhil Thesis, University of Oxford, 1992. 29
30
- [41] H. W. C. Tseung and N. Tolich, Ellipsometric measurements of the refractive indices of linear alkylbenzene and EJ-301 scintillators from 210 to 1000 nm, *Physica Scripta* **84**, 035701 (2011), arxiv:1105.2101. 31
32
33
- [42] E. Hecht, Section 4.6.2: The Fresnel Equations, in *Optics*, Pearson, , 4th ed., 2014. 34
35
- [43] M. E. Moorhead, *Reflectors in Cherenkov Detectors*, DPhil Thesis, University of Oxford, 1992. 36
37

- 1 [44] S. D. Biller, N. A. Jelley, M. D. Thorman, N. P. Fox, and T. H. Ward,
2 Measurements of photomultiplier single photon counting efficiency for the
3 Sudbury Neutrino Observatory, (1999).
- 4 [45] J. Dunger and E. Turner, Occupancy Analysis for Multiple-Hit Corrections
5 in Large Scintillation Detectors, 2016, NuPhys 2016 poster; SNO+ Internal
6 Document DocDB #4107.
- 7 [46] M. Howe *et al.*, Sudbury neutrino observatory neutral current detector
8 acquisition software overview, IEEE Transactions on Nuclear Science **51**, 878
9 (2004).
- 10 [47] J. C. Loach, *Measurement of the Flux of 8B Solar Neutrinos at the Sudbury*
11 *Neutrino Observatory*, DPhil Thesis, University of Oxford, 2008.
- 12 [48] M. Dragowsky *et al.*, The 16N calibration source for the Sudbury Neutrino
13 Observatory, Nuclear Instruments and Methods in Physics Research Section
14 A: Accelerators, Spectrometers, Detectors and Associated Equipment **481**,
15 284 (2002).
- 16 [49] S. Valder *et al.*, A Laserball Calibration Device for the SNO+ Scintillator
17 Phase, in preparation.
- 18 [50] P. G. Jones, *Background Rejection for the Neutrinoless Double Beta Decay*
19 *Experiment SNO+*, DPhil Thesis, University of Oxford, 2011.
- 20 [51] I. T. Coulter, *Modelling and Reconstruction of Events in SNO+ Related to*
21 *Future Searches for Lepton and Baryon Number Violation*, DPhil Thesis,
22 University of Oxford, 2013.
- 23 [52] W. Parker, MultiPDF Method for Position Reconstruction, 2021, SNO+
24 Internal Document DocDB #7257.
- 25 [53] W. Parker, Recoordinating ScintEffectiveSpeed & MultiPDF with Updated
26 2.2 g/l Optics, 2023, SNO+ Internal Document DocDB #7636.
- 27 [54] M. Mottram, Updated functional form energy reconstruction: EnergyRTheta-
28 Functional, 2015, SNO+ Internal Document DocDB #3488.
- 29 [55] T. Kroupová, *Improving the Sensitivity to Neutrinoless Double Beta Decay*
30 *in SNO+*, DPhil Thesis, University of Oxford, 2020.
- 31 [56] J. R. Sinclair, *Positioning and Timing Calibration of SNO+*, PhD thesis,
32 University of Sussex, 2015.
- 33 [57] K. Majumdar, *On the Measurement of Optical Scattering and Studies of*
34 *Background Rejection in the SNO+ Detector*, DPhil Thesis, University of
35 Oxford, 2015.
- 36 [58] S. S. Wilks, The Large-Sample Distribution of the Likelihood Ratio for
37 Testing Composite Hypotheses, The Annals of Mathematical Statistics **9**, 60
38 (1938).
- 39 [59] J. Caravaca, SNO+ Sensitivity to Standard Solar Neutrino Oscillations, 2020.

-
- [60] J. Dunger, *Topological and Time Based Event Classification for Neutrinoless Double Beta Decay in Liquid Scintillator*, DPhil Thesis, University of Oxford, 2018. 1
2
3
- [61] W. T. Winter, S. J. Freedman, K. E. Rehm, and J. P. Schiffer, The B8 Neutrino Spectrum, Physical Review C **73**, 025503 (2006), arxiv:nucl-ex/0406019. 4
5
- [62] I. Esteban, M. Gonzalez-Garcia, M. Maltoni, T. Schwetz, and A. Zhou, The fate of hints: Updated global analysis of three-flavor neutrino oscillations, Journal of High Energy Physics **2020**, 178 (2020). 6
7
8
- [63] J. N. Bahcall, M. Kamionkowski, and A. Sirlin, Solar neutrinos: Radiative corrections in neutrino-electron scattering experiments, Physical Review D **51**, 6146 (1995). 9
10
11
- [64] J. R. Wilson, BiPo Rejection Factors Summary, 2017, SNO+ Internal Document DocDB #4183. 12
13
- [65] R. Hunt-Stokes, Uranium and Thorium Background Analysis, 2022, SNO+ Internal Document DocDB #7664. 14
15
- [66] E. Marzec, BiPo Event Reduction using Likelihood Ratio Cut, 2013, SNO+ Internal Document DocDB #1994. 16
17
- [67] Z. Ye, Tagging in-window BiPo events with AlphaBetaClassifier – Recoordinated for 2.2 g/L LABPPO new optics, 2022, SNO+ Internal Document DocDB #7659. 18
19
20
- [68] **SNO+ Collaboration**, S. Andringa *et al.*, Current Status and Future Prospects of the SNO+ Experiment, Advances in High Energy Physics **2016**, 1 (2016). 21
22
23
- [69] W. Feller, *An Introduction to Probability Theory and Its Applications* Wiley Series in Probability and Mathematical Statistics. Probability and Mathematical Statistics, 3rd ed., rev. printing. ed. (Wiley, New York ; Chichester, 1968). 24
25
26
27
- [70] S. D. Biller and S. M. Oser, Another look at confidence intervals: Proposal for a more relevant and transparent approach, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **774**, 103 (2015). 28
29
30
31
- [71] C. Sanderson and R. Curtin, Armadillo: A template-based C++ library for linear algebra, Journal of Open Source Software **1**, 26 (2016). 32
33
- [72] C. Sanderson and R. Curtin, A User-Friendly Hybrid Sparse Matrix Class in C++, in *Mathematical Software – ICMS 2018*, edited by J. H. Davenport, M. Kauers, G. Labahn, and J. Urban, Lecture Notes in Computer Science, pp. 422–430, Cham, 2018, Springer International Publishing. 34
35
36
37
- [73] N. Barros, *Precision Measurement of Neutrino Oscillation Parameters: Combined Three-Phase Results of the Sudbury Neutrino Observatory*, PhD thesis, Universidade de Lisboa, 2011. 38
39
40

- ¹ [74] N. Vinyoles, A. Serenelli, and F. L. Villante, The B16 Standard Solar Models,
² Journal of Physics: Conference Series **1056**, 012058 (2018).
- ³ [75] A. M. Dziewonski and D. L. Anderson, Preliminary reference Earth model,
⁴ Physics of the Earth and Planetary Interiors **25**, 297 (1981).
- ⁵ [76] D. Cookman, Solar Oscillation Analysis Update, 2023, SNO+ Internal
⁶ Document DocDB #7707.
- ⁷ [77] J. Bergström *et al.*, Updated determination of the solar neutrino fluxes from
⁸ solar neutrino data, Journal of High Energy Physics **2016**, 132 (2016).
- ⁹ [78] M.-P. Lebeuf, LAB Certificate of Analysis, 2020, Example of document
¹⁰ provided by CEPSA for each tanker of LAB used to fill SNO+. SNO+ Internal
¹¹ Document DocDB #7572.
- ¹² [79] J. Caravaca, Validation of SNO+ Production Signal, 2019, SNO+ Internal
¹³ Document DocDB #5519.
- ¹⁴ [80] R. Hunt-Stokes, Private Communication, 2023.
- ¹⁵ [81] J. D. Wilson, Thermally-driven scintillator flow in the SNO+ neutrino
¹⁶ detector, Nuclear Instruments and Methods in Physics Research Section
¹⁷ A: Accelerators, Spectrometers, Detectors and Associated Equipment **1055**,
¹⁸ 168430 (2023).
- ¹⁹ [82] M. J. Martin, Nuclear Data Sheets for $A = 208$, Nuclear Data Sheets **108**,
²⁰ 1583 (2007).
- ²¹ [83] M. Shamsuzzoha Basunia, Nuclear Data Sheets for $A = 210$, Nuclear Data
²² Sheets **121**, 561 (2014).
- ²³ [84] S. Riccetto, Private Communication, 2023.
- ²⁴ [85] S. Riccetto, Full Fill Internal Po-210 Collaboration Meeting Update, 2023,
²⁵ SNO+ Internal Document DocDB #7733.
- ²⁶ [86] V. Lozza and O. Chkvorets, Neutron sources and backgrounds in SNO+,
²⁷ 2015, SNO+ Internal Document DocDB #2497.
- ²⁸ [87] S. Yu, AV Surface Po210 Analysis, 2023, SNO+ Internal Document DocDB
²⁹ #7733.
- ³⁰ [88] A. Zummo, External Background Box Analysis Plot/Results, 2022, SNO+
³¹ Internal Document DocDB #7329.
- ³² [89] T. Kaptanoglu, First look at external backgrounds in the scintillator phase,
³³ 2023, SNO+ Internal Document DocDB #7817.
- ³⁴ [90] M. J. Berger *et al.*, XCOM: Photon Cross Sections Database, NIST (2009).
- ³⁵ [91] W. J. Heintzelman, Survival to Radii Inside the PSUP of Γ s from PMT β - γ
³⁶ Events, 2012, SNO+ Internal Document DocDB #1729.

-
- [92] W. J. Heintzelman, Angular distribution of surviving Γ s from PMT β - γ events, 2013, SNO+ Internal Document DocDB #1712. 1
2
 - [93] A. S. Inácio, Using a Smaller Shell when Simulating Bi214 and Tl208 in the External Water and Ropes, 2019, SNO+ Internal Document DocDB #5154. 3
4
 - [94] A. Gelman *et al.*, Chapter: 11: Basics of Markov chain simulation, in *Bayesian Data Analysis*, , 3 ed., 2013. 5
6
 - [95] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, Asymptotic formulae for likelihood-based tests of new physics, *The European Physical Journal C* **71**, 1554 (2011), arxiv:1007.1727. 7
8
9
 - [96] JUNO Collaboration, A. Abusleme *et al.*, Feasibility and physics potential of detecting 8B solar neutrinos at JUNO, 2020, arxiv:2006.11760. 10
11
 - [97] JUNO Collaboration, J. Zhao *et al.*, Model Independent Approach of the JUNO 8B Solar Neutrino Program, 2022, arxiv:2210.08437. 12
13